International Journal of Biological, Life and Agricultural Sciences
ISSN: 2415-6612
Vol:2, No:3, 2008

Stochastic simulation of reaction-diffusion systems

Paola Lecca and Lorenzo Dematté

Abstract**—**Reactiondiffusion systems are mathematical models that describe how the concentration of one or more substances distributed in space changes under the influence of local chemical reactions in which the substances are converted into each other, and diffusion which causes the substances to spread out in space. The classical representation of a reaction-diffusion system is given by semi-linear parabolic partial differential equations, whose general form is $\partial_t \mathbf{X}(\mathbf{x},t) = \mathbf{D}\Delta\mathbf{X}(\mathbf{x},t)$, where $\mathbf{X}(\mathbf{x},t)$ is the state vector, $\mathbf{D}$ is the matrix of the diffusion coefficients and Δ is the Laplace operator. If the solute move in an homogeneous system in thermal equilibrium, the diffusion coefficients are constants that do not depend on the local concentration of solvent and of solutes and on local temperature of the medium.

In this paper a new stochastic reaction-diffusion model in which the diffusion coefficients are function of the local concentration, viscosity and frictional forces of solvent and solute is presented. Such a model provides a more realistic description of the molecular kinetics in non-homogenoeus and highly structured media as the intra- and inter-cellular spaces. The movement of a molecule A from a region i to a region j of the space is described as a first order reaction $A_i \xrightarrow{k} A_j$, where the rate constant k depends on the diffusion coefficient. Representing the diffusional motion as a chemical reaction allows to assimilate a reaction-diffusion system to a pure reaction system and to simulate it with Gillespie-inspired stochastic simulation algorithms. The stochastic time evolution of the system is given by the occurrence of diffusion events and chemical reaction events. At each time step an event (reaction or diffusion) is selected from a probability distribution of waiting times determined by the specific speed of reaction and diffusion events. *Redi* is the software tool, developed to implement the model of reaction-diffusion kinetics and dynamics. It is a free software, that can be downloaded from http://www.cosbi.eu.

To demonstrate the validity of the new reaction-diffusion model, the simulation results of the chaperone-assisted protein folding in cytoplasm obtained with *Redi* are reported. This case study is re-drawing the attention of the scientific community due to current interests on protein aggregation as a potential cause for neurodegenerative diseases.

Keywords**—**reaction-diffusion systems, Fick's law, stochastic simulation algorithm

I. Introduction

The great majority of mesoscopic reaction-diffusion models in intracellular kinetics is usually performed on the premise that diffusion is so fast that all concentrations are maintained homogeneous in space. However, recent experimental data on intracellular diffusion constants, indicate that this supposition is not necessarily valid even for small prokaryotic cells [1]. If the system is composed by a sufficiently large number of molecules, the concentration, i. e. the number of molecules per unit volume, becomes a continuum and differentiable variable of space and time. In this limit a reaction diffusion system can be modeled by using differential equations. In an unstructured solvent, ideally behaving solutes (i. e. solutes for which solute-solute interaction are negligible) obey the Fick's law of diffusion. However in biological system even for purely diffusive transport phenomena the classical Fick's diffusion is at best a first approximation [2], [3]. Spatial effects are present in many biological systems, so that the spatially homogeneous assumption will not always hold. Examples of spatial effects include mRNA movement within the cytoplasm [4], Ash 1 mRNA localization in budding yeast [5], morphogen gradients across egg-polarity genes in Drosophyla oocyte [5], and the synapse-specificity of long-term facilitation in Aplysia [6]. The intracellular medium is not a homogeneous mixture of chemical species, but a highly structured environment partitioned into compartments in which the distribution of the biomolecules could be non-homogeneous. The description of diffusion processes in this environment has to start from a model of the diffusion coefficient containing its dependency on the local concentrations of the solutes and solvent.

In order to tackle about this problem, the paper presents a new model of diffusion coefficient for a non-homogeneous non-well-stirred reaction-diffusion system. In this model the diffusion coefficient explicitly depends on the local concentration of solute, frictional coefficient and temperature. In turn, the rate of diffusion of the biochemical species are expressed in terms of this concentration-dependent diffusion coefficients. In this study purely diffusive transport phenomena of non-charged particles, and, in particular, the case in which the diffusion is driven by a chemical potential gradient in x direction only (the generalization to the three-dimensional case poses no problems) are considered. The derivation, introduced in this work, consists of five main steps: 1. calculation of the *local* virtual force F per molecules as the spatial derivative of the chemical potential 2. calculation of the particles mean drift velocity in terms of F and local frictional coefficient f; 3. estimation of the flux J as the product of the mean drift velocity and the local concentration; 4. definition of diffusion coefficients as function of local activity and frictional coefficients and concentration, and 5. calculation of diffusion rates as the negative first spatial derivative of the flux J. The determination of the activity coefficients has required the estimation of the second virial coefficient, that is calculated by using the Lennard-Jones potential to describe the inter-molecular interactions. The frictional coefficient is assumed to be linearly dependent on the local concentration of solute.

The diffusion events are modeled as reaction events and the spatial domain of the reaction chamber is divided into cubic subvolumes of size l, that from now on will be called indifferently *cells*, *meshes* or *boxes*. The movement of a molecule A from box i to box j is represented by the reaction $A_i \xrightarrow{k} A_j$, where A_i denotes the molecule A in the box i and A_j denotes the molecule A in the box j. The reaction-

Authors affiliation: The Microsoft Research - University of Trento, CoSBi, Trento (Italy). E-mail:{lecca, dematte, priami}@cosbi.eu

diffusion system is thus modeled as a purely reaction system in which the diffusion events are first order reactions whose rate coefficients ks are expressed in terms of state-dependent diffusion coefficients.

The space domain of the system is divided into a number N_s of subvolumes. The time evolution of the system is computed by a Gillespie-like algorithm [7] that at each simulation step selects in each subvolume the fastest reaction, compares the velocities of the N_s selected reactions and finally executes the reaction that is by far the fastest. To make the Gillespie approach applicable in each subvolumes, the size of the mesh has to be chosen sufficiently small so that the homogeneity and well-stirred assumption on the distribution of the molecules inside are good approximations, and sufficiently large to have a number of reaction events significantly greater than one.

II. The Model of Diffusion: Generalization of the Fick's Law

The Gibbs energy difference between regions of different concentration, i. e. the gradient of the chemical potential μ, causes diffusive transport of molecules. Let consider a solution containing N different solutes. The chemical potential μ_i of any particular chemical species i is defined as the partial derivative of the Gibbs energy G with respect to the concentration of the species i, with temperature and pressure held constant. Species are in equilibrium if their chemical potentials are equal.

$$\mu_i \equiv \frac{\partial G}{\partial c_i} = \mu_i^0 + RT \ln a_i \tag{1}$$

where c_i is the concentration of the species i, μ_i^0 is the standard chemical potential of the species i (i .e. the Gibbs energy of 1 mol of i at a pressure of 1 bar), $R = 8.314 \text{ J} \cdot \text{K}^{-1} \cdot \text{mol}^{-1}$ is the ideal gas constant, and T the absolute temperature. The quantity a_i is called *chemical activity* of component i, and it is given by

$$a_i = \frac{\gamma_i c_i}{c^0} \tag{2}$$

where γ_i is the *activity coefficient*, c^0 being a reference concentration, which, for example, could be set equal to the initial concentration. The activity coefficients express a deviation of a solution from the ideal thermodynamic behavior and in general they may depend on the concentration of all the solutes in the system. For an ideal solution, the limit of γ_i which is recovered experimentally at high dilutions is $\gamma_i = 1$. If the concentration of species i varies from point to point in space, then so does the chemical potential. For simplicity, here the case in which there is only a chemical potential gradient in the x direction only is taken into account. Chemical potential is the free energy per mole of substance, free energy is the negative of the work W which a system can perform, and work is connected to force F acting on the molecules by $dW = Fdx$. Therefore an inhomogeneous chemical potential is related to a virtual force per molecule of

$$F_i = -\frac{1}{N_A}\frac{d\mu_i}{dx} = -\frac{k_B T c^0}{\gamma_i c_i} \sum_j \frac{\partial a_i}{\partial c_j}\frac{\partial c_j}{\partial x} \tag{3}$$

where $N_A = 6.022 \times 10^{23} \text{ mol}^{-1}$ is the Avogadro's number, $k_B = 1.381 \times 10^{-23} \text{ J} \cdot \text{K}^{-1}$ is the Boltzmann's constant, and the sum is taken over all species in the system other than the solvent. This force is balanced by the drag force experienced by the solute ($F_{drag,i}$) as it moves through the solvent. Drag forces are proportional to the speed. If the speed of the solute is not too high in such a way that the solvent does not exhibit turbulence, the drag force can be written as follows

$$F_{drag,i} = f_i v_i \tag{4}$$

where $f_i \propto c_i$ is the frictional coefficient, and v_i is the mean drift speed.

Moreover, if the solvent is not turbulent, the *flux*, defined as the number of moles of solute which pass through a small surface per unit time per unit area, can be approximated as in the following

$$J_i = c_i v_i \tag{5}$$

i. e. the number of molecules per unit volume multiplied by the linear distance travelled per unit time.

Since the virtual force on the solute is balanced by the drag force (i. e. $F_{drag,i} = -F_i$), the following expression for the mean drift velocity is obtained

$$v_i = \frac{F_i}{f_i}$$

so that Eq. (5) becomes

$$J_i = -\frac{k_B T}{\gamma_i f_i} \sum_j \frac{\partial a_i}{\partial c_j}\frac{\partial c_j}{\partial x} \equiv -\sum_j D_{ij}\frac{\partial c_j}{\partial x} \tag{6}$$

where

$$D_{ij} = \frac{k_B T c^0}{\gamma_i f_i}\frac{\partial a_i}{\partial c_j} \tag{7}$$

are the diffusion coefficients. The Eq. (7) states that, in general, the flux of one species depends on the gradients of all the others, and not only on its own gradient. However, here it is supposed that the chemical activity a_i depends only weakly on the concentrations of the other solutes, i. e. it is assumed that $D_{ij} \approx 0$ for $i \neq j$ and the Fick's laws still holds. Let D_i denote D_{ii}. It is still generally the case that D_i depends on c_i in sufficiently concentrated solutions since γ_i (and thus a_i) has a non trivial dependence on c_i [8]. It is only in one very special case, namely that of an ideal solution with $\gamma_i = 1$, where the diffusion coefficient, $D_i = k_B T / f_i$, is constant. In order to find an analytic expression of the diffusion coefficients D_i in terms of the concentration c_i, let us consider that the rate of change of concentration of the substance i due to diffusion is given by

$$\mathscr{D}_i = -\frac{\partial J_i}{\partial x} \tag{8}$$

Substituting Eq. (7) into Eq. (6), and then substituting the obtained expression for J_i into Eq. (8), gives

$$\mathscr{D}_i = -\frac{\partial}{\partial x}\Big(- D_i(c_i)\frac{\partial c_i}{\partial x}\Big) \tag{9}$$

so that

$$\mathscr{D}_i = \left(\frac{\partial D_i(c_i)}{\partial x}\right)\frac{\partial c_i}{\partial x} + D_i(c_i)\frac{\partial^2 c_i}{\partial x^2} = \\ = \frac{\partial D_i(c_i)}{\partial c_j}\frac{\partial c_j}{\partial x}\frac{\partial c_i}{\partial x} + D_i(c_i)\frac{\partial^2 c_i}{\partial x^2} \quad (10)$$

Let $c_{i,k}$ denote the concentration of a substance i at coordinate x_k, and $l = x_k - x_{k-1}$ the distance between adjacent mesh points. The derivative of c_i with respect to x calculate in $x_{k-\frac{1}{2}}$ is

$$\frac{\partial c_i}{\partial x}\Big|_{x_{k-\frac{1}{2}}} \approx \frac{c_{i,k} - c_{i,k-1}}{l} \quad (11)$$

By using Eq. (11) into Eq. (6) the diffusive flux of species i midway between the mesh points $J_{i,k-\frac{1}{2}}$ is obtained:

$$J_{i,k-\frac{1}{2}} = -D_{i,k-\frac{1}{2}}\frac{c_{i,k} - c_{i,k-1}}{l} \quad (12)$$

where $D_{i,k-\frac{1}{2}}$ is the diffusion coefficient midway between the mesh points.

The rate of diffusion of substance i at the mesh point k is

$$\mathscr{D}_{ik} = -\frac{J_{i,k+\frac{1}{2}} - J_{i,k-\frac{1}{2}}}{l}$$

and thence

$$\mathscr{D}_{ik} = \frac{D_{i,k-\frac{1}{2}}}{l^2}(c_{i,k-1} - c_{i,k}) - \frac{D_{i,k+\frac{1}{2}}}{l^2}(c_{i,k+1} - c_{i,k}) \quad (13)$$

To determine completely the right-hand side of Eq. (13) is now necessary to find an expression for the activity coefficient γ_i and the frictional coefficient f_i, contained in the formula (7) for the diffusion coefficient. In fact, by substituting Eq. (2) into Eq. (7) an expression of the diffusion coefficient in terms of activity coefficients γ_i is obtained

$$D_{ii} = \frac{k_B T}{f_i}\Big(1 + \frac{c_i}{\gamma_i}\frac{\partial \gamma_i}{\partial c_i}\Big) \quad (14)$$

Let focus now on the calculation of the activity coefficients, while a way to estimate the frictional coefficients will be presented in Section II-A. By using the subscript '1' to denote the solvent and '2' to denote the solute, it can be written that

$$\mu_2 = \mu_2^0 + RT\ln\left(\frac{\gamma_2 c_2}{c^0}\right) \quad (15)$$

where γ_2 is the activity coefficient of the solute and c_2 is the concentration of the solute. Differentiating with respect to c_2 gives

$$\frac{\partial \mu_2}{\partial c_2} = RT\Big(\frac{1}{c_2} + \frac{1}{\gamma_2}\frac{\partial \gamma_2}{\partial c_2}\Big) \quad (16)$$

The chemical potential of the solvent is related to the osmotic pressure (Π) by

$$\mu_1 = \mu_1^0 - \Pi V_1 \quad (17)$$

where V_1 is the partial molar volume of the solvent and μ_1^0 its standard chemical potential. Assuming V_1 to be constant and differentiating μ_1 with respect to c_2 yield

$$\frac{\partial \mu_1}{\partial c_2} = -V_1\frac{\partial \Pi}{\partial c_2} \quad (18)$$

Now, from the Gibbs-Duhem relation [9], the derivative of the chemical potential of the solute with respect to the solute concentration is

$$\frac{\partial \mu_2}{\partial c_2} = -\frac{M(1 - c_2\overline{v})}{V_1 c_2}\frac{\partial \mu_1}{\partial c_2} = \frac{M(1 - c_2\overline{v})}{c_2}\frac{\partial \Pi}{\partial c_2} \quad (19)$$

where M is molecular weight of the solute and $\overline{v}$ is the partial molar volume of the solute divided by its molecular weight. The concentration dependence of osmotic pressure is usually written as

$$\frac{\Pi}{c_2} = \frac{RT}{M}\Big[1 + BMc_2 + O(c_2^2)\Big] \quad (20)$$

where B is the second virial coefficient (see Section II-B), and thence the derivative of Π with respect to the solute concentration is

$$\frac{\partial \Pi}{\partial c_2} = \frac{RT}{M} + 2RTBc_2 + O(c_2^2) \quad (21)$$

Introducing Eq. (21) into Eq. (19) gives

$$\frac{\partial \mu_2}{\partial c_2} = RT(1 - c_2\overline{v})\Big(\frac{1}{c_2} + 2BM\Big) \quad (22)$$

From Eq. (16) and Eq. (22) it can be obtained that

$$\frac{1}{\gamma_2}\frac{\partial \gamma_2}{\partial c_2} = \frac{1}{c_2}\Big[(1 - c_2\overline{v})(1 + 2BMc_2) - 1\Big]$$

so that

$$\int_1^{\gamma_2'}\frac{d\gamma_2}{\gamma_2} = \int_{c^0}^{c_2'}\frac{1}{c_2}\Big[(1 - c_2\overline{v})(1 + 2BMc_2) - 1\Big]dc_2$$

On the grounds that $c_2\overline{v} \ll 1$ [10], solving the integral yields

$$\gamma_2' = \exp[2BM(c_2' - c^0)] \quad (23)$$

The molecular weight $M_{i,k}$ of the species i in the mesh k can be expressed as the ratio between the mass $m_{i,k}$ of the species i in that mesh and the Avogadro's number $M_{i,k} = m_{i,k}/N_A$. If p_i is the mass of a molecule of species i and $c_{i,k}l$ is the number of molecules of species i in the mesh k, then the molecular weight of the solute of species i in the mesh k is given by

$$M_{i,k} = \frac{p_i}{N_A}\frac{l}{}c_{i,k} \quad (24)$$

Substituting this expression in Eq. (23) gives for the activity coefficient of the solute of species i in the mesh k ($\gamma_{i,k}$), the following equation

$$\gamma_{i,k} = \exp\Big(2B\ \frac{p_i}{N_A}\ \frac{l}{}\ c_{i,k}^2\Big) \quad (25)$$

A. Intrinsic viscosity and frictional coefficient

The diffusion coefficient depends on the ease with which the solute molecules can move. It is a measure of how readily a solute molecule can push aside its neighboring molecules of solvent. An important aspect of the theory of diffusion is how the magnitude of the frictional coefficient f_i of a solute of species i and, hence, of the diffusion coefficient D_i, depend on the properties of the solute and solvent molecules. Examination of well-established experimental data shows that diffusion coefficients tend to decrease as the molecular size of the solute increases. The reason is that a larger solute molecule has to push aside more solvent molecules during its progress and will therefore move slowly than a smaller molecule. A precise theory of the frictional coefficients for the diffusion phenomena in biological context cannot be simply derived from the elementary assumptions and model of the kinetic theory of gases and liquids. The Stokes's theory considers a simple situation in which the solute molecules are so much larger than the solvent molecules that the latter can be regarded as a continuum (i. e. not having molecular character). For such a system Stokes deduced that the frictional coefficient of the solute molecules is $f_i = 6\pi r_i^{(H)}\eta$, where $r_i^{(H)}$ is the hydrodynamical radius of the molecule and η is the viscosity of the solvent. For proteins diffusing in the cytosol, estimating the frictional coefficient through the Stokes's law is hard, for several reasons. First of all, the assumption of very large spherical molecules in a continuous solvent is not a realistic approximation for proteins moving through the cytosol: proteins may be not spherical and the solvent is not a continuum. Furthermore, in the protein-protein interaction, in the cytosol, water molecules should be included explicitly, thus complicating the estimation of the hydrodynamical radius. Finally, the viscosity of the solvent η within the cellular environment cannot be approximated either as the viscosity of liquid or the viscosity of gas. In both cases, the theory predicts a strong dependence on the temperature of the system, that has not been found in the cell system, where the most significant factor in determining the behavior of frictional coefficient is the concentration of solute molecules. To model the effects of non-ideality on the friction coefficient it is assumed that it linearly depends on the concentration of the solute as in sedimentation processes [11]. The equation (26) give the frictional coefficient $f_{i,k}$ of species i at mesh k. In this equation k_f is an empirical constant, whose value can be derived from the knowledge of the ratio $R = k_f/[\eta]$.

$$f_{i,k} = k_f c_{i,k} \tag{26}$$

Accordingly to the Mark-Houwink equation [9], $[\eta] = kM^{\alpha}$ is the intrinsic viscosity coefficient, α is related to the shape of the molecules of the solvent, and M is the molecular weight of the solute. If the molecules are spherical, the intrinsic viscosity is independent of the size of the molecules, so that $\alpha = 0$. All globular proteins, regardless of their size, have essentially the same $[\eta]$. If a protein is elongated, its molecules are more effective in increasing the viscosity and $[\eta]$ is larger. Values of 1.3 or higher are frequently obtained for molecules that exist in solution as extended chains. Long-chain molecules that are coiled in solution give intermediate values of α, frequently in the range from 0.6 to 0.75 [12]. For globular macromolecule, R has a value in the range of 1.4 - 1.7, with lower values for more asymmetric particles [13].

B. Calculated second virial coefficient

The mechanical statistical definition of the second virial coefficient is given by the following expression

$$B = -2\pi N_A \int_0^{\infty} r^2 \exp\Big[-\frac{u(r)}{k_B T}\Big] dr \tag{27}$$

where $u(r)$, which is given in Eq. (28), is the interaction free energy between two molecules, r is the intermolecular center-center distance, k_B is the Boltzman constant, and T the temperature. In this work, it is assumed that $u(r)$ is the Lennard-Jones pair (12,6)-potential (Eq. 28), that captures the attractive nature of the Van der Waals interactions and the very short-range Born repulsion due to the overlap of the electron clouds.

$$u(r) = 4\Big[\Big(\frac{1}{r}\Big)^{12} - \Big(\frac{1}{r}\Big)^{6}\Big] \tag{28}$$

By expanding the term $\exp\Big(\frac{4}{k_B T}\frac{1}{r^6}\Big)$ into an infinite series, the Eq. (27) becomes

$$B = -2\pi N_A \sum_{j=0}^{\infty} \frac{1}{j!}(T^*)^j \int_0^{\infty} r^{2-6j} \exp\Big[-T^*\frac{1}{r^2}\Big] dr$$

where $T^* \equiv 4/(k_B T)$ and thus

$$B = -\frac{\pi N_A}{6} \sum_{j=0}^{\infty} \frac{1}{!j} 4^j (k_B T)^{-\frac{1}{4}+\frac{1}{2}j}\ \Gamma\Big(-\frac{1}{4}+\frac{1}{2}j\Big) \tag{29}$$

The estimate of B is given by truncating the infinite series of Γ functions to $j = 4$, since, results not shown here prove that taking into account the additional terms, obtained for $j > 4$, does not significantly influence the simulation results.

III. The Optimal Size of the System's Subvolumes

The reaction chamber volume V is divided into subvolumes of volume Δ and side length l, on the basis of the kinetic and dynamical properties of the diffusion particles. The subvolumes has been chosen sufficiently small, so that the probability distributions of the reactants can be treated as uniform inside each subvolume. This means that the rate by which two molecules in a subvolume reacts does not depend on their initial locations.

Let consider diffusion as a time dependent process, in which some distribution of concentration is established at some moment, and then allowed to disperse without replenishment. The Fick's law and its analogues for the transport of other physical properties relate to the flux under the influence of a constant gradient. They therefore describe time-independent processes. They refer, for example, to the flow of particles along a constant concentration gradient which is sustained by injecting particles in one region, and drawing them off

in another. From the second Fick's law, the mean distance through which particle of solute has spread after time t is

$$l_f = 2\sqrt{\frac{D\ t}{\pi}} \tag{30}$$

where D is the diffusion coefficient of the particle.

Let t_e be the the mean free time with respect to non-reactive (elastic) collisions and t_r the mean free time with respect to reactive collisions. The net distance covered by the particle during its lifetime is

$$L = 2\sqrt{\frac{D\ t_r}{\pi}} = 2\sqrt{\frac{\pi l_f^2 t_r}{4 t_e \pi}} = l_f \sqrt{\frac{t_r}{t_e}} \tag{31}$$

In order to have a homogeneous mixing inside boxes, the length l of the box side has to fulfill the following inequality.

$$l \ll L \tag{32}$$

It is worthy of note the fact that if this inequalities is fulfilled, the particles in each box obeys the Einstein formula for the probability of fluctuations around the steady state. Note also that the rate by which two molecules in a subvolume react does not depend on their initial location if the inequality (32) is fulfilled.

In terms of the diffusion coefficient D, Eq. (31) and (32) can be re-written as

$$l \ll 2\sqrt{D t_r} \tag{33}$$

Now, in order to estimate the upper bound of l the diffusion coefficient D and the reaction time t_r have to be determined. The diffusion coefficient differs from species to species, and, in general, depends on the local concentration of solute. Since the local concentration of solute changes in time as consequence of the occurrence of the chemical reaction events and the diffusion events themselves, this would entail a dynamical change of l through the Eq. (33). This could make the algorithm of simulation more complex, so that, it is profitable to fix the value of l at the initialization time at

$$l \approx \sqrt{\langle D \rangle t_r} \tag{34}$$

where

$$\langle D \rangle = \frac{1}{R_{diff} + R_{react}} \sum_{i=1}^{M^{(diff)}} D_i^0 \tag{35}$$

and D_i^0 is the diffusion coefficient of the species i-th at time $t = 0$, and M^{diff} is the number of species that diffuse. In the next section the model of the diffusion coefficient as function of local concentration and the waiting time of reaction t_r is explained.

A. The waiting time of reaction

Let R_i be the i-th reaction channel expressed as

$$R_i : l_{i1} S_{p(i,1)} + l_{i2} S_{p(i,2)} + \cdots + l_{iL_i} S_{p(i,L_i)} \xrightarrow{r_i} \dots$$

where l_{ij} is the stoichiometric coefficient of reactant $S_{p(i,j)}$, $p(i,j)$ is the index that selects the species S that participate to R_i, L_i is the number of reactants in R_i, and r_i is the rate constant. If the fundamental hypothesis of stochastic chemical kinetics [7] holds within a box, both diffusion and reaction events waiting times are distributed according to a negative exponential distribution, so that a typical time step has size

$$t_r \approx \frac{1}{R}\Big(\sum_{\nu=1}^{R} a_\nu\Big)^{-1} = \frac{1}{R}\Big(\sum_{i=1}^{R_{diff}} a_i^{(diff)} + \sum_{i=1}^{R_{react}} a_i^{(react)}\Big) \tag{36}$$

where R is the number of events. It is given by $R = R_{diff} + R_{react}$, where R_{diff} is the number of diffusions and R_{react} is the number of reaction events [14]. The diffusion and reaction propensities are given by the following expressions, respectively

$$a_i^{(diff)} = r_i^{(diff)} \frac{\prod_{j=1}^{M_i^{(diff)}} (\#S_{p(i,j)})^{l_{ij}}}{\prod_{j=1}^{L_i^{(diff)}} l_{ij}!} \tag{37}$$

$$a_i^{(react)} = r_i^{(react)} \frac{\prod_{j=1}^{M_i^{(react)}} (\#S_{p(i,j)})^{l_{ij}}}{\prod_{j=1}^{L_i^{(react)}} l_{ij}!} \tag{38}$$

where $M_i^{(diff)}$ and $M_i^{(react)}$ are the number of chemical species that diffuse and the number of those the undergo to reactions, respectively. In general $M \neq M_i^{(diff)} + M_i^{(react)}$, since some species are involved both in diffusions and reactions. In Eq. (37), $r_i^{(diff)}$ is the kinetic rate associated to the jumps between neighboring subvolumes, whereas in Eq. (38), $r_i^{(react)}$ is the stochastic rate constants of the i-th reaction.

From Eq. (13), the rate coefficient of the first order reaction representing a diffusion is recognized to be as follows.

$$r_i^{(diff)} = \frac{D_{ii}}{l^2} \tag{39}$$

IV. The Algorithm and Data Structure

In this section the new stochastic simulation algorithm developed by the authors is illustrated. It incorporates into a Gillespie-like approach the spatial effects of diffusive phenomena accordingly to the diffusion model presented in the previous sections.

For the reader's convenience, a brief description of the Gillespie Direct and First Reaction methods is here reported. Let suppose that in the system there are R reactions and M chemical species. at any instant of time the system is decribed by the state vector $\vec{X}(t) = \{X_1(t), \dots, X_M(t)\}$ Gillespies algorithm asks two questions:

1) Which reaction occurs next?
2) When does it occur?

Both of these questions must be answered probabilistically by specifying the probability density $P(\mu, \tau)$ that the next reaction is μ and it occurs at time τ. It can be shown [7] that

$$P(\mu, \tau) = a_\mu \exp\Big(-\tau \sum_{j=1}^{R} a_j\Big) d\tau \tag{40}$$

This equation leads directly to the answers of the two afore mentioned questions. First, what is the probability distribution for reactions? Integrating $P(\mu,\tau))$ over all τ from 0 to ∞ results in

$$\Pr(\text{Reaction} = \mu) = \frac{a_\mu}{\sum_{j=1}^{R} a_j a_j} \tag{41}$$

where a_j the propensity of reaction j as in Eqs. (37) and (38).

Second, what is the probability distribution for times? Summing $P(\mu,\tau)$ over all τ results in

$$P(\tau)d\tau = \Big(\sum_{j=1}^{R} a_j a_j\Big) \exp\Big(-\tau \sum_{j=1}^{R} a_j\Big) d\tau \tag{42}$$

These two distributions lead to Gillespies direct algorithm:

1) Set initial numbers of molecules in $\vec{X}(t)$, set $t \leftarrow 0$, and the absolute simulation time T.
2) Calculate the propensity function, a_μ, for all j, $j = 1,\ldots,R$.
3) Choose j according to the distribution in Eq. (41).
4) Choose τ according to an exponential with parameter $\sum_{j=1}^{R} a_j$ (as in Eq. (42)).
5) Change the number of molecules to reflect execution of reaction μ. Set $t \leftarrow t + \tau$.
6) Go to Step 2 and repeat the procedure until $t \leq T$.

The algorithm is direct in the sense that it generates μ and τ directly. Gillespie also developed the First Reaction Method (FRM) which generates a putative time τ_j for each reaction to occur - a time the reaction would occur if no other reaction occurred first - then lets μ be the reaction whose putative time is first, and lets τ be the putative time τ_j. Formally, the algorithm for the First Reaction Method is as follows:

1) Set initial numbers of molecules in $\vec{X}(t)$, set $t \leftarrow 0$, and the absolute simulation time T.
2) Calculate the propensity function, a_μ, for all j, $j = 1,\ldots,R$..
3) For each μ, generate a putative time, τ_j, according to an exponential distribution with parameter a_j.
4) Let μ be the reaction whose putative time, τ_j, is least. 5. Let τ be τ_j.
5) Change the number of molecules to reflect execution of reaction μ. Set $t \leftarrow t + \tau$.
6) Go to Step 2 and repeat the procedure until $t \leq T$.

At first glance, these two algorithms may seem very different, but they are provably equivalent [7] that is, the probability distributions used to choose μ and τ are the same. With regard to the complexity of the procedure, this algorithm uses R random numbers per iteration (where R is the number of reactions), takes time proportional to r to update the a's, and takes time proportional to R to identify the smallest τ_j.

The design of the algorithm is inspired to the one proposed by Elf et al. [15] in the so-called *Next sub-volume method*. This method selects the next reaction and the time at which it will occur by using the Gillespie First Reaction method [7]. Each cell and the corresponding reaction time and reaction type is stored in a global priority queue that is sorted with increasing writing reaction time. From this queue at each time step, the fastest reaction (i. e. the reaction with the smallest waiting time) is picked and executed. Once the reaction has been executed the state of the cell, as well as the state of the neighboring cells that eventually have been affected by the occurrence of this reaction are updated. This approach is efficient as it does not update the state of all the cells, but only the one of the cells in which the occurrence of a reaction has produced changes in the inner amount of molecules. However, the method is centralised and sequential and does not scale to very large systems. Moreover, it cannot be easily adapted to turn parallel or distributed computing procedures to profit. Since the number of reactions involved in the system could be of the order of millions, the property of scalability is required to make large simulations feasible. The algorithm proposed by the authors overcomes the scalability's limitations of the Next sub-volume method by renouncing to the use of a global priority queue.

For each cell a set of *dependency relations* with neighbor cells is drawn; in a cell an event (reaction of diffusion) can be executed only if it is quicker than the diffusion events of the neighbor cells, since the diffusion events in and out of the cell could change the reactant concentrations, and, consequently the reaction propensities and the waiting times of the events in the neighbor cells. The algorithm has still the same average computational complexity of Elf's methods. Nevertheless, by removing the global priority queue and introducing a dependency relations graph, the algorithm gains the scalability property. The new algorithm consists of the following steps.

1) Set initial numbers of molecules in $\vec{X}(t)$, set $t \leftarrow 0$, and the absolute simulation time T. Divide the reaction chamber volume V into boxes of size l as in Eq. (34).
2) In each cell, calculate the time and the type of the next event with the FRM are and store them in a private priority queue, ordered with increasing waiting time.
3) Each cell "communicates" with its neighbors, in a hierarchical way on the basis of the dependency relations, to decide which one holds the event with the smallest waiting time, say τ_s. that will be executed next. Execute the event and update the state of the cell and the one of the neighbor cells, in the case in which the event is a diffusion, are updated.
4) Update the time variable: $t \leftarrow t + \tau_s$.
5) Go to Step 2 and repeat the steps until $t \leq T$.

A. The prototype Redi

Redi (Fig. 1) is a simulator of the stochastic kinetics of a reaction-diffusion system, that implements the model described in the previous sections. *Redi* is a command line tool, that is invoked by typing the command

```
redi.exe [params]
```

Models for the reaction-diffusion system are given in a simple input language in a text input file Named entities that appear in the system are declared in the first section of the input file. In addition, it is possible to specify a base rate for entities that undergo diffusion:

```
var Y : rate 10e-6;
```

```
var Yp : rate 10e-6;
var Z2 : weight 24.0;
var M;
var MYp;
```

In this example, Y and Yp undergo diffusion events. Their basal diffusion rate is fixed, and it is specified after the *rate* keyword. Z2 can diffuse as well, but its basal diffusion rate is not fixed, but it is computed by Eq. (**??** .

The algorithm takes as input the molecular weight of the protein to compute its chemical activity (see Eqs. (24) and (25)); the weight is expressed in kDA after the *weight* keyword. M and MYp ado not undergo diffusion event, so they have no rate or weight associated to them.

Reactions can be specified in a very intuitive format:

```
Yp + Z2 -> Y + Z2   [1.90E-03];
M + Yp <- -> MYp   [5.93E-03, 20];
```

In the first line, a bimolecular reaction is specified. The rate constant is given in scientific notation between square brackets. The second reaction is a shorthand to specify reversible reactions. In this case, it is necessary to specify a pair of rates between the square brackets for the forward and for the backward reaction, respectively.

In the input file the reaction list has to be followed by the specification of the number of molecules and spatial location of each species on the grid. For example:

```
Y [0, 0, 0, 500];
Y [0, 1, 0, 500];
Y [1, 0, 0, 500];
Y [1, 1, 0, 500];
```

After the species name, the first three number within the square brackets indicate the x, y, and z coordinates, and the fourth number is the amount of and an integer number. Although this input format may be verbose, it allows for great control and precision. We are currently studying better ways of inputting both data (location and amount of particles) and geometry (that is, for now, fixed and with a regular polyhedric shape).

The Redi package includes a number of modules providing different kinds of output, e. g. text files, writing data files to be used with professional visualization packages, visualizing in a simple 3D window concentrations in space, etc.) as well as a simple interface to write custom output modules (see an example of output visualization in Fig. 2).

V. Case study: chaperone-assisted folding

Although a protein chain can fold in its correct conformation without outside help, protein folding in a living cell is often assisted by special proteins called *molecular chaperones*. These proteins bind to partly folded polypeptide chains and help them progress along the most energetically favorable folding pathways. Chaperones are vital in the crowded conditions of the cytoplasm, since they prevent the temporarily exposed hydrophobic regions in newly synthesized protein chains from associating with each other to form proteins aggregates.

In the healthy cells, if a protein does not assume the correct 3D shape, or a cellular stress induces a right-folded protein to assume a wrong folding, the chaperones re-shape it correctly. In the case in which the protein is not correctly refolded, and the ubitiquitin-proteasome system, designed to its digestion, does not correctly work, as in many neurodegenerative disorders, the faulty proteins accumulate and damage the cell.

Fig. 1. Redi's logo.

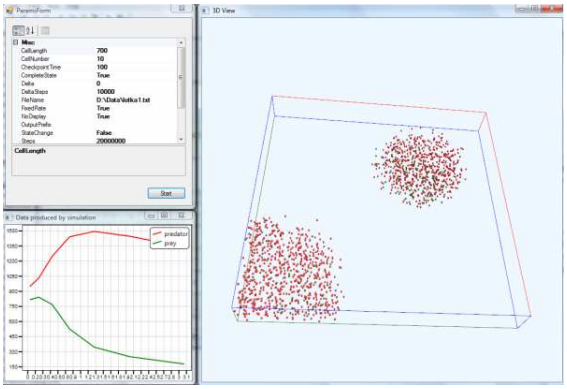

Fig. 2. Redi's screenshots of output's visualization (http://www.cosbi.eu).

Protein folding, chaperone binding, and misfolded protein accumulation - all of these processes take place inhomogeneously in the space. The spatial distribution of chaperones in the cytoplams may not be uniform, and consequently the distribution of correct and faulty proteins may be not uniform. In turn, the time evolution of spatial distribution of chaperones may affect the time evolution of the spatial distribution of faulty proteins. The reaction-diffusion systems of the case study consisting of the four reactions is shown in Table I, where $chaperone$ represents the molecular chaperone, $nascent\_protein$ presents the protein chain release from the ribosome, $right\_protein$ denotes the correctly folded protein, $misfolded\_1$ is a faulty protein generated by the first interaction with the chaperone (Reaction 2), and $misfolded\_2$ is the misfolded protein generated by the interaction between $misfolded\_1$ and $chaperone$ (Reaction 4).

Accordingly to the measurements reported in [16] the following values of diffusion coefficients have been used to simulate the system : $D^0_{protein} = D^0_{right\_protein} = D^0_{misfolded\_1} = D^0_{misfolded\_2} = 10\ \mu m^2 \text{sec}^{-1}$, and $D^0_{chaperone} = 1\ \mu m^2 \text{sec}^{-1}$. As simulation space, a square grid $9 \times 9 \mu m^2$, thus consisting of 81 cells (each cell has size $l = 1$ nm) is considered. A 2D diffusion model is simulated

TABLE I
CHAPERONE-ASSISTED PROTEIN FOLDING. REACTION 1 DESCRIBES THE FOLDING OF THE NASCENT PROTEIN INTO A CORRECTLY WORKING PROTEIN (RIGHT_PROTEIN). REACTION 2 DESCRIBES THE UNCORRECT FOLDING OF THE NASCENT PROTEIN INTO A MISFOLDED PROTEIN (MISFOLDED_1). REACTION 3 DESCRIBES THE INTERACTION BETWEEN THE CHAPERONE AND THE MISFOLDED PROTEIN, THAT, CONSEQUENTLY, IS TRANSFORMED INTO A CORRECTLY FOLDED PROTEIN. FINALLY, REACTION 4 DESCRIBES THE INTERACTION BETWEEN THE CHAPERONE AND THE MISFOLDED PROTEINS, THAT IS NOT CONVERTED INTO A CORRECTLY WORKING PROTEIN.

Reaction 1: $nascent\_protein + chaperone \xrightarrow{100\mu M^{-1}sec^{-1}} chaperone + right\_protein$

Reaction 2: $nascent\_protein + chaperone \xrightarrow{100\mu M^{-1}sec^{-1}} chaperone + misfolded\_1$

Reaction 3: $misfolded\_1 + chaperone \xrightarrow{100\mu M^{-1}sec^{-1}} chaperone + rprotein$

Reaction 4: $misfolded\_1 + chaperone \xrightarrow{100\mu M^{-1}sec^{-1}} chaperone + misfolded\_2$

and a spatially homogeneous distribution of $nascent\_protein$ and an initial null concentration of $right\_protein$ in every cell are assumed. The density (expressed in number of molecules per μm^3) and the spatial distribution of $chaperone$, $misfolded\_1$, and $misfolded\_2$ in the first instants of simulation are shown in the first plots (at time $t \approx 10^{-5}$ sec), in Fig. 3 (A), Fig. 3 (B), Fig. 3 (C), and Fig. 3 (D) respectively.

At time $t = 1.1054 \times 10^{-5}$ sec - immediately after the begging of the simulation, the correctly folded proteins are located in the regions where the concentrations of chaperones is high (see Fig. 3 (A) and (B)). The misfolded proteins produced by Reaction 2 and Reaction 4 in the first instants of the simulation are close to the chaperones (Fig. 3 (C) and (B), Fig. 4 (C) and (D), and Fig. 5 (C) and (D)). at time $t = 0.000483$ sec, the chaperones and the correctly folded proteins start to leave their initial positions to migrate toward the central area of the system (Fig. 6 (A)). The concentration of misfolded proteins (of type 1 and 2) increses and their distributions spread in the space (Fig. 6 (C) and (D)). From $t = 0.003211$ sec to $t = 0.005080$ sec, the concentration of the chaperones is non-null in all the simulation space with a peak in the right upper corner (Fig. 9 (A) and Fig. 10 (A)). The distribution of right folded proteins is similar (Fig. 9 (B) and 10 9B). The concentration of misofled proteins produced by Reaction 2 is almost null in all the space except along the borders (Fig. 9 (C) and Fig. 10 (C)). Nevertheless, the concentration of misfolded proteins produced by Reaction 4 is significantly different from zero and fairly homogeneous (Fig 9 (D) and Fig. 10 (D)). At $t = 0.007749$ sec, the chaperones shift to the upper norder of the simulation space (Fig. 11 (A)); the correctly folded proteins concentration has a maximum in the right upper corner (Fig. 11 (B)); the concentration of misfolded proteins by Reaction 2 is almost everywhere except that on the borders, whereas the distribution of misfolded produced in Reaction 4 is almost everywhere null , but it has a peak in the right upper corner (Fig. 11 (D)). Finally, at time $t = 0.014273$ secis non-null over all the space. It increases linearly from the upper border (Fig. 12 (A)). The concentration of correctly folded proteins increases from the lower left corner to the right upper corner (Fig. 12 (B)). Unlike the distribution of misfolded proteins deriving from Reaction 2, the distribution of misfolded proteins deriving from reaction 4 is different from zero everywhere (Fig. 12 (C) and increases from the left lower corner to the right upper corner (Fig. 12 (D)).

A. Spatial correlation between chaperones and proteins

The spatial correlation between the proteins and chaperones has been monitored in terms of the quantity $C_{p,c}$, which is defined by

$$C_{p,c} = \frac{\langle (\Phi_p - \langle \Phi_p \rangle)(\Phi_c - \langle \Phi_c \rangle) \rangle}{\langle \Phi_p \rangle \langle \Phi_c \rangle} \tag{43}$$

where $\Phi_p = \Phi_p(x, y, z)$ and $\Phi_c = \Phi_c(x, y, z)$ are function of spatial coordinates and denote the concentrations of nascent proteins and chaperones, respectively. The symbol $\langle \cdot \rangle$ denotes the mean value of "$\cdot$". The subscript p ranges over the following species $right\_protein$, $misfolded\_1$, and $misfolded\_2$, whereas the subscript c denotes $chaperone$. The positive value of $C_{p,c}$ means that the species p and c on average tend to be close each other in space. The average correlation between chaperones and correctly folded proteins, chaperones abd misfolded proteins derived from Reaction 2 and chaperones and misfolded proteins derived from reaction 4 decrease with increasing time (Fig. 13 (A), (B), and (C), respectively). The distribution of the intensity of these correlations in the simulation space is shown in Figs. 15 - Figs. 24. These results show that, at the beginning of the simulation, both the correctly folded and the misfolded proteins are likely to appear near the chaperones, that is they are released by the chaperones, and then they diffuse away from them, as it was obtained also in [16]. The figures 14 (A), (B), and (C) show that the total concentrations of correctly proteins, misfolded proteins (1) and (2), respectively, have a time behavior symmetric to the time behavior of their average correlations with the concentration of chaperones. In fact, the maximum of the correlation between chaperones and both correctly and misfolded proteins correspond to the onset of increase in protein concentration. The figures 13 (B) and 14 (B) show that the concentration of misfolded proteins produced in reaction 2 reaches the maximum when their correlation with the chaperones has a minimum. This behavior is due to the fact that the misfolded proteins of type 1 are released from the chaperones and then they quickly diffuse away from them. The chaperones also diffuse away from their initial positions but less quickly, so that they reach later the misfolded proteins of type 1. Once the chaperones reached the misfolded proteins, the occurrence of Reaction 4 causes the decreasing of the concentration of misfolded protein of type 1.

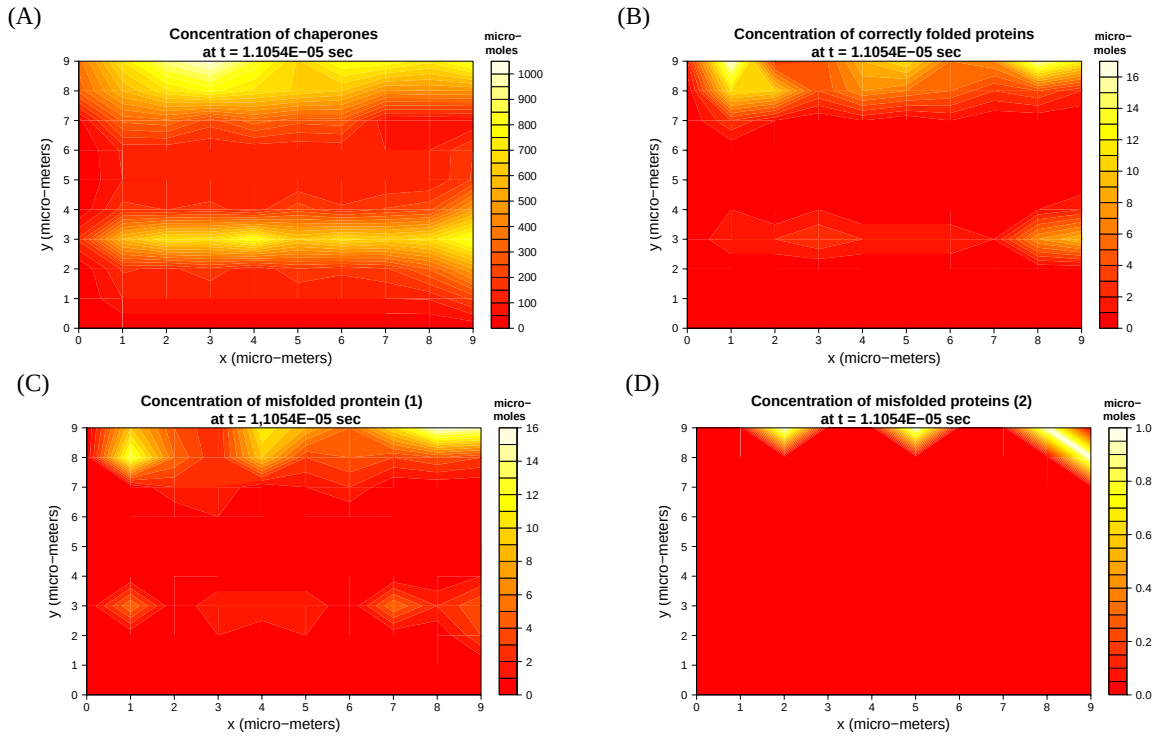

Fig. 3. Distribution of the concentration of chaperones (A), correctly folded proteins (B), misfolded proteins deriving from the Reaction 2 (C), and misfolded proteins deriving from Reaction 4 (D). The figures are snapshots of the system at time $t = 1.1054 \times 10^{-05}$ sec.

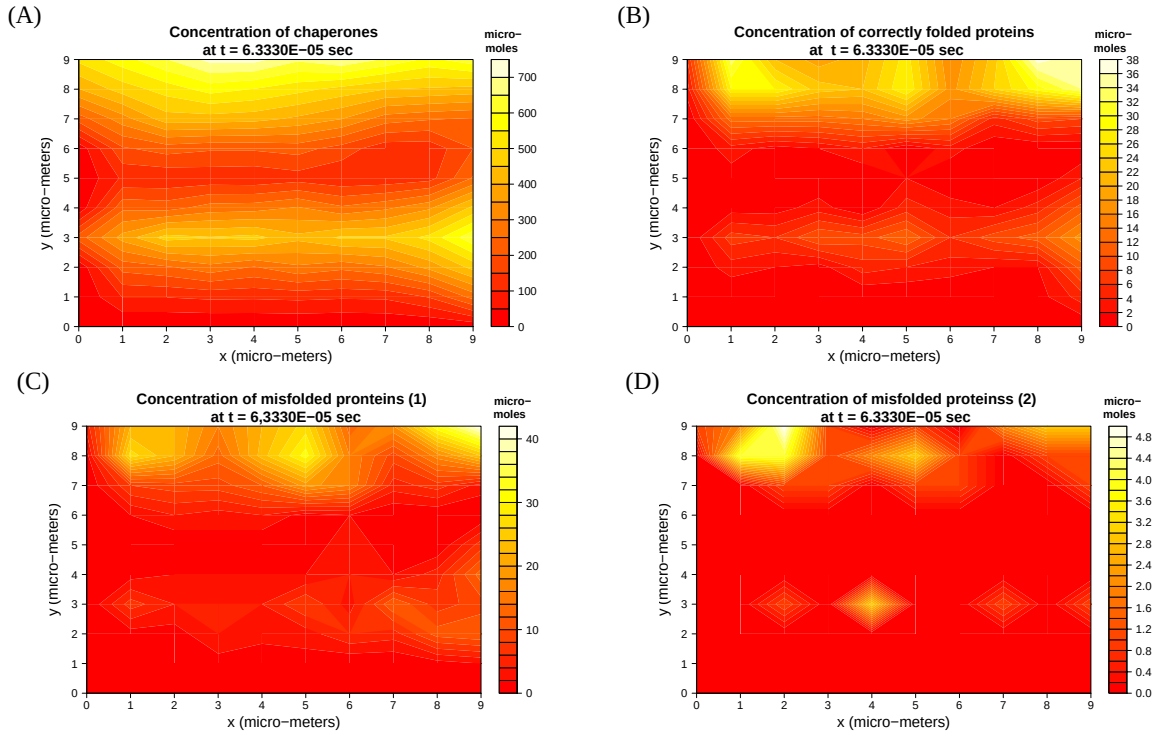

Fig. 4. Distribution of the concentration of chaperones (A), correctly folded proteins (B), misfolded proteins deriving from the Reaction 2 (C), and misfolded proteins deriving from Reaction 4 (D). The figures are snapshots of the system at time $t = 6.333 \times 10^{-05}$ sec.

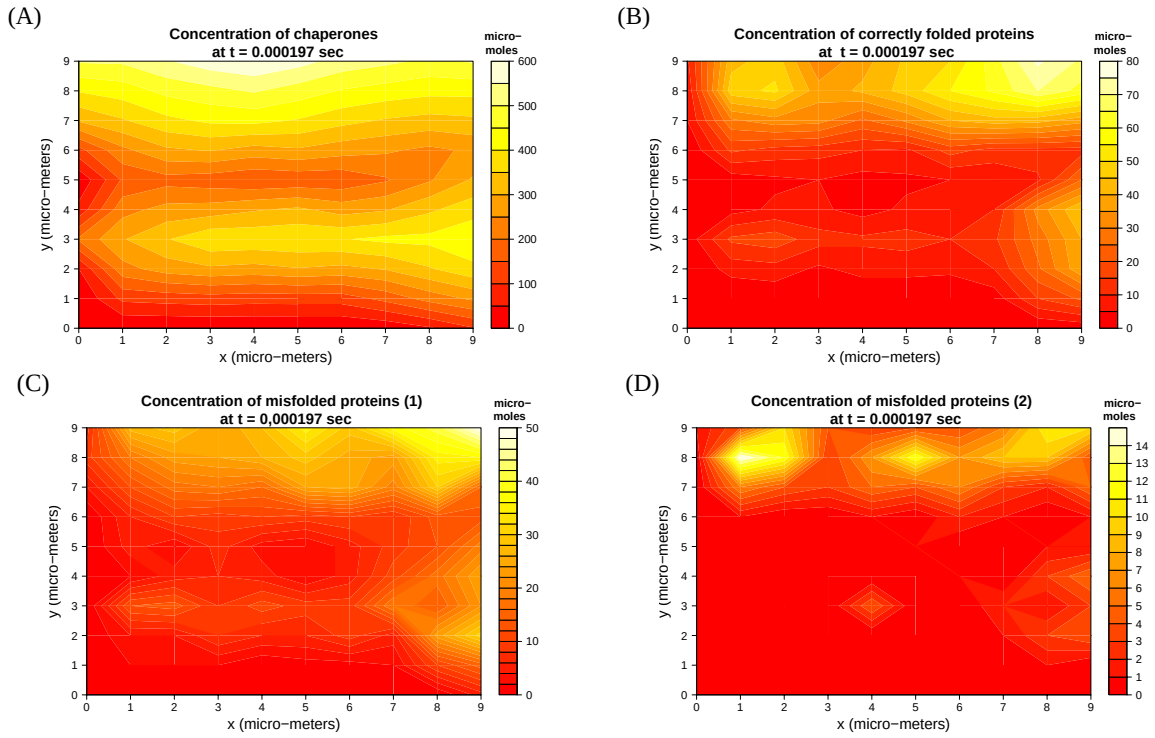

Fig. 5. Distribution of the concentration of chaperones (A), correctly folded proteins (B), misfolded proteins deriving from the Reaction 2 (C), and misfolded proteins deriving from Reaction 4 (D). The figures are snapshots of the system at time $t = 0.000197$ sec.

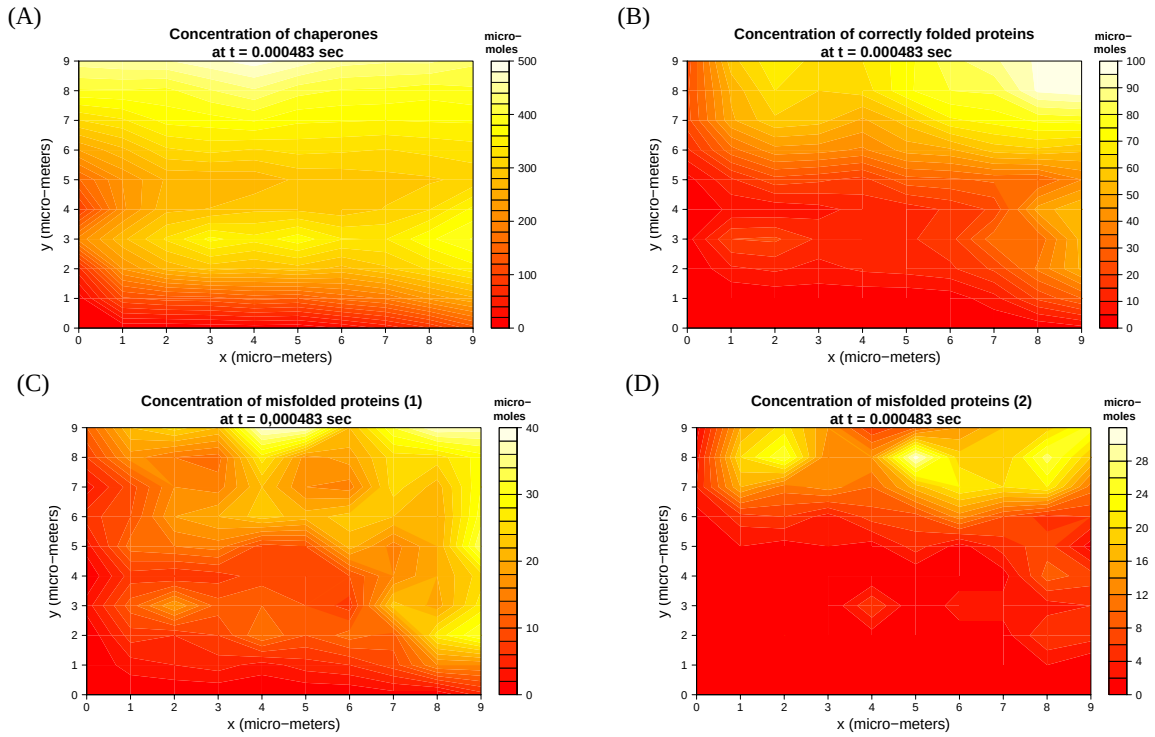

Fig. 6. Distribution of the concentration of chaperones (A), correctly folded proteins (B), misfolded proteins deriving from the Reaction 2 (C), and misfolded proteins deriving from Reaction 4 (D). The figures are snapshots of the system at time $t = 0.000483$ sec.

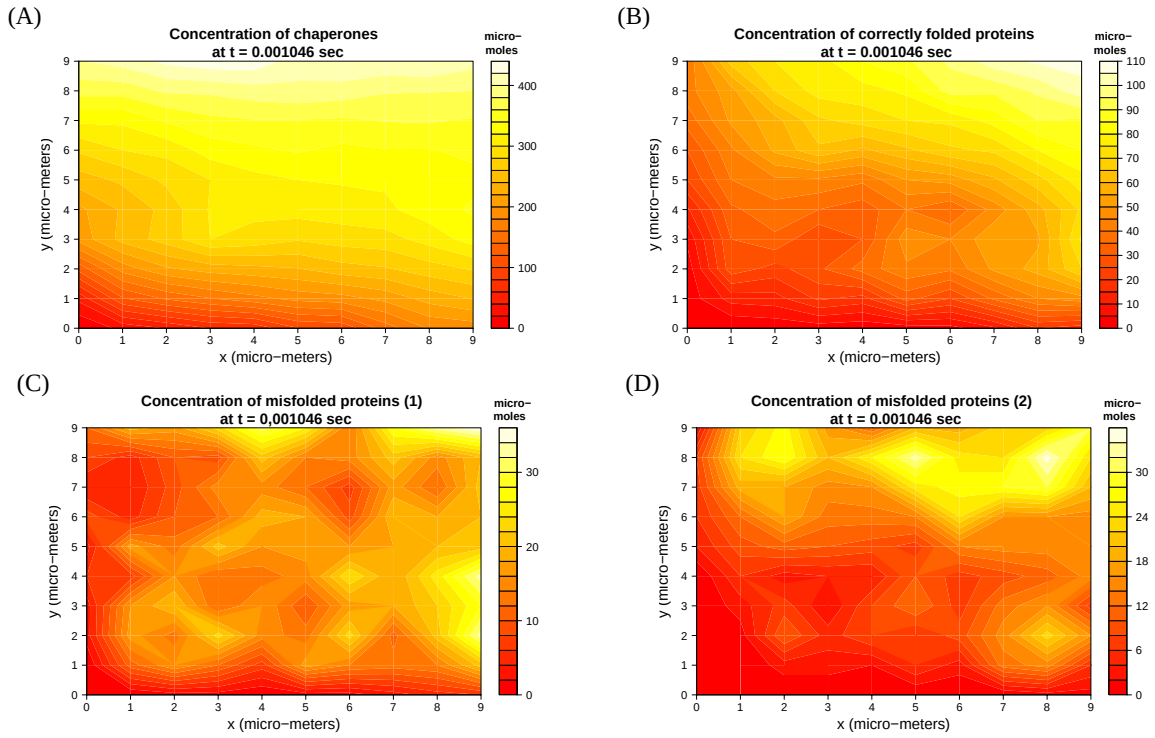

Fig. 7. Distribution of the concentration of chaperones (A), correctly folded proteins (B), misfolded proteins deriving from the Reaction 2 (C), and misfolded proteins deriving from Reaction 4 (D). The figures are snapshots of the system at time $t = 0.001046$ sec.

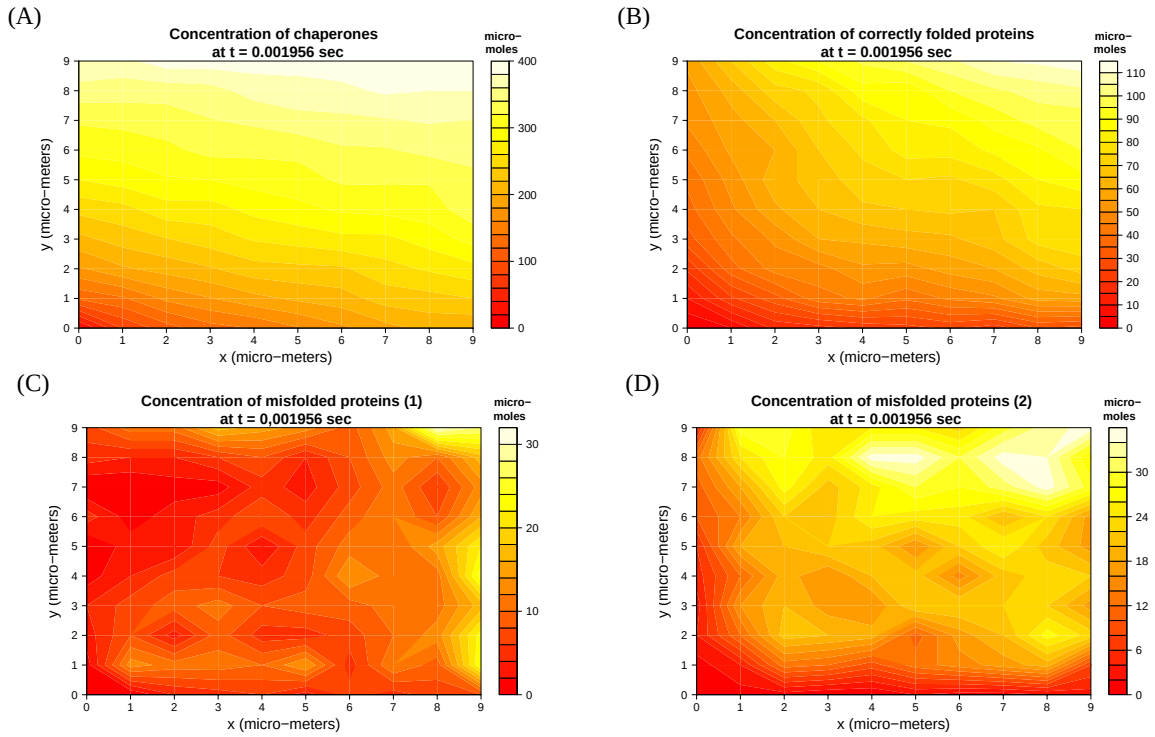

Fig. 8. Distribution of the concentration of chaperones (A), correctly folded proteins (B), misfolded proteins deriving from the Reaction 2 (C), and misfolded proteins deriving from Reaction 4 (D). The figures are snapshots of the system at time $t = 0.001956$ sec.

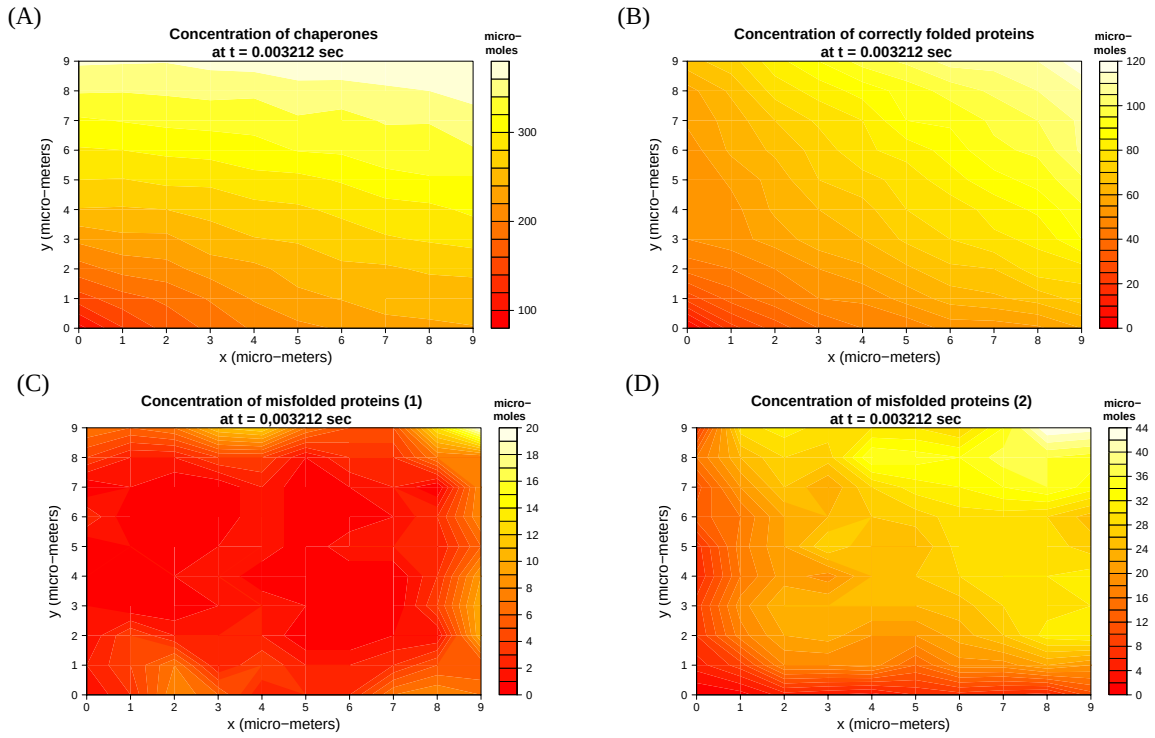

Fig. 9. Distribution of the concentration of chaperones (A), correctly folded proteins (B), misfolded proteins deriving from the Reaction 2 (C), and misfolded proteins deriving from Reaction 4 (D). The figures are snapshots of the system at time $t = 0.003212$ sec.

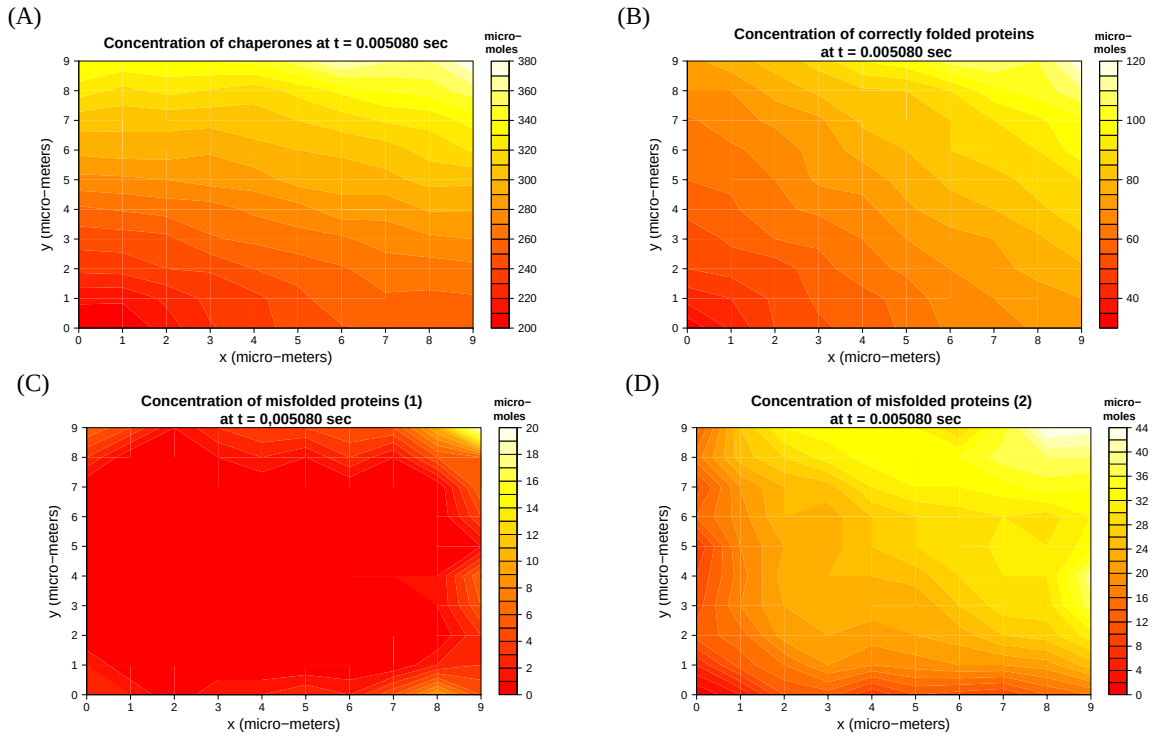

Fig. 10. Distribution of the concentration of chaperones (A), correctly folded proteins (B), misfolded proteins deriving from the Reaction 2 (C), and misfolded proteins deriving from Reaction 4 (D). The figures are snapshots of the system at time $t = 0.005080$ sec.

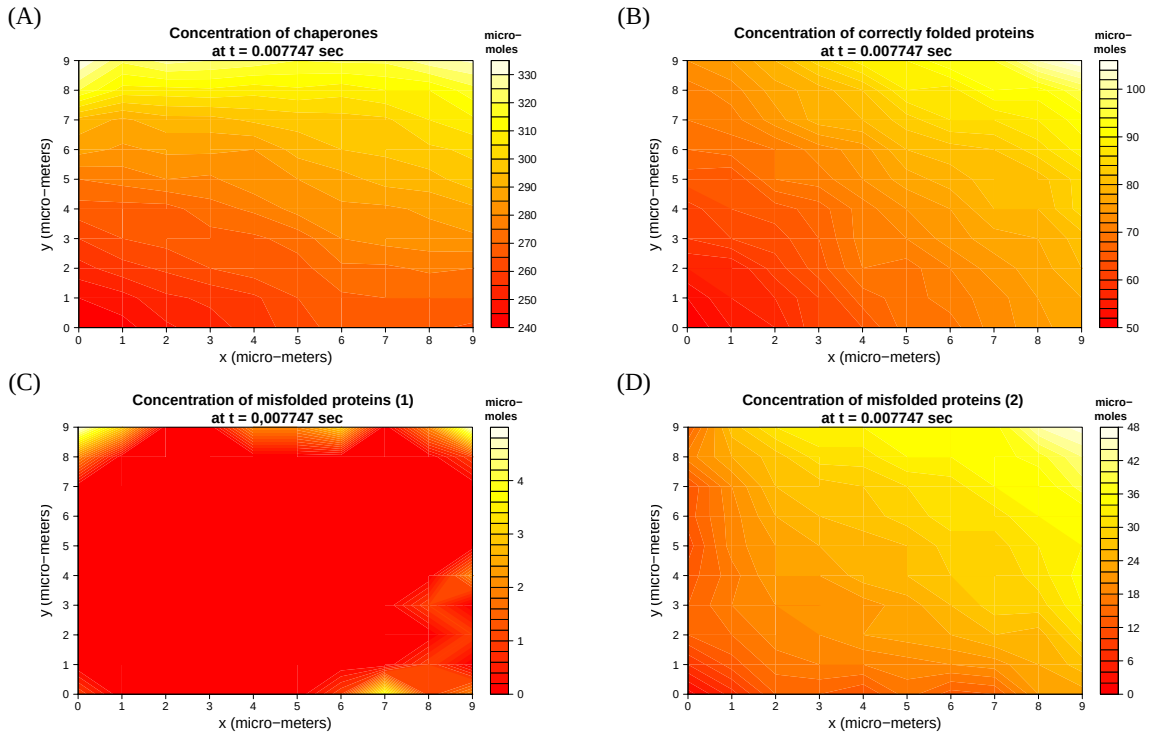

Fig. 11. Distribution of the concentration of chaperones (A), correctly folded proteins (B), misfolded proteins deriving from the Reaction 2 (C), and misfolded proteins deriving from Reaction 4 (D). The figures are snapshots of the system at time $t = 0.007747$ sec.

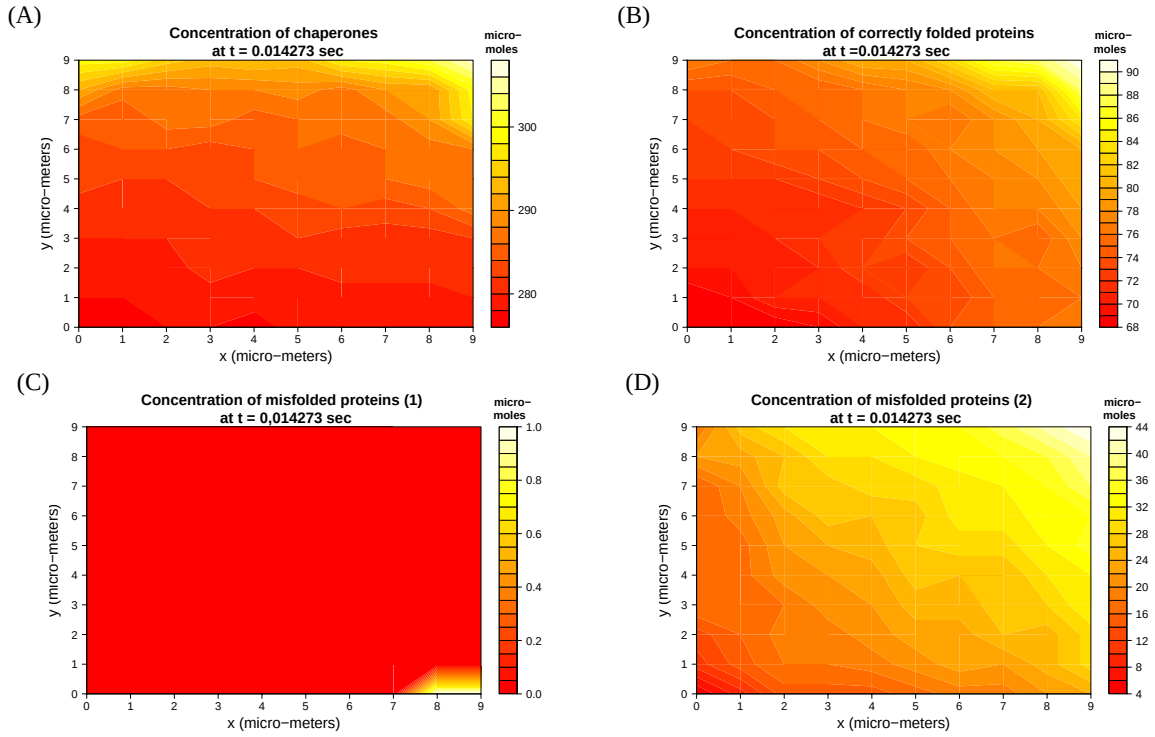

Fig. 12. Distribution of the concentration of chaperones (A), correctly folded proteins (B), misfolded proteins deriving from the Reaction 2 (C), and misfolded proteins deriving from Reaction 4 (D). The figures are snapshots of the system at time $t = 0.014273$ sec.

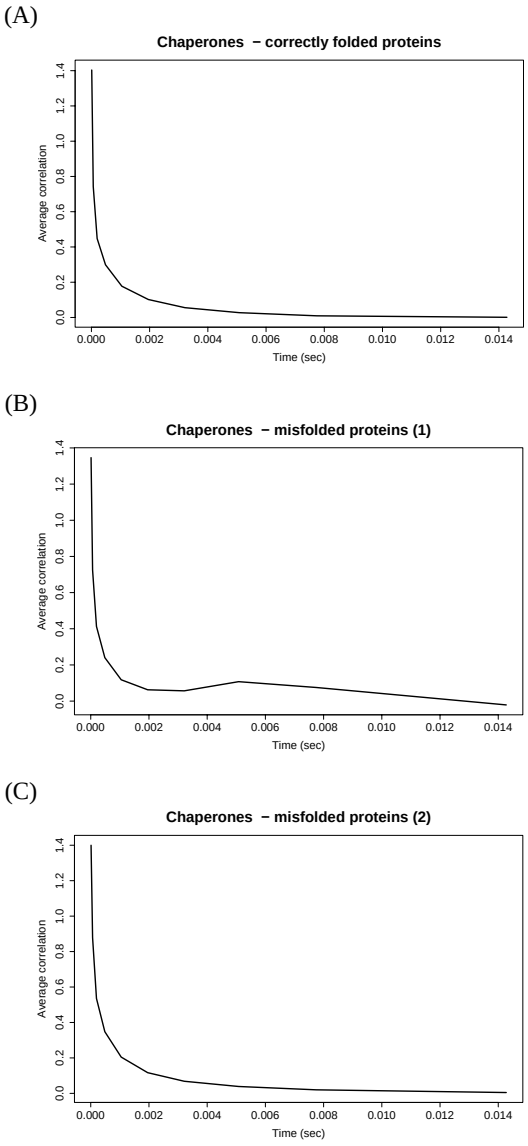

Fig. 13. Time behavior of the average correlation between chaperones and correctly folded proteins (A), chaperones and misfolded proteins produce in Reaction 2 (B), and chaperones and misfolded proteins produced in Reaction 4 (C).

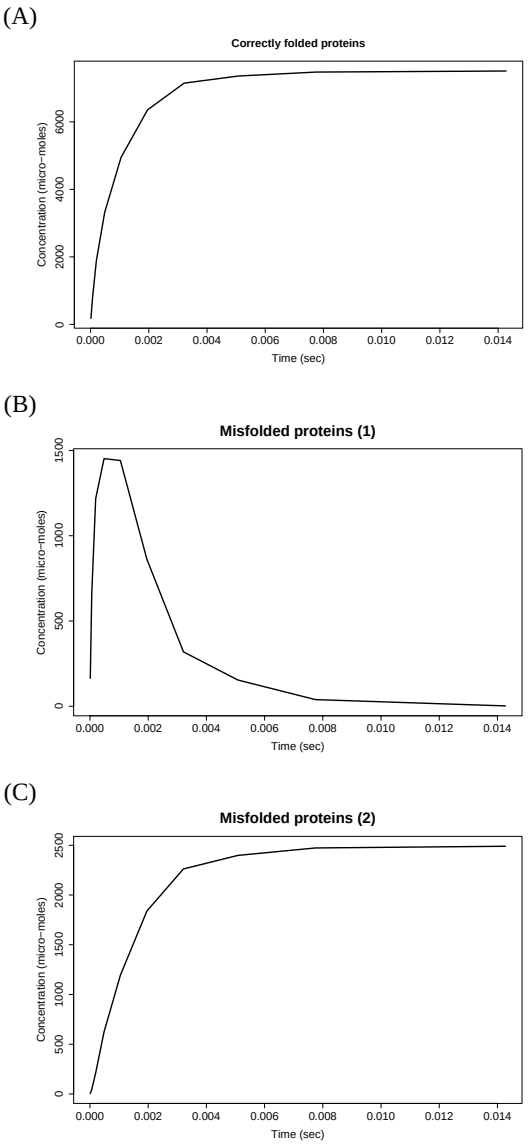

Fig. 14. Time behavior of the total concentration of correctly folded proteins (A), misfolded proteins produce in Reaction 2 (B), and misfolded proteins produced in Reaction 4 (C).

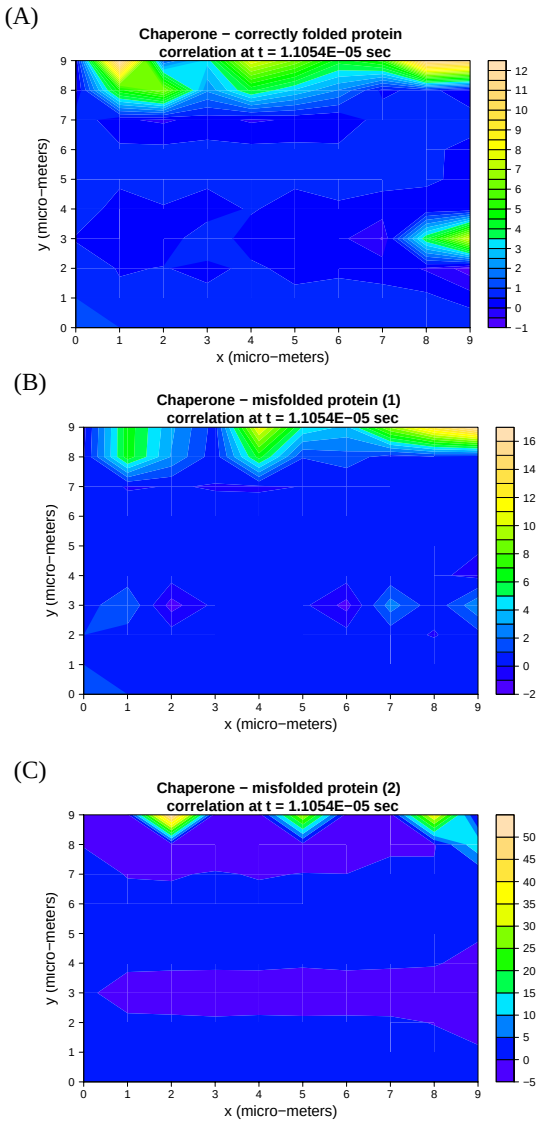

Fig. 15. Matrices of correlation (Eq, (43)) between chaperones and correctly folded proteins concentrations (A), chaperones and misfolded proteins conccentrations deriving from the Reaction 2 (B), and chaperones and misfolded proteins concentration deriving from Reaction 4 (C). The figures are snapshots of the system at time $t = 1.1054 \times 10^{-05}$ sec.

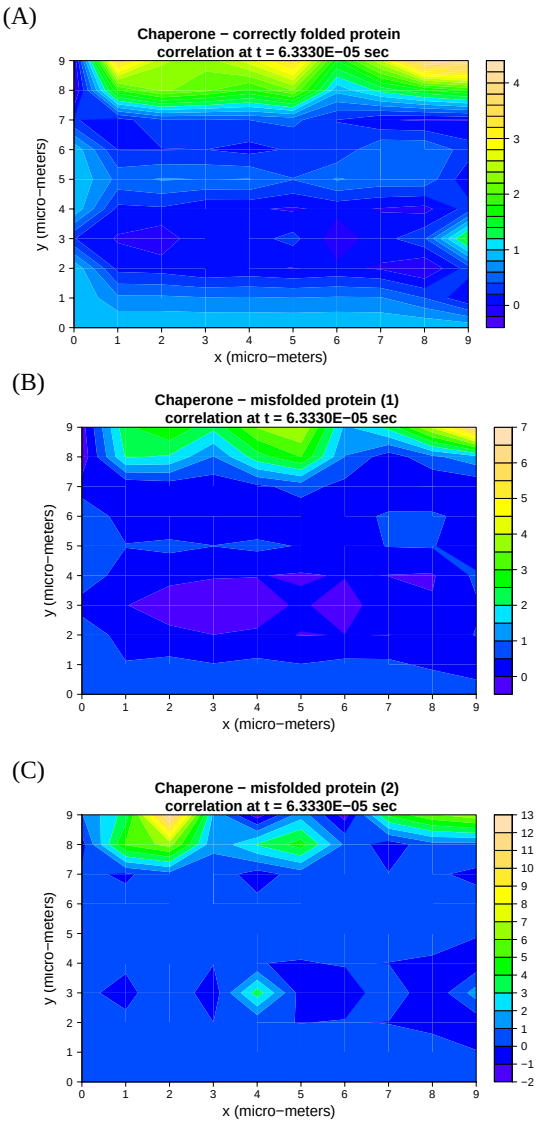

Fig. 16. Matrices of correlation (Eq, (43)) between chaperones and correctly folded proteins concentrations (A), chaperones and misfolded proteins conccentrations deriving from the Reaction 2 (B), and chaperones and misfolded proteins concentration deriving from Reaction 4 (C). The figures are snapshots of the system at time $t = 6.333 \times 10^{-05}$ sec.

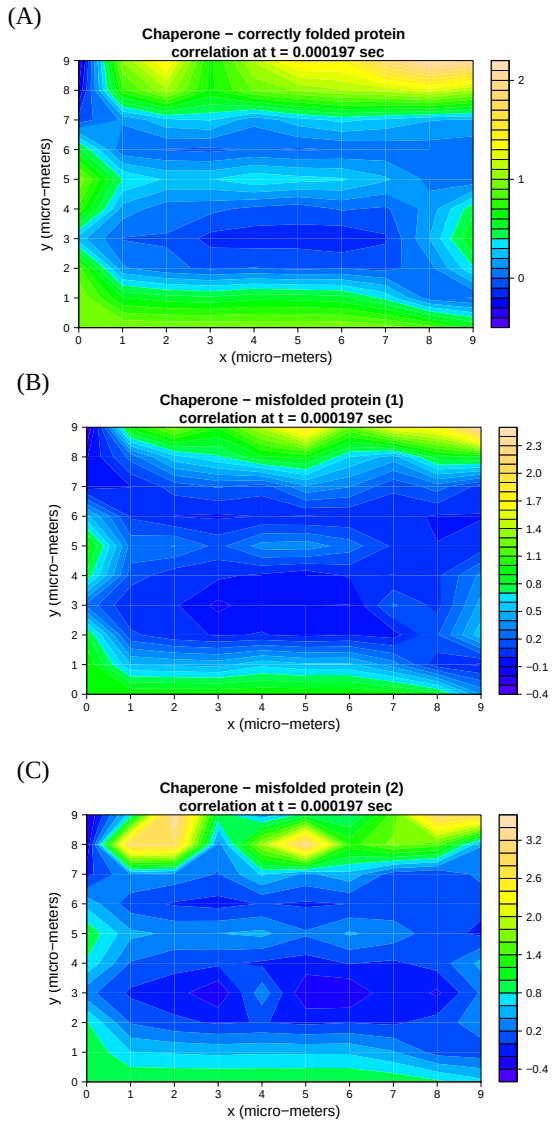

Fig. 17. Matrices of correlation (Eq, (43)) between chaperones and correctly folded proteins concentrations (A), chaperones and misfolded proteins conccentrations deriving from the Reaction 2 (B), and chaperones and misfolded proteins concentration deriving from Reaction 4 (C). The figures are snapshots of the system at time $t = 0.000197$ sec.

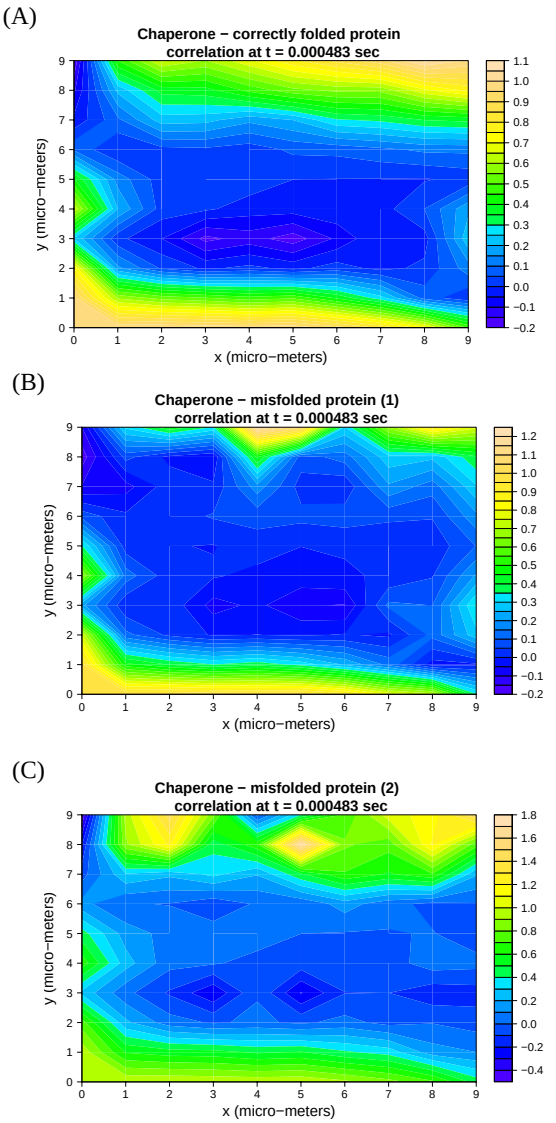

Fig. 18. Matrices of correlation (Eq, (43)) between chaperones and correctly folded proteins concentrations (A), chaperones and misfolded proteins conccentrations deriving from the Reaction 2 (B), and chaperones and misfolded proteins concentration deriving from Reaction 4 (C). The figures are snapshots of the system at time $t = 0.000483 \times 10^{-05}$ sec.

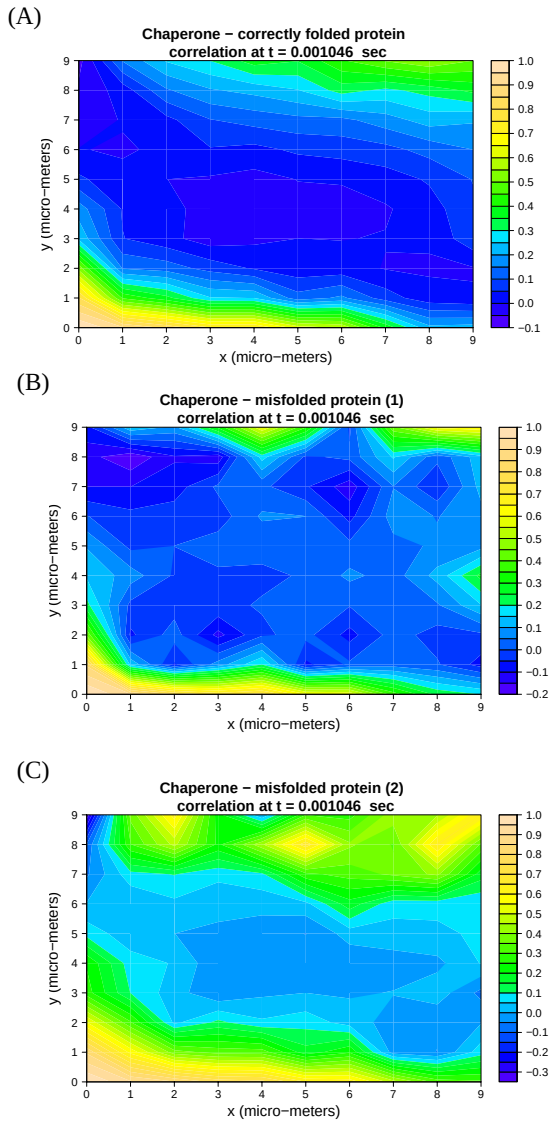

Fig. 19. Matrices of correlation (Eq, (43)) between chaperones and correctly folded proteins concentrations (A), chaperones and misfolded proteins conccentrations deriving from the Reaction 2 (B), and chaperones and misfolded proteins concentration deriving from Reaction 4 (C). The figures are snapshots of the system at time $t = 0.001046$ sec.

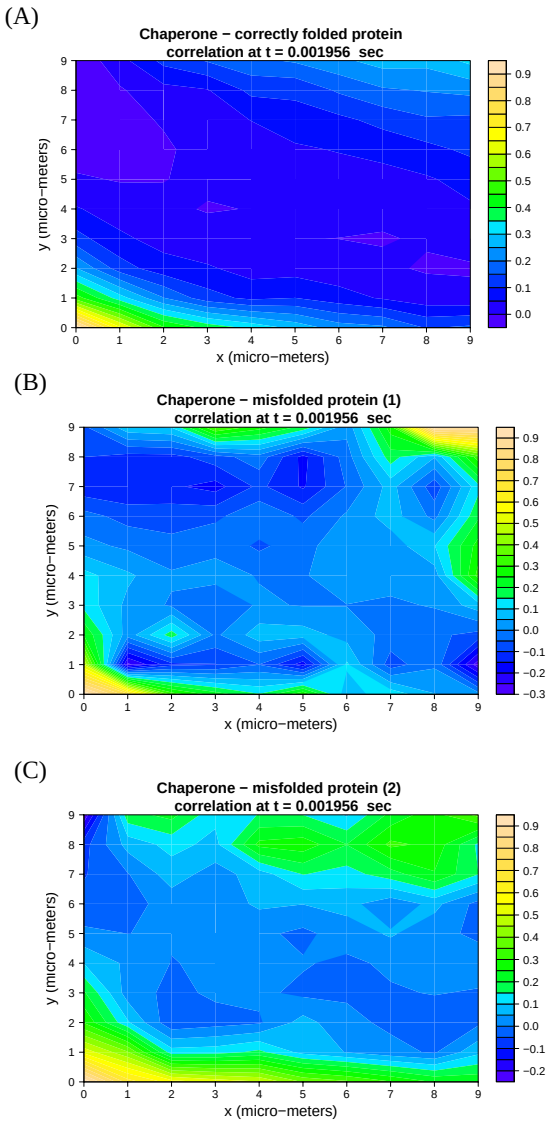

Fig. 20. Matrices of correlation (Eq, (43)) between chaperones and correctly folded proteins concentrations (A), chaperones and misfolded proteins conccentrations deriving from the Reaction 2 (B), and chaperones and misfolded proteins concentration deriving from Reaction 4 (C). The figures are snapshots of the system at time $t = 0.001956$ sec.

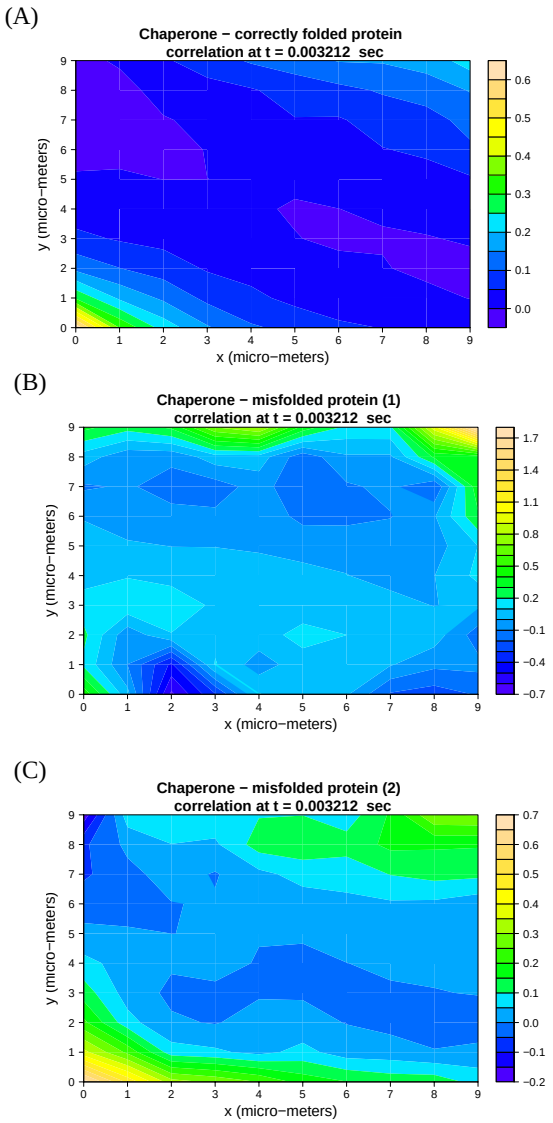

Fig. 21. Matrices of correlation (Eq, (43)) between chaperones and correctly folded proteins concentrations (A), chaperones and misfolded proteins conccentrations deriving from the Reaction 2 (B), and chaperones and misfolded proteins concentration deriving from Reaction 4 (C). The figures are snapshots of the system at time $t = 0.003212$ sec.

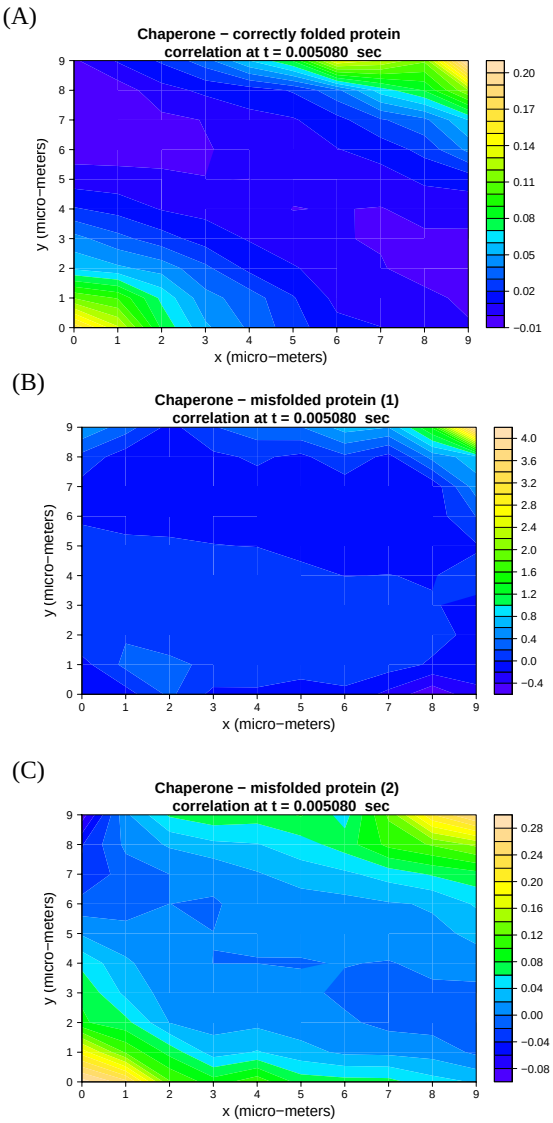

Fig. 22. Matrices of correlation (Eq, (43)) between chaperones and correctly folded proteins concentrations (A), chaperones and misfolded proteins conccentrations deriving from the Reaction 2 (B), and chaperones and misfolded proteins concentration deriving from Reaction 4 (C). The figures are snapshots of the system at time $t = 0.005080$ sec.

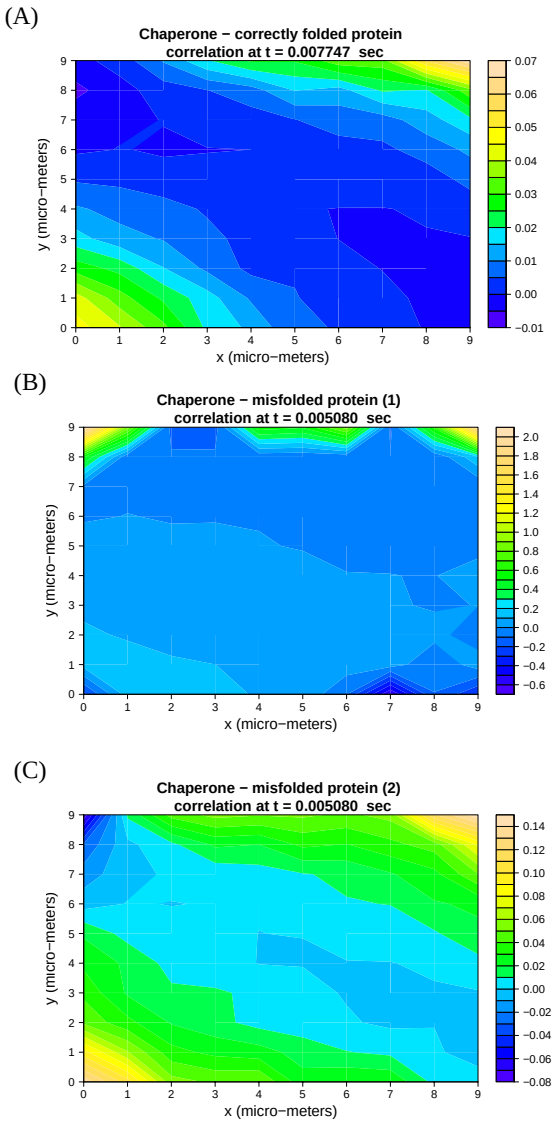

Fig. 23. Matrices of correlation (Eq, (43)) between chaperones and correctly folded proteins concentrations (A), chaperones and misfolded proteins conccentrations deriving from the Reaction 2 (B), and chaperones and misfolded proteins concentration deriving from Reaction 4 (C). The figures are snapshots of the system at time $t = 0.007747$ sec.

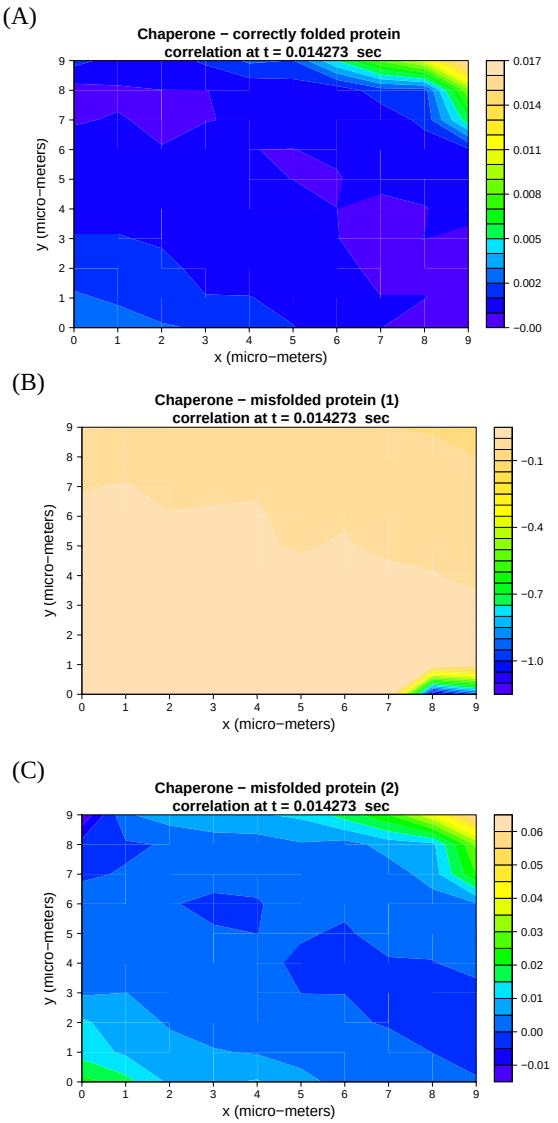

Fig. 24. Matrices of correlation (Eq, (43)) between chaperones and correctly folded proteins concentrations (A), chaperones and misfolded proteins conccentrations deriving from the Reaction 2 (B), and chaperones and misfolded proteins concentration deriving from Reaction 4 (C). The figures are snapshots of the system at time $t = 0.014273$ sec.

B. Validity of the model

In this model describing the effects of an irregular distribution of chaperones on the kinetics of the chaperone-assisted protein folding, the internal structure and mechanism of the chaperone, as well as the size and the internal dynamics of the protein folding are not treated. No external source of energy is exerted upon the system in the present simulations: the diffusive transport is caused by spatial differences of concentrations of solute. Moreover, chaperones assist not only the efficient folding of newly-translated proteins as these proteins are being synthesized on the ribosome, but they can also maintain pre-existing proteins in a stable conformation. Chaperones can also promote the disaggregation of preformed protein aggregates. The general mechanism by which chaperones carry out their function usually involves multiple rounds of regulated binding and release of an unstable conformer of target polypeptides. These reaction are not included in this simple model.

Apart from the above limitations, the model captures the essential features of the kinetics of the chaperone-assisted protein folding (see [16], [17], [18], [19], [20], [21], [22], [23]). Both the correct and misfolded proteins appear near to the chaperones, as the proteins are released from the chaperones. The correlation between chaperones and correctly folded proteins, as well as the correlation between chaperones and misfolded proteins deriving from Reaction 4, vanish at $t \approx 0.01$. This suggest that, after that time, the proteins released from the chaperones quickly diffuse away from them and aggregates at the site where the cheperones are less abundant. The diffusion of the chaperones toward those sites causes the decrement and the subsequent stabilization of the amount of misfolded proteins.

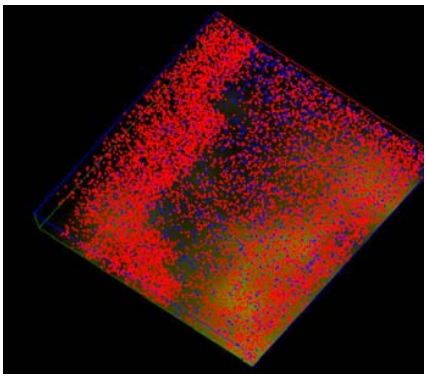

Fig. 25. A sample view of the distribution of chaperones (bluepoints)and nascent proteins (red points), right-folded proteins (yellow points), misfolded proteins of type 1 (green points) and misfolde proiteins of type 2 (magenta points).

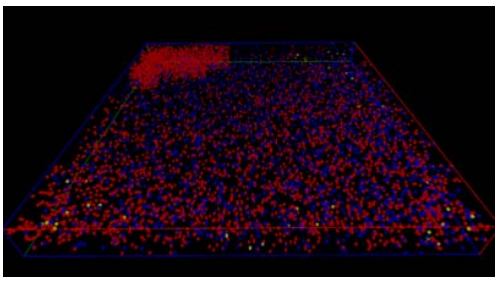

Fig. 26. Another sample view of the distribution of chaperones (bluepoints)and nascent proteins (red points), right-folded proteins (yellow points), misfolded proteins of type 1 (green points) and misfolde proiteins of type 2 (magenta points).

VI. Conclusions and future directions

The authors presented a model for the diffusion of non-charged molecules, in which the diffusion coefficients are not constant with respect to the time and space. Constant diffusion coefficients are rather more the exception than the rule in living cells and, more generally in biological tissues. The authors implemented the procedure in the framework of stochastic simulation of reaction-diffusion systems and presented the results of the method on the case study of chaperone-assisted protein folding. The software tool is equipped with a 3D visualizer that shows the spatial distribution of the diffusion molecules at every step of simulations (see Figs. 25 and 26, showing the distributions from two points of view).

Unlike the previous works as [8], [14], [24], this model provides a theoretical derivation of the molecular origins of the parameters, determining the time-behavior of the diffusive phenomena. Moreover, it provides results in agreement with experimental qualitative and quantitative data. Future work will consist in a further refinement of the procedure to make it closer to the chemistry and physics of biological transport phenomena. Some future directions will consist of a more accurate calculation of the second virial coefficient for biomolecules, especially for proteins. The use of the Lennard-Jones potential is a good approximation of the molecular interaction, but it is a drawback in describing protein-protein interaction is that water molecules must be included explicitly [25], complicating the computational task. The condition of solvated molecules is reflected also to the expression of the concentration-dependence of frictional coefficient, that will need to be accordingly modified.

Furthermore, more generally, as already mentioned, the cellular environment is a crowded solution. Namely, the cellular environments are packed with other biomolecules and this crowdedness may affect the stability and aggregation rates of proteins inside cells [26], [27], [28], [29], [30]. Unlike in typical biochemical experiments in which the proteins of interest are purified and diluted, the living cell is crowded with a wide variety of other proteins and macromolecules which generally occupy 20-30% of the total cell volume. This percentage is called *excluded volume*. The effects imposed by the excluded volume, that is caused by the volume excluded by the “inert” macromolecules, are called *macromolecular crowding* effects and those macromolecules are called *crowding agents*. The authors are currently extending the present study to develope a model whose simulations are of support to the investigation of excluded volume effectson the protein diffusion and folding.

Finally, this algorithm can be incorporated with the time extension of Gillespie algorithm, that the authors developed, in the context of process algebra languages, to treat rate coefficients depending on time [31], [32]. The algorithm which

simulates this diffusion model provides more accurate and realistic results with respect to the algorithm simulating classical Fickian diffusion and can be used to calculate and predict the time-behavior of proteins and biomolecules diffusing in a highly structured and inhomogeneous medium. Very recently, the authors in [33] have designed a parallelization of the simulation of a reaction-diffusion systems to make it more efficient and fast.

REFERENCES

[1] J. Elf, A. Doncic, and M. Ehrenberg, "Mesoscopic reaction-diffusion in intracellular signaling," *Fluctuation and noise in biological, biophysical and biomedical systems. Procs. of SPIE*, vol. 5110, 2003.
[2] P. S. Agutter and D. Wheatley, "Random walks and cell size," *BioEssays*, vol. 22, pp. 1018–1023, 2000.
[3] P. Agutter, P. Malone, and D. Wheatley, "Intracellular transport mechanisms: a critique of diffusion theory," *J. Theor. Biol.*, vol. 176, pp. 261–272, 1995.
[4] D. Fusco, N. Accornero, B. Lavoie, S. Shenoy, J. Blanchard, R. Singer, and E. Bertrand, "Single mrna molecules demonstrate probabilistic movement in living mammallian cells.," *Curr. Biol.*, vol. 13, pp. 161–167, 2003.
[5] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter, *Molecular biology of the cell*. Garland Science, 4th ed. ed., 2003.
[6] E. R. Kandel, "The molecular biology of memory storage: a dialogue between genes and synapses," *Science*, vol. 294, pp. 1030–1038, 2001.
[7] D. Gillespie, "Exact stochastic simulation of coupled chemical reactions," *Journal of Physical Chemistry*, vol. 81, pp. 2340–2361, December 1977.
[8] C. J. Roussel and M. R. Roussel, "Reaction-diffusion models of development with state-dependent chemical diffusion coefficients.," *Progress in Biophysics & Molecular Biology*, 2004.
[9] K. J. Laidler, J. H. meiser, and B. C. Sanctuary, *Physical chemistry*. Houghton Mifflin Company Boston New York, 2003.
[10] M. P. Tombs and A. R. Peacocke, *The Osmotic Pressure of Biological Macromolecules*. Monograph on Physical Biochemistry, Oxford University Press, 1975.
[11] A. Solovyova, P. Schuck, L. Costenaro, and C. Ebel, "Non ideality of sedimantation velocity of halophilic malate dehydrogenase in complex solvent," *Biophysical Journal*, vol. 81, pp. 1868–1880, 2001.
[12] K. Laidler, J. Meiser, and B. Sanctuary, *Physical Chemistry*. Houghton Mifflin Company, 2003.
[13] S. Harding and P. Johnson, "The concentration dependence of macromolecular parameters," *Biochemical Journal*, vol. 231, pp. 543–547, 1985.
[14] D. Bernstein, "Simulating mesoscopic reaction-diffusion systems using the gillespie algorithm," *PHYSICAL REVIEW E*, vol. 71, April 2005.
[15] J. Elf and M. Ehrenberg, "Spontaneous separation of bi-stable biochemical systems into spatial domains of opposite phases," *Syst. Biol.*, vol. 1, December 2004.
[16] A. R. Kinjo and S. Takada, "Competition between protein folding and aggregation wth molecular chaperones in crowed solutions: insight from mesoscopic simulations," *Biophysical Journal*, vol. 85, pp. 3521 – 3531, 2003.
[17] H. S. Chan and K. A. Dill, "A simple model of chepronin-mediated protein folding," *PROTEINS: Structure, Function, and Genetics*, vol. 24, pp. 345–351, 1996.
[18] W. A. Houry, "Chaperone-assisted protein folding," *Curr. protein Pept. Sci.*, vol. 2, no. 3, pp. 227–244, 2001.
[19] J. Frydman and F. U. Hartl, "Principles of chaperone-assisted folding: differences between in vitro and in vivo mechanisms," *Science*, vol. 272, no. 5667, pp. 1497 – 1502, 1996.
[20] T. Langer, J. martin, E. Nimmesgern, and F. U. Hartl, "The pathway of chaperone-assisted protein folding," *Fresenius' Journal of Analytical Chemistry*, vol. 343, 1992.
[21] J. Martin and F. U. Hartl, "The effect of macromolecular crowding on chaperonin-mediated protein folding," *Proc. Natl. Acad. Sci. USA*, vol. 94, pp. 1107–1112, 1997.
[22] D. Thirulamai and G. H. Lorimer, "Chaperonin-mediated protein folding," *Ann. Rev. Biophys. Biomol. Struct.*, vol. 30, pp. 245–268, 2001.
[23] D. Thirumalai and G. H. Lorimer, "Chaperonin-mediated protein folding," *Annu. Rev. Biophys. Biomol. Struct.*, vol. 30, p. 245:269, 2001.
[24] S. A. Isaacson and C. S. Peskin, "Incorporating diffusion in complex geometries into stochastic chemical kinetics simulations," *SIAM Journal of Scientific computing*, pp. 47–74, 2006.
[25] B. L. Neal, D. Asthagiri, and A. M. Lenhoff, "Molecular origins of osmotic second virial coefficients of proteins," *Biophysical Journal*, vol. 75, 1998.
[26] A. R. Kinjo and S. Takada, "Effects of macromolecular crowding on protein folding and aggregation studied bu density functional theory: statics," *Physical Review E*, vol. 66, pp. 031911: 1–9, 2002.
[27] A. R. Kinjo and S. Takada, "Effects of macromolecular crowding on protein folding and aggregation studied by density functional theory: Dynamics," *Physical review. E*, vol. 66, no. 5, pp. 051902.1–051902.10, 2002.
[28] G. Y. G. Ping and J. M. Yuan, "Depletion force from macromolecular crowding enhances mechanicsl stability of protein molecules," *Polymer*, vol. 27, p. 2564:2570, 2006.
[29] A. P. Minton, "Molecular crowding: analysis of effects of high concetrations of inert cosolutes on biochemical equilibria and rates in terms of volume exclusion," *Methods Enzymol.*, vol. 295, p. 127:149, 1998.
[30] A. P. Minton, "The influence of macromolecular crowding and macromolecular confinement on biochemical reactions in physiological media," *J. Biol. Chem.*, vol. 276, p. 10577:10580, 2001.
[31] P. Lecca, "A time-dependent extension of gillespie algorithm for biochemical stochastic π-calculus," *Proceedings of the 2006 ACM symposium on Applied computing*, 2001.
[32] P. Lecca, "Simulating the cellular passive transport of glucose using a time-dependent extension of gillespie algorithm for stochastic π-calculus.," *International Journal of Data Mining and Bioinformatics*, vol. 1, no. 4, pp. 315–336, 2007.
[33] L. Dematté and T. Mazza, "On parallel stochastic simulation of diffusive systems," in *Sixth International Conference on Computational Methods in Systems Biology*, 2008.

Paola Lecca received the Master's Degree in Theoretical Physics at the University of Trento (Italy) in 1997 and a PhD in Computer Science in 2006 at the International Doctorate School in Information and Communication Technologies of University of Trento. Her current research interests in the areas of systems biology and computational cell biology include issues related to conceptual frameworks of stochasticity in modelling and simulating biochemical networks dynamics, model's structure and model's parameter inference for optimal experimental design.

Lorenzo Dematté is currently a Doctorate Student of the International Doctoral School in Informationa and Communication technology of the University of Trento. He obtained the Master's degree (2005) in Computer Science. His research interests are in computational models for systems biology, languages for concurrent and distributed systems, and concurrency theory.