Chemical Master Equations
A Mathematical Scheme for the Multi-site Phosphorylation Case[*]

Alessandro Borri, Francesco Carravetta, Gabriella Mavelli and Pasquale Palumbo
Istituto di Analisi dei Sistemi ed Informatica "A. Ruberti", Consiglio Nazionale delle Ricerche (IASI-CNR), Viale Manzoni 30, 00185 Roma, Italy

Keywords: Systems Biology, Chemical Master Equation, Markov Processes, Enzymatic Reactions.

Abstract: The Chemical Master Equation (CME) provides a fruitful approach for the stochastic description of complex biochemical processes, because it is able to cope with random fluctuations of the chemical agents and to fit the experimental behavior in a more accurate way than deterministic concentration equations. In this work, our attention is focused on modeling and simulation of multisite phosphorylation/dephosphorylation cycles, by using the quasi-steady state approximation of enzymatic kinetics. The CME dynamics is written from the coefficients of the deterministic reaction-rate equations and the stationary distribution is computed explicitly, according to a recently developed realization scheme. Simulations are included to provide a comparison with Monte Carlo methods in terms of computational complexity.

1 INTRODUCTION

Understanding complex cellular processes is a major challenge facing present-day biology. Systems Biology is an approach that tackles the study of such complex systems and is characterized by the fact that molecular analysis - often at the genomic scale - is integrated with modeling, parameter identification, simulation analysis and control theory. Within this approach, biological processes are taken to be the results of complex, coordinated, dynamic, nonlinear interactions of a large number of components, which are affected by time and space constraints. The development of efforts in the Systems Biology research field will allow to elucidate relevant central features of the more basic complex cellular functions (such as cycle, growth, and so on) as well as to identify their system-level properties, ultimately contributing to shed more light on diseases arising as the impairment metabolic functions such as cancer, diabetes, etc. To this aim, a successful analysis of the complex mechanisms arising in System Biology requires the application of more and more efficient methods for modeling such complex cellular systems.

As far as the networks of biochemical reactions, there is an increasing interest in the development and analysis of stochastic models, due to the preminent role of fluctuations whenever a low number of molecules is involved, (Mettetal and van Oudenaarden, 2007). Indeed, in a single cell environment, the concentration of molecules can be low, so molecular species can be found with a number of copies ranging from hundreds to thousands. The small number of molecules involved and the inherent stochasticity of biochemical processes justify the necessity of the stochastic approach to cope with intracellular processes, (Van Kampen, 2007).

The biochemical complex machinery is largely based on enzymatic reactions. Among the covalent enzymatic-catalyzed modifications of substrates, phosphorylation and dephosphorylation reactions catalyzed by kinases and phosphatases, respectively, provide a reversible post-translational substrate modification that is central to cellular signalling and regulation. In (Qu et al., 2003) it has been proposed that multisite phosphorylation of target proteins by cyclin-dependent kinase proteins is the likely source of nonlinear kinetic effects in cell cycle control mechanisms. In particular, dual phosphorylation/dephosphorylation cycles are widely diffused within cellular systems and are crucial for the control of complex responses such as learning, memory, and cellular fate determination. The majority of studies on biophysical analysis of phosphorylation/dephosphorylation cycle have been performed in an averaged, deterministic framework

[*]This work is partially supported by project INTEROMICS.

Borri A., Carravetta F., Mavelli G. and Palumbo P..
Chemical Master Equations - A Mathematical Scheme for the Multi-site Phosphorylation Case.
DOI: 10.5220/0004632306810688
In *Proceedings of the 3rd International Conference on Simulation and Modeling Methodologies, Technologies and Applications* (BIOMED-2013), pages 681-688
ISBN: 978-989-8565-69-3

based on the Michaelis-Menten approach (Michaelis and Menten, 1913). As stated above about generic molecules, indeed in a single cell the substrate and enzyme concentration can be low, being thus necessary to study these cases within a stochastic framework. Examples of biochemical reactions with low number of substrates and involved enzymes present at low concentrations are the phosphorylation/dephosphorylation cycle of AMPA-Receptors (AMPARs) in a single synaptic spine and have been treated in (Castellani et al., 2001; Whitlock et al., 2006). In (Castellani et al., 2001) it has been proposed a synaptic receptors model (AMPAR model) which, at a molecular-microscopic level, is a candidate for memory storage and switching behavior, and a kinetic scheme of the enzymatic reactions describing the AMPAR double phosphorylation channel has been given. It has been stressed that the AMPAR modification, a 2-step phospho/dephosphorylation cycle, guaranties the synaptic plasticity and exhibits bistability for a wide range of parameters, consistent with values derived from biological literature.

In this paper we investigate the Chemical Master Equation (CME) approach applied to model in a stochastic framework the phosphorylation/dephosphorylation cycle. Indeed, the CMEs provides an appropriate description of complex cellular processes, (Van Kampen, 2007; Gillespie, 1977). It is a powerful method even with significant simplifications, such as spatial homogeneity of volumes where the chemical reactions under investigation are taking place. The CME is a promising approach for Systems Biology modelling for a variety of reasons, such as its capability of coping with fluctuations and chemical fluxes and of well fitting experimental data in the today widespread single cell experiments, and its capability to capture and explain the deviation from Gaussianity observed in various gene expression experiments (such as stress or metabolic response, growth of the nuclear protein amount observed in senescent cells, and so on). It results more informative of the real behavior of the system than the deterministic approach, since the results obtained in a deterministic framework cannot describe diffusion effects due to fluctuations and chemical currents capable to drive switching from one equilibrium to the other. Such mechanism has a relevant role also in neuronal plasticity phenomena, (Castellani et al., 2009). In (Bazzani et al., 2012) the stationary distribution provided by the two-dimensional chemical master equation for a well-known model of a two step phosphorylation/dephosphorylation cycle has been studied, by using the quasi-steady state approximation of enzymatic kinetics. Authors pointed out the possibility for the molecule distribution shape to be controlled (in particular changed from a unimodal distribution to a bimodal distribution) by chemical fluxes occurring in the biochemical system under investigation. This phenomenon corresponds to the typical mechanism that is adopted by such system to obtain a plastic behavior, for example when any kind of exogenous chemical messenger is present, by changing the properties of the biochemical reaction occurring inside the cell.

The paper is organized as follows: in Section 2 we review some structural properties of the Master Equation and address the efficient computation of its equilibrium solution. In Section 3, we illustrate the case study of multiple phosphorylation for an arbitrary number of sites. In Section 4 we propose some simulation results concerning the case of a triple phosphorylation framework. Section 5 offers concluding remarks.

2 THE CHEMICAL MASTER EQUATION

Consider M (bio)chemical species $Y_1,...,Y_M$ involved in q chemical reactions (Ullah and Wolkenhauer, 2011) described by the *stoichiometric coefficients* β_{ij}, for any species $j = 1,\dots,M$ and for any reaction $i = 1,\dots,q$. The Chemical Master Equation (CME) describes the time evolution of the probability of being in one of the $N \leq +\infty$ possible configurations (states) at any time (Van Kampen, 2007). More generally, the CME is the dynamic equation of a continuous-time discrete-state stochastic Markov process. If N is finite, the equations for the joint probabilities can be collected in the form of an autonomous positive linear system (Farina and Rinaldi, 2000):

$$\dot{\mathcal{P}} = G\mathcal{P}, \qquad \mathcal{P} \in \mathbb{R}^N, \tag{1}$$

where G is called *infinitesimal generator* of the Markov process.

We define as $n(t) = (n_1(t),\dots,n_M(t))$ the state of the system at time t, with i-th component $n_i(t) \in \mathbb{N}_0$ denoting the number of copies of species Y_i at time t. The generic element G_{ij} of G is the *propensity* $g^{\alpha_1,\dots,\alpha_M}_{n_1,\dots,n_M}$ of reaching the state $v_j = (n_1+\alpha_1,\dots,n_M+\alpha_M)$ from the state $v_i = (n_1,\dots,n_M)$.

Depending on whether or not the system of reactions is *closed* (the total mass is conserved) and according to the number of distinct *elements* forming the M species, the values of some components of the vector state $n(t)$ can be inferred from the others, which makes the state representation redundant. We

will henceforth assume that the M species are *independent*, i.e. the state representation is minimal (non-redundant). The interested reader is referred to (Borri et al., 2013) for a deeper discussion on these and on the following concepts.

Note that different ways of aggregation of the probabilities in the vector $\mathcal{P}$ may produce different properties of the linear system in (1). In particular, one can consider the following recursive partitioning:

$$\mathcal{P}_{n_1,\dots,n_{M-1}} \doteq \begin{bmatrix} p_{n_1,\dots,n_{M-1},0} \\ p_{n_1,\dots,n_{M-1},1} \\ \vdots \\ p_{n_1,\dots,n_{M-1},N_M} \end{bmatrix} \in \mathbb{R}^{N_M+1}, \quad (2)$$

with $0 \le n_i \le N_i$, with N_i being the maximum number of allowed copies of species Y_i, and $p_{n_1,n_2,\dots,n_M}$ being the joint probability of having n_i copies of species Y_i, for all $i = 1,\dots,M$. Then, the following vectors of probabilities can be recursively defined

$$\mathcal{P}_{n_1,\dots,n_i} \doteq \begin{bmatrix} \mathcal{P}_{n_1,\dots,n_i,0} \\ \mathcal{P}_{n_1,\dots,n_i,1} \\ \vdots \\ \mathcal{P}_{n_1,\dots,n_i,N_{i+1}} \end{bmatrix} \in \mathbb{R}^{(N_{i+1}+1)\times\cdots\times(N_M+1)}, \quad (3)$$

for $1 \le i \le M-2$. This leads to the definition of vector $\mathcal{P}$, entailing all the joint probabilities involved by the CME:

$$\mathcal{P} \doteq \begin{bmatrix} \mathcal{P}_0 \\ \mathcal{P}_1 \\ \vdots \\ \mathcal{P}_{N_1} \end{bmatrix} \in \mathbb{R}^{(N_1+1)\times\cdots\times(N_M+1)}. \quad (4)$$

The way of collecting the probabilities in $\mathcal{P}$ as in (2)–(4) induces a recursive block partitioning on matrix G, according to (1). In case the stoichiometric coefficients satisfy the *unitary step* property:

$$\beta_{ij} \in \{-1,0,1\}, i \in \{1,\dots,q\}, j \in \{1,\dots,M\}, \quad (5)$$

an interesting recursive structure of the CME dynamic equations can be pointed out:

Proposition 1. *Consider a continuous-time Markov process describing a set of q chemical reactions involving M independent (bio)chemical species and described by a set of stoichiometric coefficients $\{\beta_{ij}\}$ satisfying (5). Then the infinitesimal generator G has the following block-tridiagonal structure (see e.g. (Carravetta, 2011)):*

$$G = \Phi_{N_1}\left(\{G^1_{n_1}\},\{G^0_{n_1}\},\{G^{-1}_{n_1}\};n_1\right) = \begin{bmatrix} G^0_0 & G^{-1}_1 & \emptyset & \cdots & \cdots & \emptyset \\ G^1_0 & G^0_1 & G^{-1}_2 & \emptyset & \cdots & \emptyset \\ \emptyset & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \emptyset \\ \vdots & \ddots & \ddots & G^1_{N_1-2} & G^0_{N_1-1} & G^{-1}_{N_1} \\ \emptyset & \cdots & \cdots & \emptyset & G^1_{N_1-1} & G^0_{N_1} \end{bmatrix}, \quad (6)$$

where the $\emptyset$ entries are zero matrices of proper dimensions. Furthermore, all the non-zero blocks of G are block-tridiagonal with the same structure of G.

The operator $\Phi_\nu(\{A_i\},\{B_i\},\{C_i\};i)$ used in (6) is called *block-tridiagonal matrix builder* and is defined as follows

$$\Phi_\nu(\{A_i\},\{B_i\},\{C_i\};i) := \begin{bmatrix} B_0 & C_1 & \emptyset & \cdots & \emptyset \\ A_0 & B_1 & C_2 & \emptyset & \emptyset \\ \emptyset & \ddots & \ddots & \ddots & \emptyset \\ \vdots & \ddots & A_{\nu-2} & B_{\nu-1} & C_\nu \\ \emptyset & \cdots & \emptyset & A_{\nu-1} & B_\nu \end{bmatrix}, \quad (7)$$

where $\{A_i\}$, $\{B_i\}$, $\{C_i\}$ are sequences of suitably dimensioned square matrices, and the $\emptyset$ entries are zero matrices of proper dimensions. A network of chemical reactions satisfying (5) allows simultaneous changes of unitary amount in the state variables and is later referred as *generalized* one-step process or *unitary* process. It is possible to show that (5) is the mildest condition preserving the recursive block-tridiagonal structure of G.

The main consequence of Proposition 1 is that G is sparse, which reveals to be useful to carry out classical tasks as the computation of the stationary solution of the CME, satisfying

$$G\mathcal{P} = \mathbf{0}. \quad (8)$$

The following result, based on an interpretation of the biochemical network in terms of *Algebraic Graph Theory* (Bullo et al., 2009), characterizes the property of *uniqueness* of the CME stationary distribution, which makes the study of the equilibrium problem (8) independent of the initial state.

Proposition 2. *The stationary distribution $\mathcal{P}_{ss}$ of a continuous-time Markov process is unique if and only if the digraph associated with the Markov process has a globally reachable vertex. Under this assumption, 0 is a simple eigenvalue of G with right eigenvector u_0 and the stationary distribution is given by $\mathcal{P}_{ss} = \frac{u_0}{\mathbf{1}^T u_0}$.*

As an alternative approach to the computation of $\mathcal{P}_{ss}$ in terms of the right eigenvector u_0 of G, Propositions 1 and 2 allow an efficient computation of the solution of (8) by means of Gaussian elimination:

Proposition 3. *The algorithm of Gaussian elimination to solve the stationary equation $G\mathcal{P} = \mathbf{0}$ in unitary-step processes is performed in time $O(N^2)$, where N is the number of rows of G.*

Note that the complexity of the Gaussian Elimination for dense matrices is $O(N^3)$.

3 MULTI-SITE PHOSPHORYLATION

This section is devoted to formalizing, in the CME mathematical setting, the multi-site phosphorylation reaction framework.

3.1 Writing the Reaction-rate Equations

Consider the case of only $M = 2$ phosphorylation sites (see e.g. (Bazzani et al., 2012)), depicted in Fig. 1, involving $M+1 = 3$ species and 2 enzymes (kinase E_1 and phosphatase E_2). Such a framework is formalized by the following set of $q = 2M = 4$ reactions:

$$\begin{cases} A^{(0)} & \longrightarrow & A^{(1)}, \\ A^{(1)} & \longrightarrow & A^{(2)}, \\ A^{(1)} & \longrightarrow & A^{(0)}, \\ A^{(2)} & \longrightarrow & A^{(1)}, \end{cases}$$

where the species $A^{(0)}$, $A^{(1)}$, $A^{(2)}$ are the non-phosphorylated, phosphorylated and double-phosphorylated substrate, respectively. If one considers the double-site phosphorylation as a closed system (the total substrate concentration is constant $[A^{(0)}] + [A^{(1)}] + [A^{(2)}] = A_{tot}$), the number of independent species is $M = 2$.

We now address the general M-site phosphorylation case, illustrated in Fig. 2. We can consider the case of closed system, so that the species $A^{(i)}$ may be intended to be the i-times phosphorylated substrate, for $i = 0,...,M$, and the total substrate concentration is constant $\sum_{i=0}^{M}[A^{(i)}] = A_{tot}$. The *odd-indexed* reactions are catalyzed by the kinase E_1:

$$[A^{(i-1)}] + [E_1] \underset{K_{OFF}^{2i-1}}{\overset{K_{ON}^{2i-1}}{\rightleftarrows}} [A^{(i-1)}E_1] \xrightarrow{v_{2i-1}} [A^{(i)}] + [E_1] \tag{9}$$

with $i = 1,...,M$, while the *even-indexed* reactions are activated by the phosphatase E_2:

$$[A^{(i)}] + [E_2] \underset{K_{OFF}^{2i}}{\overset{K_{ON}^{2i}}{\rightleftarrows}} [A^{(i)}E_2] \xrightarrow{v_{2i}} [A^{(i-1)}] + [E_2] \tag{10}$$

with $i = 1,...,M$, for a total number of reactions equal to $q = 2M$.

The left-hand M pairs of reactions in (9)–(10) are considered at the equilibrium (*quasi-steady-state hypothesis*), thus the corresponding $2M$ deterministic Michaelis-Menten (MM) equations rewrite:

$$\frac{d[A^{(i-1)}E_1]}{dt} = K_{ON}^{2i-1}[A^{(i-1)}][E_1] - K_{OFF}^{2i-1}[A^{(i-1)}E_1] = 0$$
$$\frac{d[A^{(i)}E_2]}{dt} = K_{ON}^{2i}[A^{(i)}][E_2] - K_{OFF}^{2i}[A^{(i)}E_2] = 0$$

for $i = 1,...,M$. Hence, by setting the *dissociation coefficients* $k_j := K_{OFF}^{j}/K_{ON}^{j}$, $j = 1,...,2M$, one gets the equilibrium conditions:

$$\begin{cases} [A^{(i-1)}E_1]_{eq} & = \frac{[A^{(i-1)}][E_1]}{k_{2i-1}} \\ [A^{(i)}E_2]_{eq} & = \frac{[A^{(i)}][E_2]}{k_{2i}} \end{cases}$$

for $i = 1,...,M$. We now focus on the right-hand reactions in (9)–(10), where we replace the terms $[A^jE_k]$ with their equilibrium values $[A^jE_k]_{eq}$, for $j = 1,...,2M$ and $k = 1,2$. One gets:

$$\frac{[A^{(i-1)}][E_1]}{k_{2i-1}} \xrightarrow{v_{2i-1}} [A^{(i)}] + [E_1], \tag{11}$$
$$\frac{[A^{(i)}][E_2]}{k_{2i}} \xrightarrow{v_{2i}} [A^{(i-1)}] + [E_2]. \tag{12}$$

Note that the generic species $A^{(i)}$, for $i = 1,...,M$, appears in 4 distinct reactions:

$$\frac{[A^{(i-1)}][E_1]}{k_{2i-1}} \xrightarrow{v_{2i-1}} [A^{(i)}] + [E_1],$$
$$\frac{[A^{(i)}][E_2]}{k_{2i}} \xrightarrow{v_{2i}} [A^{(i-1)}] + [E_2],$$
$$\frac{[A^{(i)}][E_1]}{k_{2i+1}} \xrightarrow{v_{2i+1}} [A^{(i+1)}] + [E_1],$$
$$\frac{[A^{(i+1)}][E_2]}{k_{2(i+1)}} \xrightarrow{v_{2(i+1)}} [A^{(i)}] + [E_2],$$

which lead to the following MM equation, for any i:

$$\frac{d[A^{(i)}]}{dt} = \frac{v_{2i-1}}{k_{2i-1}}[E_1][A^{(i-1)}] - \frac{v_{2i}}{k_{2i}}[E_2][A^{(i)}] - \frac{v_{2i+1}}{k_{2i+1}}[E_1][A^{(i)}] + \frac{v_{2(i+1)}}{k_{2(i+1)}}[E_2][A^{(i+1)}]. \tag{13}$$

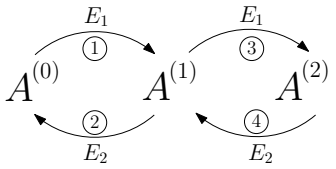

Figure 1: Double-site Phosphorylation.

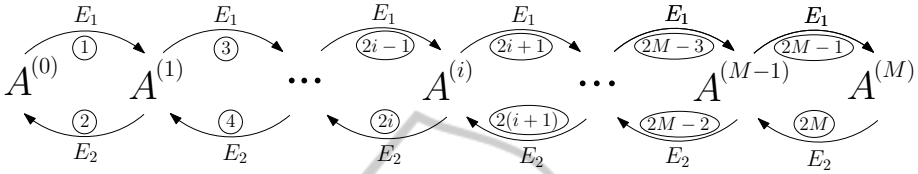

Figure 2: Multi-site Phosphorylation.

Note that $i = 0$, $i = M$ are special cases because some terms are neglected:

$$\frac{d[A^{(0)}]}{dt} = -\frac{v_1}{k_1}[E_1][A^{(0)}] + \frac{v_2}{k_2}[E_2][A^{(1)}], \tag{14}$$

$$\frac{d[A^{(M)}]}{dt} = \frac{v_{2M-1}}{k_{2M-1}}[E_1][A^{(M-1)}] - \frac{v_{2M}}{k_{2M}}[E_2][A^{(M)}]. \tag{15}$$

The following mass-balance constraints on the total substrate concentration and on the enzymes:

$$\begin{cases} \sum_{i=0}^{M}[A^{(i)}] = A_{tot}, \\ [E_1] + \sum_{i=1}^{M}\frac{[A^{(i-1)}]}{k_{2i-1}}[E_1] = E_{1,tot}, \\ [E_2] + \sum_{i=1}^{M}\frac{[A^{(i)}]}{k_{2i}}[E_2] = E_{2,tot}, \end{cases} \tag{16}$$

lead to the a minimal representation of the process by means of M independent species. For instance, we choose $x_i = [A^{(i)}]$ for $i = 1,...,M$ and, from (13), we are able to describe the system in terms of M coupled differential equations, called *reaction-rate equations*:

$$\begin{cases} \dot{x}_1 = \frac{v_1}{k_1}E_1(x)x_0 - \left(\frac{v_2}{k_2}E_2(x) + \frac{v_3}{k_3}E_1(x)\right)x_1 \\ \quad + \frac{v_4}{k_4}E_2(x)x_2 \\ \vdots \\ \dot{x}_i = \frac{v_{2i-1}}{k_{2i-1}}E_1(x)x_{i-1} - \left(\frac{v_{2i}}{k_{2i}}E_2(x) + \frac{v_{2i+1}}{k_{2i+1}}E_1(x)\right)x_i \\ \quad + \frac{v_{2(i+1)}}{k_{2(i+1)}}E_2(x)x_{i+1} \\ \vdots \\ \dot{x}_M = \frac{v_{2M-1}}{k_{2M-1}}E_1(x)x_{M-1} - \frac{v_{2M}}{k_{2M}}E_2(x)x_M \end{cases} \tag{17}$$

where $x_0 := [A^{(0)}]$ and $E_1(x)$, $E_2(x)$ are given by:

$$\begin{cases} x_0 = 1 - \sum_{i=1}^{M} x_i, \\ E_1(x) = \frac{E_{1,tot}}{1 + \sum_{i=1}^{M}\frac{x_{i-1}}{k_{2i-1}}}, \\ E_2(x) = \frac{E_{2,tot}}{1 + \sum_{i=1}^{M}\frac{x_i}{k_{2i}}}. \end{cases} \tag{18}$$

3.2 From the Reaction-rate Equations to the Master Equation

In (17)–(18), we built a minimal deterministic representation of the M-site phosphorylation by means of a system of M ordinary differential equations (ODEs) with respect to the M variables $x_1,...,x_M$, where x_i represents the normalized concentration of species Y_i, i.e. of the i-times phosphorylated substrate. It is deeply discussed in the related literature how the reaction-rate approach fails to capture the inherent randomness of biochemical phenomena, especially in the presence of a low number of molecules (Van Kampen, 2007). In the following, we build a stochastic representation of the chemical process in terms of Master Equations.

A common approach to infer the CMEs from the ODEs consists in turning the *real-valued* concentration variables x_i into the integer values n_i, representing the *copies* of species Y_i, according to the mass-balance constraint:

$$n_0 = n_{TOT} - \sum_{i=1}^{M} n_i.$$

As a first step, we need to write the stoichiometric coefficients for each reaction.

Odd-indexed Reactions. The generic reaction $2i-1$, for $i = 1,...,M$, "generates" a molecule of species Y_i and "annihilates" a molecule of species Y_{i-1}, from (11). Hence the vector of stoichiometric

coefficients $\bar{\beta}_{2i-1} = (\beta_{2i-1,1}, ..., \beta_{2i-1,M})$ for these reactions is given by:

$$\beta_{2i-1,j} = \begin{cases} 1 & j = i, \\ -1 & j = i-1, \\ 0 & \text{otherwise.} \end{cases} \quad (19)$$

Even-indexed Reactions. The generic reaction $2i$, for $i = 1, ..., M$, reverses the reaction $2i-1$, namely it generates a molecule of species Y_{i-1} and annihilates one of species Y_i, from (12). Hence the vector of stoichiometric coefficients $\beta_{2i} = (\beta_{2i,1}, ..., \beta_{2i,M})$ for these reactions is given by:

$$\beta_{2i,j} = \begin{cases} -1 & j = i, \\ 1 & j = i-1, \\ 0 & \text{otherwise.} \end{cases} \quad (20)$$

Writing the Matrix G. We are now ready to write the CME by computing the generic element g_n^α (propensity), with the shortcuts $n = (n_1, ..., n_M)$ and $\alpha = (\alpha_1, ..., \alpha_M)$. By comparison with the equation of species i in (17):

$$\dot{x}_i = \frac{v_{2i-1}}{k_{2i-1}} E_1(x) x_{i-1} - \left(\frac{v_{2i}}{k_{2i}} E_2(x) + \frac{v_{2i+1}}{k_{2i+1}} E_1(x) \right) x_i + \frac{v_{2(i+1)}}{k_{2(i+1)}} E_2(x) x_{i+1} \quad (21)$$

and taking into account the definition of the vectors of stoichiometric coefficients in (19)–(20), one can define:

$$g_n^\alpha := \begin{cases} \frac{v_{2i-1}}{k_{2i-1}} E_1(n) n_{i-1} & \alpha = \bar{\beta}_{2i-1}, i = 1, ..., M, \\ \frac{v_{2(i+1)}}{k_{2(i+1)}} E_2(n) n_{i+1} & \alpha = \bar{\beta}_{2(i+1)}, i = 1, ..., M, \\ -\sum_{i=1}^{M} \left(\frac{v_{2i}}{k_{2i}} E_2(n) + \frac{v_{2i+1}}{k_{2i+1}} E_1(n) \right) n_i & \alpha = \mathbf{0}, \\ 0 & \text{otherwise,} \end{cases} \quad (22)$$

where $\mathbf{0} = (0, ..., 0)$. Intuitively, the definition in (22) is obtained from (21) by regarding the positive addends as *generation terms* (i.e. creating a molecule) and the negative ones as *recombination terms* (i.e. annihilating a molecule) of the associated biochemical reactions. The dynamic matrix G of the CME can be written as in (6), where the scalar blocks are defined from the propensities as:

$$G_{n_1,...,n_M}^{\alpha_1,...,\alpha_M} = \begin{cases} g_{n_1,...,n_M}^{\alpha_1,...,\alpha_M} & \text{if } (\alpha_1, \ldots, \alpha_M) \neq \mathbf{0}, \\ -\sum_{\bar{\alpha}_1,...,\bar{\alpha}_M} g_{n_1,...,n_M}^{\bar{\alpha}_1,...,\bar{\alpha}_M} & \text{otherwise,} \end{cases} \quad (23)$$

and the general blocks are recursively defined in terms of the matrix builder in (7) as:

$$G_{n_1,...,n_i}^{\alpha_1,...,\alpha_i} = \Phi_{N_{i+1}} \Big(\{ G_{n_1,...,n_i,n_{i+1}}^{\alpha_1,...,\alpha_i,1} \}, \{ G_{n_1,...,n_i,n_{i+1}}^{\alpha_1,...,\alpha_i,0} \}, \{ G_{n_1,...,n_i,n_{i+1}}^{\alpha_1,...,\alpha_i,-1} \}; n_{i+1} \Big) \quad (24)$$

for $1 \leq i \leq M-1$, $0 \leq n_i \leq N_{TOT}$, $0 \leq |\alpha_i| \leq N_{TOT}$.

It can be shown (see e.g. (Van Kampen, 2007)) that the deterministic reaction-rate approach is a first-order approximation of the dynamics of the mean value of the CME.

4 SIMULATION RESULTS

In the following, we restrict our attention to the case $M = 3$ (triple phosphorylation) and we compare the performance between computing the stationary distribution explicitly, based on the results in Section 2, and obtaining it by means of statistical methods, in particular the Gillespie Stochastic Simulation Algorithm (SSA) (Gillespie, 1977).

For $M = 3$, we have $q = 6$ reactions. We assume $n_{TOT} = 40$ and $E_{1,TOT} = E_{2,TOT} = 1$. The dimension of G is $N = 41^3 > 64{,}000$, but an inspection of the associated states reveals that a large subset of those are not reachable (in the sense of Munsky et al. (Munsky and Khammash, 2008)) because they violate the mass-balance constraints, hence they have a zero steady-state probability. If one erases *a priori* the components referred to those states from the master equation, a reduced matrix $\tilde{G}$ is obtained, whose kernel is 1-dimensional. This ensures the uniqueness of the stationary distribution $\tilde{\mathcal{P}}_{ss}$, defined over the set of reachable states of the Markov Process. In this particular case, just $\tilde{N} = 12{,}341$ configurations are admissible states.

The values of the chosen parameters are: $k_1 = k_6 = 1$, $k_2 = k_3 = 1.8$, $k_4 = k_5 = 2.2$, $v_1 = v_2 = 1$, $v_3 = v_4 = 1.1$, $v_5 = v_6 = 1.2$. We apply the formalization in Section 3 and perform numerical simulations in the Matlab suite on an Apple MacBook Pro laptop with 2.5 GHz Intel Core i5 CPU and 16 GB RAM.

We started with the stochastic simulation. We considered as initial state $n(0) = (0,0,0)$, meaning that $n_0 = n_{TOT}$, i.e. all the molecules are non-phoshorylated at the beginning. We performed 10^4 Monte Carlo runs of SSA, with a time horizon of 500 seconds. The average number of reactions for each SSA run was 9925. The whole computation took more than 12 hours.

We then compared 3 methods of computing the theoretical stationary distribution:

(1) We computed the eigenvalues and eigenvectors of matrix $\tilde{G}$ by means of Matlab's *eig* routine. Ac-

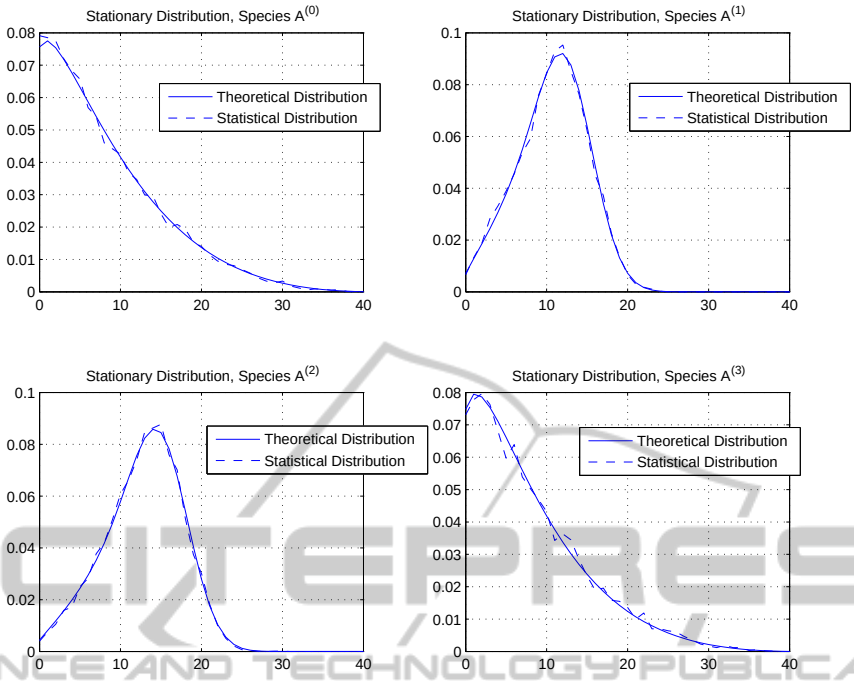

Figure 3: Steady-state marginal distributions. The solid line is the theoretical distribution computed explicitly according to the methods illustrated in the paper. The dashed line is the statistical distribution provided by 5,000 Monte Carlo runs of the Gillespie Algorithm.

cording to Proposition 2, the stationary distribution is the right eigenvector corresponding to the zero eigenvalue of $\tilde{G}$. The procedure took 2785 seconds.

(2) According to Proposition 3, we computed $\tilde{\mathcal{P}}_{ss}$ by means of an *ad hoc* implementation of the Gaussian Elimination method to compute the solution of $\tilde{G}\tilde{\mathcal{P}} = \mathbf{0}$. Although it is very accurate, this method was outperformed (in terms of time complexity) by method (1) and the elapsed time is comparable to the Gillespie method.

(3) We computed the matrix exponential $e^{\tilde{G}}$, by means of the Matlab routine *expm*, and we considered as initial condition $\tilde{\mathcal{P}}(0)$ the probability vector with value equal to 1 for the component referred to the state $n(0)$ and 0 elsewhere. We then computed the steady-state distribution by means of a fixed-point iteration for the recursion $p_{k+1} = e^{\tilde{G}} p_k$, initialized with $p_0 = \tilde{\mathcal{P}}(0)$, which is the exact discretization of the CME with unitary sampling time, and where we used the stopping condition $\|p_{k+1} - p_k\| \leq 10^{-6}$. Most of the time in this case is spent in computing the matrix exponential, computed in 357 seconds. The fixed point iteration took 88 seconds to reach the steady state.

The methods above return the same distribution, with a relative difference lower than 10^{-6} (in norm). The plots in Fig. 3 show the agreement between the statistical estimation of the steady-state solution and the theoretical stationary distribution. Due to the saving in terms of computational resources, the explicit computation is preferable.

5 CONCLUSIONS

In this work we illustrated the application of some results about the efficient computation of the stationary distribution of the Chemical Master Equation in the biochemical framework of multisite phosphorylation. Based on a recently developed representation scheme, it is shown how the proposed approach results to be accurate and with limited computational burden, especially with respect to the extensive use of statistical Monte Carlo methods.

REFERENCES

Bazzani, A., Castellani, G. C., Giampieri, E., Remondini, D., and Cooper, L. N. (2012). Bistability in the chemical master equation for dual phosphorylation cycles. *The Journal of Chemical Physics*, 136(23):235102.

Borri, A., Carravetta, F., Mavelli, G., and Palumbo, P. (2013). Some results on the structural properties and

the solution of the chemical master equation. In *Proceedings of the 2013 American Control Conference (ACC 2013)*, Washington, DC, USA.

Bullo, F., Cortes, J., and Martinez, S. (2009). *Distributed Control of Robotic Networks*. Applied Mathematics Series. Princeton.

Carravetta, F. (2011). 2-d-recursive modelling of homogeneous discrete gaussian markov fields. *IEEE Transactions on Automatic Control*, 56(5):1198–1203.

Castellani, G. C., Bazzani, A., and Cooper, L. N. (2009). Toward a microscopic model of bidirectional synaptic plasticity. *Proceedings of the National Academy of Sciences*, 106(33):14091–14095.

Castellani, G. C., Quinlan, E. M., Cooper, L. N., and Shouval, H. Z. (2001). A biophysical model of bidirectional synaptic plasticity: Dependence on ampa and nmda receptors. *Proceedings of the National Academy of Sciences*, 98(22):12772–12777.

Farina, L. and Rinaldi, S. (2000). *Positive Linear Systems: Theory and Applications*. Pure and Applied Mathematics Series. John Wiley & Sons, Inc.

Gillespie, D. T. (1977). Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry*, 81(25):2340–2361.

Mettetal, J. T. and van Oudenaarden, A. (2007). Necessary noise. *Science*, 317(5837):463–464.

Michaelis, L. and Menten, M. (1913). Kinetics of invertase action. *Biochem. Z*, 49:333–369.

Munsky, B. and Khammash, M. (2008). The finite state projection approach for the analysis of stochastic noise in gene networks. *IEEE Transactions on Automatic Control*, 53(Special Issue on Systems Biology):201–214.

Qu, Z., Weiss, J. N., and MacLellan, W. R. (2003). Regulation of the mammalian cell cycle: a model of the g1-to-s transition. *American Journal of Physiology - Cell Physiology*, 284(2):C349–C364.

Ullah, M. and Wolkenhauer, O. (2011). *Stochastic Approaches for Systems Biology*. Springer.

Van Kampen, N. G. (2007). *Stochastic Processes in Physics and Chemistry, Third Edition*. North Holland.

Whitlock, J. R., Heynen, A. J., Shuler, M. G., and Bear, M. F. (2006). Learning induces long-term potentiation in the hippocampus. *Science (New York, N.Y.)*, 313(5790):1093–1097.