Exact and Approximate Stochastic Simulations of the MAPK Pathway and Comparisons of Simulations' Results

Vilda Purutçuoğlu, Ernst Wit

Department of Mathematics and Statistics, Lancaster University

Summary

The MAPK (mitogen-activated protein kinase) or its synonymous ERK (extracellular signal regulated kinase) pathway whose components are Ras, Raf, and MEK proteins with many biochemical links, is one of the major signalling systems involved in cellular growth control of eukaryotes including cell proliferation, transformation, differentiation, and apoptosis. In this study we describe the MAPK/ERK pathway via (quasi) biochemical reactions and then implement the pathway by a stochastic Markov process. A novelty of our approach is to use multiple parametrizations in order to deal with molecules for which localization in the cell is an intricate part of the dynamic process and to describe the protein using different binding sites and various phosphorylations. We simulate the system by exact and different approximate simulations, e.g. via the Poisson τ-leap, the Binomial τ-leap and the diffusion methods, in which we introduce a new updating plan for dependent columns of the diffusion matrix. Finally we compare the results of different algorithms by the current biological knowledge and find out new relations about this complex system.

1 Introduction

The MAPK/ERK pathway whose main components are Ras, Raf, and MEK proteins (Figure 1), is one of the major signalling systems which regulates the cellular growth control of all eukaryotes. The structure of this pathway includes a number of phosphorylations on the protein level whose interactions are directed by *positive* and *negative feedback loops*. These loops cause either activation or inhibition of other proteins and are named as *oscillatory behaviour* if they are executed as combinations with time delays. The circadian rhythm is an example of an oscillatory behaviour which implies a series of coordinated feedbacks between transcriptional activation and protein degradation that generates a periodic cycle (See Section 3.1).

Because of its importance in cellular activation, the MAPK/ERK pathway has been intensively studies from different laboratories [15, 12, 13, 1]. There are a number of biological sources which give qualitative knowledge about the MAPK/ERK pathway [16, 18]. However these sources do not describe the system by an explicit set of reactions which can be more helpful to understand the actual structure. In this study we combine these underlying qualitative knowledge to represent biochemical activations of the pathway as a list of (quasi) reactions. We denote all components by simple notations, thereby, produce a basis for stochastic simulation. As a novelty, moreover, we use multiple parametrizations in order to deal with molecules for

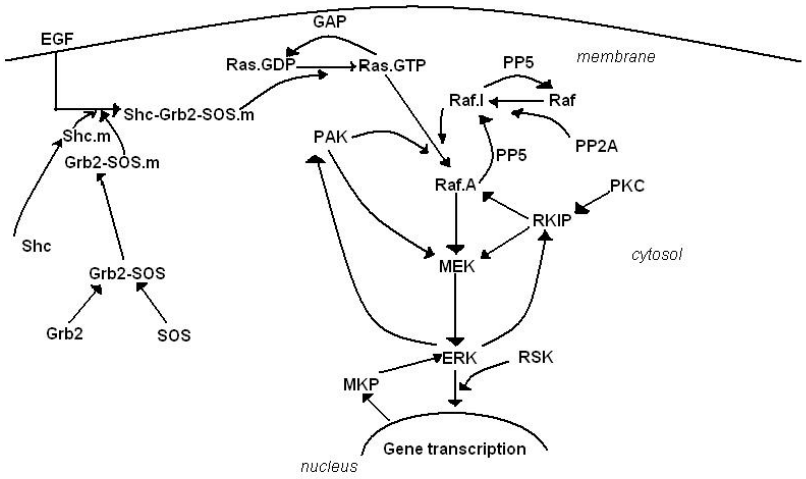


Figure 1: Simple representation of the structure of MAPK/ERK pathway.

which localization in the cell is an intricate part of the dynamic process and to describe the protein using different binding sites and various phosphorylations. The following set of equations, which describes the activation of the MAPK pathway by the EGF receptor, is an example from the reaction list of the pathway with 94 reactions and 51 substrates, representing 33 proteins and genes.

$$\text{EGF + Shc} \longrightarrow \text{EGF + Shc}_m \quad \text{(translocation)}$$

$$\text{Grb2 + SOS} \longrightarrow \text{Grb2-SOS}$$

$$\text{EGF + Grb2-SOS} \longrightarrow \text{EGF + Grb2-SOS}_m \quad \text{(translocation)}$$

$$\text{Shc}_m \text{ + Grb2-SOS}_m \longrightarrow \text{Shc-Grb2-SOS}_m$$

$$\text{Shc-Grb2-SOS}_m \text{ + Ras.GDP} \longrightarrow \text{Shc-Grb2-SOS}_m \text{ + Ras.GTP}$$

in which Grb2, SOS, Shc, EGF, Ras.GTP, and Ras.GDP are single proteins, Grb2-SOS and Shc-Grb2-SOS are protein complexes in the cytosol, and Shc-Grb2-SOS$_m$ is a protein complex near the membrane. As seen from the reactions, the translocation of substrates to the membrane is expressed by the notation m. For instance the protein Shc$_m$ denotes the Shc protein translocated from the cytosol towards the membrane. The different levels of the phosphorylation, on the other hand, are denoted by the index p or p1 and p2 where the first two abbreviations show the mono-phosphorylation and the latter implies the double-phosphorylation of a protein. For example, MEK.p2, another protein from the list, represents the double-phosphorylated MEK protein on the S218 and S222 binding sites.

Gene regulation is commonly modelled via ordinary differential equations (ODEs), which employ the law of mass action and continuous concentrations of each chemical substrate. Even though ODEs are successful to describe several reactions such as linear production and degradation, they cannot explain the small system variability of the actual reactions. For modelling biochemical systems, stochastic processes are a natural choice [4, 20]. This dynamic formalization takes into account the probabilistic manner of the different biological activations, such as the transcription of certain proteins, which occurs with low frequency in biological time [14].

2 Stochastic Simulation Algorithms (SSA)

There are various exact methods [6, 7, 5] for implementing stochastic simulations, but practically the *Direct method* [6], also known as the *Gillespie algorithm*, is the most common and usually most efficient simulator.

2.1 Gillespie algorithm

The Gillespie algorithm is strictly based on the *chemical master equation* which describes the stochastic behavior of the system [20] by

$$\frac{\partial}{\partial t}P(Y;t) = \sum_{j=1}^{M}\{a_j(Y - v_j, c_j)P(Y - v_j;t)dt - a_j(Y, c_j)P(Y;t)dt\}.$$

In this equation, M is the number of reactions $R_1, R_2, \ldots, R_M$. The N-dimensional vector $Y = (Y_1, Y_2, \ldots, Y_N)$ represents the state of the system at time t, v_j denotes the jth row of the net effect matrix V, and c_j is the stochastic rate constant of jth reaction, R_j. Accordingly, $a_j(Y, c_j)$ describes the hazard for reaction j with stochastic rate constant c_j and state Y so that the term $a_j(Y - v_j, c_j)P(Y - v_j;t)dt$ indicates the probability that the reaction R_j occurs over time interval $[t, t + dt]$ moving the state from $Y - v_j$ to Y.

This algorithm works well for simulating small systems, however, it is inefficient for developing realistic complex models since the time step for the next reaction is taken so small such that only a single reaction can occur in a given time step [5, 20]. Hence we use the following approximation techniques for simulating the signalling pathway.

2.2 The Poisson τ-leap method

An interesting idea for an approximation is to execute several reactions simultaneously over a larger time interval. If we define how many times each reaction is executed in each small time interval, we can move along the system's history axis from one time step to the next, instead of moving along from one reaction to the next. So by using these subintervals, *leaps*, a realization of state vector Y can be described. If we choose these intervals very small, the results correspond to those of the exact SSA. However if a larger time interval is selected, the results give us an approximation of the exact algorithm with a reduction of computational cost.

The Poisson τ-leap method is an approximation method whose aim is to increase intervals between sampling times efficiently under the *leap condition* [9]. The method determines a small time interval τ by

$$\tau = \min_{j\in[1,M]}\left\{\frac{\epsilon a_0(Y)}{\left|\sum_{i=1}^{N}\xi_i(Y)b_{ji}(Y)\right|}\right\}$$

in which $a_j(Y)$ is the jth reaction hazard, $b_{ji} = \partial a_j(Y)/\partial Y_i$ ($i = 1, \ldots, N$ and $j = 1, \ldots, M$), $a_0(Y) = \sum_{j=1}^{M} a_j(Y)$, $\xi(Y) = \sum_{j=1}^{M} a_j(Y)v_j$; ϵ is the specified fraction ($0 < \epsilon < 1$), in the sense that the transition probability per unit time (hazard) does not change very much, so that

the leap condition is satisfied. Under this assumption a sample value k_j, which is the number of times of the execution of the reaction R_j in the time interval [t, t+τ] given that $Y(t) = y$, is drawn from the *Poisson* distribution with mean $a_j(Y)\tau$. On the other hand the net change in the state of the system in $[t, t + \tau]$, λ, is calculated as $\lambda = \sum_j^m k_j v_j$. The method updates the current state by replacing t by $t := t + \tau$ and Y by $Y := Y + \lambda$ [9, 10].

2.3 The Binomial τ-leap method

Similar to the Poisson alternative, the Binomial τ-leap method also uses the system's history axis to produce an approximate simulation. By assuming that the leap condition holds, the maximum number $k_{max}^{(j)}$ of the jth reaction occurring in time interval $[t, t + \tau]$ given the state $Y(t)$ is determined by $\min_{v_{ij}<0}\lfloor Y_i/|v_{ij}|\rfloor$, where $\lfloor z\rfloor$ is the greatest integer in z, Y is the current state vector containing the number of molecules $Y_i(t)$ of all substrates at time t, and finally v_{ij} stands for each entry of the net effect matrix for the ith substrate and the jth reaction.

The binomial τ-leap method samples the number of the jth reaction k_j from the binomial distribution with the success probability $p = a_j\tau/k_{max}^{(j)}$ in which $\tau = f/\sum_{j=1}^{M} a_j$ and f is a coarse graining factor greater than 1. In the computation, $f = 1$ stands for the average time increment of the exact SSA [9, 19] and for the choice of *f*, like $f < 10^3$ for small steps or $f > 10^4$ for large steps, the method controls the computational time and the accuracy of the approximation. The method updates the time t and the state Y_i by setting $t_{\text{new}} = t_{\text{old}} + \tau$ and $Y_i(t_{\text{new}}) = Y_i(t_{\text{old}}) + \lambda$, where $\lambda = v_{ij}k_j$, respectively [2, 3].

The advantage of this method over the Poisson alternative is that it solves the problem of negative populations which may result in the unbounded Poisson τ-leap algorithm. This case happens, particularly, when k_j exceeds the available population size of one or more species in a single reaction [19, 2]. However it is unable to simulate several types of reaction, like transcriptional regulation, because the update regime of $k_{max}^{(j)}$ is based on only the substrates whose net effects in the jth reaction are less than 0 (See Section 3.1), whereas transcriptional regulation (DNA→DNA + RNA) has no negative net-effects.

2.4 Diffusion approximation

Under the assumption that the probability distribution of the number of the molecules of each species at t depends on the continuous t and continuous number of molecules, the stochastic model can be converted to a differential equations model. With the Fokker-Planck approach [17], this probability distribution $P(Y(t))$ can be expanded via a Taylor expansion [21, 17] and the change of state of each species at t is found by a Langevin approach, in which a correlated noise term describes the stochastic behaviour of the model over and above the drift term

$$dY(t) = \mu(Y, \Theta)dt + \beta^{\frac{1}{2}}(Y, \Theta)dW(t)$$

where $\mu(Y, \Theta) = V'a(Y, \Theta)$ and $\beta(Y, \Theta) = V'diag\{a(Y, \Theta)\}V$ are mean, or drift, and variance, or diffusion, matrices, respectively, both depending on Y and the parameter vector $\Theta = (c_1, c_2, \ldots, c_r)'$ explicitly; $dW(t)$ represents the change of a Brownian motion over time. Practically we implement the time increment by choosing the Gillespie's time step which is generated according to an exponential with parameter $\sum_{j=1}^{M} a_j$. The algorithm computes the next state at $t + dt$ by replacing $Y(t)$ by $Y(t) + dY(t)$ [8, 11].

2.5 The new updating regime for diffusion approximation

In large stochastic systems, we have observed that the distribution of the number of molecules of substrates exhibits high dependency when the system converges to the stationary distribution. This dependency structure causes a singularity in the diffusion matrix β and makes calculation of $dY(t)$ via a multivariate normal impossible. Effectively, the variance structure $\beta^{\frac{1}{2}}(Y,\Theta)dW(t)$ " lives " in a lower dimensional space. To unravel this problem we propose the following updating regime:

1. By checking the columns of β from left to right, each linearly dependent column β_i is identified.
2. The dependent columns S, totally $|S|$, are described as the linear combination of independent columns so that every $\beta_{.i} = \sum_{j\notin S, j<i} \alpha_{ij}\beta_{.j}$ The α vector shows the linear relationship between the underlying dependent column and previous independent columns and is stored for each i($i \in S$).
3. A new $(N-|S|)\times(N-|S|)$ dimensional diffusion matrix β^* is defined by eliminating $|S|$ dependent columns and rows from $\beta_{N\times N}$.
4. $N-|S|$ samples Δ_{-S} are generated from normal distribution with mean zero and variance $\beta^{*1/2}dt$ for updating $N-|S|$ linearly independent substrates via $Y_j(t+dt) = Y_j(t) + \mu_j(t)dt + \Delta_j$ where $j \notin S$.
5. The dependent columns of Y, Δ_S, are generated as $\Delta_i = \sum_{j\notin S, j<i}\alpha_{ij}\Delta_j$ ($i \in S$) using the αs from before and are updated as $Y_i(t+dt) = Y_i(t) + \mu_i(t)dt + \Delta_i$.

Under this new updating regime, it can be shown that the covariance structure β of the changes $dY(t)$ is preserved: For $i \in S$ and $j \notin S$

$$\begin{aligned}
\text{Cov}(\Delta_i, \Delta_j) &= \text{Cov}(\sum_{k\notin S, k<i}\alpha_{ik}\Delta_k, \Delta_j) \\
&= \sum_{k\notin S, k<i}\alpha_{ik}\text{Cov}(\Delta_k, \Delta_j) \\
&= \sum_{k\notin S, k<i}\alpha_{ik}\beta_{kj} \\
&= \beta_{ij}
\end{aligned}$$

for $i \in S$

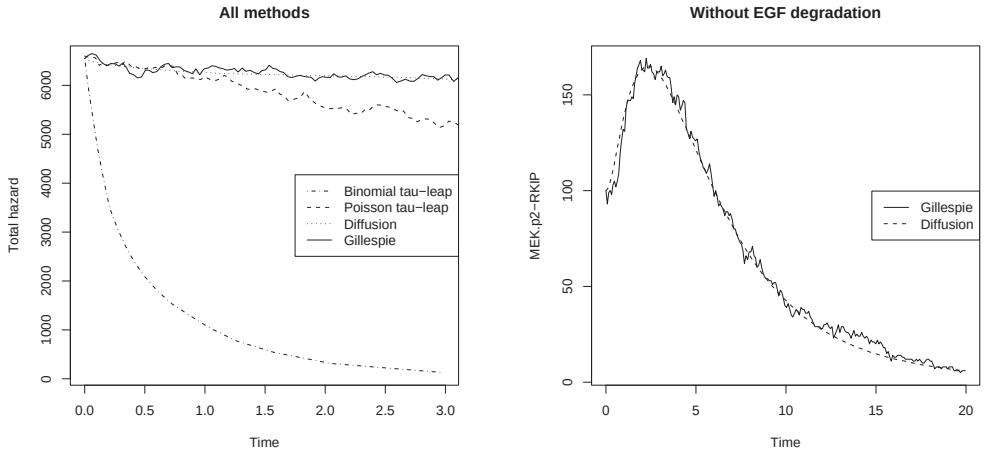
$$\begin{aligned}
\text{Cov}(\Delta_i, \Delta_i) &= \text{Cov}(\sum_{k\notin S, k<i}\alpha_{ik}\Delta_k, \Delta_i) \\
&= \sum_{k\notin S, k<i}\alpha_{ik}\text{Cov}(\Delta_k, \Delta_i) \\
&= \sum_{k\notin S, k<i}\alpha_{ik}\beta_{ki} \\
&= \beta_{ii}.
\end{aligned}$$

The results of the new algorithm are shown in Section 3.1.

3 Results

We have simulated the MAPK pathway by using exact and approximate methods mentioned above. In our computations, we have assumed that the hazards are constant for each level. We choose 3 gradations of reaction time speed, namely slow, normal, and fast. The stochastic rate constants have been calculated according to the order of each reaction, the given hazards, and the number of molecules which is initialized at 100 for all substrates. Then we have run the algorithms under three different scenarios: (i) excluding all reactions of degradation, (ii) merely including EGF degradation, and (iii) initializing the number of molecules of EGF at zero, respectively. Indeed in biochemical reactions protein degradation is much slower than the time periods during which biochemical activation and de-activation processes take place. Therefore ignoring these reactions in the MAPK pathway is realistic in simulation. Under the second scenario, we have added the *degradation* of EGF, which is a direct result of the activation of the MAPK pathway via the internalization of this receptor into vesicles. Finally under the third scenario without the presence of EGF we aim to unravel the steady state behaviour of the system, since EGF is the only protein that triggers the activation of the pathway and its inactivation has not yet been fully understood [18].

3.1 Comparison of simulation results



(a) (b)

Figure 2: (a) Total hazards of Gillespie algorithm, Poisson τ-leap method with $\delta = 0.6$, Binomial τ-leap method with $f = 10^2$, and diffusion approximation with new update plan under the first scenario and $t = 5$. (b) Gillespie and diffusion approximation with new updating regime of active MEK-RKIP complex in cytosol under the first scenario and $t = 20$. The y-axis indicates the number of molecules of this complex for the time t.

We have run simulations of the system under all simulation methods, and have compared the differences of their total hazards and changes in activities of substrates through time under the first scenario. Figure 2 shows that the Poisson τ-leap and the diffusion method with the new updating regime seem relatively smoother, whereas the Binomial τ-leap gives the worst approximation of the exact Gillespie algorithm. The comparison of number of iterations of all methods in Table 1 illustrates that the Binomial τ-leap has time increments that are too large.

Table 1: Comparison of computational times (2st column) and total number of iterations (3rd column) of the Gillespie algorithm, the Poisson τ-leap method, the Binomial τ-leap method, and the diffusion approximation with new update plan, respectively under the first scenario and $t = 5$.

	Total time	**Total iteration**
Gillespie algorithm	1 hr	30625
Poisson τ-leap method	1 hr 28 min	27851
Binomial τ-leap method	28 sec	3516
Diffusion approximation with new update plan	1 hr	30731

In this method the time step τ is decided with respect to the success probability $p = a_j\tau/k_{max}^{(j)}$ (Section 2.3). If $a_j\tau$ is bigger than $k_{max}^{(j)}$, τ is automatically made large, like in our example. Under this condition the only variable which can be adjusted for the computation of τ is the value of the coarse gaining factor f. Although for an appropriate choice of f, the result can be improved significantly, the decision about f is based on ad hoc calculations.

Furthermore if the system has many reactions of transcriptional regulation or the reaction of production, the method can not simulate the network properly because of $k_{max}^{(j)}$ updating step (Section 2.3). For instance the following reaction of transcription,

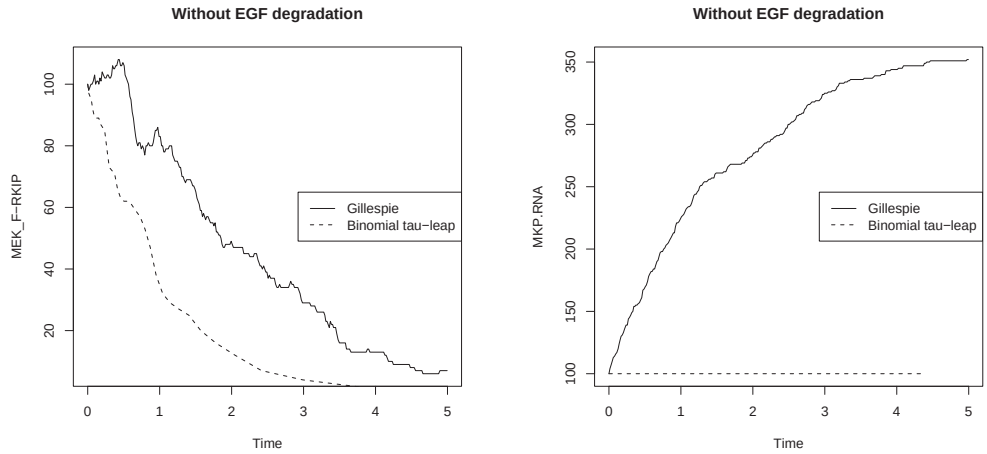
$$\text{ERK.p2-TF.p2} + \text{MKP.DNA} \longrightarrow \text{ERK.p2-TF.p2} + \text{MKP.DNA} + \text{MKP.RNA}$$

in which ERK.p2-TF.p2 is the active ERK in the nucleus, MKP.DNA and MKP.RNA are the DNA and the transcripted RNA sequences of MKP protein, respectively, cannot be executed during the algorithm since the net effect vector of the reaction $v_j = (0, \ldots, 0, 1, 0, \ldots, 0)$ has only positive entries (Figure 3b).

As seen from Table 1, on the other hand, the Poisson τ-leap method is computationally more expensive than other approximations. Indeed its result is highly dependent on the choice of ϵ which affects τ. Even though the method does not use any ad hoc calculations, it can give negative population numbers [19, 2, 3], particularly in the log-run simulation.

From the Table 1, it is also observed that the computational time of both Gillespie and diffusion algorithms are equal for $t = 5$. However since the number of iterations of the diffusion approximation is greater than that of Gillespie algorithm, it can be concluded that the speed of the former is slightly faster than the latter per iteration. Additionally from the simulation with longer time interval, it has been found that as the network becomes more complex, the computational efficiency of the diffusion approximation becomes considerably faster than the Gillespie algorithm. For instance the running time of the MAPK pathway for $t = 20$ is completed in 21 hours with Gillespie algorithm, on the other hand, the simulation takes 10 hours with diffusion approximation.

In this study because of the underlying problems of the Poisson and Binomial τ-leap methods, we have chosen the diffusion approximation to analyze the pathway as a smooth approximation of the exact simulation. Figure 2b shows the results of the active MEK-RKIP complex in cytosol produced by diffusion approximation with new updating regime. As seen from the graph, the diffusion with new updating step fits the exact result (Gillespie's results) well. The accuracy of the new procedure comes from the step that is very sensitive to the singularity of the diffusion matrix.



(a) (b)

Figure 3: (a) Inactive, mono-phosphorylated MEK and RKIP complex in cytosol and (b) RNA sequence of MKP in nucleus simulated by the Gillespie algorithm and Binomial τ-leap method with $f = 10^2$ under the first scenario and $t = 5$. The y-axes stand for the number of molecules of the underlying substrates for the time t.

3.2 Comparison of conditions and translocation of proteins

In order to compare the results under the three conditions, we have listed changes in activities of substrates with respect to their activations when EGF degradation is included. As shown in Table 2, we have found that single MEK and ERK proteins, all kinds of MEK-RKIP and ERK-RSK complexes as well as c-Fos, MKP proteins are unrelated to the availability of EGF protein in the system. Indeed a non-linear functionality between ERK and EGF has been already mentioned by earlier studies of Wiley et al. [22] and Hornberg et al. [13]. On the other hand inactive/active Ras, Raf, SOS, Shc, Grb2 proteins and their complexes either in cytosol or near the cell membrane are highly dependent on the EGF activation. Moreover as seen from Table 2, it has been observed that under the first (excluding all reactions of degradations) and the third (initializing the number of molecules of EGF to zero) scenarios, the proteins indicate exactly the same changes if the second (merely including EGF degradation) scenario is chosen as the basis condition of the pathway. This interesting finding can be explained by the oscillatory behavior of ERK which is still unclear but possibly due to the negative feedback through the induction of the expression of MAPK phosphatises [16].

Moreover co-regulation plots of biologically important pairs of substrates through time, have revealed a bi-activation relations within RSK vs c-Fos, ERK.p2 vs c-Fos, c-Fos.RNA vs ERK.p2-RSK.A, and EGF vs c-Fos proteins. The plot of the last pair is shown in Figure 4 as an illustration. These findings can imply the oscillatory behaviour of proteins like in the case of the ERK activation [16].

On the other hand for investigating the effects of the translocation of proteins under all conditions, we have plotted changes in activities of substrates with different localizations through time. We have found that the localization of molecules by multiple parametrization is necessary for explaining the dynamic behavior of the system. For instance Figure 5a and Figure 5b show different activation of Grb2-SOS protein in cytosol and near the cell membrane, respectively.

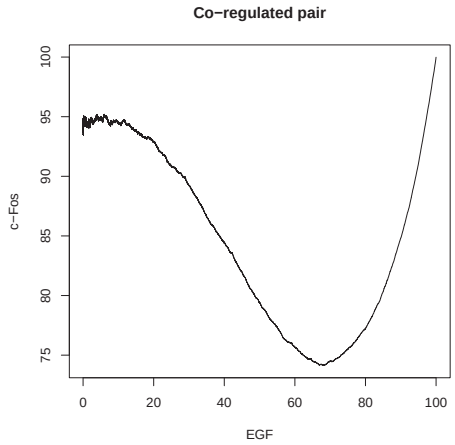
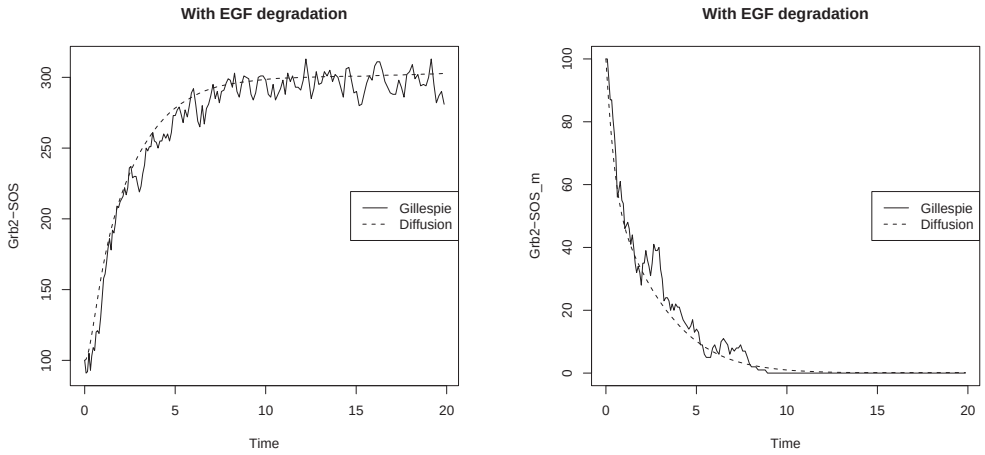


Figure 4: Plot of the co-regulation of EGF versus c-Fos proteins in the MAPK pathway under the second scenario and $t = 20$. The axes show the number of molecules of substrates through time.



(a) (b)

Figure 5: Change in activities of Grb2-SOS complex (a) in cytosol and (b) near membrane, respectively including the degradation of EGF by using both Gillespie algorithm and diffusion approximation under the second scenario and $t = 20$. The y-axes illustrate the number of molecules of the Grb2-SOS complexes for the time t.

Table 2: Changes in activities of each protein of the MAPK pathway under the first and the third scenario when the second scenario is accepted as the natural activation of the cell and $t = 20$.

	Types of proteins
Decrease in activation	Ras.GDP, Raf.I, Shc, Grb2, Grb2-SOS, MKP.RNA
Stability in activation	Raf.I-RKIP, Raf.I-RKIP$_m$, Raf.I-RKIP-Ras.GTP$_m$, MEK, MEK$_F$, MEK$_S$, MEK.p2, MEK-RKIP, RKIP, MEK$_F$-RKIP, MEK$_S$-RKIP, MEK.p2-RKIP, ERK, ERK.p1, ERK.p2, ERK.p2-TF.p2, ERK.p2-RSK.A, ERK.p2-RSK.A-TF.p2, c-Fos, c-Fos.DNA, c-Fos.RNA, c-Fos.p, MKP, MKP.DNA, TF, GAP, PP2A, PAK, PP5, RKIP.p, PKC, RSK
Increase in activation	Ras.GTP, Raf, Raf.I$_m$, Raf.I-Ras.GTP$_m$, Raf.A$_m$, Grb2$_m$, Raf.A-Ras.GTP$_m$, Shc$_m$, SOS, Grb2-SOS$_m$, Shc-Grb2$_m$, Shc-Grb2-SOS$_m$

3.3 Dynamic profiles

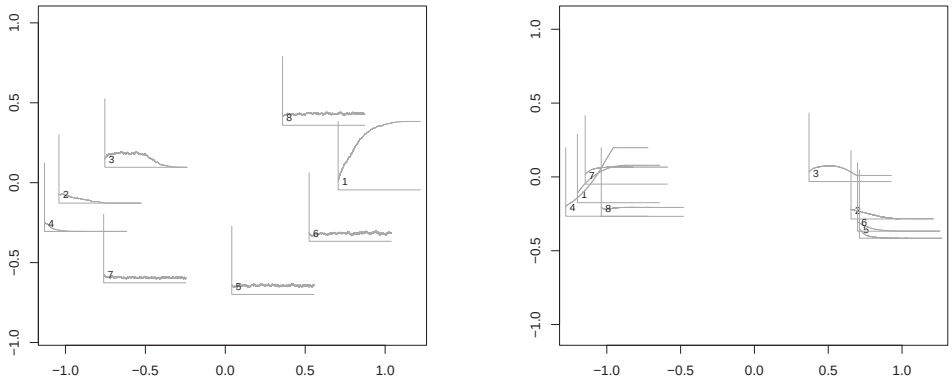
In order to find out highly variable substrates in the MAPK/ERK pathway, we have applied principal component analysis. The results from both diffusion and Gillespie algorithms show that inactive Raf (Raf.I), active MEK (MEK.p2), c-Fos phosphorylated by ERK (c-Fos.p), and RKIP phosphorylated by either PKC or ERK (RKIP.p) are associated with the most of the variability of the network. Indeed a similar conclusion was obtained in the study of Hornberg et al. (2005a): only a small group of reactions, which involve Raf, MEK, and ERK, directs the behaviour of the pathway.

For finding homogeneous subsets of substrates in the system, we have applied PAMSAM clustering [23] whose dissimilarity measure is defined via pairwise correlations. Figure 6 shows the resulting 8 clusters. The results with EGF degradation indicate that both simulation techniques partition most of the proteins in the same clusters and the substrates whose activations are similar like in the Table 2 are the members of the same or close clusters. Similarly the substrates whose activations are linear under the same conditions are gathered in the same or close classes. However the diffusion approximation has more correlated proteins than Gillespie algorithm. The reason is that the former is much smoother than the latter.

4 Conclusion

In this study, we have implemented the stochastic simulation of the MAPK pathway by using both exact and approximation algorithms. We have compared the simulation results according to their computational times and accuracies. The results show that although the Poisson τ-leap method can be accelerated by several improvements in time increments [10], it can give negative population sizes in the long-run. On the other hand the Binomial τ-leap method is fast in calculation, but it is not accurate enough to get smooth approximation of the exact algorithm and has ad hoc choices for updating states and determining the time increments.

In order to unravel the singularity problem of β, the diffusion matrix, in the diffusion approxi-



(a) (b)

Figure 6: PAMSAM clustering of 45 substrates (excluding c-Fos.DNA, GAP, PP2A, PAK, PP5, PKC) by using (a) Gillespie algorithm and (b) diffusion approximation under the second scenario and $t = 20$.

mation, we have proposed a new updating regime. The results show that our method sorts the singularity problem and gives an exact solution.

For simulating the MAPK pathway, as a novelty we have used multiple parameterizations to describe the translocation and the different levels of phosphorylation of the protein and have combined the different biological sources, which represent the pathway qualitatively, as a list of (quasi) reactions with simple notations. The analysis indicates that such kind of description, indeed, is essential to better understand the structure of the complex system and suitable for generating a stochastic simulation. Finally we have checked our conclusions with the biological knowledge about the pathway and found out several new features of the MAPK signalling system.

References

[1] L. Chang and M. Karin. Mammalian MAP kinase signalling cascades. *Nature*, 410(6824): 37-40, 2001.

[2] A. Chatterjee and D. G. Vlachos. Binomial distribution based τ-leap accelerated stochastic simulation. *Journal of Chemical Physics*, 122(2): 024112.1-024112.7, 2005.

[3] A. Chatterjee, K. Mayawala, J. S. Edwards, and D. G. Vlachos. Time accelerated Monte Carlo simulations of biological networks using the binomial τ-leap method. *Bioinformatics*, 21(9):2136-2137, 2005.

[4] N. Fedoroff and W. Fontana. Genetic networks: Small numbers of big molecules. *Science*, 297:1129-1131, 2002.

[5] M. A. Gibson and J. Bruck. Efficient exact stochastic simulation of chemical systems with many species and many channels. *Journal of Physical Chemistry*, A(104):1876-1889, 2000.

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

[7] D. T. Gillespie. A rigorous derivation of the chemical master equation. *Physica A*, 188:404-425, 1992.

[8] D. T. Gillespie. The chemical Langevin equation. *Journal of Physical Chemistry*, 113:297-306, 2000.

[9] D. T. Gillespie. Approximate accelerated stochastic simulation of chemically reacting systems. *Journal of Physical Chemistry*, 115:1716-1733, 2001.

[10] D. T. Gillespie and L. R. Petzold. Improved leap-size selection for accelerated stochastic simulation. *Journal of Physical Chemistry*, 119:8229-8234, 2003.

[11] A. Golightly and D. J. Wilkinson. Bayesian inference for stochastic kinetic models using a diffusion approximation. *Biometrics*, 61(3):781-788, 2005.

[12] J. J. Hornberg, B. Binder, F. J. Bruggeman, B. Schoeberl, R. Heinrich, H. V. Westerhoff. Control of MAPK signalling: Complexity to what really matters. *Oncogene*, 24(36):5533-5542, 2005.

[13] J. J. Hornberg, F. J. Bruggeman, H. V. Westerhoff, and J. Lankelma. Cancer: A systems biology disease. *BioSytems*, 83:81-90, 2006.

[14] D. A. Hume. Probability in transcriptional regulation and its implications for leukocyte differention and inducible gene expression. *Blood*, 96:2323-2328, 2000.

[15] W. Kolch. Meaningful relationships: The regulation of the Ras/Raf/MEK/ERK pathway by protein interactions. *Biochemical Journal*, 351:289-305, 2000.

[16] W. Kolch, M. Calder, and D. Gilbert. When kinases meet mathematics: The systems biology of MAPK signalling. *FEBS Letters*, 579:1891-1895, 2005.

[17] H. Risken. *The Fokker-Planck Equation: Methods of Solution and Applications*, Springer-Verlag, 1984.

[18] B. Schoeberl, C. Eichler-Jonsson, E. D. Gilles, and G. Müller. Computational modelling of the dynamics of the MAP kinase cascade activated by surface and internalized EGF receptors. *Nature Technology*, 20:370-375, 2002.

[19] T. Tian and K. Burrage. Binomial leap methods for simulating stochastic chemical kinetics. *Journal of Chemical Physics*, 121(21):10356-10364, 2004.

[20] T. E. Turner, S. Schnell, and K. Burrage. Stochastic approaches for modelling in vivo reactions. *Computational Biology and Chemistry*, 28:165-178, 2004.

[21] N. G. Van Kampen. *Stochastic Processes in Physics and Chemistry*, Amsterdam: North-Holland, 1981.

[22] H. S. Wiley, S. Y. Shvartsman, and D. A. Lauffenburger. Computational modelling of the EGF-receptor system: A paradigm for systems biology. *Trends in Cell Biology*, 13(1):43-50, 2002.

[23] E. C. Wit and M. D. McClure. *Statistics for Microarrays: Design, Analysis and Inference*, John Wiley and Sons, 2004.