UNIVERSIDAD AUTÓNOMA DE BAJA CALIFORNIA

LATIN AMERICAN JOURNAL OF APPLIED ENGINEERING

Computational modeling of the MAPK pathway using NetLogo

Dora L. Flores[a,*], Claudia M. Gómez[a], Edgar Manzanarez-Ozuna[a], Gustavo A. Hirata[b]

[a]*Universidad Autónoma de Baja California, Ensenada, México*
[b]*Centro de Ciencias y Nanotecnología, Universidad Nacional Autónoma de México*
**Corresponding author: dflores@uabc.edu.mx*

***Abstract*—**The signal transduction pathways have an important role in cell growth, as they are the primary means of communication of cells with their environment, which determines the activity of cells promoting responses such as proliferation, differentiation, migration, development, apoptosis, etc., there have been models representing these pathways using different techniques, such as ODEs, stochastic methods, Petri nets, pi calculation, Boolean networks, Bayesian networks, cellular automaton and agent-based systems, among others. In this research, an agent-based modeling and simulation of subsequent events after activation of the Ras molecule in the intracellular MAPK pathway section is presented, using a multi-agent programmable modeling environment called NetLogo.

***Keywords*—** Computational modeling, MAPK pathway, NetLogo, systems biology.

***Resumen*—**Las vías de transducción de señales tienen un papel importante en el desarrollo celular, ya que son la principal vía de comunicación de las células con su entorno, el cual determina la actividad de las mismas promoviendo respuestas celulares tales como: proliferación, diferenciación, migración, desarrollo, apoptosis, etc., se han realizado modelos que representan estas vías utilizando diferentes técnicas, tales como EDOs, métodos estocásticos, redes de Petri, cálculo pi, redes booleanas, redes bayesianas, autómatas celulares y sistemas basados en agentes, entre otros. En este trabajo de investigación se presenta un modelado basado en agentes y la simulación del conjunto de eventos posteriores a la activación de la molécula Ras, en la sección intracelular de la vía MAPK utilizando la herramienta llamada NetLogo.

***Palabras Claves*—**Modelado computacional, NetLogo, sistemas biológicos, vía MAPK.

I. INTRODUCTION

Signal-transduction pathways (STPs) have an important role in cell growth. They are the primary means of communication between cells and their environment. All of these pathways are highly organize and follow a broadly similar course that can be viewed as a molecular circuit [1]. These molecular circuits detect and amplify external signals that determine cellular responses such as proliferation, differentiation, migration, development, or apoptosis.

The MAPK/ERK pathway (also known as Ras-Raf-MEK-ERK pathway) is one of the most representative signaling routes, is very complex and includes many chemical components [2]. An environmental signaling molecule called growth factor that binds a membrane receptor on the cell surface and ends when the DNA in the nucleus expresses a protein promoting the cell division activates this pathway. As all the STPs, this pathway has a molecular on/off switch that, like those in a computer chip, transmit information when "on" [1]. When one of the proteins in the pathway is mutated, it become stuck in the "on" or "off" position and this can leads to an uncontrolled growth that can be associated with many forms of cancer [3].

Given the importance of the MAPK/EKR in living things, modeling this pathway, allow understand its composition as well as emerging behaviors through interactions between its components.

On the other hand, models based on agents (MBA), are formed by a set of components with rules local behavior, autonomy and self-directed interacting with each other, and

Received: 06.19.2015 Revised: 09.30.2015
Vol. 1 No. 1, 2015, pp. 11-17
Facultad de Ingeniería, Arquitectura y Diseño

live, interact and can learn from their environment and dynamically change their behavior in response to their experiences, creating from these interactions, emergent complex behaviors.

In [4-5] present a review of methodologies used for the modeling and simulation of biochemical networks illustrated for STPs. These methodologies are ODEs, stochastic methods, Petri nets, pi calculation, Boolean networks, Bayesian networks, cellular automata and agent-based systems, among others, range from qualitative to quantitative, and include structural as well as dynamic analysis techniques. All of these techniques provides benefits and also lack appropriate attributes for particular cases of simulation of biological systems.

Modeling and simulation based on agents (MSBA) have gained increasing attention in the last ten years. This trend is reflected in the increasing number of models and applications articles published in journals of scientific and technological research, the number of funded programs requiring agent-based models that incorporate elements of human and social behavior, the growing number of conferences or who have devoted issues to agent-based modeling, on-demand courses and training programs on MSBA, and the number of conference presentations on agent-based modeling [6-9].

In this paper, we propose a multi-agent system model to simulate part of the cascade of mitogen-activated protein kinase (MAPK/ERK) activated by a growth factor. In this model, we use the epidermal growth factor (EGF), that stimulates cell growth, proliferation, and differentiation by binding to its receptor. In this model, individual agents representing each of the molecules comprising the MAPK/ERK cascade, whose interaction is given by attraction (described by collision), and recognition molecules and their functions involving the development of biochemical reactions specific to them.

II. METHODOLOGY

In this section biocomputational design model based on multi-agent systems is presented. The research method is defined and the model of the MAPK/ERK pathway in which this research is based is introduced. Development environment and simulation, in which implement the proposed model is described and finally the simulation paradigm is described.

A. Research method

In [10] simulation as research model is proposed, as shown in Fig. 1. Which can be used for this project making some adjustments. Firstly determining the system under study, the abstraction of the system will be performed on a model will be applied computational through a computer program, it will be conducted simulation runs *in virtuo* generating simulated data, and finally the data obtained were compared *in virtuo* with data from *in vitro* looking for similarities to validate the model and its results.

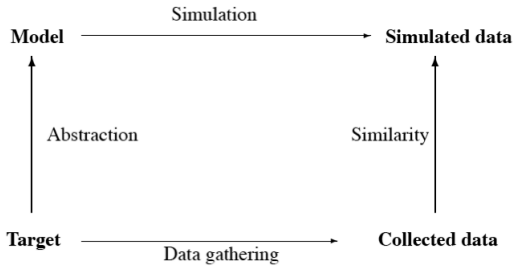


Fig. 1. The logic of simulation as a method.

Another proposed methodology to simulate systems biology is in [11]. A cycle of research begins with the selection of contradictory issues of biological significance and the creation of a model representing the phenomenon. Models can be created either automatically or manually. The model represents a computable set of assumptions and hypotheses that need to be tested or supported experimentally.

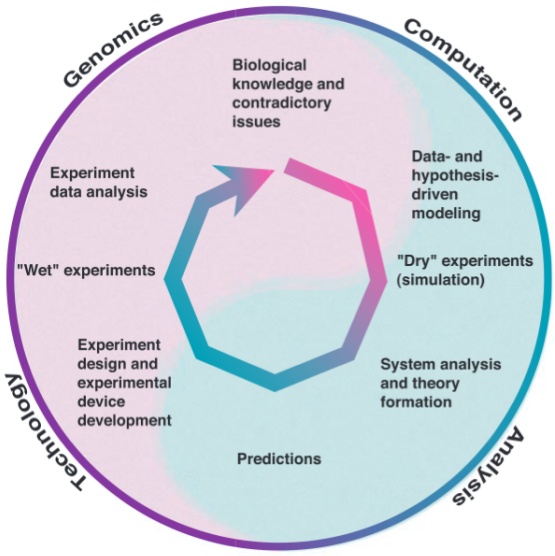


Fig. 2. The cycle for a more systematic science.

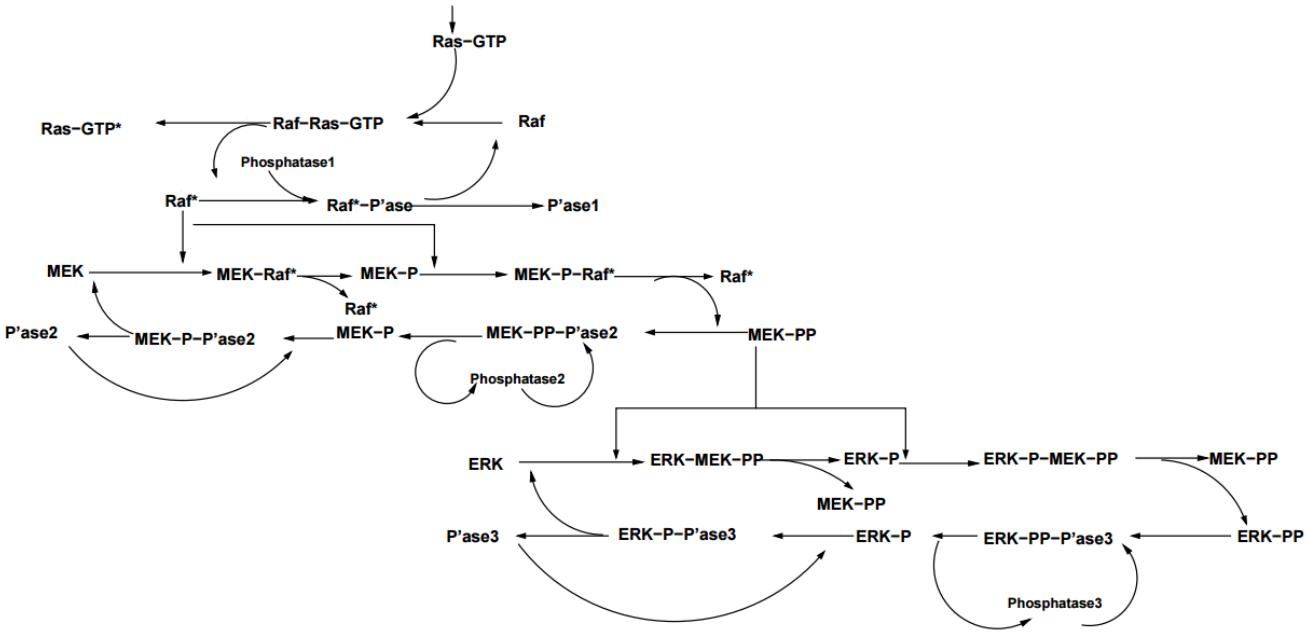
Successful experiments are those that eliminate inadequate models. Models that survive this cycle are deemed to be consistent with existing experimental evidence. While this is an idealized process of systems biology research, the hope is that advancement of research in computational science, analytical methods, technologies for measurements, and genomics will gradually transform biological research to fit this cycle for a more systematic science (Fig. 2).

B. MAPK pathway model

The mathematical model of the MAPK/ERK pathway is taken from [12]. In which it is possible to calculate the change in concentration of ninety-four components of this pathway in relation to time after stimulation of epidermal growth receptor (EGF).

In our model, simulated events are given from which the Ras protein is activated (Ras-GTP), after which it activates the protein kinase activity of Raf kinase. The next reactions are all associated with the phosphorylation and dephosphorylation of the protein components of the pathway mediated by kinases and phosphatases. Our model allows observing changes in the concentration of all the fourteen molecular species as shown in Fig. 3.

Fig. 3. Subsequent events after activation of the Ras molecule in the intracellular MAPK pathway section.



C. Modeling environment

NetLogo is a programmable modeling environment for simulating natural and social phenomena. It has been in continuous development ever since at the Center for Connected Learning and Computer-based Modeling. Is particularly well suited for modeling complex systems developing over time. Modelers can give instructions to hundreds or thousands of "agents" all operating independently. It is possible to explore the connection between the micro-level behavior of individuals and the macro-level patterns that emerge from their interaction [13].

D. In virtuo experimentation

The strength of this paradigm of simulation is based on the ability to perturb a model while it is running, dynamically modify the boundary conditions, and eliminate or add elements during the simulation [14]. *In virtuo* experimentation supports NetLogo.

III. MODEL IMPLEMENTATION

In this section all elements that comprise the proposed model are described. Such as, environment, agents with features and capabilities, the general reaction method of the agents and biochemical reactions in the MAPK cascade.

In the proposed model all the molecular species interact into the cytoplasm. Surrounded by the plasma membrane. Each of the nine molecular species (Ras, Raf, MEK, ERK, GTP, Phosphatase1 (P'ase), Phosphatase2 (P'ase2) and Phosphtase3 (P'ase3)) reacts as a part of the MAPK cascade. Each of which is represented by an agent breed, and has features (activate and phosphorylate) movement and reconnaissance capabilities that define the reactivity with other agents. The environment is divided into rectangular areas of arbitrary constant size, called patches.

These features and capabilities are characterized as rules of interaction, provide as follows: The agents have random motion in a fluid medium (Brownian motion); each agent contains a list of agents that can interact. The interaction is based on the enzyme specificity and a list of reactions that take place one the pathway is activated. Given the above, the general reaction method of the agents in this model is defined as:

1. Two agents collide if both are located in the same patch.
2. If there is enzyme specificity between agents, the reaction occurs, leading to changes in the concentrations of the agents.
3. If there is no enzyme specificity between agents, no reaction is triggered and is considered non-reactive collision.
4. Each of the agents continues to the Brownian motion.

In the multiple events in the signaling cascade between the molecular species into the environment, it is possible to observe biochemical reactions through second order, processes of association by aggregating phosphate groups (activate and phosphorylate) and dissociation through removing phosphate groups (deactivate and dephosphorylate). Each of the reactants or compounds activated is marked by an "*", and a "P" marks the phosphorylated compounds. These reactions are shown below:

The GTP molecule binds to the Ras molecule to activate and generate the Ras-GTP complex when the reaction occurs; the GTP molecule is integrated into the compound.

$$[Ras]+[GTP] \leftrightarrow [Ras\text{-}GTP^*] \quad (1)$$

The active Ras-GTP complex activates the Raf molecule when the reaction occurs; the Ras-GTP complex dissociates and loses reactivity.

$$[Raf]+[Ras\text{-}GTP^*] \leftrightarrow [Raf^*]+[Ras\text{-}GTP] \quad (2)$$

Activated Raf phosphorylates MEK molecule up in two places creating two new molecular species: MEK-P (phosphorylated MEK) and MEK-PP (doubly phosphorylated MEK). When reactions occur, Raf remains activated.

$$[MEK]+[Raf^*] \leftrightarrow [MEK\text{-}P]+[Raf^*] \quad (3)$$

$$[MEK\text{-}P]+[Raf^*] \leftrightarrow [MEK\text{-}PP]+[Raf^*] \quad (4)$$

The molecule MEK-PP phosphorylates ERK protein in two places to generate two new molecular species: ERK-P (phosphorylated ERK) and ERK-PP (doubly phosphorylated ERK). When reactions occur, MEK-PP maintains its initial state.

$$[ERK]+[MEK\text{-}PP] \leftrightarrow [ERK\text{-}P]+[MEK\text{-}PP] \quad (5)$$

$$[ERK\text{-}P]+[MEK\text{-}PP] \leftrightarrow [ERK\text{-}PP]+[MEK\text{-}PP] \quad (6)$$

Phosphatase 1 (P'ase) deactivates the Raf molecule, maintaining its initial state.

$$[Raf^*]+[P1] \leftrightarrow [Raf]+[P1] \quad (7)$$

Phosphatase 2 (P'ase2) removes up to two phosphate groups of the MEK-PP, generating molecular species MEK-P (phosphorylated MEK) and MEK. When reactions occur, P'ase2 maintains its initial state.

$$[MEK\text{-}PP]+[P2] \leftrightarrow [MEK\text{-}P]+[P2] \quad (8)$$

$$[MEK\text{-}P]+[P2] \leftrightarrow [MEK]+[P2] \quad (9)$$

Phosphatase 3 (P'ase3) removes up to two phosphate groups of the ERK-PP, generating molecular species ERK-P (phosphorylated ERK) and ERK. When reactions occur, P'ase3 maintains its initial state.

$$[ERK\text{-}PP]+[P3] \leftrightarrow [ERK\text{-}P]+[P3] \quad (10)$$

$$[ERK\text{-}P]+[P3] \leftrightarrow [ERK]+[P3] \quad (11)$$

The initial concentration for each of the molecular species and the kinetic parameters of biochemical reactions present in this model is taken from the supplementary information of [12] available on [15].

The reactions occur sequentially from number one to number eleven; in an infinite loop, which stops condition, is when the concentration of molecules GTP and MEK-PP is equal to zero.

IV. RESULTS AND DISCUSSION

The simulation was developed on NetLogo version 5.2. It has been tested with models that use hundreds of megabytes of RAM and works well; it has not been tested with models that use gigabytes of RAM.

NetLogo engine has no fixed limits on size but it has one-gigabyte ceiling on how much total RAM it can use [11]. Even when it is possible to extend the memory limit, this has not been tested by the developers of NetLogo and does not ensure a stable environment simulation.

The initial concentrations of each molecular species are shown in Table 1 given in molecules per cell. In our model the concentrations of the Ras and GTP molecules

TABLE I

INITIAL CONCENTRATION BY MOLECULAR SPECIES

Molecular species	Schoeberl (molecules/cell)	Model (molecules/cell)
GTP	6.63×10^4	6.63×10^1
Ras	1.14×10^7	1.14×10^4
Raf	4.0×10^4	4.0×10^1
MEK	2.20×10^7	2.20×10^4
ERK	2.10×10^7	2.10×10^4
P'ase	4.0×10^4	4.0×10^1
P'ase 2	4.0×10^4	4.0×10^1
P'ase 3	1.0×10^7	1.0×10^4

correspond to Ras-GDP and SOS (Son of sevenless, set of genes encoding guanine nucleotide exchange factors that act on the Ras subfamily of small GTPases) molecules respectively in the model Schoeberl.

Concentrations of molecular species in this simulation are different but proportional to those used in the model Schoeberl. In order of reduce the time and time computing resources we scale the concentrations by a 10^{-3} factor. This scale is the results of setup this model with different factors of original concentration, bigger than 0.010% arise an out of memory error.

Implementation of the model allows scaling all concentrations in the same factor, as show in Fig. 4, so that ratios are maintained. Fig. 5 depicts a setup screenshot in NetLogo. Where eight different molecules are show, Ras (yellow circles), Raf (red circles), MEK (orange circles), ERK (turquoise circles), GTP (violet triangles), P'ase (blue arrows), P'ase2 (dark green arrows), and P'ase3 (light green arrows).

In order to identify the behavior of the MAPK/ERK cascade several simulations of 20 minutes were performed at different concentrations. Since the speed of the reactions is given by the concentration gradients of the molecules, is expected in different concentrations, the system behaves differently.

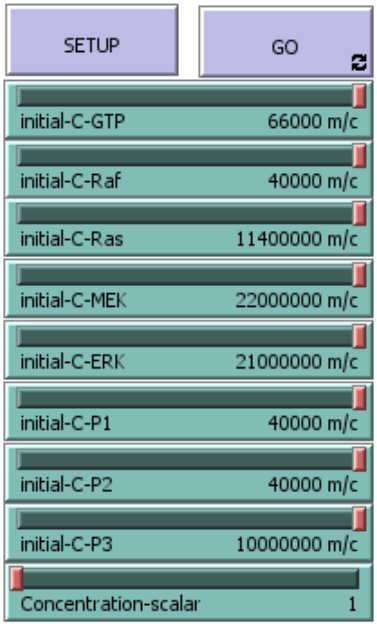


Fig. 4. Slider controls for concentrations of molecules in NetLogo simulation.

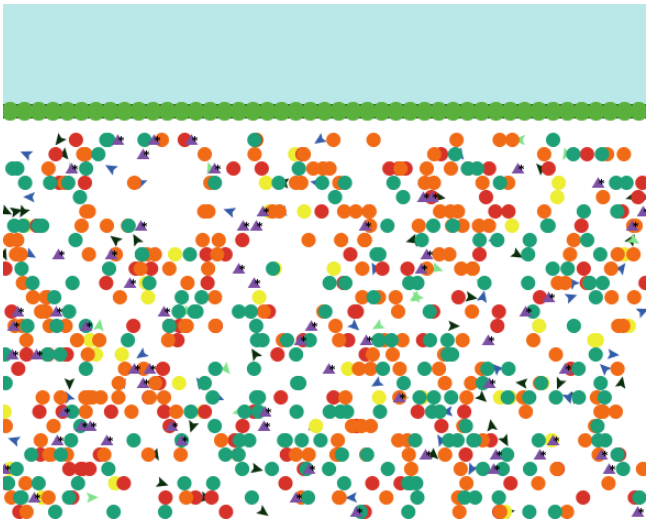


Fig. 5. A screenshot of simulation using NetLogo.

In Fig. 6, the concentrations of the ERK-PP molecules, recorded from simulation with different concentrations of molecular species are shown, in it is appreciated that with different initial values, the concentration of ERK-PP behaves differently.

There is no a pattern of behavior in order to establish relations of linearity between experiments. Conversely we can identify that while experiments 25% and 50% tend towards a linear behavior, experiments 75% and 100%, do not have this tendency. This allows us to suggest that the system is not linear and is dependent on the initial values.

In each of the experiments, the concentration of the molecule ERK-PP had a period of "latency", this ends when no longer zero and begins growth. At concentrations, 25%, 50%, 75%, and 100%, this stage ended at tick (second) 26, 74, 78, and 217, respectively. This allows us to suggest that even if concentrations have a linear distribution, so latency does not follow this pattern.

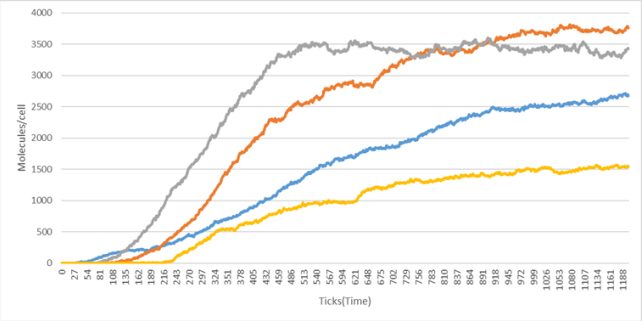


Fig. 6. Concentration of molecules ERK-PP on different initial conditions: 25% (blue line), 50% (orange line), 75% (gray line), and 100% (yellow line).

Dephosphorylation of regulatory proteins is a major mechanism for modulating signaling pathways involving protein kinases. Dephosphorylation is performed by phosphatases proteins, which are classified based on their substrate specificity.

We reduce the initial concentrations in the simulations in order to identify the impact of decrease of the deactivated Raf rate and the dephosphorylation of the molecular species MEK-P, MEK-PP, ERK-P and ERK-PP. In all of the simulations the phosphatases proportions change in 50%. It is given on the assumption that given the reduced phosphatases (inhibitors), the concentration of ERK-PP molecule tends to increase and remain doubly phosphorylated.

As shown in Fig. 7, the concentration of ERK-PP in different conditions of concentration and proportions of species phosphatases tends to be incremental, confirming the dynamics of the reactions and the negative impact of inhibition given by low gradients phosphatases.

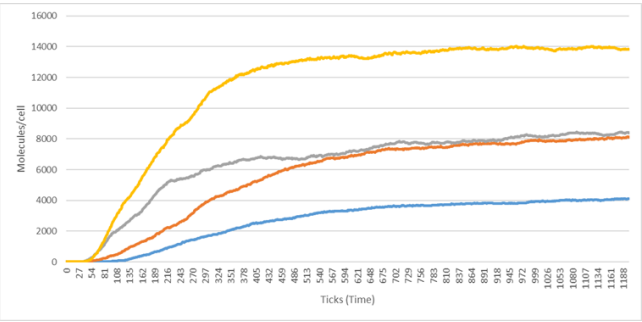


Fig. 7. Concentration of molecules ERK-PP with non-proportional phosphatases molecules 50% on different initial condition 25% (blue line), 50% (orange line), 75% (gray line), and 100% (yellow line).

The mutated Ras molecule is always active, together with guanosine triphosphate (GTP) and is called RAS-GTP. Together with Raf molecule, comprising the substrates required for the production of activated Raf molecule. As show in Equation 1. Once the reaction occurs, the mutated Ras does not dissociate and retains its reactivity. Since the activated Ras is the beginning of the cascade in our model. This will be initiated on a recurring basis and increase the concentration of the molecular species ERK-PP. Simulations were run with different initial conditions and mutated Ras molecules (Fig. 8).

In Fig. 6 and Fig. 8, the simulation results in normal and mutated Raf are shown. Clearly there is impact when Ras is mutated. This leads to a permanent increase in ERK-PP.

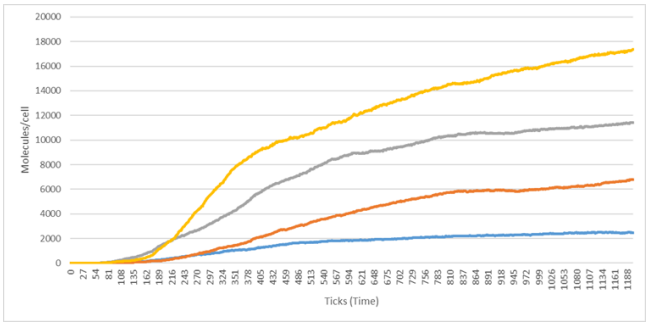


Fig. 8. Concentration of molecules ERK-PP with mutated Ras molecules on different initial conditions, 25% (blue line), 50% (orange line), 75% (gray line), and 100% (yellow line).

Fig. 9 shows the average results of the different initial conditions and characteristics of some components. The impact of the mutated Raf evidence against normal conditions on the concentration of ERK-PP is greater and incremental as time increases. The reduction of phosphatases, impacts positively on the increased concentration of ERK-PP. Both scenarios are in compliance with the assumptions in the simulation.

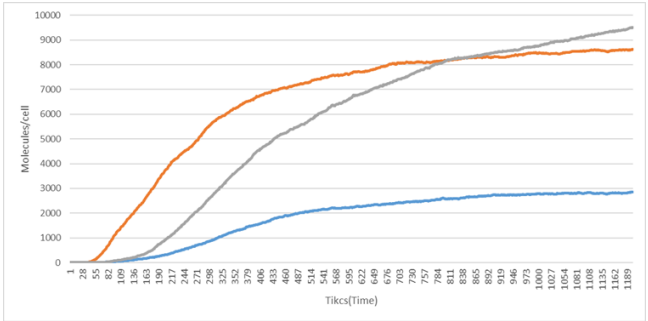


Fig. 9. Average concentration of molecules ERK-PP on different initial conditions (25%, 50%, 75%, and 100%), normal (blue line), non proportional phosphatases molecules (orange line) and mutated Ras molecules (gray line).

V. CONCLUSIONS AND FUTURE WORK

A model and simulation of MAPK pathway is presented using an agent-based model to represent mitogen activated protein kinase cascade using NetLogo.

This model represents a form non-linear behavior arising different levels of concentration of molecules with different initial conditions.

The increment in the concentration of ERK molecule grows since dephosphorylation processes are reduced is well addressed by this model.

Mutated Ras molecule means Raf molecule never loses it reactivity; therefore increment of concentration of MEK-P, MEK-PP, ERK-P, ERK-PP molecules is arising, this increase is represented appropriately by this model.

The MAPK cascade is activated by growth factors. Adding these growth factors so when they get to the membrane from the extracellular space and bound with receptors, the biochemical reactions of the MAPK cascade will occur in consequence inside the cell can extend it.

Also, ERK-PP molecule can phosphorylate and regulate many proteins that induce cellular responses, including nuclear transcription factors going into the nucleus to induce transcription of certain genes that participate in cell growth, mitosis, survival or differentiation. Integrate a fourth compartment to simulate the nucleus and add a new breed of agents representing the transcription factors going inside of the nucleus after being phosphorylated by ERK-PP.

We can introduce a new breed of agents GAP and the phosphorylation reaction of the Ras-GDP.

REFERENCES

[1] J. M. Berg, J. L. Tymoczko, L. Stryer, "Biochemistry". New York: W.H. Freeman, 2002.

[2] R. J. Orton, O. E. Sturm, V. Vyshemirsky, M. Calder, D. R. Gilbert, and W. Kolch, "Computational modelling of the receptortyrosine-kinase-activated MAPK pathway." The Biochemical journal, vol. 392, no. Pt 2, pp. 249–61, Dec. 2005.

[3] B. Vogelstein and K. W. Kinzler, "Cancer genes and the pathways they control." Nature medicine, vol. 10, no. 8, pp. 789–99, Aug. 2004.

[4] W. Materi and D. S. Wishart, "Computational systems biology in drug discovery and development: methods and applications." Drug discovery today, vol. 12, no. 7-8, pp. 295–303, Apr. 2007.

[5] N. ElKalaawy and A. Wassal, "Methodologies for the modeling and simulation of biochemical networks, illustrated for signal transduction pathways: A primer," Biosystems, vol. 129, pp. 1–18, 2015.

[6] C. M. Macal and M. J. North, "Agent-based modeling and simulation," Proceedings of the 2009 Winter Simulation Conference (WSC), pp. 86–98, Dec. 2009.

[7] F. Moreno-Arboleda; J. A. Guzman-Luna; S. A. Gómez-Arias. "Analysis and Detection in V-Formations with Outliers". Ing. Univ.; 18(1); 43-57; 2014.

[8] J. A. Betancourt-Bethencourt. "Agent-based model as a didactic tool for the subject Public Health". Educ Med Super; 28(3); 436-445; 2014.

[9] F. J. Moreno-Arboleda; J. F. Duitama-Muñoz; E. Camilo-Ospina. "A method for estimating the position and direction of a leader of a set of moving objects". Rev.fac.ing.univ. Antioquia; (62); 11-20; 2012.

[10] N. Gilbert and K. G. Troitzsch, Simulation for the Social Scientist, 1999.

[11] H. Kitano, "Systems Biology: ABrief Overview," Sciences 295, 1662-1664, 2002.

[12] B. Schoeberl, C. Eichler-Jonsson, E. D. Gilles, and G. Müller, "Computational modeling of the dynamics of the MAP kinase cascade activated by surface and internalized EGF receptors." Nature biotechnology, vol. 20, no. 4, pp. 370–375, 2002.

[13] Wilensky, U. NetLogo. http://ccl.northwestern.edu/netlogo/. Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL. 1999.

[14] J. Tisseau, "In vivo, in vitro, in silico, in virtuo. The virtuoscope," pp. 1–17, 2008.

[15] "Supplementary information: Computational modeling of the dynamics of the MAP kinase cascade activated by surface and internalized EGF receptors: Nature Biotechnology."