Constraint–driven optimization of plant defense model parameters

Dragana Miljkovic [1,2], Matjaž Depolli [1], Igor Mozetič [1], Nada Lavrač [1,2]
[1]Jožef Stefan Institute
[2]Jožef Stefan International Postgraduate School
Ljubljana, Slovenia
e-mail: {dragana.miljkovic, matjaz.depolli, igor.mozetic, nada.lavrac}@ijs.si

Tjaša Stare, Marko Petek, Kristina Gruden
National Institute of Biology
Ljubljana, Slovenia
e-mail: {marko.petek, tjasa.stare, kristina.gruden}@nib.si

***Abstract*—*Biologists have been investigating the plant defense response to virus infections for a long time. Nevertheless, its model has still not been developed. One of the reasons is the deficiency in numerical kinetic data that brings up the importance of the expert knowledge. Therefore, we based our work on acquiring domain knowledge of biological pathways which provided a basis for the construction of a dynamic mathematical model. The goal of our work was to model the major pathway of the plant defense response – the salicylic acid pathway – and determine its dynamic parameters that are in correspondence with the knowledge acquired from the biology experts. For this purpose, we first selected the Hybrid Functional Petri Net formalism to represent the model due to its intuitive graph representation important for the biologists and its mathematical capabilities necessary for the simulation. The salicylic acid model was manually constructed and curated. In addition, the knowledge related to the model variables was acquired from the biology scientists and formalized in the form of constraints. This enabled an automatic optimization search for the model parameters that violate the minimal number of constraints. If the simulation results do not match the expert expectations, the network structure and the constraint definition are revised and the optimization parameter search is repeated. The final results of our system are both simulation results and optimized model parameters, which provide an insight into the biological system. Our constraint-driven optimization approach allows for an efficient exploration of the dynamic behavior of the biological models and, at the same time, increases their reliability.*

Keywords-constraint-driven optimization; systems biology; dynamic parameters; plant defense response

I. INTRODUCTION

Plants and pathogens enter into various relations that do not necessarily damage the host plant. This interaction triggers a complex signaling network, referred to as plant defense response or plant defense signaling. For a successful defense the activation of plant defense must be rapid, efficient and targeted [1]. It was shown that salicylic acid (SA), jasmonic acid (JA) and ethylene (ET) pathways play a fundamental role in mediating the defense signaling response in plants [2].

The goal of systems biology is to build a holistic view of dynamical interactions between various biological pathways. In practice, these pathways are mostly qualitatively understood while the numerical data of the kinetic parameters are often sparse. Due to the small amount of existing quantitative data, mathematical optimization methods were recently employed in systems biology. Various local deterministic optimization techniques (Levenberg-Marquardt algorithm [3][4], Sequential Quadratic Programming [5]) and stochastic approaches (Simulated Annealing [6], Genetic Algorithms [7] and Evolutionary Algorithms [8]) are applied in systems biology.

Plant defense response, like all biological mechanisms, has several dynamic parameters that are not accessible to experimental measurements, such as speeds of reactions and inhibition thresholds. One way to estimate these parameters is to fit the model to the experimental data (if they are available) [9]. On the other hand, the absence of kinetic data for model fitting raises the importance of qualitative knowledge of domain experts. This knowledge of biological pathways can serve as a basis for the construction of a dynamic mathematical model. Most of the plant-pathogen interaction studies are focused on individual interactions or subsets of the whole plant defense mechanism [10][11]. The first attempt to model the plant defense by constructing a Boolean network and carrying out numerical simulations of plant defense model was proposed by Genoud et al. [12]. However, this model is simple, containing 18 biological entities and 12 Boolean operators, whereas to fully investigate complex biological system one needs to consider a large number of components [13].

The goal of our work is to develop the iterative process to determine the dynamic parameters of the plant defense response, in correspondence with the knowledge acquired from the biology experts. In this study we concentrate on the model plant species *Arabidopsis thaliana* and its interaction with viruses. At the level of signal perception we selected the Turnip Crinkle Virus (TCV) infection. We concentrate our study to the one of the three major pathways that is the most studied: the SA pathway. The main contributions of this paper are:

- Methodology for acquiring knowledge from the domain experts resulting in a new dynamic SA model
- Formalized biological knowledge in the form of constraints
- The dynamic model of the SA pathway.

The structure of this paper is as follows. Materials and methods section describes the methodology used to search for the model parameters through the iteration process. Every step of this methodology is presented in a separate subsection. The Results section presents the results obtained by applying the described methodology and the discussion of the results from two iteration steps. The conclusion section summarises the main advances of the present work and discusses future aspects.

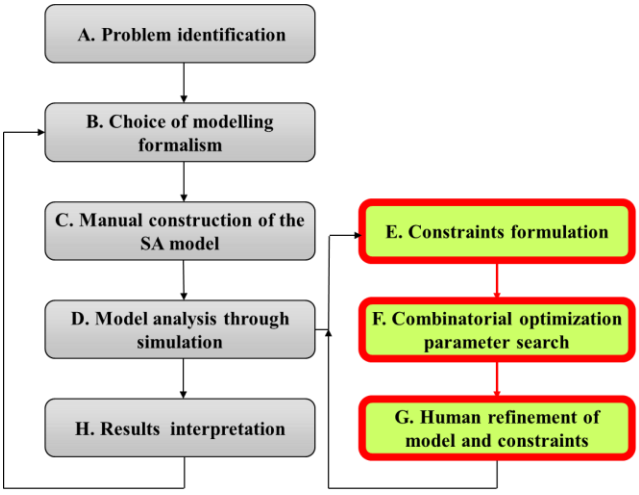

Figure 1. A schema of the utilized methodology for the constraint-driven optimization of plant defence model parameters.

II. MATERIALS AND METHODS

An overview diagram of the iterative model construction process is shown in Fig. 1. First, the SA model was manually constructed using the Hybrid Functional Petri Net formalism (HFPN) [14] and curated by biologists. Since the manual estimation of the parameters was unattainable, an automatic method was developed based on a differential evolution algorithm [15]. Here we illustrate the whole process loop of converging to the dynamic parameters of the SA pathway, that best satisfy the expert evaluation. This process includes eliciting knowledge from the biologists, formalizing it in the form of constraints, optimization of parameters that violate the minimal number of these constraints and the revision of the model structure and constraints. Eventually, the system yields both simulation results and optimized model parameters, which provide an insight into the biological system. The details of every step in this construction process are explained in the following subsections.

A. Problem identification

The final goal of developing the plant defense model is to verify whether the plant will have resistant reaction to survive the virus attack if some genes in the model are silenced. This would practically save time for the biology scientists in the design and performance of real-life experiments with plants that on average last two years. To be able to confirm or reject the hypothesis, the plant defense model has to be first developed together with the genes that are interesting candidates for silencing.

The difficulty of the plant defense modeling task is reflected in its complexity and dynamics. The plant defense is complex due to the three highly-interconnected major pathways: salicylic acid (SA), jasmonic acid (JA) and ethylene (ET), whereas the exact relationships between the biological molecules are still unclear. The dynamics is also difficult to address due to the lack of quantitative kinetic data that would reveal the unknown kinetic parameters, such as speeds of reactions and inhibition thresholds.

B. Choice of modeling formalism

A Petri net (PN) is a graphical and mathematical formalism [16], chosen as the formalism for the plant defense model representation. There are different types of PNs. Standard PNs are discrete and qualitative. But with their various extensions, PNs allow the definition of both qualitative and quantitative models. One of the PN extensions is the HFPN [14].

HFPN represents the combination of hybrid PN [17] and functional PN [18]. HFPN supports the modeling of continuous reactions controlled by switch mechanisms and the dependence of the reaction speeds on input concentrations, which is often needed to model real systems. Having in mind the overall goal of plant defense simulation model, we have selected the HFPN formalism using the Cell Illustrator software, which initially had the name Genomic Object Net [19]. This software allows to model ordinary differential equations (ODE) [20] hidden to the end user through a user-friendly HFPN graphical interface. The tool allows for an easy building of the network topology based on the experts' knowledge without available experimental data. Cell Illustrator has a graphical editor that has drawing capabilities and allows biologists to model different biological networks and simulate the dynamic interactions between the biological components. On the other hand, Cell Illustrator does not have capability for automatic optimization of dynamic parameters. For this reason, we have used additional combinatorial optimization parameter search (see subsection F.).

C. Manual construction of the SA model

To construct the plant defense response model topology, we have defined the types of biological components (molecules) and relations (reactions), shown in Fig. 2.

Biological components are grouped into four classes: small compounds or metabolites (Chorismate, etc.), proteins (Chorismate synthase, etc.), genes (EDS5 gene, etc.) and protein complexes (NPR1 oligomer, etc.). Note also that components with similar functions are grouped into a single node that represents an entire family of these components.

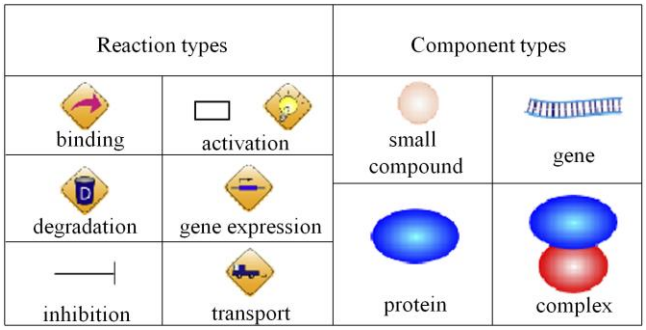

Figure 2. Types of biological molecules and reactions modelled with the Cell Illustrator.

Six types of reactions were identified in the last version of the SA model: binding, degradation, inhibition, activation, transport and gene expression. We introduce *binding* - because in some cases it is an essential means of regulation - defined as a close interaction between at least two reactants resulting in a functional active complex. *Degradation* is a diminishing of one component by processing it to smaller non-functional pieces; we introduce it to our model as a process that decreases the component's concentration. *Inhibition* is defined as a process when one component abolishes the performance of another component and prevents it from functioning. As *activation* we consider all the chemical reactions in which reactants A and B synergistically form a product C and the concentration of the product depends on the concentration of both substrates. *Transport* represents the moving of a biological component from one part of the cell to another. It is introduced in few cases when the biology experts considered it important for the plant defense response mechanism. Finally, g*ene expression* is defined as the constant or regulated activation of a gene which produces a functional protein.

D. Model analysis through simulation

Analysis of the dynamic model behavior was performed through iterative simulations of the manually constructed model. Based on the experts' evaluation the iterations were repeated until the simulation curves match their expectations. The simulation was performed initially in the Cell Illustrator software. The later simulations when the parameters were automatically estimated were executed in the simulator based on the C++ code exported from Cell Illustrator. The simulator outputs the time series curves of the dynamic behavior of the molecules of interest (see Fig. 4 and Fig. 5).

E. Constraints formulation

Formalizing the expert's knowledge into mathematical formulas is an iterative process. After evaluation together with biology experts, it turned out that there is a lot of tacit domain knowledge that can be expressed explicitly. This knowledge is accumulated in biology literature and had to be recognized as valuable in the parameter optimization search. We have explicitly focused on the knowledge related to the biological molecules and the relationships between them. The following five types of relationships, formed as constraints between the entities, were defined:

- Inequality relationship between molecules
- Growth rate of the molecules (for example, quantity of molecule 1 grows faster than that of molecule 2)
- Curve shape – e.g., it starts from zero, reaches a maximum and then drops back to zero level
- Minimal amplitude and minimal growth of the curve
- Temporal sequence in curve maxima
    - Same time (molecule 1 has the peak the same time as molecule 2)
    - Maximum before (molecule 1 has the peak before molecule 2)

F. Combinatorial optimization parameter search

Evolutionary algorithms are stochastic optimization methods utilizing the mechanisms of biological evolution in computer problem solving. One of the popular algorithms within this class is the differential evolution algorithm. The differential evolution (DE) algorithm performs a population-based search that optimizes the problem by iteratively trying to improve a candidate solution with regard to a given measure of quality. The parallel version of DE algorithm developed by Filipič and Depolli [21] was used in our work.

The optimal parameter setting of the plant defense model is defined as a combinatorial optimization problem. Criteria function for the optimization is the normalized sum of the normalized violations of constraints that are acquired from the biologists. If the criteria function has value 0 it means that all constraints are satisfied, while value 1 results from all non-satisfied constraints. All values in the range between 0 and 1 denote that a certain percentage of the time-series curves that are involved in the definition of the specific constraint do not satisfy it.

G. Human refinement of model and constraints

If the model simulation with the automatically determined parameters does not match the expectations of the biologists, the model structure and the constraints are refined.

H. Results interpretation

The simulation results are interpreted by the biology experts. The results allow for qualitative conclusions regarding the dynamic behavior of the model. This means that, in practice, it is possible to compare different curves in order to conclude which are the major components influencing the plant defense response.

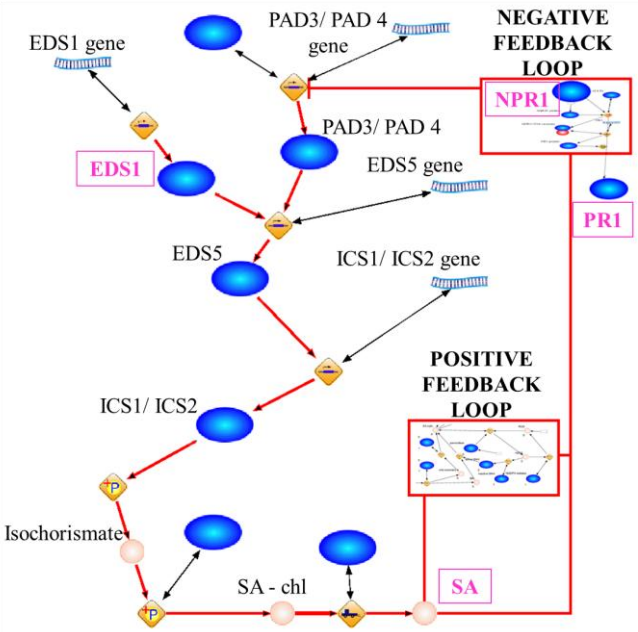

Figure 3. The simplified part of the model of a biosynthesis and signaling pathway of SA manually constructed in the Cell Illustrator.

III. RESULTS AND DISCUSSION

Here we illustrate the process of converging towards the biology experts expectations by presenting the results of two steps of the method described in the previous section.

A. Step 1 - SA model v1.0

The manually constructed SA sub-model contains 52 biological molecules and 38 reactions which includes inhibitions. This was an initial model built manually in the Cell Illustrator. The simplified version of an SA model is presented in Fig. 3. The simulator outputs 4 curves (as time series with 1000 points) for each biological molecule, which were the most interesting for the biology scientists. For this model in total 8 constraints were acquired from the biologists at the beginning of the model construction process. After the DE algorithm search was performed with a population number set to 10,000, the optimal parameters with respect to criteria function were estimated. With this set of parameters, for each violated constraint there is number showing the number of violated time points. The criteria function was calculated as a normalized sum of all of these numbers.

Below are the detailed values for individual constraints and the overall value of the criteria function:

lowerThan(e.SA_chl, e.SA_cyto) = 0.088
slowerRate(e.Chorismate, e.Prephenate) = 0.012
slowerRate(e.Chorismate, e.Phenyl_pyruvate) = 0.052
slowerRate(e.Chorismate, e.Phenylalanine) = 0.055
zeroPeakZero(e.SA_cyto) = 0.096
zeroPeakZero(e.PR1) = 0.037
equalRate(e.Prephenate, e.Phenyl_pyruvate) = 0.028
equalRate(e.Phenylalanine, e.Phenyl_pyruvate) = 0.007
finalCriteria = 0.376/ 8= 0.047

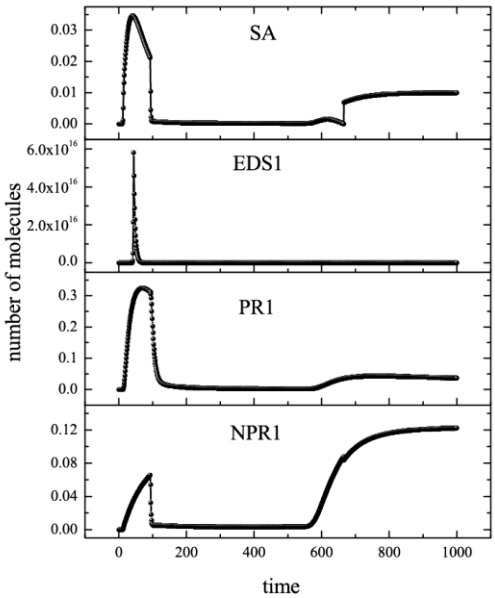

Figure 4. The dynamic behavior of the SA, EDS1, PR1 and NPR1 variables based on the optimal set estimated with respect to the criteria function calculated from the eight constraints during the step 1.

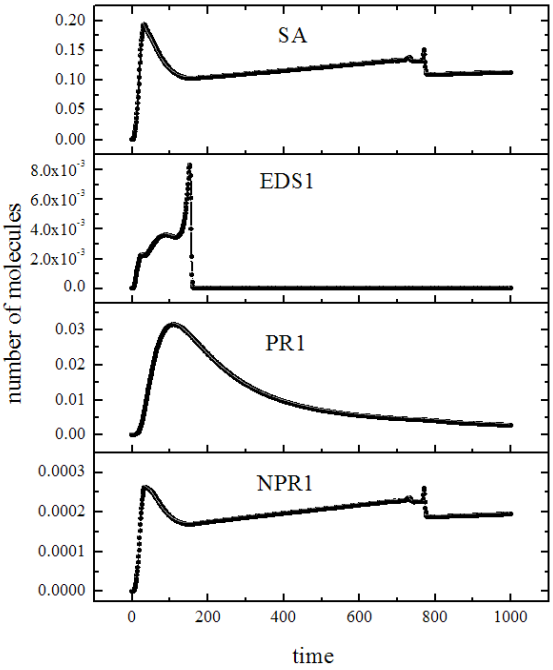

Figure 5. The dynamic behavior of the SA, EDS1, PR1 and NPR1 variables based on the optimal set estimated with respect to the criteria function calculated from the eight constraints during the step 2.

Based on this parameter set, the simulator outputs the curves of 4 biological molecules: SA, NPR1, PR1 and EDS1. Apart from the SA which is a small compound, the other three molecules are proteins. Their dynamic behavior is shown in Fig. 4. However, according to the biology experts some parts of these curves are not considered correct even though the total criteria function showed that on average 0.047 (4.7%) of each constraint is violated.

B. Step 2 - SA model v2.0

After inspection of the curves from Fig. 4, the biology experts have revised the model structure. This revision resulted in a second model version: SA model v2.0 containing 61 biological molecules and 56 reactions. Also, more constraints were specified leading to a set of 33 constraints.

The parameter search was once more performed with the same set up as in the step 1. Below are the shortened detailed values for individual constraints and the overall value of the criteria function:

equalRate(e.Prephenate,e.Phenyl_pyruvate) = 0.004
slowerRate(e.EDS1,e.EDS5) = 0.028
maxSameTime(e.ROS,e.HRT) = 0.931
maxAfter(e.HRT,e.MPK3)= 0.000
lowerThan(e.SA_chl,e.SA) = 0.001
zeroPeakZero(e.Chorismate) = 0.333
stopFast(e.virus) = 0.000
…
finalCriteria = 4.521/ 33 = 0.137

Here the total criteria function showed that on average 0.137 (13.7%) of each constraint is not satisfied. The dynamic curves of the same 4 molecules (SA, EDS1, PR1 and NPR1) are shown in Fig. 5. Even though the criteria function shows more violated constraints compared to the step 1, the biology experts were more satisfied with the presented curves in Fig. 5.

One should have in mind that the number of constraints increased in the step 2 and, therefore, the criteria function values from the step 1 and step 2 are not directly comparable. The second important part in the evaluation of the results from the parameter optimization is the manual evaluation of the biology scientists. If the final 4 curves do not match the biologists' expectations, even if the criteria function was equal to 0 in an ideal case, this is not considered as a good result and biologists would revise the constraints and model structure. The final result is a compromise between the criteria function value and the biologists expectations of the curve shapes.

The selected parameter set is large, thus, making the search space enormous. This automatically directed us to use some of the stochastic optimization methods since the deterministic methods are in the case of the large parameter sets overly time-consuming. Interesting results are obtained using our evaluation method, albeit some limitations exist. Our method is based on the knowledge of the domain experts which is still subjective, and not on the explicit and objective numerical experimental data. Nevertheless, this knowledge is still valuable and very useful for guiding the model construction. The comparison of the simulation curves with the experimental datasets remains an open challenge of our approach. The common size of the experimental datasets in plant biology is from 2 to 11 time points. The experimental curves and the simulation curves do not have the same number of time points and it is difficult to compare them. An additional problem that arises is the determination of the common start and ending in these curves.

IV. Conclusion

This paper represents the loop of converging to the dynamic parameters of the SA pathway that best satisfy the experts' evaluation. In the illustrative example of two iterations we show how the dynamic behavior of the simulated curves improves according to the experts' evaluation. In future work we plan to revise through more iterations the constraints and the SA model structure until the criteria function obtains the smallest possible value and at the same time gives the curves that meet the criteria of biologists. This methodology can be extended to any other biological pathway or mechanism.

Acknowledgment

This work was supported by the Slovenian Research Agency grants P4 0165, J4-2228, J4-4165 and P2-0103 and AD Futura scholarship.

References

[1] J. W. Moore, G. J. Loake, S. H. Spoel, "Transcription Dynamics in Plant Immunity," Plant Cell, vol. 23, Aug. 2011, pp. 2809-2820.

[2] P. Reymond and E. E. Farmer, "Jasmonate and salicylate as global signals for defense gene expression," Current Opinion in Plant Biology, vol. 1, Oct. 1998, pp. 404-411.

[3] D. Marquardt, "An Algorithm for Least-Squares Estimation of Nonlinear Parameters," SIAM Journal on Applied Mathematics, vol. 11, June 1963, pp. 431–441.

[4] K. Levenberg, "A Method for the Solution of Certain Non-Linear Problems in Least Squares," Quarterly of Applied Mathematics, vol. 2, 1944, pp. 164–168.

[5] P.T. Boggs and J.W. Tolle, "Sequential Quadratic Programming," Acta Numerica, vol. 4, 1995, pp. 1-51.

[6] S. Kirkpatrick, C. D. Gelatt, M. P. Vecchi, "Optimization by Simulated Annealing," Science, vol. 220, May 1983, pp. 671–680.

[7] M. Mitchell, "An Introduction to Genetic Algorithms", MIT Press, Cambridge, MA, 1996.

[8] A. E. Eiben and J. E. Smith, "Introduction to Evolutionary Computing", Springer-Verlag, Berlin, 2003.

[9] E. Walter and L. Pronzato, "Identification of parametric models from experimental data," MASSON, Paris, 1994.

[10] G. Olmedo, H. Guo, B.D. Gregory, S.D. Nourizadeh, L. Aguilar-Henonin, L. Hongjiang, et al. "ETHYLENE-INSENSITIVE5 encodes a 5 '-> 3 ' exoribonuclease required for regulation of the EIN3-targeting F-box proteins EBF1/2, "Proceedings of the National Academy of Sciences, vol. 103, Aug. 2006, pp. 13286-13293.

[11] P.E. Staswick, "JAZing up jasmonate signaling," Trends in Plant Science, vol. 13, Feb. 2008, pp. 66-71.

[12] T. Genoud, M.B.T. Santa Cruz, J.P. Métraux, "Numeric simulation of plant signaling networks," Plant Physiology , vol. 126, Aug. 2001, pp.1430-1437.

[13] H.A. Kestler, C. Wawra, B. Kracher, M. Kühl, "Network modeling of signal transduction: establishing the global view," Bioessays, vol. 30, Nov. 2008, pp. 1110-1125.

[14] H. Matsuno, Y. Tanaka, H. Aoshima, A. Doi, M. Matsui, S. Miyano, "Biopathways Representation and Simulation on Hybrid Functional Petri Net," In Silico Biology, vol. 3, June 2003, pp. 389-404.

[15] R. Storn and K. Price, "Differential Evolution – A Simple and Efficient Heuristic for Global Optimization over Continuous Spaces," Journal of Global Optimization, vol. 11, Dec. 1997, pp. 341 – 359.

[16] W. Reisig, "Petri nets and algebraic specifications," Theoretical Computer Science, vol. 80 , March 1991, pp. 1 – 34.

[17] M. Silva and L. Recalde. "On fluidification of Petri net models: from discrete to hybrid and continuous models," Annual Reviews in Control, vol. 28, 2004, pp. 253-266.

[18] R. Hofestädt and S. Thelen, "Quantitative modelling of biochemical networks," In Silico Biology, vol. 1, 1998, pp. 39-53.

[19] M. Nagasaki, A. Doi, H. Matsuno, S. Miyano, "Genomic Object Net:I. A platform for modeling and simulating biopathways," Applied Bioinformatics, vol. 2, 2004, pp.181-184.

[20] E. A. Coddington and N. Levinson, "Theory of Ordinary Differential Equations," McGraw-Hill, New York, 1955.

[21] B. Filipič and M. Depolli, Parallel Evolutionary Computation Framework for Single- and Multiobjective Optimization, Parallel Computing, 2009, pp. 217-240.