Plant defence model revisions through iterative minimisation of constraint violations

Dragana Miljkovic*

Department of Knowledge Technologies,
Jožef Stefan Institute,
Jamova 39, Ljubljana 1000, Slovenia
Email: dragana.miljkovic@ijs.si
*Corresponding author

Matjaž Depolli

Department of Communication Systems,
Jožef Stefan Institute,
Jamova 39, Ljubljana 1000, Slovenia
Email: matjaz.depolli@ijs.si

Tjaša Stare

Department of Systems Biology and Biotechnology,
National Institute of Biology,
Večna pot 111, Ljubljana 1000, Slovenia
Email: tjasa.stare@nib.si

Igor Mozetič

Department of Knowledge Technologies,
Jožef Stefan Institute,
Jamova 39, Ljubljana 1000, Slovenia
Email: igor.mozetic@ijs.si

Marko Petek and Kristina Gruden

Department of Systems Biology and Biotechnology,
National Institute of Biology,
Večna pot 111, Ljubljana 1000, Slovenia
Email: marko.petek@nib.si
Email: kristina.gruden@nib.si

Nada Lavrač

Department of Knowledge Technologies,
Jožef Stefan Institute,
Jamova 39, Ljubljana 1000, Slovenia
Email: nada.lavrac@ijs.si

Abstract: Biologists have been investigating plant defence response to virus infections; however, a comprehensive mathematical model of this complex process has not been developed. One obstacle in developing a dynamic model, useful for simulation, is the lack of kinetic data from which the model parameters could be determined. We address this problem by proposing a methodology for iterative improvement of the model parameters until the simulation results come close to the expectation of biology experts. These expectations are formalised in the form of constraints to be satisfied by the model simulations. In three iterative steps the model converged to satisfy the biology experts. There are two results of our approach: individual simulations and optimised model parameters, which provide a deeper insight into the biological system. Our constraint-driven optimisation approach allows for an efficient exploration of the dynamic behaviour of biological models and, at the same time, increases their reliability.

Keywords: systems biology; dynamic parameters; plant defence response; constraint-driven optimisation.

Reference to this paper should be made as follows: Miljkovic, D., Depolli, M., Stare, T., Mozetič, I., Petek, M., Gruden, K. and Lavrač, N. (2014) 'Plant defence model revisions through iterative minimisation of constraint violations', *Int. J. Computational Biology and Drug Design*, Vol. 7, No. 1, pp.61–79.

Biographical notes: Dragana Miljkovic is enrolled in the PhD programme "Information and Communication Technologies" at Jožef Stefan International Postgraduate School, Ljubljana, Slovenia. Her research interests include natural language processing applied to systems biology as well as dynamic modelling of biological responses in plants to stress. She obtained her master degree in the field of Information and Communication Technologies in Healthcare, University of Belgrade, Serbia, 2007. She was also involved in several projects related to data acquisition of electrophysiological signals and modelling dynamic responses of living organisms to their varying environment.

Matjaž Depolli received a PhD in Computer and Information Science from Jožef Stefan International Postgraduate School Ljubljana in 2010. He currently holds the position of research assistant at the Department of Communication Systems at the Jože Stefan Institute in Ljubljana. His research interests include evolutionary computation, computer simulation of physical phenomena, parallel computing, artificial intelligence, and biomedical research. He is also involved in the development of wireless body sensors, software for ECG analysis, cluster management software, and cloud services.

Tjaša Stare is enrolled in a PhD programme of Plant Biotechnology in an interdisciplinary programme of Bioscience, University of Ljubljana, Slovenia. She finished her master degree in Biochemistry and is currently employed as a young researcher in the National institute of Biology, in the department of Biotechnology and Systems Biology. Her primary research interest lies in understanding plant defence response after pathogen attack.

Igor Mozetič is a senior associate at the Jozef Stefan Institute, Department of Knowledge Technologies. He got a PhD in Computer Science in 1988 and has extensive international research and teaching experience (Univ. of Illinois at Urbana-Champaign, George Mason University, USA, TU Vienna and OFAI, Austria, and Center for Applied Mathematics and Theoretical Physics). His

current research interests are in knowledge discovery, semantic web, and data and text mining. He was a member of the W3C standardisation working group on Rule Interchange Format, and has participated in several EU projects related to knowledge discovery in biomedical and financial domains.

Marko Petek is a Research Scientist at National Institute of Biology (NIB), Ljubljana, Slovenia. He has a PhD in Biotechnology from University of Ljubljana and a BS in Pharmacy from University of Ljubljana. His current research interests and expertise are in plant-pathogen and plant-insect interactions and high-throughput gene expression methods and data analysis. He was also involved in development of data-mining tools for microarray data interpretation.

Kristina Gruden is Science Councillor and Head of DE III at National Institute of Biology (NIB), Ljubljana, Slovenia. The main topic of her research is focused on understanding of how the cell functions on the molecular level with extrapolation of this knowledge to understanding of organism responses to environmental stress. In parallel she is also involved in development of methodologies that support above mentioned research. Included are the methods of quantitative molecular biology and also methods that integrate the life sciences with statistic, computer science and mathematics.

Nada Lavrač is Head of Department of Knowledge Technologies at Jozef Stefan Institute, Ljubljana, Slovenia. Her main research interests are in machine learning, relational data mining, knowledge management, and applications of data mining in medicine and bioinformatics. She was scientific coordinator of EU projects ILPNET and SolEuNet. She is author and editor of numerous books and conference proceedings, including *Foundations of Rule Learning* (Springer 2012).

This paper is a revised and expanded version of a paper entitled 'Constraint-driven optimization of the plant defense model parameters' presented at the '3rd Workshop on Integrative Data Analysis in Systems Biology (IDASB)', Philadelphia, USA, 4–7 October 2012.

1 Introduction

Pathogens are a serious threat to living organisms and can lead to fitness costs, physiological damage or even death (Zhao et al., 2011). In particular, plants have specially evolved sophisticated mechanisms that can effectively fight-off infections with various pathogens. Upon pathogen recognition, plants trigger a complex signalling network, referred as plant defence response. For a successful defence the activation of plant defence must be rapid, efficient and targeted (Moore et al., 2011). It was shown that salicylic acid (SA), jasmonic acid (JA) and ethylene (ET) pathways play a fundamental role in mediating the defence signalling response in plants (Reymond and Farmer, 1998).

The goal of systems biology is to build a holistic view of dynamic interactions between various biological pathways. The modelling and simulation of biological pathways has attracted a considerable attention lately (Cho and Wolkenhauer, 2003; Janes and Lauffenburger, 2006; Klipp and Liebermeister, 2006). In practice, many biological 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 optimisation methods were recently employed in systems biology. Various local deterministic optimisation techniques, like Levenberg-Marquardt algorithm (Marquardt, 1963; Levenberg, 1944), Sequential Quadratic Programming (Boggs and Tolle, 1995) and stochastic approaches [Simulated Annealing (Kirkpatrick et al., 1983), Genetic Algorithms (Mitchell, 1996) and Evolutionary Algorithms (Eiben and Smith, 2003)], are applied in systems biology.

Plant defence response, like all biological mechanisms, has several dynamic parameters that are not accessible to experimental measurements, such as reaction rates and inhibition thresholds. One way to estimate these parameters is to fit the model to the experimental data (if they are available) (Walter and Pronzato, 1994). 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 the 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 defence mechanism (Olmedo et al., 2006; Staswick, 2008). The first attempt to model the plant defence by constructing a Boolean network and carrying out numerical simulations of plant defence model was proposed by Genoud et al. (2001). However, this model is simple, containing 18 biological entities and 12 Boolean operators, whereas large-scale experiments have shown that many more components are involved in defence signalling (Kestler et al., 2008).

The goal of our work is to develop the iterative process to determine the dynamic parameters of the plant defence model, in correspondence with the knowledge of biology experts. The structure of the plant defence model developed by Miljkovic et al. (2012) served as a basis for the dynamic model. The dynamic model is useful in predicting plant reactions to the conditions that have not yet been tested experimentally. The expected use of the model is to predict dynamic behaviour of selected variables by model simulation and to use the simulation results to suggest new experiments. Simulations provide faster results than the wet lab experiments that involve a time-consuming process of growing mutant plants with altered gene expressions in a specific pathway. Hence, simulation results can provide support in identifying the key components for gene mutations in the plant defence and assist in further experimental research. Such an approach to experiment design is more beneficial than a simple intuitive or experimental trial and error approach.

In this study we concentrate on the model plant species *Arabidopsis thaliana* and its interaction with viruses. At the level of signal perception we select the Turnip Crinkle Virus (TCV) infection. We focus our work to the most studied of the three major pathways: the SA pathway. The main contributions of this paper are:

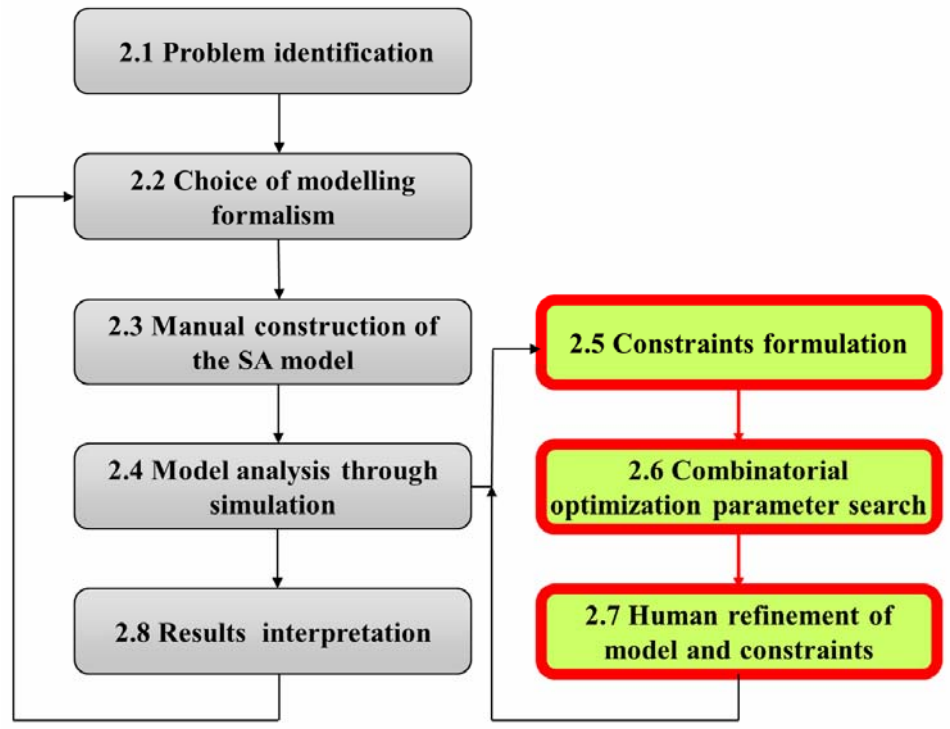
- Methodology for acquiring knowledge from the domain experts resulting in a new dynamic SA model
- Formalised biological knowledge in the form of constraints
- The dynamic model of the SA pathway.

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

2 Materials and methods

An overview diagram of the iterative model construction process is shown in Figure 1. First, the SA model was manually constructed using the Hybrid Functional Petri Net formalism (HFPN) (Matsuno et al., 2003) and curated by biologists. Since the manual estimation of the parameters was unattainable, an automatic method was developed based on a differential evolution algorithm (Storn and Price, 1997). 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 collecting knowledge from biologists, formalising it in the form of constraints, optimisation 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 optimised 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.

Figure 1 A schema of the utilised methodology for the constraint-driven optimisation of plant defence model parameters (see online version for colours)



2.1 Problem identification

The final goal of developing the plant defence 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

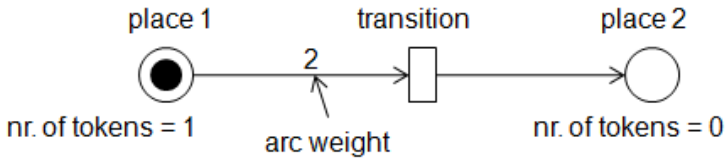
of real-life experiments with plants that on average last two years. To be able to confirm or reject a hypothesis, the plant defence model has to be first developed together with the genes that are interesting candidates for silencing.

The difficulty of the plant defence modelling task is reflected in its complexity and dynamics. The plant defence is complex due to the three highly-interconnected major pathways: salicylic acid (SA), jasmonic acid (JA) and ethylene (ET), whereas the exact relations between the involved 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 reaction rates and inhibition thresholds.

2.2 Choice of modelling formalism

A Petri net (PN) is a graphical and mathematical formalism (Reisig, 1991), chosen for the plant defence model representation. It is a bipartite graph with two types of nodes: places and transitions (see Figure 2). Places denote the resources of the system whereas transitions are the events that modify the system from one state to the other. Places can have zero or a positive number of tokens. The state of the system is called a marking and represents the distribution of tokens over the places. The definition of a PN includes the specification of an initial marking, which assigns a number of tokens to each place. In Figure 2, the initial marking is {1,0}, which corresponds to the number of tokens at places {place 1, place 2}. Weighted arcs connect places with transitions, identifying under which condition the transition can be executed. A transition is enabled if its input places contain at least a required number of tokens (defined by the arc weight). For example, the transition in Figure 2 is not enabled, because the arc weight requires the two tokens from the input place 1, while the place 1 has only one token. The firing of an enabled transition results in the consumption of the tokens from the input places and the production of tokens in its output places (this number is determined by the weights of the outgoing arcs from the transition). These changes in the number of tokens through places represent the dynamic behaviour of the system.

Figure 2 The graphical representation of basic Petri net



There are different types of PNs. Standard PNs are discrete and qualitative. But with their various extensions, PNs allow for specification of both qualitative and quantitative models. Some of the PN extensions are Hybrid Petri nets (HPN), Functional Petri nets (FPN) and HFPN which represents the combination of HPN and FPN (Matsuno et al., 2003). In the following two paragraphs we consider these extensions in more detail.

HPN introduces two types of places: discrete and continuous (Silva and Recalde, 2004). The discrete places can have tokens that are integer numbers which is identical to the basic PN. Additionally, the HPN introduces the continuous variables, which can have

tokens that are real numbers. This formalism is convenient for modelling continuous processes with the switch control (Silva and Recalde, 2004). In systems biology, mechanisms can be analysed from the switch control perspective. For example, biologists believe for several genes to be responsible for the plant defence response. By introducing the different variable types, we can emphasise the switch roll of a certain gene by modelling it with the discrete variable while the other components would be modelled with the continuous variables.

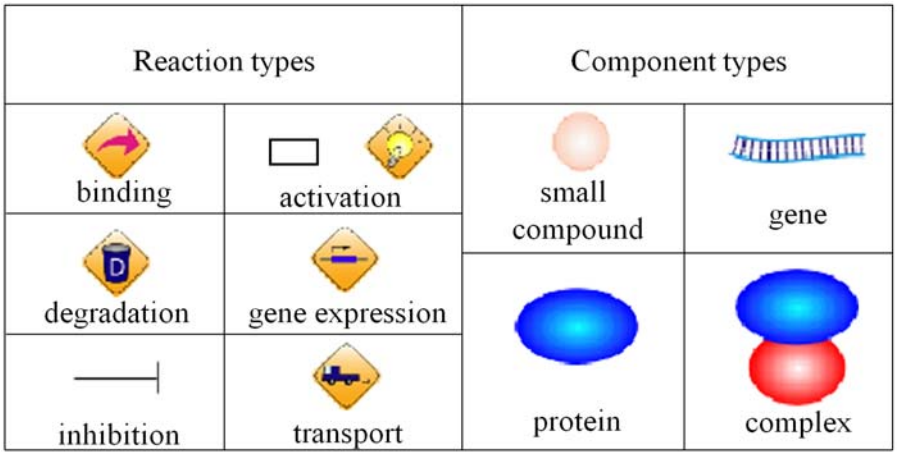
Functional Petri nets (FPN) contain continuous places and transitions (Hofestädt and Thelen, 1998). The firing of transitions has an additional parameter: speed. The speed of firing can depend on the values of the input places (Hofestädt and Thelen, 1998). An additional characteristic of this formalism is that the quantity of consumed places can be different from the quantity of produced places. The rate of the biological reaction is a crucial parameter in signalling networks and therefore was modelled in our work. It turned out that the combination of HPN and FPN formalism, i.e. Hybrid Functional Petri Net (HFPN), would be an appropriate formalism for building plant defence model. HFPN supports modelling of the continuous reactions controlled by the switch mechanisms, and at the same time, the reaction rate depends on the input concentrations, which is often the case in biological systems. As a result of this search and having in mind the overall goal of plant defence simulation model, we used the Cell Illustrator software, that implements HFPN formalism. This software started from the initial version named Genomic Object Net (Nagasaki et al., 2004). More specifically, it allows to model ordinary differential equations (ODE) (Coddington and Levinson, 1955) hidden to the end user through a user-friendly HFPN graphical interface. The software facilitates an easy building of the network topology based on the experts' knowledge. 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 optimisation of dynamic parameters. For this reason, we have used additional combinatorial optimisation parameter search (see subsection 2.6).

2.3 Manual construction of the SA model

The literature studied for the manual model construction was selected by the domain experts. We have focused on scientific papers related to the plant defence response. An ideal case a citation in one review paper or in two scientific papers was a prerequisite to include information into the SA model. If the information was available only in one publication, we critically evaluated the publication quality (e.g. high impact factor, author's relevance in the field) before including this information into the model. The information on the biosynthetic pathways and other available data was mostly acquired from biological databases such as KEGG (Kanehisa and Goto, 2000) and TAIR (Swarbreck et al., 2008). KEGG was used as a basis for constructing the metabolic pathways, the biosynthesis of the hormones and the main reactions involved in these processes. Additional reactions and genes involved were implemented according to the Aracyc database (Rhee et al., 2005). TAIR provided gene information, and synonym names were acquired from iHOP (Hoffmann and Valencia, 2004) and TAIR.

To construct the plant defence response model topology, we defined the types of biological components (molecules) and relations (reactions), shown in Figure 3. Six types of reactions are identified in the SA model: binding, degradation, inhibition, activation, gene expression and transport. 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. Gene expression is defined as the constant or regulated activation of a gene which produces a functional protein. Finally, transport represents the moving of a biological component from one part of the cell to another. It is introduced in a few cases when the biology experts considered it important for the plant defence response mechanism.

Figure 3 Types of biological molecules and reactions modelled with Cell Illustrator (see online version for colours)



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

2.4 Model analysis through simulation

Analysis of the dynamic model behaviour 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 behaviour of the molecules of interest.

2.5 Constraints formulation

Formalising 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 recognised as valuable in the parameter optimisation 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 some initial level, reaches a maximum and then drops back to approximately similar level
- Minimal amplitude and minimal growth of the curve
- Temporal sequence in curve maxima
    - Same time (molecule 1 has the peak at the same time as molecule 2)
    - Maximum before (molecule 1 has the peak before molecule 2)

2.6 Combinatorial optimisation parameter search

Evolutionary algorithms (EAs) are stochastic optimisation methods utilising the ideas of biological evolution in computer problem solving. The main advantage of EAs is their effectiveness and robustness while solving combinatorial optimisation problems which are often intractable by traditional numerical methods. They are nowadays extensively used in science, engineering, management and other domains. Nevertheless, a shortcoming of EAs is their computational complexity which derives from iterative population-based search of the solution space. On the other hand, processing a population of candidate solutions makes EAs amenable to a parallel implementation that results in significant calculation speedup.

One of the popular algorithms within the EA class is the differential evolution algorithm. The differential evolution (DE) algorithm performs a population-based search that optimises the problem defined on a continuous space by iteratively improving a set of candidate solutions with regard to a given measure of quality. The parallel version of the DE algorithm developed by Filipič and Depolli (2009) was used in our work.

The optimal parameter setting of the plant defence model is defined as a combinatorial optimisation problem. Criteria function for the optimisation is the normalised sum of the normalised 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 of 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.

2.7 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.

2.8 Results interpretation

The simulation results are interpreted by the biology experts. The results allow for qualitative conclusions regarding the dynamic behaviour 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 defence response.

3 Results and discussion

Here we provide the details on the SA model and illustrate the process of converging towards the biology experts expectations by presenting the results of three steps of the method described in the previous section.

3.1 Manually constructed SA model

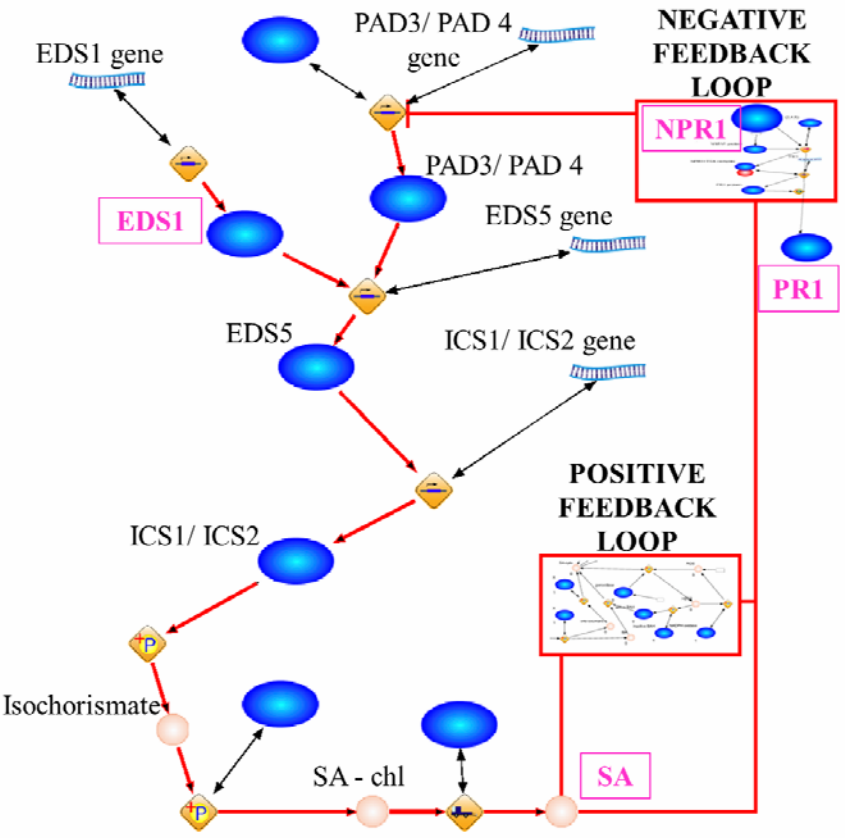
SA belongs to a wide variety of phenolic compounds with a hydroxyl group. Phenolic components are the plant's secondary metabolites with a broad spectrum of functions. SA directly or indirectly influences the seed germination, seedling establishment, cell growth, respiration, stomatal closure, senescence-associated gene expression, responses to abiotic stresses, basal thermo tolerance, nodulation in legumes fruit yield and thermogenesis (Clarke et al., 2004; Klessig and Malamy, 1994; Vlot et al., 2009) and the disease resistance. The SA pathway is the most studied in the literature and is considered to have an important role in both local and systemic acquired resistance (Vlot et al., 2009).

SA in plants is synthesised via two pathways both requiring chorismate as a substrate (Vlot et al., 2009). One pathway goes through a subset of enzymatic reactions initially catalysed by phenylalanine ammonia lyase (PAL). The other pathway, (shown in Figure 4) involves a two-step reaction catalysed with isochorismate sinthase (ICS) and isochorismate pyruvate lyase (IPL) (Vlot et al., 2009). Most of the SA synthesised in response to pathogen attack comes from the ICS/IPL pathway. *Arabidopsis thaliana* encodes two ICS enzymes, ICS1 and ICS2. When plant cell is attacked by pathogens, 90% of SA is synthesised through ICS1 (Vlot et al., 2009).

It has been experimentally shown that the SA concentration rises fast when a virus attacks the plant cell (Baebler et al., 2011; Carr et al., 2010; Uknes et al., 1993; Jeong et al., 2012). It is also known that, when the concentration of SA is too high, the plant

cell will not survive. These experimental results indicate the existence of regulatory mechanisms, more precisely, negative and positive feedback loops, which fine-tune the SA concentration in the cell and allow the plant cell to survive after the virus attack. A negative feedback loop slows down a signalling process, while the positive feedback loop accelerates it. The final cascade product regulates its own concentration by activating or inhibiting the genes involved in its biosynthesis. For example, NPR1 inhibits the expression of EDS1 and PAD3/4 (see Figure 4) (Shah, 2003), genes that activate the production of SA and consequently diminishing its own production, thus forming a negative feedback loop.

Figure 4 The simplified model of SA biosynthesis and signalling pathway represented in Cell Illustrator (see online version for colours)



The manually built SA model represents current knowledge of a plant defence signalling with an emphasis on plant-virus interaction. More specifically, we have based our model on resistant interaction between Arabidopsis and virus TCV. In Arabidopsis, the resistance to TCV is mediated by the R protein HRT (Ibdah et al., 2009) which subsequently induces the signalling cascade leading to plant defence response which

limits viral spread and multiplication. Activation of HRT protein stimulates accumulation of SA (Chandra-Shekara et al., 2004). SA accumulation results in the monomerisation and the activation of NPR1, which consequently triggers the activation of the SA dependent PR proteins (in SA model all PR proteins are represented with a single node PR1) (Maier et al., 2011; Fu et al., 2012; Moore et al., 2011).

The SA model was developed through several iterations. In the next three subsections we demonstrate how the methodology for the parameter optimisation was applied, by showing the results of the three iterative steps. This methodology (see the section Materials and methods) includes the revision of the model structure and the defined constraints after every iteration.

3.2 Step 1 – SA model v1.0

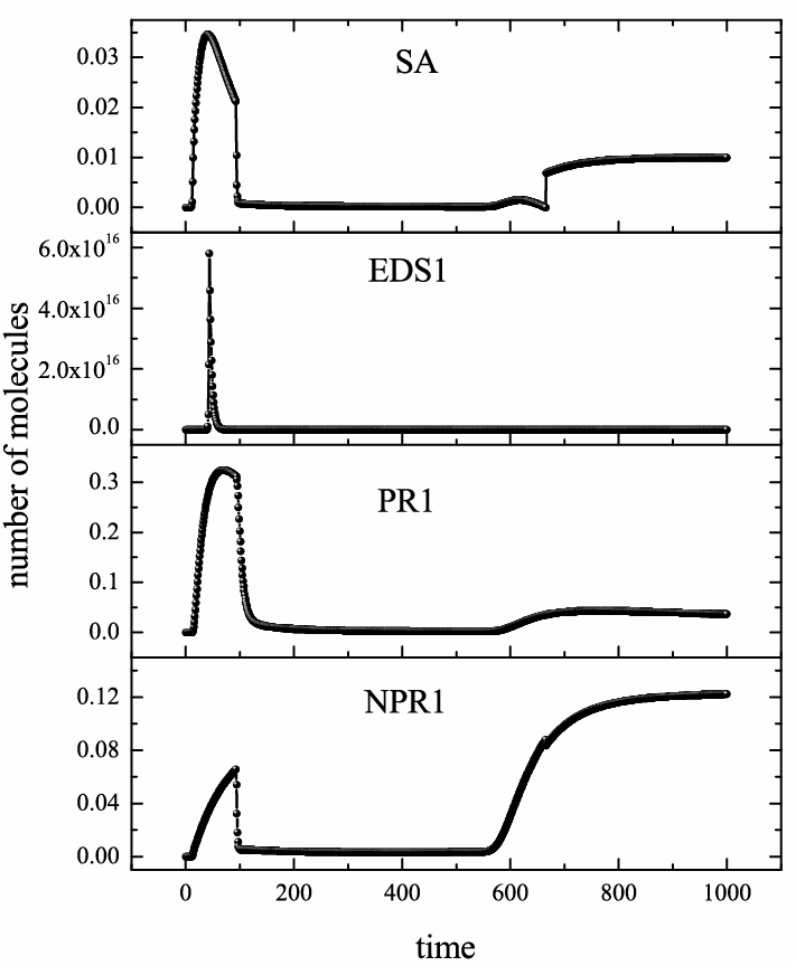
The SA model contains 52 biological molecules and 38 reactions including inhibitions. This was an initial model built in the Cell Illustrator software. The simulator outputs 4 curves (as time series with 1000 points) of each biological molecule, which were the most interesting for the biology scientists. 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 a value that represents the portion of time points in which the constraint was violated. The overall criteria function represents the average of constraint violations.

The following data represent the results of the first optimisation experiment performed on the SA model v1.0. Below are the detailed values for individual constraints and the overall value of the criteria function:

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

Based on the parameter set, the simulator outputs the curves of 4 biological molecules: SA, NPR1, PR1 and EDS1. Apart from the SA molecule, which is a small compound, the other three molecules are proteins. Their dynamic behaviour is shown in Figure 5. 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 of each constraint is violated. This headed to revision of model structure and formalising additional knowledge in the form of constraints.

Figure 5 The dynamic behaviour of the SA, EDS1, PR1 and NPR1 variables based on the optimal set estimated with respect to the criteria function calculated from the 8 constraints during the step 1



3.3 Step 2 – SA model v2.0

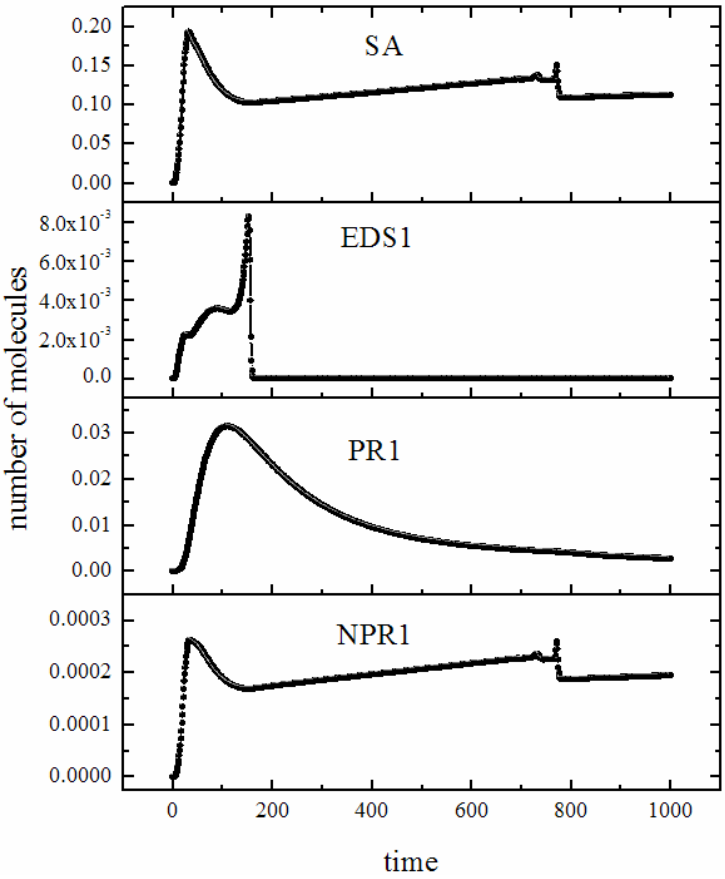
After inspection of the curves from Figure 5, the biology experts have revised the model structure. This correction resulted in a second model version: SA model v2.0 containing 61 biological molecules and 56 reactions. Also, more constraints were specified leading to a set of 33 constraints.

The parameter search was once more performed with the same set up as in the step 1. The following data represent the results of the second optimisation experiment performed on the SA model v2.0. Below are the shortened detailed values for some individual constraints and the overall value of the criteria function:

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

Here the total criteria function showed that on average 0.137 of each constraint is not satisfied. The dynamic curves of the same 4 molecules (SA, EDS1, PR1 and NPR1) are shown in Figure 6. Even though the criteria function shows more violated constraints compared to step 1, the biology experts were more satisfied with the presented curves in Figure 6 with respect to the curves in Figure 5. Nevertheless, since the criteria function was overall worse, the model structure and the constraints were revised again.

Figure 6 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 33 constraints during the step 2



3.4 Step3 – SA model v3.0

Finally, in the last iteration, the SA model converged to the v3.0 containing in total 50 biological molecules and 89 reactions. This means that during the revision of SA model v2.0 the number of biological components decreased from 61 to 50. The biologists have decided to present some of the biological molecules, belonging to the same component families and thus having the same functionality, with the single node. Also, since the number of biological molecules reduced from the SA model v2.0 to SA model v3.0, the number of constraints consequently decreased from 33 in the step 2 to 29 in the step 3. The following data represent the results of the third optimisation experiment performed on the SA model v3.0. Below are all final values for individual constraints and the overall value of the criteria function:

equalRate(Prephenate,Phenyl_pyruvate) = 0.002
equalRate(Prephenate,Phenylalanine) = 0.002
equalRate(Prephenate,Trans_cinnamic_acid) = 0.002
equalRate(Prephenate,Orto_coumaric_acid) = 0.002
equalRate(Prephenate,BA) = 0.002
equalRate(Isochorismate,SA_chl) = 0.000
slowerRate(Prephenate,Isochorismate) = 0.000
slowerRate(EDS1,EDS5) = 0.000
slowerRate(PAD3_4,EDS5) = 0.000
slowerRate(NPR1_oligomer,NPR1) = 0.000
maxSameTime(ROS,HRT) = 0.089
maxAfter(HRT,MPK3) = 0.114
maxSameTime(MPK6,MPK3) = 0.000
maxAfter(MPK6,EDS1) = 0.000
maxSameTime(PAD3_4,EDS1) = 0.000
maxAfter(PAD3_4,EDS5) = 0.011
maxAfter(EDS5,ICS1_2) = 0.000
maxAfter(ICS1_2,SA) = 0.000
maxAfter(ROS,BA2H) = 0.000
maxAfter(ROS,NPR1) = 0.000
maxAfter(NPR1,NPR1_TGA_2_4_5) = 0.011
lowerThan(SA_chl,SA) = 0.002
zeroPeakZero(Chorismate) = 1.000
zeroPeakZero(Prephenate) = 0.000
zeroPeakZero(Phenyl_pyruvate) = 0.000
zeroPeakZero(Phenylalanine) = 0.000
zeroPeakZero(SA) = 0.000
zeroPeakZero(PR1_2_5) = 0.000
stopFast(virus) = 0.007
finalCriteria = 1.243/29 = 0.043

The total criteria function in this step showed that on average 0.043 of each constraint is not satisfied. Compared to steps 1 and 2, the satisfaction of constraints was the best in the final step 3. Additionally, the satisfaction of biology experts was also considerably improved with respect to step 1 and step 2. The final dynamic curves of the same 4

molecules (SA, EDS1, PR1 and NPR1) that meet the criteria of the biological experts are shown in Figure 7. The detailed data for all three iterative optimisation steps are given for comparison in Table 1.

Figure 7 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 29 constraints during the step 3

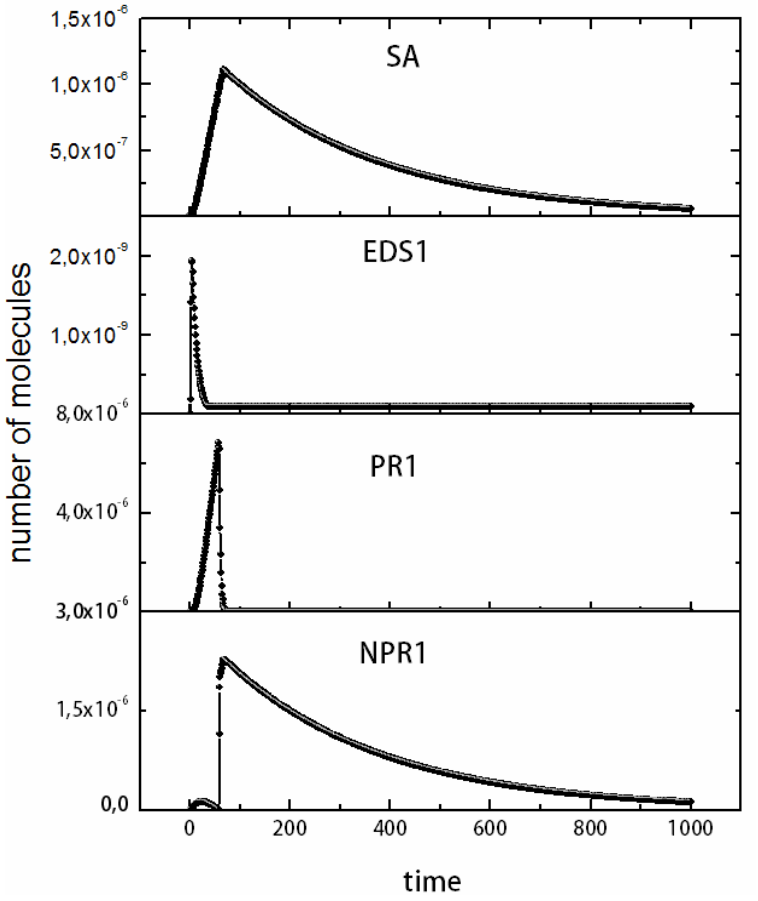


Table 1 Comparison table for three iterative optimisation steps

SA model	*Components*	*Reactions*	*Constraints*	*Averaged criteria function*
V1.0	52	38	8	0.047
V2.0	61	56	33	0.137
V3.0	50	89	29	0.042

The selected parameter set is large, thus, making the search space enormous. This automatically directed us to use some of the stochastic optimisation 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.

4 Conclusion

This paper represents the iterative process of converging to the dynamic parameters of the SA model that best satisfy the experts' evaluation. In the illustrative example of three iterations we show how the dynamic behaviour of the simulated curves improves according to the value of optimisation criteria function and the experts' evaluation. In future work we plan to develop the methodology to compare currently developed SA model and the sparse sets of experimental data and thus provide a numerical validation of this model. The methodology presented in this paper can be extended for the model development of any other biological pathway or mechanism.

Acknowledgements

This work was supported by the Slovenian Research Agency grants P4-0165, J4-2228, J4-4165 and P2-0103, and the AD Futura Agency through the scholarship to the first author of this paper.

References

Baebler, Š., Stare, K., Kovač, M., Blejec, A., Prezelj, N., Stare, T., Kogovšek, P., Pompe-Novak, M., Rosahl, S., Ravnikar, M. and Gruden, K. (2011) 'Dynamics of responses in compatible potato – potato virus Y interaction are modulated by salicylic acid', *PLoS ONE*, Vol. 6, pp.e29009–e29009.

Boggs, P.T. and Tolle, J.W. (1995) 'Sequential quadratic programming', *Acta Numerica*, Vol. 4, pp.1–51.

Carr, J.P., Lewsey, M.G. and Palukaitis, P. (2010) 'Signaling in induced resistance', *Natural and Engineered Resistance to Plant Viruses*, Pt Ii 76, pp.57–121.

Chandra-Shekara, A.C., Navarre, D., Kachroo, A., Kang, H.G., Klessig, D. and Kachroo, P. (2004) 'Signaling requirements and role of salicylic acid in HRT- and rrt-mediated resistance to turnip crinkle virus in Arabidopsis', *Plant Journal*, Vol. 40, pp.647–659.

Cho, K.H. and Wolkenhauer, O. (2003) 'Analysis and modelling of signal transduction pathways in systems biology', *Biochemical Society Transactions*, Vol. 31, pp.1503–1509.

Clarke, S.M., Mur, L.A.J., Wood, J.E. and Scott, I.M. (2004) 'Salicylic acid dependent signaling promotes basal thermotolerance but is not essential for acquired thermotolerance in *Arabidopsis thaliana*', *Plant Journal*, Vol. 38, pp.432–434.

Coddington, E.A. and Levinson, N. (1955) *Theory of Ordinary Differential Equations*, McGraw-Hill, New York.

Eiben, A.E. and Smith, J.E. (2003) *Introduction to Evolutionary Computing*, Springer-Verlag, Berlin.

Filipič, B. and Depolli, M. (2009) 'Parallel evolutionary computation framework for single- and multiobjective optimization', *Parallel Computing*, pp.217–240.

Fu, Z.Q., Yan, S., Saleh, A., Wang, W., Ruble, J., Oka, N., Mohan, R., Spoel, S.H., Tada, Y., Zheng, N. and Dong, X. (2012) 'NPR3 and NPR4 are receptors for the immune signal salicylic acid in plants', *Nature*, Vol. 486, pp.228–232.

Genoud, T., Santa Cruz, M.B.T. and Métraux, J.P. (2001) 'Numeric simulation of plant signaling networks', *Plant Physiology*, Vol. 126, pp.1430–1437.

Hofestädt, R. and Thelen, S. (1998) 'Quantitative modelling of biochemical networks', *In Silico Biology*, Vol. 1, pp.39–53.

Hoffmann, R. and Valencia, A. (2004) 'A gene network for navigating the literature', *Nature Genetics*, Vol. 36, p.664.

Ibdah, M., Chen, Y.T., Wilkerson, C.G. and Pichersky, E. (2009) 'An aldehyde oxidase in developing seeds of Arabidopsis converts benzaldehyde to benzoic acid', *Plant Physiology*, Vol. 150, pp.416–423.

Janes, K. and Lauffenburger, D. (2006) 'A biological approach to computational models of proteomic networks', *Current Opinion in Chemical Biology*, Vol. 10, pp.73–80.

Jeong, R.D., Zhu, S., Kachroo, A. and Kachroo, P. (2012) 'New insights into resistance protein-mediated signaling against turnip crinkle virus in Arabidopsis', *Journal of Plant Biochemistry and Biotechnology*, Vol. 21, pp.48–51.

Kanehisa, M. and Goto, S. (2000) 'KEGG: Kyoto Encyclopedia of genes and genomes', *Nucleic Acids Research*, Vol. 28, pp.27–30.

Kestler, H.A., Wawra, C., Kracher, B. and Kühl, M. (2008) 'Network modeling of signal transduction: establishing the global view', *Bioessays*, Vol. 30, pp.1110–1125.

Kirkpatrick, S., Gelatt, C.D. and Vecchi, M.P. (1983) 'Optimization by simulated annealing', *Science*, Vol. 220, pp.671–680.

Klessig, D.F. and Malamy, J. (1994) 'The salicylic acid signal in plants', *Plant Molecular Biology*, Vol. 26, pp.1439–1458.

Klipp, E. and Liebermeister, W. (2006) 'Mathematical modelling of intracellular signaling pathways', *BMC Neuroscience*, No. 7.

Levenberg, K. (1944) 'A method for the solution of certain non-linear problems in least squares', *Quarterly of Applied Mathematics*, Vol. 2, pp.164–168.

Maier, F., Zwicker, S., Hückelhoven, A., Meissner, M., Funk, J., Pfitzner, A.J.P., Pfitzner, U.M. (2011) 'NONEXPRESSOR OF PATHOGENESIS-RELATED PROTEINS1 (NPR1) and some NPR1-related proteins are sensitive to salicylic acid', *Molecular Plant Pathology*, Vol. 12, pp.73–91.

Marquardt, D. (1963) 'An algorithm for least-squares estimation of nonlinear parameters', *SIAM Journal on Applied Mathematics*, Vol. 11, pp.431–441.

Matsuno, H., Tanaka, Y., Aoshima, H., Doi, A., Matsui, M. and Miyano, S. (2003) 'Biopathways representation and simulation on hybrid functional Petri net', *In Silico Biology*, Vol. 3, pp.389–404.

Miljkovic, D., Stare, T., Mozetič, I., Podpečan, V., Petek, M., Witek, K., Dermastia, M., Lavrač, N. and Gruden, K. (2012) 'Signalling network construction for modelling plant defence response', *PLOS ONE*, Vol. 7, p.e51822.

Mitchell, M. (1996) *An Introduction to Genetic Algorithms*, MIT Press, Cambridge, MA.

Moore, J.W., Loake, G.J. and Spoel, S.H. (2011) 'Transcription dynamics in plant immunity', *Plant Cell*, Vol. 23, pp.2809–2820.

Nagasaki, M., Doi, A., Matsuno, H. and Miyano, S. (2004) 'Genomic Object Net:I. A platform for modeling and simulating biopathways', *Applied Bioinformatics*, Vol. 2, pp.181–184.

Olmedo, G., Guo, H.W., Gregory, B.D., Nourizadeh, S.D., Aguilar-Henonin, L., Li, H.J., An, F.Y., Guzman, P. and Ecker, J.R. (2006) '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 of the United States of America*, No. 103, pp.13286–13293.

Reisig, W. (1991) 'Petri nets and algebraic specifications', *Theoretical Computer Science*, Vol. 80, pp.1–34.

Reymond, P. and Farmer, E.E. (1998) 'Jasmonate and salicylate as global signals for defense gene expression', *Current Opinion in Plant Biology*, Vol. 1, pp.404–411.

Rhee, S.Y., Zhang, P., Foerster, H. and Tissier, C. (2005) 'AraCyc: overview of an Arabidopsis metabolism database and its applications for plant research', in Saito, K., Dixon, R. and Willmitzer, L. (Eds): *Biotechnology in Agriculture and Forestry: Plant Metabolomics*, Springer, pp.141–153.

Shah, J. (2003) 'The salicylic acid loop in plant defense', *Current Opinion in Plant Biology*, Vol. 6, pp.365–371.

Silva, M. and Recalde, L. (2004) 'On fluidification of Petri net models: from discrete to hybrid and continuous models', *Annual Reviews in Control*, Vol. 28, pp.253–266.

Staswick, P.E. (2008) 'JAZing up jasmonate signaling', *Trends in Plant Science*, Vol. 13, pp.66–71.

Storn, R. and Price, K. (1997) 'Differential evolution – a simple and efficient heuristic for global optimization over continuous spaces', *Journal of Global Optimization*, Vol. 11, pp.341–359.

Swarbreck, D., Wilks, C., Lamesch, P., Berardini, T.Z., Garcia-Hernandez, M., Foerster, H., Li, D., Meyer, T., Muller, R., Ploetz, L., Radenbaugh, A., Singh, S., Swing, V., Tissier, C., Zhang, P. and Huala, E. (2008) 'The Arabidopsis Information Resource (TAIR): gene structure and function annotation', *Nucleic Acids Research*, Vol. 36, pp.D1009–D1014.

Uknes, S., Winter, A., Delaney, T., Vernooij, B., Morse, A., Friedrich, L., Nye, G., Potter, S., Ward, E. and Ryals, J. (1993) 'Biological induction of systemic acquired resistance in Arabidopsis', *Molecular Plant-Microbe Interactions*, Vol. 6, pp.692–698.

Vlot, A.C., Dempsey, D.A. and Klessig, D.F. (2009) 'Salicylic acid, a multifaceted hormone to combat disease', *Annual Review of Phytopathology*, Vol. 47, pp.177–206.

Walter, E. and Pronzato, L. (1994) *Identification of Parametric Models from Experimental Data*. MASSON, Paris.

Zhao, Z., Xia, J., Tastan, O., Singh, I., Kshirsagar, M., Carbonell, J. and Klein-Seetharaman, J. (2011) 'Virus interactions with human signal transduction pathways', *International Journal of Computational Biology and Drug Design*, Vol. 4, pp.83–105.