

# **SEMI-AUTOMATED KNOWLEDGE ELICITATION FOR MODELLING PLANT DEFENCE RESPONSE**

Dragana Miljković

**Doctoral Dissertation**  
**Jožef Stefan International Postgraduate School**  
**Ljubljana, Slovenia, November 2013**

**Evaluation Board:**

*Prof. Dr. Kristina Gruden, Chairman, National Institute of Biology, Ljubljana, Slovenia*

*Prof. Dr. Bojan Orel, Member, Faculty of Computer and Information Science, University of Ljubljana, Ljubljana, Slovenia*

*Prof. Dr. Werner Dubitzky, Member, University of Ulster, United Kingdom*



Dragana Miljković

# **SEMI-AUTOMATED KNOWLEDGE ELICITATION FOR MODELLING PLANT DEFENCE RESPONSE**

**Doctoral Dissertation**

## **POLAVTOMATSKO ZAJEMANJE ZNANJA ZA MODELIRANJE OBRAMBNEGA ODZIVA RASTLIN**

**Doktorska disertacija**

*Supervisor:* Prof. Dr. Nada Lavrač  
*Co-Supervisor:* Asst. Prof. Dr. Igor Mozetič

Ljubljana, Slovenia, November 2013



To my grandfather



# Index

<b>Abstract</b>	<b>IX</b>
<b>Povzetek</b>	<b>XI</b>
<b>Abbreviations</b>	<b>XIII</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Biological background . . . . .	1
1.2 Overview of related work . . . . .	4
1.2.1 Biological data sources . . . . .	4
1.2.2 Modelling of biological mechanisms in plants . . . . .	5
1.2.3 Text processing in systems biology . . . . .	6
1.2.4 Optimisation methods in systems biology . . . . .	7
1.3 Hypothesis and goals . . . . .	8
1.4 Scientific contributions . . . . .	8
1.5 Organisation of the thesis . . . . .	10
<b>2 Modelling of Dynamic Systems: Background and Related Work</b>	<b>11</b>
2.1 Modelling approaches . . . . .	11
2.1.1 Concepts and classification . . . . .	11
2.1.2 Construction of mathematical models . . . . .	13
2.1.3 Selected modelling approaches . . . . .	15
2.1.4 Illustrative example . . . . .	16
2.2 Modelling approaches in systems biology . . . . .	21
2.3 Text processing in systems biology . . . . .	24
<b>3 Methodology for PDS Model Construction: An Overview</b>	<b>27</b>
<b>4 Manual Development of PDS Model Structure</b>	<b>35</b>
4.1 Definition of PDS reactions and components . . . . .	35
4.2 PDS model structure . . . . .	36
4.2.1 Directed edge-labelled graph presentation . . . . .	38
4.2.2 HFPN presentation . . . . .	40
4.2.3 Comparison with the state-of-the-art model structures of the PDS mechanism . . . . .	45
<b>5 Constraint-Driven PDS Parameter Optimisation and Model Validation</b>	<b>47</b>
5.1 PDS model represented in the HFPN formalism . . . . .	47
5.2 Iterative process of constraints formulation, combinatorial optimisation and refinement of the model and constraints . . . . .	49

5.2.1	Mathematical optimisation . . . . .	49
5.2.2	Evolutionary algorithms and DE algorithm selection . . . . .	50
5.2.3	Constraint formulation and objective function definition . . . . .	51
5.2.4	Iterative revisions of SA sub-model . . . . .	51
5.3	Model validation . . . . .	59
5.3.1	Concepts and strategies . . . . .	59
5.3.2	Sensitivity analysis: exploring the robustness of the SA sub-model . .	62
5.3.3	Comparison with experimental data . . . . .	68
<b>6</b>	<b>Model Structure Revision</b>	<b>73</b>
6.1	Bio3graph methodology . . . . .	73
6.2	Bio3graph implementation and workflow availability . . . . .	79
6.3	Evaluation of the Bio3graph results . . . . .	80
6.4	Modelling knowledge of domain experts . . . . .	81
6.5	PDS model structure extracted by Bio3graph from biological literature . . . .	82
6.6	The merged PDS model structure . . . . .	86
<b>7</b>	<b>Incremental Revision of Structures of Biological Models</b>	<b>91</b>
7.1	Methodology for incremental biological network revision . . . . .	91
7.1.1	Literature retrieval . . . . .	92
7.1.2	Graph merging . . . . .	93
7.1.3	Redundant relation removal . . . . .	93
7.1.4	Colour reset . . . . .	94
7.2	Implementation and the workflow availability . . . . .	94
7.2.1	Literature retrieval . . . . .	94
7.2.2	Graph merging . . . . .	95
7.2.3	Redundant relation removal . . . . .	95
7.2.4	Colour reset . . . . .	95
7.2.5	The workflow . . . . .	95
7.3	Use case 1: a simple plant defence model structure . . . . .	97
7.3.1	The Initial graph . . . . .	97
7.3.2	The Triplet graph . . . . .	97
7.3.3	First incremental step . . . . .	99
7.3.4	Second incremental step . . . . .	99
7.4	Use case 2: a complex plant defence model structure . . . . .	101
<b>8</b>	<b>Conclusions and Further Work</b>	<b>105</b>
8.1	Summary . . . . .	105
8.2	Further work . . . . .	108
<b>9</b>	<b>Acknowledgements</b>	<b>111</b>
<b>10</b>	<b>References</b>	<b>113</b>
	<b>Publications Related to the Dissertation</b>	<b>123</b>
	<b>Index of Figures</b>	<b>125</b>
	<b>Index of Tables</b>	<b>131</b>
	<b>Appendix</b>	<b>135</b>
A	Manually constructed PDS model structure . . . . .	135
A.1	The summary of all relations in the manual model . . . . .	135
A.2	The levels of the biological component abstraction . . . . .	155



	A.3	A manual expanded graph file at the single component level . . . . .	160
B		PDS model structure revision . . . . .	173
	B.1	Component vocabulary . . . . .	173
	B.2	Reactions vocabulary . . . . .	178
	B.3	Graph of new triplets extracted by Bio3graph . . . . .	180
	B.4	Merged PDS model structure . . . . .	184
C		Biography . . . . .	199



# Abstract

In biological sciences, the growth of experimental data is not uniform for different types of biological mechanisms, hence some biological mechanisms still have few datasets available. The thesis describes a novel methodology for the construction of biological models by eliciting the relevant knowledge from literature and domain experts. The methodology has been applied to build the model of defence response in plants, and can be used to construct models of other biological mechanisms. The methodology addresses two aspects of biological model construction: the model structure as well as the model dynamics.

The motivation of the developed methodology is to enable building biological models without or with scarce experimental data. The methodology consists of several steps, where the standard approach to the construction of dynamic models is enhanced with the following methods: a method for model structure revision by means of natural language processing techniques, a method for incremental model structure revision, and a method for automatic optimisation of model parameters guided by the expert knowledge in the form of constraints.

In the proposed approach, the initial model structure was constructed manually by defining the representation formalism, encoding the information from public databases and literature, and composing a pathway diagram. In order to complement the model structure with potentially missing relations, a new approach to automated information extraction from biological literature was developed. This approach, named Bio3graph, allows for automated extraction of biological relations in the form of triplets followed by the construction of a graph structure which can be visualised, compared to the manually constructed model structure, and examined by the experts. Using a plant defence response vocabulary of components and reaction types, Bio3graph was applied to a set of 9,586 relevant full text articles, resulting in 137 newly detected relations. The resulting pathway diagram of plant defence signalling represents a valuable source for further computational modelling and interpretation of omics data. The Bio3graph approach, implemented as an executable language processing and graph visualisation workflow, can be utilised for modelling other biological systems, given that an appropriate vocabulary is provided.

An incremental variant of the Bio3graph tool was developed to enable easy updating of a given model structure with new relations. The incremented approach allows for periodic updates of network structures based on newly published scientific literature. The incremental approach was demonstrated on two use cases. In the first use case, a simple plant defence network with 37 components and 49 relations, created manually, was extended in two incremental steps yielding the extended model with 183 relations. In the second use case, a complex model structure of defence response in *Arabidopsis thaliana*, consisting of 175 nodes and 524 relations, was incrementally updated with the information extracted from recently published articles, resulting in an enhanced network with 628 relations. The results show that by using the incremental approach it is possible to follow the development of knowledge of specific biological relations in recent literature. The implemented components offer an effective way of merging and visualising model structures and the triplet graphs generated from texts, thus enabling fast discovery of new relations.

One obstacle in developing dynamic models, useful for simulation, is the lack of kinetic data from which the model parameters could be determined. This problem was addressed

by proposing a method for iterative improvement of model parameters until the simulation results come close to the expectations of the biology experts. These expectations can be formulated as constraints to be satisfied by model simulations. To estimate the parameters of the salicylic acid pathway, the most important pathway of plant defence response, three iterative steps were performed in order to meet the expectations of the biology experts. The method enabled us to improve results of individual simulations and to optimise model parameters which provide a deeper insight into the observed biological system. As a result, the constraint-driven optimisation approach allows for efficient exploration of the dynamic behaviour of biological models and, at the same time, increases their reliability.

The thesis also contributes to publicly available biological models and scientific software. The structure of the developed models of defence response in *Arabidopsis thaliana* in the form of directed graphs is available for download. Also, the implemented Bio3graph approach for triplet extraction from literature is provided as a publicly accessible scientific workflow.

## Povzetek

V bioloških znanostih količina eksperimentalnih podatkov o različnih bioloških mehanizmih narašča neenakomerno in za nekatere biološke mehanizme je na voljo le malo podatkov. Disertacija opisuje novo metodologijo za gradnjo bioloških modelov z zajemanjem znanja iz literature in ekspertnega znanja v domenah, kjer primanjkuje eksperimentalnih podatkov. Metodologija je bila uporabljena za gradnjo modela obrambnega mehanizma rastlin, v splošnem pa jo je mogoče uporabiti tudi za gradnjo modelov drugih bioloških mehanizmov. Metodologija obravnava gradnjo bioloških modelov tako z vidika strukture modela kot tudi njegove dinamike.

Cilj razvite metodologije je omogočiti gradnjo bioloških modelov brez, oziroma z majhno količino razpoložljivih eksperimentalnih podatkov. Sestavljena je iz več korakov, pri čemer je običajni pristop h gradnji dinamičnih modelov dopolnjen z naslednjimi metodami: uporaba tehnik obdelave naravnega jezika za izboljšavo strukture modela, postopek inkrementalne izboljšave modela ter avtomatska optimizacija parametrov, vodena z ekspertnim znanjem v obliki omejitev.

V predlaganem pristopu je bila začetna struktura modela zgrajena ročno z določitvijo formalizma predstavitve modela, vnosom informacij iz javno dostopnih podatkovnih baz in literature ter z izgradnjo bioloških poti. Da bi strukturo modela lahko dopolnili z morebitnimi manjkajočimi relacijami, smo razvili nov pristop k avtomatskemu zajemanju informacij iz biološke literature. Ta pristop, ki smo ga poimenovali Bio3graph, omogoča avtomatsko luščenje bioloških relacij v obliki trojčkov ter gradnjo omrežja, ki ga lahko prikažemo v obliki grafa in primerjamo z ročno zgrajenim modelom, omogoča pa tudi obravnavo s strani ekspertov. S slovarjem komponent in vrst reakcij obrambnega mehanizma rastlin je bil pristop Bio3graph uporabljen na množici 9.586 znanstvenih člankov, kar je privedlo do odkritja 137 novih relacij. Izpopolnjen obrambni mehanizem rastlin predstavlja dragocen vir za nadaljnje računalniško modeliranje in razlago bioloških podatkov. Pristop Bio3graph, ki je implementiran kot delotok komponent za procesiranje naravnega jezika ter grafičnega prikaza omrežij, je mogoče uporabiti tudi za modeliranje drugih bioloških sistemov, če je na voljo primeren slovar komponent in relacij.

Da bi omogočili enostavno nadgradnjo strukture modela z novimi relacijami, smo razvili še inkrementalno različico pristopa Bio3graph. Le-ta omogoča ponavljajoče nadgrajevanje omrežnih struktur modelov z upoštevanjem novo objavljene znanstvene literature. Uporaba inkrementalnega pristopa je prikazana na dveh primerih. V prvem primeru je bil preprost, ročno zgrajen model obrambnega mehanizma rastlin s 37 komponentami in 49 relacijami nadgrajen v dveh zaporednih korakih, ki sta privedla do končnega, razširjenega modela s 183 relacijami. V drugem primeru pa je bila nadgrajena struktura naprednega modela obrambnega mehanizma navadnega repnjakovca (*Arabidopsis thaliana*), ki je vseboval 175 komponent in 524 relacij. Struktura modela je bila razširjena z informacijami, pridobljenimi iz novo objavljenih znanstvenih člankov, kar je privedlo do izboljšane strukture s 628 relacijami. Rezultati kažejo, da je z uporabo inkrementalnega pristopa mogoče slediti znanstvenim spoznanjem o izbranih bioloških relacijah v novo objavljeni literaturi. Implementirane komponente nudijo učinkovit način zlivanja in grafičnega prikazovanja struktur modelov ter omrežij trojčkov, pridobljenih iz besedil, kar omogoča hitro odkrivanje novih relacij.

Ena od ovir pri razvoju dinamičnih modelov, ki so primerni za simulacijo, je pomanjkanje kinetičnih podatkov, s pomočjo katerih je mogoče določiti parametre modela. Predlagana rešitev temelji na iterativni izboljšavi parametrov, ki se lahko izvaja vse dokler rezultati simulacije ne zadostijo pričakovanjem ekspertov. Takšna pričakovanja je mogoče izraziti v obliki omejitev, ki jim mora zadostiti simulacija modela. Da bi ocenili parametre poti salicilne kisline, ki je najpomembnejša pot v obrambnem mehanizmu rastlin, smo izvedli tri zaporedne korake izboljšave parametrov. Predlagana metoda je omogočila izboljšavo rezultatov posameznih simulacij ter optimizacijo parametrov modela, to pa je pripomoglo tudi k boljšemu vpogledu v opazovani biološki sistem. Zaključimo lahko, da optimizacija z omejitvami omogoča učinkovito raziskovanje dinamičnega obnašanja bioloških modelov, poveča pa tudi njihovo zanesljivost.

Disertacija doprinaša tudi k javno dostopnim biološkim modelom ter programski opremi v znanosti. Razvite strukture modelov obrambnega mehanizma navadnega repnjakovca (*Arabidopsis thaliana*) so prosto dostopne na spletu, prav tako pa je širši javnosti na voljo tudi implementacija pristopa Bio3graph v obliki delotoka.

## List of Abbreviations

ABA	=	abscisic acid
CI	=	Cell Illustrator
DE	=	differential evolution
DNA	=	deoxyribonucleic acid
EA	=	evolutionary algorithm
ET	=	ethylene
FPN	=	Functional Petri net
GA	=	gibberellic acid
HFPN	=	Hybrid Functional Petri net
HPN	=	Hybrid Petri net
HRT	=	hypersensitive response to TCV
HTML	=	hypertext markup language
IE	=	information extraction
JA	=	jasmonic acid
KEGG	=	Kyoto encyclopedia of genes and genomes orthology
mEPN	=	modified Edinburgh Pathway Notation
NCBI	=	National Center for Biotechnology Information
NLP	=	natural language processing
ODE	=	ordinary differential equation
PMC	=	PubMed Central
PN	=	Petri net
PDS	=	plant defence signalling
RNA	=	ribonucleic acid
SA	=	salicylic acid
SPN	=	Stochastic Petri net
SQUAD	=	Standardized Qualitative Dynamical Systems
TAIR	=	The Arabidopsis Information Resource
TCV	=	turnip crinkle virus
TuMV	=	turnip mosaic virus
XML	=	extended markup language





# 1 Introduction

*My goal is simple. It is complete understanding of the universe, why it is as it is and why it exists at all.*

Stephen Hawking

In recent years, computer science is rapidly developing new tools that are used to store, maintain and analyse biological information. On the other hand, with the development of new techniques like DNA microarrays, biologists are able to perform more experiments and produce higher amounts of experimental data. In such circumstances, at the intersection of computer science and biology, new disciplines have emerged, one of them being systems biology.

This chapter introduces the biological background of the topic, the context of this thesis, the terminology used and describes the motivation for the work. In addition, we present the related work relevant for the thesis, the research hypothesis and the main thesis goals followed by the list of scientific contributions. The chapter concludes by outlining the thesis structure.

## 1.1 Biological background

The term systems biology was used for the first time in 1978 (Hogeweg and Hesper, 1978). Since then it has developed to a specific branch of bioinformatics that is concerned with modelling of dynamic systems behaviour and attracts raising interest of both computer scientists and biology experts. After having used the reductionist approach in biology and formulating knowledge on separate small subsets of the global biological mechanisms, the need for unifying the knowledge and studying biological systems as a "whole" has emerged. Also, a "whole" is more than just a simple sum of its parts, since there are additional relationships between the parts that influence the dynamic behaviour of the "whole". The goal of systems biology is to study biological processes at the system level and construct their models by using different computer science and mathematical approaches (Kitano, 2002).

The basic concepts in systems biology include biological components, reactions and networks. In general, there are several types of biological components of interest:

- Nucleic acids. Five nucleic acids form DNA and RNA chains (Bean, 1973). DNA contains genetic information, i.e. a specific gene codes for a specific protein, while RNA is necessary for the process of extracting the genetic information, i.e. protein synthesis.
- Genes. A gene is a unit of inheritance. It is a sequence of DNA that codes for a specific function or a protein. The production of either RNA or protein is named gene expression, while the resulting molecule represents a gene product (Sarkar, 2008). A gene can be over- or underexpressed. The overexpressed genes produce their products in increased quantities, while the underexpressed genes produce products in a smaller

quantity than expected. Gene expression is influenced by activator or inhibitor proteins, which activate or inhibit the gene expression process.

- **Proteins.** Proteins are organic molecules made of amino-acid sequences (Branden, 1999). The functions of proteins are manifold. They have a structural function (all organisms are made of proteins), a mechanical function (muscles are getting contracted as a result of protein activity), a function in cell signalling, immune responses, cell cycle, a function of an enzyme (catalyzing biochemical reactions), etc.
- **Small compounds (metabolites).** They are organic molecules with lower molecular weight, which are not polymers by definition. They include water, different minerals and free radicals. Small compounds are the reactants and the products in metabolic networks.

All the above mentioned biological molecules participate in many biological reactions. They can also have different roles depending on the reaction type, such as protein activation, positive regulation, phosphorylation, protein inhibition, negative regulation, dephosphorylation, ubiquitination, binding, dissociation, transport, etc. Protein activation represents its transformation from an inactive into its active form which allows the protein to carry out its function. In positive regulation, a transcription factor has to bind at the promoter in order to enable or enhance gene expression (gene expression is defined as the constant or regulated activation of a gene which produces a functional protein). Phosphorylation represents the addition of a phosphate group to a protein or other organic molecule, thus turning many protein enzymes on and off changing at the same time their function and activity. Inhibition of a protein is, contrary to its activation, the process of transforming its active form into an inactive one and thus blocking its function. In negative regulation, a repressor protein binds to an operator to prevent a gene from being expressed. Dephosphorylation is removing the phosphate group from a protein or other organic molecule which modifies its performance. In ubiquitination, a protein is inactivated by attaching ubiquitin to it, which acts as a tag that signals the cell machinery to carry the protein to the proteasome for the final degradation. Binding is joining of two proteins, protein and a gene or protein and a small compound. Dissociation is a process in which the molecules separate or split into smaller particles. Transport represents the moving of a biological component from one part of the cell to another.

A biological pathway or a network represents a sequence of reactions by which one biological material is converted to another. In systems biology there are three main types of biological networks (Estrada, 2011):

- **Gene regulatory networks (or transcriptional regulatory networks).** The main components in these networks are genes. Depending on the level of abstraction these networks can describe very detailed gene expression processes, but also give a more general overview when representing only the fact that a particular biological molecule activates or inhibits some gene.
- **Signalling networks (or signal transduction networks).** The main components in these networks are proteins. In signalling networks, the product of a reaction (a protein) becomes an enzyme for the next reaction in the network. Signalling pathways consist of combinations of catalytic reactions, complex formations and transportations (Takai-Igarashi, 2005). Additionally, the reaction rate depends on the concentration of the reactants.
- **Metabolic networks.** The main components are small compounds. In metabolic networks, the product of one reaction (a small compound) becomes a reactant for the next one. The proteins in these reactions have a role of an enzyme only. Metabolic

pathways consist mostly of catalytic reactions which are speeded up by the enzymes (proteins). In metabolic networks, the speed of reaction depends on the concentration of the reacting molecules, but not to the same degree as in signalling networks.

All biological mechanisms involve one or more types of biological networks, depending on their own complexity. The particular mechanism that we investigate in this dissertation is plant defence response to pathogen attacks. Gene expressions in plants are highly responsive and easily triggered by many factors in the physical environment, including biotic factors. One of these factors are pathogens, which generally represent a serious threat to plants and can lead to fitness costs, physiological damage or even death. However, plants have developed sophisticated mechanisms that can effectively fight off infections with various pathogens. In a reverse manner, effective pathogens cause disease because they are able to avoid recognition or suppress plant defence mechanisms or both. Upon pathogen recognition, plants trigger a complex signalling network, referred to as plant defence response or plant defence signalling (PDS).

Regardless of whether the final result of the response is resistance or susceptibility, PDS is critically determined by small compounds with hormonal functions, such as salicylic acid (SA), jasmonic acid (JA) and ethylene (ET). It was shown that these three molecules play a fundamental role in mediating the defence signalling response in plants (Reymond and Farmer, 1998). Depending on their nature, pathogens trigger either SA-mediated (Glazebrook, 2005) or JA/ET-mediated (Lai et al., 2011) defence pathways in plants. Antagonism of JA towards SA and its synergism to ET was elucidated long ago (Robert-Seilanianantz et al., 2011). ET-based synergism is not only limited to JA, it can be towards SA as well (Pieterse et al., 2009). Recent studies conclude that SA/JA signalling pathways are interconnected and if the molecules are inhibited in one, the response is shifted towards the counterparts (Sato et al., 2010). For a successful defence, the activation of PDS must be rapid, efficient and targeted (Moore et al., 2011).

Understanding the mechanistic basis why a specific pathogen triggers illness in one host plant and not in another has long intrigued plant pathologists, but is still not fully resolved. There are several reasons for this. First, PDS is a very complex mechanism (Kunkel, 2002). It consists of at least three interconnected signalling pathways (the SA, JA and ET pathway) resulting in a dense biological network with complex cell mechanisms, such as positive and negative feedback loops. Second, due to the complexity, there are only subsets of the whole mechanism investigated. As a consequence, there is still a lot of information missing regarding how the mechanism functions as a whole. Third, there is a lack of kinetic biological studies that would potentially reveal new knowledge regarding the dynamic behaviour of the system.

The motivation of biology experts to develop a more comprehensive model to simulate the entire defence response of plants to virus attacks is two-fold. Firstly, such model would provide a better understanding of the complex defence response mechanism in plants which means highlighting connections in the network and understanding how the components operate. More specifically, the biology experts are interested in the investigation which intermediate biological reactions may have a larger impact on the final defence response against pathogens in plants. Secondly, the prediction of experimental results through simulation may save experimental time and indicate further research directions to the biology experts. For example, experts are interested in investigating plant behaviour when some of the important genes in the defence response are silenced. The assumption is that these investigations will result in disease resistant plants. For all these reasons, there is a need to find the appropriate means to develop the global PDS model for simulation purposes.

The first two questions that arise in the course of model construction in systems biology (Kitano, 2001), which we focus on in this thesis, are:

- Network structure identification. The structure of the system has to be defined as

a first step, which means that all the biological components of interest have to be identified and types of links between them specified.

- **Parameter identification.** After obtaining a network structure, relationships between biological components and their dynamics have to be defined as well.

To satisfy the need for discovering new knowledge in plant research, where collecting data is difficult, slow and expensive, we propose in this dissertation a method to construct the PDS model by semi-automatic elicitation of experts knowledge. By combining different methods of natural language processing (Cohen, 2004) and mathematical optimisation (Storn and Price, 1995), we show how to semi-automatically acquire the structure of biological models from the available domain resources (data and literature), determine their parameters and explore their dynamic behaviour by simulation. Additionally, we suggest a new model evaluation method, i.e. to compare the simulation curves and the experimental curves of time series data with few data points. The methodology developed in this thesis can be further extended to any biological model development.

Given that this work is relevant for both computer science and biology, Table 1 introduces the mapping of terms used in these two fields.

Table 1: Mapping between biology and computer science terminology used in this thesis.

Biology	Computer science
Biological system	Computer model, Network
Pathway	Sub-model, sub-network
Molecule, Component	Node, vertex
Reaction, Interaction	Relation, Edge, Arc, Link, Connection

## 1.2 Overview of related work

This section provides a brief overview of publicly available sources of biological data applicable in this work. The related work on modelling in plant research is also summarised. The section concludes with an overview of different text processing approaches as well as optimisation methods used in systems biology.

### 1.2.1 Biological data sources

Rapid collection of biological data is a direct consequence of enormous technological advances that resulted in extraordinary quantities of biological information. The key to handling this burgeoning information was the recruitment of computers to help systematically analyse and store the accumulating sequence and structure data into the first biological databases. Nowadays, biological databases represent archives of life sciences information from research areas including genomics, proteomics, metabolomics, microarray gene expression, etc. These databases are also an essential tool in helping scientists to understand and explain biological phenomena. The collected knowledge can assist the fight against diseases, the development of medications and discovering basic relationships amongst species.

Biological knowledge is scattered among different general and specialised databases. The stored information is generally acquired from scientific experiments, published literature and computational analyses. In the plant research field there are several databases with publicly available experimental data. Most of the experimental data are obtained for the model plant species *Arabidopsis thaliana*, but data for some other plants, such as rice, grape, potato, corn,

etc. are also available. UniProtKB (Boutet et al., 2007) is a general protein knowledgebase, containing information on many plants, like *Arabidopsis thaliana*, potato (*Solanum tuberosum*), tobacco (*Nicotiana tabacum*), tomato (*Solanum lycopersicum*), etc. AtPID (Arabidopsis thaliana Protein Interactome Database) (Cui et al., 2008) archives protein-protein interaction networks, domain architecture, orthologous information and gene ontology annotation. Arabidopsis Reactome (Tsesmetzis et al., 2008) represents a foundational knowledge base for systems biology containing data for Arabidopsis, rice, corn, etc. PlantCyc (Zhang et al., 2010) is plant metabolic pathway database containing curated information on primary and secondary metabolism in plants. KEGG (Kyoto Encyclopedia of Genes and Genomes) (Kanehisa and Goto, 2000) is a source of metabolic pathway knowledge. BioCyc (Krummenacker et al., 2005) is a set of 2,038 Pathway/Genome Databases where each of them describes the genome and metabolic pathways of a single organism. One of the Genome Databases that BioCyc incorporates is AraCyc (Mueller et al., 2003), pathway base of *Arabidopsis thaliana*. The Arabidopsis Information Resource (TAIR) (Swarbreck et al., 2008) maintains literature, protein information, genetic and molecular biology data for the model plant *Arabidopsis thaliana*.

Currently, the biggest databases of publicly available texts in biology and systems biology are PubMed and its subset PubMed Central (PMC), which contains publicly available full texts. PMC is a freely accessible online database (archive) of biomedical and life sciences literature which has been developed and managed by National Library of Medicine's National Center for Biotechnology Information (NCBI). It currently hosts more than 2.7 million articles for which full text is available, either as HTML/XML or pdf or both.

There are also many computer analysis tools that integrate the current knowledge and allow efficient exploration of experimental data. VirtualPlant (Katari et al., 2010) and On-dex (Köhler et al., 2006) provide graph-based integrations of knowledge and gene functional inferences that may be queried and filtered. Biomine (Eronen and Toivonen, 2012) is a system that puts together cross-references from several biological databases into a graph model with multiple types of edges, such as protein interactions, gene-disease associations and gene ontology annotations. Integration of data from databases into a single repository can aid the discovery of previously unknown connections spanning multiple types of relationships and databases, which can be used for model construction purposes. Furthermore, text mining tools, such as iHOP (Hoffmann and Valencia, 2004), Suiseki (Blaschke and Valencia, 2002), BioRAT (Corney et al., 2004), PLAN2L (Krallinger et al., 2009), etc. can help in the organisation of the current knowledge and discovery of new relations between biological molecules. In our work we have explored mostly KEGG, AraCyc, TAIR, PubMed and PubMed Central databases to obtain relevant information. To construct the model structure manually, KEGG and AraCyc databases along with the biological literature and iHOP text mining tool were considered while for the automated relation extraction from literature full-texts from PubMed Central were used.

### 1.2.2 Modelling of biological mechanisms in plants

The database repositories of biological data have accelerated the development of the first mathematical models of biological mechanisms and their simulations. These simulation models have enabled an initial insight into the cell dynamics and into the understanding of the organism behaviour as a unique system. However, considering numerous living species and their complexity, there are still many gaps in the experimental data obtained so far.

Mathematical models require a minimum of numerical information acquired from various transcriptomic, genomic and regulomic tools. In plant research, there are numerous studies related to plant metabolic pathways (Giersch et al., 1991; Hahn, 1987; Laisk and Walker, 1989; Poolman et al., 2001). Most studies, however, were related to photosynthesis and the pathways of central carbon metabolism, for which the majority of complex kinetic models

were developed (Rios-Esteva and Lange, 2007). The model of the Calvin cycle was one of the first models of plant metabolism constructed, possibly the most important metabolic pathway in plants (Pettersson and Ryde-Pettersson, 1988). For a summary of kinetic models applied to various aspects of plant metabolism, the reader is referred to the review study of (Morgan and Rhodes, 2002).

Signalling networks, contrary to the metabolic ones, have another dimension of the complexity. There is experimental evidence of the cross talk phenomenon where several signalling pathways communicate and influence each other's performance, which opens a new aspect in the research of signalling networks (Genoud and Métraux, 1999; Noselli and Perrimon, 2000). In a recent study, a signalling network of auxin-transport was developed using three different approaches: analytical methods, deterministic numerical simulations and stochastic numerical simulations (Twycross et al., 2010). This auxin-transport model was rather simple, including in total 6 biological components of interest.

In plant-pathogen interaction research, it is particularly hard to obtain data due to a long duration of experiments. Most of the plant-pathogen interaction studies are focused on individual interactions or subsets of the whole PDS mechanism. The models that are commonly used are static structural models with no information on their dynamics (Olmedo et al., 2006; Staswick, 2008).

The first attempt to model the PDS by constructing a Boolean network and carrying out numerical simulations of the PDS 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).

A similar work of using Boolean formalism to model only the JA signalling network as a response to different pathogens was done by (Devoto and Turner, 2005). This study introduced more biological components into the JA network than the previous one by (Genoud et al., 2001). In total there were 28 biological nodes included. However, not all nodes were representing biological components, as some were more abstract terms, for example "defence", "growth inhibition", "hook formation", etc. Reactions were represented with Boolean NOT and AND relations (18 in total). Nevertheless, this study was focused only on the JA pathway.

A recent study that merged all known data concerning protein-protein interactions in *Arabidopsis thaliana* was the Arabidopsis interactome study (Consortium, 2011). In this work, the nature of protein-protein interactions is not clearly stated, but is defined as a general interaction. When two proteins physically interact, they most probably influence each other in terms of activation or inhibition. We consulted Arabidopsis interactome to inspect the existence of the interactions between the molecules, but we could not obtain information on the reaction type, which is needed for the PDS model.

In the work of (Naseem et al., 2012), different pathways involved in the PDS mechanism (such as SA, JA, ET, auxin, gibberellic acid (GA), abscisic acid (ABA), etc.) are taken into account to establish the plant immune defence network. The resulting network contains 105 nodes and 163 relations and simulates various aspects of plant immunity. However, this network does not contain sufficiently detailed information on the particular pathways of interest in this dissertation: the SA, JA and ET pathways.

### 1.2.3 Text processing in systems biology

Scientific literature can be inspected manually or analysed by text processing and information extraction tools. There are numerous biological models which were manually constructed based on an in-depth literature survey, such as the macrophage activation model developed by (Raza et al., 2008, 2010). To construct the model structure, different information sources can be used (Subsection 1.2.1). Given that most of human biological

knowledge is still stored only in the silos of biological literature, retrieving information from the literature is necessary when building and curating the biological models.

State-of-the-art technologies enable information extraction from scientific texts in an automated way by means of text processing techniques, based on the advances in the area of natural language processing (NLP) of biology texts (see e.g., the research advances of the emerging bioNLP community<sup>1</sup>). The most common NLP approaches can be classified into three categories (Cohen and Hunter, 2008): rule-based approaches, machine-learning approaches and co-occurrence-based approaches. Examples of rule-based systems include GeneWays (Rzhetsky et al., 2004), Chilibot (Chen and Sharp, 2004), PLAN2L (Krallinger et al., 2009) and the approach proposed by (Ono et al., 2001). Combined methods, including co-occurrence-based approaches, such as Suiseki developed by (Blaschke and Valencia, 2002) and upgraded in the BioRAT system by (Corney et al., 2004), are also used in systems biology. In most systems, the information is retrieved only from abstracts of the PubMed database; an exception is the BioRAT system which can process full texts (Corney et al., 2004). A wide range of machine learning techniques is used for relations extraction in systems biology, like the Naive Bayes classifier (Craven and Kumlien, 1999), Support Vector Machines (Donaldson et al., 2003), clustering (Hasegawa et al., 2004), etc.

NLP methods are used in systems biology to generate the model structures or to enhance the existing ones. However, these text processing algorithms aim to be research assistant tools to human experts, but they can never replace human expertise. We have also used a rule-based NLP approach in our work and developed a Bio3graph tool, which was used to enhance the manually constructed PDS model structure.

#### 1.2.4 Optimisation methods in systems biology

Optimisation, in general, aims to find an optimal solution for a given problem with respect to some criteria. Formally, an optimisation problem can be defined as a task that requires optimising the objective function (also named criteria, cost, utility or fitness function). Many different problems from bioinformatics and systems biology can be defined as optimisation problems (e.g., reverse engineering of gene networks, prediction of 3D protein structure, multi parameter estimation, multi-alignment problem, etc.). The selection of the optimisation method influences remarkably the problem solution.

Optimisation methods applied in bioinformatics can be classified, according to the way solutions are found, as deterministic or stochastic. A deterministic method converges to the same solution for a given set of input parameters. A stochastic method, on the other hand, uses a random component that may result in different solutions when the same input parameters are provided. In the optimisation of parameters of biological pathways, due to small amounts of existing quantitative data, different deterministic and stochastic optimisation methods were recently employed. Various local deterministic optimisation techniques, like Levenberg-Marquardt algorithm (Marquardt, 1963; Levenberg, 1944) or Sequential Quadratic Programming (Boggs and Tolle, 1995), are applied in systems biology. For the parameter optimisation of biological models Simulated Annealing (Kirkpatrick et al., 1983), Evolutionary Algorithms (EAs) (Eiben and Smith, 2003) and Genetic Algorithms, as one type of EAs, (Mitchell, 1996) are commonly used from stochastic approaches.

EAs are among the most popular optimisation methods. These algorithms start iteration from a random population of points and perform optimisation search until some pre-defined criterion is satisfied. A subset of points is selected in every iteration, which is usually called generation. By applying some variation operators to the selected set, a new population of points is created. Several types of EAs are exploited in systems biology, such as genetic algorithms, evolutionary programming, differential evolution, etc. We have used in our

---

<sup>1</sup><http://www.bionlp.org/>

work differential evolution (DE) algorithm (Storn and Price, 1995) to optimise the model parameters.

### 1.3 Hypothesis and goals

The mechanisms triggering illnesses in host plants by specific pathogens are not fully understood for several reasons. First, the plant defence system is a highly complex network, consisting of several interconnected pathways and only small subsets of the involved mechanisms have been studied. Second, there is currently a lack of kinetic biological studies that would potentially reveal new knowledge regarding the dynamic behaviour of the system. As a result, the availability of relevant experimental data is currently very limited. However, despite the lack of sufficient data there is a strong motivation to establish a detailed PDS model that would facilitate a better understanding of the mechanisms underlying plant defence. To address this problem, we have developed a novel methodology and tools that allow the construction of PDS model using a semi-automatic knowledge elicitation approach. The overall aim of this thesis is as follows.

1. to develop a novel methodology and tools that enable the construction of models with limited experimental data, and
2. to develop a concrete PDS model with the newly developed methodologies and tools.

To achieve the main aims of this thesis, we have defined a set of specific objectives:

1. to construct the PDS biochemical network structure in two ways:
  - from scratch, without a starting network structure, and
  - from structural models published in the literature using a semi-automated information extraction approach based on natural language processing.
2. to construct a dynamic model of PDS system where we estimate concrete model parameter values using a constraint-based combinatorial parameter optimisation approach,
3. to develop a flexible methodology that facilitates the construction of biological models in various settings; this methodology consists of methods for:
  - semi-automated model structure extraction from biological literature,
  - incremental model structure revision from scientific literature, and
  - constraint-driven model parameter optimisation together with the model evaluation based on limited amount of experimental data.
4. to develop a flexible software tool that implements the methods of the new methodology.

### 1.4 Scientific contributions

This thesis contributes to the fields of systems biology and computer science. The main contributions are as follows.

1. Manually constructed model structure of PDS presented in Section 4.2 and its semi-automatically constructed extension, presented in Section 6.6.
2. Dynamic model of the SA pathway, presented in Subsection 5.2.4.



3. A novel methodology for semi-automated construction of biological models (presented in Chapter 3) consisting of:
  - a method for semi-automated model structure revision from biological literature, presented in Section 6.1,
  - an incremental method for updating the model structure from new biological literature, presented in Section 7.1, and
  - a method for constraint-driven model parameter optimisation together with the model evaluation based on few experimental data, presented in Chapter 5.
4. An executable workflow implementing the proposed PDS modelling methodology, presented in Section 6.2 and upgraded in Section 7.2.
5. Publicly available tools, models and results:
  - The Bio3graph tool, developed to extract biological relations from the literature to enrich the model structure, is publicly available at [http://kt.ijs.si/dragana\\_miljkovic/](http://kt.ijs.si/dragana_miljkovic/), together with the Supplementary material for manual model development and for automated model structure enhancement with Bio3graph.
  - The dynamic model of the SA pathway, including the results of model parameter optimisation, the code of the DE algorithm (developed by Depolli M.) and the user manual are available publicly at [http://kt.ijs.si/dragana\\_miljkovic/](http://kt.ijs.si/dragana_miljkovic/).
  - Model validation materials, including the validation results and the code in MATLAB are accessible at [http://kt.ijs.si/dragana\\_miljkovic/](http://kt.ijs.si/dragana_miljkovic/).

The main publications presenting the results of this thesis are listed below.

Miljkovic, D.; Stare, T.; Mozetič, I.; Podpečan, V.; Petek, M.; Witek, K.; Dermastia, M.; Lavrač, N.; Gruden, K. Signalling network construction for modelling plant defence response. *PLOS ONE* **7**, e51822-1e51822-18 (2012).

Miljkovic, D.; Podpečan, V.; Stare, T.; Mozetič, I.; Gruden, K.; Lavrač, N. Incremental construction of biological networks by relation extraction from literature. *Current Bioinformatics*. (In press).

Miljkovic, D.; Depolli, M.; Stare, T.; Mozetič, I.; Petek, M.; Gruden, K.; Lavrač, N. Plant defence model revisions through iterative minimization of constraint violations. *International Journal of Computational Biology and Drug Design*. (In press).

Miljkovic, D.; Podpečan, V.; Grčar, M.; Gruden, K.; Stare, T.; Petek, M.; Mozetič, I.; Lavrač, N. Modelling a biological system: network creation by triplet extraction from biological literature. In: Berthold, M. R. (ed.) *Bisociative Knowledge Discovery: An Introduction to Concept, Algorithms, Tools, and Applications*. 427-437 (Springer, Berlin, 2012).

Miljkovic, D.; Podpečan, V.; Stare, T.; Mozetič, I.; Gruden, K.; Lavrač, N. Incremental revision of biological networks from texts. In: Ortuno, F.; Rojas, I. (eds.) *Proceedings of the International Work-Conference on Bioinformatics and Biomedical Engineering (IWBBIO)*. 1-9 (Granada, Spain, 2013).

Podpečan, V.; Miljkovic, D.; Petek, M.; Stare, T.; Gruden, K.; Mozetič, I.; Lavrač, N. Integrating semantic transcriptomic data analysis and knowledge extraction from biological literature. *Proceedings of the International Conference on Bioinformatics and Biomedicine (BIBM)*. (Shanghai, China, 2013). (In press).

Miljkovic, D.; Depolli, M.; Mozetič, I.; Lavrač, N.; Stare, T.; Petek, M.; Gruden, K. Constraint-driven optimization of plant defense model parameters. In: Gao, J. (ed.) *Proceedings of the Third Workshop on Integrative Data Analysis in Systems Biology (IDASB)*. 570-574 (Danvers: Institute of Electrical and Electronics Engineers, 2012).

Miljkovic, D.; Mihăilă, C.; Podpečan, V.; Grčar, M.; Gruden, K.; Stare, T.; Lavrač, N. Workflow-based information retrieval to model plant defence response to pathogen attacks. In: Hilario, M.; Lavrač, N.; Podpečan, V.; Kok, J. N. (eds.) *Proceedings of the workshop on Third generation data mining: Towards service-oriented knowledge discovery (SoKD)*. 51-60 (Barcelona, Spain, 2010).

The rest of the published material related to the topic of this dissertation is listed at the end of the dissertation.

## 1.5 Organisation of the thesis

The rest of the thesis is organised as follows.

Chapter 2 is an introduction to the general concepts of modelling and describes the most common modelling formalisms applied in systems biology.

Chapter 3 provides a top-level overview and explanation of the methodology that we have developed to construct the PDS model. It also contains brief descriptions of each step taken during the PDS model construction process with an emphasis on the steps that are introduced as the scientific contributions of this thesis with the goal to improve and accelerate PDS model construction.

Chapter 4 presents the manually developed PDS model structure. It includes the definition of the requirements for PDS model construction and the manually developed structures of the three separate sub-models of the SA, JA and ET pathways.

Chapter 5 describes the model dynamics. It includes the definition of constraints and the iterative process of using combinatorial optimisation for model construction, its evaluation and refinement. Moreover, it presents the model validation process and the proposed evaluation method by comparing simulation results with a publicly available experimental dataset. It also discusses and interprets the results of the proposed methodology.

In Chapter 6 we present the relation extraction approach and point out its perspectives for detecting scientific expertise in a specific domain. The chapter focuses on the techniques for refining the model structure by triplet extraction from the literature and presents a revised PDS model structure.

Chapter 7 describes the incremented version of the method presented in Chapter 6. In this chapter, we present the method for incremental structure revision from the biological texts. The method is demonstrated on two use cases: a simple network and a complex PDS network.

Finally, Chapter 8 concludes the dissertation, summarises the presented work and outlines possible directions for further research.

## 2 Modelling of Dynamic Systems: Background and Related Work

*All models are approximations. Essentially,  
all models are wrong, but some are useful.*

George Edward Pelham Box

Computer modelling and simulation is a research area concerned with the design and analysis of complex systems. The overall goal of computer simulation is to represent the dynamics of a real world system in a computer model. The main topic addressed in this chapter are different modelling formalisms and the effects of the formalism selection on the simulation results. In addition, text mining approaches dealing with automated extraction of model structure are discussed.

This chapter is organised as follows. We start with the definitions of the main concepts in dynamic modelling and explain briefly the main model categories. Next, we focus on the model construction process and explain each phase. We give the main emphasis on the phase of modelling formalism selection and we detail the topic by discussing different representations of an illustrative example: a thermostat model. We provide an overview of different modelling formalisms that were applied in systems biology. We conclude this chapter by discussing the text mining approaches which enable model structure extraction and visualisation.

### 2.1 Modelling approaches

Modelling is the process of constructing a model, which is basically a representation of some system of interest. Modelling is a way of simplifying the real world that enables us to solve problems. Essentially, we do this every day without noticing. One example of such a simple model is a street map. It represents a part of the earth surface, consists of different materials, contains only streets with their names and represents earth as a flat surface. Thus, a street map is a simplified abstract representation of the earth surface used to solve problems, i.e., to answer questions, such as: "What is the shortest route?". There are also many other models that we use in daily life. For example, a diagram is a model of how something is made, a calendar is a model of a month, computer games are models of reality, etc.

#### 2.1.1 Concepts and classification

Prior to the study of modelling and simulation it is necessary to define the basic terms: system, model, experiment and simulation (Fritzson, 2004). A *system* is a real-world entity, a phenomenon or process whose properties we want to study. A *model* is a simplified, structured representation of the real-world system built to enhance our understanding of it. Models are developed to solve specific problems and to answer questions about a system without doing experiments on the real system. An *experiment* is the process of extracting information from a real-world system by changing its inputs. A *simulation* is the manipulation of the model that produces specific outcome. We can also say that it is an experiment

performed on a model that provides answers to the questions about the real-world system. Moreover, if we consider a model representing our hypothesis about the system, then the simulation could be defined as hypothesis testing. Simulations are often used when experiments with the real systems are time-consuming, dangerous or expensive. Both modelling and simulation can provide a better understanding of the dynamic behaviour of real systems. Basically, model development can be motivated from two different perspectives:

- solving a practical problem, which is mostly the goal of engineering, and
- better understanding of the system's nature, which is mostly the viewpoint of natural sciences (Fritzson, 2004).

To enrich the knowledge about different real-world systems, various modelling approaches were developed. Depending on the model representation form, there are many modelling types, such as: physical, logic and mathematical. *Physical modelling* is a way of making a smaller or larger copy of the actual object. It is often used in civil engineering. For example, a breadboard construction of a building represents a small copy of it. *Logic modelling*, used in economy and management, is a way of making a flow form as a linear sequence: inputs, activities, outputs, and outcomes (Alter and Murty, 1997). *A mathematical model* is a description of the system by using different mathematical formalisms (Dym, 2004). In mathematical formalism, statements of mathematics and logic can be thought of as statements about the consequences of certain string manipulation rules. According to the formalism, the truths expressed in logic and mathematics are syntactic forms that have no meaning unless they are given an interpretation. Mathematical modelling is becoming increasingly important since the computers enable us to give real world interpretation to the results of mathematical calculations. Mathematical models are usually composed of variables, which are abstractions of quantities of interest in the described systems, and operators that act on these variables. The operators can be algebraic, functions, differential operators, etc. Mathematical models can be classified according to different criteria, such as (Kapur, 2008):

- The scientific field for which models are built. For example, mathematical models in physics, biology, chemistry, economics, etc.
- The mathematical techniques that were used. We have mathematical models developed by means of algebra and linear matrices, differential equations, partial differential equations, etc.
- The purpose of the model. Mathematical models exist for simulation, prediction, optimisation, diagnosis, control, etc.
- The nature of the model. The models can be:
  - *linear*, where a dependent variable  $Y$  can be described as a linear function of an independent variable  $X$  in a form  $Y = \alpha * X + \beta$  ( $\alpha$  and  $\beta$  are model parameters), and *nonlinear* that are opposite to linear,
  - *static*, where the time is not taken into account, and *dynamic*, in which the time-varying interactions among variables are included,
  - *deterministic*, where the input and output variables are uniquely determined by the model parameters, and *stochastic*, where at least one of the input or output variables is probabilistic,
  - *discrete*, where the variable values are discrete (integer number) or *continuous*, where variable values are real numbers,

- *mechanistic*, where the structure of the model is determined by prior knowledge and parameters from data sets, and *empirical* (or data-based), where both the model structure and parameters are estimated from a large experimental data set (Thakur, 1991), and
- *qualitative*, where the system is described in a more abstract, descriptive way, without precise numerical information that is often not needed, and *quantitative*, which are developed when the numerical information about the system is available and they represent the system in more detail.

These categories do not exclude each other, since many models include elements of several types. Regardless of the type of the model we want to develop, there are several major phases in the model construction that are common. The next section discusses these phases in particular for mathematical model construction.

### 2.1.2 Construction of mathematical models

The phases in mathematical model construction can be defined as follows (see Figure 1).

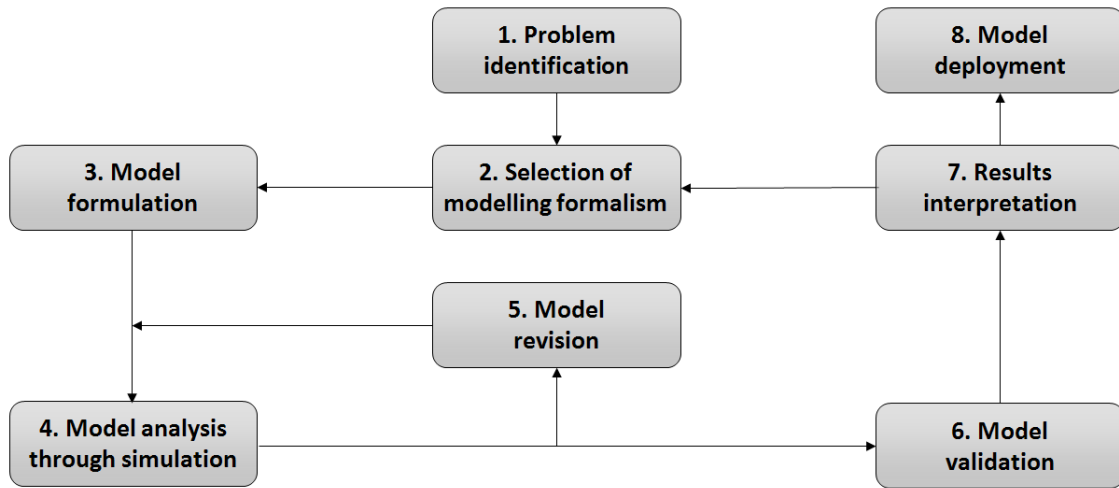


Figure 1: Phases in model construction.

- 1. Problem identification.** The situation has to be studied thoroughly in order to identify the problem precisely and understand its fundamental questions. Furthermore, the requirements that the model has to satisfy need to be defined and the hypothesis that will be tested has to be formulated. At this stage, the problem's classification should be decided, for example, whether it is deterministic or stochastic. Only with a clear, precise identification can the problem be translated into mathematical symbols and solved. Usually, this phase in model construction is accomplished in collaboration with domain experts.
- 2. Selection of modelling formalism.** The selection of the type of the modelling formalism depends on two main factors (Cantone, 2009):
  - the information that we already know about the system, and
  - the information that we want to know about the system.

The answers to the questions stated above enable us to define the abstraction level of the model prior to the selection of the particular modelling formalism. Depending on the abstraction level of the model representing the system, there are two main categories of modelling approaches, i.e., qualitative and quantitative modelling approaches, discussed in more detail in Subsection 2.1.3.

3. **Model formulation.** Assuming that all the relevant data have already been collected to gain the information on system's behaviour, at this stage the variables and their relationships have to be established. Depending on the available data sources and tools, model construction can be manual or automatic. Manual model development is time-consuming and requires a domain expert. However, the absence of numerical data requires manual model construction and curation. This way of modelling is error-prone when building complex models and is therefore more effective when combined with automatised approaches. On the other hand, most of the automatically constructed models are data-based, assuming sufficient experimental datasets available to create and validate the model. Specially for the ODE formalism, there are many tools that can automatically discover equations from the experimental data (Džeroski and Todorovski, 2008).
4. **Model analysis through simulation.** One of the most powerful analysis approaches of complex systems is model simulation. Simulation of a model is the process of acquiring information on how several model variables behave without testing the system in the real life. There are different types of simulation. In physical simulation, the actual objects are replaced with the smaller physical ones. Interactive simulation represents a special subset of physical simulation where human operators are included in the simulation loop. Simulations with humans in the loop also include computer simulations which are the most common simulation type (McHaney, 2009). Computer simulation is a conventional part of modelling in many scientific fields, such as biology, medicine, physics, chemistry, etc.
5. **Model revision.** The process of revising a model can be grouped into two major stages:
  - revision of the model structure, and
  - revision of the model dynamics.

Model structure can be revised either manually by investigating a broader literature or automatically by using different text processing methods for information extraction. Similarly, model dynamics is usually revised either through the iterative process of hand-tuning of model parameters or model parameters are calculated automatically from the experimental datasets until they satisfy certain accuracy criteria. These two stages of model revision are not totally independent since changes in the model structure influence directly the dynamic behaviour of the model.

6. **Model validation.** After the development of a suitable model for the defined problem and its simulation, it is necessary to validate the model. The model validation is performed ideally by comparing the simulation output with the real system data. There are many statistical approaches developed, such as t-tests, etc. (Kleijnen, 1995). The problem is, however, the comparison with the experimental data when they are sparse and few.
7. **Results interpretation.** The simulation results are interpreted by the domain experts.
8. **Model deployment.** The deployment of the model usually denotes the application of the model for prediction when using new data. Even when the purpose of the model is to gain the knowledge on certain phenomena, the knowledge gained has to be organised and presented in such a way that the end-users can easily use it. When the model is developed with the prediction purposes, it is also important for the end-user to become familiar with the basic steps undertaken during the model development in order to use it with understanding.

### 2.1.3 Selected modelling approaches

Below we list the most commonly used modelling approaches.

#### Qualitative modelling approaches

These approaches are usually applied in the case of missing accurate numerical information, when the dynamics of the system still needs to be modelled. Such models give a more abstract view of the studied system. Brief descriptions of some typical qualitative modelling approaches are listed below.

- **Boolean networks** are a graphical and mathematical formalism. They consist of Boolean variables connected to each other via logical operators (Ibe, 2011).
- **Process algebra** (or process calculi) was developed to support the modelling of concurrent systems. The emphasis in this approach is on modelling of communication between processes. For a more detailed introduction into process algebra, the reader is referred to (Fokink, 2007).
- **Hybrid automata** are used to model systems that have both discrete and analog components. Systems with 'discrete jumps' and 'continuous flow' can be modelled using hybrid automata. The drawback of this formalism is that it is not suitable for complex systems (Henzinger, 1996).
- **A statechart** is a visual language for describing the behaviour of the system. This behaviour is usually composed of a series of events that might happen in different possible states (Harel, 1987). This formalism is also not suitable for modelling of complex systems.
- **Petri nets (PNs)** is a graphical and mathematical formalism (Murata, 1989). Similarly to the process algebra, it was initially developed to model concurrent systems. There are several levels of the PN formalisms. While the standard PNs are discrete and qualitative, their various extensions allow the definition of both qualitative and quantitative models.

#### Quantitative modelling approaches

In contrast to qualitative modelling, these approaches require precise numerical data and represent the system at a very detailed level. Some of the commonly used quantitative modelling approaches are described below.

- **Ordinary differential equations (ODEs)** present the relations between one independent and one or more dependent variables and their derivatives. They are commonly used to model dynamic behaviour of the system (Coddington and Levinson, 1955).
- **Stochastic differential equations** are used if there is a need to model the stochastic effects of the system (Arnold, 2013).
- **Partial differential equations** present the relations between two or more independent and one or more dependent variables and their partial derivatives. They are applied when the solution depends on more than one independent variable (Renardy and Rogers, 2004).

All the listed modelling approaches have specific advantages and drawbacks. We focus our discussion on PNs, which is the formalism selected in our work, and their comparison with ODEs, which is the underlying modelling approach that is also implemented in one of the PN extensions. For example, PN models are simple and yield a broad overview of complex systems and offer an intuitive graphical representation. They do not require kinetic data, they have well-founded mathematical theory and many software tools are available. On the other hand, ODEs have higher-level modelling tools available and they take a population view of the system, rather than modelling the stochastic behaviour of system's individual components (Pettinen et al., 2005). The drawbacks of PNs are that many tools are no longer supported and the file formats cannot be automatically converted from one type of PN into another. In contrast to PN, ODEs require the exact knowledge of system dynamics and they are not visually intuitive. For a better understanding of the terminology used to explain PN and ODE, the mapping of concepts is provided in Table 2.

Table 2: The mapping between the PN and ODE terminology.

PN terminology	ODE terminology
Place	Variable
Transition	Equation
Initial marking	Initial values

#### 2.1.4 Illustrative example

In this section, the main representative formalisms, ODEs from the quantitative category, and PNs from the qualitative, are elaborated on a simple example from everyday life: a thermostat (Koutsoukos et al., 2000). A thermostat is a control unit that regulates the temperature in the room and keeps it within a certain temperature range. A thermostat is basically a simple control system with the negative feedback loop. The negative feedback loop is active when the system's output influences the input changes in such a way that it diminishes these changes. This basic control mechanism is also present in many other fields, such as economy, biology, industrial systems, etc.

There are two basic behaviours that the thermostat exhibits:

- When the room temperature drops below the lower limit, the thermostat switches on the furnace and the room temperature rises.
- When the room temperature exceeds the upper limit, the furnace is turned off and the room temperature drops.

Let us consider a simpler version of the thermostat model in which the upper and the lower temperature threshold have the same value  $T_{thr}$  (see Figure 2).

Let us first illustrate how to represent the room temperature without introducing the negative feedback loop, which denotes the *on* and *off* switching actions of the thermostat. This means that we consider only the case when the thermostat is already *on* or *off*, but we do not take into account the transitions between the *on* and *off* states. If there is a need to deal with the precise numerical information regarding the thermostat control, then we should choose the quantitative modelling approach, for example, ODE. One version of such a model is expressed by the following equations which represent the change in room temperature based on the rate at which heat is dissipating through the room walls to the outside, and the rate at which the room is heated by a heater (Koutsoukos et al., 2000).

$$\frac{dx}{dt} = a * (T_0 - x) + b * v \quad (1)$$



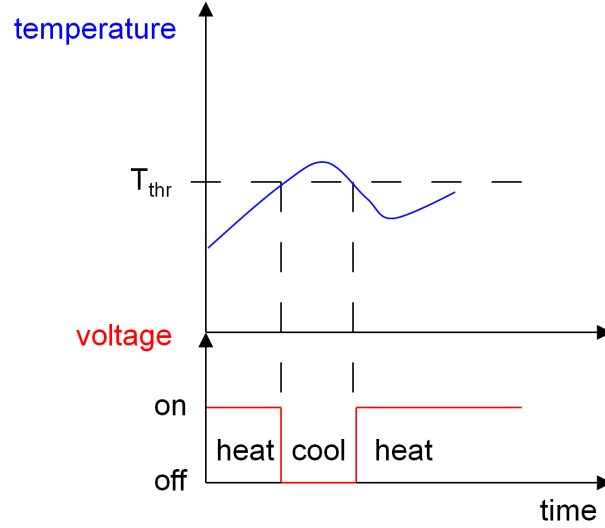


Figure 2: Dynamics of a simplified thermostat behaviour.

where

- $x$  denotes room temperature,
- $dx/dt$  denotes the rate of change of the room temperature,
- $T_0$  denotes outdoor temperature,
- $v$  denotes the voltage on the furnace control unit, and
- $a$  and  $b$  denote positive constants.

In Equation 1 the term  $a * (T_0 - x)$  describes the effect of the heat exchange between the room and the outside environment. If the room temperature is higher than the outdoor temperature, i.e.  $T_0 - x < 0$ , heat from the room will dissipate through the room wall to the outside. If the outside temperature is higher than the room temperature,  $T_0 - x > 0$ , heat from the outside is transferred into the room. The efficiency of the heat transfer through the wall is governed by constant  $a$ . High values of  $a$  mean that the insulation is very poor and heat is transferred through the wall at a high rate. Low values of  $a$  indicate good insulation. Because we normally want to heat a room when the outside is colder than the inside, we refer to the term  $a * (T_0 - x)$  as *cooling rate*. Term  $b * v$  in Equation 1 represents the *heating rate*. The higher the heating rate, the more heat is generated by the furnace and flows into the room. Constant  $b$  controls the sensitivity of the heating rate to the signal,  $v$ , from the control unit.

We now assume that the insulation of the room walls is extremely good. This means that the value of constant  $a$  is very small and hence the entire cooling rate is very small in relation to the heating rate. In this case we can approximate Equation 1 by the equation below:

$$dx/dt = b * v \quad (2)$$

So with Equation 1, we can express the *on* and *off* scenarios as follows:

- $dx/dt = b * v$ , if the thermostat or heater is *on*,
- $dx/dt = 0$ , if the thermostat or heater is *off*.

In case we do not have or we do not need the precise numerical data regarding the temperature and the furnace voltage, we should consider a qualitative approach. For example, instead of representing the temperature in terms of precise numbers, such as  $22.1 \text{ }^\circ\text{C}$ , we could have statements: too cold, normal and too warm. For the most of daily needs such statements are sufficient. These statements are one way of making abstractions from the numerical data.

Humans tend to think in qualitative ways since dealing with qualitative descriptions is more natural than dealing with numbers. Initially, the PN formalism was developed for the needs of qualitative modelling. Afterwards, various extensions were developed to bridge the gap between the qualitative and quantitative models, and some extensions can be used for the development of quantitative models. Further discussion addresses the different levels of PN formalisms, elaborated on the thermostat example (Equation 2). The differences between different modelling approaches are shown by simulating the dynamic behaviour of the simplified thermostat model.

A PN is a bipartite graph with two types of nodes: places and transitions (Figure 3). Places represent the resources of the system and transitions represent the events that transform the system from one state to another. Weighted arcs connect places with transitions, thus defining the relations between the resources and the events. At any time of the simulation of a PN, the places have zero or more tokens. The state of the system is represented by the distribution of tokens over the places and is called a marking. The definition of a PN includes the specification of the initial marking, which assigns a number of tokens to each place. In Figure 3, the initial marking is  $\{1,0\}$ , which corresponds to the number of tokens at places {place 1, place 2}. A transition is enabled if its input places contain at least the required number of tokens (defined by the weight assigned to the arc). Obviously, the transition in Figure 3 is not enabled, because the arc weight requires 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 the tokens in its output places (this number is determined by the weights of the outgoing arcs from the transition). This 'token game' represents the dynamic changes of the system.

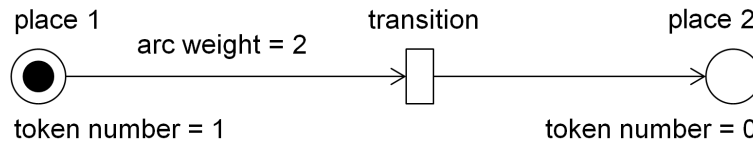


Figure 3: The graphical representation of the basic PN.

The thermostat can be modelled with the **basic PN formalism** in the following way. We will consider two places that we are interested in: the temperature of the room and the voltage. The furnace is described by one transition named "heating". In Figure 4.A, we see a graphical representation of the system with the initial marking, where the place "voltage" contains three and the place "temperature" zero tokens. The transition "heating" is enabled, which means that the heating is on, and it fires. After the firing of the transition "heating", the place "temperature" will have the increased number of tokens, which we interpret as an increase of the room temperature since the voltage of the furnace is on. The dynamic behaviour of the two places, "temperature" and "voltage", is shown in Figure 4.B.

**Stochastic Petri nets (SPN)** have the same graphical representation as the basic PNs (see Figure 4.C). The only difference between the basic PN and SPN is in the dynamics of transition firing: in SPNs the transitions fire with exponentially distributed time delay. This difference in firing delays is visible on the simulation curves in Figure 4.D whereas the firing time between the transitions of the basic PNs is always the same (Figure 4.B). SPNs are generally used to introduce the external noise (generated by fluctuations of the environment) or intrinsic noise (due to low molecular concentrations) in the model (Marsan, 1989). By using SPN, in our thermostat model we could take into account the influence of different external and internal sources to the fluctuations in temperature.

**Hybrid Petri nets (HPN)** contain both discrete and continuous places. The graphical representation of a HPN differs from the basic PN since HPN introduces two types of places: discrete and continuous. The discrete places can only have an integer number of tokens. Additionally, the HPN introduces continuous variables, where the number of tokens is a real

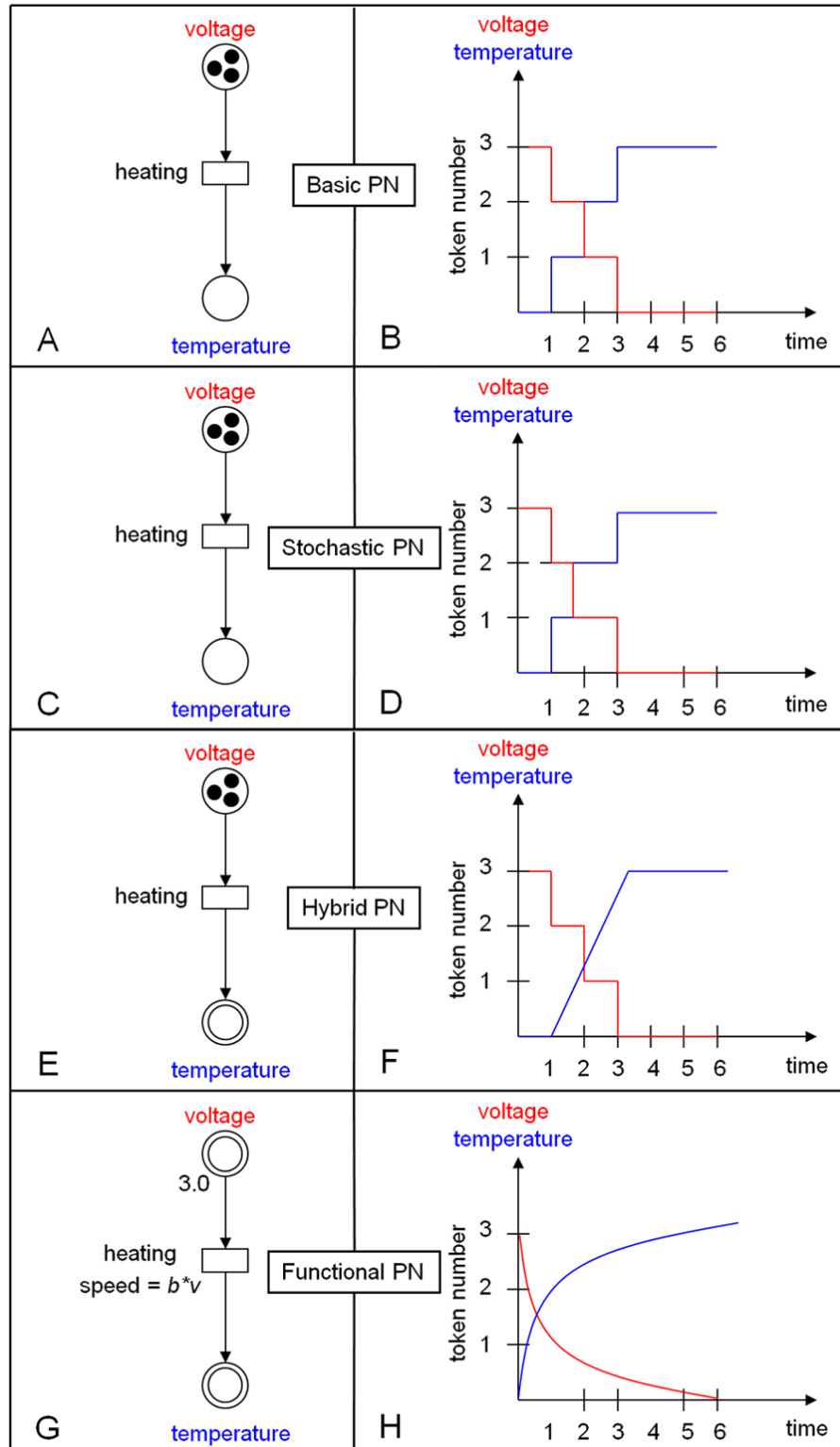


Figure 4: The simplified thermostat model represented with four different PN formalism types and the corresponding simulation curves of the temperature and voltage variables.

number. This formalism is convenient for modelling of continuous processes with the switch control (Silva and Recalde, 2004). For example, in the case of a thermostat, the voltage is represented as a discrete place with three tokens and the temperature as a continuous place (Figure 4.E). The simulation curve of the dynamic behaviour is shown in Figure 4.F.

**Functional Petri nets (FPN)** have a different graphical presentation from the basic PNs, since they contain continuous places and transitions (Figure 4.G). 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. In the thermostat example, this means that the speed of transition "heating" depends on the input place "voltage". The higher the voltage, the faster the room gets warmed up (this corresponds to the real world system). This effect is represented with the simulation results in Figure 4.H. In Figure 4.G, it can be noticed that the initial marking of the place "voltage" is now a real number (3.0).

**Hybrid Functional Petri Nets (HFPN)** represent the combination of HPN and FPN (Matsuno et al., 2003). HFPN support the modelling of continuous transitions controlled by the switch mechanisms. The transition rates, on the other hand, depend on the input values, which is often the case in real systems.

However, the simplified thermostat model (Equation 2) along with the four thermostat models represented with different PN extensions (Figures 4.A, 4.C, 4.E and 4.G) does not include the effect of the negative feedback loop. Let us now consider it. The ODE formula in Equation 1 needs to be adjusted with an additional function  $g(x)$  that represents the feedback loop as follows:

$$dx/dt = b * v * g(x) \quad (3)$$

The new term  $g(x)$  may take on different forms, depending on what kind of mechanism or behavior we want to represent. To represent a sudden *on/off* change in the room temperature rate,  $g(x)$  could be defined as step function dependent on the desired room temperature  $T_{thr}$  as follows:

$$g(x) = \begin{cases} 1, & \text{if } x \leq T_{thr} \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

To represent a continuous change of the room temperature rate, we could define  $g(x)$  as a smooth sigmoidal function:

$$g(x) = \frac{1}{1 + e^{\gamma * (x - T_{thr})}} \quad (5)$$

where  $\gamma$  is a positive constant that amplifies the difference between the room temperature  $x$  and the desired target temperature  $T_{thr}$ . Among the formalisms discussed before, two PN classes support the modelling of negative feedback loops by introducing an inhibitory arc, namely HPN and FPN. HFPN supports modelling of negative loops as well, since it integrates HPN and FPN formalisms.

The basic graphical representation of a simple system with the negative feedback loop is shown in Figures 5.A and 5.C. The inhibitory arc in both cases is directed from the place "temperature" to the heating transition. The threshold for the inhibitory arc is 1.0, which means that whenever the temperature is above 1.0, the transition "heating" will be blocked. The simulation curves for both HPN and the FPN with the negative feedback loops are shown in Figures 5.B and 5.D respectively. In these simulation curves, the impact of the negative feedback loop is clearly visible. When the temperature reaches the threshold 1.0, it stops increasing, which was not the case in the previously mentioned four examples of simulation curves in Figures 4.B, 4.D, 4.F and 4.H. Note that the FPN (and also HFPN)

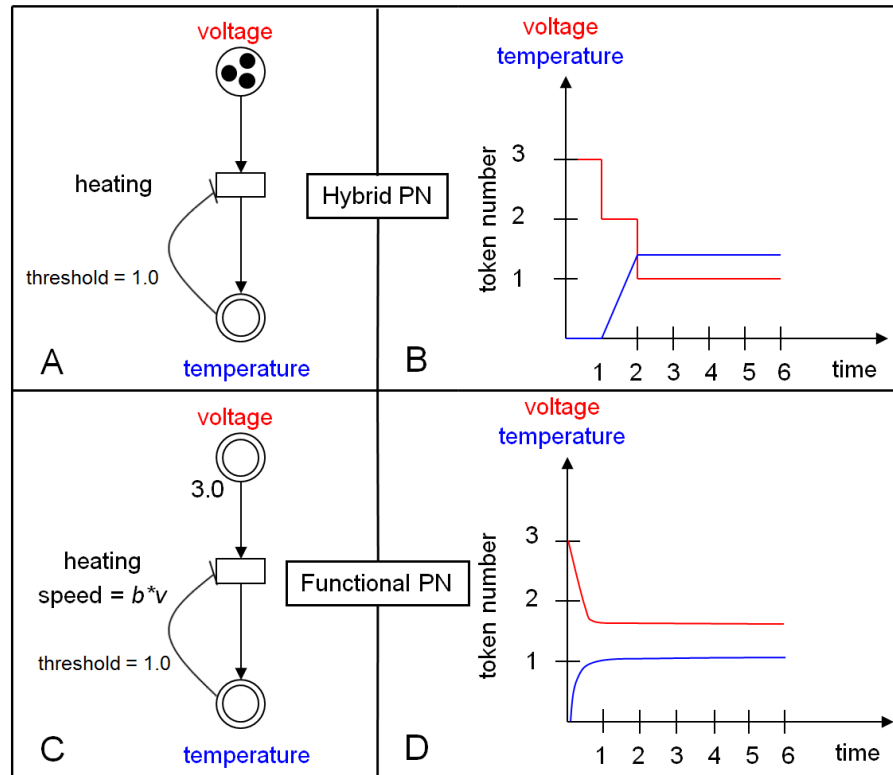


Figure 5: A graphical representation of the thermostat model with the negative feedback loop realised with the HPN and FPN modelling approach. The simulation curves are also shown.

model in Figure 5.C has exactly the same simulation output as the ODE model (Equation 2).

The thermostat model served as an example how the selection of the formalism influences the information that we can obtain from the model analysis.

## 2.2 Modelling approaches in systems biology

Many modelling approaches have been applied in systems biology. Frequently, the experiments performed in systems biology do not have the precise numerical data of the reactions kinetics. More specifically, the experimental datasets in systems biology are, on one hand, large and contain genes expression values of the whole genomes, but, on the other hand, there are only a few time points per gene.

Most commonly, when the quantitative experimental data exist, differential equations are used to build the models of biological mechanisms. In this case the calculation of equation parameters is a straightforward process. Such models are quantitative and data-based. They are used to model all three basic types of biological networks: genetic (de Hoon et al., 2003), signalling (Gong et al., 2010) and metabolic (Chassagnole et al., 2002). On the other hand, when there are few experimental data available, the qualitative modelling approach might be more appropriate.

Qualitative models of biological systems are based on graph-theoretical descriptions of the network structure. Several well-known approaches have been applied in modelling of biological systems, such as: Boolean networks, process algebra, statecharts, hybrid automata and PNs.

Boolean networks are often used to model gene regulatory networks (Albert, 2004). Other types of biological networks are also modelled with this formalism where the available experimental data, consisting of two time points, can be represented with the Boolean

values: 0 and 1 (Pinzón et al., 2009). In other cases, when there are more time points in the experimental data, the values of 0 and 1 have to be derived from the data. A widely used extension of Boolean networks combines the traditional boolean approach and a continuous system and is implemented in the SQUAD (Standardized Qualitative Dynamical Systems) software (Cara et al., 2007). Process algebras have been used to model signalling networks (Regev et al., 2001).

Statecharts have been used for modelling of gene regulatory networks (Shin and Nourani, 2010). They are used for simple network properties, but this modelling approach is not considered appropriate for complex systems.

Hybrid automata are used to model the switch control of continuous biological processes. For example, they can be used to model the system's behaviour when the regulatory gene is silenced. This means that the variable that represents the silenced gene will have a discrete value and it will be either 0 or 1. Hybrid automata are also not an appropriate formalism for the modelling of complex systems.

One of the first studies of PNs to model biological systems was by (Reddy et al., 1996). In this study the possibility of using the basic PN formalism for modelling metabolic pathways was investigated. The places in PN represent biological molecules, the transitions represent biological reactions and the tokens correspond to the presence of biological molecule in some proportion. The weights on the arcs are mapped to the number of molecules necessary to trigger the biological reaction. The same mapping is also used by all extensions of the PN formalism.

The PN formalism has numerous extensions. A detailed overview of the PN extensions and their application in systems biology is given by (Chaouiya, 2007). PN extensions enable the modelling of different characteristics of biological systems. For example, for studying the probability whether one biological reaction will occur or not, which means introducing an external or internal noise in the model, one should build the model with SPNs. A firing delay of the SPN is associated with every transition. This firing delay has an exponential distribution which may also depend on the previous marking of the system (Goss and Peccoud, 1998).

In real systems, there are many situations when a continuous flow of biological reactions is stopped and/or triggered by some inhibiting/activating molecule. To deal with continuous and discrete switch variables simultaneously, the HPNs can be used (Matsuno et al., 2000). They were used in the work of (Matsuno et al., 2000) to model the control mechanisms that are present in the gene regulatory networks. Contrary to hybrid automata, hybrid PNs are suitable for modelling of complex systems. To illustrate model complexity, Figure 6 shows one of the HPN models of early stage gene expression of bacterial virus  $\lambda$  phage from the work of (Matsuno et al., 2000). In the polygon of Figure 6 marked by a dotted line, the feedback mechanism of CI protein and Cro, shown in Figure 7, should be inserted. It is well known that the two regulatory proteins CI and Cro play significant roles in deciding the lysis (breaking down the cell) and lysogeny growth pathway (one of the methods of viral reproduction). The genes that encode CI and Cro proteins are adjacent on the  $\lambda$  phage chromosome, and each of these genes has its own promoter,  $P_{RM}$  of the gene *cI*, which codes for a protein CI, and  $P_R$  of the gene *cro* (Figure 7).

FPNs support the simulation of kinetic effects in biological networks (Hofestädt and Thelen, 1998). Furthermore, in FPNs the reaction rate depends on input concentrations, which closely matches the reality. The values of continuous variables are real numbers and represent the concentrations of the biological molecules. In practice, FPNs are equivalent to a system of differential equations. However, the main advantage of FPNs is their intuitive graphical representation.

At any level of detail, biological interactions can be modelled as networks (Alm and Arkin, 2003). For example, nodes can represent very different biological units ranging from atoms to individual organisms and the relations may describe atomic interactions in protein

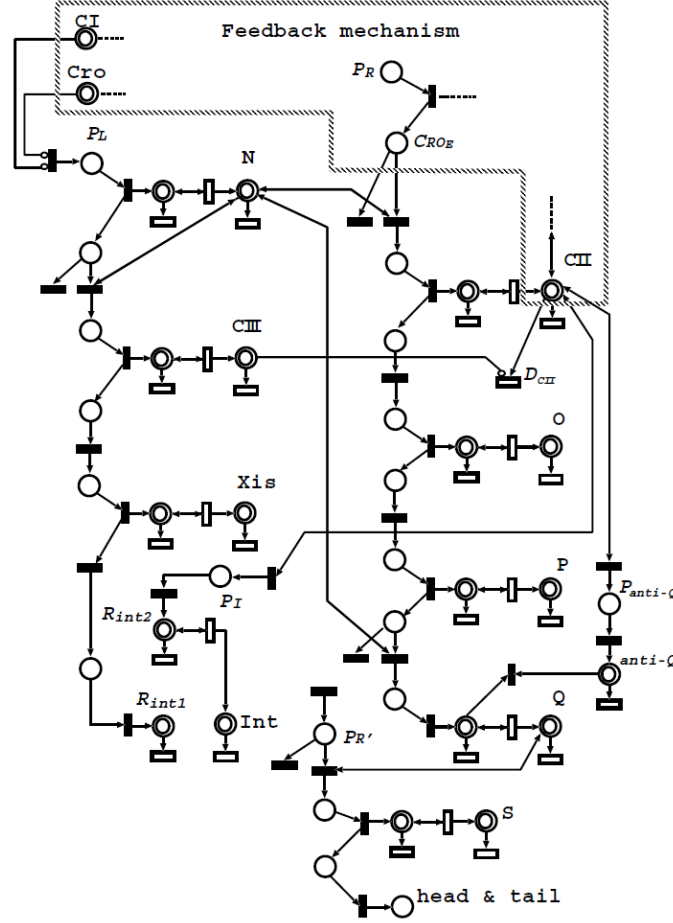


Figure 6: An example of a gene regulatory network modelled with the HPN formalism. The model represents early stage gene expression of  $\lambda$  phage (Matsuno et al., 2000). The rectangular nodes represent Petri net transitions where the black coloured are discrete and the non-coloured continuous transitions. The rounded nodes are Petri net places, where the single circle represents discrete place and the double circle continuous place.

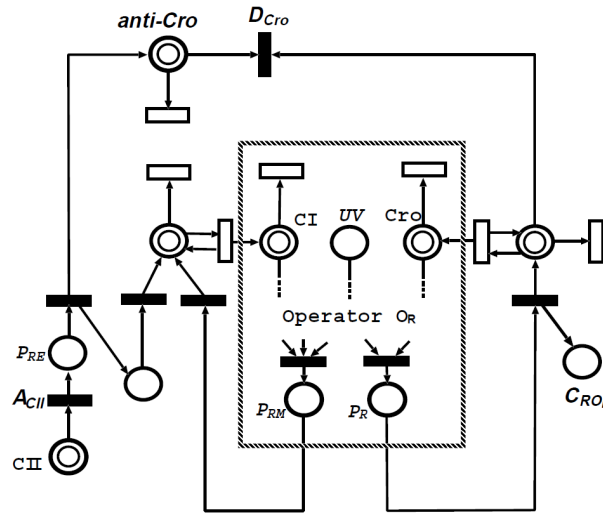


Figure 7: Feedback mechanism of CI and Cro proteins, which represents part of the early stage gene expression of  $\lambda$  phage (Matsuno et al., 2000). An explanation of types of graph nodes is the same as in Figure 6.

structure, molecular interactions or even species interactions. Network structures enable formal analysis of the modelled systems, mechanisms, and relations by using algorithms and methods from graph theory and other branches of discrete mathematics. The information obtained from such networks can be used in different ways to increase the understanding of biological systems. Several approaches have been recognised by (Alm and Arkin, 2003). For example, network structure can be used to propose hypotheses how the modelled systems are organised. Existing hypotheses can be tested and confirmed or rejected on the basis of the network data. Finally, existing open questions can be reformulated from a network perspective, for example, the role of the network structure in the evolutionary process and the role of evolution in shaping the network structure (Alm and Arkin, 2003). Studies in systems biology and graph theory have revealed that widely studied complex networks such as social networks, scientific co-authorships and the internet in fact share many features with certain biological networks, for example, the power-law node degree distribution, hierarchical modularity and small-world properties (Alm and Arkin, 2003). The architecture and physical properties of biological networks and networks in general are discussed in length by (Wuchty et al., 2006), (Alm and Arkin, 2003) and (Zhu et al., 2007).

Systems biology was proven to be successful in modelling complex biological processes (Kestler et al., 2008). Before the analysis of the network dynamics, one needs to understand the network structure (Kitano, 2002). There are various representation formalisms that can be used to represent a network structure, including the directed graphs formalism as used in the Systems Biology Graphical Notation by (Le Novère et al., 2009) or the modified EPN (mEPN) scheme proposed by (Raza et al., 2008). To construct the signalling network structure, different information sources can be used, including pathway databases such as the KEGG Pathway (Kanehisa and Goto, 2000), Reactome (Tsesmetzis et al., 2008) and BioCyc (Krummenacker et al., 2005), integrated knowledge sources such as ONDEX (Köhler et al., 2006) and Biomine (Sevon et al., 2006; Eronen and Toivonen, 2012), and the scientific literature itself.

## 2.3 Text processing in systems biology

Given that most of human biological knowledge is still stored only in the silos of biological literature, retrieving information from the literature is required when building the signalling network structure. Scientific literature can be inspected manually or analysed by natural language processing and information extraction tools. There are numerous biological models which were manually constructed based on an in-depth literature survey, such as the macrophage activation model developed by (Raza et al., 2008, 2010), or terpenoid biosynthesis pathway (Hawari and Mohamed-Hussein, 2010). On the other hand, state-of-the-art technologies enable information extraction from scientific texts in an automated way by means of text processing techniques, based on the advances in the area of natural language processing (NLP) of biology texts (see e.g., the research advances of the emerging bioNLP community at <http://www.bionlp.org/>).

Similar to our work which involves the extraction of a set of (component1, reaction, component2) triplets from biology texts, several existing NLP tools enable the extraction of interactions between the components (e.g., see the review by (Ananiadou et al., 2010)). The most common NLP approaches can be classified into three categories (Cohen and Hunter, 2008): rule-based approaches, machine-learning approaches and co-occurrence-based approaches. Examples of rule-based systems include GeneWays (Rzhetsky et al., 2004), Chilibot (Chen and Sharp, 2004), PLAN2L (Krallinger et al., 2009) and the approach proposed by (Ono et al., 2001). Combined methods, including co-occurrence-based approaches, such as the one developed by (Blaschke and Valencia, 2002) and upgraded in the BioRAT system by (Corney et al., 2004), are less appropriate for systems biology as the information retrieved is partial and can therefore not be directly transformed into a graph format used



for signalling network modelling. In most systems, the information is retrieved only from abstracts; an exception is BioRAT which can process full texts, albeit using a quite general vocabulary (Corney et al., 2004).

Most of the aforementioned approaches enable the users to query the extracted information, but do not result in an explicit network structure which can be visualised for simple inspection by the biology experts. Exceptions are the Chilibot system by (Chen and Sharp, 2004), the approach by (Blaschke et al., 1999) and the GeneWays system by (Rzhetsky et al., 2004). These systems are however not directly applicable in our context for the following reasons. The Chilibot system enables the search for relations by querying only a limited number of entities without supporting the complete network structure construction. (Blaschke et al., 1999) extract information only from abstracts. The closest to our work is the GeneWays system (Rzhetsky et al., 2004) which enables the extraction, analysis, visualisation and integration of molecular pathway data, but the system is not publicly available.



### 3 Methodology for PDS Model Construction: An Overview

*A journey of a thousand miles must begin with a single step.*

Lao-Tzu

A conventional way of manually constructing the model of a dynamic system can be accompanied by additional steps which can speed up and enhance model construction. In our work, the construction of a PDS model based exclusively on manual knowledge engineering was not feasible due to the PDS model complexity. For this reason, the laborious work of manual PDS model creation was complemented by additional automatised and semi-automatised steps. These steps address the model structure as well as model dynamics.

The purpose of this chapter is to provide a global overview of the methodology which we have developed and applied to build the PDS model (Figure 8). After the problem was defined, the most appropriate modelling formalism was selected. Next, the PDS model structure was manually constructed by eliciting knowledge from the domain experts and the literature. Since the manual estimation of the parameters was not feasible due to the PDS model complexity, an automatic method based on the DE algorithm (Storn and Price, 1995) was developed. We illustrate the whole process loop of converging to the dynamic parameters that best satisfy the expert evaluation criteria. This process includes elicitation of the knowledge from the biologists, formalisation 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 simulation results as well as optimised model parameters, which provide an insight into the biological system. The obtained structure of the PDS model was extended by applying a semi-automated approach to extract the information on relations between the biological components from the literature and various plant databases. The most relevant details of every step of the construction process are presented in the following paragraphs.

#### 1. Problem identification

In this step we have defined the requirements for the PDS model in collaboration with the experts from the National Institute of Biology, Ljubljana, Slovenia. The PDS model is intended to represent only the part of the plant cell involved in the defence response, which resulted in omitting the cell-to-cell communication within the whole plant. The model is supposed to verify whether the plant will have a resistant reaction to survive the virus attack if some genes are silenced. In practice, this greatly helps the domain experts in the design and performance of wet-lab experiments with plants. To be able to confirm or reject the hypothesis, the PDS model has to contain the genes that are candidates for silencing.

The goals of model development are the following:

1. Better understanding of the biological mechanism on the system level, which includes the relationships between entities.
2. Prediction of experimental results with the aim to:

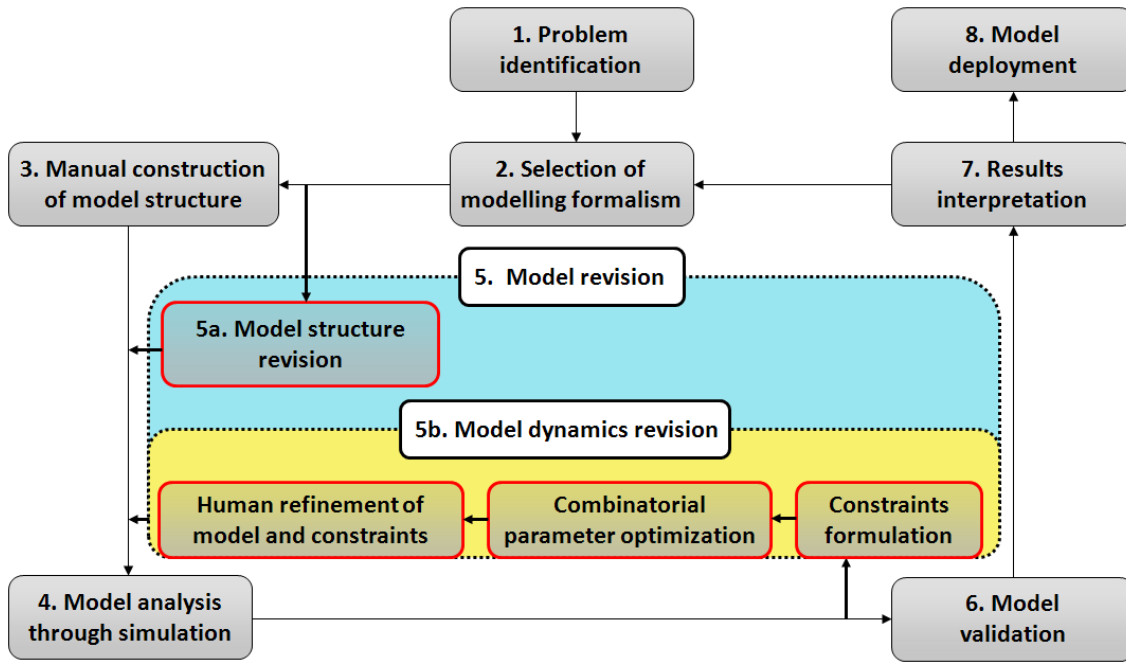


Figure 8: A schema of the developed methodology for the PDS model construction.

- (a) detect crucial reactions in the plant defence process,
- (b) predict the final defence response when some genes are silenced, and
- (c) conjecture of new connections/interactions.

The PDS model is constructed for the purpose of assisting the experiment design. For example, if the biology experts hypothesise that silencing some of the genes might be a key to improve the immune response in plants, they will perform model simulation. When different genes are knocked out in the model, the simulation will give different responses and the biology experts will decide which results appear interesting from the biological point of view. Experiments where silencing of a gene on living plants is performed are time-consuming and the duration of one experiment is two years on average. Thus, simulation of the defence response mechanism will save time and also provide directions for further research.

In particular, the PDS model should represent the resistant interaction between *Arabidopsis thaliana* and Turnip Crinkle Virus (TCV). The PDS model should have a graphical representation with two different types of nodes: component nodes and interaction nodes, which represent interactions between the biological components. Each biological molecule can be represented only with one graph node (an exception are the input and output molecules of the transport). Additionally, the effects of positive and negative feedback loops should be evident from the dynamic behaviour of the key components in the loop.

Plants do not have dedicated immune response cells. Instead, in the case of a virus attack, every cell in the plant performs the defence independently and alarms the other cells to perform a similar response. PDS is a complex mechanism which includes gene regulation, metabolic and signalling pathways (Somssich and Hahlbrock, 1998), where the reaction rates do not depend on the concentration of the reactants in the same way. In a plant cell, there are three major pathways responsible for the defence response: SA, JA and ET pathway (Reymond and Farmer, 1998). The network structure of the metabolic parts of the three pathways (SA, JA and ET) is available in KEGG. Thus, the biological molecules of the metabolic part and the relations between them are known.

The final result of the induced defence response is a group of proteins. If there are enough proteins produced, the plant will survive the virus attack. Otherwise, the virus will spread

and the plant will die. All three pathways are non-trivial to model and include complex regulatory mechanisms, such as positive and negative feedback loops. Moreover, the pathways include different molecule types, such as metabolites, genes, and proteins. Additionally, the defence response occurs in several parts of the plant cell, such as chloroplast, nucleus and cytoplasm. To simplify the PDS model, the spatial distribution is not considered for the separate compartments, but we included transport in critical points of the network.

There are many unknown issues related to the PDS mechanism. The signalling part of the SA, JA and ET pathways is not available in the databases, which implies that it is not known exactly which biological molecules are participating in the networks. This information and the information regarding the relations between the biological molecules in the signalling pathways can be retrieved from the literature. Furthermore, the kinetic parameters for neither the metabolic nor signalling parts of the pathways are available. Most of the experiments in plant research are performed with *Arabidopsis thaliana*, however, few kinetic studies of the PDS mechanism were completed. Currently, there is only one experimental dataset with five time points publicly available for *Arabidopsis thaliana* (Yang et al., 2007).

## 2. Selection of modelling formalism

The decision which modelling formalism to select depends on the requirements defined by the domain experts, currently available knowledge and open issues related to the PDS mechanism. First of all, the main limiting factor are few experimental data available. Only one dataset with five time points, available at the starting time of this work, was not enough to be used as a basis for model construction and validation. On the other hand, there is a lot of domain knowledge related to the modelling of the PDS mechanism that has not yet been formalised and systematised. For this reason, we present the model structure first in the form of a directed edge-labelled graph<sup>1</sup>, which is a common way to present biological networks in systems biology and which is, at the same time, intuitive for the domain experts. The directed graphs formalism is often used in systems biology, for example in the form of the Systems Biology Graphical Notation developed by (Le Novère et al., 2009) or the modified EPN (mEPN) scheme proposed by (Raza et al., 2008). The selected graph presentation allows also for an easy integration of the new relations which can be extracted in a semi-automated way from the literature.

A graph presentation of the model structure can be used to reveal potentially interesting new links. However, it does not contain any information on the PDS model dynamics and does not meet the requirements of domain experts that a graph should have two types of nodes (component and reaction nodes). Therefore, for simulation purposes of the PDS model, we have investigated the existing modelling formalisms. First, we have considered the qualitative modelling approaches for building the initial PDS model. While statecharts and hybrid automata were evaluated as inappropriate, the modelling formalisms that we have considered were the following: Boolean networks, process algebra and Petri nets (PNs). Intuitive graphical representation of the network structure was considered an important characteristic of the modelling formalism by domain experts. From the modelling approaches that were reviewed in Chapter 2, intuitive graphical presentation is a characteristic of PNs, statecharts and Boolean networks. The modelling formalisms that have also the mathematical apparatus developed for the model analysis are PNs and ODE. In case kinetic parameters should be included in the model, we can consider ODE, FPN or HFPN formalisms. Finally, the switch of biological reactions is an attribute of HPN, HFPN and hybrid automata formalisms. When summarising all the characteristics and the intersections of all the formalisms (see Figure 9 for an overview), the HFPN formalism proved to

---

<sup>1</sup>Directed edge-labelled graph of the PDS model structure represents biological components as nodes and biological reactions as vertices between them.

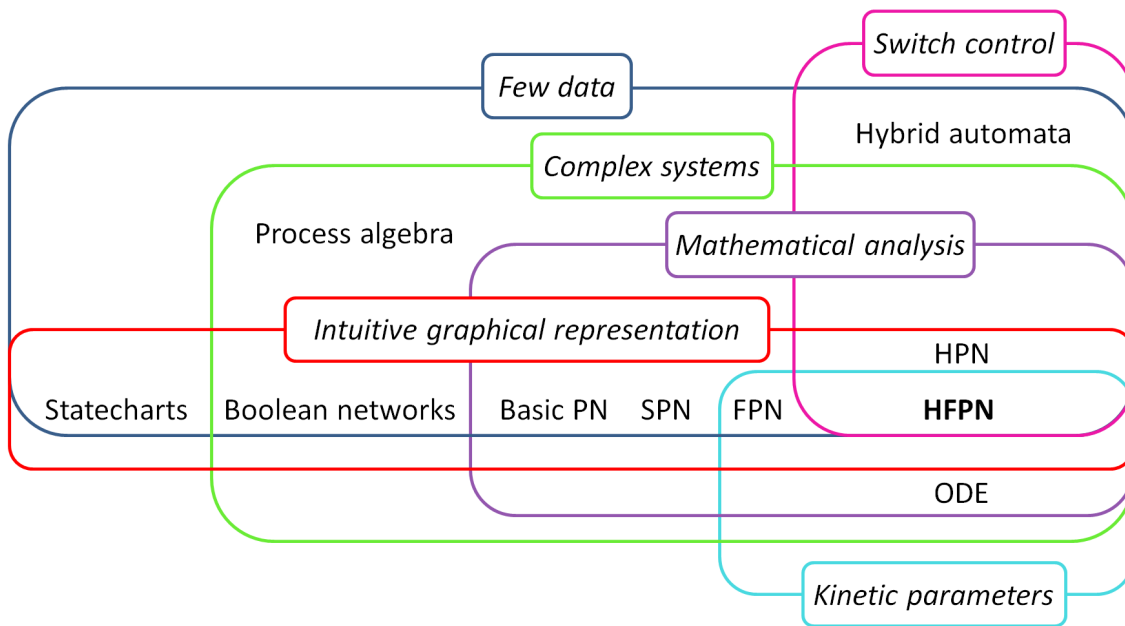


Figure 9: An overview of the modelling formalisms and the defined requirements for the PDS model.

be the most appropriate. Therefore, we have selected it for the construction of the initial PDS model.

Directed edge-labelled graph, HFPN and ODE approaches are interconnected, each of them having certain advantages in our work. More specifically, the directed edge-labelled graph approach presents the relations between components in the manner that are written in the articles, thus making easier manual construction of the PDS model structure. It is also easy to upgrade automatically so constructed model structure with triplets from texts in the form of (*subject, predicate, object*). ODEs are used for dynamics analysis but are not user friendly for the manual construction of the complex models. However, HFPN bridges these two approaches. It has a graph presentation, which is different than directed edge-labelled approach described in our work, but has intuitive presentation of biological reactions. By introducing the reaction type of graph node in HFPN it can be clearly seen which reactants trigger production of the particular product. In addition, HFPN implements ODEs which allows model simulations. The transition between HFPN and ODE is presented in Section 5.1.

### 3. Manual construction of model structure

In this step we describe several details of the manual model construction which is the most time-consuming phase in the whole model construction process. The initial PDS model structure was constructed manually by considering knowledge from the literature, different biological databases and domain experts. This initial model structure was first created in the form of a directed edge-labelled graph.

The literature, which was analysed for manual construction of the model structure, was selected by the domain experts. We have focused on scientific papers related to PDS research. Ideally, a citation in one review paper or in two scientific papers was required for the inclusion of the information into the pathway diagram. If the information was available in one publication only, we have critically assessed the publication quality (e.g., high impact factor and author’s relevance in the field) before incorporating the information into the model. The information about the biosynthetic pathways and other available data was acquired from different databases such as KEGG and TAIR (Swarbreck et al., 2008).

KEGG was used as a backbone for building the metabolic pathways, the biosynthesis of the hormones and the main reactions involved in this process. The additional reactions and the genes involved were implemented according to the Aracyc database. TAIR provided gene information, while the synonyms were acquired from iHOP (Hoffmann and Valencia, 2004) and TAIR. A list of all structure interactions was compiled, including the details of the information sources, and is available in Section A.1 in Appendix A.

Members of component families were selected according to their function. The family members were joined under a common family name if more than one component (family member) could be involved in the listed reactions. The genes involved in hormone biosynthesis are usually well described and implemented in the reaction scheme of different databases such as plant metabolic network. Therefore, the component family members of biosynthetic pathways were defined as in these databases. The other relations were manually acquired either from the studied literature (such as the ET receptors) or from the TAIR database.

After the PDS model structure was obtained, the part of the PDS model structure, which is of most interest, was selected and converted into the HFPN formalism. Once the PDS model is transformed to the HFPN environment, it is set up for the simulation studies.

#### 4. Model analysis through simulation

Analysis of the dynamic model behaviour was performed through iterative simulations of the manually constructed PDS model. Based on the experts' evaluation, the iterations were repeated until the simulation curves matched their expectations. Basically, the simulation is carried out by the software tool which is also used to construct the model.

The selected formalism to construct the PDS model for the simulation purpose is the HFPN formalism. The simulation was performed initially by Cell Illustrator (CI) (Nagasaki et al., 2004), which implements the HFPN formalism. The later simulations (when the parameters were automatically estimated) were executed in the simulator based on the C++ code exported from CI. The simulator outputs time series curves of the dynamic behaviour of molecules of interest.

While simulating the manually developed PDS model it turned out that it is practically impossible to manually determine all the parameters due to the complexity of the PDS model. The feedback loops from three sub-models merged into the global PDS model were not synchronised. To help exploring the dynamics of the model more efficiently, we introduced three steps in the process of model dynamics revision: constraints formulation, combinatorial optimisation parameter search and human refinement of the model and constraints. These steps were used for dynamics exploration of the most studied pathway related to the PDS mechanism: the SA pathway. For the other two, the JA and ET pathways, the same methodology can be applied, which is the topic of further work.

#### 5. Model revision

##### 5a. Model structure revision

The process of fusing expert knowledge and manually obtained information from the literature to build the PDS model structure turns out to be time-consuming, non-systematic and error-prone. Therefore, we have introduced one additional step that enhances the manually built model structure: extraction of relations between biological components from the literature, using the natural language processing approach.

We have developed the Bio3graph tool (described in detail in Chapter 6) that searches literature for the relations between the biological components and outputs a graph of triplets in the form (*component1*, *relation*, *component2*). These triplets were compared with the manually developed PDS model and as a result we discovered new relations with respect

to the manually constructed PDS model. The new relations were investigated further and some of them turned out to be interesting hypotheses for further biological experiments. Moreover, we have also developed an incremental version of the Bio3graph tool, which updates the network structure with new sets of triplets having an initial model as a baseline. Whenever necessary, the structures of biological networks can be quickly updated by using the incremental version of Bio3graph.

## 5b. Model dynamics revision

### Constraints formulation

The constraints in this phase are mathematical expressions defined by the domain experts. They represent the rules how the simulation output curves of certain biological components should look like. The formalisation of the expert's knowledge into constraints is an iterative process. The domain knowledge accumulated in biology literature is recognised as valuable in this phase of knowledge elicitation. The purpose of constraints is to guide and speed up the parameter optimisation search by limiting the parameter search space.

In this work we have explicitly focused on the knowledge related to biological molecules and the relations between them. Several types of relations (constraints) were defined, such as "inequality in component concentrations", "growth rate of the curve", "shape of the curve", etc. The knowledge related to the reaction rates was not directly applicable and therefore not translated into constraints. However, the definitions of some constraint types have also changed throughout several iterative steps until the domain experts approved them.

### Combinatorial parameter optimisation

We decided to use an evolutionary algorithm for parameter optimisation. Evolutionary algorithms are stochastic optimisation methods utilising the ideas of biological evolution in problem solving. One of the popular algorithms within this class is the differential evolution algorithm. The DE algorithm performs a population-based search that optimises the problem by iteratively trying to improve a candidate solution with regard to a given measure of quality. The DE algorithm developed by (Filipič and Depolli, 2009) was used in our work.

The optimal parameter setting of the PDS model is defined as a combinatorial optimisation problem. The objective function for the optimisation is the normalised sum of the normalised violations of constraints specified by the biologists. If the objective function has value 0, it means that all the constraints are satisfied. Values in the range between 0 and 1 denote that a certain percentage of time-series curves that are involved in the definition of specific constraint do not satisfy the objective function.

### Human refinement of model and constraints

In this step, the domain experts perform manual validation. If dynamic behaviour of the curves does not fit the experts' expectations, the model structure and the constraint setting are modified. That is the case, for example, when the curve has oscillations throughout the whole simulation or some unexpected peaks and sudden drops. Also, peaks of several biological components should occur during the simulation in a certain order defined by the experts.

When model simulation with automatically determined parameters did not match the expectations of the biologists, the model structure and the constraints were refined. The model structure was again thoroughly revised and the components were either added or the ones with the similar functions were merged into the same node. Also, the constraints were reviewed and additional ones were defined or the current definitions were modified if they were considered to be too strict.

## 6. Model validation



So far, the validation of the simulation results is based solely on the judgement of the domain experts to ensure that the model is close to the real-life system. However, the simulation outputs should also be compared to the experimental data. This step represents the model validation method, which was developed for the specific case of the PDS model. The comparison with the experimental data was a challenging task in our work. The comparison was performed with only one experimental dataset where each curve had only 5 time points while the simulation output curves consist of at least 1000 simulation points.

The commonly used comparison methods, such as Dynamic Time Warping, are not applicable, due to a low number of time points of the experimental curve. Therefore, we proposed a new method, based on the constraint generation and evaluation. More specifically, since the simulation output curves are a result of constraint-guided optimisation process, we proposed to evaluate these curves in a similar way, with the set of *evaluation constraints*.

We name the constraints formulated by the domain experts as *training* constraints. From the experimental dataset that contains several components of interest we generate a set of *evaluation* constraints. These constraints are of the same type as the *training* constraints. The parameter optimisation, guided by the set of *training* constraints, outputs the simulation curves that violate the least these constraints. Then, after the experts' evaluation, we calculate in which degree the simulation curves violate the set of the *evaluation* constraints.

## 7. Results interpretation

The simulation results are interpreted by the biology experts. Since the PDS model was built based exclusively on the expert knowledge and the literature, the results of the simulation have to be interpreted cautiously. The output curves of the simulation allow for qualitative conclusions regarding the dynamic behaviour of the model.

It is possible to compare different curves to conclude which are the major components influencing the PDS mechanism. For example, by comparing the growth of certain curves, we can determine the components that contribute the most to the PDS response. This knowledge is of crucial importance for the plant scientists, since the key of the PDS response lies in the speed of the PR protein production.

## 8. Model deployment

The deployment of the PDS model will be accomplished in future work through the continuation of collaboration with the experts from the National Institute of Biology (NIB), Ljubljana. The finalised version of the PDS model will be used at NIB to assist the experimental design by generating hypotheses how the PDS system will behave when particular genes are silenced. It is planned also to publish the global PDS model in the .sbml file format which is a common format for systems biology. Thus, the PDS model will be made available to the scientific community for use and further improvement.



## 4 Manual Development of PDS Model Structure

*Without labour nothing prospers.*  
Sophocles

Biological models are often constructed manually by biologists and manipulated to study the dynamics of different pathways. First the model structure is acquired followed by analysis of the model dynamics. During this process the domain experts generally rely on different biological databases and literature. In our work, the knowledge stored in biological databases was not sufficient to build the PDS model structure for more complex cell mechanisms (positive and negative feedback loops). The first task in the construction of the model structure is the definition of the types of biological components and reactions to be modelled.

In this chapter we first define the PDS taxonomy of reactions and components. The manually developed structure of the PDS model was constructed in accordance with this taxonomy. The PDS model structure was first developed in a form of a directed edge-labelled graph. To set up the PDS model for the simulation, we transform its structure to the HFPN formalism. This chapter explains the transformation process and introduces the software tool used to construct the model. It also presents the construction process of each of the three PDS pathways (SA, JA and ET pathway).

### 4.1 Definition of PDS reactions and components

When constructing the PDS model structure, we focused on metabolic, signalling and gene regulation networks of SA, JA and ET, as they play a crucial role in mediating the induced defence responses in plants (Reymond and Farmer, 1998). To identify the relevant biological reaction and component types, we have followed the modified Edinburgh Pathway Notation (mEPN) (Raza et al., 2008), slightly adapted to our requirements (Figure 10).

The following PDS component and reaction types were identified. In total, four groups of biological components were recognised (level 1 in Figure 11.A): small compounds or metabolites (Chorismate, Jasmonic acid, Linoleic acid, etc.), proteins (Chorismate synthase, EDS5, etc.), genes (EDS5, etc.) and protein complexes (NPR1 oligomer, JA-Ile/COI1/SCF complex, etc.). The following reaction types were defined (level 2 in Figure 11.B): gene expression activation, protein activation, protein phosphorylation, catalysis, translocation, gene-protein binding, protein-protein binding, protein dephosphorylation, protein inhibition, gene repression and degradation.

Let us explain the details of the PDS components and reactions taxonomy as summarised in Figure 11. To simplify manual model construction, the components with similar functions were grouped into families (level 2 in Figure 11.A). For example, the node named LOX represents the entire family of LOX proteins (LOX1 - LOX6). The individual components (for example, LOX1-LOX6) represent the lowest level (level 3) in Figure 11.A.

The reaction types were also grouped at a higher level of abstraction into three groups (see level 1 in Figure 11.B): *activation*, *binding* and *inhibition*.

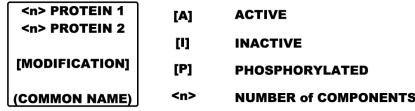
- *Activation* (A) denotes all reactions that follow the principle that, when two components X and Y are directly involved in the production of Z, the concentration of

## NOTATION SYMBOLS

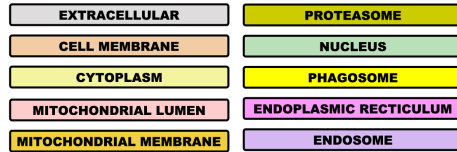
### COMPONENTS



### COMPONENT ANNOTATION



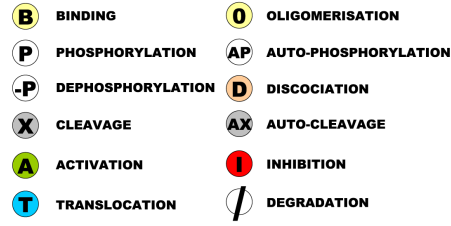
### LOCATION



### BOOLEAN OPERATORS



### TRANSITION NODES



### EDGE ANNOTATION

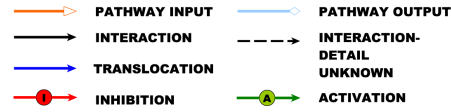


Figure 10: The presentation nodes and edges of the mEPN (Raza et al., 2008). The notation, which we have adapted for the PDS model, is grouped into four categories: components, transition nodes, Boolean operators and annotated edges.

Z depends on the concentration of both substrates (Figure 12.A). Reactions such as gene expression, protein activation, phosphorylation of proteins, catalysis and translocation are grouped under activation. Activation defines the processes that activate the next component (or raise its abundance). When a concentration of a (phosphorylated protein) increases, then the concentration of a product (which is inactive but phosphorylated protein in this case) also increases. The same holds for the translocation of components for their degradation.

- *Binding* (B) is defined as a close interaction between at least two components resulting in a complex (Figure 12.B). Binding is both the formation of a protein-protein complex or the binding of a protein to a DNA promoter region to regulate its gene expression. Binding results in the activation or inhibition of a particular target, gene or protein.
- *Inhibition* (I) is defined as a process in which one component blocks the performance of another component (Figure 12.C). Inhibition groups all biological reactions such as protein inhibition, gene repression and dephosphorylation of proteins.

Degradation was not abstracted to a higher level, but if this regulatory function was found explicitly in the literature (such as, for example, COI1 binding to SCF complex resulting in a degradation of JAZ repressors (Chini et al., 2007; Fonseca et al., 2009; Browse, 2009)), degradation was modelled as a binding reaction. In this way only few degradation reactions are modelled, while the rest is modelled in HFPN formalism as a separate reaction type.

## 4.2 PDS model structure

The PDS model structure is first presented as directed edge-labelled graph and then transformed to the HFPN formalism.

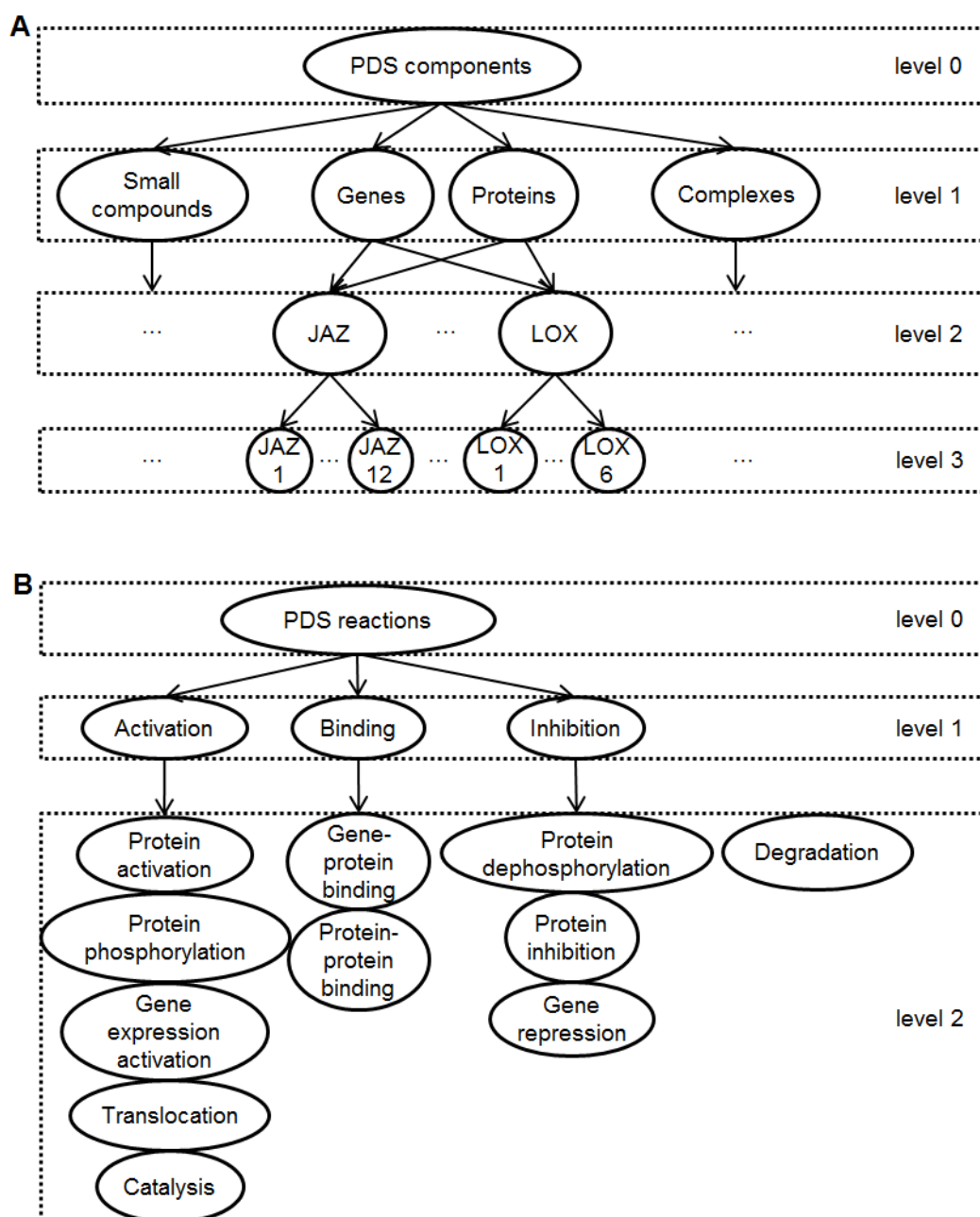


Figure 11: Taxonomy of PDS components and reactions. A) In the taxonomy of PDS components, there are four representation levels. The highest level (level 0) is the most abstract level, while the lowest one (level 3) represents single molecules. B) In the taxonomy of PDS reactions, individual reactions are represented at the lowest level (level 2) and are grouped according to their functionality into three groups at level 1: *Activation (A)*, *Binding (B)* and *Inhibition (I)*.

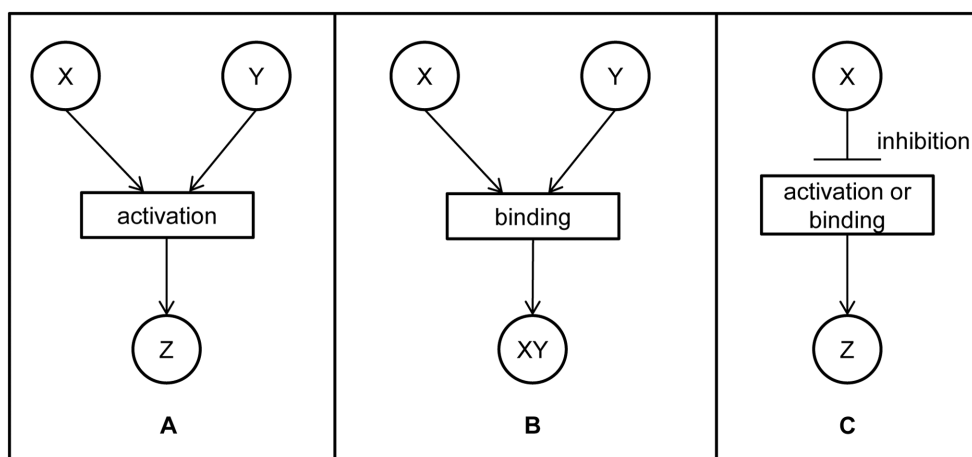


Figure 12: The PDS reaction types at level 1. There are three groups of reactions. A) *Activation (A)* denotes all the reactions directly involving two components X and Y in the production of Z, where the concentration of Z depends on the concentration of both substrates. B) *Binding (B)* results in the formation of a protein-protein complex or in the binding of a protein to a DNA promoter region to regulate its gene expression. C) *Inhibition (I)* is a process in which one component blocks the performance of another component.

#### 4.2.1 Directed edge-labelled graph presentation

As a first step, the PDS model structure is compiled manually in the form of a directed edge-labelled graph and visualised with the Biominer visualisation engine (Eronen and Toivonen, 2012). Reactions are presented as graph arcs labelled with the reaction type at the level 1 of Figure 11.B and components as graph nodes connected with arcs to each other. While building the PDS model structure, we first collected information at the level of component families (level 2 of Figure 11.A) where the members of family were joined according to their similar functions. Detailed descriptions of all the components, reactions and corresponding data sources are available in Section A.1 of Appendix A. After compiling all the necessary information, we decomposed the nodes which represent a component family into individual nodes to allow for in-depth inspection of PDS structure. Reaction decoupling into separate relations is illustrated in Figure 13. The members of the family were selected from the plant databases, Aracyc and TAIR and additional topic-related papers in the way that most of the functional analogues were listed. The list of the PDS model structure components for which decoupling was performed is available in Section A.2 of Appendix A.

We present the PDS model structure (which is decomposed to the individual nodes) in the form of directed edge-labelled graph in Figure 14. In this edge-labelled graph nodes represent the components and the edges represent the reactions. This is a very complex graph consisting of 175 nodes and 387 reactions. The components are presented at level 3 of Figure 11.A, while the reactions are abstracted to level 1 of Figure 11.B.

The edge-labelled graph shown in Figure 14 includes 175 components (31 small compounds, 135 genes/proteins and 9 complexes) and 387 reactions (231 activations, 49 bindings, 62 inhibitions and 45 produces reactions from binding reactants to their products). This graph is interactive and is visualised with the Biominer graph visualisation engine and is available in Section A.3 of Appendix A. Provided that the Java plug-in for the web browser has been installed and enabled, the reader can open and explore an interactive version of the Figure 14 at: [http://ropot.ijs.si/bio3graph/prepareVisualization.php?file=media/supplement/models/Supplement\\_file\\_2.bmg](http://ropot.ijs.si/bio3graph/prepareVisualization.php?file=media/supplement/models/Supplement_file_2.bmg).

Representing a model structure in the form of a graph based on the expert knowledge has several advantages. First, a graph representation is an intuitive way of presenting biological reactions. When model construction is based solely on the expert knowledge, it is easier

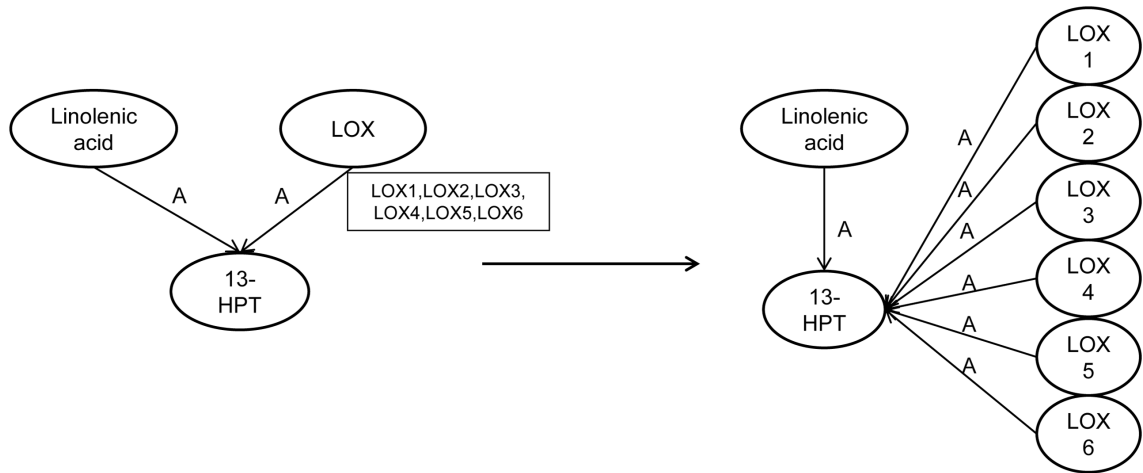


Figure 13: Principle of decomposing families of components by decoupling of reactions. The example shown in this figure illustrates the expansion of the PDS model structure from the family to the individual component level. Here, the decomposing of LOX node is done from the protein family level (level 2) to the single protein level (level 3). The final result of the expansion is a graph with 8 nodes and 7 edges.

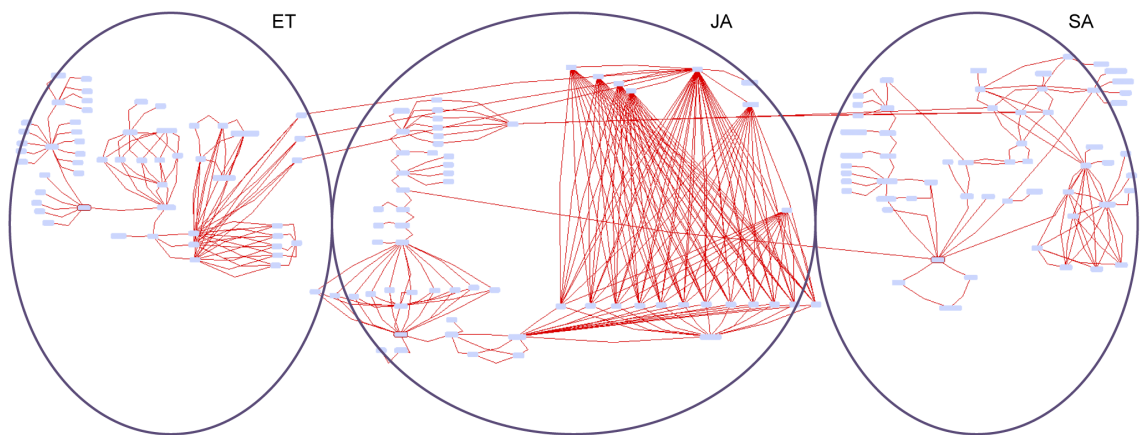


Figure 14: Manually constructed PDS model structure visualised as an edge-labelled graph. This graph, consisting of 175 nodes and 387 edges, is interactive and is visualised with the Biomine graph visualisation engine, enabling its closer inspection by zooming into its subparts and rearranging the node and the arc positions in the 2D space. The graph is organised into SA, JA and ET pathways with their crosstalk connections. The node borders of the main pathway components SA, JA and ET are coloured with red.

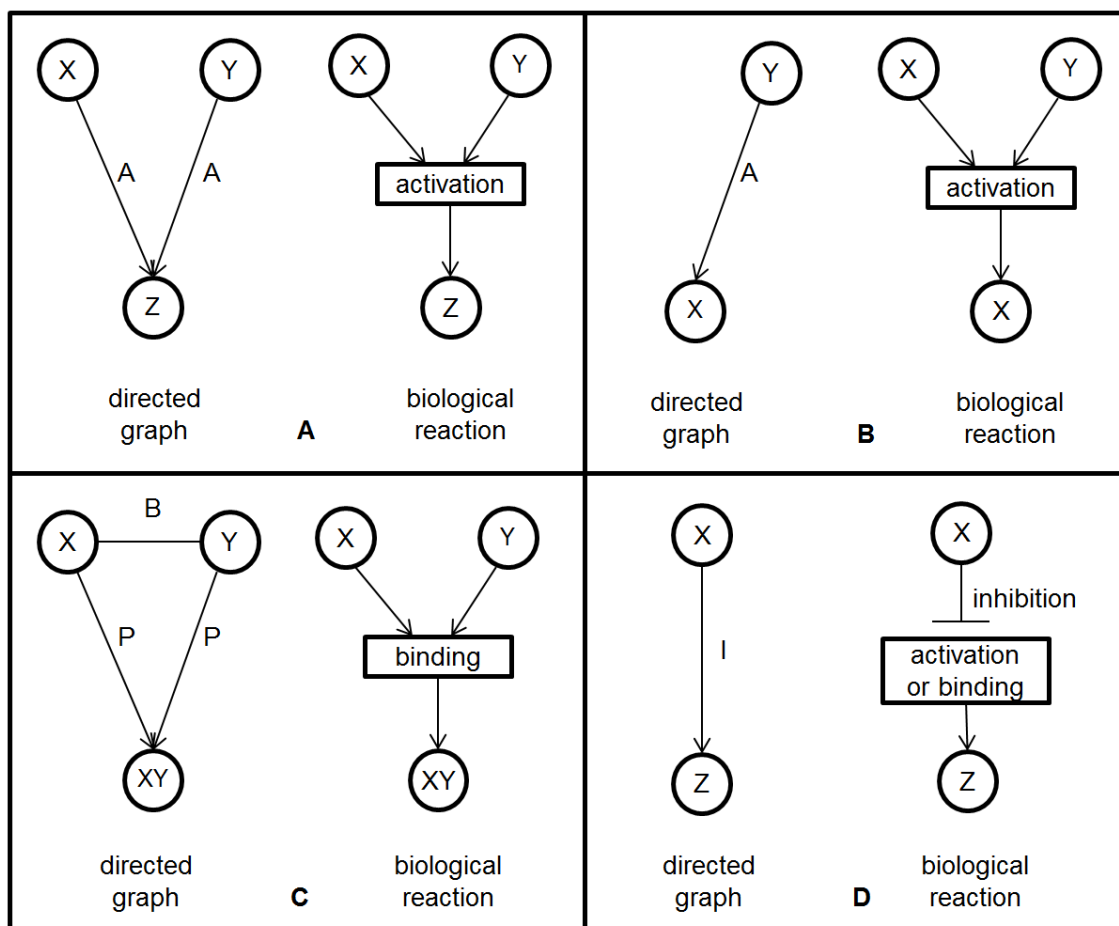


Figure 15: The principles of conversion from the edge-labelled graph format. A) *Activation* reaction (labelled A) with arcs between the reactant and the product node is transformed to a reaction between two components. B) *Activation* (labelled A) on a transcription level is a special type of *activation*, when Y induces the activation of gene X to produce protein X. In this case we omit the gene transcription level when transforming the level 2 structure from the edge-labelled graph. C) *Binding* (labelled B) relation is transformed from a B relation between the reactants and an additional relation produces (labelled as P) between the reactant and the product into a relation between two reactant nodes X and Y. D) *Inhibition* (labelled I) is the blocking of the activation or binding reaction between components by a third component X, resulting in reduced production of product Z.

to start by building the model structure in the form of a graph contrary to, for example, mathematical equations. Second, in biological literature, there are many structural models that are easy to present in the form of a directed edge-labelled graph. Thanks to that fact, many model structures in systems biology do not have to be built from scratch. Finally, graph representation is compatible with the outputs of several relation extraction algorithms from texts based on natural language processing techniques. This allows easy complementing the model structure, presented in the form of a graph, with the information automatically extracted from the literature which speeds up the process of model structure construction.

#### 4.2.2 HFPN presentation

We have selected HFPN as the formalism to model the PDS mechanism and built the model using the Cell Illustrator (CI) software. This software was developed from the initial version named Genomic Object Net (Nagasaki et al., 2004). Apart from the basic PN formalism, it allows modelling with ordinary differential equations (ODE) (Coddington and



Levinson, 1955) hidden to the end user through a user-friendly HFPN graphical interface. The software facilitates easy building of the network structure. CI 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, CI does not have the capability of automatic optimisation of dynamic parameters. However, CI has an option to export the model code into several programming languages, such as Python, C++, Visual Basic, etc. This option made the introduction of combinatorial parameter optimisation possible.

The principle of conversion from directed edge-labelled graph presentation to HFPN formalism is shown in Figure 15<sup>1</sup>. Two relations, one for each component, are merged into one reaction of two components resulting in a joint product. The selected CI graphical symbols for the different component and reaction types are presented in Figure 16.

Biological reactions are represented in CI at the modelling level 2 (Figure 11.B). Nine types of reactions are represented with six graphical symbols. Gene-protein and protein-protein binding are represented with one graphical binding symbol. Also, protein activation, protein phosphorylation and catalysis are represented with one symbol. Protein dephosphorylation, protein inhibition and gene repression are represented with an inhibition arc<sup>2</sup>.











Reaction types		Component types	
 gene-protein binding, protein-protein binding	 protein activation, phosphorylation, catalysis	 small compound	 gene
 degradation	 gene expression	 protein	 protein complex
 protein dephosphorylation, inhibition, gene repression	 transport		

Figure 16: Graphical representation of biological components and reactions modelled with the CI software.

Four types of biological components (level 2 in Figure 11.A) are presented with four different graphical symbols for 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.). While building the PDS model using the CI software, we have grouped components with similar functions into a single node that represents an entire family of these components (level 2 in Figure 11.A). We show below each sub-model separately transformed from directed edge-labelled graph to the HFPN formalism presentation and prepared for the analysis of dynamic behaviour.

### SA sub-model structure

SA belongs to a wide variety of phenolic molecules with a hydroxyl group. Phenolic components are the plant's secondary metabolites with a broad spectrum of functions. SA directly

<sup>1</sup>Graphical presentation of HFPN formalism has the form of biological reaction presentation with two types of nodes (component and reaction nodes).

<sup>2</sup>Note that among six types of graphical symbols representing reactions, only inhibition is graph arc while the rest of the reactions are the graph nodes.

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 and is considered to have an important role in both local and systemic 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 17, involves a two-step reaction catalysed with isochorismate synthase (ICS) and isochorismate pyruvate lyase (IPL) (Vlot et al., 2009). Most of the SA synthesised in response to a pathogen attack comes from the ICS/IPL pathway. *Arabidopsis thaliana* encodes two ICS enzymes, ICS1 and ICS2. When a plant cell is attacked by pathogens, 90 % of SA is synthesised through ICS1 (Vlot et al., 2009).

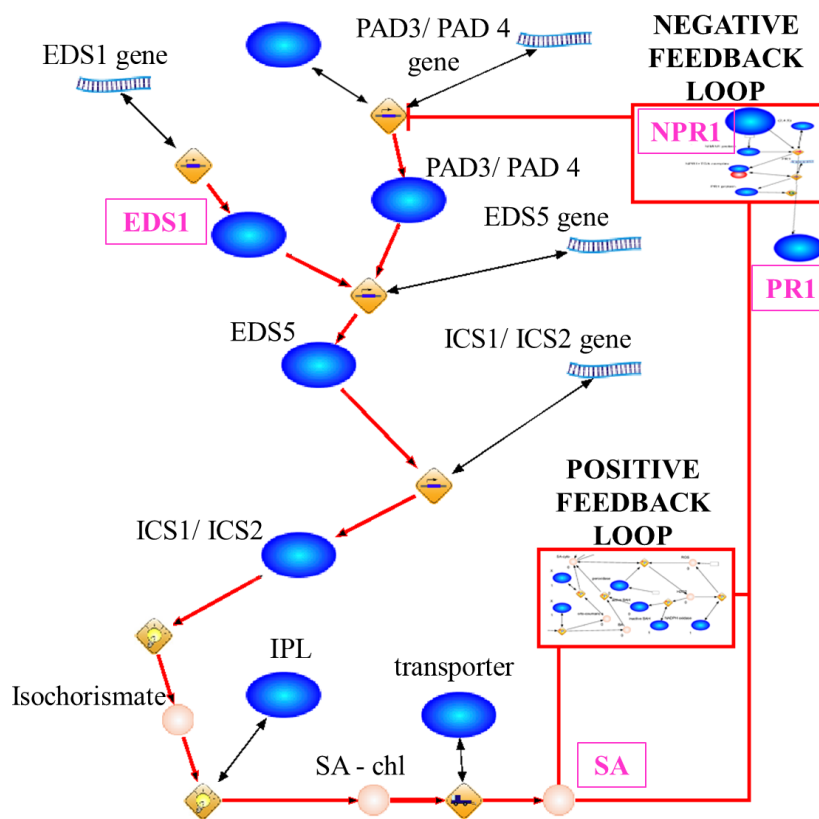


Figure 17: A simplified part of the model of the biosynthesis and signalling pathway of SA, manually constructed in the CI software. The whole SA-submodel is shown in Figure 24.

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 to survive the virus attack. The negative feedback loop slows down a signalling process, while the positive feedback loop accelerates it. In biological systems, positive feedback has an important role since it ensures a very fast response to a signal. In a system with positive feedback, the increase in some variable leads to a situation in which that quantity is further increased through its dynamics. In the SA sub-model, such increase happens to the SA component (Figure 17). However, this kind of behaviour caused

by positive feedback has a destabilizing effect and is usually accompanied by a negative feedback, i.e., saturation that limits the growth of the quantity. In the SA sub-model, the cascade product regulates its own concentration by activating or inhibiting the genes involved in its biosynthesis. For example, NPR1 inhibits the expression of the EDS1 and PAD4 genes (see Figure 17) (Shah, 2003) that activate the production of SA and consequently diminishes its own production, thus forming a negative feedback loop.

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 PDS which limits the viral spread and multiplication. The activation of the HRT protein stimulates the accumulation of SA (Chandra-Shekara et al., 2004). SA accumulation results in the monomerisation and 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 first version of the SA sub-model, named v1.0, contains in total 52 biological components and 38 reactions. This model was a subject to the iterative process of model dynamics analysis and combinatorial optimisation (see Section 5.2) resulting in the final version SA v3.0. As one component included in the final version of the SA sub-model is still not fully identified, it was labelled as X1 (Figure 24).

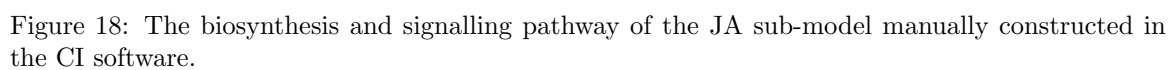
### JA sub-model structure

One of the major functions of JA is the regulation of PDS, especially in the cases of wounding of plant leaves and insect attacks. The main biosynthetic pathway for JA is the oxylipin pathway, linolenic acid being a substrate for the JA biosynthesis (Staswick, 2008). JA can be derivatised into different amino acid conjugates. Jasmonyl-isoleucine (JA-Ile) is the conjugate whose biological activity has been proven (Staswick and Tiryaki, 2004). In the presence of JA-Ile, the SCF complex, composed of SKP1 (S-phase kinase-associated protein 1), cullin, and a RING finger protein (RBX1/HRT1/ROC1), binds to the F-box protein Coronatine insensitive1 (COI1). The SCFCOI1 ubiquitine ligase binds to JAZ and presumably ubiquitinates it (Staswick, 2008; Gfeller et al., 2010; Devoto and Turner, 2005; Chini et al., 2007). When ubiquitinated JAZ repressors are targeted for degradation in 26S proteasome, they result in the de-repression of the transcription factors such as MYC2 and other beta helix-loop-helix transcription factors (Fernández-Calvo et al., 2011) which activate JA-dependant PR gene expression (Moore et al., 2011).

The JA sub-model contains 36 biological components and 65 biological reactions (Figure 18). As some components included in the JA sub-model are still not fully identified, they were labelled as X2 and X5. The main components serving the experts' evaluation how well the JA sub-model simulates the JA pathway were the complex JA-Ile/COI1/SCF, labelled as "NO DEFENCE RESPONSE" in Figure 18, and the component node THI2.1/JR1/VSP1/ATCLH1, labelled as "DEFENCE RESPONSE".

### ET sub-model structure

The ET production is influenced by environmental factors. It is usually induced by mechanical wounding or some other kind of environmental stress. L-methionine is transformed by S-adenosyl-L-methionine (SAM), 1-amino-cyclopropane-1-carboxylate synthase (ACS) and ACC oxidase (ACO), to form a gaseous hormone ET (Wang et al., 2002). When synthesised, ET binds to its receptors. There are five membrane-located receptors identified in Arabidopsis (ETR1, ETR2, EIN4, ERS1 and ERS2) (Kendrick and Chang, 2008; Zhao and Guo, 2011). The binding of ET to its receptor leads to CTR1 deactivation, which finally results in downstream activation of EIN3/EIL1/EIL2 transcription factors (Kendrick and Chang, 2008; Stepanova and Alonso, 2005). CTR1 levels are also regulated by ubiquitina-



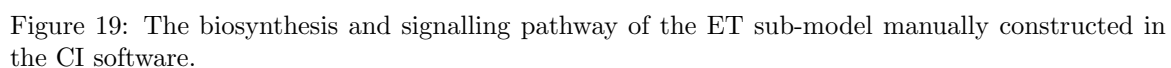
tion and 26S proteasome degradation via EBF1/EBF2 - Skp- Cullin-F-box (SCF) E3 ligase complex (Zhao and Guo, 2011). The concentration has to be well regulated, since they are the crucial positive regulators of ET signalling.

The ET sub-model includes 28 biological components and 45 biological reactions (Figure 19). Similar to the JA sub-model, the ET sub-model has two main components to test the model dynamics. These are the complex Et\_receptor/CTR1 labelled as "NO DEFENCE RESPONSE" and defence proteins GST1/b-CHI/PR4, labelled as "DEFENCE RESPONSE". When the behaviour of these two variables is opposite in approximately the same time frame, the dynamic behaviour of ET sub-model is considered correct.

### 4.2.3 Comparison with the state-of-the-art model structures of the PDS mechanism

The developed PDS model structure, shown in the form of graph in Figure 14, contains more detailed information compared to the structural models of the subsets of PDS mechanism (Olmedo et al., 2006), (Staswick, 2008) having in total 175 components and 387 reactions. Figure 6 of the study of (Olmedo et al., 2006) shows the structural model of ET pathway containing 17 components and 11 reactions. In the work of (Staswick, 2008), Figure 1 contains 13 components and 7 reactions. The work of (Naseem et al., 2012), contrary to the abovementioned studies, makes a broader overview and systematizes knowledge regarding the PDS mechanism. Their complex network, containing 105 components and 163 reactions, is of similar size as our manually built PDS model structure. However, our work focuses on the most important pathways of plant defence against pathogen attacks (SA, JA and ET) while the study of (Naseem et al., 2012) includes several other components like GA, ABA, auxin, etc. resulting in less detailed information on the particular SA, JA and ET pathways.

The PDS model structure transformed to the HFPN presentation is prepared for the analysis of model dynamics. A slightly reduced structure, compared to the directed edge-labelled graph of Figure 14, contains in total 99 components and 68 reactions. The structure of one of the first simulation models of the PDS mechanism (Genoud et al., 2001), containing 18 biological entities and 12 Boolean operators, is less complex than the PDS model structure comprising three sub-models which we have developed according to the HFPN formalism (see Subsection 4.2.2). The more complete PDS model is the one of study of (Naseem et al., 2012) with 105 components and 163 reactions. This model was built for the simulation purpose, however, as mentioned above, it has a broader overview including several other components with the price of having less detailed information of the SA, JA and ET pathways, which are the focus of our work.



## 5 Constraint-Driven PDS Parameter Optimisation and Model Validation

*With four parameters I can fit an elephant,  
and with five I can make him wiggle his trunk.*

John Von Neumann

In the manual construction of the PDS model we were faced with a problem when merging the three sub-models. As parameter hand-tuning of the merged PDS model was not feasible due to the model's complexity, it was necessary to support this process by means of parameter optimisation. When the simulation model was refined through iterative optimisation steps, the next challenge was to determine the validity of the model. The task of model validation is to determine whether the model is "good enough" to represent the real-life system. A perfect simulation model does not exist. By definition a model is a simplification of the reality and thus, only a real-system would be a "perfect" model. Therefore, we can only argue whether a model is "good enough" considering its intended use.

This chapter represents contributions of this thesis regarding model dynamics and model validation. We describe the proposed iterative methodology of model dynamics revision for acquiring knowledge from the domain experts resulting in a new dynamic model of the SA pathway, the most studied pathway of the PDS mechanism. Next, we formalised biological knowledge in the form of constraints which have never been formally defined before. We describe the basics of mathematical optimisation and present the results of several iterative steps applied to the SA sub-model. Moreover, we present model validation concepts and the most common validation strategies. We focus the discussion on the validation techniques of the models without or with few experimental data available. The investigation of the related work resulted in applying the common validation strategy to the developed SA sub-model. In addition, we have developed another, better suited validation strategy and applied it in SA sub-model validation. In future work, it is planned to apply the same strategy of constraint-driven parameter optimisation and model validation of the JA and ET sub-models.

### 5.1 PDS model represented in the HFPN formalism

In the PDS model, all components are presented with continuous node types (i.e., the concentration value of the biological components is the real number). The reaction nodes, which are continuous transitions, were assigned functions that represent the reaction rates. These functions are based on the law of mass action (Keener and Sneyd, 2009). The law of mass action states that the reaction rate is proportional to the concentrations of molecules participating in the reaction. It has the following form:

$$r_f = k_f[A][B] \quad (6)$$

where  $r_f$  is the reaction rate,  $k_f$  is the rate constant and  $[A]$  and  $[B]$  are concentrations of the reactants. The mass action kinetics is a simpler way to model biological reactions than the

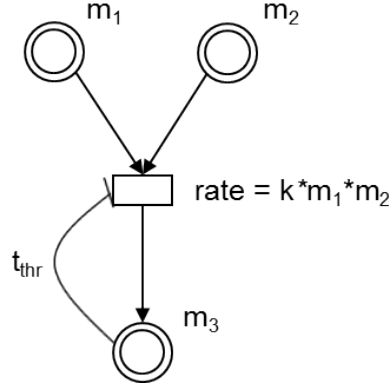


Figure 20: A simple reaction presented in the HFPN formalism.

Michaelis-Menten kinetics (Michaelis and Menten, 1913), making analysis of larger systems easier.

In the PDS model, the reaction rates have the following form:  $m_1 * m_2 * k$ , where  $m_1$  and  $m_2$  are the concentrations of the reactants and  $k$  is the reaction rate constant. We have assigned value 0.1 to most of rate constants of the activation and gene expression reaction types and value 0.01 to the degradation type. Furthermore, hand-tuned values of several inhibition thresholds are assigned to the inhibition arcs (functions of the reaction nodes and the inhibition thresholds of the JA, ET and SA pathways are presented in Figures 18, 19 and 24).

The CI software enables modelling with the ODE formalism which is hidden to the end user. If we consider a simple example of one reaction (Figure 20), we can formulate ODEs of all three biological components in the following way:

$$\frac{dm_1}{dt} = -k * m_1 * m_2 * g(m_3) \quad (7)$$

$$\frac{dm_2}{dt} = -k * m_1 * m_2 * g(m_3) \quad (8)$$

$$\frac{dm_3}{dt} = k * m_1 * m_2 * g(m_3) \quad (9)$$

where  $g(m_3)$  is the step function defined in the following way:

$$g(m_3) = \begin{cases} 1, & \text{if } m_3 \leq t_{thr} \\ 0, & \text{otherwise} \end{cases} \quad (10)$$

where  $t_{thr}$  is the inhibition threshold.

In the CI software, ODEs are simulated with the Euler integration method (Butcher, 2008). This method is the simplest and serves as the basis for more complicated methods. It belongs to the single-step integration methods and uses only the last step to estimate the value of the next one. The global error of the Euler method is proportional to its step size. Instability is a disadvantage of this method, which is resolved in more complicated ODE solvers, like Runge-Kutta, Adams-Moulton, etc. Improvement of the ODE integration method used by CI is one of the tasks for further work.

Moreover, the inhibition threshold function in the CI software is defined as a step function. This function models sudden change in system's behaviour which does not reflect properly the behaviour of a real-life system. Modification of the step function to the less abrupt function, like sigmoid for example, is also considered to be addressed in future work.

To simplify the procedure of hand-tuning the PDS model to the experts judgement, we have considered the Euler integration method and the step function in the inhibition threshold, implemented in the CI software, to be an acceptable approach.



## 5.2 Iterative process of constraints formulation, combinatorial optimisation and refinement of the model and constraints

This section introduces the mathematical optimisation approach and the definition of constraints formalised from the expert knowledge, which is one the contributions of this dissertation. Moreover, this section illustrates the iteration procedure of constraint-driven parameter optimisation of the SA sub-model.

### 5.2.1 Mathematical optimisation

In systems biology, many biological pathways are mostly qualitatively understood since the numerical data on kinetic parameters are often few. Due to the lack of existing quantitative data, employment of mathematical optimisation methods in systems biology is considered an important challenge.

Mathematical optimisation is defined as a selection of the best element from a set of alternatives with respect to some criteria. Basically, optimisation is a way to find either the minimum or the maximum of a selected property of a studied system. In this process, one first has to identify an objective, that is, a measure of the property in question, which can be quantified by a single number. The value of the objective depends on the system parameters. The task of optimisation is to find the values of the parameters that either minimise or maximise the objective. The parameters are usually constrained, most often by having at least a lower and an upper bound. Formally, an optimisation problem can be defined as a task that requires optimising the objective function (also named criteria, cost, utility or fitness function)  $f$ :

$$y = f(x) \quad (11)$$

where  $x$  is a vector of  $n$  decision variables defined over  $\mathfrak{R}$

$$x = [x_1, x_2, \dots, x_n]^T. \quad (12)$$

Decision variable vectors  $x$  that satisfy the constraints of inequality type

$$g_i(x) \geq 0, i = 1, 2, \dots, I \quad (13)$$

and constraints of equality type

$$h_j(x) = 0, j = 1, 2, \dots, J \quad (14)$$

are called the feasible solutions.

Optimisation problems are classified according to various criteria. Some classes of optimisation problems are briefly described below.

**Discrete vs. continuous** optimisation problems. This category is based on the type of decision variables. When the solutions of the discrete optimisation problem are permutations of finite numbers, the problem is called a *combinatorial optimisation problem*. In contrast, the infinite solution space characterises *continuous optimisation problems*. Solutions consisting of both integer- and real-valued decision variables also exist and are called *mixed integer optimisation problems*.

**Linear vs. nonlinear** optimisation problems. This difference is based on the type of the objective function and constraints. We refer to the problem as a *linear optimisation problem* if both the objective function and constraints are linear. Unlike linear problems, if the objective function or at least one constraint (if defined) is nonlinear, we deal with the class of *nonlinear optimisation problems*.

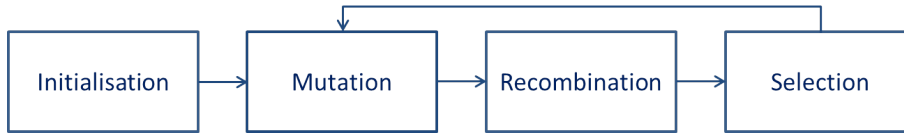


Figure 21: A scheme of an evolutionary algorithm (EA).

**Deterministic vs. stochastic** optimisation problems. When the solutions of the problem cannot be specified exactly, but only within some confidence interval, we refer to the problem as a *stochastic optimisation problem*.

**Single vs. multi-objective** optimisation problems. This group of problems is distinguished according to the number of objective functions.

Optimisation methods can be, according to the complexity of the optimisation problem, classified as exact or approximate. *Exact* (or complete) *methods* are guaranteed to result in an optimal solution for the finite size problems within a limited time interval. *Approximate methods* use different strategies to obtain a solution in a reasonable time. Various meta-heuristics belong to this class and one of the most popular among them are evolutionary algorithms (EAs).

### 5.2.2 Evolutionary algorithms and DE algorithm selection

*Evolutionary algorithms (EAs)* are stochastic optimisation methods utilizing 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 using 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 the 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 may result in a significant computational speedup. This class includes Genetic Algorithms (GA), Evolutionary Strategies, Evolutionary Programming and differential evolution (DE) algorithms. For the whole class of EA algorithms, the basic procedure is common and is shown in Figure 21.

One of the popular algorithms within the EA class is the *differential evolution* (DE) algorithm. The 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. A version of a DE algorithm developed by (Filipič and Depolli, 2009) was used in our work. A brief explanation of how the DE algorithm works is provided below.

In DE, the population of random solutions is initialised first and the solutions are evaluated with respect to the objective function  $f$ . Then, new candidates are created by means of vector addition or scalar multiplication. Next, every candidate is evaluated by comparison with its parent and the best are added to the new population.

The DE algorithm considers the candidate solutions as real vectors and not as binary strings (chromosome), which distinguishes it from GAs. It can be used for solving many practical problems where objective functions are nondifferentiable, non-continuous, non-linear, noisy, flat, multi-dimensional or have many local minima, constraints or stochasticity. The main advantage of DE algorithm is its implementation simplicity which allows it to easily solve optimisation problems requiring minimization process with real valued objective functions. For these and for additional practical reasons we have selected to use a DE algorithm in our work.

### 5.2.3 Constraint formulation and objective function definition

In systems biology, values of kinetic parameters in biological models are usually determined by the time-series data. A standard cost function represents a sum of the differences in values between experimental time-series data and simulation data, for each time point. However, this is not possible when experimental data are sparse and few. In the study of (Locke et al., 2005), a cost function was a sum of terms which quantify the agreement between a model and the qualitative experimental feature. Some of these terms represent the difference between the target phase and the average phase of the peaks of two biological components (LHY and TOC1 mRNA expression), or how broad the peak of certain components should be. Defining objective functions in the described way represents our contribution to biological science since the biological knowledge has been formalised in this process and has been made available for further exploration within the scientific community.

The formalisation of these expert terms represents an iterative process of eliciting knowledge from the experts. This knowledge is accumulated in biology literature and can guide the parameter optimisation search. Similarly to the study of (Locke et al., 2005), we have explicitly focused on the knowledge related to the biological molecules and the relationships between them. However, the main difference in our approach is that the objective function does not take into account experimental data, but is based solely on the knowledge of domain experts. The following five types of relationships, formed as constraints between the components, were defined:

- Inequality relationship between quantities of molecules.
- Growth rate of the molecules' quantities (for example, the quantity of molecule 1 grows faster than that of molecule 2).
- Curve shape (e.g., it starts at some initial level, reaches a maximum and then drops back to an approximately similar level).
- Minimal amplitude and minimal growth of a curve.
- Temporal sequence in curve maxima:
  - same time (molecule 1 has the peak at the same time as molecule 2), and
  - maximum before (molecule 1 has the peak before molecule 2).

The optimal parameter setting of the PDS model is defined as a combinatorial optimisation problem. In this approach, our contribution is the definition of the objective function as the normalised sum of the normalised violations of constraints specified by the biologists. If the objective function has value 0, it means that all constraints are satisfied, while value 1 results from all violated constraints. Values in the range between 0 and 1 denote that a certain percentage of time-series curves that are involved in the definition of the specific constraint do not satisfy it. The optimisation parameters include the reaction rate constants and the inhibition thresholds. The defined value range of the rate constants is between 0.01 and 10, while the value of the inhibition thresholds ranges between 0.1 and 1000.

We illustrate the development of the SA model through several iterative steps. In the following, we demonstrate how the parameter optimisation methodology was applied, by showing the results of the three iterative steps. This methodology includes manual revision of the SA sub-model structure and the revision of the defined constraints after every iteration.

### 5.2.4 Iterative revisions of SA sub-model

#### Step 1 - SA sub-model v1.0

The starting point of our work is the manually developed SA sub-model v1.0 that contains in total 52 biological molecules and 38 reactions including inhibitions. This was the initial model built in the CI software. Out of 38 parameters, 33 were optimised, while the other ones are left out from optimisation since they were representing initial values for the reactions. The simulator outputs 4 curves (as time series with 1,000 points with the sampling interval of 0.1) for each biological molecule, which were the most interesting for the biology experts. At the beginning of the sub-model construction process, in total 8 constraints were defined by the biologists for this sub-model. After the DE algorithm search was performed with a population number set to 10,000, the optimal parameters with respect to the objective function were estimated. With this set of parameters, for each violated constraint there is a value that represents the percentage of time points in which the constraint was violated. The overall objective function represents the average of constraint violations.

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

$$\begin{aligned} \text{lowerThan}(\text{SA\_chl}, \text{SA\_cyto}) &= 0.088 \\ \text{slowerRate}(\text{Chorismate}, \text{Prephenate}) &= 0.012 \\ \text{slowerRate}(\text{Chorismate}, \text{Phenyl\_pyruvate}) &= 0.052 \\ \text{slowerRate}(\text{Chorismate}, \text{Phenylalanine}) &= 0.055 \\ \text{zeroPeakZero}(\text{SA\_cyto}) &= 0.096 \\ \text{zeroPeakZero}(\text{PR1}) &= 0.037 \\ \text{equalRate}(\text{Prephenate}, \text{Phenyl\_pyruvate}) &= 0.028 \\ \text{equalRate}(\text{Phenylalanine}, \text{Phenyl\_pyruvate}) &= 0.007 \\ \text{finalCriterion} &= 0.376/8 = 0.047 \end{aligned}$$

Detailed values of kinetic parameters and the SA sub-model v1.0 are available publicly at [http://kt.ijs.si/dragana\\_miljkovic/](http://kt.ijs.si/dragana_miljkovic/). Based on the parameter set, the simulator outputs the curves of four 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 22. According to the biology experts, some parts of these curves are not considered correct even though the total criterion function showed that on average 0.047 of each constraint is violated. For example, in the first upper curve of the SA component (Figure 22), there is an unexpected change after the global peak drops down. Furthermore, the behaviour of the EDS1 curve (Figure 22) is not considered correct since the peak of the curve is too narrow and drops immediately down. Moreover, the raise of the NPR1 curve after the first peak does not fit the experts' expectations. For these reasons, the domain experts have revised the sub-model structure and have provided additional knowledge in the form of constraints.

## Step 2 - SA sub-model v2.0

After the inspection of the curves shown in Figure 22, the sub-model structure was revised by biology experts. This correction resulted in a second model version: SA sub-model v2.0 containing 61 biological molecules and 56 reactions. Out of 56 parameters, 51 were optimised, while the other ones are left out from optimisation since they were representing initial values for the reactions. Also, more constraints were specified, leading to a set of 33 constraints.

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

$$\begin{aligned} \text{equalRate}(\text{Prephenate}, \text{Phenyl\_pyruvate}) &= 0.004 \\ \text{equalRate}(\text{Prephenate}, \text{Phenylalanine}) &= 0.004 \end{aligned}$$

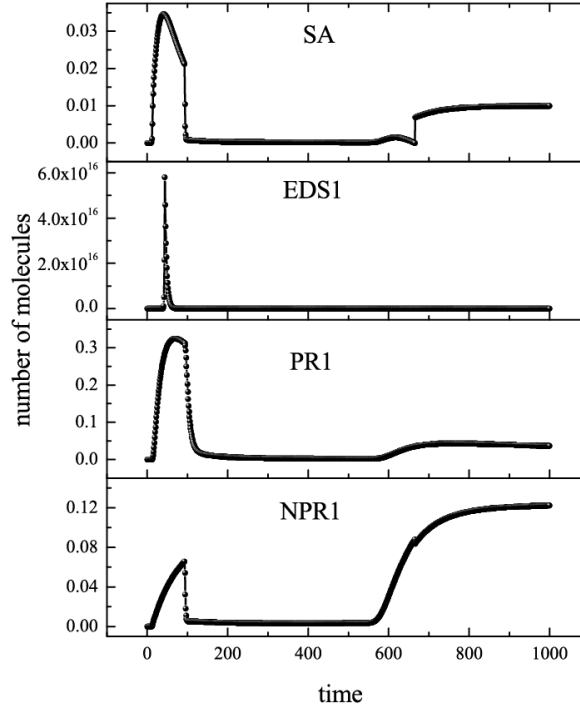


Figure 22: Dynamic behaviour of the SA, EDS1, PR1 and NPR1 variables based on the optimal parameter set estimated with respect to the objective function calculated from eight constraints during step 1.

```

equalRate(Prephenate,Trans_cinnamic_acid) = 0.003
equalRate(Prephenate,Orto_coumaric_acid) = 0.004
equalRate(Prephenate,BA) = 0.950
equalRate(Isochorismate,SA_chl) = 0.002
slowerRate(Prephenate,Isochorismate) = 0.093
slowerRate(EDS1,EDS5) = 0.028
slowerRate(PAD4,EDS5) = 0.081
slowerRate(ICS2,ICS1) = 0.004
slowerRate(NPR1_oligomer,NPR1) = 0.197
maxSameTime(ROS,HRT) = 0.930
maxAfter(HRT,MPK3) = 0.730
maxSameTime(MPK6,MPK3) = 0.094
maxAfter(MPK6,EDS1) = 0.000
maxSameTime(PAD4,EDS1) = 0.000
maxAfter(PAD4,EDS5) = 0.000
maxAfter(EDS5,ICS1) = 0.000
maxAfter(ICS2,SA) = 0.000
maxAfter(SA,H2O2) = 0.005
maxAfter(H2O2,BAH) = 0.000
maxSameTime(ROS,BAH) = 0.000
maxAfter(ROS,NPR1) = 0.000
maxAfter(NPR1,NPR_TGA_complex) = 0.000
maxSameTime(Isochorismate,Prephenate) = 0.067
lowerThan(SA_chl,SA) = 0.001
zeroPeakZero(Chorismate) = 0.333
zeroPeakZero(Prephenate) = 0.333
zeroPeakZero(Phenyl_pyruvate) = 0.333

```

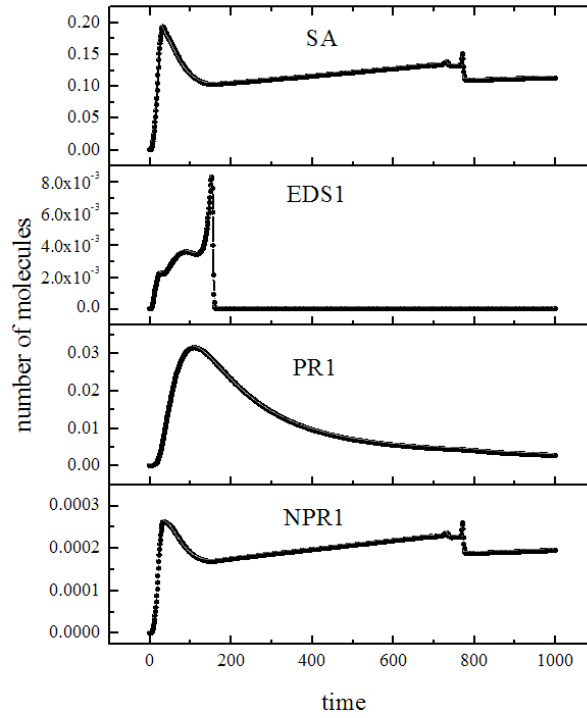


Figure 23: Dynamic behaviour of the SA, EDS1, PR1 and NPR1 variables based on the optimal set estimated with respect to the objective function calculated from 33 constraints during step 2.

$$\text{zeroPeakZero}(\text{Phenylalanine}) = 0.333$$

$$\text{zeroPeakZero}(\text{SA}) = 0.195$$

$$\text{zeroPeakZero}(\text{PR1}) = 0.028$$

$$\text{stopFast}(\text{virus}) = 0.000$$

$$\text{finalCriterion} = 4.7529/33 = 0.144$$

Detailed values of kinetic parameters and the SA sub-model v2.0 are available publicly at [http://kt.ijs.si/dragana\\_miljkovic/](http://kt.ijs.si/dragana_miljkovic/). The total criterion function shows that on average 0.144 of each constraint is not satisfied. The dynamic curves of the same 4 molecules (SA, EDS1, PR1 and NPR1) are shown in Figure 23. Even though the criterion function shows more violated constraints compared to step 1, the biology experts were more satisfied with the presented curves in Figure 23 compared to the curves in Figure 22. The curve of the SA component in Figure 23 does not have sudden drops in the last part but looks more smooth, even though it is still not optimal. The EDS1 curve in Figure 23 does not have the sudden narrow peak like in Figure 22, but the way the curve raises is not correct. The behaviour of the PR1 curve was satisfying while the NPR1 curve showed unexpected rise after the first peak, which was still considerably smaller compared to the previous iteration (Figure 22). Nevertheless, since the criterion function was overall worse, the model structure and the constraints were revised again. We have inspected the constraint definitions and the domain experts concluded that the constraint type related to the shape of the curve was defined too strictly (this means that the curve had to start from zero and go back to the zero value which was later not found as necessary). Therefore, this constraint is without strict limitations to start and drop back exactly to zero. It is only important that it rises and drops after a certain point, where the curve oscillations are penalised.

### Step 3 - SA sub-model v3.0

Finally, the last iteration resulted in the SA sub-model v3.0, containing in total 50 biological molecules and 89 reactions (the final version of the SA sub-model is shown in Figure 24).

During the revision of the SA sub-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 a single node. Also, since the number of biological molecules was reduced from the SA sub-model v2.0 to the SA sub-model v3.0, the number of constraints consequently decreased from 33 in step 2 to 29 in step 3. The following data represent the results of the third optimisation experiment performed on the SA sub-model v3.0. Out of 89 parameters, 71 were optimised, while the other ones are left out from optimisation since they were representing initial values for the reactions. The final values for individual constraints and the overall value of the criterion function are shown:

```

equalRate(Prephenate,Phenylpyruvate) = 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(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(Phenylpyruvate) = 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 above set of 29 constraints that we defined in the last iterative set is referred to as the *training constraint set*. Table 3 lists values of kinetic parameters optimised with our approach. Note that the list of these reactions is partially comparable with the reactions of the SA sub-models v1.0 (see material available online at [http://kt.ijs.si/dragana\\_miljkovic/](http://kt.ijs.si/dragana_miljkovic/)) and therefore we represent here only the final results. The total criterion function in this step showed that on average 0.043 of each constraint is not satisfied. In comparison with the first and the second step, the satisfaction of constraints was improved in the final step. Additionally, according to the experts' judgement, the curves output of step 3 were better compared to v1.0 and v2.0. The final curves of the same four molecules (SA, EDS1, PR1 and NPR1) are shown in Figure 25. Detailed data for all three iterative optimisation steps are compared in Table 4.

We joined expert knowledge with the mathematical approach to understand better the kinetic behaviour of the SA pathway, which is one the most important pathways in plant

defence. Our approach of defining an objective function consisting of domain knowledge (constraints) represents a step forward and is less time-consuming compared to the hand-tuning of model parameters. The methodology for defining constraints is similar to the one of (Locke et al., 2005). However, our approach differs from it since it is totally based on the domain knowledge, whereas the study mentioned above relies on the combination of the experimental data and expert knowledge. The only available dataset was used to give an additional comparison value to obtain a certain criterion of a model reality.

The selected parameter set is large, thus, making the search space enormous. This automatically directed us to use some stochastic optimisation methods (i.e., the DE algorithm) since the deterministic methods are overly time-consuming in the case of a large parameter set. Interesting results were obtained using our evaluation method, albeit some limitations remain. Our method is based on the subjective knowledge of domain experts, and not on the explicit and objective numerical experimental data. Nevertheless, this knowledge is valuable and very useful for guiding the sub-model construction. However, we believe that the comparison of simulation results even with the experimental dataset will provide additional confidence to the expert's validation of the SA sub-model.

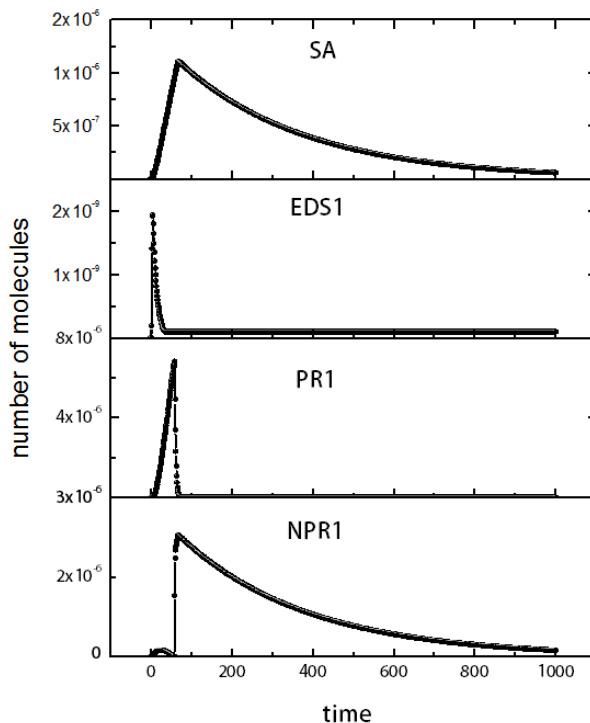


Figure 25: Dynamic behaviour of the SA, EDS1, PR1 and NPR1 variables based on the optimal parameter set estimated with respect to the objective function calculated from 29 constraints during step 3.



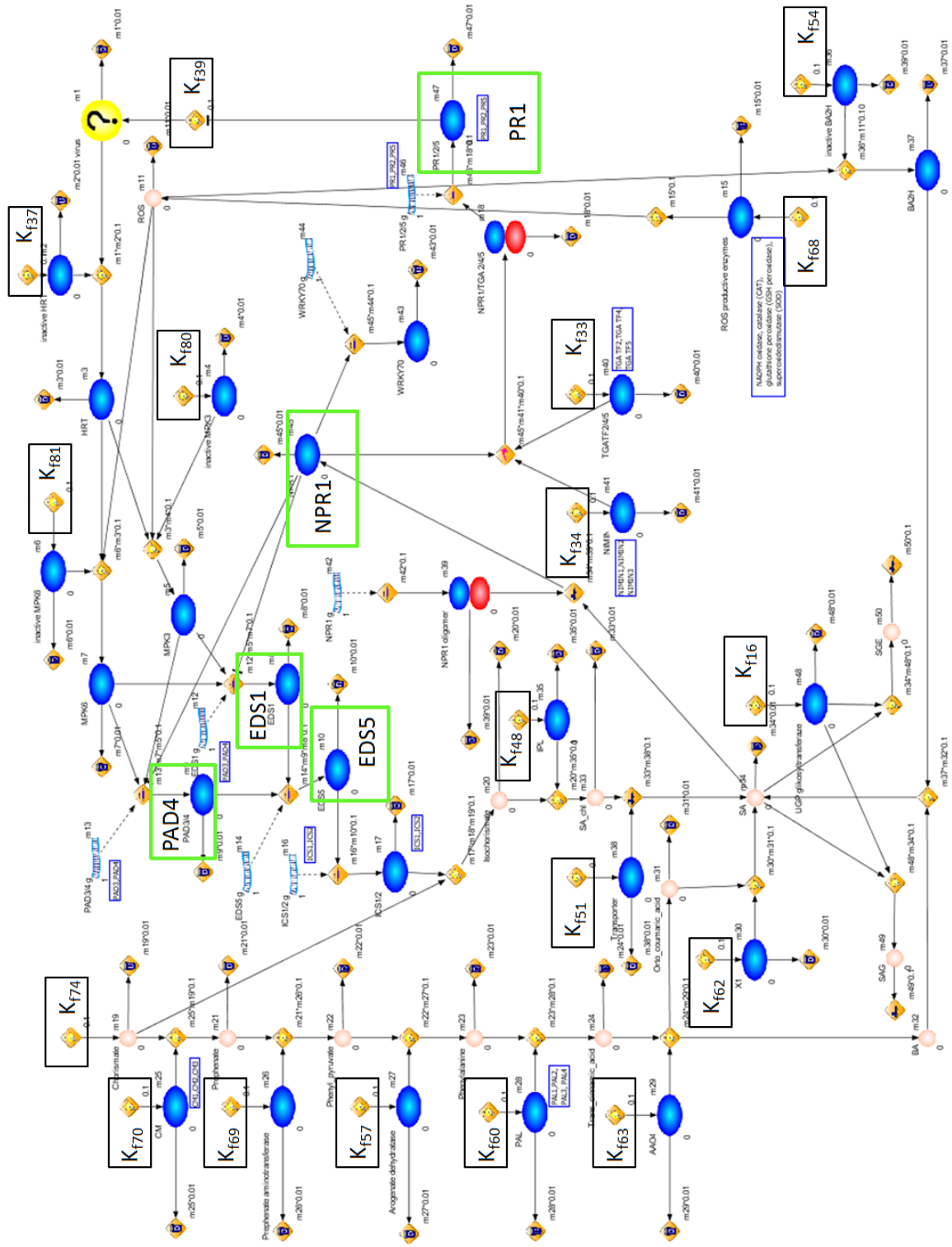


Figure 24: The biosynthesis and signalling pathway of the final SA sub-model v3.0 manually constructed in the CI. The molecules in the green squares were used for the model validation in Section 5.3, while the coefficients in the black squares were used for the local sensitivity analysis of the SA sub-model v3.0 in Subsection 5.3.2.

Table 3: A list of parameters of the SA sub-model v3.0 optimised with our approach.

Optimisation parameters	Parameter values	Optimisation parameters	Parameter values
$thr_m44_p15$	0.000535908		
$thr_m46_p4$	$6,35E-05$	$thr_m44_p2$	0.000395285
$k_{83}$	3.433	$k_{11}$	678.528
$k_{82}$	860.233	$k_{12}$	0.107657
$k_{81}$	893.532	$k_{13}$	0.1
$k_{80}$	0.1	$k_{19}$	165.033
$k_{84}$	14.614	$k_{14}$	0.1
$k_{77}$	530.164	$k_{15}$	212.282
$k_{78}$	0.338785	$k_{16}$	233.389
$k_{79}$	369.743	$k_{17}$	335.043
$k_{70}$	826.133	$k_{18}$	0.248453
$k_{72}$	339.172	$k_{10}$	0.261568
$k_{71}$	619.137	$k_{24}$	0.1
$k_{74}$	0.1	$k_{25}$	0.178708
$k_{73}$	954.421	$k_{22}$	720.195
$k_{76}$	167.399	$k_{23}$	832.662
$k_{75}$	57.601	$k_{28}$	0.1
$k_{68}$	0.1	$k_{29}$	0.1
$k_{69}$	192.478	$k_{26}$	0.1
$k_{66}$	241.386	$k_{27}$	360.421
$k_{67}$	0.380435	$k_{21}$	168.271
$k_{65}$	153.045	$k_{20}$	280.847
$k_{64}$	159.237	$k_{37}$	110.793
$k_{63}$	0.900479	$k_{38}$	0.1
$k_{62}$	0.211873	$k_{39}$	0.1
$k_{61}$	0.773124	$k_{33}$	0.557766
$k_{60}$	0.41851	$k_{34}$	0.1
$k_{59}$	0.300204	$k_{35}$	100
$k_{55}$	193.299	$k_{36}$	0.463225
$k_{56}$	11.727	$k_{30}$	0.1
$k_{57}$	981.612	$k_{32}$	0.1
$k_{58}$	133.915	$k_{31}$	288.793
$k_{52}$	52.932	$k_{48}$	0.1
$k_{51}$	0.624458	$k_{49}$	502.036
$k_{54}$	0.1	$k_{46}$	199.152
$k_{53}$	0.1	$k_{47}$	0.1
$k_{50}$	0.1	$k_{44}$	0.649051
$k_7$	228.442	$k_{45}$	0.1
$k_6$	0.1	$k_{43}$	0.1
$k_5$	238.858	$k_8$	0.503158
$k_4$	0.1	$k_{42}$	475.817
$k_3$	100	$k_9$	0.154922
$k_2$	0.833467	$k_{41}$	106.128
$k_1$	260.244	$k_{40}$	426.341
$k_{85}$	0.104416	$k_{86}$	0.420679

Table 4: A comparison of the three iterative optimization steps.

Version of the SA sub-model	Components	Reactions	Constraints	Averaged objective function
v1.0	52	38	8	0.047
v2.0	61	56	33	0.144
v3.0	50	89	29	0.043

### 5.3 Model validation

This section describes the most common validation techniques of simulation models. We specifically focus on the model validation approaches where no or few experimental data are available. The section provides the results of the SA sub-model validation by applying two validation approaches: sensitivity analysis, which is a common approach, and the developed approach that we propose in this thesis.

#### 5.3.1 Concepts and strategies

Model validation is a crucial step in the development of a simulation model. We distinguish between model verification, validation and accreditation phases. *Model verification* is concerned with determining whether the model is implemented correctly. *Model validation* estimates to which extent model simulation outputs represent a real-life system from the perspective of the expected model use. *Model accreditation* is related to the model credibility which deals with the confidence that the users have in the results of model simulation. The topic of this section is model validation.

A model is always developed with a specific purpose. Therefore, its validity should be determined concerning that purpose. For example, if a model has to be used in several conditions, the model validity has to be determined for every predefined condition. A model is valid if its accuracy is within a certain range that is required from the perspective of the intended use of the model. This accuracy range is usually defined during the early model construction phase.

The problem with the model validation is the lack of standard validity tests. There are no rules how to determine whether a model is "good enough". There is no "standard" theory or software tools available to support the validation process. However, the literature agrees that the decision whether a model is "good enough" should be a compromise accomplished between the modeller, the user and the code developer.

There are several validation techniques which can be applied in different situations. A study of (Sargent, 1998) describes different existing validation techniques. We briefly present the ones which are mostly used.

- **Animation.** The model's behaviour through time can be presented graphically.
- **Comparison to other models.** The outputs of the model, which needs to be validated, can be compared to the already existing validated models. These models can be simple cases of the developed model.
- **Extreme condition tests.** The model structure and its dynamic outputs should correspond to any extreme combination of the model parameters.
- **Face validity.** No matter whether there is an existing model for comparison (or experimental data), a domain expert should review simulation results to ensure they are reasonable. This kind of validation is used to ascertain that the simulation results are consistent with the expected system behaviour.

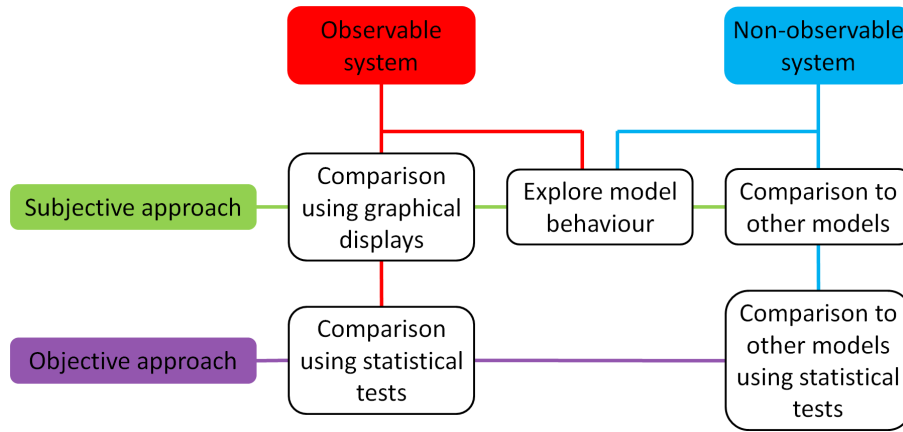


Figure 26: The classification of validation approaches for the observable and non-observable system according to (Sargent, 1998).

- **Validation by using experimental data.** If the experimental data exist, then one part is usually used to construct the model while the other smaller set is left for the model validation. This is the most common way to validate the model for which most of the statistical tests are developed.
- **Parameter variability-sensitivity analysis.** In this approach, the input values and values of internal parameters are changed to observe their influence on the simulation outputs. The corresponding changes should occur also in the real-life system.
- **Predictive validation.** The simulation model is used to predict the system behaviour under certain conditions and the predicted values are compared with the existing data.
- **Turing tests.** The domain experts are tested to distinguish between the simulation and the real-life system results.

For most of the validation techniques described above, apart from the animation and face validity, there are statistical methods available. The most influential criterion when selecting the validation approach is the system's observability (Sargent, 1998). Under observability of a system we denote the possibility to collect data on the dynamic behaviour of certain variables. Figure 26 presents the classification of subjective and objective validation approaches depending on the system's observability. "Comparison" as stated in Figure 26 means comparing/testing the simulation and the system's outputs. Comparison by using graphical displays can be performed in various ways. For example, the experts can use the graphs of simulation outputs for the subjective evaluation of the model's accuracy during the model's development or after the model construction. Also, the graphical display approach can be used in Turing tests, where the experts are asked to distinguish between the simulation and real-life system outputs. Exploring model's behaviour represents the investigation of simulation outputs using some validation techniques mentioned previously, like, for example, sensitivity analysis.

When the system is not observable (no experimental data are available), strict validation of the model is not feasible. However, if there is still domain knowledge available, a model can be validated with respect to that knowledge. If the dynamic behaviour of the model violates this knowledge, the model has to be questioned and errors should be investigated. Also, in this case, sensitivity analysis can be performed. Sensitivity analysis is defined as the investigation of the model simulation outputs when the input model parameters are changed. These changes can be either extreme or marginal. Sensitivity analysis can support the validation process in the sense whether the change of parameter values will have the effect that corresponds to the experts' expectations. Furthermore, the sensitivity analysis

might show which parameters are crucial for certain model behaviour. For example, the study of (Chu et al., 2007) explores the importance of various reaction rate coefficients in IL-6 signalling pathway by using sensitivity analysis.

When the system is observable, we can distinguish between two cases:

- (i) only output data are available, and
- (ii) input and output data are available.

For case (i), simple two-sample Student t-tests can be used if the data are normally distributed. If the data are not normally distributed, then the distribution-free tests can be applied, like rank tests. In case (ii), when both input and output experimental data are available, regression analysis can be applied. Regression analysis estimates the influence of one or more input variables to the output variable of interest.

For cases (i) and (ii), when the datasets available are big enough for model building and model testing, usually the proportion of training versus testing data is 80 %/20 % or 90 %/10 %. On the other hand, if the data amount is limited, i.e. there are not enough samples for training and testing separately, a k-fold cross-validation can be performed (Kohavi, 1995). In k-fold cross-validation, the dataset is divided into k equally-sized segments. Validation iterations are performed in such a way that during each of k iterations, a different segment is kept for the validation, where the other k-1 segments are used for building (training) the model. In order to avoid biases, during every iteration all steps of model building (training) should be performed independently (like model selection, feature selection, etc.). As an example, Figure 27 illustrates the three-fold cross-validation process. In machine learning and data mining, the most common is 10-fold cross-validation.

Having in mind that one experimental dataset with five time-points is available for validating the PDS model<sup>1</sup>, we cannot claim that the system, which we model, is non-observable. On the other hand, one dataset is not enough for reliable model validation by applying statistical approaches which are well established for validating the models of observable systems. For these reasons, we had to balance between the validation techniques for observable and non-observable systems. We have applied two subjective validation approaches for the non-observable systems and we have developed a new validation method, which requires few experimental data and is based on constraint generation approach. We briefly list below the validation approaches used in this work to validate the SA sub-model.

- (a) Face validity. We used this validation approach mostly during the model development phase. The details of iterative steps for the determination of the most optimal parameters of the SA sub-model are presented in the previous section.
- (b) Sensitivity analysis. This method is used in combination with the face validity approach where the domain experts provided feedback on the general results of the sensitivity analysis. The analysis is performed on the SA sub-model v3.0. In addition to the face validity of the sensitivity analysis results, we have explored the importance of the predefined model input parameters and their influence to the outputs of interest.
- (c) New validation method based on the k-fold cross-validation approach. This method is one of the contributions of this thesis. It can be applied only for the observable systems when at least one experimental dataset is available. To validate the SA sub-model v3.0 we have used one dataset with five time points, which is publicly available.

In the next two subsections we represent and discuss the results of validation methods (b) and (c), respectively.

---

<sup>1</sup>Apart from this dataset, there are other publicly available datasets, which are acquired during the interaction of *Arabidopsis thaliana* and viruses. However, these datasets do not have sufficient information on the PDS mechanism dynamics, i.e. they contain two time-points measured before and after the virus attack.

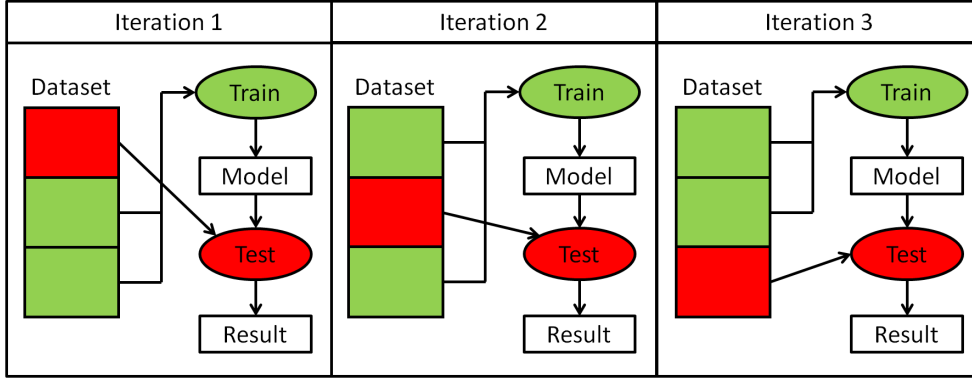


Figure 27: The iterative process of three-fold cross-validation.

### 5.3.2 Sensitivity analysis: exploring the robustness of the SA sub-model

The role of sensitivity analysis in dynamic analysis of biological models is manifold. First, based on the sensitivity analysis, the robustness of the system can be explored. If the changes of the model inputs do not affect the model outputs significantly, we can claim that the system is robust. In addition, the input perturbations can determine the key inputs that influence the model output. Then, sensitivity analysis can be used in model reduction, guiding experimental analysis and parameter estimation. For the non-observable systems, the parametric sensitivity analysis presents model validation. In that case, the produced outputs that are the result of the changes in the model inputs should correspond to the expectation of the domain experts. Thus, the parametric sensitivity analysis combined with the face validity technique represents a way to validate the models of non-observable systems.

There are numerous ways to perform sensitivity analysis. The main two approaches are:

- (a) local sensitivity analysis, and
- (b) global sensitivity analysis.

Local sensitivity analysis investigates how the small perturbations in input values influence the model outputs. When the size of perturbations is slowly increased and still the sensitivity does not change significantly, then the local analysis results are considered robust. Global sensitivity analysis, on the other hand, deals with the larger parameter changes and investigates the whole parameter space. For a non-observable system, if the domain knowledge on the system's behaviour for the extreme parameter changes is not available, local sensitivity analysis can be used. This is the case with the PDS mechanism for which we have performed local sensitivity analysis.

If we consider our PDS model represented with the HFPN formalism (which is practically a system of ODEs with graphical representation), we can write the equations with a given parameter set  $p$  and initial conditions  $y_i(0)$  in the following form:

$$\frac{dy_i}{dt} = f_i(y_i, p, t), \quad i = 1, 2, \dots, n \quad (15)$$

Mathematical presentation of the local sensitivity coefficients has the form of first-order derivatives of the model output variables with respect to the model parameters:

$$S_i = \frac{\partial y_i}{\partial p} = \lim_{\Delta p \rightarrow 0} \frac{y_i(t; p + \Delta p) - y_i(t; p)}{\Delta p} \quad (16)$$

Numerous methods are used to calculate the first-order derivatives. The most simple method, which we have also used in our work, is the method of finite difference approximation. The sensitivity, calculated according to this method, has the following form:

$$S_i = \frac{\partial y_i}{\partial p} \simeq \frac{y_i(p + \Delta p) - y_i(p)}{\Delta p} \quad (17)$$

Sensitivity is usually a function of time and is merged into a vector of sensitivities at different time points. A sensitivity coefficient is the Euclidean norm (or 2-norm) of  $S_i(t)$  (Chu et al., 2007). Thus, the sensitivity coefficient represents the impact of the input perturbation on the output value in given period of time.

### Setting the stage for sensitivity analysis

To perform the model validation, one has to determine the observable model variables. To explore the robustness of the SA sub-model v3.0, we have used the local sensitivity analysis. The input parameters, perturbed for the sensitivity analysis, are the rate coefficients of 18 biological reactions, marked in Figure 24, which are responsible for the stimulation of the SA pathway. The outputs of interest, defined by the biology experts, are the concentrations of four biological components (proteins): EDS1, PAD4, NPR1 and PR1 (also marked in Figure 24). We apply three different perturbation step sizes ( $\Delta p = 0.001 * p$ ;  $\Delta p = 0.01 * p$  and  $\Delta p = 0.1 * p$ , where  $p$  denotes an input parameter) and observe their effects on the concentration values of four output components. The perturbation of one rate coefficient is performed at a time, while the other input coefficients are kept constant on their nominal values. The duration of the model simulation for each input perturbation is set to 1,000 simulation points.

### Dynamic effects of the input parameters

At this stage we illustrate and discuss different individual effects of the particular input rate coefficients on the output values. We compare the effects on the output variables when different perturbation step sizes of inputs are applied ( $\Delta p = 0.001 * p$ ;  $\Delta p = 0.01 * p$  and  $\Delta p = 0.1 * p$ ). These results are presented in the form of the sensitivity vector norms (normalised by the largest one), which are sorted in Tables 5, 6, 7 and 8 from the most to the least important. The sensitivity coefficients (Tables 5, 6, 7 and 8) show the total effect of input perturbation on the outputs. The tables show that changes of only few out of 18 predefined input rate coefficients influence the sensitivity profiles of the output components. The input rate coefficients have different impacts on the output profiles when the perturbation size is changed. The concentration of the EDS1 protein (Table 5) is the most sensitive to the changes of rate coefficients  $k_{f37}$ ,  $k_{f33}$ ,  $k_{f57}$  and  $k_{f60}$  as denoted in Figure 24. The EDS1 time-dependent behaviour, shown in Figure 28, is a result of changes in the most important rate coefficients:  $k_{f37}$ ,  $k_{f33}$ ,  $k_{f57}$ . The coefficient  $k_{f60}$  was omitted since it is important only when the perturbation size is  $\Delta p = 0.1 * p$ , while for the other two perturbation sizes, it does not have any impact on the dynamics of the EDS1 protein. An interesting result of the sensitivity analysis is the influence of the rate coefficient  $k_{f57}$  on the EDS1 dynamic behaviour. Even though the reaction with this coefficient is further from EDS1 (from the perspective of the graph structure), there is an evident influence on its dynamic output. This influence is indirect over the negative loop where the NPR1 component blocks the production of the EDS1 after a certain threshold is achieved. The EDS1 curve changes also the shape when perturbations of  $k_{f57}$  are performed, while the trend in EDS1 dynamic behaviour stays similar when the other two coefficients,  $k_{f37}$  and  $k_{f33}$ , are perturbed. The influence of this reaction on the final performance of the PDS mechanism might be interesting for further investigations by the biology experts.

In a similar way we have inspected and plotted the impacts of the most influential parameters on the output variables PAD3/4, NPR1 and PR1 proteins. The largest impact on the PAD3/4 dynamics have the rate coefficients  $k_{f37}$ ,  $k_{f33}$  and  $k_{f63}$  (Figure 29). The

components NPR1 and PR1 are sensitive to the changes in  $k_{f16}$ ,  $k_{f48}$  and  $k_{f51}$  coefficients (Figures 30 and 31).

The domain experts have confirmed that the curves of the EDS1, PAD3/4, NPR1 and PR1 components have the expected behaviour. The shapes of the curves do not change considerably (apart from the interesting behaviour of EDS1 as a function of changes in the  $k_{f57}$  coefficient), but only the order of magnitude of the output concentration profiles. However, this change in the output concentration for the order of magnitude does not have to necessarily influence the final outcome of the PDS model, since the results of the PDS model are interpreted by the domain experts in a qualitative way (by comparing the dynamics of the curves with each other and their shapes). Moreover, having in mind that the Cell Illustrator software implements low order integration method (Euler method) the numerical results of dynamic simulation contain big error. Therefore, when interpreting the results sorted in Tables 5, 6, 7 and 8, we are interested in the relative importance of particular coefficients compared to the other ones that appear non-relevant in this analysis.

Preliminary results of local sensitivity analysis show the potential of detecting the most influential reactions of the PDS model. This kind of analysis can be used in the design of wet-lab experiments with the goal to enlarge the knowledge on the dynamics of the PDS mechanism. There is also space for different kinds of model sensitivity analyses, which we plan to perform further on the whole PDS model. For example, it might be interesting to investigate the behaviour of the model outputs when two or more input parameters are changed simultaneously. During further collaboration with the domain experts we hope to define the pairs of interesting input parameters and perform more research on model sensitivity in this direction.

Table 5: The summary of local sensitivity values calculated for the EDS1 component.

No.	EDS1					
	$\Delta p = 0.001 * p$		$\Delta p = 0.01 * p$		$\Delta p = 0.1 * p$	
1	$k_{f37}$	1	$k_{f37}$	1	$k_{f60}$	1
2	$k_{f57}$	0.896	$k_{f33}$	0.503	$k_{f33}$	0.259
3	$k_{f33}$	0.506	$k_{f57}$	0.369	$k_{f37}$	0.183
4	$k_{f68}$	0.086	$k_{f68}$	0.099	$k_{f57}$	0.176
5	$k_{f34}$	0.055	$k_{f34}$	0.064	$k_{f68}$	0.044
6	$k_{f16}$	0	$k_{f16}$	0	$k_{f34}$	0.028
7	$k_{f39}$	0	$k_{f39}$	0	$k_{f63}$	0
8	$k_{f48}$	0	$k_{f48}$	0	$k_{f16}$	0
9	$k_{f51}$	0	$k_{f51}$	0	$k_{f39}$	0
10	$k_{f54}$	0	$k_{f54}$	0	$k_{f48}$	0
11	$k_{f60}$	0	$k_{f60}$	0	$k_{f51}$	0
12	$k_{f62}$	0	$k_{f62}$	0	$k_{f54}$	0
13	$k_{f63}$	0	$k_{f63}$	0	$k_{f62}$	0
14	$k_{f69}$	0	$k_{f69}$	0	$k_{f69}$	0
15	$k_{f70}$	0	$k_{f70}$	0	$k_{f70}$	0
16	$k_{f74}$	0	$k_{f74}$	0	$k_{f74}$	0
17	$k_{f80}$	0	$k_{f80}$	0	$k_{f80}$	0
18	$k_{f81}$	0	$k_{f81}$	0	$k_{f81}$	0



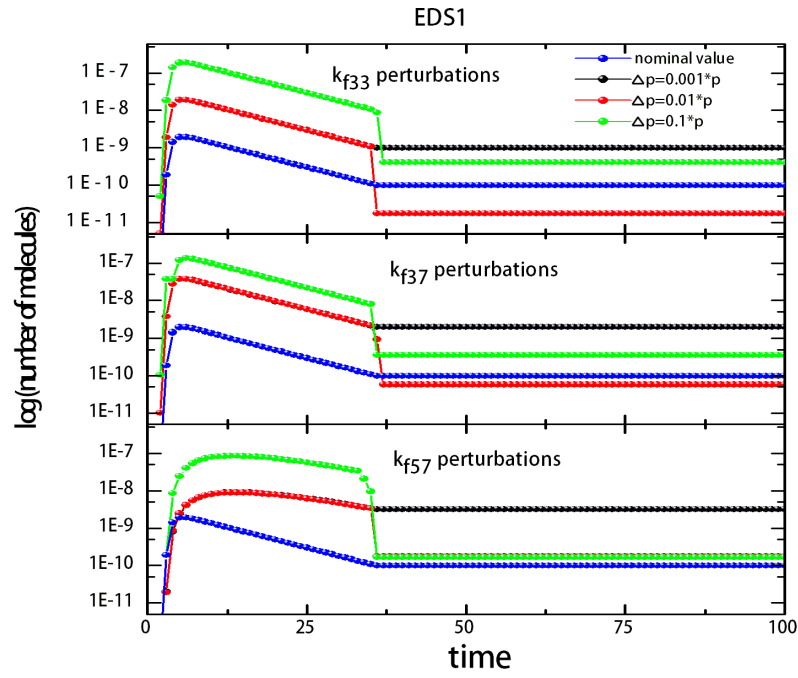


Figure 28: The influence of perturbations of rate coefficients  $k_{f33}$ ,  $k_{f37}$  and  $k_{f57}$  on the dynamic profile of the EDS1 protein.

Table 6: The summary of local sensitivity values calculated for the PAD3/4 component.

No.	PAD3/4					
	$\Delta p = 0.001 * p$		$\Delta p = 0.01 * p$		$\Delta p = 0.1 * p$	
1	$k_{f37}$	1	$k_{f37}$	1	$k_{f60}$	1
2	$k_{f63}$	0.510	$k_{f63}$	0.507	$k_{f63}$	0.257
3	$k_{f33}$	0.506	$k_{f33}$	0.503	$k_{f33}$	0.255
4	$k_{f68}$	0.084	$k_{f68}$	0.084	$k_{f37}$	0.181
5	$k_{f34}$	0.054	$k_{f34}$	0.054	$k_{f68}$	0.042
6	$k_{f16}$	0	$k_{f16}$	0	$k_{f34}$	0.027
7	$k_{f39}$	0	$k_{f39}$	0	$k_{f16}$	0
8	$k_{f48}$	0	$k_{f48}$	0	$k_{f39}$	0
9	$k_{f51}$	0	$k_{f51}$	0	$k_{f48}$	0
10	$k_{f54}$	0	$k_{f54}$	0	$k_{f51}$	0
11	$k_{f57}$	0	$k_{f57}$	0	$k_{f54}$	0
12	$k_{f60}$	0	$k_{f60}$	0	$k_{f57}$	0
13	$k_{f62}$	0	$k_{f62}$	0	$k_{f62}$	0
14	$k_{f69}$	0	$k_{f69}$	0	$k_{f69}$	0
15	$k_{f70}$	0	$k_{f70}$	0	$k_{f70}$	0
16	$k_{f74}$	0	$k_{f74}$	0	$k_{f74}$	0
17	$k_{f80}$	0	$k_{f80}$	0	$k_{f80}$	0
18	$k_{f81}$	0	$k_{f81}$	0	$k_{f81}$	0

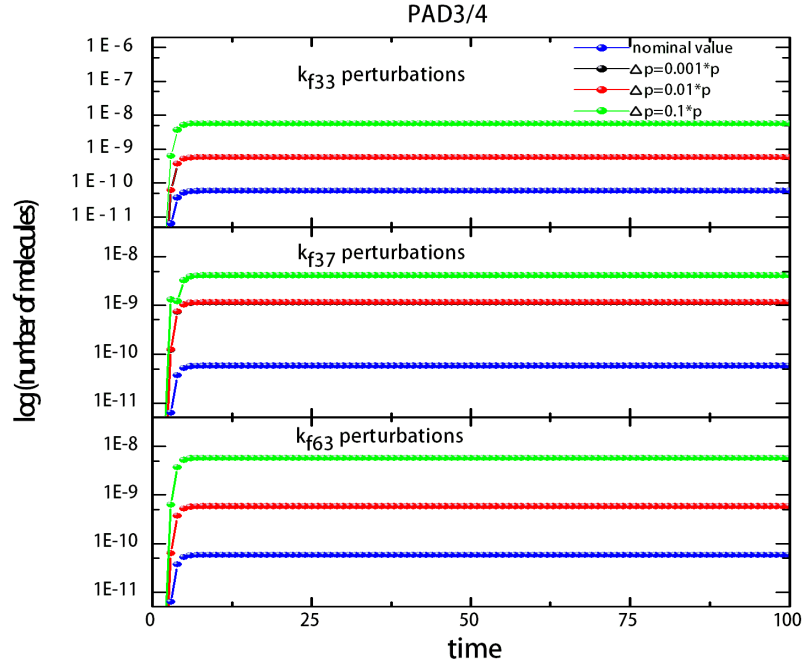


Figure 29: The influence of perturbations of rate coefficients  $k_{f33}$ ,  $k_{f37}$  and  $k_{f63}$  on the dynamic profile of the PAD3/4 protein.

Table 7: The summary of local sensitivity values calculated for the NPR1 component.

No.	NPR1					
	$\Delta p = 0.001 * p$		$\Delta p = 0.01 * p$		$\Delta p = 0.1 * p$	
1	$k_{f51}$	1	$k_{f51}$	1	$k_{f51}$	1
2	$k_{f16}$	0.775	$k_{f16}$	0.769	$k_{f16}$	0.728
3	$k_{f48}$	0.748	$k_{f48}$	0.743	$k_{f48}$	0.699
4	$k_{f39}$	0.003	$k_{f39}$	0.002	$k_{f57}$	0.002
5	$k_{f33}$	0.002	$k_{f34}$	0.001	$k_{f37}$	0.001
6	$k_{f37}$	0.002	$k_{f33}$	0.001	$k_{f33}$	0.001
7	$k_{f60}$	0.002	$k_{f37}$	0.001	$k_{f60}$	0.001
8	$k_{f34}$	0.002	$k_{f60}$	0.001	$k_{f34}$	0.001
9	$k_{f68}$	0.001	$k_{f68}$	0.001	$k_{f68}$	0
10	$k_{f54}$	0	$k_{f54}$	0	$k_{f54}$	0
11	$k_{f57}$	0	$k_{f57}$	0	$k_{f57}$	0
12	$k_{f62}$	0	$k_{f62}$	0	$k_{f62}$	0
13	$k_{f63}$	0	$k_{f63}$	0	$k_{f63}$	0
14	$k_{f69}$	0	$k_{f69}$	0	$k_{f69}$	0
15	$k_{f70}$	0	$k_{f70}$	0	$k_{f70}$	0
16	$k_{f74}$	0	$k_{f74}$	0	$k_{f74}$	0
17	$k_{f80}$	0	$k_{f80}$	0	$k_{f80}$	0
18	$k_{f81}$	0	$k_{f81}$	0	$k_{f81}$	0

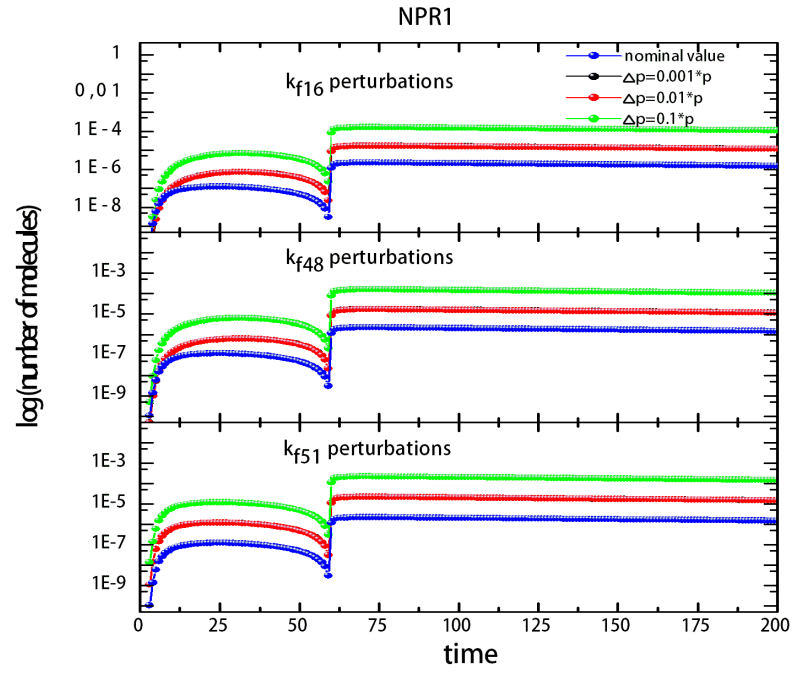


Figure 30: The influence of perturbations of rate coefficients  $k_{f16}$ ,  $k_{f48}$  and  $k_{f51}$  on the dynamic profile of the NPR1 protein.

Table 8: The summary of local sensitivity values calculated for the PR1 component.

No.	PR1					
	$\Delta p = 0.001 * p$		$\Delta p = 0.01 * p$		$\Delta p = 0.1 * p$	
1	$k_{f51}$	1	$k_{f51}$	1	$k_{f51}$	1
2	$k_{f16}$	0.717	$k_{f16}$	0.713	$k_{f16}$	0.679
3	$k_{f48}$	0.664	$k_{f48}$	0.659	$k_{f48}$	0.626
4	$k_{f39}$	0.017	$k_{f39}$	0.017	$k_{f39}$	0.017
5	$k_{f60}$	0	$k_{f37}$	0	$k_{f33}$	0
6	$k_{f33}$	0	$k_{f33}$	0	$k_{f60}$	0
7	$k_{f34}$	0	$k_{f34}$	0	$k_{f34}$	0
8	$k_{f37}$	0	$k_{f60}$	0	$k_{f37}$	0
9	$k_{f54}$	0	$k_{f68}$	0	$k_{f68}$	0
10	$k_{f57}$	0	$k_{f54}$	0	$k_{f54}$	0
11	$k_{f62}$	0	$k_{f57}$	0	$k_{f57}$	0
12	$k_{f63}$	0	$k_{f62}$	0	$k_{f62}$	0
13	$k_{f68}$	0	$k_{f63}$	0	$k_{f63}$	0
14	$k_{f69}$	0	$k_{f69}$	0	$k_{f69}$	0
15	$k_{f70}$	0	$k_{f70}$	0	$k_{f70}$	0
16	$k_{f74}$	0	$k_{f74}$	0	$k_{f74}$	0
17	$k_{f80}$	0	$k_{f80}$	0	$k_{f80}$	0
18	$k_{f81}$	0	$k_{f81}$	0	$k_{f81}$	0

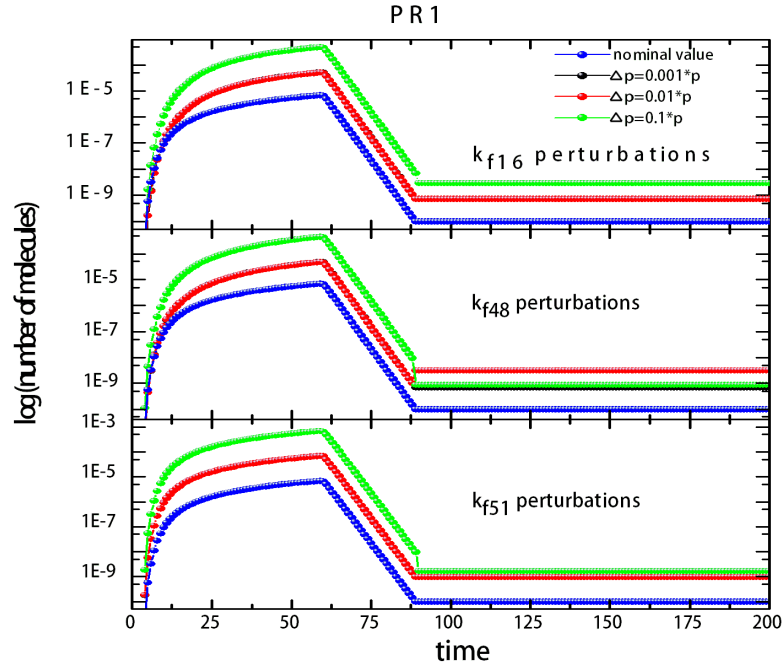


Figure 31: The influence of perturbations of rate coefficients  $k_{f16}$ ,  $k_{f48}$  and  $k_{f51}$  on the dynamic profile of the PR1 protein.

### 5.3.3 Comparison with experimental data

In this section, we present the novel validation approach, which is one of the contributions of this thesis. The main idea of this approach is based on the generation of new constraints from the experimental data for which it is necessary to have at least one experimental dataset. This approach is used to validate the SA sub-model v3.0 by using one publicly available experimental dataset. We start this section by describing the dataset, then we propose the new method and we conclude with the evaluation results.

#### Dataset

A dataset used for the validation of the SA sub-model v3.0 is published as Supplementary Table S2 in the study of (Yang et al., 2007). This study reveals the spatial and temporal behaviour of the *Arabidopsis thaliana* in response to three different inoculation types: *Turnip mosaic virus* (TuMV), *Oilseed rape mosaic virus* (ORMV) and mechanical inoculation (mock) over a 10-day time course. One of their goals was the induction of defence-related genes, where the set of 388 preselected genes was used due to the large number of samples. To examine the reaction to stress in systemic tissues, 25 *Arabidopsis thaliana* plants were used for each inoculation type (TuMV, ORMV and mock) and each of these plants had four fully developed rosette leaves inoculated. Samples were collected from five plants at each of 2, 3, 5, 7 and 10th day after inoculation of each inoculation type. The data collection process was reproduced three more times with different sets of plants. There were 144 samples analysed in total (three replications x three inoculation types x 16 time-tissue combinations). All 16 time-tissue combinations that were sampled are shown in Table 9.

Table 9: Sample types used in the study of (Yang et al., 2007) to investigate *Arabidopsis thaliana* responses to TuMV and ORMV. The "..." in the table denotes no data collected on the particular day, while "x" denotes successfully collected sample on the particular day.

Tissue	Days after inoculation				
	2	3	5	7	10
Inoculated leaf	x	x	x	x	x
Systemic rosette leaf	...	x	x	x	x
Cauline leaf	...		x	x	x
Flowers	...	x	x	x	x

We have used the time-series data only for inoculated leaves. For inoculated leaves, there were samples for 5 days for three replications (see Table 9), which were averaged per replication. The averaged data were then normalised according to the standard procedure for the microarray data.

The goal of normalisation of expression microarrays is a compensation for systematic technical differences between chips which allows for a clearer view of systematic biological differences between samples. The reason for applying normalisation is to avoid the technical variations between experimental conditions which are entirely unrelated to the differences that biologists search for. The basic assumption which allows normalisation of microarray data is that only a few hundred genes (out of tens of thousands) are expressed at different levels between several samples. A more specific assumption is that microarray measures should not be correlated with technical characteristics of probes. This implies that, according to our assumption, biological changes would be independent of technical characteristics of expression probes.

The starting assumption of most statistical approaches to normalizing expression levels is that the overall distribution of RNA levels does not change significantly between samples or across the conditions which seems reasonable for most laboratory treatments. The crucial decision researchers must make, with consequences for normalization, is on what scale to analyse their data. It is common to transform to a logarithmic (usually base 2) scale in order to make variation comparable among measures which extend over several orders of magnitude. Usually this kind of normalisation is enough, however, such conversion can increase also low intensity probes compared to the rest. Especially when a measure can be reported as zero, the logarithm is not defined. A simple solution is to add a small constant to the measures before taking the logarithm. In our work we have normalised data by transforming them to a logarithmic scale with base 2 without adding a small constant, since the measures were not reported as zero. Moreover, we have divided the expression values for the TuMV and ORMV virus with the expression values for mock inoculation to preserve only effects of these two viruses. Out of expressions of 388 genes, we have selected the ones that are in the SA v3.0 sub-model, which is 7 genes in total. The curves of time-series of the selected components processed in the way described above are shown in Figure 32 for TuMV and in Figure 33 for ORMV inoculation type.

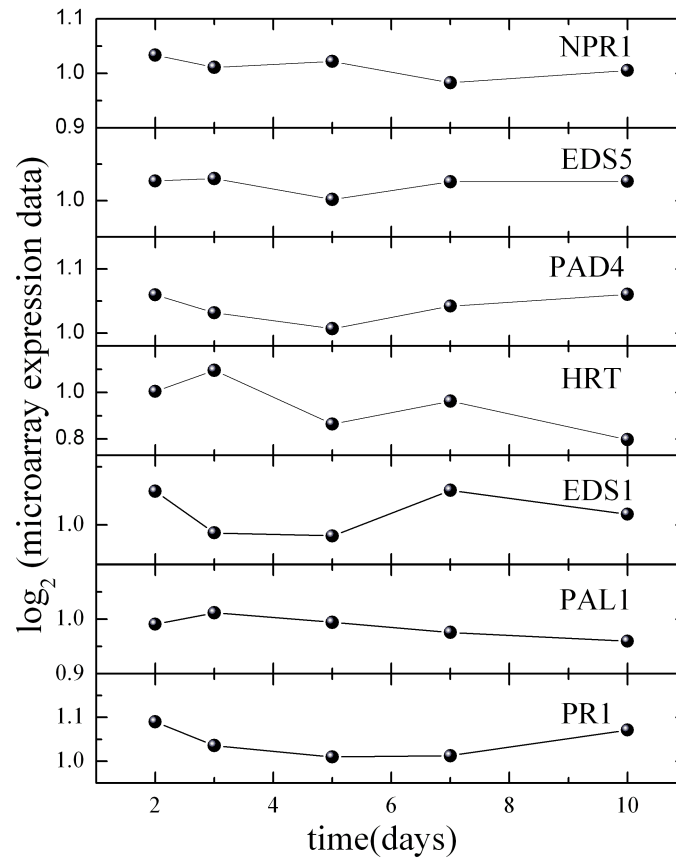


Figure 32: Processed and normalised microarray expression dataset for TuMV inoculation type for 7 components of the SA v3.0 sub-model.

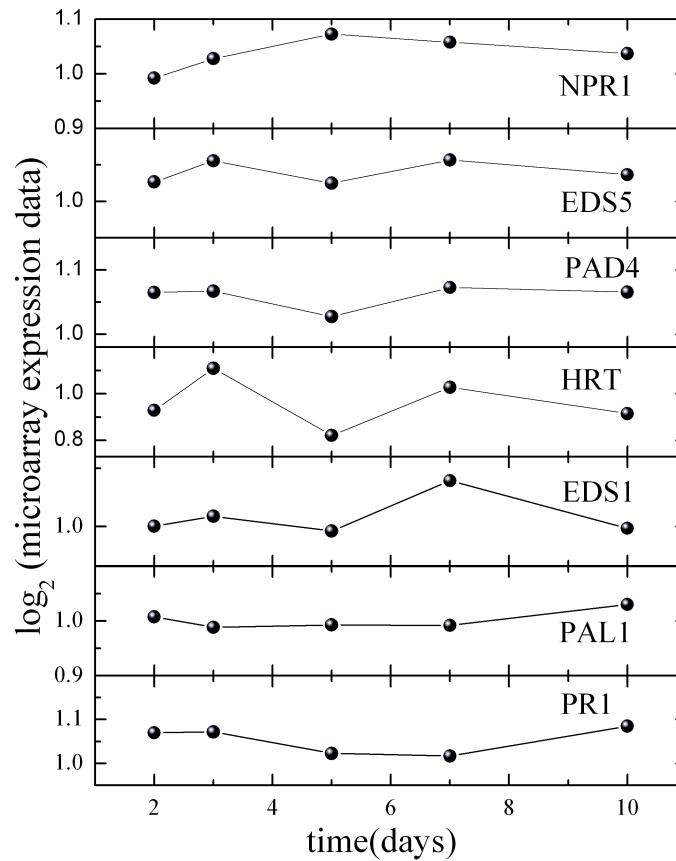


Figure 33: Processed and normalised microarray expression dataset for ORMV inoculation type for 7 components of the SA v3.0 sub-model.

## The validation method and validation results

The curves of Figures 32 and 33 were used as a base for our validation method. From Figure 32 addressing the problem of inoculation with TuMV, the following constraints were generated based on the experts' opinion. The consultation with the domain experts resulted in several additional variants of constraint types, such as obtained minimum at the same time between component1 and component2 ( $\text{minSameTime}(\text{component1}, \text{component2})$ ) or at the same time obtained maximum of component1 while the component2 is in its minimum value ( $\text{maxminSameTime}(\text{component1}, \text{component2})$ ). To the following constraint set we refer as the *evaluation set* for TuMV:

$\text{minSameTime}(\text{PAD3/4}, \text{EDS5})$ ,  
 $\text{minSameTime}(\text{EDS5}, \text{PR1/2/5})$  and  
 $\text{maxminSameTime}(\text{EDS1}, \text{NPR1})$ .

The constraints related to the inoculation with the ORMV and based on the curve shapes of Figure 33, are as follows:

$\text{minSameTime}(\text{PAD3/4}, \text{EDS5})$ ,  
 $\text{minAfter}(\text{EDS5}, \text{PR1/2/5})$ ,  
 $\text{maxAfter}(\text{NPR1}, \text{EDS1})$ ,  
 $\text{zeroPeakZero}(\text{NPR1})$ ,  
 $\text{maxminSameTime}(\text{NPR1}, \text{e.DS5})$  and  
 $\text{maxminSameTime}(\text{EDS1}, \text{PR1/2/5})$ .

To this constraint set we refer to as the *evaluation set* for ORMV. Based on these two evaluation constraint sets (for TuMV and ORMV), we have evaluated the SA v3.0 sub-model whose parameters were obtained through a combinatorial optimisation process guided by the set of validation constraints (experts' knowledge).

Table 10 represents the results of the validation method for the SA sub-model v3.0 where the best 10 optimisation results guided with the *evaluation constraint set* were evaluated (compared) with the experimental data for TuMV and ORMV inoculation.

Table 10: The results of the validation method based on the generation of constraints from the experimental datasets. The best 10 optimisation results guided with the *evaluation constraint set* are sorted in the second column starting from the best solution. The results of the second column are evaluated (compared) with the experimental data for TuMV (third column) and ORMV inoculation (forth column).

Nm.	Final criterion for <i>training</i> <i>constraint set</i>	Final criterion for <i>evaluation</i> <i>constraint set</i> for TuMV	Final criterion for <i>evaluation</i> <i>constraint set</i> for ORMV
1	0.043	0	0.095
2	0.044	0	0.05
3	0.045	0	0.002
4	0.046	0.316	0.158
5	0.053	0	0.071
6	0.065	0.316	0.158
7	0.0679	0	0.012
8	0.0681	0	0.012
9	0.0682	0.316	0.158
10	0.071	0	0.005

Generally, the PDS model structure is built for any kind of virus attack, which allows model validation with different experimental datasets. However, we observe in Table 10 that

the validation results of our SA v3.0 sub-model are different (different final criterion) when compared to TuMV and ORMV inoculation data. The main reason for this is the number of constraints in two *evaluation constraint sets*. When the number of constraints grows, the parameter search space becomes more and more narrow making the probability higher that some of these constraints will be violated to some extent. Also, there might be some small difference between plant defence reaction when TuMV and ORMV inoculation occurs. This would mean that some of the constraints in the constraint sets are contradictory. Further exploration of the constraint sets might reveal some interesting differences between the plant reactions to different viruses. The process of detecting contradictory links could be manual and automatic. The automatic system for detecting incompatible constraints will make the exploration of the validation results more efficient.



## 6 Model Structure Revision

*But what... is it good for?*

IBM executive Robert Lloyd, remarking on the microchip in 1968, the heart of today's computers.

The process of fusing expert knowledge and manually acquired information from the literature to build the PDS model structure turns out to be time-consuming. However, this manual process can be enhanced by automated methods of relation extraction from literature, which are recently popular in systems biology. The most common relation extraction methods from texts are based on natural language processing techniques. Ideally, the output of such relation extraction methods is a graph of biological components and relations between them.

This chapter is organised as follows. We present the developed Bio3graph tool that searches the biological literature for the relations between the biological components and outputs a graph of triplets in the form (*component1*, *relation*, *component2*). The power of the extracted triplets in modelling expert knowledge is demonstrated on the example of modelling knowledge of two domain experts. Moreover, the triplets, extracted by Bio3graph from a broader set of biological articles, are compared with the manually developed PDS model structure presented in Chapter 4 resulting in new, additional relations.

### 6.1 Bio3graph methodology

Manual construction of the PDS model structure is a time-consuming process, since only a limited amount of data is gathered in the available biological databases. The study of a large body of literature is therefore necessary in order to build a PDS network structure according to the most recent findings. The proposed Bio3graph methodology was developed with the purpose of automated information extraction from biological literature, aimed at complementing the manually developed PDS model structure.

An integral part of this methodology is a domain specific vocabulary that is composed of two parts: a list of components and a list of reactions together with their synonyms. The basis for the vocabulary was the list of 175 components and three reaction types defined when building the manual PDS model structure (see Figure 14). The components vocabulary consists of their short names, gene identifiers and synonyms, as annotated in TAIR (Swarbreck et al., 2008) and iHOP (Hoffmann and Valencia, 2004). As several components included in the manual PDS model structure are still not fully identified, they were labelled as X in the PDS model structure and were not included in the vocabulary. Moreover, in some cases, one biological component is represented with two nodes due to its compartmentalisation within the cell. For example, as SA accumulates in both the chloroplast and the cytosol, the node SA of the manually developed model, which stands for SA in cytosol and node SA-chl which stands for SA in chloroplast, are represented by the same component SA. In addition, most of the complexes were not included in the vocabulary except for the SCF complex. Consequently, the list of 153 biological components in the vocabulary (Section B.1 of Appendix B) contains fewer components than the vocabulary of 175 components at level 3 of Figure 11.A

used for manual PDS model structure construction. Furthermore, the vocabulary for the reaction types was developed, containing synonyms for the three reaction types: activation, inhibition and binding. Separate lists for each reaction in both the passive and the active verb form are available in Section B.2 of Appendix B.

The Bio3graph methodology consists of a series of text mining, information extraction, graph construction and graph visualisation steps, offering reusability, repeatability, and extension with additional components (Figure 34A). The name of the methodology, Bio3graph, reflects its main functionality: 'Bio3' stands for biological triplet extraction and 'graph' stands for graph construction from the extracted triplets.

- The first (NLP) part of this methodology (referred to as the triplet extraction algorithm below) concerns the extraction of relations in the triplet form (*subject, predicate, object*) thus searching for reactions between components as triplets (*component1, reaction, component2*) from publicly available biological texts by employing the above described manually developed vocabulary. Given the list of components, the algorithm detects subject and object, while the predicate represents the relation between the components as defined in the vocabulary of reaction types. For activation reaction type, an example triplet is (*PAD4, activates, EDS5*).
- The second part of this methodology concerns graph construction from the extracted triplets, and graph visualisation.

The methodology is implemented as a workflow in the Orange4WS (Podpečan et al., 2012) workflow construction and execution environment (Figure 34B). The input to the Bio3graph workflow is the collection of biological full text articles, obtained through a user-defined keyword-based search of the PMC database, accessible at: [www.ncbi.nlm.nih.gov/pmc/](http://www.ncbi.nlm.nih.gov/pmc/). The output of the workflow is a graph of triplets, automatically extracted from the articles. As an illustration, the graph consisting of all triplets extracted from the literature (consisting of 129 components and 1,132 reactions) is shown in Figure 35. The triplet extraction algorithm found relations between 129 out of 153 components listed in the vocabulary. If the Java plug-in for the web browser is installed and enabled, the reader can open and explore an interactive version of the Figure 35<sup>1</sup>.

For natural language processing we employed functions from the Natural Language Toolkit (NLTK) library (Bird et al., 2009). Additionally, the GENIA tagger (Tsuruoka et al., 2005) for biological domains was used to perform part-of-speech tagging and shallow parsing. The data were extracted from PMC using web service enabled access. Parts of the Bio3graph methodology, presented in Figure 34, are described in more detail below.

**Text pre-processing.** Full texts of scientific articles for biology domain are accessible in the publicly-available databases, such as PMC. Journal articles in the form of raw text need to be pre-processed. For example, in order to avoid a false detection of ET component by the algorithm, the phrase "et al." was transformed into "ETAL."

**Sentence splitting.** When the raw text is obtained with the previous module, the sentences are separated into lines. This step is necessary because the input into the Genia tagger module requires one sentence per line in the text file.

**Tokenization.** This is the process of splitting a sentence into words, phrases or other meaningful elements referred to as tokens. Tokenization is performed with the Genia tokenizer (Tsuruoka and Tsujii, 2005). The outputs of the tokenization process are tokens that are used for POS tagging, i.e., shallow parsing of the sentence.

**POS tagging and chunking.** Part of speech (POS) tagging is the process of labelling each word in a sentence as a noun, verb, adjective, adverb, etc. Chunking is the labelling of the sentences into the syntactically correlated groups of words such as noun phrase (NP)

<sup>1</sup>[http://ropot.ijs.si/bio3graph/prepareVisualization.php?file=media/supplement/models/Supplement\\_file\\_5.bmg](http://ropot.ijs.si/bio3graph/prepareVisualization.php?file=media/supplement/models/Supplement_file_5.bmg)

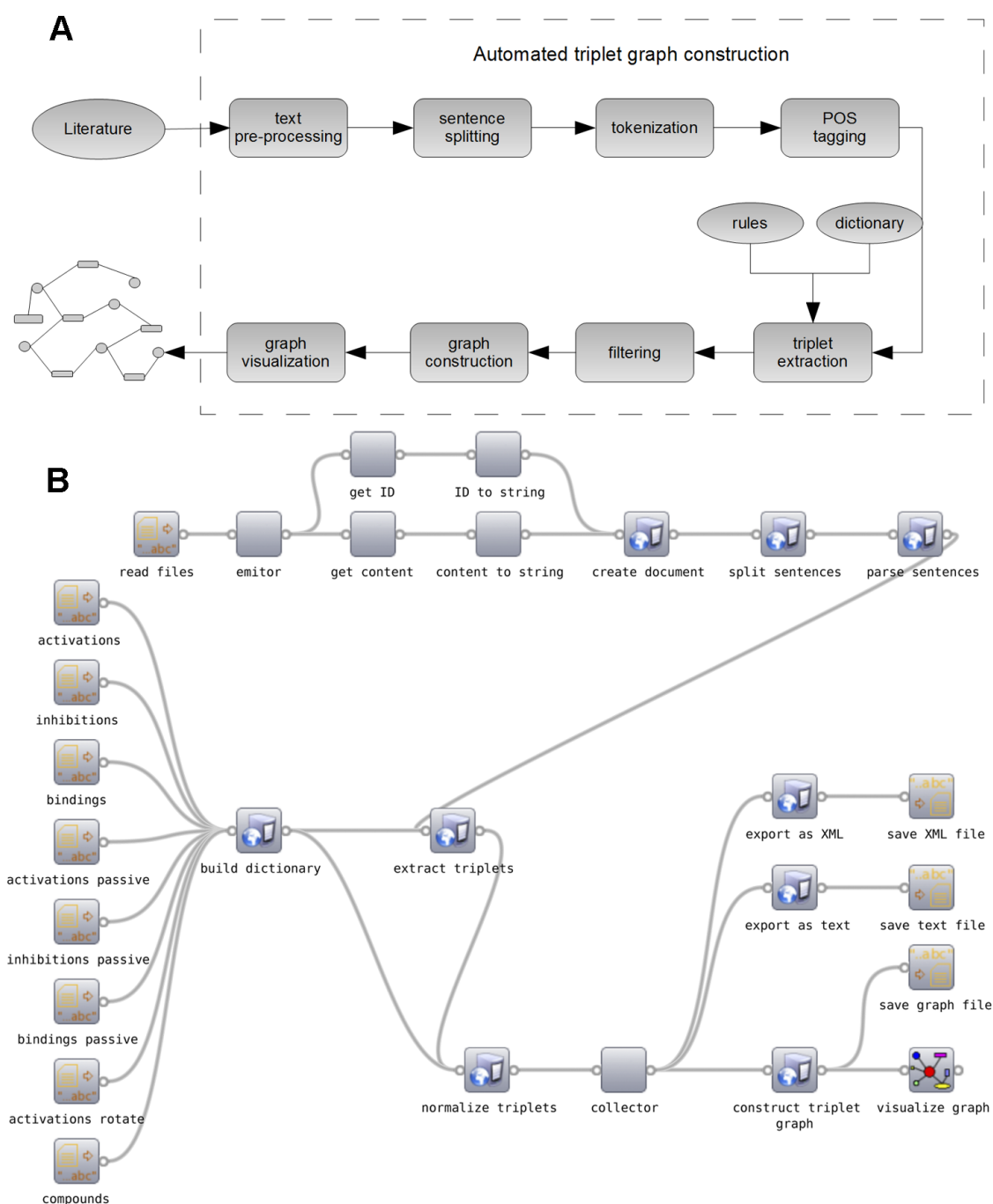


Figure 34: Overview of the Bio3graph methodology, its implementation and a sample output. A) Schematic representation of the Bio3graph methodology. Text processing is performed in a workflow according to the boxes in the schematic diagram resulting in a graph of  $(component1, reaction, component2)$  triplets. B) Bio3graph as a workflow implemented in Orange4WS.

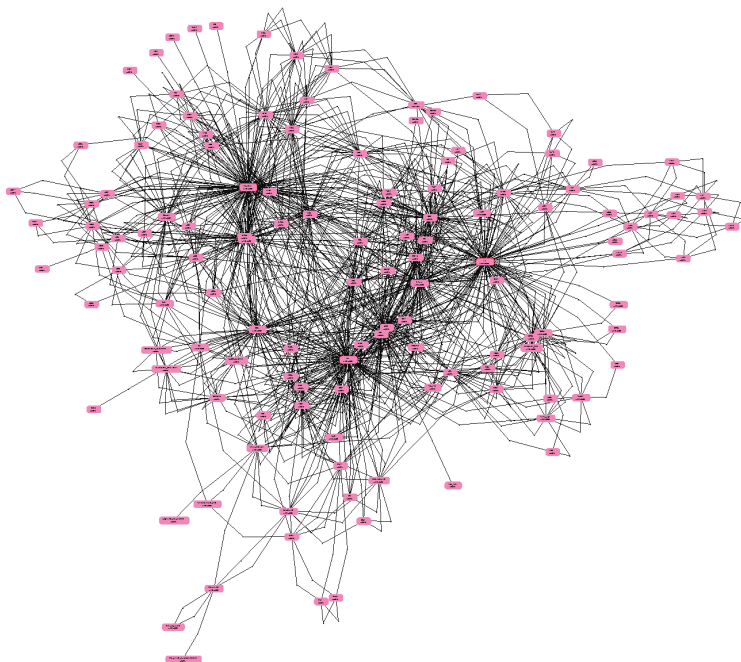


Figure 35: The triplet graph extracted and composed by Bio3graph. The output graph (consisting of 129 components and 1,132 reactions) is visualised with the Biomine graph visualisation engine.

and verb phrase (VP). For the purpose of POS tagging and chunking of biological texts we used the GENIA tagger. The output from the GENIA tagger are the chunks of words, such as Noun Phrase (NP), Verb Phrase (VP), etc. These output chunk labels are the phrase levels according to the Bracketing Guidelines for Treebank II Style Penn Treebank Project<sup>2</sup>. The sentence labelled with the chunk labels is the input to the triplet extraction module.

**Triplet extraction.** The aim of the triplet extraction algorithm is to find the triplets in the form of (*subject*, *predicate*, *object*). If the predicate is in active form, the subject is the part of the noun phrase (NP) before, and the object is the part of the noun phrase (NP) after the predicate. The opposite holds for the passive form of the predicate. The predicate is either one word that belongs to the verb phrase (VP), or it is a multi-word phrase, partially belonging to the VP. The output from the triplet extractor is the triplet list together with the sentence from which the triplet was extracted and the article PMC ID number. Triplet extraction is performed by employing rules with the help of a manually developed vocabulary of components and reactions.

- Components vocabulary. There is a list of all components of the manually developed PDS model structure.
- Reactions vocabulary. There are lists of three different types of reactions (activation, inhibition and binding) together with their synonyms and synonym phrases. For example, the activation in Figure 11.B has induces as a synonym, but also the whole phrase shows increased levels in the presence of.

We have categorised triplet extraction rules into three categories: a rule for one-word predicates, a rule for multi-word predicates and a rule for swap. We briefly describe each of them.

- One-word predicate rule. This rule deals with the predicate that is only one word, such as: activates, stimulates, reduces, etc. The algorithm for triplet extraction of this rule is shown in Figure 36. After the sentence is chunked into chunk tags with the Genia

<sup>2</sup>See publicly accessible phrase level types at <http://bulba.sdsu.edu/jeanette/thesis/PennTags.html>.

tagger, we first compare all VPs with the reactions vocabulary (step 2 in Figure 36). If at least one match is positive, we define it as a predicate. Next, we search for the subject and the object in the Noun Phrases before and after the predicate. The comparison between the NP before the detected VP and the components vocabulary (step 3 in Figure 36) is performed first. This match provides the subject. Next, the match is done between the NP, after detected VP, and the components vocabulary (step 4 in Figure 36). If the match is negative, the matching continues between the next NPs and the components vocabulary until the next VP in the sentence (step 5 in Figure 36). If the match is positive, the object is detected and the triplet is finally extracted (step 6 in Figure 36).

- Multi-word predicate rule. This rule addresses the search for triplets when the predicate is a phrase with more words, such as *is a positive regulator*, *is suppressor of*, *shows increased accumulation*, etc. The subject, the predicate and the object are searched in a similar way as in the Rule for One-Word Predicate.
- Rule for swap. The places of subject and object are swapped if the predicate is in the passive form, or if the predicate matches the *activation\_rotate* vocabulary file (see Section B.2 of Appendix B).

**Filtering.** Filtering of extracted triplets is performed in order to reduce the false negatives. The filtering box removes the triplets from the extracted ones if they belong to any of the following categories:

- Triplets with the same subject and object, for example: *(EDS1, activates, EDS1)*.
- Triplets that are extracted from 'hypothetical' sentences, such as: "It was studied whether EDS1 protein possibly activates EDS5 gene". The following set of 'hypothetical' words was defined: *possibly*, *whether*, *to determine*, *to investigate*, *to study*, *it was postulated*, *it was hypothesized*. If these 'hypothetical' words were detected in the sentence, the triplet was filtered out. Also, if the words like: *may*, *might*, *can*, *could*, *would* were detected in the VP of the predicate, the sentence is considered 'hypothetical' and the triplet is filtered out.
- Triplets extracted from the sentences related to mutant plants. A set of 'mutant plants' words was predefined, such as: *plant*, *mutant*, *line*. If these were detected in the NP of the subject or object, the triplet was filtered out.
- Too 'general' triplets. An example is the following sentence: "The activation of Salicylic acid pathway increases the activity of Jasmonic acid pathway". Triplet *(SA, activates, JA)* would be extracted from this example sentence. However, this triplet is considered to be too 'general' since it addresses not only one specific component, but the whole pathway. For this reason, the set of 'general' words was also defined for filtering: *pathway*, *signalling*, *synthesis*, *biosynthesis*, *response*, *activator*, *inhibitor* and *producer*. If some of these 'general' words were in the same NP of the subject or the object, the triplet was filtered out.
- Negation triplet. If the VP contains the words *not* or *n't*, the triplet is filtered out. Note that processing of 'contradictory triplets' is done by filtering out these negation words.

**Graph construction and visualisation.** Triplets  $T = \{(subject, predicate, object)\}$ , where *subject*, *object*  $\in Components$ , *predicate*  $\in Reactions$  obtained from the manual model or through Bio3graph are used to construct an edge-labelled directed graph  $G = (V, A)$ , where the set of vertices  $V$  is the set of all Components, and the set of arcs  $A$  is a set of all

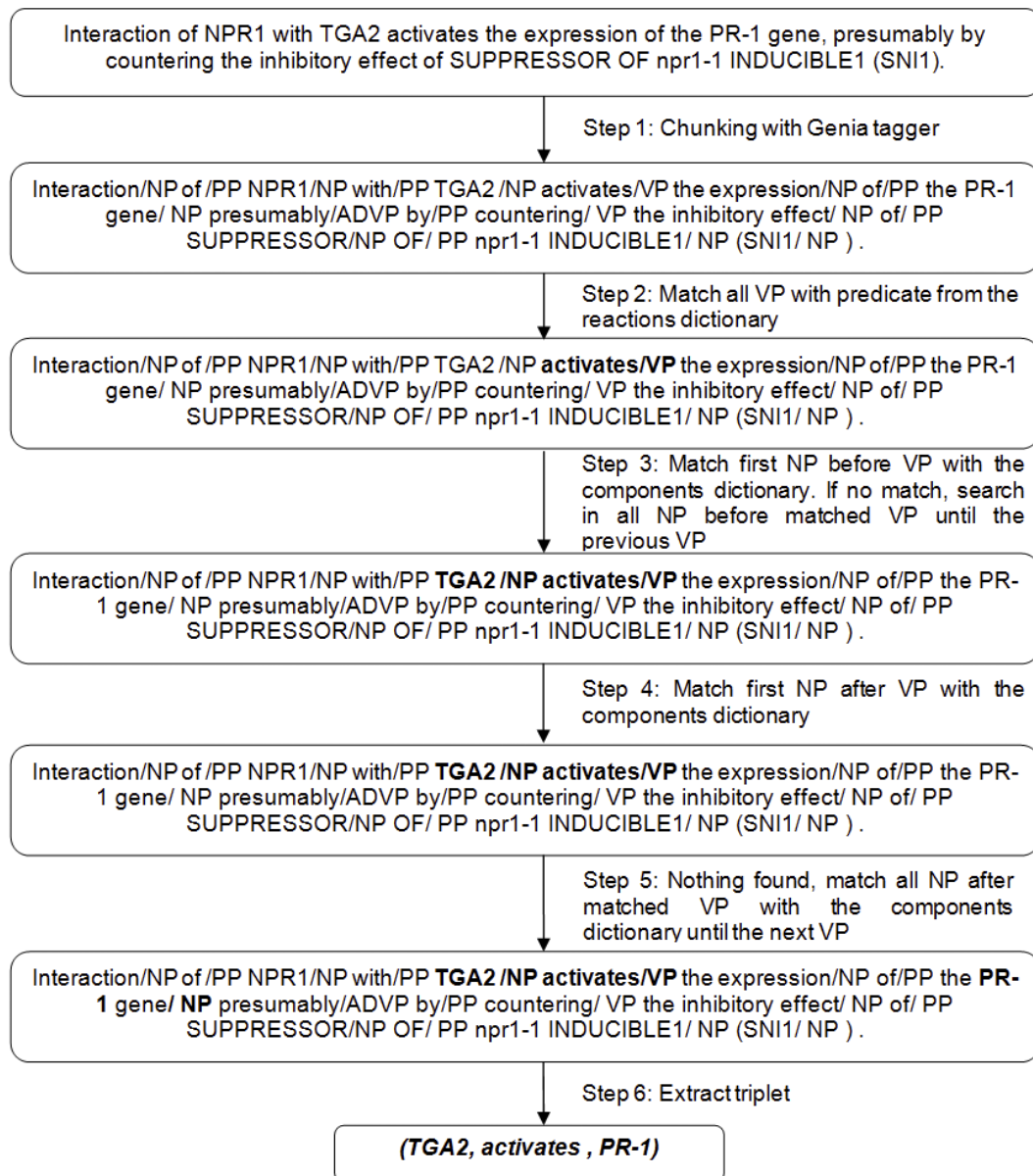


Figure 36: Illustration of the triplet extraction process. We show a part of the flow from input of POS tagging box from Figure 34 until output of triplet extraction box of the same figure. The input to the Genia POS tagger is a previously pre-processed sentence. After the shallow parsing with Genia POS tagger, the algorithm performs the step 2. The final output from the triplet extraction part of Bio3graph approach is a triplet in the form *(subject, predicate, object)* which will be then transformed and visualised as an edge-labelled graph with the Biomine visualiser.

Reactions. The weights are not assigned to arcs but in general weights can be used to reflect the reliability of the extracted triplet. Note that the graph is not necessarily connected and it does not contain any isolated vertices.

Since the extracted structure can contain a very large number of vertices and many unconnected components, it is important to use scalable graph visualisation methods, e.g., Barnes Hut n-body simulation (Barnes and Hut, 1986). We have employed a freely available platform independent graph visualisation component provided by the Biomine system (Eronen and Toivonen, 2012) which implements a variant of the force-directed layouting algorithm, and allows for the visualisation and interactive exploration of reasonably large graphs. For example, a picture of a triplet graph consisting of 175 vertices and 524 edges as drawn by the Biomine visualisation engine is shown in Figure 39.A.

## 6.2 Bio3graph implementation and workflow availability

This section discusses the implementation of the Bio3graph methodology and the availability of the Bio3graph tool. We have implemented the Bio3graph methodology in a general framework which is modular and extensible, and provides functionalities at three different levels of generality. The first level provides classes for core data structures such as Corpus, Document, and Triplet, and the related low-level language processing functions such as sentence splitting, tokenization, tagging, and parsing. The second level contains the triplet extraction algorithm, its custom vocabulary data structure, and various utility functions some of which are algorithm-specific. The third level provides post-processing such as normalization and filtering, and exporting of the results as text, XML, graph and other formats.

Our framework is implemented using the Python programming language, and relies on the publicly available Natural Language Toolkit (NLTK) (Bird et al., 2009) software for natural language processing, and the GENIA tagger (Tsuruoka et al., 2005). NLTK is a native Python suite of libraries and programs for natural language processing while the GENIA tagger provides part-of-speech tagging, shallow parsing, and named entity recognition for biomedical texts. In order to enable access to the GENIA tagger from the Python language environment we implemented a wrapper which turns the standalone tagger program into a Python library, thus allowing an easy integration with the rest of the framework. In addition, our framework also integrates the Biomine tool (Eronen and Toivonen, 2012) for graph construction and visualisation, which enables interactive graph visualization including zoom-in and zoom-out, as well as the relocation of the graph vertices and arcs.

In the triplet extraction workflow, the NLTK library provides sentence splitting (Kiss and Strunk, 2006) while the GENIA tagger is used for tokenization, POS tagging and shallow parsing (chunking) thus forming the backbone of our implementation. Because of the modular structure, the existing software libraries performing various language processing steps can be easily integrated.

In order to provide an easy, system and software independent access to our triplet extraction tool, we have developed a collection of web services that expose the relevant functions of the framework. These services were implemented using the Orange4WS (Podpečan et al., 2012) server tools as stateless SOAP web services. Currently, the web service description document (WSDL) and its related XML schema define three data structures (document, dictionary and triplet) and nine functions: `create_document`, `create_dictionary`, `split_sentences`, `parse_sentences`, `extract_triplets`, `normalize_triplets`, `construct_triplet_graph`, `triplets_to_XML` and `triplets_to_text`.

The Bio3graph workflow, shown in Figure 34.B, works as follows. First, a dictionary is created by calling the `create_dictionary` function which builds the dictionary structure according to the XML schema from the provided text files specifying the reactions and the components. Then, each loaded document is sent to the triplet extraction workflow. It consists of the following parts: creation of the document structure from raw text data, sentence

Table 11: Recall and precision analysis for 50 full-length papers.  $Recall = TP / (TP + TN)$  and  $Precision = TP / (TP + FP)$ , where  $TP$  are the true positives,  $TN$  the true negatives, and  $FP$  the false positives.  $Recall$  is the percentage of the retrieved true positive relations from the whole set of true relations.  $Precision$  is the percentage of retrieved true positive relations from the whole set of retrieved relations.

Reaction types	TP	TP+TN	TP+FN	Precision (%)	Recall (%)
Activation	142	223	311	45.7	63.7
Binding	47	80	134	35.1	58.8
Inhibition	6	9	13	46.2	66.7
All reactions	195	312	458	42.6	62.3

splitting, shallow sentence parsing, triplet extraction algorithm and triplet normalization (removal of duplicates, change of order in the case of passive predicate, and base word formatting). The resulting list of triplets is then saved into a text and XML file, transformed into a graph by Biomine and finally visualised in the Biomine interactive graph visualiser. Note, however, that the triplet extraction workflow is enclosed between the emitter and collector components, provided by Orange4WS, which enable simple, unconditional iterations. The emitter component emits elements of the input iterable object, one at a time, while the collector collects the incoming elements into a list. This allows for extracting triplets not only from a single document but from the whole corpus. The Bio3graph tool is publicly available at <http://ropot.ijs.si/bio3graph/>.

### 6.3 Evaluation of the Bio3graph results

The performance of Bio3graph was evaluated on a corpus of 50 full length articles with manually annotated correct triplets. The performance of information extraction is evaluated by calculating the precision and recall as follows:  $Recall = TP / (TP + TN)$  and  $Precision = TP / (TP + FP)$ , where  $TP$  are the true positives (the number of triplets correctly extracted by Bio3graph),  $TN$  are true negatives,  $FP$  are false positives,  $TP + TN$  is the number of manually identified correct triplets, and  $TP + FP$  is the number of triplets extracted by Bio3graph regardless if they are correct or not. The results achieved by Bio3graph on the annotated corpus are presented in Table 11 showing average precision of 42.6 % and recall of 62.3 %. The annotated texts, the simplified vocabulary, together with the Bio3graph results and a detailed summary for each of the 50 papers are available at: <http://ropot.ijs.si/bio3graph/>.

Several systems for automated information extraction have already been developed reporting remarkable precision and recall results. Most of them extract the protein-protein interactions from text abstracts or from a filtered text corpus, where only sentences with keywords were considered. For example, the Chilobot system reports a precision from 74.4 % for inhibitory relations to 79.1 % for the general protein-protein interaction, with a recall of 91.2 %. Suiseki has a recall of 70 % with the accuracy around 80 % for the best defined reactions. The methodology developed by (Ono et al., 2001) extracted protein-protein interactions for yeast organism with the precision range from 90.2 % for the 'associate' relation up to 96.1 % for the relation 'interact' and the recall for the same organism in the range from 80.9 % for the 'associate' relation up to 89.1 % for the 'interact' relation. In the same study, the recall for extracting a protein-protein interaction in *E. coli* organism ranges from 77.3 % for the 'associate' relation to 85.2 % for the 'complex' relation.

The full-length papers have generally a more complex sentence structures than the abstracts. Therefore, when processing full texts, both the precision and the recall are lower than in abstract-based relation extraction systems. The only system to which we were able



to compare Bio3graph was BioRAT. The BioRAT system achieved a precision of 51.25 % and a recall of 43.6 %. The average precision of our system is 42.6 % which is lower than the precision of BioRAT. On the other hand, recall of Bio3graph is 62.3 % which is almost 20 % higher than the recall of the BioRAT system. Since Bio3graph was validated on the whole articles and not only the abstracts, we were satisfied with a recall of 62.3 %.

In Bio3graph we chose to achieve higher recall at the cost of lower precision, given that the aim of developing Bio3graph was to add to the manually constructed structure the interactions that were missed when manually gathering the information. This means that when used in a real setting, this requires manual reviewing of more false positive triplets, instead of losing some information. As a remark, 62.3 % recall does not necessarily mean that we have not detected 37.7 % of the interactions, given that the interactions between components are often mentioned more than once in a single paper (in the abstract, results and discussion). It is very likely that if we did not extract some triplet from one part of the article, we may still find it in the other parts of the article.

The analysis of non-detected triplets shows that Bio3graph does not cover some sentence constructions. One of these constructions is: "EDS1 protein activates not only EDS5, but also activates SA". In this sentence, there are two triplets (EDS1, activates, EDS5) and (EDS1, activates, SA). However, if, for example, EDS1 is not in the vocabulary of components, and the other two components EDS5 and SA are in the vocabulary, then the algorithm would find the following triplet (EDS5, activates, SA), which is a false positive.

We have noticed that some triplets were not detected due to incorrect sentence parsing by the GENIA tagger. For example, in the sentence "Expression of NPR1 and defence genes was induced by harpin to higher levels, while only MeJA activated COI1." the triplet (MeJA, activates, COI1) cannot be extracted, because the word "activated" was labelled as a noun phrase instead as a verb phrase.

## 6.4 Modelling knowledge of domain experts

The subsection explores the automated relation extraction by Bio3graph in a bisociative setting. Bisociation can be defined as concepts that bridge two, not very connected, domains where an association bridges concepts within a given domain. In other words, bisociation enables discovery of new connections between domains. In this subsection, the emphasis is not on creative knowledge discovery, but rather on specifying and crossing the boundaries of knowledge of individual scientists. This could be used to model the expertise of virtual scientific consortia.

The goal of this experiment is to elicit differences in knowledge and interests between different scientists. We take a simplifying assumption that each scientists' knowledge corresponds to a set of papers she read. The extracted triplets and subgraph thus model her subjective, habitual knowledge (Dubitzky et al., 2012). By combining subjective knowledge bases we obtain a joint bisociative graph (BisoNet) where the intersecting subgraph represents a bridging graph pattern of bisociation. In particular, Bio3graph was used to extract triplets from a set of 122 documents, read by two biology experts:

**Reader A:** Reader A (coloured dark grey) has read 91 papers, of which 13 unique triplets were extracted automatically.

**Reader B:** Reader B (coloured medium grey) has read 31 papers, of which 21 unique triplets were extracted automatically.

**Intersections:** Eight common triplets, extracted from 91 publications read by reader A and from 31 publications read by reader B, were coloured in light grey colour.

Figure 37 shows the model extracted from 122 articles read by the two readers (two biological scientists). Besides supporting the automatic model construction, there are other

benefits from visualising knowledge of different domain experts as illustrated in Figure 5. For instance, one can clearly see which nodes are in the intersection of interest of the two experts (coloured light grey in Figure 5).

This could indicate the areas of joint interest which the two experts might want to investigate jointly in more detail, e.g., to get answers to some yet unexplored research question in the intersection of their domains of expertise. On the other hand, this visualisation enables to see also who has some unique expertise in the field, with no intersection with other experts (coloured dark and medium grey in Figure 5). If applied to modelling the knowledge of larger consortia of readers, this type of information could be used to determine the complementarities of research groups.

The proposed approach to modelling and visualisation of knowledge extracted from the literature could be used also for modelling the know-how of large project consortia where it is hard to track the expertise of all project participants. Consequently, the proposed approach to cross-context modelling may be viewed as a step towards creating virtual laboratory knowledge models.

## 6.5 PDS model structure extracted by Bio3graph from biological literature

Using the following set of keywords: "Arabidopsis thaliana" AND "defence" OR "defense" OR "ethylene" OR "jasmonate" OR "jasmonic acid" OR "salicylate" OR "salicylic acid" OR "pathogen" OR "virus", 9,586 relevant PMC articles were retrieved on April 4, 2011. PMC database was used as it enabled us to gather freely available full text articles and not only their abstracts. These articles were taken as a ground information corpus from which the triplets were extracted. Keywords were selected to obtain the most of the PDS related literature with the emphasis on the JA, ET and SA signalling pathways. Since PMC is a medically oriented database and does not cover some of plant sciences related journals, it is possible that some PDS related articles were not retrieved. Nevertheless, PMC represents the largest source of full text scientific papers and is therefore a relevant basis for our work.

The result of using the Bio3graph triplet extraction algorithm is a set of 1,132 unique triplets, identified from the total of 4,204 extracted triplets. To evaluate the correctness of the extracted triplets, we have manually inspected the sentences from which the triplets were extracted. Since some of the 1,132 triplets appear in several sentences, we have defined the term *correct triplet* in the following way: If the triplet is a true positive in at least one sentence of the whole text corpus, it is considered to be a correct triplet. The reader can open and explore an interactive graph, consisting of 377 correct triplets, if the Java plug-in for the web browser has been installed and enabled<sup>3</sup>.

Most of the relations found by the triplet extraction algorithm are the ones related to activation (out of 1,132 unique triplets in total, 736 are the activation reactions between the components). There are fewer inhibition relations and very few relations of binding type. We have already identified most of these relations when manually constructing the PDS model structure, while some of them are new. Some of the extracted triplets represent direct interactions between the components (i.e., relations between direct neighbours in the graph), while others are indirect (i.e., paths composed of a set of direct relations). A direct interaction is defined as the transduction of the signal between two components without an additional in-between component. For example, the binding of ET to its receptor ETR1 is a direct interaction while the activation of ERF proteins by ET through a signalling cascade is defined as an indirect interaction.

<sup>3</sup>[http://ropot.ijis.si/bio3graph/prepareVisualization.php?file=media/supplement/models/Supplement\\_file\\_7.bmg](http://ropot.ijis.si/bio3graph/prepareVisualization.php?file=media/supplement/models/Supplement_file_7.bmg)

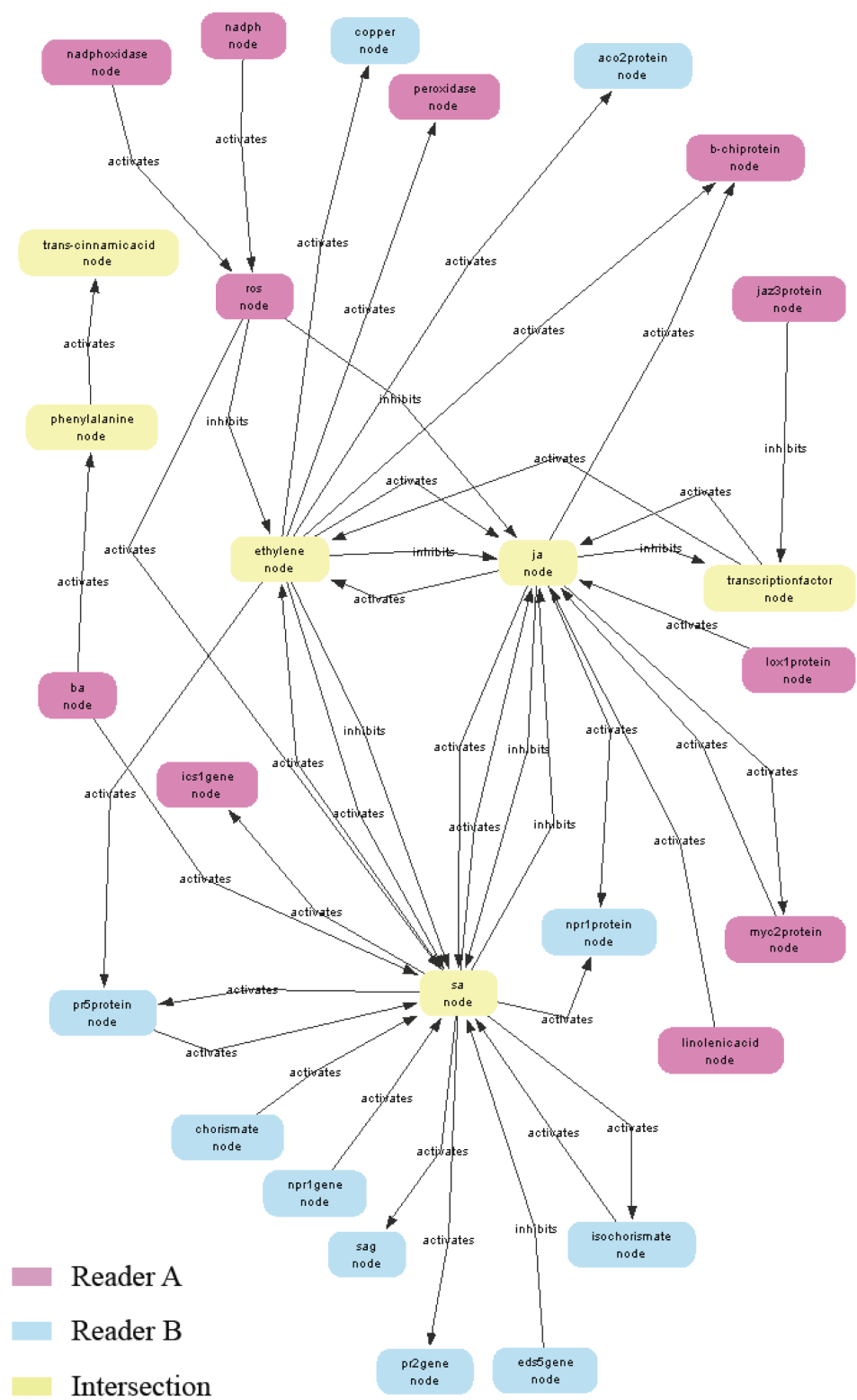


Figure 37: A model of experts' knowledge constructed from a set of triplets extracted from 122 documents, read by two different readers and displayed using the Biomine graph visualisation engine.

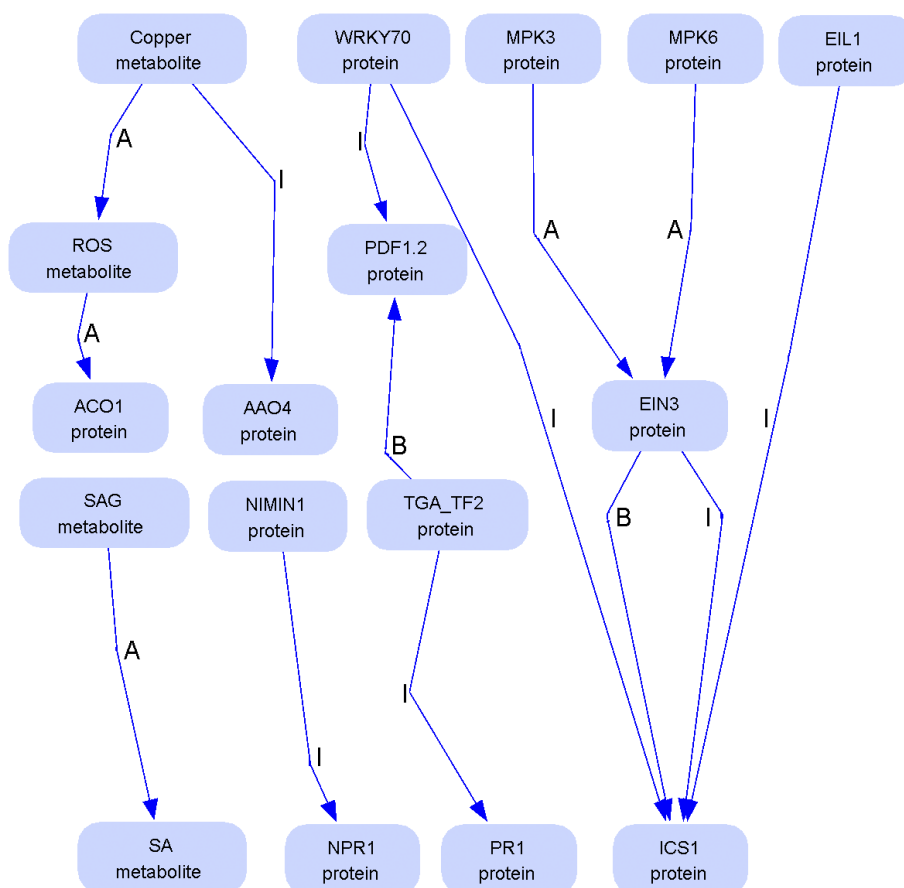


Figure 38: New direct PDS relations extracted from the biological literature. The new direct links result from the Bio3graph processing of 9,586 articles. Bio3graph extracted 14 new direct relations between the components which were not identified in the manually built PDS model structure. Note that two of these triplets are trivial (*SAG\_metabolite*, *activates*, *SA\_metabolite*) and (*NIMIN1\_protein*, *inhibits*, *NPR1\_protein*).

Table 12 gives a summary of the automatically extracted relations between the biological components, emphasizing the numbers of newly discovered direct links (last column of Table 12) discovered by the Bio3graph triplet extraction algorithm. Details of the evaluation for each extracted triplet are provided at <http://ropot.ijs.si/bio3graph>. The obtained new direct links are visualised in Figure 38, while all the correct new (direct and indirect) links discovered by the triplet extraction algorithm are available in Section B.3 of Appendix B. The reader can open and explore an interactive graph, consisting of all the correct new links, (direct and indirect), if Java plug-in for the web browser has been installed and enabled<sup>4</sup>.

Each of the subclasses of correct triplets has its own significance with respect to the PDS model structure. With correct manual direct links we confirmed the applicability of the Bio3graph approach. More importantly, reactions which have not been identified when building the manual PDS model structure (i.e., new direct and indirect links), extend our knowledge on the topic and are therefore of high importance. Indirect relations serve as a database of signal transduction knowledge. Depending on the experimental setup in which these interactions were observed, some of the relations can be redundant. In a biological experiment, a hormone can be applied to the plant in order to investigate its effect on the genes of interest. For example, JA can be applied (in the form of MeJA) to the plant and the expression of lipoxygenase (LOX) or oxide synthase (AOS) genes can be monitored in

<sup>4</sup>[http://ropot.ijs.si/bio3graph/prepareVisualization.php?file=media/supplement/models/Supplement\\_file\\_9.bmg](http://ropot.ijs.si/bio3graph/prepareVisualization.php?file=media/supplement/models/Supplement_file_9.bmg)

Table 12: Summary of PDS related triplets extracted by the Bio3graph triplet extraction algorithm from 9,586 PMC articles. In total, 1,132 triplets were extracted, out of which 377 are correct. Out of these, 14 are newly discovered direct relations and 123 are indirect, while 44 direct and 196 indirect connections were already included in the manual PDS model structure of Figure 14.

Reaction types	Total triplets	Incorrect triplets	Correct triplets	Manual indirect links	Manual direct links	New indirect links	<b>New direct links</b>
Activation	736	446	290	158	41	86	<b>5</b>
Inhibition	352	289	63	18	1	37	<b>7</b>
Binding	44	20	24	20	2	0	<b>2</b>
All reactions	1,132	755	377	196	44	123	<b>14</b>

comparison with the plants that have not been pretreated. If both genes show equal increase in the gene expression level, two triplets (*JA*, *activates*, *LOX*) and (*JA*, *activates*, *AOS*) are somewhat redundant. If AOS is also activated by LOX, increased level of AOS can be due to the JA-induced activation of LOX and not necessarily due to its activation by JA. In this case, more detailed manual inspection and biological validation has to be performed prior to the incorporation of the links into the PDS model and its simulation.

Using Bio3graph we discovered 14 new direct links, out of which two were known to the biological experts but not included in the manual model as we limited ourselves to the most important elements of PDS when building the manual model (*(SAG\_metabolite, activates, SA\_metabolite)* and *(NIMIN1\_protein, inhibits, NPR1\_protein)*). In the former, only the inverse reaction was included, i.e., *(SA\_metabolite, activates, SAG\_metabolite)*, while in the latter, the interaction in the manual model was specified as binding instead of inhibition, which has the same biological function (diminishing the concentration of active NPR1 through binding to NIMIN1). An interesting result is also the identification of components connected with two relations, from which one is a subset of the other. This is the case of ETHYLENE INSENSITIVE 3 (EIN3) and ISOCHORISMATE SYNTHASE 1 (ICS1), where the components are connected with two relations: binding (B) and inhibition (I) (Figure 38). Biologically, this is interpreted as binding of EIN3 to ICS1 causes its inactivation. Biological relevance of the most interesting new direct links is investigated in more detail below.

EIN3 and ETHYLENE INSENSITIVE 3-LIKE1 (EIL1) have been mostly studied as regulators of the ET signalling pathway. In addition to the involvement in the signal transduction of ET-mediated response, Bio3Graph search identified that EIN3 and EIL1 are negative regulators of ICS1. In the paper from which these triplets were extracted it was indeed shown that these two transcription factors inhibit gene expression of ICS1, which is one of the crucial enzymes involved in the SA biosynthesis. The reduction of SA biosynthesis results in a repression SA-mediated signal transduction which is highly-induced upon a pathogen attack (Chen et al., 2009). This study was performed on Arabidopsis plants infected with *P. syringae* bacteria. We did not consider this relation when manually building the signalling model structure, as such links between hormone signalling modules are especially difficult for researchers to explicate. Due to automated knowledge extraction, such new 'out of the box' thinking results in the discovery of additional knowledge.

Another set of new relations identified by Bio3Graph was related to the WRKY family of transcription factors and Mitogen-activated protein kinases (MAPKs or MPKs). Both are relatively large gene families with very specific functions of individual family members. Therefore, they represent a substantial challenge in manual PDS model structure construction and the information related to PDS was overlooked by the biology experts. Indeed, Bio3Graph identified several missing relations. WRKY70 is known to repress the ISO-

CHORISMATE SYNTHASE 1, SID2 (also known as ICS1) transcription, although detailed inspection of the manuscript showed that it is still not clear whether it directly binds to the SID2 promoter or not (Chen et al., 2009). The WRKY70 transcription factor also suppresses the MeJA-induced expression of PDF1.2 (Li et al., 2004) showing its importance in a cross-pathway communication. MAPKs are signal transduction components which play an important role in plant responses to biotic stress. Their performance is cascade-mediated via a complex phosphorelay mechanism. MPK3, MPK4 and MPK6 are the best characterized MAPKs in Arabidopsis (Pitzschke et al., 2009; Olmedo et al., 2006) and, were thus incorporated in the PDS model structure. With Bio3graph a new relation was identified in the literature. A modelling approach has suggested an activation reaction between EIN3 and MPK6 (Sato et al., 2010). In addition, wet-lab experiments have shown that both MPK3 and MPK6 stabilize EIN3 through phosphorylation of threonine 174 (Kendrick and Chang, 2008). MPK3 and MPK6 are therefore both positive regulators of EIN3.

Another interesting new relation identified by Bio3Graph is the inhibition of AAO4 by Cu<sup>2+</sup>. Ibdah et al. (Ibdah et al., 2009) have shown that the high concentration of Cu<sup>2+</sup> reduces the enzymatic activity of AAO4 for 95%". This fine-tuned activation of AAO4 activity is also an additional reaction revealed by our PDS model structure that could be essential in further kinetic studies.

## 6.6 The merged PDS model structure

The manual PDS model structure and the new triplets extracted from the literature were merged into a single graph consisting of 175 components and 524 reactions. This graph, visualised with the Biomine visualisation engine, is shown in Figure 39A. The merged PDS model structure is available for interactive inspection in Section B.4 of Appendix B. The reader can open and explore an interactive version of the Figure 39A, provided that the Java plug-in for the web browser has been installed and enabled<sup>5</sup>. Bio3graph found 44 direct relations from a total of 387 from the manually built PDS model structure (Figure 39B). A reason for this relatively low overlap is that Bio3graph does not process tables or figure images, out of which the information for the manual PDS model structure was also constructed. Moreover, as Bio3graph processes only one sentence at a time, it does not deal with co-reference. Considering co-reference, which could have further improved the results of triplet extraction, is a possible direction of our further research.

The merged PDS model structure thus represents a faithful representation of current knowledge on a structure of PDS with the emphasis on plant-virus interaction. More specifically, we have chosen to base 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 which limits viral spread and multiplication. Activation of HRT protein stimulates accumulation of SA (Chandra-Shekara et al., 2004). SA in Arabidopsis thaliana is synthesized via two pathways both requiring chorismate as a substrate. One pathway goes through a subset of enzymatic reactions initially catalysed by phenylalanine ammonia lyase (PAL) and its homologues (PAL 1,2,3,4). Most of the SA is however synthesized via reaction, catalysed by isochorismate synthase (ICS) and isochorismate pyruvate lyase (IPL) (Vlot et al., 2009; Wildermuth et al., 2001). Arabidopsis encodes two ICS enzymes, ICS1 and ICS2 (Vlot et al., 2009; Dempsey et al., 2011). SA accumulation results in the monomerization and the activation of NPR1, which consequently triggers the activation of the SA dependant PR proteins (Maier et al., 2011; Fu et al., 2012; Moore et al., 2011).

SA signalling is fine-tuned with negative and positive feedback loops. A negative feedback loop slows down a signalling process, while the positive feedback loop has a tendency

<sup>5</sup>[http://ropot.ijs.si/bio3graph/prepareVisualization.php?file=media/supplement/models/Supplement\\_file\\_10.bmg](http://ropot.ijs.si/bio3graph/prepareVisualization.php?file=media/supplement/models/Supplement_file_10.bmg)

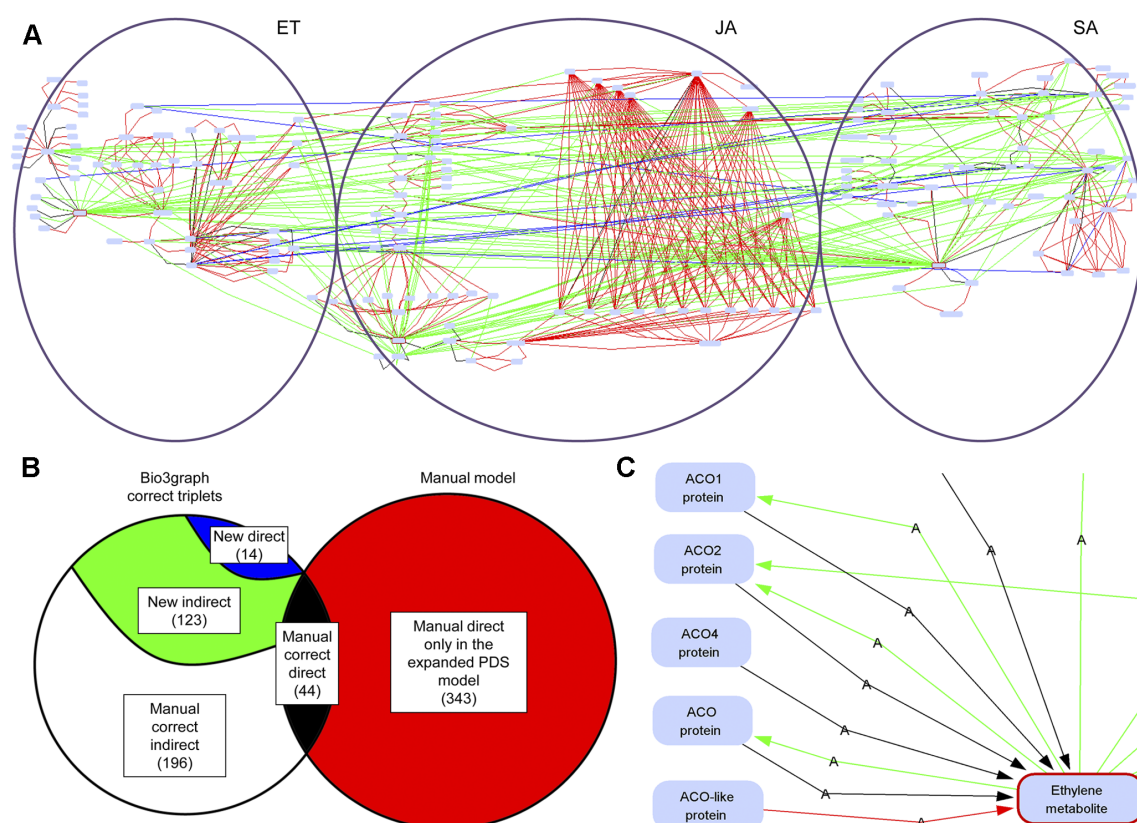


Figure 39: The final PDS model structure constructed by merging the manual and the Bio3graph graph. A) Edge-labelled graph representing the merged model. B) The Venn diagram. The relations in the manual model are all direct and are coloured in red. The intersection between the model relations and the correct triplets extracted from the literature is presented with black colour. From the correct new triplets, the indirect relations are represented with green and the direct ones with blue colour. C) Zoom-in into a part of the merged PDS structure. The links from the manual model are shown in red, while the green coloured relations represent the extracted new indirect links, blue arcs show new direct links and the black arcs show the intersection between the manual model and the correct triplets extracted with Bio3graph.

to accelerate it. The final cascade product regulates its own concentration by activating or inhibiting the genes involved in its biosynthesis. NPR1 inhibits the expression of PAD4 and EDS1 (Shah, 2003), two genes involved in the production of SA and consequently, diminishing its own production, thus forming a negative feedback loop.

The main biosynthetic pathway for JA is oxylipin pathway, linolenic acid being a substrate for JA biosynthesis (Staswick, 2008). JA can be derivatised to different amino acid conjugates. Jasmonyl-isoleucine (JA-Ile) is the conjugate whose biological activity has been proven (Staswick and Tiriyaki, 2004). In the presence of JA-Ile, the SCF complex, composed of a SKP1 (S-phase kinase-associated protein 1), cullin, and a RING finger protein (RBX1/HRT1/ROC1), binds to F-box protein Coronatine insensitive1 (COI1). SCFCOI1 ubiquitinase binds to JAZ and presumably ubiquitinates it (Staswick, 2008; Gfeller et al., 2010; Devoto and Turner, 2005; Chini et al., 2007). When ubiquitinated JAZ repressors are targeted for degradation in 26S proteasome, they result in the de-repression of the transcription factors such as the MYC2 and other beta helix-loop-helix transcription factors (Fernández-Calvo et al., 2011) which activate JA-dependant PR gene expression (Moore et al., 2011).

L-methionine is transformed by S-adenosyl-L-methionine (SAM), 1-amino-cyclopropane-1-carboxylate synthase (ACS) and ACC oxidase (ACO), to form a gaseous hormone ET (Wang et al., 2002). When synthesized, ET binds to its receptors. There are five membrane-located receptors identified in Arabidopsis (ETR1, ETR2, EIN4, ERS1 and ERS2) (Kendrick and Chang, 2008; Zhao and Guo, 2011). Binding of ET to its receptor leads to CTR1 deactivation, which finally results in downstream activation of EIN3/EIL1/EIL2 transcription factors (Kendrick and Chang, 2008; Stepanova and Alonso, 2005). CTR1 levels are also regulated by ubiquitination and 26S proteasome degradation via EBF1/EBF2 - Skp-Cullin-F-box (SCF) E3 ligase complex (Zhao and Guo, 2011). The concentration has to be well regulated, since they are the crucial positive regulators of ET signalling.

SA, JA and ET pathways do not function independently, but are rather interconnected by agonistic and antagonistic interactions to fine-tune the plant defence. These regulations are very complex and often more than one component is involved in the signal transduction (Wang et al., 2002; Koornneef and Pieterse, 2008; Pieterse et al., 2001). When Bio3graph was applied to enhance the manually built PDS model structure, most of the newly-found relations were characterized as 'indirect'. Most of them indicate a cross-talk between the sub-pathways or a feed-back regulation of the crucial components in the model. However, some new Bio3graph links are direct. A cross-talk link connects ET and SA sub-pathways: MPK3 and MPK6 activate ET signalling pathway transcription factor EIN3 which negatively regulates SA biosynthesis through binding to ICS1 (Figure 38). Using dynamic modelling approaches, the SA concentration changes can be simulated in different model topologies. If we consider only the manually built model, only NPR1 and MPK4 negatively affect the SA concentration. Removing these two proteins from the model by *in silico* knock-out results in an infinite rise of SA. We assume that in a real biological experiment with NPR1/MPK4 double mutant, the SA levels would drop thus implying the other negative regulators of SA biosynthesis are involved. Adding the cross-talk link with ET sub-pathway found by Bio3graph could improve the model to more accurately predict SA concentration changes in such knock-out plants.

Bio3graph is used also for extracting detailed information about certain protein family. Several enzymes that are members of the same family can be involved in one biological reaction. For example, according to AraCyc there are five ACC oxidases (ACO1, ACO2, ACO4, ACO and ACO-like) catalysing the last step of biosynthesis of ET (Mueller et al., 2003). Evidence for their 1-amino-cyclopropane-1-carboxylic acid oxidase activity can be obtained either from the experimental data or from a computational prediction, which is usually sequence-based. With Bio3graph we were able to extract the data for each of the family members, compare the experiments and evaluate their importance for the model. In



Figure 39.C, one can see that Bio3graph identified all ACO family members, apart from the ACO-like, activating ET production. When manually checking the papers we established which relations represent biochemical knowledge and which interactions rely only on sequence homology data. Such detailed information can be used in further dynamic modelling experiments.

As shown in Figure 39.C, five proteins belonging to ACO protein family have manually assigned activation links in our model. Triplet extraction tool Bio3graph has confirmed four out of five activation links. The activation link between the ACO-like and ET originates from the manual construction process of the structure (after expanding from level 2 to level 3 - see Figure 11.A) and has not been confirmed by Bio3graph. ACO-like therefore either has a different function than other members of this protein family or it was not explicitly determined as ET biosynthetic enzyme in the literature surveyed by Bio3graph. To determine the real function of ACO-like biological experiments should be conducted using methods that reduce or increase the expression of genes encoding ACO-like. For example, reducing ET concentration in Arabidopsis knock-out plants would confirm the involvement of ACO-like in ET biosynthesis.

Indirect links found by Bio3graph can also guide researchers to form new hypotheses and perform experiments guided by model predictions. For example, indirect link (SA, activates, EDS1) (see Section B.4 of Appendix B for a detailed view) means that it is not precisely known whether SA directly activates EDS1 or it activates some of the biosynthesis components upstream from EDS1 which results also in activation of EDS1. The exact nature of such activation can be checked and tested in the laboratory experiments. Nevertheless, this link provides a first clue about the existence of a positive feedback loop in the SA pathway.



## 7 Incremental Revision of Structures of Biological Models

*To improve is to change; to be perfect is to change often.*

Winston Churchill

After biological models are published, they do not tend to be re-examined fast. However, there is constant flow of new knowledge related to the biological mechanism. The fast updates of the biological model structures are possible with the incremental approach based on the relation extraction from literature.

This chapter is organised as follows. We describe the methodology and implementation of the incremented version of the Bio3graph tool that searches the biological literature for the relations between the biological components and outputs a graph of triplets in the form (*component1*, *relation*, *component2*). The incremental approach is demonstrated on two use cases. In the first use case, a simple plant defence network, created manually by merging three existing structural models, is extended in two incremental steps. In the second use case, a complex, published model structure of defence response in *Arabidopsis thaliana* is incrementally updated with information extracted from recently published articles. Whenever necessary, the structures of biological models can be quickly updated by using the incremental version of Bio3graph.

### 7.1 Methodology for incremental biological network revision

The work on incremental revision of biological models is based on the Bio3graph approach (described in Section 6.1) which allows for automated extraction of biological relations in the form of triplets from the literature. The Bio3graph approach to information extraction from biological literature is extended with new features which allow for periodical updates of model structures using newly published scientific literature.

The incremental approach is demonstrated on two use cases.

- A simple plant defence model structure with 37 components and 49 relations created manually by merging three existing structural models is extended in two incremental steps.
- A complex published model structure of defence response in *Arabidopsis thaliana*, containing 175 nodes and 524 relations, is incrementally updated with information extracted from recently published articles.

In the following we describe the additional steps of the incremented version of Bio3graph (see Figure 40 for schematic overview of the methodology).

The Bio3graph approach is essentially a workflow of processing components which extract triplets of the form (*component1*, *reaction*, *component2*) using natural language processing tools. Apart from the steps of Bio3graph tool, its incremental extension implements additional steps: (1) literature retrieval, (7) graph merging, (8) redundant relation removal and (9) colour reset. We define the inputs to the incremental extension as follows (see Figure 40).

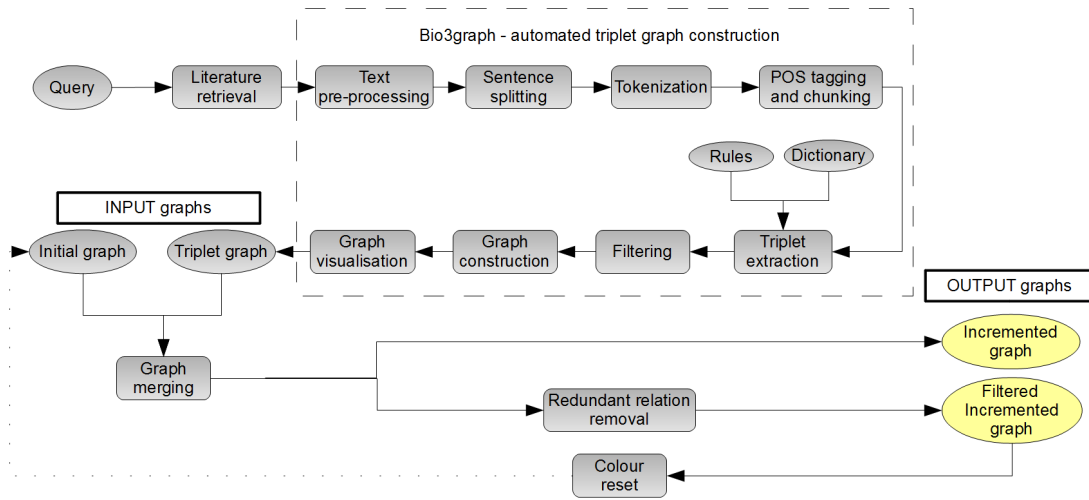


Figure 40: Scheme of the methodology for incremental construction of biological networks using information extraction from literature.

The existing model structure which is the subject of incremental enhancement is called the “Initial graph” and the result of Bio3graph is called the “Triplet graph”. The incremental extension of Bio3graph produces two outputs: “Incremented graph”, a result of merging the Initial and the Triplet graph, and “Filtered incremented graph”, a result of redundant transitive relation removal from the “Incremented graph”.

### 7.1.1 Literature retrieval

The collection of relevant scientific publications about various aspects of the selected case study topic (*Arabidopsis thaliana* defence response) was obtained from PubMed Central (PMC), a freely accessible online archive of biomedical and life sciences literature, which is developed and managed by National Library of Medicine’s National Center for Biotechnology Information (NCBI). As of May 2013, PMC database hosts more than 2.7 million articles for which full text is available, either as HTML/XML or PDF or both. NCBI also provides the Entrez Programming Utilities (E-utilities), which enable programmatic access to the Entrez query and database system covering a variety of biomedical data, including nucleotide and protein sequences, gene records, three-dimensional molecular structures, and the biomedical literature (Sayers, 2011). The E-utilities are accessible via the HTTP protocol using GET and POST commands, and return the response in a structured XML document.

PMC also provides the PMC Open Access Subset (OA), a growing collection of publications which are made available under a Creative Commons or similar license. The OA subset is a valuable source of reviewed scientific publications which are readily available for data mining, text mining, and information extraction using automated processing pipelines. To facilitate computer processing, the Open Archives Initiative service and the FTP service allow to download full-text XML as well as images, PDF, and supplementary data files for all articles in the OA subset.

To obtain sets of documents to increment model structures in both use cases we have used the PMC Advanced Search Builder to construct the query which should cover as much literature as possible regarding the defence response signalling pathways in *Arabidopsis thaliana*. The query is as follows:

```

"arabidopsis thaliana"[All Fields] AND (
  "defence"[All Fields] OR
  "defense"[All Fields] OR
  "ethylene"[All Fields] OR

```

```

"jasmonate"[All Fields] OR
"jasmonic acid"[All Fields] OR
"salicylate"[All Fields] OR
"salicylic acid"[All Fields] OR
"pathogen"[All Fields] OR
"virus"[All Fields]
)

```

The query was used for both use cases only with the following differences.

- For the first use case with the simple model all publications regardless of the publication date were collected (the query was performed in May 2012). Also, the keyword “virus” was excluded from the query and the source document set was not limited to the PMC OA subset in order to collect as much knowledge as possible (the most important non-OA publications were considered and extracted manually as PMC does not allow automated downloading of any publications outside of the OA subset). For this simple model structure the query yielded 10,299 documents out of which some were available only as PDF and were left out. In order to time-stamp them we have collected pub-date tags and extracted the earliest available date (which in most cases corresponds to the classic publication date or the electronic publication). The final corpus, containing 10,207 documents, was divided in two datasets which were used in two incremental steps of the triplet extraction by Bio3graph.
- In the second use case of the complex model structure the earliest publication date was set to the latest date of any publication used by the authors of the model (Miljkovic et al., 2012) (April 5th, 2011). The query resulted in 2,988 full-text publications which were also subject to automated triplet extraction leading to an incremental enhancement of a complex, recently published model structure.

### 7.1.2 Graph merging

In order to allow for incremental updates of an existing model structures using Bio3graph (or any other biological graph construction method) the existing graph and newly extracted graph have to be merged. The merging process produces a union of the graphs and applies colour coding to relations in order to distinguish between known, new, and duplicate relations.

All biological model structures discussed in this work (and biological network structures in general) are directed edge-labelled graphs with several types of relations. Therefore, the data structure used for merging must support the most general type of graphs which is called a multigraph. A multigraph supports duplicate relations, relations of different types and cycles.

The merging procedure merges the input graphs into a single graph using the following colour coding: existing relations originating from the existing graph are coloured in black, newly discovered relations originating from the extracted graph are coloured in red while the re-discovered existing relations originating from the extracted graph are coloured in green. Other existing information about nodes and relations is also preserved during merging.

### 7.1.3 Redundant relation removal

Automated extraction of biological relations with Bio3graph can yield relations which may not appear in the existing model (the subject of incremental revision) but do not contain new biological knowledge. Such relations, which are known as transitive relations in graph theory, represent only a shortcut of a chain of biological relations. For example,

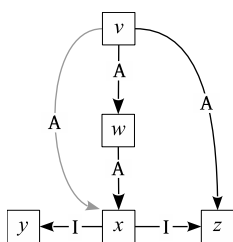


Figure 41: An example of a redundant transitive relation in a simple graph. The relation  $v \text{ Activates } x$ , shown in grey, is transitive. It does not contain new biological knowledge and is thus redundant.

the new relation  $v \text{ Activates } x$  does not represent new biological knowledge given the chain  $v \text{ Activates } w \text{ Activates } x$ .

In general, transitive relations can be removed by computing the transitive reduction of the directed graph. Transitive reduction yields a new graph on the same set of nodes with as few edges as possible to maintain the same reachability relation. For a finite, directed acyclic graph the transitive reduction is a unique subgraph which is also the minimum equivalent graph. However, the transitive reduction of directed graphs with cycles is not unique and is not necessarily a subgraph. This means that the transitive reduction of general biological model structures – which typically contain cycles – is not applicable as it may produce several equivalent model structures and also introduce new relations.

For this reason, we have developed a procedure which does not exhibit the mentioned limitations. Given an existing graph and a new graph, the procedure evaluates all relations in the new graph. For each relation in the new graph the procedure tries to find a path in the existing graph. If such a path exists, the new relation is transitive and thus redundant. If no such path exists, the new relation is not redundant as it represents a new piece of knowledge. It should be noted that we do not make any assumptions about the existing graph and that each type of relation is considered separately, i.e., the path must contain only relations of the same type.

Figure 41 shows an example of a transitive relation in a simple graph. The redundant transitive relation  $v \text{ Activates } x$  is shown in grey. On the other hand, the relation  $v \text{ Activates } z$  is not transitive as no alternative path consisting only of relations of the same type exists between  $v$  and  $z$ .

#### 7.1.4 Colour reset

The incremental revision of the Initial graph with a Triplet graph extracted from the literature can be used again in the next iteration (see Figure 40). The only requirement is that the colours of relations are reset to the default colour (black) so that merging and colour coding can be performed correctly using a next Triplet graph obtained by Bio3graph from a new set of documents.

## 7.2 Implementation and the workflow availability

Our implementation of incremental development of biological network structures is built as an extension of Bio3graph. Therefore, we only discuss the implementation of new components and the complete integrated solution as a scientific workflow as the implementation of Bio3graph is presented in length in (Miljkovic et al., 2012).

### 7.2.1 Literature retrieval

We have implemented literature retrieval in the Python programming language using the ESearch and EFetch functions provided by PMC E-utilities. The implementation accepts the

query constructed manually or by using the Advanced Search Builder and invokes ESearch to obtain the identifiers of the corresponding articles. The identifiers are then matched against the downloaded archives of the PMC OA subset and full-text XML files are extracted. Our XML parser, which is used to transform the XML files into plain text data, is set to ignore the following XML tags which do not contain relevant textual data and may contain unwanted special characters or words with excessive length (they can cause problems in some language processing components): *xref*, *table*, *graphic*, *ext-link*, *media*, and *inline-formula*.

### 7.2.2 Graph merging

The graph merging component was implemented using the NetworkX<sup>1</sup> Python library which can be natively integrated into the Bio3graph workflow in Orange4WS (Podpečan et al., 2012). To maintain the compatibility with the Bio3graph graph representation in Biomine's graph format we have also implemented a bidirectional transformation between the Biomine's graph format (Eronen and Toivonen, 2012) and NetworkX data structures which preserves all existing information concerning nodes and relations. For example, if the positions of the nodes in the visualisation canvas are available they will be preserved during merge which is essential for the efficient visual comparison of the graphs.

### 7.2.3 Redundant relation removal

The discovery of transitive relations also relies on the NetworkX library. It is implemented as a separate component which accepts the existing and the new graph and returns a list of redundant relations. In this way, the relations can be reported to the user, removed from the merged graph or even marked with a different colour in a merged graph to aid the visual evaluation of the graph.

The procedure is implemented using the path discovery procedures available in the NetworkX library. The search for an existing path in the existing graph is performed by generic function *has\_path(G, source, target)* which is essentially instantiated to the bidirectional shortest path search which executes a breadth-first search from both the source and the target and returns a list of nodes in the path or an empty list if such path does not exist.

### 7.2.4 Colour reset

Reset of the colours of relations works by modifying the attributes of the relations which are stored in the NetworkX MultiDiGraph data structure. The implemented bidirectional transformation from this data structure to the Biomine's format can be used to export the structure and properties of the reset graph into a portable text file.

### 7.2.5 The workflow

The proposed extension of Bio3graph was implemented as a scientific workflow in the same service-oriented data mining environment Orange4WS (Podpečan et al., 2012) where Bio3graph was developed and implemented. By utilising Orange4WS the following benefits were achieved. First, incremental revision and development is natively integrated with Bio3graph. Second, workflow-based implementation ensures repeatability of experiments and makes the modifications and extensions of the developed workflow easy. Finally, the workflow-based solution is shareable and can be used anywhere where Orange4WS is available.

The implementation of the incremental model structure development approach in the Orange4WS environment is shown in Figure 42. The first part of the workflow implements Bio3graph (loading of documents, preprocessing and parsing, loading of vocabularies,

---

<sup>1</sup><http://networkx.github.io>

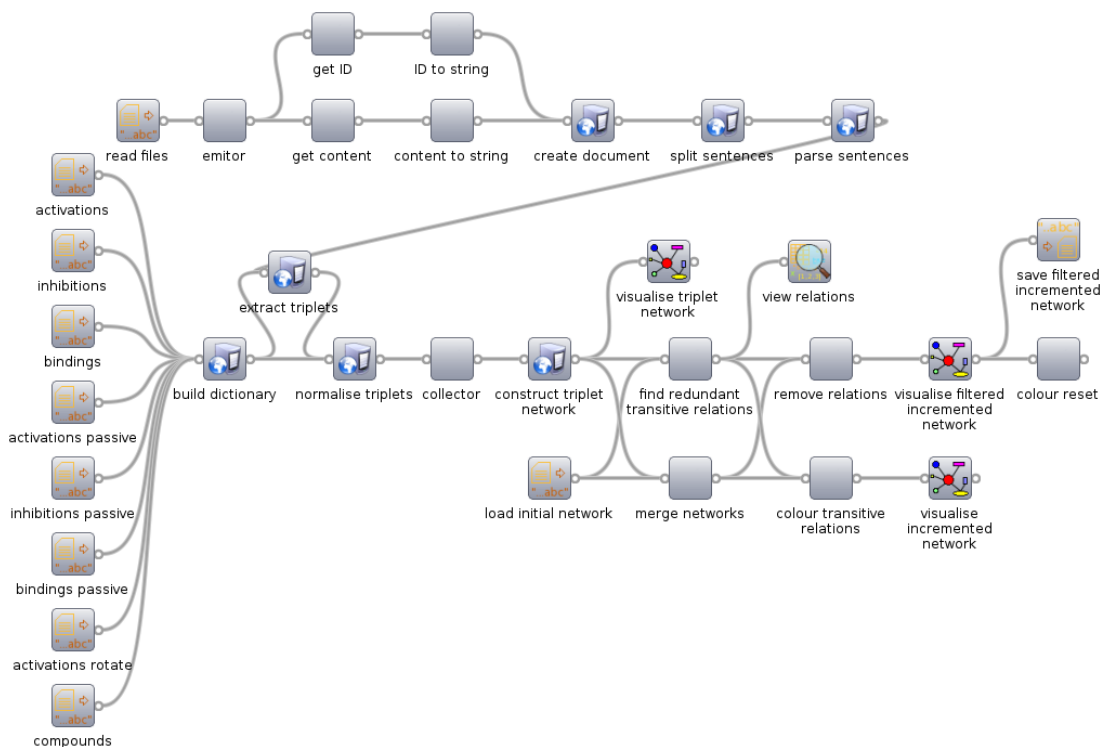


Figure 42: A screenshot of the workflow implementing the proposed incremental revision of biological network structures. The first part of the workflow implements the Bio3graph approach for automated triplet graph extraction from biological literature while the second part implements incremental extension of biological model structures.

triplet extraction, and graph construction) while the second part implements incremental development (graph merging, colour coding, removal of redundant transitive relations, and visualisation of incrementally constructed model structures). It should be noted, however, that only one incremental step is composed in the workflow.

The workflow works as follows. First, the dictionary has to be constructed as it is needed for the triplet extraction algorithm. This is accomplished by loading the dictionary files which are passed to the web service which constructs the dictionary structure. The parallel branch of the workflow is used to prepare the data. A collection of text files is sent to the *Emitor* component which simulates the for-loop by outputting the elements of the input list one by one. Each emitted document is passed to the web service which creates the document data structure. Each instance of this data structure is then forwarded to the sentence splitting component which is followed by POS tagging with the GENIA tagger. The document, tokenised and parsed, is sent to the triplet extraction web service which requires also the dictionary. The extracted triplets (if any) are subjected to the normalisation process where the names of the involved components and the reaction are replaced by the corresponding base names (for example, “influence accumulation” is replaced by “activate” and “SA” is substituted for “o-Hydroxybenzoic acid”). The extracted triplets from all documents are collected by the *Collector* component which closes the emulated for-loop. The Bio3graph part of the workflow concludes with the construction of a graph from triplets, to which we refer as a Triplet graph, and its visualisation.

The second part of the workflow, which performs incremental revision of network structure starts by loading an Initial graph from a file which will be the subject of incremental enhancement. This model and the Triplet graph are sent to the component which discovers and reports redundant transitive relations. In parallel, the graphs are merged into an Incremented graph which is colour coded marking differently the relations that belong solely to the Initial graph, the ones in the graphs’ intersection and the new ones. The discovered



redundant relations are then removed from the Incremented graph and finally the Filtered Incremented graph is visualised and saved. Alternatively, the redundant relations are not removed but coloured differently in the Incremented graph, which is useful for a visual comparison. In the very last step of the workflow the colours of relations in the Filtered Incremented graph are reset to black which makes the graph ready for the next incremental revision which can be performed by providing a new set of documents and repeating the execution of the entire workflow.

### 7.3 Use case 1: a simple plant defence model structure

This subsection presents the results of the experiment performed on a simple model which is a subset of the plant defence mechanism where the structure of the models is available in the literature.

The Initial graph in this experiment was constructed manually from the published figures (structural models) in scientific publications (Gonzalez-Garcia and Diaz, 2011; Turner et al., 2002; Shah, 2003). It was expanded in two incremental steps using Bio3graph and its incremental extension on a time-labelled collection of documents.

#### 7.3.1 The Initial graph

We have manually constructed the Initial graph from structural models published in the scientific literature. Three schemata describing the salicylic acid (SA), jasmonic acid (JA) and ethylene (ET) pathways (Gonzalez-Garcia and Diaz, 2011; Turner et al., 2002; Shah, 2003) were selected and transformed into a directed graphs with multiple relations (see Figures 43, 44 and 45). To obtain the Initial graph all three were merged into a single graph which contains 37 nodes (biological components) and 49 links. The merged graph is shown in Figure 46.A.

Among all the represented components, SA, JA and ET are the most crucial for plant defence. The types of relations between the nodes are *activation* (abbreviated as A) and *inhibition* (abbreviated as I). The nature of interactions from the schemata was easily recognisable, and the transformation was accomplished with respect to these types. Too general components such as lipid, lesion, pathogen, etc. were not implemented in the Initial graph. On the other hand, to prevent the loss of connections between components we have added several reaction products as nodes.

#### 7.3.2 The Triplet graph

Triplet extraction with Bio3graph requires a predefined vocabulary of components and reactions. We have developed the component vocabulary from the list of the Initial graph nodes that represent biological components. Small compounds and proteins were considered. In addition, we have acquired the list of component synonyms from TAIR (Swarbreck et al., 2008) and iHOP (Hoffmann and Valencia, 2004) sources. The vocabulary of reactions with reaction synonyms was used from Supporting Information S4 in (Miljkovic et al., 2012)). Besides the activation and inhibition reaction types that exist in the Initial graph, we have also taken into account the additional *binding* (abbreviated as B) reaction type.

The collection of full-text documents for triplet extraction with Bio3graph was divided into two sets according to the defined time point. We used the time point of November 2001, which is the earliest publication date of the three observed publications (Gonzalez-Garcia and Diaz, 2011; Turner et al., 2002; Shah, 2003). The first set of documents (published before November 2001) contains 1,714 publications while the second one contains 8,493 publications (published after November 2001). Using the two sets of documents two sets of triplets were obtained with the Bio3graph method. We refer to the first set as triplets before the time point and to the second set as the triplets after the time point.

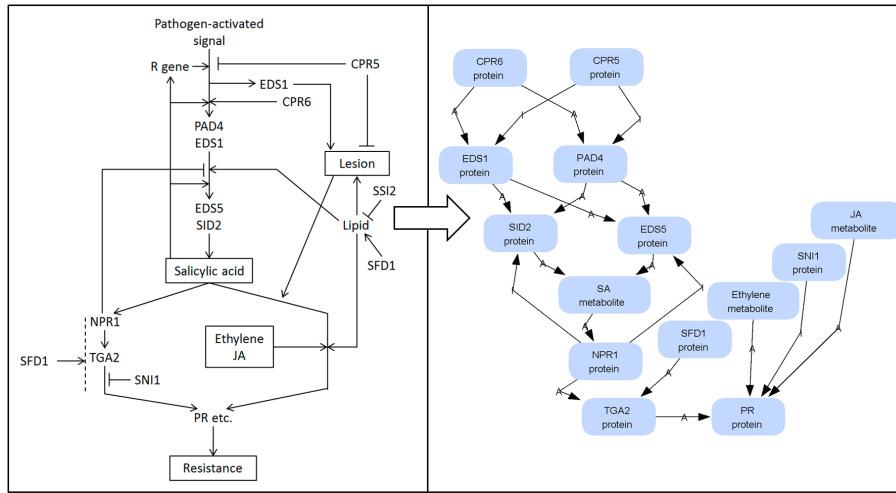


Figure 43: Transformation of the SA model structure available in literature into a directed graph with labelled edges. The model structure originates from the study of (Shah, 2003).

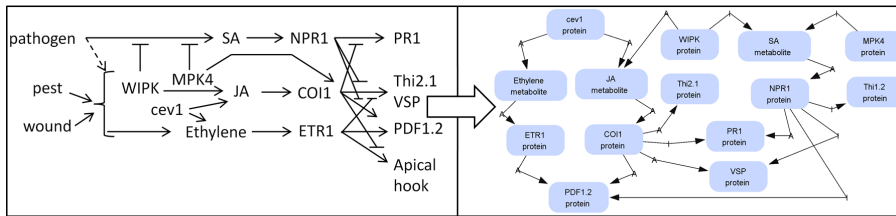


Figure 44: Transformation of the crosstalk between SA, JA and ET pathways available in literature into a directed graph with labelled edges. The model structure originates from the study of (Turner et al., 2002).

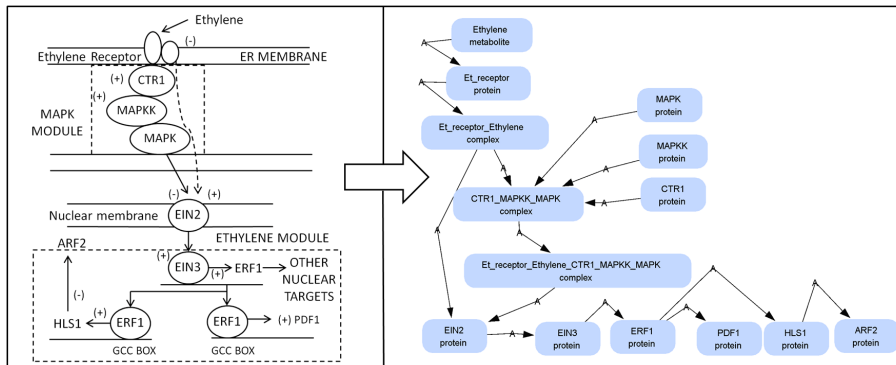


Figure 45: Transformation of the ET model structure from literature into a directed graph with labelled edges. The model structure originates from the work of (Gonzalez-Garcia and Diaz, 2011).

Some of the extracted triplets appear in several sentences but we count only the number of unique triplets. We introduce the term *correct triplet* in the following way: if the triplet is a true positive (TP) in at least one sentence of the whole text corpus, it is considered to be a *correct triplet*. The extracted triplets were inspected manually and classified as correct or false positives (FP).

The summary of triplet extraction from documents before and after the time point is presented in Table 13. The Triplet graph for the first incremental step is configured from the set of correct triplets before time point (Figure 46.B) while the Triplet graph for the second incremental improvement consists of the correct triplets after the time point.

Table 14: The summary of relations of the Incremented graph shown in Figure 46.C after the first incremental step for the use case 1. The initial links originate only from the Initial graph, while the intersection, new redundant and new links originate from the Triplet graph. The intersection links are the common relations of the Initial and the Triplet graph. The new redundant links are the transitive relations while the most interesting are the new links, which represent exclusively new relations discovered by the Bio3graph tool.

Reaction types	Initial links	Intersection links	New redundant links	New links
Activation	32	6	10	10
Inhibition	11	0	0	7
Binding	0	0	0	2
All reactions	43	6	10	19

Table 13: The summary of triplet extraction from documents before and after the time point for the use case 1 (simple plant defence model structure).

Reaction types	Triplets before time point			Triplets after time point		
	Total	Correct	FP	Total	Correct	FP
Activation	52	26	26	231	92	139
Inhibition	19	7	12	157	43	114
Binding	3	2	1	30	17	13
All reactions	74	35	39	418	152	266

### 7.3.3 First incremental step

The first incremental improvement of the Initial graph is performed with the Triplet graph consisting of correct triplets before time point of November 2001. The result of this enhancement is the Incremented graph with 37 nodes and 78 relations shown in Figure 46.C. Green, red and pink arcs represent the correct triplets discovered by Bio3graph from the biomedical texts already available at the time point, while the black arcs are the relations present in the Initial graph. The summary of relation types in the graph is shown in Table 14.

In the Incremented graph in Figure 46.C the green arcs represent the intersection between the Initial graph the Triplet graph. The red and pink arcs represent the newly discovered relations not present in the Initial graph. However, the arcs coloured in pink are transitive and thus redundant as they do not introduce new knowledge into the underlying biological model. The Initial graph, however, can contain transitive relations but they do not interfere with our transitive relation discovery procedure as described in Section 7.1 as such relations are only searched for in the new Triplet graph. The incremental extension of Bio3graph supports the removal of such relations. The result of this operation is shown in Figure 46.D and represents the final Filtered Incremented graph. The knowledge in this graph which is most interesting for a domain expert is represented by red arcs (newly discovered biological relations from the literature).

### 7.3.4 Second incremental step

The second step incremental step is performed in an analogue way as the first. The input graphs for the incremental extension of Bio3graph are as follows. The Initial graph is the Filtered Incremented graph shown in Figure 46.D, but all of its relations are now marked as known (all arc are reset to the initial black colour). The Triplet graph is constructed from the set of correct triplets after the time point of November 2001.

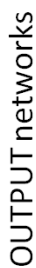


Figure 46: The enhancement of the Initial graph (A) with the correct triplets after first incremental step for the use case 1. The left side represents the input graphs for the incremental extension of Bio3graph while the right side represent the output graphs. A) The Initial graph created by merging the manually constructed three graphs from the literature shown in Figures. 43, 44 and 45. B) The Triplet graph constructed from the correct triplets extracted with Bio3graph. C) The Incremented graph obtained by merging the Initial and the Triplet graph. The relations present only in the Initial graph are coloured in black while the relations present also in the the Triplet are coloured in green, red or pink. Relations present in both Initial graph and triplet network are coloured in green, newly discovered relations are coloured in red while the newly discovered, redundant transitive relations are coloured in pink. D) The Filtered Incremented graph obtained from the Incremented graph by removing the redundant transitive relations.

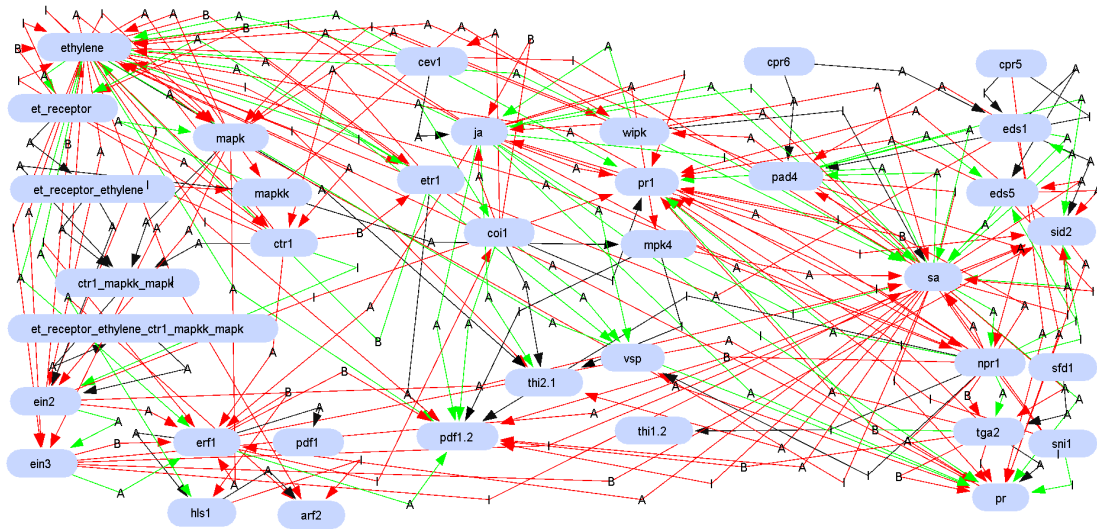


Figure 47: The final Incremented graph after two incremental steps for the use case 1. The new relations from the second set of correct triplets are shown in red. Pink arcs represent redundant transitive relations from the second set of correct triplets which are newly discovered.

The result of merging of the two input graphs is the Incremented graph with 37 nodes and 183 relations shown in Figure 47. The relations are summarised in Table 15. The removal of the redundant transitive relations which are shown in Figure 47 in pink yields the final Filtered Incremented graph which is also the final result of the first experiment.

The starting graph, constructed from the three schemata describing the SA, JA and ET pathways initially contained 49 relations. Using Bio3graph in the course of two incremental steps we have confirmed 37 relations (shown as green in Figure 47) which represent almost 75% of all relations present in the starting manually configured model structure. This shows that using Bio3graph as a starting point followed by incremental updates as new publications appear it is possible to confirm existing information but also propose new candidates (relations) for expert analysis. New candidates (shown as red arcs in Figure 47) have the potential to generate new hypothesis in biological experiments where the functionality of the link is tested.

Table 15: The summary of relations of the Incremented graph shown in Figure 47 after the second incremental step for the use case 1. The initial links originate only from the Initial graph, while the intersection, new redundant and new links originate from the Triplet graph. The intersection links are the common relations of the Initial and the Triplet graph. The new redundant links are the transitive relations while the most interesting are the new links, which represent exclusively new relations discovered by the Bio3graph tool.

Reaction types	Initial links	Intersection links	New redundant links	New links
Activation	22	26	32	34
Inhibition	9	9	1	33
Binding	0	2	0	15
All reactions	31	37	33	82

## 7.4 Use case 2: a complex plant defence model structure

To explore the potential of incrementally extending an existing, validated model using automated triplet extraction from literature we have selected a complex network structure to

complement the small-scale experiment.

In the second experiment the Initial graph was a complex plant defence model structure published in the study by (Miljkovic et al., 2012). It contains in total 175 nodes and 524 relations. Since we did not introduce new components into the graph, the vocabulary of components for the Bio3graph tool remained the same (Supporting Information S3 in (Miljkovic et al., 2012)). Also, the vocabulary of reactions was the same as in the first experiment (also available as Supporting Information S4 in (Miljkovic et al., 2012)).

The model structure, published as Supporting Information S10 in (Miljkovic et al., 2012), was used as the Initial graph and all arcs were reset to black colour. The Triplet graph was constructed from the correct triplets extracted with Bio3graph from the set of 2,988 publications which were published after the latest publication used by (Miljkovic et al., 2012) in the construction of the complex plant defence model structure. Manual validation of 399 unique triplets resulted in a set of 156 correct triplets. The Initial and the Triplet graph were merged into the Incremented graph (the summary of the relations is shown in Table 16). The evaluation of the newly discovered relations reveals that they mostly represent cross-talk connection between the SA, JA and ET pathways.

Table 16: The summary of triplets extracted from biological texts in the use case 2 (complex plant defence model structure). The initial links originate only from the Initial graph, while the intersection, new redundant and new links originate from the Triplet graph. The intersection links are the common relations of the Initial and the Triplet graph. The new redundant links are the transitive relations while the most interesting are the new links, which represent exclusively new relations discovered by the Bio3graph tool.

Reaction types	Initial links	Intersection links	New redundant links	New links
Activation	279	43	47	26
Inhibition	100	6	2	16
Binding	48	3	0	13
Produces	45	0	0	0
All reactions	472	52	49	55

While exploring the new links (red arcs) in the Incremented graph, we have observed an interesting pattern related to the discovery of binding relations. Most of the 13 new binding relations connect the components that are already connected either through activation or inhibition reaction type. The new links provide an additional explanation that first these components physically bind and then the activation or inhibition occurs.

Among the newly discovered links, biologically very interesting is (*ros*, *inhibits*, *npr1*). Earlier studies reveal that ROS, more specifically H<sub>2</sub>O<sub>2</sub>, and SA together work as a self-amplifying system (Leon et al., 1995; Rao et al., 1997). However, after consulting the publication from where the triplet was extracted (Petrov and Van Breusegem, 2012) we have found out that the new results in the study of (Peleg-Grossman et al., 2010) indicate the presence of the negative regulation of NPR1 transport by H<sub>2</sub>O<sub>2</sub>.

In addition, newly discovered triplets that are biologically interesting are (*myc2*, *inhibits*, *b-chi*) and (*myc2*, *inhibits*, *pdf1.2*) which were extracted from (Lu et al., 2013). Both links are extracted from the same sentence: “MYC2 is a negative regulator of the JA-responsive pathogen defense genes PDF1.2 and B-CHI.” In the Initial graph the relation between MYC2 and b-CHI components already exists: (*myc2*, *activates*, *b-chi*). It was acquired manually by the authors of the graph (Miljkovic et al., 2012). The discovery of the new link of a contradictory relation type indicates necessity of further exploration of the relation between MYC2 and b-CHI components. The second link (*myc2*, *inhibits*, *pdf1.2*) is also biologically interesting as it represents a cross-talk connection between JA and ET pathway where the

component of the JA pathway has diminishing influence of the product of the ET pathway.

For the final evaluation of the model structure, one should keep in mind that most of the automatically extracted relations can be considered as “indirect” and that intermediate molecules participating in the graph can be discovered by thorough inspection of the corresponding sentences or by performing additional wet-lab experiments.





## 8 Conclusions and Further Work

*Your task is not to foresee the future, but to enable it.*  
Antoine de Saint-Exupéry

The main results of this thesis are: a new methodology for constructing biological models using the expert knowledge and literature and a model of the defence response in plants, which was built by applying this methodology. Most notably, the standard approach to constructing dynamic models was upgraded with the following methods: a method for model structure revision by means of natural language processing techniques, a method for incremental development of biological model structures and a method for automatic optimisation of model parameters guided by the expert knowledge in the form of constraints.

In this chapter we summarise the most important results of the thesis. We conclude by giving directions for further work and potential improvements of the presented methodology for constructing biological models.

### 8.1 Summary

The field of systems biology, which models biological mechanisms at the system level, is faced with the rapid growth of biological data and the challenge to store, analyse and build models based on this data. However, the growth of data is not uniform for different types of biological mechanisms, hence some biological mechanisms still have a few data available. The work presented in this thesis addresses the challenge of model construction of the defence response in plants based on the expert knowledge and literature, given the lack of experimental data in this domain.

The first specific objective of this work was the manually acquired PDS model structure (described in Section 4.2) and its semi-automatically constructed extension (presented in Section 6.6). The PDS model structure was first developed in a form of a directed edge-labelled graph. The developed PDS model structure, consisting in total of 175 components and 387 reactions, contains more detailed information compared to the existing structural models of the subsets of the PDS mechanism ((Olmedo et al., 2006), (Staswick, 2008)). The complex network developed by (Naseem et al., 2012), containing 105 components and 163 reactions, is of similar size as our manually built PDS model structure. However, while our work focuses on the most important pathways of plant defence against pathogen attacks (SA, JA and ET), the study of (Naseem et al., 2012) includes several other components like GA, ABA, auxin, etc., but containing less detailed information on the particular SA, JA and ET pathways.

To set up the PDS model for the purpose of simulation, we transformed its structure to the Hybrid Functional Petri net (HFPN) formalism which resulted in the model structure containing in total 99 components and 68 reactions. The structure of one of the first simulation models of the PDS mechanism (Genoud et al., 2001), containing 18 biological entities and 12 Boolean operators, is less complex than the PDS model structure comprising three sub-models which we have developed using the HFPN formalism (see Subsection 4.2.2). A more complete PDS model is the one of (Naseem et al., 2012) with 105 components and 163

reactions. This model was built for simulation purposes, however, as mentioned above, it has a broader scope including several other components, but contains less detailed information of the SA, JA and ET pathways, which are the focus of our work. The extended PDS structure (presented in Section 6.6) contains in total 175 components and 524 reactions. This model structure is, to our knowledge, the first of this scale, comprising most of the information available on interactions between the components of SA, JA and ET pathways. Having in mind that one of the motivations to develop a PDS model was to systematize the knowledge and reveal potentially interesting links, this extended PDS structure provides a basis for such a research. We have discovered several interesting links which can be used and tested in further wet-lab experiments (Section 6.6).

A part of the PDS model, more specifically the SA sub-model, presented in the HFPN formalism, was used for the analysis of the model dynamics. The SA sub-model was developed in three iterative steps of model structure revision and constraints revision until reaching the simulation results that best satisfy the experts' expectations. The simulation curves together with the results on kinetic parameters enable deeper understanding of the plant defence mechanism. This was the second specific objective of this thesis (see Section 1.3).

The PDS model structure and the dynamic SA sub-model, were constructed in accordance with the methodology developed in this thesis (described in Chapter 3). The developed methodology was set as the third specific objective of this thesis (Section 1.3). The developed methodology aims to build the biological models without or with few experimental data. The methodology used for the PDS model construction consists of several steps. To construct the PDS model we have first defined the problem and then selected the most appropriate modelling formalism. Next, the PDS model structure was manually constructed by eliciting knowledge from the domain experts and the literature. Since manual estimation of the parameters was not feasible due to the PDS model complexity, an automatic method using a DE algorithm (Storn and Price, 1995) was developed. By iteratively applying the process, we achieved dynamic parameters' estimates that best satisfy the experts' evaluation criteria. This process included the elicitation of the knowledge from the biologists, formalisation 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. The system yielded simulation results as well as optimised model parameters, which provided an insight into the biological system. The obtained structure of the PDS model was extended by applying a semi-automated approach to extract the information on relations between the biological components from the literature. Except for the standard model construction steps, the methods that are included in this methodology are discussed below.

The main advantage of our method is its effectiveness and reusability. We have implemented the basic vocabulary and accomplished valuable results in terms of precision and recall. Its overall design allows the output of Bio3graph to be easily transferred to standard systems biology modelling formalisms. The developed Bio3graph information extraction tool can be used from scratch, without first manually building the network structure. Such an approach would be less labour-intensive, as automatic information extraction algorithms examine papers on behalf of the researchers and save experts' time from examining the full-length documents. These algorithms aim to act as research assistant tools for human experts, but they can never replace human expertise. Nevertheless, these tools always require a certain amount of effort for the manual setup of the tool (vocabulary definition like in our work, or template definition like in BioRAT (Corney et al., 2004)). Also, when the expert wants to apply the information extraction algorithm to a new domain of interest, after automatic generation of results, the expert still has to check all the results manually to assure their correctness since there is no algorithm with absolute accuracy.

The extended version of the Bio3graph tool provides efficient tracking and incremental update of existing model structures. By applying the triplet extraction incrementally on

time-labelled data, one can follow the development of knowledge about certain biological phenomena and discover new relations which can potentially enhance already developed model structures. Furthermore, the incremental extension offers the removal of transitive relations which are redundant with respect to the existing network. We have applied the extended Bio3graph method to a time-labelled collection of biomedical documents obtained from the PMC database in order to incrementally enrich two different networks. In the first use case, a simple network structure was configured from three published structural models. This network was enhanced throughout two phases which demonstrate the incremental approach. The second use case resulted in the revised PDS model structure (presented in Section 6.6). By extending this complex structure, the experts have detected several interesting links among the newly discovered relations that might be subject to further experimental validation. The results show that publicly available sources of biomedical literature, such as the PMC database, offer a good starting point for computer-assisted development of plant defence models, and that approaches such as the presented incremental method can contribute to the discovery of potentially interesting relations.

To overcome the problem of hand-tuning of model parameters, we have developed an automated method for constraint-driven model parameter optimisation. We joined expert knowledge with a mathematical optimisation approach to better understand the kinetic behaviour of the SA pathway, which is the most important pathway in plant defence. The problem of determining parameters of the SA sub-model was set up as a combinatorial optimisation problem. Our approach of defining an objective function consisting of formalised domain knowledge in the form of constraints is less time-consuming compared to the hand-tuning of model parameters. The methodology for defining constraints is similar to the one of (Locke et al., 2005). However, our approach differs from it since it is exclusively based on the domain knowledge, whereas the study mentioned above relies on the combination of the experimental data and expert knowledge. The constraint-driven optimization approach used in this work allows for an efficient exploration of the dynamic behaviour of biological models and, at the same time, increases their reliability. Moreover, we have defined the method to validate the dynamic SA sub-model based on few experimental data. In addition to that, the local sensitivity analysis of the SA sub-model has revealed its several reaction rates that biological experts found important and interesting for conducting further kinetic studies. These results correspond to the initial motives for building the PDS model which was to detect crucial reactions in the plant defence response.

This thesis also resulted in a new publicly available Bio3graph tool for relation extraction from biological texts having accomplished the forth specific objective of this thesis (Section 1.3). The advantage of our system is its public availability and simple reuse, as Bio3graph is available as a repeatable workflow in the Orange4WS publicly available data mining environment. Most importantly, by employing a different vocabulary, Bio3graph can be reused to extract the network structure of other biological mechanisms. Moreover, we have developed an incremental extension of this tool which allows easy updates of existing model structures. We expect that the extended version of the Bio3graph tool will assist the construction and enhancement of network structures that model other biological mechanisms. The obtained results show the potential of the developed method but also indicate the need for further development to improve the accuracy and utility of information extraction.

The study has offered a methodology and a semi-automatically compiled model of plant defence. However, the study encountered a number of limitations, which need to be considered. The main issue is the huge amount of manual effort that was evolved in this work. We have started by manually building the initial network structure, which required substantial human expert involvement through time-consuming acquisition and analysis of available information in the databases and literature. However, we have developed a Bio3graph approach and its incremental extension to extend the manually obtained PDS model structure based on the information extraction from literature. The Bio3graph approach to enrich the

model structure may seem error-prone, achieving 62.3 % recall and 42.6 % precision when applied on full-length articles. However, especially in the natural language processing field, the precision and recall of a specific algorithm can vary a lot depending on the text corpus that is processed (Corney et al., 2004). For example, as (Corney et al., 2004) report, the BioRAT system has a recall of 20.31 % with 55.07 % precision when applied to scientific abstracts, while the same algorithm achieves 43.6 % recall and 51.25 % precision on full-length papers. We have achieved comparable results, but have traded-off lower precision for a higher recall, as high recall is needed to extract as many new relations between biological components as possible. When estimating the parameters of the SA sub-model, we have invested manual effort and time in formulating constraints which will guide the model optimisation process. One additional limitation in our approach is that we do not have a system that will suggest new set of constraints and new model structure in order to achieve better results in the next step. On the other hand, system has means of evaluating how the current model performs dynamically which is an advantage compared to the hand-tuning of model parameters.

The methodology developed in this dissertation, despite the limitations above mentioned, when applied to the other case of biological model will be less time-consuming. It will also require less manual effort having in mind developed tools for model structure enrichment from the texts and the constraint-driven optimisation method. Moreover, we have developed a starting model structure of the plant defence and also dynamical SA sub-model, which are offered to the scientific community as a basis for further studies and refinement.

## 8.2 Further work

In this thesis, we have developed a PDS model structure, a dynamic SA sub-model and a methodology for constructing biological models based on the expert knowledge and literature. In the rest of this section, we first discuss the directions for further work related to the PDS model structure followed by a discussion of dynamic model of plant defence and the potential improvements of the developed methodology.

### PDS model structure

We plan to revise the extended PDS model structure presented in Section 6.6 in the form of a directed edge-labelled graph. This revision will also include, apart from further literature exploration, the analysis of new indirect links found by Bio3graph and detection of direct connections between biological components. This revision will also influence the PDS model structure transformed into the HFPN formalism enabling dynamic modelling. Through simulations and predictions, the identification of critical components in PDS can be achieved more efficiently. The PDS model structure in the form of the HFPN formalism will serve as a baseline for further research in the area of plant-pathogen interactions.

### Dynamic PDS model

We have developed the dynamic model of one of the three pathways of plant defence: the SA pathway. Some of the future steps in the development of the PDS model are the formalisation of constraints related to the ET and JA sub-models and parameter optimisation of the global dynamic PDS model. So defined PDS model will present a basis for further revision and fine-tuning when more experimental data become available. Moreover, it is necessary to validate the validation method to compare experimental and simulation curves that we have presented in Subsection 5.3.3. Ideally, the validation method will be deployed and evaluated in another biological domain where enough experimental data and expert knowledge exist.

## Methodology for constructing biological models

Concerning the developed methodology for constructing biological models, there are several aspects that can be addressed in the future.

To improve semi-automated model structure revision we plan to include in Bio3graph the GENIA sentence splitter (Saetre et al., 2007) which is trained on the GENIA corpus<sup>1</sup> (Kim et al., 2003) and employs a classification model based on maximum entropy modelling. Moreover, we plan to improve the triplet extraction by using fast deep parsing instead of chunking, and fine tune the rules for triplet extraction and filtering. The current implementation of Bio3graph discovers new relations, but does not enable automated discovery of new components as it employs a manually constructed vocabulary. In this version of Bio3graph it is possible to further evolve the network structure by adding new components to the vocabulary. However, we plan to implement named entity recognition and automatic discovery of synonyms which will enable automated construction of the components vocabulary.

Some graphical interfaces allow manual removal of the incorrect connections from the graph (Blaschke and Valencia, 2001). Similarly, in future work we plan to add this feature to upgrade the BioMine engine (Eronen and Toivonen, 2012). We also plan to use the developed Bio3Graph approach to upgrade the BioMine engine with an automated step of triplet extraction from literature and automated construction of the initial BioMine network, when building the BioMine database for a new domain such as plant biology, for which this resource has not yet been developed.

Regarding a method for constraint-driven parameter optimisation, we plan to investigate the possibility to implement this method in the form of a workflow. Moreover, we believe that developing a graphical user interface, which would integrate the methods for semi-automated model structure revision and incremental updates and the constraint-driven parameter optimisation, would be of a great help for the biology experts in their further work.

---

<sup>1</sup>GENIA corpus is a semantically annotated corpus for bio-textmining.



## 9 Acknowledgements

I wish to express my gratitude to my colleagues from the Jožef Stefan Institute, the thesis committee members, co-authors of my papers, collaborators from the National Institute of Biology and others who contributed to the work presented in this thesis.

First and foremost, I am deeply grateful to my supervisor Nada Lavrač. From the first coffee in Antwerp where we met she was an absolute source of motivation, inspiration, and support. She actively encouraged me in developing my own ideas and provided me with directions whenever I got stuck with a problem. She was also my moral support, not only in work issues. Thanks to her, I will never regret starting life from the very beginning when moving to Ljubljana.

My co-supervisor Igor Mozetič was one of the key figures in my work since the start of this thesis. His advice and comments were of great help for my work, especially in the beginning of my PhD journey. Among the co-authors on my papers, I want to specially thank Vid Podpečan who was the most reliable and helpful colleague that I have ever met. I am also thankful to him for reading the first "unreadable" drafts of my thesis and for his help with the translation to Slovene.

Many thanks to the colleagues from the National Institute of Biology, especially to Kristina Gruden, Tjaša Stare and Marko Petek. This thesis would not be possible without fruitful discussions with them and their enthusiasm in our collaboration. Apart from successful collaborative work, I will never forget home-made lunches on Fridays and the fact that I even had my own coffee cup at NIB.

For extremely helpful discussions, ideas and support during my PhD studies, I am thankful to Ljupčo Todorovski, Neža Mramor-Kosta, Bojan Orel, Bogdan Filipič and Matjaž Depolli.

It was a great pleasure to meet all my department colleagues, with whom I found the common language from the very first day and not only in work-related issues. I am thankful to them for the moral support, participation and ideas throughout my PhD studies, including the brownbag seminars. I enjoyed a lot our lunches "na bon", coffee breaks in the park, cycling in sunny weekends and "beer Fridays". Without Mili and Tina during these years nothing would function so well - from daily support up to best Christmas parties ever. The best way to start a day is to start it with jokes of Jolanda. My kitchen and corridor colleagues - Panče, Petra, Bernard, Martin, Bogdan, Violeta, made the hard-working days easier. My ex-office mates Darko Č., Valentin, Jovan, Darko A. and Aneta and colleagues from S39 office - Matjaž, Borut, Miha, Vid were my daily support. A huge impact on marvellous atmosphere at work and my private life had Ivica, Dragi, Nikola and Vlada. I am extremely grateful to my current office mate Nejc for the help with the final draft of the thesis and for daily support. Finally, my best friend here Biljana was with me in good and bad all this time: she made so much easier my "crisis" moments, which, I must admit, required a lot of patience from her side, and she was with me to share my happiness.

I am also very grateful to my Belgian friends. They understood very well all the problems that a PhD student is going through. I appreciate all the friendly support that I have got in my PhD time from Ozlem, Kristien, Kim, Pieter, Toon, Frederik, Miroslav, Ivan, Bogdan, Zenya, Olya, Sasha and the others. My incredible friends in Belgrade also had a huge

influence on my work. My great friend Branka knows how much her encouraging meant to me during the last four years. Ivana is probably not even aware how much she helped me by listening to me during my home visits. I will remember always Mira's special pancakes. Tanja, Dragana, Ivan, Maria, Milica, Jelena, Vladan, Jessy thank you for your understanding and for cheering me up.

The work on this dissertation would not have been possible without the grant of AD Futura Agency, Slovenian Research Agency grants P4-0165, J4-2228, J4-4165 and P2-0103.

Finally, I am sincerely grateful to my family for their support, understanding and patience. My tremendous brother taught me how to solve life problems and my cousin Sanja assured me I can accomplish everything. My deepest thanks go to my mum and dad whose love and help gave me the strength to achieve this goal.



## 10 References

- Albert, R. *Boolean Modeling of Genetic Regulatory Networks*, 459–481. Lecture Notes in Physics (Springer Berlin Heidelberg, 2004).
- Alm, E.; Arkin, A. P. Biological networks. *Current opinion in structural biology* **13**, 193–202 (2003).
- Alter, C.; Murty, S. Logic modeling: A tool for teaching practice evaluation. *Journal of Social Work Education* **33**, 103–117 (1997).
- Ananiadou, S.; Pyysalo, S.; Tsujii, J.; Kell, D. B. Event extraction for systems biology by text mining the literature. *Trends in Biotechnology* **28**, 381–390 (2010).
- Arnold, L. *Stochastic Differential Equations: Theory and Applications*. *Dover Books on Mathematics Series* (Dover Publications, 2013).
- Baebler, v.; Stare, K.; Kovač, M.; Blejec, A.; Prezelj, N.; Stare, T.; Kogovšek, P.; Pompe-Novak, M.; Rosahl, S.; Ravnikar, M.; Gruden, K. Dynamics of responses in compatible potato - potato virus y interaction are modulated by salicylic acid. *PLoS ONE* **6**, e29009 (2011).
- Barnes, J.; Hut, P. A hierarchical  $O(n \log n)$  force-calculation algorithm. *Nature* **324**, 446–449 (1986).
- Bean, C. *Introduction to Nucleic Acids* (Health Sciences Consortium, 1973).
- Bird, S.; Loper, E.; Klein, E. *Natural Language Processing with Python* (O'Reilly Media Inc, 2009).
- Blaschke, C.; Andrade, M.; Ouzounis, C.; Valencia, A. Automatic extraction of biological information from scientific text: protein-protein interactions. In: *Proceedings of International Conference on Intelligent Systems for Molecular Biology*. 60–67 (1999).
- Blaschke, C.; Valencia, A. The potential use of suiseki as a protein interaction discovery tool. *Genome Informatics* **12**, 123–34 (2001).
- Blaschke, C.; Valencia, A. The frame-based module of the suiseki information extraction system. *Ieee Intelligent Systems* **17**, 14–20 (2002).
- Boggs, P.; Tolle, J. Sequential quadratic programming. *Acta Numerica* **4**, 1–51 (1995).
- Boutet, E.; Lieberherr, D.; Tognolli, M.; Schneider, M.; Bairoch, A. Uniprotkb/swiss-prot. *Methods in molecular biology* **406**, 89–112 (2007).
- Branden, J., C. Tooze. *Introduction to Protein Structure* (Garland Science, 1999).
- Browse, J. Jasmonate passes muster: A receptor and targets for the defense hormone. *Annual Review of Plant Biology* **60**, 183–205 (2009).

- Butcher, J. *Numerical Methods for Ordinary Differential Equations* (Wiley, 2008).
- Cantone, I. *A yeast synthetic network for In-vivo Reverse-engineering and Modelling Assessment (IRMA)*. Ph.D. thesis, EUROPEAN SCHOOL OF MOLECULAR MEDICINE NAPLES SITE (2009).
- Cara, A. D.; Garg, A.; Micheli, G. D.; Xenarios, I.; Mendoza, L. Dynamic simulation of regulatory networks using squad. *BMC Bioinformatics* **8**, 462 (2007).
- Carr, J. P.; Lewsey, M. G.; Palukaitis, P. Signaling in induced resistance. *Natural and Engineered Resistance to Plant Viruses, Pt Ii* **76**, 57–121 (2010).
- Chandra-Shekara, A. C.; Navarre, D.; Kachroo, A.; Kang, H. G.; Klessig, D.; Kachroo, P. Signaling requirements and role of salicylic acid in hrt- and rrt-mediated resistance to turnip crinkle virus in arabidopsis. *Plant Journal* **40**, 647–659 (2004).
- Chaouiya, C. Petri net modelling of biological networks. *Briefings in Bioinformatics* **8**, 210–219 (2007).
- Chassagnole, C.; Noisommit-Rizzi, N.; Schmid, J. W.; Mauch, K.; Reuss, M. Dynamic modeling of the central carbon metabolism of escherichia coli. *Biotechnology and bioengineering* **79**, 53–73 (2002).
- Chen, H.; Sharp, B. M. Content-rich biological network constructed by mining pubmed abstracts. *Bmc Bioinformatics* **5** (2004). Chen, H Sharp, BM.
- Chen, H. M.; Xue, L.; Chintamanani, S.; Germain, H.; Lin, H. Q.; Cui, H. T.; Cai, R.; Zuo, J. R.; Tang, X. Y.; Li, X.; Guo, H. W.; Zhou, J. M. Ethylene insensitive3 and ethylene insensitive3-like1 repress salicylic acid induction deficient2 expression to negatively regulate plant innate immunity in arabidopsis. *Plant Cell* **21**, 2527–2540 (2009).
- Chini, A.; Fonseca, S.; Fernandez, G.; Adie, B.; Chico, J. M.; Lorenzo, O.; Garcia-Casado, G.; Lopez-Vidriero, I.; Lozano, F. M.; Ponce, M. R.; Micol, J. L.; Solano, R. The jaz family of repressors is the missing link in jasmonate signalling. *Nature* **448**, 666–U4 (2007).
- Chu, Y.; Jayaraman, A.; Hahn, J. Parameter sensitivity analysis of il-6 signalling pathways. *IET Systems Biology* **1**, 342–352 (2007).
- Clarke, S. M.; Mur, L. A. J.; Wood, J. E.; Scott, I. M. Salicylic acid dependent signaling promotes basal thermotolerance but is not essential for acquired thermotolerance in arabidopsis thaliana. *Plant Journal* **38**, 432–434 (2004).
- Coddington, E. A.; Levinson, N. *Theory of Ordinary Differential Equations* (McGraw-Hill, New York, 1955).
- Cohen, K. B.; Hunter, L. Getting started in text mining. *Plos Computational Biology* **4** (2008).
- Cohen, L., K.B. Hunter. Natural language processing and systems biology. *In Artificial Intelligence Methods and Tools for Systems Biology* **5**, 147–174 (2004).
- Consortium, A. I. M. Evidence for network evolution in an arabidopsis interactome map. *Science* **333**, 601–607 (2011).
- Corney, D. P. A.; Buxton, B. F.; Langdon, W. B.; Jones, D. T. Biorat: extracting biological information from full-length papers. *Bioinformatics* **20**, 3206–3213 (2004).

- Craven, M.; Kumlien, J. Constructing biological knowledge bases by extracting information from text sources. In: *Proceedings of International Conference on Intelligent Systems for Molecular Biology*. 77–86 (AAAI Press, 1999).
- Cui, J.; Li, P.; Li, G.; Xu, F.; Zhao, C.; Li, Y.; Yang, Z.; Wang, G.; Yu, Q.; Li, Y.; Shi, T. Atpid: Arabidopsis thaliana protein interactome database—an integrative platform for plant systems biology. *Nucleic Acids Research* **36**, D999–1008 (2008).
- de Hoon, M. J. L.; Imoto, S.; Kobayashi, K.; Ogasawara, N.; Miyano, S. Inferring gene regulatory networks from time-ordered gene expression data of *Bacillus subtilis* using differential equations. *Pacific Symposium on Biocomputing* 17–28 (2003).
- Dempsey, D.; Vlot, A.; Wildermuth, M.; Klessig, D. Salicylic acid biosynthesis and metabolism. *Arabidopsis book* **9**, e0156 (2011).
- Devoto, A.; Turner, J. G. Jasmonate-regulated arabidopsis stress signalling network. *Physiologia Plantarum* **123**, 161–172 (2005).
- Donaldson, I.; Martin, J.; de Bruijn, B.; Wolting, C.; Lay, V.; Tuekam, B.; Zhang, S.; Baskin, B.; Bader, G.; Michalickova, K.; Pawson, T.; Hogue, C. Prebind and textomy - mining the biomedical literature for protein-protein interactions using a support vector machine. *BMC Bioinformatics* **4** (2003).
- Dubitzky, W.; Kötter, T.; Schmidt, O.; Berthold, M. *Bisociative knowledge discovery : An introduction to concept, algorithms, tools, and applications*, chapter Towards creative information exploration based on Koestler’s concept of bisociation, 11–32 (Springer, 2012).
- Džeroski, S.; Todorovski, L. Equation discovery for systems biology: finding the structure and dynamics of biological networks from time course data. *Current Opinion in Biotechnology* **19**, 360–368 (2008).
- Dym, C. L. *Principles of Mathematical Modeling* (Academic Press, 2004).
- Eiben, A. E.; Smith, J. E. *Introduction to Evolutionary Computing* (Springer-Verlag, Berlin, 2003).
- Eronen, L.; Toivonen, H. Biomine: predicting links between biological entities using network models of heterogeneous databases. *BMC Bioinformatics* **13**, 119 (2012).
- Estrada, E. *The Structure of Complex Networks: Theory and Applications* (Oxford University Press, 2011).
- Fernández-Calvo, P.; Chini, A.; Fernández-Barbero, G.; Chico, J. M.; Gimenez-Ibanez, S.; Geerinck, J.; Eeckhout, D.; Schweizer, F.; Godoy, M.; Franco-Zorrilla, J. M.; Pauwels, L.; Witters, E.; Puga, M. I.; Paz-Ares, J.; Goossens, A.; Reymond, P.; De Jaeger, G.; Solano, R. The arabidopsis bhlh transcription factors myc3 and myc4 are targets of jaz repressors and act additively with myc2 in the activation of jasmonate responses. *Plant Cell* **23**, 701–715 (2011).
- Filipič, B.; Depolli, M. Parallel evolutionary computation framework for single- and multi-objective optimization. *Parallel Computing* 217–240 (2009).
- Fokkink, W. *Introduction to Process Algebra* (Springer-Verlag, 2007).
- Fonseca, S.; Chico, J. M.; Solano, R. The jasmonate pathway: the ligand, the receptor and the core signalling module. *Current Opinion in Plant Biology* **12**, 539 – 547 (2009).

- Fritzson, P. *Principles of Object-Oriented Modeling and Simulation with Modelica 2.1* (Wiley-IEEE Press, 2004).
- Fu, Z.; Yan, S.; Saleh, A.; Wang, W.; Ruble, J.; Oka, N.; Mohan, R.; Spoel, S.; Tada, Y.; Zheng, N.; Dong, X. Npr3 and npr4 are receptors for the immune signal salicylic acid in plants. *Nature* **486**, 228–32 (2012).
- Genoud, T.; Métraux, J.-P. Crosstalks in plant cell signaling: structure and function of the genetic network. *Trends in Plant Science* **4**, 503–507 (1999).
- Genoud, T.; Santa Cruz, M. B. T.; Metraux, J. P. Numeric simulation of plant signaling networks. *Plant Physiology* **126**, 1430–1437 (2001).
- Gfeller, A.; Liechti, R.; Farmer, E. E. Arabidopsis jasmonate signaling pathway. *Science Signaling* **3** (2010).
- Giersch, C.; Sivak, M.; Walker, D. Proceedings of the royal society of london : Series b : Biological sciences. *A mathematical skeleton model of photosynthetic oscillations* **245**, 77–83 (1991).
- Glazebrook, J. Contrasting mechanisms of defense against biotrophic and necrotrophic pathogens. *Annual Review of Phytopathology* **43**, 205–227 (2005).
- Gong, H.; Zuliani, P.; Komuravelli, A.; Faeder, J. R.; Clarke, E. M. Analysis and verification of the hmgb1 signaling pathway. *BMC Bioinformatics* **11**, S10 (2010).
- Gonzalez-Garcia, J. S.; Diaz, J. Information theory and the ethylene genetic network. *Plant Signal Behav* **6**, 1483–1498 (2011).
- Goss, P.; Peccoud, J. Quantitative modeling of stochastic systems in molecular biology by using stochastic petri nets. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 6750–6755 (1998).
- Hahn, B. D. A mathematical model of photorespiration and photosynthesis. *Annals of Botany* **60**, 157–169 (1987).
- Harel, D. Statecharts: A visual formalism for complex systems. *Science of Computer Programming* **8**, 231–274 (1987).
- Hasegawa, T.; Sekine, S.; Grishman, R. Discovering relations among named entities from large corpora. In: *Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics*. ACL '04 (Association for Computational Linguistics, Stroudsburg, PA, USA, 2004).
- Hawari, A. H.; Mohamed-Hussein, Z. A. Simulation of a petri net-based model of the terpenoid biosynthetic pathway. *BMC Bioinformatics* **11**, 83 (2010).
- Henzinger, T. A. The theory of hybrid automata. In: *Proceedings of 11th Annual IEEE Symposium on Logic in Computer Science*. 278–292 (IEEE Computer Society Press, 1996).
- Hofestädt, R.; Thelen, S. Quantitative modeling of biochemical networks. *In Silico Biology* **1**, 39–53 (1998).
- Hoffmann, R.; Valencia, A. A gene network for navigating the literature. *Nature Genetics* **36**, 664–664 (2004).
- Hogeweg, P.; Hesper, B. Interactive instruction on population interactions. *Computers in biology and medicine* **8**, 319–327 (1978).

- Ibdah, M.; Chen, Y. T.; Wilkerson, C. G.; Pichersky, E. An aldehyde oxidase in developing seeds of arabidopsis converts benzaldehyde to benzoic acid. *Plant Physiology* **150**, 416–423 (2009).
- Ibe, O. C. *Fundamentals of Stochastic Networks*, chapter Boolean Networks, 312 (John Wiley & Sons, 2011).
- Jeong, R. D.; Zhu, S.; Kachroo, A.; Kachroo, P. New insights into resistance protein-mediated signaling against turnip crinkle virus in arabidopsis. *Journal of Plant Biochemistry and Biotechnology* **21**, 48–51 (2012).
- Kanehisa, M.; Goto, S. Kegg: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Research* **28**, 27–30 (2000).
- Kapur, J. N. *Mathematical modelling* (New Age International Pvt Ltd Publishers, 2008).
- Katari, M.; Nowicki, S.; Aceituno, F.; Nero, D.; Kelfer, J.; Thompson, L.; Cabello, J.; Davidson, R.; Goldberg, A.; Shasha, D.; Coruzzi, G.; Gutierrez, R. Virtualplant: A software platform to support systems biology research. *Plant Physiology* **152**, 500–515 (2010).
- Keener, J. P.; Sneyd, J. *Mathematical physiology: I: Cellular physiology. Interdisciplinary applied mathematics* (Springer, New York, London, 2009).
- Kendrick, M. D.; Chang, C. Ethylene signaling: new levels of complexity and regulation. *Current Opinion in Plant Biology* **11**, 479–485 (2008).
- Kestler, H. A.; Wawra, C.; Kracher, B.; Kuhl, M. Network modeling of signal transduction: establishing the global view. *Bioessays* **30**, 1110–1125 (2008).
- Kim, J. D.; Ohta, T.; Tateisi, Y.; Tsujii, J. Genia corpus-semantically annotated corpus for bio-textmining. *Bioinformatics* **19 Suppl 1**, i180–182 (2003).
- Kirkpatrick, S.; Gelatt, C. D.; Vecchi, M. P. Optimization by simulated annealing. *Science* **220**, 671–680 (1983).
- Kiss, T.; Strunk, J. Unsupervised multilingual sentence boundary detection. *Computational Linguistics* **32**, 485–525 (2006).
- Kitano, H. (ed.) *Foundations of Systems Biology* (MIT Press, 2001).
- Kitano, H. Systems biology: A brief overview. *Science* **295**, 1662–1664 (2002).
- Kleijnen, J. P. C. Verification and validation of simulation models. *European Journal of Operational Research* **82**, 145–162 (1995).
- Klessig, D. F.; Malamy, J. The salicylic acid signal in plants. *Plant Molecular Biology* **26**, 1439–1458 (1994).
- Kohavi, R. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: *Proceedings of the 14th international joint conference on Artificial intelligence - Volume 2*. 1137–1143 (Morgan Kaufmann, 1995).
- Köhler, J.; Baumbach, J.; Taubert, J.; Specht, M.; Skusa, A.; Ruegg, A.; Rawlings, C.; Verrier, P.; Philippi, S. Graph-based analysis and visualization of experimental results with ondex. *Bioinformatics* **22**, 1383–1390 (2006).
- Koornneef, A.; Pieterse, C. M. J. Cross talk in defense signaling. *Plant Physiology* **146**, 839–844 (2008).

- Koutsoukos, X.; Antsaklis, P.; Stiver, J.; Lemmon, M. Supervisory control of hybrid systems. In: *Proceedings of IEEE*. 1026–1049 (2000).
- Krallinger, M.; Rodriguez-Penagos, C.; Tendulkar, A.; Valencia, A. Plan2l: a web tool for integrated text mining and literature-based bioentity relation extraction. *Nucleic Acids Research* **37**, W160–W165 (2009).
- Krummenacker, M.; Paley, S.; Mueller, L.; Yan, T.; Karp, P. D. Querying and computing with biocyc databases. *Bioinformatics* **21**, 3454–3455 (2005).
- Kunkel, D., B.N. Brooks. Cross talk between signaling pathways in pathogen defence. *Current Opinion in Plant Biology* **5**, 325–331 (2002).
- Lai, Z.; Wang, F.; Zheng, Z.; Fan, B.; Chen, Z. A critical role of autophagy in plant resistance to necrotrophic fungal pathogens. *The Plant journal : for cell and molecular biology* **66**, 953–968 (2011).
- Laisk, A.; Walker, D. A. A mathematical model of electron transport. thermodynamic necessity for psii regulation–‘light stomata’. *Proceedings of the Royal Society London, Series B* **237**, 417–444 (1989).
- Le Novère, N.; Hucka, M.; et al. The systems biology graphical notation. *Nature Biotechnology* **27**, 735–741 (2009).
- Leon, J.; Lawton, M. A.; Raskin, I. Hydrogen peroxide stimulates salicylic acid biosynthesis in tobacco. *Plant Physiol.* **108**, 1673–1678 (1995).
- Levenberg, K. A method for the solution of certain non-linear problems in least squares. *Quarterly of Applied Mathematics* **2**, 164–168 (1944).
- Li, J.; Brader, G.; Palva, E. T. The wrky70 transcription factor: A node of convergence for jasmonate-mediated and salicylate-mediated signals in plant defense. *Plant Cell* **16**, 319–331 (2004).
- Locke, J.; Millar, A.; Turner, M. Modelling genetic networks with noisy and varied experimental data: the circadian clock in arabidopsis thaliana. *Journal of Theoretical Biology* **234**, 383–393 (2005).
- Lu, X.; Jiang, W.; Zhang, L.; Zhang, F.; Zhang, F.; Shen, Q.; Wang, G.; Tang, K. Aaerfl positively regulates the resistance to botrytis cinerea in artemisia annua. *PLoS ONE* **8**, e57657 (2013).
- Maier, F.; Zwicker, S.; Huckelhoven, A.; Meissner, M.; Funk, J.; Pfitzner, A. J. P.; Pfitzner, U. M. Nonexpressor of pathogenesis-related proteins1 (npr1) and some npr1-related proteins are sensitive to salicylic acid. *Molecular Plant Pathology* **12**, 73–91 (2011).
- Marquardt, D. An algorithm for least-squares estimation of nonlinear parameters. *SIAM Journal on Applied Mathematics* **11**, 431–441 (1963).
- Marsan, M. A. Stochastic petri nets: An elementary introduction. In: *In Advances in Petri Nets*. 1–29 (Springer, 1989).
- Matsuno, H.; Doi, A.; Nagasaki, M.; Miyano, S. Hybrid petri net representation of gene regulatory network. *Pacific Symposium on Biocomputing* 341–352 (2000).
- Matsuno, H.; Tanaka, Y.; Aoshima, H.; Doi, A.; Matsui, M.; Miyano, S. Biopathways representation and simulation on hybrid functional petri net. *In Silico Biology* **3**, 389–404 (2003).

- McHaney, R. *Understanding Computer Simulation* (Bookboon, 2009).
- Michaelis, L.; Menten, M. Die kinetik der invertinwirkung. *Biochemistry Zeitung* **49**, 333–369 (1913).
- Miljkovic, D.; Stare, T.; Mozetič, I.; Podpečan, V.; Petek, M.; Witek, K.; Dermastia, M.; Lavrač, N.; Gruden, K. Signalling network construction for modelling plant defence response. *PLOS ONE* **7**, e51822–1e51822–18 (2012).
- Mitchell, M. *An Introduction to Genetic Algorithms* (MIT Press, Cambridge, MA, 1996).
- Moore, J. W.; Loake, G. J.; Spoel, S. H. Transcription dynamics in plant immunity. *Plant Cell* **23**, 2809–2820 (2011).
- Morgan, J. A.; Rhodes, D. Mathematical modeling of plant metabolic pathways. *Metabolic Engineering* **4**, 80–89 (2002).
- Mueller, L. A.; Zhang, P. F.; Rhee, S. Y. Aracyc: A biochemical pathway database for arabidopsis. *Plant Physiology* **132**, 453–460 (2003).
- Murata, T. Petri nets: Properties, analysis and applications. *Proceedings of the IEEE* **77**, 541–580 (1989).
- Nagasaki, M.; Doi, A.; Matsuno, H.; S., M. Genomic object net.i: a platform for modeling and simulating biopathways. *Applied Bioinformatics* **2**, 181–184 (2004).
- Naseem, M.; Philippi, N.; Hussain, A.; Wangorsch, G.; Ahmed, N.; Dandekar, T. Integrated systems view on networking by hormones in arabidopsis immunity reveals multiple crosstalk for cytokinin. *Plant Cell* **24**, 1793–1814 (2012).
- Noselli, S.; Perrimon, N. Are there close encounters between signaling pathways? *Science* **290**, 68–69 (2000).
- Olmedo, G.; Guo, H. W.; Gregory, B. D.; Nourizadeh, S. D.; Aguilar-Henonin, L.; Li, H. J.; An, F. Y.; Guzman, P.; Ecker, J. R. Ethylene-insensitive5 encodes a 5' → 3' exonuclease 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* **103**, 13286–13293 (2006).
- Ono, T.; Hishigaki, H.; Tanigami, A.; Takagi, T. Automated extraction of information on protein-protein interactions from the biological literature. *Bioinformatics* **17**, 155–161 (2001).
- Peleg-Grossman, S.; Melamed-Book, N.; Cohen, G.; Levine, A. Cytoplasmic H<sub>2</sub>O<sub>2</sub> prevents translocation of NPR1 to the nucleus and inhibits the induction of PR genes in Arabidopsis. *Plant Signal Behav* **5**, 1401–1406 (2010).
- Petrov, V. D.; Van Breusegem, F. Hydrogen peroxide-a central hub for information flow in plant cells. *AoB Plants* **2012**, pls014 (2012).
- Pettersson, G.; Ryde-Pettersson, U. A mathematical model of the calvin photosynthesis cycle. *European Journal of Biochemistry* **175**, 661–672 (1988).
- Pettinen, A.; Aho, T.; Smol, O.; Manninen, T.; Saarinen, A.; Yli-harja, O.; Linne, M. Simulation tools for biochemical networks: evaluation of performance and usability. *Bioinformatics* **21**, 357–363 (2005).
- Pieterse, C.; Ton, J.; Van Loon, L. Cross-talk between plant defense signalling pathways: boost or burden? *AgBiotechNet* **3**, 1–8 (2001).

- Pieterse, C. M. J.; Leon-Reyes, A.; Van der Ent, S.; Van Wees, S. C. M. Networking by small-molecule hormones in plant immunity. *Nature Chemical Biology* **5**, 308–316 (2009).
- Pinzón, A.; Barreto, E.; Bernal, A.; Achenie, L.; Barrios, A. F.; Isea, R.; Restrepo, S. Computational models in plant-pathogen interactions: the case of phytophthora infestans. *Theoretical biology & medical modelling* **6**, 24–35 (2009).
- Pitzschke, A.; Schikora, A.; Hirt, H. Mapk cascade signalling networks in plant defence. *Current Opinion in Plant Biology* **12**, 421–426 (2009).
- Podpečan, V.; Zemenova, M.; Lavrač, N. Orange4ws environment for service-oriented data mining. *Computer Journal* **55**, 82–98 (2012).
- Poolman, M.; Ölcer, H.; Lloyd, J. C.; Raines, C. A.; Fell, D. A. Computer modelling and experimental evidence for two steady states in the photosynthetic calvin cycle. *European Journal of Biochemistry* **268**, 2810–2816 (2001).
- Rao, M. V.; Paliyath, G.; Ormrod, D. P.; Murr, D. P.; Watkins, C. B. Influence of salicylic acid on H<sub>2</sub>O<sub>2</sub> production, oxidative stress, and H<sub>2</sub>O<sub>2</sub>-metabolizing enzymes. Salicylic acid-mediated oxidative damage requires H<sub>2</sub>O<sub>2</sub>. *Plant Physiol.* **115**, 137–149 (1997).
- Raza, S.; McDerment, N.; Lacaze, P. A.; Robertson, K.; Watterson, S.; Chen, Y.; Chisholm, M.; Eleftheriadis, G.; Monk, S.; O’Sullivan, M.; Turnbull, A.; Roy, D.; Theodoridis, A.; Ghazal, P.; Freeman, T. C. Construction of a large scale integrated map of macrophage pathogen recognition and effector systems. *Bmc Systems Biology* **4** (2010).
- Raza, S.; Robertson, K. A.; Lacaze, P. A.; Page, D.; Enright, A. J.; Ghazal, P.; Freeman, T. C. A logic-based diagram of signalling pathways central to macrophage activation. *Bmc Systems Biology* **2** (2008).
- Reddy, V. N.; Liebman, M. N.; Mavrovouniotis, M. L. Qualitative analysis of biochemical reaction systems. *Computers in Biology and Medicine* **6**, 9–24 (1996).
- Regev, A.; Silverman, W.; Shapiro, E. Representation and simulation of biochemical processes using the pi-calculus process algebra. *Pacific Symposium on Biocomputing* 459–470 (2001).
- Renardy, M.; Rogers, R. *An Introduction to Partial Differential Equations. Texts in Applied Mathematics* (Springer, 2004).
- Reymond, P.; Farmer, E. E. Jasmonate and salicylate as global signals for defense gene expression. *Current Opinion in Plant Biology* **1**, 404–411 (1998).
- Rios-Esteva, R.; Lange, B. M. Experimental and mathematical approaches to modeling plant metabolic networks. *Phytochemistry* **68**, 2351–2374 (2007).
- Robert-Seilanianantz, A.; Grant, M.; Jones, J. Hormone crosstalk in plant disease and defense: more than just jasmonate-salicylate antagonism. *Annual review of phytopathology* **49**, 317–343 (2011).
- Rzhetsky, A.; Iossifov, I.; Koike, T.; Krauthammer, M.; Kra, P.; Morris, M.; Yu, H.; Duboue, P. A.; Weng, W. B.; Wilbur, W. J.; Hatzivassiloglou, V.; Friedman, C. Geneways: a system for extracting, analyzing, visualizing, and integrating molecular pathway data. *Journal of Biomedical Informatics* **37**, 43–53 (2004).
- Saetre, R.; Yoshida, K.; Yakushiji, A.; Miyao, Y.; Matsubayashi, Y.; Ohta, T. Akane system: Protein-protein interaction pairs in the biocreative2 challenge, ppi-ips subtask. In: Hirschman, L.; Krallinger, M.; Valencia, A. (eds.) *Proceedings of the Second BioCreative Challenge Workshop* (2007).



- Sargent, R. G. Verification and validation of simulation models. In: *Proceedings of Winter Simulation Conference*. 55–64 (1998).
- Sarkar, A., S. Plutynski (ed.) *A Companion to the Philosophy of Biology* (Oxford: Blackwell, 2008).
- Sato, M.; Tsuda, K.; Wang, L.; Collier, J.; Watanabe, Y.; Glazebrook, J.; Katagiri, F. Network modeling reveals prevalent negative regulatory relationships between signaling sectors in arabidopsis immune signaling. *Plos Pathogens* **6** (2010).
- Sayers, E. A general introduction to the e-utilities. In: *Entrez Programming Utilities Help* (National Center for Biotechnology Information, Bethesda, Maryland, US, 2011).
- Sevon, P.; Eronen, L.; Hintsanen, P.; Kulovesi, K.; Toivonen, H. Link discovery in graphs derived from biological databases. In: *3rd International Workshop on Data Integration in the Life Sciences*. **4705**, 35–49 (Springer, 2006).
- Shah, J. The salicylic acid loop in plant defense. *Current Opinion in Plant Biology* **6**, 365–371 (2003).
- Shin, Y.-J.; Nourani, M. Statecharts for gene network modeling. *PLoS ONE* **5**, e9376 (2010).
- Silva, M.; Recalde, L. On fluidification of petri net models: from discrete to hybrid and continuous models. *Annual Reviews in Control* **28**, 253–266 (2004).
- Somssich, I. E.; Hahlbrock, K. Pathogen defence in plants - a paradigm of biological complexity. *Trends in Plant Science* **3**, 86–90 (1998).
- Staswick, P. E. Jazging up jasmonate signaling. *Trends in Plant Science* **13**, 66–71 (2008).
- Staswick, P. E.; Tiryaki, I. The oxylipin signal jasmonic acid is activated by an enzyme that conjugates it to isoleucine in arabidopsis. *Plant Cell* **16**, 2117–2127 (2004).
- Stepanova, A.; Alonso, J. Arabidopsis ethylene signaling pathway. *Science Signaling* cm4 (2005).
- Storn, R.; Price, K. *Differential evolution - a simple and efficient adaptive scheme for global optimization over continuous spaces* (1995).
- 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.; Huala, E. The arabidopsis information resource (tair): gene structure and function annotation. *Nucleic Acids Research* **36**, D1009–D1014 (2008).
- Takai-Igarashi, T. Ontology based standardization of petri net modeling for signaling pathways. *In Silico Biology* **5**, 529–536 (2005).
- Thakur, A. K. *Model mechanistic vs empirical*, 41–51 (Plenum Press, New York, 1991), new trends in pharmacokinetics edition.
- Tsesmetzis, N.; Couchman, M.; Higgins, J.; Smith, A.; Doonan, J. H.; Seifert, G. J.; Schmidt, E. E.; Vastrik, I.; Birney, E.; Wu, G. M.; D'Eustachio, P.; Stein, L. D.; Morris, R. J.; Bevan, M. W.; Walsh, S. V. Arabidopsis reactome: A foundation knowledgebase for plant systems biology. *Plant Cell* **20**, 1426–1436 (2008).
- Tsuruoka, Y.; Tateishi, Y.; Kim, J. D.; Ohta, T.; McNaught, J.; Ananiadou, S.; Tsujii, J. Developing a robust part-of-speech tagger for biomedical text. *Advances in Informatics, Proceedings* **3746**, 382–392 (2005).

- Tsuruoka, Y.; Tsujii, J. Bidirectional inference with the easiest-first strategy for tagging sequence data. In: *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*. 467–474 (2005).
- Turner, J. G.; Ellis, C.; Devoto, A. The jasmonate signal pathway. *The Plant Cell Online* **14**, S153–S164 (2002).
- Twycross, J.; Band, L. R.; Bennett, M. J.; King, J. R.; Krasnogor, N. Stochastic and deterministic multiscale models for systems biology: an auxin-transport case study. *BMC Systems Biology* 4–34 (2010).
- Uknes, S.; A., W.; Delaney, T.; Vernooij, B.; Morse, A.; Friedrich, L.; Nye, G.; Potter, S.; Ward, E.; Ryals, J. Biological induction of systemic acquired resistance in arabidopsis. *Molecular Plant-Microbe Interactions* **6**, 692–698 (1993).
- Vlot, A. C.; Dempsey, D. A.; Klessig, D. F. Salicylic acid, a multifaceted hormone to combat disease. *Annual Review of Phytopathology* **47**, 177–206 (2009).
- Wang, K. L. C.; Li, H.; Ecker, J. R. Ethylene biosynthesis and signaling networks. *Plant Cell* **14**, S131–S151 (2002).
- Wildermuth, M. C.; Dewdney, J.; Wu, G.; Ausubel, F. M. Isochorismate synthase is required to synthesize salicylic acid for plant defence. *Nature* **414**, 562–565 (2001).
- Wuchty, S.; Ravasz, E.; Barabási, A. *The Architecture of Biological Networks*, 165–181. Topics in Biomedical Engineering International Book Series (Springer US, 2006).
- Yang, C.; Guo, R.; Jie, F.; Nettleton, D.; Peng, J.; Carr, T.; Yeakley, J.; Fan, J.; Whitham, S. Spatial analysis of arabidopsis thaliana gene expression in response to turnip mosaic virus infection. *Molecular Plant-Microbe Interactions* **20**, 358–370 (2007).
- Zhang, P.; Dreher, K.; Karthikeyan, A.; Chi, A.; Pujar, A.; Caspi, R.; Karp, P.; Kirkup, V.; Latendresse, M.; Lee, C.; Mueller, L. A.; Muller, R.; Rhee, S. Y. Creation of a genome-wide metabolic pathway database for populus trichocarpa using a new approach for reconstruction and curation of metabolic pathways for plants. *PLANT PHYSIOLOGY* **153**, 1479–1491 (2010).
- Zhao, Q.; Guo, H. W. Paradigms and paradox in the ethylene signaling pathway and interaction network. *Molecular Plant* **4**, 626–634 (2011).
- Zhu, X.; Gerstein, M.; Snyder, M. Getting connected: analysis and principles of biological networks. *Genes & Development* **21**, 1010–1024 (2007).

## Publications Related to the Dissertation

This section lists all publications related to the thesis. It consists of published material in journals, conferences, workshops and as book chapters which are directly related to the topic of the thesis.

### Original scientific article (1.01)

Miljkovic, D.; Stare, T.; Mozetič, I.; Podpečan, V.; Petek, M.; Witek, K.; Dermastia, M.; Lavrač, N.; Gruden, K. Signalling network construction for modelling plant defence response. *PLOS ONE* **7**, e51822-1e51822-18 (2012).

Miljkovic, D.; Podpečan, V.; Stare, T.; Mozetič, I.; Gruden, K.; Lavrač, N. Incremental construction of biological networks by relation extraction from literature. *Current Bioinformatics*. (In press).

Miljkovic, D.; Depolli, M.; Stare, T.; Mozetič, I.; Petek, M.; Gruden, K.; Lavrač, N. Plant defence model revisions through iterative minimization of constraint violations. *International Journal of Computational Biology and Drug Design*. (In press).

### Published scientific conference contribution (1.08)

Podpečan, V.; Miljkovic, D.; Petek, M.; Stare, T.; Gruden, K.; Mozetič, I.; Lavrač, N. Integrating semantic transcriptomic data analysis and knowledge extraction from biological literature. *Proceeding of the International Conference on Bioinformatics and Biomedicine (BIBM)*. (Shangai, China, 2013). (In press).

Miljkovic, D.; Podpečan, V.; Stare, T.; Mozetič, I.; Gruden, K.; Lavrač, N. Incremental revision of biological networks from texts. In: Ortuno, F.; Rojas, I. (eds.) *Proceedings of the International Work-Conference on Bioinformatics and Biomedical Engineering (IWBBIO)*. 1-9 (Granada, Spain, 2013).

Miljkovic, D.; Depolli, M.; Mozetič, I.; Lavrač, N.; Stare, T.; Petek, M.; Gruden, K. Constraint-driven optimization of plant defense model parameters. In: Gao, J. (ed.) *Proceedings of the Third Workshop on Integrative Data Analysis in Systems Biology (IDASB)*. 570-574 (Danvers: Institute of Electrical and Electronics Engineers, 2012).

Miljkovic, D.; Mihăilă, C.; Podpečan, V.; Grčar, M.; Gruden, K.; Stare, T.; Lavrač, N. Workflow-based information retrieval to model plant defence response to pathogen attacks. In: Hilario, M.; Lavrač, N.; Podpečan, V.; Kok, J. N. (eds.) *Proceedings of the workshop on Third generation data mining: Towards service-oriented knowledge discovery (SoKD)*. 51-60 (Barcelona, Spain, 2010).

## Independent scientific component part or a chapter in a monograph (1.16)

Miljkovic, D.; Podpečan, V.; Grčar, M.; Gruden, K.; Stare, T.; Petek, M.; Mozetič, I.; Lavrač, N. Modelling a biological system: network creation by triplet extraction from biological literature. In: Berthold, M. R. (ed.) *Bisociative Knowledge Discovery: An Introduction to Concept, Algorithms, Tools, and Applications*. 427-437 (Springer, Berlin, 2012).

## Published scientific conference contribution abstract (1.12)

Miljkovic, D.; Depolli, M.; Petek, M.; Stare, T.; Gruden, K.; Mozetič, I.; Lavrač, N. Iterative modifications of the plant defence model based on the constraint-driven parametric optimisation. In: Ačimovič, J. (ed.). *Book of abstracts of the Casym Training Event Systems Medicine of Multifactorial Disorders Workshop & Tutorial and 8th CFGBC Symposium*. 43 (Ljubljana: Faculty of Medicine, University of Ljubljana, 2013).

Podpečan, V.; Miljkovic, D.; Langohr, L.; Gruden, K.; Mozetič, I.; Zemenova, M.; Motaln, H.; Lavrač, N. A service-oriented knowledge discovery platform with applications in system biology. In: Ačimovič, J. (ed.). *Book of abstracts of the Casym Training Event Systems Medicine of Multifactorial Disorders Workshop & Tutorial and 8th CFGBC Symposium*. 47 (Ljubljana: Faculty of Medicine, University of Ljubljana, 2013).

Stare, T.; Ramšak, Ž.; Miljkovic, D.; Petek, M.; Vodnik, D.; Mozetič, I.; Lavrač, N.; Gruden, K. Dynamics of potato (*Solanum tuberosum*) defence response to Potato virus Y (PVY) infection. In: Ačimovič, J. (ed.). *Book of abstracts of the Casym Training Event Systems Medicine of Multifactorial Disorders Workshop & Tutorial and 8th CFGBC Symposium*. 50 (Ljubljana: Faculty of Medicine, University of Ljubljana, 2013).

Stare, T.; Ramšak, Ž.; Miljkovic, D.; Petek, M.; Lavrač, N.; Mozetič, I.; Vodnik, D.; Gruden, K. Dynamics of plant defense response to virus infection. In: *Book of abstracts of the Keystone symposia on Molecular and Cellular Biology: Plant immunity : pathways and translation*. 60 (Montana, USA, 2013).

Gruden, K.; Miljkovic, D.; Stare, T.; Mozetič, I.; Podpečan, V.; Petek, M.; Witek, K.; Dermastia, M.; Lavrač, N. Constructing signalling network topology for modelling the plant defence. In: *Book of abstracts of the 13th International Conference on System Biology (ICSB)*. 199 (Toronto, Canada, 2012).

Miljkovic, D.; Depolli, M.; Petek, M.; Stare, T.; Dermastia, M.; Gruden, K.; Mozetič, I.; Lavrač, N. Constraint-driven optimization approach to build a Petri net defense model in plants. In: *Book of abstracts of the 1st Conference on Constraint-based Reconstruction and Analysis (COBRA)*. (Reykjavik: University of Iceland, Center for Systems Biology, 2011).

Miljkovic, D.; Stare, T.; Gruden, K.; Mozetič, I.; Grčar, M.; Lavrač, N. Text processing assisted development of a petri net model for plant defence response. In: *Workshop on Advances in Bio Text Mining (BIO TM)* (Ghent, Belgium, 2010).

## Index of Figures

Figure 1:	Phases in model construction. . . . .	13
Figure 2:	Dynamics of a simplified thermostat behaviour. . . . .	17
Figure 3:	The graphical representation of the basic PN. . . . .	18
Figure 4:	The simplified thermostat model represented with four different PN formalism types and the corresponding simulation curves of the temperature and voltage variables. . . . .	19
Figure 5:	A graphical representation of the thermostat model with the negative feedback loop realised with the HPN and FPN modelling approach. The simulation curves are also shown. . . . .	21
Figure 6:	An example of a gene regulatory network modelled with the HPN formalism. The model represents early stage gene expression of $\lambda$ phage (Matsuno et al., 2000). The rectangular nodes represent Petri net transitions where the black coloured are discrete and the non-coloured continuous transitions. The rounded nodes are Petri net places, where the single circle represents discrete place and the double circle continuous place. . . . .	23
Figure 7:	Feedback mechanism of CI and Cro proteins, which represents part of the early stage gene expression of $\lambda$ phage (Matsuno et al., 2000). An explanation of types of graph nodes is the same as in Figure 6. . . . .	23
Figure 8:	A schema of the developed methodology for the PDS model construction.	28
Figure 9:	An overview of the modelling formalisms and the defined requirements for the PDS model. . . . .	30
Figure 10:	The presentation nodes and edges of the mEPN (Raza et al., 2008). The notation, which we have adapted for the PDS model, is grouped into four categories: components, transition nodes, Boolean operators and annotated edges. . . . .	36
Figure 11:	Taxonomy of PDS components and reactions. A) In the taxonomy of PDS components, there are four representation levels. The highest level (level 0) is the most abstract level, while the lowest one (level 3) represents single molecules. B) In the taxonomy of PDS reactions, individual reactions are represented at the lowest level (level 2) and are grouped according to their functionality into three groups at level 1: <i>Activation (A)</i> , <i>Binding (B)</i> and <i>Inhibition (I)</i> . . . . .	37

Figure 12:	The PDS reaction types at level 1. There are three groups of reactions. A) <i>Activation</i> ( <i>A</i> ) denotes all the reactions directly involving two components X and Y in the production of Z, where the concentration of Z depends on the concentration of both substrates. B) <i>Binding</i> ( <i>B</i> ) results in the formation of a protein-protein complex or in the binding of a protein to a DNA promoter region to regulate its gene expression. C) <i>Inhibition</i> ( <i>I</i> ) is a process in which one component blocks the performance of another component. . . . .	38
Figure 13:	Principle of decomposing families of components by decoupling of reactions. The example shown in this figure illustrates the expansion of the PDS model structure from the family to the individual component level. Here, the decomposing of LOX node is done from the protein family level (level 2) to the single protein level (level 3). The final result of the expansion is a graph with 8 nodes and 7 edges. . . . .	39
Figure 14:	Manually constructed PDS model structure visualised as an edge-labelled graph. This graph, consisting of 175 nodes and 387 edges, is interactive and is visualised with the Biomine graph visualisation engine, enabling its closer inspection by zooming into its subparts and rearranging the node and the arc positions in the 2D space. The graph is organised into SA, JA and ET pathways with their crosstalk connections. The node borders of the main pathway components SA, JA and ET are coloured with red. . . . .	39
Figure 15:	The principles of conversion from the edge-labelled graph format. A) <i>Activation</i> reaction (labelled A) with arcs between the reactant and the product node is transformed to a reaction between two components. B) <i>Activation</i> (labelled A) on a transcription level is a special type of <i>activation</i> , when Y induces the activation of gene X to produce protein X. In this case we omit the gene transcription level when transforming the level 2 structure from the edge-labelled graph. C) <i>Binding</i> (labelled B) relation is transformed from a B relation between the reactants and an additional relation produces (labelled as P) between the reactant and the product into a relation between two reactant nodes X and Y. D) <i>Inhibition</i> (labelled I) is the blocking of the activation or binding reaction between components by a third component X, resulting in reduced production of product Z. . . . .	40
Figure 16:	Graphical representation of biological components and reactions modelled with the CI software. . . . .	41
Figure 17:	A simplified part of the model of the biosynthesis and signalling pathway of SA, manually constructed in the CI software. The whole SA-submodel is shown in Figure 24. . . . .	42
Figure 18:	The biosynthesis and signalling pathway of the JA sub-model manually constructed in the CI software. . . . .	44
Figure 19:	The biosynthesis and signalling pathway of the ET sub-model manually constructed in the CI software. . . . .	46
Figure 20:	A simple reaction presented in the HFPN formalism. . . . .	48
Figure 21:	A scheme of an evolutionary algorithm (EA). . . . .	50
Figure 22:	Dynamic behaviour of the SA, EDS1, PR1 and NPR1 variables based on the optimal parameter set estimated with respect to the objective function calculated from eight constraints during step 1. . . . .	53

Figure 23:	Dynamic behaviour of the SA, EDS1, PR1 and NPR1 variables based on the optimal set estimated with respect to the objective function calculated from 33 constraints during step 2. . . . .	54
Figure 25:	Dynamic behaviour of the SA, EDS1, PR1 and NPR1 variables based on the optimal parameter set estimated with respect to the objective function calculated from 29 constraints during step 3. . . . .	56
Figure 24:	The biosynthesis and signalling pathway of the final SA sub-model v3.0 manually constructed in the CI. The molecules in the green squares were used for the model validation in Section 5.3, while the coefficients in the black squares were used for the local sensitivity analysis of the SA sub-model v3.0 in Subsection 5.3.2. . . . .	57
Figure 26:	The classification of validation approaches for the observable and non-observable system according to (Sargent, 1998). . . . .	60
Figure 27:	The iterative process of three-fold cross-validation. . . . .	62
Figure 28:	The influence of perturbations of rate coefficients $k_{f33}$ , $k_{f37}$ and $k_{f57}$ on the dynamic profile of the EDS1 protein. . . . .	65
Figure 29:	The influence of perturbations of rate coefficients $k_{f33}$ , $k_{f37}$ and $k_{f63}$ on the dynamic profile of the PAD3/4 protein. . . . .	66
Figure 30:	The influence of perturbations of rate coefficients $k_{f16}$ , $k_{f48}$ and $k_{f51}$ on the dynamic profile of the NPR1 protein. . . . .	67
Figure 31:	The influence of perturbations of rate coefficients $k_{f16}$ , $k_{f48}$ and $k_{f51}$ on the dynamic profile of the PR1 protein. . . . .	68
Figure 32:	Processed and normalised microarray expression dataset for TuMV inoculation type for 7 components of the SA v3.0 sub-model. . . . .	70
Figure 33:	Processed and normalised microarray expression dataset for ORMV inoculation type for 7 components of the SA v3.0 sub-model. . . . .	70
Figure 34:	Overview of the Bio3graph methodology, its implementation and a sample output. A) Schematic representation of the Bio3graph methodology. Text processing is performed in a workflow according to the boxes in the schematic diagram resulting in a graph of ( <i>component1</i> , <i>reaction</i> , <i>component2</i> ) triplets. B) Bio3graph as a workflow implemented in Orange4WS. . . . .	75
Figure 35:	The triplet graph extracted and composed by Bio3graph. The output graph (consisting of 129 components and 1,132 reactions) is visualised with the Biomine graph visualisation engine. . . . .	76
Figure 36:	Illustration of the triplet extraction process. We show a part of the flow from input of POS tagging box from Figure 34 until output of triplet extraction box of the same figure. The input to the Genia POS tagger is a previously pre-processed sentence. After the shallow parsing with Genia POS tagger, the algorithm performs the step 2. The final output from the triplet extraction part of Bio3graph approach is a triplet in the form ( <i>subject</i> , <i>predicate</i> , <i>object</i> ) which will be then transformed and visualised as an edge-labelled graph with the Biomine visualiser. . . . .	78
Figure 37:	A model of experts' knowledge constructed from a set of triplets extracted from 122 documents, read by two different readers and displayed using the Biomine graph visualisation engine. . . . .	83

- Figure 38: New direct PDS relations extracted from the biological literature. The new direct links result from the Bio3graph processing of 9,586 articles. Bio3graph extracted 14 new direct relations between the components which were not identified in the manually built PDS model structure. Note that two of these triplets are trivial (*SAG\_metabolite*, *activates*, *SA\_metabolite*) and (*NIMIN1\_protein*, *inhibits*, *NPR1\_protein*). . . . . 84
- Figure 39: The final PDS model structure constructed by merging the manual and the Bio3graph graph. A) Edge-labelled graph representing the merged model. B) The Venn diagram. The relations in the manual model are all direct and are coloured in red. The intersection between the model relations and the correct triplets extracted from the literature is presented with black colour. From the correct new triplets, the indirect relations are represented with green and the direct ones with blue colour. C) Zoom-in into a part of the merged PDS structure. The links from the manual model are shown in red, while the green coloured relations represent the extracted new indirect links, blue arcs show new direct links and the black arcs show the intersection between the manual model and the correct triplets extracted with Bio3graph. . . . . 87
- Figure 40: Scheme of the methodology for incremental construction of biological networks using information extraction from literature. . . . . 92
- Figure 41: An example of a redundant transitive relation in a simple graph. The relation *v Activates x*, shown in grey, is transitive. It does not contain new biological knowledge and is thus redundant. . . . . 94
- Figure 42: A screenshot of the workflow implementing the proposed incremental revision of biological network structures. The first part of the workflow implements the Bio3graph approach for automated triplet graph extraction from biological literature while the second part implements incremental extension of biological model structures. . . . . 96
- Figure 43: Transformation of the SA model structure available in literature into a directed graph with labelled edges. The model structure originates from the study of (Shah, 2003). . . . . 98
- Figure 44: Transformation of the crosstalk between SA, JA and ET pathways available in literature into a directed graph with labelled edges. The model structure originates from the study of (Turner et al., 2002). . . . . 98
- Figure 45: Transformation of the ET model structure from literature into a directed graph with labelled edges. The model structure originates from the work of (Gonzalez-Garcia and Diaz, 2011). . . . . 98



- Figure 46: The enhancement of the Initial graph (A) with the correct triplets after first incremental step for the use case 1. The left side represents the input graphs for the incremental extension of Bio3graph while the right side represent the output graphs. A) The Initial graph created by merging the manually constructed three graphs from the literature shown in Figures. 43, 44 and 45. B) The Triplet graph constructed from the correct triplets extracted with Bio3graph. C) The Incremented graph obtained by merging the Initial and the Triplet graph. The relations present only in the Initial graph are coloured in black while the relations present also in the the Triplet are coloured in green, red or pink. Relations present in both Initial graph and triplet network are coloured in green, newly discovered relations are coloured in red while the newly discovered, redundant transitive relations are coloured in pink. D) The Filtered Incremented graph obtained from the Incremented graph by removing the redundant transitive relations. . . . . 100
- Figure 47: The final Incremented graph after two incremental steps for the use case 1. The new relations from the second set of correct triplets are shown in red. Pink arcs represent redundant transitive relations from the second set of correct triplets which are newly discovered. . . . . 101



## Index of Tables

Table 1:	Mapping between biology and computer science terminology used in this thesis. . . . .	4
Table 2:	The mapping between the PN and ODE terminology. . . . .	16
Table 3:	A list of parameters of the SA sub-model v3.0 optimised with our approach. . . . .	58
Table 4:	A comparison of the three iterative optimization steps. . . . .	59
Table 5:	The summary of local sensitivity values calculated for the EDS1 component. . . . .	64
Table 6:	The summary of local sensitivity values calculated for the PAD3/4 component. . . . .	65
Table 7:	The summary of local sensitivity values calculated for the NPR1 component. . . . .	66
Table 8:	The summary of local sensitivity values calculated for the PR1 component. . . . .	67
Table 9:	Sample types used in the study of (Yang et al., 2007) to investigate <i>Arabidopsis thaliana</i> responses to TuMV and ORMV. The "..." in the table denotes no data collected on the particular day, while "x" denotes successfully collected sample on the particular day. . . . .	69
Table 10:	The results of the validation method based on the generation of constraints from the experimental datasets. The best 10 optimisation results guided with the <i>evaluation constraint set</i> are sorted in the second column starting from the best solution. The results of the second column are evaluated (compared) with the experimental data for TuMV (third column) and ORMV inoculation (forth column). . . . .	71

Table 11:	Recall and precision analysis for 50 full-length papers. $Recall = TP / (TP + TN)$ and $Precision = TP / (TP + FP)$ , where $TP$ are the true positives, $TN$ the true negatives, and $FP$ the false positives. $Recall$ is the percentage of the retrieved true positive relations from the whole set of true relations. $Precision$ is the percentage of retrieved true positive relations from the whole set of retrieved relations. . . . .	80
Table 12:	Summary of PDS related triplets extracted by the Bio3graph triplet extraction algorithm from 9,586 PMC articles. In total, 1,132 triplets were extracted, out of which 377 are correct. Out of these, 14 are newly discovered direct relations and 123 are indirect, while 44 direct and 196 indirect connections were already included in the manual PDS model structure of Figure 14. . . . .	85
Table 14:	The summary of relations of the Incremented graph shown in Figure 46.C after the first incremental step for the use case 1. The initial links originate only from the Initial graph, while the intersection, new redundant and new links originate from the Triplet graph. The intersection links are the common relations of the Initial and the Triplet graph. The new redundant links are the transitive relations while the most interesting are the new links, which represent exclusively new relations discovered by the Bio3graph tool. . . . .	99
Table 13:	The summary of triplet extraction from documents before and after the time point for the use case 1 (simple plant defence model structure). . .	99
Table 15:	The summary of relations of the Incremented graph shown in Figure 47 after the second incremental step for the use case 1. The initial links originate only from the Initial graph, while the intersection, new redundant and new links originate from the Triplet graph. The intersection links are the common relations of the Initial and the Triplet graph. The new redundant links are the transitive relations while the most interesting are the new links, which represent exclusively new relations discovered by the Bio3graph tool. . . . .	101
Table 16:	The summary of triplets extracted from biological texts in the use case 2 (complex plant defence model structure). The initial links originate only from the Initial graph, while the intersection, new redundant and new links originate from the Triplet graph. The intersection links are the common relations of the Initial and the Triplet graph. The new redundant links are the transitive relations while the most interesting are the new links, which represent exclusively new relations discovered by the Bio3graph tool. . . . .	102
Table A.1:	The list of the interactions in the manually constructed structure of the SA sub-model. . . . .	135
Table A.2:	The list of the interactions in the manually constructed structure of the JA sub-model. . . . .	140
Table A.3:	The list of the interactions in the manually constructed structure of the ET sub-model. . . . .	150

Table A.4: The list of manually acquired crosstalk interactions between SA, JA and ET sub-model. . . . .	154
Table A.5: The list of manually acquired crosstalk interactions between SA, JA and ET sub-model. . . . .	156



# Appendix

## A Manually constructed PDS model structure

### A.1 The summary of all relations in the manual model

This summary includes the reactions acquired for the manual PDS model structure. Table A.1 presents list of interactions of the SA sub-model, Table A.2 lists reactions of the JA sub-model, Table A.3 includes reactions of the ET sub-model and Table A.4 presents the list of crosstalk connections.

The structure of all four tables is the same. In the first column the interaction between two biological components is given in the form of biological reaction representation, with the following structure: (reactant1 + reactant2 reaction product). For example: protein\_MYC2 + gene\_THI2.1JR1VSP1CLH1 activates protein\_THI2.1JR1VSP1CLH1. In this column the biological components are represented on the level of the family nodes. In the second column the relations after decomposition of the family nodes are shown in the edge-labelled graph format. In the last third column the source of information, related to the particular interaction, is specified.

Table A.1: The list of the interactions in the manually constructed structure of the SA sub-model.

SA biological reactions	Directed edge-labelled graph	Data source
protein_MPK3 + protein_MPK6 + gene_PAD34 activates protein_PAD34	protein_MPK3 protein_PAD4 A, protein_MPK3 protein_PAD3 A, protein_MPK6 protein_PAD4 A, protein_MPK6 protein_PAD3 A	PMID: 18378893
protein_NPR1 inhibits protein_PAD34	protein_NPR1 protein_PAD4 I, protein_NPR1 protein_PAD3 I	PMID: 12873532
complex_NPR1_oligomer + metabolite_SA activates protein_NPR1	metabolite_SA protein_NPR1 A, complex_NPR1_oligomer protein_NPR1 A	PMID: 18635760, PMID: 12837250
protein_Inactive_HRT activates protein_HRT	protein_Inactive_HRT protein_HRT A	PMID: 15546349

SA biological reactions	Directed edge-labelled graph	Data source
metabolite_Orto-coumaric_acid + protein_X1 activates metabolite_SA	metabolite_Orto-coumaric_acid metabolite_SA A, protein_X1 metabolite_SA A	KEGG
protein_AAO4 + metabolite_Trans-cinnamic_acid activates metabolite_BA + metabolite_Orto-coumaric_acid	protein_AAO4 metabolite_BA A, protein_AAO4 metabolite_Orto-coumaric_acid A, metabolite_Trans-cinnamic_acid metabolite_BA A, metabolite_Trans-cinnamic_acid metabolite_Orto-coumaric_acid A	KEGG
protein_UGP_glikosyl-transferaze + metabolite_SA activates metabolite_SGE	protein_UGP_glikosyl-transferaze metabolite_SGE A, metabolite_SA metabolite_SGE A	PMID: 18267075
protein_BA2H + metabolite_BA activates metabolite_SA	metabolite_BA metabolite_SA A, protein_BA2H metabolite_SA A	PMID: 12231938
protein_UGP_glikosyl-transferaze + metabolite_SA activates metabolite_SAG	metabolite_SA metabolite_SAG A, protein_UGP_glikosyl-transferaze metabolite_SAG A	PMID: 18267075
protein_EDS5 + gene_ICS activates protein_ICS	protein_EDS5 protein_ICS1 A, protein_EDS5 protein_ICS2 A	PMID: 12873532
complex_NPR1_TGA245 + gene_PR125 activates protein_PR125	complex_NPR1_TGA245 protein_PR1 A, complex_NPR1_TGA245 protein_PR2 A, complex_NPR1_TGA245 protein_PR5 A	PMID: 12873532



SA biological reactions	Directed edge-labelled graph	Data source
protein_NPR1 + gene_WRKY70 activates protein_WRKY70	protein_NPR1 protein_WRKY70 A	PMID: 14742872
protein_NPR1 + protein_NIMIN + protein_TGA_TF245 binds complex_NPR1_TGA245	protein_NPR1 complex_NPR1_TGA245 P, protein_NIMIN1 complex_NPR1_TGA245 P, protein_NIMIN2 complex_NPR1_TGA245 P, protein_NIMIN3 complex_NPR1_TGA245 P, protein_TGA_TF2 complex_NPR1_TGA245 P, protein_TGA_TF4 complex_NPR1_TGA245 P, protein_TGA_TF5 complex_NPR1_TGA245 P, protein_NPR1 protein_NIMIN1 B, protein_NPR1 protein_TGA_TF2 B, protein_NIMIN1 protein_TGA_TF2 B, protein_NPR1 protein_TGA_TF4 B, protein_NIMIN1 protein_TGA_TF4 B, protein_NPR1 protein_TGA_TF5 B, protein_NIMIN1 protein_TGA_TF5 B, protein_NPR1 protein_NIMIN2 B, protein_NIMIN2 protein_TGA_TF2 B, protein_NIMIN2 protein_TGA_TF4 B, protein_NIMIN2 protein_TGA_TF5 B, protein_NPR1 protein_NIMIN3 B, protein_NIMIN3 protein_TGA_TF2 B, protein_NIMIN3 protein_TGA_TF4 B, protein_NIMIN3 protein_TGA_TF5 B	PMID: 15749762, PMID: 12873532

SA biological reactions	Directed edge-labelled graph	Data source
protein_X3 + metabolite_SA_chl activates metabolite_SA	protein_X3 metabolite_SA A, metabolite_SA_chl metabolite_SA A	PMID: 18267075
metabolite_Phenylalanine + protein_PAL activates metabolite_Trans- cinnamic_acid	metabolite_Phenylalanine metabolite_Trans- cinnamic_acid A, protein_PAL1 metabolite_Trans- cinnamic_acid A, protein_PAL2 metabolite_Trans- cinnamic_acid A, protein_PAL3 metabolite_Trans- cinnamic_acid A, protein_PAL4 metabolite_Trans- cinnamic_acid A	KEGG
protein_ICS + metabolite_Chorismate activates metabolite_Isochorismate	protein_ICS1 metabolite_Isochorismate A,protein_ICS2 metabolite_Isochorismate A, metabolite_Chorismate metabolite_Isochorismate A	PMID: 12873532
protein_Prephenate- _aminotransferase + metabolite_Prephenate activates metabo- lite_Phenylpyruvate	metabolite_Prephenate metabo- lite_Phenylpyruvate A, protein_Prephenate- _aminotransferase metabo- lite_Phenylpyruvate A	KEGG
metabolite_Isochorismate + protein_IPL activates metabolite_SA_chl	metabolite_Isochorismate metabolite_SA_chl A, protein_IPL metabolite_SA_chl A	PMID: 12873532
protein_ROS_pro- ductive_enzymes activa-tes metabolite_ROS	protein_NADPH_oxi-dase metabolite_ROS A, protein_Catalase metabolite_ROS A, protein_Glutathione_pero- xidase metabolite_ROS A, pro- tein_Superoxide_dismutase metabolite_ROS A	PMID: 8612756

SA biological reactions	Directed edge-labelled graph	Data source
protein_Inactive_BA2H + metabolite_ROS activates protein_BA2H	metabolite_ROS protein_BA2H A, protein_Inactive_BA2H protein_BA2H A	PMID:12060237
protein_CM + metabolite_Chorismate activates metabolite_Prephenate	protein_CM1 metabolite_Prephenate A, protein_CM2 metabolite_Prephenate A, protein_CM3 metabolite_Prephenate A, metabolite_Chorismate metabolite_Prephenate A	KEGG
protein_Arogenate_dehy- dratase + metabo- lite_Phenylpyruvate activates metabolite_Phenylalanine	protein_Arogenate_dehy- dratase metabolite_Phenylalanine A, metabo- lite_Phenylpyruvate metabolite_Phenylalanine A	KEGG
protein_PAD34 + protein_EDS1 + gene_EDS5 activates protein_EDS5	protein_PAD4 protein_EDS5 A, protein_PAD3 protein_EDS5 A, protein_EDS1 protein_EDS5 A	PMID: 12873532, PMID: 15546349
protein_Inactive_MPK3 + protein_HRT + metabolite_ROS activates protein_MPK3	protein_HRT protein_MPK3 A,metabolite_ROS protein_MPK3 A, protein_Inactive_MPK3 protein_MPK3 A	PMID: 11875555, PMID: 21046144
protein_Inactive_MPK6 + protein_HRT + metabolite_ROS activates protein_MPK6	protein_Inactive_MPK6 protein_MPK6 A, protein_HRT protein_MPK6 A, metabolite_ROS protein_MPK6 A	PMID: 15020743, PMID: 21046144
protein_MPK3 + protein_MPK6 + gene_EDS1 activates protein_EDS1	protein_MPK3 protein_EDS1 A, protein_MPK6 protein_EDS1 A	PMID: 18705666
protein_NPR1 inhibits protein_EDS1	protein_NPR1 protein_EDS1 I	PMID: 12873532

Table A.2: The list of the interactions in the manually constructed structure of the JA sub-model.

JA biological reactions	Directed edge-labelled graph	Data source
protein_MPK4 activates protein_LOX3	protein_MPK4 protein_LOX5 A, protein_MPK4 protein_LOX4 A, protein_MPK4 protein_LOX6 A, protein_MPK4 protein_LOX1 A, protein_MPK4 protein_LOX2 A, protein_MPK4 protein_LOX3 A	PMID:20037473
protein_MYC2 + gene_THI2.1JR1VSP1-CLH1 activates protein_THI2.1JR1VSP1-CLH1	protein_MYC2 protein_JR1 A, protein_MYC2 protein_VSP1 A, protein_MYC2 protein_CLH1 A, protein_MYC2 protein_THI2.1 A	PMID: 17616737, Devoto A, Turner JG (2005) Jasmonate-regulated Arabidopsis stress signalling network. Physiologia Plantarum 123(2): 161-172
metabolite_OPDA_chl activates metabolite_OPDA	metabolite_OPDA_chl metabolite_OPDA A	Zhang H, Memelink J (2009) Regulation of Secondary Metabolism by Jasmonate Hormones. Plant-derived Natural Products, pp.181-194
protein_LOX + metabolite_Linolenic_acid activates metabolite_13-HPT	protein_LOX5 metabolite_13-HPT A, protein_LOX4 metabolite_13-HPT A, protein_LOX6 metabolite_13-HPT A, protein_LOX1 metabolite_13-HPT A, protein_LOX2 metabolite_13-HPT A, protein_LOX3 metabolite_13-HPT A, metabolite_Linolenic_acid metabolite_13-HPT A	KEGG, PMID:18583180, AraCyc database

JA biological reactions	Directed edge-labelled graph	Data source
protein_JAZ inhibits protein_THI2.1JR1- VSP1CLH1	protein_JAZ5 protein_JR1 I, protein_JAZ5 protein_VSP1 I, protein_JAZ5 protein_CLH1 I, protein_JAZ5 protein_THI2.1 I, protein_JAZ1 protein_JR1 I, protein_JAZ1 protein_VSP1 I, protein_JAZ1 protein_CLH1 I, protein_JAZ1 protein_THI2.1 I, protein_JAZ8 protein_JR1 I, protein_JAZ8 protein_VSP1 I, protein_JAZ8 protein_CLH1 I, protein_JAZ8 protein_THI2.1 I, protein_JAZ4 protein_JR1 I, protein_JAZ4 protein_VSP1 I, protein_JAZ4 protein_CLH1 I, protein_JAZ4 protein_THI2.1 I, protein_JAZ9 protein_JR1 I, protein_JAZ9 protein_VSP1 I, protein_JAZ9 protein_CLH1 I, protein_JAZ9 protein_THI2.1 I, protein_JAZ6 protein_JR1 I, protein_JAZ6 protein_VSP1 I, protein_JAZ6 protein_CLH1 I,	Devoto A, Turner JG (2005) Jasmonate-regulated Arabidopsis stress signalling network. Physiologia Plantarum 123(2): 161-172

JA biological reactions	Directed edge-labelled graph	Data source
protein_JAZ inhibits protein_THI2.1JR1- VSP1CLH1	protein_JAZ6 protein_THI2.1 I, protein_JAZ2 protein_JR1 I, protein_JAZ2 protein_VSP1 I, protein_JAZ2 protein_CLH1 I, protein_JAZ2 protein_THI2.1 I, protein_JAZ7 protein_JR1 I, protein_JAZ7 protein_VSP1 I, protein_JAZ7 protein_CLH1 I, protein_JAZ7 protein_THI2.1 I, protein_JAZ11 protein_JR1 I, protein_JAZ11 protein_VSP1 I, protein_JAZ11 protein_CLH1 I, protein_JAZ11 protein_THI2.1 I, protein_JAZ10 protein_JR1 I, protein_JAZ10 protein_VSP1 I, protein_JAZ10 protein_CLH1 I, protein_JAZ10 protein_THI2.1 I, protein_JAZ12 protein_JR1 I, protein_JAZ12 protein_VSP1 I, protein_JAZ12 protein_CLH1 I, protein_JAZ12 protein_THI2.1 I, protein_JAZ3 protein_JR1 I, protein_JAZ3 protein_VSP1 I	Devoto A, Turner JG (2005) Jasmonate-regulated Arabidopsis stress signalling network. Physiologia Plantarum 123(2): 161-172

JA biological reactions	Directed edge-labelled graph	Data source
protein_JAZ inhibits protein_THI2.1JR1- VSP1CLH1	protein_JAZ3 protein_CLH1 I, protein_JAZ3 protein_THI2.1 I, protein_JAZ3 protein_CLH1 I, protein_JAZ3 protein_THI2.1 I	Devoto A, Turner JG (2005) Jasmonate-regulated Arabidopsis stress signalling network. Physiologia Plantarum 123(2): 161-172
protein_X4 + gene_MYC2 activates protein_MYC2	protein_X4 protein_MYC2 A	Added to the model to make the transcription of Myc gene constant (the exact activator is to our knowledge still unknown)
protein_OPR3 + metabolite_OPDA activates metabolite_OPC8	protein_OPR3 metabolite_OPC8 A, metabolite_OPDA metabolite_OPC8 A	KEGG, AraCyc database, Zhang H, Memelink J (2009) Regulation of Secondary Metabolism by Jasmonate Hormones. Plant-derived Natural Products, pp.181-194
protein_AOC + metabolite_12/13_EDT activates metabolite_OPDA_chl	protein_AOC1 metabolite_OPDA_chl A, protein_AOC2 metabolite_OPDA_chl A, protein_AOC3 metabolite_OPDA_chl A, protein_AOC4 metabolite_OPDA_chl A, metabolite_12/13_EDT metabolite_OPDA_chl A	KEGG, AraCyc database, PMID19025383
protein_JAR1 + metabolite_JA activates metabolite_JA-Ile	protein_JAR1 metabolite_JA-Ile A, metabolite_JA metabolite_JA-Ile A	KEGG
protein_JMT + metabolite_JA activates metabolite_Me-JA	protein_JMT metabolite_Me-JA A, metabolite_JA metabolite_Me-JA A	KEGG, PMID: 19025383, PMID: 18583180
protein_AOS + metabolite_13-HPT activates metabolite_12/13_EDT	protein_AOS metabolite_12/13_EDT A, metabolite_13-HPT metabolite_12/13_EDT A	KEGG

JA biological reactions	Directed edge-labelled graph	Data source
protein_X2 + gene_JAZ activates protein_JAZ	protein_X2 protein_JAZ5 A, protein_X2 protein_JAZ1 A, protein_X2 protein_JAZ8 A, protein_X2 protein_JAZ4 A, protein_X2 protein_JAZ9 A, protein_X2 protein_JAZ6 A, protein_X2 protein_JAZ2 A, protein_X2 protein_JAZ7 A, protein_X2 protein_JAZ11 A, protein_X2 protein_JAZ10 A, protein_X2 protein_JAZ12 A, protein_X2 protein_JAZ3 A	PMID:21963667
complex_SCF + metabolite_JA-Ile + protein_COI1 binds complex_JA-Ile_COI1_SCF	complex_SCF complex_JA-Ile_COI1_SCF P, protein_COI1 complex_JA-Ile_COI1_SCF P, metabolite_JA-Ile complex_JA-Ile_COI1_SCF P, complex_SCF protein_COI1 B, complex_SCF metabolite_JA-Ile B, protein_COI1 metabolite_JA-Ile B	PMID: 18261950, PMID: 18583180, Zhang H, Memelink J (2009) Regulation of Secondary Metabolism by Jasmonate Hormones. Plant-derived Natural Products, pp.181-194
metabolite_OPC8 + protein_X5 activates metabolite_OPC-8:0-CoA	metabolite_OPC8 metabolite_OPC-8:0-CoA A, protein_X5 metabolite_OPC-8:0-CoA A	KEGG



JA biological reactions	Directed edge-labelled graph	Data source
protein_JAZ + complex_JA-Ile_COI1_SCF binds complex_JA- Ile_COI1_SCF_JAZ	protein_JAZ5 complex_JA- Ile_COI1_SCF_JAZ P, protein_JAZ1 complex_JA- Ile_COI1_SCF_JAZ P, protein_JAZ8 complex_JA- Ile_COI1_SCF_JAZ P, protein_JAZ4 complex_JA- Ile_COI1_SCF_JAZ P, protein_JAZ9 complex_JA- Ile_COI1_SCF_JAZ P, protein_JAZ6 complex_JA- Ile_COI1_SCF_JAZ P, protein_JAZ2 complex_JA- Ile_COI1_SCF_JAZ P, protein_JAZ7 complex_JA- Ile_COI1_SCF_JAZ P, protein_JAZ11 complex_JA- Ile_COI1_SCF_JAZ P, protein_JAZ10 complex_JA- Ile_COI1_SCF_JAZ P, protein_JAZ12 complex_JA- Ile_COI1_SCF_JAZ P, protein_JAZ3 complex_JA- Ile_COI1_SCF_JAZ P, complex_JA-Ile_COI1_SCF complex_JA- Ile_COI1_SCF_JAZ P, protein_JAZ5 complex_JA-Ile_COI1_SCF B, protein_JAZ1 complex_JA-Ile_COI1_SCF B,	PMID:18583180, PMID:18261950, PMID: 18583180, Zhang H, Memelink J (2009) Regulation of Secondary Metabolism by Jasmonate Hormones. Plant-derived Natural Products, pp.181-194

JA biological reactions	Directed edge-labelled graph	Data source
protein_JAZ + complex_JA-Ile_COI1_SCF binds complex_JA-Ile_COI1_SCF_JAZ	protein_JAZ8	PMID:18583180, PMID:18261950, PMID: 18583180, Zhang H, Memelink J (2009) Regulation of Secondary Metabolism by Jasmonate Hormones. Plant-derived Natural Products, pp.181-194
	complex_JA-Ile_COI1_SCF	
	B, protein_JAZ4	
	complex_JA-Ile_COI1_SCF	
	B, protein_JAZ9	
	complex_JA-Ile_COI1_SCF	
	B, protein_JAZ6	
	complex_JA-Ile_COI1_SCF	
	B, protein_JAZ2	
	complex_JA-Ile_COI1_SCF	
	B, protein_JAZ7	
	complex_JA-Ile_COI1_SCF	
	B, protein_JAZ11	
	complex_JA-Ile_COI1_SCF	
	B, protein_JAZ10	
metabolite_OPC4 + protein_ACXAIM1KAT activates metabolite_JA	complex_JA-Ile_COI1_SCF	KEGG
	B, protein_JAZ12	
	complex_JA-Ile_COI1_SCF	
	B, protein_JAZ3	
	complex_JA-Ile_COI1_SCF	
	B	
	metabolite_OPC4	
	metabolite_JA A,	
	protein_ACX1	
	metabolite_JA A,	
	protein_ACX3	
	metabolite_JA A,	
	protein_ACX6	
	metabolite_JA A,	
	protein_ACX5	
	metabolite_JA A,	
	protein_ACX4	
	metabolite_JA A,	
	protein_ACX2	
	metabolite_JA A,	
	protein_AIM1	
	metabolite_JA A	

JA biological reactions	Directed edge-labelled graph	Data source
protein_JAZ + complex_JA-Ile_COI1_SCF binds complex_JA- Ile_COI1_SCF_JAZ	protein_KAT2 metabolite_JA A, protein_KAT5 metabolite_JA A, protein_KAT1 metabolite_JA A	KEGG
protein_MYC2 + gene_JAZ activates protein_JAZ	protein_MYC2 protein_JAZ5 A, protein_MYC2 protein_JAZ1 A, protein_MYC2 protein_JAZ8 A, protein_MYC2 protein_JAZ4 A, protein_MYC2 protein_JAZ9 A, protein_MYC2 protein_JAZ6 A, protein_MYC2 protein_JAZ2 A, protein_MYC2 protein_JAZ7 A, protein_MYC2 protein_JAZ11 A, protein_MYC2 protein_JAZ10 A, protein_MYC2 protein_JAZ12 A, protein_MYC2 protein_JAZ3 A	Zhang H, Memelink J (2009) Regulation of Secondary Metabolism by Jasmonate Hormones. Plant-derived Natural Products, pp.181-194, PMID: 21194534
protein_OPCL1ACX- AIM1KAT + metabolite_OPC-8:0-CoA activates metabolite_OPC6	protein_OPCL1 metabolite_OPC6 A, protein_ACX1 metabolite_OPC6 A, protein_ACX3 metabolite_OPC6 A, protein_ACX6 metabolite_OPC6 A, protein_ACX5 metabolite_OPC6 A, protein_ACX4 metabolite_OPC6 A, protein_ACX2 metabolite_OPC6 A, protein_AIM1 metabolite_OPC6 A	KEGG, AraCyc database, Zhang H, Memelink J (2009) Regulation of Secondary Metabolism by Jasmonate Hormones. Plant-derived Natural Products, pp.181-194

JA biological reactions	Directed edge-labelled graph	Data source
protein_OPCL1ACX- AIM1KAT + metabolite_OPC-8:0-CoA activates metabolite_OPC6	protein_KAT2 metabolite_OPC6 A, protein_KAT5 metabolite_OPC6 A, protein_KAT1 metabolite_OPC6 A, metabolite_OPC-8:0-CoA metabolite_OPC6 A	KEGG, AraCyc database, Zhang H, Memelink J (2009) Regulation of Secondary Metabolism by Jasmonate Hormones. Plant-derived Natural Products, pp.181-194
protein_ACXAIM1KAT + metabolite_OPC6 activates metabolite_OPC4	protein_ACX1 metabolite_OPC4 A, protein_ACX3 metabolite_OPC4 A, protein_ACX6 metabolite_OPC4 A, protein_ACX5 metabolite_OPC4 A, protein_ACX4 metabolite_OPC4 A, protein_ACX2 metabolite_OPC4 A, protein_AIM1 metabolite_OPC4 A, protein_KAT2 metabolite_OPC4 A, protein_KAT5 metabolite_OPC4 A, protein_KAT1 metabolite_OPC4 A, metabolite_OPC6 metabolite_OPC4 A	KEGG
protein_MYC2 + protein_JAZ binds complex_JAZ_MYC2_TF	protein_MYC2 complex_JAZ_MYC2_TF P, protein_JAZ5 complex_JAZ_MYC2_TF P, protein_JAZ1 complex_JAZ_MYC2_TF P, protein_JAZ8 complex_JAZ_MYC2_TF P, protein_JAZ4 complex_JAZ_MYC2_TF P, protein_JAZ9 complex_JAZ_MYC2_TF P, protein_JAZ6 complex_JAZ_MYC2_TF P	PMID: 20159850,PMID: 21194534, PMID: 21335373, PMID: 19025383, Zhang H, Memelink J (2009) Regulation of Secondary Metabolism by Jasmonate Hormones. Plant-derived Natural Products, pp.181-194, Note: Negative feed back loop to JAZ is represented as binding for simulation reasons.

JA biological reactions	Directed edge-labelled graph	Data source
protein_MYC2 + protein_JAZ binds complex_JAZ_MYC2_TF	protein_JAZ2	
	complex_JAZ_MYC2_TF	
	P, protein_JAZ7	
	complex_JAZ_MYC2_TF	
	P, protein_JAZ11	
	complex_JAZ_MYC2_TF	
	P, protein_JAZ10	
	complex_JAZ_MYC2_TF	
	P, protein_JAZ12	
	complex_JAZ_MYC2_TF	
	P, protein_JAZ3	
	complex_JAZ_MYC2_TF	PMID: 20159850, PMID:
	P, protein_MYC2	21194534, PMID:
	protein_JAZ5 B,	21335373, PMID:
	protein_MYC2	19025383, Zhang H,
	protein_JAZ1 B,	Memelink J (2009)
	protein_MYC2	Regulation of Secondary
	protein_JAZ8 B,	Metabolism by Jasmonate
	protein_MYC2	Hormones. Plant-derived
	protein_JAZ4 B,	Natural Products,
	protein_MYC2	pp.181-194, Note:
	protein_JAZ9 B,	Negative feed back loop to
	protein_MYC2	JAZ is represented as
	protein_JAZ6 B,	binding for simulation
	protein_MYC2	reasons.
	protein_JAZ2 B,	
	protein_MYC2	
	protein_JAZ7 B,	
	protein_MYC2	
	protein_JAZ11 B,	
	protein_MYC2	
	protein_JAZ10 B,	
	protein_MYC2	
	protein_JAZ12 B,	
	protein_MYC2	
	protein_JAZ3 B	

Table A.3: The list of the interactions in the manually constructed structure of the ET sub-model.

ET biological reactions	Directed edge-labelled graph	Data source
protein_ACS + metabolite_SAM activates metabolite_ACC	protein_ACS1 metabolite_ACC A, protein_ACS2 metabolite_ACC A, protein_ACS3 metabolite_ACC A, protein_ACS4 metabolite_ACC A, protein_ACS5 metabolite_ACC A, protein_ACS6 metabolite_ACC A, protein_ACS7 metabolite_ACC A, protein_ACS8 metabolite_ACC A, protein_ACS9 metabolite_ACC A, protein_ACS10 metabolite_ACC A, protein_ACS11 metabolite_ACC A, metabolite_SAM metabolite_ACC A	KEGG
protein_ACO + metabolite_ACC activates metabolite_Ethylene	protein_ACO1 metabolite_Ethylene A, protein_ACO2 metabolite_Ethylene A, protein_ACO4 metabolite_Ethylene A, protein_ACO metabolite_Ethylene A, protein_ACO-like metabolite_Ethylene A, metabolite_ACC metabolite_Ethylene A	KEGG
complex_ASK1_CULLIN1- RBXE2_EBF12 inhibits protein_EIN3EIL1EIL2	complex_ASK1_CULLIN1- RBXE2_EBF12 protein_EIN3 I, complex_ASK1_CULLIN1- RBXE2_EBF12 protein_EIL1 I, complex_ASK1_CULLIN1- RBXE2_EBF12 protein_EIL2 I	PMID: 18692429

ET biological reactions	Directed edge-labelled graph	Data source
metabolite_Copper + protein_RAN1 activates metabolite_Copper_cyto	metabolite_Copper metabolite_Copper_cyto A, protein_RAN1 metabolite_Copper_cyto A	PMID: 12045274
metabolite_Copper_cyto + pro- tein_Inactive_et_receptor activates protein_Et_receptor	metabolite_Copper_cyto protein_ETR1 A, metabolite_Copper_cyto protein_ERS1 A, metabolite_Copper_cyto protein_ETR2 A, metabolite_Copper_cyto protein_ERS2 A, metabolite_Copper_cyto protein_EIN4 A, pro- tein_Inactive_et_receptor protein_ETR1 A, pro- tein_Inactive_et_receptor protein_ERS1 A, pro- tein_Inactive_et_receptor protein_ETR2 A, pro- tein_Inactive_et_receptor protein_ERS2 A, pro- tein_Inactive_et_receptor protein_EIN4 A	PMID: 12045274, PMID: 16920797
complex_Et_receptor- CTR1 inhibits protein_EIN2	complex_Et_receptor- CTR1 protein_EIN2 I	PMID: 20591837, PMID: 18692429, PMID: 21690206, PMID: 19769567
protein_EIN2 activates protein_EIN3EIL1EIL2	protein_EIN2 protein_EIN3 A, protein_EIN2 protein_EIL1 A, protein_EIN2 protein_EIL2 A	PMID: 12045274, PMID: 1103122, PMID:16920797
protein_EIN3EIL1EIL2 + gene_EBF1EBF2 activates protein_EBF1EBF2	protein_EIN3 protein_EBF1 A, protein_EIN3 protein_EBF2 A, protein_EIL1 protein_EBF1 A, protein_EIL1 protein_EBF2 A, protein_EIL2 protein_EBF1 A, protein_EIL2 protein_EBF2 A	PMID:16920797

ET biological reactions	Directed edge-labelled graph	Data source
protein_Inactive_EIN2 activates protein_EIN2	protein_Inactive_EIN2 protein_EIN2 A	PMID: 16920797
protein_EIN3EIL1EIL2 + gene_ERFEDF activates protein_ERFEDF	protein_EIN3 protein_ERF1 A, protein_EIN3 protein_EDF1 A, protein_EIN3 protein_EDF2 A, protein_EIN3 protein_EDF3 A, protein_EIN3 protein_EDF4 A, protein_EIL1 protein_ERF1 A, protein_EIL1 protein_EDF1 A, protein_EIL1 protein_EDF2 A, protein_EIL1 protein_EDF3 A, protein_EIL1 protein_EDF4 A, protein_EIL2 protein_ERF1 A, protein_EIL2 protein_EDF1 A, protein_EIL2 protein_EDF2 A, protein_EIL2 protein_EDF3 A, protein_EIL2 protein_EDF4 A	PMID: 18273012, PMID:11031228
protein_EIN5 inhibits protein_EBF1EBF2	protein_EIN5 protein_EBF1 I, protein_EIN5 protein_EBF2 I	PMID: 18692429, PMID: 16920797, PMID: 17085683
protein_ERFEDF + gene_PDF1.2 activates protein_PDF1.2	protein_ERF1 protein_PDF1.2 A, protein_EDF1 protein_PDF1.2 A, protein_EDF2 protein_PDF1.2 A, protein_EDF3 protein_PDF1.2 A, protein_EDF4 protein_PDF1.2 A	Rojo E, Solano R, Sanchez-Serrano JJ (2003) Interactions between signaling compounds involved in plant defense. Journal of Plant Growth Regulation 22(1): 82-98.



ET biological reactions	Directed edge-labelled graph	Data source
protein_SAM + metabolite_L-methionine activates metabolite_SAM	protein_SAM1 metabolite_SAM A, protein_MAT3 metabolite_SAM A, protein_MAT4 metabolite_SAM A, protein_SAM2 metabolite_SAM A, metabolite_L-methionine metabolite_SAM A	AraCyc database
protein_EIN3EIL1EIL2 + gene_GST1b-CHIPR4 activates protein_GST1b-CHIPR4	protein_EIN3 protein_GST1 A, protein_EIN3 protein_b-CHIA, protein_EIN3 protein_PR4 A, protein_EIL1 protein_GST1 A, protein_EIL1 protein_b-CHIA, protein_EIL1 protein_PR4 A, protein_EIL2 protein_GST1 A, protein_EIL2 protein_b-CHIA, protein_EIL2 protein_PR4 A	PMID: 8090746
metabolite_Ethylene inhibits com- plex_Et_receptor_CTR1	metabolite_Ethylene com- plex_Et_receptor_CTR1 I	PMID: 1204527
complex_ASK1_CULLIN1- RBXE2 + protein_EBF1EBF2 binds complex_ASK1_CULLIN1- RBXE2_EBF12	complex_ASK1_CULLIN1- RBXE2 complex_ASK1_CULLIN1- RBXE2_EBF12 P, protein_EBF1 complex_ASK1_CULLIN1- RBXE2_EBF12 P, protein_EBF2 complex_ASK1_CULLIN1- RBXE2_EBF12 P, complex_ASK1_CULLIN1- RBXE2 protein_EBF1 B, complex_ASK1_CULLIN1- RBXE2 protein_EBF2 B	PMID: 16920797

ET biological reactions	Directed edge-labelled graph	Data source
protein_CTR1 + protein_Et_receptor binds complex_Et_receptor_CTR1	protein_CTR1 complex_Et_receptor_CTR1 P, protein_ETR1 complex_Et_receptor_CTR1 P, protein_ERS1 complex_Et_receptor_CTR1 P, protein_ETR2 complex_Et_receptor_CTR1 P, protein_ERS2 complex_Et_receptor_CTR1 P, protein_EIN4 complex_Et_receptor_CTR1 P, protein_CTR1 protein_ETR1 B, protein_CTR1 protein_ERS1 B, protein_CTR1 protein_ETR2 B, protein_CTR1 protein_ERS2 B, protein_CTR1 protein_EIN4 B	PMID: 1204527, PMID: 16920797

Table A.4: The list of manually acquired crosstalk interactions between SA, JA and ET sub-model.

Crosstalk biological reactions	Directed edge-labelled graph	Data source
protein_MYC2 + gene_GST1b-CHIPR4 activates protein_GST1b-CHIPR4	protein_MYC2 protein_GST1 A, protein_MYC2 protein_b-CHI A, protein_MYC2 protein_PR4 A	PMID: 20864543, PMID: 20864543, PMID: 17616737
protein_MPK4 inhibits protein_EDS1	protein_MPK4 protein_EDS1 I	PMID: 16813576
protein_MPK4 inhibits protein_PAD34	protein_MPK4 protein_PAD4 I, protein_MPK4 protein_PAD3 I	PMID: 22408091, PMID: 16813576
metabolite_SA inhibits metabolite_OPDA	metabolite_SA metabolite_OPDA I	Rajo E, Solano R, Sanchez-Serrano JJ (2003) Interactions between signaling compounds involved in plant defense. Journal of Plant Growth Regulation 22(1): 82-98.

## **A.2 The levels of the biological component abstraction**

The Table A.5 presents the levels of the abstraction of biological components that are included in the manually constructed PDS model structure.

The first column in the table represents the family name of the biological component. The second column contains all the single members of the family that are considered important for the plant defence. The third column contains their unique ID numbers from the TAIR database.

Table A.5: The list of manually acquired crosstalk interactions between SA, JA and ET sub-model.

Family name (level 2)	Individual short name (level 3)	TAIR gene ID
SA	SA	/
HRT	HRT	AT5G43470
NPR1	NPR1	AT1G64280
NIMIN	NIMIN1	AT1G02450
NIMIN	NIMIN2	AT3G25882
NIMIN	NIMIN3	AT1G09415
PAD34	PAD4	AT3G52430
PAD34	PAD3	AT3G26830
EDS1	EDS1	AT3G48090
EDS5	EDS5	AT4G39030
PR125	PR1	AT2G14610
PR125	PR2	AT3G57260
PR125	PR5	AT1G75040
ICS	ICS1	At1g74710
ICS	ICS2	AT1G18870
WRKY70	WRKY70	At3g56400
MPK4	MPK4	At4g01370
Chorismate	Chorismate	/
Isochorismate	Isochorismate	/
BA	BA	/
BA2H	BA2H	/
ROS_productive_enzymes	NADPH_oxidase	/
ROS_productive_enzymes	Catalase	/
ROS_productive_enzymes	Glutathione peroxidase	/
ROS_productive_enzymes	Superoxide_dismutase	/
SAG	SAG	/
SGE	SGE	/
Trans-cinnamic_acid	Trans-cinnamic_acid	/
Phenylalanine	Phenylalanine	/
Phenylpyruvate	Phenylpyruvate	/
ROS	ROS	/
PAL	PAL1	AT2G37040
PAL	PAL2	AT3G53260
PAL	PAL3	AT5G04230
PAL	PAL4	AT3G10340
IPL	IPL	/
CM	CM1	AT3G29200
CM	CM2	At5g10870
CM	CM3	At1g69370
Prephenate_amino-transferase	Prephenate_amino-transferase	At2G22250

Family name (level 2)	Individual short name (level 3)	TAIR gene ID
Prephenate	Prephenate	/
Arogenate_dehydratase	Arogenate_dehydratase	At5g22630
Orto-coumaric_acid	Orto-coumaric_acid	/
UGP_glikosyltransferaze	UGP_glikosyltransferaze	AT1G01420
Ethylene	Ethylene	/
Copper	Copper	/
SAM	SAM	/
ACC	ACC	/
ACS	ACS1	AT3G61510
ACS	ACS2	AT1G01480
ACS	ACS3	AT5G28360
ACS	ACS4	AT2G22810
ACS	ACS5	AT5G65800
ACS	ACS6	AT4G11280
ACS	ACS7	AT4G26200
ACS	ACS8	AT4G37770
ACS	ACS9	AT3G49700
ACS	ACS10	AT1G62960
ACS	ACS11	AT4G08040
ACO	ACO1	AT2G19590
ACO	ACO2	AT1G62380
ACO	ACO4	AT1G05010
ACO	ACO	AT1G12010
ACO	ACO-like	AT1G77330
RAN1	RAN1	AT5G44790
Et_receptor	ETR1	AT1G66340
Et_receptor	ERS1	AT2G40940
Et_receptor	ETR2	AT3G23150
Et_receptor	ERS2	AT1G04310
Et_receptor	EIN4	AT3G04580
CTR1	CTR1	AT5G03730
EIN2	EIN2	AT5G03280
EIN3EIL1EIL2	EIN3	AT3G20770
EIN3EIL1EIL2	EIL1	AT2G27050
EIN3EIL1EIL2	EIL2	AT5G21120
ERFEDF	ERF1	AT3G23240
ERFEDF	EDF1	AT1G25560
ERFEDF	EDF2	AT1G68840
ERFEDF	EDF3	AT3G25730
ERFEDF	EDF4	AT1G13260
EBF1EBF2	EBF1	AT2G25490
EBF1EBF2	EBF2	AT5G25350
EIN5	EIN5	AT1G54490

Family name (level 2)	Individual short name (level 3)	TAIR gene ID
JA	JA	/
Linolenic_acid	Linolenic_acid	/
LOX	LOX1	AT1G55020
LOX	LOX2	AT3G45140
LOX	LOX3	AT1G17420
LOX	LOX4	AT1G72520
LOX	LOX5	AT3G22400
LOX	LOX6	AT1G67560
AOS	AOS	AT5G42650
AOC	AOC1	AT3G25760
AOC	AOC2	AT3G25770
AOC	AOC3	AT3G25780
AOC	AOC4	AT1G13280
13-HPT	13-HPT	/
12/13_EDT	12/13_EDT	/
OPDA	OPDA	/
JMT	JMT	AT1G19640
JA-Ile	JA-Ile	/
COI1	COI1	AT2G39940
JAR1	JAR1	AT2G46370
JAZ	JAZ1	AT1G19180
JAZ	JAZ2	AT1G74950
JAZ	JAZ3	At3g17860
JAZ	JAZ4	AT1G48500
JAZ	JAZ5	AT1G17380
JAZ	JAZ6	AT1G72450
JAZ	JAZ7	AT2G34600
JAZ	JAZ8	AT1G30135
JAZ	JAZ9	AT1G70700
JAZ	JAZ10	AT5G13220
JAZ	JAZ11	AT3G43440
JAZ	JAZ12	AT5G20900
MYC2	MYC2	AT1G32640
GST1b-CHIPR4	GST1	AT1G02930
GST1b-CHIPR4	PR4	AT3G04720
GST1b-CHIPR4	b-CHI	AT3G12500
PDF1.2	PDF1.2	AT5G44420
MPK6	MPK6	At2g43790
MPK3	MPK3	AT3G45640
THI2.1JR1VSP1CLH1	THI2.1	AT1G72260
THI2.1JR1VSP1CLH1	JR1	AT3G16470
THI2.1JR1VSP1CLH1	VSP1	AT5G24780
THI2.1JR1VSP1CLH1	CLH1	AT1G19670

Family name (level 2)	Individual short name (level 3)	TAIR gene ID
TGA_TF245	TGA_TF2	At5g06950
TGA_TF245	TGA_TF4	At5g10030
TGA_TF245	TGA_TF5	At5g06960
OPCL1ACXAIM1KAT	OPCL1	AT1G20510
OPCL1ACXAIM1KAT	ACX1	AT4G16760
OPCL1ACXAIM1KAT	ACX2	AT5G65110
OPCL1ACXAIM1KAT	ACX3	AT1G06290
OPCL1ACXAIM1KAT	ACX4	AT3G51840
OPCL1ACXAIM1KAT	ACX5	AT2G35690
OPCL1ACXAIM1KAT	ACX6	AT1G06310
OPCL1ACXAIM1KAT	AIM1	AT4G29010
OPCL1ACXAIM1KAT	KAT1	AT1G04710
OPCL1ACXAIM1KAT	KAT2	AT2G33150
OPCL1ACXAIM1KAT	KAT5	AT5G48880
ACXAIM1KAT	ACX1	AT4G16760
ACXAIM1KAT	ACX2	AT5G65110
ACXAIM1KAT	ACX3	AT1G06290
ACXAIM1KAT	ACX4	AT3G51840
ACXAIM1KAT	ACX5	AT2G35690
ACXAIM1KAT	ACX6	AT1G06310
ACXAIM1KAT	AIM1	AT4G29010
ACXAIM1KAT	KAT1	AT1G04710
ACXAIM1KAT	KAT2	AT2G33150
ACXAIM1KAT	KAT5	AT5G48880
OPR3	OPR3	AT2G06050
OPC8	OPC8	/
Me-JA	Me-JA	/
OPC6	OPC6	/
OPC4	OPC4	/
OPC-8:0-CoA	OPC-8:0-CoA	/
SAM	SAM1	AT1G02500
SAM	SAM2	AT4G01850
SAM	MAT3	AT2G36880
SAM	MAT4	AT3G17390
SCF	SCF	
AAO4	AAO4	AT1g04580
L-methionine	L-methionine	/

### A.3 A manual expanded graph file at the single component level

The network consists of 175 components and 387 reactions and represents the initial manually constructed PDS model structure. The graph, which should be saved as .bmg file named, for example, "graph.bmg", contains the following reactions and components:

```
# _canvas -303.0,-78.7,10401.8,2703.5
# _symmetric B
complex_ASK1_CULLIN1_RBXE2 complex_ASK1_CULLIN1_RBXE2_EBF12 P
linecolor=255/0/0 pos=4088.7,1212.2
complex_ASK1_CULLIN1_RBXE2 protein_EBF1 B
linecolor=255/0/0 pos=4001.0,1192.7
complex_ASK1_CULLIN1_RBXE2 protein_EBF2 B linecolor=255/0/0 pos=3930.3,1271.9
complex_ASK1_CULLIN1_RBXE2_EBF12 protein_EIL1 I
linecolor=255/0/0 pos=3944.6,1404.8
complex_ASK1_CULLIN1_RBXE2_EBF12 protein_EIL2 I
linecolor=255/0/0 pos=3984.7,1460.5
complex_ASK1_CULLIN1_RBXE2_EBF12 protein_EIN3 I
linecolor=255/0/0 pos=3967.2,1531.8
complex_Et_receptor_CTR1 protein_EIN2 I linecolor=255/0/0 pos=3523.9,1676.4
complex_JA-Ile_COI1_SCF complex_JA-Ile_COI1_SCF_JAZ P
linecolor=255/0/0 pos=6609.2,2482.1
complex_JA-Ile_COI1_SCF protein_JAZ1 B linecolor=255/0/0 pos=6107.5,2338.6
complex_JA-Ile_COI1_SCF protein_JAZ10 B linecolor=255/0/0 pos=6826.6,2329.1
complex_JA-Ile_COI1_SCF protein_JAZ11 B linecolor=255/0/0 pos=6902.4,2330.5
complex_JA-Ile_COI1_SCF protein_JAZ12 B linecolor=255/0/0 pos=6983.9,2326.4
complex_JA-Ile_COI1_SCF protein_JAZ2 B linecolor=255/0/0 pos=6208.1,2347.9
complex_JA-Ile_COI1_SCF protein_JAZ3 B linecolor=255/0/0 pos=6295.8,2359.1
complex_JA-Ile_COI1_SCF protein_JAZ4 B linecolor=255/0/0 pos=6373.8,2343.4
complex_JA-Ile_COI1_SCF protein_JAZ5 B linecolor=255/0/0 pos=6447.1,2361.0
complex_JA-Ile_COI1_SCF protein_JAZ6 B linecolor=255/0/0 pos=6520.7,2343.6
complex_JA-Ile_COI1_SCF protein_JAZ7 B linecolor=255/0/0 pos=6595.8,2345.9
complex_JA-Ile_COI1_SCF protein_JAZ8 B linecolor=255/0/0 pos=6673.4,2348.8
complex_JA-Ile_COI1_SCF protein_JAZ9 B linecolor=255/0/0 pos=6749.4,2323.3
complex_NPR1_TGA245 protein_PR1 A linecolor=255/0/0 pos=10011.3,1373.3
complex_NPR1_TGA245 protein_PR2 A linecolor=255/0/0 pos=10019.3,1473.3
complex_NPR1_TGA245 protein_PR5 A linecolor=255/0/0 pos=10076.7,1556.0
complex_NPR1_oligomer protein_NPR1 A linecolor=255/0/0 pos=9788.3,1207.4
complex_SCF complex_JA-Ile_COI1_SCF P linecolor=255/0/0 pos=6035.3,2531.3
complex_SCF metabolite_JA-Ile B linecolor=255/0/0 pos=5747.4,2525.8
complex_SCF protein_COI1 B linecolor=255/0/0 pos=5821.7,2623.8
metabolite_12/13_EDT metabolite_OPDA_chl A linecolor=255/0/0 pos=5114.8,1265.4
metabolite_13_HPT metabolite_12/13_EDT A linecolor=255/0/0 pos=5113.1,1116.4
metabolite_ACC metabolite_Ethylene A linecolor=255/0/0 pos=2923.9,1392.9
metabolite_BA metabolite_SA A linecolor=255/0/0 pos=8786.5,1685.6
metabolite_Chorismate metabolite_Isochorismate A linecolor=255/0/0 pos=8725.7,982.8
metabolite_Chorismate metabolite_Prephenate A
linecolor=255/0/0 pinned=1 pos=8456.1045,811.2770
metabolite_Copper metabolite_Copper_cyto A linecolor=255/0/0 pos=3349.4,923.0
metabolite_Copper_cyto protein_EIN4 A linecolor=255/0/0 pos=3523.3,1081.2
metabolite_Copper_cyto protein_ERS1 A linecolor=255/0/0 pos=3353.2,1190.4
metabolite_Copper_cyto protein_ERS2 A linecolor=255/0/0 pos=3449.8,1164.6
```



metabolite\_Copper\_cyto protein\_ETR1 A linecolor=255/0/0 pos=3210.9,1124.3  
metabolite\_Copper\_cyto protein\_ETR2 A linecolor=255/0/0 pos=3278.3,1134.0  
metabolite\_Ethylene complex\_Et\_receptor\_CTR1 I linecolor=255/0/0 pos=3297.5,1599.1  
metabolite\_Isochorismate metabolite\_SA\_chl A linecolor=255/0/0 pos=9007.9,1386.5  
metabolite\_JA metabolite\_JA-Ile A linecolor=255/0/0 pos=5359.5,2477.9  
metabolite\_JA metabolite\_Me-JA A linecolor=255/0/0 pos=5240.8,2556.7  
metabolite\_JA-Ile complex\_JA-Ile\_COI1\_SCF P linecolor=255/0/0 pos=5759.8,2429.3  
metabolite\_JA-Ile protein\_COI1 B linecolor=255/0/0 pos=5569.3,2555.0  
metabolite\_L-methionine metabolite\_SAM A linecolor=255/0/0 pos=2796.4,745.4  
metabolite\_Linolenic\_acid metabolite\_13-HPT A linecolor=255/0/0 pos=5104.3,962.9  
metabolite\_OPC-8:0-CoA metabolite\_OPC6 A linecolor=255/0/0 pos=5259.2,1758.7  
metabolite\_OPC4 metabolite\_JA A linecolor=255/0/0 pos=5138.0,2375.3  
metabolite\_OPC6 metabolite\_OPC4 A  
linecolor=255/0/0 pinned=1 pos=5245.0036,2026.1763  
metabolite\_OPC8 metabolite\_OPC-8:0-CoA A linecolor=255/0/0 pos=5267.0,1640.9  
metabolite\_OPDA metabolite\_OPC8 A linecolor=255/0/0 pos=5252.0,1525.8  
metabolite\_OPDA\_chl metabolite\_OPDA A linecolor=255/0/0 pos=5120.5,1394.9  
metabolite\_Orto-coumaric\_acid metabolite\_SA A linecolor=255/0/0 pos=8612.9,1742.1  
metabolite\_Phenylalanine metabolite\_Trans-cinnamic\_acid A  
linecolor=255/0/0 pos=8501.2,1308.2  
metabolite\_Phenylpyruvate metabolite\_Phenylalanine A  
linecolor=255/0/0 pos=8516.1,1142.6  
metabolite\_Prephenate metabolite\_Phenylpyruvate A linecolor=255/0/0 pos=8511.6,977.1  
metabolite\_ROS protein\_BA2H A linecolor=255/0/0 pos=9566.8,1181.2  
metabolite\_ROS protein\_MPK3 A linecolor=255/0/0 pos=9435.7,752.2  
metabolite\_ROS protein\_MPK6 A linecolor=255/0/0 pinned=1 pos=9694.2650,804.1757  
metabolite\_SA metabolite\_OPDA I linecolor=255/0/0 pos=6990.0,1741.2  
metabolite\_SA metabolite\_SAG A linecolor=255/0/0 pos=8903.7,1997.8  
metabolite\_SA metabolite\_SGE A linecolor=255/0/0 pos=8664.4,2014.6  
metabolite\_SA protein\_NPR1 A linecolor=255/0/0 pinned=1 pos=9584.0072,1631.6197  
metabolite\_SAM metabolite\_ACC A linecolor=255/0/0 pos=2806.8,972.5  
metabolite\_SA\_chl metabolite\_SA A linecolor=255/0/0 pinned=1 pos=8999.1799,1585.9679  
metabolite\_Trans-cinnamic\_acid metabolite\_BA A linecolor=255/0/0 pos=8612.4,1387.0  
metabolite\_Trans-cinnamic\_acid metabolite\_Orto-coumaric\_acid A  
linecolor=255/0/0 pinned=1 pos=8450.7992,1455.7455  
protein\_AAO4 metabolite\_BA A linecolor=255/0/0 pinned=1 pos=8748.5817,1512.1865  
protein\_AAO4 metabolite\_Orto-coumaric\_acid A  
linecolor=255/0/0 pinned=1 pos=8556.8009,1510.1879  
protein\_ACO metabolite\_Ethylene A linecolor=255/0/0 pos=2895.4,1641.4  
protein\_ACO-like metabolite\_Ethylene A linecolor=255/0/0 pos=2942.1,1698.6  
protein\_ACO1 metabolite\_Ethylene A linecolor=255/0/0 pos=2897.0,1475.1  
protein\_ACO2 metabolite\_Ethylene A linecolor=255/0/0 pos=2907.1,1545.1  
protein\_ACO4 metabolite\_Ethylene A linecolor=255/0/0 pos=2850.9,1586.1  
protein\_ACS1 metabolite\_ACC A linecolor=255/0/0 pos=2879.2,1023.3  
protein\_ACS10 metabolite\_ACC A linecolor=255/0/0 pos=2724.5,1091.1  
protein\_ACS11 metabolite\_ACC A linecolor=255/0/0 pos=2735.8,1026.2  
protein\_ACS2 metabolite\_ACC A linecolor=255/0/0 pos=2867.3,1085.3  
protein\_ACS3 metabolite\_ACC A linecolor=255/0/0 pos=2919.3,1130.7  
protein\_ACS4 metabolite\_ACC A linecolor=255/0/0 pos=2920.1,1204.5  
protein\_ACS5 metabolite\_ACC A linecolor=255/0/0 pos=2860.7,1249.2  
protein\_ACS6 metabolite\_ACC A linecolor=255/0/0 pos=2894.5,1303.0  
protein\_ACS7 metabolite\_ACC A linecolor=255/0/0 pos=2712.6,1280.9

protein\_ACS8 metabolite\_ACC A linecolor=255/0/0 pos=2709.3,1214.2  
 protein\_ACS9 metabolite\_ACC A linecolor=255/0/0 pos=2707.1,1151.5  
 protein\_ACX1 metabolite\_JA A linecolor=255/0/0 pos=4844.9,2327.1  
 protein\_ACX1 metabolite\_OPC4 A linecolor=255/0/0 pos=4843.6,2228.7  
 protein\_ACX1 metabolite\_OPC6 A linecolor=255/0/0 pos=4859.5,1966.6  
 protein\_ACX2 metabolite\_JA A linecolor=255/0/0 pos=4921.5,2345.8  
 protein\_ACX2 metabolite\_OPC4 A linecolor=255/0/0 pos=4929.4,2224.0  
 protein\_ACX2 metabolite\_OPC6 A linecolor=255/0/0 pos=4934.5,1983.9  
 protein\_ACX3 metabolite\_JA A linecolor=255/0/0 pos=4991.3,2364.5  
 protein\_ACX3 metabolite\_OPC4 A linecolor=255/0/0 pos=4999.5,2247.4  
 protein\_ACX3 metabolite\_OPC6 A linecolor=255/0/0 pos=5005.0,1992.9  
 protein\_ACX4 metabolite\_JA A linecolor=255/0/0 pos=5055.0,2332.0  
 protein\_ACX4 metabolite\_OPC4 A linecolor=255/0/0 pos=5052.6,2199.7  
 protein\_ACX4 metabolite\_OPC6 A linecolor=255/0/0 pos=5071.5,1979.0  
 protein\_ACX5 metabolite\_JA A linecolor=255/0/0 pos=5129.7,2300.1  
 protein\_ACX5 metabolite\_OPC4 A linecolor=255/0/0 pos=5116.1,2208.0  
 protein\_ACX5 metabolite\_OPC6 A linecolor=255/0/0 pos=5142.6,1964.7  
 protein\_ACX6 metabolite\_JA A linecolor=255/0/0 pos=5229.6,2317.0  
 protein\_ACX6 metabolite\_OPC4 A linecolor=255/0/0 pos=5187.0,2149.6  
 protein\_ACX6 metabolite\_OPC6 A linecolor=255/0/0 pos=5226.9,1958.1  
 protein\_AIM1 metabolite\_JA A linecolor=255/0/0 pos=5305.4,2310.3  
 protein\_AIM1 metabolite\_OPC4 A linecolor=255/0/0 pos=5298.2,2205.0  
 protein\_AIM1 metabolite\_OPC6 A linecolor=255/0/0 pos=5314.0,1951.1  
 protein\_AOC1 metabolite\_OPDA\_chl A linecolor=255/0/0 pos=5352.7,1274.1  
 protein\_AOC2 metabolite\_OPDA\_chl A linecolor=255/0/0 pos=5317.9,1327.5  
 protein\_AOC3 metabolite\_OPDA\_chl A linecolor=255/0/0 pos=5374.8,1363.8  
 protein\_AOC4 metabolite\_OPDA\_chl A linecolor=255/0/0 pos=5349.3,1426.0  
 protein\_AOS metabolite\_12/13\_EDT A linecolor=255/0/0 pos=5277.4,1203.9  
 protein\_Arogenate\_dehydratase metabolite\_Phenylalanine A  
 linecolor=255/0/0 pos=8337.4,1201.2  
 protein\_BA2H metabolite\_SA A linecolor=255/0/0 pos=9049.0,1747.6  
 protein\_CM1 metabolite\_Prephenate A linecolor=255/0/0 pos=8358.5,781.6  
 protein\_CM2 metabolite\_Prephenate A linecolor=255/0/0 pos=8359.4,938.1  
 protein\_CM3 metabolite\_Prephenate A linecolor=255/0/0 pos=8357.3,856.3  
 protein\_COI1 complex\_JA-Ile\_COI1\_SCF P linecolor=255/0/0 pos=5839.6,2520.8  
 protein\_CTR1 complex\_Et\_receptor\_CTR1 P  
 linecolor=255/0/0 pinned=1 pos=3531.1004,1464.2793  
 protein\_CTR1 protein\_EIN4 B linecolor=255/0/0 pos=3651.8,1317.6  
 protein\_CTR1 protein\_ERS1 B linecolor=255/0/0 pos=3483.3,1344.2  
 protein\_CTR1 protein\_ERS2 B linecolor=255/0/0 pos=3567.3,1332.9  
 protein\_CTR1 protein\_ETR1 B linecolor=255/0/0 pos=3325.0,1355.0  
 protein\_CTR1 protein\_ETR2 B linecolor=255/0/0 pos=3402.8,1349.8  
 protein\_Catalase metabolite\_ROS A linecolor=255/0/0 pos=9962.6,804.8  
 protein\_EBF1 complex\_ASK1\_CULLIN1\_RBXE2\_EBF12 P  
 linecolor=255/0/0 pos=4078.8,986.3  
 protein\_EBF2 complex\_ASK1\_CULLIN1\_RBXE2\_EBF12 P  
 linecolor=255/0/0 pos=3970.5,1127.6  
 protein\_EDF1 protein\_PDF1.2 A linecolor=255/0/0 pos=4437.3,1753.5  
 protein\_EDF2 protein\_PDF1.2 A linecolor=255/0/0 pos=4420.8,1862.8  
 protein\_EDF3 protein\_PDF1.2 A linecolor=255/0/0 pos=4477.8,1909.9  
 protein\_EDF4 protein\_PDF1.2 A linecolor=255/0/0 pos=4430.5,1967.0  
 protein\_EDS1 protein\_EDS5 A linecolor=255/0/0 pos=9452.5,1025.4

protein\_EDS5 protein\_ICS1 A linecolor=255/0/0 pinned=1 pos=9277.5793,1190.3358  
 protein\_EDS5 protein\_ICS2 A linecolor=255/0/0 pinned=1 pos=9411.1688,1188.8498  
 protein\_EIL1 protein\_EBF1 A linecolor=255/0/0 pos=3866.9,1358.0  
 protein\_EIL1 protein\_EBF2 A linecolor=255/0/0 pos=3784.7,1489.4  
 protein\_EIL1 protein\_EDF1 A linecolor=255/0/0 pos=4098.3,1718.4  
 protein\_EIL1 protein\_EDF2 A linecolor=255/0/0 pos=4118.5,1776.5  
 protein\_EIL1 protein\_EDF3 A linecolor=255/0/0 pos=4133.4,1838.0  
 protein\_EIL1 protein\_EDF4 A linecolor=255/0/0 pos=4109.5,1899.9  
 protein\_EIL1 protein\_ERF1 A linecolor=255/0/0 pos=4050.1,1672.6  
 protein\_EIL1 protein\_GST1 A linecolor=255/0/0 pos=4197.6,1524.8  
 protein\_EIL1 protein\_PR4 A linecolor=255/0/0 pos=4207.7,1395.9  
 protein\_EIL1 protein\_b-CHI A linecolor=255/0/0 pos=4192.9,1302.6  
 protein\_EIL2 protein\_EBF1 A linecolor=255/0/0 pos=3860.8,1425.1  
 protein\_EIL2 protein\_EBF2 A linecolor=255/0/0 pos=3786.2,1555.8  
 protein\_EIL2 protein\_EDF1 A linecolor=255/0/0 pos=4054.8,1781.8  
 protein\_EIL2 protein\_EDF2 A linecolor=255/0/0 pos=4071.5,1843.8  
 protein\_EIL2 protein\_EDF3 A linecolor=255/0/0 pos=3989.3,1917.7  
 protein\_EIL2 protein\_EDF4 A linecolor=255/0/0 pos=4101.9,1971.8  
 protein\_EIL2 protein\_ERF1 A linecolor=255/0/0 pos=4012.6,1731.3  
 protein\_EIL2 protein\_GST1 A linecolor=255/0/0 pos=4124.1,1558.8  
 protein\_EIL2 protein\_PR4 A linecolor=255/0/0 pos=4210.1,1461.0  
 protein\_EIL2 protein\_b-CHI A linecolor=255/0/0 pos=4145.5,1350.3  
 protein\_EIN2 protein\_EIL1 A linecolor=255/0/0 pos=3643.5,1760.9  
 protein\_EIN2 protein\_EIL2 A linecolor=255/0/0 pos=3633.9,1836.3  
 protein\_EIN2 protein\_EIN3 A linecolor=255/0/0 pos=3628.1,1911.8  
 protein\_EIN3 protein\_EBF1 A linecolor=255/0/0 pos=3883.3,1498.9  
 protein\_EIN3 protein\_EBF2 A linecolor=255/0/0 pos=3811.7,1619.7  
 protein\_EIN3 protein\_EDF1 A linecolor=255/0/0 pos=4022.9,1868.6  
 protein\_EIN3 protein\_EDF2 A linecolor=255/0/0 pos=4056.2,1931.5  
 protein\_EIN3 protein\_EDF3 A linecolor=255/0/0 pos=4023.6,1990.1  
 protein\_EIN3 protein\_EDF4 A linecolor=255/0/0 pos=4065.0,2042.1  
 protein\_EIN3 protein\_ERF1 A linecolor=255/0/0 pos=3995.4,1807.7  
 protein\_EIN3 protein\_GST1 A linecolor=255/0/0 pos=4177.2,1612.2  
 protein\_EIN3 protein\_PR4 A linecolor=255/0/0 pos=4136.9,1498.8  
 protein\_EIN3 protein\_b-CHI A linecolor=255/0/0 pos=4142.8,1422.3  
 protein\_EIN4 complex\_Et\_receptor\_CTR1 P linecolor=255/0/0 pos=3668.2,1403.9  
 protein\_EIN5 protein\_EBF1 I linecolor=255/0/0 pos=3889.0,957.1  
 protein\_EIN5 protein\_EBF2 I linecolor=255/0/0 pos=3824.2,1119.2  
 protein\_ERF1 protein\_PDF1.2 A linecolor=255/0/0 pos=4459.0,1688.9  
 protein\_ERS1 complex\_Et\_receptor\_CTR1 P linecolor=255/0/0 pos=3474.1,1440.1  
 protein\_ERS2 complex\_Et\_receptor\_CTR1 P linecolor=255/0/0 pos=3611.4,1454.6  
 protein\_ETR1 complex\_Et\_receptor\_CTR1 P linecolor=255/0/0 pos=3331.4,1450.1  
 protein\_ETR2 complex\_Et\_receptor\_CTR1 P linecolor=255/0/0 pos=3405.8,1458.6  
 protein\_Glutathione\_peroxidase metabolite\_ROS A linecolor=255/0/0 pos=9894.9,687.7  
 protein\_HRT protein\_MPK3 A linecolor=255/0/0 pos=9365.8,626.8  
 protein\_HRT protein\_MPK6 A linecolor=255/0/0 pos=9594.7,640.4  
 protein\_ICS1 metabolite\_Isochorismate A linecolor=255/0/0 pos=9117.6,1263.8  
 protein\_ICS2 metabolite\_Isochorismate A linecolor=255/0/0 pos=9191.6,1288.0  
 protein\_IPL metabolite\_SA\_chl A linecolor=255/0/0 pos=8911.3,1391.1  
 protein\_Inactive\_BA2H protein\_BA2H A linecolor=255/0/0 pos=9362.9,1532.6  
 protein\_Inactive\_EIN2 protein\_EIN2 A  
 linecolor=255/0/0 pinned=1 pos=3402.2653,1772.3357

protein\_Inactive\_HRT protein\_HRT A linecolor=255/0/0 pinned=1 pos=9791.7682,615.9390  
 protein\_Inactive\_MPK3 protein\_MPK3 A  
 linecolor=255/0/0 pinned=1 pos=9064.9240,708.8540  
 protein\_Inactive\_MPK6 protein\_MPK6 A  
 linecolor=255/0/0 pinned=1 pos=9507.8179,725.7973  
 protein\_Inactive\_et\_receptor protein\_EIN4 A linecolor=255/0/0 pos=3685.9,1095.7  
 protein\_Inactive\_et\_receptor protein\_ERS1 A linecolor=255/0/0 pos=3526.1,1154.8  
 protein\_Inactive\_et\_receptor protein\_ERS2 A linecolor=255/0/0 pos=3609.7,1146.5  
 protein\_Inactive\_et\_receptor protein\_ETR1 A linecolor=255/0/0 pos=3366.9,1130.1  
 protein\_Inactive\_et\_receptor protein\_ETR2 A linecolor=255/0/0 pos=3437.4,1093.3  
 protein\_JAR1 metabolite\_JA-Ile A linecolor=255/0/0 pos=5599.0,2407.6  
 protein\_JAZ1 complex\_JA-Ile\_COI1\_SCF\_JAZ P linecolor=255/0/0 pos=6749.7,2395.7  
 protein\_JAZ1 complex\_JAZ\_MYC2\_TF P linecolor=255/0/0 pos=6881.5,1642.8  
 protein\_JAZ1 protein\_CLH1 I linecolor=255/0/0 pos=6335.1,1479.6  
 protein\_JAZ1 protein\_JR1 I linecolor=255/0/0 pos=6253.1,1426.9  
 protein\_JAZ1 protein\_MYC2 B linecolor=255/0/0 pos=6714.2,1401.8  
 protein\_JAZ1 protein\_THI2.1 I linecolor=255/0/0 pos=6459.9,1568.9  
 protein\_JAZ1 protein\_VSP1 I linecolor=255/0/0 pos=6404.7,1518.5  
 protein\_JAZ10 complex\_JA-Ile\_COI1\_SCF\_JAZ P linecolor=255/0/0 pos=7485.7,2346.2  
 protein\_JAZ10 complex\_JAZ\_MYC2\_TF P linecolor=255/0/0 pos=7638.7,1576.4  
 protein\_JAZ10 protein\_CLH1 I linecolor=255/0/0 pos=7110.8,1458.0  
 protein\_JAZ10 protein\_JR1 I linecolor=255/0/0 pos=7008.9,1318.8  
 protein\_JAZ10 protein\_MYC2 B linecolor=255/0/0 pos=7467.3,1375.9  
 protein\_JAZ10 protein\_THI2.1 I linecolor=255/0/0 pos=7229.3,1558.8  
 protein\_JAZ10 protein\_VSP1 I linecolor=255/0/0 pos=7182.1,1508.9  
 protein\_JAZ11 complex\_JA-Ile\_COI1\_SCF\_JAZ P linecolor=255/0/0 pos=7559.9,2366.4  
 protein\_JAZ11 complex\_JAZ\_MYC2\_TF P linecolor=255/0/0 pos=7706.7,1549.9  
 protein\_JAZ11 protein\_CLH1 I linecolor=255/0/0 pos=7194.3,1452.3  
 protein\_JAZ11 protein\_JR1 I linecolor=255/0/0 pos=7088.8,1316.7  
 protein\_JAZ11 protein\_MYC2 B linecolor=255/0/0 pos=7540.1,1381.7  
 protein\_JAZ11 protein\_THI2.1 I linecolor=255/0/0 pos=7311.4,1548.7  
 protein\_JAZ11 protein\_VSP1 I linecolor=255/0/0 pos=7262.7,1501.0  
 protein\_JAZ12 complex\_JA-Ile\_COI1\_SCF\_JAZ P linecolor=255/0/0 pos=7642.2,2357.8  
 protein\_JAZ12 complex\_JAZ\_MYC2\_TF P linecolor=255/0/0 pos=7795.1,1534.1  
 protein\_JAZ12 protein\_CLH1 I linecolor=255/0/0 pos=7271.2,1441.7  
 protein\_JAZ12 protein\_JR1 I linecolor=255/0/0 pos=7174.4,1323.6  
 protein\_JAZ12 protein\_MYC2 B linecolor=255/0/0 pos=7623.6,1385.2  
 protein\_JAZ12 protein\_THI2.1 I linecolor=255/0/0 pos=7398.6,1528.3  
 protein\_JAZ12 protein\_VSP1 I linecolor=255/0/0 pos=7345.1,1485.6  
 protein\_JAZ2 complex\_JA-Ile\_COI1\_SCF\_JAZ P linecolor=255/0/0 pos=6852.0,2398.0  
 protein\_JAZ2 complex\_JAZ\_MYC2\_TF P linecolor=255/0/0 pos=6992.2,1633.8  
 protein\_JAZ2 protein\_CLH1 I linecolor=255/0/0 pos=6445.9,1457.2  
 protein\_JAZ2 protein\_JR1 I linecolor=255/0/0 pos=6355.0,1399.3  
 protein\_JAZ2 protein\_MYC2 B linecolor=255/0/0 pos=6801.4,1420.6  
 protein\_JAZ2 protein\_THI2.1 I linecolor=255/0/0 pos=6560.2,1575.7  
 protein\_JAZ2 protein\_VSP1 I linecolor=255/0/0 pos=6515.0,1512.6  
 protein\_JAZ3 complex\_JA-Ile\_COI1\_SCF\_JAZ P linecolor=255/0/0 pos=6944.8,2398.5  
 protein\_JAZ3 complex\_JAZ\_MYC2\_TF P linecolor=255/0/0 pos=7088.2,1637.9  
 protein\_JAZ3 protein\_CLH1 I linecolor=255/0/0 pos=6534.8,1449.3  
 protein\_JAZ3 protein\_JR1 I linecolor=255/0/0 pos=6442.8,1384.5  
 protein\_JAZ3 protein\_MYC2 B linecolor=255/0/0 pos=6898.2,1426.4  
 protein\_JAZ3 protein\_THI2.1 I linecolor=255/0/0 pos=6649.3,1586.6

protein\_JAZ3 protein\_VSP1 I linecolor=255/0/0 pos=6606.4,1518.7  
protein\_JAZ4 complex\_JA-Ile\_COI1\_SCF\_JAZ P linecolor=255/0/0 pos=7031.7,2387.9  
protein\_JAZ4 complex\_JAZ\_MYC2\_TF P linecolor=255/0/0 pos=7175.2,1635.4  
protein\_JAZ4 protein\_CLH1 I linecolor=255/0/0 pos=6610.4,1453.3  
protein\_JAZ4 protein\_JR1 I linecolor=255/0/0 pos=6522.9,1375.8  
protein\_JAZ4 protein\_MYC2 B linecolor=255/0/0 pos=6967.4,1372.6  
protein\_JAZ4 protein\_THI2.1 I linecolor=255/0/0 pos=6733.7,1590.8  
protein\_JAZ4 protein\_VSP1 I linecolor=255/0/0 pos=6688.8,1527.7  
protein\_JAZ5 complex\_JA-Ile\_COI1\_SCF\_JAZ P linecolor=255/0/0 pos=7098.3,2347.2  
protein\_JAZ5 complex\_JAZ\_MYC2\_TF P linecolor=255/0/0 pos=7256.9,1628.9  
protein\_JAZ5 protein\_CLH1 I linecolor=255/0/0 pos=6692.2,1467.4  
protein\_JAZ5 protein\_JR1 I linecolor=255/0/0 pos=6595.4,1368.2  
protein\_JAZ5 protein\_MYC2 B linecolor=255/0/0 pos=7049.0,1368.2  
protein\_JAZ5 protein\_THI2.1 I linecolor=255/0/0 pos=6810.2,1584.6  
protein\_JAZ5 protein\_VSP1 I linecolor=255/0/0 pos=6767.1,1525.4  
protein\_JAZ6 complex\_JA-Ile\_COI1\_SCF\_JAZ P linecolor=255/0/0 pos=7172.7,2356.2  
protein\_JAZ6 complex\_JAZ\_MYC2\_TF P linecolor=255/0/0 pos=7339.1,1618.8  
protein\_JAZ6 protein\_CLH1 I linecolor=255/0/0 pos=6766.9,1465.2  
protein\_JAZ6 protein\_JR1 I linecolor=255/0/0 pos=6674.0,1339.4  
protein\_JAZ6 protein\_MYC2 B linecolor=255/0/0 pos=7131.2,1365.3  
protein\_JAZ6 protein\_THI2.1 I linecolor=255/0/0 pos=6891.6,1571.4  
protein\_JAZ6 protein\_VSP1 I linecolor=255/0/0 pos=6844.4,1519.7  
protein\_JAZ7 complex\_JA-Ile\_COI1\_SCF\_JAZ P linecolor=255/0/0 pos=7257.0,2338.5  
protein\_JAZ7 complex\_JAZ\_MYC2\_TF P linecolor=255/0/0 pos=7418.7,1604.0  
protein\_JAZ7 protein\_CLH1 I linecolor=255/0/0 pos=6852.4,1460.6  
protein\_JAZ7 protein\_JR1 I linecolor=255/0/0 pos=6753.2,1332.6  
protein\_JAZ7 protein\_MYC2 B linecolor=255/0/0 pos=7240.0,1340.7  
protein\_JAZ7 protein\_THI2.1 I linecolor=255/0/0 pos=6972.0,1565.9  
protein\_JAZ7 protein\_VSP1 I linecolor=255/0/0 pos=6924.6,1515.9  
protein\_JAZ8 complex\_JA-Ile\_COI1\_SCF\_JAZ P linecolor=255/0/0 pos=7341.9,2352.8  
protein\_JAZ8 complex\_JAZ\_MYC2\_TF P linecolor=255/0/0 pos=7494.6,1582.2  
protein\_JAZ8 protein\_CLH1 I linecolor=255/0/0 pos=6948.2,1465.7  
protein\_JAZ8 protein\_JR1 I linecolor=255/0/0 pos=6849.2,1324.0  
protein\_JAZ8 protein\_MYC2 B linecolor=255/0/0 pos=7324.0,1407.2  
protein\_JAZ8 protein\_THI2.1 I linecolor=255/0/0 pos=7063.1,1568.4  
protein\_JAZ8 protein\_VSP1 I linecolor=255/0/0 pos=7017.3,1519.9  
protein\_JAZ9 complex\_JA-Ile\_COI1\_SCF\_JAZ P linecolor=255/0/0 pos=7417.2,2376.6  
protein\_JAZ9 complex\_JAZ\_MYC2\_TF P linecolor=255/0/0 pos=7567.2,1577.9  
protein\_JAZ9 protein\_CLH1 I linecolor=255/0/0 pos=7030.9,1465.4  
protein\_JAZ9 protein\_JR1 I linecolor=255/0/0 pos=6929.6,1323.1  
protein\_JAZ9 protein\_MYC2 B linecolor=255/0/0 pos=7394.7,1361.7  
protein\_JAZ9 protein\_THI2.1 I linecolor=255/0/0 pos=7147.8,1565.7  
protein\_JAZ9 protein\_VSP1 I linecolor=255/0/0 pos=7102.6,1515.0  
protein\_JMT metabolite\_Me-JA A linecolor=255/0/0 pos=5097.6,2624.7  
protein\_KAT1 metabolite\_JA A linecolor=255/0/0 pos=5376.1,2326.7  
protein\_KAT1 metabolite\_OPC4 A linecolor=255/0/0 pos=5368.4,2196.1  
protein\_KAT1 metabolite\_OPC6 A linecolor=255/0/0 pos=5384.9,1970.7  
protein\_KAT2 metabolite\_JA A linecolor=255/0/0 pos=5438.6,2291.0  
protein\_KAT2 metabolite\_OPC4 A linecolor=255/0/0 pos=5446.5,2183.4  
protein\_KAT2 metabolite\_OPC6 A linecolor=255/0/0 pos=5453.0,1936.9  
protein\_KAT5 metabolite\_JA A linecolor=255/0/0 pos=5530.6,2274.3  
protein\_KAT5 metabolite\_OPC4 A linecolor=255/0/0 pos=5526.7,2193.4

protein\_KAT5 metabolite\_OPC6 A linecolor=255/0/0 pos=5531.9,1961.7  
protein\_LOX1 metabolite\_13-HPT A linecolor=255/0/0 pos=5286.8,863.6  
protein\_LOX2 metabolite\_13-HPT A linecolor=255/0/0 pos=5268.3,927.5  
protein\_LOX3 metabolite\_13-HPT A linecolor=255/0/0 pos=5323.1,964.0  
protein\_LOX4 metabolite\_13-HPT A linecolor=255/0/0 pos=5274.5,1013.8  
protein\_LOX5 metabolite\_13-HPT A linecolor=255/0/0 pos=5325.5,1057.6  
protein\_LOX6 metabolite\_13-HPT A linecolor=255/0/0 pos=5291.9,1115.9  
protein\_MAT3 metabolite\_SAM A linecolor=255/0/0 pos=2971.0,814.3  
protein\_MAT4 metabolite\_SAM A linecolor=255/0/0 pos=2958.8,889.6  
protein\_MPK3 protein\_EDS1 A linecolor=255/0/0 pos=9303.0,853.7  
protein\_MPK3 protein\_PAD3 A linecolor=255/0/0 pos=9073.4,849.0  
protein\_MPK3 protein\_PAD4 A linecolor=255/0/0 pos=9213.0,824.3  
protein\_MPK4 protein\_EDS1 I linecolor=255/0/0 pos=7753.6,962.5  
protein\_MPK4 protein\_LOX1 A linecolor=255/0/0 pos=5693.6,868.2  
protein\_MPK4 protein\_LOX2 A linecolor=255/0/0 pos=5664.2,931.8  
protein\_MPK4 protein\_LOX3 A linecolor=255/0/0 pos=5719.5,971.8  
protein\_MPK4 protein\_LOX4 A linecolor=255/0/0 pos=5664.4,1022.7  
protein\_MPK4 protein\_LOX5 A linecolor=255/0/0 pos=5725.9,1060.0  
protein\_MPK4 protein\_LOX6 A linecolor=255/0/0 pos=5695.6,1119.3  
protein\_MPK4 protein\_PAD3 I linecolor=255/0/0 pos=7550.8,966.4  
protein\_MPK4 protein\_PAD4 I linecolor=255/0/0 pos=7654.2,964.2  
protein\_MPK6 protein\_EDS1 A linecolor=255/0/0 pos=9571.4,846.3  
protein\_MPK6 protein\_PAD3 A linecolor=255/0/0 pos=9317.3,784.4  
protein\_MPK6 protein\_PAD4 A linecolor=255/0/0 pos=9425.4,855.0  
protein\_MYC2 complex\_JAZ\_MYC2\_TF P linecolor=255/0/0 pos=7342.1,794.5  
protein\_MYC2 protein\_CLH1 A linecolor=255/0/0 pos=6834.6,654.4  
protein\_MYC2 protein\_GST1 A linecolor=255/0/0 pos=5846.3,951.0  
protein\_MYC2 protein\_JAZ1 A linecolor=255/0/0 pos=6662.0,1406.9  
protein\_MYC2 protein\_JAZ10 A linecolor=255/0/0 pos=7466.0,1446.7  
protein\_MYC2 protein\_JAZ11 A linecolor=255/0/0 pos=7538.0,1449.8  
protein\_MYC2 protein\_JAZ12 A linecolor=255/0/0 pos=7619.8,1446.4  
protein\_MYC2 protein\_JAZ2 A linecolor=255/0/0 pos=6798.0,1369.6  
protein\_MYC2 protein\_JAZ3 A linecolor=255/0/0 pos=6882.9,1377.0  
protein\_MYC2 protein\_JAZ4 A linecolor=255/0/0 pos=6983.5,1420.9  
protein\_MYC2 protein\_JAZ5 A linecolor=255/0/0 pos=7063.5,1416.9  
protein\_MYC2 protein\_JAZ6 A linecolor=255/0/0 pos=7149.7,1412.2  
protein\_MYC2 protein\_JAZ7 A linecolor=255/0/0 pos=7226.2,1394.8  
protein\_MYC2 protein\_JAZ8 A linecolor=255/0/0 pos=7316.3,1347.8  
protein\_MYC2 protein\_JAZ9 A linecolor=255/0/0 pos=7397.0,1427.5  
protein\_MYC2 protein\_JR1 A linecolor=255/0/0 pos=6741.2,618.0  
protein\_MYC2 protein\_PR4 A linecolor=255/0/0 pos=5850.5,855.8  
protein\_MYC2 protein\_THI2.1 A linecolor=255/0/0 pos=6966.1,728.8  
protein\_MYC2 protein\_VSP1 A linecolor=255/0/0 pos=6914.4,672.3  
protein\_MYC2 protein\_b-CHI A linecolor=255/0/0 pinned=1 pos=5952.1385,736.0863  
protein\_X4 protein\_MYC2 A linecolor=255/0/0 pos=7366.8,631.4  
protein\_NADPH\_oxidase metabolite\_ROS A linecolor=255/0/0 pos=9916.5,873.1  
protein\_NIMIN1 complex\_NPR1\_TGA245 P linecolor=255/0/0 pos=9824.8,1499.6  
protein\_NIMIN1 protein\_NPR1 B linecolor=255/0/0 pos=9708.8,1368.8  
protein\_NIMIN1 protein\_TGA\_TF2 B linecolor=255/0/0 pos=9622.8,1739.8  
protein\_NIMIN1 protein\_TGA\_TF4 B linecolor=255/0/0 pos=9772.2,1739.7  
protein\_NIMIN1 protein\_TGA\_TF5 B linecolor=255/0/0 pos=9886.6,1725.1  
protein\_NIMIN2 complex\_NPR1\_TGA245 P linecolor=255/0/0 pos=9833.9,1582.8

protein\_NIMIN2 protein\_NPR1 B linecolor=255/0/0 pos=9753.1,1459.5  
protein\_NIMIN2 protein\_TGA\_TF2 B linecolor=255/0/0 pos=9668.6,1825.0  
protein\_NIMIN2 protein\_TGA\_TF4 B linecolor=255/0/0 pos=9780.6,1850.1  
protein\_NIMIN2 protein\_TGA\_TF5 B linecolor=255/0/0 pos=9891.0,1840.2  
protein\_NIMIN3 complex\_NPR1\_TGA245 P linecolor=255/0/0 pos=9694.6,1719.7  
protein\_NIMIN3 protein\_NPR1 B linecolor=255/0/0 pos=9612.3,1561.9  
protein\_NIMIN3 protein\_TGA\_TF2 B linecolor=255/0/0 pos=9535.6,1958.0  
protein\_NIMIN3 protein\_TGA\_TF4 B linecolor=255/0/0 pos=9650.2,1916.2  
protein\_NIMIN3 protein\_TGA\_TF5 B linecolor=255/0/0 pos=9758.2,1948.9  
protein\_NPR1 complex\_NPR1\_TGA245 P linecolor=255/0/0 pos=9887.9,1416.2  
protein\_NPR1 protein\_EDS1 I linecolor=255/0/0 pos=9685.5,1103.0  
protein\_NPR1 protein\_PAD3 I linecolor=255/0/0 pos=9482.6,1106.4  
protein\_NPR1 protein\_PAD4 I linecolor=255/0/0 pos=9587.2,1100.0  
protein\_NPR1 protein\_TGA\_TF2 B linecolor=255/0/0 pos=9755.5,1628.2  
protein\_NPR1 protein\_TGA\_TF4 B linecolor=255/0/0 pos=9848.7,1652.9  
protein\_NPR1 protein\_TGA\_TF5 B linecolor=255/0/0 pos=9955.4,1639.5  
protein\_NPR1 protein\_WRKY70 A linecolor=255/0/0 pos=9873.5,1269.4  
protein\_OPCL1 metabolite\_OPC6 A linecolor=255/0/0 pos=5100.1,1815.0  
protein\_OPR3 metabolite\_OPC8 A linecolor=255/0/0 pos=5092.2,1538.7  
protein\_PAD3 protein\_EDS5 A linecolor=255/0/0 pos=9239.0,1026.0  
protein\_PAD4 protein\_EDS5 A linecolor=255/0/0 pos=9345.7,1035.0  
protein\_PAL1 metabolite\_Trans-cinnamic\_acid A linecolor=255/0/0 pos=8329.2,1308.4  
protein\_PAL2 metabolite\_Trans-cinnamic\_acid A linecolor=255/0/0 pos=8305.4,1368.5  
protein\_PAL3 metabolite\_Trans-cinnamic\_acid A linecolor=255/0/0 pos=8315.9,1430.6  
protein\_PAL4 metabolite\_Trans-cinnamic\_acid A linecolor=255/0/0 pos=8311.8,1499.3  
protein\_Prephenate\_aminotransferase metabolite\_Phenylpyruvate A  
linecolor=255/0/0 pos=8341.1,1062.6  
protein\_RAN1 metabolite\_Copper\_cyto A linecolor=255/0/0 pos=3454.9,945.3  
protein\_SAM1 metabolite\_SAM A linecolor=255/0/0 pos=2955.3,705.2  
protein\_SAM2 metabolite\_SAM A linecolor=255/0/0 pos=2922.1,770.2  
protein\_Superoxide\_dismutase metabolite\_ROS A linecolor=255/0/0 pos=9951.3,734.1  
protein\_TGA\_TF2 complex\_NPR1\_TGA245 P linecolor=255/0/0 pos=9823.4,1790.9  
protein\_TGA\_TF4 complex\_NPR1\_TGA245 P linecolor=255/0/0 pos=9954.7,1793.2  
protein\_TGA\_TF5 complex\_NPR1\_TGA245 P linecolor=255/0/0 pos=10041.4,1765.2  
protein\_UGP\_glikosyltransferaze metabolite\_SAG A linecolor=255/0/0 pos=8968.4,2172.3  
protein\_UGP\_glikosyltransferaze metabolite\_SGE A linecolor=255/0/0 pos=8702.7,2188.5  
protein\_X3 metabolite\_SA A linecolor=255/0/0 pos=8959.0,1726.5  
protein\_X5 metabolite\_OPC-8:0-CoA A linecolor=255/0/0 pos=5101.0,1674.6  
protein\_X1 metabolite\_SA A linecolor=255/0/0 pinned=1 pos=8722.2947,1632.7280  
protein\_X2 protein\_JAZ1 A linecolor=255/0/0 pos=7003.1,1936.1  
protein\_X2 protein\_JAZ10 A linecolor=255/0/0 pos=7748.7,1917.9  
protein\_X2 protein\_JAZ11 A linecolor=255/0/0 pos=7823.8,1919.2  
protein\_X2 protein\_JAZ12 A linecolor=255/0/0 pos=7906.4,1916.7  
protein\_X2 protein\_JAZ2 A linecolor=255/0/0 pos=7109.8,1932.9  
protein\_X2 protein\_JAZ3 A linecolor=255/0/0 pos=7200.8,1934.3  
protein\_X2 protein\_JAZ4 A linecolor=255/0/0 pos=7283.4,1931.2  
protein\_X2 protein\_JAZ5 A linecolor=255/0/0 pos=7360.4,1928.2  
protein\_X2 protein\_JAZ6 A linecolor=255/0/0 pos=7438.0,1926.3  
protein\_X2 protein\_JAZ7 A linecolor=255/0/0 pos=7515.9,1922.2  
protein\_X2 protein\_JAZ8 A linecolor=255/0/0 pos=7596.7,1922.5  
protein\_X2 protein\_JAZ9 A linecolor=255/0/0 pos=7674.0,1922.6  
# \_attributes complex\_ASK1\_CULLIN1\_RBXE2 pos=3989.1318,1376.2149

```
# _attributes complex_ASK1_CULLIN1_RBXE2_EBF12 pos=4128.6972,1076.2695
# _attributes complex_Et_receptor_CTR1 pos=3600.8855,1574.1671
# _attributes complex_JA-Ile_COI1_SCF pos=5964.2300,2444.9060
# _attributes complex_JA-Ile_COI1_SCF_JAZ pos=7264.2563,2445.0452
# _attributes complex_JAZ_MYC2_TF pos=7534.7484,882.6253
# _attributes complex_NPR1_TGA245 pos=9935.2417,1554.3359
# _attributes complex_NPR1_oligomer pos=9849.4582,1167.8969
# _attributes complex_SCF pos=5970.6247,2569.1960
# _attributes metabolite_12/13_EDT pos=5191.3571,1206.2083
# _attributes metabolite_13-HPT pos=5191.7756,1063.7818
# _attributes metabolite_ACC pos=2828.7497,1166.0440
# _attributes metabolite_BA pos=8747.7404,1405.9435
# _attributes metabolite_Chorismate pos=8450.8641,719.0878
# _attributes metabolite_Copper pos=3430.0817,863.4805
# _attributes metabolite_Copper_cyto pos=3355.4583,1069.2395
# _attributes metabolite_Ethylene pos=3047.2717,1576.6065 queryset=end
# _attributes metabolite_Isochorismate pos=8982.9784,1272.4562
# _attributes metabolite_JA pos=5179.4823,2428.0947 queryset=end
# _attributes metabolite_JA-Ile pos=5521.2873,2428.6930
# _attributes metabolite_L-methionine pos=2873.7460,686.8430
# _attributes metabolite_Linolenic_acid pos=5182.3691,925.4869
# _attributes metabolite_Me-JA pos=5179.6149,2534.9874
# _attributes metabolite_OPC-8:0-CoA pos=5189.1866,1694.5881
# _attributes metabolite_OPC4 pos=5179.4073,2237.7986
# _attributes metabolite_OPC6 pos=5189.4848,1808.9652
# _attributes metabolite_OPC8 pos=5191.8678,1588.6824
# _attributes metabolite_OPDA pos=5191.7220,1456.9093
# _attributes metabolite_OPDA_chl pos=5191.6693,1343.8905
# _attributes metabolite_Orto-coumaric_acid pos=8446.6786,1517.2597
# _attributes metabolite_Phenylalanine pos=8460.3831,1229.7144
# _attributes metabolite_Phenylpyruvate pos=8457.3669,1073.7248
# _attributes metabolite_Prephenate pos=8454.0794,906.3837
# _attributes metabolite_ROS pos=9845.8537,785.7078
# _attributes metabolite_SA pos=8795.1344,1926.8119 queryset=end
# _attributes metabolite_SAG pos=9017.0205,2044.4030
# _attributes metabolite_SAM pos=2872.9329,868.0667
# _attributes metabolite_SA_chl pos=8990.9766,1499.2868
# _attributes metabolite_SGE pos=8530.6824,2079.4437
# _attributes metabolite_Trans-cinnamic_acid pos=8453.1899,1397.6368
# _attributes protein_AAO4 pos=8617.8622,1514.4798
# _attributes protein_ACO pos=2747.2642,1599.6108
# _attributes protein_ACO-like pos=2806.0253,1677.6604
# _attributes protein_ACO1 pos=2780.9366,1356.7406
# _attributes protein_ACO2 pos=2752.0127,1471.7158
# _attributes protein_ACO4 pos=2725.5358,1543.9806
# _attributes protein_ACS1 pos=2986.9718,1001.6235
# _attributes protein_ACS10 pos=2642.4807,1063.7428
# _attributes protein_ACS11 pos=2657.7706,1004.1143
# _attributes protein_ACS2 pos=2996.3478,1066.0923
# _attributes protein_ACS3 pos=3016.0225,1129.2945
# _attributes protein_ACS4 pos=3020.1319,1214.3799
# _attributes protein_ACS5 pos=3004.1981,1272.1508
```



```
# _attributes protein_ACS6 pos=3009.5820,1344.7938
# _attributes protein_ACS7 pos=2630.5141,1267.7632
# _attributes protein_ACS8 pos=2623.3126,1195.8871
# _attributes protein_ACS9 pos=2629.0142,1142.0691
# _attributes protein_ACX1 pos=4601.8755,2141.2370
# _attributes protein_ACX2 pos=4733.6969,2167.3010
# _attributes protein_ACX3 pos=4862.1138,2172.3164
# _attributes protein_ACX4 pos=4974.7792,2156.7464
# _attributes protein_ACX5 pos=5103.0422,2129.6433
# _attributes protein_ACX6 pos=5253.9508,2145.3852
# _attributes protein_AIM1 pos=5408.4666,2131.2717
# _attributes protein_AOC1 pos=5501.6783,1230.2696
# _attributes protein_AOC2 pos=5499.0343,1287.1030
# _attributes protein_AOC3 pos=5500.2334,1344.1796
# _attributes protein_AOC4 pos=5500.3281,1399.3646
# _attributes protein_AOS pos=5391.6082,1207.2061
# _attributes protein_Arogenate_dehydratase pos=8214.7251,1225.5675
# _attributes protein_BA2H pos=9263.1309,1531.9092
# _attributes protein_CLH1 pos=6507.8310,693.5341
# _attributes protein_CM1 pos=8257.8162,824.4890
# _attributes protein_CM2 pos=8257.3195,956.5955
# _attributes protein_CM3 pos=8255.0435,884.4115
# _attributes protein_COI1 pos=5663.2485,2561.6248
# _attributes protein_CTR1 pos=3571.5217,1422.0962
# _attributes protein_Catalase pos=10058.7687,804.5395
# _attributes protein_EBF1 pos=3980.0435,1025.9545
# _attributes protein_EBF2 pos=3831.2808,1247.5160
# _attributes protein_EDF1 pos=4350.2870,1766.8882
# _attributes protein_EDF2 pos=4343.7037,1835.2933
# _attributes protein_EDF3 pos=4348.7575,1895.4870
# _attributes protein_EDF4 pos=4336.1263,1964.6902
# _attributes protein_EDS1 pos=9536.7463,935.5799
# _attributes protein_EDS5 pos=9358.7139,1144.0689
# _attributes protein_EIL1 pos=3785.2377,1748.6973
# _attributes protein_EIL2 pos=3785.3721,1823.0804
# _attributes protein_EIN2 pos=3509.9641,1767.6432
# _attributes protein_EIN3 pos=3793.1832,1925.2074
# _attributes protein_EIN4 pos=3676.0680,1226.9932
# _attributes protein_EIN5 pos=3787.6727,1024.8725
# _attributes protein_ERF1 pos=4346.0997,1697.6237
# _attributes protein_ERS1 pos=3434.6794,1258.3170
# _attributes protein_ERS2 pos=3555.0632,1255.1498
# _attributes protein_ETR1 pos=3167.1975,1253.2617
# _attributes protein_ETR2 pos=3289.3931,1253.1866
# _attributes protein_GST1 pos=4480.9400,1257.0682
# _attributes protein_Glutathione_peroxidase pos=9992.3760,657.0318
# _attributes protein_HRT pos=9671.5404,561.1346
# _attributes protein_ICS1 pos=9280.3357,1268.4603
# _attributes protein_ICS2 pos=9399.1534,1264.7741
# _attributes protein_IPL pos=8861.0291,1268.6910
# _attributes protein_Inactive_BA2H pos=9444.7441,1467.1812
# _attributes protein_Inactive_EIN2 pos=3278.0744,1765.9903
```

```
# _attributes protein_Inactive_HRT pos=9896.2585,610.1956
# _attributes protein_Inactive_MPK3 pos=9067.9004,654.0412
# _attributes protein_Inactive_MPK6 pos=9506.2357,674.2092
# _attributes protein_Inactive_et_receptor pos=3602.3440,1057.5273
# _attributes protein_JAR1 pos=5521.3226,2329.8808
# _attributes protein_JAZ1 pos=6256.4561,2238.8808
# _attributes protein_JAZ10 pos=7694.3421,2225.4202
# _attributes protein_JAZ11 pos=7829.9184,2226.5545
# _attributes protein_JAZ12 pos=7973.5191,2226.5061
# _attributes protein_JAZ2 pos=6455.6875,2234.2826
# _attributes protein_JAZ3 pos=6629.2943,2237.1244
# _attributes protein_JAZ4 pos=6788.3186,2234.2685
# _attributes protein_JAZ5 pos=6931.8061,2230.9638
# _attributes protein_JAZ6 pos=7082.9100,2228.2605
# _attributes protein_JAZ7 pos=7234.7819,2225.3384
# _attributes protein_JAZ8 pos=7401.4746,2226.7295
# _attributes protein_JAZ9 pos=7552.8138,2226.9592
# _attributes protein_JMT pos=5046.6098,2534.8157
# _attributes protein_JR1 pos=6326.4442,632.3070
# _attributes protein_KAT1 pos=5543.5522,2129.6971
# _attributes protein_KAT2 pos=5669.8378,2113.6469
# _attributes protein_KAT5 pos=5810.6893,2129.1855
# _attributes protein_LOX1 pos=5421.2494,851.4789
# _attributes protein_LOX2 pos=5425.5459,919.3723
# _attributes protein_LOX3 pos=5427.3340,976.7037
# _attributes protein_LOX4 pos=5426.9799,1038.7540
# _attributes protein_LOX5 pos=5431.0935,1094.9328
# _attributes protein_LOX6 pos=5430.0830,1146.9491
# _attributes protein_MAT3 pos=3067.1080,856.9350
# _attributes protein_MAT4 pos=3066.6798,920.6361
# _attributes protein_MPK3 pos=9080.4522,780.4750
# _attributes protein_MPK4 pos=5932.5236,1014.1104
# _attributes protein_MPK6 pos=9496.6856,777.7182
# _attributes protein_MYC2 pos=7175.7350,646.0556
# _attributes protein_X4 pos=7526.2104,734.2049
# _attributes protein_NADPH_oxidase pos=10013.4012,872.5681
# _attributes protein_NIMIN1 pos=9672.4716,1469.0292
# _attributes protein_NIMIN2 pos=9704.5033,1632.6185
# _attributes protein_NIMIN3 pos=9462.4531,1847.3481
# _attributes protein_NPR1 pos=9791.2031,1292.7221
# _attributes protein_OPCL1 pos=5022.8252,1807.1028
# _attributes protein_OPR3 pos=5028.1407,1585.3193
# _attributes protein_PAD3 pos=9163.0874,906.3816
# _attributes protein_PAD4 pos=9353.2502,935.6032
# _attributes protein_PAL1 pos=8178.0466,1294.5935
# _attributes protein_PAL2 pos=8177.6689,1344.4421
# _attributes protein_PAL3 pos=8178.9558,1398.3397
# _attributes protein_PAL4 pos=8180.8528,1463.9937
# _attributes protein_PDF1.2 pos=4477.4178,1813.3307
# _attributes protein_PR1 pos=10059.8979,1212.1585
# _attributes protein_PR2 pos=10098.8321,1359.1027
# _attributes protein_PR4 pos=4499.3367,1111.4144
```

```
# _attributes protein_PR5 pos=10084.7737,1461.5024
# _attributes protein_Prephenate_aminotransferase pos=8227.7854,1068.1744
# _attributes protein_RAN1 pos=3560.7868,892.5952
# _attributes protein_SAM1 pos=3061.0114,710.8046
# _attributes protein_SAM2 pos=3067.3321,793.6800
# _attributes protein_Superoxide_dismutase pos=10040.8037,727.9978
# _attributes protein_TGA_TF2 pos=9656.9518,1977.4157
# _attributes protein_TGA_TF4 pos=9863.7647,1988.3104
# _attributes protein_TGA_TF5 pos=10029.3701,1960.1179
# _attributes protein_THI2.1 pos=6729.2071,790.8070
# _attributes protein_UGP_glikosyltransferaze pos=8878.5630,2250.5117
# _attributes protein_VSP1 pos=6645.9146,742.4696
# _attributes protein_WRKY70 pos=9940.9593,1273.4311
# _attributes protein_b-CHI pos=4508.7987,954.8370
# _attributes protein_X3 pos=9134.1762,1500.0254
# _attributes protein_X5 pos=5028.8895,1695.4792
# _attributes protein_X1 pos=8456.2843,1589.7962
# _attributes protein_X2 pos=7778.0160,1592.933
```

The visualisation of this file into a graph in an interactive way is possible with the Biomine visualizer file `bmvis.jar`. A user should have installed Java software package. It is available for download at: <http://java.com/en/download/index.jsp>. The `bmvis.jar` file can be downloaded at <http://www.cs.helsinki.fi/u/phinstan/bmvis.jar> and it should be located in the same folder as the graph file. To perform visualization of the file one should do the following:

1. Open Command Prompt window
2. Change your path into the folder where the Supporting Information S2 is located.
3. Type into the Command Prompt following: "<absolute path of java.jar file >\java" -jar `bmvis.jar` "graph.bmg". An example is a following line:

```
C:\Users\Public\Desktop>"C:\Program Files (x86)\Java\jre1.6.0_22\bin\java"
-jar bmvis.jar "graph.bmg"
```

A user should pay attention to the spaces between words in the example line above and apply them in the same way when visualising the graph. In case a warning message is displayed the user should click OK and the graph will be visualized. Note that this visualisation procedure applies only for the Windows platform.



## B PDS model structure revision

### B.1 Component vocabulary

Vocabulary of the biological components was used by Bio3graph tool to determine the triplets (*component1*, *reaction*, *component2*). In this vocabulary every row represents one component with its synonyms separated by comma. The first name in the row represents the biological component name that is also visualized in the graph nodes. The vocabulary components were as follows:

SA, Salicylic acid, Salicylate, o-Hydroxybenzoic acid

HRT, HRT protein, Hypersensitive response to TCV, HYPERSENSITIVE RESPONSE TO TCV, MWF20.19, MWF20\_19, RCY1, RECOGNITION OF PERONOSPORA PARASITICA 8, RESISTANT TO CMV(Y) 1, RPP8, DISEASE RESISTANCE PROTEIN RPP8, AT5G43470

NPR1, NPR1 gene, NPR1 protein, ARABIDOPSIS NONEXPRESSER OF PR GENES 1, ATNPR1, F15H21.6, F15H21\_6, NIM1, NON-INDUCIBLE IMMUNITY 1, NONEXPRESSER OF PR GENES 1, SAI1, SALICYLIC ACID INSENSITIVE 1, AT1G64280

NIMIN1, NIMIN1 gene, NIMIN1 protein, T6A9.23, NIM1-INTERACTING 1, NIMIN-1, At1g02450

NIMIN2, NIM1-INTERACTING 2, NIMIN-2, AT3G25882

NIMIN3, NIM1-INTERACTING 3, NIMIN-3, AT1G09415

PAD4, PAD4 gene, PAD4 protein, ARABIDOPSIS PHYTOALEXIN DEFICIENT 4, AT-PAD4, F22O6.190, PHYTOALEXIN DEFICIENT 4, At3g52430

PAD3, CYP71B15, MDJ14.12, PAD3, PHYTOALEXIN DEFICIENT 3, AT3G26830

EDS1, EDS1 gene, EDS1 protein, ATEDS1, T17F15.40, ENHANCED DISEASE SUSCEPTIBILITY 1, At3g48090

EDS5, EDS5 gene, SID1 gene, EDS5 protein, SID1 protein, SALICYLIC ACID INDUCTION DEFICIENT 1, F19H22.130, F19H22\_130, ENHANCED DISEASE SUSCEPTIBILITY 5, SID1, At4g39030

PR1, PR1 gene, PR1 protein, ATPR1, PR 1, T6B13\_15, PATHOGENESIS-RELATED PROTEIN 1, PATHOGENESIS-RELATED GENE 1, T6B13.15, At2g14610, PR-1, PR-1a

PR2, PR2 gene, PATHOGENESIS-RELATED PROTEIN 2, BGL2, BG2, BETA-1, 3-GLUCANASE 2, 3-GLUCANASE, PR-2, PR 2, F28O9.110, At3g57260

PR5, PR5 protein, PR-5, PATHOGENESIS-RELATED PROTEIN 5, PATHOGENESIS-RELATED GENE 5, PR 5, At1g75040

ICS1, ICS1 gene, SID2 gene, EDS16 gene, ICS1 protein, SID2 protein, EDS16 protein, ENHANCED DISEASE SUSCEPTIBILITY TO ERYTHROSPHERE ORONTII 16, AT1G74710.1, SALICYLIC ACID INDUCTION DEFICIENT 2, ISOCHORISMATE SYNTHASE, ISOCHORISMATE SYNTHASE 1, SID2, EDS16, F25A4.31, ICS 1, ATICS1, ICS-1, ARABIDOPSIS ISOCHORISMATE SYNTHASE 1, ENHANCED DISEASE SUSCEPTIBILITY TO ERYTHROSPHERE ORONTII 16, At1g74710

ICS2, ICS2 gene, ICS2 protein, F6A14\_3, ISOCHORISMATE SYNTHASE 2, ATICS2, F6A14.3, ICS 2, ICS-2, At1g18870

WRKY70, WRKY70 gene, WRKY70 protein, WRKY DNA-binding protein 70, ARABIDOPSIS THALIANA WRKY DNA-BINDING PROTEIN 70, ATWRKY70, At3g56400

MPK4, MPK4 gene, MPK4 protein, ATPK4, ARABIDOPSIS THALIANA MAP KINASE 4, MAP KINASE 4, F2N1.1, F2N1\_1, At4g01370

Chorismate

Isochorismate, Iso-chorismic acid, Isochorismic acid

BA, Benzoic acid, benzene carboxylic acid, benzoate, Benzeneformic acid, phenylformic acid

BA2H, benzoic acid 2-hydroxylase

NADPH oxidase

Catalase, CAT  
 Glutathione\_peroxidase, GSH peroxidase  
 Superoxide\_dismutase, SOD  
 SAG, salicylic acid glucoside, Salicylic acid beta-glucoside, salicylate glucoside, 2-O-b-D glucoside  
 SGE, Salicylic acid glucose ester, salicyloyl glucose ester  
 Trans-cinnamic\_acid, Cinnamic acid, trans-Cinnamate, trans-Cinnamic acid, (E)-Cinnamate  
 Phenylalanine, L-phenylalanine, 3-Phenylalanine, Phe  
 Phenylpyruvate, Phenylpyruvate, Phenylpyruvic acid, alpha-Ketohydrocinnamic acid, keto-Phenylpyruvate, 3-Phenyl-2-oxopropanoate  
 ROS, reactive oxygen species, oxygen radicals, oxidative burst, h<sub>2</sub>O<sub>2</sub>, hydrogen peroxide, H<sub>2</sub>O<sub>2</sub>, O<sub>3</sub>, O<sub>2</sub>, O<sub>2</sub>·<sup>-</sup>  
 PAL1, PAL1 gene, PAL1 protein, CI0004, T1J8.22, PHENYLALANINE AMMONIA LYASE 1, PHE AMMONIA LYASE 1, T1J8\_22, PAL 1, PAL-1, ATPAL1, At2g37040  
 PAL2, PAL2 gene, PAL2 protein, phenylalanine ammonia-lyase 2, T4D2.190, PAL 2, PAL-2, ATPAL2, At3g53260  
 PAL3, PAL3 gene, PAL3 protein, F21E1.150, ATPAL3, PHENYL ALANINE AMMONIA-LYASE 3, F21E1\_150, PAL 3, PAL-3, At5g04230  
 PAL4, PAL4 gene, PAL4 protein, Phenylalanine ammonia-lyase 4, PAL 4, PAL-4, F14P13.6, At3g10340  
 IPL, IPL protein, isochorismate pyruvate lyase  
 CM1, ARABIDOPSIS THALIANA CHORISMATE MUTASE 1, ATCM1, CHORISMATE MUTASE 1, MXO21.4, AT3G29200  
 CM2, ATCM2, chorismate mutase 2, At5g10870  
 CM3, ATCM3, chorismate mutase 3, At1g69370  
 Prephenate\_aminotransferase, At2G22250  
 Prephenate, Prephenic acid  
 Arogenate\_dehydratase, ADT5, arogenate dehydratase 5, prephenate dehydratase, At5g22630  
 Orto-coumaric\_acid, trans-2-Hydroxycinnamate, trans-2-Hydroxycinnamic acid, 2-Hydroxycinnamate, 2-Coumaric acid, o-Coumaric acid, 2-Coumarate  
 UGP\_glikosyltransferaze, UGT72B3, UDP-GLUCOSYL TRANSFERASE 72B3, UDP-glycosyltransferase, quercetin 3-O-glucosyltransferase, AT1G01420  
 Ethylene, Ethene, Et  
 Copper, Cu, Cu<sup>2+</sup>  
 SAM, Sadenosyl methionine, S-Adenosyl methionine, S-adenosyl-L-methionine  
 ACC, 1-aminocyclopropane-1-carboxylic-acid  
 ACS1, ACS1 protein, ARABIDOPSIS THALIANA 1-AMINOCYCLOPROPANE-1-CARBOXYLATE SYNTHASE 1, 1-AMINOCYCLOPROPANE-1-CARBOXYLATE SYNTHASE 1, AT-ACS1, F2A19.110, ACC SYNTHASE 1, At3g61510  
 ACS2, ACS2 protein, 1-Amino-cyclopropane-1-carboxylate synthase 2, 1-AMINOCYCLOPROPANE-1-CARBOXYLATE SYNTHASE, F22L4.4, AT-ACC2, EC 4.4.1.14, F22L4.4, S-ADENOSYL-L-METHIONINE METHYLTHIOADENOSINE-LYASE, At1g01480  
 ACS3, ACS3 protein, F21B23.5, F21B23\_5, 1-F21B23.5, 1-AMINOCYCLOPROPANE-1-CARBOXYLATE SYNTHASE LIKE PSEUDOGENE, At5g28360  
 ACS4, ACS4 protein, ATACS4, 1-AMINOCYCLOPROPANE-1-CARBOXYLATE SYNTHASE, T30L20.7, 1-AMINOCYCLOPROPANE-1-CARBOXYLATE SYNTHASE 4, 1-AMINOCYCLOPROPANE-1-CARBOXYLIC ACID SYNTHASE POLYPEPTIDE, ACC4, At2g22810  
 ACS5, ACS5 protein, ETO2, CYTOKININ-INSENSITIVE 5, MPA24.15, ACC SYNTHASE 5, MPA24.15, ATACS5, CIN5, 1-AMINO-1-CYCLOPROPANECARBOXYLATE SYNTHASE, ETHYLENE OVERPRODUCER 2, At5g65800

ACS6, ACS6 protein, F8L21.70, F8L21\_70, 1-AMINOCYCLOPROPANE-1-CARBOXYLIC ACID (ACC) SYNTHASE 6, ATACS6, At4g11280

ACS7, ACS7 protein, 1-AMINO-CYCLOPROPANE-1-CARBOXYLATE SYNTHASE 7, T25K17.10, T25K17\_10, ATACS7, At4g26200

ACS8, ACS8 protein, 1-AMINO-CYCLOPROPANE-1-CARBOXYLATE SYNTHASE 8, T28I19.50, T28I19\_50, At4g37770

ACS9, ACS9 protein, ETO3, ETHYLENE OVERPRODUCING 3, 1-AMINOCYCLOPROPANE-1-CARBOXYLATE SYNTHASE 9, T16K5.50, At3g49700

ACS10, ACS10 protein, F16P17\_11, F16P17.11, ACC SYNTHASE 10, At1g62960

ACS11, ACS11 protein, 1-AMINOCYCLOPROPANE-1-CARBOXYLATE SYNTHASE 11, T17A2.2, T17A2\_2, At4g08040

ACO1, ACO1 protein, F3P11\_19, F3P11.19, ACC OXIDASE 1, ATACO1, At2g19590

ACO2, ACO2 protein, ACC OXIDASE 2, F24O1.40, F24O1\_40, ATACO2, ACC OXIDASE, At1g62380

ACO4, ACO4 protein, ETHYLENE-FORMING ENZYME, T7A14.12, ethylene forming enzyme, EFE, 1-AMINOCYCLOPROPANE-1-CARBOXYLATE OXIDASE, 1-AMINOCYCLOPROPANE-1-CARBOXYLIC ACID OXIDASE, EAT1, T7A14\_12, At1g05010

ACO, F12F1.12, F12F1.12 protein, F12F1\_12, AT1G12010

ACO-like, F2P24.4 protein, F2P24\_4, AT1G77330

RAN1, RAN1 protein, RESPONSIVE-TO-ANTAGONIST 1, K23L20\_14, HMA7, K23L20.14, At5g44790

ETR1, ETR1 protein, EIN1, T27F4.9, ETHYLENE INSENSITIVE 1, HISTIDINE KINASE ETR1, ETHYLENE RESPONSE 1, T27F4\_9, ETR, At1g66340

ERS1, ERS, ETHYLENE RESPONSE SENSOR, ETHYLENE RESPONSE SENSOR 1, T20B5.14, T20B5\_14, AT2G40940

ETR2, ETHYLENE RESPONSE 2, K14B15.9, AT3G23150

ERS2, ETHYLENE RESPONSE SENSOR 2, F19P19.25, F19P19\_25, AT1G04310

EIN4, ETHYLENE INSENSITIVE 4, F7O18.5, F7O18\_5, AT3G04580

CTR1, CTR1 protein, SUGAR-INSENSITIVE 1, SERINE/THREONINE-PROTEIN KINASE CTR1, SIS1, F17C15\_150, F17C15.150, AT5G03730.1, CONSTITUTIVE TRIPLE RESPONSE 1, At5g03730

EIN2, EIN2 protein, PIR2, F12E4\_10, ERA3, ORESARA 2, ATEIN2, CYTOKININ RESISTANT 1, F12E4.10, ETHYLENE INSENSITIVE 2, ORE3, CKR1, ENHANCED RESPONSE TO ABA3, ORESARA 3, ORE2, At5g03280

EIN3, EIN3 protein, ETHYLENE-INSENSITIVE3, MOE17.4, At3g20770

EIL1, ATEIL1, ETHYLENE-INSENSITIVE3-LIKE 1, T20P8.10, T20P8\_10, AT2G27050

EIL2, ETHYLENE-INSENSITIVE3-LIKE 2, T10F18.150, T10F18\_150, AT5G21120

ERF1, ERF1 gene, ERF1 protein, ETHYLENE RESPONSE FACTOR 1, ATERF1, K14B15.4, At3g23240

EDF1, EDF1 gene, EDF1 protein, TEM1, F2J7.3, F2J7\_3, TEMPRANILLO 1, TEM1, ETHYLENE RESPONSE DNA BINDING FACTOR 1, AT1G25560

EDF2, EDF2 gene, EDF2 protein, RAV2, TEM2, ATRAV2, RAP2.8, T6L1.3, T6L1\_3, TEMPRANILLO 2, RELATED TO AP2 8, ETHYLENE RESPONSE DNA BINDING FACTOR 2, REGULATOR OF THE ATPASE OF THE VACUOLAR MEMBRANE, AT1G68840

EDF3, EDF3 gene, EDF3 protein, K13N2.14, ETHYLENE RESPONSE DNA BINDING FACTOR 3, AT3G25730

EDF4, EDF4 gene, EDF4 protein, RAV1, T6J4.2, T6J4\_2, RELATED TO ABI3/VP1 1, ETHYLENE RESPONSE DNA BINDING FACTOR 4, AT1G13260

EBF1, EBF1 gene, EBF1 protein, F13B15.15, FBL6, F13B15\_15, EIN3-BINDING F BOX PROTEIN 1, At2g25490

EBF2, EBF2 gene, EBF2 protein, EIN3-BINDING F BOX PROTEIN 2, F18G18.90, F18G18\_90, EBF1/2, At5g25350

EIN5, EIN5 protein, EXORIBONUCLEASE 4, ACC INSENSITIVE 1, F20D21.30, ETHYLENE INSENSITIVE 5, ATXRN4, AIN1, F20D21\_30, XRN4, At1g54490  
 JA, Jasmonic acid, (-)-Jasmonic acid, Jasmonate, 2-(3-oxo-2-((Z)-pent-2-enyl)cyclopentane)acetic acid  
 Linolenic\_acid  
 LOX1, LOX1 protein, Lipxygenase 1, LIPOXYGENASE, ATLOX1, LOX 1, LOX-1, ARABIDOPSIS LIPOXYGENASE 1, At1g55020  
 LOX2, LOX2 protein, ARABIDOPSIS THALIANA LIPOXYGENASE 2, ATLOX2, LIPOXYGENASE 2, LOX 2, LOX-2, T14D3.80, At3g45140  
 LOX3, LOX3 protein, F28G4.10, LOX 3, LOX-3, LIPOXYGENASE 3, At1g17420  
 LOX4, ARABIDOPSIS THALIANA LIPOXYGENASE 4, ATLOX4, LIPOXYGENASE 4, LOX4, T10D10.1, T10D10\_1, AT1G72520  
 LOX5, ARABIDOPSIS THALIANA LIPOXYGENASE 5, ATLOX5, MCB17.13, AT3G22-400  
 LOX6, ARABIDOPSIS THALIANA LIPOXYGENASE 6, ATLOX6, F12B7.11, F12B7\_11, LIPOXYGENASE 6, AT1G67560 AOS, AOS protein, CYP74A, DELAYED DEHISCENCE 2, DDE2, ALLENE OXIDE SYNTHASE, CYTOCHROME P450 74A, At5g42650  
 AOC1, AOC1 protein, EARLY-RESPONSIVE TO DEHYDRATION 12, ALLENE OXIDE CYCLASE 1, ERD12, AOC 1, AOC-1, At3g25760  
 AOC2, AOC2 protein, ALLENE OXIDE CYCLASE 2, AOC 2, AOC-2, K13N2.11, At3g25770  
 AOC3, AOC3 protein, ALLENE OXIDE CYCLASE 3, K13N2.12, AOC 3, AOC-3, At3g25-780  
 AOC4, AOC4 protein, ALLENE OXIDE CYCLASE 4, T6J4\_4, T6J4.4, AOC 4, AOC-4, At1g13280  
 13-HPT, 13-HPOT, 13(S)-hydroperoxy linolenic acid  
 12/13\_EDT, allene oxide  
 OPDA, 12-OPDA, phytodienoic acid, 12-oxophytodienoic acid  
 JMT, JMT protein, F14P1.3, JASMONIC ACID CARBOXYL METHYLTRANSFERASE, F14P1.3, At1g19640  
 JA-Ile, JA-isoleucine, Jasmonate-isoleucine, N-[(1R,2R)-3-oxo-2-(2-(Z)-pentenyl)cyclopentane-1-acetyl]-(2S,3S)-isoleucine  
 COI1, COI1 protein, T28M21\_10, T28M21.10, CORONATINE INSENSITIVE 1, At2g39940  
 JAR1, JAR1 protein, JASMONATE RESISTANT 1, FIN219, F11C10.6, FAR-RED INSENSITIVE 219, At2g46370  
 JAZ1, JAZ1 gene, JAZ1 protein, T29M8\_5, T29M8.5, TIFY10A, JASMONATE-ZIM-DOMAIN PROTEIN 1, At1g19180  
 JAZ2, F25A4.8, F25A4\_8, JASMONATE-ZIM-DOMAIN PROTEIN 2, TIFY10B, AT1G74950  
 JAZ3, JAZ3 protein, JASMONATE-ZIM-DOMAIN PROTEIN 3, JAI3, JASMONATE-INSENSITIVE 3, TIFY6B, At3g17860  
 JAZ4, JASMONATE-ZIM-DOMAIN PROTEIN 4, T1N15.11, T1N15\_11, TIFY DOMAIN PROTEIN 6A, TIFY6A, AT1G48500  
 JAZ5, F28G4.16, JASMONATE-ZIM-DOMAIN PROTEIN 5, TIFY11A, AT1G17380  
 JAZ6, JASMONATE-ZIM-DOMAIN PROTEIN 6, T10D10.8, T10D10\_8, TIFY DOMAIN PROTEIN 11B, TIFY11B, AT1G72450  
 JAZ7, JASMONATE-ZIM-DOMAIN PROTEIN 7, T31E10.6, T31E10\_6, TIFY5B, AT2G34600  
 JAZ8, JASMONATE-ZIM-DOMAIN PROTEIN 8, TIFY5A, AT1G30135  
 JAZ9, F5A18.12, F5A18\_12, JASMONATE-ZIM-DOMAIN PROTEIN 9, TIFY7, AT1G70700  
 JAZ10, JAS1, JASMONATE-ASSOCIATED 1, JASMONATE-ZIM-DOMAIN PROTEIN 10, T31B5.40, T31B5\_40, TIFY DOMAIN PROTEIN 9, TIFY9, AT5G13220



JAZ11, JASMONATE-ZIM-DOMAIN PROTEIN 11, T18D12.10, TIFY3A, AT3G43440  
 JAZ12, F22D1.70, F22D1\_70, JASMONATE-ZIM-DOMAIN PROTEIN 12, TIFY3B, AT5G20900  
 MYC2, MYC2 gene, MYC2 protein, F6N18\_4, JIN1, ZBF1, JASMONATE INSENSITIVE 1, F6N18.4, ATMYC2, JAI1, RD22BP1, MYC2 TF, At1g32640  
 GST1, ARABIDOPSIS GLUTATHIONE S-TRANSFERASE 1, ARABIDOPSIS THALIANA GLUTATHIONE S-TRANSFERASE F3, ATGST1, ATGSTF3, ATGSTF6, EARLY RESPONSIVE TO DEHYDRATION 11, ERD11, F22D16.7, F22D16\_7, GLUTATHIONE S-TRANSFERASE, GLUTATHIONE S-TRANSFERASE 1, GLUTATHIONE S-TRANSFERASE 6, GSTF6, AT1G02930  
 PR4, HEL, HEL protein, PR-4, HEVEIN-LIKE, F7O18\_21, F7O18.21, PATHOGENESIS-RELATED 4, PR 4, At3g04720  
 b-CHI, b-CHI protein, basic chitinase, PATHOGENESIS-RELATED 3, PR3, PR-3, CHI-B, T2E22.18, ARABIDOPSIS THALIANA BASIC CHITINASE, PR 3, ATHCHIB, At3g12500  
 PDF1.2, PDF1.2 protein, LCR77, PDF1.2a, PLANT DEFENSIN 1.2, MFC16.8, MFC16\_8, PLANT DEFENSIN 1.2A, Low-molecular-weight cysteine-rich 77, At5g44420  
 MPK6, MPK6 protein, F18O19.10, MAP KINASE 6, ATMPK6, ARABIDOPSIS THALIANA MAP KINASE 6, MAPK6, At2g43790  
 MPK3, MPK3 protein, ATMPK3, T6D9.4, ATMAPK3, MITOGEN-ACTIVATED PROTEIN KINASE 3, AT3G45640  
 THI2.1, THI2.1 protein, THIONIN 2.1, THI2.1.1, At1g72260  
 JR1, JASMONATE RESPONSIVE 1, T2O4.6, AT3G16470  
 VSP1, ATVSP1, T4C12.50, VEGETATIVE STORAGE PROTEIN 1, AT5G24780  
 CLH1, ATCLH1, ARABIDOPSIS THALIANA CORONATINE-INDUCED PROTEIN 1, ATHCOR1, COR11, CORONATINE-INDUCED PROTEIN 1, F6F9.28, F6F9\_28, AT1G19670  
 TGA\_TF2, TGA TF2 protein, AHBP-1B, MOJ9.12, MOJ9\_12, At5g06950, TGA2  
 TGA\_TF4, TGA TF4 protein, OBF4, OCS BINDING ELEMENT 4, T31P16.20, T31P16\_20, OCS ELEMENT BINDING FACTOR 4, OCTOPINE SYNTHASE ELEMENT-BINDING PROTEIN, TGACG MOTIF-BINDING FACTOR 4, At5g10030, TGA4  
 TGA\_TF5, TGA TF5 protein, OCS-ELEMENT BINDING FACTOR 5, TGA5, MOJ9.13, MOJ9\_13, TGACG MOTIF-BINDING FACTOR 5, OBF5, At5g06960  
 OPCL1, F5M15.17, F5M15\_17, OPC-8:0 COA LIGASE1, AT1G20510  
 ACX1, ATACX1, DL4405C, FCAALL.119, ACYL-COA OXIDASE 1, ACYL-COA OXIDASE, acyl-CoA synthetase, AT4G16760  
 ACX2, ACYL-COA OXIDASE 2, ATACX2, AT5G65110  
 ACX3, ACYL-COA OXIDASE 3, ATACX3, T2D23.1, AT1G06290  
 ACX4, ACYL-COA OXIDASE 4, ATEM1.9, ATG6, ATSCX, AT3G51840  
 ACX5, ACYL-COA OXIDASE 5, T20F21.12, T20F21\_12, AT2G35690  
 ACX6, ACYL-COA OXIDASE 6, T2D23.2, T2D23\_2, AT1G06310  
 AIM1, F19B15.40, F19B15\_40, ABNORMAL INFLORESCENCE MERISTEM, AT4G29010  
 KAT2, PED1, PKT3, PEROXISOME DEFECTIVE 1, 3-KETOACYL-COA THIOLASE 2, PEROXISOMAL 3-KETOACYL-COA THIOLASE 3, AT2G33150  
 KAT1, 3-KETO-ACYL-COA THIOLASE 1, KAT1, PEROXISOMAL 3-KETOACYL-COA THIOLASE 4, PKT4, T1G11.4, T1G11\_4, AT1G04710  
 KAT5, 3-KETO-ACYL-COENZYME A THIOLASE 5, K24G6.22, K24G6\_22, PEROXISOMAL 3-KETO-ACYL-COA THIOLASE 2, PEROXISOMAL-3-KETO-ACYL-COA THIOLASE 1, PKT1, PKT2, AT5G48880  
 OPR3, DDE1, ATOPR3, F5K7.19, F5K7\_19, DELAYED DEHISCENCE 1, OXOPHYTODIE-NOATE-REDUCTASE 3, AT2G06050

OPC8, 3-oxo-2-(cis-2'-pentenyl)-cyclopentane-1-octanoate, 3-oxo-2-(cis-2'-pentenyl)-cyclopentane-1-octanoic acid, oxopentenyl-cyclopentane-octanoic acid, 8-[(1R,2R)-3-oxo-2-{(Z)-pent-2-enyl}-cyclopentyl]octanoate  
 Me-JA, MeJa, methyl jasmonate, MJ  
 OPC6, 3-oxo-2-(2-(Z)-pentenyl)cyclopentane-1-hexanoic acid, 3-oxo-2-(2-(Z)-pentenyl)cyclopentane-1-hexanoat  
 OPC4, 3-oxo-2-(2'-[Z]-pentenyl)cyclopentane-1-butanoat, 3-oxo-2-(2'-[Z]-pentenyl)cyclopentane-1-butanoic acid, 3-oxo-2-(2-(Z)-pentenyl)cyclopentane-1-butyric acid  
 OPC-8:0-CoA, oxopentenyl-cyclopentane-octanoyl-CoA, 8-[(1R,2R)-3-oxo-2-{(Z)-pent-2-enyl}-cyclopentyl]octanoyl-CoA, OPC-8:0-CoA, 3-oxo-2-(cis-2'-pentenyl)-cyclopentane-1-octanoyl-CoA  
 SAM1, ATSAM1, MAT1, METK1, S-ADENOSYLMETHIONINE SYNTHETASE 1, S-ADENOSYLMETHIONINE SYNTHETASE-1, SAM-1, SAM1, T14P4.17, T14P4\_17, AT1G02500  
 SAM2, ATSAM2, MAT2, S-ADENOSYLMETHIONINE SYNTHETASE 2, SAM-2, SAM2, T7B11.11, T7B11\_11, AT4G01850  
 MAT3, METHIONINE ADENOSYLTRANSFERASE 3, T1J8.6, T1J8\_6, AT2G36880  
 MAT4, METHIONINE ADENOSYLTRANSFERASE 4, METHIONINE OVER-ACCUMULATOR 3, MGD8.26, MTO3, S-ADENOSYLMETHIONINE SYNTHETASE 3, SAMS3, AT3G17390  
 SCF, Skp1-Cullin-F-box  
 AAO4, arabidopsis aldehyde oxidase 4, AT1g04580  
 L-methionine

## B.2 Reactions vocabulary

Vocabulary of the biological reactions was used by Bio3graph tool together with the component vocabulary to extract triples (*component1*, *reaction*, *component2*). The reaction vocabulary contains in total 6 files with synonyms for three types of reactions: activation, binding and inhibition in both active and passive form.

The synonyms used for the activation type of reaction in the active form were as follows: activate, activates, capable of activating, able to activate, elevate, elevates, increase, increases, influence accumulation, influences accumulation, influence the accumulation, influences the accumulation, induct, inducts enhance, enhances, increase, increases, accumulate, accumulates, induce, induces, catalyze, catalyzes, generate, generates, is a positive regulator, is positive regulator, encode, encodes, code for, codes for, derepress activity, derepresses activity, derepress the activity, derepresses the activity, positively signal, positively signals, signal positively, signals positively, positively regulate, positively regulates, regulate positively, regulates positively, stimulate, stimulates, trigger, triggers, facilitate, facilitates, elevate, elevates, initiate activation, initiates activation, initiate the activation, initiates the activation, produce, produces, lead to the formation, leads to the formation, lead to formation, leads to formation, leading to the formation, leading to formation, lead to the creation, leads to the creation, lead to creation, leads to creation, leading to the creation, leading to creation, lead to accumulation, lead to the accumulation, lead to an accumulation, leads to accumulation, leads to the accumulation, leads to an accumulation, lead to increase, leads to increase, lead to an increase, leads to an increase, lead to the increase, leads to the increase, result in activation, results in activation, result in the activation, results in the activation, resulting in activation, resulting in the activation result in accumulation, results in accumulation, result in the accumulation, results in the accumulation, resulting in accumulation, resulting in the accumulation, involved in activation, involved in the activation, involved in accumulation, involved in the accumulation, involved in induction, involved in the induction, involved in production, involved in the production, involved in elevation, in-

volved in the elevation, involved in synthesis, involved in the synthesis, is a precursor, is precursor, is the precursor, transduce, transduces, synthesize, synthesizes, phosphorylate, phosphorylates, oxidize, oxidizes, oxidise, oxidises, upregulate, upregulates, convert, converts, converted to, converted into, stabilize, stabilizes, stabilise, stabilises, is an enzyme involved in the biosynthetic pathway of, turn on the transcription factor, turns on the transcription factor, undergo conversion, undergoes conversion, is one of the coactivators of, is involved in the amplification loop of, is considered to be positive regulator, are considered to be positive regulators, has been identified as source of, has been identified as a source of, has been implicated as generator, has been implicated as a generator, doubled the concentration of result in synergistic activation, results in synergistic activation, resulted in synergistic activation.

The synonyms used for the activation type of reaction in the passive form were as follows: activated, elevated, inducted, enhanced, increased, accumulated, induced, catalyzed, generated, encoded, positively signalled, positively signaled, signaled positively, signalled positively, positively regulated, regulated positively, positively controlled, controlled positively, stimulated, triggered, facilitated, elevated, produced, resulted in activation, resulted in the activation, resulted in synergistic activation, resulted in the synergistic activation, resulted in accumulation, resulted in the accumulation, made by hydrolysis, transduced, synthesized by, synthesized from, phosphorylated, phosphorylated, oxidised, oxidized, oxygenated, upregulated, stabilized, stabilised, led to increase, led to an increase, led to the increase, initiated accumulation, initiated an accumulation, initiated the accumulation, translocated to the nucleus upon activation by, inducible by.

The synonyms used for the inhibition type of reaction in the active form were as follows: inhibit, inhibits, repress, represses, suppress, suppresses, is suppressor of, is a suppressor of, is the suppressor of, block, blocks, is blocker of, is a blocker of, is the blocker of, reduce, reduces, negatively signal, negatively signals, prevent, prevents, regulate negatively, regulates negatively, negatively regulate, negatively regulates, negatively control, negatively controls, control negatively, controls negatively, is a negative regulator, is negative regulator, is the negative regulator, act as a negative regulator, act as negative regulator, act as the negative regulator, inactivate, inactivates, abolish, abolishes, attenuate, attenuates, decrease, decreases, eliminate, eliminates, antagonize, antagonizes, attenuate, attenuates, alleviate, alleviates, diminish, diminishes, exert an antagonistic effect, exerts an antagonistic effect, exert antagonistic effect, exerts antagonistic effect, exert the antagonistic effect, exerts the antagonistic effect, involved in inhibition, involved in the inhibition, involved in suppression, involved in the suppression, involved in repression, involved in the repression, involved in reduction, involved in the reduction, involved in inactivation, involved in the inactivation, involved in attenuation, involved in the attenuation, prevent accumulation, prevents accumulation, prevent the accumulation, prevents the accumulation, preventing accumulation, preventing the accumulation, abrogate, abrogates, deplete, depletes, downregulate, downregulates, has an inhibitory effect, have an inhibitory effect, play an important role in the suppression, plays an important role in the suppression.

The synonyms used for the inhibition type of reaction in the passive form were as follows: inhibited, repressed, suppressed, blocked, reduced, negatively signalled, negatively signalled, negatively regulated, regulated negatively, negatively controlled, controlled negatively, prevented, inactivated, abolished, attenuated, decreased, eliminated, antagonized, attenuated, alleviated, diminished, abrogated, depleted, downregulated.

The synonyms used for the binding type of reaction in the active form were as follows: bind, binds, form complex, forms complex, form the complex, forms the complex, promote

the formation of complex, promotes the formation of complex, promote the formation of the complex, promotes the formation of the complex, promote formation of the complex, promotes formation of the complex, promote formation of complex, promotes formation of complex, make complex, makes complex, make the complex, makes the complex, compose complex, composes complex, compose the complex, composes the complex, create complex, creates complex, create the complex, creates the complex, perceive, perceives.

The synonyms used for the binding type of reaction in the passive form were as follows: binded, made complex, made the complex, composed complex, composed the complex, created complex, created the complex, formed complex, formed the complex, perceived by.

### B.3 Graph of new triplets extracted by Bio3graph

The network consists of 63 components and 137 reactions and represents the new triplets compared to the initial manually constructed PDS model structure. These new direct links are blue coloured and new indirect green coloured. The instructions for visualising the graph with the Biomine visualisation engine are provided at A.3. The graph, which should be saved as a .bmg file, contains the following reactions and components:

```
metabolite_Me-JA metabolite_JA A linecolor=0/255/0
metabolite_Ethylene protein_ACS6 A linecolor=0/255/0
metabolite_Ethylene metabolite_SA A linecolor=0/255/0
metabolite_SA protein_PAD4 A linecolor=0/255/0
protein_JAR1 protein_PR1 I linecolor=0/255/0
protein_EIN2 protein_PR1 I linecolor=0/255/0
metabolite_SA protein_EDS5 A linecolor=0/255/0
metabolite_Me-JA protein_AOS A linecolor=0/255/0
metabolite_SA protein_ACO2 I linecolor=0/255/0
metabolite_ROS metabolite_Ethylene A linecolor=0/255/0
metabolite_ROS metabolite_JA A linecolor=0/255/0
metabolite_SA protein_Catalase I linecolor=0/255/0
metabolite_JA protein_EBF1 I linecolor=0/255/0
metabolite_ACC metabolite_ROS I linecolor=0/255/0
metabolite_JA protein_ACX1 A linecolor=0/255/0
metabolite_JA protein_PR4 A linecolor=0/255/0
metabolite_Me-JA metabolite_SA A linecolor=0/255/0
metabolite_JA metabolite_SA I linecolor=0/255/0
metabolite_JA protein_AOS A linecolor=0/255/0
metabolite_SA protein_AOS I linecolor=0/255/0
metabolite_Ethylene protein_AOS A linecolor=0/255/0
metabolite_SA metabolite_ACC I linecolor=0/255/0
protein_Catalase metabolite_ROS I linecolor=0/255/0
metabolite_Ethylene protein_ERS1 A linecolor=0/255/0
metabolite_Ethylene protein_ETR2 A linecolor=0/255/0
metabolite_Ethylene protein_ERS2 A linecolor=0/255/0
protein_EIL1 metabolite_SA I linecolor=0/255/0
protein_EIN3 metabolite_SA I linecolor=0/255/0
metabolite_JA protein_LOX2 A linecolor=0/255/0
protein_Superoxide_dismutase metabolite_ROS I linecolor=0/255/0
metabolite_JA protein_LOX1 A linecolor=0/255/0
metabolite_ROS protein_PR2 A linecolor=0/255/0
metabolite_ROS protein_PR1 A linecolor=0/255/0
```

metabolite\_SA protein\_EDS1 A linecolor=0/255/0  
metabolite\_Me-JA protein\_LOX2 A linecolor=0/255/0  
metabolite\_Me-JA protein\_OPR3 A linecolor=0/255/0  
protein\_Glutathione\_peroxidase metabolite\_ROS I linecolor=0/255/0  
metabolite\_JA protein\_MPK6 A linecolor=0/255/0  
metabolite\_Ethylene protein\_JR1 I linecolor=0/255/0  
metabolite\_SA metabolite\_JA A linecolor=0/255/0  
metabolite\_SA protein\_PDF1.2 A linecolor=0/255/0  
metabolite\_SA protein\_PR1 I linecolor=0/255/0  
protein\_EDS1 metabolite\_Ethylene A linecolor=0/255/0  
protein\_PAD4 metabolite\_Ethylene A linecolor=0/255/0  
protein\_EDS1 metabolite\_ACC A linecolor=0/255/0  
protein\_PAD4 metabolite\_ACC A linecolor=0/255/0  
metabolite\_Ethylene protein\_PR2 A linecolor=0/255/0  
metabolite\_ROS protein\_Superoxide\_dismutase I linecolor=0/255/0  
metabolite\_Me-JA protein\_LOX1 A linecolor=0/255/0  
metabolite\_SA protein\_LOX1 A linecolor=0/255/0  
metabolite\_Me-JA protein\_VSP1 A linecolor=0/255/0  
metabolite\_JA protein\_CLH1 A linecolor=0/255/0  
metabolite\_Ethylene metabolite\_SA I linecolor=0/255/0  
metabolite\_Ethylene protein\_NADPH\_oxidase A linecolor=0/255/0  
metabolite\_Ethylene protein\_ACO2 A linecolor=0/255/0  
metabolite\_Me-JA metabolite\_ROS A linecolor=0/255/0  
metabolite\_Me-JA protein\_NADPH\_oxidase A linecolor=0/255/0  
metabolite\_Me-JA protein\_ETR1 A linecolor=0/255/0  
metabolite\_ACC protein\_NPR1 A linecolor=0/255/0  
metabolite\_ACC protein\_PR1 A linecolor=0/255/0  
metabolite\_SA protein\_PDF1.2 I linecolor=0/255/0  
metabolite\_Me-JA protein\_PR1 A linecolor=0/255/0  
metabolite\_Me-JA protein\_PDF1.2 I linecolor=0/255/0  
metabolite\_Ethylene protein\_PR1 A linecolor=0/255/0  
metabolite\_Ethylene protein\_NPR1 A linecolor=0/255/0  
metabolite\_SA protein\_LOX2 I linecolor=0/255/0  
metabolite\_SA protein\_AOC2 I linecolor=0/255/0  
metabolite\_SA protein\_OPR3 I linecolor=0/255/0  
metabolite\_JA protein\_ERF1 A linecolor=0/255/0  
metabolite\_SAG protein\_PR1 A linecolor=0/255/0  
metabolite\_JA protein\_b-CHI A linecolor=0/255/0  
metabolite\_Ethylene protein\_JMT I linecolor=0/255/0  
metabolite\_JA protein\_PR5 A linecolor=0/255/0  
metabolite\_Ethylene protein\_PR5 A linecolor=0/255/0  
metabolite\_SA protein\_JAZ3 I linecolor=0/255/0  
metabolite\_ROS protein\_GST1 A linecolor=0/255/0  
metabolite\_JA protein\_PR1 A linecolor=0/255/0  
metabolite\_SA protein\_THI2.1 A linecolor=0/255/0  
metabolite\_JA metabolite\_ROS A linecolor=0/255/0  
metabolite\_JA protein\_PR1 I linecolor=0/255/0  
metabolite\_ROS metabolite\_SAG A linecolor=0/255/0  
metabolite\_OPDA protein\_AOS A linecolor=0/255/0  
metabolite\_SA protein\_AOS A linecolor=0/255/0  
metabolite\_JA protein\_OPR3 A linecolor=0/255/0  
metabolite\_Me-JA metabolite\_SAG A linecolor=0/255/0

metabolite\_Me-JA protein\_WRKY70 I linecolor=0/255/0  
 protein\_EDS1 protein\_HRT A linecolor=0/255/0  
 protein\_PAD4 protein\_HRT A linecolor=0/255/0  
 protein\_ETR1 metabolite\_ROS A linecolor=0/255/0  
 metabolite\_Ethylene protein\_ETR1 A linecolor=0/255/0  
 protein\_ETR1 protein\_ERS1 I linecolor=0/255/0  
 metabolite\_Me-JA metabolite\_ACC I linecolor=0/255/0  
 metabolite\_Me-JA protein\_AOS I linecolor=0/255/0  
 protein\_EIN2 metabolite\_SA A linecolor=0/255/0  
 metabolite\_JA protein\_PR2 A linecolor=0/255/0  
 metabolite\_Ethylene metabolite\_ROS I linecolor=0/255/0  
 metabolite\_Ethylene protein\_ACO1 A linecolor=0/255/0  
 metabolite\_ROS protein\_ERF1 A linecolor=0/255/0  
 metabolite\_Ethylene protein\_ACO A linecolor=0/255/0  
 metabolite\_BA metabolite\_Ethylene A linecolor=0/255/0  
 metabolite\_ROS protein\_MPK4 A linecolor=0/255/0  
 metabolite\_SA protein\_HRT A linecolor=0/255/0  
 metabolite\_Me-JA protein\_PR5 A linecolor=0/255/0  
 protein\_NPR1 protein\_PAD4 A linecolor=0/255/0  
 metabolite\_Copper protein\_PR1 A linecolor=0/255/0  
 metabolite\_OPDA protein\_GST1 A linecolor=0/255/0  
 metabolite\_Me-JA metabolite\_ROS I linecolor=0/255/0  
 metabolite\_Me-JA protein\_PR1 I linecolor=0/255/0  
 metabolite\_Me-JA protein\_PR4 A linecolor=0/255/0  
 metabolite\_Ethylene protein\_MYC2 I linecolor=0/255/0  
 protein\_EDS5 protein\_PR1 I linecolor=0/255/0  
 metabolite\_SA protein\_b-CHI A linecolor=0/255/0  
 metabolite\_SA protein\_OPR3 A linecolor=0/255/0  
 metabolite\_SA metabolite\_OPDA A linecolor=0/255/0  
 metabolite\_JA protein\_WRKY70 I linecolor=0/255/0  
 metabolite\_ACC protein\_MPK6 A linecolor=0/255/0  
 metabolite\_ACC protein\_ERS2 A linecolor=0/255/0  
 metabolite\_ACC protein\_ETR2 A linecolor=0/255/0  
 metabolite\_JA protein\_KAT5 A linecolor=0/255/0  
 protein\_COI1 protein\_KAT5 A linecolor=0/255/0  
 protein\_WRKY70 metabolite\_SA I linecolor=0/255/0  
 metabolite\_Me-JA protein\_ERF1 A linecolor=0/255/0  
 metabolite\_SA protein\_ERF1 A linecolor=0/255/0  
 protein\_EIL1 protein\_ICS1 I linecolor=0/0/255  
 protein\_EIN3 protein\_ICS1 I linecolor=0/0/255  
 protein\_WRKY70 protein\_ICS1 I linecolor=0/0/255  
 protein\_MPK6 protein\_EIN3 A linecolor=0/0/255  
 protein\_MPK3 protein\_EIN3 A linecolor=0/0/255  
 protein\_WRKY70 protein\_PDF1.2 I linecolor=0/0/255  
 metabolite\_SAG metabolite\_SA A linecolor=0/0/255  
 protein\_TGA\_TF2 protein\_PDF1.2 B linecolor=0/0/255  
 metabolite\_ROS protein\_ACO1 A linecolor=0/0/255  
 metabolite\_Copper protein\_AAO4 I linecolor=0/0/255  
 metabolite\_Copper metabolite\_ROS A linecolor=0/0/255  
 protein\_EIN3 protein\_ICS1 B linecolor=0/0/255  
 protein\_NIMIN1 protein\_NPR1 I linecolor=0/0/255  
 protein\_TGA\_TF2 protein\_PR1 I linecolor=0/0/255

```
#_attributes metabolite_Ethylene queryset=end
#_attributes protein_ACS6
#_attributes metabolite_SA queryset=end
#_attributes protein_PAD4
#_attributes protein_JAR1
#_attributes protein_PR1
#_attributes protein_EIN2
#_attributes protein_EDS5
#_attributes metabolite_Me-JA
#_attributes protein_AOS
#_attributes protein_ACO2
#_attributes metabolite_ROS
#_attributes metabolite_JA queryset=end
#_attributes protein_Catalase
#_attributes protein_EBF1
#_attributes metabolite_ACC
#_attributes protein_ACX1
#_attributes protein_PR4
#_attributes protein_ERS1
#_attributes protein_ETR2
#_attributes protein_ERS2
#_attributes protein_EIL1
#_attributes protein_EIN3
#_attributes protein_LOX2
#_attributes protein_Superoxide_dismutase
#_attributes protein_LOX1
#_attributes protein_PR2
#_attributes protein_EDS1
#_attributes protein_OPR3
#_attributes protein_Glutathione_peroxidase
#_attributes protein_MPK6
#_attributes protein_JR1
#_attributes protein_PDF1.2
#_attributes protein_VSP1
#_attributes protein_CLH1
#_attributes protein_NADPH_oxidase
#_attributes protein_ETR1
#_attributes protein_NPR1
#_attributes protein_AOC2
#_attributes protein_ERF1
#_attributes metabolite_SAG
#_attributes protein_b-CHI
#_attributes protein_JMT
#_attributes protein_PR5
#_attributes protein_JAZ3
#_attributes protein_GST1
#_attributes protein_THI2.1
#_attributes metabolite_OPDA
#_attributes protein_WRKY70
#_attributes protein_HRT
#_attributes protein_ACO1
#_attributes protein_ACO
```

```
#_attributes metabolite_BA
#_attributes protein_MPK4
#_attributes metabolite_Copper
#_attributes protein_MYC2
#_attributes protein_KAT5
#_attributes protein_COI1
#_attributes protein_ICS1
#_attributes protein_MPK3
#_attributes protein_TGA_TF2
#_attributes metabolite_SA
#_attributes protein_AAO4
```

## B.4 Merged PDS model structure

The merged PDS model structure contains both manual and new triplets from the literature. The arcs from the manual model structure are all direct and are coloured in red. The intersection between the manual model arcs and the correct triplets extracted from the literature is represented with black coloured arcs. Further, from the correct new triplets, the indirect connections are represented with green and the direct ones with blue colour. The instructions for visualising the graph with the Biomine visualisation engine are provided at A.3. The graph, which should be saved as a .bmg file, contains the following reactions and components:

```
# _canvas -303.0,-79.0,10401.8,2711.2
# _symmetric B
complex_ASK1_CULLIN1_RBXE2 complex_ASK1_CULLIN1_RBXE2_EBF12 P
linecolor=255/0/0 pos=4085.3,1209.1
complex_ASK1_CULLIN1_RBXE2 protein_EBF1 B linecolor=255/0/0 pos=3999.9,1190.1
complex_ASK1_CULLIN1_RBXE2 protein_EBF2 B linecolor=255/0/0 pos=3928.8,1269.4
complex_ASK1_CULLIN1_RBXE2_EBF12 protein_EIL1 I
linecolor=255/0/0 pos=3943.5,1403.5
complex_ASK1_CULLIN1_RBXE2_EBF12 protein_EIL2 I
linecolor=255/0/0 pos=3980.6,1459.9
complex_ASK1_CULLIN1_RBXE2_EBF12 protein_EIN3 I
linecolor=255/0/0 pos=3966.3,1533.1
complex_Et_receptor_CTR1 protein_EIN2 I linecolor=255/0/0 pos=3528.2,1677.7
complex_JA-Ile_COI1_SCF complex_JA-Ile_COI1_SCF_JAZ P
linecolor=255/0/0 pos=6607.6,2489.8
complex_JA-Ile_COI1_SCF protein_JAZ1 B linecolor=255/0/0 pos=6109.2,2346.7
complex_JA-Ile_COI1_SCF protein_JAZ10 B linecolor=255/0/0 pos=6822.9,2338.1
complex_JA-Ile_COI1_SCF protein_JAZ11 B linecolor=255/0/0 pos=6898.0,2344.4
complex_JA-Ile_COI1_SCF protein_JAZ12 B linecolor=255/0/0 pos=6979.3,2349.6
complex_JA-Ile_COI1_SCF protein_JAZ2 B linecolor=255/0/0 pos=6209.0,2355.0
complex_JA-Ile_COI1_SCF protein_JAZ3 B linecolor=255/0/0 pos=6296.5,2366.7
complex_JA-Ile_COI1_SCF protein_JAZ4 B linecolor=255/0/0 pos=6373.6,2350.6
complex_JA-Ile_COI1_SCF protein_JAZ5 B linecolor=255/0/0 pos=6446.5,2367.9
complex_JA-Ile_COI1_SCF protein_JAZ6 B linecolor=255/0/0 pos=6519.8,2351.1
complex_JA-Ile_COI1_SCF protein_JAZ7 B linecolor=255/0/0 pos=6594.4,2352.7
complex_JA-Ile_COI1_SCF protein_JAZ8 B linecolor=255/0/0 pos=6671.3,2356.1
complex_JA-Ile_COI1_SCF protein_JAZ9 B linecolor=255/0/0 pos=6746.9,2331.0
complex_NPR1_TGA245 protein_PR1 A linecolor=255/0/0 pos=10017.9,1381.9
complex_NPR1_TGA245 protein_PR2 A linecolor=255/0/0 pos=10040.5,1501.8
```



complex\_NPR1\_TGA245 protein\_PR5 A linecolor=255/0/0 pos=10090.5,1564.9  
complex\_NPR1\_oligomer protein\_NPR1 A linecolor=255/0/0 pos=9862.0,1229.8  
complex\_SCF complex\_JA-Ile\_COI1\_SCF P linecolor=255/0/0 pos=6037.1,2541.3  
complex\_SCF metabolite\_JA-Ile B linecolor=255/0/0 pos=5748.3,2540.5  
complex\_SCF protein\_COI1 B linecolor=255/0/0 pos=5822.4,2632.3  
metabolite\_12/13\_EDT metabolite\_OPDA\_chl A linecolor=255/0/0 pos=5110.8,1260.5  
metabolite\_13-HPT metabolite\_12/13\_EDT A linecolor=0/0/0 pos=5131.0,1118.5  
metabolite\_ACC metabolite\_Ethylene A linecolor=0/0/0 pos=2922.8,1381.6  
metabolite\_ACC metabolite\_ROS I linecolor=0/255/0 pos=6348.5,943.0  
metabolite\_ACC protein\_ERS2 A linecolor=0/255/0 pos=3182.1,1183.4  
metabolite\_ACC protein\_ETR2 A linecolor=0/255/0 pos=3081.6,1190.4  
metabolite\_ACC protein\_MPK6 A linecolor=0/255/0 pos=6145.3,937.0  
metabolite\_ACC protein\_NPR1 A linecolor=0/255/0 pos=6293.7,1197.3  
metabolite\_ACC protein\_PR1 A linecolor=0/255/0 pos=6429.5,1191.1  
metabolite\_BA metabolite\_Ethylene A linecolor=0/255/0 pos=5902.0,1475.3  
metabolite\_BA metabolite\_SA A linecolor=0/0/0 pos=8776.4,1691.6  
metabolite\_Chorismate metabolite\_Isochorismate A linecolor=0/0/0 pos=8725.6,980.3  
metabolite\_Chorismate metabolite\_Prephenate A  
linecolor=255/0/0 pinned=1 pos=8456.1045,811.2770  
metabolite\_Copper metabolite\_Copper\_cyto A linecolor=255/0/0 pos=3354.4,915.5  
metabolite\_Copper metabolite\_ROS A linecolor=0/0/255 pos=6623.9,820.4  
metabolite\_Copper protein\_AAO4 I linecolor=0/0/255 pos=6010.6,1185.5  
metabolite\_Copper protein\_PR1 A linecolor=0/255/0 pos=6748.4,1008.4  
metabolite\_Copper\_cyto protein\_EIN4 A linecolor=255/0/0 pos=3525.4,1077.4  
metabolite\_Copper\_cyto protein\_ERS1 A linecolor=255/0/0 pos=3359.2,1183.4  
metabolite\_Copper\_cyto protein\_ERS2 A linecolor=255/0/0 pos=3453.0,1158.2  
metabolite\_Copper\_cyto protein\_ETR1 A linecolor=255/0/0 pos=3221.3,1094.4  
metabolite\_Copper\_cyto protein\_ETR2 A linecolor=255/0/0 pos=3283.3,1129.1  
metabolite\_Ethylene complex\_Et\_receptor\_CTR1 I linecolor=255/0/0 pos=3310.2,1609.9  
metabolite\_Ethylene metabolite\_ROS I linecolor=0/255/0 pos=6408.0,1131.3  
metabolite\_Ethylene metabolite\_SA A linecolor=0/255/0 pos=5949.5,1754.2  
metabolite\_Ethylene metabolite\_SA I linecolor=0/255/0 pos=5895.6,1775.6  
metabolite\_Ethylene protein\_ACO A linecolor=0/255/0 pos=2922.7,1635.4  
metabolite\_Ethylene protein\_ACO1 A linecolor=0/255/0 pos=2933.3,1472.9  
metabolite\_Ethylene protein\_ACO2 A linecolor=0/255/0 pos=2862.8,1543.5  
metabolite\_Ethylene protein\_ACS6 A linecolor=0/255/0 pos=3027.2,1483.7  
metabolite\_Ethylene protein\_AOS A linecolor=0/255/0 pos=4273.3,1372.2  
metabolite\_Ethylene protein\_ERS1 A linecolor=0/255/0 pos=3229.3,1468.0  
metabolite\_Ethylene protein\_ERS2 A linecolor=0/255/0 pos=3286.6,1430.6  
metabolite\_Ethylene protein\_ETR1 A linecolor=0/255/0 pos=3096.5,1424.7  
metabolite\_Ethylene protein\_ETR2 A linecolor=0/255/0 pos=3165.8,1425.2  
metabolite\_Ethylene protein\_JMT I linecolor=0/255/0 pos=4027.3,2114.4  
metabolite\_Ethylene protein\_JR1 I linecolor=0/255/0 pos=4693.1,1092.9  
metabolite\_Ethylene protein\_MYC2 I linecolor=0/255/0 pos=5064.3,1092.1  
metabolite\_Ethylene protein\_NADPH\_oxidase A linecolor=0/255/0 pos=6539.8,1185.7  
metabolite\_Ethylene protein\_NPR1 A linecolor=0/255/0 pos=6379.6,1415.3  
metabolite\_Ethylene protein\_PR1 A linecolor=0/255/0 pos=6526.0,1322.6  
metabolite\_Ethylene protein\_PR2 A linecolor=0/255/0 pos=6558.1,1494.9  
metabolite\_Ethylene protein\_PR5 A linecolor=0/255/0 pos=6533.8,1578.9  
metabolite\_Isochorismate metabolite\_SA\_chl A linecolor=255/0/0 pos=8981.9,1379.0  
metabolite\_JA metabolite\_JA-Ile A linecolor=0/0/0 pos=5363.0,2488.8  
metabolite\_JA metabolite\_Me-JA A linecolor=0/0/0 pos=5240.2,2576.6

metabolite\_JA metabolite\_ROS A linecolor=0/255/0 pos=7512.1,1619.8  
 metabolite\_JA metabolite\_SA I linecolor=0/255/0 pos=6958.6,2148.7  
 metabolite\_JA protein\_ACX1 A linecolor=0/255/0 pos=4864.2,2314.6  
 metabolite\_JA protein\_AOS A linecolor=0/255/0 pos=5302.5,1805.4  
 metabolite\_JA protein\_CLH1 A linecolor=0/255/0 pos=5855.8,1568.0  
 metabolite\_JA protein\_EBF1 I linecolor=0/255/0 pos=4596.8,1719.3  
 metabolite\_JA protein\_ERF1 A linecolor=0/255/0 pos=4748.8,2039.8  
 metabolite\_JA protein\_KAT5 A linecolor=0/255/0 pos=5495.1,2285.6  
 metabolite\_JA protein\_LOX1 A linecolor=0/255/0 pos=5330.7,1606.7  
 metabolite\_JA protein\_LOX2 A linecolor=0/255/0 pos=5369.4,1663.6  
 metabolite\_JA protein\_MPK6 A linecolor=0/255/0 pos=7361.9,1665.8  
 metabolite\_JA protein\_OPR3 A linecolor=0/255/0 pos=5108.5,2007.5  
 metabolite\_JA protein\_PR1 A linecolor=0/255/0 pos=7681.0,1795.7  
 metabolite\_JA protein\_PR1 I linecolor=0/255/0 pos=7618.6,1804.2  
 metabolite\_JA protein\_PR2 A linecolor=0/255/0 pos=7669.9,1923.6  
 metabolite\_JA protein\_PR4 A linecolor=0/255/0 pos=4830.0,1756.7  
 metabolite\_JA protein\_PR5 A linecolor=0/255/0 pos=7657.5,1994.3  
 metabolite\_JA protein\_WRKY70 I linecolor=0/255/0 pos=7540.5,1844.0  
 metabolite\_JA protein\_b-CHI A linecolor=0/255/0 pos=4836.8,1666.4  
 metabolite\_JA-Ile complex\_JA-Ile\_COI1\_SCF P linecolor=255/0/0 pos=5760.8,2448.5  
 metabolite\_JA-Ile protein\_COI1 B linecolor=0/0/0 pos=5568.3,2565.8  
 metabolite\_L-methionine metabolite\_SAM A linecolor=255/0/0 pos=2793.3,742.0  
 metabolite\_Linolenic\_acid metabolite\_13-HPT A linecolor=0/0/0 pos=5101.9,946.9  
 metabolite\_Me-JA metabolite\_ACC I linecolor=0/255/0 pos=3946.8,1865.2  
 metabolite\_Me-JA metabolite\_JA A linecolor=0/255/0 pos=5137.9,2568.6  
 metabolite\_Me-JA metabolite\_ROS A linecolor=0/255/0 pos=7551.7,1689.1  
 metabolite\_Me-JA metabolite\_ROS I linecolor=0/255/0 pos=7494.1,1701.4  
 metabolite\_Me-JA metabolite\_SA A linecolor=0/255/0 pos=6995.1,2254.5  
 metabolite\_Me-JA metabolite\_SAG A linecolor=0/255/0 pos=7112.8,2300.7  
 metabolite\_Me-JA protein\_AOS A linecolor=0/255/0 pos=5345.0,1872.0  
 metabolite\_Me-JA protein\_AOS I linecolor=0/255/0 pos=5281.3,1878.3  
 metabolite\_Me-JA protein\_ERF1 A linecolor=0/255/0 pos=4742.1,2109.0  
 metabolite\_Me-JA protein\_ETR1 A linecolor=0/255/0 pos=4202.9,1925.5  
 metabolite\_Me-JA protein\_LOX1 A linecolor=0/255/0 pos=5304.4,1693.9  
 metabolite\_Me-JA protein\_LOX2 A linecolor=0/255/0 pos=5357.7,1740.5  
 metabolite\_Me-JA protein\_NADPH\_oxidase A linecolor=0/255/0 pos=7640.3,1702.2  
 metabolite\_Me-JA protein\_OPR3 A linecolor=0/255/0 pos=5094.0,2067.6  
 metabolite\_Me-JA protein\_PDF1.2 I linecolor=0/255/0 pos=4791.7,2210.5  
 metabolite\_Me-JA protein\_PR1 A linecolor=0/255/0 pos=7677.4,1864.6  
 metabolite\_Me-JA protein\_PR1 I linecolor=0/255/0 pos=7615.7,1882.9  
 metabolite\_Me-JA protein\_PR4 A linecolor=0/255/0 pos=4825.9,1832.6  
 metabolite\_Me-JA protein\_PR5 A linecolor=0/255/0 pos=7637.0,2059.1  
 metabolite\_Me-JA protein\_VSP1 A linecolor=0/255/0 pos=5927.2,1641.9  
 metabolite\_Me-JA protein\_WRKY70 I linecolor=0/255/0 pos=7545.3,1919.8  
 metabolite\_OPC-8:0-CoA metabolite\_OPC6 A linecolor=255/0/0 pos=5135.7,1752.6  
 metabolite\_OPC4 metabolite\_JA A linecolor=255/0/0 pos=5137.8,2385.8  
 metabolite\_OPC6 metabolite\_OPC4 A  
 linecolor=255/0/0 pinned=1 pos=5245.0036,2026.1763  
 metabolite\_OPC8 metabolite\_OPC-8:0-CoA A linecolor=255/0/0 pos=5237.6,1631.1  
 metabolite\_OPDA metabolite\_OPC8 A linecolor=255/0/0 pos=5235.4,1516.9  
 metabolite\_OPDA protein\_AOS A linecolor=0/255/0 pos=5272.0,1325.3  
 metabolite\_OPDA protein\_GST1 A linecolor=0/255/0 pos=4834.6,1344.1

metabolite\_OPDA\_chl metabolite\_OPDA A linecolor=255/0/0 pos=5116.0,1387.3  
metabolite\_Orto-coumaric\_acid metabolite\_SA A linecolor=255/0/0 pos=8610.6,1746.2  
metabolite\_Phenylalanine metabolite\_Trans-cinnamic\_acid A  
linecolor=0/0/0 pos=8503.0,1306.5  
metabolite\_Phenylpyruvate metabolite\_Phenylalanine A  
linecolor=255/0/0 pos=8518.0,1140.5  
metabolite\_Prephenate metabolite\_Phenylpyruvate A linecolor=0/0/0 pos=8513.4,974.6  
metabolite\_ROS metabolite\_Ethylene A linecolor=0/255/0 pos=6469.2,1127.8  
metabolite\_ROS metabolite\_JA A linecolor=0/255/0 pos=7570.3,1612.2  
metabolite\_ROS metabolite\_SAG A linecolor=0/255/0 pos=9403.1,1416.5  
metabolite\_ROS protein\_ACO1 A linecolor=0/0/255 pos=6305.5,1053.3  
metabolite\_ROS protein\_BA2H A linecolor=255/0/0 pos=9559.3,1181.9  
metabolite\_ROS protein\_ERF1 A linecolor=0/255/0 pos=7101.2,1172.3  
metabolite\_ROS protein\_GST1 A linecolor=0/255/0 pos=7169.3,990.5  
metabolite\_ROS protein\_MPK3 A linecolor=0/0/0 pos=9421.8,742.4  
metabolite\_ROS protein\_MPK4 A linecolor=0/255/0 pos=7896.2,882.8  
metabolite\_ROS protein\_MPK6 A linecolor=0/0/0 pinned=1 pos=9694.2650,804.1757  
metabolite\_ROS protein\_PR1 A linecolor=0/255/0 pos=9967.4,995.8  
metabolite\_ROS protein\_PR2 A linecolor=0/255/0 pos=10000.7,1074.6  
metabolite\_ROS protein\_Superoxide\_dismutase I linecolor=0/255/0 pos=9951.8,696.8  
metabolite\_SA metabolite\_ACC I linecolor=0/255/0 pos=5789.7,1534.0  
metabolite\_SA metabolite\_JA A linecolor=0/255/0 pos=7012.9,2169.7  
metabolite\_SA metabolite\_OPDA A linecolor=0/255/0 pos=6966.4,1730.5  
metabolite\_SA metabolite\_OPDA I linecolor=255/0/0 pos=7018.7,1755.2  
metabolite\_SA metabolite\_SAG A linecolor=0/0/0 pos=8879.5,2034.1  
metabolite\_SA metabolite\_SGE A linecolor=255/0/0 pos=8661.9,2021.9  
metabolite\_SA protein\_ACO2 I linecolor=0/255/0 pos=5765.9,1705.7  
metabolite\_SA protein\_AOC2 I linecolor=0/255/0 pos=7170.1,1692.9  
metabolite\_SA protein\_AOS A linecolor=0/255/0 pos=7057.2,1625.9  
metabolite\_SA protein\_AOS I linecolor=0/255/0 pos=7112.1,1605.1  
metabolite\_SA protein\_Catalase I linecolor=0/255/0 pos=9436.9,1358.3  
metabolite\_SA protein\_EDS1 A linecolor=0/255/0 pos=9168.2,1429.1  
metabolite\_SA protein\_EDS5 A linecolor=0/255/0 pos=9082.3,1570.0  
metabolite\_SA protein\_ERF1 A linecolor=0/255/0 pos=6547.4,1834.3  
metabolite\_SA protein\_HRT A linecolor=0/255/0 pos=9203.1,1221.8  
metabolite\_SA protein\_JAZ3 I linecolor=0/255/0 pos=7740.4,2108.0  
metabolite\_SA protein\_LOX1 A linecolor=0/255/0 pos=7124.7,1276.2  
metabolite\_SA protein\_LOX2 I linecolor=0/255/0 pos=7118.5,1354.3  
metabolite\_SA protein\_NPR1 A linecolor=0/0/0 pinned=1 pos=9584.0072,1631.6197  
metabolite\_SA protein\_OPR3 A linecolor=0/255/0 pos=6867.6,1785.9  
metabolite\_SA protein\_OPR3 I linecolor=0/255/0 pos=6914.0,1820.4  
metabolite\_SA protein\_PAD4 A linecolor=0/255/0 pos=9063.1,1429.4  
metabolite\_SA protein\_PDF1.2 A linecolor=0/255/0 pos=6651.9,1879.5  
metabolite\_SA protein\_PDF1.2 I linecolor=0/255/0 pos=6610.2,1920.5  
metabolite\_SA protein\_PR1 I linecolor=0/255/0 pos=9422.6,1589.4  
metabolite\_SA protein\_THI2.1 A linecolor=0/255/0 pos=7787.0,1340.4  
metabolite\_SA protein\_b-CHI A linecolor=0/255/0 pos=6625.1,1406.4  
metabolite\_SAG metabolite\_SA A linecolor=0/0/255 pos=8912.4,1970.6  
metabolite\_SAG protein\_PR1 A linecolor=0/255/0 pos=9507.9,1655.4  
metabolite\_SAM metabolite\_ACC A linecolor=0/0/0 pos=2807.6,968.7  
metabolite\_SA\_chl metabolite\_SA A linecolor=255/0/0 pinned=1 pos=8999.1799,1585.9679  
metabolite\_Trans-cinnamic\_acid metabolite\_BA A linecolor=255/0/0 pos=8607.7,1384.6

metabolite\_Trans-cinnamic\_acid metabolite\_Orto-coumaric\_acid A  
linecolor=255/0/0 pinned=1 pos=8450.7992,1455.7455  
protein\_AAO4 metabolite\_BA A linecolor=255/0/0 pinned=1 pos=8748.5817,1512.1865  
protein\_AAO4 metabolite\_Orto-coumaric\_acid A  
linecolor=255/0/0 pinned=1 pos=8556.8009,1510.1879  
protein\_ACO metabolite\_Ethylene A linecolor=0/0/0 pos=2880.8,1685.1  
protein\_ACO-like metabolite\_Ethylene A linecolor=255/0/0 pos=2955.8,1712.5  
protein\_ACO1 metabolite\_Ethylene A linecolor=0/0/0 pos=2871.0,1468.6  
protein\_ACO2 metabolite\_Ethylene A linecolor=0/0/0 pos=2925.6,1556.9  
protein\_ACO4 metabolite\_Ethylene A linecolor=0/0/0 pos=2848.6,1608.3  
protein\_ACS1 metabolite\_ACC A linecolor=0/0/0 pos=2882.5,1017.5  
protein\_ACS10 metabolite\_ACC A linecolor=255/0/0 pos=2727.1,1087.7  
protein\_ACS11 metabolite\_ACC A linecolor=255/0/0 pos=2736.1,1022.2  
protein\_ACS2 metabolite\_ACC A linecolor=255/0/0 pos=2855.0,1077.0  
protein\_ACS3 metabolite\_ACC A linecolor=255/0/0 pos=2916.1,1113.6  
protein\_ACS4 metabolite\_ACC A linecolor=0/0/0 pos=2919.8,1195.5  
protein\_ACS5 metabolite\_ACC A linecolor=255/0/0 pos=2851.5,1246.7  
protein\_ACS6 metabolite\_ACC A linecolor=255/0/0 pos=2895.6,1291.7  
protein\_ACS7 metabolite\_ACC A linecolor=0/0/0 pos=2712.6,1276.0  
protein\_ACS8 metabolite\_ACC A linecolor=255/0/0 pos=2710.5,1208.8  
protein\_ACS9 metabolite\_ACC A linecolor=255/0/0 pos=2708.4,1146.7  
protein\_ACX1 metabolite\_JA A linecolor=0/0/0 pos=4840.0,2368.0  
protein\_ACX1 metabolite\_OPC4 A linecolor=255/0/0 pos=4862.1,2235.2  
protein\_ACX1 metabolite\_OPC6 A linecolor=255/0/0 pos=4860.4,1971.5  
protein\_ACX2 metabolite\_JA A linecolor=255/0/0 pos=4930.4,2359.9  
protein\_ACX2 metabolite\_OPC4 A linecolor=255/0/0 pos=4937.0,2225.8  
protein\_ACX2 metabolite\_OPC6 A linecolor=255/0/0 pos=4932.4,1985.5  
protein\_ACX3 metabolite\_JA A linecolor=255/0/0 pos=4998.9,2372.6  
protein\_ACX3 metabolite\_OPC4 A linecolor=255/0/0 pos=5000.7,2259.1  
protein\_ACX3 metabolite\_OPC6 A linecolor=255/0/0 pos=4999.5,2006.8  
protein\_ACX4 metabolite\_JA A linecolor=255/0/0 pos=5060.2,2337.5  
protein\_ACX4 metabolite\_OPC4 A linecolor=255/0/0 pos=5051.4,2210.1  
protein\_ACX4 metabolite\_OPC6 A linecolor=255/0/0 pos=5049.0,1955.9  
protein\_ACX5 metabolite\_JA A linecolor=255/0/0 pos=5137.6,2309.1  
protein\_ACX5 metabolite\_OPC4 A linecolor=255/0/0 pos=5112.2,2224.1  
protein\_ACX5 metabolite\_OPC6 A linecolor=255/0/0 pos=5142.0,1951.5  
protein\_ACX6 metabolite\_JA A linecolor=255/0/0 pos=5232.5,2322.2  
protein\_ACX6 metabolite\_OPC4 A linecolor=255/0/0 pos=5192.2,2167.9  
protein\_ACX6 metabolite\_OPC6 A linecolor=255/0/0 pos=5221.6,1968.3  
protein\_AIM1 metabolite\_JA A linecolor=255/0/0 pos=5308.4,2316.3  
protein\_AIM1 metabolite\_OPC4 A linecolor=255/0/0 pos=5297.9,2214.4  
protein\_AIM1 metabolite\_OPC6 A linecolor=255/0/0 pos=5315.0,1980.3  
protein\_AOC1 metabolite\_OPDA\_chl A linecolor=255/0/0 pos=5331.2,1259.7  
protein\_AOC2 metabolite\_OPDA\_chl A linecolor=255/0/0 pos=5380.7,1313.5  
protein\_AOC3 metabolite\_OPDA\_chl A linecolor=255/0/0 pos=5340.2,1362.7  
protein\_AOC4 metabolite\_OPDA\_chl A linecolor=255/0/0 pos=5362.0,1422.4  
protein\_AOS metabolite\_12/13\_EDT A linecolor=0/0/0 pos=5278.8,1187.5  
protein\_Arogenate\_dehydratase metabolite\_Phenylalanine A  
linecolor=0/0/0 pos=8339.5,1198.0  
protein\_BA2H metabolite\_SA A linecolor=255/0/0 pos=9044.2,1756.7  
protein\_CM1 metabolite\_Prephenate A linecolor=255/0/0 pos=8359.4,777.5  
protein\_CM2 metabolite\_Prephenate A linecolor=255/0/0 pos=8360.1,933.5

protein\_CM3 metabolite\_Prephenate A linecolor=255/0/0 pos=8358.6,851.7  
protein\_COI1 complex\_JA-Ile\_COI1\_SCF P linecolor=255/0/0 pos=5842.6,2530.1  
protein\_COI1 protein\_KAT5 A linecolor=0/255/0 pos=5752.4,2348.4  
protein\_CTR1 complex\_Et\_receptor\_CTR1 P  
linecolor=255/0/0 pinned=1 pos=3531.1004,1464.2793  
protein\_CTR1 protein\_EIN4 B linecolor=255/0/0 pos=3658.7,1316.5  
protein\_CTR1 protein\_ERS1 B linecolor=255/0/0 pos=3495.9,1336.6  
protein\_CTR1 protein\_ERS2 B linecolor=255/0/0 pos=3580.2,1328.9  
protein\_CTR1 protein\_ETR1 B linecolor=255/0/0 pos=3348.6,1348.4  
protein\_CTR1 protein\_ETR2 B linecolor=255/0/0 pos=3417.7,1347.1  
protein\_Catalase metabolite\_ROS A linecolor=255/0/0 pos=9990.3,789.3  
protein\_Catalase metabolite\_ROS I linecolor=0/255/0 pos=9940.9,831.6  
protein\_EBF1 complex\_ASK1\_CULLIN1\_RBXE2\_EBF12 P  
linecolor=255/0/0 pos=4079.4,983.7  
protein\_EBF2 complex\_ASK1\_CULLIN1\_RBXE2\_EBF12 P  
linecolor=255/0/0 pos=3966.3,1126.0  
protein\_EDF1 protein\_PDF1.2 A linecolor=255/0/0 pos=4432.8,1758.5  
protein\_EDF2 protein\_PDF1.2 A linecolor=255/0/0 pos=4420.8,1857.3  
protein\_EDF3 protein\_PDF1.2 A linecolor=255/0/0 pos=4469.2,1911.6  
protein\_EDF4 protein\_PDF1.2 A linecolor=255/0/0 pos=4429.3,1970.3  
protein\_EDS1 metabolite\_ACC A linecolor=0/255/0 pos=6172.7,1043.1  
protein\_EDS1 metabolite\_Ethylene A linecolor=0/255/0 pos=6263.0,1259.2  
protein\_EDS1 protein\_EDS5 A linecolor=255/0/0 pos=9443.8,1019.2  
protein\_EDS1 protein\_HRT A linecolor=0/255/0 pos=9631.5,726.2  
protein\_EDS5 protein\_ICS1 A linecolor=0/0/0 pinned=1 pos=9277.5793,1190.3358  
protein\_EDS5 protein\_ICS2 A linecolor=255/0/0 pinned=1 pos=9411.1688,1188.8498  
protein\_EDS5 protein\_PR1 I linecolor=0/255/0 pos=9720.0,1174.7  
protein\_EIL1 metabolite\_SA I linecolor=0/255/0 pos=6280.7,1842.3  
protein\_EIL1 protein\_EBF1 A linecolor=255/0/0 pos=3868.2,1358.3  
protein\_EIL1 protein\_EBF2 A linecolor=0/0/0 pos=3786.0,1490.7  
protein\_EIL1 protein\_EDF1 A linecolor=255/0/0 pos=4103.5,1719.6  
protein\_EIL1 protein\_EDF2 A linecolor=255/0/0 pos=4119.7,1778.6  
protein\_EIL1 protein\_EDF3 A linecolor=255/0/0 pos=4128.9,1840.2  
protein\_EIL1 protein\_EDF4 A linecolor=255/0/0 pos=4097.3,1896.6  
protein\_EIL1 protein\_ERF1 A linecolor=0/0/0 pos=4064.2,1669.6  
protein\_EIL1 protein\_GST1 A linecolor=255/0/0 pos=4199.3,1526.4  
protein\_EIL1 protein\_ICS1 I linecolor=0/0/255 pos=6486.0,1543.1  
protein\_EIL1 protein\_PR4 A linecolor=255/0/0 pos=4190.1,1398.7  
protein\_EIL1 protein\_b-CHI A linecolor=255/0/0 pos=4186.4,1299.2  
protein\_EIL2 protein\_EBF1 A linecolor=255/0/0 pos=3860.6,1425.1  
protein\_EIL2 protein\_EBF2 A linecolor=255/0/0 pos=3783.8,1556.6  
protein\_EIL2 protein\_EDF1 A linecolor=255/0/0 pos=4054.7,1771.3  
protein\_EIL2 protein\_EDF2 A linecolor=255/0/0 pos=4066.2,1833.5  
protein\_EIL2 protein\_EDF3 A linecolor=255/0/0 pos=3992.0,1924.0  
protein\_EIL2 protein\_EDF4 A linecolor=255/0/0 pos=4102.6,1968.6  
protein\_EIL2 protein\_ERF1 A linecolor=255/0/0 pos=4017.2,1721.1  
protein\_EIL2 protein\_GST1 A linecolor=255/0/0 pos=4120.0,1558.5  
protein\_EIL2 protein\_PR4 A linecolor=255/0/0 pos=4204.8,1461.8  
protein\_EIL2 protein\_b-CHI A linecolor=255/0/0 pos=4139.5,1347.7  
protein\_EIN2 metabolite\_SA A linecolor=0/255/0 pos=6140.8,1857.4  
protein\_EIN2 protein\_EIL1 A linecolor=255/0/0 pos=3648.6,1763.9  
protein\_EIN2 protein\_EIL2 A linecolor=255/0/0 pos=3634.7,1838.9

protein\_EIN2 protein\_EIN3 A linecolor=0/0/0 pos=3631.4,1915.1  
 protein\_EIN2 protein\_PR1 I linecolor=0/255/0 pos=6780.3,1498.2  
 protein\_EIN3 metabolite\_SA I linecolor=0/255/0 pos=6287.0,1947.2  
 protein\_EIN3 protein\_EBF1 A linecolor=0/0/0 pos=3884.0,1500.6  
 protein\_EIN3 protein\_EBF2 A linecolor=0/0/0 pos=3811.0,1621.3  
 protein\_EIN3 protein\_EDF1 A linecolor=255/0/0 pos=4030.5,1874.0  
 protein\_EIN3 protein\_EDF2 A linecolor=255/0/0 pos=4053.2,1940.3  
 protein\_EIN3 protein\_EDF3 A linecolor=255/0/0 pos=4020.3,1998.0  
 protein\_EIN3 protein\_EDF4 A linecolor=255/0/0 pos=4076.1,2038.5  
 protein\_EIN3 protein\_ERF1 A linecolor=0/0/0 pos=4005.6,1803.2  
 protein\_EIN3 protein\_GST1 A linecolor=255/0/0 pos=4177.9,1611.8  
 protein\_EIN3 protein\_ICS1 B linecolor=0/0/255 pos=6538.5,1667.2  
 protein\_EIN3 protein\_ICS1 I linecolor=0/0/255 pos=6479.7,1654.5  
 protein\_EIN3 protein\_PR4 A linecolor=255/0/0 pos=4138.8,1503.3  
 protein\_EIN3 protein\_b-CHI A linecolor=255/0/0 pos=4132.7,1426.1  
 protein\_EIN4 complex\_Et\_receptor\_CTR1 P linecolor=255/0/0 pos=3668.9,1404.0  
 protein\_EIN5 protein\_EBF1 I linecolor=255/0/0 pos=3890.5,955.6  
 protein\_EIN5 protein\_EBF2 I linecolor=0/0/0 pos=3823.4,1117.9  
 protein\_ERF1 protein\_PDF1.2 A linecolor=0/0/0 pos=4466.9,1692.0  
 protein\_ERS1 complex\_Et\_receptor\_CTR1 P linecolor=255/0/0 pos=3481.4,1430.8  
 protein\_ERS2 complex\_Et\_receptor\_CTR1 P linecolor=255/0/0 pos=3611.1,1454.5  
 protein\_ETR1 complex\_Et\_receptor\_CTR1 P linecolor=255/0/0 pos=3356.9,1455.2  
 protein\_ETR1 metabolite\_ROS A linecolor=0/255/0 pos=6509.5,986.3  
 protein\_ETR1 protein\_ERS1 I linecolor=0/255/0 pos=3271.4,1306.4  
 protein\_ETR2 complex\_Et\_receptor\_CTR1 P linecolor=255/0/0 pos=3420.9,1466.0  
 protein\_Glutathione\_peroxidase metabolite\_ROS A linecolor=255/0/0 pos=9870.2,655.9  
 protein\_Glutathione\_peroxidase metabolite\_ROS I linecolor=0/255/0 pos=9882.0,710.1  
 protein\_HRT protein\_MPK3 A linecolor=255/0/0 pos=9353.1,626.0  
 protein\_HRT protein\_MPK6 A linecolor=255/0/0 pos=9567.1,609.7  
 protein\_ICS1 metabolite\_Isochorismate A linecolor=0/0/0 pos=9105.7,1253.2  
 protein\_ICS2 metabolite\_Isochorismate A linecolor=0/0/0 pos=9181.9,1296.1  
 protein\_IPL metabolite\_SA\_chl A linecolor=255/0/0 pos=8897.3,1393.5  
 protein\_Inactive\_BA2H protein\_BA2H A linecolor=255/0/0 pos=9353.7,1520.3  
 protein\_Inactive\_EIN2 protein\_EIN2 A  
 linecolor=255/0/0 pinned=1 pos=3402.2653,1772.3357  
 protein\_Inactive\_HRT protein\_HRT A linecolor=255/0/0 pinned=1 pos=9791.7682,615.9390  
 protein\_Inactive\_MPK3 protein\_MPK3 A  
 linecolor=255/0/0 pinned=1 pos=9064.9240,708.8540  
 protein\_Inactive\_MPK6 protein\_MPK6 A  
 linecolor=255/0/0 pinned=1 pos=9507.8179,725.7973  
 protein\_Inactive\_et\_receptor protein\_EIN4 A linecolor=255/0/0 pos=3686.3,1092.4  
 protein\_Inactive\_et\_receptor protein\_ERS1 A linecolor=255/0/0 pos=3527.4,1153.3  
 protein\_Inactive\_et\_receptor protein\_ERS2 A linecolor=255/0/0 pos=3609.3,1146.0  
 protein\_Inactive\_et\_receptor protein\_ETR1 A linecolor=255/0/0 pos=3375.4,1124.9  
 protein\_Inactive\_et\_receptor protein\_ETR2 A linecolor=255/0/0 pos=3439.8,1085.3  
 protein\_JAR1 metabolite\_JA-Ile A linecolor=0/0/0 pos=5598.5,2426.6  
 protein\_JAR1 protein\_PR1 I linecolor=0/255/0 pos=7830.9,1769.4  
 protein\_JAZ1 complex\_JA-Ile\_COI1\_SCF\_JAZ P linecolor=255/0/0 pos=6745.9,2404.2  
 protein\_JAZ1 complex\_JAZ\_MYC2\_TF P linecolor=255/0/0 pos=6881.4,1639.6  
 protein\_JAZ1 protein\_CLH1 I linecolor=255/0/0 pos=6319.4,1494.0  
 protein\_JAZ1 protein\_JR1 I linecolor=255/0/0 pos=6237.5,1438.1  
 protein\_JAZ1 protein\_MYC2 B linecolor=255/0/0 pos=6723.7,1383.8

protein\_JAZ1 protein\_THI2.1 I linecolor=255/0/0 pos=6432.0,1574.4  
protein\_JAZ1 protein\_VSP1 I linecolor=255/0/0 pos=6389.0,1521.3  
protein\_JAZ10 complex\_JA-Ile\_COI1\_SCF\_JAZ P linecolor=255/0/0 pos=7488.3,2355.6  
protein\_JAZ10 complex\_JAZ\_MYC2\_TF P linecolor=255/0/0 pos=7654.7,1555.1  
protein\_JAZ10 protein\_CLH1 I linecolor=255/0/0 pos=7106.0,1457.6  
protein\_JAZ10 protein\_JR1 I linecolor=255/0/0 pos=7005.6,1311.4  
protein\_JAZ10 protein\_MYC2 B linecolor=255/0/0 pos=7472.1,1362.4  
protein\_JAZ10 protein\_THI2.1 I linecolor=255/0/0 pos=7237.7,1530.9  
protein\_JAZ10 protein\_VSP1 I linecolor=255/0/0 pos=7185.2,1493.2  
protein\_JAZ11 complex\_JA-Ile\_COI1\_SCF\_JAZ P linecolor=255/0/0 pos=7562.3,2376.2  
protein\_JAZ11 complex\_JAZ\_MYC2\_TF P linecolor=255/0/0 pos=7723.3,1533.6  
protein\_JAZ11 protein\_CLH1 I linecolor=255/0/0 pos=7201.5,1440.1  
protein\_JAZ11 protein\_JR1 I linecolor=255/0/0 pos=7074.4,1320.7  
protein\_JAZ11 protein\_MYC2 B linecolor=255/0/0 pos=7544.4,1366.7  
protein\_JAZ11 protein\_THI2.1 I linecolor=255/0/0 pos=7317.8,1532.2  
protein\_JAZ11 protein\_VSP1 I linecolor=255/0/0 pos=7271.9,1484.0  
protein\_JAZ12 complex\_JA-Ile\_COI1\_SCF\_JAZ P linecolor=255/0/0 pos=7644.7,2368.0  
protein\_JAZ12 complex\_JAZ\_MYC2\_TF P linecolor=255/0/0 pos=7809.9,1534.6  
protein\_JAZ12 protein\_CLH1 I linecolor=255/0/0 pos=7279.4,1428.4  
protein\_JAZ12 protein\_JR1 I linecolor=255/0/0 pos=7185.0,1317.5  
protein\_JAZ12 protein\_MYC2 B linecolor=255/0/0 pos=7626.4,1373.1  
protein\_JAZ12 protein\_THI2.1 I linecolor=255/0/0 pos=7402.6,1511.8  
protein\_JAZ12 protein\_VSP1 I linecolor=255/0/0 pos=7351.1,1471.0  
protein\_JAZ2 complex\_JA-Ile\_COI1\_SCF\_JAZ P linecolor=255/0/0 pos=6846.4,2408.7  
protein\_JAZ2 complex\_JAZ\_MYC2\_TF P linecolor=255/0/0 pos=6981.9,1626.7  
protein\_JAZ2 protein\_CLH1 I linecolor=255/0/0 pos=6435.0,1465.2  
protein\_JAZ2 protein\_JR1 I linecolor=255/0/0 pos=6322.5,1411.1  
protein\_JAZ2 protein\_MYC2 B linecolor=255/0/0 pos=6818.2,1413.2  
protein\_JAZ2 protein\_THI2.1 I linecolor=255/0/0 pos=6589.9,1578.0  
protein\_JAZ2 protein\_VSP1 I linecolor=255/0/0 pos=6500.6,1484.6  
protein\_JAZ3 complex\_JA-Ile\_COI1\_SCF\_JAZ P linecolor=255/0/0 pos=6939.0,2415.9  
protein\_JAZ3 complex\_JAZ\_MYC2\_TF P linecolor=255/0/0 pos=7097.2,1662.4  
protein\_JAZ3 protein\_CLH1 I linecolor=255/0/0 pos=6548.5,1438.6  
protein\_JAZ3 protein\_JR1 I linecolor=255/0/0 pos=6446.9,1394.5  
protein\_JAZ3 protein\_MYC2 B linecolor=255/0/0 pos=6900.1,1419.6  
protein\_JAZ3 protein\_THI2.1 I linecolor=255/0/0 pos=6665.0,1595.0  
protein\_JAZ3 protein\_VSP1 I linecolor=255/0/0 pos=6625.3,1525.4  
protein\_JAZ4 complex\_JA-Ile\_COI1\_SCF\_JAZ P linecolor=255/0/0 pos=7027.0,2404.3  
protein\_JAZ4 complex\_JAZ\_MYC2\_TF P linecolor=255/0/0 pos=7182.3,1620.3  
protein\_JAZ4 protein\_CLH1 I linecolor=255/0/0 pos=6627.9,1465.5  
protein\_JAZ4 protein\_JR1 I linecolor=255/0/0 pos=6520.8,1388.1  
protein\_JAZ4 protein\_MYC2 B linecolor=255/0/0 pos=6965.6,1366.8  
protein\_JAZ4 protein\_THI2.1 I linecolor=255/0/0 pos=6741.0,1603.4  
protein\_JAZ4 protein\_VSP1 I linecolor=255/0/0 pos=6697.8,1531.5  
protein\_JAZ5 complex\_JA-Ile\_COI1\_SCF\_JAZ P linecolor=255/0/0 pos=7100.0,2391.8  
protein\_JAZ5 complex\_JAZ\_MYC2\_TF P linecolor=255/0/0 pos=7268.6,1638.1  
protein\_JAZ5 protein\_CLH1 I linecolor=255/0/0 pos=6706.3,1472.0  
protein\_JAZ5 protein\_JR1 I linecolor=255/0/0 pos=6601.9,1356.2  
protein\_JAZ5 protein\_MYC2 B linecolor=255/0/0 pos=7037.7,1370.0  
protein\_JAZ5 protein\_THI2.1 I linecolor=255/0/0 pos=6819.9,1586.1  
protein\_JAZ5 protein\_VSP1 I linecolor=255/0/0 pos=6768.6,1545.8  
protein\_JAZ6 complex\_JA-Ile\_COI1\_SCF\_JAZ P linecolor=255/0/0 pos=7178.5,2366.1

protein\_JAZ6 complex\_JAZ\_MYC2\_TF P linecolor=255/0/0 pos=7340.0,1596.1  
protein\_JAZ6 protein\_CLH1 I linecolor=255/0/0 pos=6771.9,1442.7  
protein\_JAZ6 protein\_JR1 I linecolor=255/0/0 pos=6685.6,1339.6  
protein\_JAZ6 protein\_MYC2 B linecolor=255/0/0 pos=7169.6,1375.8  
protein\_JAZ6 protein\_THI2.1 I linecolor=255/0/0 pos=6897.1,1568.6  
protein\_JAZ6 protein\_VSP1 I linecolor=255/0/0 pos=6851.3,1518.7  
protein\_JAZ7 complex\_JA-Ile\_COI1\_SCF\_JAZ P linecolor=255/0/0 pos=7265.0,2344.6  
protein\_JAZ7 complex\_JAZ\_MYC2\_TF P linecolor=255/0/0 pos=7418.7,1585.0  
protein\_JAZ7 protein\_CLH1 I linecolor=255/0/0 pos=6859.9,1460.9  
protein\_JAZ7 protein\_JR1 I linecolor=255/0/0 pos=6766.4,1318.5  
protein\_JAZ7 protein\_MYC2 B linecolor=255/0/0 pos=7249.9,1330.3  
protein\_JAZ7 protein\_THI2.1 I linecolor=255/0/0 pos=6974.6,1558.8  
protein\_JAZ7 protein\_VSP1 I linecolor=255/0/0 pos=6927.3,1511.9  
protein\_JAZ8 complex\_JA-Ile\_COI1\_SCF\_JAZ P linecolor=255/0/0 pos=7347.8,2364.5  
protein\_JAZ8 complex\_JAZ\_MYC2\_TF P linecolor=255/0/0 pos=7495.3,1543.8  
protein\_JAZ8 protein\_CLH1 I linecolor=255/0/0 pos=6948.5,1461.1  
protein\_JAZ8 protein\_JR1 I linecolor=255/0/0 pos=6853.9,1317.6  
protein\_JAZ8 protein\_MYC2 B linecolor=255/0/0 pos=7331.3,1394.5  
protein\_JAZ8 protein\_THI2.1 I linecolor=255/0/0 pos=7060.4,1550.6  
protein\_JAZ8 protein\_VSP1 I linecolor=255/0/0 pos=7014.1,1509.9  
protein\_JAZ9 complex\_JA-Ile\_COI1\_SCF\_JAZ P linecolor=255/0/0 pos=7421.3,2388.2  
protein\_JAZ9 complex\_JAZ\_MYC2\_TF P linecolor=255/0/0 pos=7580.6,1535.9  
protein\_JAZ9 protein\_CLH1 I linecolor=255/0/0 pos=7030.2,1457.5  
protein\_JAZ9 protein\_JR1 I linecolor=255/0/0 pos=6931.2,1313.5  
protein\_JAZ9 protein\_MYC2 B linecolor=255/0/0 pos=7400.8,1349.8  
protein\_JAZ9 protein\_THI2.1 I linecolor=255/0/0 pos=7156.9,1547.5  
protein\_JAZ9 protein\_VSP1 I linecolor=255/0/0 pos=7105.7,1510.2  
protein\_JMT metabolite\_Me-JA A linecolor=0/0/0 pos=5084.1,2631.6  
protein\_KAT1 metabolite\_JA A linecolor=255/0/0 pos=5376.7,2342.6  
protein\_KAT1 metabolite\_OPC4 A linecolor=255/0/0 pos=5367.4,2201.9  
protein\_KAT1 metabolite\_OPC6 A linecolor=255/0/0 pos=5390.0,1983.2  
protein\_KAT2 metabolite\_JA A linecolor=255/0/0 pos=5429.3,2299.0  
protein\_KAT2 metabolite\_OPC4 A linecolor=255/0/0 pos=5443.3,2187.3  
protein\_KAT2 metabolite\_OPC6 A linecolor=255/0/0 pos=5457.4,1947.0  
protein\_KAT5 metabolite\_JA A linecolor=255/0/0 pos=5559.5,2276.7  
protein\_KAT5 metabolite\_OPC4 A linecolor=255/0/0 pos=5523.0,2194.1  
protein\_KAT5 metabolite\_OPC6 A linecolor=255/0/0 pos=5533.9,1970.4  
protein\_LOX1 metabolite\_13-HPT A linecolor=255/0/0 pos=5290.0,858.0  
protein\_LOX2 metabolite\_13-HPT A linecolor=255/0/0 pos=5267.2,920.1  
protein\_LOX3 metabolite\_13-HPT A linecolor=255/0/0 pos=5322.8,953.1  
protein\_LOX4 metabolite\_13-HPT A linecolor=255/0/0 pos=5271.8,1003.3  
protein\_LOX5 metabolite\_13-HPT A linecolor=255/0/0 pos=5325.6,1044.1  
protein\_LOX6 metabolite\_13-HPT A linecolor=255/0/0 pos=5296.2,1104.0  
protein\_MAT3 metabolite\_SAM A linecolor=255/0/0 pos=2970.6,806.2  
protein\_MAT4 metabolite\_SAM A linecolor=255/0/0 pos=2958.4,881.5  
protein\_MPK3 protein\_EDS1 A linecolor=255/0/0 pos=9290.0,852.5  
protein\_MPK3 protein\_EIN3 A linecolor=0/0/255 pos=6403.9,1313.1  
protein\_MPK3 protein\_PAD3 A linecolor=255/0/0 pos=9073.0,850.3  
protein\_MPK3 protein\_PAD4 A linecolor=255/0/0 pos=9201.2,820.0  
protein\_MPK4 protein\_EDS1 I linecolor=255/0/0 pos=7754.2,959.7  
protein\_MPK4 protein\_LOX1 A linecolor=255/0/0 pos=5693.0,864.7  
protein\_MPK4 protein\_LOX2 A linecolor=255/0/0 pos=5662.3,927.1



protein\_MPK4 protein\_LOX3 A linecolor=255/0/0 pos=5716.4,965.7  
protein\_MPK4 protein\_LOX4 A linecolor=255/0/0 pos=5659.3,1015.1  
protein\_MPK4 protein\_LOX5 A linecolor=255/0/0 pos=5720.1,1052.8  
protein\_MPK4 protein\_LOX6 A linecolor=255/0/0 pos=5690.1,1111.8  
protein\_MPK4 protein\_PAD3 I linecolor=255/0/0 pos=7555.8,961.8  
protein\_MPK4 protein\_PAD4 I linecolor=255/0/0 pos=7656.6,958.9  
protein\_MPK6 protein\_EDS1 A linecolor=255/0/0 pos=9556.6,852.3  
protein\_MPK6 protein\_EIN3 A linecolor=0/0/255 pos=6636.3,1271.3  
protein\_MPK6 protein\_PAD3 A linecolor=255/0/0 pos=9306.6,781.7  
protein\_MPK6 protein\_PAD4 A linecolor=255/0/0 pos=9409.8,848.8  
protein\_MYC2 complex\_JAZ\_MYC2\_TF P linecolor=255/0/0 pos=7346.3,787.3  
protein\_MYC2 protein\_CLH1 A linecolor=255/0/0 pos=6836.2,645.1  
protein\_MYC2 protein\_GST1 A linecolor=255/0/0 pos=5839.5,943.7  
protein\_MYC2 protein\_JAZ1 A linecolor=0/0/0 pos=6687.2,1418.7  
protein\_MYC2 protein\_JAZ10 A linecolor=255/0/0 pos=7470.9,1431.2  
protein\_MYC2 protein\_JAZ11 A linecolor=255/0/0 pos=7543.4,1431.9  
protein\_MYC2 protein\_JAZ12 A linecolor=255/0/0 pos=7627.8,1432.8  
protein\_MYC2 protein\_JAZ2 A linecolor=0/0/0 pos=6801.4,1364.8  
protein\_MYC2 protein\_JAZ3 A linecolor=255/0/0 pos=6888.5,1369.3  
protein\_MYC2 protein\_JAZ4 A linecolor=255/0/0 pos=6981.1,1416.1  
protein\_MYC2 protein\_JAZ5 A linecolor=255/0/0 pos=7065.9,1411.0  
protein\_MYC2 protein\_JAZ6 A linecolor=255/0/0 pos=7140.8,1417.7  
protein\_MYC2 protein\_JAZ7 A linecolor=255/0/0 pos=7237.3,1385.4  
protein\_MYC2 protein\_JAZ8 A linecolor=255/0/0 pos=7323.9,1335.9  
protein\_MYC2 protein\_JAZ9 A linecolor=255/0/0 pos=7402.6,1415.1  
protein\_MYC2 protein\_JR1 A linecolor=255/0/0 pos=6742.2,609.4  
protein\_MYC2 protein\_PR4 A linecolor=255/0/0 pos=5843.8,849.5  
protein\_MYC2 protein\_THI2.1 A linecolor=255/0/0 pos=6968.3,719.1  
protein\_MYC2 protein\_VSP1 A linecolor=255/0/0 pos=6915.5,663.9  
protein\_MYC2 protein\_b-CHI A linecolor=255/0/0 pinned=1 pos=5952.1385,736.0863  
protein\_X4 protein\_MYC2 A linecolor=255/0/0 pos=7368.6,624.7  
protein\_NADPH\_oxidase metabolite\_ROS A linecolor=0/0/0 pos=9894.9,883.7  
protein\_NIMIN1 complex\_NPR1\_TGA245 P linecolor=255/0/0 pos=9833.8,1501.3  
protein\_NIMIN1 protein\_NPR1 B linecolor=255/0/0 pos=9696.7,1367.4  
protein\_NIMIN1 protein\_NPR1 I linecolor=0/0/255 pos=9765.9,1388.1  
protein\_NIMIN1 protein\_TGA\_TF2 B linecolor=255/0/0 pos=9626.9,1747.9  
protein\_NIMIN1 protein\_TGA\_TF4 B linecolor=255/0/0 pos=9777.0,1746.7  
protein\_NIMIN1 protein\_TGA\_TF5 B linecolor=255/0/0 pos=9892.1,1731.5  
protein\_NIMIN2 complex\_NPR1\_TGA245 P linecolor=255/0/0 pos=9818.9,1584.3  
protein\_NIMIN2 protein\_NPR1 B linecolor=255/0/0 pos=9755.5,1471.3  
protein\_NIMIN2 protein\_TGA\_TF2 B linecolor=255/0/0 pos=9664.9,1833.8  
protein\_NIMIN2 protein\_TGA\_TF4 B linecolor=255/0/0 pos=9781.8,1856.0  
protein\_NIMIN2 protein\_TGA\_TF5 B linecolor=255/0/0 pos=9894.6,1843.5  
protein\_NIMIN3 complex\_NPR1\_TGA245 P linecolor=255/0/0 pos=9700.1,1728.8  
protein\_NIMIN3 protein\_NPR1 B linecolor=255/0/0 pos=9620.8,1563.9  
protein\_NIMIN3 protein\_TGA\_TF2 B linecolor=255/0/0 pos=9531.7,1965.7  
protein\_NIMIN3 protein\_TGA\_TF4 B linecolor=255/0/0 pos=9673.5,1928.5  
protein\_NIMIN3 protein\_TGA\_TF5 B linecolor=255/0/0 pos=9762.1,1956.0  
protein\_NPR1 complex\_NPR1\_TGA245 P linecolor=255/0/0 pos=9898.5,1420.1  
protein\_NPR1 protein\_EDS1 I linecolor=255/0/0 pos=9689.6,1087.8  
protein\_NPR1 protein\_PAD3 I linecolor=255/0/0 pos=9472.6,1097.0  
protein\_NPR1 protein\_PAD4 A linecolor=0/255/0 pos=9569.0,1115.6

protein\_NPR1 protein\_PAD4 I linecolor=255/0/0 pos=9596.2,1065.3  
 protein\_NPR1 protein\_TGA\_TF2 B linecolor=0/0/0 pos=9753.8,1644.7  
 protein\_NPR1 protein\_TGA\_TF4 B linecolor=255/0/0 pos=9842.4,1662.3  
 protein\_NPR1 protein\_TGA\_TF5 B linecolor=255/0/0 pos=9965.2,1642.5  
 protein\_NPR1 protein\_WRKY70 A linecolor=0/0/0 pos=9878.0,1321.0  
 protein\_OPCL1 metabolite\_OPC6 A linecolor=255/0/0 pos=5091.2,1832.2  
 protein\_OPR3 metabolite\_OPC8 A linecolor=255/0/0 pos=5081.6,1522.6  
 protein\_PAD3 protein\_EDS5 A linecolor=255/0/0 pos=9237.3,1022.8  
 protein\_PAD4 metabolite\_ACC A linecolor=0/255/0 pos=6073.6,1038.2  
 protein\_PAD4 metabolite\_Ethylene A linecolor=0/255/0 pos=6167.6,1249.5  
 protein\_PAD4 protein\_EDS5 A linecolor=0/0/0 pos=9340.0,1035.0  
 protein\_PAD4 protein\_HRT A linecolor=0/255/0 pos=9555.4,752.4  
 protein\_PAL1 metabolite\_Trans-cinnamic\_acid A linecolor=255/0/0 pos=8331.7,1305.1  
 protein\_PAL2 metabolite\_Trans-cinnamic\_acid A linecolor=255/0/0 pos=8308.9,1365.4  
 protein\_PAL3 metabolite\_Trans-cinnamic\_acid A linecolor=255/0/0 pos=8317.3,1428.1  
 protein\_PAL4 metabolite\_Trans-cinnamic\_acid A linecolor=255/0/0 pos=8315.2,1497.2  
 protein\_PDF1.2 protein\_TGA\_TF2 B linecolor=0/0/255 pos=7065.4,1907.1  
 protein\_Prephenate\_aminotransferase metabolite\_Phenylpyruvate A  
 linecolor=255/0/0 pos=8343.2,1058.7  
 protein\_RAN1 metabolite\_Copper\_cyto A linecolor=255/0/0 pos=3453.8,942.7  
 protein\_SAM1 metabolite\_SAM A linecolor=255/0/0 pos=2954.2,701.2  
 protein\_SAM2 metabolite\_SAM A linecolor=255/0/0 pos=2917.9,767.0  
 protein\_Superoxide\_dismutase metabolite\_ROS A linecolor=255/0/0 pos=10045.1,663.1  
 protein\_Superoxide\_dismutase metabolite\_ROS I linecolor=0/255/0 pos=9938.1,756.0  
 protein\_TGA\_TF2 complex\_NPR1\_TGA245 P linecolor=255/0/0 pos=9824.0,1799.9  
 protein\_TGA\_TF2 protein\_PR1 I linecolor=0/0/255 pos=9885.5,1604.2  
 protein\_TGA\_TF4 complex\_NPR1\_TGA245 P linecolor=255/0/0 pos=9960.5,1798.5  
 protein\_TGA\_TF5 complex\_NPR1\_TGA245 P linecolor=255/0/0 pos=10047.1,1768.6  
 protein\_UGP\_glikosyltransferaze metabolite\_SAG A linecolor=255/0/0 pos=8969.9,2175.1  
 protein\_UGP\_glikosyltransferaze metabolite\_SGE A linecolor=255/0/0 pos=8702.9,2192.9  
 protein\_WRKY70 metabolite\_SA I linecolor=0/255/0 pos=9345.3,1638.2  
 protein\_WRKY70 protein\_IC51 I linecolor=0/0/255 pos=9605.1,1279.8  
 protein\_WRKY70 protein\_PDF1.2 I linecolor=0/0/255 pos=7237.4,1588.3  
 protein\_X3 metabolite\_SA A linecolor=255/0/0 pos=8953.8,1733.0  
 protein\_X5 metabolite\_OPC-8:0-CoA A linecolor=255/0/0 pos=5092.7,1656.6  
 protein\_X1 metabolite\_SA A linecolor=255/0/0 pinned=1 pos=8722.2947,1632.7280  
 protein\_X2 protein\_JAZ1 A linecolor=255/0/0 pos=6997.8,1952.2  
 protein\_X2 protein\_JAZ10 A linecolor=255/0/0 pos=7782.9,1912.0  
 protein\_X2 protein\_JAZ11 A linecolor=255/0/0 pos=7851.1,1933.1  
 protein\_X2 protein\_JAZ12 A linecolor=255/0/0 pos=7928.0,1918.2  
 protein\_X2 protein\_JAZ2 A linecolor=255/0/0 pos=7123.3,1953.5  
 protein\_X2 protein\_JAZ3 A linecolor=255/0/0 pos=7206.5,1942.1  
 protein\_X2 protein\_JAZ4 A linecolor=255/0/0 pos=7283.0,1933.9  
 protein\_X2 protein\_JAZ5 A linecolor=255/0/0 pos=7354.9,1946.9  
 protein\_X2 protein\_JAZ6 A linecolor=255/0/0 pos=7424.0,1921.6  
 protein\_X2 protein\_JAZ7 A linecolor=255/0/0 pos=7490.2,1961.7  
 protein\_X2 protein\_JAZ8 A linecolor=255/0/0 pos=7587.7,1967.5  
 protein\_X2 protein\_JAZ9 A linecolor=255/0/0 pos=7726.6,1957.8  
 # \_attributes complex\_ASK1\_CULLIN1\_RBX2 pos=3989.1318,1376.2149  
 # \_attributes complex\_ASK1\_CULLIN1\_RBX2\_EBF12 pos=4128.6972,1076.2695  
 # \_attributes complex\_Et\_receptor\_CTR1 pos=3600.8855,1574.1671  
 # \_attributes complex\_JA-Ile\_COI1\_SCF pos=5964.2300,2444.9060

```
# _attributes complex_JA-Ile_COI1_SCF_JAZ pos=7264.2563,2445.0452
# _attributes complex_JAZ_MYC2_TF pos=7534.7484,882.6253
# _attributes complex_NPR1_TGA245 pos=9935.2417,1554.3359
# _attributes complex_NPR1_oligomer pos=9849.4582,1167.8969
# _attributes complex_SCF pos=5970.6247,2569.1960
# _attributes metabolite_12/13_EDT pos=5191.3571,1206.2083
# _attributes metabolite_13-HPT pos=5191.7756,1063.7818
# _attributes metabolite_ACC pos=2828.7497,1166.0440
# _attributes metabolite_BA pos=8747.7404,1405.9435
# _attributes metabolite_Chorismate pos=8450.8641,719.0878
# _attributes metabolite_Copper pos=3430.0817,863.4805
# _attributes metabolite_Copper_cyto pos=3355.4583,1069.2395
# _attributes metabolite_Ethylene pos=3047.2717,1576.6065 queryset=end
# _attributes metabolite_Isochorismate pos=8982.9784,1272.4562
# _attributes metabolite_JA pos=5179.4823,2428.0947 queryset=end
# _attributes metabolite_JA-Ile pos=5521.2873,2428.6930
# _attributes metabolite_L-methionine pos=2873.7460,686.8430
# _attributes metabolite_Linolenic_acid pos=5182.3691,925.4869
# _attributes metabolite_Me-JA pos=5179.6149,2534.9874
# _attributes metabolite_OPC-8:0-CoA pos=5189.1866,1694.5881
# _attributes metabolite_OPC4 pos=5179.4073,2237.7986
# _attributes metabolite_OPC6 pos=5189.4848,1808.9652
# _attributes metabolite_OPC8 pos=5191.8678,1588.6824
# _attributes metabolite_OPDA pos=5191.7220,1456.9093
# _attributes metabolite_OPDA_chl pos=5191.6693,1343.8905
# _attributes metabolite_Orto-coumaric_acid pos=8446.6786,1517.2597
# _attributes metabolite_Phenylalanine pos=8460.3831,1229.7144
# _attributes metabolite_Phenylpyruvate pos=8457.3669,1073.7248
# _attributes metabolite_Prephenate pos=8454.0794,906.3837
# _attributes metabolite_ROS pos=9845.8537,785.7078
# _attributes metabolite_SA pos=8795.1344,1926.8119 queryset=end
# _attributes metabolite_SAG pos=9017.0205,2044.4030
# _attributes metabolite_SAM pos=2872.9329,868.0667
# _attributes metabolite_SA_chl pos=8990.9766,1499.2868
# _attributes metabolite_SGE pos=8530.6824,2079.4437
# _attributes metabolite_Trans-cinnamic_acid pos=8453.1899,1397.6368
# _attributes protein_AAO4 pos=8617.8622,1514.4798
# _attributes protein_ACO pos=2747.2642,1599.6108
# _attributes protein_ACO-like pos=2806.0253,1677.6604
# _attributes protein_ACO1 pos=2780.9366,1356.7406
# _attributes protein_ACO2 pos=2752.0127,1471.7158
# _attributes protein_ACO4 pos=2725.5358,1543.9806
# _attributes protein_ACS1 pos=2986.9718,1001.6235
# _attributes protein_ACS10 pos=2642.4807,1063.7428
# _attributes protein_ACS11 pos=2657.7706,1004.1143
# _attributes protein_ACS2 pos=2996.3478,1066.0923
# _attributes protein_ACS3 pos=3016.0225,1129.2945
# _attributes protein_ACS4 pos=3020.1319,1214.3799
# _attributes protein_ACS5 pos=3004.1981,1272.1508
# _attributes protein_ACS6 pos=3009.5820,1344.7938
# _attributes protein_ACS7 pos=2630.5141,1267.7632
# _attributes protein_ACS8 pos=2623.3126,1195.8871
```

```
# _attributes protein_ACS9 pos=2629.0142,1142.0691
# _attributes protein_ACX1 pos=4601.8755,2141.2370
# _attributes protein_ACX2 pos=4733.6969,2167.3010
# _attributes protein_ACX3 pos=4862.1138,2172.3164
# _attributes protein_ACX4 pos=4974.7792,2156.7464
# _attributes protein_ACX5 pos=5103.0422,2129.6433
# _attributes protein_ACX6 pos=5253.9508,2145.3852
# _attributes protein_AIM1 pos=5408.4666,2131.2717
# _attributes protein_AOC1 pos=5501.6783,1230.2696
# _attributes protein_AOC2 pos=5499.0343,1287.1030
# _attributes protein_AOC3 pos=5500.2334,1344.1796
# _attributes protein_AOC4 pos=5500.3281,1399.3646
# _attributes protein_AOS pos=5391.6082,1207.2061
# _attributes protein_Arogenate_dehydratase pos=8214.7251,1225.5675
# _attributes protein_BA2H pos=9263.1309,1531.9092
# _attributes protein_CLH1 pos=6507.8310,693.5341
# _attributes protein_CM1 pos=8257.8162,824.4890
# _attributes protein_CM2 pos=8257.3195,956.5955
# _attributes protein_CM3 pos=8255.0435,884.4115
# _attributes protein_COI1 pos=5663.2485,2561.6248
# _attributes protein_CTR1 pos=3571.5217,1422.0962
# _attributes protein_Catalase pos=10058.7687,804.5395
# _attributes protein_EBF1 pos=3980.0435,1025.9545
# _attributes protein_EBF2 pos=3831.2808,1247.5160
# _attributes protein_EDF1 pos=4350.2870,1766.8882
# _attributes protein_EDF2 pos=4343.7037,1835.2933
# _attributes protein_EDF3 pos=4348.7575,1895.4870
# _attributes protein_EDF4 pos=4336.1263,1964.6902
# _attributes protein_EDS1 pos=9536.7463,935.5799
# _attributes protein_EDS5 pos=9358.7139,1144.0689
# _attributes protein_EIL1 pos=3785.2377,1748.6973
# _attributes protein_EIL2 pos=3785.3721,1823.0804
# _attributes protein_EIN2 pos=3509.9641,1767.6432
# _attributes protein_EIN3 pos=3793.1832,1925.2074
# _attributes protein_EIN4 pos=3676.0680,1226.9932
# _attributes protein_EIN5 pos=3787.6727,1024.8725
# _attributes protein_ERF1 pos=4346.0997,1697.6237
# _attributes protein_ERS1 pos=3434.6794,1258.3170
# _attributes protein_ERS2 pos=3555.0632,1255.1498
# _attributes protein_ETR1 pos=3167.1975,1253.2617
# _attributes protein_ETR2 pos=3289.3931,1253.1866
# _attributes protein_GST1 pos=4480.9400,1257.0682
# _attributes protein_Glutathione_peroxidase pos=9992.3760,657.0318
# _attributes protein_HRT pos=9671.5404,561.1346
# _attributes protein_ICS1 pos=9280.3357,1268.4603
# _attributes protein_ICS2 pos=9399.1534,1264.7741
# _attributes protein_IPL pos=8861.0291,1268.6910
# _attributes protein_Inactive_BA2H pos=9444.7441,1467.1812
# _attributes protein_Inactive_EIN2 pos=3278.0744,1765.9903
# _attributes protein_Inactive_HRT pos=9896.2585,610.1956
# _attributes protein_Inactive_MPK3 pos=9067.9004,654.0412
# _attributes protein_Inactive_MPK6 pos=9506.2357,674.2092
```

```
# _attributes protein_Inactive_et_receptor pos=3602.3440,1057.5273
# _attributes protein_JAR1 pos=5521.3226,2329.8808
# _attributes protein_JAZ1 pos=6256.4561,2238.8808
# _attributes protein_JAZ10 pos=7694.3421,2225.4202
# _attributes protein_JAZ11 pos=7829.9184,2226.5545
# _attributes protein_JAZ12 pos=7973.5191,2226.5061
# _attributes protein_JAZ2 pos=6455.6875,2234.2826
# _attributes protein_JAZ3 pos=6629.2943,2237.1244
# _attributes protein_JAZ4 pos=6788.3186,2234.2685
# _attributes protein_JAZ5 pos=6931.8061,2230.9638
# _attributes protein_JAZ6 pos=7082.9100,2228.2605
# _attributes protein_JAZ7 pos=7234.7819,2225.3384
# _attributes protein_JAZ8 pos=7401.4746,2226.7295
# _attributes protein_JAZ9 pos=7552.8138,2226.9592
# _attributes protein_JMT pos=5046.6098,2534.8157
# _attributes protein_JR1 pos=6326.4442,632.3070
# _attributes protein_KAT1 pos=5543.5522,2129.6971
# _attributes protein_KAT2 pos=5669.8378,2113.6469
# _attributes protein_KAT5 pos=5810.6893,2129.1855
# _attributes protein_LOX1 pos=5421.2494,851.4789
# _attributes protein_LOX2 pos=5425.5459,919.3723
# _attributes protein_LOX3 pos=5427.3340,976.7037
# _attributes protein_LOX4 pos=5426.9799,1038.7540
# _attributes protein_LOX5 pos=5431.0935,1094.9328
# _attributes protein_LOX6 pos=5430.0830,1146.9491
# _attributes protein_MAT3 pos=3067.1080,856.9350
# _attributes protein_MAT4 pos=3066.6798,920.6361
# _attributes protein_MPK3 pos=9080.4522,780.4750
# _attributes protein_MPK4 pos=5932.5236,1014.1104
# _attributes protein_MPK6 pos=9496.6856,777.7182
# _attributes protein_MYC2 pos=7175.7350,646.0556
# _attributes protein_X4 pos=7526.2104,734.2049
# _attributes protein_NADPH_oxidase pos=10013.4012,872.5681
# _attributes protein_NIMIN1 pos=9672.4716,1469.0292
# _attributes protein_NIMIN2 pos=9704.5033,1632.6185
# _attributes protein_NIMIN3 pos=9462.4531,1847.3481
# _attributes protein_NPR1 pos=9791.2031,1292.7221
# _attributes protein_OPCL1 pos=5022.8252,1807.1028
# _attributes protein_OPR3 pos=5028.1407,1585.3193
# _attributes protein_PAD3 pos=9163.0874,906.3816
# _attributes protein_PAD4 pos=9353.2502,935.6032
# _attributes protein_PAL1 pos=8178.0466,1294.5935
# _attributes protein_PAL2 pos=8177.6689,1344.4421
# _attributes protein_PAL3 pos=8178.9558,1398.3397
# _attributes protein_PAL4 pos=8180.8528,1463.9937
# _attributes protein_PDF1.2 pos=4477.4178,1813.3307
# _attributes protein_PR1 pos=10059.8979,1212.1585
# _attributes protein_PR2 pos=10098.8321,1359.1027
# _attributes protein_PR4 pos=4499.3367,1111.4144
# _attributes protein_PR5 pos=10084.7737,1461.5024
# _attributes protein_Prephenate_aminotransferase pos=8227.7854,1068.1744
# _attributes protein_RAN1 pos=3560.7868,892.5952
```

```
# _attributes protein_SAM1 pos=3061.0114,710.8046
# _attributes protein_SAM2 pos=3067.3321,793.6800
# _attributes protein_Superoxide_dismutase pos=10040.8037,727.9978
# _attributes protein_TGA_TF2 pos=9656.9518,1977.4157
# _attributes protein_TGA_TF4 pos=9863.7647,1988.3104
# _attributes protein_TGA_TF5 pos=10029.3701,1960.1179
# _attributes protein_THI2.1 pos=6729.2071,790.8070
# _attributes protein_UGP_glikosyltransferaze pos=8878.5630,2250.5117
# _attributes protein_VSP1 pos=6645.9146,742.4696
# _attributes protein_WRKY70 pos=9940.9593,1273.4311
# _attributes protein_b-CHI pos=4508.7987,954.8370
# _attributes protein_X3 pos=9134.1762,1500.0254
# _attributes protein_X5 pos=5028.8895,1695.4792
# _attributes protein_X1 pos=8456.2843,1589.7962
# _attributes protein_X2 pos=7778.0160,1592.9339
```

## C Biography

Dragana Miljković graduated in Medical and Nuclear Technique at the Faculty of Electrical Engineering, University of Belgrade, Serbia, in 2004. She obtained the Master of Science degree in the field of Information and Communication Technologies in Healthcare, Faculty of Electrical Engineering, University of Belgrade, Serbia, in 2007. In the period 2004-2007 she worked as Systems Developer in the Innovation Centre at University of Belgrade, Serbia, and as a Research trainee in the Biomedical Laboratory at the Faculty of Electrical Engineering, Belgrade, Serbia. From 2007 to 2009 she worked as a Research Assistant at the Faculty of Bioscience Engineering, KU Leuven, Belgium. In 2009 she started her PhD at the Jožef Stefan Postgraduate School, Ljubljana, Slovenia under the supervision of Prof. Dr. Nada Lavrač. Her research concerns modelling of biological responses in plants to stress in collaboration with the National Institute of Biology (NIB), Ljubljana, Slovenia. The results of her research work are published in several journal and conference publications.