

Text processing assisted development of a Petri Net model for plant defence response

Dragana Miljkovic¹, Tjasa Stare², Kristina Gruden², Igor Mozetic¹, Miha Grcar¹, Nada Lavrac¹

¹ Department of Knowledge Technologies, Jozef Stefan Institute, Ljubljana, Slovenia

² Department of Biotechnology and Systems Biology, National Institute of Biology, Ljubljana, Slovenia

[*Dragana.Miljkovic@ijs.si](mailto:Dragana.Miljkovic@ijs.si)

Background

Biologists have been investigating plant defence response to virus attacks for a considerable time. However, individual research groups usually focus their experimental work on a subset of the entire defence system, and the global defence response mechanism in plants is still to be developed. This broader systems approach involves complex biological networks where the knowledge about kinetic reaction parameters is often missing. Therefore, a qualitative type of model needs to be developed.

Materials and methods

In our study, we have manually developed a defence response model represented by the Petri Net formalism. We have used the Cell Illustrator software [1] which is a Hybrid Functional Petri Nets tool. The initial model was developed by manually extracting relevant biological pathways from KEGG and Arabidopsis Reactome databases. As we have realized that the knowledge stored in the databases is not sufficient to build the model for more complex cell mechanisms, such as positive and negative feedback loops, knowledge acquisition was supplemented by knowledge elicitation from biology experts and by comprehensive literature search. This process of manual information retrieval from the literature turned out to be time consuming, non-systematic and error-prone.

To overcome this knowledge acquisition bottleneck, it is necessary to employ more automated methods of extracting relevant information. With the assistance of the GATE tool for natural language processing [2], the biology experts are able to retrieve the missing information from the literature in a more systematic and efficient way. By using GATE we performed pattern-matching on a set of relevant articles. After obtaining names of the substances that are involved in reactions and names of reactions from the biology experts, sets of synonyms are created. Afterwards, the patterns are constructed, highlighted in the documents and manually examined by the experts. The relevant relations extracted from the literature are then manually added to the Petri Net model.

Results

The resulting model (Figure 1) currently contains 52 substances and 41 reactions which correspond, in the Petri Net formalism, to places and transitions, respectively. Figure 1 represents the model of salicylic acid biosynthesis which is one of the key components in signalling the defence response in plants.

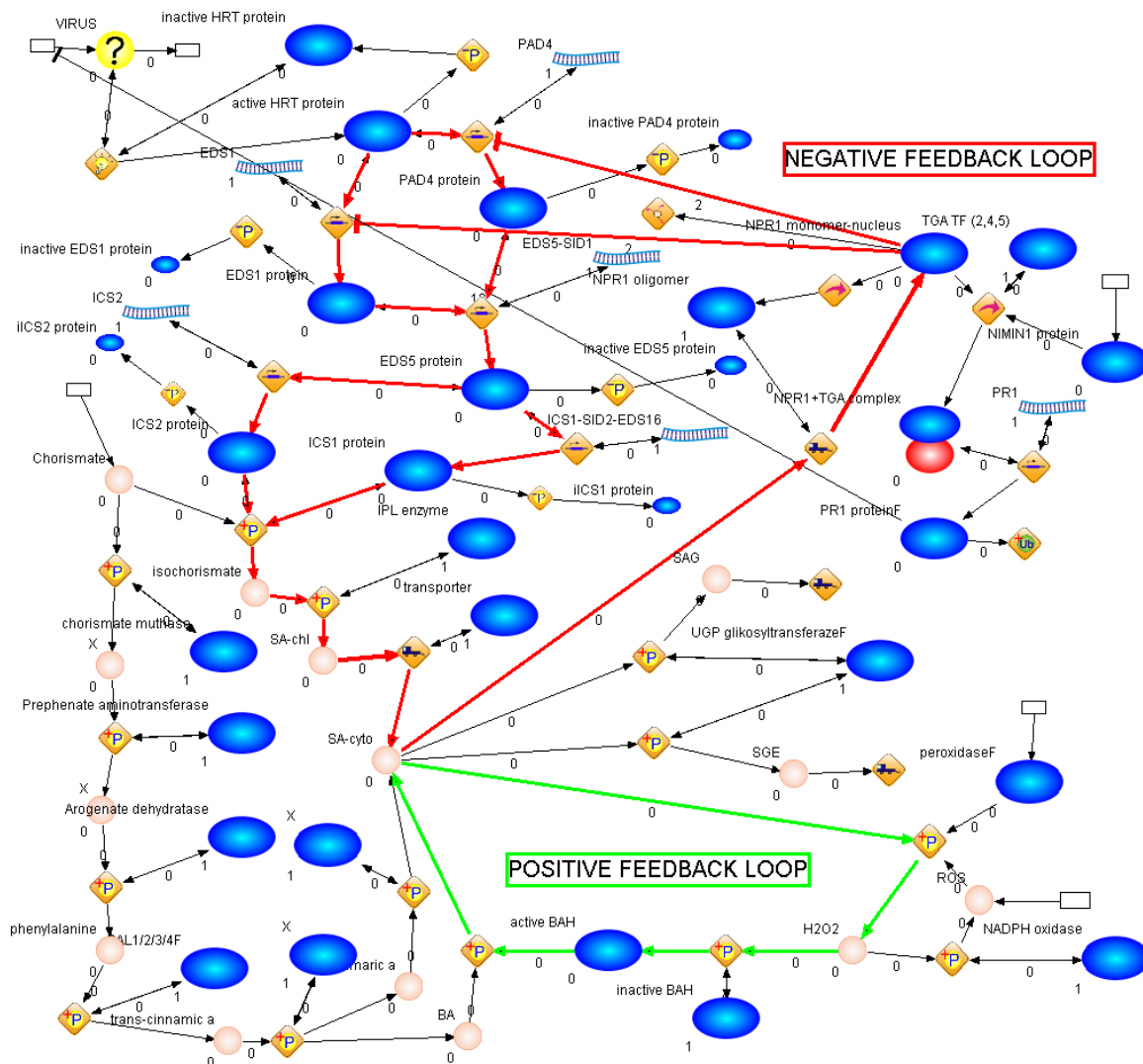
Conclusion

Text processing based approach to the development of a complex Petri Net model provides a more efficient and systematic way of retrieving the necessary information from biological literature. In future work, we plan to extend our text mining library LATINO [3] to make this process semi-automatic.

Acknowledgement

This work is partly supported by the AD Futura scholarship and the Slovenian Research Agency grants P2-0103 and J4-2228.

Figure 1. Petri Net model of salicylic acid biosynthesis in plants. Relations in negative and positive feedback loop, colored red and green respectively, are found by GATE.



References

1. Cell Illustrator 4.0 software [<http://www.cellillustrator.com/>]
2. Natural language processing software GATE [<http://gate.ac.uk/ie/>]
3. LATINO Link Analysis and Text Mining Toolbox [<http://sourceforge.net/projects/latino/>]