Formalization and Automated reasoning about a Complex Signalling Network

Annamaria Basile[1], Maria Rosa Felice[2], Alessandro Provetti[1]

[1] Dept. of Physics - Informatics section, Univ. of Messina. Messina, Italy.
[2] Dept. of Life Sciencies, Univ. of Messina. Messina, Italy.

Abstract. Tran and Baral have proposed an action language (BioSigNet-RR) that is specific for the modeling of signalling networks from Biology and for answering queries relative to the expected response to a stimulus. Translation of their action language to logic programs under Answer Set semantics yields a reasoning mechanisms that gracefully handles incomplete/partial information, updates etc. Those features are extremely important since existing regulatory networks often contain missing or suspected interaction links, or proven interactions whose outputs are uncertain. We present our application experience in developing a BioSigNet-RR formalization of the Signalling network for Arabidopsis Brassinosteroid, a complex interaction that is at the base of growth in some plant species. Such modeling exercise has involved 'filling the gaps' between the terse graphical language of signalling networks literature and the precise specification of the triggering conditions required by BioSigNet-RR. This application experience leads us to propose a new formalization style for action theories representing signaling networks that allows for the description of non-immediate effects of actions. Empirical evaluation of our declarative model has involved formulating and testing several 'what if' queries and checking the quality of the answer with domain experts.

1 Introduction

In Biology, *signalling networks* (also referred to as *signalling pathways)* are specific collections of interactions with a common theme. They are used to provide a summary, working model of the complex interactions that explain how a living cell receives and responds to signals from its environment. Modeling signalling networks is sometimes essential for understanding how cells function and it may lead to effective therapeutic strategies that correct or alter abnormal cell behaviors.

From the point of view of Artificial Intelligence signalling networks represent an interesting form of semi-formal knowledge representation: relevant cellular interactions are to be explicitly described, in a simple graphical language. However, two main issues make modeling signalling networks with action languages challenging:

1. *inhibition,* which we see as a special form of constraint, needs to be explicitly represented and reasoned about,
2. several unspoken assumptions lie in the background as they are assumed to be known by the (expert) reader. For instance, the time element, i.e., a description of the delay between stimulus and reaction, is not explicitly represented yet it is an essential element in reasoning about the long-term evolution of the cell.

This extended abstract reports the results of our work on formalization and automated reasoning with a signalling network that is considered[3]representative of the size, level of detail and complexity of such networks in the Biology literature. The chosen pathway, which is depicted in Figure 1, represents the current knowledge on the interaction that makes the arabdiposis of Brassinosteroids (represented by the lone diamond-shaped node in the pathway) stimulate growth in plants.

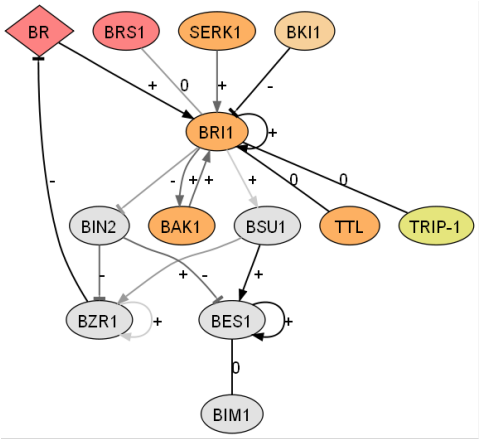

Fig. 1. Signalling network for BR, from the Web site of *Science Signaling.*

Chory et al. [3] presents the pathway and comments in detail each interaction; at this level, we can identify each arc with one relevant interaction. The nodes of the pathway represents disparate physical elements: extra-cellular signals, receptors on the cell's surface or intracellular elements able to *trasduce* signals to the nucleus. Color ans shape of each node guide the expert reader on the actual nature of the element being represented.

From the point of view of knowledge representation and reasoning, the following scientific hypotheses motivate our work.

First, there is a question about the adequacy of BioSigNet-RR [2,8] to support the formalization and reasoning about this specific signalling pathway. As we will see, some specific aspect of the cellular organization is not easily described in terms of BioSigNet-RR primitives.

Second, there is a question of adequacy of our action languages with respect to reasoning about pathway interactions. We consider the following informal test: can we apply BioSigNet-RR to formalize the type of 'what if' questions that an examiner would ask to check a student's assimilation of the material. Hence, we proceed to formulate some easy sample questions and see how the query language part of BioSigNet-RR allows to formulate it.

[3] Paccanaro and Bogre, personal communications.

Finally, our overall hypothesis is that, in the middle term, we should be able to design and implement a vertical solution by which the domain knowledge synthesized in a signalling pathway can be accessed and reasoned about automatically.

2 The representation language

Our modeling effort has adopted the BioSigNet-RR [2,8] language as it now considered the language of reference for reasoning about actions in the Biological domain. Essentially, BioSigNet-RR is an extension of the family of action languages developed by Gelfond and Lifschitz in the 90s; we refer the interested reader to [7] for a survey of the approach. Gelfond and Lifschitz proposed a sorted language, where sorts are actions and fluents, where primitives are the well-known *initially* and *causes* statements. A set of those statements is called an action theory. State is defined in terms of a set of fluents that are deemed true thereof. A declarative semantics is assigned to action theories in terms of trajectories, i.e., an iteration throughout states that the domain in undergoing. At the same time, action languages receive a semantics thanks to translation of action statements to logic programs under answer set semantics [6,1].

When the initial situation is only partially defined, or actions are unknown or even may have non-deterministic effects, alternative answer sets account for the alternative evolutions of the domain. The translation from action languages, including BioSigNet-RR, to logic programs is *modular,* in the sense that it can be done line-by-line by a parser and generator. The translation, which is described in detail in [] has been adopted as is and implemented by a Python-language program derived from Gregory Gelfond's Al2ASP project [4] (see [5] for another application project on the same guidelines).

3 Formalizing the Background knowledge

Signalling pathways are a graphical, synthetic representation of knowledge. However, to fully grasp the dynamics represented by the pathway one often needs to read attentively a natural-language background description that comes with the network.

In our project we have spent a great amount of time to understand and organize the background description of the Arabidopsis Brassinosteroid process given by Chory et al. [3]. The following phrases have been singled out and analyzed separately.

1. In the absence of steroid, BKI1, a plasma membrane-associated protein, interacts directly with the kinase domain of BRI1 to negatively regulate the signalling pathway
2. Binding of BRs to preformed BRI1 homo-oligomers leads to the dissociation of BKI1 from the plasma membrane.
3. It has been proposed that the physical interaction between BRI1 and BAK1 leads to the formation of a signalling-competent hetero-oligomer.
4. The signals transmitted from the plasma membrane-localized BRI1-BAK1 hetero-oligomer negatively regulate the activity of a glycogen synthase kinase 3 (GSK-3), called BIN2.

5. Although the mechanism is as of yet uncharacterized, inactivation of BIN2 leads to the dephosphorylation of BES1 and BZR1, members of a new family of plant-specific transcription factors.
6. BES1, and likely other family members, are further dephosphorylated through the activity of a nuclear-localized, kelch-containing protein phosphatase BSU1.
7. Current data suggest that dephosphorylated BES1 is then able to form homo or heterodimers with the basic helix -loop-helix (bHLH) transcription factor BIM1, to bind to E-box elements in the promoters of BR-regulated genes.
8. Dephosphorylated BZR1 binds to a novel element in the promoters of BR biosynthetic genes to repress their expression.
9. Because BES1 is identical to BZR1, it is expected that BES1 and BZR1 will have both activating and repressing activities.
10. Other proteins have been identified that interact genetically or physically with BRI1, but their precise functions are currently unknown.
11. In vitro and in vivo, BRI1 associates with TTL, transthyretin-like protein.
12. Overexpression of TTL causes slight dwarfing, suggesting that it may play a negative role early in the BR signalling pathway.
13. A suppressor screen using a weak allele of bri1 identified a secreted and active carboxypeptidase, called BRS1, although its molecular target in the BR signalling pathway is unknown.
14. Biochemical studies identified TRIP-1 as a BRI1 interactor; however, the knockdown in expression of TRIP family members yields a pleiotropic phenotype that is slightly reminiscent of those observed for BR biosynthetic or signalling mutants.

It relatively easy to associate each phrase to one of the arcs represented by the signalling pathway. Such association, however, is not always straightforward and will be further commented upon.

It must be pointed out that in our analysis we have discovered that the interaction which is represented by arc:

$$BR1 \overset{+/-}{\Longrightarrow} BAK1$$

is not found in the pathway depicted in Chory et al. [3], from where our work started, but is in Figure 1, which was later found on the Web site of the *Science Signaling*[4] journal.

3.1 Formalization of the Signalling Pathway for BR

The formalization of the pathway proceeds as follows. For each named cellular component, e.g., BR, we introduce two fluents[5]: *high(br)* e *low(br)*. Then, we introduce two *activation* actions: *activate(br)* and *inactivate(br)*, where the latter is an *inhibition* action that, in some sense, depresses BR.

[4] `http://www.sciencemag.org`

[5] The labels used in the signalling pathway are in lowercase, since they are constant names in the domain description.

Next, the remaining arcs are formalized, by coupling each arc to the illustrative phrase found in the description. For instance, the arc connecting BRI1 to BAK1 is connected to the phrase:

"BRI1 interacts directly with BAK1 [through a phosphorylation process]."

$$activate(bak1) \textbf{ causes } up(bri1) \tag{1}$$

"BKI1 interacts directly with the kinase domain of BRI1 to negatively regulate the signalling pathway."

$$high(bki1) \textbf{ inhibits } activate(bri1) \tag{2}$$

"Binding of BR to preformed BRI1 homo-oligomers lead to the dissociation of BKI1 from the plasma membrane."

$$binding(br, bki1) \textbf{ causes } dissociated(bki1) \textbf{ if } high(bri1) \tag{3}$$

Even though it looks like the description of a local, direct interaction, this formalization may be the most effective in capturing the rationale of the pathway. An alternative formulation, which has hitherto not been tested is the following:

$$high(br)\ high(bri1) \textbf{ triggers } dissociated(bki1)$$

"A suppressor screen using a weak allele of BRI1 identified a secreted and active carboxypeptidase, called BRS1, although its molecular target in the BR signaling pathway is unknown. It has been proposed that the physical interaction between BRI1 and BRS1 leads to the formation of a signaling-competent hetero-oligomer."

$$high(brs1) \textbf{ triggers } activate(bri1) \tag{4}$$

"In vitro and in vivo, BRI1 associates with TTL, transthyretin-like protein. Overexpression of TTL causes slight dwarfing, suggesting that it may play a negative role early in the BR signaling pathway."

$$high(bri1) \textbf{ triggers } downregulate(ttl) \tag{5}$$

"Biochemical studies identified TRIP-1 as a BRI1 interactor."

$$high(bri1) \textbf{ triggers } activate(trip1) \tag{6}$$

"The signals transmitted from the plasma membrane-localized BRI1-BAK1 heterooligomer negatively regulate the activity of a glycogen synthase kinase 3 (GSK-3), called BIN2."

$$high(bri1),\ high(bak1) \textbf{ inhibits } activate(bin2) \tag{7}$$

"Although the mechanism is as yet uncharacterized, inactivation of BIN2 leads to the dephosphorylation of BES1 and BZR1, members of a new family of plant-specific transcription factors."

$$inactivate(bin2) \textbf{ causes } low(bzr1) \tag{8}$$

$$inactivate(bin2) \textbf{ causes } low(bes1) \tag{9}$$

"Dephosphorylated BZR1 binds to a novel element in the promoters of BR biosynthetic genes to repress their expression."

$$high(bzr1) \textbf{ triggers } activate(br) \tag{10}$$

"Current data suggest that dephosphorylated BES1 is then able to form homo-or hetero-dimers with the basic helix-loop-helix (bHLH) transcription factor BIM1, to bind to E-box elements in the promoters of br-regulated genes."

$$high(bes1) \textbf{ triggers } activate(bim1) \tag{11}$$

$$high(bsu1) \textbf{ triggers } activate(bes1) \tag{12}$$

$$low(bsu1) \textbf{ inhibits } activate(bes1) \tag{13}$$

$$high(bri1) \textbf{ triggers } activate(bsu1) \tag{14}$$

$$high(serk1) \textbf{ triggers } activate(bri1) \tag{15}$$

It should be noticed again that in Chory et al. [3] the textual description seems not aligned to the graphics of the pathway. As a result, we tentatively interpret the inhibition from BRI1 to BIN2 with:

$$high(bri1) \textbf{ inhibits } (bin2) \tag{16}$$

3.2 Connecting actions to fluents

For each fluent we have had to introduce a couple of actions that represent the *upregulation* and *downregulation* of the fluent itself. Hence, we need to introduce the following two schematic rules, to be instantiated to each fluent:

$$activate(C) \textbf{ causes } high(C) \tag{17}$$

$$downregulate(C) \textbf{ causes } low(C) \tag{18}$$

In a sense, these axiom schemata naturally complement the action theory by capturing the essence of the $+/-$ labeling of the arcs. However, they introduce an extra level of complexity in the representation, since new conditions must be devised to disallow these definitions for specific values of *C,* e.g., *activate(bak1),* that do not have a direct Biological interpretation.

4 Query Formulation and informal validation of the model

To assess the adequacy of the representation language and of our specific action theory we have considered the following classroom scenario: questions about the Arabdiposis brassinosteroid process that a teacher would use to check whether her students have properly learn the material and are now able to reason about Brassinosteroids and their effects. Such questions were formulated with the goal of stressing the connections between the several cellular components of the cell.
To illustrate how natural BioSigNet-RR queries are, we now list, for each question, the expected answer in English, paired with BioSigNet-RR that captures, to some extent, the question itself.

Q: Looking at the pathway, how does *BR* affect the cell?
A: *BR* causes the activation of *BRI1* and *BAK1*, which, in turn, inhibit the activation of *BIN2.*

Our formalization is based on three distinct queries:

$$? - high(bri1) \textbf{ after } activate(br) \tag{19}$$

$$? - high(bak1) \textbf{ after } activate(br) \tag{20}$$

$$? - low(bin2) \textbf{ after } activate(br) \tag{21}$$

Q: What are the effects of activation of *BAK1*?
A: *BAK1* brings about activation of *BR1*; subsequently, *BR1* shall affect the whole cell network.

This question can be translated directly in the following formula (query):

$$? - high(bri1) \textbf{ after } activate(bak1) \tag{22}$$

Q: What are the effects of inactivation of *BIN2*?
A: inactivation of *BIN2* shall cause the inhibition of *BZR1* and *BES1.*

Again, me must resort to two separated queries:

$$? - low(bzr1) \textbf{ after } activate(bin2) \tag{23}$$

$$? - low(bes1) \textbf{ after } activate(bin2) \tag{24}$$

5 Conclusions

The formalization and deployment project described in this article can be considered successful from the point of view of assessing what can be done with action languages (and Logic Programming in general) in the context of Biological knowledge representation and automated reasoning. The overall Artificial intelligence goal, i.e. to have computers process the meaning synthesized in a signalling pathway without human intervention, is yet to be achieved, as our formalization had to deal with a time-consuming human analysis of the accompanying textual explanation, often a heavy-going technical explanation.

BioSigNet-RR has shown to be an ideal platform for formalization in this domain. However, we believe that more research is needed in order to have the action theory match the pathway. One problem that was, in our opinion, only partially solved here is that *action* is understood in two ways. The first understanding, which is probably what Gelfond-Lifschitz meant, is that of *external intervention,* i.e., alteration of the state of affairs. The second understanding, which we suspect is more frequent in Biology, is that an action may also be an alteration, called *upregulation* or downregulation, of a component of the cell. For this second understanding, a specific action sort should be devised and introduced to the formalization language.

Acknowledgments

We are grateful to Alberto Paccanaro and Laszlo Bogre at Royal Holloway, University of London for providing us careful advice and guidance in the Biological and Bioinformatics aspects of this work. Many thanks to Gregory Gelfond for providing us constant advice and support on our work to extend his AL2ASP program.

References

1. Baral, C.: Knowledge Representation, Reasoning and Declarative Problem Solving. Cambridge University Press (2003)
2. Baral, C., Chancellor, K., Tran, N., Joy, A.M., Berens, M.E.: A knowledge based approach for representing and reasoning about signalling networks. In: ISMB/ECCB (Supplement of Bioinformatics). pp. 15–22 (2004)
3. Chory, J., Belkhadir, Y., Wang, X.: Arabidopsis brassinosteroid signalling pathway. Science Signaling 364, cm5 (2006)
4. Gelfold, G.: From AL to ASP - the system al2asp. Tech. rep., Arizona State University (2011)
5. Gelfold, M., Inclezan, D.: Yet another modular action language. In: Proc. of Int'l Workshp on Software Engineering for Answer Set Programming. pp. 64–78. CEUR WS Proceedings, vol. 546. (2009)
6. Gelfond, M., Lifschitz, V.: Classical negation in logic programs and disjunctive databases. New Generation Comput. 9(3/4), 365–386 (1991)
7. Gelfond, M., Lifschitz, V.: Action languages. Electron. Trans. Artif. Intell. 2, 193–210 (1998)
8. Tran, N., Baral, C.: Hypothesizing about signalling networks. Journal of Applied Logic 7(3), 253–274 (2009)