

Completing signaling networks by abductive reasoning with perturbation experiments

Adrien Rougny¹, Yoshitaka Yamamoto^{2,3}, Hidetomo Nabeshima²,
Gauvain Bourgne⁴, Anne Poupon⁵, Katsumi Inoue⁶, Christine Froidevaux¹

¹ Laboratoire de Recherche en Informatique, CNRS, Université Paris-Sud

² University of Yamanashi

³ JST Presto

⁴ LIP6, CNRS, Université Pierre et Marie Curie

⁵ BIOS group, INRA, CNRS

⁶ National Institute of Informatics

Abstract. We propose in this paper an abductive method that allows, given a biological network, to check whether some experimental observations can be causally explained by the network and to use the unexplained observations to complete the network. We then apply the method on the FSHR-induced signaling network.

1 Introduction

Signaling networks model the flow of information occurring in cells after they have been stimulated by an extracellular signal (for instance, an hormone). Together with the rise of available high-throughput data, such networks become always larger and more complex. Consequently, automatic methods have become necessary for their analysis. Methods relying on discrete formalisms, such as logical ones, seem well suited as numerical parameters are often difficult to obtain.

One fundamental task of cell signaling biology is to test whether available experimental data can be explained by a given signaling network, and else, to modify the network (by adding or removing an edge) or clean the data so that it can be explained. Methods to accomplish this task take as input a representation of signaling networks called interaction graph (IG). In IGs, nodes are molecules (or activities that originate from molecules) and arcs are positive or negative influences that molecules have on each other.

These various works mainly differ on four aspects: (i) The semantics used to interpret IGs: In [11] and [14, 12], authors interpret the arcs of an input IG under the path semantics, introduced in [13], or under the Sign Consistency Model [15], respectively, whereas authors of [4, 10] interpret IGs as causality networks, which implies the use of a more constrained semantics. Authors of [14] also consider the problem within the boolean network semantics. (ii) The experimental data they take into account: Methods of [11, 4] take as input steady-state shift experiments, whereas methods of [14, 12, 10] take into account perturbation experiments. (iii) The modifications of the network or the cleaning of the data

they propose to explain unexplained experimental results: The methods in [11, 12] provide possible modification of the input network or the data, whereas the method in [4, 10] allows only the completion of the input network by addition of edges. (iv) The formalism they use: graph theory [14], integer linear programming [12], answer set programming [11] and first-order logic [4, 10].

In this work, we propose a method to check whether a set of perturbation experiments can be explained by a signaling network represented in SBGN-AF [16], and else, to complete the network by adding edges to the network.

SBGN-AF is a standard to represent signaling and gene regulation networks. It extends the classical IG representation by including logical operators (the AND and the OR operator) that permit to specify logical functions within the graph. Taking as input SBGN-AF maps, we extend the path semantics of [13] by considering such operators in the definition of the paths, that we formalize in first-order logic, based on the translation of SBGN-AF maps into predicates introduced in [8]. We also interpret perturbation experiments making a stronger assumption than in [14, 12, 10], resulting in a more constrained setting (cf. section 3). We perform both the explanation and the completion tasks within the same abductive framework by using the consequence finding method from SOLAR [7].

2 Paths semantics with logical operators

Positive and negative paths of an SBGN-AF map are built by transitive closure of the elementary arcs. We interpret a positive path from an activity A to an activity B (denoted by $stimulates^\bullet(A, B)$) as a possibility to explain an increase (*resp.* decrease) of B by an increase (*resp.* decrease) of A , and a negative path from A to B (denoted by $inhibits^\bullet(A, B)$) as a possibility to explain a decrease (*resp.* increase) of B by an increase (*resp.* decrease) of A . The following axioms allow building positive and negative paths using the influences and the logical operators of an SBGN-AF map.

$$stimulates(A, B) \rightarrow stimulates^\bullet(A, B) \quad (1)$$

$$inhibits(A, B) \rightarrow inhibits^\bullet(A, B) \quad (2)$$

$$stimulates^\bullet(A, B) \wedge stimulates(B, C) \rightarrow stimulates^\bullet(A, C) \quad (3)$$

$$inhibits^\bullet(A, B) \wedge inhibits(B, C) \rightarrow stimulates^\bullet(A, C) \quad (4)$$

$$stimulates^\bullet(A, B) \wedge inhibits(B, C) \rightarrow inhibits^\bullet(A, C) \quad (5)$$

$$inhibits^\bullet(A, B) \wedge stimulates(B, C) \rightarrow inhibits^\bullet(A, C) \quad (6)$$

$$and(O) \wedge input(A, B, O) \wedge stimulates^\bullet(C, A) \wedge stimulates^\bullet(C, B) \rightarrow stimulates^\bullet(C, O) \quad (7)$$

$$and(O) \wedge input(A, B, O) \wedge inhibits^\bullet(C, A) \rightarrow inhibits^\bullet(C, O) \quad (8)$$

$$and(O) \wedge input(A, B, O) \wedge inhibits^\bullet(C, B) \rightarrow inhibits^\bullet(C, O) \quad (9)$$

$$or(O) \wedge input(A, B, O) \wedge stimulates^\bullet(C, A) \rightarrow stimulates^\bullet(C, O) \quad (10)$$

$$or(O) \wedge input(A, B, O) \wedge stimulates^\bullet(C, B) \rightarrow stimulates^\bullet(C, O) \quad (11)$$

$$or(O) \wedge input(A, B, O) \wedge inhibits^\bullet(C, A) \wedge inhibits^\bullet(C, B) \rightarrow inhibits^\bullet(C, O) \quad (12)$$

Axioms (1-6) are the main transitivity axioms, while axioms (7-9) and (10-12) express the semantics of the AND and the OR logical operators, respectively.

3 Formalization of experimental observations

We consider experimental observations that originate from perturbation experiments. Such experiments consist in comparing the rate of an activity a_T between two batches of cells each having received a particular treatment. In the control batch, cells are stimulated by a set of molecules that we denote by \mathcal{IC} . In the experimental batch, cells are first treated with a set of inhibitors, that we denote by \mathcal{KO} , and then stimulated by \mathcal{IC} . We introduce a variable e that takes the value \downarrow (*resp.* \uparrow) if and only if (*iff*) the rate of a_T is lower (*resp.* higher) in the experimental batch than in the control batch. We denote such an experimental observation by the tuple $(\mathcal{IC}, \mathcal{KO}, a_T, e)$.

If, for a given experimental observation $\mathcal{E} = (\mathcal{IC}, \mathcal{KO}, a_T, e)$, $e = \downarrow$ (*resp.* $e = \uparrow$), then a_T is more (*resp.* less) inhibited or less (*resp.* more) stimulated by the initial conditions \mathcal{IC} in the experimental batch than in the control batch due to the suppression of at least one activity of \mathcal{KO} . In the cells of the experimental batch, all activities of \mathcal{KO} are suppressed, meaning that they can no longer be performed by the cells. Consequently, $e = \downarrow$ (*resp.* \uparrow) *iff* there is a lower overall stimulation (*resp.* inhibition) of a_T caused by suppressing at least one positive (*resp.* negative) path from \mathcal{IC} to a_T . Thus, $e = \downarrow$ (*resp.* \uparrow) *iff* there exists at least one positive (*resp.* negative) path outgoing from a node belonging to \mathcal{IC} , incoming to a_T , and passing through a node belonging to \mathcal{KO} . The assumption that the suppression of the nodes belonging to \mathcal{KO} does not affect a_T but rather the pathways between nodes of \mathcal{IC} and a_T is the main difference compared to [14, 12, 10], where the initial conditions are not taken into account.

To explicitly describe the role of \mathcal{IC} , we add one virtual activity node $a_{\mathcal{IC}}$ to the prior network so that for each activity $a \in \mathcal{IC}$, we add a stimulation arc from $a_{\mathcal{IC}}$ to a . According to our interpretation of perturbation experiments and our transitivity axioms, we can formalize each experimental observation $\mathcal{E} = (\mathcal{IC}, \mathcal{KO}, a_T, e)$ as the following disjunction $O_{\mathcal{E}}$:

$$\begin{aligned}
 O_{\mathcal{E}} &= \bigvee_{a_{\mathcal{KO}} \in \mathcal{KO}} \left[\begin{array}{l} (stimulates^{\bullet}(a_{\mathcal{IC}}, a_{\mathcal{KO}}) \wedge inhibits^{\bullet}(a_{\mathcal{KO}}, a_T)) \vee \\ (inhibits^{\bullet}(a_{\mathcal{IC}}, a_{\mathcal{KO}}) \wedge stimulates^{\bullet}(a_{\mathcal{KO}}, a_T)) \end{array} \right] \text{ if } e = \uparrow; \\
 O_{\mathcal{E}} &= \bigvee_{a_{\mathcal{KO}} \in \mathcal{KO}} \left[\begin{array}{l} (stimulates^{\bullet}(a_{\mathcal{IC}}, a_{\mathcal{KO}}) \wedge stimulates^{\bullet}(a_{\mathcal{KO}}, a_T)) \vee \\ (inhibits^{\bullet}(a_{\mathcal{IC}}, a_{\mathcal{KO}}) \wedge inhibits^{\bullet}(a_{\mathcal{KO}}, a_T)) \end{array} \right] \text{ if } e = \downarrow.
 \end{aligned}
 \tag{13}$$

Given an SBGN-AF map \mathcal{S} and an experimental observation $\mathcal{E} = (\mathcal{IC}, \mathcal{KO}, a_T, e)$, we want to check if \mathcal{E} can be explained by \mathcal{S} . If not, we want to find a minimal set of arcs that complete \mathcal{S} in order to explain \mathcal{E} . Both tasks can be realized within the same abductive setting, presented hereafter.

4 Abductive setting for the completion task

Let \mathcal{S} be an SBGN-AF map and $\mathcal{E} = (\mathcal{IC}, \mathcal{KO}, a_T, e)$ be an experimental observation. Let B be the background theory formed of the translation of \mathcal{S} into

predicates and axioms (1-12), and $O_{\mathcal{E}}$ be the observation formalized from \mathcal{E} . Then, solving both the explanation and the completion task consists in searching for all minimal hypotheses H such that

$$B \cup H \models O_{\mathcal{E}} \text{ and } B \cup H \not\models \square. \quad (14)$$

If $B \models O_{\mathcal{E}}$, then clearly \mathcal{S} explains \mathcal{E} .

For the computation of H , we can use the consequence finding system SOLAR [7], that allows to restrict the search space by means of the language bias \mathcal{P} describing the negations of desirable hypotheses, and seek for all the *subsumption-minimal* hypotheses belonging to \mathcal{P} . In the completion task, every added influence is either a stimulation or an inhibition. Besides, we restrict the number of added influences to at most two for each observation in order to get more realistic hypotheses that could be tested experimentally. Then, \mathcal{P} is given under the form $\{\neg\textit{stimulates}(-, -), \neg\textit{inhibits}(-, -)\}$, $Length \leq 2$, where $Length$ is the number of literals (*i.e.*, instances of $\neg\textit{stimulates}(-, -)$ and $\neg\textit{inhibits}(-, -)$) allowed in the hypothesis.

In general, SOLAR can produce a large amount of hypotheses. To reduce it, we perform a first selection that operates directly at the generation step or during a post-filtering step. We do not consider hypotheses that generate a loop in the prior map and those that contain constants mapped to logical operators or the constant a_{IC} . We then use a greedy algorithm to select hypotheses based on the decreasing number of experimental observations they can explain.

5 Application: the FSHR-induced network

We applied our method to two pathways of the FSHR-induced signaling network, namely the G protein pathway and the PI3K pathway taken from [1] (See Fig. 1). We built a dataset of 29 experimental observations by gathering and formalizing reliable experimental results from the literature related to the FSHR. For each experiment, only one activity suppressor was used. Consequently, for each experimental observation, the set \mathcal{KO} is merely a singleton.

Among the 29 different experimental observations, 17 observations could be explained by the network, and the 12 remaining ones were used to complete the network. For each of them we computed minimal hypotheses sufficient to explain it when added to the network. We ran SOLAR (ver. 2) with 12 machines (Intel Xeon E-1230 V2 (3.3GHz) and 8GB RAM) in parallel, with a limited executing time of 4 hours.

Each of the 12 observations could be explained by hypotheses containing a unique influence, although more complex hypotheses were also generated. Consequently, we chose to focus on the hypotheses containing a unique influence. Using our greedy algorithm, we ranked more than 250 hypotheses containing a unique influence generated during the abduction phase, and selected 28 among them. Results are shown in Table 5.

	1	2	3	4	5	6	7
	stimulates(p38mapk,pi3k)						
	inhibits(p38mapk,raf1)						
	inhibits(p38mapk,mek)						
	inhibits(p38mapk,erk12)						
	inhibits(rps6,rap1)						
	inhibits(mtor,camp_epac)						
	inhibits(rps6,camp_epac)						
1	{(camp, epac), pi3k, akt, ↓}						
2	{(camp, epac), pi3k, rps6, ↓}						
3	{(camp, epac), p38mapk, akt, ↓}						
4	{(camp, epac), pi3k, p70s6k, ↓}						
5	{(camp, epac), pi3k, rps6, ↓}						
6	{(fsh-fshr, epac), pka, akt, ↑}						
7	{(fsh-fshr, epac), p38mapk, akt, ↓}						
8	{(camp, epac), pka, p70s6k, ↑}						
9	{(fsh-fshr, epac), p38mapk, erk12, ↓}						
10	{(fsh-fshr, epac), camp_epac, erk12, ↓}						
11	{(fsh-fshr, epac), mek, p38mapk, ↓}						
12	{(camp, epac), pka, akt, ↑}						

Table 1: Application to the FSHR-induced network. Lines correspond to experimental observations, columns to selected hypotheses. A cell is green if the hypothesis explains the observation. Experimental observations that are explained by the network, as well as hypotheses (8-28) are omitted. *decrease* by ↓.

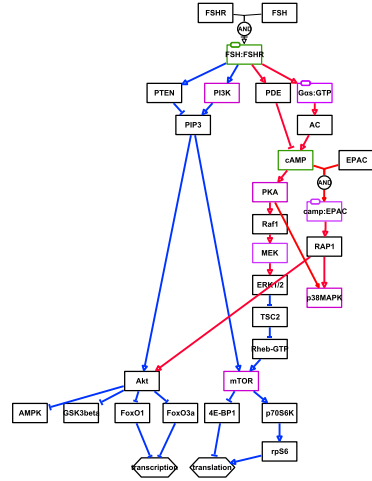


Figure 1: The FSHR-induced network, represented in SBGN-AF. The G protein pathway is represented in red and the PI3K pathway in blue.

Hypothesis (1) proposes that p38MAPK could activate PI3K. In [3], the authors make the hypothesis of such a crosstalk in Granulosa Cells. Moreover, activation of Akt in Zn^{2+} -treated cells has been shown to pass through PI3K downstream of p38MAPK [9]. This result shows that p38MAPK is able to trigger the PI3K pathway in Zn^{2+} treated cells, and thus this reinforces our hypothesis for FSH stimulated cell. Hypotheses (2-4) all suggest an inhibitory crosstalk between p38MAPK and the RAF/MEK/ERK pathway. In [5], the authors clearly state that p38MAPK inhibits the RAF/MEK/ERK pathway during muscle differentiation, thus suggesting a potential influence of p38MAPK on the RAF/MEK/ERK pathway. Hypotheses (5-28) all suggest a crosstalk between the pathway downstream of MEK and the cAMP-EPAC pathway. A crosstalk between ERK and cAMP has indeed been evidenced in [6], even if it involves a feedback loop (excluded in our work).

According to our literature review, top ranked hypotheses are more promising than low ranked ones, indicating that selecting hypotheses based on the number of observations they can explain seems to be appropriate. Interestingly, hypothesis (1) would not be generated with the interpretation of perturbation experiments from [14, 12, 10], suggesting that our stronger assumption allows to discover more unknown influences.

6 Concluding remarks

We have proposed a logical formalization of SBGN-AF maps and transitivity axioms that allow to check, given an SBGN-AF map, whether some experimental

observations can be explained by the map, and else to generate hypotheses that complete the map. Application to the FSHR-induced signaling network shows that the method leads to plausible hypotheses, some of which having already been demonstrated in other signaling systems, and thus that it is promising.

References

1. Gloaguen *et al.*: Mapping the follicle-stimulating hormone-induced signaling networks. *Frontiers in endocrinology* 2 (2011)
2. Choi *et al.*: Gonadotropin-stimulated epidermal growth factor receptor expression in human ovarian surface epithelial cells: involvement of cyclic amp-dependent exchange protein activated by camp pathway. *Endocrine-related cancer* 16(1), pp. 179–188 (2009)
3. Gonzalez-Robayna *et al.*: Follicle-stimulating hormone (fsh) stimulates phosphorylation and activation of protein kinase b (pkb/akt) and serum and glucocorticoid-induced kinase (sgk): evidence for a kinase-independent signaling by fsh in granulosa cells. *Molecular Endocrinology* 14(8), pp. 1283–1300 (2000)
4. Inoue, K., Doncescu, A., Nabeshima, H.: Completing causal networks by meta-level abduction. *Machine learning* 91(2), pp. 239–277 (2013)
5. Lee *et al.*: Activation of p38 mapk induces cell cycle arrest via inhibition of raf/erk pathway during muscle differentiation. *Biochemical and biophysical research communications* 298(5), pp. 765–771 (2002)
6. Baillie *et al.*: Phorbol 12-myristate 13-acetate triggers the protein kinase A-mediated phosphorylation and activation of the PDE4D5 cAMP phosphodiesterase in human aortic smooth muscle cells through a route involving extracellular signal regulated kinase (ERK). *Molecular Pharmacology* 60(5), pp. 1100–1111 (2001)
7. Nabeshima, H., Iwanuma, K., Inoue, K.: Solar: a consequence finding system for advanced reasoning. In: *Automated Reasoning with Analytic Tableaux and Related Methods*, pp. 257–263. Springer (2003)
8. Rougny *et al.*: Analyzing sbgn-af networks using normal logic programs. *Logical Modeling of Biological Systems*, pp. 325–361 (2013)
9. Wu *et al.*: p38 and egf receptor kinase-mediated activation of the phosphatidylinositol 3-kinase/akt pathway is required for zn2+-induced cyclooxygenase-2 expression. *AJP-Lung Cellular and Molecular Physiology* 289(5), L883–L889 (2005)
10. Yamamoto *et al.*: Completing sbgn-af networks by logic-based hypothesis finding. In: *Formal Methods in Macro-Biology*, pp. 165–179. Springer (2014)
11. Gebser *et al.*: Repair and Prediction (under Inconsistency) in Large Biological Networks with Answer Set Programming. In: *KR* (2010, April)
12. Melas *et al.*: Detecting and removing inconsistencies between experimental data and signaling network topologies using integer linear programming on interaction graphs. *PLoS computational biology* 9(9), p. e1003204 (2013)
13. Klamt *et al.*: A methodology for the structural and functional analysis of signaling and regulatory networks. *BMC bioinformatics* 7(1), p. 56 (2006)
14. Samaga *et al.*: The logic of EGFR/ErbB signaling: theoretical properties and analysis of high-throughput data. *PLoS Comput Biol* 5(8), p. e1000438 (2009)
15. Siegel *et al.*: Qualitative analysis of the relation between DNA microarray data and behavioral models of regulation networks. *Biosystems* 84(2), p. 153–174 (2006)
16. Mi *et al.*: Systems biology graphical notation: activity flow language level 1. *Nature Precedings* (2009)