Qualitative Reasoning for Understanding the Behaviour of Complex Biomolecular Networks

Ali Ayadi[1,2], Cecilia Zanni-Merk[1] and François de Bertrand de Beuvron[1]
[1]*ICUBE/SDC Team (UMR CNRS 7357)-Pole API BP 10413, Illkirch 67412, France*
[2]*LARODEC Laboratory, Institut Supérieur de Gestion de Tunis, University of Tunis, Rue de la liberté, Bardo 2000, Tunisia*

Keywords: Biomolecular Networks, Dynamical Modelling, Qualitative Reasoning, Qualitative Simulation.

Abstract: Understanding the dynamical behaviour of cellular systems requires the development of effective modelling techniques. The modeling aims to facilitate the study and understanding of the dynamic behaviour of these systems, by the simulation of their designed models. Complex biomolecular networks are the basis of these models. In this paper, we propose a method of qualitative reasoning, based on a formal logical modeling, to qualitatively simulate the biomolecular network and interpret it behaviour over time. The power of our approach is illustrated by applying it to the case study of the autoregulation of the bacteriophage T4 gene 32.

1 INTRODUCTION

In recent decades, the molecular biology has accumulated a sum of knowledge about the details of the molecular mechanisms in cells (Ingalls, 2012). For many years the biological experiments have discovered much knowledge about genes, proteins and metabolites. Indeed, with the development of high-throughput techniques, huge amounts of data has been generated on several levels (Caporaso et al., 2010). We talk about the genomics (the qualitative study of genes), the proteomics (the quantitative study of proteins) and the metabolomics (the quantitative study of metabolites) (Forbus, 1997). A major problem, which was immediately recognised, was to develop mechanisms for analysing these data, interpret and deduce important knowledge.

These advances given their advantages and disadvantages pave the way for a new discipline of molecular biology which is called systems biology. This integrative discipline aims to combine all information (from different levels) in order to understand the processes and behaviours of all cellular components while studying the interactions that take place among them. Indeed, these molecular components interact with each other, thereby forming large networks that are called complex biomolecular networks.

The complex biomolecular network consists of a set of nodes, denoting the molecular components and a set of edges, denoting the interactions among these cellular components. They are considered as systems that dynamically evolve from a state to another so that the cell can adapt itself to changes in its environment.

The key motivation behind this work is to develop a platform to simulate the state changes of the complex biomolecular networks with the hope of understanding and steering their behaviour. This issue has already been addressed in Wu et al. 's research (Wu et al., 2014b), which they introduce and define the transittability of biomolecular as the idea of steering the complex biomolecular network from an unexpected state to a desired state (Wu et al., 2014b).

In this paper, we propose a method of qualitative reasoning. Indeed, biomolecular networks consist of various subnetworks which themselves are composed of several molecular components interacting in their turn with each other, producing a complex global behaviour. Their complexity and large size have prevented a fully quantitative simulation. We consider that qualitative reasoning responds to the complexity of calculating the quantitative reasoning methods, which sometimes are impossible to implement (Fielding and Schreier, 2001).

The rest of the paper is organised as follows. In Section 2, we give some background on biomolecular networks, we discuss our motivations and we define qualitative reasoning. In Section 3, we propose a qualitative reasoning method and detail all its construction steps. In section 4, we enrich and explain this qualitative method with a concrete case study to explain how this technique can be used in practice.

Ayadi, A., Zanni-Merk, C. and Bertrand, F.
Qualitative Reasoning for Understanding the Behaviour of Complex Biomolecular Networks.
DOI: 10.5220/0006065901440149
In *Proceedings of the 8th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2016) - Volume 2: KEOD*, pages 144-149
ISBN: 978-989-758-203-5

2 BACKGROUND AND RELATED WORK

2.1 Biomolecular Networks

The cell is a complex system consisting of thousands of diverse molecular entities (genes, proteins and metabolites) which interact with each other physically, functionally and logically creating a biomolecular network (Karp, 2010; Wu et al., 2014b).

The complexity of the biomolecular network appears by its decomposition into three levels: the genome level models the genetic material of an organism, the proteome level describes the entire set of proteins and the metabolism level contains the complete set of small-molecule chemicals (Wu et al., 2014a; Hayes et al., 1978). Figure 1 depicts these levels.

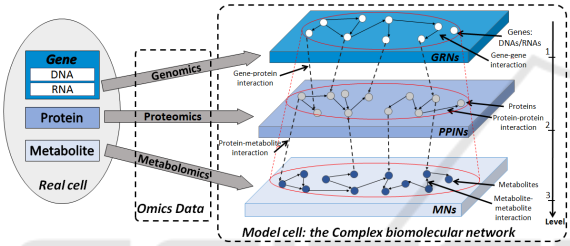

Figure 1: Multi-level modelling of a biomolecular network from a real cell.

Depending on the type of its cellular elements and their interactions, we can distinguish the three basic types of networks: the Gene Regulatory networks (GRNs), the Protein-Protein-Interaction networks (PPINs) and the Metabolic networks (MNs), that were logically and semantically formalized in (Ayadi et al., 2016).

2.2 Qualitative Reasoning

The reasoning is a mental activity that humans practice to solve difficulties they confront in their life. This reasoning is often performed in the lack of quantitative knowledge, which is called qualitative reasoning. In literature, we distinguish two types of reasoning, heuristic reasoning which is effective and the causal reasoning based on a model (Travé-Massuyès, 1997). This reasoning is based on a modeling of the system (model-based reasoning), be it a human being, a machine, etc. Such reasoning is based on a model of causal type because it combines the effects and causes, such as a causal graph. It solves a problem by reasoning about the structure and function of the object in an application environment and their behaviour over time (De Kleer and Brown, 1984; Forbus, 1997).

3 QUALITATIVE REASONING

The explicit representation of the network behaviour evolution, between two instants t_0 and t_n is essential. We must then link the dynamic model defined in (Ayadi et al., 2016) to a qualitative simulation mechanism. This simulation lets to execute the model in order to simulate the network evolution and its components over time.

We chose to use qualitative reasoning for two reasons: (1) To understand the overall functioning and properties of complex biomolecular networks, through the analysis and simulation of the dynamical model (explained in the previous section), and the interpretation of the obtained knowledge. (2) To steer these networks, in particular by allowing to evaluate at any time their simulation in a discrete time.

3.1 Basic Concepts

In the following sections, we will define the basic concepts of qualitative simulation (Travé-Massuyès, 1997) and detail the major phases of construction.

3.1.1 The Causal Graph

The qualitative simulation model is based on the development of a causal graph whose nodes denote variables which this simulation is concerned and edges denote causality relations between these variables. By analogy with the model presented in our previous work (Ayadi et al., 2016), the causal graph is itself the biomolecular network *SR* which its nodes represent causal states of network molecular components and its edges represent the types of interactions that can occur between these components.

3.1.2 Quantitative Variables & Quantity Space

A *variable* is a characteristic of interest. For example, in our case the variables of the qualitative model denote the **state of the molecular components** at a given moment denoted by $en(m,t)$. These variables are qualitative because they are represented by qualities (nominal or ordinal).

The set of these qualitative values and their corresponding intervals constitutes the *quantity space* of the variable $en(m,t)$, denoted by $EQ_{en(m,t)}$. Each variable $en(m,t)$ takes its qualitative value in its ordered set of qualitative values $EQ_{en(m,t)} = \{vq_1, vq_2, ..., vq_n\}$. In fact, the quantity space is a partition of the domain of a variable values into behaviour regions that are qualitatively homogeneous.

To resolve the conflicts of partitioning the $EQ_{en(m,t)}$, we present the following algorithm.

Algorithm 1: Pseudocode of the $EQ_{en(m,t)}$ partitioning algorithm.

Require: $m \in M$, $oe(m)$, min_m, max_m, $EQ_{en(m,t)} \leftarrow \emptyset$
Ensure: Partition of $EQ_{en(m,t)}$
if $(m \in M_P \cup M_M)$ **then**
  for all outgoing edges $i \in oe(m)$ **do**
    Read its $Threshold$;
    Sort the threshold values;
    $Threshold_1 < Threshold_2 <, ..., < Threshold_n$
    Quantitative partitioning of $EQ_{en(m,t)}$;
    $EQ_{en(m,t)} = \{[min_m; Threshold_1], [Threshold_1, Threshold_2], \cdots, [Threshold_n, max_m]\}$
    Translate quantitative measures into qualitative values;
    $EQ_{en(m,t)} = \{vq_1, vq_2, ..., vq_{n+1}\}$
    Where: $vq_1 \equiv [min_m, Threshold_1]$ and $||EQ_{en(m,t)}|| = ||oe(m)|| + 1$
  end for
  Return the quantity space
  $EQ_{en(m,t)} = \{vq_1, vq_2, ..., vq_{n+1}\}$
else
  if $(m \in M_G)$ **then**
    Boolean partitioning of $EQ_{en(m,t)}$;
    $EQ_{en(m,t)} = \{true, false\}$
    Translate boolean measures into qualitative values;
    $EQ_{en(m,t)} = \{vq_1, vq_2\}$
    Where: $vq_1 \equiv 0$, $vq_2 \equiv 1$ and $||EQ_{en(m,t)}|| = 2$
    Return the quantity space
    $EQ_{en(m,t)} = \{vq_1, vq_2\}$
  end if
end if

As defined in Algorithm 1, the partition of the quantity space $EQ_{en(m,t)}$ depends on the type of node:

- If $m \in M_G$: $en(m,t) = \{Deactivated, Activated\}$, its states can be "Activated" or "Deactivated". So, we assign to its $EQ_{en(m,t)}$ the qualitative values 0 and 1 meaning respectively "Deactivated" and "Activated".

$$en(m,t) = \{Deactivated, Activated\} \Rightarrow EQ_{en(m,t)} = \{0,1\}$$

- If $m \in M_P \cup M_M$: $EQ_{en(m,t)}$ is calculated depends on the outgoing arcs $oe(m)$ that can have the node m. In fact, for a quantity m of outgoing arcs, there will be $n+1$ qualitative values that are defined by an order relation on $EQ_{en(m,t)}$, creating an ordered set of qualitative values $EQ_{en(m,t)} = \{vq_1, vq_2, ..., vq_n\}$.

$$en(m,t) = \{[min_m, Threshold_1[, [Threshold_1, Threshold_2 [, \cdots, [Threshold_n, max_m]\} \Rightarrow EQ_{en(m,t)} = \{vq_1, vq_2, \cdots, vq_n\}$$

Figure 2 displays the execution of the $EQ_{en(m,t)}$ partitioning algorithm in both cases. In addition to the quantity space of its variables, a qualitative reasoning method also includes algebraic relations (constraints, influences, etc.) that act among these quantity space.

3.1.3 Operations and Rules

The Operations. In (Travé-Massuyès, 1997), the authors define six operations for calculating the quantity spaces of the variables. Among them, we were just use the three unary operations as shown in Table 1: the *incrementation* (*incr*), the *decrementation* (*decr*) and the *inverse* (*inv*) of a qualitative variable vq_i.

Table 1: Unary operations on quantity spaces presented in (Travé-Massuyès, 1997).

Operations on EQ
Unary operations $\forall [en(m,t)] \in EQ_{en(m,t)} = \{vq_1, vq_2, vq_3, vq_4, vq_5\}$ and $n \in \mathbb{N}$
Incrementation "*incr*" $incr_0([m]) = [en(m,t)]$ $[en(m,t)]$: vq_1 vq_2 vq_3 vq_4 vq_5 - $incr_1([en(m,t)])$: vq_2 vq_3 vq_4 vq_5 vq_5 $incr_n([en(m,t)]) = incr_{n-1}(incr_1([en(m,t)]))$
Decrementation "*decr*" $decr_0([en(m,t)]) = [en(m,t)]$ $[en(m,t)]$: vq_1 vq_2 vq_3 vq_4 vq_5 - $decr_1([en(m,t)])$: vq_1 vq_1 vq_2 vq_3 vq_4 $decr_n([en(m,t)]) = decr_{n-1}(decr_1([en(m,t)]))$
Inverse "*inv*" $[en(m,t)]$: vq_1 vq_2 vq_3 vq_4 vq_5 - $inv([en(m,t)])$: vq_5 vq_4 vq_3 vq_2 vq_1

Using these operators, we can combine several variables together to create our own operations as a specific combination table.

The Partition and Propagation Rules. Based on the work presented in (Travé-Massuyès, 1997), we adapt qualitative reasoning mechanism to calculate the qualitative value of the nodes. This mechanism is based on both the partition rules and the propagation rules (Figure 3). These rules are used to calculate the value of the target variable ($en(m,t+1)$) at the next time $t+1$ based on its qualitative value ($en(m,t)$) and the value of its predecessors ($en(Pred(m),t)$) at the current time t.

- *Partition rules* allow the translation of quantitative measures of the variables ($en(m,t)$) into qualita-

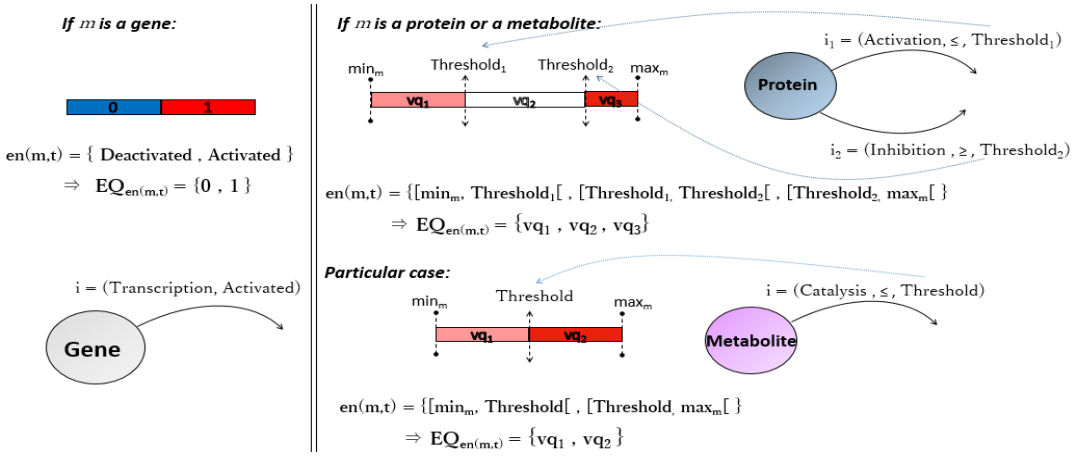

Figure 2: Description of the $EQ_{en(m,t)}$ partitioning algorithm.

tive values. They match a quantitative (real) interval with its correspond qualitative value belonging to the quantity space $EQ_{en(m,t)}$. They are defined by the pseudo code of the algorithm 1.

- *Propagation rules* calculate the propagation of the qualitative values from the sources components to the target components of the causal graph. They are defined by the aggregate functions A_m which calculates the evolution of the node status between two successive instants of the simulation (this function is detailed in (Ayadi et al., 2016)). These rules are expressed by combining the operations presented in Table 1.

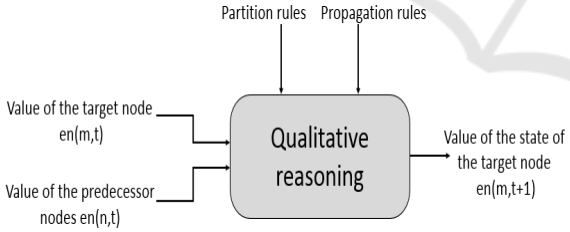

Figure 3: Qualitative reasoning mechanism.

4 AN EXAMPLE

Ribosomal proteins control its own level in the cell by itself. We call them as ribosomal regulatory protein. The gene encoding the regulatory protein is itself a target of the protein which it produces. This is so-called self-regulation or autoregulation when a protein regulates its own production. We distinguish two types of regulations:

- A negative self-regulation when the regulatory protein represses the expression of its own gene (inhibit the expression of its mRNA). This inhibition occurs when there is an accumulation of the protein concentration that exceeds a certain threshold.
- A positive self-regulation, when the regulatory protein activates the expression of its producing gene. This activation occurs when there is a lack of concentration of the protein which becomes less than a certain threshold.

We choose a special case of self-regulating ribosomal protein, the autoregulation of the bacteriophage T4 gene 32. Figure 4 displays the general model of this example. This network consists of a gene G32 coding for a protein p32 and a metabolite m32 which can catalyse the protein p32. Self-regulation of the bacteriophage T4 gene 32 depends on the concentration of the protein p32. Indeed, the concentration of p32 is regulated by itself and normally should remain between $S_{p32} = 0.2\ 10^{-6}\ Mol$ and $S_{p32} = 0.7\ 10^{-6}\ Mol$. However, when the concentration of p32 exceeds the threshold $S_{p32} = 0.7\ 10^{-6}\ Mol$, it inhibits the translation of its gene G32 making it inactive. Similarly, when the concentration of p32 decreases and becomes less than threshold $S_{p32} = 0.2\ 10^{-6}\ Mol$, it activates the translation of its gene G32 making it active. Details of this example can be found in (Lewin and Sanlaville, 1998). For the sake of simplicity, we provide the step-by-step construction of the qualitative simulation by appling it to the autoregulation of the bacteriophage T4 gene 32.

4.1 The Variables

In the example presented in Figure 4, we have three variables $en(G32,t)$, $en(p32,t)$ and $en(m32,t)$ that respectively represent the state of the gene $G32$, the protein $p32$ and the metabolite $m32$.

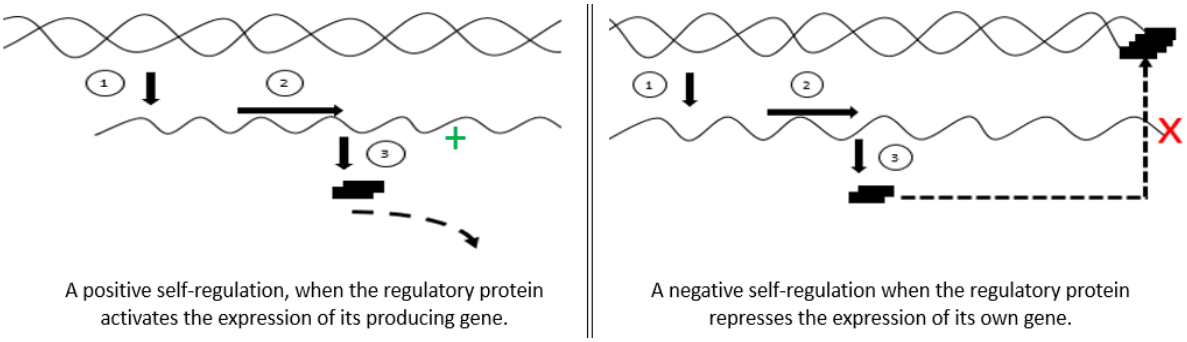

Figure 4: Autoregulation of the bacteriophage T4 gene 32.

4.2 The Causal Graph

We can use the structure of the biomolecular network as the causal graph of our example.

4.3 The Partition Rules

$$en(G32,t) \in \{Deactivated\ ,\ Activated\},$$
$$\Rightarrow EQ_{en(G32,t)} = \{0\ ,\ 1\}.$$
$$en(p32,t) \in \{[min_{p32}, 0.2[\ ,\ [0.2, 0.7[\ ,\ [0.7, max_{p32}[\},$$
$$\Rightarrow EQ_{en(p32,t)} = \{vq_1\ ,\ vq_2\ ,\ vq_3\}.$$
$$en(m32,t) \in \{[min_{m32}, 0.8[\ ,\ [0.8, max_{m32}[\},$$
$$\Rightarrow EQ_{en(m32,t)} = \{vq_1\ ,\ vq_2\}.$$

4.4 The Propagation Rules

For reasons of clarity, we note $[m]^t$ the qualitative value of the state of the component m. It means that the notation $[m]^t \equiv [en(m,t)] \in EQ_{en(m,t)}$. Now, let us define the aggregate rules of each variables.
For the variable $G32$:

$$\begin{aligned}[G32]^{t+1} = A_{G32}([G32]^t, \{i_1, i_2\}, [p32]^t)\\ if\ ([p32]^t = vq_1)\ then\\ \Rightarrow [G32]^{t+1} = 1\\ else\ if\ ([p32]^t = vq_2)\ then\\ \Rightarrow [G32]^{t+1} = [G32]^t\\ else\ if\ ([p32]^t = vq_3)\ then\\ \Rightarrow [G32]^{t+1} = 0\end{aligned}$$

For the variable $p32$:

$$\begin{aligned}[p32]^{t+1} = A_{p32}([p32]^t, \{i_3, i_4\}, [G32]^t, [m32]^t)\\ if\ ([m32]^t = vq_1)\ \wedge\ ([G32]^t = 0)\ then\\ \Rightarrow [p32]^{t+1} = [p32]^t\\ else\ if\ ([m32]^t = vq_2)\ \wedge\ ([G32]^t = 0)\ then\\ \Rightarrow [p32]^{t+1} = decr([p32]^t)\\ else\ if\ ([m32]^t = vq_1)\ \wedge\ ([G32]^t = 1)\ then\\ \Rightarrow [p32]^{t+1} = incr([p32]^t)\\ else\ if\ ([m32]^t = vq_2)\ \wedge\ ([G32]^t = 1)\ then\\ \Rightarrow [p32]^{t+1} = [p32]^t\end{aligned}$$

For the variable $M32$:

$$\begin{aligned}[m32]^{t+1} = A_{m32}([m32]^t)\\ \Rightarrow [m32]^{t+1} = [m32]^t\end{aligned}$$

4.5 The Simulation

Let us define the initial state of the network at t_0: $ER(t_0) = \langle [G32]^{t_0}, [p32]^{t_0}, [m32]^{t_0} \rangle$.
We randomly choose the initial qualitative values of the components as: $ER(t_0) = \langle 0, vq_1, vq_1 \rangle$.
Then, we have performed a series of simulations to assesses the evolution of the network over time:

$$\begin{aligned}ER(t_0+1) &= \langle [G32]^{t_0+1}, [p32]^{t_0+1}, [m32]^{t_0+1} \}\\ &= \langle 1, vq_1, vq_1 \rangle\\ ER(t_0+2) &= \langle [G32]^{(t_0+1)+1}, [p32]^{(t_0+1)+1}, [m32]^{(t_0+1)+1} \rangle\\ &= \langle 1, vq_2, vq_1 \rangle\end{aligned}$$

4.6 The Behaviour

$$\begin{aligned}CR_{[t_0,t_2]} &= \{ER(0), ER(1), ER(2)\}\\ &= \{\langle 0, vq_1, vq_1 \rangle, \langle 1, vq_1, vq_1 \rangle, \langle 1, vq_2, vq_1 \rangle\}\end{aligned}$$

Figure 5 presents the possible simulation results.

5 CONCLUSION AND FURTHER WORK

In this paper, we draw inspiration from the works of (Travé-Massuyès, 1997) to propose a qualitative reasoning method to simulate the behaviour of the

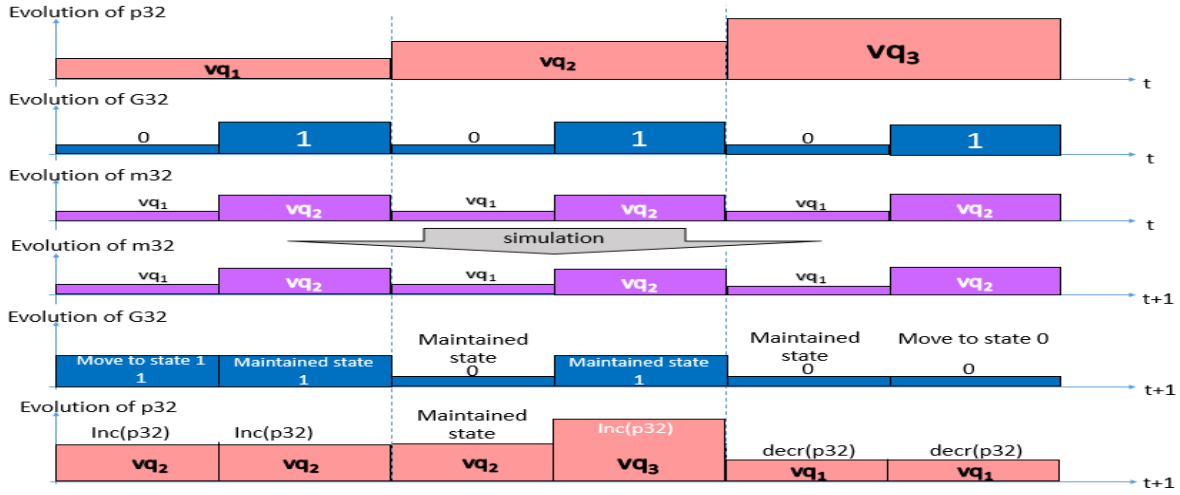

Figure 5: All possible simulation results of our example.

biomolecular network. This method is completely based on the logical formalisation presented in our previously research (Ayadi et al., 2016) that can be assimilated to a causal model.

We have applied our approach to a comprehensive model concerning the autoregulation of the bacteriophage T4 gene 32. In fact, the qualitative reasoning presented here clearly demonstrates all the elements that we need to understand the evolution of biomolecular networks.

Further work includes the translation of this logical formalism into ontologies where qualitative reasoning can be integrated to obtain an optimal model. Simulation of these models along with optimization algorithms will permit to obtain the best external stimuli to be applied to steer the network from its current state to a desired state. These results will be compared with the approach proposed by (Wu et al., 2014b).

REFERENCES

Ayadi, A., Zanni-Merk, C., de Bertrand de Beuvron, F., and Krichen, S. (2016). Logical and semantic modeling of complex biomolecular networks. *Procedia Computer Science*, 96:475 – 484. Knowledge-Based and Intelligent Information & Engineering Systems: Proceedings of the 20th International Conference KES-2016.

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., Fierer, N., Pena, A. G., Goodrich, J. K., Gordon, J. I., et al. (2010). Qiime allows analysis of high-throughput community sequencing data. *Nature methods*, 7(5):335–336.

De Kleer, J. and Brown, J. S. (1984). A qualitative physics based on confluences. *Artificial intelligence*, 24(1-3):7–83.

Fielding, N. and Schreier, M. (2001). Introduction: On the compatibility between qualitative and quantitative research methods. In *Forum Qualitative Sozialforschung/Forum: Qualitative Social Research*, volume 2.

Forbus, K. D. (1997). Qualitative reasoning.

Hayes, P. J. et al. (1978). *The naive physics manifesto*. Institut pour les études sémantiques et cognitives/Université de Genève.

Ingalls, B. (2012). Mathematical modelling in systems biology: An introduction.

Karp, G. (2010). *Biologie cellulaire et moléculaire: Concepts and experiments*. De Boeck Supérieur.

Lewin, B. and Sanlaville, C. (1998). *Gènes VI*. De Boeck Supérieur.

Travé-Massuyès, L. (1997). *Le raisonnement qualitatif pour les sciences de l'ingénieur (coll. diagnostic et maintenance)*. Hermes Science Publications.

Wu, F., Chen, L., Wang, J., and Alhajj, R. (2014a). Biomolecular networks and human diseases. *BioMed research international*, 2014.

Wu, F.-X., Wu, L., Wang, J., Liu, J., and Chen, L. (2014b). Transittability of complex networks and its applications to regulatory biomolecular networks. *Scientific reports*, 4.