

Understanding the Behaviour of Complex Biomolecular Networks by Combining Logical and Semantic Modeling

Ali Ayadi^{1,2}, Cecilia Zanni-Merk¹, and François de Bertrand de Beuvron¹

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

{ali.ayadi,cecilia.zanni-merk,debeuvron}@unistra.fr

Abstract. In literature, most researches related to the understanding of the behaviour of the cell networks focus only on some parts of the biomolecular network. However, to completely understand their behaviour over time we have to integrate the different parts of the biomolecular network and analyse them together.

The objective of the present study is to propose to the biologist a platform to simulate the state changes of biomolecular networks with the hope of steering their behaviours. In this paper, we firstly present an efficient formalism to represent the dynamic behaviour of biomolecular networks. This logical model is based on the three levels of systems theory: structural, functional and behavioural modeling. We then propose a semantic approach based on four ontologies to formalise the domain knowledge of complex biomolecular networks. Both of these approaches provide the necessary elements to model, analyse and understand the dynamic behaviour and the transition states of these networks.

Key words: *Systems biology, complex biomolecular networks, logical modeling, dynamical modeling, semantic technologies.*

1 Introduction

Systems biology is a comprehensive quantitative analysis of the manner in which all the components of a biological system interact functionally over time [1]. Yet, understanding cellular behavioural variability and its evolution over time is one of the most complex tasks facing current researchers. Indeed, with the development of high-throughput techniques such as DNA sequencing, the biological experiments have discovered much knowledge about genes, proteins and metabolites. These advances are enabling researchers to comprehensively integrate the molecular components properties in a powerful framework called the complex biomolecular network. This network consists of a set of nodes, denoting the molecular components and a set of edges, denoting the interactions among these cellular components. These networks are considered as systems that dynamically evolve from a state to another so that the cell can adapt itself to changes in its environment.

Many formalisms have been proposed in recent years for the modeling of biological networks. Among these formalisms, we can list the ordinary differential equations [2], the stochastic methods [3, 4], the boolean networks [5, 6], the Bayesian networks

[7], the Petri nets [8], etc. However, most of these researches focus only on modeling isolated parts of this network, such as the metabolic network or the gene regulatory network and do not study the dynamics of the network as a whole. Moreover, there are many other applications based on semantic technologies have focused on understanding the behaviour of biological networks such as the development of the Systems Biology ontology³, the Sequence ontology⁴, the BIOPax⁵, see also the works of Knüpfer et al. [9], [10].

In this paper, we firstly present a logical approach for modeling the dynamic behaviour of biomolecular networks. This formalism is based on the three levels of analysis of systems theory: structural, functional and behavioural modeling. Second, we propose a semantic approach to formalize the domain knowledge of complex biomolecular networks. This approach based on four ontologies provides the necessary concepts for modeling the dynamic behaviour and the state transitions of these networks.

This paper is organised as follows: Section 2 details the logical formalisation for describing the logical structure, function and behaviour of complex biomolecular networks. Section 3 is devoted to present the semantic approach that provides a rich description for modeling a biomolecular network and its state changes. Finally, in Section 4, we provide conclusions and future works.

2 Formal Modeling of Biomolecular Networks

In this section, we detail the mathematical formalism that will allow to translate, in logical terms and under certain assumptions, the dynamicity of a complex biomolecular network. The first parts of this approach have been presented in [14]. Several improvements to cope with the behavioural aspects are presented here in Section 2.1.

Indeed, a biomolecular network is a dynamical system that can be characterised by the triple $(be, do, become)$ describing its structure, function and behaviour according to the systems theory [11]. This modeling is based on three basic modeling pillars:

- The structural modeling:** to describe the architecture of the biomolecular network.
- The functional modeling:** to describe what can carry out each component of the biomolecular network, specifying the conditions for these activities.
- The behavioural modeling:** to describe how the biomolecular network and its individual components evolve over time.

Thus, the biomolecular network BN can be represented by its *structure* SR , its *function* FR and its *behaviour* $CR_{[t_0, t_n]}$ that evolves over time t (a simulation interval $[t_0, t_n]$). Therefore, mathematically, the biomolecular network BN is defined as follows:

$$BN = (SR, FR, CR_{[t_0, t_n]})$$

For lack of space, we will not detail the *structural* SR and the *functional* FR modeling. We will just remember their basic notation. For more details about these parts, please see our previous work [14].

The *structure* of the biomolecular network $SR = (M, I)$ is a graph defined by:

³ <http://www.ebi.ac.uk/sbo/main/>

⁴ <http://www.sequenceontology.org/>

⁵ <http://www.biopax.org/>

- $M = \{m_1, m_2, \dots, m_n\}$ denotes all the molecules composing the network. M is partitioned as: M_G the set of genes, M_P the set of proteins and M_M the set of metabolites.
- $I = \{i_1, i_2, \dots, i_m\}$ denotes the set of interactions between the network's molecules. Thus, for an edge $i \in I$, we note $s(i)$ the starting node and $d(i)$ the destination node. The partition of the graph nodes induces a partition of the different types of interactions. We have three interactions between molecular components of the same type, four interactions between the nodes belonging to different networks and two interactions I_{GM} and I_{MG} are not taken into account because there is no direct interaction between the genes and metabolites and vice versa.

The partition of M and I are detailed in [14]. Moreover, the *function* of the biomolecular network, denoted by FR , associates to each graph edge $i \in I$ the type of its interaction and the condition that activates it. These types of interactions belongs to the set of concepts of the Interaction Ontology proposed by Van Landeghem et al. [13].

2.1 Behavioural Modeling

Time Discretisation. Complex biomolecular networks are dynamic systems characterised by continuous interactions. The activity of these interactions can be modeled in the form of differential equations. However with the large size of these networks, solving these differential equations in continuous time leads to very significant practical difficulties and heaviness in their implementation.

Thus, in order to model the dynamic evolution of a biomolecular network and reproduce its behaviour over time, we proceed with a behavioural simulation in discrete time. This simulation allows to study the behaviour of biomolecular network through successive transitions. The state of a node (representing a cell component) at the next generation is calculated according to its state and the state of its predecessors at the current generation as well as the possible influence of each one of its incoming edges. The node states evolve synchronously. The results of the simulation will be tested and validated by biologist experts.

State of the Network. The *state of the network* at a given time if defined by a function $en(n, t)$ which associates to each node its state at the moment t .

$$en : \left(m, t \right) \mapsto \begin{cases} Activation \in \{True, False\} & \text{if } m \in M_G. \\ [c_m(t)] \in \mathbb{R} & \text{if } m \in M_P \cup M_M. \end{cases}$$

- For all $m \in M_P \cup M_M$: $en(m, t) = [c_m(t)] \in \mathbb{R}$ where: $c_m(t)$: the value of the concentration of the molecule denoted by the node m at a given time t .
- For all $m \in M_G$: $en(m, t) = Activation$ where: $Activation \in \{True, False\}$. Associating a gene with a concentration is not meaningful. Instead, a gene may have two specific states, activated or not.

We can define $ER(t)$ the state of the biomolecular network at an instant t with a set representing the states of all components in the network at any given time t .

$$ER(t) = \langle en(m_1, t), en(m_2, t), \dots, en(m_n, t) \rangle$$

Transition of the Network State. For a node $m \in M$, we define $ie(m)$ (resp. $oe(m)$) the set of incoming edges (resp. outgoing edges) on m , defined as follows:

$$ie(m) = \{i \mid d(i) = m\} \quad \text{and} \quad oe(m) = \{i \mid s(i) = m\}$$

We also define $Pred(m)$ the set of predecessor nodes on the node m such as:

$$Pred(m) = \{n \in M; \exists i \in I \mid s(i) = n \text{ and } d(i) = m\}$$

The state of a node at time $t + 1$ depends on its state at time t , as well as the possible influence of each one of its incoming edges. This influence obviously depends on the state of the starting node of the arc in question.

For each node m , we define an *aggregate function* A_m (relating to the node m) which calculates the evolution of the node status between two successive instants of the simulation. This *aggregate function* A_m depends on the current state of the node m , the state of its predecessor nodes $Pred(m)$ and the characteristics of its incoming edges $ie(m)$.

$$en(m, t + 1) = A_m(en(m, t), ie(m), en(n, t); n \in Pred(m))$$

Network Steering. A state transition in the network occurs by changing at least one of its nodes. The changes of a node state (that is, changes in the concentration of the molecule) can occur either by an *internal stimulus* (for example, due to reactions that are internal to the cell) already seen with the aggregate function (Section 2.1) or by an *external stimulus* generated outside the cell (for example, because of a medicine taken by the patient). Indeed, a stimulus is an event that can cause changes in the state of the molecule where it operates and therefore to change the state of all the biomolecular network (changing a node automatically modifies other network nodes).

An *external stimulus* S is a triplet $[t, m, \Delta_c]$, where:

- t is the time of introduction of the stimulus S .
- m is the node targeted by the stimulus S .
- Δ_c is the change in concentration caused by the stimulus S and which depends on the type of the node:
 - If $m \in M_G$, Δ_c determines the activation or deactivation of a gene: $\Delta_c \in \{Activated, Deactivated\}$.
 - Else, if $m \in M_P \cup M_M$, Δ_c represents the change of the concentration caused by the stimulus S : $\Delta_c \in \mathbb{R}$.

We denote $ER(t)$, with $t \in \mathbb{N}$, the state of the network at time $T(t) = t_0 + t \cdot \Delta T$ (where ΔT is the time step and t_0 the initial time of the simulation).

To simulate the different transition states of a biomolecular network, we give a state $ER(0)$ at time t_0 and a time step size ΔT . Then the successive states $ER(t + 1)$ are calculated from the current state $ER(t)$ according to the interactions and the aggregate functions defined by the network, and the external stimuli.

At a given time $t + 1$, for each $m \in M$ we have:

- If there are no external stimuli in time t for the node m then:

$$en(m, t + 1) = A_m(en(m, t), ie(m), en(n, t)) \quad \text{where: } n \in Pred(m)$$

– Else

- If $m \in M_G$:

$$en(m, t + 1) = \Delta_c$$

- Else (if $m \in M_P \cup M_M$):

$$en(m, t + 1) = A_m(en(m, t), ie(m), en(n, t)) + \Delta_c \quad \text{where: } n \in Pred(m)$$

Behaviour. The *behaviour* of the biomolecular network $CR_{[t_0, t_n]}$ is given by the sequence of its successive states during the simulation time.

$$CR_{[t_0, t_n]} = [ER(0), ER(1), \dots, ER(n)]$$

Indeed, the behaviour of the network extends between two distinct instants t_0 and t_n forming the simulation interval $[t_0, t_n]$.

3 A Semantic Approach for Analysing the Transittability of Complex Biomolecular Networks

Modeling the behaviour of complex biomolecular networks requires, first and foremost, to formalize the domain knowledge. However, it is not sufficient to simply describe it. Certainly, the behaviour of biomolecular networks is investigated through appropriate semantic structures for the description of their components that must not be overlooked. Thus, the use of a formalized language such as ontologies provides a rich description but also allows to perform reasoning. To do this, we propose a semantic architecture composed of four ontologies: three of them already exist in the literature, the Gene Ontology (GO) [15, 16], the Simple Event Model Ontology (SEMO) [17], the Time Ontology (TO) [19] and we are developing the Biomolecular Network Ontology (BNO). Linked together, these ontologies provide the necessary concepts for modeling the dynamic behaviour and the transition states of a complex biomolecular network.

3.1 The Gene Ontology

In this study, the Gene Ontology⁶ is considered as a core ontology. It ensures the description and the classification of cellular components. It provides a structured terminology for the description of gene functions and processes, and the relationships between these components [20].

We chose to use the Gene Ontology for the following reasons, (1) it is an initiative of several genomic databases such as the Saccharomyces Genome database (SGD), the Drosophila genome database (FlyBase), etc. to build a generic ontology for describing the role of genes and proteins, (2) it is the most developed and most used in biology (since 2000), and (3) it provides annotation files about large number of cellular entities.

⁶ <http://www.geneontology.org>

3.2 The Simple Event Model Ontology

The Simple Event Model ontology⁷ proposed by Van Hage et al. [17] provides the necessary knowledge for the description of events. The ontological architecture of the Simple Event Model ontology consists of four basic classes: *Event* that specifies what is happening, *Actor* that indicates the participants of an event, *Place* that describes the location where the event happened, and *Time* that describes the moment.

We chose to use the Simple Event Model ontology because it provides the necessary concepts to describe and model events in various subject domains.

3.3 The Time Ontology

The Time ontology⁸ developed by Hobbs and Pan [18], [19] enables a more intuitive use of the time dimension while making the most of semantic knowledge. It gives a rich vocabulary to describe the topological relationships that may exist between time points and intervals, and also provides information about time.

The main classes of this temporal ontology can be summarized as *TemporalEntity* which consists of two sub-classes *Instant* and *ProperInterval*, *DurationDescription*, *DateTimeDescription*, *TemporalUnit*, etc. Also, it contains several properties such as *hasDurationDescription*, *intervalStarts*, *hasDateTimeDescription*, etc.

We chose to use the Time Ontology because of its basic structure that is not specific to a particular application and because it is simple to adapt it in our context.

3.4 The Biomolecular Network Ontology

To study the dynamic behaviour and the transition states of a biomolecular network, it is required to model its domain knowledge. Therefore, we developed the Biomolecular Network ontology. This ontology is the major contribution of this paper, it is intended to describe exhaustively the field of complex biomolecular networks by describing the static aspect of its structure. It was defined in collaboration with domain experts.

Figure 1 presents the Biomolecular Network ontology. We use the graphical notation for OWL ontologies defined by Brockmans et al. [22] and Bārzdīņš et al. [23] where boxes are OWL classes; full lines are object properties and dotted lines are data properties. Full lines can be labelled to indicate restrictions meaning that the range of the relationship is specialized. Only a few of the object properties restrictions are displayed in Figure 1 for the sake of clarity. This domain ontology consists of four main classes:

- The class *Biomolecular_Network*: This class includes the different types of complex biomolecular networks. As mentioned earlier in Section 1, the complex biomolecular network can be composed by Gene Regulatory networks (GRNs), Protein-Protein Interaction networks (PPINs) and Metabolic networks (MNs) which correspond to the following concepts: *Genomic_Network*, *Proteomic_Network* and *Metabolomic_Network*.

⁷ <http://semanticweb.cs.vu.nl/2009/11/sem/>

⁸ <https://www.w3.org/TR/owl-time/>

These types of networks can be connected to the other ontology concepts through three properties, *has_node* that depicts its cellular components, *has_interaction* that describes the interactions linked to its components and the property *has_node only* that specifies exactly the nature and type of its components.

- The class *Node*: This class contains the different types of cellular entities M that constitute the biomolecular network. In fact, we can identify three sub-classes: the *Gene* which describes the set M_G , the *Protein* which models the set M_P and the *Metabolite* which describes the M_M . This class is connected with the *Node_State* through the property *has_state*.
- The class *Interaction*: This class covers all the diverse types of interactions that can be operated among the nodes of the biomolecular network. This class consists of two sub-classes, *Intraomic_Interactions* that covers the interactions between molecular components of the same type and the class *Interomic_Interaction* that describes the interactions between molecular components of the different type. This class is connected to the *Node* class via two properties, *has_source* and *has_end*.
- The class *Node_State* contains the possible states of the nodes. This class is composed of two sub-classes, the *Concentration* and the *Activation*.
- The class *Interaction_Type* allows to specify the types and the nature of the interaction among cellular components. This class is linked to the *BNO:Interaction* class through the properties *Has_type*. To successfully integrate the main Interaction ontology concepts (*IO:Activity_flow* and *IO:Process*) with the Biomolecular Network ontology, we create an abstract BNO UML *BNO:Interaction_Type* to generalise those two Interaction ontology concepts (Figure 1).

3.5 The Relations Among the Ontologies

Concepts in the Biomolecular Network ontology are linked to the Gene ontology classes. In fact, the concepts of the Gene Ontology are used to enrich the definitions of the concepts of the Biomolecular Network ontology by an equivalence relation *owl:equivalenceClass*. For example, as described in Figure 2(b), after inference the concept *BNO:Protein* will be specialized by the concept *GO:beta-galactosidase* (GO: 0009341) because the *BNO:Node* concept is equivalent to the concept *GO:cellular_component* (GO: 0005575). The Biomolecular Network ontology is also linked with the Simple Event Model ontology through the *BNO: Node* concept, in fact an *SEM:event* can stimulate a molecular entity (represented by the concept *BNO: Node*). The Simple Event Model ontology will be used to describe the states of *BNO:Node* and its behaviour.

Moreover, the Time ontology (TO) has been integrated in the Simple Event Model ontology. The concept *sem:Time* was made equivalent to the concept *TO:TemporalEntity* which represents the root of the Time ontology. Hence, the property *sem:hasTime* will connect the Simple Event Model ontology to the Time ontology and, as a consequence, the diverse types of temporal concepts will be defined as specializations of the class *sem:Time*. Figure 2(b) shows a use of this principle. Thus, we can exploit the wealth of temporal concepts provided by this temporal ontology to describe the *SEM:event* class. Using these relationships it is possible to merge these ontologies to formalize the necessary knowledge to study the state changes of the biomolecular network's behaviour.

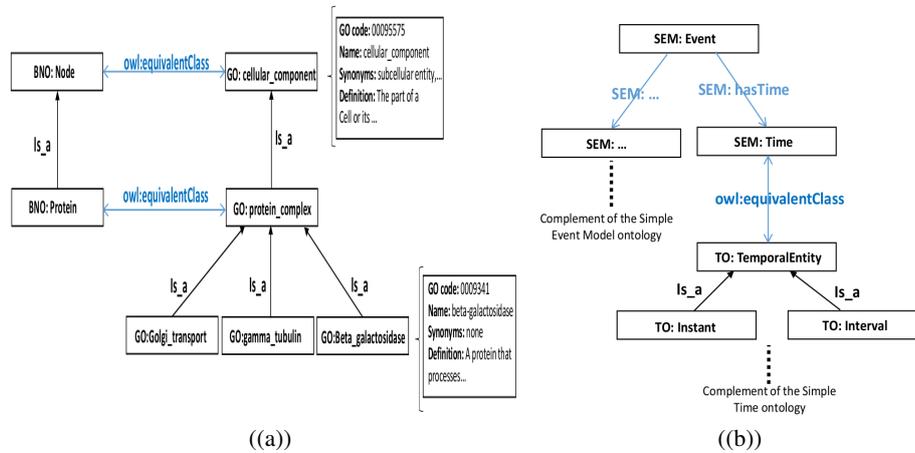


Fig. 2. Example of merging: 2(a) The Gene ontology concepts to the Biomolecular Network ontology concepts. 2(b) The Time ontology within the Simple Event Model ontology.

4 Conclusion

This paper proposes an effective approach for analysis and understanding the behaviour of complex biomolecular networks over time. This approach combines both of a logical modeling of biomolecular networks and a semantic approach that consists on four ontologies merged together.

We present a logic-based approach for modeling the dynamic behaviour of biomolecular networks. This formalism is based on the three levels of analysis of systems theory: structural, functional and behavioural modeling.

Moreover, the use of a semantic approach based on merging different ontologies can overcome issues of study the state changes of the complex biomolecular networks and their behaviour. We develop the Biomolecular Network Ontology (BNO) to describe the static structure of complex biomolecular networks and merge it with the Gene Ontology (GO) to provide structured terminologies for the description of cellular components. We also chose the Simple Event Model Ontology (SEMO) to describe events and stimuli which can stimulate the network's components and integrated the Time Ontology (TO) to study the different states of the biomolecular network and its nodes over time.

References

1. Aderem, A.: Systems Biology: Its Practice and Challenges Cell. 121(4), pp. 511-513 (2005)
2. Ratushny, A.V., Ramsey, S.A., Aitchison, J.D.: Mathematical modeling of biomolecular network dynamics. Network Biology: Methods and Applications. pp. 415-433 (2011)
3. Gillespie, D.T.: A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. Journal of computational physics. 22(4), 403-434 (1976)
4. Wilkinson, D.J.: Stochastic modelling for systems biology. CRC press. (2011)

5. Kauffman, S.A.: Metabolic stability and epigenesis in randomly constructed genetic nets. *Journal of theoretical biology*. 22(3), 437-467 (1969)
6. Zhao, Y., Kim, J., Filippone, M.: Aggregation algorithm towards large-scale boolean network analysis. *IEEE Transactions on Automatic Control*. 58(8), 1976-1985 (2013)
7. Husmeier, D.: Sensitivity and specificity of inferring genetic regulatory interactions from microarray experiments with dynamic bayesian networks. *Bioinformatics*. 19(17), 2271-2282 (2003)
8. Chaouiya, C., Kludel, H., Pommereau, F.: A modular, qualitative modeling of regulatory networks using Petri nets. In *Modeling in Systems Biology*. Springer London. pp. 253-279 (2011)
9. Courtot, M., Juty, N., Knüpfer, C., Waltemath, D., Zhukova, A., Dräger, A., ... & Hoops, S.: Controlled vocabularies and semantics in systems biology. *Molecular systems biology*, 7(1), 543 (2011)
10. Knüpfer, C., & Beckstein, C.: Function of dynamic models in systems biology: linking structure to behaviour. *Journal of biomedical semantics*, 4(1), 1 (2013)
11. Le Moigne, J. L.: *La théorie du système général: théorie de la modélisation*. jeanlouis le moigne-ae mcx. (1994)
12. Wu, F.X., Wu, L., Wang, J., Liu, J., Chen, L.: Transittability of complex networks and its applications to regulatory biomolecular networks. *Scientific reports*. 4, 4819 (2014)
13. Landeghem, S.V., Parys, T.V., Dubois, M., Inzé, D., de Peer, Y.V.: Diffany: an ontology driven framework to infer, visualise and analyse differential molecular networks. *BMC Bioinformatics*. 17(1), 1–12 (2016)
14. Ayadi, A., Zanni-Merk, C., de Bertrand de Beuvron, F., Krichen, S.: Logical and Semantic Modeling of Complex Biomolecular Networks. *Procedia Computer Science*. vol. 396, pp. 475-484 (2016)
15. Smith, B., Williams, J., Schulze-Kremer, S.: The ontology of the gene ontology. In: *AMIA*. vol. 3, pp. 609–613 (2003)
16. Consortium, G.O., et al.: The gene ontology project in 2008. *Nucleic acids research*. 36(suppl 1), D440–D444 (2008)
17. Van Hage, W. R., Malaisé, V., Segers, R., Hollink, L. and Schreiber, G.: Design and use of the simple event model (sem). *Web Semantics: Science, Services and Agents on the World Wide Web*. 9(2), 128–136 (2011)
18. Hobbs, J.R., Pan, F.: An ontology of time for the semantic web. *ACM Transactions on Asian Language Information Processing (TALIP)*. 3(1), 66–85 (2004)
19. Hobbs, J.R., Pan, F.: Time ontology in owl. *W3C working draft*. 27, 133 (2006)
20. Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., et al.: Gene ontology: tool for the unification of biology. *Nature genetics*. 25(1), 25–29 (2000)
21. Duncan, J., Eilbeck, K., Narus, S.P., Clyde, S., Thornton, S., Staes, C.: Building an ontology for identity resolution in healthcare and public health. *Online journal of public health informatics*. 7(2) (2015)
22. Brockmans, S., Volz, R., Eberhart, A., Löffler, P.: Visual modeling of OWL DL ontologies using UML. In: *International Semantic Web Conference*. pp. 198–213. Springer (2004)
23. Bārzdiņš, J., Bārzdiņš, G., Čerāns, K., Liepiņš, R., Sproģis, A.: UML style graphical notation and editor for OWL 2. In *Perspectives in Business Informatics Research*. Springer Berlin Heidelberg, 102-114 (2010)