Role of Nerve Growth Factor Signaling in Cancer Cell Proliferation and Survival Using a Reachability Analysis Approach

Gustavo Santos-García, Carolyn Talcott, Adrián Riesco, Beatriz Santos-Buitrago and Javier De Las Rivas

Abstract Systems biology attempts to understand biological systems by their structure, dynamics, and control methods. Nerve growth factor (NGF) is a neuropeptide involved in cellular signaling that binds specific cell surface receptors in order to induce cellular proliferation and survival in different cell types or cell contexts. In this paper we perform a reachability analysis and we compute common elements in all possible solutions in our cases of interest with the help of Pathway Logic, which constitutes a rewriting logic formalism that provides a knowledge base and development environment to carry out model checking, searches, and executions of signaling systems. In conclusion, we provide a symbolic system that explores complex and dynamic cellular signaling processes that induce cellular proliferation and cellular survival.

G. Santos-García(✉)
University of Salamanca, Salamanca, Spain
e-mail: santos@usal.es

C. Talcott
Computer Science Laboratory, SRI International, Menlo Park, USA
e-mail: clt@csl.sri.com

A. Riesco
Universidad Complutense de Madrid, Madrid, Spain
e-mail: ariesco@ucm.es

B. Santos-Buitrago
Seoul National University, Seoul, South Korea
e-mail: bsantosb@snu.ac.kr

J. De Las Rivas
Cancer Research Center (CSIC/USAL) and IBSAL, Salamanca, Spain
e-mail: jrivas@usal.es

Pathway Logic development has been funded in part by NIH BISTI R21/R33 grant (GM068146-01), NIH/NCI P50 grant (CA112970-01), and NSF grant IIS-0513857. This work was partially supported by NSF grant CNS-1318848. Research was supported by Spanish projects Strongsoft TIN2012-39391-C04-04, TRACES TIN2015-67522-C3-3-R, and PI12/00624 (MINECO, Instituto de Salud Carlos III) and Comunidad de Madrid project N-Greens Software-CM (S2013/ICE-2731).

M.S. Mohamad et al. (eds.), *10th International Conference on PACBB*,
Advances in Intelligent Systems and Computing 477,
DOI: 10.1007/978-3-319-40126-3_18

Keywords Signal transduction · Symbolic systems biology · Nerve growth factor · Pathway logic · Rewriting logic · Maude

1 Modeling Signaling Networks

Systems biology is an emergent field that facilitates understanding biological systems by describing their structure, dynamics, and control methods. Investigation of mammalian signaling processes, the molecular pathways by which cells detect, convert, and internally transmit information from their environment to intracellular targets such as the genome, would greatly benefit from the availability of predictive models [3, 14]. Various models for computational analysis of cellular signaling networks have been proposed to simulate responses to specific stimuli [10, 11]. However, in many cases complex cell signaling pathways have to be treated with other more qualitative modeling approaches, like logic modeling.

Symbolic models allow us to represent partial information and to model and analyze systems at multiple levels of detail, depending on the information available and the questions to be studied. Such models are based on formalisms that provide a language for representing system states and mechanisms of change such as reactions, and analysis tools based on computational or logical inference. Symbolic models can be used for simulation of system behavior.

Nerve growth factor (NGF) cellular signaling is involved in the regulation of development, maintenance, growth, proliferation, survival, and death of certain neurons [8]. This signaling is initiated by binding the ligand to two membrane-bound receptors (the tropomyosine receptor kinase A, `TrkA`, and the low-affinity NGF receptor, `NGFR`) so as to trigger a cellular signaling path (Fig. 1).

In this paper, Section 1 gives an overview to logical modeling of biological systems with rewriting logic and Pathway Logic. In Sections 2 and 3, we show the implementation of various rules and advanced logical inferences in the NGF signaling pathway. Conclusions are drawn in Section 4.

Pathway Logic Models of Signaling Networks in Cells. Rule-based modeling allows us to intuitively specify biological interactions while abstracting from the underlying combinatorial complexity. Since Ordinary Differential Equations (ODE) based systems [4] are considered in general the standard model for analyzing systems biology, it is worth briefly discussing the differences between these approaches. From the reactions point of view, ODE-based frameworks focus on diffusion-like reactions, while rule-based frameworks rely on the concept of reactive molecular collisions. On the other hand, ODE-based approaches model the average behavior of a system where the time evolution is a continuos process, while rule-based approaches model individual runs that reach different states non-deterministically chosen. Finally, ODE-based analyses are able to deal with huge numbers of molecules per species, but their complexity becomes too complex when the number of molecules is huge. Contrarily, rule-based approaches deal easily with many molecular species, but cannot deal

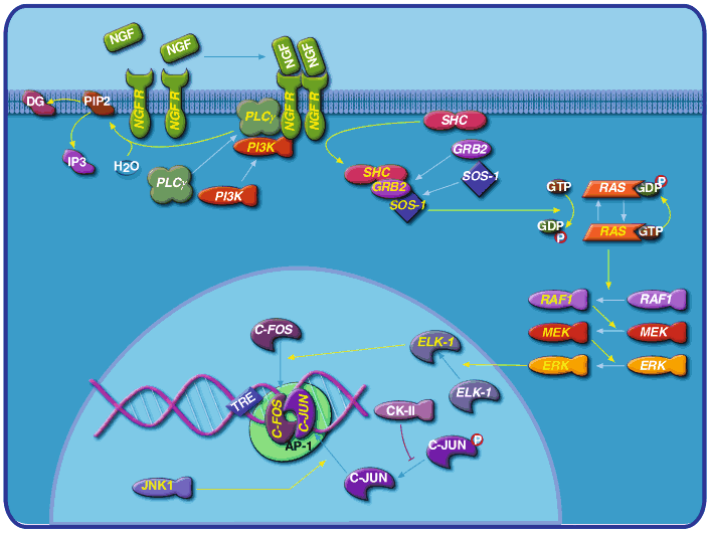

Fig. 1 Signaling of nerve growth factor pathway inside the cell (reprinted from [9]).

with very large number of molecules per species. In this work we are interested in using rule-based systems, since we want to analyze how particular dishes evolve and the particular states reachable from these dishes.[1]

Pathway Logic [12] is an approach to the modeling and analysis of molecular and cellular processes based on rewriting logic. Pathway Logic models of biological processes are developed using the Maude language. A Pathway Logic knowledge base includes data types representing cellular components such as proteins, small molecules, or complexes; compartments/locations; post-translational modifications and other dynamic events occurring in cellular reactions. The naturalness of rewriting logic for modeling and experimenting with mathematical and biological problems has been illustrated in a number of works [6, 7]. The basic idea is that we can model a cell as a concurrent system whose concurrent transitions are precisely its biochemical reactions [13]. In this way we can develop symbolic models of biological systems which can be analyzed like any other rewrite theory.

Maude [1, 2] is a high performance language and system supporting both equational and rewriting logic computation. Maude programs achieve a good agreement between mathematical and operational semantics. There are three different uses of Maude modules: (1) as programs that solve some problems; (2) as formal executable specifications that provide a rigorous mathematical model of an algorithm, a system, a language, or a formalism; and (3) as models that can be formally analyzed and verified with respect to different properties expressing various formal requirements.

[1] We use the term *dish* to refer to an initial state in analogy to an experimental setup in a Petri dish.

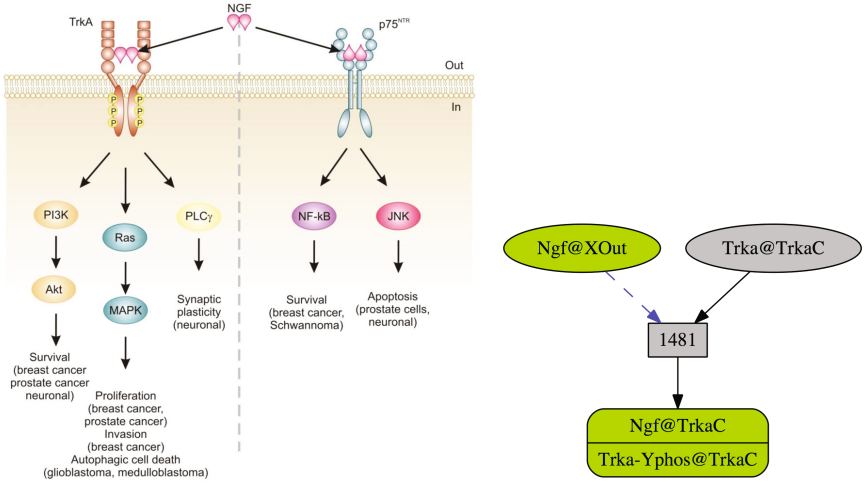

Fig. 2 (a) NGF binding to TrkA receptor mediates proliferation, differentiation and survival via activation of PI3K/Akt, Ras/MAPK and PLCγ pathways (cf. [8, Figure 1]). (b) Rule `[1481.Trka.irt.Ngf]` using Pathway Logic Assistant.

The Pathway Logic system, its documentation, a collection of examples, and related papers are available on http://pl.csl.sri.com. Models of cellular response to many different stimuli can also be found on our website.

2 Case Study: Modeling and Dynamics of NGF Signaling Pathway

2.1 Modeling: Dishes and Rewrite Rules

In this section we define some dishes and rules of the STM7 Pathway Logic knowledge base. A formal knowledge base contains information about the changes that occur in the proteins inside a cell in response to exposure to receptor ligands, chemicals, or various stresses. In our case study we will focus on models of response to nerve growth factor (NGF) stimulation. NGF cellular signaling is involved in the regulation of development, maintenance, growth, proliferation, survival, and death of certain nerve cells (neurons). The NGF signaling pathway includes reactions and circuits and, in fact, can induce cellular proliferation activating proteins ERK and AKT inside the cells (Fig. 2a).

An initial state or *dish* (called `NgfDish`) with several locations and elements is defined: the membrane (location tag `CLm`) is empty; the inside of the membrane (location tag `CLi`) contains proteins `Hras`, `Rap1a`, and `Rit1` binding to `GDP`;

the cytoplasm (location tag CLc) contains enzyme Pi3k and proteins Akt1, Aps, Araf, Arms, etc. (see the full code below); and the nucleus (location tag NUc) contains some genes (Egr1-gene and Fos-gene) and proteins (Elk1, Foxo1, and Foxo3). Moreover, there are three other locations: the outside (location tag XOut) which contains the nerve growth factor (Ngf), the NgfRC location which contains the common p75 neurotrophin nerve growth factor receptor (NgfR or $p75^{NTR}$), and the TrkaC location which contains the tropomyosin receptor kinase A (Trka):

```
eq NgfDish = PD({XOut | Ngf} {NgfRC | NgfR} {TrkaC | Trka}
  {CLm | empty}   {CLi | [Hras - GDP] [Rap1a - GDP] [Rit1 - GDP]}
  {CLc | Pi3k Akt1 Aps Araf Arms Braf Crk CrkI CrkL Erk5 Erks Frs2 Gab1
       Gab2 Ikba Ikk1 Ikk2 Irak1 Jnks Matk Mek1 Mek2 Myd88 P38s Pkcd
       Plcg1 Pxn Raf1 RapGef1 Rela Ripk2 Sh2b1 Shc1 Sqstm1 Traf6 Znf274}
  {NUc | Egr1-gene Fos-gene Elk1 Foxo1 Foxo3}) .
```

Rewrite rules describe the behavior of proteins and other components depending on modification states and biological contexts. Each rule represents a step in a biological process such as metabolic reactions or intra/inter cellular signaling reactions.

Rewrite Rule 1481. Pathway Logic contains a set of rules, derived from curated experimental findings, that provide a logical explanation of how a signal propagates in response to an NGF stimulus. Here we describe rule 1481, directly sourced from the literature. Hartman *et al.* [5] determine that PC12 cells express two distinct nerve growth factor receptors (NGFRs), p75NGFR and trkA (p140trk).

Our rewrite rule 1481 establishes: *In the presence of nerve growth factor Ngf in the outside of the cell (XOut), the protein Trka is phosphorylated on tyrosine ([Trka - Yphos]) and binds to NGF (Ngf).* In Maude syntax, this signaling process is described by the following rewrite rule:

```
rl [1481.Trka.irt.Ngf]:
   {XOut | xout Ngf}  {TrkaC | trkac Trka}
=> {XOut | xout Ngf}  {TrkaC | trkac ([Trka - Yphos] : Ngf)} .
```

Figure 2b shows this rule using the Pathway Logic Assistant. Ovals represent biomolecules participating in reactions (e.g., proteins, genes, etc.). Rectangles represent reaction rules with a label which represents its abbreviated identifier in the knowledge base. Solid arrows from an occurrence oval to a rule indicate that the occurrence is a reactant. Dashed arrows indicate that the occurrence is a modifier/enzyme/control, i.e., it is necessary for the reaction to take place but is not changed by the reaction. Solid arrows from a rule to an occurrence oval indicate that the occurrence is a product.

2.2 *Dynamics: Logical Inferences*

Thanks to the Knowledge Base provided by Pathway Logic, our analysis begins with the initial dish state `NgfDish` defined in Section 2.1. It is a well-known fact that cell proliferation and survival are connected to activation of `Akts` or `Erks`. We want to find out if there is a pathway from `NgfDish` leading to activation of `Akts` or `Erks`. In this case one can use the `search` command with a suitable search pattern and parameters (`[n]`: the first *n* solutions; `=>+`: at least one step). The target state is defined by the operator `PD`, whose argument is a "soup" of locations with their respective contents. A soup is a multiset that can include several elements regardless of their order.

The contents of each location (e.g., `XOut`) are things and/or variables (e.g., `thXOut:Things`) that contain elements according to the matching criteria of our search. In the cytoplasm, a protein `prot:BProtein` must be activated and can also have a set of other modifiers `mod:ModSet`. The search condition imposes that the variable `prot:BProtein` has membership in the sort `ErkS` or `AktS`.

```
Maude> search [1000] in QQ : NgfDish =>+ PD( loc:Locations
  {TrkaC | th:Things (Ngf : [Trka - Yphos])} {NgfRC | Ngf : NgfR}
  {XOut | thXOut:Things}  {CLm | thCLm:Things}
  {CLi | thCLi:Things}    {NUc | thNUc:Things}
  {CLc | thCLc:Things [prot:BProtein - mod:ModSet act]})
  such that (prot:BProtein :: ErkS  or  prot:BProtein :: AktS ) = true .
```

The solutions to this query given by Maude show the matching in the previous search pattern. While the terms fixed by the search pattern are not shown (e.g. `Ngf : NgfR`), the variables are presented with their corresponding values. For example, the 999th solution has the following values:

```
Solution 999 (state 48024)
loc:Locations --> empty
thXOut:Things --> Ngf
thCLm:Things --> empty
thCLi:Things --> [Hras - GDP] [Rap1a - GDP] [Rit1 - GDP]
thCLc:Things --> Aps Arms Crk CrkI CrkL Erks Erk5 Frs2 Gab1 Gab2 Ikba Ikk1
     Ikk2 Irak1 Jnks Matk Mek1 Mek2 Myd88 P38s Pi3k Pkcd Plcg1 Pxn Raf1
     RapGef1 Rela Ripk2 Shc1 Sqstm1 Traf6 [Araf - act] [Braf - act] Znf274
thNUc:Things --> Egr1-gene Fos-gene Elk1 Foxo1 Foxo3
th:Things --> [Sh2b1 - Yphos]
prot:BProtein --> Akt1
mod:ModSet --> none
```

In this solution, we observe that the variable `prot:BProtein` matches with protein `Akt1` without any set of modifications. We find out two proteins `Araf` and `Braf` in an activated form in the cytoplasm. We show evidence of a ligand/receptor effect: nerve growth factor (NGF) binds a specific cell surface receptor (NGFR). Then we ask Maude for the rule labels which have been applied to reach the final state according to the solution. One of these rules is the rewrite rule `1481.Trka.irt.Ngf` described above.

```
Maude> show path labels 48024 .
  1481.Trka.irt.Ngf    1482.NgfR.irt.Ngf    1484.Akt1.irt.NgfR.Ngf
  1488.Araf.irt.Ngf    1489.Braf.irt.Ngf    1517.Sh2b1.irt.Ngf
```

In this way, Maude allows us to explore the complete search space following a breadth-first strategy until all solutions are found.

3 Advanced Logical Inferences: Computing Similarities

A key distinguishing feature of Maude language is its systematic and efficient use of reflection (i.e. Maude's capability of handling and reasoning about terms that represent specifications described in Maude itself) through its predefined `META-LEVEL` module [2, Chapter 14]. Using the metalevel we can implement functions that manipulate and execute the modules given as argument, as well as the results obtained from these executions.

In this way, we have devised a method to compute all the solutions for a pathway given, the initial term `NgfDish`, a pattern and a condition for the reached terms, and a bound in the number of steps. Once all these terms have been computed it compares them in order to find the common structure underlying the solutions. That is, given the solutions `f(g(0), c, a)`, `f(b, c, g(c))`, and `f(b, c, g(c))`, it returns a skeleton `f(g(_), c)` that reveals the structure shared by all the solutions up to this depth.

Using this function we first tried to compute the similarities shared by all the solutions obtained for the search in the previous section. However, we realized that the state space was too big for this kind of abstraction, since too many reachable solutions are available and the relation between them is lost. Hence, we decided to start with low bounds for the depth of the search and progressively increase it to see how it evolves.

In particular, we have computed the common structure for the 536 solutions found when applying at most 5 rules and the 5718 solutions found for at most 6 rules. For the first search we obtain the following result:

```
Maude> red computeSimilarities(module, init, pattern, cond, '+, 5) .
PD( {XOut | Ngf} {NgfRC | Ngf : NgfR} {TrkaC | _} {CLm | empty}
{CLc | Ikk1 Ikk2 Irak1 Jnks Matk Myd88 P38s Pkcd Ripk2 Sqstm1 Traf6
     Znf274} {CLi | [Hras - _] [Rap1a - _] [Rit1 - _]}  {NUc | Elk1})
```

where the constants `module`, `init`, `pattern`, and `cond` stand for the module, initial term, pattern, and condition used in the previous search, respectively. Similarly, we have the following search for depth 6:

```
Maude> red computeSimilarities(module, init, pattern, cond, '+, 6) .
PD( {XOut | Ngf} {NgfRC | Ngf : NgfR} {TrkaC | _} {CLm | empty}
   {CLc | Ikk1 Ikk2 Irak1 Jnks Matk Myd88 Pkcd Ripk2 Sqstm1 Traf6
   Znf274} {CLi | [Hras - _] [Rap1a - _] [Rit1 - _]}  {NUc | _})
```

In these outputs, we find a ligand/receptor binding `NGF/NGFR`. In the search for depth 5 the proteins `Elk1` and `P38s` are present in the cytoplasm. However, in the search for depth 6, these proteins disappear as a common element. This results from the fact that a rule (`rl[618.P38s.irt.Ngf]`) is applied with activation of protein `P38s`.

4 Conclusions

Understanding of the dynamics of complex biological systems can be facilitated by developing symbolic methods. We formalize models that molecular biologists can use to think about signaling pathways and their behavior, allowing them to computationally formulate questions about their dynamics and outcomes. Rewriting logic gives us the ability to build and analyze models with multiple levels of detail; to represent biological rules; to define sorts of elements (chemicals, proteins, genes, locations, etc.) and their properties; and to precise queries using logical inference.

In this article we show an application of a rewriting logic procedure based in Maude logic language to the dynamic modeling of biological signaling pathways. As a case study, we analyze the role of nerve growth factor receptor signaling in cancer cell proliferation and survival using a reachability analysis approach. The characterization of neuron proliferation and survival is determined by protein activation of the families `AKT` and/or `ERK` [8]. We compute common elements in all possible solutions in our cases of interest. In conclusion, our results provide a reachability analysis in which a symbolic system explores complex and dynamic cellular signaling processes that induce cellular proliferation and cellular survival.

References

1. Clavel, M., Durán, F., Eker, S., Escobar, S., Lincoln, P., Martí-Oliet, N., Meseguer, J., Talcott, C.: Maude Manual (Version 2.7), March 2015. http://maude.cs.illinois.edu/w/images/1/1a/Maude-manual.pdf
2. Clavel, M., Durán, F., Eker, S., Lincoln, P., Martí-Oliet, N., Meseguer, J., Talcott, C.L.: All about maude - a high-performance logical framework, how to specify, program and verify systems in rewriting logic. In: LNCS, vol. 4350. Springer (2007)
3. Donaldson, R., Talcott, C.L., Knapp, M., Calder, M.: Understanding signalling networks as collections of signal transduction pathways. In: Quaglia, P. (ed.) Computational Methods in Systems Biology, CMSB, pp. 86–95. ACM (2010)
4. Gratie, D., Iancu, B., Petre, I.: ODE analysis of biological systems. In: Bernardo, M., de Vink, E.P., Pierro, A.D., Wiklicky, H. (eds.) 13th Int. School on Formal Methods for the Design of Computer, Communication, and Software Systems, SFM 2013. LNCS, vol. 7938, pp. 29–62. Springer (2013)
5. Hartman, D.S., McCormack, M., Schubenel, R., Hertel, C.: Multiple trkA proteins in PC12 cells bind NGF with a slow association rate. J. Biol. Chem. **267**(34), 24516–24522 (1992)

6. Martí-Oliet, N., Ölveczky, P.C., Talcott, C.L. (eds.) Logic, Rewriting, and Concurrency - Essays dedicated to José Meseguer on the occasion of his 65th birthday. LNCS, vol. 9200. Springer (2015)
7. Meseguer, J.: Twenty years of rewriting logic. J. Log Algebr. Program. **81**(7–8), 721–781 (2012)
8. Molloy, N.H., Read, D.E., Gorman, A.M.: Nerve growth factor in cancer cell death and survival. Cancers **3**(1), 510–530 (2011)
9. National Cancer Institute (US): Cancer Genome Anatomy Project. http://cgap.nci.nih.gov/Pathways (2016) (accessed February 29, 2016)
10. Santos-García, G., De Las Rivas, J., Talcott, C.L.: A logic computational framework to query dynamics on complex biological pathways. In: Adv. Intell. Syst. Comput., vol. 294, pp. 207–214. Springer (2014)
11. Santos-García, G., Talcott, C.L., De Las Rivas, J.: Analysis of cellular proliferation and survival signaling by using two ligand/receptor systems modeled by pathway logic. In: HSB 2015, pp. 226–245 (2015) (revised selected papers)
12. Talcott, C.L.: Pathway logic. In: Bernardo, M., Degano, P., Zavattaro, G. (eds.) Formal methods for computational systems biology. In: 8th Int. School on Formal Methods for the Design of Computer, Communication, and Software Systems, SFM 2008. LNCS, vol. 5016, pp. 21–53. Springer (2008)
13. Talcott, C.L., Eker, S., Knapp, M., Lincoln, P., Laderoute, K.: Pathway logic modeling of protein functional domains in signal transduction. In: Markstein, P., Xu, Y. (eds.) Proceedings of 2nd IEEE Computer Society Bioinformatics Conf, CSB 2003, Stanford, CA, August 11-14, 2003, pp. 618–619. IEEE Computer Society (2003)
14. Weng, G., Bhalla, U.S., Iyengar, R.: Complexity in biological signaling systems. Science **284**(5411), 92–96 (1999)