Research Article

Epidermal Growth Factor Signaling towards Proliferation: Modeling and Logic Inference Using Forward and Backward Search

Adrián Riesco,[1] Beatriz Santos-Buitrago,[2] Javier De Las Rivas,[3] Merrill Knapp,[4] Gustavo Santos-García,[5] and Carolyn Talcott[6]

[1]*Universidad Complutense de Madrid, Madrid, Spain*
[2]*Bio and Health Informatics Lab, Seoul National University, Seoul, Republic of Korea*
[3]*Cancer Research Center (CSIC/USAL) and IBSAL, Salamanca, Spain*
[4]*Biosciences Division, SRI International, Menlo Park, CA, USA*
[5]*University of Salamanca, Salamanca, Spain*
[6]*Computer Science Laboratory, SRI International, Menlo Park, CA, USA*

Correspondence should be addressed to Gustavo Santos-García; santos@usal.es

In biological systems, pathways define complex interaction networks where multiple molecular elements are involved in a series of controlled reactions producing responses to specific biomolecular signals. These biosystems are dynamic and there is a need for mathematical and computational methods able to analyze the symbolic elements and the interactions between them and produce adequate readouts of such systems. In this work, we use rewriting logic to analyze the cellular signaling of epidermal growth factor (EGF) and its cell surface receptor (EGFR) in order to induce cellular proliferation. Signaling is initiated by binding the ligand protein EGF to the membrane-bound receptor EGFR so as to trigger a reactions path which have several linked elements through the cell from the membrane till the nucleus. We present two different types of search for analyzing the EGF/proliferation system with the help of Pathway Logic tool, which provides a knowledge-based development environment to carry out the modeling of the signaling. The first one is a standard (forward) search. The second one is a novel approach based on *narrowing*, which allows us to trace backwards the causes of a given final state. The analysis allows the identification of critical elements that have to be activated to provoke proliferation.

1. Symbolic Systems Biology

The technological advances in the analysis of global gene expression and the growth of genomic sequence information have revolutionized research in biology and biomedicine [1, 2]. 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 [3–5].

The goal of our research is to develop abstract qualitative models of metabolic and signaling processes that can be used as the basis of further analyses by powerful tools (e.g., formal verification and model checking with formal methods). We model and analyze the biological processes by which an initial cellular system can lead to the activation of proliferation signaling by using a ligand/receptor pathway. The cells receive external signals by certain biomolecules (ligands) that are able to interact with certain receptors on the cellular surface producing some effects inside the cell [6]. A ligand/receptor system consists of a pathway with multiple elements and known reactions that have been modeled using Maude language by Pathway Logic in a natural, efficient way.

In this work, we use a knowledge base provided by the Pathway Logic system [7]. EGF signaling pathway was selected taking into account its relevance in carcinogenic processes. A novel backward search system has been developed

in accordance with the theoretical foundations provided by *narrowing*. Moreover, we present an application of classical forward search with EGF signaling pathway. Authors have published other forward search applications in [8–10].

Various approaches for computational analysis of cellular signaling networks have been proposed to simulate responses to specific stimuli [11, 12]. Simulations using in silico models founded on kinetic measurements of signaling pathways or networks allow us to achieve a detailed understanding of the biochemistry of signal transduction [13]. Ordinary differential equations or stochastic approaches are some of the quantitative methods. The use of differential equations to fit changes in the concentrations from the input to the output is an adequate approach when for a given pathway there are a large amount of quantitative information and a small number of reactions to be modeled [14, 15]. In many cases, however, qualitative approaches (e.g., logic modeling) are more convenient to model complex cell signaling pathways.

Symbolic models are based on formalisms that provide a language to represent the states of a system, mechanisms to model their changes (such as reactions), and tools for analysis based on computational or logical inference. A variety of formalisms have been used to develop symbolic models of biological systems, including Petri nets [16, 17], statecharts [18], live sequence charts [19], ambient/membrane calculi [20], and rule-based systems [21, 22].

In our domain of application, biological interactions can intuitively be specified by rule-based modeling without taking account of the underlying complexity. Rule-based approaches prevent the combinatorial explosion that results from molecular entities existing under multiple conditions. Kappa [23] and BioNetGen [24] are partly similar to Pathway Logic from the rule-based modeling formalisms.

In this paper, Section 1 gives an overview of symbolic systems biology with a rewriting logic and Pathway Logic approach. In Sections 2 and 3, we analyze the role of epidermal growth factor signaling in cancer cell proliferation using narrowing-rewriting techniques. A novel contribution for backward searches in signaling pathways is described. Conclusions are drawn in Section 4.

1.1. Cellular Signaling Networks with Pathway Logic Models. Pathway Logic (PL) provides a symbolic systems biology approach to modeling biological processes based on rewriting logic [7, 25, 26]. PL makes it possible to build and analyze models with multiple levels of detail, give information and representations of general rules, define novel data sorts and their properties, and use logical inference for executing queries.

PL primarily allows us to develop abstract qualitative models of signaling and metabolic processes. These models can be used as the basis for analysis by powerful tools to query dynamics on complex biological pathways. A number of recent contributions show their potential for inferring executable models [28] and analyzing the dynamics in different signaling transduction pathways (e.g., nerve growth factor, hepatocyte growth factor, or interleukin-6 signaling pathways) [8–10]. PL also enables developing quantitative and probabilistic models [29].

1.2. Rewriting Logic and Maude. Rewriting logic [30, 31] is logic of change that has been successfully applied to represent many different kinds of concurrent systems. A theory in rewriting logic consists of an equational theory, which allows the user to specify sorts, constructors and function symbols (possibly including some equational axioms), and equality between terms. Rewriting logic extends this equational theory by adding the notion of *rewrite rule*, which represents transitions between states.

A simple example of rewriting logic theory specification would be a vending machine that we assume to be composed of a multiset of products (e.g., apples and chocolate) and money (only quarters and dollars are allowed). The sort (i.e., the datatype) required for this theory would be the multiset, whose constructors would be the empty multiset, an apple, a chocolate, a quarter, a dollar, and the union of smaller multisets. Note that the union is commutative (it is unimportant whether we have a quarter and a dollar or a dollar and a quarter) and associative (it is unimportant whether we put together a chocolate with the multiset composed of a quarter and a dollar or we put together the multiset composed of a chocolate and a quarter with a dollar) and has the empty multiset as identity (a multiset does not change by adding the empty multiset). Possible rewrite rules would be the one that transforms a dollar into four quarters, the one that transforms a dollar into a chocolate and one quarter, or the one that transforms two quarters into an apple.

Using the idea of transition between states, it is possible to model biological systems in rewriting logic in a very natural way: while cells are just a set of multisets standing for the different components appearing in a real cell, biochemical reactions are represented by means of rules.

Rewriting logic is efficiently implemented in Maude [32, 33]. In the case of Maude, the underlying equational theory is Membership Equational Logic [34], which in addition to equations allows the user to define membership axioms stating the members of a sort. Maude provides several analysis tools for rewrite theories, including a rewrite computation, breadth-first search, and LTL model checking. Using these features, it is possible to study how our system behaves, to check whether it is possible to reach a certain state from an initial one, and to analyze whether our system verifies some temporal properties. Moreover, since rewriting logic is reflexive [35], a key distinguishing feature of Maude is its metalevel, which allows users to manipulate Maude modules and terms as standard data [36].

2. Case Study: EGF Signaling Pathway

We present in this section the case study that will be used to illustrate our technique in Section 3. It is interesting to discuss here the contributions made by these sections: Section 2 describes the pathway of interest and explains how to perform forward analysis, which requires a similar approach to the ones followed by the authors and others when analyzing other pathways. On the other hand, the backward analysis in Section 3 is completely novel; to the best of our knowledge it has never been used in a system following a transitional scheme. Even though the foundations of the analysis are the
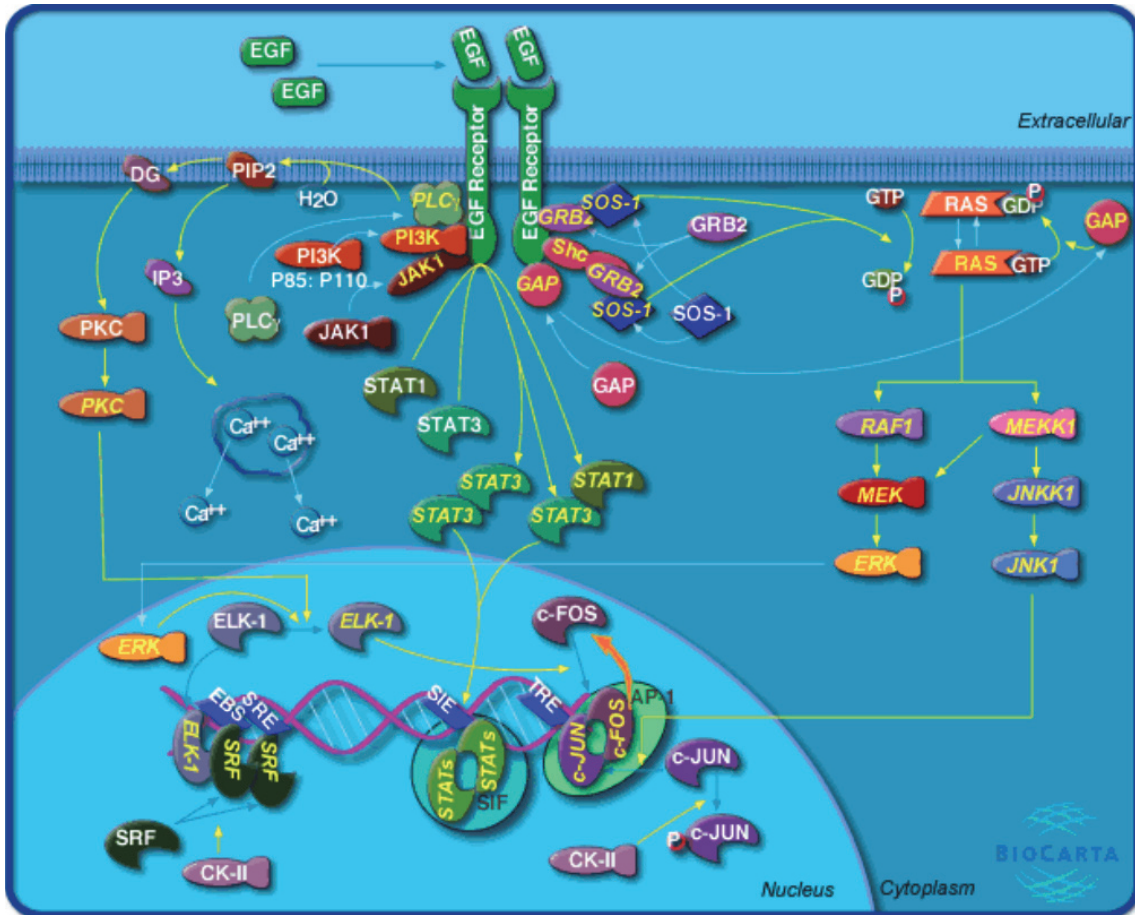
Figure 1: Human EGF signaling pathway (reprinted from Biocarta [27]).

rules described in Section 2, they are automatically transformed in order to apply the backward search, which in turn is not implemented in the core distribution of Maude and hence requires further modifications to be used. Moreover, this analysis is completely general and can be applied to any other pathway without further modifications.

2.1. Epidermal Growth Factor Signaling Pathway.

The ErbB family of the receptor tyrosine kinases contains the epidermal growth factor receptor (EGFR) [38, 39]. These receptors couple the binding of the extracellular growth factor ligands to intracellular signaling pathways that control various biologic responses such as proliferation, differentiation, cell motility, and survival [40–45]. Three major steps are involved in the activation of EGFR-dependent intracellular signaling [46]: (a) the binding of a receptor-specific ligand takes place in the extracellular portion of the EGFR or of one of the EGFR-related receptors; (b) the formation of a functionally active EGFR-EGFR dimer or a heterodimer causes the ATP-dependent phosphorylation of specific tyrosine residues in the EGFR intracellular domain; and (c) this phosphorylation triggers a complex program of intracellular signals to the cytoplasm and then to the nucleus.

EGFR activates two major intracellular pathways: (i) the RAS-RAF-MEK-MAPK pathway, which controls gene transcription, cell-cycle progression from the G1 phase to the S phase, and cell proliferation, and (ii) the PI3K-Akt pathway, which activates a cascade of antiapoptotic and prosurvival signals: bFGF, basic fibroblast growth factor, HB-EGF, heparin-binding EGF, MAPK, mitogen-activated protein kinase, PI3K, phosphatidylinositol 3,4,5-kinase, TGFa, transforming growth factor alpha, and VEGF, vascular endothelial growth factor (Figure 1).

The binding between EGFR and ligand triggers downstream intracellular signaling pathways. Some of them include the PI3K-AKT prosurvival, STAT transcription, and RAS-RAF-MEK proliferation pathways [47]. The RAS-RAF-MEK and PI3K-AKT pathways are mostly activated by the anaplastic lymphoma kinase (ALK) fusion proteins. Cell proliferation, cell motility, and carcinogenesis are driven by the amplification of the EGFR and ALK signaling pathways (Figure 2).

2.2. Modeling: Dishes and Rewrite Rules.

Some dishes and rules of the STM7 Pathway Logic knowledge base are defined below. 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
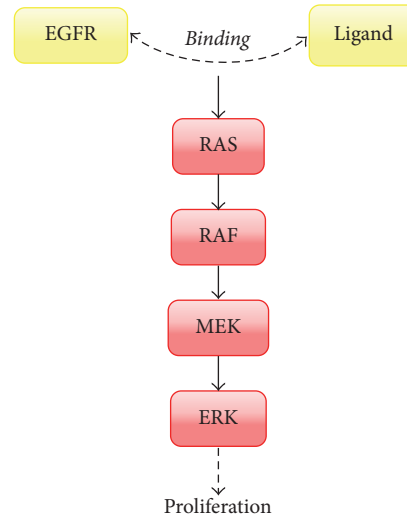
FIGURE 2: The binding between EGFR and ligand gives rise to intracellular signaling pathways including the RAS/RAF/MEK proliferation pathways [37].

epidermal growth factor (EGF) stimulation. Somatic mutations that lead to EGFR overactivity or upregulation are associated with several types of cancer (e.g., glioblastoma multiforme (GBM), lung cancer, or anal cancers). These mutations involving EGFR lead to its constant activation, which produces uncontrolled cell division. EGFR signal transduction pathways include reactions and, in fact, can induce cellular proliferation activating proteins ERK inside the cells.

An initial state or *dish* (called `EgfDish`) with several locations and elements is defined:

(i) The outside (location tag `XOut`) contains the epidermal growth factor (`Egf`).

(ii) The `EgfRC` location contains the epidermal growth factor receptor (`EgfR`).

(iii) The `CLo` location, which contains the elements stuck to the outside of the plasma membrane, is empty.

(iv) The membrane (location tag `CLm`) contains proteins `Erbb2`, `Pag1`, and `Plscr1`.

(v) The inside of the membrane (location tag `CLi`) contains several proteins binding to guanosine diphosphate GDP, `Cdc42`, `Hras`, `Kras`, and so forth, and some other proteins, `Gnai1`, `Gnai3`, `Pld1`, and so forth (see the full code shown in Box 1).

(vi) The cytoplasm (location tag `CLc`) contains enzyme `Pi3k` and some proteins, `Abl1`, `Akt1`, `Araf`, `ArhGap5`, and so forth.

(vii) The nucleus (location tag `NUc`) contains several proteins (`Atf1`, `Creb1`, `Elk1`, etc.).

In Maude syntax, this *dish* (called `EgfDish`) is expressed by the equation shown in Box 1.

Rewrite rules detail the behavior of cell components depending on biological contexts and modification states. Each rule represents an action in a biological process such as intra/intercellular signaling reactions or metabolic reactions.

2.2.1. Rewrite Rule `001.EgfR.irt.Egf`. Pathway Logic contains a set of transition rules, derived from curated experimental findings. They provide an explanation of how a signal propagates in response to an EGF stimulus. Here, we describe rule 001, directly sourced from the literature. Yarden and Sliwkowski [48] determine that when EGF and its relatives bind the ErbB family of receptors, they trigger a network of signaling pathways, culminating in responses ranging from cell division to death and from motility to adhesion.

Our rewrite rule 001 establishes the following: *in the presence of epidermal growth factor* `Egf` *in the outside of the cell (*`XOut`*), the receptor* `EgfR` *is phosphorylated on tyrosine (*`[EgfR - Yphos]`*) and binds to protein EGF (*`Egf`*)* [48, 49]. In Maude syntax, this signaling process is expressed by the rewrite rule shown in Box 2.

2.2.2. Rewrite Rule `188.Shp2.irt.Egf`. *When protein EGF (*`Egf`*) is bound to its receptor EGFR phosphorylated on tyrosine (*`[EgfR - Yphos]`*) and the cytoplasm contains the tyrosine phosphatase Shp2, then* `Shp2` *is phosphorylated on tyrosine and recruited to the* `EgfRC` *container* [50]. This signaling process is expressed by the Maude rewrite rule shown in Box 3.

2.2.3. Rewrite Rule `529.Hras.irt.Egf`. *When protein EGF (*`Egf`*) is bound to its receptor EGFR phosphorylated on tyrosine (*`[EgfR - Yphos]`*), the GRB2-associated-binding protein Gab1 or* `Gab2` *phosphorylated on tyrosine is present (*`[gab:GabS - Yphos]`*), the Ras guanyl-releasing protein 3* `RasGrp3` *or protein* `Sos1` *phosphorylated on tyrosine is present,* `Pi3k` *is present, the protein* `Shp2` *phosphorylated on tyrosine is present, and the inside of the membrane (*`Cli`*) contains* `Hras` *loaded with guanosine diphosphate GDP, then* `Hras` *switches its load to guanosine 5′-triphosphate GTP* [51]. The rewrite rule shown in Box 4 represents this signaling process.

Figure 3 shows the aforementioned rules using the Pathway Logic Assistant. An oval represents a component

```
eq EgfDish =
  PD ({XOut | Egf} {EgfRC | EgfR} {CLo | empty} {CLm | Erbb2 Pag1 Plscr1}
      {CLi | [Cdc42 - GDP] [Hras - GDP] [Kras - GDP] [Nras - GDP]
             [Rac1 - GDP] [Rala - GDP] [Ralb - GDP] [Rap1a - GDP]
             [Rap2b - GDP] [Rit1 - GDP] Gnai1 Gnai3 Pld1 Pld2 Src}
      {CLc | (Mlst8 : Mtor : Raptor) [Gsk3s - act] Abl1 Akt1 Araf ArhGap5
             ArhGef4 ArhGef7 Atf2 Bmx Braf Cbl Cblb Cin85 Crk CrkL Csk Dbl
             Dok1 Dok2 Eif4ebp1 EndA1 Eps8 Eps15 Erk5 Erks Fak1 Fak2 Flna
             Freud1 Gab1 Gab2 Git1 Grb2 Hpk1 Ipo7 IqGap1 Jak1 Jak2 Jnks Lkb1
             Mapkapk2 Mek1 Mek5 Mekk1 Mekk2 Mekk3 Mlk3 Nckipsd P38s Pdpk1
             Pi3k Pkcd Pkcz Plcg1 Plce1 Ptk6 Pxn Raf1 RalGds RapGef1 Rasa1
             RasGrp3 Rela Rictor Rin1 Rps6 Rsk1 Rsk2 Rsk3 S6k1 Sh2d3a Sh2d3c
             Shc1 Shoc2 Shp2 Sin1 Smad3 Sos1 Stat1 Stat3 Stat5a Stat5b Tnk2
             Tns3 Ube2l3 Vav1 Vav2 Vav3 Ywhaz Y196 Y1122}
      {NUc | Atf1 Creb1 Elk1 Fos HistH3 Jun Msk1 Msk2 Myc} ) .
```

Box 1

```
rl [001.EgfR.irt.Egf]:
     {XOut | xout Egf} {EgfRC | egfrc EgfR}
 =>  {XOut | xout} {EgfRC | egfrc ([EgfR - Yphos] : Egf)} .
```

Box 2

```
rl [188.Shp2.irt.Egf]:
     {EgfRC | egfrc ([EgfR - Yphos] : Egf )} {CLc | clc Shp2}
 => {EgfRC | egfrc ([EgfR - Yphos] : Egf ) [Shp2 - Yphos]} {CLc | clc} .
```

Box 3

(e.g., gene, protein) participating in a reaction. A rectangle illustrates a reaction rule with a label which represents its shortened identifier in the knowledge base. A solid arrow from an occurrence oval to a rule indicates that the occurrence is a reactant. A dashed arrow indicates that the occurrence is a modifier, enzyme, or control. That is, it is a necessary element for the reaction to take place but is kept unchanged by the reaction. A solid arrow from a rule to an occurrence oval indicates that the occurrence is a product.

2.3. Dynamics: Logical Inferences. Our analysis starts with the initial dish state EgfDish defined in Section 2.2 and derived from the knowledge base provided by Pathway Logic. It is a well-known fact that cell proliferation is connected to activation of Erks [52]. We want to find out whether there is a pathway from EgfDish leading to activation of 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., EgfRC) are elements and/or variables (e.g., thEgfRC:Things) according to the matching criteria of our search. In the nucleus, 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 (see Box 5).

The solution to this query given by Maude shows the matching in the previous search pattern. While the terms fixed by the search pattern are not shown (e.g., [EgfR - Yphos] : Egf), the variables are presented with their corresponding values. The first solution has the values shown in Box 6.

In this solution, we observe that the variable on-the-fly prot:BProtein matches protein Erks with modifications phos (TEY) and phos (SPS). We find out proteins Braf, Gsk3s, Mlk3, and Mek1 in an activated form in the cytoplasm. We show evidence of a ligand/receptor effect: epidermal growth factor (EGF) binds a specific cell surface receptor (EGFR). Then, we ask Maude for the rule labels which have been applied to reach the final state according to the solution < (see Box 7). Some of these rules are the rewrite rules 001.EgfR.irt.Egf, 188.Shp2.irt.Egf, and 529.Hras.irt.Egf described in Boxes 2, 3, and 4.

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

```
rl [529.Hras.irt.Egf]:
    {EgfRC | egfrc ([EgfR - Yphos] : Egf) [gab:GabS - Yphos]
        [hrasgef:HrasGEF - Yphos] Pi3k [Shp2 - Yphos]}
    {CLi | cli [Hras - GDP]}
=> {EgfRC | egfrc ([EgfR - Yphos] : Egf) [gab:GabS - Yphos]
        [hrasgef:HrasGEF - Yphos] Pi3k [Shp2 - Yphos]}
    {CLi | cli [Hras - GTP]} .
```
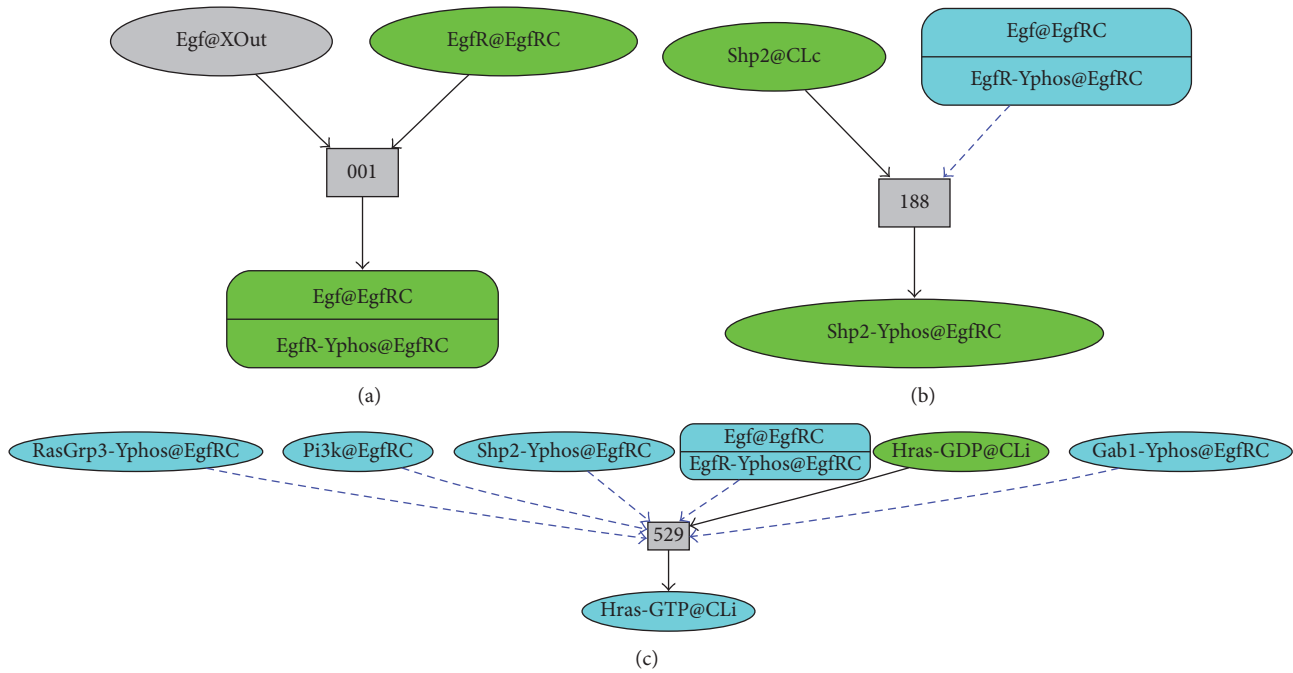
Box 4

(a)

(b)

(c)

FIGURE 3: (a) Rule [001.EgfR.irt.Egf] using Pathway Logic Assistant. (b) Rule [188.Shp2.irt.Egf] using PLA. (c) Rule [529.Hras.irt.Egf] using PLA.

3. Searching for Causes

In the previous section, we used a standard (forward) search to analyze the set of reachable states from an initial dish; now, we can use *narrowing* to analyze the possible initial states leading to a particular state. That is, in this section, we propose a backward search for analyzing the causes leading to different scenarios. We will first introduce the theoretical notions underlying our scheme, and then we will show how the search works.

Narrowing [53–55] is a generalization of the term rewriting, allowing free variables in terms and replacing pattern matching by unification in order to reduce these terms. It was first used to solve equational unification problems [56] and then generalized to deal with symbolic reachability problems [57]. More formally, the difference between a rewriting step and a narrowing step is that in both cases we use a rewrite rule $l \Rightarrow r$ to rewrite t at a position p (we express this subterm as $t|_p$), but narrowing unifies the left-hand side l and $t|_p$; that is, it uses a substitution σ such that $l\sigma =_A t|_p\sigma$ before actually performing the rewriting step, while in rewriting $t|_p$ must

be an instance of l (i.e., only matching is required). From an initial term t only containing variables (except for the function symbol at the top), we can obtain a substitution σ, using this narrowing approach, and generate a term $t\sigma$, such that $t\sigma$ can be rewritten using the traversed rules to some term t'. The current implementation of narrowing in Maude has been applied to symbolic model checking [58] and test-case generation [59], among others.

Recalling our example for the vending machine from the previous section, it is easy to see the difference between standard rewriting and narrowing. Remember that we have a rule in our system that turns two quarters into an apple; in this system, we might wonder whether it is possible to reach a state including an apple starting from a system with two quarters. Using rewriting, we reach a positive answer, as well as starting from a state with three quarters, from a state with one dollar (since we can use the rule to transform it into four quarters), and, in general, from any initial state containing two or more quarters. However, all these initial states must be concrete and contain either quarters, dollars, apples, or chocolates. Applying rewrite to an initial state containing a quarter and

```
search [1] in QQ : EgfDish =>+ PD( locOther:Locations
    {EgfRC | thEgfRC:Things [EgfR - Yphos] : Egf}
    {XOut | thXOut:Things} {CLm | thCLm:Things}
    {CLi | thCLi:Things}   {CLc | thCLc:Things}
    {NUc | thNUc:Things [prot:BProtein - mod:ModSet act]})
    such that (prot:BProtein :: ErkS) = true .
```

Box 5

```
Solution 1 (state 17589)
locOther:Locations --> {CLo | empty}
thCLm:Things --> Erbb2 Pag1 Plscr1
thCLi:Things --> Gnai1 Gnai3 Pld1 [Cdc42 - GDP] [Hras - GTP] [Kras - GDP]
      [Nras - GDP] [Rac1 - GTP] [Rala - GDP] [Rit1 - GDP] Pld2
thXOut:Things --> empty
thCLc:Things --> Abl1 Akt1 Araf Cbl Cblb Crk CrkL Csk Dbl Erk5 Fak1 Gab1
      Hpk1 IqGap1 Jak1 Jak2 Jnks Lkb1 Mek5 Mekk1 Mekk2 Mekk3 Nckipsd P38s
      Pdpk1 Pkcd Raf1 RapGef1 Rsk1 (Mlst8 : Mtor : Raptor) [Gsk3s - act]
      [Braf - act] [Mlk3 - act] [Mek1 - act] Shoc2 Sin1 Smad3 Stat1 Stat3
thNUc:Things --> Atf1 Creb1 Elk1 Fos HistH3 Ipo7 Jun Msk1 Msk2 Myc
prot:BProtein --> Erks
mod:ModSet --> phos (TEY) phos (SPS)
thEgfRC:Things --> Grb2 Pi3k [Fak2 - act] [Gab2 - Yphos] [Git1 - Yphos]
    [Pxn - Yphos] [Shp2 - Yphos] [Ptk6 - act] [Shc1 - Yphos]
    [Sos1 - Yphos] [Src - act]
```

Box 6

```
Maude> show path labels 17589 .
001.EgfR.irt.Egf   117.Grb2.irt.Egf   197.Sos1.irt.Egf   532.Ptk6.irt.Egf
683.Rac1.irt.Egf   188.Shp2.irt.Egf   075.Gab2.irt.Egf   207.Src.irt.Egf
669.Fak2.irt.Egf   398.Git1.irt.Egf   191.Shc1.irt.Egf   172.Pi3k.irt.Egf
529.Hras.irt.Egf   310.Braf.irt.Egf   639.Mek1.irt.Egf   1063.Mlk3.irt.Egf
1102.Pxn.irt.Egf   196.Erks.act.irt.Egf   1602.Erks.to.nuc.irt.Egf
1730.Erks.Ipo7.irt.Egf   1122.Erks.SPS.phos.irt.Egf
```

Box 7

a variable standing for a multiset would return a negative answer, because rewriting is not able to assign values to the variable (e.g., a quarter in our case) and hence cannot apply rules. Narrowing, on the other hand, is able to assign values to variables in order to apply rules, so a narrowing search would return a positive result and the required substitution (our variable would be just an apple in 0 steps, a quarter in 1 step, a dollar in 2 steps, etc.).

We can use the rule 01.EgfR.irt.Egf from Section 2 to illustrate the difference between rewriting and narrowing. First, for applying rewriting, we can only start from ground terms (terms without variables), so we could start with the term {XOut | Egf} {EgfRC | EgfR} and obtain the following result by applying the rule {XOut | none} {EgfRC | [EgfR - Yphos] : Egf}.

Note that starting with a ground term greatly limits the power of the search, since (i) a slight variation in this term might imply huge differences in the reachable terms and (ii) it limits the causes of the reachable states to that dish, hence ignoring other simpler or more general causes. In order to solve this problem, narrowing is applied to terms with variables that stand for potentially infinite sets of ground terms. In that case, we could start the process with the term {XOut | xout} {EgfRC | egfrc}. This way, we do not impose any constraint to the initial term. The narrowing process would find that the variable xout must be of the form xout' Egf, while egfrc must be of the form egfrc' EgfR, for xout' and egfrc' fresh variables. This way, we would obtain the term {XOut|xout'}{EgfRC|egfrc' ([EgfR - Yphos]: Egf)}, which is more general than the one above. Since the narrowing process returns the bindings applied to the initial variables, we can use it to check whether a state is reachable from a general term and then obtain the most general initial

```
Maude> (search [2] in RULES : PD(L:Locations) ~>*
 PD ({CLo | empty}
 {EgfRC | Grb2 Pi3k [Fak2 - act] [Gab2 - Yphos] [Git1 - Yphos]
        [Ptk6 - act] [Pxn - Yphos] [Shc1 - Yphos] [Shp2 - Yphos]
        [Sos1 - Yphos] [Src - act] [EgfR - Yphos] : Egf}
 {XOut | empty}
 {CLm | Erbb2 Pag1 Plscr1}
 {CLi | Gnai1 Gnai3 Pld1 Pld2 [Cdc42 - GDP] [Hras - GTP] [Kras - GDP]
        [Nras - GDP] [Rac1 - GTP] [Rala - GDP] [Rit1 - GDP]}
 {CLc | Abl1 Akt1 Araf Cbl Cblb Crk CrkL Csk Dbl Erk5 Fak1 Gab1 Hpk1
        IqGap1 Jak1 Jak2 Jnks Lkb1 Mek5 Mekk1 Mekk2 Mekk3 Nckipsd P38s
        Pdpk1 Pkcd Raf1 RapGef1 Rsk1 Shoc2 Sin1 Smad3 Stat1 Stat3
        (Mlst8 : Mtor : Raptor) [Braf - act] [Gsk3s - act] [Mlk3 - act]
        [Mek1 - act]}
 {NUc | Atf1 Creb1 Elk1 Fos HistH3 Ipo7 Jun Msk1 Msk2 Myc
        [Erks - phos (TEY) phos (SPS) act]}) .)
```

Box 8

```
Solution 2
L:Locations -->
{CLc | Abl1 Akt1 Araf Cbl Cblb Crk CrkL Csk Dbl Erk5 Fak1 Gab1 Gab2 Hpk1
       IqGap1 Jak1 Jak2 Jnks Lkb1 Mek5 Mekk1 Mekk2 Mekk3 Nckipsd P38s Pdpk1
       Pkcd Raf1 RapGef1 Rsk1 Shoc2 Sin1 Smad3 Stat1 Stat3 [Mlk3 - act]
       (Mlst8 : Mtor : Raptor)[Braf - act][Gsk3s - act][Mek1 - act]}
{CLi | Gnai1 Gnai3 Pld1 Pld2[Cdc42 - GDP][Hras - GTP][Kras - GDP]
       [Nras - GDP][Rac1 - GTP][Rala - GDP][Rit1 - GDP]}
{CLm | Erbb2 Pag1 Plscr1}
{CLo | empty}
{EgfRC | Grb2 Pi3k(Egf :[EgfR - Yphos]) [Fak2 - act] [Git1 - Yphos]
       [Ptk6 - act][Pxn - Yphos][Shc1 - Yphos][Shp2 - Yphos][Sos1 - Yphos]
       [Src - act]}
{NUc | Atf1 Creb1 Elk1 Fos HistH3 Ipo7 Jun Msk1 Msk2 Myc
       [Erks - act phos (SPS) phos (TEY)]}
{XOut | empty}
```

Box 9

term required to reach this state, hence allowing us to discard the infinite initial dishes that are a particularization of this solution.

However, due to its computational power, narrowing poses a number of important restrictions. First, it can only be applied to unconditional theories. Secondly, since unification is only available for some particular theories, only some combinations of equational axioms are allowed. Finally, it is required for all the rewrite rules to be defined at the top (the rules must be applied to the complete term and not just to the subterms). Luckily enough, the Pathway Logic specification fulfills the first two constraints, although it fails to fulfill the last one. For this reason, we have developed a way to apply narrowing to any pathway. It is worth discussing the entity and complexity of this extension, whose main points are as follows: (i) it is a metalevel function, which is able to modify the rules in a given module to make them fulfill the constraints above, and (ii) it is Full Maude [32], since narrowing is not implemented in the built-in distribution of

Maude but in its Full Maude extension. The former poses theoretical difficulties, since it requires a deep understanding of the structure of Maude modules, terms, and sorts at the metalevel, while the latter poses technical difficulties, since it requires the programmer to deal with the input/output mechanisms and the explicit database used by Full Maude. This extension is available at https://github.com/ariesco/pathway.

For example, we look for the different dishes that can produce the first result for the search in Section 2.3 by using Box 8.

Although the first solution just indicates that we can reach this result by applying no rewrite rules and starting with the same substitution we are looking for, the second solution goes one step backwards and proposes the initial state shown in Box 9.

By looking for further solutions, Maude keeps going backwards until the initial dish is as general as needed by the user. Also note that the initial dish can be further specialized to indicate that we expect some initial components or to

focus in a particular element. We consider that this technique, which has been proposed for the first time in this paper, will be very useful for analyzing the causes of many pathways.

4. Conclusions

In the last few years, the analysis of biological systems has required an overwhelming and boundless amount of quantitative data with the use of new technologies [60–63]. Obtaining quantitative data in vivo and/or in vitro is often complicated due to ethical aspects or certain limits concerning experimentation. Quantitative methods for modeling and analysis of biochemical networks are infeasible in many cases.

The dynamics of complex biological systems can be explained by developing symbolic methods [64]. Molecular biologists can formalize models to think about signaling pathways and their behavior, allowing them to computationally raise questions about their dynamics and outcomes.

Rewriting logic provides a logical framework that gives us the ability to build and analyze models with multiple levels of detail [32]. Different sorts of elements (proteins, genes, cell locations, etc.) and their properties can be defined. Biological rules can be easily represented. Maude language allows us to perform searches using logical inference.

In this paper, we have described two different searches for studying the epidermal growth factor in the Maude implementation of the Pathway Logic. The first one allows the user to study how different particular dishes can evolve by applying a breadth-first search by rewriting, hence being able to check whether a certain state is reachable. The second search presents a novel approach by applying narrowing instead of rewriting, which allows the user to trace back the possible origins of a given scenario, hence providing a causality study. In order to perform this analysis, we have extended Full Maude, the tool where the narrowing capabilities of Maude are implemented, with a command that preprocesses the Pathway Logic modules in order to fulfill the narrowing requirements. We expect that this novel approach will be useful for analyzing other pathways where the forward approach has already been used.

Competing Interests

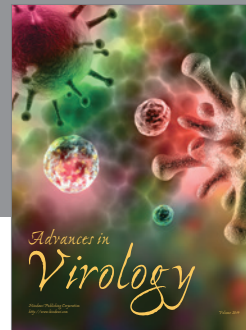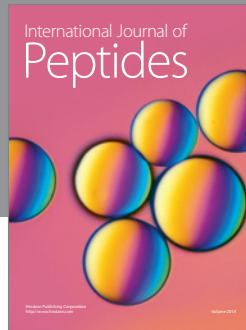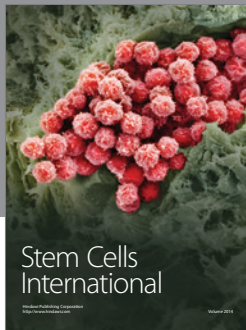The authors declare that they have no competing interests.

Acknowledgments

References

[1] J. Fisher and T. A. Henzinger, "Executable cell biology," *Nature Biotechnology*, vol. 25, no. 11, pp. 1239–1249, 2007.

[2] O. G. Vukmirovic and S. M. Tilghman, "Exploring genome space," *Nature*, vol. 405, no. 6788, pp. 820–822, 2000.

[3] R. Donaldson, C. L. Talcott, M. Knapp, and M. Calder, "Understanding signalling networks as collections of signal transduction pathways," in *Proceedings of the ACM 8th International Conference on Computational Methods in Systems Biology (CMSB '10)*, P. Quaglia, Ed., pp. 86–95, Trento, Italy, October 2010.

[4] L. M. Heiser, N. J. Wang, C. L. Talcott et al., "Integrated analysis of breast cancer cell lines reveals unique signaling pathways," *Genome Biology*, vol. 10, no. 3, article no. R31, 2009.

[5] R. S. Pais, N. Moreno-Barriuso, I. Hernández-Porras, I. P. López, J. De Las Rivas, and J. G. Pichel, "Transcriptome analysis in prenatal IGF1-deficient mice identifies molecular pathways and target genes involved in distal lung differentiation," *PLoS ONE*, vol. 8, no. 12, Article ID e83028, 2013.

[6] K. L. Carraway III and L. C. Cantley, "A neu acquaintance for ErbB3 and ErbB4: a role for receptor heterodimerization in growth signaling," *Cell*, vol. 78, no. 1, pp. 5–8, 1994.

[7] C. L. Talcott, *Formal Methods for Computational Systems Biology: 8th International School on Formal Methods for the Design of Computer, Communication, and Software Systems, SFM 2008 Bertinoro, Italy, June 2–7, 2008 Advanced Lectures*, vol. 5016 of *Lecture Notes in Computer Science*, Springer, Berlin, Germany, 2008.

[8] G. Santos-García, J. De Las, and C. L. Rivas, "A logic computational frame-work to query dynamics on complex biological pathways," in *8th International Conference on Practical Applications of Computational Biology & Bioinformatics (PACBB 2014)*, J. Saez-Rodriguez, M. P. Rocha, F. Fdez-Riverola, and J. F. De Paz Santana, Eds., vol. 294 of *Advances in Intelligent Systems and Computing*, pp. 207–214, Springer, Berlin, Germany, 2014.

[9] G. Santos-García, C. L. Talcott, A. Riesco, B. Santos-Buitrago, and J. De Las Rivas, "Role of nerve growth factor signaling in cancer cell proliferation and survival using a reachability analysis approach," in *Proceedings of the 10th International Conference on Practical Applications of Computational Biology & Bioinformatics (PACBB '16)*, M. S. Mohamad, M. P. Rocha, F. Fdez-Riverola, F. J. D. Mayo, and J. F. D. Paz, Eds., vol. 477 of *Advances in Intelligent Systems and Computing*, pp. 173–181, Springer, Sevilla, Spain, June 2016.

[10] G. Santos-García, C. L. Talcott, and J. De Las Rivas, "Analysis of cellular proliferation and survival signaling by using two ligand/receptor systems modeled by Pathway Logic," in *Hybrid Systems Biology: Fourth International Workshop, HSB 2015, Madrid, Spain, September 4-5, 2015. Revised Selected Papers*, A. Abate and D. Safránek, Eds., vol. 9271 of *Lecture Notes in Computer Science*, pp. 226–245, Springer, Berlin, Germany, 2015.

[11] A. R. Asthagiri and D. A. Lauffenburger, "A computational study of feedback effects on signal dynamics in a mitogen-activated protein kinase (MAPK) pathway model," *Biotechnology Progress*, vol. 17, no. 2, pp. 227–239, 2001.

[12] G. Weng, U. S. Bhalla, and R. Iyengar, "Complexity in biological signaling systems," *Science*, vol. 284, no. 5411, pp. 92–96, 1999.

[13] F. Eduati, J. De Las Rivas, B. Di Camillo, G. Toffolo, and J. Saez-Rodriguez, "Integrating literature-constrained and data-driven inference of signalling networks," *Bioinformatics*, vol. 28, no. 18, pp. 2311–2317, 2012.

[14] X. Liu, M. D. Betterton, A. Saadatpour, and R. Albert, "Discrete dynamic modeling of signal transduction networks," in *Computational Modeling of Signaling Networks: Methods and Protocols*,

vol. 880 of *Methods in Molecular Biology*, pp. 255–272, Humana Press, 2012.

[15] P. Smolen, D. A. Baxter, and J. H. Byrne, "Mathematical modeling of gene networks," *Neuron*, vol. 26, no. 3, pp. 567–580, 2000.

[16] S. Hardy and P. N. Robillard, "Petri net-based method for the analysis of the dynamics of signal propagation in signaling pathways," *Bioinformatics*, vol. 24, no. 2, pp. 209–217, 2008.

[17] C. Li, Q.-W. Ge, M. Nakata, H. Matsuno, and S. Miyano, "Modelling and simulation of signal transductions in an apoptosis pathway by using timed Petri nets," *Journal of Biosciences*, vol. 32, no. 1, pp. 113–127, 2007.

[18] S. Efroni, D. Harel, and I. R. Cohen, "Toward rigorous comprehension of biological complexity: modeling, execution, and visualization of thymic T-cell maturation," *Genome Research*, vol. 13, no. 11, pp. 2485–2497, 2003.

[19] A. Sadot, J. Fisher, D. Barak et al., "Toward verified biological models," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 5, no. 2, pp. 223–234, 2008.

[20] A. Regev, E. M. Panina, W. Silverman, L. Cardelli, and E. Shapiro, "BioAmbients: an abstraction for biological compartments," *Theoretical Computer Science*, vol. 325, no. 1, pp. 141–167, 2004.

[21] J. R. Faeder, M. L. Blinov, and W. S. Hlavacek, "Rule-based modeling of biochemical systems with BioNetGen," *Methods in Molecular Biology*, vol. 500, no. 1, pp. 113–167, 2009.

[22] W. Hwang, Y. Hwang, S. Lee, and D. Lee, "Rule-based multi-scale simulation for drug effect pathway analysis," *BMC Medical Informatics and Decision Making*, vol. 13, supplement 1, article S4, 2013.

[23] V. Danos and C. Laneve, "Formal molecular biology," *Theoretical Computer Science*, vol. 325, no. 1, pp. 69–110, 2004.

[24] M. L. Blinov, J. R. Faeder, B. Goldstein, and W. S. Hlavacek, "BioNetGen: software for rule-based modeling of signal transduction based on the interactions of molecular domains," *Bioinformatics*, vol. 20, no. 17, pp. 3289–3291, 2004.

[25] A. Panikkar, M. Knapp, H. Mi et al., "Applications of pathway logic modeling to target identification," in *Formal Modeling: Actors, Open Systems, Biological Systems: Essays Dedicated to Carolyn Talcott on the Occasion of Her 70th Birthday*, vol. 7000 of *Lecture Notes in Computer Science*, pp. 434–445, Springer, Berlin, Germany, 2011.

[26] C. L. Talcott, S. Eker, M. Knapp, P. Lincoln, and K. Laderoute, "Pathway Logic modeling of protein functional domains in signal transduction," in *Proceedings of the 2nd IEEE Computer Society Bioinformatics Conference (CSB '03)*, P. Markstein and Y. Xu, Eds., pp. 618–619, IEEE Computer Society, Stanford, Calif, USA, August 2003.

[27] National Cancer Institute (US), "Cancer Genome Anatomy Project," http://cgap.nci.nih.gov/Pathways.

[28] V. Nigam, R. Donaldson, M. Knapp, T. McCarthy, and C. L. Talcott, "Inferring executable models from formalized experimental evidence," in *Proceedings of the 13th International Conference on Computational Methods in Systems Biology (CMSB '15)*, O. F. Roux and J. Bourdo, Eds., vol. 9308 of *Lecture Notes in Computer Science*, pp. 90–103, Nantes, France, 2015.

[29] A. Abate, Y. Bai, N. Sznajder, C. Talcott, and A. Tiwari, "Quantitative and probabilistic modeling in pathway logic," in *Proceedings of the 7th IEEE International Conference on Bioinformatics and Bioengineering (BIBE '07)*, pp. 922–929, Boston, Mass, USA, January 2007.

[30] J. Meseguer, "Conditional rewriting logic as a unified model of concurrency," *Theoretical Computer Science*, vol. 96, no. 1, pp. 73–155, 1992.

[31] J. Meseguer, "Twenty years of rewriting logic," *The Journal of Logic and Algebraic Programming*, vol. 81, no. 7-8, pp. 721–781, 2012.

[32] M. Clavel, F. Durán, S. Eker et al., *All about Maude—A High-performance Logical Framework, How to Specify, Program and Verify Systems in Rewriting Logic*, vol. 4350 of *Lecture Notes in Computer Science*, Springer, 2007.

[33] M. Clavel, F. Durán, S. Eker et al., Maude Manual (Version 2.7) (March 2015), http://maude.cs.illinois.edu/w/images/1/1a/Maude-manual.pdf.

[34] A. Bouhoula, J.-P. Jouannaud, and J. Meseguer, "Specification and proof in membership equational logic," *Theoretical Computer Science*, vol. 236, no. 1-2, pp. 35–132, 2000.

[35] M. Clavel, J. Meseguer, and M. Palomino, "Reflection in membership equational logic, many-sorted equational logic, Horn logic with equality, and rewriting logic," *Theoretical Computer Science*, vol. 373, no. 1-2, pp. 70–91, 2007.

[36] A. Riesco, A. Verdejo, N. Martí-Oliet, and R. Caballero, "Declarative debugging of rewriting logic specifications," *The Journal of Logic and Algebraic Programming*, vol. 81, no. 7-8, pp. 851–897, 2012.

[37] K. Wu, L. House, W. Liu, and W. C. Cho, "Personalized targeted therapy for lung cancer," *International Journal of Molecular Sciences*, vol. 13, no. 12, pp. 11471–11496, 2012.

[38] R. S. Herbst, "Review of epidermal growth factor receptor biology," *International Journal of Radiation Oncology, Biology, Physics*, vol. 59, no. 2, supplement, pp. 21–26, 2004.

[39] N. Normanno, A. De Luca, C. Bianco et al., "Epidermal growth factor receptor (EGFR) signaling in cancer," *Gene*, vol. 366, no. 1, pp. 2–16, 2006.

[40] J. A. Engelman, "Targeting PI3K signalling in cancer: opportunities, challenges and limitations," *Nature Reviews Cancer*, vol. 9, no. 8, pp. 550–562, 2009.

[41] M. D. Marmor, K. B. Skaria, and Y. Yarden, "Signal transduction and oncogenesis by ErbB/HER receptors," *International Journal of Radiation Oncology Biology Physics*, vol. 58, no. 3, pp. 903–913, 2004.

[42] A. M. Krasinskas, "EGFR signaling in colorectal carcinoma," *Pathology Research International*, vol. 2011, Article ID 932932, 6 pages, 2011.

[43] K. M. Nicholson and N. G. Anderson, "The protein kinase B/Akt signalling pathway in human malignancy," *Cell Signal*, vol. 14, no. 5, pp. 381–395, 2002.

[44] T. Sasaki, K. Hiroki, and Y. Yamashita, "The role of epidermal growth factor receptor in cancer metastasis and microenvironment," *BioMed Research International*, vol. 2013, Article ID 546318, 8 pages, 2013.

[45] A. Suhardja and H. Hoffman, "Role of growth factors and their receptors in proliferation of microvascular endothelial cells," *Microscopy Research and Technique*, vol. 60, no. 1, pp. 70–75, 2003.

[46] J. A. Katzel, M. P. Fanucchi, and Z. Li, "Recent advances of novel targeted therapy in non-small cell Lung cancer," *Journal of Hematology and Oncology*, vol. 2, article no. 2, 2009.

[47] S. Marmiroli, D. Fabbro, Y. Miyata, M. Pierobon, and M. Ruzzene, "Phosphorylation, signaling, and cancer: targets and targeting," *BioMed Research International*, vol. 2015, Article ID 601543, 3 pages, 2015.

[48] Y. Yarden and M. X. Sliwkowski, "Untangling the ErbB signalling network," *Nature Reviews Molecular Cell Biology*, vol. 2, no. 2, pp. 127–137, 2001.

[49] K. Elenius, S. Paul, G. Allison, J. Sun, and M. Klagsbrun, "Activation of HER4 by heparin-binding EGF-like growth factor stimulates chemotaxis but not proliferation," *EMBO Journal*, vol. 16, no. 6, pp. 1268–1278, 1997.

[50] A. Yart, M. Laffargue, P. Mayeux et al., "A critical role for phosphoinositide 3-kinase upstream of Gab1 and SHP2 in the activation of Ras and mitogen-activated protein kinases by epidermal growth factor," *Journal of Biological Chemistry*, vol. 276, no. 12, pp. 8856–8864, 2001.

[51] A. Montagner, A. Yart, M. Dance, B. Perret, J.-P. Salles, and P. Raynal, "A novel role for Gab1 and SHP2 in epidermal growth factor-induced Ras activation," *Journal of Biological Chemistry*, vol. 280, no. 7, pp. 5350–5360, 2005.

[52] D. J. Son, J. E. Hong, J. O. Ban et al., "Synergistic inhibitory effects of cetuximab and cisplatin on human colon cancer cell growth via inhibition of the ERK-dependent EGF receptor signaling pathway," *BioMed Research International*, vol. 2015, Article ID 397563, 13 pages, 2015.

[53] J. R. Slagle, "Automated theorem-proving for theories with simplifiers commutativity, and associativity," *Journal of the Association for Computing Machinery*, vol. 21, pp. 622–642, 1974.

[54] M. J. Fay, "First-order unification in an equational theory," in *Proceedings of the 4th Workshop on Automated Deduction*, W. H. Joyner, Ed., pp. 161–167, Academic Press, Austin, Tex, USA, 1979.

[55] A. Middeldorp and E. Hamoen, "Counterexamples to completeness results for basic narrowing," in *Proceedings of the 3rd International Conference on Algebraic and Logic Programming*, vol. 632 of *Lecture Notes in Computer Science*, pp. 244–258, Springer, Pisa, Italy, 1992.

[56] J. H. Siekmann, "Unification theory," *Journal of Symbolic Computation*, vol. 7, no. 3-4, pp. 207–274, 1989.

[57] J. Meseguer and P. Thati, "Symbolic reachability analysis using narrowing and its application to verification of cryptographic protocols," *Higher-Order and Symbolic Computation*, vol. 20, no. 1-2, pp. 123–160, 2007.

[58] S. Escobar and J. Meseguer, "Symbolic model checking of infinite-state systems using narrowing," in *Proceedings of the 18th International Conference on Term Rewriting and Applications (RTA '07)*, vol. 4533 of *Lecture Notes in Computer Science*, pp. 153–168, Springer, 2007.

[59] A. Riesco, "Using narrowing to test Maude specifications," in *Rewriting Logic and Its Applications: 9th International Workshop, WRLA 2012, Held as a Satellite Event of ETAPS, Tallinn, Estonia, March 24-25, 2012, Revised Selected Papers*, F. Durán, Ed., vol. 7571 of *Lecture Notes in Computer Science*, pp. 201–220, Springer, Berlin, Germany, 2012.

[60] M. Altaf-Ul-Amin, T. Katsuragi, T. Sato, and S. Kanaya, "A glimpse to background and characteristics of major molecular biological networks," *BioMed Research International*, vol. 2015, Article ID 540297, 14 pages, 2015.

[61] F. Emmert-Streib and M. Dehmer, "Networks for systems biology: conceptual connection of data and function," *IET Systems Biology*, vol. 5, no. 3, pp. 185–207, 2011.

[62] S. Kanaya, M. Altaf-Ul-Amin, S. K. Kiboi, and F. M. Afendi, "Big data and network biology 2015," *BioMed Research International*, vol. 2015, Article ID 604623, 2 pages, 2015.

[63] F. Wu, M. Li, J. Ruan, and F. Luo, "Systems biology approaches to mining high throughput biological data," *BioMed Research International*, vol. 2015, pp. 1–2, 2015.

[64] N. Tenazinha and S. Vinga, "A survey on methods for modeling and analyzing integrated biological networks," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 8, no. 4, pp. 943–958, 2011.