On the Use of Stochastic Petri Nets in the Analysis of Signal Transduction Pathways for Angiogenesis Process

Napione Lucia[2,4], Manini Daniele[1], Cordero Francesca[1,3], Horváth András[1], Picco Andrea[2,4], De Pierro Massimiliano[1], Pavan Simona[2,4], Sereno Matteo[1], Veglio Andrea[2,4], Bussolino Federico[2,4], and Balbo Gianfranco[1]

[1] Department of Computer Science, University of Torino, Torino, Italy
[2] Institute for Cancer Research and Treatment, Candiolo (TO), Italy
[3] Department of Clinical and Biological Sciences, University of Torino, Torino, Italy
[4] Department of Oncological Sciences, University of Torino, Torino, Italy

Abstract. In this paper we consider the modeling of a selected portion of signal transduction events involved in the angiogenesis process. The detailed model of this process contains a large number of parameters and the data available from wet-lab experiments are not sufficient to obtain reliable estimates for all of them. To overcome this problem, we suggest ways to simplify the detailed representation that result in models with a smaller number of parameters still capturing the overall behaviour of the detailed one.

Starting from a detailed stochastic Petri net (SPN) model that accounts for all the reactions of the signal transduction cascade, using structural properties combined with the knowledge of the biological phenomena, we propose a set of model reductions.

1 Introduction

Formal modeling is a central theme in systems biology in which mathematical modeling and simulation can play an important role. The Petri net (PN) formalism [18] is a framework that allows the construction of a precise and clear representation of biological systems based on solid mathematical foundations. This formalism permits the study of qualitative properties related to the structure of the model (e.g., the structure of a biological pathway). The variant of PNs, called Stochastic Petri Nets (SPNs) [16,15,2] and characterized by the addition of timing and/or stochastic information, can be used for quantitative analysis (e.g., analysis that involve the rates in biochemical reactions). PNs have been first proposed for the representation of biological pathways by Reddy et al [17]. Since their introduction, many other researchers constructed PN models of biological pathways [11] with the aim of using their representations to obtain qualitative information about the behavior of these systems, mostly via simulation [12,9]. The interaction of qualitative and quantitative analysis is necessary to check a model for consistency and correctness; following this idea, Heiner et al [10] proposed a methodology to develop and analyze large biological models in a step-wise manner.

In this paper we present our experience in modeling signal transduction pathways for the angiogenesis process using SPNs. The general goal is to analyze the temporal dynamics of a few relevant biological products and this requires to build and parameterize

P. Degano and R. Gorrieri (Eds.): CMSB 2009, LNBI 5688, pp. 281–295, 2009.

the model of the phenomenon under study. A detailed model is built by biologists and then the parameters are estimated on the basis of data obtained by wet-lab experiments. It is often the case however that the amount of available wet-lab data is not sufficient to have reliable estimates of the many parameters involved in the model. The key contribution of this paper aims at alleviating this problem by providing a simplification process which transforms the detailed model into a simpler one with less parameters. The proposed simplification process is guided by qualitative properties together with knowledge on the phenomenon under study and it is validated by comparing the quantitative properties of the detailed and simplified models. Moreover, this process represents the basis of the development of arguments useful for identifying both critical complexes and interactions that play a crucial role in the biochemical system under study. With respect to the framework proposed by Gilbert et al [8], where the main idea is to illustrate the complementarity among the three different ways of modeling biochemical network , i.e. qualitative, stochastic and continuous, here we focus our attention on the definition and robustness of the simplification process to limit the complexity of the model.

Techniques that can be seen as simplification procedures have already been published in the literature. See, for example, [4,3] where approximate analysis methods based on aggregation of states are proposed. The goal of these techniques is however different from ours, since they aim at reducing the complexity of the analysis of the model and not the difficulty of its parametrization. Indeed, they result in simpler models in which the number of parameters is identical to that of the original one.

The paper is organized as follows. Section 2 provides an overview of PNs and SPNs and of their use in biochemical systems. Section 3.1 describes the angiogenesis case study. Section 3.2 presents the approach we followed to build the SPNs and Section 3.3 shows the formal and biological rules used in the simplification process as well as the resulting SPNs. The quantitative analysis performed in order to verify the mathematical robustness of the simplified model is proposed in Section 3.4. We conclude with a discussion and an outlook of future works in Section 4.

2 Modeling Formalism and Solution Techniques

The descriptions commonly applied in biology, where the relations among components are expressed by biochemical reactions, or by interactions of genes as well as by cell population interactions, are easy to transform into PNs in which places correspond to genes/proteins/compounds (substrates) and transitions to their interactions.

2.1 Petri Net Representation for Biochemical Entities Interactions

PNs are a graphical language for the formal description of distributed systems with concurrency and synchronization. PNs are bipartite graphs with two types of nodes, namely places and transitions, connected by directed arcs. The state of the system is given by the distribution of tokens over the places of the net. The dynamics of the model (starting from an initial marking) is captured by state changes due to firing of transitions and by the consequent movement of tokens over the places.

Definition 1. *[PN - syntax]*A Petri Net graph is a tuple $(P, T, W, \mathbf{m}_0)$, where

- P is a finite set of *places*;
- T is a finite set of *transitions*;
- P and T are such that $P \cap T = \emptyset$;
- $W : (P \times T) \cup (T \times P) \rightarrow \mathbb{N}$ defines the arcs of the net and assigns to each of them a multiplicity;
- $\mathbf{m}_0$ is the *initial* marking which associates with each place a number of *tokens*.

When applied in systems biology, places represent biochemical entities (enzymes, compounds, etc.) and transitions represent the interactions among entities [17]. The quantities of the entities are represented by tokens in the places. The biological system we consider consists of biochemical reactions similar to those reported in Fig. 1-a where we show the PN representation scheme we adopted to describe all reactions of this type. Fig. 1-b represents the state evolution due to firing of transition k_{53}.

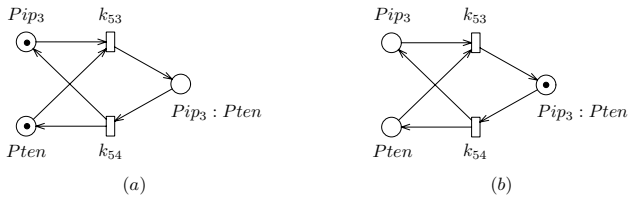

Fig. 1. PN representation of reactions $Pip3 + Pten \underset{k_{54}}{\overset{k_{53}}{\rightleftarrows}} Pip3 : Pten$

2.2 Analysis Techniques Based on Structural Properties

The PN graph inspection can provide several functional properties of the model, whose validity is true independently of the initial state of the system: such properties are, for instance, the boundedness and the existence of structural deadlocks and traps [18,2]. In deriving such kinds of information an important role is played by the so called net's *invariants*. There exist two kinds of invariants: place invariants (P-invariants) and transition invariants (T-invariants) [18]. In this paper we deal with P-invariants only. A P-invariant is a weighted sum of tokens contained in a subset of places of the net that remains constant through the entire evolution of the model, starting from an initial marking. The subset of places used for computing the P-invariant is the support (i.e., the set of nonzero components) of a P-semiflow $\mathbf{f}$ [14], which is a vector of nonnegative weights assigned to all the places of the net. A P-semiflow $\mathbf{f}$ is an integer and nonnegative solution of the matrix equation $\mathbf{fC} = 0$, where $\mathbf{C}$ is the incidence matrix of the net, obtained by properly using the information provided by the flow relation $\mathbf{W}$.

The interpretation of a P-invariant in a biological context, where tokens represent compounds, enzymes etc., is relatively simple: the places that support the semiflow $\mathbf{f}$ represent a portion of the PN where a given kind of correlated matter is preserved.

2.3 Quantitative Temporal Analysis

To study the temporal dynamics of a biological system it is natural to apply an extension of PNs that allows to introduce in the model temporal specifications. SPNs are time extensions of PNs in which exponentially distributed random delays (interpreted as durations of certain activities) are associated with the firings of the transitions. SPNs are qualitatively equivalent to PNs, meaning that for their structural analysis it is sufficient to disregard their time specifications. The temporal stochastic behaviour of an SPN is isomorphic to that of a continuous time Markov chain (CTMC). This CTMC can be built automatically from the description of the SPN and corresponds to the behaviour of the biological system described by the Master Chemical Equations [7]. This "stochastic approach" based on SPNs, adopts a discrete view of the quantity of the entities and sees their temporal behaviour as a random process.

Another possibility is to adopt a "deterministic approach" in which the temporal behaviour of the entities is seen as a continuous and completely predictable process. In our context we make use also of the deterministic approach because it allows for faster and simpler evaluation of the simplification process we propose for SPNs.

The deterministic approach translates the interactions into a set of coupled, first-order, ordinary differential equations (ODE) with one equation per entity. These equations describe how the quantities of the entities change based on the speed and the structure of the interactions among reactants. Referring again to the reactions considered in Figure 1, the corresponding ODEs are

$$\begin{aligned}
\frac{dX_{Pip3}(t)}{dt} &= -k_{53}X_{Pip3}(t)X_{Pten}(t) + k_{54}X_{Pip3:Pten}(t),\\
\frac{dX_{Pten}(t)}{dt} &= -k_{53}X_{Pip3}(t)X_{Pten}(t) + k_{54}X_{Pip3:Pten}(t),\\
\frac{dX_{Pip3:Pten}(t)}{dt} &= k_{53}X_{Pip3}(t)X_{Pten}(t) - k_{54}X_{Pip3:Pten}(t)
\end{aligned}$$

where $X_i(t)$ denotes the quantity of reactant i at time t. Having the ODEs and information on the initial amount of the different entities, numerical integration of the ODEs is applied to calculate the quantities at a given time instant.

3 A Stochastic Petri Nets Based Approach Applied to Signal Transduction Pathways for the Angiogenesis Process

One main objective in systems biology is to model and analyze temporal dynamics of the phenomenon under study. By using SPNs as the formalism for the construction of the model, the analysis is performed in two steps: the first provides qualitative information on the structure of the model and the second investigates quantitative properties including statistical indices describing the temporal behavior of the system. Here we use this approach to study the angiogenesis process.

3.1 Biological Case Study Definition

Angiogenesis, defined as the formation of new vessels from the existing ones, is a topic of great interest in all areas of human biology, particularly to scientists studying vascular development, vascular malformation and cancer biology. Angiogenesis is

a complex process involving the activities of many growth factors and relative receptors, which trigger several signaling pathways resulting in different cellular responses. The Vascular Endothelial Growth Factor (VEGF) family proteins are widely regarded as the most important growth factors involved in angiogenesis. VEGF-A, a member of VEGF family, has been most carefully studied and is thought to be of singular importance. VEGF receptor-2 (KDR in humans) is thought to mediate most of VEGF-A's angiogenic functions, including cell proliferation, survival, and migration. Although the core components of the main KDR-induced pathways have been identified, the molecular mechanisms involved need to be characterized in fine details in order to better understand the flow of information. Indeed, a strong body of evidences indicates the presence of common adaptor/effector proteins involved in the survival and proliferation pathways induced by VEGF-A/KDR axis, pointing out the difficulty to isolate a specific pathway and suggesting the presence of common nodes which contribute to create an intricate signaling network. In particular, the phosphorylated active receptor, indicated as KDR^*, catalyzes phosphorylation of several intracellular substrates including the adaptor protein $Gab1$ [13,5]. The main pathway through which VEGF-A induces cell proliferation involves the activation of $PLC\gamma$ [19]. Activation of $PLC\gamma$ promotes phosphatidylinositol 4,5-bisphosphate (Pip_2) hydrolysis giving rise to 1,2-diacylglycerol (DAG). VEGF-A-induced cell survival is dependent on the activity of $Pi3K$ [6]. The activated $Pi3K$ phosphorylates Pip_2 generating phosphatidylinositol-3,4,5-triphosphate (Pip_3). This recruits Akt to the membrane where it is activated trough phosphorylation. Activated Akt induces cell survival. Taking into account these notions, we wrote a system of biochemical reactions based on the available biological information together with further supposed mechanisms which could contribute to underline the presence of additional molecular nodes in the context of VEGF-A-induced proliferation and survival pathways.

3.2 Model Construction

In this section we discuss the approach we followed to represent the signal transduction cascade by SPN. Consider the detailed biological model depicted in Fig. 2.

These reactions describe KDR-proximal signaling events in the context of the survival and proliferation signal modules induced by receptor activation. In particular, reactions are split into four blocks. The First Block represents the earliest signaling events which include KDR^* (we use the star to denote that proteins are active), $Gab1$, and Pip_3. The Second Block concerns the regeneration of Pip_2, a common substrate for the two signal modules that we are considering. In this block Pip_2 recovery was considered to result from the contribution of Pten-dependent dephosphorylation of Pip_3 in combination with DAG catabolism (here recapitulated in the pseudo-enzyme E). The Third Block includes the reactions describing the survival pathway triggered by the $PI3K$/Akt axis. The Fourth Block represents the proliferation pathway involving $PLC\gamma$ activation. Using the reaction representations outlined in Section 2.1 and the GreatSPN tool [1] the SPN model of the angiogenesis process was built as illustrated in Fig. 3. Exploiting the block organization and the structure of the model we analyzed the biochemical reactions in order to identify possible pathways and sub-pathways that describe embedded behaviors of the complete model. We denoted the reactions by means of their kinetic constants.

KDR-Receptor (First Block)

$$Kdr^* + Gab1 \underset{k_1}{\overset{k_0}{\rightleftarrows}} Kdr^*{:}Gab1$$
$$Kdr^*{:}Gab1 \overset{k_2}{\rightarrow} Kdr^*{:}Gab1^*$$
$$Gab1 + Pip3 \underset{k_4}{\overset{k_3}{\rightleftarrows}} Gab1{:}Pip3$$
$$Kdr^* + Gab1{:}Pip3 \underset{k_6}{\overset{k_5}{\rightleftarrows}} Kdr^*{:}Gab1{:}Pip3$$
$$Kdr^*{:}Gab1{:}Pip3 \overset{k_7}{\rightarrow} Kdr^*{:}Gab1^*{:}Pip3$$
$$Kdr^*{:}Gab1^*{:}Pip3 \underset{k_9}{\overset{k_8}{\rightleftarrows}} Gab1^*{:}Pip3 + Kdr^*$$
$$Kdr^*{:}Gab1^* + Pip3 \underset{k_{11}}{\overset{k_{10}}{\rightleftarrows}} Kdr^*{:}Gab1^*{:}Pip3$$

Survival (Third Block)

$$Gab1^*{:}Pip3 + Pi3k \underset{k_{13}}{\overset{k_{12}}{\rightleftarrows}} Gab1^*{:}Pip3{:}Pi3k$$
$$Gab1^*{:}Pip3{:}Pi3k + Kdr^* \underset{k_{15}}{\overset{k_{14}}{\rightleftarrows}} Kdr^*{:}Gab1^*{:}Pip3{:}Pi3k$$
$$Kdr^*{:}Gab1^* + Pi3k \underset{k_{17}}{\overset{k_{16}}{\rightleftarrows}} Kdr^*{:}Gab1^*{:}Pi3k$$
$$Kdr^*{:}Gab1^*{:}Pi3k \overset{k_{18}}{\rightarrow} Kdr^*{:}Gab1^*{:}Pi3k^*$$
$$Kdr^*{:}Gab1^*{:}Pi3k^* + Pip2 \underset{k_{20}}{\overset{k_{19}}{\rightleftarrows}} Kdr^*{:}Gab1^*{:}Pi3k^*{:}Pip2$$
$$Kdr^*{:}Gab1^*{:}Pi3k^*{:}Pip2 \overset{k_{21}}{\rightarrow} Kdr^*{:}Gab1^*{:}Pi3k + Pip3$$
$$Kdr^*{:}Gab1^*{:}Pip3 + Pi3k \underset{k_{23}}{\overset{k_{22}}{\rightleftarrows}} Kdr^*{:}Gab1^*{:}Pip3{:}Pi3k$$
$$Kdr^*{:}Gab1^*{:}Pip3{:}Pi3k \overset{k_{24}}{\rightarrow} Kdr^*{:}Gab1^*{:}Pip3{:}Pi3k^*$$
$$Kdr^*{:}Gab1^*{:}Pip3{:}Pi3k^* + Pip2 \underset{k_{26}}{\overset{k_{25}}{\rightleftarrows}} Kdr^*{:}Gab1^*{:}Pip3{:}Pi3k^*{:}Pip2$$
$$Kdr^*{:}Gab1^*{:}Pip3{:}Pi3k^*{:}Pip2 \overset{k_{27}}{\rightarrow} Kdr^*{:}Gab1^*{:}Pip3{:}Pi3k + Pip3$$
$$Pip3 + Akt \underset{k_{29}}{\overset{k_{28}}{\rightleftarrows}} Pip3{:}Akt$$
$$Pip3{:}Akt \overset{k_{30}}{\rightarrow} Pip3 + Akt^*$$

Pip2 Regeneration (Second Block)

$$Pip3 + Pten \underset{k_{54}}{\overset{k_{53}}{\rightleftarrows}} Pip3{:}Pten$$
$$Pip3{:}Pten \overset{k_{55}}{\rightarrow} Pip2 + Pten$$
$$Pten + Pip2 \underset{k_{57}}{\overset{k_{56}}{\rightleftarrows}} Pten{:}Pip2$$
$$Pten{:}Pip2 + Pip3 \underset{k_{59}}{\overset{k_{58}}{\rightleftarrows}} Pten{:}Pip2{:}Pip3$$
$$Pten{:}Pip2{:}Pip3 \overset{k_{60}}{\rightarrow} Pten{:}Pip2 + Pip2$$
$$Dag + E \underset{k_{62}}{\overset{k_{61}}{\rightleftarrows}} Dag{:}E$$
$$Dag{:}E \overset{k_{63}}{\rightarrow} Pip2 + E$$

Proliferation (Fourth Block)

$$Kdr^* + Plc_\gamma \underset{k_{32}}{\overset{k_{31}}{\rightleftarrows}} Kdr^*{:}Plc_\gamma$$
$$Kdr^*{:}Plc_\gamma \overset{k_{33}}{\rightarrow} Kdr^*{:}Plc_\gamma^*$$
$$Kdr^*{:}Plc_\gamma^* + Pip2 \underset{k_{35}}{\overset{k_{34}}{\rightleftarrows}} Kdr^*{:}Plc_\gamma^*{:}Pip2$$
$$Kdr^*{:}Plc_\gamma^*{:}Pip2 \overset{k_{36}}{\rightarrow} Kdr^*{:}Plc_\gamma + Dag$$
$$Kdr^*{:}Gab1^* + Plc_\gamma \underset{k_{38}}{\overset{k_{37}}{\rightleftarrows}} Kdr^*{:}Gab1^*{:}Plc_\gamma$$
$$Kdr^*{:}Gab1^*{:}Plc_\gamma \overset{k_{39}}{\rightarrow} Kdr^*{:}Gab1^*{:}Plc_\gamma^*$$
$$Kdr^*{:}Gab1^*{:}Plc_\gamma^* + Pip2 \underset{k_{41}}{\overset{k_{40}}{\rightleftarrows}} Kdr^*{:}Gab1^*{:}Plc_\gamma^*{:}Pip2$$
$$Kdr^*{:}Gab1^*{:}Plc_\gamma^*{:}Pip2 \overset{k_{42}}{\rightarrow} Kdr^*{:}Gab1^*{:}Plc_\gamma + Dag$$
$$Kdr^*{:}Gab1^*{:}Pip3 + Plc_\gamma \underset{k_{44}}{\overset{k_{43}}{\rightleftarrows}} Kdr^*{:}Gab1^*{:}Pip3{:}Plc_\gamma$$
$$Kdr^*{:}Gab1^*{:}Pip3{:}Plc_\gamma \overset{k_{45}}{\rightarrow} Kdr^*{:}Gab1^*{:}Pip3{:}Plc_\gamma^*$$
$$Kdr^*{:}Gab1^*{:}Pip3{:}Plc_\gamma^* + Pip2 \underset{k_{47}}{\overset{k_{46}}{\rightleftarrows}} Kdr^*{:}Gab1^*{:}Pip3{:}Plc_\gamma^*{:}Pip2$$
$$Kdr^*{:}Gab1^*{:}Pip3{:}Plc_\gamma^*{:}Pip2 \overset{k_{48}}{\rightarrow} Kdr^*{:}Gab1^*{:}Pip3{:}Plc_\gamma + Dag$$
$$Gab1^*{:}Pip3 + Plc_\gamma \underset{k_{50}}{\overset{k_{49}}{\rightleftarrows}} Gab1^*{:}Pip3{:}Plc_\gamma$$
$$Gab1^*{:}Pip3 : Plc_\gamma + Kdr^* \underset{k_{52}}{\overset{k_{51}}{\rightleftarrows}} Kdr^*{:}Gab1^*{:}Pip3{:}Plc_\gamma^*$$

Fig. 2. Reactions of the detailed model

In the model Akt and DAG have been considered as the end points of the survival and proliferation pathways, respectively. Taking into account these end points in combination with the notion that Akt activation is strictly Pip_3-dependent, we examined the signal

transduction cascade focusing our attention mainly on the reactions that lead to the production of Pip_3 (i.e. k_{21} and k_{27}) and DAG (i.e. k_{36}, k_{42}, and k_{48}).

This analysis (supported also by a careful drawing of the SPN) allowed us to recognize different sub-pathways that lead to the survival or proliferation effects. In the context of the survival signal module we identified two sub-pathways, each one characterized by the presence of a distinguishing complex, KDR^*:$Gab1^*$ or KDR^*:$Gab1^*$:Pip_3, belonging to the First Block. Actually, the sub-pathways that determine the survival behavior are three since an additional element, $Gab1^*$:Pip_3, also contributes to the formation of KDR^*:$Gab1^*$:$Pi3k$:Pip_3 complex already involved in one of the identified sub-pathways. Summarizing there are three sub-pathways that lead to survival effect starting from: KDR^*:$Gab1^*$, KDR^*:$Gab1^*$:Pip_3 and $Gab1^*$:Pip_3. Considering the proliferation module, we identified four different sub-pathways that are distinguished by the compounds belonging to the First Block, i.e.: KDR^*, KDR^*:$Gab1^*$, KDR^*:$Gab1^*$:Pip_3 or $Gab1^*$:Pip_3. Notice that the distinguishing elements of the detected sub-pathways are the same within the survival and proliferation modules, with the exception of the compound KDR^*.

Referring again to the SPN of Fig. 3, we can notice that the time evolution of this SPN is intuitively portrayed by a top-down view. On the top is depicted the place KDR^* that represents the starting point of the signal cascade induced by its ligand. From the KDR^* cascade start all the sub-pathways that characterize the proliferation and survival pathways. The places describing DAG and Pip_3 are aligned on the bottom of the net. It is interesting to note that the sub-pathways we identified in the detailed model are represented in the SPN with structures, such as that outlined by a dashed box in Fig. 3, which correspond to the reaction groups separated by continuous lines in Fig. 2. We denote these Sub-Components by SC. Each SC involves:

- the binding between an enzyme and other species present in the cascade (e.g. transitions $k_{31}k_{32}$);
- the enzyme activation (e.g. transition k_{33});
- the recruitment of the Pip_2 (e.g. transitions $k_{34}k_{35}$);
- the production of the molecules representing the pathway end point and the enzyme deactivation (e.g. transition k_{36}).

3.3 Model Simplification

The SPN we obtained requires a simplification process to take place in order to limit the complexity of the parameterization and analysis of the model as we pointed out before. The computation of the P-semiflows of this SPN show that the net is bounded (the net is covered by P-semiflows, i.e., every place of the net is member of the support of one P-semiflow, at least). Interpreting the P-semiflows in biological terms, we can recall again that this means that all the compounds associated with the places of the net, independently of their original amounts, cannot grow indefinitely during the evolution of the model out of its initial state

The presence of repeated structures in the SPN corresponds to the fact that the biological model is characterized by the existence of several similar reaction groups, and this observation can be used to identify simplification steps to be applied to the detailed

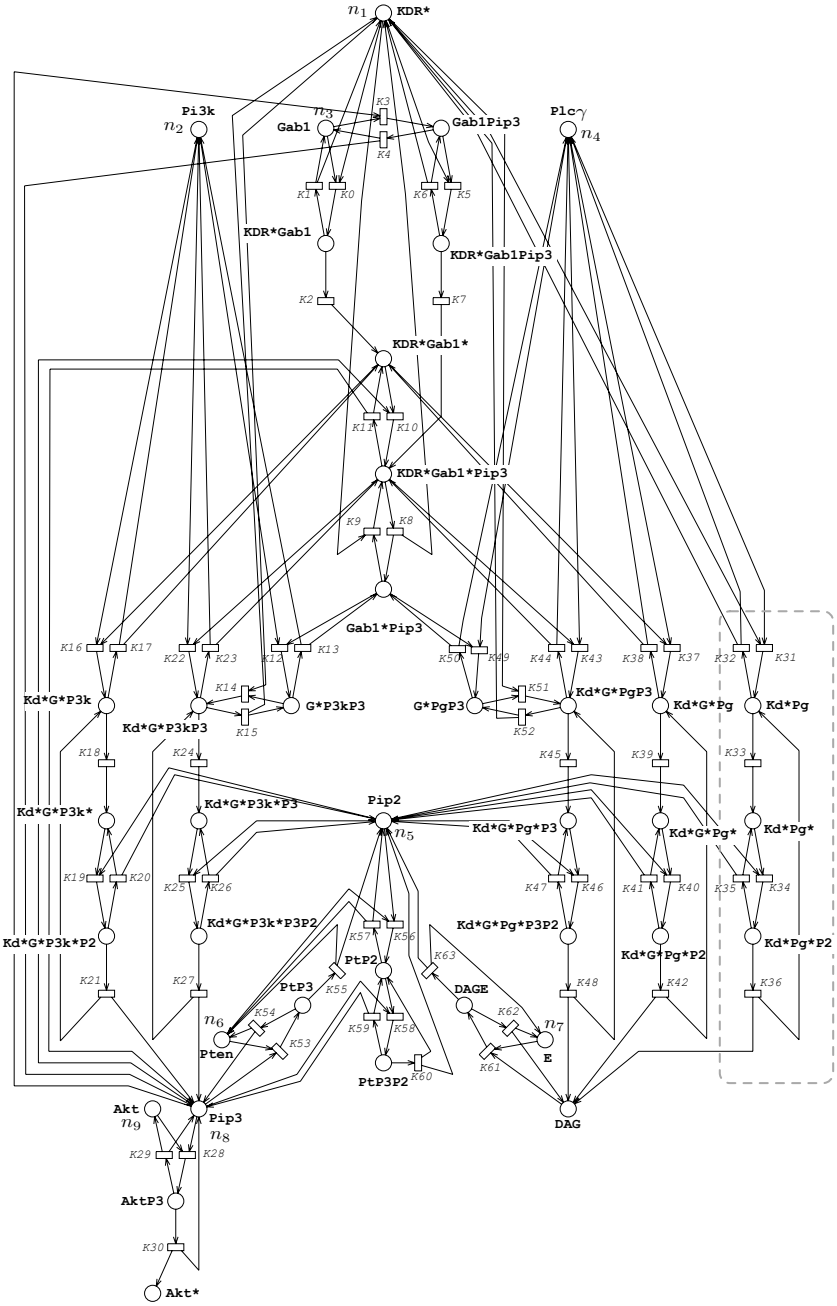

Fig. 3. SPN representing the detailed model. Compound symbols: $KDR \equiv Kd$, $Gab1 \equiv G$, $Pi3k \equiv P3k$, $Plc_\gamma \equiv Pg$, $Pip_3 \equiv P3$, $Pip_2 \equiv P2$, $Pten \equiv Pt$.

model. In particular, the SCs shown in the previous section are enzymatic kinetics reaction groups. Each of these groups can be written as the reactions set (1) where the enzyme E binds reversibly to the compound C.

$$\begin{aligned} E + C &\rightleftarrows EC \rightarrow E^*C \\ E^*C + S &\rightleftarrows E^*CS \\ E^*CS &\rightarrow EC + P \end{aligned} \tag{1}$$

This complex EC irreversibly becomes E^*C, which means that the enzyme is activated. E^*C binds reversibly to the substrate S (forming E^*CS), before converting it into a product P and releasing the complex EC. In order to simplify the detailed model we can represent each SC by the following couple of "merging" pseudo-reactions:

$$\begin{aligned} E + S &\rightarrow E^*S \\ E^*S + S' &\rightleftarrows E + S + P \end{aligned} \tag{2}$$

KDR-Receptor (First Block)	**Survival (Third Block)**
$Kdr^* + Gab1 \overset{k_0}{\underset{k_1}{\rightleftarrows}} Kdr^*{:}Gab1$ $Kdr^*{:}Gab1 \overset{k_2}{\rightarrow} Kdr^*{:}Gab1^*$ $Gab1 + Pip3 \overset{k_3}{\underset{k_4}{\rightleftarrows}} Gab1{:}Pip3$ $Kdr^* + Gab1{:}Pip3 \overset{k_5}{\underset{k_6}{\rightleftarrows}} Kdr^*{:}Gab1{:}Pip3$ $Kdr^*{:}Gab1{:}Pip3 \overset{k_7}{\rightarrow} Kdr^*{:}Gab1^*{:}Pip3$ $Kdr^*{:}Gab1^*{:}Pip3 \overset{k_8}{\underset{k_9}{\rightleftarrows}} Gab1^*{:}Pip3 + Kdr^*$ $Kdr^*{:}Gab1^* + Pip3 \overset{k_{10}}{\underset{k_{11}}{\rightleftarrows}} Kdr^*{:}Gab1^*{:}Pip3$	$Gab1^*{:}Pip3 + Kdr^* + Pi3k \overset{k_{12}}{\rightarrow} Kdr^*{:}Gab1^*{:}Pip3{:}Pi3k^*$ $Kdr^*{:}Gab1^*{:}Pip3{:}Pi3k^* + Pip2 \overset{k_{13}}{\rightarrow} Gab1^*{:}Pip3 + Kdr^* + Pi3k + Pip3$ ——— $Kdr^*{:}Gab1^* + Pi3k \overset{k_{14}}{\rightarrow} Kdr^*{:}Gab1^*{:}Pi3k^*$ $Kdr^*{:}Gab1^*{:}Pi3k^* + Pip2 \overset{k_{15}}{\rightarrow} Kdr^*{:}Gab1^* + Pi3k + Pip3$ ——— $Kdr^*{:}Gab1^*{:}Pip3 + Pi3k \overset{k_{16}}{\rightarrow} Kdr^*{:}Gab1^*{:}Pip3{:}Pi3k^*$ $Kdr^*{:}Gab1^*{:}Pip3{:}Pi3k^* + Pip2 \overset{k_{17}}{\rightarrow} Kdr^*{:}Gab1^*{:}Pip3 + Pi3k + Pip3$ ——— $Pip3 + Akt \overset{k_{18}}{\underset{k_{19}}{\rightleftarrows}} Pip3{:}Akt$ $Pip3{:}Akt \overset{k_{20}}{\rightarrow} Pip3 + Akt^*$
Pip2 Regeneration (Second Block)	**Proliferation (Fourth Block)**
$Pip3 + Pten \overset{k_{29}}{\underset{k_{30}}{\rightleftarrows}} Pip3{:}Pten$ $Pip3{:}Pten \overset{k_{31}}{\rightarrow} Pip2 + Pten$ ——— $Pten + Pip2 \overset{k_{32}}{\underset{k_{33}}{\rightleftarrows}} Pten{:}Pip2$ $Pten{:}Pip2 + Pip3 \overset{k_{34}}{\underset{k_{35}}{\rightleftarrows}} Pten{:}Pip2{:}Pip3$ $Pten{:}Pip2{:}Pip3 \overset{k_{36}}{\rightarrow} Pten{:}Pip2 + Pip2$ ——— $Dag + E \overset{k_{37}}{\underset{k_{38}}{\rightleftarrows}} Dag{:}E$ $Dag{:}E \overset{k_{39}}{\rightarrow} Pip2 + E$	$Kdr^* + Plc_\gamma \overset{k_{21}}{\rightarrow} Kdr^*{:}Plc_\gamma^*$ $Kdr^*{:}Plc_\gamma^* + Pip2 \overset{k_{22}}{\rightarrow} Kdr^* + Plc_\gamma^* + DAG$ ——— $Kdr^*{:}Gab1^* + Plc_\gamma \overset{k_{23}}{\rightarrow} Kdr^*{:}Gab1^*{:}Plc_\gamma^*$ $Kdr^*{:}Gab1^*{:}Plc_\gamma^* + Pip2 \overset{k_{24}}{\rightarrow} Kdr^*{:}Gab1^* + Plc_\gamma + DAG$ ——— $Kdr^*{:}Gab1^*{:}Pip3 + Plc_\gamma \overset{k_{25}}{\rightarrow} Kdr^*{:}Gab1^*{:}Pip3{:}Plc_\gamma^*$ $Kdr^*{:}Gab1^*{:}Pip3{:}Plc_\gamma^* + Pip2 \overset{k_{26}}{\rightarrow} Kdr^*{:}Gab1^*{:}Pip3 + Plc_\gamma + DAG$ ——— $Gab1^*{:}Pip3 + Kdr^* + Plc_\gamma \overset{k_{27}}{\rightarrow} Kdr^*{:}Gab1^*{:}Pip3{:}Plc_\gamma^*$ $Kdr^*{:}Gab1^*{:}Pip3{:}Plc_\gamma^* + Pip2 \overset{k_{28}}{\rightarrow} Gab1^*{:}Pip3 + Kdr^* + Plc_\gamma + DAG$

Fig. 4. Reactions after first step of simplification

By exploiting this representation, we rewrite the reactions of the Third and the Fourth Blocks as shown in Fig. 4, and we use them to simplify the original SPN obtaining the net depicted in Fig. 5, that is still covered by P-semiflows, meaning that these transformations are acceptable also from a qualitative point of view. Note that in this new

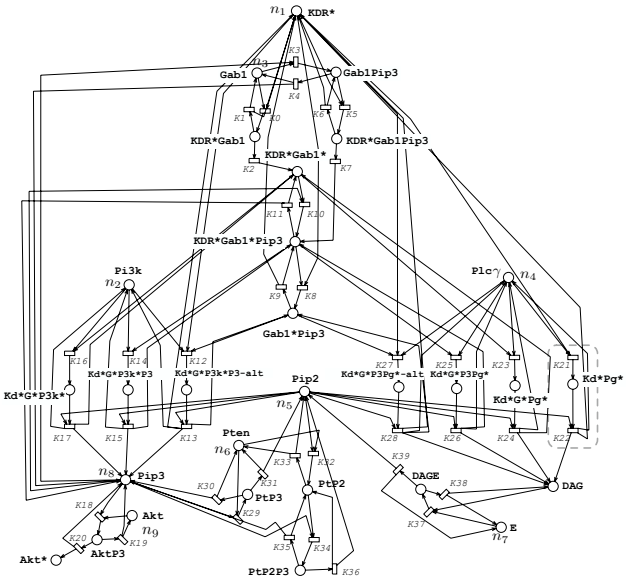

Fig. 5. SPN obtained after the first step of simplification (Simpl1)

SPN any SC is represented by a sequence of transition-place-transition, such as the one outlined by the dashed box, again. These SCs are defined by reaction groups separated by continuous line in Fig. 4. A further simplification step can be performed on the basis of the following reaction:

$$E + S + S' \rightarrow E + S + P \tag{3}$$

The application of this observation to reactions of the Third and Fourth Blocks provides the reactions illustrated in Fig. 6, used to define the new simplified net illustrated in Fig. 7 (which is again covered by P-semiflows). Note that in this last SPN any SC is represented by a singular transition such as the one outlined again by the dashed box. Moreover reactions $k_{29}k_{30}$, k_{31} and $k_{37}k_{38}$, k_{39} of the Second Block are simplified following the scheme: $E + S \rightarrow E + P$. The structural criteria that we used to guide the simplification process helped us to verify that the reduced models maintain the biological significance of the original one and provide a good approximation of its behavior. Notice that the reaction substitutions represented by Eq. (2 and 3) can be seen as patterns (or net substructures) that can be replaced every time they occur in the detailed and intermediate nets of the simplification process. In reality, it has not been possible to perform these transformations "mechanically". Instead, the invariant conditions had to be checked for each simplification substitution. At the end of this process, considering that the P-semiflows of the simplified SPNs have smaller supports, we found the same eight P-semiflows in all the three nets:

– one P-semiflow including KDR^* and the complexes containing it that are present in the sub-pathways (both in proliferation and in survival) that bring to the recruiting of substrate Pip_2;

KDR-Receptor (First Block)	Survival (Third Block)
$Kdr^* + Gab1 \underset{k_1}{\overset{k_0}{\rightleftarrows}} Kdr^*{:}Gab1$ $Kdr^*{:}Gab1 \underset{k_3}{\overset{k_2}{\rightleftarrows}} Kdr^*{:}Gab1^*$ $Gab1 + Pip3 \underset{k_4}{\overset{k_3}{\rightleftarrows}} Gab1{:}Pip3$ $Kdr^* + Gab1{:}Pip3 \underset{k_6}{\overset{k_5}{\rightleftarrows}} Kdr^*{:}Gab1{:}Pip3$ $Kdr^*{:}Gab1{:}Pip3 \underset{k_8}{\overset{k_7}{\rightleftarrows}} Kdr^*{:}Gab1^*{:}Pip3$ $Kdr^*{:}Gab1^*{:}Pip3 \underset{k_9}{\overset{k_8}{\rightleftarrows}} Gab1^*{:}Pip3 + Kdr^*$ $Kdr^*{:}Gab1^* + Pip3 \underset{k_{11}}{\overset{k_{10}}{\rightleftarrows}} Kdr^*{:}Gab1^*{:}Pip3$	$Gab1^*{:}Pip3 + Kdr^* + Pi3k + Pip2 \overset{k_{12}}{\rightarrow} Gab1^*{:}Pip3 + Kdr^* + Pi3k + Pip3$ $Kdr^*{:}Gab1^*{:}Pip3 + Pi3k + Pip2 \overset{k_{13}}{\rightarrow} Kdr^*{:}Gab1^*{:}Pip3 + Pi3k + Pip3$ $Kdr^*{:}Gab1^* + Pi3k + Pip2 \overset{k_{14}}{\rightarrow} Kdr^*{:}Gab1^* + Pi3k + Pip3$ $Pip3 + Akt \underset{k_{16}}{\overset{k_{15}}{\rightleftarrows}} Pip3{:}Akt$ $Pip3{:}Akt \overset{k_{17}}{\rightarrow} Pip3 + Akt^*$
Pip2 Regeneration (Second Block)	**Proliferation (Fourth Block)**
$Pip3 + Pten \overset{k_{22}}{\rightarrow} Pten + Pip2$ $Pten + Pip2 \underset{k_{24}}{\overset{k_{23}}{\rightleftarrows}} Pten{:}Pip2$ $Pten{:}Pip2 + Pip3 \underset{k_{26}}{\overset{k_{25}}{\rightleftarrows}} Pten{:}Pip2{:}Pip3$ $Pten{:}Pip2{:}Pip3 \overset{k_{27}}{\rightarrow} Pten{:}Pip2 + Pip2$ $Dag + E \overset{k_{28}}{\rightarrow} Pip2 + E$	$Kdr^* + Plc_\gamma + Pip2 \overset{k_{18}}{\rightarrow} Kdr^* + Plc_\gamma + DAG$ $Kdr^*{:}Gab1^* + Plc_\gamma + Pip2 \overset{k_{19}}{\rightarrow} Kdr^*{:}Gab1^* + Plc_\gamma + DAG$ $Kdr^*{:}Gab1^*{:}Pip3 + Plc_\gamma + Pip2 \overset{k_{20}}{\rightarrow} Kdr^*{:}Gab1^*{:}Pip3 + Plc_\gamma + DAG$ $Gab1^*{:}Pip3 + Kdr^* + Plc_\gamma + Pip2 \overset{k_{21}}{\rightarrow} Gab1^*{:}Pip3 + Kdr^* + Plc_\gamma + DAG$

Fig. 6. Reactions after second step of simplification

- one P-semiflow including $Gab1$ and the complexes containing it that are present in the sub-pathways (both in proliferation and in survival) that bring to the recruiting of substrate Pip_2;
- one P-semiflow including Akt and the complexes that lead to its activation;
- one P-semiflow including both Pip_3 and Dag, and Pip_2 that is the common substrate in both pathways. This semiflow includes also the cascade complexes containing Pip_3;
- each enzyme present in the model ($Pi3k$, $Plc\gamma$, $Pten$, E) has a semiflow including the complexes containing it.

The consistency among the structural properties of all the nets allowed us to consider the simplified models valid from a qualitative point of view.

3.4 Model Analysis for Accuracy Assessment

The simplification process proposed in Section 3.3 results in SPNs which maintain the qualitative properties of the original SPN, but are approximations of the detailed model from a quantitative point of view. In this section we report in silico experiments that were performed in order to check the validity of the simplifications from the point of view of quantitative properties. Indeed, before using the simplified models in a parameter identification experiment which uses real data coming from wet-lab experiments, it is necessary to make sure that an overall agreement exists between the quantitative temporal behaviours of the detailed and the simplified models for a wide range of model parameters. This test allows to build confidence on the fact that the reduced model is

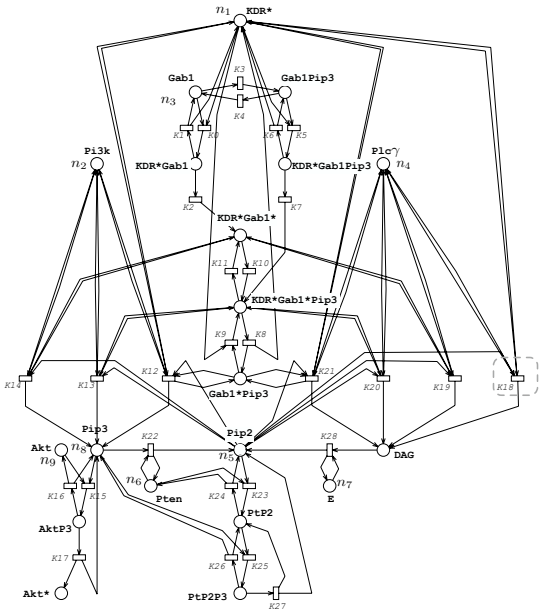

Fig. 7. SPN obtained after the second step of simplification (Simpl2)

suited for a preliminary analysis of the angiogenesis process as it can be done with in silico experiments.

This accuracy assessment was performed applying the "deterministic approach" described in Section 2.3. We compared the temporal behaviours of the detailed SPN with those of the simplified SPNs obtained for several different initial markings and several sets of transition rate parameters. Throughout the comparisons we concentrated on three important entities, in particular Pip_3 and Dag, and the substrate Pip_2 which is common to both the survival and the proliferation pathways. Hereafter we report on two cases which illustrate the obtained results. The initial condition is identical for the two cases: for all three SPNs (see Figures 3, 5 and 7) we use the initial marking $n_1 = 2$, $n_2 = n_3 = n_4 = 1$, $n_5 = 20$, $n_6 = n_7 = n_8 = n_9 = 1$ which reflects the concentration differences that are likely to exist in wet-lab experiments. Different sets of transition rates are used in the two cases to push the behaviours of the models in opposite directions. In the first case the rates are such that the transitions along the survival pathway are ten times faster than all the others. Figure 8A depicts the temporal behaviour for the detailed model. With these parameters the concentration of Pip_3 increases, the concentration of Pip_2 decreases and the concentration of DAG remains low. The temporal behaviour of the simplified models, depicted in Figures 8B-C, shows the same major characteristics. In the second case the transitions along the proliferation pathway are ten times faster than all the others. Figures 8E-F-G depict the temporal behaviour for the three models. Also in this case, the major characteristic, i.e. the fact that the concentration of DAG prevails over the concentration of Pip_2 and Pip_3, is maintained. The general agreement among the results of all these models was also tested by

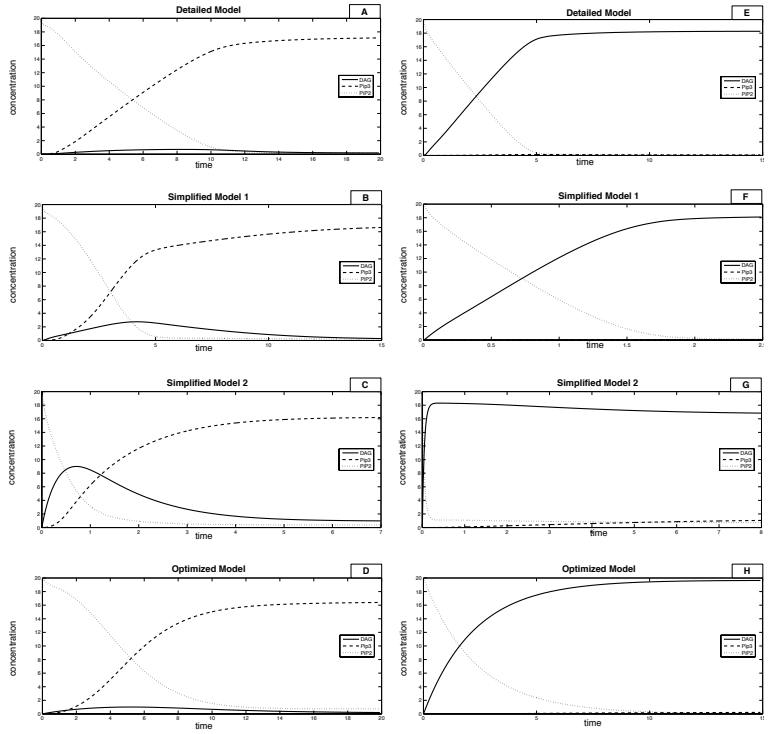

Fig. 8. The left (right) column shows the ODE results from the first (second) set of parameters

computing the steady state distribution of tokens in the places corresponding to compounds Pip_3 and DAG and to the substrate Pip_2, obtained from the solutions of the CTMCs corresponding to all the SPN models we have constructed (that we do not report in detail in this paper, due to space constraints).

For the first set of experiments reported here, the rates of corresponding transitions in the simplified and detailed models were set equal independently of the fact that in the simplified SPNs they often represent the compound effect of a few transitions of the detailed one. As a result, even if the overall characteristics are maintained, the shape of the curves can be rather different and the dynamics take place on different time scales. Focusing our attention on the curves corresponding to the first set of parameters (Figures 8A-B-C), we can notice that the crossing of the two concentration curves for Pip_2, which decreases, and Pip_3 that grows takes place at time instants that are not of the same order of magnitude in all the three cases. A more important difference, however, is observed when we concentrate on the dynamics of DAG that shows a small initial growth followed by a descent to a value next to zero. For this case, the most simplified model predicts an important initial climb that makes the shape of the curve quite different from the others (see Figure 8C). Turning our attention to the curves corresponding to the second set of parameters (Figures 8E-F-G), we can notice that all the models predict a crossing between the concentration of Pip_2 which decreases and the

concentration of DAG that instead grows. With these parameters the concentration of Pip_3 remains always extremely small. In this case the predictions of the more compact model are quantitatively quite different since the crossing is reported to happen much sooner and the shapes of the curves are very different.

In order to mimic better the temporal behaviour of the detailed SPN, we applied an optimization technique to determine the transition rates for the simplified SPNs. We used the nonlinear optimization function of MatLab to find these transition rates with which the curves resulting from the simplified SPNs became closer to the curves resulting from the detailed SPN. The results are illustrated for the SPN obtained after the second step of simplification (Simpl2) in Figure 8D and 8H. It can be seen that by properly setting the transition rates, even the simplest of our SPNs (Simpl2) can mimic quite precisely the behaviour of the detailed model (compare the diagrams A with D and E with H of Figure 8).

4 Discussion

In this paper we showed how to use structural properties of SPNs and biochemical properties of the system in guiding the simplification process. The procedure was presented through a case study, namely, the model of signaling transduction pathways involved in the angiogenesis process. We showed that the procedure results in simplified SPNs that are able to mimic precisely the temporal behavior of the detailed SPN.

One non trivial step is to determine the transition rates in the simplified SPNs in such a way that the resulting temporal behavior is a good approximation of that of the detailed SPN. In this work we faced this problem by applying optimization.

In the future, on the basis of the simplification schemes presented in this paper, we plan to work on generalizations of these reduction steps which will allow to operate on other portions of the detailed model and on the identification of rules concerning the relations existing among the corresponding rates of the detailed and simplified models. In addition, we will study the possibility of defining formally the quantitative characteristics that have to be maintained by the simplification process. In particular, temporal logics will be considered to this purpose.

Furthermore, we will consider the study of the whole VEGF-induced intracellular network, including signal modules that were not considered here, such as the migration pathway. This could contribute to a better understanding of the intricate signaling induced by VEGF-A during the angiogenesis.

Acknowledgment

Cordero is the recipient of research fellowship supported by grants from Italian Association for Cancer Research and Regione Piemonte.

References

1. Baarir, S., Beccuti, M., Cerotti, D., De Pierro, M., Donatelli, S., Franceschinis, G.: The GreatSPN tool: recent enhancements. SIGMETRICS Perform. Eval. Rev. 36(4), 4–9 (2009)
2. Balbo, G.: Introduction to stochastic petri nets. In: Brinksma, E., Hermanns, H., Katoen, J.-P. (eds.) EEF School 2000 and FMPA 2000. LNCS, vol. 2090, p. 84. Springer, Heidelberg (2001)

3. Busch, H., Sandmann, W., Wolf, V.: A numerical aggregation algorithm for the enzyme-catalyzed substrate conversion. In: Priami, C. (ed.) CMSB 2006. LNCS (LNBI), vol. 4210, pp. 298–311. Springer, Heidelberg (2006)
4. Calder, M., Vyshemirsky, V., Gilbert, D., Orton, R.: Analysis of Signalling Pathways Using Continuous Time Markov Chains. In: Priami, C., Plotkin, G. (eds.) Transactions on Computational Systems Biology VI. LNCS (LNBI), vol. 4220, pp. 44–67. Springer, Heidelberg (2006)
5. Dance, M., Montagner, A., Yart, A., Masri, B., Audigier, Y., Perret, B., Salles, J.P., Raynal, P.: The adaptor protein Gab1 couples the stimulation of vascular endothelial growth factor receptor-2 to the activation of phosphoinositide 3-kinase. Journal of Biological Chemestry 281, 23285–23295 (2006)
6. Gerber, H.P., McMurtrey, A., Kowalski, J., Yan, M., Keyt, B.A., Dixit, V., Ferrara, N.: Vascular endothelial growth factor regulates endothelial cell survival through the phosphatidylinositol 3'-kinase/Akt signal transduction pathway. requirement for Flk-1/KDR activation. Journal of Biological Chemestry 273, 30336–30343 (1998)
7. Gillespie, D.T.: A rigorous derivation of the master chemical equation. Physica 188, 404–425 (1992)
8. Glbert, D., Heiner, M., Lehrack, S.: A unifying framework for modelling and analysing biochemical pathways using Petri nets. In: Proc. Int. Conf. Computational Methods in System Biology, pp. 200–216 (2007)
9. Goss, P., Pecoud, J.: Quantitative modeling of stochastic systems in molecular biology by using stochastic Petri nets. Proc. Natl. Acad. Sci. 95(12), 6750–6755 (1998)
10. Heiner, M., Koch, I., Will, J.: Model validation of biological pathways using Petri nets demonstrated for apoptosis. BioSystems 75, 10–28 (2004)
11. Hofestädt, R.: A Petri net application of metabolic processes. Journal of System Analysis, Modeling and Simulation 16, 113–122 (1994)
12. Hofestädt, R., Thelen, S.: Quantitative modeling of biochemical networks. Silico Biology 1(6) (1998)
13. Laramée, M., Chabot, C., Cloutier, M., Stenne, R., Holgado-Madruga, M., Wong, A.J., Royal, I.: The scaffolding adapter Gab1 mediates Vascular Endothelial Growth factor signaling and is required for endothelial cell migration and capillary formation. Journal of Biological Chemistry 282, 7758–7769 (2007)
14. Memmi, G., Vautherin, J.: Advanced algebraic techniques. In: Brawer, W., Reisig, W., Rozenberg, G. (eds.) APN 1986, Part I. LNCS, vol. 254. Springer, Heidelberg (1987)
15. Molloy, M.K.: On the integration of delay and throughput measures in distributed processing models. Ph.D. Thesis, UCLA (1981)
16. Natkin, S.: Les réseaux de Petri stochastiques et leur application à l'évaluation des systèmes informatiques. Thèse de Docteur Ingégneur, CNAM (1980)
17. Reddy, V., Mavrovouniotis, M., Liebman, M.: Petri net representation in metabolic pathways. In: Proc. Int. Conf. Intelligent Systems for Molecular Biology, pp. 328–336 (1993)
18. Reisig, W.: A Primer in Petri Net Design. Springer Compass International, Heidelberg (1992)
19. Takahashi, T., Ueno, H., Shibuya, M.: VEGF activates protein kinase C-dependent, but Ras-independent Raf-MEK-MAP kinase pathway for DNA synthesis in primary endothelial cells. Oncogene 18, 2221–2230 (1999)