Biochimica et Biophysica Acta 1844 (2014) 289–298

Contents lists available at ScienceDirect

Biochimica et Biophysica Acta

journal homepage: www.elsevier.com/locate/bbapap

Hybrid modeling of the crosstalk between signaling and transcriptional networks using ordinary differential equations and multi-valued logic ☆

Faiz M. Khan [a], Ulf Schmitz [a], Svetoslav Nikolov [c], David Engelmann [b], Brigitte M. Pützer [b], Olaf Wolkenhauer [a,d,1], Julio Vera [a,e,*,1]

[a] Department of Systems Biology and Bioinformatics, Institute of Computer Science, University of Rostock, Rostock, Germany
[b] Institute of Experimental Gene Therapy and Cancer Research, Rostock University Medical Center, Rostock, Germany
[c] Bulgarian Academy of Sciences, Institute of Mechanics, Sofia, Bulgaria
[d] Stellenbosch Institute for Advanced Study (STIAS), Wallenberg Research Centre at Stellenbosch University, Stellenbosch, South Africa
[e] Laboratory of Systems Medicine, Department of Dermatology, Faculty of Medicine, University of Erlangen-Nurnberg, Erlangen, Germany

ARTICLE INFO

Article history:
Received 1 March 2013
Received in revised form 9 May 2013
Accepted 10 May 2013
Available online 18 May 2013

Keywords:
Signaling
E2F1
Non-linear motif
Feedback loop
Chemoresistance
Cancer

ABSTRACT

A decade of successful results indicates that systems biology is the appropriate approach to investigate the regulation of complex biochemical networks involving transcriptional and post-transcriptional regulations. It becomes mandatory when dealing with highly interconnected biochemical networks, composed of hundreds of compounds, or when networks are enriched in non-linear motifs like feedback and feedforward loops. An emerging dilemma is to conciliate models of massive networks and the adequate description of non-linear dynamics in a suitable modeling framework. Boolean networks are an ideal representation of massive networks that are humble in terms of computational complexity and data demand. However, they are inappropriate when dealing with nested feedback/feedforward loops, structural motifs common in biochemical networks. On the other hand, models of ordinary differential equations (ODEs) cope well with these loops, but they require enormous amounts of quantitative data for a full characterization of the model. Here we propose hybrid models, composed of ODE and logical sub-modules, as a strategy to handle large scale, non-linear biochemical networks that include transcriptional and post-transcriptional regulations. We illustrate the construction of this kind of models using as example a regulatory network centered on E2F1, a transcription factor involved in cancer. The hybrid modeling approach proposed is a good compromise between quantitative/qualitative accuracy and scalability when considering large biochemical networks with a small highly interconnected core, and module of transcriptionally regulated genes that are not part of critical regulatory loops. This article is part of a Special Issue entitled: Computational Proteomics, Systems Biology & Clinical Implications. Guest Editor: Yudong Cai.

1. Introduction

In cells, biological processes are driven by complex networks integrating genes, transcripts, like mRNAs and microRNAs (miRNAs), proteins and small molecules. Concentrations and activity of these biomolecules are continuously changing in a concerted manner in response to internal and external cell signals. Those changes are regulated by multiple nested biological circuits that may contain feedback and feedforward loops. These non-linear biological circuits give rise to important cell features like robustness against noise, adaptation despite environmental changes or hysteretic responses with multistability [1]. In the last decade it has been found that some biochemical networks are extremely large and complex: they are commonly integrated by hundreds of compounds and are enriched in non-linear motifs like feedback and feedforward loops. Under these conditions, the analysis of biochemical networks evades human intuition. However, mathematical modeling is an appropriate tool to help in understanding those networks [2–4]. Depending on the features of the biochemical network under consideration and the available experimental data, a variety of frameworks for mathematical modeling are available. These range from simple and abstract approaches to biologically detailed ones, from deterministic to probabilistic or from spatio-temporal continuous to discrete ones [5].

Models in ordinary differential equations (ODEs), accounting for the variation in time of variables representing the concentration or activation state of biochemical molecules, have been used to describe biological networks for decades [6]. These ODE models allow the quantitative simulation of the concentration changes or activity profiles of proteins, genes, RNAs and other molecules over time. These features make it possible to compare the model predictions with most of the

☆ This article is part of a Special Issue entitled: Computational Proteomics, Systems Biology & Clinical Implications. Guest Editor: Yudong Cai.

* Corresponding author at: Department of Systems Biology and Bioinformatics, Institute of Computer Science, University of Rostock, Rostock, Germany. Tel.: +49 16095759623.
E-mail address: julio.vera.g@gmail.com (J. Vera).

[1] Equal contributors.

1570-9639/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.bbapap.2013.05.007

standard experimental measurements. When biological compartments are considered, they can also account for a qualitative description of spatial features [7]. Finally, ODE models are the almost "natural" modeling framework when investigating the features of biochemical networks containing regulatory circuits like feedback- and feedforward loops [8]. However, a complete characterization of the model requires to specify the values of several kinds of model parameters (for example, initial protein concentration and rate constants), which in biochemical mid-size networks becomes a computationally intensive task that requires large amounts of quantitative experimental data and sophisticated optimization algorithms. For larger networks parameter estimation becomes cumbersome and difficulties to identify unique values for model parameters emerge [9–11].

An alternative to the ODE models is the *discrete modeling* approach which allows the modeling of interactions among a large number of proteins, genes and other biomolecules to analyze system behavior and make predictions on biologically relevant scenarios [12–14]. It is a qualitative approach that depends only on the network structure (i.e. parameter free) with the simplifying assumption that networks nodes, accounting for the expression level or activation state of biological molecules, exist only in a well-defined set of possible discrete numerical values [10]. Boolean models are the simplest discrete models, in which each element of the network can have one of two possible states at any time (1: ON, expressed or active; 0: OFF, non-expressed or inactive). This assumption is supported by biological evidences, e.g. when genes/proteins exhibit ON/OFF switch like behavior. This can be adopted by assuming a reduced set of biologically meaningful values for the model variables [15,16,10]. A number of recent papers illustrate how discrete logic models nicely capture the behavior of large systems where the interactions are simple [17–19]. However, in some cases discrete logic models do not reproduce some of the time-dependent features associated to non-linear biological circuits, like those containing nested feedback loops.

The recent development of new experimental techniques facilitates high-throughput quantitative proteomics measurements and increases the complexity of the biological networks under investigation. Therefore, it is necessary to find suitable modeling strategies that realize a compromise between the ability to simulate the behavior of large biochemical networks, the complexity associated to the existence of multiple regulatory loops in them and the diversity of sources of experimental data. Hybrid models are mathematical and computational constructs that combine, in a single modeling framework, interdependent variables that distributed over discrete/continuous or deterministic/stochastic domains [20]. Hybrid models account, in an integrative manner, for different spatio-temporal scales of the same biological phenomenon, which are described using different, interconnected modeling frameworks. This approach seems especially valuable to deal with the multifactorial and multi-level nature of biological phenomena like cancer emergence and progression [21].

Here we propose, discuss and analyze the construction of hybrid models, composed of ODE and logic sub-modules, as a strategy to handle large scale, non-linear biochemical networks associated to cancer and other complex diseases. We illustrate the construction of this kind of models using as example a regulatory network centered in E2F1, a key cancer-related transcription factor.

2. Material and methods

For the implementation of our hybrid modeling approach we used built-in ODE solvers from Matlab to simulate the dynamics of the core regulatory module (MathWorks, MA) and CellNetAnalyzer (CNA) to simulate the logical part of the target genes and phenotypical read-outs [14]. To connect both subparts, we develop a Matlab script that discretizes the values of critical variables of the core regulatory model and uses these values as the input vector for the simulations with CNA. The output of the model simulations (continuous and discrete variable values) is stored in a matrix, whose content is visualized using color coded surface plots.

3. Results

3.1. Methodology proposed

Here, we address the problem of modeling large biochemical networks involving dozens to hundreds of receptors, kinase proteins, transcription factors, mRNAs and microRNAs. Their expression and activity are typically regulated by multiple, cross-talking pathways that may even contain overlapping regulatory loops. Our strategy is to organize and divide the network into three parts with distinctive regulatory features (Fig. 1A): 1) the *core regulatory module* of the network, a highly interconnected subnetwork that contains the signaling and transcriptional pathways enriched in feedback and feedforward loops, and therefore is expected to display a high non-linear behavior; 2) the *target genes module*, which accounts for dozens to hundreds of genes whose expression is directly regulated by the proteins and transcription factors included in the core regulatory module and can be experimentally quantified using high-throughput transcriptomics and proteomics techniques; and 3) a set of *phenotypical read-outs*, phenomenological variables accounting for relevant phenotypes triggered upon activation of groups of genes in the target genes module. Input signals external to the networks are assumed to regulate or interact with key compounds of the core regulatory module; although in some cases they could also be direct regulators of the target genes.

In our approach the core regulatory network is modeled using ordinary differential equations, while the target genes module and the phenotypical read-outs are encoded using discrete logic modeling (Boolean or multi-valued logic). Both parts of the model are connected using a discretization interface. Given the set of continuous time-dependent variables of the core regulatory module ($X_{i,C}, i = 1,2,\ldots k$), we define a set of auxiliary discrete variables ($X_{i,D}, i = 1,2,\ldots k$) for the discretization interface. Furthermore, we discretize the time by considering a discrete and finite set of p time points within the duration of the simulation ($t_j, j = 1,2,\ldots p$). For every time point considered, the interface assigns values to the auxiliary discrete variables, $X_{i,D}(t_j)$, following a set of discretization rules that depend on the current values of the continuous variables, $X_{i,C}(t_j)$, and n_i physiologically relevant thresholds, $TH_{i,m}, m = 1,2,\ldots n_i$. The number of thresholds for each continuous variable defines the number of ordinal states for the auxiliary discrete variables. The structure of the discretization rules is displayed in Table 1.

To simulate the whole model, first the core module is initialized and simulated with a set of values defined for the model inputs (Fig. 1). At each of the p time points defined, the values of the auxiliary discrete variables, $X_{i,D}(t_j)$, are calculated following the discretization rules. Those discrete values are passed as inputs to the logical module accounting for the target genes and are used to update the values of its discrete variables following a logical steady-state analysis. Subsequently, the updated values for the model variables in the transcriptional circuitry module are used to update the phenotypical read-outs. In this manner and for every discrete time point (t_j), the model simulation generates: 1) a set of continuous values for the variables of the regulatory core module, $X_{i,C}(t_j)$; 2) a set of discrete values for the auxiliary variables, $X_{i,D}(t_j)$; 3) values for the discrete variables accounting for the expression of the target genes, $Tgt_i(t_j)$; and 4) Boolean values for the phenotypic read-outs (Fig. 1A).

This approach has the advantage that the complex and highly regulated part of the network, which is enriched in regulatory loops is encoded with a modeling approach that provides many computational and analytical tools to investigate its properties [7,22]. On the other hand, the use of logic modeling to describe the associated target genes allows the simulation of massive transcriptional networks with dozens to hundreds of biologically relevant genes with low computational burden [23,24]. In this line, the data produced in high-throughput transcriptomics