Gene 533 (2014) 379–384

Contents lists available at ScienceDirect

Gene

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

Methods paper

Biological networks 101: Computational modeling for molecular biologists

Jetse Scholma [a], Stefano Schivo [b], Ricardo A. Urquidi Camacho [a], Jaco van de Pol [b], Marcel Karperien [a], Janine N. Post [a,*]

[a] Developmental BioEngineering, MIRA Institute for Biomedical Technology and Technical Medicine, University of Twente, 7522NH Enschede, The Netherlands
[b] Formal Methods and Tools, Faculty of EEMCS, University of Twente, 7522NH Enschede, The Netherlands

ARTICLE INFO

Article history:
Accepted 3 October 2013
Available online 12 October 2013

Keywords:
Signal transduction
Gene expression
Computational modeling
Biologist
Chondrocyte

ABSTRACT

Computational modeling of biological networks permits the comprehensive analysis of cells and tissues to define molecular phenotypes and novel hypotheses. Although a large number of software tools have been developed, the versatility of these tools is limited by mathematical complexities that prevent their broad adoption and effective use by molecular biologists. This study clarifies the basic aspects of molecular modeling, how to convert data into useful input, as well as the number of time points and molecular parameters that should be considered for molecular regulatory models with both explanatory and predictive potential. We illustrate the necessary experimental preconditions for converting data into a computational model of network dynamics. This model requires neither a thorough background in mathematics nor precise data on intracellular concentrations, binding affinities or reaction kinetics. Finally, we show how an interactive model of crosstalk between signal transduction pathways in primary human articular chondrocytes allows insight into processes that regulate gene expression.

1. Introduction

In the past years it has become increasingly clear that a thorough understanding of cells, tissues and disease pathologies is a prerequisite for the development of effective therapies and drugs (Lenas et al, 2009a, b). This insight is causing a gradual shift towards systems biology, that aims to unravel the molecular mechanisms of biological systems as a whole, rather than focusing on the individual components. Systems biology follows an empirical cycle in which computational modeling plays an important role (Fig. 1). To support biologists in the construction of computational models, computer scientists have contributed a large number of software tools, gathered for instance in the SBML database, listing over 250 software packages (SBML). Still, modeling is applied only sparsely in biological research. In this paper we will provide experimental biologists with guidelines for model construction.

The research question provides boundary conditions for the level of detail that is required in the model. The most precise models are detailed descriptions of reaction mechanisms, including the complete sequence of elementary reactions. These models are usually based on mass action kinetics and reaction steps are mathematically described using ordinary differential equations. Time and concentrations are continuous in the model, and the behavior of the network can be solved analytically for small models or can be evaluated numerically. Copasi is an example of a supporting tool that enables the construction and analysis of ODE models (Hoops et al., 2006). Although ODE models provide useful insights in biology (Kogan et al., 2012), they are parameter-intensive and require data on intracellular concentrations, binding affinities and reaction kinetics. Therefore, ODE models are often restricted to single pathways or small subnetworks. Being centered on mathematical equations, ODE models are not intuitively accessible without a mathematical background and/or previous modeling experience. On the other side of the spectrum are logic-based models, such as Boolean models, which can be evaluated in a stepwise manner. These models abstract from the details of reaction mechanisms, concentrations and time. “When A is active, B becomes active” is an example of an interaction in such models. The simplicity of the interactions makes these models suitable for construction of very large networks that can qualitatively capture biological phenomena surprisingly well (Kerkhofs et al., 2012). Boolean networks can for example be constructed using GINsim (Naldi et al., 2009).

Between these two extremes is a large range of models that abstract from biological reality with respect to reaction mechanisms, time or concentrations. For research questions concerning networks of signaling pathways that lead to gene expression patterns, a well-chosen abstraction level shall enable the construction of models that preserve the relevance of timing information and support multi-level concentrations in the study of biological systems. In this paper we focus on such timed

Abbreviations: ANIMO, Analysis of Networks with Interactive MOdeling; *AXIN2*, Axin 2 gene product; *B2M*, beta-2-microglobulin mRNA; ERK, Extracellular Signal-Related Kinase 1, MAPK3; FZD, Frizzled; GINsim, Gene Interaction Network simulation; hrs, hours; IL-1b, interleukin 1, beta protein; *IL-1B*, interleukin 1, beta mRNA; IL-1BR, interleukin 1, beta receptor; JNK, c-Jun N-terminal kinase protein; KEGG, Kyoto Encyclopedia of Genes and Genomes; MAPK, mitogen-activated protein kinase; MMP, matrix metalloprotease; ODE, Ordinary Differential Equation; PP96, Proteome profiler 96; Wnt, wingless-type MMTV integration site family.

* Corresponding author. Tel.: +31 53 4894205; fax: +31 53 489 2319.
E-mail address: j.n.post@utwente.nl (J.N. Post).

0378-1119/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.gene.2013.10.010

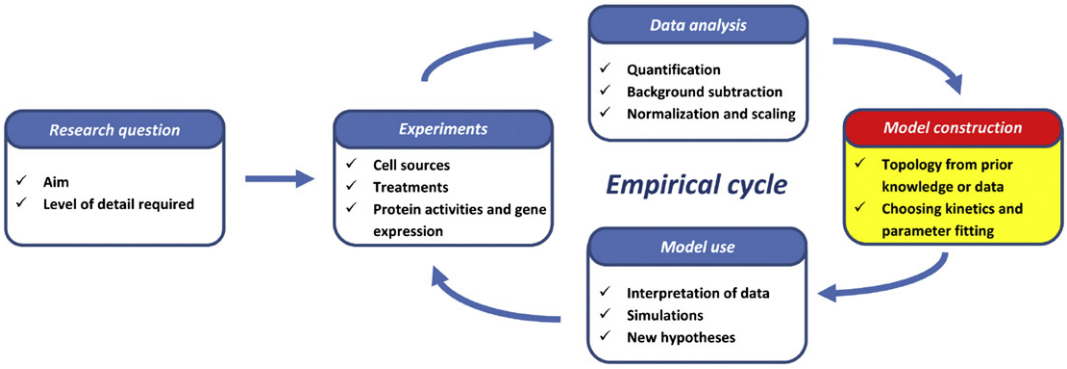

Fig. 1. Workflow describing the empirical cycle in systems biology. Starting from a research question, experiments and data provide input for a model, which in turn can be used for *in silico* experiments and the generation of new hypotheses.

multi-level models with abstracted reaction kinetics. Example tools for such models include ANIMO (Schivo et al., 2012a,b) and Cell Illustrator (Nagasaki et al., 2003). We use ANIMO (Box 1) as the modeling tool in this paper, as it was developed for use by biologists who have no prior modeling experience. Furthermore, the visualization stays as close as possible to biological traditions and it provides a good match with the level of detail found in most experimental data.

The next section of this paper briefly addresses aspects of the experimental design that need to be considered when the aim is to generate data to construct a model. Then, we will give detailed descriptions of the steps from experimental data to an *in silico* model. This section is followed by the construction of a small model based on our own research, illustrating how a model can be used to generate new hypotheses.

2. Experimental preconditions

Many biological events can be interpreted as changes in activity. For example, changes in concentration, phosphorylation or localization of a protein, or changes in gene expression are causal factors with respect to downstream effects. As such, the state or concentration of the molecules involved can be described in terms of an activity. The more active the molecule is, the stronger it will affect downstream processes. We propose a number of guidelines for experimental design: (1) In the process of choosing the most suitable molecules to measure, include molecules that either have downstream effects in the model or can be used as an output of the model. Successive iterations of the empirical cycle can be performed to expand an existing model (Fig. 1). (2) Inclusion of overlapping treatment conditions can be used to normalize experimental data between different days or assay batches. (3) For each of the measurements, a positive control that gives an indication of the potential maximum intensity in the biological system needs to be included. In this way, activity of data can be scaled between 0 and 100% to construct a nondimensional model, omitting the need for precise intracellular concentrations. (4) A negative control ($t = 0$) gives insight in background activity levels.

Single time-point measurements give poor insight in the dynamic behavior of the system. To decide how many time points should be measured and what the optimum time range is, the following factors need consideration. Ideally, measurements are obtained at time points starting from $t = 0$ until the system reaches a steady state. For most primary effects in signal transduction networks, this means measuring more time-points in the first 2–30 min after stimulation. When peak dynamics are expected, 3 time points are the absolute minimum to describe each peak, one before the peak, one as close as possible to the actual peak and one after the peak. Five time-points and more allow finding, and describing a peak in more detail, especially in the presence of experimental noise. If no peak dynamics are expected, at least 4 time points should be measured. Try to avoid having the highest measurement value as the first or last value in your time series, as it will lead to uncertainty about the actual behavior of the system.

Effects of an experimental treatment can be categorized as primary (or direct) effects or higher order (or indirect) effects. The latter are effects in which feedback is involved. For signal transduction, it is often sufficient to have time points up to 240/480 min for primary effects. Primary effects on gene expression typically take 4–12 h. When you are interested in higher order effects, the time range of these effects has to be taken into account. For signal transduction this can mean measurements up to 24/48 h; for gene expression involving higher order effects, e.g. in the case of cell differentiation, effects can take up to several weeks. The corresponding effort to understand the whole chain of events leading to a specific endpoint rapidly increases when longer time-courses and higher order effects are to be captured in a model.

Box 1

ANIMO: Analysis of Networks with Interactive MOdeling

Cytoscape (Shannon et al., 2003) has been designed for the visualization of (static) molecular interaction networks. ANIMO, a plugin to Cytoscape, was recently developed to turn Cytoscape networks into dynamic functional models (Schivo et al., 2012a,b). These dynamics are introduced by enriching interactions in the network with cause-and-effect relationships, such as "A activates B". Each interaction requires a single parameter that determines the speed or strength of the interaction. ANIMO works with non-dimensional data that are scaled to a hypothetical maximum, for instance the maximum experimental intensity. Every node in the network represents the whole population of a molecular species and is characterized by its current activity level. Upstream molecules exert their downstream effects as long as their current activity level is greater than zero, whereas the activity of downstream molecules is influenced through incoming interactions. User-input is automatically translated into an underlying formal model based on Timed Automata (Schivo et al., 2012a). ANIMO is suited for construction of timed models with 2–100 activity levels and timing can be abstracted to time steps instead of real–time. As such, it covers a modeling area between Boolean models and ODE based models.

3. Modeling 101

Constructing a model in ANIMO starts with drawing a network topology. This topology consists of nodes, corresponding to molecules, and arrows, corresponding to interactions (activations or inhibitions).