Symbolic Model Checking of Biochemical Networks

Nathalie Chabrier and François Fages*

Projet Contraintes, INRIA-Rocquencourt,
BP105, 78153 Le Chesnay Cedex, France,
{Nathalie.Chabrier,Francois.Fages}@inria.fr

Abstract. Model checking is an automatic method for deciding if a circuit or a program, expressed as a concurrent transition system, satisfies a set of properties expressed in a temporal logic such as CTL. In this paper we argue that symbolic model checking is feasible in systems biology and that it shows some advantages over simulation for querying and validating formal models of biological processes. We report our experiments on using the symbolic model checker NuSMV and the constraint-based model checker DMC, for the modeling and querying of two biological processes: a qualitative model of the mammalian cell cycle control after Kohn's diagrams, and a quantitative model of gene expression regulation.

1 Introduction

In recent years, Biology has clearly engaged an elucidation work of high-level biological processes in terms of their biochemical basis at the molecular level. The mass production of post genomic data, such as ARN expression, protein production and protein-protein interaction, raises the need of a strong parallel effort on the formal representation of biological *processes*. Metabolism networks, extracellular and intracellular signaling pathways, and gene expression regulation networks, are very complex dynamical systems. Annotating data bases with qualitative and quantitative information about the dynamics of biological systems, will not be sufficient to integrate and efficiently use the current knowledge about these systems. The design of formal tools for *modeling* biomolecular processes and for *reasoning* about their dynamics seems to be a mandatory research path to which the field of formal verification in computer science may contribute a lot.

Several formalisms have been proposed in recent years for the modeling of biochemical networks. Regev and Shapiro [22] were the first to propose the use of a formal concurrent language, namely Milner's π-calculus, for the modeling of a biochemical processes such as the RTK/MAPK pathway. The bio-calculus of [21] introduces a more biology-oriented syntax for a similar calculus. More

* This work has been done in the framework of the INRIA Cooperative Research Action "Process Calculi and Biology of Molecular Networks", ARC CPBIO, http://contraintes.inria.fr/cpbio

C. Priami (Ed.): CMSB 2003, LNCS 2602, pp. 149–162, 2003.

recently, quantitative modelings of biochemical processes have been developed with hybrid Petri nets [19, 16], hybrid concurrent constraint languages [4], and hybrid automata [1, 14].

In this paper we propose to go beyond simulation and to focus on the issue of providing automated methods for *querying and validating formal models* in systems biology. More specifically, we propose,

- first, the use of the temporal logics CTL as a query language for models of biological processes,
- second, the use of concurrent transition systems for the modeling of biological processes,
- and third, the use of symbolic model checking techniques for automatically evaluating CTL queries in both qualitative and quantitative models.

Our approach will be illustrated by two examples: a qualitative model of the mammalian cell cycle control after Kohn's diagrams [6, 17], and a quantitative model of gene expression.

1.1 Example 1: The Mammalian Cell Cycle Control

In this example, the main actors are genes, proteins with their phosphorylation sites, multimolecular complexes, and membranes. The molecules interact together to produce new proteins (synthesis), form multimolecular complexes (complexation), modify proteins (phosphorylation and dephosphorylation) degrade or transport molecules.

The cell cycle in eukaryotes is divided into four phases. Between two cell divisions, the cell is in a gap phase called G_1. The synthesis phase S starts with the replication of the nucleus. A second gap phase G_2 precedes the fourth phase: the mitose phase M during which the cell divides into two cells. The gap phase G_1 is mainly responsible for the duration of the cell cycle, it is in fact a growing phase of the cell and may contain a quiescent phase G_0 in which the cell can stay for long period of time or forever (stable state) without further division. Each phase is characterized by the activity of two major types of proteins: cyclins and cyclin-dependent kinases (Cdk). Cdk activity requires binding to a cyclin, and is controlled by specific inhibitors and by stimulatory or inhibitory phosphorylations by several kinases or phosphatases which in turn may produce positive feedback loops.

A state of the cell is defined by the values of the actors: either the presence or absence of molecules, or their number, or their concentration in each part of the cell, and by general data like the pH and the temperature. Note that a set of states can be just represented by partial information on the actual values of state variables, like for instance intervals or constraints between variables.

The biological queries one can consider about the cell cycle control are of different kinds: