Modeling, inference and simulation of biological networks using Constraint Logic Programming (CLP)

Eric Fanchon (1), Fabien Corblin (1,2) and Laurent Trilling (2)
(1) Institut de Biologie Structurale, CNRS-CEA-U. Joseph Fourier, 41 rue J. Horowitz, 38027 Grenoble Cedex 1. (2) LSR-IMAG, U. Joseph Fourier, BP 53, 38041 Grenoble Cedex 9.

Biology is now entering a new era in which molecular components have to be integrated into a *system* in order to reach new levels of understanding. Our objective consists in developing a computing tool allowing on one hand to infer models from properties which can be incomplete and qualitative, on the other hand to perform simulations or predictions starting from these (partially known) models. Such a tool should allow biologists to specify a network from the available data in order to obtain a class of models consistent with the data. More generally, the tool should be highly flexible to support the *exploration* of model properties in the context of incomplete knowledge.

The concept of interaction network is a fundamental one in systems biology. Our work is based on the "asynchronous multivalued logic networks" proposed by R. Thomas, E. H. Snoussi *et al*. (1,2). This formalism has been used to model genetic, neuronal and immunological networks. Formally, it can be viewed as a discrete abstraction of a special class of Piecewise-Linear Differential Equations (PLDEs). It allows a qualitative analysis of the dynamical behavior of such differential systems. Another benefit of this type of formalism lies in the discreteness which lends itself very well to computational implementations. The interaction graph associated to the PLDE system defines the architecture of the network. The parameters characterize the strength of the (non-linear) interactions. Recently, this formalism has been extended by de Jong *et al.* (3) to take into account the so-called 'singular states' and 'sliding modes'. Singular states are states of reduced dimensionality located at thresholds or intersection of thresholds, and sliding modes are trajectories that slide along a threshold (or intersection of thresholds). This extended formalism is sound in the sense that every continuous trajectory of the original PLDEs is associated to a qualitative (discrete) trajectory of the discrete network.

We show that logic networks of this type can be described formally and exploited via a Constraint Logic Programming (CLP) implementation. The CLP approach rests on the cooperation of solvers on various fields (tree, list, rational, real, boolean). Its advantages are that (i) the implementation is expressed in a very similar way to the *formal specification*, thus guaranteeing the correctness of the implementation, (ii) it is *iterative* - when new information become available, new constraints can be added to reduce further the space of possible models; (iii) many different queries can easily be posed to this formal specification due to its logical form. For example, queries equivalent to simulation (parameters known / computation of behavior) as well as inference of model parameters (information on behavior / computation of parameter values). Situations that are intermediate between simulation and inference are frequent. Indeed, the experimental characterization of *behaviors* (trajectories in phase space) is itself often partial, and a current challenge in the field is to be able to exploit all available partial knowledge to get more precise models.

These principles are applied to the study of adhesion between human endothelial cells. The work is done in collaboration with experimental biologists (4). A submodel extracted from a larger network is presented (2 variables, 7 discrete parameters).

We explain briefly the architecture of the implementation in the declarative language prolog IV (5). A preliminary version of the tool has been published (6) which did not take into account the existence of sliding modes. This was too restrictive and a full implementation is now available. A set of logical predicates defines the discrete *transition rules* corresponding to the type of networks studied (asynchronous multivalued networks with singular states). A given network is described by a set of discrete equations and a set of inequalities between parameters; these entities are derived from the

architecture of the given network (number of nodes/genes and pattern of interactions: activation or repression of gene g_i by gene g_j with threshold θ_{ij}). Observational knowledge is also described by constraints (logical predicates). This can be a direct measurement of a kinetic parameter, or knowledge about the behaviour of the system, such as, for example: "when the system is perturbed and set into state S_p, it returns to stable state S_0 by going through at least one state in which the concentration of such protein P is above such threshold θ". As illustrated in this example, this knowledge can be incomplete. It can nevertheless be formalized into a logical expression (after discretization) and exploited to make deductions about, for example, the possible values of the model parameters. As said in (ii) above, each new observation allows to add a new constraint which, in general, reduces the space of solutions. Likewise, hypotheses can be expressed as formal prolog queries in order to test their consequences. This provides a flexible tool to query model properties or, more generally, properties of a given network architecture.

In the cell adhesion study, we exploit *behavioral* information resulting from the observation of the response of the cell culture after a perturbation. To illustrate the strength and the flexibility of the CLP approach, results will be commented concerning : some general properties of the model; the existence of stationary states; and the use of behavioral information to reduce the space of possible models. In particular, it is shown that imposing the existence of a path from the perturbed state to the adherent state eliminates a large number of models. If times permit, results from larger published models of developmental biology will also be presented: segmentation of the drosophila embryo (7) and the drosophila gap-gene system (8).

(1) E. H. Snoussi, Dyn. Stab. Sys., **4**, 189 (1989).

(2) R. Thomas and M. Kaufman. Multistationarity, the Basis of Cell Differentiation and Memory. II. Logical Analysis of Regulatory Networks in Term of Feedback Circuits. Chaos, **11**, 180-195 (2001).

(3) H. de Jong, J.-L. Gouzé, C. Hernandez, M. Page, T. Sari, and J. Geiselmann. Qualitative Simulation of Genetic Regulatory Networks Using Piecewise-Linear Models. Bulletin of Mathematical Biology, **66**, 301-340 (2004).

(4) B. Hermant, S. Bibert, E. Concord, B. Dublet, M. Weidenhaupt, T. Vernet and D. Gulino-Debrac. Identification of Proteases Involved in the Proteolysis of Vascular Endothelium Cadherin during Neutrophil Transmigration. The Journal of Biological Chemistry, **278**, 14002-14012 (2003).

(5) A. Colmerauer. Prolog - Constraints Inside, Manuel de Prolog, PROLOGIA, Case 919, 13288 Marseille cedex 09, France (1996).

(6) E. Fanchon, F. Corblin, L. Trilling, B. Hermant and D. Gulino. Modeling the Molecular Network Controlling Adhesion Between Human Endothelial Cells: Inference and Simulation Using Constraint Logic Programming. Computational Methods in Systems Biology 2004, V. Danos and V. Schachter (Eds.), LNBI **3082**, 104–118, Springer-Verlag (2005).

(7) Sanchez L, Thieffry D. A logical analysis of the Drosophila gap-gene system. J. Theor Biol., **211**, 115-141 (2001).

(8) Sanchez L, Thieffry D. Segmenting the fly embryo: a logical analysis of the pair-rule cross-regulatory module. J. Theor. Biol., **224**, 517-537 (2003).