Systems Biology, Models, and Concurrency*

Walter Fontana

Department of Systems Biology, Harvard Medical School, 200 Longwood Avenue, Boston MA 02115, USA
walter@hms.harvard.edu

(*) Joint work with Vincent Danos, Jérôme Feret, Russ Harmer, and Jean Krivine

Abstract

Models will play a central role in the representation, storage, manipulation, and communication of knowledge in systems biology. Models capable of fulfilling such a role will likely differ from the familiar styles deployed with great success in the physical sciences. Molecular systems at the basis of cellular decision processes are concurrent and combinatorial. Their behavior is as much constrained by relationships of causality between molecular interactions as it is by chemical kinetics. Understanding how such systems give rise to coherent behavior and designing effective interventions to fight disease will require a notion of model that is akin to the concept of program in computer science. I will discuss recent progress in implementing a platform and tools for formal analysis that bring us closer to this vision. Protein interactions are represented by means of rules expressed in a formal language that captures a very simple, yet effective and biologically meaningful level of abstraction. Models, then, are collections of rules operating on an initial set of agents, in complete analogy to rules of organic chemical reactions. I will describe tools for analyzing and navigating rule collections as well as exploring their dynamics. We draw on concepts familiar to computer science, especially event structures, and adapt them to biological needs with the goal of formalizing the notion of "pathway". The challenges are many, but a roadmap for the future is discernible. Computer science will play a central role in providing an additional foundational layer, both theoretical and practical, that neither physics nor chemistry can offer on their own in the future definition of the biological sciences.

Categories and Subject Descriptors J.3 [*Life and medical sciences*]: Biology and genetics

General Terms Experimentation, Languages

Keywords systems biology, models, concurrency, Kappa

Summary

In the past decades, molecular biology and the genome projects have opened many doors and worn out some hinge concepts in the process. We don't quite know anymore what a gene is. It used to be a stretch of DNA coding for one protein. But it turns out that most genes code for many variants of a protein. We used to think that a protein has a particular function. But a protein may be involved in so many processes and combine with so many other proteins into complexes with new interaction properties that the notion of a function has become elusive. We used to think that cells have types defined by genomic states, but there is too much consequential diversity within cells of the "same" type. We used to think of pathways that process information like pathways that transform metabolites. But now we know that few, if any, molecular signaling pathways are standalone entities. The cell resembles a combinatorially excitable medium more than a hardwired chip. Over here, we are tracking genetic diversity at the single nucleotide level and soon human genomes will be sequenced over night for less than the cost of a transatlantic airline ticket. Over there is the promise of curing hideous disease, the potential of turning the molecular biology toolbox into a living technology, and the spell of providing answers to Big Questions: What is a cell "thinking"? What does it mean to be "evolvable"? In the middle of it all is a big gaping hole. Genome data is very powerful for identifying disease, but the genome is not the level at which we intervene to cure disease. To cure disease, we must engage with the dynamical systems of protein interactions that conspire with the genome in generating phenotype. The loose collection of wide-ranging efforts to fill this gap is called systems biology. Models will play an important role in this effort. They will be vehicles for the organization, storage and communication of rapidly evolving knowledge, the design of experiments, and a deeper understanding of biology.

We believe it is essential that models play a double role in systems biology. First, a model should be a datastructure that contains a transparent, formal, and executable representation of the facts it rests upon. In our case, these facts are about molecular interaction mechanisms. A model should be an environment in which data that is about interaction actually interacts. Second, a model must be equipped with analytical tools for revealing the causal structures resulting from a particular system of facts. These structures shape dynamics in non-intuitive ways. Such a notion of model differs from common practice in physics. In fact, it resembles much more the notion of a program in computer science. If so, then modeling is not unlike programming. Conversely, if a program is to *be* a biological model, it cannot be written in an arbitrary language. The analysis (static and otherwise) of the program must reveal, not obfuscate, the causal mechanisms that constrain the dynamics of the system it represents. Our approach is to express facts (or hypotheses) about molecular agents in a formal language, effectively turning empirical data directly into instructions. We have recently implemented a platform (Danos *et al* 2007) for defining and analyzing concurrent models of combinatorially complex signaling systems based on a dialect of Kappa, a formalism with clear appeal to biologists, originally proposed by Vincent Danos and Cosimo Laneve (Danos and Laneve 2004). The rule-centric nature of Kappa—analogous to reaction schemes in organic chemistry—curbs the combinato-

POPL'08, January 7–12, 2008, San Francisco, California, USA.
ACM 978-1-59593-689-9/08/0001.

rial explosion that wipes out traditional kinetic approaches based on differential equations or their stochastic Petri Net counterparts. The reduction of concepts from concurrency to biological practice is neither simple to implement nor easy for biologists to grasp. It deals with unfamiliar concepts, whose clarification took a long time even within their domain of origin. In particular, we adapted the notion of an event structure (Nielsen and Winskel 1995) to formalize the intuitive notion biologists hold of a "pathway to an observable". Techniques from abstract interpretation (Cousot and Cousot 1977; Danos *et al* 2008) are deployed to generate all "local views" of molecular entities made possible by a given set of rules, as well as to generate maps that depict causal and conflictual relationships between rules. These maps, in juxtaposition with specific event structures, can be used to understand system dynamics. We have implemented a fully scalable continuous-time Monte Carlo execution schedule (Gillespie simulator) (Danos *et al* 2007b) for generating and sampling traces in accordance with standard stochastic chemical reaction kinetics. This suite of techniques is being tested in collaboration with laboratories to elucidate large cellular decision systems. Future challenges call for extensions of the language to express additional physical aspects, such as geometric properties of agents and topological properties of reaction media (e.g. compartments), without losing formality and analytical capabilities.

References

V. Danos, J. Feret, W. Fontana, R. Harmer, and J. Krivine. Rule-based Modelling of Cellular Signalling. In: L. Caires and V.T.Vasconcelos (Eds.): CONCUR 2007 - Concurrency Theory, *Lecture Notes in Computer Sciences*, **4703**, 17-41 (2007).

V. Danos and C. Laneve. Formal molecular biology. *Theoretical Computer Science*, **325(1)**, 69-110 (2004).

M. Nielsen and G. Winskel. Models For Concurrency. In: Handbook of Logic and the Foundations of Computer Science, vol. 4, 1-148, Oxford University Press (1995).

P. Cousot and R. Cousot. Abstract interpretation: a unified lattice model for static analysis of programs by construction or approximation of fixpoints. In: POPL 1977 Conference Record of the Fourth Annual ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages, 238-252, ACM Press, New York, (1977)

V. Danos, J. Feret, W. Fontana and J. Krivine. Abstract interpretation of cellular signalling networks. In: VMCAI 2008 Ninth International Conference on Verification, Model Checking and Abstract Interpretation. *Lecture Notes in Computer Sciences*. To appear. (2008)

V. Danos, J. Feret, W. Fontana and J. Krivine. Scalable simulation of cellular signaling networks. In: APLAS 2007 Fifth ASIAN Symposium on Programming Languages and Systems. *Lecture Notes in Computer Sciences*, **4807**. To appear. (2007)

About the speaker

Walter Fontana holds an MS in biochemistry and a PhD in theoretical chemistry from the University of Vienna, Austria. He did postdoctoral research in molecular evolution and complex systems at Los Alamos National Laboratory in New Mexico. He subsequently returned to Vienna for a few years, but decided to resign tenure for a 6-year term-limited appointment as research professor in residence at the Santa Fe Institute, a think tank in New Mexico. From 1999 to 2000 he was also a Member in residence at the Institute for Advanced Study in Princeton, New Jersey. He is professor of systems biology at Harvard Medical School since 2004. His current theoretical research focus is on biological information processing and his experimental focus is on aging in *C.elegans*, a worm made of 959 cells about 1mm long, and 70 microns in diameter.