Molecules as Automata

Luca Cardelli

Microsoft Research, Cambridge, U.K.
luca@microsoft.com

Extended Abstract

Molecular biology investigates the structure and function of biochemical systems starting from their basic building blocks: macromolecules. A macromolecule is a large, complex molecule (a protein or a nucleic acid) that usually has inner mutable state and external activity. Informal explanations of biochemical events trace individual macromolecules through their state changes and their interaction histories: a macromolecule is endowed with an identity that is retained through its transformations, even through changes in molecular energy and mass. A macromolecule, therefore, is qualitatively different from the small molecules of inorganic chemistry. Such molecules are stateless: in the standard notation for chemical reactions they are seemingly created and destroyed, and their atomic structure is used mainly for the bookkeeping required by the conservation of mass.

Attributing identity and state transitions to molecules provides more than just a different way of looking at a chemical event: it solves a fundamental difficulty with chemical-style descriptions. Each macromolecule can have a huge number of internal states, exponentially with respect to its size, and can join with other macromolecules to from even larger state configurations, corresponding to the product of their states. If each molecular state is to be represented as a stateless chemical species, transformed by chemical reactions, then we have a huge explosion in the number of species and reactions with respect to the number of different macromolecules that actually, physically, exist. Moreover, macromolecules can join to each other indefinitely, resulting in situations corresponding to infinite sets of chemical reactions among infinite sets of different chemical species. In contrast, the description of a biochemical system at the level of macromolecular states and transitions remains finite: the unbounded complexity of the system is implicit in the potential molecular interactions, but does not have to be written down explicitly. Molecular biology textbooks widely adopt this finite description style, at least for the purpose of illustration.

Many proposal now exist that aim to formalize the combinatorial complexity of biological systems without a corresponding explosion in the notation. Macromolecules, in particular, are seen as stateful concurrent agents that interact with each other through a dynamic interface. While this style of descriptions is (like many others) not quite accurate at the atomic level, it forms the basis of a formalized and growing body of biological knowledge.

The complex chemical structure of a macromolecule is thus commonly abstracted into just internal states and potential interactions with the environment.

A. Corradini and U. Montanari (Eds.): WADT 2008, LNCS 5486, pp. 18–20, 2009.

Each macromolecule forms, symmetrically, part of the environment for the other macromolecules, and can be described without having to describe the whole environment. Such an open system descriptive style allows modelers to extend systems by composition, and is fundamental to avoid enumerating the whole combinatorial state of the system (as one ends up doing in closed systems of chemical reactions). The programs-as-models approach is growing in popularity with the growing modeling ambitions in systems biology, and is, incidentally, the same approach taken in the organization of software systems. The basic problem and the basic solution are similar: programs are finite and compact models of potentially unbounded state spaces.

At the core, we can therefore regard a macromolecule as some kind of automaton, characterized by a set of internal states and a set of discrete transitions between states driven by external interactions. We can thus try to handle molecular automata by some branch of automata theory and its outgrowths: cellular automata, Petri nets, and process algebra. The peculiarities of biochemistry, however, are such that until recently one could not easily pick a suitable piece of automata theory off the shelf. Many sophisticated approaches have now been developed, and we are particularly fond of stochastic process algebra. In this talk, however, we do our outmost to remain within the bounds of a much simpler theory. We go back, in a sense, to a time before cellular automata, Petri nets and process algebra, which all arose from the basic intuition that automata should interact with each other. Our main criterion is that, as in finite-state automata, we should be able to easily and separately draw the individual automata, both as a visual aid to design and analysis, and to emulate the illustration-based approach found in molecular biology textbooks.

With those aims, we investigate stochastic automata collectives. Technically, we place ourselves within a small fragment of a well-know process algebra (stochastic pi-calculus), but the novelty of the application domain, namely the mass action behavior of large numbers of well-mixed automata, demands a broader outlook. By a collective we mean a large set of interacting, finite state automata. This is not quite the situation we have in classical automata theory, because we are interested automata interactions. It is also not quite the situation with cellular automata, because our automata are interacting, but not necessarily on a regular grid. And it is not quite the situation in process algebra, because we are interested in the behavior of collectives, not of individuals. And in contrast to Petri nets, we model separate parts of a system separately. By stochastic we mean that automata interactions have rates. These rates induce a quantitative semantics for the behavior of collectives, and allow them to mimic chemical kinetics. Chemical systems are, physically, formed by the stochastic interactions of discrete particles. For large number of particles it is usually possible to consider them as formed by continuous quantities that evolve according to deterministic laws, and to analyze them by ordinary differential equations. However, one should keep in mind that continuity is an abstraction, and that sometimes it is not even a correct limit approximation.

In biochemistry, the stochastic discrete approach is particularly appropriate because cells often contain very low numbers of molecules of critical species: that is a situation where continuous models may be misleading. Stochastic automata collectives are hence directly inspired by biochemical systems, which are sets of interacting macromolecules, whose stochastic behavior ultimately derives from molecular dynamics. Some examples of the mismatch between discrete and continuous models are discussed.