Languages for Systems Biology
GC1 In Vivo ⇔ In Silico

Luca Cardelli

Microsoft Research

I propose to study languages that can precisely and concisely represent biological processes such as the one described below. I give a very specific example, for concreteness and for shock value. But the range of phenomena and problems that fit in this endeavor is much larger. The domain is that of systems biology [13], which aims to represent not only cellular-level phenomena, such as the one below, but also phenomena at the level of tissues, organs, organisms, and colonies. Descriptive formalisms are needed to represent and relate many levels of abstraction.

The given example concerns the "algorithm" that a specific virus follows to reproduce. It is a sequence of steps that involve the dynamic merging and splitting of compartments, the transport of materials, and the transcription and interpretation of digital information. The algorithm is informally described in English below. What are appropriate languages and semantic models that can accurately and concisely describe such an algorithm, at a high level of abstraction but *in its entirety*? Formal modeling (e.g., at the level that can drive a simulator) is becoming of central importance in biology, where complex processes need to be analyzed for hypothesis testing. The area is increasing concerned with the discrete, although stochastic and perturbation-proof, processing of information.

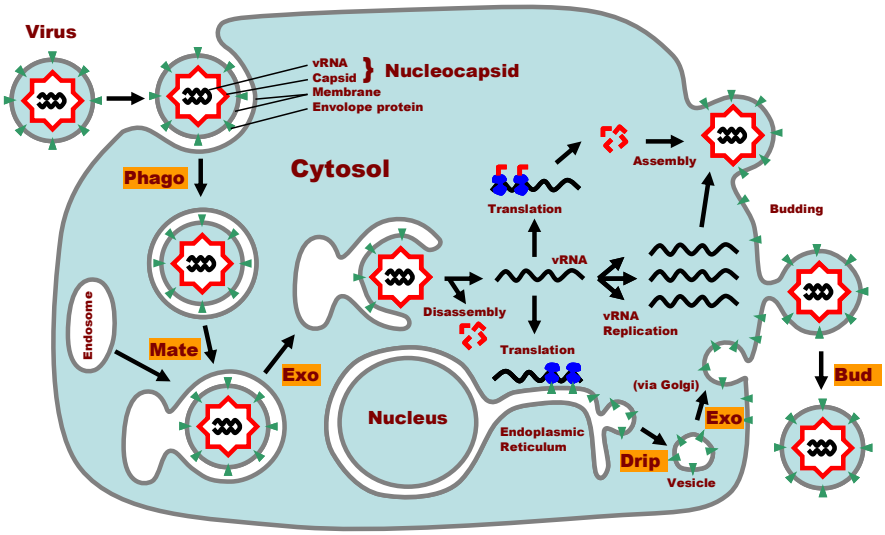

Figure 1 Semliki Forest Virus Infection and Reproduction ([1] p.279)

Figure 1. A virus is too big to cross a cellular membrane. It can either punch its RNA through the membrane or, as in this example, it can enter a cell by utilizing standard cellular endocytosis machinery. The virus consists of a capsid containing the viral RNA (the nucleocapsid). The nucleocapsid is surrounded by a membrane that is similar to the cellular membrane (in fact, it is obtained from it "on the way out"). This membrane is however enriched with a special protein that plays a crucial trick on the cellular machinery, as we shall see shortly. The virus is brought into the cell by phagocytosis, wrapped in an additional membrane layer; this is part of a standard transport pathway into the cell. As part of that pathway, an endosome merges with the wrapped-up virus. At this point, usually, the endosome causes some reaction to happen in the material brought into the cell. In this case, though, the virus uses its special membrane protein to trigger an exocytosis step that deposits the naked nucleocapsid into the cytosol. The careful separation of internal and external substances that the cell usually maintains has now been subverted. The nucleocapsid is in direct contact with the inner workings of the cell, and can begin doing damage. First, the nucleocapsid disassembles itself, depositing the viral RNA into the cytosol. This vRNA then follows three distinct paths. First it is replicated (either by cellular proteins, or by proteins that come with the capsid), to provide the vRNA for more copies of the virus. The vRNA is also translated into proteins, again by standard cellular machinery. Some proteins are synthesized in the cytosol, and form the building blocks of the capsid: these self-assemble and incorporate a copy of the vRNA to form a nucleocapsid. The virus envelope protein is instead synthesized in the Endoplasmic Reticulum, and through various steps (through the Golgi apparatus) ends up lining transport vesicles that merge with the cellular membrane, along another standard transport pathway. Finally, the newly assembled nucleocapsid makes contact with sections of the cellular membrane that are now lined with the viral envelope protein, and buds out to recreate the initial virus structure outside the cell.

Are existing languages and semantic models adequate to represent these kinds of situations? Many classical approaches are relevant, but I believe the current answer must be: definitely not. Biologists are busy inventing their own abstact notations [8][9][10]. There are, of course, some proposals from computing as well [2][3][5][6][7]. The systems to be described are massively concurrent, heterogeneous, and asynchronous (notoriously the hardest ones to cope with in programming), with stochastic behavior and high resilience to drastic changes of environment conditions. What organizational principles make these systems work predictably? [11][12]

Answers to these questions should be of great interest to computing, for the organization of complex software systems. But that may come later: the proposal here is exclusively to model biological systems in order to understand how they work. The fundamental connection to computing (shared by systems biologists) is that many levels of organization are much more akin to software systems than to physical systems, both in hierarchical complexity and in algorithmic-like information-driven behavior. Hence the emphasis on the central role that languages may play.

References

[1] B.Alberts, D.Bray, J.Lewis, M.Raff, K.Roberts, J.D.Watson. **Molecular Biology of the Cell**. Third Edition, Garland.

[2] L.Cardelli. **Brane Calculi – Interactions of Biological Membranes**. http://research.microsoft.com/Users/luca/Papers/Brane%20Calculi.pdf.

[3] V.Danos and C.Laneve. **Formal Molecular Biology**. Theoretical Computer Science, to Appear.

[4] R.Milner. **Communicating and Mobile Systems: The π-Calculus**. Cambridge University Press, 1999.

[5] C.Priami. **The Stochastic pi-calculus**. The Computer Journal 38: 578-589, 1995.

[6] C.Priami, A.Regev, E.Shapiro, and W.Silverman. **Application of a stochastic name-passing calculus to representation and simulation of molecular processes**. Information Processing Letters, 80:25-31, 2001.

[7] A.Regev, E.M.Panina, W.Silverman, L.Cardelli, E.Shapiro. **BioAmbients: An Abstraction for Biological Compartments**. Theoretical Computer Science, to Appear.

[8] K. W. Kohn: **Molecular Interaction Map of the Mammalian Cell Cycle Control and DNA Repair Systems.** Molecular Biology of the Cell, 10(8):2703-34, Aug 1999.

[9] H. Kitano: **A graphical notation for biochemical networks**. BIOSILICO 1:169-176, 2003.

[10] **Systems Biology Markup Language.** http://www.sbml.org

[11] Hartwell LH, Hopfield JJ, Leibler S, Murray AW: **From molecular to modular cell biology.** Nature. 1999 Dec 2;402(6761 Suppl):C47-52.

[12] McAdams HH, Arkin A.: **It's a noisy business! Genetic regulation at the nanomolar scale**. Trends Genet. 1999 Feb;15(2):65-9.

[13] **4th International Conference on Systems Biology**. http://icsb2003.molecool.wustl.edu