

**Theory and Software Infrastructure for Systems Biology**

**Doyle, J.<sup>1</sup>; Prajna, S.<sup>1</sup>, Papachristodoulou, A.<sup>1</sup>, Tanaka, R.<sup>1</sup>, Khammash, M.<sup>2</sup>, El-Samad, H.<sup>2</sup>, Hucka, M.<sup>1</sup>, Doyle, F.<sup>2</sup>, Gillespie, D.<sup>1</sup>, Petzold, L.<sup>2</sup>, Carlson, J.<sup>2</sup>, Yi, T-M<sup>1</sup>, Goncalves, J<sup>1</sup>**

**<sup>1</sup>California Institute of Technology, Pasadena, CA, USA; <sup>2</sup>University of California, Santa Barbara, CA, USA**

This poster will sketch an overview of the research of a team from Caltech and UCSB addressing the challenge of scalable scientific theory, software, and networking infrastructure for biology. Deep insights about the fundamental nature of biological complexity can now be drawn from the convergence of three research themes. Molecular biology has provided a detailed description of the much of components of biological networks, and the organizational principles of these networks are becoming increasingly apparent. In addition, advanced technology has provided engineering examples of networks with complexity approaching that of biology (which will also be enablers for systems biology). While the components are entirely different, there is striking convergence at the network level of architecture and the role of protocols in structuring complex system modularity. Finally, there is a new mathematical framework for the study of complex networks that suggests that this apparent network-level evolutionary convergence both within biology and between biology and technology is not accidental, and follows necessarily from the requirements that both biology and technology be efficient and robust.

One aim of this research program is to build on the mathematics of systems engineering to create a coherent and complete theoretical infrastructure proceeding from experimental data to modeling, analysis, inference, and with tight feedback to experimentation and modeling throughout. Both data and modeling assertions and questions must be described in a common framework that is biologically natural, yet can be turned over to powerful algorithms for resolution. (A key element in our software infrastructure is the Systems Biology Markup Language, SBML, which is described in another poster). Is a model consistent with experimental data, which may come from extremely heterogeneous sources? If so, is it robust to additional perturbations that are plausible but untested? Are different models at multiple scales of resolution consistent? What is the most promising experiment to refute a model? Put in natural terms (which are typically stochastic, nonlinear, nonequilibrium, uncertain, hybrid and so on), such questions are conventionally viewed as computationally intractable (NP/coNP hard), and biologists are forced to translate into unnatural terms in order to use available algorithms. This is undesirable and potentially unnecessary in light of our new research results.

The mathematics and algorithms described in this poster build on robust control and dynamical systems theory, stochastic processes, numerical analysis, computational complexity, real semi-algebraic geometry, semidefinite programming, and duality. The crucial insight is that biological systems have evolved in ways that are highly robust to uncertainty in their environment and components yet may be fragile to novel perturbations. This “robust yet fragile” feature must be explicitly exploited in systems biology, just as it has been implicitly critical in the success of existing research. Using the same theory, new Internet protocols have been developed at Caltech that enable for the first time controlled high-speed exchange of data crucial for enabling high-end distributed scientific computing applications. These are currently being extensively deployed worldwide by a team led by Caltech researchers, and the implications for computation and collaboration in systems biology will be briefly sketched.

*Funding from NSF, AFOSR, ARO, JST (Japan), NIH, and DARPA is acknowledged.*