Integrating an Ensemble of Distributed Biochemical Network Models

V.A. Shiva Ayyadurai, C. Forbes Dewey, Jr.
Biological Engineering Division, Massachusetts Institute of Technology, Cambridge, MA 02138 USA

A new system for integrating an ensemble of distributed biochemical network models is presented. Rapid growth in the number of biochemical network models, created in different formats, across different computing systems, with minimal input and output information, necessitates the need for such a system in order to build large scale models in a flexible and scalable manner. The EGFR pathway of Kholodenko is used to demonstrate the viability of this new system. Current approaches for integrating biochemical network models: 1. are monolithic; 2. require manual integration; 3. are single-format dependent (e.g. SBML); 4. ignore the constant maintenance required to update and re-integrate changing models; and, 5. are constrained by requiring all models to be located on one central computer.

The system, known as Cytosolve, offers a scaleable computational architecture to overcome these limitations. The novel elements of this system are: 1. Distributed Solver which calculates and updates species concentrations for each time step; 2. Mass Balance Manager which dynamically couples and calculates species concentrations that are common across models; 3. Pathway Ontology, which is MIRIAM compliant, that manages and reconciles nomenclature and units differences across models; and 4. Network Communications Monitor which optimizes latency time between each time step. The system supports both well-mixed and compartmentalized model representations.

Cytosolve is applied to an ensemble of distributed models comprising the Kholodenko EGFR short-term signaling pathway. Unlike monolithic approaches which require all models to be manually "wired" together and be resident on one central computer, each model in the Cytosolve distributed approach resides on its own native computing environment, in its own native format and is dynamically integrated. In Figs. 1 and 2, two specific species concentration profiles from the Cytosolve distributed approach are compared with the monolithic approach of the fully integrated model.

The results demonstrate that Cytosolve solutions are nearly the same as the monolithic approach. Relative to run-time latency, Cytosolve minimizes this effect for each time step by optimizing communication among the ensemble of models. Our previous research indicates that run-time latency is only a small multiple of the overall time required to solve the model using the monolithic approach; at the same time, there is no requirement to invest in creating a single monolithic model from the various components. It is expected that the parallel approach will yield additional run-time computational savings for very large models. Cytosolve also supports ease of maintenance, since each model of the ensemble can change independently without requiring any manual re-integration.

Fig. 1 - Bound: EGF-EGFR

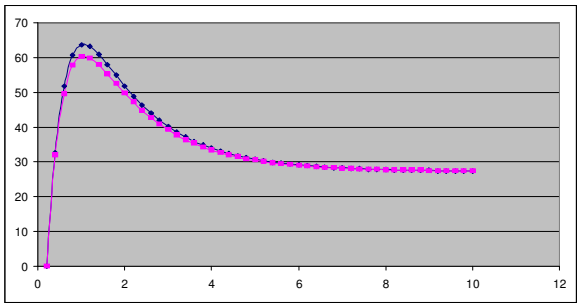
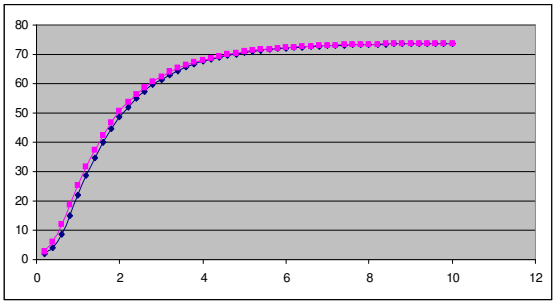


Fig. 2 - Dimerized: (EGF-EGFR)2



Plots of species concentrations of bound EGF-EGFR and Dimerized (EGF-EGFR)2 complex in Fig. 1 and Fig. 2, respectively. The blue and purple lines in each diagram are the species concentrations calculated using the Cytosolve distributed approach and monolithic approach, respectively.