

# Integrating bio-ontologies with a workflow/Petri Net model to qualitatively represent and simulate biological systems

**Mor Peleg, Irene S. Gabashvili, and Russ B. Altman**, *Stanford Medical Informatics, Stanford University, Stanford, CA, USA*

[peleg@smi.stanford.edu](mailto:peleg@smi.stanford.edu), [isg@smi.stanford.edu](mailto:isg@smi.stanford.edu), [russ.altman@stanford.edu](mailto:russ.altman@stanford.edu)

The volume of data contained in biological literature is increasingly growing. The diversity of data is very large, and includes information such as genetic sequence polymorphism, molecular function, clinical phenotypes, and cellular location. Data exists in many levels of detail, ranging from atomic mechanisms to cellular-level processes. A challenge facing researchers is the ability to piece together the wealth of dispersed data relating to their domain of interest. Thus, our goal is to develop technology for representing qualitative, noisy, and sparse biological results in support of the eventual goal of fully accurate quantitative models. We developed data structures for storing data related to a biological domain in a way that allows the knowledge to be systematically evaluated and examined by scientists as well as computer algorithms. In an earlier work [1], we concentrated on representation of structural, functional, and dynamic aspects of biological systems and querying the system to discover relationships among processes and structural components (e.g., biopolymers that have the same set of roles and are not inhibited by the same inhibitor). We have extended this work by representing the interactions of mutated gene products with other cellular components. This representation and its querying tools may aid scientists to gain insights into the effects of specific mutations on molecular, cellular, and organism-level processes.

In developing our knowledge model, we integrated existing models of system structure, and behavior with biomedical ontologies, creating extensions, where needed. We based our model on the Workflow model of the Workflow management coalition [2]. This model can represent the ordering among processes that take place in a modeled system, and their participants (e.g., substrates, products). Processes can be hierarchically decomposed to control complexity for human understanding. In addition, the Workflow model represents compositional relationships among participants (e.g., a complex is composed of subunits) and functional roles that participants play in the modeled system.

We use Transparent Access to Multiple Biological Information Sources (TAMBIS) [3] and the Unified Medical Language System (UMLS) [4] as controlled vocabularies of biological and medical concepts, respectively, that can be mapped to the participants, roles, and processes in the workflow model. The resulting model is a biological workflow model.

We extended the biological workflow model by adding an ontology that defines: (1) cellular location, (2) types of participation relationships of participants in processes (e.g.,

inhibition, catalysis), (3) types of evidence that supports facts in the knowledge base (e.g., in-vivo, inferred from other species), (4) nucleic acid 3D structure linked to secondary and primary structural blocks, (5) alleles, mutations, and their relationships with sequence components, and (6) specialization relationships (i.e., initiator tRNA is a kind of tRNA) and compositional relationships among participants (i.e., relating domains to a molecule, relating subunits to a complex, relating molecules to collections of molecules that share similar function, structure, or cellular localization).

We implemented our framework as a frame-based knowledge base using the Protégé-2000 tool [5]. Using Protégé's Axiom Language, we composed queries about structural and functional aspects of biological systems – such as relationships between defective processes and the clinical phenotype of the mutation that is causing it.

The Workflow model can be mapped to Petri Nets [6] – a mathematical model that represents concurrent systems, which allows verification of formal properties as well as qualitative simulation. Formal properties that can be checked include (1) checking that there is no infinite accumulation of metabolites (boundedness), and (2) ensuring that all modeled processes can be traversed (liveness). We use existing tools for verifying and simulating the Petri Nets that correspond to the Workflow model of a biological system.

To illustrate the power of this model, we have used it to represent data taken from the literature in two distinct domains: (1) the process of invasion of erythrocyte cells by malaria parasites, and (2) mutations in tRNA molecules and their affects on the process of protein translation. Using the queries that we constructed, we could find mutations that cause both certain clinical phenotypes (e.g., Juvenile myopathy, encephalopathy, lactic acidosis and stroke (MELAS)) as well as abnormal function (e.g., misreading). Mapping the workflow model to Petri Nets enabled us to verify the soundness of some dynamic aspects of tRNA biology and to simulate system behavior in the presence of different mutations. Our model is available at <http://smi.stanford.edu/projects/helix/malaria.html>

1. Peleg M, Yeh I, Altman RB. Modeling biological processes using Workflow and Petri Net models. *Bioinformatics* 2002;18(6):825-837.
2. Layna Fischer (Ed). Workflow Handbook: Future Strategies Inc., Lighthouse Point, FL, USA.; 2002.
3. Baker PG, Goble CA, Bechhofer S, Paton NW, Stevens R, Brass A. An ontology for bioinformatics applications. *Bioinformatics* 1999;15(6):510-520.
4. Lindberg C. The Unified Medical Language System (UMLS) of the National Library of Medicine. *J Am Med Rec Assoc* 1990;61(5):40-42.
5. Noy NF, Ferguson RW, Musen MA. The knowledge model of Protege-2000: Combining interoperability and flexibility. 2d International Conference on Knowledge Engineering and Knowledge Management (EKAW'2000); Juan-les-Pins, France; 2000.
6. Peterson JL. Petri Net Theory and the Modeling of Systems. Englewood Cliffs, NJ: Prentice-Hall; 1981.