The Algorithmic Biology of Interacting Biological Pathways

Jian-Qin Liu and Su-Shing Chen
Universirty of Florida
Corresponding author: suchen@cise.ufl.edu

Abstract

Biological systems are holistic. A biological system, such as a human body, must contain many pathways interacting in an extremely complicated manner. A challenging question is how we can model a biological system or at what level? Our approach is to start at the pathway level, which may go up to the system top level and down to the bottom sequence level. In mathematics, we will use the graph and hypergraph theories to model pathways and logical formalism (or formal systems) to study the interaction mechanisms.

A hypergraph is a generalization of a graph, where edges can connect any number of vertices. Formally, an hypergraph is a pair (*X*,*E*) where *X* is a set of elements, called *nodes* or *vertices*, and a *E* is a subset of the power set S(X) called *hyperedges*. While graph edges are pairs of nodes, hyperedges are arbitrary sets of nodes, and can therefore contain an arbitrary number of nodes.A formal system consists of two components, a formal language plus a set of inference rules or transformation rules. A formal system may be formulated purely abstractly, or it may be intended to serve as a description of some domain of real phenomena or some aspect of objective reality (e.g. biological pathways) .In mathematics, formal proofs are the product of formal systems, consisting of axioms and rules of deduction. Theorems are then recognized as the possible 'last lines' of formal proofs. The point of view that this sums up all there is to mathematics is often called formalism, but is more properly referred to finitism. Perhaps, algorithmic biology can be considered in this perspective. David Hilbert founded metamathematics as a discipline for discussing formal systems. Any language that one uses to talk about a formal system is called a metalanguage. The metalanguage may be nothing more than ordinary natural language, or it may be partially formalized itself, but it is generally less completely formalized than the formal language component of the formal system under examination, which is then called the object language, that is, the object of the discussion in question. Here we are discussing a biological pathway language!

In order to stay out of the possible controversies implied by Gödel incompleteness, precise definitions are adopted here based on categories. As a reasonable way to lay the foundation for the formal system, we are building on the data-checkable evidence of biological pathway in which the most obvious merit -- instinct rationality brought by formal systems -- is a key to dispose of the controversies even though the empirical situation remains. A category $\wp$ is given as the construct:

$$<\mathrm{Ob}(\wp), \mathrm{Mor}(X,Y), \vartheta>$$

where Ob($\wp$) denotes the set of pathway objects, Mor(X,Y) is the morphisms of X into Y (X, Y $\in$ Ob($\wp$)) and ϑ is the law of composition designed by rule Q below.

In a related aspect, graph rewriting has very broad meanings, which imply the rich parallel features of its computingparadigms. Designated as a class of special directed hypergraph that can be converted into string representation by rewriting, a Cell system of interacting pathways is proposed as the core of our formal system. The model can be formalized as shown in the following construct:

Cell = < V, T, D, V', E, Y, Z, PTs, Q >

Where

V -- the alphabet set.
T -- the terminal set and $T \subset V$.
D -- the set of {0, 1}.
V' -- the set of vertexes; $V' \subset V$;
E -- the set of edges;
Y -- the set of the hypergraphs <HE, V'> in which HE is the set of hyperedges in Y and corresonding HR (hyperedge replacement) and VR (vertex replacement) are defined based on HE;
Z -- the set of local concentrations and $Z \subset V$.
PTs -- the set of pathways -- the directed graphs that have inputs and outputs in V' and contain hyperedges in HE with interactions which fall into HE. It is defined as:
[pathway] ::= [pathway] ∪ [single-hyperedge with two vertex] under the operation of Q as follows:

Q -- the set of operators for operations on hypergraphs in HE from V and E i.e., the objects in PTs; and we also have the following:

Q = {Q1, Q2, Q3, Q4}

The operation processes carried out by the operator set of Q are formalized as four rules in terms of graph rewriting on hypergraphs:

(1) Q1 (the rule of "interaction"): $\alpha \to \chi, \beta \to \delta, \varepsilon \to \phi, \ldots, \gamma \to \eta$ where $\alpha, \chi, \beta, \delta, \varepsilon, \phi, \ldots, \gamma, \eta$ refer to pathways as we defined in PTs.

(2) Q2 (the rule of "feedback"): $\varphi \to \kappa, \lambda \to \mu$, where $\varphi, \kappa, \lambda, \mu$ refer to the pathways and $\varphi = \mu, \kappa \neq \lambda$

(3) Q3 (the rule of "addition of new pathways"): $\nu \to \pi\theta$

(4) Q4 (the rule of “deletion”): $\rho \to \sigma$

In this formal system, V is used as the symbolic representation for all of the chemicals in cells, e.g., the names of the reactants in biochemical reactions, which lead to biological pathways. D is introduced for digital denotation in tapes of equivalent Turing-machines; Z means the concentration values of chemicals in cells that are represented as symbolic forms; and the measurement D is decided by the threshold quantity represented by symbols, which belong to the set of V. The formal system is a kind of algorithmic machines reflecting the idea "from pathways to pathways" where pathway rewriting is conceptualized by operators defined by rules.