KAMI: a bio-curation tool for cellular signalling

Eugenia Oshurko*, Sébastien Légaré, Russ Harmer
ievgenia.oshurko@ens-lyon.fr

Abstract

Rule-based modelling has proven to be a successful approach for studying complex systems of cellular signalling. A rule-based language (Kappa, kappalanguage.org) has been actively developed and used in recent years. While being able to deal with the problem of combinatorial explosion in the number of molecular species particular to classical modelling techniques, Kappa stays unsuited for building and curating big explanatory models—for what we refer to as bio-curation. To tackle exactly this problem we propose a tool called KAMI (Knowledge Aggregator and Model Instantiator), which allows gradual semi-automatic aggregation of protein-protein interactions of different provenance, their annotation, visualisation and further instantiation to concrete rule-based models.

KAMI

KAMI (Knowledge Aggregator and Model Instantiator) proposes a bio-curation framework providing:

- De-contextualized model representation focusing on mechanisms of individual protein-protein interactions (PPIs).
- Semi-automatic aggregation of individual PPIs.
- Means to instantiate knowledge in different contexts: cell types, wild type vs disease.
- Automatic generation of executable models from different instantiations (e.g. Kappa rules, ODEs).

Interaction

KAMI uses the background knowledge about generic interaction mechanisms of conserved protein domains to validate input PPIs and automatically sharpen input knowledge. At the current stage KAMI is equipped with the semantics of: protein kinase domain, phosphatase domain, SH2 domain.

Semantics is built in the semantic nuggets SN and the semantic action graph SA.

Knowledge Aggregation Pipeline

KAMI provides a format for programmatic input of PPIs, which implements an intuitive intermediary representation that frees the user from the need to formulate knowledge in the abstract graphical form of nuggets.

The schematic pipeline of knowledge aggregation in KAMI is the following:

Model Instantiation

A gene node in the meta-model represents a neighbourhood in sequence space of a gene. By associating regions, residues and states to a specific gene we represent a feasible neighbourhood of its products. Nuggets in KAMI represent potential interactions and necessary conditions for them to happen. It implies that:

- The necessary conditions may not always be realized: splice variants may not have required regions or sites, mutants can have key residue replacement which prevents interactions, e.g. we can “unfold” GRB2 in the action graph into two products: wild type protein GRB2WT and a mutant GRB2SHD, whose key residue S90 was replaced by D as follows:

    KAMI allows the user to define gene products and to apply the respective “unfolding” to the action graph, which induces automatic propagation to the nuggets.

- As the result, necessary conditions for interactions in some nuggets may fail, e.g. the given protein definition of GRB2 WT will invalidate nugget N0 from our previous example for the mutant GRB2SHD.

This process of unfolding and propagation to nuggets of given definition of gene products is called instantiation. Instantiated models in KAMI can be further translated into Kappa rules. For example, the instantiated nugget N1 gives a rise to a single Kappa rule for the interaction of GRB2WT but no rule for GRB2SHD:

    GRB2WT($pY_{[0]}$, Y1092$^\text{WT}$) →
    "EGFR binds to the SH2 domain of GRB through a binding site with phosphorylated residue Y1092"