

STKE Connections Map Metadata Schema

Expressing Qualitative Signal Transduction Pathways in SBML

Eighth Forum on Software Platforms for Systems Biology, Nov. 10, 2003
Fourth International Conference on Systems Biology

Shai Sachs

ssachs@akazaresearch.com

Akaza Research, Cambridge, MA

<http://www.akazaresearch.com>



Science's Signal Transduction Knowledge Environment

American Association for the Advancement of Science

<http://stke.sciencemag.org>



Project Goals

- Create a data exchange format for the Connections Map (CM) built around SBML.
 - » Distribute machine-readable CM data sets for use in research.
 - » License CM data sets to affiliated companies and institutions.
- Contribute to the development of SBML.
 - » Contribute insights gained from this project to the SBML community.
 - » Provide another working real-world implementation of SBML.
 - » Demonstrate the feasibility of porting existing repositories to SBML.

Reasons for Choosing SBML

- Ability to express many CM concepts with native SBML support.
- Degree of third-party support in software and online data repositories.
- Speed of development of the standard and open-source nature of development efforts.
- CellML not chosen because it has less third-party support, and it is evolving at a slower pace.
 - » However, CellML could adequately represent many CM concepts.

Mapping from CM to SBML

- Mapping CM concepts to SBML:
 - » Pathways correspond to <model>.
 - » Nodes correspond to <species>.
 - » Relations correspond to <reaction>.
 - » Database primary keys used as “id” attributes.
 - Primary keys modified slightly to conform to SId grammar.
- Use as many native SBML elements as possible, without losing semantics of CM concepts.
- Incorporate “SBML-style” elements, using <listOf___> elements to enclose lists.
- Separate information from different entities into different elements.
 - » Node Metadata in <species> : <annotation> : <node>.
 - » Relation Metadata in <reaction> : <annotation> : <relation>.

Example: Relations Mapped to Reactions

```
<reaction id="stkecm_CMR_10003" reversible="false">
  <annotation>
    <relation sign="+" method="stkecm_CV_1"
      reliability="Demonstrated" />
  </annotation>
  <listOfProducts>
    <speciesReference species="stkecm_CMC_10000" />
  </listOfProducts>
  <listOfModifiers>
    <modifierSpeciesReference species="stkecm_CMN_999"/>
  </listOfModifiers>
</reaction>
```

(Fas to Akt protein kinase stimulatory relation in Fas Signaling Pathway in Cardiomyocytes.)

Canonical vs. Specific Pathways

- Specific pathways are associated with an organism; canonical pathways are not.
- Each specific pathway has a “parent” canonical pathway.
- Similar distinctions are made for components.
- For specific pathways, a <compartment> element describes the organism and tissue in which the pathway occurs and in which all species exist.
- For canonical pathways, a “canonical compartment” is created. All species exist within this compartment.

Controlled Vocabularies Proposal

- Separate Controlled Vocabulary file from SBML file.

- Main SBML File:

```
<listOfControlledVocabularies>
```

```
  <controlledVocabularyReference type="component_type"
```

```
    location="http://stke.sciencemag.org/cm/vocabularies/component_type.xml"
```

```
  />
```

```
  <controlledVocabularyReference type="organism"
```

```
    location="http://stke.sciencemag.org/cm/vocabularies/organism.xml" />
```

```
</listOfControlledVocabularies>
```

- External Controlled Vocabulary File (e.g. "component_type.xml")

```
<controlledVocabulary type="component_type">
```

```
  <listOfTerms>
```

```
    <term id="1" value="diglyceride" parentID="2" />
```

```
    <term id="2" value="lipid" />
```

```
  </listOfTerms>
```

```
</controlledVocabulary>
```

Controlled Vocabularies for Complex Types

- Store relevant component information in a “Component Library” file separate from the main SBML file.

```
<componentLibrary>
```

```
  <listOfComponents>
```

```
    <component id="stkecm_CMC_13132" name="Interleukin-13" abbreviation="IL-13"
      dataType="simple" type="stkecm_CV_1" organism="stkecm_CV_2">
```

```
      <description>IL-13 is a cytokine produced by activated T helper 2 cells, mast cells, and
        basophils.</description>
```

```
    </component>
```

```
  </listOfComponents>
```

```
</componentLibrary>
```

- Use component ids and `<componentLibraryReference>` to refer to a component within the main SBML file.
- Advantages:
 - » Facilitates development of component libraries at many granularities or in many organisms.
 - » Creates a single point of access for all component information.
 - » Enables non-STKE users to use component information without supporting SBML.

Level 2 Could Not Express Some CM Concepts

- Qualitative descriptions of modifier effect: “+”, “-”, “0”, “?”.
- Complex components – <compartment> element does not suffice for expressing containment relationship between complex and simple components.
- Homologs – Relationships between canonical and specific components.
 - » If X is a canonical component, and is the parent for specific components Y and Z , then Y and Z are homologs.
- Synonyms – multiple names for a single component.
- Groups of Relations
 - » Feedback / Feedforward groups
 - » Coupled relations
 - » Alternative relations
- Citations

Proposed Level 3 Features Will Provide More Support for CM Concepts

- Geometric Layout features will express simple layout parameters for nodes.
 - » (x, y) position on a grid.
- Complex Species feature will enable description of complex and token components.
- Model Composition may enable description of the relationship between canonical and specific pathways, consequently expressing homologs more naturally.
- Level 3 Controlled Vocabularies feature will express description of several attributes.
 - » Name of organism in which pathway occurs.
 - » Name of tissue cell in which pathway occurs.
 - » Subcellular localization of a node.
 - » Modifier effect (“+”, “-”, “0”, “?”).

Conclusions

- SBML expresses many CM concepts naturally.
- However, some concepts had to be expressed within <annotation> elements.
 - » Many software tools could be modified slightly to incorporate many key features of CM, including qualitative relation descriptions and complex/token components.
 - » A C library (perhaps an extension of libSBML) for reading CM SBML documents could facilitate these modifications.
- SBML framework focuses on quantitative more than qualitative expression of biochemical pathways.
- Separating a Controlled Vocabulary from an SBML files facilitates development of new versions of a Controlled Vocabulary over time.
- Level 3 of SBML will provide many needed features which will enable more natural mapping from STKE CM to SBML.