Internal Discussion Document
Possible extension to the Systems Biology Markup Language
Complex Species and Species Graphs

Andrew Finney
afinney@cds.caltech.edu
ERATO Kitano Systems Biology Workbench Development Group
Control and Dynamical Systems 107-81
California Institute of Technology, Pasadena, CA 91125

Version of March 14, 2001

Contents

1 Disclaimer

This document is intended for internal discussion amongst members of the ERATO Kitano Systems Biology Workbench Development Group and selected collaborators. It is very unlikely that any final version of this document will have any IPR or other release restrictions.

This document has not been reviewed in detail by the group and at this stage just describes the author's ideas. In addition this document is incomplete: it contains editorial notes which act as place holders for future work.

2 Introduction

This document describes a possible extension of SBML for inclusion in level 2 to enable the reactants and products of reactions to be graphs of species. This document, together with a more formal description of the extension, will be integrated at a later date into a document on level 2.

3 Motivation

This extension is proposed to enable large sets of seemingly identical reactions involving very similar species to be expressed in a concise form. Typically this occurs in computational biology when a species has several binding sites each of which can be unbound or bound to a number of different species. Such a species has a large number of states. In SBML level 1 all these different states have to be defined as separate species. Then all the reactions between these states have to be enumerated. This level 2 extension is designed to reduce this combinatorial explosion. However the enumeration of reactions is still required when those reactions have differing rate laws or rate law parameter values.

An additional motivation was to create modular framework for describing species to make SBML more amenable to simulators such as Stochsim and MCell.

4 Overview

In this extension species are either 'simple', i.e. defined as in level 1, or 'complex'. A complex species can have one or more named binding sites. A simple species effectively has one unnamed binding site.

By using this feature a species graph can be constructed where the arcs are pairs of binding sites between species. Such a species graph can define a complex chemical object. The features that the arcs represent are not defined formally by this extension and can include both covalent or non-convalent bonds. The reactants and products of reactions are either simple species, as in level 1, or species graphs.

To reduce the combinatorial explosion it is clear that reactions should apply to sets of species rather than specific species. In this scheme a set of species graphs can be defined by an incomplete species graph. An incomplete species graph has a missing arc for one or more binding sites. The set of species graphs defined as all the complete species graphs that can be created from the incomplete species graph by adding valid arcs and nodes in place of the missing arcs. This extension attempts to ensure that the set of complete species is computable from a wide range of incomplete species graphs.

5 Complex species

A complex specie has one or more named binding sites. For each binding site there is a set of one or more simple species or binding sites of complex species that can bind to that site. Binding sites may be used simply as a way to delineate parts of large molecules for the purposes of abstraction i.e. they don't have to have any real chemical significance.

A simple specie is defined as in level 1. A complex specie might be defined as follows:

```
<complexSpecie name="A">
```

```
    <listOfBindingSites>
        <bindingSite name="topleft">
            <listOfBindingSpecies>
                <simpleBinding species="a">
                <complexBinding species="B" site="left">
            </listOfBindingSpecies>
        </bindingSite>
        <bindingSite name="topright">
            <listOfBindingSpecies>
                <simpleBinding species="w">
            </listOfBindingSpecies>
        </bindingSite>
        <bindingSite name="bottomleft">
            <listOfBindingSpecies>
                <simpleBinding species="x">
            </listOfBindingSpecies>
        </bindingSite>
        <bindingSite name="bottomright">
            <listofBindingSpecies>
                <simpleBinding species="y">
            </listOfBindingSpecies>
        </bindingSite>
    </listOfBindingSites>
</complexSpecie>
```

In the above example we are defining a complex specie A with four binding sites. The *topleft* binding site can bind to either species a or the *left* site of B complex specie. For each of the remaining sites there is only one species that the site can bind with. These species are all different simple species.

6 Species Graph

In a pathway a reaction consists of a set of reactants and products. Reactants and products are either simple species or a species graph. A species graph is a graph where the nodes are instances of species and the arcs are either pairs of binding sites when two complex species are bound together or a single binding site when a complex specie is bound to a simple specie. (Alternatively one can think of a simple specie has having a single unnamed binding site).

A complex specie in a reaction may not be completely specified i.e. the state of a binding site can be left open in which case the reaction is deemed to apply to all states of that binding site. These incomplete or ambiguous specie graphs in reactions define a set of complete species graphs.

Alternatively it is possible to define in a species graph that a given binding site is unoccupied or free.

The individual complex specie (specie instances) that comprise a species graph are named. These names have scope across a whole reaction. The individual simple specie in a species graph are not named i.e. each occurrence of a simple specie in a reaction must be viewed as a separate chemical entity.

Using the example complex specie above a valid species graph might be as follows. (another complex specie 'B' is used , but not defined, that has only one binding site 'left').

```
<listOfComplexSpeciesInstances>
    <ComplexSpeciesInstance name="A1" specie="A"/>
    <ComplexSpeciesInstance name="B1" specie="B"/>
</listOfComplexSpeciesInstances>
...
<reactionSpeciesGraph name="X">
    <listOfBindings>
        <!-- note that bottomright is missing implying that the reaction is     -->
        <!-- unaffected by or/and doesn't change the state of that binding site -->
        <binding>
            <!-- example of two complex species binding together -->
            <complexLink componentInstance="A1" bindingSite="topleft"/>
            <complexLink componentInstance="B1" bindingSite="left"/>
        </binding>
```

```
        <binding>
            <!-- example of a complex species binding to a simple species -->
            <complexLink componentInstance="A1" bindingSite="topright"/>
            <simpleLink species="w"/>
        <binding>
            <!-- example of a unoccupied binding site on a complex species -->
            <complexLink componentInstance="A1" bindingSite="bottomleft"/>
        </binding>
    </listOfBindings>
</reactionSpeciesGraph>
```

It is hopefully obvious that this shows that *A1* and *B1* are instances of *A* and *B* complex species respectively and are bound together via their *topleft* and *left* sites respectively. What about the remaining sites on *A1*? The bottom left site as written above must be unoccupied or free. The bottom right is not listed indicating that the reaction in which the species graph occurs is independent of the state of that site. Therefore this incomplete species graph represents a set of two complete species graphs. The binding site of *A1* could be bound to *y* or it could be free. The remaining site on *A1* is bound to the species *w*.

7 Explanation and Reasoning

We have named sites so that it is possible to identify which sites are involved in a reaction and thus identify unspecified binding states. We specify the species that can attach to a site to simplify the process of enumerating the set of complete species graphs from an incomplete species graph. The complex species in a species graph are named so that (1) it is possible to specify a cycle of bindings and (2) so that species bindings unaffected by a reaction can be traced through a reaction. Species graphs are named so that their concentrations can occur in kinetic laws.

8 Relationship between species graphs and reactions

This formalization is meant to describe the state of species graphs *between* reactions rather than during the reaction i.e. it is not meant to describe enzyme bindings to reactants. However this is really a question of how a given biochemical model is defined and not something constrained by the language.

9 Initial Species Graphs

The extension will include a new section to specify the initial concentrations of specific species graphs. This will be separate from species definitions. These species graphs will be complete. I have called these "initial species graphs"

10 Example Model

The example model here is a subset of a model developed by Borisuk and Tyson (J Theor Biol 1998 195 69-85). This example was chosen because:

- it contains an enumeration of states of a species graph namely the M-phase promoting factor (MPF)
- the rate laws for pairs of reactions between these states are identical

The subset of the published model used here is shown in figure 1. The level 1 representation of this model is as follows. (in SBML all reactions are reversible by default).

To simplify the extension case all the initial dimers are unphosphorylated.

```
<model name="tyson1998subset">

<listOfCompartments>
```

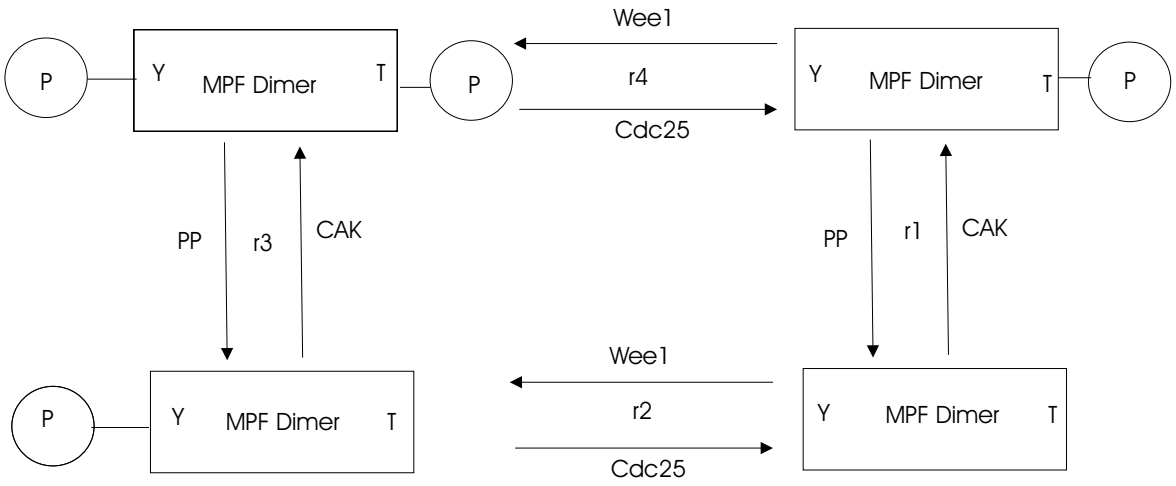

Figure 1: *A subset of the tyson model*

```
    <compartment name="cell"/>
</listOfCompartments>

<listOfParameters>
    <parameter name="Kpp" value="0.004"/>
    <parameter name="Kcak" value="0.64"/>
    <parameter name="V251" value="0.017"/>
    <parameter name="V252" value="0.17"/>
    <parameter name="VWee1" value="0.01"/>
    <parameter name="VWee2" value="1"/>
</listOfParameters>

<listOfSpecies>
    <specie name="Wee1" compartment="cell" initialAmount="1"/>
    <specie name="Wee1P" compartment="cell" initialAmount="1"/>
    <specie name="Cdc25" compartment="cell" initialAmount="1"/>
    <specie name="PP" compartment="cell" initialAmount="1"/>
    <specie name="CAK" compartment="cell" initialAmount="1"/>
    <specie name="Dimer" compartment="cell" initialAmount="1"/>
    <specie name="PDimer" compartment="cell" initialAmount="0"/>
    <specie name="DimerP" compartment="cell" initialAmount="0"/>
    <specie name="PDimerP" compartment="cell" initialAmount="0"/>
</listOfSpecies>

<listOfReactions>
    <reaction name="r1">
        <listOfReactants>
            <speciesReference specie="Dimer"/>
        </listOfReactants>
        <listOfProducts>
            <speciesReference specie="DimerP"/>
        </listOfProducts>
        <kineticLaw formula="Kcak * Dimer - Kpp * DimerP"/>
    </reaction>

    <reaction name="r2">
        <listOfReactants>
            <speciesReference specie="Dimer"/>
        </listOfReactants>
        <listOfProducts>
            <speciesReference specie="PDimer"/>
        </listOfProducts>
        <kineticLaw formula=
            "Dimer * (VWee1 * Wee1P + VWee2 * Wee1) - (V251 * Cdc25 + V252 * Cdc25P) * PDimer"/>
    </reaction>
```

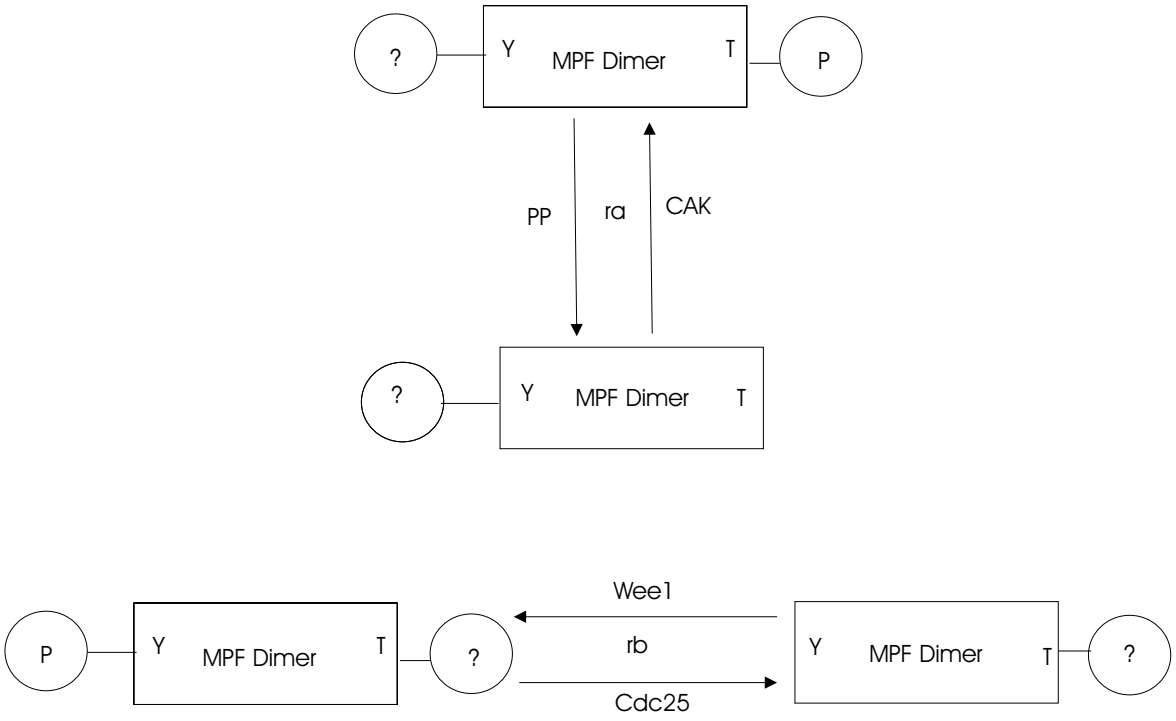

Figure 2: *A formulation of the Tyson model subset abstracting the dimer phosphorylation reactions*

```
    <reaction name="r3">
        <listOfReactants>
            <speciesReference specie="PDimer"/>
        </listOfReactants>
        <listOfProducts>
            <speciesReference specie="PDimerP"/>
        </listOfProducts>
        <kineticLaw formula="Kcak * PDimer - Kpp * PDimerP"/>
    </reaction>

    <reaction name="r4">
        <listOfReactants>
            <speciesReference specie="DimerP"/>
        </listOfReactants>
        <listOfProducts>
            <speciesReference specie="PDimerP"/>
        </listOfProducts>
        <kineticLaw formula=
            "DimerP * (VWee1 * Wee1P + VWee2 * Wee1) - (V251 * Cdc25 + V252 * Cdc25P) * PDimerP"/>
    </reaction>
</listOfReactions>
</model>
```

The formulation of this model using a species graph representation is shown in figure 2. The proposed SBML representation of this model is as follows.

```
<model name="tyson1998subsetabstracted">

<listOfCompartments>
    <compartment name="cell"/>
</listOfCompartments>

<listOfParameters>
    <parameter name="Kpp" value="0.004"/>
```

```
    <parameter name="Kcak" value="0.64"/>
    <parameter name="V251" value="0.017"/>
    <parameter name="V252" value="0.17"/>
    <parameter name="VWee1" value="0.01"/>
    <parameter name="VWee2" value="1"/>
</listOfParameters>

<listOfSpecies>
    <specie name="Wee1" compartment="cell" initialAmount="1"/>
    <specie name="Wee1P" compartment="cell" initialAmount="1"/>
    <specie name="Cdc25" compartment="cell" initialAmount="1"/>
    <specie name="PP" compartment="cell" initialAmount="1"/>
    <specie name="CAK" compartment="cell" initialAmount="1"/>
    <specie name="P" compartment="cell" initialAmount="0"/>
</listOfSpecies>

<listOfComplexSpecies>
    <complexSpecie = "Dimer">
        <listOfBindingSites>
            <bindingSite name="Y">
                <listOfBindingSpecies>
                    <simpleBinding species="P"/>
                </listOfBindingSpecies>
            </bindingSite>
            <bindingSite name="T">
                <listOfBindingSpecies>
                    <simpleBinding species="P"/>
                </listOfBindingSpecies>
            </bindingSite>
        </listOfBindingSpecies>
    </complexSpecie>
<listOfComplexSpecies>

<!-- only dimer state at start of simulation... ->
<listOfInitialSpeciesGraphs>
    <initialSpeciesGraph initialAmount="1" compartment="cell">
        <listOfComplexSpecieInstances>
            <complexSpecieInstance name="X" complexSpecie="Dimer"/>
        </listOfComplexSpecieInstances>
        <listOfBindings>
            <binding>
                <complexLink complexSpecieInstance="X" site="Y"/>
            </binding>
            <binding>
                <complexLink complexSpecieInstance="X" site="T"/>
            </binding>
        </listOfBindings>
    </initialSpeciesGraph>
</listOfInitialSpeciesGraphs>

<listOfReactions>
    <reaction name="ra">

        <listOfComplexSpeciesInstances>
            <complexSpecieInstance name="X" complexSpecie="Dimer"/>
        </listOfComplexSpeciesInstances>

        <listOfReactants>
            <reactionSpeciesGraph name="DimerIn">
                <listOfBindings>
                    <complexLink complexSpeciesInstance="X" site="T"/>
                <listOfBindings>
            </reactionSpeciesGraph>
        </listOfReactants>

        <listOfProducts>
            <reactionSpeciesGraph name="DimerOut">
                <listOfBindings>
                    <complexLink complexSpeciesInstance="X" site="T"/>
```

```
                    <simpleLink species="P"/>
                <listOfBindings>
            </reactionSpeciesGraph>
        </listOfProducts>
        <kineticLaw formula="Kcak * DimerIn - Kpp * DimerOut"/>
    </reaction>

    <reaction name="rb">
        <listOfComplexSpeciesInstances>
            <complexSpecieInstance name="X" complexSpecie="Dimer"/>
        </listOfComplexSpeciesInstances>

        <listOfReactants>
            <reactionSpeciesGraph name="DimerIn">
                <listOfBindings>
                    <complexLink complexSpeciesInstance="X" site="Y"/>
                <listOfBindings>
            </reactionSpeciesGraph>
        </listOfReactants>

        <listOfProducts>
            <reactionSpeciesGraph name="DimerOut">
                <listOfBindings>
                    <complexLink complexSpeciesInstance="X" site="Y"/>
                    <simpleLink species="P"/>
                <listOfBindings>
            </reactionSpeciesGraph>
        </listOfProducts>
        <kineticLaw formula=
            "DimerIn * (VWee1 * Wee1P + VWee2 * Wee1) - (V251 * Cdc25 + V252 * Cdc25P) * DimerOut"/>
    </reaction>

</listOfReactions>
</model>
```

Although it is probably difficult to justify the additional complexity of the extension for this model it is not difficult to see that another phosphorylation state of the MPF dimer would have justified the use of the extension (another 8 reactions and 4 species would have been required in the level 1 model as opposed to just one more reaction in the case of the extension based model). It would be interesting to attempt to model such a system in both level 1 and the extension if a studied naturally occurring case exists. It is important to note the verbosity of the extension based model has been, perhaps artificially, reduced by having an initial concentration of only one of the MPF dimer states.

11 Translation to Level 1

To map back to simple SBML level 1 structures requires that the set of complete species graphs, implied by the incomplete species occurring in reaction elements, are enumerated. Each complete species graph becomes a species. Translation software would have problems with recursive complex species definitions even though they are highly advantageous for describing macromolcules for example. Perhaps for each complex specie we could have some upper limit on the number of times the specie can occur in any species graph as a simple constraint on the translation.

To make such a scheme work a canonical representation for species graphs is required. I assume that this is an ongoing research topic in its own right.

12 Discussion

Its not clear yet whether the proposed extension either efficiently addresses the motivations for its creation or matches the scheme used in other simulators. In particular this scheme is only useful when a rate law can be applied across a range of similar reactions.

I feel that there should be a demonstration of a real need for this extension before it can be included in

SBML Level 2. I would like to have suggestions for reliably simplifying this extension.

13 Acknowledgements

I would like to thank Tau-Mu Yi and Baltazar Aguda for their advice during the writing of this document.