Internal Discussion Document
Possible extension to the Systems Biology Markup Language
Complex Species: species with multiple states

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 16, 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 species to have multiple states. 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, or when a species can have different geometric states. In short in many existing models species can have 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 specie can have one or more named sets of one or more named states. The sets of states are called *features*. The specific states of complex specie instances are specified in reactions as part of a species reference. The initial amounts of different complex specie states, at the start of a simulation, have to be specified.

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 specie states can be defined by an incomplete specie reference. An incomplete specie reference has one or more missing features i.e. the state of those features is not specified. The set of complete specie references is derived from an incomplete species reference.

5 Complex species

A complex specie has one or more named features. For each feature there is a set of one or more states. A simple specie is defined as in level 1. A complex specie might be defined as follows:

```
<complexSpecie name="receptor">
    <listOfFeatures>
        <feature name="geometry">
            <listOfStates>
                <state name="native"/>
                <state name="unfolded"/>
                <state name="partiallyFolded"/>
            </listOfStates>
        </feature>
        <feature name="leftBindingSite">
```

```
            <listOfStates>
                <state name="unbound"/>
                <state name="OHbound"/>
            </listOfStates>
        </feature>
    </listOfFeatures>
</complexSpecie>
```

In the above example we are defining a complex specie *receptor* with two features. The *geometry* feature represents the structure of the species and has the states *native*, *unfolded* and *partiallyFolded*. The *leftBindingSite* feature represents a binding state of the *receptor*.

5.1 Initial Species States

The extension will include a new element inside complex species elements to specify the initial concentrations of specific complex species states.

6 Species reference to Complex Species

In a pathway a reaction consists of a set of reactants and products. Reactants and products are species references. In this extension a species reference can consist of one or more feature states which specify the state of a complex specie that the reaction applies to. A feature state is pair of state and feature names.

The instances of complex specie that occurring in a reaction are listed separately from the specie references. This is to enable more than one complex specie state to occur in a reaction and still allow the tracking of the specie through the reaction. The specie reference has a name so that the concentration of that specie can be used in the reaction.

Its simple to create a representation of a set of species by having a complex specie reference with an incomplete list of feature states. The reference then represents an enumeration of all the states of the missing features.

For example here's an example of a specie reference referring to a receptor defined in the previous example.

```
<listOfComplexSpeciesInstances>
    <complexSpecieInstance name="r" complexSpecie="receptor"/>
</listOfComplexSpeciesInstances>

...

<complexSpecieReference name="X" specieInstance="r">
    <listOfFeatureStates>
        <featureState feature="geometry" state="native"/>
    </listOfFeatureStates>
</complexSpecieReference>
```

This species reference specifies the set of two receptors where the geometry is native and the leftBindingSite is either bound or bound to OH.

7 Explanation and Reasoning

We have named features so that it is possible to identify which features are involved in a reaction and thus identify unspecified feature states. We specify the states that a feature can have to simplify the process of enumerating the set of complete species graphs from an incomplete species graph.

8 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)

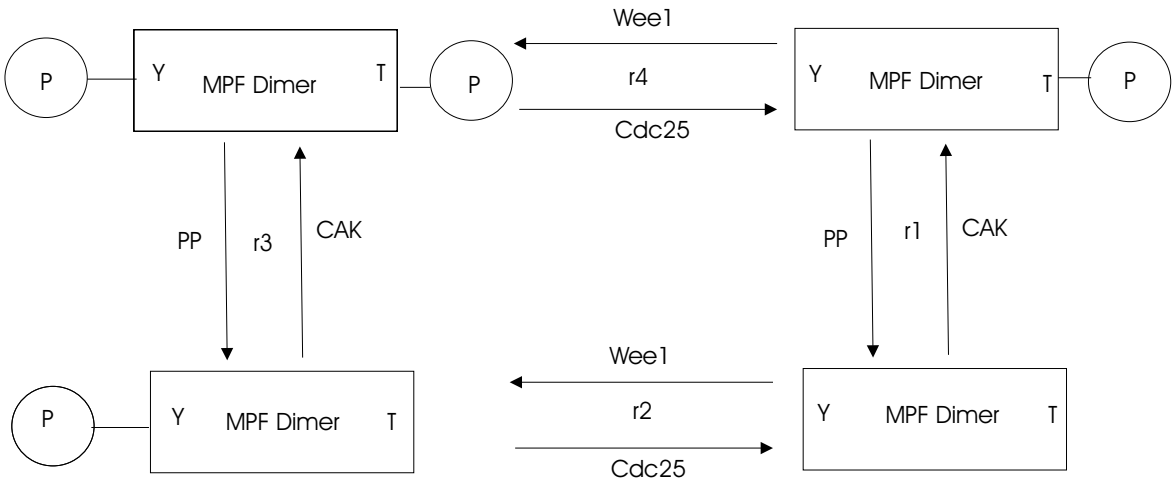

Figure 1: *A subset of the tyson model*

- 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>
    <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>

    <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"/>
</listOfSpecies>

<listOfComplexSpecies>
    <complexSpecie = "Dimer">
        <listOfFeatures>
            <feature name="Y">
                <listOfStates>
                    <state name="boundP"/>
                    <state name="unbound"/>
                </listOfStates>
```

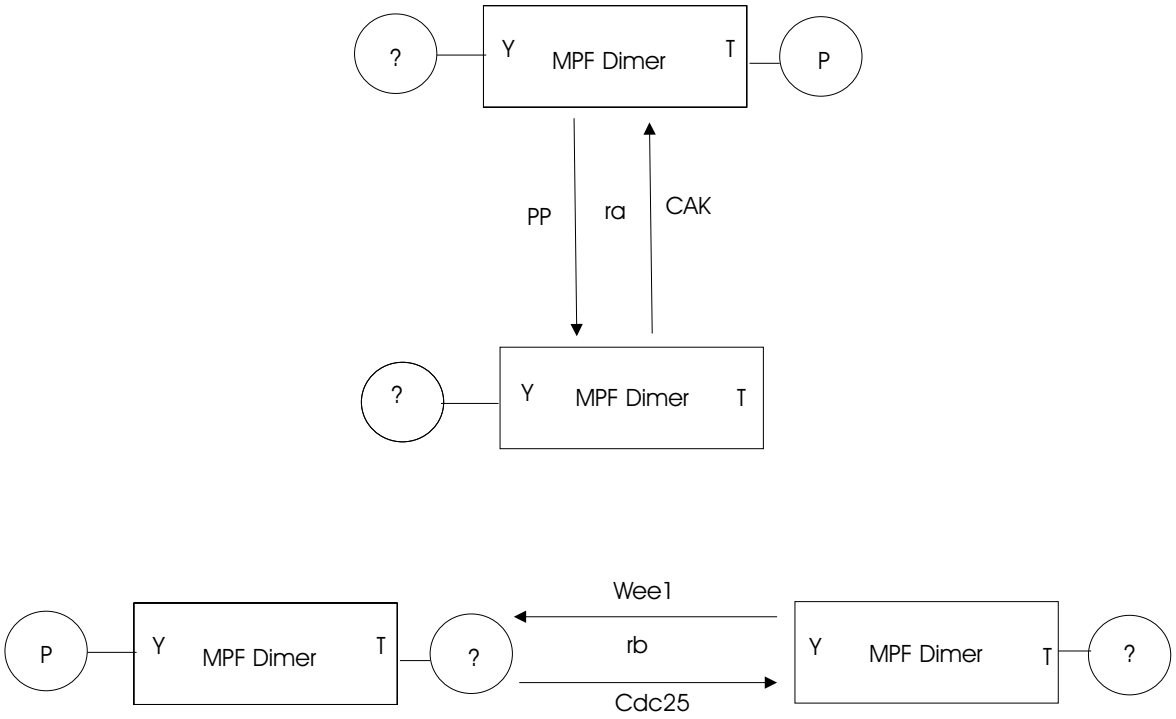

Figure 2: *A formulation of the Tyson model subset abstracting the dimer phosphorylation reactions*

```
            </feature>
            <feature name="T">
                <listOfStates>
                    <state name="boundP"/>
                    <state name="unbound"/>
                </listOfStates>
            </feature>
        </listOfFeatures>
    </complexSpecie>

    <!-- only dimer state at start of simulation... ->
    <listOfInitialStates>
        <initialState initialAmount="1" compartment="cell">
            <featureState name="Y" state="unbound"/>
            <featureState name="T" state="unbound"/>
        </initialState>
    </listOfInitialStates>
<listOfComplexSpecies>

<listOfReactions>
    <reaction name="ra">
        <listOfComplexSpeciesInstances>
            <complexSpecieInstance name="dimer" complexSpecie="Dimer"/>
        </listOfComplexSpeciesInstances>

        <listOfReactants>
            <complexSpecieReference name="DimerIn" complexSpecieInstance="dimer">
                <featureState name="T" state="unbound"/>
            </complexSpecieReference>
        </listOfReactants>

        <listOfProducts>
            <complexSpecieReference name="DimerOut" complexSpecieInstance="dimer">
                <featureState name="T" state="boundP"/>
            </complexSpecieReference>
        </listOfProducts>
```

```
            <kineticLaw formula="Kcak * DimerIn - Kpp * DimerOut"/>
        </reaction>

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

            <listOfReactants>
                <complexSpecieReference name="DimerIn" complexSpecieInstance="dimer">
                    <featureState name="Y" state="unbound"/>
                </complexSpecieReference>
            </listOfReactants>

            <listOfProducts>
                <complexSpecieReference name="DimerOut" complexSpecieInstance="dimer">
                    <featureState name="Y" state="boundP"/>
                </complexSpecieReference>
            </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.

9 Translation to Level 1

Translation of models using this extension to level 1 should be reasonably straight forward: most of the effort will focus on enumerating reactions.

10 Discussion

Its not clear yet whether the proposed extension either efficiently addresses the motivations for its creation. In particular this scheme is only useful when a rate law can be applied across a range of similar reactions. I would like to have suggestions for reliably simplifying this extension.

11 Acknowledgements

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