

Meeting Minutes 22 January 2001

Biochemical Pathway Models in CellML

Authors:

Melanie Nelson (Physiome Sciences Inc.)

Warren Hedley (Bioengineering Institute, University of Auckland)

Contributors:

Poul Nielsen (Bioengineering Institute, University of Auckland)

David Bullivant (Bioengineering Institute, University of Auckland)

1 Introduction

CellML is intended to be able to represent many different types of models. Therefore, its basic structure is rather general, and it relies primarily on the explicit definition of mathematics using MathML to specify models. It will always be possible to specify a model purely in terms of the mathematics, without using any of the elements defined in this document. However, in some types of models, some information is lost in reducing the model to pure mathematics. For instance, in biochemical pathway models, it will not always be straightforward, or even possible, to unambiguously determine from the mathematical rate laws which variables represent inhibitors or activators in the reactions.

Therefore, some additional elements were needed in CellML to fully capture the information in biochemical pathway models. Three fundamental representations of these models must be supported:

- **Mathematical Equations:** these equations define the kinetic rate laws that model the chemical reactions occurring in the pathway.
- **Chemical Expressions:** these are the stoichiometric equations (such as $A + B \rightleftharpoons 2C + D$) used by chemists to represent reactions.
- **Pathway Diagrams:** these are the stylized drawings commonly used by biochemists and cell biologists to represent the pathway of reactions. Some examples of pathway diagrams are shown in Section 4.

It is important that CellML be able to store the information needed to unambiguously reproduce any of these representations of a model. It is also important to minimize duplication of information within the model definition, because duplication can lead to inconsistencies. Therefore, we must integrate the information needed to support the three types of model representation.

2 Required Information about Reactions

The following information is required to allow CellML to store the information needed for all three representations of a pathway model:

- Which variables represent participants in the reaction, and what role or roles each participant plays in the reaction.
- The stoichiometry of each participant, for each role that the participant assumes in the reaction. (A species may have a stoichiometry of 1 as a reactant, but a stoichiometry of 2 as an inhibitor; i.e., one equivalent of the species is required for each unitary reaction equivalent, but two equivalents of the species are required for inhibition).

- Mathematics defining how the reactions progress. Modellers must be free to define arbitrary rate laws for reactions (i.e., they should not be restricted to choosing from some predefined set of kinetic rate laws).

Note that this list defines the set of information that CellML must be able to store, not the list of information required for a valid CellML model. Not all models must have all of this information. For instance, qualitative pathway models would not have mathematics defining how the reactions progress. A model consisting of nothing but the mathematics defining the progression of the reactions would also be completely valid CellML.

3 The Solution

A solution that meets the requirements described in Section 2 was reached after several iterations. Readers interested in seeing some of the discarded possibilities, along with the explanation of why these were found to be insufficient, should consult the [January 19, 2001](#)¹, and [January 20, 2001](#)², meeting minutes.

3.1 Basic Format

A new element is defined specifically to store the reaction information. This element is called the `<reaction>` element, and it may only appear inside of a `<component>` element. It is possible for a single `<component>` element to contain more than one `<reaction>` element, although this practice is not recommended. The `<reaction>` element may contain a `reversible` attribute, the boolean value of which indicates whether or not the reaction is reversible. The default value of the `reversible` attribute is "yes".

The reaction element contains multiple `<variable_ref>` components, each of which references one of the variables that participates in the reaction. Recommended best practice is to create a `<variable_ref>` element for each variable representing a chemical species that participates in a reaction, as well as one for the variable representing the rate of the reaction. The required `variable` attribute is the only attribute on the `<variable_ref>` element. Its value is the name of the referenced variable. This variable must be declared in the current `<component>` element.

Each `<variable_ref>` element contains one or more empty `<role>` elements. There are four possible attributes on the `<role>` element. The required `role` attribute specifies the way in which the variable participates in the reaction. There are currently seven values allowed for this attribute: "reactant", "product", "catalyst", "activator", "inhibitor", "modifier", and "rate". These are defined in Section 3.2. The `direction` attribute should only be used on `<role>` elements in reversible reactions. It may have values of "forward", "reverse", or "both", and indicates for which direction the role is relevant. It may only be defined on a `<role>` element within a reversible reaction, and has a default value of "forward". The optional `delta_variable` attribute indicates which variable is used to store the change in concentration of the species represented by the variable referenced by the current `<variable_ref>` element. The optional `stoichiometry` attribute stores the stoichiometry of the current variable relative to the other reaction participants. Section 3.2 contains detailed rules for the use of these attributes.

The `<variable_ref>` elements may also contain `<math>` elements, which enclose equations defined using MathML. The equations that define the reaction rate should be placed inside the `<variable_ref>` element that references the variable representing the rate of the reaction. MathML appearing in a `<variable_ref>` corresponding to a variable representing a chemical species should only be used to relate the reaction rate to the change in concentration of the species. This allows modellers to define an arbitrary mathematical

¹http://www.cellml.org/private/progress_reports/20010119_meeting_minutes.html

²http://www.cellml.org/private/progress_reports/20010120_meeting_minutes.html

equation to relate the reaction rate to the concentration change of the species. However, it is strongly recommended that the **stoichiometry** and **delta_variable** attributes be used instead if the concentration change is simply the reaction rate multiplied by the stoichiometry. (The rules for deriving this mathematical relationship from the **stoichiometry** attribute are defined in Section 3.2.)

Section 4 contains several examples of **<reaction>** elements for various types of reactions.

3.2 Rules for Use

CellML must support both quantitative and qualitative reaction models. This, in conjunction with the natural complexity of biochemical reactions, leads to a reasonably complicated set of rules for the use of the various parts of a **<reaction>** element. These rules, and the formal definitions of the allowed values of the various attributes described in Section 3.1 are formally specified below.

3.2.1 The **reversible** attribute on the **<reaction>** element

The **reversible** attribute on the **<reaction>** element indicates whether or not the reaction can proceed in both directions. It has two allowed values: "yes" and "no". When not specified, the default value is "yes". However, it is recommended best practice to always explicitly define the value of this attribute.

If the **reversible** attribute has a value of "yes", it is assumed that all reactants in the forward direction are products in the reverse direction, and vice versa. Therefore, it is not necessary to explicitly specify this with separate **<role>** elements for the reverse direction. In other words, the CellML fragment shown in Figure 1 is exactly equivalent to the fragment in Figure 2. The more compact formulation shown in Figure 1 is the recommended best practice. In fact, use of the more verbose formulation is specifically not recommended, because it creates the possibility for inconsistencies between the **delta_variable** and **stoichiometry** attributes specified on the forward roles and those specified on the reverse roles. If the more verbose formulation is used, the values of the **delta_variable** and **stoichiometry** attributes for the forward and reverse reactant and product roles for a given species variable must be identical.

3.2.2 Allowed values of the **role** attribute

There are seven allowed values for the **role** attribute on a **<role>** element:

- **"reactant"**: the species represented by the referenced variable is one of the species consumed or transformed by the reaction (in the direction given by the **direction** attribute). Reactants are also often called substrates.
- **"product"**: the species represented by the referenced variable is one of the species produced by the reaction (in the direction given by the **direction** attribute).
- **"catalyst"**: the species represented by the referenced variable catalyzes the reaction. In biochemical pathways, such a species will almost always be an enzyme, and will almost always occur with a **stoichiometry** equal to "1".
- **"activator"**: the species represented by the referenced variable enhances the reaction. Activators can occur with any stoichiometry. An activator will usually be a small molecule that increases the activity of an enzyme catalyzing the reaction. However, the detailed reaction representing this activation of the enzyme may not be included in the model. Instead, the activator may be represented as directly affecting the kinetics of the catalyzed reaction.
- **"inhibitor"**: the species represented by the referenced variable inhibits the reaction. Inhibitors can occur with any stoichiometry. An inhibitor will usually be a species that inhibits the activity of an enzyme catalyzing the reaction. However, the detailed reaction representing this inhibition of

```
<reaction reversible="yes">
  <variable_ref variable="A">
    <role
      role="reactant" direction="forward"
      delta_variable="delta_A" stoichiometry="1" />
  </variable_ref>
  <variable_ref variable="B">
    <role
      role="reactant" direction="forward"
      delta_variable="delta_B" stoichiometry="1" />
  </variable_ref>
  <variable_ref variable="C">
    <role
      role="product" direction="forward"
      delta_variable="delta_C" stoichiometry="1" />
  </variable_ref>
  <variable_ref variable="r">
    <role role="rate" direction="forward" />
    <math xmlns="http://www.w3.org/1998/Math/MathML">
      ...
    </math>
  </variable_ref>
</reaction>
```

FIGURE 1: A CellML fragment showing the recommended best practice for defining the simple reversible reaction $A + B \rightleftharpoons C$. The MathML block defining the rate law is omitted.

```
<reaction reversible="yes">
  <variable_ref variable="A">
    <role
      role="reactant" direction="forward"
      delta_variable="delta_A" stoichiometry="1" />
    <role
      role="product" direction="reverse"
      delta_variable="delta_A" stoichiometry="1" />
  </variable_ref>

  <variable_ref variable="B">
    <role
      role="reactant" direction="forward"
      delta_variable="delta_B" stoichiometry="1" />
    <role
      role="product" direction="reverse"
      delta_variable="delta_B" stoichiometry="1" />
  </variable_ref>

  <variable_ref variable="C">
    <role
      role="product" direction="forward"
      delta_variable="delta_C" stoichiometry="1" />
    <role
      role="reactant" direction="reverse"
      delta_variable="delta_C" stoichiometry="1" />
  </variable_ref>

  <variable_ref variable="r">
    <role role="rate" direction="forward" />
    <math xmlns="http://www.w3.org/1998/Math/MathML">
      ...
    </math>
  </variable_ref>
</reaction>
```

FIGURE 2: A CellML fragment showing the more verbose method for defining the simple reversible reaction $A + B \rightleftharpoons C$, in which the reactant and product roles are specified in both the forward and reverse direction. The MathML block defining the rate law is omitted. Use of this formulation is not recommended.

the enzyme may not be included in the model. Instead, the inhibitor may be represented as directly affecting the kinetics of the catalyzed reaction.

- **"modifier"**: the species represented by the referenced variable modifies the reaction in some unspecified way.
- **"rate"**: the referenced variable represents the rate of the reaction. A variable assigned this role cannot be assigned any other roles (i.e., it is an error to have more than one **<role>** element on a `variable_ref` if the `role` attribute of one of the **<role>** elements equals **"rate"**).

The `role` attribute is required on all **<role>** elements. There is no default value.

3.2.3 Allowed values of the `direction` attribute

There are three allowed values for the `direction` attribute on a **<role>** element:

- **"forward"**: the value of the `role` attribute is the role of the referenced variable in the reaction when running in the "favoured" direction. The favoured direction is the one in which the reactants are being consumed (i.e., the time-derivatives of their concentrations are negative), as defined by the kinetic rate law.
- **"reverse"**: the value of the `role` attribute is the role of the referenced variable in the reaction when running opposite to the "favoured" direction. In this direction, the reactants (as defined by the kinetic rate law) are being produced.
- **"both"**: the value of the `role` attribute is the role of the referenced variable in both directions of the reaction. This value for the `direction` attribute may only appear on roles of **"catalyst"**, **"activator"**, **"inhibitor"**, or **"modifier"**. It is not chemically sensible to say that a species is a **"reactant"** in both directions. Nor does it make sense to declare that a species is a **"product"** in both directions.

The `direction` attribute is optional, and defaults to **"forward"**.

3.2.4 Allowed use of the `stoichiometry` attribute

- The `stoichiometry` attribute may take any numeric value. In most cases, the value will be an integer. However, a valid CellML model may use fractional stoichiometries.
- The `stoichiometry` attribute is optional, and has no default value. The absence of a `stoichiometry` attribute formally implies nothing, but would usually mean that the stoichiometry of the reaction is unknown. Note that the absence of a stoichiometry value specifically does **not** imply a stoichiometry of **"1"**.
- It is possible to define an arbitrary function to relate the change in concentration of a chemical species to the reaction rate. If this is done, it is an error to use the `stoichiometry` attribute on the reactant or product role for this chemical species. It is allowable to use the `stoichiometry` attribute on other roles of the species.
- When creating a chemical expression representation of the model, CellML-compliant processing applications are only required to correctly represent stoichiometry as defined by the `stoichiometry` attribute. Stoichiometry defined using explicit MathML may be ignored when *representing* the model, but must be used when *running* the model, provided the equations defined by the modeller are self-consistent.

3.2.5 Proper use of the `delta_variable` attribute

The following rules define how a valid CellML document may use the `delta_variable` attribute :

- The `delta_variable` attribute is optional and has no default value.
- The `delta_variable` attribute may only appear on `<role>` elements in which the `role` attribute equals "reactant" or "product". It is only in these roles that a chemical species may undergo a change in concentration.
- Any variable for which a `delta_variable` attribute is declared must also have a `stoichiometry` attribute on the parent `<role>` element. This is because the combination of the `delta_variable` attribute and the `stoichiometry` attribute implies a mathematical relationship between the variable named in the `delta_variable` attribute and the variable assigned the role of "rate", as defined in Section 3.2.6. However, a `stoichiometry` attribute may be declared in the absence of a `delta_variable` attribute.
- If a `delta_variable` is declared for any reaction participant, a `<variable_ref>` element must be provided for the variable that represents the reaction rate. This `<variable_ref>` must contain exactly one `<role>` element, with a `role` attribute equal to "rate". Note that the reverse is not true: a variable may be assigned a role of "rate" even if the "reactant" and "product" variables do not have declared `delta_variable` attributes. In this case, the modeller must provide explicit mathematics relating the "rate" variable to the change in concentration of the various chemical species.

3.2.6 Math implied by the `delta_variable` and `stoichiometry` attributes

The use of the `delta_variable` and `stoichiometry` attributes on a `<role>` element implies the following mathematical relationship between the declared `delta_variable` and the rate variable:

- For reactants: $\text{delta_variable} = (\text{stoichiometry})(\text{rate})$
- For products: $\text{delta_variable} = -(\text{stoichiometry})(\text{rate})$

The two CellML fragments shown in Figure 3 and Figure 4 are mathematically equivalent. The representation in Figure 3 is the recommended best practice, because processing applications are not required to be able to extract the stoichiometry from an explicit MathML definition such as the one shown in Figure 4. Explicit mathematics should only be used in cases where the math implied from the recommended formulation would be incorrect, i.e., in the rare cases when a more complex function is needed to relate the change in concentration of a species to the reaction rate. In these cases, it is recommended best practice to put the mathematical expression used to define the change in concentration inside the `<variable_ref>` element for the variable representing the concentration.

It is an error to explicitly declare mathematics that conflicts with or duplicates implied mathematics. Therefore, a modeller cannot declare a `stoichiometry` attribute and `delta_variable` attribute in addition to explicit math relating the change in concentration of the referenced species to the reaction rate.

3.3 Resolution of inconsistencies

Duplication of information is avoided as much as possible. However, because of the requirement that modellers be free to define arbitrary rate laws, it was not possible to eliminate all information duplication. For instance, it would place an undue burden on processing applications to expect software to be able to deduce all information about a reaction from kinetic laws of arbitrary form, even though most information is in fact represented in these laws. Therefore, there is a possibility that the information in the mathematics

```

<reaction reversible="yes">
  <variable_ref variable="A">
    <role
      role="reactant" direction="forward"
      delta_variable="delta_A" stoichiometry="1" />
    </variable_ref>
    ...
  <variable_ref variable="r">
    <role role="rate" direction="forward" />
    <math xmlns="http://www.w3.org/1998/Math/MathML">
      ...
    </math>
  </variable_ref>
</reaction>

```

FIGURE 3: A CellML fragment showing the recommended method for defining the change in concentration of a chemical species A with respect to the reaction rate r . The MathML block defining the rate law is omitted.

and the information in the `<reaction>` element may be inconsistent. It is anticipated that most modellers will define CellML models using some sort of processing software, which can reasonably be expected to write consistent CellML. However, since CellML is a text-based format, modellers may also create or edit models by hand, and in doing so risk creating inconsistent models. The following rules govern the required behavior of CellML-compliant processing software in the event that information in the mathematics and the information in the reaction element do not agree:

- Preference is given to mathematics explicitly defined using MathML when **running** a simulation with the model.
- It is left to the processing software to decide which information to use in **representing** the model. Software is free to ignore the mathematics when creating a pathway diagram or chemical expression rendering of the model. However, software should clearly document which information is used to create representations of the model.
- Processing software may check for inconsistencies between the mathematics and the information in the `<reaction>` element. However, it is not required to do so, and it is left to the processing software to determine what to do if an inconsistency is found.

3.4 Qualitative vs. quantitative pathway models

Qualitative pathway models consist solely of information about how the different chemical species in the pathway relate. There are no kinetic rate laws describing the progress of the reactions that relate the species. However, the stoichiometry of the reactions may be known. Because there is no math in a qualitative model, a CellML-compliant processing application is not required to be able to run a simulation using a qualitative model. However, some applications may support simple simulations using such models.

Any model in which the change of concentration of a chemical species participating in a reaction is implicitly or explicitly defined (using a `delta_variable` attribute or an equation setting the value of a

```
<reaction reversible="yes">

  <variable_ref variable="A">
    <role
      role="reactant" direction="forward" />
    <math xmlns="http://www.w3.org/1998/Math/MathML">
      <apply><eq />
      <ci> delta_A </ci>
      <apply><times />
      <cn cellml:units="dimensionless"> 1.0 </cn>
      <ci> r </ci>
    </apply>
  </math>

</variable_ref>

...

<variable_ref variable="r">
  <role role="rate" direction="forward" />
  <math xmlns="http://www.w3.org/1998/Math/MathML">
    ...
  </math>
</variable_ref>

</reaction>
```

FIGURE 4: A CellML fragment showing the method for defining the change in concentration of a chemical species A with respect to the reaction rate r with explicit mathematics. The MathML block defining the rate law is omitted. Use of this formulation is not recommended.

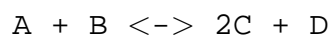
delta variable) is quantitative. All others are qualitative.

4 Examples

This section contains examples demonstrating the recommended use of `<reaction>` and `<role>` elements to define some common reactions.

4.1 Simple Reversible Reaction

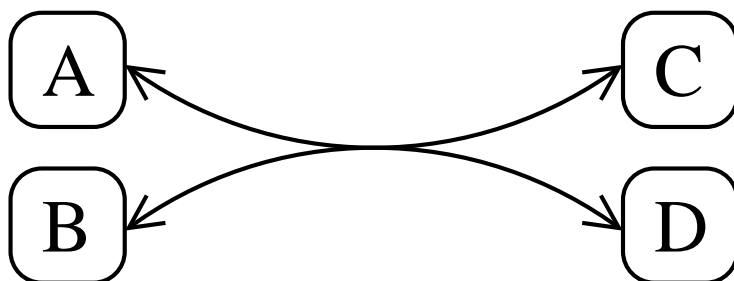
- **Equation**



- **Description**

A simple reversible reaction.

- **Picture**



- **CellML Definition**

```
<reaction reversible="yes">
  <variable_ref variable="A">
    <role
      role="reactant" direction="forward"
      delta_variable="delta_A" stoichiometry="1" />
  </variable_ref>
  <variable_ref variable="B">
    <role
      role="reactant" direction="forward"
      delta_variable="delta_B" stoichiometry="1" />
  </variable_ref>
  <variable_ref variable="C">
    <role
      role="product" direction="forward"
      delta_variable="delta_C" stoichiometry="2" />
  </variable_ref>
  <variable_ref variable="D">
    <role
      role="product" direction="forward"
      delta_variable="delta_D" stoichiometry="1" />
  </variable_ref>
</reaction>
```

```

<variable_ref variable="r">
  <role
    role="rate" direction="forward" />
  <math xmlns="http://www.w3.org/1998/Math/MathML">
    ...
  </math>
</variable_ref>

</reaction>

```

4.2 Simple Reversible Reaction with Catalysis

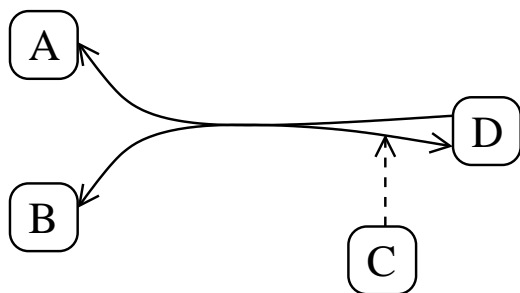
- Equation

$A + B \rightleftharpoons D$ (catalyzed in the forward direction by C)

- Description

A reversible reaction where C acts as a catalyst in the forward direction only.

- Picture



- CellML Definition

```

<reaction reversible="yes">

  <variable_ref variable="A">
    <role
      role="reactant" direction="forward"
      delta_variable="delta_A" stoichiometry="1" />
  </variable_ref>

  <variable_ref variable="B">
    <role
      role="reactant" direction="forward"
      delta_variable="delta_B" stoichiometry="1" />
  </variable_ref>

  <variable_ref variable="C">
    <role role="catalyst" direction="forward" />
  </variable_ref>

  <variable_ref variable="D">
    <role
      role="product" direction="forward"
      delta_variable="delta_D" stoichiometry="1" />
  </variable_ref>

  <variable_ref variable="r">

```

```

    <role
      role="rate" direction="forward" />
    <math xmlns="http://www.w3.org/1998/Math/MathML">
      ...
    </math>
  </variable_ref>

</reaction>

```

4.3 Autophosphorylation

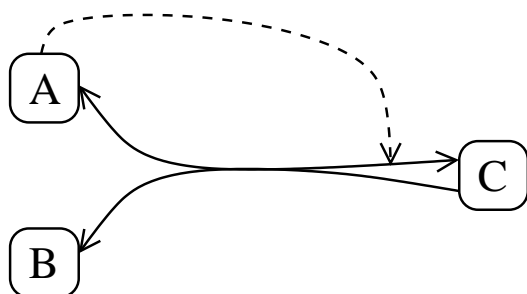
- **Equation**

$A + B \rightleftharpoons C$ (catalyzed by A in the forward direction)

- **Description**

A autocatalyzes a reaction (e.g., an autophosphorylation, where B is phosphate).

- **Picture**



- **CellML Definition**

```

<reaction reversible="yes">
  <variable_ref variable="A">
    <role
      role="reactant" direction="forward"
      delta_variable="delta_A" stoichiometry="1" />
    <role role="catalyst" direction="forward" />
  </variable_ref>
  <variable_ref variable="B">
    <role
      role="reactant" direction="forward"
      delta_variable="delta_B" stoichiometry="1" />
  </variable_ref>
  <variable_ref variable="C">
    <role
      role="product" direction="forward"
      delta_variable="delta_C" stoichiometry="1" />
  </variable_ref>
  <variable_ref variable="r">
    <role role="rate" direction="forward" />
  </variable_ref>
</reaction>

```

4.4 Simple Reaction with Catalysis and Substrate Inhibition

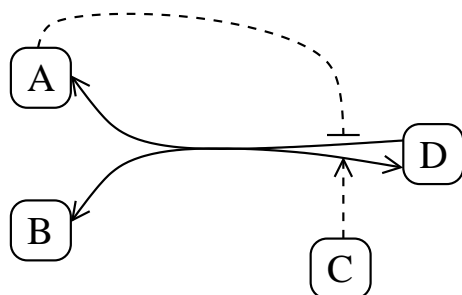
- **Equation**

$A + B \rightleftharpoons D$ (catalyzed by C, inhibited by A)

- **Description**

A reversible reaction in which the reverse reaction is inhibited by one of the substrates.

- **Picture**



- **CellML Definition**

```
<reaction reversible="yes">
  <variable_ref variable="A">
    <role
      role="reactant" direction="forward"
      delta_variable="delta_A" stoichiometry="1" />
    <role
      role="inhibitor" direction="reverse"
      delta_variable="delta_A" stoichiometry="1" />
  </variable_ref>
  <variable_ref variable="B">
    <role
      role="reactant" direction="forward"
      delta_variable="delta_B" stoichiometry="1" />
  </variable_ref>
  <variable_ref variable="C">
    <role
      role="catalyst" direction="forward" />
  </variable_ref>
  <variable_ref variable="D">
    <role
      role="product" direction="forward"
      delta_variable="delta_D" stoichiometry="1" />
  </variable_ref>
  <variable_ref variable="r">
    <role
      role="rate" direction="forward" />
    <math xmlns="http://www.w3.org/1998/Math/MathML">
      ...
    </math>
  </variable_ref>
</reaction>
```

4.5 Simple Reaction with Catalysis and Product Inhibition

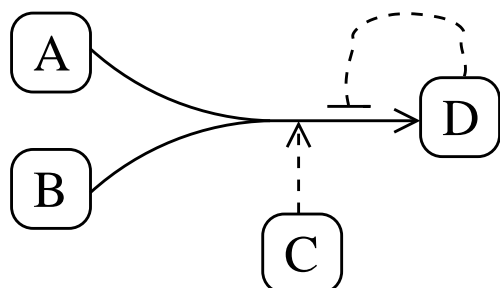
- **Equation**

$A + B \rightarrow D$ (catalyzed by C, inhibited by D)

- **Description**

An irreversible, catalyzed reaction that exhibits product-inhibition.

- **Picture**



- **CellML Definition**

```
<reaction reversible="no">
  <variable_ref variable="A">
    <role role="reactant" delta_variable="delta_A" stoichiometry="1" />
  </variable_ref>
  <variable_ref variable="B">
    <role role="reactant" delta_variable="delta_B" stoichiometry="1" />
  </variable_ref>
  <variable_ref variable="C">
    <role role="catalyst" />
  </variable_ref>
  <variable_ref variable="D">
    <role role="product" delta_variable="delta_D" stoichiometry="1" />
    <role role="inhibitor" stoichiometry="1" />
  </variable_ref>
  <variable_ref variable="r">
    <role role="rate" />
    <math xmlns="http://www.w3.org/1998/Math/MathML">
      ...
    </math>
  </variable_ref>
</reaction>
```