



# Converting KEGG pathway database to SBML

Akira Funahashi<sup>1,2,\*</sup>, Akiya Jouraku<sup>2,\*</sup>, Hiroaki Kitano<sup>1,2,3,4</sup>

<sup>1</sup>JST/ERATO-SORST Kitano Symbiotic Systems Project, Tokyo, Japan

<sup>2</sup>School of Fundamental Science and Technology, Keio University, Kanagawa, Japan

<sup>3</sup>The Systems Biology Institute, Tokyo, Japan

<sup>4</sup>Sony Computer Science Labs, Tokyo, Japan

\* These authors contributed equally to this work.

## Abstract

Systems biology is characterized by the synergistic integration of theory, computational modeling, and experiment. Though software infrastructure is one of the most critical components of systems biology research, there is no common infrastructure or standard to enable integration of computational resources. The Systems Biology Markup Language (SBML) was developed to help address this problem. SBML is an open, XML-based format for representing biochemical reaction networks. Several dozen simulation and analysis packages already support SBML and more in the process of being extended to support it.

Identification of gene-regulatory logic and biochemical networks is a major challenge of systems biology. Several attempts are underway to create large-scale, comprehensive databases of gene-regulatory and biochemical networks. Making the contents of these databases available in SBML format is useful for the following reasons:

- (1) it will enable researchers to apply many SBML-aware software tools to the networks in those databases, and
- (2) the feedback from developing the translation tools will provide valuable feedback for the continued evolution of SBML.

As a first attempt at writing translation tools, we have decided to convert the KEGG (Kyoto Encyclopedia of Genes and Genomes) database. We have implemented a converter called **KEGG2SBML** that automatically converts KEGG pathway database files into SBML.

With this converter (KEGG2SBML), **we have succeeded in converting 10,860 KEGG pathways into SBML Level 1 and Level 2 documents.** All converted SBML documents are freely available. KEGG2SBML is available as an open-source package.

## KEGG : Kyoto Encyclopedia of Genes and Genomes

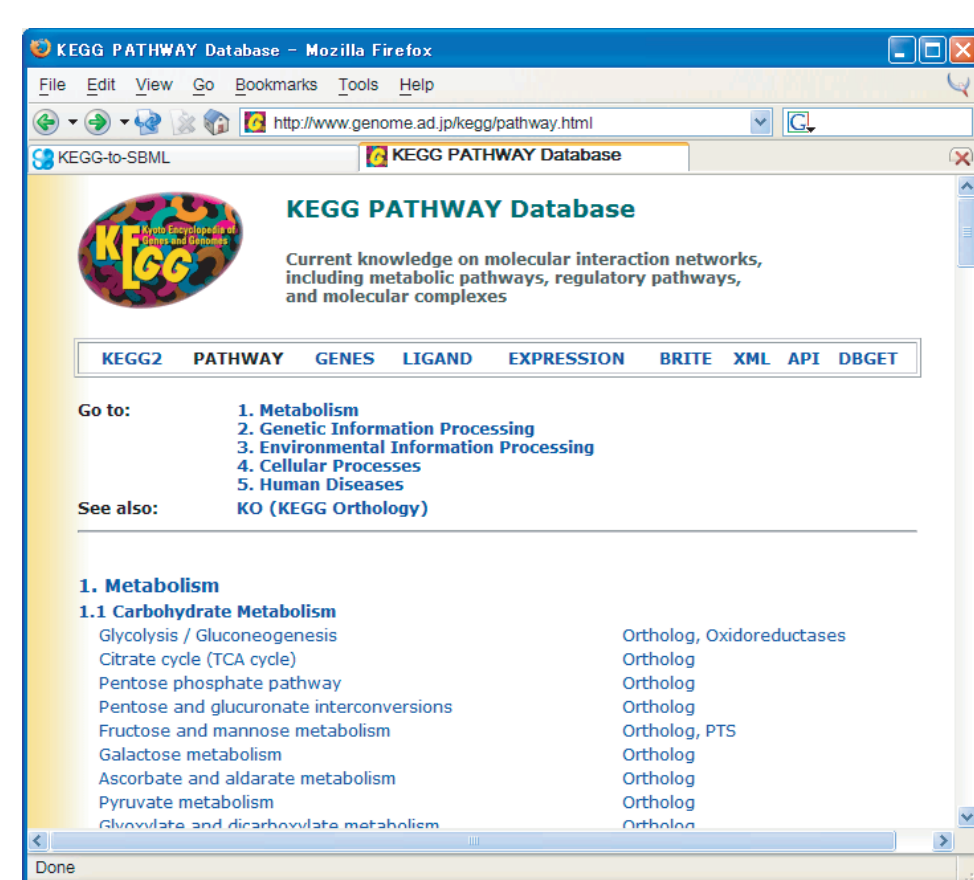
• KEGG database contains more than 13,000 metabolic pathways

for more than 150 organisms.

• <http://www.genome.ad.jp/kegg/>

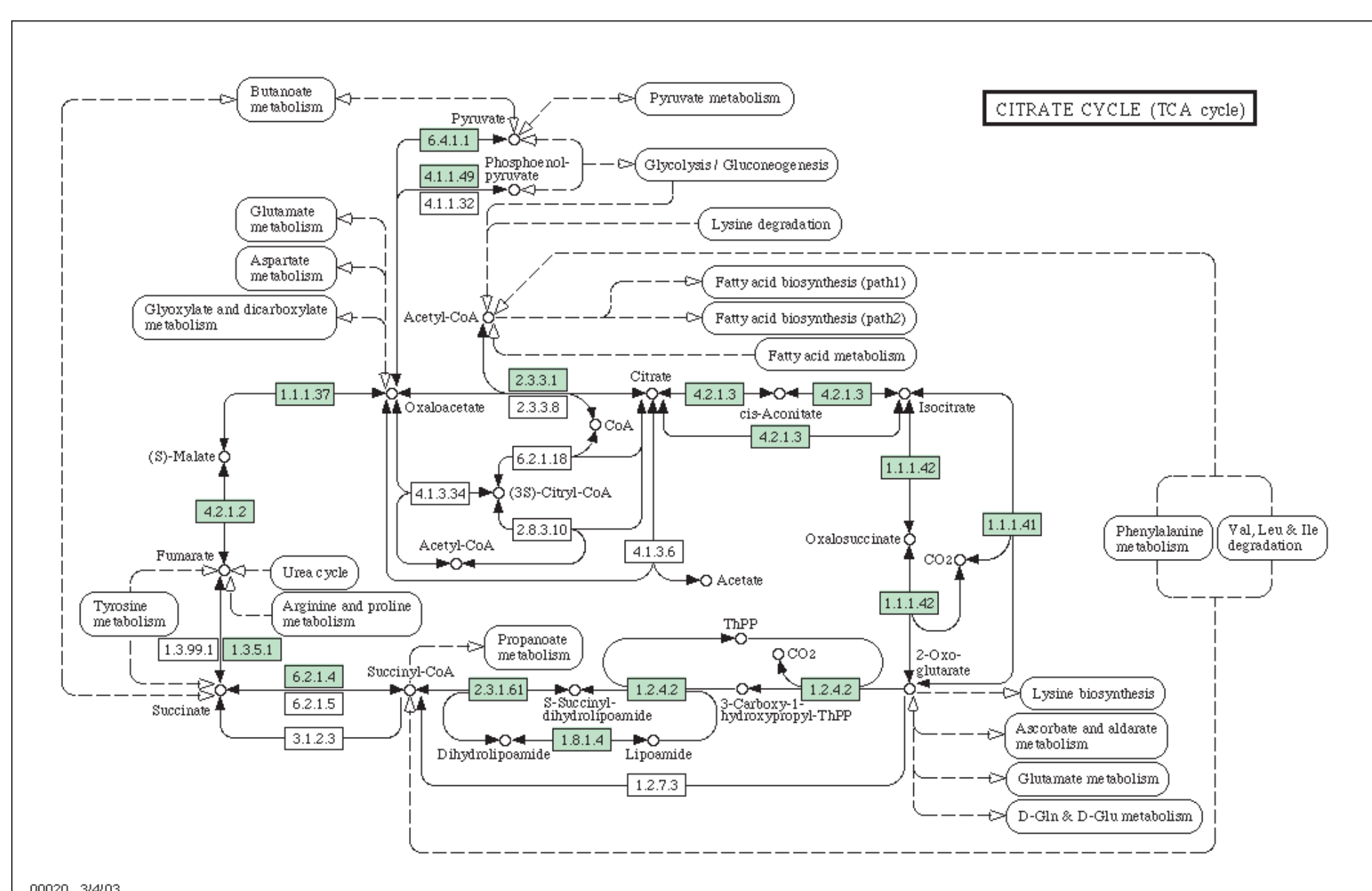
• Pathway Database

1. Metabolism
2. Genetic Information Processing
3. Environmental Information Processing
4. Cellular Processes
5. Human Diseases

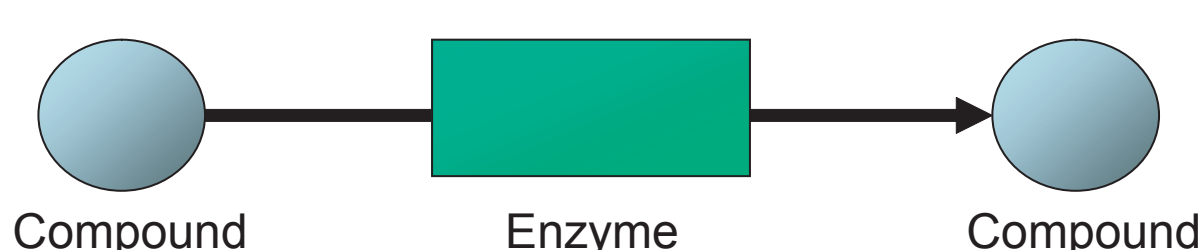


## KEGG metabolic pathways

KEGG database contains variety of metabolic pathways, which consists of compounds and enzymes. For example, the citrate cycle (TCA cycle) of *Saccharomyces cerevisiae* is represented as shown below. Compounds are represented as circle nodes, and enzymes are represented as rectangle nodes. A reaction between these compounds is represented as an arrow, and the enzyme which corresponds to this reaction is placed on the arrow.



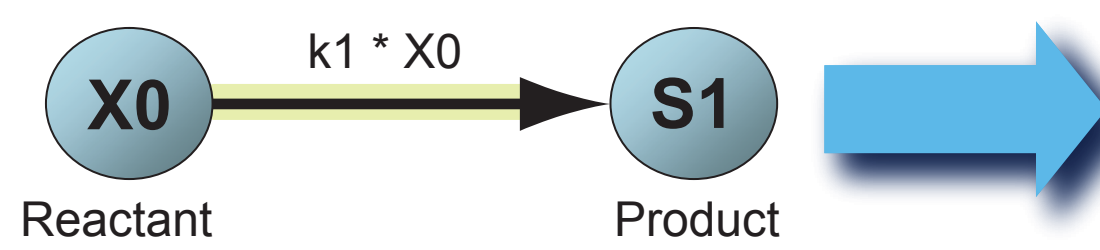
Citrate cycle (TCA cycle) of *Saccharomyces cerevisiae*



Example of a reaction used in KEGG metabolic pathways

## SBML

The Systems Biology Markup Language (SBML) is an XML dialect for representing and exchanging quantitative and qualitative models of biochemical reaction networks. SBML is applicable to metabolic networks, cell-signaling pathways, genomic regulatory networks, and many other areas in systems biology.



Biochemical reaction

```

<listOfSpecies>
  <species name="X0"/>
  <species name="S1"/>
</listOfSpecies>
<listOfReactions>
  <reaction name="reaction_1"/>
    <listOfReactants>
      <speciesReference species="X0"/>
    </listOfReactants>
    <listOfProducts>
      <speciesReference species="S1"/>
    </listOfProducts>
    <kineticLaw formula="k1 * X0"/>
    </kineticLaw>
  </reaction>
</listOfReactions>

```

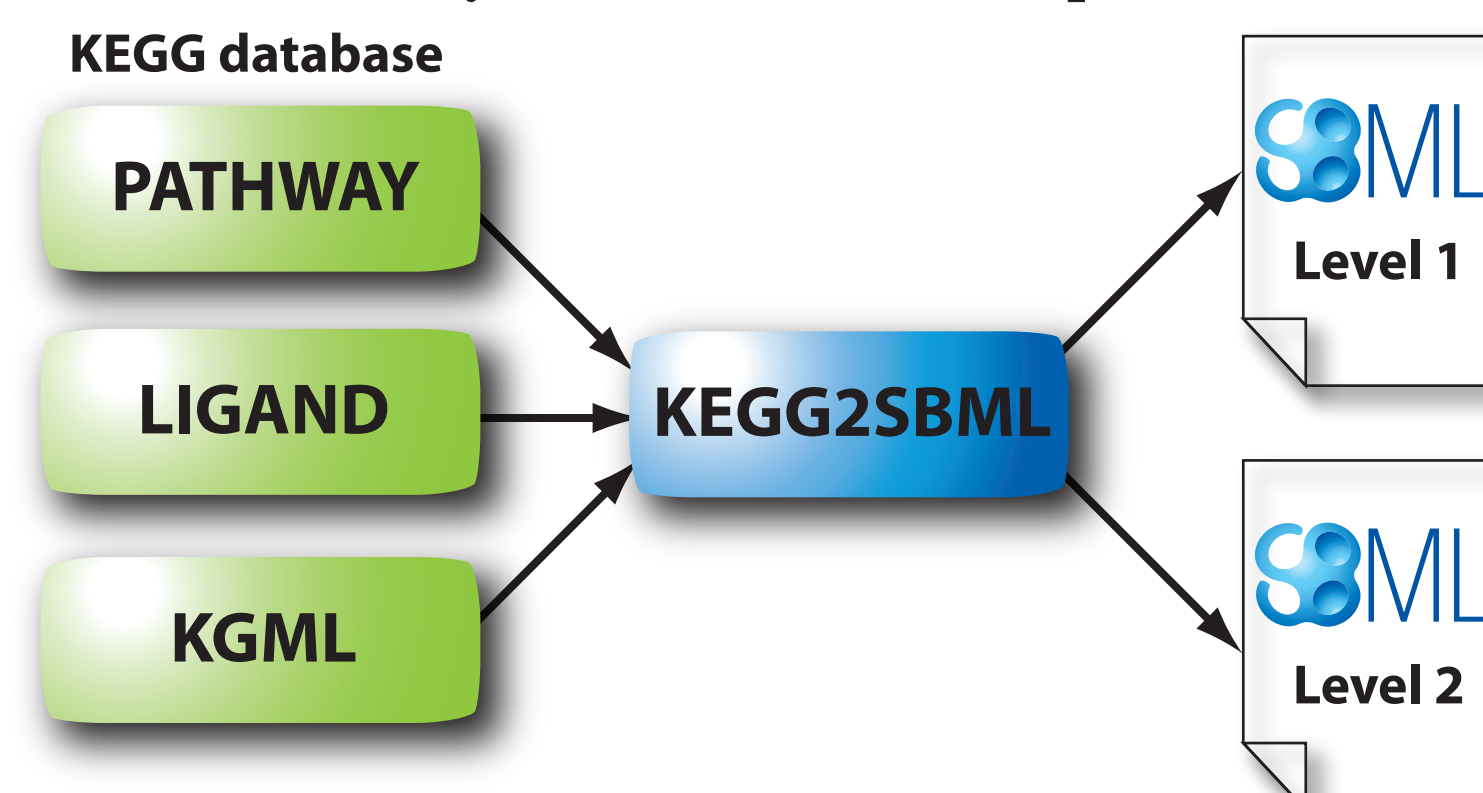
SBML representation

## KEGG to SBML converter

We have implemented a converter called KEGG2SBML that automatically converts KEGG pathway database files into SBML Level 1 and Level 2 files. KEGG2SBML uses PATHWAY database, LIGAND database and KEGG Markup Language (KGML) as an input to generate SBML documents. LIGAND database is a collection of information about biochemical compounds and reactions, and KGML is a specification of graph objects in KEGG PATHWAY database. Further, KEGG2SBML can parse diagram layout information from KEGG and add it to SBML; the result can be used in CellDesigner, a process network diagram editor we have also developed. KEGG2SBML is implemented with Perl5 (>= 5.6.1), and requires following libraries:

- expat >= 1.95.2
- XML::Parser >= 2.1.9
- libxml-perl >= 0.07

KEGG2SBML has been tested on UNIX based operating systems such as Linux and FreeBSD. The converter may run on other UNIX platforms and Cygwin under Windows.



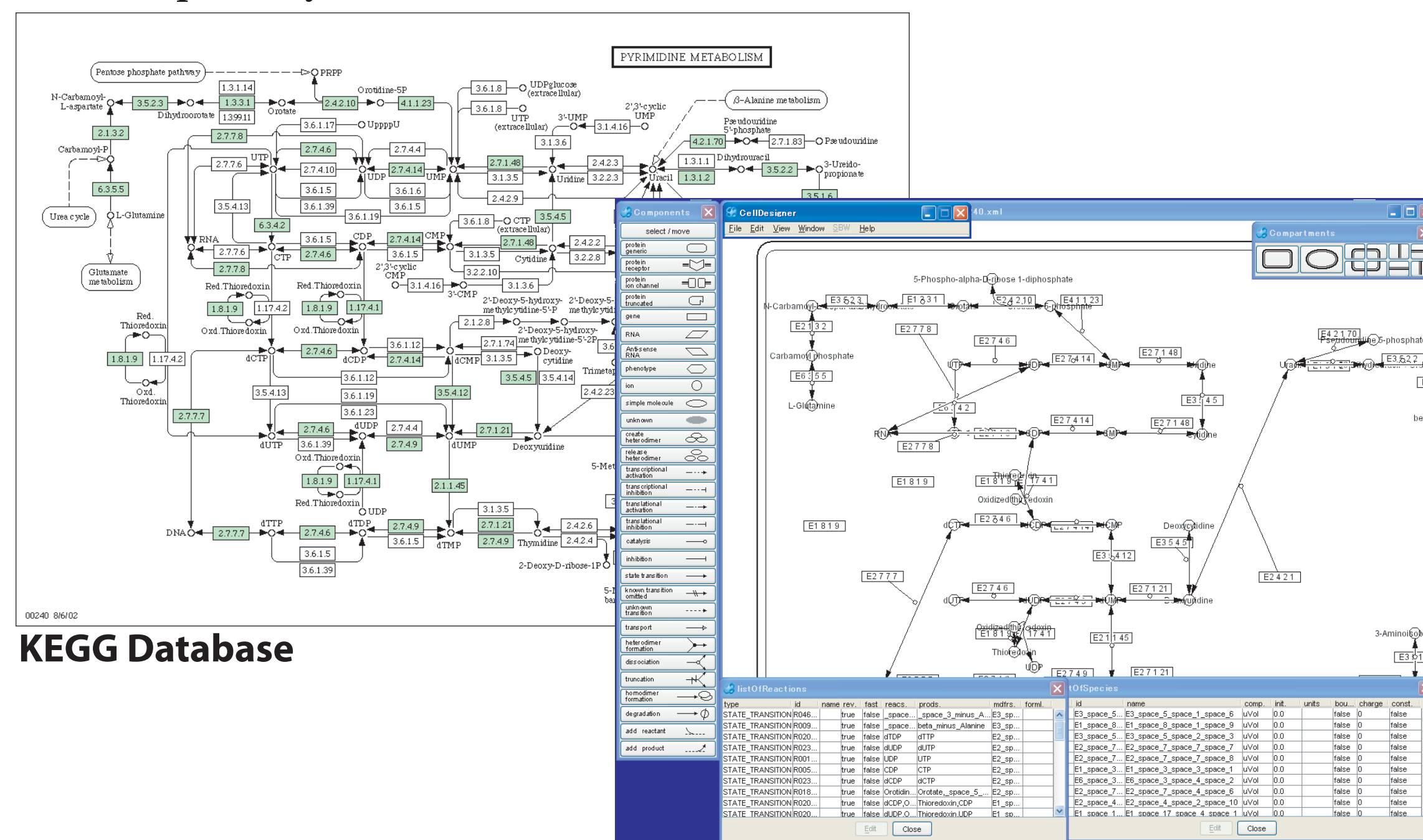
Structure of KEGG-to-SBML converter

## KEGG2SBML is freely available as an open-source package from

<http://sbml.org/software/kegg2sbml/>

## Converted SBML models

We have succeeded in converting 10,860 KEGG metabolic pathways into SBML Level 1 and Level 2 documents. Existing SBML-aware applications can directly use these pathways.



Converted SBML models are available from <http://systems-biology.org>

