

# **SYSTEMATIC ANALYSIS AND VISUALIZATION OF PATHWAYS USING KEGG AND SBML**

Michael Dondrup

MCB Research Course: Integrative Bioinformatics

November 24, 2009

# Outline

1 Introduction

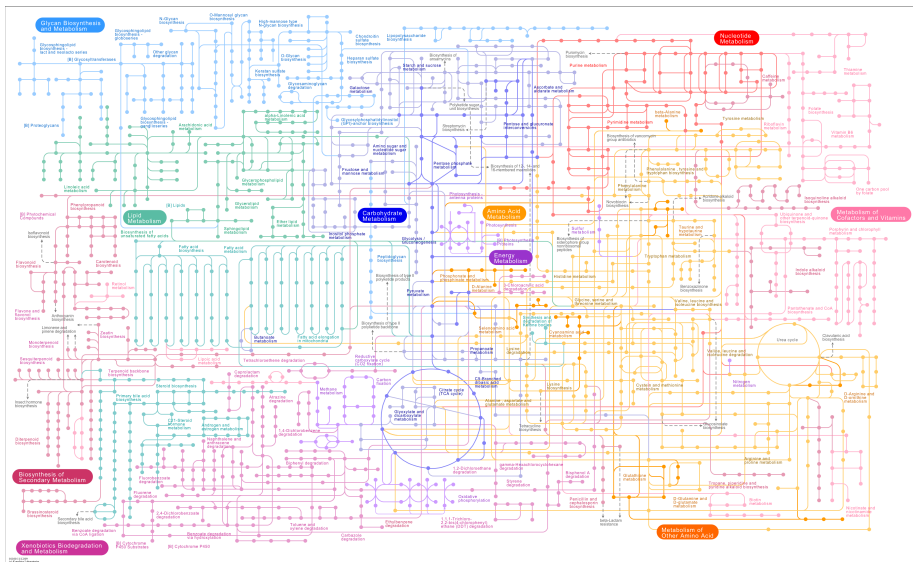
2 Conversion

3 Exercises

## KEGG: Kyoto Encyclopedia of Genes and Genomes

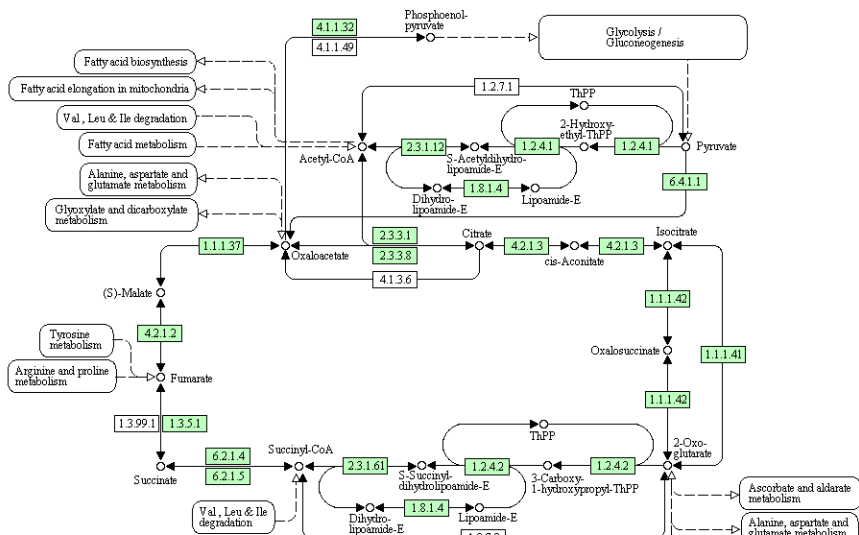
- <http://www.genome.jp/kegg/>
- Large data collection of genes, functional categories (BRITE), genome annotations, compounds
- biochemical reactions: enzymes and compounds (LIGAND)
- Visualization of biochemical pathways, know as KEGG pathway maps
- Organism specific pathway maps with hyperlinks
- uses Enzyme Commission numbers (EC number)
- contains metabolic and non-metabolic (e.g. structural, cancer, regulatory) pathways
- free for academic use!

# KEGG in 2 minutes



# KEGG in 2 minutes

## CITRATE CYCLE (TCA CYCLE)



## SBML: Systems Biology Markup Language

- <http://sbml.org>
- a language for modelling biological processes
- machine readable text: XML
- standardised exchange format
- can be use to model metabolic reactions
- and other general networks
- graphical editors exist eg.: CellDesigner  
<http://www.celldesigner.org/>
- we will use SBML as an exchange format for CellDesigner and ProMeTra

# SBML

The screenshot displays the SBML Editor software interface. At the top, there is a toolbar with various icons for file operations and editing. Below the toolbar, the main workspace shows a metabolic pathway diagram with the following components:

- Species:** Oxaloacetate, Pyruvate, S-Malate, and Kumarate (represented by ovals).
- Proteins:** mdh2, pycB, and fumX (represented by rectangles).
- Reactions:** mdh2 catalyzes the conversion of Oxaloacetate to S-Malate; pycB catalyzes the conversion of Pyruvate to Oxaloacetate; fumX catalyzes the conversion of S-Malate to Kumarate.

On the left side, there is a navigation pane with a tree view containing "Model", "Compartments", "Species", and "Reactions". Below it is a "Layer" section with a "base" layer selected.

At the bottom, there is a table with the following columns: class, id, name, compar..., positio..., quantit..., initi..., subst...:

class	id	name	compar...	positio...	quantit...	initi...	subst...
PROTEIN	pycB	pycB	uVol	inside	Amount	0.0	
SIMPLE...	_sp...	C00149	uVol	inside	Amount	0.0	
SIMPLE...	Fum...	C00122	uVol	inside	Amount	0.0	
SIMPLE...	Oxa...	C00036	uVol	inside	Amount	0.0	
SIMPLE...	Pyru...	C00022	uVol	inside	Amount	0.0	

Below the table, there are "Edit" and "Export" buttons. To the right of the table, there are "Edit Notes" and "Edit Protein Notes" buttons. The "Species" tab is selected, and the following text is displayed:

Species (id=fumX, name=fumX; aae00020.xml)  
Protein (id=p0, name=fumX)

None of the conversion tools can provide 100% correct conversion, and both have their problems

KEGG Converter:

- <http://www.grissom.gr/keggconverter>
- Java application and web-site

KEGG2SBML:

- originally from <http://sbml.org/Software/KEGG2SBML>
- Perl program, we provide an improved version for this course:
- <http://www.bccs.uib.no/~mdo041/mcbcourse/>
- This is a commandline program

# Installing kegg2sbml

- Download the file named kegg2hbml-1.6.0.tar.gz to O: `http://www.bccs.uib.no/~mdo041/mcbcourse/kegg2sbml/`
- Extract the archive to O: by double clicking it
- Open a command prompt (type "cmd" in Start – Run)
- type:
  - o: `<Enter>`
  - `cd kegg2sbml-1.6.0<Enter>` depending on where you extracted it
- type `perl -I.\lib bin\kegg2sbml <Enter>` to test it and see a list of options
- to try that at home, you can get perl from: `http://cpan.org` or try ActivePerl: `http://www.activestate.com/activeperl/`

# Installing the ligand database

- download the file "keggligand.tar.gz" to O: from `http://www.bccs.uib.no/~mdo041/mcbcourse/kegg2sbml/`
- and extract it to drive O:, there is a `O:\KEGG\ligand` directory
- you could also have kegg2sbml download all files itself, but that takes very long
- see the example output of `kegg2sbml -help` for how to get the database from kegg

# Converting pathways

- `O:\kegg2sbml-1.6.0> perl -I.\lib\ bin\kegg2sbml -i -fn -a O:\mypathwaydata -g O:\KEGG\ligand hsa00010 hsa00020`
- the program will download the pathway coordinates from kegg
- the result is save in the SBML folder
- open the results in CellDesigner and adjust the positions of the boxes
- download the same pathways from the KEGG converter site and compare the results

# Remarks

- this does currently not work for non-metabolic pathways
- we are building a more convenient solution in the eSysbio project
- we will update the course page to let you know
- for now: it's a good idea to build your pathway model yourself in cell designer
- based on the output of one of the converters