

Editing Pathway/Genome Databases I

By Ron Caspi

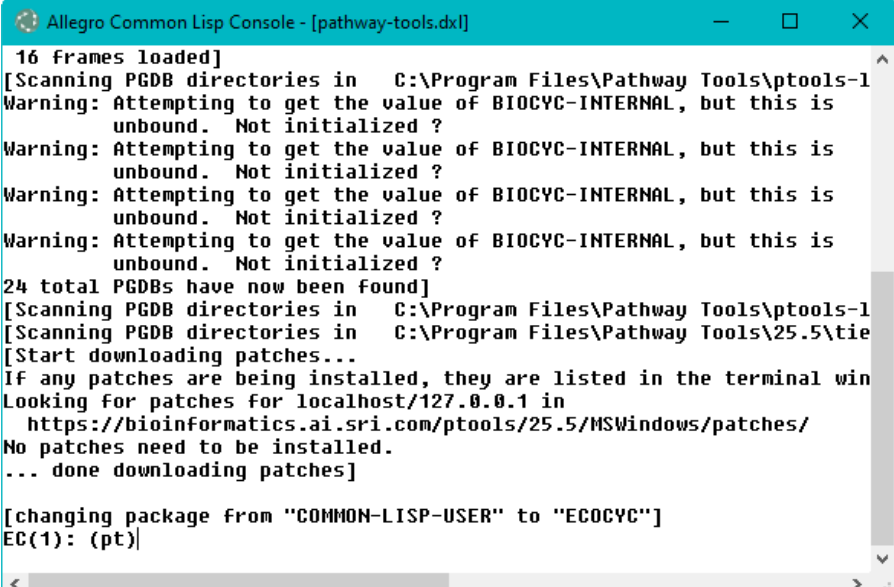
ron.caspi@sri.com



A copy of this presentation could be found at
<http://bioinformatics.ai.sri.com/ptools/tutorial/sessions/curation>

Starting Pathway Tools

- Start the software with the Lisp option
- The Lisp console will show up
- At the prompt type (pt) (followed by Enter) to start the graphical user interface (GUI)
- If you close the GUI, the Lisp console will continue to run
- Exit by typing (exit)



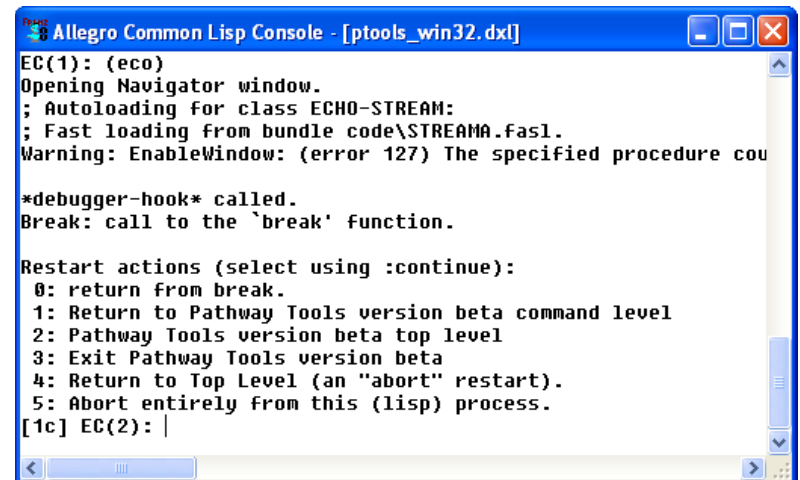
```
Allegro Common Lisp Console - [pathway-tools.dxl]
16 frames loaded]
[Scanning PGDB directories in C:\Program Files\Pathway Tools\ptools-1
Warning: Attempting to get the value of BIOCYC-INTERNAL, but this is
unbound. Not initialized ?
Warning: Attempting to get the value of BIOCYC-INTERNAL, but this is
unbound. Not initialized ?
Warning: Attempting to get the value of BIOCYC-INTERNAL, but this is
unbound. Not initialized ?
Warning: Attempting to get the value of BIOCYC-INTERNAL, but this is
unbound. Not initialized ?
24 total PGDBs have now been found]
[Scanning PGDB directories in C:\Program Files\Pathway Tools\ptools-1
[Scanning PGDB directories in C:\Program Files\Pathway Tools\25.5\tie
[Start downloading patches...
If any patches are being installed, they are listed in the terminal win
Looking for patches for localhost/127.0.0.1 in
  https://bioinformatics.ai.sri.com/ptools/25.5/MSWindows/patches/
No patches need to be installed.
... done downloading patches]

[changing package from "COMMON-LISP-USER" to "ECOCYC"]
EC(1): (pt)|
```

Working in a Lisp environment: breaks

- A break shifts the focus from the main GUI to the Lisp console
- To generate a break: BREAK key on your keyboard. If you don't have one, type (break) at the listener pane and hit enter
- When Lisp encounters an unrecognized command, it breaks
- A break is NOT a crash

Lisp presents several recovery options from a break
Type `:cont x` where x is the number of the best option



```
Allegro Common Lisp Console - [ptools_wln32.dxl]
EC(1): (eco)
Opening Navigator window.
; Autoloading for class ECHO-STREAM:
; Fast loading from bundle code\STREAMA.fasl.
Warning: EnableWindow: (error 127) The specified procedure cou

*debugger-hook* called.
Break: call to the `break' function.

Restart actions (select using :continue):
0: return from break.
1: Return to Pathway Tools version beta command level
2: Pathway Tools version beta top level
3: Exit Pathway Tools version beta
4: Return to Top Level (an "abort" restart).
5: Abort entirely from this (lisp) process.
[1c] EC(2): |
```

Bug reports

If you get a break as a result of a bug, get the evaluation stack by typing

`:zo :count :15` at the lisp prompt

Copy the output, and send it by email to ptools-support@ai.sri.com



```
Allegro Common Lisp Console [ptools_win32.dxl]
*debugger-hook* called.
Break: call to the 'break' function.

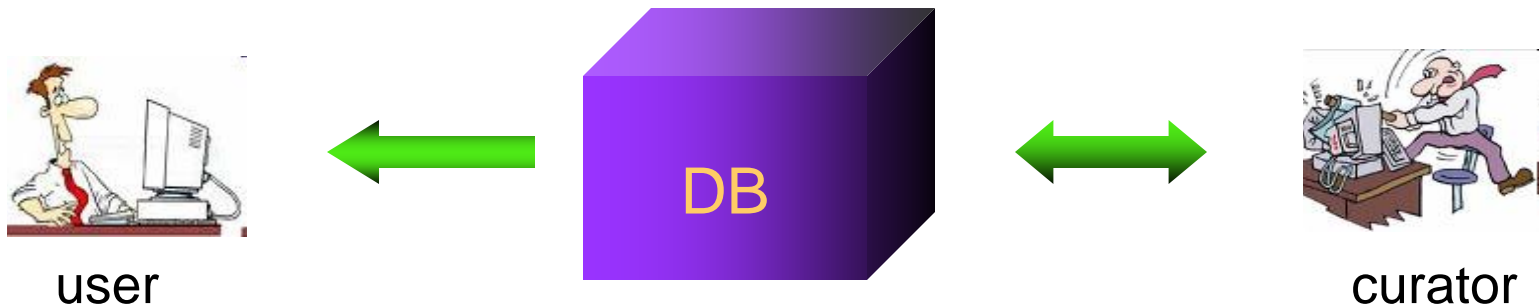
Restart actions (select using :continue):
0: return from break.
1: Return to Pathway Tools version beta command level
2: Pathway Tools version beta top level
3: Exit Pathway Tools version beta
4: Return to Top Level (an "abort" restart).
5: Abort entirely from this (lisp) process.
[1c] EC(2): :zo :count :all
Evaluation stack:
(BREAK)
[... EXCL::%EVAL ]
->(EVAL (BREAK))
((METHOD CLIM:READ-FRAME-COMMAND (ECOCVC)) #<ECOCVC @ #x2391bf6a>
 :STREAM
 #<CLIM:INTERACTOR-PANE
 @
 #x2391ba22>)
((:INTERNAL (:EFFECTIVE-METHOD 1 T NIL NIL T) 0)
 #<ECOCVC @ #x2391bf6a> . #<non-lisp object @ #x3>)
((METHOD CLIM:READ-FRAME-COMMAND :AROUND
 (CLIM:STANDARD-APPLICATION-FRAME))
 #<ECOCVC @ #x2391bf6a> :STREAM
 #<CLIM:INTERACTOR-PANE @ #x2391ba22>)
((:INTERNAL (:EFFECTIVE-METHOD 1 T T T) 0) #<ECOCVC @ #x2391bf6a>
 :STREAM
 #<CLIM:INTERACTOR-PANE
 @
 #x2391ba22>)
((METHOD CLIM:DEFAULT-FRAME-TOP-LEVEL
 (CLIM:STANDARD-APPLICATION-FRAME))
 #<ECOCVC @ #x2391bf6a>)
((:INTERNAL (:EFFECTIVE-METHOD 1 T T NIL NIL) 0)
 #<ECOCVC @ #x2391bf6a>)
((METHOD CLIM:RUN-FRAME-TOP-LEVEL (CLIM:STANDARD-APPLICATION-FRAME))
 #<ECOCVC @ #x2391bf6a>)
((:INTERNAL (:EFFECTIVE-METHOD 1 T NIL NIL T) 0)
 #<ECOCVC @ #x2391bf6a> . #<non-lisp object @ #x1>)
((METHOD CLIM:RUN-FRAME-TOP-LEVEL :AROUND
 (CLIM:STANDARD-APPLICATION-FRAME))
 #<ECOCVC @ #x2391bf6a>)
((:INTERNAL (:EFFECTIVE-METHOD 1 T NIL T T) 0)
 #<ECOCVC @ #x2391bf6a> . #<non-lisp object @ #x1>)
(ECO)
[... EXCL::%EVAL ]
(EVAL (ECO))
(TPL:TOP-LEVEL-READ-EVAL-PRINT-LOOP)
(TPL:START-INTERACTIVE-TOP-LEVEL
 #<EXCL:TERMINAL-SIMPLE-STREAM [initial terminal io] fd 0/1 @
 #x2008e242>
 #<Function TOP-LEVEL-READ-EVAL-PRINT-LOOP> NIL)
[1c] EC(3): |
```

Pathway Tools in editing mode

The database can be accessed by two distinct modes

- Navigator mode allows no modification of the DB
- Editing mode allows complete modification of the DB

Editing is currently available in **Desktop mode only**, though online editors are being developed



Installing an editable PGDB

In order to be able to perform editing, you must have an editable PGDB installed on your system (built-in PGDBs can't be edited).



In the following exercises we will be using the PGDB for *Arthrospira platensis* NIES-39.

Your installation has this PGDB built-in. To make an editable copy of it:

- Select this PGDB
- File → Save PGDB As...
- Type “Test” for New PGDB ID and click outside that box.
- Click OK and wait for the process to complete.
- You will now have two PGDBs for this organism. Open the one with “Source” listed as “User”.

Organisms	Pathways	Genes (ORF %)	Genome Size (bp)	Citations	Source	Version	Registry Download Date
<i>Arabidopsis thaliana</i> col					User (MySQL)	24.0	
<i>Arthrospira platensis</i> NIES-39	228	6,579 (38.4%)	6,788,435	1759	User	25.5	
<i>Arthrospira platensis</i> NIES-39	228	6,579 (38.4%)	6,788,435	1759	Built-In	25.5	
<i>Escherichia coli</i> K-12 substr. MG1655	363	4,735 (10.9%)	4,641,652	41490	Built-In	25.5	
<i>Helicobacter pylori</i> 26695					Registry	25.5	23-Feb-2022 14:49:12
MetaCyc	2,980	14,343 (0.5%)	0	70088	Built-In	25.5	

Copyright Notice

Saving/undoing changes



The user **must** save changes explicitly

- File → Save Current DB
or
- Save DB button on upper right corner
or
- $\wedge s$

“Undo” is called Revert Current DB in PTools lingo. It only works with unsaved changes, and it reverts **all** unsaved changes (no step-by step undo).

Storing databases in MySQL or Oracle enables the following commands:

- List Unsaved Changes in Current DB
- Checkpoint Current DB Updates to File
- Restore Updates from Checkpoint File
- Refresh All Open DBs

Other editing-related DB commands under the File menu

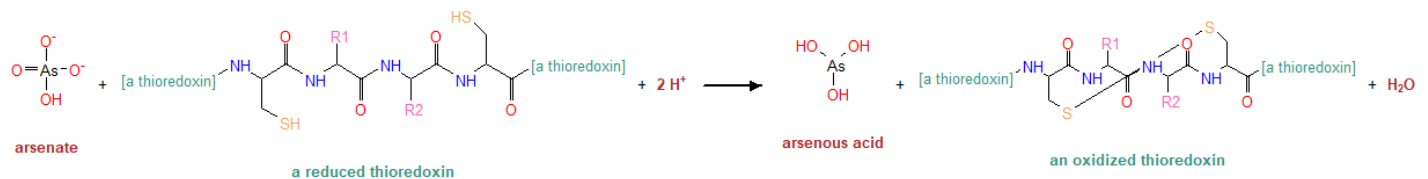
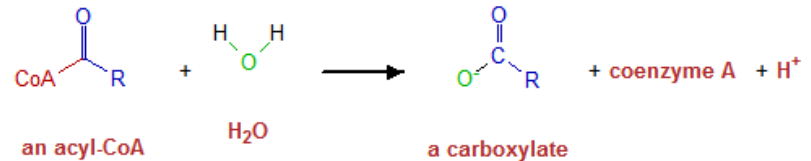
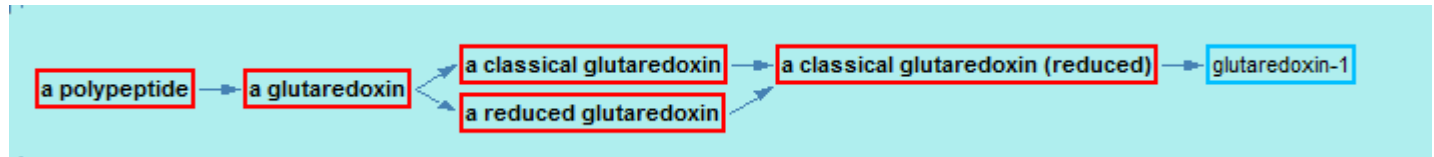
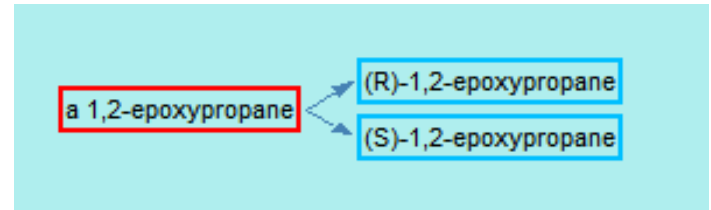
- Create new version for selected DBs (also modifies the default-version file to have the new version opened automatically)
- Save DB as... (makes a new copy that can be opened in the same session as the source PGDB)

- Delete a DB



Classes and instances are like folders and files

- Instances describe specific objects (e.g. L-lysine)
- Classes describe groups of biological objects (e.g. “an amino acid”)
- Classes can contain other objects, while instances can’t
- Every compound with an “R” in its structure should be a class
- Proteins or modified proteins that are substrates of MetaCyc reactions are always classes



What's a frame? And why does it need an ID?



- Every object in the database is a “frame”, and each frame has a unique ID within the database. Instance frame IDs are usually assigned automatically and are not very meaningful.

Examples: CPD-23 PWYQT-7 RXN0-555 MONOMER-387 CPLXI-345.

The **prefix** describes the type of object. Frame names generated in a PGDB other than MetaCyc include **one or two characters** that identify the source database.

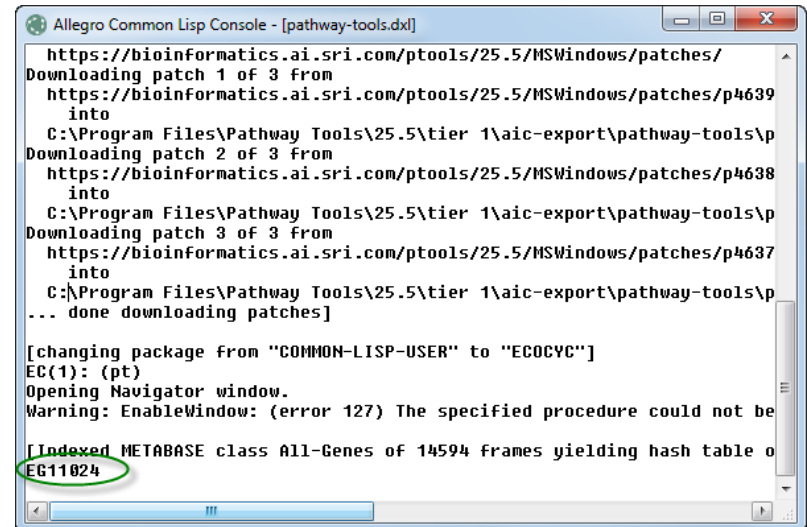
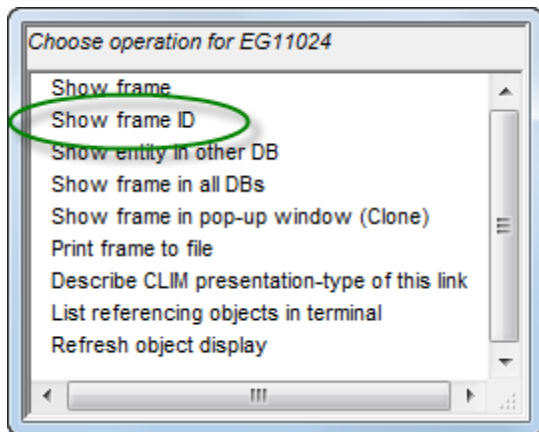
Legacy objects in Pathway Tools (created before current naming standards) usually deviate from these guidelines.

- **Class frame names are usually created by humans and use language.** E.g. Thioglucosides, Amino-Acid-Biosynthesis.
- Object **names** include **common names and synonyms**. They are useful for both humans and computer searches. Unlike frame IDs, names may be not unique.

Printing frame IDs in the Lisp Console

Curation involves working with frame IDs. A convenient way to do it is to print them to the Lisp console and copy them from there.

- Right-click on an object name and select Show → Show frame name
- Move to the Lisp console and copy the name to the clipboard.
- If collecting multiple names, paste it into a text editor



```
Allegro Common Lisp Console - [pathway-tools.dxl]
https://bioinformatics.ai.sri.com/ptools/25.5/MSWindows/patches/
Downloading patch 1 of 3 from
https://bioinformatics.ai.sri.com/ptools/25.5/MSWindows/patches/p4639
into
C:\Program Files\Pathway Tools\25.5\tier 1\aic-export\pathway-tools\p
Downloading patch 2 of 3 from
https://bioinformatics.ai.sri.com/ptools/25.5/MSWindows/patches/p4638
into
C:\Program Files\Pathway Tools\25.5\tier 1\aic-export\pathway-tools\p
Downloading patch 3 of 3 from
https://bioinformatics.ai.sri.com/ptools/25.5/MSWindows/patches/p4637
into
C:\Program Files\Pathway Tools\25.5\tier 1\aic-export\pathway-tools\p
... done downloading patches]

[changing package from "COMMON-LISP-USER" to "ECOCYC"]
EC(1): (pt)
Opening Navigator window.
Warning: EnableWindow: (error 127) The specified procedure could not be
[Index]
[Index] METABASE class All-Genes of 14594 frames yielding hash table o
EG11024
```

The screenshot shows the output of the Lisp console. The text "EG11024" at the bottom is circled in green, indicating the frame ID that was printed.

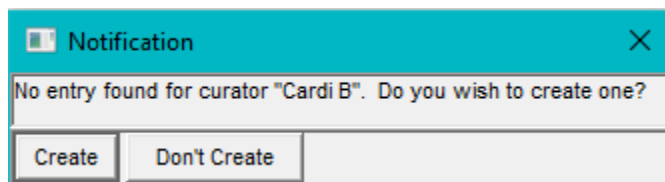
The author credit system



Author credit gives you credit for your work.

Associating objects with their author makes it easy to find the objects you worked on.

Author credit is not properly stored unless a curator frame has been created and configured properly.



Helicobacter pylori 26695 Curator: Ron Caspi

Email: caspi@ai.sri.com

Affiliations: SRI International

Note: Listed below are contributions the curator has made to HpyCyc. They are sorted, with the most recent at the top.

Stats: Pathways: 48, Proteins: 536, RNAs: 59, Reactions: 134, Compounds: 1584, Misc.: 19

Pathways

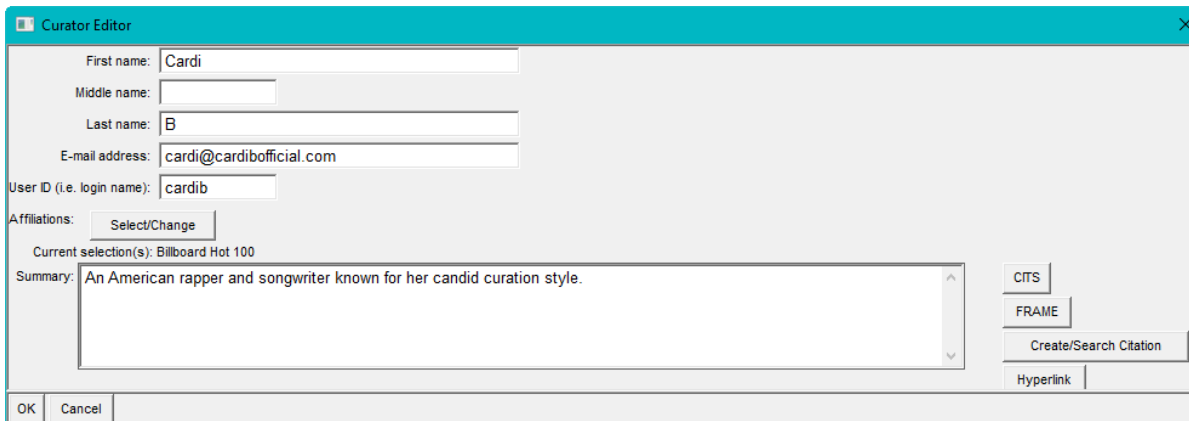
Shown below are the following credit types: Created: 25, Revised: 23

Created:

ATP biosynthesis on 19-Dec-2017,
formaldehyde oxidation VII (THF pathway) on 19-Jun-2017,
lipoprotein posttranslational modification on 10-Apr-2017,
(5Z)-dodec-5-enoate biosynthesis II on 24-Jan-2017,
biosynthesis of Lewis epitopes (H. pylori) on 21-Sep-2016,
L-cysteine biosynthesis VI (from L-methionine) on 30-Mar-2016,
crotonate fermentation (to acetate and cyclohexane carboxylate) on 02-Nov-2013,
superpathway of demethylmenaquinol-6 biosynthesis II on 17-Sep-2013,
demethylmenaquinol-6 biosynthesis II on 17-Sep-2013,
superpathway of adenosine nucleotides *de novo* biosynthesis I on 18-Feb-2013,
superpathway of guanosine nucleotides *de novo* biosynthesis I on 15-Feb-2013,
superpathway of pyrimidine deoxyribonucleotides *de novo* biosynthesis on 31-Jan-2013,
pyrimidine deoxyribonucleotides *de novo* biosynthesis I on 16-Jan-2013,
2'-deoxy- α -D-ribose 1-phosphate degradation on 08-Jan-2013,
purine deoxyribonucleosides degradation I on 08-Jan-2013,
hydroxymethylpyrimidine salvage on 23-Sep-2011,
thiamine diphosphate salvage II on 14-Sep-2011,
tetrahydrofolate biosynthesis on 22-Sep-2010,
autoinducer AI-2 biosynthesis I on 09-Mar-2009,
CMP-pseudamine biosynthesis on 25-Feb-2009,
superpathway of adenosine nucleotides *de novo* biosynthesis II on 13-Jan-2009,
superpathway of guanosine nucleotides *de novo* biosynthesis II on 13-Jan-2009,
acyl carrier protein activation on 13-Aug-2008,
UMP biosynthesis I on 24-Oct-2007,
GDP-mannose biosynthesis on 13-Sep-2007

Creating curator frame – part 1

- Start by checking if there already is a frame for your organization (Tools → Search → Organizations)
- If there isn't one, create one (File → Create → Organization).
- Next, create a frame for yourself (File → Create → Curator)
- User ID must be longer than 3 characters

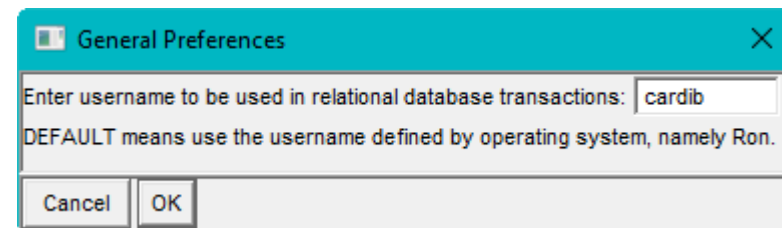
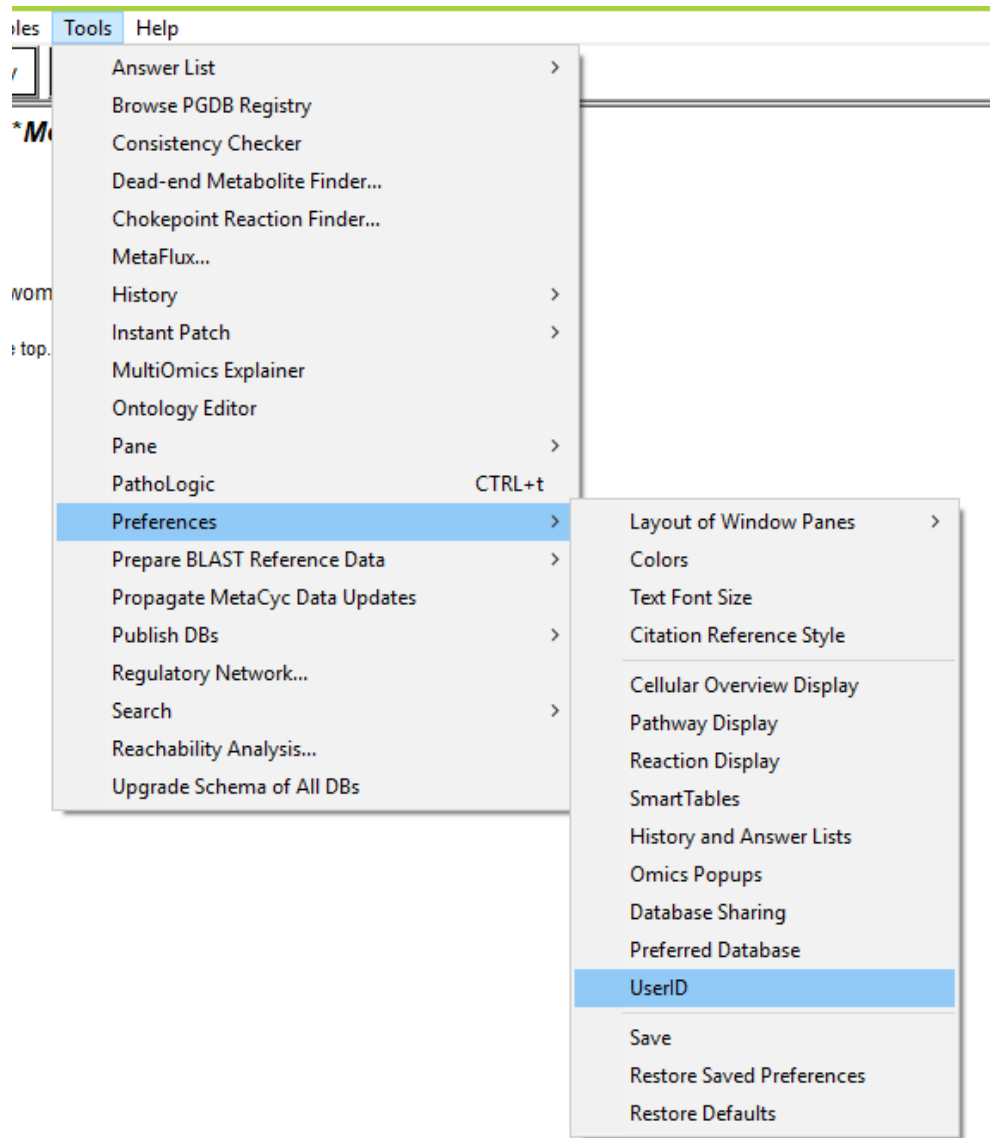


The screenshot shows the 'Curator Editor' window with the following fields and options:

- First name: Cardi
- Middle name: (empty)
- Last name: B
- E-mail address: cardi@cardibofficial.com
- User ID (i.e. login name): cardib
- Affiliations: Select/Change
- Current selection(s): Billboard Hot 100
- Summary: An American rapper and songwriter known for her candid curation style.
- Buttons: CITS, FRAME, Create/Search Citation, Hyperlink
- Bottom buttons: OK, Cancel



Creating curator frame – part 2



What sort of objects are usually edited?

The following are just a few examples

- metabolites
- reactions
- pathways
- enzymes: (creating complexes, assigning to correct reactions)
- regulatory information
- transcription units

The editors

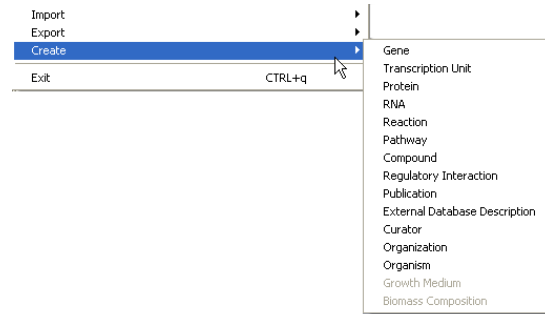
- PGDB Info Editor
- Compound Editor and Compound Structure Editors
- Reaction Editor
- Pathway Editor, Pathway Info Editor
- Signaling Pathway Editor
- Protein Editor and Protein Subunit Structure Editor
- Synonym Editor
- Publication Editor
- Curator/Organization Editors
- Gene Editor
- Isoform/Coding-Segment Editor
- RNA Editor
- Transcription Unit Editor
- Regulatory Interaction Editor
- External Database Editor
- Organism Editor
- Frame Editor
- Ontology Editor



Invoking the Editors

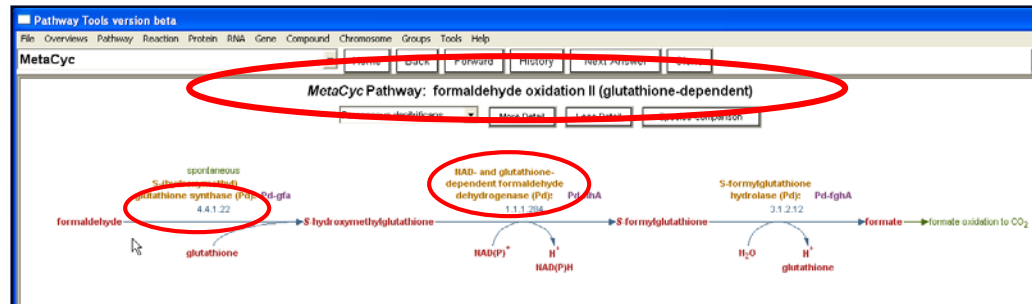
Creating a new Object:

Use the **New** command under certain top menus, or the **Create** command under the File menu

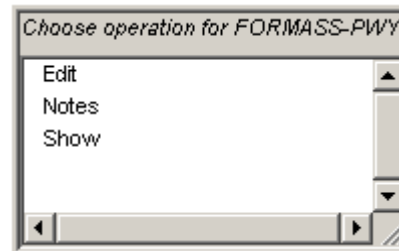


Editing an existing Object:

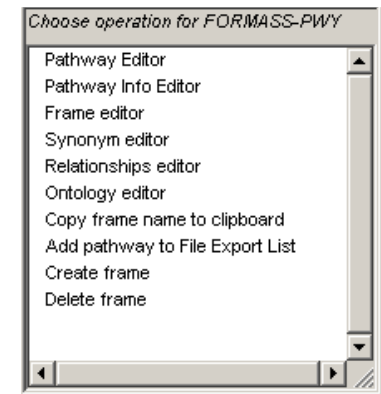
Right-Click on the **any clickable name**, select Edit, then the appropriate editor



Right-click



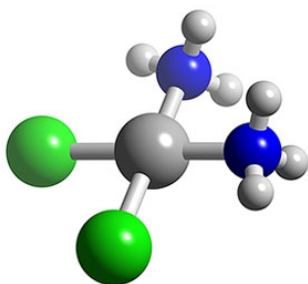
left-click on "Edit"



COMPOUNDS

The compound editor

- Create or edit a compound (but not its structure)
- Specify Class
- Common Name and Synonyms
- Comments, citations
- Links to other DBs



Edit FRUCTOSE-16-DIPHOSPHATE

Class: all carbohydrates->a carbohydrate->a glycan->a carbohydrate derivative->a sugar phosphate->a hexose phosphate->a hexose 6-phosphate, all carbohydrates->a carbohydrate->a glycan->a carbohydrate derivative->a sugar phosphate->a sugar bisphosphate

Common Name:

Synonyms:

<input type="text" value="fructose 1,6-bisphosphate"/>	<input type="button" value="Make this the Common Name"/>
<input type="text" value="fructose 1,6-bisphosphate"/>	<input type="button" value="Make this the Common Name"/>
<input type="text" value="D-fructos 1,6-bisphosphate"/>	<input type="button" value="Make this the Common Name"/>
<input type="text" value="FBP"/>	<input type="button" value="Make this the Common Name"/>
<input type="text"/>	
<input type="text"/>	
<input type="text"/>	

Abbreviated name:

Systematic name:

N-1 name: N-1 name: N name:

Citations:

Summary:

Links to other databases:

Database	ID	Relationship
IAF1260	fdp	Same Entity
HMDB	HMDB01056	Same Entity
ChemSpider	4574223	Same Entity
PubChem-compound	5460765	Same Entity
ChEBI	32966	Same Entity
Wikipedia	Fructose_1,	Same Entity
KEGG LIGAND	C00354	Same Entity
CAS	488-69-7	Same Entity
----		Same Entity

Does this compound have no plausible structure?

Credits:

Date: none yet

Curators:

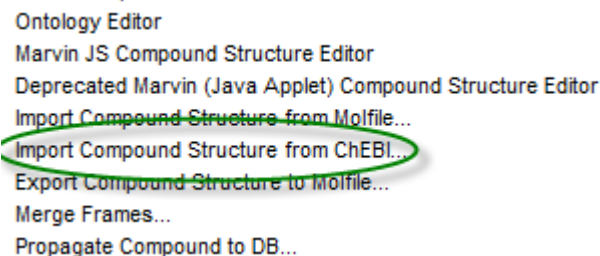
Organizations:

Current selection(s):

OK Cancel

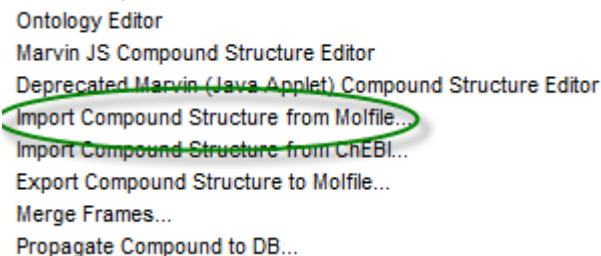
Entering compound structures

- Option 1: if the compound is present in ChEBI (<https://www.ebi.ac.uk/chebi/>) and you enter the link, you can import the structure directly from ChEBI



Ontology Editor
Marvin JS Compound Structure Editor
Deprecated Marvin (Java Applet) Compound Structure Editor
~~Import Compound Structure from Molfile...~~
Import Compound Structure from ChEBI...
Export Compound Structure to Molfile...
Merge Frames...
Propagate Compound to DB...

- Option 2: install a freeware desktop program such as ChemSketch, save structure as Mol file, and import into PTools



Ontology Editor
Marvin JS Compound Structure Editor
Deprecated Marvin (Java Applet) Compound Structure Editor
Import Compound Structure from Molfile...
~~Import Compound Structure from ChEBI...~~
Export Compound Structure to Molfile...
Merge Frames...
Propagate Compound to DB...

<https://www.acdlabs.com/resources/freeware/chemsketch/>

The Marvin structure editor

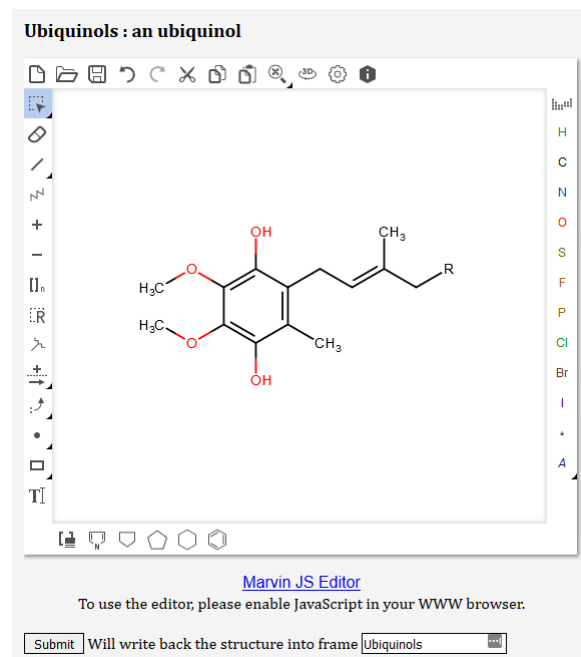
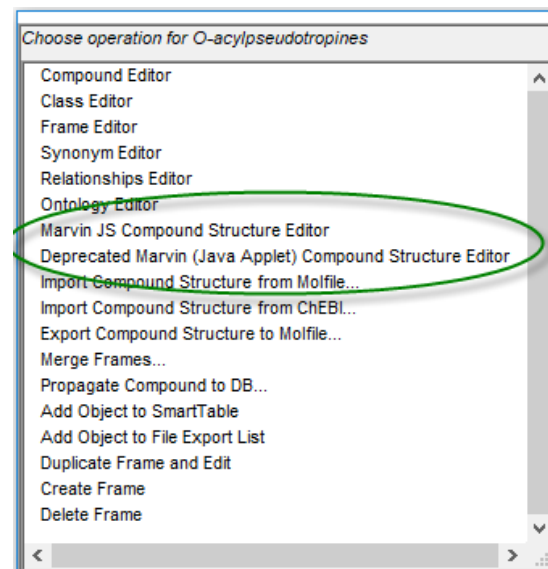
- Pathway Tools supports the **Marvin JS structure editor**, produced by ChemAxon, which needs to be obtained from them.

<https://chemaxon.com/products/marvin-js>

<http://bioinformatics.ai.sri.com/ptools/installation-guide/released/marvin-js.html>

Other compound-related functions

- Exporting to other DBs
- Duplicate Frame and Edit



The PGDB info editor

To access: go to the PGDB home page (File-> Summarize Current Database or Organism Set)

Right-click the organism name, and choose Edit → PGDB Info Editor

This is the place to:

- Create a comment for the PGDB home page
- Specify PGDB authors
- Modify NCBI taxonomy
- Specify a footer
- Set the tier level
- Enter **MIGS** Data
- Enter **Annotation** Data

The screenshot shows the PGDB Info Editor window with the following fields and content:

- PGDB Info Editor** (Title Bar)
- PGDB Info** | **MIGS Data** | **Annotation Data** (Tabs)
- Synonyms:** Bacillus anthracis AmesGenbank entry NC_00399, Bacillus anthracis AmesGenbank entry NC_00399, Bacillus anthracis strain Ames, Bacillus anthracis Ames, Bacillus anthracis str. Ames
- Taxon:** Bacillus anthracis Ames | **NCBI Taxonomy Browser** | Current taxonomic lineage: cellular organisms -> Bacteria -> Firmicutes -> Bacilli -> Bacillales -> Bacillaceae -> Bacillus -> Bacillus anthracis -> Bacillus anthracis Ames
- Citations:** 12721629
- Summary:** The primary data source for this dataset is the full genome sequence of [FRAME: TAX:198094] [CITS: [12721629]], derived from Genbank accession NC_003997. This dataset was created using the PathoLogic component of the Pathway Tools program [CITS: [10370234]].
- PGDB Tier:** 2
- Genome Source:** Genbank entry NC_003997
- PGDB Authors:** J. Bashkin, Jonathan Wagg, Nan Guo, Peter Karp, Ron Caspi
- Project Home Page URL:**
- Project Primary Contact Email:**
- Copyright string:** Copyright 2004-2016 SRI International.
- Footer citation for web pages:**
- OK** | **Cancel** (Buttons)

The synonym editor



Lets you easily edit the synonyms and set the common name

The more synonyms, the more likely users are to find an object

The image is a screenshot of a software dialog box titled "Edit Synonyms for EG11028". The dialog box has a light blue header bar with a close button (X) in the top right corner. Below the header, there is a text input field labeled "Common Name:" containing the text "trpE". Underneath this is a section labeled "Synonyms:" which contains a list of text input fields. The first three fields contain the text "anth", "tryD", and "tryp-4". To the right of each of these three fields is a button labeled "Make this the Common Name". Below the "Synonyms:" section is another text input field labeled "Abbreviated Name:". At the bottom of the dialog box, there are two buttons: "OK" and "Cancel".

Correct assignment of reactions to enzymes

After running PathoLogic, some enzymes will be assigned to reactions incorrectly. A curator needs to remove incorrect assignments and attach correct ones.

Example: gene NIES39_RS19045 (*ubiE*) was curated by RefSeq as bifunctional demethylmenaquinone methyltransferase/2-methoxy-6-polyprenyl-1,4-benzoquinol methylase UbiE, and thus assigned multiple reactions of EC 2.1.1.163, demethylmenaquinone methyltransferase.

Since cyanobacteria produce plastoquinone and not menaquinone, this can't be true.

Click on the UniProt link and you will find that in fact this is MenG and should be assigned as EC 2.1.1.329 (a BLAST search against any cyanobacterium would have produced this result as well). This reaction is part of the phylloquinol biosynthesis pathway.

UbiE or MenG?

UbiE/MenG has been assigned 5 incorrect reactions. Each one is involved in an incorrectly-predicted pathway. We will

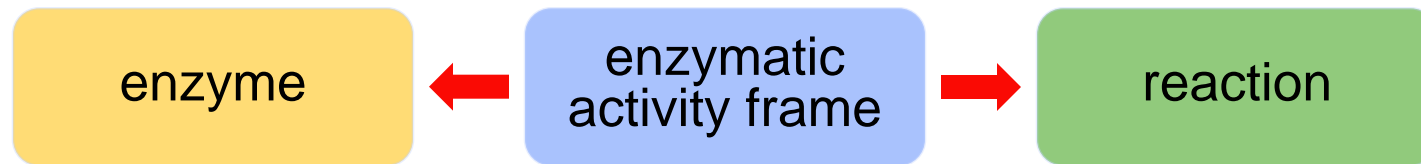
1. Change the gene name to menG
2. Change the enzyme name to demethylphyloquinol methyltransferase
3. Delete the incorrect pathways
4. Delete the incorrect reactions
5. Attach the enzyme to the correct reaction

Other things to do here (but we won't today)

1. Import the plastoquinone biosynthesis pathway from MetaCyc
2. Find enzymes known to participate in that pathway and attach them to the appropriate reactions

How enzymatic activities are handled

Each enzymatic activity is defined by a new database object that points to an enzyme and a reaction



The enzymatic activity object is a frame, like everything else in the database. It's frame ID is in the form ENZRXXN-XXX

Enzymatic activities also have a common name and potential synonyms (e.g. 2-oxoglutarate synthase)

Deleting incorrect enzymatic activities

Right-click the enzymatic activity name, and select Edit -> Delete

Enzymatic reaction of: acetyl-CoA:ACP transacylase (beta-ketoacyl-acyl carrier protein synthase III)

Synonyms: acetyl-CoA:[ACP] S-acetyltransferase; [acyl-carrier-protein] S-acetyltransferase

acyl-carrier protein + acetyl-CoA

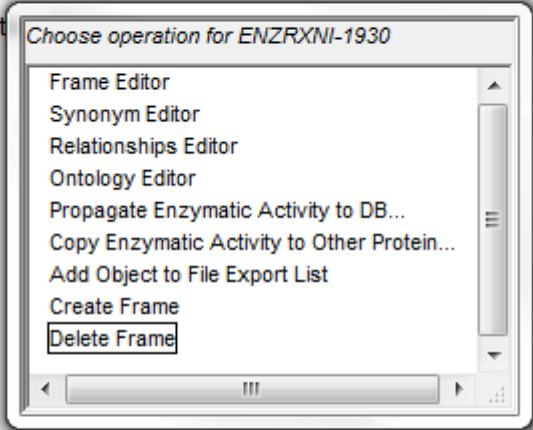
↓ 2.3.1.38

an acetyl-[acp] + coenzyme A

The reaction is favored in the direction shown.

In Pathways: superpathway of fatty acid biosynthesis initiation (E. coli)

Created by: caspi on 17-Sep-2013

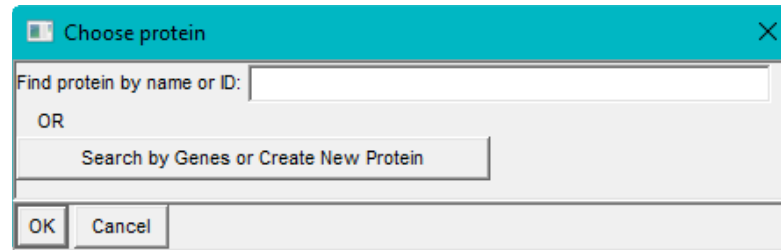


The screenshot shows a context menu titled "Choose operation for ENZRXNI-1930" with the following options: Frame Editor, Synonym Editor, Relationships Editor, Ontology Editor, Propagate Enzymatic Activity to DB..., Copy Enzymatic Activity to Other Protein..., Add Object to File Export List, Create Frame, and Delete Frame. The "Delete Frame" option is highlighted with a black border.

Attaching a reaction to an enzyme

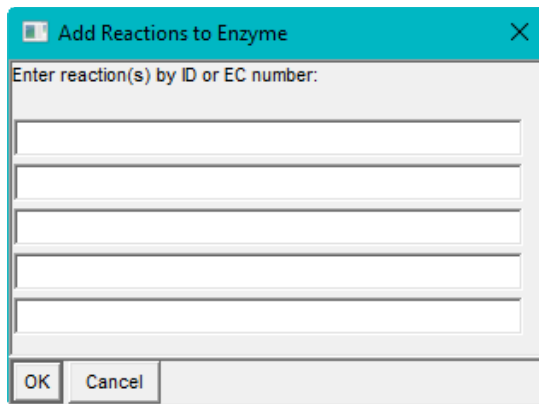
To add an enzyme to a reaction: First copy the frame ID of the enzyme, then

- Right click the reaction, choose Edit → Create/Add enzyme and paste the ID.



Or

- Copy the frame ID of the reaction, then
- Right click the enzyme, choose Edit → Add Reaction(s) and paste the ID.



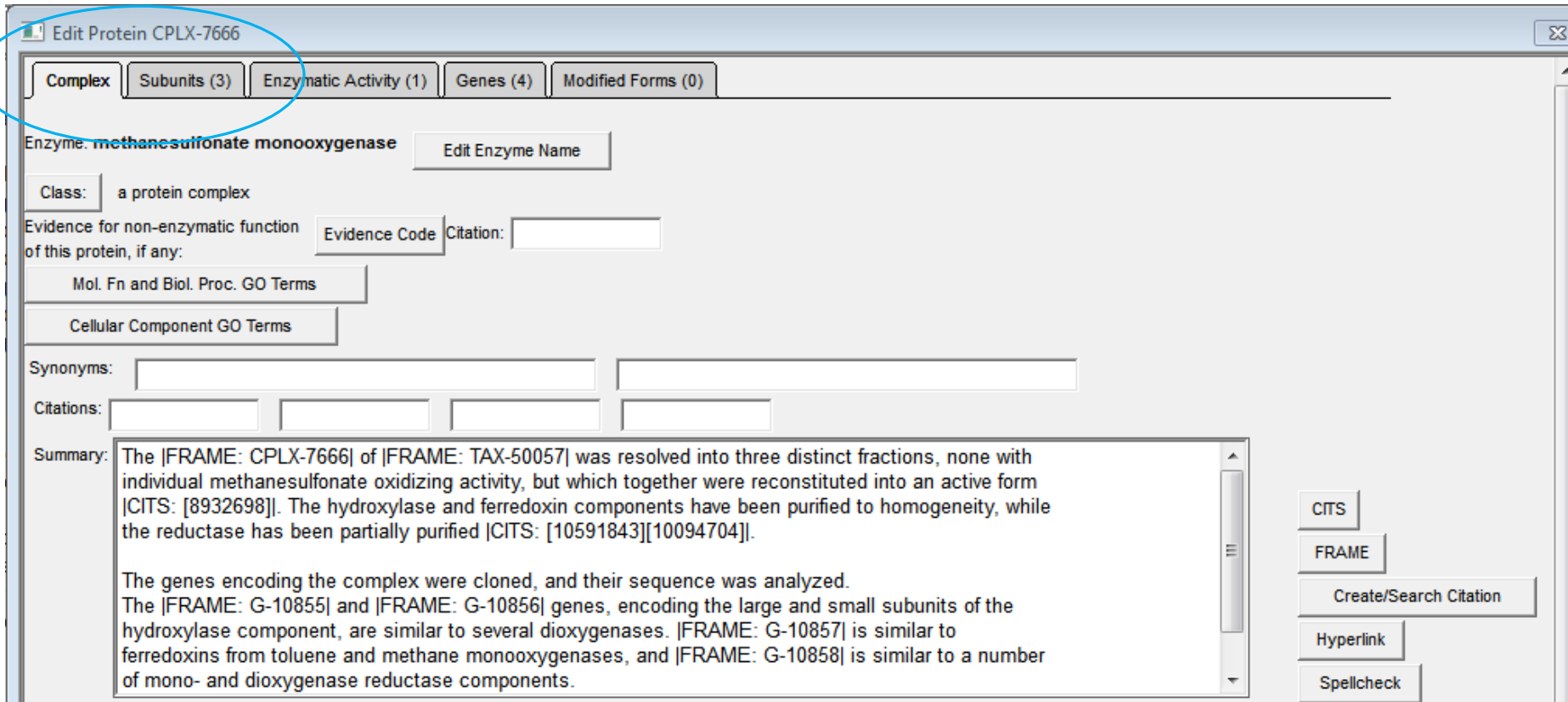
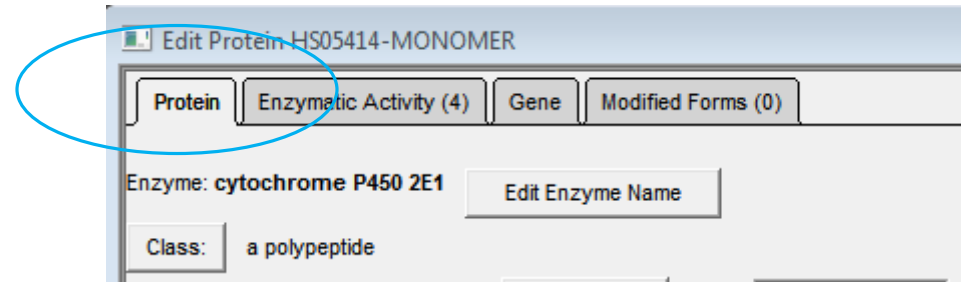
Connecting a reaction with a protein will open the Protein Editor and require you to enter an **enzymatic activity name** in the editor. More about that in the next slides.

PROTEINS

The protein editor

(monomer version)

(protein complex version)



Protein editor – first tab

For an example of a complex, open CPLX2764-6 in the *A. platensis* PGDB

The screenshot displays the 'Edit Protein CPLX2764-6' window. At the top, there are tabs for 'Complex', 'Subunits (3)', 'Enzymatic Activity (1)', 'Genes (3)', and 'Modified Forms (0)'. The 'Complex' tab is active. Below the tabs, the 'Enzyme' field is set to 'urease' with an 'Edit Enzyme Name' button. The 'Class' field is 'a protein complex'. There are two 'Evidence for characterization other than enzymatic activity, if any:' fields, each with an 'EV-COMP-AINF' dropdown, a 'Citation:' label, and an empty text box. Below these are 'Synonyms:' and 'Citations:' fields, each with two empty text boxes. A large 'Summary:' text area is present. On the right side, there are buttons for 'CITS', 'FRAME', 'Create/Search Citation', 'Hyperlink', and 'Import MetaCyc Comment'. At the bottom, there are 'Molecular Weight (kD, experimental):' and 'pt:' fields, each with a 'Citation:' label and an empty text box. Below that is a 'Links to other databases:' section with a 'Database' dropdown, an 'ID' text box, and a 'Relationship' dropdown with a '?' icon. The 'Complex Copy Number:' field is also present. At the very bottom are 'OK' and 'Cancel' buttons.

Protein editor - subunits tab (for complexes)

- Edit the copy number of each subunit
- Specify UniProt ID or links to other databases
- Specify experimental MW and any useful info that may apply (GO terms, features, copy number)

Sections below are for the following subunits:
ureA gene product: urease subunit gamma
NES39_RS13785 gene product: urease subunit beta
ureC gene product: urease subunit alpha

Subunit: ureA gene product

Name: Coefficient:

Synonyms:

Protein Accession:

Citations:

Summary:

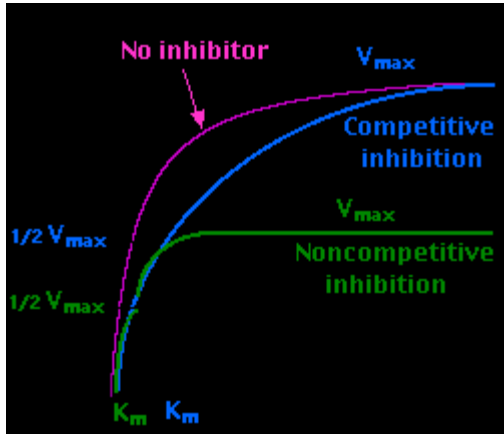
Molecular Weight (kD, experimental): Citation:

is initial methionine cleaved?

Links to other databases:

Database	ID	Relationship
InterPro	IPR012010	In Family
InterPro	IPR002026	In Family
InterPro	IPR036463	In Family
Pfam	PF00547	In Family
String	696747.NIE:	Same Entity
UniProt	D4ZSS5	Same Entity

Protein editor - enzymatic activity tab



- Activators/Inhibitors/Cofactors/Alternative substrates:
- Activator (allosteric)
 - Activator (allosteric)**
 - Activator (nonallosteric)
 - Activator (mechanism unknown or not curated)
 - Inhibitor (competitive)
 - Inhibitor (noncompetitive)
 - Inhibitor (uncompetitive)
 - Inhibitor (mixed)
 - Inhibitor (irreversible)
 - Inhibitor (allosteric)
 - Inhibitor (mechanism unknown or not curated)
 - Inhibitor (other)
 - Cofactor or prosthetic-group
 - Alt. substrate for L-ornithine
 - Alt. substrate for urea
 - Alt. substrate for H2O
 - Alt. substrate for L-arginine



Edit Protein CPLX-6621

Complex | Subunits (1) | **Enzymatic Activity (1)** | Gene | Modified Forms (0)

Add New Activity

Enzyme activity name: arginase

Reaction (shown in EC left-to-right direction): L-arginine + H₂O ↔ urea + L-ornithine

Evidence for this activity: EV-EXP-IDA-PURIFIED-PROTEIN Citation: 2241902 EV-EXP-IDA-PURIFIED-PROTEIN Citation: 2515788

Synonyms:

Citations:

Summary:

Reaction Direction: No Direction Stored Citation: Activity is physiologically relevant?

Reaction Location: cytosol (default) Add another location for this activity

Activators/Inhibitors/Cofactors/Alternative substrates:	Physiologically relevant?	K _i (μM)	Citation(s)
Activator (allosteric) L-ornithine	<input type="checkbox"/>		11370664
Activator (allosteric) Mn2+	<input type="checkbox"/>		2515788
Inhibitor (noncompetitive) L-arginino-succinate	<input checked="" type="checkbox"/>		2241902
Inhibitor (noncompetitive) L-canavanine	<input type="checkbox"/>		2241902
Inhibitor (noncompetitive) D-octopine	<input type="checkbox"/>		2241902
Inhibitor (competitive) L-lysine	<input checked="" type="checkbox"/>		2241902
Inhibitor (competitive) L-homoarginine	<input type="checkbox"/>		2241902
Activator (mechanism unknown or not curated)	<input type="checkbox"/>		
Activator (mechanism unknown or not curated)	<input type="checkbox"/>		

T_{opt} (°C): Citation:

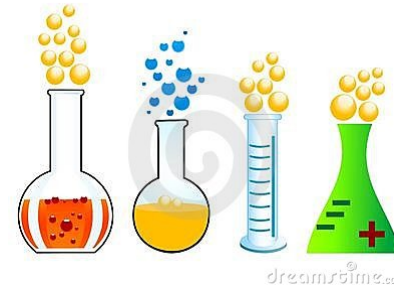
pH_{opt}: 10.5 Citation: 2515788 10.5 Citation: 2515788

Substrate	K _m	Units	Citation	V _{max}	Units	Citation
L-arginine	<input type="text"/>	μM		<input type="text"/>	U/mg (= μmol /mg /min)	
L-ornithine	<input type="text"/>	μM		<input type="text"/>	U/mg (= μmol /mg /min)	
urea	<input type="text"/>	μM		<input type="text"/>	U/mg (= μmol /mg /min)	

OK Cancel

REACTIONS

The reaction editor



With the Reaction Editor you can:

- Enter or edit a reaction equation
- Specify EC numbers (official?)
- Enter a common name (if no full EC number exists)
- Set Conversion Type
- Specify location information (transport, cellular location)
- Specify reaction direction



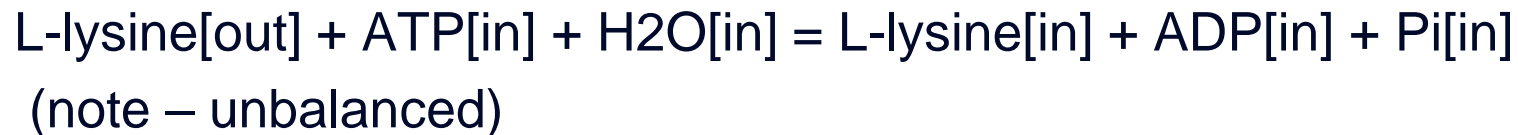
Reaction editor examples

Regular chemical reaction:



Transport reaction:

first, set reaction type to transport



balance, then add reaction locations

Duplication is bad



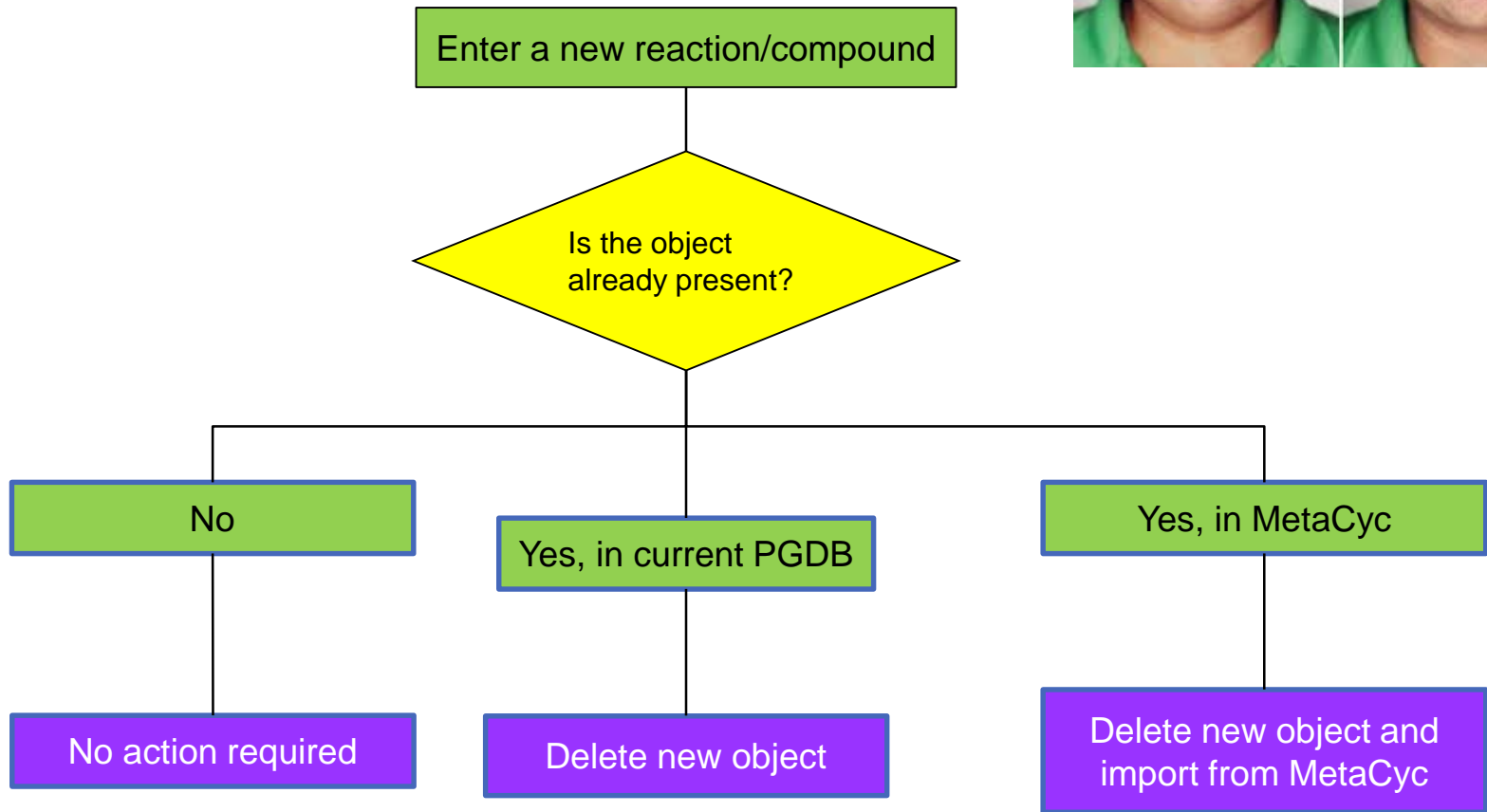
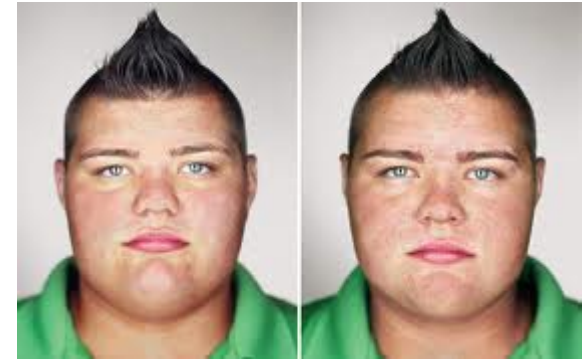
Avoid Duplication! Reuse information whenever possible

- A PGDB should not describe the same biological or chemical entity more than once
- You should not recreate an entry already present in MetaCyc
- Some tools help prevent the inadvertent creation of duplicate compounds and reactions



Duplicates can be cute, but should be avoided

The duplicate checker



Compound duplicate checker - import

If the compound duplicate checker catches a duplicated compound upon its creation, you can delete the new compound and import the MetaCyc compound instead.

Resolve Compound Duplicates

The compound that you are creating or editing (CPD30-4157) may be a duplicate of an existing compound. Duplicating existing compounds can be very problematic for producing accurate comparative analyses, so we strongly recommend that whenever possible you reuse existing compounds in the current PGDB. Below are a list of Compounds from current kb (YEASTCYC) and MetaCyc (if any) that may be potential duplicates of the compound you entered. Please select:

Show -- To view the compound in Pathway Tools navigator
Delete -- To delete the compound you are editing
(This option may not be available if the compound is involved in reactions or is an effector of enzymes)
Merge -- To merge compounds, select a compound from the drop down list and click on "Merge" to merge it with compound from that row
Replace entered compound with Y and exit -- To replace the entered compound with Y and exit this window (the entered compound will be deleted and the selected compound will be displayed in the Pathway Tools navigator)

The below two options are only available if the program finds duplicate matches in MetaCyc for the entered compound:
Import X from MetaCyc deleting the entered compound and exit -- To import compound from MetaCyc deleting the entered compound and exiting this window
Import X from MetaCyc without deleting the entered compound -- To import the compound from MetaCyc without deleting the entered compound

ENTERED COMPOUND:

achilleol B (CPD30-4157)

Molecular Weight: 426.724

Synonyms:
achilleol B

POSSIBLE METACYC CANDIDATES:

achilleol B (CPD-13880)

Molecular Weight: 426.724

Synonyms:
achilleol B

This compound is involved in 1 reactions.

This compound is an effector of enzyme(s) in the database.

Delete CPD30-4157 Show

Import CPD-13880 from MetaCyc, deleting the entered compound, and exit Show

Import CPD-13880 from MetaCyc, without deleting the entered compound

Compound duplicate checker - merge

If the compound duplicate checker catches an existing duplication, you can merge the two compounds, keeping the MetaCyc Frame ID.

Resolve Compound Duplicates

The compound that you are creating or editing (CPD-13880) may be a duplicate of an existing compound. Duplicating existing compounds can be very problematic for producing accurate comparative analyses, so we strongly recommend that whenever possible you reuse existing compounds in the current PGDB. Below are a list of Compounds from current kb (YEASTCYC) and MetaCyc (if any) that may be potential duplicates of the compound you entered. Please select:

Show -- To view the compound in Pathway Tools navigator
Delete -- To delete the compound you are editing
(This option may not be available if the compound is involved in reactions or is an effector of enzymes)
Merge -- To merge compounds, select a compound from the drop down list and click on "Merge" to merge it with compound from that row
Replace entered compound with Y and exit -- To replace the entered compound with Y and exit this window (the entered compound will be deleted and the selected compound will be displayed in the Pathway Tools navigator)

The below two options are only available if the program finds duplicate matches in MetaCyc for the entered compound:
Import X from MetaCyc deleting the entered compound and exit -- To import compound from MetaCyc deleting the entered compound and exiting this window
Import X from MetaCyc without deleting the entered compound -- To import the compound from MetaCyc without deleting the entered compound

ENTERED COMPOUND:

achilleol B (CPD-13880)
Molecular Weight: 426.724
Synonyms:
achilleol B

POSSIBLE YEASTCYC CANDIDATES:

achilleol B (CPD30-4157)
Molecular Weight: 426.724
Synonyms:
achilleol B

Merge CPD30-4157 with:

Replace entered compound with CPD30-4157 and exit

Delete CPD-13880

Delete CPD30-4157

Show

Merge

Show

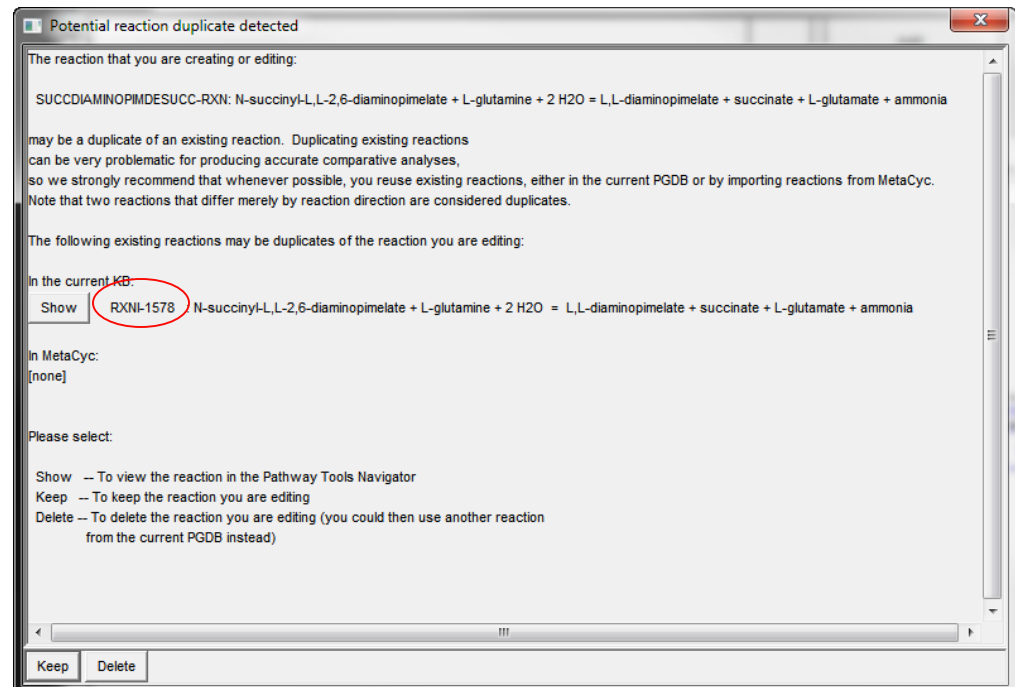
Reuse existing reactions instead of creating duplicates!

If the reaction is already present in your PGDB, you will see a window like this one



You should choose the option “Delete”.

If you want to use the reaction in a pathway, press “show”, then copy the frame ID of the existing reaction before you press Delete and close this window, so you could use it later when specifying the pathway.

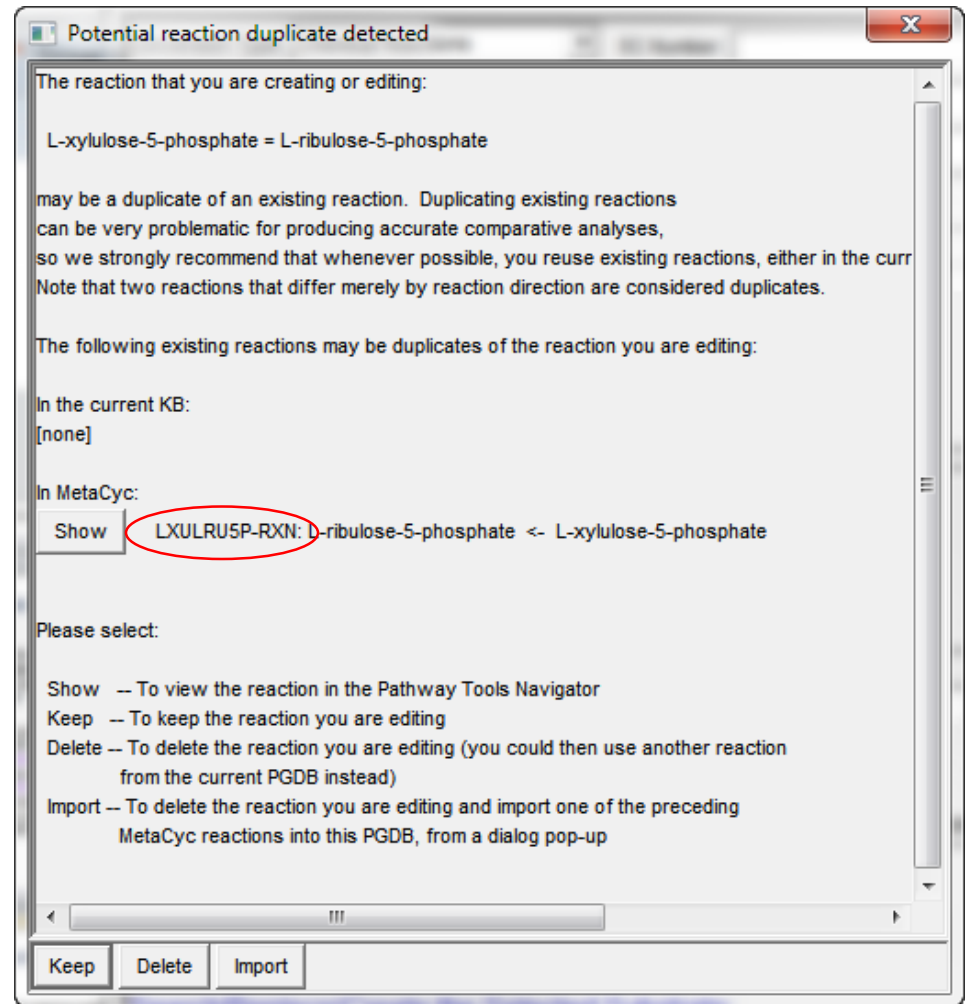


Import MetaCyc reactions instead of creating duplicates!

If the reaction is already present in MetaCyc (but not the current PGDB), you will see a window like this one

You should choose the option “**Import**”.

If you want to use the reaction in a pathway, copy the frame ID of the existing reaction **before** you press the Import button, so you could use it later when specifying the pathway.



PATHWAYS

The pathway info editor

- Class (variant class)
- Common Name
- Synonyms
- Evidence Codes
- Citations
- Summary
- External Links
- Hypothetical reactions
- Key reaction
- Rate-limiting steps
- Enzymes not in use
- Author credit

Pathway Info Editor for GLYCOLYSIS

Class: Generation of Precursor Metabolites and Energy->Glycolysis **This class is a variant pathway class.**

Common Name: glycolysis I

Synonyms:

Embden-Meyerhof pathway	Make this the Common Name
glucose degradation	Make this the Common Name
Embden-Meyerhof-Parnas pathway	Make this the Common Name
EMP pathway	Make this the Common Name
glycolysis (plasticid)	Make this the Common Name

Evidence for Pathway Existence: EV-EXP-TAS Citation: EcoSal Evidence Code Citation:

Citations:

Summary:

ii) it is an amphibolic pathway (pathway that involves both catabolism and anabolism) because it can reversibly produce hexoses from various low-molecular weight molecules.

Because various degradation pathways feed into glycolysis at many different points, glycolysis or portions of it run in the forward or reverse direction, depending on the carbon source being utilized, in order to satisfy the cell's need for precursor metabolites. This switching of direction is possible because all but two of the enzymatic reactions comprising glycolysis are reversible, and the conversions catalyzed by the two exceptions are rendered functionally reversible by other enzymes (IFRAME:F16B-CPLX) and (IFRAME:PEPSYNTH-CPLX) that catalyze different irreversible reactions flowing in the opposite direction.

About This Pathway

Glucose is not shown here as a component of glycolysis because when used by <i>E. coli</i> as a source of carbon and energy, glucose enters the cell via a phosphotransferase system (transport of glucose, IFRAME:CPLX-157), the first intracellular species, therefore, being glucose-6-phosphate.

<i>E. coli</i> does constitutively produce IFRAME:GLUCOKIN-MONOMER (the intracellular enzyme that converts glucose to glucose-6-phosphate) but it is not needed for the utilization of either exogenous or endogenous glucose [CITS: [9023215]]. Under anabolic stress conditions, it may be required to supplement levels of glucose 6-phosphate [CITS: [7786044]].

Links to other databases:

Database	ID	Relationship
		Same Entity

Check box if this is an engineered pathway

Hypothetical Reactions: Select/Change Key Reactions: Select/Change

Rate-Limiting Steps: Select/Change

Enzymes Not Used in this Pathway: *alkaline phosphatase* Change

Credits:

Date	Curators	Organizations
Revised: 26-Jan-2007	Select/Change Create	Select/Change Create
	Current selection(s): Ingraham JL	Current selection(s): UC Davis
	Select/Change Create	Select/Change Create
	Current selection(s):	Current selection(s):

Update Last-Curated Date ?

OK Cancel

Evidence codes for pathways

<http://bioinformatics.ai.sri.com/evidence-ontology/>

Experimental evidence codes:

IDA: inferred from direct assay

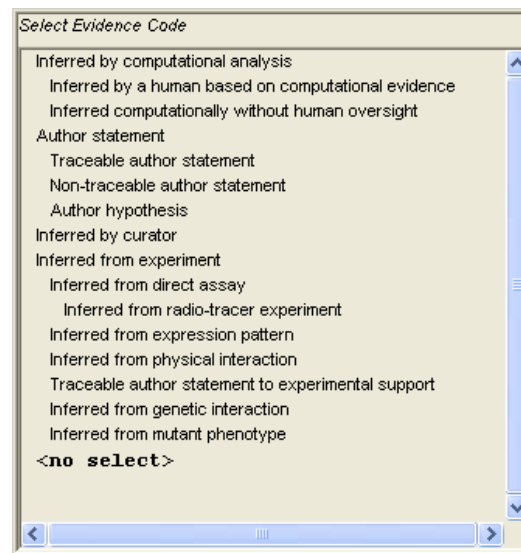
IEP: inferred from expression pattern

IPI: inferred from physical interaction

TAS: traceable author statement

IGI: inferred from genetic interaction

IMP: inferred from mutant phenotype



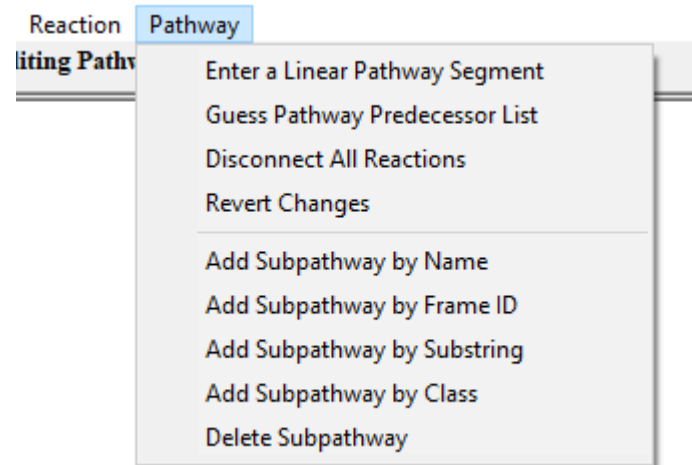
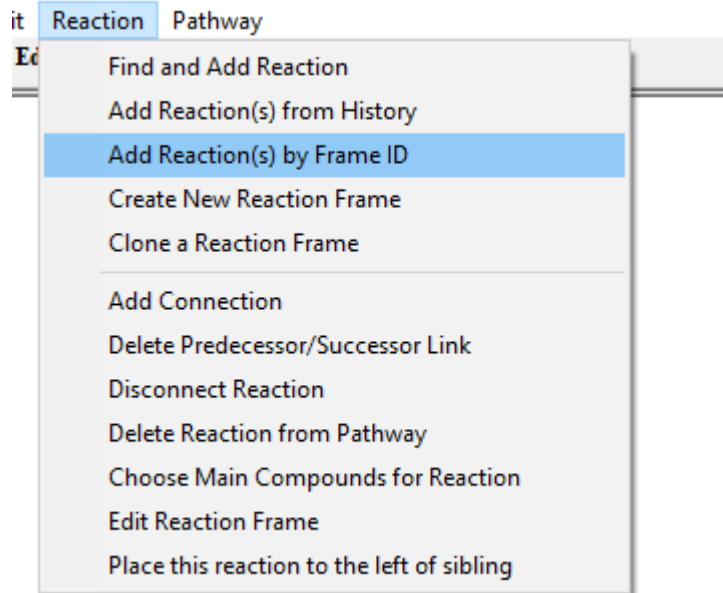
Full documentation for an evidence code is displayed in the Navigator (click the code icon)

The pathway editor

Graphically create and modify pathways

Reaction Menu: add reactions one by one

Pathway Menu: add sub-pathways to create a superpathway



Connecting reactions

Pathway Editor

Exit Reaction Pathway

Editing Pathway L-valine biosynthesis

BRANCHED-CHAINAMINOTRANSFERAL-RXN
2.6.1.42
L-valine + 2-oxoglutarate <--> L-glutamate + 3-methyl-2-oxobutanoate

DIHYDROXYISOVALDEHYDRAT-RXN
4.2.1.9
(2R)-2,3-dihydroxy-3-methylbutanoate
-> 3-methyl-2-oxobutanoate + H2O

2 pyruvate $\xrightarrow[\text{H}^+]{\text{CO}_2}$ (S)-2-acetolactate $\xrightarrow[\text{H}^+]{\text{NADP}^+}$ (2R)-2,3-dihydroxy-3-methylbutanoate

Additional Functionality can be accessed by right-clicking on compounds or reaction lines

- Add Connection
- Delete Predecessor/Successor Link
- Disconnect Reaction
- Delete Reaction from Pathway
- Choose Main Compounds for Reaction
- Edit Reaction Frame
- Place this reaction to the left of sibling

(example: L-valine biosynthesis)



Pathway editor limitations

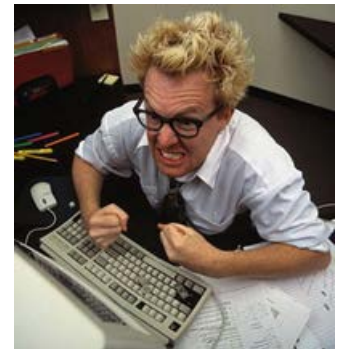
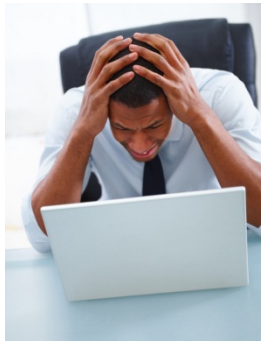
Complex situations can cause ambiguity:

- reaction directionality not specified
- reaction directionality opposite to direction in pathway
- dialog box for disambiguating
- pathway drawn in bizarre arrangement

- **Fix:**

- try disconnecting reactions, specifying main compounds, and adding them in different order

Limitation: a reaction can appear only once in a pathway.



Homework

Please download “1. Reactions and pathways handout.pdf” from <http://bioinformatics.ai.sri.com/ptools/tutorial/sessions/curation> and complete the exercise before our next Zoom session.

If you run into difficulties or have any questions, contact me at ron.caspi@sri.com. I will be happy to answer any questions.