

Editing Pathway/Genome Databases

By Ron Caspi

ron.caspi@sri.com



This presentation can be found at
[http://bioinformatics.ai.sri.com/ptools/tutorial/sessions/curation/
Curation of genes, enzymes and Pathways/](http://bioinformatics.ai.sri.com/ptools/tutorial/sessions/curation/Curation_of_genes,_enzymes_and_Pathways/)

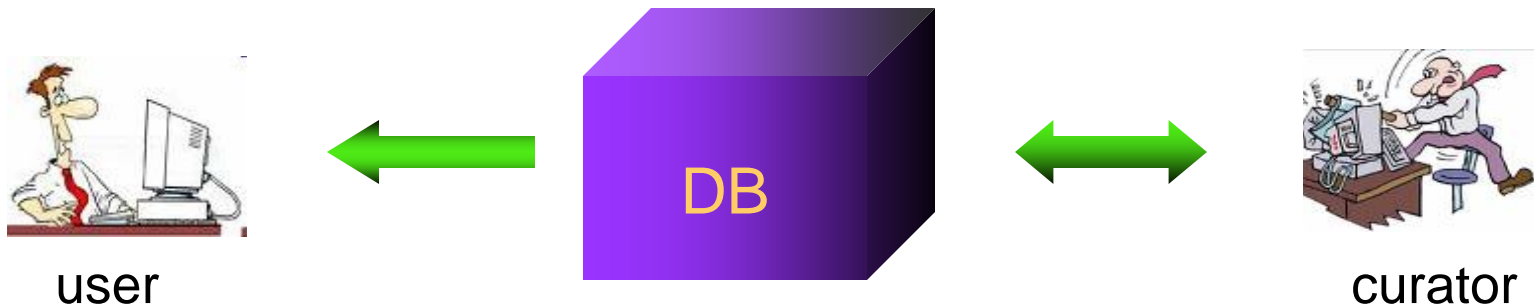
A lot more information is available in the Curator's Guide, at
<https://bioinformatics.ai.sri.com/ptools/curatorsguide.pdf>

Pathway Tools in Editing Mode

The database can be accessed by two distinct modes

- The Navigator allows limited interaction with the DB
- The Editors allow complete modification of the DB

Editing is available in **Desktop mode only**



Installing *H. Pylori* PGDB

In order to be able to perform editing, you must have a PGDB other than MetaCyc or EcoCyc installed on your system.



- In the following exercises we will be using the PGDB for *Helicobacter pylori* 26695.
- To install this PGDB: open the PGDB Registry
(Tools → browse PGDB Registry)
- Type pylori and hit Enter
- Double-click on *Helicobacter Pylori* 26695 to move it to the bottom field, then click on “Fetch and install selected PGDBs”.
- Click all OK buttons until it is installed, then close the Registry by clicking the Cancel button.

Saving/Undoing Changes



The user must save changes explicitly

- File => **Save Current DB** (Control-S works too)
or
- **Save DB** button on upper right

“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

Other Editing-related DB commands under the File menu

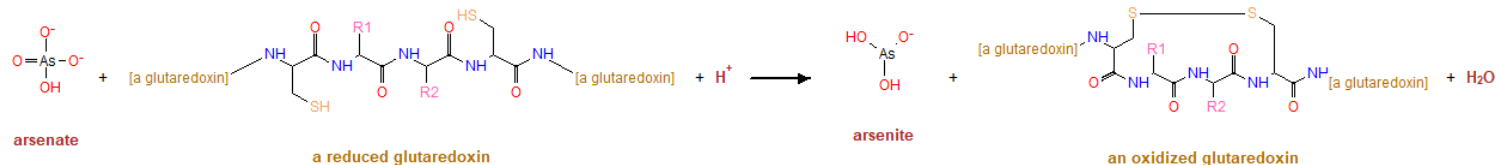
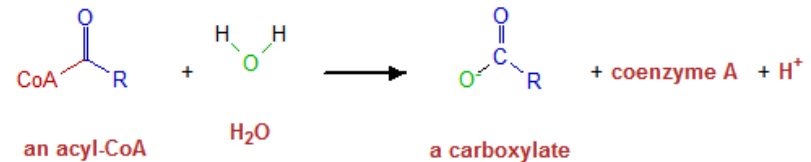
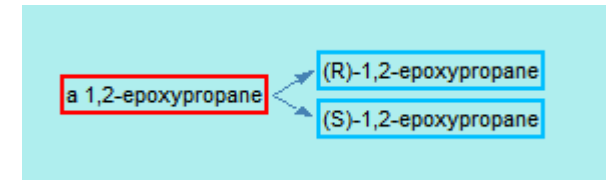
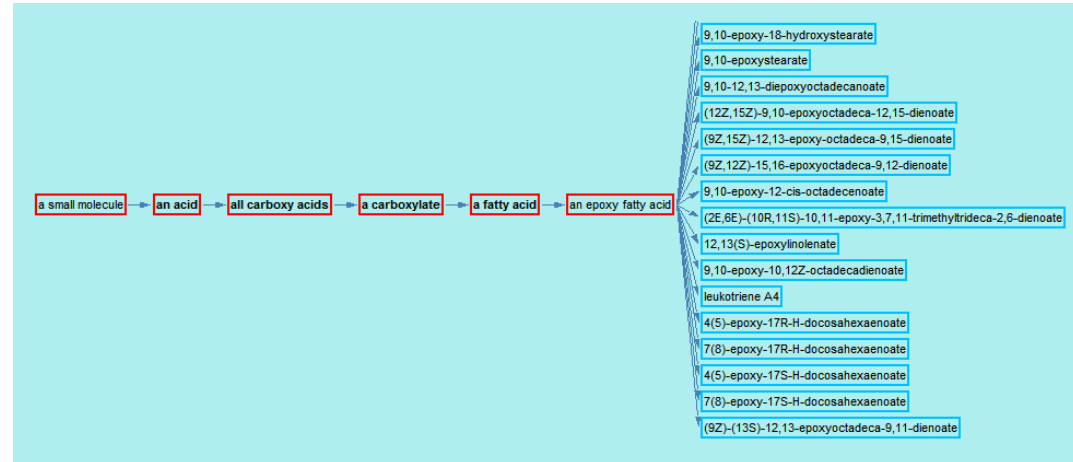
- Create new version for selected DBs (and modifies the default-version file to have the new version opened automatically)
- Save DB as (makes a new copy that can be opened in the same session as the source PGDB)
- Refresh All Open DBs (only MySQL/Oracle DBs)

- Delete a DB



Classes and Instances

- Instance frames describe specific objects (e.g. L-lysine)
- Class frames 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



Object Names and Frame Names



- The **frame name** (also known as frame ID) is a unique ID of the object within the database. Instance frame IDs are usually assigned automatically and are not intended for human consumption.

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

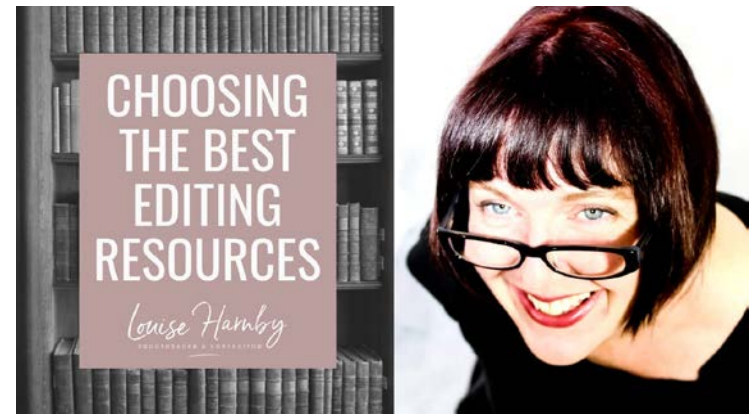
The **prefix** describes the type of object. Frame names generated in a PGDB other than MetaCyc include **one or two characters** that identify the source database.

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.

There are many, many Pathway Tools Editors

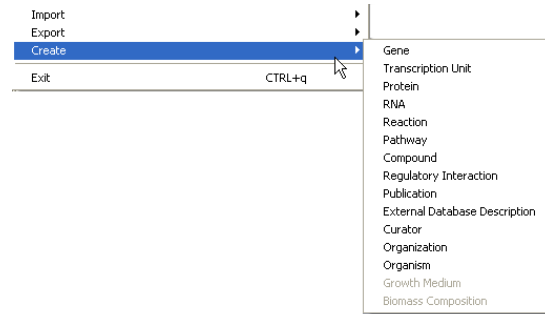
- PGDB Info Editor
- Compound Editor and Compound Structure Editors
- Reaction Editor
- Pathway Editor, Pathway Info Editor
- Signaling Pathway Editor
- Protein Editor and Protein Subunit Structure Editor
- Synonym Editor
- Publication Editor
- Curator/Organization Editors
- Gene Editor
- Isoform/Coding-Segment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



Invoking the Editors

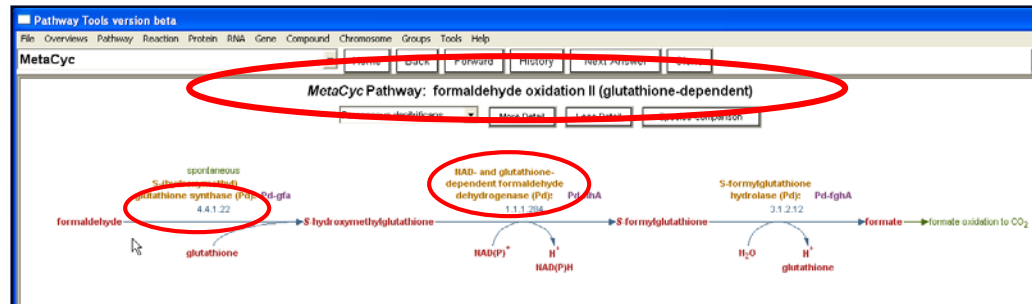
Creating a new Object:

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

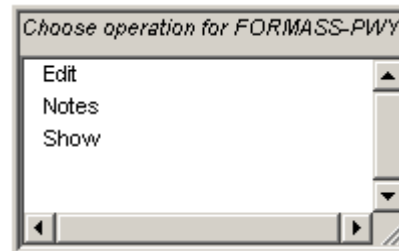


Editing an existing Object:

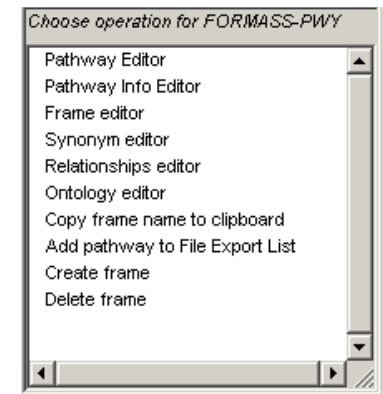
Right-Click on the **any clickable name**, select Edit, then the appropriate editor



Right-click



left-click on "Edit"

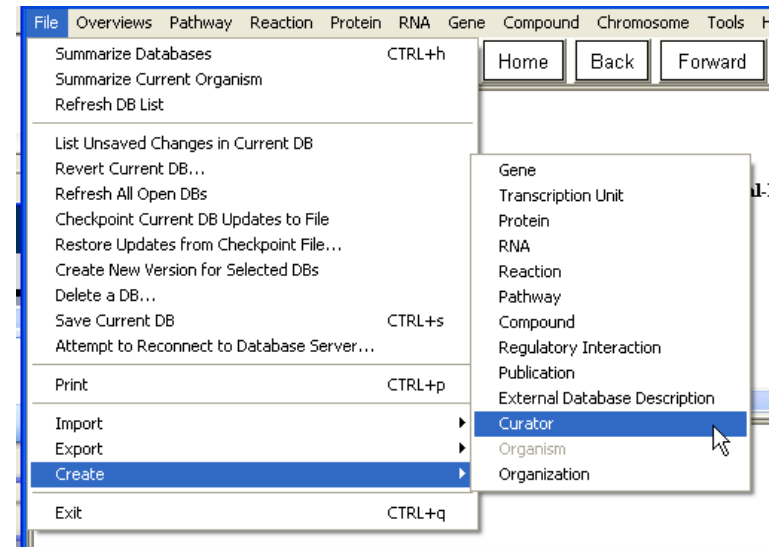


Author Credit System



Assigning author credit is the only way to ensure information about the origin of a frame is maintained, even if the frame is exported to other databases.

Author credit is not properly stored unless a curator frame with login information has been created, and the login information was specified as a username under Preferences.



Determined by Windows login, not permanent



Created by: caspi on 12-Aug-2009

Determined by the Author Credit System, permanent

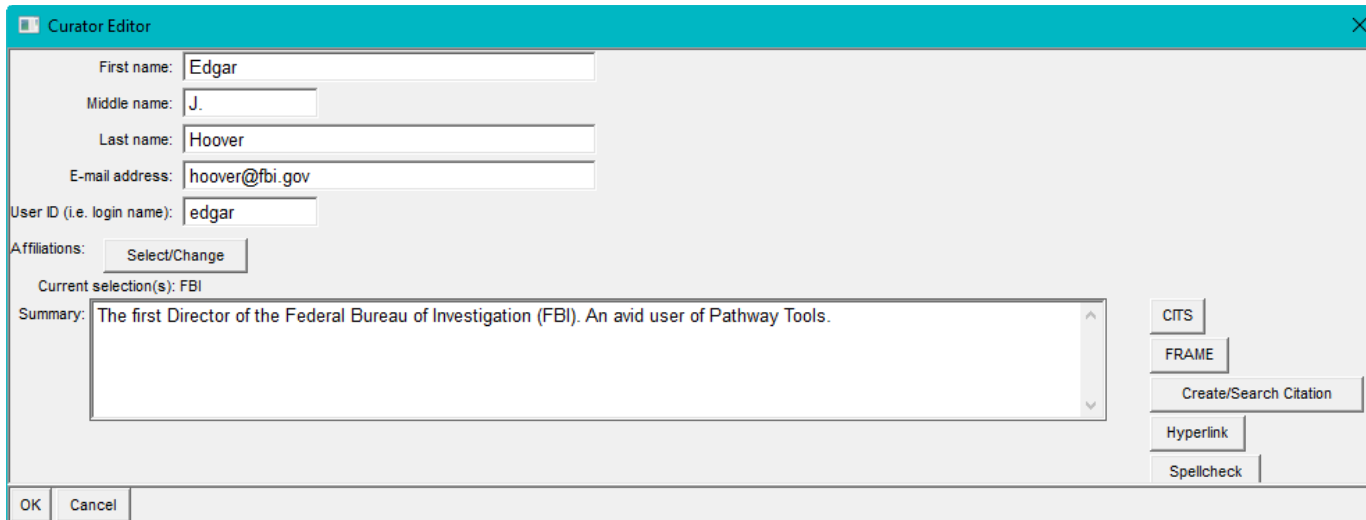


Credits: Created 12-Jun-2013 by Hoover JE, FBI

Organization and Curator Editors

User ID must be longer than 3 characters

If your User ID should be different from your workstation login, specify it under Preferences -> User ID (see next slide).

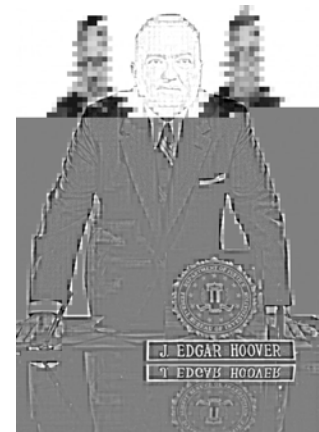
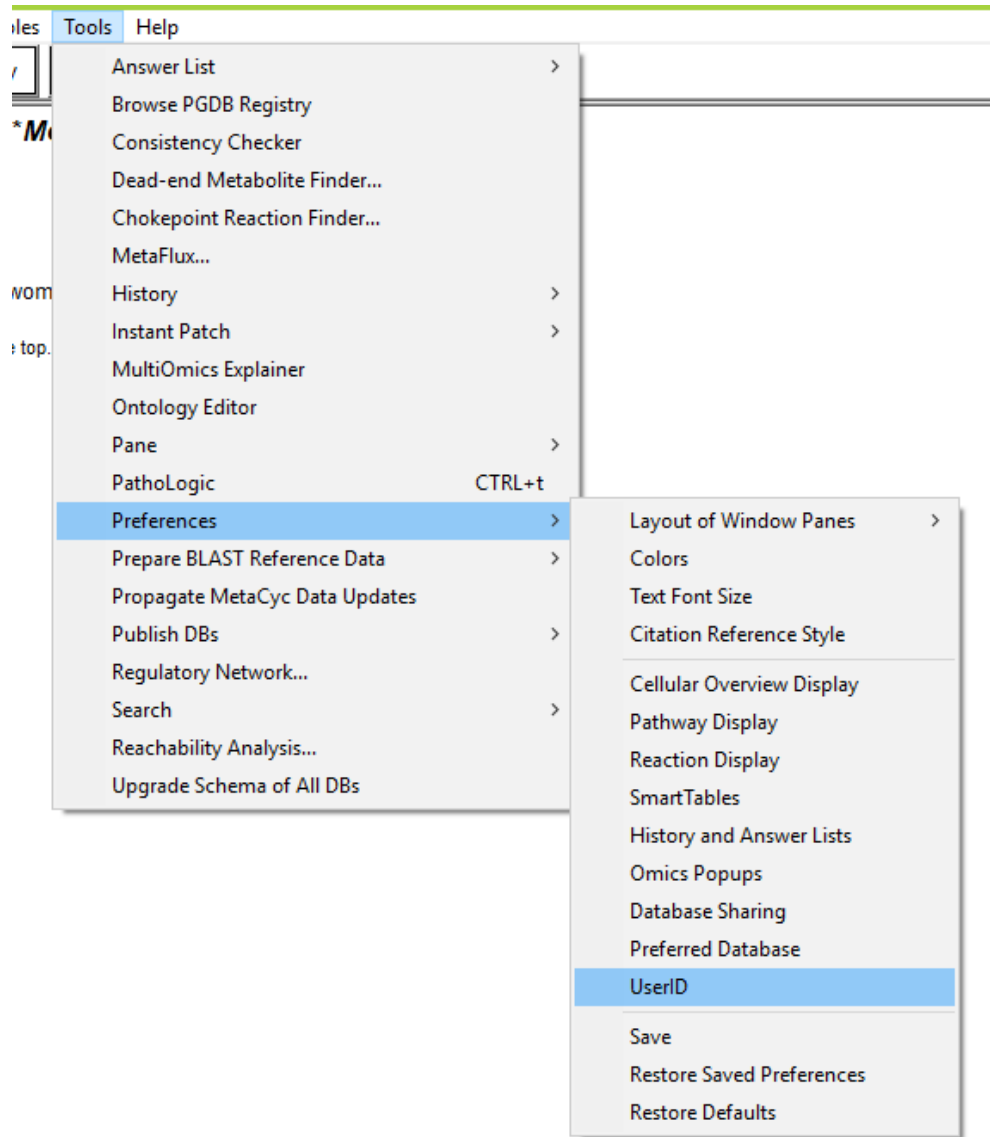


The screenshot shows a window titled "Curator Editor" with the following fields and controls:

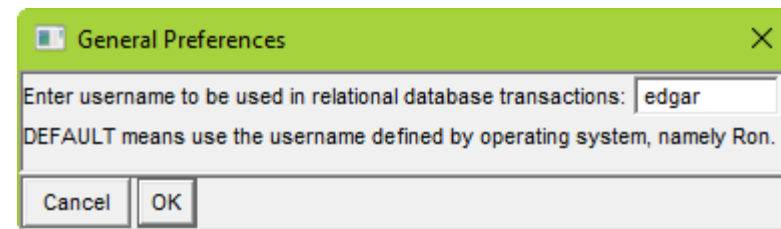
- First name: Edgar
- Middle name: J.
- Last name: Hoover
- E-mail address: hoover@fbi.gov
- User ID (i.e. login name): edgar
- Affiliations: Select/Change
- Current selection(s): FBI
- Summary: The first Director of the Federal Bureau of Investigation (FBI). An avid user of Pathway Tools.
- Buttons: CITS, FRAME, Create/Search Citation, Hyperlink, Spellcheck
- Bottom buttons: OK, Cancel



Specifying the User ID

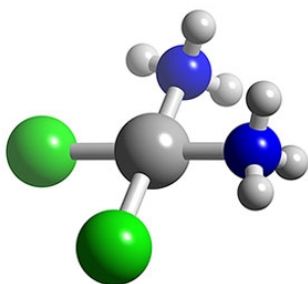


This needs to be done only if the Pathway Tools User ID has to be different from the workstation login.



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="button" value="Make this the Common Name"/>
<input type="text"/>	<input type="button" value="Make this the Common Name"/>
<input type="text"/>	<input type="button" value="Make this the Common Name"/>

Abbreviated name:

Systematic 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:

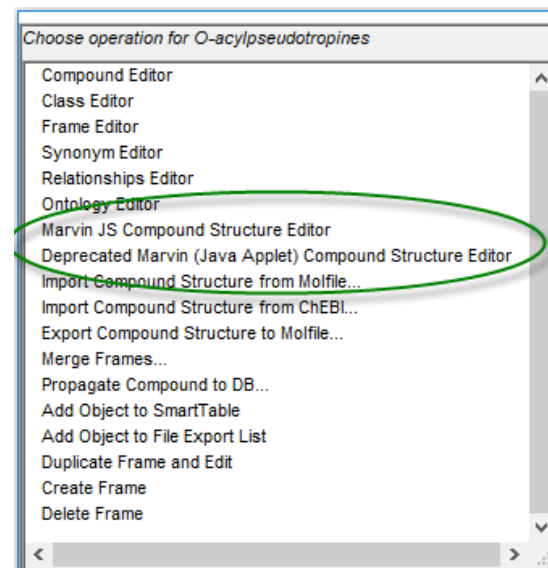
Current selection(s):

More Compound Editing

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

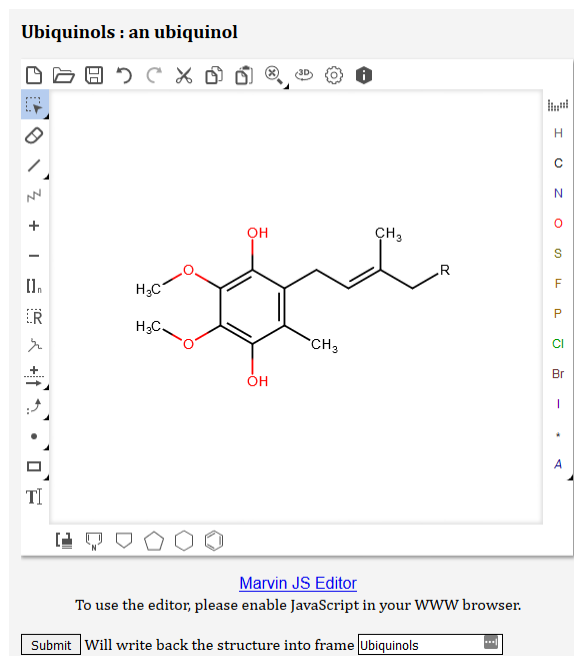
<https://chemaxon.com/products/marvin-js>

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



Other commands:

- Export/Import to Mol files
- Import from ChEBI
- Exporting to other DBs
- Merging
- Duplicate Frame and Edit



The PGDB Info Editor

To access: go to the PGDB home page (click on the Home button, then on the organism name)

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

This is the place to:

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

The screenshot shows the PGDB Info Editor interface. It has three tabs: PGDB Info (selected), MIGS Data, and Annotation Data. The main content area includes:

- Synonyms:** A list of text boxes containing: Bacillus anthracis AmesGenbank entry NC_00399, Bacillus anthracis AmesGenbank entry NC_00399, Bacillus anthracis strain Ames, Bacillus anthracis Ames, and Bacillus anthracis str. Ames.
- Taxon:** A text box with "Bacillus anthracis Ames" and a "NCBI Taxonomy Browser" button. To the right, a "Current taxonomic lineage" path is shown: cellular organisms -> Bacteria -> Firmicutes -> Bacilli -> Bacillales -> Bacillaceae -> Bacillus -> Bacillus anthracis -> Bacillus anthracis Ames.
- Citations:** A text box with "12721629" and two empty text boxes.
- Summary:** A scrollable text area containing: "The primary data source for this dataset is the full genome sequence of [FRAME: TAX:198094] [CITS: [12721629]], derived from Genbank accession NC_003997." and "This dataset was created using the PathoLogic component of the Pathway Tools program [CITS: [10370234]]."
- PGDB Tier:** A dropdown menu set to "2".
- Genome Source:** A text box with "Genbank entry NC_003997".
- PGDB Authors:** A list of text boxes containing: J. Bashkin, Jonathan Wagg, Nan Guo, Peter Karp, and Ron Caspi.
- Project Home Page URL:** An empty text box.
- Project Primary Contact Email:** An empty text box.
- Copyright string:** A text box with "Copyright 2004-2016 SRI International." and "(in HTML format)" below it.
- Footer citation for web pages:** An empty text box.

At the bottom, there are "OK" and "Cancel" buttons.

The Synonym Editor

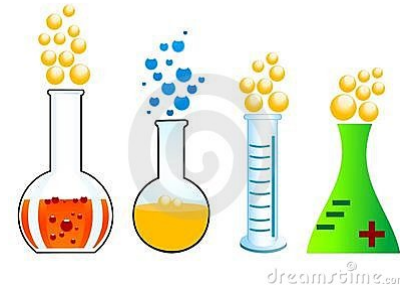


Lets you easily edit the synonyms and set the common name

Common Name:	Value
Common Name:	good
Synonyms:	
excellent	Make this the Common Name
superb	Make this the Common Name
outstanding	Make this the Common Name
magnificent	Make this the Common Name
exceptional	Make this the Common Name
marvellous	Make this the Common Name
wonderful	Make this the Common Name
first-rate	Make this the Common Name
first-class	Make this the Common Name
sterling	Make this the Common Name
fabulous	Make this the Common Name
fantastic	Make this the Common Name
terrific	Make this the Common Name
awesome	Make this the Common Name
wicked	Make this the Common Name

OK Cancel

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

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 (plastidic)	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 (JFRAME:F16B-CPLX) and (JFRAME: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, JFRAME: CPLX-157), the first intracellular species, therefore, being glucose-6-phosphate.

<i>E. coli</i> does constitutively produce JFRAME: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: Change
alkaline phosphatase

Credits:

Revised	Date	Curators	Organizations
	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
	none yet	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

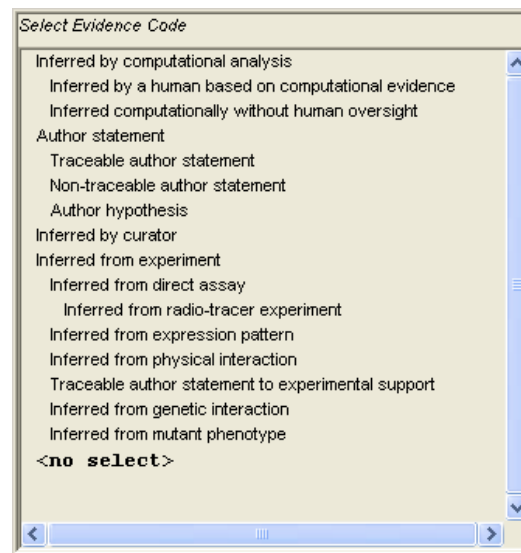
IEP: inferred from expression pattern

IPI: inferred from physical interaction

TAS: traceable author statement

IGI: inferred from genetic interaction

IMP: inferred from mutant phenotype



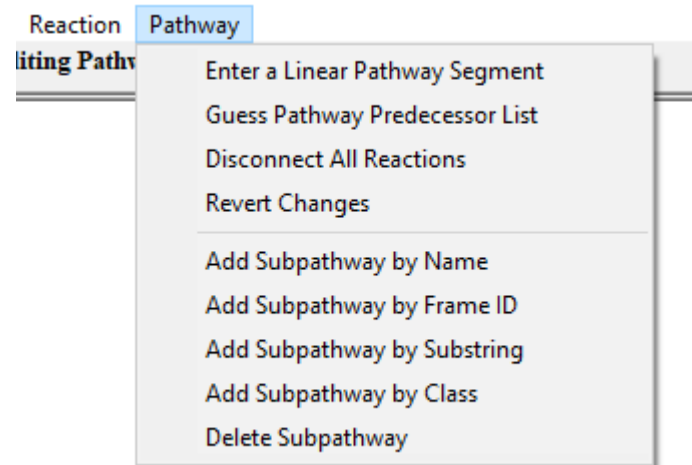
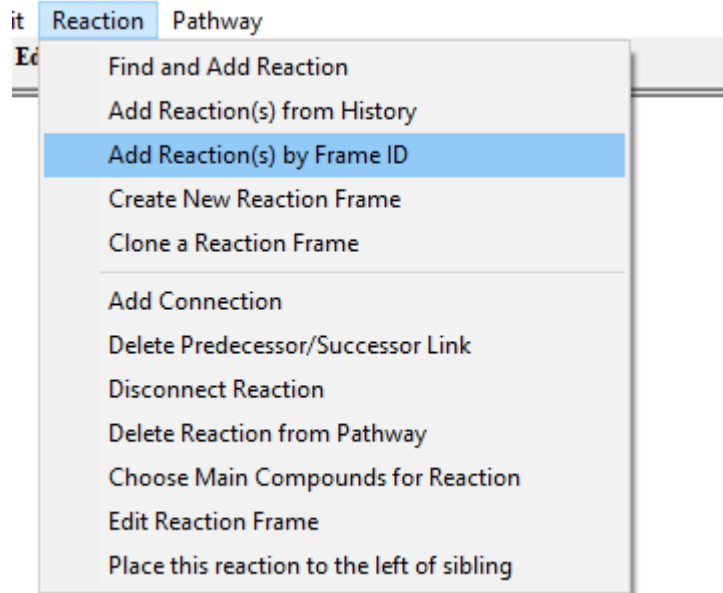
Full documentation for an evidence code is displayed in the Navigator (click the code icon)

The Pathway Editor

Graphically create and modify pathways

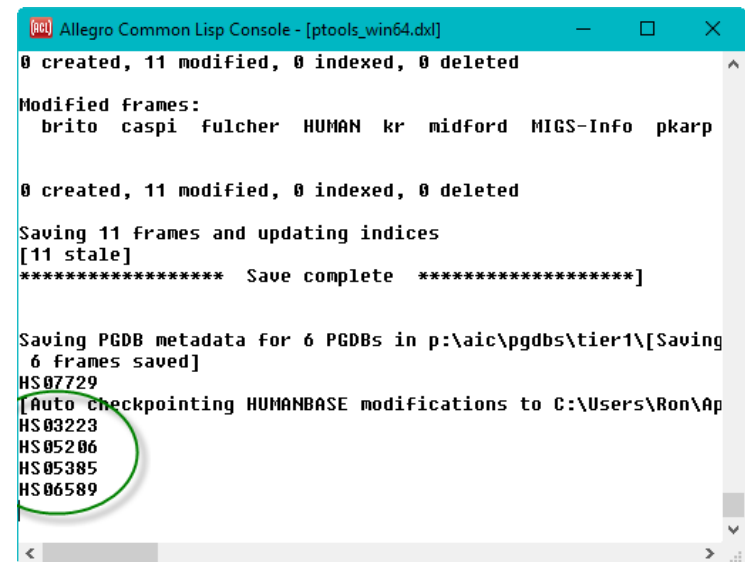
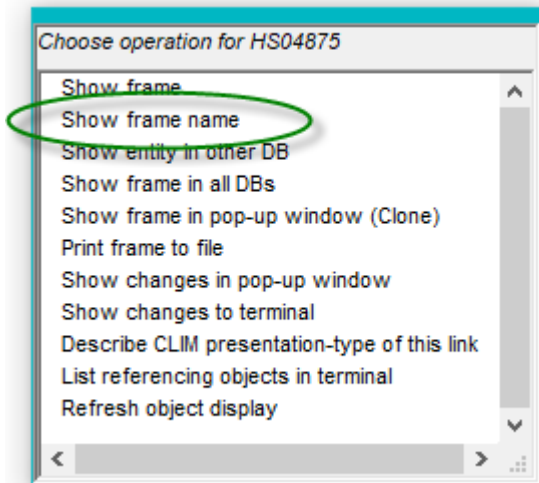
Reaction Menu: add reactions one by one

Pathway Menu: add sub-pathways to create a superpathway



Printing Frame IDs in the Lisp Console

A convenient way to work with Pathway Tools frame IDs is to print them to the Lisp console using the right click → Edit → Show frame name command, then copy them from the console and paste them into the respective editors.



Connecting Reactions

Pathway Editor
Exit Reaction Pathway

Editing Pathway ascorbate degradation to D-ribulose-5-phosphate

LXULRUSP-RXN
5.1.3.22
L-ribulose-5-phosphate = L-xylulose-5-phosphate

RIBULP3EPIM-RXN
5.1.3.1
D-ribulose-5-phosphate = D-xylulose-5-phosphate

RIBULPEPIM-RXN
5.1.3.4
L-ribulose-5-phosphate = D-xylulose-5-phosphate

RXNO-705
4.1.1.85
3-keto-L-gulonate 6-phosphate
= L-xylulose-5-phosphate + CO₂

L-ascorbate $\xrightarrow{H_2O}$ 3-keto-L-gulonate \xrightarrow{ATP} 3-keto-L-gulonate 6-phosphate \xrightarrow{ADP}

Additional Functionality can be accessed by right-clicking on compounds

- Add Connection
- Add Link from/to Pathway
- Add Link from/to Reaction or text
- Delete Polymerization Link**
- Assign Polymerization Name

- Add Connection
- Place this Compound at Cycle Top**
- Add Link from/to Pathway
- Add Link from/to Reaction or text



Pathway Editor Limitations

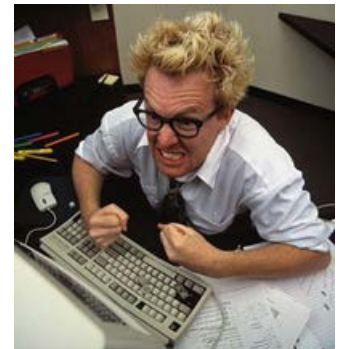
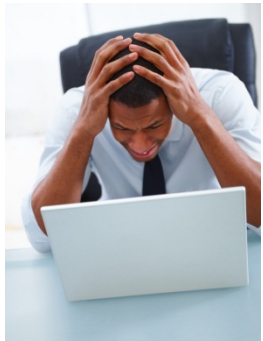
Complex situations can cause ambiguity:

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

- **Fix:**

- try disconnecting reactions and adding them in different order

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



Overview of Creating a Pathway



- Identify all the **metabolites**. Define any missing ones.
- Find the individual **reactions** in the PGDB/MetaCyc and create new reactions if necessary.
- Compose the **pathway** from the individual reactions using the pathway editor.
- Assign a **class** to the pathway.
- Add a **summary**, **citations**, and an **evidence code**.
- Assign the appropriate **enzymes**, create complexes when appropriate.
- Curate information about enzymes and genes, including **evidence codes** for the enzymatic reactions.

Lisp Breaks

- When Lisp encounters an unrecognized command it breaks
- A break is NOT a crash
- When a break occurs, the control moves from the GUI to the Lisp console



To generate a break: type (break) at the listener pane, hit enter

Lisp presents several recovery options

Type `:cont x` where x is the number of the best option

A screenshot of the 'Allegro Common Lisp Console' window. The window title is 'Allegro Common Lisp Console - [ptools_win32.dxl]'. The console output shows the following text:

```
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 :all` at the lisp prompt

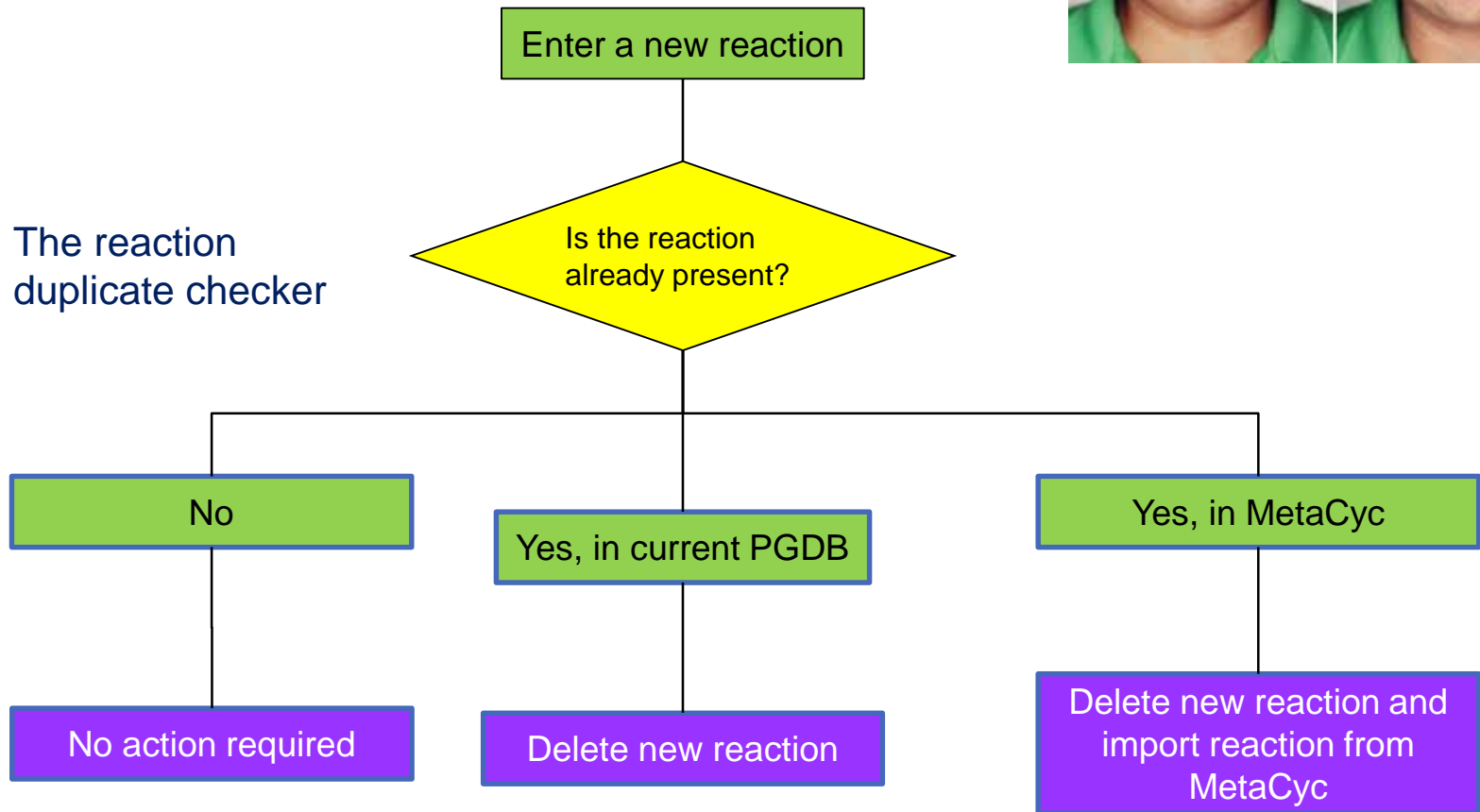
Copy the output, and send it by email to ptools-support@ai.sri.com



```
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): |
```

The Reaction Duplicate Checker



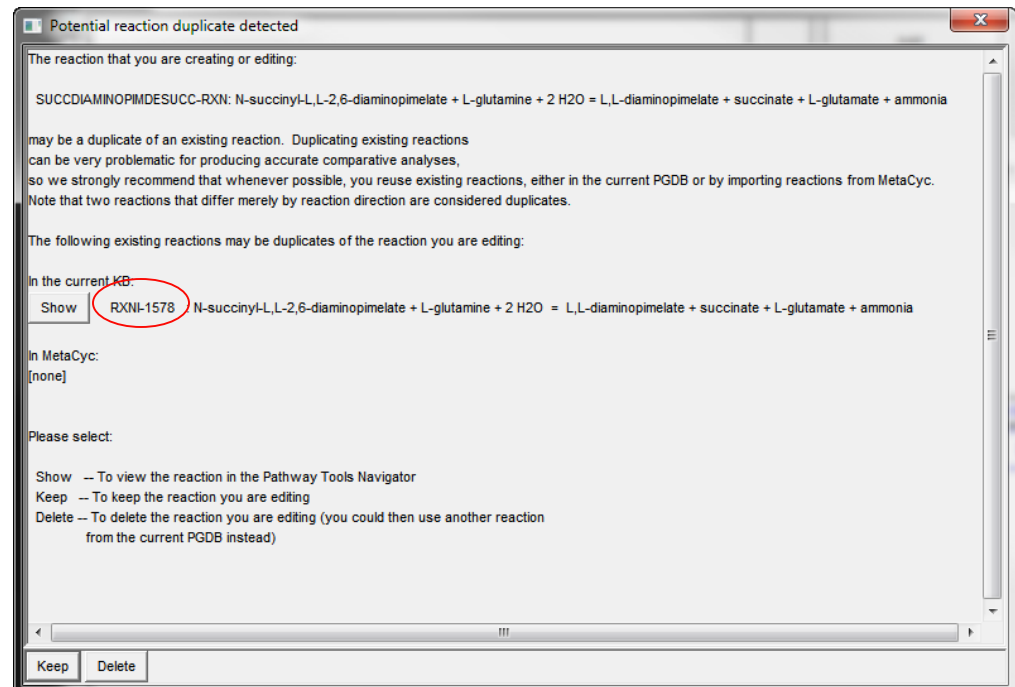
Reuse existing reactions instead of creating duplicates!

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



You should choose the option “Delete”.

Before you delete and close this window, write down the frame ID of the identical reaction, so you could use it later when specifying the pathway.

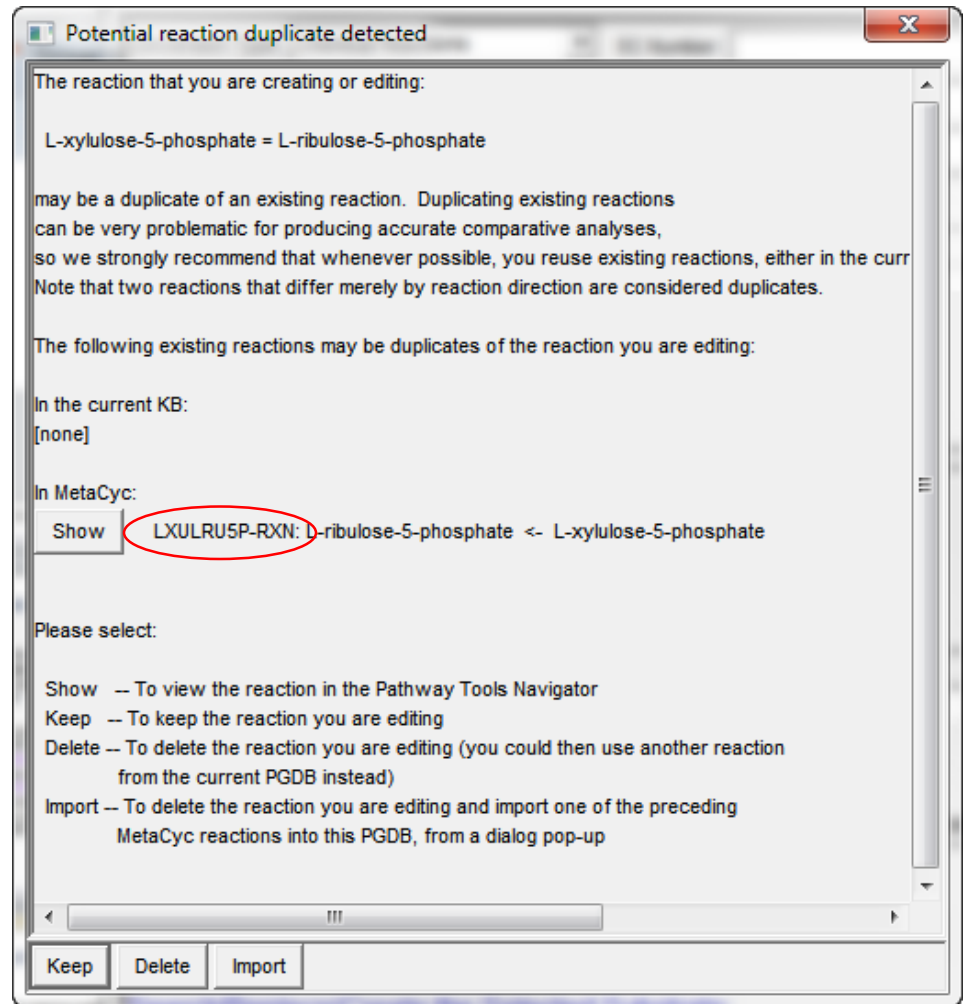


Import MetaCyc reactions instead of creating duplicates!

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**”.

Before you import and close this window, **write down** the frame ID of the identical reaction, so you could use it later when specifying the pathway.



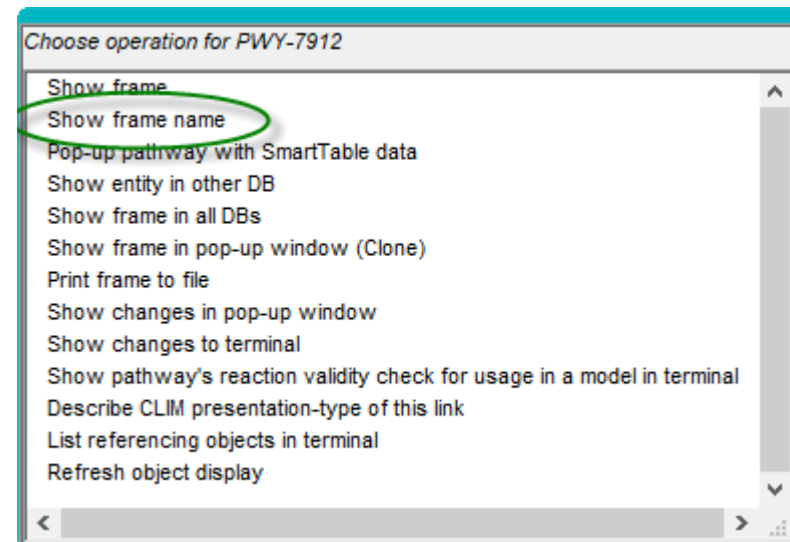
How to get a frame ID?

Navigate to the page of the frame (from the duplicate checker you can click on the Show button)

Right-click on the object name and select Shows →Frame name

Move to the Lisp console and copy the name to the clipboard.

If collecting multiple names, paste it into a text editor



Fill Reaction frame ID's in your handout

Reaction	Frame ID
alpha-L-gulose + 2 NAD+ = 2 pyruvate + 2 NADH + 4 H+	
alpha-L-gulose + NAD = L-gulono-1,4-lactone + NADH + H+	
L-gulono-1,4-lactone + O2 = L-ascorbate + hydrogen peroxide + H+	

Don't forget to include **spaces** between chemical names and terms such as “+” and “=”

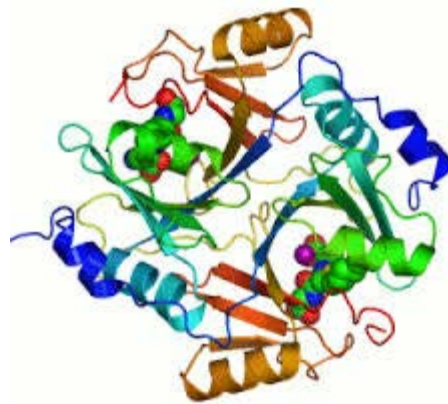
To show a frame ID of an object, you can right-click on it and select Show → Show frame name. The frame ID will be printed in the Lisp console, from which you can copy it.

Editing Pathway/Genome Databases – Lab Section

Exercise 1: Creating reactions and pathways

- Create organization and curator frames, specify your Pathway Tools username
- Enter new reactions
- Import reactions from MetaCyc
- Construct a new pathway with these reactions

Curating Genes and Enzymes



The Gene and Isoform/Coding-Segment Editors

- Enter synonyms/accession numbers
- Enter Links to other databases
- Define transcription direction
- Modify start/end positions
- Define introns
- Create new isoforms
- Define frame shifts

Gene	Start base	End base	Gap size
1	1	210	908 bp. Please specify interpretation.
2	1119	1278	264 bp. Please specify interpretation.
3	4223	4372	369 bp. Please specify interpretation.
4	4762	4922	407 bp. Please specify interpretation.
5	5330	5506	87 bp. Please specify interpretation.
6	6394	6535	3165 bp. Please specify interpretation.
7	9701	9888	500 bp. Please specify interpretation.
8	10389	10530	87 bp. Please specify interpretation.
9	11418	11754	
10			
11			
12			
13			
14			
15			
16			
17			
18			
19			

Base numbers are: **Relative to start of gene**

If specifying multiple coding segments, please provide the appropriate interpretation:
 -- RNA splicing: The RNA transcript is spliced to remove introns.
 -- Protein splicing: The immature pro-peptide is spliced to remove intrins.
 -- Ribosomal slippage: The ribosome slips during translation to generate a programmed internal frame shift.
 You do not need to supply an interpretation here if you are merely using this form

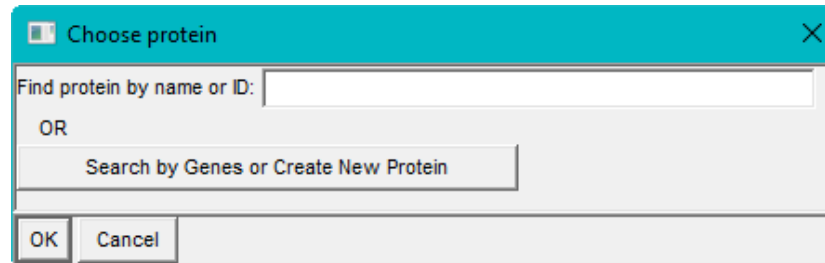
Interpretation: **Not specified**



Adding an Enzyme to a Reaction

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

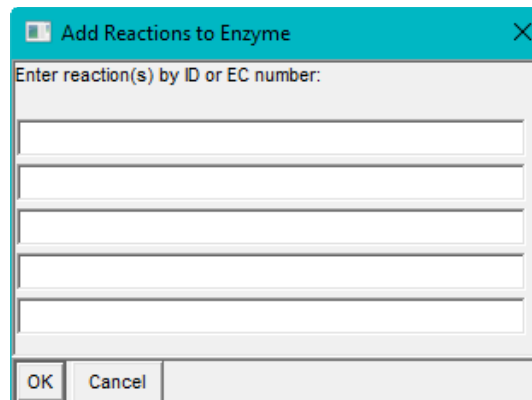
- Right click the reaction, choose Edit → Create/Add enzyme and paste the ID.



The 'Choose protein' dialog box has a teal title bar with a close button. It contains a text input field labeled 'Find protein by name or ID:'. Below this is the text 'OR' and a button labeled 'Search by Genes or Create New Protein'. At the bottom are 'OK' and 'Cancel' buttons.

Or

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



The 'Add Reactions to Enzyme' dialog box has a teal title bar with a close button. It contains a text input field labeled 'Enter reaction(s) by ID or EC number:'. Below this are five empty text input fields. At the bottom are 'OK' and 'Cancel' buttons.

Creating Protein Complexes

Right-click on a protein and select

Edit → Protein Subunit Structure Editor

Change “Macromolecule Type” from polypeptide to protein complex

Example: a simple homohexamer

Specify Protein Subunit Structure

Protein: serine acetyltransferase

Macromolecule Type: protein complex Number of distinct subunits: 1

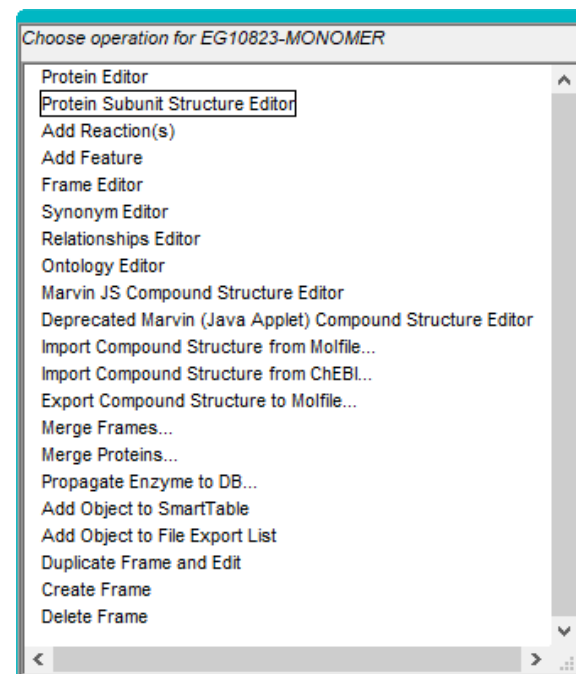
Specific Class(es), if any:

e.g. A homotetramer counts as 1 gene product, not 4 -- the number supplied here should match the number of subunits supplied below.

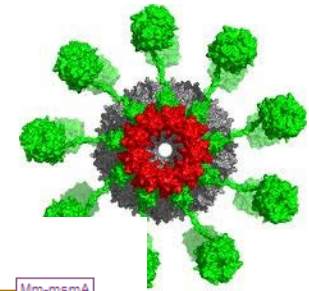
For a complex of complexes, check the "Complex?" box below for each subunit that is a complex, and enter the number of distinct subunits and the components for each. The coefficient can be omitted if it is not known. The Status column below tells if a protein already exists or will be created.

Subunit	Complex?	Gene or #Subunits	Coefficient	Status
serine acetyltransferase	<input checked="" type="checkbox"/>	Gene: cysE	6	Already exists (edit name to create a new object)

OK Cancel

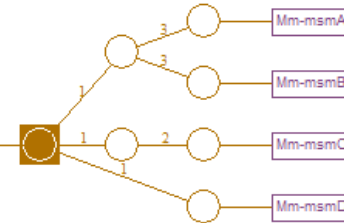


Specifying Multiple Subunits



Gene-Reaction Schematic ?

1,14.13.111 : methanesulfonate + NADH + oxygen -> form...



Specify Protein Subunit Structure ✕

Protein: methanesulfonate monooxygenase

Species: Select

Macromolecule Type: Number of distinct subunits:

Specific Class(es), if any:

e.g. A homotetramer counts as 1 gene product, not 4 -- the number supplied here should match the number of subunits supplied below.
For a complex of complexes, check the "Complex?" box below for each subunit that is a complex, and enter the number of distinct subunits and the components for each. The coefficient can be omitted if it is not known. The Status column below tells if a protein already exists or will be created.

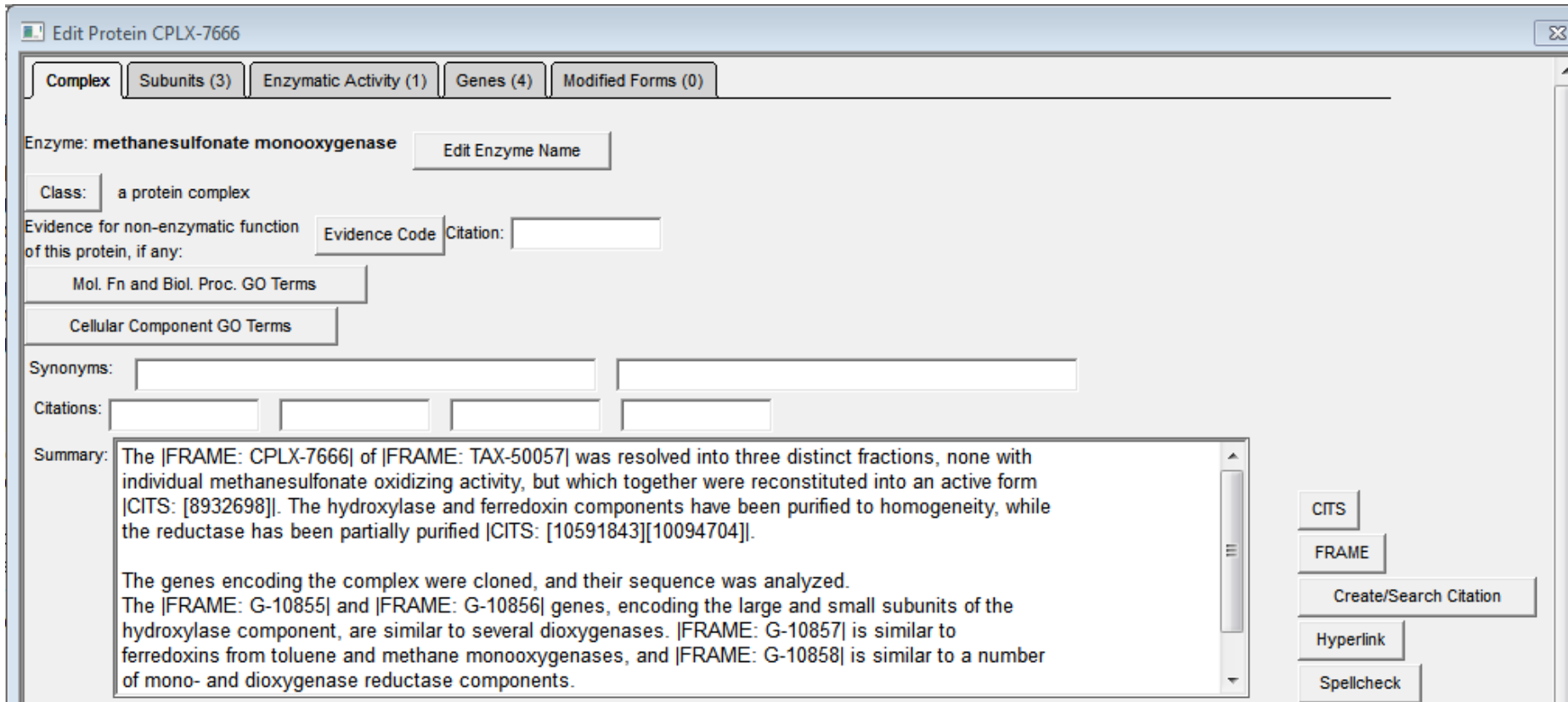
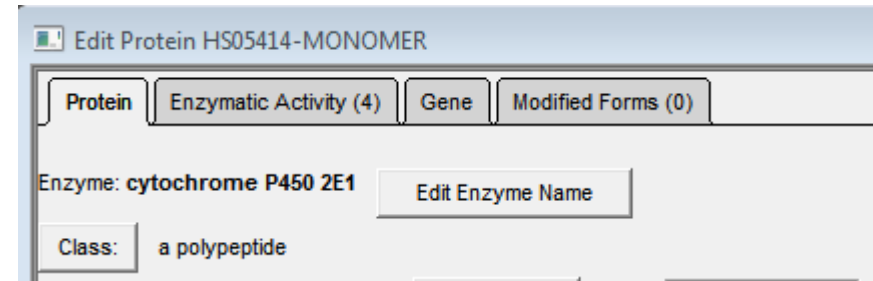
Subunit	Complex?	Gene or #Subunits	Coefficient	Status
<input type="text" value="methanesulfonate monooxygenase hydroxylase component"/>	<input checked="" type="checkbox"/>	#Subunits: <input type="text" value="2"/>	<input type="text" value="1"/>	Already exists (edit name to create a new object)
<input type="text" value="methanesulfonate monooxygenase hydroxylase componen"/>	<input type="checkbox"/>	Gene: <input type="text" value="msmA"/>	<input type="text" value="3"/>	Already exists (edit name to create a new object)
<input type="text" value="methanesulfonate monooxygenase hydroxylase componen"/>	<input type="checkbox"/>	Gene: <input type="text" value="msmB"/>	<input type="text" value="3"/>	Already exists (edit name to create a new object)
<input type="text" value="methanesulfonate monooxygenase ferredoxin component"/>	<input checked="" type="checkbox"/>	#Subunits: <input type="text" value="1"/>	<input type="text" value="1"/>	Already exists (edit name to create a new object)
<input type="text" value="methanesulfonate monooxygenase ferredoxin component s"/>	<input type="checkbox"/>	Gene: <input type="text" value="msmC"/>	<input type="text" value="2"/>	Already exists (edit name to create a new object)
<input type="text" value="methanesulfonate monooxygenase reductase component"/>	<input type="checkbox"/>	Gene: <input type="text" value="msmD"/>	<input type="text" value="1"/>	Already exists (edit name to create a new object)

OK Cancel

The Protein Editor

(monomer edition)

(protein complex edition)



Protein Editor – First Tab

For an example of a complex, open CPLXI-62 in the *H. pylori* PGDB

Edit Protein CPLXI-62

Complex | Subunits (4) | Enzymatic Activity (1) | Genes (4) | Modified Forms (0)

Enzyme: **2-oxoglutarate:acceptor oxidoreductase**

Class: a protein complex

Evidence for non-enzymatic function of this protein, if any: Evidence Code: Citation:

Mol. Fn and Biol. Proc. GO Terms

Cellular Component GO Terms

Synonyms:

Citations:

Summary: 2-Oxoglutarate:acceptor oxidoreductase (OOR) of *H. pylori* is a heterotetramer consisting of the following four subunits, OorA, OorB, OorC and OorD.

2-Oxoglutarate:acceptor oxidoreductase (OOR) catalyzes the reversible oxidative decarboxylation of 2-oxoglutarate to form succinyl-CoA [CITS: [9495749]]. OOR is one of three atypical enzymes in the proposed complete citric acid cycle of *H. pylori*, replacing the typical citric acid cycle enzyme α -ketoglutarate dehydrogenase, which is not found in the annotated genome of *H. pylori* [CITS: [9495749][10809701].

OOR from *H. pylori* has been partially purified and characterized [CITS: [9495749]]. OOR from *H. pylori* and other bacteria are extremely oxygen labile, unlike their 2-oxoacid dehydrogenase multienzyme complex counterparts]

Molecular Weight (kD, experimental): Citation: pl: Citation:

Links to other databases:

Database	ID	Relationship
----	<input type="text"/>	Same Entity

Credits:

Date	Curators	Organizations
Created	<input type="button" value="Select/Change"/> <input type="button" value="Create"/>	<input type="button" value="Select/Change"/> <input type="button" value="Create"/>
24-Apr-2003	Current selection(s): Krieger CJ	Current selection(s): SRI International
----	<input type="button" value="Select/Change"/> <input type="button" value="Create"/>	<input type="button" value="Select/Change"/> <input type="button" value="Create"/>
none yet	Current selection(s):	Current selection(s):

Update Last-Curated Date ?

Protein Editor - Subunits Tab

- 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

Sections below are for the following subunits:
HP0588 gene product: delta subunit of 2-oxoglutarate:acceptor oxidoreductase
HP0589 gene product: alpha subunit of 2-oxoglutarate:acceptor oxidoreductase
HP0590 gene product: beta subunit of 2-oxoglutarate:acceptor oxidoreductase
HP0591 gene product: gamma subunit of 2-oxoglutarate:acceptor oxidoreductase

Subunit: **HP0588 gene product**

Name: Coefficient:

Gene Classes:

Mol. Fn and Biol. Proc. GO Terms

Cellular Component GO Terms

Synonyms:

Citations:

Summary:

Molecular Weight (kD, experimental): Citation:

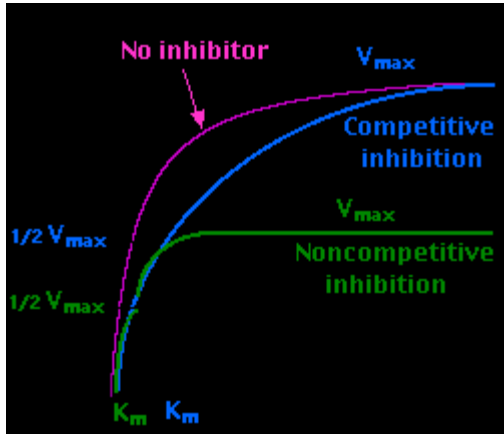
is initial methionine cleaved?

Links to other databases:

Database	ID	Relationship
InterPro	IPR017896	In Family
InterPro	IPR017900	In Family
Pfam	PF13237	In Family
Prosite	PS00198	In Family
Prosite	PS51379	In Family
Protein Model Portal	O25310	Same Entity
Mint	MINT-17035	Same Entity
String	85962.HP05	Same Entity
Database of Interacting Proteins	DIP-3580N	Same Entity
UniProt	O25310	Same Entity
-----		Same Entity

Features: There are no features currently associated with this protein

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"/>	<input type="text"/>	U/mg (= μmol /mg /min)	<input type="text"/>
L-ornithine	<input type="text"/>	μM	<input type="text"/>	<input type="text"/>	U/mg (= μmol /mg /min)	<input type="text"/>
urea	<input type="text"/>	μM	<input type="text"/>	<input type="text"/>	U/mg (= μmol /mg /min)	<input type="text"/>

OK Cancel

Citations

- Citation boxes
- The CITS field

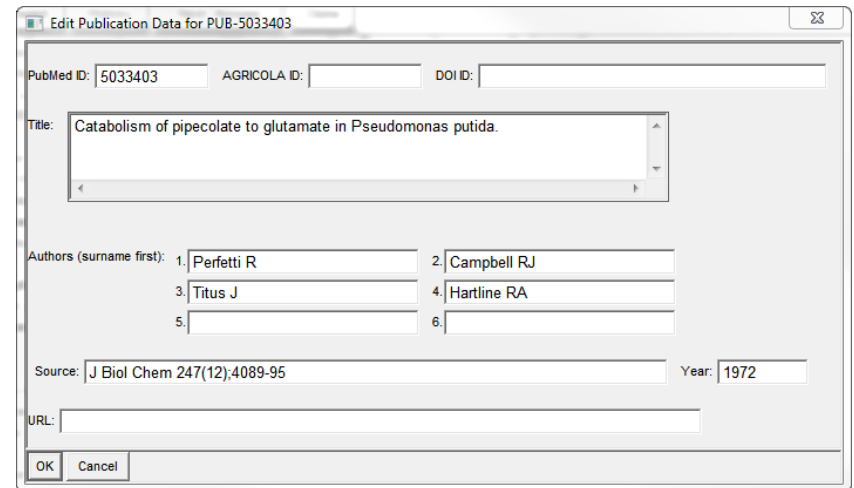


PubMed citations:

- Use PubMed IDs
- Automatically imported when exiting editor
- You can invoke the Publication Editor by right clicking on a citation

Non PubMed citations:

- Enter an ID in the form **Smith06** in a citation box, invoke editor by clicking out of the box. Click on “Search or Create Publication Frame”.
- If you have a DOI number, enter it and click outside the DOI ID box, and it will be retrieved automatically.
- If there is no DOI, type in the details.



Edit Publication Data for PUB-5033403

PubMed ID: 5033403 AGRICOLA ID: DOI ID:

Title: Catabolism of pipecolate to glutamate in *Pseudomonas putida*.

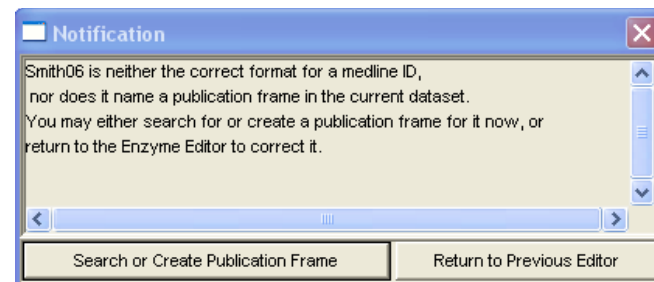
Authors (surname first):

1. Perfetti R	2. Campbell RJ
3. Titus J	4. Hartline RA
5.	6.

Source: J Biol Chem 247(12):4089-95 Year: 1972

URL:

OK Cancel



Notification

Smith06 is neither the correct format for a medline ID, nor does it name a publication frame in the current dataset. You may either search for or create a publication frame for it now, or return to the Enzyme Editor to correct it.

Search or Create Publication Frame Return to Previous Editor

Editing Pathway/Genome Databases – Lab Section

Exercise 2 : Curating enzymes, entering citations, and exporting pathways

- Assign enzymatic activities to proteins
- Define protein complexes
- Create a publication frame
- Export a pathway to a file

Exercise 3 : Constructing Superpathways

- Construct a superpathway

Exercise 4 : Creating a complicated protein complex

- Create a protein complex that involves modified proteins.