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/

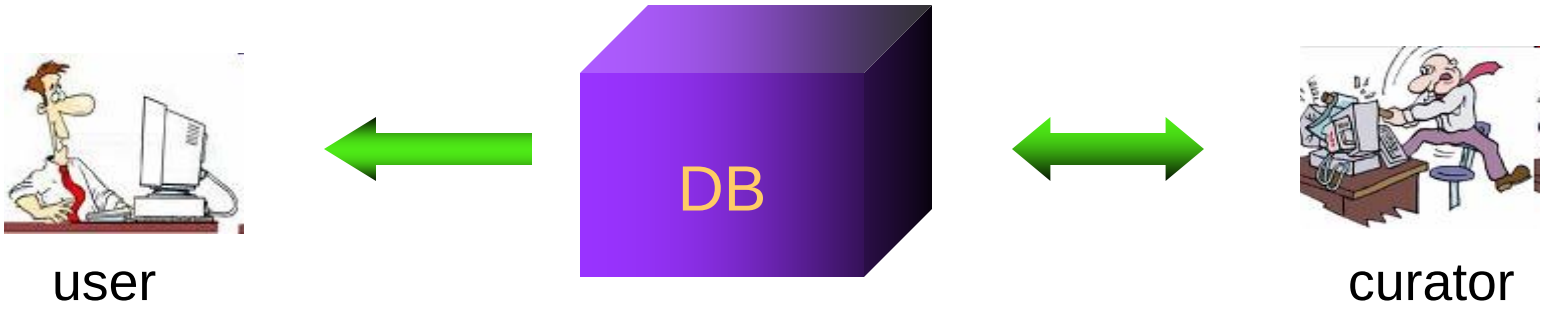
A lot more information is available in the Curator's Guide, at
https://bioinformatics.ai.sri.com/ptools/curatorsguide.pdf

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:

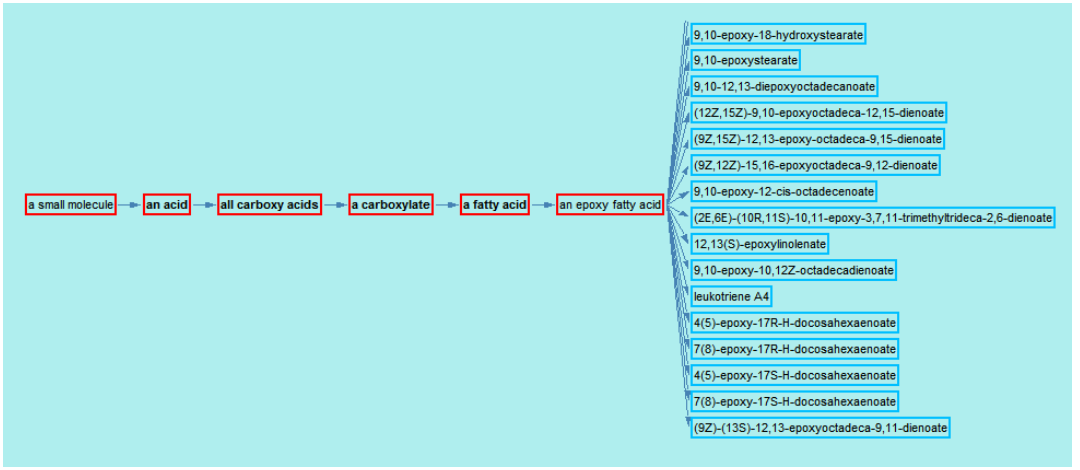
- List Unsaved Changes in Current DB
- Checkpoint Current DB Updates to File
- Restore Updates from Checkpoint File

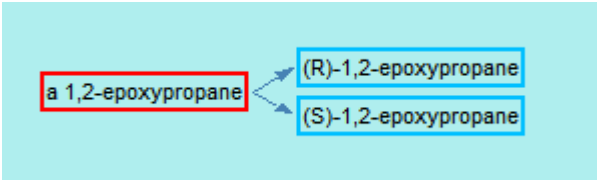
Other Editing-related DB commands under the File menu

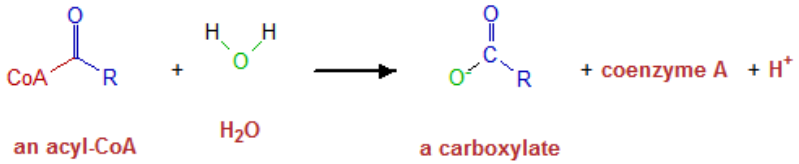
- Create new version for selected DBs (and modifies the default-version file to have the new version opened automatically)
- Save DB as (makes a new copy that can be opened in the same session as the source PGDB)
- Refresh All Open DBs (only MySQL/Oracle DBs)
- Delete a DB

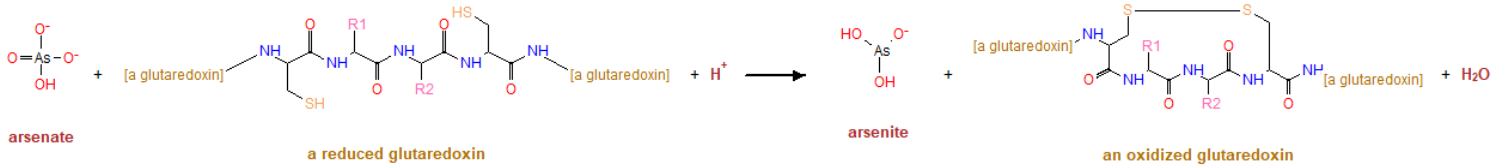
Classes and Instances

- 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 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.

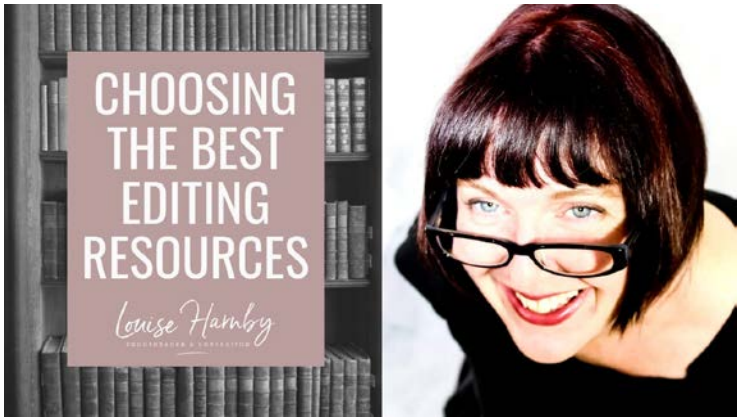
  Legacy objects in Pathway Tools (created before current naming standards) usually deviate from these guidelines.

- 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.

There are many, many Pathway Tools Editors

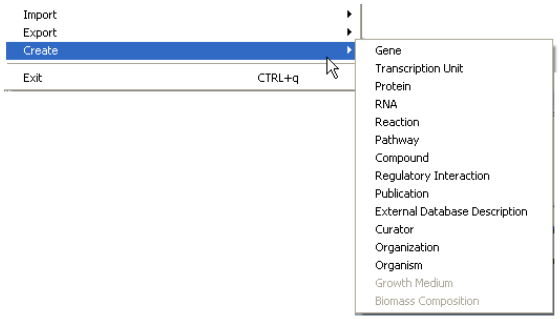
- PGDB Info Editor
- Compound Editor and Compound Structure Editors
- Reaction Editor
- Pathway Editor, Pathway Info Editor
- Signaling Pathway Editor
- Protein Editor and Protein Subunit Structure Editor
- Synonym Editor
- Publication Editor
- Curator/Organization Editors
- Gene Editor
- Isoform/Coding-Segment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



Invoking the Editors

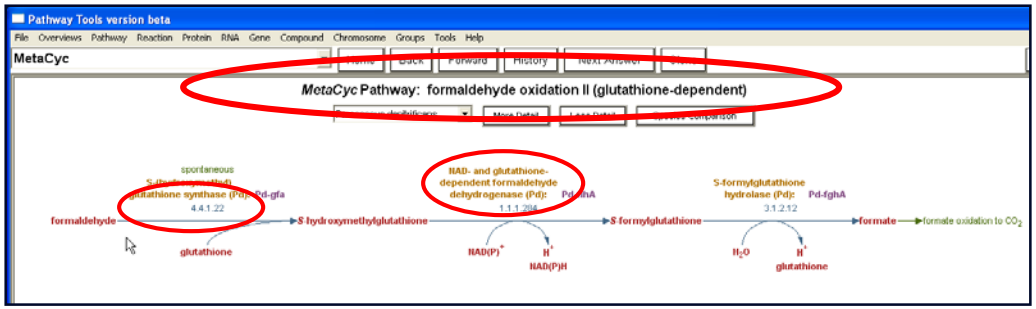
Creating a new Object:

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

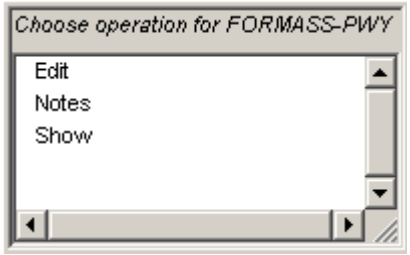


Editing an existing Object:

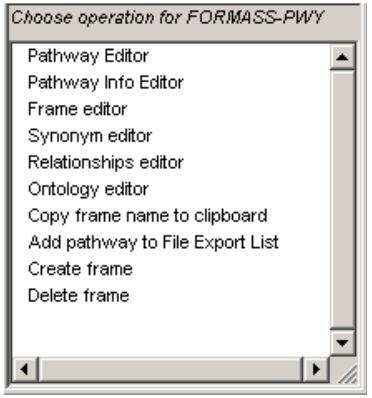
Right-Click on the any clickable name, select Edit, then the appropriate editor



Right-click



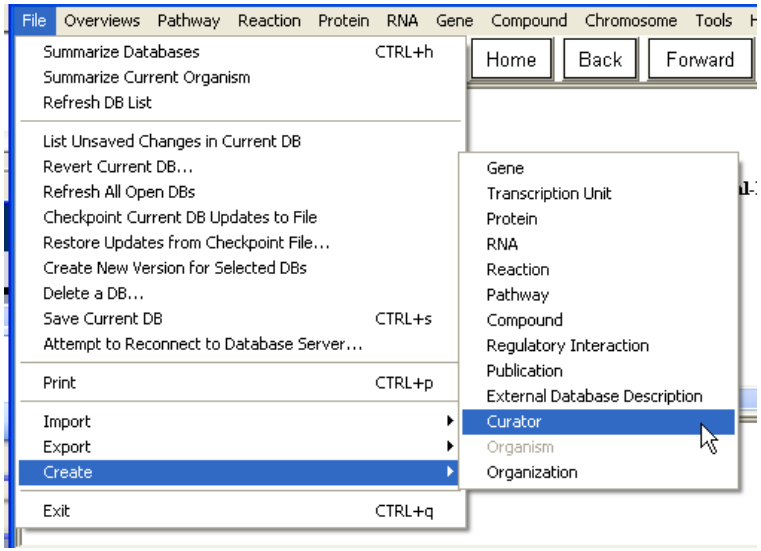
left-click on “Edit”



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.



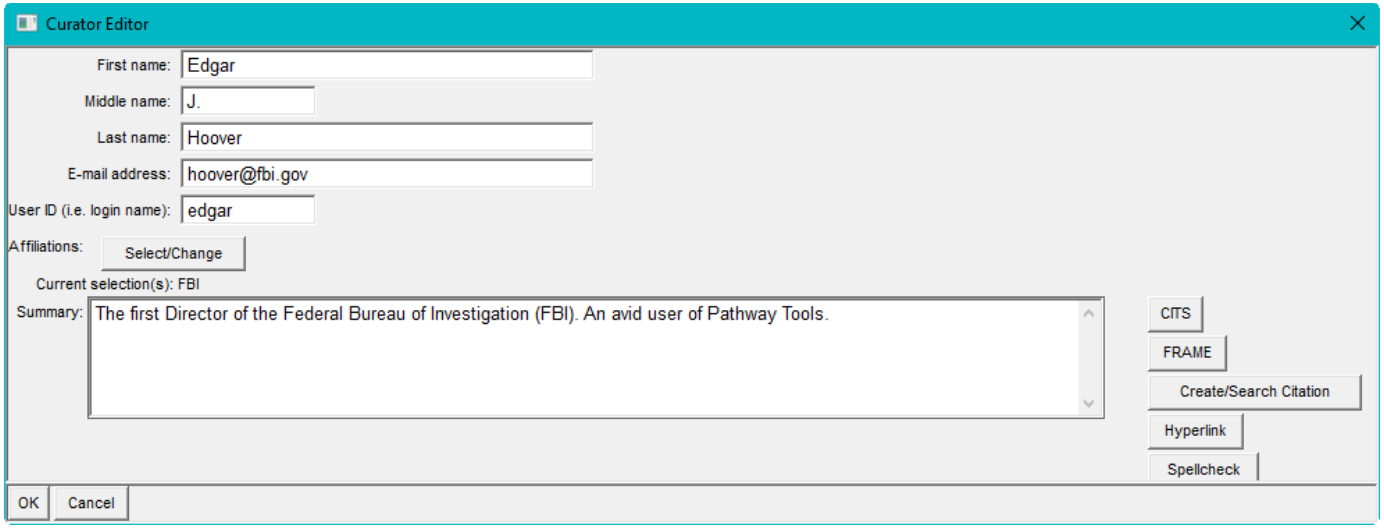
Determined by Windows login, not permanent → Created by: caspi on 12-Aug-2009

Determined by the Author Credit System, permanent → Credits: Created 12-Jun-2013 by **Hoover JE, FBI**

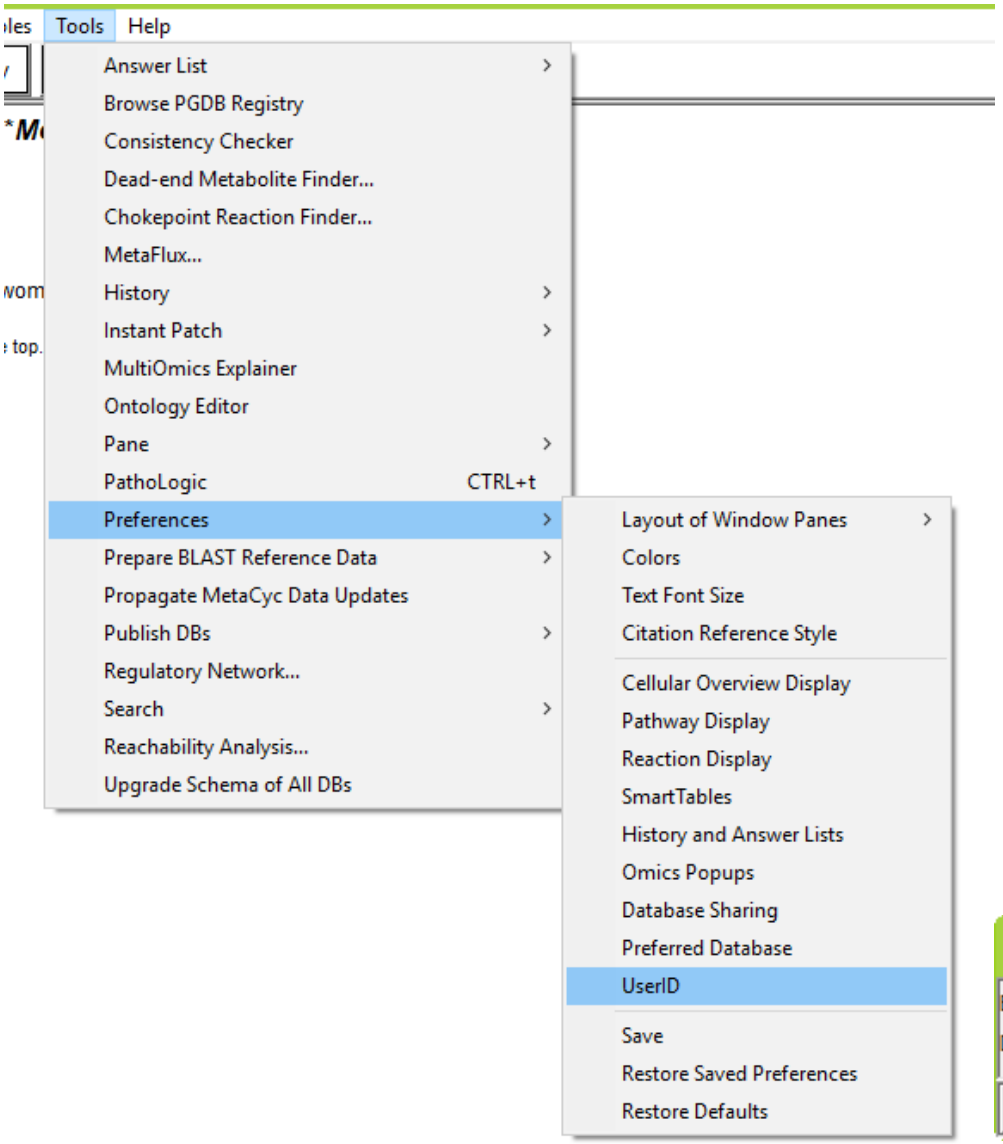
Organization and Curator Editors

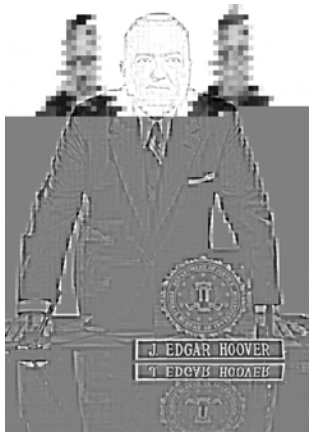
User ID must be longer than 3 characters

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

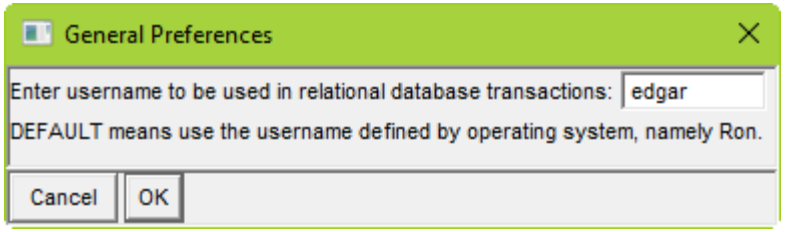


Specifying the User ID





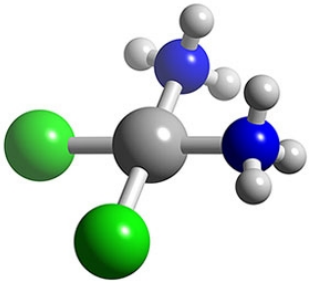
This needs to be done only if the Pathway Tools User ID has to be different from the workstation login.

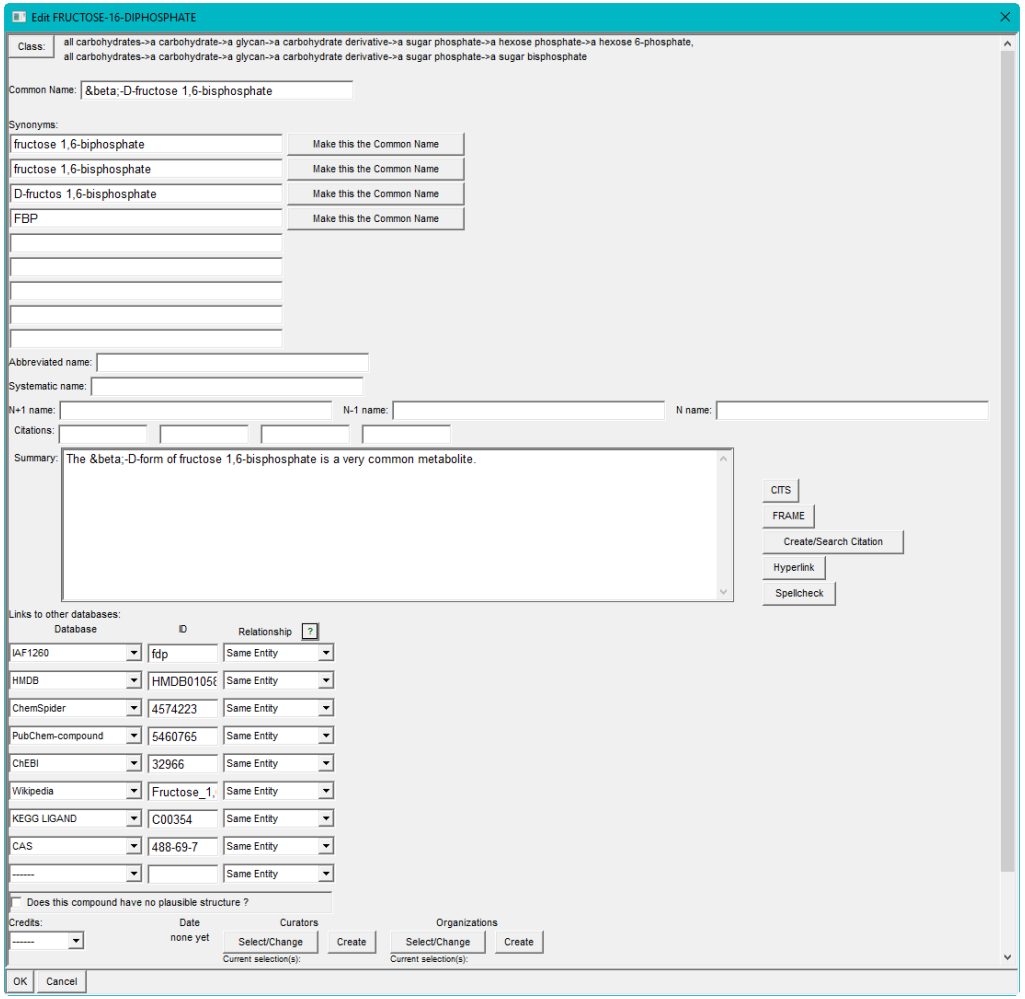


The Compound Editor

- Create or edit a compound (but not its structure)

- Specify Class
- Common Name and Synonyms
- Comments, citations
- Links to other DBs



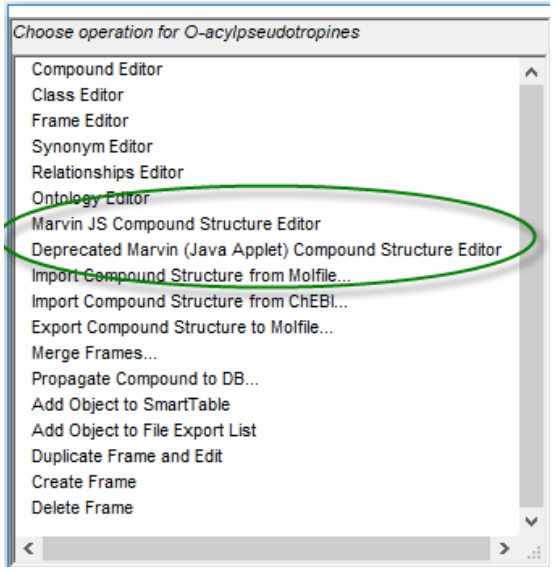


More Compound Editing

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

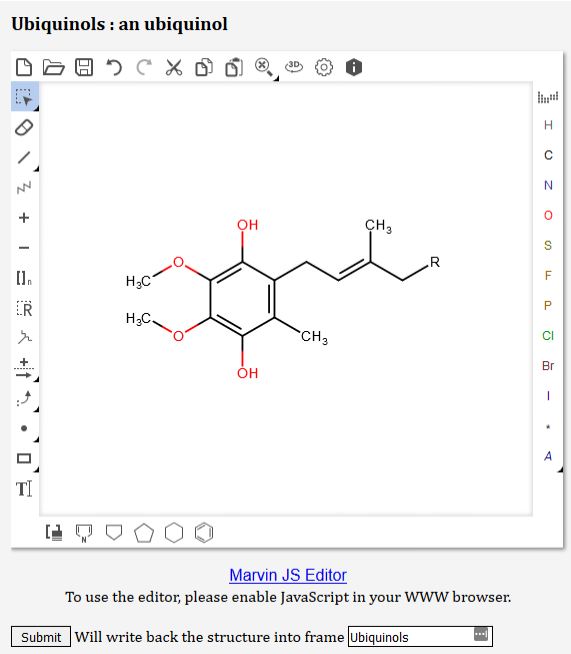
https://chemaxon.com/products/marvin-js

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



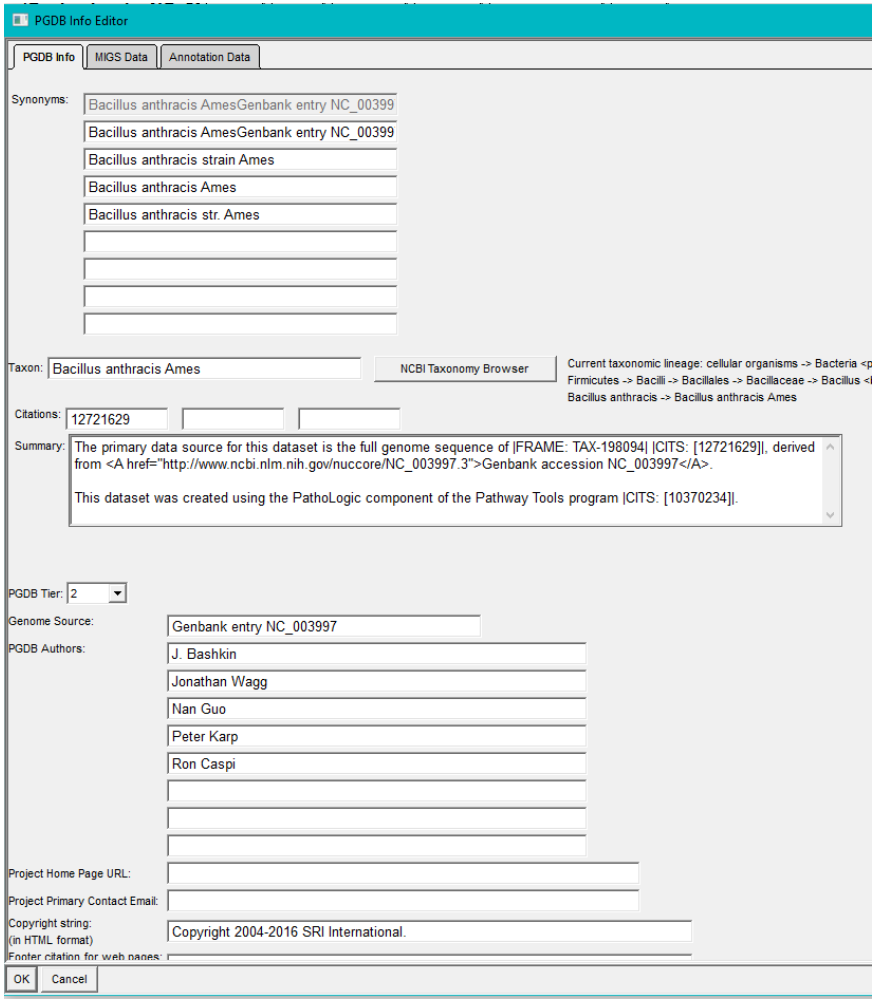
The PGDB Info Editor

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

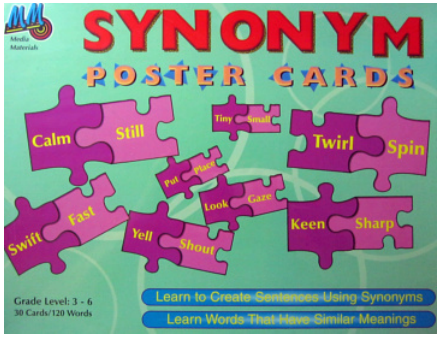
Right-click the organism name, and choose Edit → PGDB Info Editor

This is the place to:

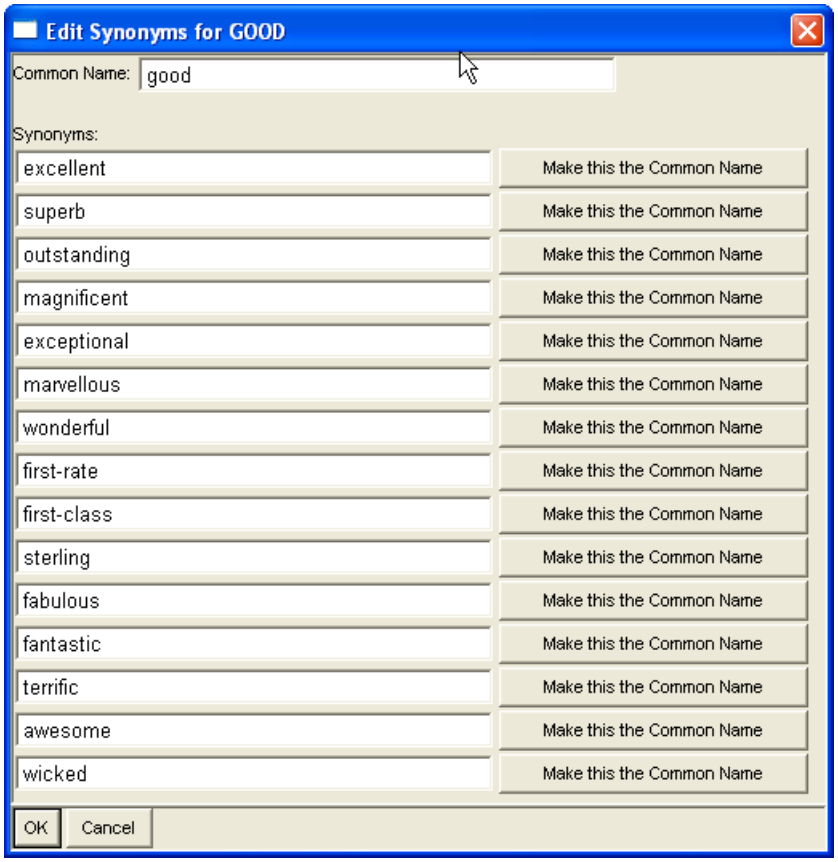
- Specify PGDB authors
- Modify NCBI taxonomy
- Specify a footer
- Set the tier level
- Create a comment for the PGDB home page
- Enter MIGS Data
- Enter Annotation Data



The Synonym Editor



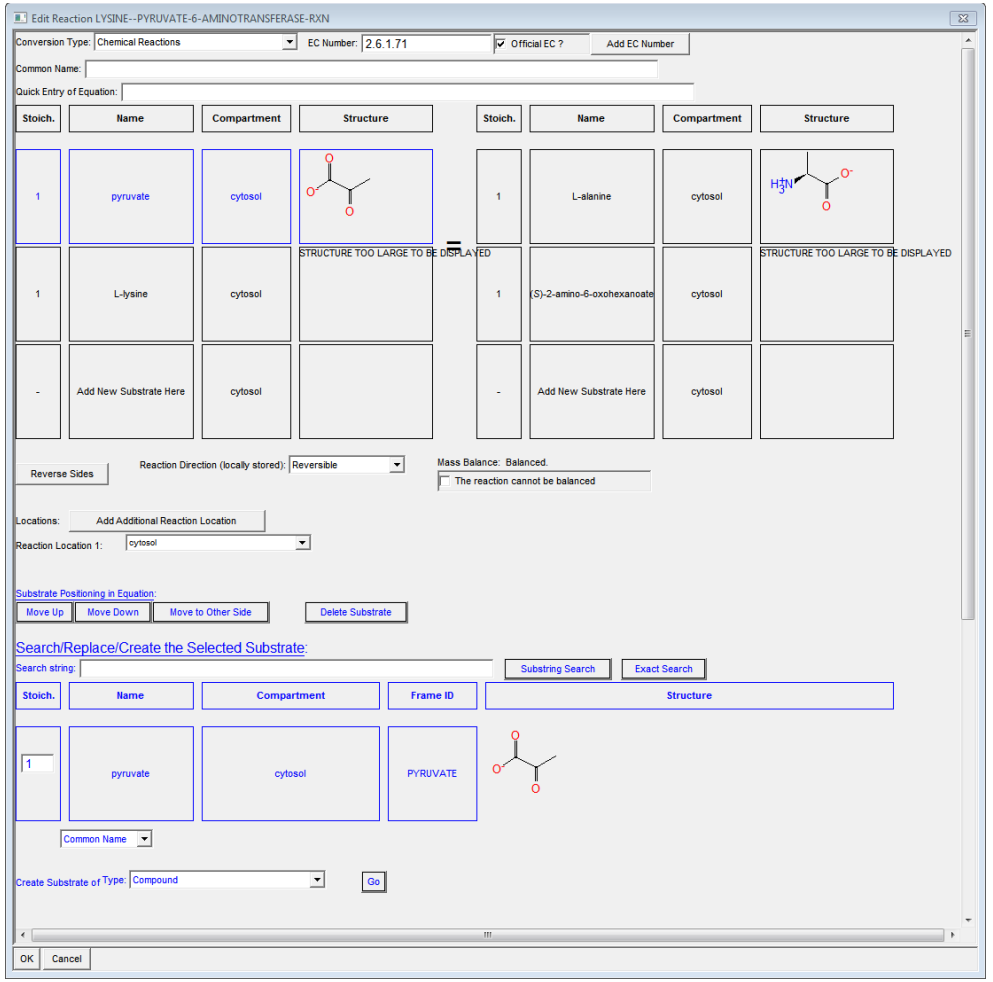
Lets you easily edit the synonyms and set the common name



The Reaction Editor

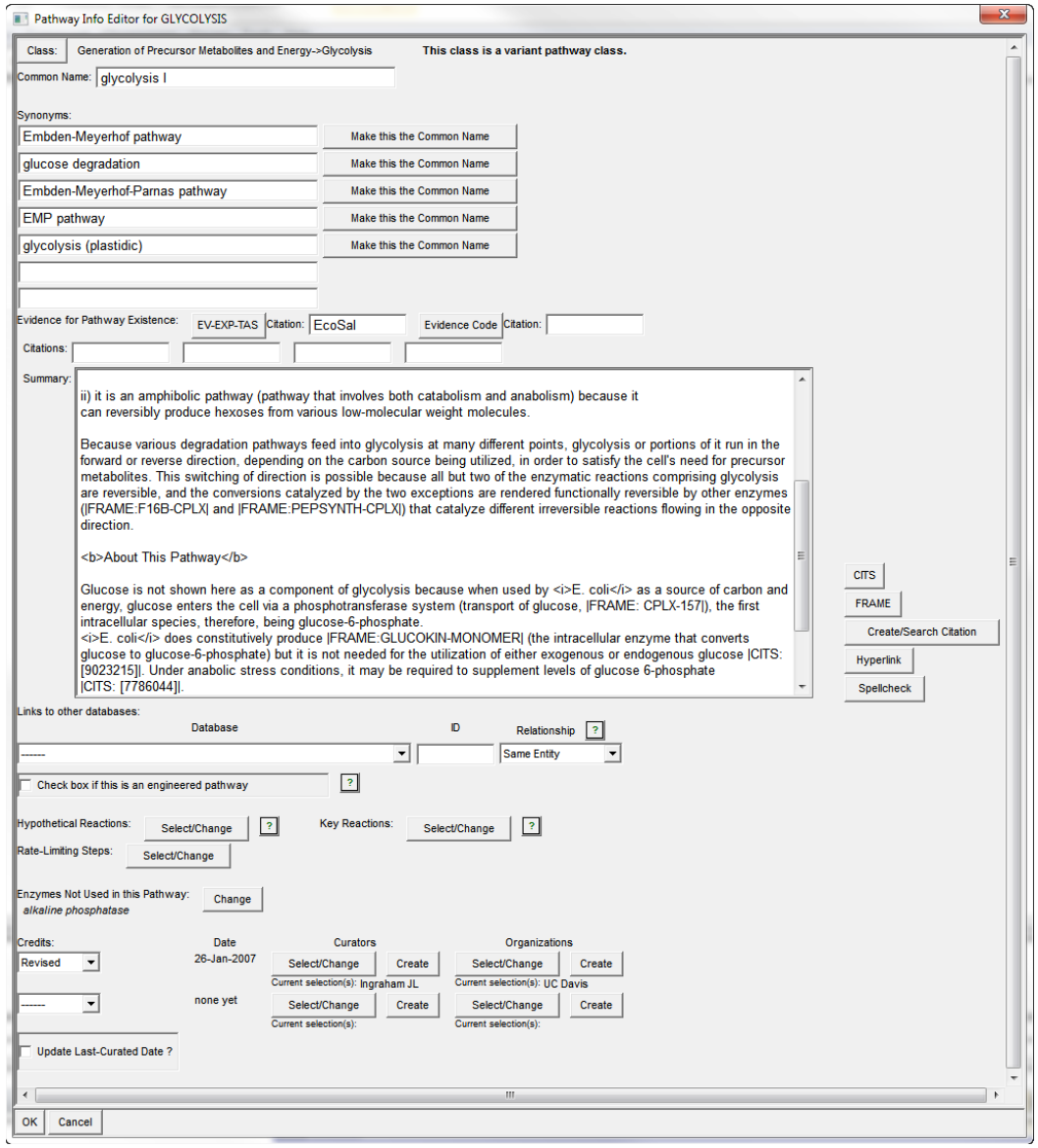
With the Reaction Editor you can:

- 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

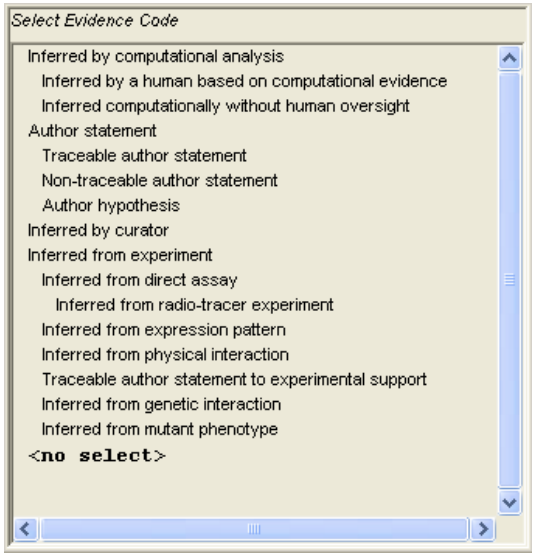


Evidence Codes for Pathways

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



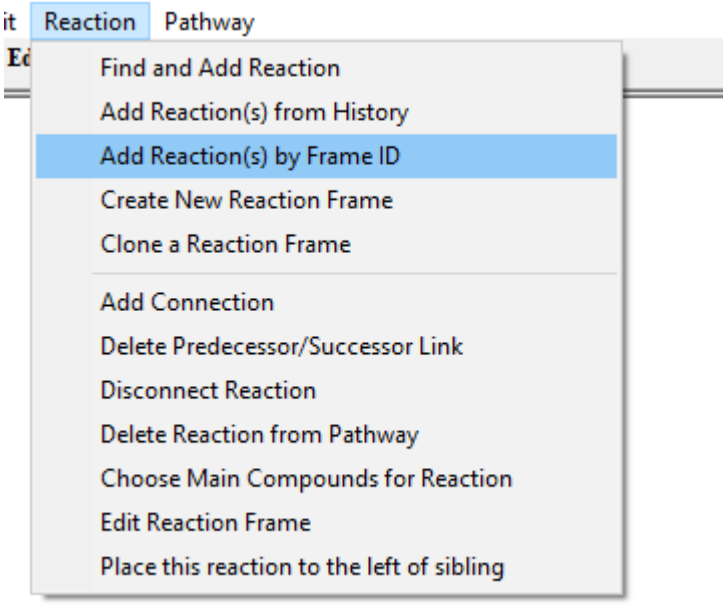
Full documentation for an evidence code is displayed in the Navigator (click the code icon)

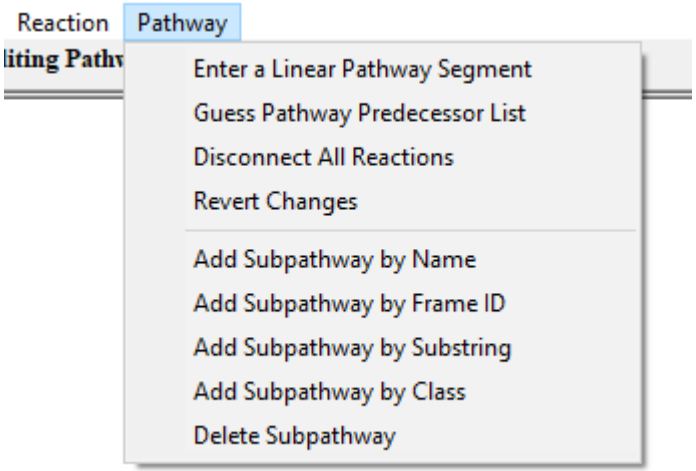
The Pathway Editor

Graphically create and modify pathways

Reaction Menu: add reactions one by one

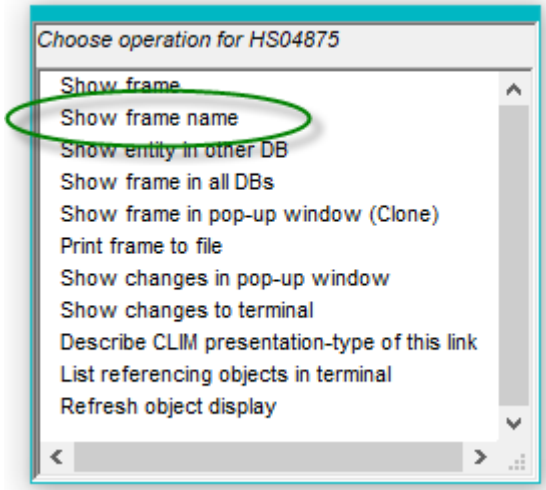
Pathway Menu: add sub-pathways to create a superpathway

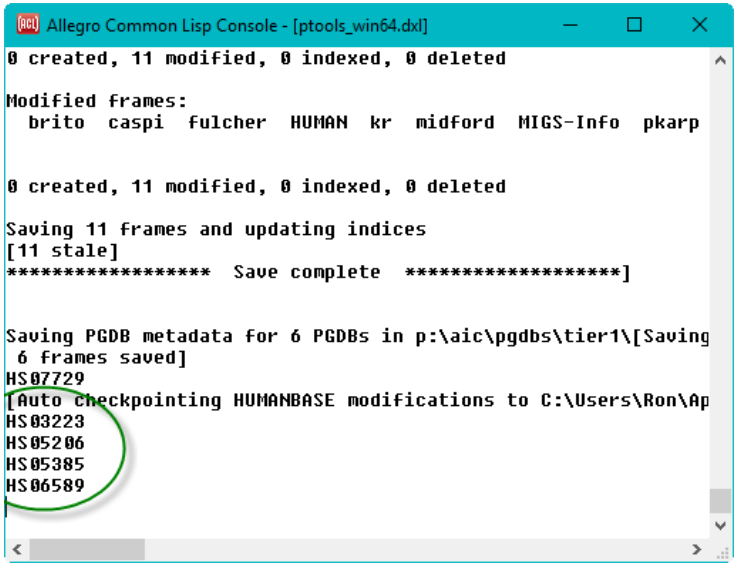




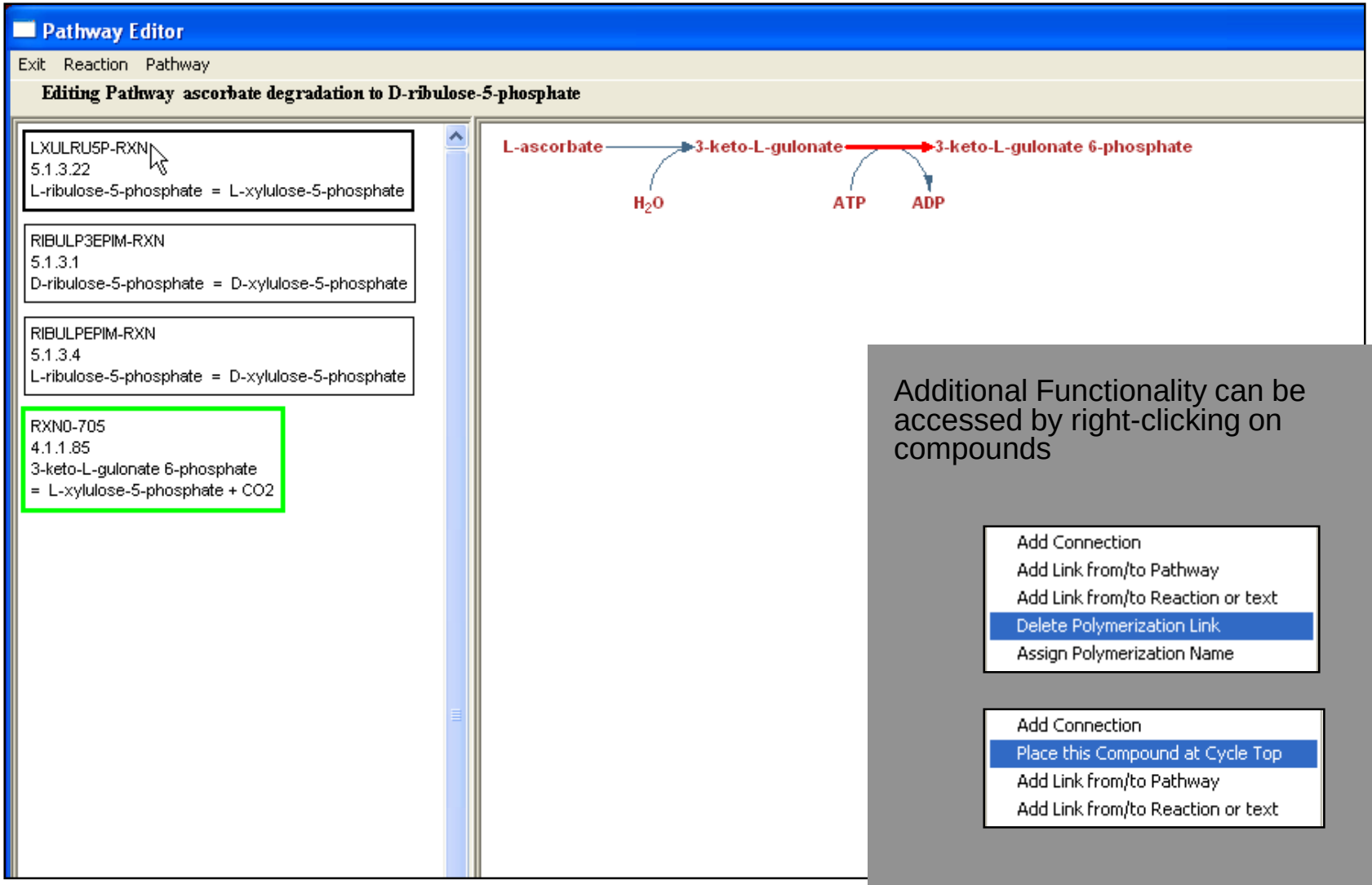
Printing Frame IDs in the Lisp Console

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.





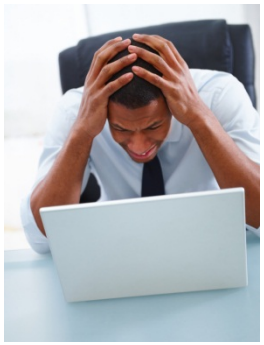
Connecting Reactions



Pathway Editor Limitations

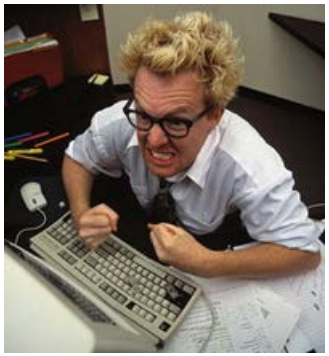
Complex situations can cause ambiguity:

- 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.

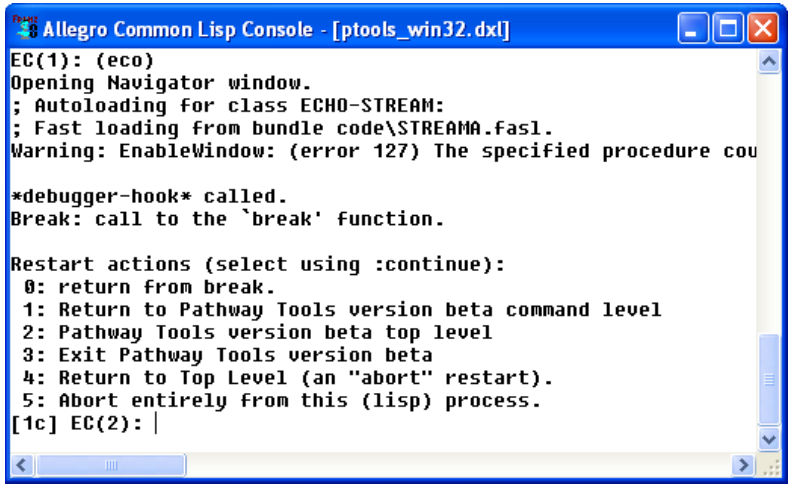
Lisp Breaks

- 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



Bug Reports

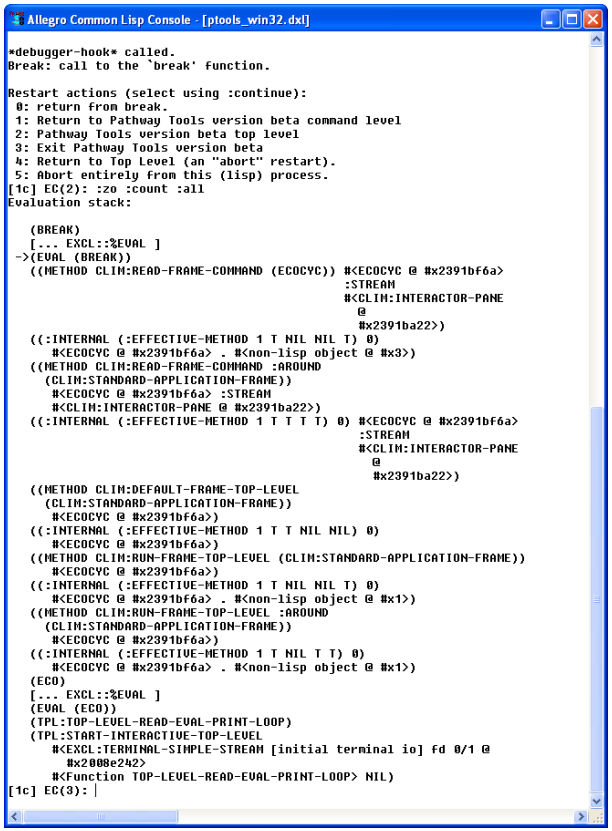
If you get a break as a result of a bug, get the evaluation stack by typing

:zo :count :all at the lisp prompt

Copy the output, and send it

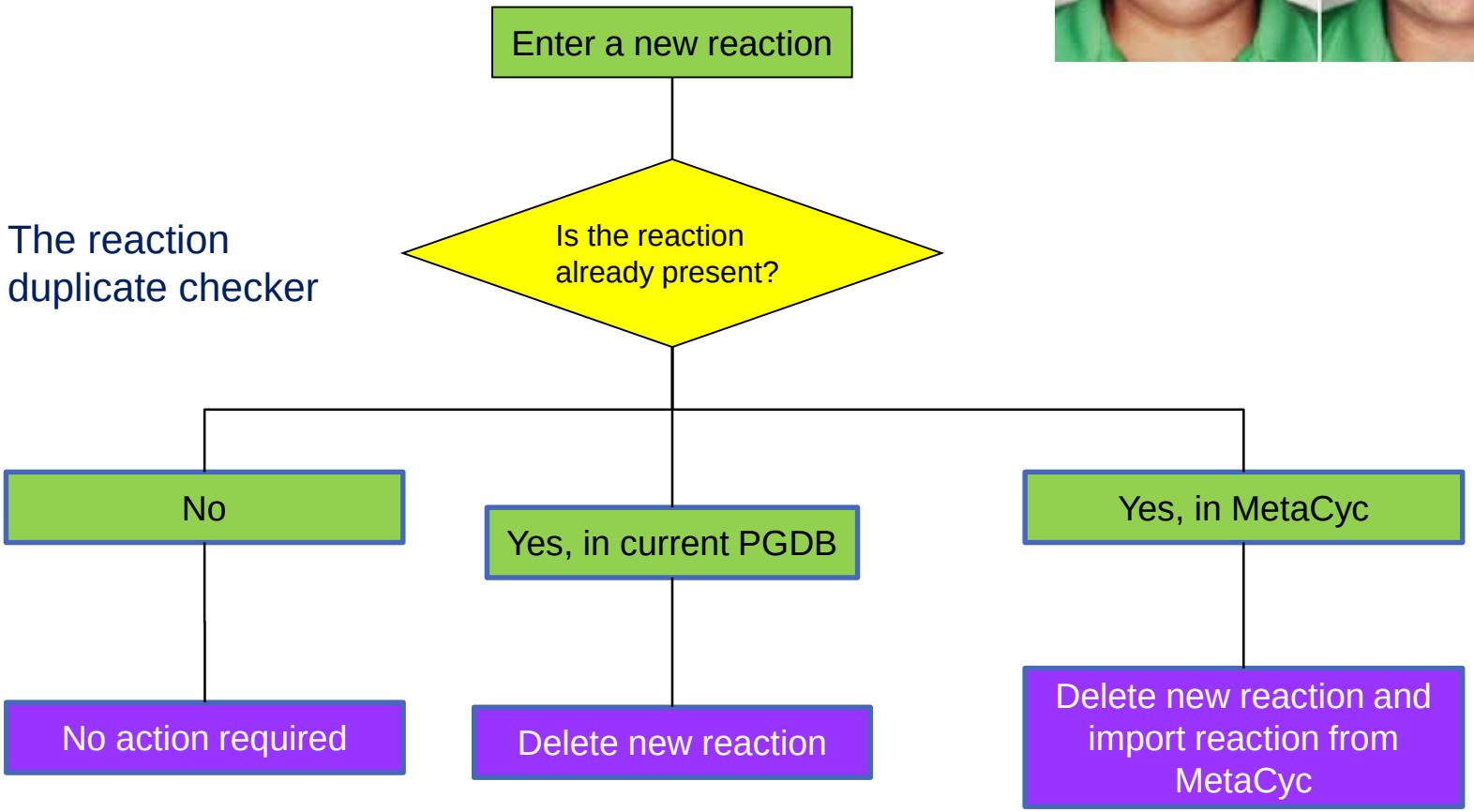
by email to

ptools-support@ai.sri.com



The Reaction Duplicate Checker

The reaction duplicate checker

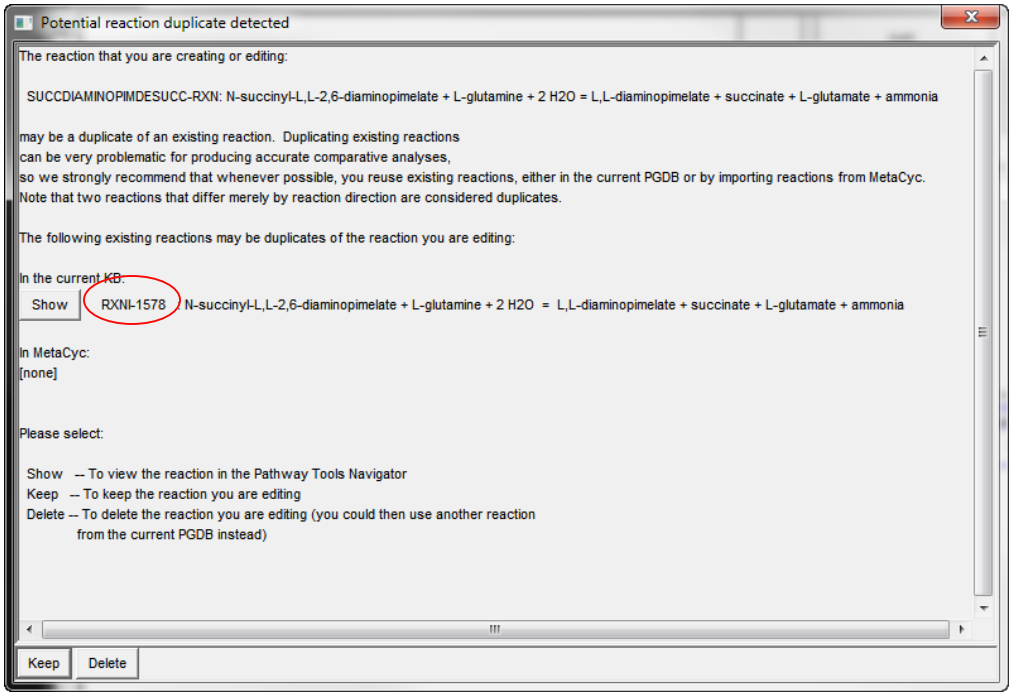


Reuse existing reactions instead of creating duplicates!

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.

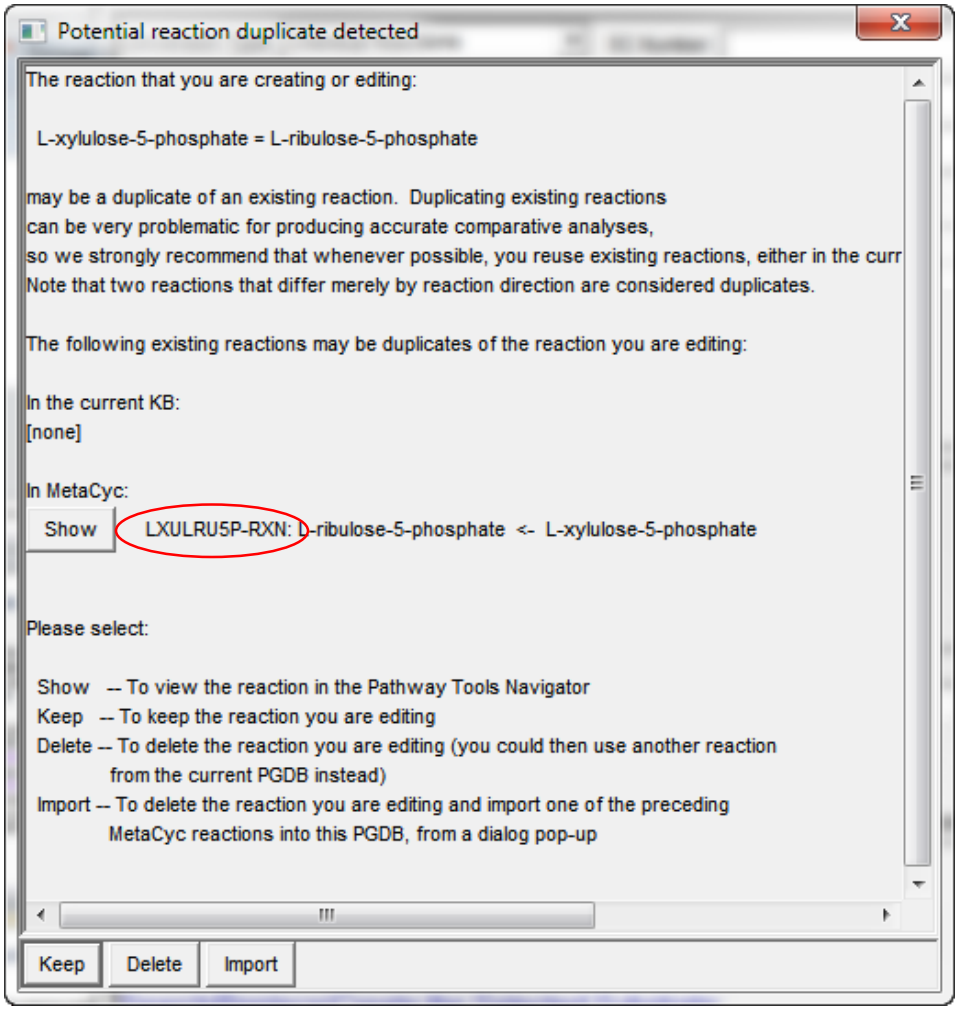


Import MetaCyc reactions instead of creating duplicates!

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.



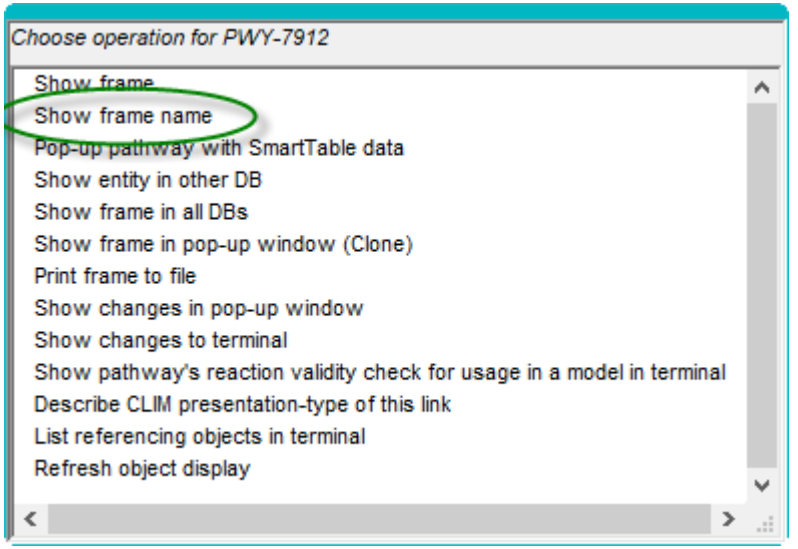
How to get a frame ID?

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 "="

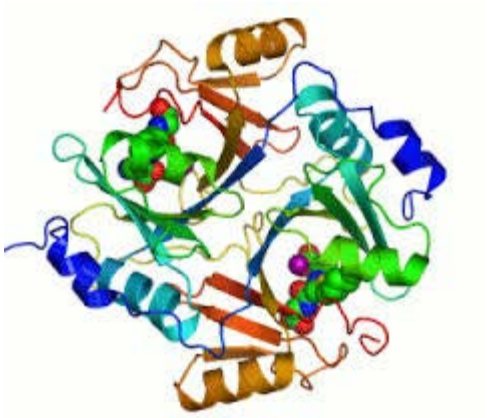
To show a frame ID of an object, you can right-click on it and select Show →Show frame name. The frame ID will be printed in the Lisp console, from which you can copy it.

Editing Pathway/Genome Databases – Lab Section

Exercise 1: Creating reactions and pathways

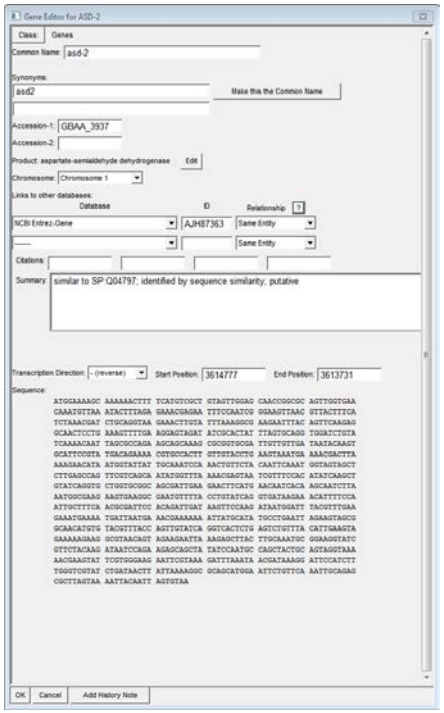
- Create organization and curator frames, specify your Pathway Tools username
- Enter new reactions
- Import reactions from MetaCyc
- Construct a new pathway with these reactions

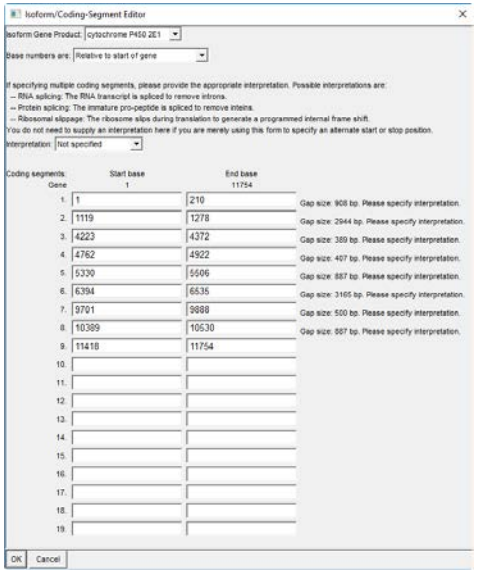
Curating Genes and Enzymes

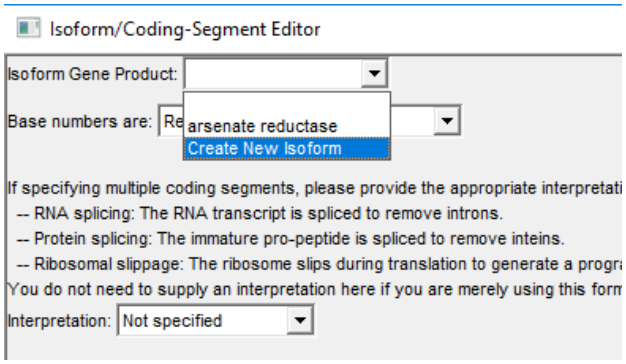


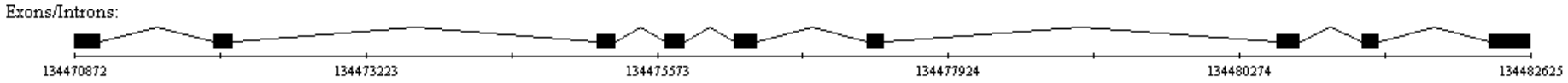
The Gene and Isoform/Coding-Segment Editors

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









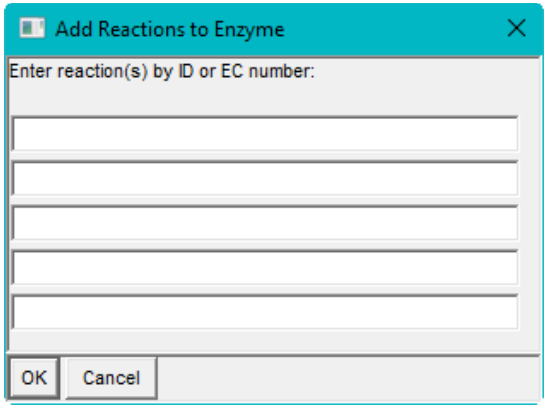
Adding an Enzyme to a Reaction

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.

Or

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

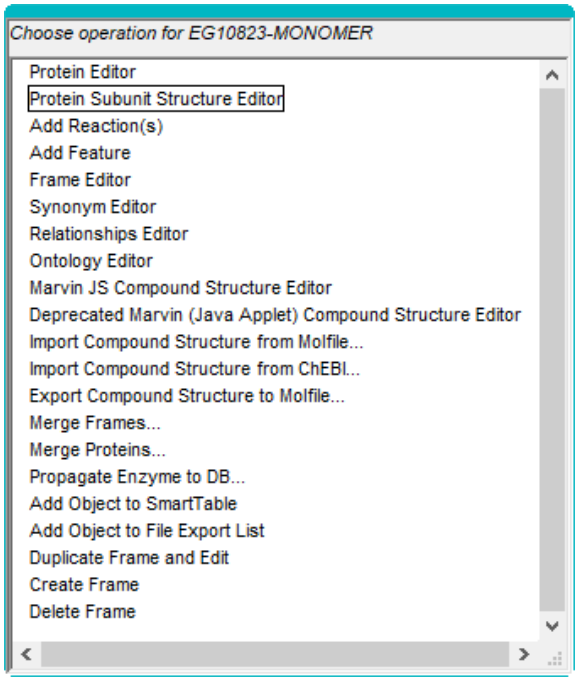


Creating Protein Complexes

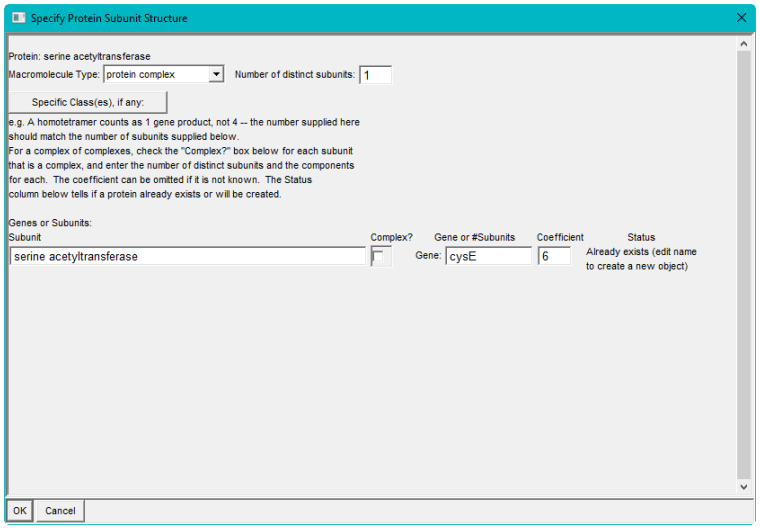
Right-click on a protein and select

Edit → Protein Subunit Structure Editor

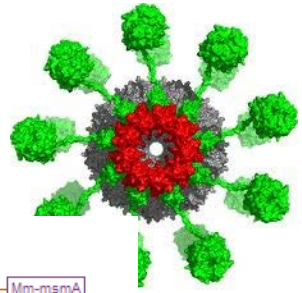
Change “Macromolecule Type” from polypeptide to protein complex

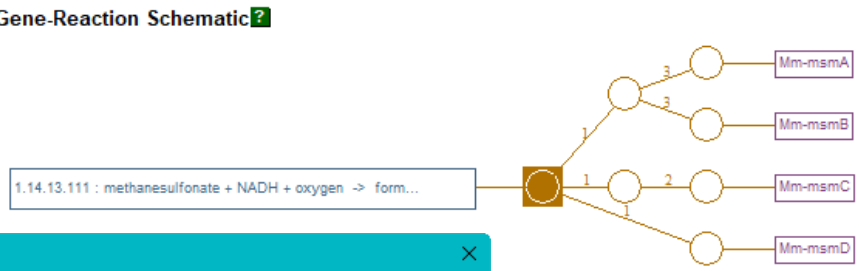


Example: a simple homohexamer



Specifying Multiple Subunits





Specify Protein Subunit Structure

Protein: methanesulfonate monooxygenase

Species: Methylosulfonomonas methylovora Select

Macromolecule Type: protein complex Number of distinct subunits: 3

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.

Genes or Subunits:

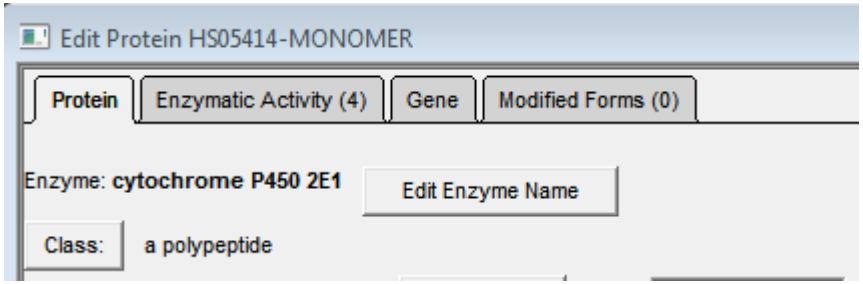
Subunit	Complex?	Gene or #Subunits	Coefficient	Status
methanesulfonate monooxygenase hydroxylase component	☑	#Subunits: 2	1	Already exists (edit name to create a new object)
methanesulfonate monooxygenase hydroxylase componen	☐	Gene: msmA	3	Already exists (edit name to create a new object)
methanesulfonate monooxygenase hydroxylase componen	☐	Gene: msmB	3	Already exists (edit name to create a new object)
methanesulfonate monooxygenase ferredoxin component	☑	#Subunits: 1	1	Already exists (edit name to create a new object)
methanesulfonate monooxygenase ferredoxin component s	☐	Gene: msmC	2	Already exists (edit name to create a new object)
methanesulfonate monooxygenase reductase component	☐	Gene: msmD	1	Already exists (edit name to create a new object)

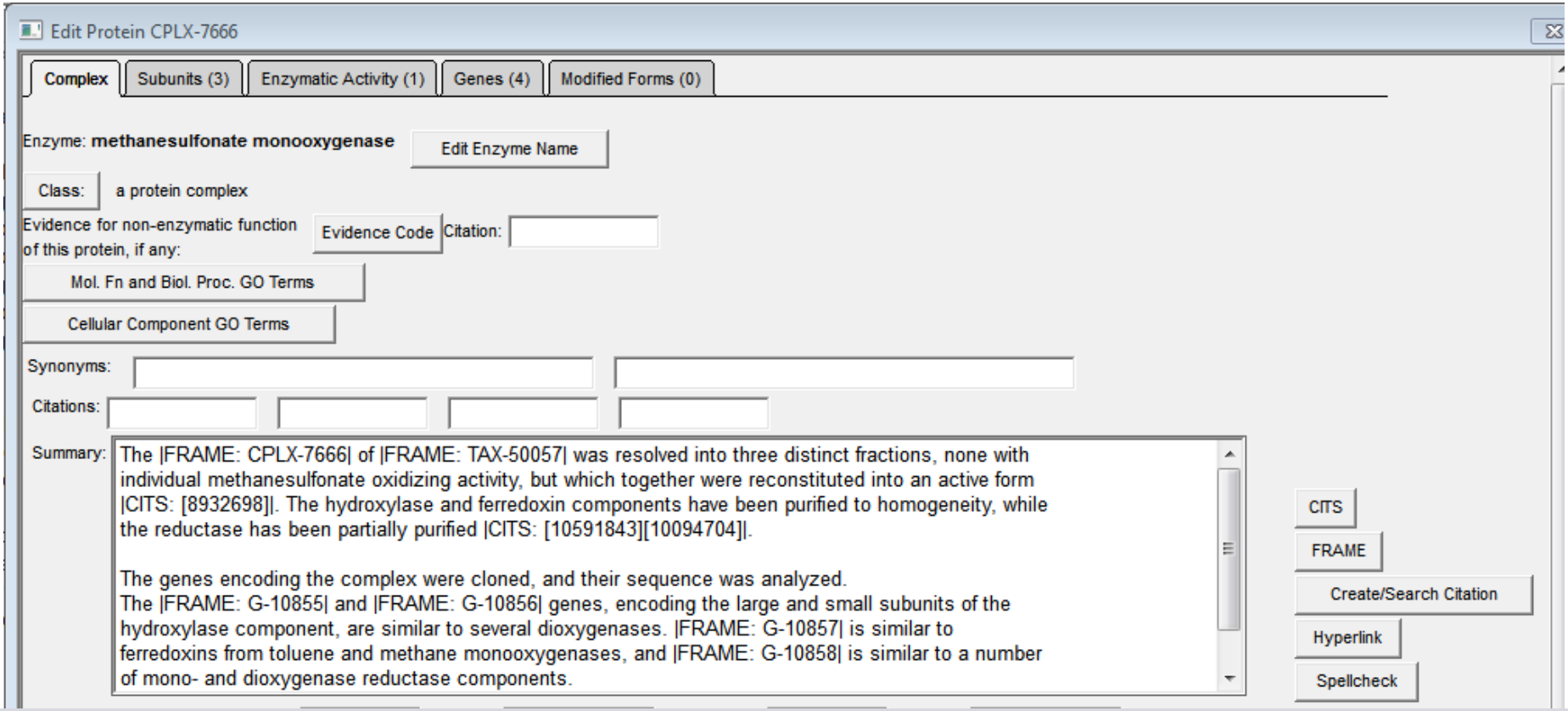
OK Cancel

The Protein Editor

(monomer edition)

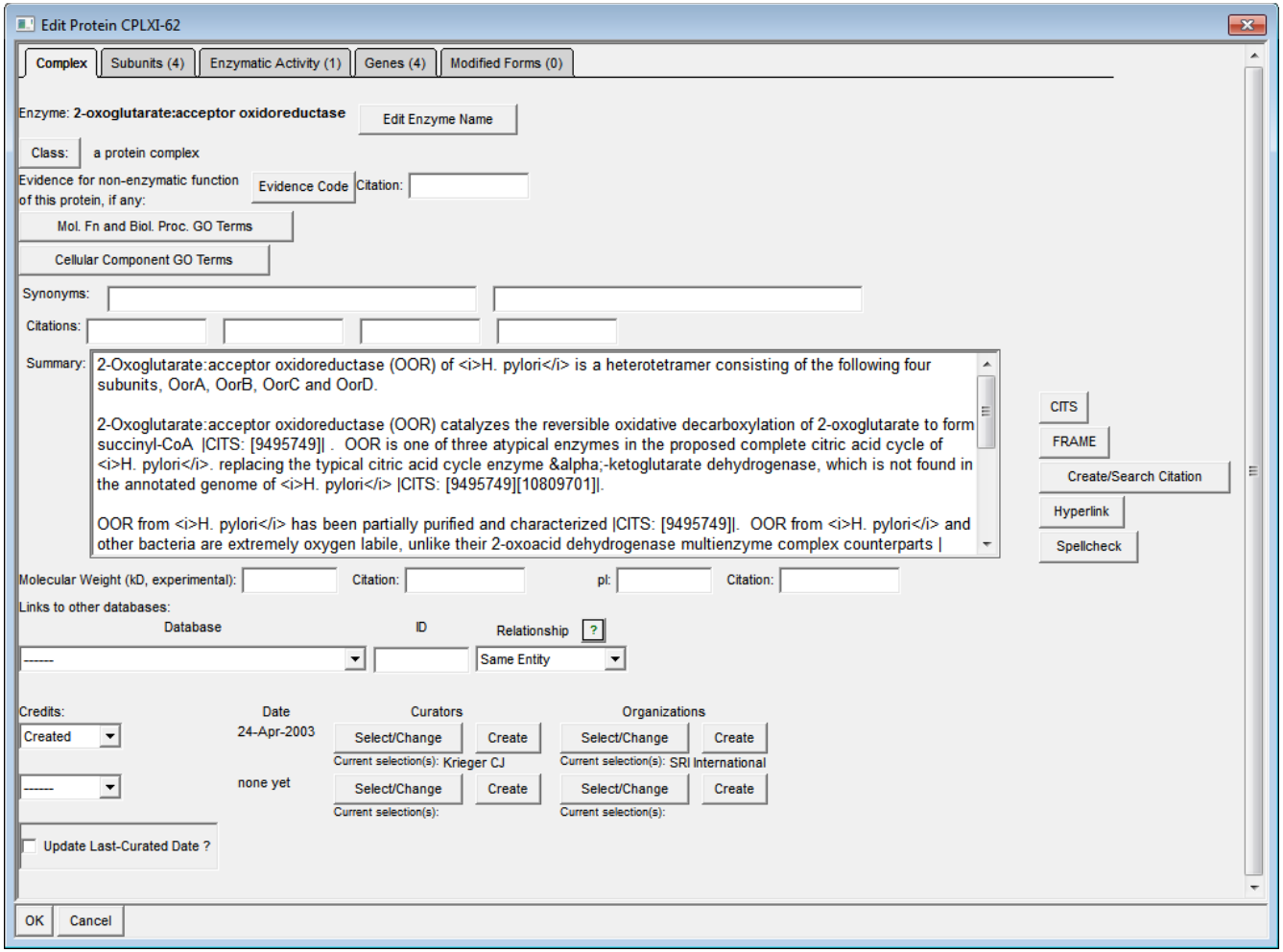
(protein complex edition)





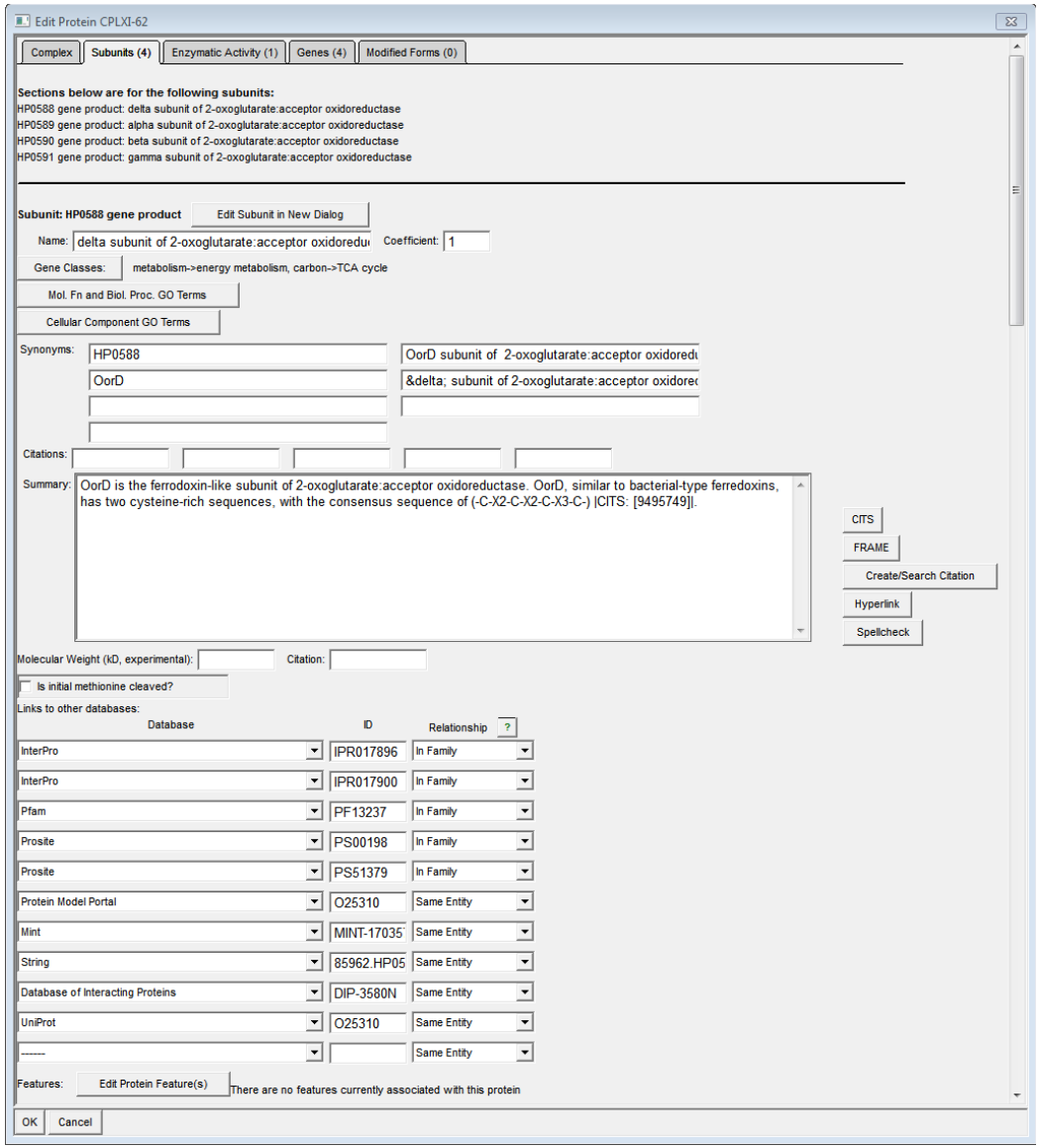
Protein Editor – First Tab

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

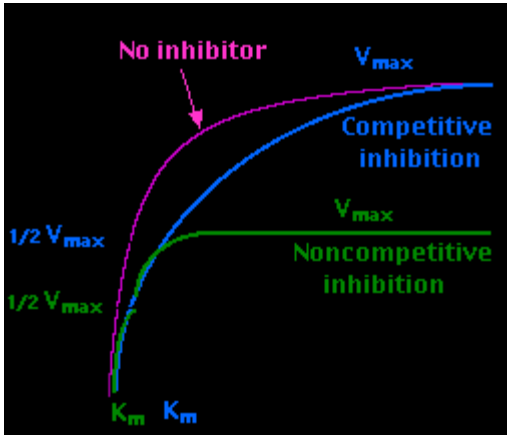


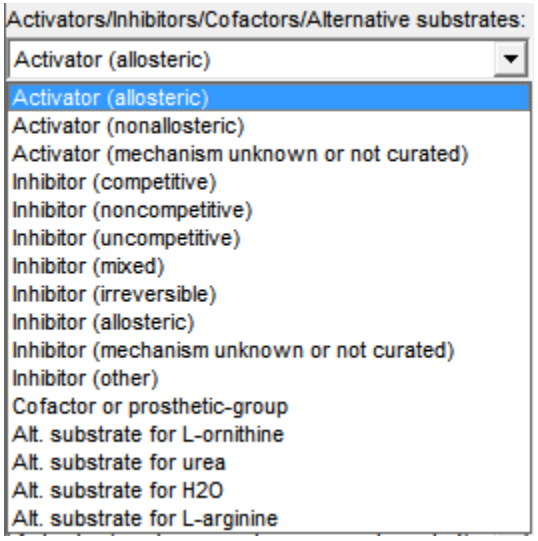
Protein Editor - Subunits Tab

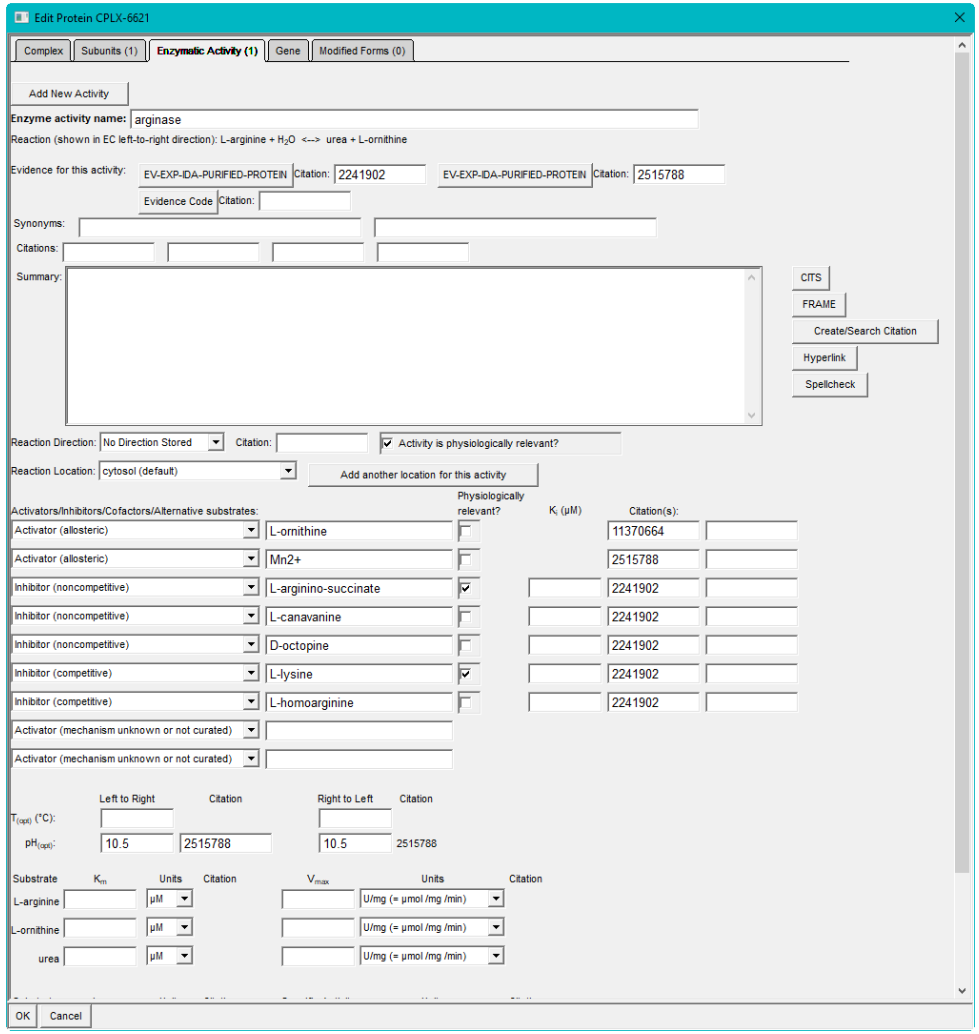
- Edit the copy number of each subunit
- Specify UniProt ID or links to other databases
- Specify experimental MW and any useful info that may apply



Protein Editor - Enzymatic Activity Tab





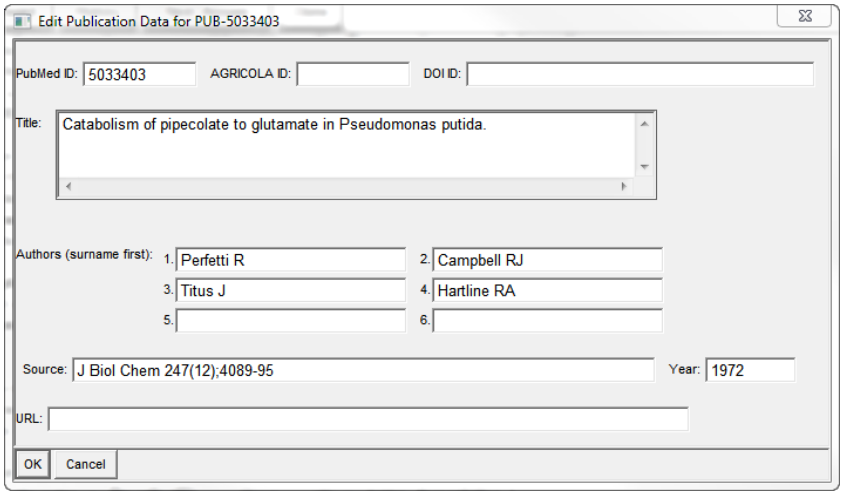


Citations

- Citation boxes
- The CITS field

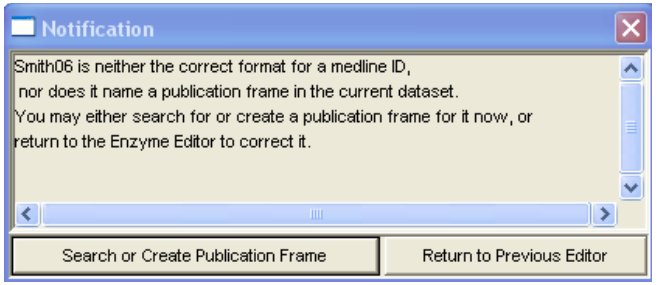
PubMed citations:

- 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.



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.