



# An Advanced Web Query Interface for Biological Databases

Peter D. Karp, Mario Latendresse  
 Bioinformatics Research Group, SRI International, Menlo Park CA USA  
[bioinformatics.ai.sri.com/ptools/](http://bioinformatics.ai.sri.com/ptools/) BioCyc.org



Peter Karp  
 SRI International  
 333 Ravenswood Ave, Menlo Park CA  
 94025  
 650-859-4358

## 1. Introduction

The **Pathway Tools** software [1] is used to create, query, curate, and analyze Pathway/Genome Databases (PGDBs). A PGDB integrates information about genomes, metabolic pathways, and regulatory networks.

Pathway Tools operations include:

- Query and update genome and pathway annotation
- Prediction of metabolic network from genome sequence
- Visual analysis of omics datasets
- Comparative analysis
- Network analysis and debugging

A large number of PGDBs are accessible on the Internet from Pathway Tools based Web sites that provide the exploration capabilities described here.

### Pathway Tools Availability

- Freely available to academics and non-profits. Licensed by 2,100 groups.
- Runs on PC/Linux, PC/Windows, Macintosh

### References:

- [1] Karp, P.D. et al. *Briefings in Bioinformatics* 2010 11:40-79.
- [2] Latendresse, M. and Karp, P.D., *Database* 2010.

Supported by grants GM75742 and GM080746 from the National Institutes of Health.

Simple initial query:

- Find Compounds whose molecular weight is between 400 and 420
- Show a result table that includes the compound name, chemical formula, and molecular weight



## 2. Structured Advanced Query Form

Writing precise and complicated queries against biological databases (DBs) usually requires programming skill

Most Web-based DB query forms are quite restricted in power of queries they can support

Pathway Tools Structured Advanced Query Form (SAQP) [2]:

- User friendly
- Supports precise and complicated queries
- Supports multiple conditions and connectives
- Supports connections between object types
- Schema driven
- Interactive documentation on DB schema

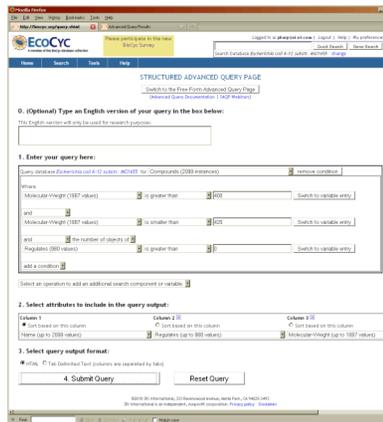
Where to find the SAQP on the Web:

- <http://bioinformatics.ai.sri.com/ptools/>
- Search -> Advanced at any Pathway Tools driven Web site

## 3. Example Queries and Their Results

Add additional condition to previous query:

- The multivalued slot **Regulates**, which lists regulation interactions that the compound regulates, must have more than 0 values



Advanced Query Results

Your query in BiVoilà is: `molwt >= 400 & molwt <= 420 & (REGULATES > 0)`

This query resulted in a single table of 41 rows.

Head #	Regulates #	Substrate	MW
1	1	1,4-Diphosphoglycerate	400.00
2	1	2-Oxoglutarate	400.00
3	1	3-Phosphoglycerate	400.00
4	1	4-Phosphoglycerate	400.00
5	1	5-Phosphoglycerate	400.00
6	1	6-Phosphoglycerate	400.00
7	1	7-Phosphoglycerate	400.00
8	1	8-Phosphoglycerate	400.00
9	1	9-Phosphoglycerate	400.00
10	1	10-Phosphoglycerate	400.00
11	1	11-Phosphoglycerate	400.00
12	1	12-Phosphoglycerate	400.00
13	1	13-Phosphoglycerate	400.00
14	1	14-Phosphoglycerate	400.00
15	1	15-Phosphoglycerate	400.00
16	1	16-Phosphoglycerate	400.00
17	1	17-Phosphoglycerate	400.00
18	1	18-Phosphoglycerate	400.00
19	1	19-Phosphoglycerate	400.00
20	1	20-Phosphoglycerate	400.00
21	1	21-Phosphoglycerate	400.00
22	1	22-Phosphoglycerate	400.00
23	1	23-Phosphoglycerate	400.00
24	1	24-Phosphoglycerate	400.00
25	1	25-Phosphoglycerate	400.00
26	1	26-Phosphoglycerate	400.00
27	1	27-Phosphoglycerate	400.00
28	1	28-Phosphoglycerate	400.00
29	1	29-Phosphoglycerate	400.00
30	1	30-Phosphoglycerate	400.00
31	1	31-Phosphoglycerate	400.00
32	1	32-Phosphoglycerate	400.00
33	1	33-Phosphoglycerate	400.00
34	1	34-Phosphoglycerate	400.00
35	1	35-Phosphoglycerate	400.00
36	1	36-Phosphoglycerate	400.00
37	1	37-Phosphoglycerate	400.00
38	1	38-Phosphoglycerate	400.00
39	1	39-Phosphoglycerate	400.00
40	1	40-Phosphoglycerate	400.00
41	1	41-Phosphoglycerate	400.00

## 4. Pathway/Genome Databases on the Internet

Available PGDBs include the following. Web links to all available from <http://BioCyc.org/otherpgdbs.shtml> or find them using Google

### Microbial

- BioCyc – 670 microbial PGDBs including EcoCyc (*E. coli* K-12)
- BeoCyc – 33 Bioenergy organisms
- ScoCyc – *Streptomyces coelicolor*
- TBCCyc – *Mycobacterium tuberculosis*
- PseudoCyc – *Pseudomonas aeruginosa*
- PATRIC – Bacterial pathogens
- ApiCyc – 8 apicomplexan PGDBs

### Fungal

- YeastCyc – *Saccharomyces cerevisiae*
- CalbiCyc – *Candida albicans*

### Plant

- SolCyc – 7 Solanacea PGDBs
- ChlamyCyc – *Chlamydomonas reinhardtii*
- AraCyc – *Arabidopsis thaliana*
- RiceCyc – *Oryza sativa*
- PlantCyc – General plant PGDB

### Mammalian

- MouseCyc, HumanCyc, CattleCyc

Add two additional conditions to previous query:

- The regulation interaction must have a parent class whose name includes the string "enzym", indicating that it involves regulation of enzyme activity
- The mode of regulation must be positive (activation)



Advanced Query Results

Your query in BiVoilà is: `molwt >= 400 & molwt <= 420 & (REGULATES > 0) & (PARENT_CLASS >= 'enzym') & (MODE >= 'activation')`

This query resulted in 4 single tables of 4 rows.

Head #	Regulates #	Substrate	MW
1	1	1,4-Diphosphoglycerate	400.00
2	1	2-Oxoglutarate	400.00
3	1	3-Phosphoglycerate	400.00
4	1	4-Phosphoglycerate	400.00