

# Computing Metabolic Routes in the Human Microbiome



Peter D. Karp and Markus Krümmenacker

Bioinformatics Research Group, SRI International, Menlo Park CA USA

BioCyc.org

bioinformatics.ai.sri.com/ptools/



## 1. BioCyc Microbial Genome Portal

Genomes plus metabolic reconstructions for 14,000 prokaryotes at BioCyc.org

- 2,400 organisms from Human Microbiome Project

Most databases derived through computational prediction of the metabolic pathways of each organism from genome sequence.

Some databases received manual curation; information on pathways, enzymes, reactions, and metabolites manually entered from experimental literature.

### The MetaCyc Database (version 22.5)

- 2,600 metabolic pathways curated from 2,900 organisms
- 15,200 reactions, 15,090 metabolites

MetaCyc provides extensive data fields including:

- Mini-reviews equivalent to 8,600 textbook pages summarize experimental findings for enzymes and pathways
- Extensive literature references
- Cofactors, activators, inhibitors, alternative substrates
- Kinetic data, reaction atom mappings

### References

[1] Caspi, R., et al., "The MetaCyc database of metabolic pathways and enzymes," *Nucleic Acids Res.* 46:D633-9 2018.

Supported by grants GM092729, GM75742, and GM080746 from the National Institutes of Health.

## 2. Pathway Tools Software

Pathway Tools capabilities include:

### 1. Provide full-featured pathway/genome informatics portal

- Genome browser, sequence search/alignment/extraction
- Pathway viewer, pathway collages, zoomable metabolic maps
- Extensive search tools

### 2. Computationally infer new information for a sequenced genome or metagenome

- Infer metabolic pathways
- Infer pathway hole fillers (genes coding for missing enzymes in predicted pathways)
- Infer operons

### 3. Creation of metabolic-flux models with MetaFlux

- Flux-balance analysis (FBA) models
- Software suggests model refinements, greatly accelerating rate at which FBA models can be constructed

### 5. Analyze metabolomics and gene expression datasets

- Paint high-throughput datasets onto metabolic maps and pathway diagrams
- Omics Dashboard tool
- Enrichment analysis of metabolite sets and gene-expression data

### Pathway Tools Availability

- Freely available to academics and non-profits. Licensed by 9,000 groups.
- Runs on Linux, Windows, Macintosh

### References

[1] Karp, P.D. et al, *Briefings in Bioinformatics* 2015.

## 3. Metabolic Route Searching

**The problem:** How does a microbial community convert metabolite X to metabolite Y?

### Single-Organism Metabolic Route Search:

- Find routes from source to target compound in single-organism metabolic network
  - A route is a linear sequence of reactions
- Application 1: Route finding in existing metabolic networks
- Application 2: Design of engineered metabolic networks (add reactions from MetaCyc)
- Publication: Latendresse et al, *Bioinformatics* 30:2043 2014

Inputs specified by user:

- BioCyc organism whose network is to be searched
- Maximum route length
- Cost of adding in-organism reactions to the route
- Cost of adding MetaCyc reactions to the route
- Cost of each atom lost along the route

Approach:

- Branch-and-bound algorithm finds optimal routes
- Optimality criterion: Minimize route cost  $R_c C_c + R_m C_m + k C_a$

Where

- $R_c$  is the number of in-organism reactions in the route and  $C_c$  is the cost of an in-organism reaction
- $R_m$  is the number of MetaCyc reactions in the route and  $C_m$  is the cost of a MetaCyc reaction
- $k$  is the number of atoms lost along the route and  $C_a$  is the cost of losing an atom

Results:

- Implemented as RouteSearch tool within Pathway Tools
  - BioCyc web site: Metabolism -> Metabolic Route Search
- RouteSearch finds solutions to known metabolic engineering problems among its top-scored solutions
- Solutions presented via a graphical interface that speeds user understanding

### Multi-Organism Metabolic Route Search (MORS):

- Find efficient routes from source to target compound via metabolism of multiple organisms

Inputs specified by user:

- Same inputs as above plus:
- Set of BioCyc organisms whose networks are to be searched

Approach:

- Compute union of all metabolic reactions across all designated organisms
- Branch-and-bound algorithm finds optimal routes across multi-organism network
- Additional constraint: minimize *organism switches*
- Ignore transport considerations

### Example routes in gut microbiome plus *Homo sapiens*

- L-tryptophan to indoxyl sulfate – MORS finds known route
  - First reaction: 162 microbes Last two reactions: *Homo sapiens*



- L-carnitine to TMAO – MORS finds known route
  - First reaction: 78 microbes Last reaction: *Homo sapiens*

