Metabolic pathway analysis

Examples of metabolic pathways

Methionine Biosynthesis in S. cerevisiae

Methionine Biosynthesis in E. coli

Alternative methionine pathways

KEGG "consensus pathway" for Methionine metabolism
Lysine biosynthesis in Escherichia coli

1. L-aspartyl-4-P
2. L-Aspartate dihydropicolinic acid
3. Tetrahydrodipicolinate
4. N-succinyl-epsilon-keto-L-alpha-aminopimelic acid
5. Succinyl diaminopimelate
6. LL-diaminopimelic acid
7. Mesodiaminopimelic acid

8. DapF

9. L-aspartic semialdehyde tetrahydrodipicolinate N-succinyltransferase
10. DapD
11. Aspartate kinase III
12. DapC
13. Succinyl diaminopimelate aminotransferase
14. DapE
15. N-succinyldiaminopimelate desuccinylase
16. DapB

17. Aspartate semialdehyde dehydrogenase
18. Asd

19. Pyruvate
20. NADP+ or NAD+
21. NADPH or NADH; H+
22. Succinyl CoA
23. ADP
24. ATP
25. Alpha-ketoglutarate
26. Glutamate
27. Succinate
28. H2O
29. NADP+; Pi
30. NADPH; H+

Lysine biosynthesis in Saccharomyces cerevisiae

1. DapA

2. Dihydrodipicolinate synthase
3. LysA

4. Diaminopimelate decarboxylase

5. Lysine biosynthesis in KEGG (yeast enzymes in green)

6. EcoCyc example - proline utilization

7. EcoCyc example - proline biosynthesis

8. Ecocyc - metabolic overview
KEGG example: proline and arginine metabolism (E.coli)

- Where is proline?
- How is proline synthesized in E.coli?
- How is proline catabolized in E.coli?
- Is it obvious that reactions 1.5.99.8 and 1.5.1.2 have distinct side reactants?

Pathway reconstruction by reaction clustering

Jacques van Helden
jvanheld@ucmb.ulb.ac.be

Graph-based analysis of biochemical networks

A graph of compounds and reactions

Reactions from KEGG
- 10,166 compounds
  (only 4302 used by one reaction)
- 5,283 reactions
- 10,685 substrate → reaction (7,297 non-trivial)
- 10,621 reaction → product (6,828 non-trivial)

Metabolic Pathways as subgraphs

Reconstructing a pathway from a subset of reactions

Input:
- a set of reactions (the seed reactions)

Output:
- a metabolic pathway including
  - the seed reactions, together with their substrates and products
  - optionally, some additional reactions, interalated to improve the pathway connectivity
  - the pathway can either be connected, or contain several unconnected components

Seed nodes

Escherichia coli
- 4219 Genes (Blattner)
- 967 enzymes (Swissprot)
- 159 pathways (EcoCyc)
**Linking seed nodes**

- Compound
- Reaction
- Seed Reaction
- Direct link

**Enhance linking by intercalating reactions**

- Compound
- Reaction
- Seed Reaction
- Direct link
- Intercalated reaction

**Subgraph extraction**

**Validation of the method**

- Take a set of experimentally characterized pathways, and for each one
  - Select a subset of enzymes
  - Use the reactions catalysed by these enzymes as seed nodes
  - Extract the subgraph
  - Compare with known pathway

**Example: reconstitution of lysine pathway**

- Gap size: 0
  - all Ecs from original pathway are provided as seeds
- Seeds
  - 1.2.1.11
  - 2.3.1.117
  - 2.6.1.17
  - 2.7.2.4
  - 3.5.1.18
  - 4.1.1.20
  - 4.2.1.52
  - 5.1.1.7
- Result:
  - Inferring reaction orientation (reverse or forward)
  - Ordering
Example: reconstitution of lysine pathway

- Gap size: 1
- 5 seed reactions
- Result
  - Inferring missing steps
  - Inferring reaction orientation
  - Ordering

- Gap size: 2
- 4 seed reactions
- Result
  - E.coli pathway found
  - Alternative pathways also returned

- Gap size: 3
- 3 seed reactions
- Result
  - E.coli pathway is not found, because the program finds shortcuts between the seed reactions

Applications of pathway reconstruction

- We have the complete genome for dozens of bacteria, for which there is almost no experimental characterization of metabolism
- For these genomes, enzymes have been predicted by sequence similarity
- In some cases, one expects to find the same pathways as in model organisms, but in other cases, variants or completely distinct pathways
- For each known pathway from model organisms
  - Select the subset of reactions for which an enzyme exists in the target organism
  - If a reasonable number of reactions are present
    - Using these as seeds, reconstruct a pathway
    - Preferentially (but not exclusively) intercalate reactions for which an enzyme has been detected in the target organism

Graph-based analysis of biochemical networks

From gene expression data to pathways

Jacques van Helden
jvanheld@uclm.ulb.ac.be

Reaction clustering and gene expression data

- Many biochemical pathways are co-regulated at the transcriptional level.
- Starting from the observation that a group of genes is co-regulated, try to find if they could be involved in a common pathway.
### Study case: cluster of co-regulated genes

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>YFG1030W</td>
<td>mei10</td>
</tr>
<tr>
<td>YFG1209W</td>
<td>mei13</td>
</tr>
<tr>
<td>YGL001C</td>
<td>mei14</td>
</tr>
<tr>
<td>YHR131C</td>
<td>mei16</td>
</tr>
<tr>
<td>YLR153W</td>
<td>MET7</td>
</tr>
<tr>
<td>YNL075W</td>
<td>mei5</td>
</tr>
<tr>
<td>YOR031C</td>
<td>mei6</td>
</tr>
</tbody>
</table>

### Gene expression data: cell cycle

**KEGG - reaction coloring in pathway maps**

**KEGG - reaction coloring in pathway maps**

**KEGG - reaction coloring in pathway maps**

**KEGG - reaction coloring in pathway maps**

**KEGG - gene search in pathway maps**

**KEGG - reaction coloring in pathway maps**

**KEGG - reaction coloring in pathway maps**

**KEGG - reaction coloring in pathway maps**

**KEGG - reaction coloring in pathway maps**

**Building pathways from gene clusters**

- **Gene cluster**
  - Identify genes coding for enzymes
  - Identify subset of catalyzed reactions
  - Interconnect these reactions to find all possible pathways

- **Pathway reconstruction**
  - Potential pathway

- **Gene expression profiler**
  - Cluster of co-regulated genes

**Pathway found in Spellman’s “MET” cluster**

**Analysis of Gene Expression Data**

- **Gene cluster**
  - 20 genes
  - 7 enzymes
  - 8 reactions

- **Pathway Diagram**
  - Known Pathways: 2 matching pathways

**Comparison with Sulfur assimilation**

**Comparison with methionine biosynthesis**

**Summary**

- Starting from an unordered cluster of genes, one gets an ordered set of reactions, connected to form a pathway.
- Should permit discovery of novel pathways, that are not stored in any pathway database yet.
- Interpretation of intercalated reactions
  - enzyme is not regulated
  - DNA chip defect for that gene
  - gene was not on the DNA chip
  - enzyme remains to be identified in that organism
Analysis of data from Gasch et al.

- 6152 yeast genes
- Various conditions of stress (heat shock, osmotic shock, peroxide, amino acid starvation, nitrogen depletion
- Steady-state growth on alternative carbon sources
- Overexpression studies

Selected experiments

Represse by mannose (at least 3-fold)

Induced by galactose (at least 2-fold)

Represse by glucose (at least 2-fold)