Genome-scale Metabolic Reconstructions
LEARNING OBJECTIVES

Each student should be able to:

• Explain the process of creating a genome-scale metabolic reconstruction
GENOME-SCALE METABOLIC RECONSTRUCTIONS

- Overview
- Draft Reconstruction
- Refinement of Reconstruction
- Conversion of Reconstruction into Computable Format
- Network Evaluation
- Data Assembly and Dissemination

Phylogenetic Coverage of Genome-scale Network Reconstructions

A GEnome scale Network Reconstructions (GENREs) serves as a structured knowledge base of established biochemical facts, while a GEnome scale Models (GEMs) is a model which supplements the established biochemical information with additional (potentially hypothetical) information to enable computational simulation and analysis.

Expansion of Metabolic Networks and Global Reactome Coverage Over Time
Reconstruction Process: 96 Step Protocol

GENOME-SCALE METABOLIC RECONSTRUCTIONS

- Overview
  - Draft Reconstruction
  - Refinement of Reconstruction
  - Conversion of Reconstruction into Computable Format
  - Network Evaluation
  - Data Assembly and Dissemination

Draft Reconstruction

1. Obtain genome annotation
2. Identify candidate metabolic functions
3. Obtain candidate metabolic reactions
4. Assembly of draft reconstruction
5. Collect experimental data
## Genome Databases

<table>
<thead>
<tr>
<th>Name</th>
<th>Link</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Comprehensive Microbial Resource (CMR)</td>
<td><a href="http://cmr.jcvi.org/cgi-bin/CMR/CmrHomePage.cgi">http://cmr.jcvi.org/cgi-bin/CMR/CmrHomePage.cgi</a></td>
<td></td>
</tr>
<tr>
<td>Genomes OnLine Database (GOLD)</td>
<td><a href="http://www.genomesonline.org/">http://www.genomesonline.org/</a></td>
<td></td>
</tr>
<tr>
<td>TIGR</td>
<td><a href="http://www.tigr.org/db.shtml">http://www.tigr.org/db.shtml</a></td>
<td></td>
</tr>
<tr>
<td>SEED database32</td>
<td>theseed.uchicago.edu/FIG/index.cgi</td>
<td>Comparative genomics tool</td>
</tr>
</tbody>
</table>

Draft Reconstruction:
Obtain Genome Annotation

1. Automatic Annotation of Genome Sequences
   b. MetaSHARK - http://bioinformatics.leeds.ac.uk/shark/

2. Existing Databases:
   a. TIGR-CMR Comprehensive Microbial Resource
      http://cmr.jcvi.org/tigr-scripts/CMR/CmrHomePage.cgi
   b. National Center for Biotechnology Information (NCBI)
   c. EcoCyc - http://ecocyc.org
   d. Vega - http://vega.sanger.ac.uk/index.html

3. The following information should be retrieved for each gene: genome position, coding region, strand, locus name, alias, gene function, protein classification (Enzyme Commission (E.C.) number).
Network Reconstruction

Objective:
Create a biochemically, genetically and genomically (BiGG) structured knowledge base

Reconstruction and Use of Microbial Metabolic Networks: the Core Escherichia coli Metabolic Model as an Educational Guide by Orth, Fleming, and Palsson (2010)
Draft Reconstruction

1. Obtain genome annotation
2. Identify candidate metabolic functions
3. Obtain candidate metabolic reactions
4. Assembly of draft reconstruction
5. Collect experimental data
### Constraint-based Metabolic Reconstructions & Analysis

#### Organism Name
- **Gene Symbol**
- **Gene Function**
- **Gene Locus**
- **Gene Information**

#### Results: 1 to 20 of 1230

1. **Gene**: galU
   - **Function**: glucose-1-phosphate dehydrogenase
   - **Information**: Escherichia coli K-12 sub. MG1655
   - **Locus**: galU

2. **Gene**: cdpA
   - **Function**: cysteine desulfurase
   - **Information**: Escherichia coli K-12 sub. MG1655
   - **Locus**: cdpA

#### No transcriptional regulators

Only genes with *metab* in description

Draft Reconstruction

1. Obtain genome annotation
2. Identify candidate metabolic functions
3. Obtain candidate metabolic reactions
4. Assembly of draft reconstruction
5. Collect experimental data
## Desired Reaction Information

1. Reaction Name*
2. Reaction Description*
3. Reaction Formula*
4. Gene-reaction Association*
5. Genes (Gene Locus)*
6. Proteins
7. Cellular Subsystem*  
   (e.g. Glycolysis)
8. Reaction Direction*
9. Flux Lower Bound*
10. Flux Upper Bound*
11. Confidence Score (1-5)
12. EC Number
13. Notes
14. References

* Required

Reconstruction and Use of Microbial Metabolic Networks: the Core Escherichia coli Metabolic Model as an Educational Guide by Orth, Fleming, and Palsson (2010)
List Of Standards That Have Been Used In Numerous Metabolic Reconstructions

- **Naming Conventions**
  - Reaction abbreviations are capitalized.
  - Use reaction names suffix standards (See next slide)
  - Try to construct the root of the reaction abbreviation based on the enzyme name, for example AKGDHe = Alpha-ketoglutarate Dehydrogenase (in the extracellular compartment).
  - Metabolites are lower case.
  - Metabolite formulas in the charged state are based on the chemical structure at a pH of 7.2. The charge state can be defined using tools (such as pKaDB).
  - Do not use wildcard characters in abbreviations: no apostrophes, no parentheses, etc. The exceptions to this are the use of parentheses in sink and demand reactions.

- **Notes Fields (reactions and compounds):**
  - Add references whenever possible (e.g. PMID, KEGG ID, PubChem ID, PubSubstance ID), if these identifiers are not available, make sure to state this explicitly.
  - Add any detailed descriptions necessary to understand any specific rationale for the addition.
  - Reactions must always be charge balanced. If not balanced, state why.
  - Always add your full name or the initials to the note field. This increases traceability.

# Reaction Names Suffix Standards

<table>
<thead>
<tr>
<th>Reaction Type</th>
<th>Suffix</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABC transporter</td>
<td>-abc</td>
<td>ALAabc</td>
</tr>
<tr>
<td>Transport reactions</td>
<td>-t</td>
<td>GLCT1</td>
</tr>
<tr>
<td>Reversible reactions</td>
<td>-r</td>
<td>GLCT1r</td>
</tr>
<tr>
<td>Irreversible reactions</td>
<td>-i</td>
<td>PTRCT3i</td>
</tr>
<tr>
<td>Dehydrogenase reactions</td>
<td>-DH</td>
<td>PDH</td>
</tr>
<tr>
<td>Synthetase reactions</td>
<td>-S</td>
<td>ATPS</td>
</tr>
<tr>
<td>Kinase reactions</td>
<td>-K</td>
<td>ACKr</td>
</tr>
<tr>
<td>Chloroplast reactions</td>
<td>-h</td>
<td>HEX1h</td>
</tr>
<tr>
<td>Endoplasmic Reticular reactions</td>
<td>-er</td>
<td>CERASE124er</td>
</tr>
<tr>
<td>Extracellular reactions</td>
<td>-e</td>
<td>AKGDHe</td>
</tr>
<tr>
<td>Golgi reactions</td>
<td>-g</td>
<td>S6T12g</td>
</tr>
<tr>
<td>Lysosomal reactions</td>
<td>-l</td>
<td>10FTHFtl</td>
</tr>
<tr>
<td>Mitochondrial reactions</td>
<td>-m</td>
<td>AKGDm</td>
</tr>
<tr>
<td>Nucleus reactions</td>
<td>-n</td>
<td>UMPK3n</td>
</tr>
<tr>
<td>Peroxisomal reactions</td>
<td>-x</td>
<td>SCP3x</td>
</tr>
<tr>
<td>Periplasmic reactions</td>
<td>-pp</td>
<td>PPTHpp</td>
</tr>
<tr>
<td>Vacuole</td>
<td>-v</td>
<td>GLCGS5Dv</td>
</tr>
</tbody>
</table>

**Escherichia coli K-12 substr. MG1655** Enzyme: UTP–glucose-1-phosphate uridylyltransferase

**Summary:**
UTP-glucose-1-phosphate uridylyltransferase (GalU) carries out a key step in the generation of UDP-D-glucuronate as part of the larger system of colonic acid building blocks biosynthesis.

GalU catalyzes the addition of UTP to α-D-glucose 1-phosphate to yield UDP-D-glucose [Welshborn94].

GalU is predicted to be part of a GalU/GalF complex based on research done in the uropathogenic *E. coli* strain VW187 (O7:K1), where predicted uridylyltransferase subunit with GalU was shown to interact physically and functionally with GalU [Marolda96].

*galU* mutants are unable to utilize galactose as the sole source of carbon [Fukayama62, Sundararaj62].

http://biocyc.org/ecoli/new-image?object=E611319
Constraint-based Metabolic Reconstructions & Analysis

Escherichia coli K-12 subsp. MG1655 Enzyme: UTP-glucose-1-phosphate uridylyltransferase

Enzymatic reaction of: UTP-glucose-1-phosphate uridylyltransferase

Synonyms: UDPG synthetase, UDP-glucose pyrophosphorylase

\[ \text{g-D-glucose 1-phosphate + UTP + H}^+ \rightleftharpoons \text{UDP-g-glucose + diphosphate} \]

The reaction direction shown, that is, \( A + B \rightleftharpoons C + D \) versus \( C + D \rightleftharpoons A + B \), is in accordance with the Enzyme Commission system.

The reaction is favored in the direction shown.

In Pathways: colanic acid building blocks biosynthesis, galactose degradation I (Leloir pathway)

Summary:

\textit{In vitro}, the reaction is reversible with an equilibrium constant of 5.0, i.e. degradation of UDP-glucose is favored. However, the concentration of pyrophosphate \textit{in vivo} is low, and thus the reaction likely proceeds in the direction of UDP-glucose formation.
Desired Metabolite Information

1. Metabolite Name*
2. Metabolite Description*
3. Metabolite Neutral Formula
4. Metabolite Charged Formula*
5. Metabolite Charge*
6. Metabolite Compartment*
7. Metabolite KEGGID
8. Metabolite PubChemID
9. Metabolite CheBI ID
10. Metabolite Inchi String
11. Metabolite Smile

* Required

Draft Reconstruction

1. Obtain genome annotation
2. Identify candidate metabolic functions
3. Obtain candidate metabolic reactions
4. Assembly of draft reconstruction
5. Collect experimental data
# Draft Reconstruction: Assembly Of Reaction Spreadsheet

<table>
<thead>
<tr>
<th>Rxn name</th>
<th>Rxn description</th>
<th>Formula</th>
<th>Gene-reaction association</th>
<th>Genes</th>
<th>Proteins</th>
<th>Subsystem</th>
<th>Rev</th>
<th>LB</th>
<th>UB</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALLULPE</td>
<td>Allulose 6-phosphate epimerase</td>
<td>allul6p[c] &lt;-&gt; f6p[c]</td>
<td>(b4085)</td>
<td>b4085</td>
<td></td>
<td>Alternate Carbon Metabolism</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ALLabcpp</td>
<td>D-allose transport via ABC system (periplasm)</td>
<td>all-D[p] + atp[c] + h2o[c] -&gt; adp[c] + all-D[c] + h[c] + pi[c]</td>
<td>(b4087) and (b4086) and (b4088)</td>
<td>b4086 b4087 b4088</td>
<td></td>
<td>Transport, Inner Membrane</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ALLtex</td>
<td>Allose transport via diffusion (extracellular to periplasm)</td>
<td>all-D[e] &lt;-&gt; all-D[p]</td>
<td>(b2215) or (b0241) or (b1377) or (b0929)</td>
<td>b0241 b0929 b1377 b2215</td>
<td></td>
<td>Transport, Outer Membrane Porin</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ALPATE160pp</td>
<td>apolipoprotein N-acyltransferase (phosphatidylethanolamine, periplasm)</td>
<td>alpp[p] + pe160[p] -&gt; 2agpe160[p] + lpp[p]</td>
<td>(b1677) and (b0657)</td>
<td>b0657 b1677</td>
<td></td>
<td>Unassigned</td>
<td>0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1. The draft reconstruction includes a list of candidate genes and reactions
2. Not all of the spreadsheet cells will be filled at this time
3. Some functions could be missing because of the limited search criteria
Draft Reconstruction:

**Assembly Of Metabolite Spreadsheet**

<table>
<thead>
<tr>
<th>Metabolite name</th>
<th>Metabolite description</th>
<th>Metabolite neutral formula</th>
<th>Metabolite charged formula</th>
<th>Metabolite charge</th>
<th>Metabolite Compartment</th>
<th>Metabolite KEGGID</th>
<th>Metabolite PubChemID</th>
<th>Metabolite CheBI ID</th>
<th>Metabolite Inchi String</th>
</tr>
</thead>
<tbody>
<tr>
<td>ala-B[p]</td>
<td>beta-Alanine</td>
<td></td>
<td>C3H7NO2</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ala-D[c]</td>
<td>D-Alanine</td>
<td></td>
<td>C3H7NO2</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ala-D[e]</td>
<td>D-Alanine</td>
<td></td>
<td>C3H7NO2</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ala-D[p]</td>
<td>D-Alanine</td>
<td></td>
<td>C3H7NO2</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ala-L[c]</td>
<td>L-Alanine</td>
<td></td>
<td>C3H7NO2</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ala-L[e]</td>
<td>L-Alanine</td>
<td></td>
<td>C3H7NO2</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ala-L[p]</td>
<td>L-Alanine</td>
<td></td>
<td>C3H7NO2</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1. The draft metabolite spreadsheet should include a list of candidate metabolites
2. Not all of the spreadsheet cells will be filled at this time
3. Some metabolites could be missing because of the limited search criteria
Draft Reconstruction

1. Obtain genome annotation
2. Identify candidate metabolic functions
3. Obtain candidate metabolic reactions
4. Assembly of draft reconstruction
5. Collect experimental data
# Biochemical Databases

<table>
<thead>
<tr>
<th>Name</th>
<th>Link</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transport Classification Database (TCDB)</td>
<td><a href="http://www.tcdb.org/">http://www.tcdb.org/</a></td>
<td>TCDB is a curated database of factual information from over 10,000 published references.</td>
</tr>
</tbody>
</table>

### Protein Location Databases

<table>
<thead>
<tr>
<th>Name</th>
<th>Link</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>PA-SUB</td>
<td><a href="http://www.cs.ualberta.ca/~bioinfo/PA/Sub/">http://www.cs.ualberta.ca/~bioinfo/PA/Sub/</a></td>
<td>Proteome Analyst specialized Subcellular Localization server (SVM based).</td>
</tr>
</tbody>
</table>

### Bio-numbers

<table>
<thead>
<tr>
<th>Name</th>
<th>Link</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>CyberCell Database (CCDB)</td>
<td><a href="http://redpoll.pharmacy.ualberta.ca/CCDB/cgi-bin/STAT_NEW.cgi">http://redpoll.pharmacy.ualberta.ca/CCDB/cgi-bin/STAT_NEW.cgi</a></td>
<td></td>
</tr>
<tr>
<td>B10NUMB3R5</td>
<td><a href="http://bionumbers.hms.harvard.edu/">http://bionumbers.hms.harvard.edu/</a></td>
<td></td>
</tr>
</tbody>
</table>

GENOME-SCALE
METABOLIC RECONSTRUCTIONS

• Overview
• Draft Reconstruction
• Refinement of Reconstruction
• Conversion of Reconstruction into Computable Format
• Network Evaluation
• Data Assembly and Dissemination

Refinement of Reconstruction

• The entire draft reconstruction needs to be curated and refined.

• The metabolic functions and reactions collected in the draft reconstruction are individually evaluated against organism-specific literature (and expert opinion).

• Information about biomass composition, maintenance parameters and growth conditions need to be collected.

• Refine and assemble the curated reconstruction in a pathway-by-pathway manner, starting from the canonical pathways. Peripheral pathways and reactions/gene products without clear pathway assignment are added in a later step.

Refinement of Reconstruction

6. Determine and verify substrate and cofactor usage.
7. Obtain a neutral formula for each metabolite in the reaction.
8. Determine the charged formula for each metabolite in the reaction.
9. Calculate reaction stoichiometry.
10. Determine reaction directionality.
11. Add information for gene and reaction localization.
12. Add subsystem information to the reaction.
13. Verify GPR association.
15. Determine and add the confidence score.

16. Add references and notes.
17. Repeat Steps 6-17 for all those draft reconstruction genes.
18. Add spontaneous reactions.
19. Add extracellular and periplasmic transport reactions.
20. Add exchange reactions.
22. Draw metabolic map (optional).
34. Add NGAM Reaction (ATPM).
35. Add demand reactions.
36. Add sink reactions.
37. Determine growth medium requirements.
Refinement of Reconstruction:
Determine And Verify Substrate And Cofactor Usage

• If no organism-specific information can be found in the literature, information from phylogenetically close organisms can be used but should be marked as such.

• Reactions containing generic terms, such as protein, DNA, electron acceptor, and so on, should not be included, as they are not specific enough and normally serve in databases as space holders until more knowledge and biochemical evidence become available.

• Substrate and cofactor specificity of enzymes may differ between organisms. Organism-unspecific databases, such as KEGG and BRENDA, list all possible transformations of an enzyme that have been identified in any organism.

• Information about substrate and cofactor utilization can be obtained from organism-specific biochemical studies and may also be listed in organism-specific databases (e.g., Ecocyc).

### KEGG Gene Information

**Escherichia coli K-12 MG1655: b1236**

<table>
<thead>
<tr>
<th>Entry</th>
<th>CDS</th>
<th>KO00007</th>
</tr>
</thead>
</table>

**Gene name**: galU

**Definition**: glucose-1-phosphate uridylyltransferase (EC:2.7.7.9)

**Orthology**: KO0063 and glucose-1-phosphate uridylyltransferase (EC:2.7.7.9)

**Organism**: eco, Escherichia coli K-12 MG1655

**Pathway**:
- eco00040: Pentose and glucuronate interconversions
- eco00052: Galactose metabolism
- eco00500: Starch and sucrose metabolism
- eco00520: Amino sugar and nucleotide sugar metabolism
- eco01100: Metabolic pathways
- eco01110: Biosynthesis of secondary metabolites

**Module**: eco_M00362, Nucleotide sugar biosynthesis, prokaryotes

**Class**:
- Metabolism; Carbohydrate Metabolism; Pentose and glucuronate interconversions [PATH:eco00040]
- Metabolism; Carbohydrate Metabolism; Galactose metabolism [PATH:eco00052]
- Metabolism; Carbohydrate Metabolism; Starch and sucrose metabolism [PATH:eco00500]
- Metabolism; Carbohydrate Metabolism; Amino sugar and nucleotide sugar metabolism [PATH:eco00520]

[BRUTE hierarchy](http://www.genome.jp/dbget-bin/www_bget?eco:b1236)
### KEGG Enzyme Information

**Enzyme:** EC 2.7.7.9

<table>
<thead>
<tr>
<th>Entry</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Name</td>
<td><strong>UTP:--glucose-1-phosphate uridylyltransferase;</strong> UDP glucose pyrophosphorylase; glucose-1-phosphate uridylyltransferase; UTPG phosphorylase; UTPG pyrophosphorylase; uridine 5'-diphosphoglucose pyrophosphorylase; uridine diphosphoglucose pyrophosphorylase; uridine diphosphate-D-glucose pyrophosphorylase; uridine-diphosphate glucose pyrophosphorylase</td>
</tr>
<tr>
<td>Class</td>
<td>Transferrases; Transferring phosphorus-containing groups; Nucleotidyltransferases</td>
</tr>
<tr>
<td>Sysname</td>
<td><strong>UTP:alpha-D-glucose-1-phosphate uridylyltransferase</strong></td>
</tr>
<tr>
<td>Reaction (TUBIN)</td>
<td><strong>UTP + alpha-D-glucose 1-phosphate = diphosphate + UDP-glucose</strong></td>
</tr>
<tr>
<td>Reaction (KEGG)</td>
<td>[RN:R00289]</td>
</tr>
</tbody>
</table>

*http://www.genome.jp/dbget-bin/www_bget?ec:2.7.7.9*
# KEGG Reaction Information

**Reaction: R00289**

<table>
<thead>
<tr>
<th>Entry</th>
<th>Reaction</th>
</tr>
</thead>
<tbody>
<tr>
<td>R00289</td>
<td></td>
</tr>
</tbody>
</table>

**Name:** UTP:alpha-D-glucose-1-phosphate uridylyltransferase  
**Definition:** UTP + D-Glucose 1-phosphate $\leftrightarrow$ Tdiphosphate + UTP-glucose

**Equation:** C00075 + C00103 $\leftrightarrow$ C00013 + C00029

**KEGG Reaction Information**

- **Entry:** R00289
- **Name:** UTP:alpha-D-glucose-1-phosphate uridylyltransferase
- **Definition:** UTP + D-Glucose 1-phosphate $\leftrightarrow$ Tdiphosphate + UTP-glucose
- **Equation:** C00075 + C00103 $\leftrightarrow$ C00013 + C00029
- **RPair:** RP00196 C00029_C00103 main  
  - RP00382 C00013_C00075 leave  
  - RP00546 C00029_C00075 trans
- **Enzyme:** EC:2.7.7.9
- **Pathway:**  
  - rn00040 Pentose and glucuronate interconversions  
  - rn00052 Galactose metabolism  
  - rn00050 Starch and sucrose metabolism  
  - rn00520 Amino sugar and nucleotide sugar metabolism  
  - rn01100 Metabolic pathways  
  - rn01110 Biosynthesis of secondary metabolites
- **Orthology:** K00963 UTP-glucose-1-phosphate uridylyltransferase [EC:2.7.7.9]

Brenda Enzyme Information

EC 2.7.7.9 - UTP-glucose-1-phosphate uridylyltransferase

http://www.brenda-enzymes.info/php/result_flat.php?ecno=2.7.7.9&Suchword=&organism%5B%5D=Escherichia+coli&show_tm=0
Refinement of Reconstruction

6. Determine and verify substrate and cofactor usage.
7. Obtain a neutral formula for each metabolite in the reaction.
8. Determine the charged formula for each metabolite in the reaction.
9. Calculate reaction stoichiometry.
10. Determine reaction directionality.
11. Add information for gene and reaction localization.
12. Add subsystem information to the reaction.
13. Verify GPR association.
15. Determine and add the confidence score.
16. Add references and notes.
17. Repeat Steps 6-17 for all those draft reconstruction genes.
18. Add spontaneous reactions.
19. Add extracellular and periplasmic transport reactions.
20. Add exchange reactions.
22. Draw metabolic map (optional).
34. Add NGAM Reaction (ATPM).
35. Add demand reactions.
36. Add sink reactions.
37. Determine growth medium requirements.
Refinement of Reconstruction:

**Obtain a Neutral and Charged Formula for each Metabolite in the Reaction**

- In databases, metabolites are generally listed with their uncharged formula.
- In contrast, in medium and in cells, many metabolites are protonated or deprotonated.
- The protonation state, and thus, the charged formula, depends on the pH of interest. Often metabolic networks are reconstructed assuming an intracellular pH of 7.2.
- The intracellular pH of bacterial cells may vary depending on, e.g., environmental conditions.
- The pH of organelles may be different, e.g., peroxisome and lysosome.
- The protonated formula is calculated based on the $\text{pK}_a$ value of the functional groups.
- Software packages, such as Pipeline Pilot and $\text{pK}_a$ DB, can predict the $\text{pK}_a$ values for a given compound ([http://www.chemaxon.com/marvin/help/calculations/pKa.html](http://www.chemaxon.com/marvin/help/calculations/pKa.html)).
List Of Functional Groups, Their Charge Formula And The Corresponding $pK_a$

<table>
<thead>
<tr>
<th>Molecule/group</th>
<th>Acid</th>
<th>Base</th>
<th>$pK_a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acetic acid</td>
<td>$\text{H}_2\text{C}_2\text{O}_2\text{OH}$</td>
<td>$\text{H}_2\text{C}_2\text{O}_2^- + \text{H}^+$</td>
<td>4.76</td>
</tr>
<tr>
<td>Carboxyl group</td>
<td>$\text{R}-\text{COH}$</td>
<td>$\text{R}-\text{COO}^- + \text{H}^+$</td>
<td>1.8–2.4</td>
</tr>
<tr>
<td>Ammonium</td>
<td>$\text{NH}_4^+$</td>
<td>$\text{NH}_3 + \text{H}^+$</td>
<td>9.22</td>
</tr>
<tr>
<td>Amino group</td>
<td>$\text{R}-\text{NH}_2$</td>
<td>$\text{R}-\text{NH}_3^- + \text{H}^+$</td>
<td>8–11</td>
</tr>
<tr>
<td>Bicarbonate</td>
<td>$\text{HO}_2\text{CO}$</td>
<td>$\text{HO}_2\text{CO}^- + \text{H}^+$</td>
<td>3.77</td>
</tr>
<tr>
<td>Glycine</td>
<td>$\text{CH}_2\text{OH}$</td>
<td>$\text{CH}_2\text{OH}^- + \text{H}^+$</td>
<td>2.34</td>
</tr>
<tr>
<td>Phosphoric acid</td>
<td>$\text{H}_3\text{PO}_4$</td>
<td>$\text{H}_3\text{PO}_4^- + \text{H}^+$</td>
<td>2.14</td>
</tr>
</tbody>
</table>

Neutral and Charged Formula for each Metabolite in the Reaction

<table>
<thead>
<tr>
<th>Substrates</th>
<th>C6H12O6</th>
<th>C10H16N5O13P3</th>
<th>C6H13O9P0</th>
<th>C10H15N5O10P2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neutral Formulae</td>
<td>C6H12O6</td>
<td>C10H16N5O13P3</td>
<td>C6H13O9P0</td>
<td>C10H15N5O10P2</td>
</tr>
<tr>
<td>Charged Formulae</td>
<td>C6H12O6</td>
<td>C10H16N5O13P3</td>
<td>C6H13O9P0</td>
<td>C10H15N5O10P2</td>
</tr>
</tbody>
</table>

REI601M, Introduction to Systems Biology, Dr. Innes Thiele 2012, https://systemsbiology.hi.is/wiki/REI601M
Example of Finding the Metabolite Charge

- Go to the KEGG website and enter the KEGGID
  ✓ http://www.genome.jp/kegg/
- Download the “mol” file (Copy text to file; include all spaces)
Example of Finding the Metabolite Charge

- Open the file in MarvinSpace (free to academic institutions)
- Under the "calculations" menu:
  - calculations -> protonation -> pKa
- Click OK on the pKa options window
MarvinSketch Windows Showing pH Values

Percentage of metabolites that exists at a given pH value

pH Value
Marvin Tools: Example #2

By Dr. Wenfeng Guo

http://www.ebi.ac.uk/chebi/searchId.do?chebiId=CHEBI:15351
Marvin Tools: Example #2 (II)

By Dr. Wenfeng Guo
Marvin Tools: Example #2 (III)

You can also cut and paste into MarvinSketch.
Acetyl-CoA (CHEBI:15351)
Refinement of Reconstruction

6. Determine and verify substrate and cofactor usage.
7. Obtain a neutral formula for each metabolite in the reaction.
8. Determine the charged formula for each metabolite in the reaction.
9. Calculate reaction stoichiometry.
10. Determine reaction directionality.
11. Add information for gene and reaction localization.
12. Add subsystem information to the reaction.
13. Verify GPR association.
15. Determine and add the confidence score.
16. Add references and notes.
17. Repeat Steps 6-17 for all those draft reconstruction genes.
18. Add spontaneous reactions.
19. Add extracellular and periplasmic transport reactions.
20. Add exchange reactions.
22. Draw metabolic map (optional).
34. Add NGAM Reaction (ATPM).
35. Add demand reactions.
36. Add sink reactions.
37. Determine growth medium requirements.
Refinement of Reconstruction:

**Calculate Reaction Stoichiometry**

- The reaction stoichiometry can be determined by counting different elements on the left- and right-hand side of the reaction.
- Addition of protons and water may be required in this step, as some databases and many biochemical textbooks omit these molecules from the reactions.
- It is therefore necessary to balance every element and charge on both sides of the reaction.
- It should be noted that unbalanced reactions may lead to the synthesis of protons or energy (ATP) out of nothing.

Calculate Reaction Stoichiometry

Substrates
- Glc: C_6H_{12}O_6^0
- ATP: C_{10}H_{16}N_5O_{13}P_3^0
- G6P: C_6H_{13}O_{9}P^0
- ADP: C_{10}H_{15}N_5O_{10}P_2^0

Neutral formulae
- Glc: C_6H_{12}O_6^0
- ATP: C_{10}H_{12}N_5O_{13}P_3^{4-}
- G6P: C_6H_{11}O_{9}P^{-2}
- ADP: C_{10}H_{12}N_5O_{10}P_2^{3-}

Charged formulae

Stoichiometry
- 1 Glc + 1 ATP = C_{16}H_{24}O_{19}P_3, 4e^-
- 1 G6P + 1 ADP + 1 H^+ = C_{16}H_{23}O_{19}P_3, 5e^-

Reefinement of Reconstruction

6. Determine and verify substrate and cofactor usage.
7. Obtain a neutral formula for each metabolite in the reaction.
8. Determine the charged formula for each metabolite in the reaction.
9. Calculate reaction stoichiometry.
10. Determine reaction directionality.
11. Add information for gene and reaction localization.
12. Add subsystem information to the reaction.
13. Verify GPR association.
15. Determine and add the confidence score.
16. Add references and notes.
17. Repeat Steps 6-17 for all those draft reconstruction genes.
18. Add spontaneous reactions.
19. Add extracellular and periplasmic transport reactions.
20. Add exchange reactions.
22. Draw metabolic map (optional).
34. Add NGAM Reaction (ATPM).
35. Add demand reactions.
36. Add sink reactions.
37. Determine growth medium requirements.
Refinement of Reconstruction:

Determine Reaction Directionality

- Use biochemical data and literature if available.
- Alternatively, the standard $\Delta_f G^\circ$ and of $\Delta_r G^\circ$ can be calculated based on group contribution theory for most KEGG reactions from Web GCM.
- If data on reaction of interest are not available, the following rule of thumb may be applied: (1) reactions involving transfer of phosphate from ATP to an acceptor molecule should be irreversible (with the exception of the ATP synthetase, which is known to occur in reverse); and (2) reactions involving quinones are generally irreversible.
- Assigning the wrong direction to a reaction may have significant impact on the model's performance. In general, one should leave a reaction reversible if no information is available and the aforementioned rules of thumb do not apply.
- Models with too many reversible reactions (too loose constraints) may have the so-called futile cycle that can overcome the proton gradient by freely exchanging metabolites and protons across compartments.

Determine Reaction Directionality

<table>
<thead>
<tr>
<th>Substrates</th>
<th>Neutral Formulae</th>
<th>Charged Formulae</th>
<th>Stoichiometry</th>
<th>Directionality</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( \text{glc} ) ( \text{C}<em>6\text{H}</em>{12}\text{O}_6^0 )</td>
<td>( \text{atp} ) ( \text{C}<em>{10}\text{H}</em>{16}\text{N}<em>5\text{O}</em>{13}\text{P}_3^0 )</td>
<td>( 1 \text{glc} + 1 \text{atp} )</td>
<td>( 1 \text{glc} + 1 \text{atp} \rightarrow 1 \text{g6p} + 1 \text{adp} + 1 \text{H}^+ )</td>
</tr>
<tr>
<td></td>
<td>( \text{g6p} ) ( \text{C}<em>6\text{H}</em>{13}\text{O}_9\text{P}^0 )</td>
<td>( \text{adp} ) ( \text{C}<em>{10}\text{H}</em>{15}\text{N}<em>5\text{O}</em>{10}\text{P}_2^0 )</td>
<td>( 1 \text{g6p} + 1 \text{adp} + 1 \text{H}^+ )</td>
<td></td>
</tr>
</tbody>
</table>

\( \text{C}_{16}\text{H}_{24}\text{O}_{19}\text{P}_3^-, 4\text{e}^- \) \( \Rightarrow \) \( \text{C}_{16}\text{H}_{23}\text{O}_{19}\text{P}_3^-, 5\text{e}^- \)
Refinement of Reconstruction

6. Determine and verify substrate and cofactor usage.
7. Obtain a neutral formula for each metabolite in the reaction.
8. Determine the charged formula for each metabolite in the reaction.
9. Calculate reaction stoichiometry.
10. Determine reaction directionality.
11. Add information for gene and reaction localization.
12. Add subsystem information to the reaction.
13. Verify GPR association.
15. Determine and add the confidence score.
16. Add references and notes.
17. Repeat Steps 6-17 for all those draft reconstruction genes.
18. Add spontaneous reactions.
19. Add extracellular and periplasmic transport reactions.
20. Add exchange reactions.
22. Draw metabolic map (optional).
34. Add NGAM Reaction (ATPM).
35. Add demand reactions.
36. Add sink reactions.
37. Determine growth medium requirements.
Refinement of Reconstruction:

Determine Gene And Reaction Localization

• The use of algorithms such as PSORT and PASUB can be considered if no experimental data are available.

• In the absence of appropriate data, proteins should be assumed to reside in the cytosol.

## List Of Cellular Compartments Used In Reconstructions

<table>
<thead>
<tr>
<th>Compartment</th>
<th>Commonly used symbol</th>
<th>Achaea</th>
<th>Bacteria</th>
<th>Eukaryotic pathogens</th>
<th>Fungi</th>
<th>Photosynthetic eukarya</th>
<th>Baker’s yeast</th>
<th>Human</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extracellular space</td>
<td>[e]</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Periplasm</td>
<td>[p]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cytoplasm</td>
<td>[c]</td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Nucleus</td>
<td>[n]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mitochondrion</td>
<td>[m]</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Chloroplast</td>
<td>[h]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Lysosome*</td>
<td>[l]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vacuole</td>
<td>[v]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Golgi apparatus</td>
<td>[g]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Endoplasmatic reticulum</td>
<td>[r]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Peroxisome</td>
<td>[x]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Flagellum</td>
<td>[f]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Glyoxysome</td>
<td>[o]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Glycosome</td>
<td>[y]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Acidocalcisome</td>
<td>[a]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
</tbody>
</table>

### Reaction/Metabolite Requirements

- **Substrates**
  - **glc**: $C_6H_{12}O_6^0$
  - **atp**: $C_{10}H_{16}N_5O_3^0$
  - **g6p**: $C_6H_{13}O_9P^0$
  - **adp**: $C_{10}H_{15}N_5O_{10}P_2^0$

- **Neutral formulae**
  - $C_6H_{12}O_6^0$
  - $C_{10}H_{12}N_5O_{13}P_3^0$
  - $C_6H_{13}O_9P^0$
  - $C_{10}H_{15}N_5O_{10}P_2^0$

- **Charged formulae**
  - $C_6H_{12}O_6^0$
  - $C_{10}H_{12}N_5O_{13}P_3^{4-}$
  - $C_6H_{11}O_9P^{2-}$
  - $C_{10}H_{12}N_5O_{10}P_2^{3-}$

- **Stoichiometry**
  - $1\text{ glc} + 1\text{ atp}$
  - $1\text{ g6p} + 1\text{ adp} + 1\text{ h}^+$

  $C_{16}H_{24}O_{19}P_3^-, 4\text{e}^-$
  $=\quad C_{16}H_{23}O_{19}P_3^-, 5\text{e}^-$

- **Directionality**
  - $1\text{ glc} + 1\text{ atp} \rightarrow 1\text{ g6p} + 1\text{ adp} + 1\text{ h}^+$

- **Location**
  - Cytosol: $1\text{ glc} + 1\text{ atp} \rightarrow 1\text{ g6p} + 1\text{ adp} + 1\text{ h}^+$

---

REI601M, Introduction to Systems Biology, Dr. Innes Thiele 2012, https://systemsbiology.hi.is/wiki/REI601M
Refinement of Reconstruction

6. Determine and verify substrate and cofactor usage.
7. Obtain a neutral formula for each metabolite in the reaction.
8. Determine the charged formula for each metabolite in the reaction.
9. Calculate reaction stoichiometry.
10. Determine reaction directionality.
11. Add information for gene and reaction localization.
12. Add subsystem information to the reaction.
13. Verify GPR association.
15. Determine and add the confidence score.
16. Add references and notes.
17. Repeat Steps 6-17 for all those draft reconstruction genes.
18. Add spontaneous reactions.
19. Add extracellular and periplasmic transport reactions.
20. Add exchange reactions.
22. Draw metabolic map (optional).
34. Add NGAM Reaction (ATPM).
35. Add demand reactions.
36. Add sink reactions.
37. Determine growth medium requirements.
Refinement of Reconstruction:
Add Subsystem Information To The Reaction

• This information will be of great help for the debugging, network visualization (Paint4Net), and network evaluation work.

• The subsystem assignment can be done based on, e.g., biochemical textbooks or KEGG maps. Note that a reaction or an enzyme can appear in multiple KEGG maps; therefore, the subsystem should reflect its primary function.

• See http://www.genome.jp/kegg/pathway.html

Refinement of Reconstruction

6. Determine and verify substrate and cofactor usage.
7. Obtain a neutral formula for each metabolite in the reaction.
8. Determine the charged formula for each metabolite in the reaction.
9. Calculate reaction stoichiometry.
10. Determine reaction directionality.
11. Add information for gene and reaction localization.
12. Add subsystem information to the reaction.
13. Verify GPR association.
15. Determine and add the confidence score.
16. Add references and notes.
17. Repeat Steps 6-17 for all those draft reconstruction genes.
18. Add spontaneous reactions.
19. Add extracellular and periplasmic transport reactions.
20. Add exchange reactions.
22. Draw metabolic map (optional).
34. Add NGAM Reaction (ATPM).
35. Add demand reactions.
36. Add sink reactions.
37. Determine growth medium requirements.
Refinement of Reconstruction: Verify GPR Association

- The genome annotation often provides information about the GPR association, i.e., it indicates which gene has what function.

- The verification and refinement necessary in this step includes determining:
  - if the functional protein is a heteromeric enzyme complex;
  - if the enzyme (complex) can carry out more than one reaction and
  - if more than one protein can carry out the same function (i.e., isozymes exist).

- Linear pathways, such as fatty acid oxidation, have often been combined into few lumped reactions. The genes associated with these reactions are all required, with the exception of isozymes. Subsequently, the GPR association should reflect the requirement for all genes within the lumped reaction by using the Boolean rule AND.

Examples of GPR Associations and their Representation in Boolean Format

Refinement of Reconstruction

6. Determine and verify substrate and cofactor usage.
7. Obtain a neutral formula for each metabolite in the reaction.
8. Determine the charged formula for each metabolite in the reaction.
9. Calculate reaction stoichiometry.
10. Determine reaction directionality.
11. Add information for gene and reaction localization.
12. Add subsystem information to the reaction.
13. Verify GPR association.
15. Determine and add the confidence score.
16. Add references and notes.
17. Repeat Steps 6-17 for all those draft reconstruction genes.
18. Add spontaneous reactions.
19. Add extracellular and periplasmic transport reactions.
20. Add exchange reactions.
22. Draw metabolic map (optional).
34. Add NGAM Reaction (ATPM).
35. Add demand reactions.
36. Add sink reactions.
37. Determine growth medium requirements.
Refinement of Reconstruction:

Add Metabolite Identifier

- Metabolite identifiers are necessary to enable the use of reconstructions for high-throughput data mapping (e.g., metabolomic or fluxomic data) and for comparison of network content with other metabolic reconstructions.

- Each metabolite should be associated with at least one of the following identifiers:
  - ChEBI (http://www.ebi.ac.uk/chebi/)
  - KEGG (http://www.genome.jp/kegg/)
    - In many cases, having one of the identifiers is sufficient to automatically obtain the other two identifiers.

- Database-independent representations of the exact chemical structure of metabolites include:
  - InCHI strings (http://www.iupac.org/home/publications/e-resources/inchi.html)

- Databases containing the atoms, bonds, connectivity and coordinates of a molecule, include:
  - Molfiles (MDL file format, http://www.symyx.com/),

Refinement of Reconstruction

6. Determine and verify substrate and cofactor usage.
7. Obtain a neutral formula for each metabolite in the reaction.
8. Determine the charged formula for each metabolite in the reaction.
9. Calculate reaction stoichiometry.
10. Determine reaction directionality.
11. Add information for gene and reaction localization.
12. Add subsystem information to the reaction.
13. Verify GPR association.
15. Determine and add the confidence score.
16. Add references and notes.
17. Repeat Steps 6-17 for all those draft reconstruction genes.
18. Add spontaneous reactions.
19. Add extracellular and periplasmic transport reactions.
20. Add exchange reactions.
22. Draw metabolic map (optional).
34. Add NGAM Reaction (ATPM).
35. Add demand reactions.
36. Add sink reactions.
37. Determine growth medium requirements.
Refinement of Reconstruction:
Determine And Add The Confidence Score

• The confidence score provides a fast way of assessing the amount of information available for a metabolic function, pathway or the entire reconstruction.

• Every network reaction should have a confidence score reflecting the information and evidence currently available.

• The confidence score ranges from 0 to 4, where 0 is the lowest and 4 is the highest evidence score.

• It should be noted that multiple information types result in a cumulative confidence score. For example, a confidence score of 4 may represent physiological and sequence evidence.

# Confidence Scoring System Currently Employed for Metabolic Reconstructions

<table>
<thead>
<tr>
<th>Evidence type</th>
<th>Confidence score</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Biochemical data</td>
<td>4</td>
<td>Direct evidence for gene product function and biochemical reaction: protein purification, biochemical assays, experimentally solved protein structures and comparative gene-expression studies</td>
</tr>
<tr>
<td>Genetic data</td>
<td>3</td>
<td>Direct and indirect evidence for gene function: knockout characterization, knock-in characterization and overexpression</td>
</tr>
<tr>
<td>Physiological data</td>
<td>2</td>
<td>Indirect evidence for biochemical reactions based on physiological data: secretion products or defined medium components serve as evidence for transport and metabolic reactions</td>
</tr>
<tr>
<td>Sequence data</td>
<td>2</td>
<td>Evidence for gene function: genome annotation and SEED annotation</td>
</tr>
<tr>
<td>Modeling data</td>
<td>1</td>
<td>No evidence is available, but reaction is required for modeling. The included function is a hypothesis and needs experimental verification. The reaction mechanism may be different from the included reaction(s)</td>
</tr>
<tr>
<td>Not evaluated</td>
<td>0</td>
<td>ADO</td>
</tr>
</tbody>
</table>

Refinement of Reconstruction

6. Determine and verify substrate and cofactor usage.
7. Obtain a neutral formula for each metabolite in the reaction.
8. Determine the charged formula for each metabolite in the reaction.
9. Calculate reaction stoichiometry.
10. Determine reaction directionality.
11. Add information for gene and reaction localization.
12. Add subsystem information to the reaction.
13. Verify GPR association.
15. Determine and add the confidence score.

16. Add references and notes.
17. Repeat Steps 6-17 for all those draft reconstruction genes.
18. Add spontaneous reactions.
19. Add extracellular and periplasmic transport reactions.
20. Add exchange reactions.
22. Draw metabolic map (optional).
34. Add NGAM Reaction (ATPM).
35. Add demand reactions.
36. Add sink reactions.
37. Determine growth medium requirements.
### Reaction Spreadsheet

**ecoli_iaf1260.xls**

<table>
<thead>
<tr>
<th>Rxn name</th>
<th>Rxn description</th>
<th>Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>FUM</td>
<td>fumarase</td>
<td>fum[c] + h2o[c] ↔ mal-L[c]</td>
</tr>
<tr>
<td>FUMT2_2pp</td>
<td>Fumarate transport via proton symport (2 H) (periplasm)</td>
<td>fum[p] + 2h[p] → fum[c] + 2h[c]</td>
</tr>
<tr>
<td>FUMT2_3pp</td>
<td>Fumarate transport via proton symport (3 H) (periplasm)</td>
<td>fum[p] + 3h[p] → fum[c] + 3h[c]</td>
</tr>
<tr>
<td>FUMtex</td>
<td>Fumarate transport via diffusion (extracellular to periplasm)</td>
<td>fum[e] ↔ fum[p]</td>
</tr>
</tbody>
</table>

### Gene-reaction association

<table>
<thead>
<tr>
<th>Gene-reaction association</th>
<th>Genes</th>
<th>Proteins</th>
<th>Subsystem</th>
</tr>
</thead>
<tbody>
<tr>
<td>(b2801)</td>
<td>b2801</td>
<td>Transport, Inner Membrane</td>
<td></td>
</tr>
<tr>
<td>(b1612 or b1611 or b4122)</td>
<td>b1611 b1612 b4122</td>
<td>Citric Acid Cycle</td>
<td></td>
</tr>
<tr>
<td>(b3528)</td>
<td>b3528</td>
<td>Transport, Inner Membrane</td>
<td></td>
</tr>
<tr>
<td>(b4138 or b0621 or b4123)</td>
<td>b0621 b4123 b4138</td>
<td>Transport, Inner Membrane</td>
<td></td>
</tr>
<tr>
<td>(b2215 or b0241 or b1377 or b0929)</td>
<td>b0241 b0929 b1377 b2215</td>
<td>Transport, Outer Membrane Porin</td>
<td></td>
</tr>
</tbody>
</table>

### Table

<table>
<thead>
<tr>
<th>Reversible</th>
<th>LB</th>
<th>UB</th>
<th>Objective</th>
<th>Confidence Score</th>
<th>EC Number</th>
<th>Notes</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-1000</td>
<td>1000</td>
<td>0</td>
<td>0</td>
<td>4.2.1.2</td>
<td>LOCUS:b1611#ABBREVIATION:fumC#ECNUMBERS:4.2.1.2#</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1000</td>
<td>0</td>
<td>0</td>
<td>JLR</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1000</td>
<td>0</td>
<td>4</td>
<td>JLR</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>-1000</td>
<td>1000</td>
<td>0</td>
<td>0</td>
<td>LOCUS:b0241#ABBREVIATION:phoE#</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1000</td>
<td>0</td>
<td>0</td>
<td>2.3.1.157</td>
<td>LOCUS:b3730#ABBREVIATION:gluU#ECNUMBERS:2.7.7.23#</td>
<td></td>
</tr>
</tbody>
</table>
## Metabolite Spreadsheet

### metabolite neutral formula

<table>
<thead>
<tr>
<th>Metabolite name</th>
<th>Metabolite description</th>
<th>Metabolite neutral formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>ala-D[p]</td>
<td>D-Alanine</td>
<td></td>
</tr>
<tr>
<td>ala-L[c]</td>
<td>L-Alanine</td>
<td></td>
</tr>
<tr>
<td>ala-L[e]</td>
<td>L-Alanine</td>
<td></td>
</tr>
<tr>
<td>ala-L[p]</td>
<td>L-Alanine</td>
<td></td>
</tr>
<tr>
<td>alaala[c]</td>
<td>D-AlanyL-D-alanine</td>
<td></td>
</tr>
<tr>
<td>alaala[e]</td>
<td>D-AlanyL-D-alanine</td>
<td></td>
</tr>
</tbody>
</table>

### metabolite charged formula

<table>
<thead>
<tr>
<th>Metabolite charged formula</th>
<th>Metabolite charge</th>
<th>Metabolite Compartment</th>
<th>Metabolite KEGGID</th>
</tr>
</thead>
<tbody>
<tr>
<td>C3H7NO2</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C3H7NO2</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C3H7NO2</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C3H7NO2</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C3H7NO2</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C6H12N2O3</td>
<td>0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### metabolite PubChemID

<table>
<thead>
<tr>
<th>Metabolite PubChemID</th>
<th>Metabolite CheBI ID</th>
<th>Metabolite Inchi String</th>
<th>Metabolite Smile</th>
</tr>
</thead>
</table>
Refinement of Reconstruction

6. Determine and verify substrate and cofactor usage.
7. Obtain a neutral formula for each metabolite in the reaction.
8. Determine the charged formula for each metabolite in the reaction.
9. Calculate reaction stoichiometry.
10. Determine reaction directionality.
11. Add information for gene and reaction localization.
12. Add subsystem information to the reaction.
13. Verify GPR association.
15. Determine and add the confidence score.
16. Add references and notes.
17. Repeat Steps 6-17 for all those draft reconstruction genes.
18. Add spontaneous reactions.
19. Add extracellular and periplasmic transport reactions.
20. Add exchange reactions.
22. Draw metabolic map (optional).
34. Add NGAM Reaction (ATPM).
35. Add demand reactions.
36. Add sink reactions.
37. Determine growth medium requirements.
Refinement of Reconstruction: 
Add Spontaneous Reactions

- The biochemical literature and databases (e.g., KEGG and BRENDA) are to be used to identify candidate spontaneous reactions that are to be included.

- Only include those reactions, which have at least one metabolite present in the reconstruction to minimize the number of dead ends.

- Associate the spontaneous reactions with an artificial gene (s0001) and protein (S0001).

Refinement of Reconstruction

6. Determine and verify substrate and cofactor usage.
7. Obtain a neutral formula for each metabolite in the reaction.
8. Determine the charged formula for each metabolite in the reaction.
9. Calculate reaction stoichiometry.
10. Determine reaction directionality.
11. Add information for gene and reaction localization.
12. Add subsystem information to the reaction.
13. Verify GPR association.
15. Determine and add the confidence score.

16. Add references and notes.
17. Repeat Steps 6-17 for all those draft reconstruction genes.
18. Add spontaneous reactions.
19. Add extracellular and periplasmic transport reactions.
20. Add exchange reactions.
22. Draw metabolic map (optional).
34. Add NGAM Reaction (ATPM).
35. Add demand reactions.
36. Add sink reactions.
37. Determine growth medium requirements.
Refinement of Reconstruction:

Add Extracellular, Periplasmic Transport Reactions, and Exchange Reactions

- Every metabolite taken up from the medium or is secreted into the medium should include a transport reaction (extracellular space to periplasm and periplasm to cytoplasm).
- The transport reactions for metabolites that can diffuse through the membranes must be included. Small, hydrophilic compounds can diffuse through the outer membrane.
- Exchange reactions need to be added for all extracellular metabolites.

Refinement of Reconstruction

6. Determine and verify substrate and cofactor usage.
7. Obtain a neutral formula for each metabolite in the reaction.
8. Determine the charged formula for each metabolite in the reaction.
9. Calculate reaction stoichiometry.
10. Determine reaction directionality.
11. Add information for gene and reaction localization.
12. Add subsystem information to the reaction.
13. Verify GPR association.
15. Determine and add the confidence score.
16. Add references and notes.
17. Repeat Steps 6-17 for all those draft reconstruction genes.
18. Add spontaneous reactions.
19. Add extracellular and periplasmic transport reactions.
20. Add exchange reactions.
22. Draw metabolic map (optional).
34. Add NGAM Reaction (ATPM).
35. Add demand reactions.
36. Add sink reactions.
37. Determine growth medium requirements.
Refinement of Reconstruction:

Add Intracellular Transport Reactions

• When multi-compartment networks are constructed, intracellular transport reactions need to be added for all the metabolites that are supposed to 'move' between compartments.

• Minimize the number of intracellular transport reactions to the ones that really need to be there.

• If too many transport reactions are added in a reconstruction, they can cause cycles (futile cycles or Type III pathways). This is a common problem in reconstructions with multiple compartments.

• For the directionality of intracellular transport reactions, one should consider the nature of the pathway in the compartment. If the pathway is biosynthetic, it is very likely that (i) the precursor(s) is only imported, (ii) the product(s) of the pathway is only exported from the compartment and (iii) intermediates are not transported at all.

• Many transport reactions are in symport or antiport with protons, cations or other metabolites.

• To minimize the error and increase consistency, one can adopt the intracellular transport mechanism from a corresponding transport reaction from extracellular/periplasmic space to cytoplasm if it is known (and it is not an ABC transport reaction); otherwise (facilitated) diffusion reaction may be assumed as the mechanism.

Refinement of Reconstruction

6. Determine and verify substrate and cofactor usage.
7. Obtain a neutral formula for each metabolite in the reaction.
8. Determine the charged formula for each metabolite in the reaction.
9. Calculate reaction stoichiometry.
10. Determine reaction directionality.
11. Add information for gene and reaction localization.
12. Add subsystem information to the reaction.
13. Verify GPR association.
15. Determine and add the confidence score.
16. Add references and notes.
17. Repeat Steps 6-17 for all those draft reconstruction genes.
18. Add spontaneous reactions.
19. Add extracellular and periplasmic transport reactions.
20. Add exchange reactions.
22. Draw metabolic map (optional).
34. Add NGAM Reaction (ATPM).
35. Add demand reactions.
36. Add sink reactions.
37. Determine growth medium requirements.
Constraint-based Metabolic Reconstructions & Analysis

H. Scott Hinton, 2016

Lesson: Genome-scale Metabolic Reconstructions

Refinement of Reconstruction:

Draw Metabolic Map

• Paint4Net Developed by Andrejs Kostromins

• Paint4Net v1.0 is the COBRA Toolbox extension for visualization of constraints-based reconstruction and analysis (COBRA) models and reconstructions in the MATLAB environment.

• Uses the Bioinformatics toolbox to visualize COBRA models and reconstructions as a hypergraph.

• The Paint4Net v1.0 contains two main commands:
  
  • **draw_by_rxn**
    • For visualization of all or a part of a COBRA model by specified list of reactions.

  • **draw_by_met**
    • For visualization of the connectivity of a particular metabolite with other metabolites through reactions of a COBRA model

Assessing the Metabolic “Environment” or “Connectivity” of A Metabolite (KEGG Map)

Refinement of Reconstruction

6. Determine and verify substrate and cofactor usage.
7. Obtain a neutral formula for each metabolite in the reaction.
8. Determine the charged formula for each metabolite in the reaction.
9. Calculate reaction stoichiometry.
10. Determine reaction directionality.
11. Add information for gene and reaction localization.
12. Add subsystem information to the reaction.
13. Verify GPR association.
15. Determine and add the confidence score.
16. Add references and notes.
17. Repeat Steps 6-17 for all those draft reconstruction genes.
18. Add spontaneous reactions.
19. Add extracellular and periplasmic transport reactions.
20. Add exchange reactions.
22. Draw metabolic map (optional).
34. Add NGAM Reaction (ATPM).
35. Add demand reactions.
36. Add sink reactions.
37. Determine growth medium requirements.
Refinement of Reconstruction:
Determine Biomass Composition

- The biomass reaction accounts for all known biomass constituents and their fractional contributions to the overall cellular biomass.
- Needs to be determined experimentally for cells growing in log phase.
- It may not be possible to obtain a detailed biomass composition for the target organism. In this case, one can estimate the relative fraction of each precursor from the genome (e.g., by using the Comprehensive Microbial Resource (CMR) database).
- The contribution of fatty acids and phospholipids needs to be determined from experiments. The model compounds will not represent all possible combinations but only average compounds that are consistent with the experimental data individual.

Important Role of the Biomass Objective Function

• If a biomass precursor is not accounted for in the biomass reactions, the synthesis reactions may not be required for growth (i.e., it is nonessential). Therefore, associated genes may not be assumed as essential. Subsequently, the presence or absence of a metabolite in the biomass reaction may affect the in silico essentiality of reactions and their associated gene(s).

• Also, the fractional contribution of each precursor has a minor role for gene and reaction essentiality studies. When one wishes to predict the optimal growth rate accurately, the fractional distribution of each compound has an important role.

• The unit of the biomass reaction is h⁻¹, as all biomass precursor fractions are converted to mmol·gDW⁻¹. Therefore, the biomass reaction sums the mole fraction of each precursor necessary to produce 1 g dry weight of cells.

Definition of Biomass Reaction

Quantifying Macromolecular Content of a cell

Quantifying Building Blocks of Macromolecules

Refinement of Reconstruction:

**Determine Biomass Composition**

24. Determine the chemical composition of the cell, i.e., protein, RNA, DNA, lipids, and cofactor content.
25. Determine the amino acid content either experimentally or by estimation.
26. The molar percentage and molecular weight of each amino acid must be used to calculate the weight per mol protein.
27. Determine the nucleotide content either experimentally or by estimation.
28. Calculate the fractional distribution of each nucleotide to the biomass composition.
29. Determine the lipid content.
30. Determine the content of the soluble pool (polyamines and vitamins and cofactors).
31. Determine the ion content.
32. Determine GAM.
33. Compile and add biomass reaction to the reconstruction.
Determine the Chemical Composition of the Cell

Example of Biomass Composition Determination for *Pseudomonas putida KT 2440*

Chemical composition of *E. coli* adopted from and utilized as a template for *P. putida KT2440*, since no extensive information was available.

Phospholipid contributions to the biomass function where PE is Phosphatidylethanolamine, PG is phosphatidylglycerol, and CL is cardiolipin.

Protein composition in *P. putida* broken down by monomer contributions in mmol/gDW.

dNTP composition of the entire *P. putida* chromosomal genome. The data are obtained from direct measurements, literature, or can be estimated from genome information.

### Ecoli_iaf1260 Core Biomass Objective Function Spreadsheet

<table>
<thead>
<tr>
<th>Ecoli CORE Biomass Objective Function</th>
<th>box</th>
<th>Massaction coefficient</th>
<th>v(vt)</th>
<th>w(vt)</th>
<th>composition (molar fraction)</th>
<th>molarG (Calc.)</th>
<th>metabolite</th>
<th>location</th>
<th>type</th>
<th>formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reagent</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NH4 pg 50</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H2O</td>
<td>0.5754</td>
<td>0.35</td>
<td></td>
<td></td>
<td></td>
<td>0.039 0.51039 0.059</td>
<td>cytoplasm</td>
<td>AA</td>
<td>C0H0N2C02</td>
<td></td>
</tr>
<tr>
<td>NH4 pg 36</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.095 0.29572 0.062</td>
<td>cytoplasm</td>
<td>AA</td>
<td>C0H1N4C02</td>
<td></td>
</tr>
<tr>
<td>H2O</td>
<td>0.0453</td>
<td>0.054</td>
<td></td>
<td></td>
<td></td>
<td>0.0453 0.24105 0.072</td>
<td>cytoplasm</td>
<td>AA</td>
<td>C0H1N4C03</td>
<td></td>
</tr>
<tr>
<td>NH4 pg 36</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.046 0.24105 0.072</td>
<td>cytoplasm</td>
<td>AA</td>
<td>C0H1N4C04</td>
<td></td>
</tr>
<tr>
<td>NH4 pg 36</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.047 0.24105 0.072</td>
<td>cytoplasm</td>
<td>AA</td>
<td>C0H1N4C05</td>
<td></td>
</tr>
<tr>
<td>NH4 pg 36</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.048 0.23710 0.072</td>
<td>cytoplasm</td>
<td>AA</td>
<td>C0H1N4C06</td>
<td></td>
</tr>
<tr>
<td>NH4 pg 36</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.049 0.23710 0.072</td>
<td>cytoplasm</td>
<td>AA</td>
<td>C0H1N4C07</td>
<td></td>
</tr>
<tr>
<td>NH4 pg 36</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.049 0.23710 0.072</td>
<td>cytoplasm</td>
<td>AA</td>
<td>C0H1N4C08</td>
<td></td>
</tr>
<tr>
<td>NH4 pg 36</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.050 0.23710 0.072</td>
<td>cytoplasm</td>
<td>AA</td>
<td>C0H1N4C09</td>
<td></td>
</tr>
<tr>
<td>NH4 pg 36</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.051 0.23710 0.072</td>
<td>cytoplasm</td>
<td>AA</td>
<td>C0H1N4C10</td>
<td></td>
</tr>
<tr>
<td>NH4 pg 36</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.052 0.23710 0.072</td>
<td>cytoplasm</td>
<td>AA</td>
<td>C0H1N4C11</td>
<td></td>
</tr>
<tr>
<td>NH4 pg 36</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.053 0.23710 0.072</td>
<td>cytoplasm</td>
<td>AA</td>
<td>C0H1N4C12</td>
<td></td>
</tr>
<tr>
<td>NH4 pg 36</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.054 0.23710 0.072</td>
<td>cytoplasm</td>
<td>AA</td>
<td>C0H1N4C13</td>
<td></td>
</tr>
<tr>
<td>NH4 pg 36</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.055 0.23710 0.072</td>
<td>cytoplasm</td>
<td>AA</td>
<td>C0H1N4C14</td>
<td></td>
</tr>
<tr>
<td>NH4 pg 36</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.056 0.23710 0.072</td>
<td>cytoplasm</td>
<td>AA</td>
<td>C0H1N4C15</td>
<td></td>
</tr>
<tr>
<td>NH4 pg 36</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.057 0.23710 0.072</td>
<td>cytoplasm</td>
<td>AA</td>
<td>C0H1N4C16</td>
<td></td>
</tr>
<tr>
<td>NH4 pg 36</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.058 0.23710 0.072</td>
<td>cytoplasm</td>
<td>AA</td>
<td>C0H1N4C17</td>
<td></td>
</tr>
<tr>
<td>NH4 pg 36</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.059 0.23710 0.072</td>
<td>cytoplasm</td>
<td>AA</td>
<td>C0H1N4C18</td>
<td></td>
</tr>
<tr>
<td>NH4 pg 36</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.060 0.23710 0.072</td>
<td>cytoplasm</td>
<td>AA</td>
<td>C0H1N4C19</td>
<td></td>
</tr>
<tr>
<td>NH4 pg 36</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.061 0.23710 0.072</td>
<td>cytoplasm</td>
<td>AA</td>
<td>C0H1N4C20</td>
<td></td>
</tr>
<tr>
<td>NH4 pg 36</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.062 0.23710 0.072</td>
<td>cytoplasm</td>
<td>AA</td>
<td>C0H1N4C21</td>
<td></td>
</tr>
<tr>
<td>NH4 pg 36</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.063 0.23710 0.072</td>
<td>cytoplasm</td>
<td>AA</td>
<td>C0H1N4C22</td>
<td></td>
</tr>
<tr>
<td>NH4 pg 36</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.064 0.23710 0.072</td>
<td>cytoplasm</td>
<td>AA</td>
<td>C0H1N4C23</td>
<td></td>
</tr>
<tr>
<td>NH4 pg 36</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.065 0.23710 0.072</td>
<td>cytoplasm</td>
<td>AA</td>
<td>C0H1N4C24</td>
<td></td>
</tr>
<tr>
<td>NH4 pg 36</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.066 0.23710 0.072</td>
<td>cytoplasm</td>
<td>AA</td>
<td>C0H1N4C25</td>
<td></td>
</tr>
<tr>
<td>NH4 pg 36</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.067 0.23710 0.072</td>
<td>cytoplasm</td>
<td>AA</td>
<td>C0H1N4C26</td>
<td></td>
</tr>
<tr>
<td>NH4 pg 36</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.068 0.23710 0.072</td>
<td>cytoplasm</td>
<td>AA</td>
<td>C0H1N4C27</td>
<td></td>
</tr>
<tr>
<td>NH4 pg 36</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.069 0.23710 0.072</td>
<td>cytoplasm</td>
<td>AA</td>
<td>C0H1N4C28</td>
<td></td>
</tr>
<tr>
<td>NH4 pg 36</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.070 0.23710 0.072</td>
<td>cytoplasm</td>
<td>AA</td>
<td>C0H1N4C29</td>
<td></td>
</tr>
<tr>
<td>NH4 pg 36</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.071 0.23710 0.072</td>
<td>cytoplasm</td>
<td>AA</td>
<td>C0H1N4C30</td>
<td></td>
</tr>
<tr>
<td>NH4 pg 36</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.072 0.23710 0.072</td>
<td>cytoplasm</td>
<td>AA</td>
<td>C0H1N4C31</td>
<td></td>
</tr>
<tr>
<td>NH4 pg 36</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.073 0.23710 0.072</td>
<td>cytoplasm</td>
<td>AA</td>
<td>C0H1N4C32</td>
<td></td>
</tr>
</tbody>
</table>

Refinement of Reconstruction:

Determine Biomass Composition

24. Determine the chemical composition of the cell, i.e., protein, RNA, DNA, lipids, and cofactor content

25. Determine the amino acid content either experimentally or by estimation

26. The molar percentage and molecular weight of each amino acid must be used to calculate the weight per mol protein

27. Determine the nucleotide content either experimentally or by estimation

28. Calculate the fractional distribution of each nucleotide to the biomass composition

29. Determine the lipid content

30. Determine the content of the soluble pool (polyamines and vitamins and cofactors)

31. Determine the ion content

32. Determine GAM

33. Compile and add biomass reaction to the reconstruction
Determining the Amino Acid Content

Example of Biomass Composition Determination for *Pseudomonas putida* KT2440

Chemical composition of *E. coli* adopted from [11] and utilized as a template for *P. putida* KT2440, since no extensive information was available.

Phospholipid contributions to the biomass function where PE is phosphatidylethanolamine, PG is phosphatidylglycerol, and CL is cardiolipin.

Protein composition in *P. putida* broken down by monomer contributions in mmol/gDW.

dNTP composition of the entire *P. putida* chromosomal genome. The data are obtained from direct measurements, literature, or can be estimated from genome information.

Constraint-based Metabolic Reconstructions & Analysis

H. Scott Hinton, 2016

http://cmr.jcvi.org/cgi-bin/CMR/GenomePage.cgi?org=ntec01
Create Codon Usage Table
# Codon Usage Table

<table>
<thead>
<tr>
<th>2nd</th>
<th>1st</th>
<th>C</th>
<th>A</th>
<th>G</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>TTT</td>
<td>Phe</td>
<td>1.84%</td>
<td>0.69%</td>
</tr>
<tr>
<td></td>
<td>TTC</td>
<td>Phe</td>
<td>1.36%</td>
<td>0.07%</td>
</tr>
<tr>
<td></td>
<td>TTA</td>
<td>Leu</td>
<td>1.66%</td>
<td>0.58%</td>
</tr>
<tr>
<td></td>
<td>TTG</td>
<td>Leu</td>
<td>1.48%</td>
<td>0.66%</td>
</tr>
<tr>
<td>C</td>
<td>CTT</td>
<td>Leu</td>
<td>1.05%</td>
<td>0.72%</td>
</tr>
<tr>
<td></td>
<td>CTC</td>
<td>Leu</td>
<td>1.24%</td>
<td>0.55%</td>
</tr>
<tr>
<td></td>
<td>CTA</td>
<td>Leu</td>
<td>0.27%</td>
<td>0.91%</td>
</tr>
<tr>
<td></td>
<td>CTG</td>
<td>Leu</td>
<td>0.57%</td>
<td>2.31%</td>
</tr>
<tr>
<td>A</td>
<td>ATT</td>
<td>Ile</td>
<td>3.44%</td>
<td>0.98%</td>
</tr>
<tr>
<td></td>
<td>ATC</td>
<td>Ile</td>
<td>2.05%</td>
<td>2.20%</td>
</tr>
<tr>
<td></td>
<td>ATA</td>
<td>Ile</td>
<td>0.16%</td>
<td>0.55%</td>
</tr>
<tr>
<td></td>
<td>ATG</td>
<td>Met</td>
<td>2.70%</td>
<td>1.31%</td>
</tr>
<tr>
<td>G</td>
<td>GTT</td>
<td>Val</td>
<td>1.72%</td>
<td>1.84%</td>
</tr>
<tr>
<td></td>
<td>GTC</td>
<td>Val</td>
<td>1.49%</td>
<td>3.08%</td>
</tr>
<tr>
<td></td>
<td>GTA</td>
<td>Val</td>
<td>1.05%</td>
<td>2.04%</td>
</tr>
<tr>
<td></td>
<td>GTG</td>
<td>Val</td>
<td>0.23%</td>
<td>3.74%</td>
</tr>
</tbody>
</table>

21 total ORFs, 7204 codons

Fields: [Codon Triple] [Codon Frequency] [Number of Codon Occurrences]
## Codon Usage

**Selected Organism:** *Escherichia coli K12-MG1655*  
**Selected Annotation:** Primary Annotation  
**Selected Molecule Name:** Chromosome Escherichia coli K12-MG1655  
**Roles:** Amino acid biosynthesis: Aromatic amino acid family

<table>
<thead>
<tr>
<th>AA Name</th>
<th>Codon Triplet</th>
<th>Number of Triplet Occurrences</th>
<th>% Triplet Frequency</th>
<th>% Triplet Frequency for AA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ala (A)</td>
<td>GCC</td>
<td>222</td>
<td>3.08%</td>
<td>28.76%</td>
</tr>
<tr>
<td>Ala (A)</td>
<td>GCA</td>
<td>147</td>
<td>2.04%</td>
<td>19.04%</td>
</tr>
<tr>
<td>Ala (A)</td>
<td>GCG</td>
<td>270</td>
<td>3.74%</td>
<td>34.97%</td>
</tr>
<tr>
<td>Ala (A)</td>
<td>GCT</td>
<td>133</td>
<td>1.84%</td>
<td>17.23%</td>
</tr>
<tr>
<td><strong>Totals for Ala (A)</strong></td>
<td></td>
<td><strong>772</strong></td>
<td><strong>10.7%</strong></td>
<td><strong>100%</strong></td>
</tr>
<tr>
<td>Arg (R)</td>
<td>CGA</td>
<td>18</td>
<td>0.24%</td>
<td>4.41%</td>
</tr>
<tr>
<td>Arg (R)</td>
<td>AGG</td>
<td>3</td>
<td>0.04%</td>
<td>0.74%</td>
</tr>
<tr>
<td>Arg (R)</td>
<td>CGT</td>
<td>149</td>
<td>2.06%</td>
<td>36.52%</td>
</tr>
<tr>
<td>Arg (R)</td>
<td>CGG</td>
<td>30</td>
<td>0.41%</td>
<td>7.35%</td>
</tr>
<tr>
<td>Arg (R)</td>
<td>AGA</td>
<td>9</td>
<td>0.12%</td>
<td>2.21%</td>
</tr>
<tr>
<td>Arg (R)</td>
<td>CGC</td>
<td>199</td>
<td>2.76%</td>
<td>48.77%</td>
</tr>
<tr>
<td><strong>Totals for Arg (R)</strong></td>
<td></td>
<td><strong>408</strong></td>
<td><strong>5.63%</strong></td>
<td><strong>100%</strong></td>
</tr>
<tr>
<td>Asn (N)</td>
<td>AAC</td>
<td>125</td>
<td>1.73%</td>
<td>47.17%</td>
</tr>
<tr>
<td>Asn (N)</td>
<td>ATT</td>
<td>140</td>
<td>1.94%</td>
<td>52.83%</td>
</tr>
<tr>
<td><strong>Totals for Asn (N)</strong></td>
<td></td>
<td><strong>265</strong></td>
<td><strong>3.67%</strong></td>
<td><strong>100%</strong></td>
</tr>
</tbody>
</table>
Refinement of Reconstruction:

**Determine Biomass Composition**

24. Determine the chemical composition of the cell, i.e., protein, RNA, DNA, lipids, and cofactor content

25. Determine the amino acid content either experimentally or by estimation

26. The molar percentage and molecular weight of each amino acid must be used to calculate the weight per mol protein

27. Determine the nucleotide content either experimentally or by estimation

28. Calculate the fractional distribution of each nucleotide to the biomass composition

29. Determine the lipid content

30. Determine the content of the soluble pool (polyamines and vitamins and cofactors)

31. Determine the ion content

32. Determine GAM

33. Compile and add biomass reaction to the reconstruction
Flow Chart to Calculate the Fractional Contribution of a Precursor to the Biomass Reaction

a. The fractional contribution of alanine.

b. To convert the molar percentage into weight of alanine per mole protein, the molar percentage is multiplied by the molecular weight of alanine. Note that the polymerization of amino acid leads to the loss of a water molecule, which needs to be considered when calculating the molecular weight. Once the weight of amino acid per mole protein is obtained for all amino acids, they are summed to obtain the weight of protein per mole protein.

c. The weight of alanine per mole protein is converted into weight alanine per weight protein by multiplying with the sum of all amino acids' weight.

d. The weight of alanine is multiplied by the cellular content of protein and divided by its molecular weight to obtain the mole alanine per cell dry weight. Multiplying this molar contribution by a factor of 1,000 will result in a final unit of mmol alanine per gram of dry weight.

Refinement of Reconstruction:

Determine Biomass Composition

24. Determine the chemical composition of the cell, i.e., protein, RNA, DNA, lipids, and cofactor content

25. Determine the amino acid content either experimentally or by estimation

26. The molar percentage and molecular weight of each amino acid must be used to calculate the weight per mol protein

27. Determine the nucleotide content either experimentally or by estimation

28. Calculate the fractional distribution of each nucleotide to the biomass composition

29. Determine the lipid content

30. Determine the content of the soluble pool (polyamines and vitamins and cofactors)

31. Determine the ion content

32. Determine GAM

33. Compile and add biomass reaction to the reconstruction
Determine the Nucleotide Content
Example of Biomass Composition Determination for Pseudomonas putida KT 2440

Chemical composition of E. coli adopted from 11 and utilized as a template for P. putida KT2440, since no extensive information was available.

Phospholipid contributions to the biomass function where PE is phosphatidylethanolamine, PG is phosphatidylglycerol, and CL is cardiolipin.

Protein composition in P. putida broken down by monomer contributions in mmol/gDW.

dNTP composition of the entire P. putida chromosomal genome. The data are obtained from direct measurements, literature, or can be estimated from genome information.

Determine the Nucleotide Content and Calculate the Fractional Distribution of Each Nucleotide

- Experimental determination of the nucleotide content,
  ✓ Obtain data for each deoxynucleotide triphosphate (dATP, dCTP, dGTP and dTTP) and each nucleotide triphosphate (ATP, CTP, GTP and UTP).

- Estimation of nucleotide composition from genome information
  ✓ For example, use CMR database. From the Genome Tools tab, select Summary Information, followed by DNA Molecule Info. The number of each dNTP (i.e., dATP, dCTP, dGTP and dTTP) present in the genome is listed on the summary page.

  ✓ To determine the RNA composition of the cell, the codon usage that was accessed for the amino acid content in Step 25 can be used. It must be remembered that RNA incorporates U instead of T; therefore, the codon usage needs to be read with every T replaced by a U.

  ✓ Tabulate the frequency of each nucleotide.

- Calculate the fractional distribution of each nucleotide to the biomass composition
**CMR database** -> **Genome Tools Tab**  
-> **Summary Information** -> **DNA Molecule Info**

<table>
<thead>
<tr>
<th>Molecule Name</th>
<th>chromosome Pseudomonas putida KT2440</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type</td>
<td>chromosome</td>
</tr>
<tr>
<td>Topology</td>
<td>circular</td>
</tr>
<tr>
<td>GenBank Accession.Version</td>
<td>AE015451.1</td>
</tr>
<tr>
<td>Sequence Length</td>
<td>6181803 bp</td>
</tr>
<tr>
<td>Primary Annotation Coding Regions</td>
<td>5415589 bp</td>
</tr>
<tr>
<td>Primary Annotation Intergenic Regions</td>
<td>758274 bp</td>
</tr>
<tr>
<td>Primary Annotation: Number of Genes</td>
<td>5437</td>
</tr>
<tr>
<td>Primary Annotation: Number of Genes assigned to role ids</td>
<td>3050</td>
</tr>
<tr>
<td>Primary Annotation: Number of Genes not assigned to role ids</td>
<td>0</td>
</tr>
<tr>
<td>Primary Annotation: Conserved Hypothetical Genes</td>
<td>299</td>
</tr>
<tr>
<td>Primary Annotation: Hypothetical Genes</td>
<td>598</td>
</tr>
<tr>
<td>Number of A</td>
<td>1186504 bp</td>
</tr>
<tr>
<td>Number of T</td>
<td>1192024 bp</td>
</tr>
<tr>
<td>Number of G</td>
<td>1913381 bp</td>
</tr>
<tr>
<td>Number of C</td>
<td>1889954 bp</td>
</tr>
<tr>
<td>Number of A+T</td>
<td>2378528 bp</td>
</tr>
<tr>
<td>Number of G+C</td>
<td>3803335 bp</td>
</tr>
</tbody>
</table>

**Summary Information**
Refinement of Reconstruction:

Determine Biomass Composition

24. Determine the chemical composition of the cell, i.e., protein, RNA, DNA, lipids, and cofactor content
25. Determine the amino acid content either experimentally or by estimation
26. The molar percentage and molecular weight of each amino acid must be used to calculate the weight per mol protein
27. Determine the nucleotide content either experimentally or by estimation
28. Calculate the fractional distribution of each nucleotide to the biomass composition
29. Determine the lipid content
30. Determine the content of the soluble pool (polyamines and vitamins and cofactors)
31. Determine the ion content
32. Determine GAM
33. Compile and add biomass reaction to the reconstruction
Determine the Lipid Content
Example of Biomass Composition Determination for *Pseudomonas putida* KT 2440

Chemical composition of *E. coli* adopted from 11 and utilized as a template for *P. putida* KT2440, since no extensive information was available.

![Table A](image1)

**Phospholipid contributions to the biomass function where PE is phosphatidylethanolamine, PG is phosphatidylglycerol, and CL is cardiolipin.**

Protein composition in *P. putida* broken down by monomer contributions in mmol/gDW.

![Table B](image2)

Phospholipid composition of the entire *P. putida* chromosomal genome. The data are obtained from direct measurements, literature, or can be estimated from genome information.

![Table D](image3)

Determine Biomass Composition:

Determine the Lipid Content

- Determine the contributions from fatty acids and phospholipids.
  
  i. Determine the average molecular weight of a fatty acid in the cell by incorporating the average fatty acid composition of the cell (requires experimental data, e.g., from literature).
  
  ii. The average molecular weight of each fatty acid must be used.
  
  iii. Add the weight contributions of each fatty acid to determine the average molecular weight for the fatty acid chain.
  
  iv. Use this weight to calculate the average molecular weight of various lipids within the cell. Carry out such a computation by adding the molecular weight of the core structure of the molecule and the molecular weight of the fatty acids attached to the core structure based on the average molecular weight of one fatty acid that was determined above.
  
  v. The molar percentages of the three major phospholipids, phosphatidylethanolamine, phosphatidylglycerol and cardiolipin, may be found in the literature.
  
  vi. Then determine the phospholipid contributions to the biomass function.

Refinement of Reconstruction:

Determine Biomass Composition

24. Determine the chemical composition of the cell, i.e., protein, RNA, DNA, lipids, and cofactor content
25. Determine the amino acid content either experimentally or by estimation
26. The molar percentage and molecular weight of each amino acid must be used to calculate the weight per mol protein
27. Determine the nucleotide content either experimentally or by estimation
28. Calculate the fractional distribution of each nucleotide to the biomass composition
29. Determine the lipid content
30. Determine the content of the soluble pool (polyamines and vitamins and cofactors)
31. Determine the ion content
32. Determine GAM
33. Compile and add biomass reaction to the reconstruction
Determine Biomass Composition:
The soluble pool: contains polyamines, vitamins and cofactors (e.g. E.coli)

<table>
<thead>
<tr>
<th>Abbr</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>putre</td>
<td>Putrescine</td>
</tr>
<tr>
<td>spmd</td>
<td>Spermidine</td>
</tr>
<tr>
<td>accoa</td>
<td>Acetyl-CoA</td>
</tr>
<tr>
<td>coa</td>
<td>Coenzyme A (CoA)</td>
</tr>
<tr>
<td>succoa</td>
<td>Succinyl-CoA</td>
</tr>
<tr>
<td>malcoa</td>
<td>Malonyl-CoA</td>
</tr>
<tr>
<td>nad</td>
<td>Nicotinamide adenine dinucleotide</td>
</tr>
<tr>
<td>nadh</td>
<td>Nicotinamide adenine dinucleotide - reduced</td>
</tr>
<tr>
<td>nadp</td>
<td>Nicotinamide adenine dinucleotide phosphate</td>
</tr>
<tr>
<td>nadph</td>
<td>Nicotinamide adenine dinucleotide phosphate - reduced</td>
</tr>
<tr>
<td>udcpdp</td>
<td>Undecaprenyl diphosphate</td>
</tr>
<tr>
<td>10fthf</td>
<td>10-Formyltetrahydrofolate</td>
</tr>
<tr>
<td>thf</td>
<td>5,6,7,8-Tetrahydrofolate</td>
</tr>
<tr>
<td>mlthf</td>
<td>5,10-Methylenetetrahydrofolate</td>
</tr>
<tr>
<td>5mthf</td>
<td>5-Methyltetrahydrofolate</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Abbr</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>chor</td>
<td>Chorismate</td>
</tr>
<tr>
<td>enter</td>
<td>Enterochelin</td>
</tr>
<tr>
<td>gthrd</td>
<td>Reduced glutathione</td>
</tr>
<tr>
<td>pydx5p</td>
<td>Pyridoxal 5’-phosphate (Vitamin B6)</td>
</tr>
<tr>
<td>amet</td>
<td>S-Adenosyl-L-methionine</td>
</tr>
<tr>
<td>thmpp</td>
<td>Thiamine diphosphate</td>
</tr>
<tr>
<td>adocbl</td>
<td>Adenosylcobalamin</td>
</tr>
<tr>
<td>q8h2</td>
<td>Ubiquinol-8</td>
</tr>
<tr>
<td>2dmmql8</td>
<td>2-Demethylmenaquinol 8</td>
</tr>
<tr>
<td>mql8</td>
<td>Menaquinol 8</td>
</tr>
<tr>
<td>hemeO</td>
<td>Heme O</td>
</tr>
<tr>
<td>pheme</td>
<td>Protoheme</td>
</tr>
<tr>
<td>sheme</td>
<td>Siroheme</td>
</tr>
<tr>
<td>ribflv</td>
<td>Riboflavin</td>
</tr>
<tr>
<td>fad</td>
<td>Flavin adenine dinucleotide oxidized</td>
</tr>
</tbody>
</table>

Determine Biomass Composition:

Determine the Content of the Soluble Pool
(polyamines and vitamins and cofactors)

\[ \text{Chemical formula} \]

\[ \times V_{\text{DW}} \]

\[ \times M_W \]

\[ \times 10^3 \times M_{\text{DW}} \]

\[ \times N_A \]

\[ \times M_{\text{DW}} \]

\[ X \text{ mmol liter}^{-1} \]

\[ X \text{ molecules} / \text{cell} \]

\[ X \text{ mg g}_{\text{DW}}^{-1} \]

\[ (X \times V_{\text{DW}} \times M_{\text{DW}} \times 10^{-3}) \text{ g}_{\text{DW}}^{-1} \]

\[ (X \times 10^3(N_A \times M_{\text{DW}})) \text{ mmol g}_{\text{DW}}^{-1} \]

\[ X/M_W \text{ mmol g}_{\text{DW}}^{-1} \]

\[ X \times 10^3 \text{ g}_{\text{DW}}^{-1} \]

Conversion factors:

- Average cell aqueous volume: \( V_c = 6.7 \times 10^{-18} \text{ liter cell}^{-1} \)
- Average dry mass: \( M_{\text{DW}} = 3 \times 10^{-13} \text{ g}_{\text{DW}} \text{ cell}^{-1} \)
- Avogadro's number: \( N_A = 6.02 \times 10^{23} \text{ molecules mol}^{-1} \)

---

Refinement of Reconstruction:  
**Determine Biomass Composition**

24. Determine the chemical composition of the cell, i.e., protein, RNA, DNA, lipids, and cofactor content
25. Determine the amino acid content either experimentally or by estimation
26. The molar percentage and molecular weight of each amino acid must be used to calculate the weight per mol protein
27. Determine the nucleotide content either experimentally or by estimation
28. Calculate the fractional distribution of each nucleotide to the biomass composition
29. Determine the lipid content
30. Determine the content of the soluble pool (polyamines and vitamins and cofactors)
31. Determine the ion content
32. Determine GAM
33. Compile and add biomass reaction to the reconstruction
Determine Biomass Composition:

Determine the Ion Content

- Calculate the molar fraction of the ions.
- Assume that concentration data are available or can be estimated for each ion.
- Convert the reported concentration ($c_i$) for each ion species $i$ into mM. Add all the ion species (total ion concentration, $c_{total}$). Calculate the molar fraction ($f_i$) of each ion species $i$ by dividing $c_i$ with $c_{total}$:

\[
 f_i = \frac{c_i}{c_{total}} \quad \text{where} \quad c_{total} = \sum c_i
\]

Refinement of Reconstruction:

Determine Biomass Composition

24. Determine the chemical composition of the cell, i.e., protein, RNA, DNA, lipids, and cofactor content
25. Determine the amino acid content either experimentally or by estimation
26. The molar percentage and molecular weight of each amino acid must be used to calculate the weight per mol protein
27. Determine the nucleotide content either experimentally or by estimation
28. Calculate the fractional distribution of each nucleotide to the biomass composition
29. Determine the lipid content
30. Determine the content of the soluble pool (polyamines and vitamins and cofactors)
31. Determine the ion content
32. Determine GAM
33. Compile and add biomass reaction to the reconstruction
Determination of **Growth-associated Maintenance (GAM) Cost**

**a** Biosynthetic cost: required energy (in ~P) per cellular content of macromolecules:

- **Protein**
  - % of dry weight
  - mmol Protein / gDW = Σ mmol amino acids / gDW
  - mmol ~P / mmol
  - *C_P

- **RNA**
  - % of dry weight
  - mmol RNA / gDW = Σ mmol NTP / gDW
  - mmol ~P / mmol
  - *C_R

- **DNA**
  - % of dry weight
  - mmol DNA / gDW = Σ mmol dNTP / gDW
  - mmol ~P / mmol
  - *C_D

Total ~P required / gDW: Σ mmol ~P / gDW

**b**

<table>
<thead>
<tr>
<th></th>
<th>wt %</th>
<th>Total mmol</th>
<th>mmol ~P / mmol</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Protein</td>
<td>0.563</td>
<td>5.197</td>
<td>C_P = 4.324</td>
<td>22.472</td>
</tr>
<tr>
<td>DNA</td>
<td>0.031</td>
<td>0.101</td>
<td>C_D = 1.385</td>
<td>0.138</td>
</tr>
<tr>
<td>RNA</td>
<td>0.21</td>
<td>0.649</td>
<td>C_R = 0.406</td>
<td>0.284</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td>22.873</td>
</tr>
</tbody>
</table>

Refinement of Reconstruction:

**Determine Biomass Composition**

24. Determine the chemical composition of the cell, i.e., protein, RNA, DNA, lipids, and cofactor content
25. Determine the amino acid content either experimentally or by estimation
26. The molar percentage and molecular weight of each amino acid must be used to calculate the weight per mol protein
27. Determine the nucleotide content either experimentally or by estimation
28. Calculate the fractional distribution of each nucleotide to the biomass composition
29. Determine the lipid content
30. Determine the content of the soluble pool (polyamines and vitamins and cofactors)
31. Determine the ion content
32. Determine GAM
33. Compile and add biomass reaction to the reconstruction
Determine Biomass Composition:

Compile and Add Biomass Reaction To The Reconstruction

- All precursors are assembled in one single reaction, the biomass reaction, which is then added to the reaction list of the reconstruction.

- Add GAM to biomass reaction as follows:
  - $x \text{ ATP} + x \text{ H}_2\text{O} \rightarrow x \text{ ADP} + x \text{ P}_i + x \text{ H}^+$,
  - where $x$ is the number of required phosphate bonds.

- CRITICAL STEP: It is to be noted that some metabolites might be produced. For instance, in the *E. coli* biomass reaction, proton ($\text{H}^+$), orthophosphate ($\text{P}_i$) and some other metabolites are produced. These metabolites originate mainly from the growth-associated ATP hydrolysis.

$$Z(\text{ecoli_core_model}) = (1.496) \text{ 3pg} + (3.7478) \text{ accoa} + (59.8100) \text{ atp} + (0.3610) \text{ e4p} + (0.0709) \text{ f6p} + (0.1290) \text{ g3p} + (0.2050) \text{ g6p} + (0.2557) \text{ gln-L} + (4.9414) \text{ glu-L} + (59.8100) \text{ h2o} + (3.5470) \text{ nad} + (13.0279) \text{ nadph} + (1.7867) \text{ oaa} + (0.5191) \text{ pep} + (2.8328) \text{ pyr} + (0.8977) \text{ r5p} \rightarrow (59.8100) \text{ adp} + (4.1182) \text{ akg} + (3.7478) \text{ coa} + (59.8100) \text{ h} + (3.5470) \text{ nadh} + (13.0279) \text{ nadp} + (59.8100) \text{ pi}$$

iaf1260 BIOMASS OBJECTIVE FUNCTION
(Ec_biomass_iAF1260_core_59p81M)

\[
Z = 0.000223 \text{10fthf}[c] + 0.000223 \text{2ohph}[c] + 0.5137 \text{ala-L}[c] + 0.000223 \text{amet}[c] + 0.2958 \text{arg-L}[c] + 0.2411 \text{asn-L}[c] + 0.2411 \text{asp-L}[c] + 59.984 \text{atp}[c] + 0.004737 \text{ca2}[c] + 0.004737 \text{cl}[c] + 0.000576 \text{coa}[c] + 0.003158 \text{cobalt2}[c] + 0.1335 \text{ctp}[c] + 0.003158 \text{cu2}[c] + 0.09158 \text{cys-L}[c] + 0.02617 \text{datp}[c] + 0.02702 \text{dctp}[c] + 0.02702 \text{dgtp}[c] + 0.02617 \text{dttp}[c] + 0.000223 \text{fad}[c] + 0.007106 \text{fe2}[c] + 0.007106 \text{fe3}[c] + 0.2632 \text{gln-L}[c] + 0.2632 \text{glu-L}[c] + 0.6126 \text{gly}[c] + 0.2151 \text{gtp}[c] + 54.462 \text{h2o}[c] + 0.09474 \text{his-L}[c] + 0.2905 \text{ile-L}[c] + 0.1776 \text{k}[c] + 0.01945 \text{kdo2lipid4}[e] + 0.4505 \text{leu-L}[c] + 0.3432 \text{lys-L}[c] + 0.1537 \text{met-L}[c] + 0.07895 \text{mg2}[c] + 0.000223 \text{mlthf}[c] + 0.003158 \text{mn2}[c] + 0.003158 \text{mobd}[c] + 0.01389 \text{murein5px4p}[p] + 0.001831 \text{nad}[c] + 0.00447 \text{nadp}[c] + 0.011843 \text{nh4}[c] + 0.02233 \text{pe160}[c] + 0.04148 \text{pe160}[p] + 0.02632 \text{pe161}[c] + 0.04889 \text{pe161}[p] + 0.1759 \text{phe-L}[c] + 0.000223 \text{pheme}[c] + 0.2211 \text{pro-L}[c] + 0.000223 \text{pydx5p}[c] + 0.000223 \text{ribflv}[c] + 0.2158 \text{ser-L}[c] + 0.000223 \text{sheme}[c] + 0.003948 \text{so4}[c] + 0.000223 \text{thf}[c] + 0.000223 \text{thmpp}[c] + 0.2537 \text{thr-L}[c] + 0.05684 \text{trp-L}[c] + 0.1379 \text{tyr-L}[c] + 5.5e-005 \text{udcpdp}[c] + 0.1441 \text{utp}[c] + 0.4232 \text{val-L}[c] + 0.003158 \text{zn2}[c] \rightarrow 59.81 \text{adp}[c] + 59.81 \text{h}[c] + 59.806 \text{pi}[c] + 0.7739 \text{ppi}[c]
\]
Refinement of Reconstruction

6. Determine and verify substrate and cofactor usage
7. Obtain a neutral formula for each metabolite in the reaction
8. Determine the charged formula for each metabolite in the reaction.
9. Calculate reaction stoichiometry
10. Determine reaction directionality
11. Add information for gene and reaction localization
12. Add subsystem information to the reaction
13. Verify GPR association
14. Add metabolite identifier
15. Determine and add the confidence score
16. Add references and notes
17. Repeat Steps 6–17 for all those draft reconstruction genes
18. Add spontaneous reactions
19. Add extracellular and periplasmic transport reactions
20. Add exchange reactions
21. Add intracellular transport reactions
22. Draw metabolic map (optional)
23-33. Determine biomass composition
34. Add NGAM Reaction (ATPM)
35. Add demand reactions
36. Add sink reactions
37. Determine growth medium requirements
Refinement of Reconstruction:

**Add Non-GAM (NGAM) Reactions**

- Add the following reaction to the reconstruction reaction list:
  - \( \text{ATPM}: \text{1 ATP} + \text{1 H}_2\text{O} \rightarrow \text{1 ADP} + \text{1 P}_i + \text{1 H}^+ \).
  - Represents NGAM requirements of the cell to maintain, e.g., turgor pressure.
- The value for the reaction rate can be estimated from growth experiments. For example, based on such measurements, the reaction flux rate was constrained to 8.39 mmol gDW\(^{-1}\) h\(^{-1}\) in the *E. coli* metabolic model.
- The best way to obtain accurate information regarding GAM and NGAM is by plotting growth data obtained from chemostat growth experiments. GAM and NGAM can be directly read from the plot.

Refinement of Reconstruction

6. Determine and verify substrate and cofactor usage.
7. Obtain a neutral formula for each metabolite in the reaction.
8. Determine the charged formula for each metabolite in the reaction.
9. Calculate reaction stoichiometry.
10. Determine reaction directionality.
11. Add information for gene and reaction localization.
12. Add subsystem information to the reaction.
13. Verify GPR association.
15. Determine and add the confidence score.
16. Add references and notes.
17. Repeat Steps 6–17 for all those draft reconstruction genes.
18. Add spontaneous reactions.
19. Add extracellular and periplasmic transport reactions.
20. Add exchange reactions.
22. Draw metabolic map (optional).
23–33. Determine biomass composition.
34. Add NGAM Reaction (ATPM).
35. Add demand reactions.
36. Add sink reactions.
37. Determine growth medium requirements.
Refinement of Reconstruction: Add Demand Reactions

• Demand reactions are unbalanced network reactions that allow the accumulation of a compound, which otherwise is not allowed in steady-state models because of mass-balancing requirements (i.e., in steady state the sum of influx equals the sum of efflux for each metabolite).

• In general, metabolic reconstructions contain only few demand reactions.

• Most of the demand reactions will be added in the gap-filling process.

• At this stage, demand functions should only be added for compounds that are known to be produced by the organism, e.g., certain cofactors, lipopolysaccharide and antigens, but
  ✓ for which no information is available about their fractional distribution to the biomass or
  ✓ which may only be produced in some environmental conditions. By including a demand reaction for a particular metabolite one can turn otherwise blocked reactions (cannot carry flux) into active reactions (can carry flux).

• During the debugging- and network-evaluation process, demand reactions may temporarily be added to the model to test or verify certain metabolic functions.

Refinement of Reconstruction

6. Determine and verify substrate and cofactor usage.
7. Obtain a neutral formula for each metabolite in the reaction.
8. Determine the charged formula for each metabolite in the reaction.
9. Calculate reaction stoichiometry.
10. Determine reaction directionality.
11. Add information for gene and reaction localization.
12. Add subsystem information to the reaction.
13. Verify GPR association.
15. Determine and add the confidence score.
16. Add references and notes.
17. Repeat Steps 6-17 for all those draft reconstruction genes.
18. Add spontaneous reactions.
19. Add extracellular and periplasmic transport reactions.
20. Add exchange reactions.
22. Draw metabolic map (optional).
34. Add NGAM Reaction (ATPM).
35. Add demand reactions.
36. Add sink reactions.
37. Determine growth medium requirements.
Refinement of Reconstruction:

Add Sink Reactions

- Sink reactions are similar to demand reactions but are defined to be reversible and thus provide the network with metabolites.

- These sink reactions are of great use for compounds that are produced by nonmetabolic cellular processes but that need to be metabolized.

- Adding too many sink reactions may enable the model to grow without any resources in the medium. Therefore, sink reactions have to be added with care. As for demand reactions, sink reactions are mostly used during the debugging process.

- They help in identifying the origin of a problem (e.g., why a metabolite cannot be produced).

- These sink reactions are functionally replaced by filling the identified gap.

Refinement of Reconstruction

6. Determine and verify substrate and cofactor usage.
7. Obtain a neutral formula for each metabolite in the reaction.
8. Determine the charged formula for each metabolite in the reaction.
9. Calculate reaction stoichiometry.
10. Determine reaction directionality.
11. Add information for gene and reaction localization.
12. Add subsystem information to the reaction.
13. Verify GPR association.
15. Determine and add the confidence score.
16. Add references and notes.
17. Repeat Steps 6-17 for all those draft reconstruction genes.
18. Add spontaneous reactions.
19. Add extracellular and periplasmic transport reactions.
20. Add exchange reactions.
22. Draw metabolic map (optional).
34. Add NGAM Reaction (ATPM).
35. Add demand reactions.
36. Add sink reactions.
37. Determine growth medium requirements.
Refinement of Reconstruction:

Determine Growth Medium Requirements

- Information about growth-enabling media should be collected before the conversion and debugging stage. The following information should be collected:
  
  1. Which metabolites are present?
  2. Are there any auxotrophies?
  3. The definition of a base medium composition, e.g., water, protons, ions and so on.
  4. Information about rich medium composition.

- Uptake or secretion rates should be documented and collected.

GENOME-SCALE METABOLIC RECONSTRUCTIONS

• Overview
• Draft Reconstruction
• Refinement of Reconstruction
→ Conversion of Reconstruction into Computable Format
• Network Evaluation
• Data Assembly and Dissemination

Reconstruction Process: 96 Step Protocol


1. Draft Reconstruction
1. Obtain genome annotation.
2. Identify candidate metabolic functions.
3. Obtain candidate metabolic reactions.
4. Assembly of draft reconstruction.
5. Collect of experimental data.

2. Refinement of reconstruction
6. Determine and verify substrate and cofactor usage.
7. Obtain neutral formula for each metabolite.
8. Determine the charged formula.
9. Calculate reaction stoichiometry.
10. Determine reaction c/e/rectinality.
11. Add information for gene and reaction localization.
12. Add subsystems information.
15. Determine and add confidence score.
16. Add references and notes.
17.旗 information from other organisms.
18. Repeat Step 6 to 17 for all genes.
19. Add spontaneous reactions to the reconstruction.
20. Add extracellular and periplasmic transport reactions.
22. Add intracellular transport reactions.
23. Draw metabolic map (optional).
33. Add biomass reaction.
34. Add ATP maintenance reaction (ATPM).
35. Add demand reactions.
36. Add sink reactions.
37. Determine growth medium requirements.

3. Conversion of reconstruction into computable format
38. Initialize the COBRA toolbox.
39. Load reconstruction into Matlab.
40. Verify S matrix.
41. Set objective function.
42. Set simulation constraints.

4. Network evaluation
43-44 Test if network is mass- and charge balanced.
45. Identify metabolic dead-ends.
46-48 Gap analysis.
49. Add missing exchange reactions to model.
50. Set exchange constraints for a simulation condition.
51-58 Test for stoichiometrically balanced cycles.
59. Re-compute gap list.
60-65 Test if biomass precursors can be produced in standard medium.
66. Test if biomass precursors can be produced in other growth media.
67-75 Test if model can produce known secreted products.
76-78 Check for blocked reactions.
79-80 Compute single gene deletion phenotypes.
81-82 Test for known inabilities of the organism.
83. Compare predicted physiological properties with known properties.
84-87 Test if the model can grow fast enough.
88-94 Test if the model grows too fast.

Data assembly and Dissemination
95. Print Matlab model content.
96. Add gap information to the reconstruction output.
Stage 3:

Conversion from Reconstruction to Mathematical Model

38. Initialize the COBRA toolbox
39. Load reconstruction in Matlab
40. Verify S matrix
41. Set objective function
42. Set simulation constraints
Assembly and Representation

Conversion from Reconstruction to Mathematical Model

38. Initialize the COBRA toolbox
   • initCobraToolbox.m

39. Load reconstruction in Matlab
   • model = xlsmodel(RxnFileName, MetFileName);
   • model = xls2model('Model_Filename.xls');

40. Verify S matrix
   • Spy(S)

41. Set objective function
   • model = changeObjective(model, 'ObjectiveFunction');
Stage 3: Conversion from Reconstruction to Mathematical Model.

38. Initialize the COBRA toolbox
39. Load reconstruction in Matlab
40. Verify S matrix
41. Set objective function
42. Set simulation constraints
Conversion from Reconstruction to Mathematical Model:

Set Simulation Constraints

1. Use the following function to set the constraints of the model:

   ```javascript
   model = changeRxnBounds(model, rxnNameList, value, boundType);
   ```

2. The list of reactions for which the bounds should be changed is given by 'rxnNameList', whereas an array contains the new boundary reaction rates ('value'). This type of bound can be set to lower bound ('l') or upper bound ('u'). Alternatively, both bounds can be changed ('b').

3. Use the following command to list all constrained reactions that are greater than a minimal value ('MinInf') and smaller than a maximal value ('MaxInf'):

   ```javascript
   printConstraints(model, MinInf, MaxInf)
   ```

4. In addition, there is a function available that lists all reactions and their flux values in a solution ('fluxData'):

   ```javascript
   printFluxVector(model, fluxData)
   ```

GENOME-SCALE METABOLIC RECONSTRUCTIONS

- Overview
- Draft Reconstruction
- Refinement of Reconstruction
- Conversion of Reconstruction into Computable Format
- Network Evaluation
- Data Assembly and Dissemination

Reconstruction Process: 96 Step Protocol

1. Draft Reconstruction
   1. Obtain genome annotation.
   2. Identify candidate metabolic functions.
   3. Obtain candidate metabolic reactions.
   4. Assembly of draft reconstruction.
   5. Collect of experimental data.

2. Refinement of reconstruction
   6. Determine and verify substrate and cofactor usage.
   7. Obtain neutral formula for each metabolite.
   8. Determine the charged formula.
   9. Calculate reaction stoichiometry.
   10. Determine reaction c/reactivity.
   11. Add information for gene and reaction localization.
   12. Add subsystems information.
   15. Determine and add confidence score.
   16. Add references and notes.
   17. Flag information from other organisms.
   18. Repeat Step 6 to 17 for all genes.
   19. Add spontaneous reactions to the reconstruction.
   20. Add extracellular and periplasmic transport reactions.
   22. Add intracellular transport reactions.
   23. Draw metabolic map (optional).
   24.-32. Determine biomass composition.
   33. Add biomass reaction.
   34. Add ATP maintenance reaction (ATPM).
   35. Add demand reactions.
   36. Add sink reactions.
   37. Determine growth medium requirements.

3. Conversion of reconstruction into computable format
   38. Initialize the COBRA toolbox.
   39. Load reconstruction into Matlab.
   40. Verify S matrix.
   41. Set objective function.
   42. Set simulation constraints.

4. Network evaluation
   43.-44. Test if network is mass- and charge balanced.
   45. Identify metabolic dead-ends.
   49. Add missing exchange reactions to model.
   50. Set exchange constraints for a simulation condition.
   51-58. Test for stoichiometrically balanced cycles.
   59. Re-compute gap list.
   60-65. Test if biomass precursors can be produced in standard medium.
   66. Test if biomass precursors can be produced in other growth media.
   67-75. Test if model can produce known secretion products.
   76-78. Check for blocked reactions.
   81-82. Test for known inabilities of the organism.
   83. Compare predicted physiological properties with known properties.
   84-87. Test if the model can grow fast enough.
   88-94. Test if the model grows too fast.

Data assembly and Dissemination
95. Print Matlab model content.
96. Add gap information to the reconstruction output.

Stage 4: 
Network Evaluation: "Debugging Mode"

- The fourth stage in the reconstruction process consists of network verification, evaluation and validation.
- Common error modes in metabolic reconstructions are listed in Table.
- The metabolic model is tested for its ability to synthesize biomass precursors (such as amino acids, nucleotides triphosphates and lipids).
- This evaluation generally leads to the identification of missing metabolic functions in the reconstruction, so-called network gaps, which can then be added.
- The reconstruction process is an iterative procedure.

<table>
<thead>
<tr>
<th>Error Mode</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wrong reaction constraints</td>
<td>Check reaction constraints if they are applied correctly</td>
</tr>
<tr>
<td>Missing transport reactions</td>
<td>Add transport reactions</td>
</tr>
<tr>
<td>Missing exchange reactions</td>
<td>Add exchange reactions</td>
</tr>
<tr>
<td>Cofactor cannot be consumed or produced</td>
<td>Follow Figure 13 (Thiele, 2010)</td>
</tr>
<tr>
<td>Shuttling of compounds across compartment</td>
<td>Adjust reversibility of transport reactions</td>
</tr>
</tbody>
</table>

Stage 4: Network Evaluation

43-44. Test if network is mass-and charge balanced.
45. Identify metabolic dead-ends.
49. Add missing exchange reactions to model.
50. Set exchange constraints for a simulation condition.
51-58. Test for stoichiometrically balanced cycles.
59. Re-compute gap list.
60-65. Test if biomass precursors can be produced in standard medium.
66. Test if biomass precursors can be produced in other growth media.
67-75. Test if the model can produce known secretion products.
76-78. Check for blocked reactions.
81-82. Test for known incapability’s of the organism.
83. Compare predicted physiological properties with known properties.
84-87. Test if the model can grow fast enough.
88-94. Test if the model grows too fast.
Network Evaluation:

Test if Network is Mass-and Charge Balanced

• Check for stoichiometrically unbalanced reactions.

• Use the “CheckMassChargeBalance” function to check for unbalanced reactions.

  \[
  \text{[massImbalance,imBalancedMass,imBalancedCharge,imBalancedBool,Elements]} = \text{checkMassChargeBalance(model)}
  \]

• In case of unbalanced reactions, the function returns a structure containing the name of the unbalanced reaction and which elements are unbalanced ('UnbalancedRxns').

checkMassChargeBalance Example

• Change the formula for two reactions in “ecoli_iaf1260_MB.xls”
  ✓ Arsenate reductase (ASR) - Add an H$_2$O
    ▪ Model reaction number = 371; Spreadsheet row number = 372
    ▪ From: aso4[c] + 2 gthrd[c] → aso3[c] + gthox[c] + h2o[c]
    ▪ To: aso4[c] + 2 gthrd[c] → aso3[c] + gthox[c] + 2 h2o[c]
  ✓ Arginine succinyltransferase (AST) - Add a proton
    ▪ Model reaction number = 372; Spreadsheet row number = 373
    ▪ To: arg-L[c] + succoa[c] → coa[c] + 2 h[c] + sucarg[c]

• Change the metabolite charged formula
  ✓ Acetate (ac[c]) - Add an oxygen atom
    ▪ Model metabolite number = 242; Spreadsheet row number = 295
    ▪ From: C2H3O2
    ▪ To: C2H3O3
checkMassChargeBalance Example Code

MassChargeBalance_iaf1260_MB.m

% MassChargeBalance_iaf1260_MB.m

clear;

% Input the modified E.coli core model

model = xls2model('ecoli_iaf1260_MB.xls');

% Check mass & charge balance

MassChargeBalance_iaf1260_MB.m Output

>> [...] = checkMassChargeBalance(model)
Assuming biomass reaction is: Ec_biomass_iAF1260_core_59p81M
ATP maintenance reaction is not considered an exchange reaction by default.

Checked element H
Checking element C
Checking element O
Checking element P
Checking element S
Checking element N
Checking element Mg
Checking element X
Checking element Fe
Checking element Zn
Checking element Co
Checking element R

Element Matrix

\[
\begin{array}{c}
H = 1 \\
C = 2 \\
O = 3 \\
P = 4 \\
S = 5 \\
N = 6 \\
Mg = 7 \\
X = 8 \\
Fe = 9 \\
Zn = 10 \\
Co = 11 \\
R = 12 \\
\end{array}
\]

massImbalance =

\[
\begin{array}{c}
(371,1) \quad 2 \\
(372,1) \quad 1 \\
(144,3) \quad 1 \\
(167,3) \quad -1 \\
(187,3) \quad 1 \\
(195,3) \quad -1 \\
(198,3) \quad 1 \\
(199,3) \quad 1 \\
(227,3) \quad 1 \\
(276,3) \quad 1 \\
(277,3) \quad 1 \\
(371,3) \quad 1 \\
(386,3) \quad 1 \\
(429,3) \quad 1 \\
(507,3) \quad 1 \\
(1409,3) \quad 1 \\
(1708,3) \quad 1 \\
(2011,3) \quad 1 \\
(2324,3) \quad 1 \\
\end{array}
\]

2 Extra Protons in Reaction 371
1 Extra Protons in Reaction 372
1 Extra Oxygen in Reaction 371
MassChargeBalance_iaf1260_MB.m Example:

**Printing “UnbalancedRxns” Matrix Formulas**

Use “printRxnFormula” function to find the reaction formulas for the identified reactions

```
>> printRxnFormula(model,model.rxns(371))
ASR  aso4[c] + 2.000000 gthrd[c] \rightarrow 2.000000 h2o[c] + aso3[c] + gthox[c]  (b3503) and (b1064)
ans =
    'aso4[c] + 2 gthrd[c] \rightarrow 2 h2o[c] + aso3[c] + gthox[c]'

>> printRxnFormula(model,model.rxns(372))
AST  succoa[c] + arg-L[c] \rightarrow 2.000000 h[c] + coa[c] + sucarg[c]  (b1747)
ans =
    'succoa[c] + arg-L[c] \rightarrow 2 h[c] + coa[c] + sucarg[c]'
```

Reaction Name
2 Extra protons plus 1 extra oxygen implies an extra H₂O

Reaction Formula
### Remaining Matrix Formulas

**MassChargeBalance_iaf1260_MB.m Example:**

```matlab
printRxnFormula(model, model.rxns(Reaction Index))
```

<table>
<thead>
<tr>
<th>Reaction Index</th>
<th>Reaction</th>
<th>Reaction Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>144</td>
<td>ACACCT</td>
<td>acac(c) + acoa(c) -&gt; aacoa(c) + ac(c)</td>
</tr>
<tr>
<td>167</td>
<td>ACKr</td>
<td>atp(c) + ac(c) &lt;=&gt; adp(c) + actp(c)</td>
</tr>
<tr>
<td>187</td>
<td>ACODA</td>
<td>h2o(c) + acorn(c) -&gt; ac(c) + orn(c)</td>
</tr>
<tr>
<td>195</td>
<td>ACS</td>
<td>atp(c) + ac(c) + coa(c) -&gt; amp(c) + ppi(c) + acoa(c)</td>
</tr>
<tr>
<td>198</td>
<td>ACT2rpp</td>
<td>h[p] + ac[p] &lt;=&gt; h[c] + ac[c]</td>
</tr>
<tr>
<td>199</td>
<td>ACT4pp</td>
<td>ac[p] + na1[p] -&gt; ac[c] + na1[c]</td>
</tr>
<tr>
<td>227</td>
<td>AGDC</td>
<td>h2o(c) + acgam6p[c] -&gt; ac(c) + gam6p[c]</td>
</tr>
<tr>
<td>276</td>
<td>ALDD2x</td>
<td>h2o(c) + nad(c) + acald[c] -&gt; 2 h[c] + nadh[c] + ac[c]</td>
</tr>
<tr>
<td>277</td>
<td>ALDD2y</td>
<td>h2o(c) + nadp(c) + acald[c] -&gt; 2 h[c] + nadph[c] + ac[c]</td>
</tr>
<tr>
<td>386</td>
<td>BUTCT</td>
<td>acoa(c) + but[c] -&gt; ac(c) + btcoa[c]</td>
</tr>
<tr>
<td>429</td>
<td>CITL</td>
<td>cit[c] -&gt; ac[c] + oaa[c]</td>
</tr>
<tr>
<td>507</td>
<td>CYSS</td>
<td>acser[c] + h2s[c] -&gt; h[c] + ac[c] + cys-L[c]</td>
</tr>
<tr>
<td>1409</td>
<td>HXCT</td>
<td>acoa[c] + hxa[c] -&gt; ac[c] + hxcoa[c]</td>
</tr>
<tr>
<td>1708</td>
<td>NACODA</td>
<td>h2o[c] + acg5sa[c] -&gt; ac[c] + glu5sa[c]</td>
</tr>
<tr>
<td>2011</td>
<td>POX</td>
<td>h2o[c] + pyr[c] + q8[c] -&gt; co2[c] + ac[c] + q8h2[c]</td>
</tr>
<tr>
<td>2324</td>
<td>UHGADA</td>
<td>h2o[c] + u3aga[c] -&gt; ac[c] + u3hga[c]</td>
</tr>
</tbody>
</table>

ac[c] is involved in every unbalanced equation; A good candidate to check for an incorrect metabolite charged formula.
Stage 4: Network Evaluation

43-44. Test if network is mass-and charge balanced.
45. Identify metabolic dead-ends.
49. Add missing exchange reactions to model.
50. Set exchange constraints for a simulation condition.
51-58. Test for stoichiometrically balanced cycles.
59. Re-compute gap list.
60-65. Test if biomass precursors can be produced in standard medium.
66. Test if biomass precursors can be produced in other growth media.
67-75. Test if the model can produce known secretion products.
76-78. Check for blocked reactions.
81-82. Test for known incapability’s of the organism.
83. Compare predicted physiological properties with known properties.
84-87. Test if the model can grow fast enough.
88-94. Test if the model grows too fast.
A dead-end metabolite can only be produced or consumed in a given network. Although many dead-end metabolites that create network gaps can be connected to the network by re-evaluating genomic and experimental data, some dead-end metabolites will remain in the refined, curated reconstruction. These dead-end metabolites can be categorized into two groups, depending on the type of reactions that could connect them to the remaining network: knowledge gaps and scope gaps. The knowledge gaps represent the missing biochemical knowledge for the target organism. In contrast, the scope gaps include reactions and cellular processes, which are currently not accounted for in the metabolic reconstruction.

Identifying Gaps: Connectivity-based Approach

• There are at least two approaches to identify gaps in the reconstruction. In the connectivity-based approach, one can count the nonzero entries in each row of the stoichiometric (S) matrix and identify those metabolites, which are only produced or consumed.

• In the example, metabolite D is only produced by reaction $v_3$ and the S matrix contains only one entry in the row corresponding to metabolite D.

Identifying Gaps: Functionality-based Approach

- A second approach is based on model functionality; in this approach the model capability to carry flux through every network reaction is tested. This approach identifies blocked reactions, which are directly or indirectly associated with one or more dead-end metabolites.

- In the shown example, one would not identify metabolite E as a dead-end metabolite with the connectivity-based approach, as it is produced and consumed in the network. However, testing for flux through reactions containing E will show that reaction v_3 and b_3 cannot carry any flux in this model.

- Reactions v_3 and b_3 cannot carry any flux in this network as the metabolite 'E' is unbalanced.
- These reactions are also called 'blocked reactions'.
- Topological analysis would not have identified 'E' as a dead-end metabolite, as reaction v_3 is producing the metabolite.
- Flux variability analysis can be used to identify block reaction in the network.

Network Evaluation:

**Gap Types**

- Gaps in metabolic reconstructions are manifested as:
  - metabolites which cannot be produced by any of the reactions or imported through any of the available uptake pathways in the model are called **root no-production metabolites** (e.g., metabolite A); or
  - metabolites that are not consumed by any of the reactions in the network or exported based on any existing secretion pathways are called **root no-consumption metabolites** (e.g., metabolite B).

- The lack of flow in root no-production metabolites and root no-consumption metabolites is propagated downstream/upstream respectively giving rise to additional metabolites that cannot carry any flow. We refer to these metabolites that are indirectly prevented from carrying flow as:
  - **downstream no-production metabolites** (e.g., metabolite C) and
  - **upstream no-consumption metabolites** (e.g., metabolite D).

- By restoring connectivity for the root problem metabolites, most upstream/downstream metabolites are automatically fixed.

Network Evaluation:

Identify Metabolic Dead-ends: gapFind

• Use “gapFind” to identify the gaps

\[ \text{allGaps, rootGaps, downstreamGaps} = \text{gapFind(model, true, false)} \]

• where
  ✓ allGaps - all gaps found by GapFind
  ✓ rootGaps - all root no-production (and consumption) gaps
  ✓ downstreamGaps - all downstream gaps

% GapFindExample.m

% Input the E.coli core model
model = readCbModel('ecoli_textbook');

% Run gapFind
[allGaps,rootGaps,downstreamGaps] = gapFind(model,true,false)

FBAsolution = optimizeCbModel(model,'max');

% Plot connectivity to downstream gaps. Radius = 1
[involvedRxns,involvedMets,deadEnds] = draw_by_met (model,{'fru[e]'},...
   true,1,'struc',{},FBAsolution.x);
[involvedRxns,involvedMets,deadEnds] = draw_by_met (model,{'fum[e]'},...
   true,1,'struc',{},FBAsolution.x);
[involvedRxns,involvedMets,deadEnds] = draw_by_met (model,{'gln-L[e]'},...
   true,1,'struc',{},FBAsolution.x);
[involvedRxns,involvedMets,deadEnds] = draw_by_met (model,{'mal-L[e]'},...
   true,1,'struc',{},FBAsolution.x);

>> GapFindExample

allGaps =
'fru[e]'  
'fum[e]'  
'gln-L[e]'  
'mal-L[e]'

rootGaps =
Empty cell array: 0-by-1

downstreamGaps =
'fru[e]'  
'fum[e]'  
'gln-L[e]'  
'mal-L[e]''}
"gapFind" Example Metabolite Connectivity

Note that there are no inputs to any of the green metabolites since they cannot be secreted.

Secretion is a downstream process, thus a downstream gap.
E. coli Core Model

Downstream Gaps


http://systemsbiology.ucsd.edu/Downloads/E_coli_Core


Stage 4: Network Evaluation

43-44. Test if network is mass-and charge balanced.
45. Identify metabolic dead-ends.
49. Add missing exchange reactions to model.
50. Set exchange constraints for a simulation condition.
51-58. Test for stoichiometrically balanced cycles.
59. Re-compute gap list.
60-65. Test if biomass precursors can be produced in standard medium.
66. Test if biomass precursors can be produced in other growth media.
67-75. Test if the model can produce known secretion products.
76-78. Check for blocked reactions.
81-82. Test for known incapability’s of the organism.
83. Compare predicted physiological properties with known properties.
84-87. Test if the model can grow fast enough.
88-94. Test if the model grows too fast.
Network Evaluation:

Perform Gap Analysis

46. **Identify candidate reactions to fill gaps.** Use primary literature and genome annotation tools to find candidate genes and reactions to fill the gap. Also, use KEGG maps, biochemical textbooks or other available biochemical maps to identify the metabolic 'environment' of the dead-end metabolite. If the genome annotation of the target organism is present in KEGG, one can highlight the dead-end metabolite on the map. This may give an indication of which enzyme(s) may be able to produce or synthesize the dead-end metabolite and thus provide a good starting point for literature and/or genome search.

47. **Add gap reactions to the reconstruction.** If experimental and/or annotation data support gap reactions or they are needed for modeling purposes, the reaction(s) should be added to the reconstruction.

**CRITICAL STEP** Adding new reactions to the network may cause new gaps. When adding reactions, make sure that all the metabolites are connected to the network.

48. **Add notes and references to dead-end metabolites.** Each dead-end metabolite should be documented. The note for the remaining dead-end metabolites should distinguish between knowledge and scope gap for future reference.

**CRITICAL STEP** The more detailed and carefully the gap-filling steps are completed, the easier and faster the debugging process will be.

Maltose-6-phosphate is highlighted on the KEGG map for “Starch and Sucrose Metabolism”. All annotated *E. coli* genes (MG1655) in KEGG are colored green. Enzymes that are currently not annotated or not found are shown with white boxes.

Maltose-6-phosphate is a dead-end metabolite in *E. coli*’s metabolic reconstruction. The enzyme 3.2.1.122 is currently not annotated.

There are only two enzymes in the KEGG database that seem to produce/consume Maltose-6-phosphate: 2.7.1.69 and 3.2.1.122. In contrast, D-Glucose-6-Phosphate is highly connected in the *E. coli* reconstruction.

Pathway Databases
http://en.wikipedia.org/wiki/Metabolic_pathway

- BioCyc: Metabolic network models for hundreds of organisms
  http://www.biocyc.org/

- KEGG: Kyoto Encyclopedia of Genes and Genomes
  http://www.genome.jp/kegg/

- Reactome, a database of reactions, pathways and biological processes
  http://www.reactome.org/ReactomeGWT/entrypoint.html

- MetaCyc: A database of non-redundant, experimentally elucidated metabolic pathways (1800+ pathways from more than 2200 different organisms).
  http://metacyc.org/

- Metabolism, Cellular Respiration and Photosynthesis - The Virtual Library of Biochemistry and Cell Biology
  http://www.biochemweb.org/metabolism.shtml

- PathCase Pathways Database System
  http://nashua.case.edu/PathwaysWeb/

- Interactive Flow Chart of the Major Metabolic Pathways
  http://www2.ufp.pt/~pedros/bq/integration.htm

- DAVID: Visualize genes on pathway maps
  http://david.abcc.ncifcrf.gov/

- Wikipathways: pathways for the people
  http://www.wikipathways.org/index.php/WikiPathways

- ConsensusPathDB
  http://cpdb.molgen.mpg.de/
Stage 4: Network Evaluation

43–44. Test if network is mass-and charge balanced.
45. Identify metabolic dead-ends.
49. Add missing exchange reactions to model.
50. Set exchange constraints for a simulation condition.
51–58. Test for stoichiometrically balanced cycles.
59. Re-compute gap list.
60–65. Test if biomass precursors can be produced in standard medium.
66. Test if biomass precursors can be produced in other growth media.
67–75. Test if the model can produce known secretion products.
76–78. Check for blocked reactions.
81–82. Test for known incapability’s of the organism.
83. Compare predicted physiological properties with known properties.
84–87. Test if the model can grow fast enough.
88–94. Test if the model grows too fast.
Add Missing Exchange Reactions and Set Exchange Constraints

49. Add missing exchange reactions to model. The gap-filling process may have resulted in the inclusion of further transport reactions. Thus, exchange reactions need to be added to the reconstruction.

50. Set exchange constraints for a simulation condition. Determine an environmental condition, in which most network evaluation tests should be carried out initially ('standard condition'). Use

\[
\text{model} = \text{changeRxnBounds(model,rxnNameList,value,boundType)}
\]

to set the constraints. Reactions whose bounds should be changed are listed in 'rxnNameList'. The new value for each reaction is contained in the array 'value'. Finally, the type of constraint has to be defined in the list 'boundType'. The possible types are: 'l' for lower bound, 'u' for upper bound and 'b' if both reaction bounds should be set to the specified value.

Stage 4: Network Evaluation

43-44. Test if network is mass- and charge balanced.
45. Identify metabolic dead-ends.
49. Add missing exchange reactions to model.
50. Set exchange constraints for a simulation condition.
51-58. Test for stoichiometrically balanced cycles.
59. Re-compute gap list.
60-65. Test if biomass precursors can be produced in standard medium.
66. Test if biomass precursors can be produced in other growth media.
67-75. Test if the model can produce known secretion products.
76-78. Check for blocked reactions.
81-82. Test for known incapability’s of the organism.
83. Compare predicted physiological properties with known properties.
84-87. Test if the model can grow fast enough.
88-94. Test if the model grows too fast.
Types of Extreme Pathways

- Type I extreme pathways have exchange fluxes across the system boundaries that correspond to non-currency metabolites.
- Type II extreme pathways have only currency metabolites that cross system boundaries.
- Type III extreme pathways do not contain any exchange fluxes, and thus correspond to internal loops.

Type III Extreme Pathway (Loops) Removal During Simulations

- Jan Schellenberger wrote a function that removes thermodynamically infeasible loops from models:

- An allowLoops option is included in the following Cobra functions.
  - optimizeCbModel
  - fluxVariability
  - sampleCbModel

- When loops are not allowed the function run significantly slower.
Stage 4: Network Evaluation

43-44. Test if network is mass-and charge balanced.
45. Identify metabolic dead-ends.
49. Add missing exchange reactions to model.
50. Set exchange constraints for a simulation condition.
51-58. Test for stoichiometrically balanced cycles.
59. Re-compute gap list.
60-65. Test if biomass precursors can be produced in standard medium.
66. Test if biomass precursors can be produced in other growth media.
67-75. Test if the model can produce known secretion products.
76-78. Check for blocked reactions.
81-82. Test for known incapability’s of the organism.
83. Compare predicted physiological properties with known properties.
84-87. Test if the model can grow fast enough.
88-94. Test if the model grows too fast.
Lesson: Genome-scale Metabolic Reconstructions & Analysis

Stage 4:

Network Evaluation

43-44. Test if network is mass-and charge balanced.
45. Identify metabolic dead-ends.
49. Add missing exchange reactions to model.
50. Set exchange constraints for a simulation condition.
51-58. Test for stoichiometrically balanced cycles.
59. Re-compute gap list.
60-65. Test if biomass precursors can be produced in standard medium.
66. Test if biomass precursors can be produced in other growth media.
67-75. Test if the model can produce known secretion products.
76-78. Check for blocked reactions.
81-82. Test for known incapability's of the organism.
83. Compare predicted physiological properties with known properties.
84-87. Test if the model can grow fast enough.
88-94. Test if the model grows too fast.
Network Evaluation:

Test if Biomass Precursors can be Produced in Standard Medium

- Test the model's ability to produce each individual biomass component in standard medium condition (e.g., minimal medium M9 supplemented with D-glucose).

  ✓ Growth on minimal medium M9 was simulated by maximizing flux through a defined biomass objective function and allowing the uptake of the desired carbon source, NH₄, SO₄, O₂, and Pᵢ and the free exchange of H⁺, H₂O, and CO₂ (Joyce, A. R., J. L. Reed, et al. (2006). "Experimental and computational assessment of conditionally essential genes in *Escherichia coli.*" Journal of Bacteriology 188(23): 8259-8271.)

- The capability to produce biomass precursors also needs to be tested in other growth media. Therefore, the correctness of the network content is evaluated with respect to all the known growth conditions of the target organism. This includes all the known carbon, nitrogen, sulfur and phosphorus sources.

M9 Minimal Medium

- One liter of M9 medium (Sigma catalog no. 6030) contains:
  - $\text{Na}_2\text{HPO}_4 \cdot 7\text{H}_2\text{O}$ (6.8g), $\text{KH}_2\text{PO}_4$ (3g), $\text{NaCl}$ (0.5g), $\text{NH}_4\text{Cl}$ (1g), $\text{MgSO}_4$ (2 mM), $\text{CaCl}_2$ (0.1 mM)
- Growth on minimal medium was simulated by maximizing flux through a defined biomass objective function and allowing the uptake of
  - $\text{NH}_4$, $\text{SO}_4$, $\text{O}_2$, and $\text{P}_i$ and the free exchange of $\text{H}^+$, $\text{H}_2\text{O}$, and $\text{CO}_2$
- All exchange reaction lower constraints, except the following, should be greater than zero
  - $-1000 \leq \text{NH}_4$, $\text{SO}_4$, $\text{O}_2$, and $\text{P}_i \leq 0$
  - $-1000 \leq \text{H}^+$, $\text{H}_2\text{O}$, and $\text{CO}_2 \leq 1000$
  - $-1000 \leq \text{Carbon source} \leq 0$
  - Use the following commands to change the constraints
    - `model = changeRxnBounds(model,'EX_xxx(e)',-1000,'l')`
    - `model = changeRxnBounds(model,'EX_xxx(e)', 1000,'u')`
  - Verify that no other metabolites are allowed to be uptaken
    - No other metabolites should have a negative lower constraint
    - Check using the "printConstraints(model, -1001, 1)" command
Recipe for M-9 Minimal Media

- 5X M9 basis
  - Na$_2$HPO$_4$.12 H$_2$O 85.7 g
  - KH$_2$PO$_4$ 15.0 g
  - NaCl 2.5 g
  - Dissolve above components in 1000 ml of milli-Q and autoclave

- 5 g (NH$_4$)$_2$SO$_4$ in 15 ml of H$_2$O

- Trace elements
  - 1 g EDTA
  - 29 mg ZnSO$_4$.7H$_2$O
  - 198 mg MnCl$_2$.4H$_2$O
  - 254 mg CoCl$_2$.6H$_2$O
  - 13.4 mg CuCl$_2$
  - 147 mg CaCl$_2$
  - Dissolve in 100 ml of milli-Q and autoclave

- 20% (w/v) glucose: 25 g in 100 ml of milliQ and filter with 0.22 micron filter
- 0.1 M CaCl$_2$.2H$_2$O: 1.47 g in 100 ml milliQ and filter with 0.22 micron filter
- 1M MgSO$_4$.7H$_2$O: 24.65 g in 100 ml milliQ and filter with 0.22 micron filter
- 10 mM FeSO$_4$.7H$_2$O: 140 mg in 50 ml of milliQ (prepare fresh)
- 1% thiamine: 500mg in 10 ml of milliQ (prepare fresh)
- Proportions for 1 liter M-9 media
  - 200 ml of M-9 basis; 3 ml of (NH$_4$)$_2$SO$_4$; 1 ml of CaCl$_2$.2H$_2$O; 1 ml trace elements; 20 ml glucose; 1ml MgSO$_4$.7H$_2$O; 1 ml FeSO$_4$.7H$_2$O
  - 2ml thiamine; 1ml antibiotic (standard conc.)

**Minimal Nutrients for *E.coli* iaf1260**

\[
\begin{align*}
\text{EX}_{\text{glc}}(e) &= -10, \text{EX}_{\text{o2}}(e) = -1000 \\
\text{EX}_{\text{ca2}}(e) &= -0.00440206 \\
\text{EX}_{\text{cl}}(e) &= -0.00440206 \\
\text{EX}_{\text{co2}}(e) &= 21.9456 \\
\text{EX}_{\text{cobalt2}}(e) &= -0.0029347 \\
\text{EX}_{\text{cu2}}(e) &= -0.0029347 \\
\text{EX}_{\text{fe2}}(e) &= -0.00701801 \\
\text{EX}_{\text{fe3}}(e) &= -0.00660355 \\
\text{EX}_{\text{glc}}(e) &= -10 \\
\text{EX}_{\text{h2o}}(e) &= 46.4241 \\
\text{EX}_{\text{h}}(e) &= 8.53495 \\
\text{EX}_{\text{k}}(e) &= -0.165042 \\
\text{EX}_{\text{mg2}}(e) &= -0.00733676 \\
\text{EX}_{\text{mn2}}(e) &= -0.0029347 \\
\text{EX}_{\text{mobd}}(e) &= -0.0029347 \\
\text{EX}_{\text{nh4}}(e) &= -10.0215 \\
\text{EX}_{\text{o2}}(e) &= -19.9695 \\
\text{EX}_{\text{pi}}(e) &= -0.893343 \\
\text{EX}_{\text{so4}}(e) &= -0.232555 \\
\text{EX}_{\text{zn2}}(e) &= -0.0029347 \\
\text{Ec}_{\text{biomass}} &= 0.929292 \\
\end{align*}
\]

The metabolite molybdate (mobd) is not used in any reactions other than the biomass objective function and transport reactions which allow it to diffuse in and out of the cell.
Network Evaluation:

Test if Biomass Precursors Can Be Produced in Standard Medium (II)

60. Obtain the list of biomass components:

61. Add demand function for each biomass precursor ('metaboliteNameList'):

62. For each biomass component, perform the following test: Change objective function to the demand function ('rxnName'):

63. Maximize ('max') for new objective function (Demand function)

✓ Case 1, the model can produce biomass component (FBAsolution.obj > 0), proceed with the next biomass component.

✓ Case 2, the model cannot produce biomass component (FBAsolution.obj = 0). Follow steps 64 and 65

All this can be accomplished using the "biomassPrecursorCheck" function.

model=readCbModel('ecoli_textbook');
[missingMets,presentMets] = biomassPrecursorCheck(model)

DM_3pg[c] 3pg[c] ->
DM_accoa[c] accoa[c] ->
DM_atp[c] atp[c] ->
DM_e4p[c] e4p[c] ->
DM_f6p[c] f6p[c] ->
DM_g3p[c] g3p[c] ->
DM_g6p[c] g6p[c] ->
DM_gln-L[c] gln-L[c] ->
DM_glu-L[c] glu-L[c] ->
DM_h2o[c] h2o[c] ->
DM_nad[c] nad[c] ->
DM_nadph[c] nadph[c] ->
DM_ooa[c] ooa[c] ->
DM_pep[c] pep[c] ->
DM_pyr[c] pyr[c] ->
DM_r5p[c] r5p[c] ->

missingMets = 'atp[c]' 'nadph[c]'
presentMets = '3pg[c]' 'accoa[c]' 'e4p[c]' 'f6p[c]' 'g3p[c]' 'g6p[c]' 'gln-L[c]' 'glu-L[c]' 'h2o[c]' 'nad[c]' 'ooa[c]' 'pep[c]' 'pyr[c]' 'r5p[c]'

This function may identify metabolites that are typically recycled within the network such as ATP, NAD, NADPH, ACCOA.

"biomassPrecursorCheck" Example

Demand reactions are created for each element in the biomass function to check to see if the precursors can be synthesized

Different name than in the Cobra Documentation
Network Evaluation:

Test if Biomass Precursors Can Be Produced in Standard Medium (II)

64. Identify reactions that are mainly responsible for synthesizing the biomass component.

65. For each of these reactions, follow the paths outlined in the debugging flowchart.


- ‘rxn’ stands for reaction;
- ‘conf’ stands for confidence score;
- ‘met’ stands for metabolite.
Stage 4: Network Evaluation

43-44. Test if network is mass- and charge balanced.
45. Identify metabolic dead- ends.
49. Add missing exchange reactions to model.
50. Set exchange constraints for a simulation condition.
51-58. Test for stoichiometrically balanced cycles.
59. Re-compute gap list.
60-65. Test if biomass precursors can be produced in standard medium.
66. Test if biomass precursors can be produced in other growth media.
67-75. Test if the model can produce known secretion products.
76-78. Check for blocked reactions.
81-82. Test for known incapability’s of the organism.
83. Compare predicted physiological properties with known properties.
84-87. Test if the model can grow fast enough.
88-94. Test if the model grows too fast.
Stage 4: Network Evaluation

43-44. Test if network is mass-and charge balanced.
45. Identify metabolic dead-ends.
49. Add missing exchange reactions to model.
50. Set exchange constraints for a simulation condition.
51-58. Test for stoichiometrically balanced cycles.
59. Re-compute gap list.
60-65. Test if biomass precursors can be produced in standard medium.
66. Test if biomass precursors can be produced in other growth media.
67-75. Test if the model can produce known secretion products.
76-78. Check for blocked reactions.
81-82. Test for known incapability’s of the organism.
83. Compare predicted physiological properties with known properties.
84-87. Test if the model can grow fast enough.
88-94. Test if the model grows too fast.
Network Evaluation:

Test If The Model Can Produce Known Secretion Products

- Collect a list of known secretion bioproducts and medium conditions.
- The secretion of by-products from the model can be determined using either the “productionEnvelope” (one secreted bioproduct) or “multiProductionEnvelope” (all secreted bioproducts) functions.
- Identify missing secreted bioproducts.
Production Envelope of Secreted Metabolites

% Secretion_multiProductionEnvelope.m

clear;

model = readCbModel('ecoli_textbook');

model = changeRxnBounds(model, 'EX_glc(e)', -5, 'l');
model = changeRxnBounds(model, 'EX_o2(e)', -20, 'l');

deletions = {};

biomassRxn = {'Biomass_Ecoli_core_N(w/GAM)_Nmet2'};

[biomassValues, targetValues] = multiProductionEnvelope(model, deletions, biomassRxn)
Stage 4:
Network Evaluation

43-44. Test if network is mass-and charge balanced.
45. Identify metabolic dead-ends.
49. Add missing exchange reactions to model.
50. Set exchange constraints for a simulation condition.
51-58. Test for stoichiometrically balanced cycles.
59. Re-compute gap list.
60-65. Test if biomass precursors can be produced in standard medium.
66. Test if biomass precursors can be produced in other growth media.
67-75. Test if the model can produce known secretion products.
76-78. Check for blocked reactions.
81-82. Test for known incapability’s of the organism.
83. Compare predicted physiological properties with known properties.
84-87. Test if the model can grow fast enough.
88-94. Test if the model grows too fast.
Network Evaluation:

**Check For Blocked Reactions**

- Reactions that cannot carry any flux in any simulation conditions are called blocked reactions. These reactions are directly or indirectly associated with dead-end metabolites, which cannot be balanced and give rise to blocked compounds.

- The function "findBlockedReactions" described in the protocols paper does not work.

- Use the Matlab script called "findBlockedReactionTest.m"

- The exchange reactions need to be able to uptake metabolites to get an accurate output. Normally several of the exchanged reactions in the E.coli textbook model are not allowed to uptake metabolites. They include: EX_fru(e), 'EX_fum(e), 'EX_gln_L(e), 'EX_mal_L(e)

- The pathways of the blocked reactions can be traced to find the problem. A single reaction can block many other reactions

---

clear;
model=readCbModel('ecoli_textbook'); % Input the E.coli core model
model = changeRxnBounds(model,'GLUDy',0,'b'); % Test for blocked reaction

% Open all exchange reactions
[selExc,selUpt] = findExcRxns(model); % Find exchange reactions
model = changeRxnBounds(model,model.rxns(selExc),-1000,'l'); % Change lower bounds
model = changeRxnBounds(model,model.rxns(selExc),1000,'u'); % Change upper bounds

tol = 1e-10;
%blockedReactions=[]; % Creates type problem in Matlab
[minFlux,maxFlux] = fluxVariability(model,0);
cnt = 1;
for i=1:length(minFlux)
    if (maxFlux(i) < tol && maxFlux(i) > -tol && minFlux(i) < tol && minFlux(i) > -tol)
        blockedReactions(cnt) = model.rxns(i);
        cnt = cnt + 1;
    end
end

findBlockedReactionTest.m
Stage 4: 
Network Evaluation

43-44. Test if network is mass-and charge balanced.
45. Identify metabolic dead-ends.
49. Add missing exchange reactions to model.
50. Set exchange constraints for a simulation condition.
51-58. Test for stoichiometrically balanced cycles.
59. Re-compute gap list.
60-65. Test if biomass precursors can be produced in standard medium.
66. Test if biomass precursors can be produced in other growth media.
67-75. Test if the model can produce known secretion products.
76-78. Check for blocked reactions.

81-82. Test for known incapability’s of the organism.
83. Compare predicted physiological properties with known properties.
84-87. Test if the model can grow fast enough.
88-94. Test if the model grows too fast.
Network Evaluation: **Compute Single Gene Deletion Phenotypes**

- Analysis of false-positive and false-negative predictions will help to further refine the network content if the information is available.

- Phenotyping data (e.g., biolog data), or gene essentiality data, can be used to improve the network content.

- The “singleGeneDeletion” can be used to compare experimental data with predicted behavior of single gene knockouts.

- This function allows the use of different methods ('method') for optimization, e.g., FBA, minimization of metabolic adjustment (MOMA) or linear MOMA. The list of genes that shall be deleted is given by 'geneList'.

- Calculates the growth rate of the wild-type strain ('grRateWT') of each deletion strain ('grRateKO'), as well as the relative growth rate ratios ('grRatio').

- Test to see if known incapacities and the physiological properties of the organism can be reproduced by the model.


[The Biolog OmniLog® incubates and monitors 50 microplates, or 1,920 phenotypic assays simultaneously to measure physiological responses in diverse microbial cells.

http://www.biolog.com/pdf/pm_lit/00A%20037rA%20PM%20Microbiology%202011.pdf]
Single Reaction Deletion

```
% SingleReactionDeletionTest.m

clear;

% Input model
model = readCbModel('ecoli_textbook');

[grRatio, grRateKO, grRateWT, hasEffect, delRxns, fluxSolution] = singleRxnDeletion(model, 'FBA');

% [grRatio, grRateKO, grRateWT, hasEffect, delRxns, fluxSolution] = singleRxnDeletion(model, 'MOMA');
% [grRatio, grRateKO, grRateWT, hasEffect, delRxns, fluxSolution] = singleRxnDeletion(model, 'lMOMA');

% Constraint-based Metabolic Reconstructions & Analysis

```

### Table: Single Reaction Deletion Results

<table>
<thead>
<tr>
<th>Reactions</th>
<th>grRateWT</th>
<th>grRateKO</th>
<th>grRatio</th>
</tr>
</thead>
<tbody>
<tr>
<td>'ACALD'</td>
<td>0.873921507</td>
<td>0.873921507</td>
<td>1</td>
</tr>
<tr>
<td>'ACALDt'</td>
<td>0.873921507</td>
<td>0.873921507</td>
<td>1</td>
</tr>
<tr>
<td>'ACKr'</td>
<td>0.873921507</td>
<td>0.873921507</td>
<td>1</td>
</tr>
<tr>
<td>'ACONTa'</td>
<td>0.873921507</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>'ACONTb'</td>
<td>0.873921507</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>'ACt2r'</td>
<td>0.873921507</td>
<td>0.873921507</td>
<td>1</td>
</tr>
<tr>
<td>'ADK1'</td>
<td>0.873921507</td>
<td>0.873921507</td>
<td>1</td>
</tr>
<tr>
<td>'AKGDH'</td>
<td>0.873921507</td>
<td>0.858307408</td>
<td>0.982133294</td>
</tr>
<tr>
<td>'AKGt2r'</td>
<td>0.873921507</td>
<td>0.873921507</td>
<td>1</td>
</tr>
<tr>
<td>'ALCD2x'</td>
<td>0.873921507</td>
<td>0.873921507</td>
<td>1</td>
</tr>
<tr>
<td>'ATPM'</td>
<td>0.873921507</td>
<td>0.916647464</td>
<td>1.048889925</td>
</tr>
<tr>
<td>'ATPS4r'</td>
<td>0.873921507</td>
<td>0.374229875</td>
<td>0.428219093</td>
</tr>
<tr>
<td>'Biomass_Ecoli_core_N(w/GAM)_Nmet2'</td>
<td>0.873921507</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>'CO2t'</td>
<td>0.873921507</td>
<td>0.461669614</td>
<td>0.528273547</td>
</tr>
<tr>
<td>'CS'</td>
<td>0.873921507</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>'CYTBD'</td>
<td>0.873921507</td>
<td>0.21166295</td>
<td>0.24219904</td>
</tr>
<tr>
<td>'D_LACt2'</td>
<td>0.873921507</td>
<td>0.873921507</td>
<td>1</td>
</tr>
<tr>
<td>'ENO'</td>
<td>0.873921507</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>'ETOHt2r'</td>
<td>0.873921507</td>
<td>0.873921507</td>
<td>1</td>
</tr>
</tbody>
</table>
```
Constraint-based Metabolic Reconstructions & Analysis

H. Scott Hinton, 2016

Lesson: Genome-scale Metabolic Reconstructions

Single Gene Deletion

% SingleGeneDeletionTest.m

clear;

% Input model

model=readCbModel('ecoli_textbook');

[grRatio,grRateKO,grRateWT,delRxns,hasEffect] = singleGeneDeletion(model,'FBA');

% [grRatio,grRateKO,grRateWT,delRxns,hasEffect] = singleGeneDeletion(model,'MOMA');

% [grRatio,grRateKO,grRateWT,delRxns,hasEffect] = singleGeneDeletion(model,'lMOMA');
Double Gene Deletion

% DoubleGeneDeletionTest.m

clear;

% Input model

model=readCbModel('ecoli_textbook');

[grRatio, grRateKO, grRateWT] = doubleGeneDeletion(model,'FBA');

% [grRatio, grRateKO, grRateWT, delRxns, hasEffect] = doubleGeneDeletion(model,'MOMA');

% [grRatio, grRateKO, grRateWT, delRxns, hasEffect] = doubleGeneDeletion(model,'lMOMA');

imagesc(grRatio)
xlabel('Gene Knockout #1');
ylabel('Gene Knockout #2');
Stage 4: Network Evaluation

43-44. Test if network is mass-and charge balanced.
45. Identify metabolic dead-ends.
49. Add missing exchange reactions to model.
50. Set exchange constraints for a simulation condition.
51-58. Test for stoichiometrically balanced cycles.
59. Re-compute gap list.
60-65. Test if biomass precursors can be produced in standard medium.
66. Test if biomass precursors can be produced in other growth media.
67-75. Test if the model can produce known secretion products.
76-78. Check for blocked reactions.
81-82. Test for known incapability's of the organism.
83. Compare predicted physiological properties with known properties.
84-87. Test if the model can grow fast enough.
88-94. Test if the model grows too fast.
Network Evaluation

Test if the Model Can Grow Fast Enough

- Check boundary constraints
  - `printConstraints(model, MinInf, MaxInf)` – % example `printConstraints(model,-1001,1001)`

- Check reaction directionality
  - `printRxnFormula(model)`

- Determine the reduced cost associated with network reactions when optimizing for objective function.
  - `FBAsolution = optimizeCbModel(model, osenseStr, primalOnlyFlag)`
  - set `primalOnlyFlag` to 'false' to get the reduced cost returned with the optimal solution (`FBAsolution.w`). When maximizing the objective function 'osenseStr' will be 'max', whereas minimization is defined by 'min'.
  - Find the reactions with the lowest reduced cost values. Increase flux through those reactions, if possible, by removing upper bounds. This will lead to increased flux through the objective reaction.

Stage 4: Network Evaluation

43-44. Test if network is mass-and charge balanced.
45. Identify metabolic dead-ends.
49. Add missing exchange reactions to model.
50. Set exchange constraints for a simulation condition.
51-58. Test for stoichiometrically balanced cycles.
59. Re-compute gap list.
60-65. Test if biomass precursors can be produced in standard medium.
66. Test if biomass precursors can be produced in other growth media.
67-75. Test if the model can produce known secretion products.
76-78. Check for blocked reactions.
81-82. Test for known incapability’s of the organism.
83. Compare predicted physiological properties with known properties.
84-87. Test if the model can grow fast enough.
88-94. Test if the model grows too fast.
Network Evaluation

Test if the Model Grows too Fast

- Check boundary constraints
  - \texttt{printConstraints(model,MinInf,MaxInf)}

- Check reaction directionality
  - \texttt{printRxnFormula(model)}

- Use single-reaction deletion to identify single reactions that may enable the model to grow too fast.
  - \[ [\text{grRatio}, \text{grRateKO}, \text{grRateWT}] = \	ext{singleRxnDeletion}(model, \text{‘FBA’},) \]
  - The function will return the wild-type growth rate (‘grRateW’), the growth rate of the reaction-deleted network (‘grRateKO’) and the relative growth rate ratio (‘grRatio’). However, it is most likely that multiple reactions contribute to this observation, and thus, they are not identified by this method.

- The reduced cost analysis can be used to identify those reactions that can reduce the growth rate (positive cost value).
  - \texttt{FBAsolution = optimizeCbModel(model,osenseStr,primalOnlyFlag)}
  - set \texttt{primalOnlyFlag} to ‘false’ to get the reduced cost returned with the optimal solution (FBAsolution.w). Set ‘osenseStr’ to ‘max’.
  - Find the reactions with the lowest reduced cost values. Increase flux through those reactions, if possible, by removing upper bounds. This will lead to increased flux through the objective reaction.

GENOME-SCALE METABOLIC RECONSTRUCTIONS

- Overview
- Draft Reconstruction
- Refinement of Reconstruction
- Conversion of Reconstruction into Computable Format
- Network Evaluation
- Data Assembly and Dissemination

**Reconstruction Process: 96 Step Protocol**


1. **Draft Reconstruction**
   1. Obtain genome annotation.
   2. Identify candidate metabolic functions.
   3. Obtain candidate metabolic reactions.
   4. Assembly of draft reconstruction.
   5. Collect of experimental data.

2. **Refinement of reconstruction**
   6. Determine and verify substrate and cofactor usage.
   7. Obtain neutral formula for each metabolite.
   8. Determine the charged formula.
   9. Calculate reaction stoichiometry.
   10. Determine reaction c/rectangularity.
   11. Add information for gene and reaction localization.
   12. Add subsystems information.
   15. Determine and add confidence score.
   16. Add references and notes.
   17. Add information from other organisms.
   18. Repeat step 6 to 17 for all genes.
   19. Add spontaneous reactions to the reconstruction.
   20. Add extracellular and periplasmic transport reactions.
   22. Add intracellular transport reactions.
   23. Draw metabolic map (optional).
   24-32 Determine biomass composition.
   33. Add biomass reaction.
   34. Add ATP maintenance reaction (ATPM).
   35. Add demand reactions.
   36. Add sink reactions.
   37. Determine growth medium requirements.

3. **Conversion of reconstruction into computable format**
   38. Initialize the COBRA toolbox.
   39. Load reconstruction into Matlab.
   40. Verify S matrix.
   41. Set objective function.
   42. Set simulation constraints.

4. **Network evaluation**
   43-44 Test if network is mass- and charge balanced.
   45. Identify metabolic dead-ends.
   46-48 Gap analysis.
   49. Add missing exchange reactions to model.
   50. Set exchange constraints for a simulation condition.
   51-58 Test for stoichiometrically balanced cycles.
   59. Re-compute gap list.
   60-65 Test if biomass precursors can be produced in standard medium.
   66. Test if biomass precursors can be produced in other growth media.
   67-75 Test if model can produce known secretions products.
   76-78 Check for blocked reactions.
   79-80 Compute single gene deletion phenotypes.
   81-82 Test for known inabilities of the organism.
   83. Compare predicted physiological properties with known properties.
   84-87 Test if the model can grow fast enough.
   88-94 Test if the model grows too fast.

5. **Data assembly and Dissemination**
   95. Print Matlab model content.
   96. Add gap information to the reconstruction output.
Stage 5: Data Assembly and Dissemination

95. Print Matlab model content.
   • Make the final reconstruction available to the research community in at least two formats: Excel spreadsheet and SBML
   • Excel spreadsheet Cobra function
     
     writeCBmodel(model,'xls','FileName')
   
   • SBML Cobra function
     
     writeCBmodel(model,'xls','FileName')

96. Add gap information to the reconstruction output.
   • Completed in Steps 45-48

GENOME-SCALE METABOLIC RECONSTRUCTIONS

- Overview
- Draft Reconstruction
- Refinement of Reconstruction
- Conversion of Reconstruction into Computable Format
- Network Evaluation
- Data Assembly and Dissemination

EXTRAS
E. coli Core Model


http://systemsbiology.ucsd.edu/Downloads/E_coli_Core
Constraint-based Metabolic Reconstructions & Analysis

Full E.coli model "ecoli_iaf1260.xml"
### MassChargeBalance_iaf1260_MB.m Example:

**“UnbalancedRxns” Matrix**

<table>
<thead>
<tr>
<th>Reaction Indices</th>
<th>H</th>
<th>C</th>
<th>O</th>
<th>P</th>
<th>S</th>
<th>N</th>
<th>Mg</th>
<th>X</th>
<th>Fe</th>
<th>Zn</th>
<th>Co</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td>371</td>
<td></td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>372</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>144</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>167</td>
<td></td>
<td>-1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>187</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>195</td>
<td></td>
<td>-1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>198</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>199</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>227</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2324</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Secreted Metabolites
Addition of Constraints

• Types of constraints
  ✓ Mass balance
  ✓ Steady-state
  ✓ Thermodynamics (e.g., reaction directionality)
  ✓ Environmental constraints (e.g., presence/absence of nutrient)
  ✓*Regulatory (e.g., on/off gene expression)
Addition of Constraints (II)

Constraint-based Metabolic Reconstructions & Analysis
H. Scott Hinton, 2016