

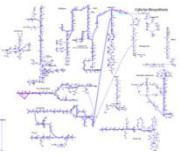
Genome-scale Metabolic Reconstructions



LEARNING OBJECTIVES

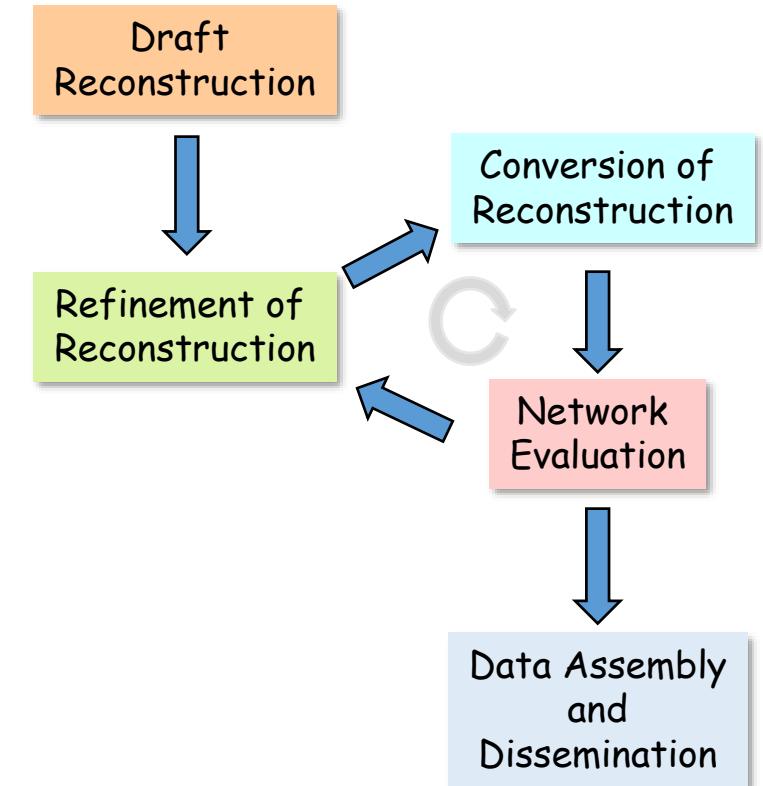
Each student should be able to:

- Explain the process of creating a genome-scale metabolic reconstruction

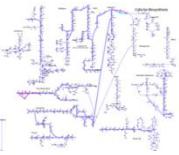


GENOME-SCALE METABOLIC RECONSTRUCTIONS

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

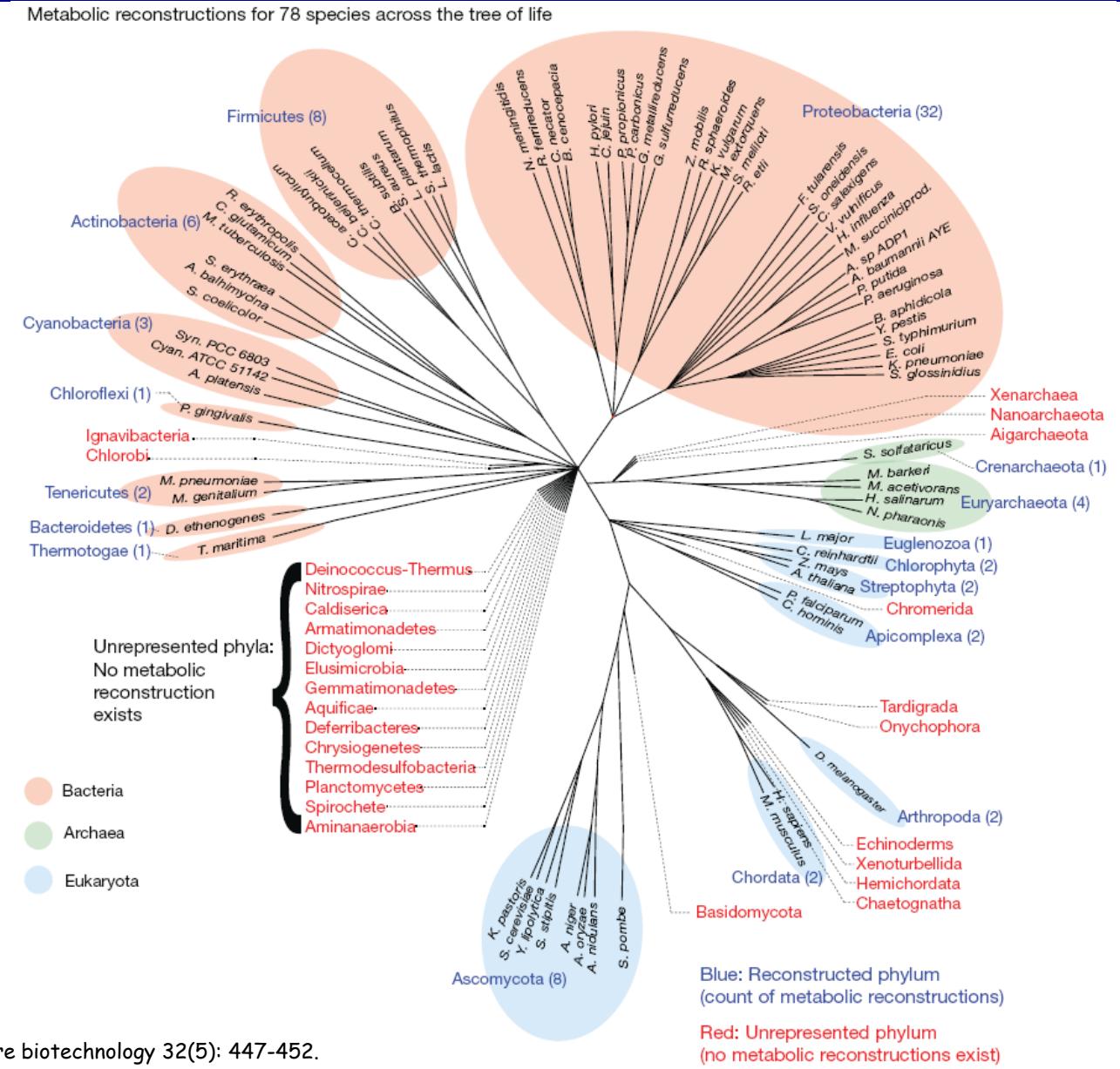


Thiele, I. and B. O. Palsson (2010). "A protocol for generating a high-quality genome-scale metabolic reconstruction." Nature protocols 5(1): 93-121.

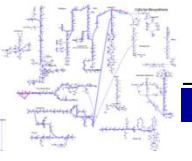


Phylogenetic Coverage of Genome-scale Network Reconstructions

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.

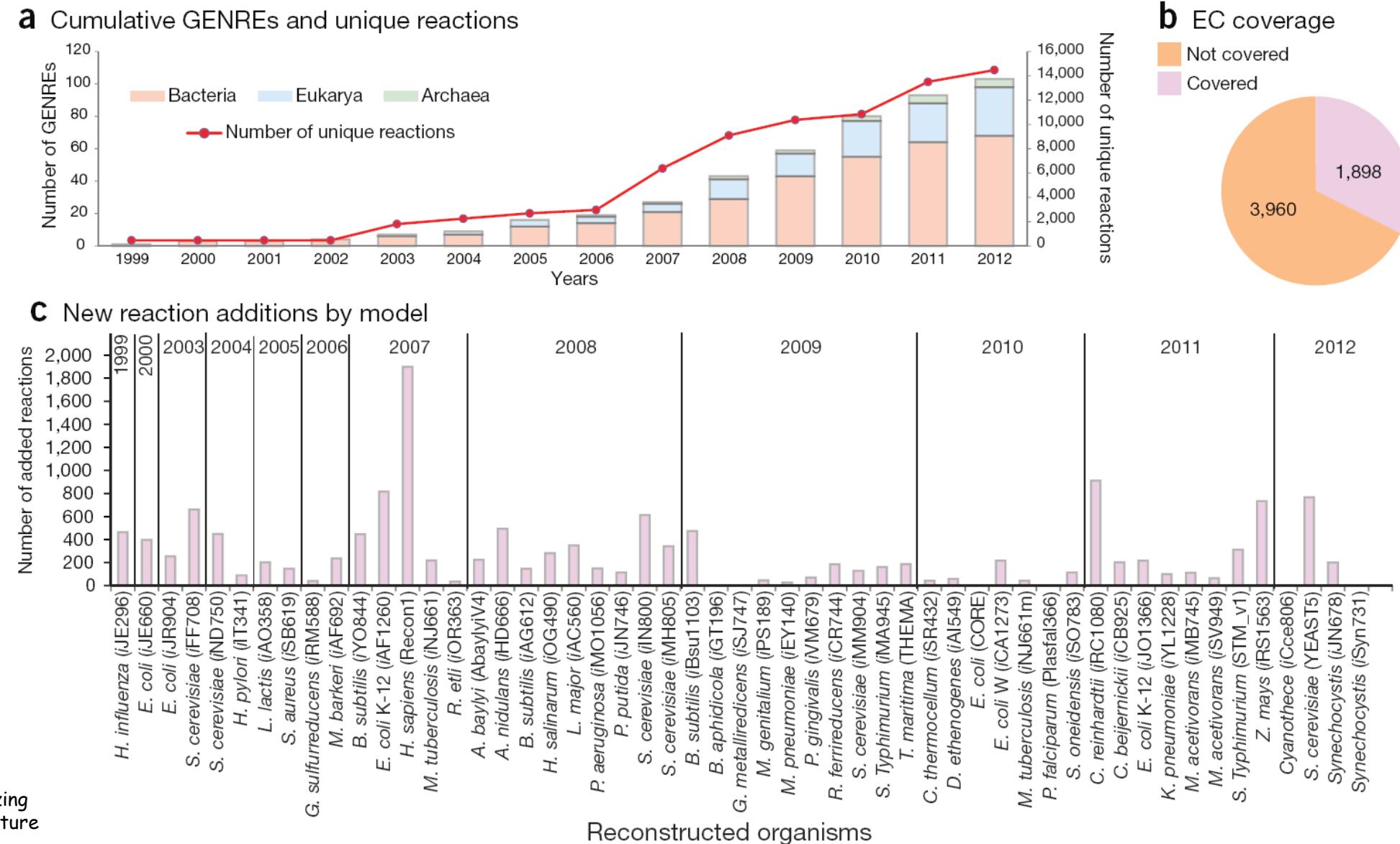


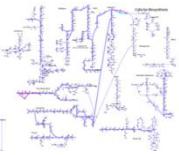
Monk, J., J. Nogales, et al. (2014). "Optimizing genome-scale network reconstructions." *Nature biotechnology* 32(5): 447-452.



Expansion of Metabolic Networks and Global Reactome Coverage Over Time

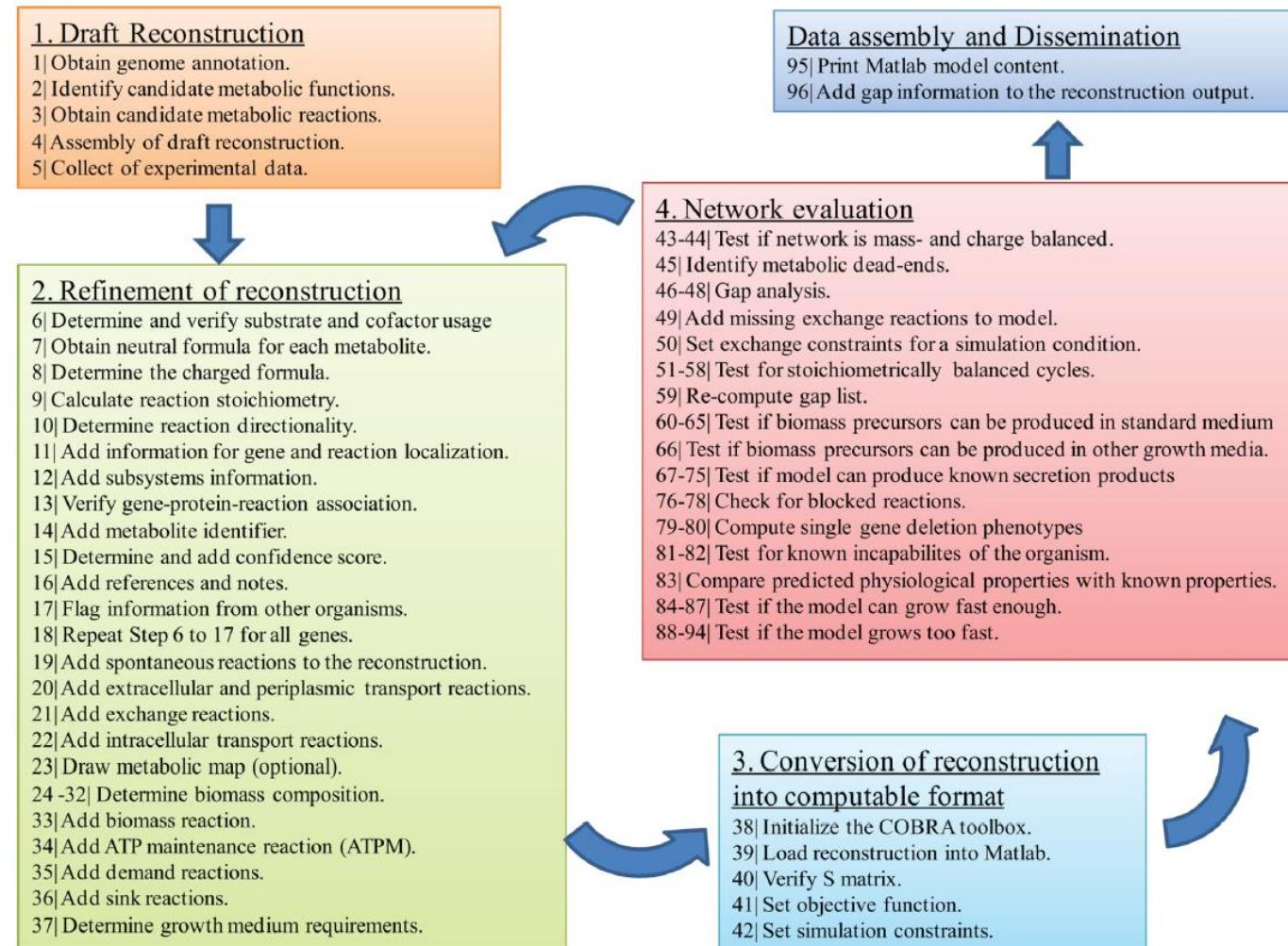
Monk, J., J. Nogales, et al. (2014). "Optimizing genome-scale network reconstructions." *Nature biotechnology* 32(5): 447-452.

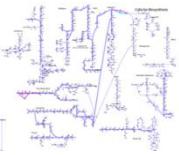




Reconstruction Process: 96 Step Protocol

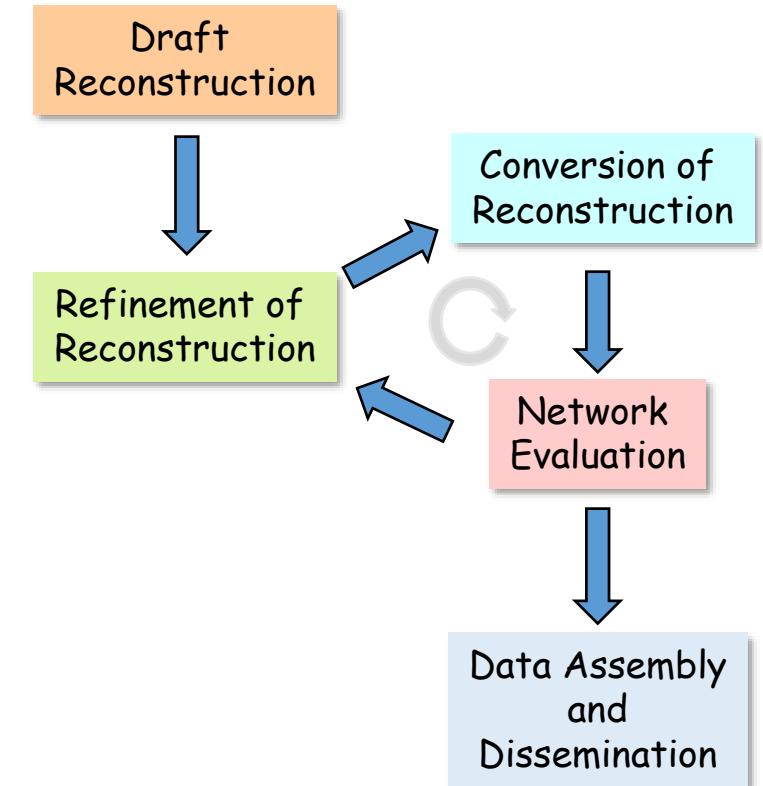
Thiele, I. and B. O. Palsson (2010). "A protocol for generating a high-quality genome-scale metabolic reconstruction." *Nature protocols* 5(1): 93-121.





GENOME-SCALE METABOLIC RECONSTRUCTIONS

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



Thiele, I. and B. O. Palsson (2010). "A protocol for generating a high-quality genome-scale metabolic reconstruction." Nature protocols 5(1): 93-121.



Draft Reconstruction

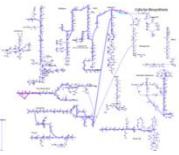
1. Obtain genome annotation
2. Identify candidate metabolic functions
3. Obtain candidate metabolic reactions
4. Assembly of draft reconstruction
5. Collect experimental data



Genome Databases

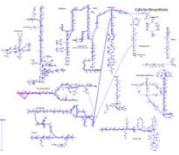
Name	Link	Comment
Comprehensive Microbial Resource (CMR)	http://cmr.jcvi.org/cgi-bin/CMR/CmrHomePage.cgi	
Genomes OnLine Database (GOLD)	http://www.genomesonline.org/	
TIGR	http://www.tigr.org/db.shtml	
NCBI Entrez Gene	http://www.ncbi.nlm.nih.gov/sites/entrez	
SEED database32	theseed.uchicago.edu/FIG/index.cgi	Comparative genomics tool

Thiele, I. and B. O. Palsson (2010). "A protocol for generating a high-quality genome-scale metabolic reconstruction." *Nature protocols* 5(1): 93-121.



Draft Reconstruction: Obtain Genome Annotation

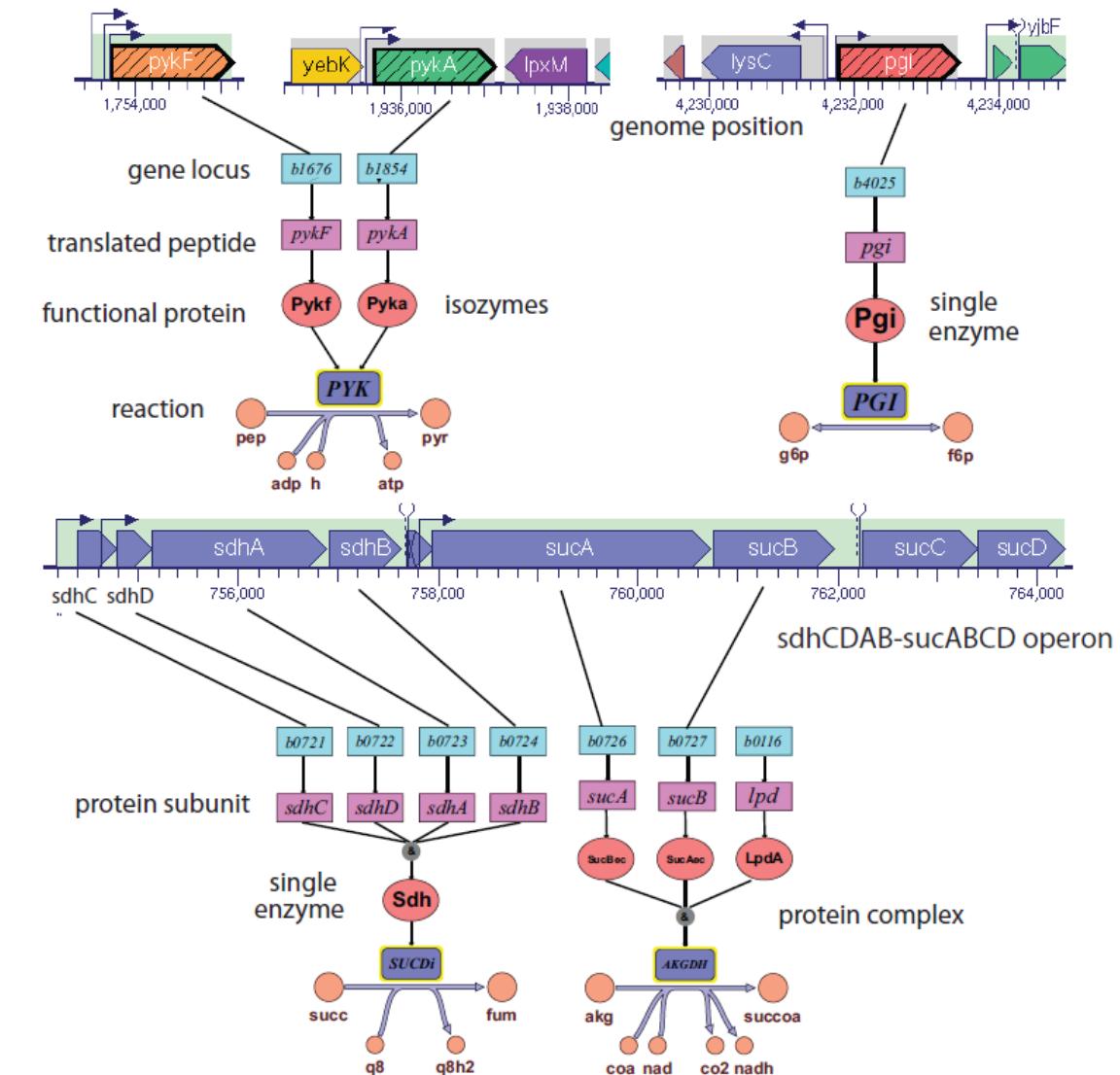
1. Automatic Annotation of Genome Sequences
 - a. Pathway Tools (Pathologic) - <http://bioinformatics.ai.sri.com/ptools/>
 - b. MetaSHARK - <http://bioinformatics.leeds.ac.uk/shark/>
2. Existing Databases:
 - a. TIGR-CMR Comprehensive Microbial Resource
<http://cmr.jcvi.org/tigr-scripts/CMR/CmrHomePage.cgi>
 - b. National Center for Biotechnology Information (NCBI)
<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?CMD=search&DB=gene>
 - 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:

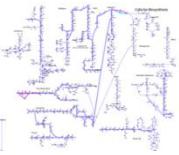
Create A biochemically, genetically and genetically (BiGG) structured knowledge base





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



NCBI Resources How To

Gene Gene Escherichia coli str. K-12 substr. MG1655 AND metab* NOT regulator

Save search Limits Advanced

Display Settings: Summary, 20 per page, Sorted by Relevance

Results: 1 to 20 of 1230

Organism Name

No transcriptional regulators

Only genes with metab* in description

Gene Symbol

Gene Function

Gene Information

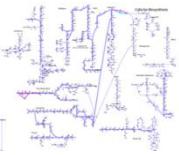
Gene Locus

1. acyl-CoA thioesterase, involved in phospholipid metabolism [Escherichia coli str. K-12 substr. MG1655]
Other Aliases: b0736, ECK0725, JW0726
Genomic context: Chromosome
Annotation: NC_000913.2 (773975..774379)
ID: 948907

2. glucose-1-phosphate uridylyltransferase [Escherichia coli str. K-12 substr. MG1655]
Other Aliases: b1236, ECK1231, JW1224, verA, ychD
Genomic context: Chromosome
Annotation: NC_000913.2 (1290680..1291588)
ID: 945730

3. cysteine sulfinate desulfinate [Escherichia coli str. K-12 substr. MG1655]
Other Aliases: b2810, ECK2806, JW2781, ygdJ
Genomic context: Chromosome
Annotation: NC_000913.2 (2941359..2942564)
ID: 947275

http://www.ncbi.nlm.nih.gov/gene?term=Escherichia%20coli%20str.%20K-12%20substr.%20MG1655%20AND%20metab*%20NOT%20regulator



NCBI Resources How To

Gene Gene Search

Limits Advanced

Display Settings: Full Report Send to:

galU glucose-1-phosphate uridylyltransferase [*Escherichia coli* str. K-12 substr. MG1655]

Gene ID: 945730, updated on 9-Jun-2012

Summary

Gene symbol galU
Gene description glucose-1-phosphate uridylyltransferase
Primary source EcoGene:EG11319 ← Additional Sources
Locus tag b1236
See related ECOCYC:EG11319 ←
Gene type protein coding
RefSeq status PROVISIONAL
Organism *Escherichia coli* str. K-12 substr. MG1655 (strain: K-12, substrain: MG1655, old-name: *Escherichia coli* K12)
Lineage Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia
Also known as ECK1231; JW1224; verA; ychD

Genomic context

Sequence: NC_000913.2 (1290680..1291588)

NC_000913.2

[1288468 → ↗ galU ↗ 1293367]

rssA → rssB → galU → hns → talk →

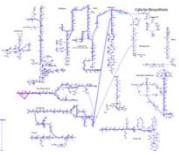
<http://www.ncbi.nlm.nih.gov/gene/945730>



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

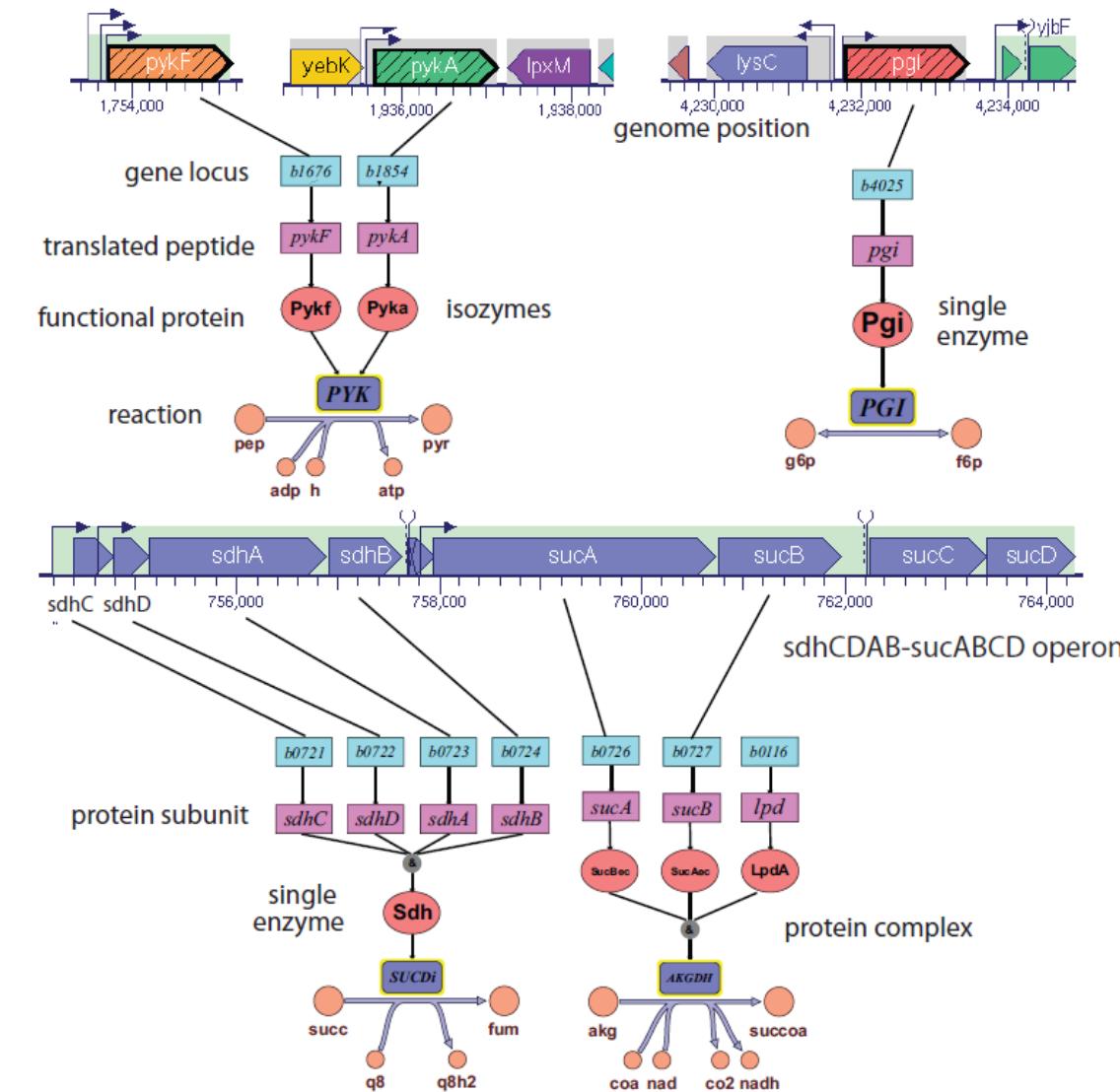




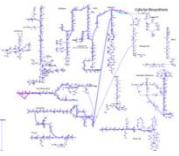
Desired Reaction Information

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



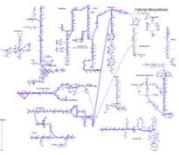
Reconstruction and Use of Microbial Metabolic Networks: the Core Escherichia coli Metabolic Model as an Educational Guide by Orth, Fleming, and Palsson (2010)



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.

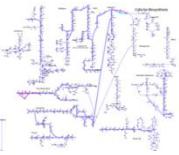
Thiele, I. and B. O. Palsson (2010). "A protocol for generating a high-quality genome-scale metabolic reconstruction." *Nature protocols* 5(1): 93-121, Supplementary Methods.



Reaction Names Suffix Standards

Reaction Type	Suffix	Example
ABC transporter	-abc	ALAabc
Transport reactions	-t	GLCt1
Reversible reactions	-r	GLCt1r
Irreversible reactions	-i	PTRCt3i
Dehydrogenase reactions	-DH	PDH
Synthetase reactions	-S	ATPS
Kinase reactions	-K	ACKr
Chloroplast reactions	-h	HEX1h
Endoplasmic Reticular reactions	-er	CERASE124er
Extracellular reactions	-e	AKGDHe
Golgi reactions	-g	S6T12g
Lysosomal reactions	-l	10FTHfl
Mitochondrial reactions	-m	AKGdm
Nucleus reactions	-n	UMPK3n
Peroxisomal reactions	-x	SCP3x
Periplasmic reactions	-pp	PPTHpp
Vacuole	-v	GLCGSDv

Thiele, I. and B. O. Palsson (2010). "A protocol for generating a high-quality genome-scale metabolic reconstruction." Nature protocols 5(1): 93-121, Supplementary Methods.



EcoCyc
A member of the BioCyc database collection

Pathway Tools Adv Tutorial
Aug 1st-3rd 2012
[Click here for details](#)

LOGIN | Why Login? | Create New Account

Quick Search Gene Search

Searching *Escherichia coli* K-12 substr. MG1655 [change organism database](#)

Home | Search | Tools | Help | Gene

If you [log in](#) you can add this to a group.

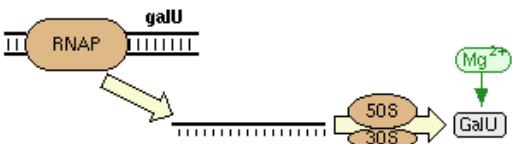
***Escherichia coli* K-12 substr. MG1655 Enzyme: UTP--glucose-1-phosphate uridylyltransferase**

Protein Sequence Nucleotide Sequence Nucleotide Sequence, Advanced

Gene: [galU](#) Accession Numbers: EG11319 (EcoCyc), b1236, ECK1231

Synonyms: ychD

Regulation Summary Diagram: ?



Summary:

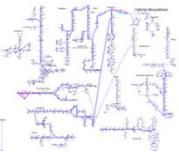
UTP-glucose-1-phosphate uridylyltransferase (GalU) carries out a key step in the generation of UDP-D-glucuronate as part of the larger system of [colanic acid building blocks biosynthesis](#).

GalU catalyzes the addition of UTP to α-D-glucose 1-phosphate to yield UDP-D-glucose [[Weissborn94](#)].

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 [[Fukasawa62](#), [Sundararaj62](#)].

<http://biocyc.org/ecoli/new-image?object=EG11319>



EcoCyc
A member of the BioCyc database collection

Pathway Tools Adv Tutorial
Aug 1st-3rd 2012
[Click here for details](#)

LOGIN | Why Log
Quick Sea
Searching *Escherichia coli* K-12 substr. MG1655 [change](#)

Home | Search | Tools | Help | Gene

If you [log in](#) you can add this to a group.

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

Gene-Reaction Schematic: [?](#)

2.7.7.9 — galU

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

Synonyms: UDPG synthetase, UDP-glucose pyrophosphorylase

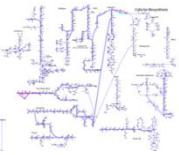
$\alpha\text{-D-glucose 1-phosphate} + \text{UTP} + \text{H}^+ \rightleftharpoons \text{UDP-D-glucose} + \text{diphosphate}$

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:
In vitro, the reaction is reversible with an equilibrium constant of 5.0, i.e. degradation of UDP-glucose is favored [[Kamogawa65](#)]. However, the concentration of pyrophosphate *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

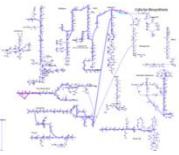
Gene	b2388	Locus	Genomics ORF annotation	
Peptide	glk	Gene	Transcriptomics mRNA levels	
Protein	Glk	Protein	Proteomics protein levels	
		Reaction	"Fluxomics" flux measurements	
Substrates	glc	atp	g6p	adp
Neutral	C ₆ H ₁₂ O ₆ ⁰	C ₁₀ H ₁₆ N ₅ O ₁₃ P ₃ ⁰	C ₆ H ₁₃ O ₉ P ⁰	C ₁₀ H ₁₅ N ₅ O ₁₀ P ₂ ⁰
Charged	C ₆ H ₁₂ O ₆ ⁰	C ₁₀ H ₁₂ N ₅ O ₁₃ P ₃ ⁴⁻	C ₆ H ₁₁ O ₉ P ²⁻	C ₁₀ H ₁₂ N ₅ O ₁₀ P ₂ ³⁻
Stoichiometry	C ₁₆ H ₂₄ O ₁₉ P ₃ , 4e ⁻	1 glc + 1 atp	==	C ₁₆ H ₂₃ O ₁₈ P ₃ , 5e ⁻
Directionality	1 glc + 1 atp	→	1 g6p + 1 adp + 1 h ⁺	
Location	cytosol: 1 glc + 1 atp	→	1 g6p + 1 adp + 1 h ⁺	

Thiele, I. and B. O. Palsson (2010). "A protocol for generating a high-quality genome-scale metabolic reconstruction." Nature protocols 5(1): 93-121.



Draft Reconstruction

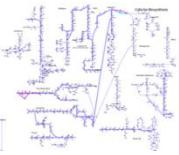
1. Obtain genome annotation
2. Identify candidate metabolic functions
3. Obtain candidate metabolic reactions
- 4. Assembly of draft reconstruction
5. Collect experimental data



Draft Reconstruction: Assembly Of Reaction Spreadsheet

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

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

Metabolite name	Metabolite description	Metabolite neutral formula	Metabolite charged formula	Metabolite charge	Metabolite Compartment	Metabolite KEGGID	Metabolite PubChemID	Metabolite CheBI ID	Metabolite Inchi String
ala-B[p]	beta-Alanine		C3H7NO2	0					
ala-D[c]	D-Alanine		C3H7NO2	0					
ala-D[e]	D-Alanine		C3H7NO2	0					
ala-D[p]	D-Alanine		C3H7NO2	0					
ala-L[c]	L-Alanine		C3H7NO2	0					
ala-L[e]	L-Alanine		C3H7NO2	0					
ala-L[p]	L-Alanine		C3H7NO2	0					

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

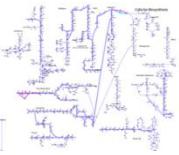




Biochemical Databases

Name	Link	Comment
KEGG	http://www.genome.jp/kegg/	
BRENDA	http://www.brenda-enzymes.info/	
Transport DB	http://www.membranetransport.org/	
PubChem	http://pubchem.ncbi.nlm.nih.gov/	
Transport Classification Database (TCDB)	http://www.tcdb.org/	TCDB is a curated database of factual information from over 10,000 published references.
pK _a Plugin	http://www.chemaxon.com/product/pka.html	Free for academic users
pK _a DB	http://www.acdlabs.com/products/phys_chem_lab/pka/	Commercial software package to determine acid-base ionization/dissociation constant, pKa

Thiele, I. and B. O. Palsson (2010). "A protocol for generating a high-quality genome-scale metabolic reconstruction." *Nature protocols* 5(1): 93-121.



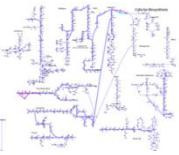
Protein Location Databases

Name	Link	Comment
PSORT	http://www.psort.org/psortb/	Support vector machine (SVM) based.
PA-SUB	http://www.cs.ualberta.ca/~bioinfo/PA/Sub/	Proteome Analyst specialized Subcellular Localization server (SVM based).

Bio-numbers

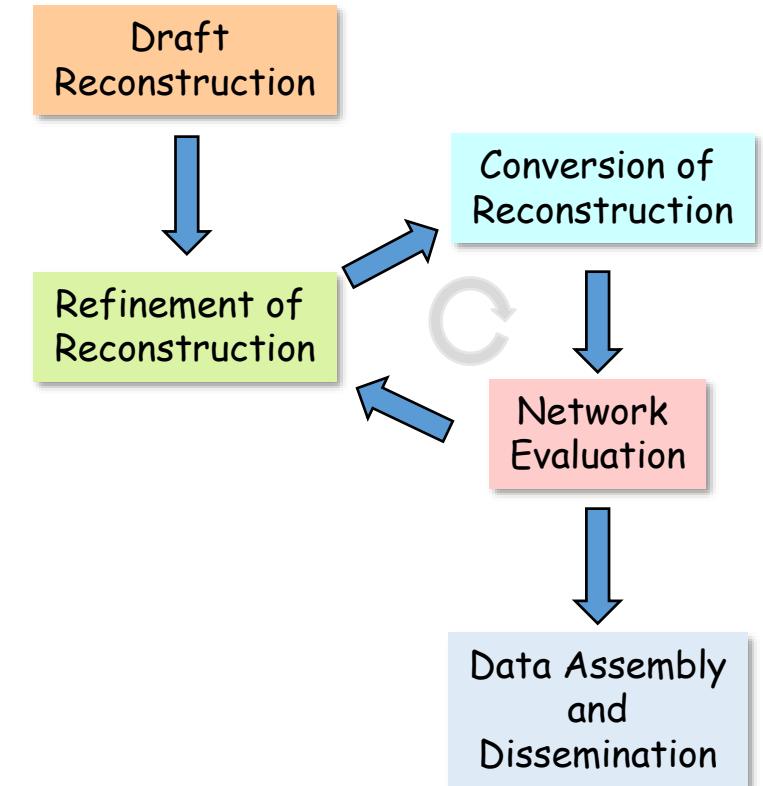
Name	Link	Comment
CyberCell Database (CCDB)	http://redpoll.pharmacy.ualberta.ca/CCDB/cgi-bin/STAT_NEW.cgi	
B10NUMB3R5	http://bionumbers.hms.harvard.edu/	

Thiele, I. and B. O. Palsson (2010). "A protocol for generating a high-quality genome-scale metabolic reconstruction." *Nature protocols* 5(1): 93-121.

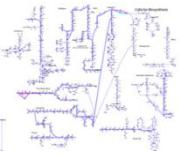


GENOME-SCALE METABOLIC RECONSTRUCTIONS

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



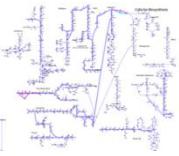
Thiele, I. and B. O. Palsson (2010). "A protocol for generating a high-quality genome-scale metabolic reconstruction." *Nature protocols* 5(1): 93-121.



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

Thiele, I. and B. O. Palsson (2010). "A protocol for generating a high-quality genome-scale metabolic reconstruction." *Nature protocols* 5(1): 93-121.



Refinement of Reconstruction

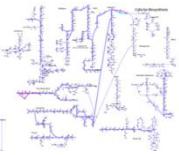
- 6. Determine and verify substrate and cofactor usage.
- 7. Obtain a neutral formula for each metabolite in the reaction
- 8. Determine the charged formula for each metabolite in the reaction.
- 9. Calculate reaction stoichiometry.
- 10. Determine reaction directionality
- 11. Add information for gene 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: 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).

Thiele, I. and B. O. Palsson (2010). "A protocol for generating a high-quality genome-scale metabolic reconstruction." *Nature protocols* 5(1): 93-121.

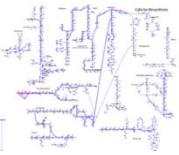


KEGG Gene Information

KEGG Escherichia coli K-12 MG1655: b1236 Help

Entry	b1236	CDS	T00007
Gene name	galU		
Definition	glucose-1-phosphate uridylyltransferase (EC:2.7.7.9)		
Orthology	K00963 UTP--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] BRITE hierarchy		

http://www.genome.jp/dbget-bin/www_bget?eco:b1236

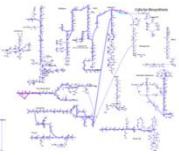


KEGG Enzyme Information

KEGG ENZYME: 2.7.7.9 Help

Entry	EC 2.7.7.9 Enzyme
Name	UTP---glucose-1-phosphate uridylyltransferase; UDP glucose pyrophosphorylase; glucose-1-phosphate uridylyltransferase; UDPG phosphorylase; UDPG pyrophosphorylase; uridine 5'-diphosphoglucose pyrophosphorylase; uridine diphosphoglucose pyrophosphorylase; uridine diphosphate-D-glucose pyrophosphorylase; uridine-diphosphate glucose pyrophosphorylase
Class	Transferases; Transferring phosphorus-containing groups; Nucleotidyltransferases BRITE hierarchy
Sysname	UTP:alpha-D-glucose-1-phosphate uridylyltransferase
Reaction (IUBMB)	UTP + alpha-D-glucose 1-phosphate = diphosphate + UDP-glucose [RN:R00289]
Reaction (KEGG)	R00289 Show all

http://www.genome.jp/dbget-bin/www_bget?ec:2.7.7.9

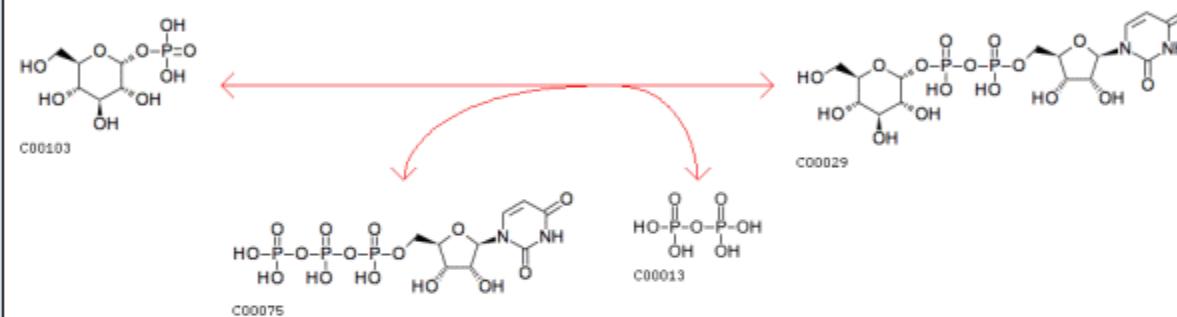


KEGG Reaction Information

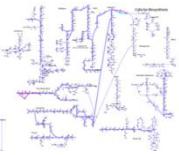


REACTION: R00289

Help

Entry	R00289	Reaction
Name	UTP:alpha-D-glucose-1-phosphate uridylyltransferase	
Definition	UTP + D-Glucose 1-phosphate \leftrightarrow Diphosphate + UDP-glucose	
Equation	$C00075 + C00103 \leftrightarrow C00013 + C00029$	
		
RPair	RP00196 C00029_C00103 main RP00382 C00013_C00075 leave RP04546 C00029_C00075 trans	
Enzyme	2.7.7.9	
Pathway	rn00040 Pentose and glucuronate interconversions rn00052 Galactose metabolism rn00500 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]	

http://www.genome.jp/dbget-bin/www_bget?reaction+R00289



Brenda Enzyme Information

BRENDA
The Comprehensive Enzyme Information System
EC 2.7.7.9 - UTP-glucose-1-phosphate uridylyltransferase

EC NUMBER	COMMENTARY
2.7.7.9	-

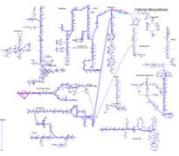
RECOMMENDED NAME GeneOntology No.
UTP-glucose-1-phosphate uridylyltransferase [GO:0003983](#)

REACTION	REACTION DIAGRAM	COMMENTARY	ORGANISM	LITERATURE
UTP + alpha-D-glucose 1-phosphate = diphosphate + UDP-glucose		mechanism	Escherichia coli	643748
UTP + alpha-D-glucose 1-phosphate = diphosphate + UDP-glucose		-	-	-

REACTION TYPE ORGANISM COMMENTARY LITERATURE
nucleotidyl group transfer - - -

PATHWAY	KEGG Link	MetaCyc Link
colanic acid building blocks biosynthesis	-	COLANSYN-PWY
galactose degradation III	-	PWY-3821
glycogen biosynthesis II (from UDP-D-Glucose)	-	PWY-5067
L-ascorbate biosynthesis VI	-	PWY3DJ-35471
stachyose degradation	-	PWY-6527
sucrose biosynthesis	-	SUCSYN-PWY
sucrose degradation III	-	PWY-621
sucrose degradation VI (anaerobic)	-	PWY-3801

http://www.brenda-enzymes.info/php/result_flat.php4?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.
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:
Obtain a Neutral and Charged Formula
for each Metabolite in the Reaction**

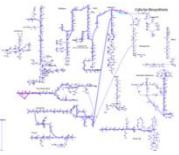
- In databases, metabolites are generally listed with their uncharged formula.
- In contrast, in medium and in cells, many metabolites are protonated or deprotonated.
- The protonation state, and thus, the charged formula, depends on the pH of interest. Often metabolic networks are reconstructed assuming 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 pK_a value of the functional groups.
- Software packages, such as Pipeline Pilot and pK_a DB, can predict the pK_a values for a given compound (<http://www.chemaxon.com/marvin/help/calculations/pKa.html>).



List Of Functional Groups, Their Charge Formula And The Corresponding pK_a

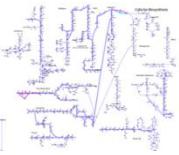
Molecule/group	Acid	Base	pK_a
Acetic acid			4.76
Carboxyl group			1.8–2.4
Ammonium			9.25
Amino group			8–11
Bicarbonate			3.77 10.2
Glycine			2.34 9.6
Phosphoric acid			2.14 6.86 12.4

Thiele, I. and B. O. Palsson (2010). "A protocol for generating a high-quality genome-scale metabolic reconstruction." *Nature protocols* 5(1): 93-121.



Neutral and Charged Formula for each Metabolite in the Reaction

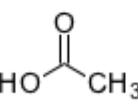
Substrates	glc	atp	g6p	adp
Neutral formulae	$C_6H_{12}O_6^0$	$C_{10}H_{16}N_5O_{13}P_3^0$	$C_6H_{13}O_9P^0$	$C_{10}H_{15}N_5O_{10}P_2^0$
	 C00031	 C00002	 C00092	 C00008
Charged formulae	$C_6H_{12}O_6^0$	$C_{10}H_{12}N_5O_{13}P_3^{4-}$	$C_6H_{11}O_9P^{2-}$	$C_{10}H_{12}N_5O_{10}P_2^{3-}$
	 C00031	 C00002	 C00092	 C00008



Example of Finding the Metabolite Charge

KEGG

COMPOUND: C00033

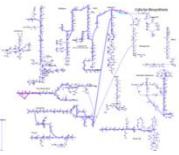
Entry	C00033	Compound
Name	Acetate; Acetic acid; Ethanoic acid	
Formula	C ₂ H ₄ O ₂	
Exact mass	60.0211	
Mol weight	60.052	
Structure	 C00033	
Mol file KCF file DB search Jmol KegDraw		

Empty lines are required

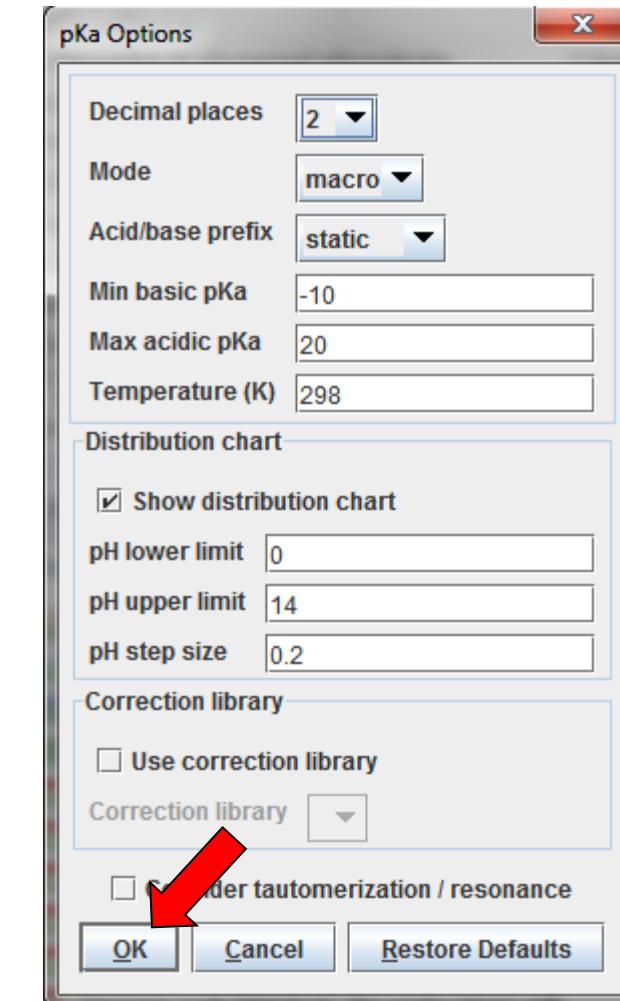
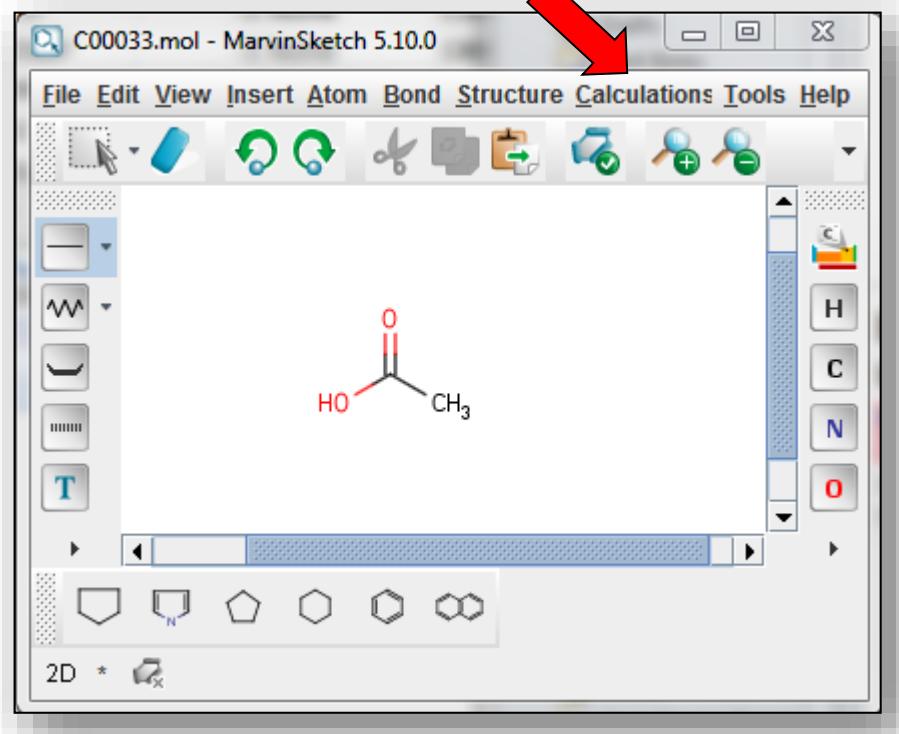
```
4 3 0 0 0 0 0 0 0 0999 V2000
24.5700 -15.7500 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
25.7840 -16.4527 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
23.3619 -16.4527 0.0000 O 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
24.5700 -14.3503 0.0000 O 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
1 2 1 0 0 0
1 3 1 0 0 0
1 4 2 0 0 0
M END
```

Acetate "mol" File

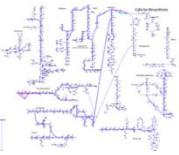
- Go to the KEGG website and enter the KEGGID
 - ✓ <http://www.genome.jp/kegg/>
- Download the "mol" file (Copy text to file; include all spaces)



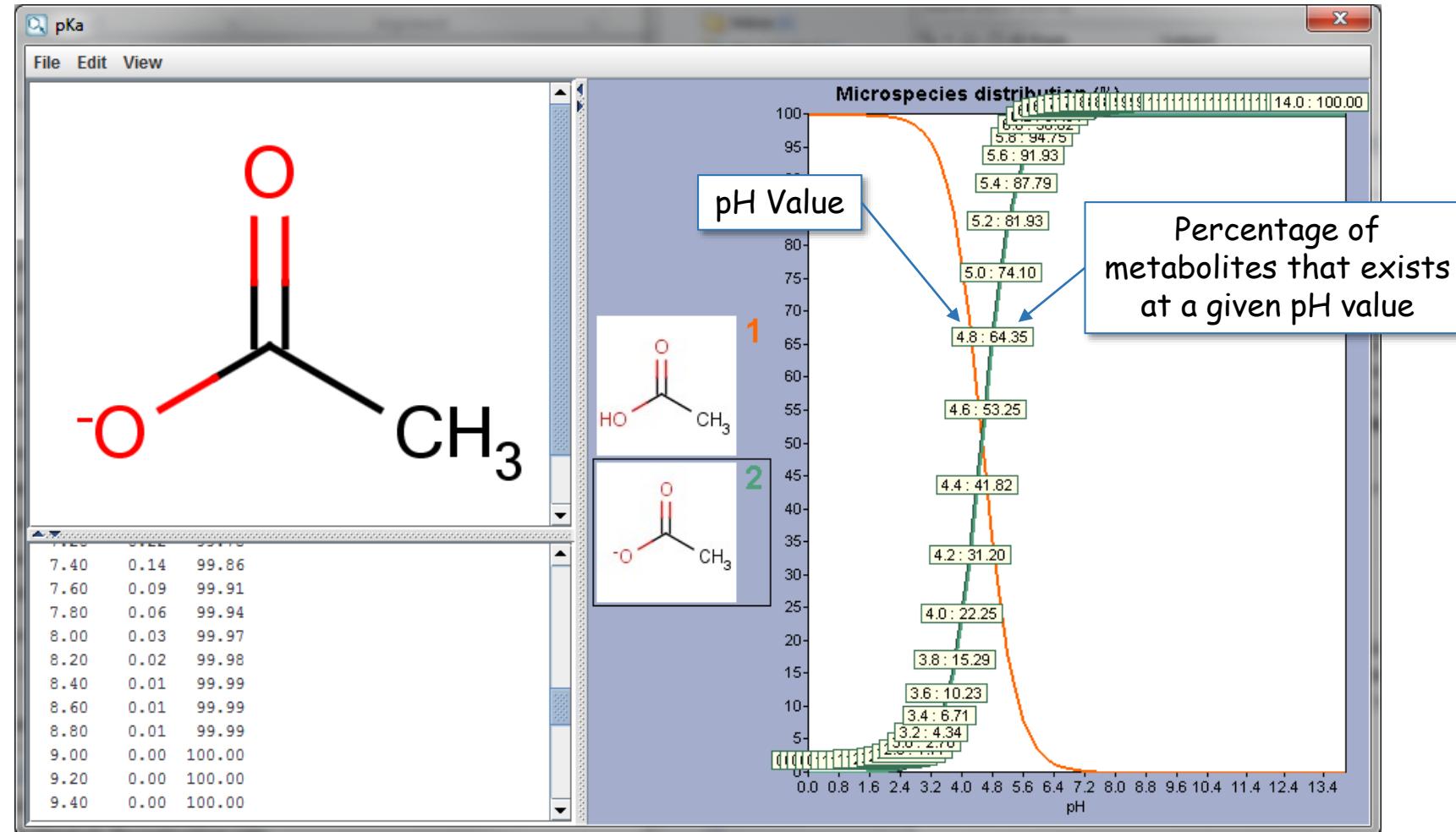
Example of Finding the Metabolite Charge

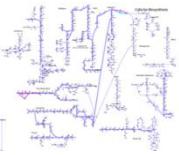


- Open the file in MarvinSpace (free to academic institutions)
 - ✓ <http://www.chemaxon.com/products/marvin/marvinspace/>
- Under the "calculations" menu:
 - ✓ calculations → protonation → pKa
- Click OK on the pKa options window



MarvinSketch Windows Showing pH Values





Marvin Tools: Example #2

EMBL-EBI

Enter Text Here Terms of Use | Privacy | Cookies

Databases Tools Research Training Industry About Us Help Site Index

EBI > Databases > Small Molecules > ChEBI > Main

acetyl-CoA (ChEBI:15351)

Search for ★★★ only

Main ChEBI Ontology Automatic Xrefs

ChEBI Name [?](#) acetyl-CoA
ChEBI ID [?](#) ChEBI:15351
Definition [?](#) An acyl-CoA having acetyl as its S-acetyl component.
Stars [?](#) ★★★ This entity has been manually annotated by the ChEBI Team.
Secondary ChEBI IDs [?](#) ChEBI:40470, ChEBI:13712, ChEBI:22192, ChEBI:2408
See structure as: Image Applet
[Download Molfile](#)

• Find compounds which contain this structure
• Find compounds which resemble this structure

Formula [?](#) C₂₃H₃₈N₇O₁₇P₃S Source KEGG COMPOUND

Net Charge [?](#) 0

Average Mass [?](#) 809.57208

COPY ONE OF THESE 3 ITEMS AND PASTE IN MARVIN VIEW

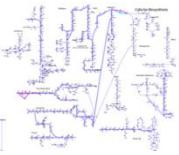
InChI [?](#) [Download](#) InChI=1/C23H38N7O17P3S/c1-12(31)51-7-6-25-14(32)4-5-26-21(35)18(34)23(2,3)9-44-50(41,42)47-49(39,40)43-8-13-17(46-48(36,37)38)16(33)22(45-13)30-11-29-15-19(24)2-10-28-20(15)30/h10-11,13,16-18,22,33-34H,4-9H2,1-3H3,(H,25,32)(H,39,40)(H,41,42)(H2,24,27,28)(H2,36,37,38)t13-16,-17,-18,-22-/m1/s1

InChIKey [?](#) [Download](#) InChIKey=ZSLZBFCDCINBPY-ZSJPKINUSA-N

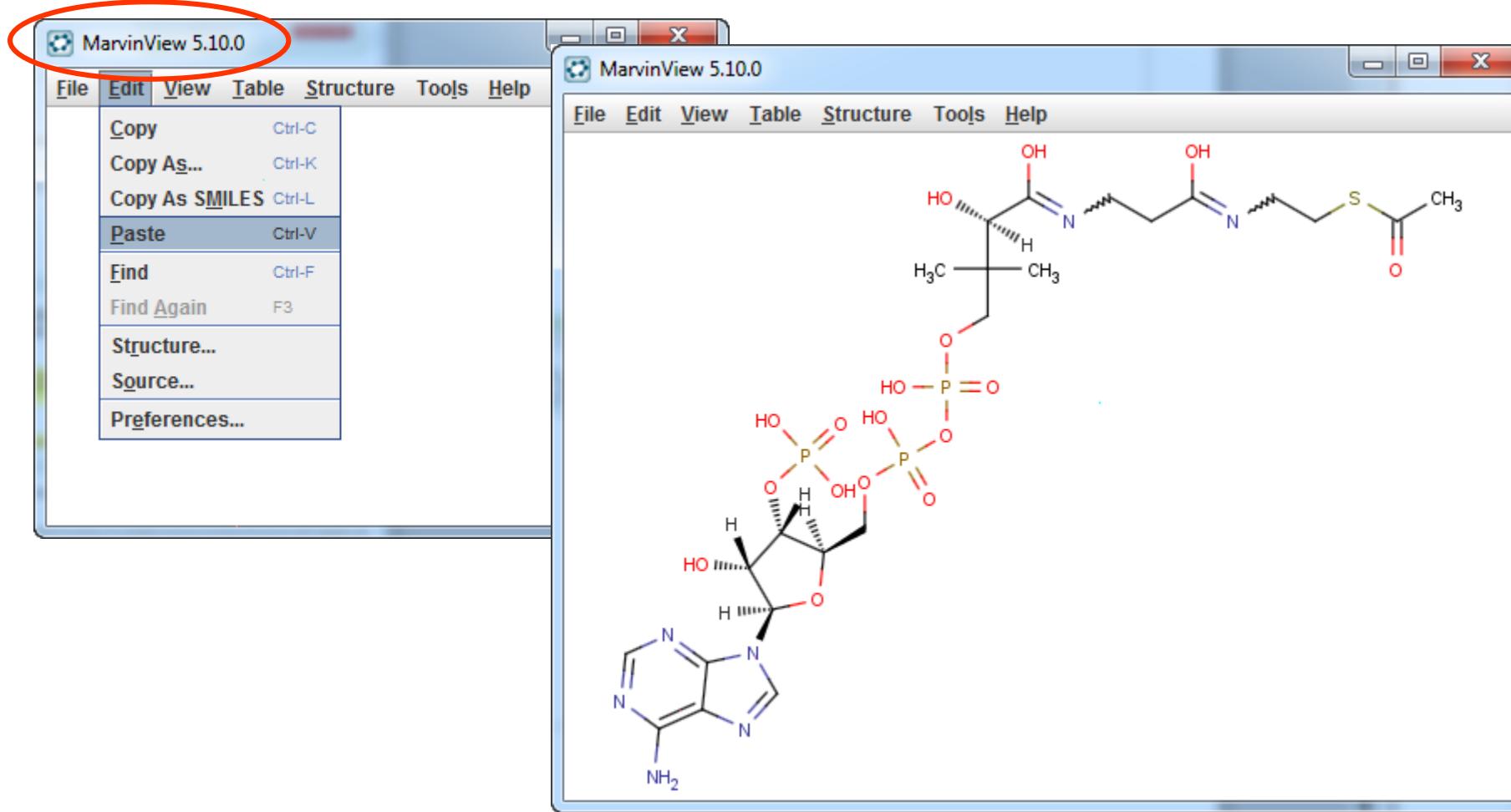
SMILES [?](#) [Download](#) CC(=O)SCCNC(=O)CCNC(=O)[C@H](O)C(C)(C)COP(O)=OOP(O)(=O)OC[C@H]1O[C@H](C[C@H](O)[C@@H](O)OP(O)(O)=O)n1cnc2c(N)ncnc12

By Dr. Wenfeng Guo

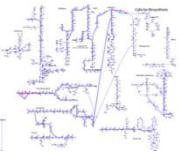
<http://www.ebi.ac.uk/chebi/searchId.do?chebiId=ChEBI:15351>



Marvin Tools: Example #2 (II)

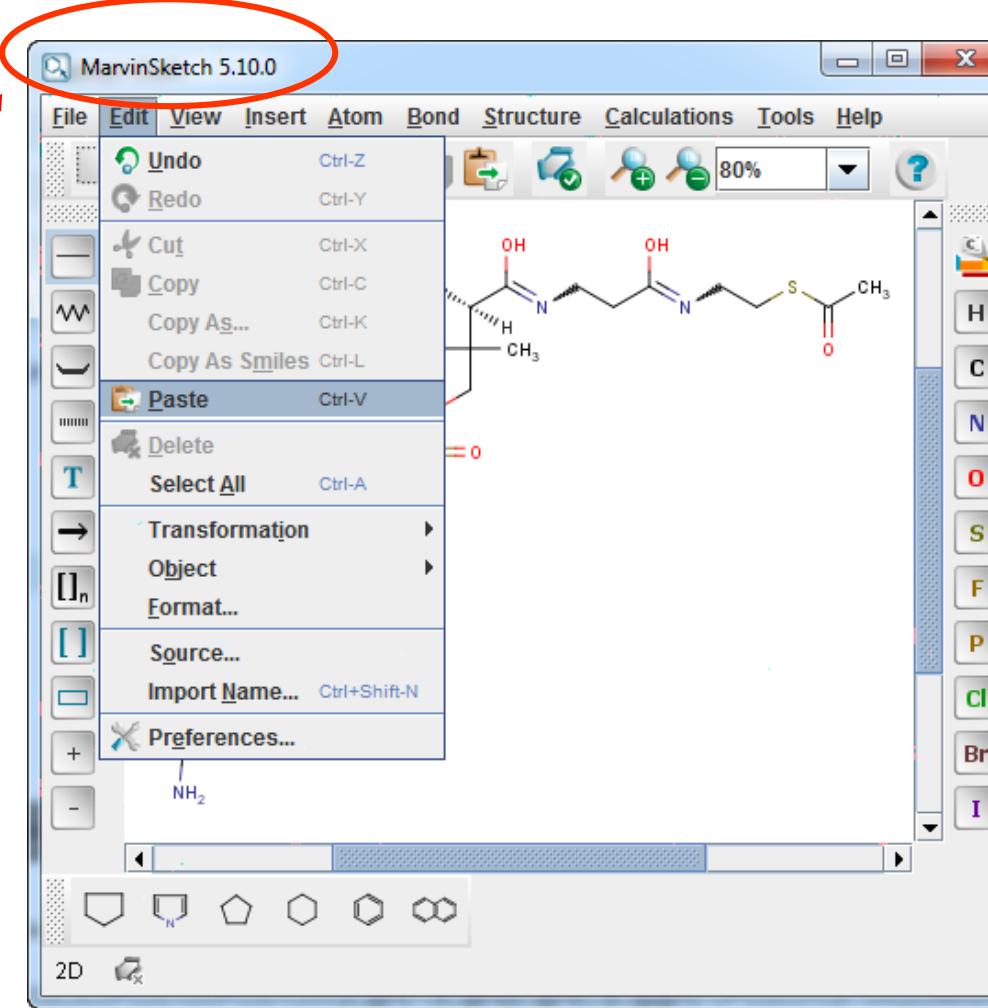


By Dr. Wenfeng Guo



Marvin Tools: Example #2 (III)

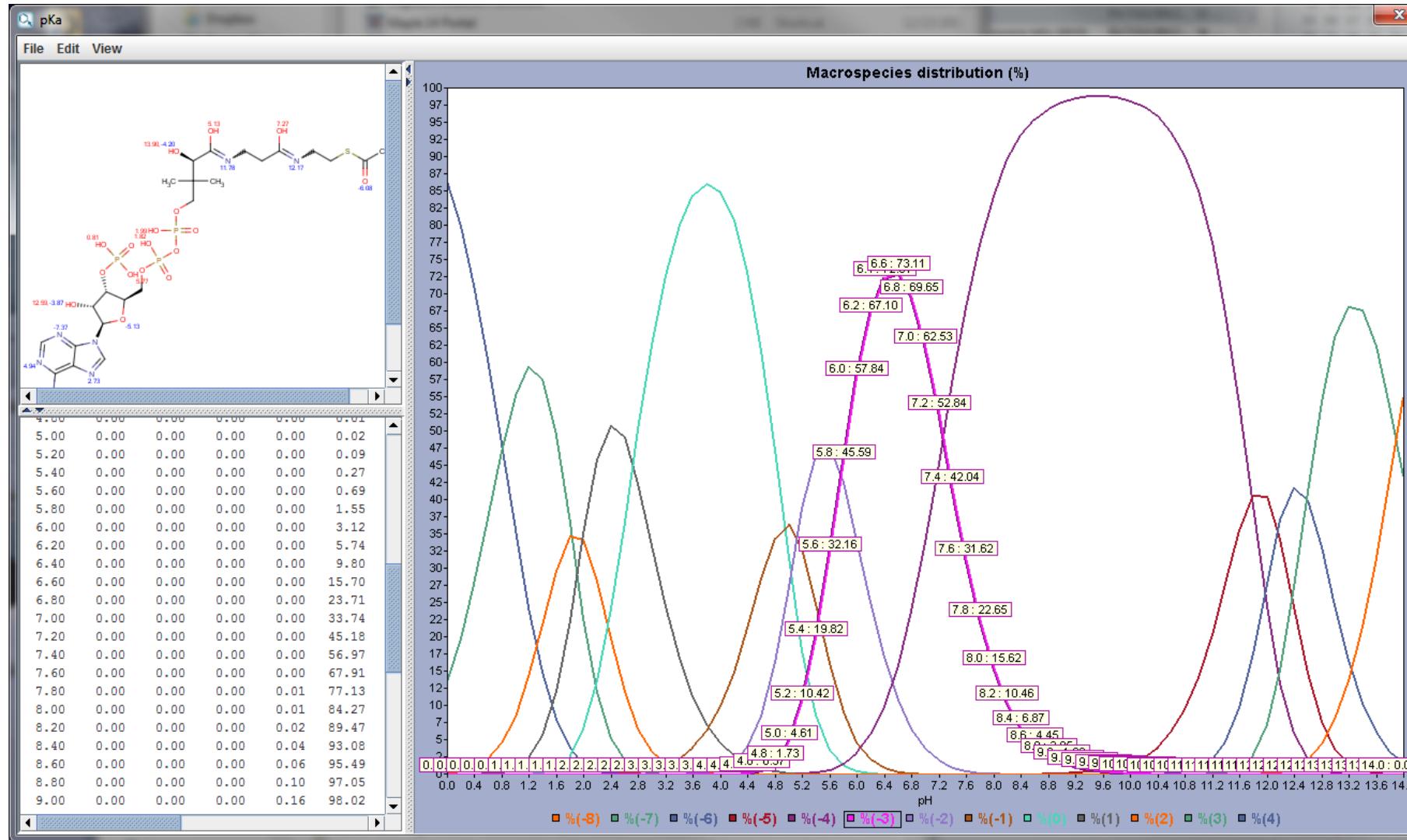
You can also cut
and paste into in
MarvinSketch

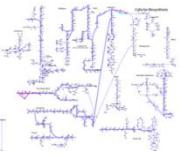


By Dr. Wenfeng Guo



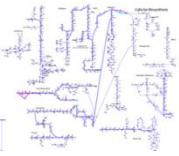
Acetyl-CoA (ChEBI:15351)





Refinement of Reconstruction

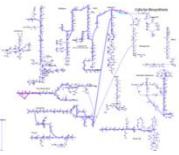
6. Determine and verify substrate and cofactor usage.
7. Obtain a neutral formula for each metabolite in the reaction
8. Determine the charged formula for each metabolite in the reaction.
9. Calculate reaction stoichiometry. →
10. Determine reaction directionality
11. Add information for gene and reaction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

Thiele, I. and B. O. Palsson (2010). "A protocol for generating a high-quality genome-scale metabolic reconstruction." *Nature protocols* 5(1): 93-121.



Calculate Reaction Stoichiometry

Substrates

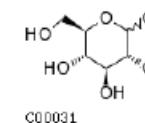
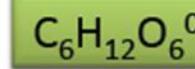
glc

atp

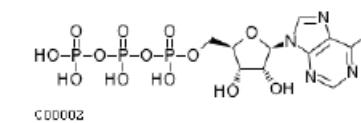
g6p

adp

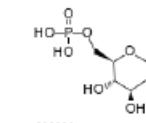
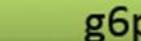
Neutral formulae



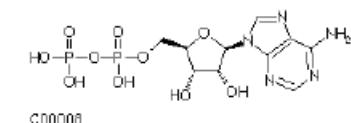
C00031



C00002

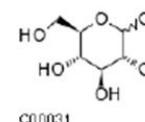
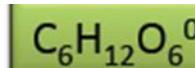


C00092

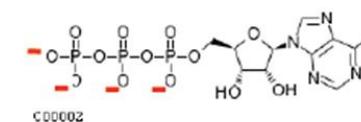
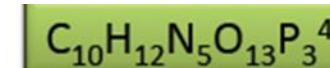


C00008

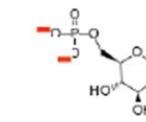
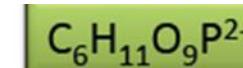
Charged formulae



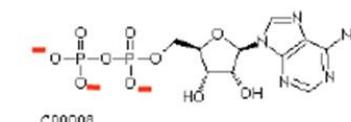
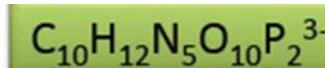
C00031



C00002

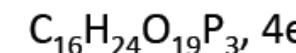
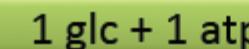


C00092

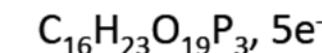


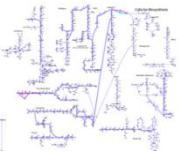
C00008

Stoichiometry



==



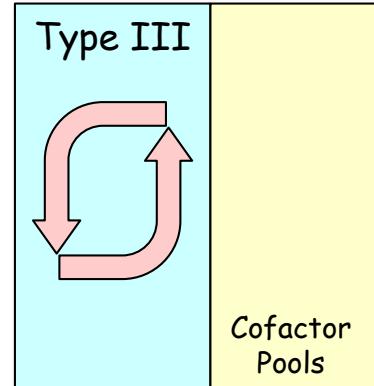


Refinement of Reconstruction

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

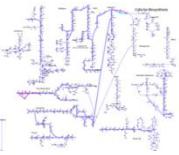


Refinement of Reconstruction: Determine Reaction Directionality



- 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

Thiele, I. and B. O. Palsson (2010). "A protocol for generating a high-quality genome-scale metabolic reconstruction." Nature protocols 5(1): 93-121.



Determine Reaction Directionality

Substrates

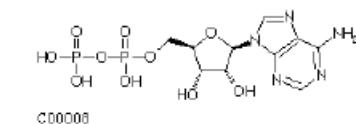
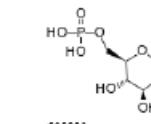
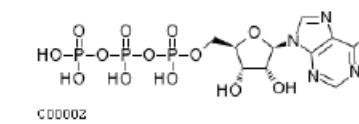
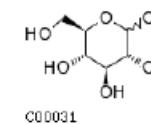
glc

atp

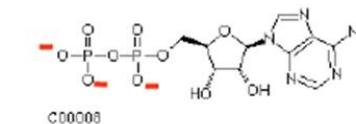
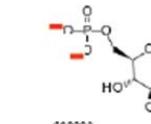
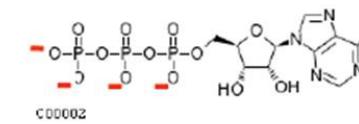
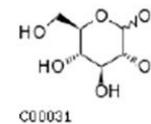
g6p

adp

Neutral formulae

 $C_6H_{12}O_6^0$ $C_{10}H_{16}N_5O_{13}P_3^0$ $C_6H_{13}O_9P^0$ $C_{10}H_{15}N_5O_{10}P_2^0$ 

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

 $1 \text{ glc} + 1 \text{ atp}$ $1 \text{ g6p} + 1 \text{ adp} + 1 h^+$ $C_{16}H_{24}O_{19}P_3, 4e^-$

==

 $C_{16}H_{23}O_{19}P_3, 5e^-$

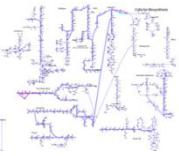
Directionality

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



Refinement of Reconstruction

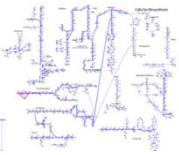
6. Determine and verify substrate and cofactor usage.
7. Obtain a neutral formula for each metabolite in the reaction
8. Determine the charged formula for each metabolite in the reaction.
9. Calculate reaction stoichiometry.
10. Determine reaction directionality
11. Add information for gene and reaction 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: Determine Gene And Reaction Localization

- The use of algorithms such as PSORT and PASUB can be considered if no experimental data are available.
 - ✓ PSORT - Gardy, J.L. et al. PSORTb v.2.0: expanded prediction of bacterial protein subcellular localization and insights gained from comparative proteome analysis. *Bioinformatics* (Oxford, England) 21, 617-623 (2005).
 - ✓ PASUB - Lu, Z. et al. Predicting subcellular localization of proteins using machine-learned classifiers. *Bioinformatics* (Oxford, England) 20, 547-556 (2004).
 - ✓ Internet-accessible tools - Emanuelsson, O., Brunak, S., von Heijne, G. & Nielsen, H. Locating proteins in the cell using TargetP, SignalP and related tools. *Nat. Protoc.* 2, 953-971 (2007).
- In the absence of appropriate data, proteins should be assumed to reside in the cytosol.

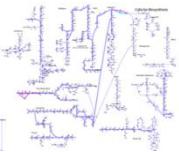
Thiele, I. and B. O. Palsson (2010). "A protocol for generating a high-quality genome-scale metabolic reconstruction." *Nature protocols* 5(1): 93-121.



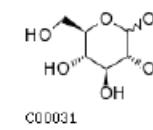
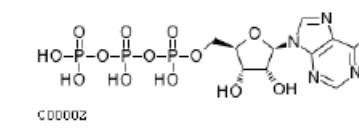
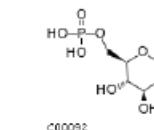
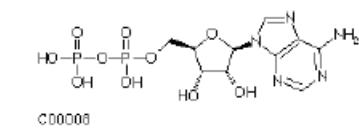
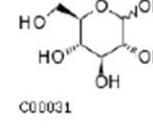
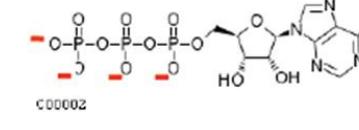
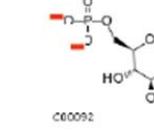
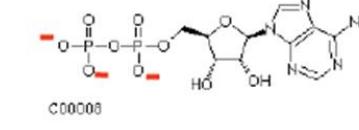
List Of Cellular Compartments Used In Reconstructions

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

Thiele, I. and B. O. Palsson (2010). "A protocol for generating a high-quality genome-scale metabolic reconstruction." *Nature protocols* 5(1): 93-121, Supplementary Methods.



Reaction/Metabolite Requirements

Substrates	glc	atp	g6p	adp
Neutral formulae	$C_6H_{12}O_6^0$	$C_{10}H_{16}N_5O_{13}P_3^0$	$C_6H_{13}O_9P^0$	$C_{10}H_{15}N_5O_{10}P_2^0$
				
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	$1 \text{ glc} + 1 \text{ atp}$		$1 \text{ g6p} + 1 \text{ adp} + 1 h^+$	
	$C_{16}H_{24}O_{19}P_3, 4e^-$	$=$	$C_{16}H_{23}O_{19}P_3, 5e^-$	
Directionality	$1 \text{ glc} + 1 \text{ atp} \rightarrow$		$1 \text{ g6p} + 1 \text{ adp} + 1 h^+$	
Location	$\text{cytosol: } 1 \text{ glc} + 1 \text{ atp} \rightarrow$		$1 \text{ g6p} + 1 \text{ adp} + 1 h^+$	



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 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>

1. Metabolism

1.1 Carbohydrate Metabolism

Glycolysis / Gluconeogenesis
Citrate cycle (TCA cycle)
Pentose phosphate pathway
Pentose and glucuronate interconversions
Fructose and mannose metabolism
Galactose metabolism
Ascorbate and aldarate metabolism
Starch and sucrose metabolism
Amino sugar and nucleotide sugar metabolism
Pyruvate metabolism
Glyoxylate and dicarboxylate metabolism
Propanoate metabolism
Butanoate metabolism
C5-Branched dibasic acid metabolism
Inositol phosphate metabolism

1.2 Energy Metabolism

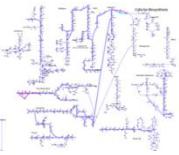
Oxidative phosphorylation
Photosynthesis
Photosynthesis - antenna proteins
Carbon fixation in photosynthetic organisms
Carbon fixation pathways in prokaryotes
Methane metabolism
Nitrogen metabolism
Sulfur metabolism

Thiele, I. and B. O. Palsson (2010). "A protocol for generating a high-quality genome-scale metabolic reconstruction." *Nature protocols* 5(1): 93-121.



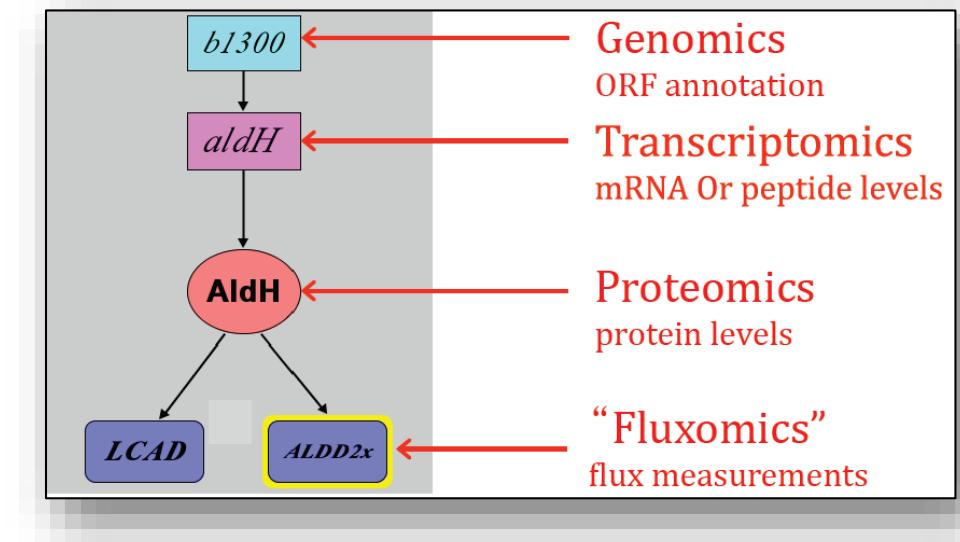
Refinement of Reconstruction

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



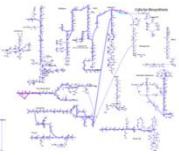
Refinement of Reconstruction: Verify GPR Association

- 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).

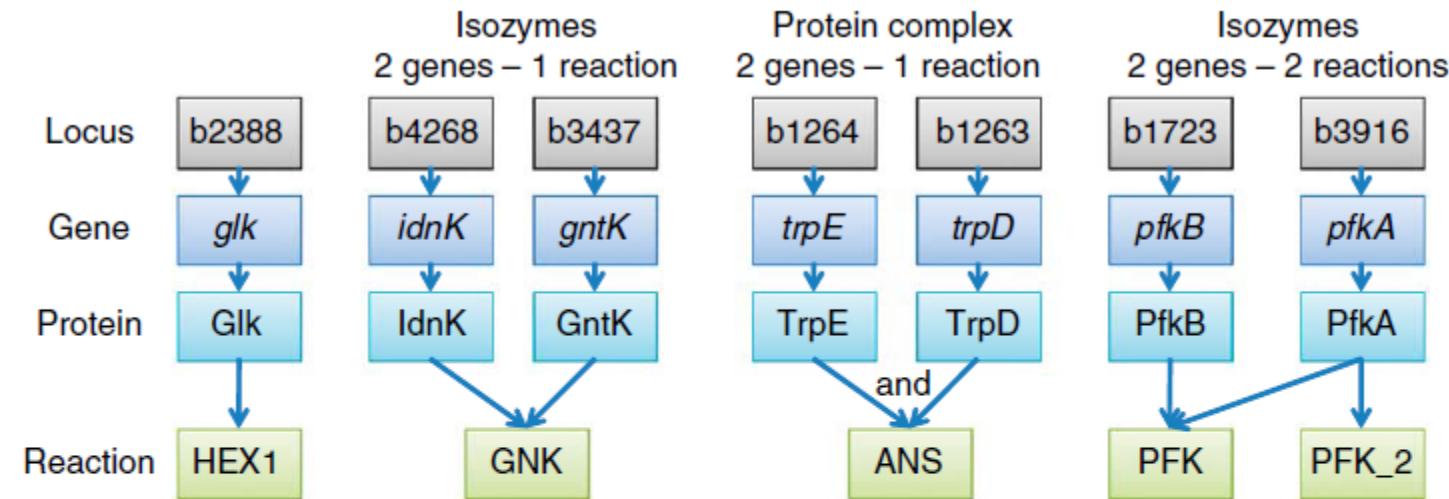


- Linear pathways, such as fatty acid oxidation, have often been combined into few lumped reactions. The genes associated with these reactions are all required, with the exception of isozymes. Subsequently, the GPR association should reflect the requirement for all genes within the lumped reaction by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

Thiele, I. and B. O. Palsson (2010). "A protocol for generating a high-quality genome-scale metabolic reconstruction." *Nature protocols* 5(1): 93-121.



Examples of GPR Associations and their Representation in Boolean Format



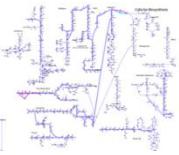
Reaction abbreviation	Reaction name	E. C.number	GPR
HEX1	Hexokinase (D-glucose:ATP)	2.7.1.1	(b2388)
GNK	Gluconokinase	2.7.1.12	(b3437) or (b4268)
ANS	Anthranoate synthase	4.1.3.27	(b1264) and (b1263)
PFK	Phosphofructokinase	2.7.1.11	(b1723) or (b3916)
PFK_2	Phosphofructokinase (2)	2.7.1.11	(b3916)

Thiele, I. and B. O. Palsson (2010). "A protocol for generating a high-quality genome-scale metabolic reconstruction." Nature protocols 5(1): 93-121.



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 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/>)
 - PubChem (<http://pubchem.ncbi.nlm.nih.gov/>)
 - 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:
 - SMILES (http://en.wikipedia.org/wiki/Simplified_molecular-input_line-entry_system)
 - 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/>),

Thiele, I. and B. O. Palsson (2010). "A protocol for generating a high-quality genome-scale metabolic reconstruction." *Nature protocols* 5(1): 93-121.



Refinement of Reconstruction

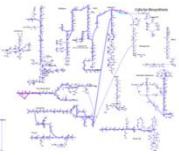
6. Determine and verify substrate and cofactor usage.
7. Obtain a neutral formula for each metabolite in the reaction
8. Determine the charged formula for each metabolite in the reaction.
9. Calculate reaction stoichiometry.
10. Determine reaction directionality
11. Add information for gene and reaction 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: 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.

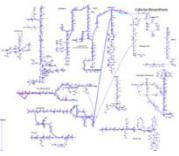
Thiele, I. and B. O. Palsson (2010). "A protocol for generating a high-quality genome-scale metabolic reconstruction." *Nature protocols* 5(1): 93-121.



Confidence Scoring System Currently Employed for Metabolic Reconstructions

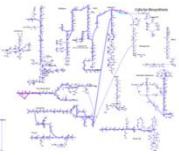
Evidence type	Confidence score	Examples
Biochemical data	4	Direct evidence for gene product function and biochemical reaction: protein purification, biochemical assays, experimentally solved protein structures and comparative gene-expression studies
Genetic data	3	Direct and indirect evidence for gene function: knockout characterization, knock-in characterization and overexpression
Physiological data	2	Indirect evidence for biochemical reactions based on physiological data: secretion products or defined medium components serve as evidence for transport and metabolic reactions
Sequence data	2	Evidence for gene function: genome annotation and SEED annotation
Modeling data	1	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)
Not evaluated	0	

Thiele, I. and B. O. Palsson (2010). "A protocol for generating a high-quality genome-scale metabolic reconstruction." *Nature protocols* 5(1): 93-121.



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



Reaction Spreadsheet

ecoli_iaf1260.xls

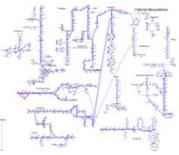
Rxn name	Rxn description	Formula
FUCtpp	L-fucose transport via proton symport (periplasm)	fuc-L[p] + h[p] \leftrightarrow fuc-L[c] + h[c]
FUM	fumarase	fum[c] + h2o[c] \leftrightarrow mal-L[c]
FUMt2_2pp	Fumarate transport via proton symport (2 H) (periplasm)	fum[p] + 2 h[p] \rightarrow fum[c] + 2 h[c]
FUMt2_3pp	Fumarate transport via proton symport (3 H) (periplasm)	fum[p] + 3 h[p] \rightarrow fum[c] + 3 h[c]
FUMtex	Fumarate transport via diffusion (extracellular to periplasm)	fum[e] \leftrightarrow fum[p]



Gene-reaction association	Genes	Proteins	Subsystem
(b2801)	b2801		Transport, Inner Membrane
(b1612) or (b1611) or (b4122)	b1611 b1612 b4122		Citric Acid Cycle
(b3528)	b3528		Transport, Inner Membrane
(b4138) or (b0621) or (b4123)	b0621 b4123 b4138		Transport, Inner Membrane
(b2215) or (b0241) or (b1377) or (b0929)	b0241 b0929 b1377 b2215		Transport, Outer Membrane Porin



Reversible	LB	UB	Objective	Confidence Score	EC Number	Notes	References
1	-1000	1000	0	0		JLR	
1	-1000	1000	0	0	4.2.1.2	LOCUS:b1611#ABBREVIATION:fumC#ECNUMBERS:4.2.1.2#	
0	0	1000	0	4		JLR	
0	0	1000	0	4		JLR	
1	-1000	1000	0			LOCUS:b0241#ABBREVIATION:phoE#	
0	0	1000	0	0	2.3.1.157	LOCUS:b3730#ABBREVIATION:gImU#ECNUMBERS:2.7.7.23#	



Metabolite Spreadsheet

ecoli_iaf1260.xls

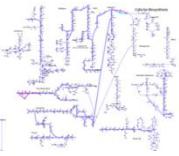
Metabolite name	Metabolite description	Metabolite neutral formula
ala-D[p]	D-Alanine	
ala-L[c]	L-Alanine	
ala-L[e]	L-Alanine	
ala-L[p]	L-Alanine	
alaala[c]	D-Alanyl-D-alanine	
alaala[e]	D-Alanyl-D-alanine	



Metabolite charged formula	Metabolite charge	Metabolite Compartment	Metabolite KEGGID
C3H7NO2	0		
C6H12N2O3	0		



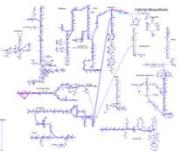
Metabolite PubChemID	Metabolite CheBI ID	Metabolite Inchi String	Metabolite Smile



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 Spontaneous Reactions

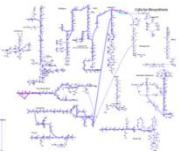
- The biochemical literature and databases (e.g., KEGG and BRENDA) are to be used to identify candidate spontaneous reactions that are to be included.
- Only include those reactions, which have at least one metabolite present in the reconstruction to minimize the number of dead ends.
- Associate the spontaneous reactions with an artificial gene (s0001) and protein (S0001).

Thiele, I. and B. O. Palsson (2010). "A protocol for generating a high-quality genome-scale metabolic reconstruction." Nature protocols 5(1): 93-121.



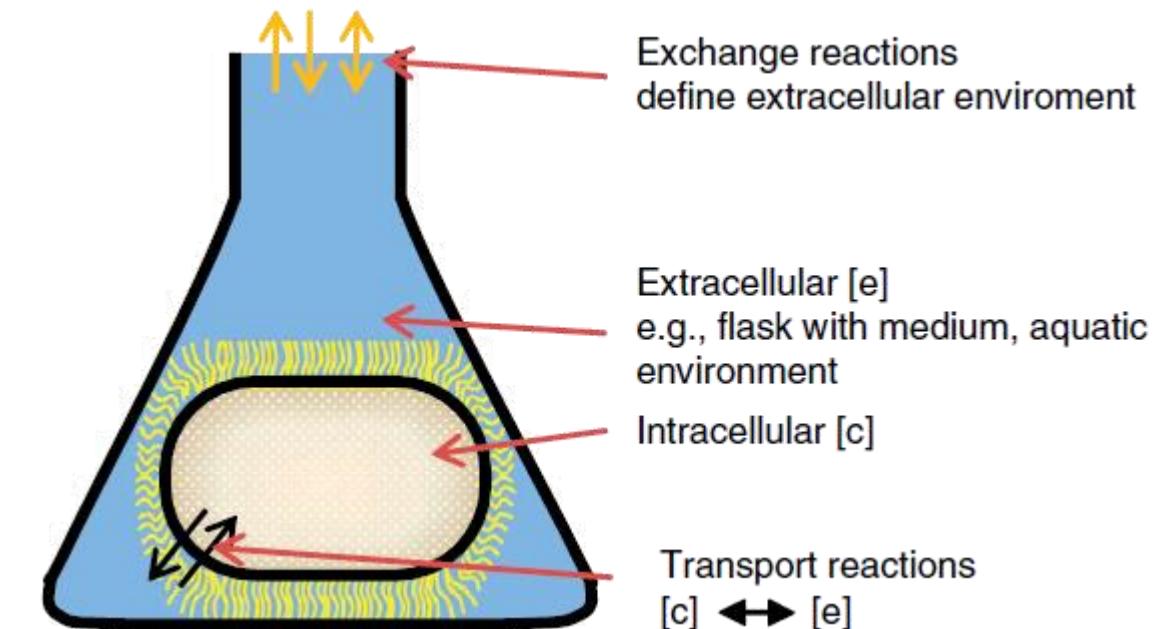
Refinement of Reconstruction

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 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.



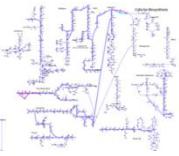
Thiele, I. and B. O. Palsson (2010). "A protocol for generating a high-quality genome-scale metabolic reconstruction." *Nature protocols* 5(1): 93-121.



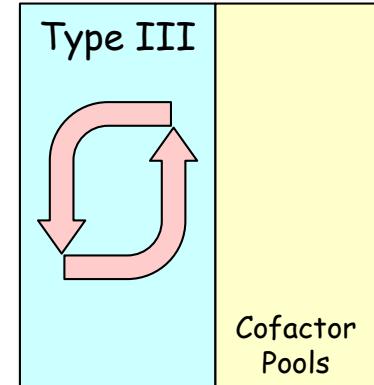
Refinement of Reconstruction

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 Intracellular Transport Reactions



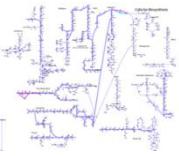
- When multi-compartment networks are constructed, intracellular transport reactions need to be added for all the metabolites that are supposed to 'move' between compartments.
- Minimize the number of intracellular transport reactions to the ones that really need to be there.
- If too many transport 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.

Thiele, I. and B. O. Palsson (2010). "A protocol for generating a high-quality genome-scale metabolic reconstruction." *Nature protocols* 5(1): 93-121.



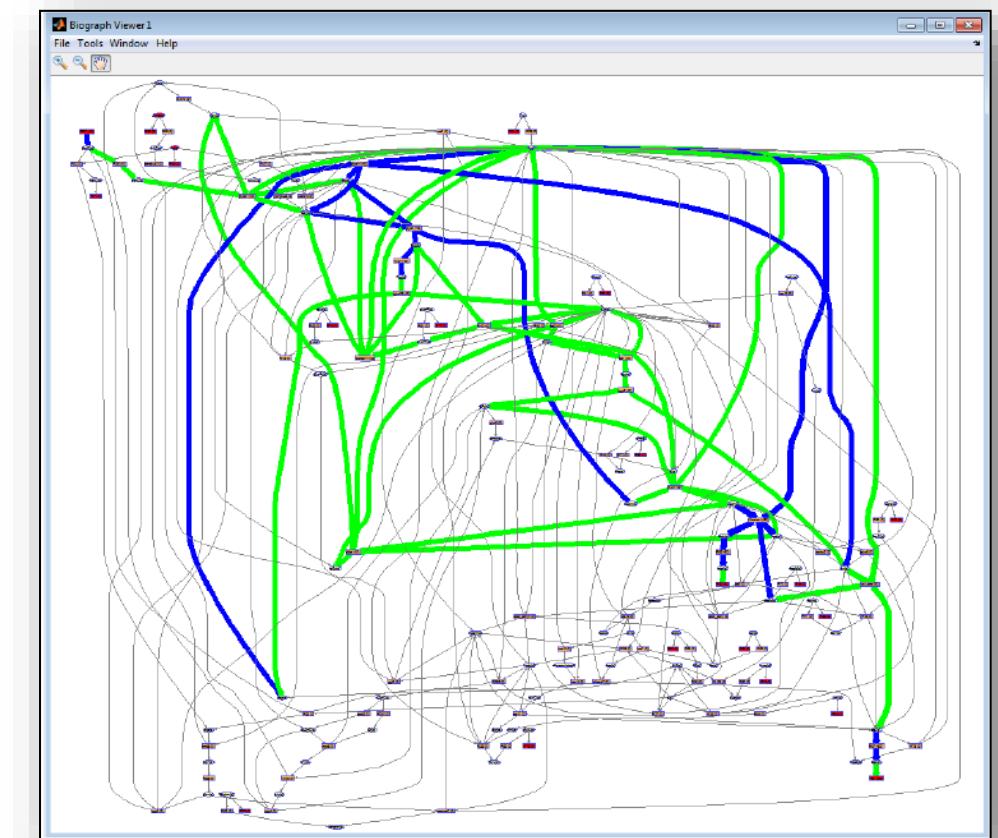
Refinement of Reconstruction

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

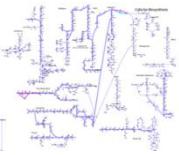


Refinement of Reconstruction: Draw Metabolic Map

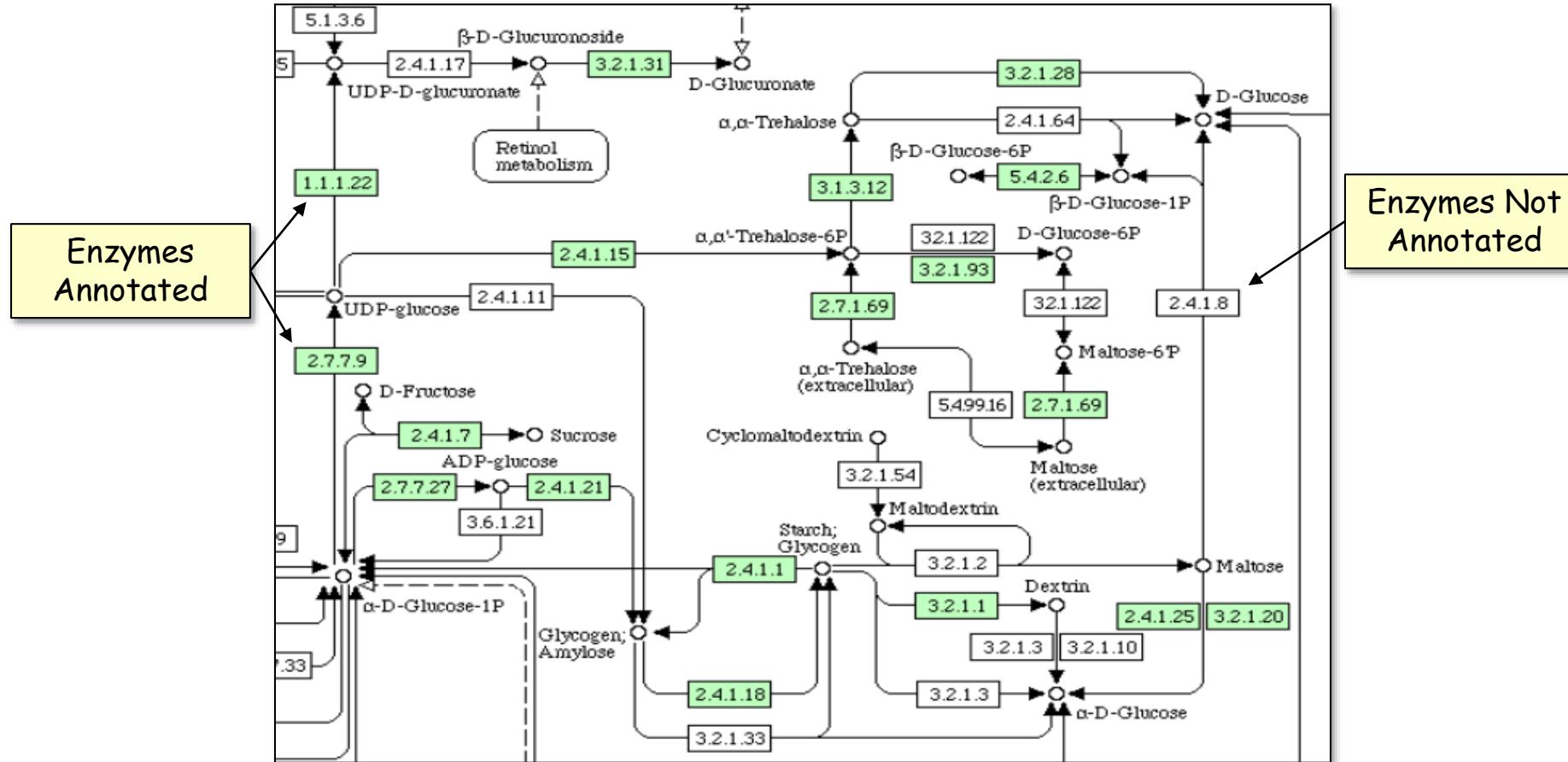
- Paint4Net Developed by Andrejs Kostromins
- Paint4Net v1.0 is the COBRA Toolbox extension for visualization of constraints-based reconstruction and analysis (COBRA) models and reconstructions in the MATLAB environment.
- Uses the Bioinformatics toolbox to visualize COBRA models and reconstructions 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



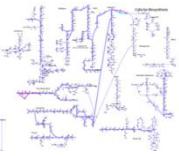
Kostromins, A. and E. Stalidzans (2012). "Paint4Net: COBRA Toolbox extension for visualization of stoichiometric models of metabolism." Bio Systems.



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

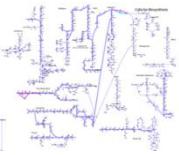


Thiele, I. and B. O. Palsson (2010). "A protocol for generating a high-quality genome-scale metabolic reconstruction." Nature protocols 5(1): 93-121, Supplementary Methods.



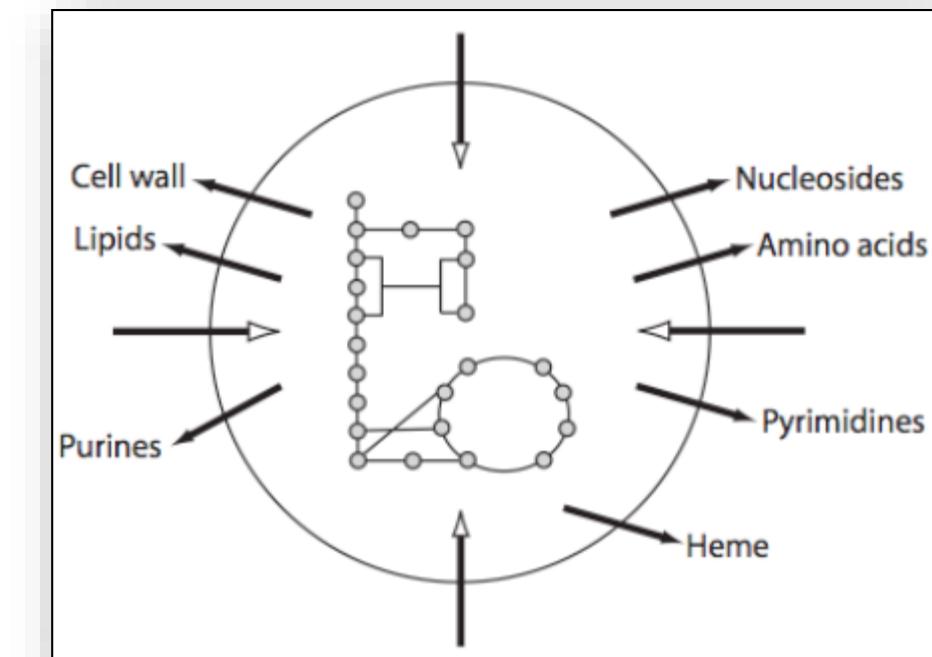
Refinement of Reconstruction

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

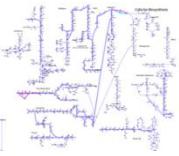


Refinement of Reconstruction: Determine Biomass Composition

- 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.



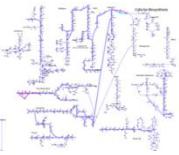
Thiele, I. and B. O. Palsson (2010). "A protocol for generating a high-quality genome-scale metabolic reconstruction." *Nature protocols* 5(1): 93-121.



Important Role of the Biomass Objective Function

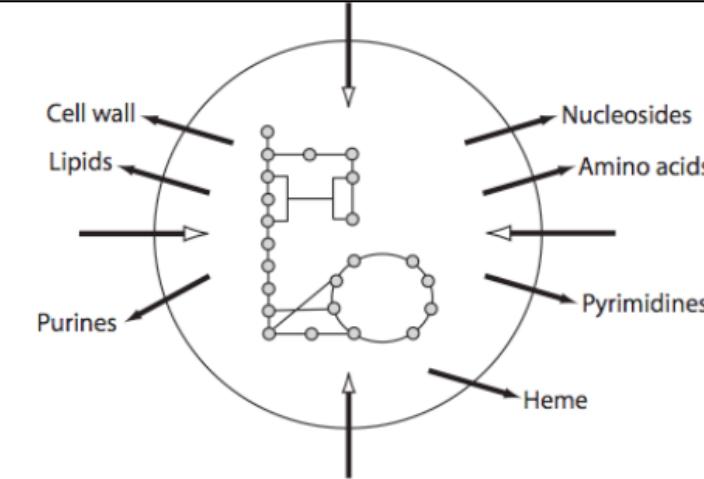
- If a biomass precursor is not accounted for in the biomass reactions, the synthesis reactions may not be required for growth (i.e., it is nonessential). Therefore, associated genes may not be assumed as essential. Subsequently, the presence or absence of a metabolite in the biomass reaction may affect 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 $\text{mmol}\cdot\text{gDW}^{-1}$. Therefore, the biomass reaction sums the mole fraction of each precursor necessary to produce 1 g dry weight of cells.

Thiele, I. and B. O. Palsson (2010). "A protocol for generating a high-quality genome-scale metabolic reconstruction." *Nature protocols* 5(1): 93-121.

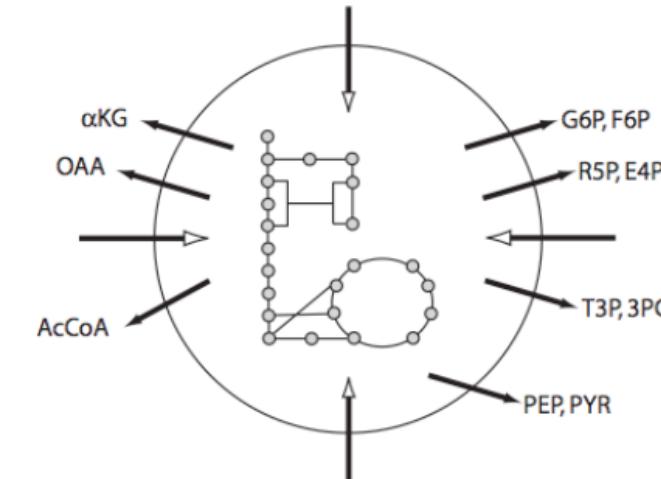


Definition of Biomass Reaction

Quantifying
Macromolecular Content
of a cell



Quantifying Building
Blocks of
Macromolecules



REI601M, Introduction to Systems Biology, Dr. Innes Thiele, 2012, <https://systemsbiology.hi.is/wiki/REI601M>



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

A	Cellular Component	Cellular Content % (w/w)	B	Monomer	(mmol/g _{DW})	Monomer	(mmol/g _{DW})
	Protein	55%		Ala	0.482	Leu	0.507
	RNA	20.5%		Arg	0.286	Lys	0.145
	DNA	3.1%		Asn	0.128	Met	0.098
	Lipids	9.1%		Asp	0.230	Phe	0.154
	LPS	3.4%		Cys	0.045	Pro	0.212
	Peptidoglycan	2.5%		Glu	0.243	Ser	0.243
	Glycogen	2.5%		Gln	0.203	Thr	0.206
	Polyamines	0.4%		Gly	0.348	Trp	0.063
	Other	3.5%		His	0.102	Tyr	0.110
	Total	100.00%		Ile	0.197	Val	0.314

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

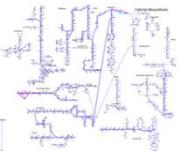
Phospholipid contributions to the biomass function where PE is Phosphatidylethanolamine, PG is phosphatidylglycerol, and CL is cardiolipin.

C	Phospholipids	(mol/mol)	Average MW	Content % (w/w)	mmol/g _{DW}	D	DNA Monomer	Number of bp	Content % (mol/mol)	(mmol/g _{DW})
	PE	64.95%	699.1	58.20%	0.0325		dATP	1186504	19.19%	0.0122
	PG	21.50%	700.3	19.30%	0.0324		dCTP	1889954	30.57%	0.0197
	CL	10.06%	1508	19.45%	0.0151		dGTP	1913381	30.95%	0.0195

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.

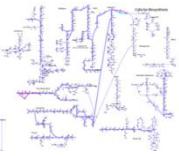
Thiele, I. and B. O. Palsson (2010). "A protocol for generating a high-quality genome-scale metabolic reconstruction." Nature protocols 5(1): 93-121, Supplementary Methods.



Ecoli_iaf1260 Core Biomass Objective Function Spreadsheet

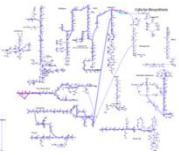
E. coli CORE Biomass Objective Function										
	Notes	Macromolecule	overall wt%	overall wt% (wild type)	composition (molar fraction)	mmol/gDW (Calc.)	metabolite	location	type	formula
Reactant										
NH pg. 96		Protein	0.5794	0.55	0.096	0.513689	ala-L	cytoplasm	AA	C3H7NO2
NH pg. 96					0.055	0.295792	arg-L	cytoplasm	AA	C6H15N4O2
NH pg. 96					0.045	0.241055	asn-L	cytoplasm	AA	C4H8N2O3
NH pg. 96					0.045	0.241055	asp-L	cytoplasm	AA	C4H6NO4
NH pg. 96					0.017	0.091580	cys-L	cytoplasm	AA	C3H7NO2S
NH pg. 96					0.049	0.263160	gln-L	cytoplasm	AA	C5H10N2O3
NH pg. 96					0.049	0.263160	glu-L	cytoplasm	AA	C5H8NO4
NH pg. 96					0.115	0.612638	gly	cytoplasm	AA	C2H5NO2
NH pg. 96					0.018	0.094738	his-L	cytoplasm	AA	C6H9N3O2
NH pg. 96					0.054	0.290529	ile-L	cytoplasm	AA	C6H13NO2
NH pg. 96					0.084	0.450531	leu-L	cytoplasm	AA	C6H13NO2
NH pg. 96					0.064	0.343161	lys-L	cytoplasm	AA	C6H15N2O2
NH pg. 96					0.029	0.153686	met-L	cytoplasm	AA	C5H11NO2S
NH pg. 96					0.035	0.185265	phe-L	cytoplasm	AA	C9H11NO2
NH pg. 96					0.041	0.221055	pro-L	cytoplasm	AA	C5H9NO2
NH pg. 96					0.040	0.215792	ser-L	cytoplasm	AA	C3H7NO3
NH pg. 96					0.047	0.253687	thr-L	cytoplasm	AA	C4H9NO3
NH pg. 96					0.011	0.056843	trp-L	cytoplasm	AA	C11H12N2O2
NH pg. 96					0.026	0.137896	tyr-L	cytoplasm	AA	C9H11NO3
NH pg. 96					0.079	0.423162	val-L	cytoplasm	AA	C5H11NO2
GC content f DNA			0.0327	0.031	0.246	0.026166	dntp	cytoplasm	DNA	C10H12N5O12P3
GC content f					0.254	0.027017	dotp	cytoplasm	DNA	C9H12N3O13P3
GC content f					0.254	0.027017	dgtp	cytoplasm	DNA	C10H12N5O13P3
GC content f					0.246	0.026166	dtp	cytoplasm	DNA	C10H13N2O14P3
NH pg. 98	RNA		0.216	0.205	0.200	0.133508	ctp	cytoplasm	RNA	C9H12N3O14P3
NH pg. 98					0.322	0.215096	ctp	cytoplasm	RNA	C10H12N5O14P3
NH pg. 98					0.216	0.144104	utp	cytoplasm	RNA	C9H11N2O15P3
NH pg. 98					0.262	0.174831	atp**	cytoplasm	RNA	C10H12N5O13P3
NH pg. 4, set murein			0.0263	0.025	1	0.013894	murein5px4p	periplasm	murein	C77H117N15O40
NH pg. 4, ani LPS			0.0358	0.034	1	0.019456	kdo2lipid4	extra-cellular face	LPS	C84H148N2O37P2
NH pg. 4, set lipid			0.0959	0.091	0.4590	0.063814	pe160	cytoplasm / periplasm	lipid	C37H74N108P1
NH pg. 4, set					0.5410	0.075214	pe161	cytoplasm / periplasm	lipid	C37H70N108P1

Feist, A. M., C. S. Henry, et al. (2007). "A genome-scale metabolic reconstruction for Escherichia coli K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information." Molecular Systems Biology 3: 121, Supplementary Information 3.



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 Amino Acid 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.

Cellular Component	Cellular Content % (w/w)
Protein	55%
RNA	20.5%
DNA	3.1%
Lipids	9.1%
LPS	3.4%
Peptidoglycan	2.5%
Glycogen	2.5%
Polyamines	0.4%
Other	3.5%
Total	100.00%

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

Phospholipids	(mol/mol) %	Average MW	Content % (w/w)	mmol/g _{DW}
PE	64.95%	699.1	58.20%	0.0325
PG	21.50%	700.3	19.30%	0.0324
CL	10.06%	1508	19.45%	0.0151

Monomer	(mmol/g _{DW})	Monomer	(mmol/g _{DW})
Ala	0.482	Leu	0.507
Arg	0.286	Lys	0.145
Asn	0.128	Met	0.098
Asp	0.230	Phe	0.154
Cys	0.045	Pro	0.212
Glu	0.243	Ser	0.243
Gln	0.203	Thr	0.206
Gly	0.348	Trp	0.063
His	0.102	Tyr	0.110
Ile	0.197	Val	0.314

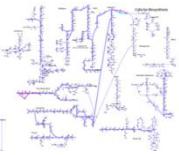
Protein composition in *P. putida* broken down by monomer contributions in mmol/gDW.



DNA Monomer	Number of bp	Content % (mol/mol)	(mmol/g _{DW})
dATP	1186504	19.19%	0.0122
dCTP	1889954	30.57%	0.0197
dGTP	1913381	30.95%	0.0195
dTTP	1192024	19.28%	0.0123
Total	6181863	100.00%	

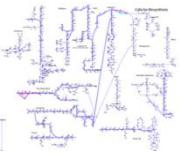
dNTP composition of the entire *P. putida* chromosomal genome. The data are obtained from direct measurements, literature, or can be estimated from genome information.

Thiele, I. and B. O. Palsson (2010). "A protocol for generating a high-quality genome-scale metabolic reconstruction." Nature protocols 5(1): 93-121, Supplementary Methods.

Escherichia coli K12-MG1655 Genome Page 

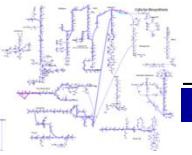
<p>Gene Search</p> <p>Search by: Locus Match: <input checked="" type="radio"/> Exact <input type="radio"/> Inexact Keywords/Accession: <input type="text"/> <input type="button" value="Search"/></p> <p>Region View</p> <p>Molecule: Chromosome Escherichia coli <input type="button" value="Search"/> Coordinates: <input type="text"/> - <input type="text"/> <input type="button" value="Search"/></p> <p>Sequence Retrieval</p> <p>Molecule: Chromosome Escherichia coli <input type="button" value="Search"/> Coordinates: <input type="text"/> - <input type="text"/> Strand: <input checked="" type="radio"/> Forward <input type="radio"/> Reverse <input type="button" value="Search"/></p> <p>Gene Retrieval By Coords</p> <p>Molecule: Chromosome Escherichia coli <input type="button" value="Search"/> Coordinates: <input type="text"/> - <input type="text"/> <input type="button" value="Search"/></p> <p>Region Comparison</p> <p>Locus: <input type="text"/> <input type="button" value="Search"/></p>	<p>General Information for Escherichia coli K12-MG1655</p> <p>Sequencing Center: Univ of Wisconsin Funding Center: NIH-NHGRI Publication: PubMed Abstract Sequence & Annotation Acquisition: CMR Batch Download Other: CMR Gene Attribute Download GenBank FTP Download Sequencing Center Genome Page Background Information Completed Genome: Yes CMR Version Number: Version 1.0 Date added to the CMR: January-December 2000 GenBank Accession.Version: U00096.2 Genome Properties: View All Properties</p> <p>Taxonomy of Escherichia coli K12-MG1655</p> <p>Taxon ID: 168927 Kingdom: Bacteria Intermediate Rank 1: Proteobacteria Intermediate Rank 2: Gammaproteobacteria Intermediate Rank 3: Enterobacteriales Intermediate Rank 4: Enterobacteriaceae Intermediate Rank 5: Escherichia</p>	<p>Statistics for Escherichia coli K12-MG1655</p> <p>DNA Molecule Summary</p> <table border="1"><tr><td>Total Number of all DNA molecules:</td><td>1</td><td>100.00%</td></tr><tr><td>Total Size of all DNA molecules:</td><td>4639221 bp</td><td>100.00%</td></tr><tr><td>Number of Primary Annotation coding bases:</td><td>4070314 bp</td><td>87.73%</td></tr><tr><td>Number of JCVI Annotation coding bases:</td><td>4233750 bp</td><td>91.25%</td></tr><tr><td>Number of G+C bases:</td><td>2356208 bp</td><td>50.78%</td></tr></table> <p>Primary Annotation Summary</p> <table border="1"><tr><td>Total genes:</td><td>4289</td><td>100.00%</td></tr><tr><td>Protein coding genes:</td><td>4289</td><td>100.00%</td></tr><tr><td>Genes assigned a role category:</td><td>2052</td><td>47.84%</td></tr><tr><td>Genes not assigned a role category:</td><td>665</td><td>15.50%</td></tr><tr><td>Conserved hypothetical genes:</td><td>898</td><td>20.93%</td></tr><tr><td>Hypothetical genes:</td><td>674</td><td>15.71%</td></tr></table> <p>JCVI Automated Annotation Summary</p> <table border="1"><tr><td>Total genes:</td><td>5397</td><td>100.00%</td></tr><tr><td>Protein coding genes:</td><td>5287</td><td>97.96%</td></tr><tr><td>Genes assigned a role category:</td><td>2067</td><td>39.09%</td></tr><tr><td>Genes not assigned a role category:</td><td>669</td><td>12.65%</td></tr><tr><td>Conserved hypothetical genes:</td><td>909</td><td>17.19%</td></tr><tr><td>Hypothetical genes:</td><td>1642</td><td>31.05%</td></tr><tr><td>tRNA genes:</td><td>88</td><td>1.63%</td></tr><tr><td>rRNA genes:</td><td>22</td><td>0.40%</td></tr></table>	Total Number of all DNA molecules:	1	100.00%	Total Size of all DNA molecules:	4639221 bp	100.00%	Number of Primary Annotation coding bases:	4070314 bp	87.73%	Number of JCVI Annotation coding bases:	4233750 bp	91.25%	Number of G+C bases:	2356208 bp	50.78%	Total genes:	4289	100.00%	Protein coding genes:	4289	100.00%	Genes assigned a role category:	2052	47.84%	Genes not assigned a role category:	665	15.50%	Conserved hypothetical genes:	898	20.93%	Hypothetical genes:	674	15.71%	Total genes:	5397	100.00%	Protein coding genes:	5287	97.96%	Genes assigned a role category:	2067	39.09%	Genes not assigned a role category:	669	12.65%	Conserved hypothetical genes:	909	17.19%	Hypothetical genes:	1642	31.05%	tRNA genes:	88	1.63%	rRNA genes:	22	0.40%
Total Number of all DNA molecules:	1	100.00%																																																									
Total Size of all DNA molecules:	4639221 bp	100.00%																																																									
Number of Primary Annotation coding bases:	4070314 bp	87.73%																																																									
Number of JCVI Annotation coding bases:	4233750 bp	91.25%																																																									
Number of G+C bases:	2356208 bp	50.78%																																																									
Total genes:	4289	100.00%																																																									
Protein coding genes:	4289	100.00%																																																									
Genes assigned a role category:	2052	47.84%																																																									
Genes not assigned a role category:	665	15.50%																																																									
Conserved hypothetical genes:	898	20.93%																																																									
Hypothetical genes:	674	15.71%																																																									
Total genes:	5397	100.00%																																																									
Protein coding genes:	5287	97.96%																																																									
Genes assigned a role category:	2067	39.09%																																																									
Genes not assigned a role category:	669	12.65%																																																									
Conserved hypothetical genes:	909	17.19%																																																									
Hypothetical genes:	1642	31.05%																																																									
tRNA genes:	88	1.63%																																																									
rRNA genes:	22	0.40%																																																									

<http://cmr.jcvi.org/cgi-bin/CMR/GenomePage.cgi?org=ntec01>



Create Codon Usage Table

The screenshot shows the JCVI CMR Codon Usage Query Page. At the top, there is a navigation bar with links for Home, Genome Tools, Searches, Comparative Tools, Lists, Downloads, and Carts. A search bar is also present. On the left, a sidebar titled "Codon Usage Table" includes a "CMR Blast" section and a "Get Seq/Gene By Coords" section. The main content area is divided into two sections: "1. Select a genome and DNA molecule:" and "2. Optional - Restrict the codon table to one or more role categories:". In the first section, a dropdown menu lists various Escherichia coli strains, with "Escherichia coli K12-MG1655" selected. In the second section, a dropdown menu lists biological process categories like "Amino acid biosynthesis" and "Central intermediary metabolism". An "Add >>" button is located at the bottom of the second section.

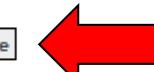


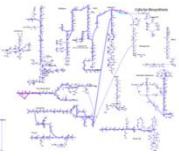
Codon Usage Table

	2nd								
1st	T	C	A	G					3rd
T	TTT Phe (F) 1.84% 133	TCT Ser (S) 0.69% 50	TAT Tyr (Y) 1.58% 114	TGT Cys (C) 0.63% 46	T				
	TTC Phe (F) 1.36% 98	TCC Ser (S) 0.97% 70	TAC Tyr (Y) 0.95% 69	TGC Cys (C) 0.69% 50	C				
	TTA Leu (L) 1.66% 120	TCA Ser (S) 0.58% 42	TAA STOP 0.16% 12	TGA STOP 0.12% 9	A				
	TTG Leu (L) 1.48% 107	TCG Ser (S) 0.86% 62	TAG STOP 0% 0	TGG Trp (W) 0.63% 46	G				
C	CTT Leu (L) 1.05% 76	CCT Pro (P) 0.72% 52		CGT Arg (R) 2.06% 149	T				
	CTC Leu (L) 1.24% 90	CCC Pro (P) 0.55% 40	CAT His (H) 1.34% 97	CGC Arg (R) 2.76% 199	C				
	CTA Leu (L) 0.27% 20	CCA Pro (P) 0.91% 66	CAC His (H) 1.16% 84	CGA Arg (R) 0.24% 18	A				
	CTG Leu (L) 5.27% 380	CCG Pro (P) 2.31% 167	CAA Gln (Q) 1.69% 122	CGG Arg (R) 0.41% 30	G				
A	ATT Ile (I) 3.44% 248	ACT Thr (T) 0.98% 71	AAT Asn (N) 1.94% 140	AGT Ser (S) 0.81% 59	T				
	ATC Ile (I) 2.06% 149	ACC Thr (T) 2.20% 159	AAC Asn (N) 1.73% 125	AGC Ser (S) 1.61% 116	C				
	ATA Ile (I) 0.16% 12	ACA Thr (T) 0.55% 40	AAA Lys (K) 3.30% 238	AGA Arg (R) 0.12% 9	A				
	ATG Met (M) 2.70% 195	ACG Thr (T) 1.31% 95	AAG Lys (K) 0.76% 55	AGG Arg (R) 0.04% 3	G				
G	GTT Val (V) 1.72% 124	GCT Ala (A) 1.84% 133	GAT Asp (D) 3.56% 257	GGT Gly (G) 2.49% 180	T				
	GTC Val (V) 1.49% 108	GCC Ala (A) 3.08% 222	GAC Asp (D) 1.77% 128	GGC Gly (G) 3.26% 235	C				
	GTA Val (V) 1.05% 76	GCA Ala (A) 2.04% 147	GAA Glu (E) 4.55% 328	GGA Gly (G) 0.81% 59	A				
	GTG Val (V) 2.33% 168	GCG Ala (A) 3.74% 270	GAG Glu (E) 2.09% 151	GGG Gly (G) 1.15% 83	G				

21 total ORFs, 7204 codons

Fields: [Codon Triplet] [% Codon Frequency] [Number of Codon Occurrences]

[View the Downloadable Codon Table](#)

Codon Usage Selected Organism: *Escherichia coli K12-MG1655*

Selected Annotation: Primary Annotation

[Download](#)Selected Molecule Name: Chromosome *Escherichia coli K12-MG1655*

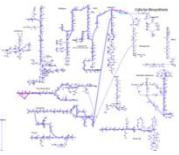
Roles: Amino acid biosynthesis: Aromatic amino acid family

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

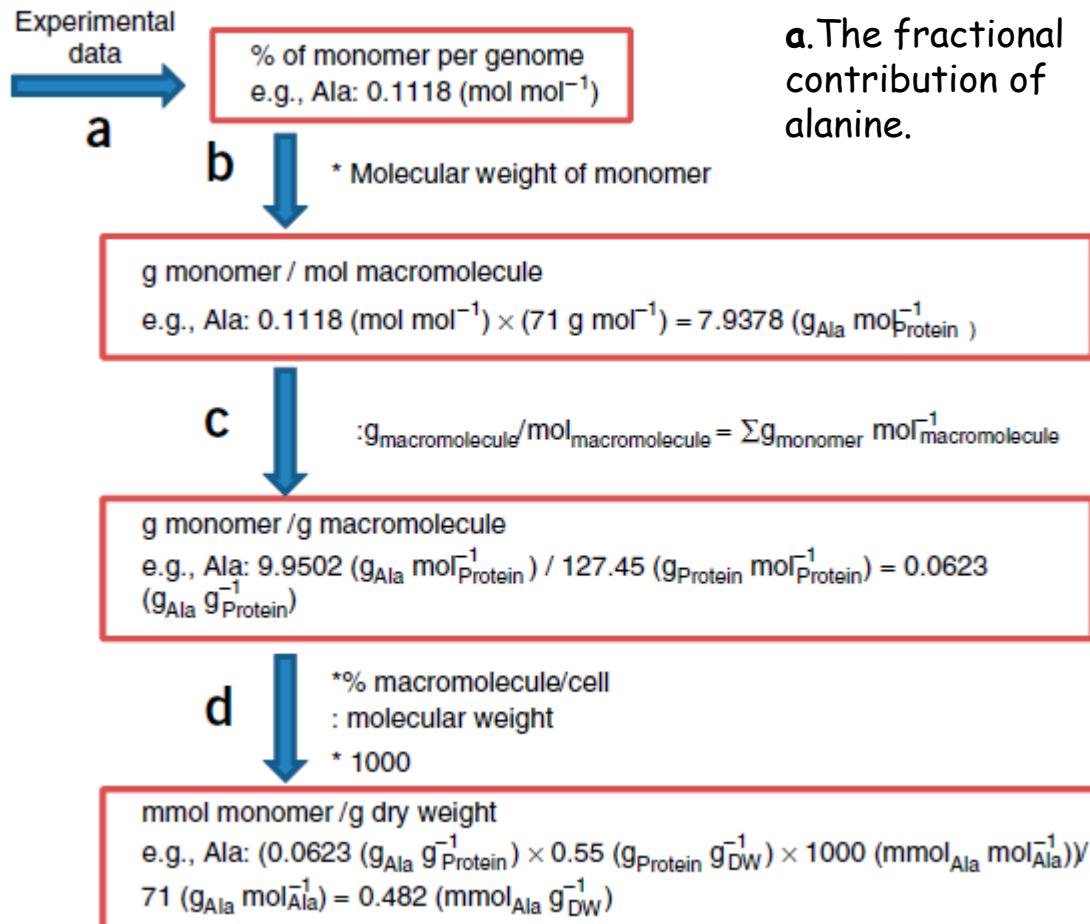


Refinement of Reconstruction: Determine Biomass Composition

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.

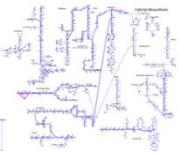
d. The weight of alanine is multiplied by the cellular content of protein and divided by its molecular weight to obtain the mole alanine per cell dry weight. Multiplying this molar contribution by a factor of 1,000 will result in a final unit of mmol alanine per gram of dry weight.

Thiele, I. and B. O. Palsson (2010). "A protocol for generating a high-quality genome-scale metabolic reconstruction." Nature protocols 5(1): 93-121.



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.

Cellular Component	Cellular Content % (w/w)
Protein	55%
RNA	20.5%
DNA	3.1%
Lipids	9.1%
LPS	3.4%
Peptidoglycan	2.5%
Glycogen	2.5%
Polyamines	0.4%
Other	3.5%
Total	100.00%

Monomer	(mmol/g _{DW})	Monomer	(mmol/g _{DW})
Ala	0.482	Leu	0.507
Arg	0.286	Lys	0.145
Asn	0.128	Met	0.098
Asp	0.230	Phe	0.154
Cys	0.045	Pro	0.212
Glu	0.243	Ser	0.243
Gln	0.203	Thr	0.206
Gly	0.348	Trp	0.063
His	0.102	Tyr	0.110
Ile	0.197	Val	0.314

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

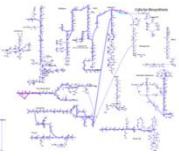
Phospholipids	(mol/mol)	Average MW	Content % (w/w)	mmol/g _{DW}
PE	64.95%	699.1	58.20%	0.0325
PG	21.50%	700.3	19.30%	0.0324
CL	10.06%	1508	19.45%	0.0151

DNA Monomer	Number of bp	Content % (mol/mol)	(mmol/g _{DW})
dATP	1186504	19.19%	0.0122
dCTP	1889954	30.57%	0.0197
dGTP	1913381	30.95%	0.0195
dTTP	1192024	19.28%	0.0123
Total	6181863	100.00%	

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

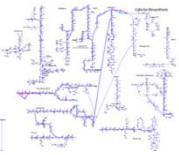
dNTP composition of the entire *P. putida* chromosomal genome. The data are obtained from direct measurements, literature, or can be estimated from genome information.

Thiele, I. and B. O. Palsson (2010). "A protocol for generating a high-quality genome-scale metabolic reconstruction." Nature protocols 5(1): 93-121, Supplementary Methods.



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

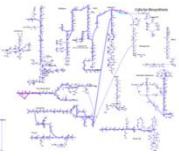
- Experimental determination of the nucleotide content,
 - ✓ Obtain data for each deoxynucleotide triphosphate (dATP, dCTP, dGTP and dTTP) and each nucleotide triphosphate (ATP, CTP, GTP and UTP).
- Estimation of nucleotide composition from genome information
 - ✓ For example, use CMR database. From 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

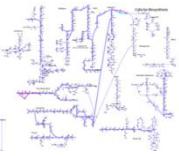
Selected Organism: *Pseudomonas putida KT2440*[Download](#)

Molecule Name	chromosome <i>Pseudomonas putida KT2440</i>	
Type	chromosome	
Topology	circular	
GenBank Accession.Version	AE015451.1	
Sequence Length	6181863 bp	100.00%
Primary Annotation Coding Regions	5415589 bp	87.60%
Primary Annotation Intergenic Regions	766274 bp	12.39%
Primary Annotation: Number of Genes	5437	100.00%
Primary Annotation: Number of Genes assigned to role ids	3850	70.81%
Primary Annotation: Number of Genes not assigned to role ids	0	0.00%
Primary Annotation: Conserved Hypothetical Genes	989	18.19%
Primary Annotation: Hypothetical Genes	598	10.99%
Number of A	1186504 bp	19.19%
Number of T	1192024 bp	19.28%
Number of G	1913381 bp	30.95%
Number of C	1889954 bp	30.57%
Number of A+T	2378528 bp	38.47%
Number of G+C	3803335 bp	61.52%



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

Cellular Component	Cellular Content % (w/w)
Protein	55%
RNA	20.5%
DNA	3.1%
Lipids	9.1%
LPS	3.4%
Peptidoglycan	2.5%
Glycogen	2.5%
Polyamines	0.4%
Other	3.5%
Total	100.00%

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

Monomer	(mmol/g _{DW})	Monomer	(mmol/g _{DW})
Ala	0.482	Leu	0.507
Arg	0.286	Lys	0.145
Asn	0.128	Met	0.098
Asp	0.230	Phe	0.154
Cys	0.045	Pro	0.212
Glu	0.243	Ser	0.243
Gln	0.203	Thr	0.206
Gly	0.348	Trp	0.063
His	0.102	Tyr	0.110
Ile	0.197	Val	0.314

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

Phospholipids	(mol/mol)	Average MW	Content % (w/w)	mmol/g _{DW}
PE	64.95%	699.1	58.20%	0.0325
PG	21.50%	700.3	19.30%	0.0324
CL	10.06%	1508	19.45%	0.0151

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

DNA Monomer	Number of bp	Content % (mol/mol)	(mmol/g _{DW})
dATP	1186504	19.19%	0.0122
dCTP	1889954	30.57%	0.0197
dGTP	1913381	30.95%	0.0195
dTTP	1192024	19.28%	0.0123
Total	6181863	100.00%	

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

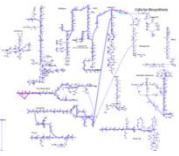
Thiele, I. and B. O. Palsson (2010). "A protocol for generating a high-quality genome-scale metabolic reconstruction." Nature protocols 5(1): 93-121, Supplementary Methods.



Determine Biomass Composition: Determine the Lipid Content

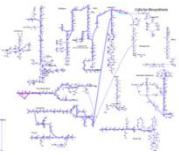
- Determine the contributions from fatty acids and phospholipids.
 - i. Determine the average molecular weight of a fatty acid in the cell by incorporating the average fatty acid composition of the cell (requires experimental data, e.g., from literature).
 - ii. The average molecular weight of each fatty 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.

Thiele, I. and B. O. Palsson (2010). "A protocol for generating a high-quality genome-scale metabolic reconstruction." *Nature protocols* 5(1): 93-121.



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

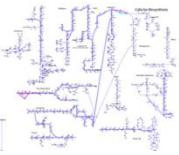


Determine Biomass Composition:
The soluble pool: contains polyamines,
vitamins and cofactors (e.g. *E.coli*)

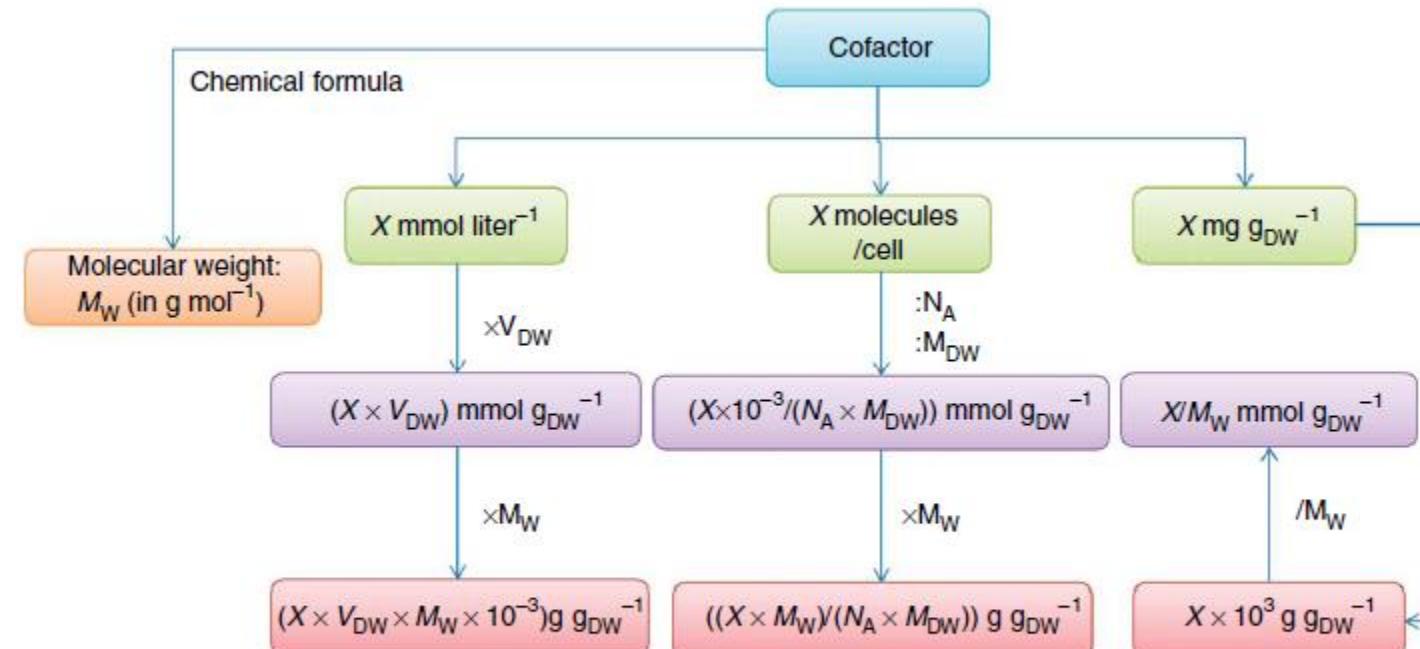
Abbr	Name
putre	Putrescine
spmd	Spermidine
accoa	Acetyl-CoA
coa	Coenzyme A (CoA)
succoa	Succinyl-CoA
malcoa	Malonyl-CoA
nad	Nicotinamide adenine dinucleotide
nadh	Nicotinamide adenine dinucleotide - reduced
nadp	Nicotinamide adenine dinucleotide phosphate
nadph	Nicotinamide adenine dinucleotide phosphate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

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

Thiele, I. and B. O. Palsson (2010). "A protocol for generating a high-quality genome-scale metabolic reconstruction." Nature protocols 5(1): 93-121, Supplementary Methods.



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



Conversion factors^a:

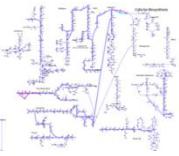
Average cell aqueous volume: $V_c = 6.7 \times 10^{-16} \text{ liter cell}^{-1}$

$V_{\text{DW}} = V_c / M_{\text{DW}} = 2.23 \times 10^{-13} \text{ liter g}_{\text{DW}}^{-1}$

Average dry mass: $M_{\text{DW}} = 3 \times 10^{-13} \text{ g}_{\text{DW}} \text{ cell}^{-1}$

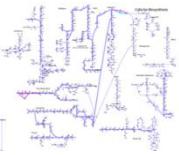
Avogadro's number: $N_A = 6.02 \times 10^{23} \text{ molecules mol}^{-1}$

Thiele, I. and B. O. Palsson (2010). "A protocol for generating a high-quality genome-scale metabolic reconstruction." Nature protocols 5(1): 93-121.



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



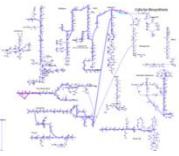
Determine Biomass Composition: Determine the Ion Content

- 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} :

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

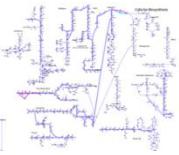
Type	Concentration (CCDB)	Concentration (mM)	Species*	Molar fraction (mM species/mM total)
Number of K ions	$9 * 10^7$ (200-250 mM)	225	k	0.7142
Number of Fe ions	$7 * 10^6$ (18 mM)	9	fe2	0.0286
		9	fe3	0.0286
Number of Mg ions	$4 * 10^6$ (10 mM)	10	mg2	0.0317
Number of Cl ions	$2 * 10^6$ (6 mM)	6	cl	0.0190
Number of Ca ions	$2 * 10^6$ (6 mM)	6	ca2	0.0190
Number of Na ions	$2 * 10^6$ (5 mM)	5	na	0.0159
Number of PO ₄ ions	$2 * 10^6$ (5 mM)	5	pi	0.0159
Number of Cu ions	$1.7 * 10^6$ (4 mM)	4	cu2	0.0127
Number of Mn ions	$1.7 * 10^6$ (4 mM)	4	mn2	0.0127
Number of Mo ions	$1.7 * 10^6$ (4 mM)	4	mobd	0.0127
Number of Zn ions	$1.7 * 10^6$ (4 mM)	4	zn2	0.0127
Number of Cobaltions	$1.7 * 10^6$ (4 mM) ^a	4	cobalt2	0.0127
Number of NH ₄ ions	$6 * 10^6$ (15 mM) ^b	15	nh4	0.0476
Number of SO ₄ ions	$2 * 10^6$ (5 mM) ^c	5	so4	0.0159
	Total ion concentration	315		

Thiele, I. and B. O. Palsson (2010). "A protocol for generating a high-quality genome-scale metabolic reconstruction." Nature protocols 5(1): 93-121.



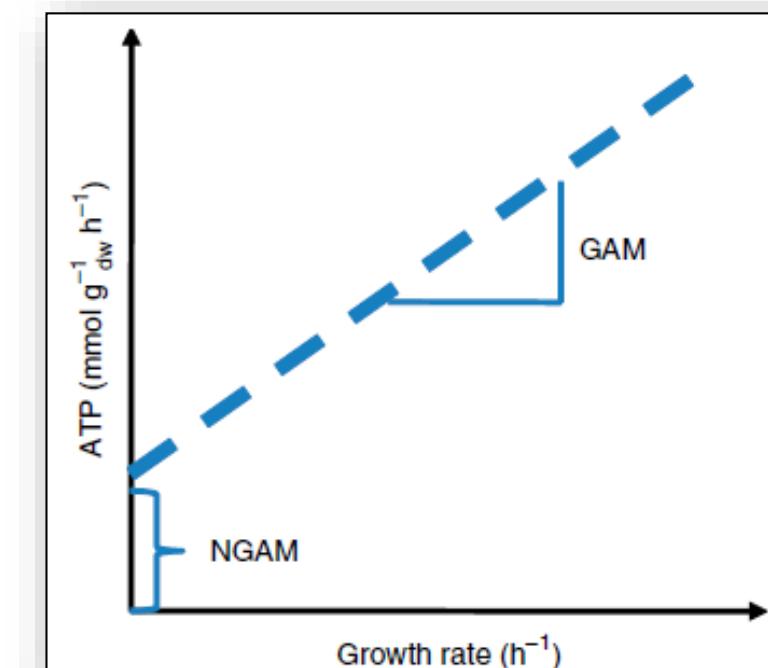
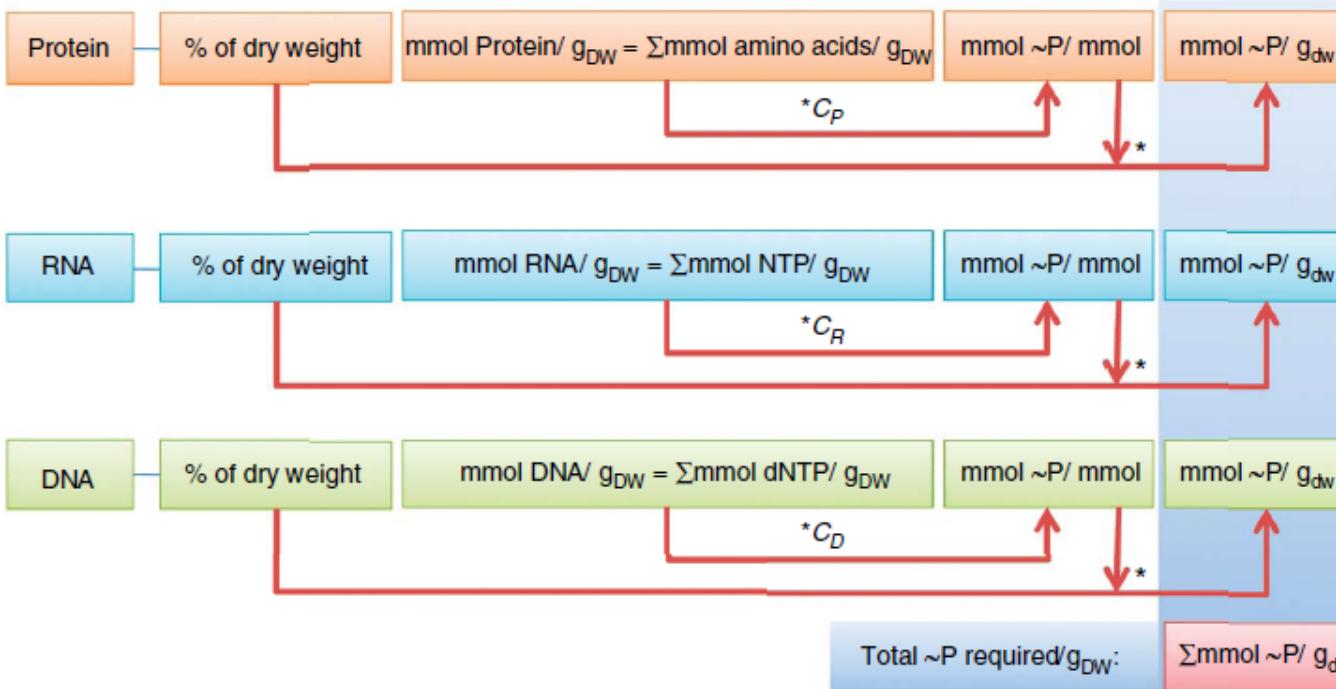
Refinement of Reconstruction: Determine Biomass Composition

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 $\sim\text{P}$) per cellular content of macromolecules:



b

	wt %	Total mmol	mmol $\sim\text{P}/\text{mmol}$	Total
Protein	0.563	5.197	$C_P = 4.324$	22.472
DNA	0.031	0.101	$C_D = 1.365$	0.138
RNA	0.21	0.649	$C_R = 0.406$	0.264
		Total		22.873

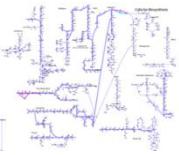
Growth-associated maintenance:
Hydrolysis of 22.873 mmol ATP g_{DW} $^{-1}$
Added to biomass reaction:
 $x \text{ ATP} + x \text{ H}_2\text{O} \rightarrow x \text{ ADP} + x \text{ P}_i + x \text{ H}$,
where x is 22.873

Thiele, I. and B. O. Palsson (2010). "A protocol for generating a high-quality genome-scale metabolic reconstruction." Nature protocols 5(1): 93-121.



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 (H^+), orthophosphate (P_i) and some other metabolites are produced. These metabolites originate mainly from the growth-associated ATP hydrolysis

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

Thiele, I. and B. O. Palsson (2010). "A protocol for generating a high-quality genome-scale metabolic reconstruction." *Nature protocols* 5(1): 93-121.



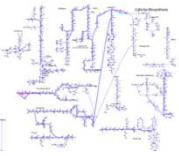
iaf1260 BIOMASS OBJECTIVE FUNCTION (Ec_biomass_iAF1260_core_59p81M)

$Z = 0.000223 \text{ 10fthf}[c] + 0.000223 \text{ 2ohph}[c] + 0.5137 \text{ ala-L}[c] + 0.000223 \text{ amet}[c] + 0.2958 \text{ arg-L}[c] + 0.2411 \text{ asn-L}[c] + 0.2411 \text{ asp-L}[c] + 59.984 \text{ atp}[c] + 0.004737 \text{ ca2}[c] + 0.004737 \text{ cl}[c] + 0.000576 \text{ coa}[c] + 0.003158 \text{ cobalt2}[c] + 0.1335 \text{ ctp}[c] + 0.003158 \text{ cu2}[c] + 0.09158 \text{ cys-L}[c] + 0.02617 \text{ datp}[c] + 0.02702 \text{ dctp}[c] + 0.02702 \text{ dgtp}[c] + 0.02617 \text{ dtpp}[c] + 0.000223 \text{ fad}[c] + 0.007106 \text{ fe2}[c] + 0.007106 \text{ fe3}[c] + 0.2632 \text{ gln-L}[c] + 0.2632 \text{ glu-L}[c] + 0.6126 \text{ gly}[c] + 0.2151 \text{ gtp}[c] + 54.462 \text{ h2o}[c] + 0.09474 \text{ his-L}[c] + 0.2905 \text{ ile-L}[c] + 0.1776 \text{ k}[c] + 0.01945 \text{ kdo2lipid4}[e] + 0.4505 \text{ leu-L}[c] + 0.3432 \text{ lys-L}[c] + 0.1537 \text{ met-L}[c] + 0.007895 \text{ mg2}[c] + 0.000223 \text{ mlthf}[c] + 0.003158 \text{ mn2}[c] + 0.003158 \text{ mobd}[c] + 0.01389 \text{ murein5px4p}[p] + 0.001831 \text{ nad}[c] + 0.000447 \text{ nadp}[c] + 0.011843 \text{ nh4}[c] + 0.02233 \text{ pe160}[c] + 0.04148 \text{ pe160}[p] + 0.02632 \text{ pe161}[c] + 0.04889 \text{ pe161}[p] + 0.1759 \text{ phe-L}[c] + 0.000223 \text{ pheme}[c] + 0.2211 \text{ pro-L}[c] + 0.000223 \text{ pydx5p}[c] + 0.000223 \text{ ribflv}[c] + 0.2158 \text{ ser-L}[c] + 0.000223 \text{ sheme}[c] + 0.003948 \text{ so4}[c] + 0.000223 \text{ thf}[c] + 0.000223 \text{ thmpp}[c] + 0.2537 \text{ thr-L}[c] + 0.05684 \text{ trp-L}[c] + 0.1379 \text{ tyr-L}[c] + 5.5e-005 \text{ udcpdp}[c] + 0.1441 \text{ utp}[c] + 0.4232 \text{ val-L}[c] + 0.003158 \text{ zn2}[c] \rightarrow 59.81 \text{ adp}[c] + 59.81 \text{ h}[c] + 59.806 \text{ pi}[c] + 0.7739 \text{ ppi}[c]$



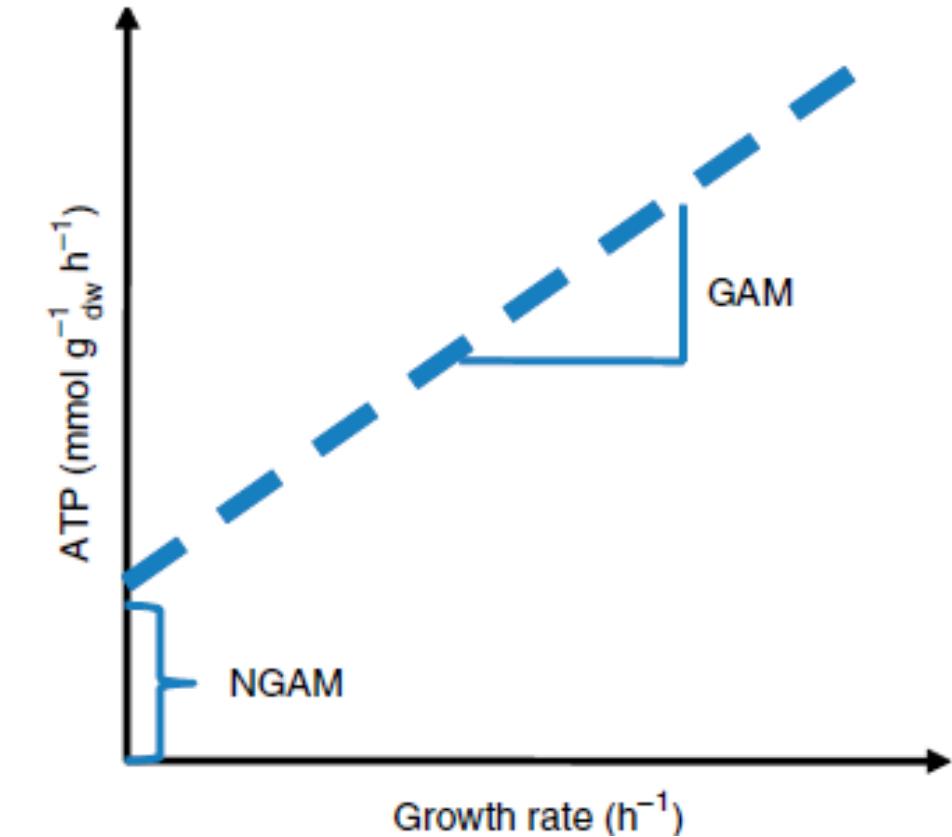
Refinement of Reconstruction

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

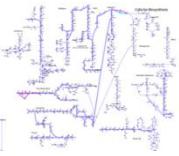


Refinement of Reconstruction: Add Non-GAM (NGAM) Reactions

- Add the following reaction to the reconstruction reaction list:
 - ✓ ATPM: $1 \text{ ATP} + 1 \text{ H}_2\text{O} \rightarrow 1 \text{ ADP} + 1 \text{ P}_i + 1 \text{ H}^+$.
 - ✓ 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 $8.39 \text{ mmol gDW}^{-1} \text{ h}^{-1}$ 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.



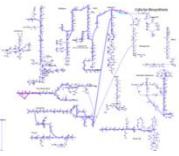
Thiele, I. and B. O. Palsson (2010). "A protocol for generating a high-quality genome-scale metabolic reconstruction." Nature protocols 5(1): 93-121.



Refinement of Reconstruction

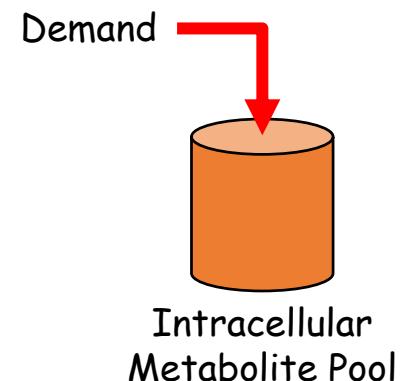
6. Determine and verify substrate and cofactor usage.
7. Obtain a neutral formula for each metabolite in the reaction
8. Determine the charged formula for each metabolite in the reaction.
9. Calculate reaction stoichiometry
10. Determine reaction directionality
11. Add information for gene and reaction 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



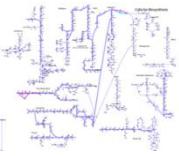


Refinement of Reconstruction: Add Demand Reactions

- 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.

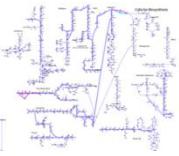


Thiele, I. and B. O. Palsson (2010). "A protocol for generating a high-quality genome-scale metabolic reconstruction." *Nature protocols* 5(1): 93-121.



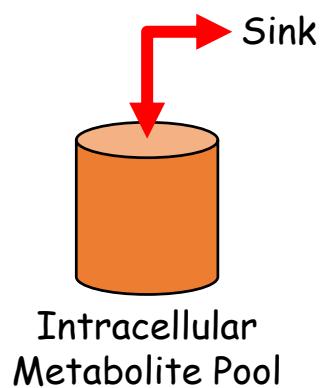
Refinement of Reconstruction

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

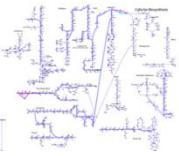


Refinement of Reconstruction: Add Sink Reactions

- 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.



Thiele, I. and B. O. Palsson (2010). "A protocol for generating a high-quality genome-scale metabolic reconstruction." *Nature protocols* 5(1): 93-121.



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

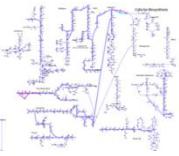




Refinement of Reconstruction: Determine Growth Medium Requirements

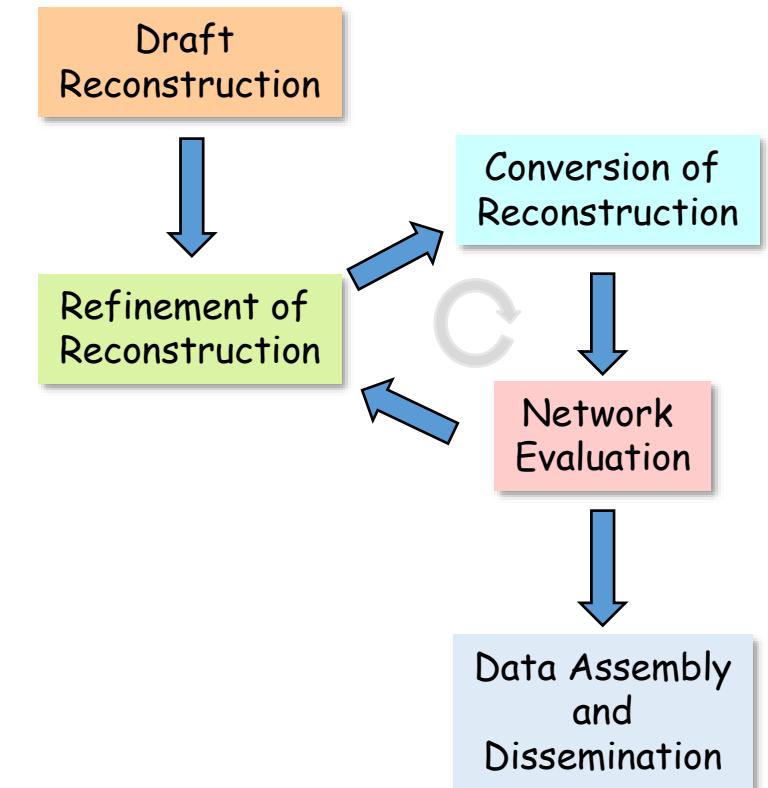
- Information about growth-enabling media should be collected before the conversion and debugging stage. The following information should be collected:
 1. Which metabolites are present?
 2. Are there any auxotrophies?
 3. The definition of a base medium composition, e.g., water, protons, ions and so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

Thiele, I. and B. O. Palsson (2010). "A protocol for generating a high-quality genome-scale metabolic reconstruction." *Nature protocols* 5(1): 93-121.

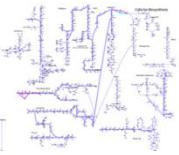


GENOME-SCALE METABOLIC RECONSTRUCTIONS

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

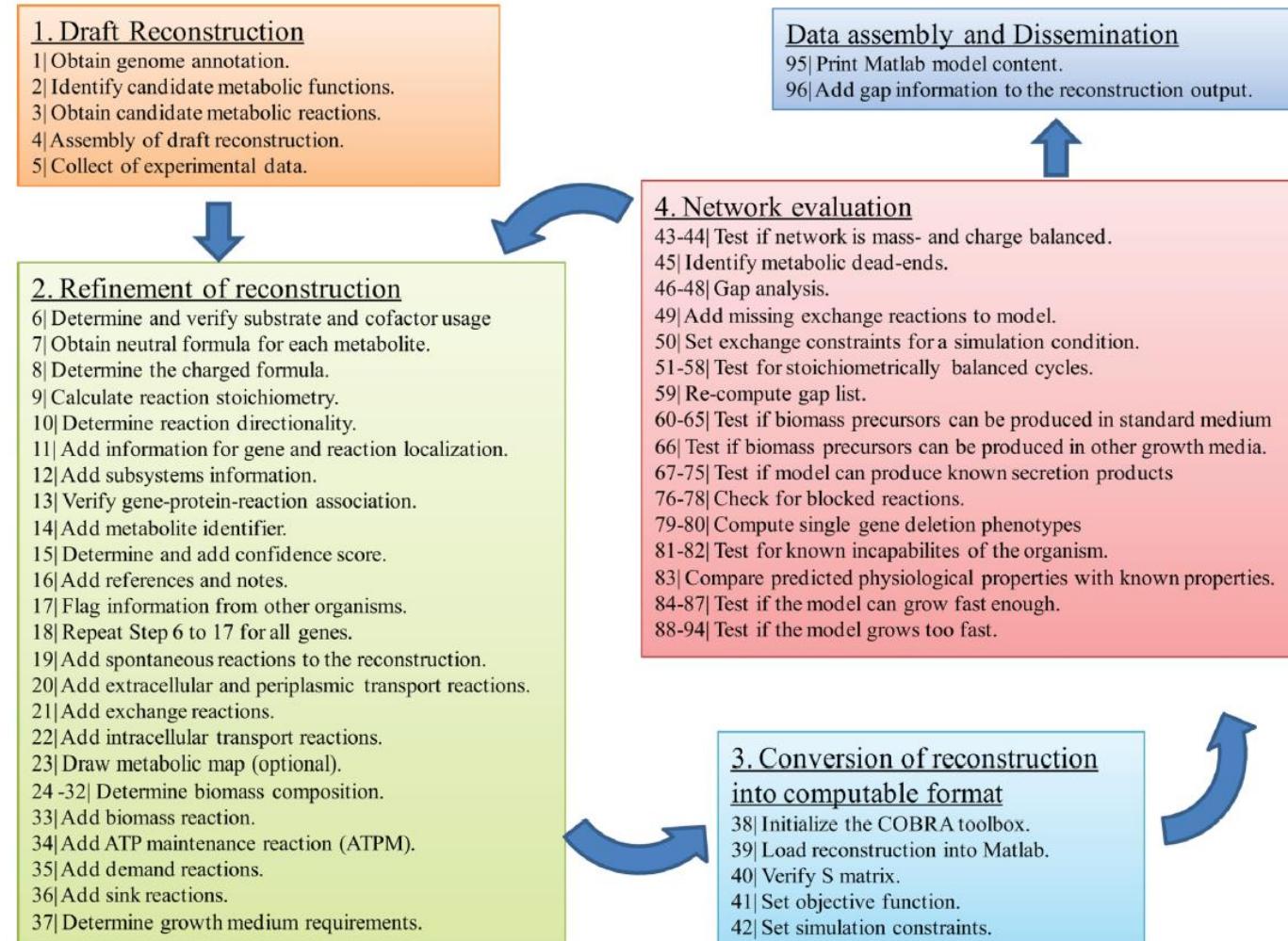


Thiele, I. and B. O. Palsson (2010). "A protocol for generating a high-quality genome-scale metabolic reconstruction." *Nature protocols* 5(1): 93-121.



Reconstruction Process: 96 Step Protocol

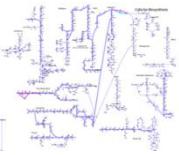
Thiele, I. and B. O. Palsson (2010). "A protocol for generating a high-quality genome-scale metabolic reconstruction." *Nature protocols* 5(1): 93-121.





Stage 3:
**Conversion from Reconstruction
to Mathematical Model**

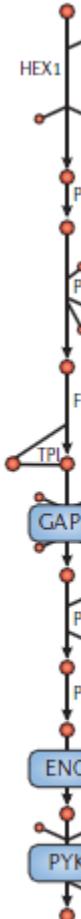
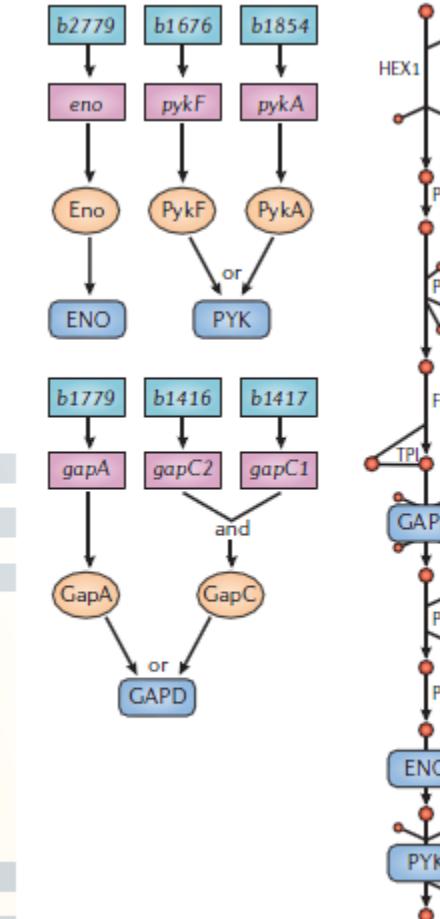
38. Initialize the COBRA toolbox
39. Load reconstruction in Matlab
40. Verify S matrix
41. Set objective function
42. Set simulation constraints



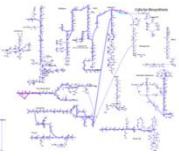
Assembly and Representation

Abbreviation	Glycolytic reactions	Genes
HEX1	[c]GLC + ATP → G6P + ADP + H	glk
PGI	[c]G6P ↔ F6P	pgi
PFK	[c]ATP + F6P → ADP + FDP + H	pfkA, pfkB
FBA	[c]FDP ↔ DHAP + G3P	fbaA, fbaB
TPI	[c]DHAP ↔ G3P	tpiA
GAPD	[c]G3P + NAD + PI ↔ 13DPG + H + NADH	gapA, gapC1, gapC2
PGK	[c]13DPG + ADP ↔ 3PG + ATP	pgk
PGM	[c]3PG ↔ 2PG	gpmA, gpmB
ENO	[c]2PG ↔ H ₂ O + PEP	eno
PYK	[c]ADP + H + PEP → ATP + PYR	pykA, pykF

	HEX1	PGI	PFK	FBA	TPI	GAPD	PGK	PGM	ENO	PYK
ATP	-1	0	-1	0	0	0	1	0	0	1
GLC	-1	0	0	0	0	0	0	0	0	0
ADP	1	0	1	0	0	0	-1	0	0	-1
G6P	1	-1	0	0	0	0	0	0	0	0
H	1	0	1	0	0	0	1	0	0	-1
F6P	0	1	-1	0	0	0	0	0	0	0
FDP	0	0	1	-1	0	0	0	0	0	0
DHAP	0	0	0	1	-1	0	0	0	0	0
G3P	0	0	0	1	1	-1	0	0	0	0
NAD	0	0	0	0	0	-1	0	0	0	0
PI	0	0	0	0	0	-1	0	0	0	0
13DPG	0	0	0	0	0	1	-1	0	0	0
NADH	0	0	0	0	0	1	0	0	0	0
3PG	0	0	0	0	0	0	1	-1	0	0
2PG	0	0	0	0	0	0	0	1	-1	0
PEP	0	0	0	0	0	0	0	0	1	-1
H ₂ O	0	0	0	0	0	0	0	0	1	0
PYR	0	0	0	0	0	0	0	0	0	1



Reed, J. L., I. Famili, et al. (2006). "Towards multidimensional genome annotation." *Nature reviews. Genetics* 7(2): 130-141.



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)`

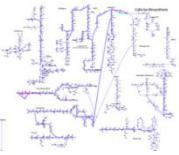
41. Set objective function

- `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:

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

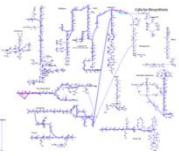
2. The list of reactions for which the bounds should be changed is given by 'rxnNameList', whereas an array contains the new boundary reaction rates ('value'). This type of bound can be set to lower bound ('l') or upper bound ('u'). Alternatively, both bounds can be changed ('b').
3. Use the following command to list all constrained reactions that are greater than a minimal value ('MinInf') and smaller than a maximal value ('MaxInf'):

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

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

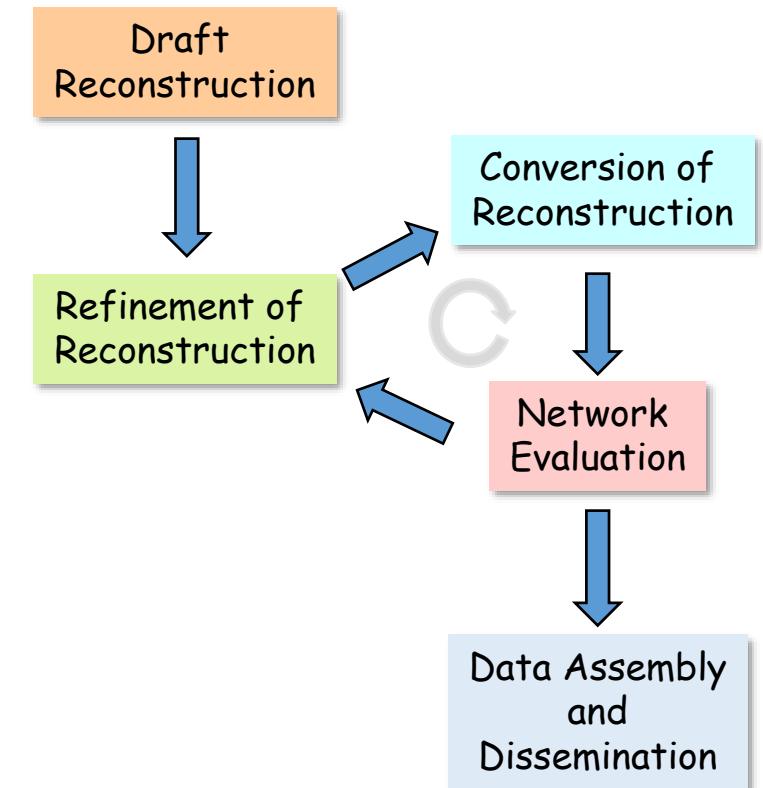
```
printFluxVector (model, fluxData)
```

Thiele, I. and B. O. Palsson (2010). "A protocol for generating a high-quality genome-scale metabolic reconstruction." Nature protocols 5(1): 93-121.

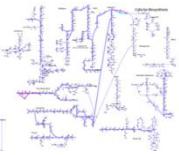


GENOME-SCALE METABOLIC RECONSTRUCTIONS

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

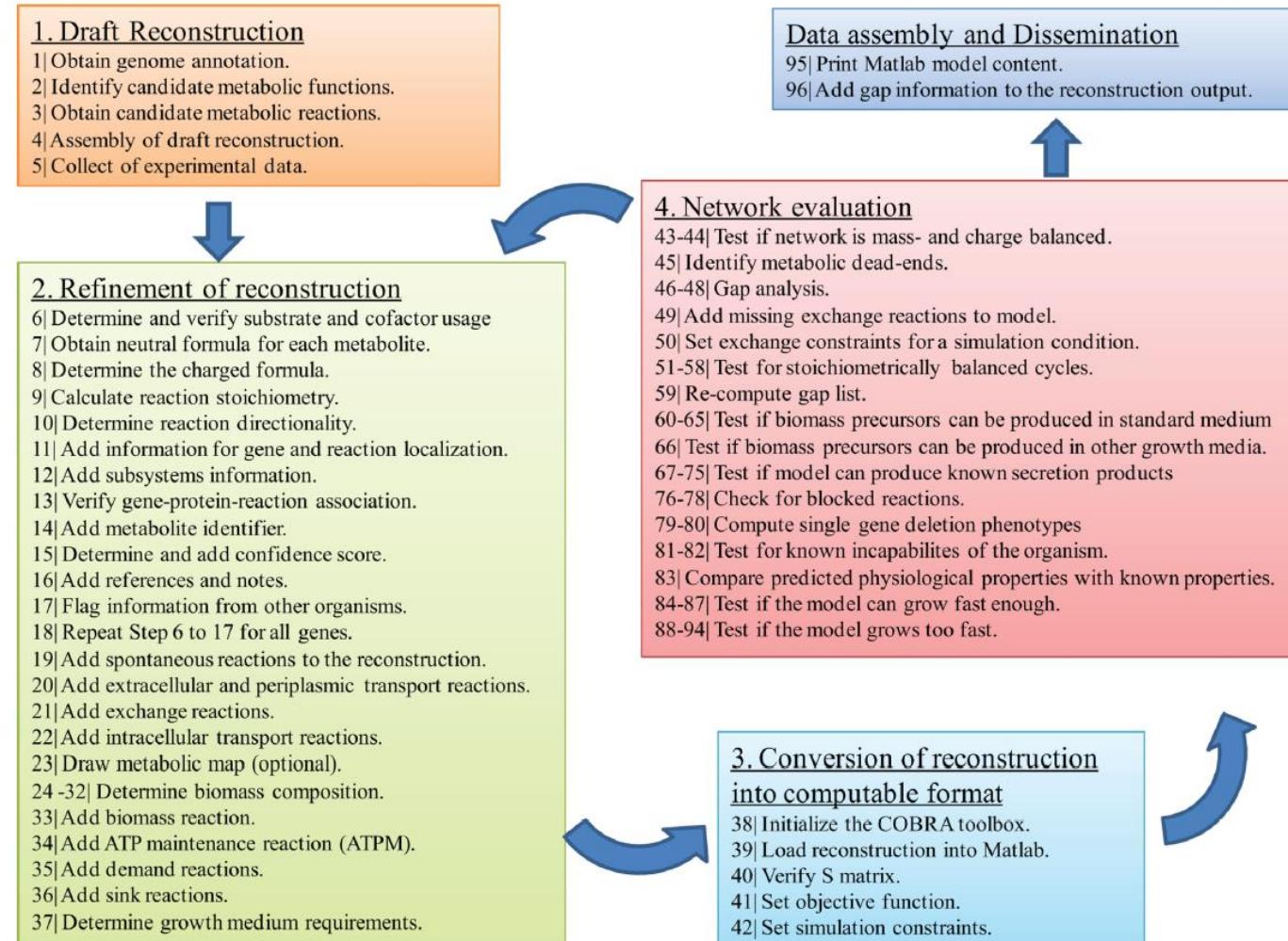


Thiele, I. and B. O. Palsson (2010). "A protocol for generating a high-quality genome-scale metabolic reconstruction." Nature protocols 5(1): 93-121.



Reconstruction Process: 96 Step Protocol

Thiele, I. and B. O. Palsson (2010). "A protocol for generating a high-quality genome-scale metabolic reconstruction." *Nature protocols* 5(1): 93-121.





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.

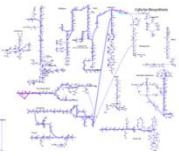
Error Mode	Action
Wrong reaction constraints	Check reaction constraints if they are applied correctly
Missing transport reactions	Add transport reactions
Missing exchange reactions	Add exchange reactions
Cofactor cannot be consumed or produced	Follow Figure 13 (Thiele, 2010)
Shuttling of compounds across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

Thiele, I. and B. O. Palsson (2010). "A protocol for generating a high-quality genome-scale metabolic reconstruction." *Nature protocols* 5(1): 93-121.



Stage 4: Network Evaluation

- 43-44. Test if network is mass-and charge balanced.
45. Identify metabolic dead-ends.
- 46-48. Perform 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 the model can produce known secretion products.
- 76-78. Check for blocked reactions.
- 79-80. Compute single gene deletion phenotypes.
- 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.



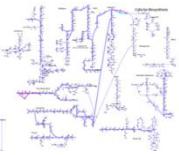
Network Evaluation: Test if Network is Mass-and Charge Balanced

- Check for stoichiometrically unbalanced reactions.
- Use the "CheckMassChargeBalance" function to check for unbalanced reactions.

```
[massImbalance,imBalancedMass,imBalancedCharge,imBalancedBool,Elements]= checkMassChargeBalance(model)
```

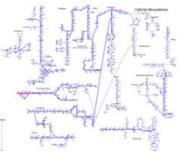
- In case of unbalanced reactions, the function returns a structure containing the name of the unbalanced reaction and which elements are unbalanced ('UnbalancedRxns').

Thiele, I. and B. O. Palsson (2010). "A protocol for generating a high-quality genome-scale metabolic reconstruction." Nature protocols 5(1): 93-121.



checkMassChargeBalance Example

- Change the formula for two reactions in "ecoli_iaf1260_MB.xls"
 - ✓ Arsenate reductase (ASR) - **Add an H₂O**
 - 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
 - From: arg-L[c] + succoa[c] → coa[c] + h[c] + sucarg[c]
 - 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

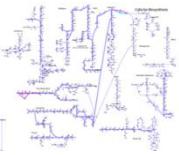
```
% MassChargeBalance_iaf1260_MB.m

clear;

% Input the modified E.coli core model
model = xls2model('ecoli_iaf1260_MB.xls');

% Check mass & charge balance
[massImbalance,imBalancedMass,imBalancedCharge,imBalancedBool,Elements] = checkMassChargeBalance(model)
```

Modified model to include both changed reactions and the changed metabolite



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.

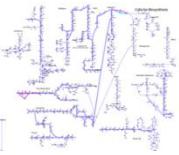
Checked element H
Checking element C
Checking element O
Checking element P
Checking element S
Checking element N
Checking element Mg
Checking element X
Checking element Fe
Checking element Zn
Checking element Co
Checking element R

Element Matrix

H = 1
C = 2
O = 3
P = 4
S = 5
N = 6
Mg = 7
X = 8
Fe = 9
Zn = 10
Co = 11
R = 12

massImbalance =

(371,1)	2	2 Extra Protons in Reaction 371
(372,1)	1	
(144,3)	1	
(167,3)	-1	
(187,3)	1	1 Extra Protons in Reaction 372
(195,3)	-1	
(198,3)	1	
(199,3)	1	
(227,3)	1	
(276,3)	1	
(277,3)	1	
(371,3)	1	1 Extra Oxygen in Reaction 371
(386,3)	1	
(429,3)	1	
(507,3)	1	
(1409,3)	1	
(1708,3)	1	
(2011,3)	1	
(2324,3)	1	



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 $\text{aso4}[c] + 2.000000 \text{gthrd}[c] \rightarrow 2.000000 \text{h2o}[c] + \text{aso3}[c] + \text{gthox}[c]$ (b3503) and (b1064)

ans =

$'\text{aso4}[c] + 2 \text{gthrd}[c] \rightarrow 2 \text{h2o}[c] + \text{aso3}[c] + \text{gthox}[c]'$

Reaction Index
Reaction Name
2 Extra protons plus 1 extra oxygen implies an extra H_2O
Reaction Formula

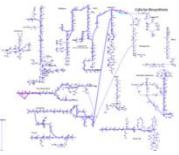
```
>> printRxnFormula(model, model.rxns(372))
```

AST $\text{succoa}[c] + \text{arg-L}[c] \rightarrow 2.000000 \text{h}[c] + \text{coa}[c] + \text{sucarg}[c]$ (b1747)

ans =

$'\text{succoa}[c] + \text{arg-L}[c] \rightarrow 2 \text{h}[c] + \text{coa}[c] + \text{sucarg}[c]'$

1 Extra proton



MassChargeBalance_iaf1260_MB.m Example:

Remaining Matrix Formulas

printRxnFormula(model,model.rxn(Reaction Index))

Reaction Index	Reaction	Reaction Formula
144	ACACCT	acac[c] + accoa[c] → aacoa[c] + ac[c]
167	ACKr	atp[c] + ac[c] ⇌ adp[c] + actp[c]
187	ACODA	h2o[c] + acorn[c] → ac[c] + orn[c]
195	ACS	atp[c] + ac[c] + coa[c] → amp[c] + ppi[c] + accoa[c]
198	ACT2rpp	h[p] + ac[p] ⇌ h[c] + ac[c]
199	ACT4pp	ac[p] + na1[p] → ac[c] + na1[c]
227	AGDC	h2o[c] + acgam6p[c] → ac[c] + gam6p[c]
276	ALDD2x	h2o[c] + nad[c] + acald[c] → 2 h[c] + nadh[c] + ac[c]
277	ALDD2y	h2o[c] + nadp[c] + acald[c] → 2 h[c] + nadph[c] + ac[c]
386	BUTCT	accoa[c] + but[c] → ac[c] + btcoa[c]
429	CITL	cit[c] → ac[c] + oaa[c]
507	CYSS	acser[c] + h2s[c] → h[c] + ac[c] + cys-L[c]
1409	HXCT	accoa[c] + hxo[c] → ac[c] + hxcoa[c]
1708	NACODA	h2o[c] + acg5sa[c] → ac[c] + glu5sa[c]
2011	POX	h2o[c] + pyr[c] + q8[c] → co2[c] + ac[c] + q8h2[c]
2324	UHGADA	h2o[c] + u3aga[c] → ac[c] + u3hga[c]

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



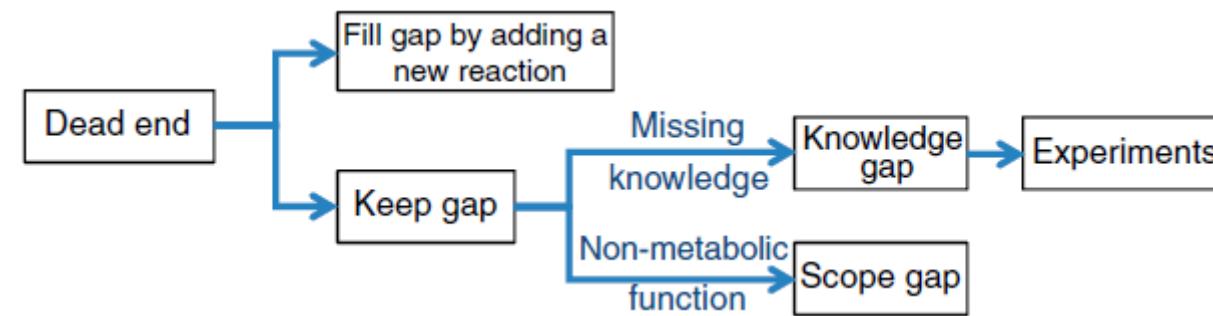
Stage 4: Network Evaluation

- 43-44. Test if network is mass-and charge balanced.
-  45. Identify metabolic dead-ends.
- 46-48. Perform 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 the model can produce known secretion products.
- 76-78. Check for blocked reactions.
- 79-80. Compute single gene deletion phenotypes.
- 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.

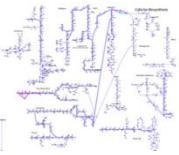


Network Evaluation: Identify Metabolic Dead-ends

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.

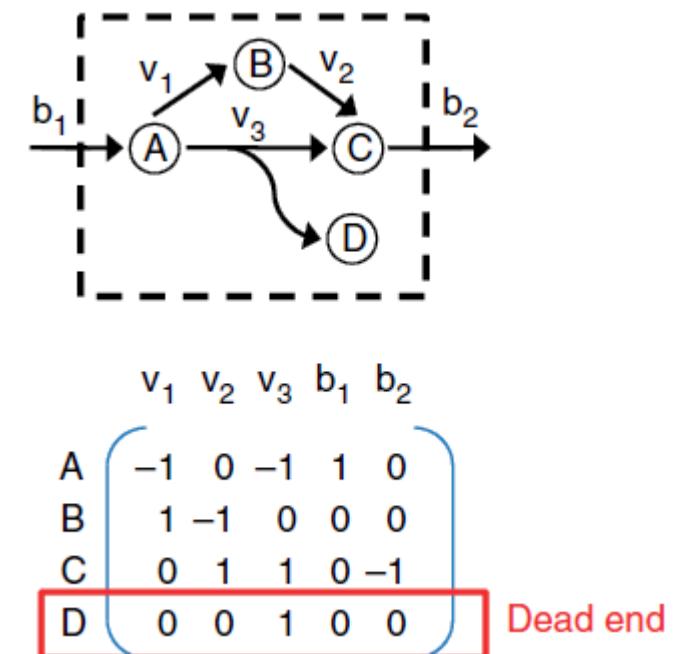


Thiele, I. and B. O. Palsson (2010). "A protocol for generating a high-quality genome-scale metabolic reconstruction." *Nature protocols* 5(1): 93-121.

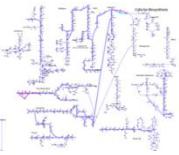


Identifying Gaps: Connectivity-based Approach

- 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.

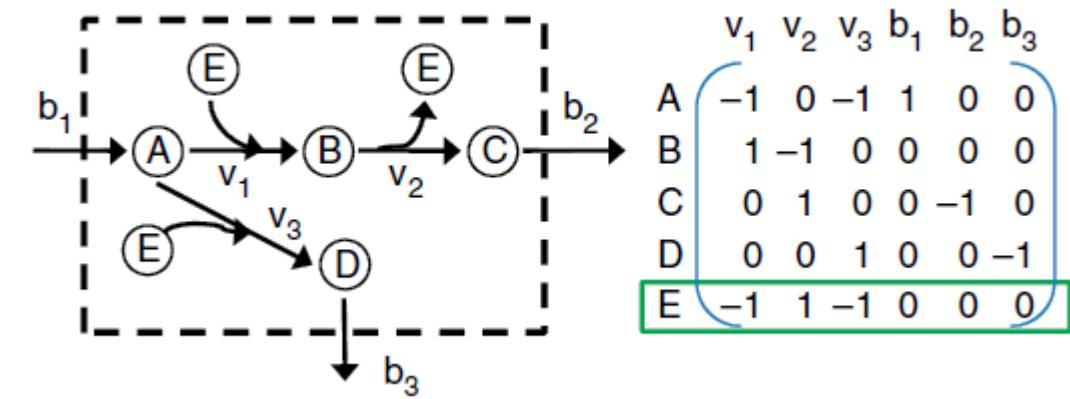


Thiele, I. and B. O. Palsson (2010). "A protocol for generating a high-quality genome-scale metabolic reconstruction." Nature protocols 5(1): 93-121.



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.



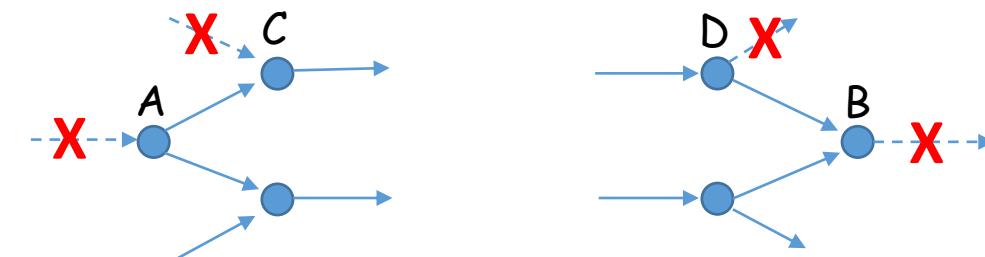
- Reactions v_3 and b_3 cannot carry any flux in this network as the metabolite 'E' is unbalanced.
- These reactions are also called 'blocked reactions'.
- Topological analysis would not have identified 'E' as a dead-end metabolite, as reaction v_3 is producing the metabolite.
- Flux variability analysis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

Thiele, I. and B. O. Palsson (2010). "A protocol for generating a high-quality genome-scale metabolic reconstruction." Nature protocols 5(1): 93-121.

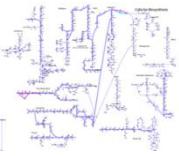


Network Evaluation: Gap Types

- 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.



Satish Kumar, V., M. S. Dasika, et al. (2007). "Optimization based automated curation of metabolic reconstructions." BMC Bioinformatics 8: 212.



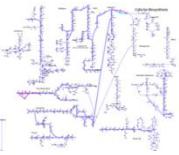
Network Evaluation: Identify Metabolic Dead-ends: gapFind

- Use "gapFind" to identify the gaps

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

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

Satish Kumar, V., M. S. Dasika, et al. (2007). "Optimization based automated curation of metabolic reconstructions." BMC Bioinformatics 8: 212.



Network Evaluation: gapFind Example

```
% GapFindExample.m
clear;

% 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
[invovledRxns,involvedMets,deadEnds]= draw_by_met (model,{ 'fru[e]' },...
    true,1,'struc',{ '' },FBAsolution.x);
[invovledRxns,involvedMets,deadEnds]= draw_by_met (model,{ 'fum[e]' },...
    true,1,'struc',{ '' },FBAsolution.x);
[invovledRxns,involvedMets,deadEnds]= draw_by_met (model,{ 'gln-L[e]' },...
    true,1,'struc',{ '' },FBAsolution.x);
[invovledRxns,involvedMets,deadEnds]= draw_by_met (model,{ 'mal-L[e]' }, ...
    true,1,'struc',{ '' },FBAsolution.x);
```

```
>> GapFindExample
```

```
allGaps =
```

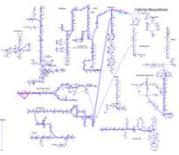
```
'fru[e]'
'fum[e]'
'gln-L[e]'
'mal-L[e]'
```

```
rootGaps =
```

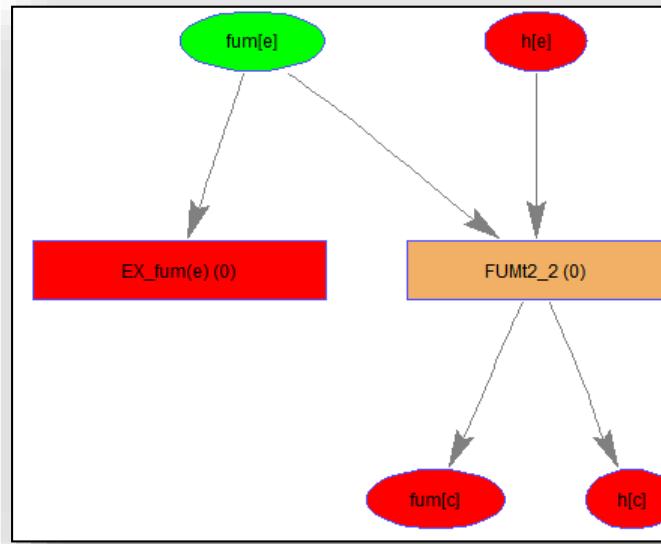
Empty cell array: 0-by-1

```
downstreamGaps =
```

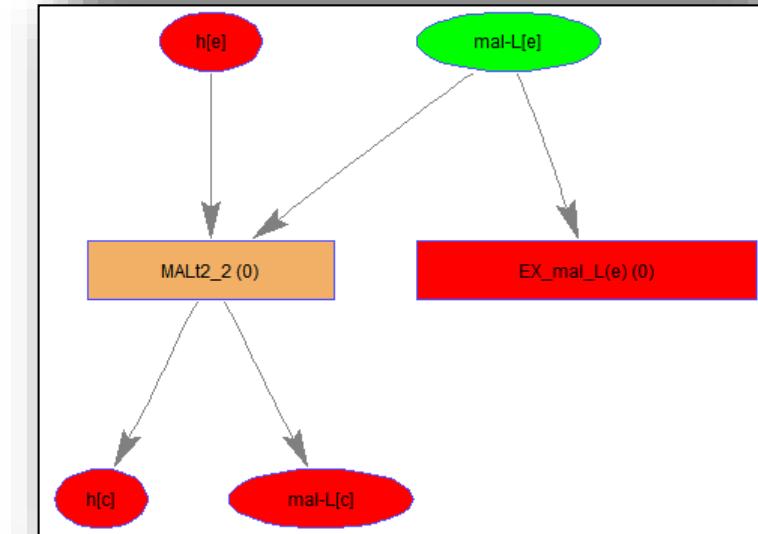
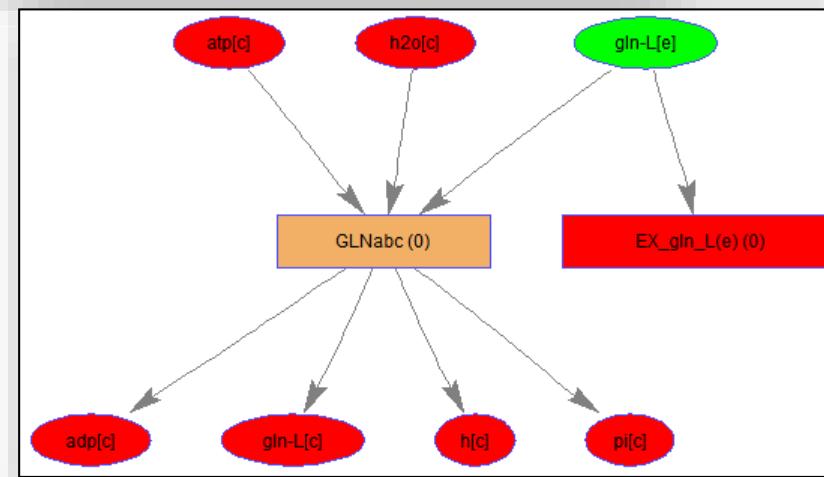
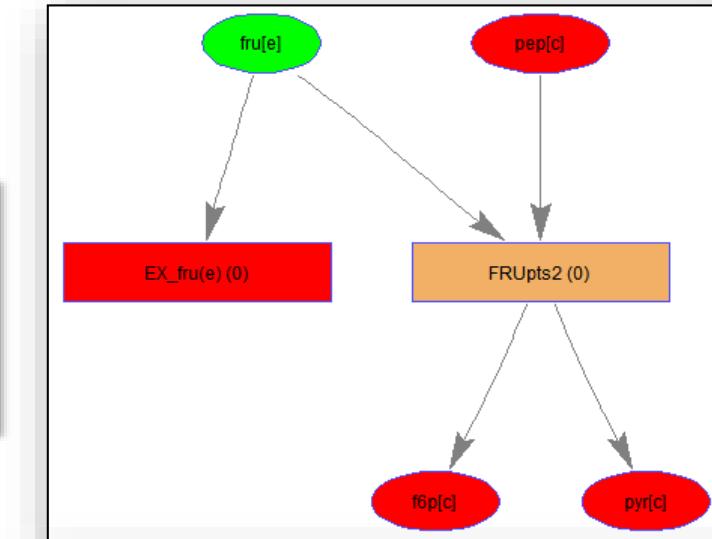
```
'fru[e]'
'fum[e]'
'gln-L[e]'
'mal-L[e]'
```

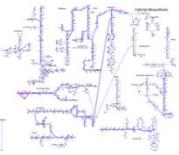


“gapFind” Example Metabolite Connectivity



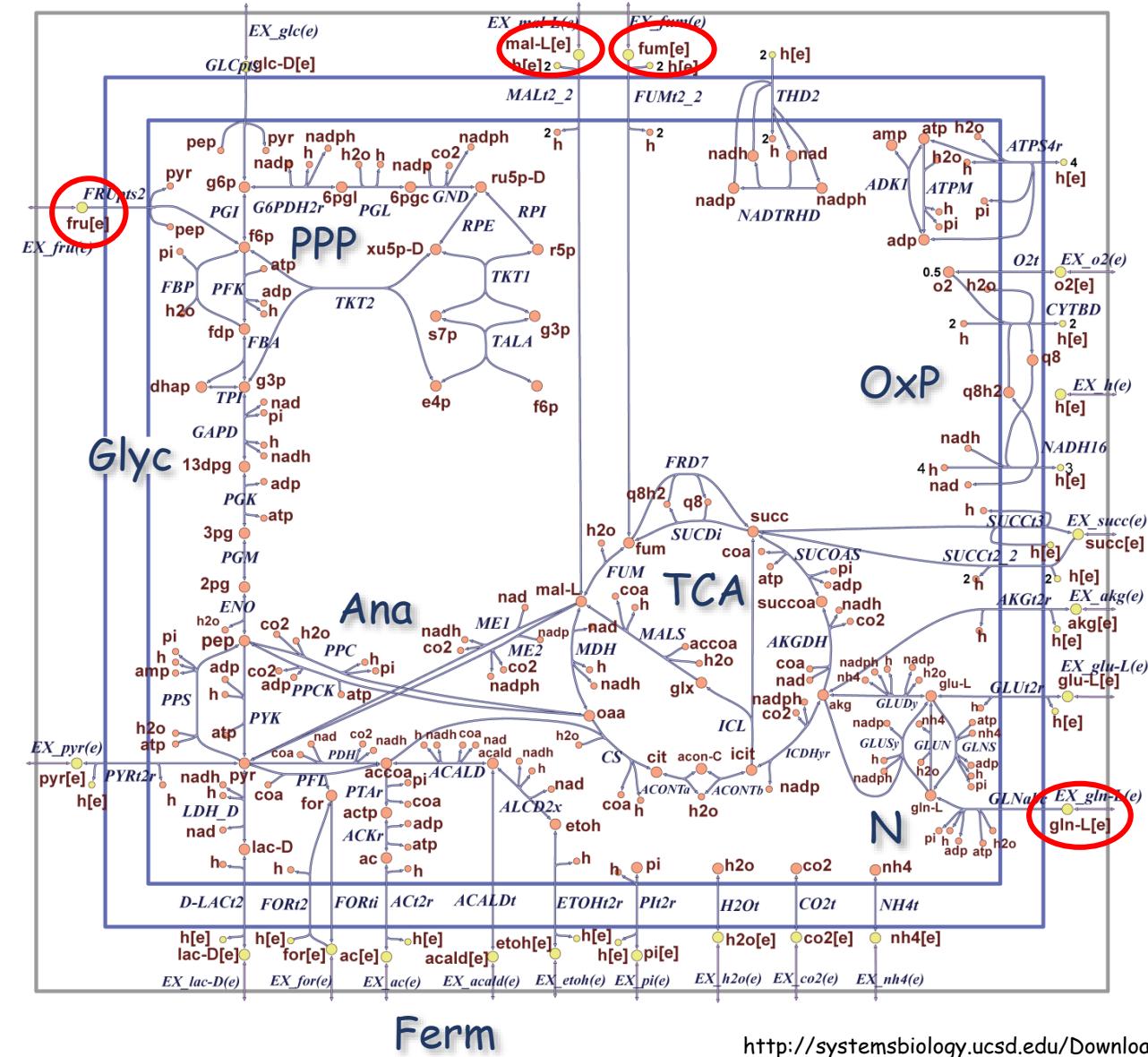
Note that there are no inputs to any of the green metabolites since they cannot be secreted.
Secretion is a downstream process, thus a downstream gap

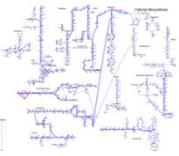




E.coli Core Model Downstream Gaps

Orth, J. D., I. Thiele, et al. (2010). "What is flux balance analysis?" *Nature biotechnology* 28(3): 245-248.





Stage 4: Network Evaluation

- 43-44. Test if network is mass-and charge balanced.
45. Identify metabolic dead-ends.
- 46-48.** Perform 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 the model can produce known secretion products.
- 76-78. Check for blocked reactions.
- 79-80. Compute single gene deletion phenotypes.
- 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.

Thiele, I. and B. O. Palsson (2010). "A protocol for generating a high-quality genome-scale metabolic reconstruction." *Nature protocols* 5(1): 93-121.



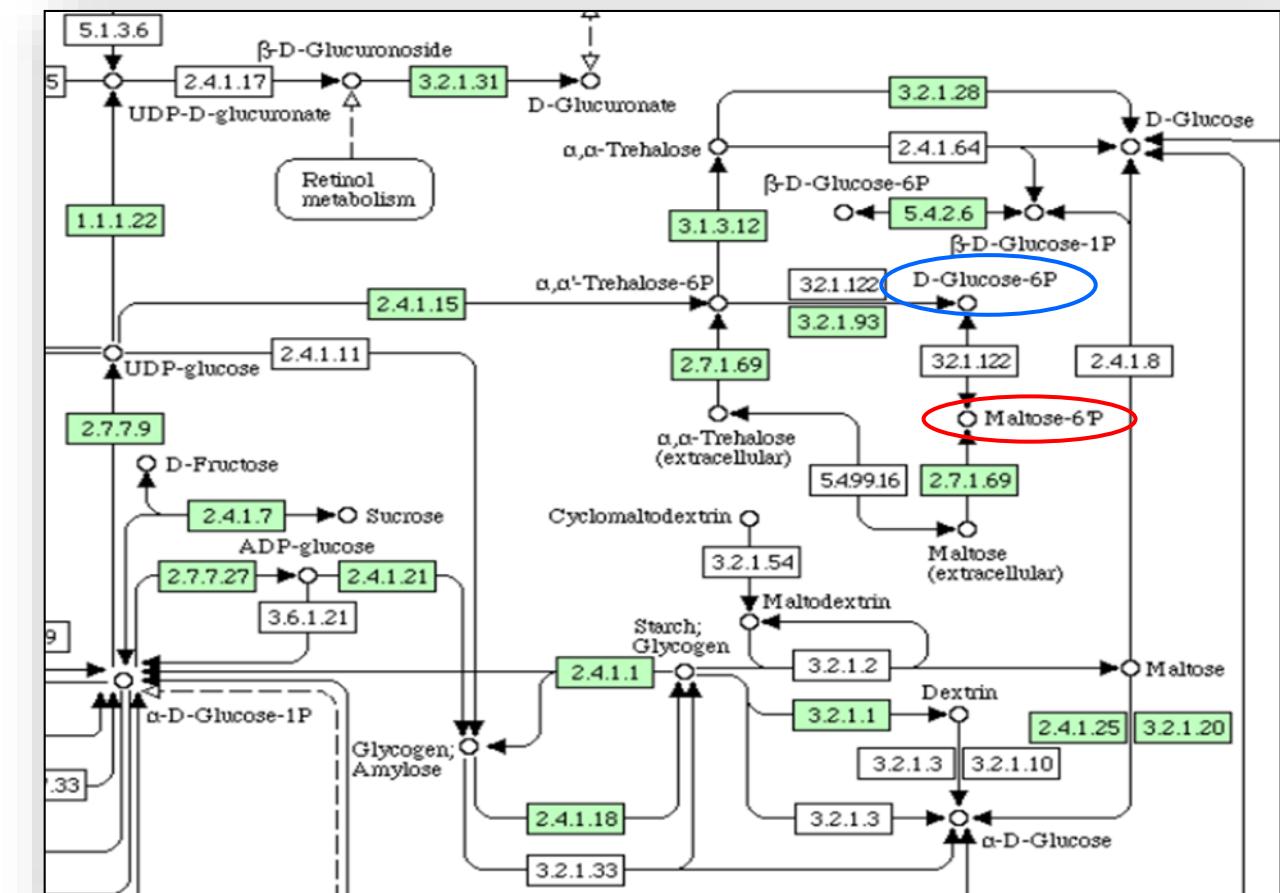
Using KEGG Pathways

<http://www.genome.jp/kegg/pathway.html>

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.



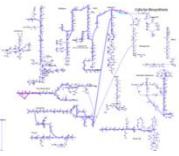
Thiele, I. and B. O. Palsson (2010). "A protocol for generating a high-quality genome-scale metabolic reconstruction." Nature protocols 5(1): 93-121, Supplementary Methods.



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.
- 46-48. Perform 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 the model can produce known secretion products.
- 76-78. Check for blocked reactions.
- 79-80. Compute single gene deletion phenotypes.
- 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

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

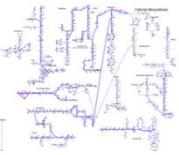
to set the constraints. Reactions whose bounds should be changed are listed in 'rxnNameList'. The new value for each reaction is contained in the array 'value'. Finally, the type of constraint has to be defined in the list 'boundType'. The possible types are: 'l' for lower bound, 'u' for upper bound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

Thiele, I. and B. O. Palsson (2010). "A protocol for generating a high-quality genome-scale metabolic reconstruction." Nature protocols 5(1): 93-121, Supplementary Methods.

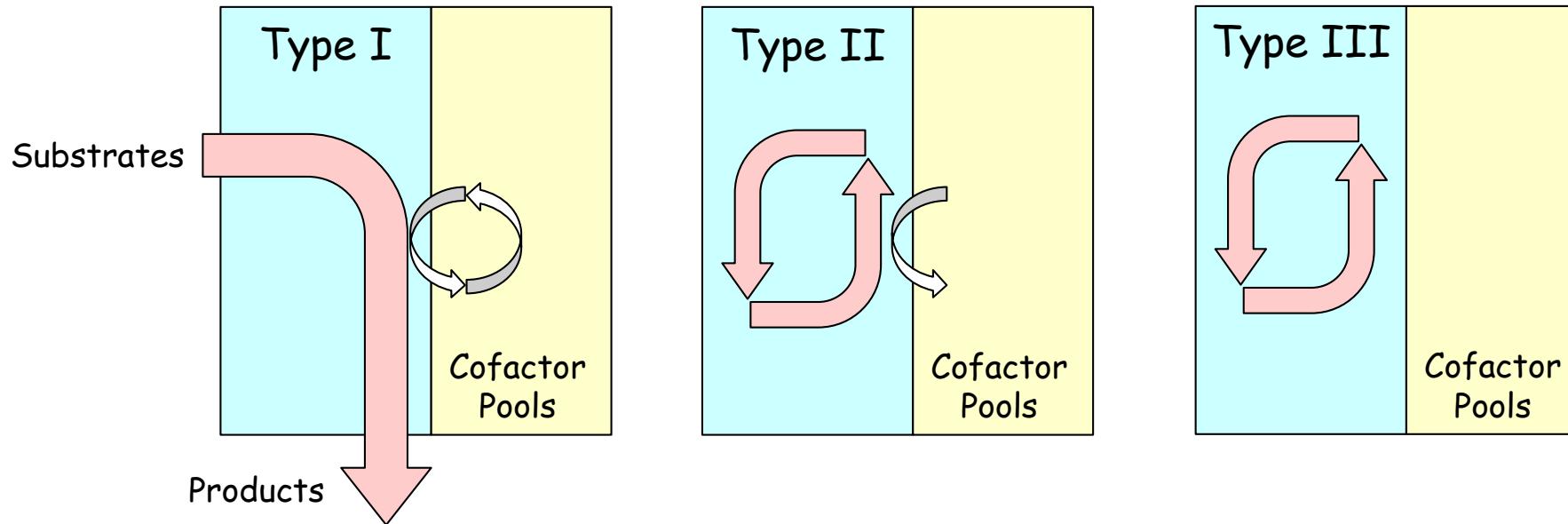


Stage 4: Network Evaluation

- 43-44. Test if network is mass-and charge balanced.
45. Identify metabolic dead-ends.
- 46-48. Perform 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 the model can produce known secretion products.
- 76-78. Check for blocked reactions.
- 79-80. Compute single gene deletion phenotypes.
- 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.

Price, N. D., I. Famili, et al. (2002). "Extreme pathways and Kirchhoff's second law." Biophysical journal 83(5): 2879-2882.



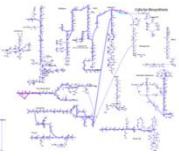
Type III Extreme Pathway (Loops) Removal During Simulations

- Jan Schellenberger wrote a function that removes thermodynamically infeasible loops from models:
 - ✓ Schellenberger, J., N. E. Lewis, et al. (2011). "Elimination of thermodynamically infeasible loops in steady-state metabolic models." *Biophysical journal* 100(3): 544-553.
- 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.
-  46-48. Perform 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 the model can produce known secretion products.
- 76-78. Check for blocked reactions.
- 79-80. Compute single gene deletion phenotypes.
- 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.
- 46-48. Perform 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 the model can produce known secretion products.
- 76-78. Check for blocked reactions.
- 79-80. Compute single gene deletion phenotypes.
- 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.

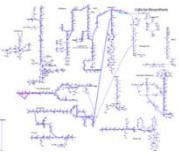


Network Evaluation:

Test if Biomass Precursors can be Produced in Standard Medium

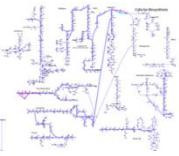
- Test the model's ability to produce each individual biomass component in standard medium condition (e.g., minimal medium M9 supplemented with D-glucose).
 - ✓ Growth on minimal medium M9 was simulated by maximizing flux through a defined biomass objective function and allowing the uptake of the desired carbon source, NH_4^+ , SO_4^{2-} , O_2 , and P_i and the free exchange of H^+ , H_2O , and CO_2 (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.

Thiele, I. and B. O. Palsson (2010). "A protocol for generating a high-quality genome-scale metabolic reconstruction." *Nature protocols* 5(1): 93-121.



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), KH_2PO_4 (3g), NaCl (0.5g), NH_4Cl (1g), MgSO_4 (2 mM), 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
 - ✓ NH_4 , SO_4 , O_2 , and P_i and the free exchange of H^+ , H_2O , and CO_2
- All exchange reaction lower constraints, except the following, should be greater than zero
 - ✓ $-1000 \leq \text{NH}_4, \text{SO}_4, \text{O}_2, \text{and P}_i \leq 0$
 - ✓ $-1000 \leq \text{H}^+, \text{H}_2\text{O}, \text{and CO}_2 \leq 1000$
 - ✓ $-1000 \leq \text{Carbon source} \leq 0$
 - ✓ 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
 - $\text{Na}_2\text{HPO}_4 \cdot 12 \text{H}_2\text{O}$ 85.7 g
 - KH_2PO_4 15.0 g
 - NaCl 2.5 g
 - Dissolve above components in 1000 ml of milli-Q and autoclave
- 5 g $(\text{NH}_4)_2\text{SO}_4$ in 15 ml of H₂O
- Trace elements
 - 1 g EDTA
 - 29 mg $\text{ZnSO}_4 \cdot 7\text{H}_2\text{O}$
 - 198 mg $\text{MnCl}_2 \cdot 4\text{H}_2\text{O}$
 - 254 mg $\text{CoCl}_2 \cdot 6\text{H}_2\text{O}$
 - 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 $\text{CaCl}_2 \cdot 2\text{H}_2\text{O}$: 1.47 g in 100 ml milliQ and filter with 0.22 micron filter
- 1M $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$: 24.65 g in 100 ml milliQ and filter with 0.22 micron filter
- 10 mM $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$: 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 $(\text{NH}_4)_2\text{SO}_4$; 1 ml of $\text{CaCl}_2 \cdot 2\text{H}_2\text{O}$; 1 ml trace elements; 20 ml glucose; 1ml $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$; 1 ml $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$ 2ml thiamine; 1ml antibiotic (standard conc.)



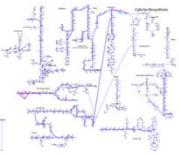
Minimal Nutrients for *E.coli* iaf1260

$\text{EX_glc}(e) = -10, \text{EX_o2}(e) = -1000$

$\text{EX_ca2}(e)$	-0.00440206
$\text{EX_cl}(e)$	-0.00440206
$\text{EX_co2}(e)$	21.9456
$\text{EX_cobalt2}(e)$	-0.0029347
$\text{EX_cu2}(e)$	-0.0029347
$\text{EX_fe2}(e)$	-0.00701801
$\text{EX_fe3}(e)$	-0.00660355
$\text{EX_glc}(e)$	-10
$\text{EX_h2o}(e)$	46.4241
$\text{EX_h}(e)$	8.53495
$\text{EX_k}(e)$	-0.165042
$\text{EX_mg2}(e)$	-0.00733676
$\text{EX_mn2}(e)$	-0.0029347

$\text{EX_mobd}(e)$	-0.0029347
$\text{EX_nh4}(e)$	-10.0215
$\text{EX_o2}(e)$	-19.9695
$\text{EX_pi}(e)$	-0.893343
$\text{EX_so4}(e)$	-0.232555
$\text{EX_zn2}(e)$	-0.0029347
Ec_biomass	0.929292

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.

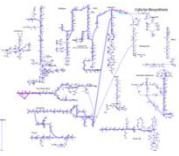


Network Evaluation: Test if Biomass Precursors Can Be Produced in Standard Medium (II)

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.

Thiele, I. and B. O. Palsson (2010). "A protocol for generating a high-quality genome-scale metabolic reconstruction." *Nature protocols* 5(1): 93-121.



"biomassPrecursorCheck" Example

```
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_oaa[c]	oaa[c]	->
DM_pep[c]	pep[c]	->
DM_pyr[c]	pyr[c]	->
DM_r5p[c]	r5p[c]	->

Different name
than in the Cobra
Documentation

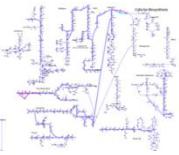
Demand reactions are
created for each element
in the biomass function to
check to see if the
precursors can be
synthesized

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]'
'oaa[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.

precursorCheck.m

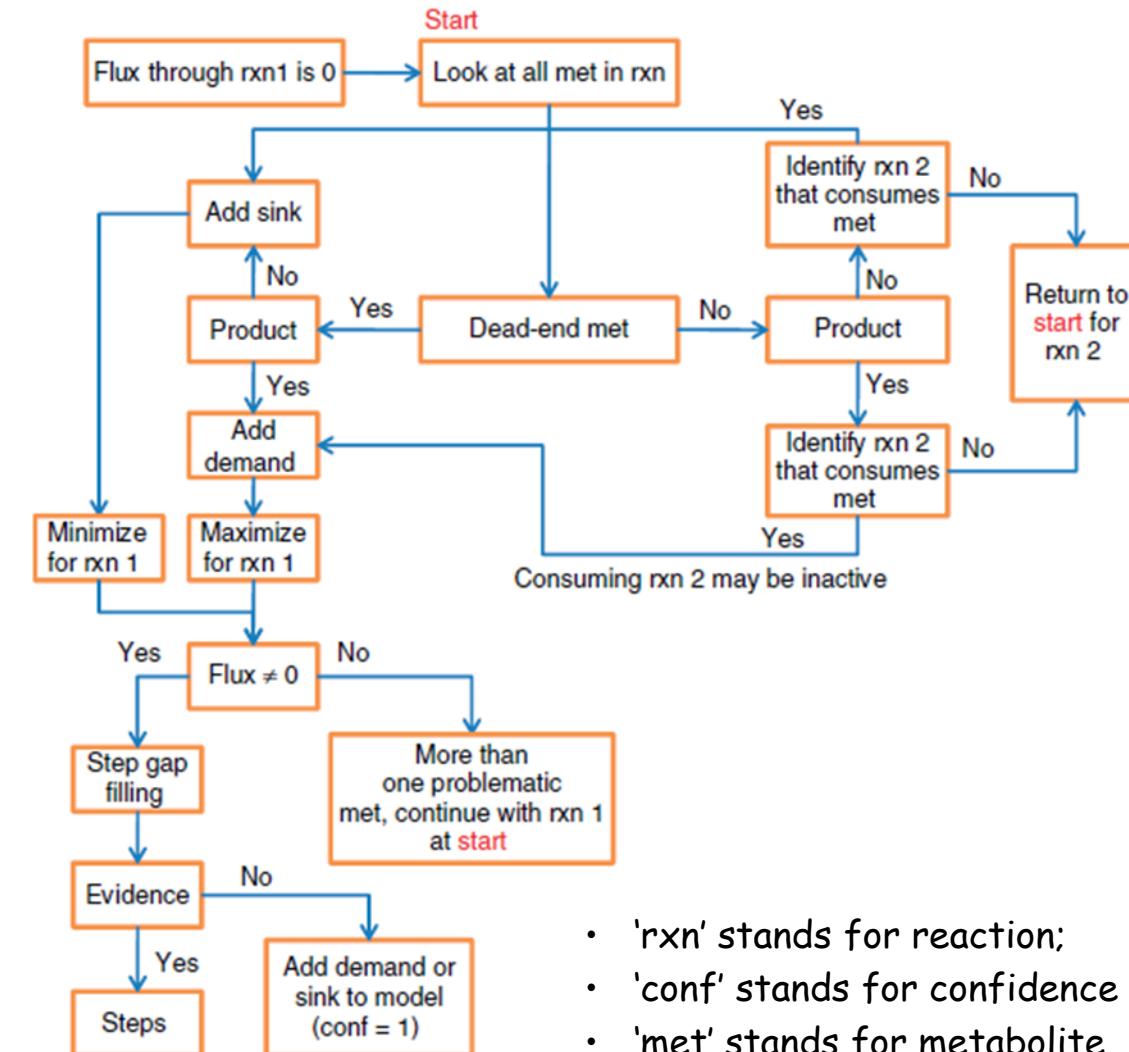


Network Evaluation:

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.



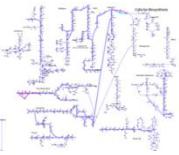
- 'rxn' stands for reaction;
- 'conf' stands for confidence score;
- 'met' stands for metabolite.

Thiele, I. and B. O. Palsson (2010). "A protocol for generating a high-quality genome-scale metabolic reconstruction." Nature protocols 5(1): 93-121.



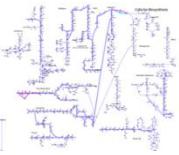
Stage 4: Network Evaluation

- 43-44. Test if network is mass-and charge balanced.
45. Identify metabolic dead-ends.
- 46-48. Perform 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 the model can produce known secretion products.
- 76-78. Check for blocked reactions.
- 79-80. Compute single gene deletion phenotypes.
- 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.



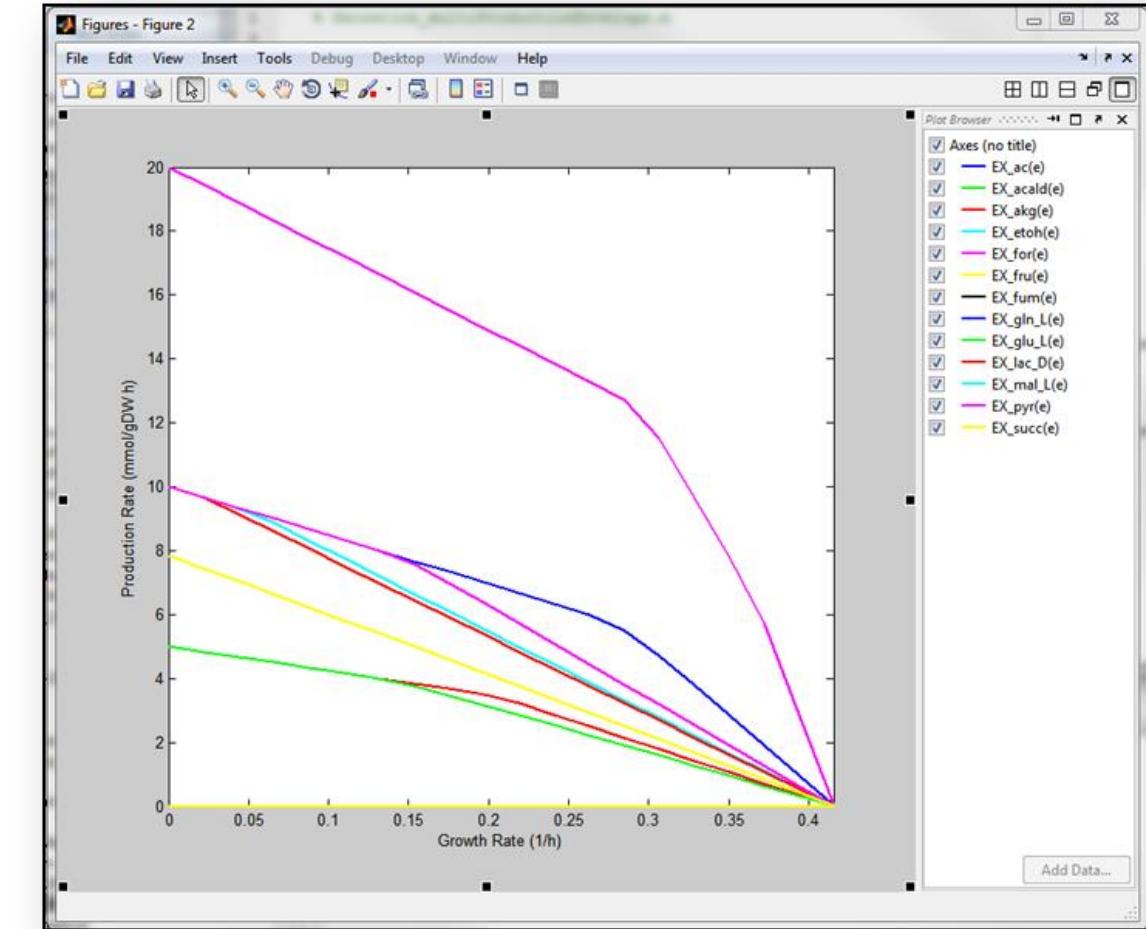
Stage 4: Network Evaluation

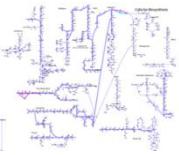
- 43-44. Test if network is mass-and charge balanced.
45. Identify metabolic dead-ends.
- 46-48. Perform gap analysis.
49. Add missing exchange reactions to model.
50. Set exchange constraints for a simulation condition.
- 51-58. Test for stoichiometrically balanced cycles.
59. Re-compute gap list.
- 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 the model can produce known secretion products.
- 76-78. Check for blocked reactions.
- 79-80. Compute single gene deletion phenotypes.
- 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 bioprodut) or "multiProductionEnvelope" (all secreted bioproducts) functions.
- Identify missing secreted bioproducts.





Production Envelope of Secreted Metabolites

Secretion_multiProductionEnvelope.m

```
% Secretion_multiProductionEnvelope.m

clear;

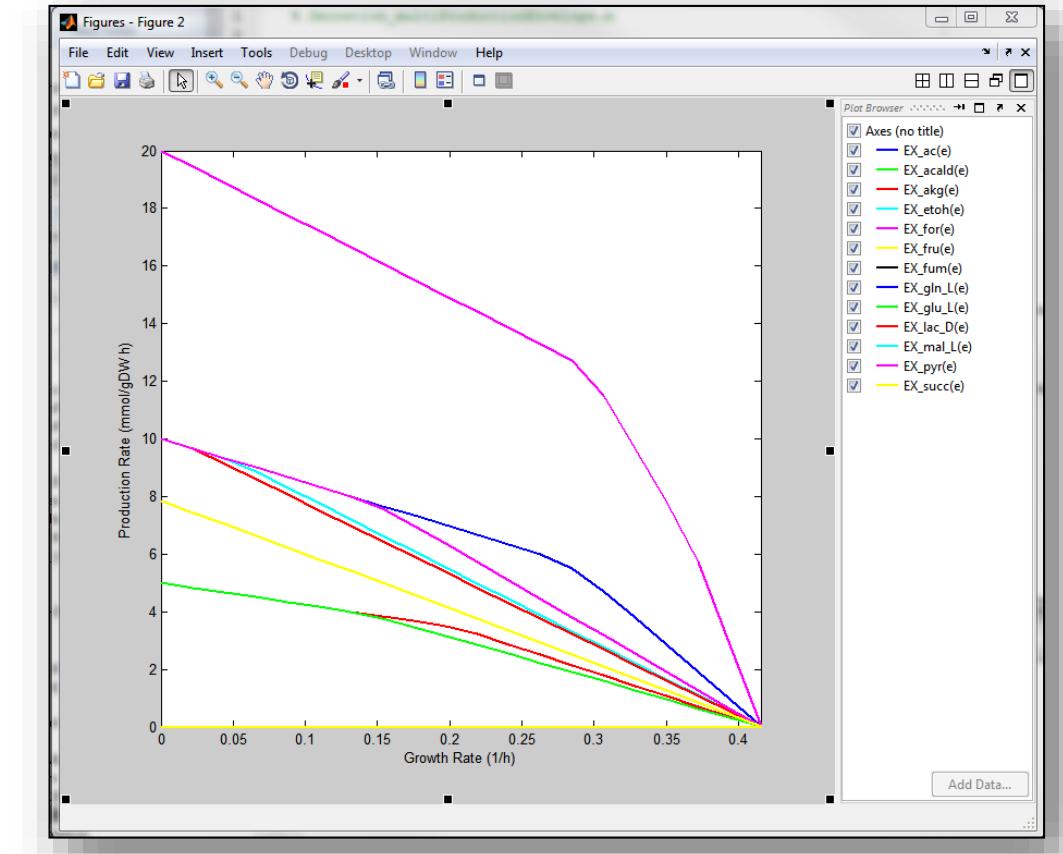
model=readCbModel('ecoli_textbook');

model = changeRxnBounds(model,'EX_glc(e)',-5,'1');
model = changeRxnBounds(model,'EX_o2(e)',-20,'1');

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.
- 46-48. Perform 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 the model can produce known secretion products.
-  76-78. Check for blocked reactions.
- 79-80. Compute single gene deletion phenotypes.
- 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.

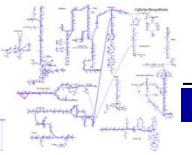


Network Evaluation:

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

Thiele, I. and B. O. Palsson (2010). "A protocol for generating a high-quality genome-scale metabolic reconstruction." *Nature protocols* 5(1): 93-121.



```
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

blockedReactions
```

```
>> findBlockedReactionTest

blockedReactions =

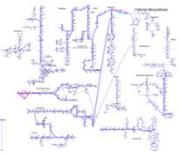
'GLUDy'
```

findBlockedReactionTest.m



Stage 4: Network Evaluation

- 43-44. Test if network is mass-and charge balanced.
45. Identify metabolic dead-ends.
- 46-48. Perform 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 the model can produce known secretion products.
- 76-78. Check for blocked reactions.
-  79-80. Compute single gene deletion phenotypes.
-  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

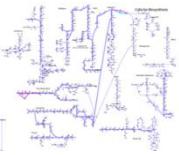
- Analysis of false-positive and false-negative predictions will help to further refine the network content if the information is available.
- Phenotyping data (e.g., biolog data), or gene essentiality data, can be used to improve the network content.
- The "singleGeneDeletion" can be used to compare 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 incapacabilities 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

Thiele, I. and B. O. Palsson (2010). "A protocol for generating a high-quality genome-scale metabolic reconstruction." *Nature protocols* 5(1): 93-121.



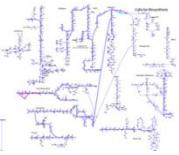
Single Reaction Deletion

```
% 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');
```

Reactions	grRateWT	grRateKO	grRatio
'ACALD'	0.873921507	0.873921507	1
'ACALDt'	0.873921507	0.873921507	1
'ACKr'	0.873921507	0.873921507	1
'ACONTa'	0.873921507	0	0
'ACONTb'	0.873921507	0	0
'AAct2r'	0.873921507	0.873921507	1
'ADK1'	0.873921507	0.873921507	1
'AKGDH'	0.873921507	0.858307408	0.982133294
'AKGt2r'	0.873921507	0.873921507	1
'ALCD2x'	0.873921507	0.873921507	1
'ATPM'	0.873921507	0.916647464	1.048889925
'ATPS4r'	0.873921507	0.374229875	0.428219093
'Biomass_Ecoli_core_N(w/GAM)_Nmet2'	0.873921507	0	0
'CO2t'	0.873921507	0.461669614	0.528273547
'CS'	0.873921507	0	0
'CYTBD'	0.873921507	0.21166295	0.24219904
'D_LACT2'	0.873921507	0.873921507	1
'ENO'	0.873921507	0	0
'ETOHT2r'	0.873921507	0.873921507	1

SingleReactionDeletionTest.xlsx



Single Gene Deletion

```
% SingleGeneDeletionTest.m
```

```
clear;
```

```
% Input model
```

```
model=readCbModel('ecoli_textbook');
```

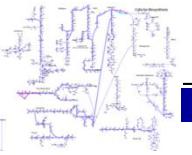
```
[grRatio,grRateKO,grRateWT,delRxns,hasEffect] = singleGeneDeletion(model,'FBA');
```

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

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

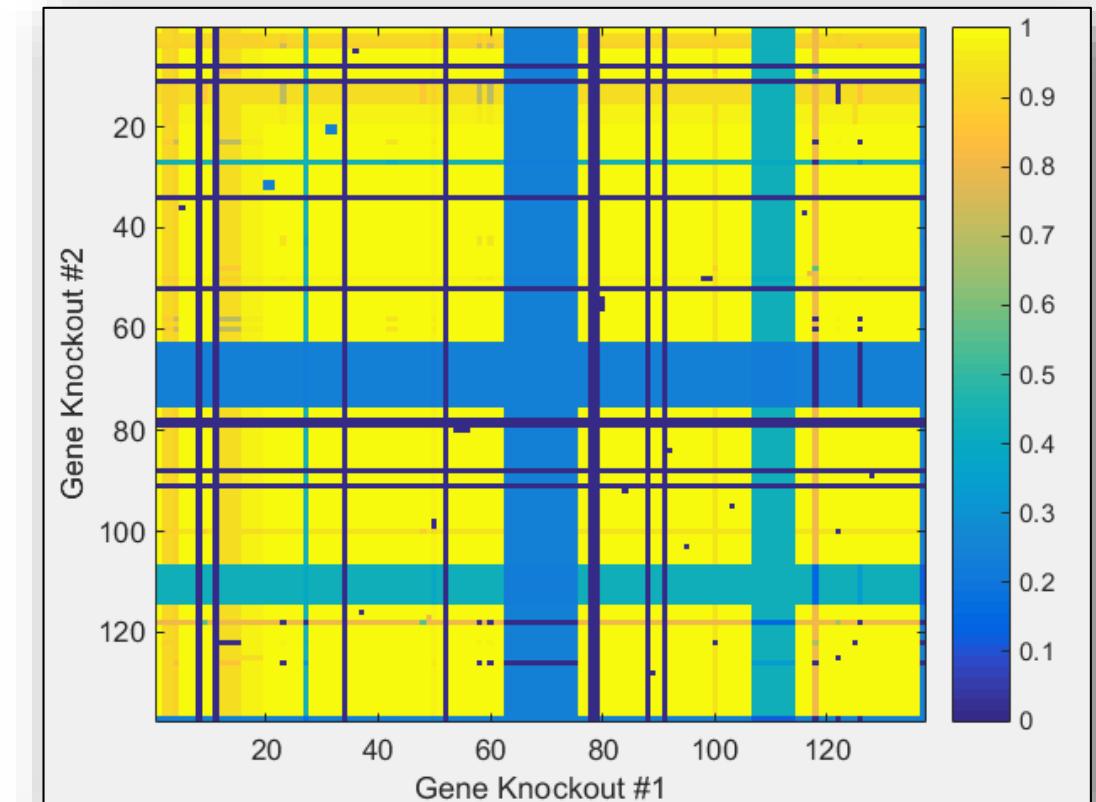
genes	grRateWT	grRateKO	grRatio	delRxns	hasEffect
'b0008'	0.873921507	0.873921507	1	FALSE	[]
'b0114'	0.873921507	0.796695925	0.911633275	TRUE	'PDH'
'b0115'	0.873921507	0.796695925	0.911633275	TRUE	'PDH'
'b0116'	0.873921507	0.782351053	0.895218903	TRUE	2x1 cell
'b0118'	0.873921507	0.873921507	1	FALSE	[]
'b0351'	0.873921507	0.873921507	1	FALSE	[]
'b0356'	0.873921507	0.873921507	1	FALSE	[]
'b0451'	0.873921507	0	0	TRUE	'NH4t'
'b0474'	0.873921507	0.873921507	1	TRUE	'ADK1'
'b0485'	0.873921507	0.873921507	1	FALSE	[]
'b0720'	0.873921507	0	0	TRUE	'CS'
'b0721'	0.873921507	0.814297508	0.931774194	TRUE	'SUCDI'
'b0722'	0.873921507	0.814297508	0.931774194	TRUE	'SUCDI'
'b0723'	0.873921507	0.814297508	0.931774194	TRUE	'SUCDI'
'b0724'	0.873921507	0.814297508	0.931774194	TRUE	'SUCDI'
'b0726'	0.873921507	0.858307408	0.982133294	TRUE	'AKGDH'
'b0727'	0.873921507	0.858307408	0.982133294	TRUE	'AKGDH'
'b0728'	0.873921507	0.858307408	0.982133294	TRUE	'SUCAOS'
'b0729'	0.873921507	0.858307408	0.982133294	TRUE	'SUCAOS'

SingleGeneDeletionTest.xlsx



Double Gene Deletion

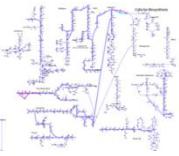
```
% DoubleGeneDeletionTest.m  
  
clear;  
  
% Input model  
  
model=readCbModel('ecoli_textbook');  
  
[grRatio,grRateKO,grRateWT] = doubleGeneDeletion(model,'FBA');  
  
%[grRatio,grRateKO,grRateWT,delRxns,hasEffect] = doubleGeneDeletion(model,'MOMA');  
  
%[grRatio,grRateKO,grRateWT,delRxns,hasEffect] = 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.
- 46-48. Perform 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 the model can produce known secretion products.
- 76-78. Check for blocked reactions.
- 79-80. Compute single gene deletion phenotypes.
- 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.

Thiele, I. and B. O. Palsson (2010). "A protocol for generating a high-quality genome-scale metabolic reconstruction." *Nature protocols* 5(1): 93-121.



Stage 4: Network Evaluation

- 43-44. Test if network is mass-and charge balanced.
45. Identify metabolic dead-ends.
- 46-48. Perform 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 the model can produce known secretion products.
- 76-78. Check for blocked reactions.
- 79-80. Compute single gene deletion phenotypes.
- 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.

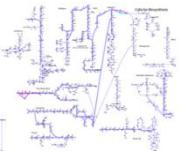


Network Evaluation

Test if the Model Grows too Fast

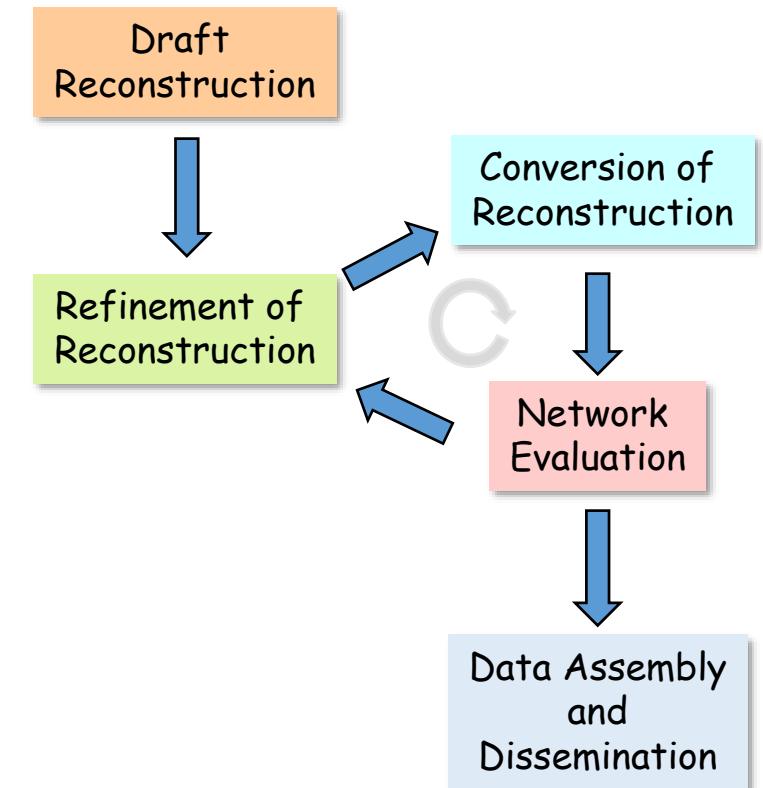
- Check boundary constraints
 - ✓ `printConstraints(model,MinInf,MaxInf)`
- Check reaction directionality
 - ✓ `printRxnFormula(model)`
- Use single-reaction deletion to identify single reactions that may enable the model to grow too fast.
 - ✓ `[grRatio,grRateKO,grRateWT] = singleRxnDeletion(model, '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)..
 - ✓ `FBAsolution = optimizeCbModel(model,osenseStr,primalOnlyFlag)`
 - ✓ set `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.

Thiele, I. and B. O. Palsson (2010). "A protocol for generating a high-quality genome-scale metabolic reconstruction." *Nature protocols* 5(1): 93-121.

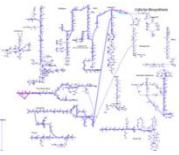


GENOME-SCALE METABOLIC RECONSTRUCTIONS

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

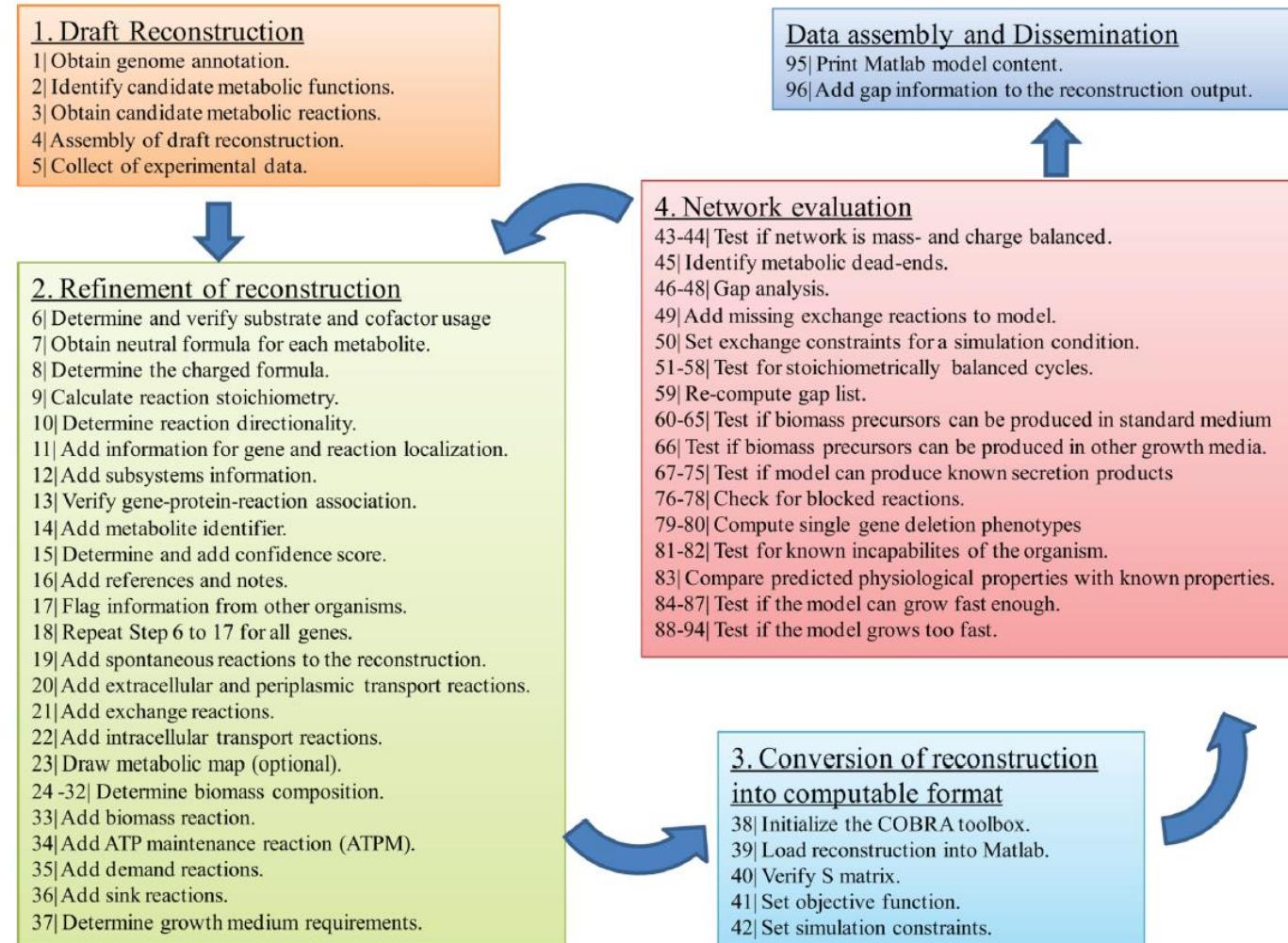


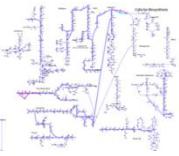
Thiele, I. and B. O. Palsson (2010). "A protocol for generating a high-quality genome-scale metabolic reconstruction." *Nature protocols* 5(1): 93-121.



Reconstruction Process: 96 Step Protocol

Thiele, I. and B. O. Palsson (2010). "A protocol for generating a high-quality genome-scale metabolic reconstruction." *Nature protocols* 5(1): 93-121.





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')
```

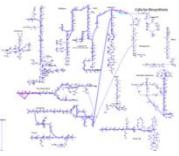
- SBML Cobra function

```
writeCBmodel(model,'xls','FileName')
```

96. Add gap information to the reconstruction output.

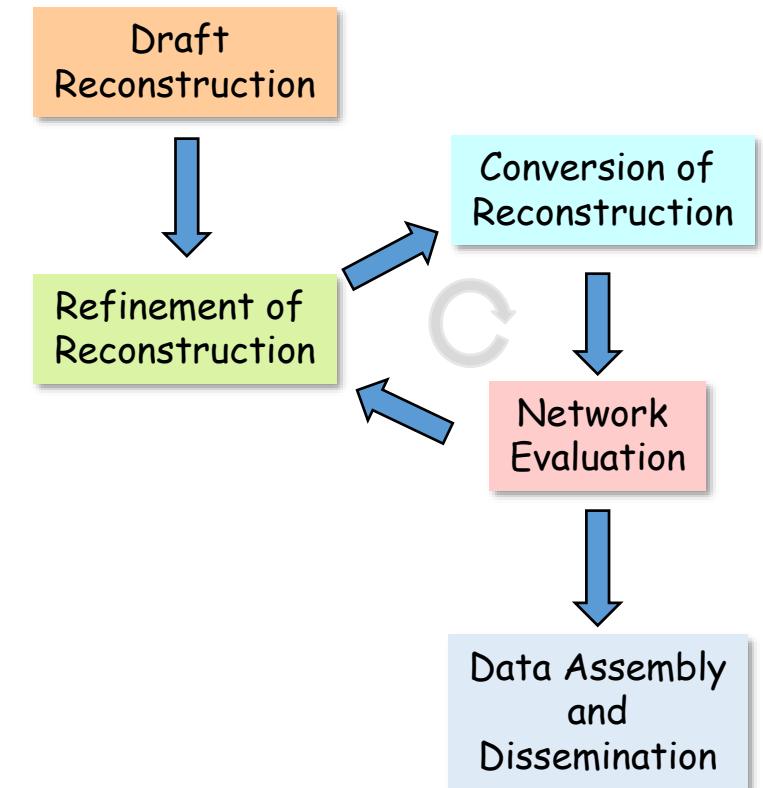
- Completed in Steps 45-48

Thiele, I. and B. O. Palsson (2010). "A protocol for generating a high-quality genome-scale metabolic reconstruction." *Nature protocols* 5(1): 93-121.

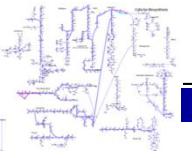


GENOME-SCALE METABOLIC RECONSTRUCTIONS

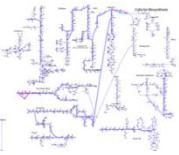
- Overview
- Draft Reconstruction
- Refinement of Reconstruction
- Conversion of Reconstruction into Computable Format
- Network Evaluation
- Data Assembly and Dissemination



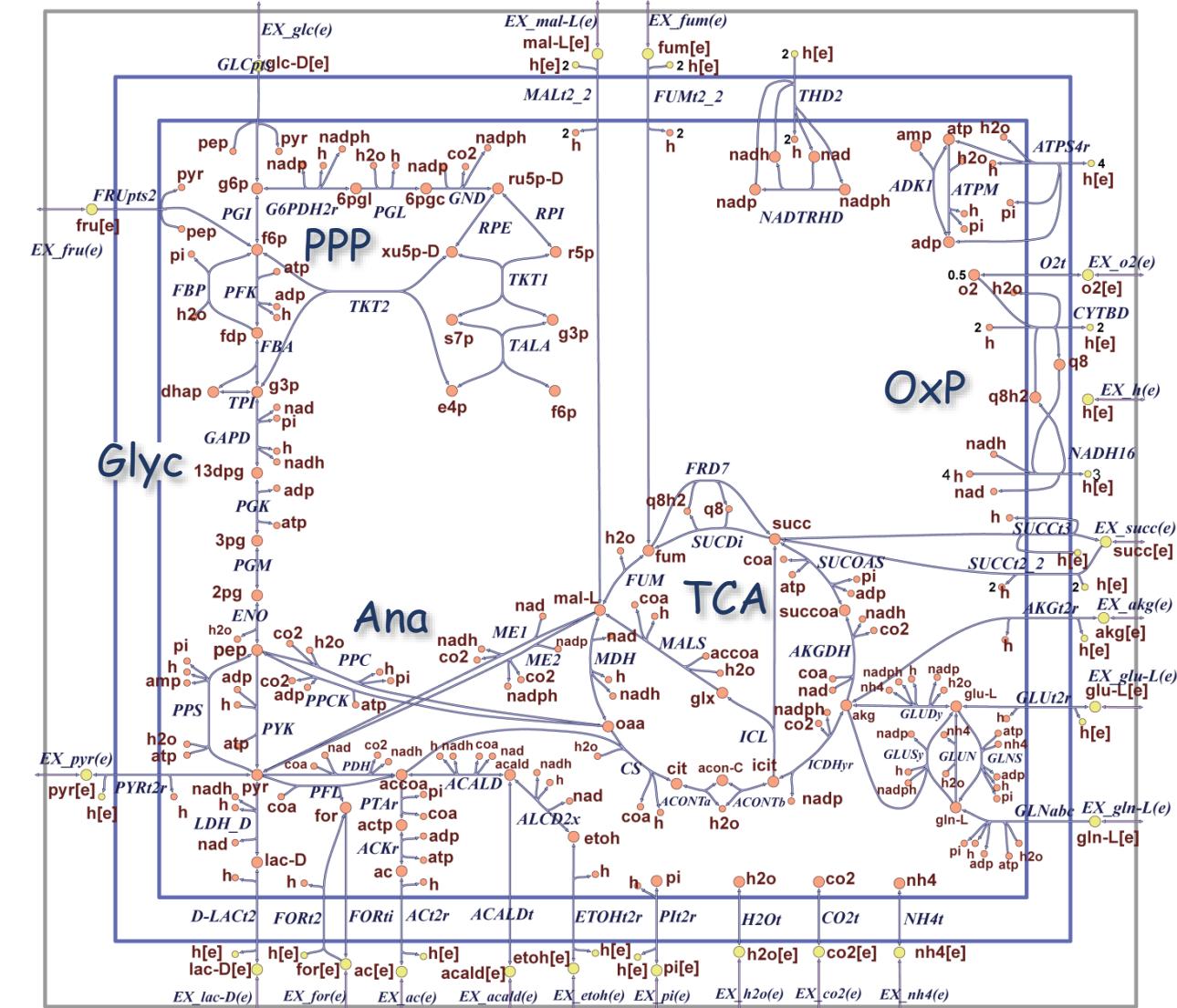
Thiele, I. and B. O. Palsson (2010). "A protocol for generating a high-quality genome-scale metabolic reconstruction." Nature protocols 5(1): 93-121.



EXTRAS



E.coli Core Model



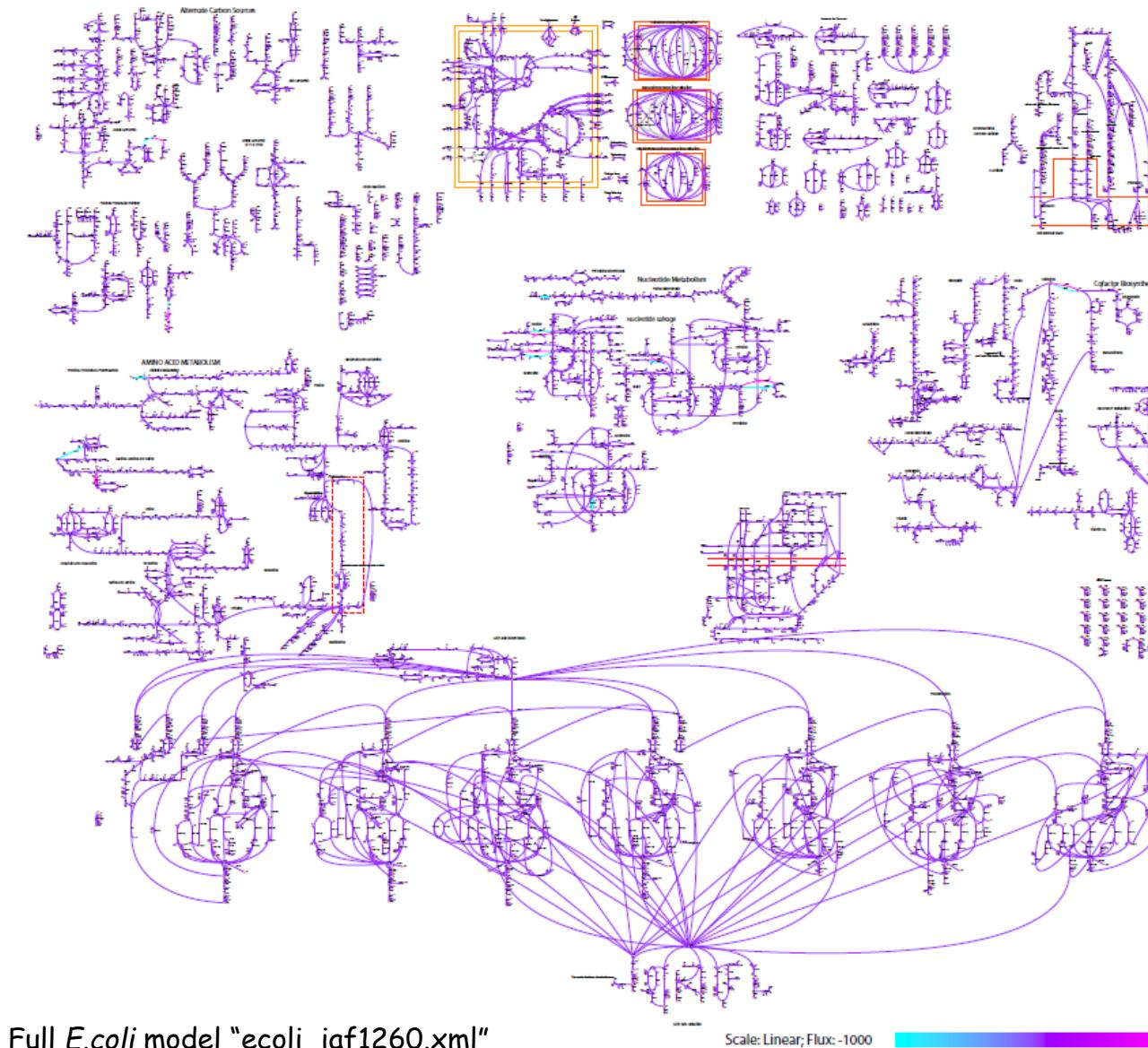
Orth, J. D., I. Thiele, et al. (2010). "What is flux balance analysis?" *Nature biotechnology* 28(3): 245-248.

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



Constraint-based Metabolic Reconstructions & Analysis

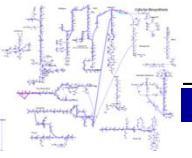
H. Scott Hinton, 2016 -185-



Full *E.coli* model "ecoli_iaf1260.xml"

Scale: Linear; Flux: -1000

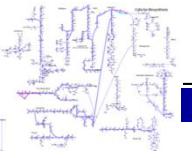




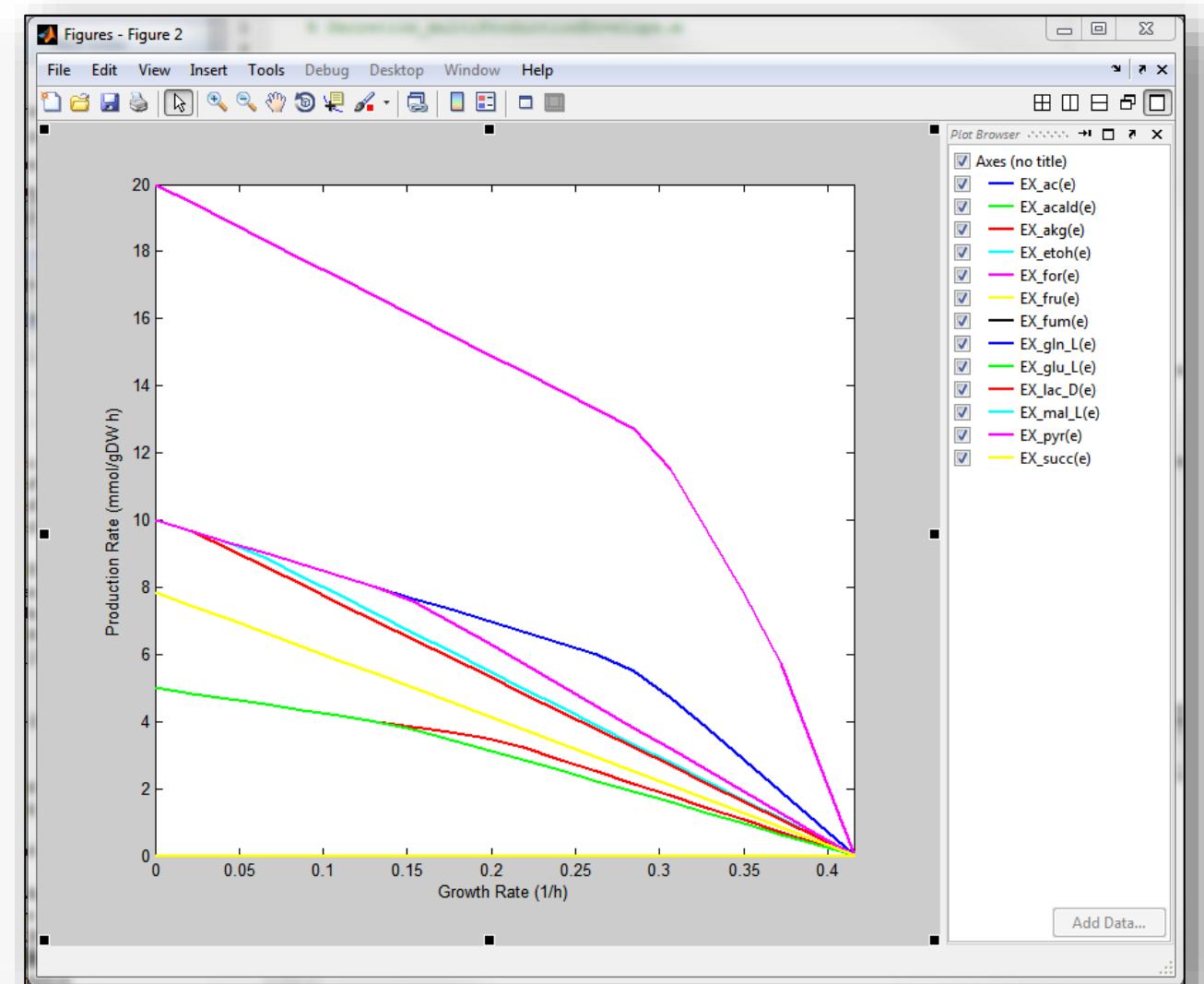
MassChargeBalance_iaf1260_MB.m Example:

“UnbalancedRxns” Matrix

Reaction Indices	H	C	O	P	S	N	Mg	X	Fe	Zn	Co	R
371	2											
372	1											
144		1										
167		-1										
187		1										
195		-1										
198		1										
199		1										
227		1										
...												
2324		1										



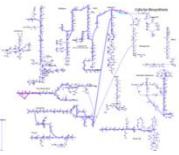
Secreted Metabolites





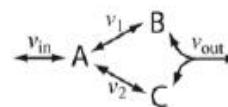
Addition of Constraints

- 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)

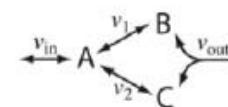


Addition of Constraints (II)

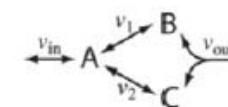
Metabolic Network



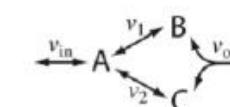
Thermodynamics (reversibility)



Maximum enzyme capacity



Mass balance of metabolites



Kinetics

