Model Creation-Enhancement

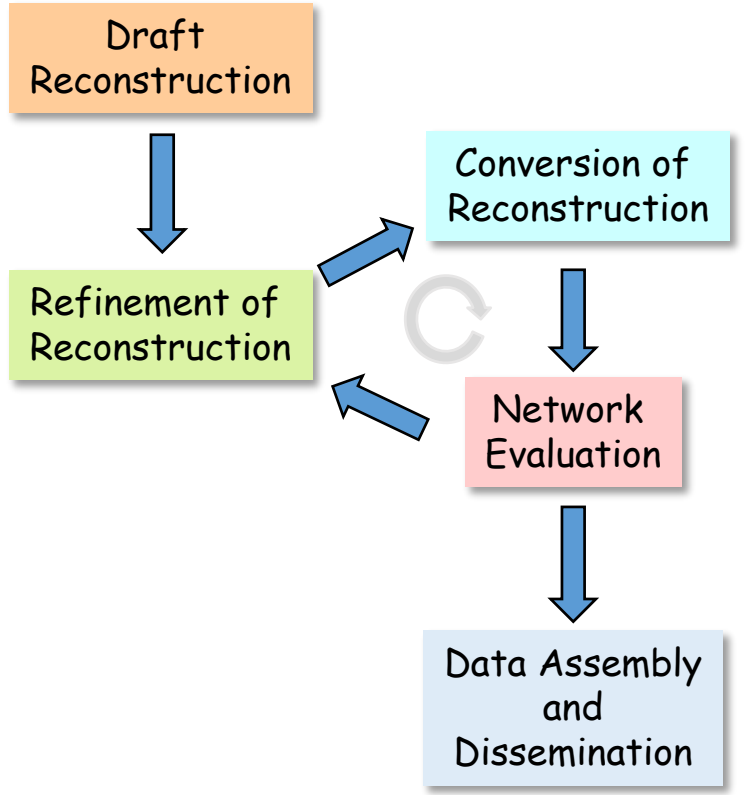
Learning Objectives

Each student should be able to:

- Explain the process of creating a genome-scale metabolic reconstruction
- Explain the reaction requirements for a COBRA model
- Explain the metabolite requirements for a COBRA model
- Explain the role of the biomass function
- Explain the role of the demand and sink reactions
- Explain the types of extreme pathways
- Demonstrate the ability to create a COBRA model
- Demonstrate the ability to add reactions to an existing COBRA model

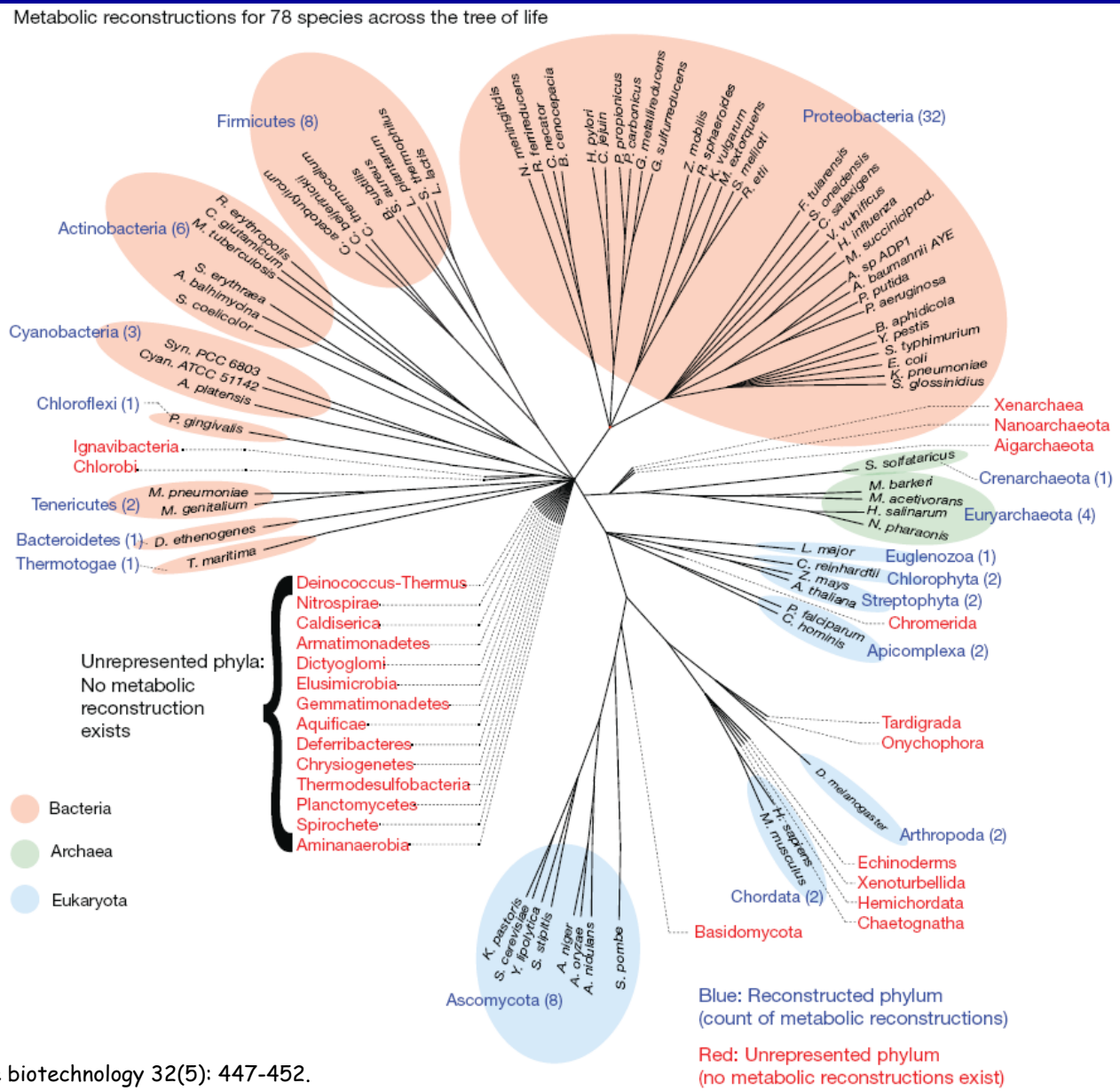
Lesson Outline

- Overview
- Fundamentals of Genome-scale Model Construction
  - ✓ Reconstruction Process
  - ✓ Required Reaction Information
  - ✓ Required Metabolite Information
  - ✓ Required Gene Information
  - ✓ Biomass Composition
  - ✓ Demand/Sink Reactions
  - ✓ Extreme Pathways
- Model Construction with COBRApy
  - ✓ Creating New Models
  - ✓ Adding to Existing Models



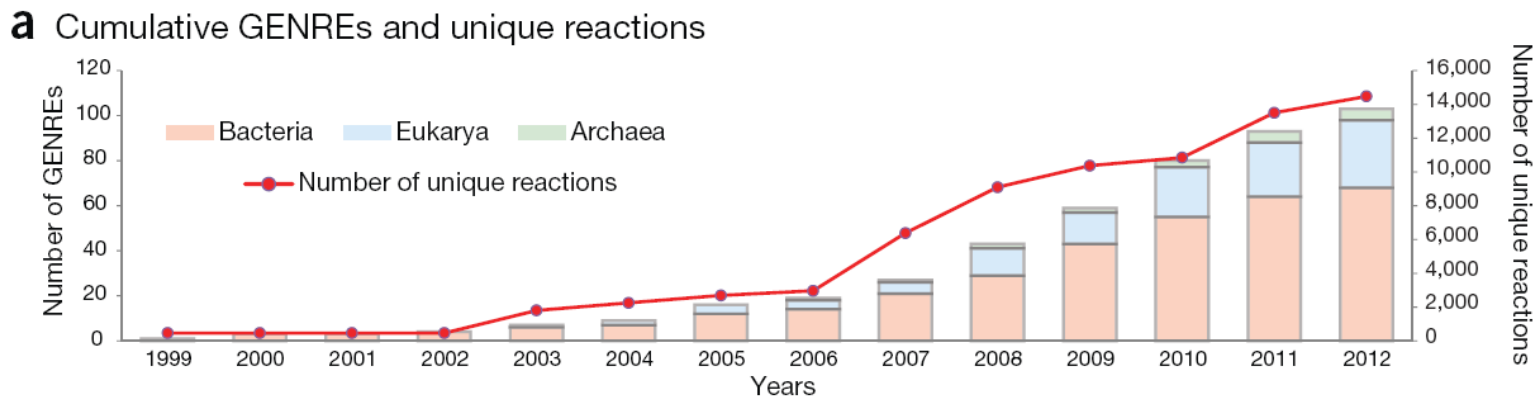
Phylogenetic Coverage of Genome-scale Network Reconstructions

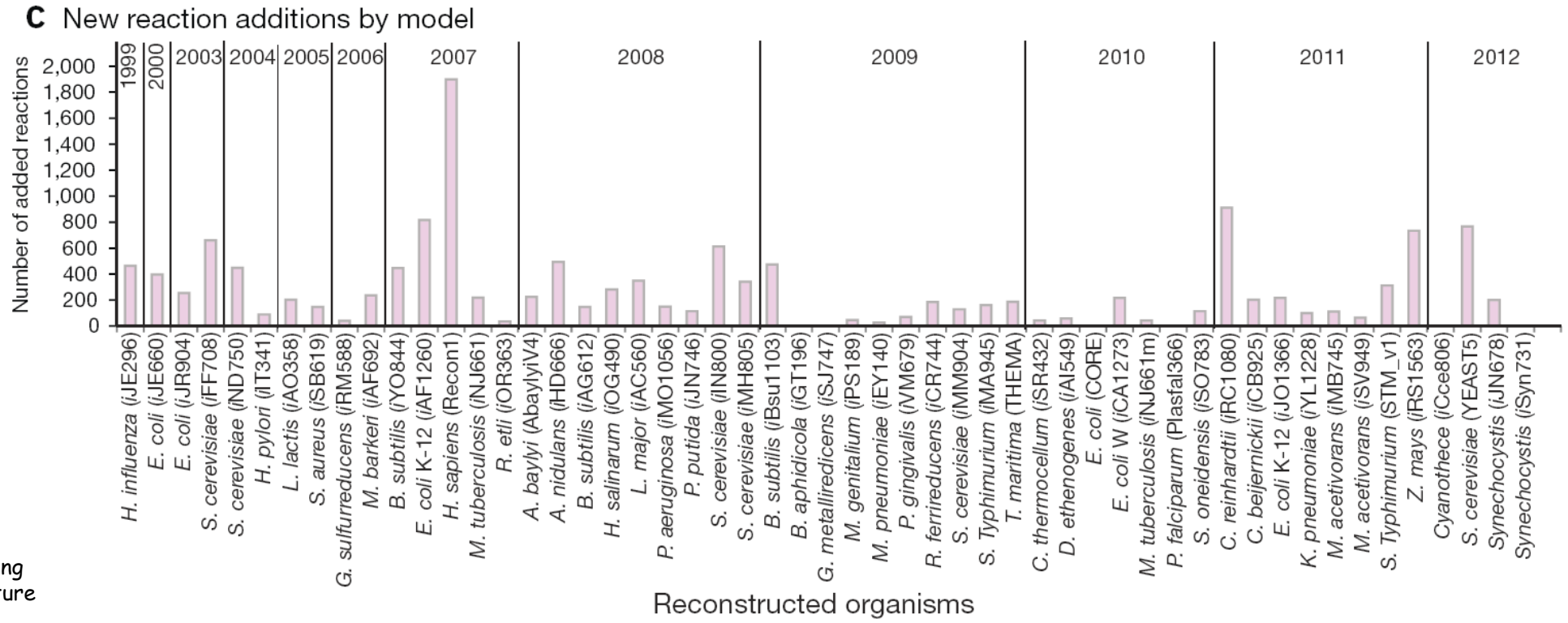
A GEnome scale Network Reconstructions (GENREs) serves as a structured knowledge base of established biochemical facts, while a GEnome scale Models (GEMs) is a model which supplements the established biochemical information with additional (potentially hypothetical) information to enable computational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

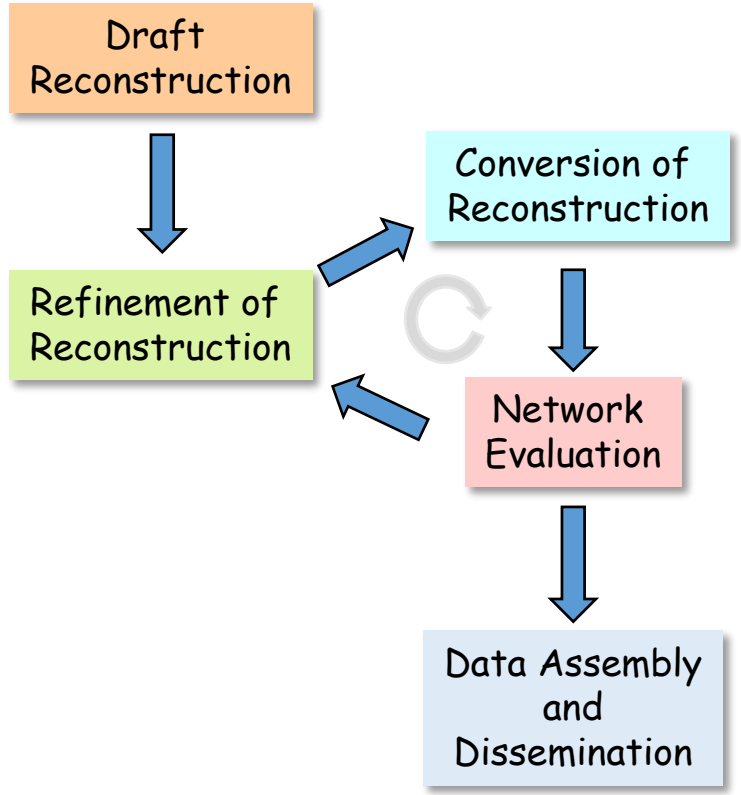




Monk, J., J. Nogales, et al. (2014). "Optimizing genome-scale network reconstructions." Nature biotechnology 32(5): 447-452.

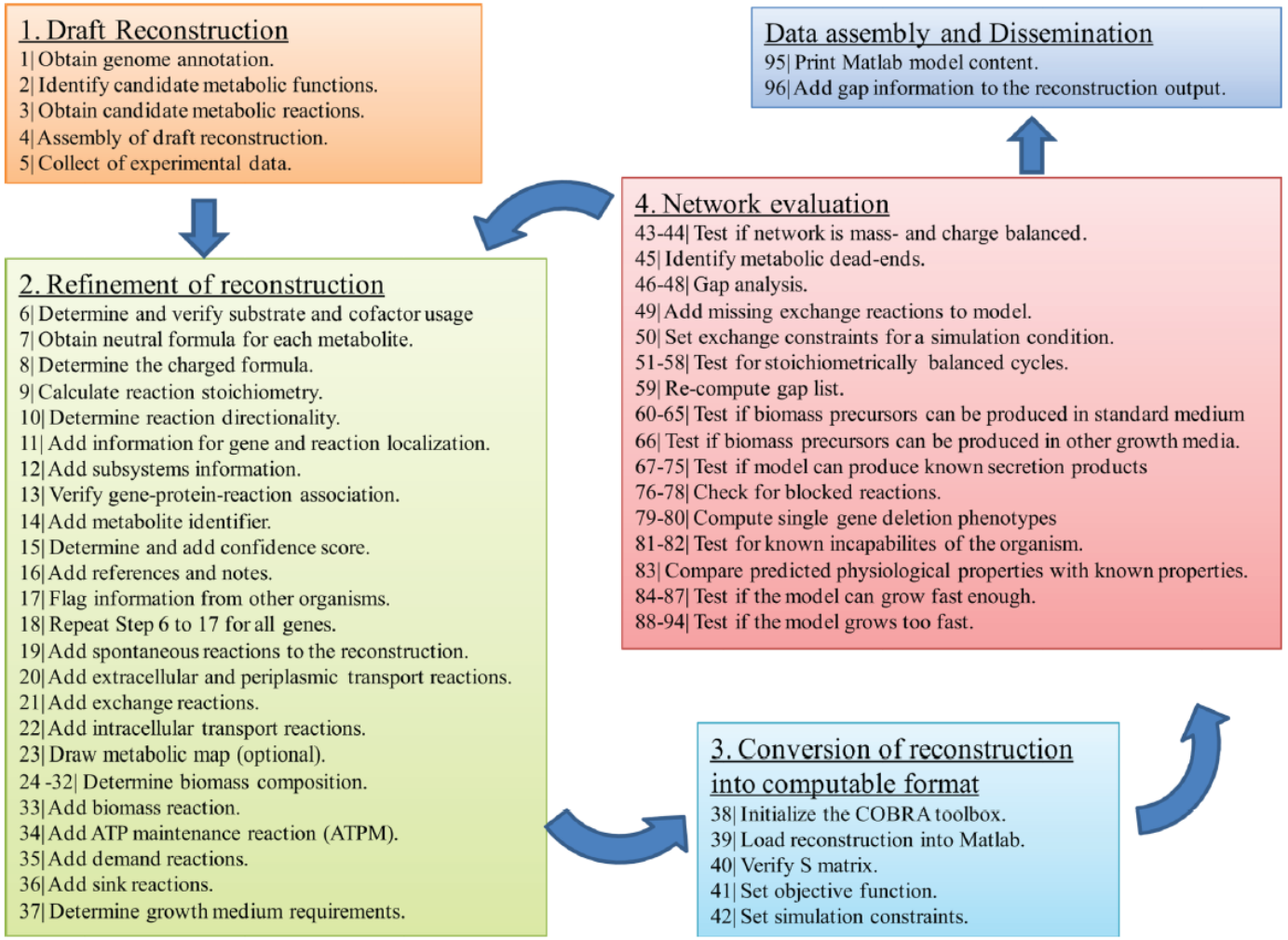
Lesson Outline

- Overview
- Fundamentals of Genome-scale Model Construction
  - ✓ Reconstruction Process
  - ✓ Required Reaction Information
  - ✓ Required Metabolite Information
  - ✓ Required Gene Information
  - ✓ Biomass Composition
  - ✓ Demand/Sink Reactions
  - ✓ Extreme Pathways
- Model Construction with COBRApy
  - ✓ Creating New Models
  - ✓ Adding to Existing Models



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.



PROTOCOL

A protocol for generating a high-quality genome-scale metabolic reconstruction

Ines Thiele[1,2] & Bernhard Ø Palsson[1]

[1]Department of Bioengineering, University of California, San Diego, La Jolla, California, USA. [2]Current address: Center for Systems Biology, Faculty of Industrial Engineering, Mechanical Engineering and Computer Science, University of Iceland, Reykjavik, Iceland. Correspondence should be addressed to B.Ø.P. (palsson@ucsd.edu).

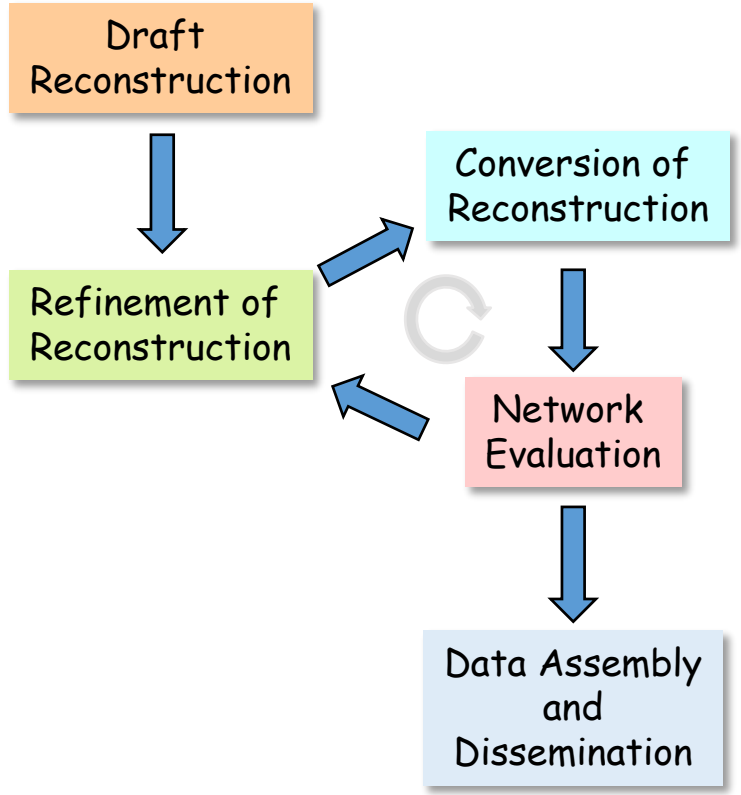
Published online 7 January 2010; doi:10.1038/nprot.2009.203

Network reconstructions are a common denominator in systems biology. Bottom–up metabolic network reconstructions have been developed over the last 10 years. These reconstructions represent structured knowledge bases that abstract pertinent information on the biochemical transformations taking place within specific target organisms. The conversion of a reconstruction into a mathematical format facilitates a myriad of computational biological studies, including evaluation of network content, hypothesis testing and generation, analysis of phenotypic characteristics and metabolic engineering. To date, genome-scale metabolic reconstructions for more than 30 organisms have been published and this number is expected to increase rapidly. However, these reconstructions differ in quality and coverage that may minimize their predictive potential and use as knowledge bases. Here we present a comprehensive protocol describing each step necessary to build a high-quality genome-scale metabolic reconstruction, as well as the common trials and tribulations. Therefore, this protocol provides a helpful manual for all stages of the reconstruction process.

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

Lesson Outline

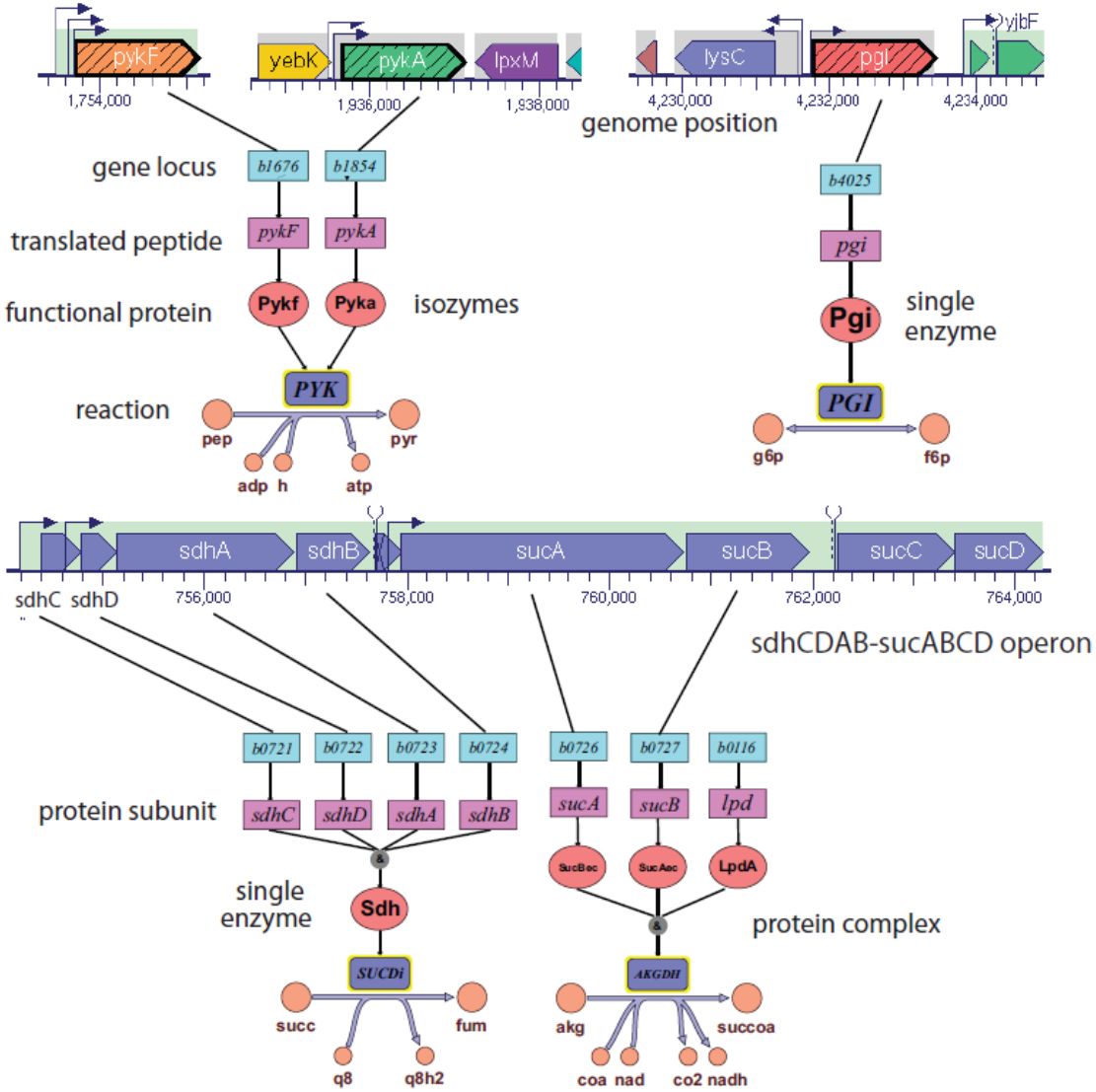
- Overview
- Fundamentals of Genome-scale Model Construction
  - ✓ Reconstruction Process
  - ✓ Required Reaction Information
  - ✓ Required Metabolite Information
  - ✓ Required Gene Information
  - ✓ Biomass Composition
  - ✓ Demand/Sink Reactions
  - ✓ Extreme Pathways
- Model Construction with COBRApy
  - ✓ Creating New Models
  - ✓ Adding to Existing Models



Desired Reaction Information

1. Reaction ID* - r_ID
2. Reaction Name* - model.reactions.r_ID.name
3. Reaction Formula* - model.reactions.r_ID.reaction
4. Gene-reaction Association* - model.reactions.r_ID.gene_name_reaction_rule
5. Genes (Gene Locus) * - model.reactions.r_ID.gene_reaction_rule
6. Subsystem * - model.reactions.r_ID.subsystem
7. Reaction Direction* - model.reactions.r_ID.reversibility
8. Flux Lower Bound* - model.reactions.r_ID.lower_bound
9. Flux Upper Bound* - model.reactions.r_ID.upper_bound
10. Annotations - model.reactions.r_ID.annotation
11. Notes - model.reactions.r_ID.notes
12. Mass balance - model.reactions.r_ID.check_mass_balance
13. Reactants - model.reactions.r_ID.reactants
14. Products - model.reactions.r_ID.products

* Required



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 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.
- **Annotations and Notes:**
  - ✓ 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.

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.

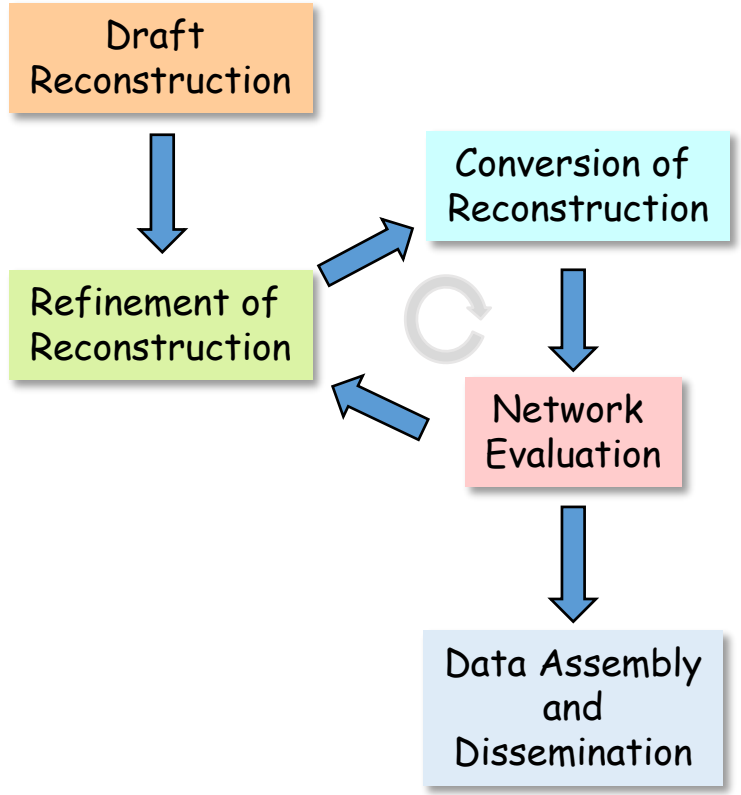
Reaction Names Suffix Standards

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	10FTHFtl
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.

Lesson Outline

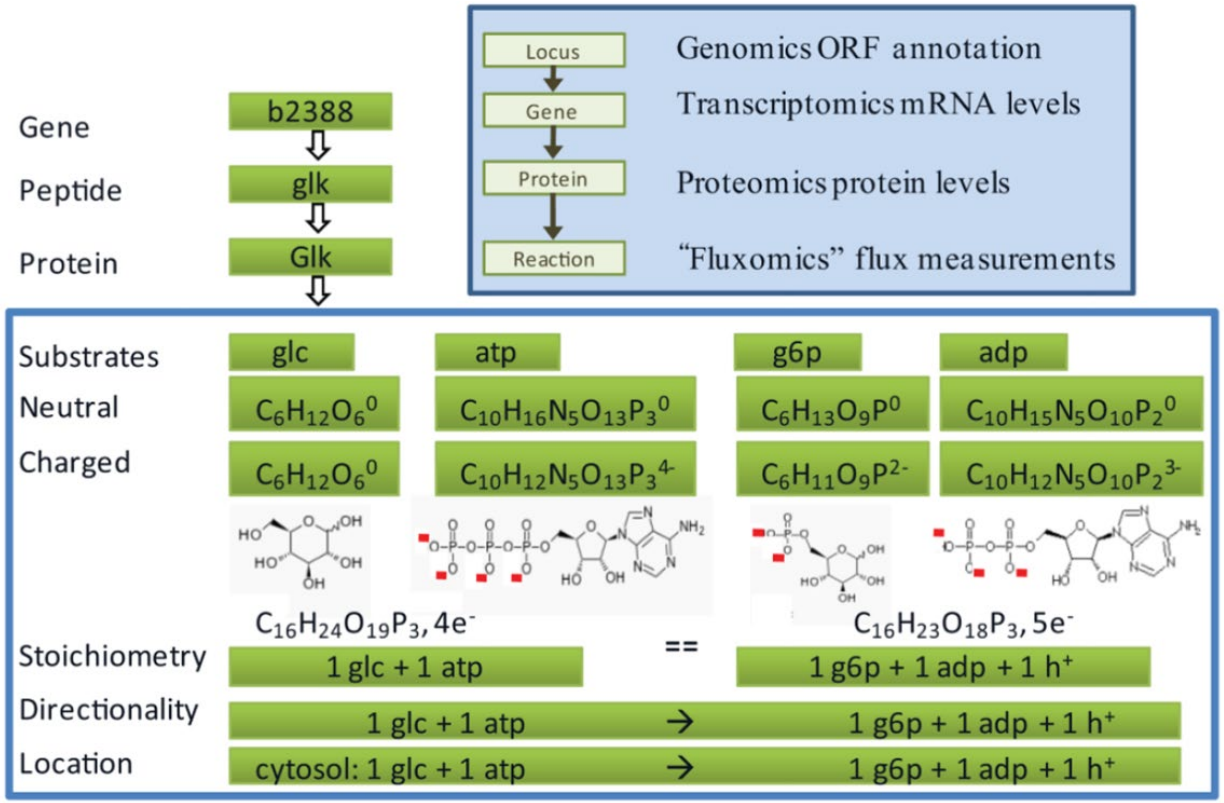
- Overview
- Fundamentals of Genome-scale Model Construction
  - ✓ Reconstruction Process
  - ✓ Required Reaction Information
  - ✓ Required Metabolite Information
  - ✓ Required Gene Information
  - ✓ Biomass Composition
  - ✓ Demand/Sink Reactions
  - ✓ Extreme Pathways
- Model Construction with COBRApy
  - ✓ Creating New Models
  - ✓ Adding to Existing Models



Desired Metabolite Information

1. Metabolite ID* - m_ID
2. Metabolite Name* - model.metabolites.m_ID.name
3. Metabolite Formula* - model.metabolites.m_ID.formula
4. Metabolite Charge* - model.metabolites.m_ID.charge
5. Metabolite Compartment* - model.metabolites.m_ID.compartment
6. Metabolite Elements - model.metabolites.m_ID.elements
7. Metabolite Formula Weight - model.metabolites.m_ID.formula_weight
8. Metabolite Summary - model.metabolites.m_ID.summary()
9. Metabolite Annotations - model.metabolites.m_ID.annotation
10. Metabolite Notes - model.metabolites.m_ID.notes

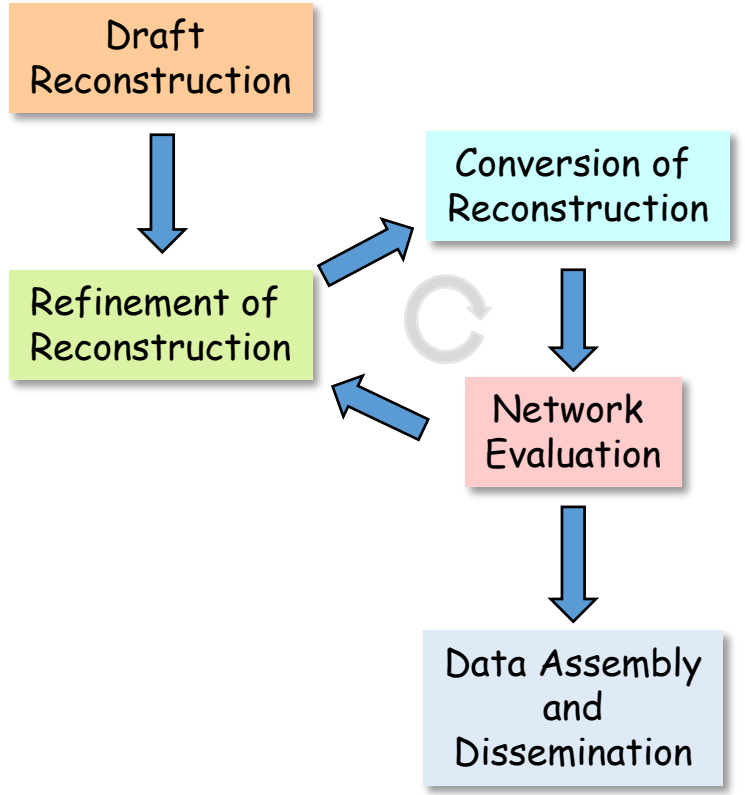
* Required



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

Lesson Outline

- Overview
- Fundamentals of Genome-scale Model Construction
  - ✓ Reconstruction Process
  - ✓ Required Reaction Information
  - ✓ Required Metabolite Information
  - ✓ Required Gene Information
  - ✓ Biomass Composition
  - ✓ Demand/Sink Reactions
  - ✓ Extreme Pathways
- Model Construction with COBRApy
  - ✓ Creating New Models
  - ✓ Adding to Existing Models



Gene Information

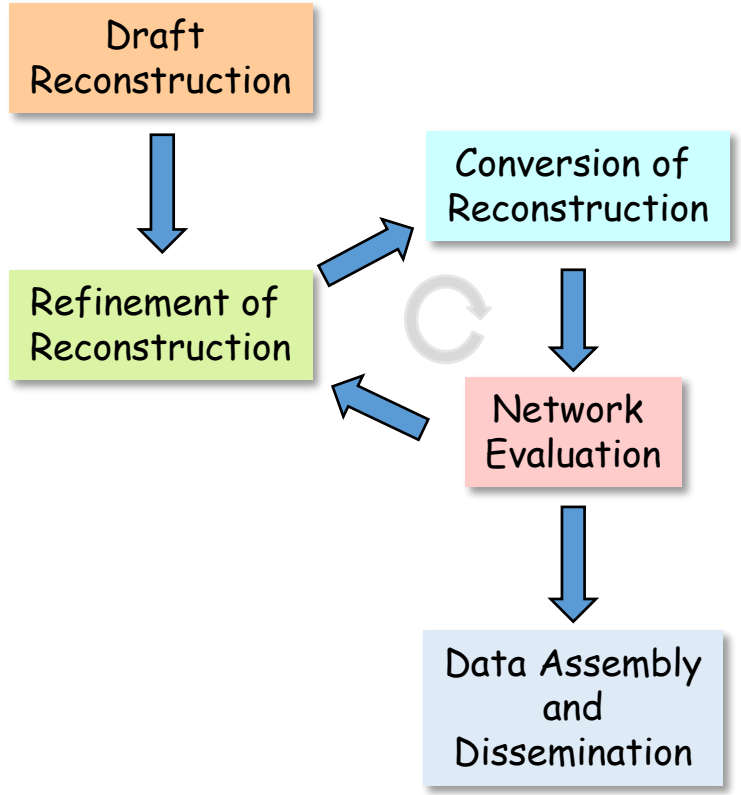
1. Gene ID (g_ID) – model.genes.g_ID
2. Gene Name - model.genes.g_ID.name
3. Gene Functional - model.genes.g_ID.functional
4. Gene Associated Reactions - model.genes.g_ID.reactions

Genes within a COBRA model automatically generate their specific information from the reaction 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.

Lesson Outline

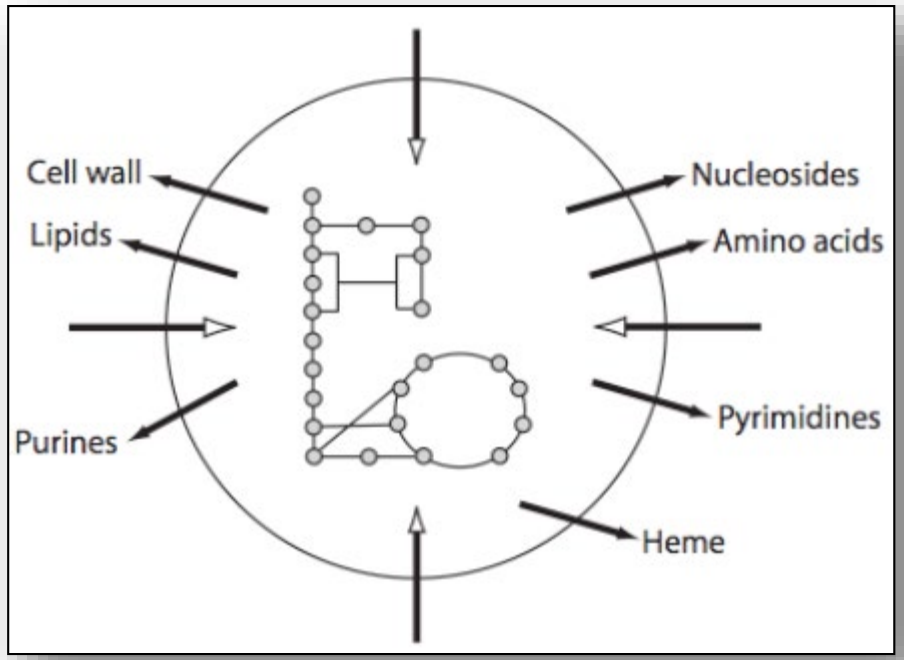
- Overview
- Fundamentals of Genome-scale Model Construction
  - ✓ Reconstruction Process
  - ✓ Required Reaction Information
  - ✓ Required Metabolite Information
  - ✓ Required Gene Information
  - ✓ Biomass Composition
  - ✓ Demand/Sink Reactions
  - ✓ Extreme Pathways
- Model Construction with COBRApy
  - ✓ Creating New Models
  - ✓ Adding to Existing Models



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.

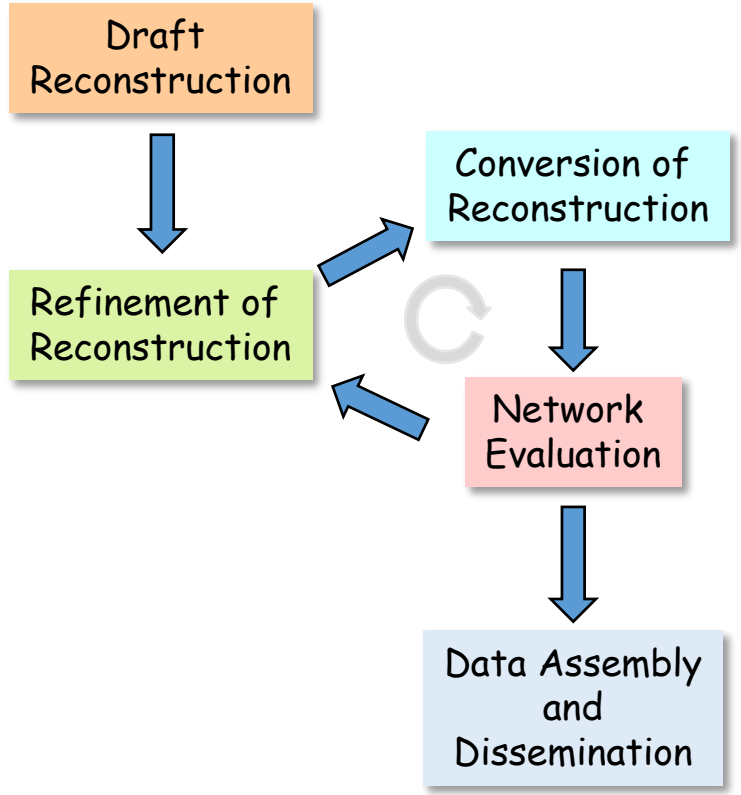
Important Role of the Biomass Objective Function

- 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).
- To predict the optimal growth rate accurately, the fractional distribution of each compound has an important role.
- The unit of the biomass reaction is h^{-1}, as all biomass precursor fractions are converted to $mmol \cdot gDW^{-1}$. 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.

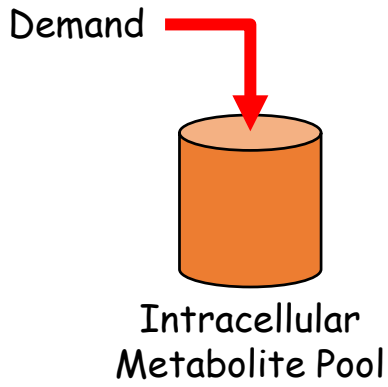
Lesson Outline

- Overview
- Fundamentals of Genome-scale Model Construction
  - ✓ Reconstruction Process
  - ✓ Required Reaction Information
  - ✓ Required Metabolite Information
  - ✓ Required Gene Information
  - ✓ Biomass Composition
  - ✓ Demand/Sink Reactions
  - ✓ Extreme Pathways
- Model Construction with COBRApy
  - ✓ Creating New Models
  - ✓ Adding to Existing Models



Refinement of Reconstruction:
Add Demand Reactions

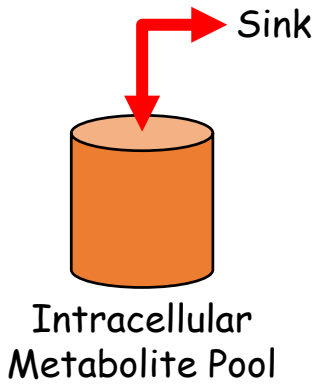
- Demand reactions are unbalanced network reactions that allow the accumulation of a compound, which otherwise is not allowed in steady-state models because of mass-balancing requirements (i.e., in steady state the sum of influx equals the sum of efflux for each metabolite).
- In general, metabolic reconstructions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

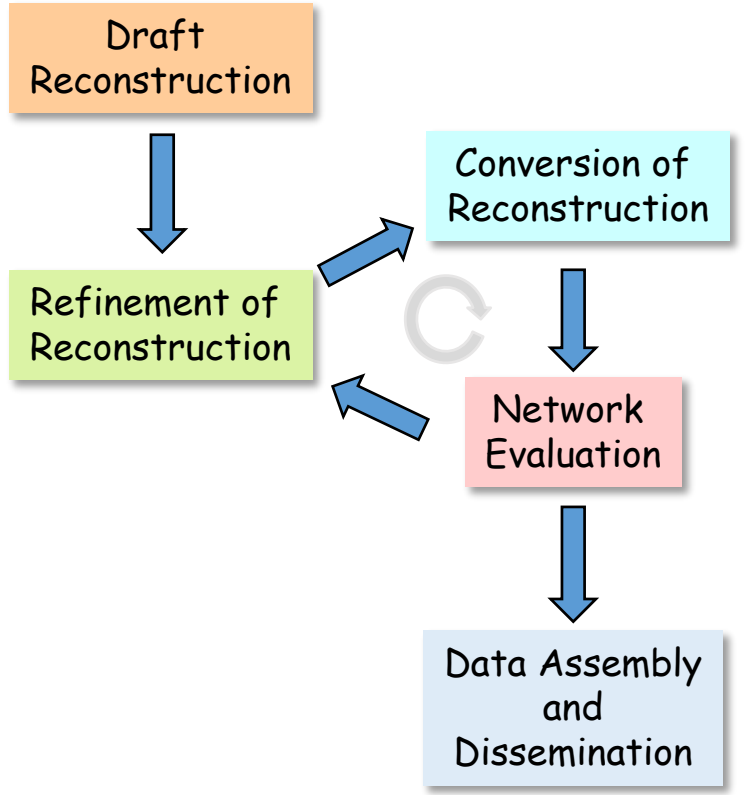
- Sink reactions are similar to demand reactions but are defined to be reversible and thus provide the network with metabolites.
- These sink reactions are of great use for compounds that are produced by non-metabolic cellular processes but that need to be metabolized.
- Adding too many sink reactions 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.



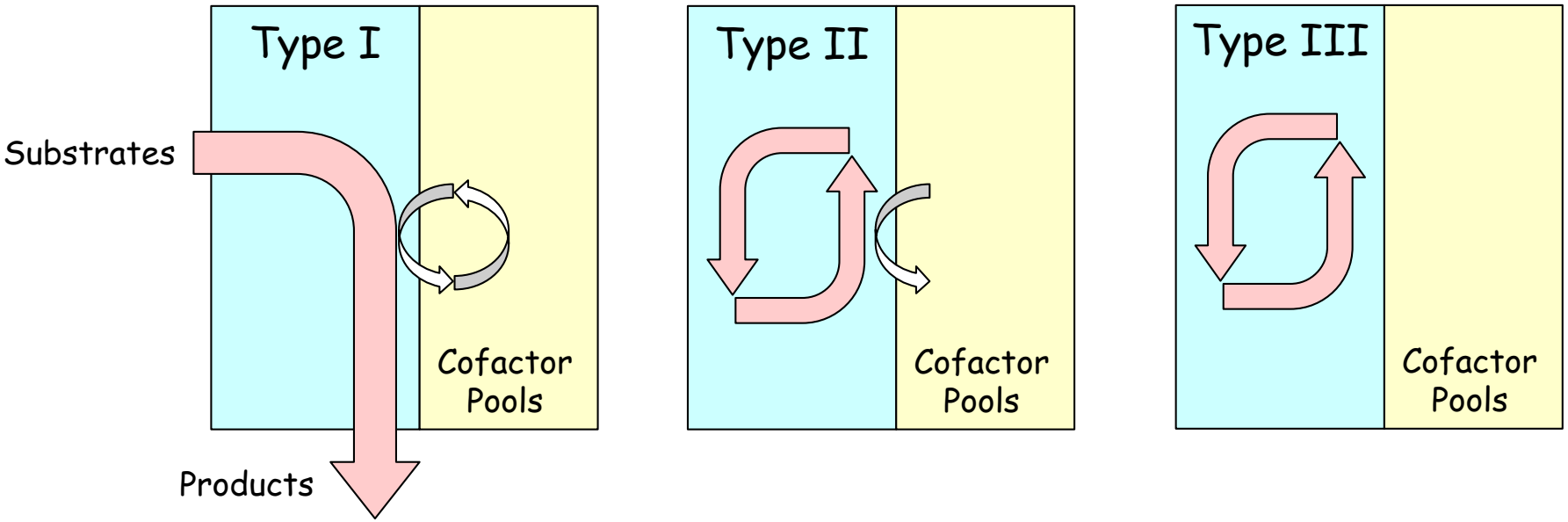
Thiele, I. and B. O. Palsson (2010). "A protocol for generating a high-quality genome-scale metabolic reconstruction." Nature protocols 5(1): 93-121.

Lesson Outline

- Overview
- Fundamentals of Genome-scale Model Construction
  - ✓ Reconstruction Process
  - ✓ Required Reaction Information
  - ✓ Required Metabolite Information
  - ✓ Required Gene Information
  - ✓ Biomass Composition
  - ✓ Demand/Sink Reactions
  - ✓ Extreme Pathways
- Model Construction with COBRApy
  - ✓ Creating New Models
  - ✓ Adding to Existing Models



Types of Extreme Pathways

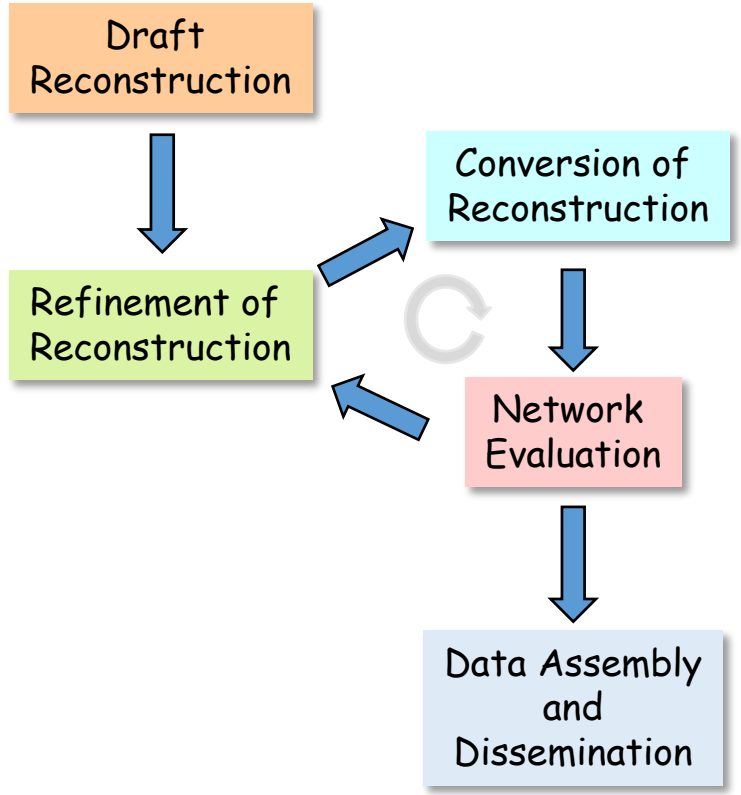


- Type I extreme pathways have exchange fluxes across the system boundaries that correspond to non-currency metabolites.
- Type II extreme pathways have only currency metabolites that cross system boundaries.
- Type III extreme pathways do not contain any exchange fluxes, and thus correspond to internal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

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

Lesson Outline

- Overview
- Fundamentals of Genome-scale Model Construction
  - ✓ Reconstruction Process
  - ✓ Required Reaction Information
  - ✓ Required Metabolite Information
  - ✓ Required Gene Information
  - ✓ Biomass Composition
  - ✓ Demand/Sink Reactions
  - ✓ Extreme Pathways
- Model Construction with COBRApy
  - ✓ Creating New Models
  - ✓ Adding to Existing Models



Creating New Models

Building_model.ipynb

Building a Model

Building a model component by component

Model, Reactions and Metabolites

(Adapted from https://cobrapy.readthedocs.io/en/latest/building_model.html)

This simple example demonstrates how to create a model, create a reaction, and then add the reaction to the model.

We'll use the '3OAS140' reaction from the *E.coli* iJO1366 model:

```
1.0 malACP_c + 1.0 h_c + 1.0 ddcaACP_c → 1.0 co2_c + 1.0 ACP_c + 1.0 3omrsACP_c
```

First, create the model.

In [1]:
```
from cobra import Model, Reaction, Metabolite

model = Model('example_model')
```

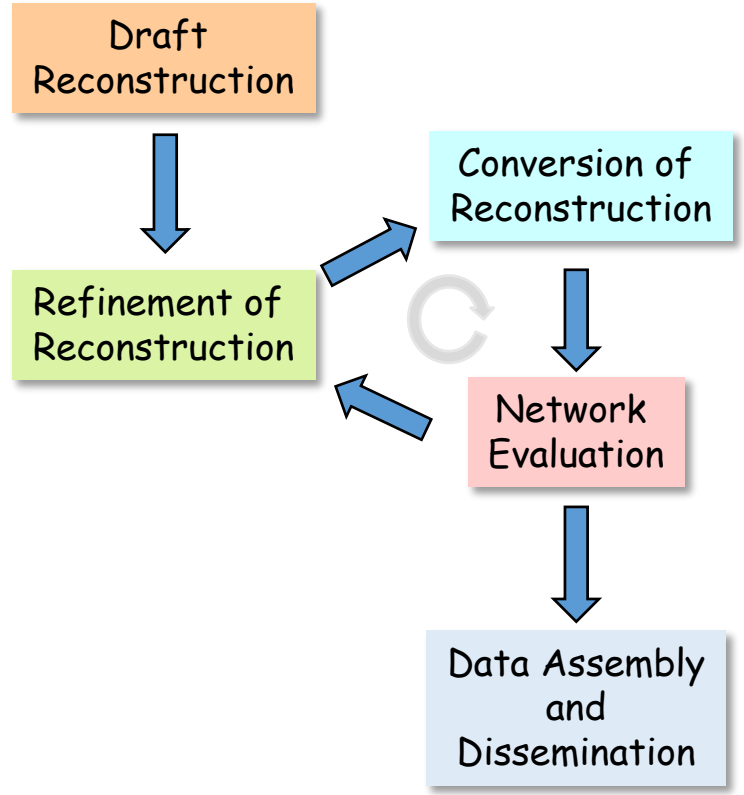
```
Set parameter Username
Academic license - for non-commercial use only - expires 2022-10-10
```

Looking at the new models content, it should be empty with no reactions, metabolites, genes or objective function.

In [2]:
```
model
```

Lesson Outline

- Overview
- Fundamentals of Genome-scale Model Construction
  - ✓ Reconstruction Process
  - ✓ Required Reaction Information
  - ✓ Required Metabolite Information
  - ✓ Required Gene Information
  - ✓ Biomass Composition
  - ✓ Demand/Sink Reactions
  - ✓ Extreme Pathways
- Model Construction with COBRApy
  - ✓ Creating New Models
  - ✓ Adding to Existing Models



Adding Heme Oxygenase to an iJO1366 *E. coli* Model

Biliverdin IXα is produced when heme undergoes reductive ring cleavage at the α-methene bridge catalyzed by heme oxygenase (HOXG). Biliverdin IXα, through interaction with biliverdin reductase, initiates signaling pathways leading to anti-inflammatory responses and suppression of cellular pro-inflammatory events. The use of biliverdin IXα as a cytoprotective therapeutic has been suggested. [1,2]

In this notebook we will add the HOXG reaction to an iJO1366 *E. coli* model.

Begin by setting up by loading the appropriate COBRApy and Python packages.

In [1]:
```
from cobra import Model, Reaction, Metabolite
from cameo import models
from cameo.visualization.plotting.with_plotly import PlotlyPlotter
plotter = PlotlyPlotter()
import cobra.test
from cobrapy_bigg_client import client
```

Since a HOXG reaction can be found in the *Synechocystis sp. PCC 6803* iJN678 model, let's load both the iJO1366 and iJN678 models.

In [2]:
```
#model_iJO1366 = models.bigg.iJO1366 # Cameo tool for loading BIGG models
#model_iJN678 = models.bigg.iJN678 # Cameo tool for loading BIGG models
model_iJO1366 = client.download_model('iJO1366', save=False)
model_iJN678 = client.download_model('iJN678', save=False)
model_iJO1366.solver = 'glpk'
```

```
Set parameter Username
Academic license - for non-commercial use only - expires 2022-10-10
```

Adding_HOXG_to_iJO1366.ipynb

Reflective Questions

1. How many steps are required for the genome-scale metabolic reconstruction process?
2. What information is required about reactions for the genome-scale metabolic reconstruction process?
3. What information is required about metabolites for the genome-scale metabolic reconstruction process?
4. What information is required about genes for the genome-scale metabolic reconstruction process?
5. What is a gene locus?
6. What is a gene-reaction association?
7. What is the difference in the naming convention for reactions and metabolites?
8. Are there wildcard characters in the naming convention?
9. What is the purpose of the biomass function?
10. What growth phase should be used to determine biomass fractional contributions?
11. What are the units for growth?
12. What is the danger of using sink reactions?
13. What type of extreme pathways corresponds to internal loops?

References

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.
2. Monk, J., J. Nogales, et al. (2014). "Optimizing genome-scale network reconstructions." Nature biotechnology 32(5): 447-452
3. Reconstruction and Use of Microbial Metabolic Networks: the Core Escherichia coli Metabolic Model as an Educational Guide by Orth, Fleming, and Palsson (2010)
4. Price, N. D., I. Famili, et al. (2002). "Extreme pathways and Kirchhoff's second law." Biophysical journal 83(5): 2879-2882
5. Feist, A. M. and B. O. Palsson (2010). "The biomass objective function." Current opinion in microbiology 13(3): 344-349.
6. Kumar, V. S. and C. D. Maranas (2009). "GrowMatch: an automated method for reconciling in silico/in vivo growth predictions." PLoS computational biology 5(3): e1000308.
7. Gianchandani, E. P., M. A. Oberhardt, et al. (2008). "Predicting biological system objectives de novo from internal state measurements." BMC Bioinformatics 9: 43.
8. Schuetz, R., L. Kuepfer, et al. (2007). "Systematic evaluation of objective functions for predicting intracellular fluxes in Escherichia coli." Molecular Systems Biology 3: 119.
9. Izard, J. and R. J. Limberger (2003). "Rapid screening method for quantization of bacterial cell lipids from whole cells." Journal of microbiological methods 55(2): 411-418.
10. Burgard, A. P. and C. D. Maranas (2003). "Optimization-based framework for inferring and testing hypothesized metabolic objective functions." Biotechnology and bioengineering 82(6): 670-677.
11. Benthin, S., Nielsen, J. & Villadsen, J. A simple and reliable method for the determination of cellular RNA content. Biotechnol. Tech. 5, 39–42 (1991).
12. Herbert, D., Phipps, P.J. & Strange, R.E. Chemical analysis of microbial cells. Methods Microbiol. 5, 209–344 (1971).