Construction and Validation of Genome Scale Metabolic Models
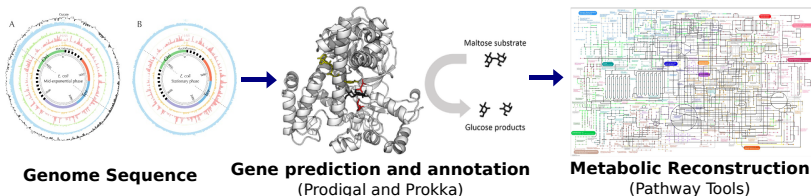
Dr.Dipali Singh

Quadram Institute Bioscience
Norwich Research Park
Norwich, UK

September 19, 2024

Genome-Scale Metabolic Model

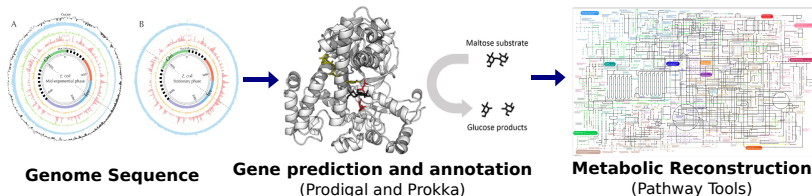Large size models: usually with 100s to 1000s reactions and metabolites

GSM describes the metabolic interactions in a given organism based on reaction network predicted from enzymes encoded by the genome

Genome Sequence   **Gene prediction and annotation**   **Metabolic Reconstruction**
(Prodigal and Prokka)   (Pathway Tools)

Genome-Scale Metabolic Model

Large size models: usually with 100s to 1000s reactions and metabolites

GSM describes the metabolic interactions in a given organism based on reaction network predicted from enzymes encoded by the genome

Genome Sequence

Gene prediction and annotation
(Prodigal and Prokka)

Metabolic Reconstruction
(Pathway Tools)

Easy !!

- Choose your favourite organism
- Go to your favourite data base
- Save the reactions in a suitable format
- Job done ($\approx$ 1 minute with a local db)

Unfortunately, NO

Easy !!

- Choose your favourite organism
- Go to your favourite data base
- Save the reactions in a suitable format
- Job done ($\approx$ 1 minute with a local db)

Unfortunately, NO

Easy !!

- Choose your favourite organism
- Go to your favourite data base
- Save the reactions in a suitable format
- Job done ($\approx$ 1 minute with a local db)

Unfortunately, NO

Any GSM reconstruction based solely upon data base is likely to exhibit a number of problems:

- inconsistent naming of metabolites and reactions identifiers
- incorrect reaction stoichiometries and reversibility
- missing gene-protein-reaction (GPR) associations

Therefore, in order to represent a realistic representation of a given organism, an initial draft model will require refinement and curation by the user

Any GSM reconstruction based solely upon data base is likely to exhibit a number of problems:

- inconsistent naming of metabolites and reactions identifiers
- incorrect reaction stoichiometries and reversibility
- missing gene-protein-reaction (GPR) associations

Therefore, in order to represent a realistic representation of a given organism, an initial draft model will require refinement and curation by the user

Constructing a genome scale model

Any GSM reconstruction based solely upon data base is likely to exhibit a number of problems:

- inconsistent naming of metabolites and reactions identifiers
- incorrect reaction stoichiometries and reversibility
- missing gene-protein-reaction (GPR) associations

Therefore, in order to represent a realistic representation of a given organism, an initial draft model will require refinement and curation by the user

Problems in construction

Non-specific metabolites e.g. :

- "Some-tRNA"
- "Long-Chain-Fatty-Acids"
- "An alcohol"

Inconsistent metabolites identifiers e.g. :

- multiple identifiers for same metabolite
- example of Ribose: D-Ribofuranose, D-Ribopyranose, CPD0-1108, CPD0-1110 etc
- artificially created network discontinuity

Incorrect stoichiometries e.g.

- "3.2.1.58−RXN": NOTHING −> NOTHING

Non-specific metabolites e.g. :

- "Some-tRNA"
- "Long-Chain-Fatty-Acids"
- "An alcohol"

Inconsistent metabolites identifiers e.g. :

- multiple identifiers for same metabolite
- example of Ribose: D-Ribofuranose, D-Ribopyranose, CPD0-1108, CPD0-1110 etc
- artificially created network discontinuity

Incorrect stoichiometries e.g.

- "3.2.1.58−RXN": NOTHING −> NOTHING

Problems in construction

Non-specific metabolites e.g. :

- "Some-tRNA"
- "Long-Chain-Fatty-Acids"
- "An alcohol"

Inconsistent metabolites identifiers e.g. :

- multiple identifiers for same metabolite
- example of Ribose: D-Ribofuranose, D-Ribopyranose, CPD0-1108, CPD0-1110 etc
- artificially created network discontinuity

Incorrect stoichiometries e.g.

- "3.2.1.58−RXN": NOTHING −> NOTHING

Polymeric species, consisting of an undefined number of
monomeric units, can give rise to mass inconsistancies e.g

- BioCyc database reports a starch synthesis reaction as:

    " ADP − D − GLUCOSE" −>"ADP" + "Starch"

- single glucose moiety is added to starch

- Amylase reaction as:

    " Starch" −>"MALTOPENTAOSE"

- with an overall conversion of 1 glucose moiety into five
glucose moieties

The problem with polymers

Polymeric species, consisting of an undefined number of monomeric units, can give rise to mass inconsistancies e.g

- BioCyc database reports a starch synthesis reaction as:

  " ADP−D−GLUCOSE" −>"ADP" + "Starch"

- single glucose moiety is added to starch

- Amylase reaction as:

  " Starch" −>"MALTOPENTAOSE"

- with an overall conversion of 1 glucose moiety into five glucose moieties

The problem with polymers

Polymeric species, consisting of an undefined number of monomeric units, can give rise to mass inconsistancies e.g

- BioCyc database reports a starch synthesis reaction as:
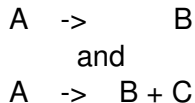
  " ADP−D−GLUCOSE" −>"ADP" + "Starch"

- single glucose moiety is added to starch

- Amylase reaction as:

  " Starch" −>"MALTOPENTAOSE"

- with an overall conversion of 1 glucose moiety into five glucose moieties

Clearly:

$$A \rightarrow B$$
and
$$A \rightarrow B + C$$

Cannot both be true.
(Violation of mass conservation)
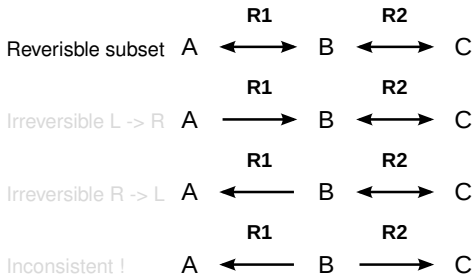
Even without empirical formulae such sets of reactions can be identified by a combination of:

- Analysis of left null-space
- Linear programming
- Provides an automatic method for identification of the polymer problem.
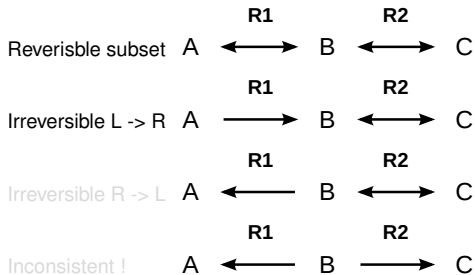- See: Gevorgyan *et al* 2008, Bioinformatics

Inconsistent subsets:

Inconsistent subsets:

Inconsistent subsets:
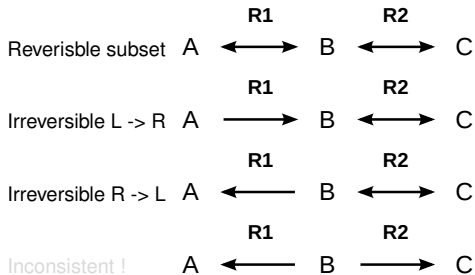
Inconsistent subsets:

		R1		**R2**	
Reverisble subset	A	← →	B	← →	C
Irreversible L -> R	A	→	B	← →	C
Irreversible R -> L	A	←	B	← →	C
Inconsistent !	A	←	B	→	C
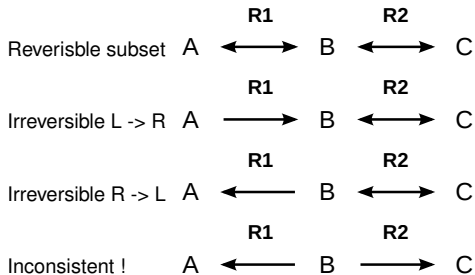
Violations of conservation of energy:

- Identify from LP:

  1. Constrain all transporters to zero flux.

  2. Set a demand for ATP and/or NAD(P)H.

  3. If the LP has a viable solution an inconsistency exists.

  4. All reactions in the solution must now be examined.

Mis-annotation:

- Reactions absent that should be present
- Example, Campylobacter model was not able to produce asparagine
- Due to missing reaction in the model
- Reactions present that should be absent
- Alternatively, the model was able to synthesis niacinamide, though Campylobacter is auxotrophic to niacinamide
- This was due to over-predicted reaction from Bioinformatice tools

Bioinformatic tools along with experimental data can be used for the curation

Mis-annotation:

- Reactions absent that should be present
- Example, Campylobacter model was not able to produce asparagine
- Due to missing reaction in the model

- Reactions present that should be absent
- Alternatively, the model was able to synthesis niacinamide, though Campylobacter is auxotrophic to niacinamide
- This was due to over-predicted reaction from Bioinformatice tools

Bioinformatic tools along with experimental data can be used for the curation

Mis-annotation:

- Reactions absent that should be present
- Example, Campylobacter model was not able to produce asparagine
- Due to missing reaction in the model
- Reactions present that should be absent
- Alternatively, the model was able to synthesis niacinamide, though Campylobacter is auxotrophic to niacinamide
- This was due to over-predicted reaction from Bioinformatice tools

Bioinformatic tools along with experimental data can be used for the curation

Problems in construction

Mis-annotation:

- Reactions absent that should be present
- Example, Campylobacter model was not able to produce asparagine
- Due to missing reaction in the model
- Reactions present that should be absent
- Alternatively, the model was able to synthesis niacinamide, though Campylobacter is auxotrophic to niacinamide
- This was due to over-predicted reaction from Bioinformatice tools

Bioinformatic tools along with experimental data can be used for the curation

Mis-annotation:

- Reactions absent that should be present
- Example, Campylobacter model was not able to produce asparagine
- Due to missing reaction in the model
- Reactions present that should be absent
- Alternatively, the model was able to synthesis niacinamide, though Campylobacter is auxotrophic to niacinamide
- This was due to over-predicted reaction from Bioinformatice tools

Bioinformatic tools along with experimental data can be used for the curation

Mis-annotation:

- Reactions absent that should be present
- Example, Campylobacter model was not able to produce asparagine
- Due to missing reaction in the model
- Reactions present that should be absent
- Alternatively, the model was able to synthesis niacinamide, though Campylobacter is auxotrophic to niacinamide
- This was due to over-predicted reaction from Bioinformatice tools

Bioinformatic tools along with experimental data can be used for the curation
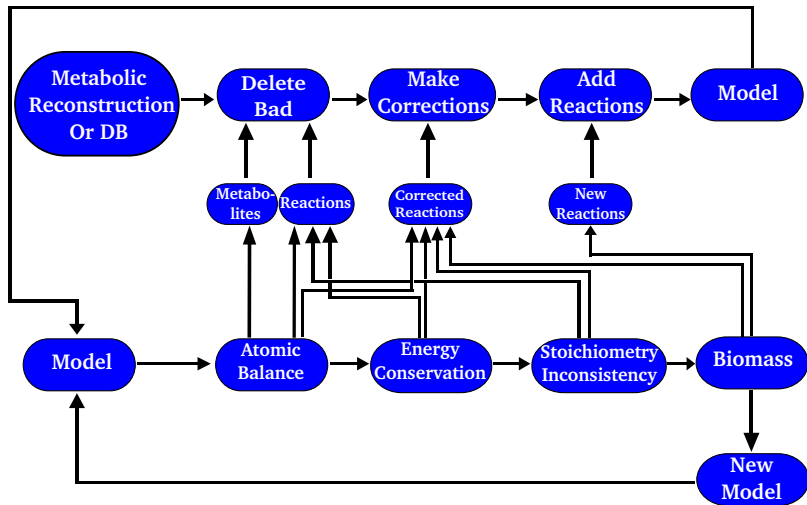
Problems in construction

Mis-annotation:

- Reactions absent that should be present
- Example, Campylobacter model was not able to produce asparagine
- Due to missing reaction in the model
- Reactions present that should be absent
- Alternatively, the model was able to synthesis niacinamide, though Campylobacter is auxotrophic to niacinamide
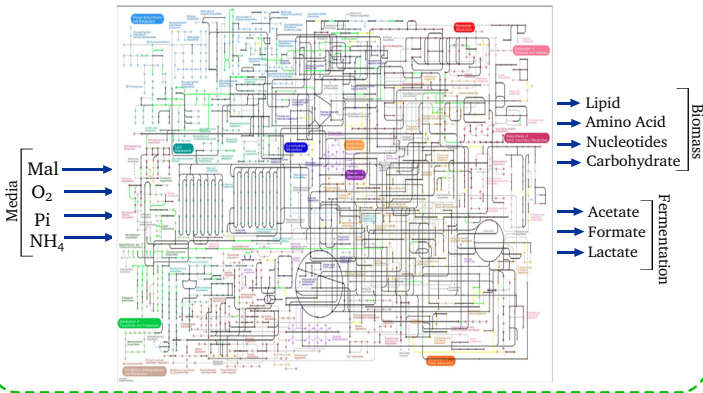- This was due to over-predicted reaction from Bioinformatice tools

Bioinformatic tools along with experimental data can be used for the curation

Genome-Scale Metabolic Model

For our convevience:

- Environmental/external metabolites are distinguished from internal metabolites by using prefix 'x_'

- Transport reactions exchange metabolites between the model and the environment.

    - Transport reactions are differentiated from other reactions by using suffix '_tx'

    - All transport reactions are defined with external species on the left side such that positive flux represents transport of metabolite into the system and negative flux represents transport of metabolites out of the system.

Naming Convention: Transporters

For our convevience:

- Environmental/external metabolites are distinguished from internal metabolites by using prefix 'x_'

- Transport reactions exchange metabolites between the model and the environment.

    - Transport reactions are differentiated from other reactions by using suffix '_tx'

        - Biomass transporters: '_bm_tx'
        - Media transporters: '_mm_tx'

    - All transport reactions are defined with external species on the left side such that positive flux represents transport of metabolite into the system and negative flux represents transport of metabolites out of the system.

Naming Convention: Transporters

For our convevience:

- Environmental/external metabolites are distinguished from internal metabolites by using prefix 'x_'

- Transport reactions exchange metabolites between the model and the environment.

    - Transport reactions are differentiated from other reactions by using suffix '_tx'

        - Biomass transporters: '_bm_tx'
        - Media transporters: '_mm_tx'

    - All transport reactions are defined with external species on the left side such that positive flux represents transport of metabolite into the system and negative flux represents transport of metabolites out of the system.

Naming Convention: Transporters

For our convevience:

- Environmental/external metabolites are distinguished from internal metabolites by using prefix 'x_'

- Transport reactions exchange metabolites between the model and the environment.

    - Transport reactions are differentiated from other reactions by using suffix '_tx'

        - Biomass transporters: '_bm_tx'
        - Media transporters: '_mm_tx'

    - All transport reactions are defined with external species on the left side such that positive flux represents transport of metabolite into the system and negative flux represents transport of metabolites out of the system.

Naming Convention: Transporters

For our convevience:

- Environmental/external metabolites are distinguished from internal metabolites by using prefix 'x_'

- Transport reactions exchange metabolites between the model and the environment.

    - Transport reactions are differentiated from other reactions by using suffix '_tx'

        - Biomass transporters: '_bm_tx'
        - Media transporters: '_mm_tx'

    - All transport reactions are defined with external species on the left side such that positive flux represents transport of metabolite into the system and negative flux represents transport of metabolites out of the system.

Naming Convention: Transporters

For our convevience:

- Environmental/external metabolites are distinguished from internal metabolites by using prefix 'x_'

- Transport reactions exchange metabolites between the model and the environment.

  - Transport reactions are differentiated from other reactions by using suffix '_tx'

    - Biomass transporters: '_bm_tx'
    - Media transporters: '_mm_tx'

  - All transport reactions are defined with external species on the left side such that positive flux represents transport of metabolite into the system and negative flux represents transport of metabolites out of the system.

Reactions between compartments can be differentiated using suffix. e.g. eukaryotic models

- Reactions in cytosol: '_Cyto'

- Reactions in mitochondria: '_Mito'

- Reactions in mitochondria: '_Plas'

- Constructing GSMs from a database is easy

- Constructing meaningful GSMs from databases require rigorous and methodical curation

- Lastly, not all the published models are correct. Do your quality checks before using them!

- Constructing GSMs from a database is easy

- Constructing meaningful GSMs from databases require rigorous and methodical curation

- Lastly, not all the published models are correct. Do your quality checks before using them!

Constructing models - Summary

- Constructing GSMs from a database is easy

- Constructing meaningful GSMs from databases require rigorous and methodical curation

- Lastly, not all the published models are correct. Do your quality checks before using them!