Construction of Genome Scale Models

Mark Poolman

January 17, 2018

Wednesday L7

Contrast with small models

Small Models	Genome-scale Models
Defined	Undefined (1) We don't define the contents.
Defined	Undefined (2) We don't know the contents.
Easy to analsyse	Hard to analyse The tech. won't always scale.
Easy to analyse	Hard to analyse Our understanding won't always scale.

Contrast with small models

Small Models	Genome-scale Models
Defined	Undefined (1) We don't define the contents.
Defined	Undefined (2) We don't know the contents.
Easy to analsyse	Hard to analyse The tech. won't always scale.
Easy to analyse	Hard to analyse Our understanding won't always scale.

Contrast with small models

Small Models	Genome-scale Models
Defined	Undefined (1) We don't define the contents.
Defined	Undefined (2) We don't know the contents.
Easy to analsyse	Hard to analyse The tech. won't always scale.
Easy to analyse	Hard to analyse Our understanding won't always scale.

Contrast with small models

Small Models	Genome-scale Models
Defined	Undefined (1) We don't define the contents.
Defined	Undefined (2) We don't know the contents.
Easy to analsyse	Hard to analyse The tech. won't always scale.
Easy to analyse	Hard to analyse Our understanding won't always scale.

Contrast with small models

Small Models	Genome-scale Models
Defined	Undefined (1) We don't define the contents.
Defined	Undefined (2) We don't know the contents.
Easy to analsyse	Hard to analyse The tech. won't always scale.
Easy to analyse	Hard to analyse Our understanding won't always scale.

Constructing a genome scale model

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

Constructing a genome scale model

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

Constructing a genome scale model

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

Problems in construction

1) Mis-annotation:

- Reactions absent that should be present.
- Reactions present that should be absent.

Little can be done from the structural modelling perspective.

Primarily a bioinformatics problem .

Problems in construction

1) Mis-annotation:

- Reactions absent that should be present.
- Reactions present that should be absent.

Little can be done from the structural modelling perspective.

Primarily a bioinformatics problem .

Problems in construction

2) Non-specific metabolites e.g. :

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

Problems in construction

3) Incorrect stoichiometries e.g.

- “3.2.1.58−RXN”: NOTHING −> NOTHING
- “3.6.3.4−RXN’ : "CU+2" + "WATER" + "ATP" −> "CU+2" + "|Pi|" + "ADP"
- "UROPORIIIMETHYLTRANSA−RXN":
- "UROPORPHYRINOGEN−III" + 2 "S−ADENOSYLMETHIONINE"
- −> 2 "ADENOSYL−HOMO−CYS" + "CPD−9038"

Problems in construction

3) Incorrect stoichiometries e.g.

- “3.2.1.58−RXN”: NOTHING −> NOTHING
- “3.6.3.4−RXN’ : "CU+2" + "WATER" + "ATP" −> "CU+2" + "|Pi|" + "ADP"
- "UROPORIIIMETHYLTRANSA−RXN":
- "UROPORPHYRINOGEN−III" + 2 "S−ADENOSYLMETHIONINE"
- −> 2 "ADENOSYL−HOMO−CYS" + "CPD−9038"

Problems in construction

3) Incorrect stoichiometries e.g.

- ‘‘3.2.1.58−RXN’’: NOTHING −> NOTHING
- ‘‘3.6.3.4−RXN’ : "CU+2" + "WATER" + "ATP" −> "CU+2" + "|Pi|" + "ADP"
- "UROPORIIIMETHYLTRANSA−RXN":
- "UROPORPHYRINOGEN−III" + 2 "S−ADENOSYLMETHIONINE"
- −> 2 "ADENOSYL−HOMO−CYS" + "CPD−9038"

Problems in construction

3) Incorrect stoichiometries e.g.

- ‘‘3.2.1.58−RXN’’: NOTHING −> NOTHING
- ‘‘3.6.3.4−RXN’ : "CU+2" + "WATER" + "ATP" −> "CU+2" + "|Pi|" + "ADP"
- "UROPORIIIMETHYLTRANSA−RXN":
- "UROPORPHYRINOGEN−III" + 2 "S−ADENOSYLMETHIONINE"
- −> 2 "ADENOSYL−HOMO−CYS" + "CPD−9038"

The problem with polymers 1

Consider:

- " RXN−1827":

  "1−4−alpha−D−Glucan" + "WATER" −>"ALPHA−MALTOSE" + "1−4−alpha−D−Glucan"

  - Is the real stoichiometry ?
  - " WATER" −> "ALPHA−MALTOSE"
  - " WATER" −> "ALPHA−MALTOSE" + "1−4−alpha−D−Glucan"
  - "1−4−alpha−D−Glucan" +"WATER" −> "ALPHA−MALTOSE"

Note: Only a human can tell which is correct, the computer can't

The problem with polymers 1

Consider:

- “ RXN−1827”:

  “1−4−alpha−D−Glucan” + “WATER” −>“ALPHA−MALTOSE” + “1−4−alpha−D−Glucan”

  - Is the real stoichiometry ?
  - “ WATER” −> “ALPHA−MALTOSE”
  - “ WATER” −> “ALPHA−MALTOSE” + “1−4−alpha−D−Glucan”
  - “1−4−alpha−D−Glucan” +“WATER” −> “ALPHA−MALTOSE”

Note: Only a human can tell which is correct, the computer can’t

The problem with polymers 1

Consider:

- “RXN−1827”:

  “1−4−alpha−D−Glucan” + “WATER” −>“ALPHA−MALTOSE” + “1−4−alpha−D−Glucan”

  - Is the real stoichiometry ?
  - “WATER” −> “ALPHA−MALTOSE”
  - “WATER” −> “ALPHA−MALTOSE” + “1−4−alpha−D−Glucan”
  - “1−4−alpha−D−Glucan” +“WATER” −> “ALPHA−MALTOSE”

Note: Only a human can tell which is correct, the computer can’t

The problem with polymers 2

Now consider:

- “GLYCOGENSYN−RXN”:

  “1−4−alpha−D−Glucan” + “ADP−D−GLUCOSE” −> “ADP” + “1−4−alpha−D−Glucan”

- **Correct stoichiometry is:**

  “ADP−D−GLUCOSE” −> “ADP” + “1−4−alpha−D−Glucan”

- **But combining this with the previous reaction results in a net stoichiometry:**

  “ADP−D−GLUCOSE” −> “ADP” + “ALPHA−MALTOSE”

Creating six carbon atoms from nothing !!

The problem with polymers 2

Now consider:

- “GLYCOGENSYN−RXN”:

  “1−4−alpha−D−Glucan” + “ADP−D−GLUCOSE” −> “ADP” + “1−4−alpha−D−Glucan”

- **Correct stoichiometry is:**

  “ADP−D−GLUCOSE” −> “ADP” + “1−4−alpha−D−Glucan”

- **But combining this with the previous reaction results in a net stoichiometry:**

  “ADP−D−GLUCOSE” −> “ADP” + “ALPHA−MALTOSE”

Creating six carbon atoms from nothing !!

The problem with polymers 2

Now consider:

- “ GLYCOGENSYN−RXN”:

  “1−4−alpha−D−Glucan” + “ADP−D−GLUCOSE” −> “ADP” + “1−4−alpha−D−Glucan”
- **Correct stoichiometry is:**

  “ ADP−D−GLUCOSE” −> “ADP” + “1−4−alpha−D−Glucan”
- **But combining this with the previous reaction results in a net stoichiometry:**

  “ ADP−D−GLUCOSE” −> “ADP” + “ALPHA−MALTOSE”

Creating six carbon atoms from nothing !!

The problem with polymers 2

Now consider:

- “ GLYCOGENSYN−RXN”:

  “1−4−alpha−D−Glucan” + “ADP−D−GLUCOSE” −> “ADP” + “1−4−alpha−D−Glucan”
- **Correct stoichiometry is:**

  “ ADP−D−GLUCOSE” −> “ADP” + “1−4−alpha−D−Glucan”
- **But combining this with the previous reaction results in a net stoichiometry:**

  “ ADP−D−GLUCOSE” −> “ADP” + “ALPHA−MALTOSE”

Creating six carbon atoms from nothing !!

The problem with polymers - a solution

For homopolymers rewrite reaction in terms of monomeric units, e.g. Glc for 1-4-alpha-D-Glucan:

- "GLYCOGENSYN−RXN": "ADP−D−GLUCOSE" −> "ADP" +"1−4−alpha−D−Glucan"
- "RXN−1827": "1−4−alpha−D−Glucan" −> 1/2 "ALPHA−MALTOSE"

Problem is harder for heteropolymers

The problem with polymers - a solution

For homopolymers rewrite reaction in terms of monomeric units, e.g. Glc for 1-4-alpha-D-Glucan:

- “GLYCOGENSYN−RXN”: “ADP−D−GLUCOSE” −> “ADP” +“1−4−alpha−D−Glucan”
- “RXN−1827”: “1−4−alpha−D−Glucan” −> 1/2 “ALPHA−MALTOSE”

Problem is harder for heteropolymers

The problem with polymers - a solution

For homopolymers rewrite reaction in terms of monomeric units, e.g. Glc for 1-4-alpha-D-Glucan:

- “GLYCOGENSYN−RXN”: “ADP−D−GLUCOSE” −> “ADP” +“1−4−alpha−D−Glucan”
- “RXN−1827”: “1−4−alpha−D−Glucan” −> 1/2 “ALPHA−MALTOSE”

Problem is harder for heteropolymers

The problem with polymers - a solution

For homopolymers rewrite reaction in terms of monomeric units, e.g. Glc for 1-4-alpha-D-Glucan:

- “GLYCOGENSYN−RXN”: “ADP−D−GLUCOSE” −> “ADP” +“1−4−alpha−D−Glucan”
- “RXN−1827”: “1−4−alpha−D−Glucan” −> 1/2 “ALPHA−MALTOSE”

Problem is harder for heteropolymers

Stoichiometric inconsistencies

Clearly:

A -> B

and

A -> B + C

Cannot both be true.
(Violation of mass conservation)

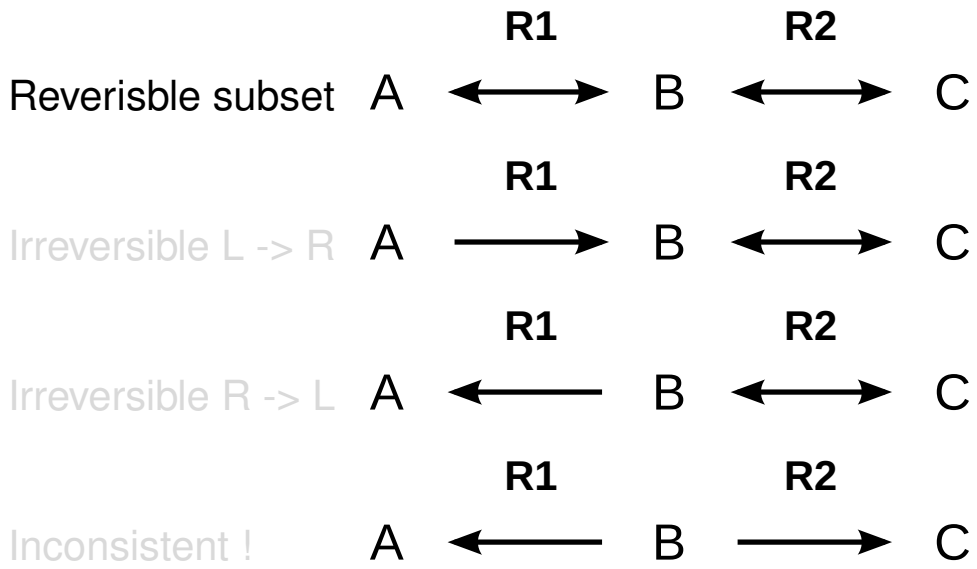
Stoichiometric inconsistencies

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

Reaction irreversibility/thermodynamics

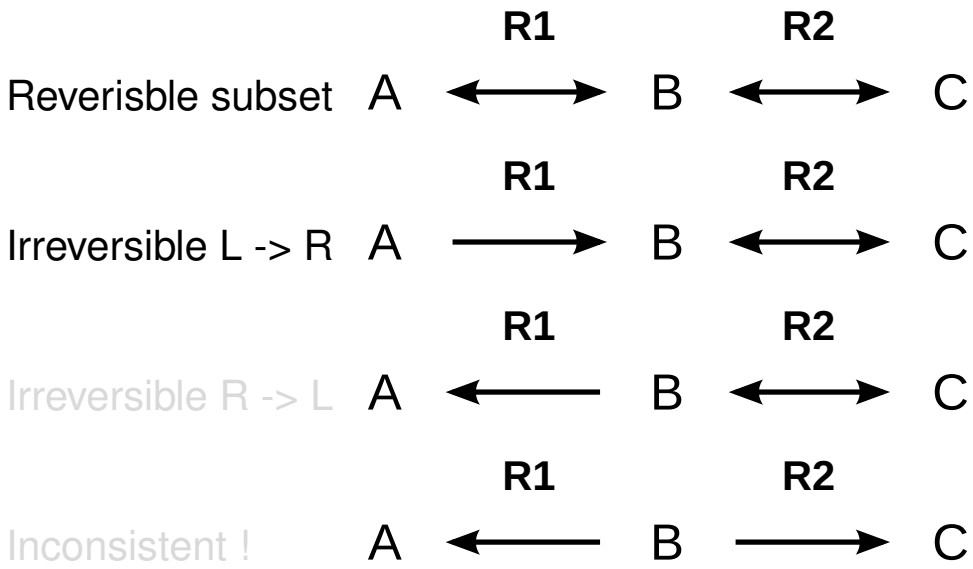
Inconsistent subsets:



Identify from null-space in conjunction with a knowledge of reaction (ir)reversibilities.

Reaction irreversibility/thermodynamics

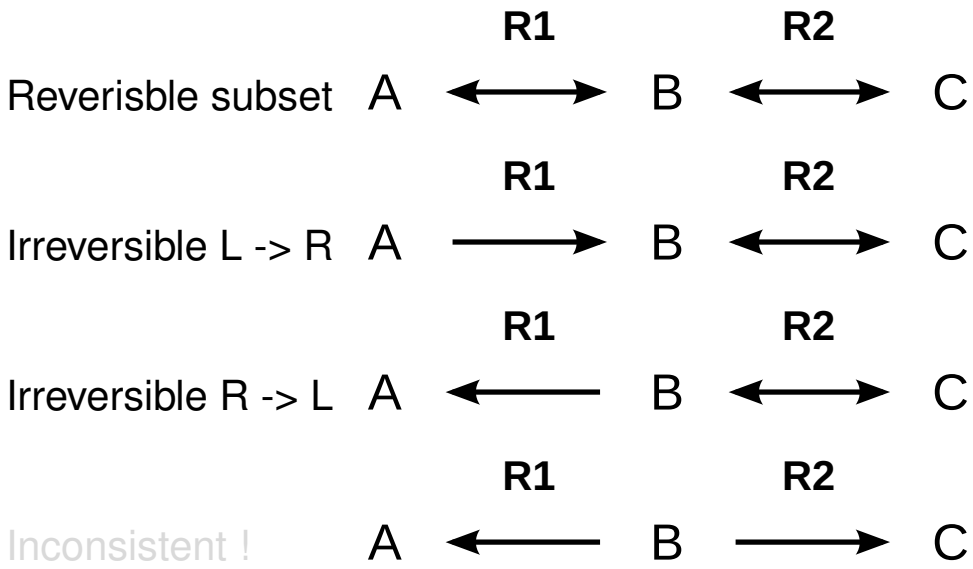
Inconsistent subsets:



Identify from null-space in conjunction with a knowledge of reaction (ir)reversibilities.

Reaction irreversibility/thermodynamics

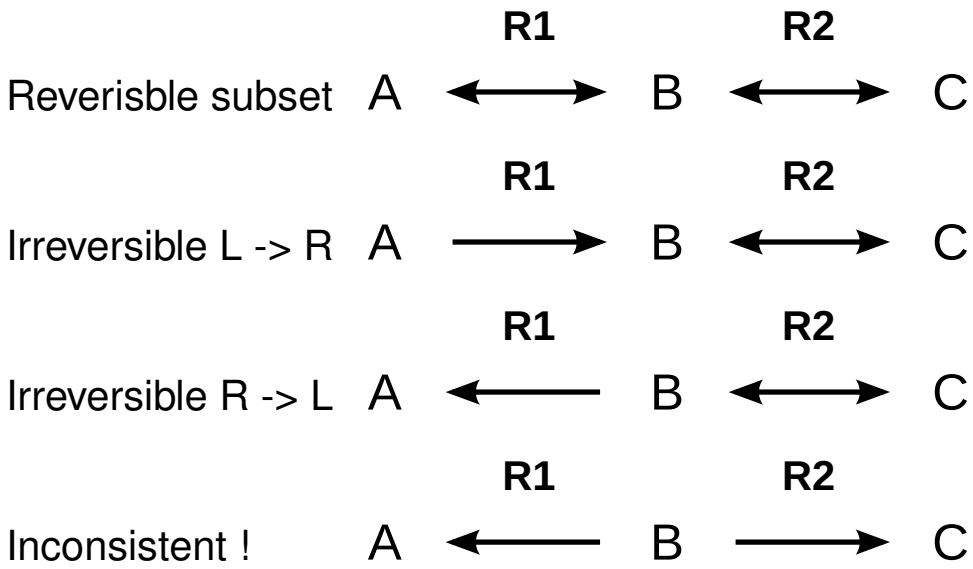
Inconsistent subsets:



Identify from null-space in conjunction with a knowledge of reaction (ir)reversibilities.

Reaction irreversibility/thermodynamics

Inconsistent subsets:



Identify from null-space in conjunction with a knowledge of reaction (ir)reversibilities.

Reaction irreversibility/thermodynamics

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.

Constructing models - Summary

Constructing GSMs from a database is easy.

Constructing meaningful GSMs from databases is hard.

Requires rigorous and methodical QC

Technicalities - Modularised models

Components:

- From the DB
- Defined transporters.
- Additional reactions of interest.

ScrumPy provides the `Include()` directive to accomplish this:

```
External("PROTON","WATER")

Include(AutoGeobac.spy,
        Transporters.spy,
        Biomass.spy,
        ExtraReacs.spy,
        ETC_core.spy
)
```

TODO: Technicalities - Naming conventions

Be consistent !

- Only use quoted identifiers for items in the DB
- Use “_tx” suffix to denote a transporter.
- Name transporters according to what they transport.
- Use _bm_tx suffix to denote biomass export.

TODO: Technicalities - Transport fluxes

Be consistent !

- Positive flux *always* denotes gain.
- Negative flux *always* denotes loss.