Construction of Genome Scale Models

Mark Poolman

Oxford Brookes University

October 17, 2012

Contrast with small models

Small	Genome Scale

Contrast with small models

Small	**Genome Scale**
Defined	Undefined (we can't specify what's in it)

Contrast with small models

Small	**Genome Scale**
Defined	Undefined (we can't specify what's in it)
Defined	Undefined (we can't know what's in it)

Contrast with small models

Small	**Genome Scale**
Defined	Undefined (we can't specify what's in it)
Defined	Undefined (we can't know what's in it)
Comprehensible	Incomprehensible

Contrast with small models

Small	**Genome Scale**
Defined	Undefined (we can't specify what's in it)
Defined	Undefined (we can't know what's in it)
Comprehensible	Incomprehensible
Easy to analyse	Hard to analyse

Contrast with small models

Small	**Genome Scale**
Defined	Undefined (we can't specify what's in it)
Defined	Undefined (we can't know what's in it)
Comprehensible	Incomprehensible
Easy to analyse	Hard to analyse

Lack of scaling (techniques and tech.)

Contrast with small models

Small	**Genome Scale**
Defined	Undefined (we can't specify what's in it)
Defined	Undefined (we can't know what's in it)
Comprehensible	Incomprehensible
Easy to analyse	Hard to analyse

Lack of scaling (techniques and tech.)

Questions and answers less obvious.

Construct your model

Construct your model

- Access a data-base for your organism.

Construct your model

- Access a data-base for your organism.
- Extract all the reactions.

Construct your model

- Access a data-base for your organism.
- Extract all the reactions.
- Save them in a suitable format.

Construct your model

- Access a data-base for your organism.
- Extract all the reactions.
- Save them in a suitable format.
- Job Done ($\approx$ 1 minute with a local DB)

Construct your model

- Access a data-base for your organism.
- Extract all the reactions.
- Save them in a suitable format.
- Job Done ($\approx$ 1 minute with a local DB)
- Simples !

Problems in construction

1) Mis-annotation
Little can be done to identify/solve these problems, primarily a bioinformatics problem.

Problems in construction

2) Non-specific metabolites e.g. :

- “Some-tRNA”

Problems in construction

2) Non-specific metabolites e.g. :

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

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
  ~

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"
  ~

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"
  ~
- (unbalanced C)

The problem with polymers 1

Consider:

- "RXN-1827":
  "1-4-alpha-D-Glucan" + "WATER" -> "ALPHA-MALTOSE" + "1-4-alpha-D-Glucan"
  ~

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 ?

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"

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"

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"

The problem with polymers 2

Now consider:

- "GLYCOGENSYN-RXN":
  "1-4-alpha-D-Glucan" + "ADP-D-GLUCOSE" -> "ADP" + "1-4-alpha-D-Glucan"
  ~

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:

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"

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:

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"`

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:

- `2 “ADP-D-GLUCOSE” -> 2 “ADP” + “1-4-alpha-D-Glucan”`

The problem with polymers - a solution

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

- `2 “ADP-D-GLUCOSE” -> 2 “ADP” + “1-4-alpha-D-Glucan”`
- Problem is worse for heteropolymers.

Stoichiometric inconsistencies

Clearly:

- `A -> B`

Stoichiometric inconsistencies

Clearly:

- `A -> B`
- and

Stoichiometric inconsistencies

Clearly:

- `A -> B`
- and
- `A -> B + C`

Stoichiometric inconsistencies

Clearly:

- `A -> B`
- and
- `A -> B + C`
- Cannot both be true.

Stoichiometric inconsistencies

Clearly:

- `A -> B`
- and
- `A -> B + C`
- Cannot both be true.
- Violate mass conservation.

Stoichiometric inconsistencies

Even without empirical formulae such sets of reactions can be identified by a combination of:

- Analysis of left null-space

Stoichiometric inconsistencies

Even without empirical formulae such sets of reactions can be identified by a combination of:

- Analysis of left null-space
- Linear programming

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.

Stoichiometric inconsistencies

Even without empirical formulae such sets of reactions can be identified by a combination of:

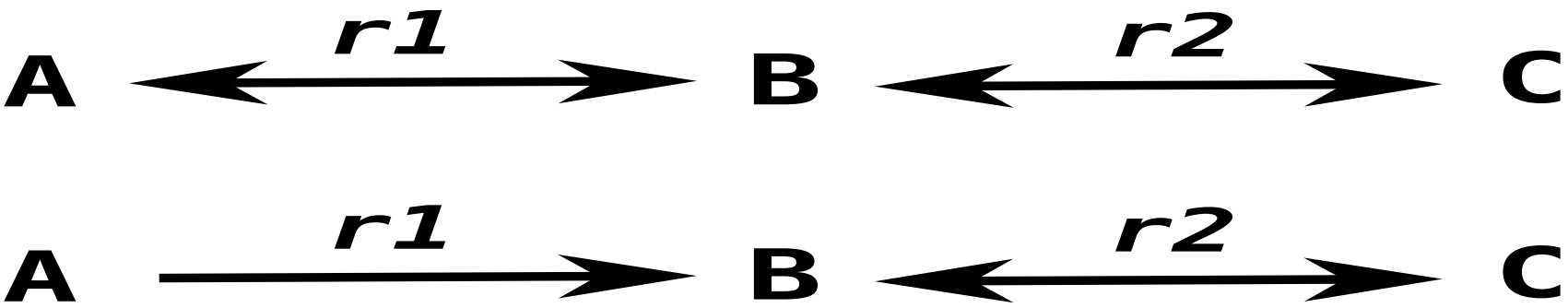
- Analysis of left null-space
- Linear programming
- Provides an automatic method for identification of the polymer problem.
- See: Gevorgyan *et al* 2008, Bioinformatics

Thermodynamic inconsistencies

Remember Enzyme (or reaction) subsets. Consider:

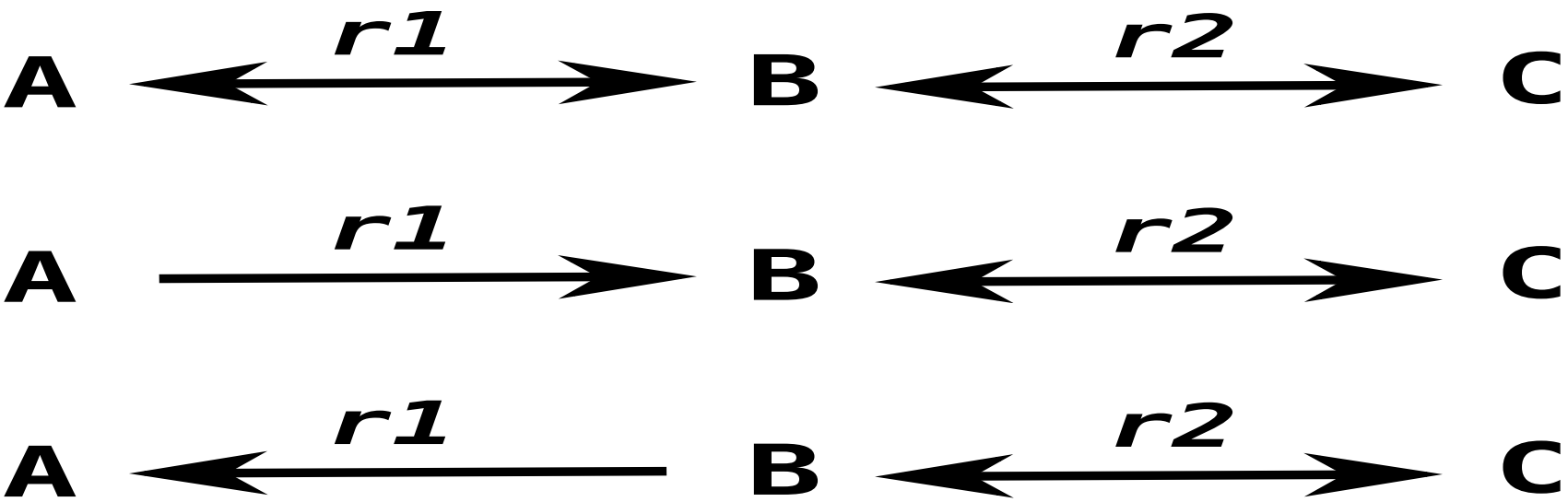
Thermodynamic inconsistencies

Remember Enzyme (or reaction) subsets. Consider:



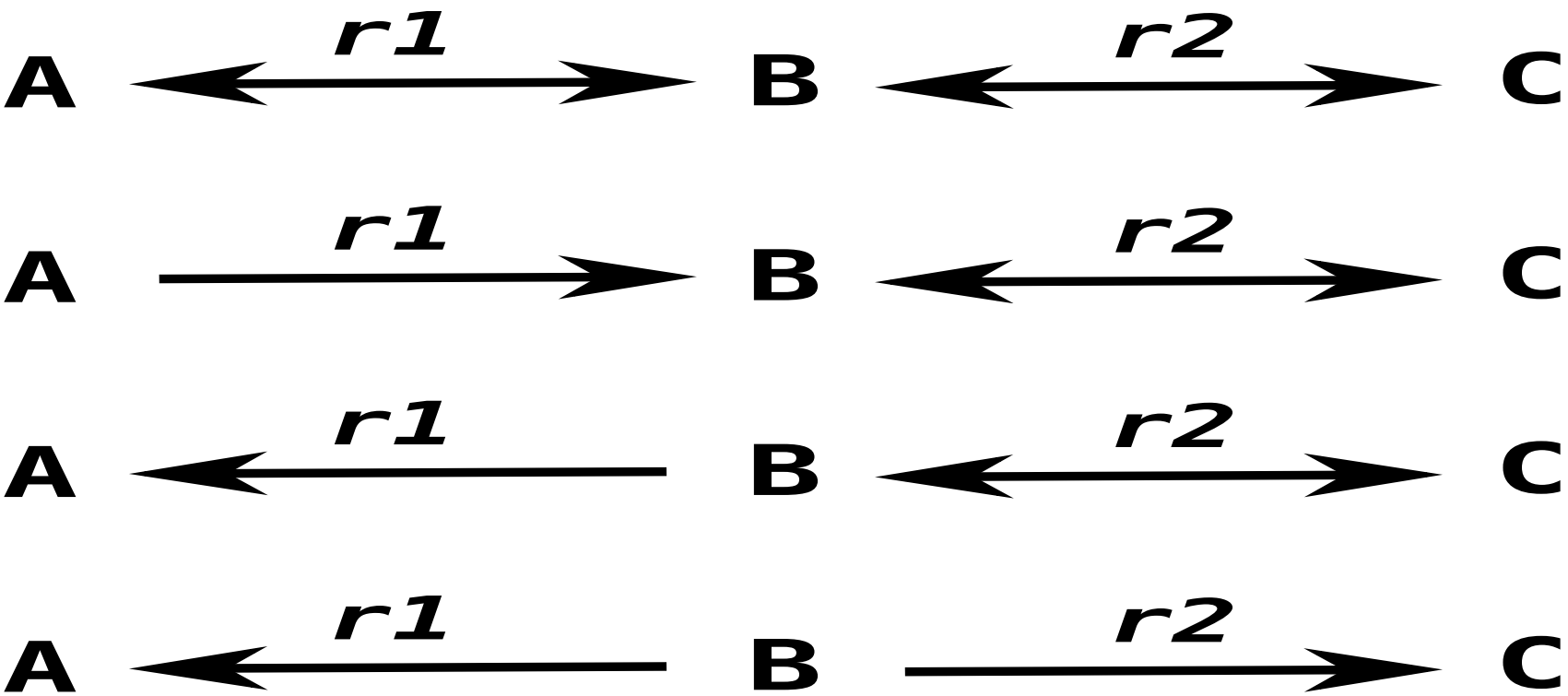
Thermodynamic inconsistencies

Remember Enzyme (or reaction) subsets. Consider:



Thermodynamic inconsistencies

Remember Enzyme (or reaction) subsets. Consider:



Reaction irreversibility/thermodynamics

Must be considered, wait for next talk.

Additional challenges with eukaryotes

Major complication is compartmentation. Databases are very sparse regarding:

- Enzyme targeting.

Additional challenges with eukaryotes

Major complication is compartmentation. Databases are very sparse regarding:

- Enzyme targeting.
- Transporters.

Additional challenges with eukaryotes

Major complication is compartmentation. Databases are very sparse regarding:

- Enzyme targeting.
- Transporters.
- Reaction/metabolite location.

Construct your model

- Access a data-base for your organism.

Construct your model

- Access a data-base for your organism.
- Extract all the reactions.

Construct your model

- Access a data-base for your organism.
- Extract all the reactions.
- Save them in a suitable format.

Construct your model

- Access a data-base for your organism.
- Extract all the reactions.
- Save them in a suitable format.
- Job Done ($\approx$ 1 minute with a local DB)

Construct your model

- Access a data-base for your organism.
- Extract all the reactions.
- Save them in a suitable format.
- Job Done ($\approx$ 1 minute with a local DB)
- Spend six months fixing it !