

CS-E5885 Modeling biological networks

Chemical reaction network models

Harri Lähdesmäki

Department of Computer Science
Aalto University

January 11, 2022

Contents

- ▶ Introduction to modeling biological networks
- ▶ Chemical reaction network models
 - ▶ Mathematical/Matrix formalism
- ▶ Stoichiometry
- ▶ Conservation laws
- ▶ Reachability
- ▶ Reading (see references at the end):
 - ▶ Chapters 1 and 2 from (Wilkinson, 2011)

Why?

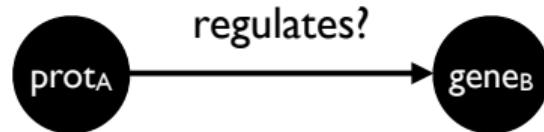
- ▶ Computational systems biology aims to provide a system-level understanding of a biological system
- ▶ Understand the molecular-level mechanisms underlying biological systems using a combination of experimental and computational techniques

Why?

- ▶ Computational systems biology aims to provide a system-level understanding of a biological system
 - ▶ Understand the molecular-level mechanisms underlying biological systems using a combination of experimental and computational techniques
1. Formulate biological networks in mathematical terms
 2. Measure biological systems/networks at molecular-level
 3. Reconstruct biological network models (or recalibrate their parameters) from experimental data and/or database information
 4. Predict computationally how a network behaves under novel experimental conditions
 5. Test the computational predictions experimentally
 6. Understand which network mechanisms are altered in a disease
 7. Design drugs/external conditions/synthetic “circuits” or molecules/etc. to alter biological networks
 8. ...

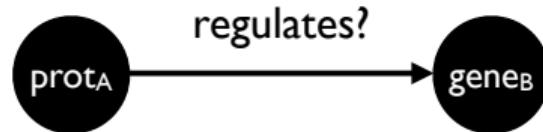
Why?

- In the past, traditional experimental techniques have focused on individual interactions, e.g.



Why?

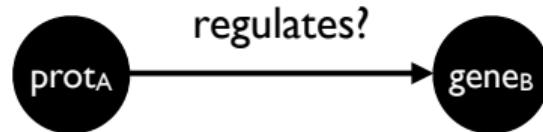
- ▶ In the past, traditional experimental techniques have focused on individual interactions, e.g.



- ▶ Modern molecular biology experiments can use highly parallel (high-throughput) measurements to quantify “all” interactions at once
 - ▶ For example, a single ChIP-seq measurement can reveal interactions for a protein A and all genes B_i , $i = 1, \dots, N$ (e.g. for human $N = 20000$)

Why?

- In the past, traditional experimental techniques have focused on individual interactions, e.g.



- Modern molecular biology experiments can use highly parallel (high-throughput) measurements to quantify “all” interactions at once
 - For example, a single ChIP-seq measurement can reveal interactions for a protein A and all genes B_i , $i = 1, \dots, N$ (e.g. for human $N = 20000$)
- Here the goal is to understand how a biological system as a whole functions over time (holistic approach)

Why modeling?

- ▶ Mathematical representation of the current knowledge of a biological system/network
 - ▶ Elements: genes, transcripts, proteins, metabolites, post-translational modifications, drugs, other small molecules, ...
 - ▶ Interactions: protein-protein, protein-DNA, protein-RNA, RNA-RNA, protein-drug, etc.
 - ▶ Dynamics

Why modeling?

- ▶ Mathematical representation of the current knowledge of a biological system/network
 - ▶ Elements: genes, transcripts, proteins, metabolites, post-translational modifications, drugs, other small molecules, ...
 - ▶ Interactions: protein-protein, protein-DNA, protein-RNA, RNA-RNA, protein-drug, etc.
 - ▶ Dynamics
- ▶ Test current knowledge
 - ▶ Compare the current state of knowledge with experimental data

Why modeling?

- ▶ Mathematical representation of the current knowledge of a biological system/network
 - ▶ Elements: genes, transcripts, proteins, metabolites, post-translational modifications, drugs, other small molecules, ...
 - ▶ Interactions: protein-protein, protein-DNA, protein-RNA, RNA-RNA, protein-drug, etc.
 - ▶ Dynamics
- ▶ Test current knowledge
 - ▶ Compare the current state of knowledge with experimental data
- ▶ Learn network models from experimental data
 - ▶ Different “modeling” aims:
 - ▶ Mapping large-scale molecular networks: interactions
 - ▶ Modeling dynamics of biological networks
 - ▶ Structure selection and parameter inference

Why modeling?

- ▶ Mathematical representation of the current knowledge of a biological system/network
 - ▶ Elements: genes, transcripts, proteins, metabolites, post-translational modifications, drugs, other small molecules, ...
 - ▶ Interactions: protein-protein, protein-DNA, protein-RNA, RNA-RNA, protein-drug, etc.
 - ▶ Dynamics
- ▶ Test current knowledge
 - ▶ Compare the current state of knowledge with experimental data
- ▶ Learn network models from experimental data
 - ▶ Different “modeling” aims:
 - ▶ Mapping large-scale molecular networks: interactions
 - ▶ Modeling dynamics of biological networks
 - ▶ Structure selection and parameter inference
- ▶ Use models predictively
 - ▶ Predictive behavior / “virtual experiments”

Stochastic modeling

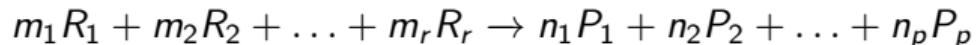
- ▶ Biological systems are sometimes (mistakenly) considered as deterministic systems
- ▶ Dynamics of biological systems at the level of individual molecules are intrinsically stochastic in nature
- ▶ Biological measurements are also noisy, often considerably so
- ▶ Statistics and stochastic modeling play an important role in modeling biological systems

Mathematical formalisms for biological systems

- ▶ A biological system can be represented in a number of ways:
 - ▶ Verbal description
 - ▶ Diagrams, graphs
 - ▶ Interaction and dependency graphs
 - ▶ Regression models
 - ▶ Probabilistic graphical models
 - ▶ Ordinary/Stochastic/Partial differential equations (ODEs/SDEs/PDEs)
 - ▶ Coupled chemical reactions

Chemical reactions

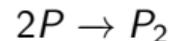
- ▶ Chemical reactions define a full, stochastic dynamic model of a biophysical system
- ▶ A chemical reaction with stoichiometric coefficients



- ▶ The chemical species which react: R_1, \dots, R_r
- ▶ Their proportions: m_1, \dots, m_r
- ▶ The chemical species that are produced: P_1, \dots, P_p
- ▶ Their proportions n_1, \dots, n_p
- ▶ A system of coupled chemical reactions consists of a set of chemical reactions

An example: protein dimerisation

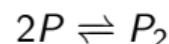
- ▶ Protein dimerisation



- ▶ Dissociation



- ▶ A reaction that can happen in both directions is called reversible and denoted as



Examples of biological systems/networks

- ▶ *Transcriptional networks
- ▶ Epigenetic mechanisms
- ▶ *Signaling / Protein-protein interaction networks
- ▶ *Metabolic networks
- ▶ Neuronal synaptic networks
- ▶ Ecological networks
- ▶ *Epidemiological networks
- ▶ ...
- ▶ For each biological system, one needs to choose an appropriate level of modeling details

Transcriptional networks: A simplistic view

- ▶ Transcription of genetic material into RNA molecules
 - ▶ A process of making mRNA copies of protein-coding genes and various other non-coding RNA elements

Transcriptional networks: A simplistic view

- ▶ Transcription of genetic material into RNA molecules
 - ▶ A process of making mRNA copies of protein-coding genes and various other non-coding RNA elements
- ▶ A simplistic view: transcription is driven largely by transcription factor (TF) proteins

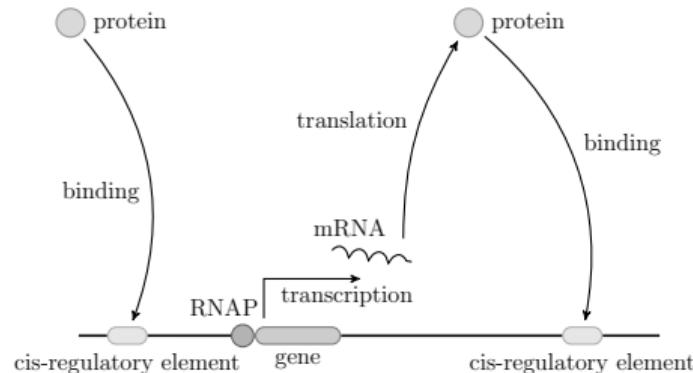


Figure: From (Äijö, 2009)

Transcriptional networks: A simplistic view

- Transcriptional regulators (genes/proteins) form cascades/networks

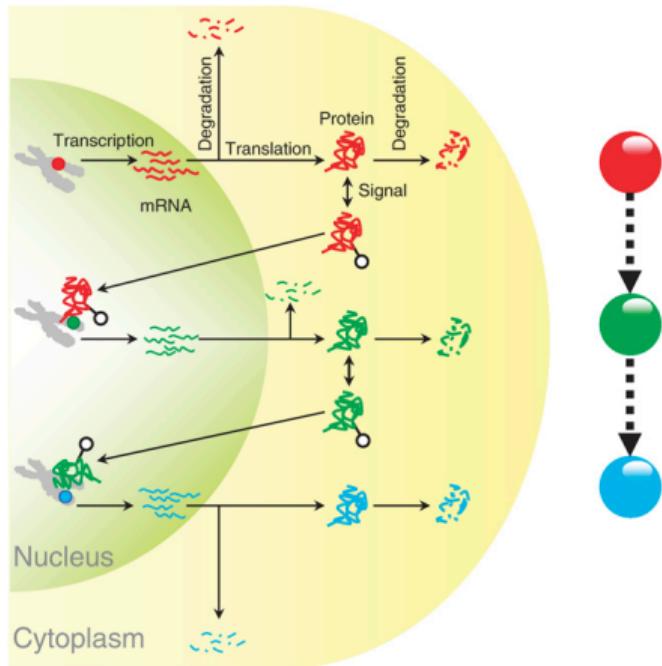


Figure: From (Jothi et al., 2009)

Transcriptional networks: A more detailed view; initiation and elongation

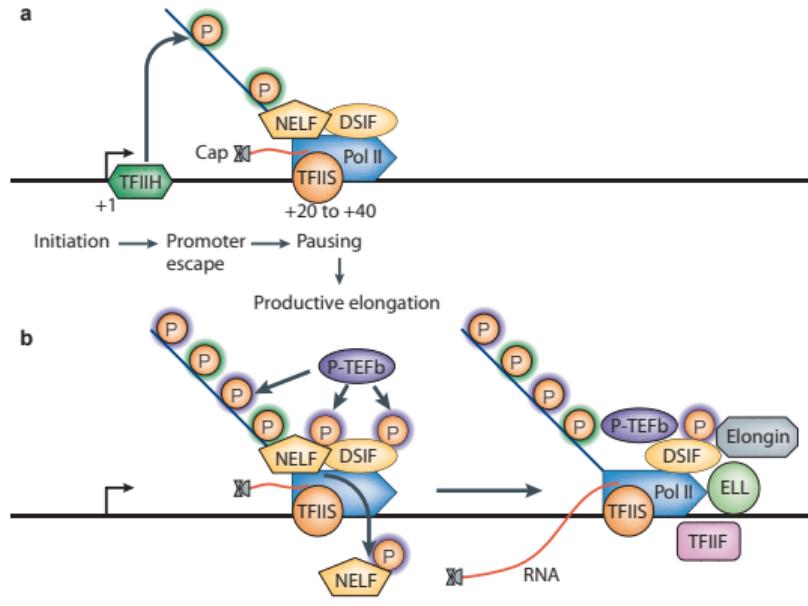
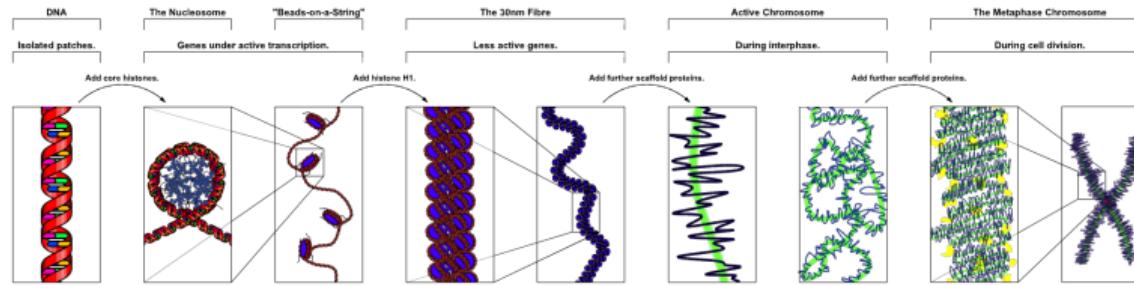


Figure: From Saunders et al., 2006

Transcriptional networks: A more detailed view; chromatin structure and epigenetics



Transcriptional networks: A more detailed view; chromatin structure and epigenetics

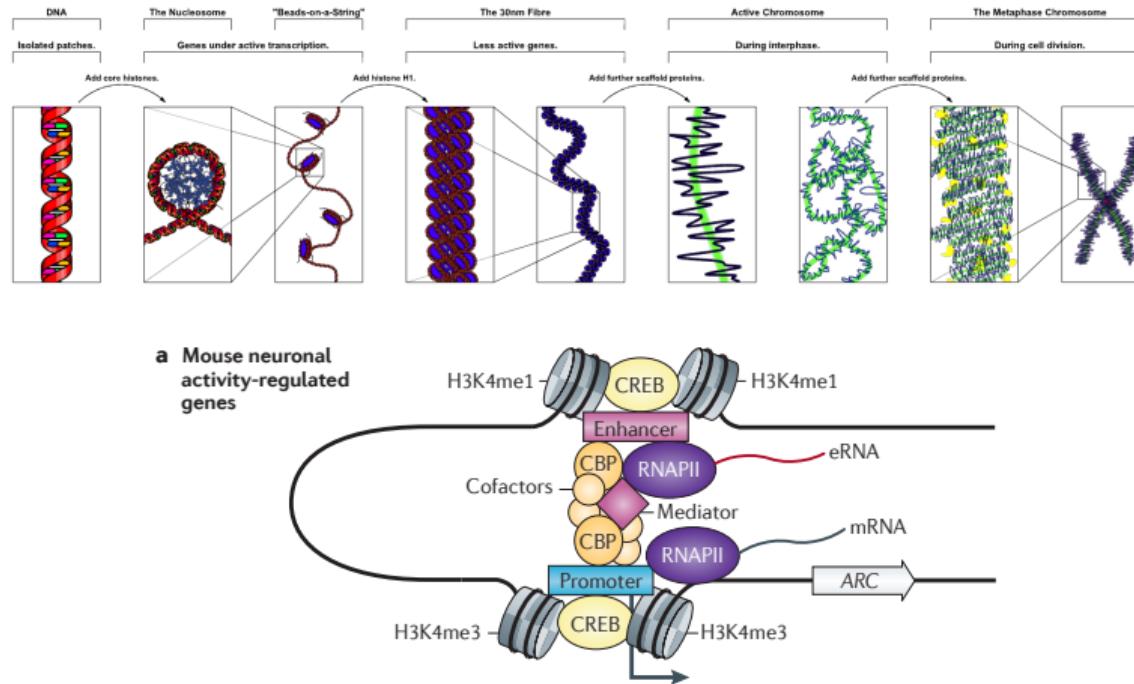


Figure: From Wikipedia and (Ong and Corces, 2012)

Transcriptional networks: A more detailed view

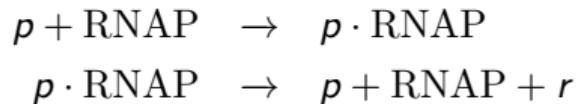
- ▶ A more detailed view
 - ▶ Chromatin accessibility
 - ▶ Nucleosome locations
 - ▶ Post-translational modifications
 - ▶ Sequence dependent protein-DNA binding
 - ▶ DNA methylation states
 - ▶ Histone modifications
 - ▶ DNA looping/3-D structures
 - ▶ RNAPII recruitment
 - ▶ Elongation initiation
 - ▶ ...
- ▶ Important to choose ‘a good’ modeling framework with an appropriate level of approximation!

Prokaryotic transcription

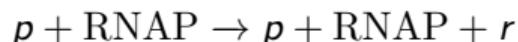
Figure 1.4 from (Wilkinson, 2011)

Prokaryotic transcription (2)

- ▶ Gene g (not modeled explicitly), promoter p , transcript r , RNA polymerase RNAP (see Fig. 1.4)

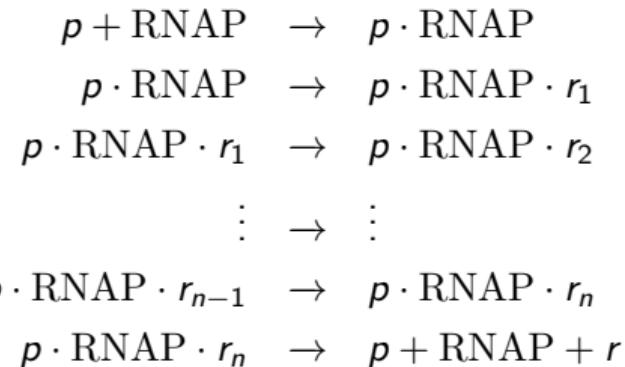


- ▶ Reactions do not necessarily form a closed system
- ▶ Linear chain of reactions can sometimes be summarized as



Prokaryotic transcription (3)

- ▶ Transcription including elongation: RNA polymerase moves along the DNA and transcribes the transcript one residue r_i at a time



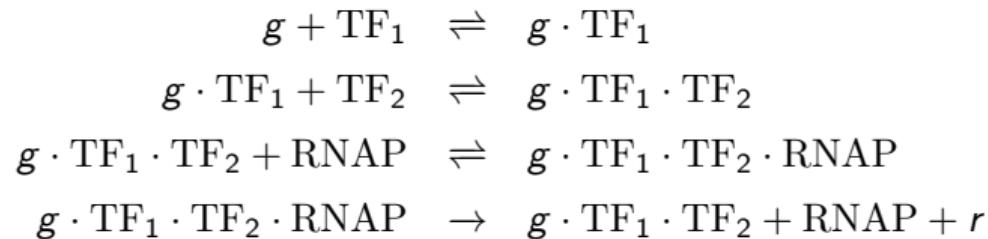
- ▶ Gene/promoter is blocked when a single RNAP binds/transcribes a gene/promoter

Eukaryotic transcription

Figure 1.5 from (Wilkinson, 2011)

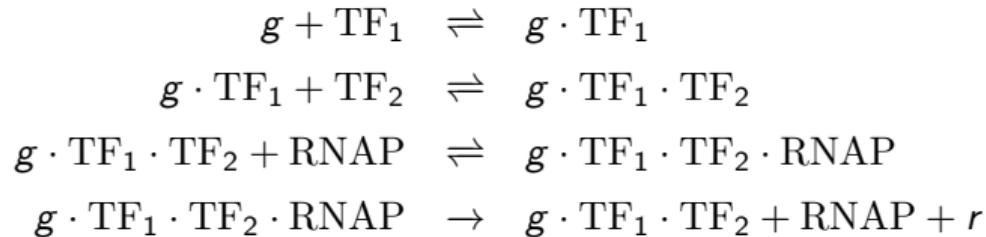
Eukaryotic transcription (2)

- ▶ Eukaryotic transcription is much more complex: see Fig. 1.5 for a simple model with transcription factors TF_1 and TF_2



Eukaryotic transcription (2)

- ▶ Eukaryotic transcription is much more complex: see Fig. 1.5 for a simple model with transcription factors TF_1 and TF_2



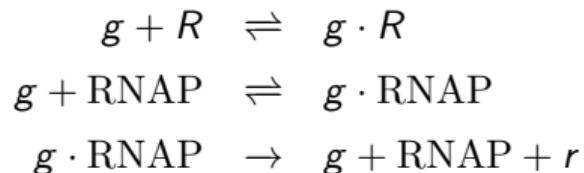
- ▶ Recall from the introductory slides that even this is a huge simplification
- ▶ But coupled chemical reactions provide a flexible modeling framework for complex biological systems too

Prokaryotic transcription repression

Figure 1.6 from (Wilkinson, 2011)

Prokaryotic transcription repression (2)

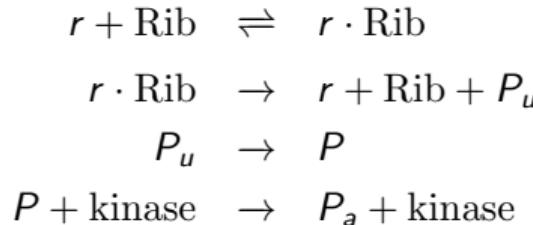
- ▶ Transcriptional regulation necessarily involves feedback (a definition for biological networks)
- ▶ An example of repression: see Fig. 1.6



- ▶ g and $g \cdot R$ are different chemical species

Translation

- ▶ Simplified reactions to produce an unfolded protein P_u from an mRNA molecule with the help of ribosome Rib, folded protein P , and active form of the protein P_a with a kinase

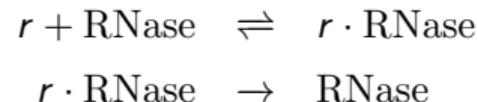


mRNA degradation

- ▶ Simply

$$r \rightarrow \emptyset$$

or



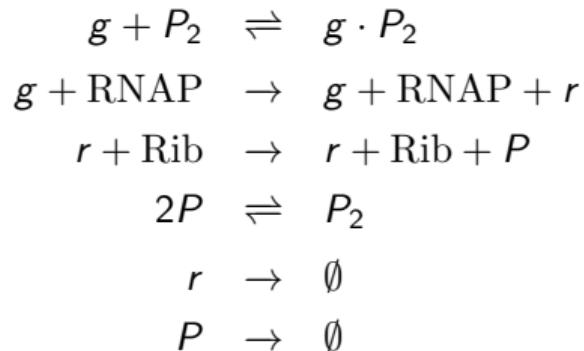
- ▶ Similar reactions for protein degradation and mRNA transport

Prokaryotic auto-regulation

Figure 1.7 from (Wilkinson, 2011)

Prokaryotic auto-regulation (2)

- ▶ Combine the previous building blocks of simple reactions into an auto-regulatory model (see Fig. 1.7)

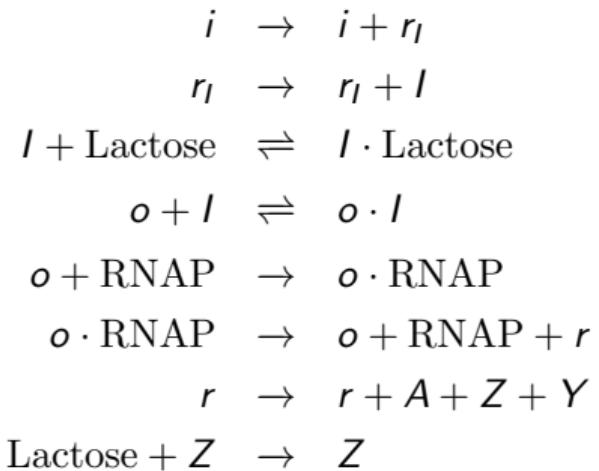


Lac operon

Figure 1.8 from (Wilkinson, 2011)

Lac operon (2)

- ▶ See Fig. 1.8



Coupled chemical reactions: summary

- ▶ Coupled chemical reaction network model formulation
- ▶ A flexible way of specifying a quantitative model is to write down coupled chemical reactions corresponding to a model
- ▶ We still need to have
 - ▶ Initial amounts of all chemical species
 - ▶ Rate laws for every reaction: quantify the “probability” of a certain reaction to happen

Graphical representations: graphs

- ▶ Displaying a model graphically helps in understanding
- ▶ Can be formalized using directed graphs $\mathcal{G}(V, E)$, where $V = \{v_1, \dots, v_n\}$ and $E = \{(v_i, v_j) \mid v_i, v_j \in V, v_i \rightarrow v_j\}$ are nodes and the directed edges respectively

Figure 2.1 from (Wilkinson, 2011)

Petri nets

- ▶ Petri nets are a mathematical framework to systems modeling together with
 - ▶ An associated graphical representation
 - ▶ A matrix formalism
- ▶ Petri net (graphical representation) corresponding to the prokaryotic auto-regulatory model is shown in Figure 2.3
 - ▶ Rectangular boxes represent reactions
 - ▶ Circles/Edges in and out of boxes correspond to reactants and products
 - ▶ Weights on edges specify the stoichiometries (proportions or molecular counts)

Petri nets (2)

Figure 2.3 from (Wilkinson, 2011)

Matrix formalism of Petri nets

- ▶ Formal definition $N = (P, T, \text{Pre}, \text{Post}, M)$
 - ▶ P is the list of chemical species: the number of species is u
 - ▶ T is the list of reactions: the number of reactions is v
 - ▶ Matrix Pre defines the stoichiometry of reactants
 - ▶ Size is reactions-by-chemical species, v -by- u
 - ▶ Pre_{ij} many copies of molecule P_j are needed in reaction i
 - ▶ Matrix Post defines the stoichiometry of products
 - ▶ Size is reactions-by-chemical species, v -by- u
 - ▶ Post_{ij} many copies of molecule P_j are produced in reaction i
 - ▶ M is an initial state (vector)
- ▶ The reaction i can happen only if $M_j \geq \text{Pre}_{ij}$ for all j

Matrix formalism of Petri nets (2)

- ▶ Stoichiometry of the prokaryotic auto-regulation example

Table 2.1 from (Wilkinson, 2011)

Matrix formalism of Petri nets (3)

- ▶ Petri net (matrix formalism) corresponding to the prokaryotic auto-regulation example

From (Wilkinson, 2011), equations below Table 2.1.

Matrix formalism (4)

- ▶ The molecular counts decrease and increase according to matrices Pre and $Post$, respectively
- ▶ Reaction matrix: $A = Post - Pre$
 - ▶ Rows represent the effect of individual reactions
 - ▶ See Table 2.2 in (Wilkinson, 2011) for an example
- ▶ Stoichiometry matrix $S = A^T$ (size: u -by v)

Matrix formalism (4)

- ▶ The molecular counts decrease and increase according to matrices *Pre* and *Post*, respectively
- ▶ Reaction matrix: $A = Post - Pre$
 - ▶ Rows represent the effect of individual reactions
 - ▶ See Table 2.2 in (Wilkinson, 2011) for an example
- ▶ Stoichiometry matrix $S = A^T$ (size: u -by- v)
- ▶ Given an initial state M and transitions vector $r \in \mathbb{Z}_+^v$ the new state M^* is

$$M^* = M + Sr,$$

where v is the number of reactions

Conservation law

- ▶ **Definition:** P -invariant is a non-zero vector y ($\in \mathbb{R}^u$) that is a solution to the matrix equation $Ay = 0$
- ▶ P -invariant defines the conservation laws of the network
 - ▶ Physical interpretation: total number of copies of molecular species which remain constant
- ▶ E.g. in the previous example $y = (1, 1, 0, 0, 0)^T$ is a P -invariant
 - ▶ I.e. $g \cdot P_2 + g = \text{Constant}$
 - ▶ This P -invariance simply says that the prokaryotic auto-regulation systems contains only one genome (or gene g)

Conservation law

- ▶ **Definition:** P -invariant is a non-zero vector y ($\in \mathbb{R}^u$) that is a solution to the matrix equation $Ay = 0$
- ▶ P -invariant defines the conservation laws of the network
 - ▶ Physical interpretation: total number of copies of molecular species which remain constant
- ▶ E.g. in the previous example $y = (1, 1, 0, 0, 0)^T$ is a P -invariant
 - ▶ I.e. $g \cdot P_2 + g = \text{Constant}$
 - ▶ This P -invariance simply says that the prokaryotic auto-regulation systems contains only one genome (or gene g)
- ▶ If y is a P -invariant then the linear combination of states, $y^T M$, is conserved

$$\begin{aligned} y^T M^* - y^T M &= y^T (M^* - M) = y^T S r \\ &= r^T S^T y = r^T A y = 0 \end{aligned}$$

Conservation law (2)

- ▶ **Definition:** T -invariant is a non-zero, non-negative, integer-valued vector x ($\in \mathbb{Z}_+^v$) that is a solution to the matrix equation $Sx = 0$
- ▶ Physical interpretation:
 - ▶ Corresponds to sequence of transitions that return to the initial state (recall: $M^* = M + Sr$)

Conservation law (2)

- ▶ **Definition:** T -invariant is a non-zero, non-negative, integer-valued vector x ($\in \mathbb{Z}_+^v$) that is a solution to the matrix equation $Sx = 0$
- ▶ Physical interpretation:
 - ▶ Corresponds to sequence of transitions that return to the initial state (recall: $M^* = M + Sr$)
- ▶ E.g. reversible reactions
- ▶ T -invariance is trivial to verify but less straightforward to find because solution x is required to be integer-valued

Reachability

- ▶ **Definition:** a state M^* is reachable from state M if there exists a finite sequence of reactions so that $M^* = M + Sr$
- ▶ Note that existence of solution to $M^* = M + Sr$ does not guarantee reachability
 - ▶ Each reaction in the finite sequence must have enough reactants for the reactions to be possible

Represent chemical reactions in computer

- ▶ SBML – systems biology markup language

```
1  <?xml version="1.0" encoding="UTF-8"?>
2  <sbml xmlns="http://www.sbml.org/sbml/level1"
3      level="1" version="2">
4      <model name="gene_network_model">
5          <listOfUnitDefinitions>
6              ...
7          </listOfUnitDefinitions>
8          <listOfCompartments>
9              ...
10         </listOfCompartments>
11         <listOfSpecies>
12             ...
13         </listOfSpecies>
14         <listOfParameters>
15             ...
16         </listOfParameters>
17         <listOfRules>
18             ...
19         </listOfRules>
20         <listOfReactions>
21             ...
22         </listOfReactions>
23     </model>
24 </sbml>
```

Structure of SBML

Summary

- ▶ Computational systems biology = (dynamical) systems theory + molecular cell biology
- ▶ Mathematical models: abstraction of biological systems
 - ▶ Coupled chemical reactions
 - ▶ Graphical representations: graphs and Petri nets
 - ▶ Stoichiometry
 - ▶ Conservation laws
 - ▶ Reachability

References

- ▶ Darren J. Wilkinson, *Stochastic Modelling for Systems Biology*, Chapman & Hall/CRC, 2011