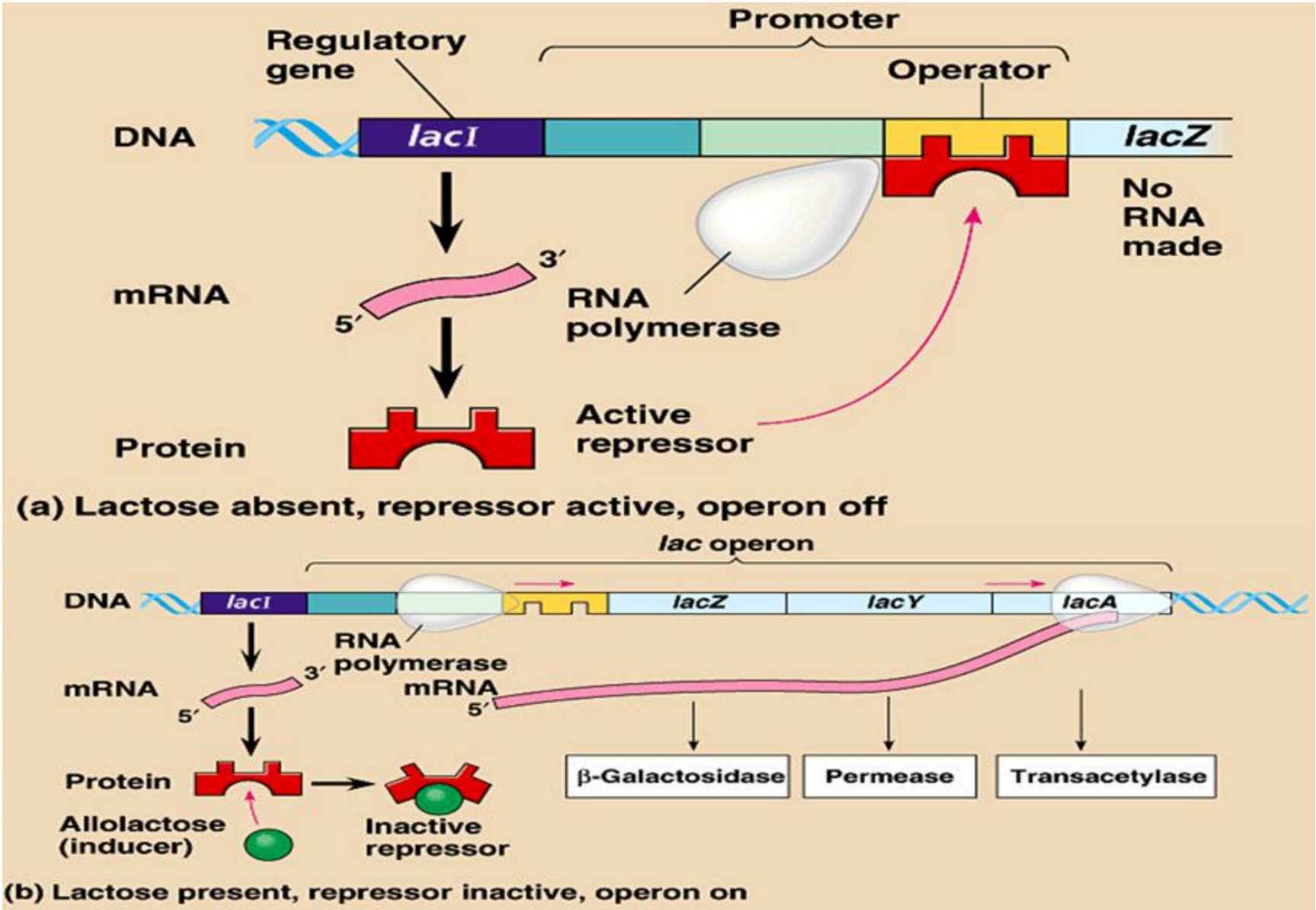
Basics of Boolean modeling

Matthew Macauley

Department of Mathematical Sciences
Clemson University
http://www.math.clemson.edu/~macaule/

Algebraic Biology

The *lac* operon in *E. coli*



Downsides of an ODE model

- Very mathematically technical.
- Too hard to solve explicitly. Numerical methods are needed.
- MANY experimentally determined parameters.
- Often, rate constants aren't known even up to orders of magnitude.

$$\frac{dM}{dt} = \alpha_M \frac{1 + K_1(e^{-\mu\tau_M} A_{\tau_M})^n}{K + K_1(e^{-\mu\tau_M} A_{\tau_M})^n} + \Gamma_0 - \widetilde{\gamma}_M M$$

$$\frac{dB}{dt} = \alpha_B e^{-\mu\tau_B} M_{\tau_B} - \widetilde{\gamma}_B B$$

$$\frac{dA}{dt} = \alpha_A B \frac{L}{K_L + L} - \beta_A B \frac{A}{K_A + A} - \widetilde{\gamma}_A A$$

$$\frac{dP}{dt} = \alpha_P e^{-\mu(\tau_B + \tau_P)} M_{\tau_B + \tau_P} - \widetilde{\gamma}_P P$$

$$\frac{dL}{dt} = \alpha_L P \frac{L_e}{K_{L_e} + L_e} - \beta_{L_e} P \frac{L}{K_{L_e} + L} - \alpha_A B \frac{L}{K_L + L} - \widetilde{\gamma}_L L$$

A Boolean approach

Let's assume everything is Boolean (0 or 1):

- Expression levels are high or basal (thousands of times lower).
- Gene products are present or absent.
- Enzyme concentrations are high or low.
- The probability of a repressor being activated is high ($p \approx 1$) or low ($p \approx 0$).
- The operon is ON or OFF.

Motivating example

A statement like

> *"mRNA will be transcribed ($M = 1$) if the transcription factor is present ($C = 1$) and the repressor protein is inactivated ($R = 0$)"*

can be modeled as

$$M(t+1) = C(t) \wedge \overline{R(t)}$$

We will assume that time is discretized: $t = 0, 1, 2, \ldots$.

A toy model of the *lac* operon

Parameters (constants):

- L_e: extracelllar lactose
- G_e: extracelllar glucose

Variables:

- M: mRNA
- E: gene products
- L: intracellular lactose

Each variable has an update function:

- mRNA is transcribed ($M = 1$) if there is no extracellular glucose ($G_e = 0$) and either intracellular ($L = 1$) or extracellular lactose (L_e) is present:

$$M(t+1) = f_M = \overline{G_e} \wedge (L(t) \vee L_e).$$

- The LacY and LacZ gene products ($E = 1$) will be translated if there are high levels of mRNA ($M = 1$):

$$E(t+1) = f_E = M(t).$$

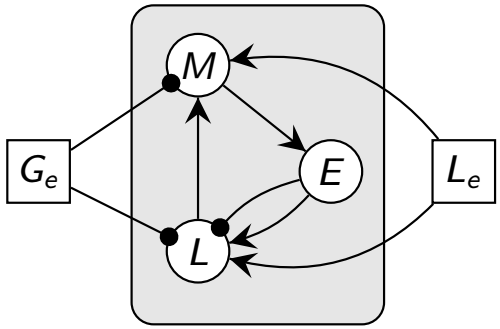
- Lactose will be in the cell ($L = 1$) if there is no extracellular glucose ($G_e = 0$), and either of the following holds:
  - Extracellular lactose is present ($L_e = 1$) and *lac* permease is available ($E = 1$).
  - Intracellular lactose is present ($L = 1$) but β-galactosiadase is absent ($E = 0$).

$$L(t+1) = f_L = \overline{G_e} \wedge \left[(L_e \wedge E(t)) \vee (L(t) \wedge \overline{E(t)})\right].$$

How to analyze a Boolean model

Our Boolean model is:

$$f_M = \overline{G_e} \wedge (L(t) \vee L_e)$$
$$f_E = M(t)$$
$$f_L = \overline{G_e} \wedge \big[(L_e \wedge E(t)) \vee (L(t) \wedge \overline{E(t)})\big].$$



We will update these functions synchronously:

$$f : (M(t), E(t), L(t)) \longmapsto (M(t+1), E(t+1), L(t+1)).$$

The state space (or phase space) is the directed graph (V, T), where

$$V = \big\{(M, E, L) \mid M, E, L \in \{0, 1\}\big\}, \qquad T = \big\{(x, f(x)) \mid x \in V\big\}.$$

We need to compute this for all 4 possible parameter vectors $(L_e, G_e) \in \{0, 1\}^2$.

At the bare minimum, we should expect:

- Lactose absent $\Rightarrow$ operon OFF,
- Lactose present, glucose absent $\Rightarrow$ operon ON,
- Lactose and glucose present $\Rightarrow$ operon OFF.

How to visualize the dynamics of a Boolean model

We can plot the state space using one of several software packages:

- The BoolNet library in R.
- Cyclone, available at `https://github.com/discretedynamics/cyclone`.
- GINsim (Gene Interaction Network simulation), available at `http://ginsim.org/`.

The function input of these varies, from Boolean functions (BoolNet) to polynomials (Cyclone), to truth tables (GINsim).

Boolean operation	logical form	polynomial form
AND	$x \wedge y$	xy
OR	$x \vee y$	$x + y + xy$
XOR	$x \oplus y$	$x + y$
NOT	$\overline{x}$	$1 + x$

A truth table is just a tabular representation of the entire function.

x	y	$x \wedge y$	$x \vee y$	$x + y + xy$
0	0	0	0	0
0	1	0	1	1
1	0	0	1	1
1	1	1	1	1

Installing Cyclone: Simulation and Analysis of Finite Dynamical Systems

Cyclone was written by Elena Dimitrova, Adam Knapp, Brandlyn Stigler, and Michael Stillman.

It can be downloaded from: `https://github.com/discretedynamics/cyclone`.

Open a terminal (Mac or Linux) and navigate to the zipfile. Then type:

```
> unzip cyclone-master.zip
> cd cyclone-master
> mkdir -p build
> cd build
> cmake ..
> make
```

On a Mac, if you don't have `cmake`, go download and install it, and then type

```
> PATH="/Applications/CMake.app/Contents/bin":"$PATH"
```

into the command-line.

Running Cyclone: Simulation and Analysis of Finite Dynamical Systems

Create a text file titled `lac-toy.pds` with the following contents:

```
# lac operon toy example
NUMBER OF VARIABLES: 5
NUMBER OF STATES: 2
M = (NOT Ge) AND (L | Le)
E = M
L = (NOT Ge) AND ((Le AND E) | (L AND (NOT E)))
Le = Le
Ge = Ge
```

Next, run the following command:

```
> ./simFDS lac-toy.pds
```

This will create the two text files:

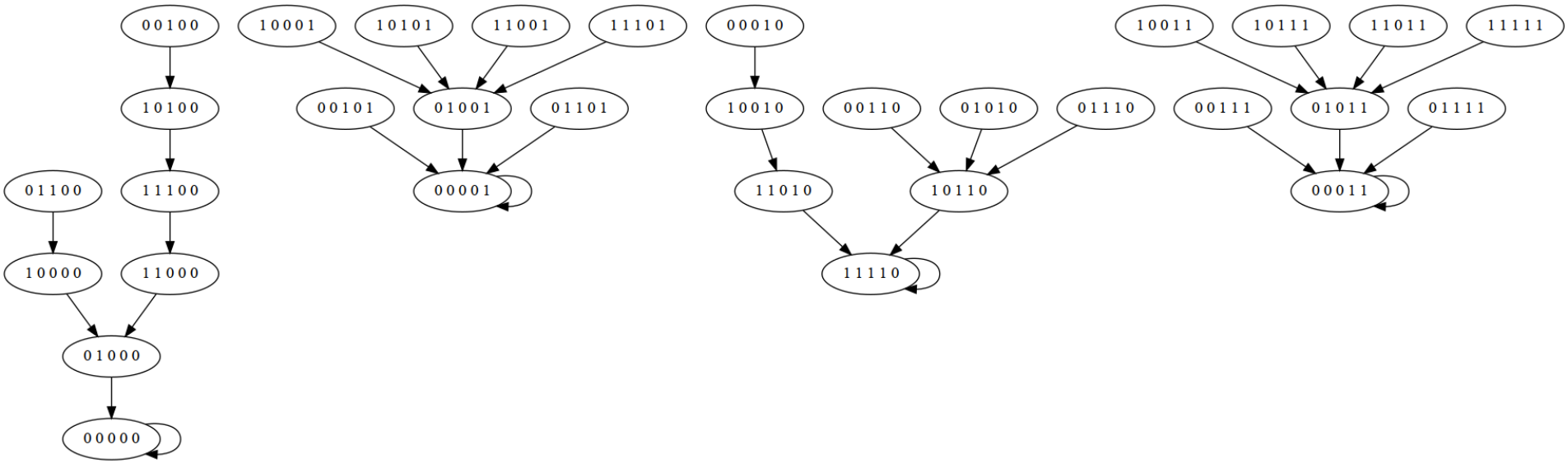
```
lac-toy-limitcycles.tex
lac-toy-statespace.dot
```

Visualizing and analyzing cyclone output

This command requires **Graphviz** to be installed:

```
> dot -Tpng -o lac-toy.png lac-toy-statespace.dot
```

It creates a png file of the state space:



Recall the variable order (M, E, L, L_e, G_e).

- A fixed point with $M = E = L = 0$ means that the operon is OFF.
- A fixed point with $M = E = L = 1$ means that the operon is ON.

Summary so far

Gene regulatory networks consist of a collection of gene products that interact with each other to control a specific cell function.

Classically, these have been modeled quantitatively with differential equations (continuous space, continuous time).

Boolean models take a different approach. They are discrete-space, discrete-time models that are inherently qualitative.

The state space graph encodes the dynamics. The most important features are the fixed points, and a necessary step in model validation is to check that they are biologically meaningful.

The model of the lac operon shown here is a "toy model." Next, we will see more complicated models of the lac operon that capture intricate biological features of these systems.

Modeling with Boolean logic is a relatively new concept, first done in the 1970s. It is a popular research topic in the field of systems biology.

A more refined model

Our first model only used 3 variables: mRNA (M), enzymes (E), and lactose (L).

Let's propose a new model with 5 variables:

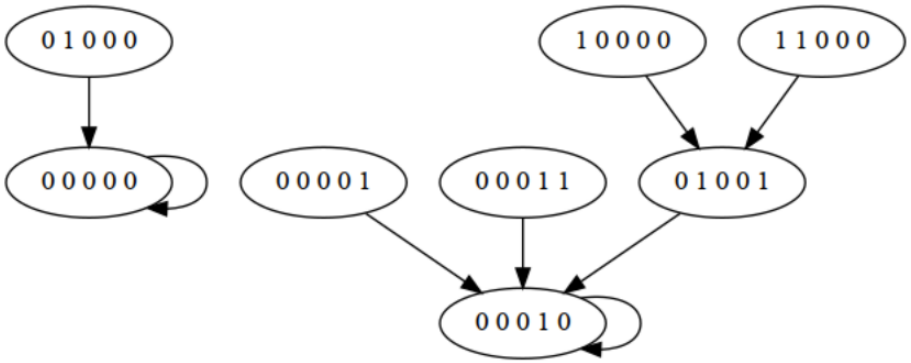
- M (mRNA): $f_M = A$
- B (β-galactosidase): $f_B = M$
- A (allolactose): $f_A = A \vee (L \wedge B)$
- L (intracellular lactose): $f_L = P \vee (L \wedge \overline{B})$
- P (*lac* permease): $f_P = M$

Assumptions

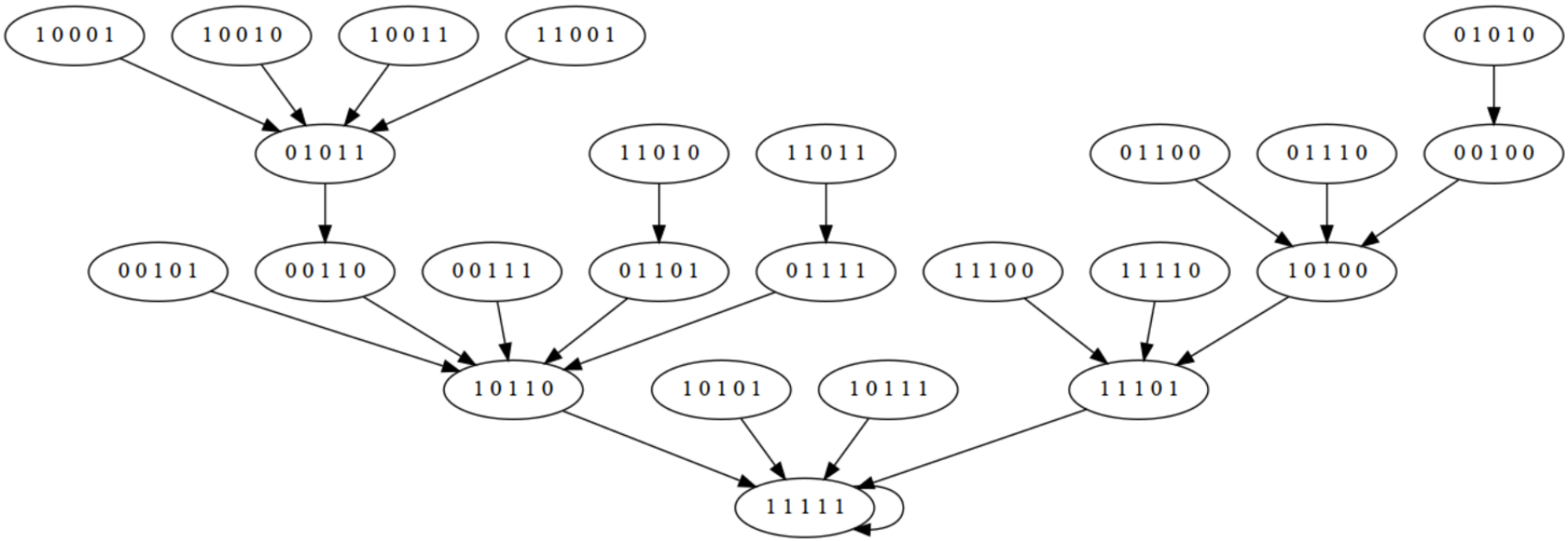
- Extracellular lactose is always available.
- Extracellular glucose is always unavailable.
- Translation and transcription require one unit of time.
- Protein and mRNA degradation require one unit of time.
- Lactose metabolism require one unit of time.

Model dynamics in Cyclone

Here is the state space of our 5-variable model:



$$f_M = A$$
$$f_B = M$$
$$f_A = A \vee (L \wedge B)$$
$$f_L = P \vee (L \wedge \overline{B})$$
$$f_P = M$$



Model dynamics in the BoolNet package of **R**

lac-operon

```
1  targets, factors
2  M, A
3  B, M
4  A, A | (L & B)
5  L, P | (L & !B)
6  P, M
7
```

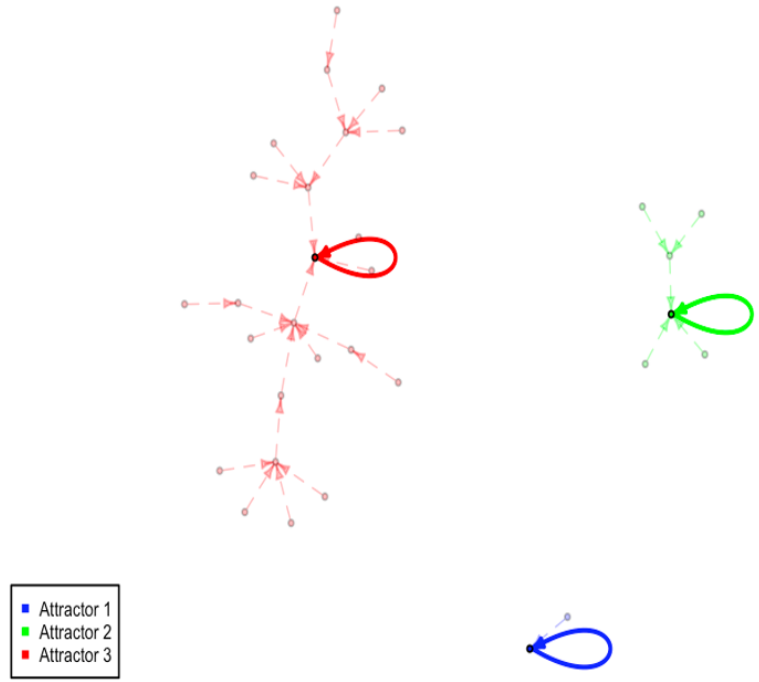
7:1 Text File

Console Terminal

~/code/BoolNet/

```
> lacModel <- loadNetwork("lac-operon")
> getAttractors(lacModel)
Attractor 1 is a simple attractor consisting of 1 state(s) and has a basin of 2 state(s):

 |--<----|
 V       |
 00000   |
 V       |
 |-->----|


Genes are encoded in the following order: M B A L P

Attractor 2 is a simple attractor consisting of 1 state(s) and has a basin of 6 state(s):

 |--<----|
 V       |
 00010   |
 V       |
 |-->----|


Genes are encoded in the following order: M B A L P

Attractor 3 is a simple attractor consisting of 1 state(s) and has a basin of 24 state(s):

 |--<----|
 V       |
 11111   |
 V       |
 |-->----|


Genes are encoded in the following order: M B A L P
```

```
> lacModel <- loadNetwork("lac-operon")
> lacAttractors <- getAttractors(lacModel)
> plotStateGraph(lacAttractors)
>
```



Problems with our refined model

Boolean model

- M (mRNA): $f_M = A$
- B (β-galactosidase): $f_B = M$
- A (allolactose): $f_A = A \vee (L \wedge B)$
- L (intracellular lactose): $f_L = P \vee (L \wedge \overline{B})$
- P (*lac* permease): $f_P = M$

Problems

- The fixed point $(M, B, A, L, P) = (0, 0, 0, 0, 0)$ should not happen with lactose present but not glucose. [though let's try to justify this. . .]
- The fixed point $(M, B, A, L, P) = (0, 0, 0, 1, 0)$ is not biologically feasible: it would describe a scenario where the bacterium does not metabolize intracellular lactose.

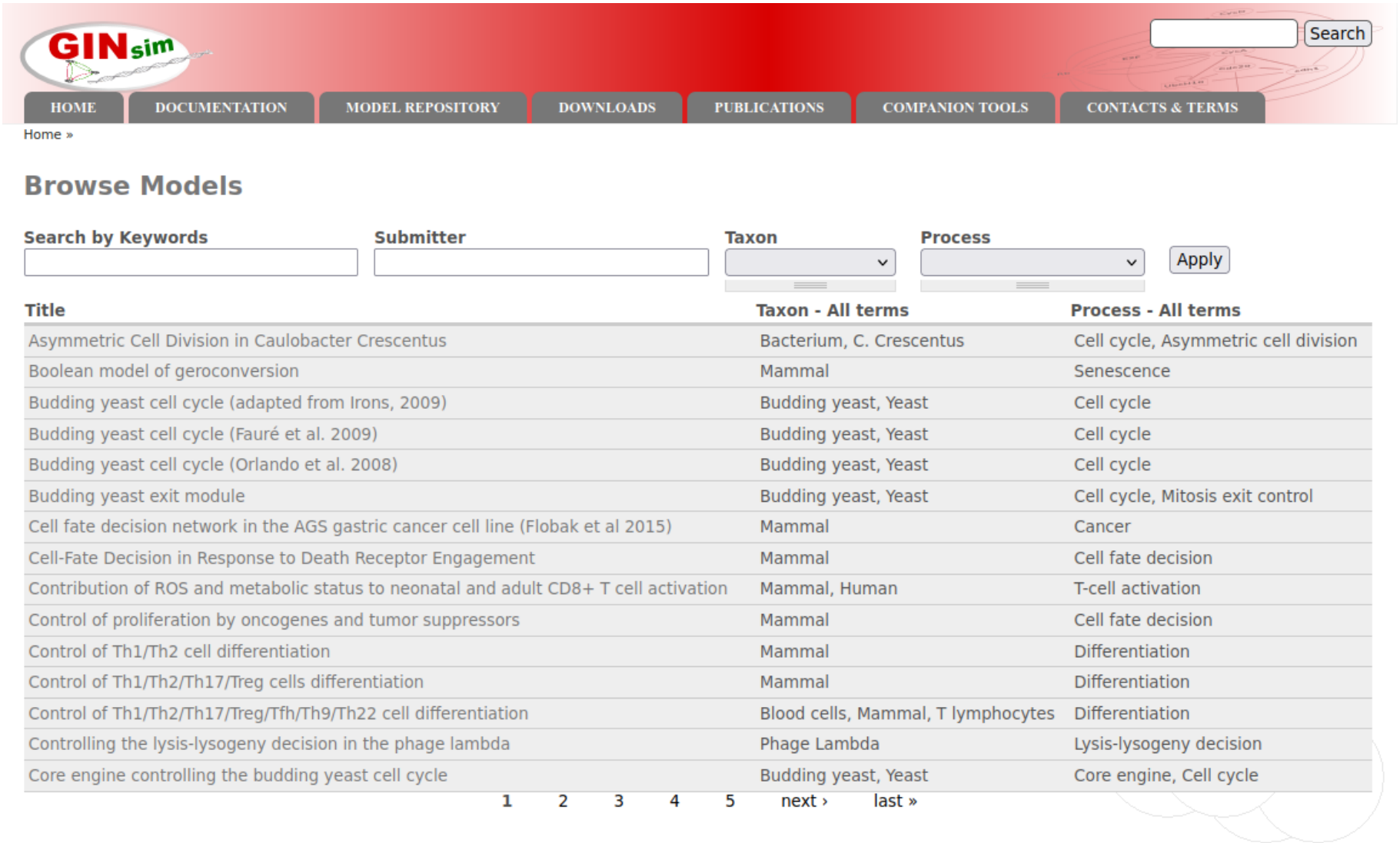
Conclusion

The model fails initial testing and validation, and is in need of modification. (HW)

Examples of Boolean models of (mostly) molecular networks

- A. Veliz-Cuba and B. Stigler. Boolean models can explain bistability in the *lac* operon. *J. Comp. Biol.*, **18**(6):783–794, 2011.
- R. Albert, H.G. Othmer. The topology of the regulatory interactions predicts the expression pattern of the segment polarity genes in Drosophila melanogaster. *J. Theor. Biol.*, **223**(1):1–18, 2003.
- L. Mendoza, D. Thieffry, and E. Alvarez-Buylla. Genetic control of flower morphogenesis in Arabidopsis thaliana: a logical analysis. *Bioinformatics*, **15**(7):593–606, 1999.
- M. Davidich and S. Bornholdt. Boolean network model predicts cell cycle sequence of fission yeast. *PloS ONE*, **3**(2):e1672, 2008.
- S. Kauffman, C. Peterson, B. Samuelsson, and C. Troein. Random Boolean network models and the yeast transcriptional network. *Proc. Natl. Acad. Sci.*, **100**(25):14796–14799, 2003.
- A. Fauré, A. Naldi, C. Chaouiya, and D. Thieffry. Dynamical analysis of a generic Boolean model for the control of the mammalian cell cycle. *Bioinformatics*, **22**(14):e124–e131, 2006.
- B. Stigler and H. Chamberlin. A regulatory network modeled from wild-type gene expression data guides functional predictions in caenorhabditis elegans development. *BMC Syst. Biol.*, **6**(1):77, 2012.
- A. Jenkins and M. Macauley. Bistability and asynchrony in a Boolean model of the L-arabinose operon in *Escherichia coli*. *Bull. Math Biol.*, **79**(8):1778–1795, 2017.
- L. Mendoza. A network model for the control of the differentiation process in Th cells. *Biosystems* **84**(2):101–114, 2006.
- I. Deal, M. Macauley, and R. Davies. Boolean models of the transport, synthesis, and metabolism of tryptophan in Escherichia Coli. *Bull. Math. Biol.* **85**(4), 29 pp., 2023.
- R. Robeva and D. Murrugarra. The spruce budworm and forest: a qualitative comparison of ODE and Boolean models. *Lett. Biomathematics* **3**(1):75–92, 2016.
- F. Hammami, et al. Analysis of a logical regulatory network reveals how Fe-S cluster biogenesis is controlled in the face of stress. *Microlife* **4**:uqad003, 2023.
- E.E. Allen, J.S. Fetrow, L.W. Daniel, S.J. Thomas, and D.J John. Algebraic dependency models of protein signal transduction networks from time-series data. *J. Theor. Biol.*, **238**(2):317–330, 2006.

GINsim logical model repository



`http://ginsim.org/models_repository`