Lecture 14:

Mathematical representation of biological processes:

Dynamic signalling

and

Gene Expression Regulation

Images and text from: Tyson JJ - *Sniffers, buzzers, toggles and blinkers.* Curr Opin Cell Biol. 2003 Apr;15(2):221-31.
E. Klipp, Systems Biology in Practice, Wiley-VCH, 2005 – Chapter 8

PART 1

Modelling of molecular networks:

Simple modules for building complex dynamic networks

- Linear and hyperbolic
- Sigmoidal (**BUZZER**)
- Perfectly adapted (**SNIFFER**)
- Positive feedback
  - Mutual activation (**ONE WAY SWITCH**)
  - Mutual inhibition (**TOGGLE SWITCH**)
- Negative feedback
  - homeostasis
  - oscillations (**BLINKER**)

Signal response curves

Introduction: molecular networks vs. circuit design

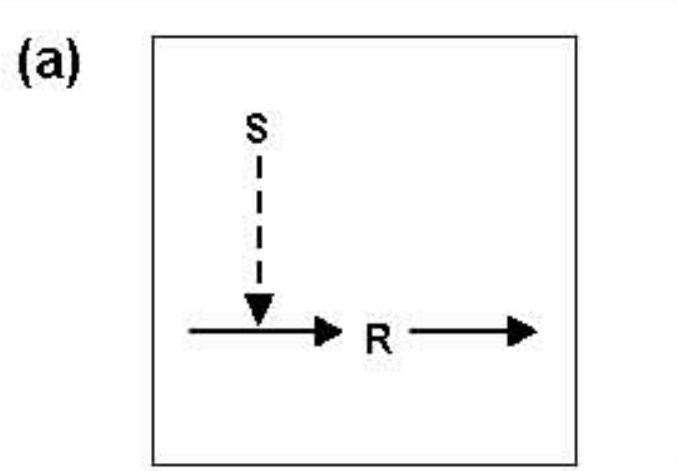
- A molecular network looks strikingly similar to the wiring diagram of a modern electronic gadget. Instead of resistors, capacitors and transistors hooked together by wires, one sees genes, proteins and metabolites hooked together by chemical reactions and intermolecular interactions.

- Complex molecular networks, like electrical circuits, seem to be constructed from simpler modules: sets of interacting genes and proteins that carry out specific tasks and can be hooked together by standard linkages.

- Simple signalling pathways can be embedded in networks using positive and negative feedback to generate more complex behaviours — toggle switches and oscillators — which are the basic building blocks of the exotic, dynamic behaviour shown by nonlinear control systems.

Linear signal-response curve (1)

- A simple example of protein dynamics: protein synthesis and degradation
- Using the law of mass action, we can write the rate equation
- S = signal strength (e.g. concentration of mRNA)
- R = response magnitude (e.g. concentration of protein)



$$\frac{dR}{dt} = k_0 + k_1 S + k_2 R$$

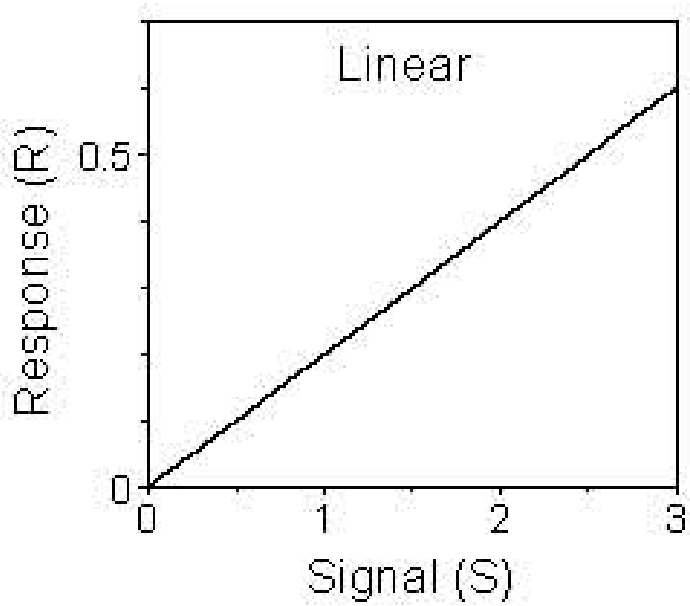
Linear (2)

- A steady-state solution of a differential equation, dR/dt = f(R), is a constant, R_{ss}, that satisfies the algebraic equation f(R)=0.

$$R_{ss} = \frac{k_0 + k_1 S}{k_2}$$

- In most cases, such simple components are embedded in more complex pathways, to generate signal-response curves of more adaptive value.

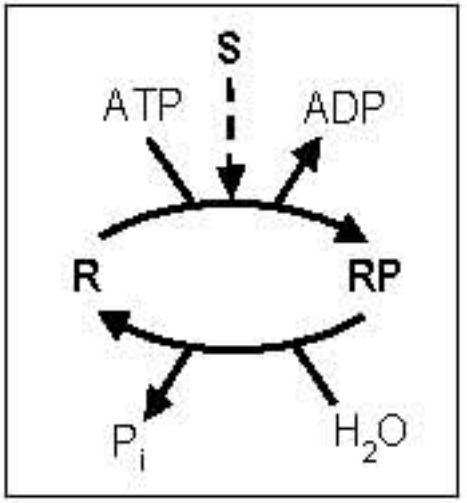
- Diagram representing the strength of the signal response in as function of the signal strength



Hyperbolic signal-response curve (1)

- Another simple example of protein dynamics: protein phosphorylation and de-phosphorylation
- RP = the phosphorylated form of the response element (which we suppose to be the active form)
- R_P =[RP], and $R_T = R + R_P$ = total concentration of the response element.

(b)



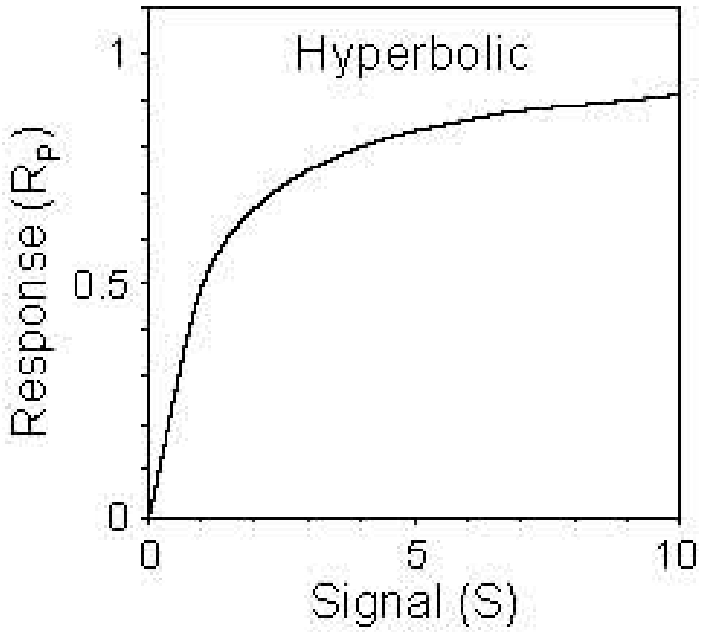
$$\frac{dR}{dt} = k_0 + k_1 S + k_2 R$$

Hyperbolic (2)

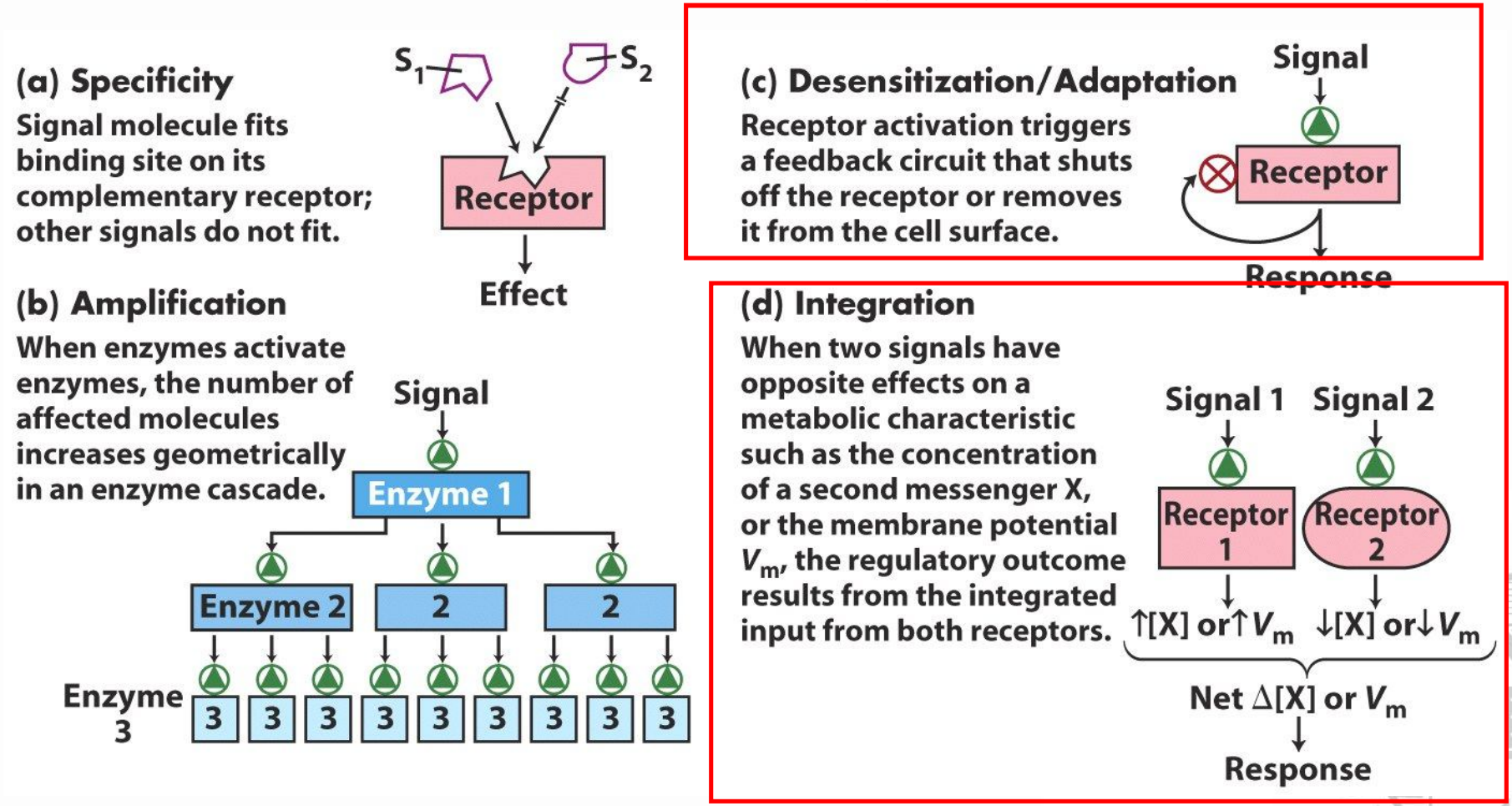
- A steady-state solution , for f(R)=0.

$$R_{P,\,ss} = \frac{RtS}{(k_2 / k_1) + S}$$

- In most cases, such simple components are *integrated* in more complex pathways, for example to generate signal-response curves of more *adaptive* value.



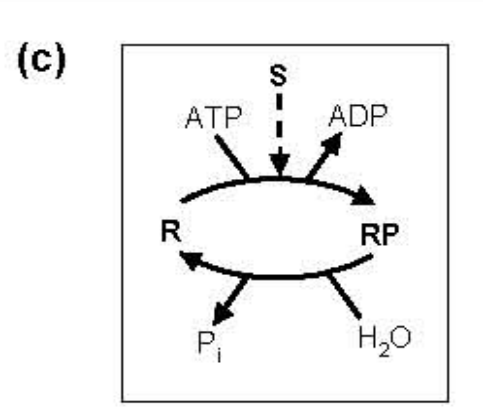
Integration of simple modules to obtain adaptation



From: D. L. Nelson, Lehninger Principles of Biochemistry, IV Edition – Chapter 12 Biosignaling

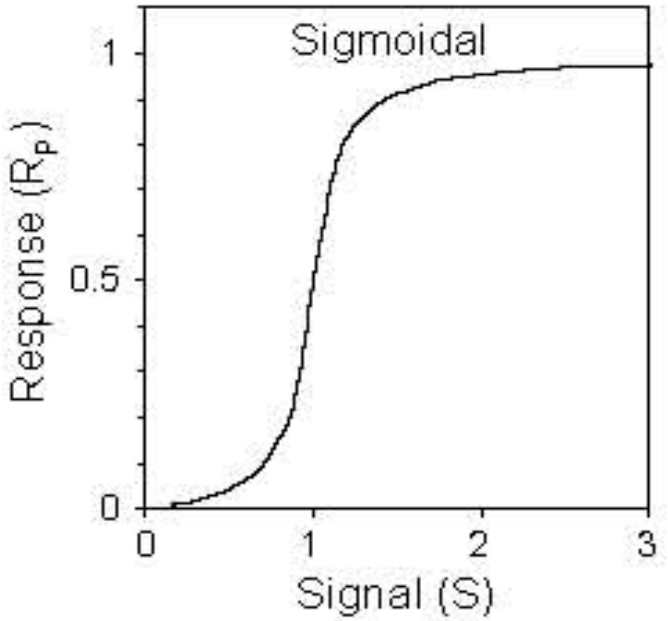
Sigmoidal signal-response curve (1)

- The sigmoidal case (c) is a modification of case (b), where the phosphorylation and de-phosphorylation reactions are governed by Michaelis-Menten kinetics.

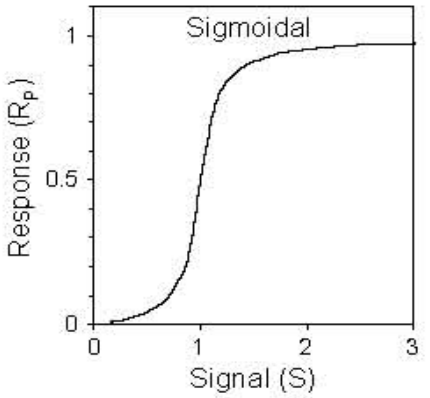


- The steady-state solution of the quadratic equation is the Goldbeter-Koshland function.

- The sigmoid shape represents a switch (for a small difference in signal the response at a certain threshold switches: from null to high, from off to on, from 0 to 1)



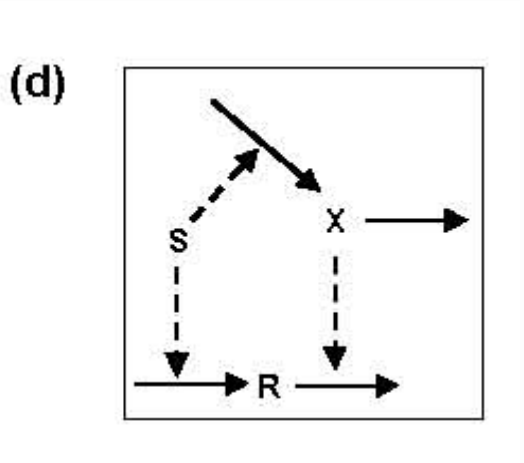
Sigmoidal signal-response curve (2): BUZZER

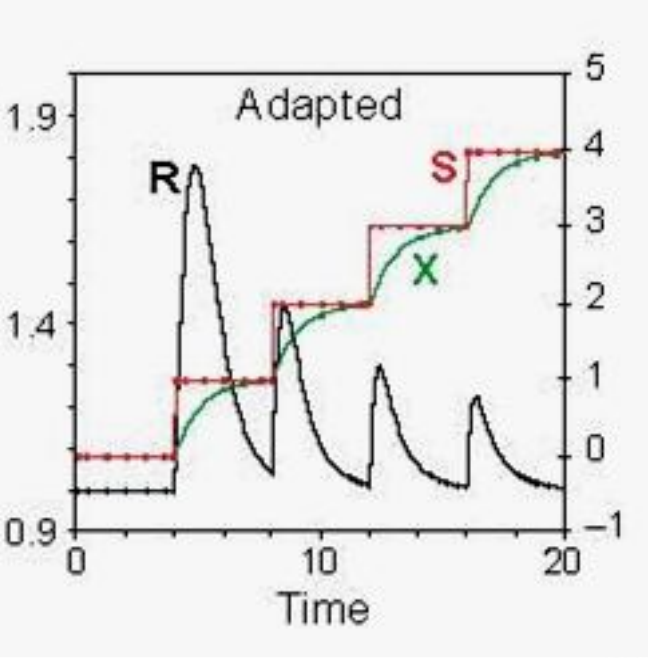


- The Goldbeter–Koshland function, although switch-like, shares with linear and hyperbolic curves the properties of being graded and reversible.
- **Graded** = the response increases continuously with signal strength. A slightly stronger signal gives a slightly stronger response.
- **Reversible** = if the signal strength is changed from $S_{initial}$ to S_{final}, the response at S_{final} is the same whether the signal is being increased ($S_{initial} < S_{final}$) or decreased ($S_{initial} > S_{final}$)
- Although continuous and reversible, a sigmoidal response is abrupt. Like a buzzer or a laser pointer, to activate the response one must push hard enough on the button, and to sustain the response one must keep pushing.
- When one lets up on the button, the response switches off at precisely the same signal strength at which it switched on.

Perfect adaptation: SNIFFER

- By supplementing the simple linear response element (Figure a) with a second signalling pathway (through species X), we obtain perfect adaptation to the signal.
- **Perfect adaptation:** although the signalling pathway exhibits a transient response to changes in signal strength, its steady-state response Rss is independent of S.
- This is typical of chemotactic systems, which respond to an abrupt change in attractants or repellents, but then adapt to a constant level of the signal.
- Our own sense of smell operates this way, so we refer to this type of response as a '**sniffer**.'





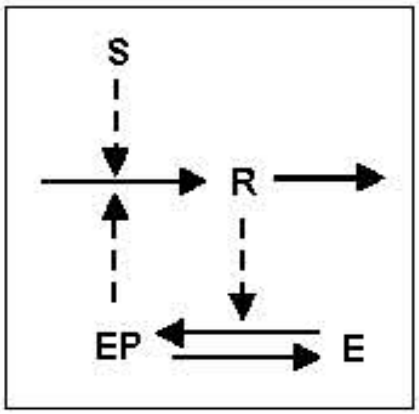
Positive feedback: switches

- In Figure d the signal influenced the response via two parallel pathways that push the response in opposite directions (an example of feed-forward control).

- Alternatively, some component of a response pathway may feed back on the signal. Feedback can be positive, negative or mixed.

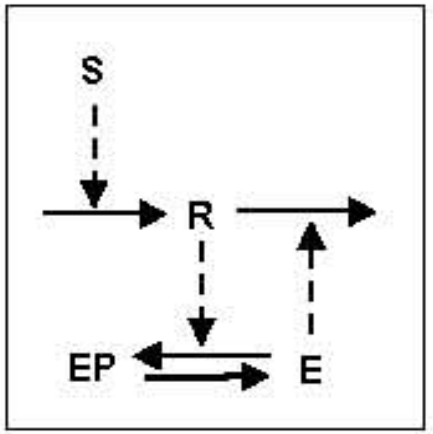
- There are two types of **positive feedback**:
- In Figure e, R activates protein E (by phosphorylation), and EP enhances the synthesis of R.
- In Figure f, R inhibits E, and E promotes the degradation of R; hence, R and E are mutually antagonistic.

(e)



mutual activation

(f)

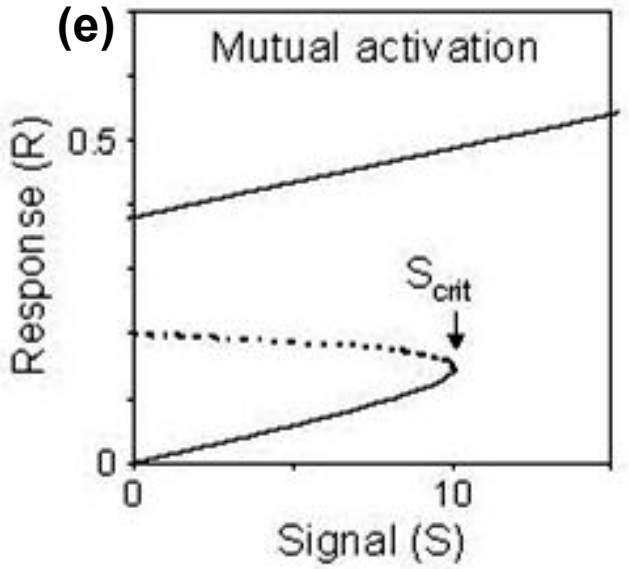


mutual inhibition

Irreversible switch: ONE WAY SWITCH

- In either case (mutual activation or antagonism), positive feedback may create a **discontinuous switch**, meaning that the cellular response changes abruptly and irreversibly as signal magnitude crosses a critical value.

- In Figure e, as signal strength (S) increases, the response is low until S exceeds some critical intensity, S_{crit}, at which point the response increases abruptly to a high value.



- Then, if S decreases, the response stays high (i.e. the switch is **irreversible**; unlike a sigmoidal response, which is reversible).

- *Notice that, for S values between 0 and S_{crit}, the control system is 'bistable' — that is, it has two stable steady-state response values (on the upper and lower branches — the solid lines) separated by an unstable steady state (on the intermediate branch — the dashed line).*

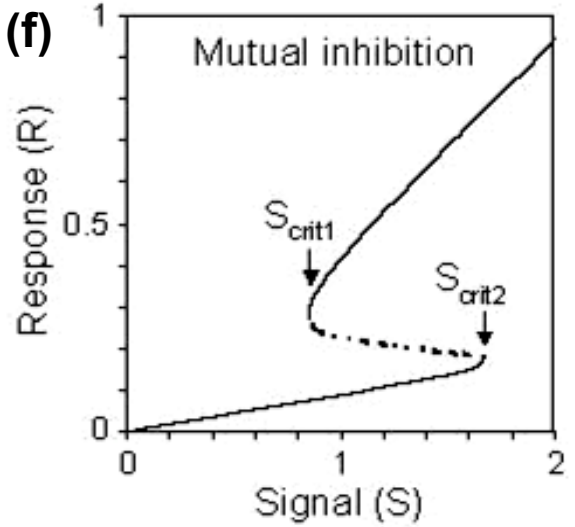
Reversible switch: TOGGLE SWITCH

- In the **toggle switch**, if S is decreased enough, the switch will go back to the off-state, as in Figure f.

- For intermediate stimulus strengths

$$S_{crit1} < S < S_{crit2}$$

the response of the system can be either small or large, depending on how S was changed.



- This sort of two-way, discontinuous switch is often referred to as **hysteresis**.

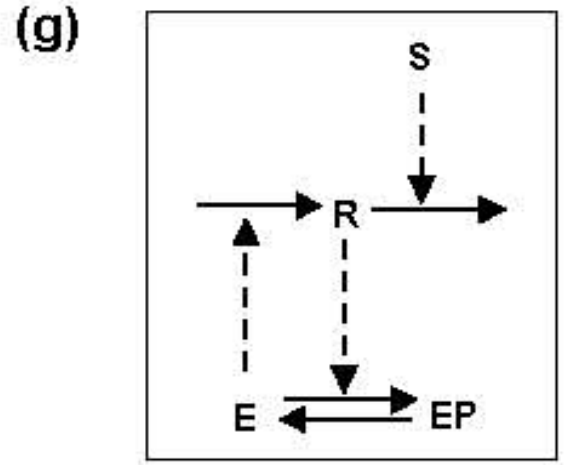
- Examples include: the lac operon in bacteria, the activation of M-phase-promoting factor (MPF) in frog egg extracts, and the autocatalytic conversion of normal prion protein to its pathogenic form.
- While one-way switches play roles in processes characterized by a point-of-no-return, like frog oocyte maturation in response to progesterone or apoptosis (regulated cell death).

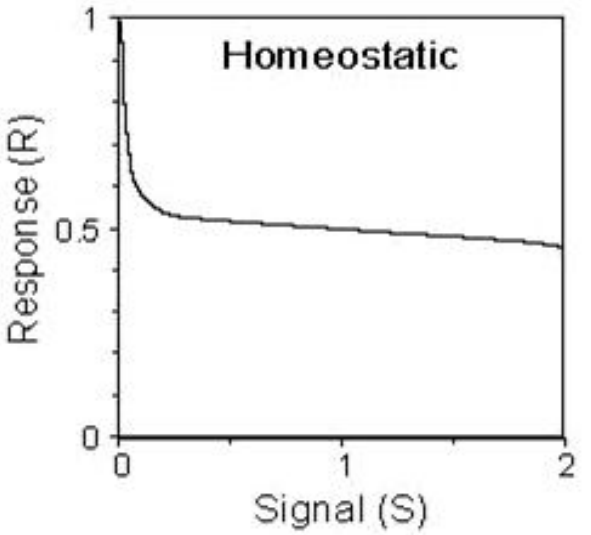
Some Math: bifurcations

- *The signal-response curves in Figure e and f would be called '**one-parameter bifurcation diagrams**' by an applied mathematician.*
- *The parameter is signal strength (manipulable by the experimenter). The steady-state response, on the Y axis, is an indicator of the behaviour of the control system as a function of the signal.*
- *At S_{crit}, the behaviour of the control system changes abruptly and irreversibly from low response to high response (or vice versa).*
- *Such points of qualitative change in the behaviour of a nonlinear control system are called **bifurcation points**, in this case, a 'saddle-node bifurcation point'.*
- *Other, more esoteric bifurcation points, are associated with more complex signal-response relationships.*

Negative feedback: homeostasis

- In **negative** feedback, the response counteracts the effect of the stimulus.
- In Figure g, the response element, R, inhibits the enzyme catalysing its synthesis
- The steady state concentration of R is confined to a narrow window for a broad range of signal strengths, because the supply of R adjusts to its demand.
- This type of regulation, commonly employed in biosynthetic pathways, is called **homeostasis**.
- *(It is a kind of imperfect adaptation, but it is not a sniffer because stepwise increases in S do not generate transient changes in R.)*





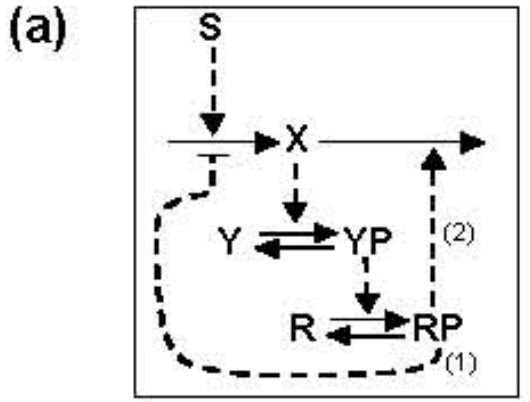
Negative feedback: BLINKER

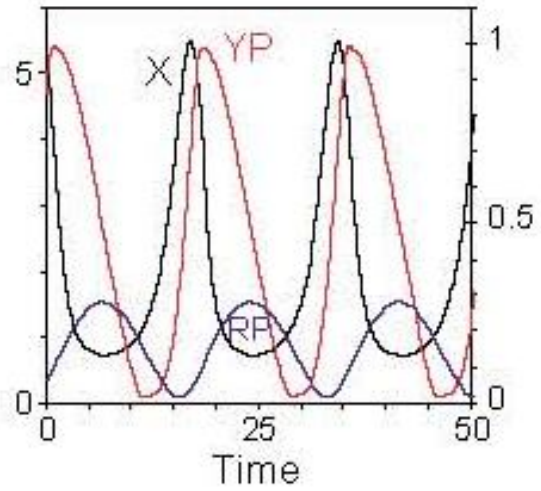
- Negative feedback can also create an oscillatory response. A two-component, negative feedback loop, X→R—|X, can exhibit **damped** oscillations to a stable steady state but not **sustained** oscillations

- Sustained oscillations require at least three components: X→Y→R—|X. The third component (Y) introduces a time delay in the feedback loop, causing the control system repeatedly to overshoot and undershoot its steady state

- Negative feedback has been proposed as a basis for oscillations in protein synthesis, MAPK signalling pathways, and circadian rhythms

- *Elowitz and Leibler designed an artificial genetic network consisting of three operons that repress one another in a loop. Individual bacteria containing these plasmids showed periodic expression of a fluorescent reporter protein, qualifying this case as a literal '**blinker**'.*





PART 2

Modelling molecular networks:

Gene expression regulation

Modelled with:

- Ordinary differential equations
- Directed and undirected graphs
- Bayesian networks
- Boolean networks

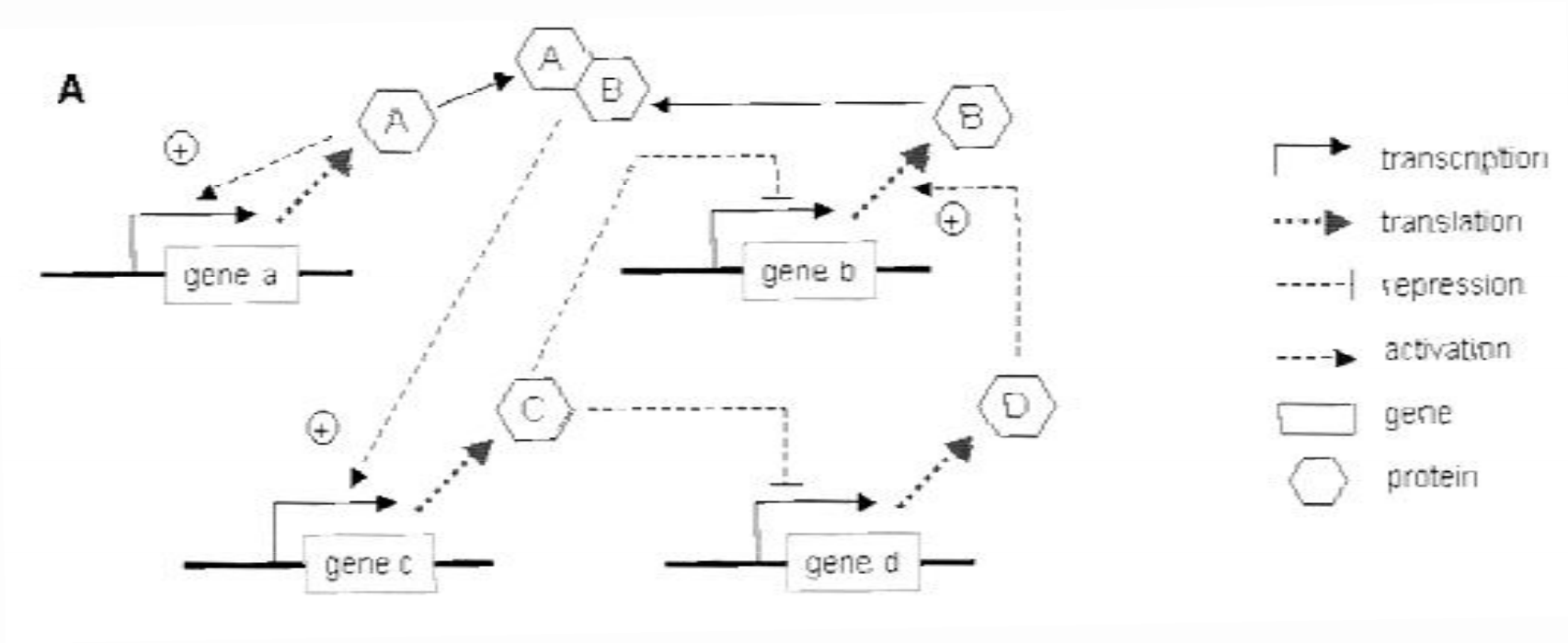
- and (not covered here): stochastic equations, partial differential equations, rule-based formalisms and many others…

Images and text from: E. Klipp, Systems Biology in Practice, Wiley-VCH, 2005 – Chapter 8

One example, many approaches (1)

- A network of interactions among DNA, RNA, proteins, and other molecules realises the regulation of gene expression.
- There is forward flow of information from gene to mRNA to protein according to the dogma of molecular biology.
- Moreover, positive and negative feedback loops and information exchange with signalling pathways and energy metabolism ensure the appropriate regulation of the expression according to the actual state of the cell and its environment
- Modelling of gene expression is an example of a scientific field where one may obtain results with different techniques.

One example, many approaches (2)



- The example presented here contains four genes, a through d, which code for the proteins A through D.
- mRNA is not shown for sake of simplicity.
- The proteins A and B may form a heterodimer that activates the expression of gene c.
- Protein C inhibits the expression of genes band d, which are in this way co-regulated.
- Protein D is necessary for the transcription of protein B.

Description with: Ordinary Differential Equations (1)

- Gene expression can be mathematically described with systems of ordinary differential equations in the same way as metabolism (Chapter 5) or signalling (Chapter 6)
- In general, one considers:

$$\frac{dx_i}{dt} = f_i(x_1,\ldots,x_n) \quad i = 1,\ldots,n$$

- The variables x_i represent the concentrations of mRNAs, proteins, or other molecules. The functions f_i comprise the rate equations that express the changes of x_i due to transcription, translation, or other individual processes.

Description with: Ordinary Differential Equations (2)

The dynamics of the system depicted in Fig. 8.2 can be described in several ways depending on the desired particularization. If we consider only the mRNA abundances a, b, c, and d, we get:

$$\frac{da}{dt} = f_a(a) \qquad f_a(a) = v_a - k_a \cdot a$$

$$\frac{db}{dt} = f_b(b, c, d) \qquad f_b(b, c, d) = \frac{V_b \cdot d^{n_d}}{(K_b + d^{n_d})(K_{Ic} + c^{n_c})} - k_b \cdot b$$

$$\frac{dc}{dt} = f_c(a, b, c) \qquad f_c(a, b, c) = \frac{V_c \cdot (a \cdot b)^{n_{ab}}}{K_c + (a \cdot b)^{n_{ab}}} - k_c \cdot c$$

$$\frac{dd}{dt} = f_d(c, d)\,. \qquad f_d(c, d) = \frac{V_d}{K_{Ic} + c^{n_c}} - k_d \cdot d\,. \qquad (8\text{-}3)$$

← If we also consider the depicted regulatory interactions

Here, ka, kb, kc, and kd are the first-order rate constants of the degradation of a, b, c, and d, respectively.

Description with: Ordinary Differential Equations (3)

- The ODE formalism allows involving more details, e.g. we can distinguish between the processes determining the velocity of translation (basic rate V_b and inhibition by protein C), transcription (dependence on mRNA concentration b and on the activator concentration D), and degradation or consumption on both levels (degradation of band B and formation of complex AB).

- **The advantage:** ODE systems take into account detailed knowledge about gene regulatory mechanisms such as individual kinetics, individual interactions of proteins with proteins or proteins with mRNA.

- **The disadvantage:** the current lack of exactly this type of knowledge - the lack of kinetic constants due to measurement difficulties and uncertainties in the function of many proteins and their interactions.

Description with: Ordinary Differential Equations (4)

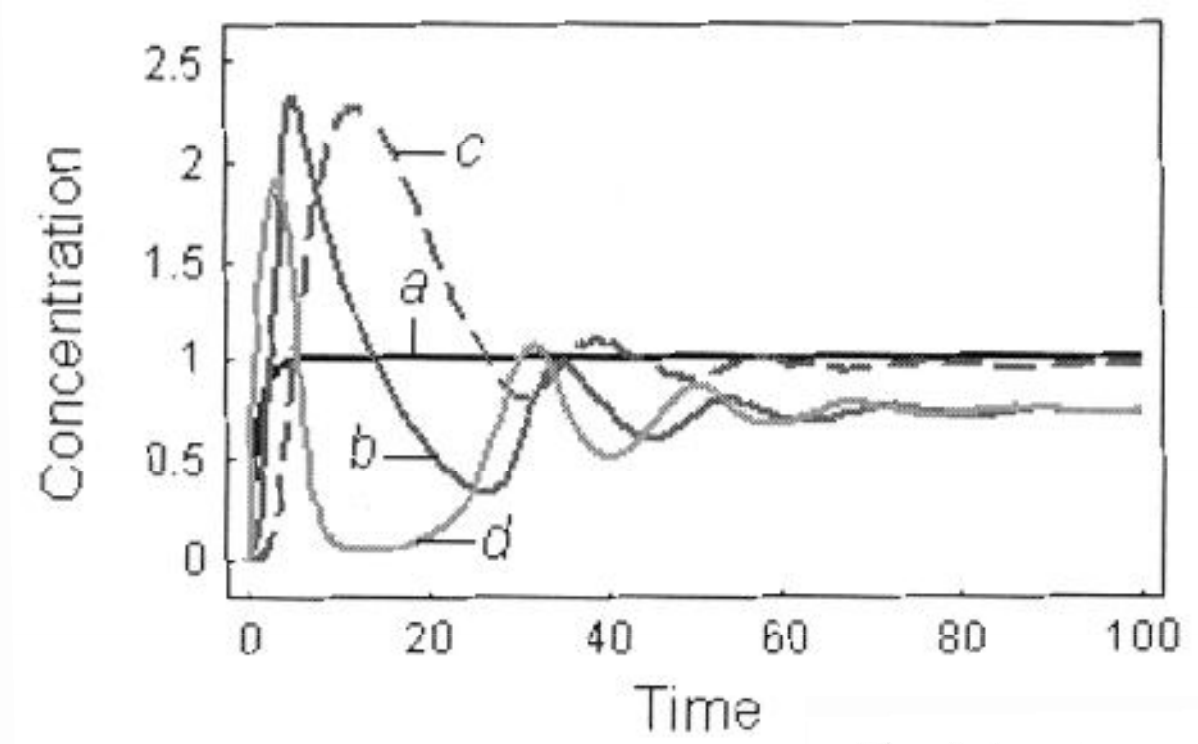
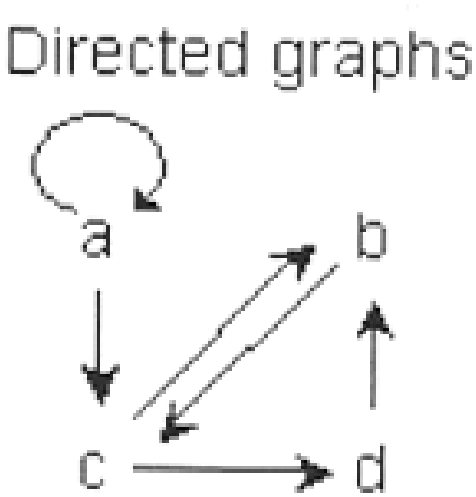


Fig. 8.3 Dynamics of the mRNA concentrations of the system presented in Example 8-1 according to Eq. (8-4). Parameters: $\nu_a = 1$, $k_a = 1, V_b = 1$, $K_b = 5$, $K_{ic} = 0.5$, $n_c = 4$, $k_b = 0.1$, $V_c = 1$, $K_c = 5$, $k_c = 0.1, V_d = 1$, $k_d = 1$. Initial conditions: $a(0) = b(0) = c(0) = d(0) = 0$.

Description with:

Directed and Undirected graphs (1)

- A directed graph G is a tuple <V, E>, where V denotes a set of vertices and E a set of edges.
- The vertices i∈V correspond to the genes (or other components of the system) and the edges correspond to their regulatory interactions.
- In a general way, one may express an edge as a tuple <*i,j, properties*>.
- Properties can indicate whether *j* activates (+) or inhibits (-) *i* .



$V = \{a,b,c,d\}$

$E = \{(a,a,+),(a,c,+),(b,c,+),(c,b,-),(c,d,-),(d,b,+)\}$

Description with:

Directed and Undirected graphs (2)

- Many databases are organized as richly annotated directed graphs (e.g., Transfac and KEGG).
- Directed graphs cannot represent the dynamics of a network, but they allow certain predictions about network properties:
  - Tracing paths between genes yields the sequence of regulatory events, shows redundancy in the regulation, or indicates missing regulatory interactions.
  - A cycle in the network may indicate feedback regulation.
  - Comparison of gene regulatory networks of different organisms may reveal evolutionary relations and reveal targets for pharmaceutical applications .
  - The network complexity can be measured by the connectivity

Description with: Bayesian networks (1)

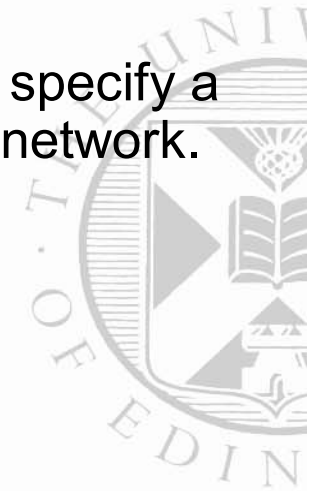
- A Bayesian network is based on the representation of the regulatory network as a directed acyclic graph G = <V, E>, where the vertices i $\in$ V represent genes and edges denote regulatory interactions.

- Variables x_i belonging to the vertices i denote a property relevant to the regulation, e. g., the expression level of a gene or the amount of active protein.

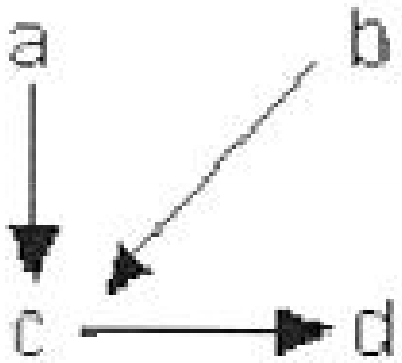
- A conditional probability distribution $p(x_i|L(x_i))$ is defined for each x_i, where $L(x_i)$ are the parent variables be the direct regulators of i.

- The directed graph *G* and the conditional distribution together specify a joint probability distribution $p(x)$ that determines the Bayesian network. The joint probability distribution can be decomposed into:

$$p(x) = \prod_i p(x_i \mid L(x_i))$$



Description with: Bayesian network (2)



Note:
some interactions are neglected:

- inhibition of b by c,

- activation of b by d

in order to get a network without cycles.

$p(x_a)$

$p(x_b)$

$p(x_c|x_a, x_b)$,

$p(x_d|x_c)$.

Description with: Bayesian networks (3)

- The directed graph expresses dependencies of probabilities: the expression level of a gene represented by a child vertex depends on the expression levels of genes belonging to the parent vertices.
- It also implies conditional independence *i* $(x_i; y \mid z)$, meaning that x_i is independent of the set of variables y given the set of variables z.

For the network given in Fig. 8.2c the conditional independence relations are $i(x_a; x_b)$ and $i(x_d; x_a, x_b | x_c)$. The joint probability distribution of the network is

$$p(x_a, x_b, x_c, x_d) = p(x_a) \cdot p(x_b) \cdot p(x_c | x_a, x_b) \cdot p(x_d | x_c).$$

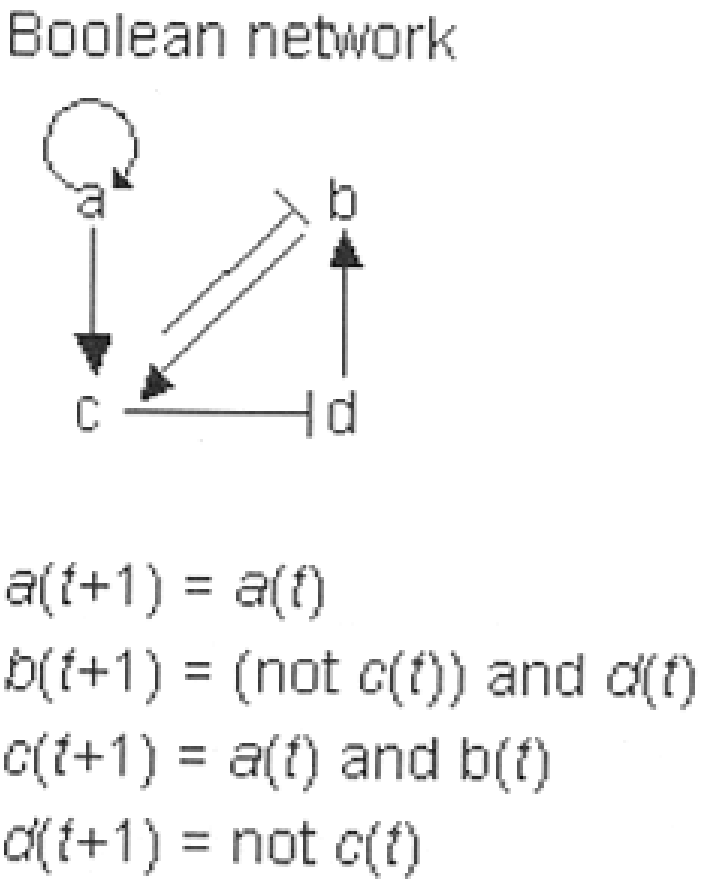
- Bayesian networks have been used to deduce gene regulatory networks from gene expression data. The aim is to find the network (or equivalence class of networks) that best explains the measured data.
- Another problem is the determination of initial probability distributions.

Description with: Boolean networks (1)

- In a Boolean network, the expression level of each gene is assigned to a binary variable: a gene is considered to be either on (1) or off (0) – (i. e., it is transcribed or not).
- The states of the genes are updated simultaneously in discrete time steps.
- The new state can depend on the previous state of the same gene or other genes. These dependencies cause the Boolean network.
- The following termini are used:
  - the N gene are the N nodes of the network,
  - the k interactions regulating the expression of a certain gene are the k inputs of that node,
  - the binary expression value of each gene is its output.

Description with: Boolean networks (2)

- Since every node can be in one of two different states, a network of N genes can assume 2^N different states. An N-dimensional vector of variables can describe the state at time t.
- The value of each variable at time t+1 depends on the values of its inputs and it can be computed by means of the Boolean rules.
- For a node with k inputs, the number of possible Boolean rules is 2^{2^k}.
- Although a Boolean network is a very simplified representation of the gene regulatory network, it enables a first computation of gene expression dynamics.



Description with: Boolean networks (3)

- The sequence of states given by the Boolean transitions represents the trajectory of the system. Since the number of states in the state space is finite, the number of possible transitions is also finite.
- Therefore, each trajectory will lead either to a steady state or to a state cycle. These states are called attractors.
- Transient states are those states that do not belong to an attractor.
- All states that lead to the same attractor constitute the basin of attraction.

Boolean networks (4)

For the network presented in Fig. 8.2d the following Boolean rules apply:

$$a(t+1) = f_a(a(t)) = a(t) \qquad \text{Rule 1 for } k=1$$

$$b(t+1) = f_b(c(t), d(t)) = (\text{not } c(t)) \text{ and } d(t) \qquad \text{Rule 2 for } k=2$$

$$c(t+1) = f_c(a(t), b(t)) = a(t) \text{ and } b(t) \qquad \text{Rule 2 for } k=2$$

$$d(t+1) = f_d(c(t)) = (\text{not } c(t)) \qquad \text{Rule 0 for } k=1$$

The temporal behavior is determined by the sequence of states (a, b, c, d) given an initial state (compare also Section 10.3.3).
From Tab. 8.1 it is easy to see that this network has two different types of stationary behavior. If the initial state of a is 0, then the system evolves towards the steady state 0101, meaning that genes a and c are off, while genes b and d are on. If the initial state of a is 1, then the system evolves towards a cyclic behavior including the following sequence of states: $1000 \rightarrow 1001 \rightarrow 1101 \rightarrow 1111 \rightarrow 1010 \rightarrow 1000$.

Tab. 8.1 Successive states in the Boolean network.

0000 → 0001	1000 → 1001
0001 → 0101	1001 → 1101
0010 → 0000	1010 → 1000
0011 → 0000	1011 → 1000
0100 → 0001	1100 → 1011
0101 → 0101	1101 → 1111
0110 → 0000	1110 → 1010
0111 → 0000	1111 → 1010

Reading

1. Tyson JJ, Chen KC, Novak B. - *Sniffers, buzzers, toggles and blinkers: dynamics of regulatory and signaling pathways in the cell.* - Curr Opin Cell Biol. 2003 Apr;15(2):221-31. Review. PMID: 12648679

✓ *This article is also a good introduction to the quest for a shared representational language that can help physiologists/biologists and mathematicians/computer scientists compare wet lab results and numerical models.*

2. E. Klipp, *Systems Biology in Practice, Wiley-VCH, 2005*
   - *Chapter 8*
   - *and chapter 3 for an introduction on maths*

Appendix: Hill coefficient (1)

Cooperative Ligand Binding Can Be Described Quantitatively

Cooperative binding of oxygen by hemoglobin was first analyzed by Archibald Hill in 1910. From this work came a general approach to the study of cooperative ligand binding to multisubunit proteins.

For a protein with n binding sites, the equilibrium of Equation 5–1 becomes

$$\mathrm{P} + n\mathrm{L} \rightleftharpoons \mathrm{PL}_n \qquad (5\text{–}12)$$

and the expression for the association constant becomes

$$K_\mathrm{a} = \frac{[\mathrm{PL}_n]}{[\mathrm{P}][\mathrm{L}]^n} \qquad (5\text{–}13)$$

The expression for θ (see Eqn 5–8) is

$$\theta = \frac{[\mathrm{L}]^n}{[\mathrm{L}]^n + K_\mathrm{d}} \qquad (5\text{–}14)$$

Rearranging, then taking the log of both sides, yields

$$\frac{\theta}{1-\theta} = \frac{[\mathrm{L}]^n}{K_\mathrm{d}} \qquad (5\text{–}15)$$

$$\log\left(\frac{\theta}{1-\theta}\right) = n \log [\mathrm{L}] - \log K_\mathrm{d} \qquad (5\text{–}16)$$

where $K_\mathrm{d} = [\mathrm{L}]^n_{0.5}$.

Images from:
D. L. Nelson, Lehninger Principles of Biochemistry, IV Edition – Chapter 5

Appendix: Hill coefficient (2)

Equation 5–16 is the **Hill equation,** and a plot of $\log [\theta/(1-\theta)]$ versus $\log [L]$ is called a **Hill plot.** Based on the equation, the Hill plot should have a slope of n. However, the experimentally determined slope actually reflects not the number of binding sites but the degree of interaction between them. The slope of a Hill plot is therefore denoted by n_H, the **Hill coefficient,** which is a measure of the degree of cooperativity. If n_H equals 1, ligand binding is not cooperative, a situation that can arise even in a multisubunit protein if the subunits do not communicate. An n_H of greater than 1 indicates positive cooperativity in ligand binding. This is the situation observed in hemoglobin, in which the binding of one molecule of ligand facilitates the binding of others. The theoretical upper limit for n_H is reached when $n_H = n$. In this case the binding would be completely cooperative: all binding sites on the protein would bind ligand simultaneously, and no protein molecules partially saturated with ligand would be present under any conditions. This limit is never reached in practice, and the measured value of n_H is always less than the actual number of ligand-binding sites in the protein.

An n_H of less than 1 indicates negative cooperativity, in which the binding of one molecule of ligand *impedes* the binding of others. Well-documented cases of negative cooperativity are rare.

Images from:
D. L. Nelson, Lehninger Principles of Biochemistry, IV Edition – Chapter 5