4 Network Motifs

4.1 Regulation networks

Cellular processes are fine-tuned by complex regulations e.g. signaling pathways, the transcription network, and circuits for cell cycle control, the regulation of growth or stress response. Signaling systems can be described as a network, with nodes representing biochemical substances or complexes (sometimes in different modifications). Transcriptional regulation of genes by transcription factors can be schematically represented by networks. In this case nodes represent proteins, and arrows display the bounding ability of transcription factors to the promoter region. The arrows can be quantified by gene input functions. The structure of the network is determined by binding sites (their sequences can be determined transcription factor binding can be examined by *in vivo* experiments) in the regulatory regions of the genome.

An arrow in the network declares that a substance affects another substance (e.g. it catalyses the production /degradation).The inputs of multiple arrows pointing to a single node have to be processed and this can be described (simplification!) by boolean functions (e.g. logical AND or OR). (Substance levels are often balanced by opposing processes like synthesis and degradation.)
Signaling pathways detect input stimuli and translate the information (concentrations, modifications, localization of proteins, e.c.) to output signals effecting downstream processes, (e.g. processes like gene expression). For this reason they can be looked upon as information-processing devices. The ***input-output relation*** of a signaling system can conduct information-processing tasks such as classification, regression, data compression or transduction of signals.

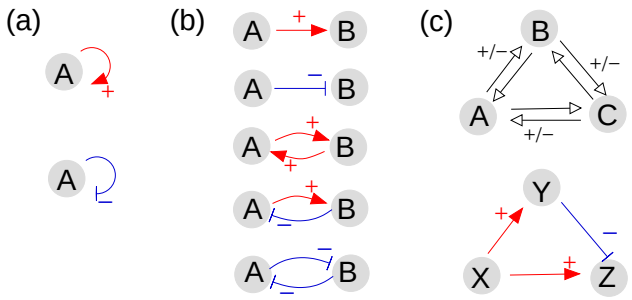

Figure 1: Regulatory patterns with one, two or three nodes. (a) Autoregulation: positive = red; negative = blue (b)A Model with two nodes (c) Models with three nodes can have 6 arrows. Below three arrows are selected, one obtains the inconsistent feed-forward loop type I(this is found as a network motif in transcription networks) (for further information see Figure 6).

The network structure is not random, e.g. the transcription network of *E. coli* contains ***dense overlapping regulons***, strongly interconnected subnetworks that respond to a set of input stimuli and control the expression of functionally related genes. The negative autoregulation or the feed-forward loop (see Figure 5) are examples for typical network motifs (see below). They often appear in clusters, which have been described as generalized motifs.

4.2 Erdös-Renyi random graphs and network motifs

Patterns that appear much (=significantly) more often than expected by chance are called network motifs. To declare what is significant, we need a null model, a graph in which structure only appear "by chance". A simple possibility are Erdös-Renyi random graphs.

We consider N nodes with random connections between them. Each possible edge is realised independently with probability p. After deciding about each possible edge, one obtains a graph, which is a realisation of

the random graph. Alternatively, one may allow for exactly k edges and distribute them randomly. For large graphs, both random graphs have similar statistical properties.
A random graph with N nodes has $N * (N-1)$ divided edges among different nodes and N self-edges $\Rightarrow$ And we get N^2 possible edges.

If k edges are realised, the probability for an possible edge is $p = \frac{k}{N^2}$. Random graphs can serve as a null hypothesis to decide if certain propertis of natural networks are statistically significant.

Random graphs with preserved degrees Random graphs with predefined degrees (# of incoming and outgoing edges per node) can be obtained by repeated flipping of (randomly chosen) edge pairs:
(see Fig.2.)

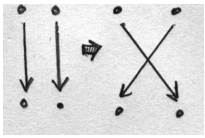

Figure 2: This operation preserves all degrees, but randomises other statistical properties of the graph.

4.3 Example: self-regulation as a network motif

A transcription network (E.coli) contains $N = 424$ nodes, $k = 519$ edges and 40 self-edges. Is this number significantly high? (In other words: is self-regulation a network motif?)

To decide this, we count the self-edges in a random graph (3) with the same numbers of nodes and edges.

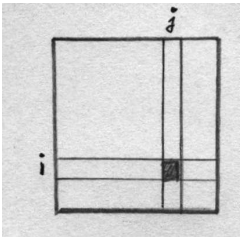

Figure 3: The adjacency matrix a_{ij} represents the structure of a graph. A value $a_{ij} = 1$ indicates an edge from the node j to node i. If k entries (= k edges) are randomly distributed the number of self-edges (entire on the diagonal) is approximately binomially distributed, prob$(n) \approx \binom{k}{n} p^n (1-p)^{k-n}$, where $p = \frac{1}{N} = \frac{N}{N^2} = \frac{\text{number of diagonal elements}}{\text{number of all matrix elements}}$.

The number of self-edges in the random graph has the mean value $\langle n \rangle = \binom{k}{N}$ and standard deviation $\sigma_n = \sqrt{\frac{k}{N}}$. With the above numbers, we obtain the Z-score:

$$\frac{n_{observed} - \langle n_{random} \rangle}{\sigma_{n,random}} \approx 32, \text{which is highly significant.}$$

Self-regulation is a network motif in the transcription networks, i.e. a local structure that appears significantly more often than in a comparable random network. The frequence occurence may limit at a selection pressure that led to the evolution and preservation of the network structures. The degree of significance depends on the random graph considered.

Possible functions of the self-regulation:

- self-activation may lead to bistable behavior

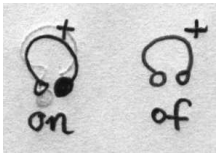

Figure 4: self-activation=left; self-inhibition=right

- self-inhibition may stabilize the node at a certain expression value, or speed up its response.

Efficient and fast response A fast response can be important for cells: for instance, if the input signal indicates that a nutrient is available, while the output starts the production of enzymes needed to use this nutrient. The response time in the linear scheme with linear kinetics is determined by the degradation constant k_2: a way to speed up the response would be to increase the turnover; but in order to reach the same steady state level r^{steady} in the model, also the synthesis rate v needs to be increased. Therefore, a fast response would be paid by an permanent higher production of R, which costs energy and material. Negative feedback can speed up the response without this disadvantage: at low levels of r, the value of v is high; later, self-inhibition kicks in, synthesis occurs at a lower rate, and the system relaxes towards the steady state as before.

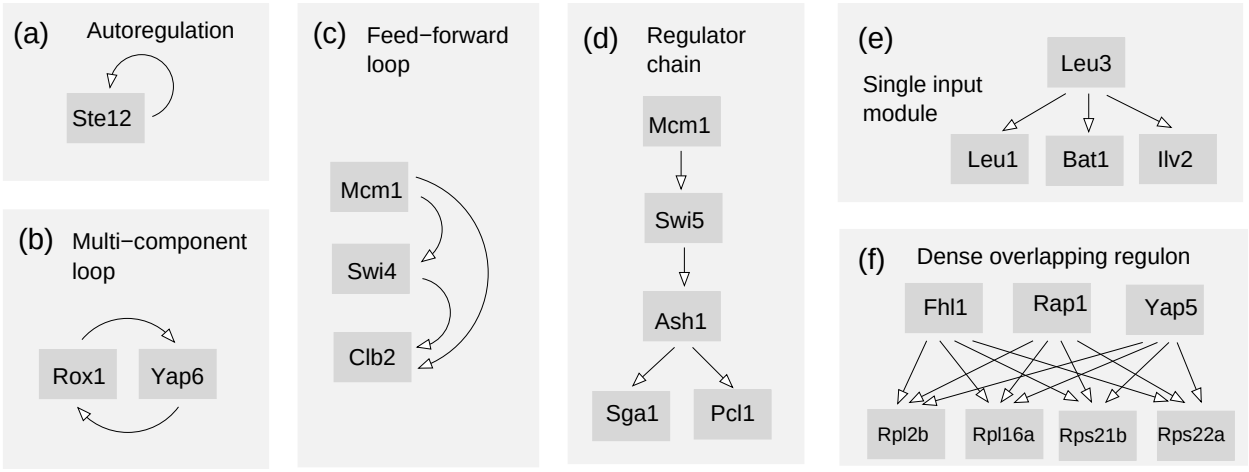

Figure 5: Network motifs in the transcription network of the yeast *S. cerevisiae*. After Lee et al.

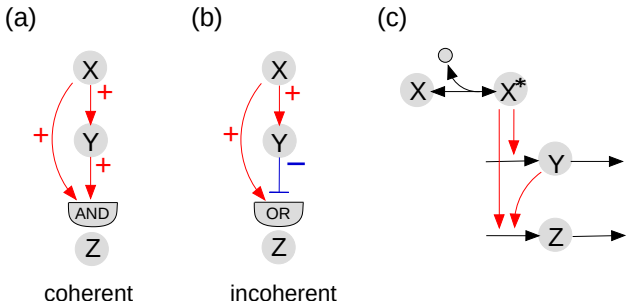

Figure 6: Feed-forward loops. There is a direct regulation of the output gene Z by the input gene X via Y: (a) Coherent feed-forward loop type 1 (with AND gate). (b) Incoherent feed-forward loop type 1 (with OR gate). (c) Possible realisation by a kinetic model. X is activated (X*) by a small effector molecule.

4.4 The Feed-forward Loop

The feed-forward loop (FFL) is a common motif in transcription networks, that consits of three interacting genes: an input gene X regulates the output gene Z via an intermediate gene Y. Each edge can be positive or negative. The two inputs of Z can be described by boolean functions like the logical AND or OR (examples shown in Fig. 6).

The transcription network in E.coli ($N = 424$ genes, $k = 591$ nodes) contains 42 FFL, and no feedback-3-loop. In an Erdös-Renyi random graph, we expect about 0.6 ± 0.8 FFL.

In a ***coherent FFL type 1***, all regulations are activating, while in an ***incoherent FFL type 1***, Y inhibits Z. (Other types of feed-forward loops, with different sign combinations, occur rarely and will not be considered here.)

At first sight, the function of the branch via gene Y in the feed-forward loop is not obvious: in the coherent FFL, the two branches seem redundant, and in the incoherent FFL, they neutralize each other. But like the adaption motif, this argument holds only in steady-state situations. If (for an incoherent FFL type 1) the input X is suddenly switched on, gene Y turns up with a delay. That means Z is first activated via the direct branch and later inhibited again by Y. Furthermore a step in the input X is translated into a peak-like behavior of Z.

Dynamical models and measurements in gene circuits in *E. coli* have shown that feed-forward loops can (i) serve as sign-sensitive delays, (ii) generate temporal pulses, and (iii) accelerate the response to an input signal. This behavior depends on the kinetic parameter or the boolean paradigm.

4.5 Dynamic Model of the Feed-Forward Loop

To translate the boolean structure of a feed-forward-loop into a simple kinetic model, we assume that gene X is expressed constitutively and its activity x is controlled by the concentration of a ligand, whereas the activities of Y and Z are determined by their expression.
If transcription and translation are packed into a single step, we get the rate equations

$$\frac{dy}{dt} = f_{\mathrm{y}}(x) - \beta_{\mathrm{y}}\, y \tag{1}$$

$$\frac{dz}{dt} = f_{\mathrm{z}}(x, y) - \beta_{\mathrm{z}}\, z. \tag{2}$$

y and z denote protein levels,whereas f_{y} and f_{z} are the production rates, and β_{y} and β_{z} are degradation constants.

To study the dynamic behavior, it is practical to describe the transcription of Y by a step-like gene input function:

$$f_{\mathrm{y}}(x) = \alpha_{\mathrm{y}}\, \Theta(x > x_0). \tag{3}$$

The truth function $\Theta(\cdot)$ produces a value of 1 if the inequality in the argument is satisfied, otherwise it is 0. As long as x stays below the threshold value x_0, Y is not transcribed; otherwise, Y is transcribed with constant rate α_{y}.

Let us consider two cases, (i) a coherent FFL with a logical AND input function for the production of Z, and (ii) an incoherent FFL with a logical AND function. The corresponding input functions for gene Z are:

$$\begin{aligned} \text{coherent, AND:} \quad & f_{\mathrm{z}}(x, y) = \alpha_{\mathrm{z}}\, \Theta(x > x_0 \text{ AND } y > y_0) \\ \text{incoherent, AND:} \quad & f_{\mathrm{z}}(x, y) = \alpha_{\mathrm{z}}\, \Theta(x > x_0 \text{ AND NOT } y > y_0). \end{aligned} \tag{4}$$

Figure (7) shows simulation results from model (of equation 1), with input functions (3) and (4) and a predefined, pulse-like input $x(t)$.
Characteristic features of thefeed-forward loop are:

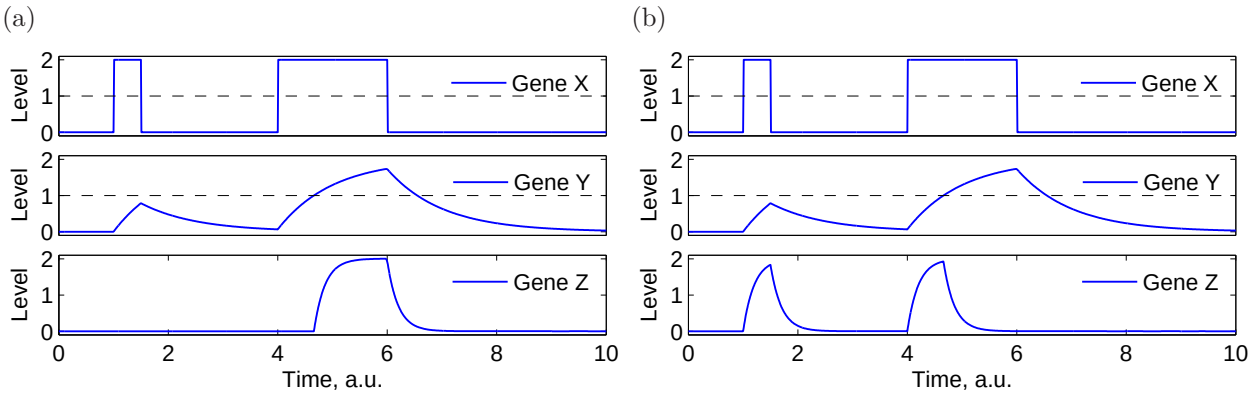

Figure 7: Dynamic behavior of two types of feed-forward-loop (FFL). (a) Coherent FFL type 1 with AND logic (see Fig. 6, (a)). Time curves for active input X (top), intermediate gene Y (center), output Z (bottom) are shown (arbitrary units). The FFL filters out the short pulse; the response to the longer pulse is delayed, but the response to the end of the pulse is immediate. (b) Incoherent FFL type 1 with AND logic (see Fig. 6, (b)). The onset of each input pulse leads to a pulse of Z of fixed maximal length.

- The **coherent-AND FFL** shows a delayed response to the onset and an immediate response to the end of pulses. For this reason short pulses are filtered out.

- The **incoherent-AND FFL** responds immediately to an input pulse, but this response is switched off only after a while. In return input pulses are translated into standard pulses of similar length (pulse generator).

4.6 Single-input-module

The basis for this model is the assumption of a co-regulation by a single signal (occurs, for instance, in metabolic pathways with product inhibition) (see Fig.9).

Therefore temporal gene expression programs offer the advantage to regulate the strength and chronology of enzyme activation/deactivation.

'First in, last out'-production. The genes are first activated one after the other and later inactivated in reverse order. (Fig.10). This is used by the cell for a just-in-time production of e.g. important metabolites.

4.7 Multi-Z-FFL

The priciple of this kind of regulation are coherent FFL that reacts with an or-logic and reversed threshold values (see Fig. 11).

This means there is not a single co-regulating signal (Single-Input-Module principle). The gene that is activated at first is inactivated at first, too. (Fig.12)⇒ **'first in, first out'-production**. This can be helpful if parts of a protein complex (e.g., bacterial flagellae) are produced in a certain order, but with approximately fixed stoichiometry.

4.8 Adaption Motif (*)

An important property of signaling systems is their transient response to changing inputs. The system shown in Figure 13 has a remarkable property called ***precise adaption***: after a jump of the input value, it

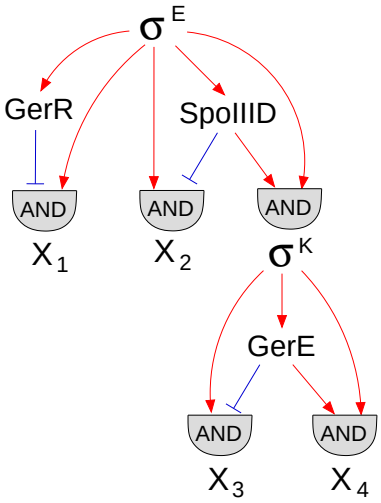

Figure 8: Gene regulation network coordinating sporulation in the bacterium *B. subtilis*. Some microbes can transform themselves into spores, which are much more resistant to adverse environmental conditions than the livin bacteria. When sporulation is triggered in the soil bacterium *B. subtilis*, a gene regulatory network produces several waves of gene expression, in which many genes are regulated in a temporally coordinated manner. The system contains a number of feed-forward loops that activate the downstream target genes. Activation of the Master regulator σ^{E} triggers waves of expression in different groups of target genes (denoted by X_1, X_2, X_3, X_4). After Eichenberger et al.

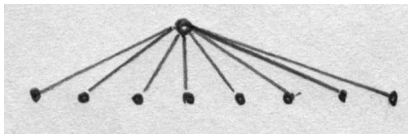

Figure 9: Co-regulation by a single signal.

shows a transient dynamics, but in the long run, it always returns to exactly the same steady state. Precise adaption combines sensitivity to temporal changes with robustness against the baseline value and plays a vital role in the bacterial chemotaxis.

In the ***adaption motif*** shown in Figure 13, the input X activates the production of Z, but inhibits it again via activation of Y. With mass-action kinetics and linear activation, we get the equations:

$$\frac{dy}{dt} = \alpha_y\, x - \beta_y\, y \tag{5}$$

$$\frac{dz}{dt} = \alpha_z\, x - \beta_z\, y\, z.$$

For x>0, they lead to steady state:

$$y^{st} = \frac{\alpha_{\mathrm{y}}}{\beta_{\mathrm{y}}} x, \qquad z^{st} = \frac{\alpha_{\mathrm{z}}\, \beta_{\mathrm{y}}}{\beta_{\mathrm{z}}\, \alpha_{\mathrm{y}}}. \tag{6}$$

The steady-state level of Z depends only on kinetic constants. In steady state, the activation and inactivation are balanced. When the input suddenly changes, activation responds faster, which leads to a transient peak (Fig. 13 (b)).

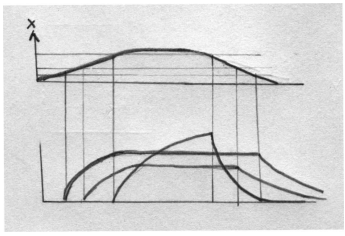

Figure 10: Graph of a 'first in, last out'-production

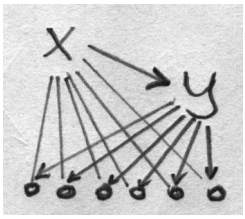

Figure 11: Principle of a multi-z-feed forwrd loop

4.9 Negative Feedback (*)

Negative feedback is very common in transcription networks and it is also important for metabolic pathway regulation. For instance, to prevent overproduction and stabilize the amount of amino acids, the first enzyme of amino acid synthesis is often inhibited by the final product.
The effect of negative feedback on cellular dynamics can be: (i) the stabilization of a state of acellular network; (ii) the reduction of the variance of fluctuations and the variability of steady states; (iii) the production of pulse-like overshoots; (iv) the induction of sustained oscillations; (v) and the acceleration of response times.

In the model (see Fig. 14), all reactions follow irreversible mass-action kinetics $v_i = k_i\, s_{i-1}$, the external substrate level $s_0 = 1$ is kept constant, and all other metabolites start at levels $s_i = 0$. The first reaction is inhibited allosterically by one of the downstream metabolites. The feedback inhibition by the n^{th} metabolite can be implemented as $v_1 = \frac{s_1 k_1}{(1+s_n/K_1}$).

Without feedback, metabolite concentrations reach a steady state after a short transition period (Fig. 14 (a)).

By adding an inhibition by the second metabolite, the level of the first metabolite shows an overshooting response (Fig. 14 (b)) and a inhibition by the last metabolite leads to damped oscillations(Fig. 14 (c)).

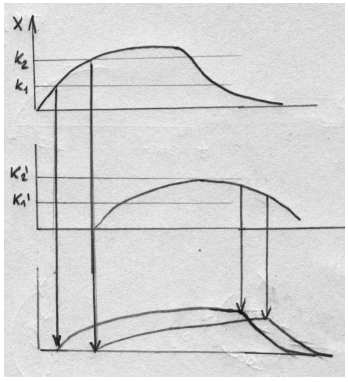

Figure 12: Graph of a 'first in, first out'-production

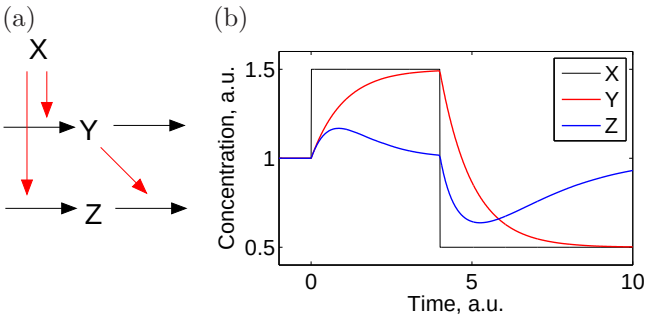

Figure 13: The adaption motif. (a) The signal X catalyzes the production of Y and Z;and Y catalyzes the degradation of Z (b) Temporal profile of the adaption motif: a step-like input level x=black induces a sustained response of y=red; z is the output level = blue and shows a transient response before returning to its steady state value

The example also shows that negative autoregulation can speed up the system's response to external changes. The response time $\tau_{(1/2)}$, defined as the time at which the last metabolite S_r reaches its half-maximal level, decreases from a) to c) shown in Fig. 14.

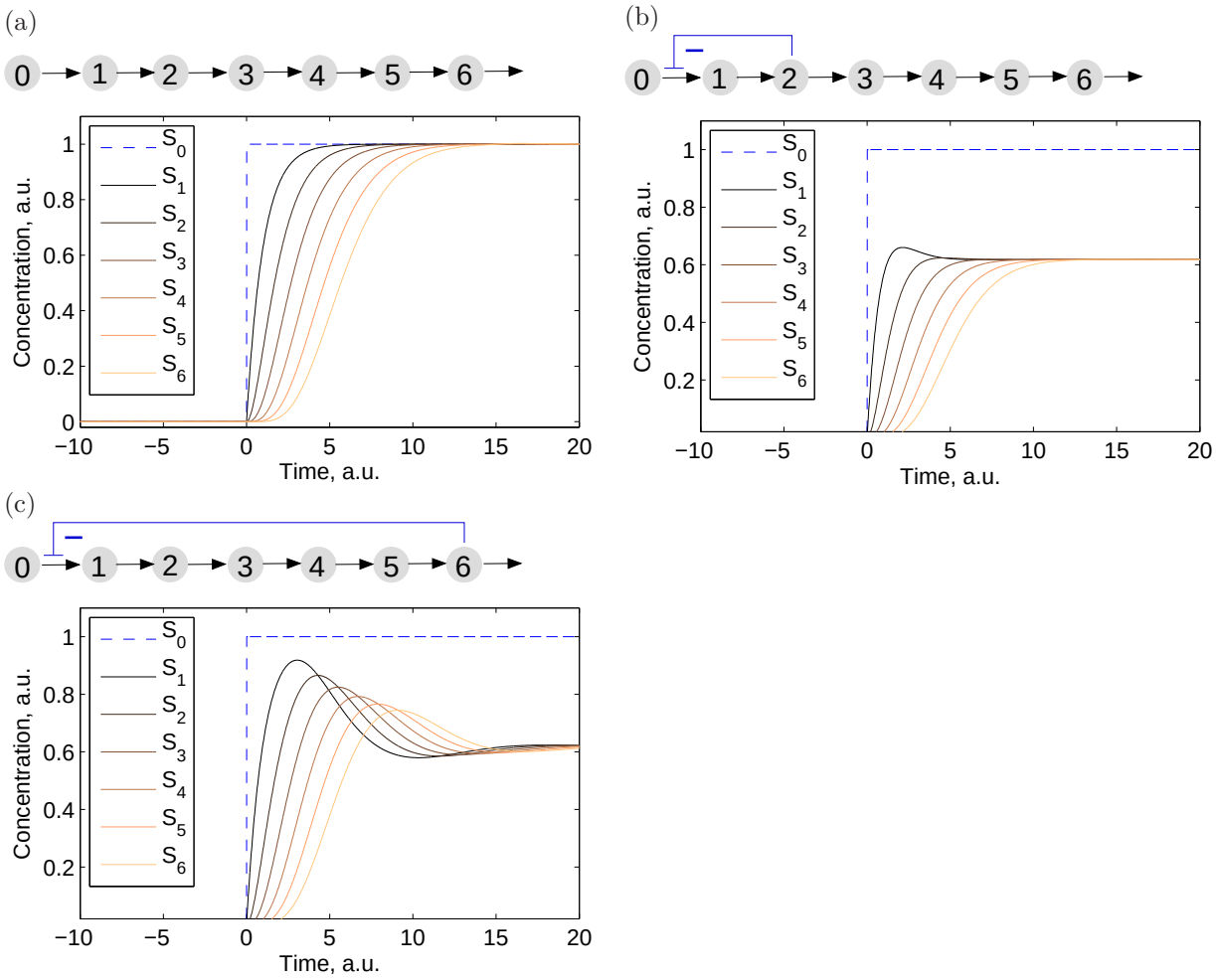

Figure 14: Negative feedback in an unbranched metabolic chain. (a) Concentration time series [(- -) is the external substrate ($s_0 = 1$) that becomes available at time $t = 0$]. (b) Negative feedback by the second metabolite (c) Negative feedback by the last metabolite -¿overshoot and damped oscillations.