Cellular Reprogramming and Controllability of Complex Systems

University of Maryland and ISR

April 25, 2016

Roger Brockett

John A. Paulson School of Engineering and Applied Science

Harvard University

Metadata on the Talk

In this talk I will give a brief description of certain aspects of cell biology and use these to define some problems related to the control of the fate of cells. The centerpiece here is the problem of cellular reprogramming, e.g. steering a group of skin cells so that they become skeletal muscle cells.

The Background: This talk is based on a ongoing, multi-person, experimental and theoretical research effort centered at the University of Michigan, under the leadership of Indika Rajapakse with DARPA sponsorship.

Expectations: I intend to sketch aspects of a specific "complex system" and describe
some of the ways that are being used to influence the evolution an ensemble of simpler units. I will put disproportionate emphasis on one or two questions which can be reduced to a clean mathematical form. This is not meant to mislead but rather to suggest to engineers that "Systems Biology" is not exclusively for others.

About the Talk

We first define the problem of cellular reprogramming and explain how one might think of it in mathematical terms, including a suggestions about ways to influence the evolution of cell type (control inputs).

After that we discuss periodicity as a central feature of cell dynamics and introduce a class of mathematical models of the type that have been shown to support oscillations. .

Given the large number of nonlinear processes involved, it is hardly surprising that looking for simplified models suitable for capturing the features of the system that will facilitate reprogramming leads in a different direction from traditional model reduction and linear system identification.

Why reprogram a cell?

There are several compelling reasons for wanting to reprogram a colony of cells. Two especially compelling ones are:

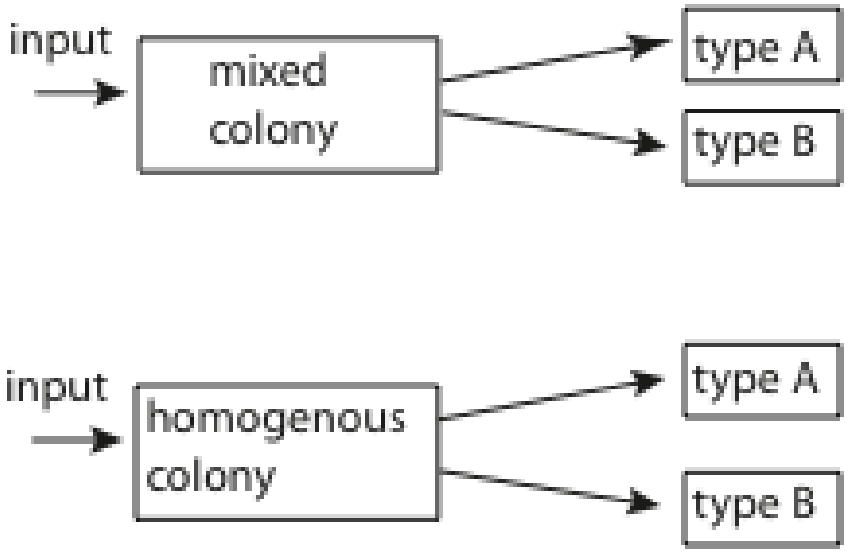
1. Cancer cells have been accidently reprogrammed by possible exposure to a carcinogen or radiation or some other insult. A possible path to recovery is to reprogram the cells, restoring the cell line to a healthy state.

2. As a way to modify skin cells to create replacements for damaged tissue by reprogramming the skin cells to prepare them for a new role such as a role in muscle tissue.

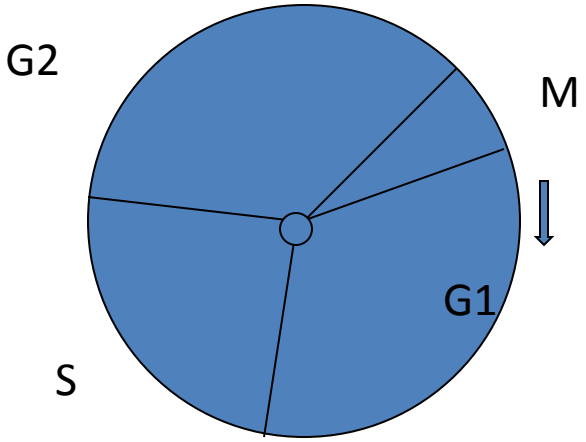
Dealing with colonies, not individual cells

As in certain other areas that have recently become popular topics, when control is applied to cell biology it usually involves controlling populations, not individual cells. This kind of ensemble control occurs in the control of swarms, quantum systems, demographics, etc.

Sometimes a stimulus is applied to a mixed colony of cells with the goal of accentuating the differences. Other times the goal is to create change in an nearly homogenous colony, accentuating normal differences.



The Standard Picture of a Proliferating Cell Cycle



G1 phase. Metabolic changes prepare the cell for division. At a certain point - the restriction point - the cell is committed to division and moves into the S phase.

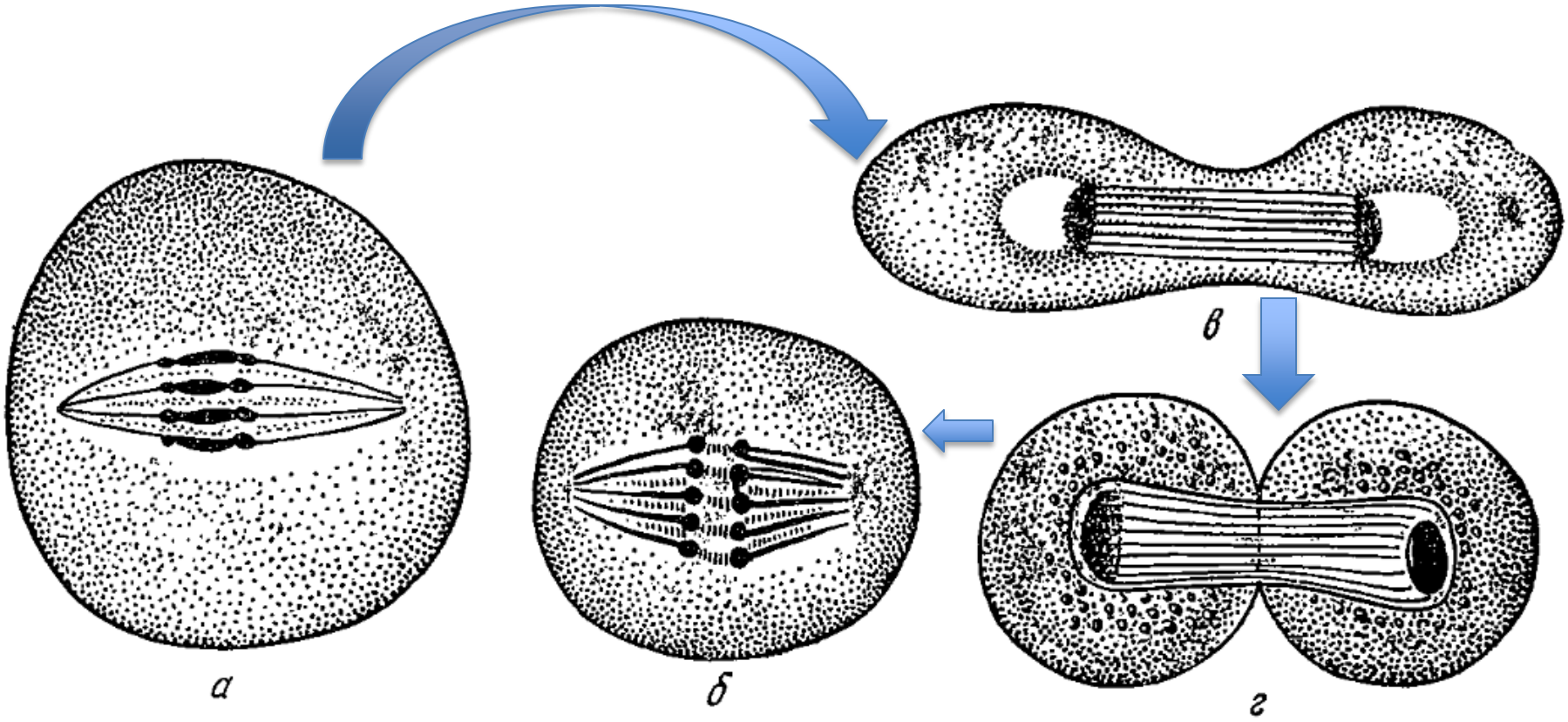
S phase. DNA synthesis replicates the genetic material. Each chromosome now consists of two sister chromatids.

G2 phase. Metabolic changes assemble the cytoplasmic materials necessary for mitosis and cytokinesis.

M phase. A nuclear division (mitosis) followed by a cell division (cytokinesis).

Across species, the period of the cell cycle varies but often it is on the order of 24 hours.

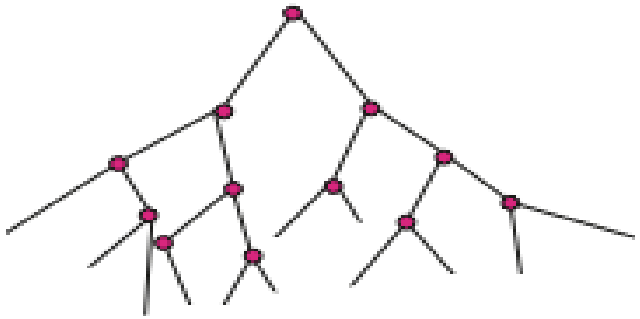
From the web



Mathematical Challenge: The cycle "repeats" but not the way earth revolving about the sun does!

The process of moving around the cycle is first to prepare the cell for division by processes involving generation of proteins, lipids, and genetic material and processes related to the geometric reconfiguration of the internal elements. When division occurs two daughter cells are generated. These may or may not be functionally different from the mother and/or each other.

Nature orchestrates the processes involved so as to generate daughter cells of a desired type. Determining the appropriate way to introduce steering (control) has been the subject of study for several decades. The possibility of using quantitative system identification techniques to gain more insight into how this can be done effectively and efficiently is more recent.



Engineering analogy: under stress the electric power grid may lose synchronization and break into islands

$$\dot{x} = f(x) \quad \Longrightarrow \quad \begin{aligned} \dot{x}_1 &= f_1(x_1) \\ \dot{x}_2 &= f_2(x_2) \end{aligned}$$

The initial conditions for the divided system come from the pre breakup state of the original system. There does not seem to be a general theory which predicts if and when a breakup will occur. This seems to be a more complicated version of the problem of computing the domain of attraction. Perhaps this is an area in which interesting developments can be expected in the near future.

Cell Types: From stem cells to differentiated cells

There are approximately 200 distinct cell types in the human body. These include various types of skin cells, muscle cells, nerve cells red blood cells, etc. In addition there are the undifferentiated cells called stem cells. Of course all cells from a particular person have the same DNA but the different cell types make use of the DNA in different ways. Stem cells are said to be pluripotent whereas the specialized cells are said to be differentiated.

Importantly, all the cells in a particular animal have the same DNA. The differences between the differentiated cells arise from epigenetic factors such as differences in the way that the DNA is folded or modified by attached proteins. More on epigenetics later.

A little vocabulary

DNA provides a cookbook but does not specify what is to be cooked in a given situation.

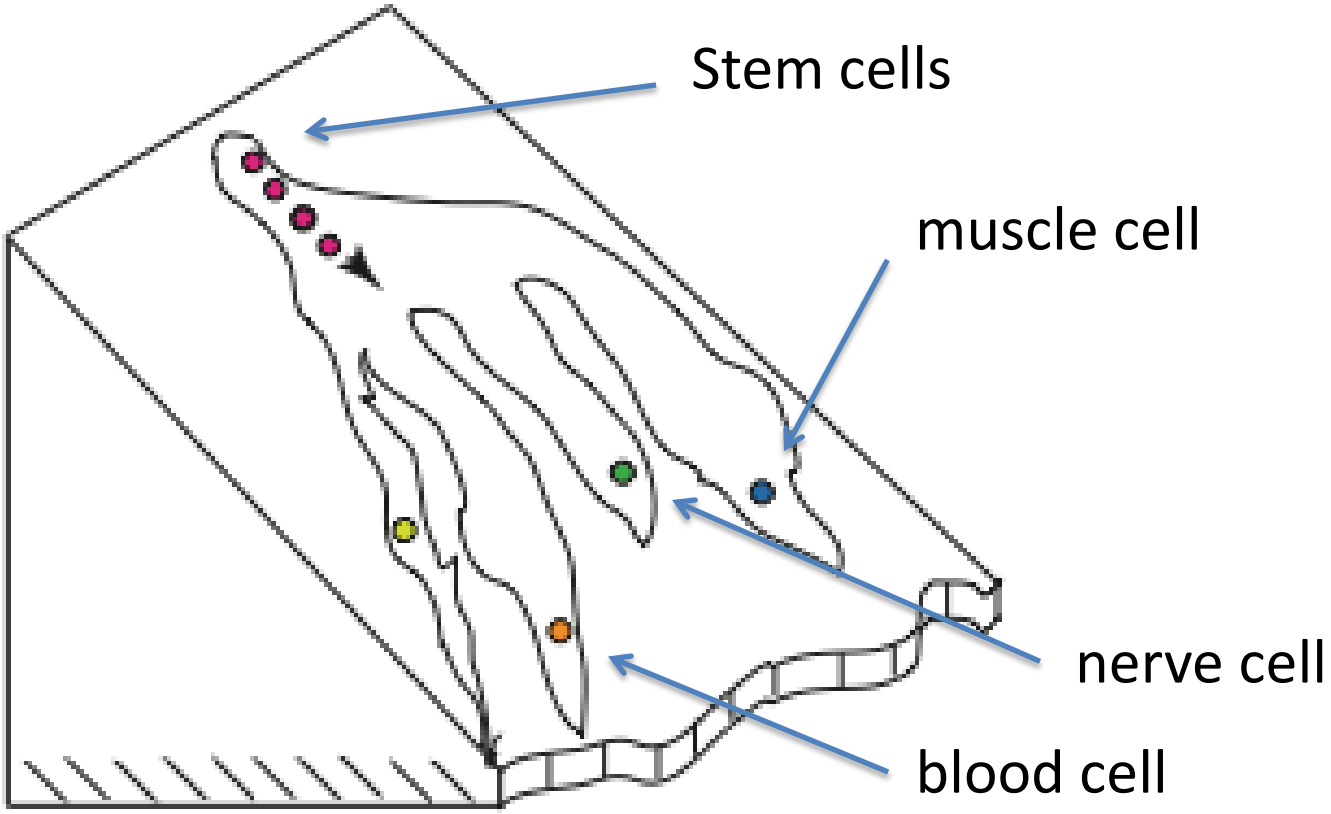
A **transcription factor** (or DNA-binding **factor**) is a protein that binds to specific DNA sites, thereby controlling the rate of **transcription** of genetic information from DNA to messenger RNA, ultimately, controlling the rate at which specific proteins are created.

A possibly helpful analogy

DNA	=	cookbook
transcription factor (s)	=	bookmark (s)
RNA	=	gathering the ingredients
protein production	=	cooking and serving the result
protein degradation	=	cleaning up

Transcription can regulate protein production up or down.

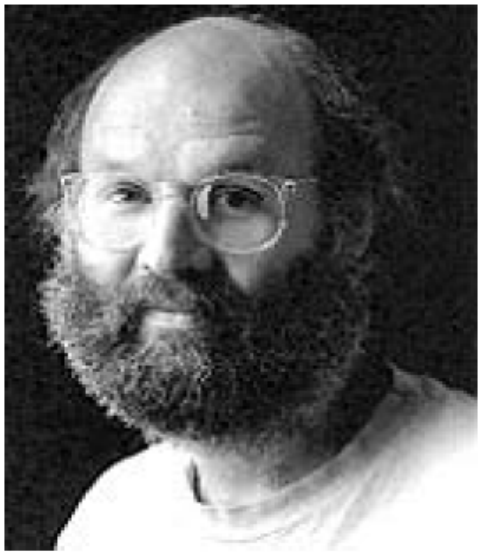
The Waddington (1953) picture: Falling into a cell type



In this picture, a cell type corresponds to a locally stable equilibrium. Stem cells lie at the highest potential with differentiated cells further “down the valley”.

Demonstrating that reprogramming is possible

Around 1990 H. Weintraub (L) successfully reprogrammed human skin cells turning them into muscle cells via the over- expression of one particular transcription factor, MyoD, thus becoming the first to demonstrate that the natural course of cell development and differentiation could be altered. In 2007, Yamanaka (R) et al. changed the paradigm further by successfully reprogramming human skin cells to embryonic-stem-cell-like state using the four TFs {Oct4, Sox2, Klf4, Myc}, showing that the cell state could even be pushed back to a pluripotent state [2].



How well can the process be understood?

The pioneering work was based on a deep understanding of existing results and considerable trial and error, particularly in the case of Yamanaka. In attempting to optimize the efficiency of the reprogramming process it would be helpful to have a list of the potentially useful control inputs and a quantitative model of the system. However the system is complicated!

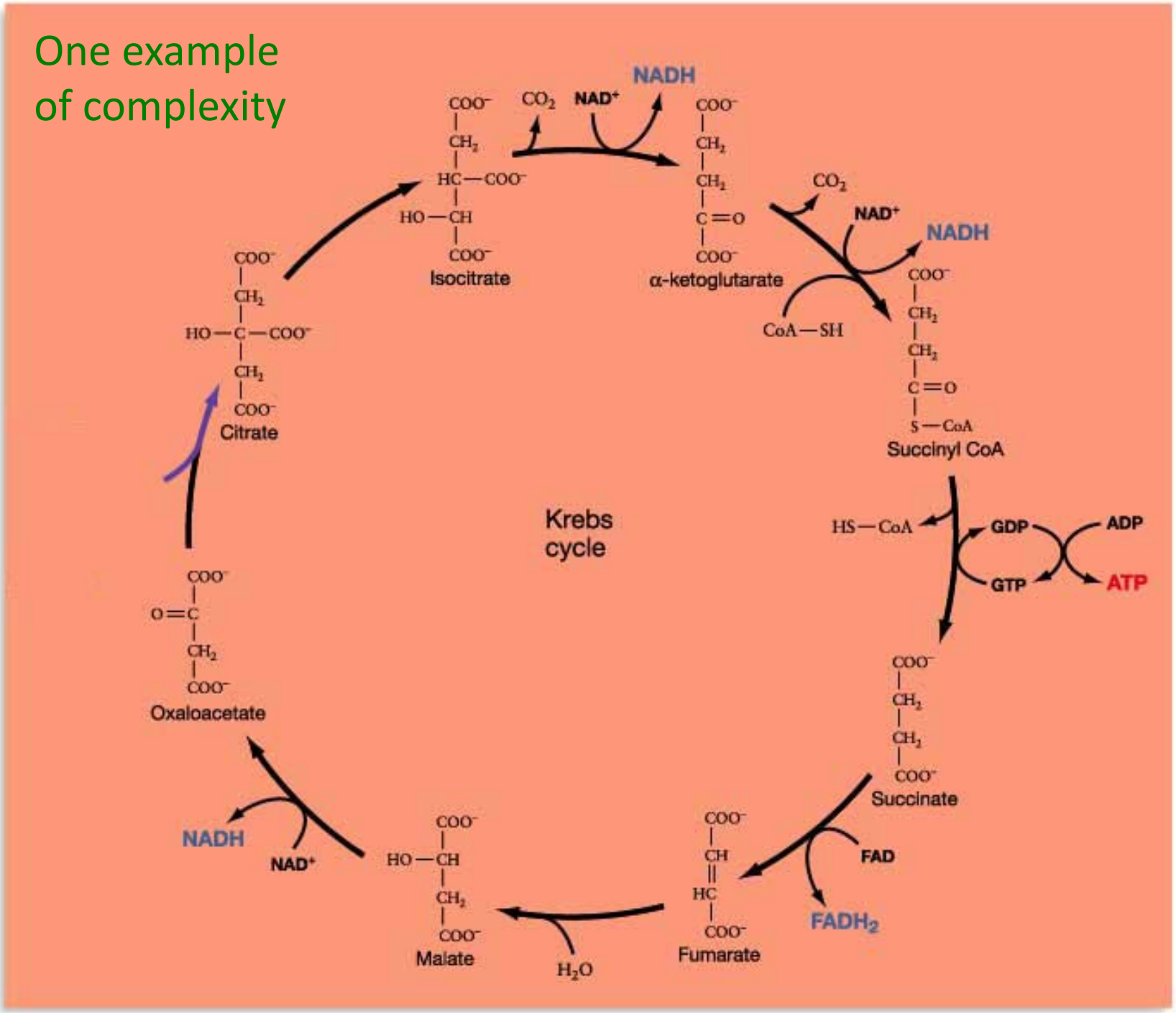
Quoting from: A Different Way of Doing Things

From The Scientist April 1, 2016

By Kivanç Birsoy and David M. Sabatini

"Cellular metabolism comprises an elaborate network of thousands of biochemical reactions that allow a cell to grow, divide, and respond to its environment. More than 100 years of research has identified some 3,000 enzymes and nutrient transporters, but only recently has it become clear that cancer cells exploit these metabolic components to support their own proliferation and survival."

One example
of complexity



Data for modeling the inner workings of cells

There are a growing number of ways to probe cells so as to learn more about the dynamical relationships that they utilize. One dominant feature is the presence of nearly periodic trajectories. For the most part these have periods of about 24 hours. By modeling and then measuring these one can hope to better evaluate the gains and time constants of various components. Another way is to refine models is to measure the steady state change in the state that is generated by a constant input. Such experiments can provide some quantitative information about cell dynamics, even though the in is a steady state measurement.

Organization based on Oscillations

Quoting from: A circadian gene expression atlas in mammals: …. Ray Zhang et al. PNAS, 2014.

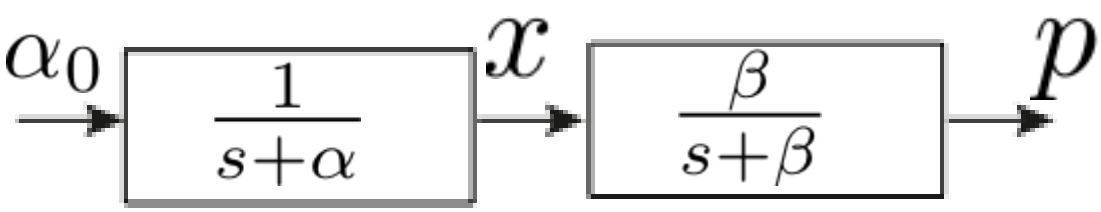
"We generated high-resolution multi-organ expression data showing that nearly half of all genes in the mouse genome oscillate with circadian rhythm somewhere in the body. Such widespread transcriptional oscillations have not been previously reported in mammals. Applying pathway analysis, we observed new clock-mediated spatiotemporal relationships. Moreover, we found a majority of best-selling drugs in the United States target circadian gene products. Many of these drugs have relatively short half-lives, and our data predict which may benefit from timed dosing."

Oscillations and synchronization

As noted, there are thousands of individual processes involved in the maintenance and growth of individual cells. Some of these processes can be described as ordinary chemical reactions while others involve transcription and thus differ from the reactions taught in high school chemistry. These processes must be sequenced appropriately in normal operation. An important part of the mechanisms for doing this are the various biological oscillators existing it the cell. Many details about how they work together remain unknown but most operate with a period of about 24 hours. These are not simply the response to an external periodic input but are self sustained but with the ability to adjust so as to synchronize with the **circadian** oscillation.

Oscillations involving gene expression

The greater the expression of a given transcription factor the higher the rate of generation of the corresponding protein. If this were all there was to it we would have a protein concentration p and a gene expression level x related as in



second order but more like heat flow equations

$$\dot{x} = -\alpha x + \alpha_0 \qquad \dot{p} = -\beta p + \beta_0$$

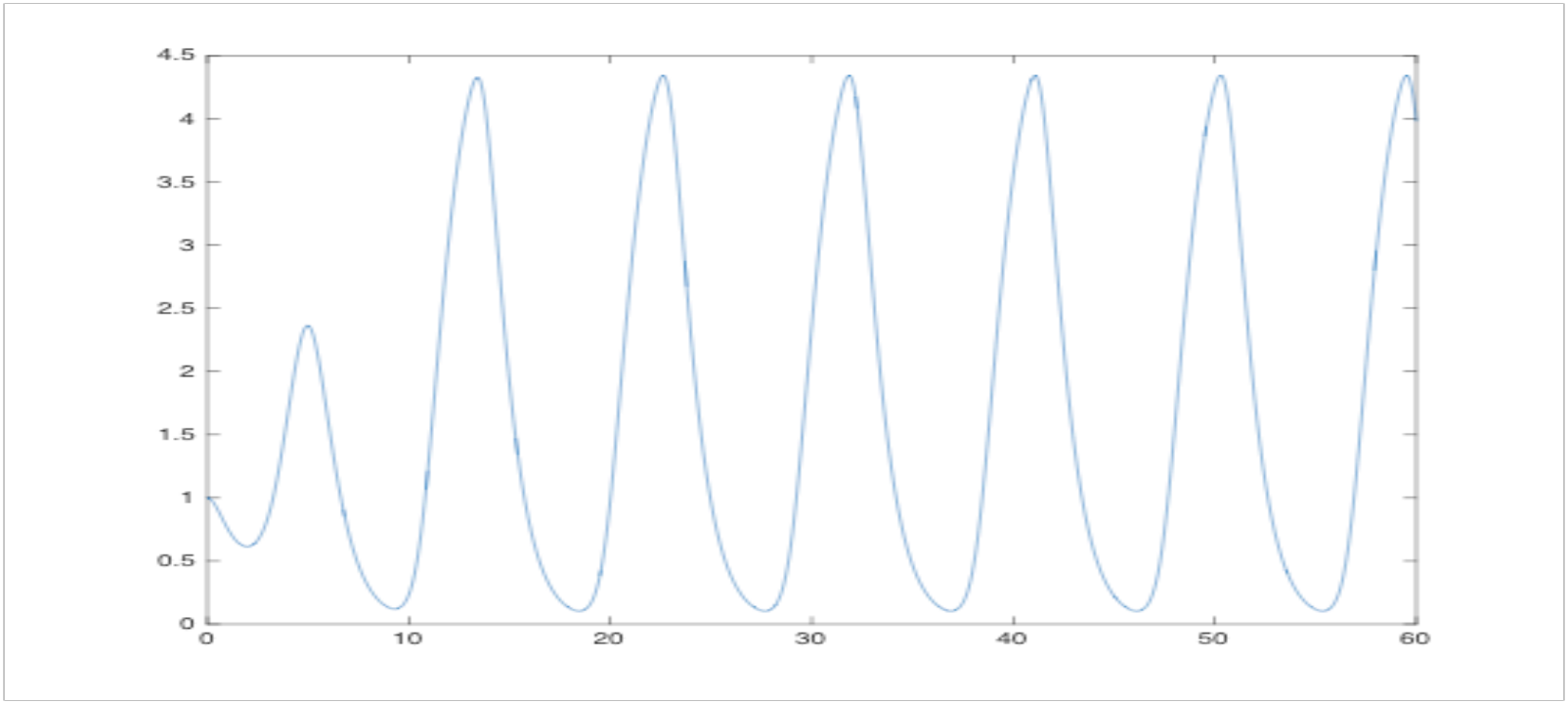
However, there can be cross coupling of relationships of this kind involving the suppression of of the value of α_0

The (simplified) repressilator model

$$\ddot{p}_1 + (1+\beta)\dot{p}_1 + \beta p_1 = \frac{\alpha\beta}{1+p_3^\gamma} + \beta\alpha_0$$
$$\ddot{p}_2 + (1+\beta)\dot{p}_2 + \beta p_2 = \frac{\alpha\beta}{1+p_1^\gamma} + \beta\alpha_0$$
$$\ddot{p}_3 + (1+\beta)\dot{p}_3 + \beta p_3 = \frac{\alpha\beta}{1+p_2^\gamma} + \beta\alpha_0$$

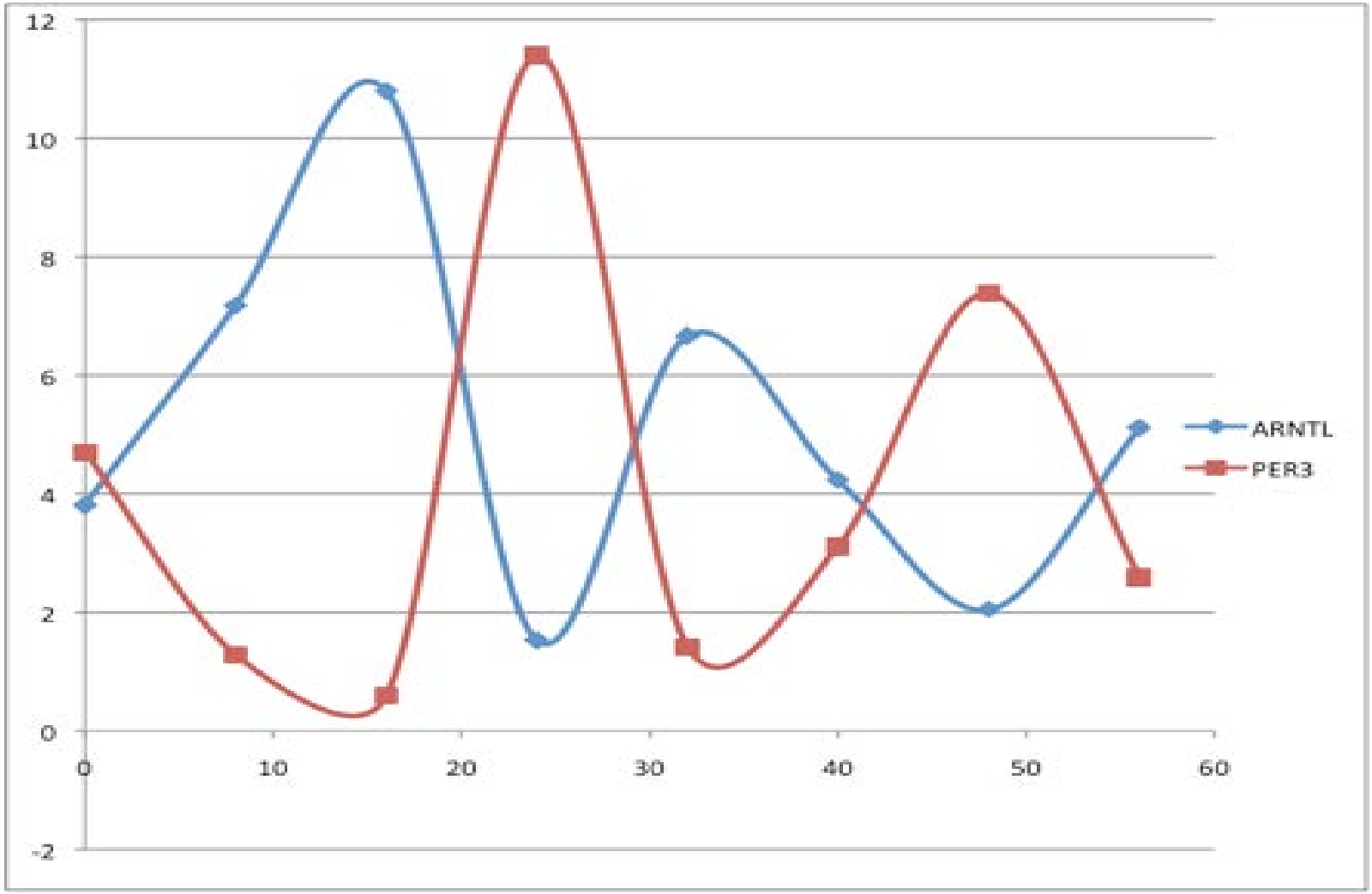
The nonlinearity comes from chemical kinetics (Hill's function). These equations can have a constant solution and/or oscillatory ones.

One solution of the repressilator equations



Oscillations are not well approximated by sine waves so harmonic balance is not effective and phase is undefined.

Data from Rajapakse lab

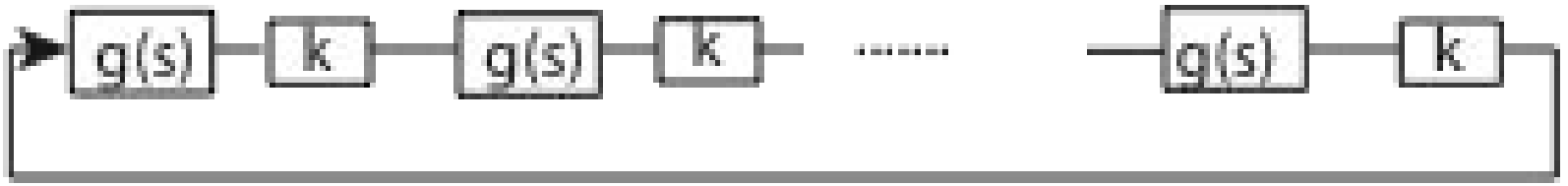


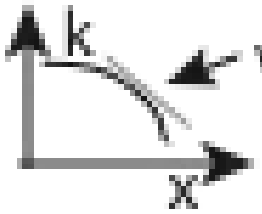
This data is from a colony of cells which were initialized at t=0. As time goes on they become less well synchronized so smoothing occurs.

A mathematical/statistical challenge

Give the data graphed in the previous slide, suggest a parametrized model describing the loss of synchronization and use it to find the maximum likelihood fit the the parameters of the oscillator.

A more general version of Elowitz-Leibler





variable slope

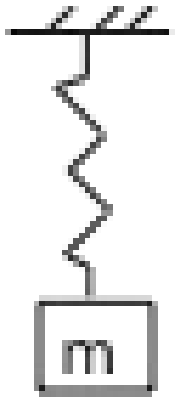
$(g(s))^n + k^n = 0$

$(s+1)^{-3} + k^3 = 0$

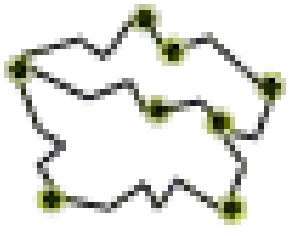
$s^3 + 3s^2 + 3s + 1 + (1/k)^3 = 0$

Imaginary roots at $1/k = 2.82$

Compare with mass-spring physics



$m d^2x/dt^2 + kx = 0$



$$\ddot{x} + f(x) = 0$$

Synchronization with an external stimulus

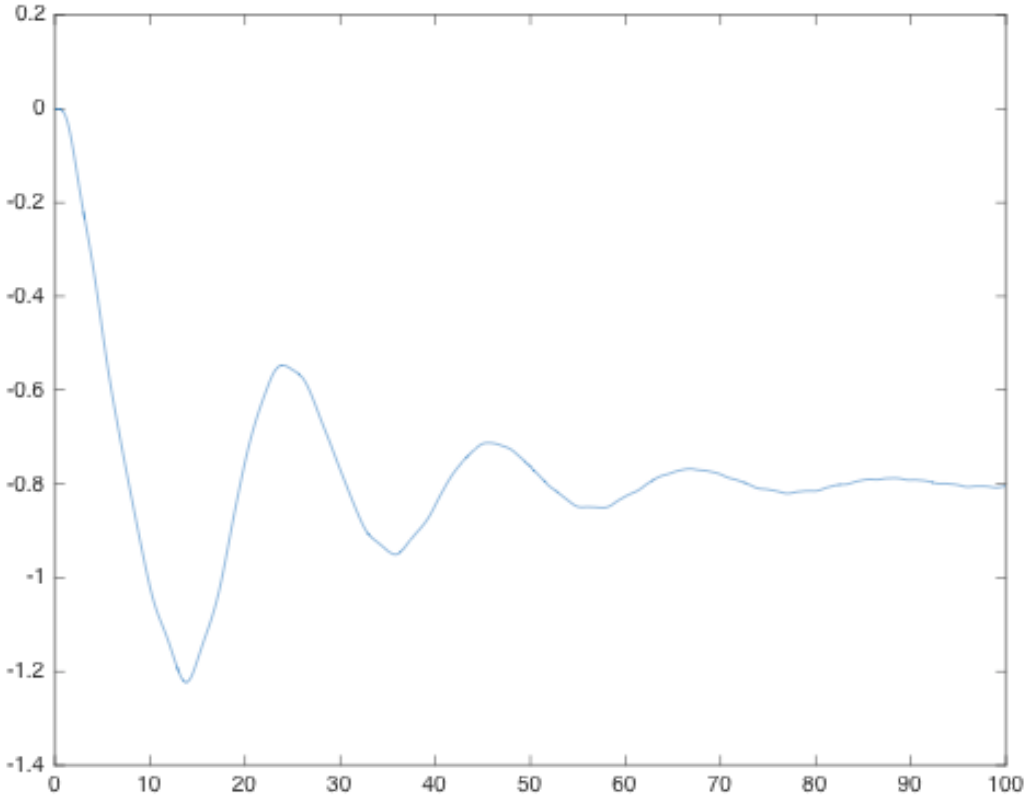
We distinguish between driven oscillations and self sustained ones. The earth's rotation provides a number of signals which oscillate with a period of about 24 hours. How can one model a mechanism by which an oscillator, or family of loosely coupled oscillators, synchronize with an external signal without being able to extract significant energy from the external signal? (In engineering terms, it must have a high input impedance.)

A lot more could be said about the tree-like structure of the clock networks of plants and animals reported in the literature. (suprachiasmatic nucleus in humans, etc.)

Synchronization with an external stimulus

$$\ddot{x} + \dot{x} f(x) + (1.8 + \beta)x = 0$$
$$\dot{\beta} = -m\beta + kx\cos t - \dot{x}\sin t$$

external signal



Showing β as a function of time

Showing how an oscillator having a natural frequency of about 1.35 locks in with an external signal of frequency 1. In one interpretation, beta would be the concentration of a catalyst and the equation for its derivative derived from chemical kinetics.

Data from steady state relationships

Given a description of the general form
$\dot{x} = f(x, u)$
with x a high dimensional vector, and given that it is possible to measure a vector of responses $y = Hx$, one approach to learning quantitative information about f is to let u be a constant and measure the steady state value of y. The equilibrium relationship is, to first order,

$$y = H \left(\frac{\partial F}{\partial x}\right)^{-1} \left(\frac{\partial f}{\partial u}\right) u$$

This is easier than determining all of f

Possible Measurements

Estimates for the entries of the matrix
$G = H\left(\frac{\partial F}{\partial x}\right)^{-1}\left(\frac{\partial f}{\partial u}\right)$
can obtained by carrying out a series of experiments involving just using just one nonzero component of u at a time. Because the dimension of x is typically much higher than the dimensions of u or y this information is not enough to construct a complete model.

How to adapt this to the situation where steady state is an oscillation?

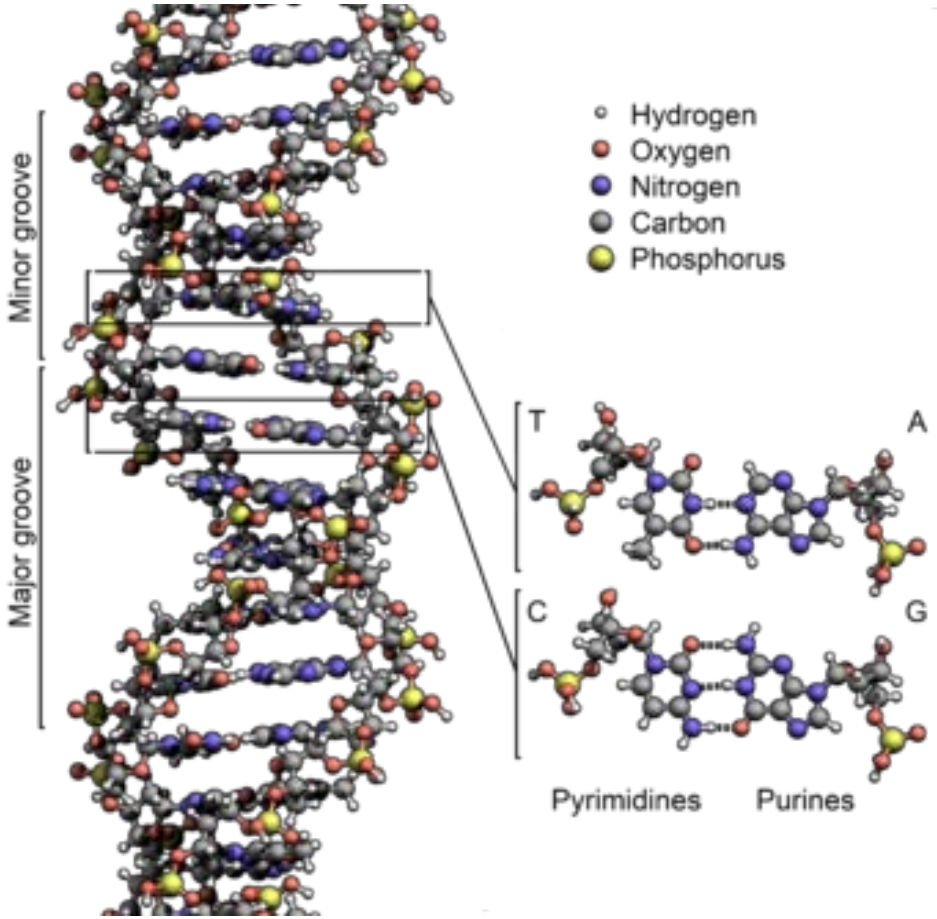
More detailed data based modeling

The steady state model can be refined by incorporating transient data. This involves making measurements at more points in time. In some respects system identification is a well understood subject, however when the system involves thousands of variables and a parametrized model is not available, the general theory is not much help. Moreover, a complete model might not be useful because of its complexity. Practical work focuses on identifying a reduced complexity model.

Data guided controllability for reprogramming

Through the magic of high resolution genomic sequencing it is possible to identify segments of DNA with similar function. The reference below calls these "topologically associated domains".

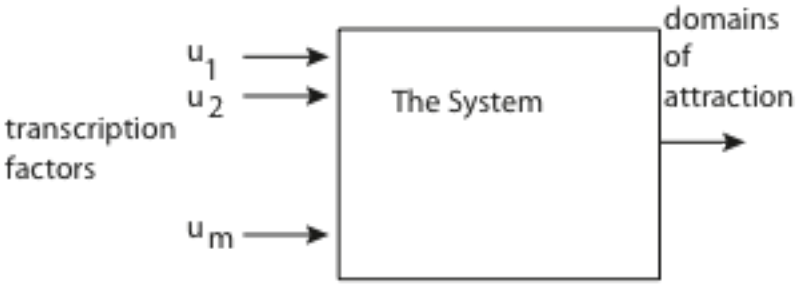
A short segment of DNA

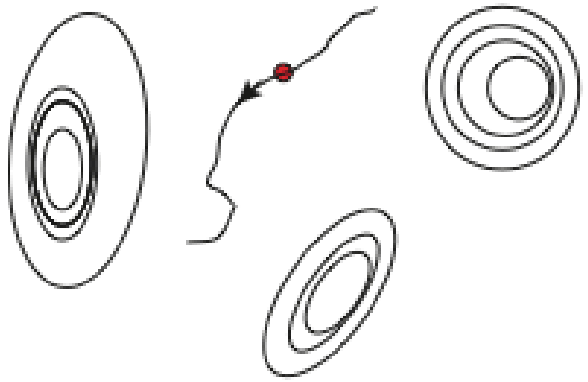


Data guided Controllability: Learning from the Human Genome, Geoff Patterson, et al. (Submitted)

Data guided controllability for reprogramming

A measure of the activity of a subset of the topologically associated domains is taken as an approximate state vector. Recall that the goal of reprogramming is to steer the system from one oscillating domain to another. Thus it makes sense to think of reprogramming as steering from some point on one oscillation to a desired domain of attraction. The controls are to be taken to be subsets of potentially significant transcription factors.





various domains of attraction as suggested by Waddington

Results, glossing over many many details

Briefly, through the use of system identification and simulation it was possible to rate different combinations of transcription factors as to there effectiveness as reprogramming stimuli.

Data guided Controllability: Learning from the Human Genome, Geoff Patterson, et al. (Submitted)