Protein production

What can we do with ODEs

Modeling protein production with ODE

We have seen various examples of “wiring” networks.

Once you know how a regulatory network is wired in a particular way, systems of ODEs can be written and solved, providing a powerful tool for predicting how these regulatory networks will behave inside the cell.

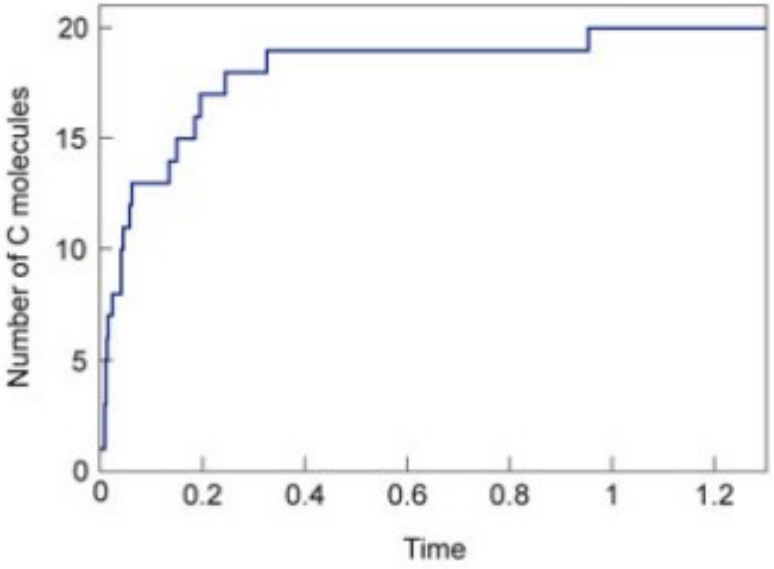
This lecture: two rather simple but very important examples (an unregulated gene and a negatively autoregulated gene).

The same methods are used to analyse much more complicated networks with many tens of genes and proteins.

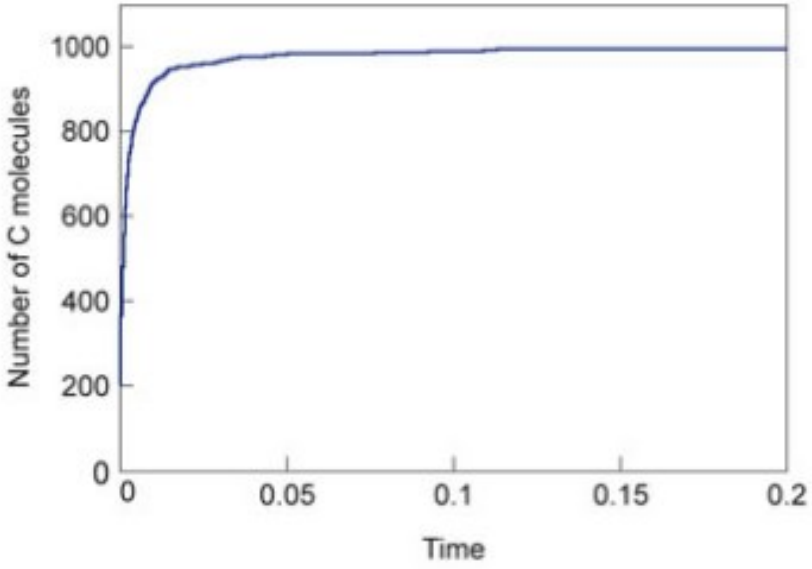
Memo: Ignoring “space”, and compartmentalisation, we are making the important assumption that the interior of the cell (or a particular cellular compartment) is “well mixed” (this is ok on certain timescales, hence for certain processes, but will not always be the case!).

Consider: an *A* molecule collides with a *B* molecule,
the two can react to produce a molecule of type *C*.

$$A + B \xrightarrow{q} C$$



Number of C molecules as a function of time, for $q = 1$, starting with $N_A = 20$, $N_B = 20$.



The same as shown left but starting with $N_A = 1000$, $N_B = 1000$.

Consider: an *A* molecule collides with a *B* molecule,
the two can react to produce a molecule of type *C*.

In a small interval of time *dt*, the probability of a *C* molecule being produced is

qN_AN_B/V ,

where V is the volume of the system, N_A is the number of *A* molecules and N_B is the number of *B* molecules.

Note: the probability scales with $1/V$ since a pair of *A* and *B* molecules will be less likely to meet each other in a larger volume.

$$A + B \xrightarrow{q} C$$

Figures previous slide show simulations of this process. (*q* and *V* have been set = 1)

If the numbers are small, the stochasticity in the process is important (randomness in the A+B collision and reaction times) and discrete events can be seen. Over next lectures, we will learn how to simulate this class of processes via an efficient algorithm.

If the numbers are large, then many events take place in a small time interval. We can assume the process continuous, and write:

$$\begin{aligned} \frac{dc_A}{dt} &= \frac{dc_B}{dt} = -qc_a c_B \\ \frac{dc_C}{dt} &= qc_a c_B. \end{aligned}$$

Note: this is a second order or bimolecular reaction (it involves two reacting molecules), the dimensions of the rate constant are (concentration^{-1})(time^{-1}). We also need to specify initial conditions, e.g. $c_A(0) = c_B(0) = c_0$
and $c_C(0) = 0$.

Can write ordinary differential equation methods to understand how cells control the production of protein molecules from their genes.

Here, we are interested in how the concentration, c_P, of a specific protein molecule, P, changes with time inside the cell. Protein P is produced from its gene, gP, by transcription (to make messenger RNA) followed by translation (to make an amino-acid chain) and protein folding.

We could model all of these processes in detail but for now let's just suppose that protein P is produced at a constant rate, k, as long as the gene, gP, is active. This reaction is zeroth order: the protein P is created at a constant rate that does not depend on any other variables in the model. The dimensions of the rate constant for this reaction are therefore (concentration)(time^{-1}).

We write this as a chemical reaction,

$$\text{source} \xrightarrow{k} \text{P}$$

"source" is actually the gene, gP, plus the whole machinery of transcription and translation. Here we just put this into a 'black box' and assume that protein P is produced at a constant rate.

Protein molecules are also removed from the cell; This could be because another protein molecule actively degrades them or because the cell is growing and dividing into daughter cells (and every time the cell divides, a given protein molecule has a chance of being lost). For now, let's just assume that there is a fixed probability per unit time, μ, that any given molecule of P is removed.
We can also write this as a chemical reaction:

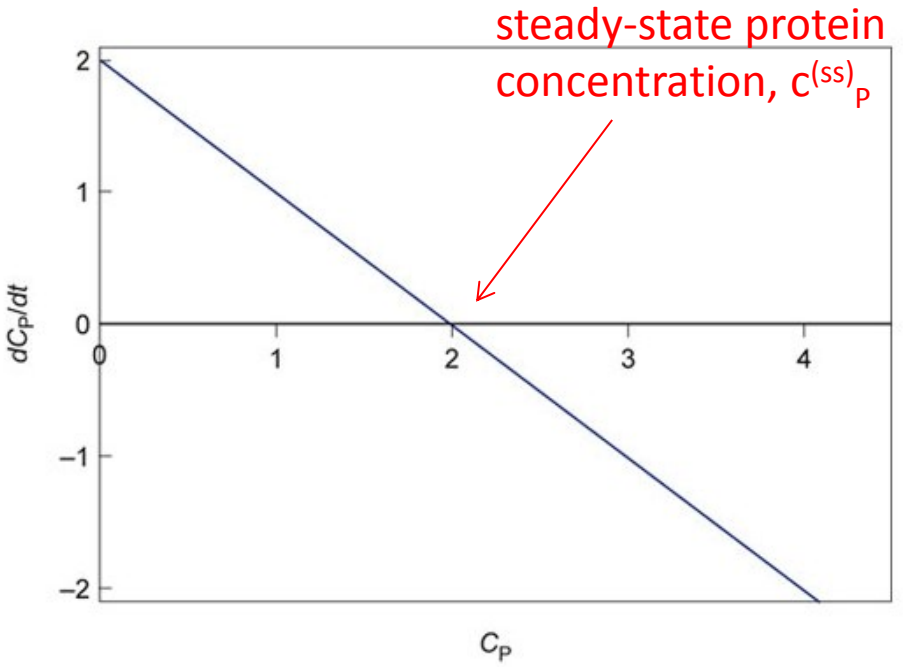
$$\mathrm{P} \xrightarrow{\mu} \mathrm{sink}$$

[a first-order or unimolecular reaction: a single molecule of P reacts. For unimolecular reactions, the dimensions of the rate constant are (time)$^{-1}$. The "sink" here is another black box, could be removal into a daughter cell or degradation into unspecified products]

we can write a differential equation for the rate of change of the concentration c_P of P molecules:

$$\frac{dc_P(t)}{dt} = k - \mu c_P(t)$$

If you know *k* and μ you can plot the rate of concentration change:
(kere *k=2* and μ =1)



Steady-state concentrations are a very important property of regulatory networks, and quite often this is all that people focus on when they study a model for a particular regulatory network.

The value of $c^{(ss)}_P$ depends on both k and μ. If protein removal is due to cell division and if the average time between cell divisions (the cell cycle time) is τ, then

$$\mu = \frac{\ln 2}{\tau}$$

For the bacterium E. coli on a good food source, τ is about 30 min, so μ is about 0.02/min.
Protein production rates, k ,vary greatly, from virtually zero to about 50/min.

So the number of protein molecules in a cell (assuming that there is only one copy of the gene) can vary from zero to several thousand.

For the simple model discussed here, we also solve the model for the time-dependent protein concentration, $c_P(t)$. This is important because genes can be turned on or off in response to signals, and we'd like to know how fast the cell can respond to a given signal.

The time-dependent solution for protein concentration in this model can be found by simple integration:

$$c_{\mathrm{P}}(t) = \frac{k}{\mu}(1 - \mathrm{e}^{-\mu t}) + c_{\mathrm{P}}(0)\mathrm{e}^{-\mu t}$$

Let's suppose that protein P is an enzyme that allows the cell to metabolise lactose. Initially, the gene, gP, is repressed because a repressor protein is bound to its promoter. We assume that initially no protein P is present: $c_P(0) = 0$.
At time zero, the cell detects some lactose and the repressor leaves the promoter, so the gene becomes activated.
How quickly can the cell produce protein P and start metabolising lactose? If $c_P(0) = 0$, then the dynamics is given by:

$$c_{\mathrm{P}}(t) = \frac{k}{\mu}(1 - \mathrm{e}^{-\mu t})$$

We define the rise time, t_{rise}, as the time it takes for protein P to reach half of its steady-state value. Setting $c_P(t)$ to $c^{(ss)}_P/2$ and solving for t_{rise}, we obtain:

$$t_{rise} = -\frac{1}{\mu} \ln \left[1 - \frac{\mu c_{\mathrm{P}}^{(ss)}}{2k} \right]$$

which becomes, when we substitute in $c^{(ss)}_P = k/\mu$:

$$t_{rise} = \ln(2)/\mu$$

i.e. the response time of this simple network is determined only by the protein-removal rate. For bacteria, protein removal is usually due to cell growth and division.

As we saw earlier, the removal rate, μ, is typically $\ln(2)/\tau$, where τ is the cell cycle time. So the response time for bacterial gene networks is typically of the order of the cell cycle time, which is at least 30 min.

ODE for negatively autoregulated gene

Genes can be turned on and off by the binding of specific proteins to the DNA in the promoter region. In many cases, proteins actually turn off their own production (i.e. the protein product of a gene is a repressor that binds to its own gene and turns off protein production). This is an example of negative feedback and is called negative autoregulation.

We saw in lecture 3 that for *E.coli*, and probably for other organisms too, negative autoregulation happens much more often than one would expect if the regulatory "connections" between genes were chosen at random.

Why has negative autoregulation been selected by evolution as a favoured regulatory motif?

To try to understand this, let's write down the equivalent differential equation model for a protein that represses its own production.

We anticipate a result from the next lectures:
For a protein binding to a DNA binding site, the probability that the binding site is occupied is

$$p_{bound} = \frac{\left(\frac{c}{c_0}\right) \exp -\beta\Delta\epsilon}{1 + \left(\frac{c}{c_0}\right) \exp -\beta\Delta\epsilon}$$

where c/c_0 is the concentration of protein (relative to some standard value, c_0) and $\Delta\varepsilon$ is the change in energy when the protein binds. We can define a dissociation constant, K_d, as

$$K_d = c_0 e^{\beta\Delta\epsilon}$$

For low concentrations (where c/c_0 is very small), we can see that the probability p_{bound} that the binding site is bound becomes proportional to the inverse of the dissociation constant: $p_{bound} \rightarrow c/K_d$.
This shows us that K_d is actually just the equilibrium constant for the dissociation of the protein from its binding site. The reason why this proportionality does not hold at higher concentrations is that the binding site becomes saturated with protein.

The more strongly the protein binds to its DNA binding site, the more negative $\Delta\varepsilon$ will be. Strong negative autoregulation (large negative $\Delta\varepsilon$ therefore corresponds to a small value of K_d. Combining the equations above, we get

$$p_{bound} = \frac{\left(\frac{c_P}{K_d}\right)}{1 + \left(\frac{c_P}{K_d}\right)}$$

and the probability that the binding site is unoccupied is given by:

$$p_{unbound} = 1 - p_{bound} = \frac{1}{1 + \left(\frac{c_P}{K_d}\right)}$$

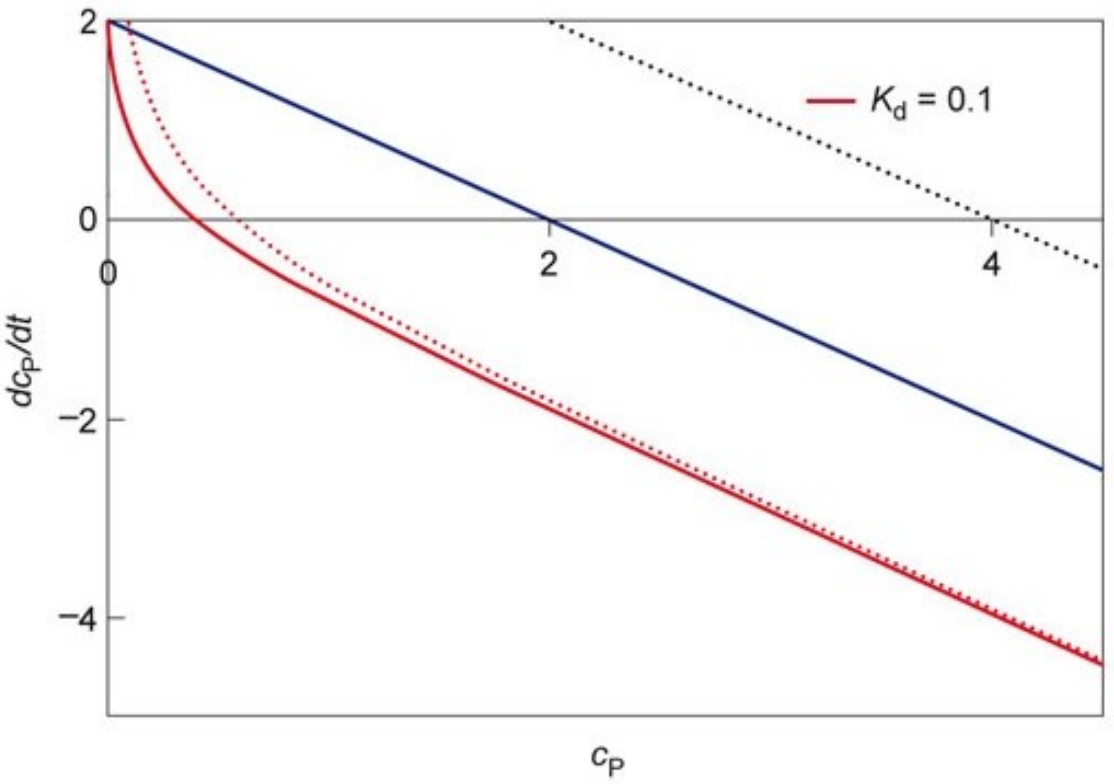
Returning to our differential equation for the production and degradation of protein, the production rate is now proportional to the probability that the promoter binding site is not occupied by protein:

$$\frac{dc_P(t)}{dt} = kp_{unbound} - \mu c_P = \frac{k}{1 + \left(\frac{c_P}{K_d}\right)} - \mu c_P$$

We now have a nonlinear differential equation for the concentration of protein, $c_P(t)$. Let's find out what the steady-state protein concentration is.

Plot of the rate of change of c_P versus c_P, for two values of the production rate k. Also plotted are the results for a gene without negative autoregulation.

[The solid lines are for k = 2 and μ = 1, and the dotted lines are for k = 4 and μ = 1. The blue lines show the result without negative feedback (for the same k and μ).]



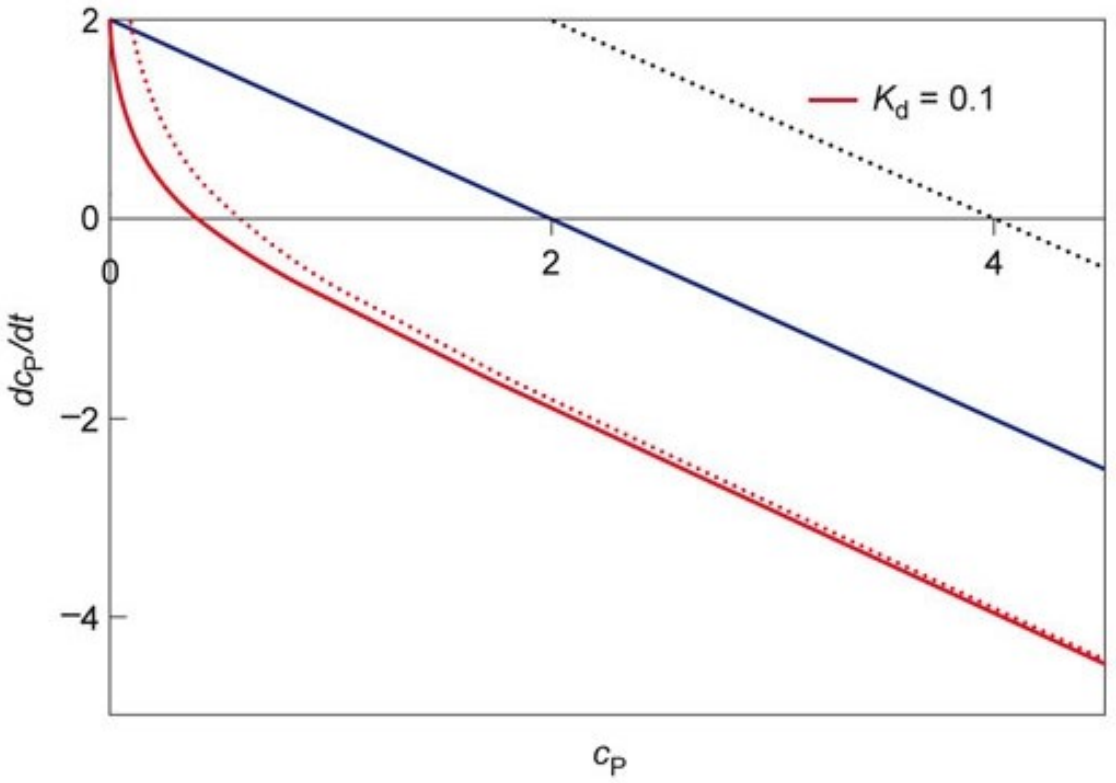
We see that as in the non-regulated case, when the protein concentration c_P is low production dominates, while when the protein concentration is high protein degradation dominates over production.

Again for one particular value of protein concentration production and degradation are balanced ($d c_P/dt = 0$), and this is the steady-state protein concentration.

Negative autoregulation affects the steady-state protein concentration in two important ways.

First, the steady-state protein concentration is lower for the negatively autoregulated gene (shown in red) than for the unregulated gene (shown in blue).

Second, when we compare the results for two different values of the production rate, k (solid and dotted lines), we can see that for the unregulated gene the steady-state protein concentration depends strongly on k (in fact, we know from our calculations above that it is proportional to k); while for the negatively autoregulated gene, $c^{(ss)}_P$ changes only a little when k is changed by a factor of two.



Both of these effects have important implications for the performance of the gene, as we shall see.

More on negative autoregulation.

To get an expression for the steady-state protein concentration $c^{(ss)}_P$ for the negatively autoregulated gene, we set the rate of change of $c_P(t)$ to zero:

$$\frac{dc_{\mathrm{P}}(t)}{dt} = \frac{k}{1 + \left(\frac{c_{\mathrm{P}}}{K_d}\right)} - \mu c_{\mathrm{P}} = 0$$

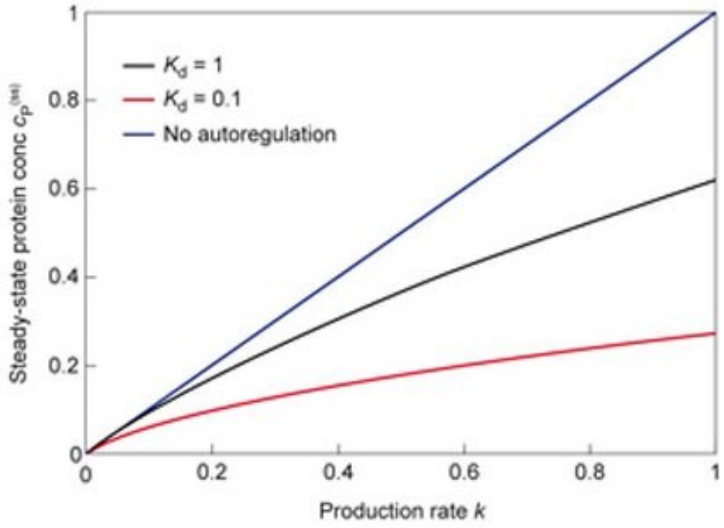
Obtaining:

$$c_{\mathrm{P}}^{(ss)} = \frac{K_d}{2}\left[-1 + \sqrt{1 + \frac{k}{\mu K_d}}\right]$$

For very strong autoregulation (where K_d is very small), the result reduces to:

$$c_{\mathrm{P}}^{(ss)} = \frac{K_d}{2}\left[-1 + 2\sqrt{\frac{k}{\mu K_d}}\right] \simeq \sqrt{\frac{kK_d}{\mu}}.$$

shows $c^{(ss)}_P$ as a function of the protein production rate, k, for several values of the dissociation constant, K_d. As the negative autoregulation gets stronger (as K_d decreases), the curves become flatter: the steady-state protein concentration becomes less dependent on the protein production rate.



More on negative autoregulation.

In the cell, the protein production rate depends on the concentration of RNA polymerase, as well as the concentration of ribosomes, mRNA degradation enzymes, etc. All of these factors vary from cell to cell and over time inside any given cell. We therefore expect the protein production rate to fluctuate within and between cells. For a gene without negative autoregulation, this will cause the protein concentration to fluctuate, since $c^{(ss)}_P$ is proportional to the production rate k. This fluctuation problem can be avoided using negative autoregulation. Because the curve of $c^{(ss)}_P$ versus k is much flatter in the case of negative autoregulation, the steady-state protein concentration will remain stable even if the intracellular environment (i.e. the protein production rate) fluctuates. In other words, negative autoregulation can make the performance of a gene robust to changes in protein production rate.

You may have noticed that for negative autoregulation $c^{(ss)}_P$ does depend on the dissociation constant, K_d. Is this a problem for robustness? Probably not: we expect K_d to fluctuate much less than k because K_d depends only on how strongly the protein binds to its DNA binding site, which is determined by the structure of the protein and the sequence of the binding site.

More on negative autoregulation.

Negative autoregulation also has an important effect on the rise time, t_{rise}: the time the cell needs to turn the gene on (to the half-maximal protein level). We saw that for the unregulated gene this time was fixed by the protein-removal rate, $t_{rise} = \ln(2)/\mu$.

What happens for a negatively autoregulated gene?

To calculate t_{rise}, in principle, we should solve the full equation, but this is tricky analytically. If we look at early times, when c_P is small, we can approximate $c_P(t)/K_d < 1$ then

$$t_{rise} = -\frac{1}{\mu}\ln\left[1 - \frac{\mu c_{\mathrm{P}}^{(ss)}}{2k}\right]$$

and if we also assume that autoregulation is strong we can substitute the previous result for $c^{(ss)}_P$, obtaining

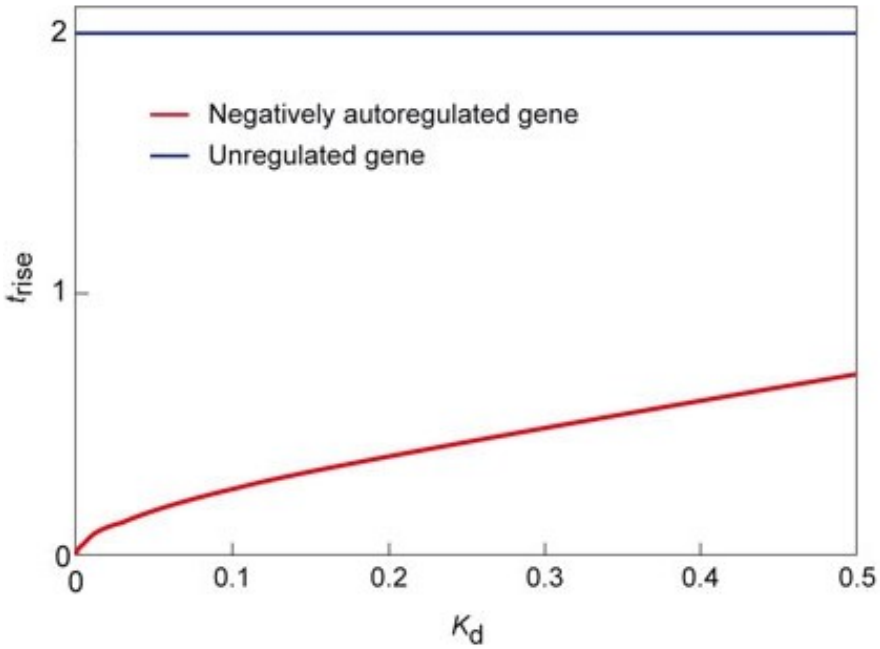
$$t_{rise} = -\frac{1}{\mu}\ln\left[1 - \frac{\mu}{2k}\sqrt{\frac{kK_d}{\mu}}\right] = \frac{1}{\mu}\ln\left[\frac{2}{2 - \sqrt{\frac{kK_d}{\mu}}}\right]$$

More on negative autoregulation.

$$t_{rise} = -\frac{1}{\mu}\ln\left[1 - \frac{\mu}{2k}\sqrt{\frac{kK_d}{\mu}}\right] = \frac{1}{\mu}\ln\left[\frac{2}{2 - \sqrt{\frac{kK_d}{\mu}}}\right]$$

As K_d decreases (i.e. as the negative autoregulation becomes stronger), t_{rise} decreases. This important result shows that negative autoregulation can help cells to respond more rapidly to changes in their environmental conditions than they would be able to without regulation.

The units chosen in the plot here are rather arbitrary.



To get a feeling for some real numbers, we have already seen that a typical protein-removal rate μ in a bacterial cell would be 0.02/min, so the rise time for a typical protein without negative autoregulation would be ln(2)/μ (~ 35 min). While protein production rates and protein-DNA dissociation constants can vary enormously, a realistic value for k might be 0.2 molecules/min per cell volume and K_d might be 0.02 molecules per cell volume (for a protein that binds very strongly to its DNA binding site). The value of t_{rise} for a negatively autoregulated gene, assuming these parameter values, would then be 12.7 min: almost a factor of three faster than the gene without negative autoregulation.

How are these measurements done, at population level?

Aside from noise and fluctuations, which we address later, how is the type of mRNA present in a sample (a population) of cells measured?
DNA microarray chips can be used. These are large arrays (tens of thousands) of pixels (dots). Each pixel represents part of a gene, by having of the order of $10^6 - 10^9$ single-stranded DNA-mers, that are identical copies from the DNA of the gene.

The chip size is of the order of 1 cm^2. The analysis consists of taking a cell sample, extracting all mRNA in this (hopefully) homogeneous sample, and translating it to cDNA (DNA that is complementary to the RNA, and thus identical to one of the strands on the original DNA). The cDNA is labeled with a fluorescent marker. The solution of many cDNAs is now flushed over the DNA chip, and the cDNAs that are complementary to the attached single-stranded DNA-mers will bind to them. The DNA chip is washed and images (with pixel resolution) and the fluorescent light intensity thus measures the effective mRNA concentration.

Biochemical Noise

Small number stochasticity

Small numbers

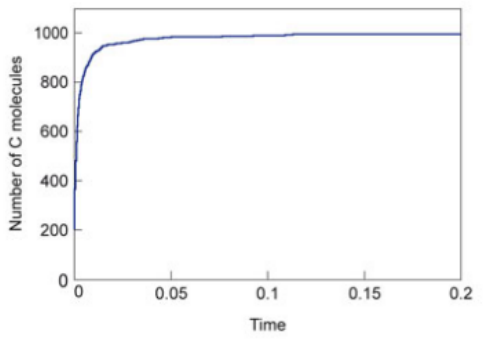
Cells with identical genes and environmental factors can differ chemically.

One way in which this can come about is because of reactions involving small numbers of molecules.

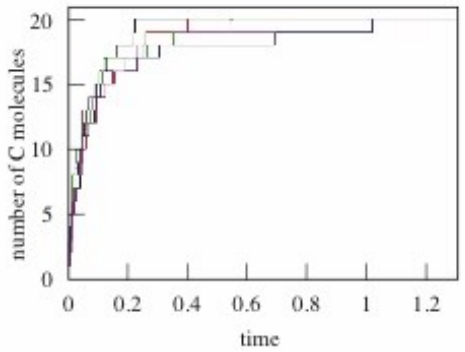
This can be addressed very quantitatively.

$$A + B \xrightarrow{q} C$$

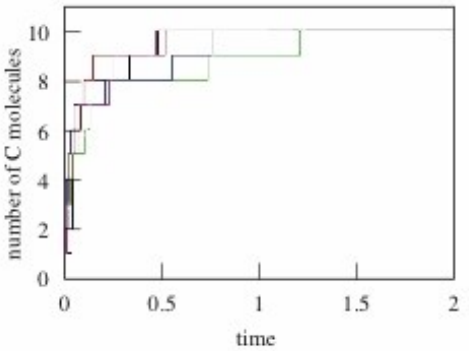
Rate constant q is set to 1. Each simulation run is repeated five times (results shown in different colours) .



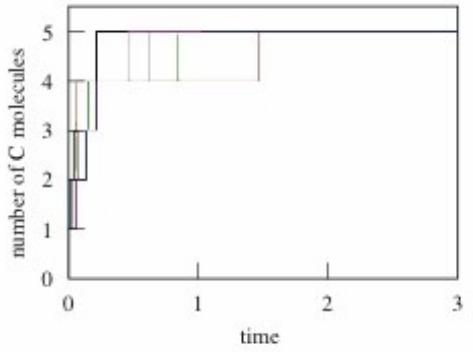
starting with $N_A = 1000, N_B = 1000$



starting with $N_A = 20, N_B = 20$



starting with $N_A = 10, N_B = 10$



starting with $N_A = 5, N_B = 5$

Why is a chemical reaction “noisy” when the number of molecules is small?

The reason is that chemical reactions are stochastic, or random. That is, the outcome is governed by probabilities, and there are sufficiently few molecules that there is no single overwhelmingly favoured outcome.

In our box of A and B molecules (the cell), we do not know the exact positions and velocities of all of the molecules and so we do not know the exact time when a pair of A and B molecules will meet and react. The exact times when reactions happen and the exact sequence of reactions that happen can be different in repeat runs of the same experiment.

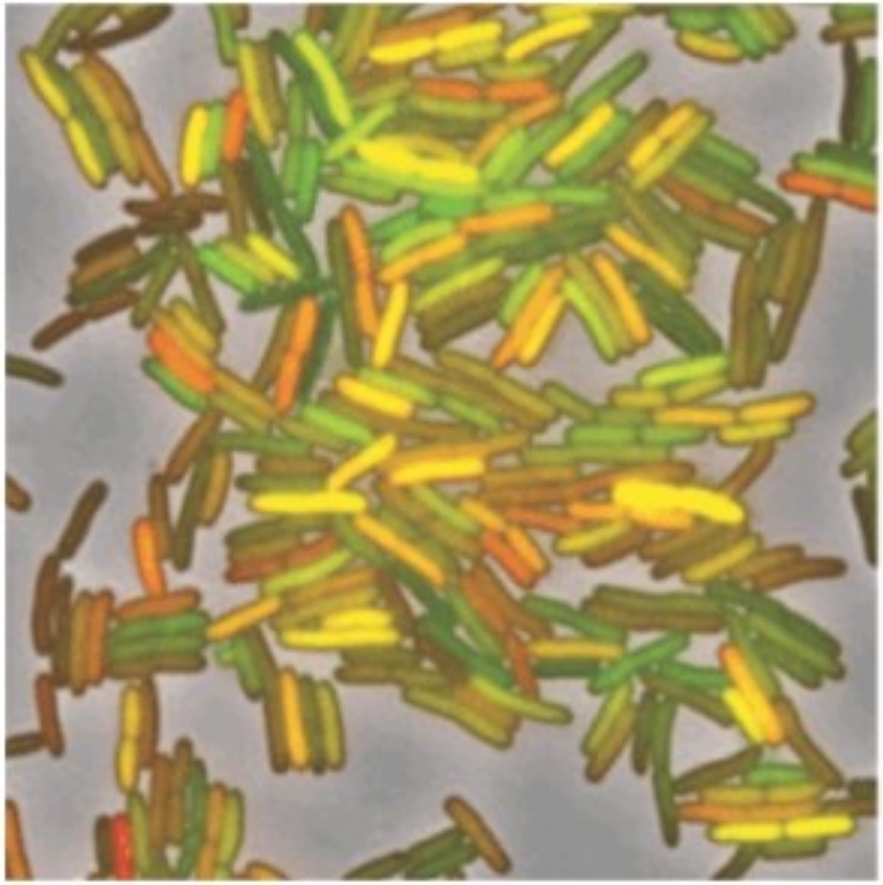
Why is this relevant? Even in something as small as a bacterial cell, there are many billions of molecules, so why would these stochastic effects be important?
In fact, stochastic effects can be very important in cells, because even though the total number of molecules in a cell is large, the number of molecules involved in a particular biochemical reaction network can be very small.
For example, in slow-growing cells, there is only one copy of the DNA (so the number of molecules of a particular gene may actually be only one). The number of messenger RNA molecules in the cell corresponding to a particular gene can also be very small for weakly expressed genes, and some proteins are only present in small numbers. Biochemical reaction networks involving genes, mRNA or proteins that are present in small numbers per cell are likely to be dramatically affected by small-molecule number fluctuations.

We call these stochastic fluctuations “biochemical noise”.

Individual cells are not identical

That biochemical noise really is significant for biological cells was illustrated in an important experiment by Michael Elowitz et al. in 2002.

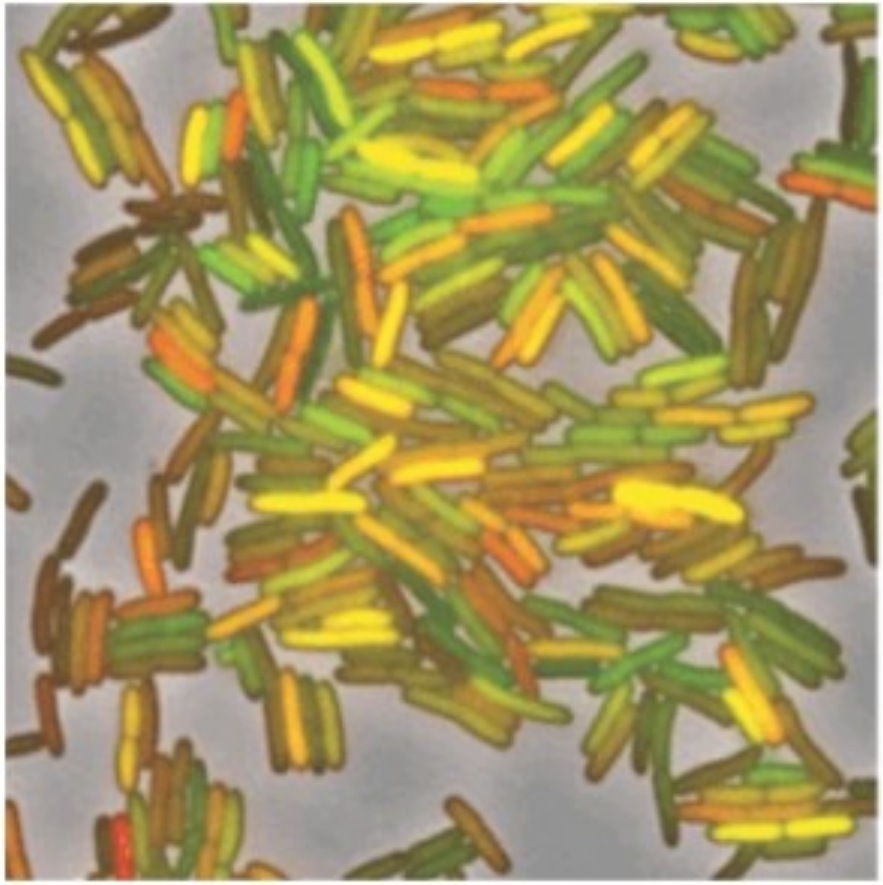
They engineered *Escherichia coli* bacteria to carry two different coloured fluorescent reporter genes. These genes encode proteins that do not interfere with any cellular functions but when excited by light of the right wavelength they fluoresce (i.e. they emit light of a longer wavelength).
This can be detected in an epifluorescence microscope.



Elowitz et al. thus measured the relative amounts of the two fluorescent proteins in individual bacterial cells. The question that they wanted to answer was: if two cells are genetically identical and experience the same environmental conditions, will they produce the same amount of the two fluorescent proteins?

Individual cells are not identical

Image shows the results of one of their experiments. This is an overlay of micrographs of a group of *E. coli* cells growing on a semi-solid gel under the microscope. These cells all grew from a single "ancestor" at the start of the experiment so they **are genetically identical. Presumably also all subject to very similar stimuli.**



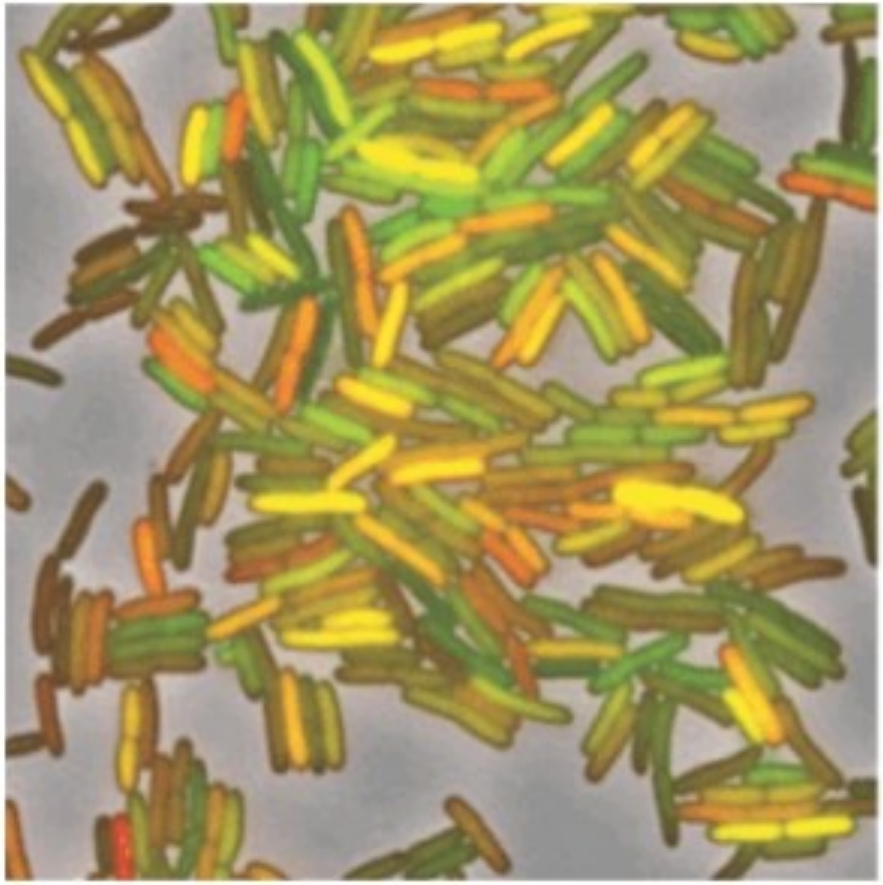
The colours show the relative amounts of the two fluorescent proteins present in each cell: green represents protein 1 and red represents protein 2. Cells that are coloured yellow contain approximately equal amounts of proteins 1 and 2.

It is clear from this image that these "identical" cells are different colours, showing that they are very far from identical in their levels of production of the fluorescent proteins. Elowitz et al. also showed that cells that produce the reporter proteins at low levels (small number of molecules) have much more "noisy" levels of expression than cells that produce the proteins at high levels (a large number of molecules). This is what we would expect if differences between cells are caused by small molecule number noise since $s = 1/\sqrt{N}$ is larger for small N.

Intrinsic and Extrinsic Noise

Are the differences between cells really caused by small molecule number noise in the chemical reactions involved in protein production (transcription and translation)?
Or are the different colours caused by differences between the cells? For example, we can see in the microscopy image that some cells are short because they have just been generated, while others are much longer and are about to divide.

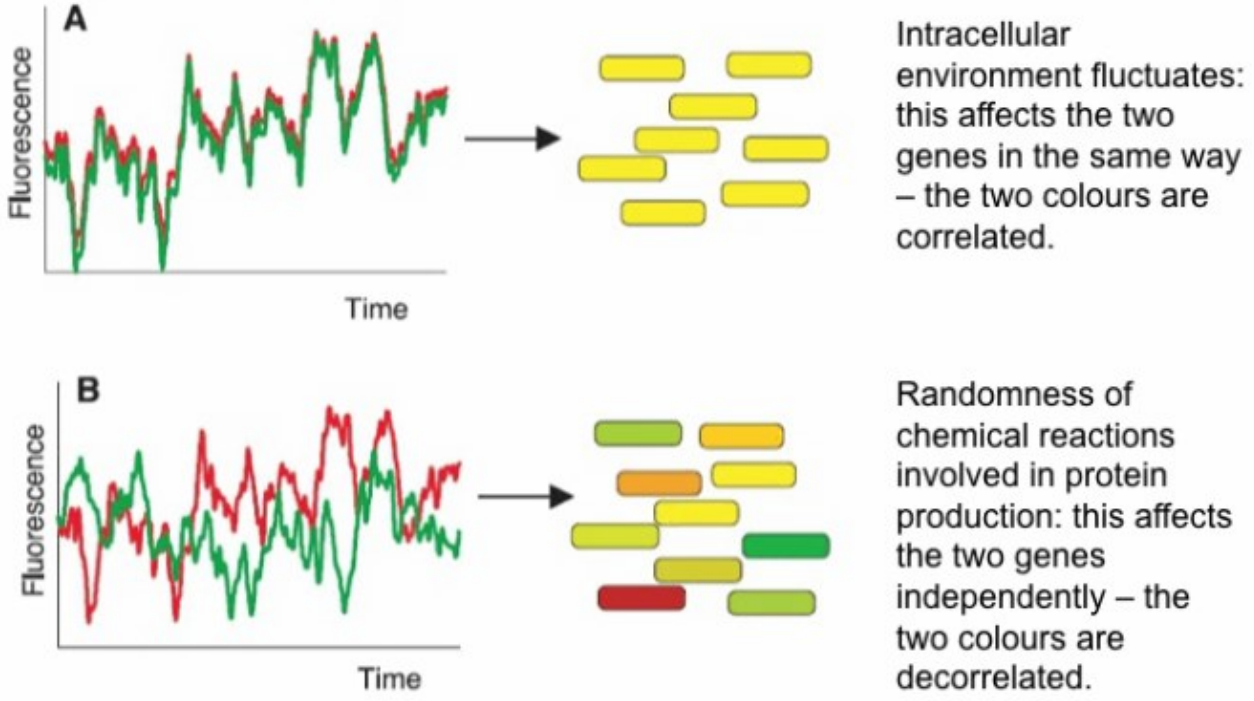
Perhaps this affects the level of protein expression? Cells could also contain different concentrations of RNA polymerase or ribosomes, which would cause them to produce more or less fluorescent protein.



Intrinsic and Extrinsic Noise

Precisely to explore the origins of the different amounts of the proteins, Elowitz et al. used two fluorescent proteins (in different colours) instead of just one. Within each cell, the genes encoding the two proteins should experience the same cell volume, RNA polymerase, ribosome concentration, etc.

So if the differences in protein expression are caused by differences between cells, the levels of the two colours should be correlated cells with a lot of protein 1 should also have a lot of protein 2. However, if chemical reaction stochasticity is responsible for the differences in protein expression, we would not expect the levels of protein 1 and protein 2 in individual cells to be correlated.



Intrinsic and Extrinsic Noise

In fact, by measuring the amount of correlation between the levels of proteins 1 and 2 in individual cells in their experiments, Elowitz et al. could measure how much of the cell-to-cell variation is caused by differences between cells (which they called extrinsic noise) and how much is caused by chemical reaction stochasticity (which they called intrinsic noise).

In their experiments, both sources of noise played a significant role.

Why does it matter that genetically identical cells can have different levels of protein expression? One reason is that biochemical noise limits how precisely cells can control their own behaviour. If a cell needs to control precisely the concentration of a particular protein, either it must produce a large number of molecules (which is expensive) or it must use a biochemical control circuit (such as a negative feedback loop) to reduce the noise.

On a more positive note, biochemical noise may actually be useful for cells in some cases. For example, bacterial populations are often exposed to environmental stress (attack by antibiotics, changes in food availability, etc). If all of the cells in the population are identical in their protein composition, the stress may wipe them all out; but if there is large variability in protein composition among cells, it is possible that a few cells will happen to have the right protein levels to survive the stress. The population can then regrow from these cells once the stress is over.

Elements of modeling noise

Noise - “Birth-death” model for gene expression

For stochastic chemical reactions, we cannot predict exactly which reaction will happen when, or which cell in a population will contain which exact numbers of molecules of proteins, mRNA, etc.
However, we can make predictions about probability distributions.

For example, we might predict the probability that a randomly selected cell in a population will have 100 molecules of a particular protein, even though we cannot predict which cell this will be.
The quantity we are interested in is therefore p(N, t): the probability that our system contains N molecules of protein P at time t.

We can write down an equation for p(N, t) for the simple “one-step model” of gene expression that we discussed above, in which we include chemical reactions for protein production and degradation:

$$\text{source} \xrightarrow{k} \text{P}$$
$$\text{P} \xrightarrow{\mu} \text{sink}$$

We assume that these reactions are “Poisson processes”. This means that if we observe the system for a very short time interval from time t to time t + dt, the probabilities that the first reaction (production), or the second (degradation) happen will be:

$$Prob(produce) = kdt \qquad Prob(degrade) = \mu N dt$$

Noise - “Birth-death” model for gene expression

Considering all the (4) ways in which the number of proteins can change, we have:

$$\frac{dp(N,t)}{dt} = kp(N-1) + \mu(N+1)p(N+1) - kp(N) - \mu N p(N)$$

You can check that the steady state solution is:

$$p(N) = \frac{1}{N!}\left(\frac{k}{\mu}\right)^N e^{-\frac{k}{\mu}}$$

… a Poisson distribution!

As (k/μ) increases, the average number of molecules increases. The mean and standard deviation σ_N of the distribution p(N) are given by:

$$\langle N \rangle = \frac{k}{\mu}$$

$$\sigma_N = \sqrt{\langle N^2 \rangle - \langle N \rangle^2} = \sqrt{\frac{k}{\mu}}$$

Can estimate the importance of stochastic effects considering the ratio of the standard deviation to the mean:

$$\frac{\sigma_N}{\langle N \rangle} = \sqrt{\frac{\mu}{k}} = \frac{1}{\sqrt{\langle N \rangle}}$$

Noise - A two-step model for protein production

The model that we have just been considering may be too simple.
In reality, the production of protein from a gene does not happen in a single step. We can make our model slightly more realistic by making a two-step model that includes both transcription and translation.

The reaction scheme for this model would be

$$\begin{aligned} &\text{source} \longrightarrow \mathrm{M} \\ &\mathrm{M} \longrightarrow \text{sink} \\ &\mathrm{M} \longrightarrow \mathrm{M} + \mathrm{P} \\ &\mathrm{P} \xrightarrow{\mu} \text{sink} \end{aligned}$$

Here, M represents mRNA and P represents protein.

It is possible to write down a chemical master equation also for this model, and to solve it for the steady state probability distribution. In this case, there is a probability distribution for the number of messenger RNA molecules as well as for the number of protein molecules.

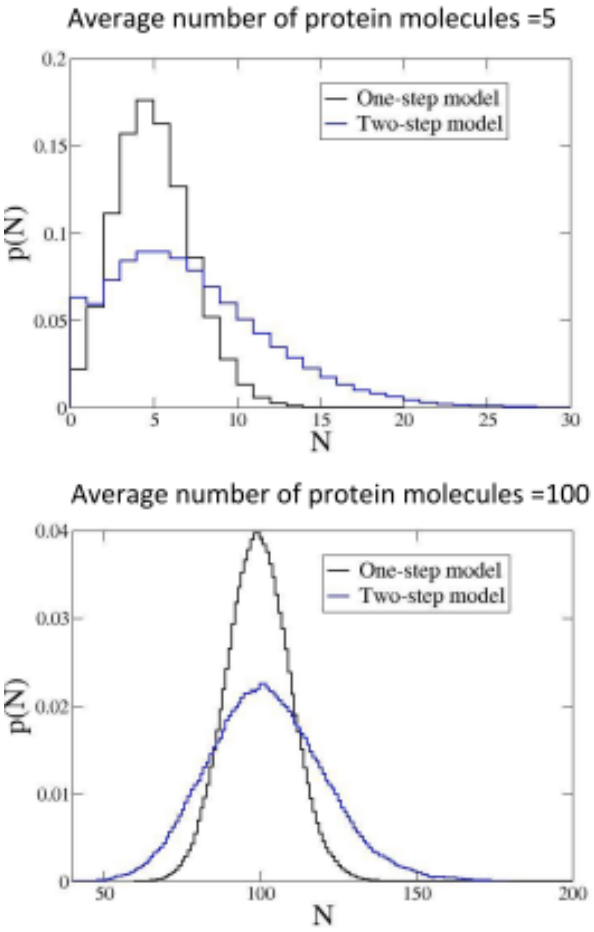
For mRNA we only need to consider the top two reactions (since the bottom two reactions do not change the number of mRNA molecules), which are identical to our previous simpler model. So we expect the probability distribution for the number of mRNA molecules to be a Poisson distribution. However the bottom two reactions, which control the production and degradation of protein, are now different from our simple model. This means that the probability distribution of protein may be different from a Poisson distribution in this model.

Noise - A two-step model for protein production

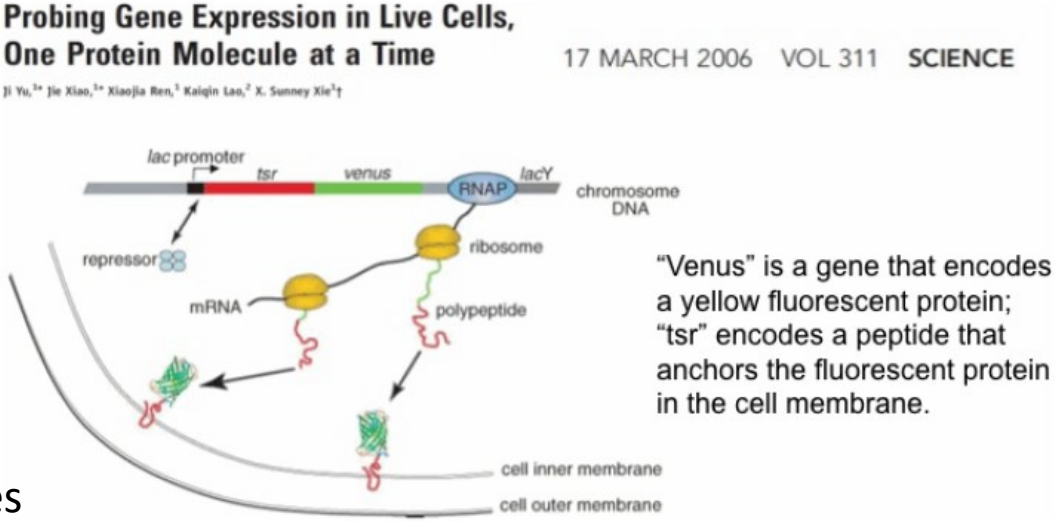
Let's set the parameters (translation rate/mRNA decay rate) so that five proteins are made on average per mRNA molecule (although some mRNA molecules will produce more and some less).

We can compare this with the previous one-step model by fixing the transcription rate so that the average protein number is the same in both models. The results are shown in the plot:

We can see immediately that the distribution is broader in the two-step model, i.e. this model predicts more noisy protein expression than the one-step model. The reason for this is that the extra chemical reaction step amplifies the noise: the number of mRNA molecules is itself noisy, and then on top of this each mRNA molecule can produce a variable number of proteins.



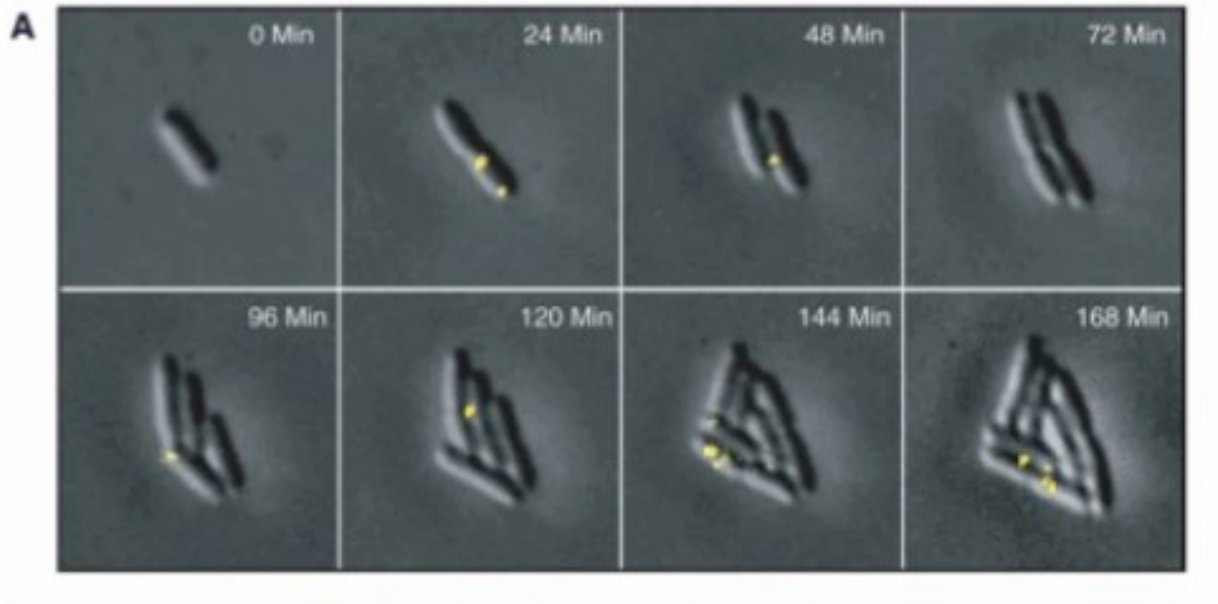
Visualising Noise in expression



How can we test whether these are good models for noisy gene expression in real cells?
One way to do this is actually to carry out single molecule experiments, in other words to watch, under the microscope, the production of single protein molecules in individual cells.

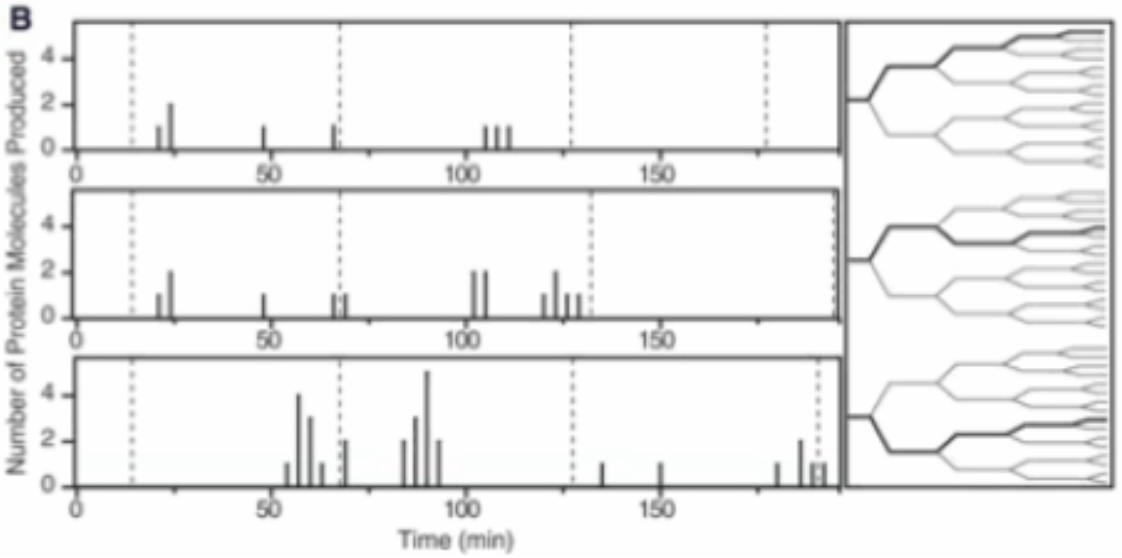
Since protein molecules are very small, this is a very challenging task. However, in 2006, Yu et al. managed to design an appropriate experiment. They made a strain of *E. coli* that produced a yellow fluorescent protein attached to a polypeptide (a chain of amino acid molecules), which could anchor this complex in the cell's lipid membrane.
When the fluorescent protein is anchored in the membrane, it diffuses around much less, making it easier to see single molecules under the microscope. In this system, using advanced fluorescent microscopy, it is possible to see individual fluorescent protein molecules as dots within the cell membrane. Yu et al. could then grow cells under the microscope and track the moments when individual dots appeared in the membrane. In this way, they could see the production of individual protein molecules in real time.
To keep the protein numbers low, the researchers included a binding site for the Lac repressor protein. When this repressor protein is bound to the operator site in front of the gene that encodes the fluorescent protein, no protein will be produced.

Visualising Noise in expression



The bacterial cells in the series of images grow from a single cell during the experiment. The yellow dots show individual protein molecules bound to the cell membrane. By tracking the appearance of these dots, Yu et al. were able to monitor the moments when protein molecules appeared in the membrane.

Visualising Noise in expression



The three plots show the rate of protein production (number of new proteins per 3 minute time interval) for three different cell "lineages", which are illustrated on the right.

This was done for different cell lineages, as shown in the plot, which indicates the number of protein molecules that were produced in a 3 min interval. The dotted vertical lines show the moments when the cell divided into two daughter cells.

What's really striking about Yu et al.'s results is that for most of the time, no protein molecules are being produced. **Protein production occurs in short bursts**, with long intervals where nothing happens. This is probably because most of the time the Lac repressor protein is bound to the DNA, thereby preventing protein expression. The bursts of expression take place during the rare moments when a stochastic fluctuation causes the repressor to fall off its DNA binding site. Yu et al.'s setup therefore allows us to see stochastic chemical reactions happening inside biological cells, in real time and at single-molecule resolution.

We have focused here on noise in gene expression, but the stochasticity of chemical reactions is also important in many other cell functions.

Single-molecule experiments have revealed the effects of biochemical noise in the molecular machines that drive the flagellar motor that allows cells to swim and in the bacterial membrane receptors that sense environmental gradients.

Other experiments have found important effects of biochemical noise in the development of fruit-fly embryos and the mechanisms that control whether or not cells proliferate.

It seems that noise is everywhere.