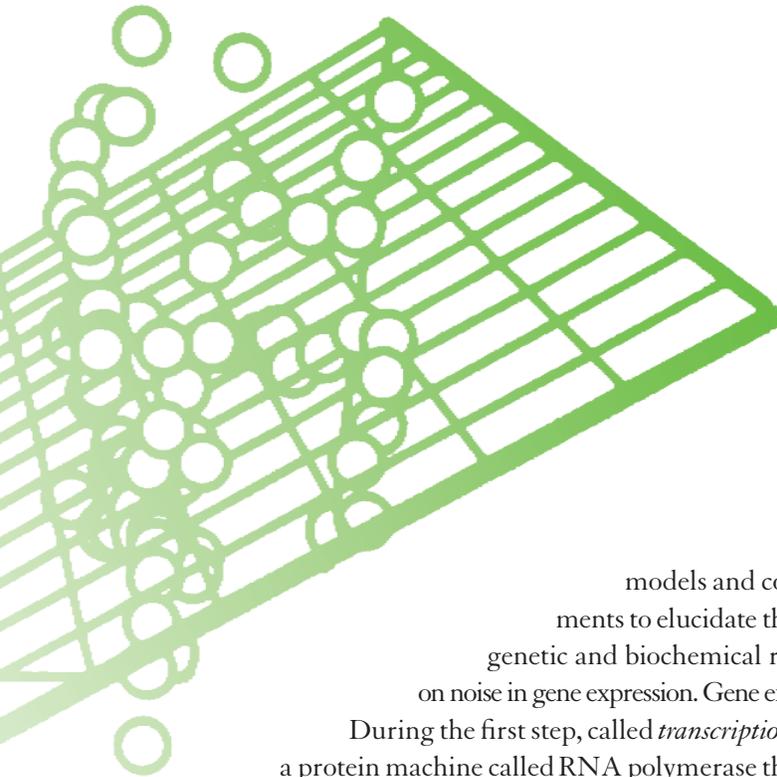


Alexander van Oudenaarden

Noisy Genes

Noise is often perceived as being undesirable and unpredictable. The experimental physicist spends a lot of time trying to lower the noise floor of the experimental set-up to a level that enables the detection of tiny signals. Although noise is therefore often a source of frustration, noise can be controlled and damped by using, for example, clever electronics, mechanical dampers, or shielded rooms. In biology, however, noise is intrinsic to living systems and cannot be controlled by the experimentalist. Living systems are inherently noisy, and are optimized to function in the presence of fluctuations. In this context, evolution plays the role of the experimentalist in trying to control the noise. During evolution biological cells have been fine tuned and optimized to function in noisy environments, but it's not clear what the biological function of noise is. Does noise increase or decrease the fitness of a cell? Some organisms can exploit fluctuations to introduce diversity into a population, as occurs with certain viruses. In contrast, stability against fluctuations is essential for controlling cell differentiation, as in a developing embryo.



IN OUR GROUP we both develop theoretical models and conduct quantitative experiments to elucidate the functional role of noise in genetic and biochemical reactions. Here, I will focus on noise in gene expression. Gene expression is a two step process.

During the first step, called *transcription*, the DNA code is read by a protein machine called RNA polymerase that synthesizes a new molecule: messenger RNA (mRNA). In the second step, called *translation*, the mRNA code is translated into a protein by the ribosomes. In our experiments we work with bacterial cells such as *Escherichia coli* (*E. coli*). These small cylindrical cells have a length of about 1 μm and a diameter of about 0.5 μm . The volume of these cells is only 1 femtoliter (10^{-15} L). *Figure 1a* shows a phase contrast microscopy image of several *E. coli* cells. All the cells in this image originated from a single cell and are therefore genetically identical. However, although the cells are genetically identical they show a large variability in fluorescence (*Fig. 1b*). The fluorescence intensity is proportional to the concentration of Green Fluorescent Protein (GFP) in the cell. This protein is intrinsically fluorescent and is therefore an ideal reporter for measuring protein concentration in a single living cell. GFP originates from a jelly fish and was isolated by the biologists. GFP is now widely used in a variety of other organisms, such as bacteria, yeast, flies, worms, and mice (the favorite model organisms in biology). What is the source of the large variability in GFP concentration between genetically identical twin cells? The small bacterial cells are tiny noisy biochemical reactors because only a small number of molecules are available for transcription and translation. For example, the cell contains only one DNA molecule (chromosome), and contains about 30 RNA polymerase

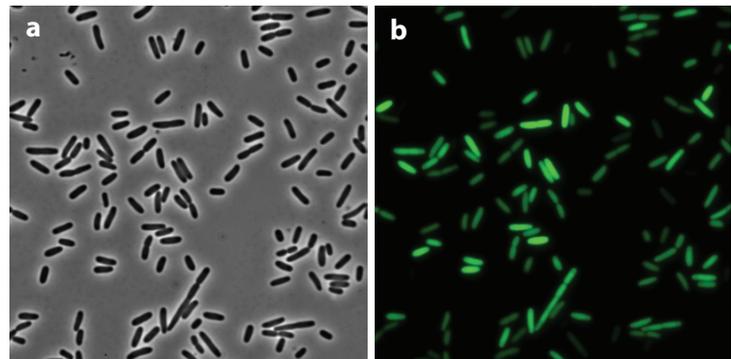
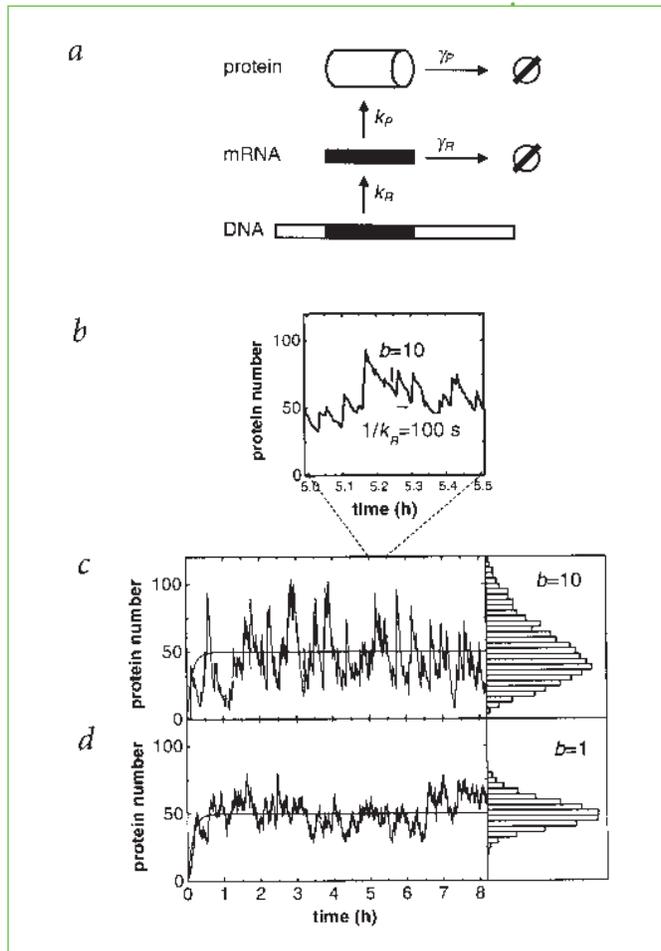


FIGURE 1

a. Phase contrast microscopy image of genetically identical *Escherichia coli* cells. The scale bar is 2 μm . *b.* Fluorescence microscopy image of same cells as in *a.* The fluorescence intensity reflects the GFP concentration in the cells.

The burst size effect. (FIGURE 2)

a. mRNA molecules are transcribed at rate k_R from the template DNA strand. Proteins are translated at a rate k_P off each mRNA molecule. Proteins and mRNA degrade at rates γ_P and γ_R respectively. Degradation into constituents is denoted by \emptyset . **b.** Typically, mRNA is unstable when compared to the protein product of a gene. However, during its brief lifetime an mRNA molecule can inject a large burst of proteins into the cytoplasm. A Monte Carlo timecourse over a 30 min. time interval shows bursts of protein creation of average size $b = k_P/\gamma_R$ occurring at average rate k_R . **c., d.** The following examples both achieve the same mean protein concentration, but with different noise characteristics. **(c)** A gene with low transcription but high translation rates produces bursts that are large, variable, and infrequent, resulting in strong fluctuations. **(d)** Conversely, a gene with high transcription and low translation rates produces bursts that are small and frequent, causing only weak fluctuations in protein concentration and therefore producing a smaller variation in the population.



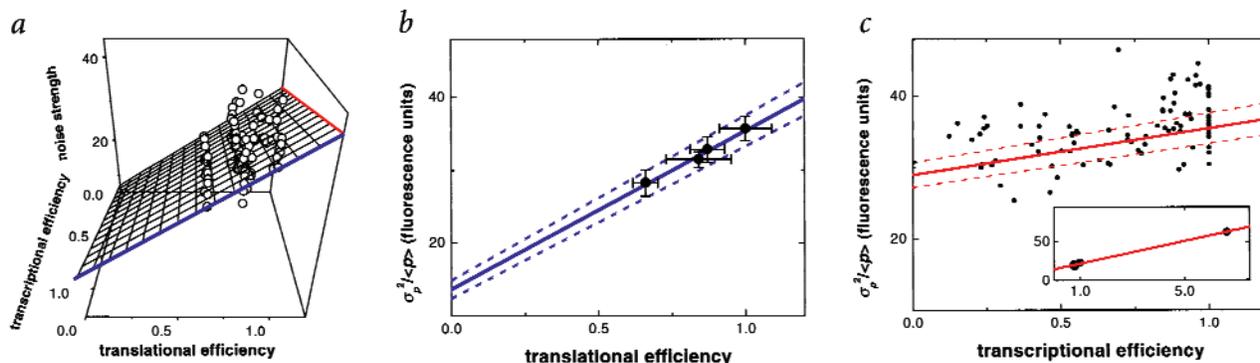
molecules and about 100 ribosomes. A significant fraction of the noise in gene expression comes from the statistical fluctuations due to the small number of molecules. Low numbers lead to relatively large fluctuations. Stochastic fluctuations in gene expression lead to the fascinating phenomenon of non-genetic individuality: even in the case that two individuals are genetically identical, protein concentrations between the two individuals can vary significantly because of the stochastic nature of protein synthesis.

Figure 2 shows a Monte Carlo simulation for the gene expression of a single gene. The simulations include four important reactions: transcription, translation, decay of the mRNA molecule and decay of the protein molecule. In most biological systems the proteins are much more stable than the mRNA molecules ($\gamma_R \gg \gamma_P$). This biological problem shows close similarity with several problems in statistical physics. Using statistical physics the noise in a single gene can be deduced analytically. In our work we define the noise strength as the ratio of the variance σ_p^2

and the mean $\langle p \rangle$. Note that the variance σ_p^2 is the square of the standard deviation σ_p . Using the master equation the noise strength is determined analytically:

$$\frac{\sigma_p^2}{\langle p \rangle} \approx b + 1$$

where $b = k_P/\gamma_R$. The burst size b is equivalent to the number of proteins that is synthesized from a single mRNA transcript. If b is large, a single mRNA molecule is recycled several times before the mRNA molecule degrades. If $b = 1$, each mRNA molecule is only translated once. The analytical expression predicts that the noise strength increases as the mRNA molecules are recycled further. This effect is shown in the Monte Carlo simulations in Fig. 2. It is clear that cells with a burst size of 10 have a much larger noise (reflected in the broad histogram) than cells with a burst size of 1. This difference can be understood by analyzing the dynamics in more detail (inset b). The two simulations in Fig. 2 have the same average number of protein $\langle p \rangle$. This analysis predicts that the noise in a cell is strongly dependent on b , and therefore on the translation rate k_P , but is independent of the transcription rate k_R . This asymmetry between transcription and translation can be experimentally tested.



We tested this prediction by quantifying the cell to cell variation in GFP expression as a function of the transcription and translation rates. The transcription and translation rates were changed by introducing mutations in the chromosome of the bacterium. We introduced a single copy of the GFP gene into the chromosome of the bacterium. The mutations were introduced in such a way that they specifically affect the transcription and translation rates of GFP and not of any other gene in the chromosome.

The experimental results are plotted in *Fig. 3*. We find that the noise, defined as the ratio between variance and mean, increases significantly when the translation efficiency (translation rate) increases. However we find a much weaker dependence as a function of transcription rate. This observation is consistent with the above prediction.

We are living in a noisy environment. We have shown that in addition to external noise sources, cells are intrinsic noise generators. Noise is often harmful, as it garbles cell signals, corrupts biological clocks and disrupts the fine-tuned process of development. Cell signaling pathways and developmental switches have evolved so as to minimize the disruptive effect of such fluctuations, in ways that are only now beginning to be understood.

FURTHER READING

- M. Thattai & A. van Oudenaarden, *PNAS* **98**, 8614 (2001).
 M. Thattai & A. van Oudenaarden, *Biophysical Journal* **82**, 2943 (2002).
 E. Ozbudak, M. Thattai, I. Kurtser, A. D. Grossman & A. van Oudenaarden, *Nature Genetics* **31**, 69 (2002).

ALEXANDER VAN OUDENAARDEN is the *W. M. Keck Foundation Career Development Professor in Biomedical Engineering and Assistant Professor of Physics at MIT*. His research interests focus on genetic noise, as described in this article, and genetic networks. Another research focus is cell motility powered by actin polymerization. The philosophy of the van Oudenaarden lab (web.mit.edu/biophysics) is to combine the physics mentality of quantitative and predictive modeling with the powerful experimental tools of modern biology to elucidate important open problems in current genetics and cell biology. Professor van Oudenaarden is an Alfred Sloan Research fellow; was awarded the National Science Foundation's CAREER award; and is the first physicist to be appointed a *W. M. Keck Foundation Career Development Professor in Biomedical Engineering*.

FIGURE 3

a. Each data point is the summarized result of an entire histogram corresponding to a flow cytometer run of a population of typically 10^4 – 10^5 cells. The noise strength of the population (z , in arbitrary fluorescence units) is plotted as a function of transcriptional efficiency (x) and translational efficiency (y). These data are fitted to a plane of the form $z = a_0 + a_x x + a_y y$ using a least square routine, giving $a_0 = 7.1 \pm 0.9$, $a_x = 6.5 \pm 0.4$, $a_y = 21.8 \pm 0.9$. The ratio $a_y/a_x = 3.4$ gives the relative effect of translational versus transcriptional efficiency on noise strength. *b., c.* For clarity, the three-dimensional data are projected parallel to the fit plane onto the boundary planes (*b.*) $x = 1$, noise strength as a function of translation, and (*c.*) $y = 1$, noise strength as a function of transcription. The intersection of the fit plane with each boundary plane is shown as a solid line; dotted lines indicate one standard deviation interval.