# Stochastic Biomodelling

Andrzej Mizera (andrzej.mizera@uni.lu)
Faculty of Science, Technology and Communication
University of Luxembourg

# Course content

1. Introduction
   - Stochasticity in biological processes
   - Deterministic vs stochastic biomodelling
2. Prerequisite:
   - Crash course on probability theory
3. Stochastic modelling of chemical kinetics: the chemical master equation (CME)
4. Stochastic simulation of the CME – Gillespie's direct method algorithm
5. Practicals:
   - Implementing the Gillespie's algorithm in MATLAB and investigating its characteristics on various biochemical systems
   - Comparing the obtained simulation results with the solutions in the deterministic formulation

Some words on what is the role of mathematical modelling in Systems Biology

# Course schedule

- **Week 1**
  - Monday 8.2.2016 at 15 - 17
  - Wednesday 10.2.2016 at 15 - 17
  - Thursday 11.2.2016 at 15 - 17

- **Week 2**
  - Monday 15.2.2016 at 15 - 17
  - Wednesday 17.2.2016 at 15 - 17

All lectures will take place in Auditorium XX (Agora building, 1st floor).

# Course materials

- Lecture slides
- Materials for practicals
  - MATLAB m-files with solved programming tasks
- Additional reading: research papers

# Literature

The content of this course is mostly based on the following two books:

1. D. J. Wilkinson. Stochastic Modelling for Systems Biology. 2nd Ed., CRC Press, Boca Raton, 2012.

2. E. Klipp, W. Liebermeister, C. Wierling, A. Kowald, H. Lehrach, and R. Herwig. Systems Biology. A Textbook. Wiley–VCH Verlag GmbH & Co. KGaA, Weinheim, 2009.

# Acknowledgments


- Probability theory based on slides of Dr Tatiana Baumuratova, Systems Biology Group, FSTC, LSRU, University of Luxembourg)

- DTMC/CTMC: based on the slides by Ion Petre on the course "Advanced computational modelling" given at Åbo Akademi University in the period 2012-2013 (lectures 5 & 7).

# Stochastic Biomodelling

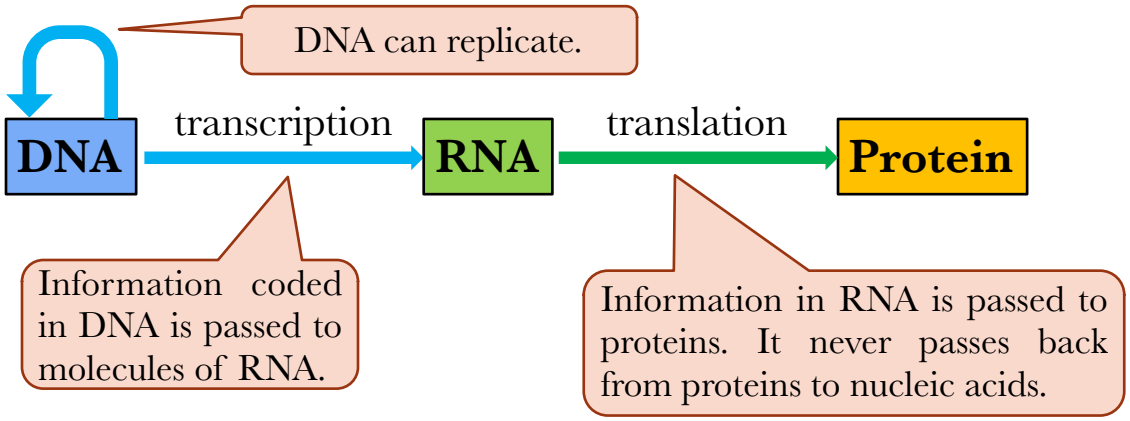
## Introduction: Stochasticity in biological processes

# Stochasticity

"**Stochastic**" is synonymous with "random". The word is of Greek origin and means "pertaining to chance".

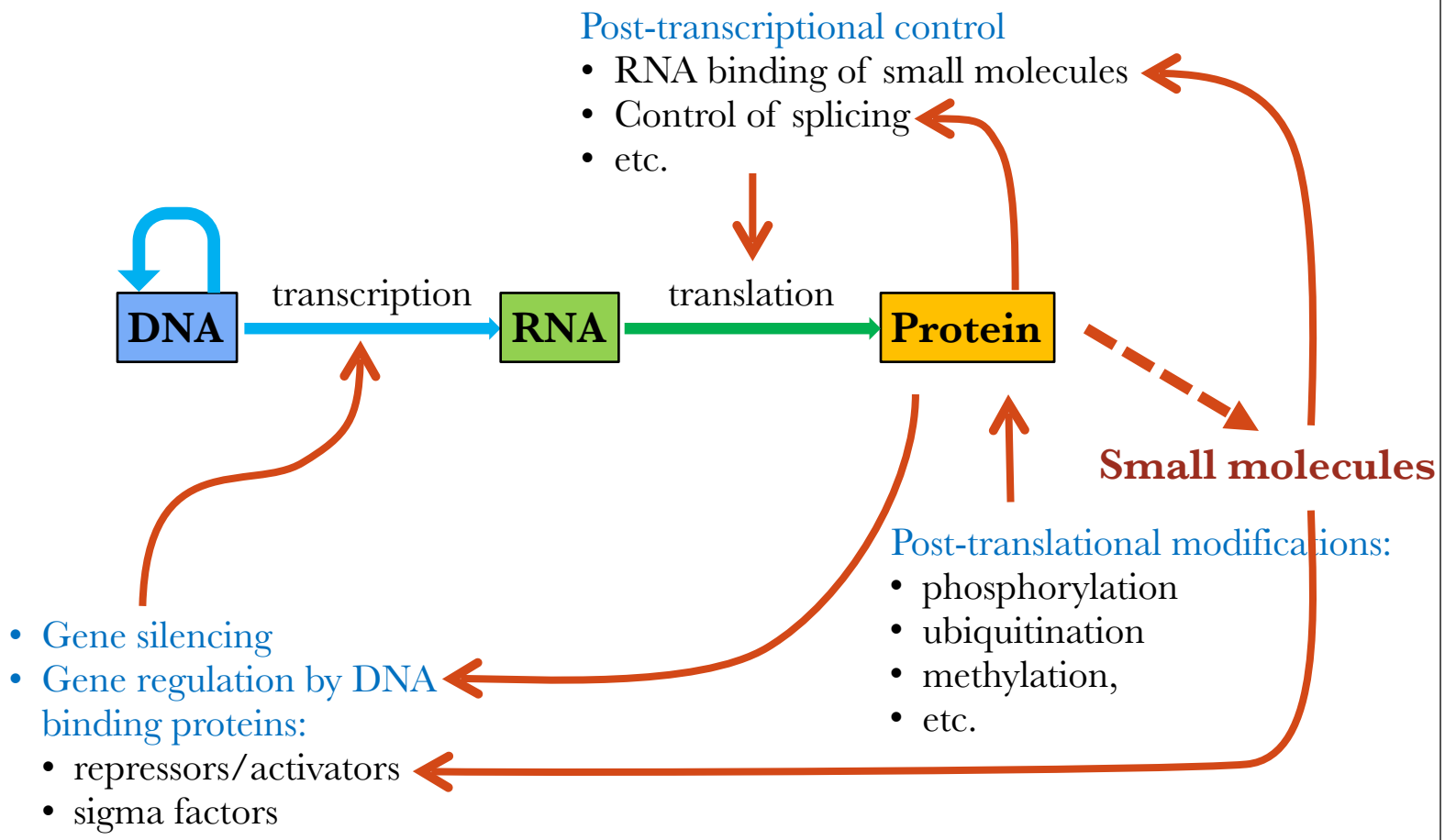
Origlio, Vincenzo. "Stochastic." From MathWorld – A Wolfram Web Resource, created by Eric W. Weisstein. http://mathworld.wolfram.com/Stochastic.html.

# The Central Dogma




- The **central dogma of molecular biology** is an explanation of the flow of genetic information within a biological system.
- Regulation occurs at all points.
- Virtually all cells of a human being have the same DNA.
- Differentiation is a **control problem**.

# Is this a complete picture?



# Stochasticity in gene expression

Cell

Leading Edge
Review

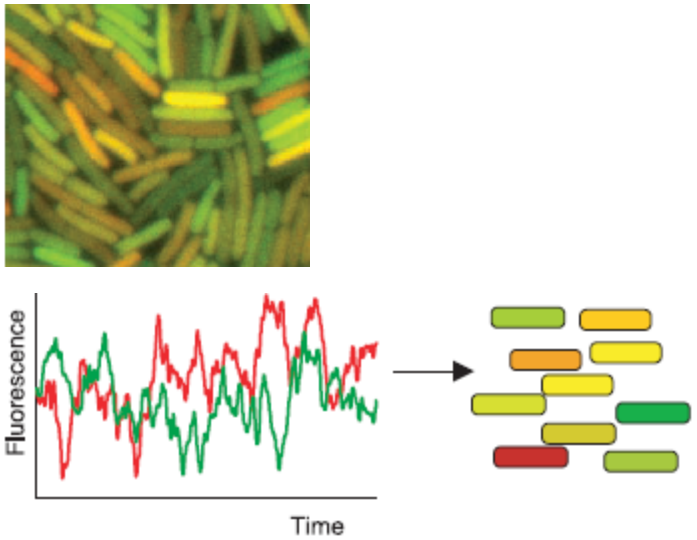
## Nature, Nurture, or Chance: Stochastic Gene Expression and Its Consequences

Arjun Raj[1] and Alexander van Oudenaarden[1,*]
[1]Department of Physics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
*Correspondence: avano@mit.edu
DOI 10.1016/j.cell.2008.09.050

Gene expression is a fundamentally stochastic process, with randomness in transcription and translation leading to cell-to-cell variations in mRNA and protein levels. This variation appears in organisms ranging from microbes to metazoans, and its characteristics depend both on the biophysical parameters governing gene expression and on gene network structure. Stochastic gene expression has important consequences for cellular function, being beneficial in some contexts and harmful in others. These situations include the stress response, metabolism, development, the cell cycle, circadian rhythms, and aging.

# Stochasticity in gene expression

- Different cells behave differently, even if they are genetically identical and have been grown under the same conditions
- The variability of protein levels, metabolic state, and cell morphology is caused by internal factors, like cell cycle phase and by changes in the environment (e.g., nutrients, temperature, cell density).
- However, cells also generate random behaviour by themselves: random fluctuations in gene expression and in other microscopic processes enable cells to switch spontaneously between different states, which allow them to create diversity in their population.




*Escherichia coli* bacteria expressing different fluorescent proteins. With two proteins shown in red and green, the total brightness depends on the sum of both proteins. Different expression levels are partly caused by random events in the cell.

**Source:** M. B. Elowitz, et.al., *Science* (2002)

*Proc. Natl. Acad. Sci. USA*
Vol. 94, pp. 814–819, February 1997
Biochemistry

# Stochastic mechanisms in gene expression

**(prokaryotic genetics/transcriptional regulation/simulation of genetic regulation/stochastic behavior)**

HARLEY H. MCADAMS*† AND ADAM ARKIN‡

*724 Esplanada Way, Stanford, CA 94305; and ‡Department of Developmental Biology, Beckman Center, Stanford University School of Medicine, Stanford, CA 94305

*Communicated by Lucy Shapiro, Stanford University School of Medicine, Stanford, CA, December 4, 1996 (received for review September 14, 1996)*

**ABSTRACT** **In cellular regulatory networks, genetic activity is controlled by molecular signals that determine when and how often a given gene is transcribed. In genetically controlled pathways, the protein product encoded by one gene often regulates expression of other genes. The time delay, after activation of the first promoter, to reach an effective level to control the next promoter depends on the rate of protein accumulation. We have analyzed the chemical reactions controlling transcript initiation and translation termination in a single such "genetically coupled" link as a precursor to modeling networks constructed from many such links. Simulation of the processes of gene expression shows that proteins are produced from an activated promoter in short bursts of variable numbers of proteins that occur at random time intervals. As a result, there can be large differences in the time between successive events in regulatory cascades across a cell population. In addition, the random pattern of expression of competitive effectors can produce probabilistic outcomes in switching mechanisms that select between alternative regulatory paths. The result can be a partitioning of the cell population into different phenotypes as the cells follow different paths. There are numerous unexplained examples of phenotypic variations in isogenic populations of both prokaryotic and eukaryotic cells that may be the result of these stochastic gene expression mechanisms.**

coded by one gene regulates expression of other genes. The time delay in genetically coupled links (Fig. 1) depends on the time required for protein concentration growth, after promoter activation, to the concentration range that controls the next level in the cascade. Conversely, the time delay after the controlling promoter turns off depends on the time for the protein concentration to decay below the effective range. Fig. 1*B* shows a common architecture for such genetically coupled links. In these links, for appropriate combinations of input signals, transcripts are initiated and the protein product accumulates when production exceeds degradation; the increasing protein concentration simply broadcasts the information that the promoter is "on." The message is "received" or detected by the concentration-dependent response at the protein signal's site(s) of action, stimulating a response at each site in accord with that site's chemical behavior. (We use the term "protein signal" to mean the regulatory protein concentration *in its effective form* at its site of action.)

In this paper we examine the properties of a single genetically coupled link as a precursor to modeling networks constructed from many such links. Specifically, we ask what determines the time required for protein concentration to grow to effective signaling levels after a promoter is activated and how statistical variations in this time can affect observed cellular phenomena across a cell population. It has been proposed that the pattern of protein concentration growth is

# Stochastic mechanisms in gene expression

- There are numerous unexplained observations of phenotypic variation in isogenic or clonal populations.
- **Examples:**
  - distinctive individual chemotactic responses observed in clonal bacterial cells grown in homogenous conditions that persist over the cell lifetimes
  - phase variation in *Escherichia coli* expression of type 1 pili in isogenic bacterial populations
  - distribution of generation times of cells in growing *E. coli* cultures

# Stochastic mechanisms in gene expression

Quantitative analysis of the mechanisms underlying all these phenomena requires a statistical description of outcomes and explicit modelling of the stochastic mechanisms in the control logic.

# Stochastic mechanisms in gene expression

- Single genetically coupled link as a precursor to modelling networks constructed from many such links is considered.

- For the same link in different cells of the same genotype, there will be wide ***random variations*** in both the ***times to produce a given protein concentration*** or in the ***number of proteins*** produced when the promoter is transiently activated.

- **Questions**

  - What determines the time required for protein concentration to grow to effective signalling levels after a promoter is activated?

  - How statistical variations in this time can affect observed cellular phenomena across a cell population?

# Stochastic mechanisms in gene expression

- To answer these questions, we need to ***formalise*** and ***quantify*** the notion of randomness in genetic regulatory mechanisms.

- It has been proposed that the pattern of protein concentration growth is ***stochastic***, exhibiting short bursts of ***variable numbers of proteins*** at ***varying time intervals***.

# Stochastic mechanisms in gene expression

- Implications of this noisy pattern of gene expression for cellular regulation include:
  - the **time for the cell to execute cascaded functions can vary widely** across isogenic cells in a population (due to switching delay for genetically coupled links);
  - the overall ***regulatory circuit design*** is probably strongly driven by the needed ***determinism*** in outcome for circuits constructed from these ***highly noisy components***;
  - ***stochastic simulation techniques*** must be used to model regulated networks where high noise levels in parts of the network can produce statistical variation in phenotypic outcomes.

# Stochastic effects in signalling

T. Lipniacki, B. Hat, J. R. Faeder, W. S. Hlavacek. Stochastic effects and bistability in T cell receptor signaling. *Journal of Theoretical Biology*, 254:110-122, 2008.

"The stochastic dynamics of T cell receptor (TCR) signaling are studied using a ***mathematical model intended to capture kinetic proofreading*** (sensitivity to ligand–receptor binding kinetics) and ***negative*** and ***positive feedback regulation*** mediated, respectively, by the phosphatase SHP1 and the MAP kinase ERK. …"

"Analysis of the model indicates that ***TCR signaling dynamics are marked by significant stochastic fluctuations and bistability***, which is caused by the competition between the positive and negative feedbacks. ***Stochastic fluctuations are such that single-cell trajectories differ qualitatively from the trajectory predicted in the deterministic approximation of the dynamics***. ***Because of bistability, the average of single-cell trajectories differs markedly from the deterministic trajectory. Bistability combined with stochastic fluctuations allows for switch-like responses to signals, which may aid T cells in making committed cell-fate decisions.***"

# Stochastic Biomodelling

## Introduction: Deterministic vs stochastic biomodelling

# Mathematical/computational biomodelling

- Existing mathematical/computational approaches used to model biological processes differ in their underlying assumptions and the level of resolution they can provide.

- A broad classification of these methods separates the resulting models into **two classes**:
  1. ***deterministic*** (macroscopic) and
  2. ***stochastic*** (mesoscopic),

  where each of these two classes embodies various subclasses with their different mathematical formalisms.

# Deterministic vs. stochastic biomodelling

- Ignoring quantum mechanical effects, biological systems are often viewed as **deterministic**, with their **dynamics entirely specified**, given
  - sufficient information on the state of the system (position, orientation and momentum of every single molecule) and
  - a complete understanding of the chemistry and physics of the interactions between biomolecules.

- Unfortunately, we are still unable to model biological systems of realistic complexity and size using such a molecular dynamic approach.

- Therefore, current models admit far-reaching simplifications, which result in a higher level view of the system being modelled.

# Biomodelling with differential equations: some physical difficulties

- Assumes that the time evolution of a chemically reacting system is both continuous and deterministic
- Difficulties with this assumption:
  - the time evolution is **NOT** continuous: molecular population levels increase and decrease only with discrete amounts
  - the time evolution is **NOT** deterministic (even when ignoring the quantum effects and assuming classical mechanics for the molecular kinetics)
    - it is only deterministic in the full position-momentum phase space (knowing the positions and velocities of all molecules)
- However:
  - in many cases the time evolution of a chemical system can be treated as continuous and deterministic
  - **Solution in these cases: deterministic models!**
  - the difficulties come when some species populations are small, or in conditions of chemical instability
  - **Solution in these cases: stochastic models!**

# Deterministic vs. stochastic modelling

T. Lipniacki and M. Kimmel. Deterministic and Stochastic Models of NFκB Pathway. *Cardiovasular Toxicology* 7:215-234, 2007.

> *" … ordinary differential equations sometimes are considered adequate to describe cell population models, i.e., models which predict the average behavior in the population. However, the average behavior may be very different from the behavior of any cell in the population. If, for example, half of cells in the population choose proliferation pathway and the other half the apoptotic pathway, then the average will not correspond to any biological process."*

# Mathematical models

## Stochastic model

- Given the current state of the system, **many** possible future behaviour are possible
- Well-suited to model ***individual***, rather than average behaviour
- Probability distributions dictate the behaviour of the system
- Typical
  - Number of molecules are modelled
  - Reactions are taking place following "collisions" among the reactants
  - Markov processes

## Deterministic model

- Given the current state of the system, all future behaviour of the system is **uniquely** defined
- Usually the model reflects the ***average*** behaviour of the observed system
- Typical methods used: differential or difference equations
- Typical:
  - Concentrations of molecules are modelled
  - Reactions are taking place diffusion-like (gradient-like)
  - Differential equations

# Mathematical models

## Stochastic model

- The objects
  - the ***number of copies*** of all species of interest
  - the rates of all reactions
- Main assumptions
  - The system is well-stirred
  - The system is at thermodynamical equilibrium
- Methods
  - ***Probability theory***
  - Some expertise from modelling in physics, especially in quantum physics

## Deterministic model

- The objects
  - the ***concentrations*** of all species of interest
  - the rates of all reactions
- Main assumptions
  - The system is well-stirred
  - The system is at thermodynamical equilibrium
- Methods
  - ***Mathematical analysis (continuous mathematics)***
  - Arguably the ***most developed part of mathematics***
  - Great expertise from modelling in physics, chemistry, engineering

# Writing the model

## Stochastic model

- It is the description of a continuous time, discrete state Markov process
- *Grand probability function*: $P(X_1,X_2,\ldots,X_n,t)$ is the probability that at time t there are $X_1$ molecules of species $S_1$, …, $X_n$ molecules of species $S_n$
- The ***grand probability function*** may be obtained through a differential equation: the ***chemical master equation***
  - Reason what is the probability of being in a certain state after one step

## Deterministic model

- The reaction rate gives the amount with which the concentration of every metabolite involved in the reaction changes per unit of time
  - For a consumed metabolite, the change will be $-v(t)$
  - For a produced metabolite, the change will be $v(t)$

# Stochastic model of biochemical reactions

- The state of a biochemical network ***evolves continuously*** through time with ***discrete changes*** in state occurring as a result of ***reaction events***. These reaction events are ***random*** and are governed by ***probabilistic laws***.

- Speaking about the ***state of a biochemical network*** we consider:
  - ***reactions*** that take place
  - ***changes*** of the network state as results of reactions.

- But:
  - ***Which*** reaction will occur?
  - At ***what time***?

**Answers to these questions can be obtained with the use of probability theory.**