Randomness and information in biological data
BIO-369

Prof. Anne-Florence Bitbol

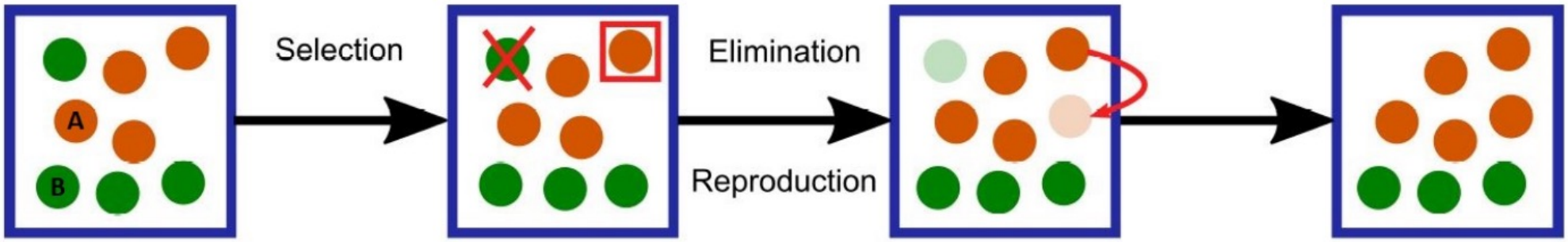
EPFL

Lecture 5

Outline of the course

Random walks – Reminder

- **The Moran model in population genetics**



Schematic of one step of the Moran process
The population comprises two types of individuals, type A (orange) and type B (green)

Outline of the course

What is the average number of copies of a given protein per cell in bacteria?

0%	A.	A few copies
0%	B.	A few dozens of copies
0%	C.	A few hundreds of copies
0%	D.	A few thousands of copies
0%	E.	A few tens of thousands of copies
0%	F.	A few hundreds of thousands of copies

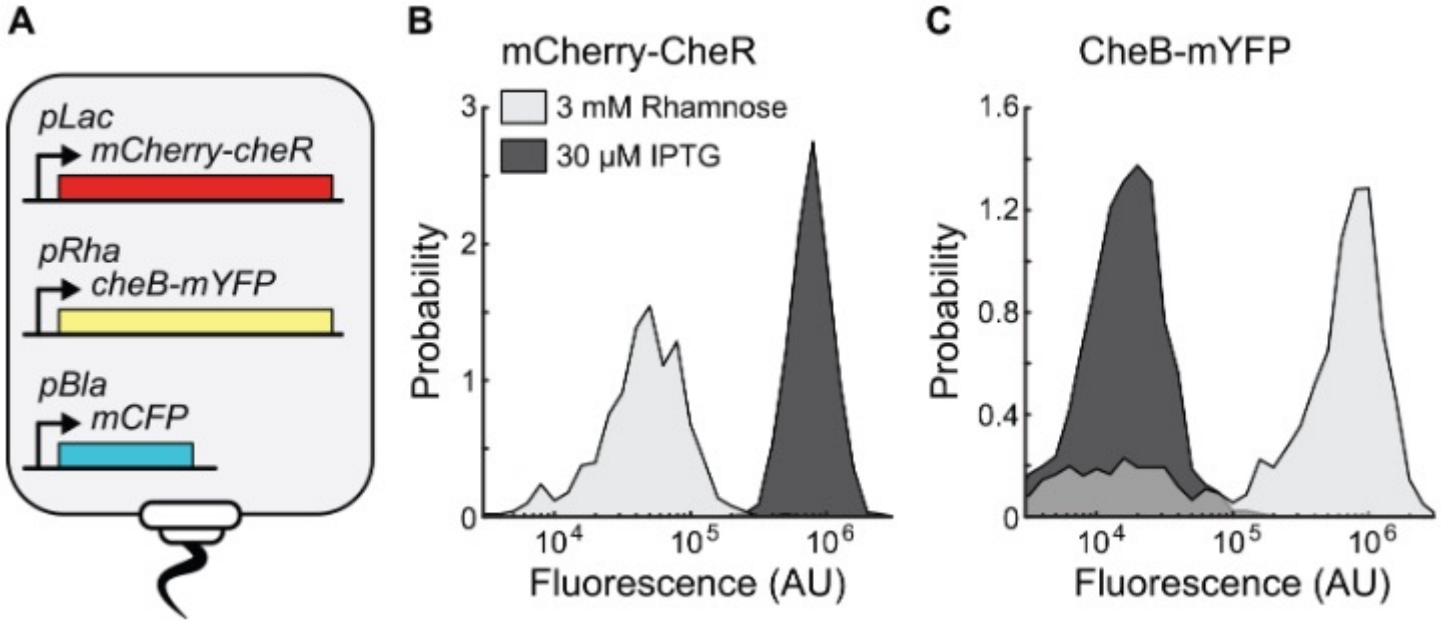
To answer, please:

- Connect to http://ttpoll.eu
- Enter the session ID **bio369**
- Select your answer

Motivation

- **Protein abundances are heterogeneous across cells**

Expression of fluorescently labeled chemotaxis proteins with inducible promoters under two conditions

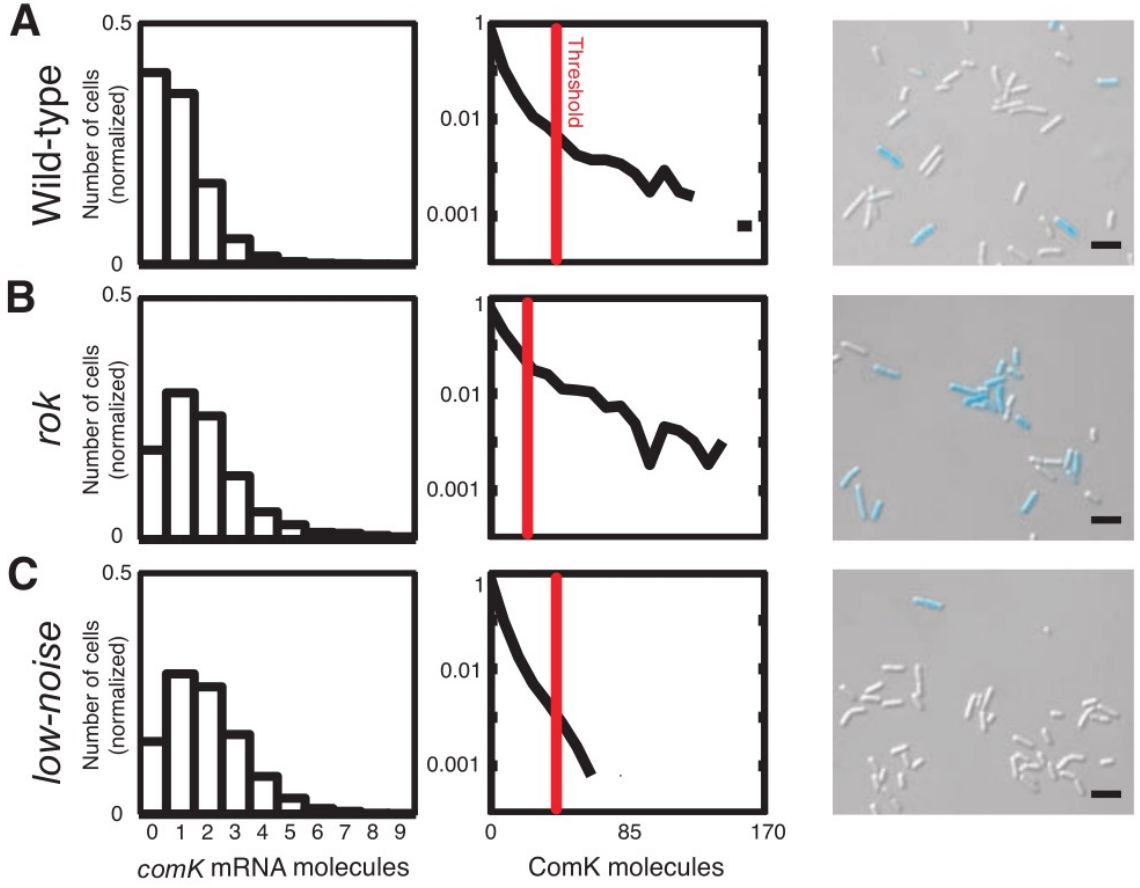


Dufour et al 2016

Motivation

- **Protein abundances can have major consequences**

  Noise in the expression of the gene comK determines the transition to competence for DNA uptake in *Bacillus subtilis* as it enters the stationary growth phase (left and middle: model; right: scale bars 4 μm)



Maamar et al 2007

rok strain: higher abundance of ComK (no Rok = transcriptional repressor of comK)

Low-noise strain: *rok* + reduced translational efficiency (in prokaryotes, translation is noisier than transcription)

During dt, how can the number n of protein copies vary?

0%	A.	It stays constant
0%	B.	It increases by 1
0%	C.	It decreases by 1
0%	D.	A, B and C are all possible
0%	E.	B and C are possible, but not A

To answer, please:

- Connect to http://ttpoll.eu
- Enter the session ID **bio369**
- Select your answer

After a long time, what will happen to the number of protein copies?

0% A. It will either go to 0 or reach a maximum value and stay there forever

0% B. It will stabilize at a steady-state value

0% C. It will fluctuate around a steady-state mean value

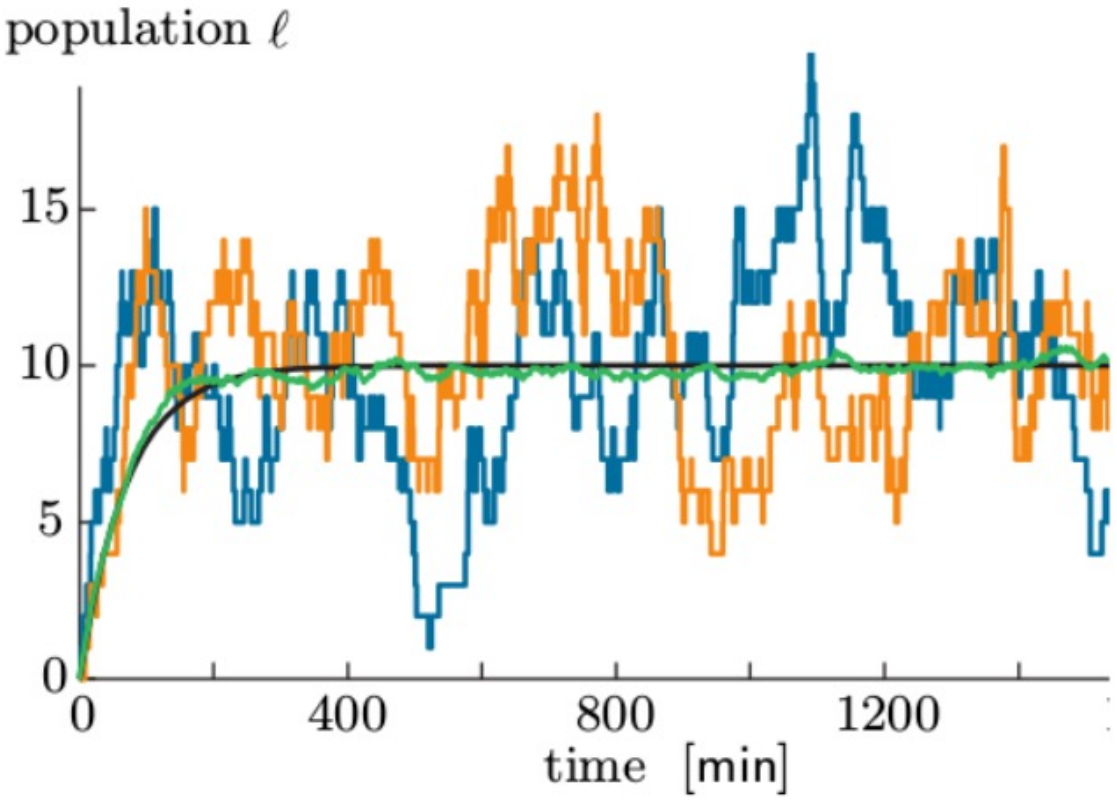
0% D. It depends

To answer, please:

- Connect to http://ttpoll.eu
- Enter the session ID **bio369**
- Select your answer

Protein abundances in single cells

- **Protein abundance as a random walk**

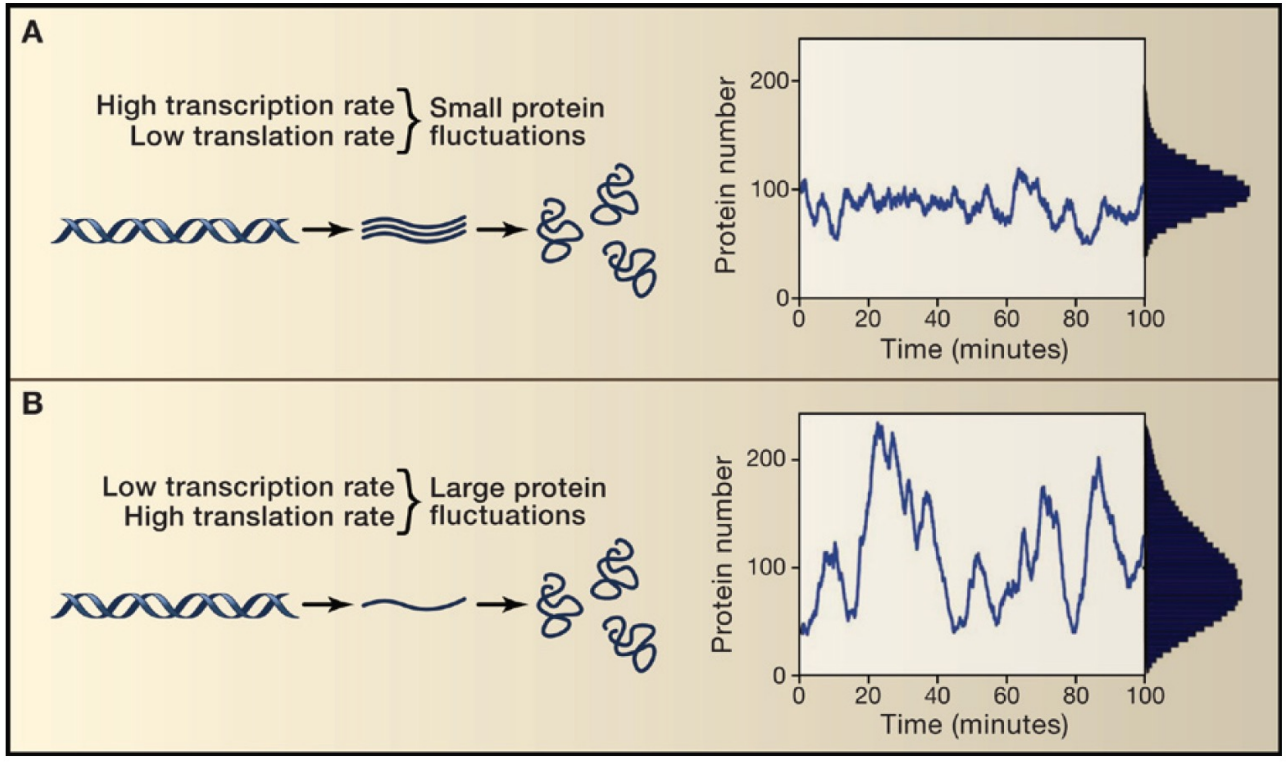


Simulation of a simple model for the abundance of a protein expressed in small copy numbers
Orange / blue: number of proteins versus time in individual realizations ℓ starting from 0 protein
Green: mean over 200 such replicates; Black: deterministic approximation

Gene expression and protein abundances

- **Transcription and translation**

Raj & van Oudenaarden 2008

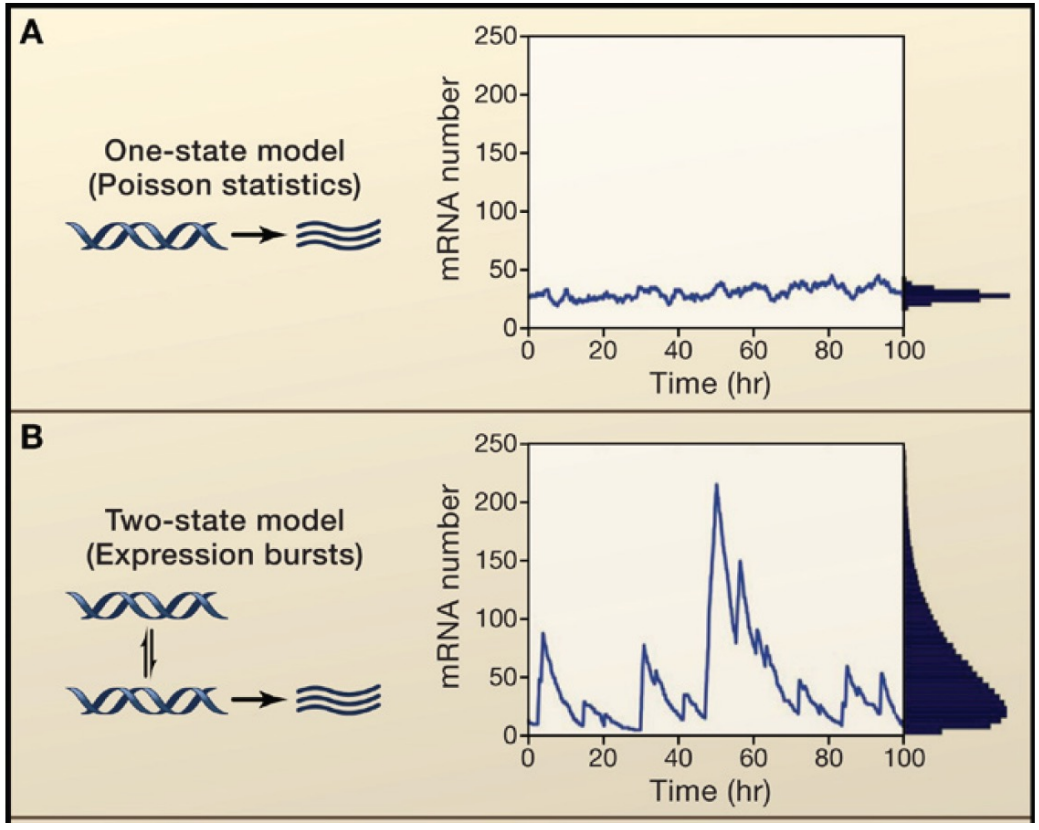


Noise in prokaryotic gene expression depends on the rates of transcription and translation
Translation is bursty

Gene expression and protein abundances

- **Gene regulation: expressed and repressed gene states**

Raj & van Oudenaarden 2008

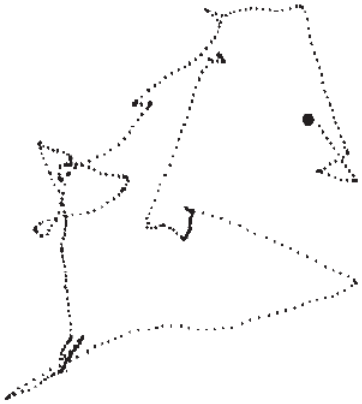


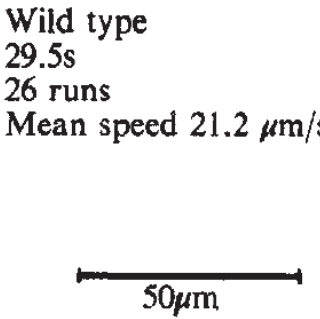
In eukaryotes, genes can be in ON (expressed) or OFF (repressed) states, i.e. be transcribed or not. Transcription is bursty because of the existence of these two states (→ problem class)

Random walks

Bacterial chemotaxis

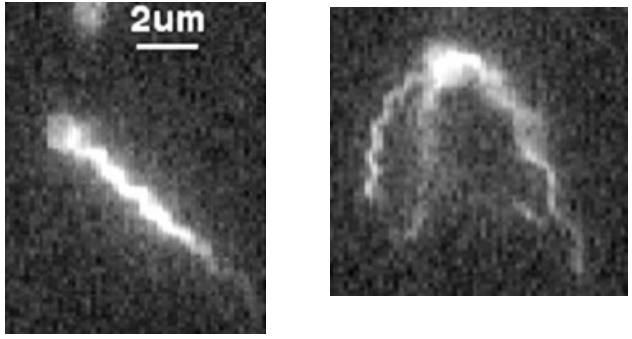
- Motion of swimming, flagellated bacteria: 3D random walk with “runs” and “tumbles”





Berg & Brown (1972)

- Runs → CCW rotation of the flagella
- Tumbles → CW rotation of the flagella

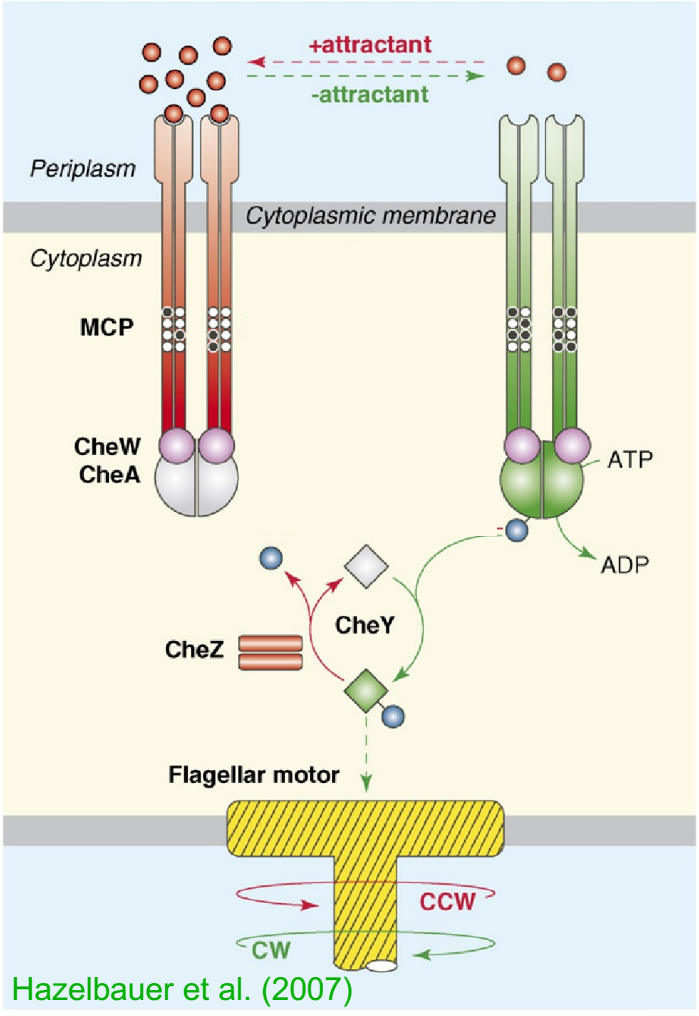


Turner et al. (2000)

- Gradients of chemicals can induce a change of tumbling frequency & CW bias
  → Chemotaxis: bacteria are attracted or repelled by various chemicals

Random walks

- **Chemotaxis transduction pathway in *Escherichia coli***

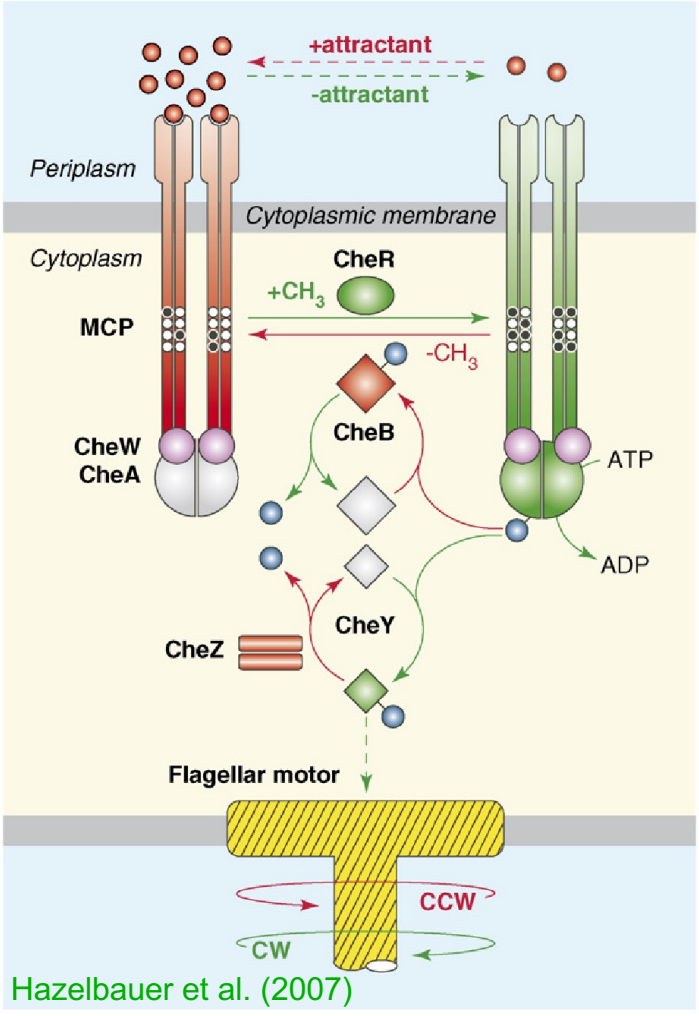


Hazelbauer et al. (2007)

- **If fewer attractant molecules bind to MCP chemoreceptors (transmembrane proteins):**
- The autophosphorylation activity of the histidine kinase CheA (bound to receptors) increases
- CheA-P phosphorylates CheY, a cytoplasmic protein (response regulator)
- CheY-P binds to FliM protein of flagellar motor → switching to CW rotation: tumbling
- **If more attractants bind to chemoreceptors:**
- The activity of CheA decreases CheY-P is dephosphorylated by CheZ (phosphatase)
- The motor reverts to its default state, CCW → running

Random walks

- **Chemotaxis transduction pathway in *Escherichia coli***



Hazelbauer et al. (2007)

- **If fewer attractant molecules bind to MCP chemoreceptors (transmembrane proteins):**
- The autophosphorylation activity of the histidine kinase CheA (bound to receptors) increases
- CheA-P phosphorylates CheY, a cytoplasmic protein (response regulator)
- CheY-P binds to FliM protein of flagellar motor → switching to CW rotation: tumbling
- **If more attractants bind to chemoreceptors:**
- The activity of CheA decreases CheY-P is dephosphorylated by CheZ (phosphatase)
- The motor reverts to its default state, CCW → running

Outline of the course

Outline of the course

II Extracting information from biological data

1 Quantifying randomness and information in data: entropy
- 1.1 Notion of entropy
- 1.2 Interpretation of entropy
- 1.3 Entropy in neuroscience data: response of a neuron to a sensory input

2 Quantifying statistical dependence
- 2.1 Covariance and correlation
- 2.2 Mutual information
- 2.3 Identifying coevolving sites in interacting proteins using sequence data

3 Inferring probability distributions from data
- 3.1 Model selection and parameter estimation: maximum likelihood
- 3.2 Introduction to maximum entropy inference
- 3.3 Predicting protein structure from sequence data

4 Finding relevant dimensions in data: dimension reduction
- 4.1 Principal component analysis
- 4.2 Beyond principal component analysis

5 Introduction to Bayesian inference