The Mystery of Gene Expression

Emanuele Leoncini

Junior Seminar

June 18th 2013


RAP

Réseaux , Algorithmes et Probabilités

Réseaux , Algorithmes et Probabilités

- Communication networks (models and algorithms)
- New math tools for probabilistic models of complex networks

- Communication networks (models and algorithms)
- New math tools for probabilistic models of complex networks

Bike-sharing system Vélib

Stochastic modeling of biological phenomena

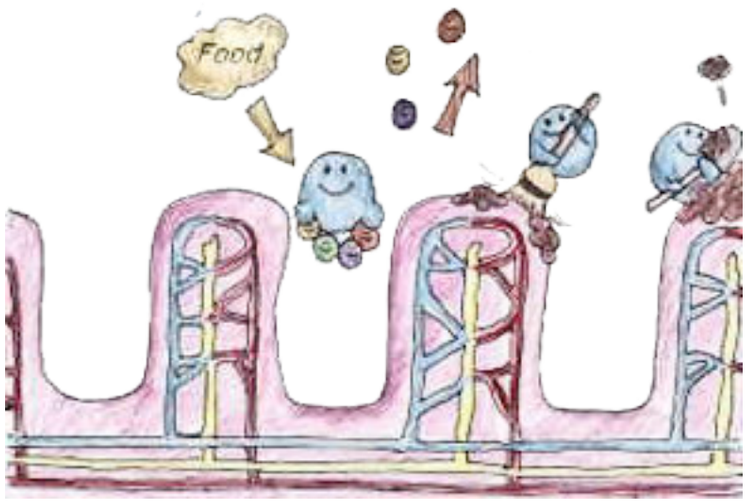
Bacteria

Bacteria

- first life-form on Earth (~4 billion years ago)
- $5 \cdot 10^{30}$ bacteria on Earth
- independent “simple” organisms

Bacteria

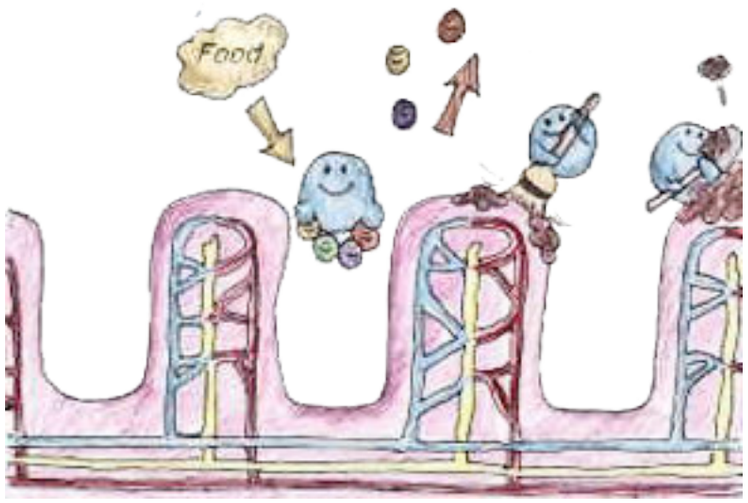
- first life-form on Earth (~4 billion years ago)
- $5 \cdot 10^{30}$ bacteria on Earth
- independent “simple” organisms



The Good (probiotics)

Bacteria

- first life-form on Earth (~4 billion years ago)
- $5 \cdot 10^{30}$ bacteria on Earth
- independent “simple” organisms

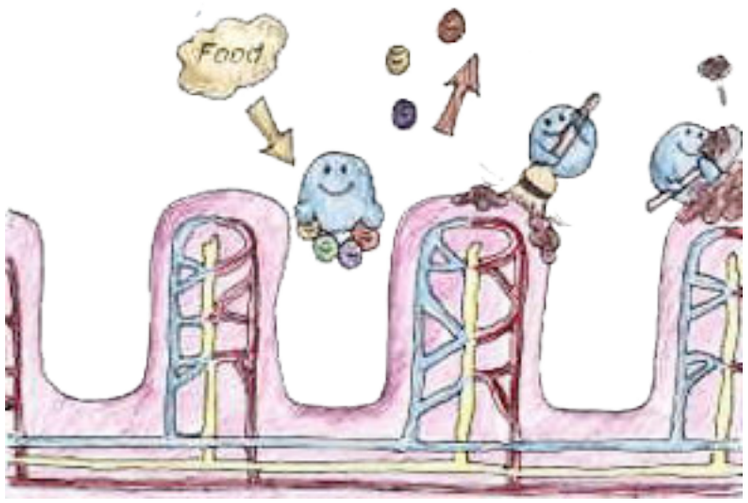


The Good (probiotics)

The Bad (predator)

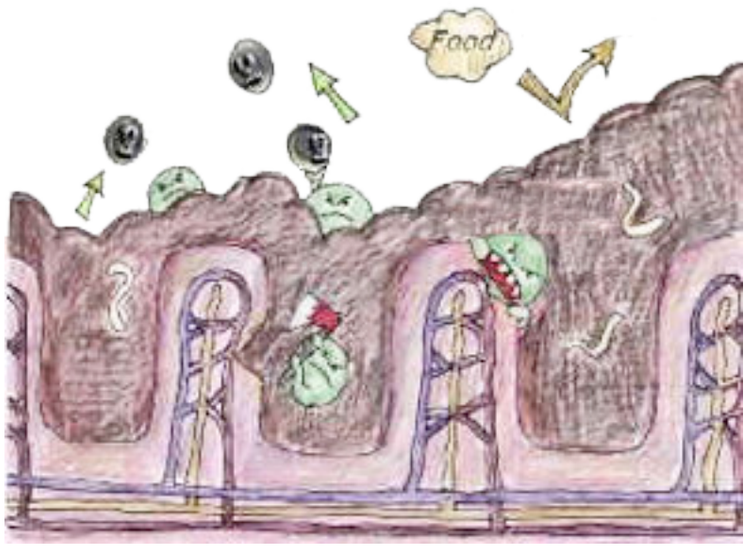
Bacteria

- first life-form on Earth (~4 billion years ago)
- $5 \cdot 10^{30}$ bacteria on Earth
- independent “simple” organisms



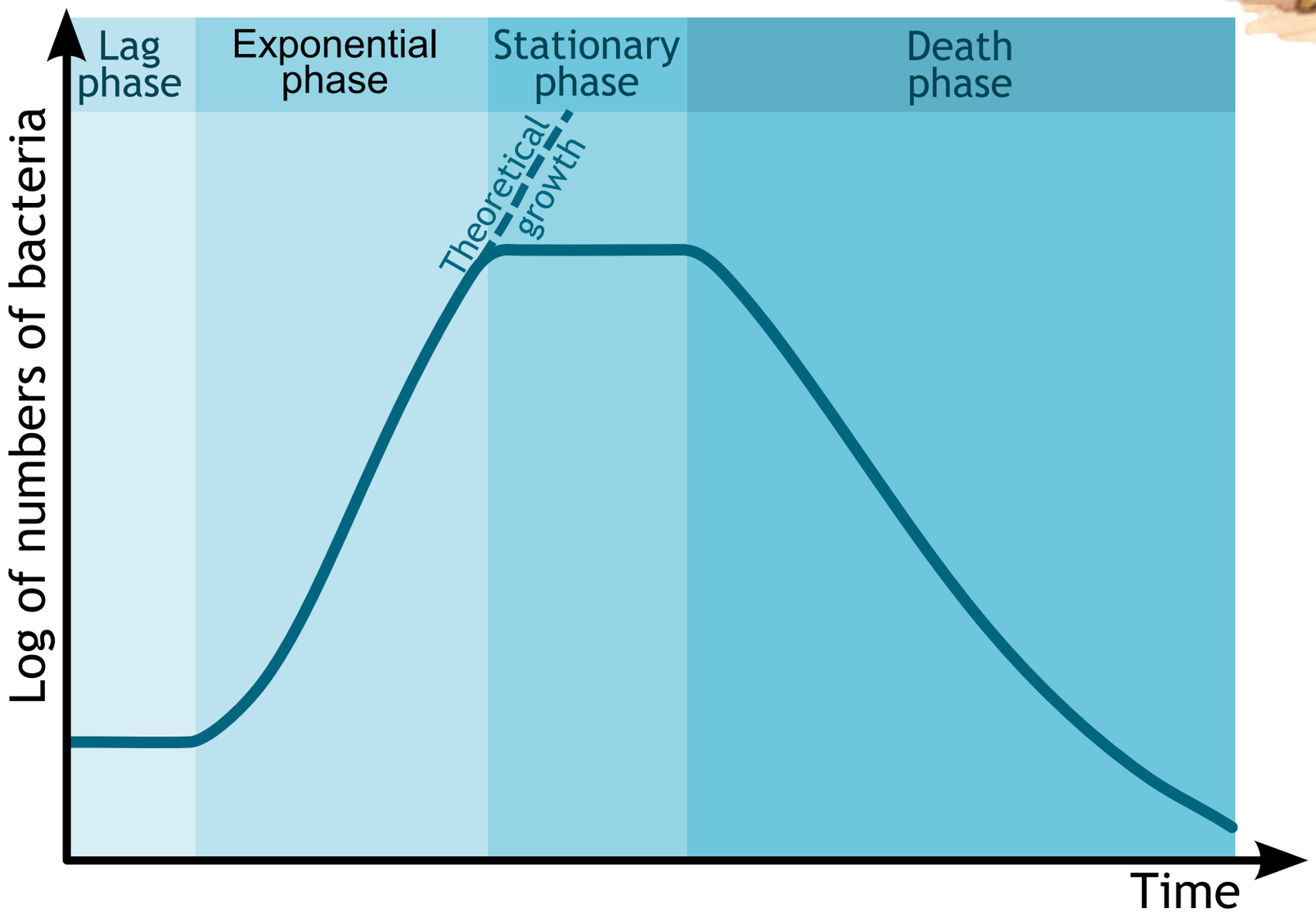
The Good (probiotics)

The Bad (predator)

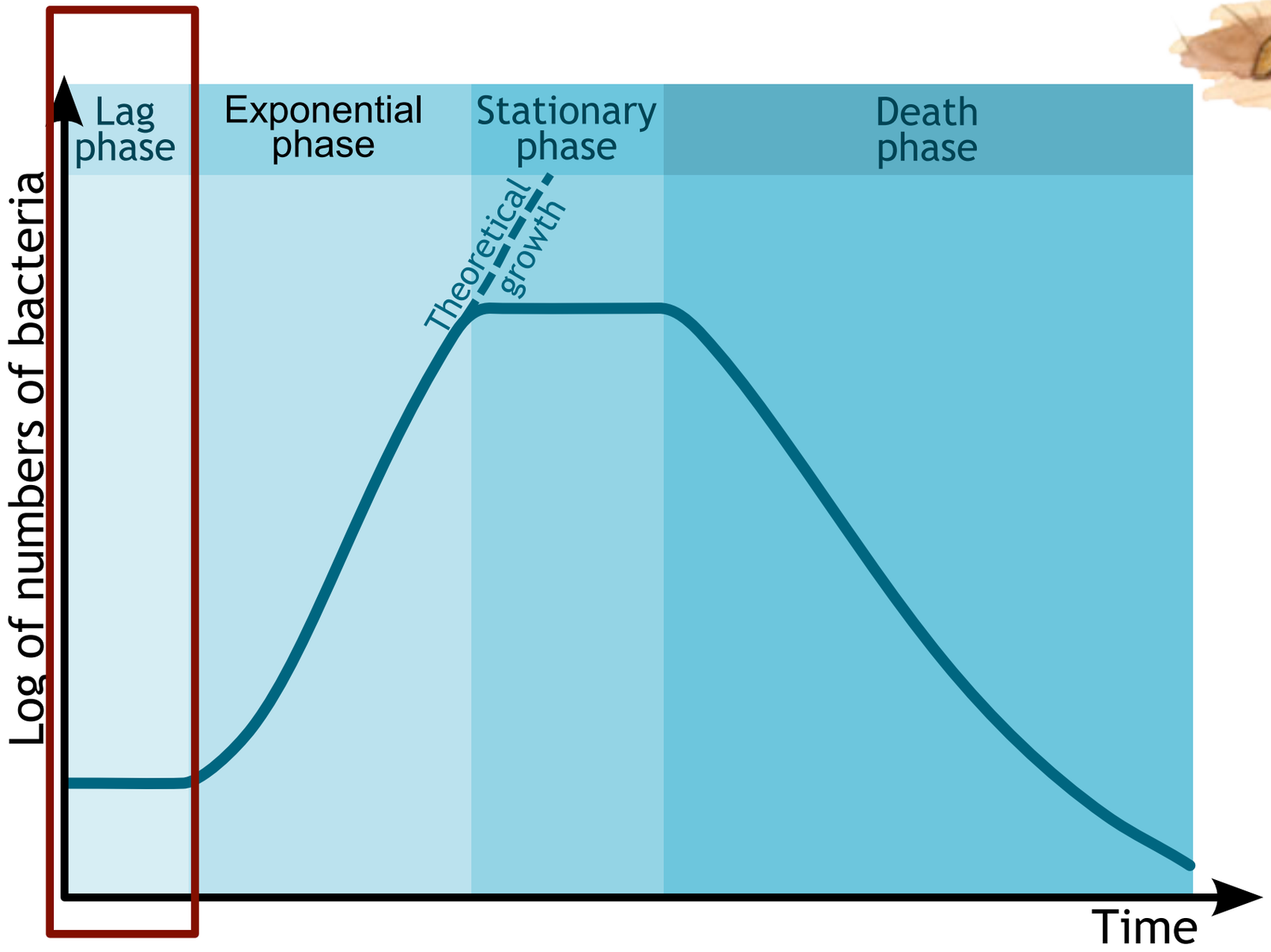


The Ugly (pathogen)

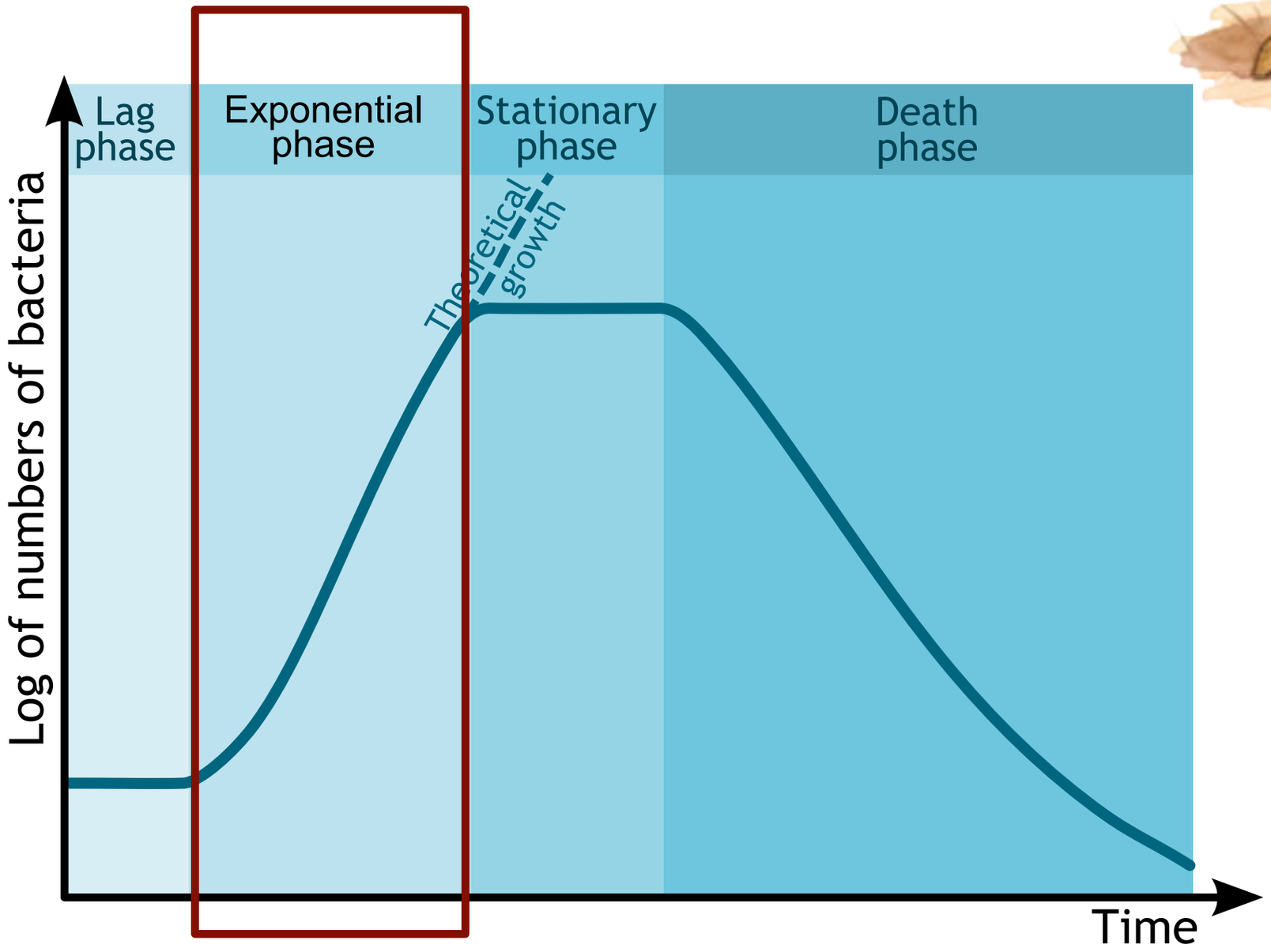
Bacteria behaviour



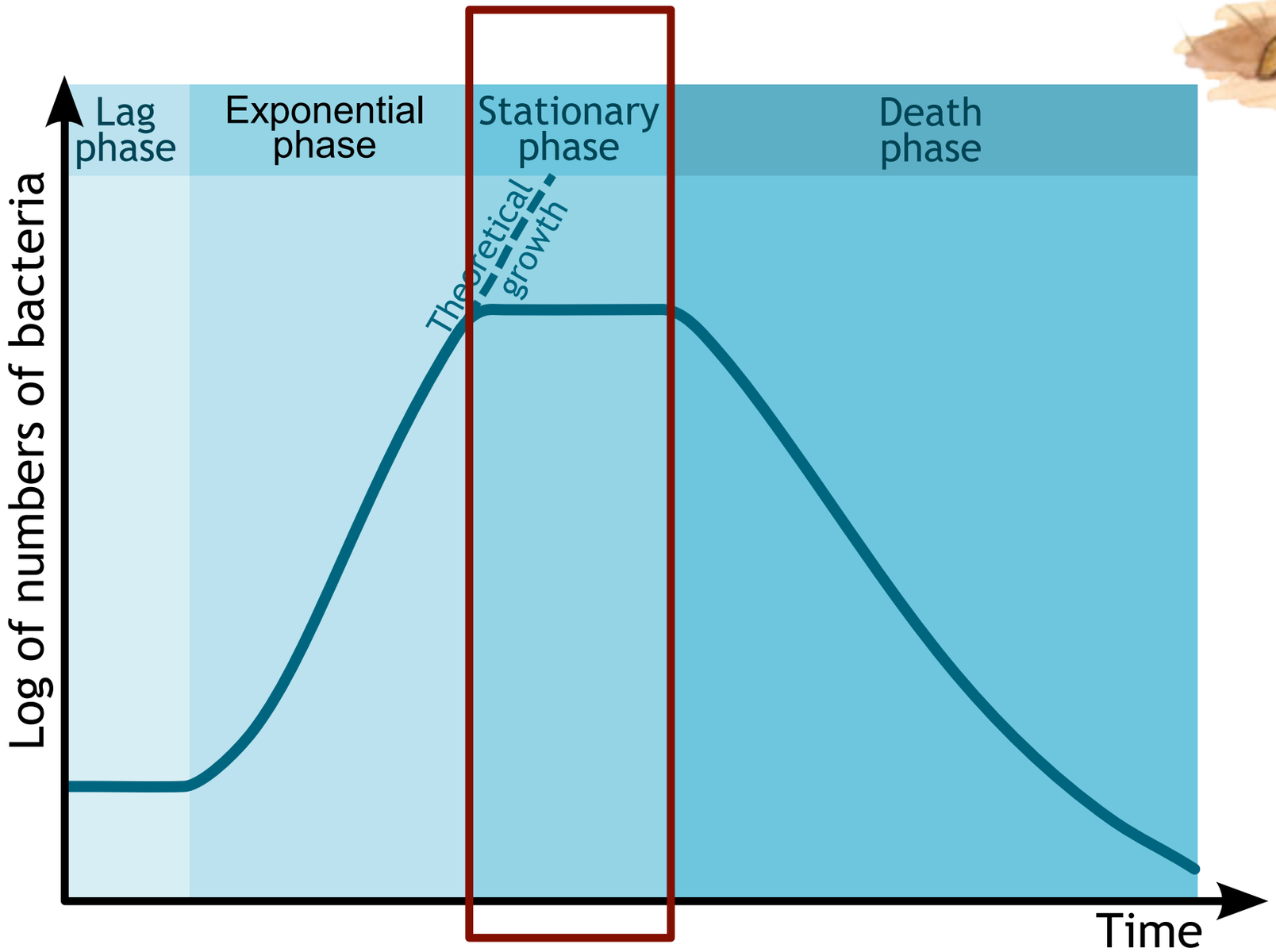
Bacteria behaviour

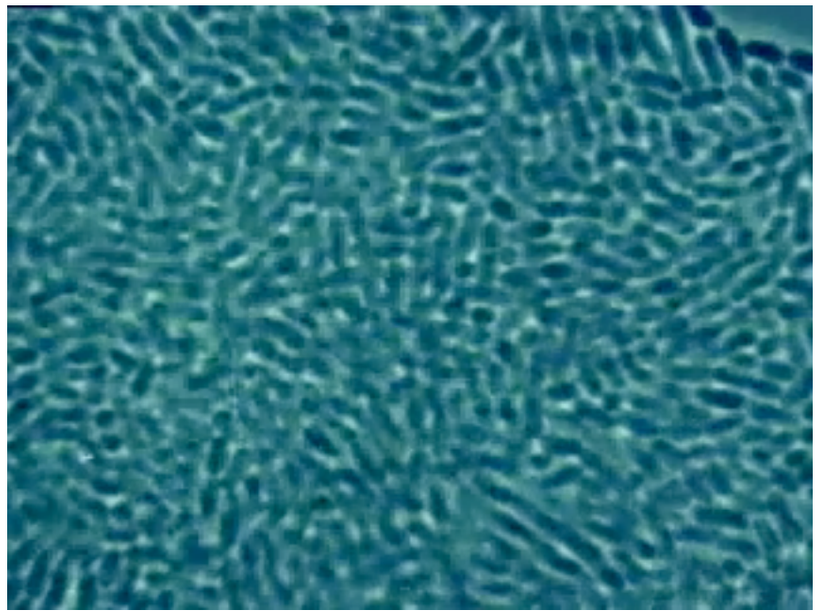


Bacteria behaviour

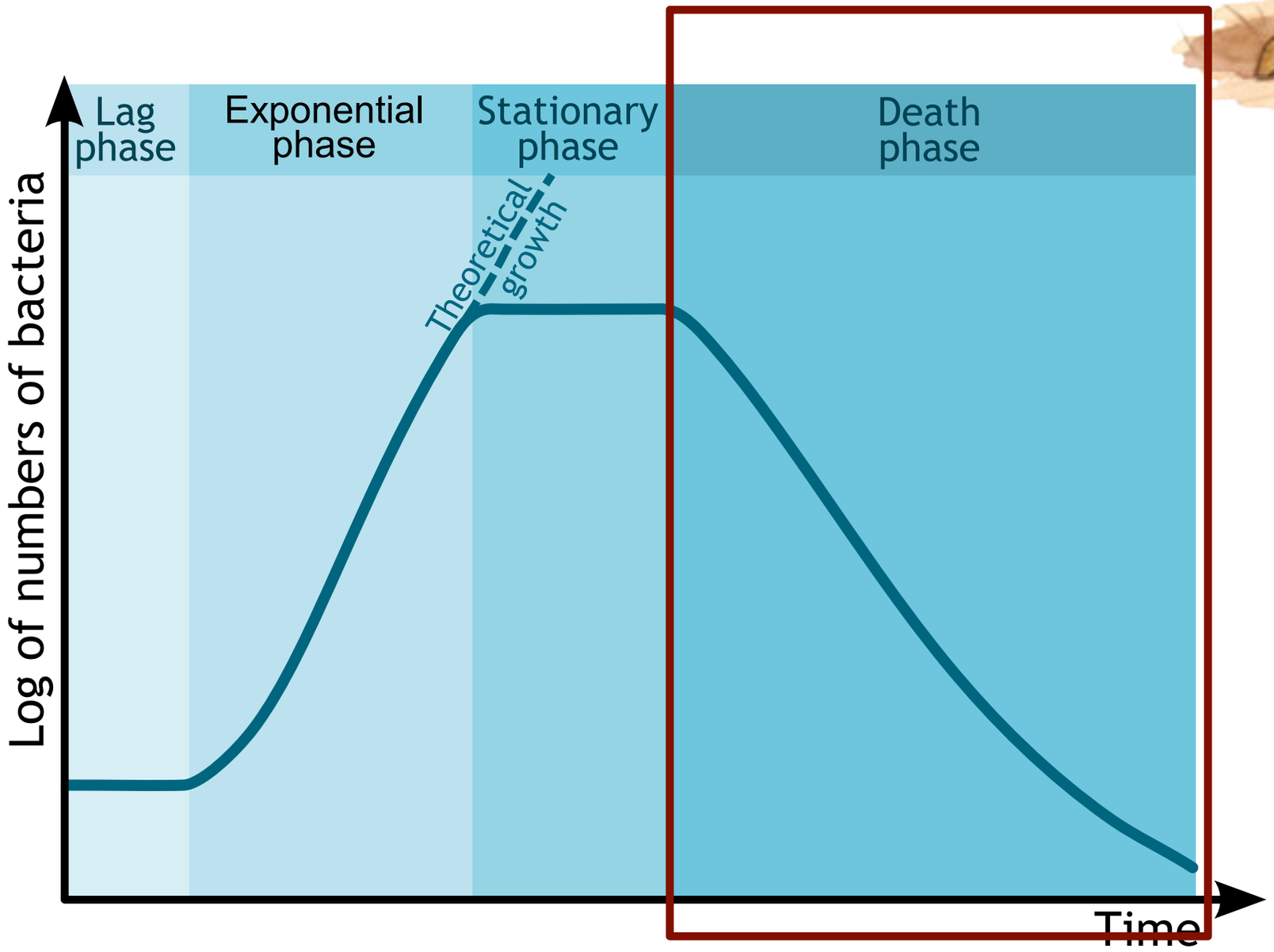


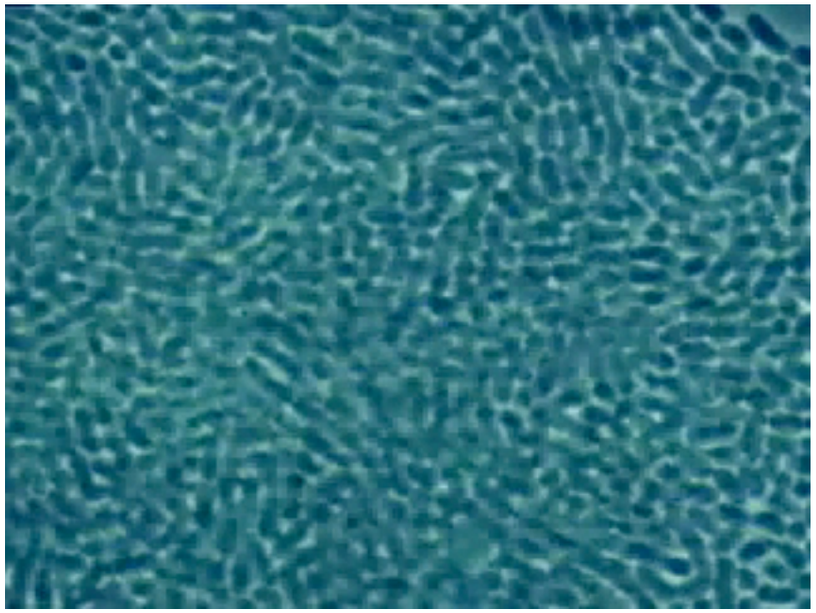
Bacteria behaviour



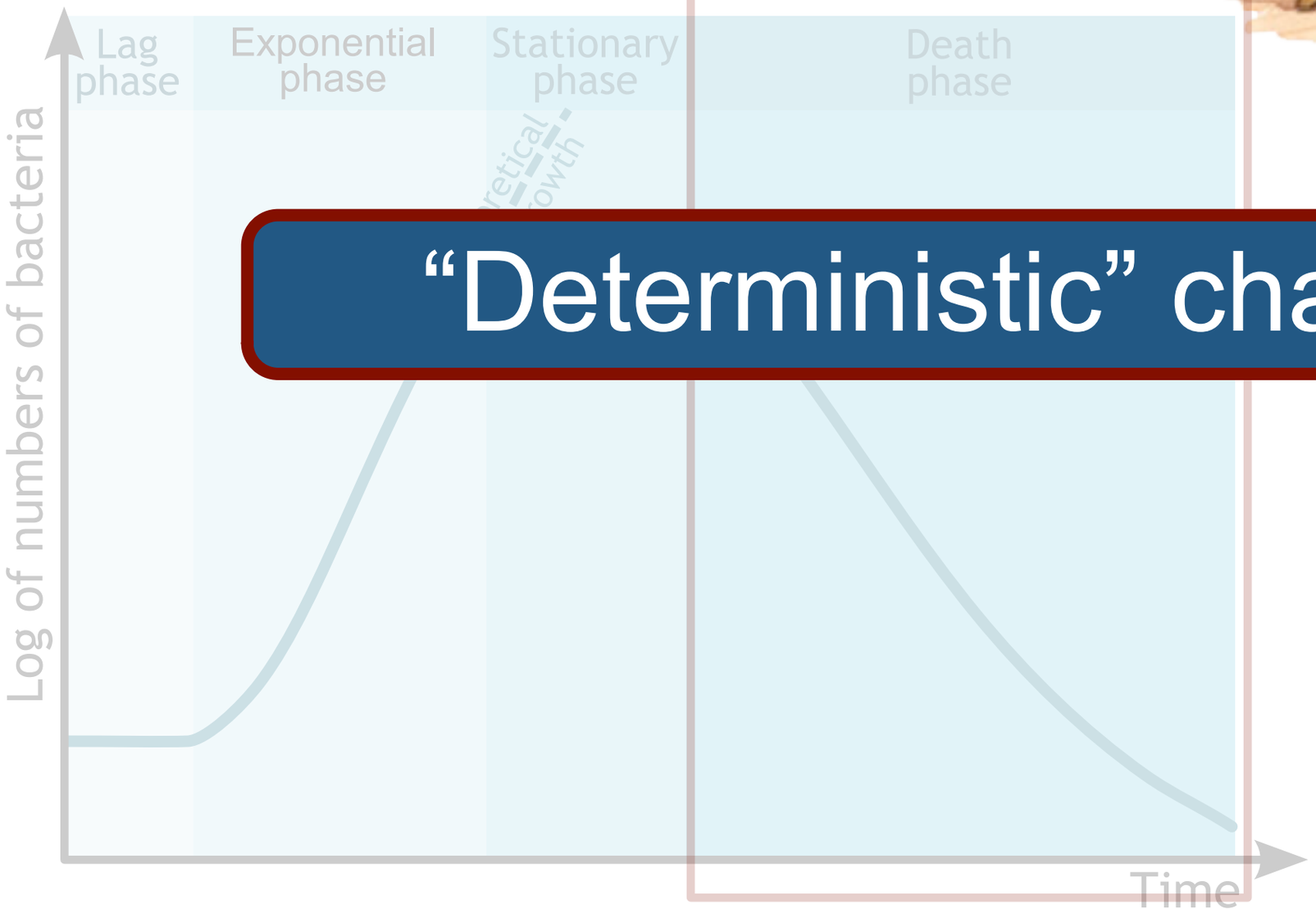


Bacteria behaviour





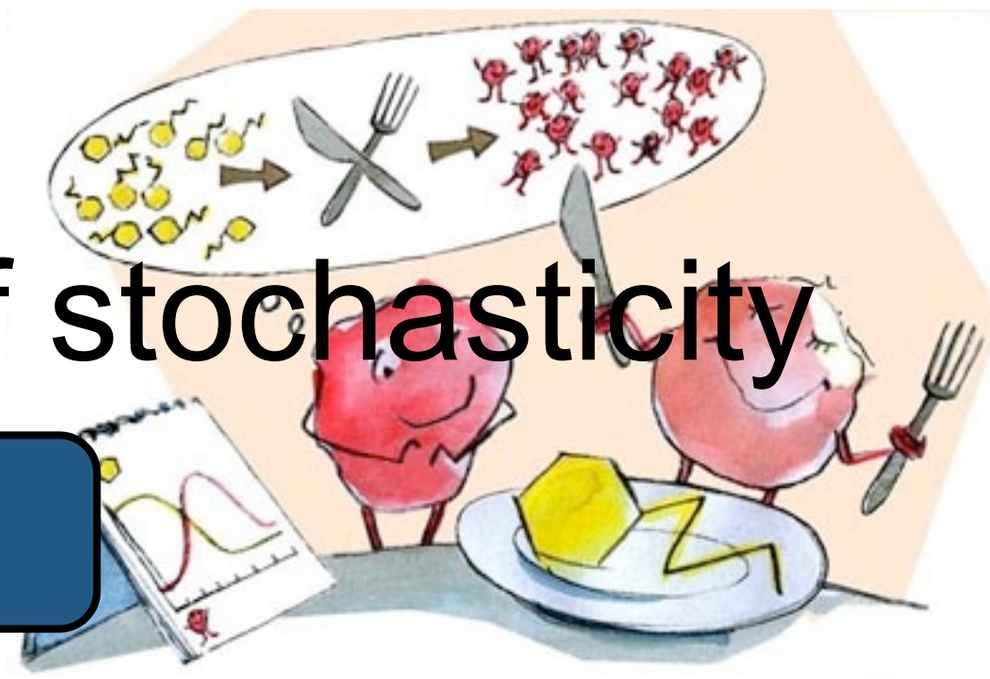
Bacteria behaviour



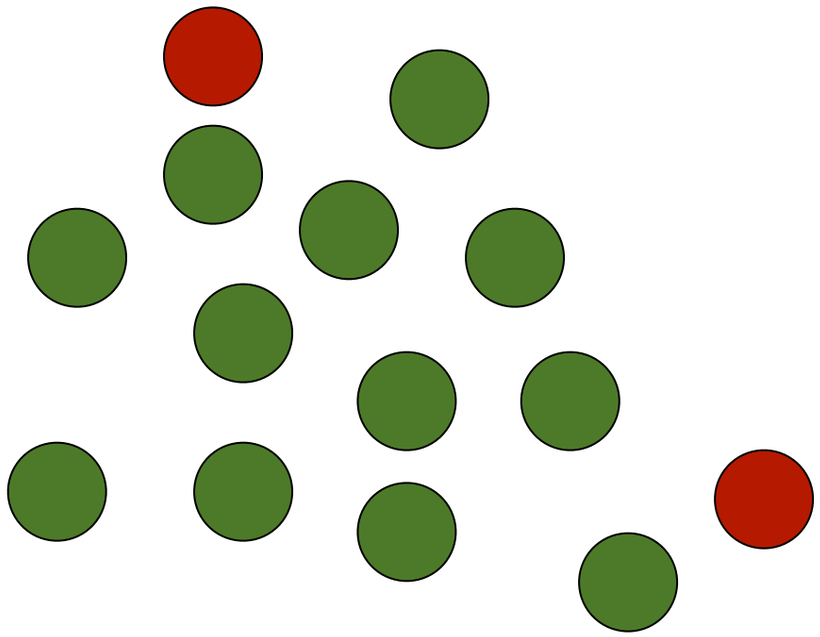
“Deterministic” character

Bacteria -- Two levels of stochasticity

Bacteria -- Two levels of stochasticity

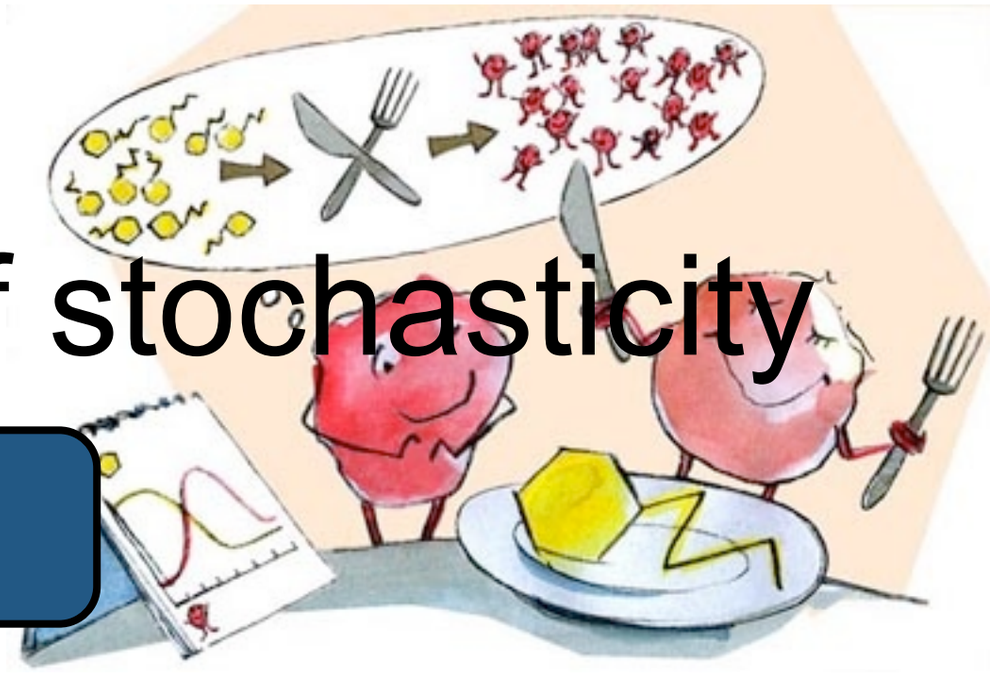


1. Stochastic Decision

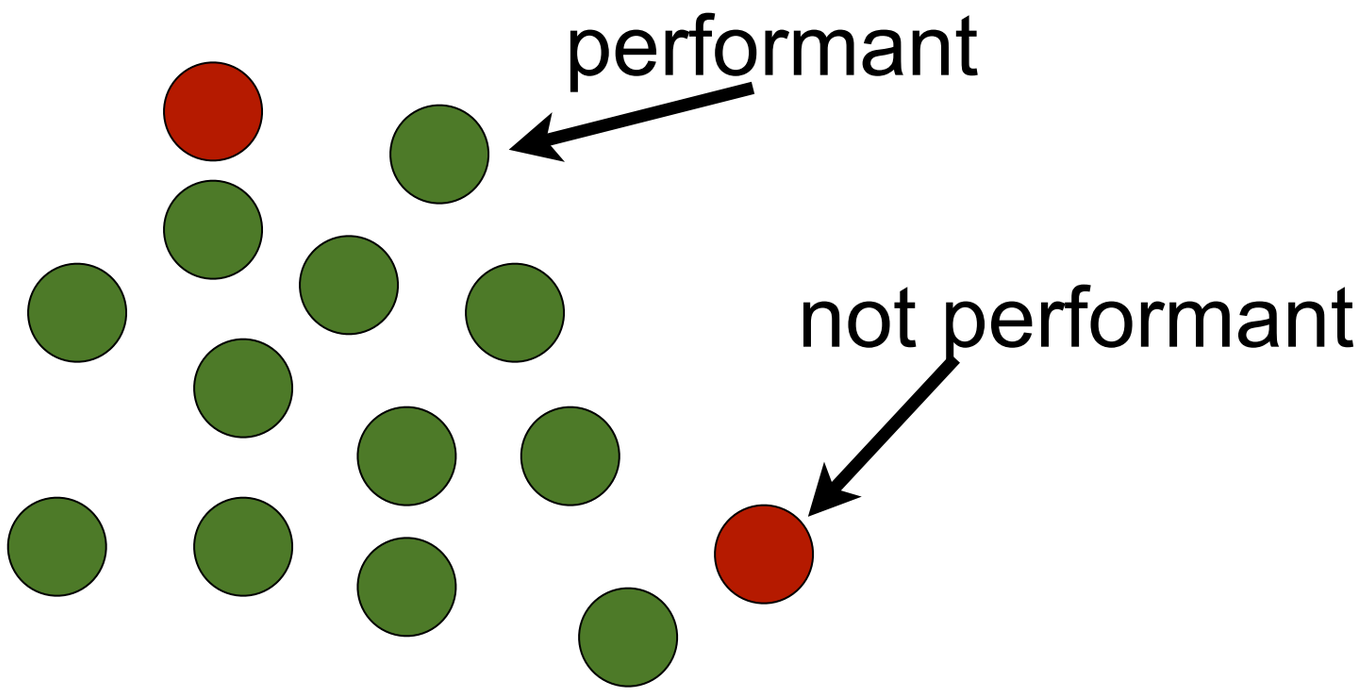


Abundant nutriment

Bacteria -- Two levels of stochasticity

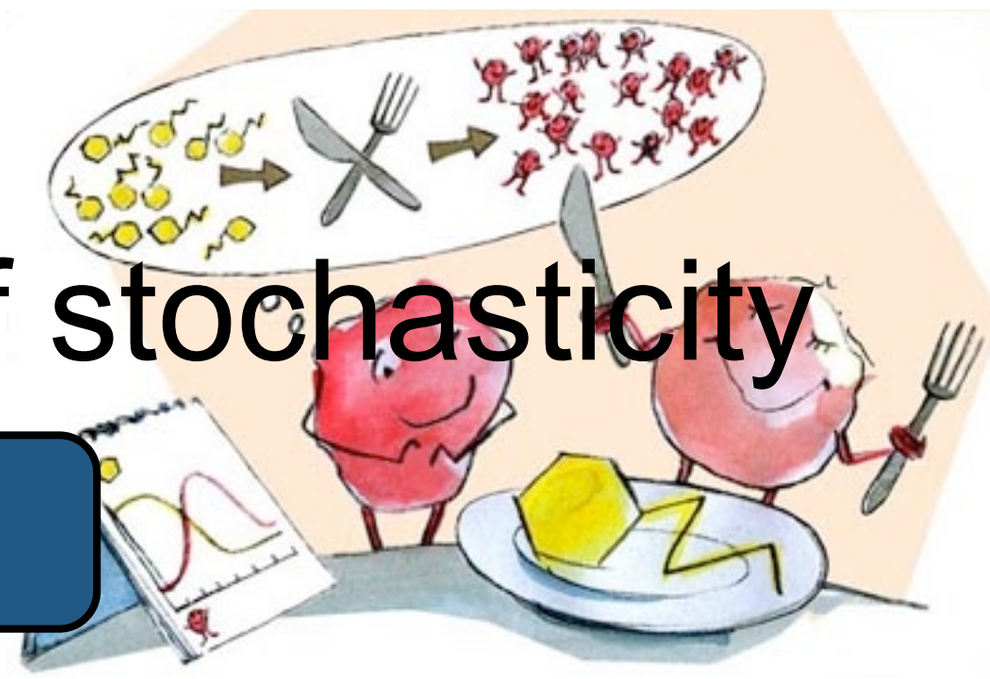


1. Stochastic Decision

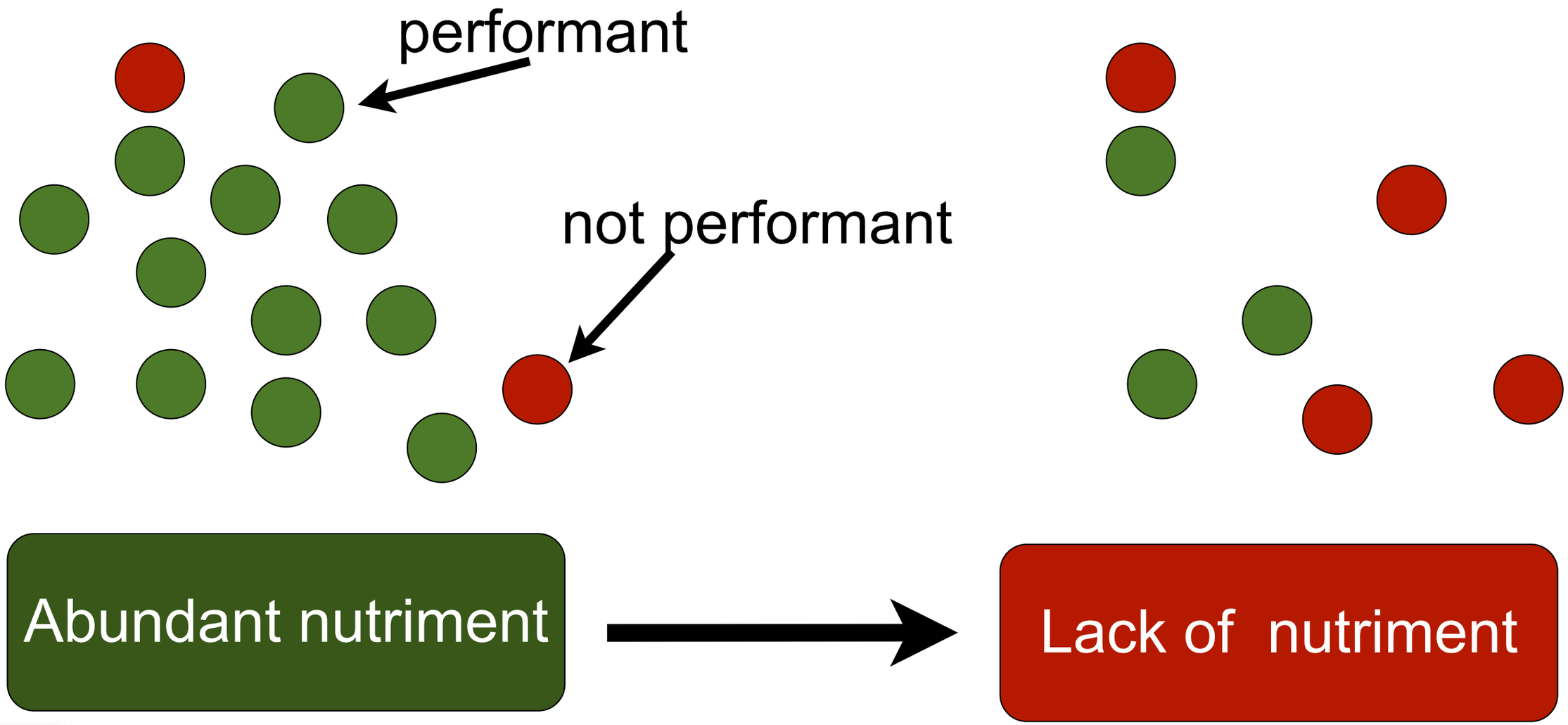


Abundant nutriment

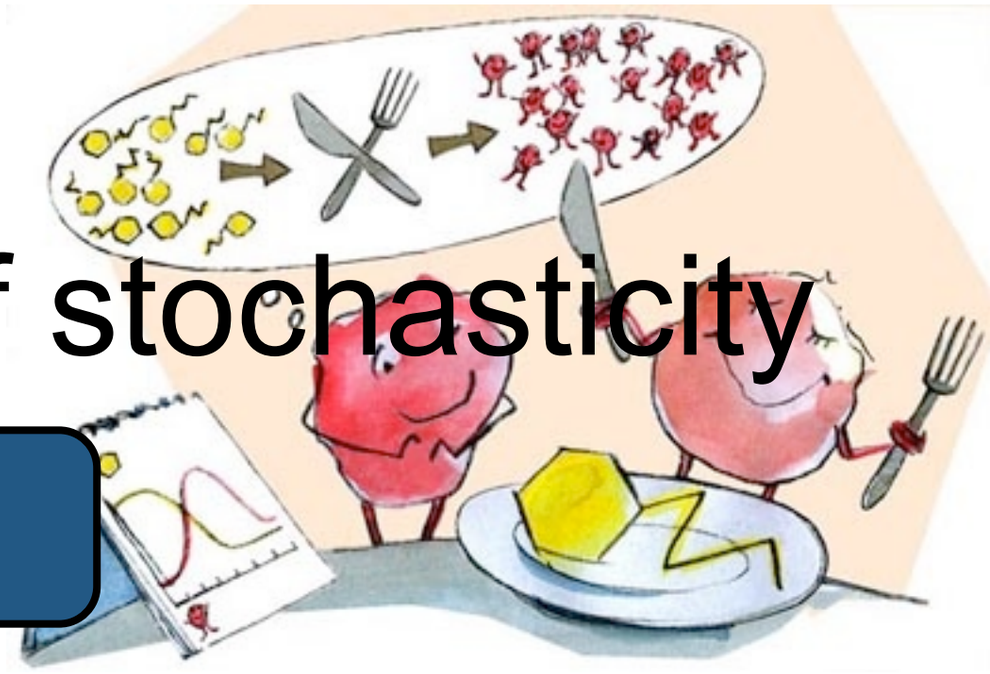
Bacteria -- Two levels of stochasticity



1. Stochastic Decision



Bacteria -- Two levels of stochasticity



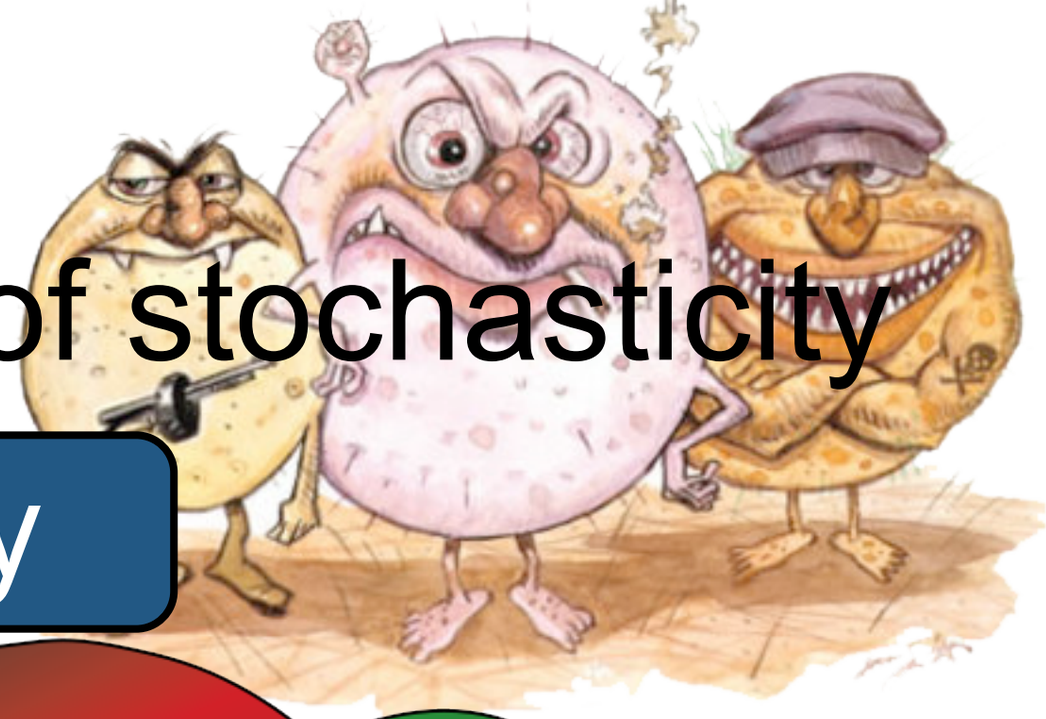
1. Stochastic Decision

Pros:
- flexibility
- simple

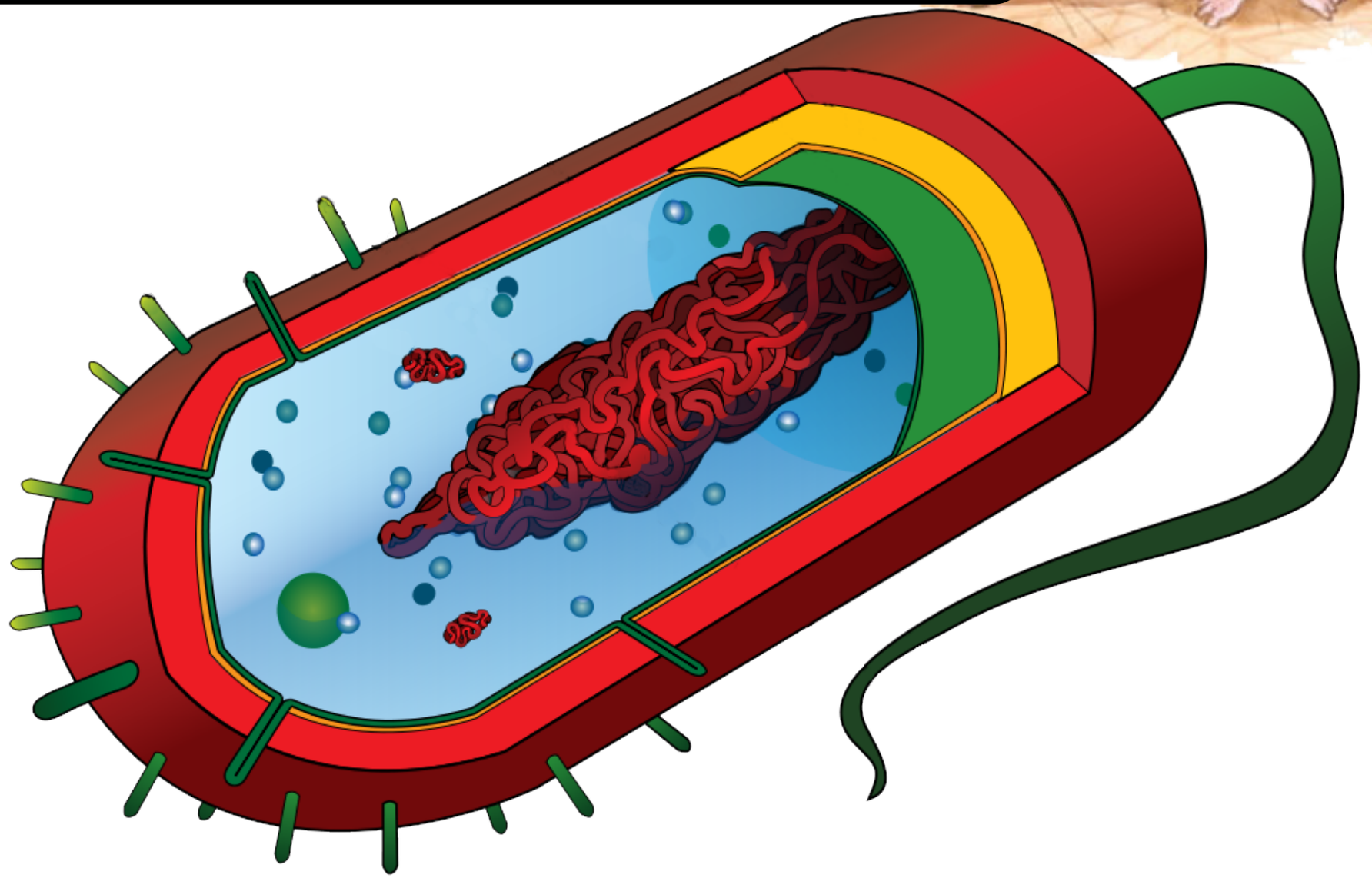
Cons:
- fluctuations

Abundant nutriment → Lack of nutriment

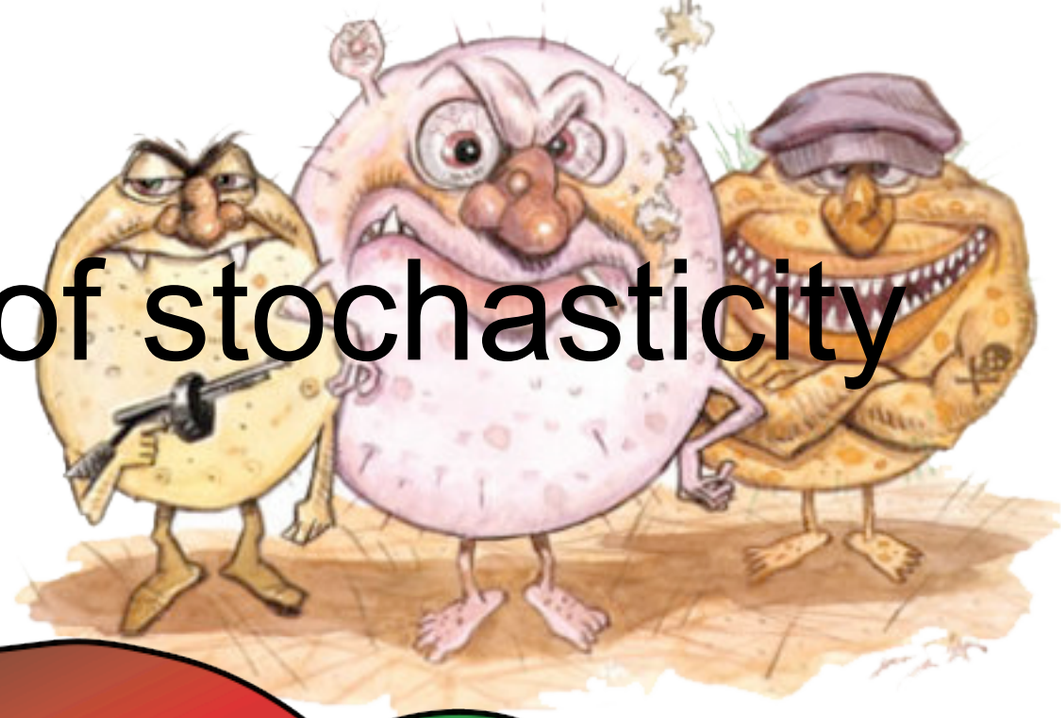
Bacteria -- Two levels of stochasticity

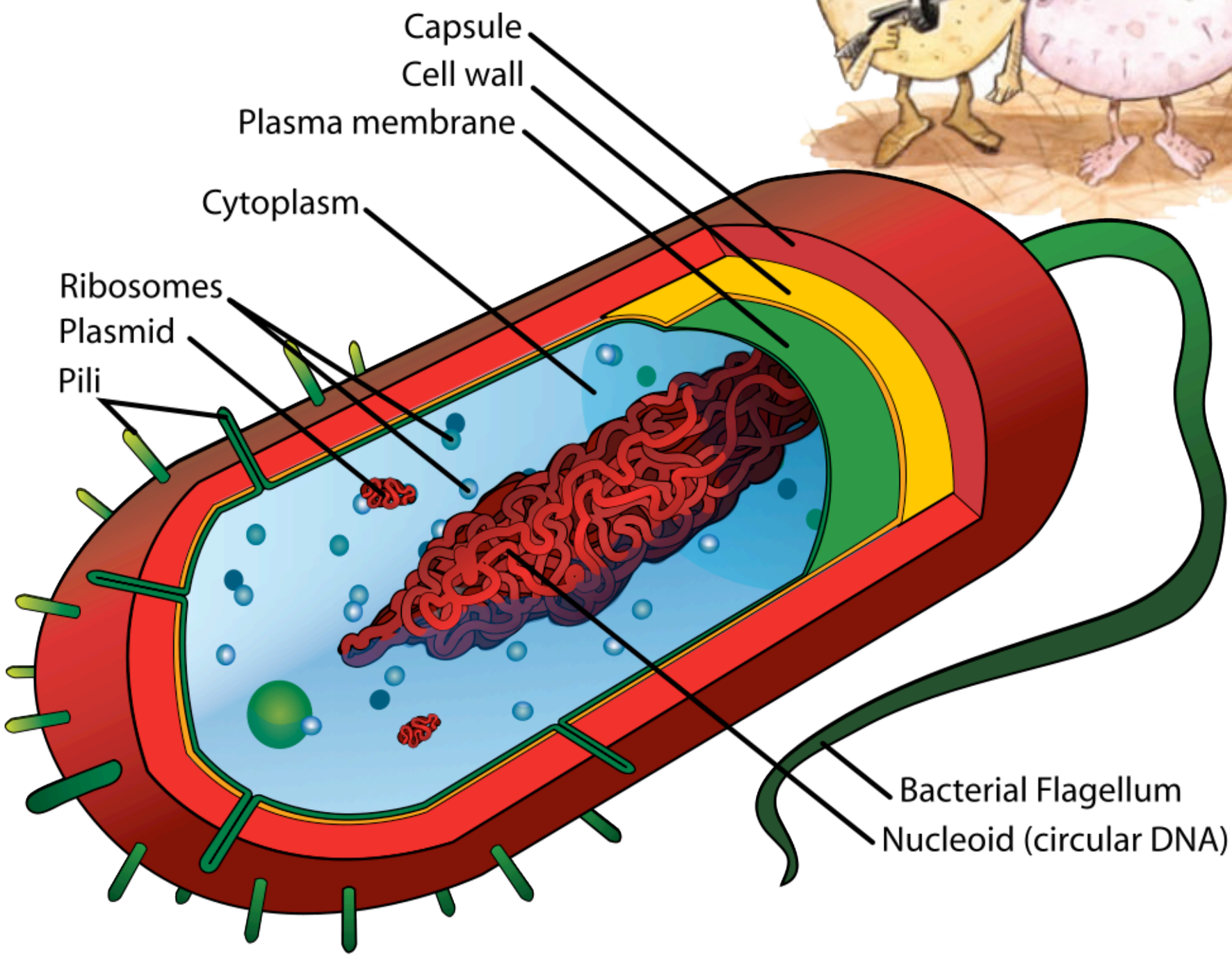


2.Structural stochasticity

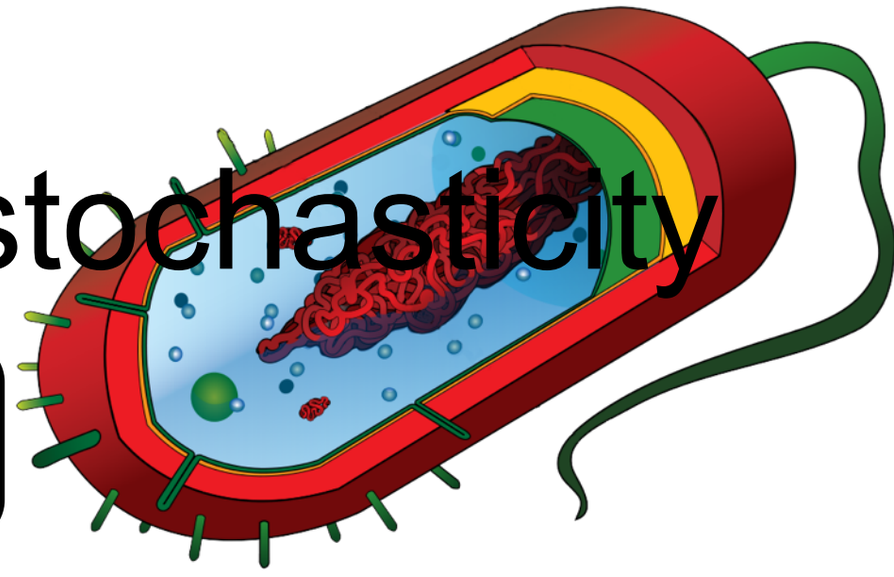


Bacteria -- Two levels of stochasticity



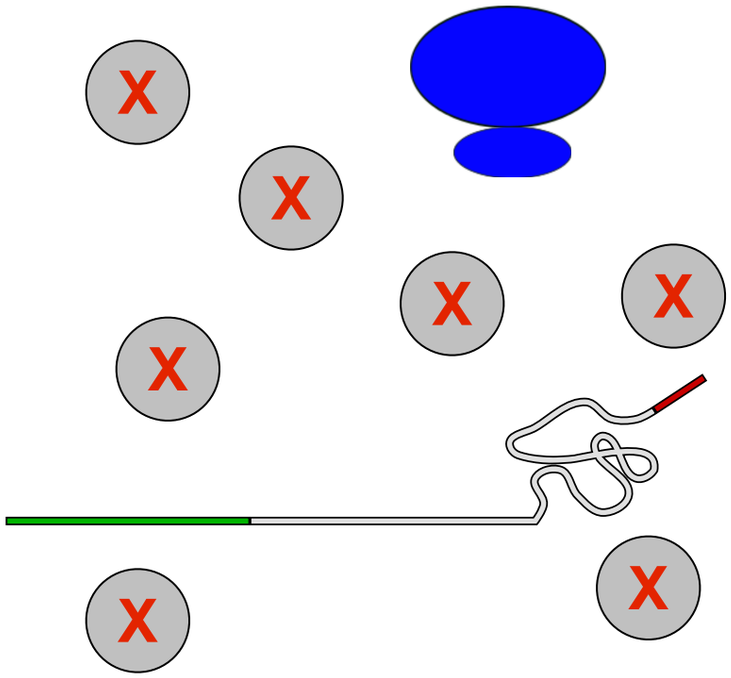


Bacteria -- Two levels of stochasticity

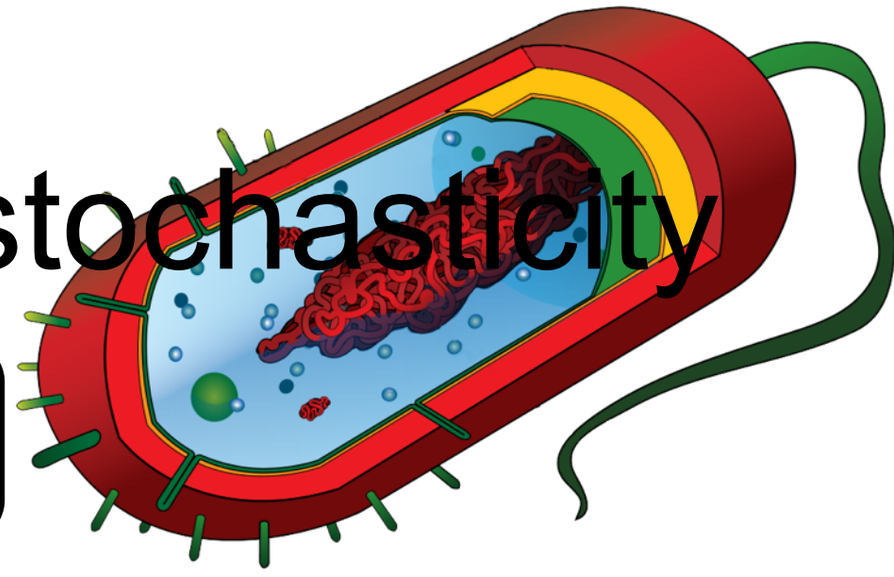


2.Structural stochasticity

- No spatial organization
- Reactions: stochastic encounters

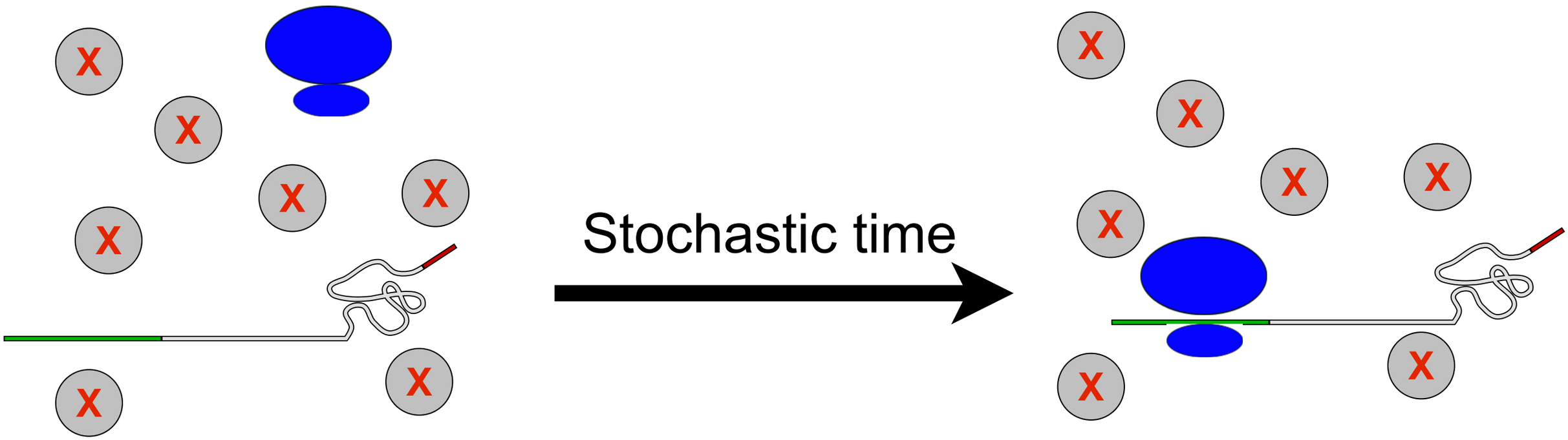


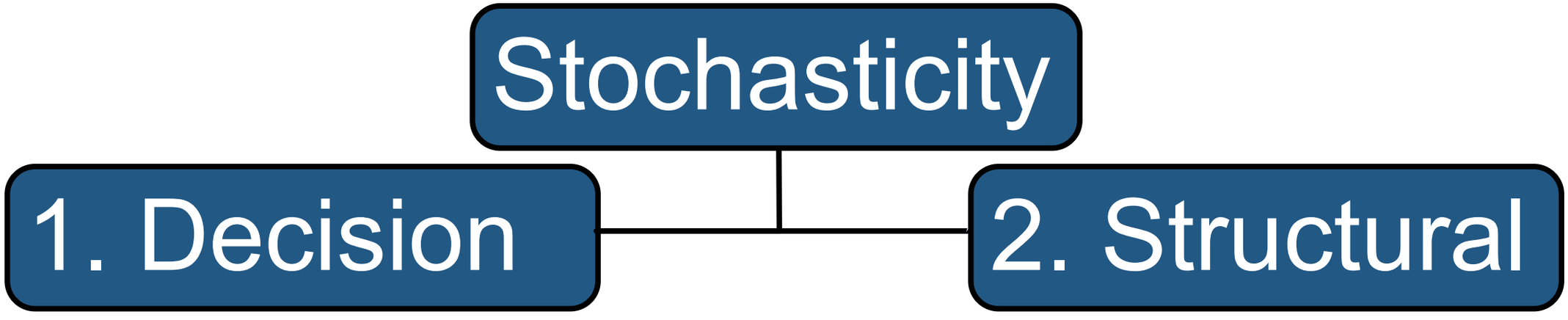
Bacteria -- Two levels of stochasticity



2.Structural stochasticity

- No spatial organization
- Reactions: stochastic encounters



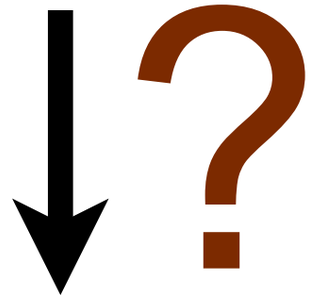


In order to assure:

- flexibility
- performance

in stochastic environment

Cell structure design



Fluctuations

Deterministic behaviour robust to fluctuations

Gene Expression

Gene Expression:

Gene Expression:

- Proteins: the core of biologic processes (enzymes, DNA duplication, cell machinery...)

Gene Expression:

- Proteins: the core of biologic processes (enzymes, DNA duplication, cell machinery...)
- Lack of one protein can have serious consequences

Gene Expression:

- Proteins: the core of biologic processes (enzymes, DNA duplication, cell machinery...)

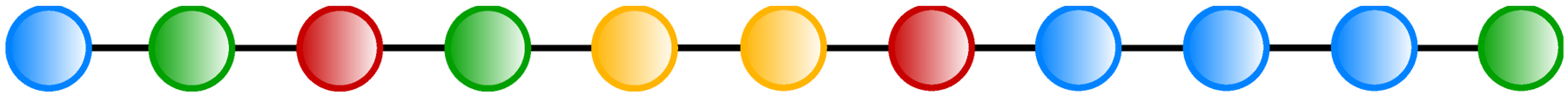
- Lack of one protein can have serious consequences

A highly consuming process:

- >80% of cell resources
- ~3.5 millions of proteins
- ~2000 types of proteins constantly produced

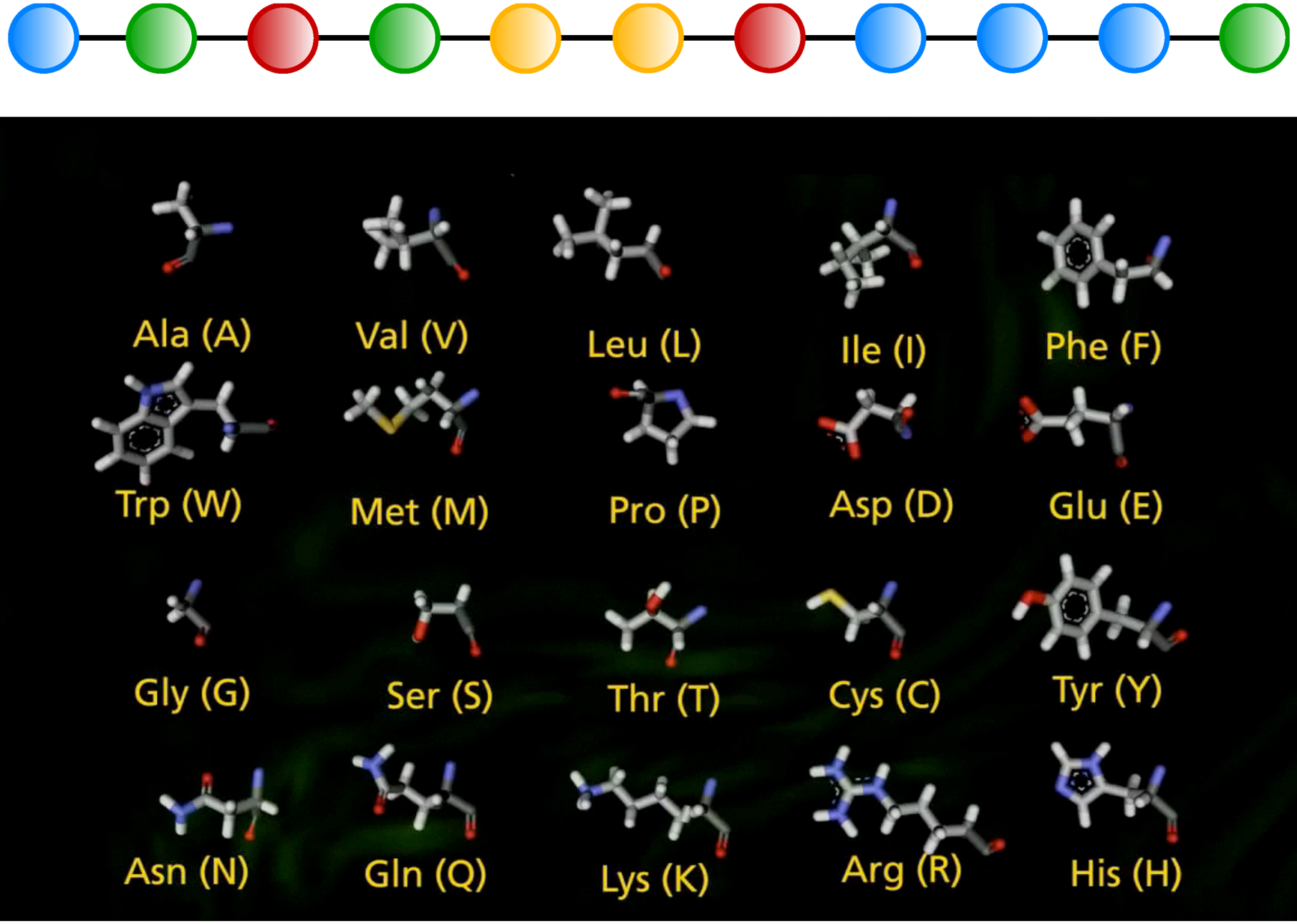
What is a protein?

Protein: chain of elementary bricks (**amino acids**)



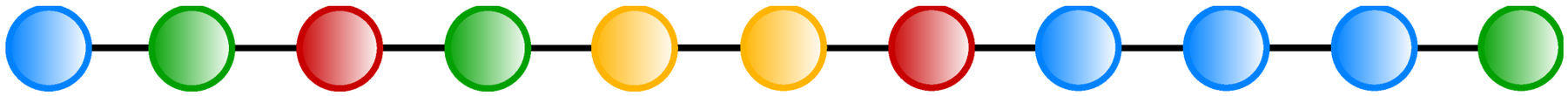
What is a protein?

Protein: chain of elementary bricks (**amino acids**)



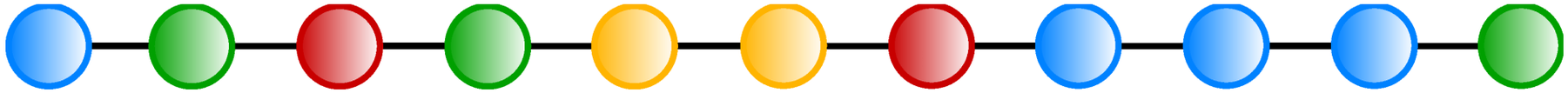
What is a protein?

Protein: chain of elementary bricks (**amino acids**)

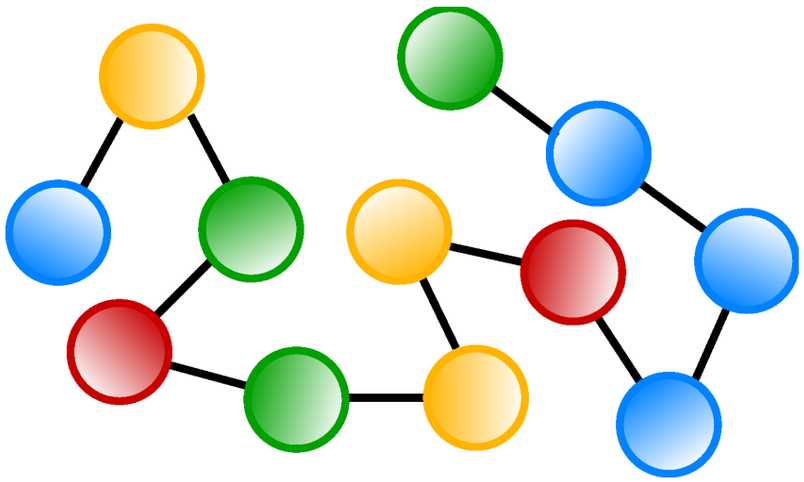


What is a protein?

Protein: chain of elementary bricks (**amino acids**)

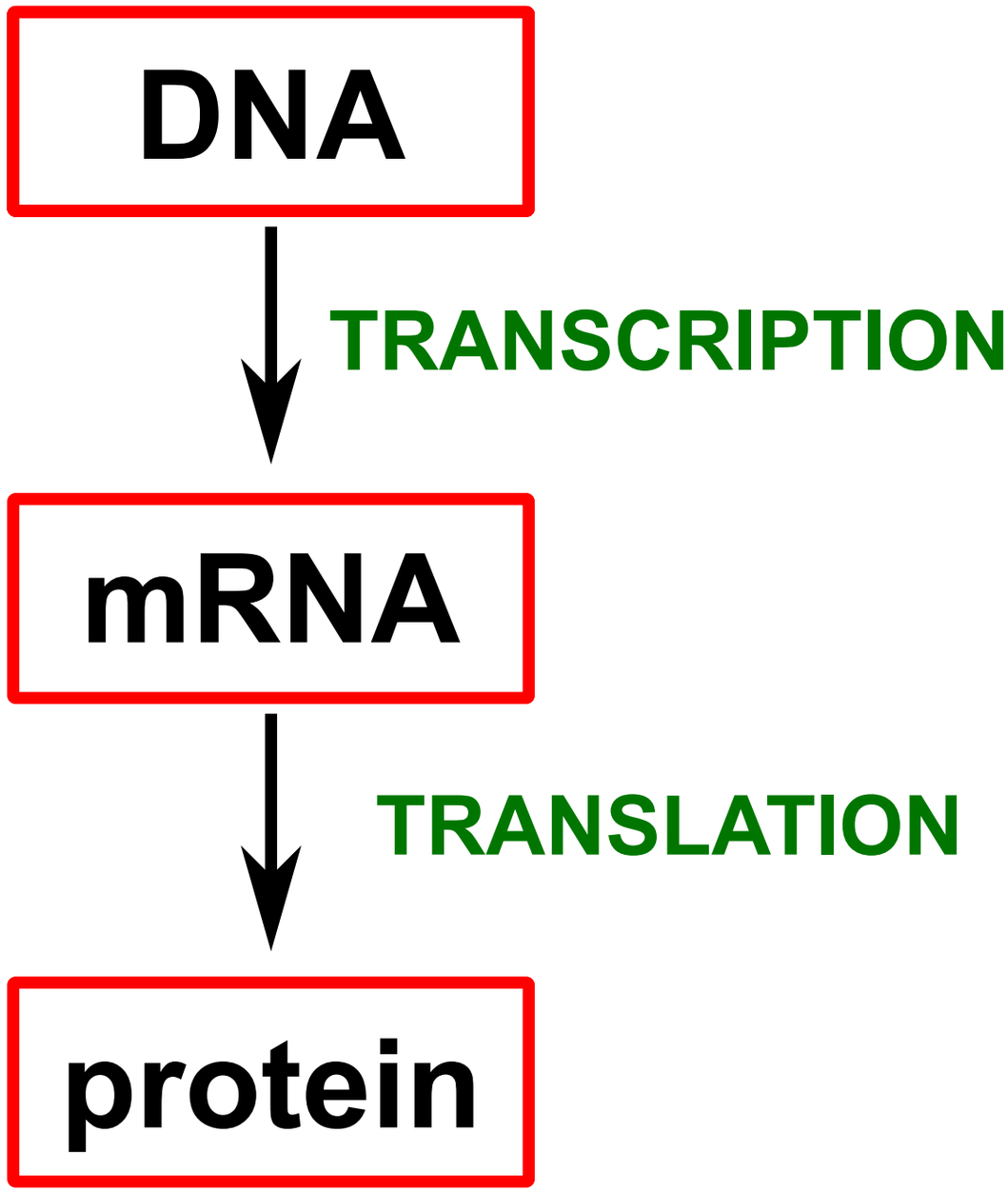


3D conformation

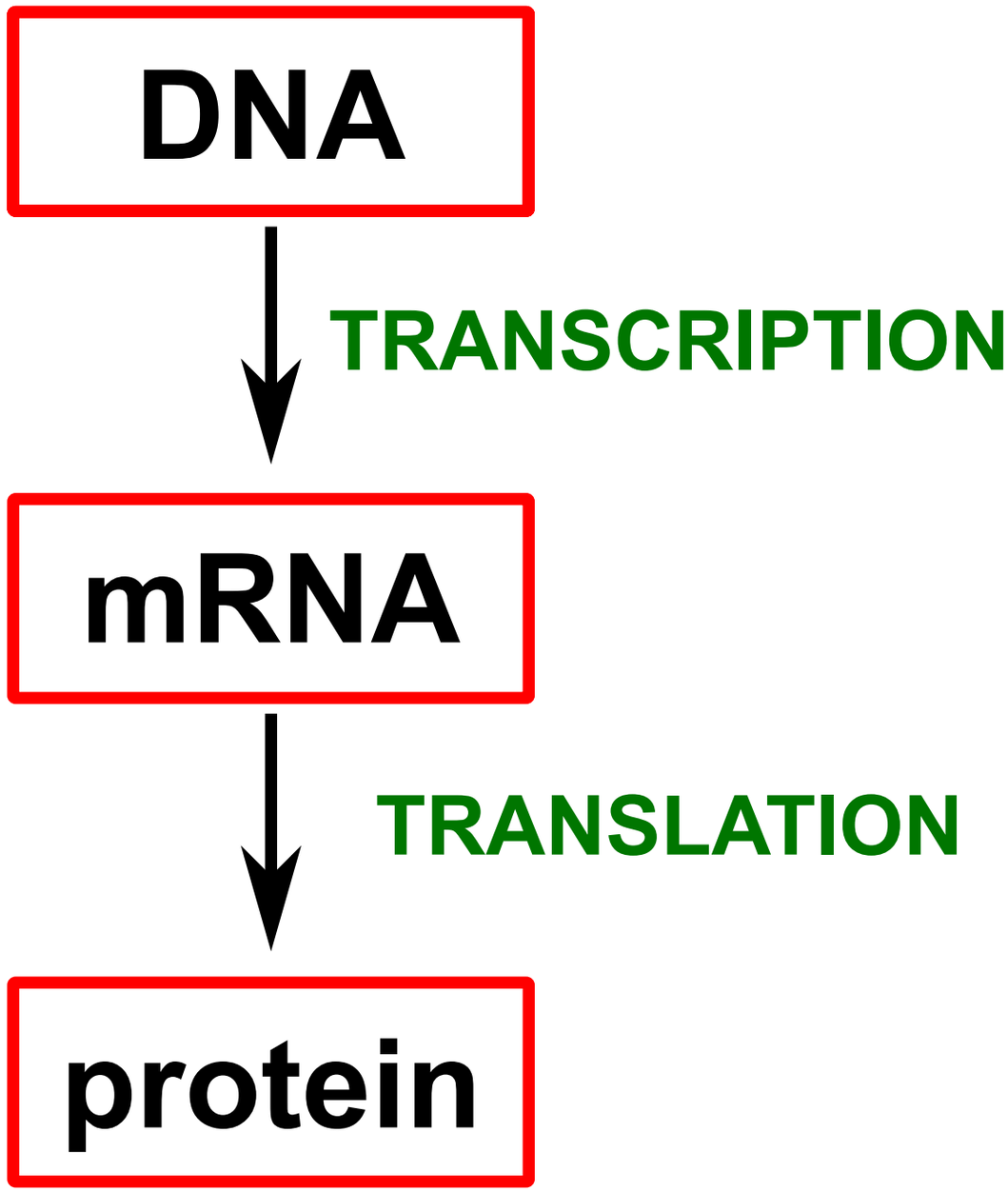


both determining the protein function

Central Dogma of Molecular Biology



Central Dogma of Molecular Biology



"It states that such [sequential] information cannot be transferred from protein to either protein or nucleic acid"

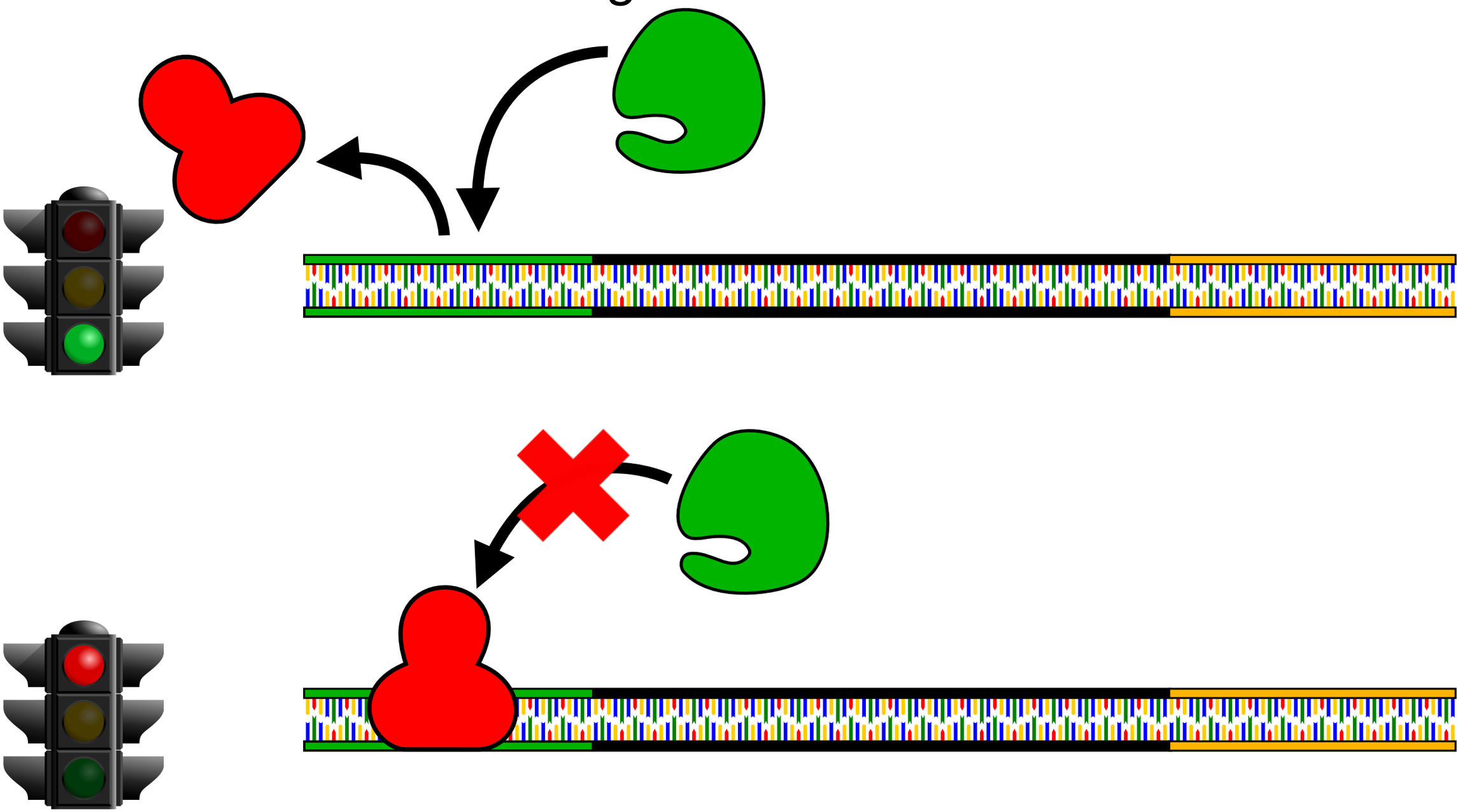
Crick (1958)

DNA

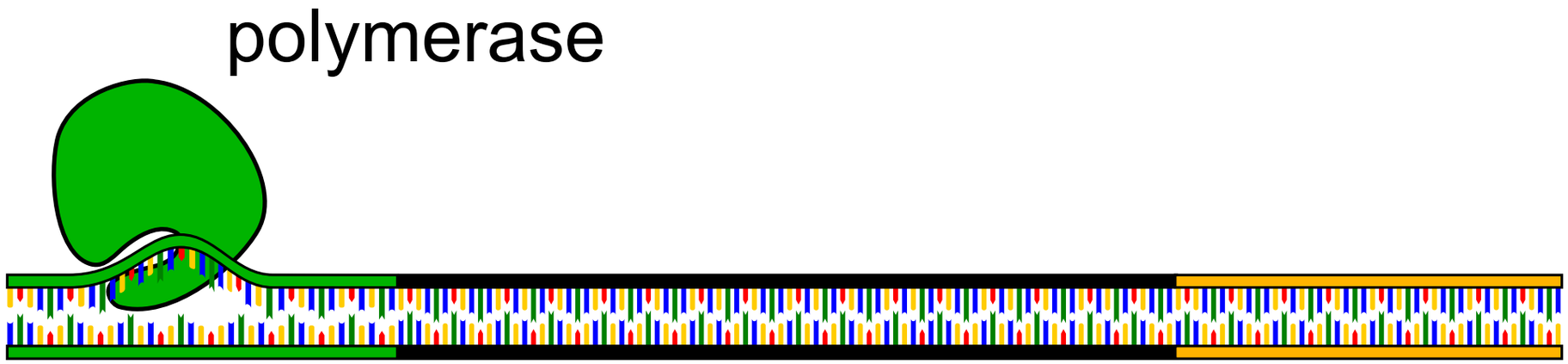
Gene: portion of DNA encoding for a specific protein

Gene activation

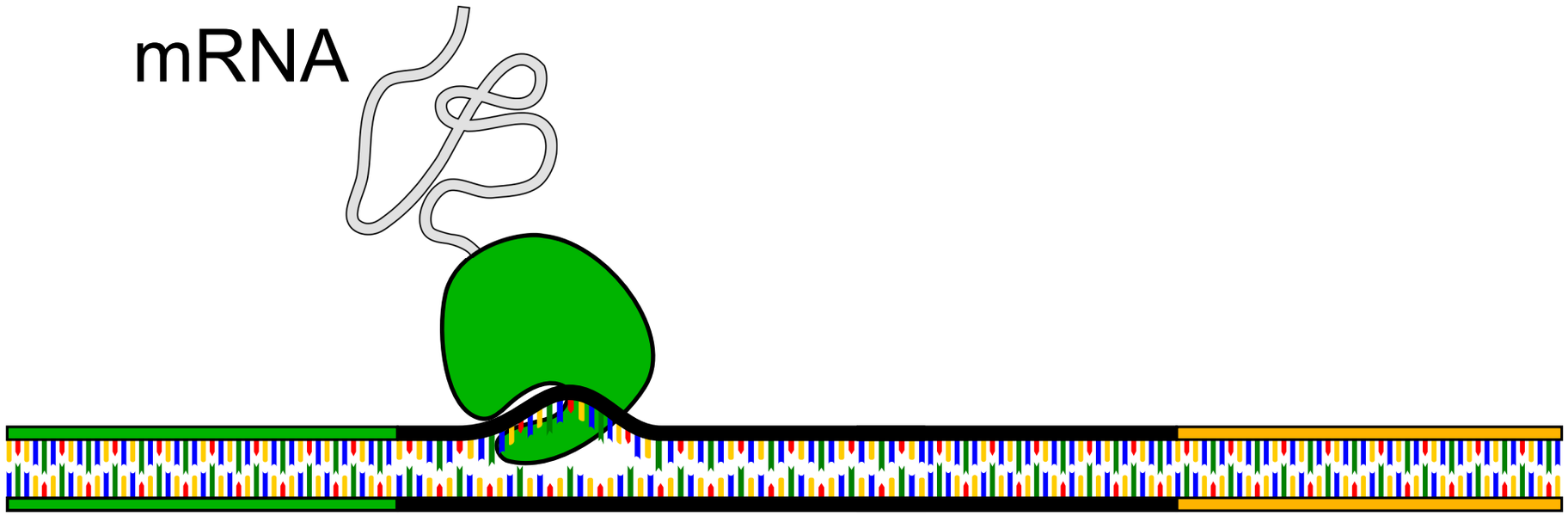
Two states of gene: **active** and **inactive**



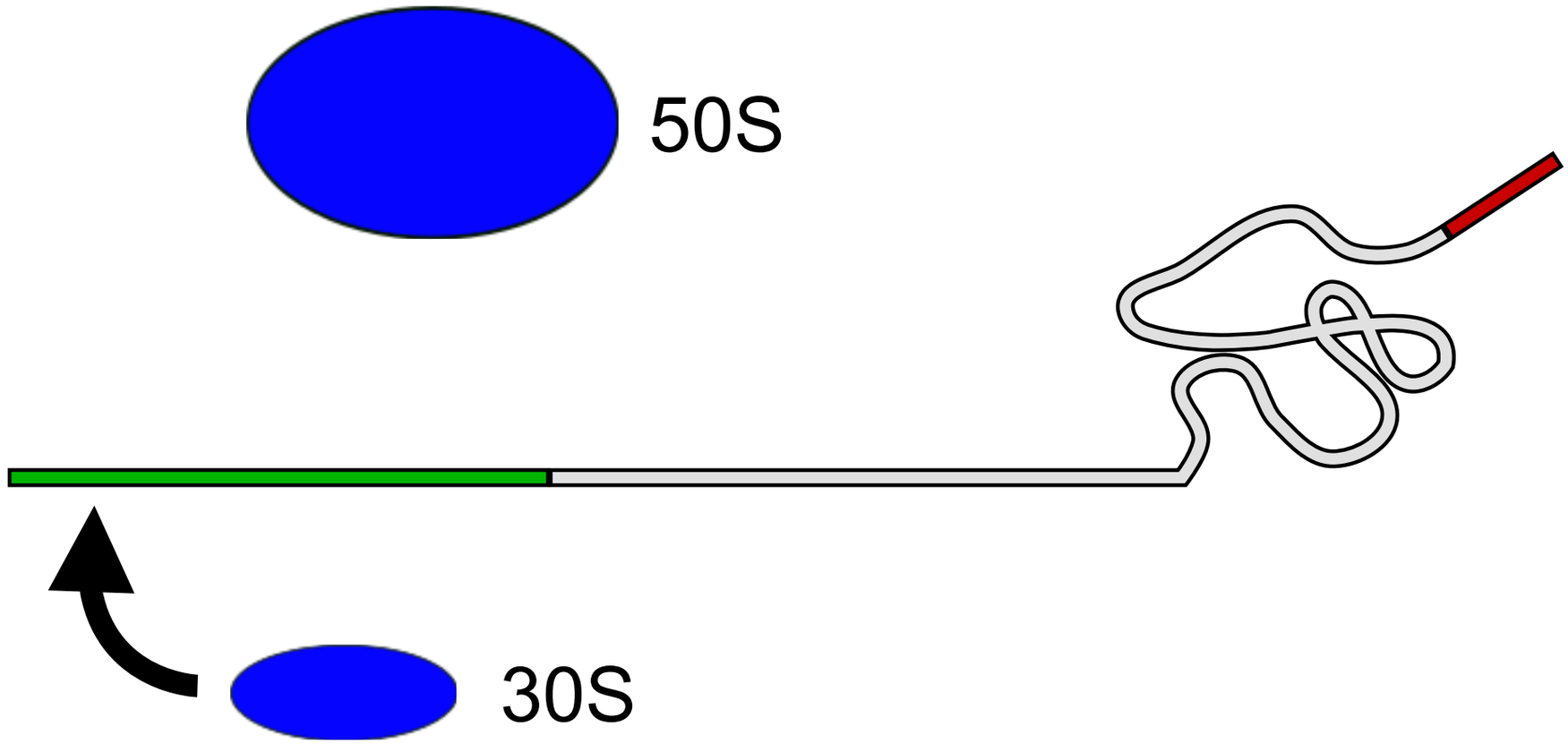
Transcription: initiation



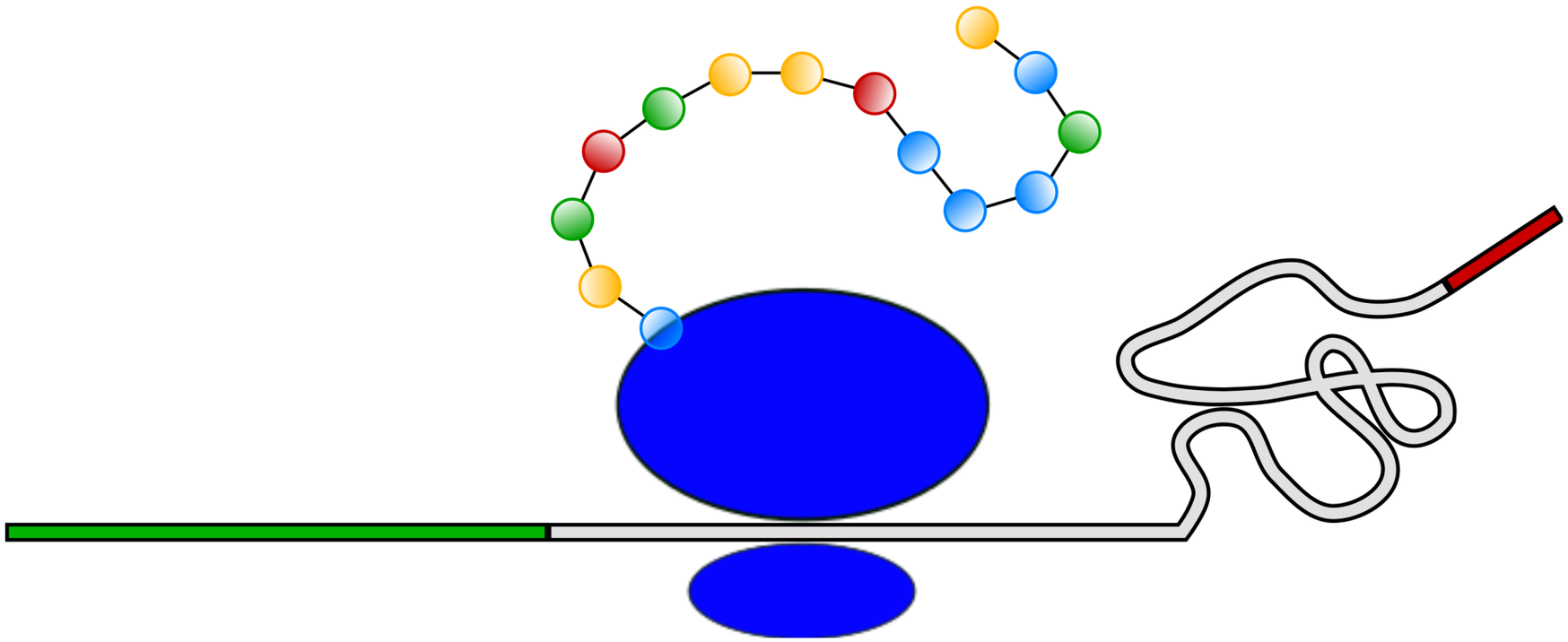
Transcription: elongation



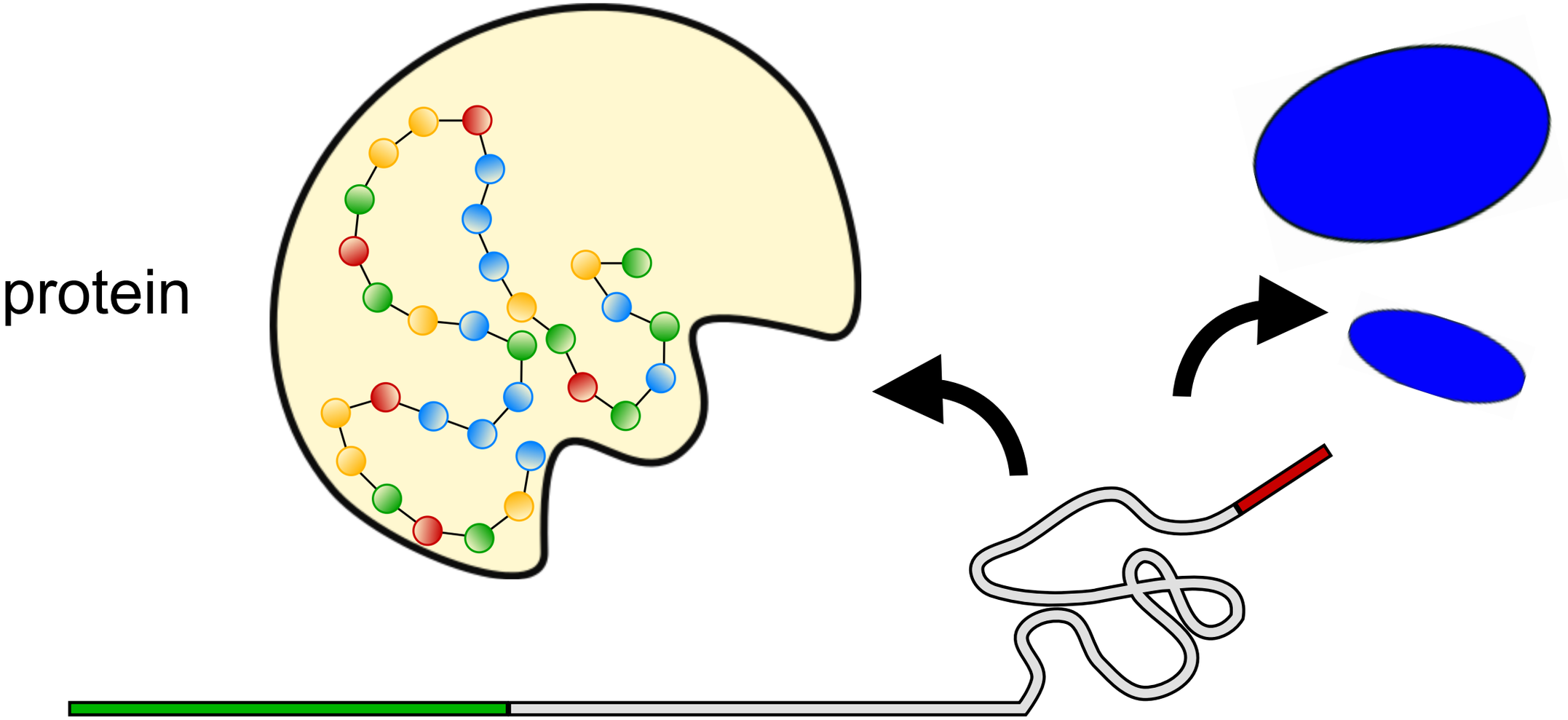
Translation: initiation



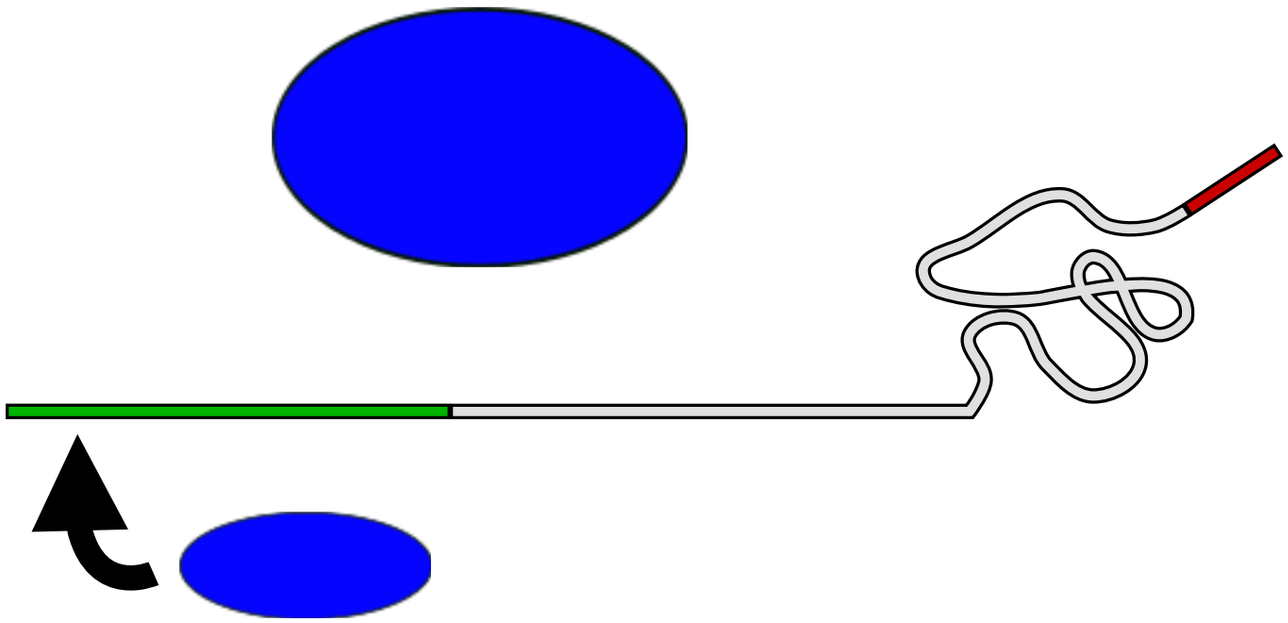
Translation: elongation

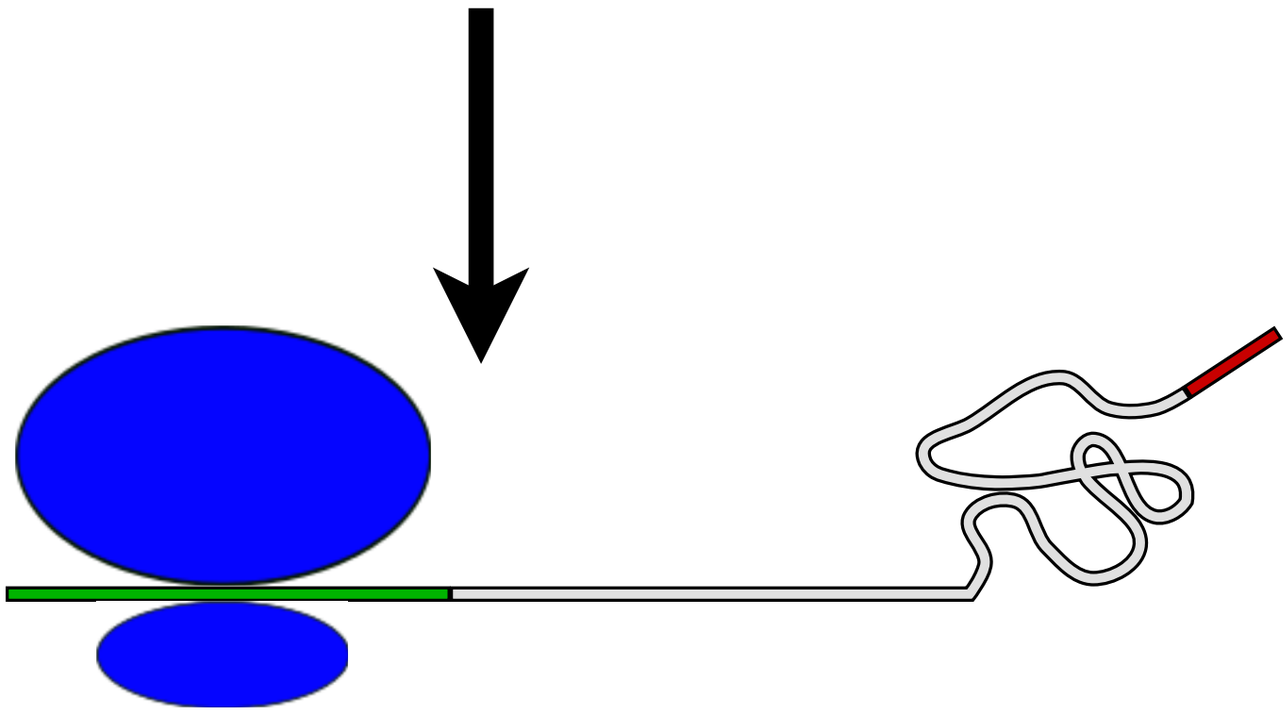


Translation: termination

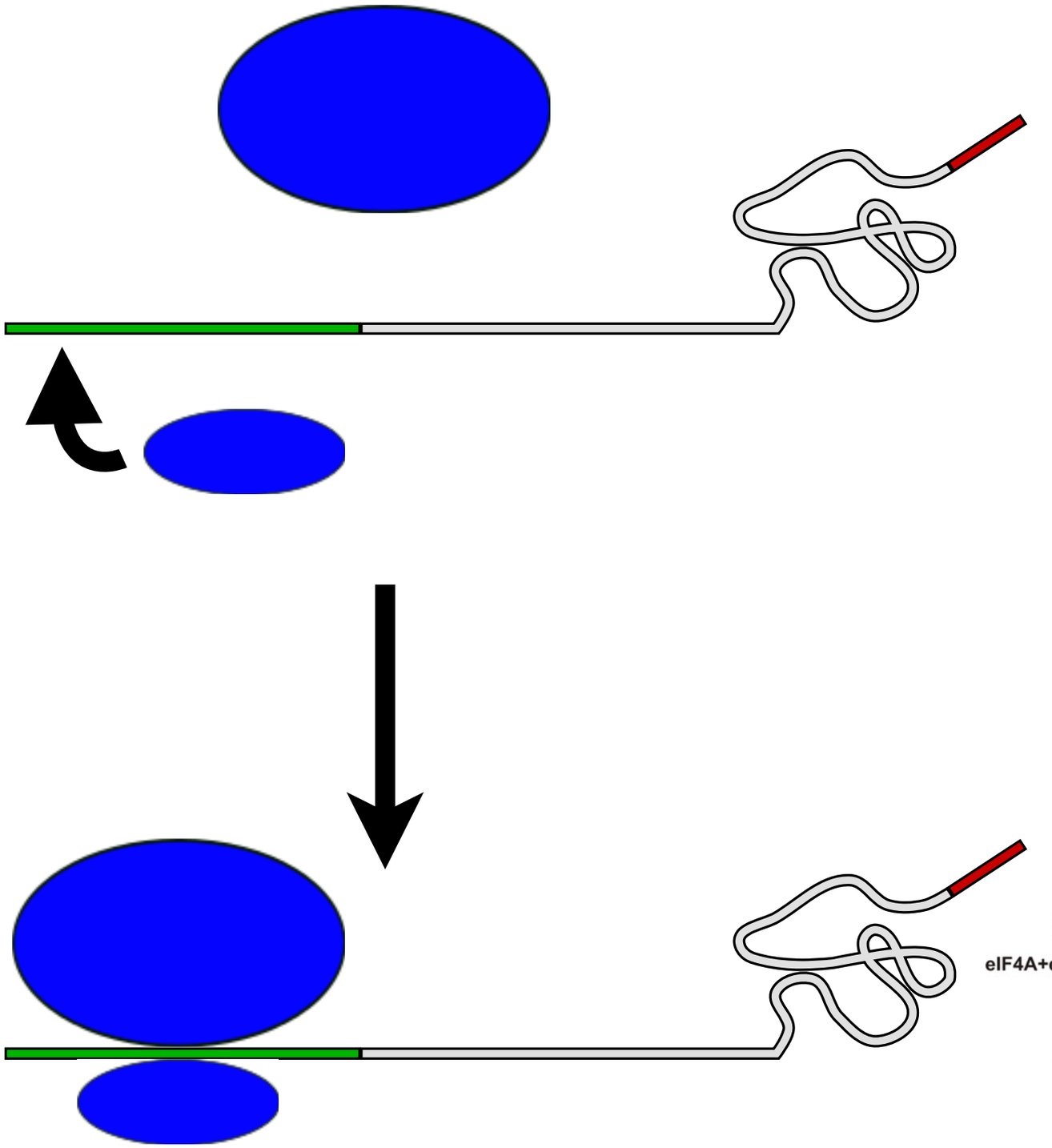


Complex mechanisms

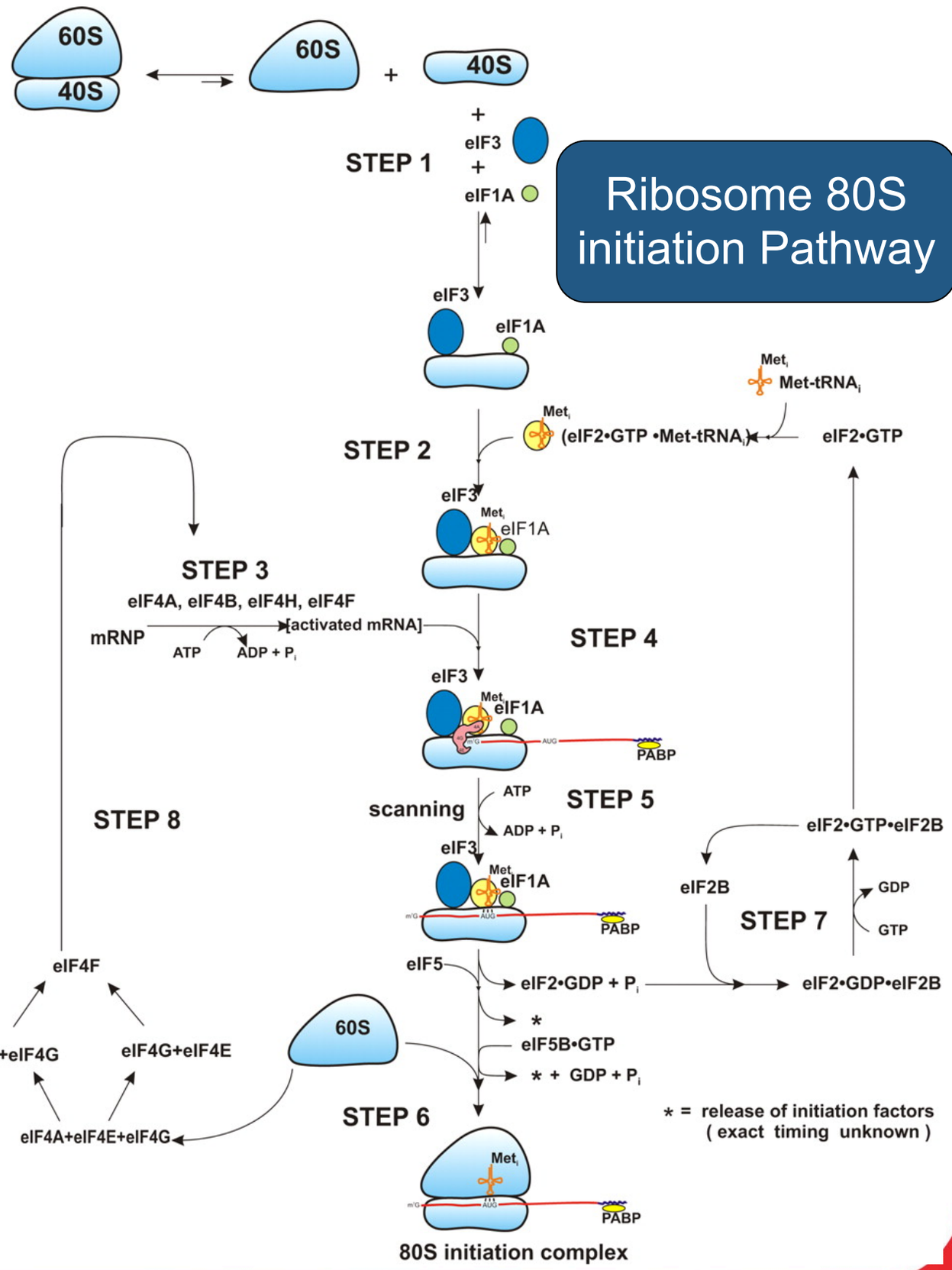




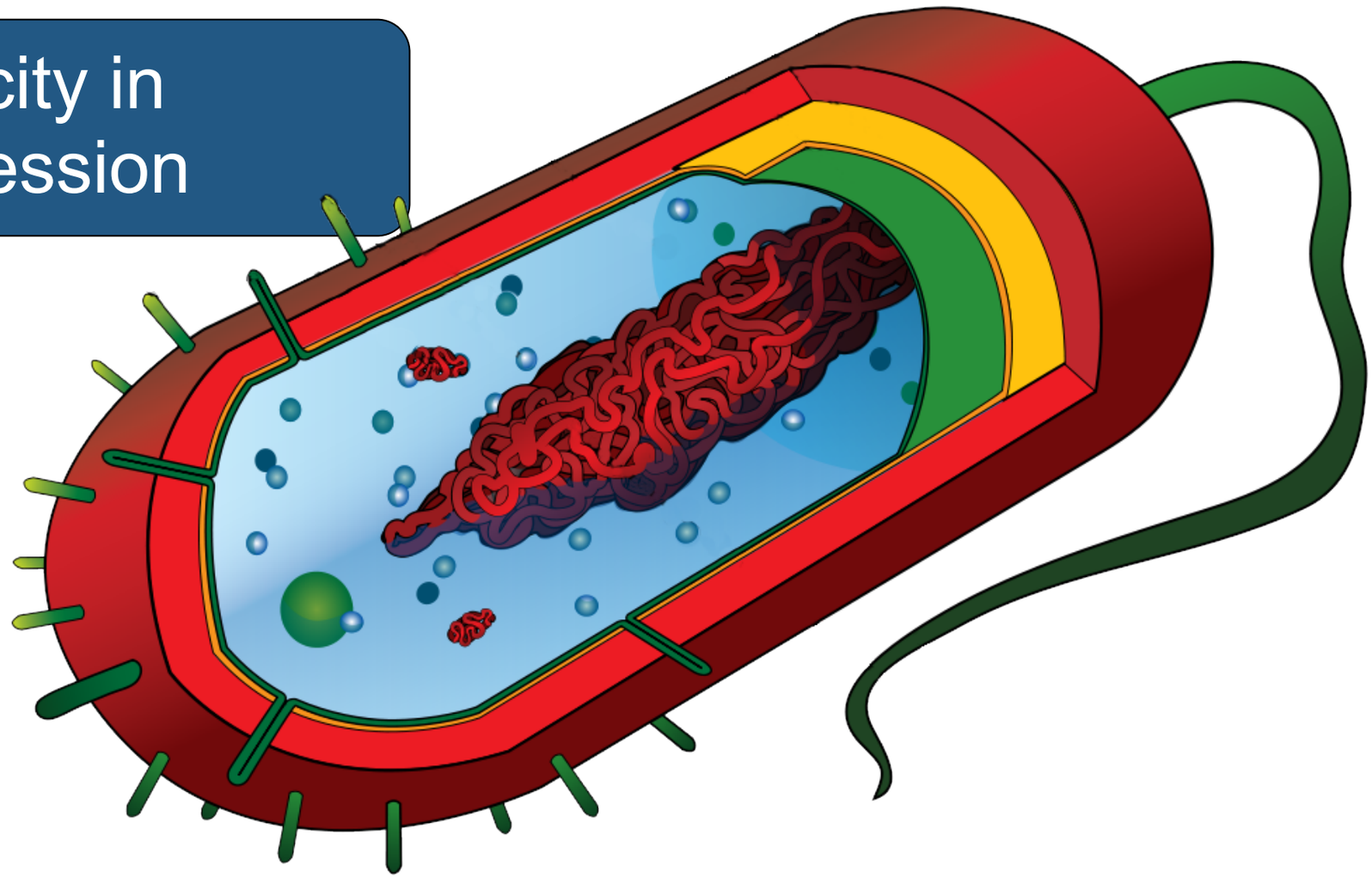
Complex mechanisms

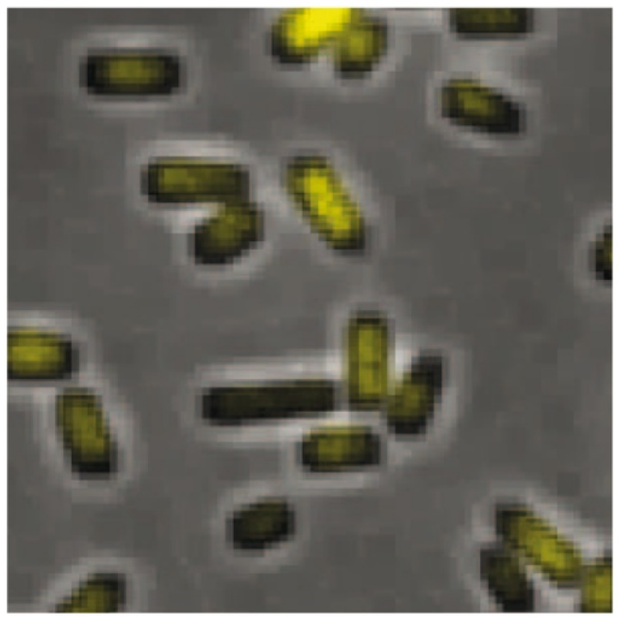


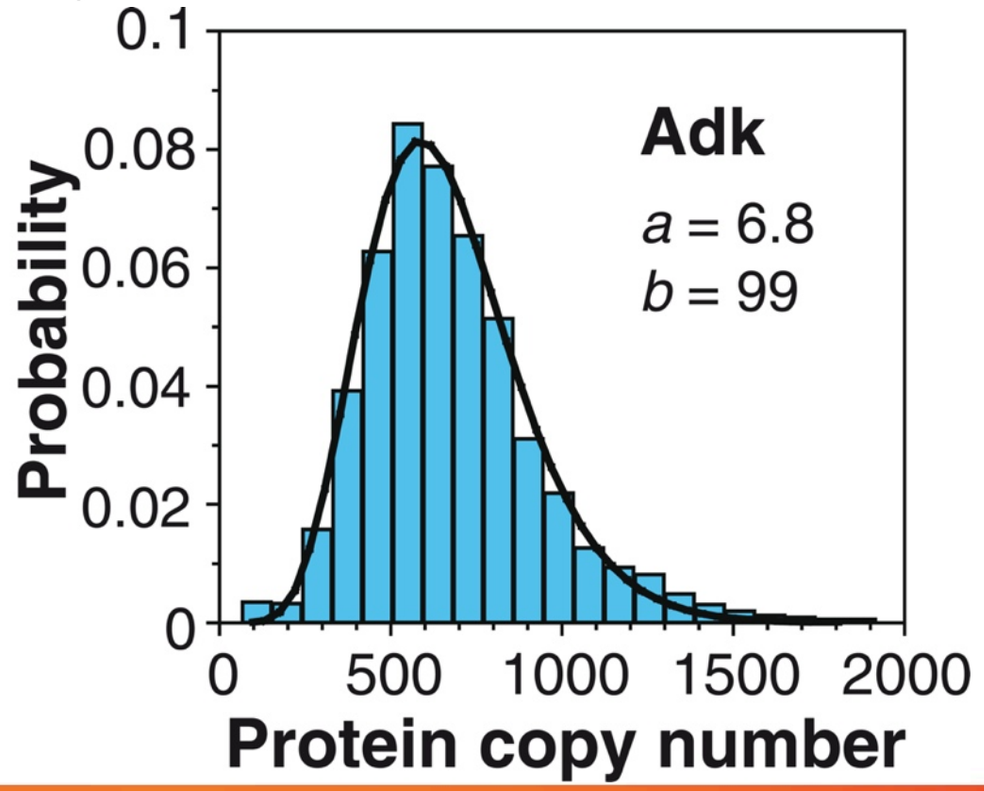
Ribosome 80S initiation Pathway



Stochasticity in gene expression







How to analyze gene expression?

Experiments

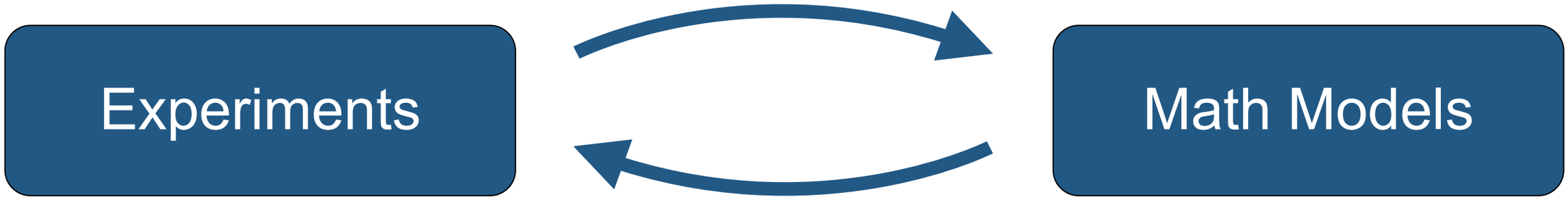
Pros:

- Fine description
- Finding of new phenomena

Cons:

- Expensive
- Hard to reproduce (sometimes not conclusive)

How to analyze gene expression?



Pros:

- Fine description
- Finding of new phenomena

Cons:

- Expensive
- Hard to reproduce (sometimes not conclusive)

Pros:

- Synthesis
- Reproducibility
- Cheap

Cons:

- Time consuming
- Simple but exhaustive ?

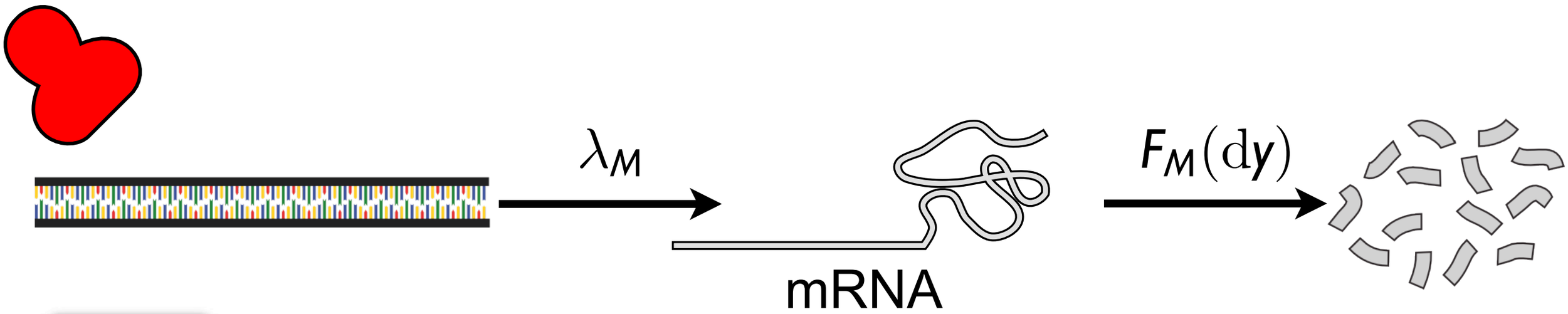
Model

Protein production

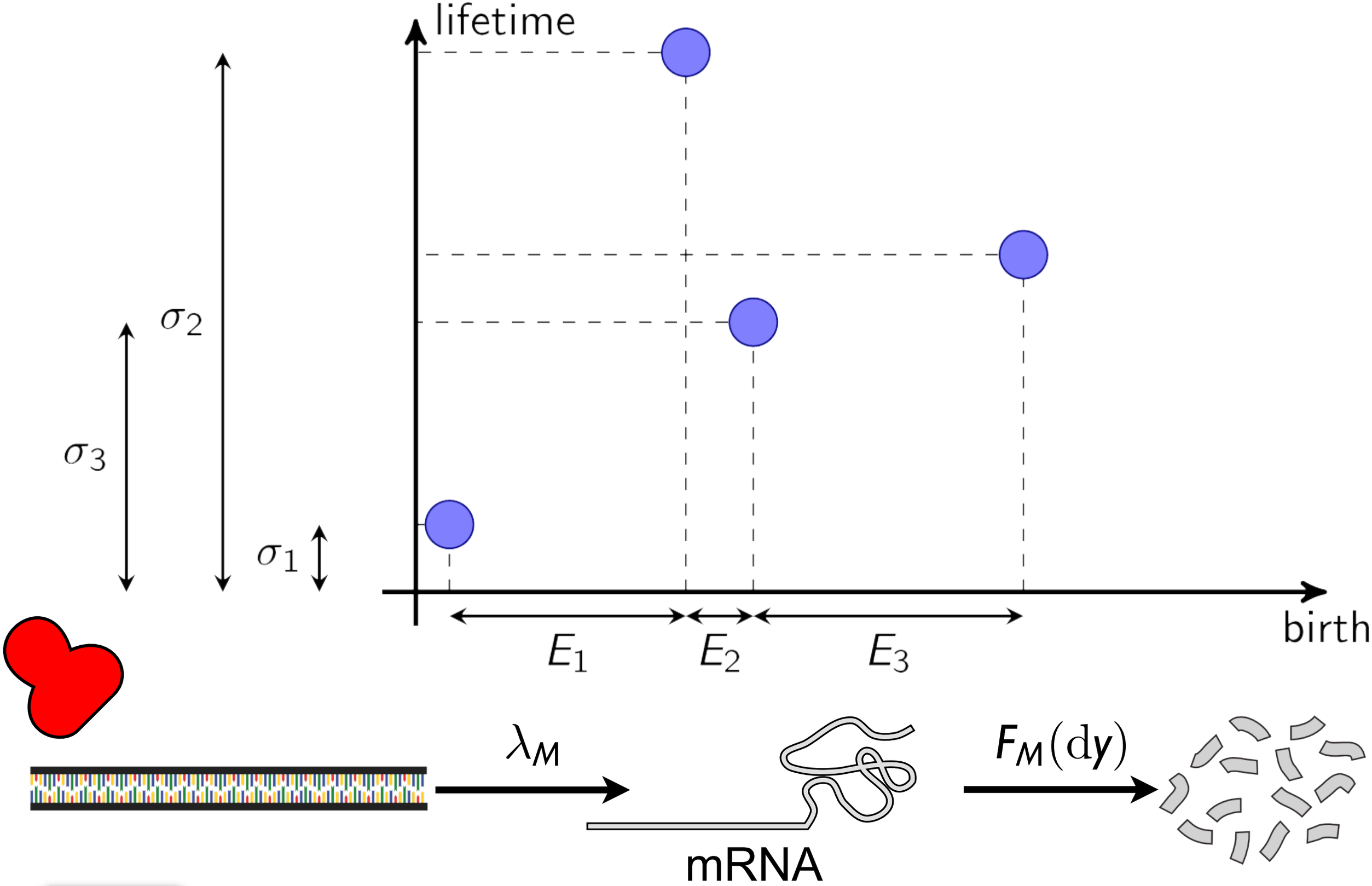
- stochastic
- discrete numbers of components

Protein production

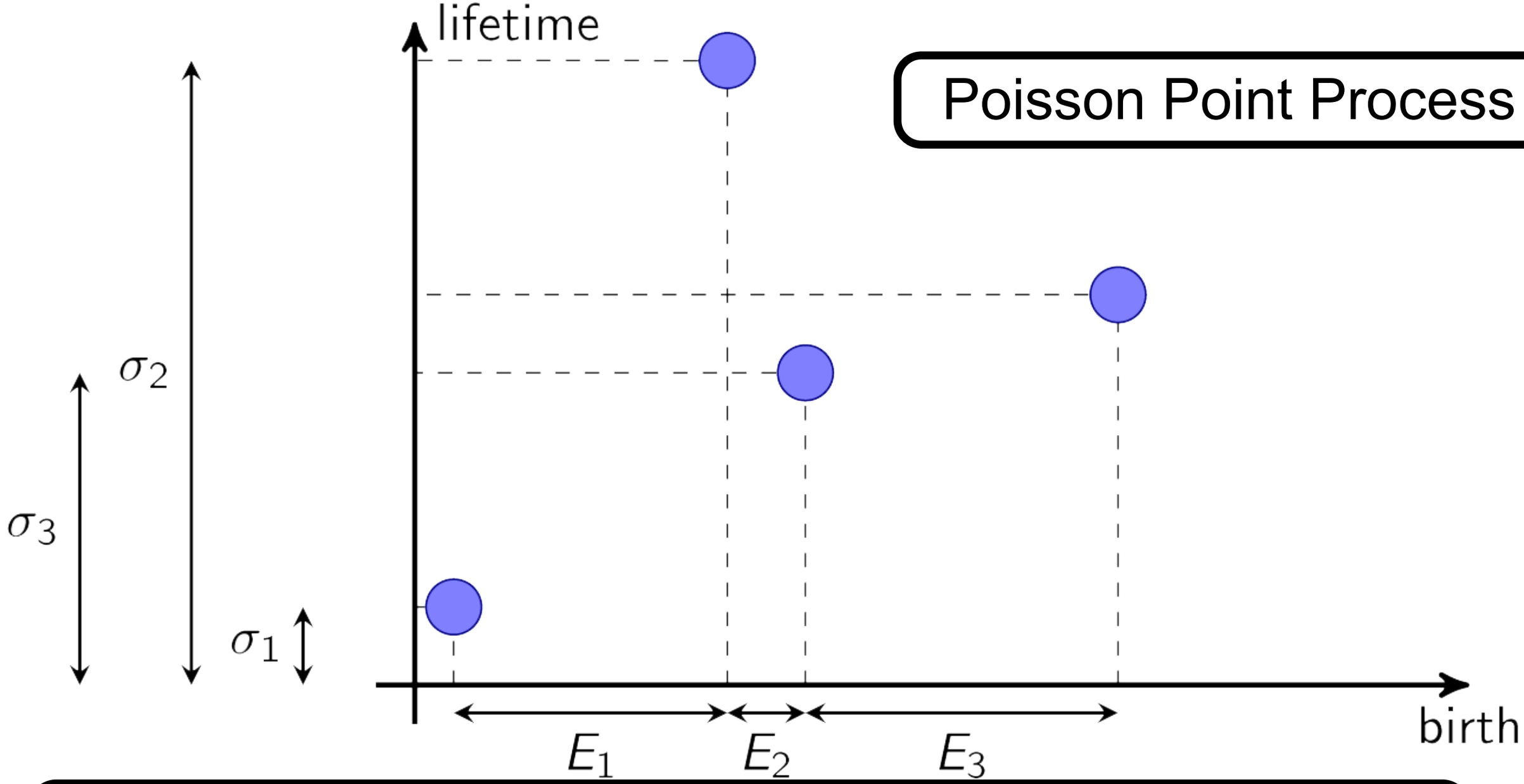
- stochastic
- discrete numbers of components



Protein production



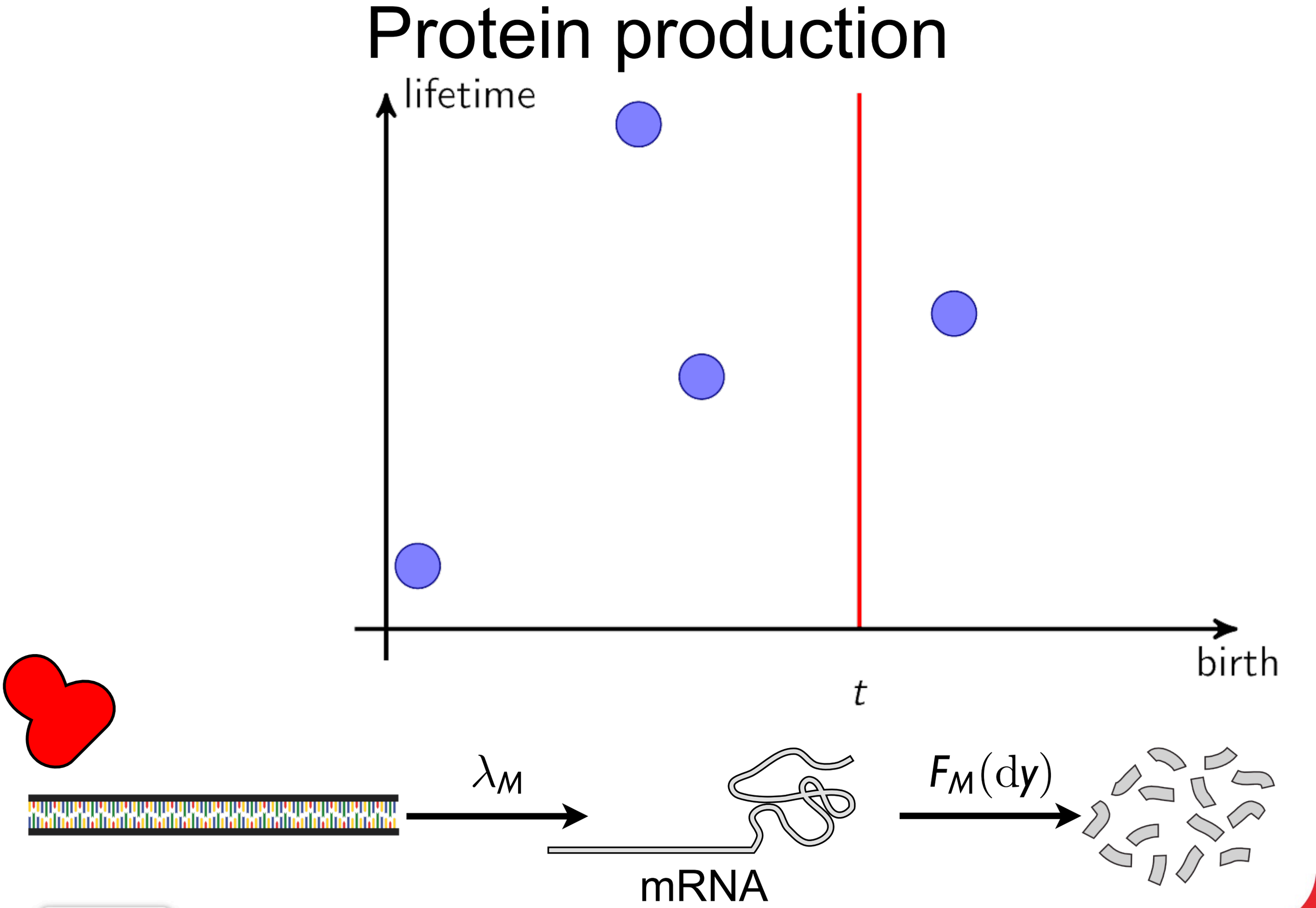
Protein production



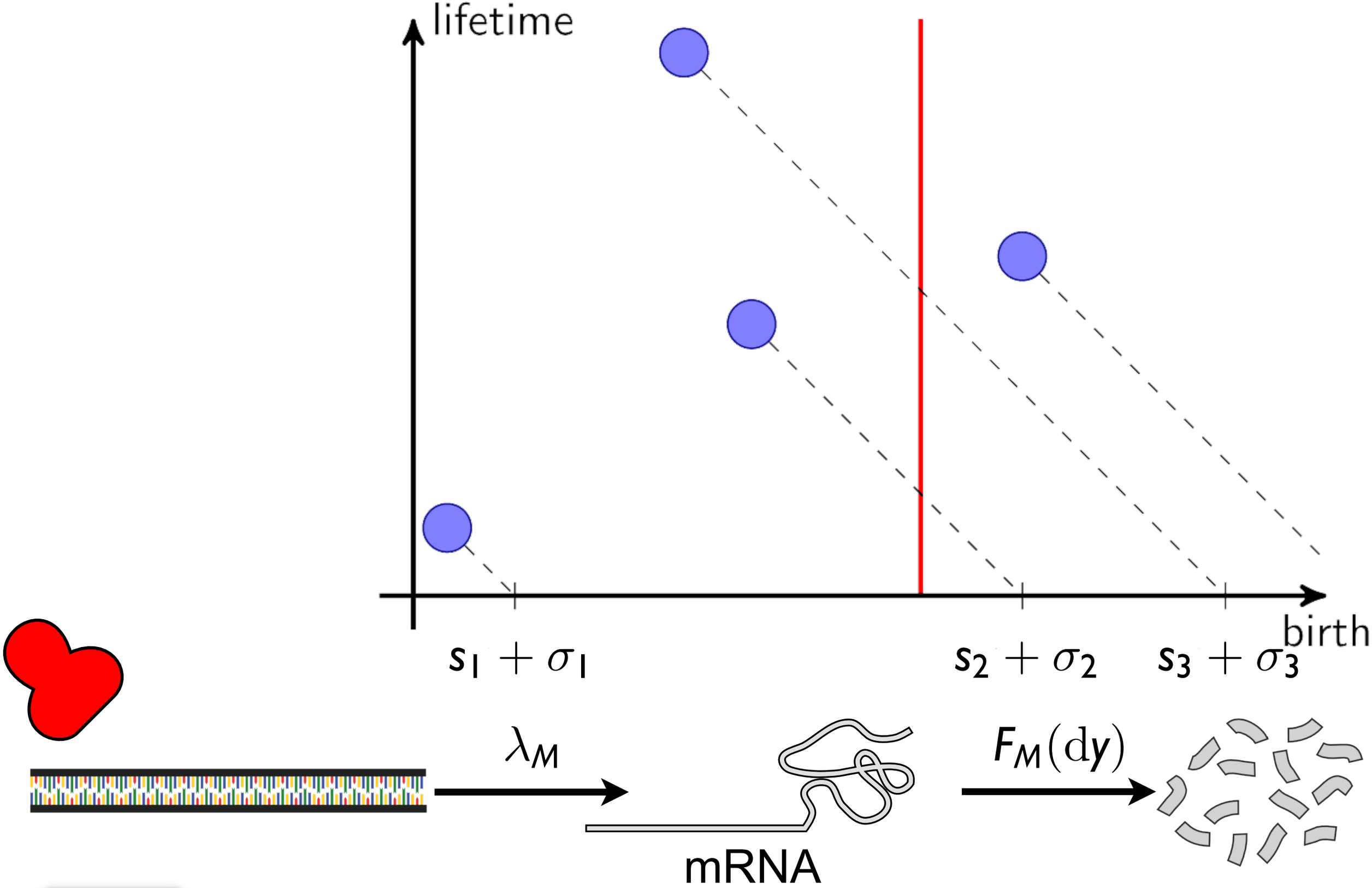
E_i exponentially distributed with parameter λ_M

σ_i general distribution with density $F_M(\mathrm{d}y)$

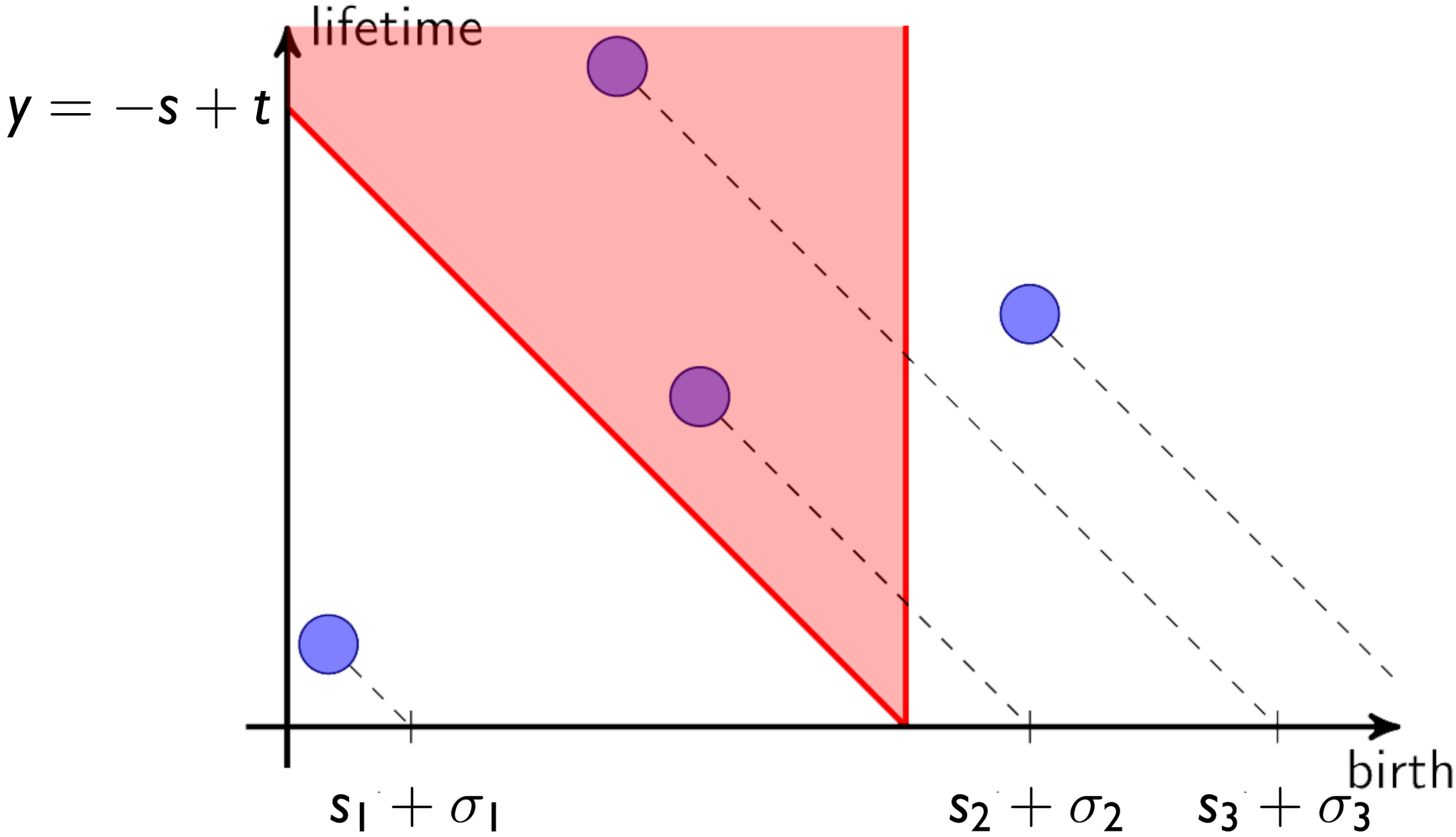
Protein production

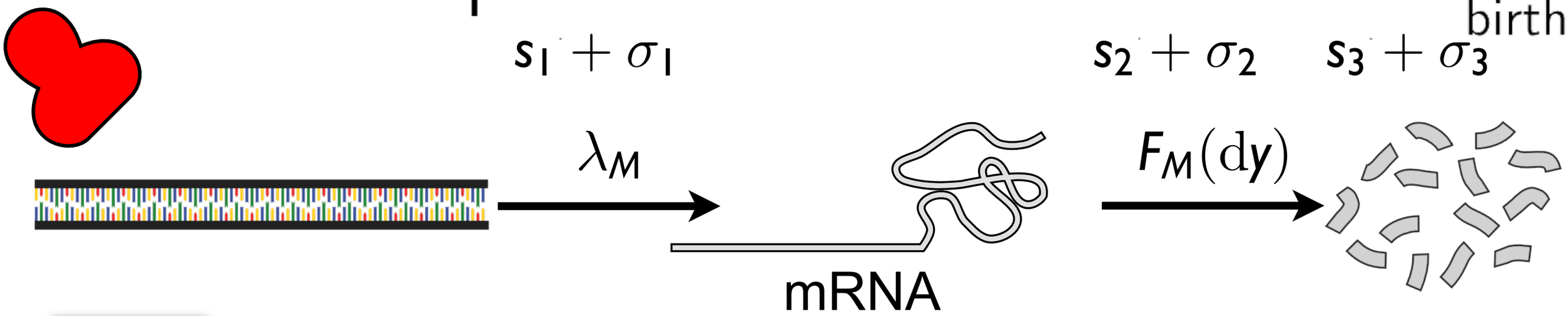


Protein production



Protein production



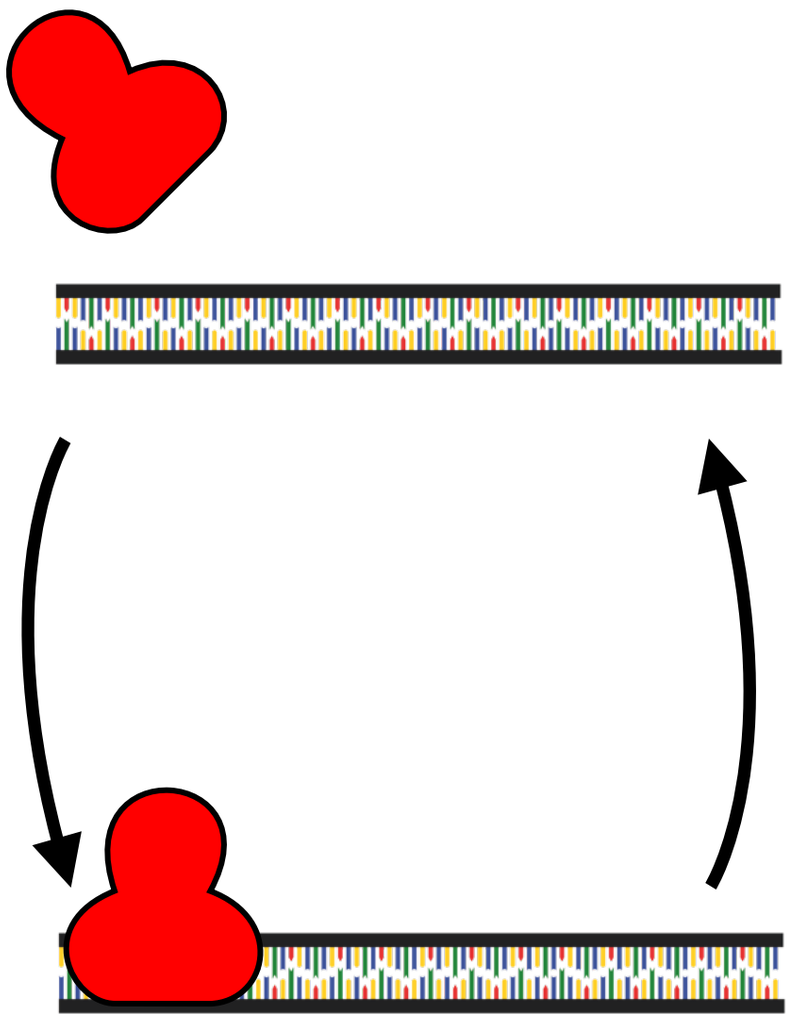


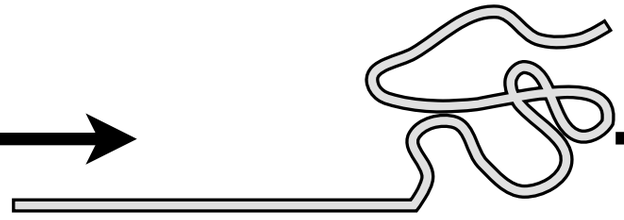
Protein production

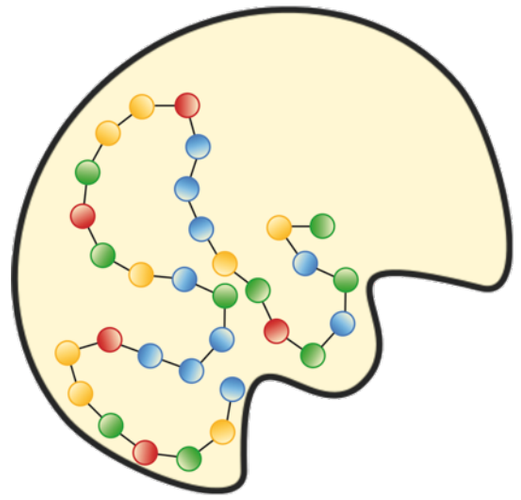
$Y(t) \in \{0, 1\}$ → $M(t) \in \mathbb{N}$ → $P(t) \in \mathbb{N}$

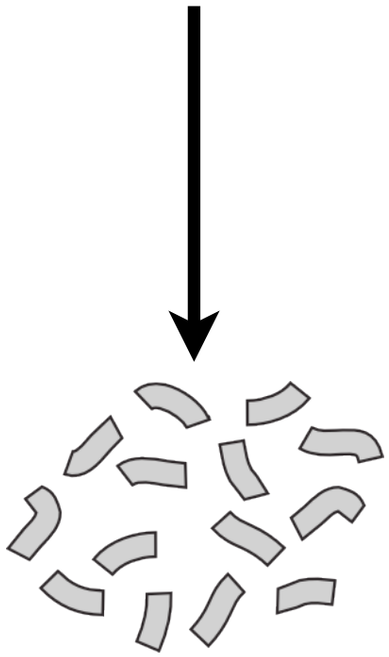
$M(t) \in \mathbb{N}$ → $\varnothing$

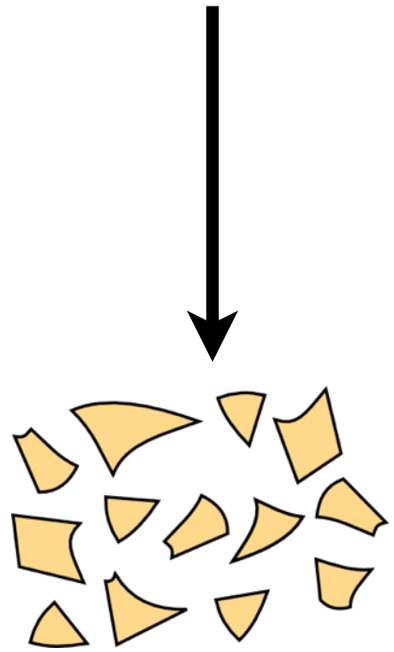
$P(t) \in \mathbb{N}$ → $\varnothing$











Protein production

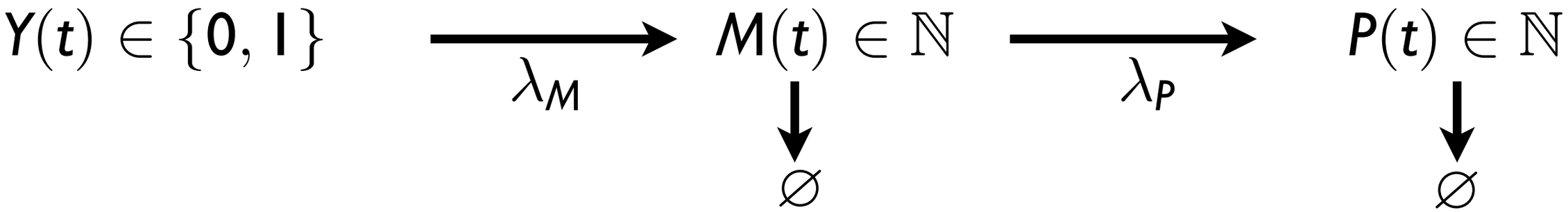
$$Y(t) \in \{0, 1\} \longrightarrow M(t) \in \mathbb{N} \longrightarrow P(t) \in \mathbb{N}$$

$$M(t) \downarrow \varnothing \qquad P(t) \downarrow \varnothing$$

$$M = \iint_{\mathbb{R}\times\mathbb{R}_+} \mathbb{1}_{\{u \le 0 \le u+v\}} \mathcal{N}_{\lambda_M}(\mathrm{d}u, \mathrm{d}v)$$

$$P = \iint_{\mathbb{R}\times\mathbb{R}_+} \mathcal{N}_{\lambda_M}(\mathrm{d}u, \mathrm{d}v) \left[\iint_{\mathbb{R}\times\mathbb{R}_+} \mathbb{1}_{\{u \le x \le u+v\}} \mathbb{1}_{\{x \le 0 \le x+y\}} \mathcal{N}^u_{\lambda_P}(\mathrm{d}x, \mathrm{d}y) \right]$$

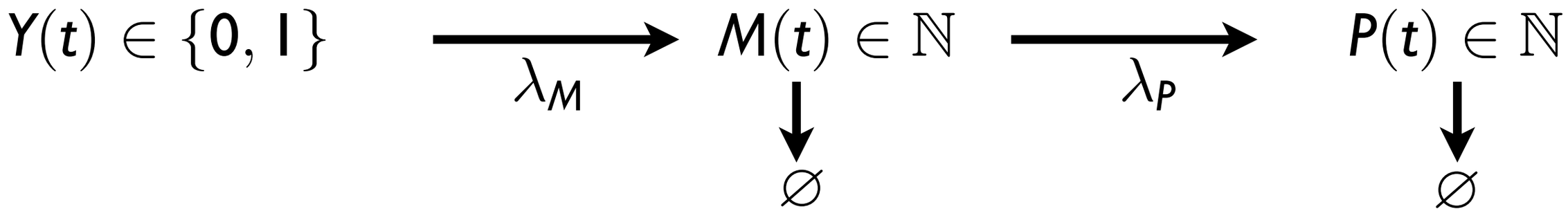
Protein production



Target protein copies:

$$\mathbb{E}[P] = 100$$

Protein production



Target protein copies:

$$\mathbb{E}[P] = 100$$

Two possible strategies:

small λ_M (few mRNAs)
large λ_P

large λ_M (many mRNAs)
small λ_P

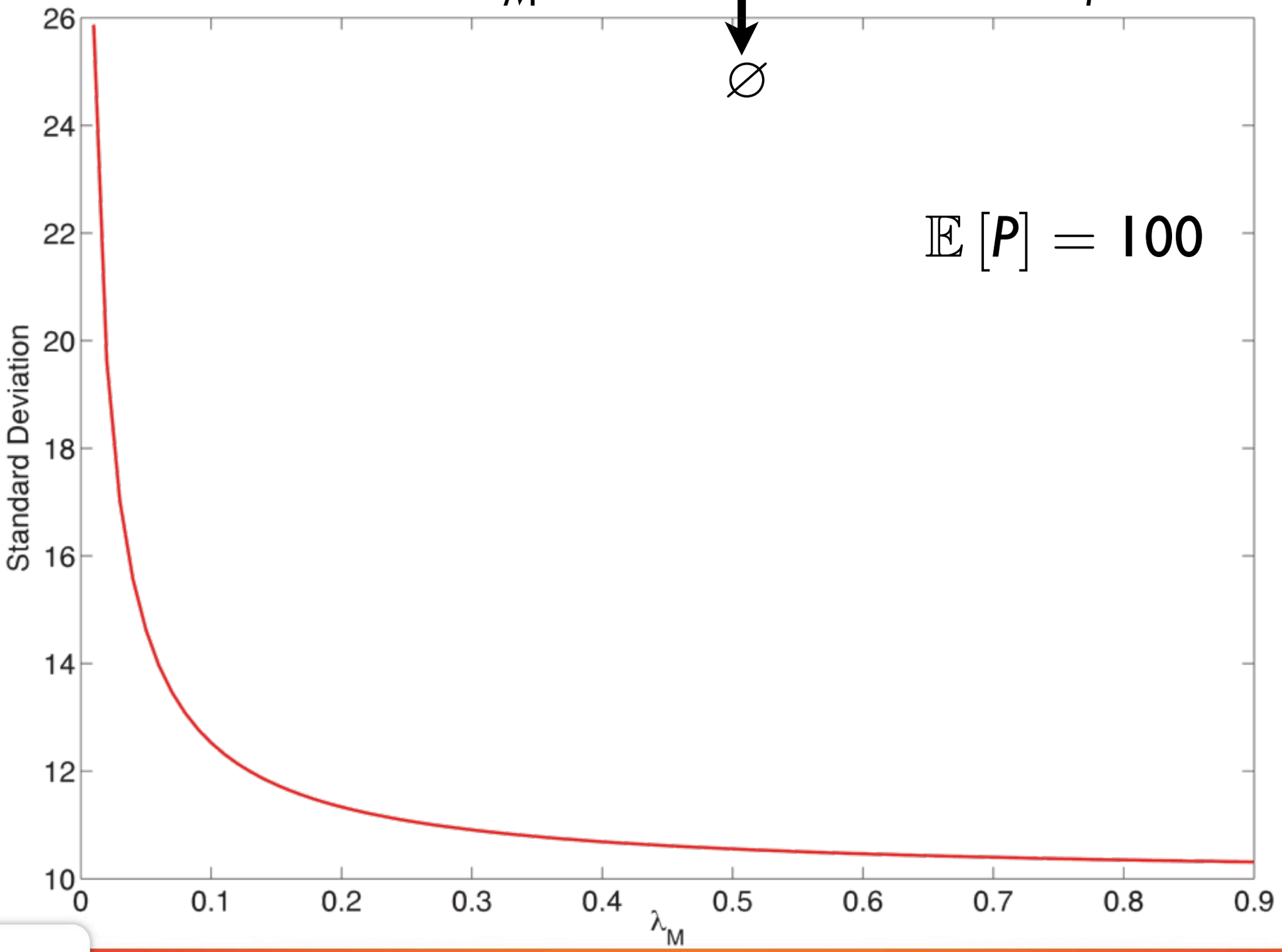
Protein production

$Y(t) \in \{0, 1\}$ $\xrightarrow{\lambda_M}$ $M(t) \in \mathbb{N}$ $\xrightarrow{\lambda_P}$ $P(t) \in \mathbb{N}$

$M(t) \to \varnothing$

$P(t) \to \varnothing$

$\mathbb{E}[P] = 100$



A few results:

- Quantitative characterization of fluctuations

$$\mathrm{var}(P) = \mathbb{E}[P]\left[1 + \frac{\lambda_3}{(\mu_2+\mu_3)} + \frac{\lambda_2\lambda_3(1-\delta_+)(\Lambda+\mu_2+\mu_3)}{(\mu_2+\mu_3)(\Lambda+\mu_2)(\Lambda+\mu_3)}\right]$$

- Rigorous (and controlled) analysis:
identification of the crucial steps in gene expression

- Counter-intuitive (or surprising) results

- Identification of critical behaviour

- Model as hypothesis-testing framework

Gene expression: work to do...

- Interaction between proteins: how does it impact on fluctuations?
- More realistic (treatable) model
- Control in stochastic environment

More in general...

- Deeper cooperation between maths and biology

Gene expression: work to do...

- Interaction between proteins: how does it impact on fluctuations?
- More realistic (treatable) model
- Control in stochastic environment

More in general...

- Deeper cooperation between maths and biology

Crit Rev Biochem Mol Biol. 2011 April ; 46(2): 137–151. doi:10.3109/10409238.2011.556597.

Mathematical modeling of gene expression: a guide for the perplexed biologist

Gene expression: work to do...

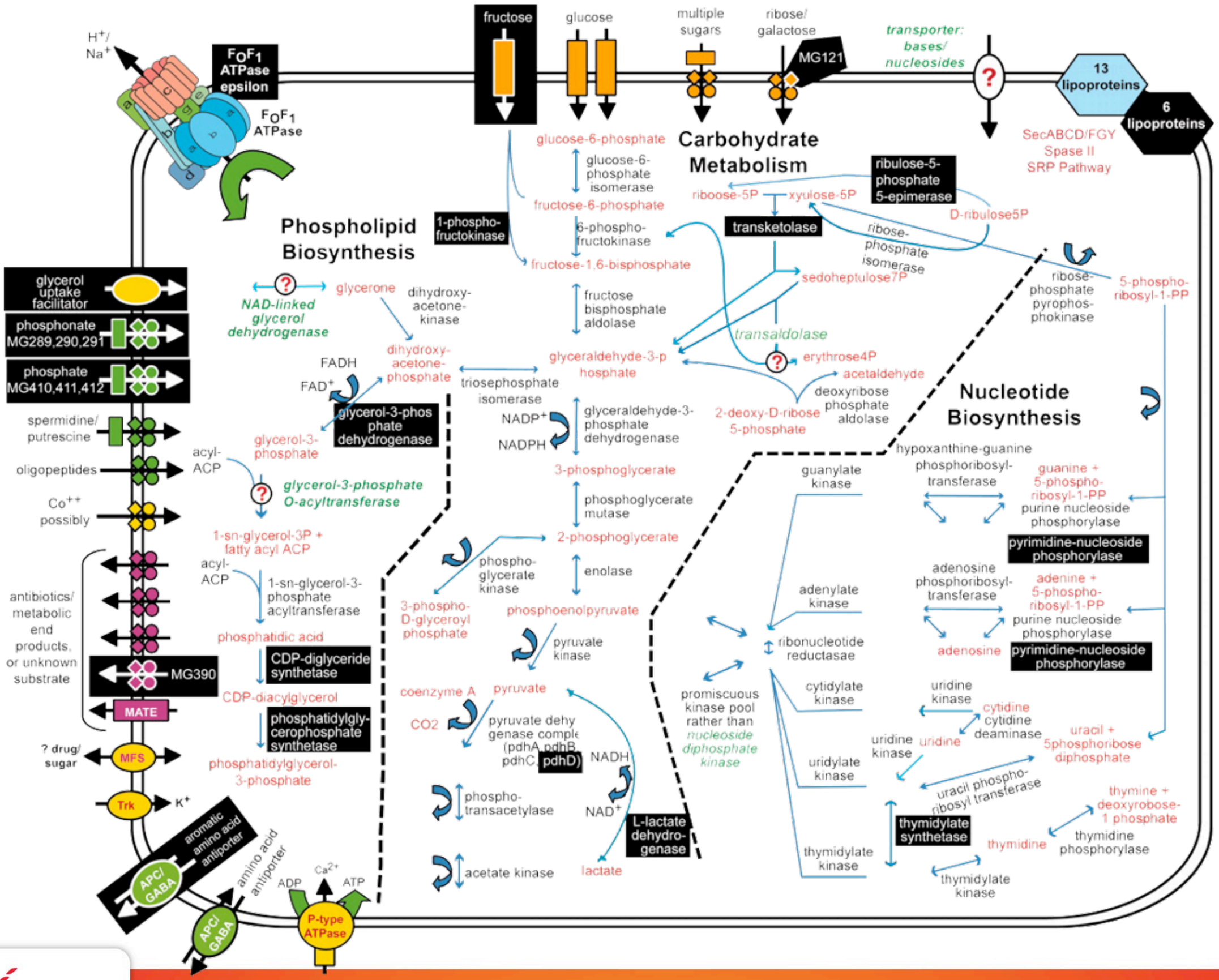
- Interaction between proteins: how does it impact on fluctuations?
- More realistic (treatable) model
- Control in stochastic environment

More in general...

- Deeper cooperation between maths and biolo

Crit Rev Biochem Mol Biol. 2011 April ; 46(2): 137–151. doi:10.3109/10409238.2011.556

Mathematical modeling of gene expression: a guide fo perplexed biologist


H+/ Na+
FoF1 ATPase epsilon
FoF1 ATPase
fructose
glucose
multiple sugars
ribose/ galactose
MG121
transporter: bases/ nucleosides
?
13 lipoproteins
6 lipoproteins
SecABCD/FGY Spase II SRP Pathway
Carbohydrate Metabolism
glucose-6-phosphate
glucose-6-phosphate isomerase
fructose-6-phosphate
1-phospho-fructokinase
6-phospho-fructokinase
fructose-1,6-bisphosphate
fructose bisphosphate aldolase
glyceraldehyde-3-phosphate
triosephosphate isomerase
dihydroxy-acetone-phosphate
glycerol-3-phosphate dehydrogenase
FADH
FAD+
glycerol-3-phosphate
NADP+
NADPH
glyceraldehyde-3-phosphate dehydrogenase
3-phosphoglycerate
phosphoglycerate mutase
2-phosphoglycerate
enolase
phosphoenolpyruvate
pyruvate kinase
pyruvate
pyruvate dehydrogenase complex (pdhA pdhB pdhC pdhD)
coenzyme A
CO2
NADH
NAD+
L-lactate dehydro-genase
lactate
phospho-transacetylase
acetate kinase
phospho-glycerate kinase
3-phospho-D-glyceroyl phosphate
Phospholipid Biosynthesis
NAD-linked glycerol dehydrogenase
glycerone
dihydroxy-acetone-kinase
glycerol-3-phosphate O-acyltransferase
acyl-ACP
1-sn-glycerol-3P + fatty acyl ACP
1-sn-glycerol-3-phosphate acyltransferase
phosphatidic acid
CDP-diglyceride synthetase
CDP-diacylglycerol
phosphatidylglycerophosphate synthetase
phosphatidylglycerol-3-phosphate
ribulose-5-phosphate 5-epimerase
riboose-5P
xyulose-5P
transketolase
D-ribulose5P
ribose-phosphate isomerase
sedoheptulose7P
transaldolase
erythrose4P
acetaldehyde
2-deoxy-D-ribose 5-phosphate
deoxyribose phosphate aldolase
ribose-phosphate pyrophos-phokinase
5-phospho-ribosyl-1-PP
Nucleotide Biosynthesis
hypoxanthine-guanine phosphoribosyl-transferase
guanine + 5-phospho-ribosyl-1-PP
purine nucleoside phosphorylase
pyrimidine-nucleoside phosphorylase
guanylate kinase
adenylate kinase
adenosine phosphoribosyl-transferase
adenine + 5-phospho-ribosyl-1-PP
purine nucleoside phosphorylase
adenosine
pyrimidine-nucleoside phosphorylase
ribonucleotide reductasae
promiscuous kinase pool rather than nucleoside diphosphate kinase
cytidylate kinase
uridine kinase
cytidine
cytidine deaminase
uridine
uridine kinase
uracil + 5phosphoribose diphosphate
uridylate kinase
uracil phospho-ribosyl transferase
thymidylate synthetase
thymine + deoxyrobose-1 phosphate
thymidine
thymidine phosphorylase
thymidylate kinase
thymidylate kinase
glycerol uptake facilitator
phosphonate MG289,290,291
phosphate MG410,411,412
spermidine/ putrescine
oligopeptides
Co++ possibly
antibiotics/ metabolic end products, or unknown substrate
MG390
MATE
? drug/ sugar
MFS
Trk
K+
aromatic amino acid antiporter
APC/ GABA
amino acid antiporter
APC/ GABA
ADP
Ca2+
ATP
P-type ATPase

Sophisticated

Many phenomena to be understood

...but we love it!

Sophisticated

Many phenomena to be understood

Thanks.

...but we love it!