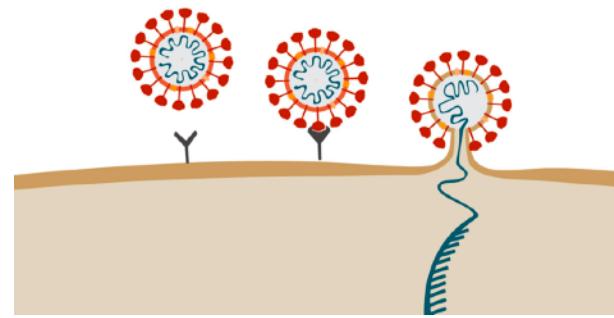
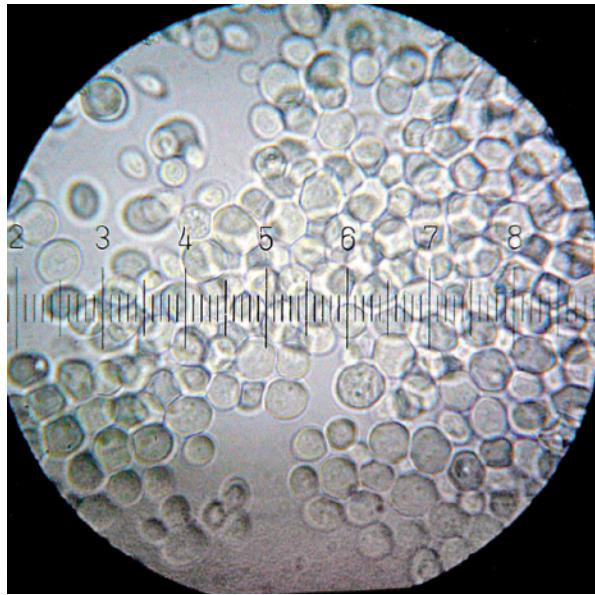


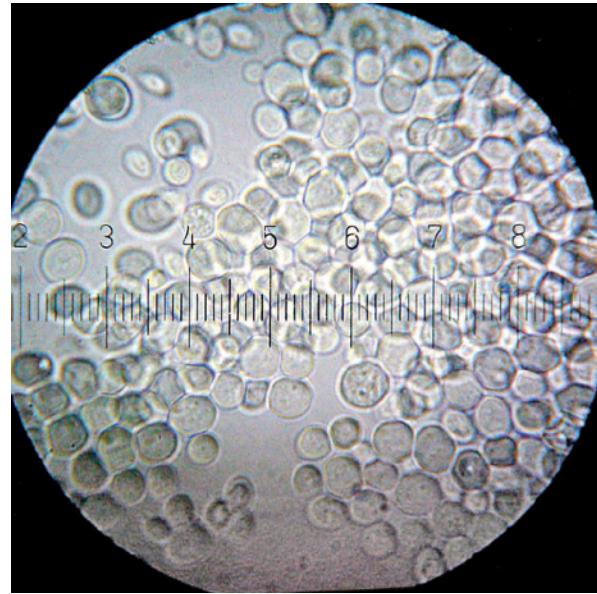
# All that glitters is not Deep Learning in Life Sciences (but sometimes it is!)

Jakub M. Tomczak

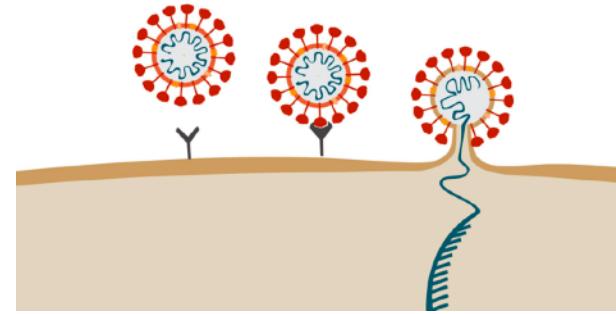
# TYPICAL PROBLEMS IN LIFE SCIENCES



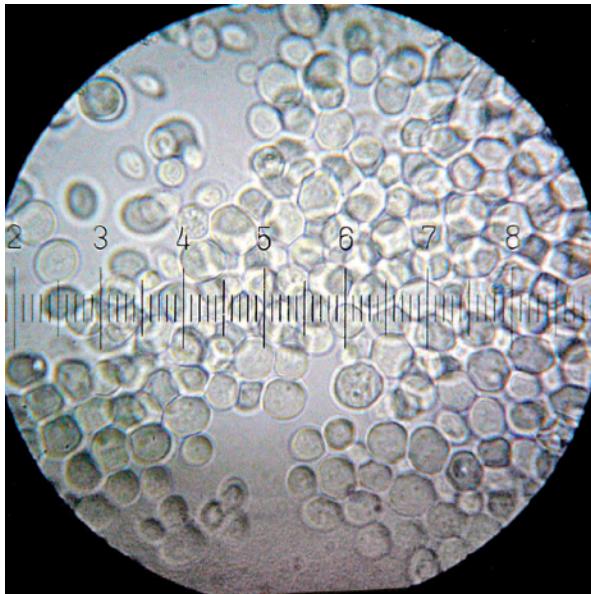
# TYPICAL PROBLEMS IN LIFE SCIENCES



How to model  
biochemical  
processes?

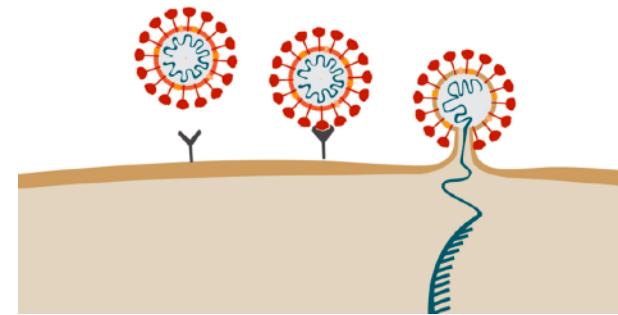


# TYPICAL PROBLEMS IN LIFE SCIENCES

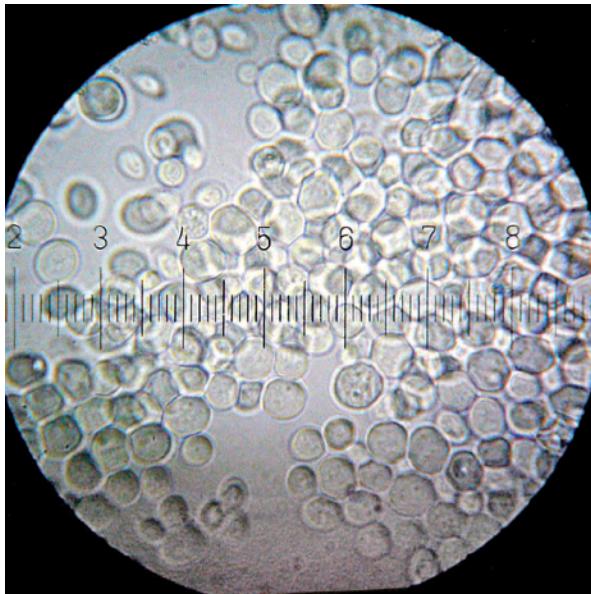


How to model biochemical processes?

How many cells do we see?

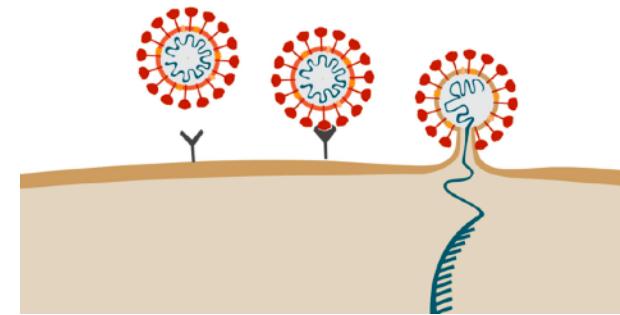


# TYPICAL PROBLEMS IN LIFE SCIENCES



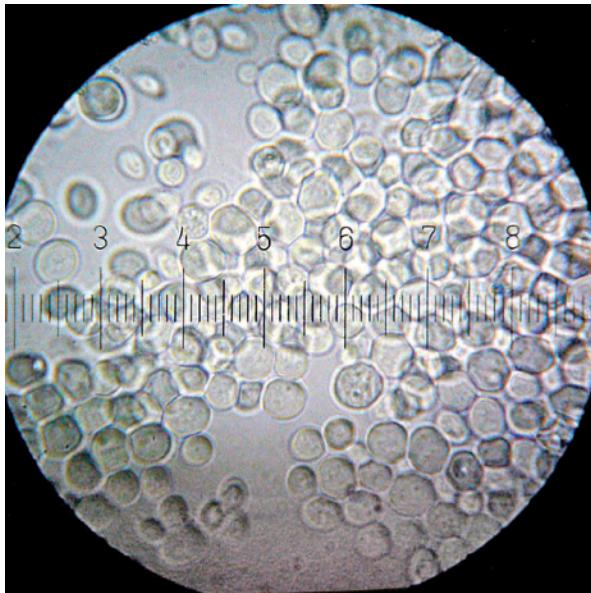
How to model biochemical processes?

How many cells do we see?



How fast are enzymes catalyzed?

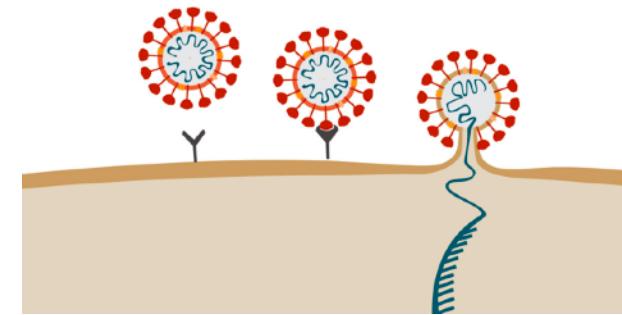
# TYPICAL PROBLEMS IN LIFE SCIENCES



**Metabolism**

How to model biochemical processes?

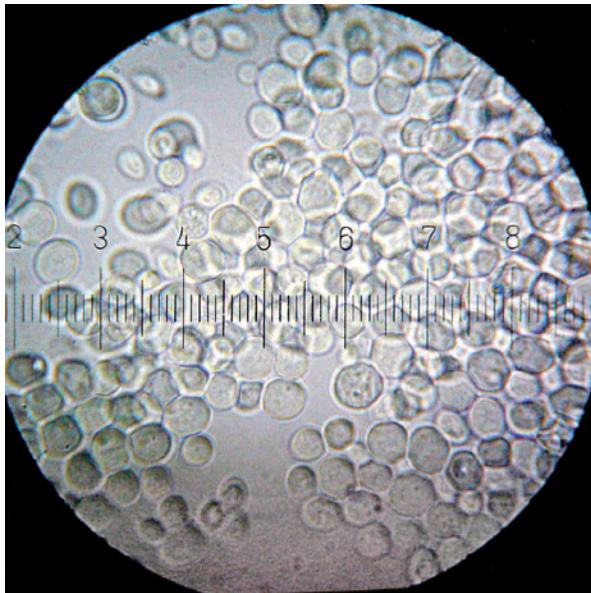
How many cells do we see?



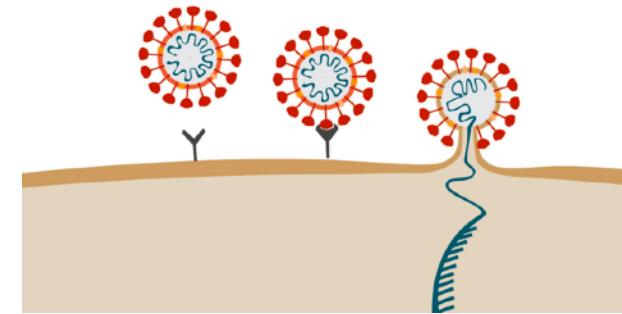
How fast are enzymes catalyzed?

**Enzyme kinetics**

# TYPICAL PROBLEMS IN LIFE SCIENCES



How to model  
biochemical  
processes?



How many cells  
do we see?

Biochemistry  
Anatomy  
Physiology

Metabolism      Ecology      Virology      Enzyme kinetics

Cell biology

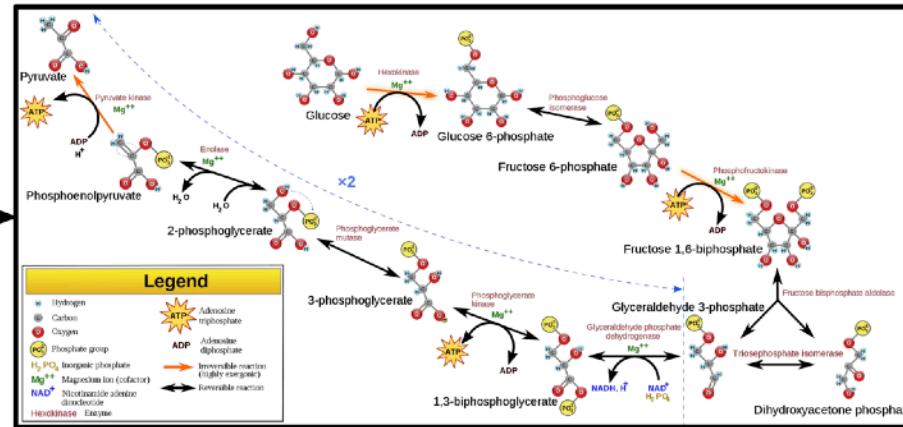
Microbiology      Botany      ...

# EXAMPLE: GLYCOLYSIS



Input  
(nutrients)

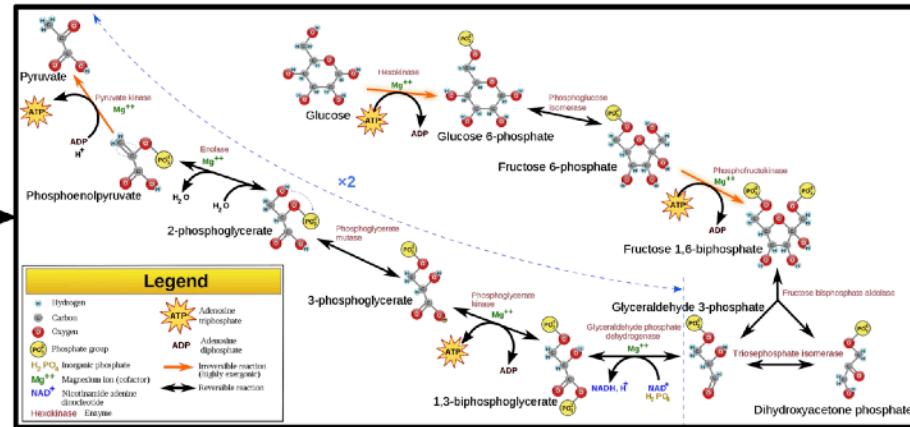
# EXAMPLE: GLYCOLYSIS



Input  
(nutrients)

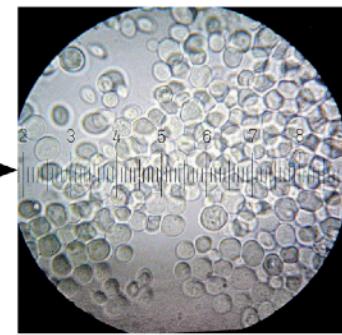
Biochemical processes  
(enzymes + products)

# EXAMPLE: GLYCOLYSIS



Input  
(nutrients)

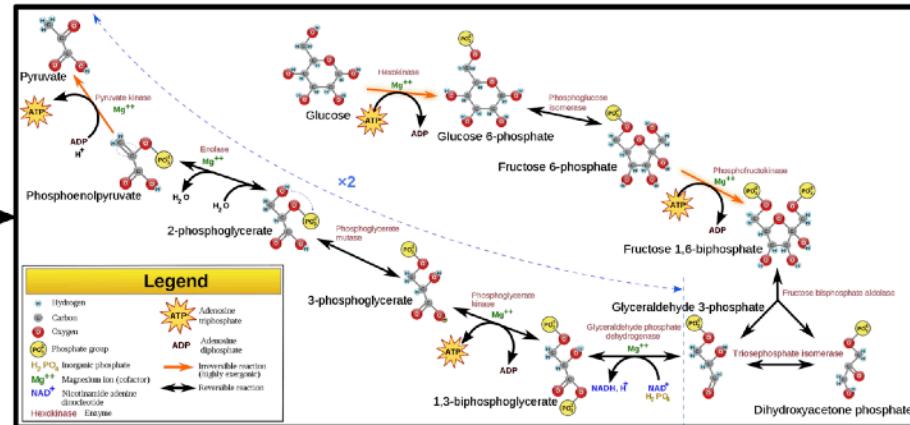
Biochemical processes  
(enzymes + products)



Output  
(living cells)

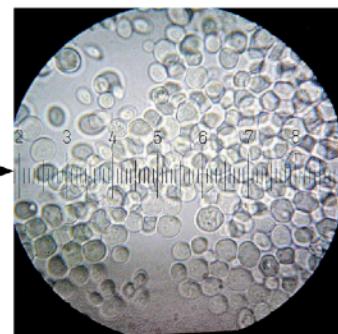
# EXAMPLE: GLYCOLYSIS

How to understand the phenomenon?



Input  
(nutrients)

Biochemical processes  
(enzymes + products)

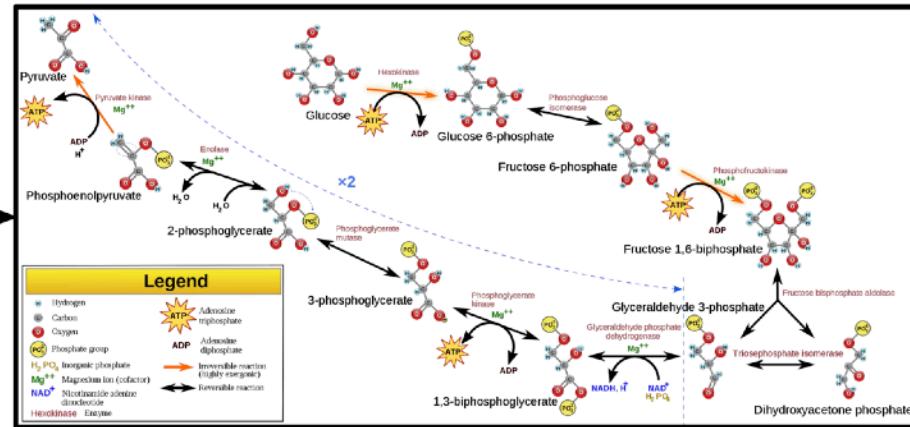


Output  
(living cells)

# EXAMPLE: GLYCOLYSIS

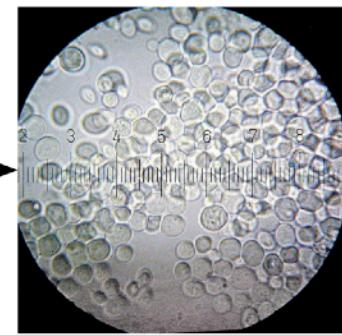
How to understand the phenomenon?

Model and identify each reaction



Input  
(nutrients)

Biochemical processes  
(enzymes + products)

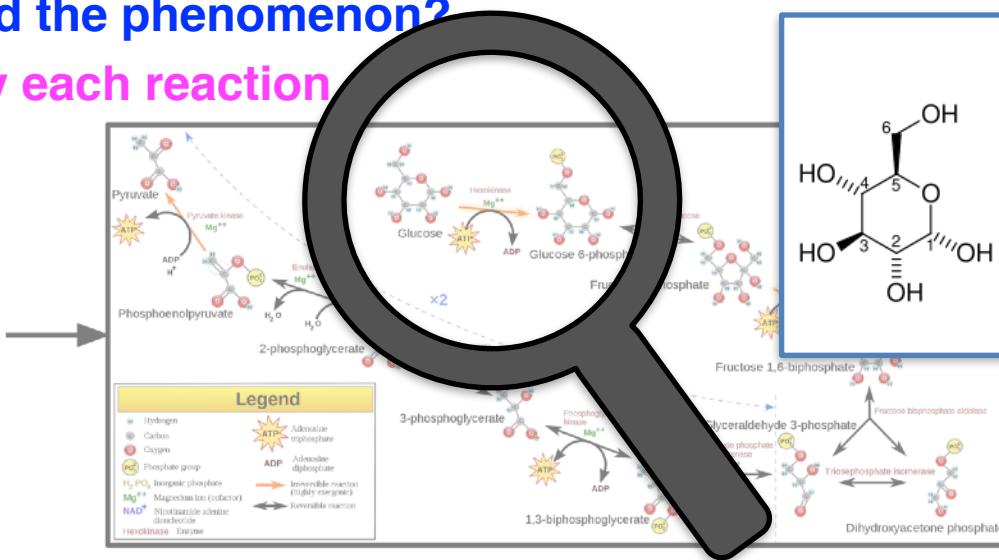


Output  
(living cells)

# EXAMPLE: GLYCOLYSIS

How to understand the phenomenon?

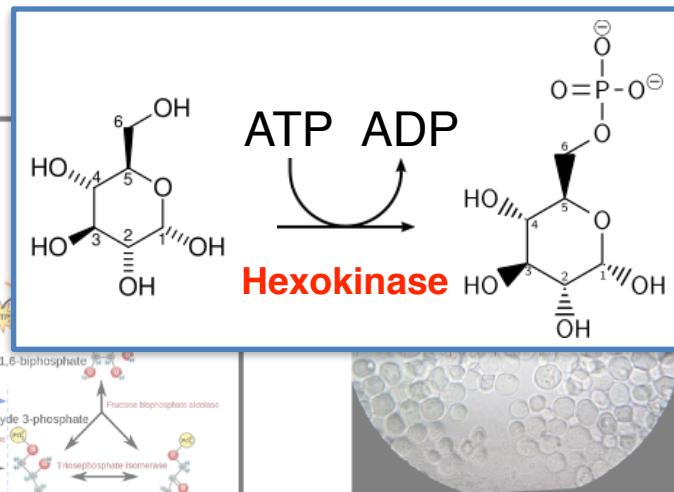
Model and identify each reaction



Input  
(nutrients)

Biochemical processes  
(enzymes + products)

Output  
(living cells)



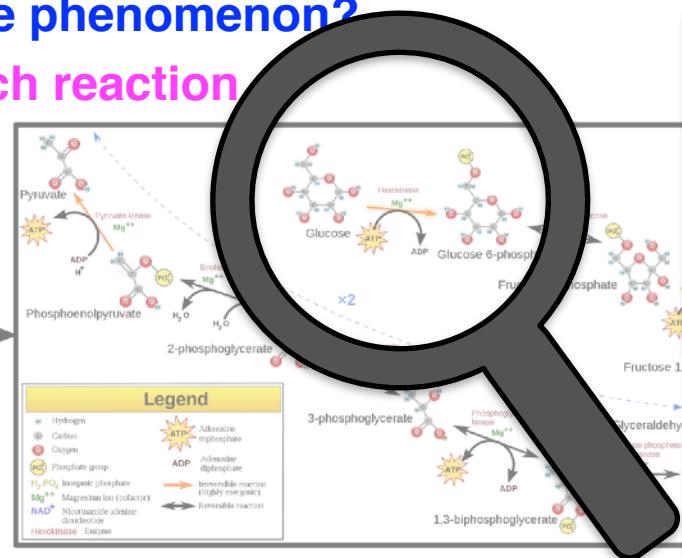
# EXAMPLE: GLYCOLYSIS

How to understand the phenomenon?

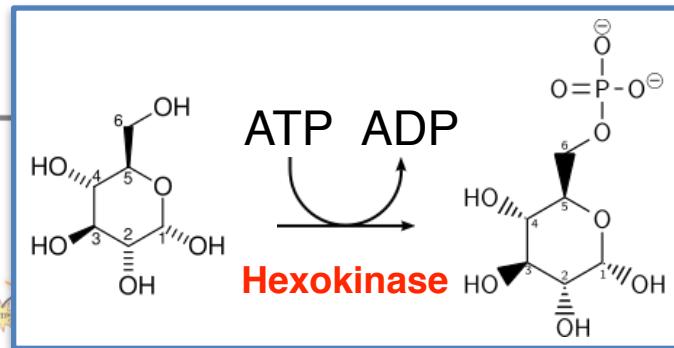
Model and identify each reaction



Input  
(nutrients)



Biochemical processes  
(enzymes + products)



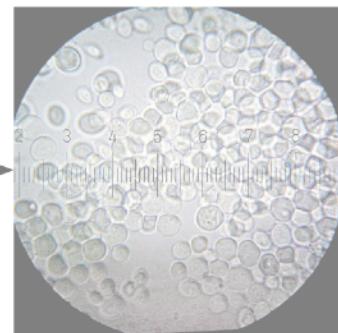
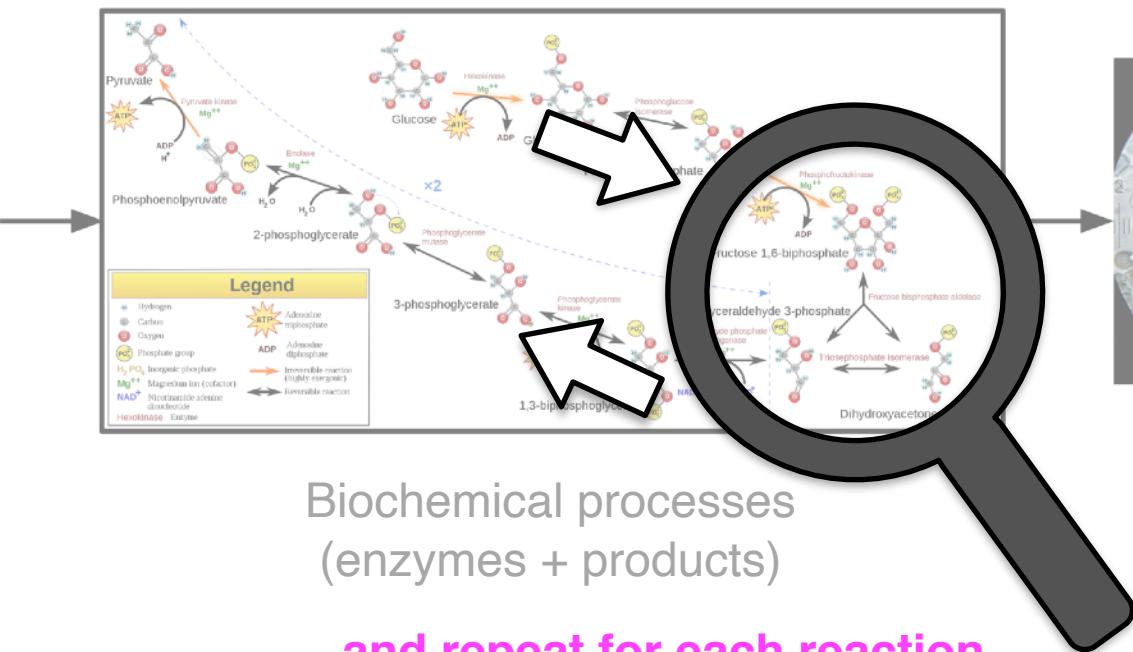
$$v = \frac{E_0 k_{cat} S}{K_M + S}$$

Output  
(living cells)  
+ DATA (measurements)

# EXAMPLE: GLYCOLYSIS

How to understand the phenomenon?

Model and identify each reaction



Input  
(nutrients)

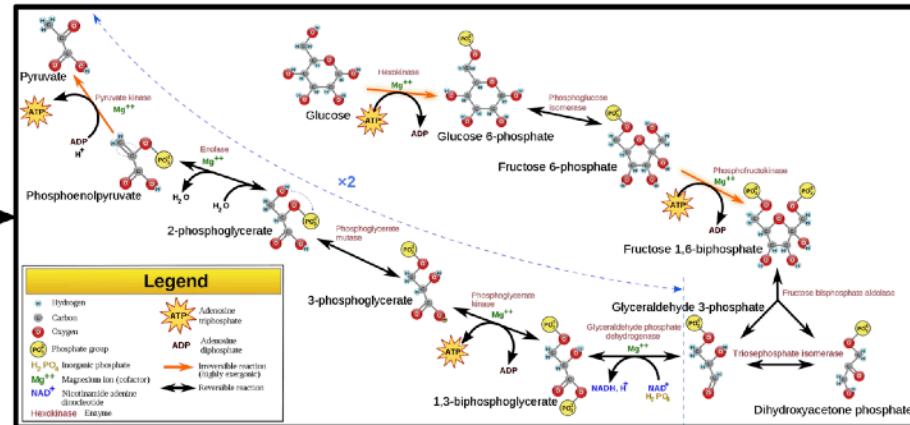
Biochemical processes  
(enzymes + products)

Output  
(living cells)

and repeat for each reaction  
(we must be **EFFICIENT!**)

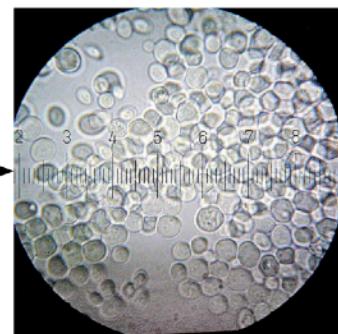
# EXAMPLE: GLYCOLYSIS

How to understand the phenomenon?



Input  
(nutrients)

Biochemical processes  
(enzymes + products)

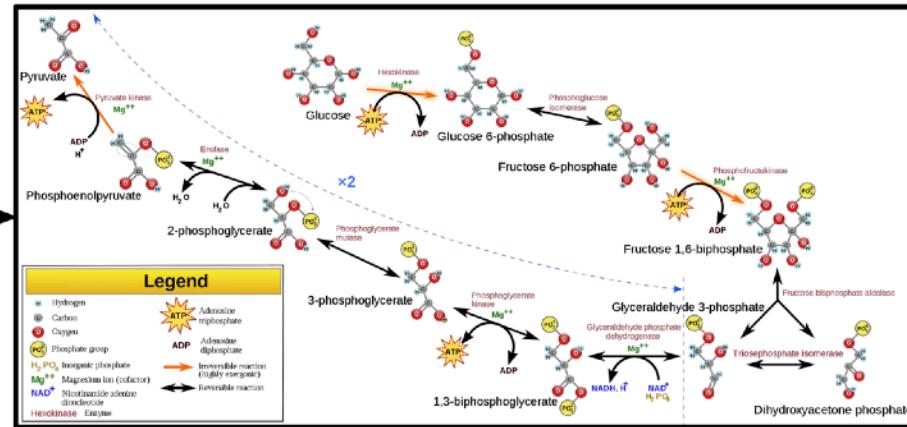


Output  
(living cells)

# EXAMPLE: GLYCOLYSIS

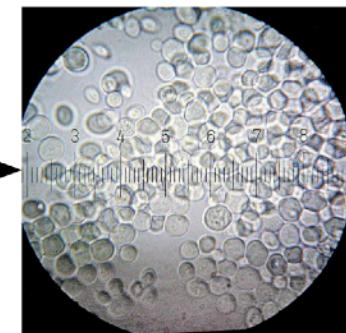
How to understand the phenomenon?

Model and identify the whole network at once



Input  
(nutrients)

Biochemical processes  
(enzymes + products)

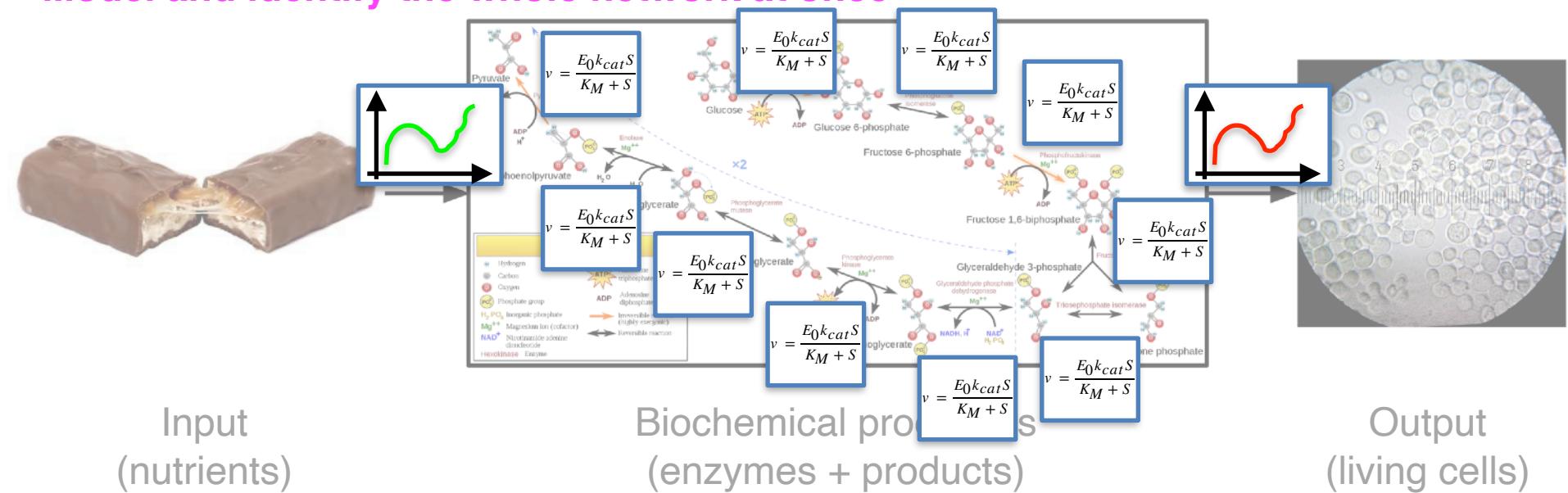


Output  
(living cells)

# EXAMPLE: GLYCOLYSIS

How to understand the phenomenon?

Model and identify the whole network at once



Input  
(nutrients)

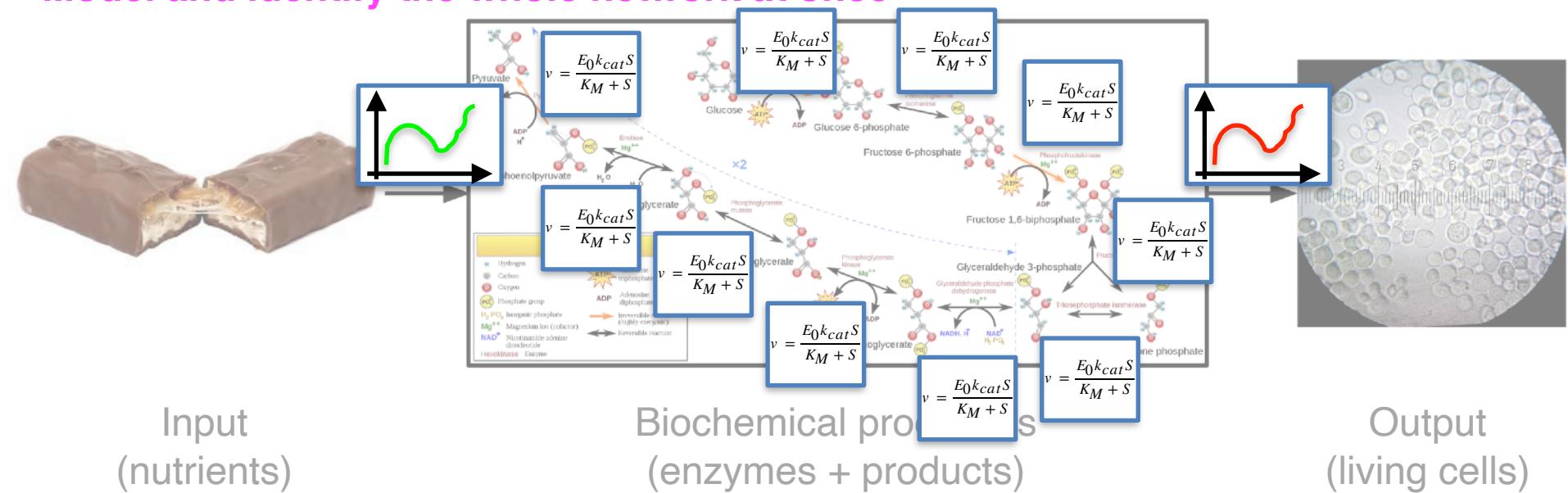
Biochemical processes  
(enzymes + products)

Output  
(living cells)

# EXAMPLE: GLYCOLYSIS

How to understand the phenomenon?

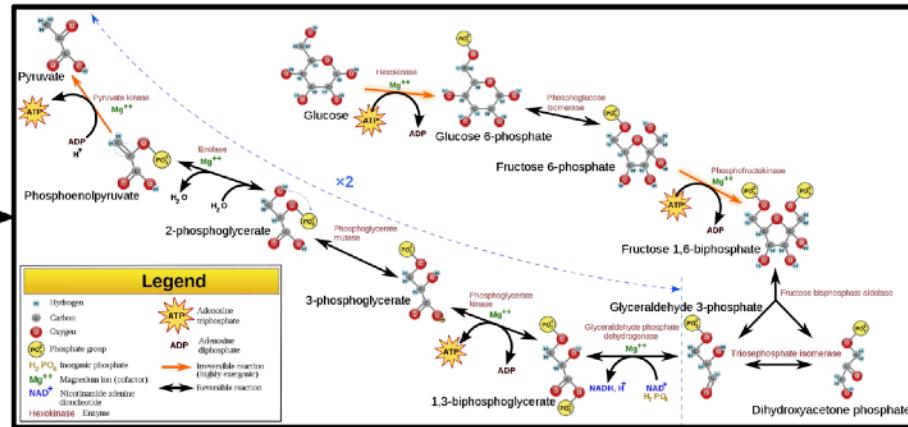
Model and identify the whole network at once



We know the mathematical description and want to identify the parameters of the network with limited measurements.

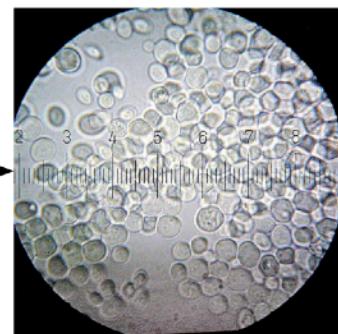
# EXAMPLE: GLYCOLYSIS

How to understand the phenomenon?



Input  
(nutrients)

Biochemical processes  
(enzymes + products)

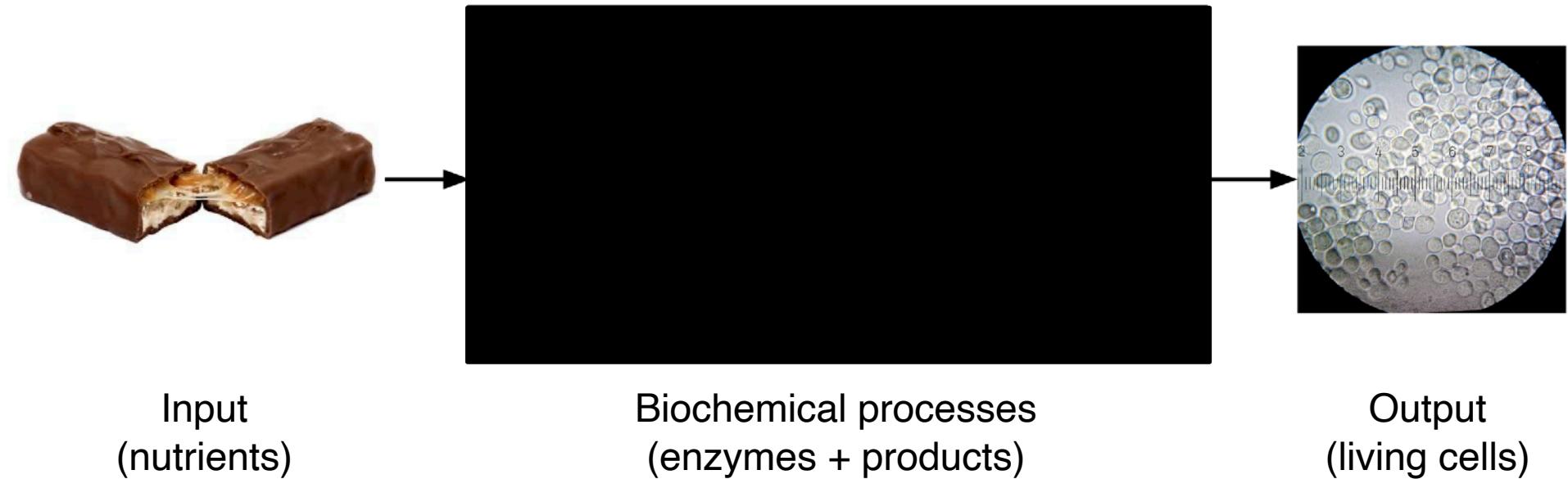


Output  
(living cells)

# EXAMPLE: GLYCOLYSIS

How to understand the phenomenon?

We don't know the “inside” and treat it as a black-box.



# EXAMPLE: GLYCOLYSIS

How to understand the phenomenon?

We don't know the “inside” and treat it as a black-box.



By learning the input-output dependency, we can understand (to some degree) the phenomenon or use the model to study it.

**Enzyme kinetics:** how the chemical reactions are catalyzed by enzymes.

**Goal:** Find the reaction rate (i.e., the speed at which a chemical reaction takes place) of *a single reaction*.

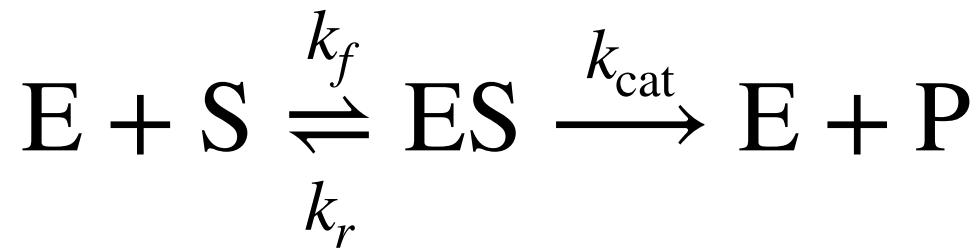
**Why?**

- Understanding the catalytic mechanism of an enzyme.
- Understanding the role of an enzyme in a chemical reaction.
- Understanding how an enzyme activity is controlled.
- Understanding how a drug (inhibitor) slows down the reaction.

## ENZYME KINETICS: MICHAELIS-MENTEN MODEL

The commonly used model in enzyme kinetics is the **Michaelis-Menten model**.

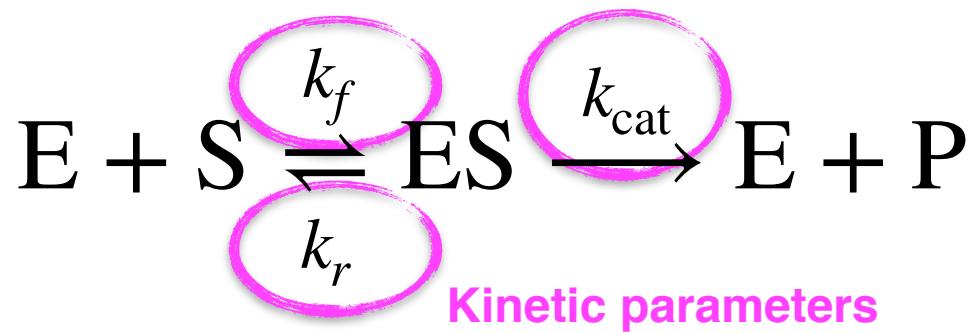
We consider a reversible reaction where an enzyme (E) binds to a substrate (S) to form a complex (ES) to irreversibly release a product (P) and free the enzyme:



# ENZYME KINETICS: MICHAELIS-MENTEN MODEL

The commonly used model in enzyme kinetics is the **Michaelis-Menten model**.

We consider a reversible reaction where an enzyme (E) binds to a substrate (S) to form a complex (ES) to irreversibly release a product (P) and free the enzyme:



## ENZYME KINETICS: MICHAELIS-MENTEN MODEL

We consider a reversible reaction where an enzyme (E) binds to a substrate (S) to form a complex (ES) to irreversibly release a product (P) and free the enzyme.

Considering the system in a quasi-steady-state, we get:

$$v = \frac{dP}{dt} = \frac{V_{max}S}{K_M + S} = \frac{E_0 k_{cat} S}{K_M + S}$$

# ENZYME KINETICS: MICHAELIS-MENTEN MODEL

We consider a reversible reaction where an enzyme (E) binds to a substrate (S) to form a complex (ES) to irreversibly release a product (P) and free the enzyme.

Considering the system in a quasi-steady-state, we get:

$$v = \frac{dP}{dt} = \frac{V_{max}S}{K_M + S} = \frac{E_0 k_{cat} S}{K_M + S}$$

Annotations for the Michaelis-Menten equation:

- max. velocity** points to  $V_{max}S$
- init concentration of enzyme** points to  $E_0$
- catalytic rate constant** points to  $k_{cat}S$
- Michaelis constant** points to  $K_M$

# ENZYME KINETICS: MICHAELIS-MENTEN MODEL

We consider a reversible reaction where an enzyme (E) binds to a substrate (S) to form a complex (ES) to irreversibly release a product (P) and free the enzyme.

Considering the system in a quasi-steady-state, we get:

$$v = \frac{dP}{dt} = \frac{V_{max}S}{K_M + S} = \frac{E_0 k_{cat} S}{K_M + S}$$

Annotations for the Michaelis-Menten equation:

- max. velocity**: Points to  $V_{max}S$ .
- init concentration of enzyme**: Points to  $E_0$ .
- catalytic rate constant**: Points to  $k_{cat}$ .
- catalytic efficiency**: Points to  $\frac{k_{cat}}{K_M}$ .
- Michaelis constant**: Points to  $K_M$ .

## ENZYME KINETICS: MICHAELIS-MENTEN MODEL

We consider a reversible reaction where an enzyme (E) binds to a substrate (S) to form a complex (ES) to irreversibly release a product (P) and free the enzyme.

Considering the system in a quasi-steady-state, we get:

$$v = \frac{dP}{dt} = \frac{V_{max}S}{K_M + S} = \frac{E_0 k_{cat} S}{K_M + S}$$

Solution:

$$v = V_{max} \left( 1 - \exp(-bS) \right)$$

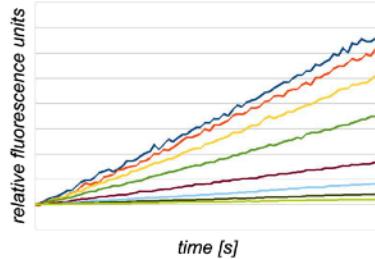
# ENZYME KINETICS: MICHAELIS-MENTEN MODEL

How to find the kinetic parameter values? **The standard approach.**

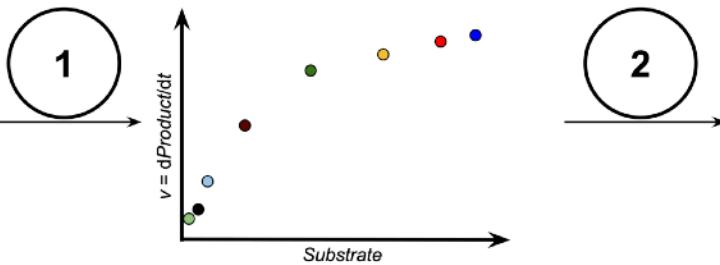
# ENZYME KINETICS: MICHAELIS-MENTEN MODEL

How to find the kinetic parameter values? **The standard approach.**

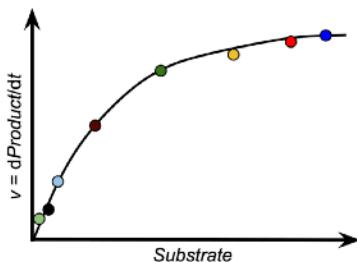
Measurements at different substrate concentrations



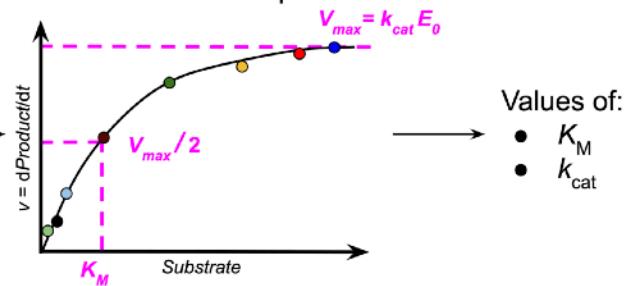
Michaelis-Menten plot



Michaelis-Menten plot



Michaelis-Menten plot

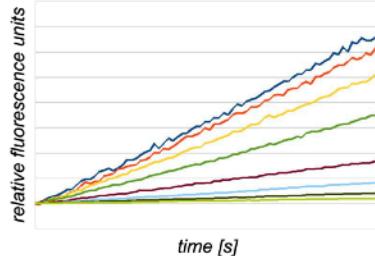


Values of:  
•  $K_M$   
•  $k_{cat}$

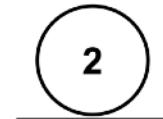
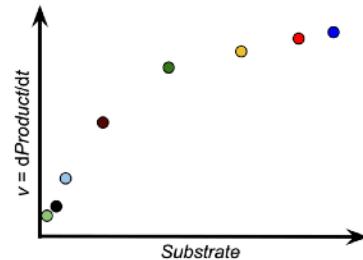
# ENZYME KINETICS: MICHAELIS-MENTEN MODEL

How to find the kinetic parameter values? **The standard approach.**

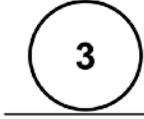
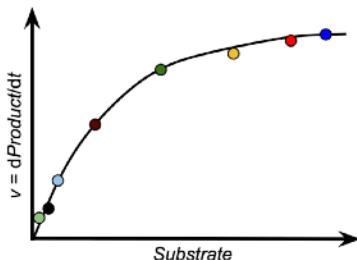
Measurements at different substrate concentrations



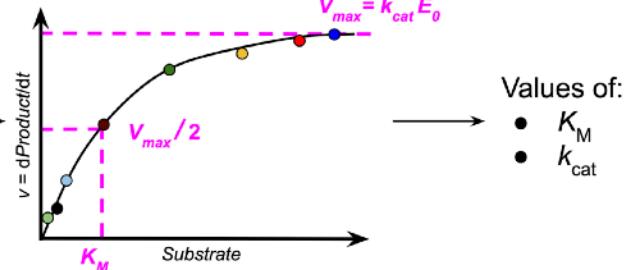
Michaelis-Menten plot



Michaelis-Menten plot



Michaelis-Menten plot



Values of:  
•  $K_M$   
•  $k_{cat}$

## Pros:

- easy
- pretty accurate

## Cons:

- super laborious
- wastes a lot of substrate
- time-consuming

# ENZYME KINETICS: MICHAELIS-MENTEN MODEL

How to find the kinetic parameter values? Our approach: ABC.

# ENZYME KINETICS: MICHAELIS-MENTEN MODEL

How to find the kinetic parameter values? **Our approach: ABC.**

**Our main motivation: Use (cheap) computations instead  
of laborious and costly work in a lab.**

**Proposition: Use Approximate Bayesian Computation.**

# ENZYME KINETICS: MICHAELIS-MENTEN MODEL

How to find the kinetic parameter values? **Our approach: ABC.**

1. Initialize  $\theta_t := \theta_0$ .

2. For  $t \in \{0, 1, \dots, T - 1\}$ :

(i) (**Generate**) Sample  $\theta' \sim q(\theta | \theta_t)$ .

(ii) (**Evaluate**) Calculate the distance:

$$\Delta(\theta') = \|x - f(\theta')\|^2$$

(iii) (**Select**)

If  $\Delta(\theta') < \varepsilon$ , then  $\theta_{t+1} := \theta'$ .

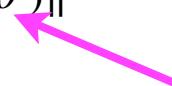
Else  $\theta_{t+1} := \theta_t$ .

# ENZYME KINETICS: MICHAELIS-MENTEN MODEL

How to find the kinetic parameter values? Our approach: ABC.

1. Initialize  $\theta_t := \theta_0$ .
2. For  $t \in \{0, 1, \dots, T - 1\}$ :
  - (i) (**Generate**) Sample  $\theta' \sim q(\theta | \theta_t)$ .
  - (ii) (**Evaluate**) Calculate the distance:

$$\Delta(\theta') = \|x - f(\theta')\|^2$$

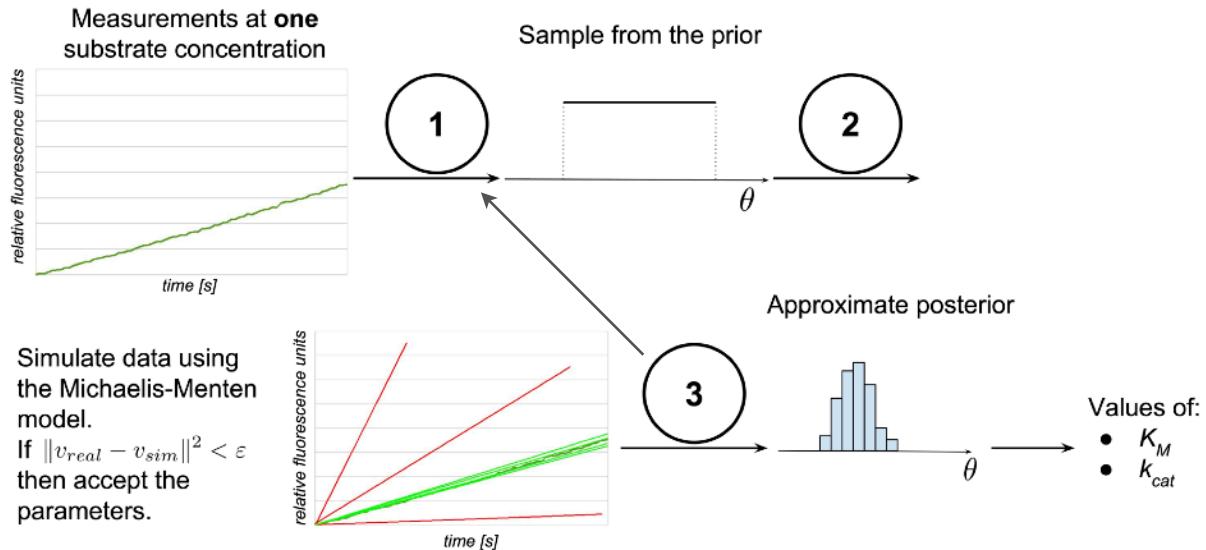
(iii) (**Select**)  simulator (the MM model)

If  $\Delta(\theta') < \varepsilon$ , then  $\theta_{t+1} := \theta'$ .

Else  $\theta_{t+1} := \theta_t$ .

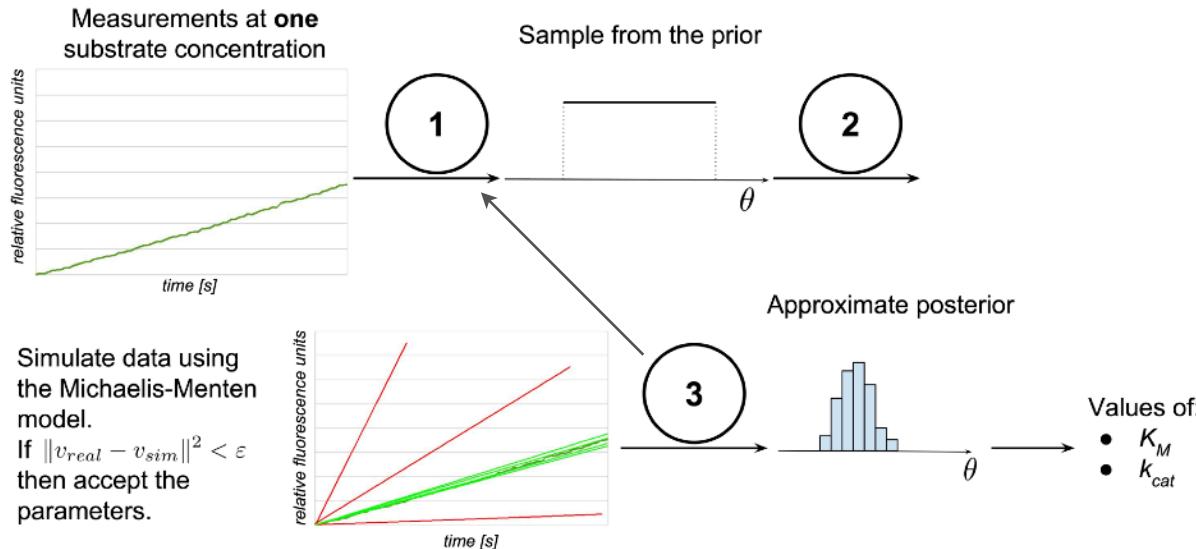
# ENZYME KINETICS: MICHAELIS-MENTEN MODEL

How to find the kinetic parameter values? Our approach: ABC.



# ENZYME KINETICS: MICHAELIS-MENTEN MODEL

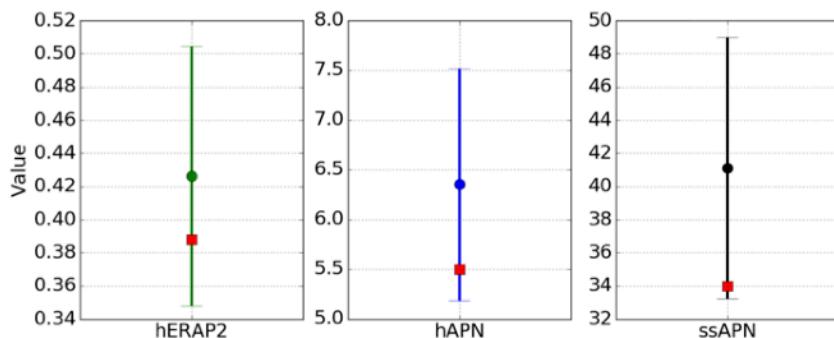
How to find the kinetic parameter values? Our approach: ABC.



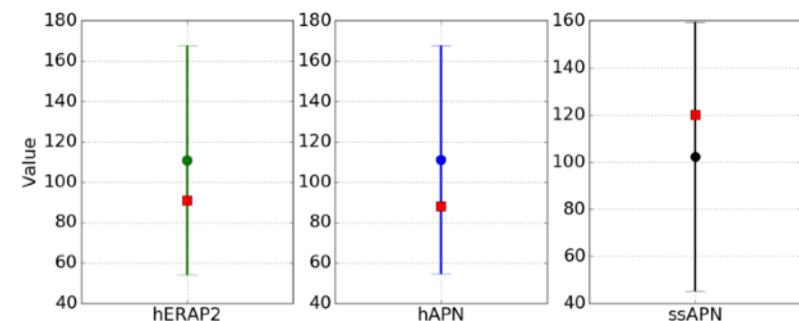
# ENZYME KINETICS: THREE AMINOPEPTIDASES

human aminopeptidase (*hAPN*), *Sus scrofa APN* (*ssAPN*) and human endoplasmic reticulum aminopeptidase 2 (*hERAP2*)

■ standard approach



$$k_{cat}$$

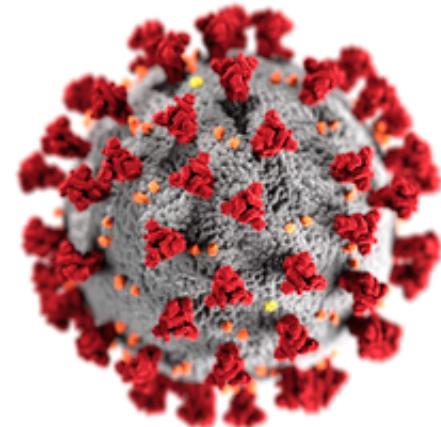


$$K_M$$

# ENZYME KINETICS: SARS-COV-1 & SARS-COV-2

In our recent study, we presented:

- An analysis of the active site of PLpro (enzyme) in SARS-CoV-1 (SARS) and SARS-CoV-2 (CoV2).
- A kinetic analysis of the Ub-AMC hydrolysis by PLpro from SARS and CoV2
- Ebselen and structural analogues of ebselen as potent covalent inhibitors of PLproCoV2



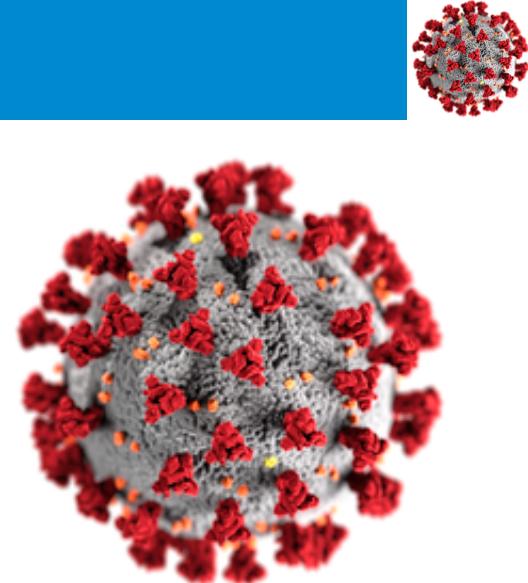
# ENZYME KINETICS: SARS-COV-1 & SARS-COV-2

In our recent study, we presented:

- An analysis of the active site of PLpro (enzyme) in SARS-CoV-1 (SARS) and SARS-CoV-2 (CoV2).

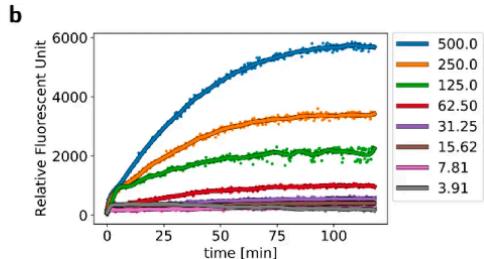
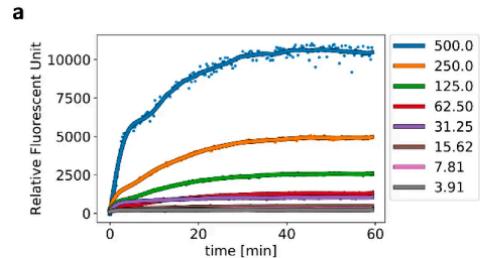
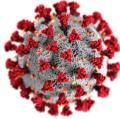
- **A kinetic analysis of the Ub-AMC hydrolysis by PLpro from SARS and CoV2**

- Ebselen and structural analogues of ebselen as potent covalent inhibitors of PLproCoV2



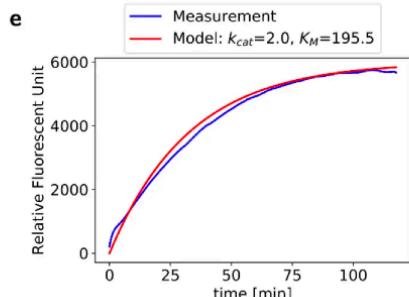
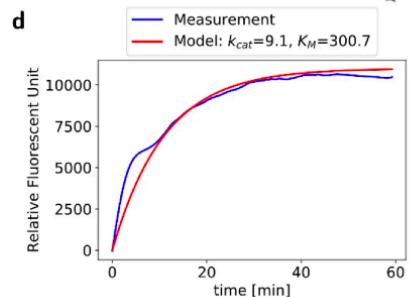
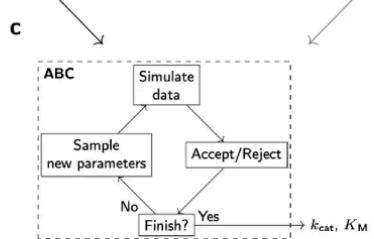
<sup>41</sup> Weglarz-Tomczak, E. et al. (2021). Identification of ebselen and its analogues as potent covalent inhibitors of papain-like protease from SARS-CoV-2. *Scientific reports*, 11(1), 1-10.

# ENZYME KINETICS: SARS-COV-1 & SARS-COV-2

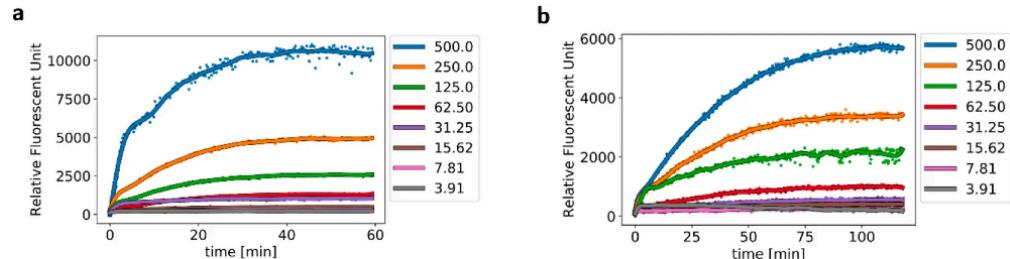
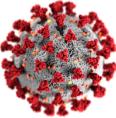


**f**

	$k_{cat} [s^{-1}]$	$K_M [\mu M]$	$k_{cat}/K_M [s^{-1}M^{-1}]$
PL <sup>pro</sup> SARS	$9.1 \pm 0.5$	$300.7 \pm 20.2$	<b><math>0.030 \pm 0.003</math></b>
PL <sup>pro</sup> CoV2	$2.0 \pm 0.3$	$195.5 \pm 5.2$	<b><math>0.010 \pm 0.001</math></b>

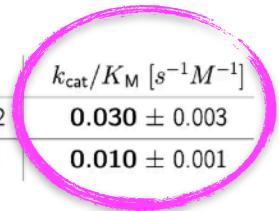


# ENZYME KINETICS: SARS-COV-1 & SARS-COV-2

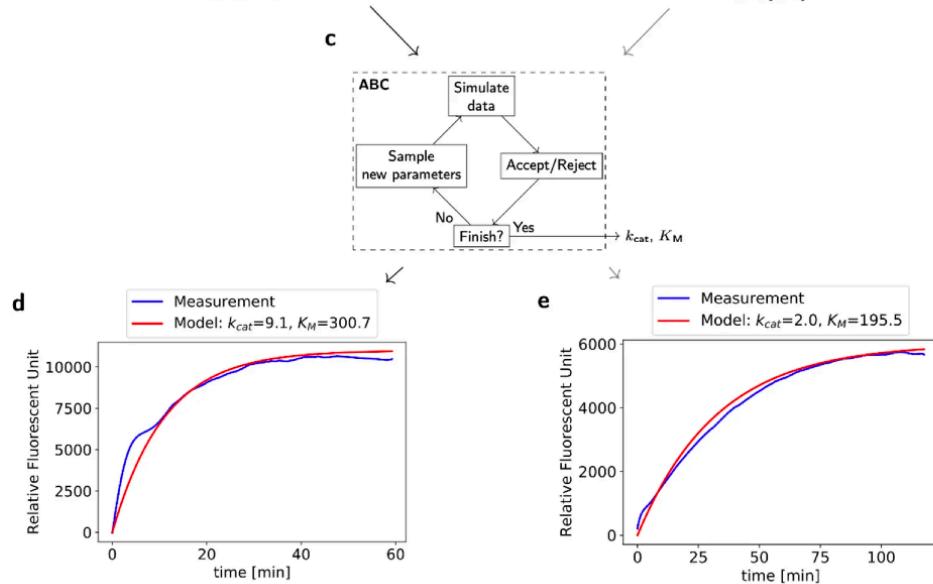


**f**

	$k_{cat} [s^{-1}]$	$K_M [\mu M]$	$k_{cat}/K_M [s^{-1}M^{-1}]$
PL <sup>pro</sup> SARS	$9.1 \pm 0.5$	$300.7 \pm 20.2$	<b><math>0.030 \pm 0.003</math></b>
PL <sup>pro</sup> CoV2	$2.0 \pm 0.3$	$195.5 \pm 5.2$	<b><math>0.010 \pm 0.001</math></b>



We see that SARS-CoV-1 is 3 times faster!



This confirms a known fact: once infected, SARS-CoV-1 was overall more aggressive and the disease developed faster.

# IDENTIFICATION OF WHOLE NETWORKS

Let us look at a network of reactions.

# IDENTIFICATION OF WHOLE NETWORKS: GENE REPRESSILATOR MODEL

Let us look at a network of reactions.

First, we focus on the well-known gene repressor model:

$$\frac{dm_1}{dt} = -m_1 + \frac{\alpha}{1 + p_3^n} + \alpha_0$$

$$\frac{dp_1}{dt} = -\beta(p_1 - m_1)$$

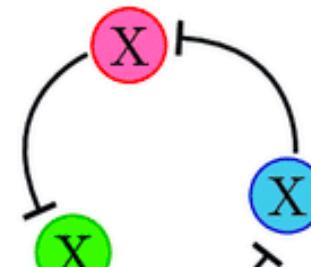
$$\frac{dm_2}{dt} = -m_2 + \frac{\alpha}{1 + p_1^n} + \alpha_0$$

$$\frac{dp_2}{dt} = -\beta(p_2 - m_2)$$

$$\frac{dm_3}{dt} = -m_3 + \frac{\alpha}{1 + p_2^n} + \alpha_0$$

$$\frac{dp_3}{dt} = -\beta(p_3 - m_3)$$

a)

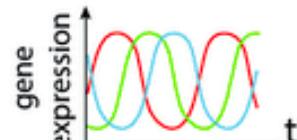


b)

gene expression



gene expression



# IDENTIFICATION OF WHOLE NETWORKS: GENE REPRESSILATOR MODEL

Let us look at a network of reactions.

First, we focus on the well-known gene repressor model:

$$\frac{dm_1}{dt} = -m_1 + \frac{\alpha}{1 + p_3^n} + \alpha_0$$

$$\frac{dp_1}{dt} = -\beta(p_1 - m_1)$$

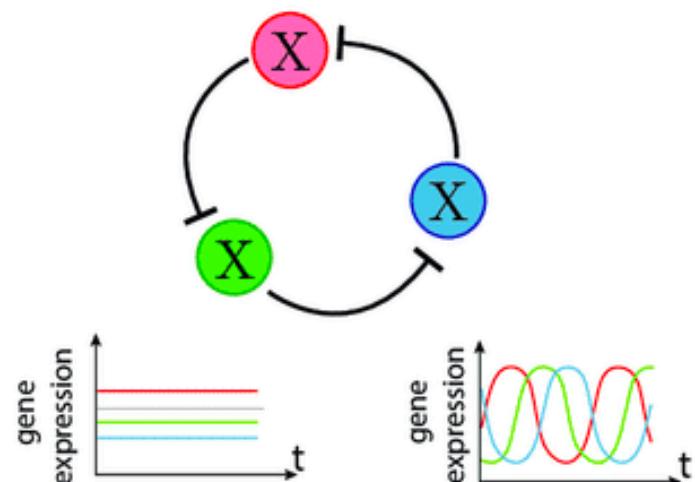
$$\frac{dm_2}{dt} = -m_2 + \frac{\alpha}{1 + p_1^n} + \alpha_0$$

$$\frac{dp_2}{dt} = -\beta(p_2 - m_2)$$

$$\frac{dm_3}{dt} = -m_3 + \frac{\alpha}{1 + p_2^n} + \alpha_0$$

$$\frac{dp_3}{dt} = -\beta(p_3 - m_3)$$

a)



b)

**GOAL: Find parameters  $[\alpha, \alpha_0, \beta, n]$  by observing only mRNA ( $m$ ), i.e., gene expression, NOT proteins ( $p$ ).**

# IDENTIFICATION OF WHOLE NETWORKS: GENE REPRESSILATOR MODEL

What we know:

- We know the model (i.e., **ODEs**).
- For given parameter values, we can always **run** a numerical integrator.
- There are **four parameters**.

# IDENTIFICATION OF WHOLE NETWORKS: GENE REPRESSILATOR MODEL

What we know:

- We know the model (i.e., **ODEs**).
- For given parameter values, we can always **run** a numerical integrator.
- There are **four parameters**.

This problem could be solved using **computational methods**, namely, derivative-free optimization (or, as previously presented, ABC).

# IDENTIFICATION OF WHOLE NETWORKS: GENE REPRESSILATOR MODEL

What we know:

- We know the model (i.e., **ODEs**).
- For given parameter values, we can always **run** a numerical integrator.
- There are **four parameters**.

This problem could be solved using **computational methods**, namely, derivative-free optimization (or, as previously presented, ABC).

For instance, we can use **population-based algorithms**.

## IDENTIFICATION OF WHOLE NETWORKS: POPULATION-BASED OPT.

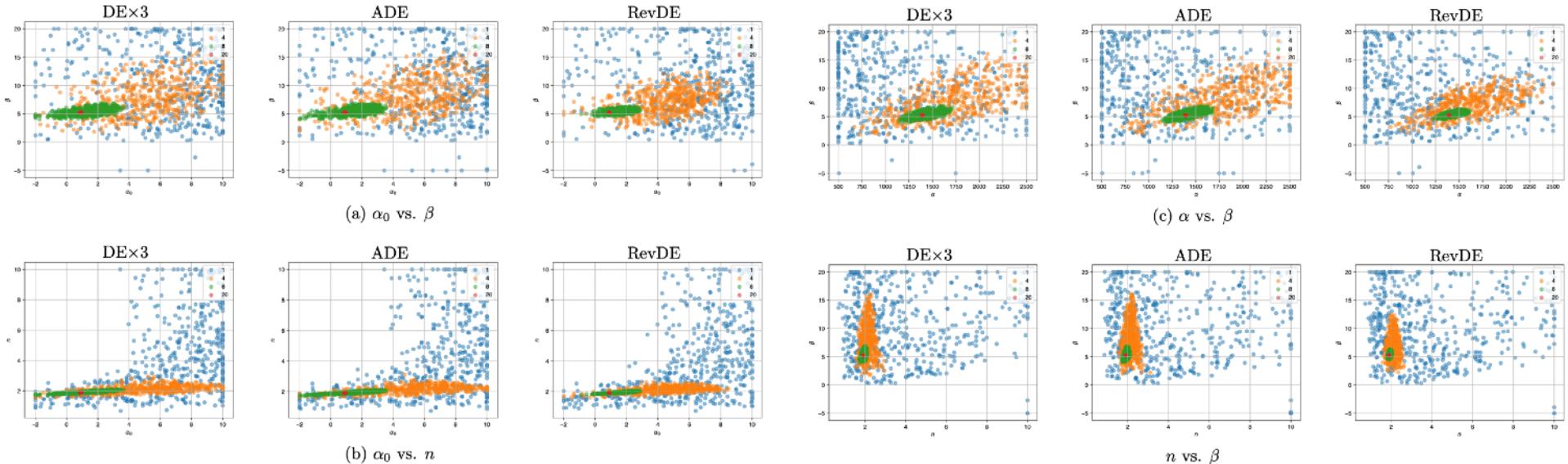
- **The key idea:** Run an algorithm multiple times in parallel and exchange information about the objective among solutions.

- **The key idea:** Run an algorithm multiple times in parallel and exchange information about the objective among solutions.
- The general scheme:
  1. **(Init)** Initialize a population of solutions,  $\mathcal{P}_t := \mathcal{P}_0$ , and evaluate.
  2. Repeat until STOP:
    - (i) **(Generate)** Generate new solutions,  $\mathcal{S}_{t+1}$ .
    - (ii) **(Evaluate)** Evaluate new solutions.
    - (iii) **(Select)** Select  $\mathcal{P}_{t+1}$  from  $\mathcal{P}_t$  and  $\mathcal{S}_{t+1}$ .

- **The key idea:** Run an algorithm multiple times in parallel and exchange information about the objective among solutions.
- The general scheme:
  1. **(Init)** Initialize a population of solutions,  $\mathcal{P}_t := \mathcal{P}_0$ , and evaluate.
  2. Repeat until STOP:
    - (i) **(Generate)** Generate new solutions,  $\mathcal{S}_{t+1}$ .  
 $x_{new} = x_1 + \gamma(x_2 - x_3)$   
**differential mutation**
    - (ii) **(Evaluate)** Evaluate new solutions.
    - (iii) **(Select)** Select  $\mathcal{P}_{t+1}$  from  $\mathcal{P}_t$  and  $\mathcal{S}_{t+1}$ .

**Select best performing candidates  
from the old population and new points.**

# IDENTIFICATION OF WHOLE NETWORKS: GENE REPRESSILATOR MODEL

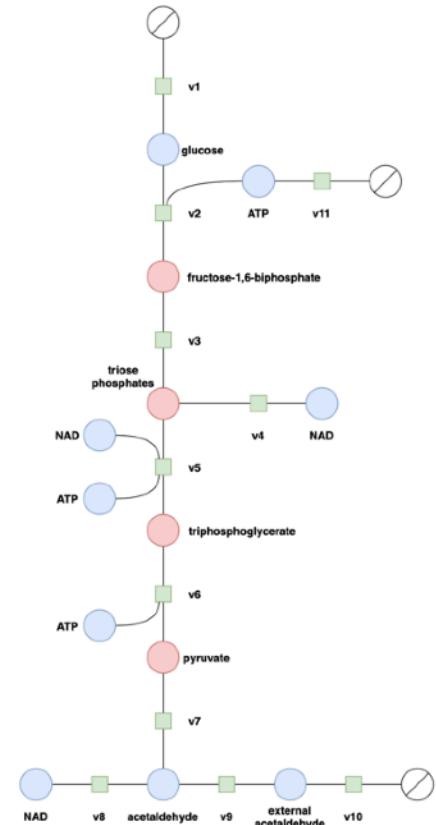


- the 1st generation
- the 4th generation
- the 8th generation
- the 20th generation

# IDENTIFICATION OF WHOLE NETWORKS: GLYCOLYSIS

Now we are ready to attack the larger problem.

We consider the problem **glycolysis** of the baker's yeast.



# IDENTIFICATION OF WHOLE NETWORKS: GLYCOLYSIS

Now we are ready to attack the larger problem.

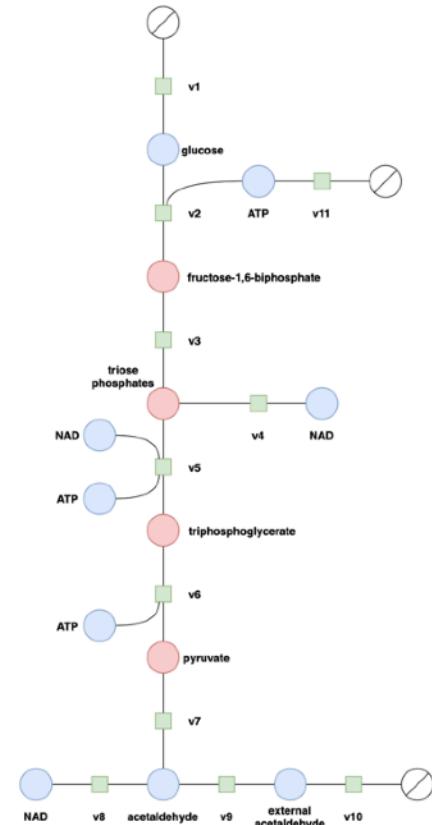
We consider the problem **glycolysis** of the baker's yeast.

The whole glycolysis is extremely complex process.

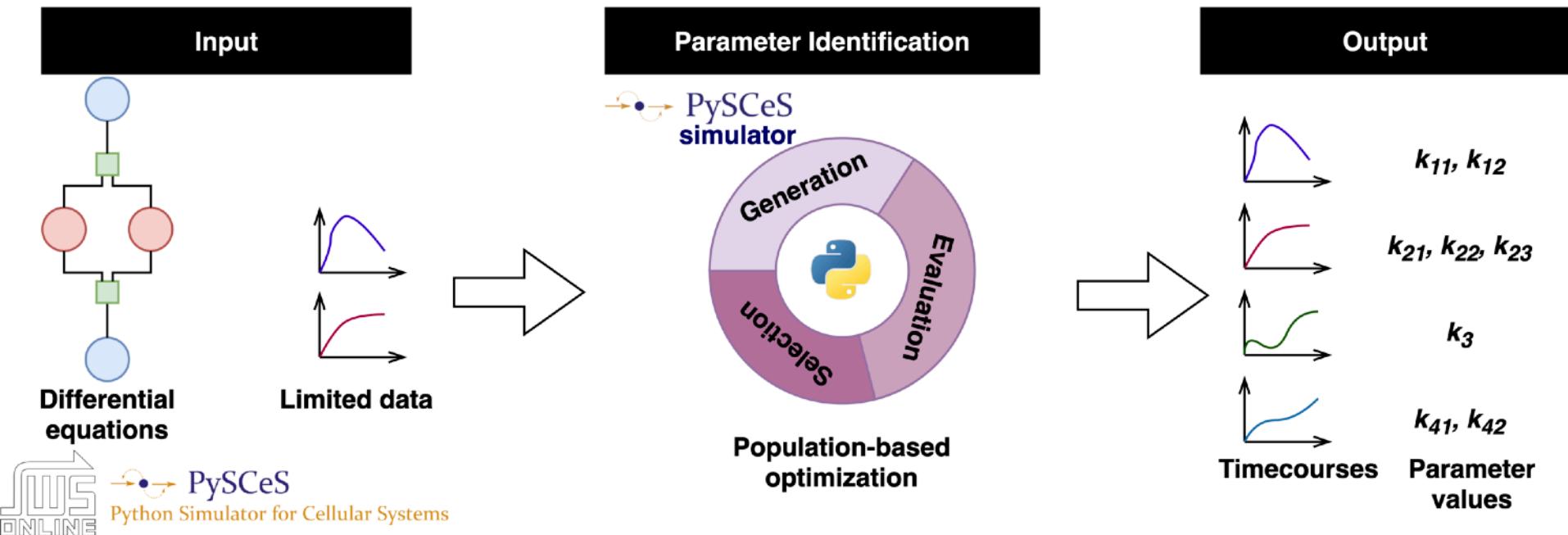
In our studies, we used a simplified model:

- 11 reactions;
- 9 metabolites;
- **18 kinetic parameters.**

We assume that **5 metabolites are observed.**



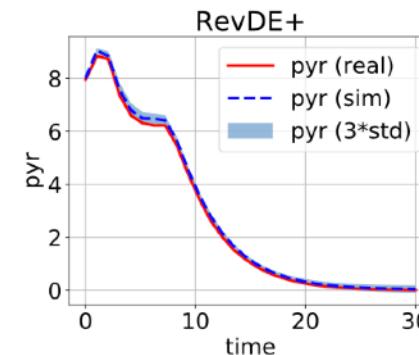
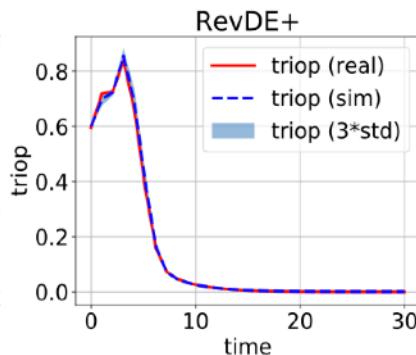
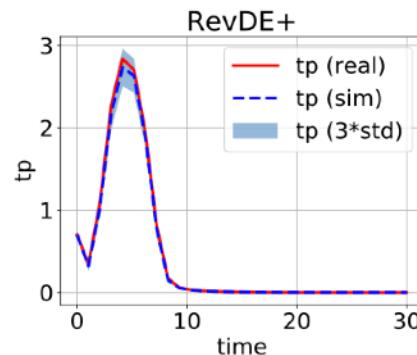
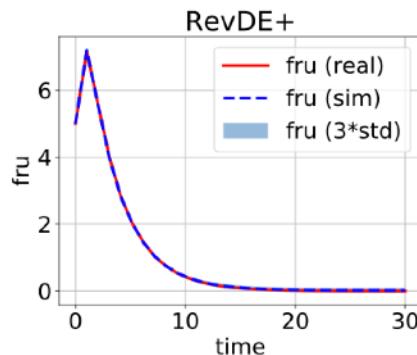
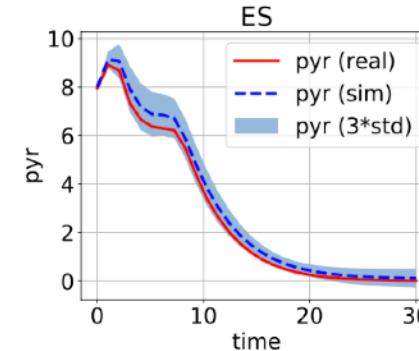
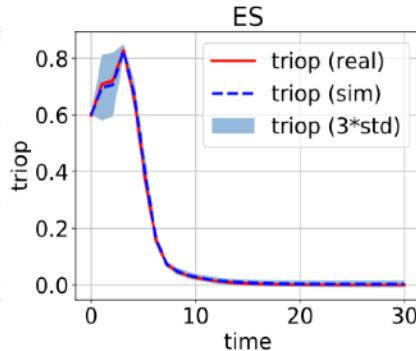
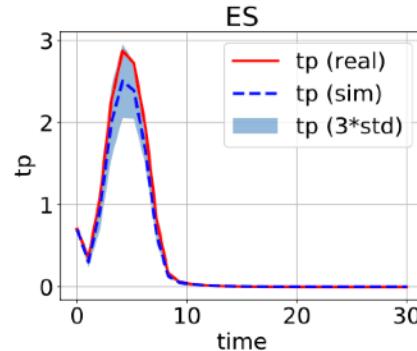
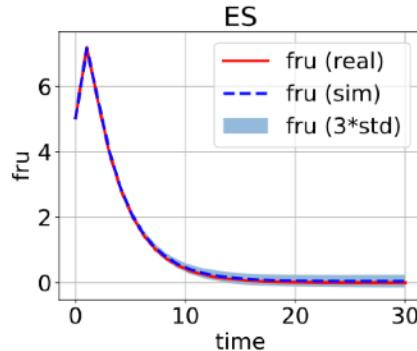
# IDENTIFICATION OF WHOLE NETWORKS: GLYCOLYSIS



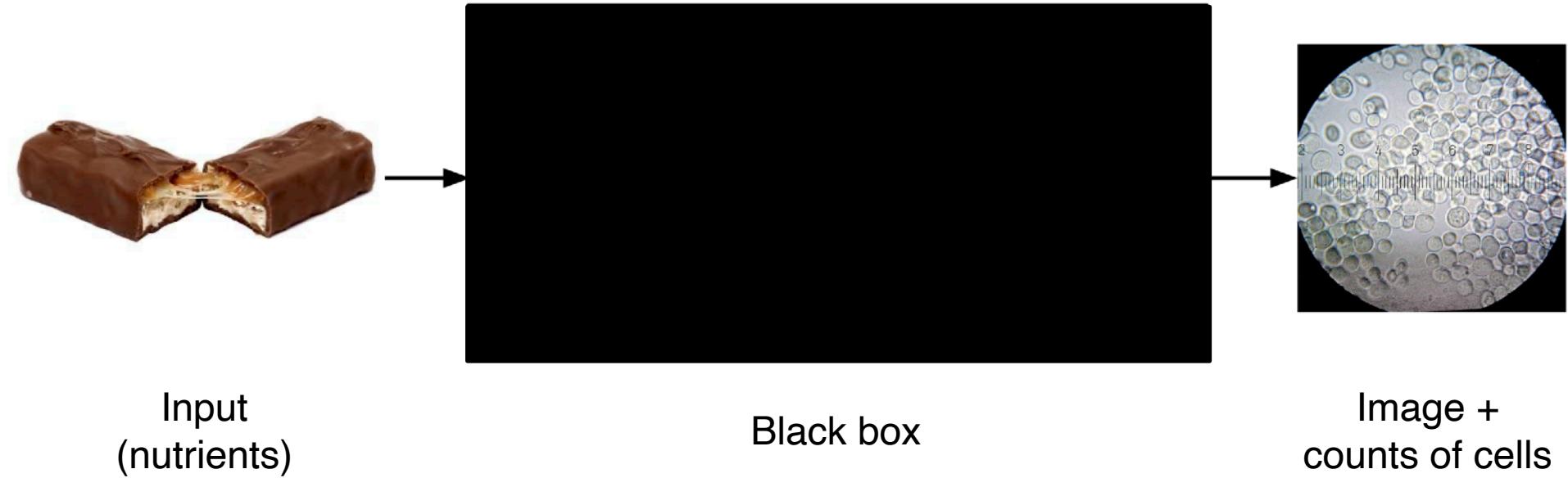
Code: <https://github.com/jmtomczak/pop4sb>

# IDENTIFICATION OF WHOLE NETWORKS: GLYCOLYSIS

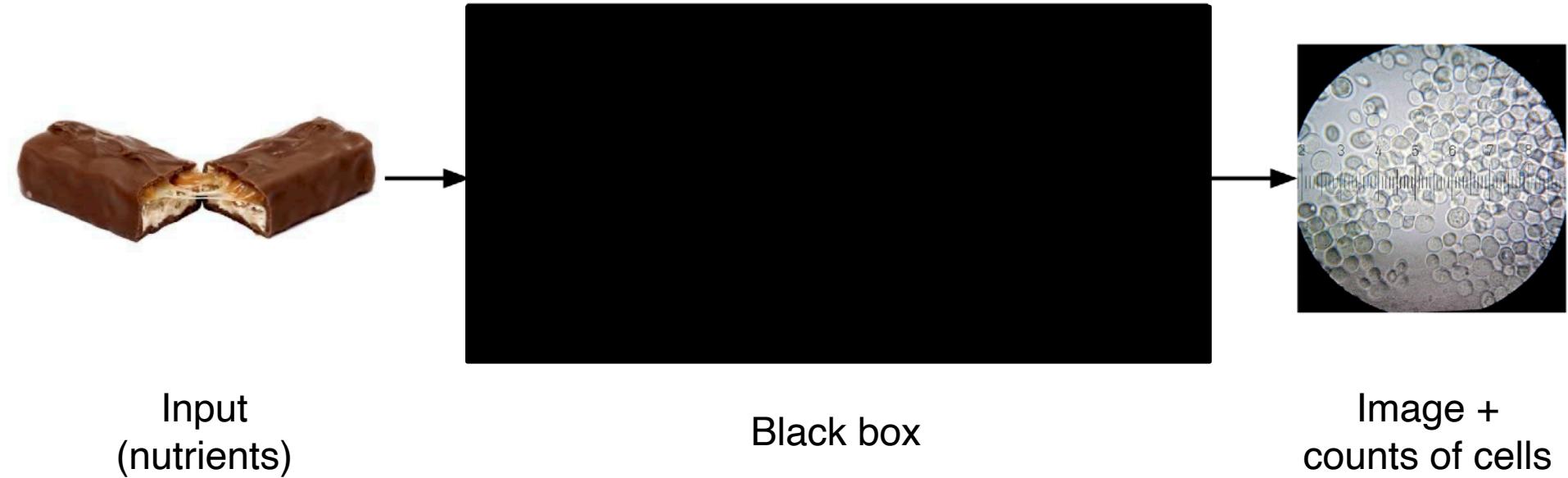
## Unobserved metabolites



# (AUTOMATIC) CELL COUNTING: THE CASE OF CANCER CELLS



# (AUTOMATIC) CELL COUNTING: THE CASE OF CANCER CELLS



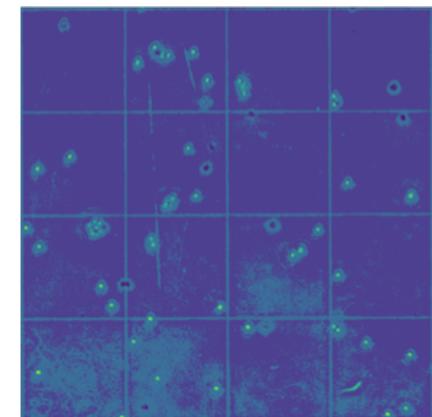
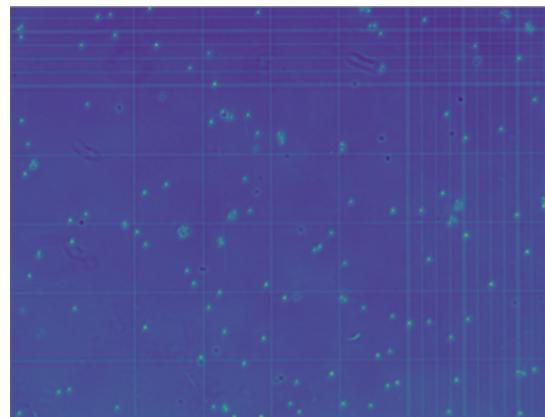
## ASSUMPTIONS:

- (i) We don't know how the cancer develops.
- (ii) We give different nutrients to determine how they influences cancer.

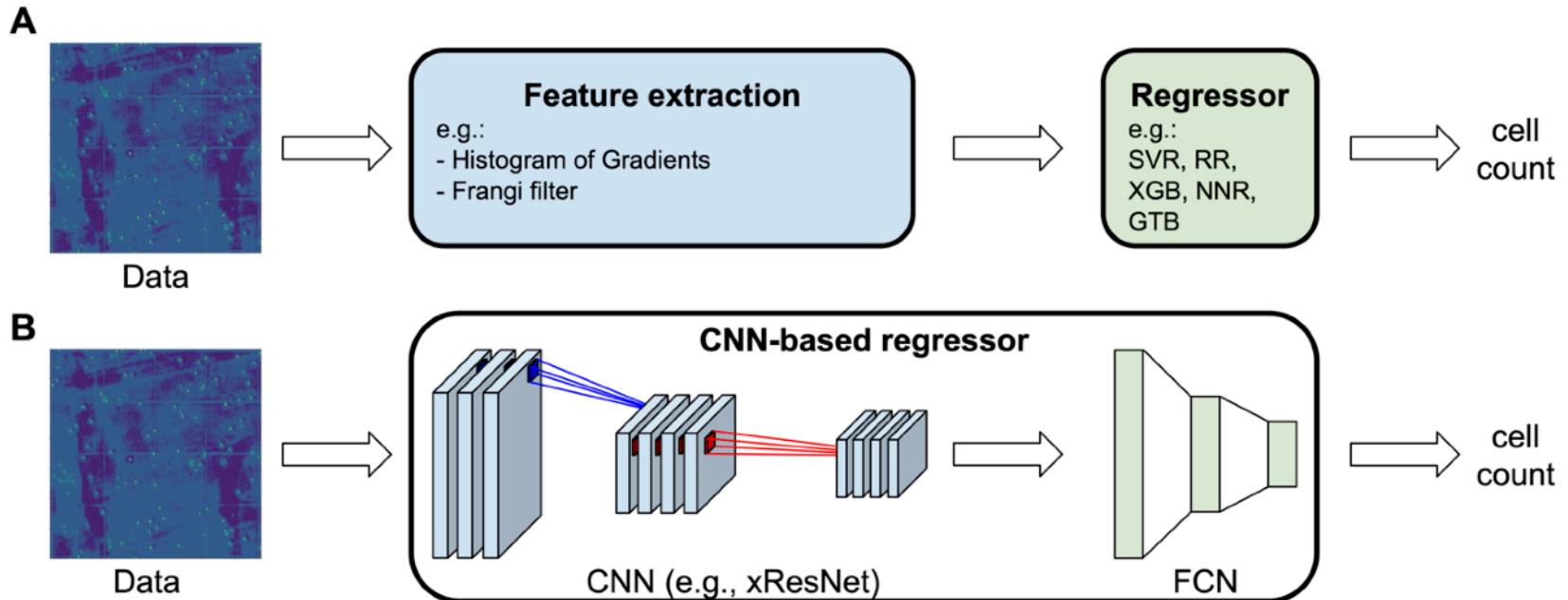
# (AUTOMATIC) CELL COUNTING: THE CASE OF CANCER CELLS

Data:

- a human osteosarcoma (U2OS) and a human leukemia (HL-60)
- 165 images (133 training, 32 test)
- 700px by 700px
- Collected at the UvA  
(led by E.W.-T.)



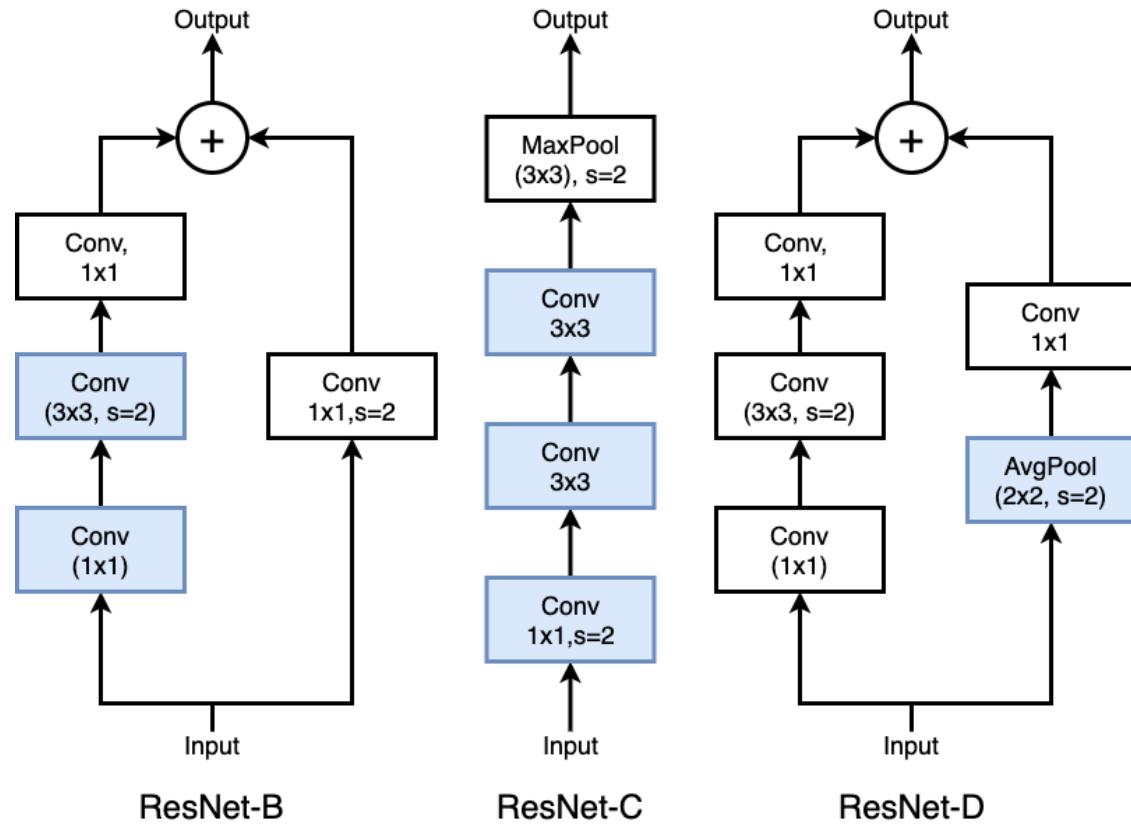
# (AUTOMATIC) CELL COUNTING: THE CASE OF CANCER CELLS



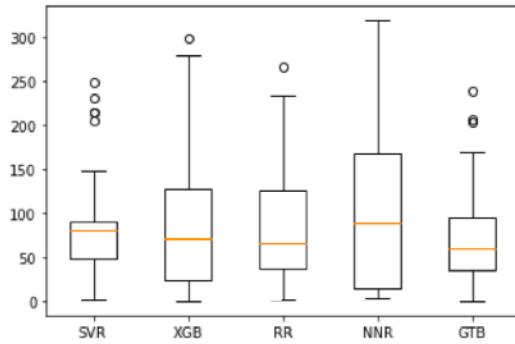
**A:** Machine learning pipeline.    **B:** Deep learning approach.

# (AUTOMATIC) CELL COUNTING: THE CASE OF CANCER CELLS

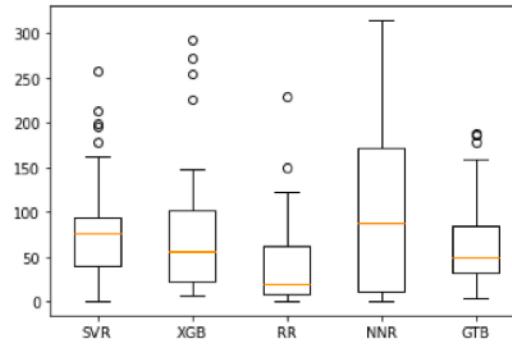
We used **xResNet**  
+ transfer learning.



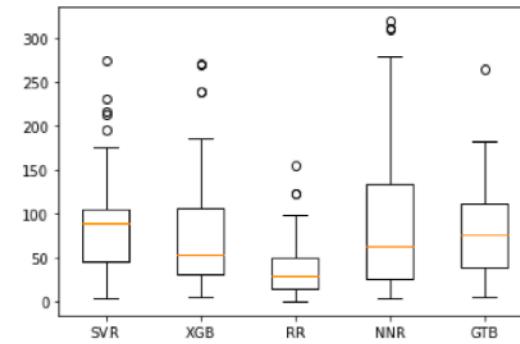
# (AUTOMATIC) CELL COUNTING: THE CASE OF CANCER CELLS



*Image*



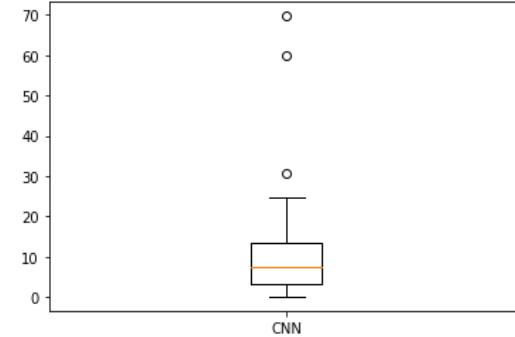
*HOG*



*Frangi*

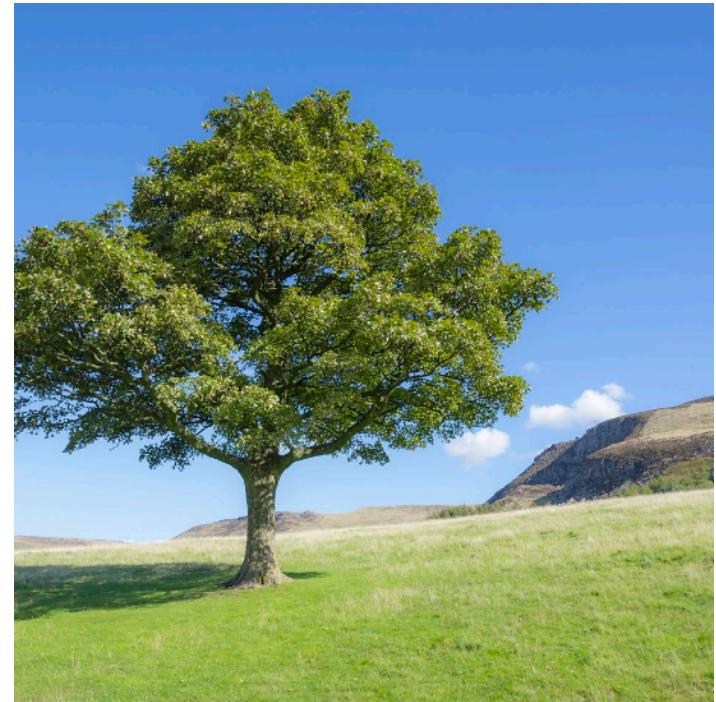
Machine learning pipeline: min. avg. error = 40

CNN w/ TL: avg. error = 33  
CNN w/ TL: avg. error = 12



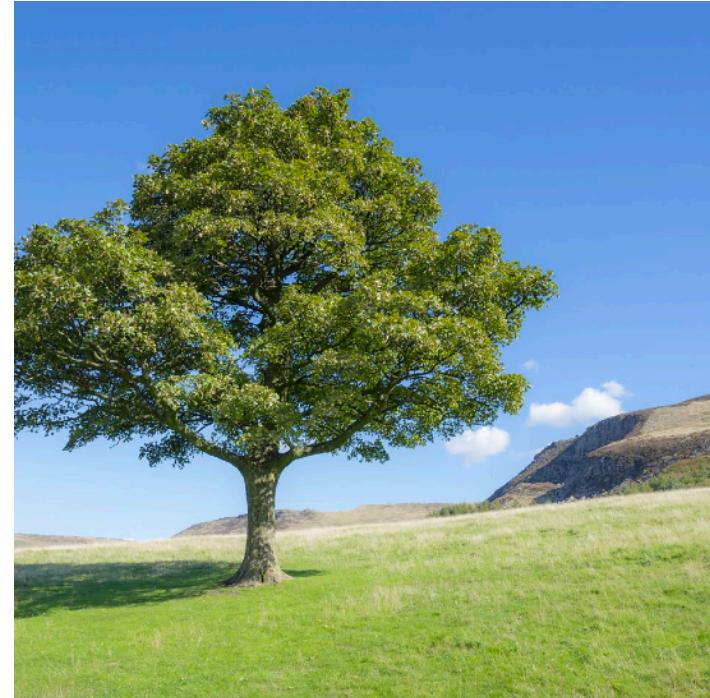
# CONCLUSION

- Life sciences are fascinating!



# CONCLUSION

- Life sciences are fascinating!
- Computational methods give a great opportunity to study our reality.
- AI-powered tools are useful from nano to macro scale.



# CONCLUSION

- Life sciences are fascinating!
- Computational methods give a great opportunity to study our reality.
- AI-powered tools are useful from nano to macro scale.
- We should always try to use as much of prior knowledge as possible.



# CONCLUSION

- Life sciences are fascinating!
- Computational methods give a great opportunity to study our reality.
- AI-powered tools are useful from nano to macro scale.
- We should always try to use as much of prior knowledge as possible.
- Deep learning is not an answer to all questions.



# THANK YOU FOR YOUR ATTENTION

Jakub M. Tomczak  
Computational Intelligence group  
Vrije Universiteit Amsterdam

**Webpage:** <https://jmtomczak.github.io/>

**Github:** <https://github.com/jmtomczak>

**Twitter:** <https://twitter.com/jmtomczak>