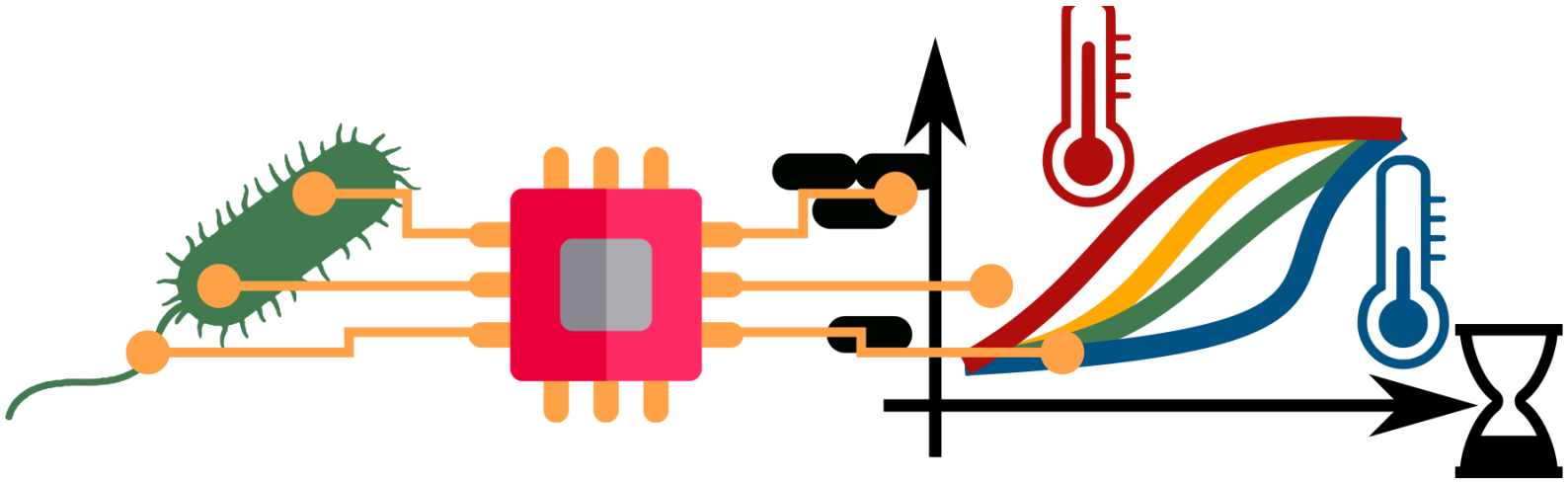
Recombinant Expression Simulation: Theoretical Background

- The BioLabSim.nrw consortium



Outline

1. Theory
    - Microbial Growth
        - » Growth Phenotypes
        - » Mathematical Models
        - » High-Throughput Measurements
        - » Biotechnological Relevance
    - Bacterial Gene Expression
        - » Expression factors
        - » Contribution to expression strength
        - » Expression Strength Prediction
2. Simulation
    - Life presentation of the simulation

Teaching Objectives

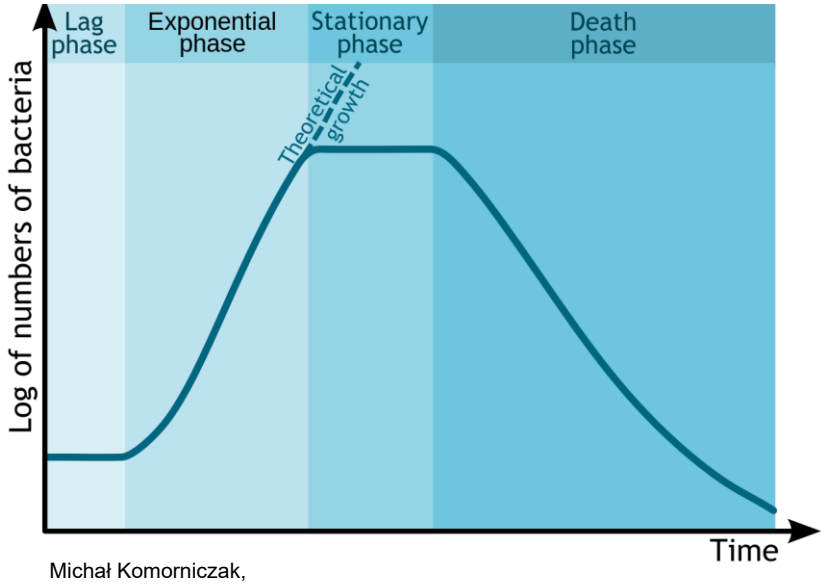
1. Distinguish a suitable growth law for data analysis.
   - Knowledge of growth types
   - Knowledge of equations and their variables

2. Identify factors of gene expression on a DNA sequence.
   - Knowledge of expression factors

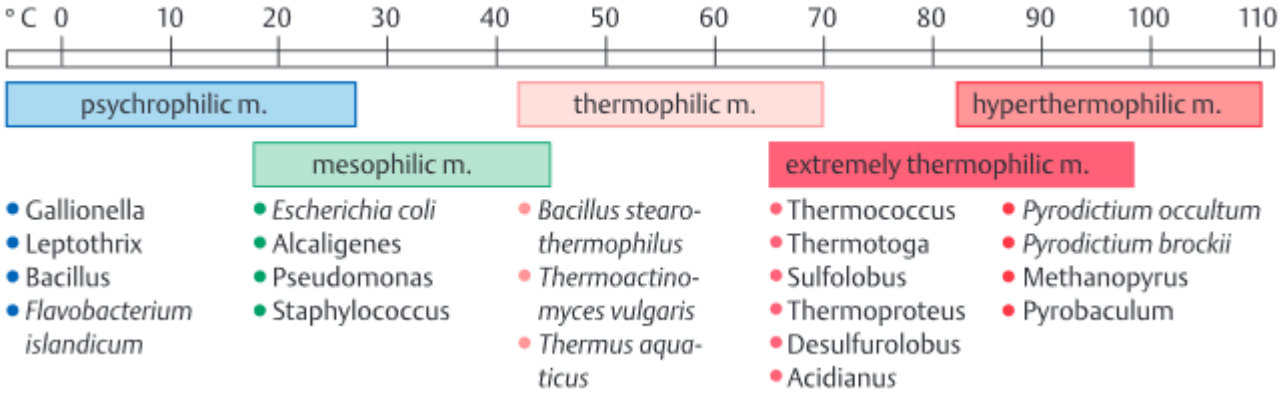
3. Judge whether a phenomenon is better described by a mechanistic or statistic model.
   - Knowledge of modelling principles

Microbial Growth Phenotypes

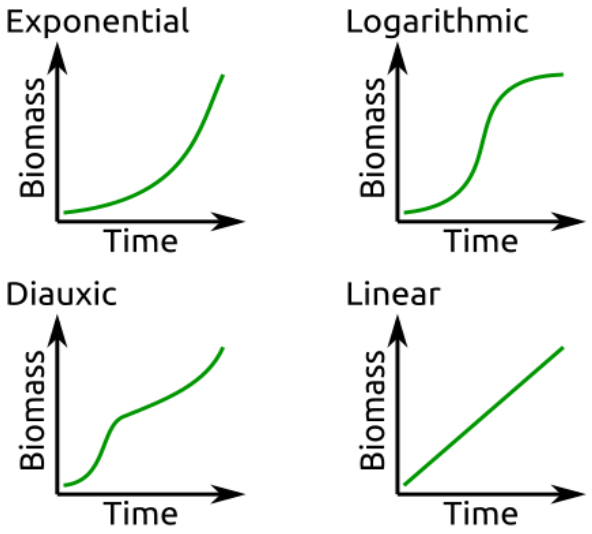
- Environmental conditions determine growth rate:
  - ➢ Temperature, Substrate, pH, etc.
- Growth is separated in different phases
- In non-optimal conditions and environmental changes microorganisms adapt their growth.



Michał Komorniczak, https://en.wikipedia.org/wiki/Bacterial_growth#/media/File:Bacterial_growth_en.svg

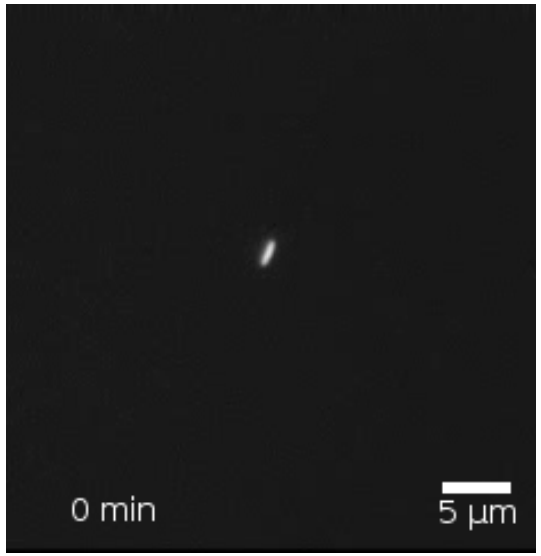


Schmid, Schmidt-Dannert, Hammelehle (2014) Biotechnology, An illustrated primer, ISBN-10: 9783527335152



Growth Models: The Exponential Function

- In optimal conditions microorganisms growth exponentially
- Many processes initially follow exponential dynamics:
  - Covid19 spread, Computing power, etc.
- Growth rate remains constant.



Stewart EJ, Madden R, Paul G, Taddei F (2005) (https://commons.wikimedia.org/wiki/File:E.coli-colony-growth.gif), https://creativecommons.org/licenses/by-sa/4.0/legalcode

Mechanistic Growth Model:

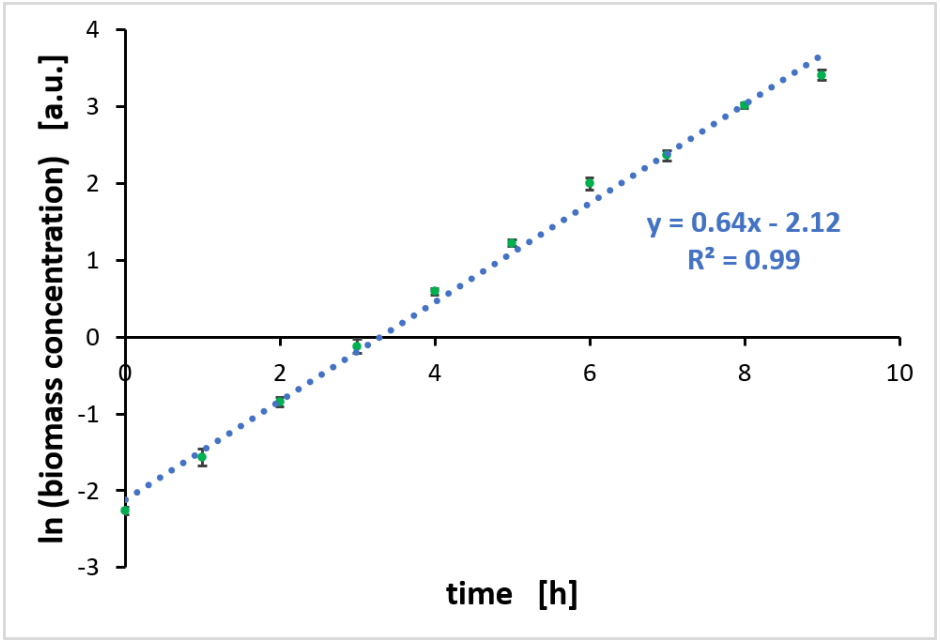
$$X \xrightarrow{\mu} 2X$$

X: Biomass, gDW/L
μ: growth rate, /h
t: time, h

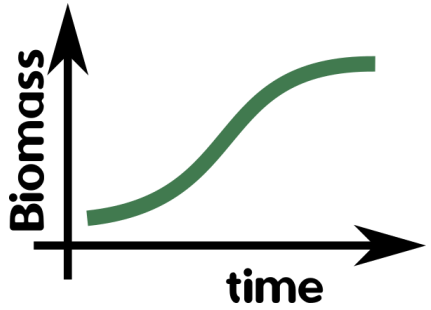
$$\frac{dX}{dt} = \mu X$$

$$X(t) = X_0 e^{\mu t}$$

$$\ln(X(t)) = \ln(X_0) + \mu(t - t_0)$$



Growth Models: The Logistic Function



- Environmental capacity limits cause transition from exponential to logistic growth.
- Capacity limits: substrates, space, waste accumulation, etc.
- Growth rate remains constant

$$\frac{dX}{dt} = \mu X\left(1 - \frac{X}{C}\right)$$

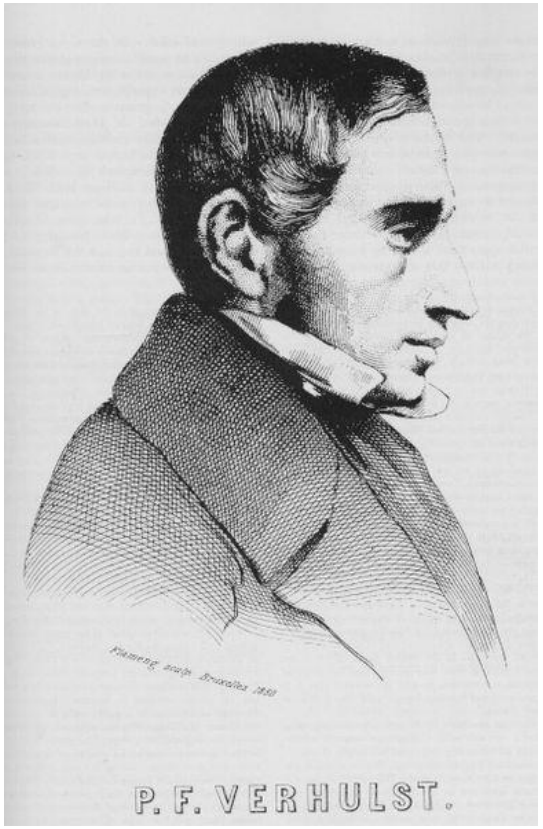
X: Biomass, gDW/L

μ: growth rate, /h

C: capacity limit, gDW/L

t: time, h

$$X(t) = \frac{C}{1 + \left(\frac{C - X_0}{X_0}\right)e^{-\mu t}}$$



source: Pierre Francois Verhulst.jpg, https://commons.wikimedia.org/w/index.php?title=File:Pierre_Francois_Verhulst.jpg&oldid=141824277

- Growth rate is estimated with parameter search to minimize error.

Growth Models: The Monod Equation

- Growth rate is variable and depends on environment, e.g. substrate level, temperature.
- Monod assumed that growth is substrate limited by enzymatic actions.
- Similar to Michaelis-Menten equation.

$$X + S \xrightarrow{\mu} 2X$$

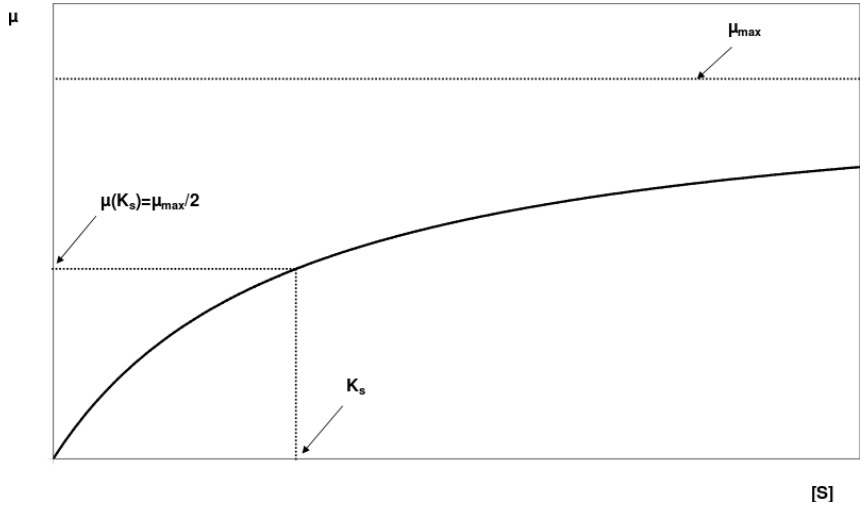
$$\mu = \frac{\mu_{max} S}{K_S + S}$$

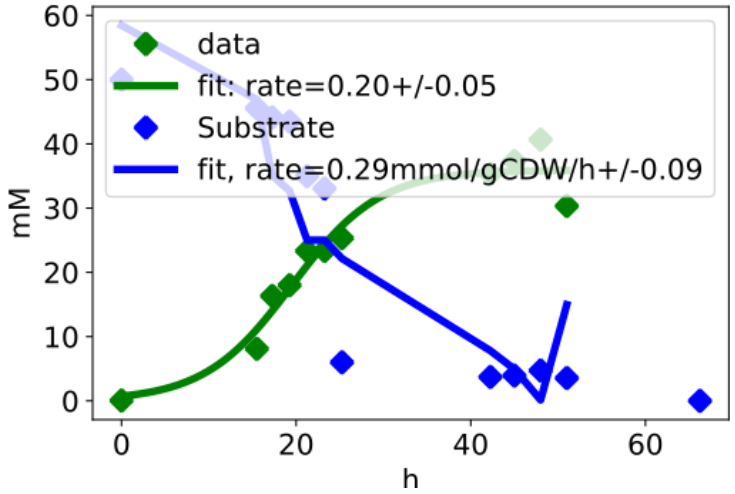
μ: growth rate, /h

μ_{max}: max rate, /h

S: substrate, mM

K_S: substrate affinity, mM



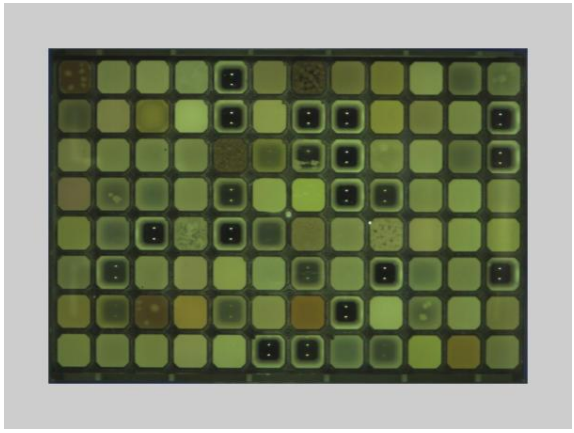


- Substrate measurements required for parameter estimation.

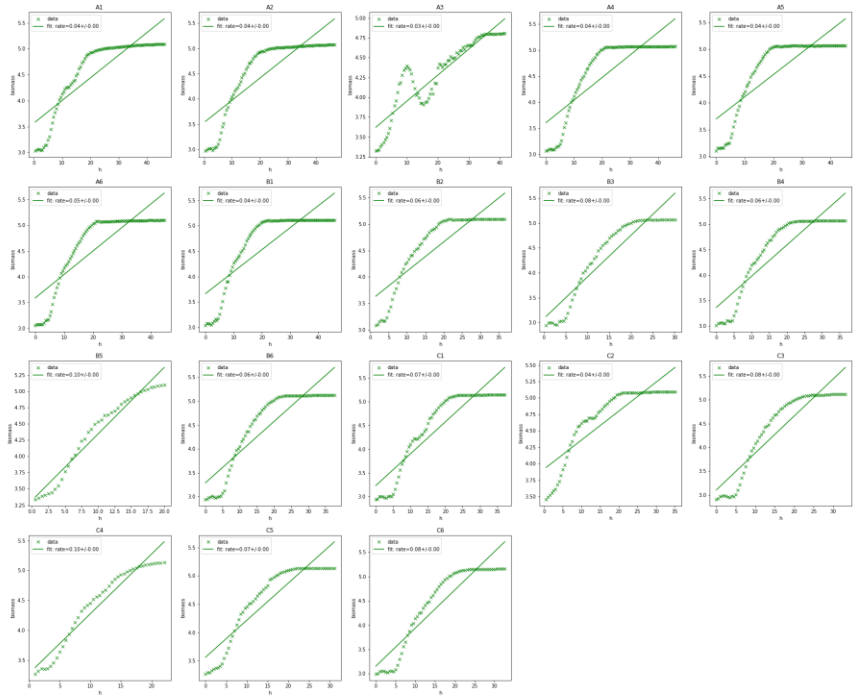
File:Monod 3.svg, https://commons.wikimedia.org/w/index.php?title=File:Monod_3.svg&oldid=427056782 (last visited April 20, 2022).

High-Throughput Measurements

- Organisms with high growth rates need to be selected based on mutations and medium.
- 96-well plate measurement of growth by Enzyscreen GrowthProfiler

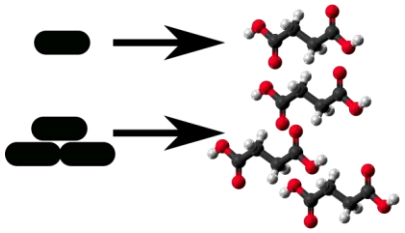


- Data export as csv-files
- Automated analysis in Jupyter Notebooks

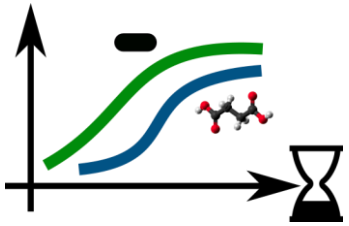


Source: https://www.enzyscreen.com/growth_profiler.htm

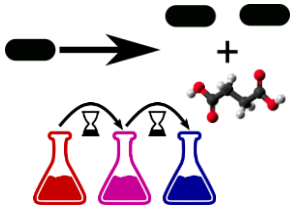
Biotechnological Relevance



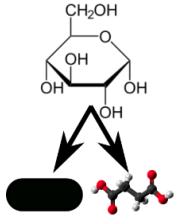
High cell densities lead to high productivity.



Target products produced in different growth phases.



Growth-coupled production enables optimization with ALE.



Efficiency tradeoff by substrate use in biomass and product.

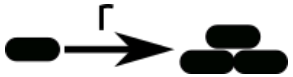
Optimal growth requires energy for heating/cooling.

Biomass waste disposal strategy for GMOs.

Growth Curve Data Analysis Example

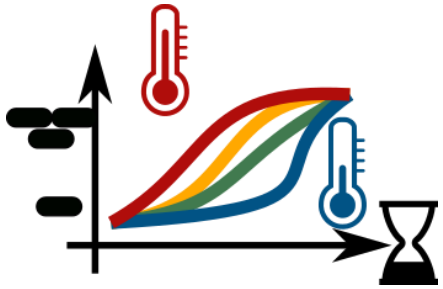
Objective

Optimal Temperature



Growth Rate & Max Biomass

Procedure



Analysis

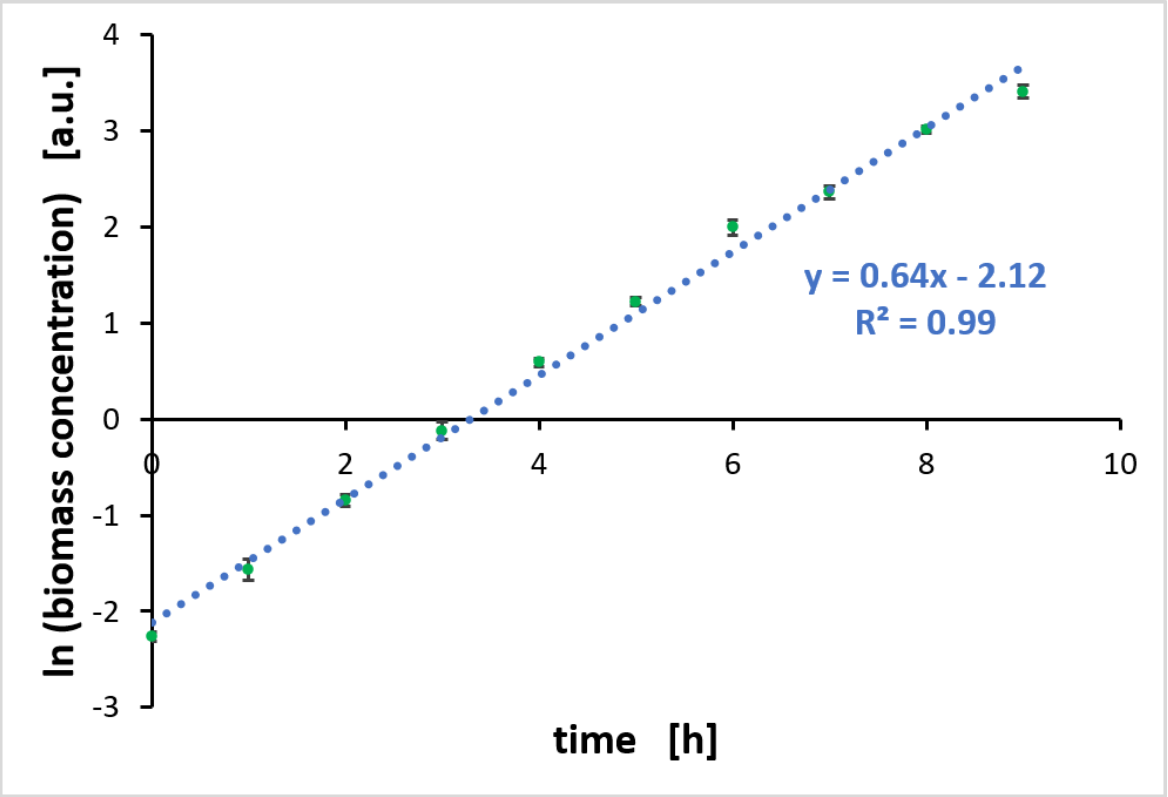
- Scrambled code-lines provided
- User rearranges line order for functional code

- Standard for quick laboratory data analysis.
- Users implement Excel linear regression.

Excel-basee Growth Rate Calculation

1. Determine mean value and standard deviation (only necessary for biological replicates)
2. Plot the Logarithm of biomass concentrations against time
3. Perform a linear regression with the linear range

→ The slope of the regression line is equal to the growth rate



Example cultivation from BioLabSim at 35 °C.
Shown is the mean value and the standard deviation of three biological replicates.

Python-based Growth Rate Calculation

1. Rearrange code lines

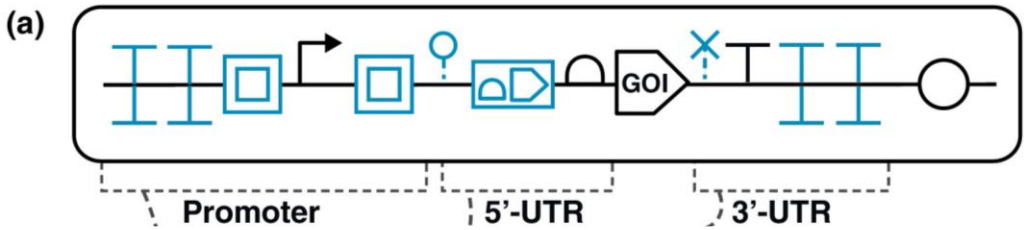
```
# Rearrange the correct code sequence for plotting the growth curves with the logarithm of biomass.
Time, Biomass = my_data[:,0], my_data[:,1:]
DataFile = 'Strain_characterization_1.csv'
LnBiomass = np.log(Biomass)
plt.scatter(Time, X, label=Exp) for Exp,X in enumerate(LnBiomass.T)]
plt.legend(bbox_to_anchor=(1.05, 1), loc='upper left')
my_data = np.genfromtxt(DataFile, delimiter=',', skip_header=1)
```

2. Extract the highest growth slope for ln(biomass)
3. Extract the range of linear slope for ln(biomass)
4. Rearrange code lines for linear regression and max biomass extraction

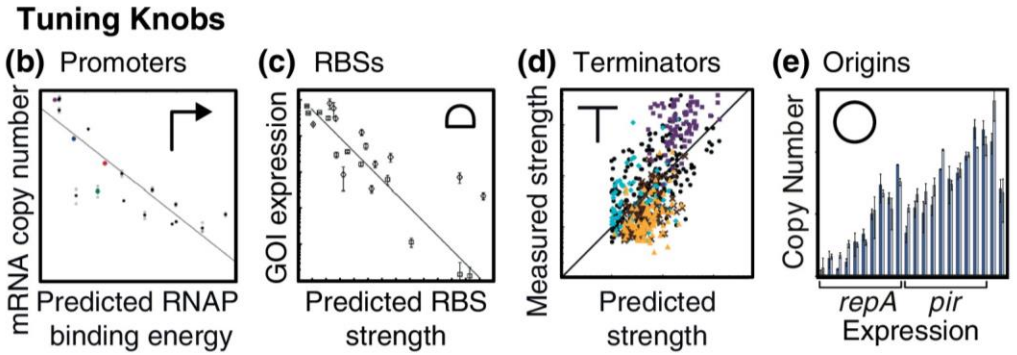
Outline

1. Theory
    - Microbial Growth
        - » Growth Phenotypes
        - » Mathematical Models
        - » High-Throughput Measurements
        - » Biotechnological Relevance
    - **Bacterial Gene Expression**
        - » Expression factors
        - » Contribution to expression strength
        - » Expression Strength Prediction
2. Simulation
    - Life presentation of the simulation

Bacterial Gene Expression



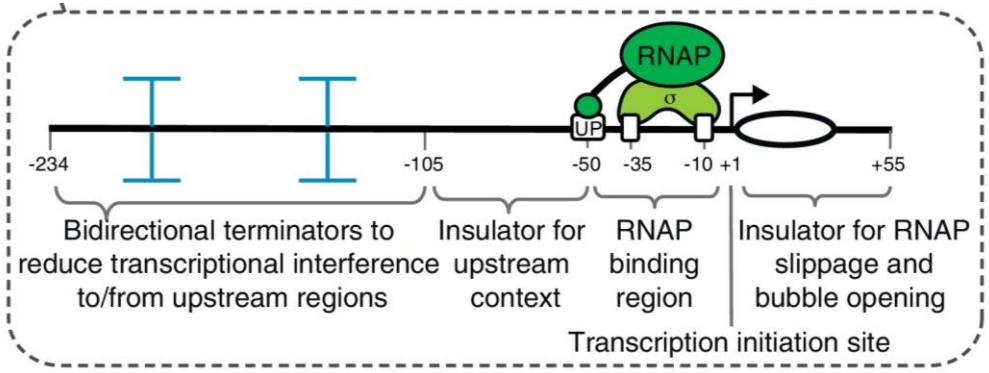
- Three levels of gene expression regulation:
  - Transcription initiation by the promoter
  - Translation initiation by the 5' UTR
  - mRNA stability by the 3' UTR

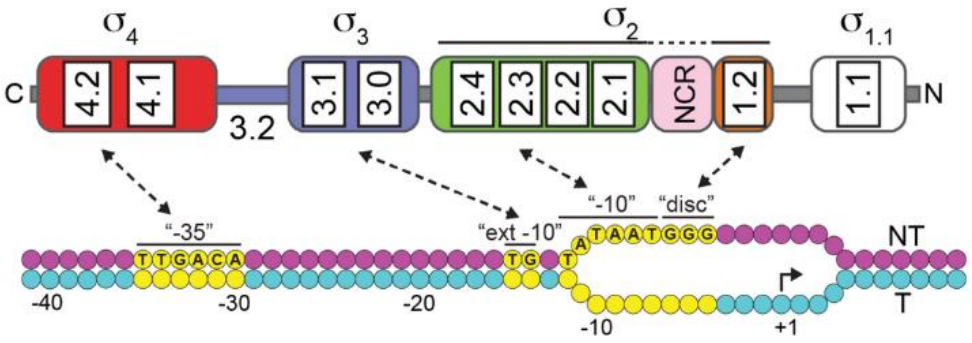


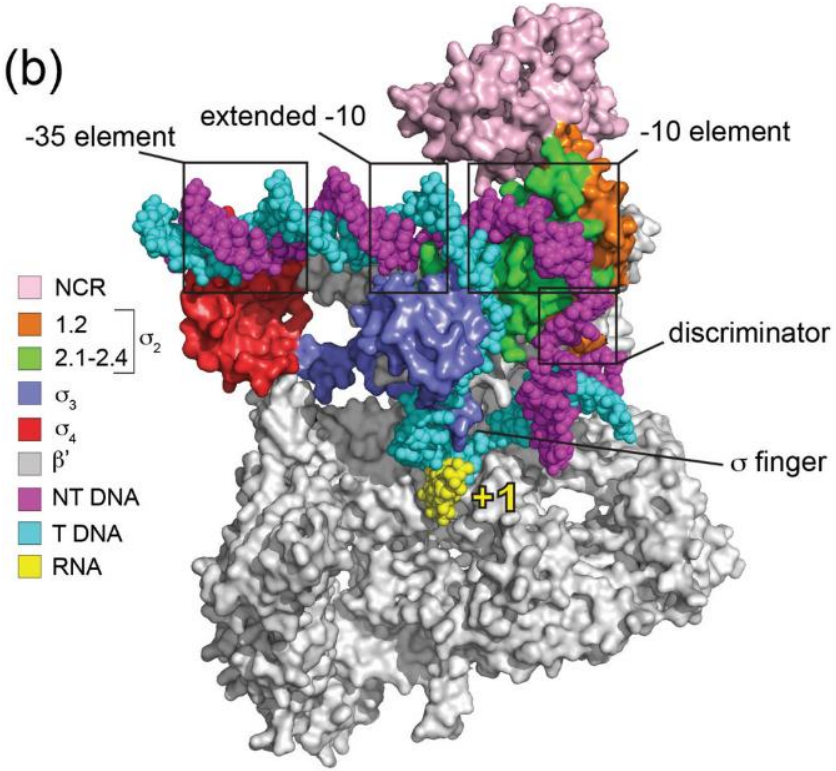
Nielsen, A., Voigt, C. (2013).
Advances in genetic circuit design: Novel biochemistries, deep part mining, and precision gene expression.
Current Opinion in Chemical Biology, *17*(6), 878–892. doi: 10.1016/j.cbpa.2013.10.003

Sigma Factors and Sequence Boxes

- The RNA-polymerase function is extending a nucleotide sequence based on homology.
- A σ-factor guides the RNA-polymerase to the promoter of a gene.
- Different σ-factors have different recognition sequences and expression efficiencies.
- A useful biotechnical used σ–factor in *E. coli* and *P. putida* is σ70.
- Important σ70 recognition sites are the boxes:
  - -35: TTGACA
  - -10: TATAAT

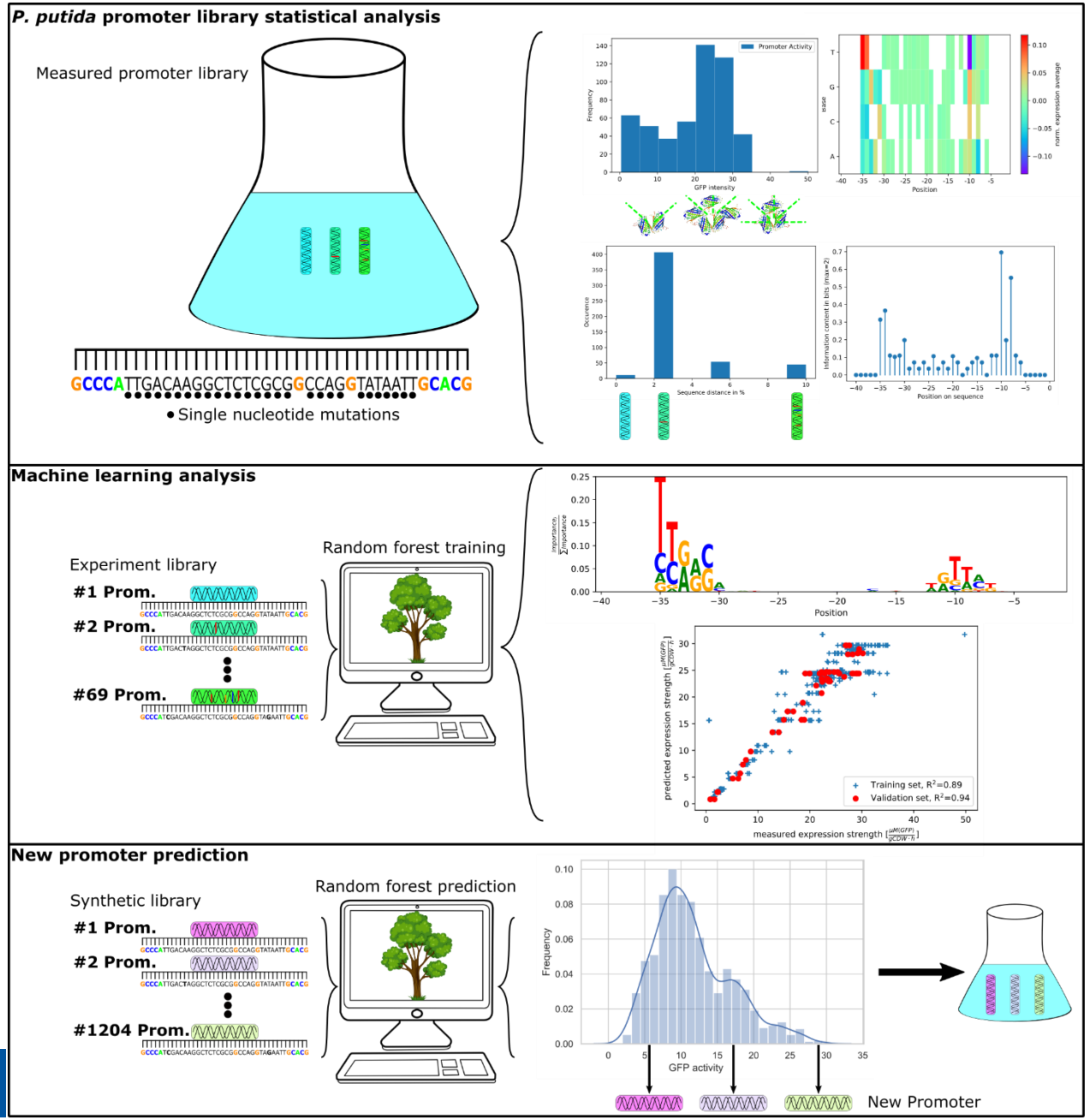






M.S. Paget (2015). Bacterial Sigma Factors and Anti-Sigma Factors: Structure, Function and Distribution *Biomolecules*, 5(3), 1245–65. doi: 10.3390/biom5031245

The Random Forest – Applications

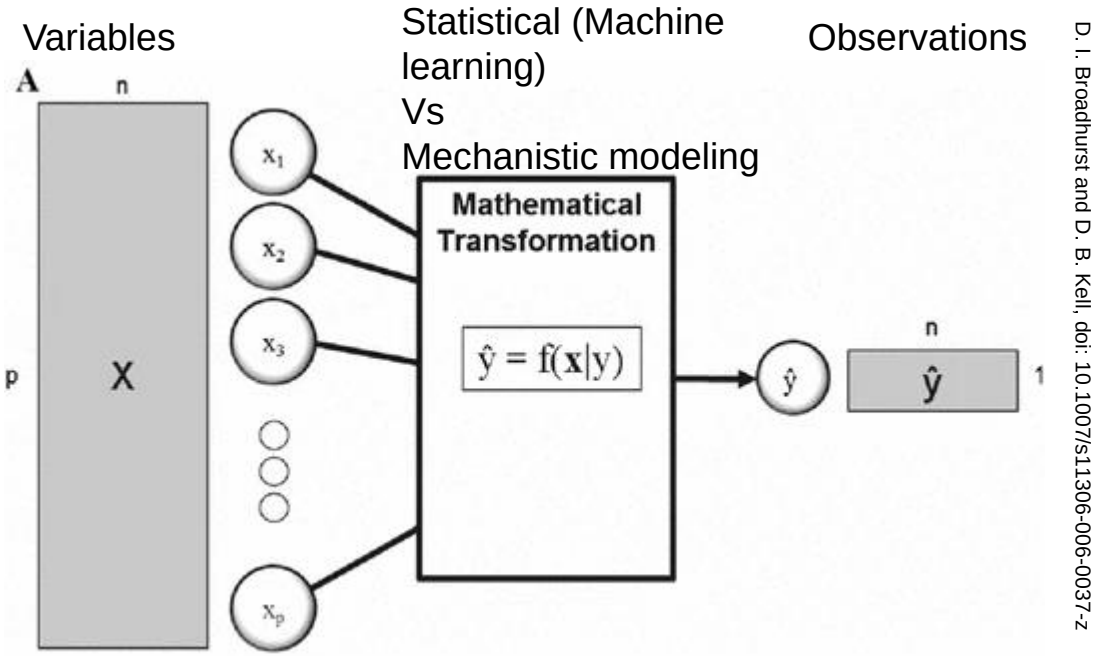


Promoter Library Analysis

1. Promoter library with mutations in promoter sequence and GFP activity.
2. RF-Learning with sequence as binary input, and additional variables, e.g., GC-content.
3. Output is GFP activity.
4. Prediction quality control with R^2 and RMSE
5. Variable importance: nucleotide contribution to the prediction.
6. Identification of novel strong promoters.

U. Liebal et al. (2021) Insight to Gene Expression From Promoter Libraries With the Machine Learning Workflow Exp2Ipynb. *Front. Bioinform.*, doi: 10.3389/fbinf.2021.747428

Principles of Regression



D. I. Broadhurst and D. B. Kell, doi: 10.1007/s11306-006-0037-z

Regression:

- **Prediction** of continuous observations with variables.

Mathematical transformation:

- **Statistical modeling** with generic formula (machine learning, AI).
- **Mechanistic modeling** with biological interactions and laws.

Examples

Statistical Model

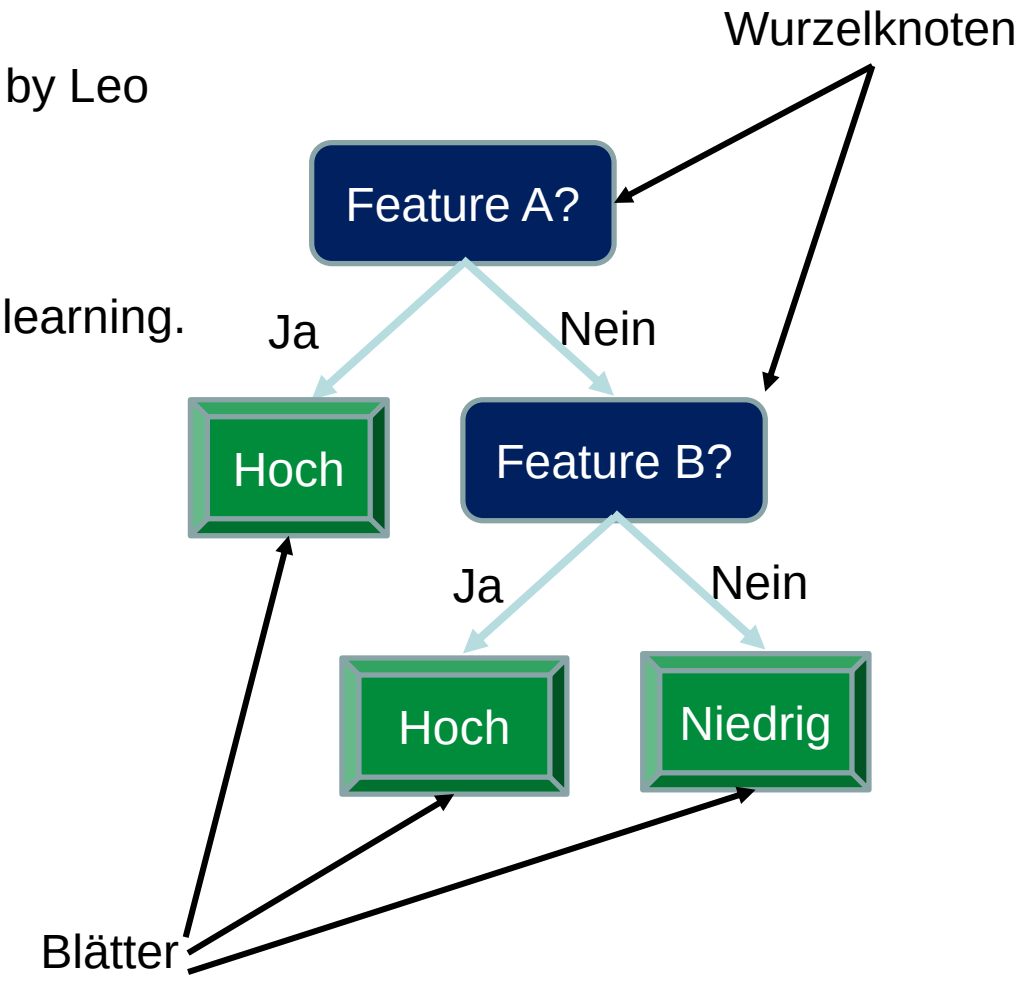
- X: Sequence Library with mutations in promoter region
- Y: Expression rate, measured with GFP
- Regression model: Random Forest

Mechanistic Modeling

- X: time of culture growth
- Y: Biomass
- Regression model: Logistic Function

Promoter Activity Prediction with Random Forest

- Automated method for classification and regression.
- Most famous as ‘CART: Classification And Regression Tree’ by Leo Breiman, 1984.
- Visually ordered overview to formal rules.
- Binary rules at roots lead to leave-outcomes.
- A decision tree can be generated automatically via machine learning.



Construction of a decision tree

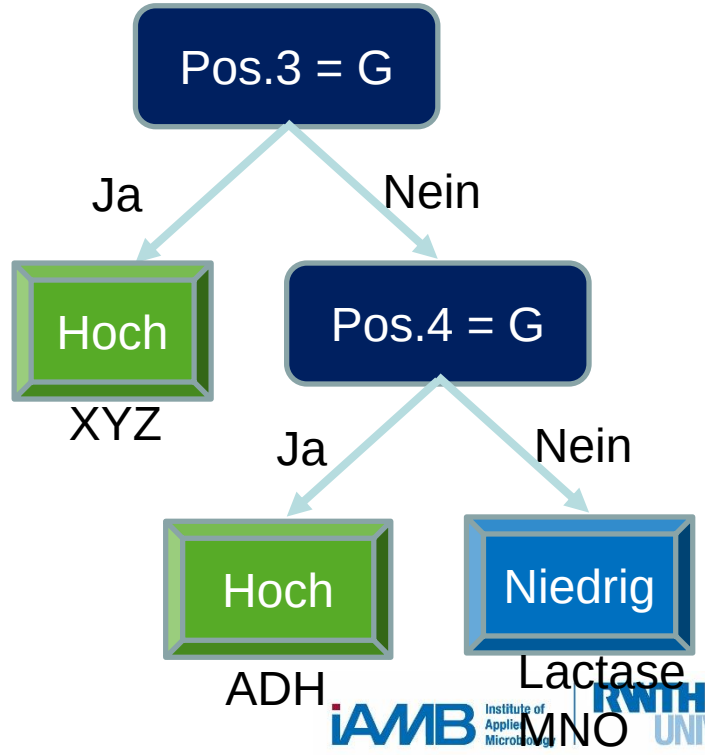
- Step 1: determine starting root

Effective separation at position 3.
If: Pos3 = ‘G’ then expression *High*.
Else: different level of expression → no clear decisions

- Step 2: data separation

If: Pos.4 = ’G’ then expression *High.*
Else: expression *Low*.

Name	Position 1	2	3	4	Expr. Level
Lactase	A	G	T	A	Niedrig
ADH	A	C	T	G	Hoch
XYZ	A	T	G	C	Hoch
MNO	A	G	T	A	Niedrig



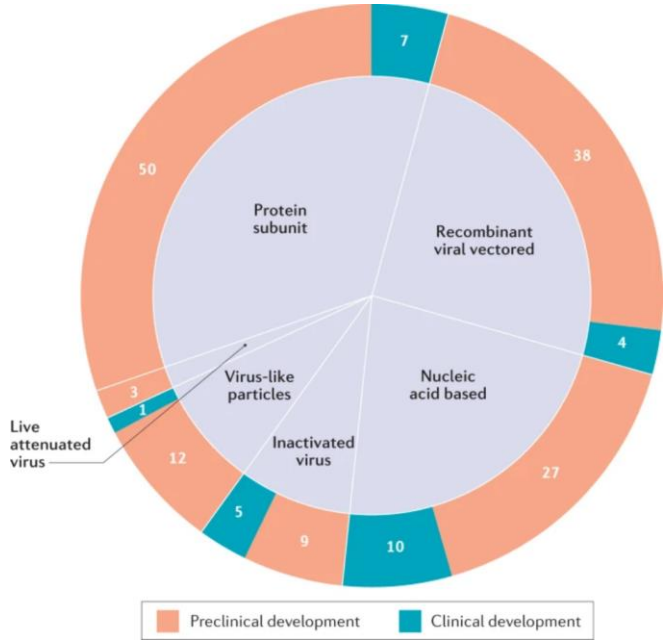
The expression of all sequences can be determined following the roots to the leafes.

Outline

1. Theory
    - Microbial Growth
        - » Growth Phenotypes
        - » Mathematical Models
        - » High-Throughput Measurements
        - » Biotechnological Relevance
    - Bacterial Gene Expression
        - » Expression factors
        - » Contribution to expression strength
        - » Expression Strength Prediction
2. **Simulation**
    - Life presentation of the simulation

Simulation of Recombinant Expression

Develop an expression system for recombinant production of COVID19 protein subunits for vaccination.

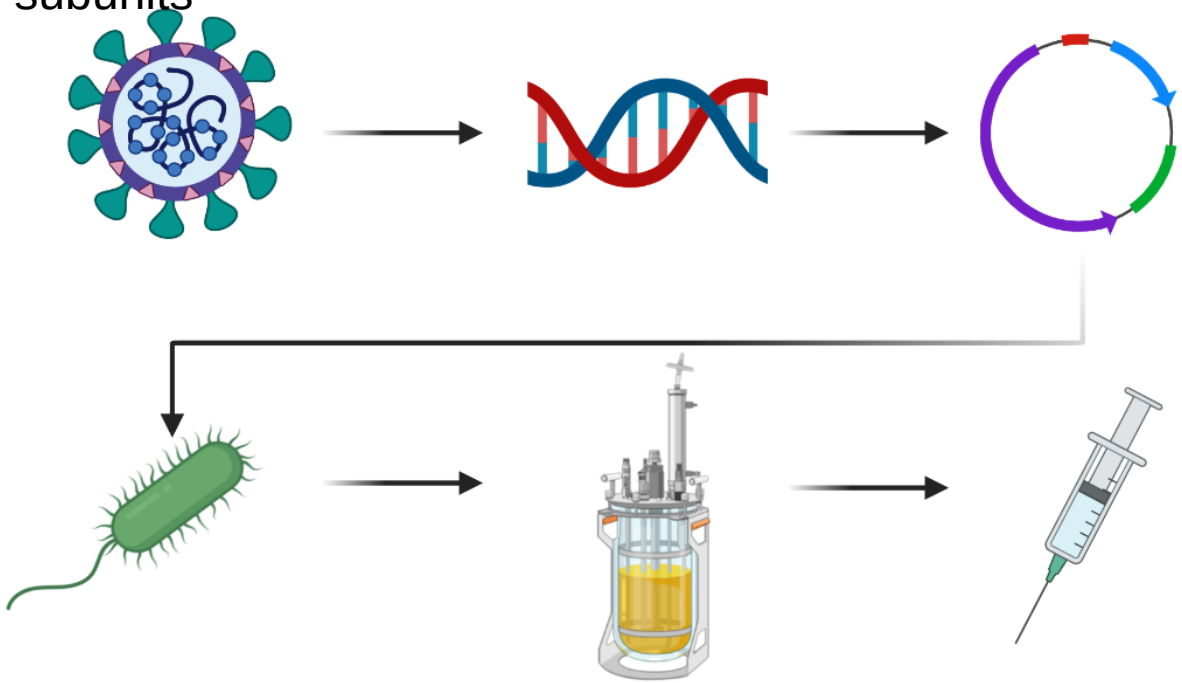


Source: Jeyanathan et al., 2020, https://doi.org/10.1038/s41577-020-00434-6

The majority of vaccine (candidates) is based on recombinant expression of protein subunits.

Your tasks:

1. Characterize two strains for their growth properties
2. Generate promoter sequences
3. Clone promoter sequences into strain
4. Measure expression rate
5. Conduct batch fermentation to harvest protein subunits



Authors and License

Check the BioLabSim.nrw members at:

https://git.rwth-aachen.de/ulf.liebal/biolabsim/-/blob/91dd40b2f40e6a24263af96c2d5b9817733f4a25/AUTHORS.md