

# lec1.tex

Henrik Linder

August 29, 2022

In ML lab we work with data from titanic.

E.coli is the base bacterium in biology, about  $1\text{ }\mu\text{m}$  in diameter. Cell division every 30 mins. Ca 5 million base pairs in its DNA. Error rate about  $1 / 10^{-9}$

ABout  $10^9$  base pairs in humans,  $10^{10}$  in trees.

About the same number of genes in most organisms, but different genome length.

E.coli will sense sugar close by, it will build a flagellum that spins at 100 Hz and propells forward to eat the sugar with a pump (mouth). They are all proteins, about 1 nm. Proteins made from about 300 aminoacids (AA). About 4500 protein types, about 4500 genes. (In E.coli, that is) about  $3 \cdot 10^6$  proteins per cell.  $10 - 10^5$  protein copy numbers.

Cells change protein composition based on the environment.

They have protein receptors on the membrane, a signal the goes into the cell. EM-fields are useless on this scale because the ions tend to screen the radiation.

How do cells make proteins?

Proteins are encoded in genes.

(Whats the specific sequence of 300 AAs that define the 3D-shape of the protein.)

RNA - polymerase (RNAP) will bind to the DNA, and transcribe to messenger (mRNA).

Ribosomes then translate the mRNA and make a protein.

Cells regulate transcription to change the protein composition.

Cells use special proteins th regulate genes. These are called transcriptions Factors (TFs), about 300 of them.

External signals from the environment: S1, S2, S3, get transcribed to 300 TFs: TF1, TF2, ..., TF300 these proteins then sit on the genes and affect the expression of these genes. It sits on the DNA and makes it more or less likely for the RNAP to attach. (Activator or repressor.)

TFs can attach to eachother and make complicated expressions, (AND/OR/XOR conditionals)

Shorthand: TFy increases Y: TFy  $\rightarrow$  Y Shorthand: TFy decreases Y: TFy  $\rightarrow$   $\neg$  Y

Switch TF:  $\mu$ s Find binding site: 1-10 s transcribe: about minute hours to get full expression.

Gene regulatory network: genes can affect eachother

4500 nodes,  $10^4$  links, sparse network.

We will make an ER-network.

The epigenetics is an unsolved problem, even though the human genome is done.

## **1 From book, ch 1**

Another observation is that the number of regulatory genes for prokaryotes increases as the square of the number of genes that should be regulated [488, 489]. In other words, doubling the size of a system requires four times more regulators. Thus the regulatory network of a living cell is a well-integrated system and not modular in any simple sense of this word (see Fig. 1.8).