# “Life” $\leftrightarrow$ Entropy

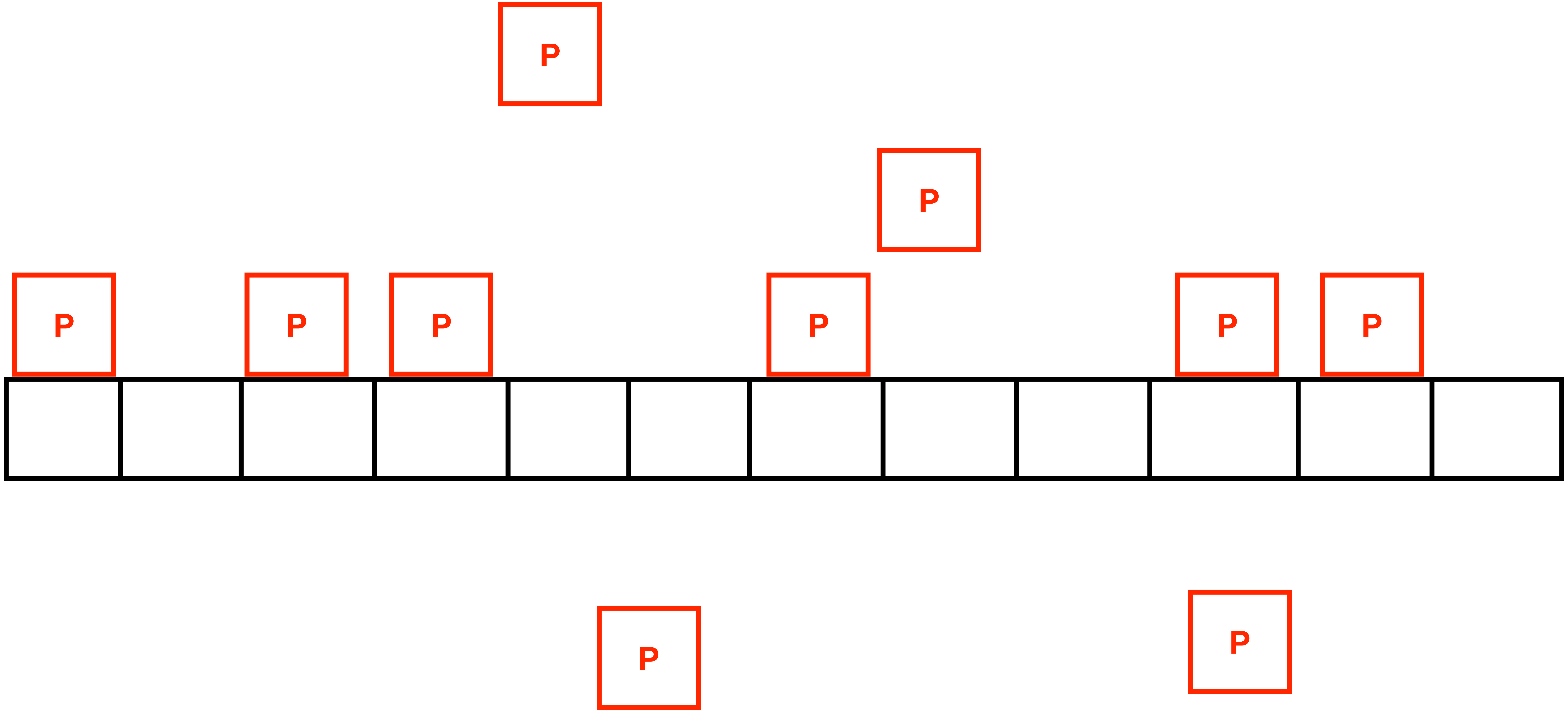
**Diversity/Variability**

**Flexibility/order**

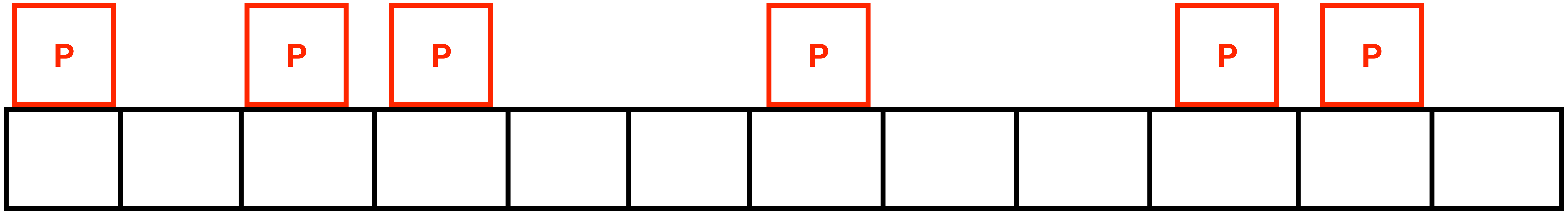
**Information**

**Evolution (Arrow of time!)**

Consider a bunch of proteins near DNA; can you predict what fraction of the DNA would be covered by proteins?

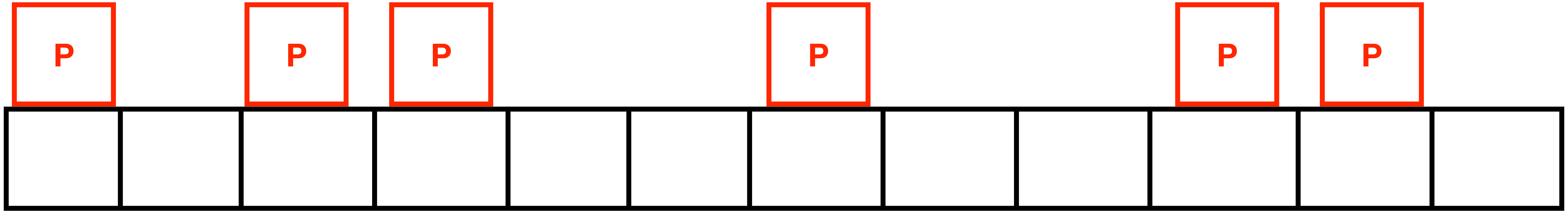


Every time a protein binds, it gains an energy, say $-\epsilon$




What would that imply?

## If all possible N sites are bound by proteins, the system would gain an energy $-\epsilon N$




So, that would be the prediction: DNA is fully covered

# But that is now what you see if you do a measurement!

**Only a fraction (say, 70%) of the sites are covered by proteins!**

(Even in vitro! Where there is no complexity of in vivo)

**Only a fraction (say, 70%) of the sites are covered by proteins!**

**Why?**

# Can we predict the fraction?

# Idea: what matters is not just the energy.

Thermal energy leads to many different bound/unbound states

Calculating entropy is the way to “count” these many different configurations

# What is entropy?

# How do we know something is “ordered”?

# Imagine visiting library every day, and taking the photo of the book shelf

What would you call as an “ordered” book shelf or disordered book shelf?

If the book shelf was perfectly ordered, you will always see only one arrangement of books

If you see different arrangement photos different days, things are less ordered. More entropy

**Entropy is proportional to the number of different arrangements seen!**

$$\text{Entropy, } S \propto \ln W$$

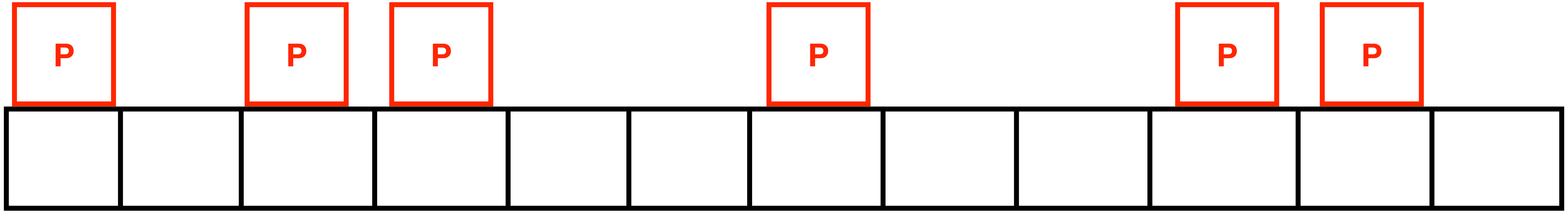
$$W = \text{Number or arrangments}$$

(W, also called, number of micro-states)

# Entropy, $S = k_B \ln W$

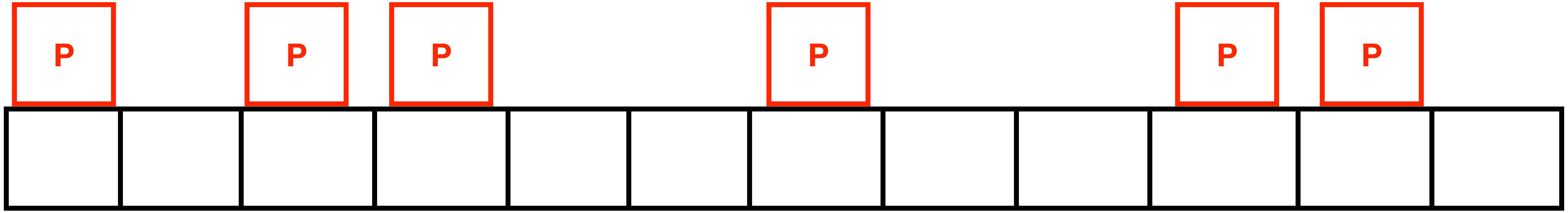
# N bindings sites, m proteins, how many ways can we arrange?

Can you count?



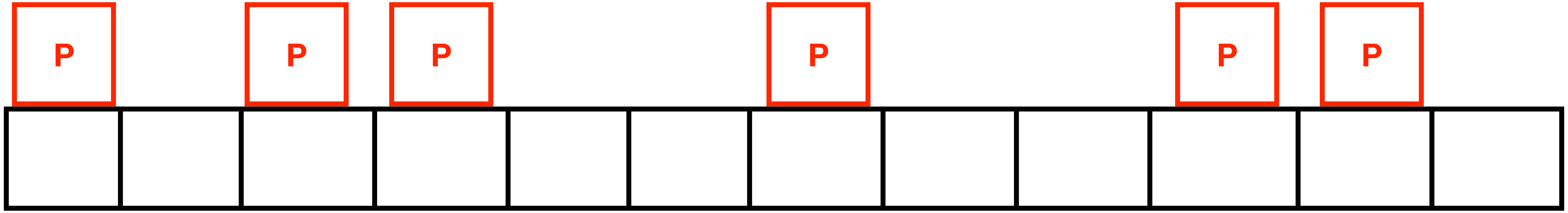

What is entropy?

# N bindings sites, m proteins, how many ways can we arrange?



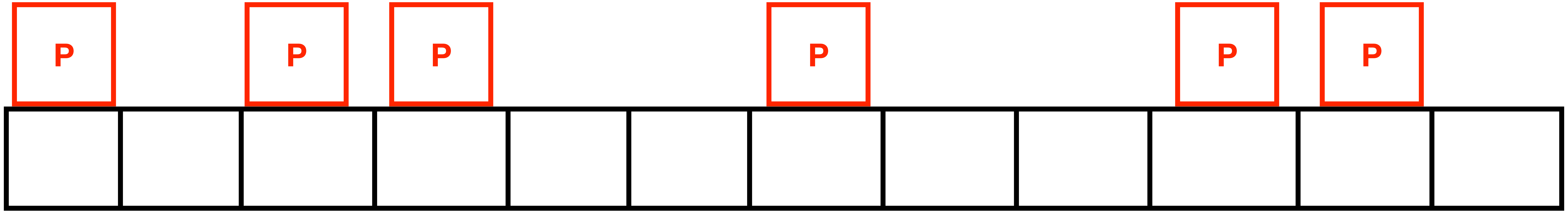

$$W = \binom{N}{m} = \frac{N!}{m!(N-m)!}$$

# Entropy of protein arrangements




$$S = k_B \ln W = k_B \ln \left( \frac{N!}{m!(N-m)!} \right)$$

# Entropy of protein arrangements

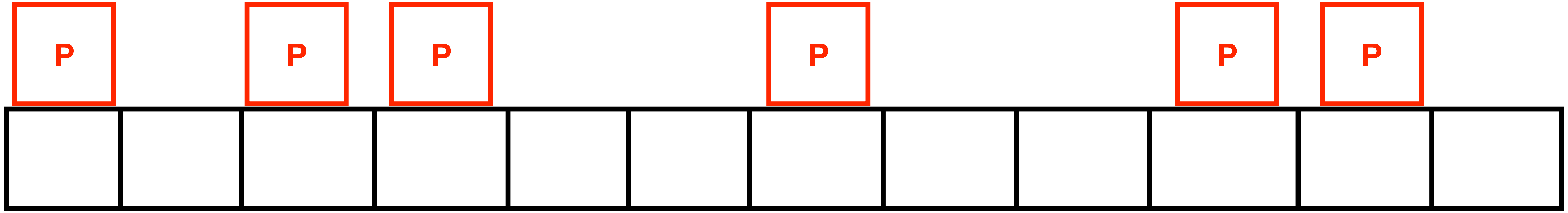



$$S = k_B \ln W = k_B \ln \left( \frac{N!}{m!(N-m)!} \right)$$

$$\ln(N!) \approx N \ln N - N$$

(Stirling's approximation!)

# Entropy of protein arrangements




$$\frac{S}{k_B} = \ln\left(\frac{N!}{m!(N-m)!}\right) = -N[c\ln c + (1-c)\ln(1-c)]$$

$$c = \text{density of bound proteins} = \frac{m}{N}$$

The system minimises free energy

$$\text{Free Energy} = E - TS$$

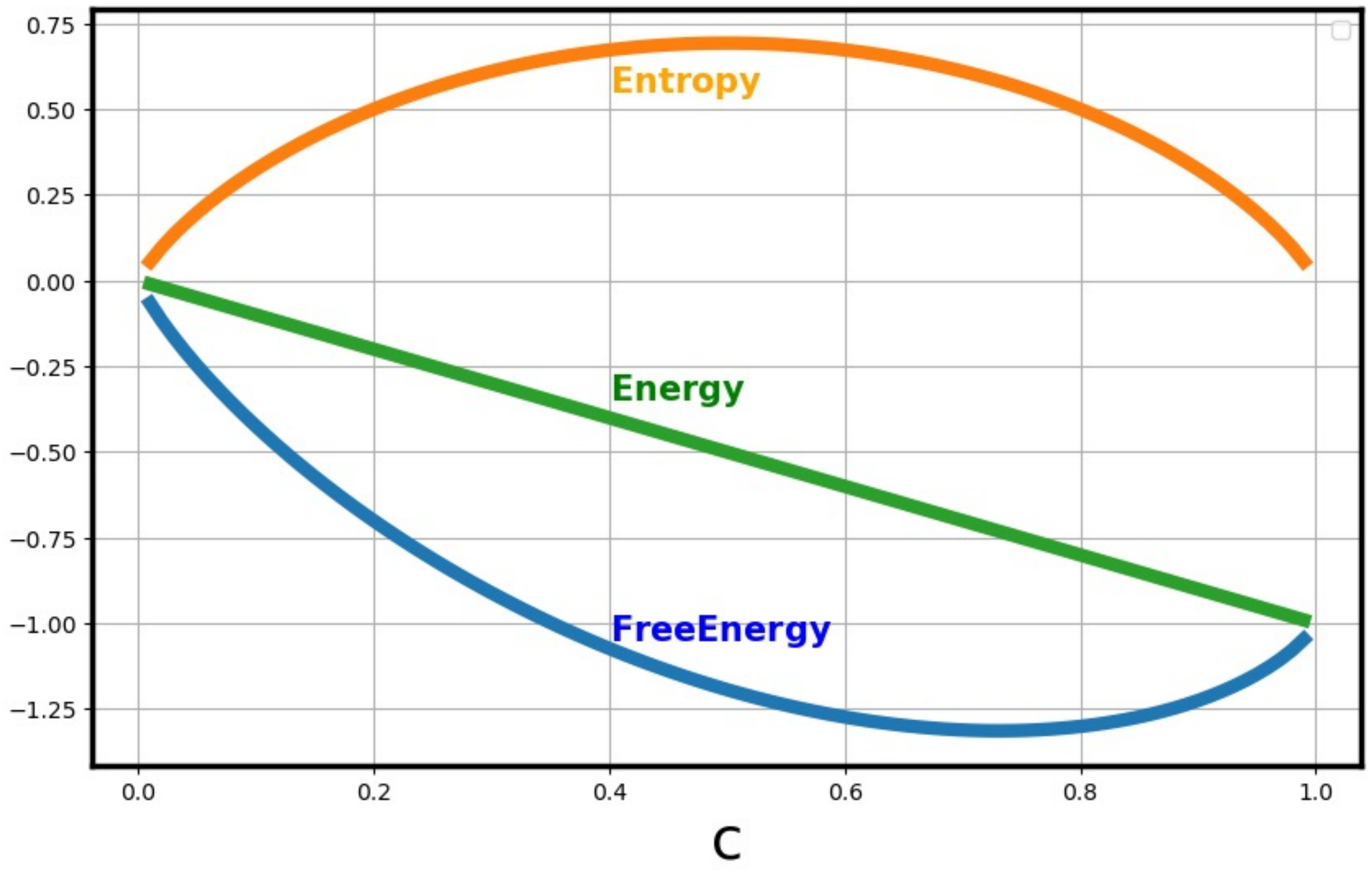
$$\text{Energy when m proteins are bound} = -m\epsilon$$

$$\text{Entropy}, S = -Nk_B[c \ln c + (1-c)\ln(1-c)]$$

$$\text{Entropy}, S = -Nk_B[c\ln c + (1-c)\ln(1-c)]$$

$$\text{Energy when m proteins are bound} = -Nc\epsilon$$

$$\text{Free Energy} = -Nc\epsilon + Nk_BT[c\ln c + (1-c)\ln(1-c)]$$



# Any in vitro (closed) system goes to the “state” where its free energy is minimum!

This is applicable to proteins, vesicle formation, ANYthing you take!

Any in vitro (closed/equilibrium) system goes to the “state” where its free energy is minimum!

BUT what happens in vivo is NOT understood!

Why a biological system in vivo takes the “state” that it takes is not explainable by any one theory so far!

If you find such a theory, you will probably get a Nobel prize!

Living beings are “open” systems. Eat food/ takes energy!

Constantly out of equilibrium!

Life is away from “thermodynamic equilibrium”.

Equilibrium is “death”

Now, instead of arrangement of books, let us considered arrangement of DNA sequence

# Consider DNA sequence for two different proteins. Protein X and protein Y

Imagine we know the sequences across evolution for both proteins

Bacteria, Yeast, worm, fruit fly, mouse

Assume: the sequences have slightly changed during evolution

Which one has changed more?

Which one is more “conserved” over evolution?

How will you quantify?

Count the different “arrangements” of bases seen => entropy

$$\text{Entropy}, S = -k_B \sum_i P_i \ln P_i$$

$$\text{Entropy}, S = -k_B \sum_i P_i \ln P_i$$

If all W arrangements are equally probably, $P_i = (1/W)$

| | | |
|---|---|---|
| A | A | T |
| A | G | T |
| A | T | A |
| A | C | G |
| A | T | T |

| | | |
|---|---|---|
| A | G | T |
| A | C | T |
| A | A | C |
| A | T | T |
| A | G | T |

A A T
A G T
A T A
A C G
A T T

A G T
A C T
A A C
A T T
A G T

Location/position

Base

| | 1st | 2nd | 3rd |
|---|---|---|---|
| A | 5/5 | 1/5 | 1/5 |
| T | 0 | 2/5 | 3/5 |
| G | 0 | 1/5 | 1/5 |
| C | 0 | 1/5 | 0 |

Location/position

Base

| | 1st | 2nd | 3rd |
|---|---|---|---|
| A | 1 | 1/5 | 0 |
| T | 0 | 1/5 | 4/5 |
| G | 0 | 2/5 | 1/5 |
| C | 0 | 1/5 | 0 |

Table: Probability of finding a base B at a particular position i: $P_i^B$

$$S_i = -\sum_B P_i^B \ln P_i^B$$

A A T
A G T
A T A
A C G
A T T

A G T
A C T
A A C
A T T
A G T

Location/position

Base

| | 1st | 2nd | 3rd |
|---|---|---|---|
| A | 5/5 | 1/5 | 1/5 |
| T | 0 | 2/5 | 3/5 |
| G | 0 | 1/5 | 1/5 |
| C | 0 | 1/5 | 0 |

Location/position

Base

| | 1st | 2nd | 3rd |
|---|---|---|---|
| A | 1 | 1/5 | 0 |
| T | 0 | 1/5 | 4/5 |
| G | 0 | 2/5 | 1/5 |
| C | 0 | 1/5 | 0 |

Table: Probability of finding a base B at a particular position i: $P_i^B$

$$S_i = -\sum_B P_i^B \ln P_i^B$$

A A T
A G T
A T A
A C G
A T T

A G T
A C T
A A C
A T T
A G T

Location/position

Base

| | 1st | 2nd | 3rd |
|---|---|---|---|
| A | 5/5 | 1/5 | 1/5 |
| T | 0 | 2/5 | 3/5 |
| G | 0 | 1/5 | 1/5 |
| C | 0 | 1/5 | 0 |
| S(i) | 0 | 1.33 | 0.95 |

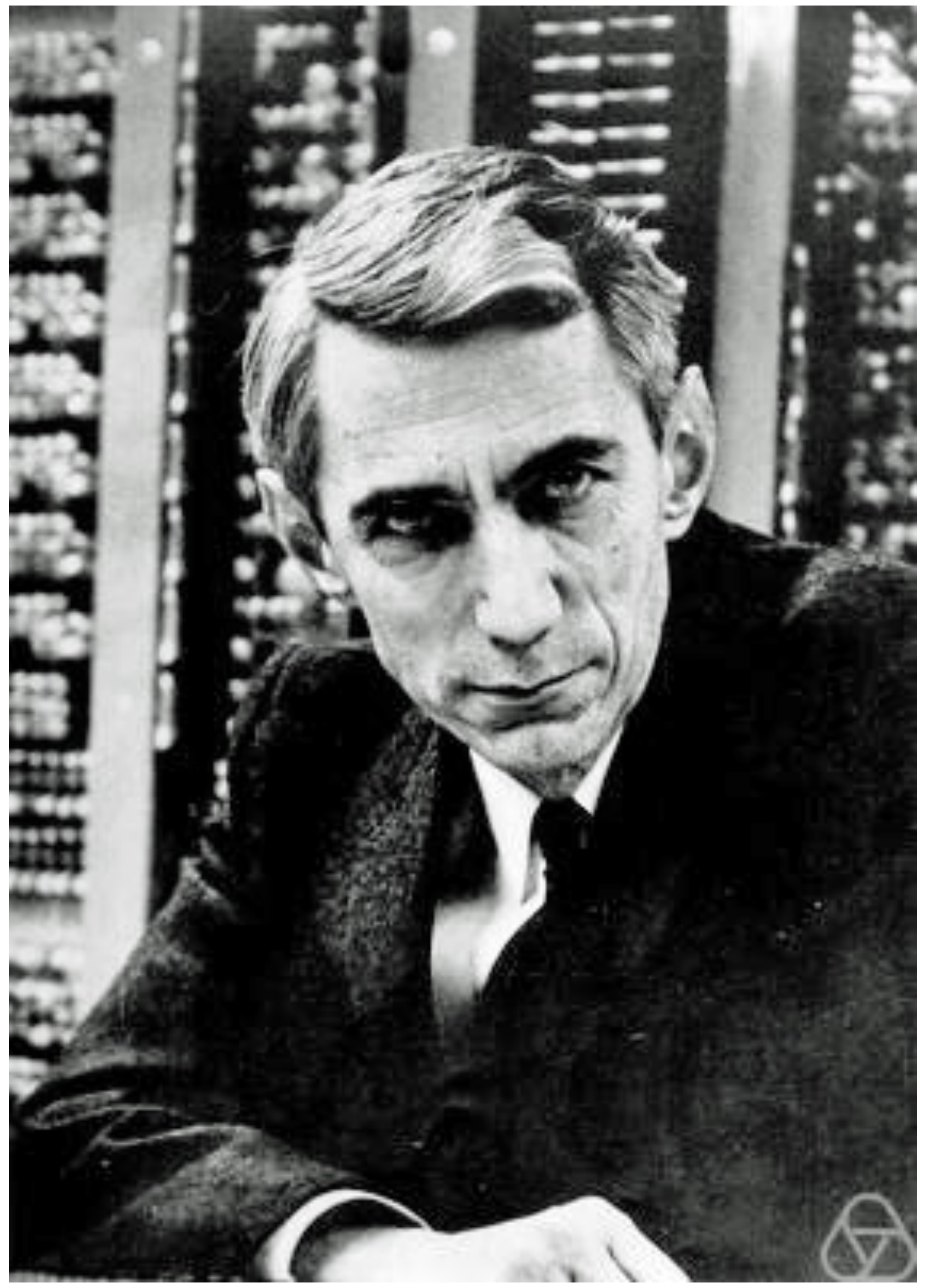
Location/position

Base

| | 1st | 2nd | 3rd |
|---|---|---|---|
| A | 1 | 1/5 | 0 |
| T | 0 | 1/5 | 4/5 |
| G | 0 | 2/5 | 1/5 |
| C | 0 | 1/5 | 0 |
| S(i) | 0 | 1.33 | 0.5 |

This is how you quantify “order” by calculating entropy

$$\text{Entropy}, S = -\sum_i P_i \ln P_i$$



Reprinted with corrections from *The Bell System Technical Journal,*
Vol. 27, pp. 379–423, 623–656, July, October, 1948.

## A Mathematical Theory of Communication

By C. E. SHANNON

### INTRODUCTION

THE recent development of various methods of modulation such as PCM and PPM which exchange bandwidth for signal-to-noise ratio has intensified the interest in a general theory of communication. A basis for such a theory is contained in the important papers of Nyquist[1] and Hartley[2] on this subject. In the present paper we will extend the theory to include a number of new factors, in particular the effect of noise in the channel, and the savings possible due to the statistical structure of the original message and due to the nature of the final destination of the information.

Claude Shannon: father of information theory

Just like two mobile phones communicate, cells also communicate

Instead of electro magnetic waves, cells use chemicals diffusing and reacting!

Thermodynamic entropy and information entropy meet in biology

Nature of evolution => creating diversity; by changing underlying DNA sequence

If we were to get DNA sequence of two human population, can we use entropy to estimate which population is “ancient” and which one is new?

Life is away from “thermodynamic equilibrium”.

Equilibrium is “death”

Entropy production is related to non-equilibrium processes => Living systems

Life <-> entropy!

# Summary

- Thermodynamic entropy
- Free energy
- Information entropy
- Variability

“Life” $\leftrightarrow$ Entropy