

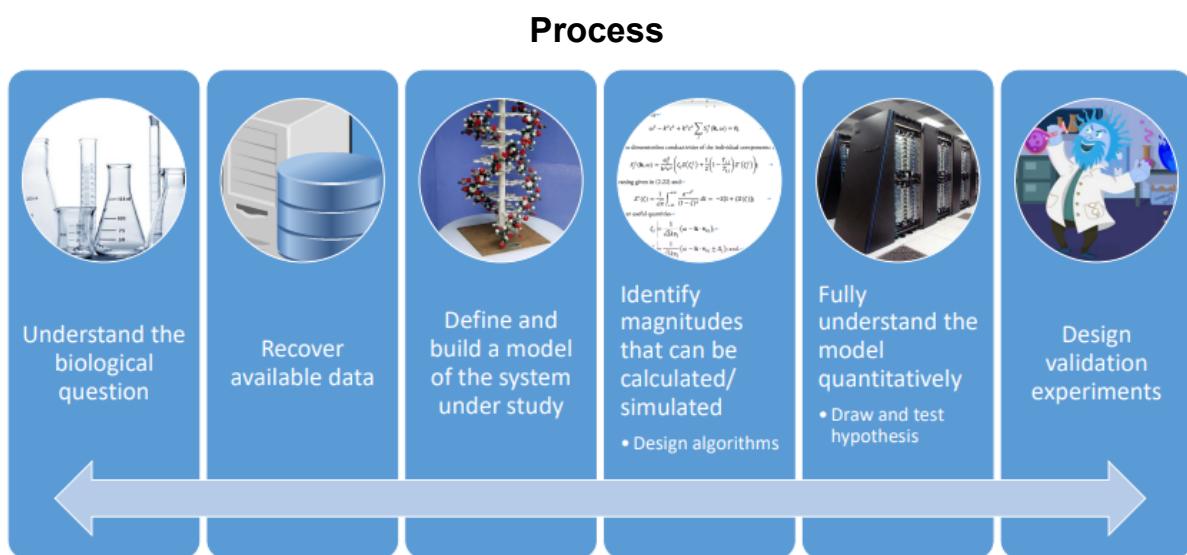
# Unit 0. Introduction

Biophysics is an interdisciplinary science that applies the approaches and methods of physics to study biological systems.

It does cover all scales of biological organization, from molecular (we will focus on this one) to organismic and populations:

- **Molecular scale:** Understand thermodynamic and kinetics aspects of the structure to function relationships in biomolecules

It overlaps with many other fields (reason why this subject is called biophysics, it includes many fields).



**Step 1:** We must confirm that it is a molecular question and that it is quantitative (if the numbers obtained in the result are not as good as expected, we can make a qualitative analysis).

*Note:* IC<sub>50</sub> values are common in drug design experiments, but have no direct physical meaning. We can not use IC<sub>50</sub> to compare 2 labs.

**Step 2:** There are plenty of DBs. The data should be FAIR:

- Findable
- Accessible
- Interoperable
- Re-usable

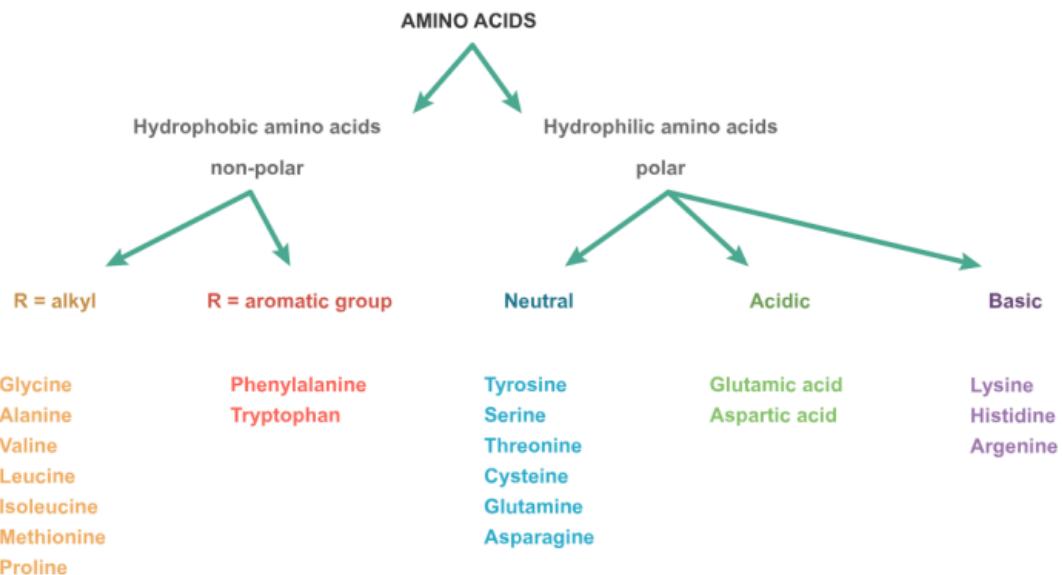
**Step 3:** Define which model we want to use to make the analysis (it should answer the biological question).

## AA

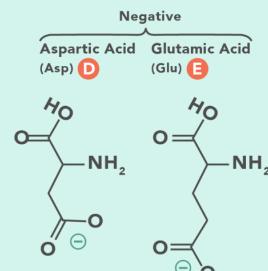
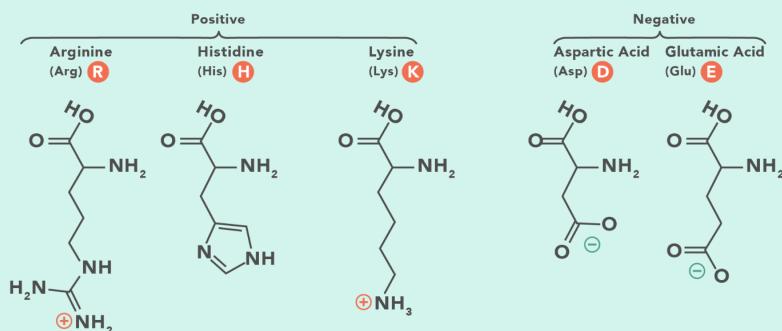
The general structure of the aa is  $\text{NH}_2 - \text{CH}(\text{R}) - \text{COOH}$  and a variable side chain.

At neutral pH:

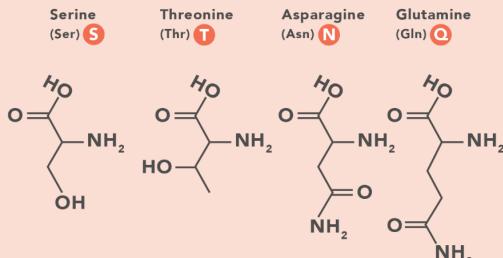
- $\text{NH}_2$  protonated to  $\text{NH}_3^+$  and  $\text{COOH}$  deprotonated to  $\text{COO}^-$



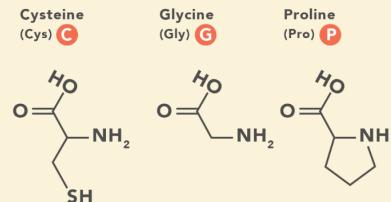
A. Amino Acids with Electrically Charged Side Chains



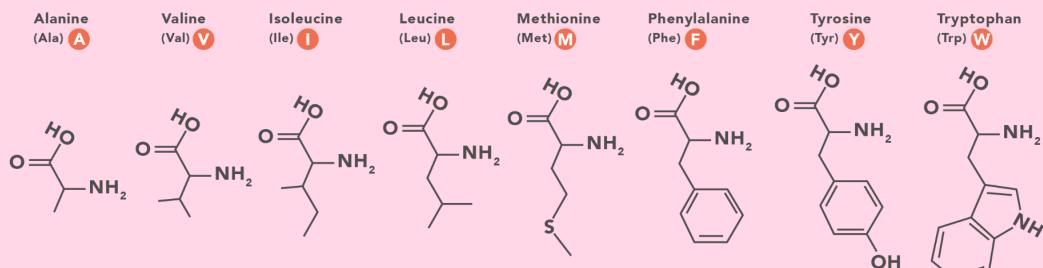
B. Amino Acids with Polar Uncharged Side Chains



C. Special Cases



D. Amino Acids with Hydrophobic Side Chains



# Unit 1. Classical Thermodynamics

Thermodynamics is the study of the effects of work, heat and energy on a system

**Open system:** Mass and energy can enter and leave the system (biological systems)

**Closed system:** Only energy can enter and leave the system (labs are considered closed)

**Isolated system:** Nothing can enter or leave

**Energy:** Capability to produce work of any kind. The energy depends on the mass. If I double the mass, I have also doubled the energy.

**Temperature:** Measure of "heat content". It does not depend on the mass.

**Heat (q) and work (w):** Are ways of transferring energy. Example of moving a table or increasing the temperature.

The amount of heat or work transferred depends on how we do things (depends on the process).

**State function:** Any magnitude that does not depend on the process, only on initial and final states

**Internal energy (E):** Energy contained in the system. Example: Potential, kinetic, chemical  
At closed systems,  $\Delta E$  corresponds to heat exchanged at constant volume ( $q_v$ )

**Enthalpy (H):** The total heat contained in the system ( $H = E + PV$ )

At closed systems,  $\Delta H$  corresponds to heat exchanged at constant pressure ( $q_p$ ).

We know that chemical bonds contain energy and in chemical reactions, bonds are destroyed and made. At constant pressure,  $\Delta H$  is the difference in energy between the broken bonds and the created bonds.

Breaking bonds:  $\Delta H > 0$  and  $\Delta S > 0$

Creating bonds:  $\Delta H < 0$  and  $\Delta S < 0$

**Note:** Capital letters are for state functions

## 1st Thermodynamics principle

The energy is not created or destroyed, it is transferred

$$\Delta E = q + w$$

**Reversible and irreversible (spontaneous) processes:** A process is considered reversible when it can be (infinitesimally) reversed without significant change in the thermodynamic properties of the system and environment.

In other words, when the system and the environment are in equilibrium.

**Entropy (S):** Measure of the probability of the states. In other words, it is a measure of how dispersed energy is. It is the heat exchanged in a reversible process divided by the temperature.

$$S = q_{rev}/T$$

Universe  $\Delta S = 0$  for reversible processes and  $\Delta S > 0$  for irreversible processes

## 2nd Thermodynamics principle

The entropy of the universe is always  $\geq 0$

**Gibbs free energy (G):** The available energy to make work

$\Delta G = 0$  for reversible processes and  $\Delta G < 0$  for irreversible ones

$\Delta G > 0$  indicates that the opposite reaction is spontaneous.

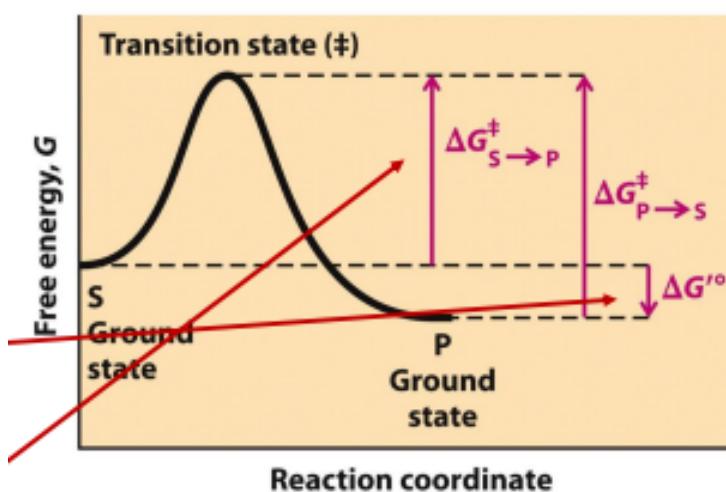
## Chemical processes

We work at constant V and P!

Values at 1M concentration of all reactants are taken as the “reference” values  $\Delta H^\circ$ ,  $\Delta S^\circ$ ,  $\Delta G^\circ$  and, these values should be corrected with the actual concentration in which we are working.

Reactions proceed in the direction of decreasing  $\Delta G$ . This does not imply that the reaction is actually produced at a reasonable rate (Kinetics is not included here).

It needs to be excited to a certain level (transition state). This is how our body is regulated.



## Thermodynamics in biology. Metabolism

Knowing the values of Gibbs, allows us to know the direction of the process.

If we want to stop a process, we need to block a stage that is highly spontaneous.

## From macroscopic to microscopic scales

Normal thermodynamics measures “macroscopic” properties of a system.

At the microscopic levels the concepts remain but their interpretation is different:

- Internal energy: Kinetic energy of molecules + interactions between components
- Pressure: Collisions with system walls
- Temperature: Kinetic energy of molecules

The distribution of properties is not uniform, because molecules are in different states with different populations and energies (any measure is always an average of those states).

## Kinetic theory of gases

The pressure that a gas exerts is caused by the collisions of its molecules with the walls of the container.

The pressure is proportional to the number of molecules per volume unit and to the average translational kinetic energy of molecules.

Assumptions:

- Large number of molecules
- Large separation between molecules
- They move randomly
- Obey Newton's Law
- Collide elastically with each other and with the wall (so they do not lose momentum)
- Consists of identical molecules

$$P = \frac{F}{A} = \frac{F}{d^2} = \frac{1}{3} \left( \frac{N}{d^3} m \bar{v^2} \right) = \frac{1}{3} \left( \frac{N}{V} \right) m \bar{v^2} = \frac{2}{3} \left( \frac{N}{V} \right) \left( \frac{1}{2} m \bar{v^2} \right)$$

## Microscopic temperature

Measure of the kinetic energy. We know that:

We also know the ideal gas equation (expressed in molecules instead of mols. R -> k<sub>B</sub>):

$$PV = \frac{2}{3} N \left( \frac{1}{2} m \bar{v^2} \right)$$

$$PV = N k_B T$$

If we combine them, we obtain:

$$k_B = R / N_A$$

$$T = \frac{2}{3k_B} \left( \frac{1}{2} m \bar{v^2} \right)$$

$$k_B = 1.3806 \cdot 10^{-23} \text{ J/K}$$

$$R = 8.314 \text{ J / (K mol)}$$

$$R = 1.987 \text{ Kcal / (K mol)}$$

$$N_A = 6.022 \cdot 10^{23} \text{ mol}^{-1}$$

## How are molecules distributed

The Boltzmann distribution is a probability distribution that gives the probability of a certain state as a function of that state's energy and temperature of the system to which the distribution is applied. It is given as:

$$p_i = \frac{e^{-E_i/k_B T}}{\sum_{j=1}^M e^{-E_j/k_B T}}$$

Where  $p$  is the probability of state  $i$ ,  $\epsilon$  the energy of state  $i$ ,  $k$  the Boltzmann constant,  $T$  the temperature of the system and  $M$  is the number of all states accessible to the system. The denominator is a normalization.

$p$  is the probability of being in one state, so it is related directly to the amount of particles that are in that state. Remember that states that have the same energy level have the same probability of existing.

Boltzmann distribution corresponds to the maximum probability (hence the maximum Entropy!!)

Population ratios can be calculated as:

$$\frac{N_j}{N_i} = \frac{p_j}{p_i} = e^{-\Delta E_{ij}/k_B T}$$

## How are velocities distributed

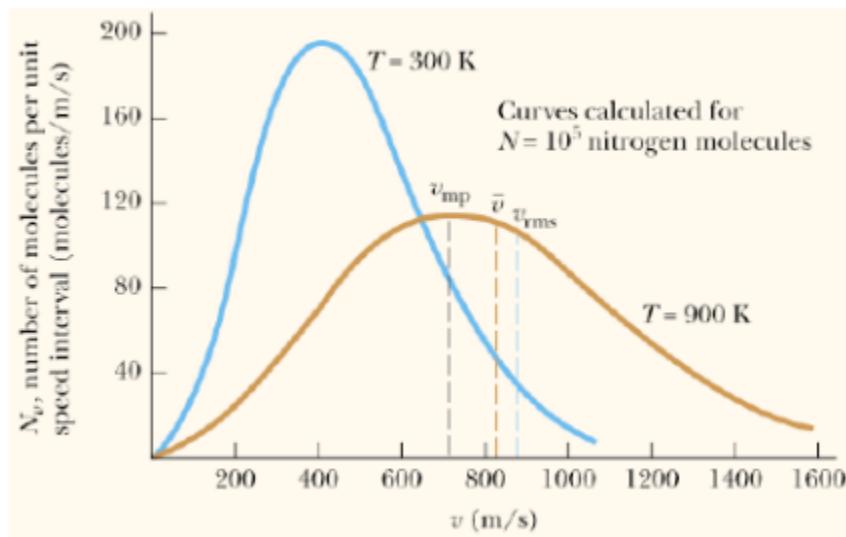
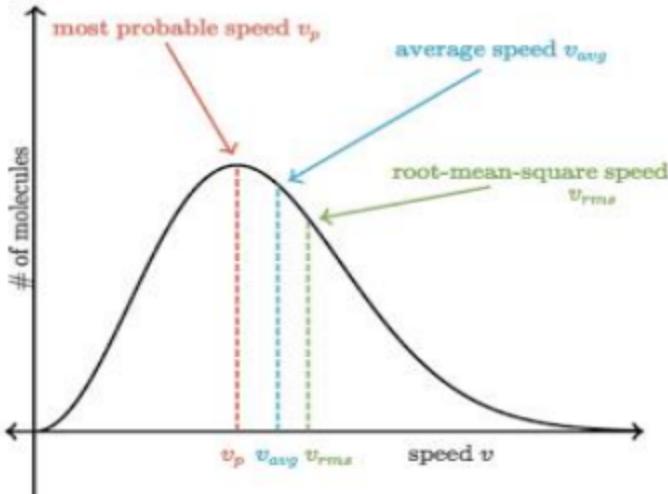
Velocities follow the Maxwell-Boltzmann distribution:

$$f(v) = \sqrt{\left(\frac{m}{2\pi kT}\right)^3} 4\pi v^2 e^{-\frac{mv^2}{2kT}}$$

The Maxwell-Boltzmann distribution shows that the molecular velocities depend on mass and temperature.

For a given temperature, the fraction of velocities greater than a reference value increases with the reduction of the mass.

This explains why light molecules, such as hydrogen and helium, escape easily from the atmosphere. Heaviest molecules as nitrogen and oxygen are retained.



If the temperature increases, all the states go together in terms of energy and thus the velocities will be more uniform (flat distribution)

## Statistical thermodynamics

Boltzmann distribution predicts populations of states in systems at thermal equilibrium  
The deviation of Boltzmann distribution leads naturally to the introduction of partition function

### Population of a state

The average number of molecules that occupy it. On average, there are "n" molecules in a state of energy " $\epsilon$ ".

The populations of the states depend only on " $\epsilon$ " that is function of T

All states having the same energy level and T, are equal probable.

## Configurations and weights

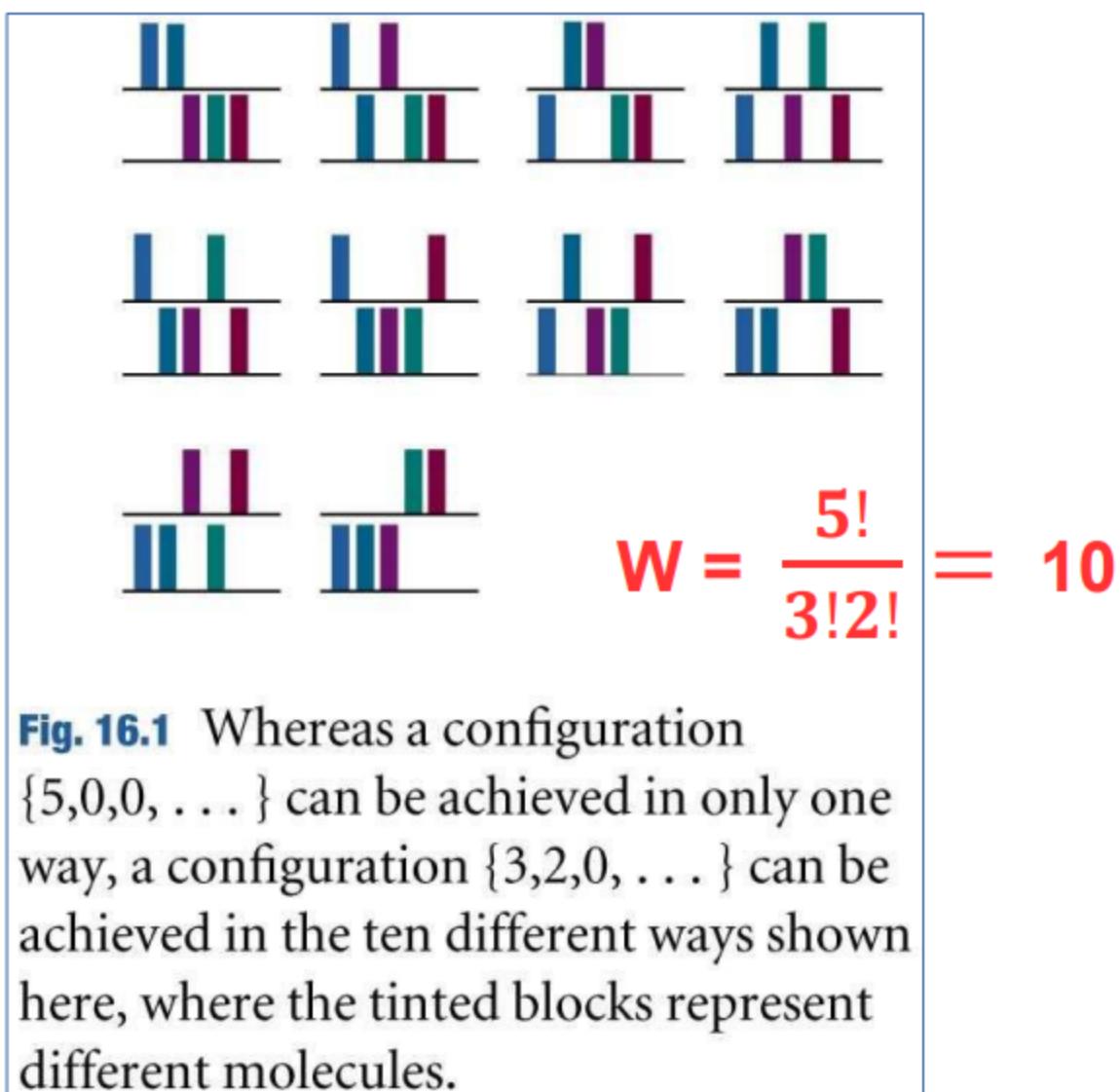
At any instant, molecules could occupy one energy level.

The specification of the set of populations  $n_0, n_1, n_2, \dots$  in the form  $\{n_0, n_1, n_2, \dots\}$  is a statement of the instantaneous configuration of the system.

Weight of a configuration ( $W$ ) =  $N$  possible microstates for a configuration:

$$W = \frac{N!}{n_0!n_1!n_2!n_3!\dots}$$

*Example: 5 molecules, 2 energy levels*



## Configuration and weights: The optimal distribution

We could find the most probable distribution imposing a system with:

Constant total number of molecules:  $N = \sum_i n_i$

Constant total energy:  $E = \sum_i p_i \varepsilon_i$

A compatible configuration having the maximum weight  $\delta W = 0$

**Boltzmann distribution:**  $p_i = \frac{n_i}{N} = \frac{e^{-\beta \varepsilon_i}}{\sum_j e^{-\beta \varepsilon_j}}$        $\beta = \frac{1}{k_B T}$

## The molecular partition function

Boltzmann distribution is a probability distribution. It gives the probability of a certain state in function of that state's Energy and the Temperature of the system

The normalization term is called Molecular (canonical) Partition Function

Molecular partition function defines the statistical properties of the system from the microscopic states.

The molecular partition function gives an indication of the number of states that are thermally accessible to a molecule at the temperature of the system:

- At  $T = 0$ , only the ground level is accessible and  $q = g$
- At very high temperatures, virtually all states are accessible, and  $q$  is correspondingly large

$$p_i = \frac{n_i}{N} = \frac{e^{-\beta \varepsilon_i}}{\sum_j e^{-\beta \varepsilon_j}} \quad \beta = \frac{1}{k_B T}$$

$$Z = \sum_j e^{-\beta \varepsilon_j} = \sum_l g_l e^{-\beta \varepsilon_l}$$

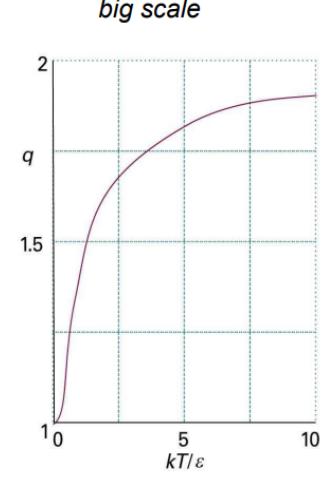
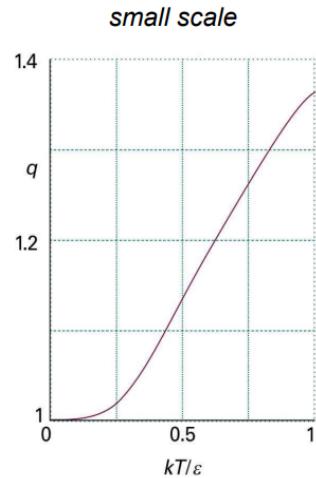
Counting either for "j" microstates or for "l" energy levels  $g_j$  degenerated

## Example: Two level system

$$\varepsilon = \varepsilon \quad \text{---}$$

$$\varepsilon = 0 \quad \text{---}$$

$$Z = \sum_{i=0}^{\varepsilon} e^{-\beta \varepsilon_i} = 1 + e^{-\beta \varepsilon}$$



$$\varepsilon = \varepsilon \quad \text{---}$$

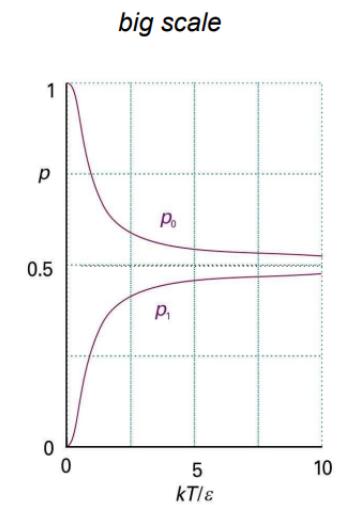
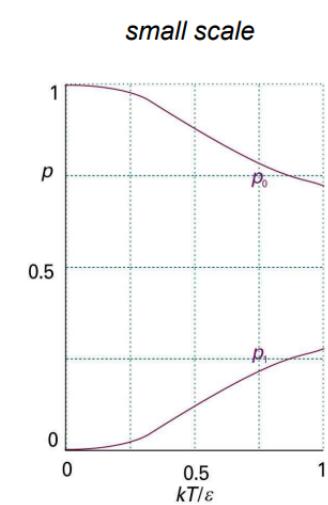
$$\varepsilon = 0 \quad \text{---}$$

$$p_i = \frac{n_i}{N} = \frac{e^{-\beta \varepsilon_i}}{q} = \frac{e^{-\beta \varepsilon_i}}{1 + e^{-\beta \varepsilon}}$$

$$Z = \sum_{i=0}^1 e^{-\beta \varepsilon_i} = 1 + e^{-\beta \varepsilon}$$

$$p_0 = \frac{n_0}{N} = \frac{1}{1 + e^{-\beta \varepsilon}}$$

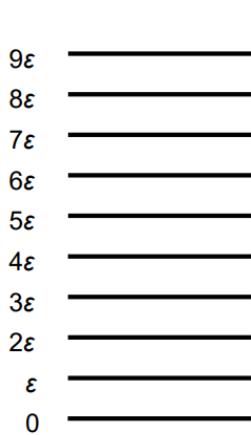
$$p_1 = \frac{n_1}{N} = \frac{e^{-\beta \varepsilon}}{1 + e^{-\beta \varepsilon}}$$



The fraction of populations of the two states of a two-level system as a function of temperature.

## Example: Uniform distribution of energy levels

A system with equally spaced non-degenerate energy levels



$$p_i = \frac{n_i}{N} = \frac{e^{-\beta \varepsilon_i}}{\sum_{i=0}^9 e^{-\beta \varepsilon_i}}$$

Temperature (in  $\beta\varepsilon$ )

Partition function



The equally spaced array of energy levels used in the calculation of the partition function.

## Formulas

Internal energy  $E = \frac{-N}{Z} \left( \frac{\partial Z}{\partial \beta} \right) = \frac{N}{Z} \sum_i \varepsilon_i e^{-\beta \varepsilon_i}$

$$p_i = \frac{1}{Z} e^{-\beta \varepsilon_i}$$

$$\ln p_i = -\beta \varepsilon_i - \ln Z$$

Pressure  $P = k_B N T \left( \frac{\partial \ln Z}{\partial V} \right)$

$$-\sum_i p_i \ln p_i = \sum_i (p_i \beta \varepsilon_i + p_i \ln Z) =$$

Entropy  $S = E/T + k_B N \ln Z$

$$\beta \sum_i p_i \varepsilon_i + \ln Z = \beta E/N + \ln Z$$

Gibbs  $G = -k_B N T \left[ \ln Z - \left( \frac{\partial \ln Z}{\partial V} \right) \right]$

$$N \ln Z = -\beta E - N \sum_i p_i \ln p_i$$

"Statistical" entropy

$$S = \frac{E}{T} + k_B N \ln Z = -k_B N \sum_i p_i \ln p_i$$

$$S = -k_B N \sum_i p_i \ln p_i$$

$$S = -k_B N \sum_i \frac{n_i}{N} \ln \frac{n_i}{N} = -k_B N \left( \frac{1}{N} \sum_i n_i \ln n_i - \ln N \right)$$

$$S = k_B \ln W$$

$$\ln W = N \ln N - \sum_i n_i \ln n_i$$

## Biophysics seminar 1: Thermodynamics

Entropy is always increasing because it is always the most likely situation from a statistical point of view (example of 2 dice).

**Stirling engine:** We apply heat to the bottom plate and we give a little push to the wheel. This will move the piston and the foam will go up. The air in the chamber gets in contact with the hot plate and gets warmed up.

As the air in the chamber gets hot, it expands itself. This will move the piston.

Since the piston is connected to the wheel, it will start moving,

The wheel is connected to the block of foam, when it moves it pushes the foam down.

The hot air in the chamber transfers heat to the cold plate. As the air in the chamber gets cold, it compresses itself. This moves the piston down.

### Questions regarding the Stirling engine:

#### 1. What is the role of entropy in the functioning of a Stirling engine?

Expanding and compressing the gas. When increasing the temperature, the air expands and when the temperature decreases, the air will compress. This is the fact that makes the

Entropy is increasing from the first state to the other one.

#### 2. Do you think this engine can run forever from the heat of a cup of hot water? Why?

No, because at a certain time the cup of hot water will cool down and therefore the air will not expand any more.

Also, at some point there will be an equilibrium (there will be no difference in the temperature).

#### 3. If we put the engine in a very hot room, do you think it will be able to run using the heat of a cup of hot water? Why?

No, because the top platform will be hot enough to make the air expand and therefore the piston will never go down.

#### 4. Do you think this engine can run using a source of cold instead of a source of heat? Why?

It depends if we are in a hot or cold room.

If we are in a really cold room, it won't work. This is due to the fact that both plates will be at the same temperature.

If we are in a warm room, it will work. This is due to the fact that one plate Will be warmer than the other and therefore the air will expand (when in contact with the warm plate) or compress (when in contact with the cold plate).

Note that the mechanism will work “backwards”.

### Questions regarding dice probabilities and entropy:

1. **What number is the most likely when rolling two dice? Why?**

7. Because its the number that has a higher combination of numbers that gives as a result 7.

2. **How many microstates are available for that number?**

1/6, 2/5, ¾, 4/3, 5/2, 6/1

So, there are 6 microstates.

3. **What number is the most unlikely when rolling two dice? Why?**

2 and 12. Since there is only 1 combination of numbers that give as a result that number.

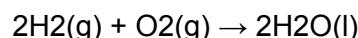
4. **How many microstates are available for that number?**

1

### Questions regarding enthalpy:

1. **The combustion of hydrogen is an exothermic process. Assume that this process takes place at constant pressure.**

- **Write the corresponding balanced equation.**



- **Say if  $\Delta H > 0$  or  $\Delta H < 0$ .**

Smaller than 0

- **Say what bonds contain more energy, the ones of products or the ones of the reactants.**

The reactants

2. **Consider two complementary molecules of single stranded DNA inside the cell nucleus. If they hybridize, will this process increase or decrease the enthalpy of the DNA molecules?**

It will decrease the enthalpy. Because we are creating bonds.

**3. The hydrolysis of ATP in our cells is an exothermic process:**

- Say if for this process  $\Delta H > 0$  or  $\Delta H < 0$ .

$\Delta H < 0$

- Can you deduce if the  $\Delta H$  for ATP synthesis will be  $> 0$  or  $< 0$ ? Why?

Bigger than 0

**Imagine a protein with a lot of alpha helices:**

- Does it has hydrogen bonds? Yes, alpha helices are bounded by hydrogen bonds.
- What happens if we increase the temperature of the protein? The protein will denaturalize (lose the structure because the hydrogen bonds break)
- What is the  $\Delta H$  of this process? It is positive because we are breaking bonds.

**Questions regarding the Gibbs free energy equation:**

**1. Consider a block of ice and a glass of liquid water:**

- What is the  $\Delta H$  for going from solid ice to liquid water? We are breaking bonds and thus, it will be positive.
- What is the  $\Delta S$  for going from solid ice to liquid water? It will increase
- Under what conditions becomes spontaneous going from solid ice to liquid water? When the temperature is high enough.

**2. Consider two water molecules floating in vaccum, if they create a hydrogen bond between themselves:**

- What is the  $\Delta H$  of the process? It is negative
- What is the  $\Delta S$  of the process? It is negative
- What value is larger in absolut values?  $|\Delta H|$  or  $|T \cdot \Delta S|$ ? The enthalpy, because the process is spontaneous.

## Kinetics

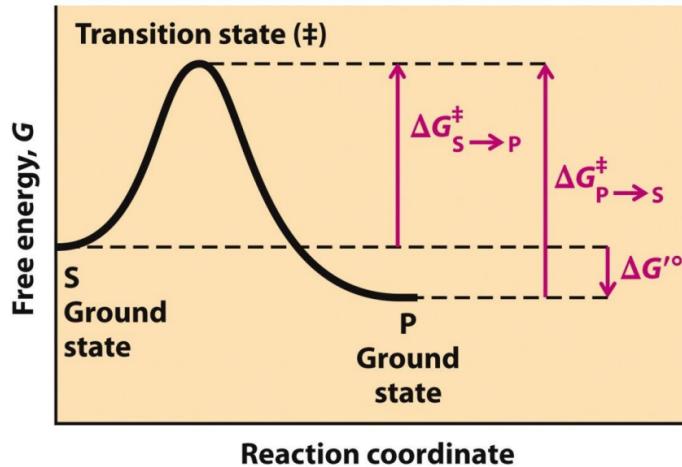
Thermodynamics allows us to know whether a process is “in equilibrium” ( $\Delta G = 0$ ) or “spontaneous” ( $\Delta G < 0$ ), but does not predict whether the process takes place in a reasonable time frame.

Chemical kinetics aims to understand the actual rate of the process.

Reaction progress requires (in most cases) that the system overcomes an energy barrier (transition state). Even when the system reaches the necessary energy there is still just a probability of going forward.

Example of breaking a bond and also separating the atoms that were attached.

Reaction progress is conditioned by reactant, and product concentrations, T, P, ...



## Kinetics in Biology

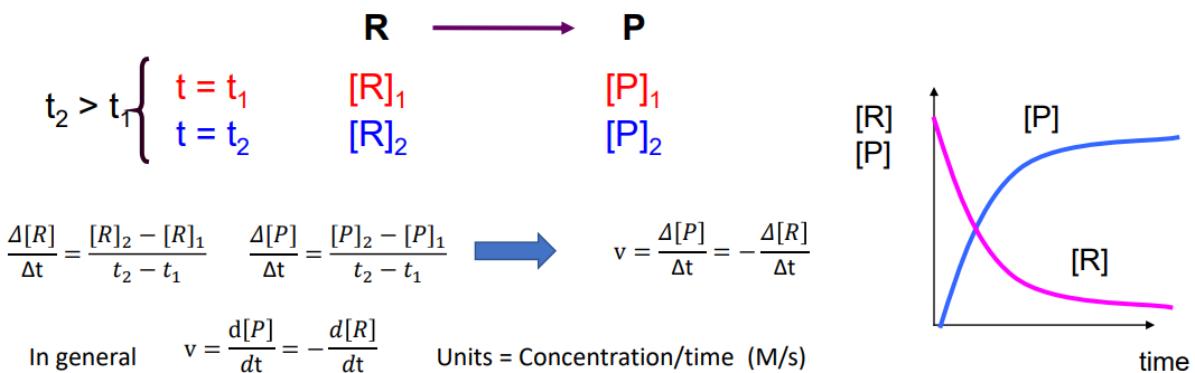
Relevant processes:

- Chemical reactions
- Transport across membranes
- Physical movement

All biological processes are controlled through kinetics regulation

## Concepts and nomenclature

**Reaction rate:** The amount of product that is formed per time or the amount of reactant that disappears per time. So, it's the variation of concentration of a reactant or product per time unit.



$$\text{The general case for } aA + bB \rightarrow cC \quad v = -\frac{1}{a} \frac{dA}{dt} = -\frac{1}{b} \frac{dB}{dt} = \frac{1}{c} \frac{dC}{dt}$$

**Order of reaction:** The number of different concentrations required to evaluate reaction rate.

More specifically, the reaction order is the exponent to which the concentration of that species is raised, and it indicates to what extent the concentration of a species affects the rate of a reaction, as well as which species has the greatest effect.

**Half live:** Time to reduce the concentration of reactants to 50%

**Chemical mechanism:** The process of chemical transformation that occurs

**Kinetic mechanism:** The expression of how and when the different components interact

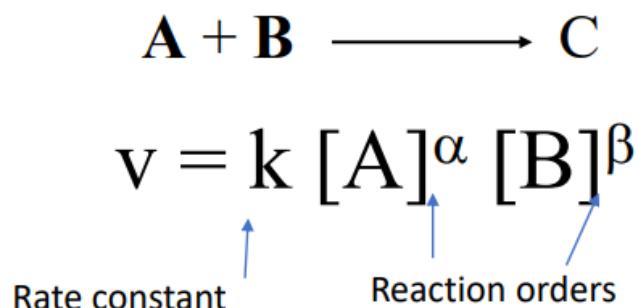
**Rate law or rate equation:** The mathematical expression that relates component's concentrations and global rate.

## Expressions for reaction rates. Effect of concentrations (reaction orders)

**Order:** Empirical magnitude (experimental) determined through the rate law.

**Partial order:** Exponent of each component in the rate law.

**Global order:** Sum of all partial orders of the rate law.

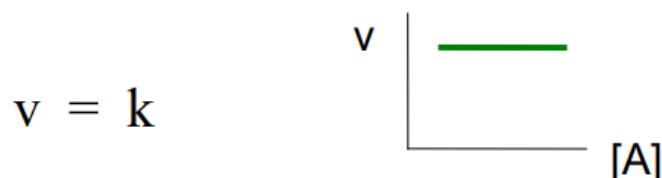


Alpha and beta could be an integer (positive or negative), fractional numbers or zero.

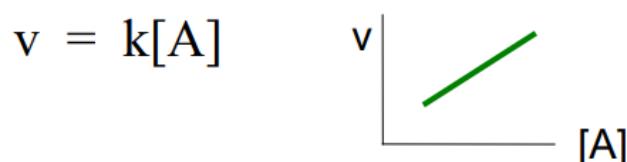
They do not have to coincide with the stoichiometric coefficients, however at the microscopic level they correspond to the number of molecules participating in the reaction.

### Rate laws at different reaction orders

**Zero order ( $n=0$ ):** The velocity of the reaction is independent of the concentration of the reactants



**First order ( $n=1$ ):** The velocity of the reaction is proportional to the concentration of one of the reactants.

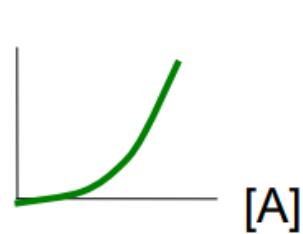


**Second order (n=2):** The velocity of the reaction is proportional to the concentration of two reactants.

- Or proportional to the square of one of them.

$$v = k[A][B]$$

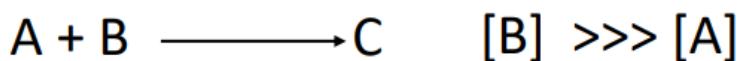
$$v = k[A]^2$$



Order	Differential form	Integrated form	Half live	Units of k	Graphic representation
0	$-\frac{d[A]}{dt} = k$	$[A] = [A]_0 - k_A t$	$\frac{[A]_0}{2k}$	$M s^{-1}$	
1	$-\frac{d[A]}{dt} = k[A]$	$[A] = [A]_0 e^{-k_A t}$	$\frac{\ln 2}{k}$	$s^{-1}$	
2	$-\frac{d[A]}{dt} = k[A]^2$	$\frac{1}{A} = \frac{1}{A_0} + kt$	$\frac{1}{[A]_0 k}$	$M^{-1} s^{-1}$	
2'	$-\frac{d[A]}{dt} = k[A][B]$	$\frac{1}{[A]_0[B]_0} \ln \frac{[B]_0[A]}{[A]_0[B]} = kt$	$\frac{1}{[A]_0[B]_0}$	$M^{-1} s^{-1}$	

### Reactions of “pseudo” first order

Bimolecular reactions (usually second order) can be simplified keeping one of the concentrations constant (and much larger)



$$v = k [A][B] \rightarrow (k [B]) [A] = k' [A]$$

where  $k'$  is the ‘pseudo’ first order constant depending on  $[B]$

## Methods to determine reaction orders

**Integration method:** The experimental values of concentration or pressure (for gasses) as a function of time are plotted with the different integrated equations of velocity.

The order of reaction is found by the representation that yields a constant value during all the profile for the velocity constant

**Half live method:** The experimental values of concentration or pressure (for gasses) are used to calculate the different expressions of half live.

The expression that gives a constant value for the velocity constant indicates the order of the reaction.

**The method of initial rates:** It is based on the measure of different initial rates ( $v_0$ ) with experiments that change the concentration of one of the reactants.

A plot of  $\log v_0$  as a function of  $\log [A]_0$  gives the order of the reactant A.

The same procedure is followed for all reactants.

## Effect of temperature

The progress of the reaction requires the system to surmount an energy barrier

- Hint: Increase of T increases the probability of exploring high energy states

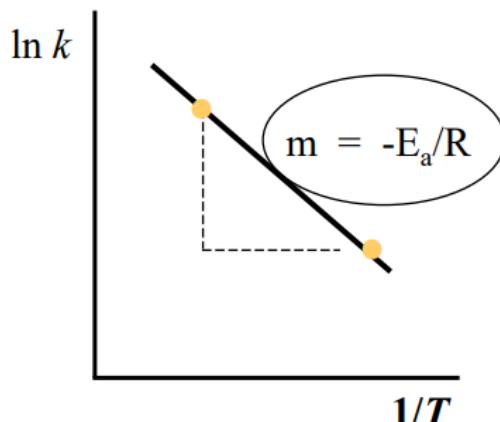
Arrhenius obtained an empirical relation between T and k (spoiler: Boltzmann strikes again)

$$k = A e^{-E_a/RT}$$

Where  $E_a$  is called the activation energy, and A is a measure of the probability of collisions between reactants being productive.

$$\ln k = \ln A - \frac{E_a}{RT}$$

$$\ln \frac{k_1}{k_2} = - \frac{E_a}{R} \left( \frac{1}{T_1} - \frac{1}{T_2} \right)$$



## Reaction rates (from the microscopic view)

There are two main theories to understand (and evaluate) rate constants from the microscopic view.

**Collision theory:** Reactant molecules must collide (in a specific orientation) before a reaction can occur. Using Maxwell-Boltzmann distribution and defining a threshold energy (the activation energy  $E_a$ ) for reactants gives:

$$k = Z^0 e^{-E_a/RT}$$

**Transition state theory:** It explains the progress of the reaction in “thermodynamic terms”. This is the preferred approach in biology

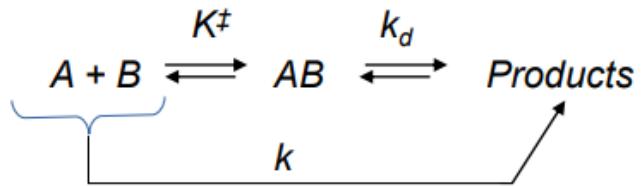
### Transition state theory

TST considers the formation of an intermediate  $AB^\ddagger$

$AB^\ddagger$  is an activated complex in thermal equilibrium with A and B.

Now, the reaction is divided in two steps:

- The formation of  $AB^\ddagger$  considered as an equilibrium process
- The decomposition of  $AB^\ddagger$  to form products.



$[AB^\ddagger]$  decomposition rate ( $k_d$ ) depends on the vibrational state of  $AB$ . It comes from kinetic quantum theory.

Classic Thermodynamics is used to deal with  $AB^\ddagger$

$$k_d = q \frac{k_B T}{h} \quad \Delta G^\ddagger = - RT \ln K^\ddagger$$

q: progress probability (~1)  
h: Plank constant

$$K^\ddagger = \frac{[AB^\ddagger]}{[A][B]}$$

## Transition state theory

Macroscopic law

$$v = k [A][B]$$

Microscopic law

$$v = k_d [AB^\ddagger]$$

using  $K^\ddagger = \frac{[AB^\ddagger]}{[A][B]}$

$$v = k_d K^\ddagger [A][B]$$

$$k = k_d K^\ddagger$$

using  $k_d = q \frac{k_B T}{h}$

$$\Delta G^\ddagger = -RT \ln K^\ddagger$$

$$k = q \frac{k_B T}{h} e^{-\frac{\Delta G^\ddagger}{RT}}$$

using  $\Delta G^\ddagger = \Delta H^\ddagger - T \Delta S^\ddagger$

$$k = q \frac{k_B T}{h} e^{\frac{\Delta S^\ddagger}{R}} e^{-\frac{\Delta H^\ddagger}{RT}}$$

$\Delta S^\ddagger$  has a large influence independent on T

## Catalysis

A catalyst is a substance that changes the velocity of a chemical reaction. Catalysts don't appear in the stoichiometric equation.

It modifies the reaction mechanism to decrease the energy of activation

A small quantity of catalysts is enough to produce a considerable effect. They can increase the reaction rate from 10 to  $10^{12}$  times. The catalyst is not destroyed, it is regenerated.

It does not affect the position of the equilibrium (it works on both directions of the reaction)

Most biological processes have negligible rates without catalysis.

- Biological catalysts are mostly proteins (Enzymes), but also RNAs (Ribozymes)
- Control of catalytic effect is key to control biological processes

## Enzymes use binding energy to decrease activation barriers

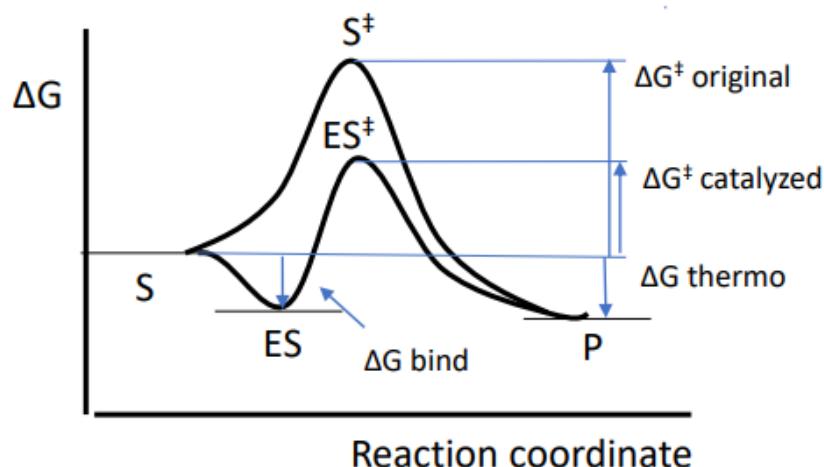
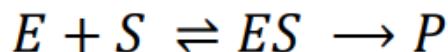
First step is the formation of ES

- The binding energy ( $\Delta G < 0$ ) is invested in decreasing the original activation energy

Binding decreases the degrees of freedom of S and reduces  $\Delta S \ddagger$  ( $ES \rightarrow ES\ddagger$ )

- Effects of "approximation" and "orientation"

Residues in the protein can, additionally, provide acid-base or electrostatic catalysis



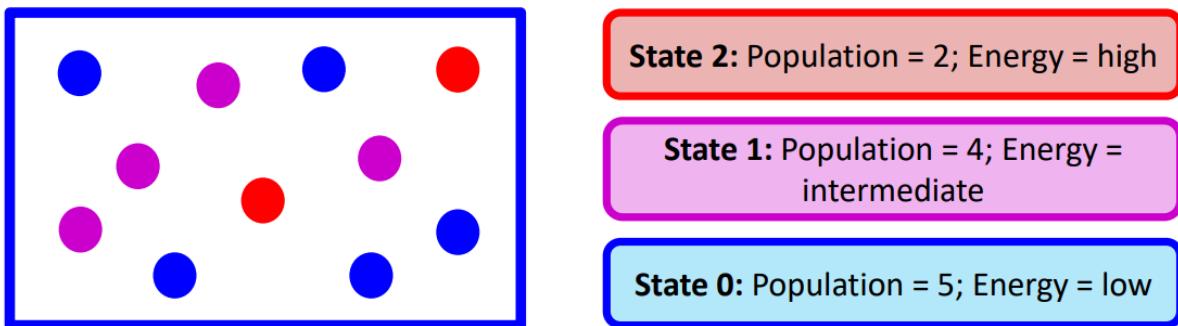
## Summary

- (Chemical) Kinetics tries to understand how the velocity of (chemical) processes.
- Reaction rates are proportional to the concentration of reactants powered to some exponent (reaction order).
- For elementary reactions reaction order correspond to the number of molecules involved
- Temperature affects rate constants following Arrhenius law
- From the microscopic point of view, another consequence of Boltzmann distribution
- Transition State theory provides a way to understand energies involved on kinetics
- Catalysis allows otherwise slow reactions to take place in reasonable times.
- Biological processes are always controlled by regulation of enzyme activity

## Biophysics seminar 1: Statistical Thermodynamics

We classify molecules depending on their kinetic energy into molecular states (molecules have different amounts of kinetic energy). Molecular states are defined by 2 properties:

- The population of particles in that state ( $N_i$ )
- The energy of the particles in that state ( $E_i$ )

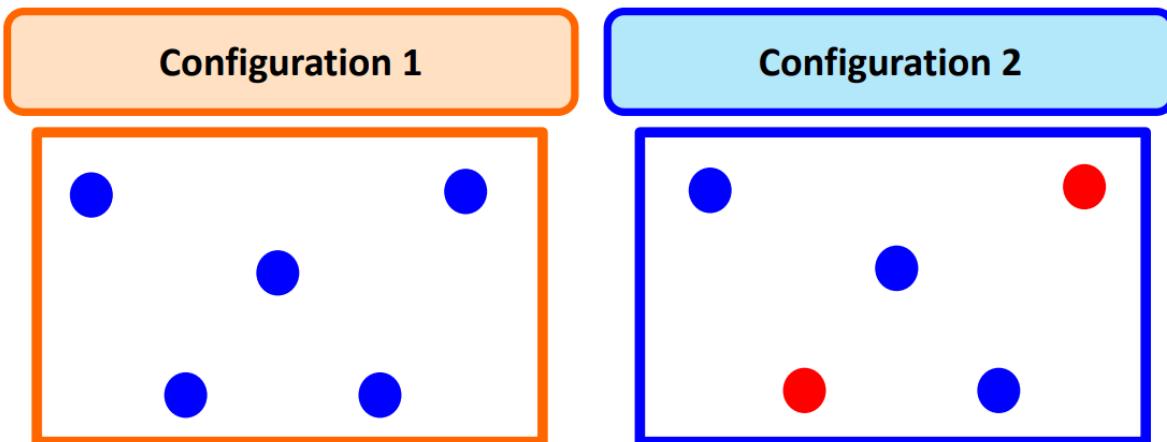


## Configurations

A configuration is one of the ways in which molecules in the system can be distributed among the different energy levels.

Imagine an isolated chemical system with 2 main properties:

- A given number of molecules ( $N$ )
- A given amount of energy ( $E$ )



We can think of configurations as macrostates. Then, the configuration with more available microstates should be the most likely.

In the example above, the second configuration has more available microstates:

- Configuration 1 has 1 microstate
- Configuration 2 has 10 microstates

Keep in mind that all microstates can happen with the same probability.

To calculate the microstates of a configuration, we can use the following formula ( $W$  == weight of the configuration == number of microstates):

$$W = \frac{N!}{N_0! N_1! N_2! \dots N_i!}$$

Where:

- $N$  is the number of particles in the system
- $N_0$  is the number of particles in the molecular state 0
- $N_1$  is the number of particles in the molecular state 1
- $N_2$  is the number of particles in the molecular state 2
- $N_i$  is the number of particles in the molecular state i

Once we have calculated the microstates available for a configuration, we calculate the entropy of that configuration using:

$$S = k \cdot \log(W)$$

**Exercise: Calculate the number of available microstates for the following configurations and their respective entropy.**

Keep in mind that energy from state 0 is smaller than the one of state 1, which is smaller than the one from state 2...

$$C1: \{N_0 = 20; N_1 = 0; N_2 = 0\}$$

$$C2: \{N_0 = 12; N_1 = 5; N_2 = 3\}$$

$$C3: \{N_0 = 7; N_1 = 7; N_2 = 7\}$$

**C1 has 1 microstate and entropy 0**

**C2 has 7.054.320 microstates and entropy  $9.45 \cdot 10^{-23}$**

**C3 has 399.072.960 microstates and entropy  $1.187 \cdot 10^{-22}$**

We see that configurations that spread more particles across their molecular states have more microstates.

If configurations with particles in many different levels are the most likely, why are there systems (for example, ice at very low temperatures) where most particles are at the same level with very low energy? Because the energy of the configuration is the addition of the energy of its particles. If the configuration has more energy than the system it cannot happen.

**Exercise. Estimate the energy of the configurations of the previous exercise.**

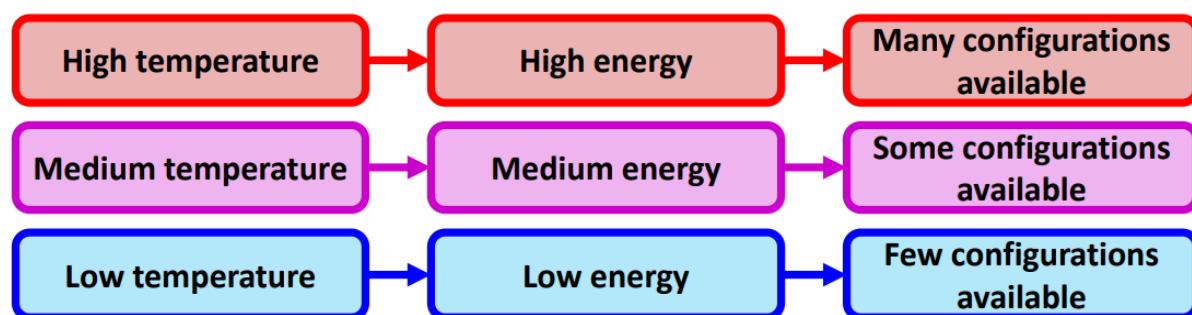
C1 has low energy -> Available for all systems

C2 has medium energy -> Available for some systems

C3 has high energy -> Available for almost no systems

In this explanation, we are focusing on the kinetic energy of the particles.

Can you think of any macroscopic property of a chemical system that is related to the kinetic energy of its particles? Temperature. It affects the system



The Boltzmann distribution explains the relation between populations in molecular states and temperature.

$$\frac{N_i}{N} = \frac{e^{-E_i/kT}}{\sum_n e^{-E_n/kT}}$$

Where:

- $N_i$  is the number of particles in a molecular state  $i$
- $N$  is the number of particles in the whole chemical system
- $E_i$  is the energy in the molecular state  $i$
- $k$  is the Boltzmann constant
- $T$  is temperature
- $E_n$  represents the energy of all the molecular states in the system

The Boltzmann distribution tells us the probability of finding a certain molecular state.

The denominator is called “molecular partition function” and it is represented with a “q”.

We will see how the value of the molecular partition function changes its value according with the temperature of the system:

- Being minimum at zero temperature
- Being maximum at infinite temperature

## Example

$$\text{C3: } \{N_0 = 7; N_1 = 7; N_2 = 7\} \longrightarrow N_2/N = 7/21 = 1/3$$

$$\frac{N_i}{N} = \frac{e^{-E_i/kT}}{\sum_n e^{-E_n/kT}}$$

$$\downarrow$$

$$\frac{N_2}{N} = \frac{e^{-E_2/kT}}{e^{-E_0/kT} + e^{-E_1/kT} + e^{-E_2/kT}}$$

Let's assume that:

- T = 298 K (room temperature)
- E0 = 0 kJ/mol; E1 = 2 kJ/mol; E2 = 4 kJ/mol
- k = 1.380649·10<sup>-23</sup> J/K
- **Haurem d'utilitzar el número d'avogadro per passar de mols a molècules.**
- També s'ha de passar de Kj a Joules

**Is this configuration possible?** The population of one state is inversely proportional to the energy of the state. High energy states are unlikely. **Per tal de que sigui possible, s'ha de cumplir la igualtat de la funció de Boltzmann.**

High energy state

Low energy state

$$\frac{\downarrow\uparrow N_i}{N} = \frac{e^{-\uparrow\uparrow E_i/kT}}{\sum_n e^{-E_n/kT}}$$

$$\uparrow\uparrow \frac{N_i}{N} = \frac{e^{-\downarrow\downarrow E_i/kT}}{\sum_n e^{-E_n/kT}}$$

Aquí podem veure com si incrementem l'energia del estat, tindrem una menor probabilitat (i al contrari).

As we decrease temperature, less states become available. At zero kelvin only one state is possible: the state of zero energy.

### Temperature = 0

$$\frac{N_i}{N} = \frac{e^{-E_i/kT}}{\sum_n e^{-E_n/kT}}$$

if  $E_i > 0 \rightarrow \lim_{T \rightarrow 0} \frac{E_i}{kT} = \infty$   
 if  $E_i = 0 \rightarrow \lim_{T \rightarrow 0} \frac{0}{kT} = 0$

$$\frac{N_i}{N} = \frac{e^{-\infty}}{\sum_n e^{-\infty}} = \frac{0}{0 + 0 + 0 + 0 \dots}$$

$$\rightarrow \frac{N_i}{N} = \frac{e^{-\infty}}{\sum_n e^{-\infty}} = \frac{0}{0 + 0 + 0 + 0 + 1 \dots}$$

$$\rightarrow \frac{N_i}{N} = \frac{e^{-0}}{\sum_n e^{-\infty}} = \frac{1}{0 + 0 + 0 + 0 + 1 \dots}$$

→ For any molecular state with  $E_i > 0$ :

$$\frac{N_i}{N} = \frac{e^{-\infty}}{\sum e^{-\infty}} = \frac{0}{1}$$

→ For the molecular state with  $E_0 = 0$ :

$$\frac{N_i}{N} = \frac{e^{-0}}{\sum e^{-\infty}} = \frac{1}{1}$$

Denominator is 1 because there is the state of minimum energy

When  $E_0=0$ , we have only one state (state of minimum energy).

### Temperature = $\infty$

$$\frac{N_i}{N} = \frac{e^{-E_i/kT}}{\sum_n e^{-E_n/kT}} \rightarrow \frac{N_i}{N} = \frac{e^{-0}}{\sum_n e^{-0}}$$

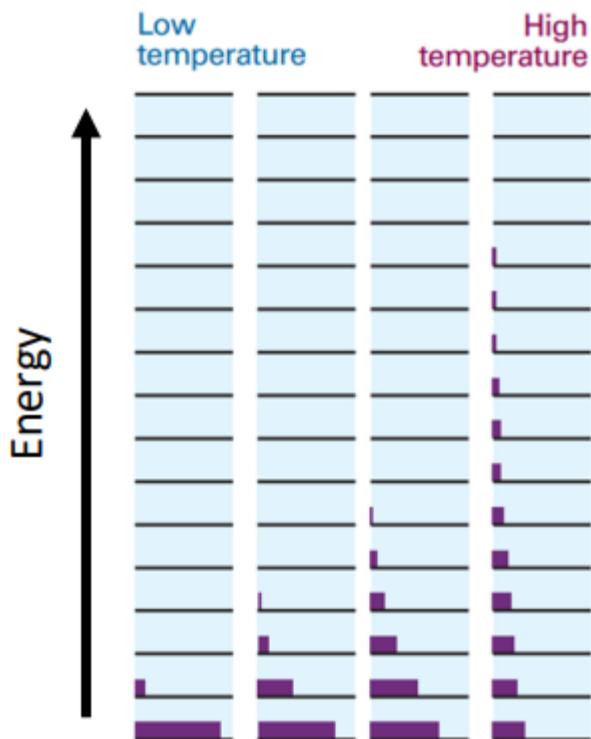
$$\frac{N_i}{N} = \frac{1}{1 + 1 + 1 + 1 \dots}$$

Here, all states have the same number of molecules and therefore they are all equally probable (this can only happen when temperature is infinite. So it is not possible)

### Temperature at microscopic scale

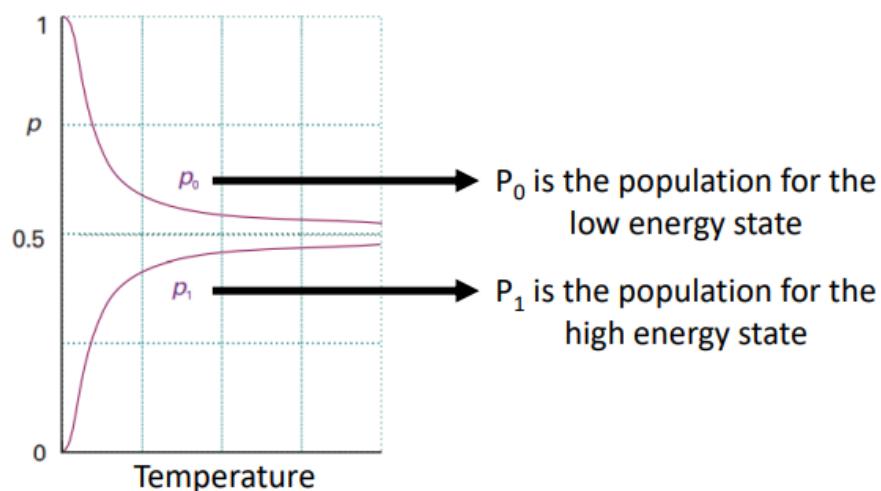
Besides being a measure of the average kinetic of particles, it can also be understood as a measure of molecule dispersion across molecular states

As you increase temperature, more molecular states are available, but you will always have most of your population in the low energy states. You can see this in the next plot, where each compartment represents a molecular state and purple bars represent how populated each state is.



**If temperature is infinite you will still have molecules in your system with low energy, even with energy = 0**

Imagine a system where only two states are possible, if temperature becomes infinite you will still have 50% of your molecules at low energy



## Seminar 4. Introduction to biopython

The most important object from the biopython.PDB module is the structure object. It represents the 3D structure of a molecule

This is how a structure object is organized:

*Structure - Model - Chain - Residue - Atom*

### Structure

It's the biggest object in the scale.  
It contains one or more models within itself

To create a structure object, you have to parse a PDB file:

```
>>> from Bio.PDB import *
>>> parser = PDBParser()
>>> structure = parser.get_structure("2dkt",
"2dkt.pdb")
```

We define the name and then the path in the *parser.get\_structure()* method.

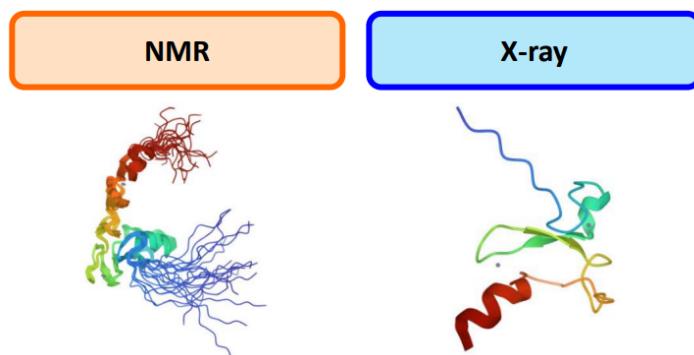
You can access the models by indexing the structure object, or iterate them by using the *get\_models()* function:

- *model = structure[0]*
- *for model in structure.get\_models()*

### Object

It contains one or more chains within itself

NMR experiments can produce several models for the same structure, while X-ray only produces one model.



Thus, the model object will only make sense for NMR structures. For X-ray structures there will be only one model.

You can access the chains by indexing the model object with the chain id, or iterate them by using the `get_chains()` function:

- `chain = model["A"]`
- `for chain in model.get_chains()`

## Chain

It contains one or more residues within itself.

Structures usually contain several chains, which can correspond with:

- Protein subunits for complexes
- DNA strands in a DNA molecule

Chains are identified with an ID which is a letter. You can check what protein subunit corresponds with each chain ID in the PDB webpage.

You can access the residues by indexing the chain object with the residue ID, or iterate them by using the `get_residues()` function:

- `residue = chain[94]`
- `for residues in chain.get_residues()`

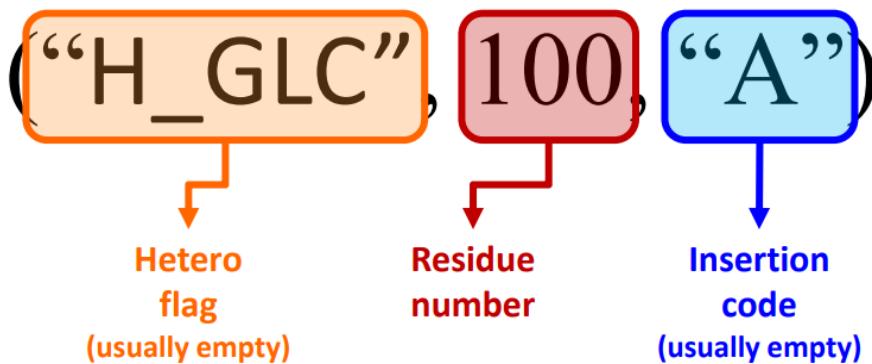
## Residue

It contains one or more atoms within itself

The residue object represents one monomer in the different macromolecules we can manipulate:

- Amino acids for protein structures
- Nucleotides for DNA or RNA structures

The residue object has a complicated ID, which is a tuple of 3 elements:



You can access the atoms by indexing the residue object with the atom identifier, or iterate them by using the `get_atoms()` function:

- `atom_CA = residue["CA"]`
- `for atom in residue.get_atoms()`

## Atom

The atom object is very useful to deal with molecule geometry, since it provides the coordinates for the position of each atom and allows the user to:

- Get coordinates from an atom
  - `CA.get_coord()`
- Calculate distances (in Angstroms)
  - `distance = atomA - atomB`

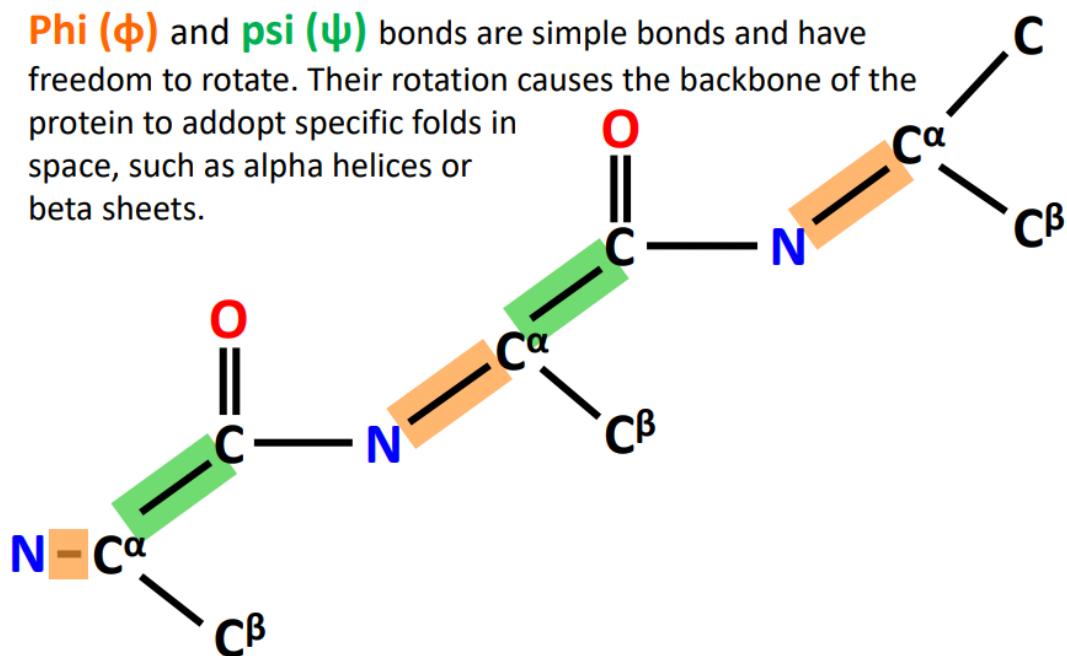
Calculate the dihedrals between 4 atoms:

```
>>> v1 = chain[res_id][atom_name1].get_vector()
>>> v2 = chain[res_id+1][atom_name2].get_vector()
>>> v3 = chain[res_id+1][atom_name3].get_vector()
>>> v4 = chain[res_id+1][atom_name1].get_vector()
>>> phi_dihedral = calc_dihedral(v1, v2, v3, v4)
```

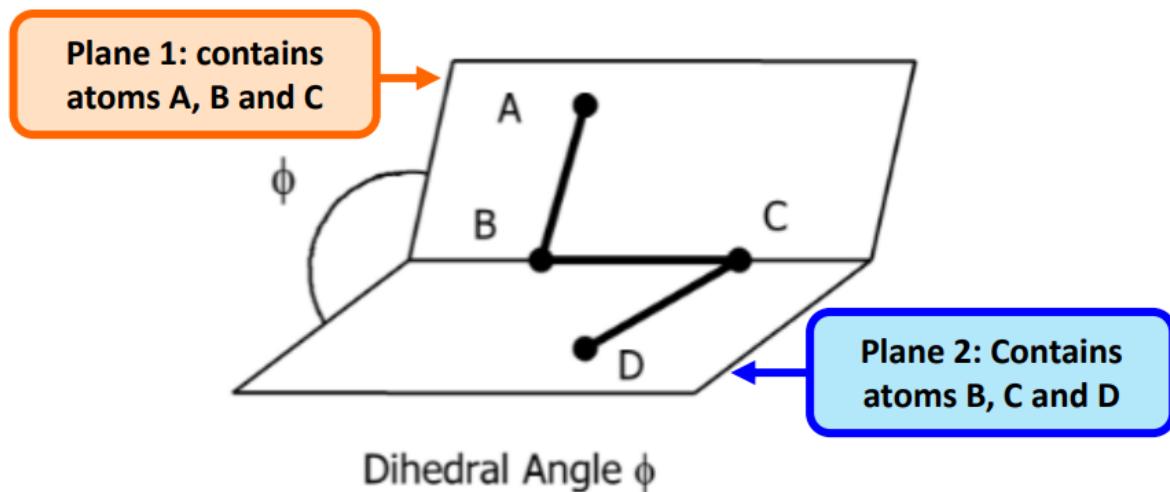
## Protein geometry

Peptidic bonds have resonance (they share the double bond with the C=O and C-N). This gives them double bond character and, hence, they can not rotate.

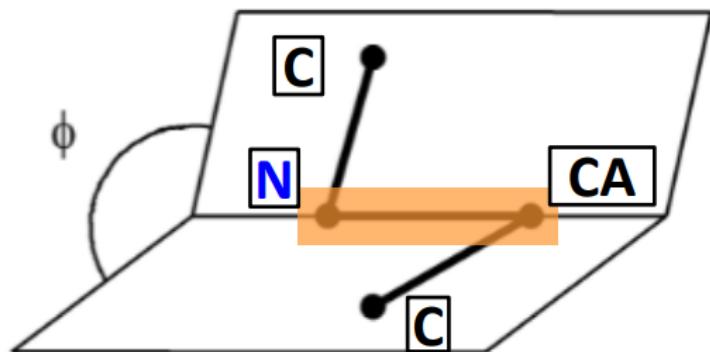
**Phi ( $\phi$ )** and **psi ( $\psi$ )** bonds are simple bonds and have freedom to rotate. Their rotation causes the backbone of the protein to adopt specific folds in space, such as alpha helices or beta sheets.



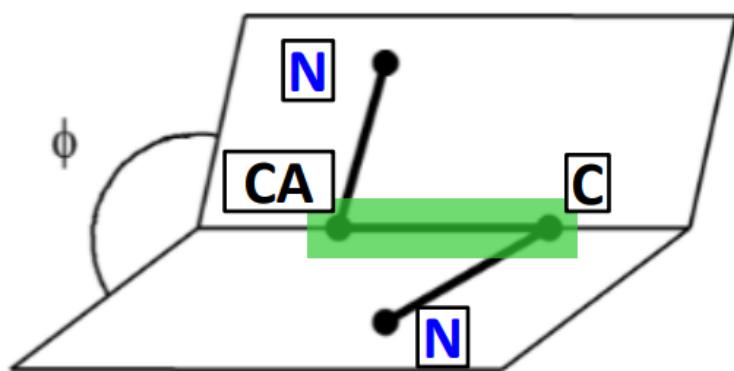
The dihedrals are the angle created by two planes. To calculate the dihedral of a bond, you select the atoms involved in the bond (B and C) and one atom before and another atom after (A and D).



For the Phi angle, the involved atoms are the carbonyl carbon and the nitrogen of one amino acid; and the alpha carbon and carbonyl carbon of the next amino acid.



For the Psi ( $\psi$ ) angle, the involved atoms are the nitrogen of one amino acid; and the alpha carbon, carbonyl carbon and nitrogen of the next amino acid.



The **Ramachandran** plot is a representation of the Phi and Psi dihedrals in a protein.

## Unit 2. Macromolecular energies

Urea and GdmCl unfold proteins. Because its structure is really similar to the peptide bond. The amount of energy to go from folded to unfolded is really small (10-15 kcal).

### Summary of energy components when a protein folds

Entropic (overall, it is favorable to fold the protein):

- Conformational entropy ( $\Delta G$  increases). If the initial state (unfolded) has more states, it is bigger (we are decreasing the entropy). So its not favorable
- Hydrophobic effect ( $\Delta G$  decreases) (Solvation entropy). The solvent has to make a cavity to surround the folded protein. Thus it has an entropic cost.  
The thing is that we are removing everything from the protein that does not interact correctly with the solvent, the hydrophobic residues (it hides inside the protein structure).

Enthalpy (overall, it is favorable to fold the protein):

- Van der Waals ( $\Delta G$  decreases)
- Hydrogen bonds ( $\Delta G$  remains equal)
- Electrostatic ( $\Delta G$  decreases)

Solvation enthalpy ( $\Delta G$  decreases):

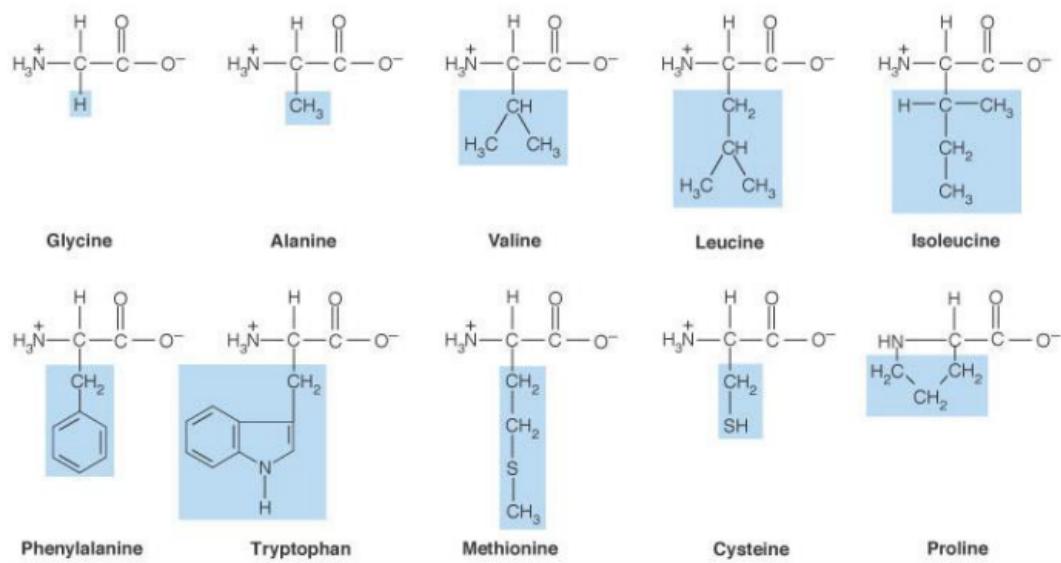
- Electrostatics
- Van der Waals

So, the only thing that goes against the folding is the conformational energy.

## Amino acids

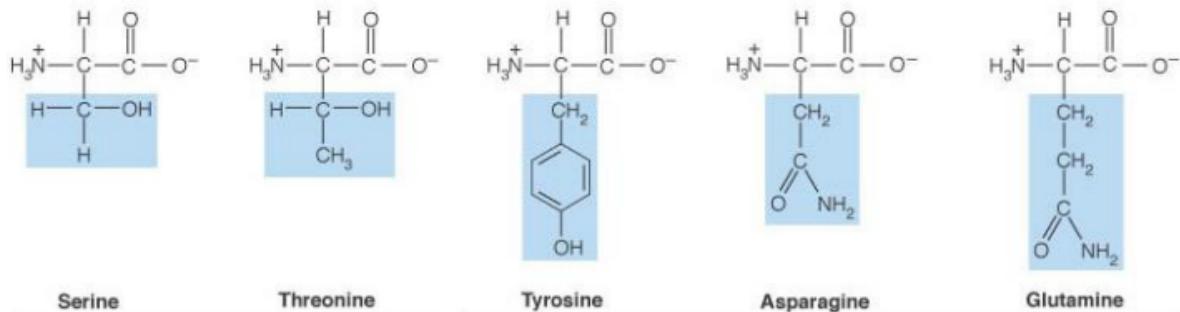
Neutral Nonpolar AA (Gly, Ala, Val, Leu, Iso, Phe, Trp, Met, Cys, Pro):

- They don't want to interact with the solvent because they are hydrophobic (they can't do polar interactions like hydrogen bonds and electrostatics). So, the only thing that they can do is Van der Waals and the hydrophobic effect.



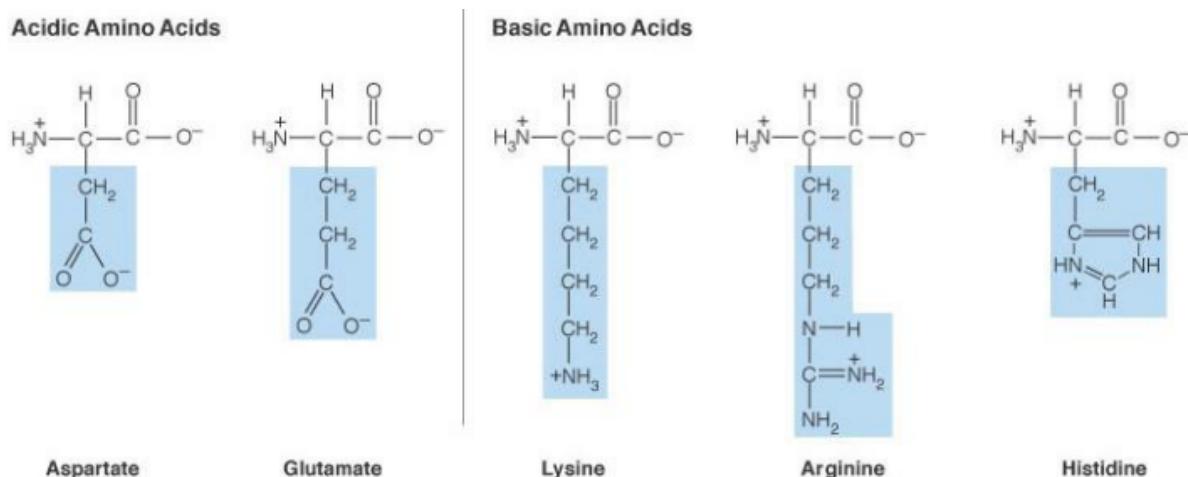
### Neutral Polar AA (Ser, Thr, Tyr, Asparagine, Glutamine):

- They can do polar interactions like Hydrogen Bonds and Van der Waals. They can not do electrostatics interactions because they have no charge.
- Tyrosine can be classified as Nonpolar



### Acidic/Basic AA (Lys, Arg, His, Aspartate, Glutamate):

- They can do electrostatic interactions, hydrogen bonds and Van der Waals
- Histidine can be in the other group (don't know which)



## Conformational entropy

Entropy decreases on folding

Folded structures are almost unique but unfolded ones are complex ensembles

- Increased chain flexibility increases the number of "unfolded" microstates (increases entropy).

Flexible residues (Gly) destabilize (positive entropy)

Rigid residues (Cysteine because it has the disulfuric bonds) stabilize.

## Hydrophobic effect

Solvent entropy increases when the number of contacts with non-polar (hydrophobic) residues decrease

Its effect is slightly higher than conformational entropy.

It is the driving force on protein folding or formation of protein complexes

Roughly proportional to molecular surface

- Aprox 30 cal/Å<sup>2</sup> apolar surface (100 res 4,860 Å<sup>2</sup> => -145 kcal/mol)

## Van der Waals energies

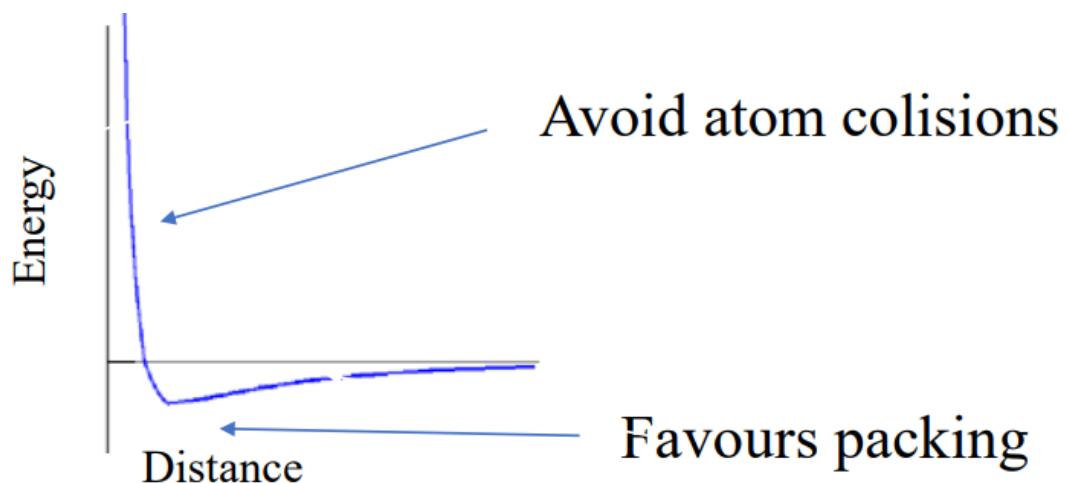
Caused by a combination of transient interactions between electronic clouds ("dispersion"), and repulsion between them.

When 2 atoms are getting closer, there is favorable interaction, but if they get really close, there is a repulsion.

Small but cumulative (~ -1000 kcal/mol average protein)

Control surface-surface interaction

This is not a very important force because all atoms have VdW interactions and, hence, a mutation is not going to affect the VdW force.



It favors the package, there are no spaces inside the molecule.

## Hydrogen bonds

Dipolar interactions (N, O, S)-H with electron pairs in N, O and S.

Directional: geometry defines the intensity of the interaction

Usual energy around 1-5 kcal/mol

In practical terms, H-bonds do not contribute to the stability of the protein.

Remember that we are always thinking about the difference between the folded and the unfolded state of the protein (not in the stability of the protein).

In the unfolded state, the protein is completely exposed to the solvent (water) and will make the maximum amount of H-bonds.

Water is the molecule that makes more H-bonds. Thus, H-bonds do not contribute.

If the folded state manages to maintain the number of H-bonds, we won't lose or win anything. If we lose H-bonds, we lose stability.

So, all H-bonds should be fulfilled in a folded structure.

## Electrostatics

In terms of numbers, it is the strongest interaction.

Largest enthalpic component of solvation. So, it is not just interaction with charged residues but also with the water.

Unpaired ionic residues are strongly destabilizers unless exposed to solvent and buried ionic pairs are strong stabilizers.

## Nucleic acids specifics

Poly-anions: strong electrostatics (more important than in proteins):

- Repulsion between main-chain phosphates
- Highly sensitive to ionic environment

Base-base interactions:

- Hydrogen bonds:
  - Basis of all recognition events (interchain, external) but does not contribute to stability as mentioned before.
  - Strong in gas phase, lower in solution
- Stacking (interaction between 2 different steps of the DNA):
  - It is a mix of VdW (majority) and electrostatics
  - Major contribution to stability (~ - 2 kcal/mol/step)
  - Highly cooperative

# Energy Evaluation



Quality

- Different approaches: different levels of accuracy
  - Quantum mechanics
    - Rigorous, allow chemical processes, limited to small systems
  - Classical mechanics
    - Less rigorous but accurate, no size limit, does not allow for chemical transformations
  - Statistical/coarse-grained potentials
    - Knowledge-based, accurate for low resolution and non-realistic representations
  - Empirical scores
    - Simple parameter set fitted to experiment

## Quantum mechanics

It's rigorous and universal. But it is too costly and can only be applied to small systems (active sites, chemical transformations).

But it's the best way to compute energy. Allows for reorganization of bonds and non-equilibrium structures. To evaluate rate constants, the analysis of transition states and reorganization of bonds is required.

Classical Force Fields assume equilibrium and fixed structures. They can define that a structure is "chemically correct", but can not identify correct conformations (can't distinguish between folded and unfolded structures). This can only be obtained by evaluating entropy or comparison of known structures (statistical potentials).  
CF can not calculate solvation energy because solvation energy contains entropic terms.

## Classical mechanics

It uses empirical approximations (require parameterization), it is computationally very efficient and has no size limit

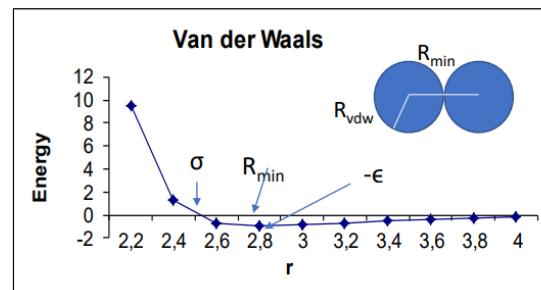
But it's less accurate than quantum mechanics and solvation and entropy cannot be included!!

The energy can be decomposed in:

- Bonded terms:
  - E str, E binding, E torsion
- Non bonded terms:
  - Electrostatic and VdW
- Others

# Van der Waals

- Empirical
- “Lennard-Jones” or “6-12” potentials
- Several alternatives



$$E_{vdw} = 4\epsilon \left( \left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^6 \right) \quad E_{vdw} = \epsilon \left( \left(\frac{R_{min}}{r}\right)^{12} - 2 \left(\frac{R_{min}}{r}\right)^6 \right) \quad E_{vdw} = \frac{A}{r^{12}} - \frac{C}{r^6}$$

- Parameters for atom pairs obtained from combination of atomic ones

$$\epsilon_{ij} = \sqrt{\epsilon_i \epsilon_j} \quad \epsilon_{ij} = \sqrt{\epsilon_i \epsilon_j} \quad A_{ij} = (\epsilon_i \epsilon_j)^{1/2} (R_{vdw}^i + R_{vdw}^j)^{12}$$

$$\sigma_{ij} = \sqrt{\sigma_i \sigma_j} \quad R_{min,ij} = R_{vdw,i} + R_{vdw,j} \quad C_{ij} = 2 \times (\epsilon_i \epsilon_j)^{1/2} (R_{vdw}^i + R_{vdw}^j)^6$$

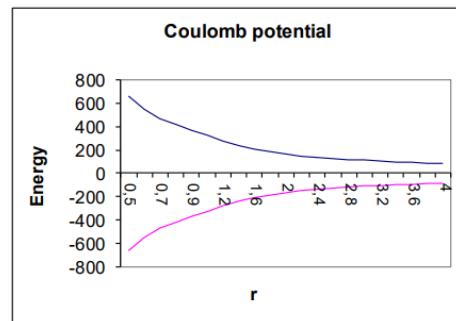
# Electrostatics

- Stronger interaction
- All atoms have a partial charge
  - Dipolar interactions (directional)
  - Hydrogen bond energies
- Coulomb's Law

$$E_{elec} = \frac{1}{4\pi\epsilon_0} \frac{q_i q_j}{\epsilon_r r} = 332.16 \frac{q_i q_j}{\epsilon_r r}$$

- $q_i, q_j$ : Partial charges
- $\epsilon_r$ : Relative dielectric constant
- $r$ : distance
- Dielectric strategies
  - Fixed value: 1 for vacuum, 80 for fully solvated
  - Linear:  $\epsilon_r = k \cdot r$
  - Mehler-Solmajer

$$\epsilon_r = \frac{86.9525}{1 - 7.7839 e^{-0.3153 r}} - 8.5525$$



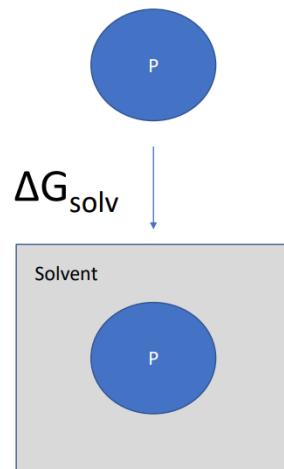
## Solvation energy

Defined as the energy to move a solute from vacuum to solvent

Two major components:

- Electrostatic (Enthalpy)
- Hydrophobic (Solvent entropy)
  - Entropy cannot be calculated in single system

Key to understand macromolecular stability and binding



### Approaches to the electrostatic component

Explicit solvent

- Solvent molecules are present in the system

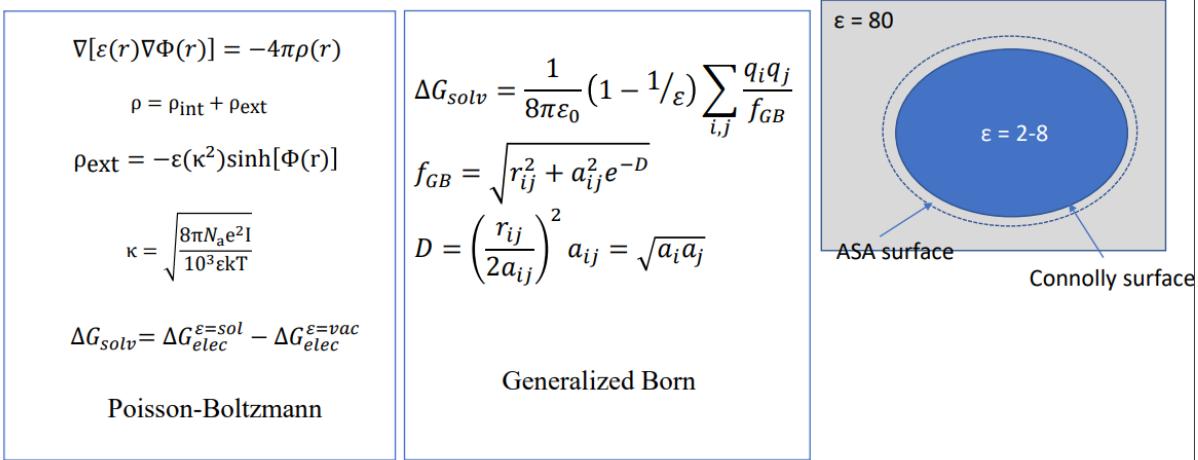
$$E_{\text{solv}} = \sum_i^{\text{solute}} \sum_j^{\text{solvent}} \frac{q_i q_j}{\epsilon r_{ij}}$$

- However, solvent molecules reorient, a single value is meaningless, values should be averaged over time

$$E_{\text{solv}} = \left\langle \sum_i^{\text{solute}} \sum_j^{\text{solvent}} \frac{q_i q_j}{\epsilon r_{ij}} \right\rangle$$

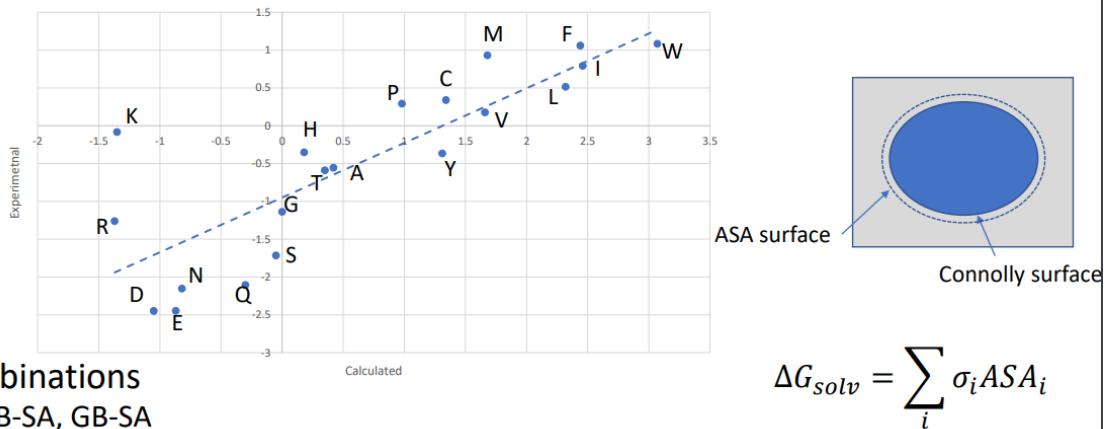
## Approaches to the electrostatic Solvation

- Implicit solvent. Classical continuous models



## Approaches to the hydrophobic component

- Experimental  $\Delta G_{solv}$  is (almost) linearly correlated with ASA Surface

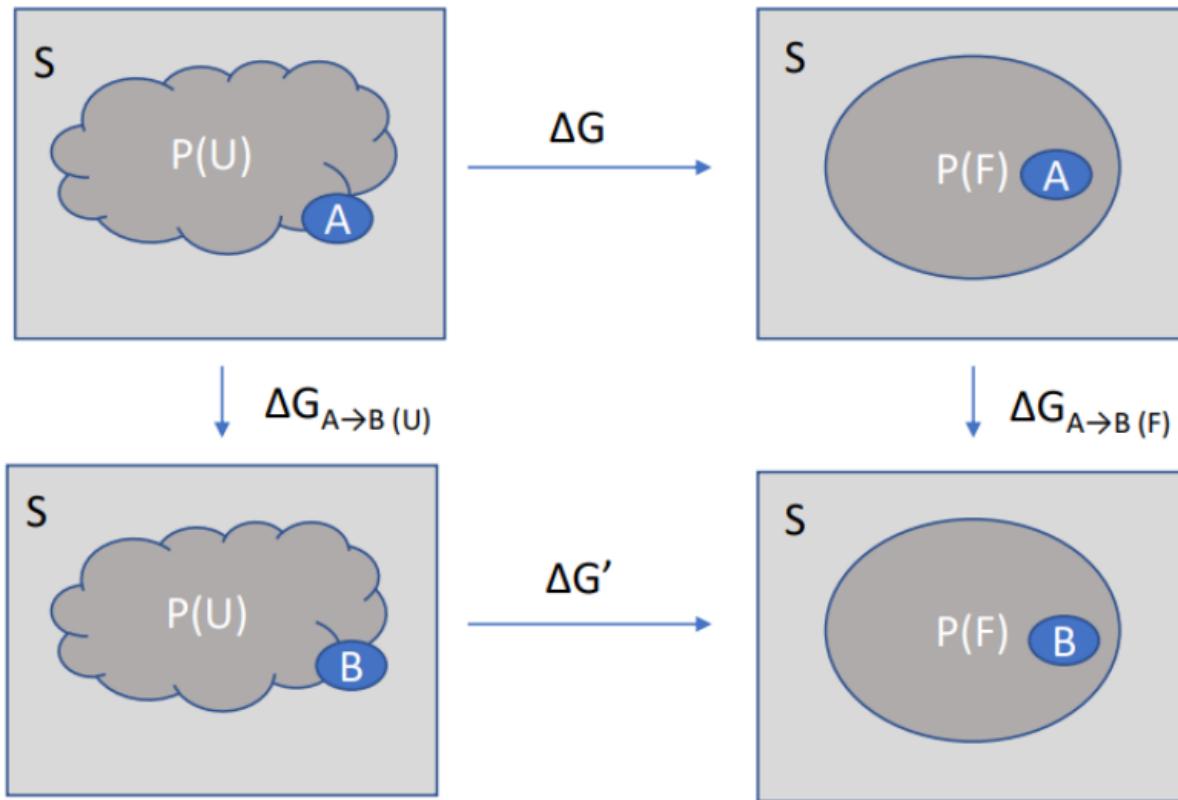


# Which level and when?

- Quantum Mechanics
  - Very accurate energies
  - Chemical processes
  - Distorted structures
- Classical Mechanics
  - Near-equilibrium structures
  - Conformational Changes, Folding
  - Refinement of experimental structures
- Statistical/knowledge based
  - Non natural (simplified) representations
  - Scoring based on similarity
  - Uses “non-energy” terms
- Empiric scores
  - Closely related series of compounds

## Seminar 5: Mutational Analysis

Before understanding the effect of a mutation in the stability of a protein, we are going to understand protein folding and protein stability.



$\Delta G$  is a state function. Thus, it's the same going from:

- P(U)<sub>A</sub> to P(F)<sub>A</sub>
- P(U)<sub>A</sub> to P(U)<sub>B</sub> to P(F)<sub>B</sub> to P(F)<sub>A</sub>

We can think of the folding process as a chemical reaction, where the reactants are the unfolded state and the products are the folded state. As any chemical reaction, there are thermodynamic variables associated with this process:

- Change in free energy
- Change in enthalpy
- Change in entropy

## Properties of the folded and unfolded state

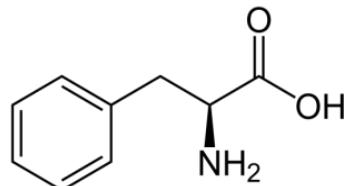
### Unfolded:

- All residues are exposed to solvent
- All H-bonds are made with water
- Many protein conformations are available

### Folded:

- Some residues are exposed to the solvent and some others are buried
- Many H-bonds are made within the protein
- Few protein conformations are available

**Example 1: A phenylalanine in a peptidic chain goes from unfolded to folded in a buried position**



	Unfolded	Folded	Overall $\Delta G$
Polar int.	$-\Delta G$	$-\Delta G$	$0 \Delta G$
Electrostatic int.	$0 \Delta G$	$0 \Delta G$	$0 \Delta G$
VdW int.	$-\Delta G$	$-\Delta G$	$0 \Delta G$
Solvation	$+\Delta G$	$-\Delta G$	$-\Delta G$
Total	$\Delta G$	$\Delta G$	$\Delta G$

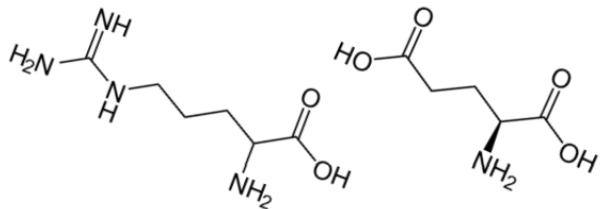
**Polar interaction:** We check if it can make H bonds. It can make it in both stages in this case.

**Electrostatic:** La carga es 0, per tant 0. Es necesita dues molecules amb carrega

**VdW:** Sempre posem negatiu. Totes les molecules en poden fer

**Solvation:** Es hidrofobic, per tant si esta desplegat sera inestable.

**Example 2: An arginine and glutamate pair go from unfolded to a buried position**



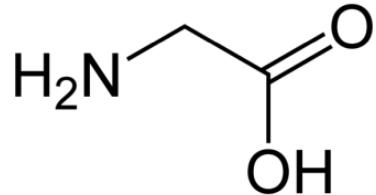
	Unfolded	Folded	Overall $\Delta G$
Polar int.	- - $\Delta G$	- - $\Delta G$	0 $\Delta G$
Electrostatic int.	0 $\Delta G$	- $\Delta G$	- $\Delta G$
VdW int.	- - $\Delta G$	- - $\Delta G$	0 $\Delta G$
Solvation	- - $\Delta G$	0 $\Delta G$	+ + $\Delta G$
Total			nu se $\Delta G$

**Polar interaction:** Poden fer H bonds entre ells!

**Electrostatic int:** Sí que poden fer. En el unfolded no interaccionen, pero en el folded si que interaccionen.

**Solvation:** Quan interaccionen amb l'aigua, serà favorable (son hidrofílics i a sobre tenen carga, que li mola molt a l'aigua). Quan esta folded, posem 0 perque la carga elèctrica de les dues molecules son oposades (es neutralitzen).

**Example 3: 3 glycines go from unfolded to an exposed position**



	Unfolded	Folded	Overall $\Delta G$
Polar int.	-	-	0
Electrostatic int.	0	0	0
VdW int.	-	-	0
Solvation	+	+	0
Total			++

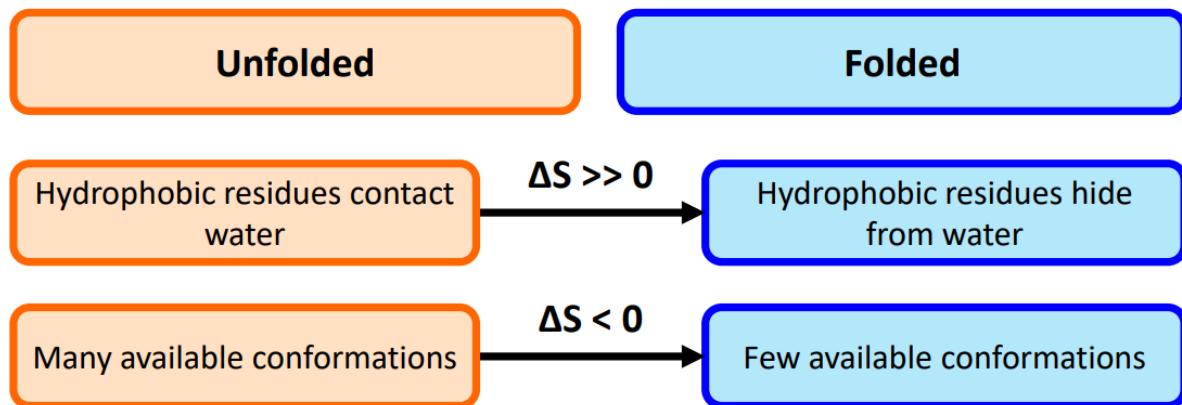
Poden fer H bonds.

No tenen carga

Total + perque les glicines tenen més tendencia a "bend". Son més flexibles (perque la side chain és més petita).

## Glycines and conformational flexibility

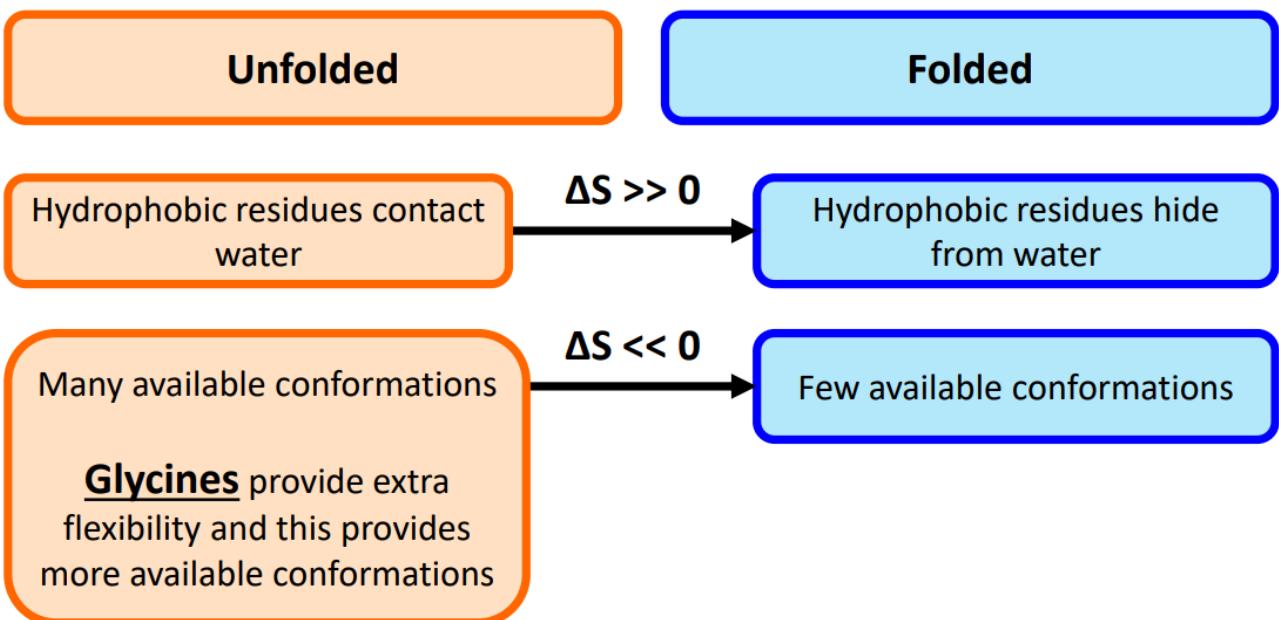
During folding, the hydrophobic effect increases entropy, while the loss of conformational flexibility decreases entropy.



Overall,  $\Delta S > 0$ . This favors the spontaneous folding.

Glycines don't have side chains (one lone hydrogen). This enables glycine to adopt many rotation angles that other amino acids cannot undergo.

- The side chains of AA are big and they can clash with the rest of the protein.
- A hydrogen atom is way smaller than any other side chain. Then, it will be able to adopt many conformations where the hydrogen can fit, but other side chains don't.



Since  $\Delta S$  goes down,  $\Delta G$  goes up and the folding process becomes less favorable.

## Mutation Analysis: The concepts

The objective of a mutation analysis is to determine the effect of a mutation in the stability of the protein. We quantify this with the  $\Delta\Delta G$ .

We represent the change in free energy for the folding of a protein as  $\Delta G$ . Remember that the  $\Delta$  symbol is used to indicate change. Then, the change in  $\Delta G$  when a mutation occurs is the  $\Delta\Delta G$  ( $\Delta\Delta G = \Delta G' - \Delta G$ ).

The  $\Delta\Delta G$  tells you how the folding of one protein is affected when a mutation happens.

If  $\Delta\Delta G < 0$

The mutant protein is more stable, the  $\Delta G$  of the mutant is lower

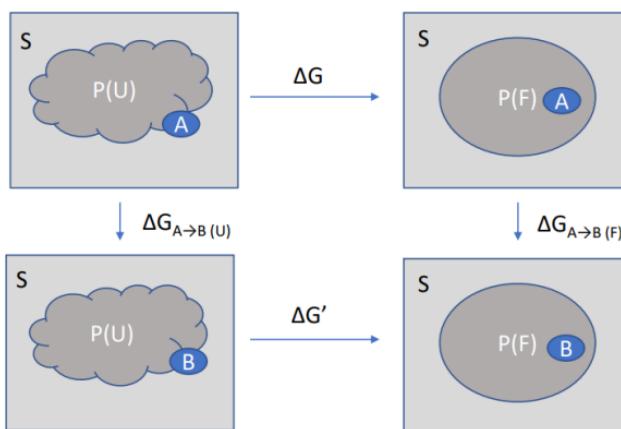
If  $\Delta\Delta G > 0$

The mutant protein is less stable, the  $\Delta G$  of the mutant is higher

As we said before, since  $\Delta G$  is a state function, it is pathway independent and this enables us to make the following calculations:

$$\Delta\Delta G_{ab} = \Delta G_b - \Delta G_a$$

$$\Delta\Delta G_{ab} = \Delta G_{ab(F)} - \Delta G_{ab(U)}$$



$$\Delta\Delta G_{ab} = \Delta G_{ab(F)} - \Delta G_{ab(U)}$$

$\Delta G_{ab(F)}$  is the  $\Delta G$  of going from wild type to mutant in the folded protein

$\Delta G_{ab(U)}$  is the  $\Delta G$  of going from wild type to mutant in the unfolded protein

We can represent  $\Delta G_{ab}$  (either for folded or unfolded proteins) as an addition of the following terms:

- Internal energy of the mutated residue ( $\Delta\Delta G_{internalAB}$ )
- Internal energy of the protein ( $\Delta\Delta G_{internalP}$ )
- Interaction between the mutation and the protein ( $\Delta\Delta G_{interP,AB}$ )
- Solvation of the protein ( $\Delta\Delta G_{solvationP}$ )
- Solvation of the mutated residue ( $\Delta\Delta G_{solvationAB}$ )

If we write this equation splitting  $\Delta G_{ab}$  into all the terms we defined in the previous slide, we get this expression:

$$\begin{aligned}\Delta\Delta G_{A \rightarrow B} = & \Delta\Delta G_{A \rightarrow B(F)}^{internalAB} + \Delta\Delta G_{A \rightarrow B(F)}^{internalP} + \Delta\Delta G_{A \rightarrow B(F)}^{interP,AB} + \Delta\Delta G_{A \rightarrow B(F)}^{solvationP} + \Delta\Delta G_{A \rightarrow B(F)}^{solvationAB} \\ & - \Delta\Delta G_{A \rightarrow B(U)}^{internalAB} - \Delta\Delta G_{A \rightarrow B(U)}^{internalP} - \Delta\Delta G_{A \rightarrow B(U)}^{interP,AB} - \Delta\Delta G_{A \rightarrow B(U)}^{solvationP} - \Delta\Delta G_{A \rightarrow B(U)}^{solvationAB}\end{aligned}$$

These terms cancel each other out. We assume that there are no differences in these energies between the folded and the unfolded state.

Thus, we obtain:

$$\Delta\Delta G_{A \rightarrow B} = \Delta\Delta G_{A \rightarrow B(F)}^{interP,AB} + \Delta\Delta G_{A \rightarrow B(F)}^{solvationAB} - \Delta\Delta G_{A \rightarrow B(U)}^{interP,AB} - \Delta\Delta G_{A \rightarrow B(U)}^{solvationAB}$$

We can decompose the interaction terms into the different interactions that can be made by one amino acid and the rest of the protein:

- Polar interactions
- Electrostatic interactions
- Van der Waals interactions

Thus, we can use a similar table that we saw before

	WT	Mutant	Overall $\Delta G$
Polar int.			
Electrostatic int.			
VdW int.			
Solvation			
Total			

## Exercises

Analyze qualitatively the energetic consequences of the following sequence variants.

Consider 2 cases:

- The original residues are exposed to solvent
- The original residues are buried

### Q-S to E-S, exposed

Glutamin (Q) & Serine (S) & Glutamate (E)

Folded protein	WT	Mutant	Overall ΔG
Polar int.	-	-	0
Electrostatic int.	0	0	0
VdW int.	-	-	0
Solvation	-	--	-
Total			-

Polar:

Electrostatic: Negatiu quan son opositius

Unfolded protein	WT	Mutant	Overall ΔG
Polar int.	0	0	0
Electrostatic int.	0	0	0
VdW int.	-	-	0
Solvation	-	--	-
Total			-

Polar interaction es 0 perque en el unfolded state no hi ha interacció entre els AA de la proteïna.

Result = 0 (menos menos menos)

**Q-S to E-S, buried**  
 Glutamin (Q) & Serine (S) & Glutamate (E)

Folded protein	WT	Mutant	Overall ΔG
Polar int.	-	-	0
Electrostatic int.	0	0	0
VdW int.	-	-	0
Solvation	+	++	+
Total			+

Extra + porque el glutamic acid está cargado

Unfolded protein	WT	Mutant	Overall ΔG
Polar int.	0	0	0
Electrostatic int.	0	0	0
VdW int.	-	-	0
Solvation	-	--	-
Total			-

Extra - porque a la agua le encanta lo que está cargado.

Resultat= Mutation decreases protein stability ((+) + (-))

**W-W to G-G, buried**

Tryptophan (W) &amp; Glycine (G)

Folded protein	WT	Mutant	Overall ΔG
Polar int.			
Electrostatic int.			
VdW int.			
Solvation			
Total			

Unfolded protein	WT	Mutant	Overall ΔG
Polar int.			
Electrostatic int.			
VdW int.			
Solvation			
Total			

Remember that Tryptophan (W) is the AA with the biggest side chain and Glycine (G) is the AA with the smallest side chain.

G-G has way more possible conformations in the unfolded state than W-W. This makes  $\Delta S$  get lower in the G-G pair, thus leading to a higher  $\Delta G$ .

If the mutant has higher  $\Delta G$ , then the mutation decreases protein stability and  $\Delta\Delta G > 0$

### F-K to E-K, exposed

Folded protein	WT	Mutant	Overall ΔG
Polar int.			
Electrostatic int.			
VdW int.			
Solvation			
Total			

Unfolded protein	WT	Mutant	Overall ΔG
Polar int.			
Electrostatic int.			
VdW int.			
Solvation			
Total			

**Is protein stability good or bad for biological systems?**

**Is maximum stability for proteins the best for biological systems?**

Superstable proteins are a problem for the cell, because it cannot degrade them, leading to the accumulation of such proteins in the cell. Alzheimer's disease and bovine spongiform encephalopathy.

# Folding

Anfinson did the first experiment related to the folding process. He used urea (to unfold) and a reducing agent (to remove the disulfide bonds) to unfold the ribonuclease A.

Then, when performing a dialysis to remove the urea and the reducing agent, the activity of the ribonuclease was recovered.

The activity of an enzyme in a solution with urea is equal to 0.

At that time, they thought that the cell had enough information to correctly fold the protein. But he demonstrated that the protein folds by itself.

## The folding problem:

Proteins and Nucleic Acids are synthesized as linear flexible polymers and they fold by themselves.

Proteins are kept unfolded before reaching their final location

Folding is spontaneous, quick and cooperative :

- However, protein can get trapped in "misfolded" structures (Amyloids, "folding diseases") and it is nearly irreversible.
- Proteins unfold/fold reversibly under specific conditions
- Folding "in vivo" is efficient

Nucleic acids unfold/fold reversibly (better than proteins)

Changes in media conditions (pH, T, solvents,...) denature (unfold) proteins and nucleic acids

Levinthal paradox (calculates how long it would take for the protein to fold if it had to test all possible conformations):

- 12 conf x torsion → 144 confs for each residue. 100 residues →  $12^{200}$  conformations ( $6.8 \times 10^{215}$ )

Most macromolecules have a defined folded structure (but not all)

Folded state is expected to be thermodynamic minimum ( $\Delta G$ )

Folding should follow a preferred kinetic pathway. So the folding is not at random, there is one mechanism that the protein follows.

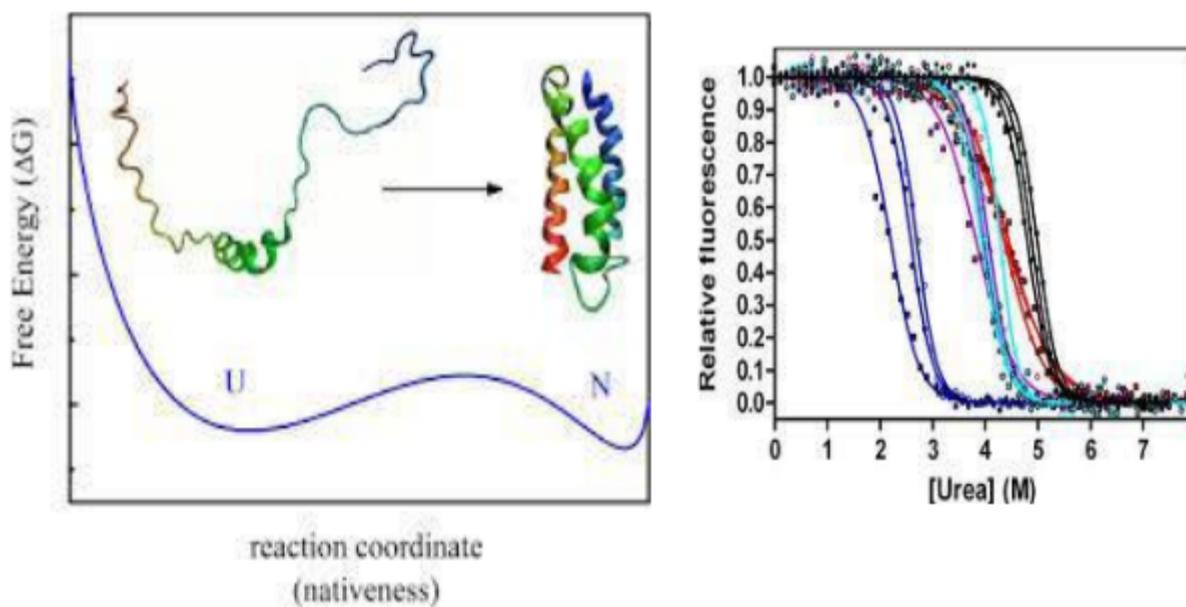
Fluorescence is an indicator of the state of the protein.

The aromatic residues emit fluorescence

Fluorescence is extremely dependent of the environment of the amino acid:

- If the amino acid is inside the protein there is a big fluorescence
- If it's outside (in contact with the water), there is small fluorescence.

So, we can measure the state of the protein.

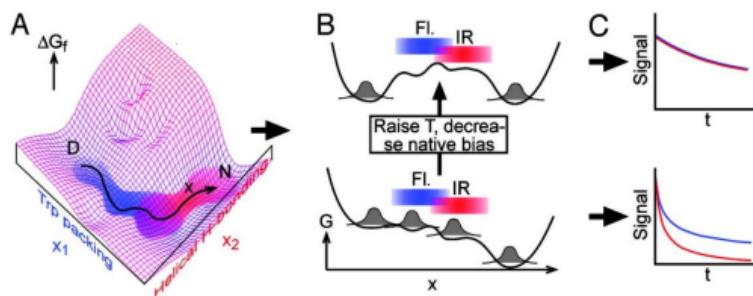


### Exceptions to the rule

The 2 states model (there are no intermediates) fails for some proteins!

Downhill folding

- No energy barrier
- Usually small  $\alpha$ -rich proteins



Intrinsically disordered proteins (they don't fold. The  $\Delta G$  of the folding process is positive):

- Fold on binding to targets (they only fold when they find the correct partner to form the complex)
  - Conformational selection (their conformation depends on the partner)
- Alternative folded structures
- Usually studied by NMR
- Hybrid proteins

## Entropic funnel model

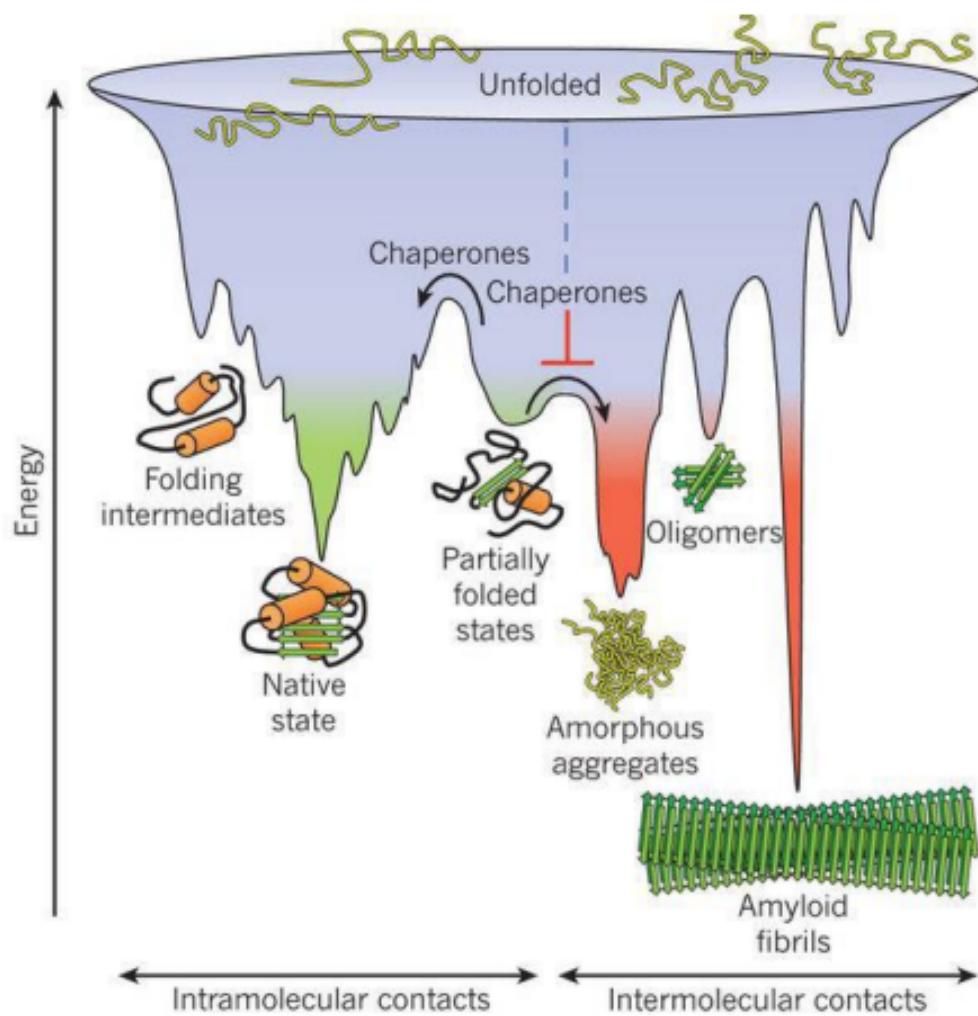
Concept of “Energy Landscape”.

In the unfolded estate there are a lot of possibilities (microstates), there is a high entropy.  
When we start folding the protein, we lose entropy but we gain enthalpy (because there are more interactions such as H bonds, WdV, electrostatic...).  
So, the free energy is improving because the enthalpy has a higher roll (we should also add the entropy of the solvent, which is not represented in this plot).

Entropy and enthalpy compensate building a “funnel”

Lower energy limits the conformational space

The model can be adapted to most cases



## Time scales

Secondary structure formation << 0.01 s

“Dry Molten globule”/folding intermediates 0.05-0.1 s

- “molten globule” is a kind of folding intermediate that has structure but is not compacted (because it's full of water).

Compaction 0.1 – 1 s

Isomerization Pro, -SS- formation (1-5 s)

## *In vivo* folding

*in vivo* folding is more efficient but not quicker

Relies in keeping unfolded states stable

Chaperons

- Stabilize unfolded states, avoiding aggregation (when transporting, for example)
- Non-specific
- Also help to fold

Helper enzymes

- Accelerate particularly slow changes
- Peptidyl-prolyl cis-trans isomerases
- Protein disulfide isomerases

## Bioinformatics in folding

Protein Structure prediction

- Ab-initio MD simulations (atomistic and coarse-grained)
- Contact predictions -> AlphaFold
  - Fold recognition
- Comparative modelling
- Fragment based prediction
- Disorder prediction
- Now almost superseeded by AI!!

Rational Design

Directed evolution

De-novo design

# Binding

## Molecular recognition

Everything happens in biology because 2 molecules recognize each other and form a complex. All macromolecules work through recognition processes

- Protein-ligand
  - Enzymes, membrane receptors, transport proteins, Drugs
- Protein-protein
  - Regulation of enzyme activity (signal transduction), multi-subunit protein and complexes
- Protein-NA
  - NA metabolism, gene regulation
- NA-Ligand
  - Drugs
- NA-NA
  - Transcription, replication, protein synthesis, ...

Recognition is selective

- Depends on the participating groups.

Recognition is dynamic

- Induced fit / Conformational selection
- Complexes can be permanent, but most of them are transient. They are formed when they are necessary.

## Energy considerations

No new energetic considerations

- Entropic (Conformational, hydrophobic)
- Enthalpic
  - Vdw (Shape and contacts)
  - Hbond (Define geometry, must be completed)
  - Electrostatic (Severe solvation penalty, high energies)

Now we are comparing the solvation in the complex state and the solvation when the proteins are separated.

Structural complementarity

- Both shape and interactions
- The basis of most protein-protein complex predictions

Complex formation implies to bury new interactions

- Unstable (hydrophobic) surfaces in water may indicate binding regions.

## DNA-DNA specialized interactions

We can have duplex, triplex and quadruplex.

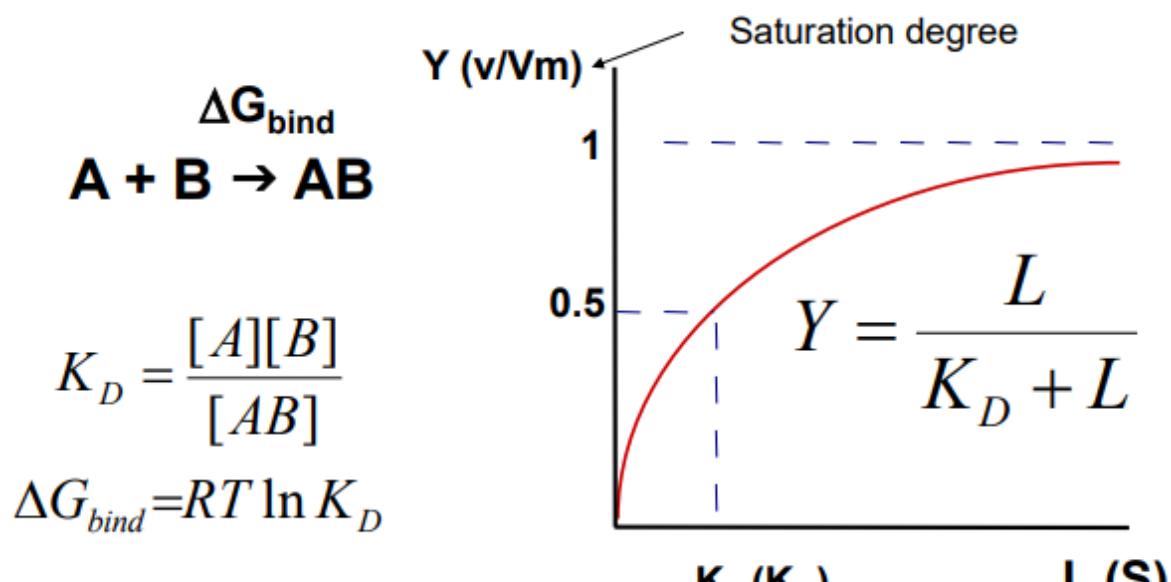
## Protein-DNA specialized interactions

### Thermodynamic cycle for binding

We want to determine the free energy of the binding process when forming complexes.

Effects of mutations on binding come from the differences due to amino acid changes

### Binding (reaction rates) measurements

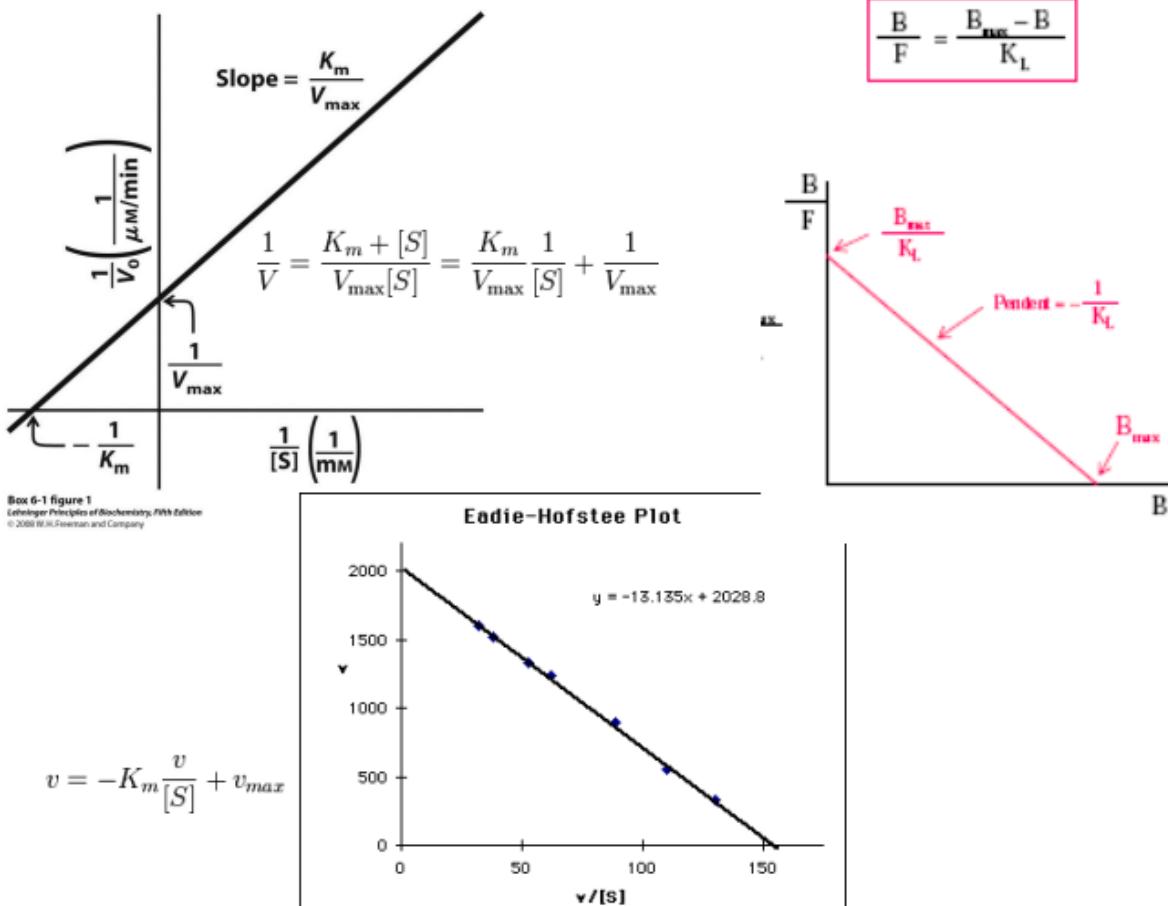


Kd is a dissociation constant !!

L is "free" ligand

Less Kd means stronger binding.

Kd is the concentration of substrate that gives 50% of the saturation.



## Other parameters

**K<sub>i</sub>**: Inhibition constants (usually from kinetic measurements). True KD for inhibitory compounds binding to the enzyme

**IC<sub>50</sub>**: Concentration of inhibitor that induces 50% of the effect (Incorrect but usual!)

$$IC_{50} = K_I \left( 1 + \frac{S}{K_M} \right)$$

For a competitive inhibitor

## **Molecular interaction potential**

A 3D potential similar to electrostatic potential but indicating binding energy around receptor

Binding energies for a given probe are calculated in a 3D grid placed around the receptor

Classical (force field) or QM approaches can be used

Defines spatial regions where binding is favorable

The combination of MIPs for simple probes define a “pharmacophore”

MIPs are calculated (once) over a regular 3D grid

- Contours made on grids allows for visual analysis
- Interaction of complex molecules are evaluated by interpolation on grid

## **Summary**

1. Molecular recognition is the basis of any biological function
2. Recognition is a consequence of the 3D structure
  - a. Structural complementarity
  - b. Match of interacting groups
3. Recognition is dynamic (conformational selection applies)
4. Same types of interactions and energy considerations as macromolecular stability
5. Evaluation of Binding requires to consider both unbound and bound states
6. Molecular interaction potentials help to understand binding.

# Dynamics

## Conformational ensembles

“Ensemble”: set of structures that represents ALL possible microscopic states of the system

Thermodynamics can be deduced from the average of “ensemble” properties.

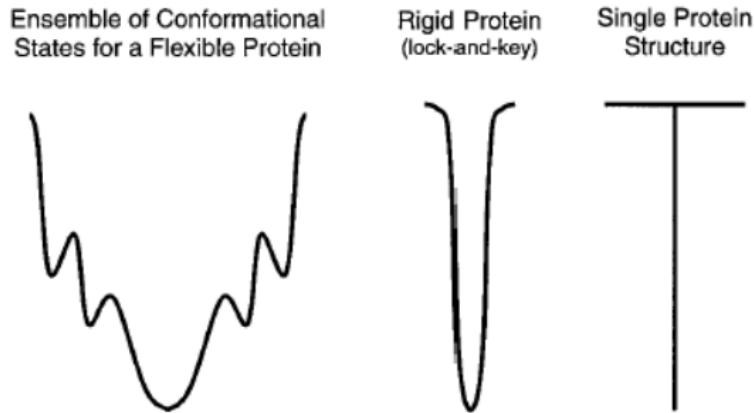
Ensembles help to understand macromolecules behavior

- Induced fit vs conformational selection (when binding to a molecule, we select a state of the funnel). Maybe a conformation is not favorable but this can change when another protein binds.

Types of ensembles

- Canonical ensemble (**NVT**). So, our ensemble is made in the same condition where all the states share the same Temperature, volume and number of particles.
- Isothermal-isobaric ensemble (**NPT**)

The things inside the parentheses are constant parameters



This is a dynamic funnel. If an enzyme binds, then the funnel changes.

## Obtaining ensembles

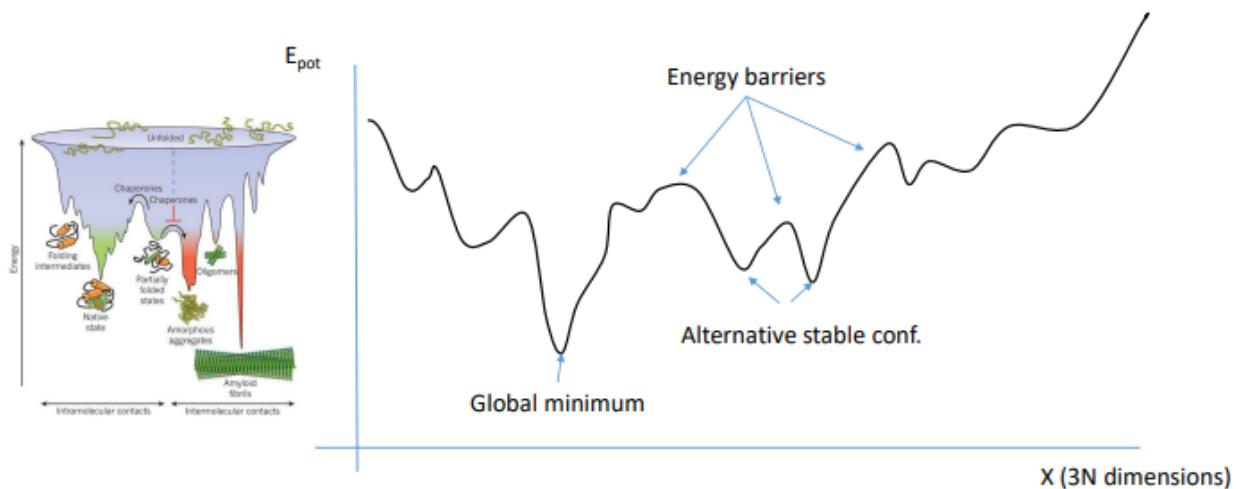
We need to understand proteins as ensembles and, hence, we need more than one conformation and a little idea of the difference of energy between different conformations.

Experimental:

- PDB analysis
- RMN, SAXS
- Protein Ensemble Database

Theoretical:

- Simulation:
  - Molecular dynamics
  - Monte-Carlo



The initial structure corresponds to a single point in the hypersurface. Thus, at the beginning of an experiment we do not have this plot, we only have a single point. We will discover it once we start “walking”.

Landscape can be discovered by conformational sampling

- Simple algorithms: Molecular Mechanics/ Monte Carlo / Molecular Dynamics
- Enhance sampling algorithms aim to improve coverage (SBIO)

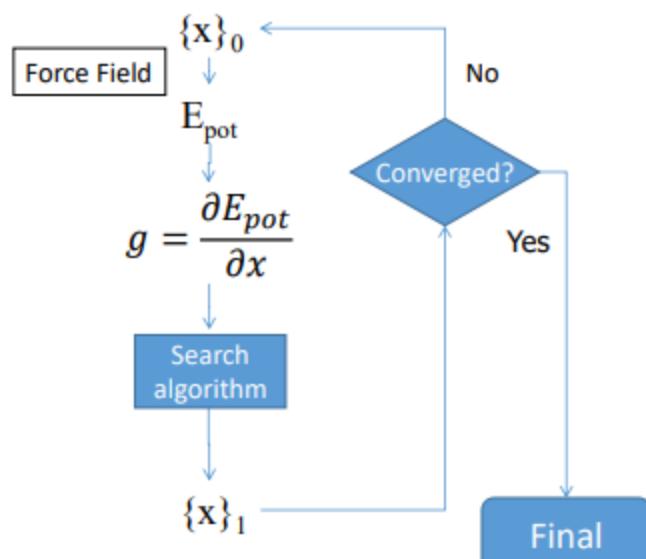
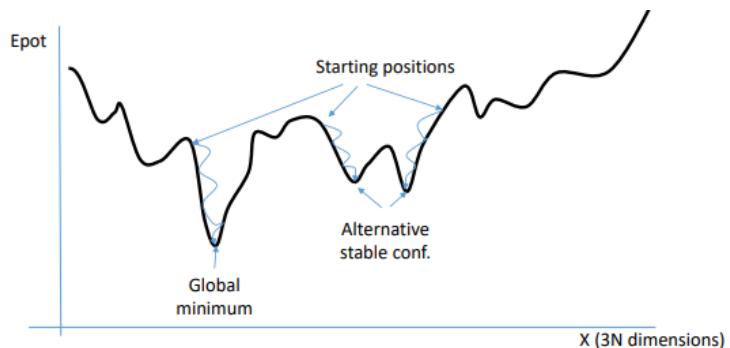
## Molecular Mechanics

Moves to conformations with lower energy values

- Fast and cheap
- Find local minima. Result always depends on the initial conformation
- Used before simulations to “relax” the system. Because we arrive at a minimum and, hence, the math around that minimum will be more stable. So, it's a trick to make our simulations more stable.

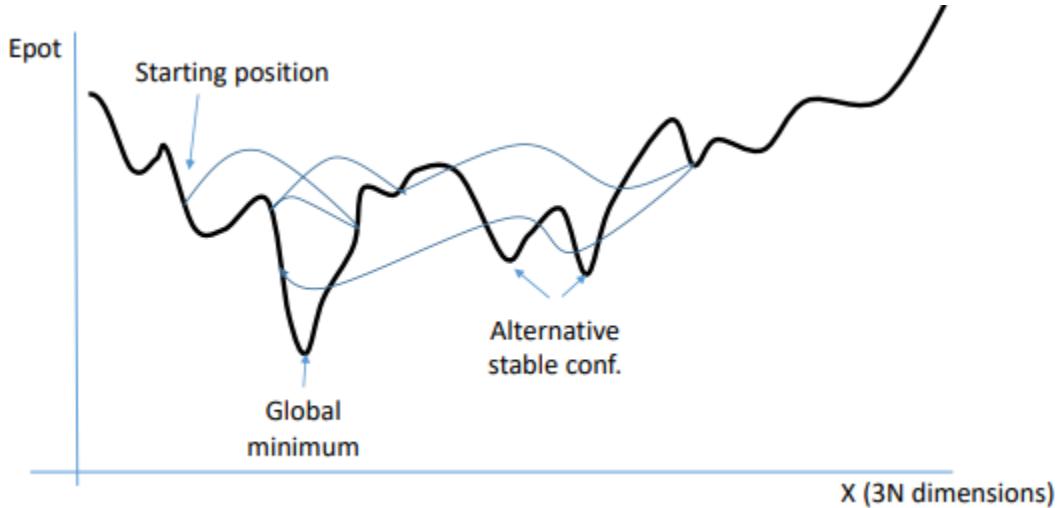
For this reason, it is always the first step for the preparation of the set up of any kind of experiment.

Relax experimental structures implies to drive them to an energy minimum as close as possible to the original coordinates.

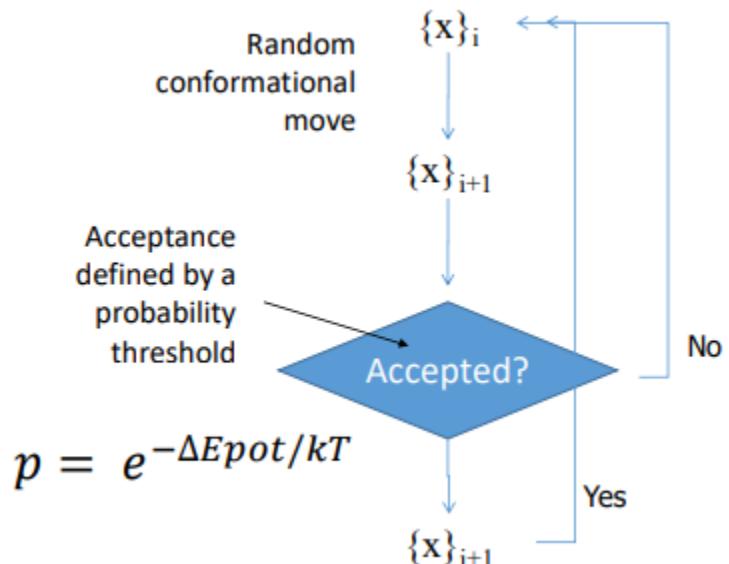


## Monte Carlo

Jumps randomly and accepts changes to lower or to **slightly** higher energies (considering temperature).

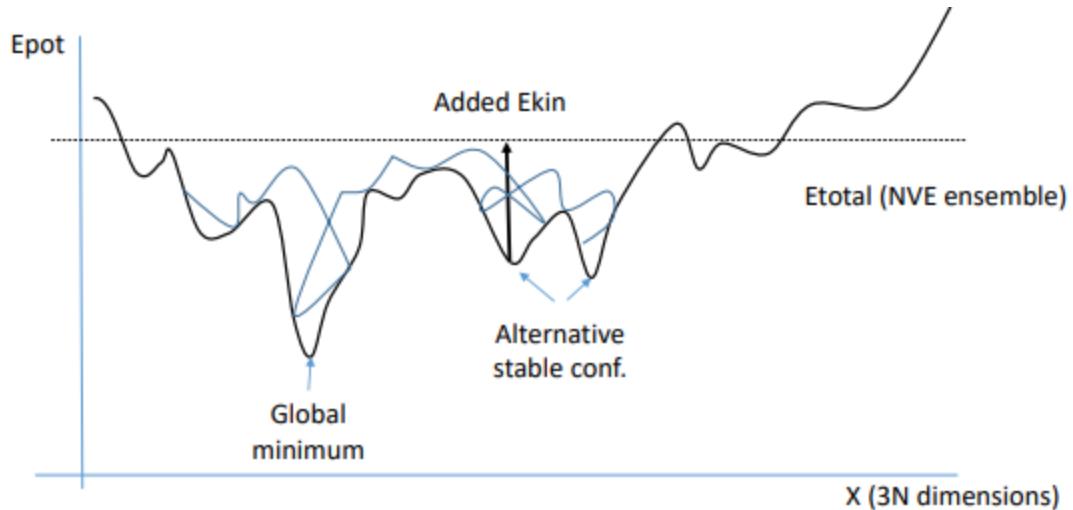


- Very efficient in exploring large changes (no stopping barriers)
- Can include “non-physical” transitions in combination with MD simulations:
  - pH const. dynamics
  - Chemical transformations
- Time is not considered
- Simulators are less optimized than MD.



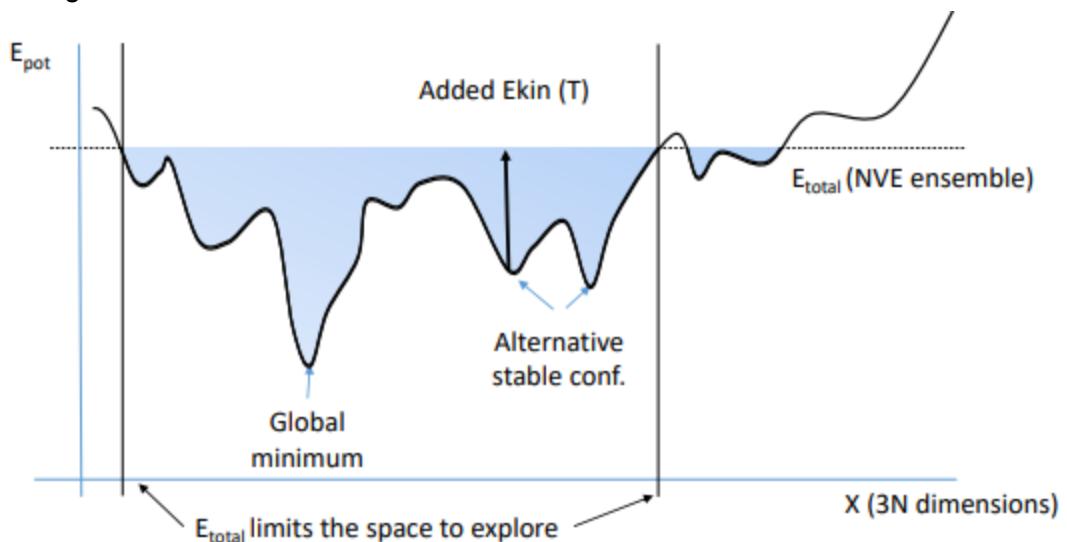
## Molecular Dynamics Algorithm

Uses extra (kinetic) energy to jump over energy barriers



## Molecular Dynamics (NVE)

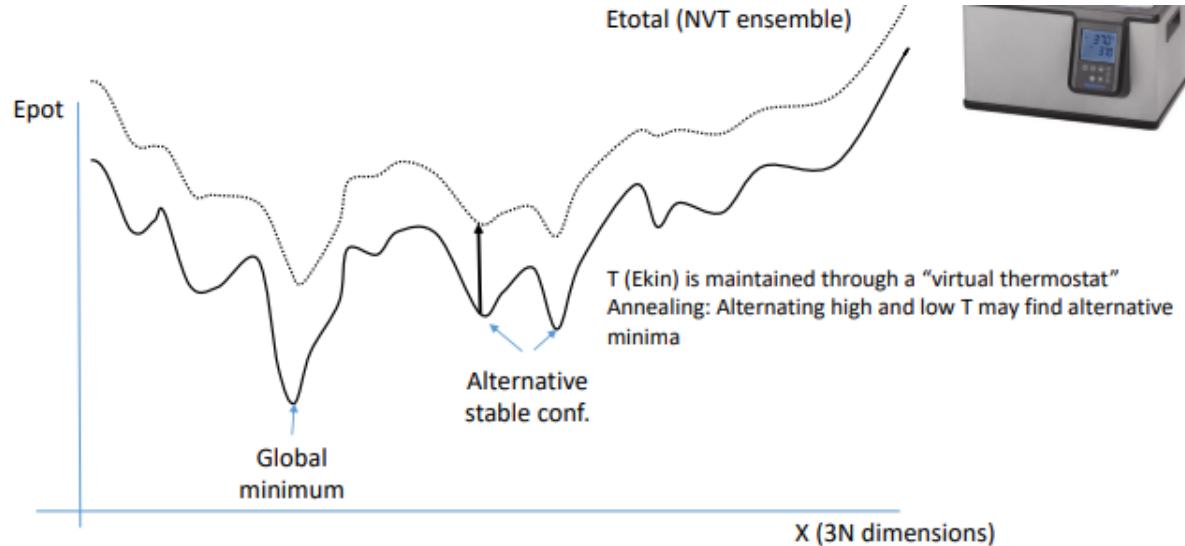
Depending on the starting point and the amount of energy, I will be able to visit more space. It does not make sense to use a really big energy (big temperature) because the proteins are at 30 degrees.



## Molecular dynamics (NVT)

If temperature is constant, then the kinetic is also constant.

So, this is more realistic.



## Which algorithm?

### Molecular mechanics:

- Mostly used for the initial minimization of the system ("relaxation") to assure a mathematical minimum on the energy surface.

### Monte Carlo:

- Efficient conformational sampling
- Mostly used in combination with MD to sample alternative systems (ionization states, etc)

### Molecular dynamics:

- Behavior of the systems for a long time.
  - Folding/unfolding
  - Conformational dynamics
- Advanced optimization
- Flexibility analysis

## Using ensembles

Analysis of a single system along time is equivalent to the analysis of many copies of the same system (ergodic principle)

- Simulation snapshots considered as individual states

Statistical thermodynamics allows to obtain  $\Delta S$  and  $\Delta G$

## **Summary**

- Conformational ensembles are the true representation of a macromolecule
- Particularly relevant on IDPs
- Macroscopic properties are best obtained by averaging ensemble properties
- Statistical thermodynamics
- Simulation methods allow to explore the conformational space and build a valid ensemble

## Enzyme Catalysis

A catalyst is a substance that changes the velocity of a chemical reaction and they do not appear in the stoichiometric equation (they participate but they do not react). They are not consumed either.

It modifies the reaction mechanism to decrease the energy of activation. A small quantity of catalysts is enough to produce a considerable effect, because it is not consumed. They can increase the reaction rate from 10 to  $10^{12}$  times.

In biology, almost no reaction has a significant rate, so they need an enzyme. Thermodynamics can not be modified, but the kinetics yes.

Since the thermodynamics is not affected, it also does not affect the position of the equilibrium (it works on both directions of the reaction, but not necessarily with the same efficiency). Most biological processes have negligible rates without catalysis, so we need a catalysis.

- Biological catalysts are mostly proteins (Enzymes), but also RNAs (Ribozymes)
- Control of catalytic effect is key to control biological processes

## Some nomenclature

**Enzyme:** Protein able to catalyze chemical transformations. They are key to control metabolism, and cell regulation. Any other molecule that is a catalyst, but not a protein, it is not an enzyme.

### Substrate → Product:

- Metabolite that become transformed by enzyme catalyzed reactions into products
- Enzymes catalyze both directions, so product can become substrates

**Cofactors / Coenzymes / Prosthetic groups:** Non protein components the are necessary for the catalytic products

**Inhibitors:** Molecules that decrease enzyme efficiency

**Activators:** Molecules that enhance enzyme efficiency. Activators may be “essential” when they are necessary for the catalysis

## Biological catalysts

We have enzymes (proteins) and ribozymes (RNA).

- Enzymes catalyze a wide range of chemical processes
- Ribozymes catalyze a limited number of processes like
  - RNA cleavage
  - Intron processing
  - Protein synthesis (ribosome)

## Enzymes use binding energy to decrease activation barriers

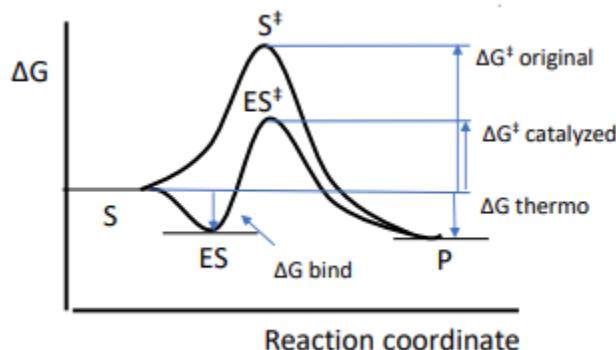
First step is the formation of ES

- The binding energy ( $\Delta G < 0$ ) is invested in decreasing the original activation energy

Binding decreases the degrees of freedom of S and reduces  $\Delta S \ddagger$  ( $ES \rightarrow ES\ddagger$ )

- Effects of "approximation" and "orientation"

Residues in the protein can, additionally, provide acid-base or electrostatic catalysis



## Modes of catalysis

Approximations & Orientation

- The enzyme provides binding energy to compensate for entropy loss and assure the proper orientation for the chemical transformation.  
If they are in the correct orientation, the enhanced rate increases a lot.

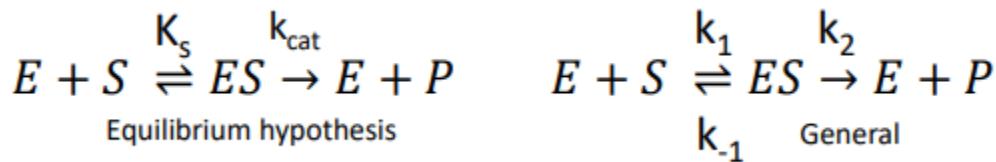
Active catalysis. The enzyme provides selective stabilization energy for the transition State.

- Electrostatic
- Acid-Base catalysis

Covalent catalysis

- The enzyme participates in the reaction using covalent intermediates.

## Basic enzyme mechanism ( $S \rightarrow P$ )



- In the usual experimental conditions:  $E \ll S \rightarrow S \approx S_T$
- And stationary state:  $dE/dt = dES/dt = 0$

$$\frac{dE}{dt} = -k_1 E S + k_{-1} ES + k_2 ES \quad 0 = k_1 E S - (k_{-1} + k_2) ES \quad v = -\frac{dS}{dT} = \frac{dP}{dt} = k_2 ES$$

$$\frac{dES}{dt} = k_1 E S - k_{-1} ES - k_2 ES \quad \frac{k_{-1} + k_2}{k_1} = \frac{E S}{ES} = K_M \approx K_S \quad v = \frac{k_2 E_T S}{K_M + S}$$

$$E_T = E + ES \quad ES = E \frac{S}{K_M} \quad E = \frac{E_T}{1 + \frac{S}{K_M}} \quad V_{max} = k_2 E_T$$

$$v = \frac{V_{max} S}{K_M + S}$$

$k$  = Kinetic constants

$K$  = Thermodynamic constants (equilibrium constant)

## Measurable kinetic parameters

Michaelis-Menten equation is valid for many enzyme mechanisms

- The microscopic meaning of  $K_M$  and  $V_{max}$  depends on the actual mechanism (In general  $V_{max} = k_{cat} E_T$ )

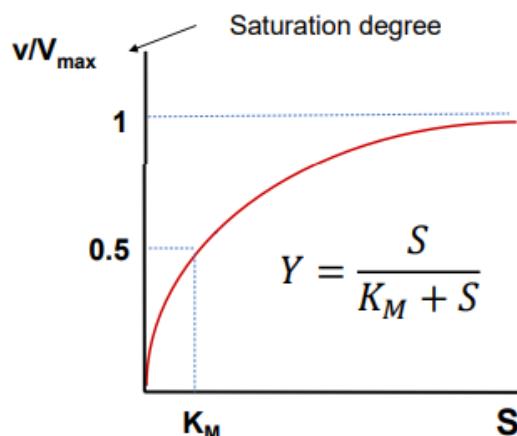
Several approaches exist to linearize MM equation, however the proper evaluation of parameters is done by non-linear regression.

$k_{cat}$  ( $V_{max}$ ) gives a measure of the efficiency of the catalysis

$K_M$  gives a measure of the affinity of the E and S

- $K_M$  corresponds to the S giving 50% of saturation ( $S_{50}$ )

Overall enzyme effectivity is measured by  $k_{cat}/K_M$



## Enzyme regulation

Regulation is a key feature of enzymes.

- Long term regulation is done at the level of transcription, modulating the amount of enzyme expressed

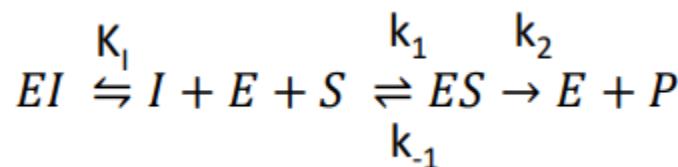
### Enzyme inhibition

- Reduction of enzyme activity due to the binding to another component (inhibitor). Can be partial or complete.
- Many mechanisms
  - Irreversible. Inhibitor forms a stable complex that inactivates the enzyme permanently
  - Reversible. Inhibitor forms a complex that inactivates the enzyme. Activity is recovered by reducing inhibitor concentration
- Most common: Competitive inhibition. The inhibitor binds to the substrate site. It does not change the catalytic constant, because if we put enough substrate we will reach the maximum velocity.

### Allosteric regulation

- The enzyme has at least two stable conformations with different kinetic properties.
- Shift of their relative stability by any ligand can lead to inhibition or to activation

## Competitive inhibition

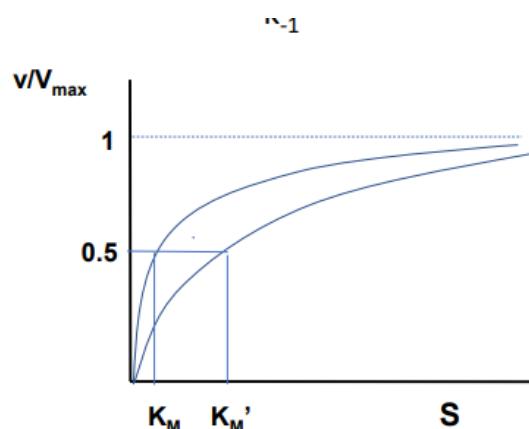


$$\frac{dES}{dt} = k_1 E S - k_{-1} ES - k_2 ES = 0$$

$$E_T = E + ES + EI$$

$$\frac{ES}{ES} = K_M \quad K_I = \frac{E}{EI} \quad E = \frac{E_T}{1 + \frac{S}{K_M} + \frac{I}{K_I}}$$

$$v_I = \frac{k_{cat} E_T S}{K_M \left(1 + \frac{I}{K_I}\right) + S} \quad K'_M = K_M \left(1 + \frac{I}{K_I}\right)$$



$$Inh\ Degree = 1 - \frac{v_I}{v} = 1 - \frac{K_M + S}{K_M \left(1 + \frac{I}{K_I}\right) + S} = \frac{\frac{K_M}{K_I} I}{K_M \left(1 + \frac{I}{K_I}\right) + S} = \frac{I}{K_I \left(1 + \frac{S}{K_M}\right) + I} \quad IC_{50} = K_I \left(1 + \frac{S}{K_M}\right)$$

**Degree of inhibition:** 0 means no inhibition and 1 means complete inhibition.

## Practical considerations

- Only  $K_s$  and  $K_i$  are actual equilibrium constants and can be related to thermodynamic  $\Delta G$ 's
- $K_m$  is just a kinetic parameter, and it is only an estimation of  $K_s$ , however in most cases is assumed as a binding affinity
- $V_{max}$  depends on E concentration and can change between experiments.  $k_{cat}$  is the true constant.
- IC<sub>50</sub> also depends on the experiment settings, but is very common in drug design publications

# Unit 3.3. Biological membranes and transport

## Compartmentation and biological membranes

Organized complexes formed mainly by lipids and proteins:

- Phospholipids have a polar head (many times charged) and a lipid chain that is hydrophobic.

Limit metabolic processes (they limit compartments of the cell):

- This allows specialized organelles

Regulate composition of intracellular medium according to metabolic states because they allow some particles to enter the compartments and not others.

Support special processes like signal transduction (hormones, and electrical signals, respiratory chain, photosynthesis)

They generate concentration gradients because they don't allow the entrance of some molecules. If the concentration of one substance is different in the 2 sides of the membrane and that substance can not cross, then we have a storage of energy.

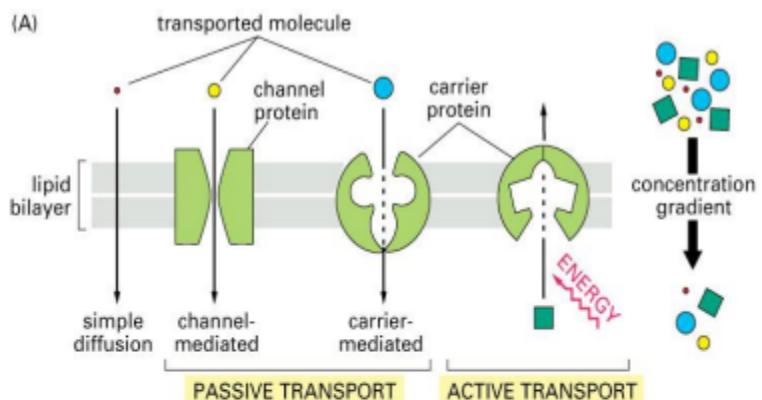
We have to think of the membrane as a solution that can have hydrophobic molecules inside...

## Transport

Since most molecules can not cross the membrane, we need a transporter.

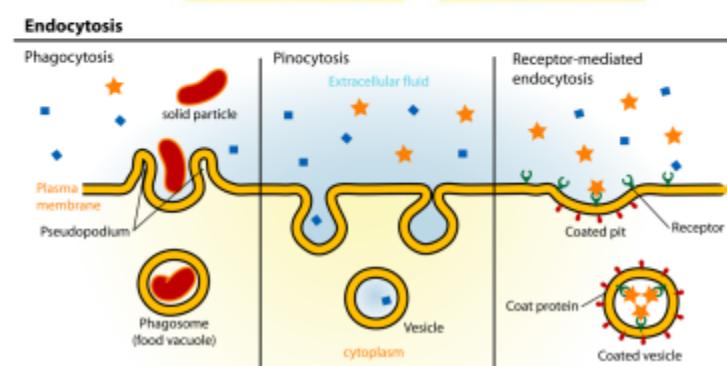
Passive transport (diffusion):

- No energy is required (only a difference of concentration). Water and hydrophobic elements.



Active transport:

- Energy consumption is coupled to transport. Implies a reaction (enzymes) that catalyze a reaction and a movement.



Bulk transport:

- Implies the formation of cell structures to transport materials

## Thermodynamics and kinetics

Molecules move in the direction of decreasing concentrations

- (a.k.a. in the direction that increases entropy).

Equilibrium correspond to equal concentrations

For charged solutes the possible electrostatic potential (~-70mV) should be considered.  
Thus, the concentration will not be equal in both sides.

$$\Delta G = RT \ln \frac{C_2}{C_1} + ZF\Delta V$$

Z: Solute charge

F: Faraday constant = 96500 C/mol.  
(charge of 1mol of  $e^-$ )

Fick's first law (diffusion)

$$\frac{dN}{dt} = -DA \frac{\Delta c}{\Delta x}$$

$$J = \frac{1}{A} \frac{dN}{dt}$$

$$J = -D \frac{\Delta c}{\Delta x} = -p\Delta c$$

N: transported molecules, D: Diffusion constant, A: Section, J: Flux, p: membrane permeability

Protein mediated transport shows enzyme-like kinetics

## Diffusion coefficients

Diffusion coefficients decrease with increasing particle size

Intracellular media decrease diffusion

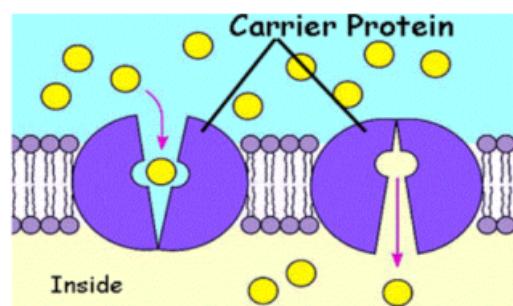
- Intracellular Crowding. If there is more density, it's more difficult to move.

Diffusion coefficients increase with temperature

## Passive transport. Facilitated diffusion

There is 2 mechanism that allow the facilitated diffusion:

- Channel proteins (it is adapted to the size of the molecule so that it can only pass a concrete molecule).
  - Provide water filled pores to allow hydrophilic molecules to diffuse
- Carrier proteins: They have 2 possible conformations. The change of conformations allows the molecules to pass. What triggers the change of conformation? The thermodynamics of the molecules that passes
  - Bind actively molecules and change conformation to allow transport
  - Solute selective



No additional energy is required  
 All mediated transport show saturation  
 Regulation is possible

Membrane proteins are strange because they have hydrophobic outside and hydrophilic inside.

## Active transport

Transport is done against concentration gradient

Types

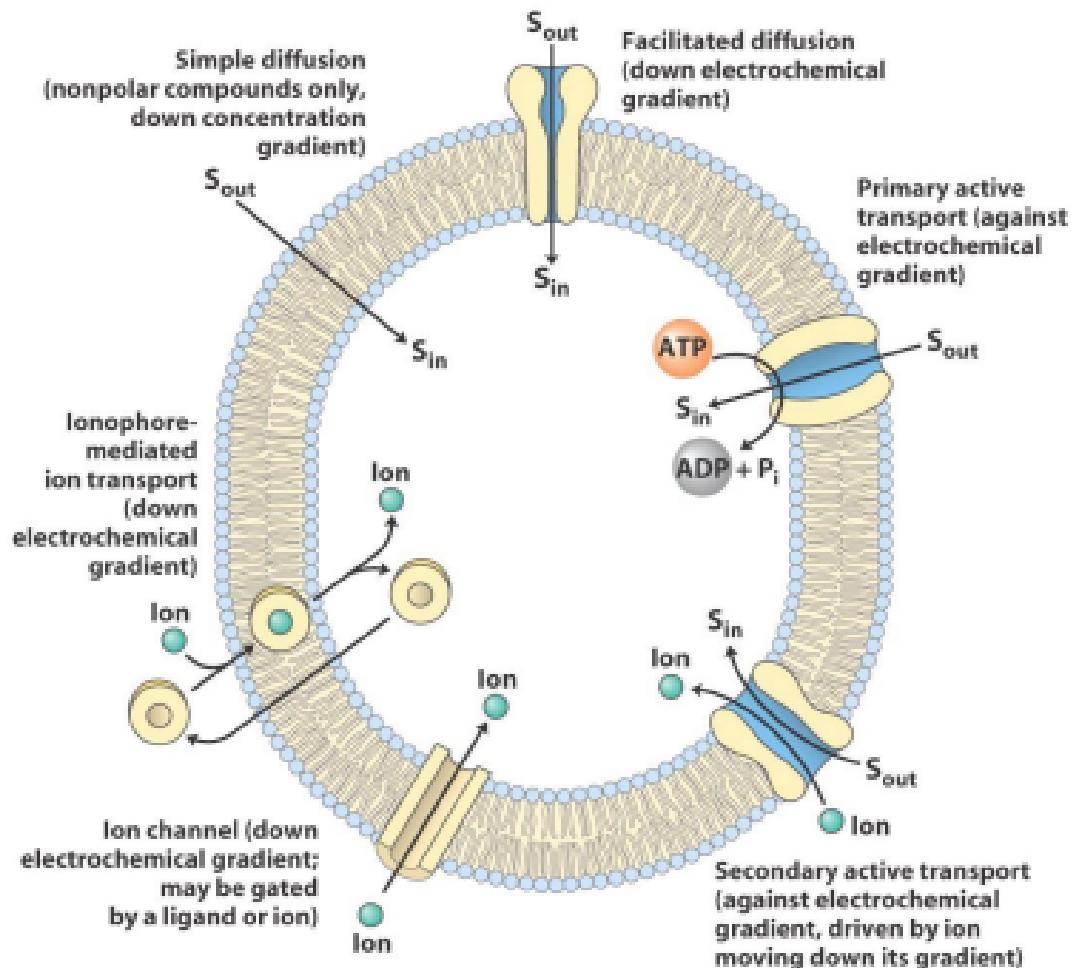
- Uniport: 1 molecule transported
- Symport: 2 molecules transported in the same direction
- Antiport: 2 molecules transported in opposite directions (no hydrolysis of ATP)

Electrogenic/electroneutral: Depending on the charges of transported molecules

- Electrogenic transport is sensitive to membrane potential

Primary: Source of energy is ATP hydrolysis coupled to transport

Secondary: Source of energy is the concentration or electrochemical gradient produced by another transport. They depend on another transported protein.



## **Sodium-Potassium pump**

Ubiquitous system for the continuous exchange of 3 internal Na<sup>+</sup> x 2 external K<sup>+</sup>

Usually called Na<sup>+</sup> /K<sup>+</sup> ATPase due to the hydrolysis of ATP required

Responsible of generating the membrane potential (~- 70mV)

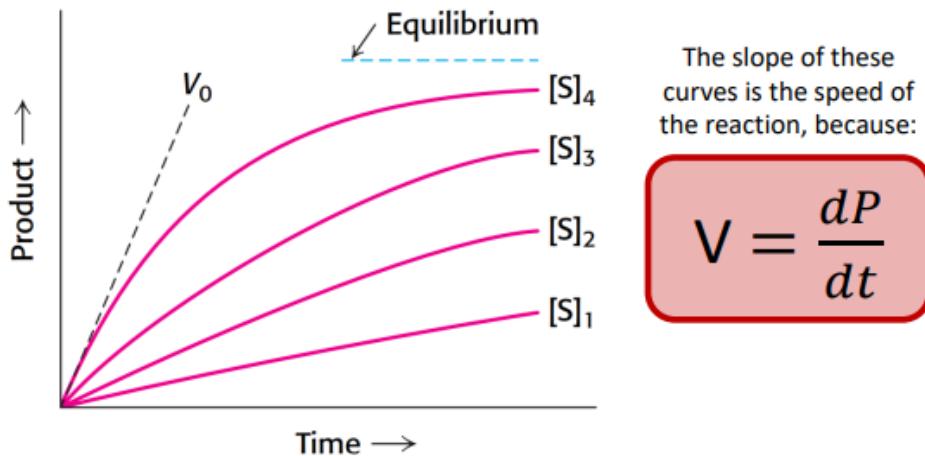
Source of energy for secondary transporters (like Na-Glucose transporter)

## Seminar 9. Enzyme kinetics

In both, kinetics and enzyme kinetics we will study the speed of chemical reactions  
In enzyme kinetics substrates (S) are transformed into products (P)

$$V = -\frac{dS}{dt} = \frac{dP}{dt}$$

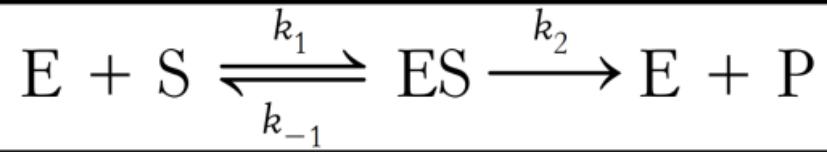
Given a constant amount of enzyme, we see how the initial speed of a reaction depends on the concentration of substrate



If we increase the amount of substrate, we see how the speed of the reaction goes up, until we reach maximum speed

**Why does the speed of the reaction stop increasing while we increase substrate concentration?** At maximum speed all enzyme molecules are occupied by substrates.  
**How would you improve maximum speed?** The reaction cannot go any faster unless you add more enzyme molecules to the system.

To explain this behavior, Michaelis and Menten developed their model for enzyme kinetics



In chemical kinetics we saw that we can express the speed of a reaction as the product of the reactants and a constant:



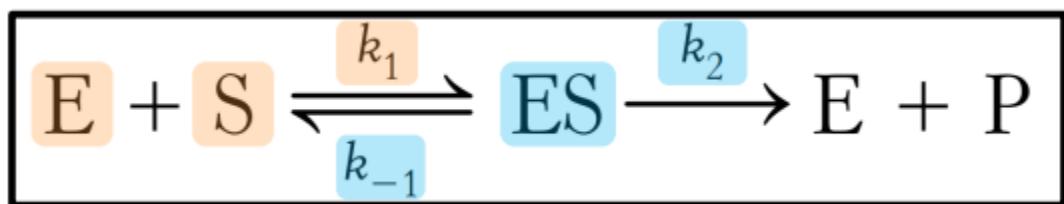
$$V = k \cdot [A]^x \cdot [B]^y$$

If we adapt the formula of classical kinetics to the Michaelis-Menten system, we get the following expression:

$$V = k_2 \cdot [ES]$$

$[ES]$  is very hard to know, so we will try to express this value as a combination of other values that we can know.

To quantify  $[ES]$  we can consider what elements in the system contribute to the **formation** of ES and the **decomposition** of ES

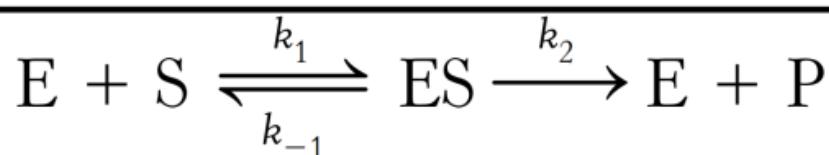


**Formation of ES:**  
 $k_1 \cdot [E] \cdot [S]$

**Decomposition of ES:**  
 $k_{-1} \cdot [ES] + k_2 \cdot [ES]$

We are going to combine these two equations by assuming that the chemical system is in a moment where formation and decomposition of ES is in equilibrium (ES is constant). Therefore:

$$\frac{d[ES]}{dt} = 0$$



$$\frac{d[ES]}{dt} = 0 = k_1 \cdot [E] \cdot [S] - K_{-1} \cdot [ES] - K_2 \cdot [ES]$$

$$0 = k_1 \cdot [E] \cdot [S] - [ES] \cdot (K_{-1} + K_2)$$

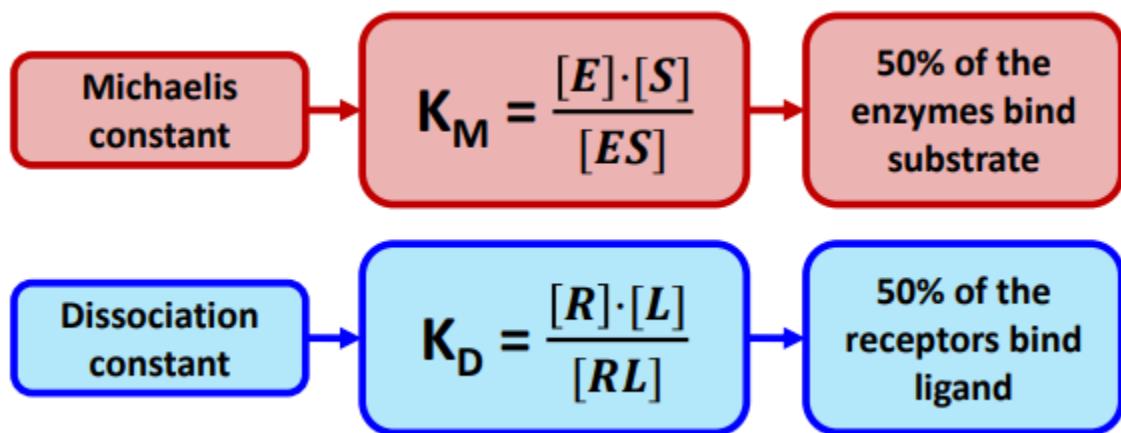
$$\frac{[E] \cdot [S]}{[ES]} = \frac{k_{-1} + k_2}{k_1} = KM$$

## Michaelis Constant (KM)

KM is a ratio of concentrations, but it is also the concentration of substrate at which the reaction takes place at 50% of maximum speed.

$$\frac{[E] \cdot [S]}{[ES]} = \frac{k_{-1} + k_2}{k_1} = KM$$

KM is very similar to another constant we have seen recently in class, can you guess what constant is this? The dissociation constant



Remember that we are still trying to express [ES] as something we can measure easily, we can apply the following operation:

$$KM = \frac{[E] \cdot [S]}{[ES]} \longrightarrow [ES] = \frac{[E] \cdot [S]}{KM}$$

We can estimate [S] because:

$[S] \ggg [ES]$  and  $[S]_{total} \approx [S]$

We cannot estimate [E] because:

$[E]$  similar to  $[ES]$  and  $[E]_{total} \neq [E]$

Then, we have to express [E] as a combination of parameters that can be easily measured:

$$[E] = [E]_{Total} - [ES]$$

Now we have 3 different equations that we can integrate together to come up with one expression that explains the speed of a reaction

$$\begin{array}{l}
 V = k_2 \cdot [ES] \\
 [ES] = \frac{[E] \cdot [S]}{K_M} \\
 [E] = [E]_{total} - [ES]
 \end{array}
 \quad \rightarrow \quad
 \begin{aligned}
 V &= k_2 \cdot \frac{([E]_{total} - [ES]) \cdot [S]}{K_M} \\
 &\downarrow \\
 V &= k_2 \cdot \frac{[E]_{total} \cdot [S]/K_M}{1 + [S]/K_M} \\
 &\downarrow \\
 V &= k_2 \cdot [E]_{total} \cdot \frac{[S]}{[S] + KM}
 \end{aligned}$$

This final formula enables us to express the speed of a reaction using concentrations and constants that can be easily measured.

The maximum speed of the reaction will involve that  $[E]_{total}$  is involved in the reaction

$$V = k_2 \cdot [ES] \longrightarrow V_{max} = k_2 \cdot [E]_{total}$$

We can relate that to the formula of the speed of the reaction:

$$\begin{aligned}
 V &= k_2 \cdot [ES] \longrightarrow V_{max} = k_2 \cdot [E]_{total} \\
 &\quad \left[ \begin{array}{l} V = k_2 \cdot [E]_{total} \cdot \frac{[S]}{[S] + KM} \\ \longrightarrow V = V_{max} \cdot \frac{[S]}{[S] + KM} \end{array} \right]
 \end{aligned}$$

**See that when  $[S] = K_M$ ;  $V = V_{max} \cdot 1/2$ .**

$$V = V_{max} \cdot \frac{[S]}{[S]+KM}$$

$$V = V_{max} \cdot \frac{K_M}{K_M+KM} \longrightarrow V = V_{max} \cdot \frac{1}{2}$$

## **Justify briefly the following sentences (all them are true)**

**Only quantum mechanics can be used to evaluate kinetic rate constants theoretically.**  
To evaluate rate constants the analysis of transition states and possibly the reorganization of bonds is required. Classic forcefields assume equilibrium structures and a fixed chemical structure, and cannot be used. Only QM allows for reorganization of bonds and non-equilibrium structures

**The protein folding process implies finding a preferred kinetic pathway from the unfolded to native states.**

To evaluate all possible conformations is not feasible due to the large degrees of freedom.  
To achieve a folding in normal times requires a preferential pathway.

**NPT is the most realistic simulation ensemble**

NPT corresponds to constant temperature and pressure, very much similar to laboratory experiments, so is the most realistic ensemble

**The choice of the simulation temperature in NVE molecular dynamics defines the limits of the conformational space.**

NVE corresponds to constant volume and constant total Energy. Once the temperature is set the total energy of the system cannot be changed. Since from that on only those conformations with an excess of kinetic energy can be visited, the selection of the temperature and total energy define which conformations are available

**Protein folding in vivo is much more efficient than in vitro**

In vivo, chaperones stabilize unfolded structures until they fold properly. This increases the efficiency of the folding (all synthesized proteins fold correctly).

**Folding “in vivo” is helped by accelerating the isomerization of Pro residues.**

Pro residues can exist in both cis and trans backbone conformations, and the cis-trans isomerization is slow. Fast folding is helped by isomerases, enzymes able to catalyze the isomerization of Pro residues,

**Protein stability can be evaluated experimentally by following the denaturation process**

Following the denaturation process we can get the relations U/F at different degrees of unfolding, and deduce the corresponding  $\Delta G$  that can be then extrapolated to obtain protein stability

**Replacement of positive ionic residues located at the protein surface by negative or neutral polar residues is acceptable**

At the surface polar and ionic interactions are made with the solvent, so these changes have little energetic implications.

**Following enzyme catalyzed chemical transformations requires the use of quantum mechanics methods**

Only quantum mechanics methods can analyze bond reorganization, so chemical processes. The other methods assume that bonds are fixed.

**Macromolecules' stable conformations are thermodynamic energy minima**

A "Stable conformation" requires a minimum in the energy landscape, as at minima there is no energy gradient that induces changes.

**Increasing the yield of protein folding "in vivo" requires to stabilize unfolded structures.**

Proteins can be trapped in misfolded conformations or aggregates. A strategy followed "in vivo" by chaperones, is to keep stable the unfolded conformations, and provide "second opportunities" to increase the overall yield.

**Classic force fields cannot be used to evaluate whether a protein fold is correct**

Classic forcefields define that the structure is "chemically correct", but has no terms to identify correct conformations. This can only be obtained by evaluating entropy (which is not available) or by comparison with known structures (using statistical potentials)

**Monte Carlo simulations is not the preferred approach to relax experimental structures**

Relax experimental structures implies to drive them to an energy minimum as close as possible to the original coordinates. MC does random changes that can significantly alter the structure. The preferred method is molecular mechanics.

**NPT are the recommended ensemble in Molecular Dynamics simulations**

NPT keeps constant the number of particles, pressure and temperature. This is the most realistic situation.

**A hyperbolic shape in the plot of binding degree against ligand concentration indicates the participation of a protein in the process**

A hyperbolic shape implies that beyond some ligand concentration, binding does not increase (Saturation). This behavior is indicative of a limited number of binding sites, so a protein should participate. Otherwise the increase would be linear

**The value of the dissociation constant of a protein-ligand complex can be obtained or estimated from the concentration of ligand giving 50% of saturation**

The shape of a binding curve is hyperbolic, and can be adjusted to  $Y = L/(Kd+L)$ .

For  $Y = 0.5$   $L$  should be  $Kd$

**Knowledge of protein structures is required to evaluate whether a protein fold is correct**

The definition of "correct fold" depends on the comparison with known structures as forcefields cannot distinguish between folded and unfolded structures

**An initial relaxation of the structure is required before starting a simulation calculation.**

To avoid numerical instability during simulation the initial structure should be in a mathematical minimum according to the forcefield to be used. An initial relaxation of the structure using molecular mechanics allows to achieve this.

**In an enzyme process a concentration of substrate of 100nM gives a reaction rate that corresponds to the 50% of the rate obtained at 10mM concentration of the same substrate. This finding confirms that 100nM is a good estimate of the KM kinetic parameter.**

As 10mM is much larger than 100nM, it can be assumed that corresponds to the maximum velocity. Hence the concentration of substrate giving 50% of such corresponds to the KM

**Indicate whether the following sentences are true or false. Justify briefly the answer**

**The NPT ensemble is not realistic, as all biological processes take place at constant volume.**

False. Biological processes take place at a constant pressure and temperature, so NPT is actually the most realistic.

**Molecular mechanics is a required initial step in all simulations**

True. Molecular mechanics assures that the structures are a mathematical minimum within the forcefield in use. It is required to avoid numerical instabilities in the simulation.

**Chaperones base their action in accelerating the folding process.**

False. Chaperones stabilize unfolded structures to assure that they fold properly even at a lower rate.

**Classic forcefields can be used to evaluate kinetic rate constants**

False. The evaluation of kinetic rate constants requires to analyze processes where the molecular structure changes (breaking or forming covalent bonds) or are severely distorted (transition states). Classic forcefields require a fixed molecular structure and equilibrium conformations and cannot be used to analyze chemical transformations (Quantum mechanics is required)

**The addition of denaturants or the increase in temperature allow to evaluate the stability of protein-protein complexes.**

True. The addition of any compound (a denaturant) or a change in conditions (e.g. Temperature) that leads to the unfolding of the protein can be used to follow the process, obtaining the apparent  $\Delta G$  at each experimental point (from U/F).

**A larger content of buried hydrophobic residues increases the protein stability**

True. A larger content of buried hydrophobic residues imply a larger hydrophobic effect (related to the hydrophobic solvation energy of such residues) and therefore a larger protein stability.

**Solvation energy in water cannot be calculated using classical forcefields**

True. Solvation energy contains entropic terms, and entropy cannot be calculated from an FF

**Monte Carlo simulations are limited to the conformational space between energy barriers**

False. MC simulations do random conformational changes, so they are not limited by energy barriers

**Flexible residues like Gly decrease protein stability**

True. They increase the degrees of freedom of the unfolded form, and then the loss in conformational entropy on folding is larger.

**Hydrogen bonds do not contribute to the overall stability in nucleic acid double helices**

True. Hydrogen bonds have the same energy in the double helix or when they are made with water in the unfolded chain. Most stability comes from stacking interactions (hydrophobic + Vdw) between base-pair steps.

**Statistical potentials are needed to identify properly folded structures**

True. Normal FF cannot distinguish properly folded structures. Statistical potentials do so by comparing with known protein structures.

**Protein folding is driven by entropic contributions**

True. Protein folding is energetically driven by the hydrophobic effect that comes from the gain in entropy of the solvent.

**NVT ensembles can be obtained by molecular dynamics simulations at constant pressure**

False. NVT corresponds to constant VOLUME and TEMP. Constant pressure simulations are used for the NPT ensemble.

**Calculation of binding energies requires only the calculation of interaction energies between interacting molecules.**

False. Also, solvation energies of both the complex and the unbound components is relevant.

**When analyzing reaction rates, the concentration of substrate known as KM corresponds to a fully saturated protein**

False. KM corresponds to the concentration that gives 50% of saturation.

**Chaperones accelerate protein folding “in vivo” by catalyzing the isomerization of Pro residues.**

False. Chaperones stabilize unfolded structures avoiding mis-folding

**Classify the following energy terms according to their influence in macromolecule stability (favorable (F) / unfavorable (U) / indifferent (I))**

**Macromolecule conformational entropy:** U. Conformation entropy decreases on folding (less available conformations)

**Electrostatic interaction between main chain atoms of nucleic acids:** U. Main chains in Nucl. Acids have a strong negative charge, the interaction is repulsive

**Electrostatic interaction between surface residues in proteins:** I. At the surface electrostatics is small due to the effect of solvent

**Electrostatic interactions inside the protein:** F. An electrostatic interaction in an environment without solvent is the strongest interactions in energetic terms.

**Entropy of solvent:** F. This is the hydrophobic effect that is the principal driving force for folding

**Van der Waals energies:** F. VdW interactions are favorable as they improve contacts.

**Hydrogen bonds:** I. Hydrogen bonds contribute equally to the system energy in both unfolded and folded structures, so they are indifferent in energetics terms. They contribute to the proper orientation of macromolecular interactions.

**5) (1 point) Identify related terms in the two columns**

- |                                   |                                  |
|-----------------------------------|----------------------------------|
| a. Solvent entropy                | A. Hooke equation                |
| b. Bond angles                    | B. Generalized-Born equation     |
| c. Van der Waals energy           | C. Hydrophobic effect            |
| d. Electrostatic solvation energy | D. Quantum mechanics calculation |
| e. Bond breaking energy barrier   | E. 12-6 Potential                |

**a-C, b-A, c-E, d-B, e-D**

- |  |   |
|--|---|
| A. Interaction energy between polar residues and solvent | A. 12-6 potentials                          |
| B. Hydrophobic solvation                                 | B. Accessible Surface Areas (ASA)           |
| C. Van der Waals interaction energy                      | C. Quantum mechanics                        |
| D. Entropic effects                                      | D. Generalized-Born equation                |
| E. Kinetic rate constants                                | E. Molecular Dynamics/Statistical Mechanics |

**A-D. Generalized-Born is a measure of the electrostatic component of solvation**

**B-B ASA can be used to estimate the hydrophobic component of solvation**

**C-A 12-6 potentials are the usual type of equations used to evaluate VdW energy**

**D-E Only simulation data (in combination with statistical mechanics) can evaluate entropy**

**E-C Only QM can evaluate chemical transformations and hence rate constants**