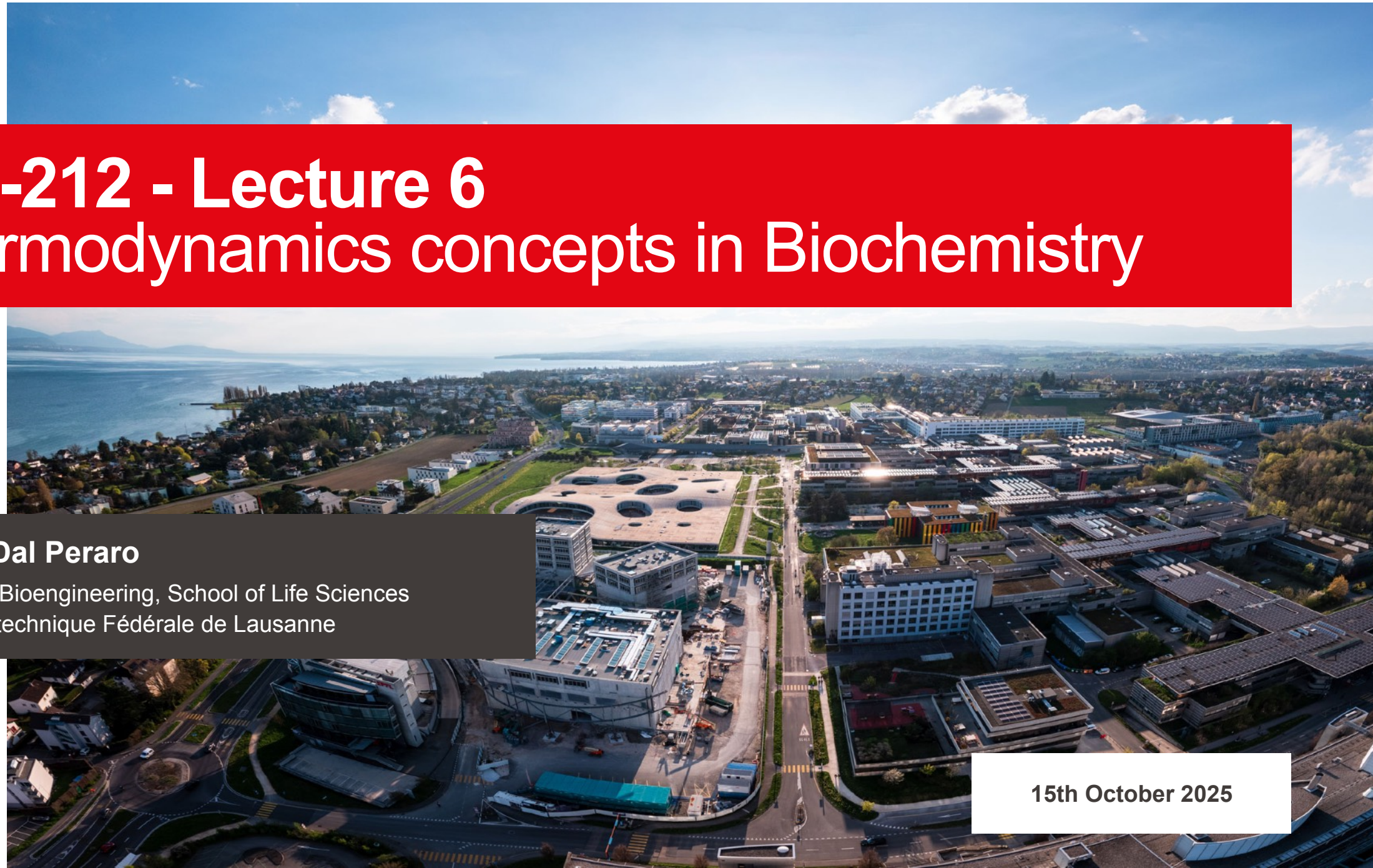
BIO-212 - Lecture 6
Thermodynamics concepts in Biochemistry



Matteo Dal Peraro

Institute of Bioengineering, School of Life Sciences
École Polytechnique Fédérale de Lausanne

15th October 2025

About me

- **physicist**, studied at University of Padova, Italy

- became a **biophysicist** (PhD at SISSA in Trieste)

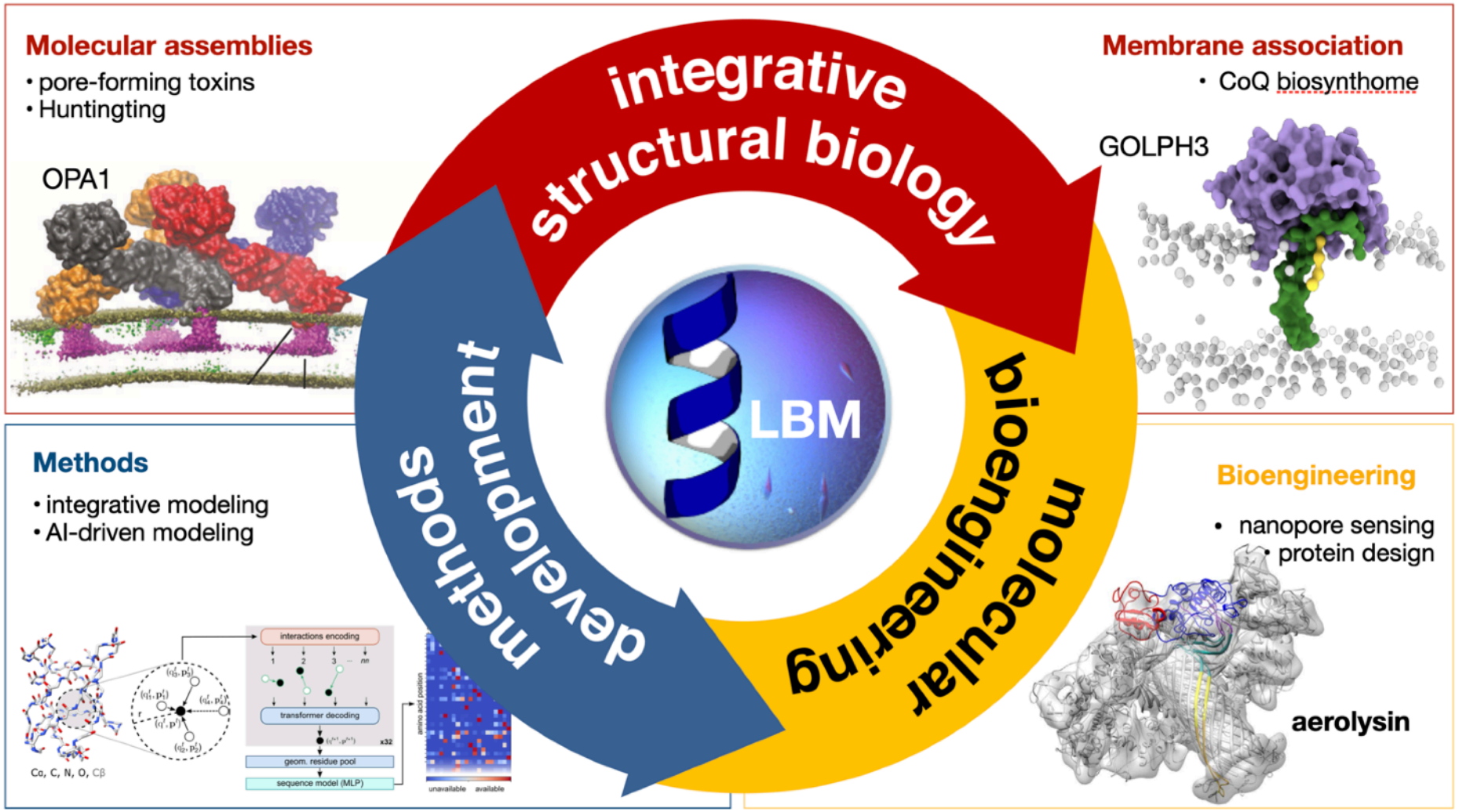
- postdoc at UPenn Chemistry, Philadelphia USA

- professor at SV, Institute of Bioengineering (IBI)

- Director of the Institute of Bioengineering (IBI)
- office AAB 048 - matteo.dalperaro@epfl.ch

About my laboratory

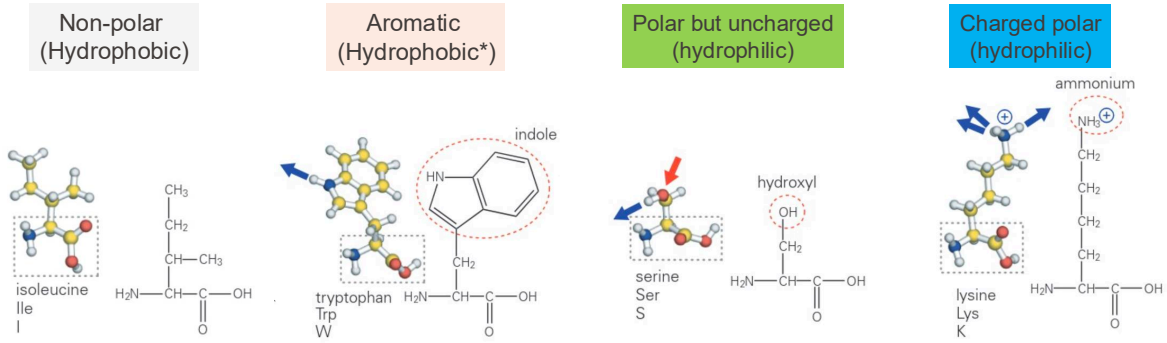
- **Laboratory of Biomolecular Modeling** (LBM) - AAB 0th floor, AI 2nd floor
- **goal:** understanding the physico-chemical principles of biological function and use them for bioengineering (e.g., drug and protein design, nanopores)



Lecture 5 – Quick Summary

Amino acid physicochemical properties

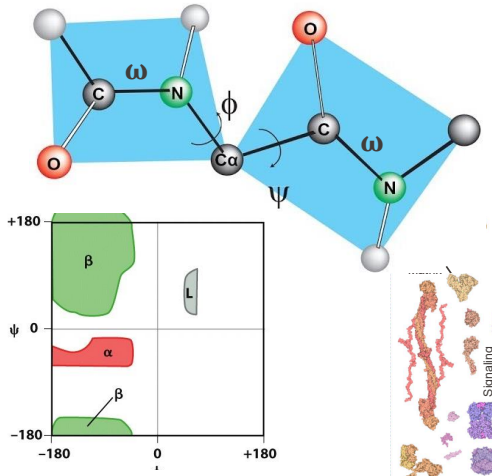
- proteins are made by amino acids

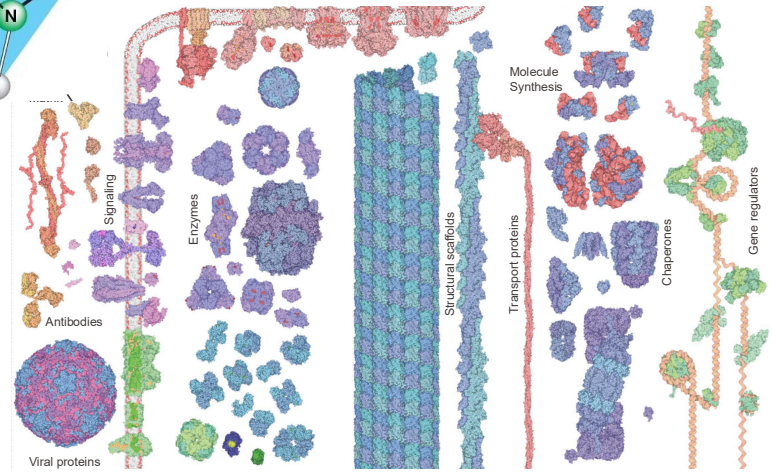


- their chemical features determines how proteins interact with other biomolecules and how they fold

Protein structure and diversity

- proteins are polymers





- Their chemical properties also determine their structure
- They have a hierarchy of motifs to generate the most diverse architectures

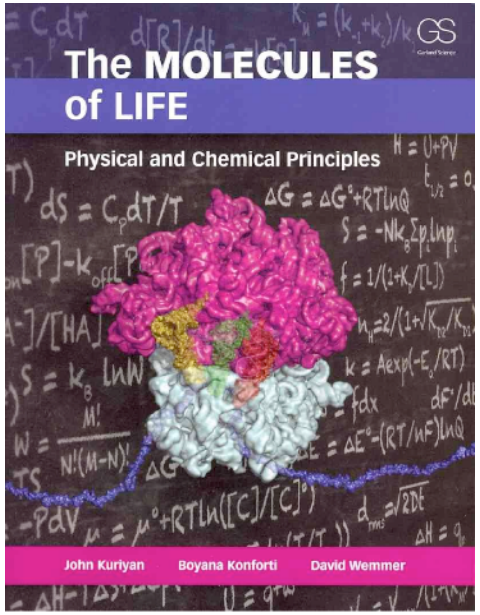
Lecture 6 - Outline

Today:

- Protein structure and folding
- Thermodynamics concept for biochemistry

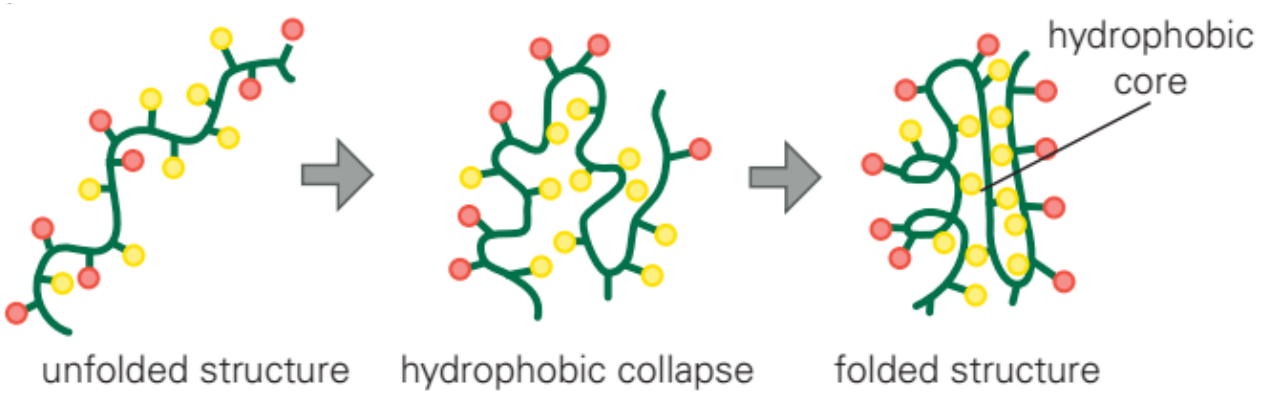
Reading suggestions:

- The Molecules of Life (Chapters 7-10)



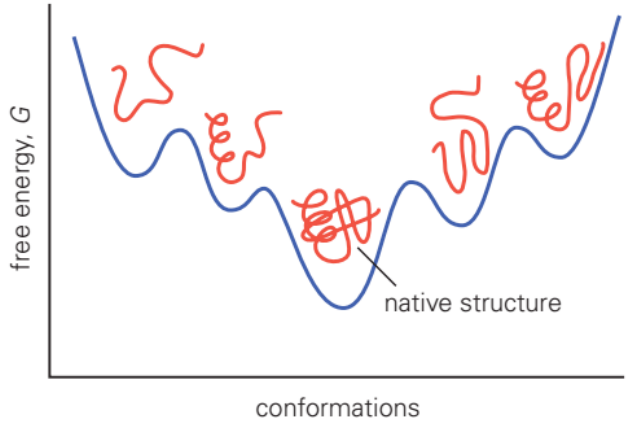
Sequence Determines Structure

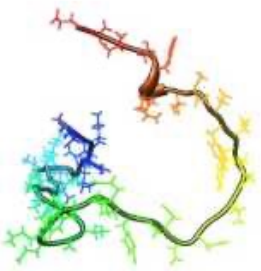
Proteins fold into defined 3D structures

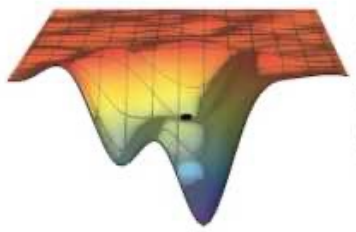


the folding process is energetically driven and proteins tend to fold to what we call an energy global minima (spontaneously or aided by molecular chaperones)

thermodynamic hypothesis







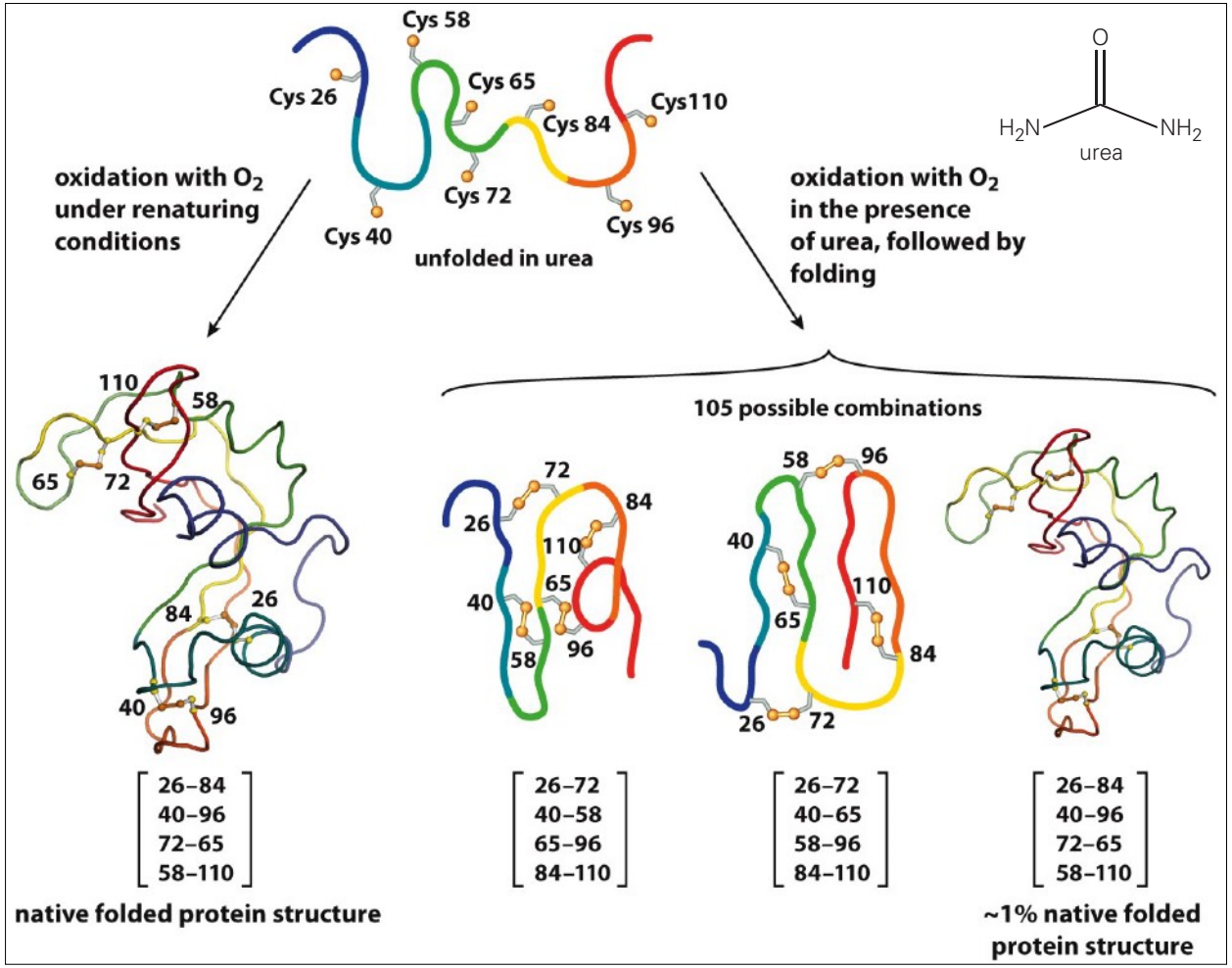
Sequence Determines Structure

The Anfinsen experiment (1954)

A Nobel prize (1972) experiment that by measuring enzymatic activity figured out the principles of protein folding

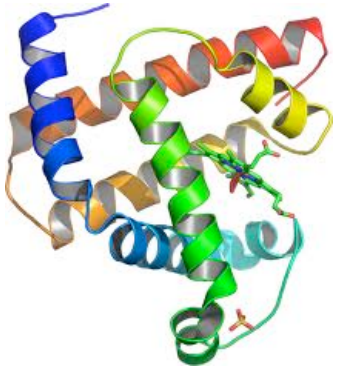
Ribonuclease refolded and oxidized recovers 100% activity

Ribonuclease oxidized and refolded in presence of urea recovers 1% activity

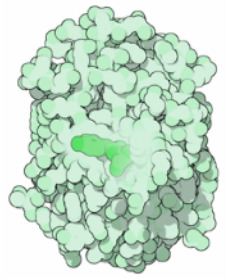


Protein structures

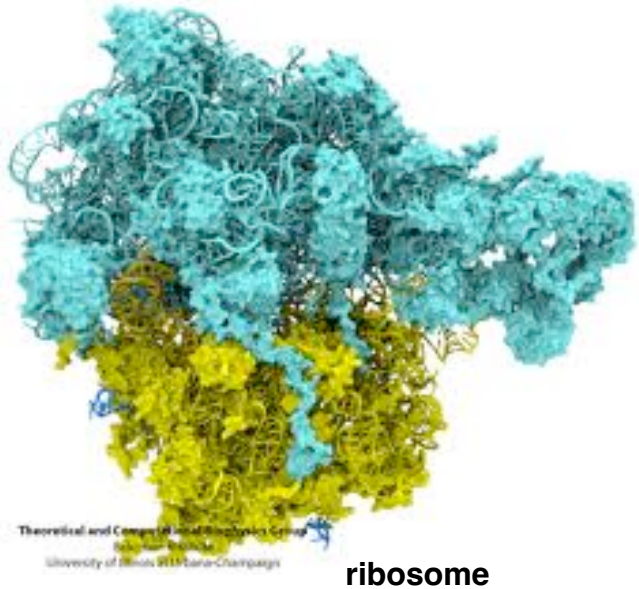
Matteo Dal Peraro



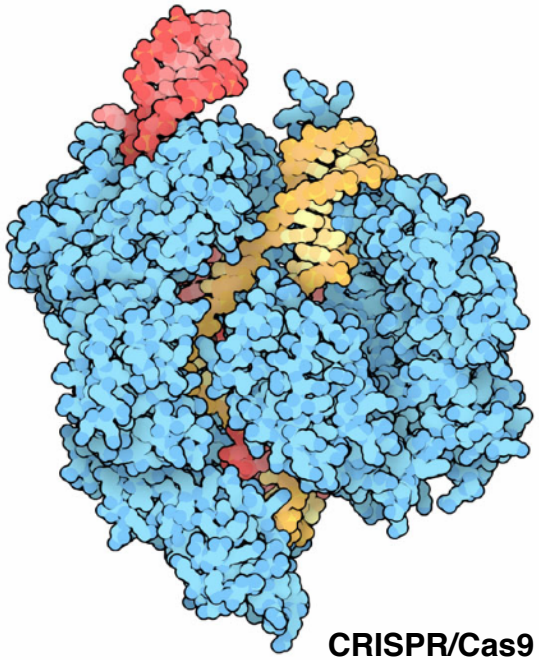
myoglobin 1959
(Nobel in Chemistry 1962)



GFP
(Nobel in Chemistry 2008)



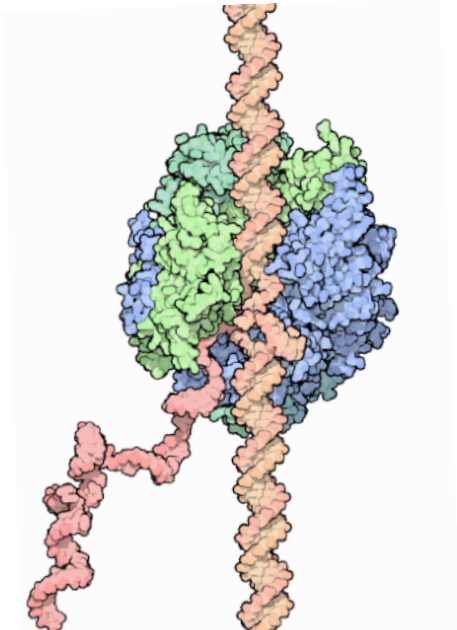
ribosome
(Nobel in Chemistry 2009)



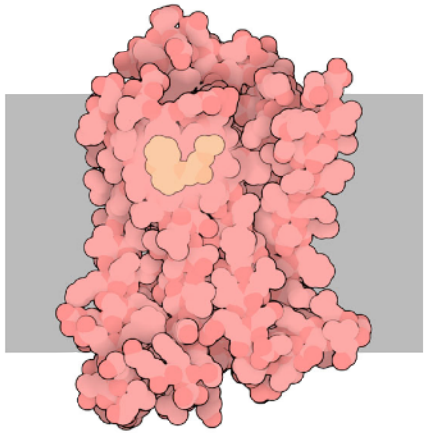
CRISPR/Cas9
(Nobel in Chemistry 2020)

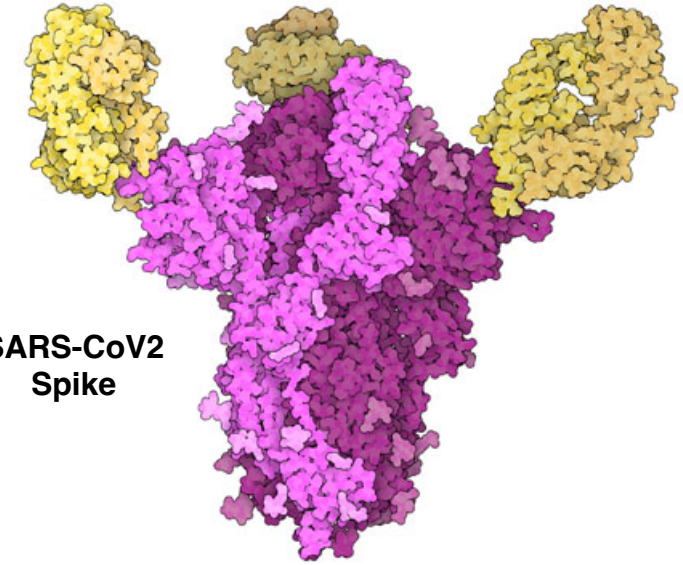
KcsA potassium channel
(Nobel in Chemistry 2003)

RNA polymerase II
(Nobel in Chemistry 2006)



GPCRs
(Nobel in Chemistry 2012)





SARS-CoV2 Spike

http://www.rcsb.org

Structure *is* function

If you want to understand function, study structure

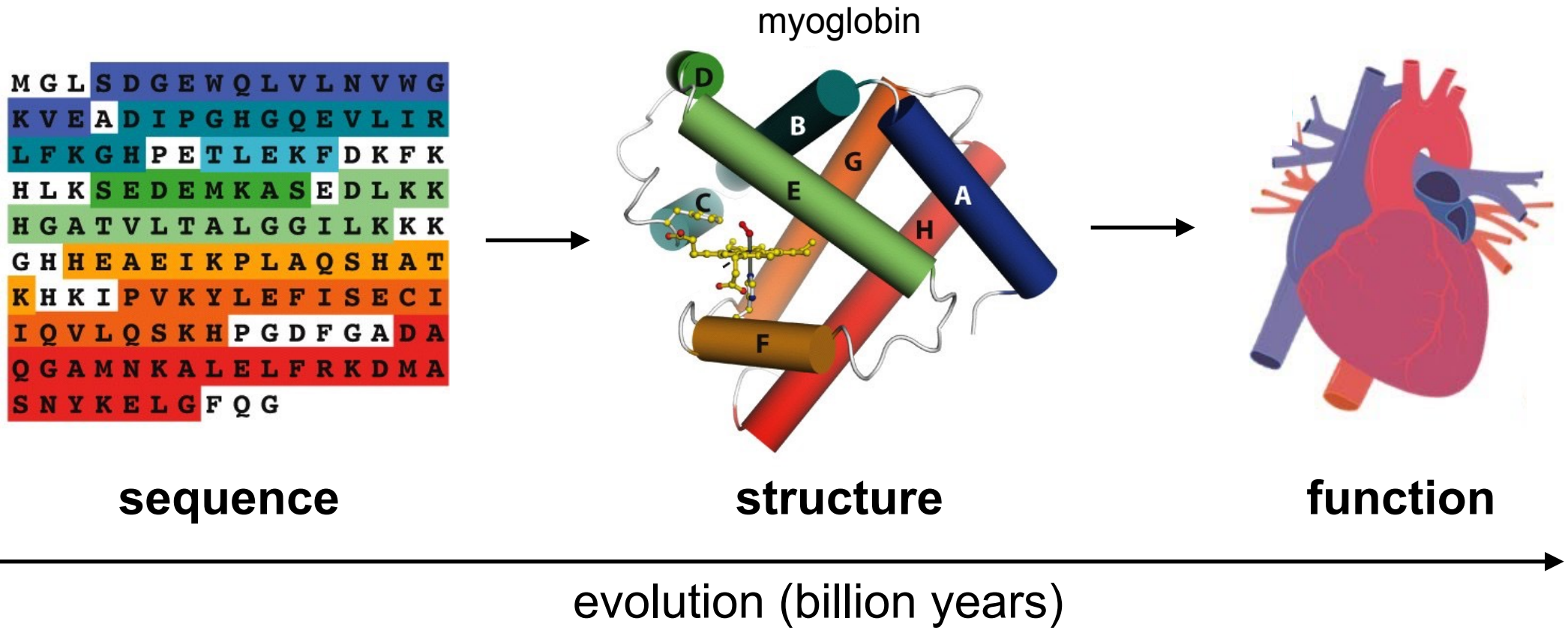
F. Crick

Pauling and Corey (1951)

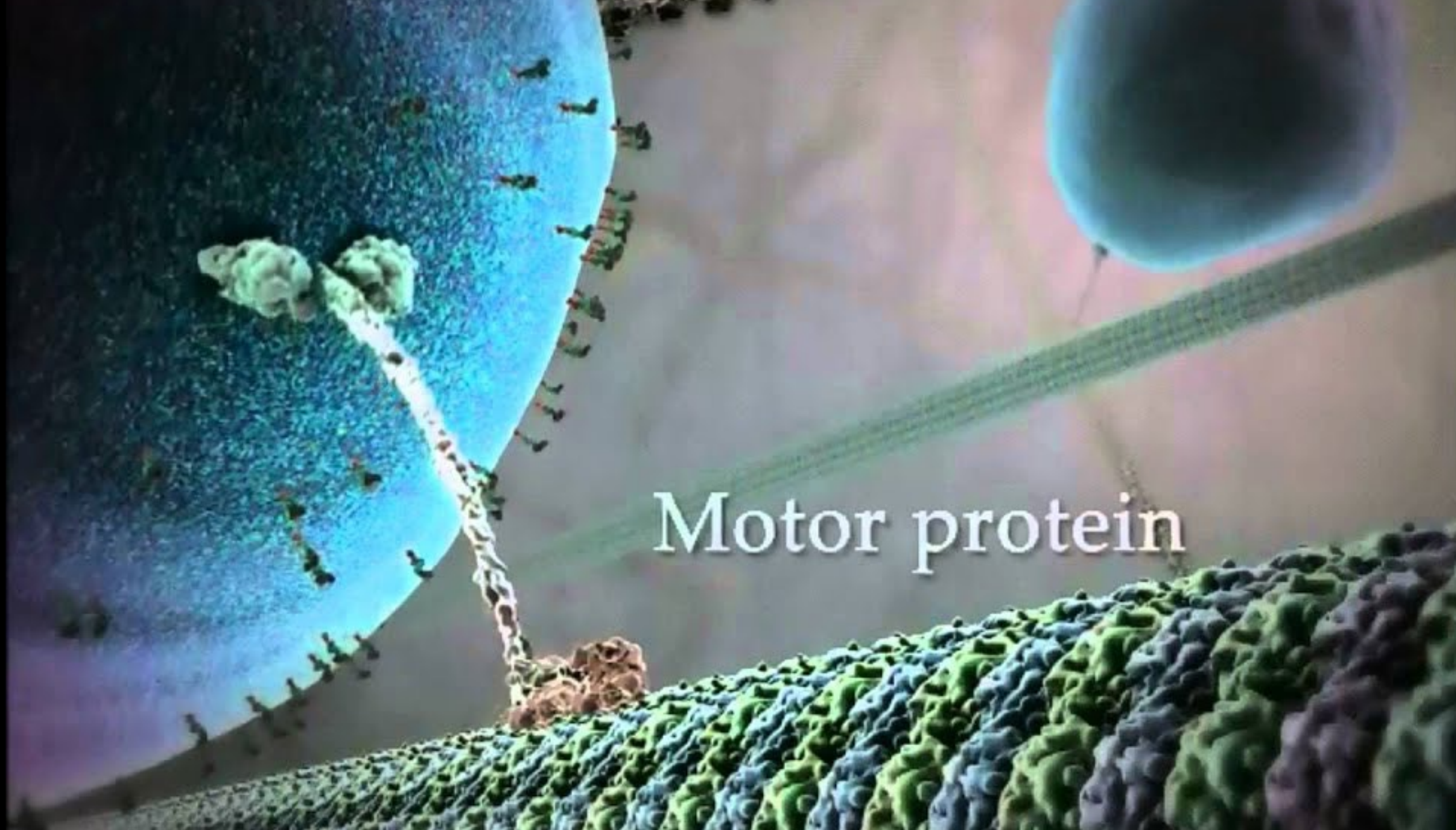
Franklin, Watson and Crick (1953)

Perutz and Kendrew (1959)

The paradigm of protein science



- Anfinsen discovered it in a key experiment already in 1954


Motor protein

Spontaneous Reactions

Favorable energy changes in energy *U* are not sufficient to indicate the direction of the spontaneous change

$$A + B \rightarrow C + D \quad \Delta U = \text{negative} \quad (\text{energy released})$$

Examples for spontaneous macroscopic reactions:

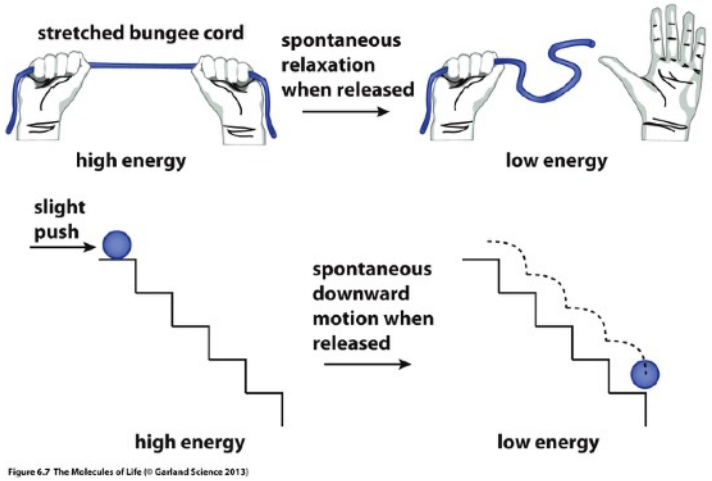
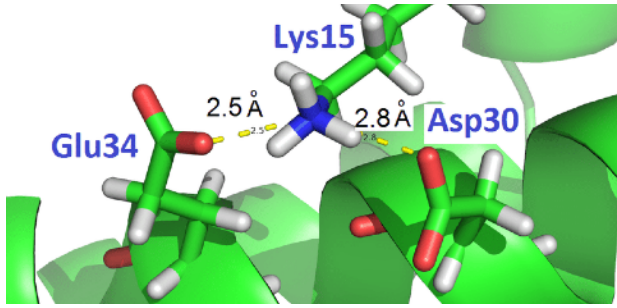


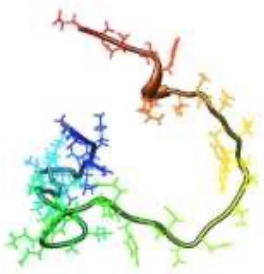
Figure 6.7 The Molecules of Life (© Garland Science 2013)

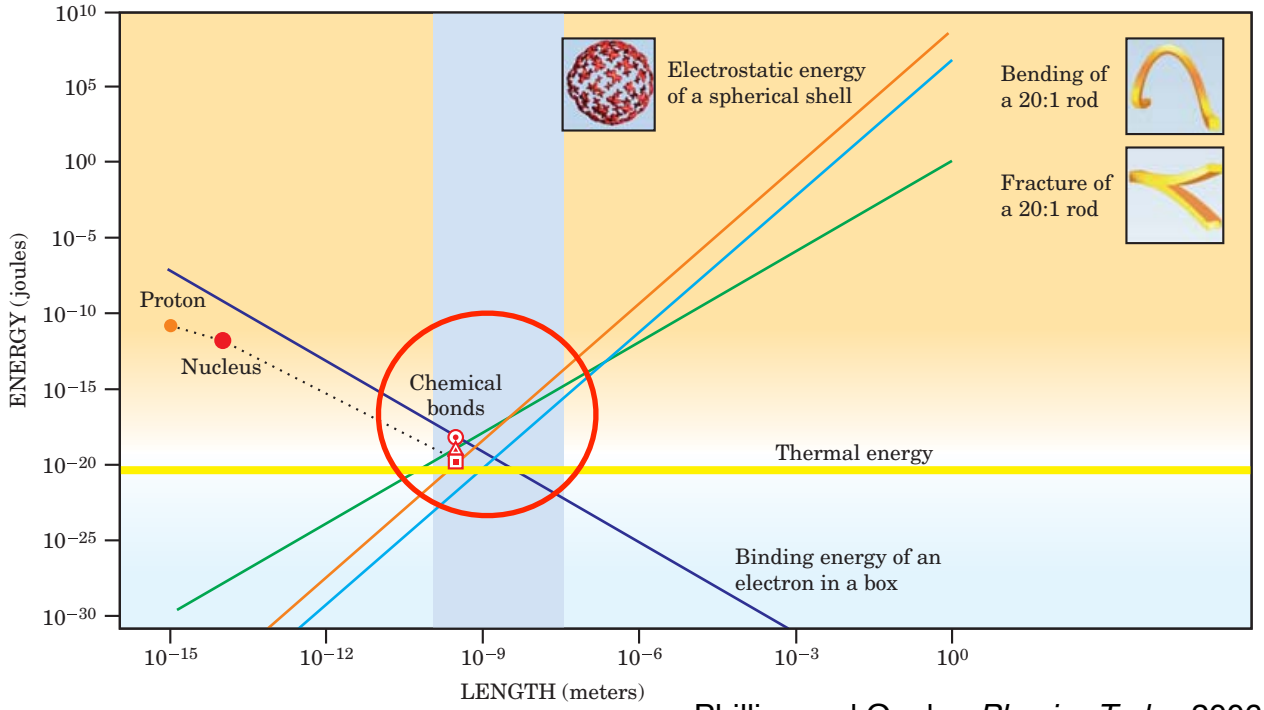
This intuition for microscopic systems is not always correct, because:
- The collective behavior also depends on **entropy**
- there are **kinetic barriers** for the reactions

The intriguing nature of biological interactions

- biological systems are subjected to deterministic forces (enthalpy) and thermal forces (entropy)
- at the dimension scale of biological systems these are however on the same order of magnitude
- all transformations in cells are thus determined by this subtle interplay, defined by the free energy of the system (**G=H-TS**) (accuracy in a noisy world, and use of thermal fluctuations to deploy biological function)







Phillips and Quake, *Physics Today* 2006

E_{det}/k_BT

k_BT = 4.1 pN*nm
= 0.6 kcal/mol
= 2.5 kJ/mol
= 0.025 eV

at room temperature (300K)

Gibbs Free Energy

How do we know when a biological process is at equilibrium or at least tend to the equilibrium conditions? We introduce a new function that will include the 1st and 2nd law, ie. conservation of energy and maximisation of entropy. This is the free energy (**Gibbs free energy, G** at constant pressure)

$$G = H - TS \qquad \text{with} \quad H = U + pV$$

which always decreases when a process occurs spontaneously and it is at a minimum at equilibrium at constant pressure and temperature

$$dG = dH - TdS \qquad \text{with} \quad dG \leq 0$$

- G is only dependent on the system
- It accounts for both enthalpic and entropic contributions

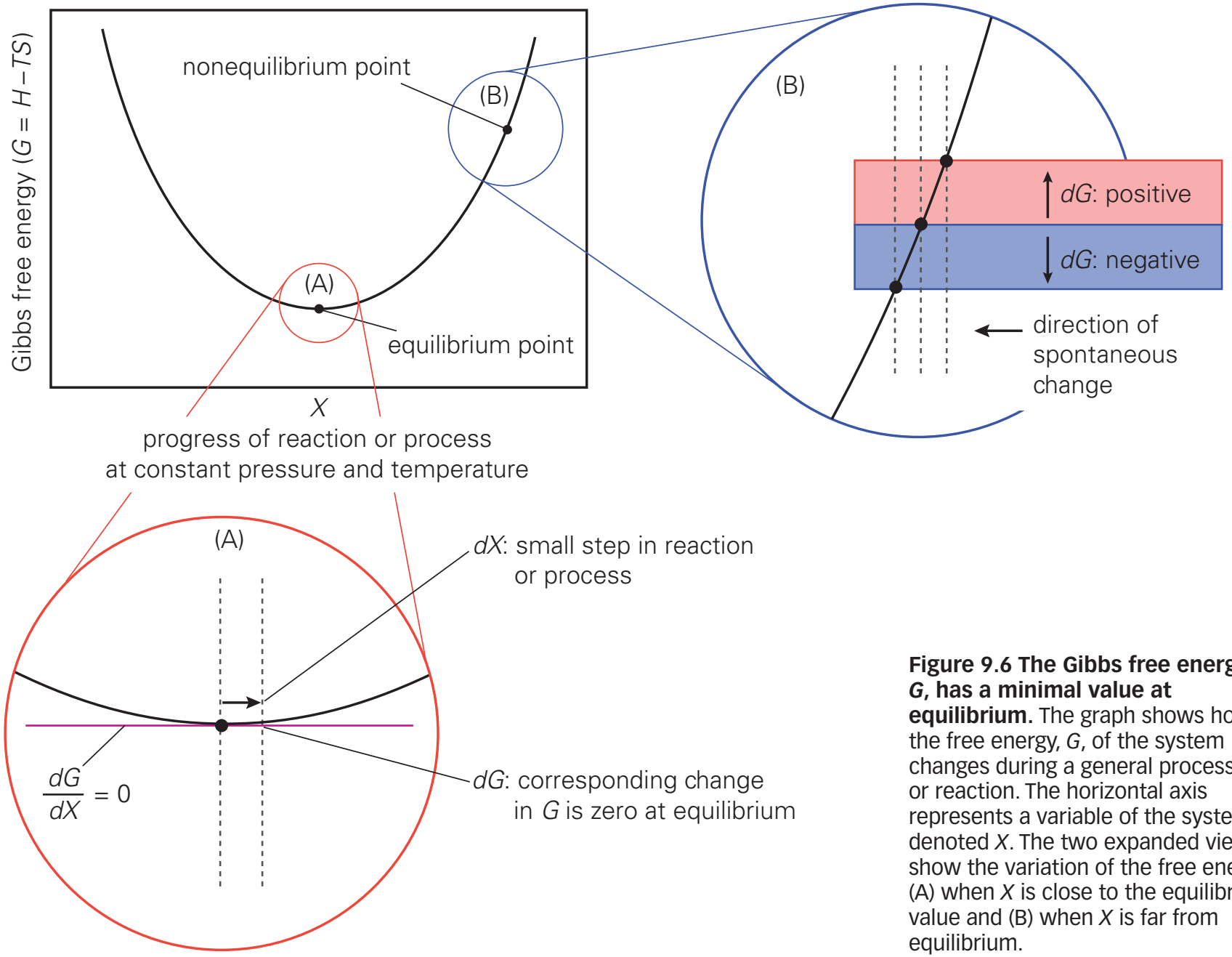


Figure 9.6 The Gibbs free energy, *G*, has a minimal value at equilibrium. The graph shows how the free energy, *G*, of the system changes during a general process or reaction. The horizontal axis represents a variable of the system, denoted *X*. The two expanded views show the variation of the free energy (A) when *X* is close to the equilibrium value and (B) when *X* is far from equilibrium.

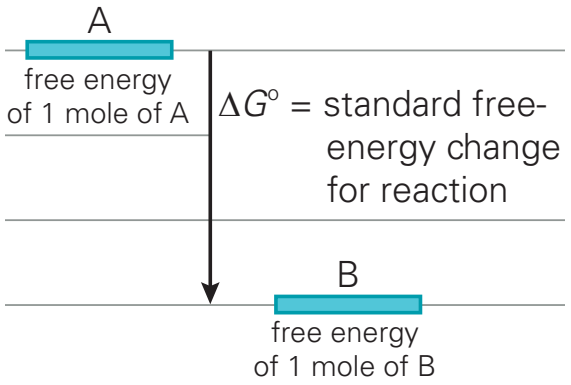
Free energy of chemical reactions

Consider ATP hydrolysis:

$$\mathrm{ATP} + \mathrm{H_2O} \rightarrow \mathrm{ADP} + \mathrm{P_i}$$

$$\Delta G = \int_{\text{reactants}}^{\text{products}} dG = G(\text{products}) - G(\text{reactants}) = G(\mathrm{ADP} + \mathrm{P_i}) - G(\mathrm{ATP} + \mathrm{H_2O})$$

Usually we refer to the standard ΔG^o at standard conditions: ie pressure 1 atm, 1 M of solute, apart for water (55 M), and room temperature 298 K. Thus at these conditions for ATP hydrolysis is $\Delta G^0 = -28\ \mathrm{kJ \cdot mol^{-1}}$

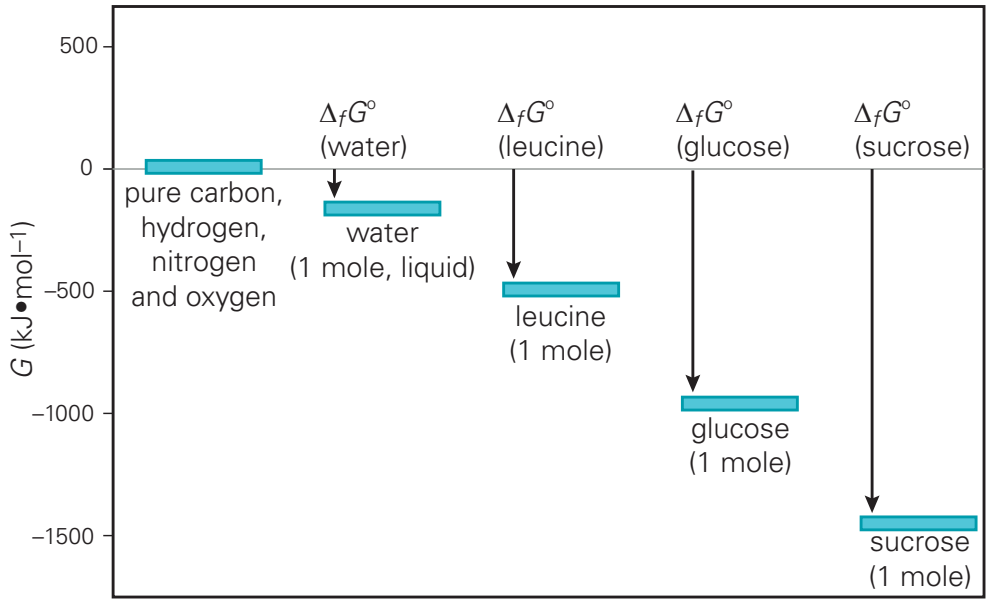


Free Energy of chemical reactions

Compound	$\Delta G^{o\prime}$ (kJ•mol^{-1})
acetate$^-$	–369.2
CO_2 (gas)	–394.4
CO_2 (aqueous solution)	–386.2
carbonate ion	–587.1
ethanol	–181.5
fructose	–915.4
fructose-6-phosphate^{2-}	–1758.3
α-D-glucose	–917.2
glucose-6-phosphate^{2-}	–1760.2
H^+ (aqueous solution)	0.0
H_2 (gas)	0.0
H_2O (liquid)	–237.2
isocitrate^{3-}	–1160.0
lactate$^-$	–516.6
OH^-	–157.3
pyruvate$^-$	–474.5
succinate^{2-}	–690.2

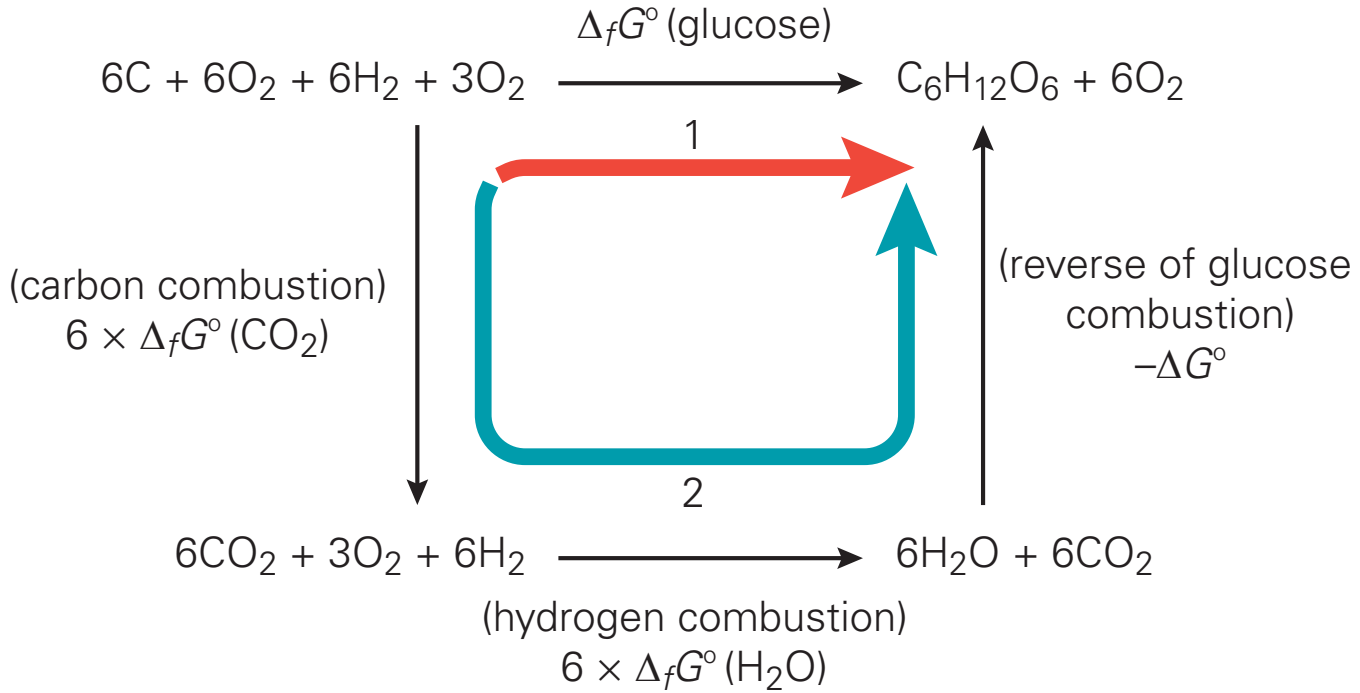
How to calculate these free energies? Using the free energy of formation of the molecules involved in the reactions, starting from the composing elements

$$\Delta G^o = \sum_{\text{all products}} \Delta_f G^o(\text{product}) - \sum_{\text{all reactants}} \Delta_f G^o(\text{reactant})$$

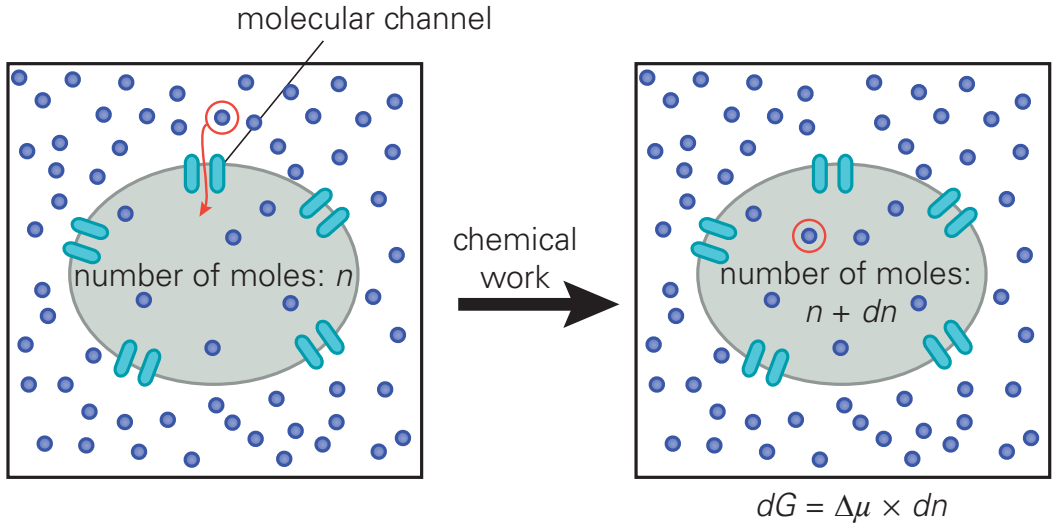


Free energy of chemical reactions

Since it might be difficult to measure the free energy of formation for some given products in one single step, the reaction is broken down in intermediate steps involving less complex reactions and thermodynamic cycles are used to calculate the final free energy (this is working because **G is a state function**)



Chemical work



- the most important source of free energy for the cell comes from chemical work - e.g. transformation or transfer of matter - changes in the free energy of the system that result from changes in the number of molecules.

$$\Delta G = \int \Delta\mu \; dn$$

- the **chemical potential** takes into account when a system loses or gains molecules, or molecules change states.

- chemical potential of a type of molecule is simply the free energy of one mole of these molecules under the specified conditions (units are in fact J/mole)

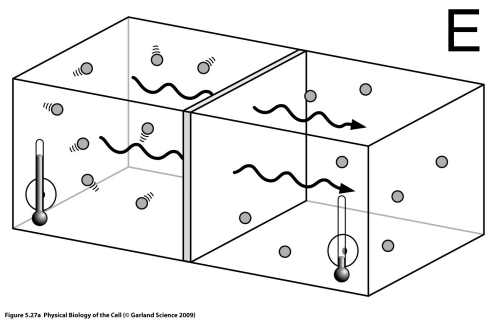
$$\mu_i = \left(\frac{\partial G}{\partial N_i} \right)_{T,P,N_{i \neq j}} \approx \frac{\Delta G}{\Delta N_i} = G \text{ (for one molecule)}$$

Chemical potential

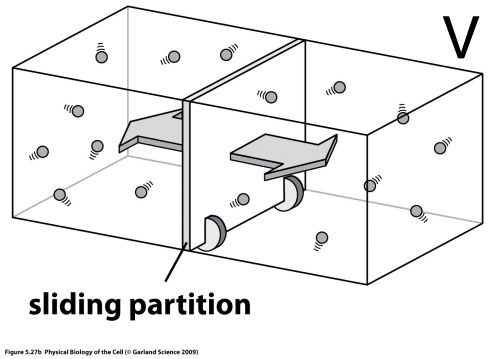
- it is the thermodynamic quantity that determines the equilibrium conditions, the tendency of a molecule to escape from a region to another and/or to react from a given state to another

- like T is the entropic force driving energy transfer, the chemical potential drives the net transfer or transformation of particles

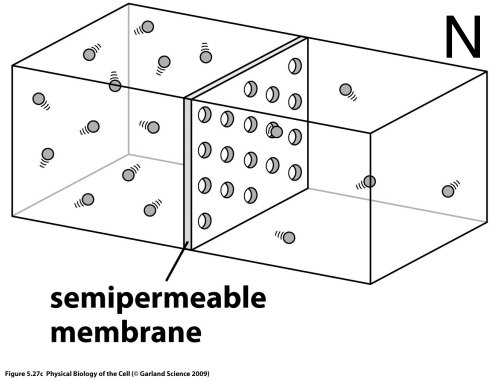
- from the maximization of entropy, you obtain the **equilibrium condition**:



$$T_A = T_B$$



$$p_A = p_B$$



$$\mu_A = \mu_B$$

Chemical potential driving forces

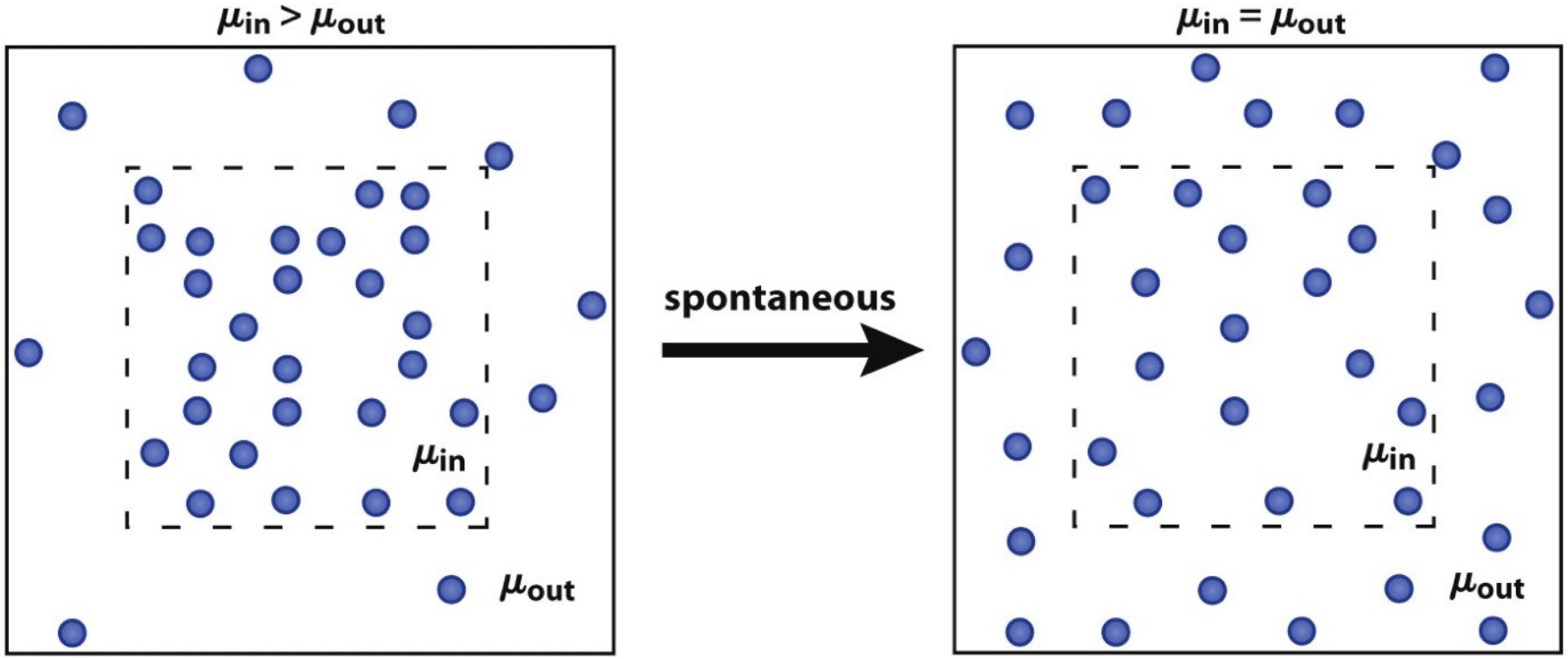


Figure 10.1 The Molecules of Life (© Garland Science 2013)

$$dG = (\mu_{in} - \mu_{out})\, dN_{in} < 0$$

therefore molecules move spontaneously from regions of high chemical potential to regions of low chemical potential

Chemical potential and concentration solution

- chemical potential of a solute is related to the **logarithm of the concentration** of the solute as a consequence of the positional entropy of the solute
- if you consider a ideal solute solution ($N_{waters} >> N_{molecules}$)

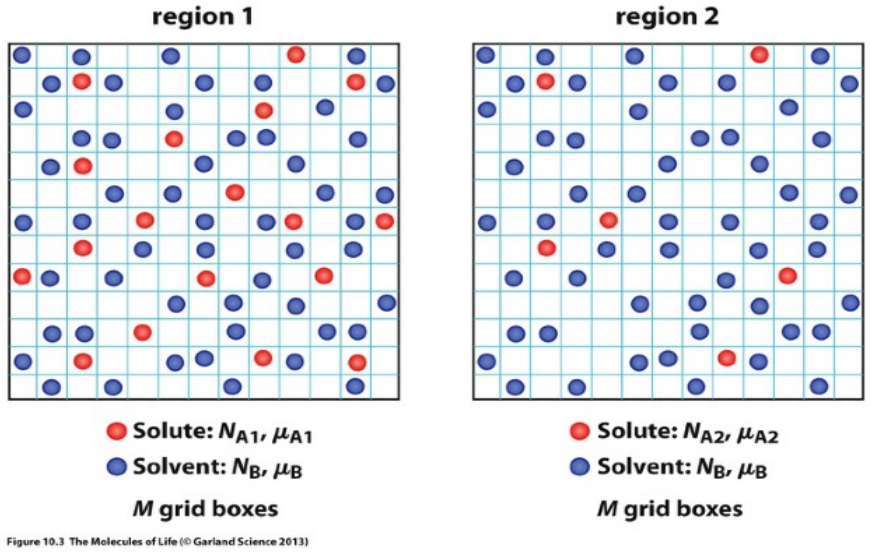


Figure 10.3 The Molecules of Life (© Garland Science 2013)

$$\Delta\mu = \mu_2 - \mu_1 = RT\ln\left(\frac{C_2}{C_1}\right)$$

we expect molecules to move spontaneously from region 1 (high concentration) to region 2 (low concentration), which is in the direction of of decreasing chemical potential.

in reality the chemical potential is also always defined with respect to standard conditions

$$\mu = \mu^\circ + RT\ln\left(\frac{C}{C^\circ}\right) = \mu^\circ + RT\ln C$$

- see full derivation at page 417

gas constant $R = N_A k_B = 8.3145$ J/K*mol

Equilibrium Constants

for a generic reaction $\upsilon_A A + \upsilon_B B \rightleftharpoons \upsilon_C C + \upsilon_D D$

where ν_i are the stoichiometric coefficients, imposing the free energy is at the minimum ie dG=0, equilibrium conditions, you obtain that

$$\upsilon_A \mu_A + \upsilon_B \mu_B = \upsilon_C \mu_C + \upsilon_D \mu_D$$

Therefore, isolating the chemical potential at standard conditions and using the relationship between chemical potential and concentration:

$$\upsilon_C \mu_C^o + \upsilon_C \mu_D^o - \upsilon_A \mu_A^o - \upsilon_B \mu_B^o = -RT\ln\frac{[C]_{eq}^{\upsilon_C}\,[D]_{eq}^{\upsilon_D}}{[A]_{eq}^{\upsilon_A}\,[B]_{eq}^{\upsilon_B}} = \Delta G^o$$

Defining the **equilibrium constant K_{eq}**

$$K_{eq} = \frac{[C]_{eq}^{\upsilon_C}\,[D]_{eq}^{\upsilon_D}}{[A]_{eq}^{\upsilon_A}\,[B]_{eq}^{\upsilon_B}}$$

we obtain that $\Delta G^o = -RT\ln K_{eq}$ and $K_{eq} = e^{-\left(\frac{\Delta G^o}{RT}\right)}$

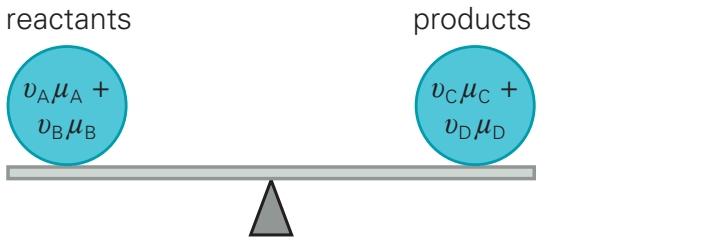
see full derivation at page 422

Equilibrium Constants

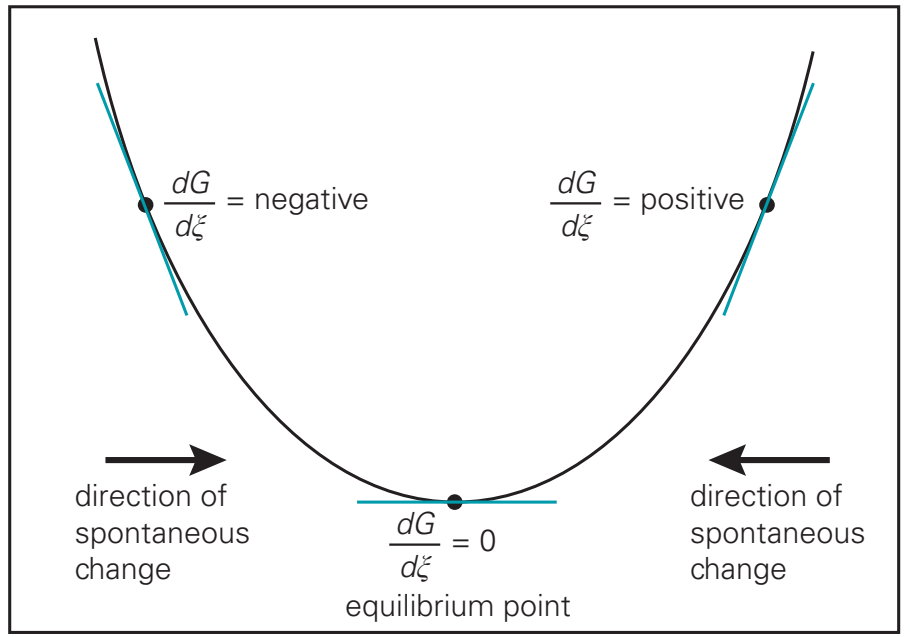
Remember that the equilibrium constant K_{eq} is adimensional because also in the case the stoichiometric coefficients do not cancel out - all concentrations are refereed with respect to the standard conditions

$$K_{eq} = \frac{\left(\frac{[C]_{eq}}{[C]^o}\right)^{\upsilon_C}\left(\frac{[D]_{eq}}{[D]^o}\right)^{\upsilon_D}}{\left(\frac{[A]_{eq}}{[A]^o}\right)^{\upsilon_A}\left(\frac{[B]_{eq}}{[B]^o}\right)^{\upsilon_B}}$$

$$\Delta G^o = -RT \ln K_{eq}$$



pictorial representation
on how to reach equilibrium ->



$$dN_i = \nu_i d\xi$$

ξ:: reaction coordinate

Example of ATP hydrolysis

$$ATP + H_2O \rightarrow ADP + P_i \qquad \Delta G^o = -28\ kJ \bullet mol^{-1}$$

Equilibrium constant (water is neglected as concentration does not change):

$$K = \frac{[ADP][P_i]}{[ATP]}$$

We can compute K using $K_{eq} = e^{-\left(\frac{\Delta G^o}{RT}\right)}$

$$K = e^{+28/2.478} = e^{+11.3} \approx \left(10^{0.43}\right)^{11.3} \approx 10^5$$

Assuming that $[P_i]$ is constant 10^{-2} M (10 mM within physiological levels)

$$\frac{[ADP]}{[ATP]} = 10^7$$

What do you think its the extent of this reaction ?

Example of ATP hydrolysis

$$ATP + H_2O \rightarrow ADP + P_i \qquad \Delta G^o = -28\ kJ \bullet mol^{-1}$$

Equilibrium constant (water is neglected as concentration does not change):

$$K = \frac{[ADP][P_i]}{[ATP]}$$

We can compute K using $K_{eq} = e^{-\left(\frac{\Delta G^o}{RT}\right)}$

$$K = e^{+28/2.478} = e^{+11.3} \approx \left(10^{0.43}\right)^{11.3} \approx 10^5$$

Assuming that $[P_i]$ is constant 10^{-2} M (10 mM within physiological levels)

$$\frac{[ADP]}{[ATP]} = 10^7$$

Thus at equilibrium almost all the ATP will indeed have converted to ADP, except for roughly 1 part per 10 million. The substantial free energy difference between ATP and its hydrolysis products does indeed **drive the reaction nearly to completion.**

Mass action ratio

If we are far from equilibrium, dealing with observed concentrations not at equilibrium conditions, how do we calculate ΔG?

$$\Delta G = \left[\upsilon_C \mu_C^o + \upsilon_D \mu_D^o - \upsilon_A \mu_A^o - \upsilon_B \mu_B^o \right] + RT \ln \frac{[C]_{obs}^{\upsilon_C} [D]_{obs}^{\upsilon_D}}{[A]_{obs}^{\upsilon_A} [B]_{obs}^{\upsilon_B}}$$

$$\Rightarrow \boxed{\Delta G = \Delta G^o + RT \ln Q}$$

and we can define the **reaction quotient** as Q: $$Q = \frac{[C]_{obs}^{\upsilon_C} [D]_{obs}^{\upsilon_D}}{[A]_{obs}^{\upsilon_A} [B]_{obs}^{\upsilon_B}}$$

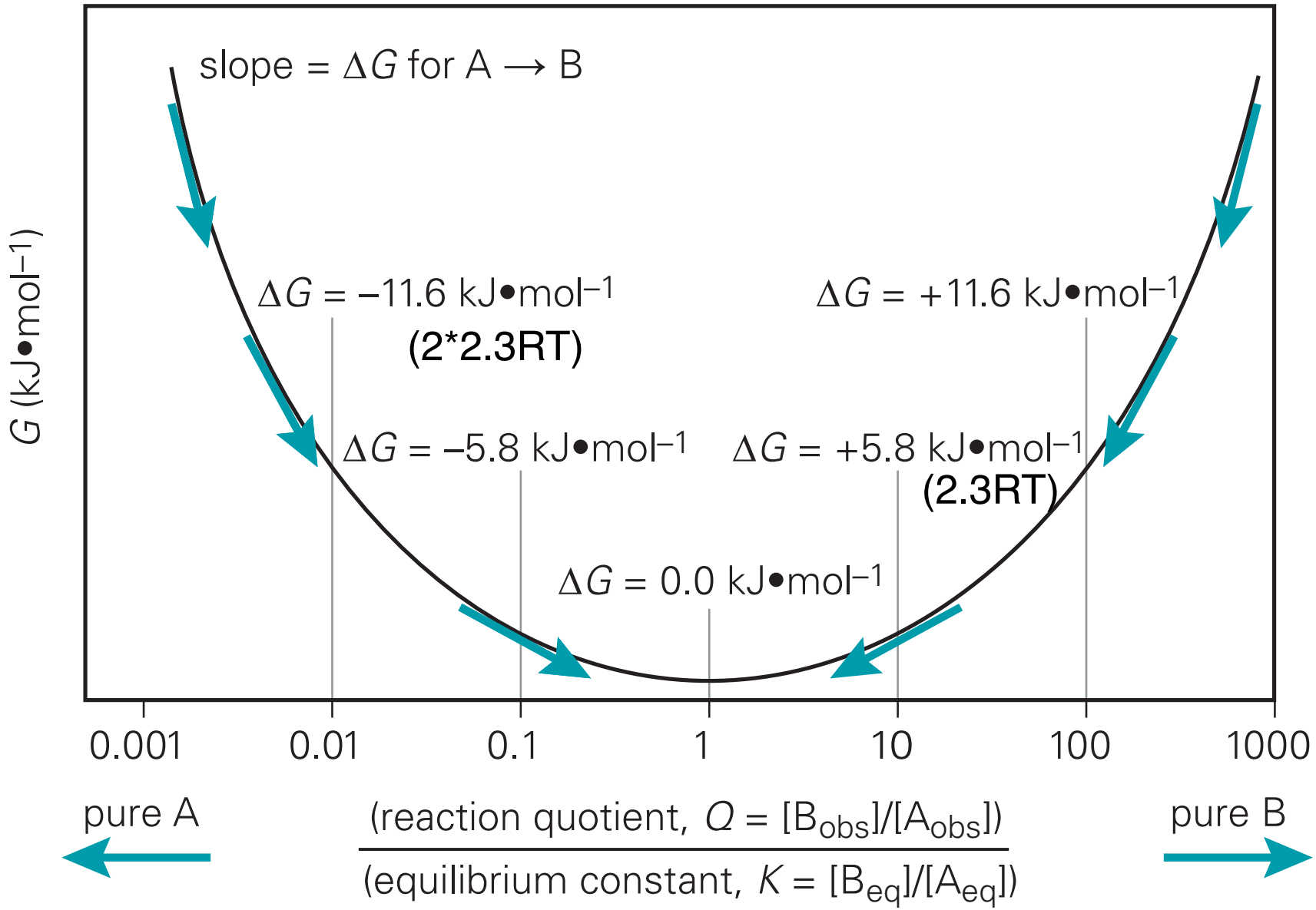
introduction the equilibrium constant for better comparison:

$$\Delta G = \Delta G^o + RT \ln Q = -RT \ln K + RT \ln Q$$

$$\Rightarrow \boxed{\Delta G = +RT \ln \frac{Q}{K} = +2.3RT \log_{10} \frac{Q}{K}}$$

the **ratio Q/K, the mass action ratio** gives us a way to understand where a reaction is going. If Q/K < 1 implies that ΔG < 0 and the reaction is going forward. When Q=K the reaction is at equilibrium.

$$\Delta G = +RT\ln\frac{Q}{K} = +2.3RT\log_{10}\frac{Q}{K}$$



However, in living cells, the ratio of ATP to ADP is held nearly constant, the reaction does not go to completion because the action of metabolic reactions that consume sugars and other sources of energy are used to synthesize ATP.

$$ATP + H_2O \rightarrow ADP + P_i \qquad \Delta G^o = -28\ kJ \bullet mol^{-1}$$

$$K = \frac{[ADP][P_i]}{[ATP]} \qquad \frac{[ADP]}{[ATP]} = 10^7$$

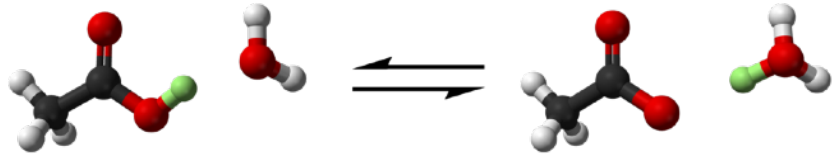
Q	Q/K	ΔG (kJ•mol^{-1})	[ATP]/[ADP]	Relevant condition
10^5	1	0	10^{-7}	equilibrium
10^3	10^{-2}	–11	10^{-5}	
1	10^{-5}	–28	10^{-2}	standard condition
10^{-3}	10^{-8}	–46	10^1	mitochondrial matrix
10^{-5}	10^{-10}	–57	10^3	cytoplasm

For instance ATP concentrations in the cytoplasm are held so that [ATP]/[ADP] is ~1000, storing more free energy than in standard conditions. Thus, the ATP hydrolysis reaction can be coupled to other reactions that are uphill in terms of free energy, making such transformations possible.

Acid-base equilibrium

Acids dissociate in water to release protons

$$HA \rightleftharpoons H^+ + A^-$$



Acid dissociation constant:

$$K_a = \frac{[H^+][A^-]}{[HA]}$$

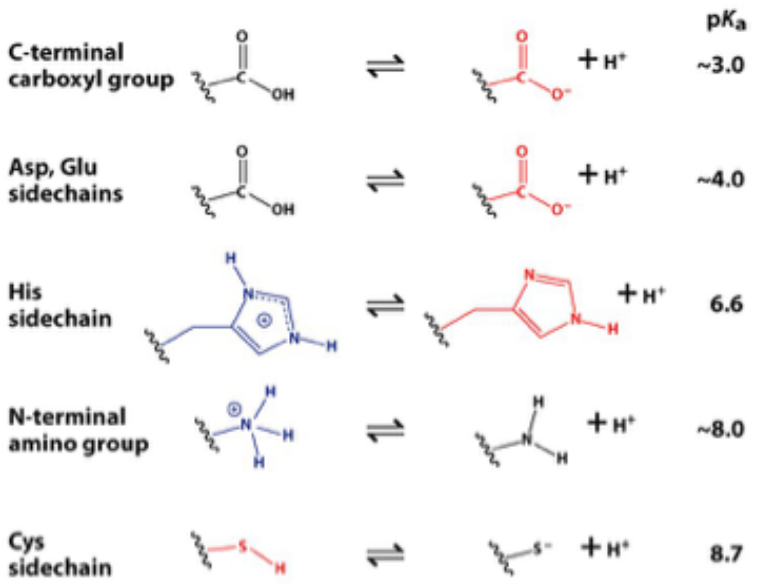
Applying the $\log_{10}$ and recalling that $pH = -\log_{10}([H^+])$ and $pK_a = -\log_{10}(K_a)$

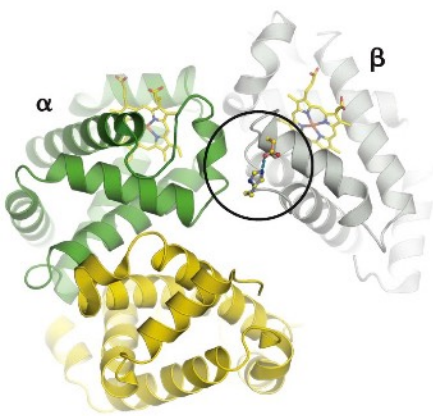
We obtain the Henderson-Hasselbalch equation:

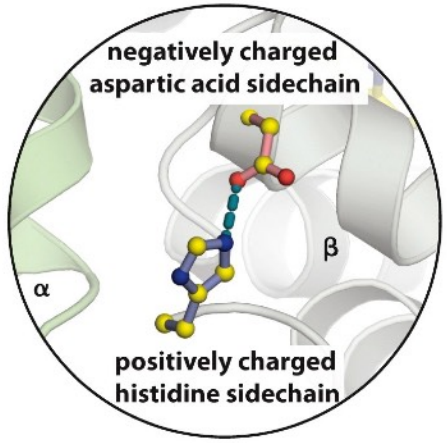
$$pH = pK_a + \log_{10}\left(\frac{[A^-]}{[HA]}\right)$$

We will revise it in the exercise session

Acid-base equilibrium and protonation states







With the **Henderson-Hasselbalch** equation and a few concepts from the **free energy and equilibria** you will be able to compute pKas for residues within the protein environment (see exercises)

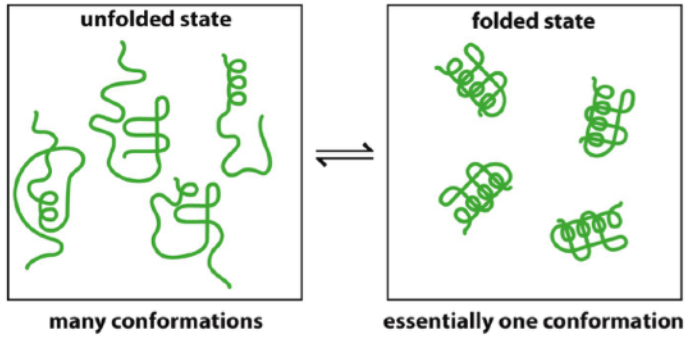
His for example has a pKa of 6 and at pH 7 we have that $\frac{[\mathrm{His}]}{[\mathrm{His}^+]} = 10$
and thus $\Delta G^{o\prime} = -RT\ln K = -8.314 \times 300 \times \ln 10 = -5.7\ \mathrm{kJ \bullet mol^{-1}}$
for the deprotonation reaction of His ($His^+ \longrightarrow His + H^+$).
If His is close to an Asp in a protein environment its pKa can raise up !

Free energy change in protein folding

Two concepts we are going to use to describe protein folding

$$K = e^{-\Delta G^{o}/RT} \qquad \Delta G^{o} = \Delta H^{o} - T\Delta S^{o}$$

Protein folding reaction can be simplified as transition from the unfolded state (U) to the folded state (F).



Thus that the folding reaction is

$$\mathrm{U} \rightleftharpoons \mathrm{F}$$

$$K_{\text{folding}} = \frac{[\mathrm{F}]}{[\mathrm{U}]} \quad (\text{at equilibrium})$$

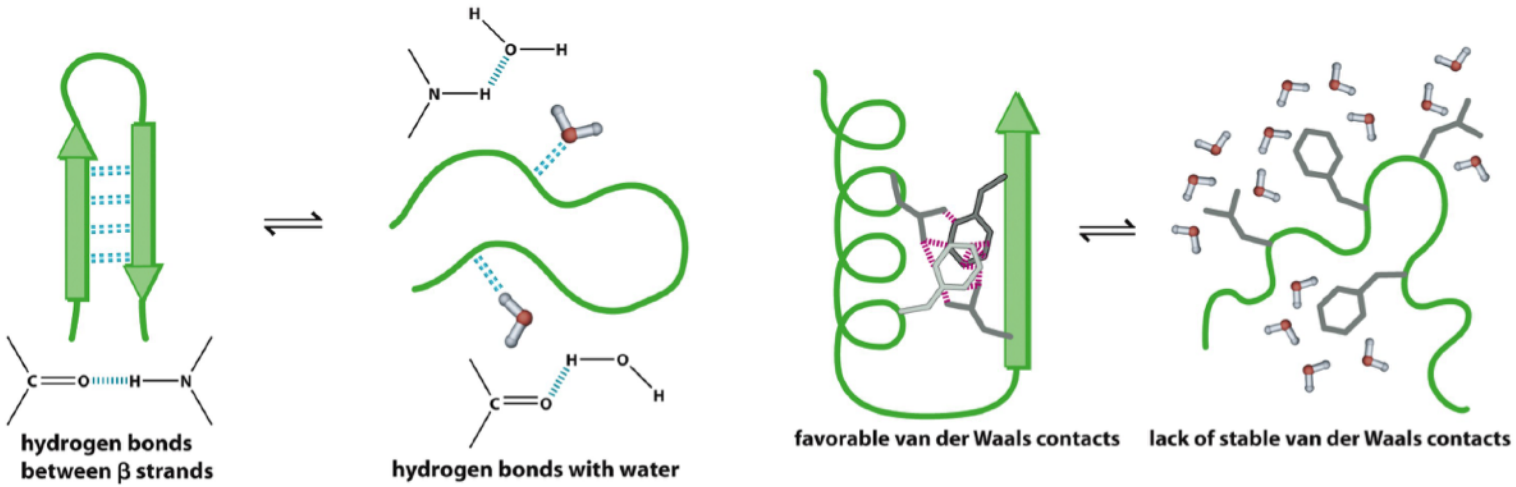
for the unfolding reaction we have $K_{\text{unfolding}} = \frac{[\mathrm{U}]}{[\mathrm{F}]} = K_{\text{folding}}^{-1}$

Energetic components in protein folding

Protein folding results from a balance between enthalpy and entropy

$$\Delta G^o = \Delta H^o - T\Delta S^o$$

As you know there are several types of molecular interactions in a folded protein



Which component gives a higher contribution and why ?

a) Enthalpy

b) Entropy

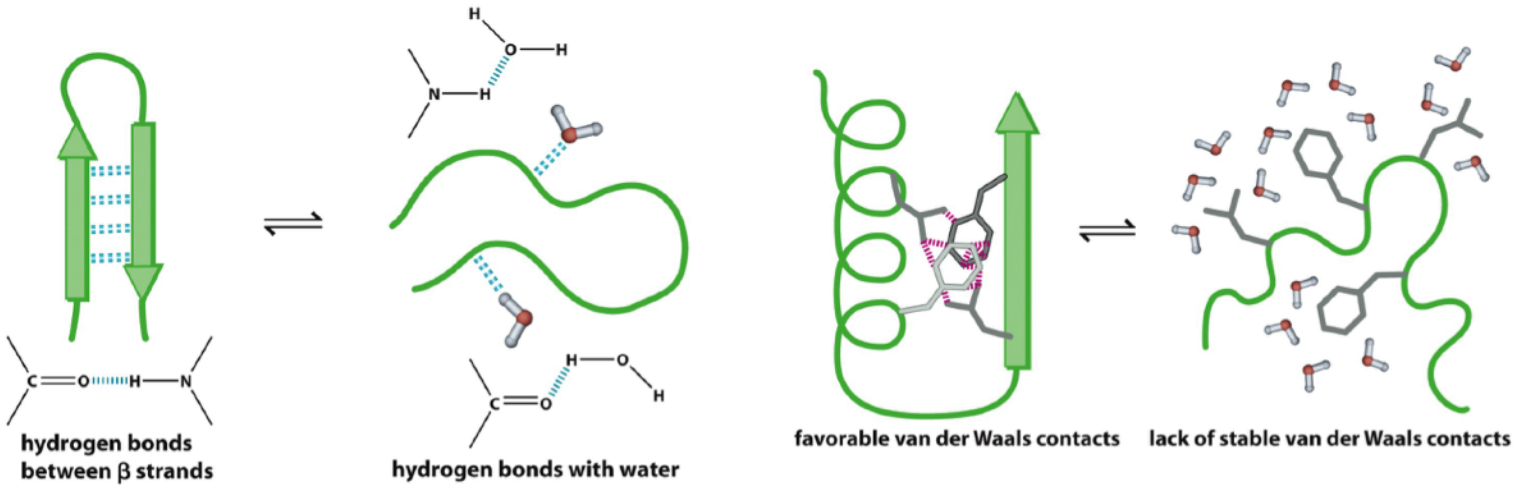
c) Both the same

Energetic components in protein folding

Protein folding folding result rom a balance between enthalpy and entropy

$$\Delta G^o = \Delta H^o - T\Delta S^o \quad < 0$$

As you know there are several types of molecular interactions in a folded protein



Enthalpy: H-bonds do not have a major effect, in U many with water are replaced in F within the protein. vdW interactions are more favourable in F but are not so strong - overall you get a small favourable H balance ($\Delta H<0$)

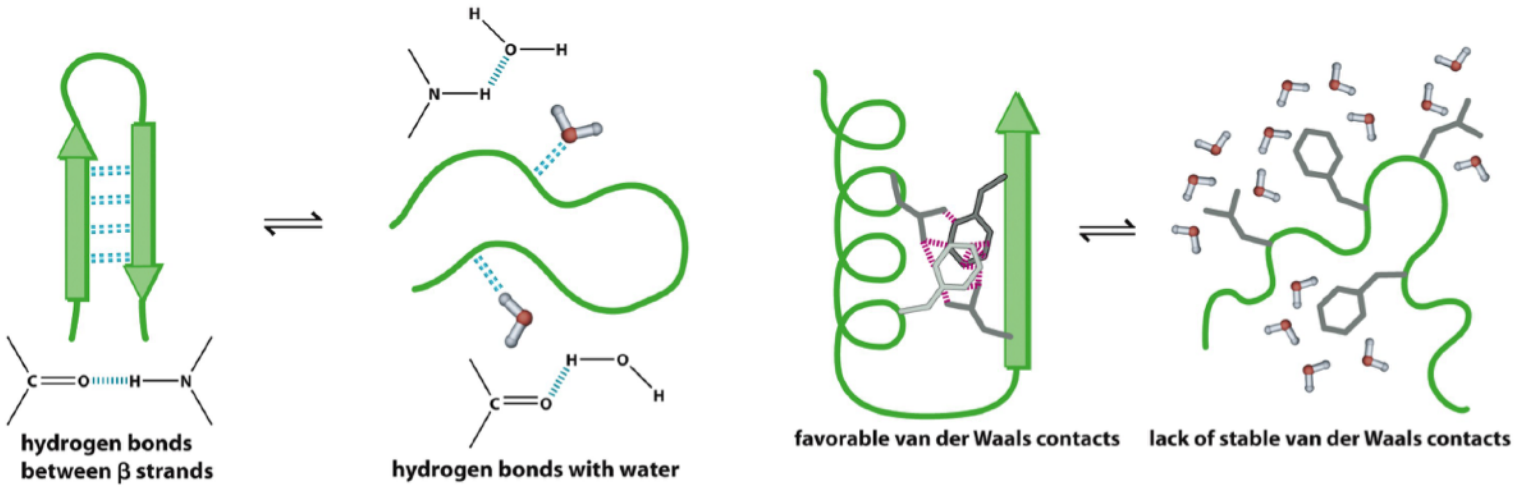
This can be around ΔH = -150 kJ/mol for a small size protein (~50 aa)

Energetic components in protein folding

Protein folding folding result rom a balance between enthalpy and entropy

$$\Delta G^o = \Delta H^o - T\Delta S^o \quad < 0$$

As you know there are several types of molecular interactions in a folded protein



Entropy: U state has high entropy (many different unfolded configurations).
The F state has low entropy (only one configuration, S=0)

$$\Delta S = R\ln N_C \qquad \Delta S^o = 0 - S^o_{\text{unfolded}} = -S^o_{\text{unfolded}}$$

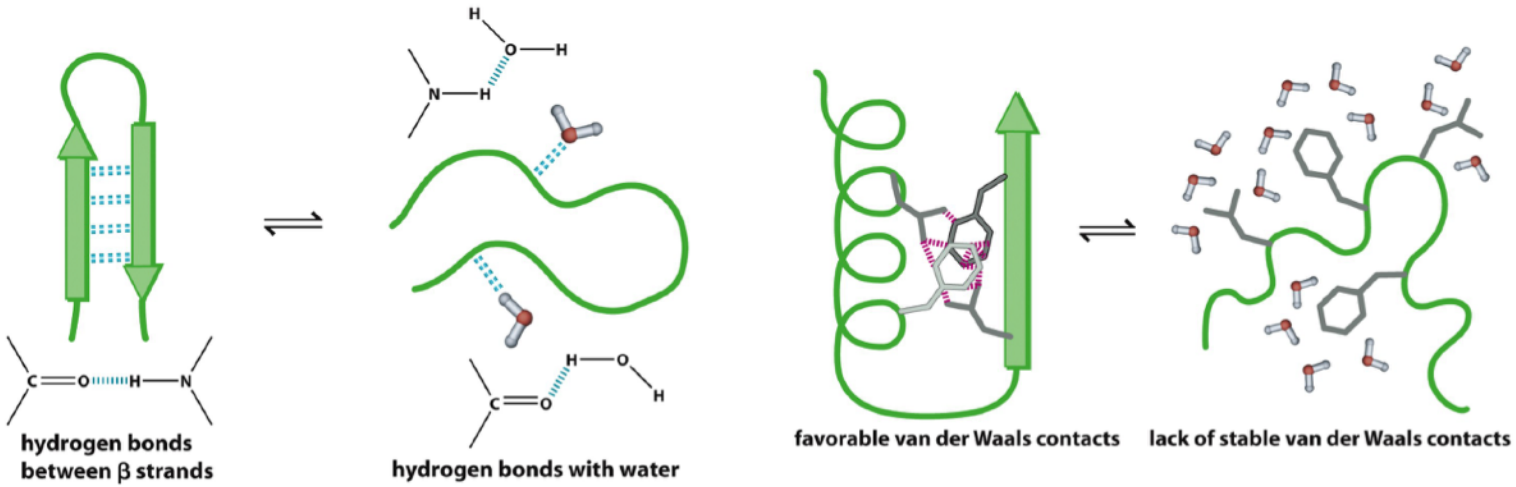
- Thus $-T\Delta S^0$ can be around +200 kJ/mol for a small size protein

Energetic components in protein folding

Protein folding folding result rom a balance between enthalpy and entropy

$$\Delta G^o = \Delta H^o - T\Delta S^o$$

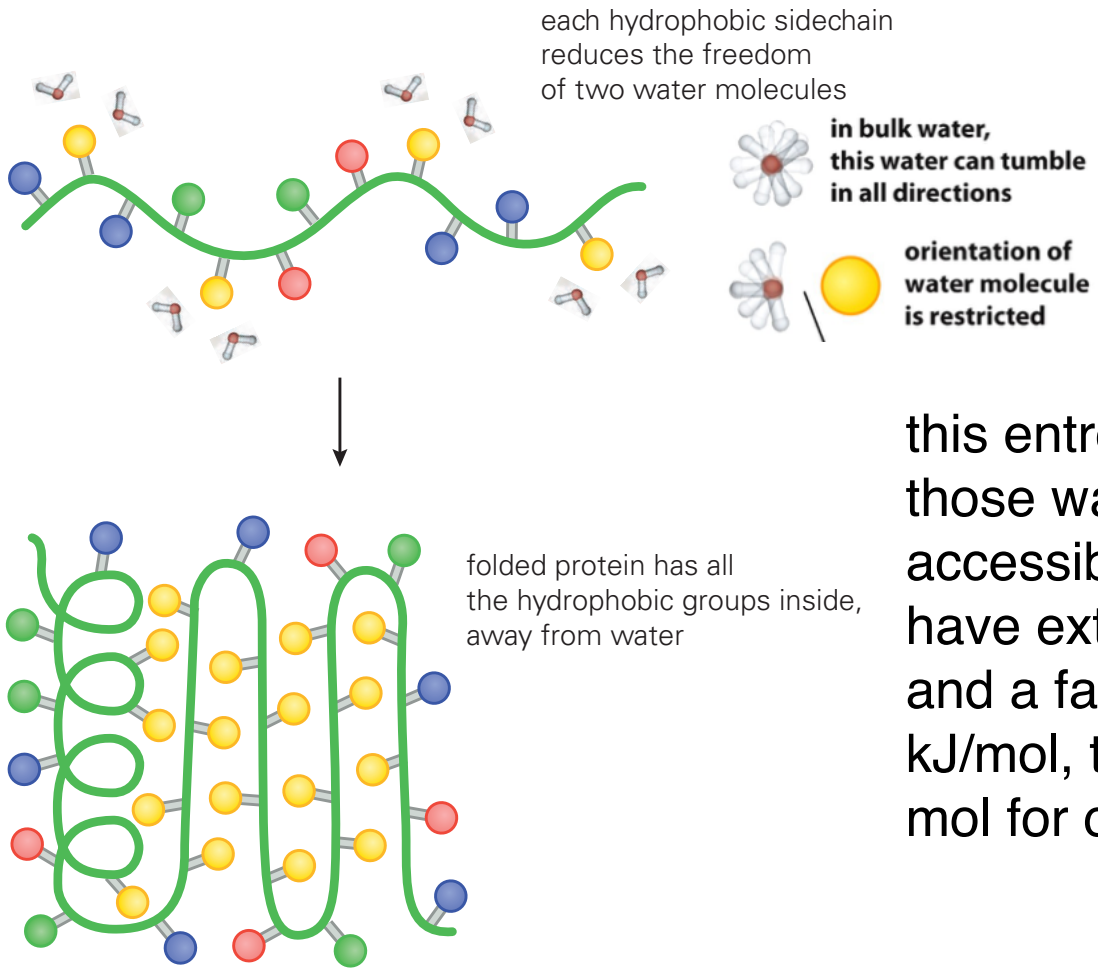
As you know there are several types of molecular interactions in a folded protein



Thus the overall ΔG is positive, ΔH - TΔS = -150 - (-200) ~50 kJ/mol, which is not what we expect as the protein actually folds, ie ΔG < 0. Something is missing - namely the hydrophobic effect

Energetic components in protein folding

Increase of entropy on the solvent (e.g. water) upon folding due to the so-called **hydrophobic effect,** provides another favorable term to the free energy of folding

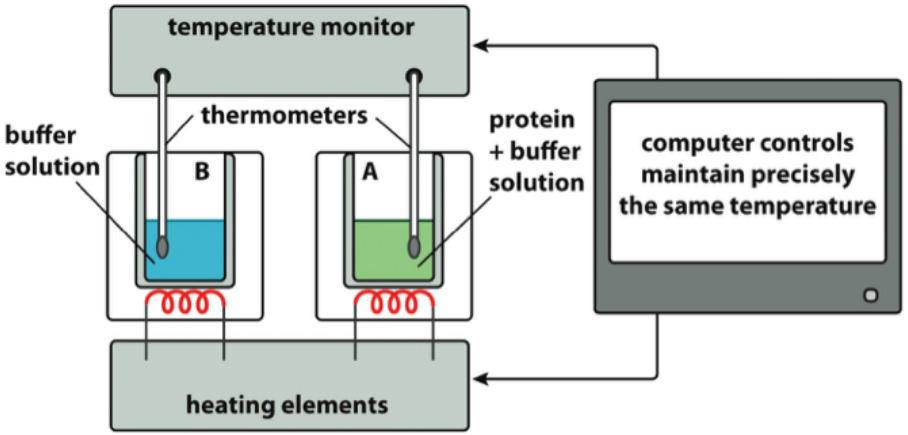


this entropy contributions is low in U state

this entropy contribution is high as those waters can have more accessible conformations, thus we have extra $\Delta S > 0$ term from water and a favourable term $-T\Delta S = -100$ kJ/mol, thus overall the $\Delta G \sim -50$ kJ/mol for our exemplary small protein

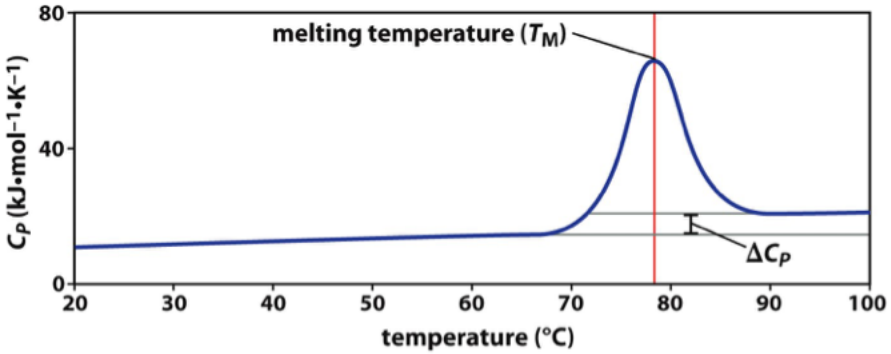
Thermodynamic properties of proteins

Enthalpic and entropic energy contributions for protein folding can be in fact measured by using calorimeters, which measure the heat capacity



Instrument :
Differential Scanning Calorimeter

$C_p = \Delta H / \Delta T$ or $\Delta H = C_p \Delta T$



Data: Melting Curve

Temperature scan of the sample

- Measure heat capacity associated to protein folding
- Determines the melting temperature

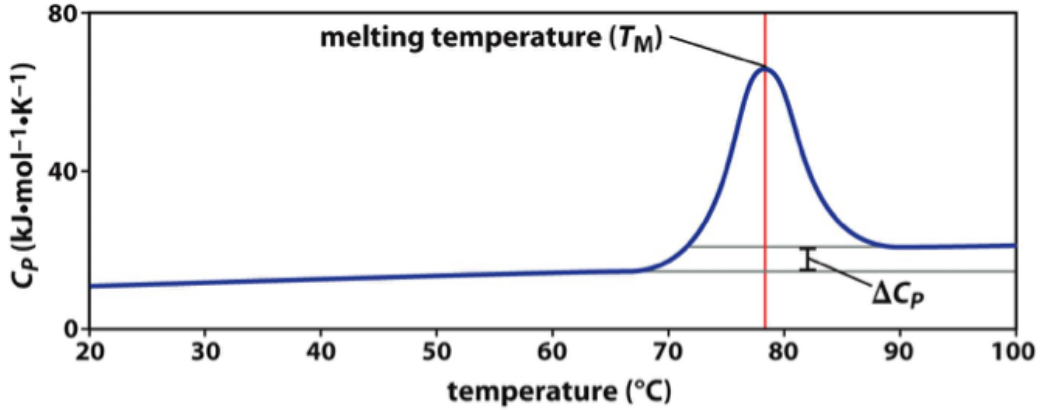
Thermodynamic properties of proteins

with calorimetry we follow the unfolding reaction

$$\mathrm{F} \rightleftharpoons \mathrm{U}$$

$$K_{\text{unfolding}} = \frac{[\mathrm{U}]}{[\mathrm{F}]}$$

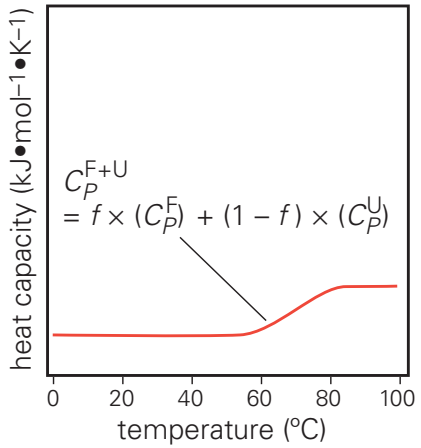
$$\Delta G^{\circ}_{\text{unfolding}} = \Delta H^{\circ}_{\text{unfolding}} - T\Delta S^{\circ}_{\text{unfolding}}$$



As the temperature increases more of the protein unfolds, until at a certain temperature (T_M - melting temperature), the ΔG° becomes zero (as remember that $\Delta G^{\circ} = -RT \ln K_{eq}$)

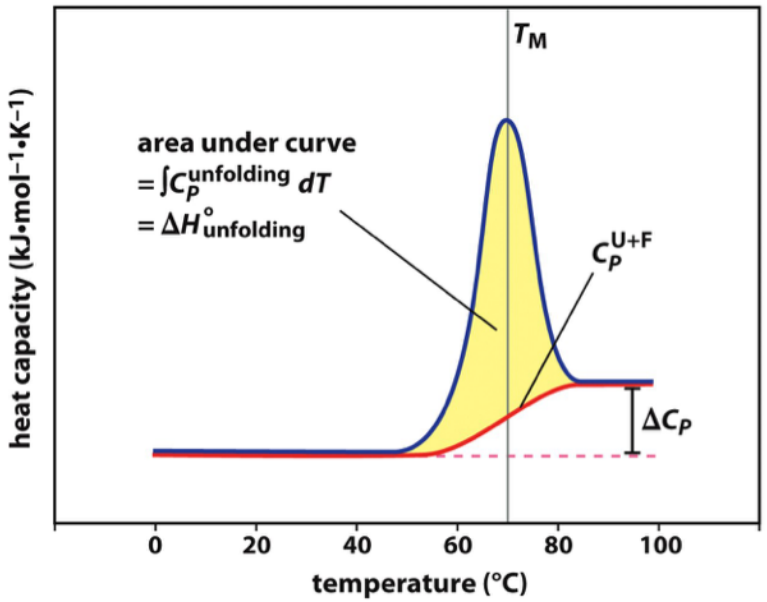
The measured heat capacity includes contributions of both the folded population(f) and the unfolded population(1-f)

$$C_P^{\mathrm{F+U}} = f\, C_P^{\mathrm{F}} + (1-f)\, C_P^{\mathrm{U}}$$



Thermodynamic properties of proteins

But looking at the data something is missing which is the peak reporting the heat capacity due to the interactions that are broken passing from the folded to unfold state



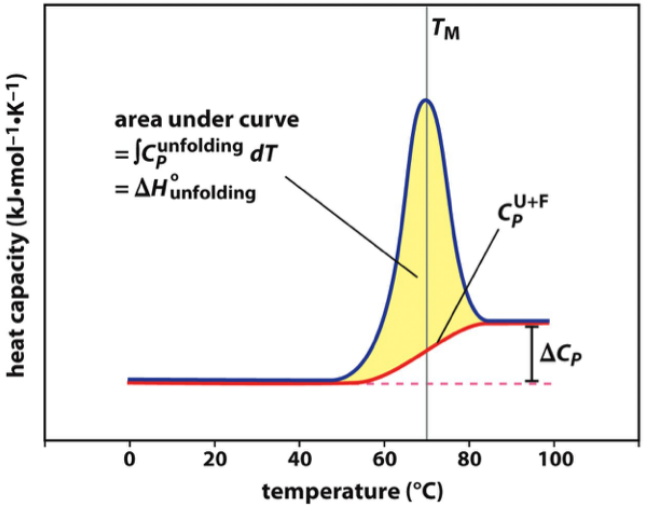
$$C_P^{\text{observed}} = C_P^{\text{F+U}} + C_P^{\text{unfolding}} = f\ C_P^{\text{F}} + (1-f)\ C_P^{\text{U}} + C_P^{\text{unfolding}}$$

Thus to estimate only the energy that accounts for the unfolding process we have to integrate $C_p^{\text{unfolding}}$ obtaining ΔH

$$\int C_P^{\text{unfolding}}\ dT = H_{\text{unfolded}} - H_{\text{folded}} = \Delta H_{\text{unfolding}}$$

Thermodynamic properties of proteins

$$\int C_P^{\text{unfolding}}\, dT = H_{\text{unfolded}} - H_{\text{folded}} = \Delta H_{\text{unfolding}}$$



At melting temperature $K(\text{at } T_M) = \dfrac{[\text{U}]}{[\text{F}]} = 1$ thus $\Delta G^\circ(T_M) = -RT \ln K(T_M) = 0$

if the free energy is zero we can not only estimate the enthalpy contribution to unfolding but also the entropic one - thus at the melting T we can completely characterise the thermodynamics of folding, in fact:

$$\Delta G^\circ_{\text{unfolding}}\ (\text{at } T_M) = \Delta H^\circ_{\text{unfolding}} - T_M \Delta S^\circ_{\text{unfolding}} = 0$$

$$\Rightarrow \Delta S^\circ_{\text{unfolding}} = \frac{\Delta H^\circ_{\text{unfolding}}}{T_M}$$

Thermodynamic properties of proteins

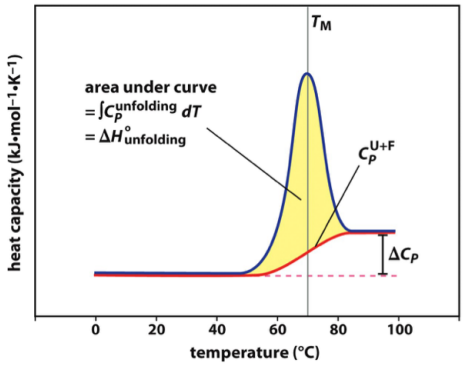
But how to characterise the process
at any given T different from T_M?
For this we can benefit from the fact that

$$\Delta C_P = C_P^{\mathrm{U}} - C_P^{\mathrm{F}}$$

and thus:

$$\Delta H^{\circ}_{\mathrm{unfolding}}(T) = \Delta H^{\circ}_{\mathrm{unfolding}}(T_{\mathrm{M}}) + \Delta C_P (T - T_{\mathrm{M}})$$

$$\Delta S^{\circ}_{\mathrm{unfolding}}(T) = \Delta S^{\circ}_{\mathrm{unfolding}}(T_{\mathrm{M}}) + \Delta C_P \ln\left(\frac{T}{T_{\mathrm{M}}}\right)$$



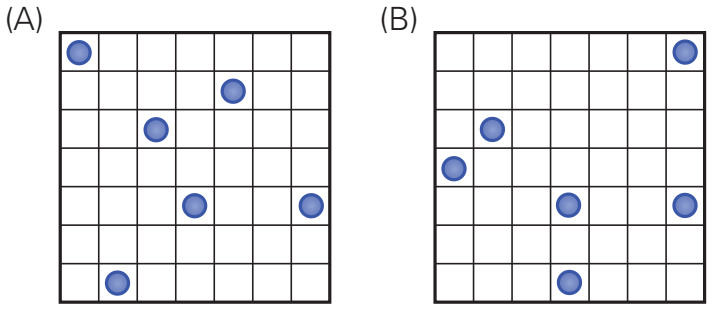
in case ΔC_p were zero, then the equations above would have been independent from the temperature.
$\Delta C_p > 0$ is correlated with the extent of interaction of hydrophobic groups with water

Protein	Molecular weight	ΔH° (25°C) [kJ•mol^{-1} per residue]	ΔS° (110°C) [J•mol^{-1} per residue]	ΔC_P [J•K^{-1}•mol^{-1} per residue]
Protein G-B1	7200	1.4	16.1	53
Parvalbumin	11,500	1.4	16.8	46
Cytochrome *c*	12,400	0.64	17.8	67
Ribonuclease A	13,600	2.4	17.8	44
Hen lysozyme	14,300	2.0	17.6	52
Staph. nuclease	16,800	0.85	17.5	61
Myoglobin	17,900	0.04	17.9	75
Papain	23,400	0.93	17.0	60
β-Papain	23,800	1.3	17.9	58
α-Chymotrypsin	25,200	1.1	18.0	58
Average		1.2 ± 0.7	17.4 ± 0.6	57 ± 9

Connecting macrostates with microstates

Macroscopic measurements like in the case of the unfolding process can be conducted to the microscopic features of the systems. We have to use a statistic approach to describe them and use **entropy** which describe their **randomness**.

This is related to the **multiplicity** of the system. The system is characterized by a **state** with some global properties like temperature, pressure, and number of molecules. A **microstate** is a specific configuration of molecules that is consistent with the state. Each state corresponds to many different microstates.

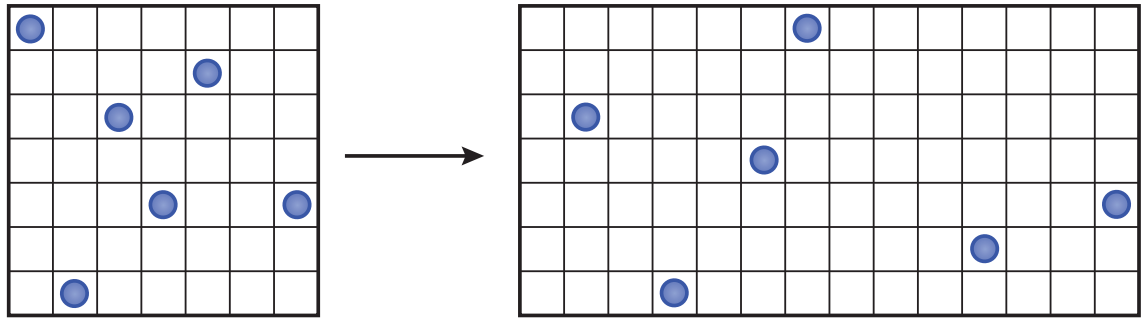


microstate A microstate B

M potential configurations
N molecules

$$W(M,N) = \frac{M!}{N!(M-N)!}$$

Connecting macrostates with microstates



$M = 49$ $\qquad$ $M = 98$

$$W(49,6) = \frac{49!}{6!(49-6)!} = \frac{6.08\times10^{62}}{720\times6.04\times10^{52}} \approx 1.4\times10^{7}$$

using Stirling's approx $\ln n! \approx n\ln n - n$

$$\ln W(98,6) = 98\times\ln(98) - 98 - \ln(720) - (92\times\ln 92 - 92)$$

$$\Rightarrow W(98,6) = e^{20.72} = (10^{0.434})^{20.72} \approx 10^{9}$$

a system will tend to states of increased multiplicity or with maximal number of microstates.

Since *W* is a very high number for molecular systems, it is handier to use *lnW*, which has the nice property to be **extensive**

$$W_{A+B} = W_A \times W_B$$

$$\Rightarrow \ln W_{A+B} = \ln W_A + \ln W_B$$

Entropy (statistical definition)

$$S = k_B \ln W$$

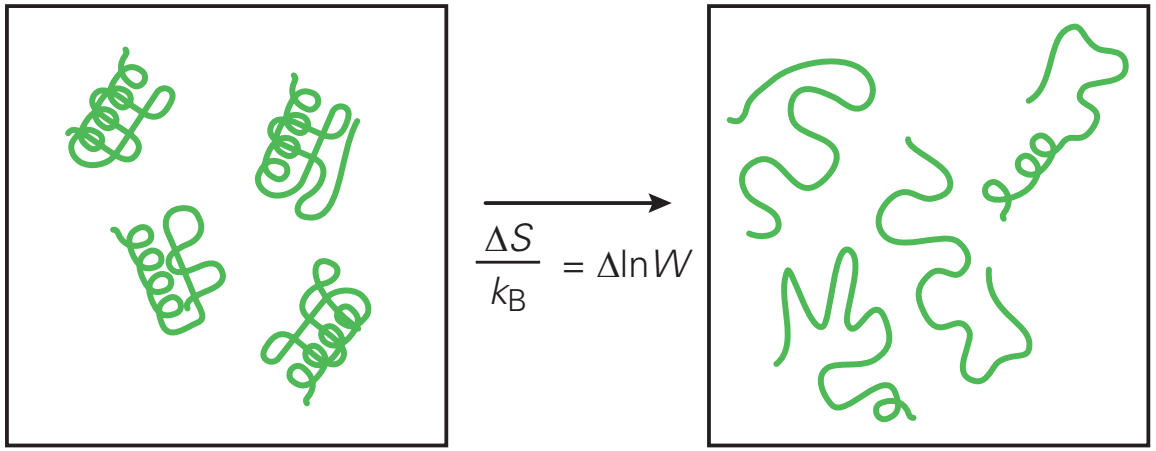
which is therefore extensive and a state function, and has the unit of an energy/temperature, J/K, like the Boltzmann constant.
This is equivalent to the **thermodynamic definition** of S:

$$\Delta S = \frac{q_{rev}}{T}$$

that you have seen derived from the study of heat engines (see page 330 Chapter 7 for a demonstration for ideal gas).
Thus spontaneous processes will increase entropy and at equilibrium S will be maximal

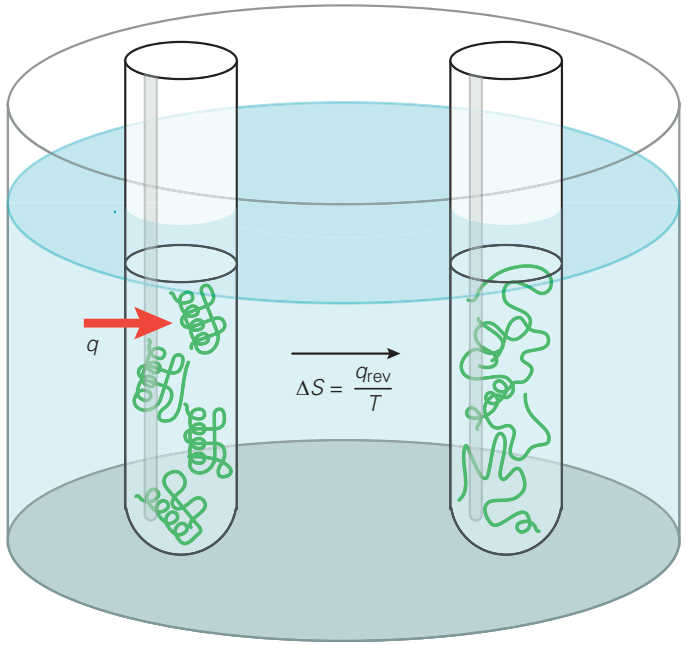
2nd law of thermodynamics - maximal entropy principle

statistical definition of S



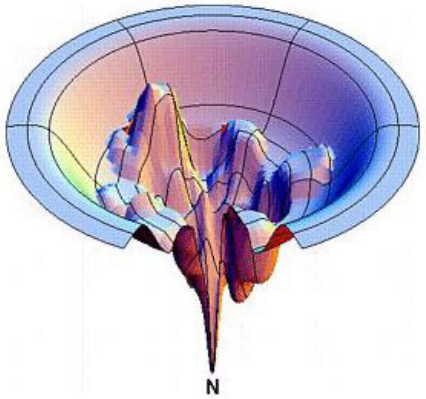
W = number of conformations or configurations

thermodynamic definition of S



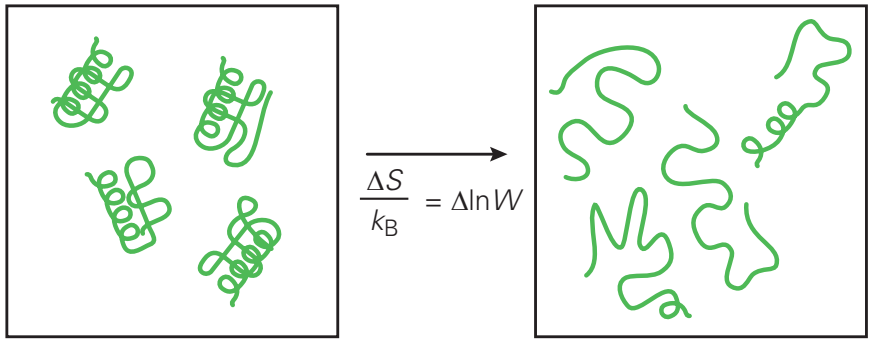
W is difficult to be estimated, but heat can be measured and this equivalence provides a way to reconnect *S* to the molecular features of the system - the link is T, temperature

Boltzmann Distribution



At low T → they all roll into the lowest valley (energy U dominates)

At high T → they spread out, exploring higher hills (entropy S dominates)



W = number of conformations or configurations

Temperature measures how much they care about exploring versus settling

$$P(U) \propto e^{-U/k_B T}$$

mathematical way of saying low energy is good, but higher energy can still happen — less often, by an amount that depends on temperature

consider our folding problem

$$\mathrm{U} \rightleftharpoons \mathrm{F}$$

$$K_{\text{folding}} = \frac{[\mathrm{F}]}{[\mathrm{U}]} \text{ (at equilibrium)} = \frac{P_F}{P_U} = e^{-(E_F - E_U)/k_B T}$$

that's the microscopic explanation of $K = e^{-\Delta G^{\circ}/RT}$

Boltzmann Distribution and Energetic Levels

If you have a system with N molecules and total energy U, when N is large (~N_A), it is difficult to know how the energy is distributed through N atoms. Thus we can only describe in statistical terms the population of a state, i.e. the N_i – number of molecules that will be found in an energy level with energy U_i.

$$P(N_i) = \frac{N_i}{N} = \frac{e^{-U_i/k_B T}}{\sum_i e^{-U_i/k_B T}}$$

thus $$N_i = \frac{N e^{-U_i/k_B T}}{Q}$$

where Q is the **partition function**

$$Q = \sum_i e^{-U_i/k_B T}$$

and k_B is the **Boltzmann constant** (k_B = 1.381 x 10^{-23} J/K)

but Q is constant at a given temperature, thus $$N_i \propto e^{-U_i/k_B T}$$

thus you can estimate the ratios between different populations at different energy levels using the following relation:

$$\frac{N_2}{N_1} = e^{-\Delta U/k_B T}$$

$$\Delta U = U_2 - U_1$$

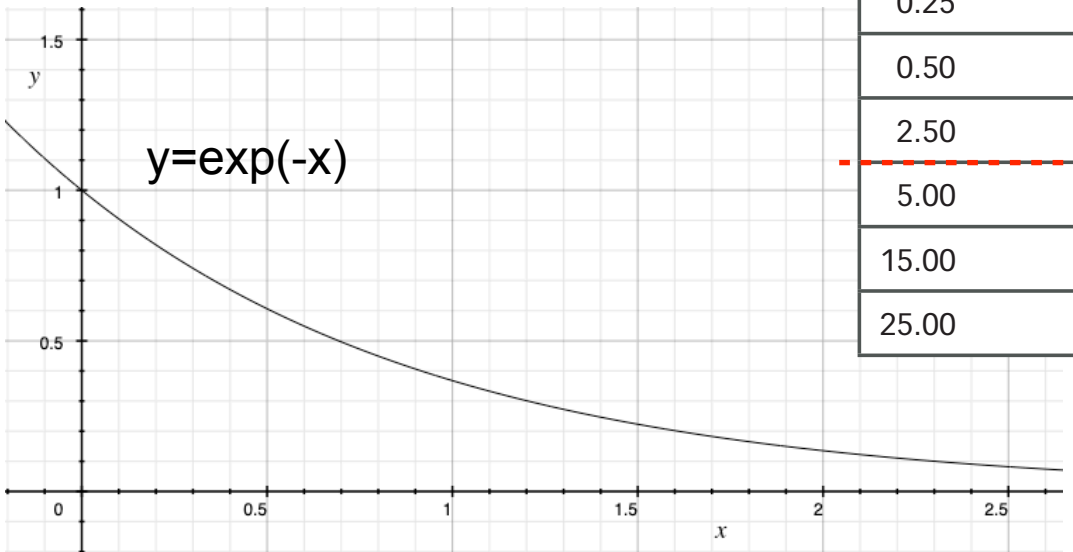
- Remember that the gas constant R is the "molar" form of k_B, in fact:
- $R = N_A k_B$ = 8.3145 J/K*mol, thus if you work with KJ/mol you have to use RT in the Boltzmann distribution

Boltzmann Distribution and Energetic Levels

we can simply access the ratios between populations of different energy levels using the following relation:

$$\frac{N_2}{N_1} = e^{-\Delta U / k_B T} \qquad \Delta U = U_2 - U_1 \qquad \frac{N_2}{N_1} = e^{-\Delta U / 2.529}$$

using as unit kJ/mol



ΔU (kJ•mol^{-1})	$\frac{\Delta U}{k_B T}$ (T = 300 K, $k_B T \approx$ 2.5 kJ•mol^{-1})	$e^{\frac{-\Delta U}{k_B T}}$
0.25	0.1	0.90
0.50	0.5	0.61
2.50	1.0	0.37
5.00	2.0	0.13
15.00	6.0	0.00067
25.00	10.0	0.0000045

$k_B T$

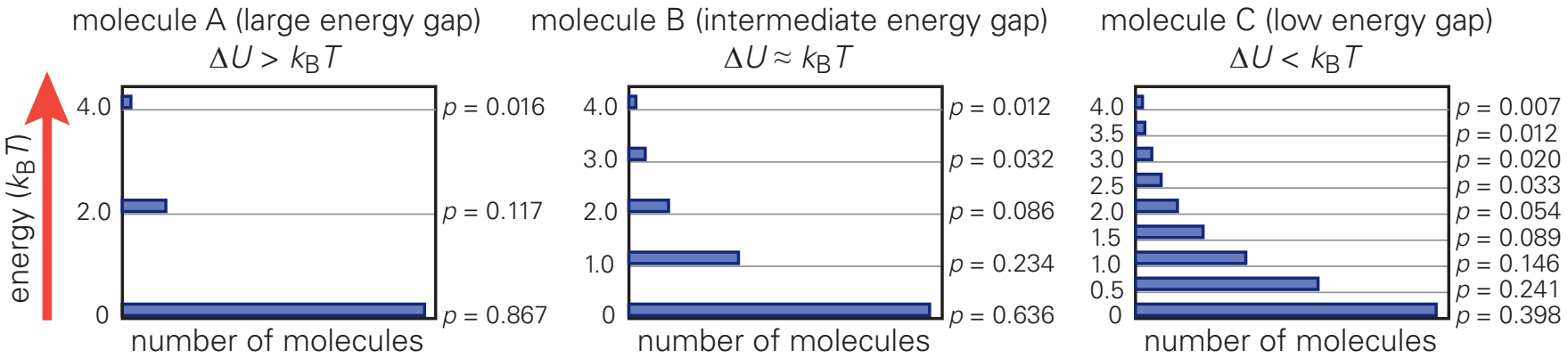
Boltzmann Distribution and Energetic Levels

we can simply access the ratios between populations of different energy levels using the following relation:

$$\frac{N_2}{N_1} = e^{-\Delta U / k_B T} \qquad \Delta U = U_2 - U_1$$

$$\frac{N_2}{N_1} = e^{-\Delta U / 2.529}$$

using as unit kJ/mol



3 molecules with different accessible energy levels - levels that are less spaced (< kT) are more accessible

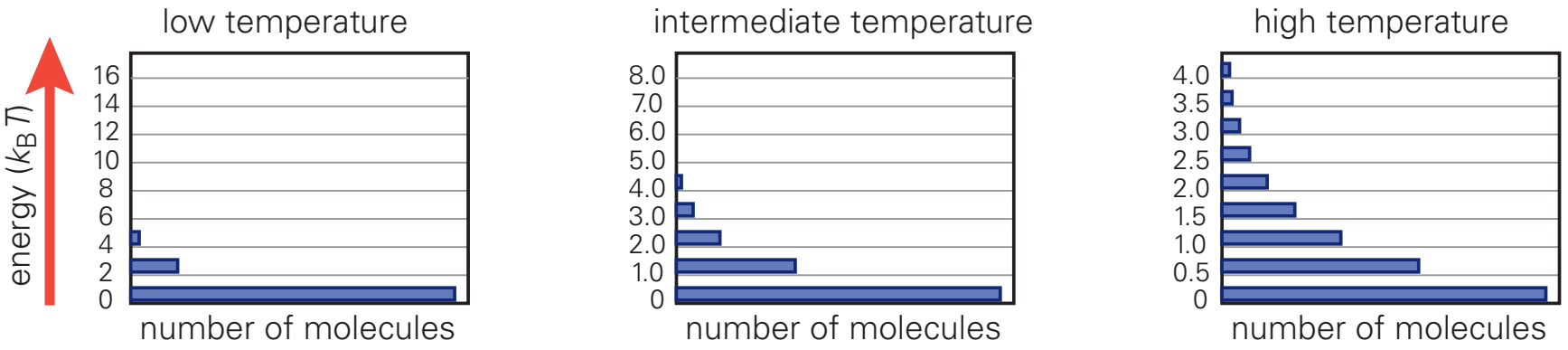
Boltzmann Distribution and Energetic Levels

we can simply access the ratios between populations of different energy levels using the following relation:

$$\frac{N_2}{N_1} = e^{-\Delta U / k_B T} \qquad \Delta U = U_2 - U_1$$

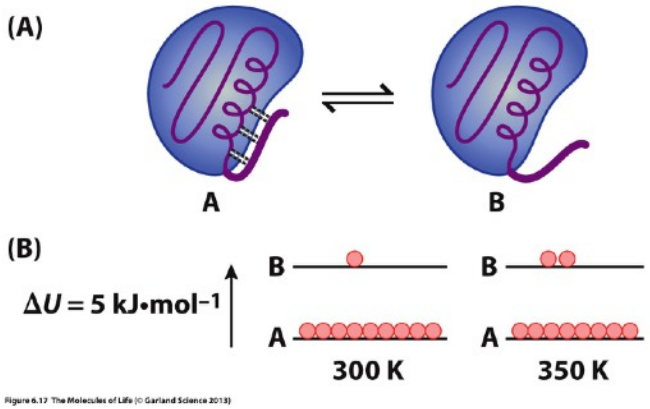
$$\frac{N_2}{N_1} = e^{-\Delta U / 2.529}$$

using as unit kJ/mol



same molecule at different T, the occupancy of energy levels increases with T

Boltzmann Distribution in macromolecules



Protein molecules take up energy as they unfold

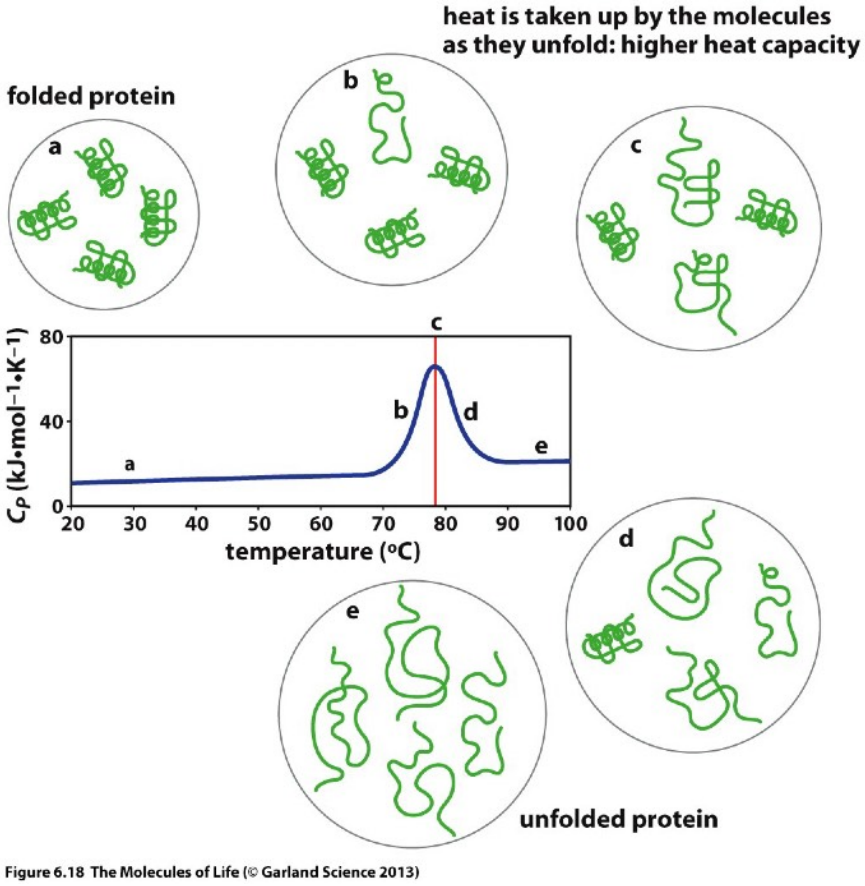
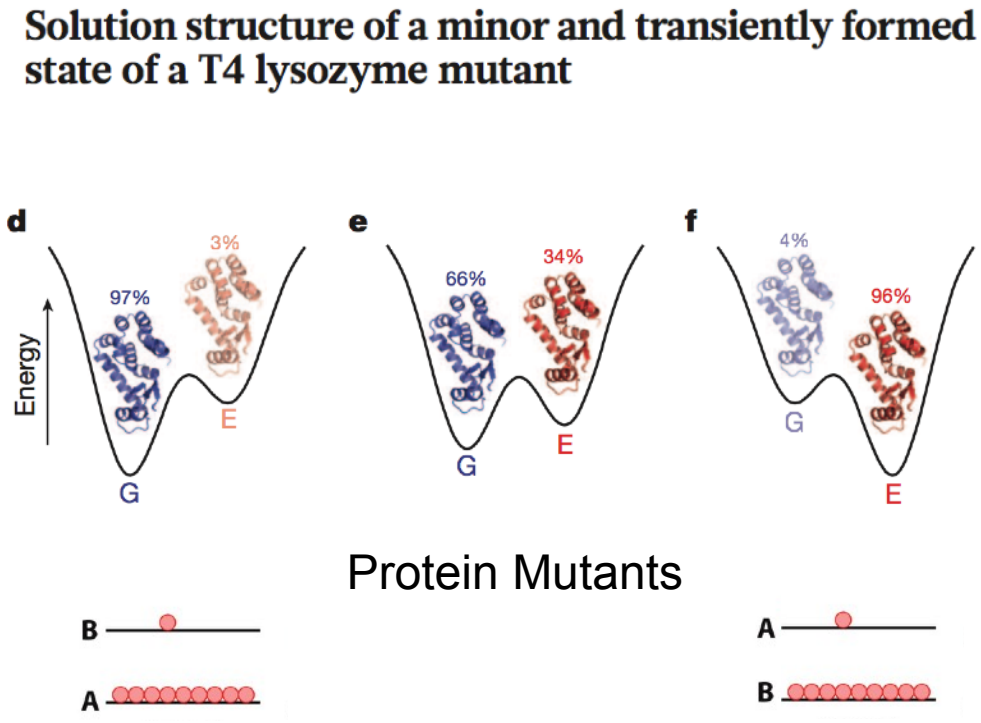


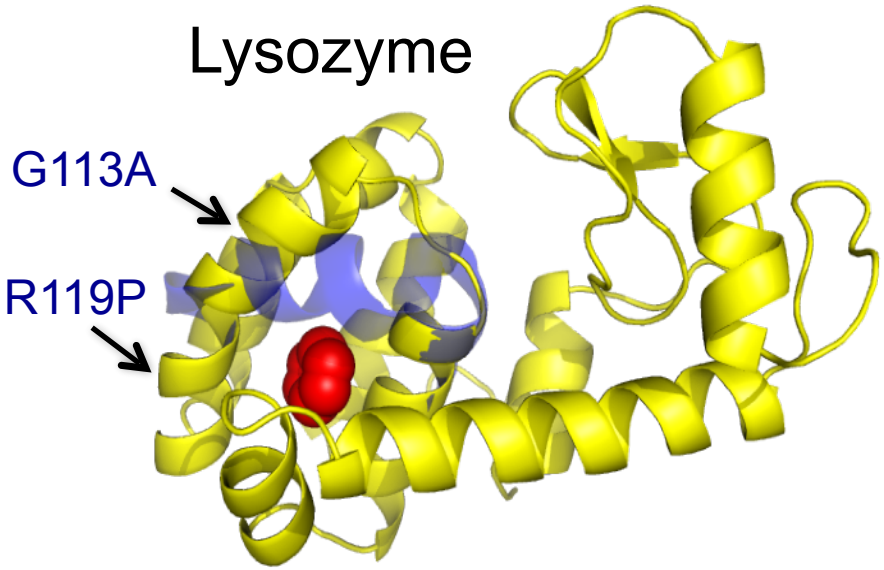
Figure 6.18 The Molecules of Life (© Garland Science 2013)

Shifting the distribution of populations with temperature

One can see how the formalism of the Boltzmann distribution helps us to describe what occurs in proteins and other biomolecules.

Engineering conformational changes





Ground State (xtal-3dmv)
Excited State
Benzene
Excited State Stabilizing Mutations

Proteins can co-exist in multiple conformational states (e.g. ground and excited)
With rational mutations one can shift the occupation of these states

Bouvignies, G., Vallurupalli, P., Hansen, D. *et al.* Solution structure of a minor and transiently formed state of a T4 lysozyme mutant. *Nature* **477**, 111–114 (2011). https://doi.org/10.1038/nature10349