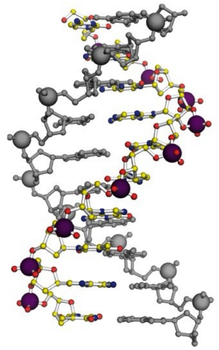
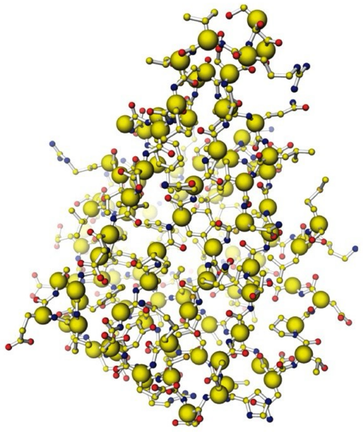
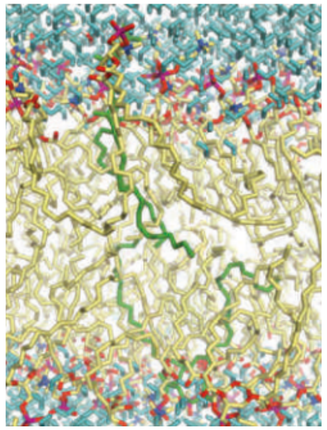


Chimie Biologique I
Biological Chemistry I
BIO-212

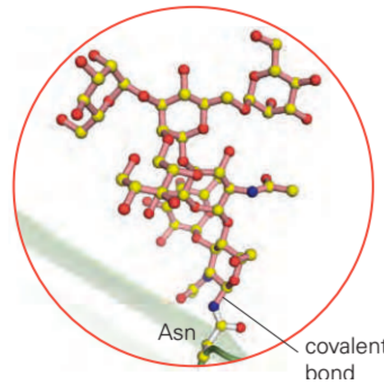


Welcome !!!!!!

Matteo Dal Peraro, IBI-SV
Aleksandar Antanasijevic, GHI-SV



EPFL



About me …

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

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

- postdoc at UPenn Chemistry, Philadelphia USA

- associate professor at SV, Institute of Bioengineering (IBI)

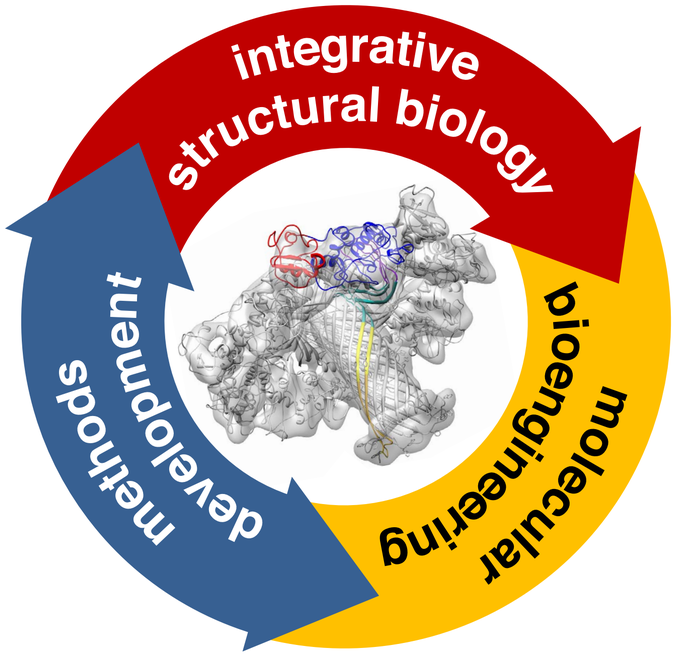
- associate director of the Institute of Bioengineering (IBI)
- office AAB 048 - matteo.dalperaro@epfl.ch

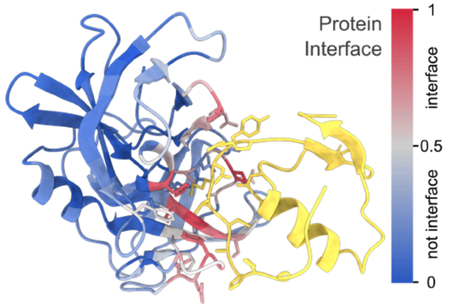
BioEngineering

EPFL

About my lab …

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

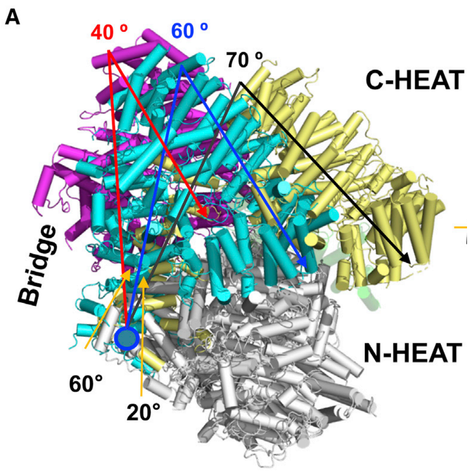




- **deep learning for molecular modeling and design**

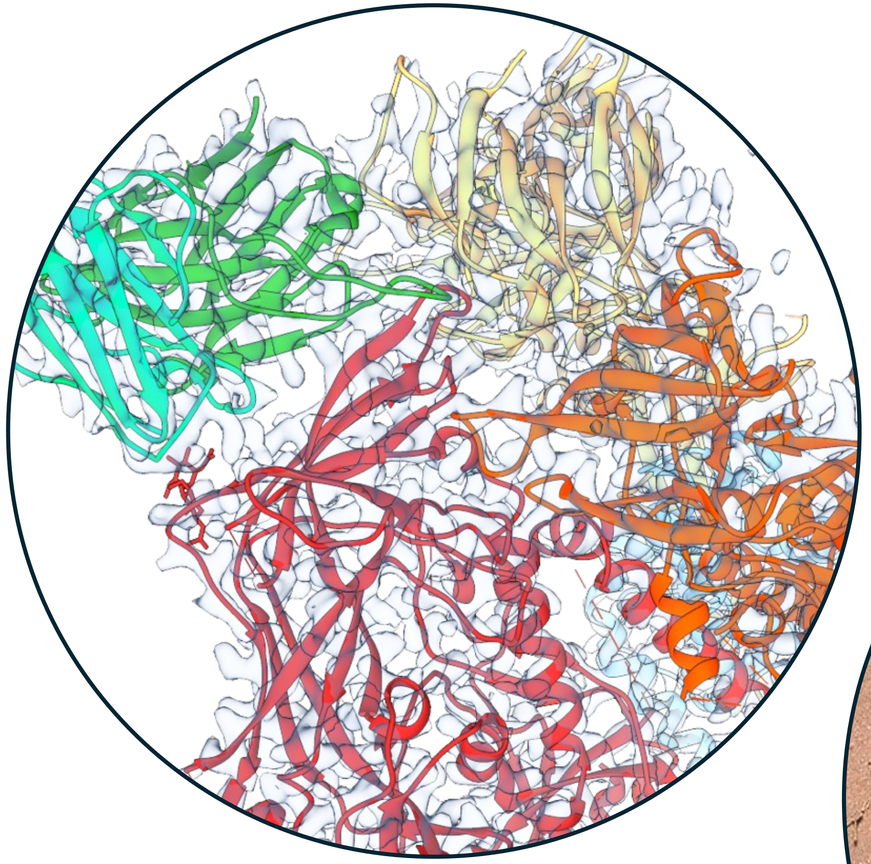
http//pesto.epfl.ch

- **large molecular assembly**



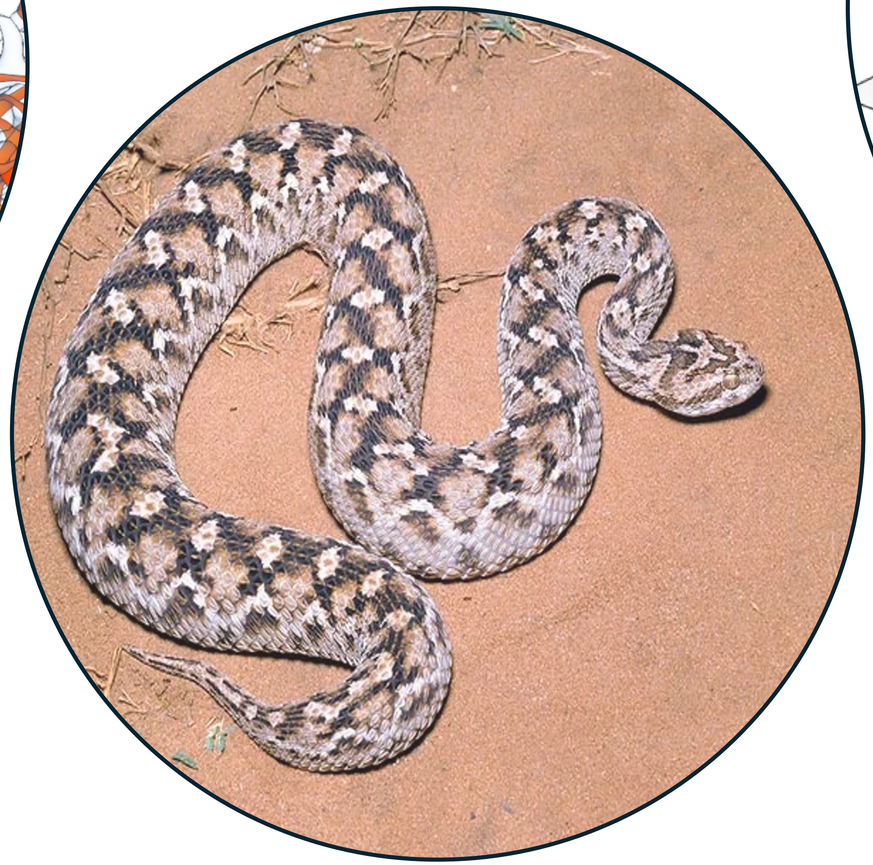
- **biological nanopore sensing**

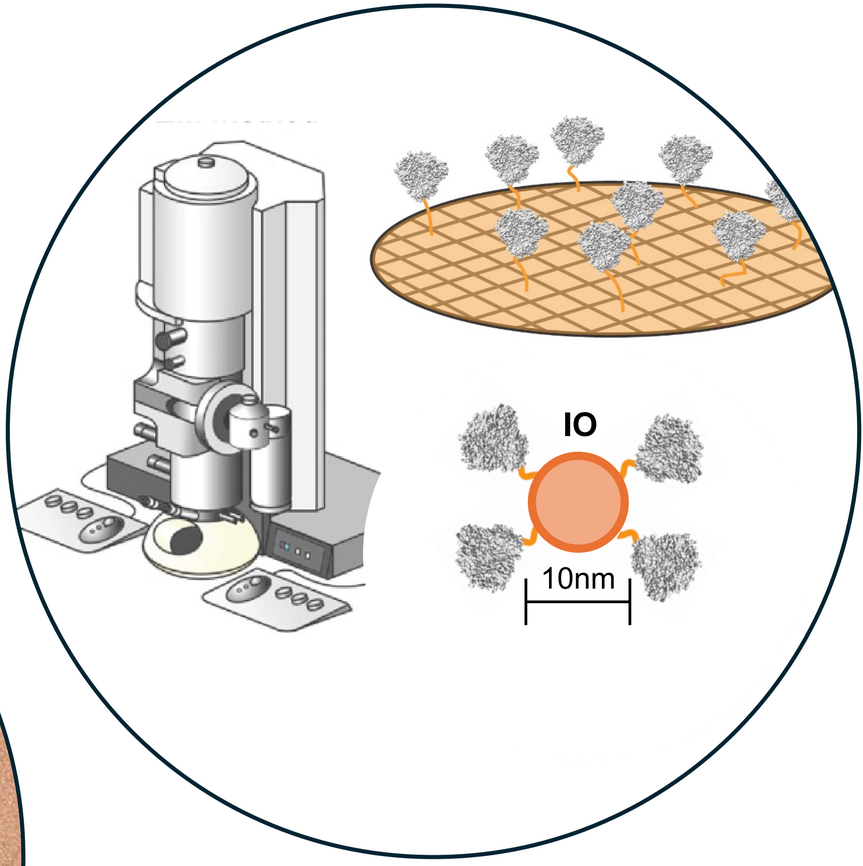
Laboratory of Virology and Structural Immunology (Aleksandar Antanasijevic)



Structural Antivenomics

- Venom – antivenom interactions
- Novel antivenom formulations





Viruses and Antibodies

- Structural analyses of antibodies
- Immunogen design and evaluation

Technology development

- Custom EM grids and nanobeads
- Tools for studying mucosal Abs

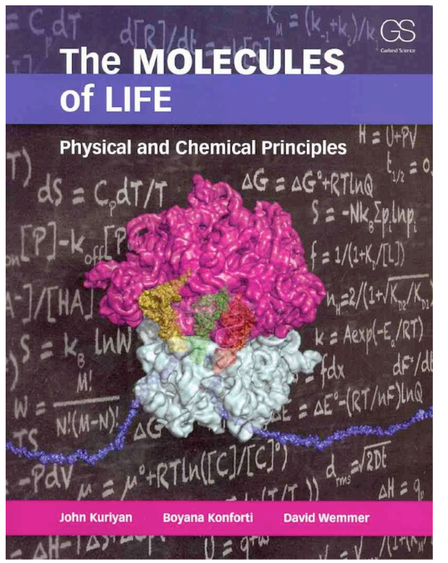
Overview

- Material (slides, exercises, Q&A forum …) on Moodle

- Organization of exercises: ~10 TAs:
  - Jana, Fernando, Evgenia, Edoardo, Arthur, Alissa, Evgenia, Camilla, Kiruthika, Yangyang, Yash

- Exam: 3h written exam, questions will be provided in both English and French and you may answer in either language, you can bring anything, mix of QCM and open exercises

- Textbook (but not everything will be there): Kuriyan et al. ‘The Molecules of Life’



Vision and Rules

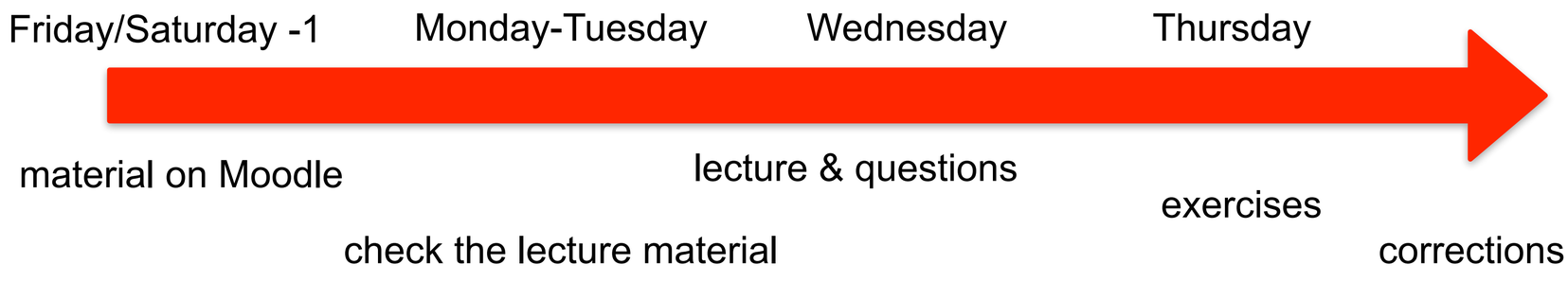
- build an understanding of biology from the molecular level
- understanding of the energetic principles that govern molecular interactions in biomolecules
- leverage protein structures to understand biological function
- learn about experimental and computational methods to analyze proteins

- You are encouraged to ask questions
- Take notes
- Read the book if you can - expand what we discuss in class
- Attend lecture regularly
- Attend exercises regularly
- Provide early feedback - week 4/5 feedback questionnaire (provided by EPFL)

Exercise Session Guidelines

- 4 ECTS = 2h exercise session (8-10 am Thursdays morning)
- Give a look ahead of time if you can
- pair up with classmates and discuss
- discuss with the TAs - do not be shy, they are there to help
- work on series before checking the solutions

Week 1 – 11/9	MDP	Intro - Biomolecular interactions - Nucleic acids
Week 2 – 18/9	AA	Lipids and Carbohydrates
Week 3 – 25/9	AA	Amino-Acids and Proteins
Week 4 – 2/10	MDP	Thermodynamics and Energetics
Week 5 – 9/10	AA	Macromolecular Structure and Function
Week 6 – 16/10	MDP	Protein Structure Prediction and Protein Design
Week 7 – 30/10	MDP+TAs	Practical
Week 8 – 6/11	AA	Production and Purification of Biomolecules
Week 9 – 13/11	AA	Introduction to Biophysical Methods
Week 10 – 20/11	MDP	Biomolecule Assembly - Entropy, Free Energy, Folding
Week 11 – 27/11	MDP	Recognition and Binding
Week 12 – 4/12	AA	Measuring Biomolecular Interactions
Week 13 – 11/12	AA	Kinetics and Catalysis
Week 14 – 18/12	MDP-AA	Biomolecular Engineering



Common Problems in this class

- ”I don’t understand the subject, too difficult”
- “The pace is too fast”
- “The slides are not very helpful”
- ”We have heard this subject in other courses”
- “The lectures and the exercises are disconnected”
- “The exercise sessions are too short”
- “The lectures are not well structured”
- “ I do not know what to learn and remember (for the exam)”

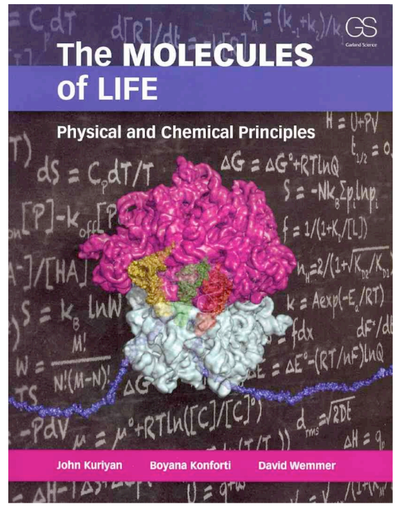
Lecture 1 - Outline

Today:

- The molecules of life
- Energetic principles of molecular interactions
- The building blocks:
  - Nucleic Acids

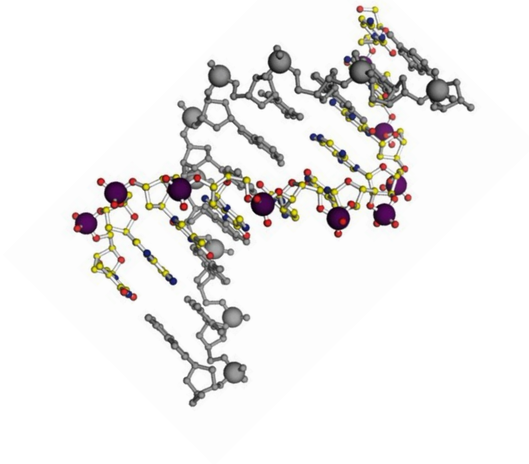
Reading suggestions:

-The Molecules of Life (Chapters 1-2)



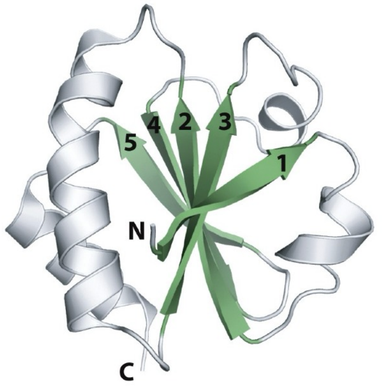
The essentials of life

Heritable information



- DNA
- RNA

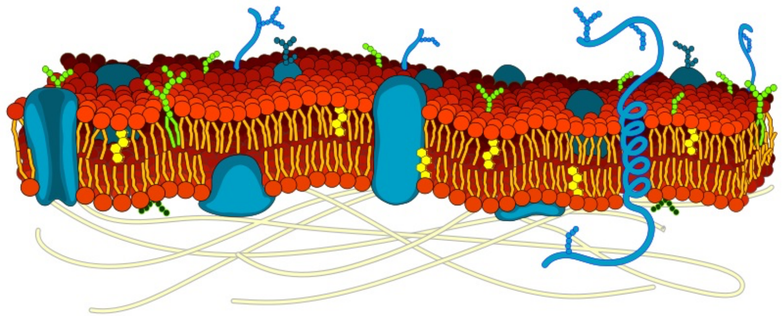
Molecular machinery



- Proteins
- Metabolites

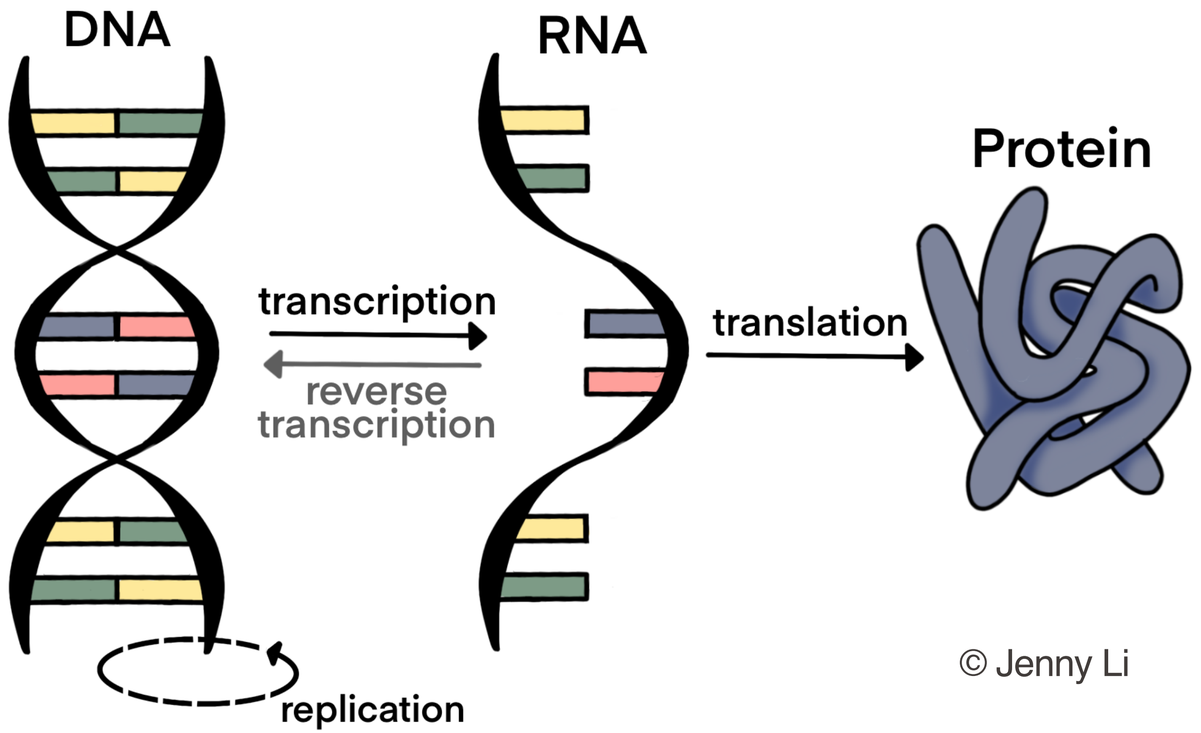
Boundary

- Lipids
- Cell wall



inspired from M. Hecht

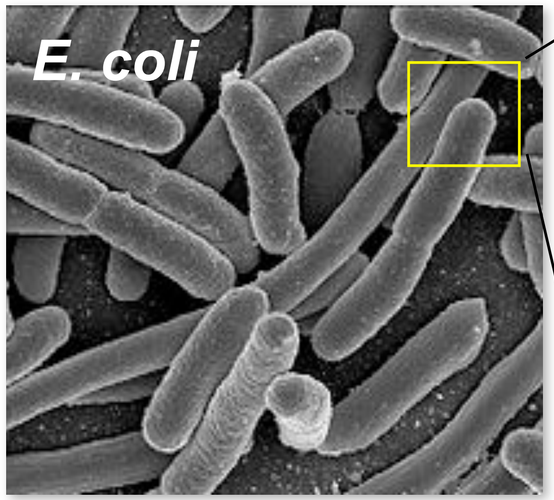
The central dogma of molecular biology

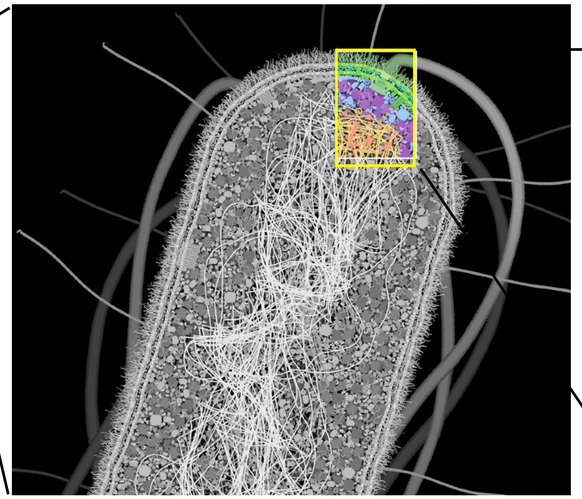


- Genetic information is used to build molecules
- Proteins are the workhorse of the cell

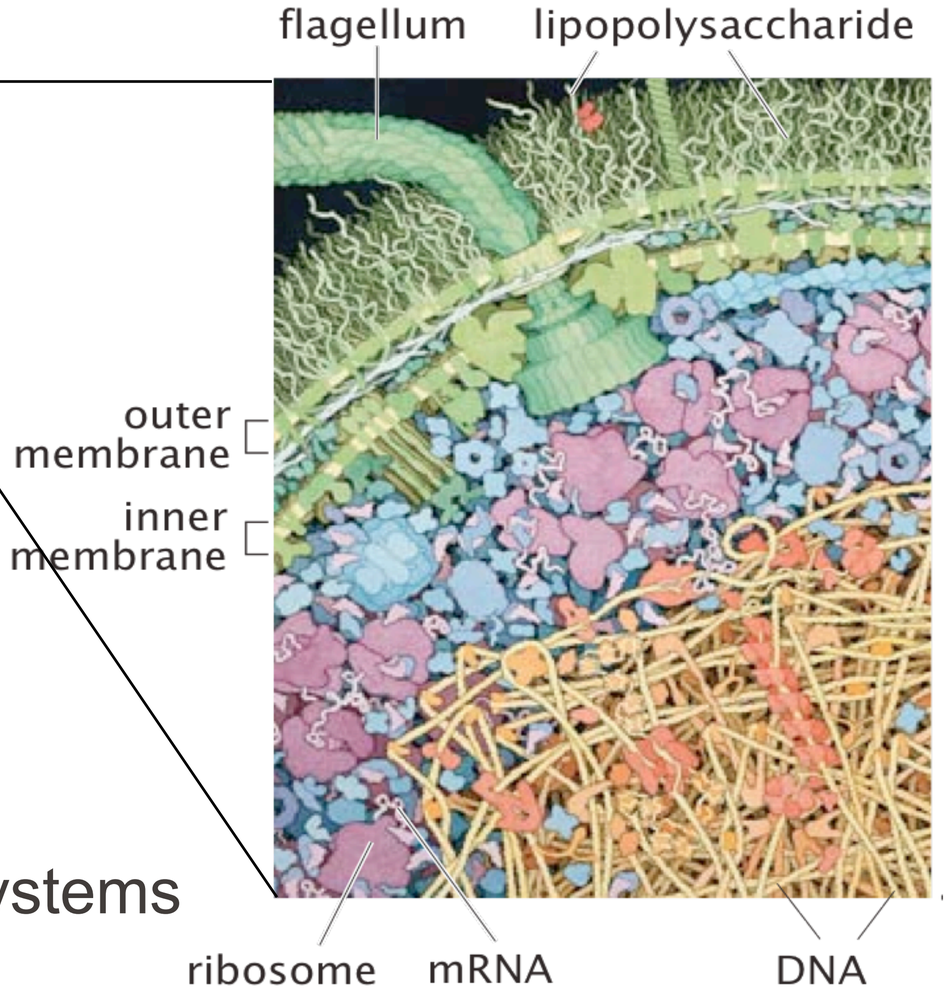
Matteo Dal Peraro

Cells: the minimal unit of life





$V_{E.coli} \approx 1\ \mu m^3 = 1\ fL$

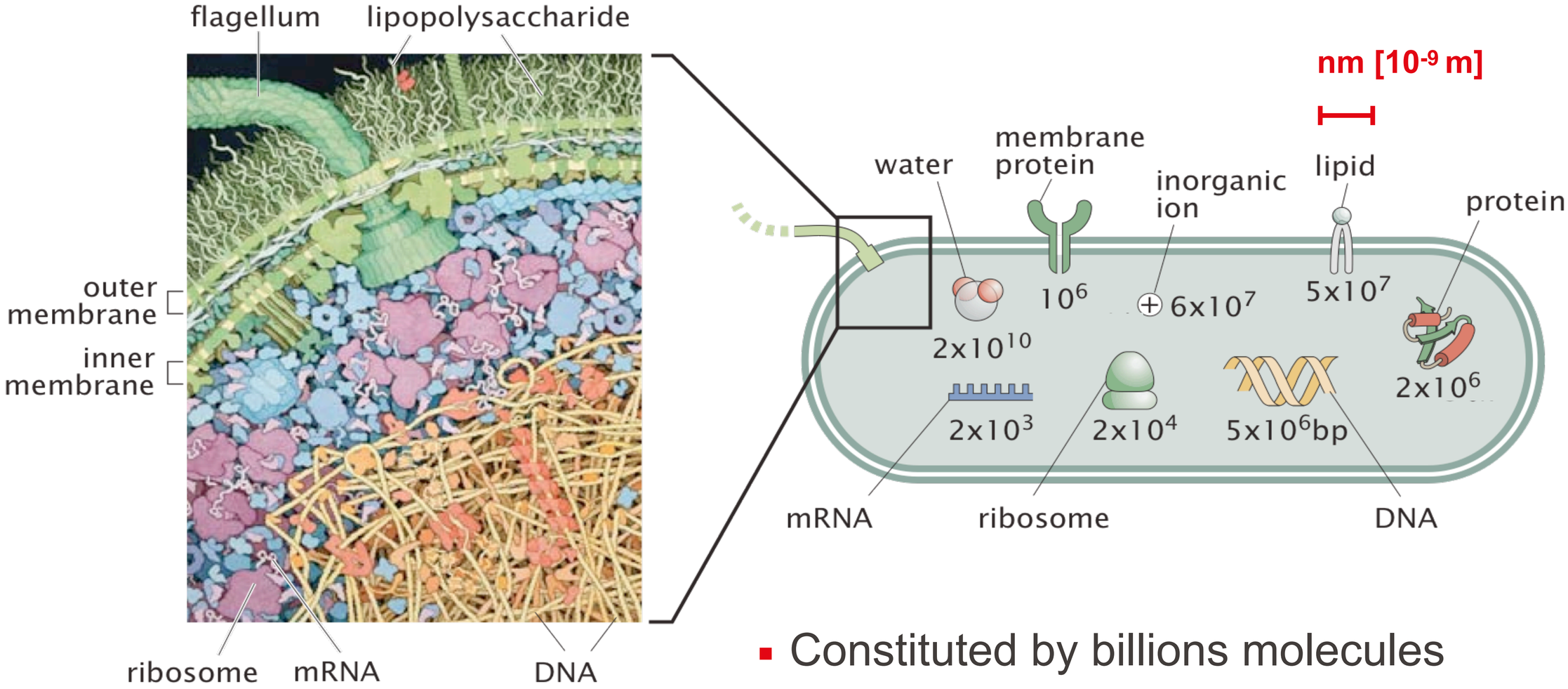


- Cells are small, complex and crowded systems
- Molecular self-sustained factories

from *Physical Biology of the Cell*

Matteo Dal Peraro

Census of an *E. coli* cell



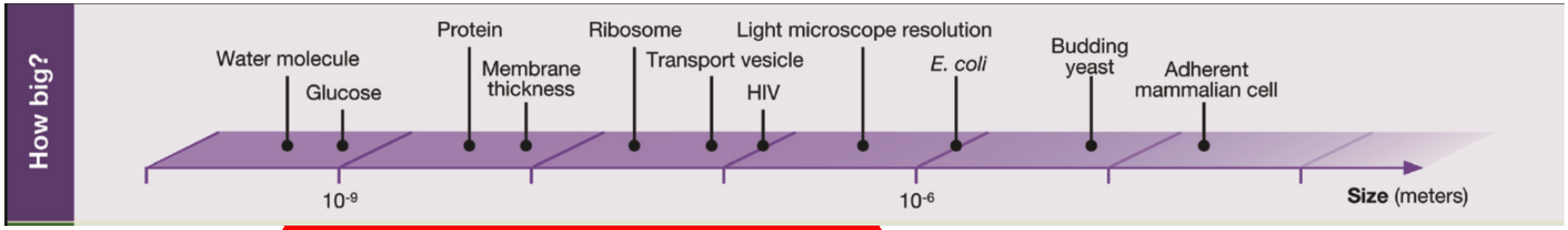
from *Physical Biology of the Cell*

- Constituted by billions molecules
- They regulate all the cellular functions
- They are governed by the laws of physics

Numbers of Life

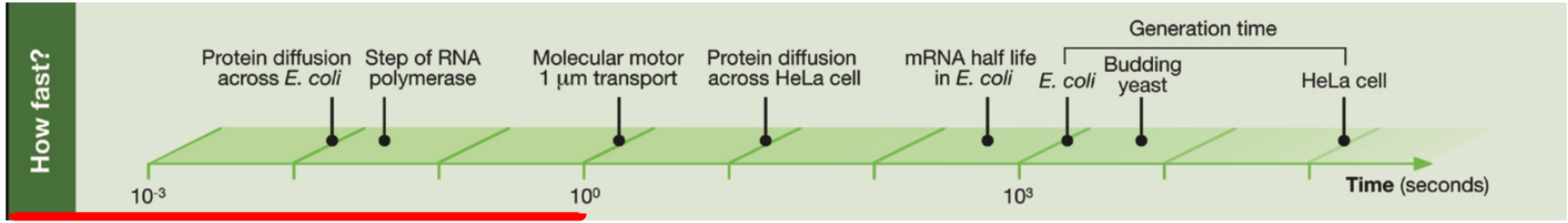
these are the relevant scales important when we talk about biomolecules

size



Å nm

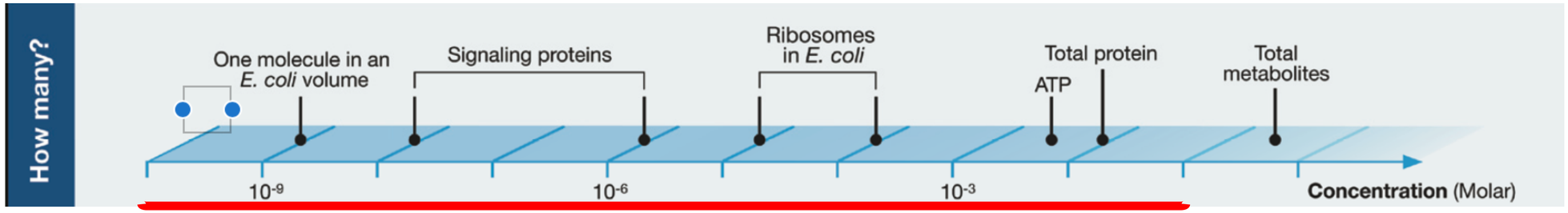
time



ns — ms

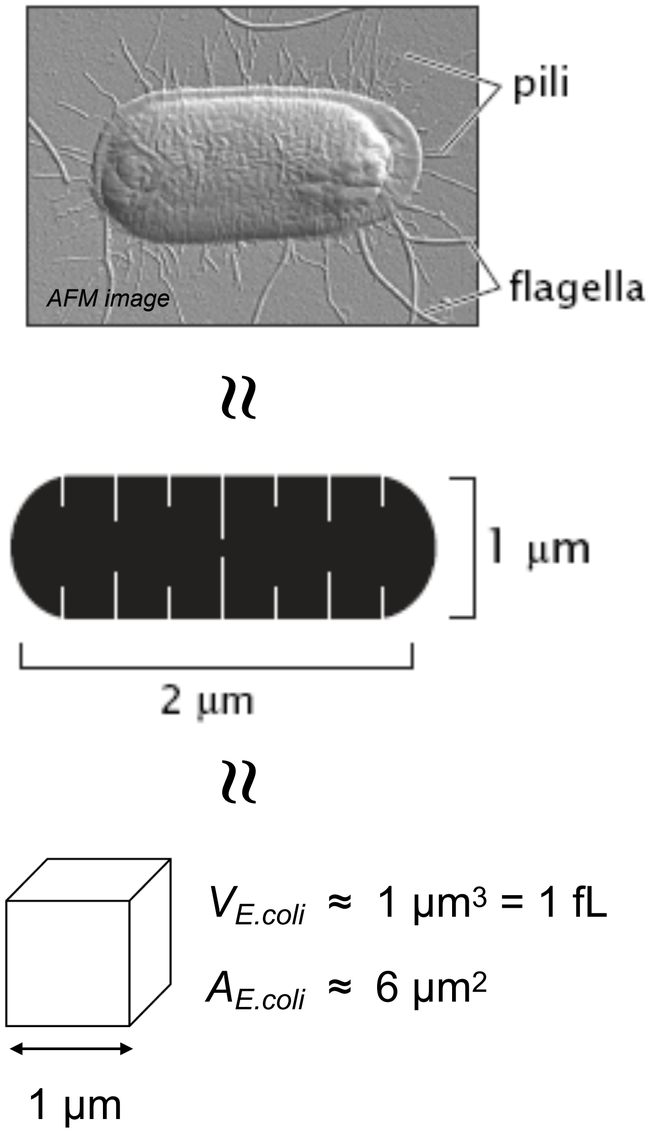
protein conformational changes

concentration



E.coli as molecular ruler

- it is a good representative of biological cells (e.g. DNA-based genome, transcription machinery, lipid bilayer membranes)
- cell size and molecular population can be used as a biological ruler (≈ 1 μm^3)
- cellular equilibrium and dynamics depends on the concentration inside the cell (*in vivo* conditions)



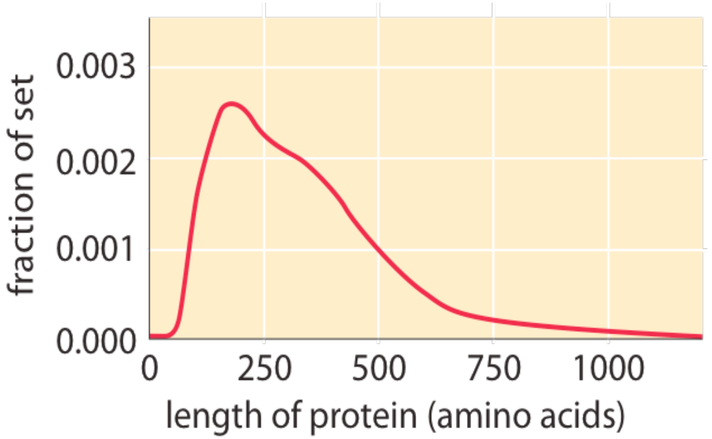
Counting up *E.coli*

Estimate the number of proteins: $N_{protein}$ = [$m_{total\ protein}$ / $m_{per\ protein}$]

$V_{E.coli}$ ≈ 1fL, assume $\rho_{E.coli} \approx \rho_{H20}$ =1 g/mL => $m_{E.coli}$ ≈ 1 pg

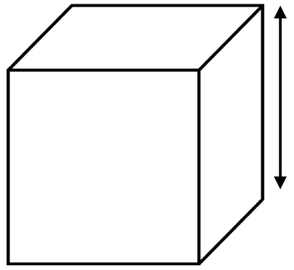
- experimentally is known that dry weight is 30% total weight, and proteins take up to 50% of dry weight, thus $m_{total\ protein}$ ≈ 0.15 pg

- average protein is ~30 kDa (300 AA, m_{AA}≈100 Da >> $m_{per\ protein}$~ 30 kDa); being a Da ≈ 1.66×10^{-24} g we obtain that $m_{per\ protein}$= 5×10^{-20} g, thus

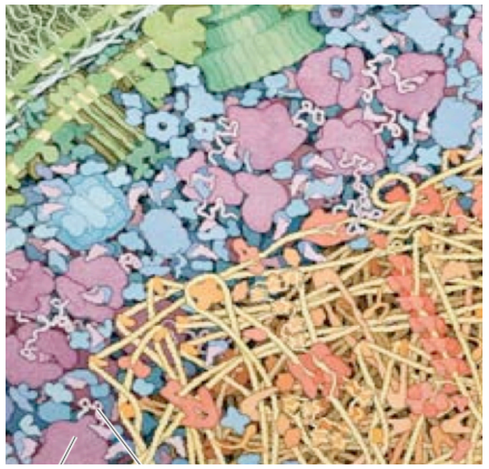


$N_{protein}$ = $m_{total\ protein}$ / $m_{per\ protein}$ ≈ (15×10^{-14} g) / (5×10^{-20} g) ≈ **3×10^{6}**

of which 1/3 are typically membrane proteins **$N_{cytoplasmic} \approx 2\times10^{6}$, $N_{membrane} \approx 10^{6}$**

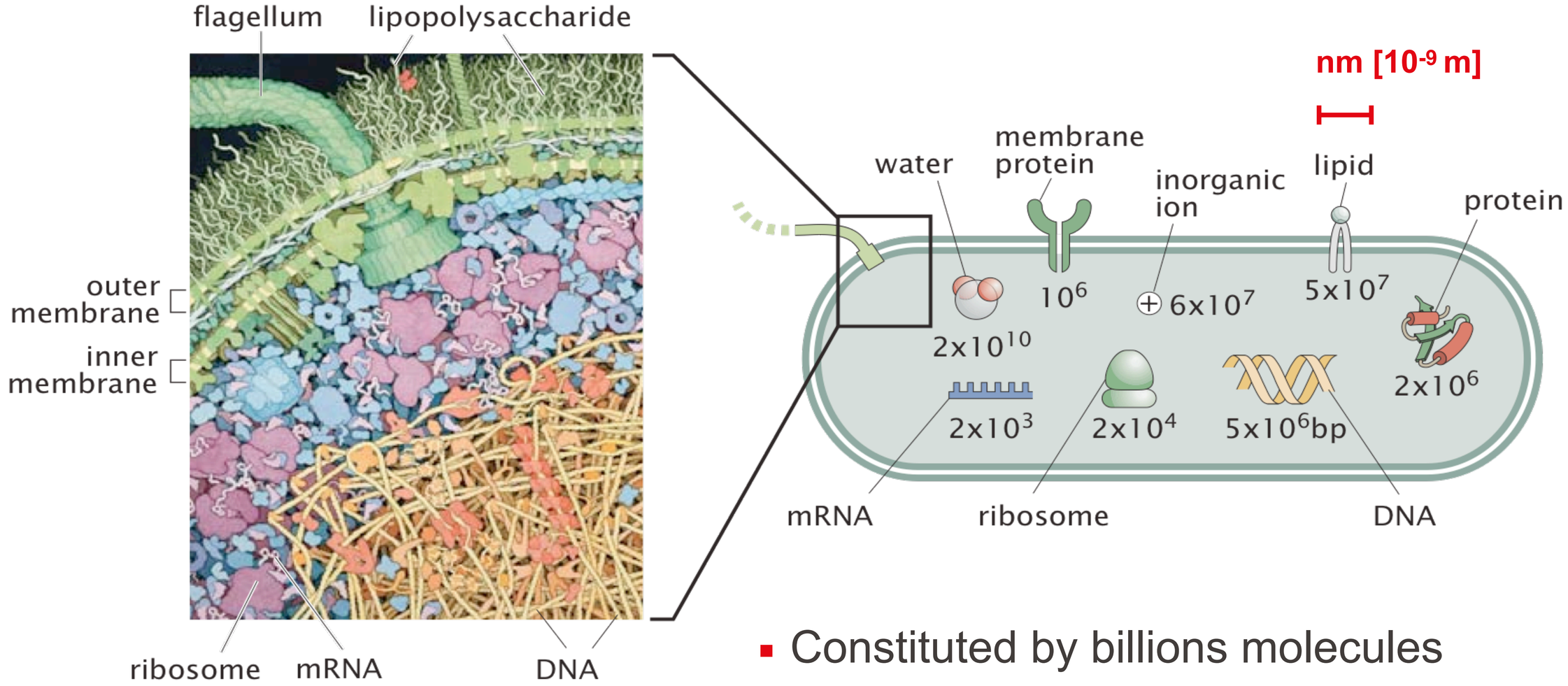


in 1 μm ~ 100 proteins each with 10 nm linear space, given ~2 nm of radius per protein, the space between 2 protein is ~ 5 nm



Matteo Dal Peraro

Census of an *E. coli* cell

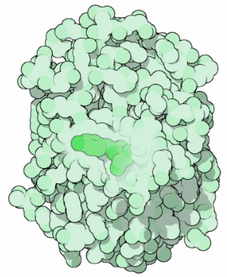


from *Physical Biology of the Cell*

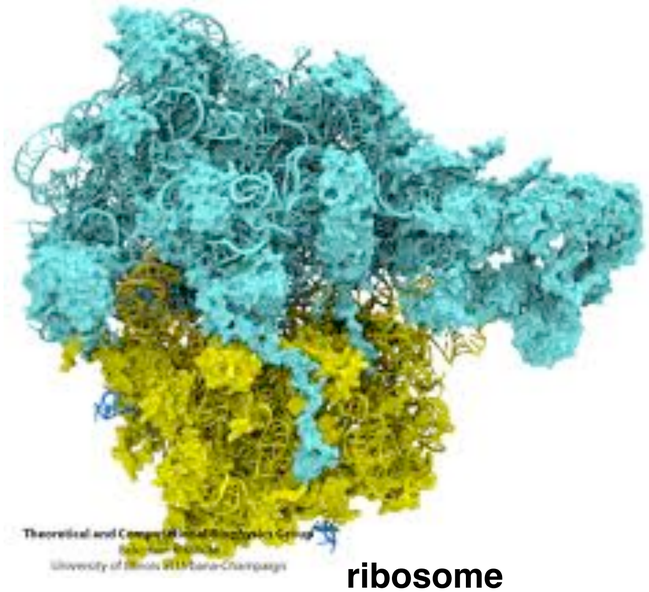
- Constituted by billions molecules
- They regulate all the cellular functions
- They are governed by the laws of physics

Biomolecular structures

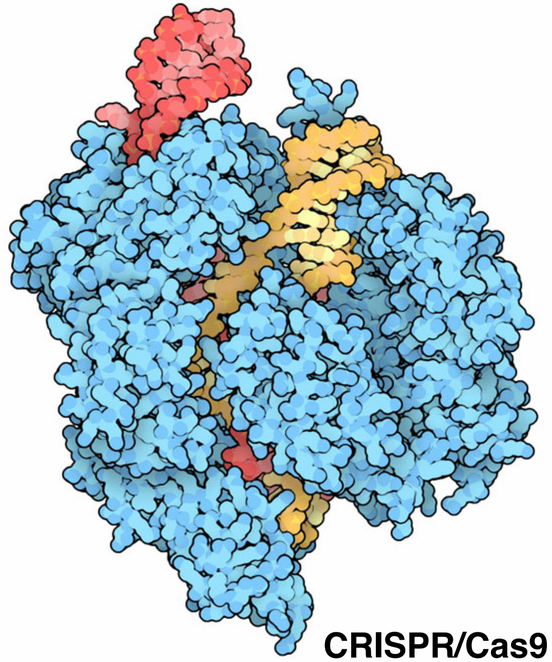
myoglobin 1959
(Nobel in Chemistry 1962)



GFP
(Nobel in Chemistry 2008)



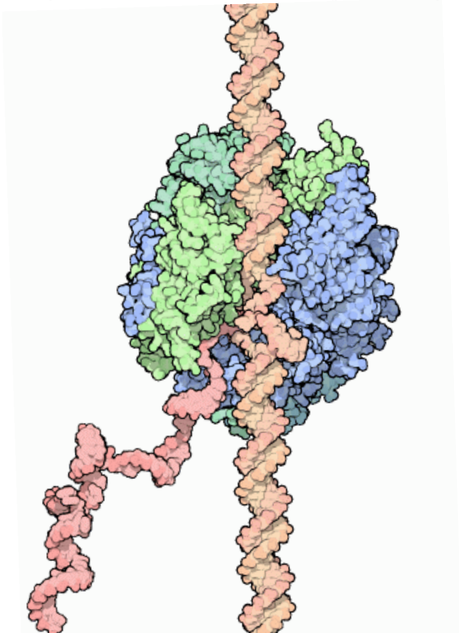
ribosome
(Nobel in Chemistry 2009)



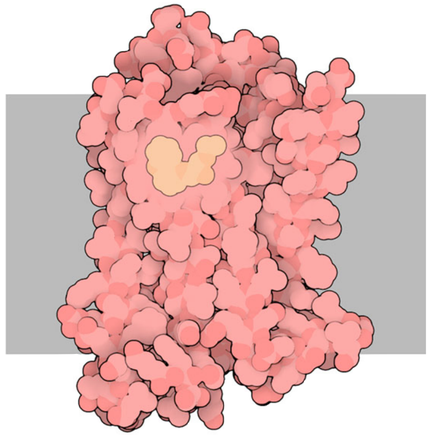
CRISPR/Cas9
(Nobel in Chemistry 2020)

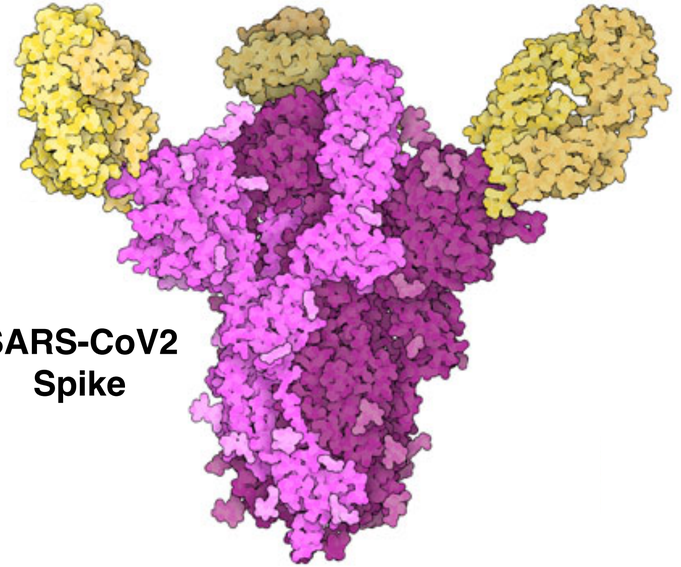
KcsA potassium channel
(Nobel in Chemistry 2003)

RNA polymerase II
(Nobel in Chemistry 2006)



GPCRs
(Nobel in Chemistry 2012)





SARS-CoV2 Spike

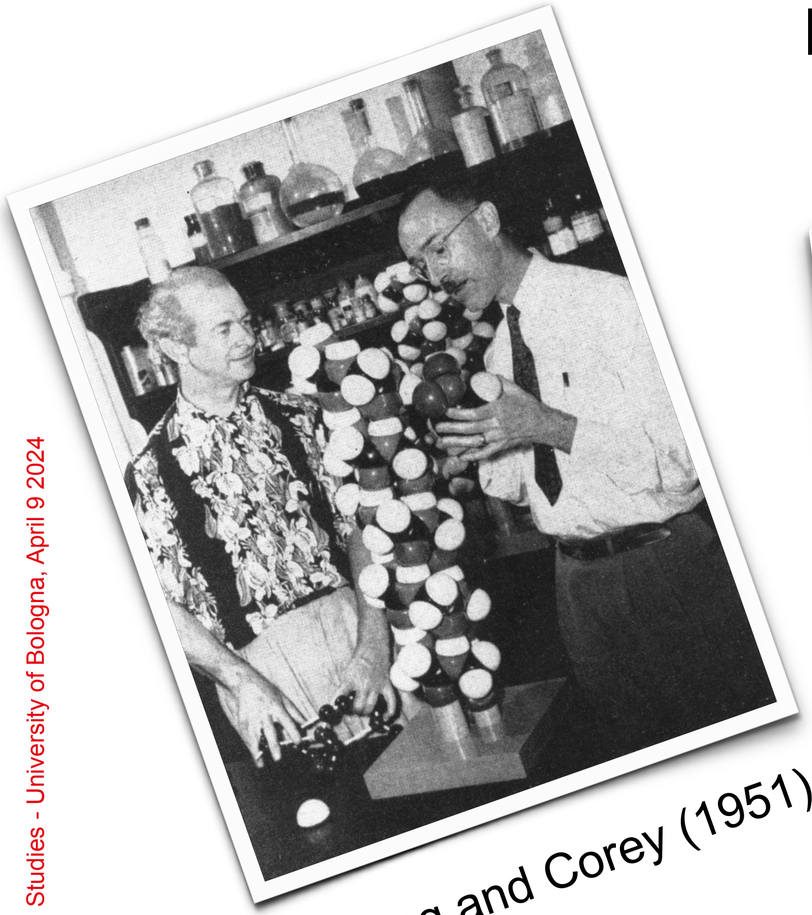
http://www.rcsb.org

Matteo Dal Peraro

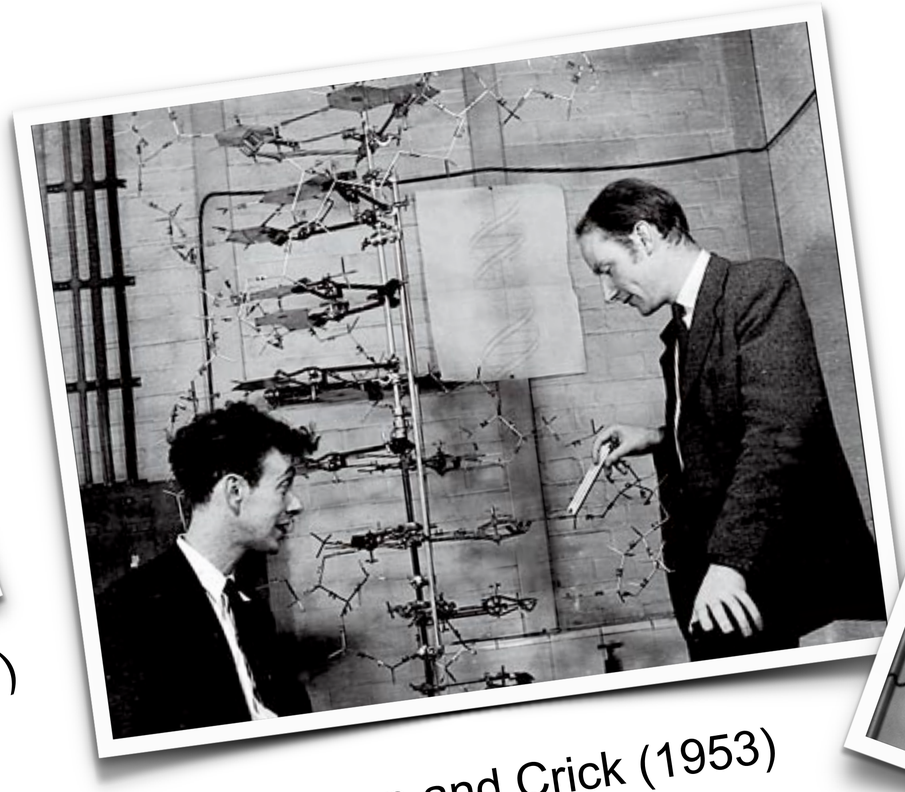
Structure *is* function

If you want to understand function, study structure

F. Crick



Pauling and Corey (1951)



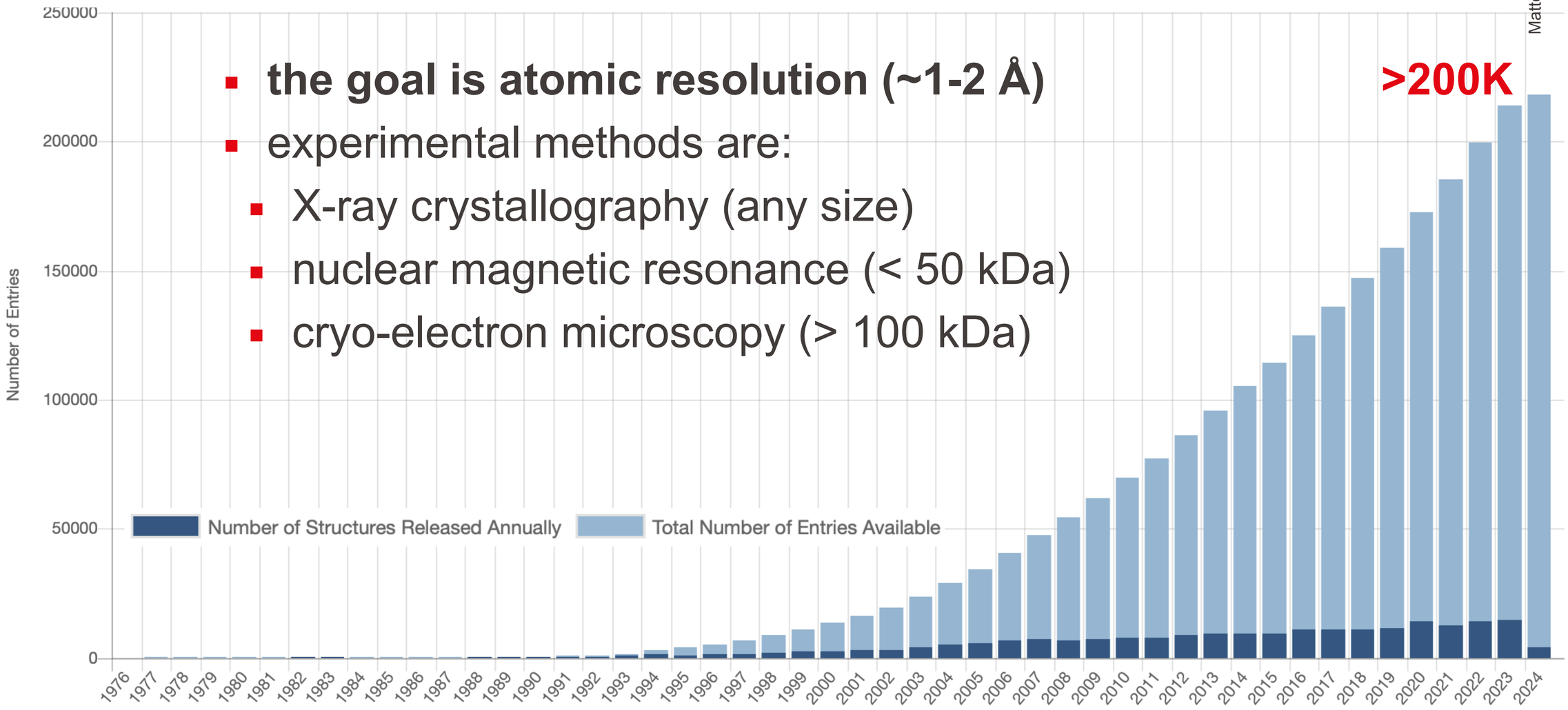
Watson and Crick (1953)

Perutz and Kendrew (1959)

Structure determination is key

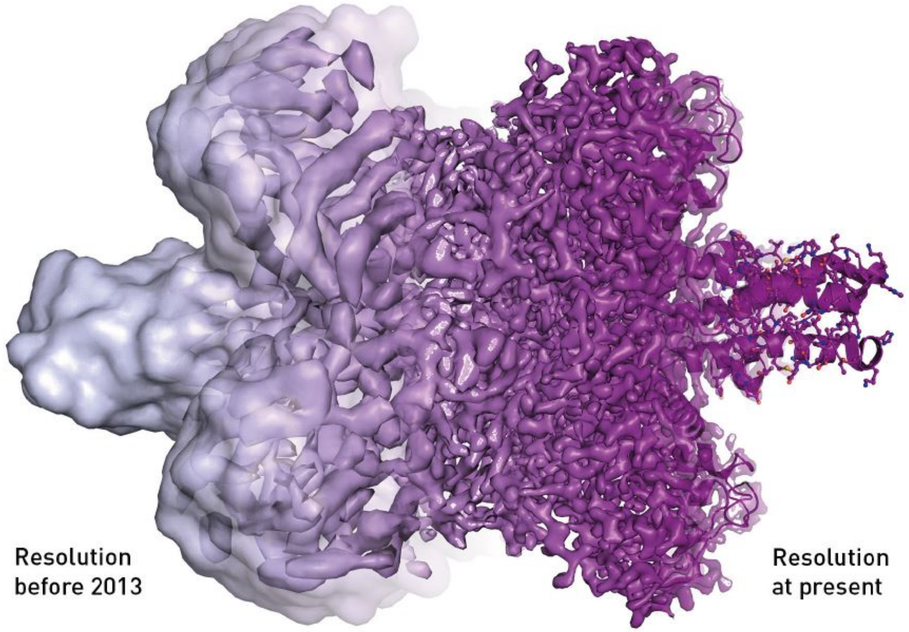
- **the goal is atomic resolution (~1-2 Å)**
- experimental methods are:
  - X-ray crystallography (any size)
  - nuclear magnetic resonance (< 50 kDa)
  - cryo-electron microscopy (> 100 kDa)

>200K



- **The Protein DataBank — PDB** **http://www.rcsb.org**

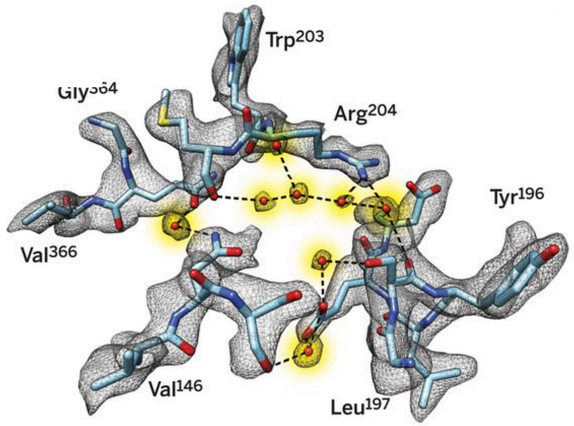
Experimental revolution — cryoEM

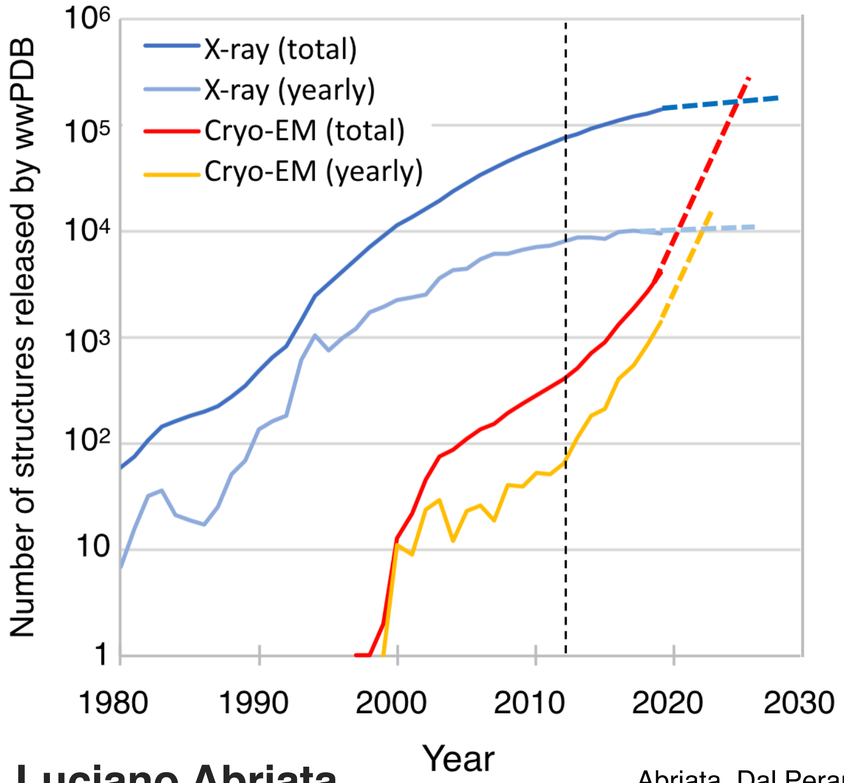


- progress in cryo-electron microscopy led to much improved resolution
- cryo-EM is becoming the new gold standard in structural biology

Krios G4 Cryo-TEM

resolution record 1.22 Å
size record ~50 kDa





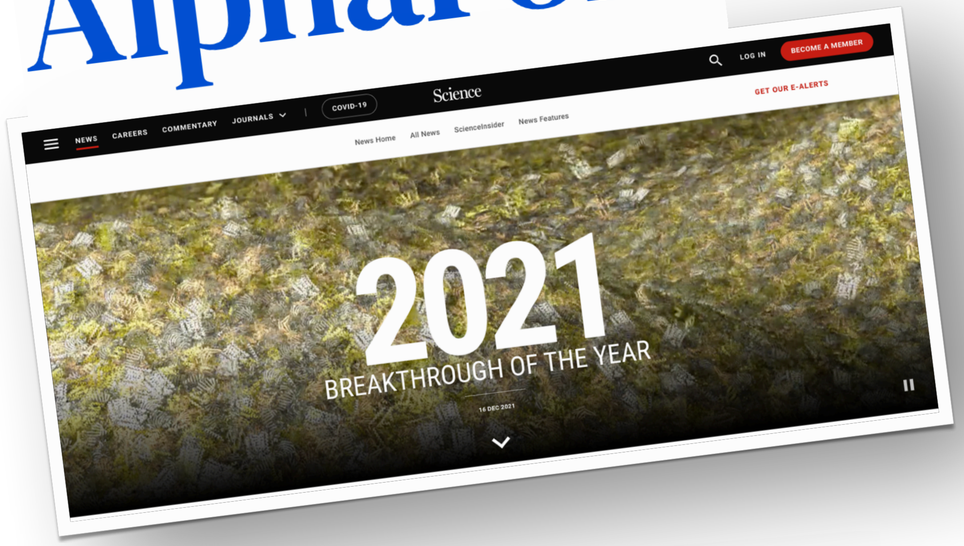
Luciano Abriata

Abriata, Dal Peraro, **JCIM** 2020

Data revolution — machine learning

DeepMind

AlphaFold



FOCUS | 11 JANUARY 2022

Method of the Year 2021: Protein structure prediction

Protein structure prediction is our Method of the Year 2021, for the remarkable levels of accuracy achieved by deep learning-based methods in predicting the 3D structures of proteins and protein complexes, essentially solving this long-standing challenge.

NEWS | BIOLOGY

'The game has changed.' AI triumphs at solving protein structures

In milestone, software predictions finally match structures calculated from experimental data

One of biology's biggest mysteries 'largely solved' by AI

By Helen Briggs
BBC science correspondent

NEWS | 30 November 2020

'It will change everything': DeepMind's AI makes gigantic leap in solving protein structures

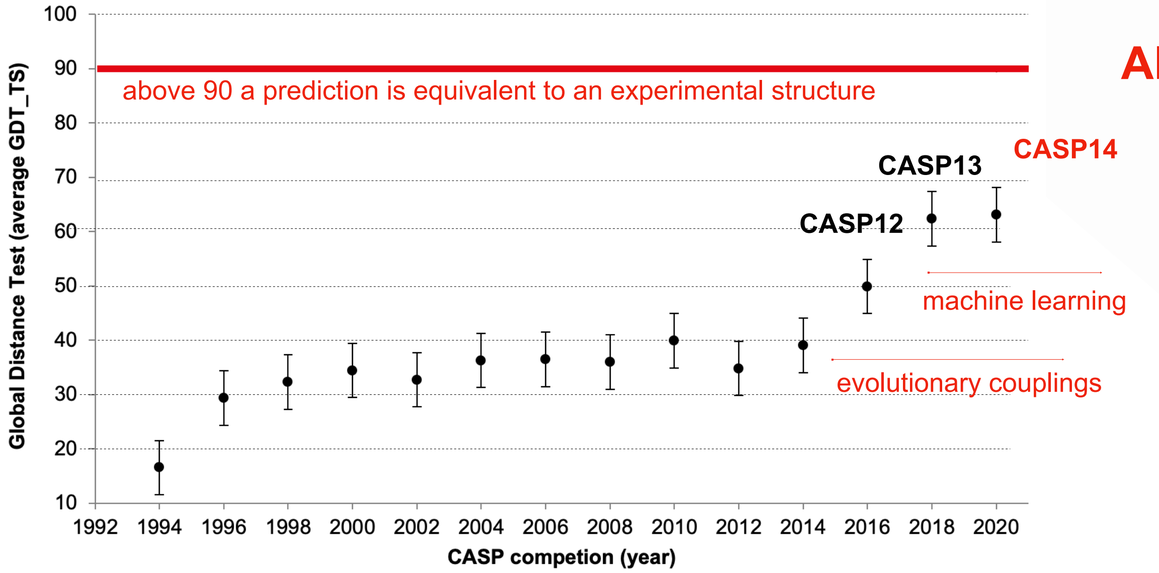
Google's deep-learning program for determining the 3D shapes of proteins stands to transform biology, say scientists.

'THE ENTIRE PROTEIN UNIVERSE': AI PREDICTS SHAPE OF NEARLY EVERY KNOWN PROTEIN

DeepMind's AlphaFold tool has determined around 200 million protein structures, which are now available to scientists in a database.

AlphaFold2 solved a 70yo problem

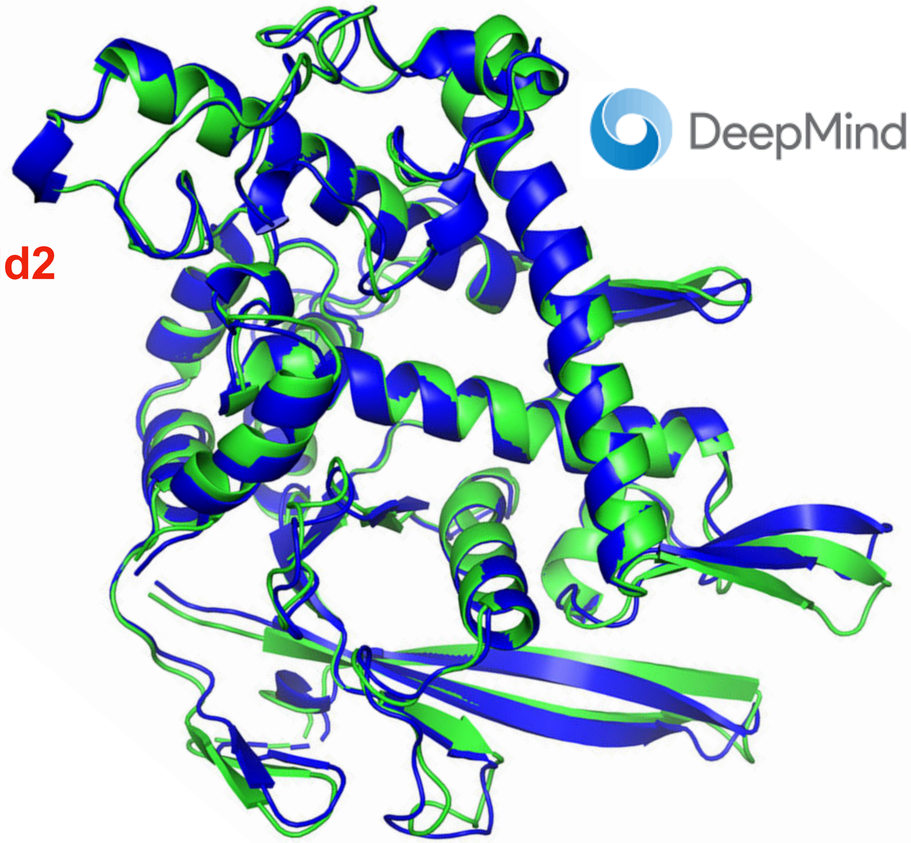
Critical Assessment of protein Structure Prediction — CASP



Abriata and Dal Peraro, **Proteins** 2019
Abriata, Tamo' and Dal Peraro, **Proteins** 2018

Luciano Abriata

AlphaFold2



experimental structure
computational prediction

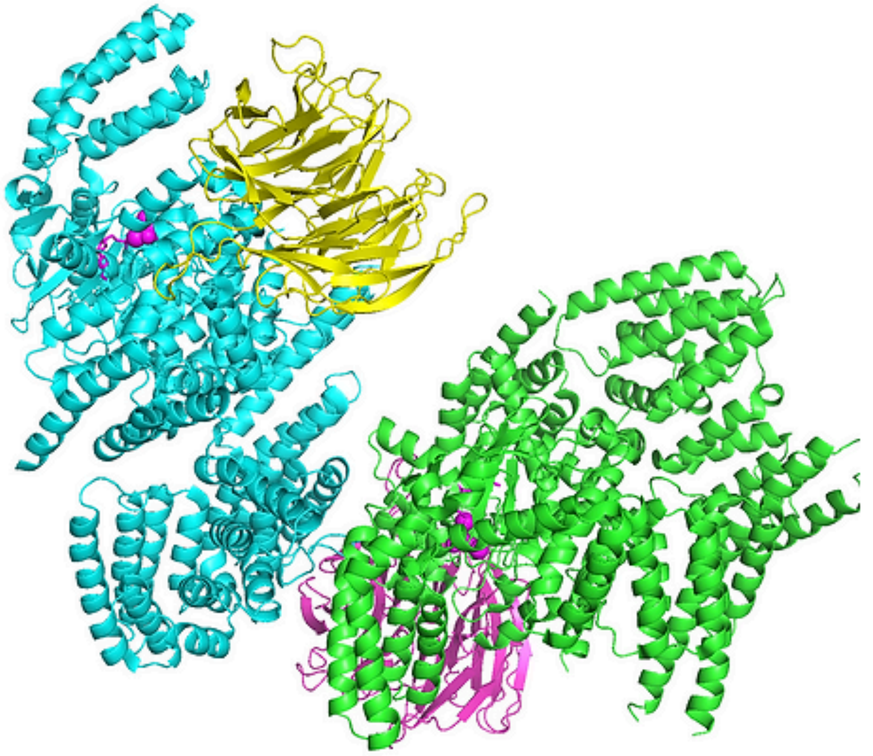
- experimental-like accuracy
- >200 M predicted models available in UniProt

EPFL MAKE Project

https://www.designinglifewithai.com/

Designing Life with AI

We're thrilled to introduce "Designing Life with AI" at EPFL, where AI and protein design intersect, involving faculty, professors, and 40 students collaborating on topics like binder design and phosphosite engineering to kinase remodeling. After a year of innovative research, our projects are now being tested in the wet-lab, and we're working on creating a pipeline and resources for new students, aiming to expand our project and make EPFL a hub for protein design.



prof. Sahand Jamal Rahi
Laboratory of the Physics of Biological Systems

prof. Paolo De Los Rios
Laboratory of Statistical Biophysics

prof. Francesco Stellacci
Supramolecular NanoMaterials and Interfaces Laboratory

prof. Bruno Correia
Laboratory of Protein Design and Immunoengineering

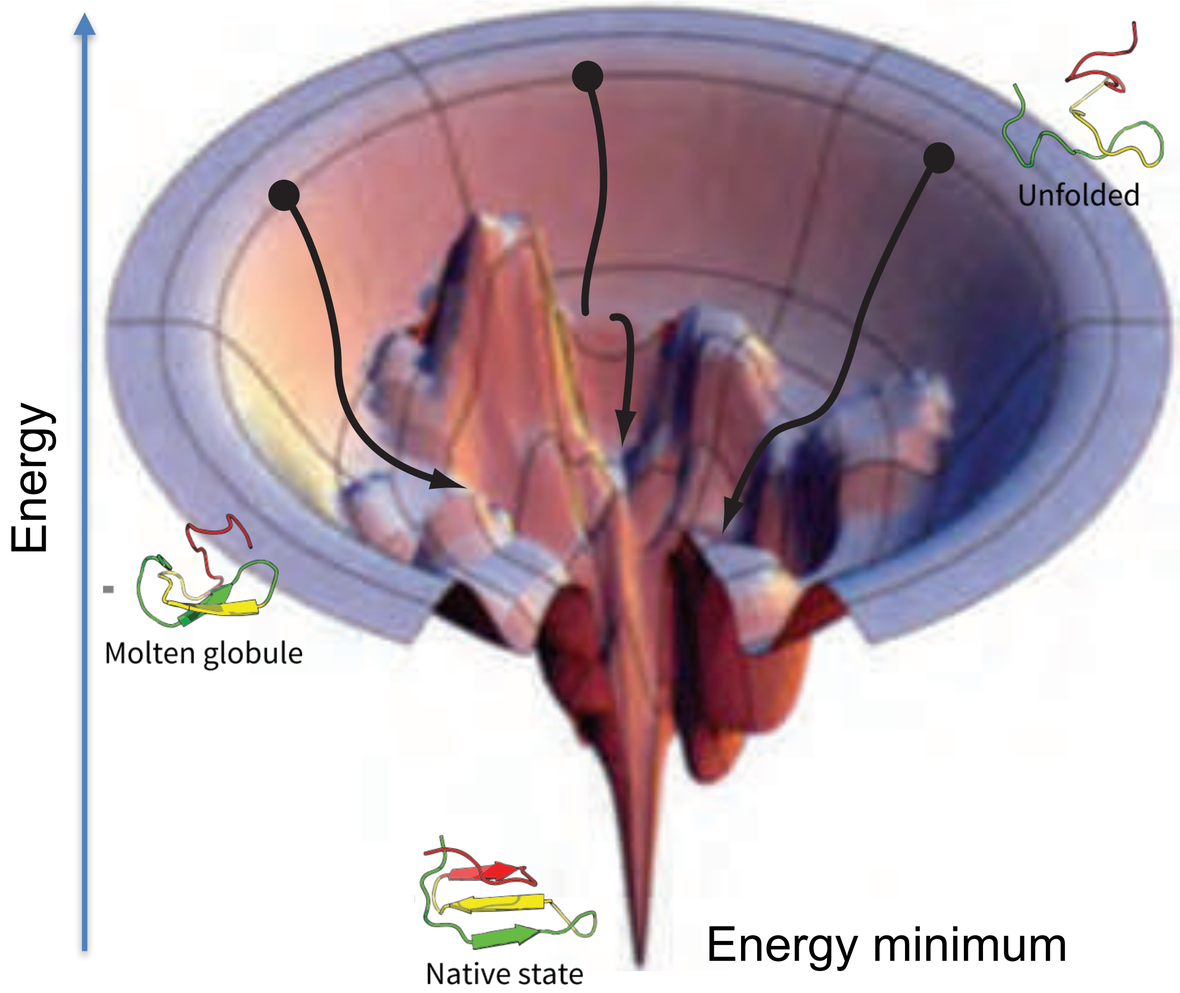
prof. Li Tang
Laboratory of Biomaterials for Immunoengineering

prof. Matteo Dal Peraro
Laboratory for Biomolecular modeling

prof. Patrick Barth
Laboratory of Protein and Cell Engineering

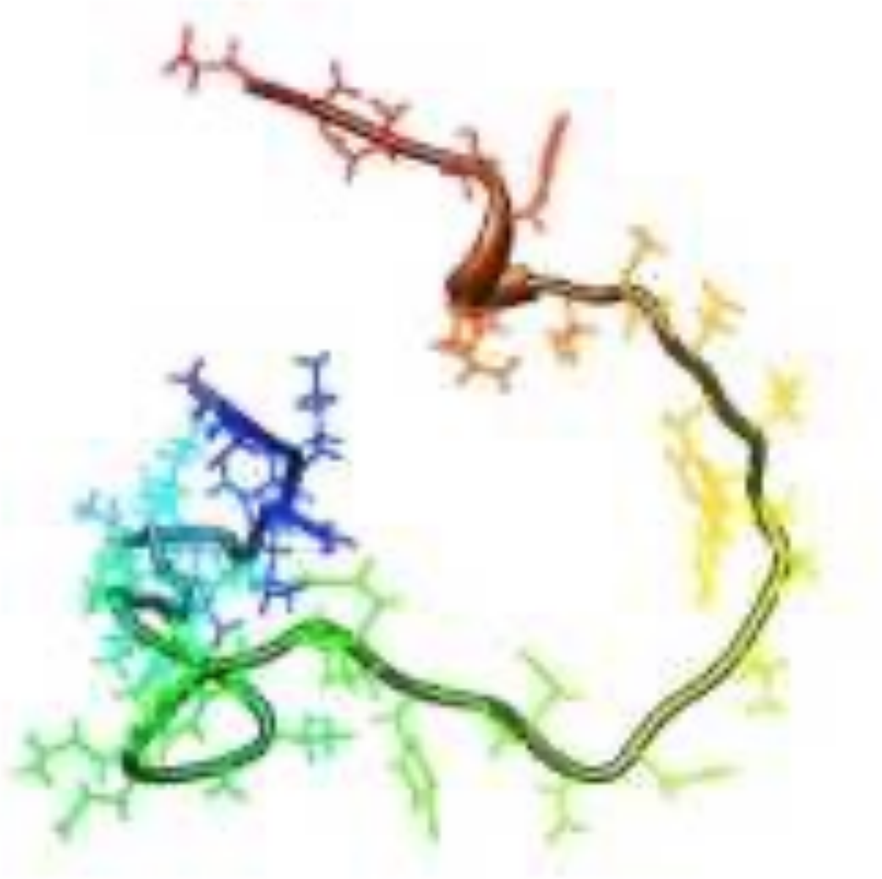
prof. Angela Steinauer
Laboratory of Biomolecular Engineering and Nanomedicine

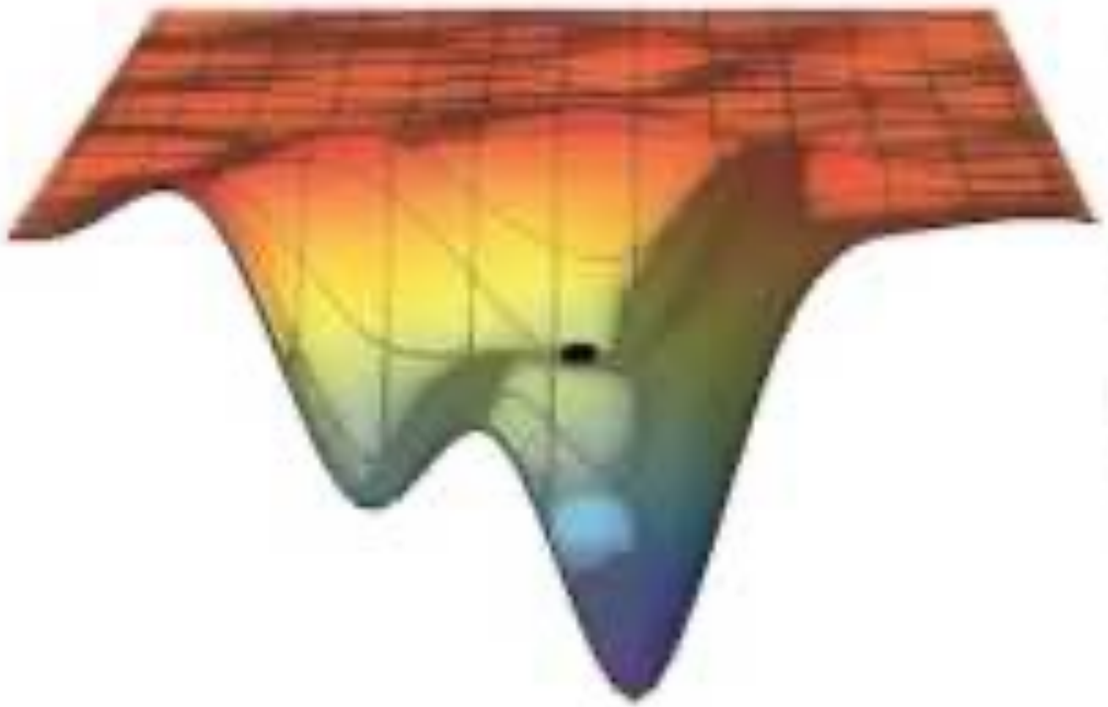
Energy as the main driver of biological processes



Changes that drive a molecular system to lower energies determines what happens at the molecular and higher levels in biology (eg protein folding)

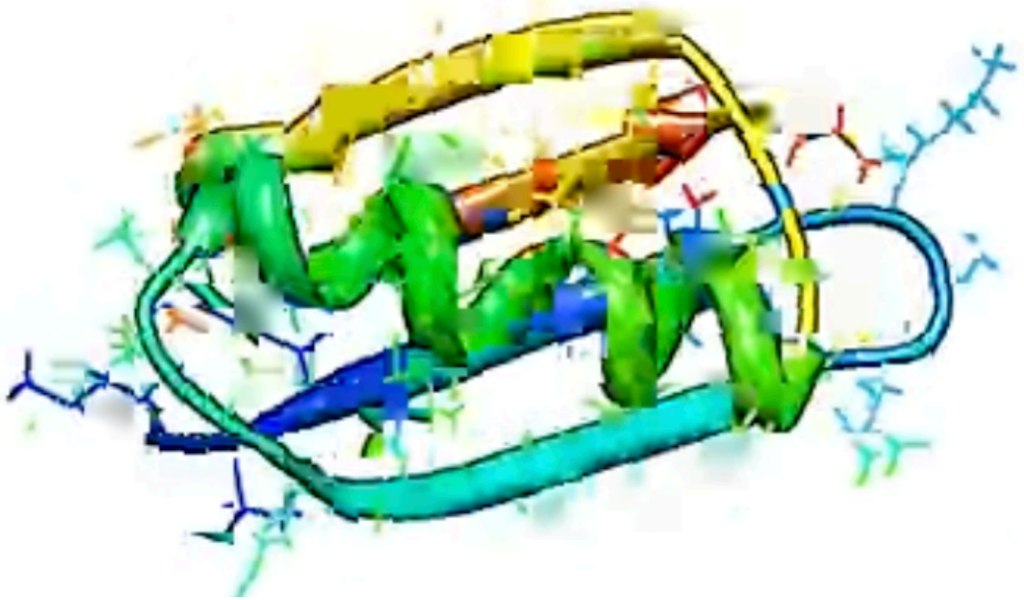
Protein folding

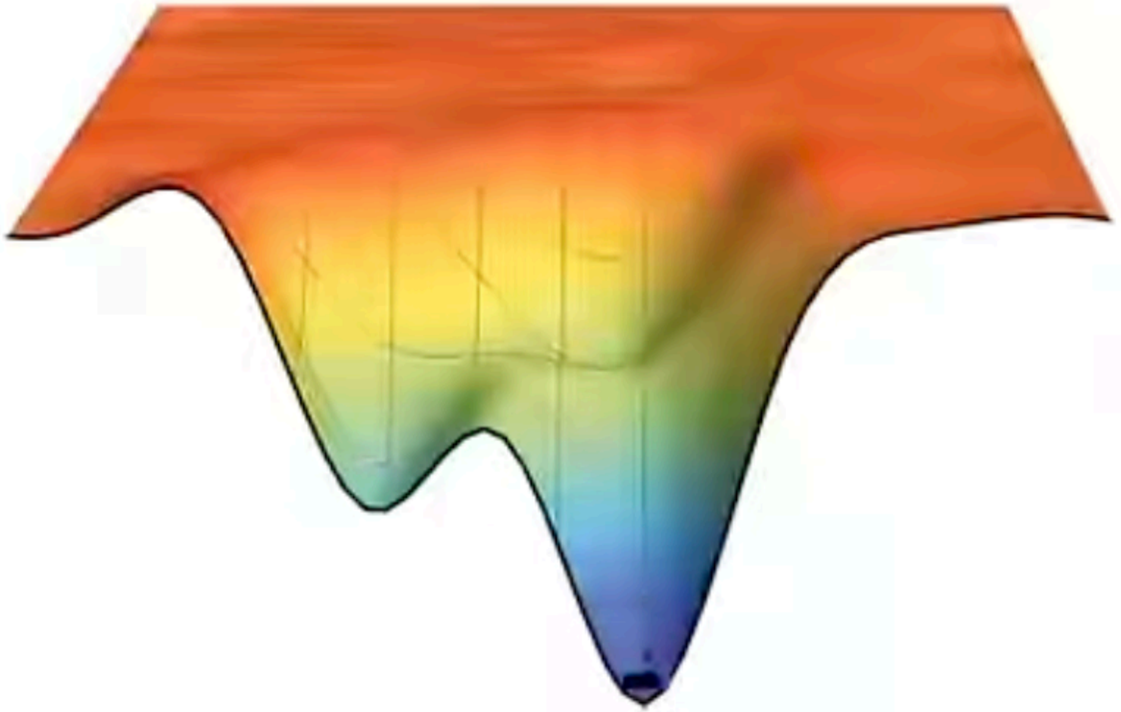




- Stoony Brook University (2021)

Protein folding

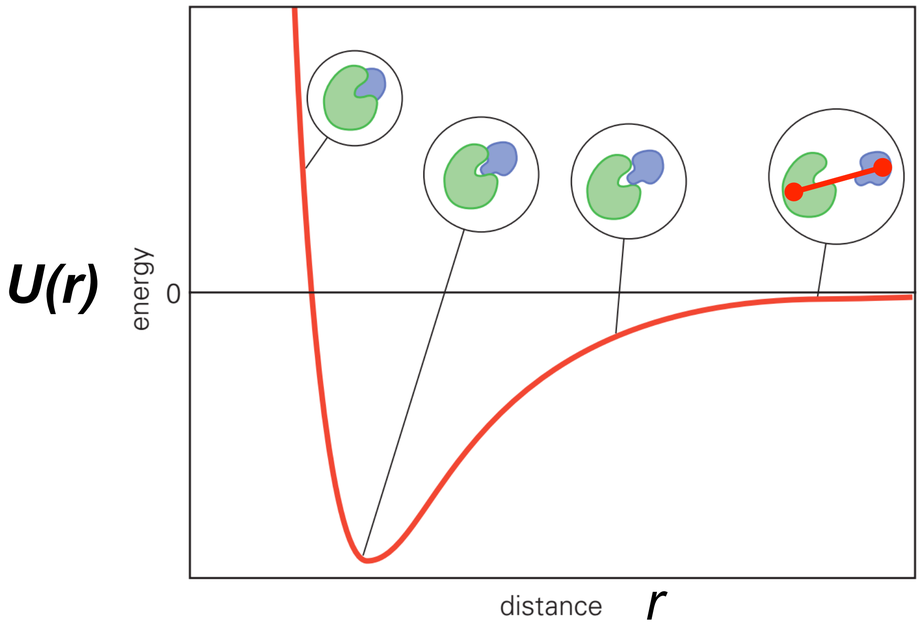




- Stoony Brook University (2021)

Molecular Interactions in Biomolecules

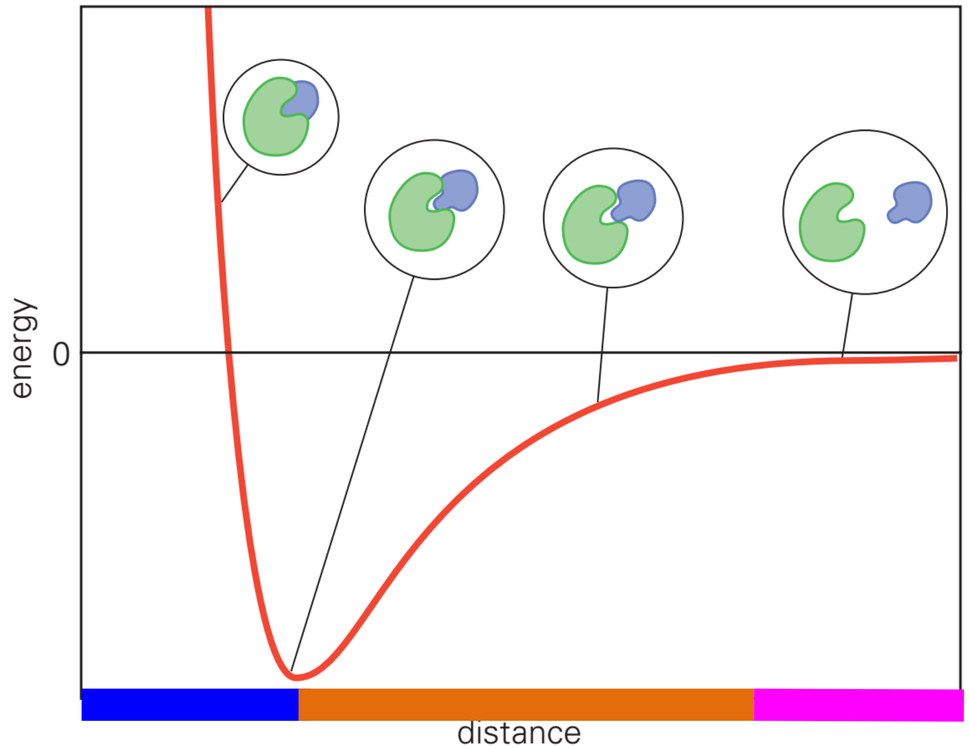
There are two levels:
- interactions that happens within a biomolecule in a **covalently** by chemical bonds)
- interaction between different biomolecules (**intermolecular**) or within the same molecule (**intramolecular**), determined by **noncovalent** interactions



The energy of non-covalent interactions is dependent on the **distances** between molecules - ***U(r)***

Molecular Interactions in Biomolecules

The energy of interaction between two molecules is determined by **<u>noncovalent</u>** interactions



r_0 = equilibrium distance

Nature of the Interaction

Repulsive

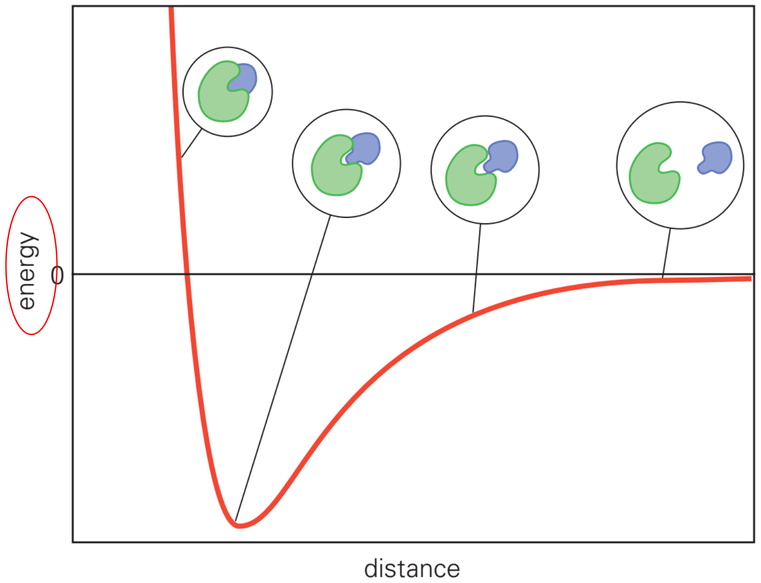
(due to the electronic repulsion
Pauli exclusion principle)

Attractive

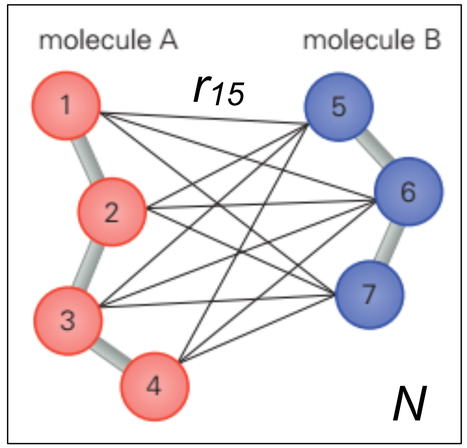
Non-Interacting

Molecular Interactions in Biomolecules

The energy of interaction between two molecules is determined by **noncovalent** interactions



Atomic pairwise Interactions



$$U(r) \sim \sum_{i<j}^{N} U(r_{ij})$$

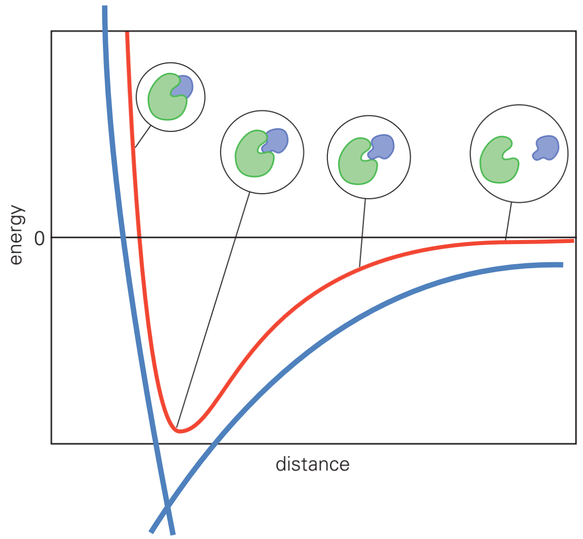
But how do we compute this energy ?

-Given that we can describe our system of interest at the atomic level,
we can compute the sum of the **pairwise interactions** between all the different pairs of atoms.

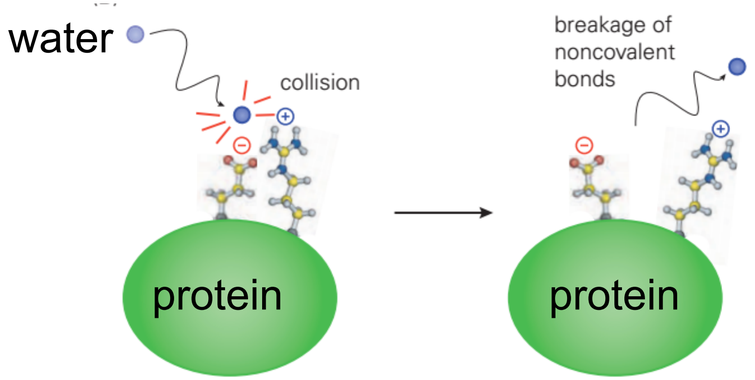
-This approach is applicable to all the building blocks of life (based on classical mechanics, which is a good empirical approximation as quantum mechanics more accurately describes molecular interactions).

Molecular Interactions in Biomolecules

<u>noncovalent</u> interactions



Noncovalent interactions are transient



Noncovalent interactions are broken and remade simply due to thermal fluctuations (related to the thermal energy of every degree of freedom, k_BT **~ 2.5 kJ/mol** at 300 K)

"everything that living things do can be understood in terms of the jigglings and wigglings of atoms." R. Feynman

-Important types of noncovalent interactions in biomolecules:

- **van der Waals interactions -**
- **Ionic interactions**
- **Hydrogen bonds**

$$U(r) \sim -\sum_{i<j}^{N} \frac{1}{r_{ij}^{n}}$$

Molecular Interactions in Biomolecules

van der Walls interactions (London forces)

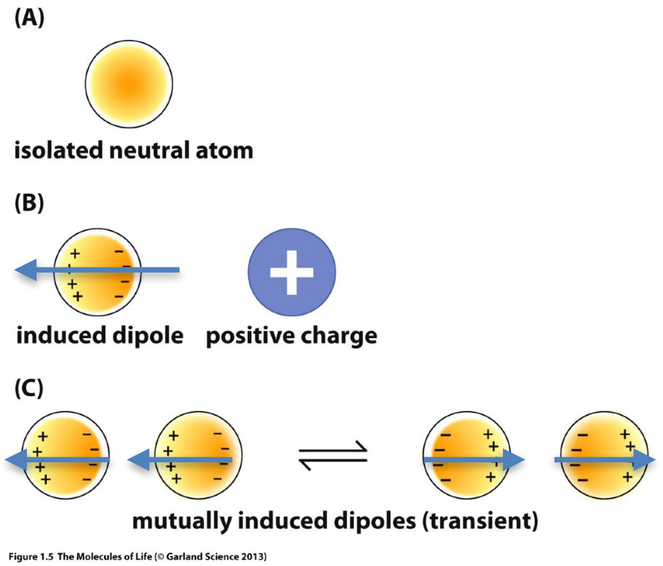
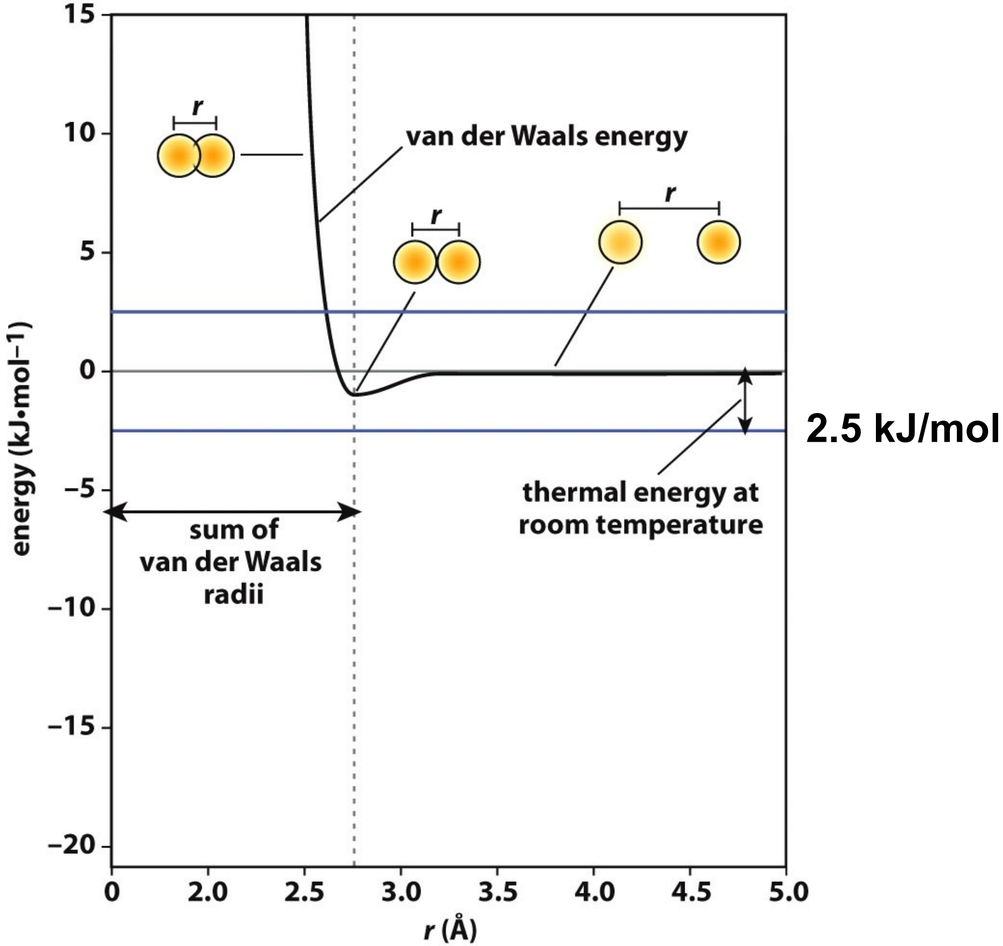


Figure 1.5 The Molecules of Life (© Garland Science 2013)

$$\mu_{ind} = \alpha \mathbf{E}$$

$$\alpha : polarizability$$

Atom	van der Waals radius (Å)	Electro-negativity (Pauling scale)
O	1.5	3.4
Cl	1.9	3.2
N	1.6	3.0
S	1.8	2.6
C	1.7	2.6
P	1.8	2.2
H	1.2	2.1



- due to induced dipoles in atoms

- Notice the quantitative information of such potential (real energies and distances).

- In this potential there is a repulsive, attractive and non interacting region.

- The radius of each atom determines the distance of minimal energy (sum of vdW radii).

Molecular Interactions in Biomolecules

van der Walls interactions (London forces)

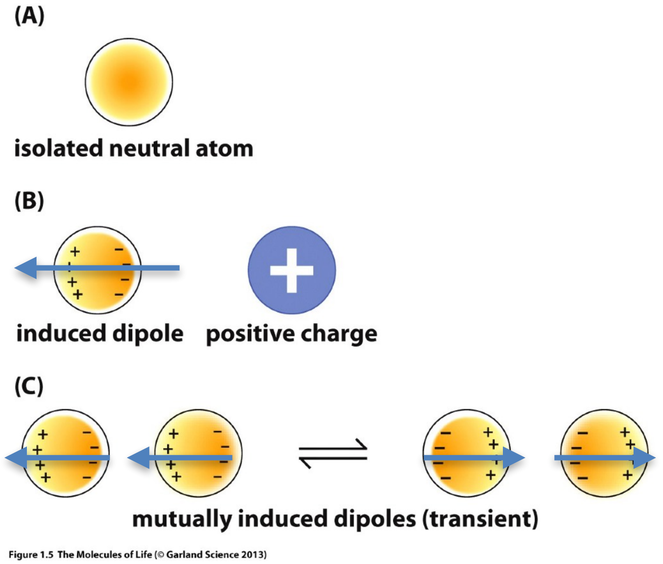
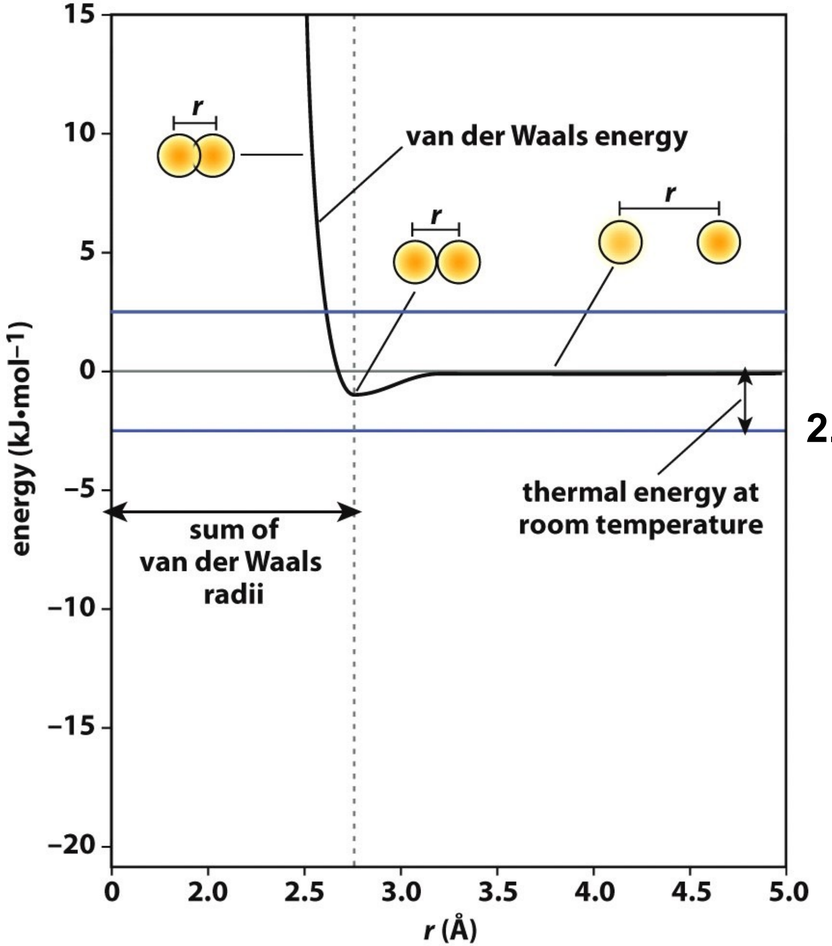


Figure 1.5 The Molecules of Life (© Garland Science 2013)

$$\mu_{ind} = \alpha \mathbf{E}$$

$$\alpha : polarizability$$

Atom	van der Waals radius (Å)	Electro-negativity (Pauling scale)
O	1.5	3.4
Cl	1.9	3.2
N	1.6	3.0
S	1.8	2.6
C	1.7	2.6
P	1.8	2.2
H	1.2	2.1



- due to induced dipoles in atoms

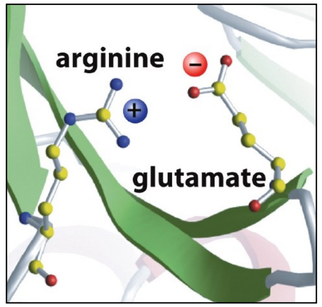
-Individual vdW interactions are very weak - attractive part goes like $U(r) \sim -1/r^6$.

-Magnitude of the thermal energy is greater than a vdW interaction.

-Many add up to significant energies for the stabilization of biomolecules

Molecular Interactions in Biomolecules

Ionic interactions: Simplest kind of interaction is between two charged atoms (if charges are opposite are called **salt bridges**)



Energy potential for ion pair interactions:

-Similar distance dependency to van der Waals

-Stabilization energy is much greater ***- U(r) ~ -1/r***

-The equilibrium distance is smaller

$$U_{electrostatic}(r) = \frac{1}{4\pi\epsilon_0}\frac{q_i q_j}{r_{ij}}$$

$$= 1391\ kJ \cdot mol^{-1}\frac{q_i q_j}{r_{ij}[\mathring{A}]}$$

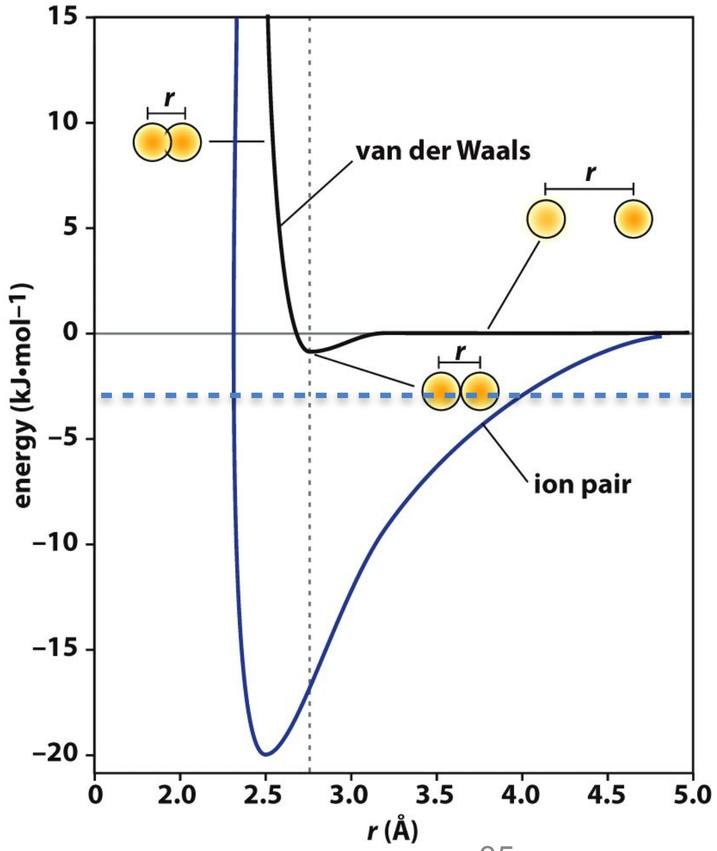


Figure 1.9 The Molecules of Life (© Garland Science 2013)

Molecular Interactions in Biomolecules

Ionic interactions: Simplest kind of interaction is between two charged atoms (if charges are opposite are called **salt bridges**)

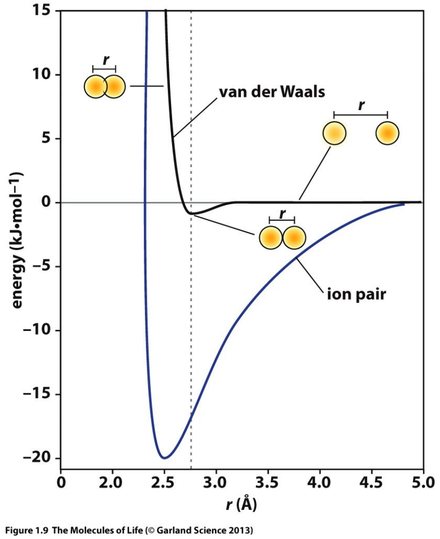


Figure 1.9 The Molecules of Life (© Garland Science 2013)

Energy potential for ion pair interactions:

-Similar distance dependency to van der Waals

-Stabilization energy is much greater

-The equilibrium distance is smaller

Ionic interactions are dependent on the environment:
- Water and ions reduce electrostatic interaction strength

$$U(r) = \frac{1}{4\pi\epsilon_0}\frac{1}{D}\frac{q_1 q_2}{r}$$

$$= 1391\ kJ \cdot mol^{-1}\frac{q_i q_j}{r_{ij}[\mathring{A}]}\frac{1}{D}$$

in water

D = dielectric constant

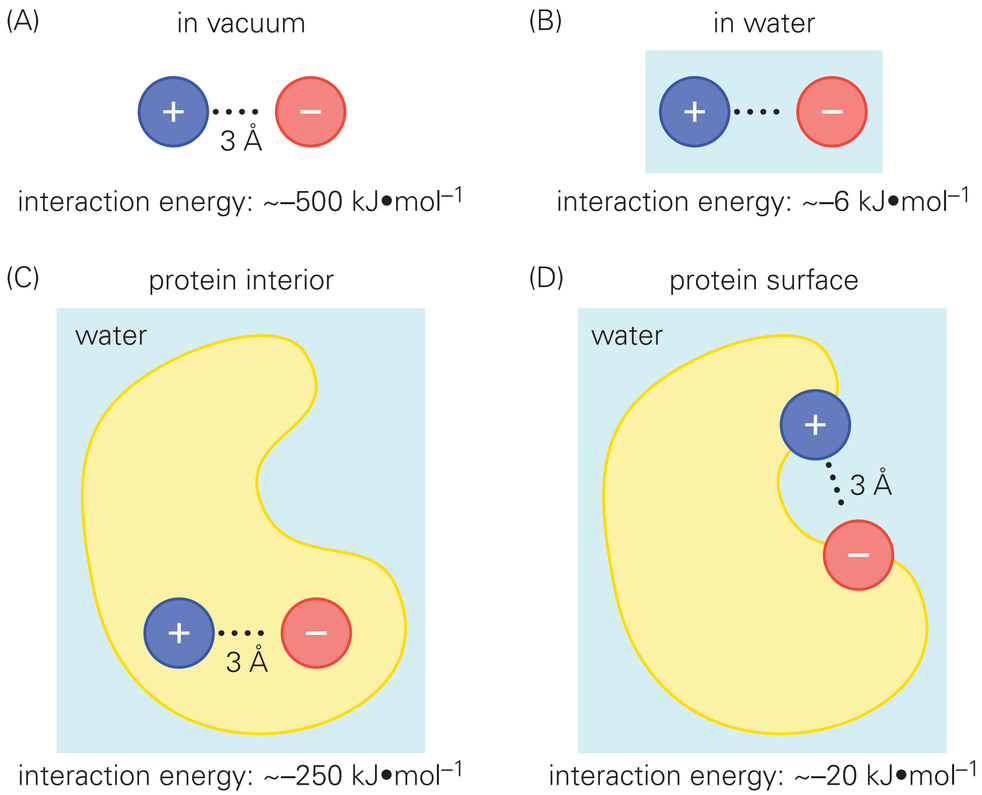
[air(T=0°C) D=1.00059]

Liquid	T (°C)	D
Heptane	0	1.958
Heptane	30	1.916
Methanol	25	33
Formamide	20	109
Formic acid	16	58
Nitrobenzene	25	35
HCN	0	158
HCN	20	114
Glycol	25	37
Water	0	88.00
Water	25	78.54

Discussion

How are salt bridges affected by the environment ?

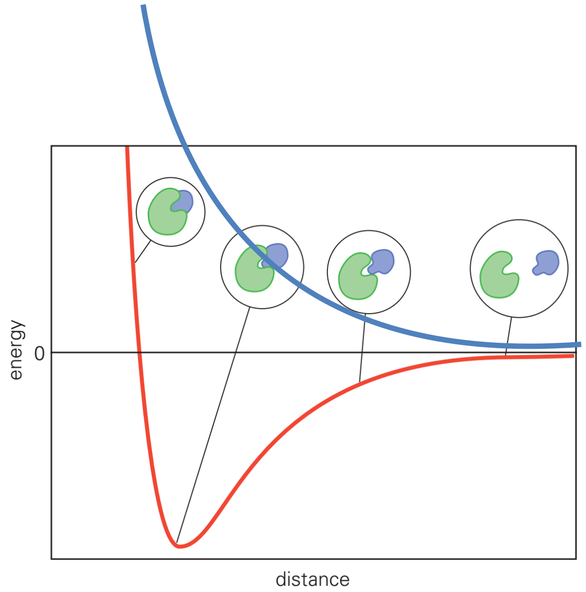
$$U(r) = 1391\ kJ \cdot mol^{-1} \frac{q_i q_j}{r_{ij}[\mathring{A}]} \frac{1}{D}$$



Discussion

How does the energy potential go if charges are of the same sign?

$$U(r) = 1391\ kJ \cdot mol^{-1} \frac{q_i q_j}{r_{ij}[\mathring{A}]} \frac{1}{D}$$

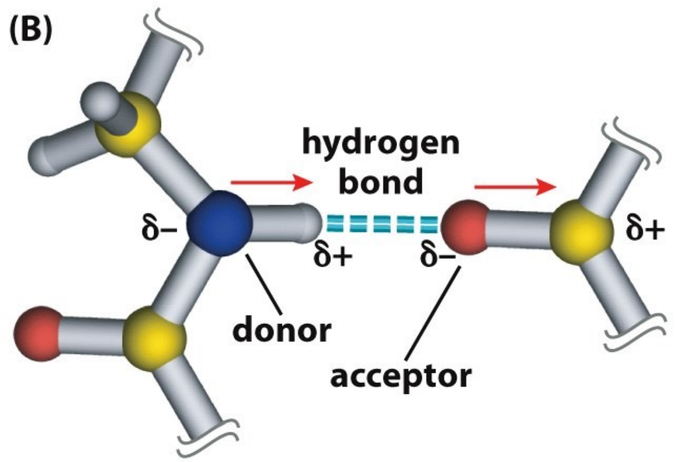


Molecular Interactions in Biomolecules

Hydrogen Bonds

- Interactions between polar groups in which a hydrogen atom with a partial positive charge is close to an atom with a partial negative charge.

H-bond



Atom	van der Waals radius (Å)	Electro-negativity (Pauling scale)
O	1.5	3.4
Cl	1.9	3.2
N	1.6	3.0
S	1.8	2.6
C	1.7	2.6
P	1.8	2.2
H	1.2	2.1

- H-bonds are distance- and angle-dependent (position of donor and acceptor atoms)

- Typical distances between 2.4-2.7 Å and angles depends nature of donor/acceptor

- Hydrogen bonds arise from the polarization of atoms involved in covalent bonds (think in terms of dipole-dipole interactions)

- H-bonds are weaker than salt bridges - they go like $U(r) \sim -1/r^3$

Molecular Interactions in Biomolecules

Hydrogen Bonds

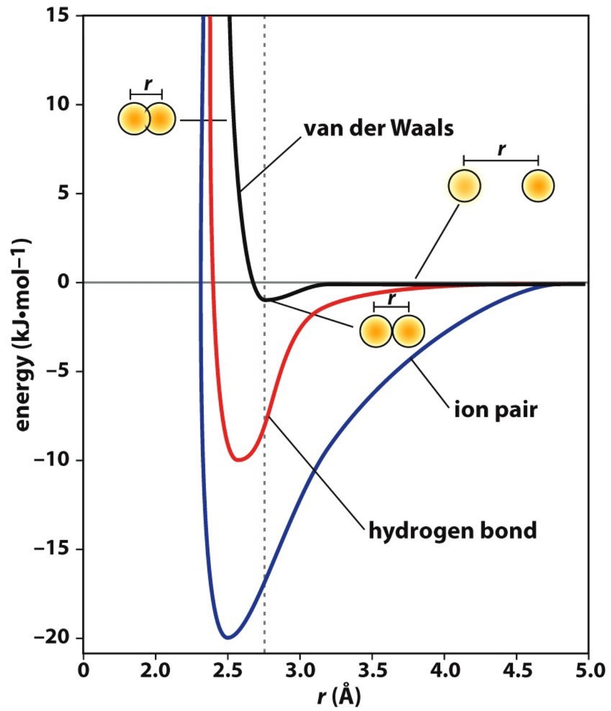


Figure 1.12 The Molecules of Life (© Garland Science 2013)

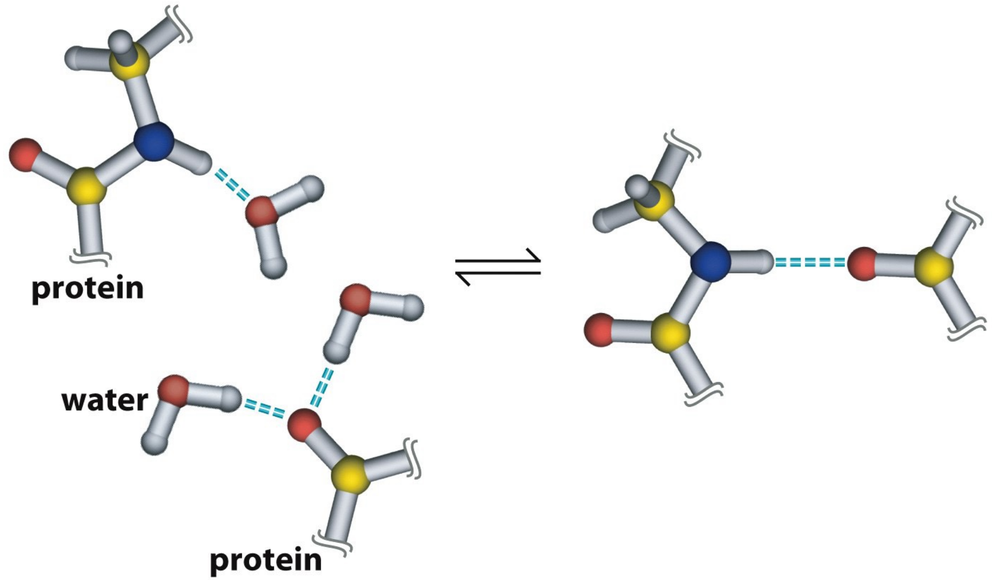


Figure 1.13 The Molecules of Life (© Garland Science 2013)

- H-bonds are also dependent on the environment – water molecules will weaken effective hydrogen bonds - solvation effect

Extra explanation

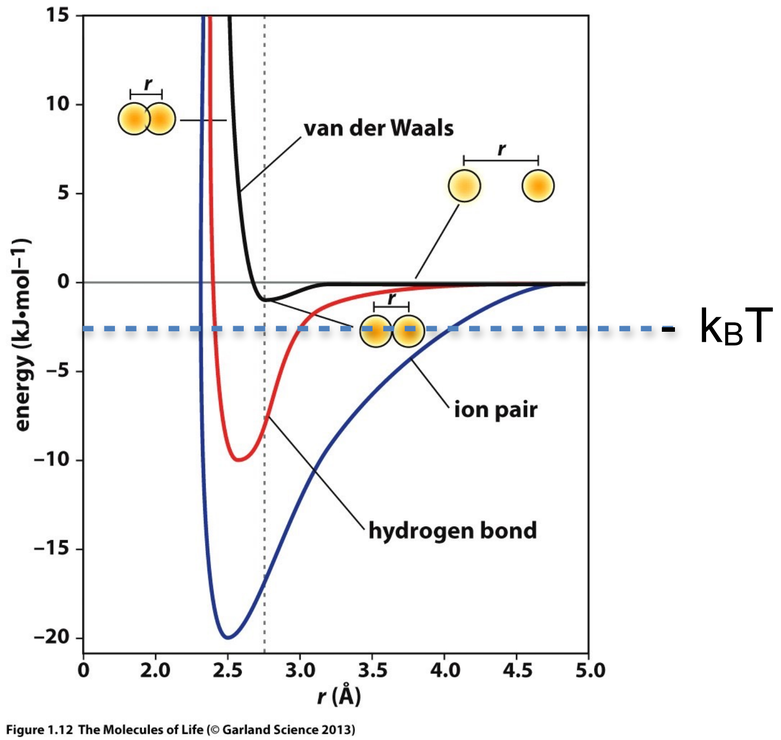


Figure 1.12 The Molecules of Life (© Garland Science 2013)

- what is important in these plots is the interacting energy (aka binding energy) of each interaction
- the binding energy is nothing else that the energy it takes to break the interaction
- in quantitative terms it is the difference between the energy at the minimum of the potential and the energy at very long distances when the two molecules do not feel each other
- this is in the order of 1 kJ/mol for vdW interactions, 10 kJ/mol for H-bonds and 20 kJ/mol or more for salt bridges.
- if you want to break these interactions this is the amount of energy you need to inject into the system
- now you have to compare these binding energies with the thermal energy k_BT, and you understand why it is easier to break a vdW interaction in biomolecules rather to break a salt bridge

Molecular Interactions in Biomolecules

Atomic Interactions

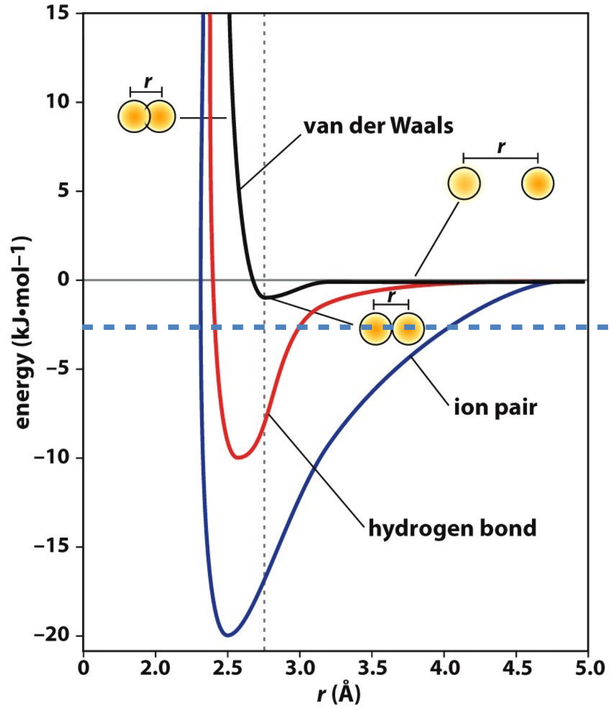
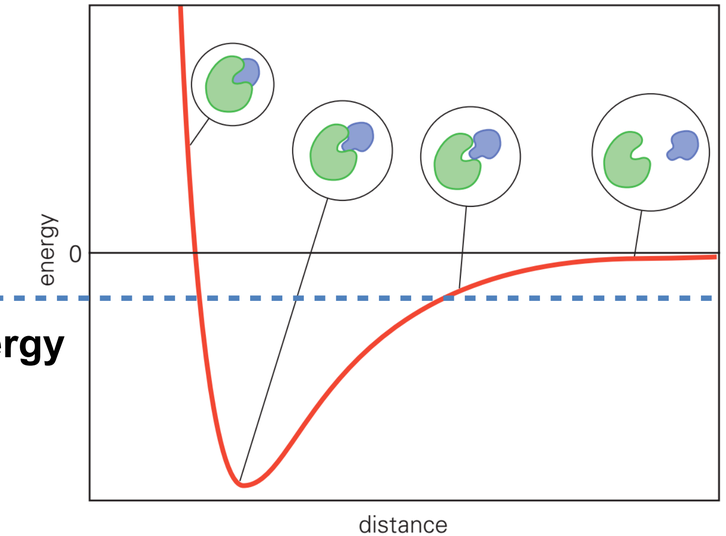


Figure 1.12 The Molecules of Life (© Garland Science 2013)

Macromolecular interactions



thermal energy

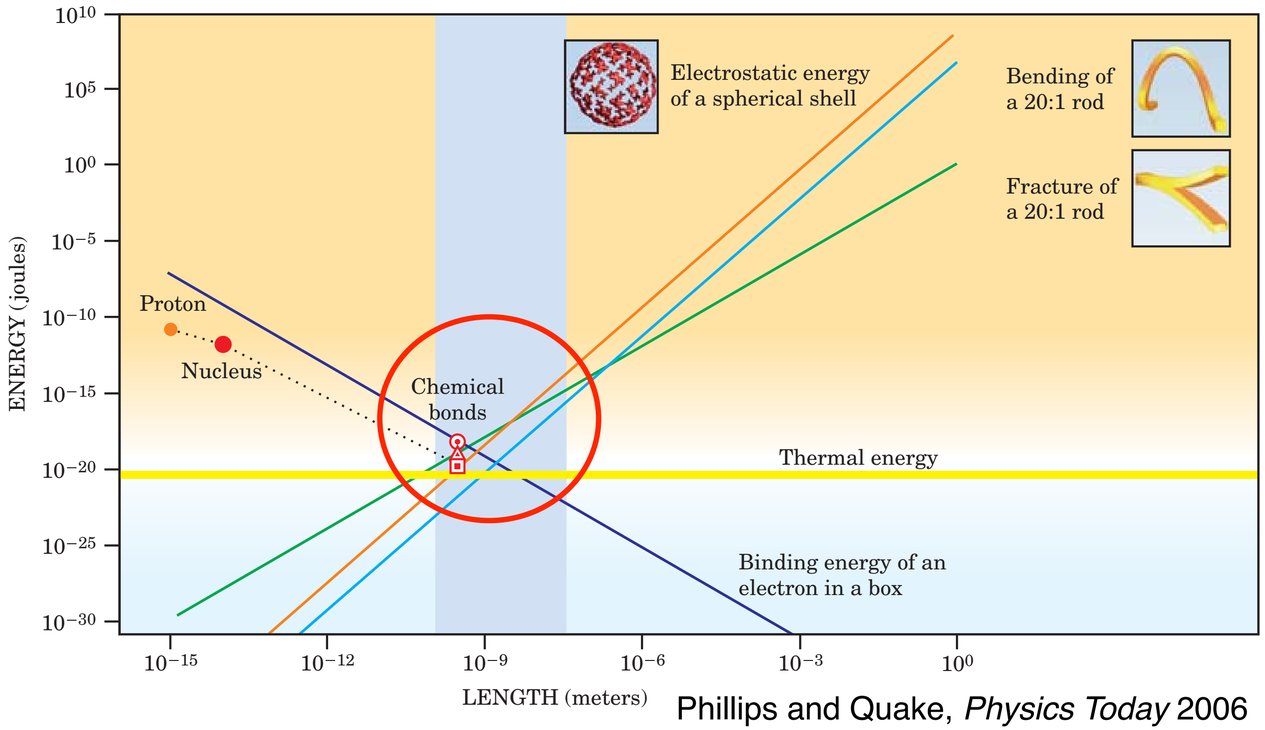
2.5 kJ/mol

$$U(r) \sim \sum_{i<j}^{N} U(r_{ij})$$ (addictive)

We can now understand some of the energetic principles that govern the interactions between macromolecules and determine how complex biological processes occur

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)

Molecular Interactions in Biomolecules

Some important take-home messages:

-The energy of interaction between different biomolecules is determined by **noncovalent interactions**

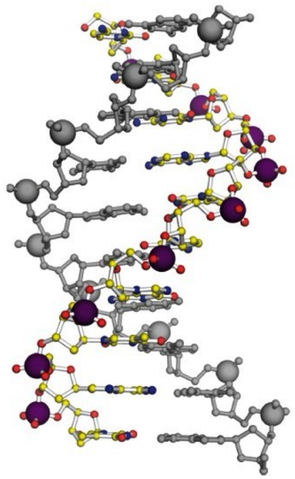
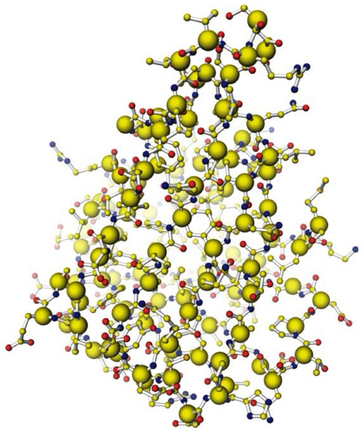
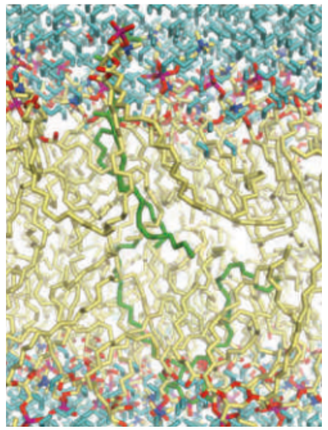
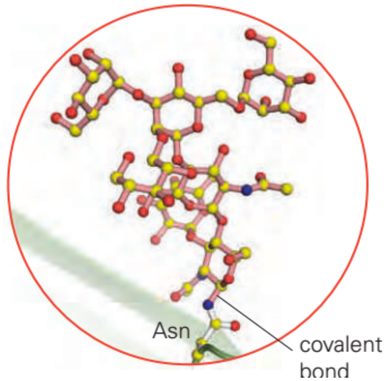
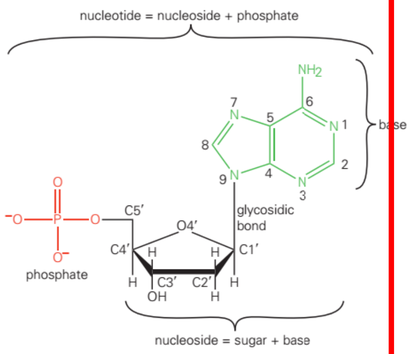
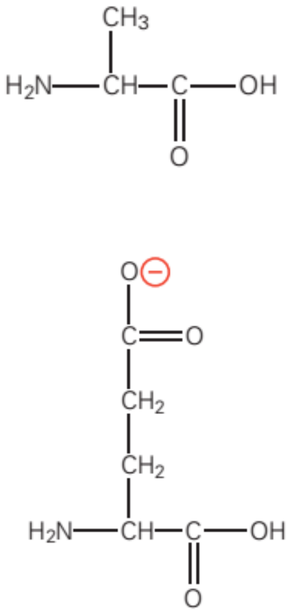
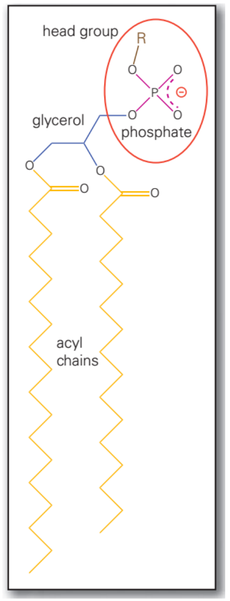
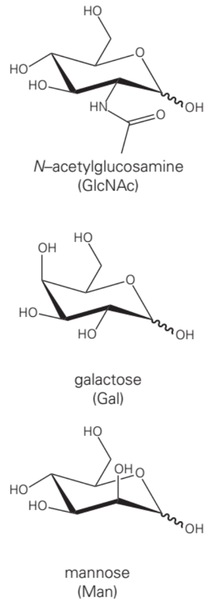
- Neutral atoms attract each other at short distances through **van der Waals interactions**

- **Ionic interactions** between charged molecules can be very strong, but are attenuated by water molecules (screening effect)

- **Hydrogen bonds** are very common in biological macromolecules and are a consequence of polarization of covalent bonds

- Always consider these molecular interactions with respect to the **thermal energy** level

The Molecules of Life

	Nucleic Acids	Proteins	Lipids	Glycans
Macromolecular Structure				
Building Block				

DNA in the cell

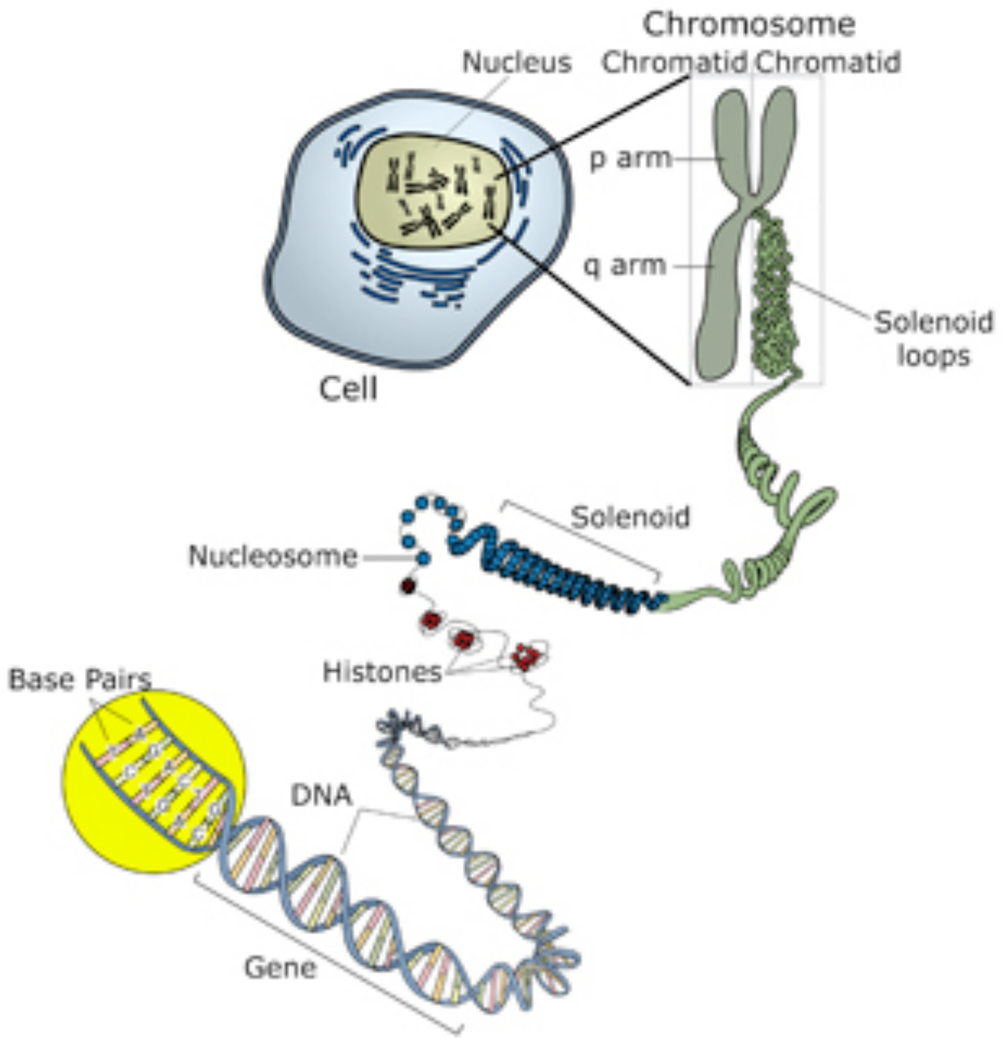
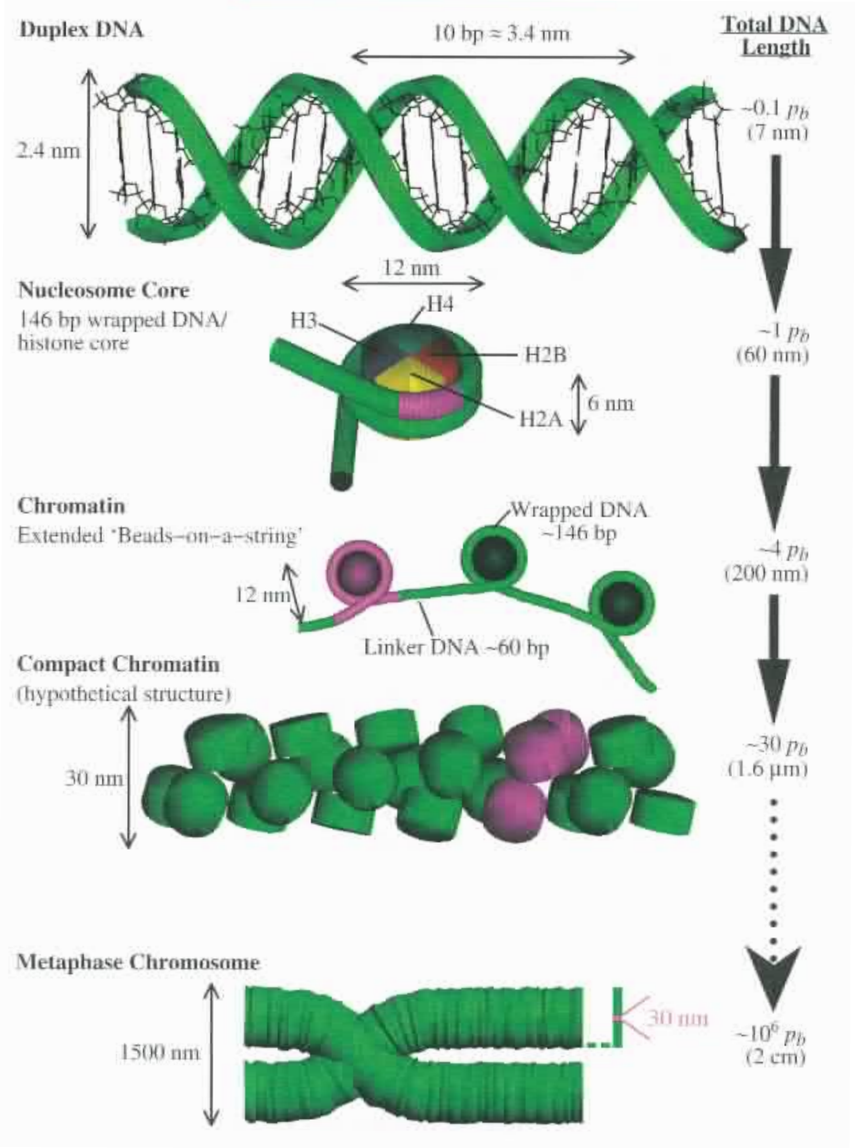


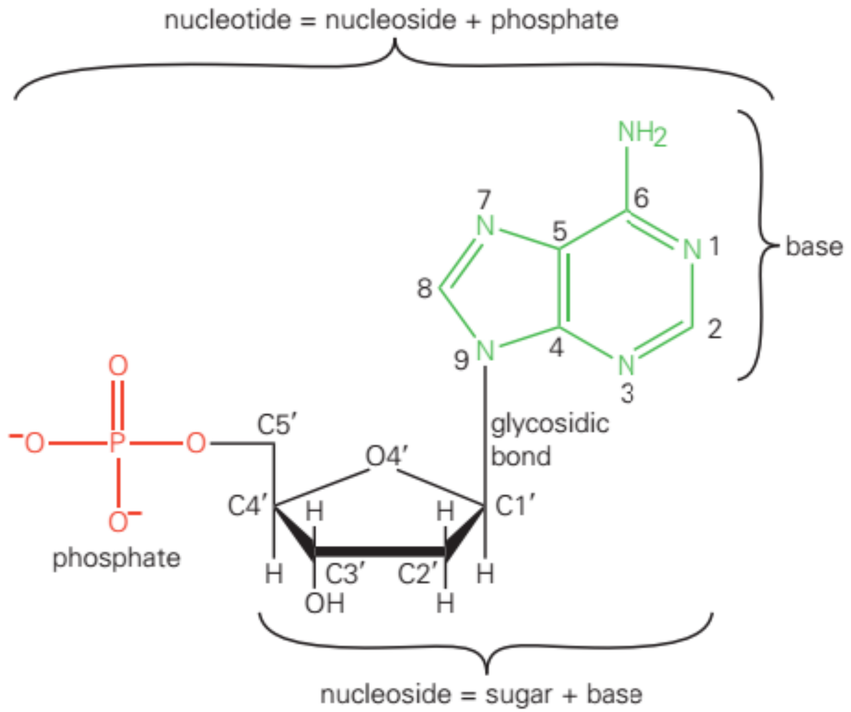
Image adapted from: National Human Genome Research Institute.



Nucleic Acids

-DNA and RNA are both polymers of **nucleotides**.

Structure of a nucleotide

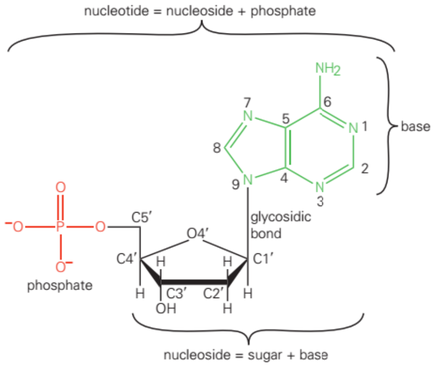


-Key functional groups:

- Five carbon sugar – **pentose** in black
- nitrogen-containing aromatic ring system, i.e. **base** – adenine in green
- **phosphate** group in red (ranging from 1 to 3)

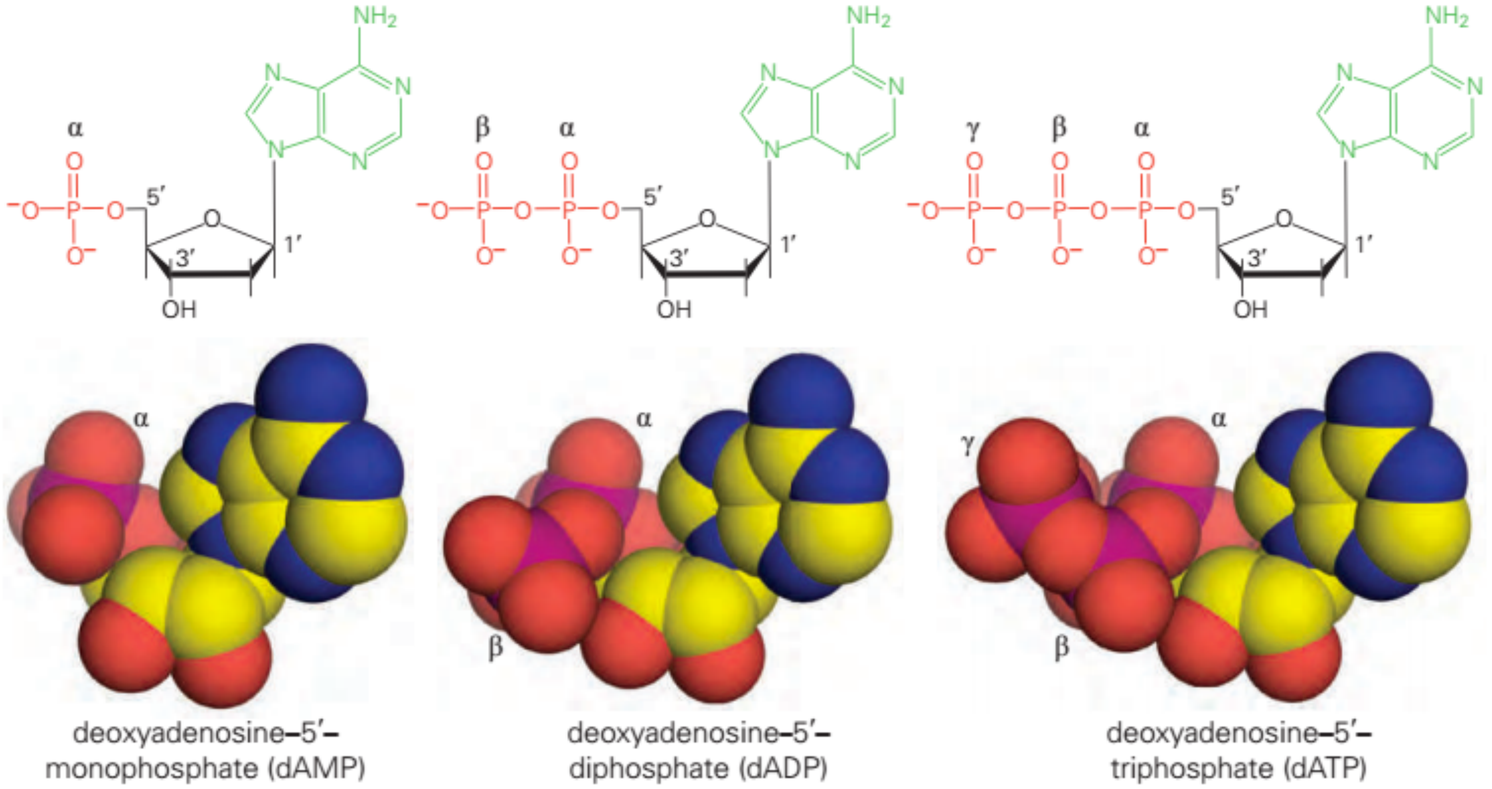
-Notice the linkages between the groups

Nucleic Acids – The phosphate



-Nucleotides with one, two or three phosphate groups are referred to as nucleotide mono, di or triphosphate.

-The three phosphate groups are called alpha, beta and gamma



Nucleic Acids are Polymers

-Nucleotides are joined together in DNA and RNA by the formation of a phosphodiester linkage between the 3’ carbon of one nucleotide and the 5’ of another

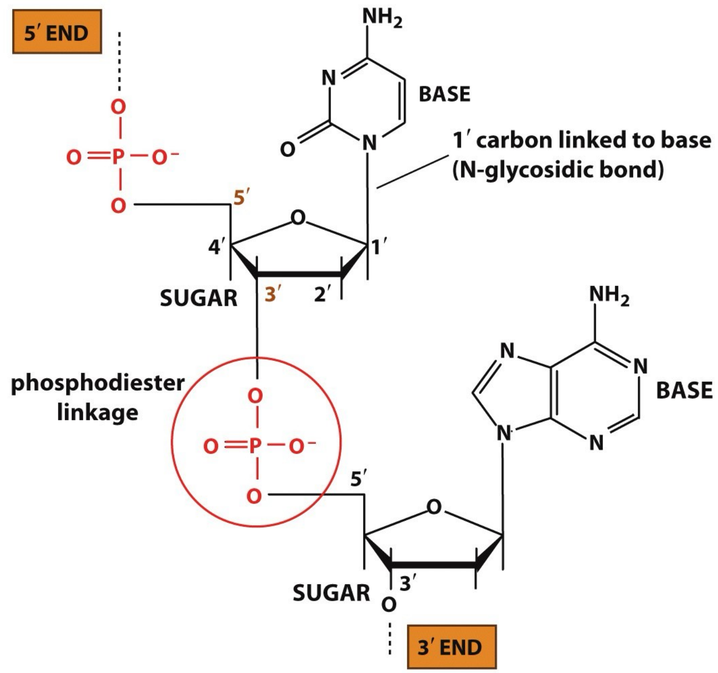


Figure 1.20 The Molecules of Life (© Garland Science 2013)

-The phosphate groups are negatively charged (anion nature)
– important determinant for the 3D structure of DNA and RNA

Nucleic Acids are Polymers

-The synthesis of new molecules of DNA and RNA involves the stepwise addition of nucleotide to one end of the chain.

-The triphosphate group is high in energy and its hydrolysis drives the reaction

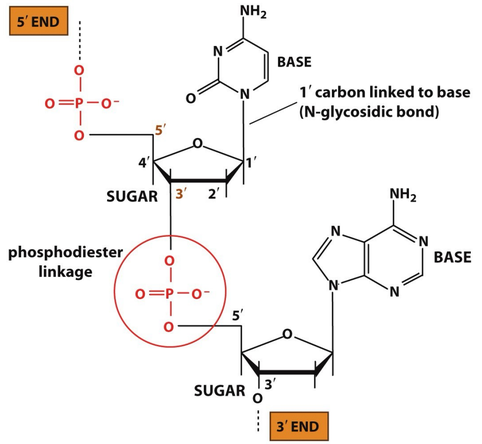


Figure 1.20 The Molecules of Life (© Garland Science 2013)

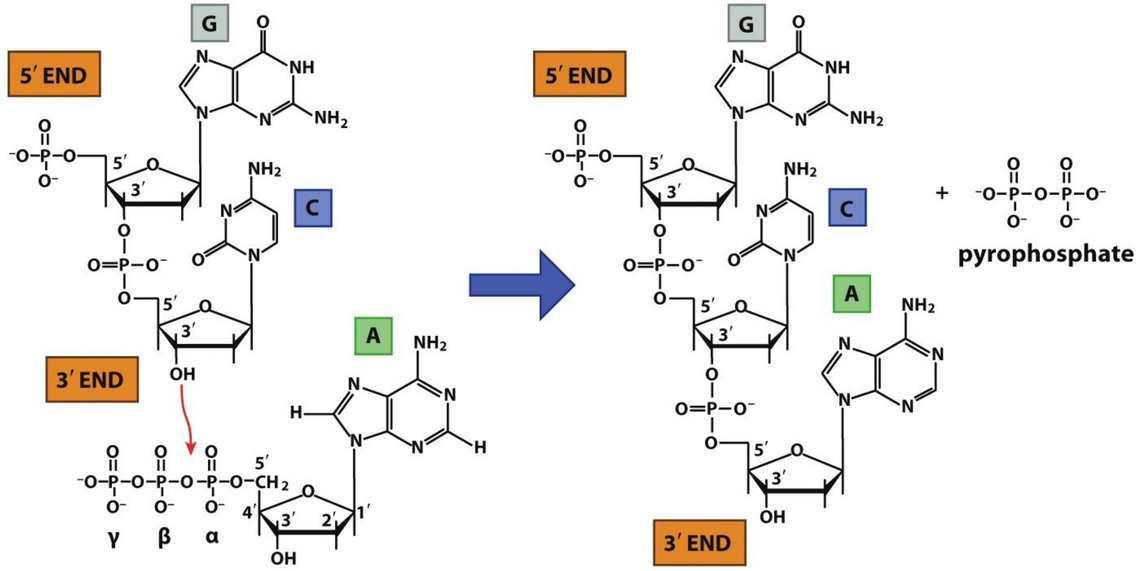


Figure 1.21 The Molecules of Life (© Garland Science 2013)

Nucleic Acids are Polymers

-The synthesis of new molecules of DNA and RNA involves the stepwise addition of nucleotide to one end of the chain.

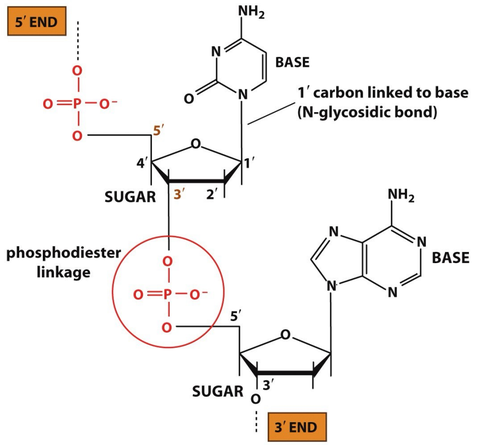


Figure 1.20 The Molecules of Life (© Garland Science 2013)

-The triphosphate group is high in energy and its hydrolysis drives the reaction

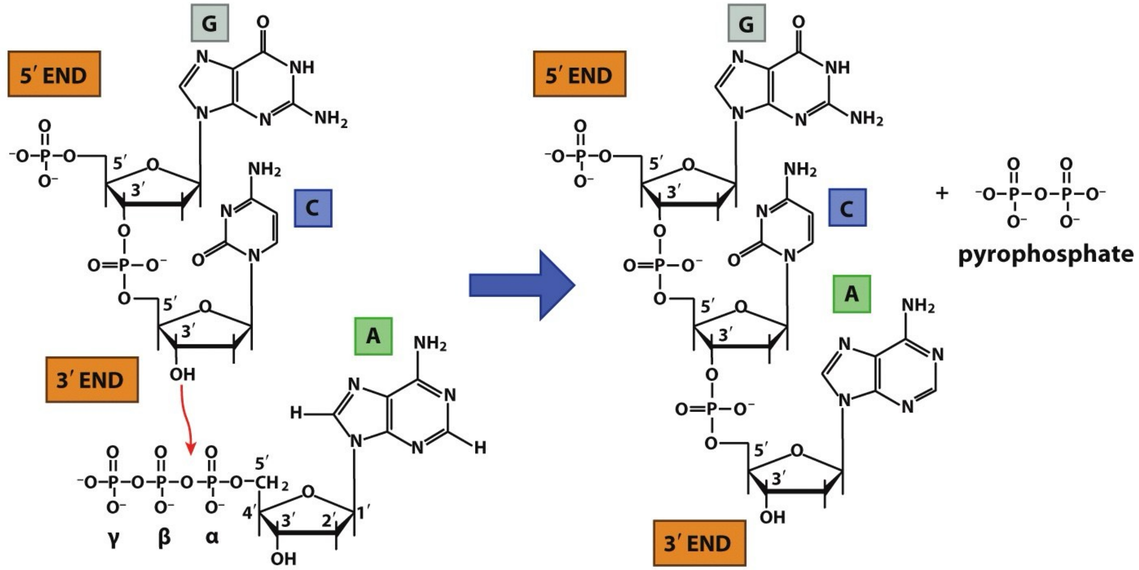


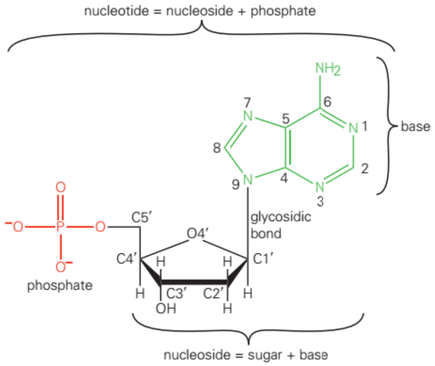
Figure 1.21 The Molecules of Life (© Garland Science 2013)

-DNA and RNA synthesis are template directed – DNA polymerases use a template strand to select each nucleotide to be added to the growing chain

-3’->5’ phosphodiester linkage imposes directionality

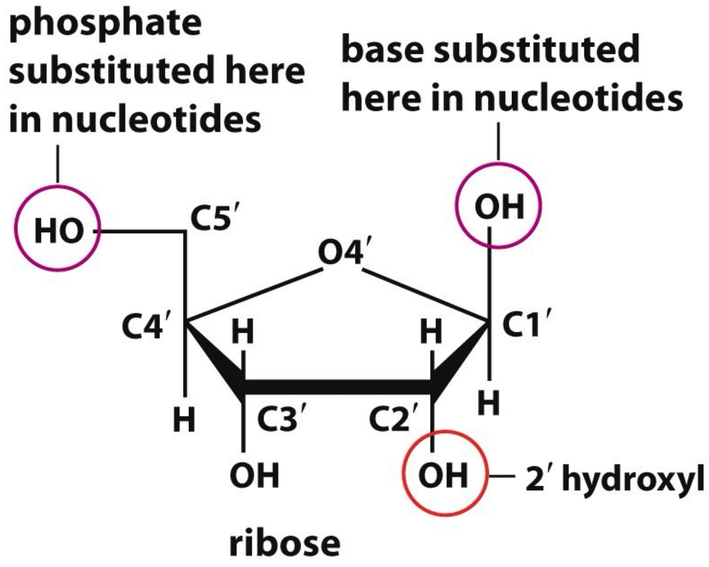
-By convention DNA sequences are written from 5’ to the 3’ end

Nucleic Acids – The pentose

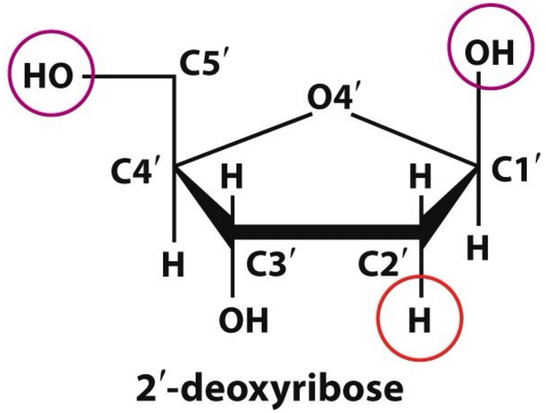


-Sugars used in RNA are derived from ribose.

-Sugars used in DNA are derived from 2’-deoxyribose



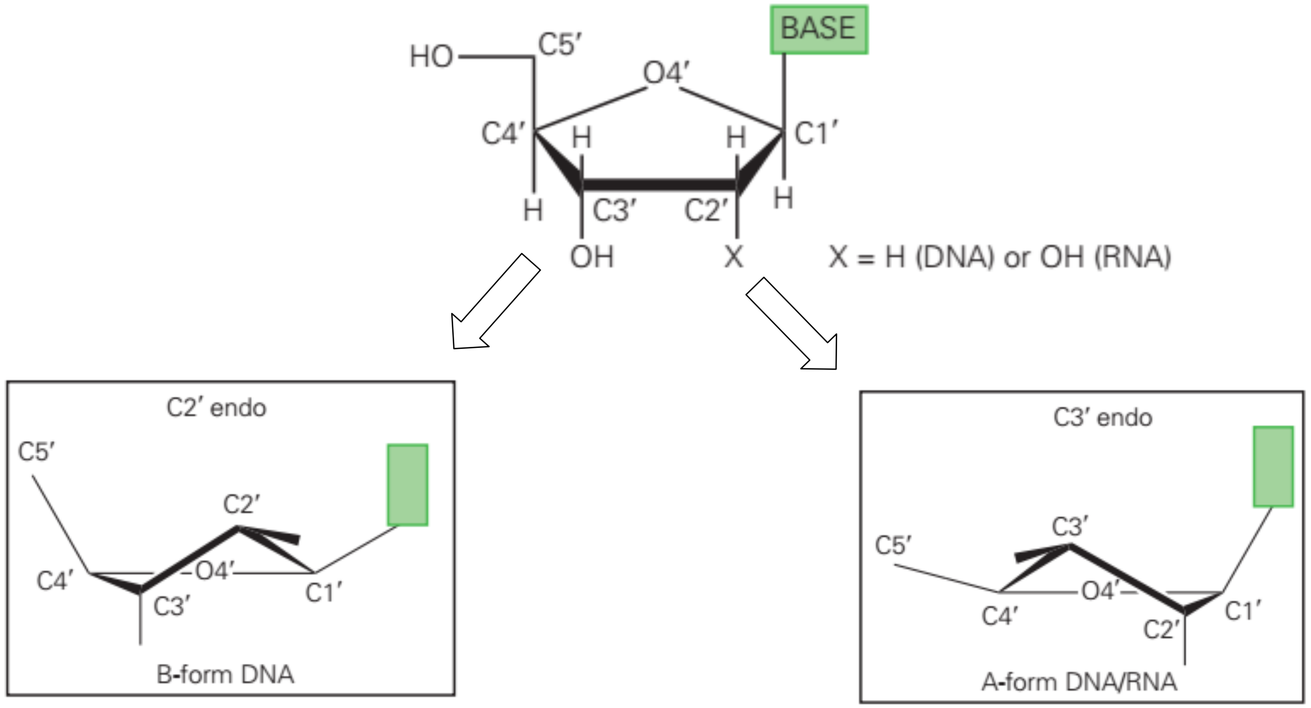
ribonucleotides



2’-deoxyribonucleotides

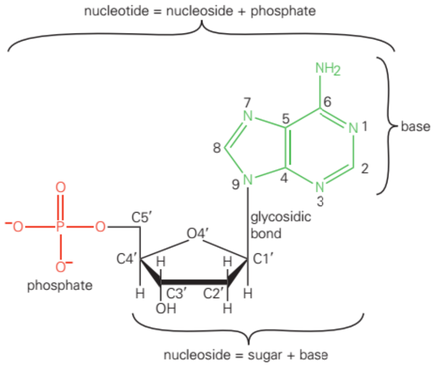
Nucleic Acids – The pentose

-In DNA/RNA molecules the pentose adopts a so-called sugar pucker conformation



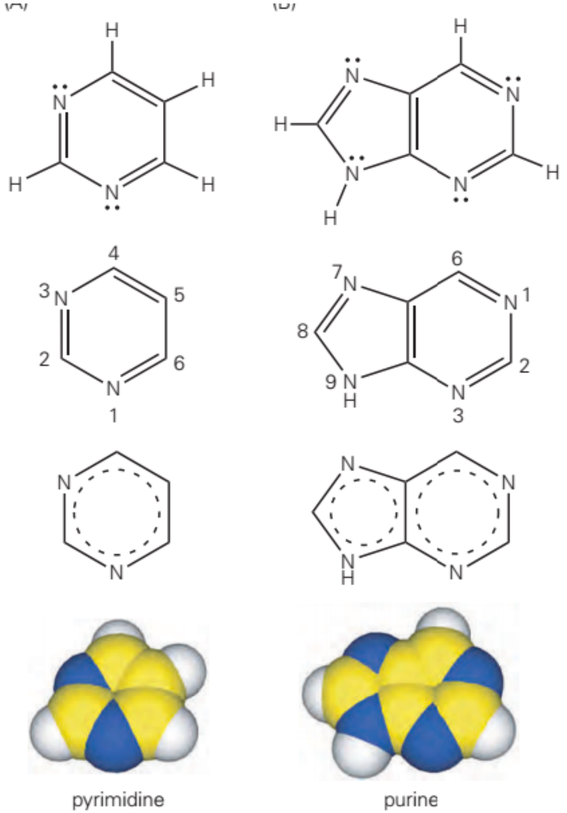
-In energetically favorable conformations four of the atoms of the pentose ring are roughly coplanar and one is out of the plane

Nucleic Acids – The base



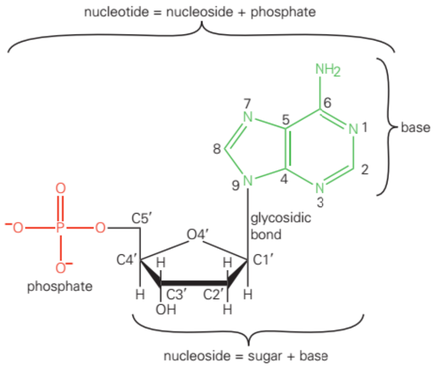
-DNA and RNA are built with 5 different bases

-The name “base” comes from its chemical composition – the ring systems contain lone pairs of electrons in the nitrogens being able to act as electron pair donors – so called Lewis bases



Nucleotide bases in RNA and DNA are substituted forms of two heterocyclic molecules known as **pyrimidine** and **purine**

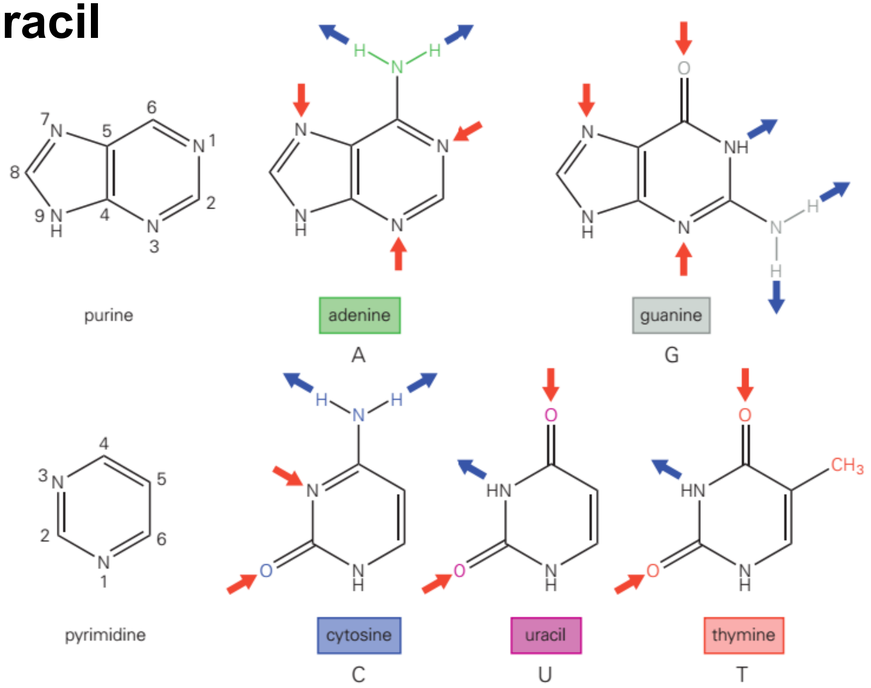
Nucleic Acids – The base



-DNA contains two substituted purines (**adenine** and **guanine**)

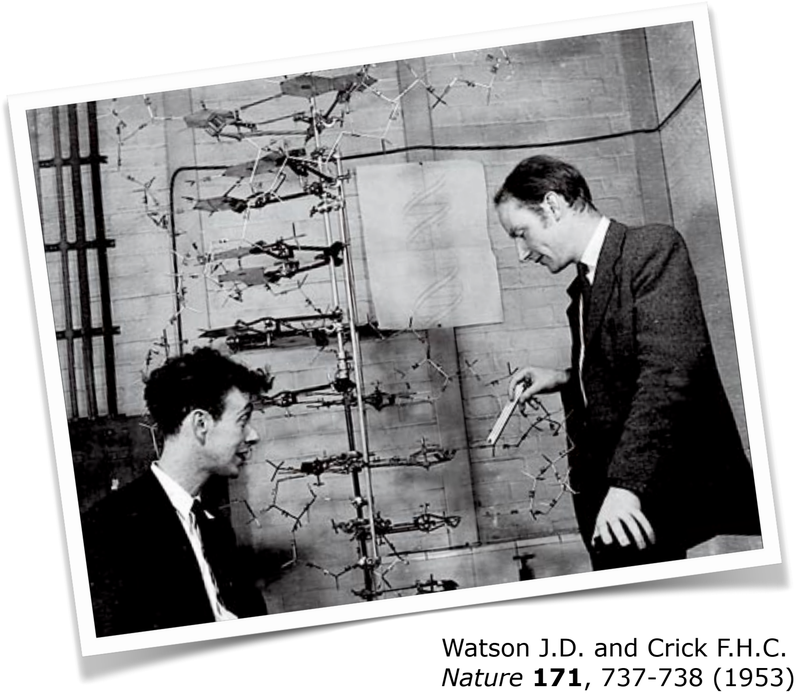
-DNA contains two substituted pyrimidines (**cytosine** and **thymine**)

-In RNA **thymine** is replaced by **uracil**

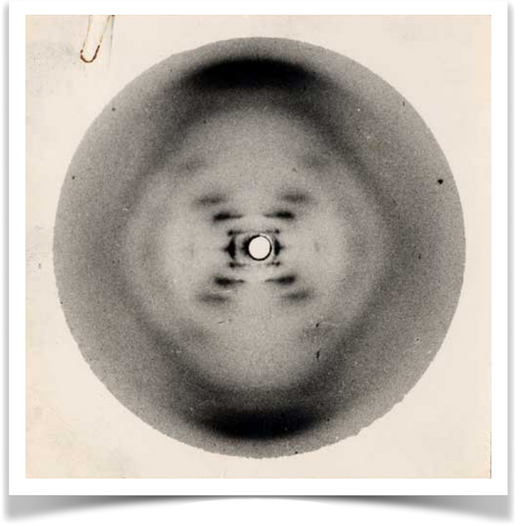


-Blue arrows point to hydrogen bond donor groups
-Red arrows point to hydrogen bond acceptor groups

Key for the ability of RNA and DNA to serve as templates for the transfer of genetic information



Watson J.D. and Crick F.H.C.
Nature **171**, 737-738 (1953)



Rosalind Franklin's X-ray image of DNA

historical detour

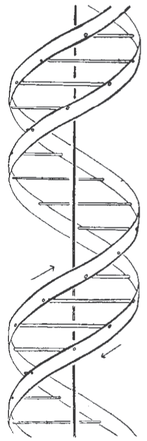
MOLECULAR STRUCTURE OF NUCLEIC ACIDS

A Structure for Deoxyribose Nucleic Acid

WE wish to suggest a structure for the salt of deoxyribose nucleic acid (D.N.A.). This structure has novel features which are of considerable biological interest.

A structure for nucleic acid has already been proposed by Pauling and Corey[1]. They kindly made their manuscript available to us in advance of publication. Their model consists of three inter-twined chains, with the phosphates near the fibre axis, and the bases on the outside. In our opinion, this structure is unsatisfactory for two reasons: (1) We believe that the material which gives the X-ray diagrams is the salt, not the free acid. Without the acidic hydrogen atoms it is not clear what forces would hold the structure together, especially as the negatively charged phosphates near the axis will repel each other. (2) Some of the van der Waals distances appear to be too small.

Another three-chain structure has also been suggested by Fraser (in the press). In his model the phosphates are on the outside and the bases on the inside, linked together by hydrogen bonds. This structure as described is rather ill-defined, and for this reason we shall not comment on it.



This figure is purely diagrammatic. The two ribbons symbolize the two phosphate–sugar chains, and the horizontal rods the pairs of bases holding the chains together. The vertical line marks the fibre axis

We wish to put forward a radically different structure for the salt of deoxyribose nucleic acid. This structure has two helical chains each coiled round the same axis (see diagram). We have made the usual chemical assumptions, namely, that each chain consists of phosphate di-ester groups joining β-D-deoxy-ribofuranose residues with 3′,5′ linkages. The two chains (but not their bases) are related by a dyad perpendicular to the fibre axis. Both chains follow right-handed helices, but owing to the dyad the sequences of the atoms in the two chains run in opposite directions. Each chain loosely resembles Furberg's[2] model No. 1; that is, the bases are on the inside of the helix and the phosphates on the outside. The configuration of the sugar and the atoms near it is close to Furberg's 'standard configuration', the sugar being roughly perpendicular to the attached base. There is a residue on each chain every 3·4 A. in the *z*-direction. We have assumed an angle of 36° between adjacent residues in the same chain, so that the structure repeats after 10 residues on each chain, that is, after 34 A. The distance of a phosphorus atom from the fibre axis is 10 A. As the phosphates are on the outside, cations have easy access to them.

The structure is an open one, and its water content is rather high. At lower water contents we would expect the bases to tilt so that the structure could become more compact.

The novel feature of the structure is the manner in which the two chains are held together by the purine and pyrimidine bases. The planes of the bases are perpendicular to the fibre axis. They are joined together in pairs, a single base from one chain being hydrogen-bonded to a single base from the other chain, so that the two lie side by side with identical *z*-co-ordinates. One of the pair must be a purine and the other a pyrimidine for bonding to occur. The hydrogen bonds are made as follows: purine position 1 to pyrimidine position 1; purine position 6 to pyrimidine position 6.

If it is assumed that the bases only occur in the structure in the most plausible tautomeric forms (that is, with the keto rather than the enol configurations) it is found that only specific pairs of bases can bond together. These pairs are: adenine (purine) with thymine (pyrimidine), and guanine (purine) with cytosine (pyrimidine).

In other words, if an adenine forms one member of a pair, on either chain, then on these assumptions the other member must be thymine; similarly for guanine and cytosine. The sequence of bases on a single chain does not appear to be restricted in any way. However, if only specific pairs of bases can be formed, it follows that if the sequence of bases on one chain is given, then the sequence on the other chain is automatically determined.

It has been found experimentally[3,4] that the ratio of the amounts of adenine to thymine, and the ratio of guanine to cytosine, are always very close to unity for deoxyribose nucleic acid.

It is probably impossible to build this structure with a ribose sugar in place of the deoxyribose, as the extra oxygen atom would make too close a van der Waals contact.

The previously published X-ray data[5,6] on deoxy-ribose nucleic acid are insufficient for a rigorous test of our structure. So far as we can tell, it is roughly compatible with the experimental data, but it must be regarded as unproved until it has been checked against more exact results. Some of these are given in the following communications. We were not aware of the details of the results presented there when we devised our structure, which rests mainly though not entirely on published experimental data and stereo-chemical arguments.

It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material.

Full details of the structure, including the conditions assumed in building it, together with a set of co-ordinates for the atoms, will be published elsewhere.

We are much indebted to Dr. Jerry Donohue for constant advice and criticism, especially on inter-atomic distances. We have also been stimulated by a knowledge of the general nature of the unpublished experimental results and ideas of Dr. M. H. F. Wilkins, Dr. R. E. Franklin and their co-workers at

James Watson Explains DNA Basepairing

Reprinted from Nature, Vol. 287, No. 5784, pp. 755-758, October 23 1980

Crystal structure analysis of a complete turn of B-DNA

Richard Wing*, Horace Drew, Tsunehiro Takano, Chris Broka, Shoji Tanaka, Keiichi Itakura† & Richard E. Dickerson

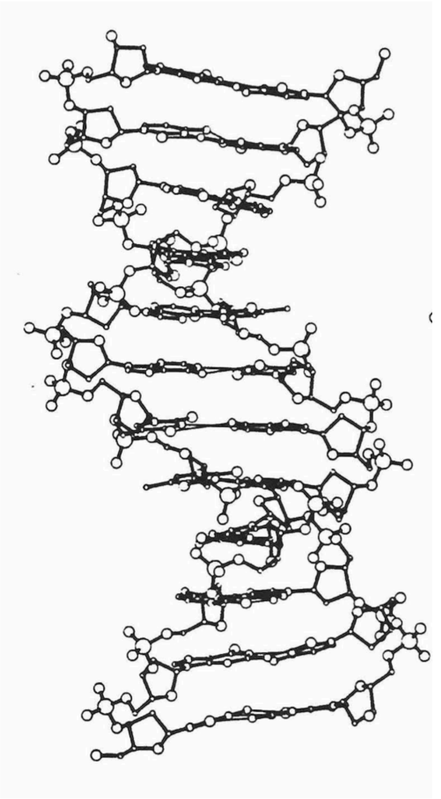
Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, California 91125

DNA is probably the most discussed and least observed of all biological macromolecules. Although its role in biology is a central one, with many examples such as operators and restriction sites where specific base sequences have control functions or interact with specific enzymes, the structures that DNA can adopt have been based until now only on sequence-averaged fibre diffraction patterns. Recent improvements in triester synthesis methods have made possible the preparation of sufficient homogeneous DNA of predetermined sequence for crystallization and X-ray structure analysis. We report here the first single-crystal structure analysis of more than a complete turn of right-handed B-DNA, with the self-complementary dodecamer sequence d(CpGpCpGpApApTpTpCpGpCpG) or CGCGAATTCGCG.

helix axis. Intensities remain strong in all directions out to 2.9 Å, and then exhibit a rapid decline until essentially no data can be obtained beyond 1.9 Å. Of the 5,691 possible reflections to 1.9 Å resolution, 2,818 were found to have an intensity greater than 2σ and were used in the analysis. Two isomorphous heavy atom derivatives were used: *cis*-dichlorodiamino platinum (II) obtained by diffusion, and a 3-Br derivative obtained by *de novo* synthesis of the dodecamer with 5-bromocytosine in the third position along each chain. The 1-Br derivative was crystallized but proved not to be isomorphous, and the 9-Br derivative was synthesized but not needed. Isomorphism in the *cis*-Pt derivative began to fail beyond 4-Å resolution, but the 3-Br derivative remains isomorphous to 2.7 Å.

The present report describes the partially refined structure obtained from multiple isomorphous replacement (MIR) analysis at 2.7 Å (mean figure of merit 57%), followed by Jack–Levitt refinement procedures[3] using 2,725 2σ intensities between 8.0 and 1.9 Å. The current residual error or R factor is 24.8% for a DNA molecule of 486 atoms and 9 initial water molecules. The structure of the DNA itself is essentially correct and is reported now because of its general interest. Refinement will continue with the addition of more solvent and spermine atoms, and some improvement in local nucleotide conformations.

A skeletal drawing of CGCGAATTCGCG is presented in Fig. 1, and a space-filling version from the same orientation in



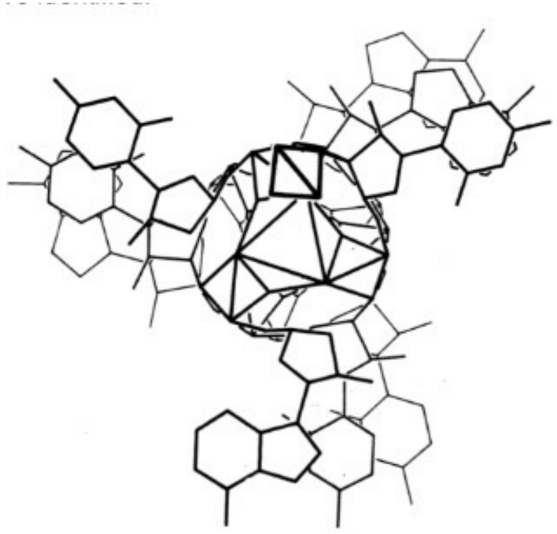


Figure 6, reproduced from: A proposed structure for the nucleic acids. Pauling and Cory (1953) PNAS 39, 84-97.

historical detour

Nucleic Acids - 3D structure

-DNA forms a double helix with antiparallel strands

-Two strands together wind up to form a right-handed double-helix

-Bases are on the inside of the helix and the phosphate backbone group are on the outside. Allowing for interactions with ions and water and minimizing repulsion between phosphates

- Base pairing holds the DNA strands together and is strictly complementary

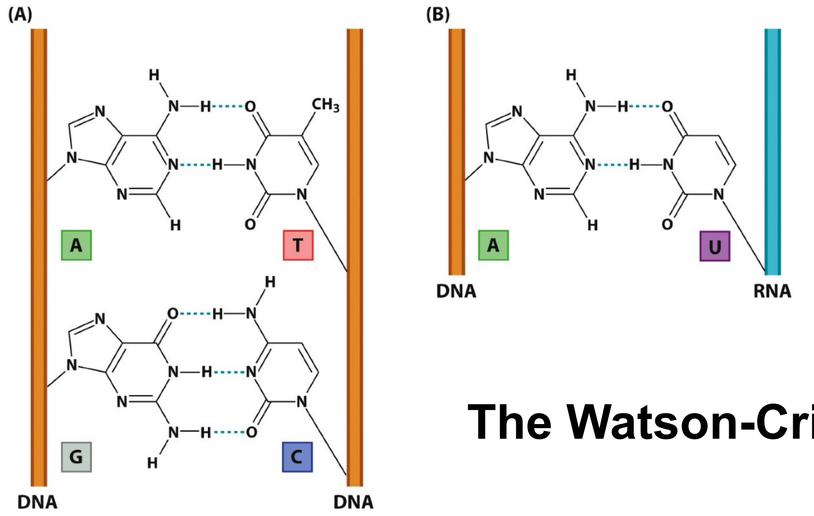


Figure 1.24 The Molecules of Life (© Garland Science 2013)

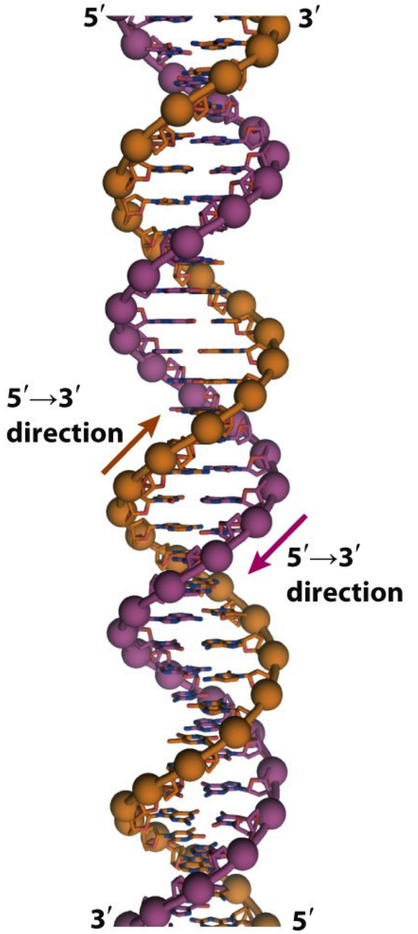


Figure 1.23 The Molecules of Life (© Garland Science 2

Phosphate groups
in spheres

The Watson-Crick base pairs: A-T, G-C and A-U

Discussion : DNA 3D structure

Associate the different molecular interactions discussed above to the DNA structure.

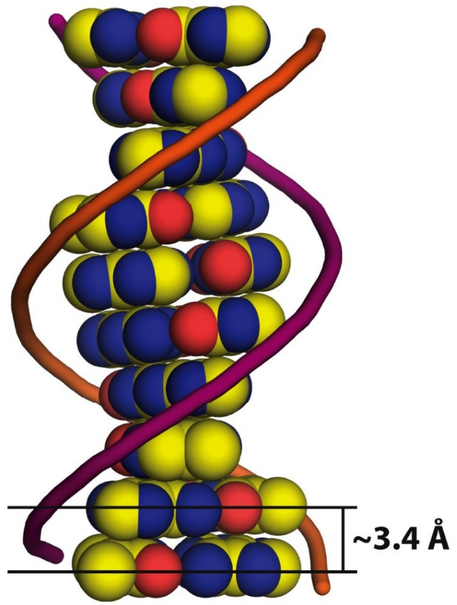
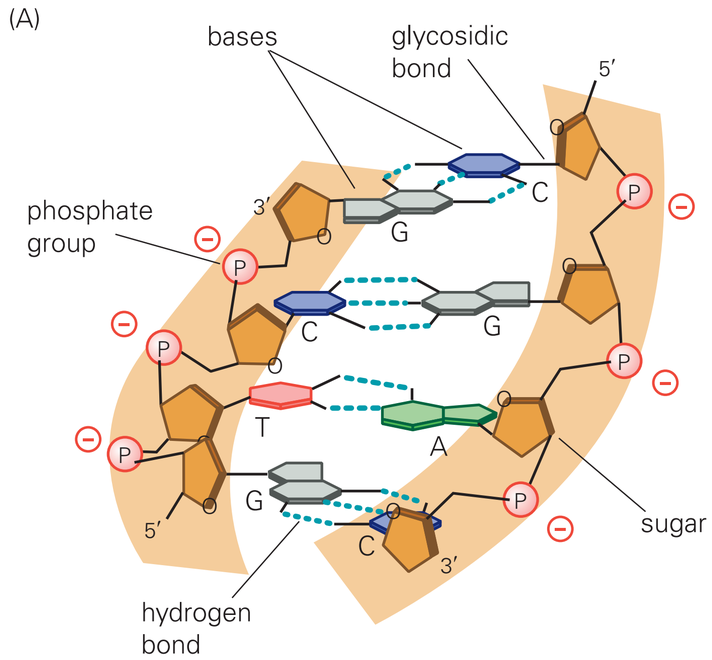
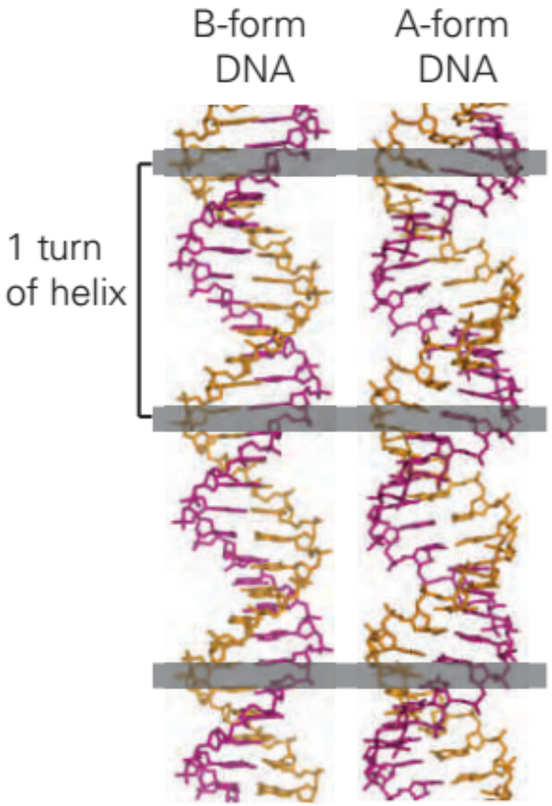


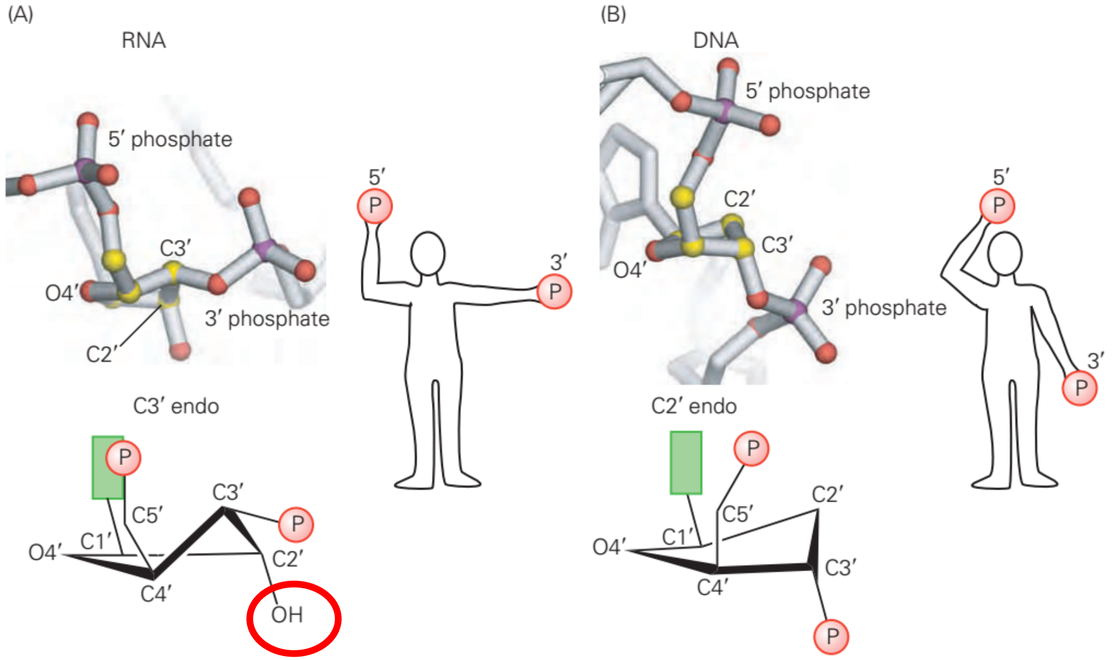
Figure 1.27 The Molecules of Life (© Garland Science 2013)



Nucleic Acids - 3D structure

-The difference between the A and B forms arise from the sugar pucker

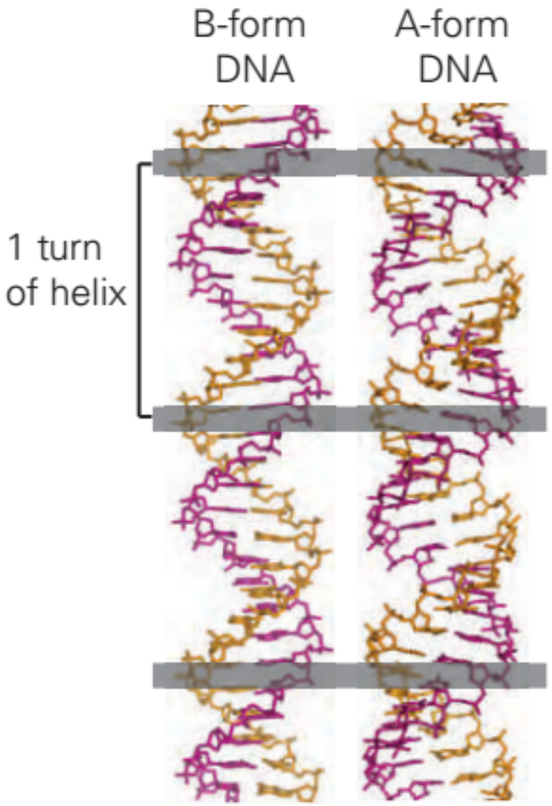


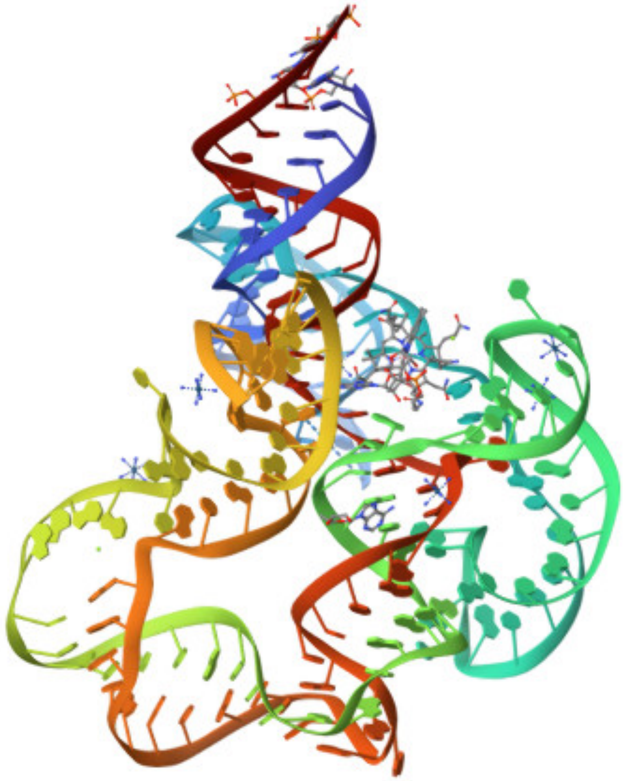


-For RNA the hydroxyl group changes the conformation of the nucleotide conformation with larges impact in the overall structure

Nucleic Acids - 3D structure

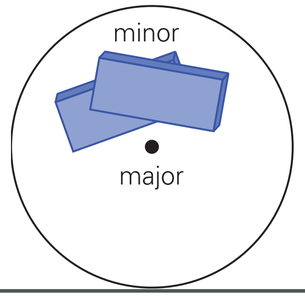
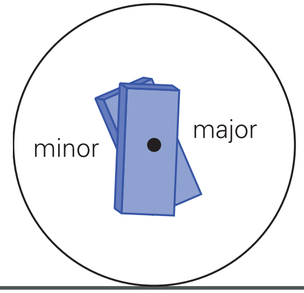
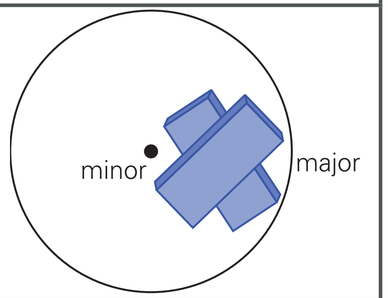
-The difference between the A and B forms arise from the sugar pucker





-For RNA the hydroxyl group changes the conformation of the nucleotide conformation with larges impact in the overall structure

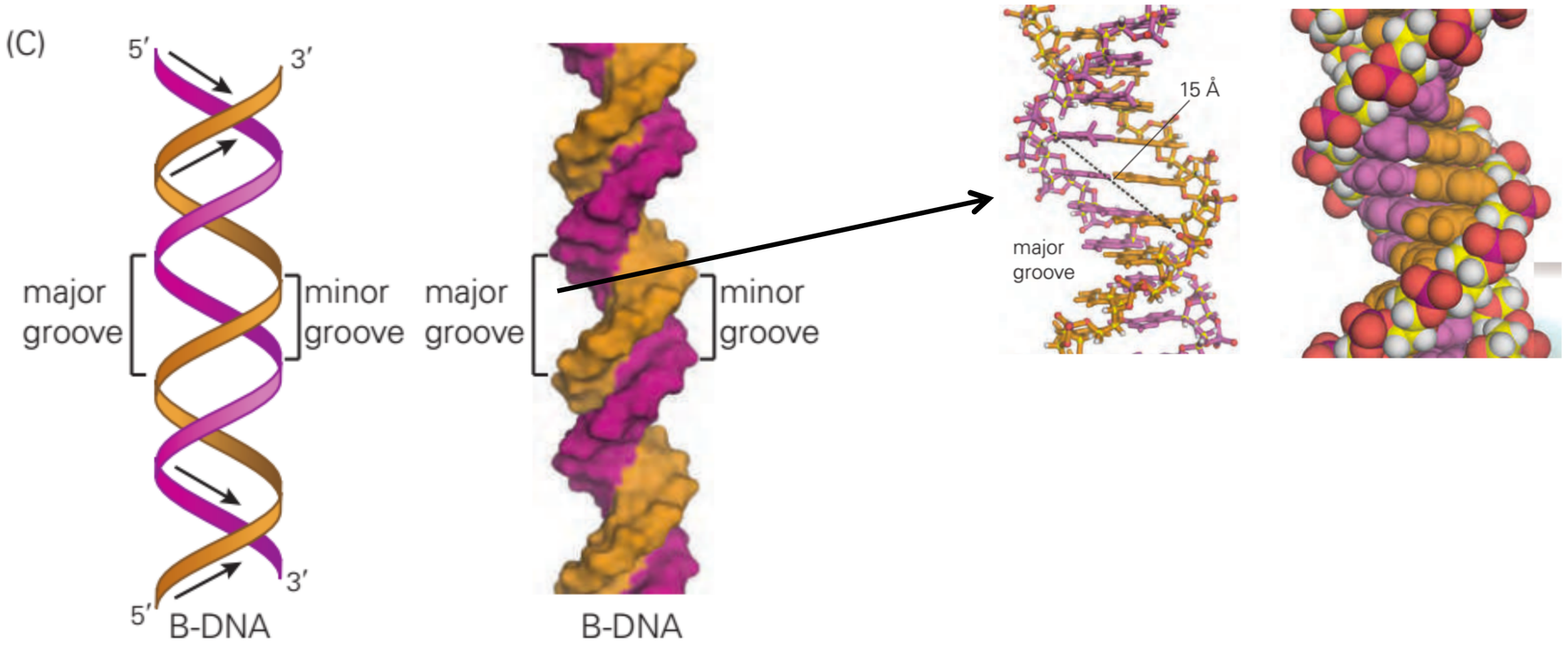
Table 2.1 Structural features of A-, B-, and Z-form helices.

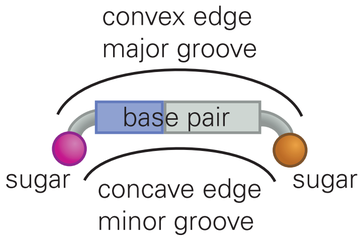
Helical form	A	B	Z
Helical sense	Right	Right	Left
Diameter	~ 26 Å	~ 20 Å	~ 18 Å
Base pairs per turn	~ 11	~ 10	~ 12
Helical twist (rotation per base pair for A and B, per two-base repeat for Z)	~ 34°	~ 36°	~ 60° (CpGp)
Helix pitch (rise per helical turn)	~ 25 Å	~ 33 Å	~ 46 Å
Helix rise (along helix axis; per base pair for A and B, per two-base repeat for Z)	~ 2.3 Å	~ 3.3 Å	~ 7.4 Å (CpGp)
Base tilt (with respect to helix axis)	~ 20°	~ 0°	~ – 9°
Base orientation (with respect to sugar)	Anti	Anti	C anti/G syn
Base pair positions (helix axis indicated by black dot)			
Features of base pair positions	Base pairs displaced from axis; deep major groove, less accessible	Base pairs on axis; both major and minor grooves accessible	Base pairs stick out into the major groove, the minor groove is deep and narrow

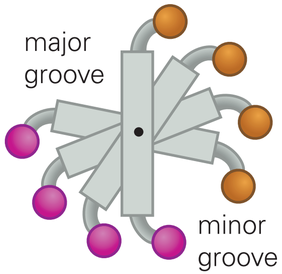
(Adapted from R.E. Dickerson et al., and M.L. Kopka, *Science* 216: 475–482, 1982. With permission from AAAS.)

Nucleic Acids - 3D structure

-One of the most important features of DNA double helices are the grooves.
(we focus on the B-DNA only)

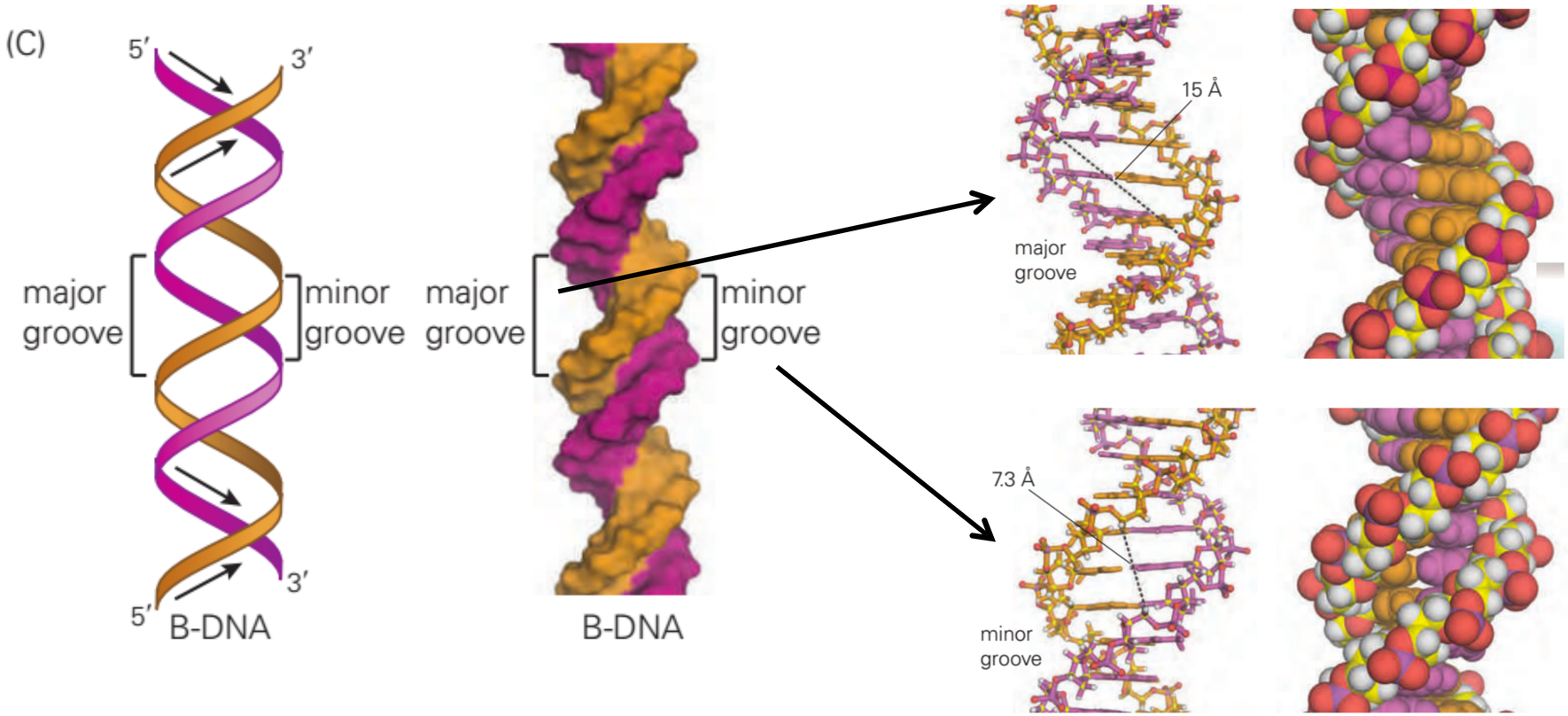






Nucleic Acids - 3D structure

-One of the most important features of DNA double helices are the grooves.



-The major and minor groove present very distinct structural features

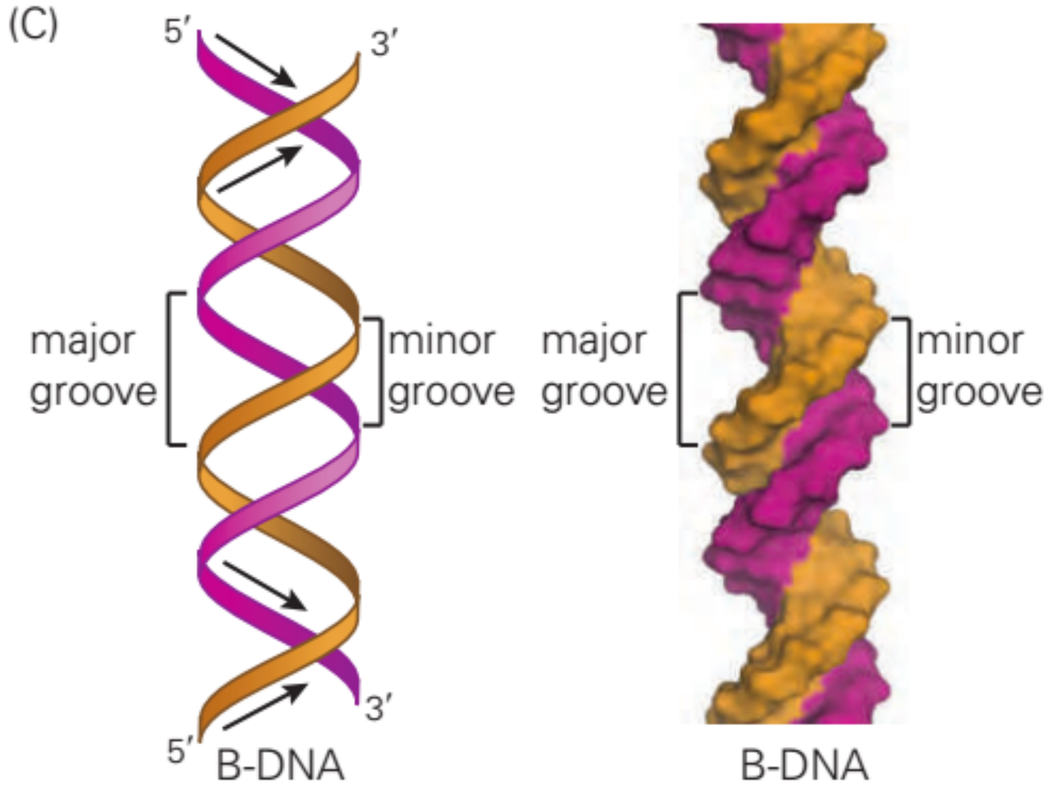
-Very important for the recognition of DNA by proteins.

Statement:

The major groove has a greater information content than the minor groove.

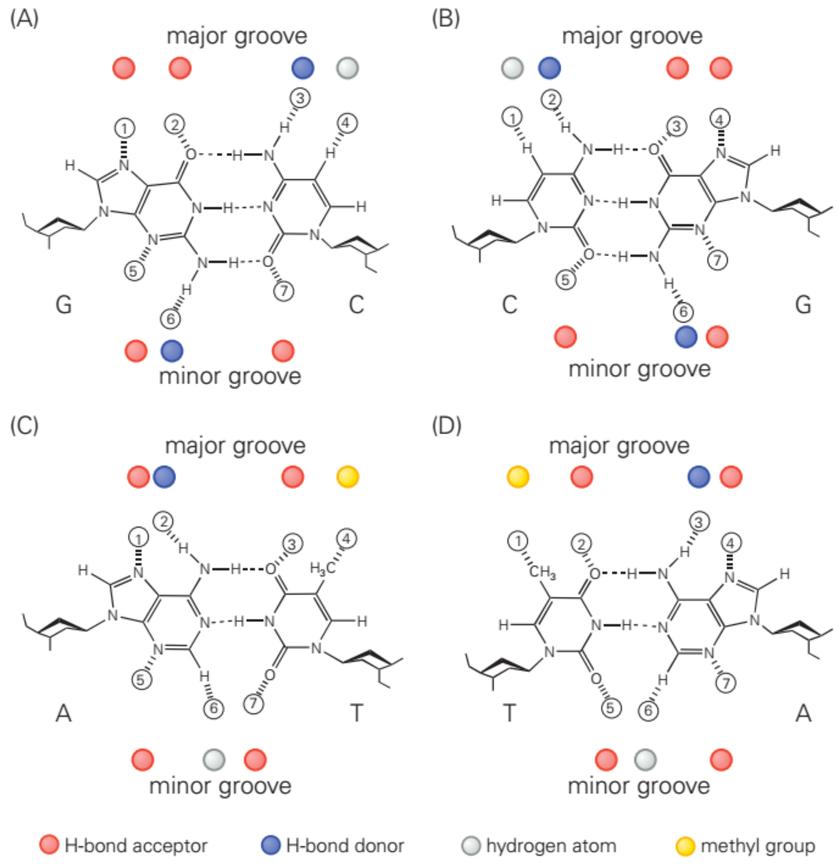
-Yes or No ?

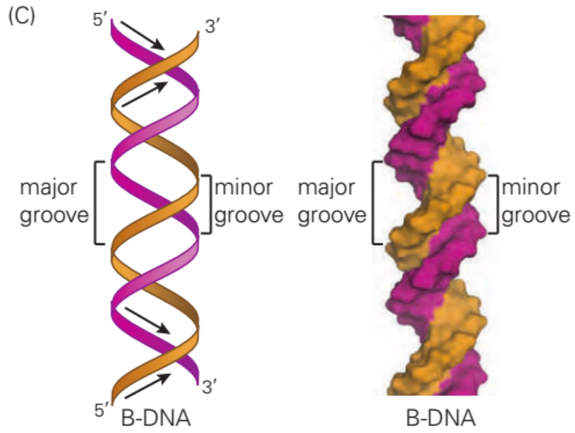
- Why ?

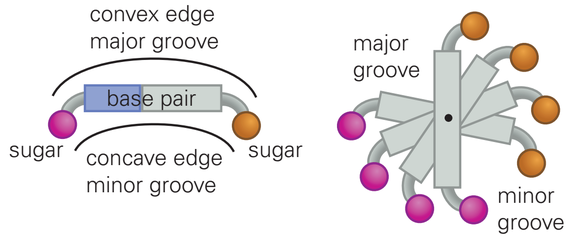


Nucleic Acids - 3D structure

-Potential interaction sites at the edges of Watson-crick base pairs





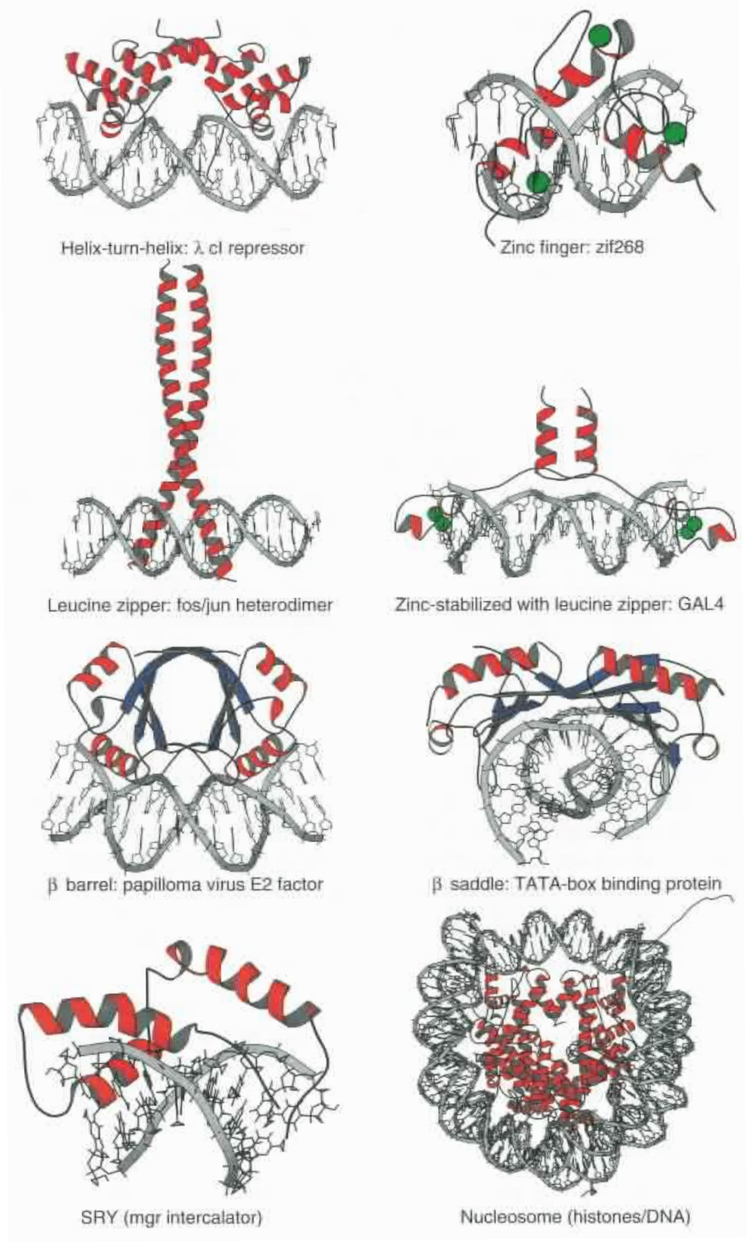


-Four type of interactions are possible.

-The major and minor grooves can be identified by looking at the connections of the base pairs with the sugars. Major groove on the convex edge and the minor groove in on the concave edge.

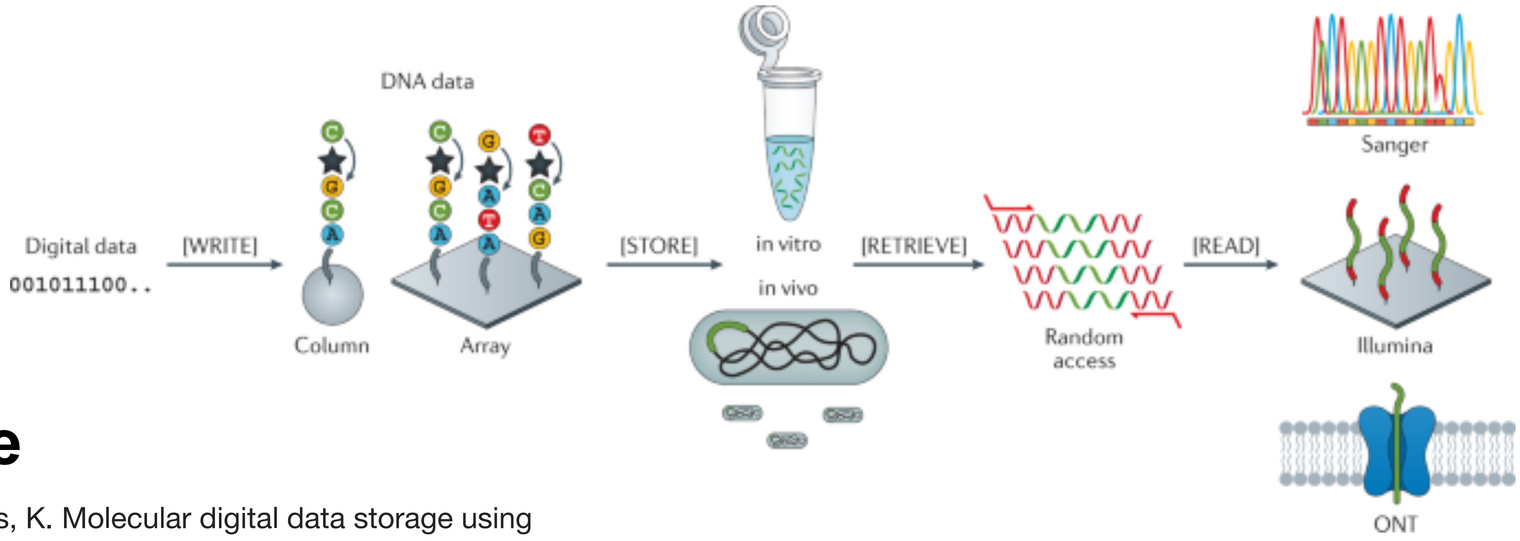
-Notice the chemical diversity of the major grooves vs the minor grooves.

DNA/protein interactions



Complex	Binding Motif[a]	Binding Groove[b]	Details of Complex
λ repressor	HTH	Mgr	Canonical HTH; homodimers; 2 helices of Cro dimer cradle Mgr, stabilized by direct H-bond and vdW contacts; little DNA distortion.
CAP repressor	HTH	Mgr	About 90° bend.
trp repressor	HTH	Mgr	Indirect, water-mediated base contacts.
Purine rep.	HTH	Mm	α-helices inserted in mgr.
Yeast MATα2	HTH	Mgr	Homeobox domains bind as monomers.
Zif268	Zn	Mgr	Zinc finger subfamily; each Zn finger recognizes 3 bps.
GATA-1	Zn	Mm	Transcription factors subfamily; single domain coordinated by 4 cysteines.
GAL4	Zn	Mgr	Metal binding subfamily; each of two Zn ions, coordinated by 6 cysteines, recognizes 3 bps.
GCN4	Leu/Zip	Mgr	Canonical; basic region/leucine zipper (α helices) motif; slight DNA bending.
fos/jun	Leu/Zip	Mgr	α-helices resemble GCN4; unstructured basic region folds upon DNA binding.
fos/jun/NFAT	Leu/Zip	Mgr	α-helices bend to interact with NFAT.
MetJ	β-ribbon	Mgr	Two anti-parallel β-strands in Mgr; bends each DNA end by 25°.
papillomavirus E2 DNA target	β-barrel	Mgr	Domed β-sheets form an 8-strand β-barrel dimer interface with 2 α-helices in Mgr; strong tailored fit for every base of the recognition element; bent DNA; compressed mgr; DNA target crystallized without protein.
TBP	β-saddle	mgr	Ten-β-strand saddle binds in Mgr; significant distortion, ≈ 90° bend.
p53 tumor supp.	Loop/other	Mm	Binds to DNA via protruding loop and helix anchored to anti-parallel β-barrel.
SRY	Loop/other	mgr	Isoleucine intercalated into mgr.
NFAT	Loop/other	Mm	Flexible binding loop stabilized by DNA.
histones	Loop/other	Mm	Nonspecific PO_4 interactions.
distamycin (drug)		mgr	Selective to AT bps; binds in mgr without distortion.

DNA in biotechnology

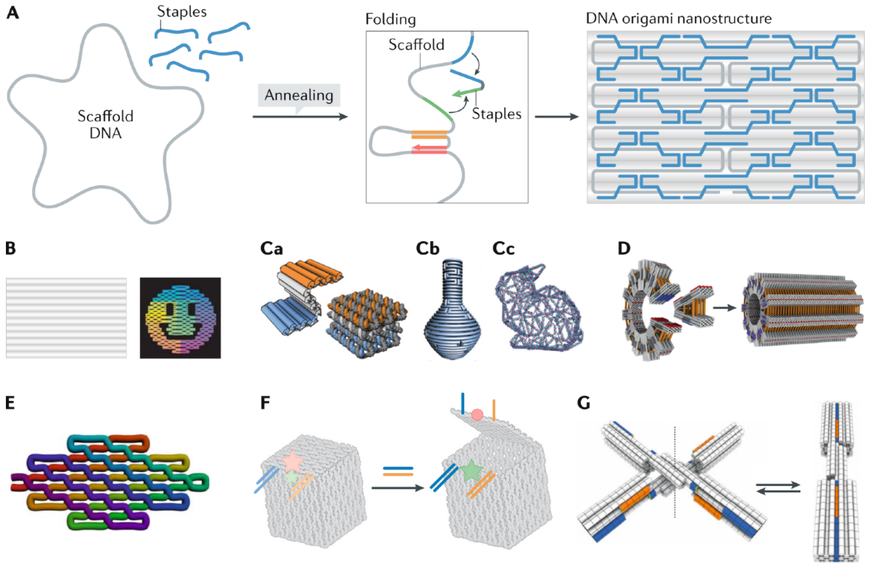


Data storage

Ceze, L., Nivala, J. & Strauss, K. Molecular digital data storage using DNA. *Nat Rev Genet* **20**, 456–466 (2019). https://doi.org/10.1038/s41576-019-0125-3

DNA origami

Dey, S., Fan, C., Gothelf, K.V. *et al.* DNA origami. *Nat Rev Methods Primers* **1**, 13 (2021). https://doi.org/10.1038/s43586-020-00009-8



Nucleic Acids – Take Home Messages

-DNA and RNA are the informational polymers in the cell – encode genetic information in a way that can be read by macromolecular machines, to direct the synthesis of other molecules.

-Nucleotides have pentose sugars attached to nitrogenous bases and phosphate groups.

-The nucleotide bases in DNA and RNA are substituted pyrimidines or purines.

-4 deoxyribonucleotides in DNA (A,T,G,C) and four ribonucleotides in RNA (A,U,C,G)

-DNA and RNA are synthesized in 5’ to 3’ direction by sequential reactions that are driven by hydrolysis of nucleotide triphosphates

-DNA forms a double helix with antiparallel strands

-Double helix involves complementary base pairing (A-T and C-G) and is stabilized by, hydrogen bonds, base pair stacking and electrostatic interactions

-B-form DNA allows sequence specific recognition of the major groove by proteins. Each base pair has a unique set of interacting elements in the major groove but not in the minor groove.

Questions ?????

- now
- Moodle forum
- @ matteo.dalperaro@epfl.ch
- all the TAs

Next week - Lecture 2

Lipids and glycans