Welcome to lecture series “Bioinformatics” !

Moodle / exam / exercises

- Lecture material at:
  https://moodle.uni-due.de/
  → Biologie
  → Bioinformatik (AG Hoffmann)
  → Bioinformatics for Computer Scientists, Chemists et al
- “Material” forum: handouts, exercises
- “Anonymous forum”: your questions (better than e-mail)
- Exam probably February 9, 2021, 14:30-16:30, either PC-hall A-009 (vis-à-vis Ikea) or online
- Computer exam: books, scripts, notes allowed as aiding material
- For computer scientists: programming project possible
- Exercises: 2 hrs every 2 weeks, 14:15-15:45; starting Thu, Nov 18

Literature

- Zvelebil & Baum: Understanding Bioinformatics, Garland Science (English)
- Merkl & Waack: Bioinformatik Interaktiv, Wiley-VCH (German, also as e-book)
- Selzer, Marhöfer & Rohwer: Applied Bioinformatics – An Introduction, Springer (German/English)
- Lesk: Introduction to Bioinformatics, 4th ed., Oxford University Press

What is “Bioinformatics”?

Attempt of a definition:

Bioinformatics is the science of the computational methods used to process biological data.

Response to mass of biomolecular data.

What is “Bioinformatics”?

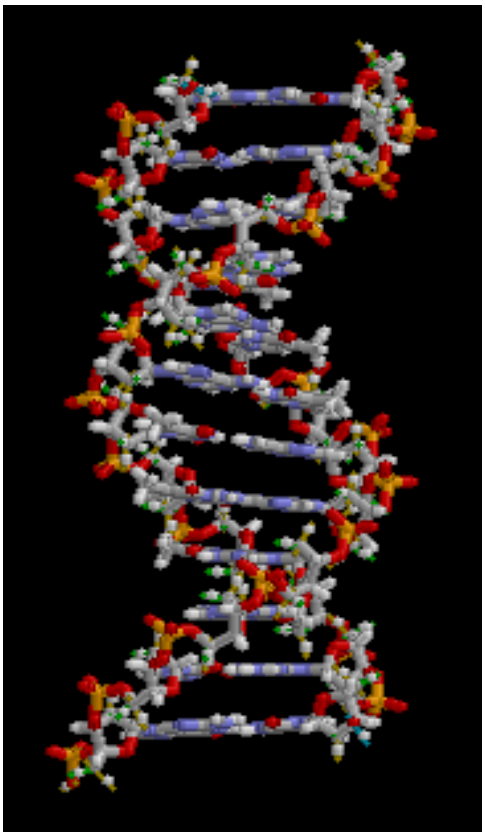
Attempt of a definition:

Bioinformatics is the science of the computational methods used to process biological data.

Response to mass of biomolecular data.

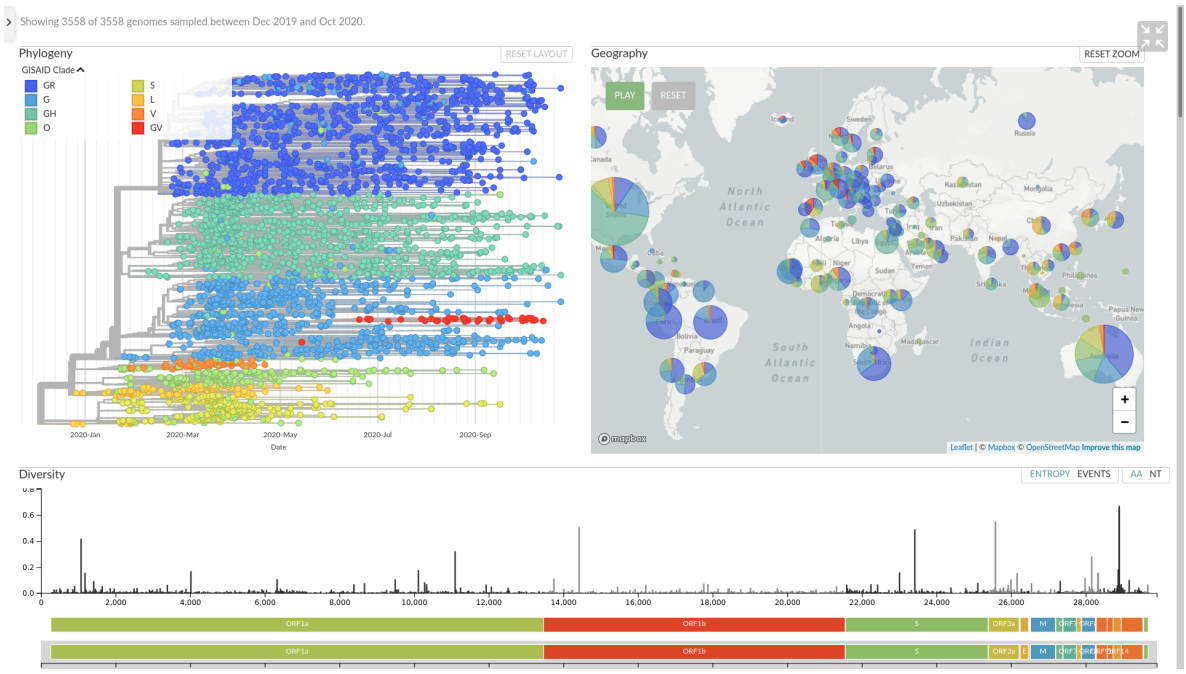
1. The boundaries of bioinformatics to allied sciences are not well defined. Such allied sciences are, for instance, statistics, mathematics, computer science, biology, computational chemistry, medicinal chemistry, ...

Bioinformatics: focus on molecules



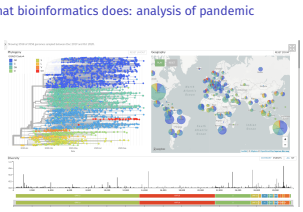
All life forms have a common language: molecules.

What bioinformatics does: analysis of pandemic



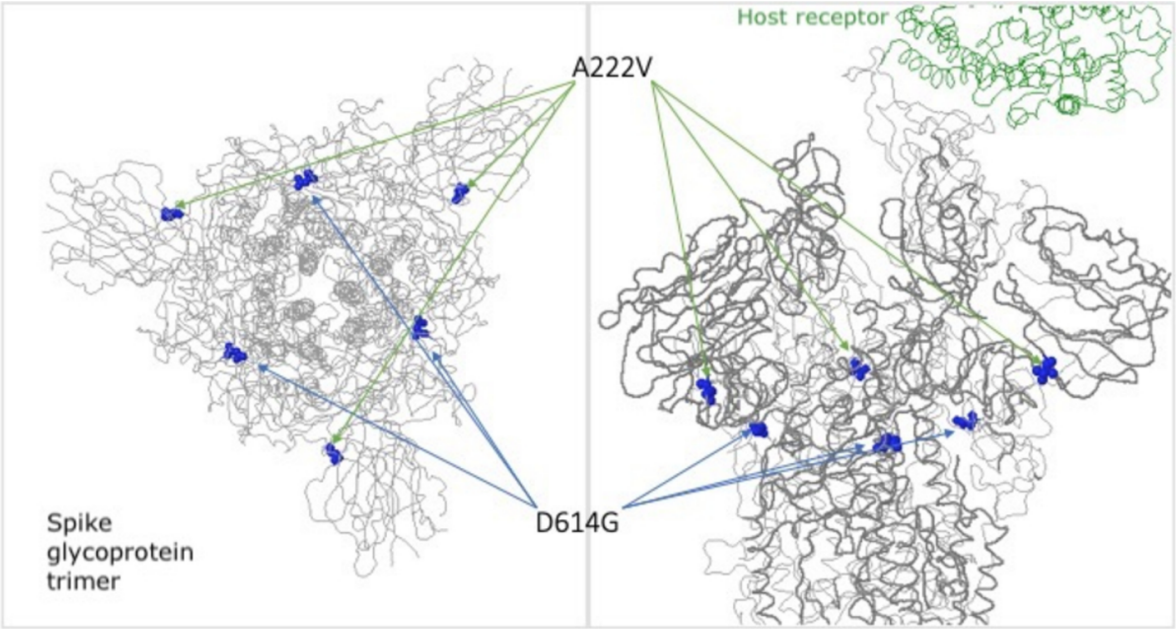
2020-11-03

└─What bioinformatics does: analysis of pandemic

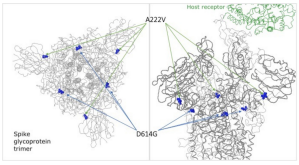


From `https://www.gisaid.org/`.

What bioinformatics does: identify new SARS-CoV-2 strains

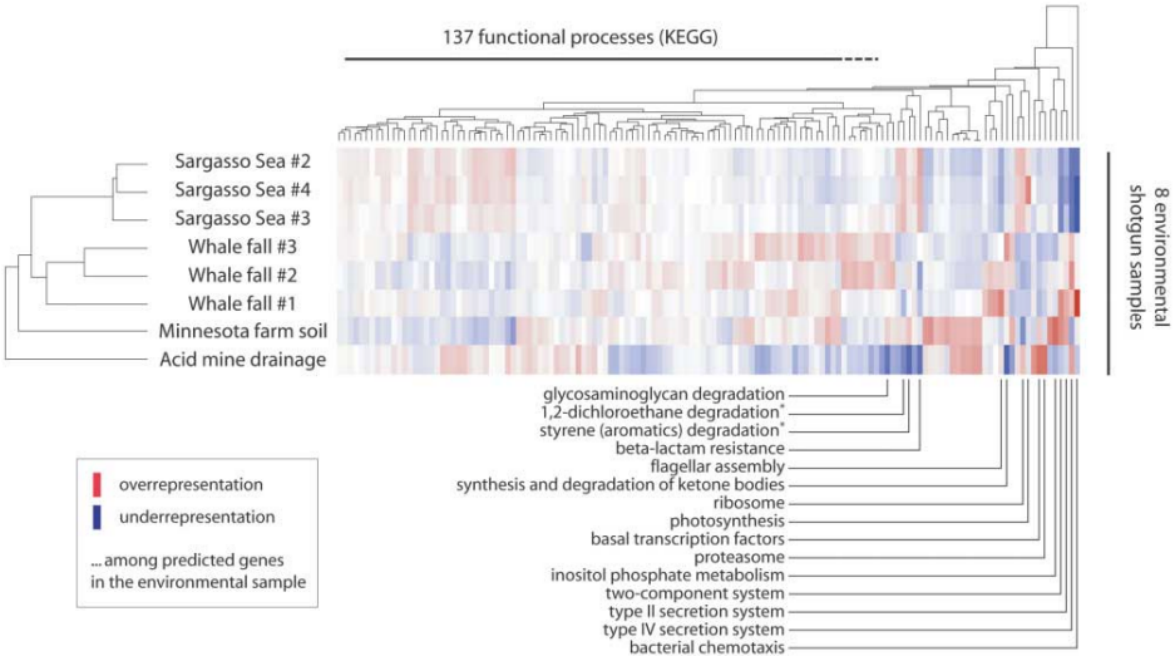


What bioinformatics does: identify new SARS-CoV-2 strains



From `https://www.gisaid.org/`.

What bioinformatics does: a molecular view of ecosystems



Topics of lecture series “Bioinformatics”

- Life: bio-essentials
- Modern methods in biology: sources of data
- Organisation of biological data
- **Usage of sequence data**

Life – a kind of definition

No generally accepted definition of life

Important to life in my view:

Complex, adaptive, non-equilibrium structures that replicate.

Complexity

Smallest unit of life:

cell

- Each cell: large number of specifically interacting molecules
- Organisms: large numbers of specifically interacting cells

Life "controlled" by a program –

the **genome**

A specific genome defines a **genotype**

("This organism has genotype X.")

Human genome: $\approx 6 \cdot 10^9 bit$

Smallest viruses: $\approx 10^3 bit$

Genome is not sufficient to **predict** development of organism, because ...

Adaptivity

... **interaction** with **environment** is also important:

Uptake/release of energy, matter, information (examples?)

⇒ Specific reaction, specific development of **individual**

Adaptivity of species:

evolution

Non-Equilibrium

Life is not at the equilibrium = not stable

All life forms ...

- start life (cell division, ...)
- live for some time (minutes to $10^3 yr$)
- divide or die

Death:

- loss of structure
- stop of replication
- (exact life span may be difficult to determine in practice)

Structure

Multiscale structure from >100 m to 10^{-9} m:

- societies, communities, eco-systems
- organisms
- organs, tissues, ...
- cells (see p. 36)
- cellular substructures (organelles, ...)
- molecular complexes (ribosomes, spliceosomes, ...; p. 37)
- macromolecules (DNA, RNA, proteins)
- small molecules (metabolites, building blocks, ...)

Structure is an aspect of the **phenotype** (set of features of organism)

Genotype and environment control phenotype

Replication

Life forms can **reproduce** ("replicate")!

Reproduction is usually not accurate, e.g. because of

- mutations (changes) of genome
- recombination between parent genomes ⇒
  new mixtures

⇒selection of adapted individuals ⇒

evolution

The building blocks and how they interact

Some key players in life processes:

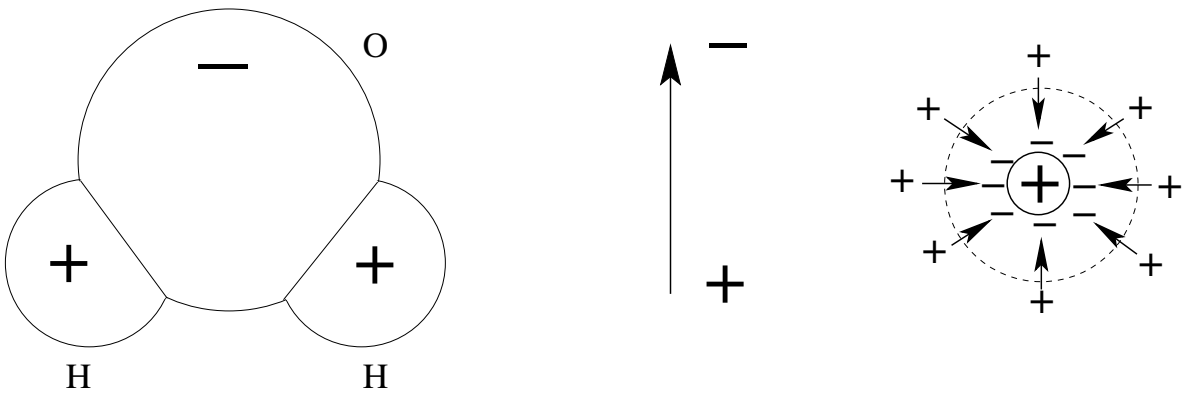
- water
- “small molecules”: amino acids, nucleotides, sugars, ...
- lipids
- DNA
- RNA
- proteins
- complexes, higher order structures

Water: polar medium

Medium associated with all life forms we know

On p.37 space between large molecules filled with water

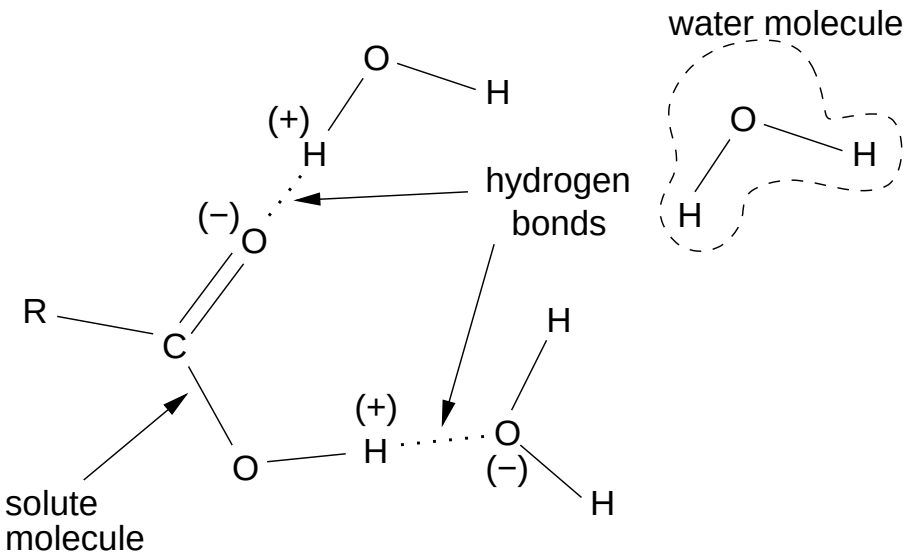
Polar: interacts favorably with charges



Charge–dipole interaction lowers potential (Coulomb) energy

Water: hydrogen bonds

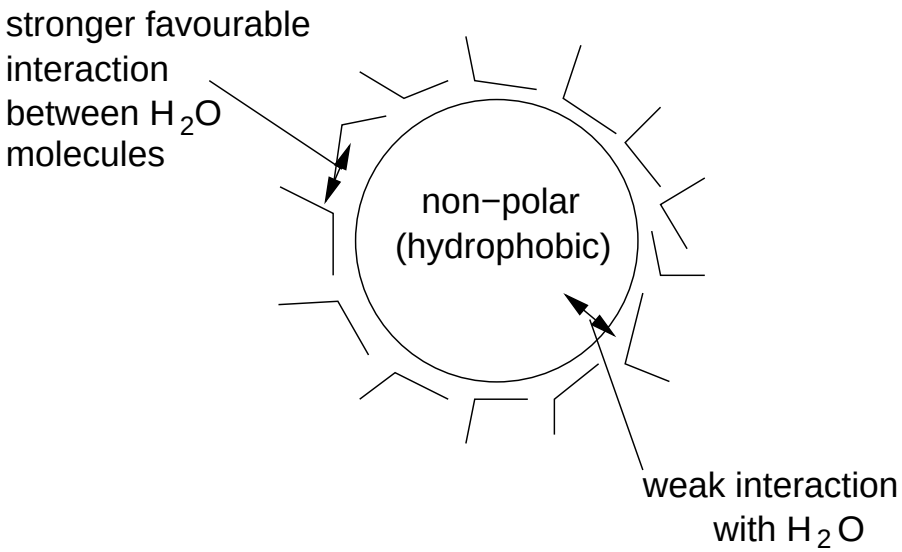
Hydrogen bonds ("H-bonds") are directed:

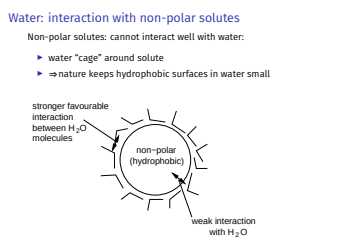


Water: interaction with non-polar solutes

Non-polar solutes: cannot interact well with water:

- water "cage" around solute
- ⇒nature keeps hydrophobic surfaces in water small





1. A well-known example of the "hydrophobic effect" is what happens to oil drops in water. If you let a few drops of cooking oil fall into water they have the tendency to unite to form a larger drop. In this way the overall surface of oil exposed to the water is minimized (the surface of many small drops is larger than the surface of one large drop). This is one of the effects responsible for the very important phenomenon called "protein folding", i.e. the formation of compact structures from extended polypeptide chains (we come back to this later).

"Small molecules"

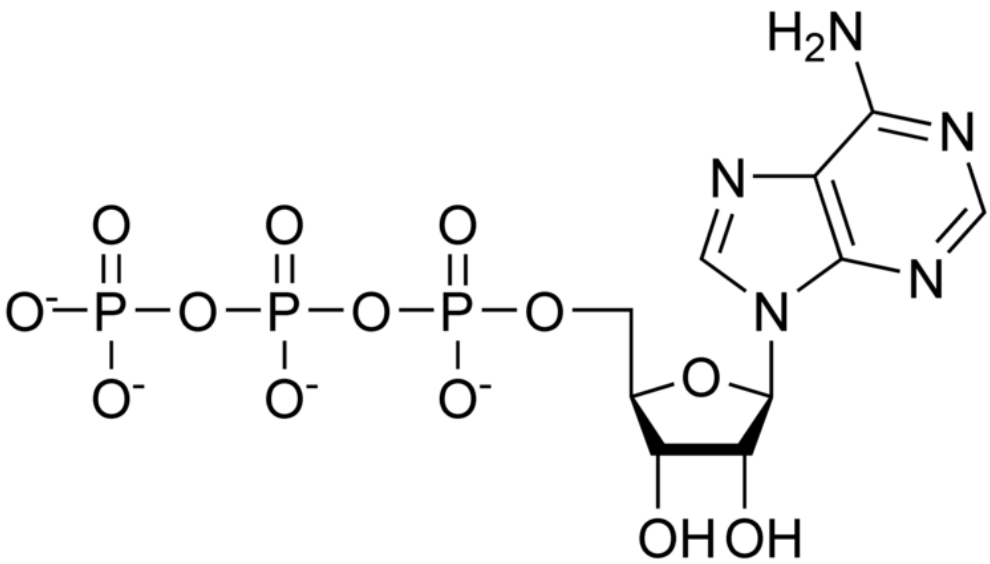
Various functions:

- metabolites
- signalling
- building blocks for macromolecules
- co-factors
- special case: drugs
- ...

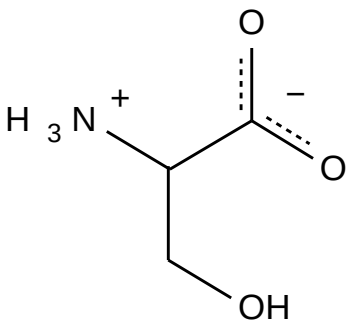
Very diverse class of molecules

“Small molecules”: example nucleotide

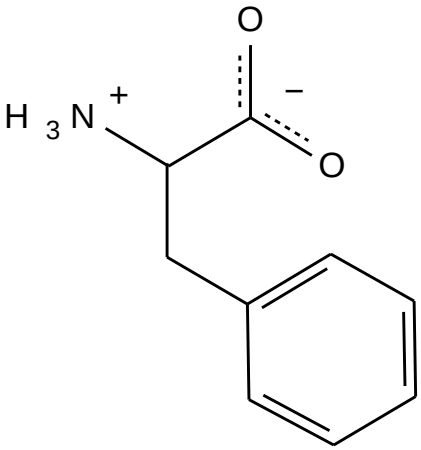
Adenosine triphosphate (ATP):



“Small molecules”: example amino acids



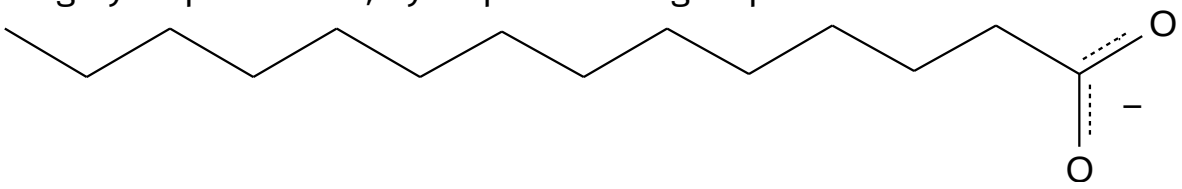
Serine (Ser, S)



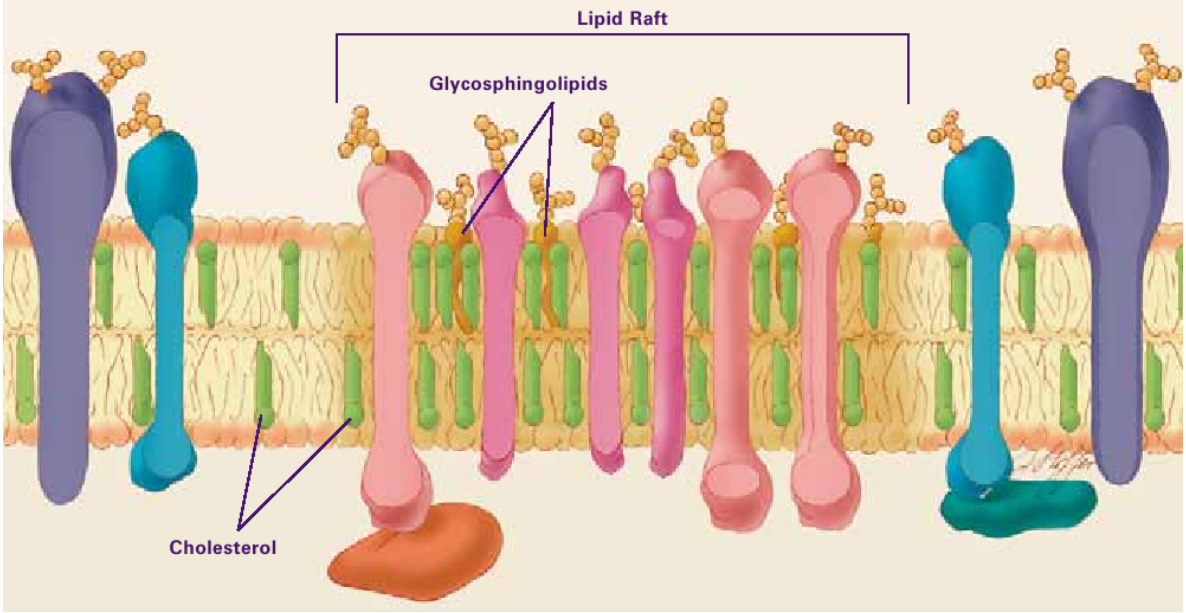
Phenylalanine (Phe, F)

Lipids

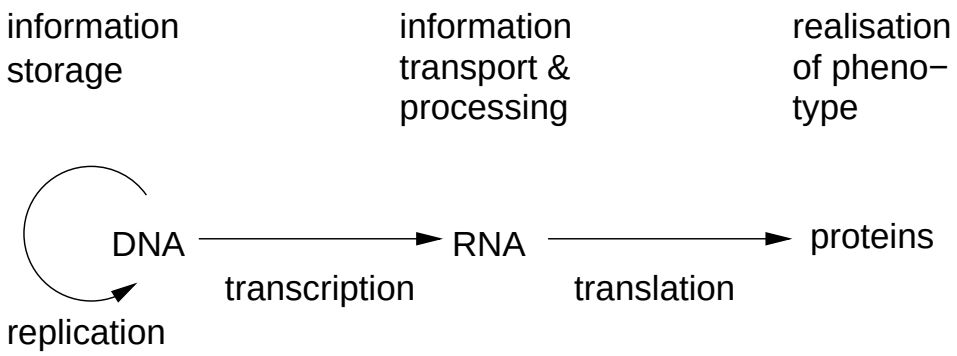
long hydrophobic tail, hydrophilic headgroup:

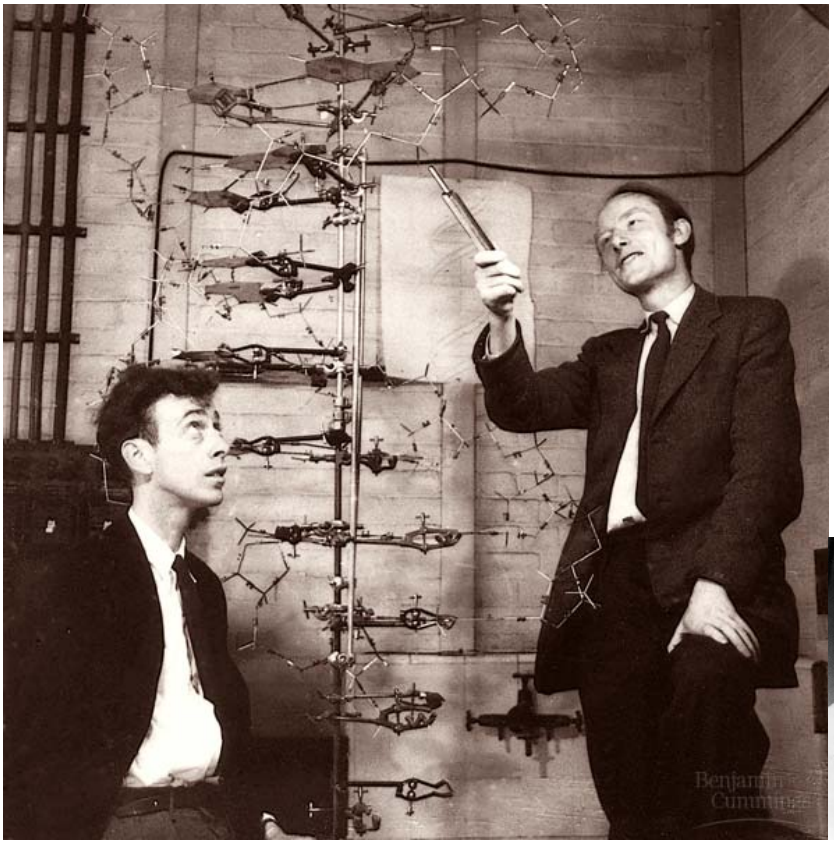


Lipids: membranes



The “central dogma” of molecular biology





2020-11-03

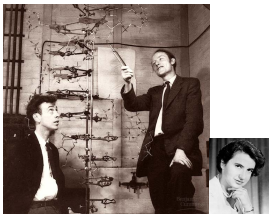
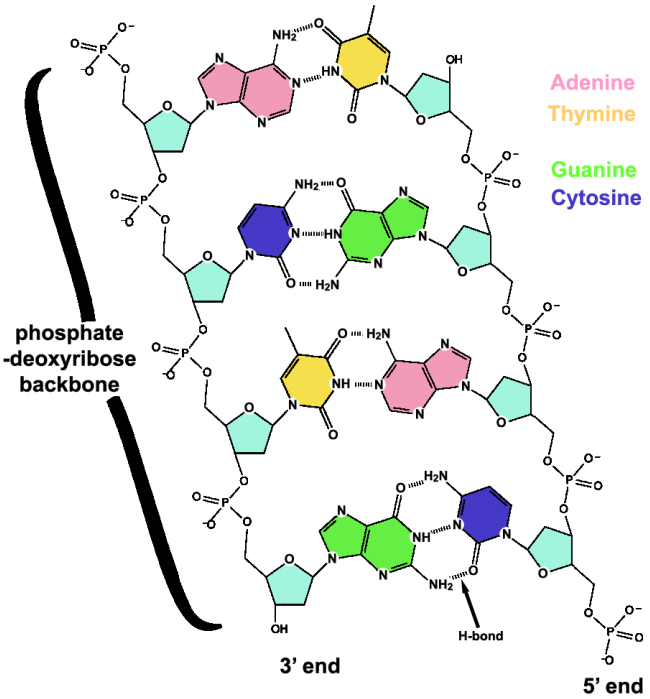


Photo of Rosalind Franklin: By Jewish Chronicle Archive/Heritage-Images http://www.britannica.com/EBchecked/topic-art/217394/99712/Rosalind-Franklin, Fair use, https://en.wikipedia.org/w/index.php?curid=24959067

Deoxyribonucleic acid (DNA)

- Central information storage
- Very long polymers
- Duplex forming (the famous double helix):



Ribonucleic acid (RNA)

Multipurpose class, e.g. information transport,

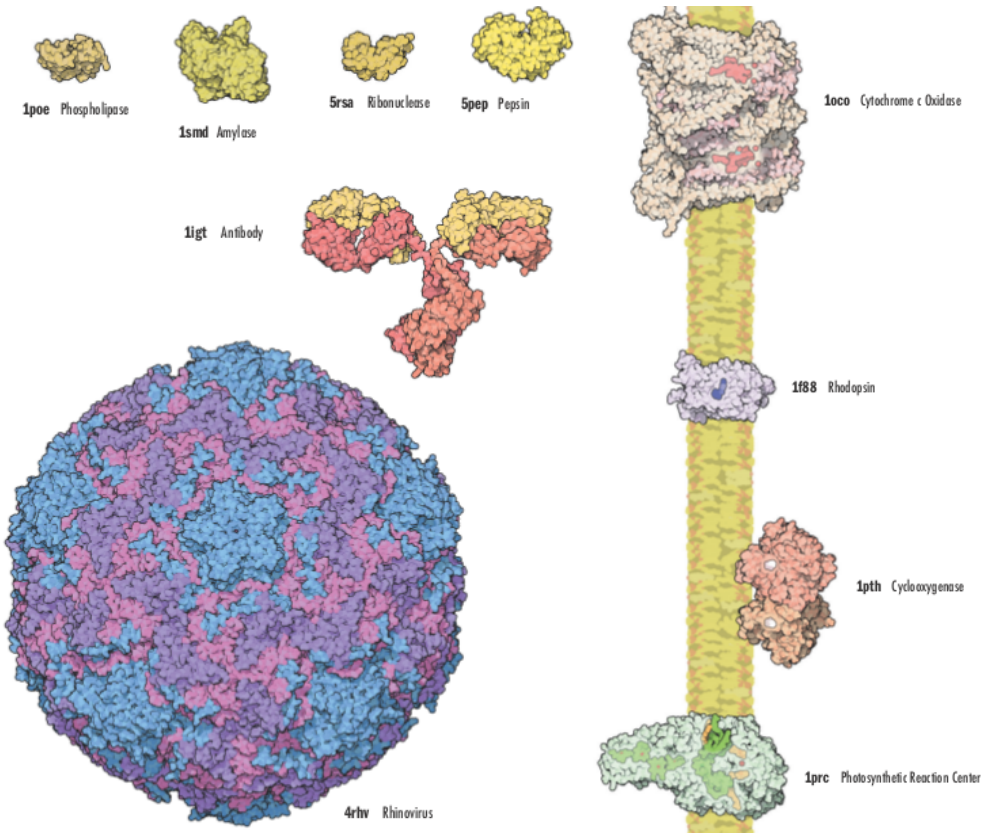
forms “duplexes” with itself and DNA

Functions in nature and technology:

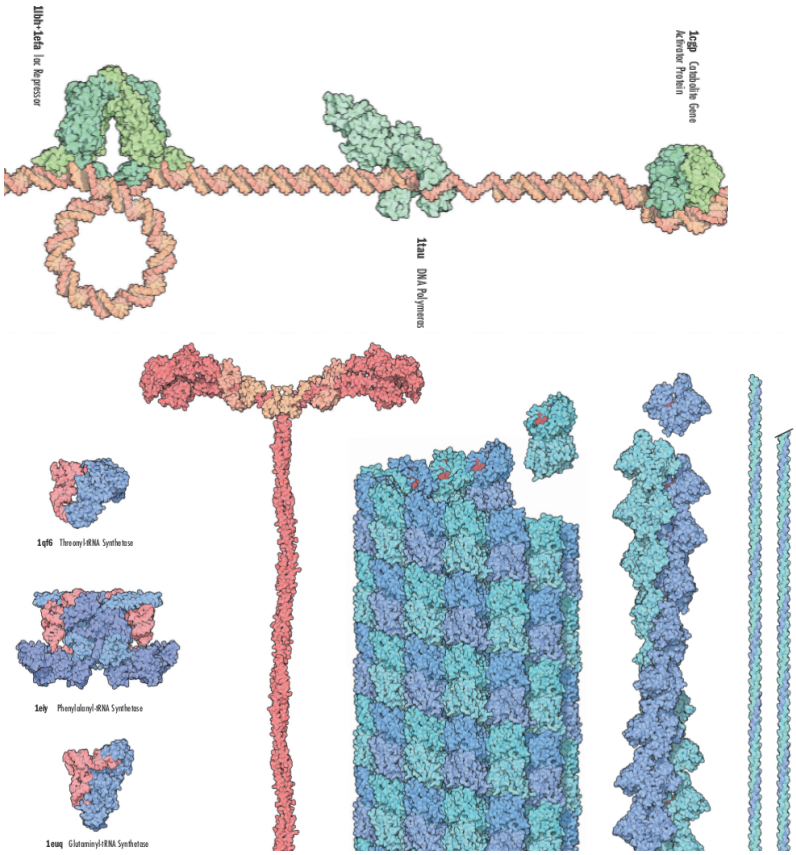
- messenger RNA (mRNA): transport of information from DNA to ribosomes
- transfer RNA (tRNA): carry amino-acid building blocks to ribosomes
- ribozymes: RNA based enzymes
- small RNA: silencing of genes etc.
- ...

(see paintings by David S. Goodsell)

Biomacromolecules I (Goodsell poster)



Biomacromolecules II



Proteins

Mainly **polypeptides**:

chains of amino acids connected by peptide bonds

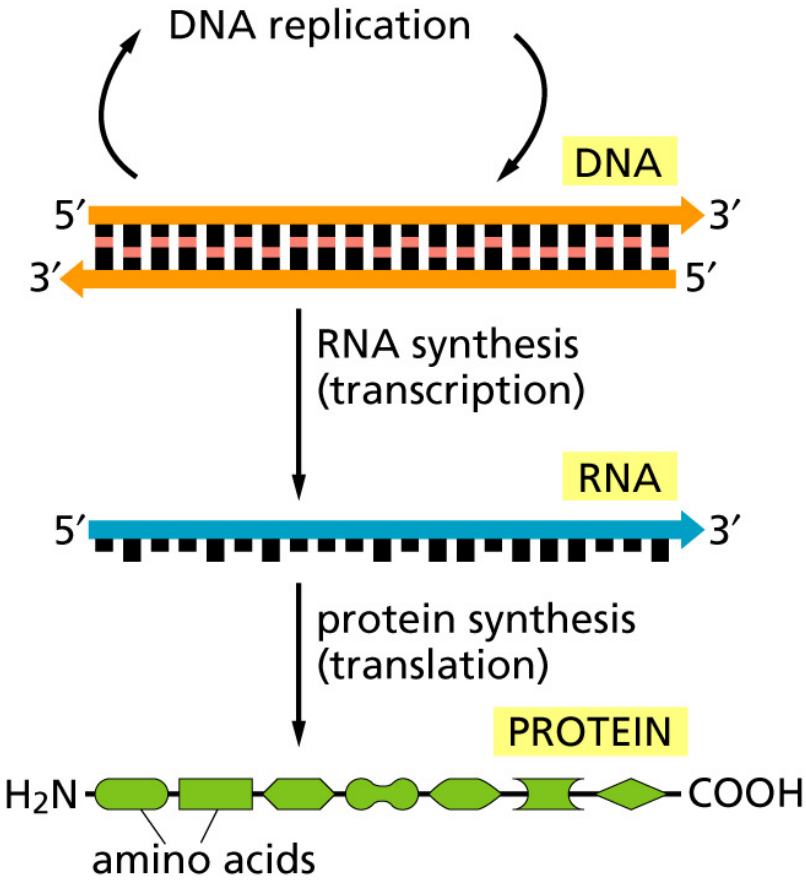
(often additional non-peptide groups included)

Various functions:

- enzymes: catalysis of chemical reactions
- transport: channels in membranes, carriers
- structure: collagen, actin, ...
- ...

Much more about proteins later.

The central dogma with molecules



Complexes, higher order structures

Building blocks interact ⇒

more or less **transient** complexes, organelles, ...

Examples:

- ribosomes (RNA, proteins)
- nucleosomes (DNA, proteins)
- ...

Life and information processing

Life can be interpreted in terms of information processing:

storage, retrieval, copying, translating and editing of information.

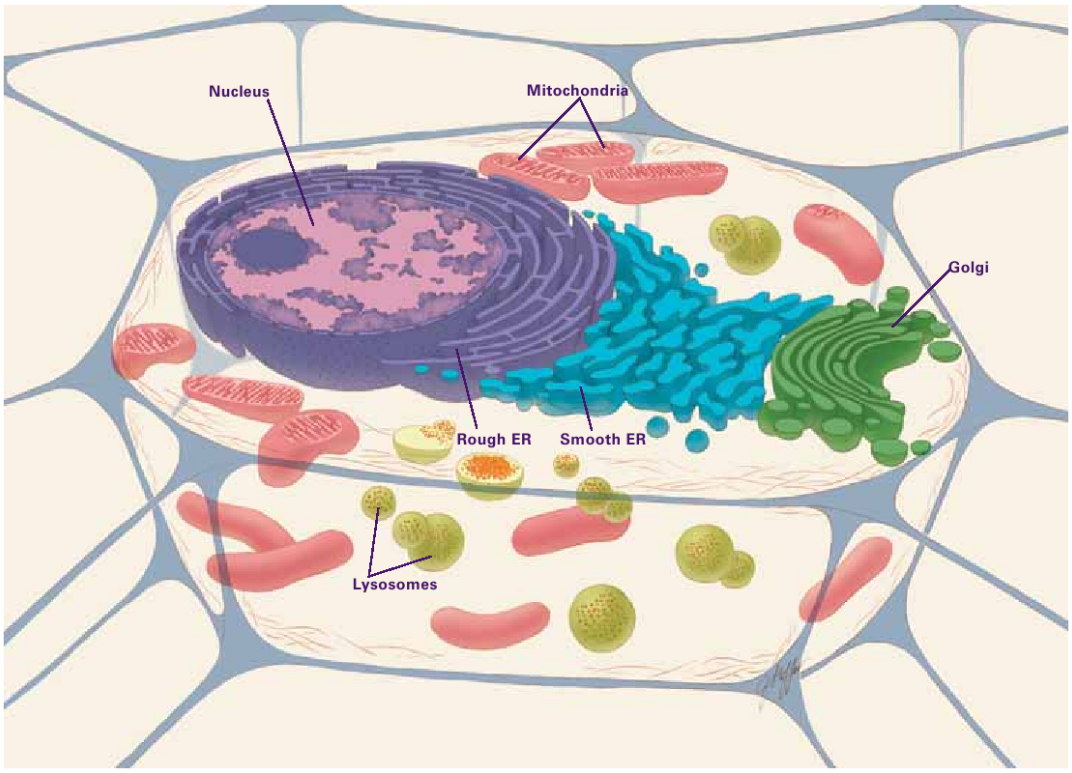
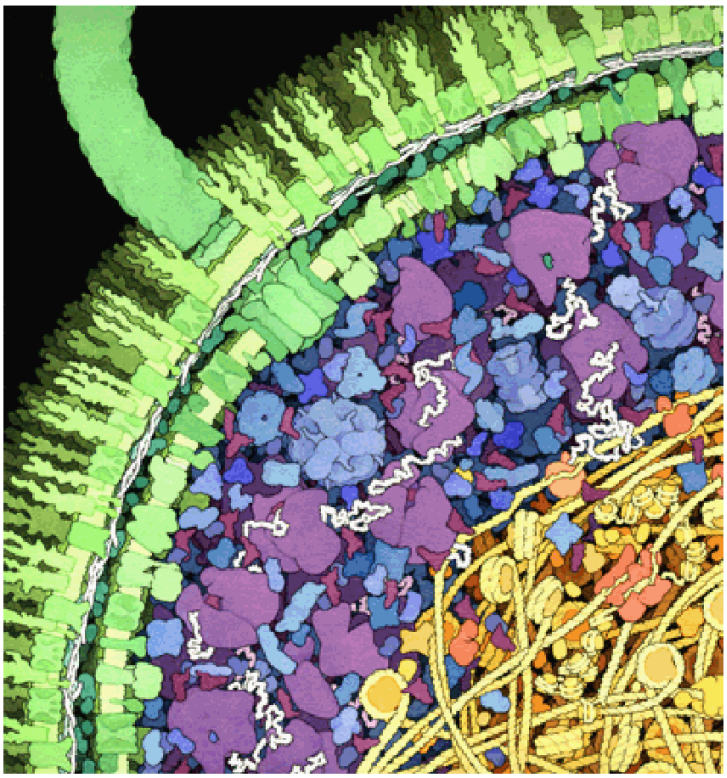
Essential aspects of life can be modelled by computer science:

- sequences of DNA, RNA or proteins as character strings
- interactions as networks (“graphs”)
- evolutionary relationships as trees
- protein folding etc. as optimization
- transmission of information via signaling pathways
- ...

Summary

- What is life
- Classes of biomolecules
- Interactions and biological structures
- Describing molecular biology by data structures

Eucaryotic cell



Procaryotic cell