CSI5126. Algorithms in bioinformatics

Essential **Cellular Biology**

Marcel Turcotte

uOttawa

School of Electrical Engineering and Computer Science (EECS)
University of Ottawa

Version September 11, 2018

Summary

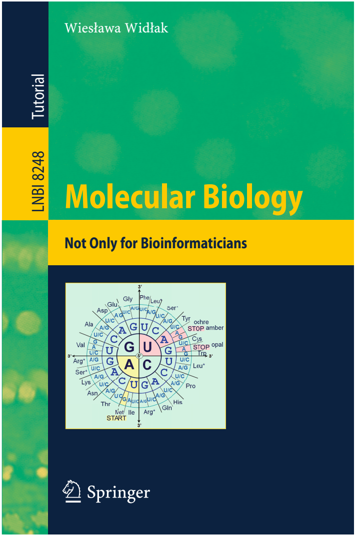
This lecture presents the **cell**, the **kinds of cells**, their **organization** and **composition**. Concepts from **molecular evolution** are introduced. It presents the **macromolecules** of the cell, with their basic organization. Throughout the presentation, we will highlight the importance of the notions for bioinformatics.

General objective

- **Describe** the organization of the cell and macromolecules

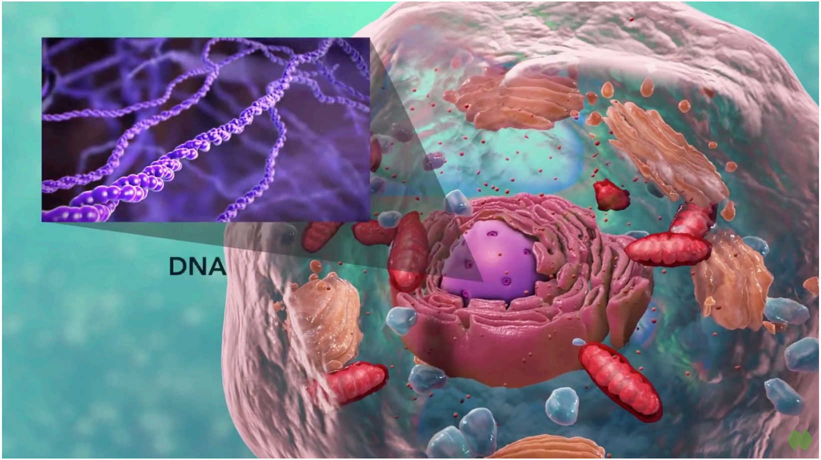
Reading

- Wiesława Widłak (2013). Molecular Biology: Not Only for Bioinformaticians (Vol. 8248). Springer. Chapters 1, 2, and 3.



link.springer.com/book/10.1007/978-3-642-45361-8

Cell Structure



https://www.youtube.com/watch?v=URUJD5NEXC8

Cells: building blocks of living organisms

Two **kinds** of cells (with and without nucleus)

Prokaryote (procaryote, prokaryotic cell, procaryotic organism):
Cell or organism **lacking** a membrane-bound, structurally **discrete nucleus** and other sub-cellular compartments. Bacteria are prokaryotes.

Cells: building blocks of living organisms

Two **kinds** of cells (with and without nucleus)

Prokaryote (procaryote, prokaryotic cell, procaryotic organism): Cell or organism **lacking** a membrane-bound, structurally **discrete nucleus** and other sub-cellular compartments. Bacteria are prokaryotes.

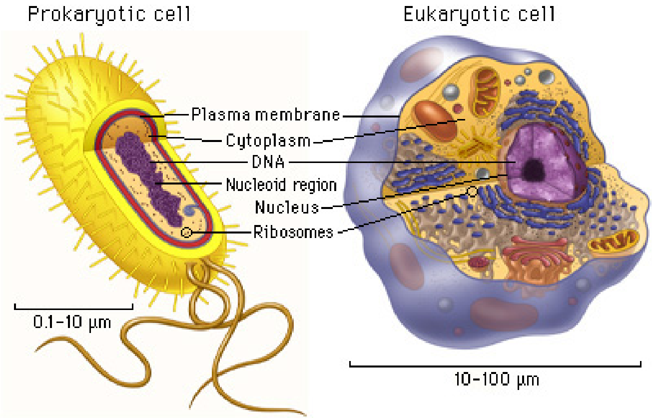
Eukaryote (eucaryote, eukaryotic cell, eucaryotic cell): Cell or organism **with** a membrane-bound, structurally **discrete nucleus** and other well-developed sub-cellular compartments. Eukaryotes include all organisms except viruses, bacteria, and cyanobacteria (blue-green algae).

Cells: building blocks of living organisms

- Eukaryotic cells are generally larger than prokaryotic cells.
- The packaging of the genetic information (DNA) is much more **structured** and compact in **Eukaryotes** compared to **Prokaryotes**.

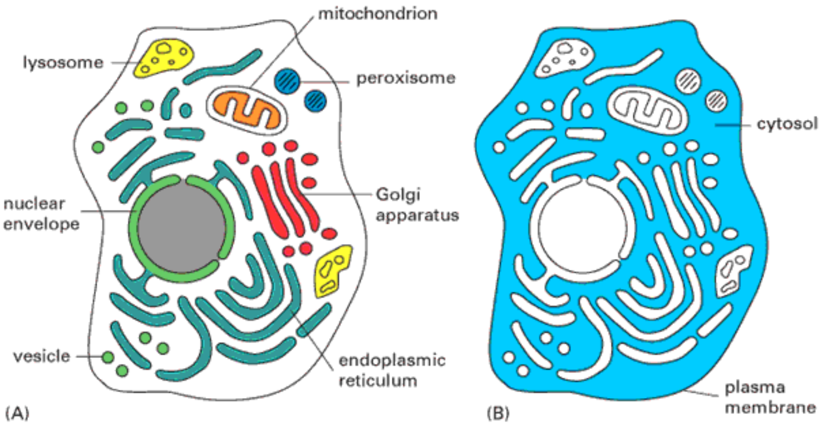
Cell theory: 1939 by Matthias Schleiden and Theodor Schwann.

Prokaryotic vs eukaryotic cell



www.phschool.com/science/biology_place/biocoach/cells/common.html

Organisation of an eukaryotic cell



Organelle genomes

- Organelles are discrete structures having **specialized functions**.

Organelle genomes

- Organelles are discrete structures having **specialized functions**.
- **Mitochondria** are **energy-generating** organelles (cellular power plants).

Organelle genomes

- Organelles are discrete structures having **specialized functions**.
- **Mitochondria** are **energy-generating** organelles (cellular power plants).
- Mitochondria **contain DNA** and a small number of genes, which are sometimes called extrachromosomal genes or mitochondrial genes.

Organelle genomes

- Organelles are discrete structures having **specialized functions**.
- **Mitochondria** are **energy-generating** organelles (cellular power plants).
- Mitochondria **contain DNA** and a small number of genes, which are sometimes called extrachromosomal genes or mitochondrial genes.
- Several organelles are believed to be engulfed prokaryotes (**endosymbiotic theory** made popular by Lynn Margulis)

Organelle genomes

- Organelles are discrete structures having **specialized functions**.
- **Mitochondria** are **energy-generating** organelles (cellular power plants).
- Mitochondria **contain DNA** and a small number of genes, which are sometimes called extrachromosomal genes or mitochondrial genes.
- Several organelles are believed to be engulfed prokaryotes (**endosymbiotic theory** made popular by Lynn Margulis)
- Mitochondrial genes are **inherited from the mother only**.

Bioinformaticist's point of view

- The **organization of genes** (genome structure) is quite different between the two kinds of cell.

Bioinformaticist's point of view

- The **organization of genes** (genome structure) is quite different between the two kinds of cell.
- Consequently the **gene-finding algorithms must be adapted**.

Bioinformaticist's point of view

- The **organization of genes** (genome structure) is quite different between the two kinds of cell.
- Consequently the **gene-finding algorithms must be adapted**.
- Eukaryotic cells being more complex provide a richer set of problems:
  e.g. **protein sub-cellular localisation problem**.

Bioinformaticist's point of view

- The **organization of genes** (genome structure) is quite different between the two kinds of cell.
- Consequently the **gene-finding algorithms must be adapted**.
- Eukaryotic cells being more complex provide a richer set of problems:
  e.g. **protein sub-cellular localisation problem**.
- During the sequence assembly, one has to consider the possibility of contamination, mtDNA/nuclear DNA, bacterial DNA.

Resources

- Texas Education Agency
  Advanced Biotechnology Collection on iTunes U
  - https://itunes.apple.com/ca/itunes-u/tea-advanced-biotechnology/id876525204?mt=10
  - Specifically the **Cell Structure and Function** segment
- Help Me Understand **Genetics**
  - https://ghr.nlm.nih.gov/primer
- **BBC** The Cell The Hidden Kingdom
  - https://www.youtube.com/watch?v=aDuwkdQzb2g
- http://learn.genetics.utah.edu

(3) kingdoms of life

Prokarya: the cells of those organisms, **prokaryotes**, do not have a nucleus. Representative organisms are *cyanobacteria* (blue-green algae) and *Escherichia coli* (a common bacteria).

(3) kingdoms of life

Prokarya: the cells of those organisms, **prokaryotes**, do not have a nucleus. Representative organisms are *cyanobacteria* (blue-green algae) and *Escherichia coli* (a common bacteria).

Eukarya: the cells of those organisms, **eukaryotes**, all have a nucleus. Representative organisms are *Trypanosoma brucei* (unicelluar organism which can cause sleeping sickness) and *Homo sapiens* (multicellular organism).

(3) kingdoms of life

Prokarya: the cells of those organisms, **prokaryotes**, do not have a nucleus. Representative organisms are *cyanobacteria* (blue-green algae) and *Escherichia coli* (a common bacteria).

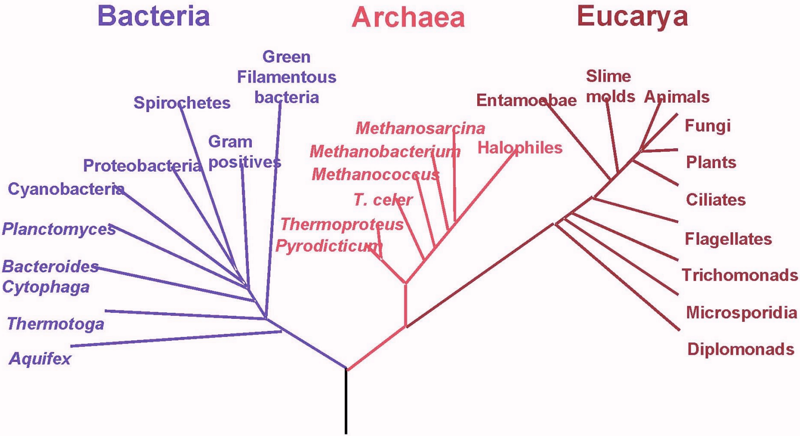
Eukarya: the cells of those organisms, **eukaryotes**, all have a nucleus. Representative organisms are *Trypanosoma brucei* (unicelluar organism which can cause sleeping sickness) and *Homo sapiens* (multicellular organism).

Archaea: (archaebacteria) like the prokaryotes **they lack the nuclear membrane** but have **transcription** and **translation mechanisms close to those of the eukaryotes**.

(3) **kingdoms of life**: Archaea

Methanococcus jannaschii is an **methane producing archaebacterium** which had its complete genome sequenced in 1996. This organism was discovered in 1982 in white smoker of a hot spot at the bottom of the Pacific ocean: depth **2600 meters**, **temperature 48-94° C (thermophilic)**, optimum at 85° C, 1.66 Mega bases, 1738 genes. 56% of its genes are unlike any known eukaryote or prokaryote, one kind of DNA polymerase (other genomes have several).

Phylogenetic Tree of Life



Phylogenetic tree

- "The objectives of phylogenetic studies are (1) to reconstruct the correct **genealogical ties between organisms** and (2) to estimate the **time of divergence** between organisms since they last shared a common ancestor."

⇒ Li, W.-H. and Graur, D. (1991) *Fundamentals of Molecular Evolution*. Sinauer.

Phylogenetic tree

- "The objectives of phylogenetic studies are (1) to reconstruct the correct **genealogical ties between organisms** and (2) to estimate the **time of divergence** between organisms since they last shared a common ancestor."
- "A phylogenetic tree is a **graph** composed of nodes and branches, in which only one branch connects any two adjacent nodes."

⇒ Li, W.-H. and Graur, D. (1991) *Fundamentals of Molecular Evolution*. Sinauer.

Phylogenetic tree

- "The objectives of phylogenetic studies are (1) to reconstruct the correct **genealogical ties between organisms** and (2) to estimate the **time of divergence** between organisms since they last shared a common ancestor."
- "A phylogenetic tree is a **graph** composed of nodes and branches, in which only one branch connects any two adjacent nodes."
- "The **nodes** represents the **taxonomic units**, and the **branches** define the **relationships** among the units in terms of descent and ancestry."

⇒ Li, W.-H. and Graur, D. (1991) *Fundamentals of Molecular Evolution*. Sinauer.

Phylogenetic tree

- "The objectives of phylogenetic studies are (1) to reconstruct the correct **genealogical ties between organisms** and (2) to estimate the **time of divergence** between organisms since they last shared a common ancestor."
- "A phylogenetic tree is a **graph** composed of nodes and branches, in which only one branch connects any two adjacent nodes."
- "The **nodes** represents the **taxonomic units**, and the **branches** define the **relationships** among the units in terms of descent and ancestry."
- "The **branch length** usually represents the **number of changes** that have occurred in that branch." (or some amount of time)

⇒ Li, W.-H. and Graur, D. (1991) *Fundamentals of Molecular Evolution*. Sinauer.

Bioinformaticist's point of view

- **Bench-marking** (cross-validation) and **molecular evolution**
- **Molecular sequence alignment** : are the sequences evolutionary related?
- **Large phylogeny problem**: Reconstructing phylogenetic trees from molecular sequence data
- **Small phylogeny problem**: Reconstructing ancestral molecular sequences

Nothing in Biology Makes Sense Except in the Light of Evolution

Theodosius Dobzhansky

What about **virus**?

What about **virus**?

- Virus are **agents** infecting the cells of living organisms.

What about **virus**?

- Virus are **agents** infecting the cells of living organisms.
- Are not able to replate by themselves – therefore, must "hijack" the machinery of a living organism.

What about **virus**?

- Virus are **agents** infecting the cells of living organisms.
- Are not able to replate by themselves – therefore, must "hijack" the machinery of a living organism.
- Simple structure consisting of **nucleic acids** and **proteins**.

What about **virus**?

- Virus are **agents** infecting the cells of living organisms.
- Are not able to replate by themselves – therefore, must "hijack" the machinery of a living organism.
- Simple structure consisting of **nucleic acids** and **proteins**.
- Small number of genes: mainly for the protein that forms the capsid (envelop).

What about **virus**?

- Virus are **agents** infecting the cells of living organisms.
- Are not able to replate by themselves – therefore, must "hijack" the machinery of a living organism.
- Simple structure consisting of **nucleic acids** and **proteins**.
- Small number of genes: mainly for the protein that forms the capsid (envelop).
- Can be **DNA** or **RNA**-based.

What about **virus**?

- Virus are **agents** infecting the cells of living organisms.
- Are not able to replate by themselves – therefore, must "hijack" the machinery of a living organism.
- Simple structure consisting of **nucleic acids** and **proteins**.
- Small number of genes: mainly for the protein that forms the capsid (envelop).
- Can be **DNA** or **RNA**-based.
- RNA virus encode an enzyme, called a reverse transcriptase, allowing to copy their genome to DNA, and insert it into the host.

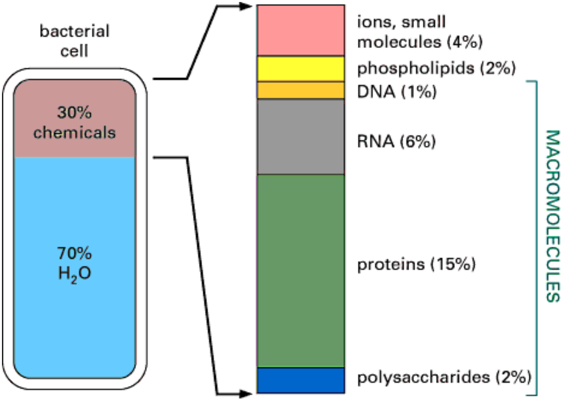
What about **virus**?

- Virus are **agents** infecting the cells of living organisms.
- Are not able to replate by themselves – therefore, must "hijack" the machinery of a living organism.
- Simple structure consisting of **nucleic acids** and **proteins**.
- Small number of genes: mainly for the protein that forms the capsid (envelop).
- Can be **DNA** or **RNA**-based.
- RNA virus encode an enzyme, called a reverse transcriptase, allowing to copy their genome to DNA, and insert it into the host.
- Virus that infect bacteria are called phages or bacteriophages.

What about **virus**?

- Virus are **agents** infecting the cells of living organisms.
- Are not able to replate by themselves – therefore, must "hijack" the machinery of a living organism.
- Simple structure consisting of **nucleic acids** and **proteins**.
- Small number of genes: mainly for the protein that forms the capsid (envelop).
- Can be **DNA** or **RNA**-based.
- RNA virus encode an enzyme, called a reverse transcriptase, allowing to copy their genome to DNA, and insert it into the host.
- Virus that infect bacteria are called phages or bacteriophages.
- Viroids don't even have a capsid – consists of a single-stranded RNA.

Composition of the Cell



⇒ **DNA**, **RNA** and **proteins** will be the main focus of the course.

Macromolecules: **DNA** (deoxyribonucleic acid), **RNA** (ribonucleic acid) and **Protein**

Bioinformatics is mainly concerned with three classes of molecules:

- **DNA**, **RNA** and **proteins** — collectively called **macromolecules** or **biomolecules**.

Macromolecules: **DNA** (deoxyribonucleic acid), **RNA** (ribonucleic acid) and **Protein**

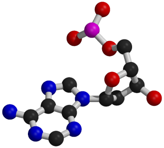
All three classes of macromolecules are **polymers**, that is they are composed of smaller units (molecules), called **monomers**, that are **linked sequentially** one to another forming **unbranched linear structures**.

Macromolecules: DNA (deoxyribonucleic acid), RNA (ribonucleic acid) and Protein

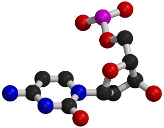
Generally speaking, the units (monomers) consits of two distinct parts, one that is **common** to all the monomers and defines the **backbone** of the molecule, and another part that confers the **identity** of the unit, and therefore its **properties**.

```
[ ]-[ ]-[ ]-[ ]-[ ]- ... -[ ]-[ ]
 |   |   |   |   |         |   |
 *   @   *   #   +         +   @
```

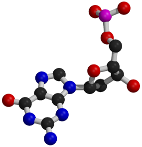
DNA's building blocks: ACGT



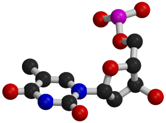
Adenine (A)



Cytosine (C)

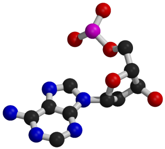


Guanine (G)

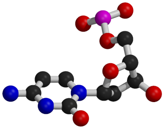


Thymine (T)

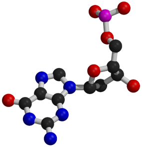
DNA's building blocks: ACGT



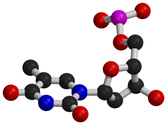
Adenine (A)



Cytosine (C)



Guanine (G)



Thymine (T)

$\Rightarrow$ **Identify** the common and unique parts of each monomer.

(20) Amino Acids (Naturally Occuring)


A (Ala)


D (Asp)


E (Glu)


K (Lys)


P (Pro)


W (Trp)


V (Val)


R (Arg)


C (Cys)

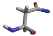

G (Gly)


I (Ile)


M (Met)


S (Ser)


Y (Tyr)


N (Asn)


Q (Gln)


H (His)


L (Leu)


F (Phe)


T (Thr)

⇒ Stick (licorice) representation.

Structure

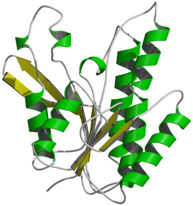
It's useful to distinguish between four **levels of abstraction** or **structure**: **primary**, **secondary**, **tertiary** and **quaternary** structure.

1, 2, 3, ...

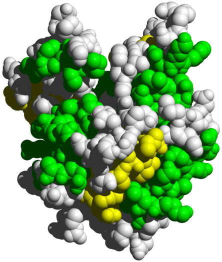
EARRVLVYGGRGALGSRCVQNW ...(236) ...

(a) primary structure

(b) secondary structure



(c) tertiary structure - ribbon



(d) tertiary structure - all atoms

Bioinformaticist's point of view

- A large number of computational problems are related to the **primary sequence**: sequence assembly, sequence alignment, phylogenetic tree inference, gene-finding, sequence motif discovery, etc.
- Predicting the **secondary, tertiary, and quaternary (docking) structure** are problems, on its own.
- These **abstractions** are allowing us to formulate efficient algorithms - understanding the implications is paramount.

Macromolecules: DNA (deoxyribonucleic acid), RNA (ribonucleic acid) and Protein

The primary structure or **sequence** is an ordered list of characters, from a given alphabet, written contiguously from left to right.
DNA : 4 letters alphabet,
$\Sigma = \{A, C, G, T\}$
RNA : 4 letters alphabet,
$\Sigma = \{A, C, G, U\}$
Proteins : 20 letters alphabet,
$\Sigma = \{A, C, D, E, F, G, H, I, K, L, M, N, P, Q, R, S, T, V, W, Y\}$

Examples

In the case of **nucleic acids** (DNA and RNA), the building blocks are called **nucleotides**, whilst in the case of **proteins** they are called **amino acids**.

Examples of DNA, RNA and protein sequences.

```
> Chimpanzee Chromosome 1; A DNA sequence (size = 245,522,847 nt)
TAACCCTAACCCTAACCCTAACCCTAACC ... TCTCATGACAGTGAGTGAGTTCTCATGATC
```

Examples

In the case of **nucleic acids** (DNA and RNA), the building blocks are called **nucleotides**, whilst in the case of **proteins** they are called **amino acids**.

Examples of DNA, RNA and protein sequences.

```
> Chimpanzee Chromosome 1; A DNA sequence (size = 245,522,847 nt)
TAACCCTAACCCTAACCCTAACCCTAACC ... TCTCATGACAGTGAGTGAGTTCTCATGATC
```

```
> A01592; An RNA sequence (coding Beta Globin gene) (size = 441 nt)
AUGGUGCACCUGACUCCUGAGGAGAAGUCUGC ... GCAAGGUGAACGUGGAUGAAGUUGGUGGUG
```

Examples

In the case of **nucleic acids** (DNA and RNA), the building blocks are called **nucleotides**, whilst in the case of **proteins** they are called **amino acids**.

Examples of DNA, RNA and protein sequences.

```
> Chimpanzee Chromosome 1; A DNA sequence (size = 245,522,847 nt)
TAACCCTAACCCTAACCCTAACCCTAACC ... TCTCATGACAGTGAGTGAGTTCTCATGATC
```

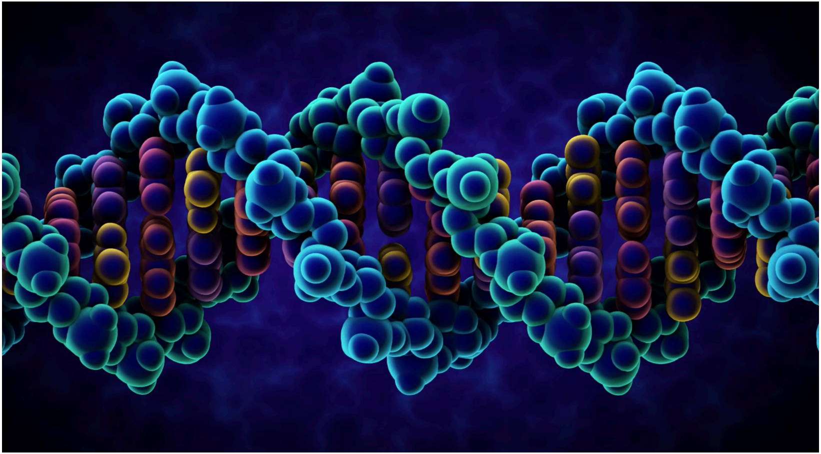
```
> A01592; An RNA sequence (coding Beta Globin gene) (size = 441 nt)
AUGGUGCACCUGACUCCUGAGGAGAAGUCUGC ... GCAAGGUGAACGUGGAUGAAGUUGGUGGUG
```

```
> Beta Globin; A protein sequence (size = 147 aa)
MVHLTPEEKSAVTALWGKVNVDEVGGEAL ... FFESFGDLSTPDAVMGNPKVKAHGKKVLGA
```

Bioinformaticist's point of view

- **Exact string** (sequence) comparison, **approximate matching** ($k-$mismatches), comparison under the **edit-distance**, **significance** of match, **multi-way** sequence comparison
- Finding **repeats**, **approximate repeats**, finding interesting **patterns**
- Secondary, tertiary and quaternary structure inference

DNA

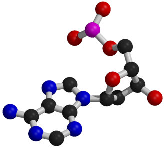


https://www.youtube.com/watch?v=o_-6JXLYS-k

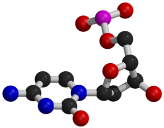
Deoxyribonucleic acids (**DNA**)

- **DNA** was discovered by **Johann Friedrich Miescher** in **1869**. Who discarded the possibility that DNA might be related to heredity!
- The **double-helical structure** of DNA was proposed in **1953** by James Watson and Francis Crick (who died on July 28, 2004).
- This discovery is often referred to as the **most important breakthrough in biology of the 20th century**.
- The proposed model finally explained Chargaff's rule (same amount of adenine and thymine, same amount of guanine and cytosine).
- More importantly, the model finally explains **how DNA and heredity are linked!** (replication)

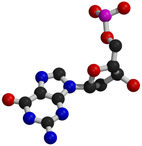
DNA's building blocks: ACGT



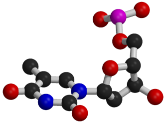
Adenine (A)



Cytosine (C)



Guanine (G)



Thymine (T)

DNA/RNA's building blocks

- The **common part** of the nucleotides is formed of a **deoxy-ribose** (pentose, sugar) and a **phosphate** group.

DNA/RNA's building blocks

- The **common part** of the nucleotides is formed of a **deoxy-ribose** (pentose, sugar) and a **phosphate** group.
- The part that is **unique** is called the (nitrogenous) **base**.

DNA/RNA's building blocks

- The **common part** of the nucleotides is formed of a **deoxy-ribose** (pentose, sugar) and a **phosphate** group.
- The part that is **unique** is called the (nitrogenous) **base**.
- If you look carefully you'll see big (two rings) and small (one ring) bases, respectively called **purines** (A,G) and **pyrimidines** (C,T).

DNA/RNA's building blocks

- The **common part** of the nucleotides is formed of a **deoxy-ribose** (pentose, sugar) and a **phosphate** group.
- The part that is **unique** is called the (nitrogenous) **base**.
- If you look carefully you'll see big (two rings) and small (one ring) bases, respectively called **purines** (A,G) and **pyrimidines** (C,T).
- In the case of **DNA**, the bases are Adenine (**A**), Cytosine (**C**), Guanine (**G**) and Thymine (**T**).

DNA/RNA's building blocks

- The **common part** of the nucleotides is formed of a **deoxy-ribose** (pentose, sugar) and a **phosphate** group.
- The part that is **unique** is called the (nitrogenous) **base**.
- If you look carefully you'll see big (two rings) and small (one ring) bases, respectively called **purines** (A,G) and **pyrimidines** (C,T).
- In the case of **DNA**, the bases are Adenine (**A**), Cytosine (**C**), Guanine (**G**) and Thymine (**T**).
- In the case of **RNA**, the bases are Adenine (**A**), Cytosine (**C**), Guanine (**G**) and Uracil (**U**).

DNA/RNA's building blocks

- The length of a DNA/RNA molecule is often expressed in **bases**, e.g. a 10 **mega base** long region.
- Or, since nucleic acids molecules **hybridize** (bind together) to form a duplex (double helical) structure, the length of a molecule is often expression is base pairs to avoid confusion, e.g. a 10 **mega base** pairs region.

DNA/RNA's building blocks

- **DNA** stands for deoxyribonucleic acid, and **deoxy** comes from the fact that the **C2' carbon of the sugar has no oxygen**; while RNA has one. **RNA**'s O2' oxygen is key to its **functional versatility**!
- The other difference is the use of **T** (thymine) in the case of **DNA** vs **U** (uracil) in the case of **RNA**.
- Nucleotides are **always attached one to another in the same way** (well, almost always): the **C3' atom** of the nucleotide i is covalently linked to the **phosphate group** of the nucleotide $i + 1$.

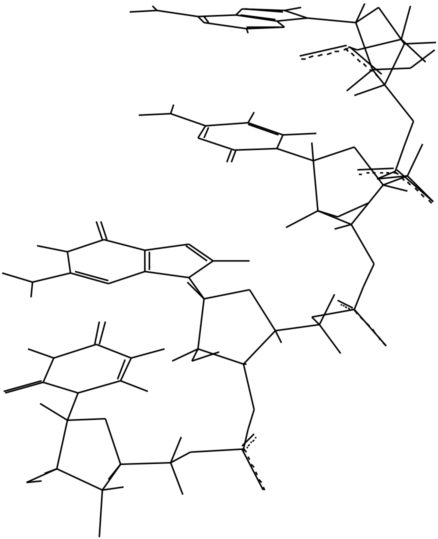
DNA/RNA's building blocks

- The **orientation of a DNA molecule** is important; just like the orientation of words are important in natural languages.

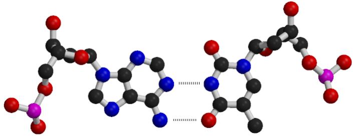
DNA/RNA's building blocks

- The **orientation of a DNA molecule** is important; just like the orientation of words are important in natural languages.
- The convention is to enumerate the string from its **5' end**; this correspond to the order into which information is process for certain key steps, to be described later. The features that are occurring **before the 5'** are said to be **upstream** while those occurring after the **3'** end are **downstream**, upstream and downstream signals.

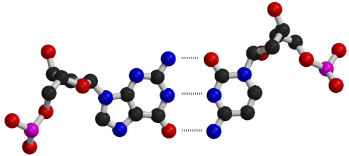
DNA strand



Watson-Crick (Canonical) **base pairs**



(Adenosine) A : T (Thymine)



(Guanine) G : C (Cytosine)

⇒ One of the two base pairs is stronger that the other, which one?

Watson-Crick (Canonical) **base pairs**

In the case of **DNA**, bases interact, i.e. form hydrogen bonds, primarily through the following set of rules:

- **A** interacts with **T** (and vice versa)
- **G** interacts with **C** (and vice versa)

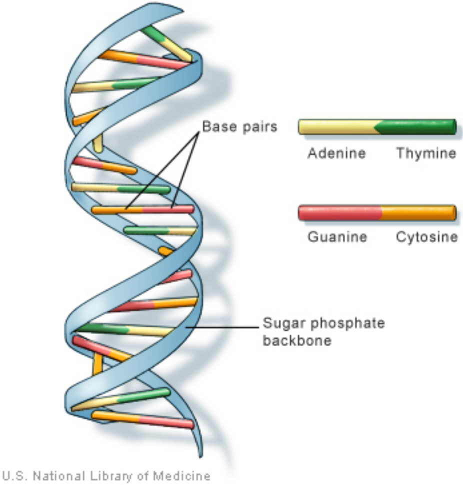
Watson-Crick (Canonical) **base pairs**

In the case of **DNA**, bases interact, i.e. form hydrogen bonds, primarily through the following set of rules:

- **A** interacts with **T** (and vice versa)
- **G** interacts with **C** (and vice versa)

Those rules are the consequence of the fact that A:T and G:C pairs position the backbone atoms roughly at the same three-dimensional location and therefore both produces the same double helical structure; isosteric base pairs.

- **DNA** molecules generally form **right-hand** side helices in **B form**, while **RNA** are **A form**, also right-hand side. A left-hand side helix exists that is called **Z DNA**.
- **DNA** molecules **cannot exist as a single strand**, they are degraded, i.e. cut into pieces.
- A **DNA** molecule is made of **two complementary strands running in opposite directions**.


Base pairs
Adenine
Thymine
Guanine
Cytosine
Sugar phosphate backbone
U.S. National Library of Medicine

DNA and Heredity

- **DNA structure explains how information can be copied from one generation to the next**, or simply from one parent cell to its daughter cells during replication.

Before replication

```
5'  - GATACA -> 3' A
      ||||||
3' <- CTATGT -  5' B
```

$\Rightarrow$

A is as a template to produce **B'**

```
5'  - GATACA -> 3' A


5'  - GATACA -> 3' A
      ||||||
3' <- CTATGT -  5' B'
```

DNA and Heredity

Before replication

```
5'  - GATACA -> 3' A
      ||||||
3' <- CTATGT -   5' B
```

$\Rightarrow$

B is as a template to produce **A'**

```
5' -  TGTATC -> 3' B


5' -  TGTATC -> 3' B
      ||||||
3' <- ACATAG -> 5' A'
```

DNA and Heredity

Parent cell (AB)

```
5'  - GATACA -> 3' A
      ||||||
3' <- CTATGT -  5' B
```

Daughter cell AB'

```
5'  - GATACA -> 3' A
      ||||||
3' <- CTATGT -  5' B'
```

Daughter cell A'B

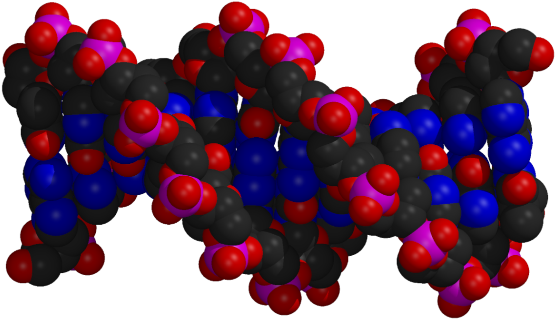
```
5' -  TGTATC -> 3' B
      ||||||
3' <- ACATAG -> 5' A'
```

Two daughter cells, identical to their parent.
(semi-conservative process)

Remarks

- Complex organisms are growing from a single cell to billions of cells. Each cell contains an **exact copy** of the **DNA** of its **parent cell**.
- The information is redundant, the information on the second strand can be inferred from the information on the first strand. This is the basis of **DNA repair mechanisms**. A base that is deleted can be replaced. A mismatch can be detected.

CPK representation of a fragment of a **DNA helix** (B form)

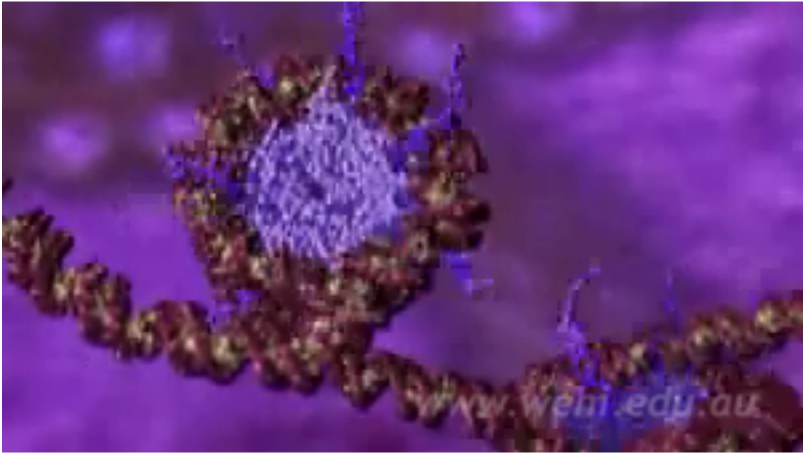


```
TAAGTTATTA                                AAAAAAATAC
||||||||||  ...  (580,074 bp)  ...  ||||||||||
ATTCAATAAT                                TTTTTTTATG
```

About CPK

CPK stands for **Corey-Pauling-Koltun** representation. **Every atom** is represented as a sphere, with **radius proportional to its van der Walls radius**. The usual color scheme is to represent carbon atoms in black, nitrogen in blue, oxygen in red and phosphorus atoms in pink.

Chromosome



https://youtu.be/0jPcT1uUZiE?list=PLD0444BD542B4D7D9

About the animation

- **Histone proteins** attach to the **DNA**.
- **Histones interact one with another** to form a complex called **nucleosome**, but also forcing the DNA to wrap around it.
- The histone, nucleosome and DNA models were derived from their PDB (http://www.rcsb.org/pdb/) structures and other published data.

About the animation

- **Histone proteins** attach to the **DNA**.
- **Histones interact one with another** to form a complex called **nucleosome**, but also forcing the DNA to wrap around it.
- The histone, nucleosome and DNA models were derived from their PDB (http://www.rcsb.org/pdb/) structures and other published data.
- Macromolecular structures cannot be directly oberserved. A molecular bond is between 1 and 2 Å (angstrom – 10^{-10} m) long, wave length in the visible spectrum are 400 to 700 nm (10^{-9} m).

Bioinformaticist's point of view

- Given DNA sequence information alone, **predict the locations where the histones will be binding**.
- Knowing the location of the histones might help predicting the **location of genes** as well as the **location of regulatory elements**.
- The **three-dimensional organization of the genome** is a hot topic.

Summary

- Two kinds of cells: **prokaryotic** and **eukaryotic**.
- **Eukaryotic cells have organelles**, and some organelles, such as the mitochondria, contain DNA.
- Three Kingdom of life: **Prokarya**, **Eukarya**, and **Archea**
- A **phylogeny** specifies the relationships between organisms and time of divergence.
- Three kinds of macromolecules: **DNA**, **RNA**, and **proteins**.
- **Macromolecules are linear (unbranched) polymers**, such that all the monomers have a common and a specific part (remember the analogy with the linked nodes).

References

W.-H. Li and D. Graur.
Fundamentals of Molecular Evolution.
Sinauer, 1991.

R. Durbin, S. Eddy, A. Krogh, and G. Mitchison.
Biological Sequence Analysis.
Cambridge University Press, 1998.

D. Gusfield.
Algorithms on Strings, Trees, and Sequences: Computer Science and Computational Biology.
Cambridge University Press, 1997.

N. C. Jones and P. A. Pevzner.
An introduction to bioinformatics algorithm.
MIT Press, 2004.

Pensez-y!

L'impression de ces notes n'est probablement pas nécessaire!