

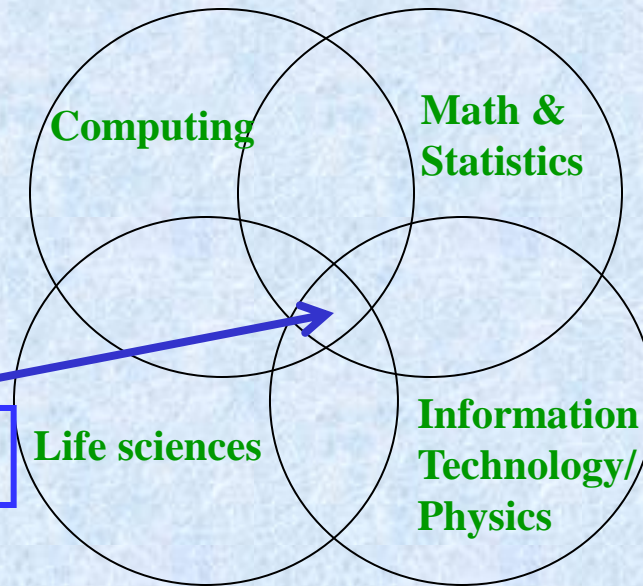
Bioinformatics

Bioinformatics is a field of science in which **biology, computer science, and information technology** merge into a single discipline to analyze **biological information** using **computers and statistical techniques**.

The term bioinformatics was coined by Paulien Hogeweg in 1979 for the study of informatics in biological systems.

Next Generation Sequencing

Bioinformatics



Biological experiments can generate large amounts of data (macromolecular sequences, structures, expression profiles, pathways etc).

Bioinformatics is about **acquiring, managing, analyzing and understanding** those data.

Bioinformatics: major aspects

- 1. Well organized databases**
- 2. Computationally derived hypothesis**
- 3. Web servers (tools/online applications)**
- 4. Virtual screening of compounds for drug development**

Example 1

Food pattern and weight control

- | | |
|---|-------------|
| 1. Rice 50%; wheat 25%; meat 10%; fruits 10%; vegetables: 5%: | Not control |
| 2. Rice 40%, wheat 5%; meat 15%; fruits 5%; vegetables: 35%: | Not control |
| 3. Rice 25%, wheat 10%, meat 10%, fruits 25%, vegetables 30%: | Control |
| 4. Rice 15%, wheat 15%, meat 40%, fruits 20%, vegetables 10%: | Not control |
| 5. Rice 15%, wheat 20%, meat 5%; fruits 15%; vegetables 45%: | Control |

**Known
experimentally**

Test case:

Rice 20%; wheat 10%, meat 10%, fruits, 20%, vegetables 40%

Computer science (statistical methods)

?

Statistical approach

Rice + wheat \leq 35%

AND

Meat $<$ 15%

AND

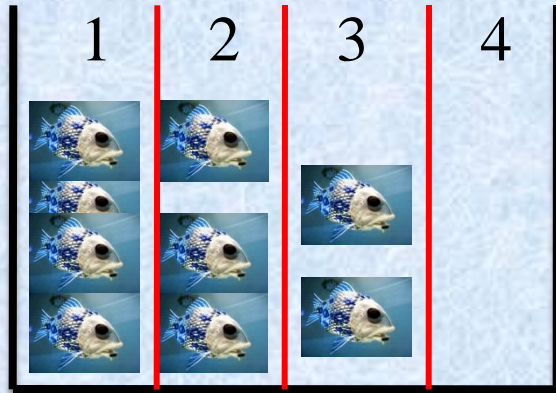
Vegetables \geq 30%

Answer: Control

**Bioinformatics can handle large amount of data
and provide possible solutions**

Example 2

Finding fish in a pond



Experiments:

All compounds (trial and error)

35 million compounds, ZINC database

2.2 million: Enamine

35,000: Natural compounds

Finding suitable drug for a disease

Disease

Compound 1 X

Compound 2 X

Drug O

Solution: BIOINFORMATICS

Computational techniques assist one in searching drug target and designing drug.

Initiative for Parallel Bioinformatics (IPAB)

Organizer: Tokyo Institute of Technology

Target: cYes Kinase

Library: Enamine (2.2 million compounds)

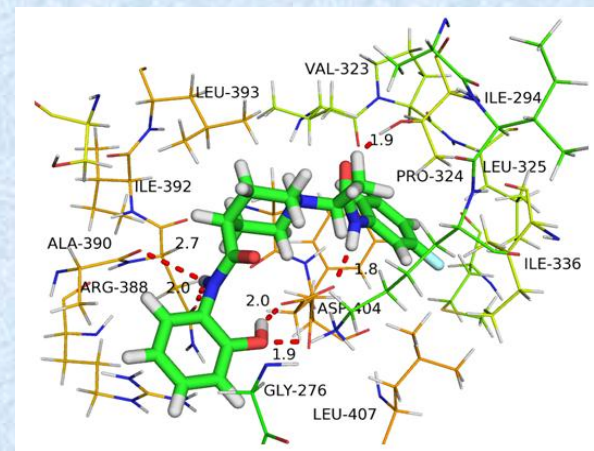
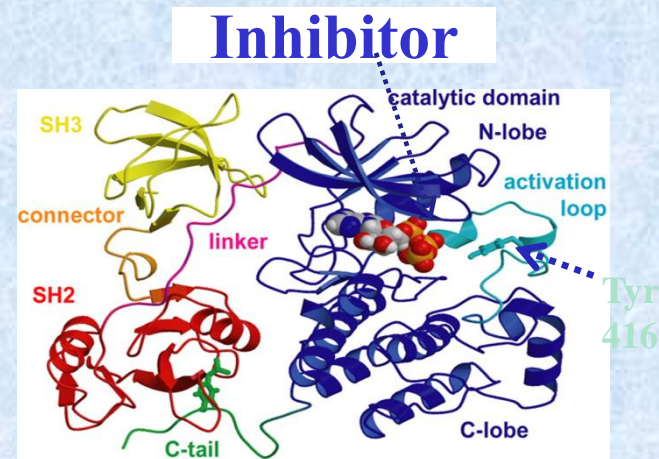
Identification: 120 compounds

Test: 50 compounds

Inhibition: 4 Compounds

Probable hit: 1 compound

2015: 200 compounds; inhibition: 5; hits: 2

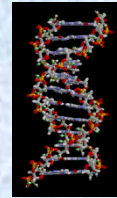


Complexity of biological systems

Nucleotide sequence

ATCGGGATAT

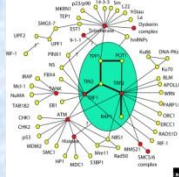
Nucleotide structure



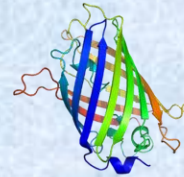
Gene expression

... GTGCATCTGACTCTGAGGAGAAG ... DNA
 (transcription)
 ↓
 ... GUGCAUCUGACUCCUGAGGAGAAG ... RNA
 (translation)
 ↓
 ... V H L T P E E K ... protein

Protein interactions



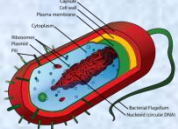
Protein structure



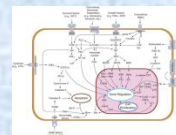
Protein sequence

ATCHKLVIYAT

Cell



Cell signaling



Tissues



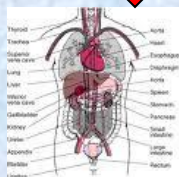
Organism



Physiology



Organs



The genetic material

DNA (deoxyribonucleic acid) is the genetic material.

The information stored in DNA allows the organisms that are able to regulate their internal chemical composition, growth and reproduction.

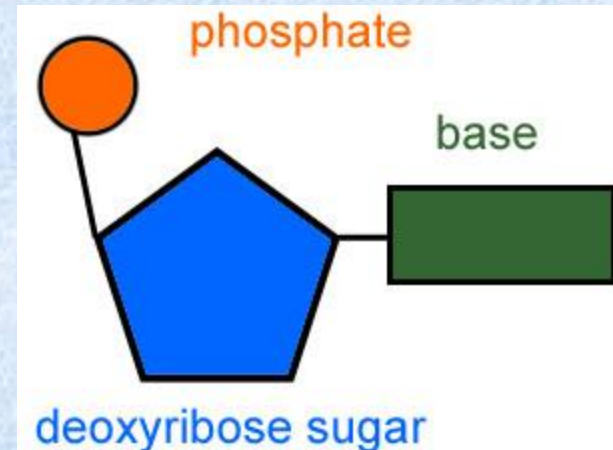
It allows to inherit our mother's curly hair, father's eyes etc.

The various units that govern those characteristics at the genetic level are called **genes**.

Genes contains specific sequence of nucleotides.

Nucleotide

- (i) phosphate group,
- (ii) central deoxyribose sugar and
- (iii) base

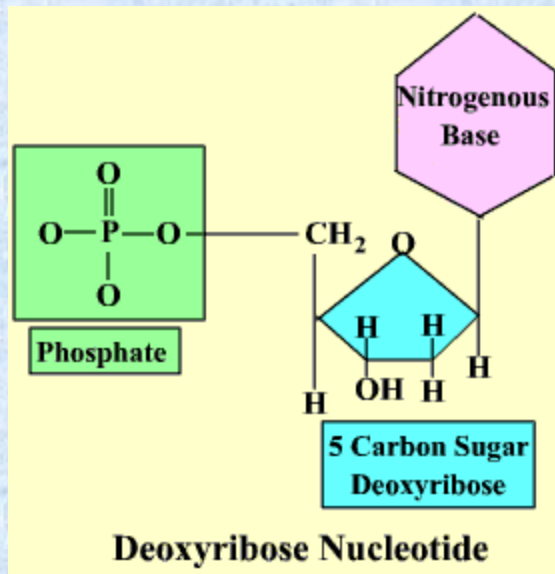


Four different bases

DNA is made and resides in the nucleus of living cells.

DNA gets its name from the sugar molecule contained in its backbone (deoxyribose);

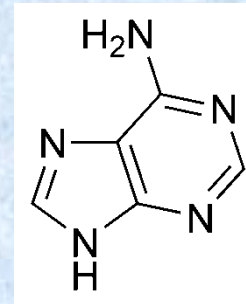
Four different nucleotide bases occur in DNA: **adenine (A), cytosine (C), guanine (G), and thymine (T).**



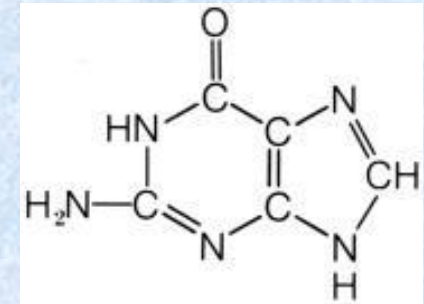
1. Methylation protects the DNA.
2. Maintains the fidelity of DNA replication.

Purines

The purines have a double-ring shape, with a pentagon attached to one side of a hexagon



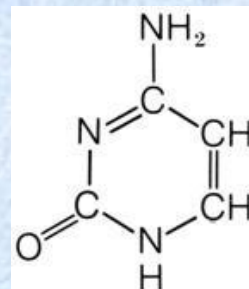
Adenine (A)



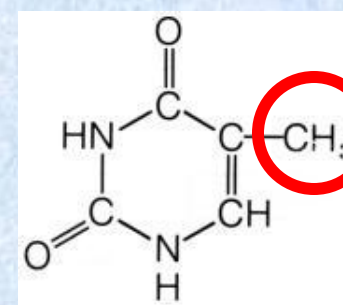
Guanine (G)

Pyrimidines

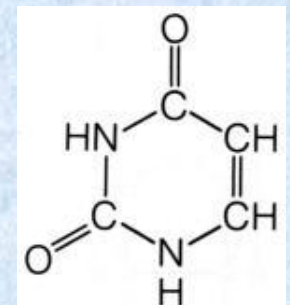
The pyrimidines have a hexagon shape, generally made with four carbons and two nitrogen



Cytosine (C)

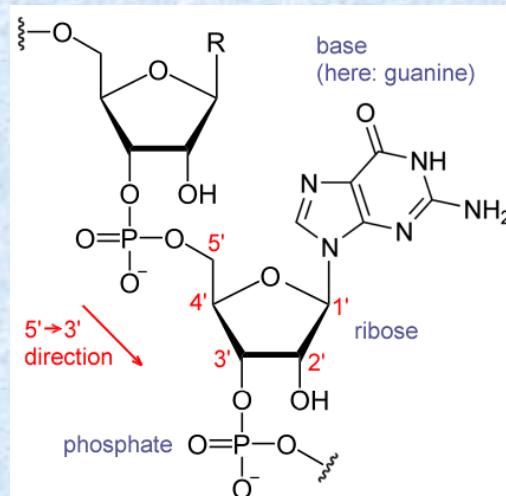
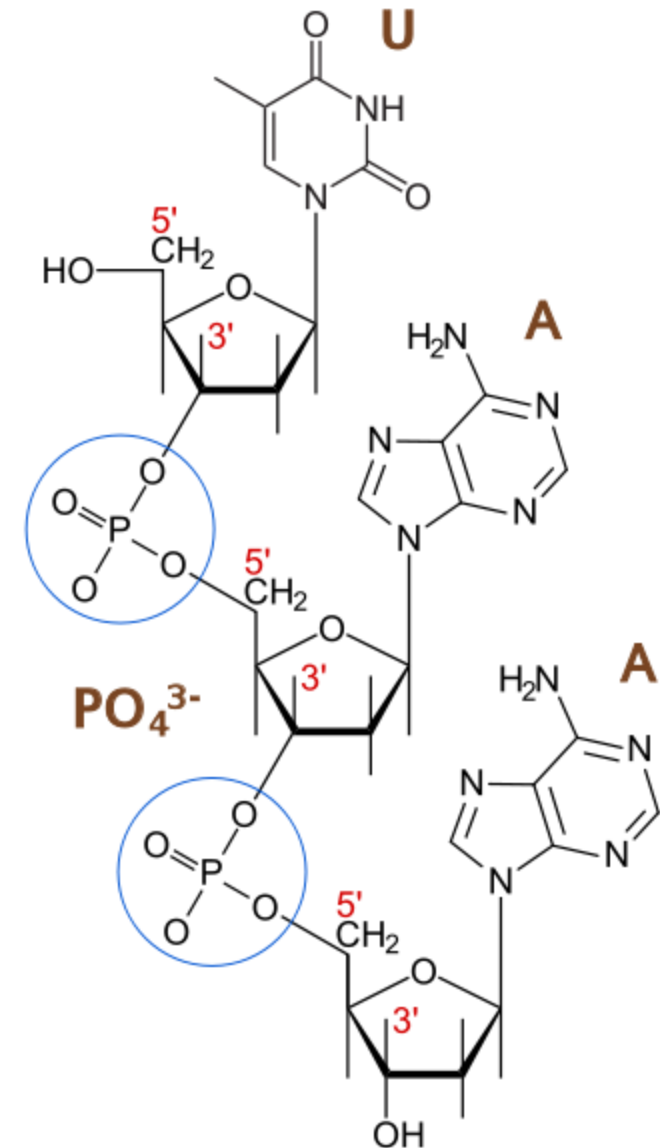
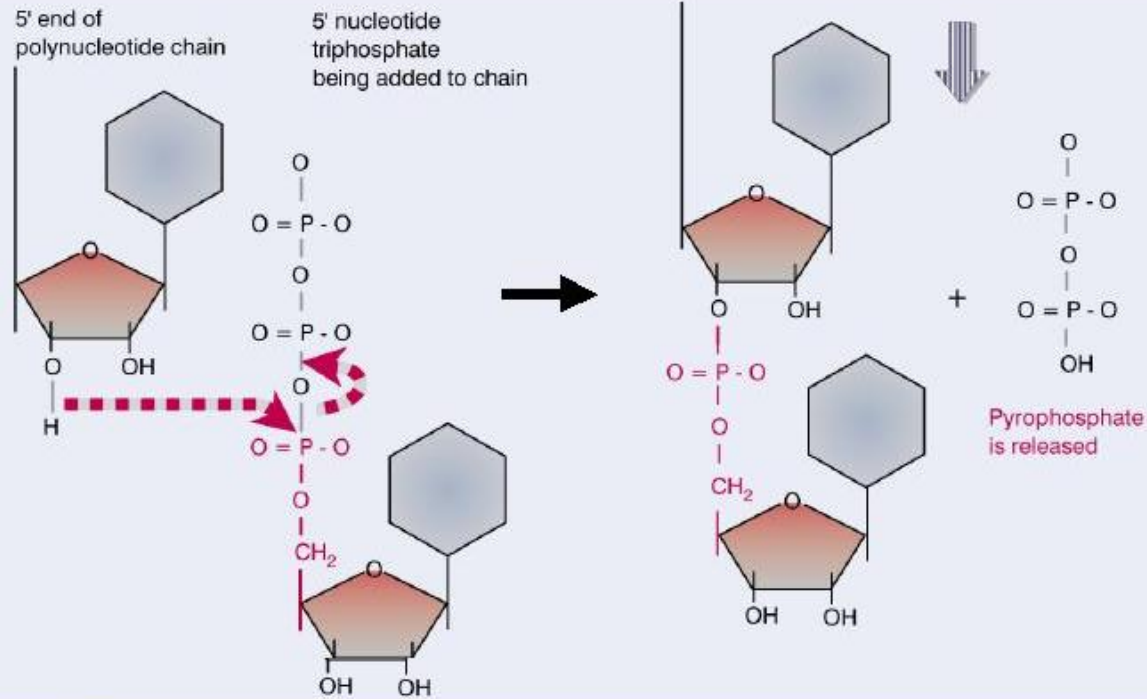


Thymine (T)



Uracil (U)

Phosphodiester bond formation



DNA

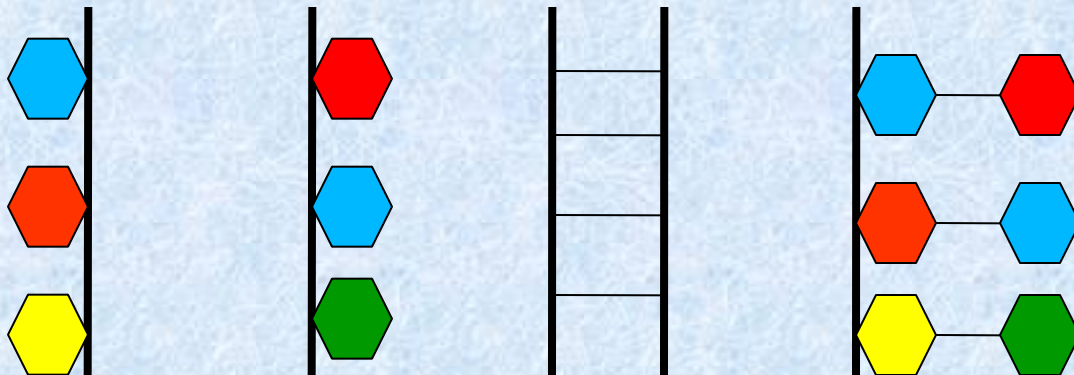
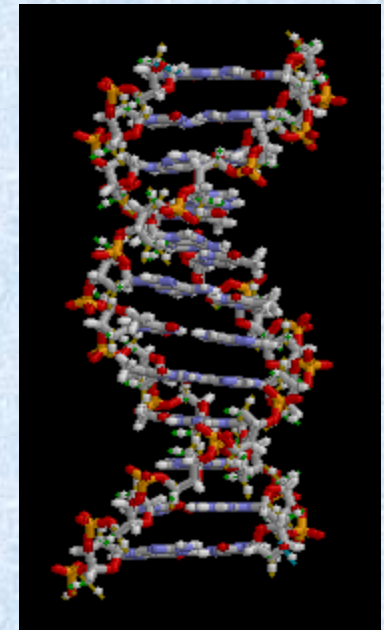
DNA molecules within cells typically exist as double-stranded molecules.

The information content on one of those strands is essentially redundant with the information on the other (1953).

The redundant strand is not the same – it is **complementary**.

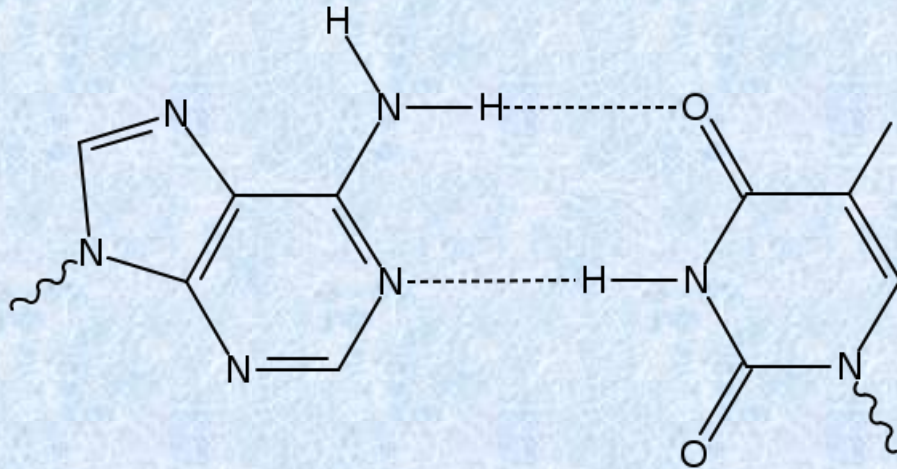
For every G on one strand, a C is found on its complementary and vice-versa

For every A on one strand, a T is found on its complementary and vice-versa



Watson and Crick, 1962, Nobel prize

DNA base pairing



Adenine

Thymine

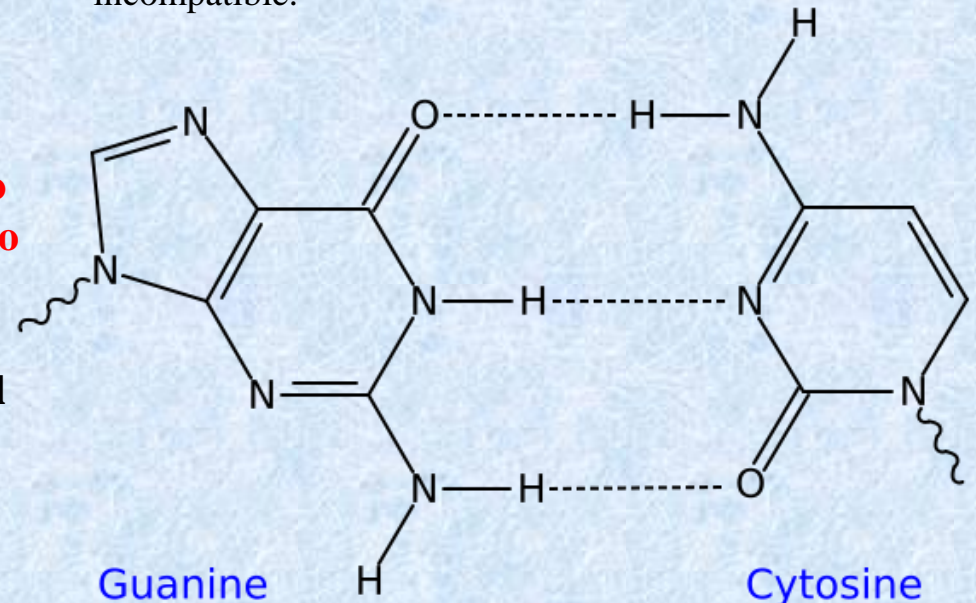
This bonding is specific, and **adenine always bonds to thymine** (and vice versa) and **guanine always bonds to cytosine** (and vice versa).

Right spacing and interaction between their chemical groups to form stable **base pairs**.

The nitrogenous base G with its two-ringed structure is simply too large to pair with a two-ringed A or another G in the space that usually exists between the two DNA strands.

Same way, it is too small between T and T/C.

Space is not a barrier to interaction between G's and T's or A's and C's but their chemical nature are incompatible.



Guanine

Cytosine