STATISTICAL PHYSICS OF BIOMACROMOLECULES

TENTATIVE

1. Introduction
2. Basic general observables and Simplest polymer model :

   FREELY JOINTED CHAIN
3. Polymer with local rigidity : KRATKY-POROD & WORM-LIKE CHAIN
4. Self avoiding random walk : FRACTAL DIMENSION, FLORY APPROXIMATION & MAPPING ON CRITICAL SYSTEMS
5. Nature of interactions in aqueous solutions
6. Polymer collapse
7. Proteins

1. Introduction

What is a polymer ?

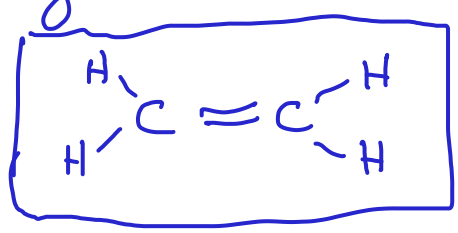
"Polymers" are chains of "monomers". Monomers are some basic building blocks that have the ability to link to each other.

Although the monomers in a polymer are typically very similar to each other, monomers in different polymers can be vastly different.

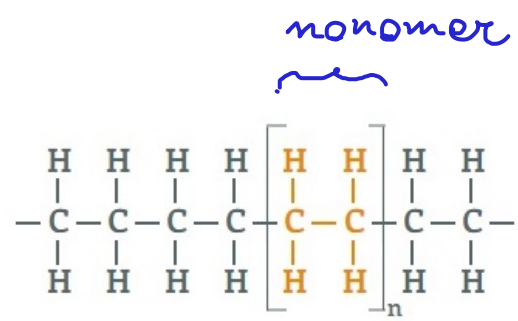
Exemples :

PE : polyethylene

Ethylene :

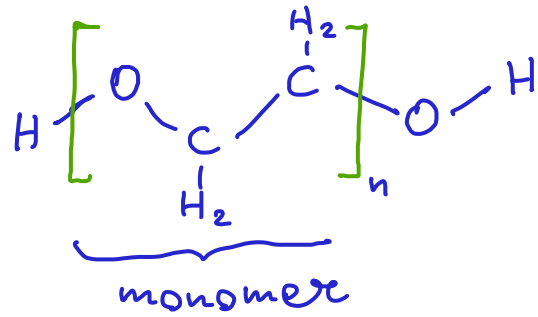


poly → (break the double bond)



Molecular Structure of Polyethylene

PEG : polyethylene glycol



PET

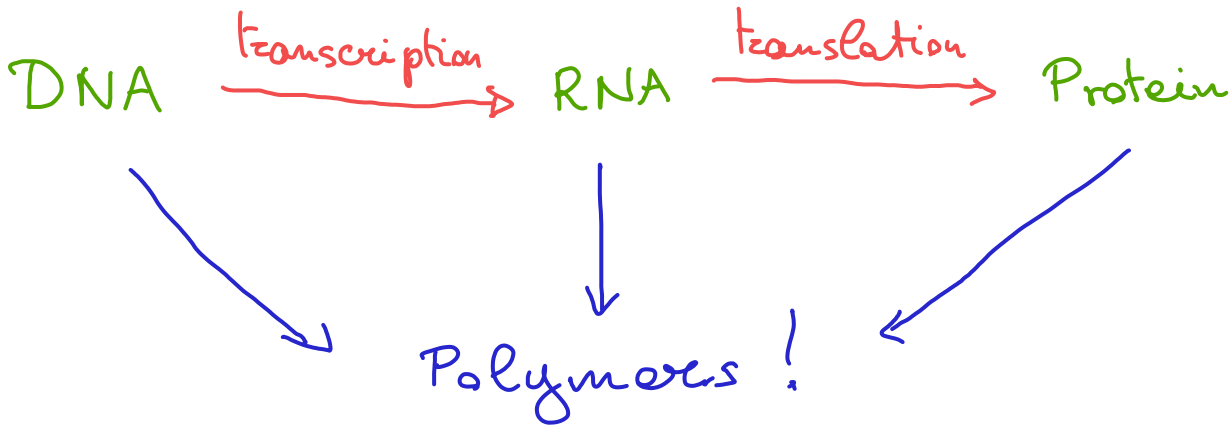
PE, PEG and PET are among the building blocks of most plastics and rubbers (of course the chemical industry has invented many many others!).

This is the reason why the first developments in the statistical physics of polymers have been done by people with strong interests in (and financial backing from) the chemical industry.

In a biological setting, which is the one that we are mostly interested in here, polymers are ubiquitous.

Let me just recall to you the "central dogma" of biology:

Information is stored in DNA, it is transcribed in RNA which is ultimately translated into proteins.



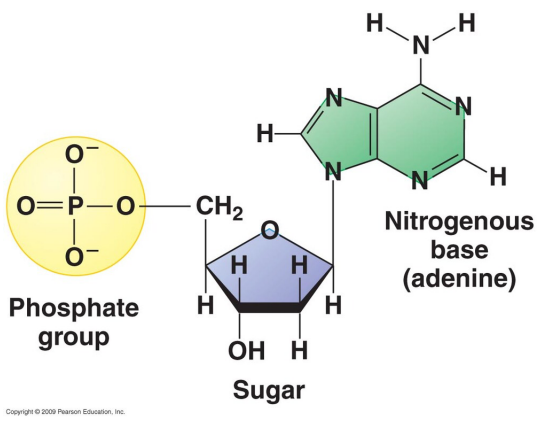
We are going to address these processes in greater detail during the course. Enough to say that what is considered as one of the laws of biology has to do with polymers from start to end.

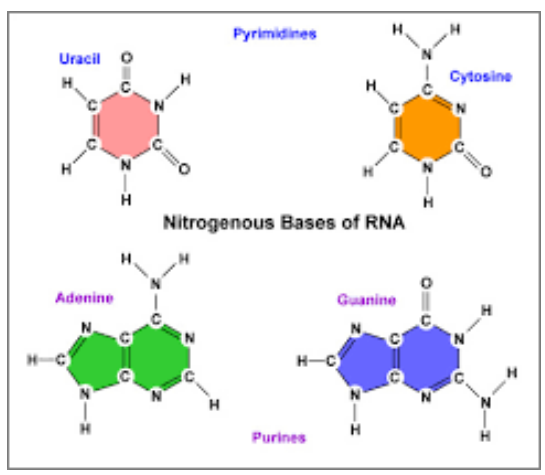
Let us look at these three classes of polymers in more chemical detail.

DNA :

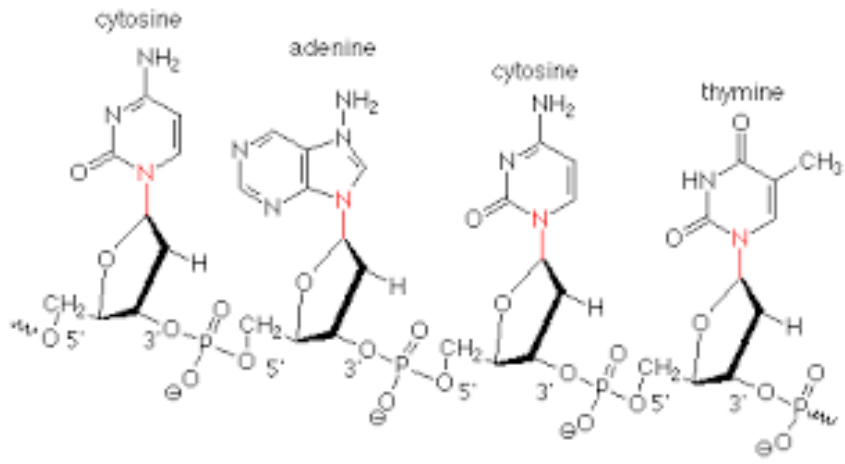
The fondamental biological information (the "blueprint" of the cell) is stored in DNA.

DNA is made of four basic building blocks, the four bases

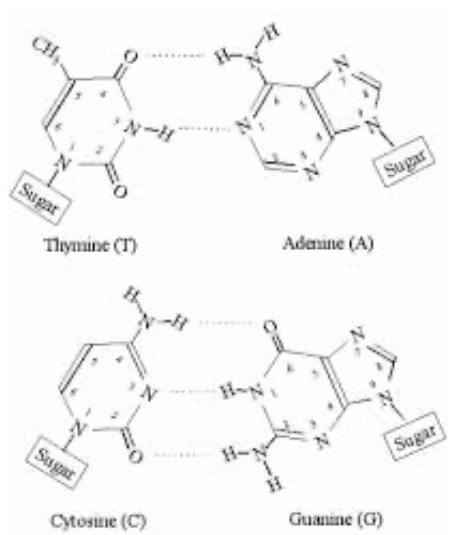




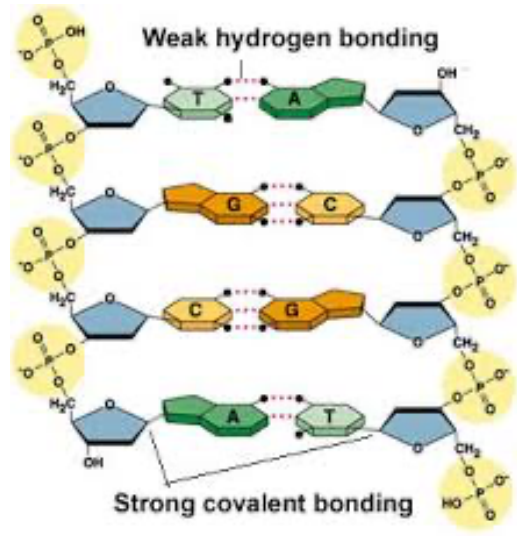
These bases are first concatenated in a precise order (otherwise there would be no information!) by covalently binding one to the other



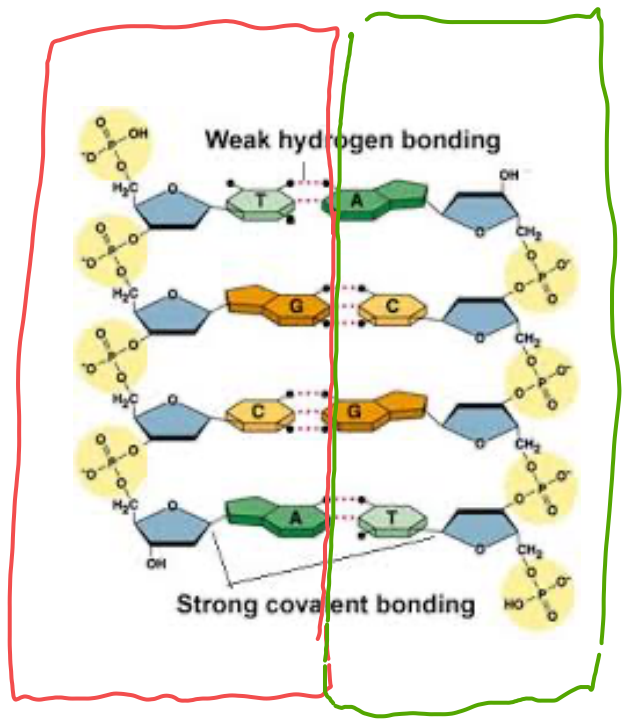
This process composes a long polymer, so called "single-strand" DNA (ssDNA). ssDNA is thus a polymer made of different monomers (it is a "heteropolymer", whereas polymers made by identical units are called "homopolymers").
Actually, DNA is more complex. Indeed its structure is known as "double helix". Why "double"?
Because it is composed by two ssDNA polymers, that bind to each other using what is called "Watson-Crick pairing", namely, there is an energetic favorability of having A in front of T, and C in front of G



because of optimal hydrogen-bond pairing (we will see it later in the course)

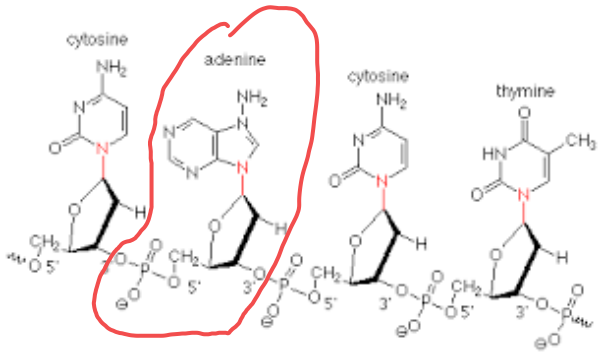


The final result is that there are two polymers, two ssDNA strands, which are complementary (each T, A, C, G on one strand is paired to a A, T, G, C on the other, respectively).

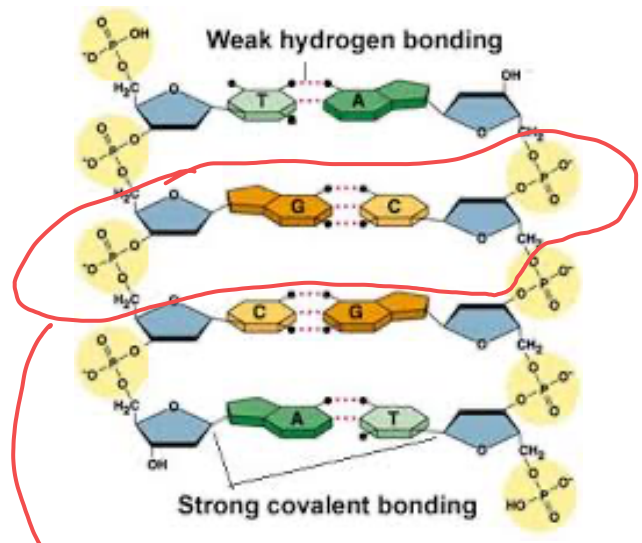


Together, these two ssDNAs form a double-stranded DNA (dsDNA), which is, chemically, the DNA that you find in cells

Also dsDNA is, in itself, a polymer.

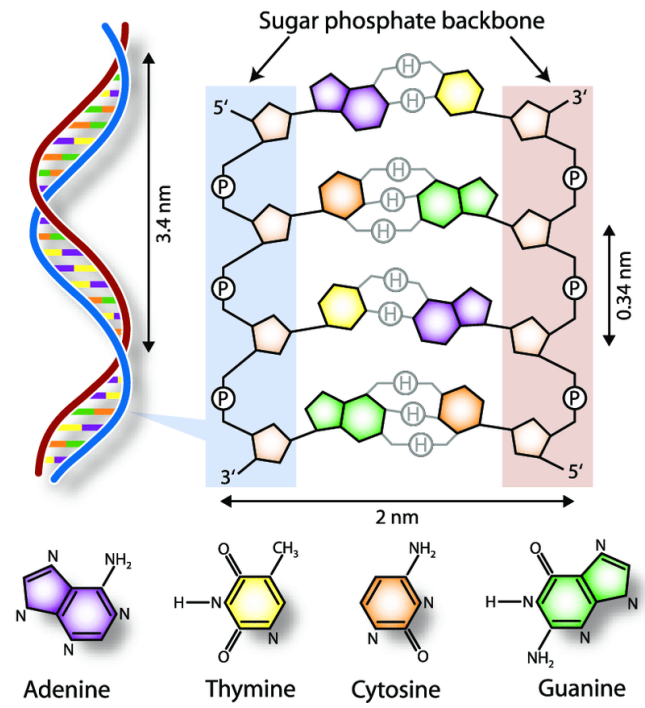


monomer for ssDNA



monomer for dsDNA: it is the "base-pair"

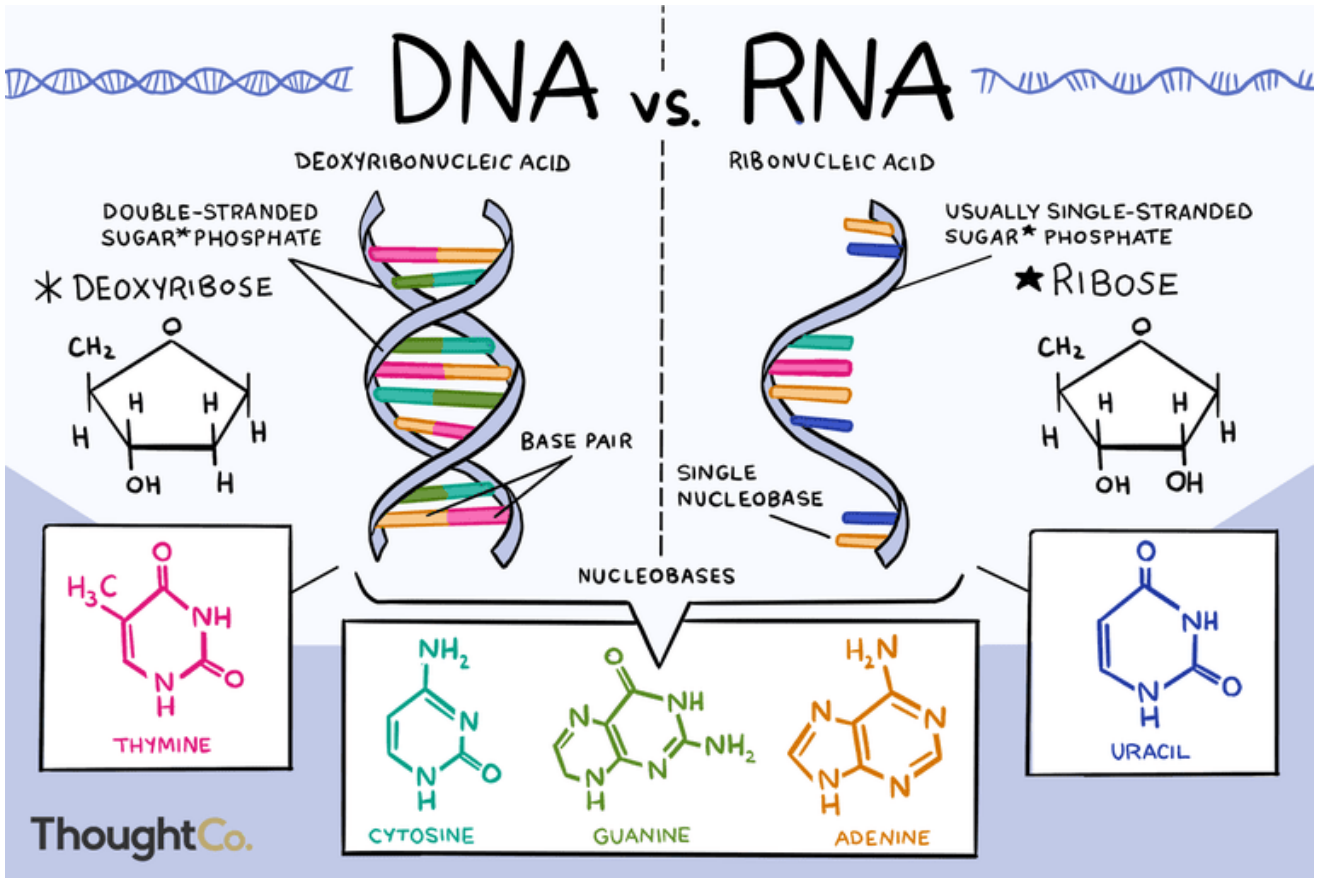
dsDNA then actually locally organizes into a "double helix" [experiments with X-Ray crystallograpy by Rosalind Franklin, interpreted as double helix by Francis Crick and James Watson, who won the Nobel Prize in 1962. Franklin could not win it, because she died in 1958. Yet her role has been acknowledged only many years later.]



The double helix satisfies both the properties of the covalent bonds, that come from molecular orbitals that are not isotropic (actually are anisotropic and chiral) and the packing properties of the bases ("base stacking").

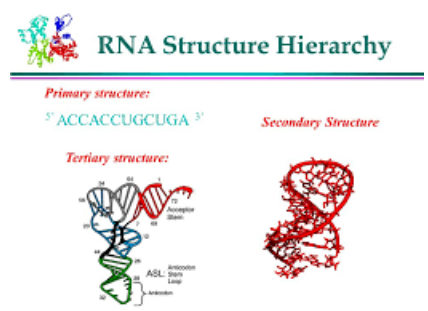
RNA:

In the process of expression, a gene (the "code" for a protein) must first be transcribed from DNA to RNA. RNA is almost identical to DNA, although the bases are slightly different:



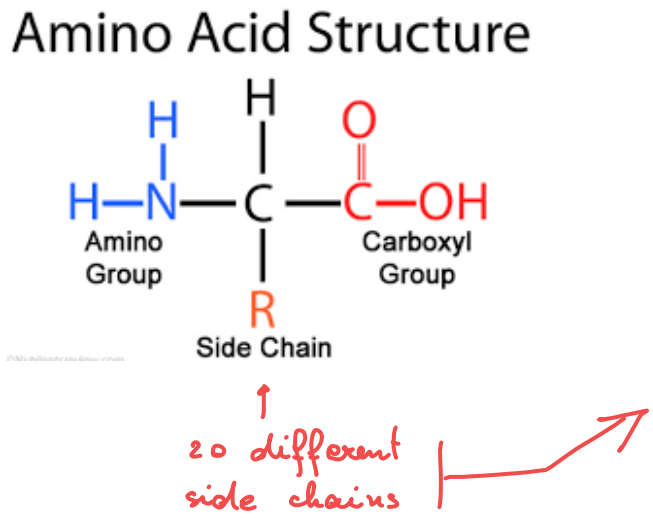
The process is called transcription because RNA is essentially identical to DNA (it is a "transcript" with the same alphabet/language.

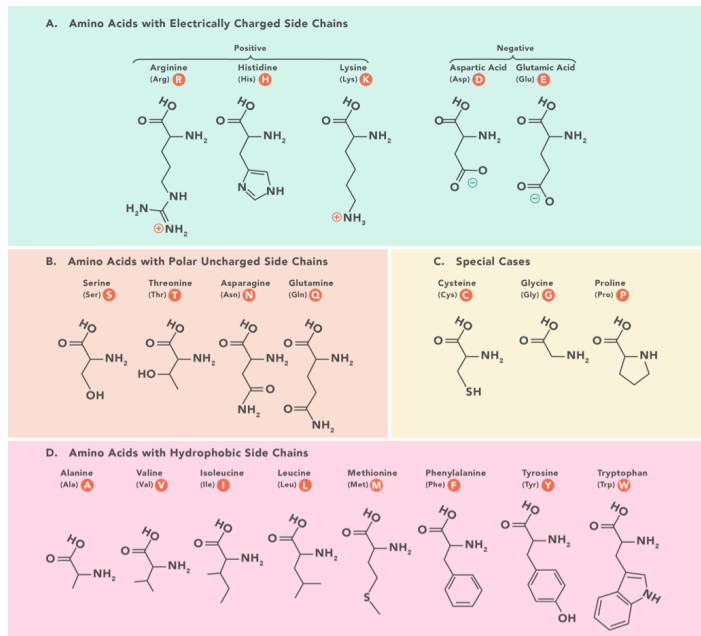
RNA is predominantly single-stranded, unless there is complementarity (A-U, C-G) in different parts of the same strand, and so it can fold on itself and form secondary structures



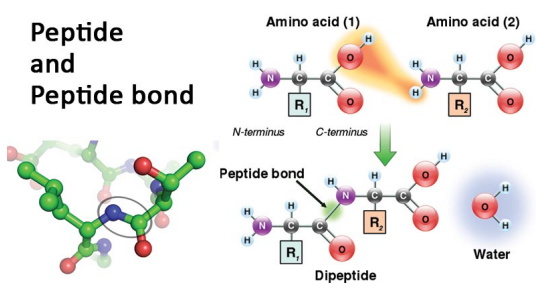
Proteins :

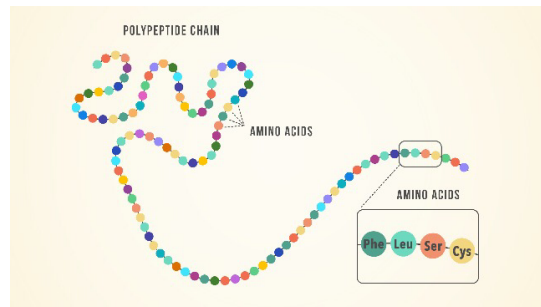
The end-product of gene expression, as captured by the CENTRAL DOGMA, are proteins. The production of proteins from RNA is called "translation" because proteins are polymers made of a different alphabet : the amino-acids.
There are 20 biologically relevant aminoacids :





Their polymerisation takes place by "peptide bond" formation.





How is the language of RNA translated in the language of aminoacids ? By codons.

Let's make a simple calculation:

There are 20 amino-acids. There are 4 bases:

1 base cannot code 20 amino-acids

2 bases can code $16 = 4^2$ amino-acids, not 20

3 bases can code $4^3 = 64$ symbols > 20

One codon = 3 bases

1st letter	Second Letter: U	Second Letter: C	Second Letter: A	Second Letter: G	3rd letter
U	UUU Phe	UCU Ser	UAU Tyr	UGU Cys	U
	UUC Phe	UCC Ser	UAC Tyr	UGC Cys	C
	UUA Leu	UCA Ser	UAA Stop	UGA Stop	A
	UUG Leu	UCG Ser	UAG Stop	UGG Trp	G
C	CUU Leu	CCU Pro	CAU His	CGU Arg	U
	CUC Leu	CCC Pro	CAC His	CGC Arg	C
	CUA Leu	CCA Pro	CAA Gln	CGA Arg	A
	CUG Leu	CCG Pro	CAG Gln	CGG Arg	G
A	AUU Ile	ACU Thr	AAU Asn	AGU Ser	U
	AUC Ile	ACC Thr	AAC Asn	AGC Ser	C
	AUA Ile	ACA Thr	AAA Lys	AGA Arg	A
	AUG Met	ACG Thr	AAG Lys	AGG Arg	G
G	GUU Val	GCU Ala	GAU Asp	GGU Gly	U
	GUC Val	GCC Ala	GAC Asp	GGC Gly	C
	GUA Val	GCA Ala	GAA Glu	GGA Gly	A
	GUG Val	GCG Ala	GAG Glu	GGG Gly	G

Each amino-acid can be coded by multiple codons, and three codons are Stop codons that indicate the end of the region to be translated

We are going to deal later in this course with the machines that transcribe DNA into RNA and that translate RNA into proteins.

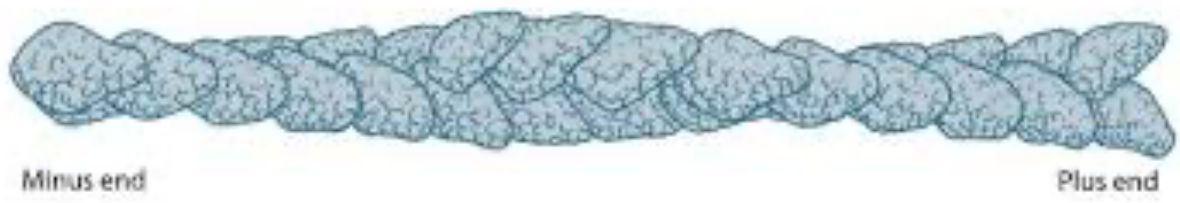
Proteins have a rather complex life: from their synthesis at the ribosome, they have to fold to a specific conformation, and they will have to be unfolded for degradation or for transport.

A lot of bad things can happen along the way.

We will look at that later on in the course.

There are polymers that use as monomers entire proteins:
actin filaments, microtubules, flagella

Actin filaments



Actin Filament

Microtubules

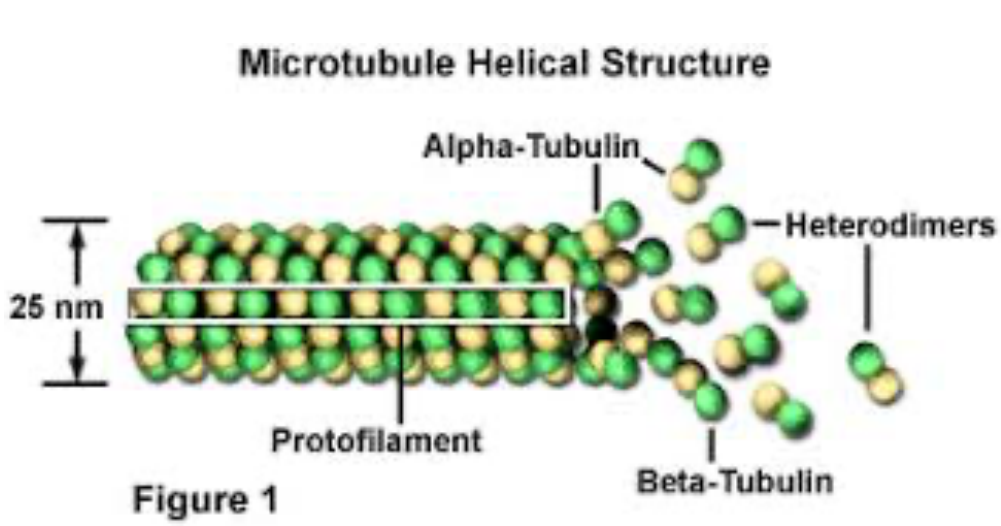
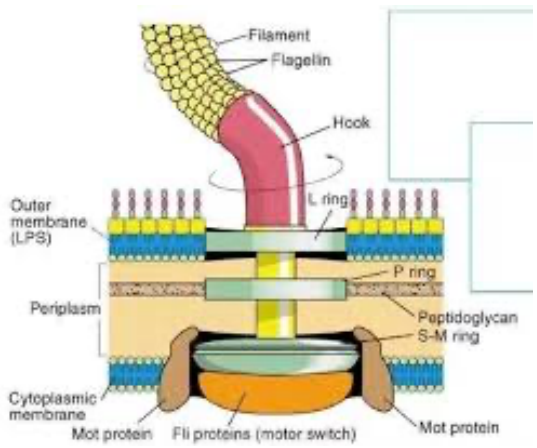
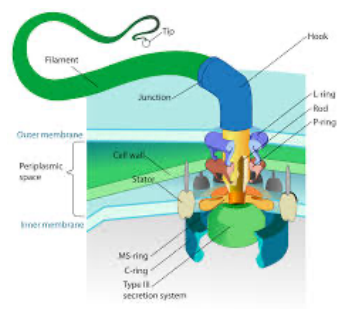


Figure 1

Flagella



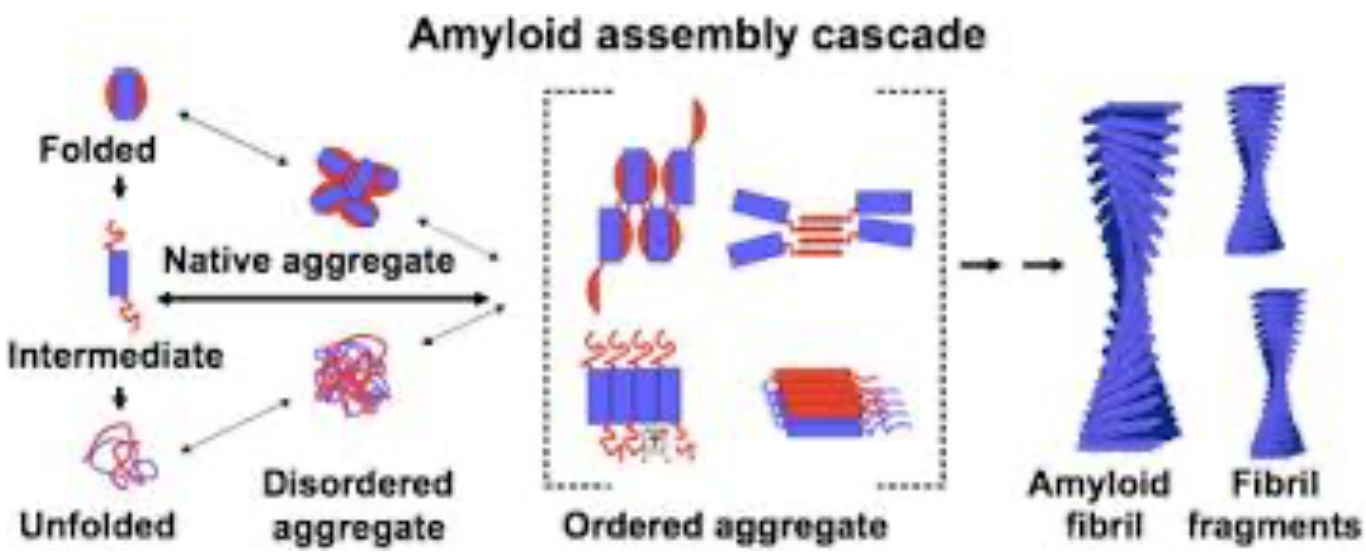


In all these polymers the monomers (proteins themselves) bind to each other not by covalent bonds but by other forms of interactions (electrostatics, hydrophobic, etc) that we are going to study later in the course.

Toxic biological polymers

Proteins can also wrongly polymerize in the cell!

Protein aggregates in the form of fibers correlate (which is different to "cause") several neurodegenerative diseases, such as Alzheimer, Parkinson, Huntington etc.



Conclusions

Polymers are long one-dimensional macromolecules made by the concatenation of monomers. They are of different sizes (amino-acid ≈ 4 Å ; actin monomer ≈ 4 nm) and features. They represent the major class of biological macromolecules.