MOLECULAR BIOLOGY OF THE CELL

Giacomo Castagnetti – Genomics – 2021/2022

INDEX

- DNA
    - Structure
        - Backbone
        - Nucleotides
        - Nucleosides
    - Nomenclature
    - Polymer
    - Tautomers
    - Energy carriers
    - B-DNA
    - A-DNA
    - Z-DNA
    - Supercoiling
    - Chromatin
        - Bacteria
        - Eukaryotes
            - Heterochromatin
            - Karyotype
            - Chromosomes
- RNA
    - Chemical modifications
    - Hydroxyl group
        - dsRNA
    - RNA folding
        - Covariations
    - Non-Watson-Crick base pairs
        - Wobble base pair
        - Hoogsteen base pair
    - Stabilisation of the tertiary structure
        - Divalent cations
        - Coaxial stacking
        - Ribose zipper
        - Triple helixes
    - Complexity
- Cell cycle
    - Phases
    - CdKs (cyclin-dependent kinases)
        - Phosphorylation
        - Cyclins
        - Decision making
        - Complexes
        - Activation
        - Inhibitors
        - Network of switches
    - Checkpoints
    - Transcription of specific genes
        - Cancer

- DNA replication
    - Phases
        - Initiation
        - Elongation
        - Termination
    - Effectors
        - DNA polymerase
            - Structure
            - Accuracy
            - Processivity
            - Types of DNA polymerase
        - Helicases
            - Structure
        - ssDNA binding proteins
        - Topoisomerases
        - Slider clamp
            - Structure
        - Clamp loader
    - Origin of replication
        - Bacteria (E. coli)
        - Eukaryotes
    - Primases
        - Bacteria
        - Eukaryotes
    - Okazaki fragments
        - Bacteria
        - Eukaryotes
    - Replication fork
        - Bacteria
        - Eukaryotes
    - Termination
        - Bacteria
        - Eukaryotes
            - The end-replication problems
    - Regulation of replication
        - Bacteria
        - Eukaryotes
- Mitosis
    - Phases
        - Prophase
        - Prometaphase
        - Metaphase
        - Anaphase
        - Telophase
        - Cytokinesis
    - Cohesins
    - M-CdK

    - Condensin
    - Mitotic spindle
        - Structure
        - Centrosomes
            - Centrioles
        - Motor proteins
        - Self-organisation
        - Kinetochore
            - Bi-orientation and tension
            - Sensing tension
    - APC/C checkpoint
- Transcription
    - Overlook
        - Phases
            - Initiation
            - Elongation
            - Termination
    - Eukaryotic transcription
        - Effectors
    - Bacterial transcription
        - Effectors
            - α C and N terminal
        - Operons
        - Types of promoters
        - Important convention
        - Initiation
            - Sigma factors
            - Promoter
                - Consensus sequences
                - Spacing between boxes
                - Strength of a promoter
            - Regulation on sigma factors
                - Pro-sigma factors
                - Anti-sigma factors
            - Abortive initiation
        - Elongation
            - Errors
            - Supercoils
        - Termination
- Regulation of transcription
    - Activators and repressors
        - Cis elements
        - Trans elements
    - Sequence readout in regulation
        - Specificity
        - Affinity
        - Spacing between consensus

        - Sequence arrangement
        - Domains used to bind DNA
            - Helix-turn-helix (HTH)
            - Zinc fingers
            - Basic leucine zipper
            - Basic helix-loop-helix (bHLH)
            - Beta readout
            - Loop readout
    - Examples of regulation pathways in bacteria
        - Trp (tryptophan) repressor
        - Cap activator
        - LacI repressor
            - Logic gate
        - MerR activator
        - NtrC enhancer
        - AraC
        - Transcriptional attenuation (co-translational regulation)
            - Trp operon
    - Signal transduction
    - Eukaryotic signalling
        - NF-kB
    - Transcription motives
        - Positive feedback loop
        - Negative feedback loop
        - Flip-flop devices
        - Feed forward loop
- RNA processing
    - Cleavage
        - RNAse III
        - RNAse P
        - tRNA processing
            - Modification of bases
    - Capping
    - Polyadenylation
        - Coupling
    - RNA editing
        - Deamination
    - RNA decay
        - Bacteria
        - Eukaryotes
- Translation
    - Process
    - tRNA
        - Modifications
    - Genetic code
        - Reading frames
        - Codon usage

- Aminoacyl-tRNA
  - tRNA synthetases classes
  - Transamidases and desulfurases
- Ribosome
  - A-P-E
- Phases
  - Initiation
  - Elongation
  - Termination
- Polysomes
- Hybrid state
- Translation factors
  - GTPases
    - GAPs and GEFs
  - Molecular mimicry
- Initiation
    - Bacteria
    - Eukaryotes
  - Initiator tRNA
    - Bacteria
    - Eukaryotes
  - Start site
    - Bacteria
      - Initiation factors
    - Eukaryotes
      - Pre-initiation complex
      - Initiation complex
- Elongation
  - Bacteria
    - Cognate amino acid
      - Minihelix
    - Peptide bond
      - Translocation
- Termination
  - Class I release factors
  - Class II release factors
- Circularisation

- Regulation of translation
  - UTR regulation
    - 5’ UTR
    - 3’ UTR
    - Internal ribosome entry sites (IRES)
  - mRNA stability
  - Ribosome stall
  - Recoding
    - Nonsense suppression
      - Incorporation of nonstandard amino acids

        - Frameshifting
            - +1 frameshift
            - -1 frameshift
        - Antibiotics
    - Co-translational protein folding
    - Post-translational regulation
        - Molecular chaperone
            - Hsp70
            - Hsp60
        - Proteasome
- Regulatory RNAs or riboregulation
    - Specificity of regulatory RNAs
        - Trans-regulation
        - Cis-regulation
            - Riboswitches
            - CRISPR
    - Eukaryotic small RNAs
        - siRNA pathway
        - miRNA pathway

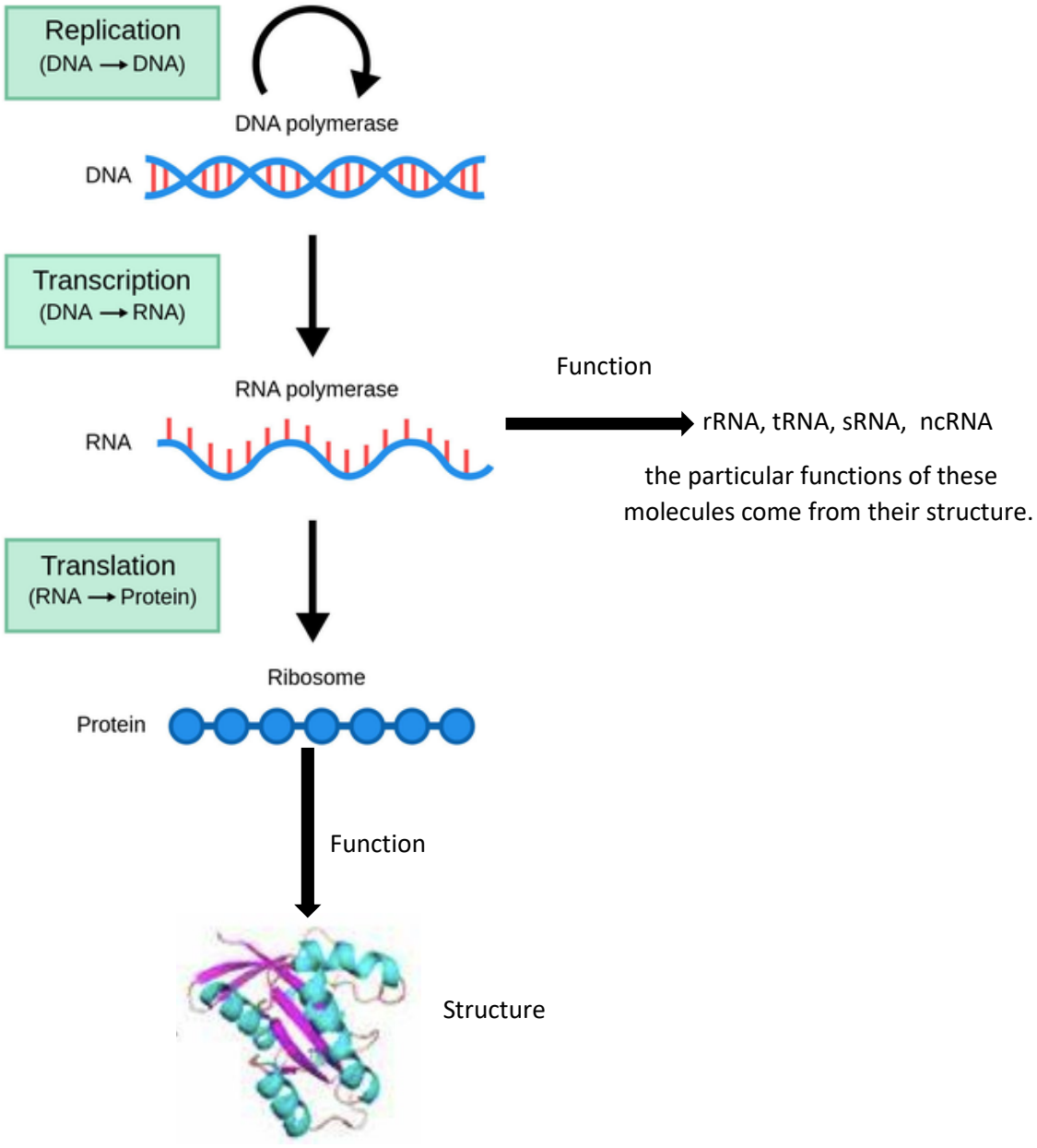
INTRODUCTION

Molecular biology consists in the study of how biomolecular mechanisms work, not why.

Living organisms

1. All living organisms are made of cells and have a metabolism
2. All living organisms are born, grow and eventually die
3. Have a physical barrier that separates the internal and external environment
4. Have all the molecular building blocks
5. The genetic information is stored in the genome
6. Present gene expression mechanisms
7. Have molecular processes that consent gene expression (transcription and translation)
8. Can replicate their genome
9. Living organisms require a source of energy and respect the thermodynamics principles

Life is a flux of information



Biomolecules

The biomolecules are the building blocks of life, 70% of a cell is made of water, then the most abundant macromolecules are proteins, followed by lipids sugars and nucleic acids.
every biomolecule is in the order of magnitude of 1nm.

- **DNA**: very stable molecule, perfect to store information; duplication of DNA is a semi-conservative process, following the central dogma of molecular biology DNA is transcribed in RNA.
- **RNA**: very reactive molecule, perfect to satisfy very specific functions, the additional hydroxyl group on carbon 5 makes RNA very labile; depending on the structure of the transcript we get different functions.
- **Proteins**: through the genetic code we can go from RNA to proteins (translation)

Each step of the flow of information is highly regulated, both during and after every phase regulatory factors and enzymes act to change the expression of the genes (transcriptional, post-transcriptional, translational, post-translational); this mechanism is the base for cellular differentiation in pluricellular organisms and of evolution: organisms regulate their gene expression in response to environmental signals.

Mutations

Mutations are changes in the genetic sequence of an individual that is transmitted to the next generation (hereditable).
Mutations can cause changes to the phenotype or not, silent mutations are the most common, because they affect non structural genes that are more prone to keep these changes throughout the generations, in this way mutations accumulate in the genetic information of species.

The evolutionary distance of two individuals of the same species is generally in constant enlargement and, when it becomes too wide or when specific mutations change the interaction between the individuals and the environment, the spread of a gene in the population gets affected and it could take over to the whole population or there could be the formation of a new species.

The equilibrium of the alleles of a gene in a population is regulated by the rules of natural selection (survival of the fittest).

Mutations in important genes usually cause the loss of function of the gene and so there are not compatible with life, in general, living organisms tend to conserve the function of important gene: the more a gene is conserved, the more it is important, that's why we can find the same genes among very genetically distant individuals, the most efficient genes for life are kept safe and lay the bases for evolution.

Phylogenetic trees

If we look at phylogenetic genes the special distance between the branches can give us an idea of the genetic distance.

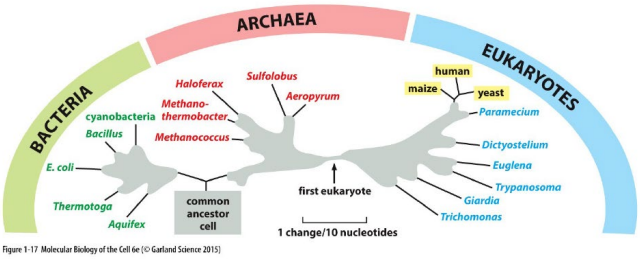


Figure 1-17 Molecular Biology of the Cell 6e (© Garland Science 2015)

the 3 realms have almost no common genes, the similarity of DNA sequences tells us how related organisms are, the more similar the sequence the more closely related.

Thanks to the genetic diversity a population can react better to the changes of the environment, there is more chance that a slice of the population will be stronger in the new conditions.

Sources of mutations

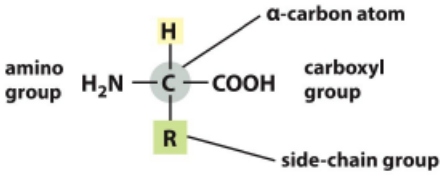
- Intragenic mutation: a point mutation causes the change of function of a gene.
- Gene duplication: a gene is copied, 1 copy will keep its function while the other accumulates mutations.
    - after a small amount of time after the duplication the 2 copies are still very similar, they can be considered homologs and usually they are present in organisms of the same species; they are defined as paralogs.
    - After a lot of time the 2 copies are related but they are different; if they belong tgo different species they are called orthologs.
- DNA segment shuffling: exchange of genetic material between two genes.
- Horizontal gene transfer: transfer of genetic information between two living individuals, it causes the transformation of the organism.
- Vertical gene transfer: recombination of genes in the filial generation due to sexual reproduction.

Model organisms

By studying molecular biology of bacteria we can learn concepts true also for human chromosomes (we have the same building blocks); model organisms consent to get information for a vast set of organisms.
We exploit the presence of orthologs to know the function of certain genes in organisms similar to those under experimentation (That is to say that we can predict the function of a gene by its sequence).

Proteins

Amino acids are the building blocks of proteins:



AMINO ACID			SIDE CHAIN
Aspartic acid	Asp	D	negative
Glutamic acid	Glu	E	negative
Arginine	Arg	R	positive
Lysine	Lys	K	positive
Histidine	His	H	positive
Asparagine	Asn	N	uncharged polar
Glutamine	Gln	Q	uncharged polar
Serine	Ser	S	uncharged polar
Threonine	Thr	T	uncharged polar
Tyrosine	Tyr	Y	uncharged polar

POLAR AMINO ACIDS

AMINO ACID			SIDE CHAIN
Alanine	Ala	A	nonpolar
Glycine	Gly	G	nonpolar
Valine	Val	V	nonpolar
Leucine	Leu	L	nonpolar
Isoleucine	Ile	I	nonpolar
Proline	Pro	P	nonpolar
Phenylalanine	Phe	F	nonpolar
Methionine	Met	M	nonpolar
Tryptophan	Trp	W	nonpolar
Cysteine	Cys	C	nonpolar

NONPOLAR AMINO ACIDS

amino acids are joined by peptide bonds to form polypeptides.

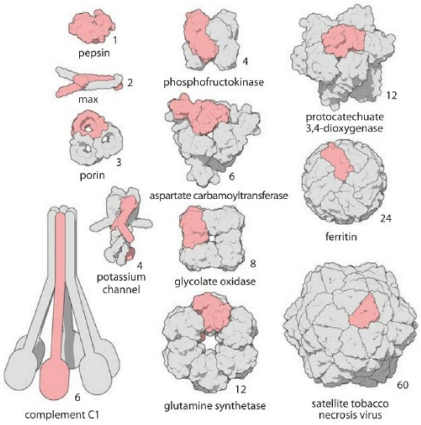
The interactions of the sidechains of a polypeptide determine its conformation and therefore its function.

The levels of organisation of an amino acid sequence are:

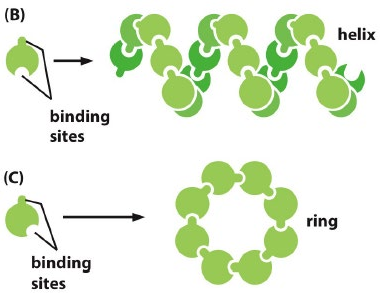
- **Primary structure**: linear sequence of amino acids
- **Secondary structure**: polypeptide backbone can form an alpha helix structure (right ended helix) or beta sheets (planar structures, made by antiparallel alignment of polypeptides); in these structure are mainly involved hydrogen bonds and hydrophilic and hydrophobic

interactions (in the centre of the polypeptide are concentrated the hydrophobic elements while the hydrophilic ones are on the surface, in contact with water).
amphipathic alpha helices can form coiled coils, to minimise the exposure of hydrophobic amino acid side chains to aqueous environment.

- **Tertiary structure**: the secondary structures of the same polypeptide can interact with each other to form a three dimensional shape, the conformation of the protein.
  Folding is driven by non-covalent interactions of the atoms in the polypeptide and, usually, polypeptides are characterised by compact regions that can fold on their own and that fulfil a specific function; domains can be considered as building blocks for proteins, a specific combination of domains affords a new function.
  Once a protein has evolved to a stable fold with useful properties, its structure could be modified during evolution to enable it to perform new functions, this process is called divergent evolution; in rare cases two proteins can reach the same conformation and function, even if they come from very different genes, this usually occurs because the molecular mechanism result to be very efficient and convenient for the survival of the organism, we call this convergent evolution.
  Through experimental evidences we can say that proteins with more than 25% sequence identity usually share the same function (this is useful to identify orthologs).
- **Quaternary structure**: the combination of multiple polypeptides in a single protein affords the final form of the protein.
  A protein formed by identical polypeptides is called homo-oligomer while a protein formed by different subunits is called hetero-oligomer.
  Usually proteins and their subunits have a globular form and while the single polypeptides have a dimension in the order of magnitude of 1nm a protein is usually around 5nm.
  Examples of quaternary structures:



polypeptides interact thorough their surface, so this is the principal factor that determines their arrangement, if a monomer has 2 binding sites the quaternary structure can result in different forms depending on the relative position of the two binding sites:



moreover the position of the binding sites on the single polypeptide depends on its folding, so a slightly different folding procedure can afford very different quaternary structures

Functions of the proteins:

Enzymes

Enzymes catalyse reactions that make or break covalent bonds, they direct the vast majority of chemical processes in the cell.
When they are used to make bonds they have an anabolic function while when they break bonds they have a catabolic function.

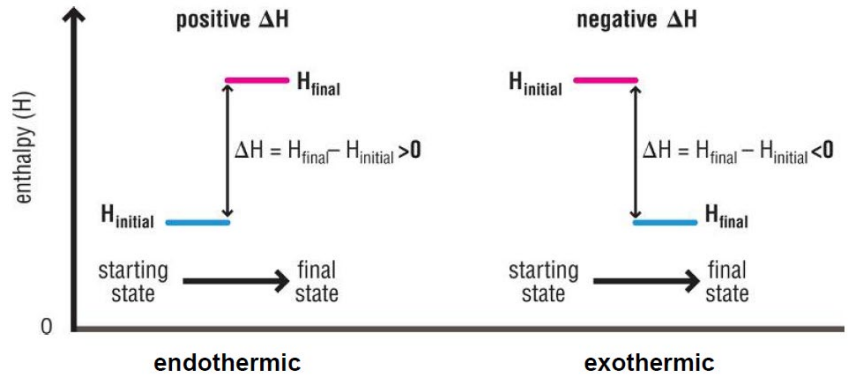
Enzymes act by lowering the activation energy of a reaction, to go deeper into their mechanisms we need some previous concepts

- **Thermodynamics**

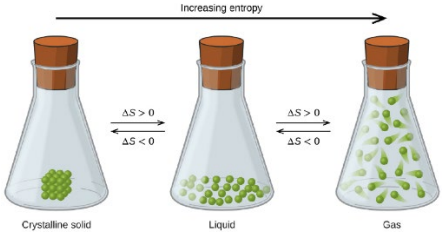
  Types of systems:
  Open system: both energy and mass can be exchanged between the system and the environment.
  Closed system: energy, but not mass, can be exchanged between the system and the environment.
  Isolated system: neither energy nor mass can be exchanged between the system and the environment.

  First principle of thermodynamics: The internal energy U of an isolated system is constant

  Enthalpy (H)
  Enthalpy is defined as: $\boldsymbol{H=U+PV}$, it is a state function and its variation determines the likelihood of a process occurring.
  Exothermic reaction: heat is transferred from the system to the environment $\Delta \boldsymbol{H}<0$.
  Endothermic reaction: heat is transferred from the environment to the system $\Delta \boldsymbol{H}>0$.



  Entropy (S)
  Entropy is defined as the molecular disorder of a system, it is a state function



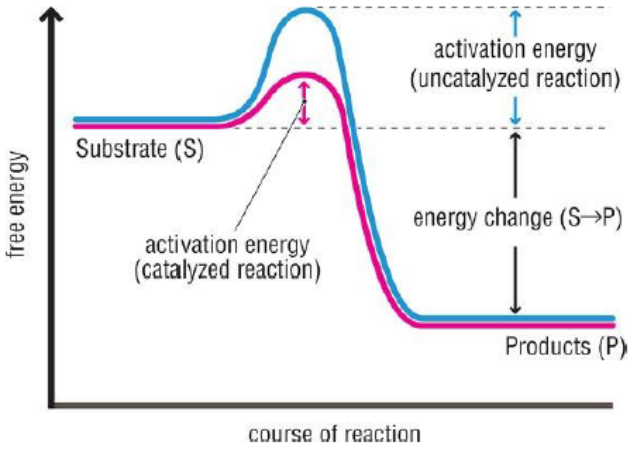
  The greater the disorder, the greater the entropy

  Second principle of thermodynamics: The entropy of an isolated system increases during a spontaneous reaction

  Third principle of thermodynamics: The entropy of a perfect crystal is zero at T=0K

Lowering the energy corresponds to lower the entropy, adding energy corresponds to increase entropy.
A folded protein has a lower entropy in respect to an unfolded one, this lowering of energy is a consequence of the transfer of heat from the protein to the environment after the folding:

Gibbs free energy (G)
The Gibbs free energy is defined as the amount of heat produced in a chemical reaction that can be transformed in work (ex. Folding of a protein), it is a state function defined as
ΔG=ΔH-TΔS
ΔG<0: the transformation can occur **spontaneously** and it frees energy.
ΔG>0: the transformation **doesn't occur spontaneously** and it needs energy.
unfavourable reactions can still occur when coupled to favourable ones so that the overall ΔG is negative (ex. DNA synthesis).
When two molecules bind, the interactions between the molecules and water are broken, and interaction between the two molecules form. This frees water molecules and increases overall entropy of the system.
water molecules trapped in between interacting molecules have low energy and low entropy, so they decrease the entropy of the system; generally, binding reactions are exothermic and make favourable contributions to free energy.

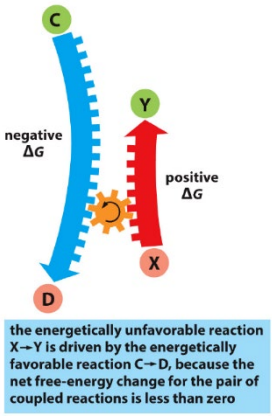


$$G_f - G_i = \Delta G < 0$$

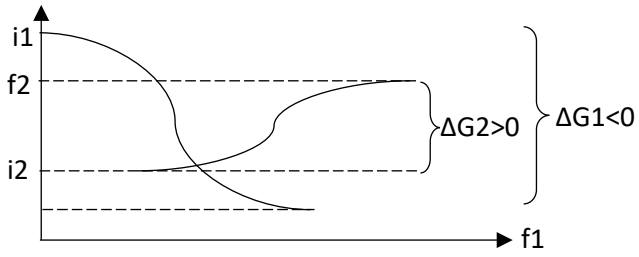
Even if in this case $\Delta G < 0$ we need to surpass a little hill, it is the activation energy (Ea), enzymes are very effective and helpful in lowering the activation energy, without modifying the ΔG

- **Reaction coupling**

Through reaction coupling we consider a couple of reaction and the sum of their free energy.
if the freed energy of the spontaneous reaction is enough to supply energy to the non spontaneous reaction the non spontaneous reaction can take advantage of this and occur easily.

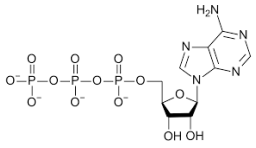


Let's consider a couple of reactions:



In this case the second equation has a gap which is thinner than the first one, so both reaction can occur.

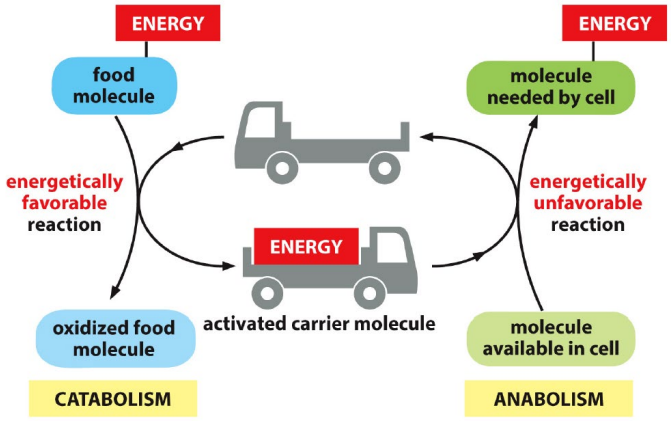
Reaction coupling uses **energy carrier**:

ATP:  ,NADPH, NADH, GTP

phosphate groups can be hydrolysed to provide energy or polymerised to store energy that has been produced.

Ex. DNA biosynthesis
Adding nucleotides to DNA requires energy, cells couple that unfavourable reaction with an energy generating reaction: hydrolysis of a nucleotide triphosphate provides nucleotide monophosphate and energy both needed for the reaction.
NTPs are hydrolysed to nucleotide monophosphates and pyrophosphate; pyrophosphate hydrolysis to monophosphates releases additional energy and is coupled to bond formation.

Upon these mechanisms the **metabolism** of the cell is built:

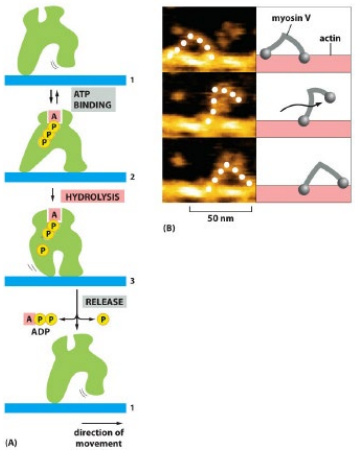


Enzymes usually work in sequence, they are organised in **metabolic pathways**, where the products of a reaction are the reactant of the next one.
It is enough that one enzyme is lacking lo inhibit the function of the pathway.

Structure

Proteins constitute the cytoskeleton (actin, microtubules and intermediate filaments), it is useful to provide roads on which move biochemical components, it is essential for some cellular processes such as cellular division and it gives the structural characteristics and the shape to the cell (it is capable of changing shape).

Movement



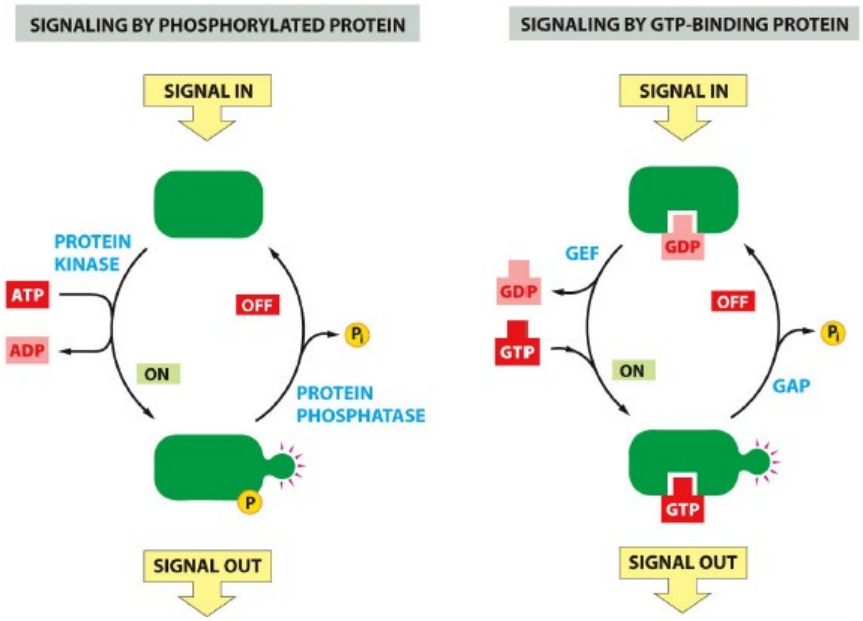
Ex. Myosin walks on actin
in the first stage the binding site for ATP is free and just one leg of the myosin is bound to the actin.
then ATP binds to the protein allowing it to connect the other leg to the actin filament.
due to this expenditure of energy ATP becomes ADP and one leg is disconnected.
finally the ADP leaves the molecule and we reach the initial condition.

It is important to highlight the allosteric interaction in this process: the link of a molecule (ATP) to a specific binding site of the protein (myosin) causes a change in the shape of the whole protein.

Signalling

Proteins can be used to sent signal through out the body and to get signals from the external environment in order to react to them; signals can be transduced both in and out from the cell.
the proteins that fulfil this role are receptors regulators and hormones.

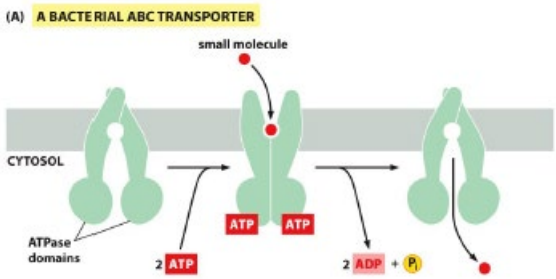
We have two main types of signalling:



in the first case a kinase phosphorylates the proteins and a phosphatase dephosphorylate them; by modifying the proteins a cascade signal can start and influence gene expression.

In the second case it is displayed the GTP cycle, it has the same principle but slightly different phases.

Transport



through the linkage of ATP in an allosteric site the selective membrane channel changes its shape and permits to a molecule to surpass the cellular membrane

Characteristics of the proteins:

Disordered regions

Disordered regions of proteins are important, they permit to bind flexibly to a wide range of molecule, to store different quantities of energy by binding a wide range of high energy bonds, to have a flexible active site and to have a dense interaction between polypeptide that can create barriers.

Disulphide bonds

Disulphide bonds can be formed between the same polypeptide and between different polypeptides, they are capable of stabilising the protein fold but they are present only at certain levels of pH, usually only extracellular proteins are able to carry disulphide bonds.

Assembly factors

Proteins frequently undergo post-transcriptional modification that allow to reach the final structure of the proteins.
the main protagonists of these modifications are chaperons, cofactors, proteolytic and modifying enzymes.

Chaperons and cofactors are molecules that bring noncovalent interactions into the structure allowing a specific folding and specific functions.

Proteolytic enzymes are useful to transform a precursor of a protein into its final form through some cuts that drastically modify the surface of the protein.

Binding to other molecules

As we said proteins interact with all the other biological entities through their surface, so we can define for each protein the properties of affinity and specificity.

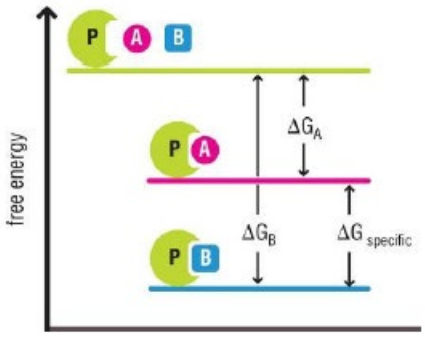
Affinity measures the strength of the bond between the protein and the ligand.
$K_d = [A] + [B]$ in order to have $[A] + [B] \leftrightarrow [AB]$ (50%)
affinity can be defined by 2 variables, Kon and Koff ($K_d = K_{on} + K_{off}$):

- Kon (forward reaction) = how likely A and B interact together (if it is low the two molecules interact a lot)
- Koff (back reaction) = how likely 2 molecules detach one from the other

Specificity is the property to bind selectively to one or few types of molecules.
specificity can be determined by comparing binding constants of different interactions:
$K_d^{specific}/K_d^{non\ specific}$ gives a measure of the specificity

Specificity and affinity are linked to each other and they are both linked to the thermodynamics laws:

1. Energy of the products < energy of the reactants
2. If ΔG is wide the specificity is high



Ex. Histones are not specific for sequences, the specificity is low and the affinity is high.

DNA-protein interaction

All the information contained in DNA should be available for expression, if a specific protein is gets bound to the double helix the information is no more available.

DNA-protein interaction follows two principles:

- Polar interaction: reaction with the minus charges of the phosphate group in the backbone of DNA, at physiological pH phosphate group is deprotonated
- Hydrophobic interaction: DNA bases are hydrophobic, this interaction occurs only in single strand DNA

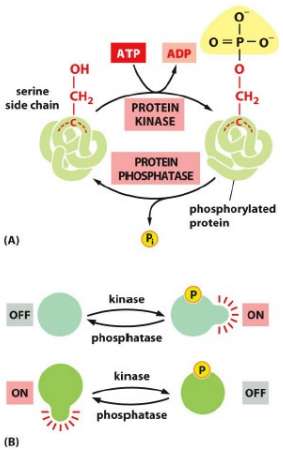
Due to these factors, a protein that wants to bind to DNA has to have a proper **shape** and I has to be **positively charged**.

Post translational configuration

Post-translational modifications of proteins

modification	description	common function of modification	icon
lipid modification	attachment of lipid to protein (examples include palmitylation and myristoylation)	localization to the membrane	
glycosylation	attachment of carbohydrates to side chains; typically Ser, Thr or Asn	protection from degradation, recognition by other proteins	
phosphorylation	attachment of phosphate group to Ser, Thr, His, Tyr, His, Asp	regulation of interactions or activity	
acetylation	attachment of an acetyl group to Lys side chains	regulation of interactions	
methylation	attachment of one or more methyl groups to Lys or Arg side chains	regulation of interactions	
ubiquitination	attachment of small protein, ubiquitin, to lysine side chains (one or more ubiquitins may be attached)	regulation of interactions, targeting of protein for degradation	ubiquitin

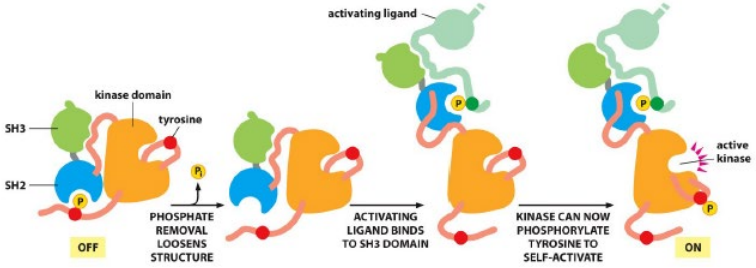
Ex. Phosphorylation



This can change the conformation and interaction of proteins, the kinases and phosphorylation can be controlled by the cell in response to signals coming from external environment.
phosphorylation is a fast way to respond to a signal.

We may have multiple signals that converge into a protein achieving a logic microprocessor: a biologic molecule able to interpret more than one condition.

Ex. Kinase protein



1. Kinases themselves are regulated, usually they are off.
2. When it is dephosphorylated it changes shape (first condition to activate it)
3. Due to the change, now the protein can interact with an activation ligand (second condition)
4. The kinase can self phosphorylate itself and it becomes active

This protein is a signal integrating device: $contion\ 1\ \wedge\ condition\ 2$

Cooperative binding

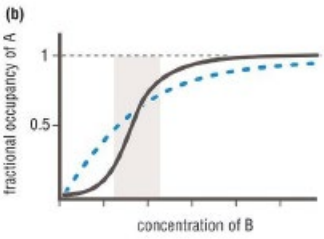
Ex. 2 DNA binding proteins

In cooperative binding the binding of A makes the binding of B more favoured (this is the case of positive cooperativity, it could be also negative: attach of B is discouraged).
This mechanism affects the specificity:
In non-cooperative cases A can binds with low specificity to α and B binds with the same specificity to β; if we consider cooperative binding a high affinity can be reached if 2 condition are respected:

- α and β are close in order to allow the cooperativity (this is a pretty rare event)
- A and B have low specificity.

The cooperative proteins are allowed to have a low specificity because they rely on the low probability to find α and β close to each other, as a consequence they can have a high affinity and build a solid bond.

Cooperativity also changes binding curves, a sigmoidal curve is characteristic of cooperativity:



in non cooperative interaction we need a big amount of A and B.
In the cooperative case there is a narrow range of concentration in which we go from 0% to 100%: once we reach the concentration needed to create the bond of the first protein the second one follows easily.

Ex. Allostery
Allostery allows to control the binding activity of a protein, through the bound of a molecule in the regulatory site we are able to change the shape and the activity of the protein.

When we have multimers frequently we have cooperativity and allostery together:
allostery causes the change in shape of on monomer of the protein, deactivating it, now the protein is much more keen on the deactivation of the second monomer because of cooperative interaction (the bound of the second inhibitor is favoured).

Carbohydrates and sugars

Carbohydrates are sugar based molecules, they serve as energy sources, energy stores, mechanical support and have other biological functions.

Monosaccharides

Carbohydrates are built from simple monosaccharides: $(CH_2O)_n$
Monosaccharides are polyhydroxy aldoses or ketoses, constituted by a sequence of chiral carbons (asymmetric), so each sugar can exist in D and L isomer.
D and L designation is determined by the arrangement of atoms around the asymmetric carbon furthest from the C=O (in nature usually there exist only D-isomers).

Sugars are named for the number of carbons: glucose and fructose (6 carbons) are hexoses, while ribose (5 carbons) is a pentose.

Every sugar has its circular structure, this happens when a hydroxyl group reacts with the ketone/aldehyde group, these rings are stable and therefore very common in nature.

The hexoses cyclise as six-member pyranose rings while pentoses cyclise as five-member furanose rings; pyranose and furanose rings have isomers (anomers) which differ in the orientation of the hydroxyl group on the C1 carbon (the one that was not anomeric in the linear form) and they are called α (down) and β (up).

polysaccharides

The cyclic monosaccharides are the building blocks of complex carbohydrates, they can be joined by a condensation reaction between two hydroxyl groups: this results in a disaccharide with a glycosidic bond that keeps together the tow subunits.

Additional monosaccharides can be chained together, forming polysaccharides.
Sugars have many hydroxyl groups and new units can be added through any of these, making branched polymers like glycogen.

Interaction with proteins and lipids

Carbohydrates are often attached to proteins and lipids:
- carbohydrate + protein = glycoprotein
- carbohydrate + lipid = glycolipid

Cell surfaces are rich in carbohydrates, which are attached to cell surface proteins, these carbohydrates menage to stabilise the cell surface proteins.

Glycosylation is important for cell-cell interactions: glycosylation patterns can be recognised by lectins (special proteins) which mediate the interactions (ex. ABO blood types).

Lipids and fatty acids

Lipids form the basis of cell membranes, steroids and fat & oils used for energy storage; they are strongly hydrophobic and are comprised of fatty acids.

Fatty acids have a hydrophobic hydrocarbon chain (tail) with a hydrophilic carboxylic acid group at one end.
Saturated fatty acids have all the carbons linked by single bonds while unsaturated fatty acids have one or more double bonds in the hydrocarbon chain.
Bonds can be in cis (bent) or trans (straight) conformation, which affects the kink of the hydrocarbon chain.

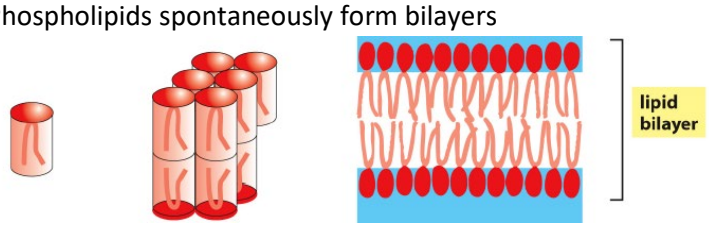
There exist many types of lipids, usually the basic structure consists in fatty acids attached to a polar component (amphipathic molecules).

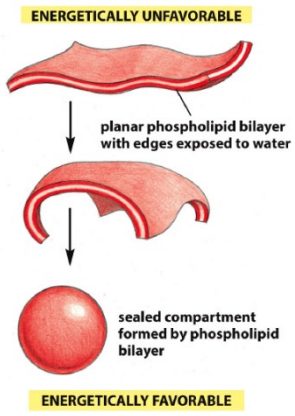
- Triglycerides
- Phospholipids are common in cell membranes, they consist in a molecule of glycerol linked with 2 fatty acids and one phosphate group to which it is attached another functional group. Depending on the group linked to the phosphate we have different types of phospholipids.
- Glycolipids are fatty acids linked with sugars.

Saturation of fatty acids affects how they interact with one another: saturated tails are very flexible and can interact with one another while kinked unsaturated tails are bulkier and cannot pack as closely.
Loosely packed unsaturated fats are more fluid at room temperature (vegetable oil) as opposed to saturated fats (animal fats).
Vegetable oils can be artificially hydrogenated, reducing the double bonds, so they can be solid at room temperature, harmful trans fat, there the double bonds are in the trans configuration, are produced by this artificial hydrogenation

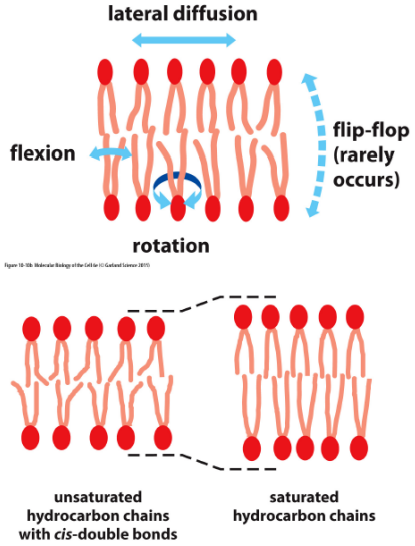
Membranes

Phospholipids spontaneously form bilayers





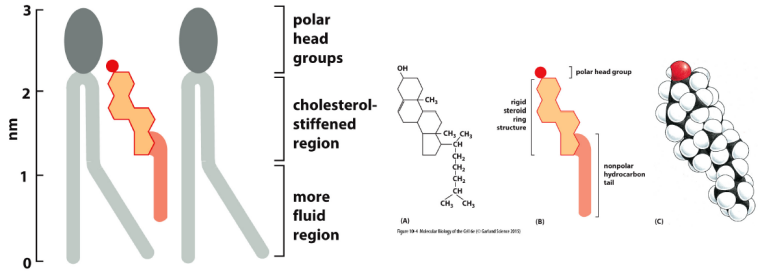
the lipid bilayer is a two-dimensional fluid:



- A lipid molecule can experience different motions:
  - Rotation
  - Flexion
  - Lateral diffusion
  - Flip-flop (rare)
- Membranes with a high grade of unsaturated lipids are thinner because the tails are more spread and can intersect better with the opposite ones, the lipids are also more spread apart.

Cholesterol

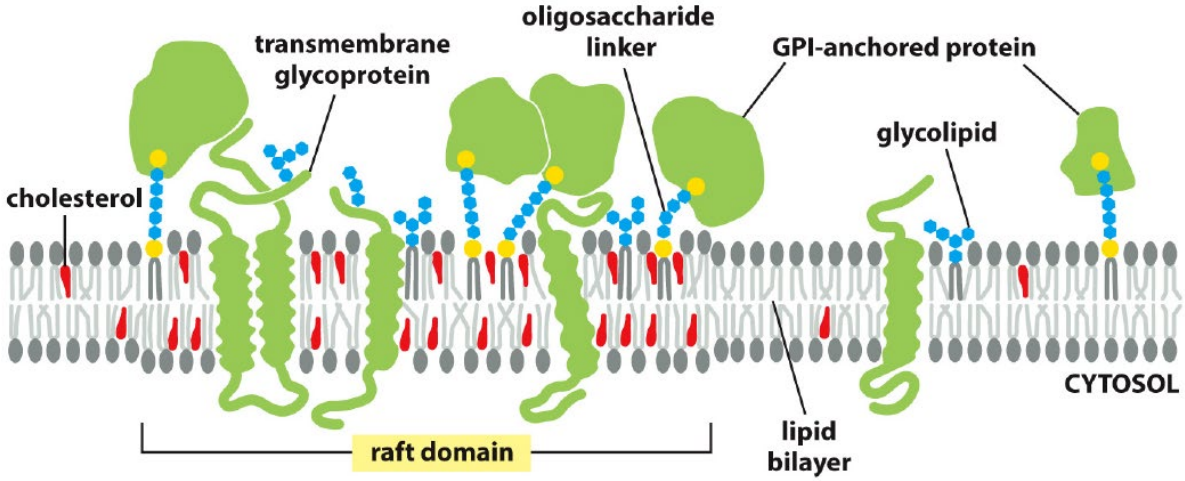
In this arrangement, cholesterol has an important role, it enhances the permeability barrier of the lipid bilayer: the hydroxyl group inserts close to the polar head groups of phospholipids while the rigid plate-like steroid ring interacts with the regions closest to the head group immobilising them. In this way the bilayer results less deformable (but not any less fluid), and the permeability to small water molecules decreases.
Cholesterol also helps the hydrocarbon chains to not crystallise.



Rafts

Cell surfaces are rich in proteins that are stabilised thanks to the attachment of carbohydrates; this is very important for cell-cell interactions.
furthermore, in all organisms nutrients and waste material must pass through the membrane and the molecules facing the extracellular milieu give a signature of the cell to the environment.

This multitude of molecules floating in the bilayer interacts and weak protein-protein, protein-lipid and lipid-lipid interactions reinforce one another to partition the interacting components into **raft domains**.
Cholesterol, glycolipids and some proteins are enriched in these domains

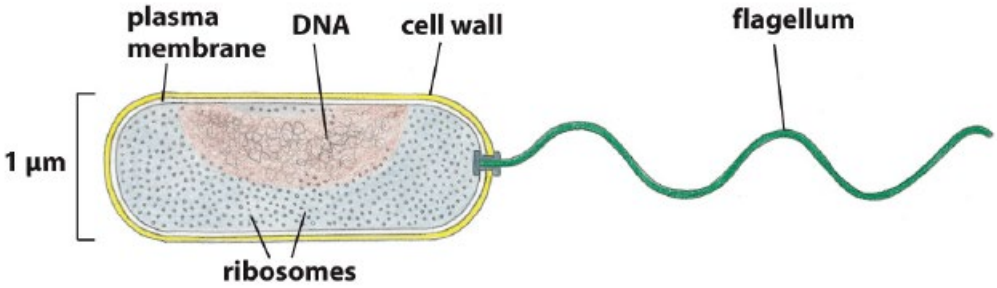


Because of their compositions rafts domains have an increased membrane thickness.

BACTERIAL CELLS

The first necessary information to state is that the belief that prokaryotes are unicellular cells that live alone in the environment is wrong, indeed prokaryotes can be considered as pluricellular organisms: prokaryotes live in populations and sometimes they act for the interest of the population rather than for personal evolution (ex. altruistic- cheater genes), furthermore bacterial populations present differentiation in shape and function among the cells that compose them (just as in humans) and they have signal transmission and other molecular mechanisms that allow them to react to environmental changes.

Structure



- 1um size.
- Prokaryotes and archea don't have a nucleus, their DNA is free in the cytoplasm and it contains from 1000 to 6000 gene.
- The internal environment is occupied by cytoplasm.
- Prokaryotes have the cellular membrane and the cell wall, which can identify the cell as gram negative or gram positive depending on its composition.
- Many prokaryotes have a flagella that, through its rotational movement, allows the bacteria to move.
- In prokaryotes as in eukaryotes ribosomes are present, in order to accomplish protein synthesis.
- Bacteria have a huge biochemical diversity: there exist various types of metabolisms, form and sizes of bacterial cells, as we said they are even able to differentiate
  ex. Amebae
  Amoeba grows like a fungus (in a population of cells) and all the cells are photosynthetic, some cells are able to specialise and become able to fix hydrogen and other can develop into resistant speres, these various types of cells achieve to boost the wealth of the population.

Nucleoid

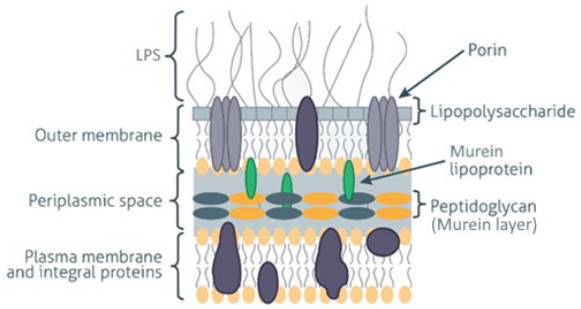
Even if prokaryotes do not have a nucleus, their DNA is organised and compacted through binging proteins: NAPS(nucleoid associated proteins).
these proteins allow to define a specific area in the cytoplasm adhibited to DNA storage: the nucleoid.

NAPS are present in a wide variety of species, that's because they are vital and very efficient in their job: DNA compaction is a fundamental and obscure aspect of life, it is even more important than DNA regulation, that's because these proteins are so widespread.

Bacterial wall

Depending on the reaction with the Gram stain bacteria can be divided into two macro categories.

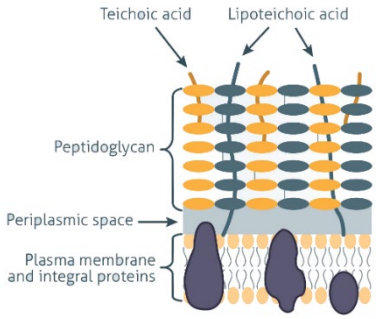
Gram -: not stained by Gram stain because they do not have an exposed cell wall, they have a 2 layer membrane around the wall.



Layer of **LPS** (lipopoly-saccharides), important for recognition of signals and cells in the environment.
Outer membrane (**OM**) is the surface of interaction of the bacterium.
The space between the membrane and the cell wall is very important, it is called periplasm.

The cell wall is constituted of **peptidoglycan**, **porins** are distributed all over the outer membrane to allow the communication with the external environment and **membrane proteins** are spred all over the bilayers, in particular the **Murein lipoprotein** has a very important role, it tightly links the two layers and provides structural integrity to the outer membrane.

Gram +: stained by the Gram stain, it present an exposed robust layer of peptidoglycan that reacts with the Gram stain.



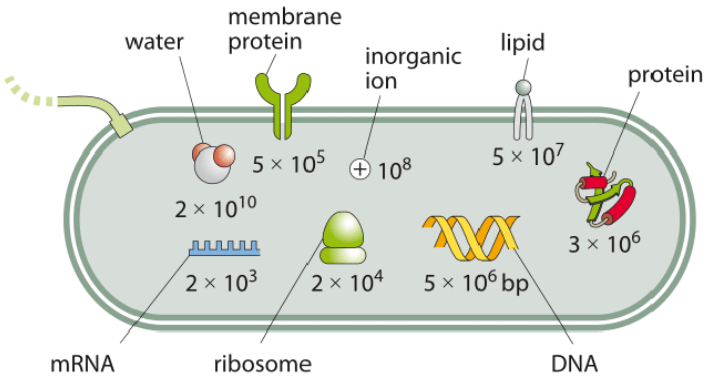
Teichoic acids are polyglycerol phosphate polymers, they are involved in the regulation of cell morphology, cell division and they provide flexibility to the cell-wall by attracting cations such as calcium and potassium.

Also Gram + bacteria present the **periplasmic space**.

The membrane presents a lot of **proteins**.

Numbers

(A) bacterial cell (specifically, *E. coli*: $V \approx 1\ \mu m^3$; $L \approx 1\ \mu m$; $\tau \approx 1$ hour)



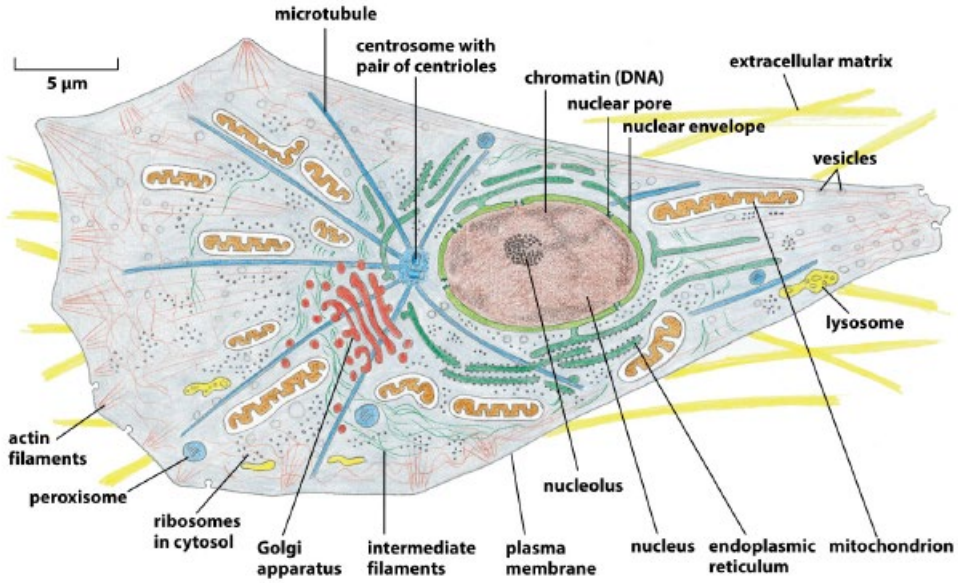
Ex. Compute the molar concentration of 1 molecule inside a bacterium.

$$M = \frac{1\ molecule}{fl} * \frac{1\ mol}{6*10^{23}\ molecule} * \frac{1\ fl}{10^{-15} l} = 1.7 * 10^{-9} M = 1.7\ nM$$

EUKARYOTIC CELL

The key difference between eukaryotes and prokaryotes is that eucaryotes present compartmentalization, each compartment affords a function.

Structure



Eukaryotes are generally plastic and malleable, only if these cells have a cell wall (plants) or if they are strongly held in place by tissues they remain with a specific form; in all the other cases is very common for an eukaryotic cell to change shape.

Cytoskeleton

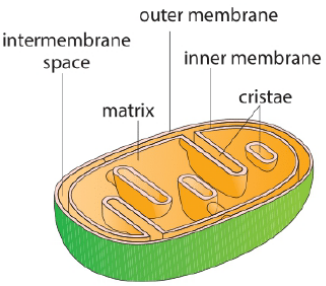
The cytoskeleton is a structure of protein filaments spread all over the whole volume of the cell, it allows the cell to acquire a shape (mechanical strength) and to move; it is constituted by 3 types of filaments:

- Microtubules: they start from the centrosomes (2 per cell) and provide structural support to the cell, they are also the “ropes” through which the organelles can move (organelles are shuttled around the cell)
- Actin filaments: actin filaments can be polymerised and hydrolysed in order to lengthen or shrink; they allow the cell to move in the environment thanks to the change in shape (ex. Microvilli)
- Intermediate filaments: they provide mechanical strength

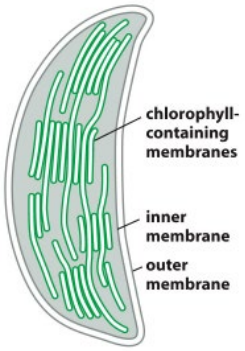
All these filaments are made of simple building blocks such as keratin.

Mitochondria and chloroplasts

- Mitochondria: mitochondria allow the cell to perform the cellular respiration: sugars and oxygen are consumed to gain energy and CO2.



- Chloroplasts: chloroplasts are membrane systems that are able to capture light and use its energy for the synthesis of the carbohydrates and oxygen.



Mitochondria and chloroplasts are two of the few organelles that contain some personalised DNA, separated from the chromosomes of the cell; actually, we can identify a proper genome of a mitochondrion, it is very small, even smaller than the genome of a bacterium and due to its presence the eukaryotic genome can be defined as a **hybrid genome**.

Mitochondria are the remanence of a predatory event: eucaryotic cells predated an aerobic bacterium, managed to live in symbiosis with it, without altering its function and its genome that has been inherited separately from the eukaryotic chromosomes through generations.

Due to its plasticity it is very common, for an eukaryotic cell, to predate bacteria; in some cases the incorporated cells are not digested, on the other hand an exchange of services begins: the bacterium provides nutriments to the eukaryotic cell while the bigger cell protects the bacterium.

In the case of mitochondria the bacterium lost the cell wall and the biggest part of its genome (that's why mitochondrial genes are more similar to prokaryotic genes rather than eukaryotic ones), and enabled the eukaryotic cell to adopt an aerobic metabolism.

After this event another symbiosis phenomenon (**endosymbiosis**) occurred when some eukaryotic cells implemented chloroplasts; this remanence of a photosynthetic bacterium allowed the cell to exploit its properties and gain oxygen and energy.

A similar mechanism is very common to these days, it is called **kleptoplasty**: some animals such as sea slugs still use are able to digest photosynthetic organisms (algae) and keep the chloroplasts intact to maintain and exploit their function.

Endoplasmic reticulum, Golgi apparatus and lysosomes

The endoplasmic reticulum (ER) surrounds the nucleus, it is a complex membrane system, it is the unit that processes proteins and it is the largest organelle in eukaryotic cells.

It is divided in:

- Rough ER: punctuated by ribosomes, that are strictly associated with the membranes of the reticulum, it is perfect for an efficient protein synthesis.
- Smooth ER: without ribosomes.

The functions of ER, other than producing and manipulating proteins are to synthesise lipids and act as a deposit of calcium.

ER usually makes contact with the **Golgi apparatus**, which is a membrane system that uses vesicles: Vesicles can be generated by the cell membrane through invagination, and they reach the Golgi apparatus thanks to the mediation of the **endosome**.
the endosome sorts the material that comes from the external environment and decide if sending it back to the outside, taking it inside a **lysosome** to degrade it or if taking it in the Golgi apparatus.
In the Golgi apparatus the proteins can be modified and successively transferred to the ER.

N.B. all these transfers are accomplished through the cytoskeleton!!

Nucleus

The nucleus is the organelle that contains the chromosomal DNA, it is delimited by a membrane, and it contains way more genetic information than a mitochondria and also more information than a bacteria, eukaryotes are the organisms with the higher number of genes.

The DNA is highly organised and compacted inside the nucleus, it is called chromatin and it is divided in euchromatin (actively coding DNA) and heterochromatin (inactive DNA).

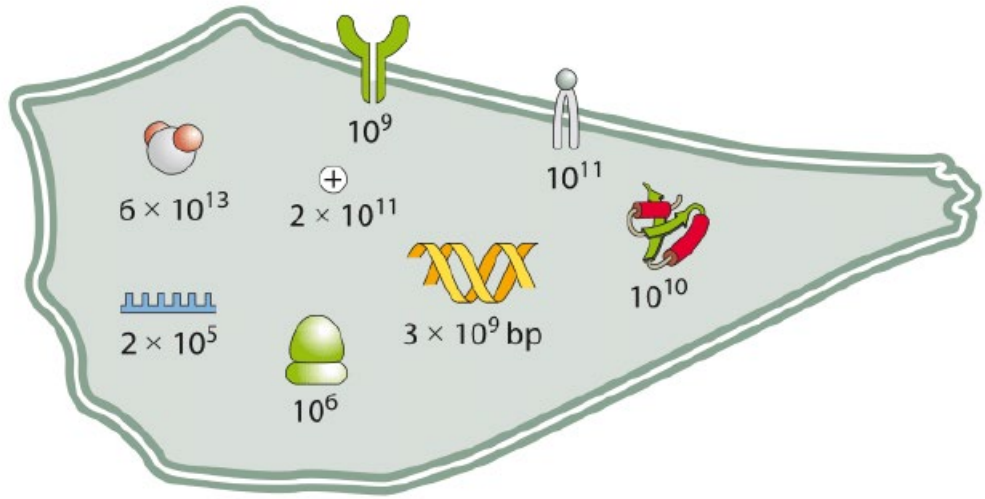
As we described when we talked about the central dogma of molecular biology the information of DNA flows through RNA to afford proteins, in eukaryotic cell the synthesis of the proteins occurs outside the nucleus, so it is necessary a way to transfer information from the inside of the nucleus to the outside on the other hand bacterial translation is cotranscriptional, it occurs almost simultaneously to the transcription so it is very fast to act on the genes expression.

In eukaryotes to activate a gene we need to transcribe it, the RNA needs to be exported from the nucleus through nuclear pores and it has to be translated.
the export of material outside of the nucleus consists in an active transport through the nuclear membrane that involves the nuclear pores, it is very selective (only processed RNA can pass) and it requires energy.

Numbers

Eukaryotic cells are at least 10x bigger than prokaryotes, in each direction; this consists in a volume which is 1000x bigger.

(C) mammalian cell (specifically, HeLa: $V \approx 3000\ \mu m^3$; $L \approx 20\ \mu m$; $\tau \approx 1$ day)



Molarity of a molecule inside an eukaryotic cell:

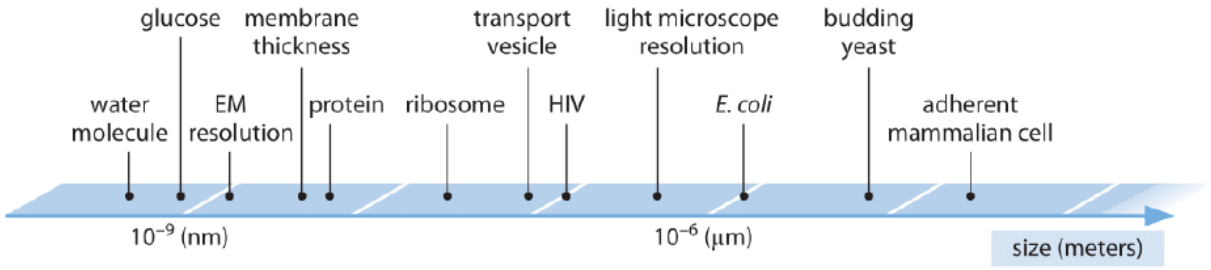
$$M = \frac{1\ molecules}{fl} * \frac{1\ mol}{6 * 10^{23}\ molecules} * \frac{1\ fl}{3000\ mol} = 1 * 10^{-12} M = 1\ pM$$

If we think abound an enzyme that has a binding affinity of 1 nM its activity in a bacteria or in an eukaryote is very different: in a bacteria it is enough to have 1 molecule of substrate to guarantee the action of the enzyme while in eukaryotes we need much more molecules.

PROTISTS

A group of diverse eukaryotic, predominantly unicellular microscopic organisms, they may share certain morphological and physiological characteristics with animals or plants or both.

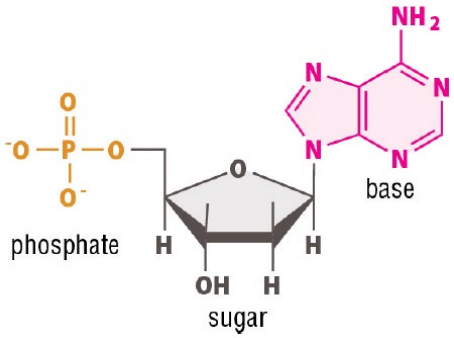
SIZES



property	*E. coli*	budding yeast	mammalian (HeLa line)
cell volume	0.3–3 μm³	30–100 μm³	1000–10,000 μm³
proteins per μm³ cell volume	2–4 × 10⁶		
mRNA per cell	10^3–10^4	10^4–10^5	10^5–10^6
proteins per cell	~10^6	~10^8	~10^{10}
mean diameter of protein	4–5 nm		
genome size	4.6 Mbp	12 Mbp	3.2 Gbp
number protein coding genes	4300	6600	21,000
regulator binding site length	10–20 bp	5–10 bp	
promoter length	~100 bp	~1000 bp	~10^4–10^5 bp
gene length	~1000 bp	~1000 bp	~10^4–10^6 bp (with introns)
concentration of one protein per cell	~1 nM	~10 pM	~0.1–1 pM
diffusion time of protein across cell ($D \approx 10$ μm²/s)	~0.01 s	~0.2 s	~1–10 s
diffusion time of small molecule across cell ($D \approx 100$ μm²/s)	~0.001 s	~0.03 s	~0.1–1 s
time to transcribe a gene	<1 min (80 nts/s)	~1 min	~30 min (incl. mRNA processing)
time to translate a protein	<1 min (20 aa/s)	~1 min	~30 min (incl. mRNA export)
typical mRNA lifetime	3 min	30 min	10 h
typical protein lifetime	1 h	0.3–3 h	10–100 h
minimal doubling time	20 min	1 h	20 h
ribosomes/cell	~10^4	~10^5	~10^6
transitions between protein states (active/inactive)	1–100 μs		
time scale for equilibrium binding of small molecule to protein (diffusion limited)	1–1000 ms (1 μM–1 nM affinity)		
time scale of transcription factor binding to DNA site	~1 s		
mutation rate	10^{-8}–10^{-10}/bp/replication		

DNA

Structure



DNA is a polymer of nucleotides, there exist 4 nucleotides for DNA.

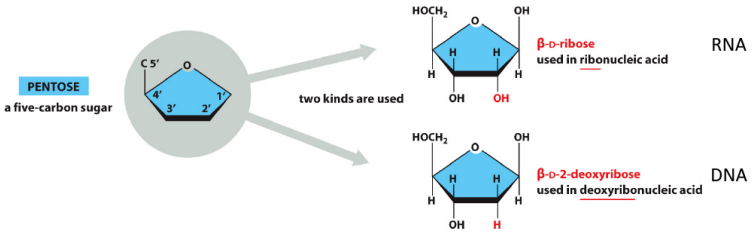
Each nucleotide comprise a base, a sugar and one or more phosphate groups.

The nucleotides are linked by phosphodiesters bonds

Backbone

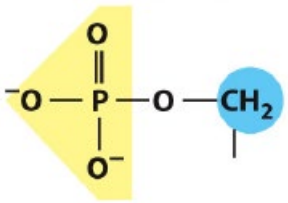
Sugar + phosphate = backbone, it is always the same for the DNA of the whole organism and for every living being on earth.

- Sugar:



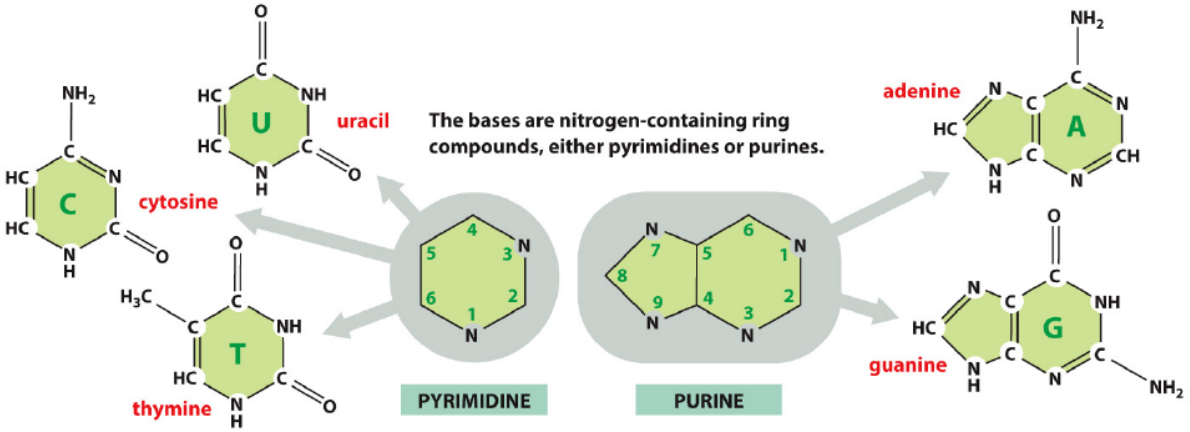
in the sugars the carbons are numbered 1' to 5', RNA has an additional OH group in 2' carbon, this makes the molecule highly reactive with respect to the DNA.

- Phosphate: the phosphates are normally joined to the hydroxyl of the 5' carbon of ribose or deoxyribose sugar. mono-, di- and triphosphates are common but polyphosphate compounds have different functions in respect to DNA, for instance ATP is useful as an energy storage

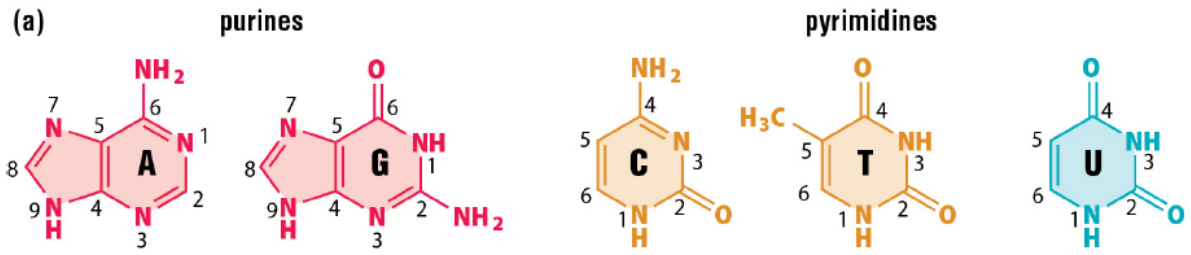


Nucleotides

Sugar + phosphate + base = nucleotide



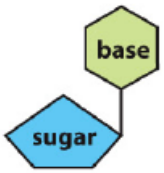
bases are planar rings that are typically uncharged under physiological conditions



DNA contains the bases adenine, guanine, cytosine and thymine while RNA does not have thymine, but uses uracil instead.

Nucleosides

Base + sugar = nucleoside.
The link between the sugar and the base is a glycosidic bond, and it involves C1' of the sugar and the N1 of the pyrimidine or N9 of a purine.



Nomenclature

A nucleoside or nucleotide is named according to its nitrogenous base, however single letter abbreviations are used as shorthand for the base alone, the nucleoside and the nucleotide.

Base	Nucleoside	Nucleotide	abbreviation
Adenine	Adenosine	Adenylate	A
Guanine	Guanosine	Guanylate	G
Cytosine	Cytidine	Cytidylate	C
Uracil	Uridine	Uridylate	U
Thymine	thymidine	thymidylate	T

(note that if you want to refer to DNA nucleosides you need to add deoxy- if front of the name)

Depending on other characteristics of each nucleotide we have these distinctions:

- S = strong G or C
  W = weak A or T(U)
- R = purine A or G
  Y = pyrimidine C or T
- K = keto G or T
  M = amino A or C

Consensus sequence: a sequence with a function

ex. 5' – AWSSCGYT – 3'
Consensus: 5' – A(A/T)(G/C)(G/C)CG(C/T)T - 3'
any sequence with alle these combinations will afford the same function.

Polymer

Nucleotides are joined by a phosphodiester bond between the 3' hydroxyl of one sugar and the phosphate attached to the 5' hydroxyl of the next sugar.
nucleic acids strands are thus directional, one end has an exposed 3' hydroxyl, the other end has an exposed 5' phosphate.

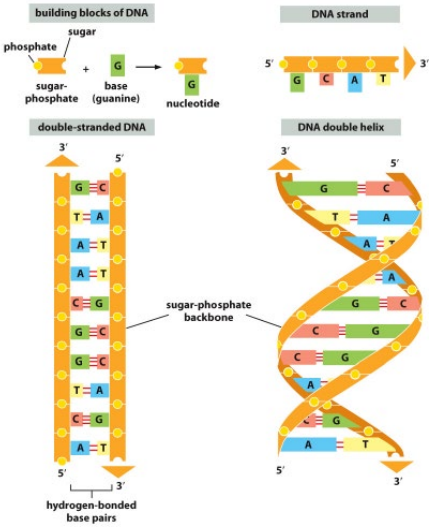
By convention, nucleotide sequences are written in the 5' to 3' direction, this is very important because also the majority of enzymes that interact with DNA do so in a specific direction.

Two DNA strands associate via hydrogen bonds to form double stranded DNA, bases pair precisely with their complementary base:

- A pairs with T, with 2 H-bonds (weaker)
- C pairs with G with 3 H-bonds (stronger)

In general purines (R) pair with pyrimidines (Y), these are called Watson-Crick base pairs.
The sequence of one strand dictates the sequence of the other strand, thus the strands are complementary to one another and they both carry the same information.

The two strands are antiparallel, the 5' end of one strand pairs with the 3' end of the other.

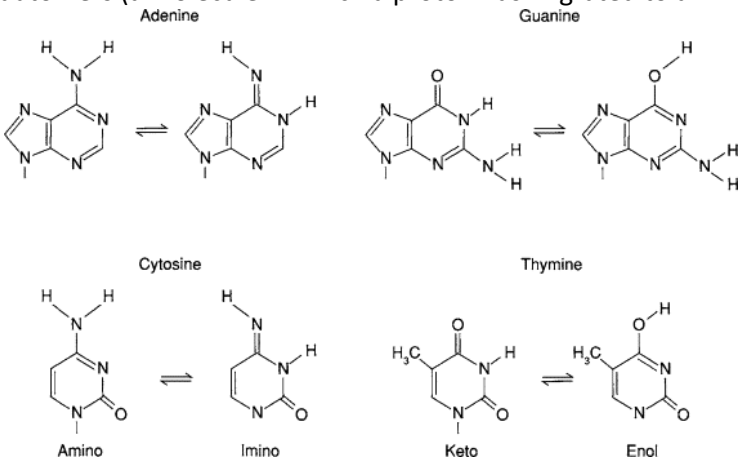


remember: the stability of DNA is enhanced by the presence of H-bonds between the bases of adjacent nucleotides.

The most energetically favourable formation of double stranded DNA is for the two strands to wind around one another in a right-handed double helix.

Tautomers

Bases can also exist (less than 0.01%) as tautomers (a molecule in which a proton has migrated to a different place).
Tautomers have implications for the accuracy of DNA replication (DNA polymerase has difficulty in pairing the right base), and can therefore provide genetic variation.



Amino bases like adenine and cytosine go from an amino group to an imino.

Keto bases go from a keto group to an enol.

Energy carriers

Nucleotides are also important energy carriers: ATP (very used in reaction coupling), CoA, FADH, NADH, NADPH.

If we think from a logic point of view, these molecules are at the heart of energy (thermodynamics) and information: which are two fundamental aspects of life.

This link between thermodynamics and information is also found in physics.

B-DNA

The B-DNA is the most common form of DNA, it consists in a right-handed double helix, it has a diameter of 2 nm, the distance between 2 adjacent base pairs is 0.34 nm and the helix repeats every 10.5 base pairs.
One of the most important characteristics of this structure is that it is characterised by the presence of two groove: the major groove (1.3 nm) and the minor groove (0.9 nm).

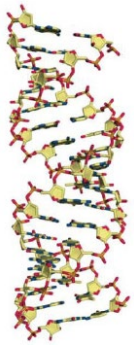
Enzymes are able to exploit the structural characteristics of DNA to interact with it:

- Shape readout: the negative charge coming from the phosphate group in the backbone is used to create polar interaction, this is a not specific readout
- Specific readout: a specific readout of the DNA double helix is possible thanks to the major groove.
  The major groove enables the protein to recognise the quantities of electron donors and acceptors in the sequence; the result is a precise electron donor and acceptor profile (array) that identifies a sequence and a protein can recognise it.
  This process is possible only thanks to the structure of proteins: α helixes and β sheets are perfect to fit the major groove and give to the protein the contact surface needed to detect electron donor and acceptors.
  ex.
  GAATTC
  CTTAAG
  The recognised signals through groove readout are the combination of base pairs, not their order so in this case the possible attachments of the protein are in both the red spots (they are symmetrical).
  On the other hand the minor groove is much more narrow and the proteins struggle to get information from it, the only things that can be inferred are the strong or weak base pairs.
- Sequence-dependent shape readout: through the sequence-dependent shape readout we can extract much more information from the minor groove.
  certain weak nucleotide sequences can narrow the minor groove (ex. AAAAAA), this causes:
  - Major groove enlarges.
  - The electrostatic potential of the minor groove increases, it becomes more negative and provides a specific docking environment for 2 amino acids: the lysine and the arginine which possess a long arm with a positive charge at the end.

A-DNA

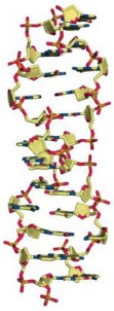
In this form DNA is right handed, it has 11 bases per turn so it results a little bit enlarged and squeezed.

This form is induced by DNA binding proteins and it is very common in double stranded RNA



Z-DNA

This left-handed helix can result from methylation of cytosine, torsional stress and high salt concentrations



Supercoiling

If a filament of DNA is under tension it starts twisting on itself forming supercoils to relieve the tension.

Supercoiling can be positive or negative, depending on the direction of the DNA twisting:

- **Clockwise** winding of DNA, tending to separate the strands leads to **negative** supercoiling (DNA is **underwound**)
- **Counter clockwise** winding induces **positive** supercoiling that enhances the forces that keep together the two strands of DNA (many bacteria exploit this characteristic to keep DNA together) (DNA is **overwound**)

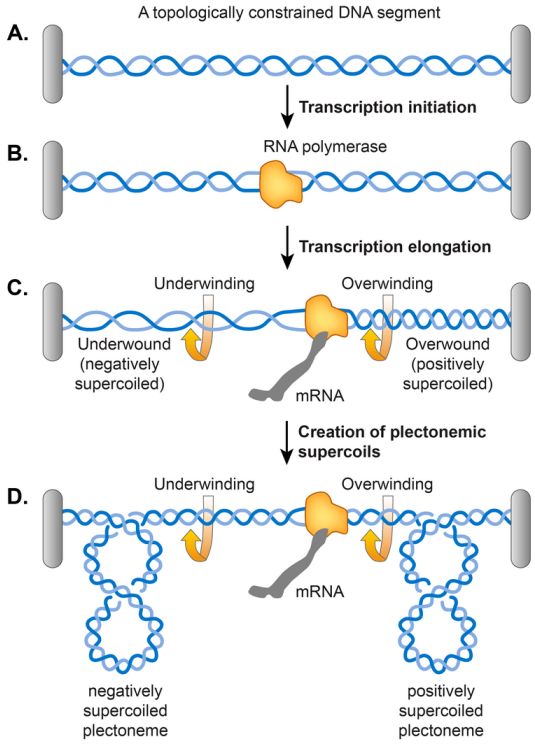
These structures can help to compact DNA and have also a role in the interaction between DNA and enzymes; in particular, duplication needs the DNA to be underwound (the opening of DNA is favoured), if the DNA is overwound it is required a lot of energy to open it; furthermore when an enzyme walks on a strand it leaves positive supercoils ahead and negative supercoils behind.

To manipulate supercoils the cell has available some specific tools (enzymes), in principle, DNA is cut and twisted, so that the double helix structure is put under pressure, essentially the number of bases per turn changes.

Topoisomerases are enzymes designed to control supercoils:

- Topoisomerase type I: relieves negative supercoils of DNA, it adds a turn in the DNA helix. It nicks DNA, cutting a single strand and reconnecting it once the other strand has passed between the cut.

- Topoisomerase type II: relieves positive supercoils in DNA using ATP, it resolves the overlap of two double strand DNA filaments.
  It cuts double stranded DNA, reconnecting the strands once an entire supercoil has been produced.



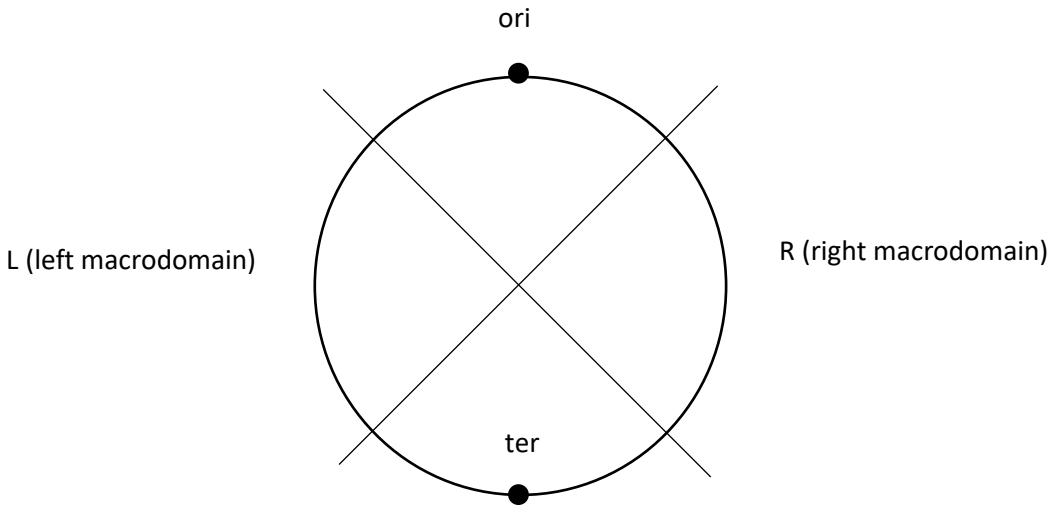
Chromatin

We need a high degree of compact to fit the E. Coli genome inside a cell, we need to compact it nearly 1000 times its original length.

Bacteria

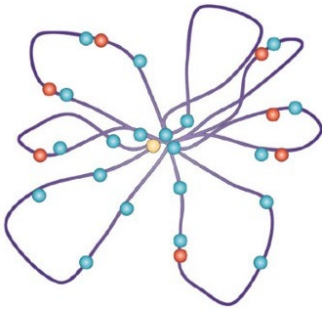
Chromatin is compacted DNA, also bacteria have it, it's just less ordered.
NAPs (nucleoid-associated proteins) are small positively charged proteins, they are fundamental for chromatin and they are organised in 3 levels:

III. **Nucleoid**: area of the cell in which DNA is situated in a compact structure, the bacterial chromosome is divided into 4 macrodomains:



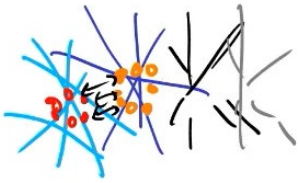
ori = origin of replication
ter = termination of the replication (the replication forks rejoin)
usually this chromosome is negatively supercoiled

II. **HDR**: **high density regions** are a bunch of microdomains held together by some NAPs, called **histone like proteins**, that interact with microdomains and with each other to form an higher level structure:



nearly 250 kbp

The loops of this structure usually present negative supercoils that keep them compacted but, once they need to be transcribed, they can be selectively decompacted. Multiple HDR can interact between each other, some forces act through the centre of these complexes keeping them together in a stable structure that develops in depth (in respect to the drawing above), reaching the third level of organisation.



At the heart of HDR there are RNA polymerase's transcription foci, that's important because:

a. We have a confinement of the transcription, all the factor stay together, reaching a good efficiency
b. Transcription changes the shape supercoils, so it changes one of the main factor of DNA compaction; this means that transcription of a DNA region determines the compaction of DNA and therefore another region will be discouraged to transcribe.

Chromatin shape and transcription are not neutral one to the other.
Moreover, these limited regions are functionally separated, a core is separated from the other yielding sort of **compartmentalization**.

I. **Microdomains**: formed by NAPs

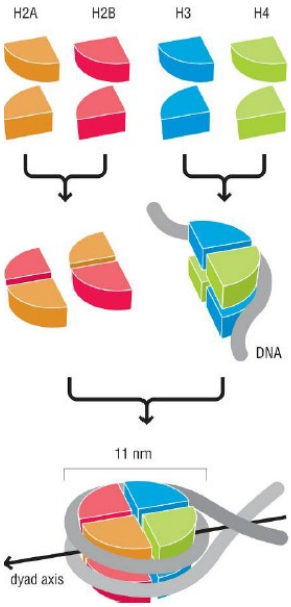
loop bridge bend(IHF) microdomain (1-15 kbp)

All these naps together contribute to form the basic structure of the bacterial nucleoid which is microdomain.

Eukaryotes

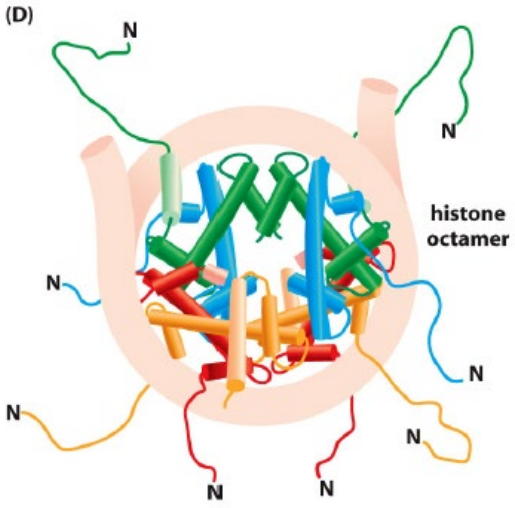
Eukaryotic chromatin needs a way bigger degree of compaction, to achieve it, it has more levels of organisation:
eukaryotic chromatin is organised in chromosomes, each chromosome is as big as a bacteria and it is formed by a centromere, arms and telomeres

I. **Nucleosome**: nucleosomes are complexes of DNA wrapped twice around specific proteins (histones), each nucleosome is distanced from the others by 60 bp sequences, called spacers or linkers and they are wrapped by 150bp.



histones:

each nucleosome is formed by an octamer of histones, there exist **4 core histones**, they arose early in eukaryotic evolution and, as they are crucial for survival, are very highly conserved: H2A, H2B, H3, H4.
Each nucleosome is formed by the association of 2 H3-H4 dimers, then the DNA start winding in a left handed wrap.
Then the 2 H2A and H2B dimers bind to form a 9 nm disk, to which DNA is wrapped twice, for a total of 156 bp (in the not tight form).
Now the **histone octamer** is complete:

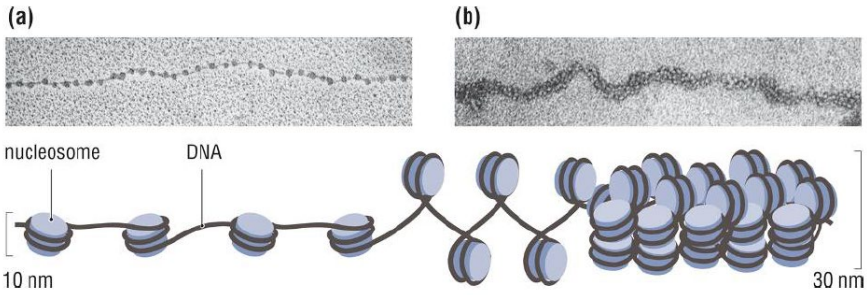


The two ends of the filament exit from the same side, this will be crucial for the regulation of DNA expression.
Furthermore as you can see histones present long arms with a positive charge at the end, they are arginine and lysine amino acids, they are useful to secure in place the DNA strands (negatively charged)

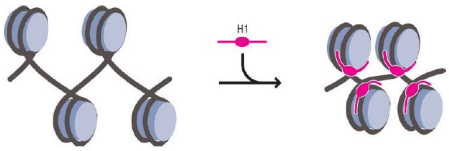
a. **compaction**: the more DNA is wrapped around the histones the more compact the fibre is.
b. **Negative supercoils**: When the histones are removed negative supercoils remain in the filament, these are important for transcription and duplication: in fact to transcribe DNA we need to open it and to act on the single strands, negative supercoils give to the double strand the **tendency to open up**.
c. **Histone tails**: one of the most important things is histone tails, each core histone has an N-terminal tail (up to 25 amino acids) that extends outwards between the DNA coils; their role is further compact DNA and to accept **post-transcriptional modifications**, mainly acetylation (loosens the chromatin), deacetylation and methylation (attacks only deacetylated regions and it turns off genes), this

mechanism allows epigenetic regulation to occur, creating the distinction between euchromatin and heterochromatin (for this reason bacterial chromatin is very accessible to transcription factors while in eukaryotes it is needed to decondense it first).

II. **30 nm fibre**: the series of nucleosomes forms a 10nm fibre, at this point we can exploit the fact that DNA filament exits from the nucleosome from the same side to create a 30 nm fibre that further compacts DNA.

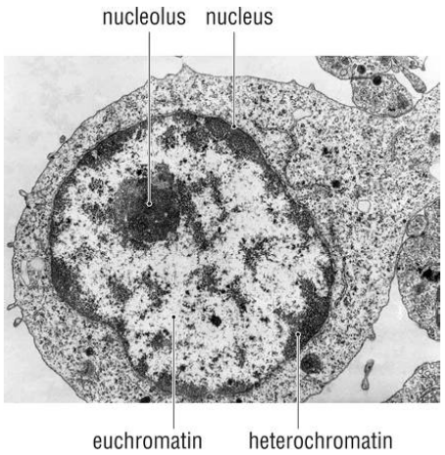


H1 is the **linker histone** (fifth histone protein), the molecule that binds to both the DNA filaments that are exiting from the nucleosome, H1 is able to tighten up DNA by pulling the portions in between the nucleosomes inside the nucleosomes.
The result is that the nucleosomes get closer and they get wrapped by more base pairs, while a rigid fibre forms.



III. **700 nm coil**: the 30nm fibre is then compacted into chromosomes in which large loops of chromatin are anchored to a central scaffold; usually loops of DNA protrude out from the compacted structure to be more accessible to transcription factors.

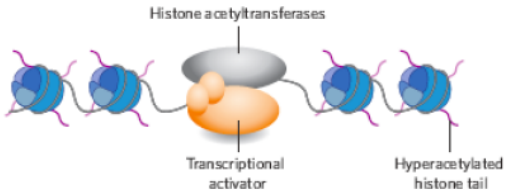
Eukaryotic chromatin is a very busy environment, there are also a lot of non-histone proteins and histone modifying enzymes.



This is the image of the environment of the nucleus, some regions stain lightly, these are the less compacted ones, that are part of the euchromatin; while darker regions are more compact and they are part of the heterochromatin.

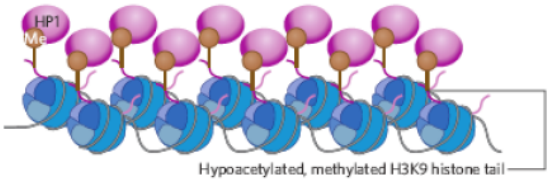
Inside the nucleus there is also very electron dense structure, the nucleolus, it is not enclosed in a membrane and it is just the site where the ribosomal RNA transcription occurs (most intense transcription activity).
rRNA is transcribed very actively, from specific genes and the area is so dark because these genes are often part of the heterochromatin.

Euchromatin



- **decondensed**, uncompacted
- histone tails **hyperacetylated**
- within chromosome arms
- unique sequences
- gene-rich
- **transcriptionally active**
- accessibile to transcription factors
- replicated during S phase
- ricombines during meiosis

Heterochromatin



- highly **compacted**
- histone tails **hypoacetylated**
- predominant at **centromeres** and **telomeres**
- rich in sequence **repeats**
- **poorly accessible** to transcription factors
- poor in genes
- replicated only in late S phase
- does not recombine during meiosis
- may present **histone substitutions**

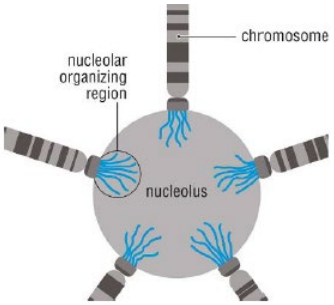
Heterochromatin

How comes that a compacted structure as heterochromatin (nucleolus) allows the transcription of rRNA?

In order to generate all the ribosomes of the cell we need many repetitions of rRNA, so we also need many copies of genes that code for rRNA, it turns out that heterochromatin is perfect when we have to use many similar genes.

Similar genes usually tend to recombine between each other so if they are in heterochromatin the **recombination** is very **discouraged** even during mitosis, in this way these super important sequences are conserved.

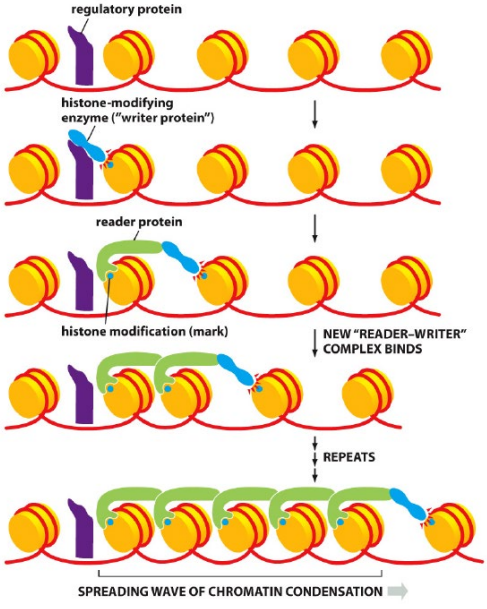
rRNA genes sequences usually are placed at the tip of the chromosomes, close to the telomers, so the nucleolus consists in the association of multiple chromosome tips, that are gathered together by some transcription factors and then decondensed in order to be transcribed.



The heterochromatin proteins and factors of eukaryotes are **very conserved** between all the organisms so we can study the factors that are found in yeast.

The main organiser is Swi6 (or HP1 in drosophila), which is a protein that has 2 domains:

- Chromodomain: it is able to **bind** one of the methylated tails of histone 3.
- Chromo shadow domain: responsible for the **recruitment** of the methyltransferase, which is an enzyme able to methylase another histone.



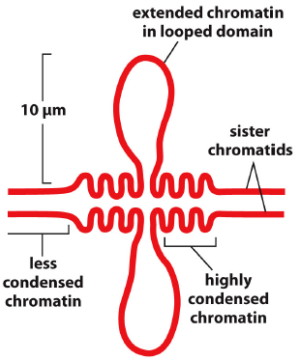
Imagine we have an **initial regulatory event**, when a histone is methylated Swi6 can bind to it, through the chromodomain while the chromo shadow domain methylates another histone, that will cause the binding of another Swi6 (positive feedback loop).

Through this mechanism the heterochromatin regions can **spread**.

Imagine that a gene is placed is in a region close to the start of a methylation event (euchromatin-heterochromatin **boundary**), this means that the gene can be silenced or activated in different cells, randomly, this will cause a chimeric expression-inexpression of the gene.

Karyotype

When chromosomes are condensed, after the S phase, they are visible through microscopy, the display of compacted chromosome is known as karyotype.

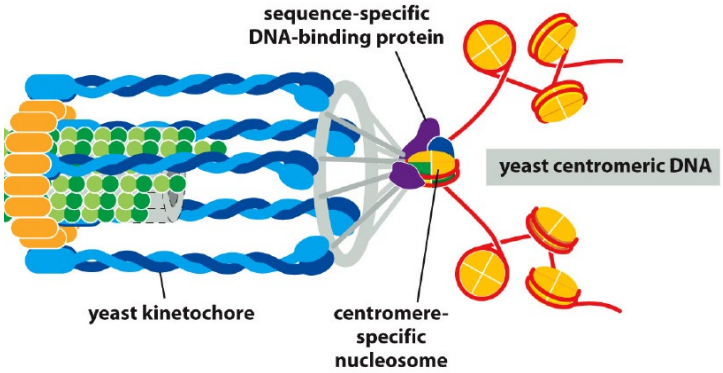


In the karyotype in the chromosome there are darker and clearer bands, again caused by a different compaction degree of the chromatin; patches of highly condense chromatin are interspaced by less condense chromatin and loop domains that are more active in transcription.

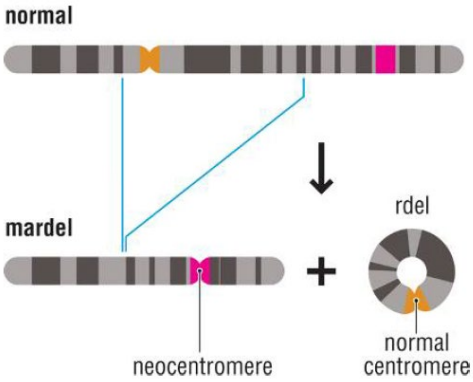
Chromosomes

Chromosomes are formed by:

- **Chromosome arms**: can contain euchromatin.
- **Centromeres**: just 1 for each chromosome, the region that surrounds it is called **pericentromeric** region which is also composed of heterochromatin but it is different in terms of histone composition; the centromere has **histone substitution**, there is a completely different set of histones, in particular **CENP-A** (centromeric protein A) that is responsible for forming the peculiar chromatin structure in the centromere.
  These hist substitutions are important for providing molecular marks that define the centromere, then they are used to attach the kinetochore, for example CENP-A substitutes histone 3 and it defines a centromere specific nucleosome, this particular nucleosome recruits specific DNA binding proteins that are associated with the kinetochore, in order to provide a stable dock for microtubules.



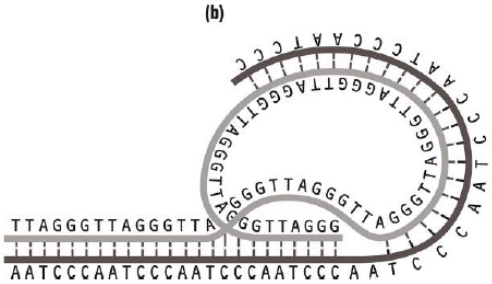
Some experiment have told us that it is **not the DNA sequence that defines a centromere**, but it is the chromatin structure, frequently centromeres are characterised by repeated sequences such as alpha-satellite repeats.
Imagine to have a chromosome with a normal centromere and some alpha-satellite repeats, until the normal centromere of the chromosome is present, the alpha-satellite will not become a centromere but as soon as the chromosome is rearranged and the normal centromere is deleted the alpha satellite now can become a **neocentromere**.
Centromeres are not fixed and repeated sequences can become alternative centromeres.



Other particular histone substitutions are useful to carry out particular functions, for example H2A is important because it is phosphorylated at sites of DNA ds breaks, and it recruits the enzymes involved in the DNA repair.

- **Telomers**: telomeric DNA is very important, it is represented by short repeats of a given sequence, they tend to be species specific (in human the repeat is TTAGGG), these repeats can be very long, up to 30 kb of repeats.
  One common theme in eukaryotes is that these repeats are **guanine rich**, and they have **3' end** that stretch out.
  The 3' ends allow the replication of chromosome ends preventing the consumption of the telomeric sequences that protect important information, but also they can be very dangerous, ssDNA is a sign of danger for the cell and that's because:
  - it could invade a double stranded sequence and recombine with important genes.
  - ssDNA is often sign of stalled replication fork or other DNA polymerase problems.

  Therefore at telomeres and in any other occasion where ssDNA is present there are proteins (**single stranded binding proteins**) that associate with ssDNA in order to protect it from being degraded and from interacting with other DNA.
  In order to further protect chromosome ends, and not mistaking them for double stranded breaks the 3' end of the telomeric repeat can invade the same telomere sequence and so form a knot or a loop structure.



Also telomeric chromatin is characterized by several factors which associates with those repeats, the repeats are used as a template by a particular enzyme, the **telomerase**, which is a reverse transcriptase that makes DNA out of an RNA template (reverse of RNA pol); telomerase uses a short RNA primer carried along with the enzyme and it binds to the telomeric repeats, each time it associates with a telomer telomerase uses the primer to extend the telomer adding more telomeric repeats.

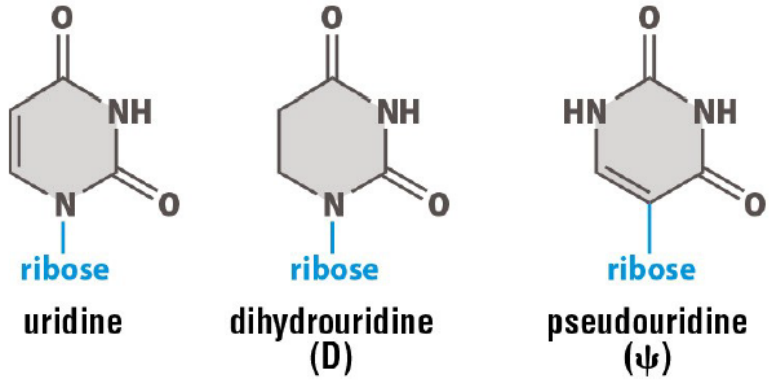
RNA

RNA is chemically similar to DNA, the differences are:

- RNA uses uracil instead of thymine (it pairs with adenine).
- RNA has a hydroxyl group at C2'.
- There exist many modified RNA bases that have a role in the RNA function (this phenomenon happens very less in DNA because it must be a reliable store for information)

Chemical modifications

Many RNA molecules, especially directly functional ones (tRNA and mRNA), require chemical modification of their bases after synthesis (post-translational modifications).
These modifications have a regulatory function and are usually conserved among species, reflecting their crucial role (for example 10% of the nucleotides of tRNA are modified).



Lack of double bond

very important: it makes RNA way more stable and protects it from overreactions

Ex. Vaccines
In the past researchers thought that these modifications were useless, in reality they can have also some important regulatory roles.
The discovery of modified bases gave a good alternative to DNA vaccines, the DNA has been replaced by stable molecules of RNA (pseudouridine is also a pretty natural modification).

Hydroxyl group

The 2' hydroxyl group is important for RNA structure and function:

- It favours the formation of A-type helixes in double stranded RNA because the hydroxyl group tends to clash with the phosphate backbone.
- The hydroxyl group is very reactive (it favours the attack to the phosphodiester bond) so the stability of the polymer is much lower due to this additional group, RNA is a suitable molecule for short term functions.
- The sugar pucker is different, the sugar pucker is the conformation of the sugar ring, the hydroxyl group causes the presence of the **C3' endo conformation of the sugar pucker** while in DNA we have a C2' endo sugar pucker.

dsRNA

These changes (in particular the C3' endo pucker) favour an **A-type helix** in double stranded RNA, rather than the B-type helix of DNA, major and minor groove are more similar in width, but the

major groove is way deeper than the minor groove; these factors make the double stranded RNA more thick and 'spring like' (the succeeding nucleosides are packed closer to each other).

The compact character of the A-type helix permits in translation, during the pairing codons-anticodons, the formation of a mini A-double helix of RNA between the two complementary sequences.

Single stranded RNA is present in the cells without any problems, while the presence of double stranded RNA, longer than 30-40 base pairs activates metabolic responses that destroy it.
This occurs because many viral viruses genomes are constituted by dsRNA, so the cell has perceives it as non-self and uses protective mechanisms against it.

RNA folding

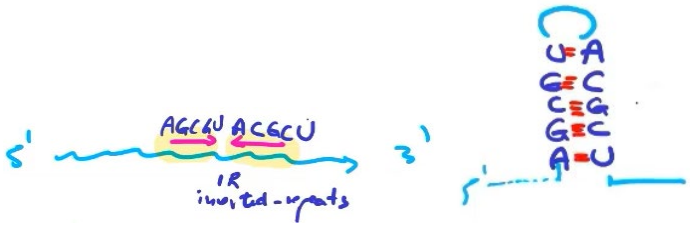
Non-coding RNAs often fold into complex, molecule-specific, three-dimensional structures and their final structure is very difficult to predict and also it is very dynamic.

- Primary structure: RNA sequence 5' to 3' direction

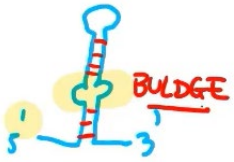
- Secondary structure: short double stranded helical regions, usually referred to as stem-loop regions.
  the main secondary structures of RNA are:

stem-loop (hairpin) **pseudoknot** **kissing loops**

These structures are formed by inverted repeats (IR), which are complementary adjacent regions of RNA



These bases are allowed to pair following both Watson and Crick bp and non-canonical ones.
These structures are usually characterised by a high **G-C content** because this is the most stable base pair, it has more hydrogen bonds (It affords a final structure with a ΔG more negative, according to the temperature of the system).
(stem-loops can be used as thermometers, when a certain temperature is reached the folded structure can be disfavoured)
another possible structure of a stem loop is the following:



in which a few nucleotides within a stem do not base pair with the corresponding reverse complement in the stem, however the structure of the stem doesn't change much.

Loops and bulges are usually very short (tetraloops or pentaloops) and their bases are **hydrophobic** so they tend to **point inside** of the structure.
Sometimes the secondary structure **expose** a few **bases outside** and these can happen to be paired with nucleotides of the 3' end of the RNA (**pseudoknots**) or with other bases pointing outside from other loops (**kissing loops**).
The effect of these phenomena is the folding in a peculiar tertiary structure, intimately linked with the function of the molecule.
To predict the secondary structure it is enough to search for

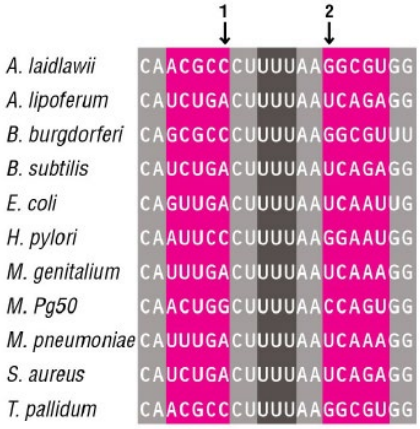
- Tertiary structure: arrangement of the double helices and single stranded regions in the final configuration of RNA.

An increasing number of genes are being discovered that produce non-coding RNA; short (6-8 bp) non-coding RNAs (like miRNA) have conserved secondary structure while long (<30 bp) non-coding RNAs have more complex folding possibilities.

Covariations

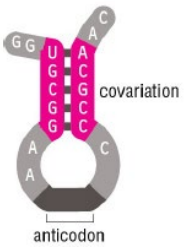
Secondary structure can be predicted by **bioinformatic analysis**, while tertiary structure is not predictable.

In order to find a secondary structure we just need to look for inverted repeats that are close enough (stems) and with a sufficient GC content; however since short repeats are enough to form a hairpin, we may find a I.R. by chance if the sequence is short.



To avoid false positive detection of stem loops, we can base our search on multiple genomes, this is because each secondary structure likely corresponds to a function in the genome; if the **function** deriving from the secondary structure we are looking for is **important**, it will be more likely that the secondary **structure** will be **conserved** among multiple species.

If the secondary structure is important it should be conserved as a structure, rather than as a sequence; we may not have same nucleotide sequence but we may look for presence of covariations of sequences that are inverted repeats.



If the inverted repeats are present in same spot in different genomes (orthologs) this is a sign that those inverted repeats will fold in a secondary structure.

The differences in the nucleotide sequence are called **covariations** because left arm variates with the right one, consistently, in order to still form a stem; we can use covariations to predict the conservation of a secondary structure (structure conservation).

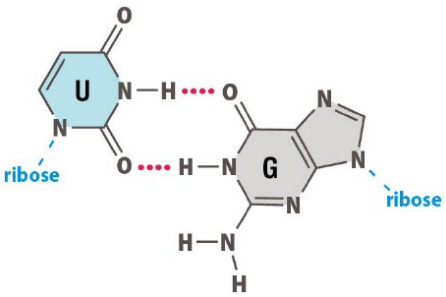
Non-Watson-Crick base pairs

RNA has many biochemical oddities that lead to unusual base pairs; some of these strange base pairs are prom by modifications of bases, invariably, this non-Watson-Crick base pairs introduce conformational stress to the complementary strands.

Wobble base pair

A biochemical oddity that explains the degeneracy of the genetic code is the wobble.
The wobble base pair consists in the pairing of uracil and guanosine (G•U), in a canonical CG the two

rings of the bases are more or less aligned on the seme plain, this leads to the canonical position of the two sugars in the double stranded molecule.



In the wobble pair the uracil can form 2 hydrogen bond interaction with guanine base, however the position of the base, and therefore the ribose linked to it, is offset with respect to the canonical; for this reason the interaction energy and the strength of the bond is weaker than an AU pair, even if both these pirs have 2 H bons.

RNA base pairs strength:
G≡C, C≡G > A=U, U=A > G•U, U•G

As a result of the wobble U can pair with A and with G, therefore U covers all the purines, this explains very well some aspects of the genetic code.

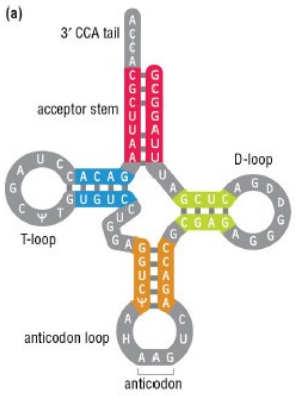
Hoogsteen base pair
Rotation of purines around glycosidic bons, the pairing is flipped.

Stabilisation of the tertiary structure
Many interactions act in order to form and stabilise the tertiary structure

Divalent cations
In order to stabilise the tertiary structure of RNA the repelling forces between the negatively charged backbone have to be neutralised.

The divalent ions allow to compact the structure, cations such as Mg2+ can bridge negative charges of two distant backbones, acting as a bridge they allow interactions between different domains in the RNA tertiary structure.

Coaxial stacking
Coaxial stack maximises the hydrophobic interactions of close hydrophobic regions



A good example of coaxial stacking is tRNA, it is one of the simplest examples of tertiary structures in RNA and it is a short molecule that can be represented on paper (secondary structure) by a cloverleaf model, in which we can find 3 super conserved stem loops.

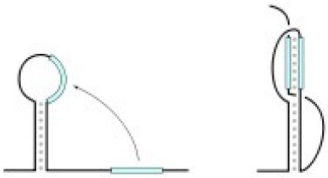
This is just a model convenient for us, the real structure in the organisms, is a much different structure similar to a T bone; this simple tertiary struct can be explained quite well and it is one of the few that we have understood.

In tRNA coaxial stacking contributes to the folding of the tertiary structure, it is the same phenomenon that grants the stab of double stranded DNA (hydrophobic bases interact and they **stack on top of each other** in double stranded helix).

In RNA coaxial stacking results to be energetically favourable, hairpins that are close by can stack on the same axis, one on top of the other, this limits the explosion of bases to the aqueous environment.

We can see in the picture how the stem loops pile up one to the other in the tertiary structure, the final structure is kind of an L because they pile up in 2 couples, in orthogonal direction.

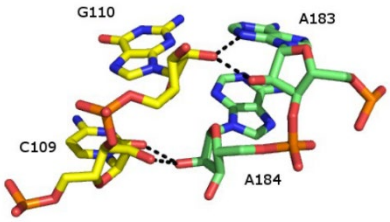
Another thing we can infer from the picture is that the side loop are not really visible, as we said before the bases belonging to the loops tend to point inwards, unless we have some sort of modification that can apply a **constraint**; for example, the anticodon loop has an H (hyper modification), which is a modification that distorts the filament in order to make the three bases point out.



Another example of coaxial stacking is the pseudoknot: the 3' end of a RNA molecule can base pair with perfect match with single stranded RNA in the loop of a hairpin in the same molecule.
When pseudoknot forms we have **two regions of dsRNA** (stem and pseudoknot) which on paper they are represented separated, while in their tertiary structure (because of thermodynamics) these two double stranded regions tend to stack one on top of the other.

Ribose zipper

The ribose zipper derives from the fact that the hydroxyl group on 2' carbon in ribose can act both as hydrogen bond donor and acceptor, this allows formation of bifurcated hydrogen bonds with other hydroxyl groups and this results in ribose zippers.



Two ribose sugars can interact by H bonding one with the other, thanks to the 2' OH two single complementary strands of RNA can form hydrogen bonds **not through the bases** but **through the backbone**; it is a sort of back to back attraction.

Triple helixes

It is quite common to have short stretches of triple RNA helixes, they are due to the formation of H bonds between 3 strands of RNA.

Triple interactions are possible because the grooves of a double stranded RNA helix are very shallow (especially the minor groove), and a third strands of RNA can enter those grooves and form hydrogen bond with the existing bases; multiple of these interactions can form a triple helix.

Ex. CGG base triples can form, the additional G positions next to a canonical pair

These interactions are not very extended, but there exist very short stretches that have super important biological roles.

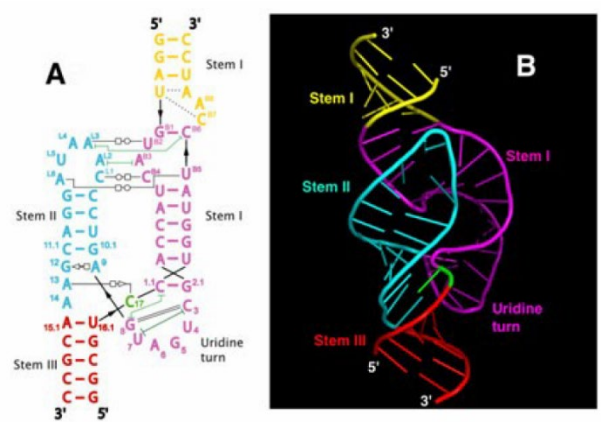
Ex. **A minor motif** is an adenosine inserted in a double helical region that is stabilised by a network of H bonds this particular motif is present in the core of the ribosome and it has an important role.
In the A site (detection) there are 2 adenosines, iper-conserved, they are part of the rRNA and they are able to enter exactly in the minor groove of the codon-anticodon minihelix, at nucleotide 1 and 2; this allows the ribosome to inspect the formation of codon anticodon interaction.
If the interaction is correct a network of hydrogen bonds is created between the adenosine and the minihelix (A minor motif), it stabilises the structure and cause a conformational change in the ribosome that is sensed and that allows the prosecution of the translation process.

Another possible triple helix is the AUU, also in this case the interactions are stabilised by the formation of hydrogen bonds

Complexity

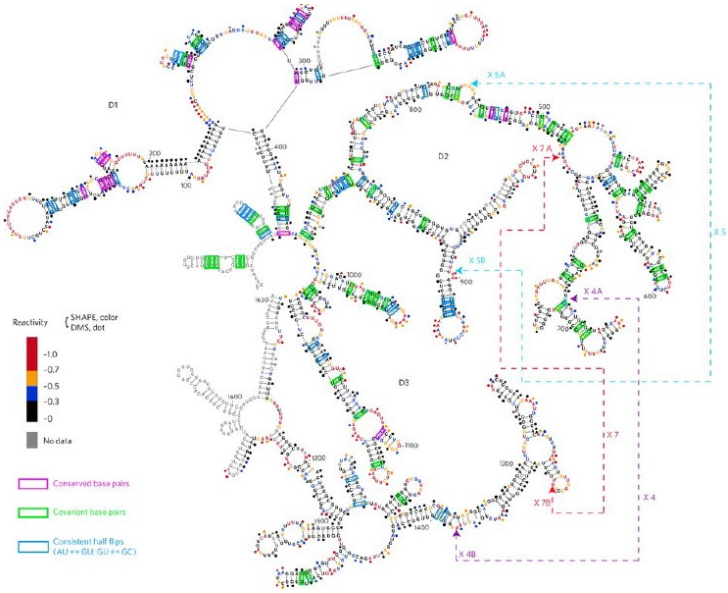
Apart from few simple RNA molecules scientists are not able to describe and predict the proper structure of RNA.

One of the known RNA molecules is the hammerhead ribozyme structure, it is a molecule that has the ability to cut RNA (it has a catalytic core).



The secondary structure model can be displayed on paper but it is pretty useless, all the interaction are represented through lines, this method is poorly efficient already with such a small molecule.

Big non-coding RNAs are very difficult to understand, also because RNA struct can be very dynamic.



CELL CYCLE

The cell cycle is at the hearth of cellular replication: each cell divides to produce two daughter cells.

The cell cycle is also important for unicellular organisms, and it is conserved in all kingdoms of life, it is crucial to divide if and only if an organism has the necessary resources to do it.

The cell cycle is a highly regulated process, its regulation is important to control proliferation of cells and a bad functioning can cause cancer (in pluricellular organisms).

Once a cell enters the cell cycle it cannot exit, if a cell gets stuck in the cell cycle it dies.

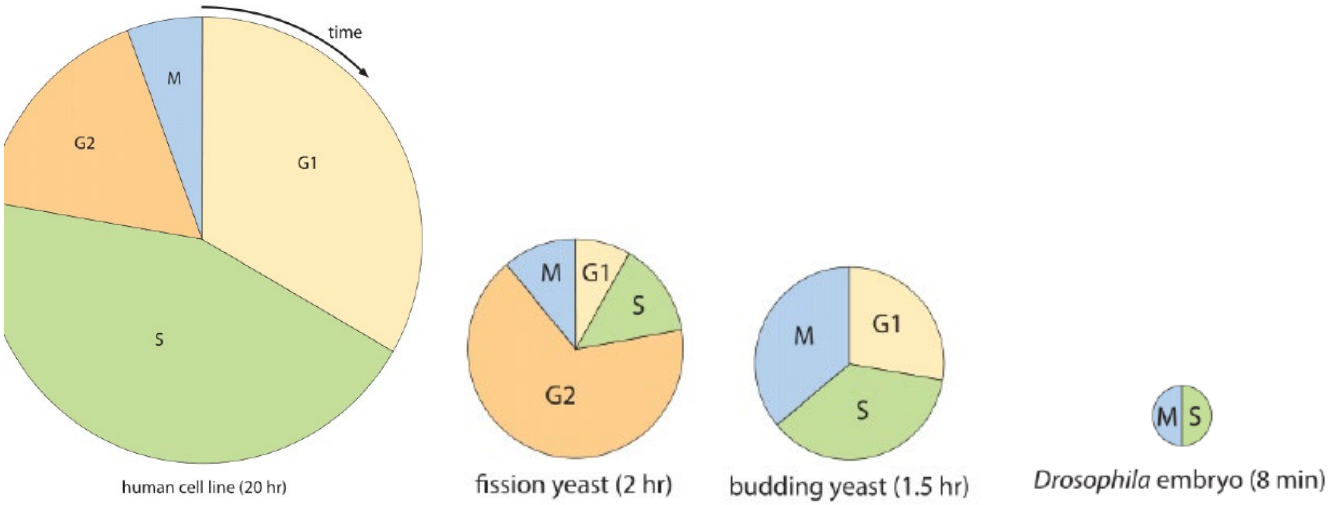
Steps:

- Interphase:
    - G1 phase: growth before chromosome duplication, at this stage a cell can exit the cycle and become quiescent (G0 phase).
    - S phase: DNA synthesis, two identical sister chromatids are created.
    - G2 phase: growth and preparation before the separation of the sister chromatids.
- Mitosis
    - M phase: mitosis and cytokinesis.

Gap phases

G1 and G2 are 2 gap phases, needed but they can be very short, each cell spends a different amount of time in them depending on its function.

Different cells have different cell cycles:



the dimension of the circumference correspond to the duration of the cycle.

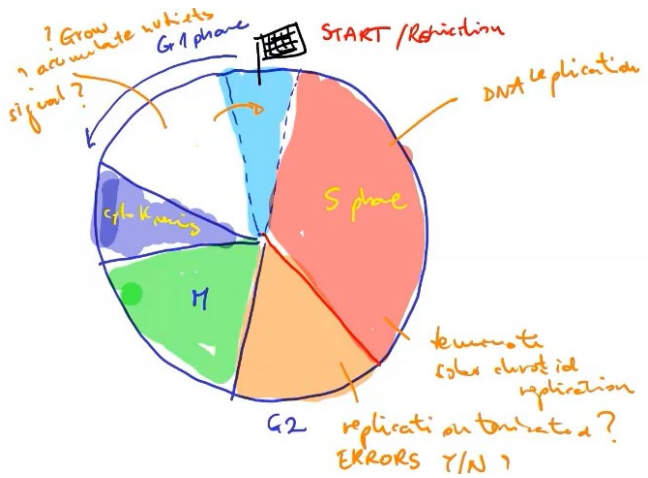
Independently from the duration of each phase, every organism has to regulate the access to each phase.

Checkpoints

During cell cycle there are a lot of checkpoints, in these checkpoints the cell takes the decision whether to proceed in the cycle or correct some errors, or storing more nutrients.

Phases

- **G1 phase**:
    - **Early G1**: cell prepares for the next duplication cycle, collecting nutrients and signals. All cells keep on replicating, it is just the timing that changes, usually a cell can just pause in the cycle.
      There is just one exception: **G0 phase (resting phase)**, common for unicellular organisms that do not have the favourable conditions to replicate or for some types of specialised cells in pluricellular organisms (nerve cells).
      G0 phase can go on and on until the right conditions are met, while G1 phase cannot be paused forever and, most importantly, during G1 phase the cell is well-awake and keeps on growing.
    - **Start**: the start of the cell cycle is a checkpoint, the cell decides to start the cell cycle if the **nutrients** are enough, if the **conditions** are favourable, if the cell has the right **dimension** (grown enough) and if there are **signals** that push the cell toward the S phase.
      Once the cell pass this point there is no turning back.
- **S phase (synthesis)**: DNA is replicated, while chromosomes are not visible the sister chromatids are synthesised, and they are held together by cohesins until mitosis.
  S phase terminates only if the replication has finished.
- **G2 phase**: checks if the S phase has been carried out correctly:
    1. DNA replication is terminated
    2. The chromatin is intact (no breaks)
    3. There are no errors
- **M phase**: the components of the cell are split in two halves.
- **Cytokinesis**: the membrane of the cell is cleaved.



CdKs (cyclin-dependent kinases)

Every phase depends on one of these kinases, different kinases pair up with different factrs to form different complexes for different phases of the cell cycle:

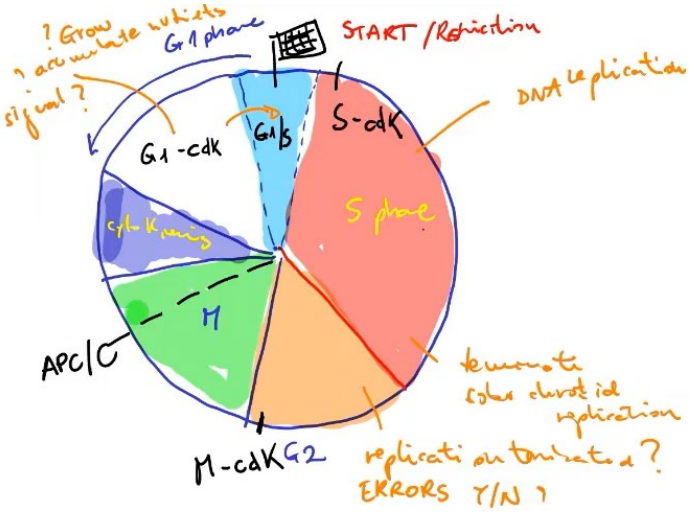
- G1-CdK complex
- G1S-CdK complex
- S-CdK complex
- M-CdK complex

Phosphorylation

As main function, kinases are able to phosphorylate groups of proteins, activating or inactivating these proteins, moreover the kinases activate sequentially one protein after the other (the first is activated and, if the first is activated, the second is activated), in this way cell cycle always go in one direction.

Cyclins

The levels of expression of CdKs do not change, they are always present in the cell, they are activated by cofactors: cyclins.
Cyclins are able to activate the enzymatic activity of kinases and can also control their specificity, each cyclin-CdK complex controls the activation of the next one.



In early G1 phase we have the G1 complex that prepares to activate the G1-S complex that will allow to activate the S complex that will phosphorilate a series of factors that will allow the replication of DNA.

Once the replication has finished the M complex is activated and it will stay high and act for the whole duration of the phase.

Decision making

The activity of these complexes varies between phases:
the level/concentration of the kinases does not vary, it varies their activity; on the other hand the levels of concentration of the cyclins change (levels of cyclins = levels of kinase activity).
if the levels of the cyclins change, the kinase activity changes.

- G1 cyclins: control the cell cycle, preparing the cell for all successive steps and govern the activity of G1/S cyclins.
- G1/S cyclins: trigger progression through the start checkpoint, they are responsible of the decision to remain or not in the G1 phase.
- S cyclins: stimulate chromosome duplication and contribute to some early mitotic events.
- M-cyclins: stimulate entry into mitosis and control it.

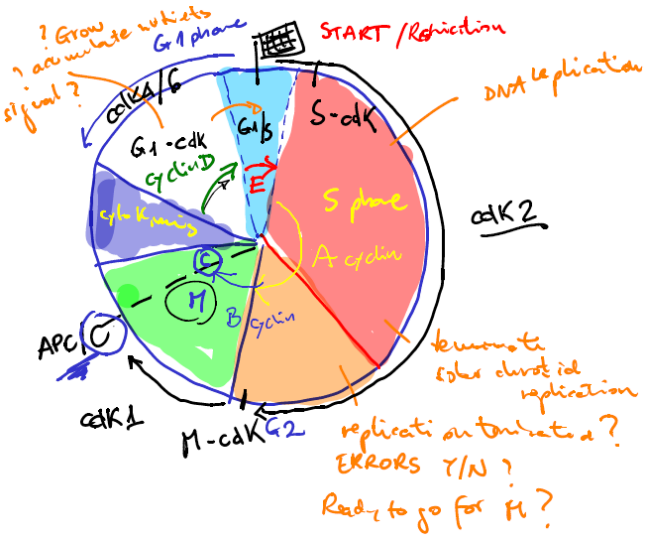
Complexes

The different complexes are represented by different combinations between kinases and cyclins.

How to remember them?
imagine you pass from start, soon after the start of the CC we can start enumerating cyclins in alphabetical order:

- From S to M: cyclin A
- To enter M: cyclin B

- In order to segregate chromosomes (terminate M phase): APC/C complex (not a cyclin) it is a factor that is important for the metaphase to anaphase transition, it gives the signal to disrupt the cohesins, it is the true climax of the cell cycle.
- early G1 phase: cyclin D
- Start phase: cyclin E



half of the cell cycle is governed by CdK-2 (from start to G2).

The most important part of it is governed by CdK-1 (mitosis).

And the last part is governed by CdK-4/6 (early G1).

Activation

How cyclins govern the activity of kinases?
cyclins can interact with DdKs through a domain that has hydrophobic patches, this specific interaction causes the activation of the kinase.

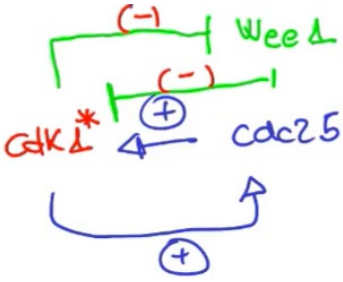
Once a cyclin interacts with a kinase, it takes out a T-loop from the kinase, this partially activates the kinase.
After that additional factors, called CdK-activating kinase, phosphorylates the T-loop (can only phosphorylated if it is exposed by a cyclin) and leads to a fully activated CdK that starts a dramatic cascade of phosphorylation activity.

One case in which this kind of activation mechanism reaches an extreme: this is the case of **M-CdK activation**.
the process is really absurd, we have to think about the complex CdK-1/cyclin B: cyclin B partially activates CdK-1 then a CAK (CdK-activating kinase) can fully activate CdK-1; however once Cdk-1 is phosphorylated in a particular activation site, the kinase Wee1 phosphorylates CdK-1 inactivating it. This mechanism has a reason, in fact a second enzyme, Cdc25, which is a phosphatase is able to remove the inhibitory phosphate and activating again CdK-1.

The important thing is that the activity of CdK-1 is inhibitory towards Wee1 and stimulatory towards Cdc25, if we have active CdK-1 it will have further activate the activity of Cdc25 and inhibit Wee1.

This is due to a molecular circuit which allows cells to respond dynamically to different stimuli.

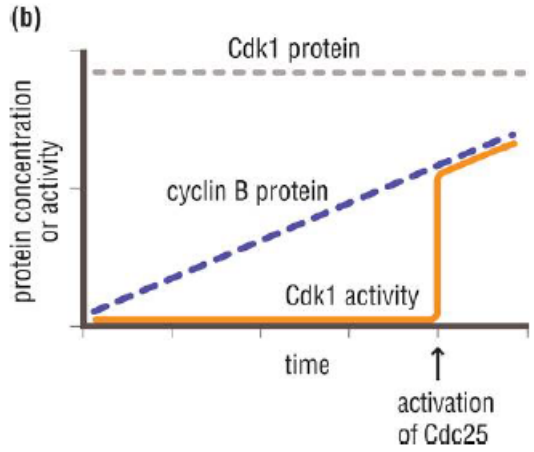
In the case of CdK-1 a double positive feedback loop is carried out: the first signal (activation of CdK-1) allows the cell to increment the activity of Cdc25 that stimulates the activation of CdK-1; at the same time the activity of Wee1 is discouraged by the active CdK and so it will not act to deactivate CdK-1.
both paths end up with the activation of CdK-1.



In positive feedback two signals aliment each other exponentially until there is a complete switch of condition and there is no come back.

As soon as we have a little bit of active CdK-1 then there is an explosion of the activity of CdK (on-off switch).

This mechanism is used from the cell to decide when to start and to be able to start the process in a coordinated and fast way.



if we look at the kinetics of expression we can see that the levels of the kinase does not change, while the concentration of cyclin B increases as the cells switches the phase.

Without the double positive loop the increasing in cyclin B would cause an increasing in the kinase activity, while, considering the role of Wee1, the activity of CdK-1 is inhibited.

Once the cyclin B reaches a certain concentration threshold, it wins the competition with Wee1 and CdK1 will become active, starting the positive loop with Cdc25, causing a rapid cascade reaction. In this way the process result fast and coordinated in time ending up with the switch from G2 to M.

Inhibitors

There are additional regulators: cyclin kinase inhibitors (CKIs), they either bind the CdK-cyclin complex or they just bind to the CdK.

Ex. P27 is an inhibitor that binds to the CdK-cyclin complex (G1-S or S phase), when it is bound the ATP binding is inhibited and therefore the activity is inhibited.

Ex. Inhibitors the bind monomeric CdKs (G1 phase), these inhibitors inhibit the binding of the cyclin and so the activation of the kinase.

Frequently these inhibitors are removed in order to pass to the next phase, this occurs through degradation; if we consider p27, to progress to S phase, G1/S promotes phosphorylation that degrades the inhibitor.
Not only phosphorylation is important but also ubiquitination, it is a a post translational modification involving binding of small ubiquitin molecules to a protein; ubiquitin promotes degradation of the protein.
These post translational modification control not only the activity but also the stability of the proteins.

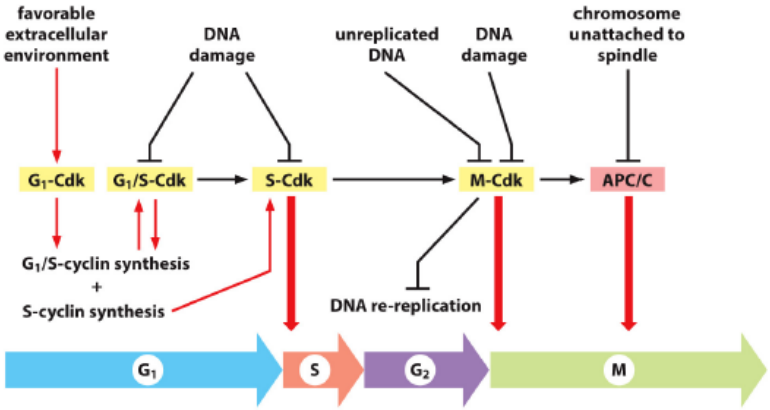
In cell cycle we have a lot of ubiquitin dependent degradation, one nice example is APC/C: usually it is inactive, it gets activated by the cofactor Cdc20 and then APC/C catalyses ubiquitination of 2 important proteins.
Securin: it is the protein that is important for keeping the two sister chromatids together, it prevents the degradation of the cohesins; so, by targeting it with ubiquitin, the cell is able to degrade first Securin and then cohesins.
M cyclins (cyclin B): after M phase, it is necessary to progress to the cytokinesis, in order to do so the cyclin B is ubiquitinated and degraded, as a result CdK-1 is deactivated (it still remains present).

Degradation is a good way to prevent the cell to go back in the cycle.

Network of switches

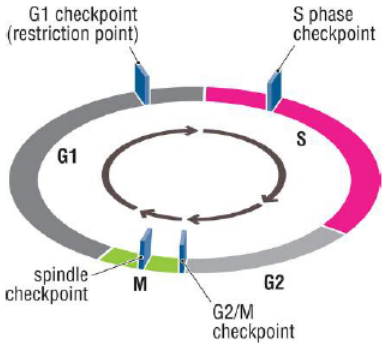
We can view the cell cycle as a network of biochemical switches that are activated only if certain conditions are met, one after the other.

- Early G1: if there are the favourable conditions the expression of cyclin D is promoted and the G1 CdKs are activated (CdK-4/6); this activation causes the synthesis of G1/S cyclin and S cyclin (cyclin E and A).
- Start checkpoint: G1/S cyclins pair with G1/S CdKs, if there is no DNA damage (useless and harmful to duplicate damaged DNA) the S-CdK is activated.
- S phase: if the DNA duplication is finished and it is error free the S-CdK activity promotes expression of cyclin B which can activate M-CdK complex.
- M phase: M-Cdk promotes the activity of APC/C if and only if the chromosomes are attached correctly to the spindle.



Checkpoints

Checkpoints are quality control mechanisms that half the call cycle when thigs go wrong and stimulate mechanisms to correct the problem.
In each checkpoint the cell needs to answer some questions:



- **G1 checkpoint**: is the environment favourable? Do I have enough nutrients? Am I grown enough? Do I have a stimulus to progress?
  in order to check if the cell has enough nutrients, there are extrinsic regulators of the cell cycle on top of these intrinsic conditions.
  nutrients and dimension of the cell are intrinsic conditions.
  extrinsic conditions are mitogens and growth factors, that act on G1-CdK and that decide to pass or not the restriction point.
  growth factors (like hormones) stimulate the increasing in cell size while mitogens stimulate the cell division.
  the growth rate of cell is regulated by kinase TOR: activated TOR promotes proteins synthesis that promotes growth rate.
  in higher eukaryotes TOR is activated by a signal transduction cascade, where cells have receptors and growth factors bind to these receptors, and they activate a signaling cascade of phosphorylation leading to the activation of TOR.
  a similar process takes place in yeast, in this case we don't have the receptor and the

signalling cascade, but the factors are internalised in the cell thanks to transporters and the presence of amino acids that stimulate the activity of TOR.
these grow factor receptors can be considered oncogenes: a gain of function of these receptors could make them active without being bound to growth factor (uncontrolled proliferation).

- **S phase checkpoint**: is DNA damaged?
  we do not want to replicate damaged DNA so this has to be checked before starting the replication.
  the damage check is carried out thanks to ssDNA, single stranded DNA is a sign of danger for cells; first of all it is quickly bound to single strand DNA binding proteins which interact with a series of factors that can halt the cell cycle.
  practically all DNA damages end up with the binding of ssDNA binding proteins and the cell cycle is halted.
- **G2/M checkpoint**: is the DNA replicated? Is the environment favourable? Are there errors in the DNA? Again the DNA check is carried out, we don't want to segregate damaged chromosomes
- **Spindle checkpoint**: are the chromosomes attached correctly to the spindle?
  this is the climax of the cell cycle: in the metaphase plate, each pair of sister chromatids needs to be contacted by the microtubules coming from opposite poles, the centromeres need to be contacted from opposite poles by microtubules to be sure that then the sister chromatids, once cohesins are cleaved, migrate to opposite poles.
  tension governs this checkpoint: a physical force coming from opposite direction is sensed on the chromosomes.

if problems occur and the checkpoints are not passed timely enough the cell commits suicide (apoptosis), if cell cycle is halted in a non-physiological way the cell decides to kill itself.

This is quite reasonable thinking about a multicellular organisation, cells have to be really well coordinated; if a cell loses control it is better to eliminate it.

Transcription of specific genes

All cyclin-CdK complexes and their kinase activity is important, because phosphorylation events that are promoted by all the different cyclins regulate other factors and especially transcription regulators.
in this way cyclin-CdK complexes are able to control the transcription, and therefore the expression, of specific genes that control the cell cycle.

These genes can:

- Boost cell cycle and promotes the cell division
- Halt the cell cycle and even that provoke the suicide of the cell

This is the real link of cell cycle with the cancer.

Ex. Complex made by Rb (oncosuppressor) that binds to E2F, which is a promoter of transcription. E2F activates gene expression, provoking cell proliferation; with cell proliferation we refer to the starting of a new cell cycle, an uncontrolled cell proliferation leads to cancer.
after the RbE2F complex is formed the G1-CdK complex (cyclin D with CdK 4/6) cause changes in Rb, through phosphorylation, E2F is released and promotes transcription.

Binding of Rb to E2F halts the cell cycle (oncosuppressor) while the phosphorylation of Rb by G1-CdK promotes the cell cycle (oncogene).

Oncosuppressor: gene whose loss of function produces cancer, negative controller of cell cycle.

Oncogene: gene whose gain of function promotes cancer, positive controller of cell cycle.

Ex. **EGF** (EGFR is an oncogene)

Growth factors: increase the cell mass, when it has reached a certain mass mitogens are stimulated.
Mitogen: promotes proliferation of the cell.

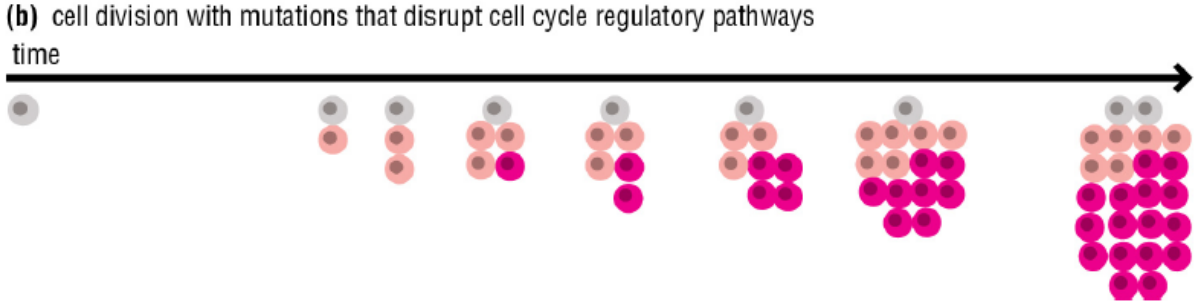
EGF is a small secreted molecule that directly stimulates cell division by activating cyclin-CdKs complexes.
EGF binds to a receptor (EGFR) and activates a cascade reaction of phosphorylation events, we call this signal transduction:
The molecules responsible for the start of the cascade do not need to enter inside the cell, but they cause a change in conformation of a receptor that transduces the signal inside the cell, giving rise to the cascade reaction.
eventually a factor is activated and it is translocated inside the nucleus, and can promote regulation of gene expression.

Ex. **TGFβ** (oncosuppressor)

It is secreted by a variety of cells and binds the TGFβ-receptor activating a signal transduction pathway that leads to the inhibition of Rb phosphorylation, it therefore prevents EGF activation (release from Rb) and halts the cellular proliferation.

Cancer

Cancer primarily occurs as a result of uncontrolled cell division.



once a first mutation causes the loss of control over the cell cycle, frequently additional mutations get accumulated, this is the real danger of cancer.

There are positive regulators of the cell cycle can be referred to as oncogenes, they promote the proliferation when they have a gain of function.
On the other hand negative regulator are tumour suppressor, when they have a loss of function they are no more able to stop the proliferation of cells (minus times minus equal plus).

Ex. p53

P53 is an oncosuppressor and it is part of the response to DNA damage; more than 50% of cancers have a mutation (loss of function) on p53.
If we have undamaged DNA a protein called MDM2 is able to ubiquitinate p53, doing so p53 is degraded and the cell cycle can continue.

If we have damaged DNA several factors are activated and promote the dissociation of MDM2 from p53, through the phosphorylation of these two players.

This causes the formation of multimers of phosphorylated p53, these multimers interact with chromatin and stimulate gene expression of genes that halt the cell cycle or that promote cell death; in this way, if we have DNA damage we want to stop the cycle until the errors are fixed.
The loss of function of p53 will prevent the cell to stop the cell cycle, the cell will continue proliferating even if it has errors in DNA causing the accumulation of other mutations.

DNA REPLICATION

Takes place during S phase, it is a semi-conservative process, the resulting two filaments of DNA are formed by one strand coming from the original molecule and one newly synthesised strand.

The DNA replication starts from discrete points, called “origins of replication”, at these sites DNA must be opened to create the **replication bubble**.

From the origin the bubble expands in both direction, by unwinding the double helix, two replication forks are set up (replication is bidirectional).

Replication is carried out by using ssDNA as a template for the new strand and it always occurs in the same direction : 5’ → 3’.

In each replication fork we have 2 antiparallel strands that are duplicated simultaneously but in very different ways: one of them is called leading strand while the other one lagging strand.

Leading strand: the replication is carried out in a continuous manner, the strand used as a template is unwound in 5’ to 3’ direction so the enzymes involved in the replication can progress linearly

Lagging strand: the replication is discontinuous, small DNA fragments are synthesised starting from the replication fork margin going back to the origin of the bubble; each fragment is called Okazaki fragment.

Phases

Initiation

The first phase of DNA replication if the initiation, it is composed of the following steps:

1. The origin of replication is recognized by an initiator protein that opens up the double helix.
2. **Helicases** are recruited to unwind the helix as the duplication progresses.
3. ssDNA is exposed and it gets contacted from ssDNA binding proteins that participate to signalling and control of the replication (it has to be triggered, or fired, just once per cell cycle).
4. To start the actual synthesis of the new strand a **primer** is needed, it is a short RNA stretch that pairs with a template DNA and that allows the polymerase to start working; it is synthesised by a **primase**, which is a special kind of polymerase that is able to synthesise in vivo, de novo, a short strand of nucleotides (RNA in bacteria).

elongation

the replication machinery progresses along the genome, in a processive (constant) manner, the polymerase can synthesise long stretches of DNA, without detaching from the template.

5. The **sliding clamp** is recruited it is essential to load the DNA polymerase on the DNA and to keep it in place, it grants the processivity of the replication.
   It is a very conserved molecule and it has a ring shape, for the replication are needed two sliding clamps for each replication fork, each one encloses a DNA strand and binds to the DNA polymerase locking it in place (it can’t fall off) through protein-protein interactions.
   In order to be casted around DNA the ring (sliding clamp) needs to be opened, the clamp loader is a spiral molecule that is able to mount the sliding clamp in place.
6. The sliding clamp recruits the polymerase, the replisome complex is complete.
7. Replication machinery moves along the DNA, copying the strands

Termination

Termination occurs when two different forks meed, when the fork reaches the end of a linear chromosome, or when polymerase meets the previously replicated strand.

8. The replication complexes are disassembled.
9. RNA primers are removed and replaced with DNA.
10. DNA ligase connects adjacent stretches of DNA.

Effectors

Follows a detailed description of all the enzymes involved in DNA replication

DNA polymerase

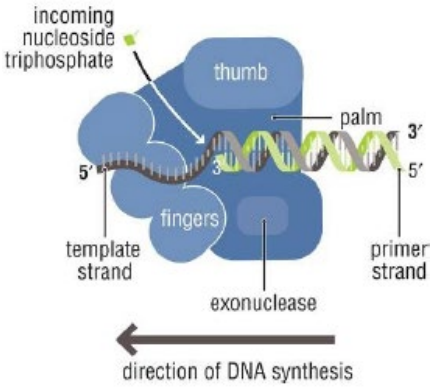
Synthesises DNA in 5'→3' direction, so it moves from 3' to 5' ends on the template filament, and the 5' end of the new nucleotide is connected to the 3' end of the strand.

Replicative polymerase needs to be accurate, it doesn't have to incorporate wrong nucleotides; this is achieved by the inspection of the template bases and thanks to the repair activity of proofreading.

There exist multiple types of DNA polymerases, each one with a different function, bacterial polymerases are super-fast: they synthesise 1000 nucleotides per second, while eukaryotic polymerases are much slower, 50 nucleotides per second, but it is much more accurate (sometimes cells may need inaccurate polymerases to overcome some errors, so each polymerase has different characteristics).

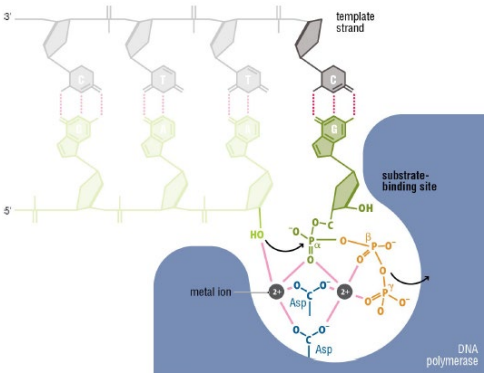
Structure

DNA polymerases have 3 domains: the thumb, the fingers and the palm



The growing double-stranded DNA fits into the palm while the ssDNA winds though the fingers. The finger domain helps position the incoming nucleotide; the thumb domain holds the elongated dsDNA.

In the active site there are 2 aspartate resides amino acids, their negative charges aid the coordination of 2 magnesium ions; many enzymes that deal with DNA has as cofactors divalent ions as magnesium, manganese, calcium.
These ions help the nucleophilic attach by the 3' OH group onto the α-phosphate of the incoming nucleotide (the closest phosphate to the alpha carbon); the nucleophilic attach releases pyrophosphates (phosphate α and β), this reaction is energetically very favourable (hydrolysis of phosphate is used in reaction coupling) and results in the formation of a phosphodiester bond.

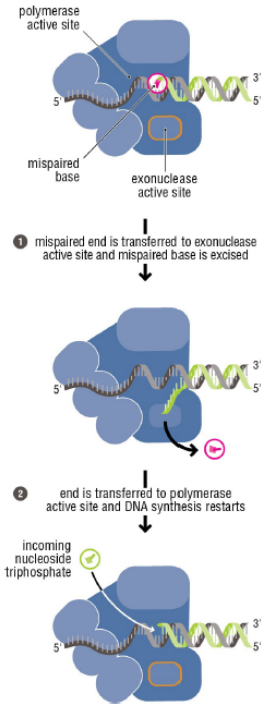


Accuracy

The DNA polymerase manages to be accurate (1 error per 10^5 nucleotides) thanks to:

- Proofreading: The proofreading is a process carried out by the DNA polymerase that consists in the removal and replacement of a wrong nucleotide that has just been synthesised.
  If the wrong nucleotide is incorporated a little conformational stress causes the slowing down of the polymerase, this allows the activation of the exonuclease activity.
  The exonuclease activity is built in the enzymes and it is actually always active, but it prevails only if the speed of synthesis is lower than the speed of exciding; it occurs in 3'→5' direction (opposite to synthesis).
  phases of the repair:
    - Energy is spend and the newly synthesized filament is unwind and flap out
    - The 3' end is transferred from the active site to the exonuclease catalytic site
    - The wrong nucleotide is removed

  This process has a high energy cost with respect to other repair mechanisms.



- Conformational constraint recognition: it doesn't require energy, it occurs during the synthesis and it prevents the wrong nucleotide to be incorporated.
  This is achieved because mismatches introduce conformational stress, the misalignment of the two base pairs causes a rejection so the nucleophilic attach is disfavoured.
  At the catalytic site there is conformational constrain that allow the nucleophilic attach only if the proper complementary base (nucleotide) pairs with the template filament.

Processivity

DNA replication is a processive synthesis, this means that the polymerase walks long distances. Many thousands of nucleotides are synthesised without reassembling the polymerase on the template strand, this is good because it allows to save a lot of resources (energy, primers, time and so on).

One exception of a non-processive DNA polymerase is the polymerase α in eukaryotes, which has to synthesise a short DNA sequence after the RNA primer and then it falls off from the template.

Other non-processive DNA nucleases are the ones in charged of DNA repair.

Types of DNA polymerase

Bacteria:

- DNA pol III: both leading and lagging strands

Eukaryotes:

- DNA pol δ: leading strand
- DNA pol ε: lagging strand

DNA polymerase families

family	polymerase	source	function
A	pol 1	*E. coli*	gap repair
	Taq	*Thermus aquaticus*	replication
	T7	phage T7	replication
B	pol α	eukaryotes	primase and repair
	pol δ	eukaryotes	replication
	pol ε	eukaryotes	replication
	pol ζ (Rev3)	eukaryotes	translesion synthesis
	pol II	*E. coli*	
	T4	phage	
	RPB69	phage	
	pol B	archaea	replication
C	pol III	*E. coli*	replication
D	pol D	archaea	replication
X	pol β	eukaryotes	gap repair
	pol λ	eukaryotes	gap repair
	pol μ	eukaryotes	gap repair
	pol σ	eukaryotes	gap repair
Y	pol η (Rad 30)	eukaryotes	translesion synthesis
	pol ι	eukaryotes	translesion synthesis
	pol κ	eukaryotes	translesion synthesis
	REV 1	eukaryotes	translesion synthesis
	Din B (Pol IV)	*E. coli*	translesion synthesis
	UmuCD (Pol V)	*E. coli*	translesion synthesis
	Dbh	archaea	translesion synthesis
	Dpo4	archaea	translesion synthesis
RT	reverse transcriptase	retrovirus	copy genome
		eukaryotes	copy retrotransposons
	telomerase	eukaryotes	elongate telomeres

Helicases

Responsible of the unwinding of the DNA that precedes the polymerase in order to make available ssDNA.

The quaternary structure of helicases is very conserved between prokaryotes and eukaryotes while the functioning is different: in eukaryotes the enzyme is assembled and moves on the leading strand while in bacteria it moves on the lagging strand.
In both cases the direction of movement is the same but the enzymatic direction is opposite (5'→3' for bacteria and 3'→5' for eukaryotes).

Structure

Helicases are hexamers, forming a donut shaped ring, in which ssDNA is at the centre.

The helicase of Escherichia coli is called DnaB: homo hexamer, formed by six identical units.
In eukaryotes the helicase is called MCM, formed by different subunits, it is a hetero hexamer.

ssDNA binding proteins

they bind to ssDNA present in the replication bubble and prevent the single stranded filament to form secondary structures; any secondary structure could slow down the polymerase and cause the take over of the exonuclease activity.

Topoisomerases

These enzymes help to relieve supercoiling, unwinding DNA causes torsional stress that must be relieved or the progression of the replication fork could be impaired.
Topoisomerases work along with the replisome in order to relax supercoils.

Slider clamp

Main effector that allows the processivity of the replication, it keeps the DNA polymerase tethered to DNA.

Structure

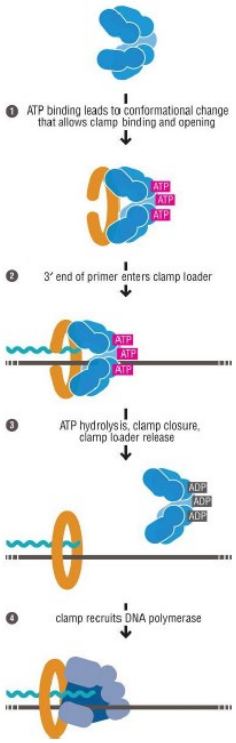
Bacteria: heteromeric, formed by 2 units (trimers).
Eukaryotes: it is called PCNA, it is an heteromeric protein, formed by 3 units (dimers).

Even if the components are different the quaternary structure is very conserved (in fact they have the same function).
A hole is present at the centre of the structure (35 A of diameter) which is highly stable and it is difficult to get it off DNA.

The contact with the DNA polymerase is achieved by the presence of a conserved repeated motive of 8 amino acids that contact the polymerase.

Clamp loader

The clamp loader opens up the sliding clamp ring consuming energy, it has a ATPase activity which is fundamental for the loading of the sliding clamp.



The ATP hydrolysis causes a conformational change in the sliding clamp that allow it to be loaded onto the DNA.

Mechanism

1. The clamp loader is activated by ATP binding.
2. Once activated the clamp loader undergoes a conformational change that allow it to bind the sliding clamp and open it.
3. This complex now gains affinity for the 3’ end of the RNA primer.
4. The attachment of the clamp loader promotes ATP hydrolysis that promotes the closure of the ring and the clamp loader dissociation
5. The sliding clamp recruits the polymerases.

Every time the synthesis restarts energy has to be spent, that’s why processivity is important.

Origin of replication

DNA replication starts from discrete point: origin of replication; initiator proteins bind to the origin (ATP hydrolysis), this is the event that defines an origin of replication.

Bacteria (E. coli)

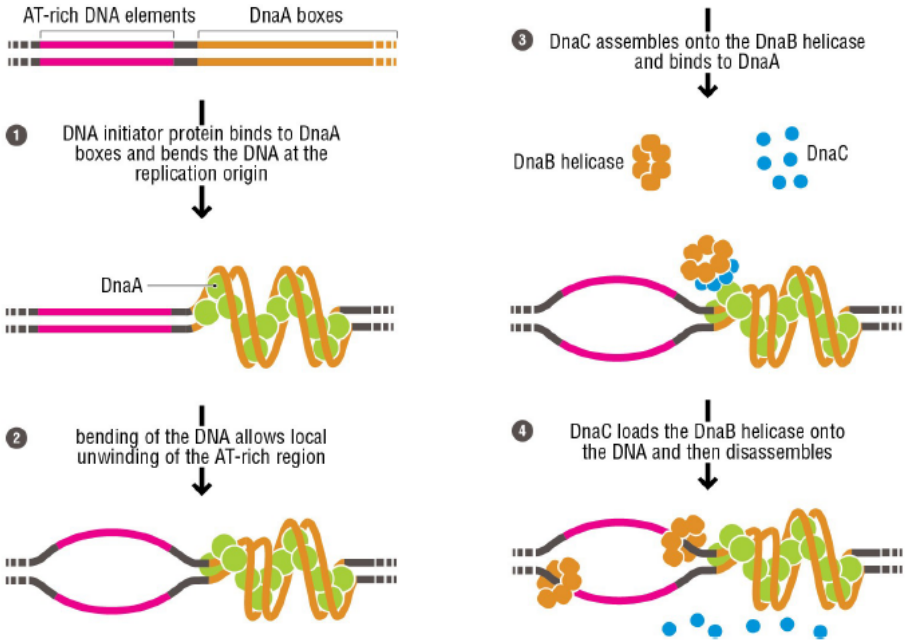
DnaA is the initiator, it binds to the origin introducing a torsional stress.

The sequence corresponding to the origin is characterised by two elements:
- An AT rich sequence, AT pairs have 2 hydrogen bonds, whereas GC pairs have 3, so AT rich regions are easier to unwind than GC rich areas.
- A 245bp sequence, with seven 9bp DnaA boxes that permit to the initiator to bind.
DnaA multimerises (in a spiral filament) and becomes able to bind to DnaA boxes, that are recognised each by a monomer of DnaA.

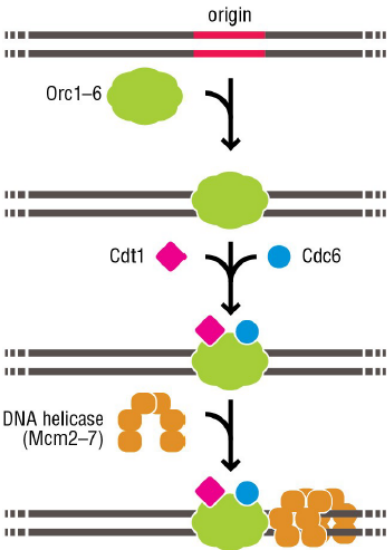
The multimerization is ATP dependent, it needs ATP to introduce torsional stress in the DNA sequence, in a weak region of DNA (AT rich) and double filament is opened.

Then helicases (DnaB) are recruited through DnaC; DnaC allows the interaction between DnaB and DnaA, the helicases are loaded on the single stranded DNA filament and then DnaC detaches.

(in bacteria DnaB is recruited on lagging strand)



Eukaryotes



Eukaryotic origins are way more complex, in simple eukaryotes like yeast the mechanism is quite similar to bacteria, an initiator complex called ORC binds to a consensus sequence, flanked by AT rich elements; then the initiator protein then recruits other proteins and the helicase at the origin of replication (in this case helicase is recruited on the leading strand).

In higher eukaryotes the initiator doesn't have a consensus, the sequence dependent nature of the origin is lost.

The origin depends on the chromatin state rather than on the DNA sequence, for example, in drosophila the initiator (ORC) is recruited in regions where histone tails are hyperacetylated.

This different organization can be explained by the size of the genomes: bacterial genome is small, circular and it has just one origin of replication, eukaryotes genomes, on the other hand, are much bigger (10^3 times bigger) and have many origins in each chromosome to assure a decent replication duration.

The origin needs to fire just once per cell cycle, so the starting of the replication has to be highly regulated.

Primases

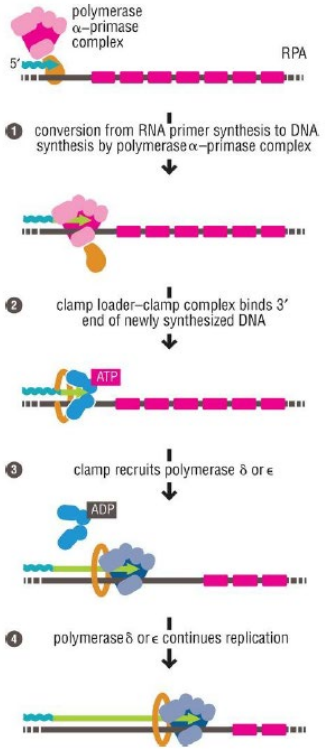
Primases are particular DNA polymerases, responsible for the synthetisation of RNA primers that will be used from the replicative polymerases.

Bacteria

The primase act simply, it consists in a DNA dependent RNA synthase, it uses DNA in order to synthesise short stretches of RNA, about 10 to 30bp, therefore it is weakly processive.
The power of the primase is that it is able to synthesise "de novo" a stretch of nucleotide, with just a DNA template.

Eukaryotes

Eukaryotic primase is formed by 4 subunits: 2 subunits constitute the primase activity (form the RNA primer) and associated with this complex there is a DNA polymerase, DNA pol α.
Therefore in eukaryotes the primer is chimeric, it contains an initial RNA stretch and then pol α takes over and continues synthetising a DNA primer.



This process initiates the polymerase switching in eukaryotes: the primase complex starts synthesizing the RNA primer then there is a conversion and the polymerase activity switches to pol α which now will start synthesise DNA, using the previously synthesised RNA as primer.

Then the clamp loader is recruited at the joining of the primer with the template strand, this allows the recruitment and loading of the sliding clamp and then of the replicative DNA polymerases (pol δ and ε), entering processive polymerisation.

This whole process occurs pretty frequently on the lagging strand, every time an okazaki fragmed is formed.

Okazaki fragments

Because DNA polymerase sysnthesises only in 5’ to 3’ direction one of the two strands of a replication fork will be in opposite verse with respect to the progression of the replisome, this strand is called lagging strand.

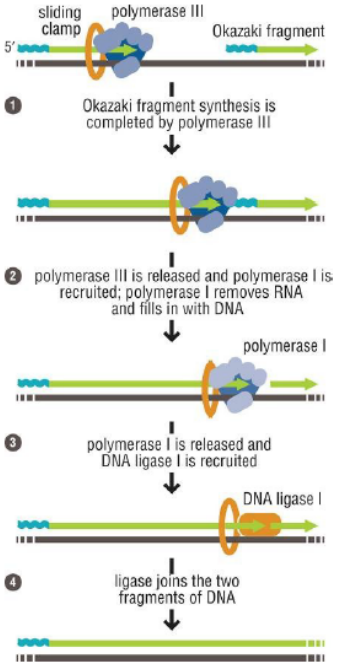
The Synthesis of DNA on the lagging strands occurs through short discontinuous fragments: the Okazaki fragments.
Polymerases on the lagging strand pocede downstream with respect to the flow of replication and, each time they encounter an already formed double strand DNA they have to stop, they have reached the border between two Okazaki fragments.

Bacteria

In prokaryotes, when DNA pol III is blocked from the primer of the previous Okazaki fragment, it dissociates from the DNA and it is switched of, while the sliding clamp remains attached.

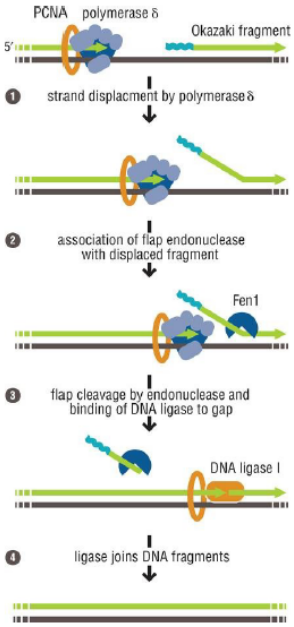
The clamp loader will recruit a new polymerase: polymerase I which eliminates the RNA bases and synthesises DNA, but it cannot attach the end of the synthesised DNA to the next DNA fragment, leaving a nick in the DNA.

DNA ligase seals the two portions of DNA with a phosphodiester bond.



Eukaryotes

In eukaryotes, when pol δ meets the RNA primer, it continues to synthesise DNA while the primer gets displaced, a flap is formed.



Flap: portion of RNA and a bit of DNA unpaired to the complementary strand but still attached by a phosphodiester bond.

The flap is removed by an enzyme, flap endonuclease (Fen1) and substituted by a DNA sequence thanks to pol I; if the flap is too long for Fen1, Dna2 can help to cleave the long flap, which is then processed by Fen1 (Fen1 is conserved in archea and eukaryotes, not in bacteria).

The final nick is sealed by DNA ligase I.

Replication fork

Each replication fork proceeds in a single direction, as a whole, even if it is actually walking on two strands of DNA that are antiparallel.

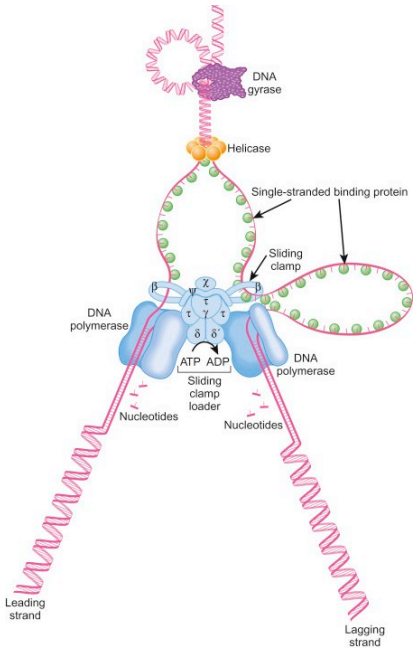
This is possible thanks to the replisome machinery: each replication fork is provided with a replisome, we can model it with a scheme that is called “trombone”.

Trombone loop: the lagging strand opens up and closes exactly like a trombone (extends and collapses), while the replisome proceeds along the strand; in this way the enzymes proceed in the same direction.

Bacteria

The bacterial replisome has two copies of DNA polymerase III, sliding clamps and clamp loader.

The polymerases are linked by two tau subunits which associates with the clamp loader and helicase.



Eukaryotes

Eukaryotic replication proteins are not physically associated in a replisome, but clamp loader, sliding clamp and primase are all part of the replication machinery at the fork.

When the lagging polymerase reaches the previous Okazaki fragment it dissociates, at the same time the clamp loader loads a sliding clamp just next to a new primer and another polymerase is recruited.

Termination

Bacteria

In bacterial circular chromosomes, termination of replication occurs at the *ter* site, on the opposite side of the chromosome from *ori*.

Ter is bound to Tus, a replicator terminator protein, when the replication forks hit Tus they stop, and the replication complex disassembles.

The remaining short piece of unreplicated DNA is filled in by DNA polymerase I.

The daughter chromosomes are entangled (intersected), this is resolved by topoisomerase II in a process called decatenation.

Eukaryotes

In eukaryotes the termination doesn't occur in specific sites and there is no need to decatenate because the molecule is linear.

When a replication fork reaches an already duplicated site it disassembles and the remaining DNA is filled in (magically).

Even if the molecule is linear, after the duplication the two filaments are wound one to the other so topoisomerases need to resolve this structure.

The end-replication problems

Toward the end of the eukaryotic replication 2 problems come up:

1. The last primer of the last Okazaki fragment cannot be substituted because DNA polymerases cannot synthesise with just a template.
   so at each DNA replication cycle fragments of DNA close to the end of the chromosome are lost.
2. The replisome cannot reach the very end of the chromosome, so an entire portion of the chromosome is lost.

Telomeres are able to solve these problems; a telomere is a repetition of short junk DNA sequences, positioned at the end of each arm of each chromosome, they have the role to preserve the important information at contained in the core of the chromosomes.

Telomeres range in length from 100bp to 20000bp, and they get shorter after each round of replication, they can be elongated by adding new specific telomere sequences at the ends, through a specific enzyme: telomerase.

Telomerase is a special DNA polymerase made of a protein and a RNA molecule (ribozyme).
The RNA provides the template for synthesis of telomere repeats (TTAGGG in humans), while the protein is an enzyme called telomerase reverse transcriptase (TERT), it is well conserved in eukaryotes, and it is able to synthetise more telomeric sequences at the end of the chromosomes.

Telomerase binds to single stranded telomere DNA (3' end) and the DNA sequence base-pairs with the telomerase RNA; the enzyme uses the 3' end of DNA as primer and the RNA as a template to synthesise DNA (each time the same small RNA sequence is used as a template).

The result is the elongation of the 3' strand of the chromosome, which can be used as a template to balance the loss of information of the replication.

Regulation of replication

As we mentioned many times the regulation of the firing of DNA replication is essential to assure just one replication per cell cycle.

Bacteria

- **ATP association control**:
  In E.coli the initiation at *ori* is primed by DnaA, binding to DnaA boxes, but this happens only if DnaA is associated to ATP and the hydrolysis of ATP is stimulated by the loading of the

sliding clamp; when the DnaA-ADP complex is formed (by hydrolysis of ATP) then the complex dissociates from the origin and duplication starts.
One simple way to control the firing of the origin in bacteria is simply to associate the DnaA to ADP, the complex will not have the capability to hydrolyse and start the replication (note: this enzyme is simply associated to ATP, not phosphorylated).

- **DNA methylation**:
  Another mechanism of control of replication is the exploiting of the methylation, in particular Dam methylation, or DNA adenine methylase, which is an enzyme that forms a covalent link of a methyl group with an adenine base; in particular Dam recognises a particular sequence, very frequent in every genome: GATC.
  This sequence is repeated 11 times at the origin of replication in Escherichia coli, and it can be hemi-methylated or fully methylated by Dam.

  Hemi-methylated: 

  Fully methylated: 

  Bacteria often use this mechanism to protect from bacteriophages, by methylation they are able to recognise their DNA from non-self one; frequently methylases come with specific endonucleases are responsible to cut the non-methylated sequences.
  This mechanism is used to **regulate replication** because the origins are full of GATC sequences, since DNA duplication is semi conservative, these sites will result hemi-methylated right after the duplication.
  After some time from the duplication the Dam methylase will have methylated all the sequences and the DNA will be **fully methylated**.
  This is the **information** that is used to **prevent accidental firing** of the origin: DnaA can only bind to fully methylated DnaA boxes so, until all the boxes will be hemi-methylated, the replication will not start
  In order to **control the timing**, there is a particular protein, **SecA**, that binds to the hemi-methylated sequences, preventing the GATC sites to become fully methylated, only once the cell is ready to duplicate the DNA, SecA will dissociate from DNA, Dam will fully methylate the DNA and the replication will start.
  (exception: E. coli in a rich environment can start a duplication before the end of the previous one, this phenomenon is called nested replication)

Eukaryotes

The initiator complex is ORC, it needs to bind to the origin and then it recruits factors that help the recruitment of helicases.

There is a complex, pre-replication complex (pre-RC), that is formed before the S phase, in G1 phase, just before passing the restriction point.

Pre-RC is formed by ORC, helicase MCM and other factors.

To initiate duplication the pre-replication complex needs to be activated through phosphorylation (modification at protein level)

The signal is triggered by the S-cdk complex, which has kinase activity and it will phosphorylate ORC and other cofactors that will trigger the helicase phosphorylation and, therefore, their activation; the replication starts.

The phosphorylated ORC will stay at the origin up to the completion of DNA replication, this will prevent that the occupation of the origin region by a non-phosphorylated ORC, the replicative function of ORC is inhibited.

The phosphorylation series of S-cdk involves proactive phosphorylation (helicases) and inhibitory phosphorylation (ORC).

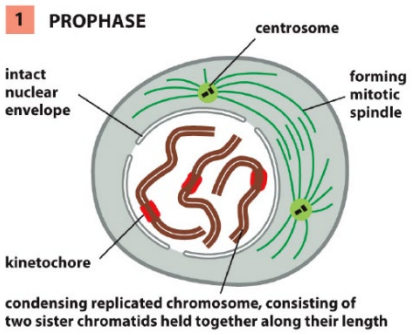
MITOSIS

The mitosis is the climax of the cell cycle, it starts after the G2 chekpoint.

Phases

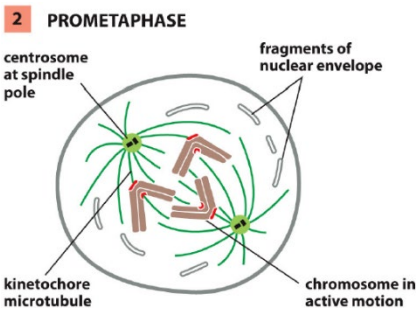
Prophase

At prophase the replicated **chromosomes**, each consisting of 2 sister chromatids, **condense** while the nuclear **membrane** starts **melting down**.
Outside the nucleus, the **mitotic spindle assembles** between the two centrosomes, which have replicated and moved apart.



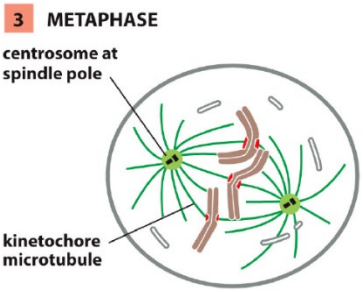
Prometaphase

In prometaphase the nuclear envelope is completely broken down, chormosomes can now attach to spindle microtubules via their kinetochores and undergo acrive movement (they are pushed toward the centre of the cell).



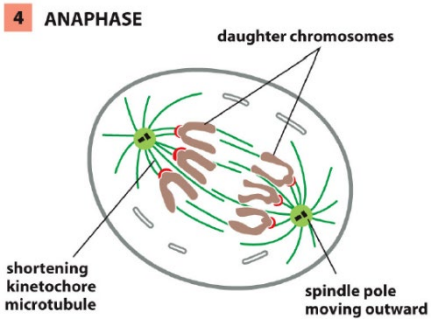
Metaphase

At metaphase, the chormosomes are aligned on the metaphase plate, midway between the spindle poles.
The kinetochore microtubules attach sister chormatids to opposite poles of the spindle (APC/C checks the integrity of the spindle attachment).



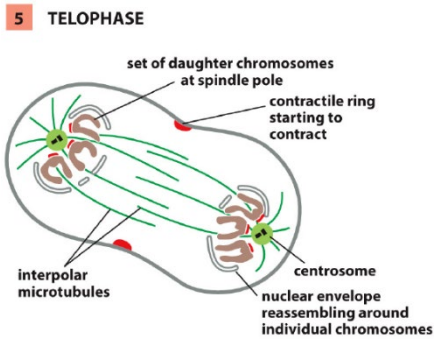
Anaphase

At anaphase, the sister chromatids synchronously separate to form two daughter chromosomes, and each is pulled slowly toward the spindle pole it faces.
The kinetochore microtubules get shorter, and the spindle poles also move apart; both processes contribute to chromosome independent segregation.



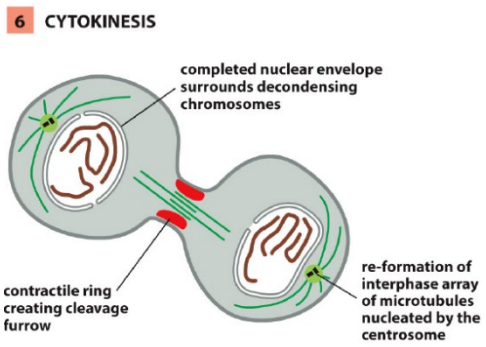
Telophase

During telophase, the two sets of daughter chormosomes arrive at the poles of the spindle and decondense. A new nuclear envelope assemples around each set, completing the formation of two nuclei and marking the end of mitosis.
the division of the cytoplasm begins with contraction of the contractile ring.



Cytokinesis

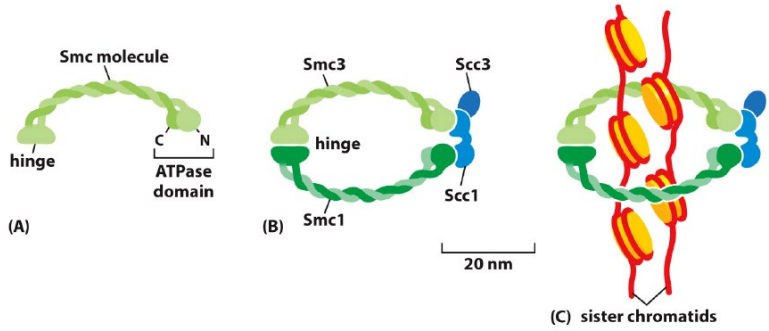
During cytokinesis, the cytoplasm is divided in two by a contractile ring of actin and myosin filaments, which pinches the cell to create two identical daughters.



Cohesins

The cohesins are the proteins responsible to keep the sister chromatids together and they facilitate the attachment of the sister chromatids to the poles of the mitotic spindle.

Cohesins are complexes formed by 4 proteins: Smc1 and Smc3 proteins (odd), large proteins formed each by coiled-coil, with a terminal domain that has an ATPase, and two additional less important factors.

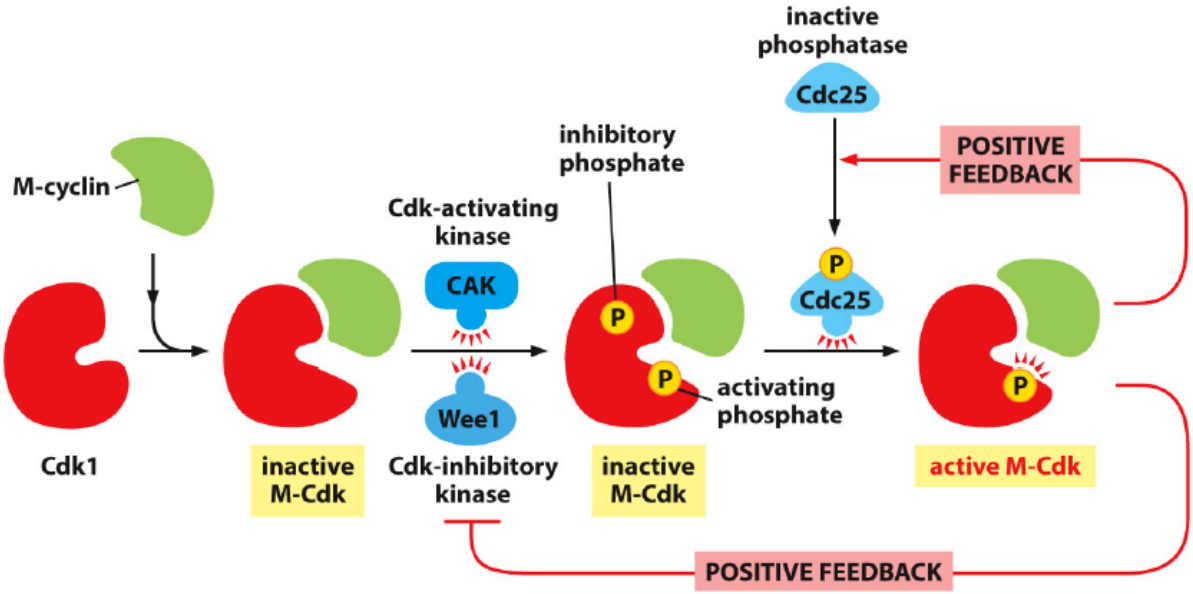


The result is a ring structure that encircles the sister chromatids, they are loaded around the DNA thanks to their ATPase activity.

M-CdK

M-CdK complex, formed by CdK-1 and cyclin B, drives the entry into mitosis; we already discussed the mechanism in the “activation” paragraph, in CdK complexes chapter.

Just to recap:

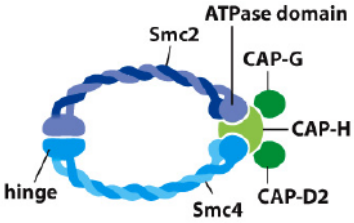


1. Cyclin B partially activates Cdk1
2. M-CdK complex is activated through phposphorilation thanks to CAK
3. M-CdK is inhibited through phosphorylation thanks to Wee1, in this way the levels of cyclin can increase while the activation degree of the kinases remains low.
4. The levels of cyclin B eventually reach a threshold above which the first kinases start activating.
5. The active M-CdK complex give rise to a double positive feedback: it promotes the activity of Cdc25, an enzyme that is able to activate the inactive form of M-CdK, and it inhibits the activity of Wee1.

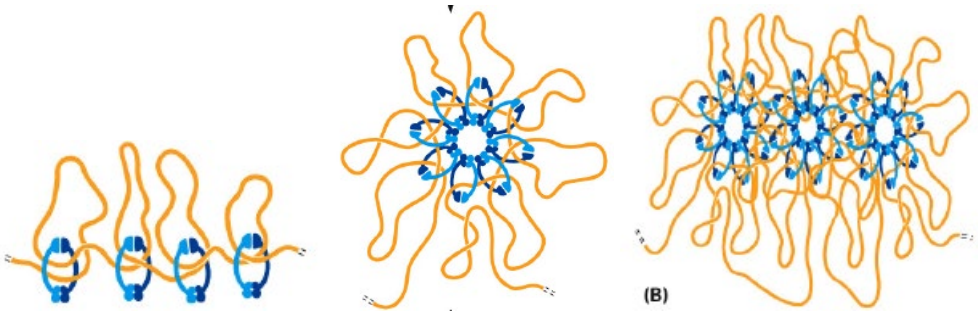
Condensin

The cascade reaction of CdK-1 also stimulates the coiling of chromosomes, after the duplication they are ordered, disentangled and separated.

Condensation and resolution depend (in part) from **condensin**, a five-subunit complex similar to cohesin, formed by Smc2 and Smc4, coiled-coil proteins (even) with an ATPase domain at one end, that are held together by three additional subunits.



Condensin forms ring like structures that allows to form DNA loops, then, by protein-protein interaction, these loops are condensed and clusters are formed; the DNA gets compacted.



Mitotic spindle

Microtubules become extremely dynamic during M phase, they grow and shrink dramatically.

Structure

The core of the mitotic spindle is a **bipolar array of microtubules**.

There are many different types of microtubules: interpolar microtubules, kinetochore microtubules and astral microtubules.

The plus ends of the **interpolar microtubules** overlap with their cognate plus ends of the possite pole in the cytoplasm, resulting in an antiparallel array in the midzone.

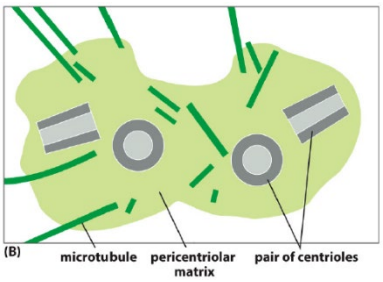
The plus ends of the **kinetochore microtubules** are attached to sister-chromatid paris at large protein structures called kinetochores contacting the centromeres of each sister chromatid.

Many spindles contain also **astral microtubules**, radiating outwards from the poles and contacting the cell cortex with their plus ends, helping to position the spindle in the cell by pushing in each direction.

The result is a wide interaction of opposing forces thea establishes a precise structure.

Centrosomes

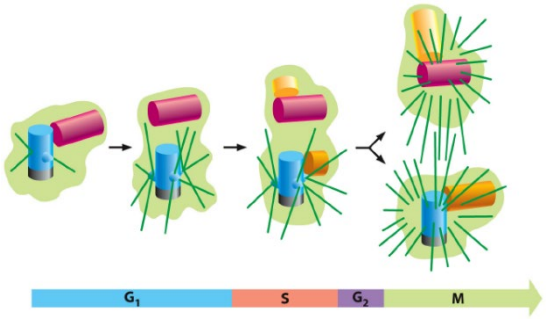
The **minus ends** of the microtubules are nucleated (grouped) in the centrosomes, they are the organelle in wich the spindle poles are focused.



The centrosome consists in the **pericentriolar matrix**, surrounding a **pair of centrioles** aligned at right angles to each other (90°), and the **microtubules** that potrube out from this structure.

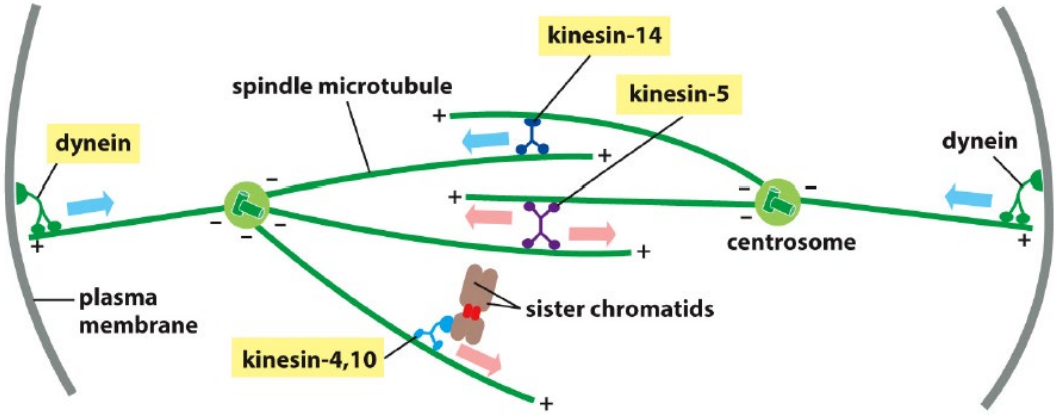
Centrioles

Usually a cell is provided with just one pair of centrioles, to carry out the mitosis two pairs are needed, so the cell starts the preparation early in the cell cycle: in G1 phase.
Once the two couples are obtained they remain coupled in a single centrosomal complex until the mitosis starts, at that point the complex splits and the two centromeres start migrating to the opposite poles of the cell.



Motor proteins

Since tension is created, different motor proteins connect to the microtubules and help focusing the mitotic spindle, they are called kinesins.



Four major types:

- **Kinesin 5**: it has **2 motor domains**, each of which associates with a different antiparallel **interpolar microtubule**, coming from opposite cell poles.
  every motor domain of every kinesin have a **direction**:
  - **Minus end** direction
  - **Plus end** direction

  Both motordomains of kines 5 have **plus end direction**, this means that kines 5 **pushes** the 2 **poles apart**.
- **Kinesin 14**: it has only 1 motor domain, but still connects two antiparallel interpolar microtubules: it is bound to 1 microtubule and the motor domain is on another interpolar microtubule projecting out from the opposite pole.
  The motor of kinesin 14 has **a minus end direction** movement, this means that kinesin 14 tends to **pull** the two **poles together**.

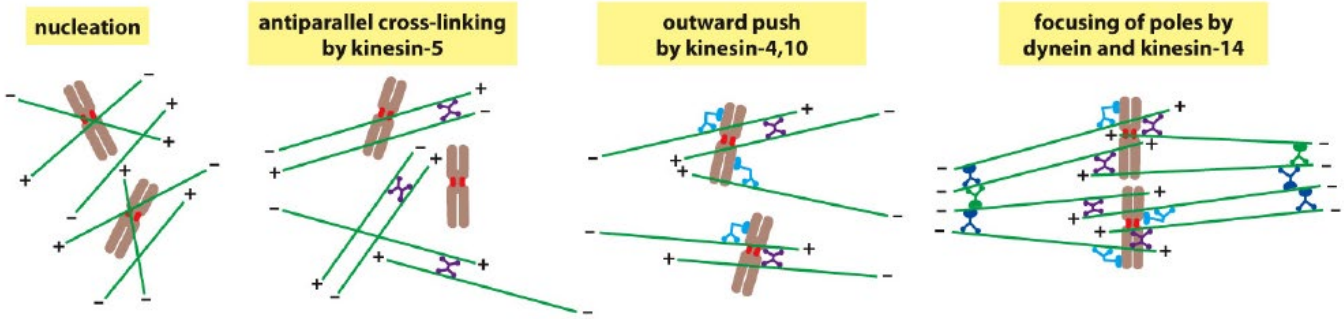
- **Dynein (motordomain protein)**: connected to the cell cortex, its motor domains have minus direction and therefore they pull the centrosome to which they are attached toward the membrane.
- **Chromo kinesins (kinesin 4 and 10)**: they have one motor domain that has plus end direction and they are associated with chromosome arms, their movement pushes the chromosomes away from the poles, toward the metaphase plate.

Self-organisation

The spindle formation is a self assembling structure with a bottom-up organisation, it uses a trial and error strategy:
First of all in order to complete the formation of the mitotic spindle the nuclear envelope has to be broken down, so that the astro microtubules reach the cell cortex and focus the spindle.
The melting of the nuclear membrane causes an increase in the dynamic instability of microtubules, the **catastropy-rescue** balance is instaurated (catastropy=plus end of the microtubule shrinks down; rescue=plus end is built up).

During M phase there is an enormous increase in the dynamic insability, by trial and error the microtubules menage to explore the space and get in contact with the chromosomes or with each other just by chance (by increasing the noise of the system the likelyhood of possible contacts increases).

After **microtubules nucletation** (formation of microtubules), the microtubules will enter the catastropy-resue balance and, eventually, two interpolar microtubules will be crosslinked and fostered by kinesin 5.
Now kinesin 5 will start pushing the two poles apart, during this process some microtubules, again by chance, will hit the chromosome arms and then kinesin 4 and 10 will start pushing them away from the poles.
Eventually, kinesin 14 and the dynein will act balancing the other forces and focussing the spindle at the centre of the cell.

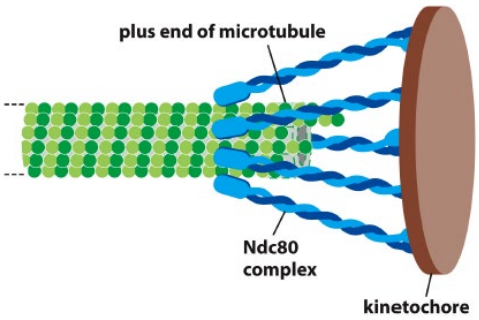


Kinetochore

The formation of the spindle is accompanied not only by a higher dynamic instability of the microtubules but also by the **stabilisation** that is obtained when the microtubules make contact with something.

For example, when a **microtubule contacts a chromosome**, that microtubule will be stabilised, this is the **role of the kinetochore**; each chromosome in mitosis is formed by two sister chromatids, each of which has a kinetochore associated with the correspective centromere.

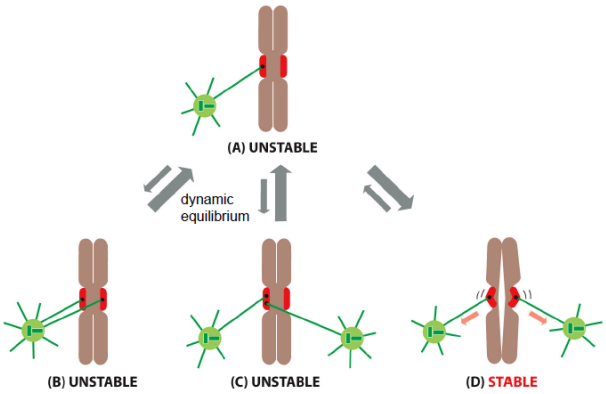
Kinetochore has a basketlike structure that has a plate and a complex of coiled coil molecules, NDC80, that is able to make contact with a microtubule and to stabilise its plus end.



Once a couple of sister chromatids has contacted one microtubule it has been "grabbed from one end"; now the cell, before entering the anaphase and segregating the sister chromatids, has to be sure that the centromeres (therefore kinetochores) of both sister chromatids contacted by microtubules that project out from opposite poles, the chromosome has to be "pulled from both directions".

Bi-orientation and tesion

Ideally in the metaphase plate the kinetochores of sister chromatids should look in opposite direction, reaching the bi-orientation layout, to be able to check this characteristic the blind cell exploits the way the kinetochores are built.
Kinetochores are organised in a back to back fashion at the centromeres, they are opposite faces of the same coin, this reduces the probability of having the two centromeres contacted by the microtubules projecting out from the same pole.

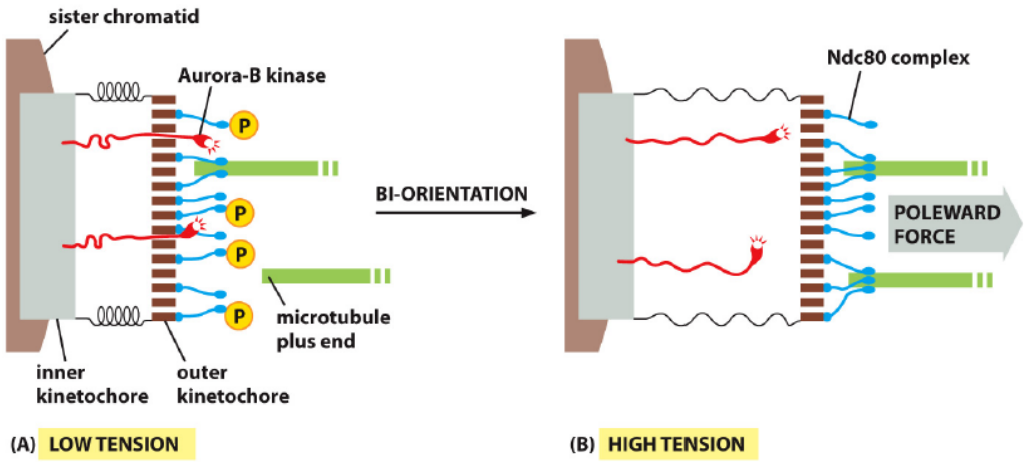


What must be achieved in order to enter anaphase (crucial checkpoint) is **tension**, the tension on the sister chromatids has to be sensed.
Tension stabilises the whole arrangement, if tension occurs then the contact of microtubules is stabilised by kinetochores, if tension does not occur then the contact of microtubules to kinetochores is reduced, the situation stays unstable and the microtubules will keep on triyng to make contact with something else by chance.

Sensing tension

In order to have tension it is needed not only the force that pulls the chromatids apart (**microtubules**) but also an opposite force that keeps them together (**cohesins**).

This tension is sensed at the kinetochore (it is the one that decides whether to stabilise or not the attachment of a microtubule): each kinetochore is formed by two structures, the inner kinetochore that makes contact with the sister chromatids and then an elastic connection to an outer plate.
The Aurora-B kinase is a kinase that has a long linker that connects the outer and inner kinetochore, its enzymatic activity is elapsed at the outer tip.



This structure is able to sense tension simply because of the variable distance between inner kinetochore and outer kinetochore:
If there is there is high tension, the distance will be large, if there is no tension the distance is shorter.

When the distance is short the active site of Aurora-B is outside of the outer kinetochore and it is able to phosphorylate the NDC80 (the coiled coil protein that are able to contact and stabilise the plus end of microtubules); **when NDC80 is phosphorylated it cannot stabilise the microtubules**.
If eventually a suffiecient number of microtubules attaches to the NDC80 complex and this interaction is stable enough, a pulling force starts extending the distance between inner and outer kinetochore and thereby Aurora-B will not be able to phorphorylate NDC80 and this will stabilise the connection with the microtubules.

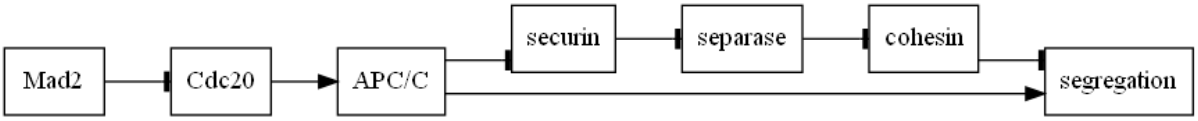
The kinase activity of Aurora-B is reduced if tension is created, this is the sensor that starts the anaphase.

Note: this is not the only force that acts on chromosomes as we said before also the chromosome arms are contacted by microtubules, in this case the involved **force pushes the arm away** from the poles and this force instaurates an equilibrium with the pulling forces acting on the centromere. that's way in anaphase the chromosomes are U shaped).

APC/C checkpoint

The decision making for the metaphase to anaphase transition is APC/C.
The decision is mediated by degradation of proteins, in fact once **APC/C binds** to **Cdc20** it becomes active and it is able tu **ubiquitinate securin**, that is an **inhibitor** of **separase** which **cuts cohesins**.
In presence of active APC/C the output of this circuit is the melting down of cohesins and the release of the tension, causing the fast migration of the chromosomes to the poles.

Mad2 together with other proteins inhibiths Cdc20; Mad2 binds to unattached kinetochores, it is a check, useful to prevent the anaphase if there are kinetochores not contacted by microtubules (if we enter anaphase with a non-contacted sister chromatid it will not segregate).
If there is no Mad2 to bind to kinetochores this is a signal for activation of APC/C.



TRANSCRIPTION

The transcription process is the first step of the gene expression, it generates RNA molecules from a DNA sequence, exploiting RNA polymerases.

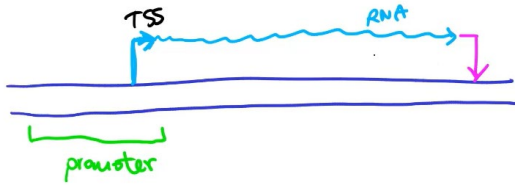
RNA is single stranded (dsRNA is a sign of danger), it is very reactive (OH group) and it has a 3D structure, it is synthesised by RNA polymerases which are very similar to DNA polymerases, they synthesise in 5'→3' direction.

Transcription is a very regulated process, in multicellular organisms each cell expresses different genes in different moments and at different rates, to be capable of this, regulators are fundamental to direct the activity of RNA polymerases.

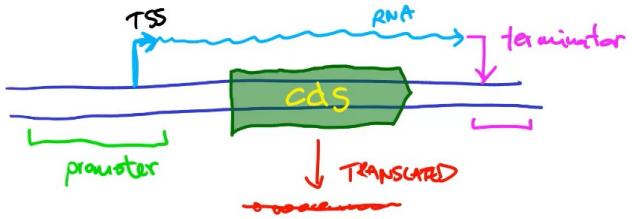
Overlook

The promoter is the first requirement for transcription, it is a nucleotide sequence in DNA that identifies the region that needs to be transcribed, it is bound with several factors including RNA polymerase.

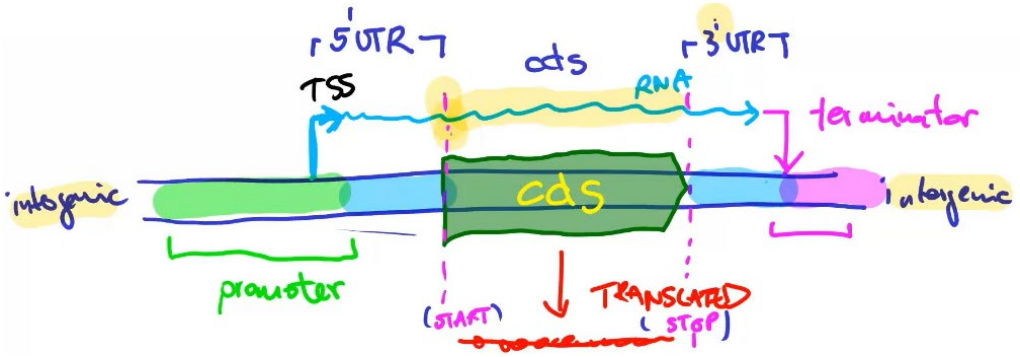
The first nucleotide incorporated in the transcipt is called TSS (transciption start site).



The gene that codes for a protein is called CDS (coding sequence), promoter and termitator (stop sign is detected) are the end parts of the gene (they are part of it tho).



Also the coding region has a start and stop signs, these translational start and stop are nested into transcriptional start and stop.
In the trascript generated by RNA polymerase we have to distinguish 3 regions: the coding sequence, a 5' UTR and a 3' UTR; UTR stands for untranslated region, they are stretches of RNA that will not be translated.



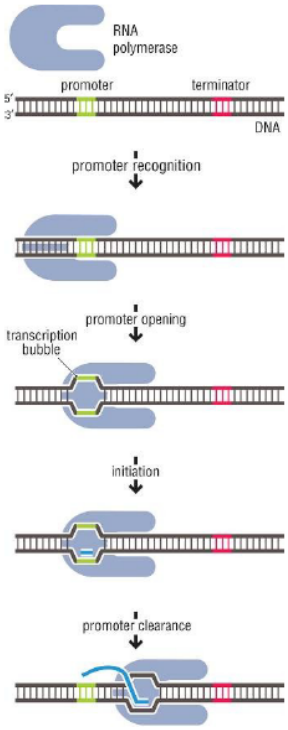
Intragenic regions are sequences in between genes

Phases

Initiation

Inititation consists in the recognition of the promoter and the ssDNA template filament is made avaliable.

Two phases and two complexes formed by the RNA polymerase with the DNA.



- **Close promoter complex**: polymerase is associated with the promoter but the double strand has not be unwound yet, the two filament are close and there is no ssDNA avaliable to be transcibed.
- **Open promoter complex**: the transcription bubble is formed, ssDNA is used as a template to synthesise RNA; the transciption bubble is a short stretch of ssDNA that will follow along the RNA polymerase during the elongation process.

The formation of the open promoter complex allows the start of the anabolic activity of RNA polymerase.

- **Abortive initiation**: RNA polymerase starts synthesising short transcripts but it is not able to leave the promoter, this phenomenon is called abortive initiation.
- **Promoter clearence**: if the RNA transcript is sufficiently long then there is a conformational change in the polymerase and the association with the DNA and the transcription bubble becomes stronger and the synthesis becomes processive, leaving the promoter and starting the elongation.

A lot of gene regulation happens at this phase, this is because blocking the expression of a gene at the beginning of its transcription is surely the most efficient and energy saving way to do it.

Elongation

The polymerase moves along the DNA with the transcription bubble and the RNA filament that is elongating, the complex proceeds in direction 3'→5', while the synthesis has direction 5'→3'. Since the transcription bubble moves, a constant unwinding (ahead) and winding (behind) of the DNA occurs at the borders of the bubble.

Termination

When the stop signal is encountered the polymerase dissociates from DNA and the RNA filament is released.

Eukaryotic transcription

Eukaryotic transcription is way more complex than bacterial one, since the chromatin is organised in nucleosomes there are additional constraints:

- Uncast of nucleosomes ahead of the transcription bubble.
- Formation of nucleosomes behind the transcription bubble.

Nucleosome remodelling enzymes that allow the disassembling and repositioning of nucleosomes, for example, histone chaperones that capture the nucleosome subunits and other enzymes that modify the histone tail to allow a processive process.

The modificaiton of histione tails are also a great way to regulate transcription in eukaryotes.

Effectors

Bacterial and archea have just one RNA polymerase, however all the polymerases are multysubunit enzymes, they have a very complex quaternary structure and they are formed by polypeptides generated by different genes.

In eukaryotes there are multiple polymerases, different genes use different polymerases:

- **RNA polymerase I**: transcribes ribosomial RNA (rRNA), works on ribosomal genes.
- **RNA polymerase III**: transcribes transfer RNA (tRNA), these two polymerases are dedicated exclusively in the transcription of RNAs involved in the translation.
- **RNA polymerase II**: messanger RNA and small regulatory RNA, transcribes all the other types of RNA, it is the most important polymerase.

Plants have additional polymerases (IV and V).

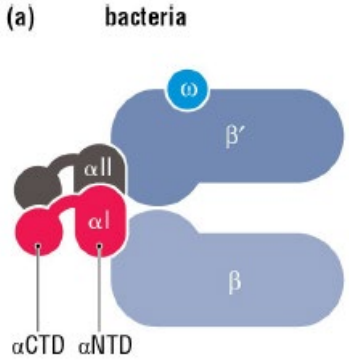
Eukaryotic polymerases have a core structure similar to procaryotic one but they are more complex and they are formed by **12 subunits**.

Bacterial transcription

Effectors

The bacterial polymerase is formed by 5 subunits:

- **2α**
- **β**
- **β'**
- **ω**



which form the **core enzyme (2αββ'ω)** that can pair with

- **σ**

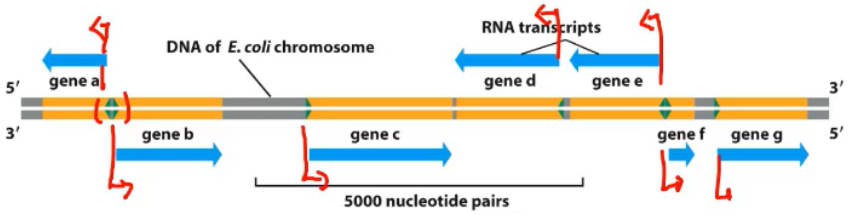
to form the **holo enzyme (2αββ'ωσ).**

α C and N terminal

The two alpha subunits in bacteria are devided into two subdomains, the C terminal domain and the N teminal domain.
The N terminal domain (αNTD) interacts with β and β' while the C terminal domain (αCTD) is very important for regulation because it contact and bind the DNA, it can reach a specific sequence or interact with different proteins, changing the affinity with DNA; the two domains are joined by a flexible linker.

Operons

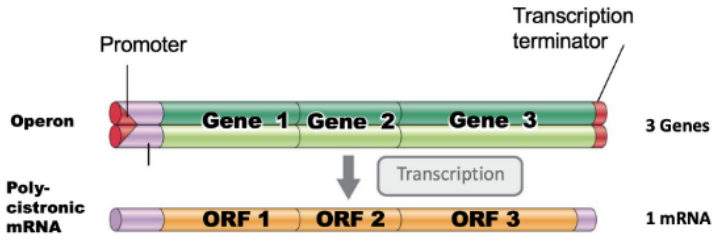
Promoter and teminator are part of the information of the gene, they also define the organisation of the genome since they define intergenic regions.



Each promoter defines a transcriptional unit.

Because bacterial genomes are small, frequently transcriptional units contain more than one gene, this means that multiple genes have the same promoter, forming a unit called **multi cistronic**

operon.
An operon is a series of genes that are regulated the same because they are generated from a single transcript, so the regulation of the transcription of that transcript regulates the entire series of genes.

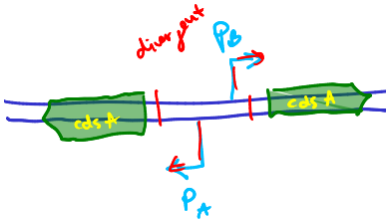


Each gene is called cistron.

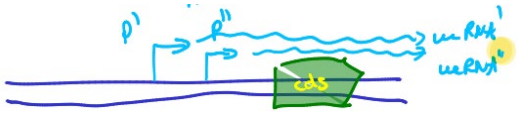
Monocistronic operons on the other hand are transcriptional units with one promoter and one gene.
Multicistronic operons are important because they allow various genes to be regulated with the same signal, and this makes a lot of sense especially if those genes are part of the same pathway.

Types of promoters

- divergent: promoters point out to different directions

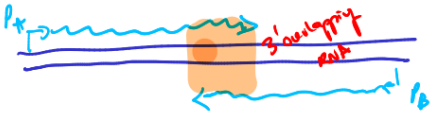


- overlapping:



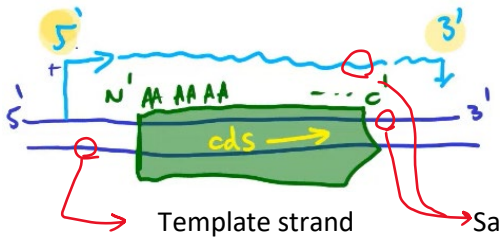
- overlapping antisense: causes a 5' overlap of RNA

- antisense promoters with overlapping transcript: causes a 3' overlap of RNA



Sometimes the antisense overlapping RNAs can form dsRNA, that, as we well mentioned before, is a sign of danger for the cell and gets degradated quickly, this is another way that cells have to control gene expression.

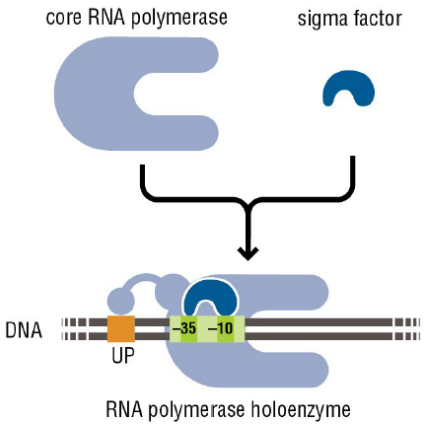
Important convention



By convention we represent the RNA transcript of a DNA template with the same direction of the antisense strand, from the opposite side with respect to the template strand (not antiparallel).

Initiation

RNA polymerase by itself is not able to recognise promoter sequences, it needs help, this help is provided by the sigma factors.



Thanks to the sigma factor the holoenzyme is brought onto specific sequences that are found in the promoter and these sequences have motives that are conserved.

The more a promoter sequence is conserved or close to the conserved sequence, the tighter is the association with RNA polymerase and the higher is the transcription rate.

Sigma factors

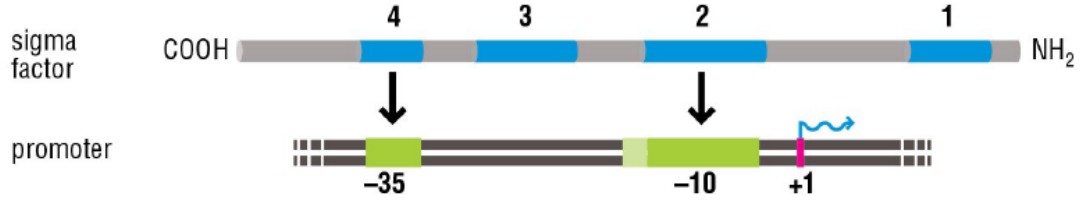
The nucleotide sequence in the promoter is read out by the sigma factor

Bacteria have different types of sigma factors: the **housekeeping σ factor** controls transcription of genes that are always expressed, while **alternative σ factors** are used as a regulatory mechianism, in fact, by swapping the sigma factor the consensus motive that is regognised is different, in this way the transcription machinery is redirected from one type of promoter to other types of promoter.

Promoter

Consensus sequences

σ^{70} or vegetitative σ factor or σ_v is the housekeeping factor, it is responsible for the recognition of two elements (boxes) that define the **core promoter**, the -10 element and the -35 element.
The **-10 element** has the consensus sequence **TATAAT**, it is recognised by **domain 2** of the sigma factor (there also exist extended -10 elements: TGnTATAAT that provide a further boost to the transcription).
The **-35 element** has the consensus sequence **TTGACA**, it is recognised by **domain 4** of the sigma factor.
These sequences are placed respectively 10 and 35 nucleotides upstream from the start of the transcription.



Almost every gene has this vegetative promoter.

Alternative sigma factors are needed in particular growth conditions, they have a completely different consensus that will attach porely to vegetative promoters and will interact with the genese that have alternative promoters.

Spacing between boxes

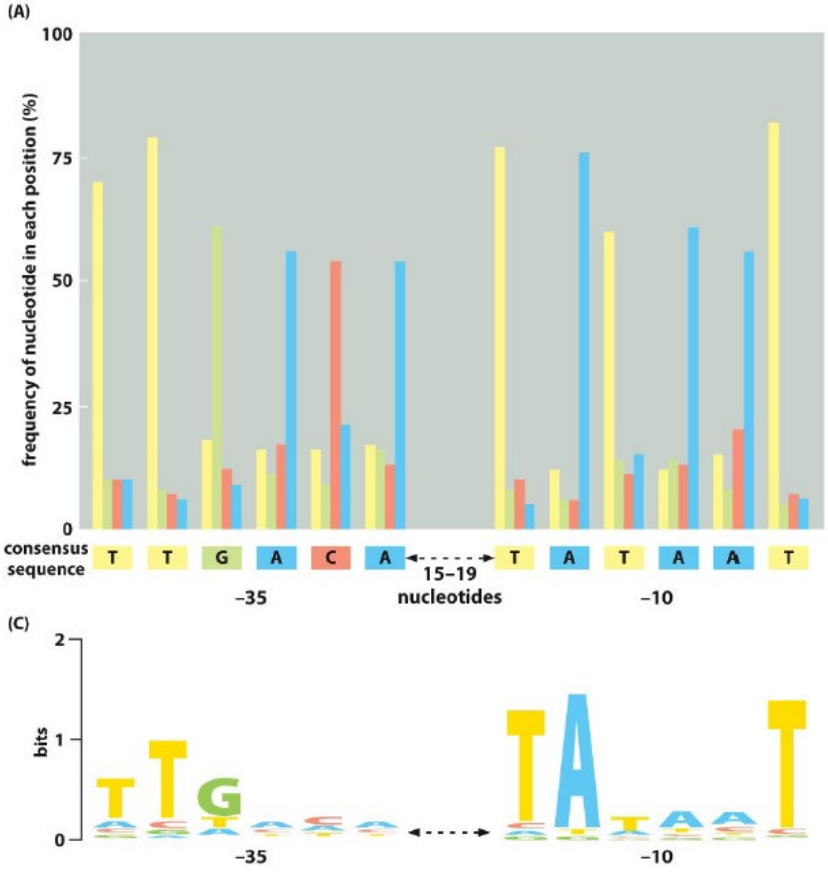
Another important parameter of a promoter is the distance between the -35 and -10 boxes, for the maximum affinity it has to be between 15 and 20 bp (peak at 17), over 20 bp the nucleotide sequence cannot be considered a promoter anymore.

The more a promoter deviates from the consensus and from the ideal distance between the two boxes, the more difficult will be for the RNA pol to be recruited.

Strength of a promoter

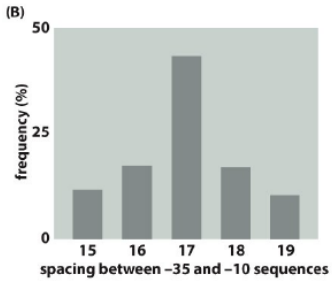
the strength of a promoter is a measure proportional to the rate of transcription of the gene, stronger is the promoter, tighter is the association of DNA with pol, higher is the transcription rate.

The optimal conditions are met very rarely, that's because cells do not need every promoter always fully active*, substitutions of certain nucleotides in the sequence yield a gradient of affinity between the DNA and the RNA polymerase (sequence readout).
The effect of specific nucleotide deviations from the ideal consensus sequence in specific positions can be computed statistically.



These diagrams show the most important characteristics that practically **define** any promoter, in -35 box the "core traits" are: TTG_ _ _; while in the -10 box the fundamental nucleotides are: TA_ _ _ T. The other positions are not strict and a substitution will affect the strength of the promoter without switching it off.

The same mechanism happens with the spacing between the two boxes:



The strongest promoter has 17 bp between -35 and -10 boxes

*These mechanism is the first step of regulation, in a system with no stimuli the cell still needs to express more certain products and less others, this can be achieved by using different promoters for each gene, their sequences are fixed (not dynamic) and can be changed just by mutations.

Regulation on sigma factors

The expression of genes can be regulated by regulating the σ factors.

Pro-sigma factors

Pro sigma factors are sigma factor themselves bound to a prodomain (cis), to reach their active form, the prodomain needs to be cleaved, by regulating the rate fo cleavage transcription can be regulated (post translational regulation).

Once the prodomain is cleaved the sigma factor can associate with the core enzyme.

Anti-sigma factors

Aniti sigma factors are proteins that bind to sigma (trans), and inhibit its function.

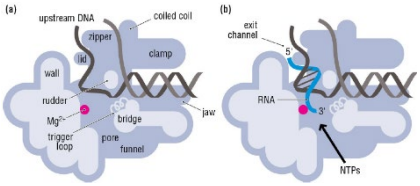
Ex. Flagellum

the early flagellum genes that code for the basal body that inserts in the cellular membrane and wall are expressed through σ^{70}, and when the basal body has been synthesised the late flagellum genes that use σ_f can be expressed.

the regulation mechanism is based on the anti-sigma factor **FlgM** and on the usage of the basal body as a pore: while the σ^{70} is working some anti-sigma factors are able to completely inhibit the action of σ_f; once the basal body is built it acts like a pore, FlgM passes through the pore, leaving the cell and dissociating form σ_f that is activated and starts coding for the late flagellum genes, allowing the flagellum to be put together.

Abortive initiation

1. **Close complex**: the polymerase with sigma factor is able to recognise a promoter and associating to it.
2. **Transcription bubble**: region of unwound DNA 14 bp wide, to open the double helix, in bacteria there is no need of ATP (in eukaryotes with pol II the transcription bubble is formed using an helicase and it requires energy); the **open complex** is formed
3. **Abortive initiation**: the polymerase tempts to produce full length RNAs but only manages to synthesise short stretches, 2-9 nucleotides, with no promoter clearance.



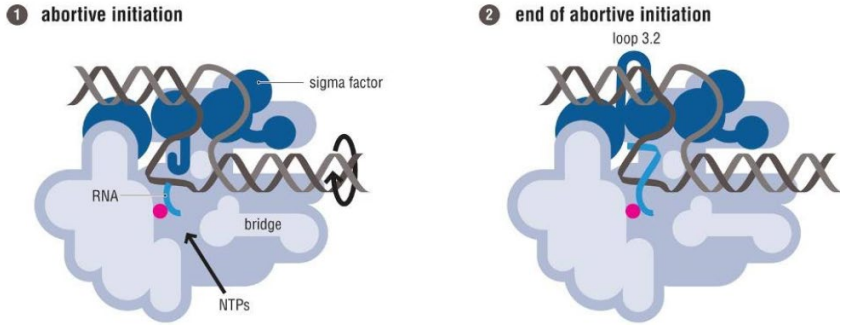
certain modifications to RNA polymerase can affect the processivity of the transcription.

addition of nucleotides occurs at 3' end (5'→3' synthesis) and it is biochemistry behind the binding is very similar to the one behind DNA nucleotide addition: nucleophilic attach on 3' end of the growing transcript, mediated by aspartate and magnesium ions.

the end of abortive initiation is achieved when the transcript gets long enough, the responsibility for abortive initiation is the sigma factor: the sigma factor has a loop that obstruct partially the exit of the RNA transcript from the reactive site; as long as this loop is not swung out abortive initiation persists.

In order to swing out the loop there must be enough nucleotides available and the rate of

transcription has to be fast and efficient enough to swing out the loop.

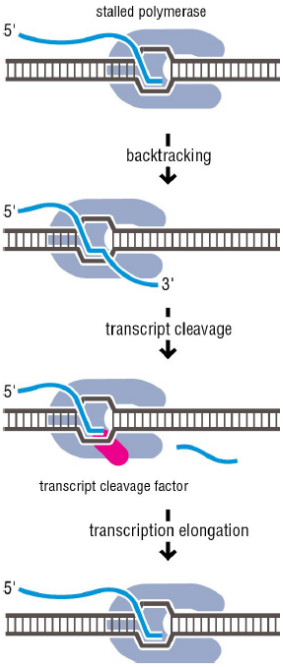


Elongation

The displacement of the loop leads to a conformational change that associates the polymerase with higher affinity to the DNA, the promoter clearance occurs and the elongation starts.
During elongation the sigma factor is lost, because it is not needed anymore (in eukaryotes RNA pol II is phosphorylated as it enters the transcription) and the transcription bubble moves along with the polymerase, the bubble if formed by a triplex (1 ssDNA filament and a duplex of RNA-DNA which is elongated) and proceed about 50 nt/s.

Errors

Polymerase sometimes makes errors, so there is an error correction activity:

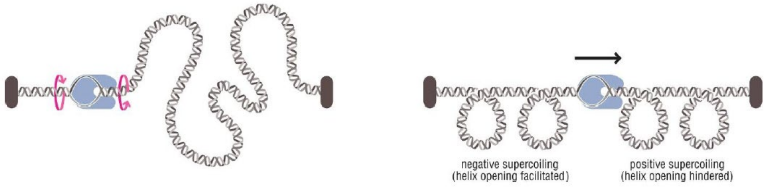


- the RNA polymerase can pause or slow down, this happens either because there are secondary struct in the RNA that may affect the stability of the polymerase, or because of other factors that can control the rate of elongation, or because the wrong nucleotide has been inserted in the sequence.
- When this happens the polymerase can backtrack a couple of nucleotides, now the 3' end that was contained in the transcription bubble protrudes out of the core of the enzyme.
- This stimulates the endonuclease activity of RNA polymerase, this activity can cut the protruding 3' end that contains the mismatched base.
- After the cut the 3'end is captured back in the transcription bubble and elongation restarts.

Supercoils

As transcription moves along DNA, that is generally constrained (circular or blocked by proteins), it generates tortional stress: positive supercoiling forms ahead of the polymerase while negative supercoiling forms behind.

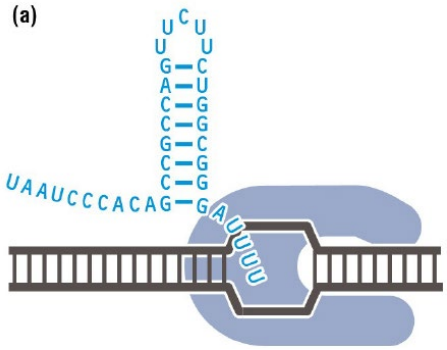
The supercoiling needs to be relieved, otherwise for the polymerase will be harder and harder to proceed; DNA gyrases are type 2 topoisomerases that cut double stranded DNA in both filaments and relieve positive supercoils while topoisomerase 1 cleaves one strand of a DNA filament and relieves negative supercoils.



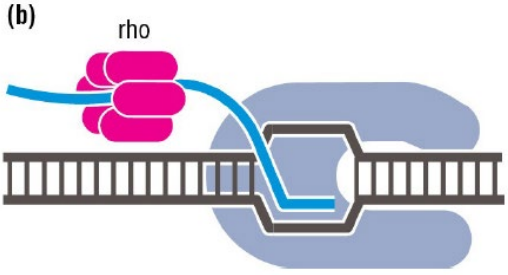
Termination

The sign for the termination of transcription is in the DNA and the trigger is the new synthesised RNA itself, in particular, in bacterial there are two types of terminators:

- **Intrinsic**: the DNA sequence contains the information for the formation of a **hairpin** followed by a **uracil** rich tract.
  To build this structure the DNA will have an **inverted repeat** that will be **GC rich** to allow the formation of a more stable stem-loop (more hydrogen bonds between the two strands) and **thymine** rich sequence; hence we are able to predict the termination of transcription just by looking at the DNA sequence.



  The formation of the secondary structure manages to be the signal of termination because it generates a **pulling force** on the RNA-DNA hybrid in the transcription bubble.
  The the hairpin that is forming leverages on the structure of the RNA polymerase and a traction force is produced on the RNA filament in the transcription bubble, if that RNA stretch is uracil rich the pair energy with the DNA (ΔG) it is not favourable with respect to the force expressed by the hairpin.
  Once the RNA is pulled out from the reactive core of the polymerase the transcription terminates.
- **Non-intrinsic Rho dependent**: depends on a termination factor called Rho which forms hexamers ring shaped.
  Rho hexamers are able to bind C-rich areas of RNA and produce a pulling force through ATP hydrolysis that separates the RNA-DNA duplex.
  The C-rich RNA wraps up on the ring-like structure formed by hexamers of rho and the ATP hydrolysis causes a conformational change in the structure of the ring that generates the pulling force (of course also a poly-U can help).



REGULATION OF TRANSCRIPTION

Regulating the transcription is necessary to respond to environmental changes.
Only symbionts are less interested in regulation because they live in a stable environment (usually these organisms have small genomes), in all other cases regulation is fundamental.

Imagine you could use monosaccharides and disaccharides to gain energy, you have to spend more energy to hydrolyse disaccharides so it is stupid to consume disaccharides if monosaccharides are present; so disaccharides enzymes have to be silenced in order to use less energy.

Regulation is also fundamental for differentiation and development, it works like a microprocessor, an environmental input enters in the regulatory machine and a transcriptional output is produced.

Activators and repressors

The first level of transcriptional regulation is the promoter strength, it is constitutive, meaning that it is constant in time.

The second and most important level of regulation is dynamic and it can happen in various phases:

- At transcript initiation (most prevalent)
- At elongation or termination
- Regulation from transcribed RNA itself

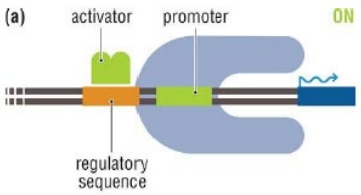
Activators and repressors are proteins that **up regulate** (boost or increase) or **down regulate** (turn off or stop) transcription.

Activators and repressor **bind to DNA** and their binding effects the transcription, the elements to which activator and repressor bind are called **regulatory elements** or sequences, in bacteria they are called **operator** or operator elements or operator sites.

These elements usually are close to the promoter, in this case they are called **cis elements**, although sometimes elements can be very distant from the gene they regulate, in fact **trans elements** can also lie on a different chromosome.

Cis elements

Activators: usually the regulatory sequence is part of the promoter, but not in the core promoter (-35 and -10); it lies some nucleotides **upstream**.
In the case of activating promoters a regulatory sequence is bound with specificity by a given activator, the binding promotes the recruitment of RNA polymerase by protein-protein interaction.
If a promoter is weak, an activator can help to recruit more the RNA polymerase

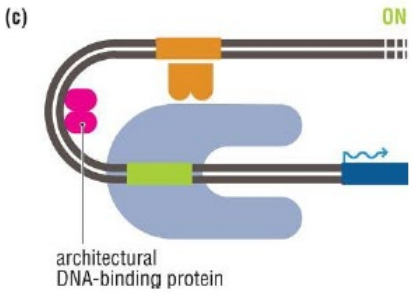


Repressors: frequently the binding site for the regulator is **within the core promoter**, the repressor will sit on the boxes, preventing the RNA polymerase to bind.

In general if the regulatory sequence is upstream (**40-200 BP**) likely it is an activator element, while if it overlaps the core promoter it usually binds a repressor.
In these cases the only thing to regulate is the connection between a stimulus and the conformational change of the activator or repressor that will allow the recognition of the sequence.

Trans elements

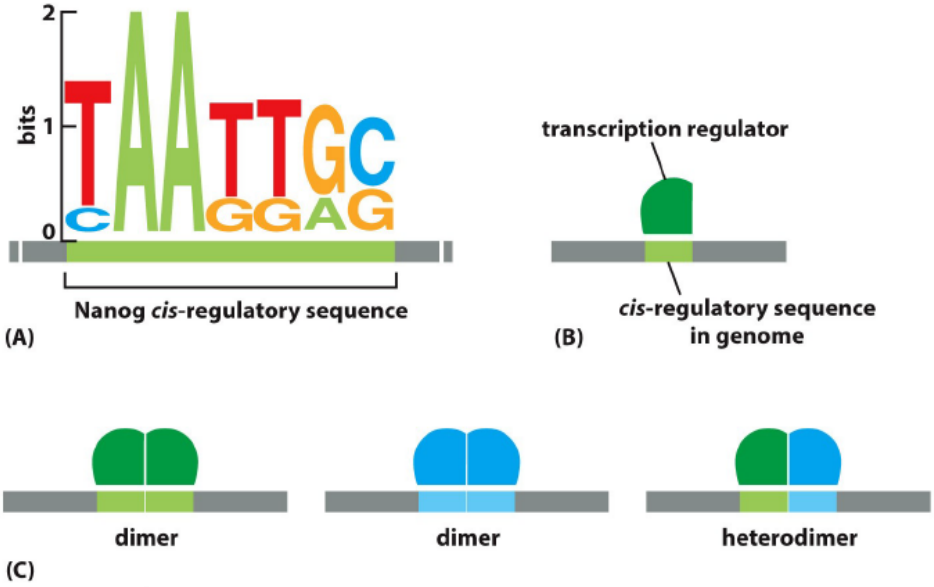
The regulatory element menages to be far away from the promoter because of DNA looping, when the element is bound to the regulator it can promote the recruitment of the polymerase by protein-protein interaction.
For example in bacteria NAPs induce looping and sequences that were far away can result to be in close contact.
Trans or distal regulatory elements are called enhancers and silencers, they can be either upstream or downstream of thousands of bases.



(In eukaryotes usually there are co-activators and co-repressors that act on the polymerase, they cannot bind DNA alone but they are recruited by specific regulatory proteins.)

Sequence readout in regulation

In order to specifically bind regulators to the promoter, we need a sequence specific DNA readout.



The distance from the consensus sequence defines the affinity of the regulator to the regulatory sequence, the consensus sequences are pretty short (their single repetition is pretty common by chance) because the sequence readout can only inspect few base pairs (major groove), so to increase the specificity usually **inverted repeats** are arranged next to each other.

In these cases the conservation of the consensus in not the only parameter, with more than 1 binding site, with an inverted repeat, the **regulator can dimerise** and **bind** to the sequence with **more specificity and affinity**.

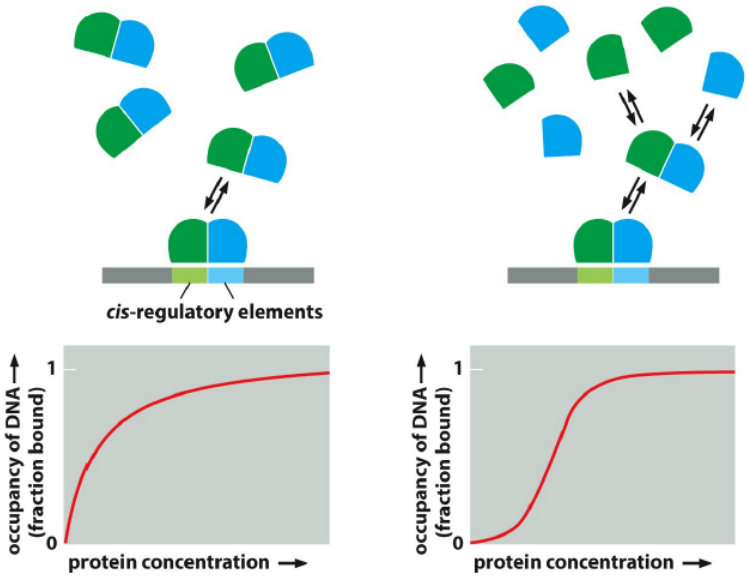
Specificity

The **specificity** is increased because an inverted repeat is less likely than a single consensus sequence; regulatory elements have a mutation rate that is higher than the coding sequences but lower than junk DNA, this is because there is an evolutionary pressure to conserve these sequences.

Affinity

The **affinity** is increased through 2 phenomena:

- **Cooperativity**: 2 weak interactions close by can become strong if there is protein-protein interaction between the two monomers, this phenomenon is known as **cooperativity** (deviation from the consensus is more tolerated).
  Now suppose there are 2 different regulator that can homodimerize, the dimerization also allows to combine different monomers together in heteromeric dimerization, in this way a completely new regulatory sequence can be regulated.
  Combinatorial effect: with 2 regulators we can recognise 3 regulatory elements, from which 3 different regulatory paths can start, this is a very energy saving mechanism.
  All or none effect: in cooperative binding the **second hemi** portion is bound with higher affinity if first is already bound, this means that in cooperative binding the all or none phenomenon is favoured, there is a certain threshold after which the kinetics of binding increases very fast and saturation is achieved (on-off switches).

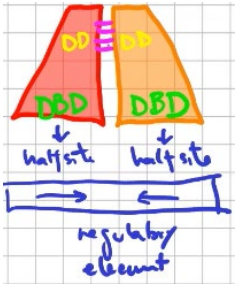


- **Allostery**: proteins have an active site and a regulatory site, the binding of a ligand to the regulatory site will activate the active site through a conformational change.
  In transcriptional regulation the active site is the DNA binding site, the conformations of the DNA binding motif changes upon the binding of the sigma molecule in the regulatory site of the protein.

Spacing between consensus

Another important characteristic of the sequence readout in regulators is the **spacing between the centre of the two consensus motifs**, each sequence is bound by a monomer of the dimer and it is separated from the other by 10/11 nucleotides, this is important because it is a **constraint** that allows to find regulatory sequences through bioinformatics analysis.

The 10-11 nucleotides (pitch of the helix) are the spacing needed for the helix to make a complete turn, and so expose the sequence to the same side of the DNA, allowing the binding of two adjacent enzymes.



All these factors regarding the binding of a regulator to a regulatory sequence can interact between each other, for example allostery can affect the ability of a regulator to read sequences that are 10 nucleotides apart: imagine a homodimer, each monomer has 2 domains a DNA binding domain (DBD) that will contact DNA by protein-DNA interaction and by sequence readout at its correspondent half-site (part of the regulatory element); each

monomer has also a dimerization domain is responsible for forming the dimer, allowing protein-protein interaction.
If we consider allostery, the two monomers can bind their ligand in the allosteric site and this may cause a conformational change in the dimer, it will be "bent" and it will be switch off.

We have a sort of on-off switch, according to the state of the transcriptional regulator, we define the regulator APO when it is not bound to the ligand and HOLO when it is bound to its ligand, the ligand can have positive or negative effect on binding affinity.

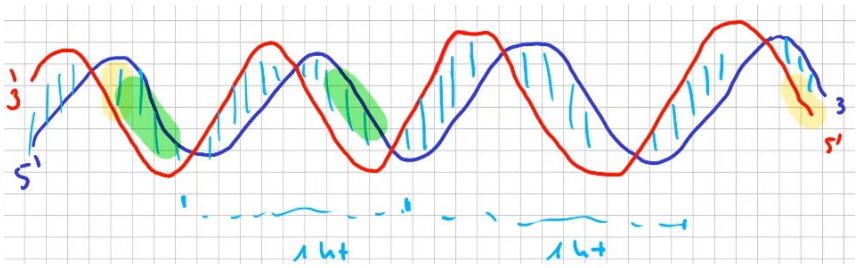
Sequence arrangement

Now let's reason on an hypothetic sequence to understand better how the nucleotide interact with a homodimer:

n n n G A C C A n n n n n T G G T C n n n
n n n C T G G T n n n n n A C C A G n n n

To write the inverted repeat of a sequence we have to consider also the polarity of DNA so we have to write the reverse complement of the sequence, in this case the consensus is: GACCA-n5-TGGTG.

the bases are arranged in this way because if two identical monomers are put one in from of the other they are specular, they will read DNA in opposite directions but they will read the same sequence, that's why we need the reverse complement of the sequence on the same strand.



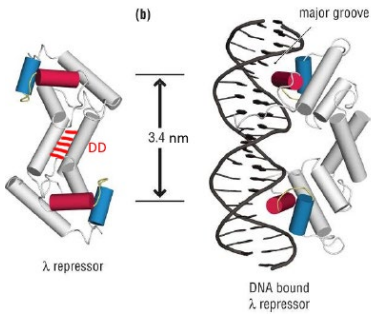
Imagine that two alpha helixes, each brought by a monomer of the regulator, bind to the green areas, they will have to be properly spaced to fit with two adjacent major grooves.
Now we can understand how, when a ligand binds to a regulator, the change of conformation can cause a loss or gain of the possibility to bind strongly with DNA.

Domains used to bind DNA

Few **conserved** domains are used in all forms of life to interact with DNA

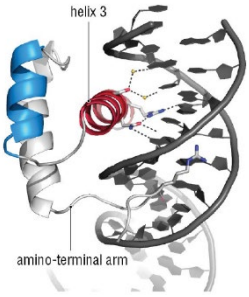
Helix-turn-helix (HTH)

Formed by 3 α helixes, one of them is the recognition helix, it is the one that is responsible for the sequence specific readout, it fits perfectly in the major groove.



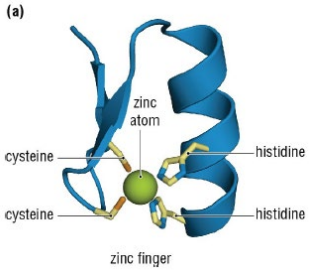
Examples:
Fage repressors: homodimers, composed of a dimerization domain and a HTH domain, the DD sets the 2 recognition helixes apart in space of 3.4 nm to perfectly fit two adjacent major grooves.
Cap: in detail later.
Homeodomain: found in eukaryotes, helix 1 and 2 interact between

each other while helix 3 enters the major groove, of course there can be different resedues in the third helix and these resedues are responsible for the sequence specific contact with the base pairs that are found in the major groove.
The readout is nothing but formation of hydrogen bonds, so hydrogen donor and acceptor groups or hydrophobic interactions affect the sequence recognition.
The N terminal arm (positively charged) of the homeodomain can make contact with the minor groove.



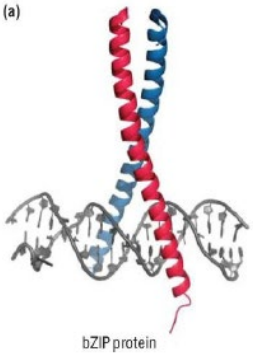
Zinc fingers

Small domains, constituted by two antiparallel β strands (mini β sheet) which carry two cysteine and then an α helix.
The structure of this domain is kept in place by a zync atom which is coordinated by the cysteines and 2 hisdidine of the alpha helix; depending on the resedues of the alpha helix, different sequence can be read out in the major groove.
One interesting thing is that there can be multiple zinc finger domains that wrap around DNA and zinc fingers can also be ingeneered by biotechnologists for specific seuqnce recognition.



Basic leucine zipper

Two coiled coils form a tweesers like structure, they bind simmetrically to DNA half sites, each of them is formed by a DNA binding domain and a **bZIP** region.
The bZIP region is leucine rich, leucine is an hydrophobic resedues so multiple leucine molecules like to interact between each other; in the DBD there are positive charged amino acids, the one that can interact better with the negative backbone of DNA, with specific readout sequences, while the leucine zipper domain is a **dimerisation domain**, it allows to keep together the two monomers by **hydrophobic interactions**, in this way also heterodimer can easily be formed.



Basic helix-loop-helix (bHLH)

Formed by coiled coils, 2 coils bind the major groove and they form a 4 helix bundle separated by a loop in between the DNA binding helix and the dimerization helix (the DNA binding strategy is the same as leucine zipper).

Beta readout

In some cases alpha helixes are not used in the recognition domain but beta sheets are used, for example in MetJ repressor in E.coli, it is a dimer and the DBD is formed by 2 beta strands.

Loop readout

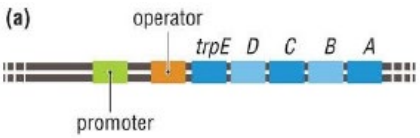
A more complex situation happens in a mammalian transcription factor, it is a homotetramer, in this case there are no alpha or beta strands that fit into DNA but the loops of the fold can fit into the DNA helix and read the nucleotide sequence.
This is not a very common mechanism because loops are not very stable, alpha helixes are much more stable and they can carry out a precise and reliable recognition, loops are generally less efficient.

Examples of regulation pathways in bacteria

Trp (tryptophan) repressor

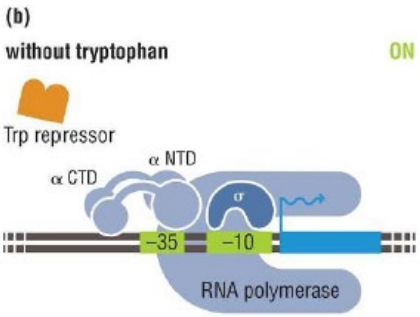
Trp is a great example of repression that works by promoter occlusion, Trp repressor is the regulator molecule that acts on the tryptophan synthesis pathway in bacteria.

Trp is a helix turn helix protein, it dimerises and in its functional form it is a homodimer; the operator (regulatory element) recognised by Trp repressor is downstream from the core promoter of Trp operon, the Trp operon is the set of genes needed for the synthetises of tryptophan (all transcribed and expressed together).

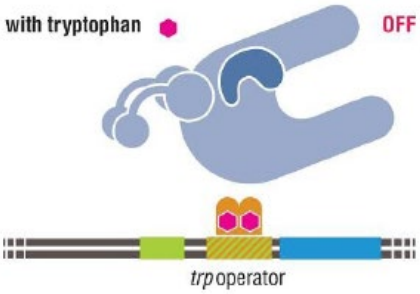


Just by reasoning about this pathway we can easily understand that tryptophan synthesis will have to be regulated based on the quantity of tryptophan in the cell, if there already is a sufficient amount of tryptophan there is no need to synthetise it.

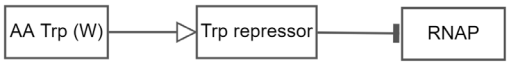
The regulation of this operon through Trp repressor will depend on bioavailability of tryptophan.
Tryptophan level high-> transcription off
Tryptophan level low-> on



This regulation is achieved through **allostery regulation of Trp repressor**: Trp repressor in **apo** form is not able to bind to DNA, therefore RNA polymerase can bind to promoter and preform transcription.



The **holo** form of Trp repressor consists in the binding of the repressor with the tryptophan (the ligand is the tryptophan), in this way a conformational change in the repressor occurs and the affinity of the repressor toward DNA is highly increased; Trp binds to the **Trp operator** that **overlaps the -10 element** of the core promoter of the operon, as a result the transcription of tryptophan is blocked.



In this case the W is also called co-repressor, because it is needed in order for the Trp repressor to be turned on (it can bind to DNA) and so inhibit the transcription.
The opposite type of regulation uses an inducer, a molecule that activates the repressor and by doing so it detaches the repressor from DNA.
In both cases we are talking about promoter occlusion by a repressor, the only thing that changes is the function of the holo repressor.

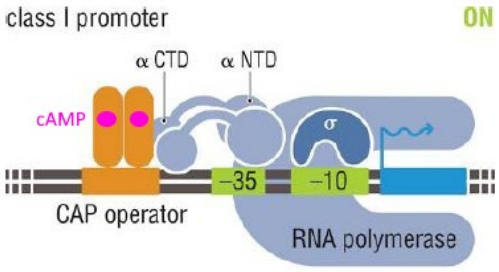
Cap activator

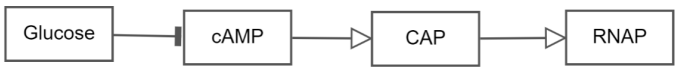
Cap is a catabolite activator protein, in order to be called activator it should not occlude the core promoter of the gene it regulates, therefore its regulatory element will be upstream from the promoter.

Cap is also called cAMP (cyclic adenosine mono phosphate) receptor protein or CRP, and it is able to bind cAMP.

Cap regulates positively 2 mechanisms:

- **Class I promoter**: Operators for CAP are found in a lot of genes in E. coli and these genes are activated when carbon sources are low, in class I promoter the regulatory element is upstream with respect to the core promoter.
  If glucose is low, then cAMP increases, the presence of cAMP is a signal of low easily metabolizable carbon sources. cAMP binds to CAP and it has a positive activity on the DNA binding activity of CAP, the binding of CAP enhances the recruitment of RNA polymerase by protein-protein interaction, it will make contact with the αCTD domain of RNA polymerase.





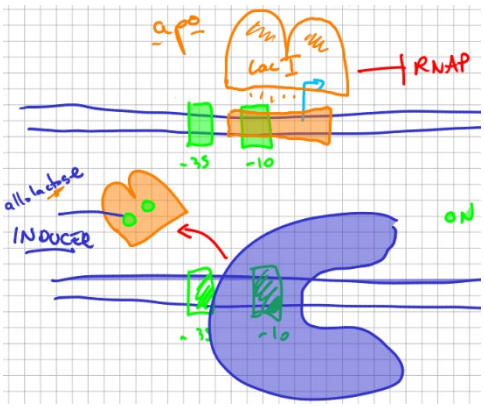
  The cAMP is defined as co-activator because it has a positive effect on the activator (it is needed to activate the activator); as in Trp repressor, also in this case it is the holo form that binds to DNA, the only difference is the region in which the repressor binds.
- **Class II promoter**: the Cap operator overlaps partially the -35 box, so it causes σ-4 domain (factor that recognises promoter sequences) to bind weakly to -35 element; however there are protein-protein contacts with other domains of the RNA polymerase that help out its recruitment onto the promoter.

The two promoter classes differ simply by the position of the operator element, if it is upstream of the core promoter, then the recruitment is achieved by protein-protein with the αCTD otherwise if it overlaps the -35 box the interactions are made with αNTD and with the remaining subunits.

LacI repressor

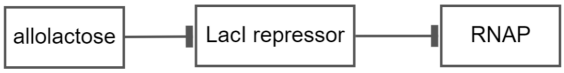
LacI works together with CAP, allolactose is the ligand of the repressor, in particular it is an inducer of LacI; LacI regulates the lac operon, the series of genes that allow to metabolise lactose.

More or less the lac operator overlaps the -10 box, so LacI works with promoter occlusion, until now it seems pretty similar to Trp, the only difference is that the binding affinity is the opposite of Trp with respect to the presence or not of the ligand (LacI uses an inducer, Trp a co-repressor).



LacI binds DNA in its apo form, active without the ligand, occluding the core promoter and inhibiting transcription; if allolactose is present LacI detaches from DNA and transcription is on, inactive form of the repressor.

In this case transcription is de-repressed.



Logic gate

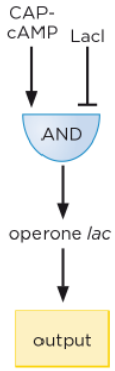
Since the LacI repressor and the CAP activator both work with sugars, they are in close relation: signal integration allows to combine different inputs and stimuli into one single output, this is achieved through complex promoters and regulation; the result is a biological microprocessor, a logic gate that can interpret a wide range of variables.

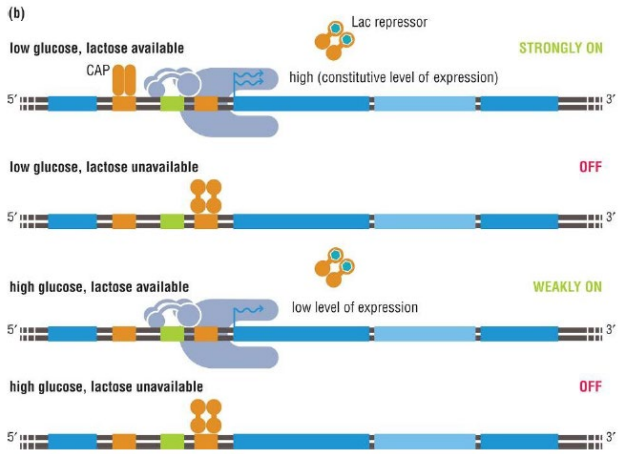
The lac operon comprehends a series of genes important for the metabolism of lactose, which is a disaccharide, formed by one molecule of glucose and one of galactose, it can be used as a source of energy but it requires a lot of energy to be hydrolysed so it is not the preferred source if a monosaccharide is available.

If the cell has both glucose and lactose it has to chose a metabolic pathway to gain energy that is different from the other combinations of the availability of the two nutrients.
To be able to take this wide range of decisions the CAP regulation and the LacI regulation interact between each other.

If the lac operon was regulated just through allolactose availability, when lactose and glucose are present the lactose would be metabolised, even if there is no need for that, so we have somehow to combine together the two regulation mechanisms, onto the lac operon; multiple regulatory elements are integrated on the same promoter.

The logic gate of CAP and LacI is the following:

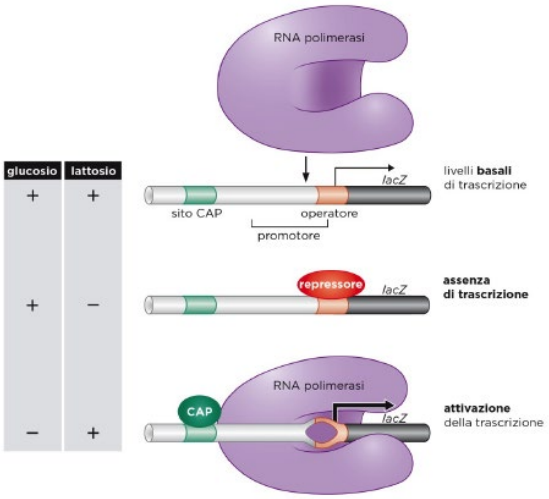




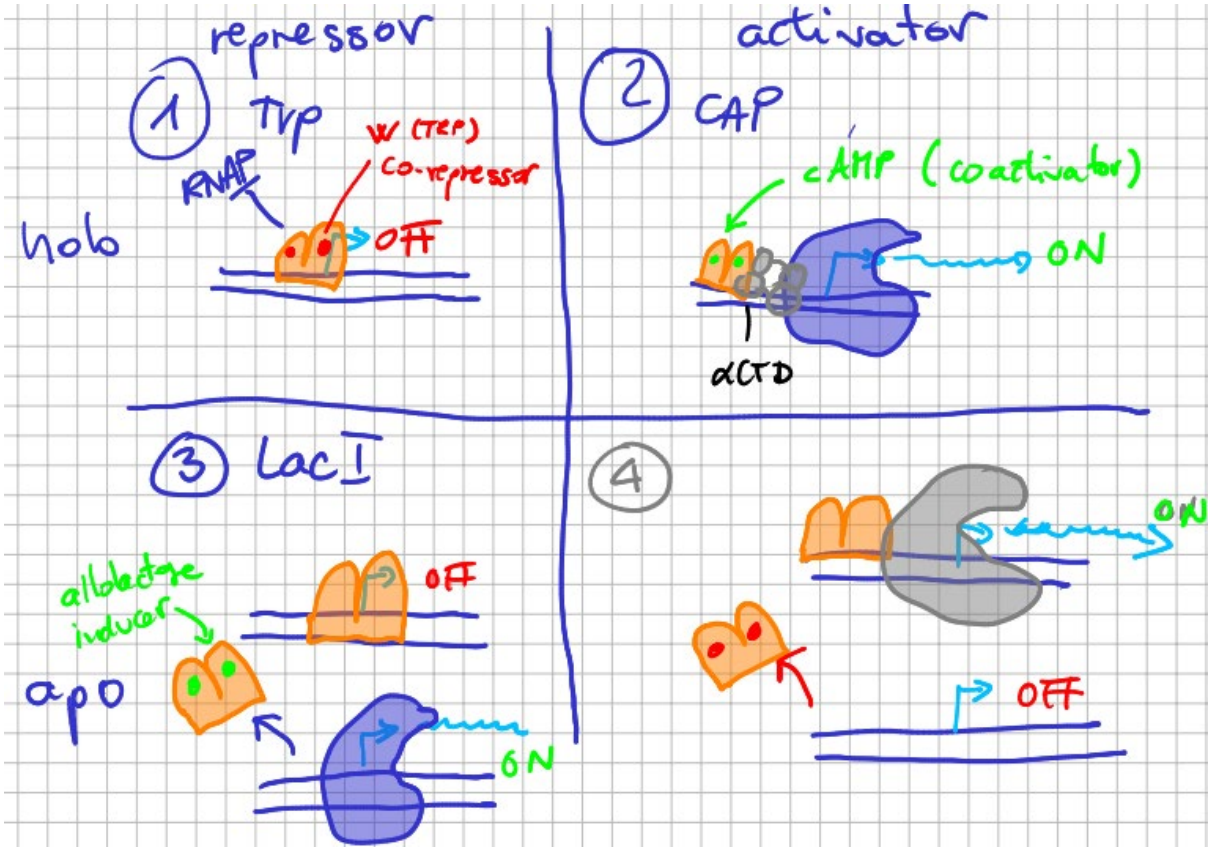
To activate this gate we need the active CAP (it's an activator) and inactive LacI (it's a repressor), if these conditions are met the transcription is on, full power.

Let's examine all the cases from a biological point of view:

- If lactose is not present the lac operon will always have to be off (no presence of allolactose, the repressor is bound to DNA→LacI on, AND condition not met).
- If glucose is low and lactose is present the lac operon must be on, at full power; we can see how in this case we have the situation we described before: low glucose means high cAMP and so the CAP activator is active, high lactose means high allolactose and so repressor detached from DNA
- If glucose is high and lactose is present the lac operator will be weakly on, that's because the LacI is off, it is bound to the inducer (presence of allolactose) but the promoter CAP is not bound to DNA, it does not boost the transcription!!!!

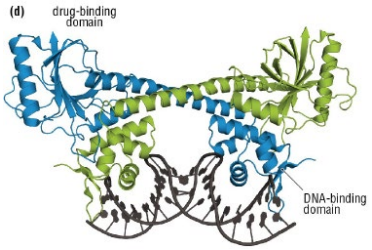


In this way we managed to obtain intermediate responses, these are fundamental for the equilibrium of the cell, in this case when there are both glucose and lactose a little bit of lactose can be metabolised, maybe just in case all the glucose is metabolised.



MerR activator

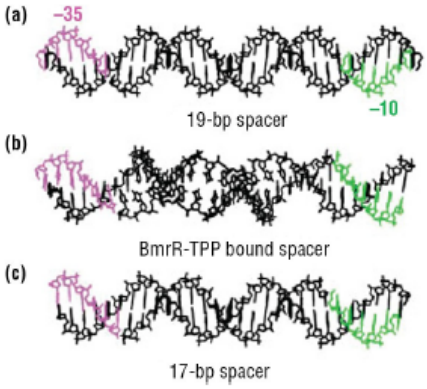
This activator does not work by pol recruitment, MerR is a family of regulator that influences the recognition of the -35 and -10 boxes more efficient, it changes the topology of DNA.



MerR is a dimer, each monomer has an alpha helix that recognises the sequences in the major groove, and it makes contact with the backbone in the minor groove.

The insertion of the dimer into the DNA helix distorts the helix, this distortion can be used to change the relative position of the -10 box with the -35 element.

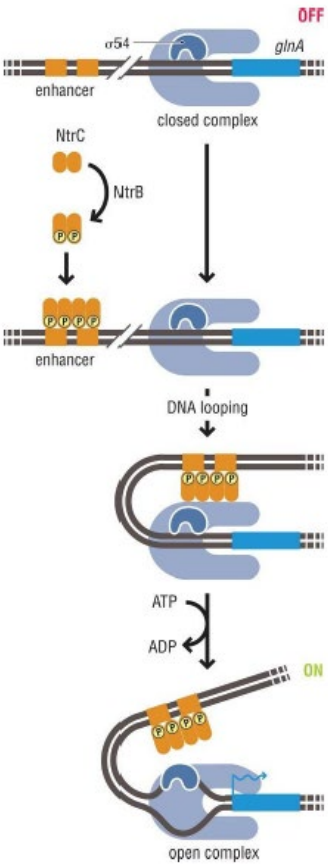
The distance between the two element is a very important variable for the recognition, the most efficient distance is 17 bp, in this way the two boxes are placed perfectly for the recognition caried out by σ4 and σ2 domains of sigma.



At MerR regulated promoters the spacing of -10 and -35 is suboptimal, they are usually separated by 19 bp, this means not only that they are further away but also that they are slightly shifted with respect to the side of the double helix.
When MerR binds the DNA gets distorted, this distortion provokes a topological rearrangement of -10 and -35 one with respect to the other and the spacing between them becomes exactly equal to 17 bp, the promoter now is stronger.

NtrC enhancer

NtrC is part of the nitrogen metabolism, it regulates the RNA polymerase that uses **σ54** (an alternative sigma factor) and it binds at enhancer elements.



NtrC is not usually able to bind DNA, when a proper stimulus arrives NtrC is phosphorylated and this phosphorylation allows it to bind the enhancer element; the binding site of NtrC to far from the promoter of the gene to directly regulate it.

Meanwhile at the promoter of the targeted gene the σ54 is waiting at the close complex stage, it is waiting the binding with the phosphorylated NrtC because it needs ATP hydrolysis to drive promoter opening.

Through **DNA looping** NtrC is put in proximity with the close complex and protein-protein interactions of the polymerase with NtrC stimulate ATP hydrolysis and the ATP hydrolysis is used to form the open complex and then transcription starts.

araC

araC is another example of DNA looping, it has a regulatory region that affects 2 divergent promoters, one is responsible for the expression of the regulator itself and the opposite promoter is responsible for the transcription of the operon that codes for the genes that are regulated by the same regulator that is transcribed in opposite direction.
This is because if a regulator is needed in certain contidions also the regulation of this regulator will be governed by the same conditions.

Very frequently the genes of the regulators are either within the regulated operon or close by and they can also be autoregulated.

Transcriptional attenuation (co-translational regulation)

In bacteria the transcription can be controlled by the rate of translation of the transcript itself.
This mechanism is possible because in bacteria there is no physical separation between transcription and translation, in bacteria as soon as the 3' end of a transcript is available the translation starts, translation and transcription occur simultaneously.
The regulation of transcription through translation is called transcriptional attenuation.

Trp operon

Again we will go through the Trp regulation, but in this case we are not considering transcriptional repression, we are considering the effects of translation on the transcription (transcriptional attenuation).

This mechanism involves **translation** and the formation of alternative **secondary structures in RNA**.
The Trp operon RNA has a **leader sequence**, the leader sequence is found at the start of the transcript and usually it is translated.

Frequently leader sequences are cut off from the final peptide because they are not used for the function of the protein but they confer some desired features, for example some leader sequences are responsible for the secretion of a protein or for targeting a protein to the nucleus or other organelles and then they are cleaved off.
The only role of the leader sequences is to be translated.

In this case the precise function of the leader sequence of Trp operon is that of modulating the speed of translation, it is able to slow down the translational speed.
Inside the Trp leader sequence we find **2 codons** one after the other for **tryptophan**, W is a very rare amino acid, so finding two codons in a row is an oddity; therefore, usually the availability of tryptophan amino acids and tryptophanyl tRNA is pretty low.

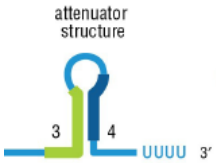
The function of the attenuator is governed by these two codons.

After the leading sequence the RNA transcript of the trp operon has 4 different regions that can form alternative stem loops, in particular we can have 2 situations:

- Pairing of region 2 with 3:

A harmless stem loop structure is formed, transcription continues without any problem.

- Pairing of region 3 with 4:

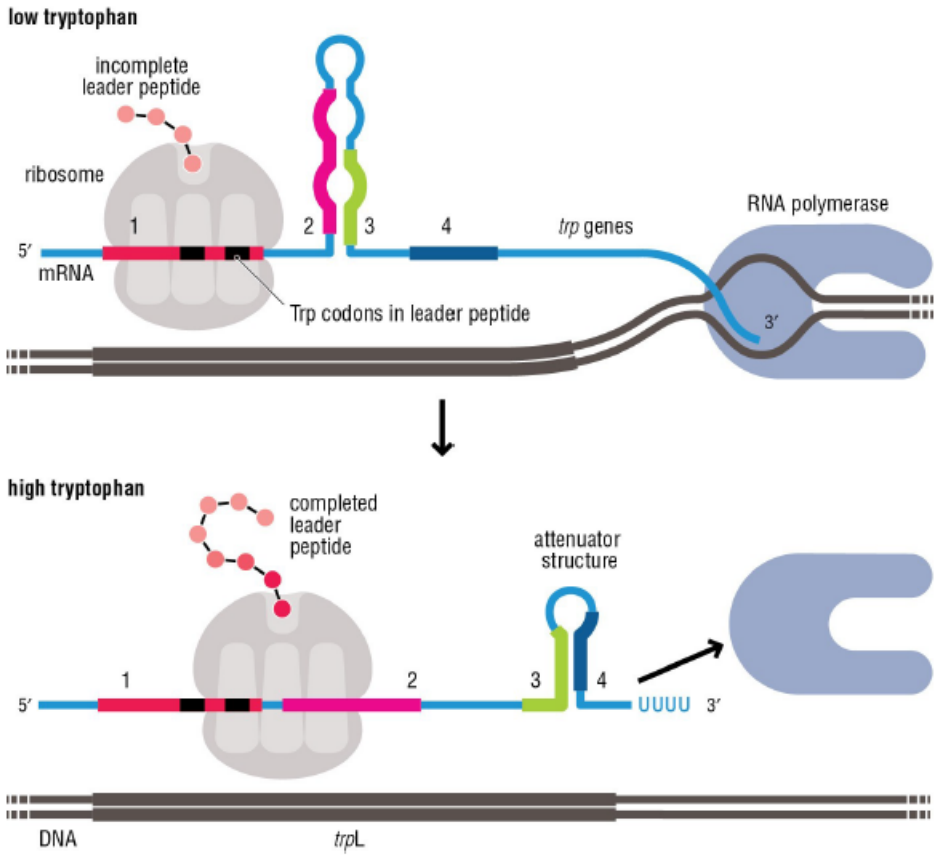


the stem loop formed corresponds to an intrinsic terminator, if this structure is formed it will apply a leverage on the RNA polymerase and transcription will be attenuated.

The transcription and translation of the Trp operon causes the synthesis of the enzymes needed for the tryptophan synthesis so they are needed when there are low W levels; in low tryptophan transcription of Trp genes is on while in high tryptophan transcription of Trp operon should be low or off.

The rarity of tryptophan also affects the availability of tryptophanyl tRNA (W tRNA), the tRNA charged with W.
If the **tryptophan levels are high** it is very easy to form W tRNA, so it will be easy to incorporate W in the growing polypeptide chain, therefore, even if two tryptophan amino acids in a row are needed the ribosome will have no problem in finding them and translation will move fast across the leader sequence.
If translations proceeds fast in its first phases the hairpin between region 2 and 3 is not formed and therefore the only possible structure that can be formed is the intrinsic terminator (this is happening while RNA pol is still transcribing RNA, all of this happens at the beginning of the transcription); once the intrinsic terminator is formed RNA polymerase falls off from DNA because of the leverage that is applied on the RNA-DNA duplex in the active site of the polymerase.

On the other hand if **tryptophan levels are low** this means that there will not be much W tRNA available so having 2 W codons in a row could be a problem, translation will require more time to find a second W, so the ribosome pauses and this allows to region 2 to pair region 3 to form the 2-3 hairpin and preventing the 3-4 stem to form.



Signal transduction

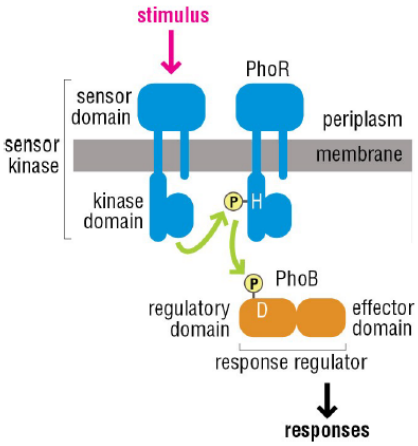
Process trough which a signal or a stimulus is transduced into a response.

If the signal is a membrane permeant molecule there is no problem, the molecule can enter the cell and bind the regulator (ex. Metal ions).

Frequently signals cannot go through the membrane, it is needed either a transporter or other system able to transduce the signal across the membrane and then start a series of events triggering or repressing transcription.

In eukaryotes there are very complex pathways, a very simple example common to eukaryotes and bacteria is a phosphorylation mechanism, called 2 component system.
A 2 component system is formed by a response regulator which is the transcriptional regulator inside the cell and the sensor kinase in the membrane.

Response regulator itself is composed of 2 domains, a regulatory domain that gets the signal and an effector domain which is a DNA binding domain; the affinity to DNA of the effector domain is changed when the regulatory domain is phosphorylated.

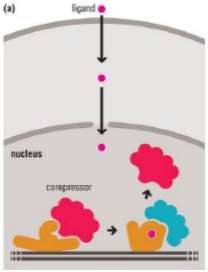


The sensor kinase is a membrane associated protein, composed by 2 domains, a sensor domain, usually extracellular, and a kinase domain, intracellular; the sensor domain is responsible for perceiving the signal, if the signal is sensed a conformational change occurs in the kinase domain and it auto phosphorylates, by auto phosphorylation the kinase domain becomes active and it can phosphorylate the regulatory domain of the response regulator.

With this system the signal does not have to enter the cell, it is received outside the cell and it is transduced by a 2 component domain.

Eukaryotic signalling

In eukaryotes there is an additional barrier for the signal to achieve a transcriptional regulation, it needs to also cross the nuclear membrane, because transcription occurs in the nucleus.



In the simplest possible case the ligand is able to cross the membrane (permeant) and enter the nucleus through the nuclear pores; when it is inside the nucleus it can bind to **nuclear receptor proteins** that respond specifically to the ligand, changing their conformation and then recruiting or not the co-repressor or the co-activator.

The majority of the ligands are not able to cross the nuclear membrane and once they are in the cytoplasm they provoke some conformational change in the cytosolic nuclear receptor protein and this change allows the translocation inside the nucleus of the nuclear receptor protein.

A lot of regulation processes in eukaryotes occur because there is an inactive regulator in the cytoplasm then a signal is sensed, a conformational change occurs and the regulator becomes active and this simply allows the regulator to enter the nucleus.

NF-kB

Important for mammalian immune responses; a signal binds to a receptor that can sense pathogens or dangerous conditions.
The binding activates a cascade of phosphorylation events that amplifies the signal, at the end the phosphorylation provokes the ubiquitination of the inhibitor protein that was keeping a regulator in the cytoplasm.
The regulator is free and it is able to translocate into the nucleus to regulate transcription.

Transcription motives

All these motives are elements where different regulators are hooked up one with the other in logic circuits that can generate many different outputs, the discipline that studies this is systems biology.

All these interactions and circuits are based on few simple rules that allow to develop complexity, the single elements are very simple.

In the case of transcription regulation the systems are robust and they can resist noise to output reliable and consistent outputs.

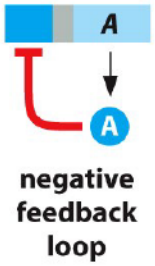
Positive feedback loop

A regulator auto regulates positively itself, the more regulator there is the more regulator is produced.

The result is the explosion of the transcription of a set of genes, this system is useful to go from one state to the next, it switches from state A to a state B, with no way back.

These circuits generate de-stability.

Negative feedback loop

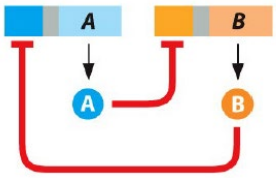


negative feedback loop

In a negative feedback the presence of a regulator inhibits its transcription, while its absence allows the transcription to occur.

As a result, when there is no regulator the regulator is produced, while when the regulator is present it is repressed, in time we have a oscillatory kinetic, better described as homeostasis, it is needed to the cell in order to respond a stimulus and then go back to its equilibrium state; cells live in a range of conditions, so equilibrium is fundamental.

Flip-flop devices



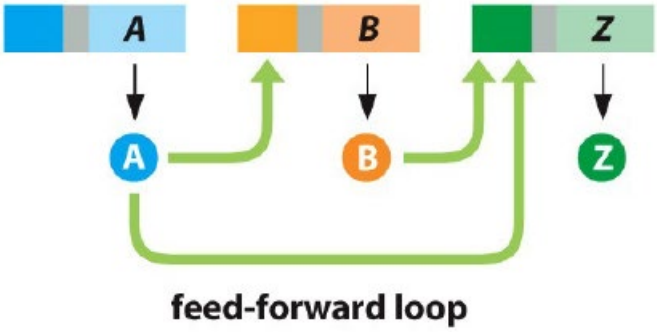
flip-flop device
(indirect positive feedback loop)

In this case if A is present B is not produced.

If a signal inhibits A, B is produced and inhibits A, taking over in the circuit, even when the signal will go away the system will stay in state B.

Again we have de-stability.

Feed forward loop

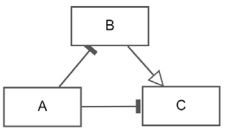


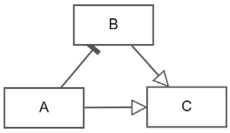
feed-forward loop

3 nodes motives, the first and the second usually are regulators and the third one is the output.

All these circuits have a direct way and an indirect way, the direct way goes from the first regulator to the output while the indirect passes through the second regulator.

Upon these rule, 8 possible circuits can be built, 4 coherent and 4 incoherent; in a coherent circuit the outcome of the direct way is equal to the indirect way, while in the incoherent they are different.

Ex: coherent: 

uncoherent: 

Coherent and incoherent can be used either to generate pulses or to buffer noise.

Incoherent: even if signal persists in an incoherent path at a certain point the expression of the second regulator B will affect the initial regulation of A on C because they have opposite effects. In this way a constant signal is traduced to an impulse.

Coherent: if a signal is weak and flickering the circuit will not respond to it, it buffers noise, responding only to full and nice signals

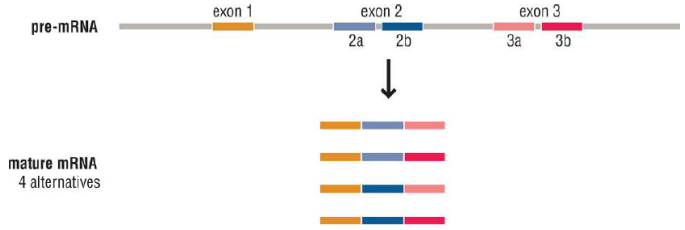
RNA PROCESSING

After transcription the RNA usually it is not ready to perform its functions, it may be processed in many different ways; we will go through the processing of a pre-mature RNA to a mature RNA.

- **Cleavage**: the RNA is cleaved in order to form the mature RNA.
- **Splicing**: a particular type of cleavage is splicing, it consist in the removal of a RNA sequence that is not going to be translated, called **intron**, the remaining sequences, called **exons**, are joined with each other to form the mature RNA.
- **5' capping**: is a modification of the 5' end that consists in the addition of a methyl guanosine cap, a structure that is not conformed to 5'→3' direction rules; this modification is important for stabilising the eukaryotic mRNA.
  Moreover it will be very important for eukaryotes in order to make their mRNA translatable, it will provide the molecular determinant for the assembly of the translation initiation factors.
- **Poly adenylation**: consists in the appending at the 3'end of a poly A tail, this tail is not genetically encoded (it doesn't come from a DNA sequence) but it is added by an enzyme, also polyadenylation is important for stabilisation and for RNA translation.
- **Editing**:
  - Base modification (ex. from an adenine to an inosine)
  - Base insertion and deletions

Processing is important because:

1. It generates diversity, for example in the alternative splicing:
   it is an advanced version of splicing, it splices together different exons it can chose which exons keep to form many different isoforms of a gene.
   In this way many protein isoforms (with different function) can derive from the same gene (it is very useful for complex organisms).



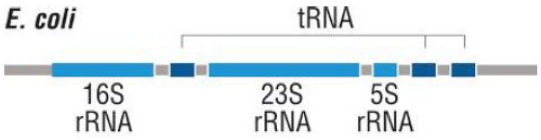
2. It regulates gene expression, for example the 5'cap and the poly A tail stabilise the messenger RNA and increase translation rate.
3. It Acts as quality control, detecting defective RNAs and degrading them.

All these actions are post-transcriptional.

Modifications involved in processing involve a lot do different complexes, localised in different sites of the cell, usually these processes are primed and triggered by **ribonucleoproteins (RNPs)** they are factors that contain both proteins and RNA.

Sometimes these complexes contain guide RNA and other non-coding RNA whose role is to provide sequence specific recognition of the target RNA, guiding the protein effector that modifies the sequence.

Cleavage



Example: in E. coli the ribosomal rRNA genes are organised in a cluster, rDNA codes for rRNA genes; these rRNA genes provide the structural backbone of the ribosome, in particular the small and large subunit.
These genes are interspaced with non-coding DNA portions and also there is a tRNA locus, transcribed together with rRNA in a long precursor, the transcription of this sequence generates a pre-RNA.

Having the genes in a single operon is very useful because In every ribosome we always need the same amount of parts for the ribosome so it is convenient to produce them together.

The initial transcript needs to be processed into the proper parts, this is achieved by ribonucleases that mature and cleave RNA.

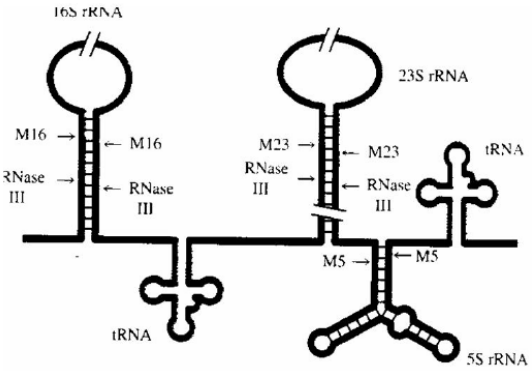
There exist a lot of different ribonuclease:

- **Exonucleases**: eat up the RNA and usually they start from 3' end, going to 5', they are non-sequence specific
- **Endonucleases**: can be specific or can be guided to a specific point through a guide RNA. They cleave RNA within the strand, they are divided into 2 groups:
    - ssRNA cleave: RNAse III
    - dsRNA cleave: RNAse E, RNAse P

RNAse III

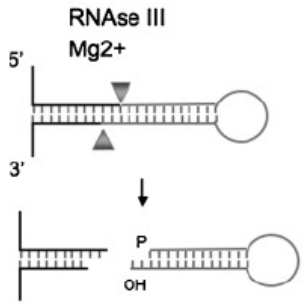
The maturation of the long precursor of E.coli rRNA genes starts with RNAse III, it is a conserved enzyme throughout all living organisms that cuts double stranded RNA; in eukaryotes it is involved in the maturation of micro RNAs and non-coding regulatory RNAs.

How come RNAse III is important?
The long precursor folds up in a peculiar secondary structure and the ribosomal genes will be found in the loops of different stem loop structure.



The processing by RNAse III happens on the stem and allow to keep only the loops, the rRNA genes are cut out from the longer precursor.

Another important aspect of RNAse III (true for all RNAse III nucleases) is that it has very peculiar signature, after the processing it leaves a 2 nucleotide overhang at the 3' end.
This overhang can be used by other proteins that recognise the product of the cleavage; this is important to further process the transcript (micro RNA and interference RNA).



RNAse P

Another endonuclease is involved in the processing of this particular transcript, RNAse P, it cuts ssRNA, it has an RNA component itself.

Notice that the bacterial RNAse P enzyme , formed of protein and RNA is able to cut RNA with just the RNA component, this is because of the 2'OH group that confers catalytic properties to RNA (the

protein part just enhances the efficiency); in eukaryotes RNAse P RNA component cannot cut RNA alone.

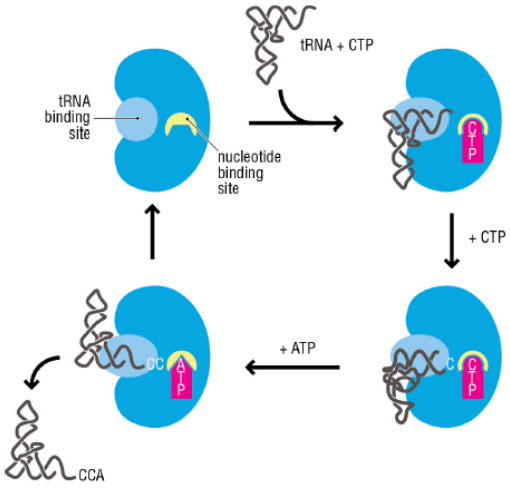
Finally, sometimes, the processing, generation and separation of rRNA from tRNA (same precursor) is mediated by splicing.
Splicing can also be promoted just by the RNA itself, this phenomenon is called self-splicing and the parts of RNA that get removed are called self-splicing introns; because RNA is a reactive molecule, some introns can adopt a peculiar structure that allows self-processing (frequently found in bacteria, in eukaryotes there are also proteins).

RNAse E cleaves ss RNA and it can be guided by proteins (in details later).

tRNA processing

once the tRNA transcript has been isolated and it has folded, it needs the addition of a CCA motif at 3' end.

tRNAs have a conserved CCA sequence at 3' end, it is the attachment site for the amino acid which is covalently linked to it (we don't want this very important molecule to be wrongly formed).
Sometimes CCA is genetically encoded, more often it is added later (processing mechanism) by a **CCA adding enzyme**.



This enzyme has a **tRNA binding site** and a **nucleotide binding site**, the addition is catalysed in the nucleotide binding site: tRNA fits in its binding site, then the modification of 3' end can occur with the attachment of 2 **CTP** in the nucleotide binding site, resulting in the binding of 2 consecutive cytosines and finally adenine is added in the same region, using ATP.

So the nucleotide binding pocket can adopt 3 different conformations, this controls weather C or A is added.

Sometimes the CCA adding enzyme can also degrade non-well-formed tRNAs, it targets them for degradation by adding an additional CCA tail (CCACCA) that will stimulate the degradation of the molecule.

Modification of bases

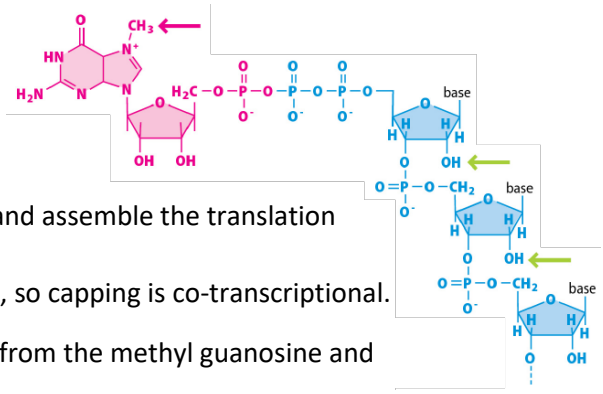
tRNAs has also more than 80 modification biochemicals, they generate diversity, they change shape or struct or functionality of tRNA, some of them are huge like the addition of an amino acid to a base or small like methylation.

Some of These modifications are crucial, for example they change the shape of tRNA, in fact a hyper modified base is present close by the anticodon loop in tRNA and it promotes the extrusion of anticodon loop.

Another important modification is pseudo-uridination, that makes the RNA molecule more stable and less prone to be degraded.

Capping

5' end capping involves the eukaryotic messenger RNA, after the synthesis of pre-mRNA the 5' end is 3 phosphate, so we can link to it another modified nucleotide, a **methyl guanosine base**, Through 5'-5' bond; sometimes also the second and the third bases of mRNA are methylated at 2' carbon.

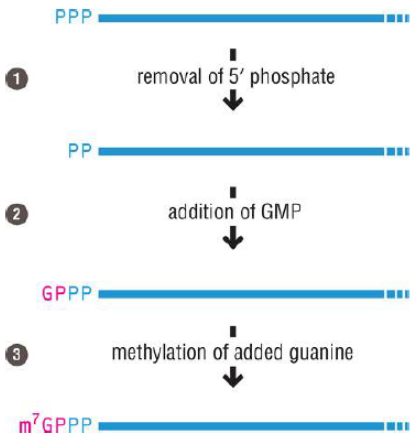


Capping has a lot of functions:

- Prevents RNA to be degraded, 5’ to 3’.
- Provides a tag for translation machinery to find the start and assemble the translation initiation factors.
- Important for Elongation and termination of transcription, so capping is co-transcriptional.

At the end the 5’-5’ bond is formed by 3 phosphates, one coming from the methyl guanosine and two coming from the transcript.

The capping process starts as soon as 20/30 nucleotides of mRNA emerge from RNA polymerase II reaction centre, note that this process doesn’t happen in pol I and pol III, therefore rRNA and tRNA do not have the cap.



The cap is added in 3 successive steps.

1. The first nucleotide is 3 phosphate, in order to bind the cap we need to remove one of the phosphates.
2. Guanyl transferase enzyme adds a GMP (guanosine monophosphate) with 5’ to 5’ direction, forming GPPP bond
3. the guanine base is methylated.

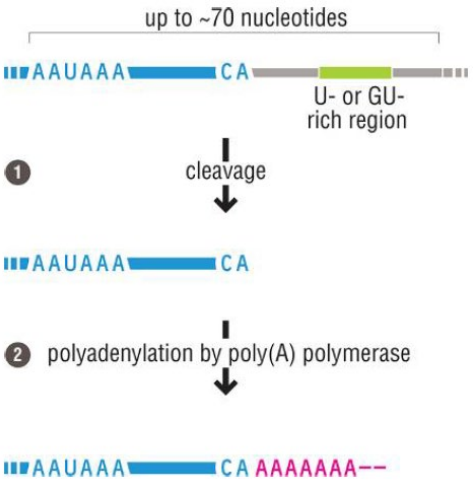
(For this is responsible a tail of pol II)

Polyadenylation

Addition of a poly A tail at 3’ end, it stabilises the 3’ end and it is important because RNAses usually degrade 3’ to 5’.

The adenine sequence is not genetically encoded, it is not transcribed, it is added afterwards.

The transcript contains polyadenylation signs, they indicate where the pre messenger RNA has to be cleaved and where the poly A tail has to be added.



Usually there is a conserved small region in the transcript that is constituted by a hexamer: **AAUAAA**.
Following downstream there is a **U or GU rich region**. In between we have **CA nucleotides**.

This indicates that cleavage has to occur at CA region, once the transcript is cleaved poly-A polymerase adds one after the other 200 adenosines (it is not a template synthesis).

All messenger RNAs generated by RNA pol 2 are capped and poly adenylated, this enzymes governs and couples the two processes.
Exception: histone mRNAs are not polyadenylated, they are however stabilised by stem loop structures at 3’ end.

Poly a tail is also important because it important because permits circularisation of RNA during translation.

Coupling

Capping and poly-A are coupled in transcription to RNA pol II activity

Pol II has a CTD domain (different from α-ctd in bacteria) that can be phosphorylated, it is responsible form mediating mRNA processing.

The phosphorylation of CTD governs both processivity of pol II, capping and poly adenylation.

When pol II starts elongation, leaving the promoter, CTD starts to be increasingly phosphorylated, the more phosphorylated CTD the more the transcript is elongated, the less CTD is phosphorylated, the closer pol II is to the promoter.

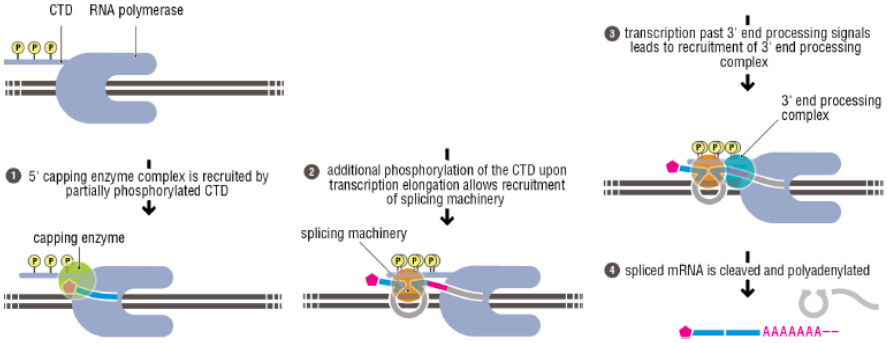
As soon as pol II enters elongation CTD is partially phosphorylated, when this happens the partial phosphorylation recruits the capping enzyme onto the polymerase; it is created a Docking site for the capping enzyme.

As soon as transcript is generated and exits the reactive centre of the enzyme (20-30 nucleotides long) it already finds the capping enzyme sitting there and capping is promoted.

As polymerase moves along the CTD gets even more phosphorylated, higher levels of phosphorylation promote recruitment of other factors, for example, splicing factors; so the transcript is co-transcriptionally spliced.

The highly phosphorylated CTD tail is also a signal for the recruitment of the 3' end processing complex, which is nothing but the poly adenylation complex, the presence of AAUAAA followed by a uracil rich tract allows polyadenylation to occur.

Both capping, splicing and the cleavage responsible of the poly adenylation are co transcriptional (The poly A tail is added afterwards).



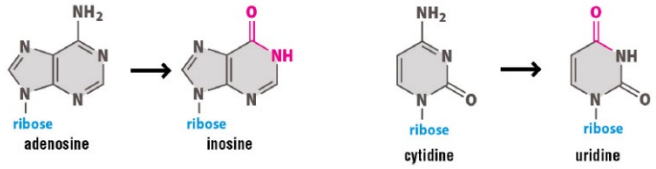
RNA editing

Deamination

A common RNA modification is the deamination of adenosine or cytidine.
Adenosine base can be deaminated into **inosine** (inosine pairs with cytosine)
Cytidine can be deaminated into **uridine** (U pairs with A or with a wobble with G)

These modifications are important because by changing the coding region, the final protein sequence changes; by changing a base on the mRNA it is possible to change a codon that will recruit a different tRNA.

These modifications are important and may not happen by chance, but with a functional role.



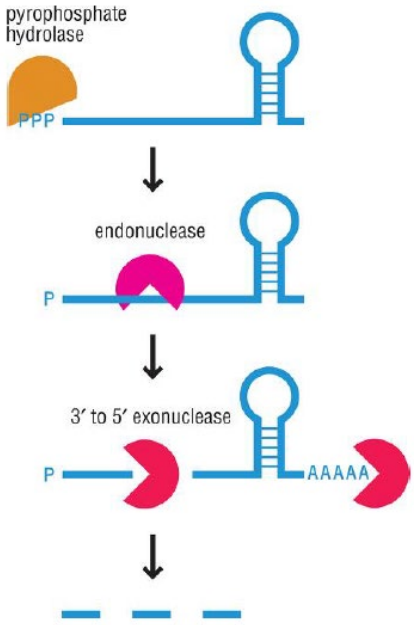
RNA decay

It is very important because:

- If we have defective RNA we have to destroy it, not translating it, its protein product could be deleterious.
- The translation machinery could remain stuck on a damaged mRNA.
- Decay pathway is used as post transcriptional regulation, it is a very fast regulation, if we have a RNA product ready it is very easy to cleave it.

Bacteria

Bacterial degradation is initiated by an endonuclease, **RNAse E,** that is able to cleave the transcript, if the transcript is cleaved the exonucleases can degrade it.



The first cleave is necessary because often the 3'end is protected by a secondary structure that is able to fence of the exonucleases, if RNAse E cleaves the mRNA degraded by exonucleases and by addition of poly a tail at 3' end.

*Poly a tail in bacteria is added to RNAs that need to be destroyed (*opposite to eukaryotes).

RNAse E does not degrade intact RNA, that's because intact RNAs have a signature, they have a 3 phosphate group at the 5' end, when a RNA is cleaved for the first time it remains either with 0 or 1 phosphate group; so, to place a cut, RNAse E checks for the presence of the 3 phosphate and it does not cut if there is an intact 5' end.

For the first cut one or two phosphates have to be removed and then the transcript can be recognised by RNAse E and it can be cleaved; the phosphate removal at 5' end is promoted by the **pyrophosphate hydrolase** that jumps in action when there are signs that a transcript has to be degraded.

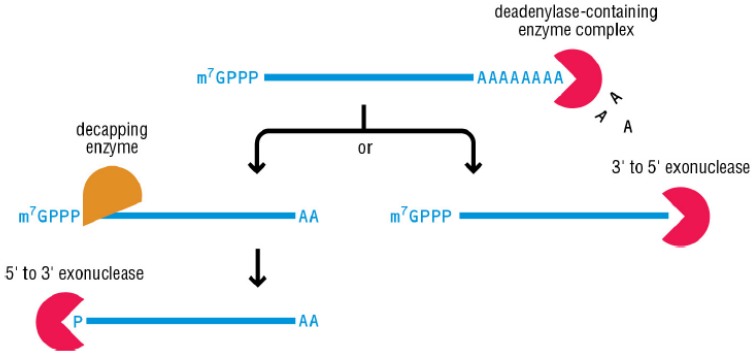
Another way to adopt RNAse E is to use a small RNA that can drive the enzyme to a spot by sequence specific recognition of the target, and then RNAse E can cut and promote the decay of mRNA.

RNAse E is part of the bacterial degradosome complex, it is a complex of proteins that is responsible for the RNA decay in bacteria.

Eukaryotes

The process is similar to prokaryotes, it uses different players.

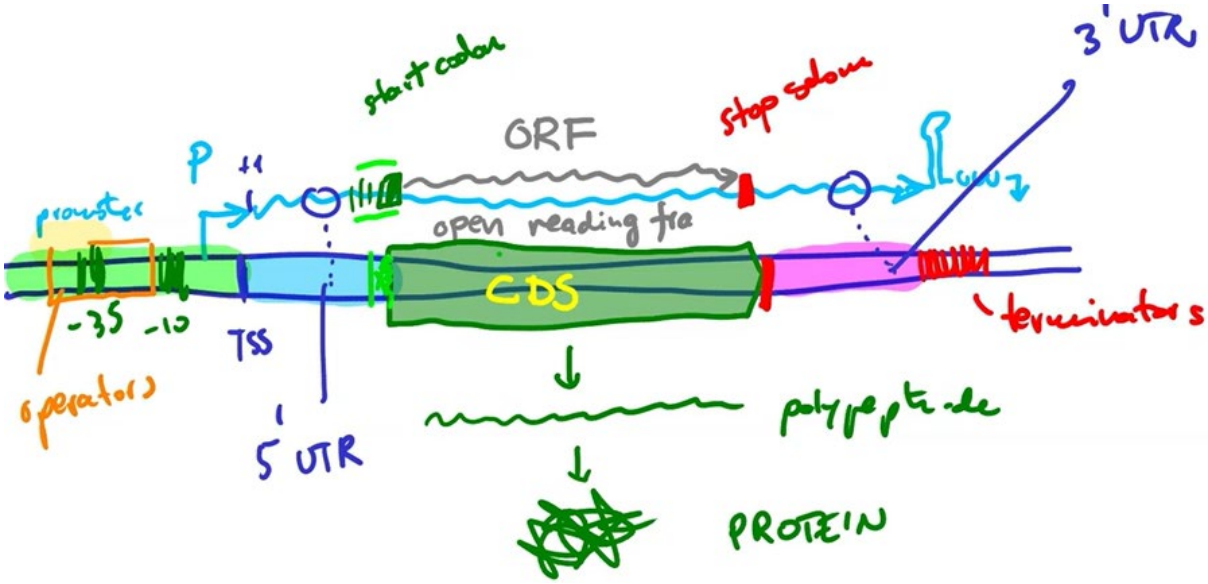
If 3' and 5' are protected the first step is to remove the poly A tail, then the transcript can be digested 3' to 5'; while this happens a de-capping enzyme is recruited and then 5' to 3' exonucleases can eat up the RNA in the opposite direction XRN1.



TRANSLATION

As for transcription, in translation RNA has signs that tell the translation machinery where to start translation and where to stop it.

Recall on the sequence gene that codes for a protein:



If we refer to the RNA sequence that codes for the protein we call it ORF, or open reading frame, while if we map the sequence onto DNA we call it about CDS, or coding sequence; the only difference between these two sequences is just that in DNA instead of uracil there will be thymine (the direction is always the one of the coding sequence, 5'→3', and we just map it on the DNA in the same verse).

Process

Translation is the production of a polypeptide from a mRNA.

Proteins are made of 20 amino acids so the information in the mRNA must be translated into a correct amino acid to a specific adaptor molecule: the **transfer RNA** (tRNA).

The tRNA has the responsibility to link 2 different worlds and perfectly match them, it can recognise and pair by complementarity the codons in the mRNA, and it is able to match the correct amino acid (cognate) onto the correct tRNA (at least 20 amino acids).

This link is super important because from this depends the whole translation process; the binding is made possible thanks to aminoacyl-tRNA synthetases, these are very important enzymes that have a proofreading activity that match the right cognate amino acid to the right tRNA, meaning the tRNA that has the anticodon that matches the codon that synthetises for the same amino acid.

During translation the amino acids are put il the right sequence and they are linked one to the other.

The machinery that links together the amino acids is the ribosome, it is a huge complex, made of RNA and proteins, it has 2 subunits the small subunit and the large subunit, they are both important and they have different roles but in both cases their function is carried out by RNA.

- **Small subunit**: responsible for the proper decoding, it grants the accuracy of the decoding.

- **Large subunit**: responsible for making the covalent bonds between one and the next amino acid, it forms the polypeptide chain.

During the translation the ribosome moves along the RNA in 5' to 3' at **15 amino acids per seconds**, note that each amino acid is codified by 3 nucleotides, this means that the ribosome speed is about **50 nucleotides per second**; it has a quite high error rate, one amino acid out of a thousand is wrong.

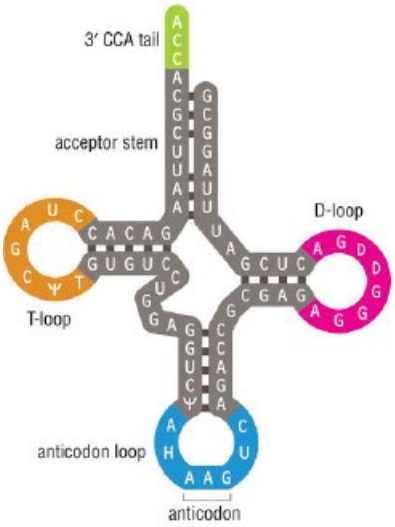
The ribosome is **assisted** by a series of protein factors called **translation factors**, they are called GTPases, there is a lot of energy consumption in the form of ATP and GTP to ADP and GDP, translation is the most energy consuming process of a cell.

Translation has 4 phases: initiation, elongation, termination and **ribosome recycling**

tRNA

tRNA is probably one of the ancient molecules we deal with.

It must be bifunctional, it links the amino acids with the nucleotide information.



Lucky enough these components are carried at opposite ends of the tRNA molecule; if we consider the tertiary structure amino acids are covalently linked to the CCA tail (it is a 3' stem) at one end while the three nucleotides of the anticodons are carried in the anticodon loop.

Each tRNA is formed by 75-95 nucleotides and each tRNA is specific to certain amino acids.

When a tRNA is loaded with an amino acid it is called **aminoacyl tRNA**.

Modifications

The tRNA is heavily modified, the modified bases have a very important function such has conferring a specific tertiary structure or reacting with specific molecules.

They are situated in the other 2 structures of tRNA, the **D loop** and the **T loop**:
D loop is called so after the **dihydrouridine**, one of the modifications found in it.
T loop has a **pseudo uridine** and it is called T because of a **ribothymidine**.

To remember the relative positionment of the structures fo tRNA we can think that the CCA tail is on 3' end, and that the **T**ail is on the same side of **T** loop.

Another important modification is placed immediately **after the anticodon**, it is symbolised by H and it stands for a **hypermodified purine**, it is needed and conserved for two reasons:

- It prevents the wrong base pairing of the fourth position with the codon of mRNA, it molecularly define the anticodon and prevents the paring of an additional nucleotide with the RNA
- It imposes the conformational change that flips out the bases of the anticodon, usually in loops the bases do not tend to stick out to the water environment because they are hydrophobic so this hypermodified purine imposes a conformational change that flips out

the bases of the anticodon so they become freely available to make contact with the codon, forming the minihelix.

Genetic code

The genetic code is degenerate, we have 20 natural amino acids but since they are codified by triplets there can be 64 possible combinations, therefore more than 1 codon codes for 1 amino acid (every possible codon is used).

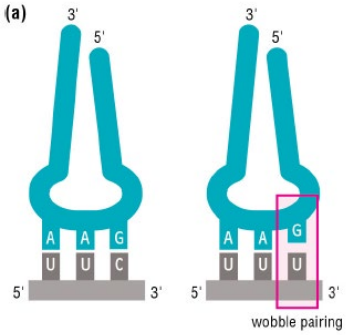
Among all the codons we must have something that tells us where the start is and where the stop is: there is one start codon, AUG, and 3 stop codons, UAA, UAG, UGA.

GCA GCC GCG GCU	AGA AGG CGA CGC CGG CGU	GAC GAU	AAC AAU	UGC UGU	GAA GAG	CAA CAG	GGA GGC GGG GGU	CAC CAU	AUA AUC AUU	UUA UUG CUA CUC CUG CUU	AAA AAG	AUG	UUC UUU	CCA CCC CCG CCU	AGC AGU UCA UCC UCG UCU	ACA ACC ACG ACU	UGG	UAC UAU	GUA GUC GUG GUU	UAA UAG UGA
Ala	Arg	Asp	Asn	Cys	Glu	Gln	Gly	His	Ile	Leu	Lys	Met	Phe	Pro	Ser	Thr	Trp	Tyr	Val	stop
A	R	D	N	C	E	Q	G	H	I	L	K	M	F	P	S	T	W	Y	V	

We have combinations of degeneracy 4 and combinations of degeneracy 2; some cases have degeneracy 6 but they are just the combination of a set of degeneracy 4 and a set of degeneracy 2.

Apart from methionine there is just one other amino acid that is codified by just one codon, which is UGG that codes for Trp.

- **Degeneracy 2:** the last nucleotide changes, only purines or pyrimidines are interchangeable between each other in this set; this means that each set can be identified as Y (pyrimidine) or R (purine).
  The reason that explains this phenomenon is **wobbling**:
  The wobble pair is a particular base pair that is non-Watson crick that forms between **uracil and guanine,** it is formed by 2 hydrogen bonds and the energy is weaker than a AU pair because the plane of the two bases is shifted a little bit.
  Wobbling is allowed only in the third position of a codon, if we have a U on the RNA sequence we can pair it with A or G, thanks to the wobbling, as you can see A and G are purines so we have just described de behaviour of a R codon.
  If we have a G in the RNA it can pair with C or U, through wobble or Watson crick.
  In the third position wobble pair is allowed and this wobble explains the degeneracy 2 of the genetic code.



(b)

1st base of anticodon	3rd base of codon
C	G
A	U
U	A or G
G	U or C
I	U, C or A

- **Degeneracy 4:** the last nucleotide changes, in all its possibilities, this is caused because sometimes **inosine (I)** is present in the anticodon, it extends the possibility to **pair with more nucleotides**

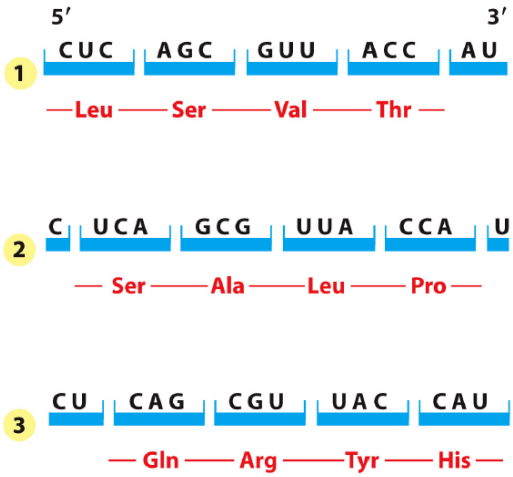
If we take these factors into consideration and if we consider that degeneracy 6 is only degeneracy 4 and 2 together (these amino acids are just coded by more combinations, maybe also very different between each other), we can understand that a single tRNA cannot recognise all the codons for a specific amino acid; in degeneracy 6 we surely need different tRNA because very different codons have to be recognised.

tRNA that carry the same amino acids are called iso-acceptors.

In total around 40 different tRNAs are needed for the 61 codons (wobbling inosine and modifications allow this arrangement).

Reading frames

The fact that the ribosome proceeds sequentially on the sequence and that it needs to read out codon per codon properly means that we have 3 different frames in the RNA sequence.



If we read a sequence 3 nucleotides by 3 nucleotides, we can start either from the first or the second or from the third, if we start from the fourth we would be again in the first frame.

Therefore one very important thing ribosome has to do is to maintain the right frame, frame maintenance means that it has to move 3 nucleotides by 3 nucleotides, if by chance it reads 4 nucleotides it falls off of frame and therefore the sequence will end up completely different.

To find the right frame we can look at the punctuation (the start signal).

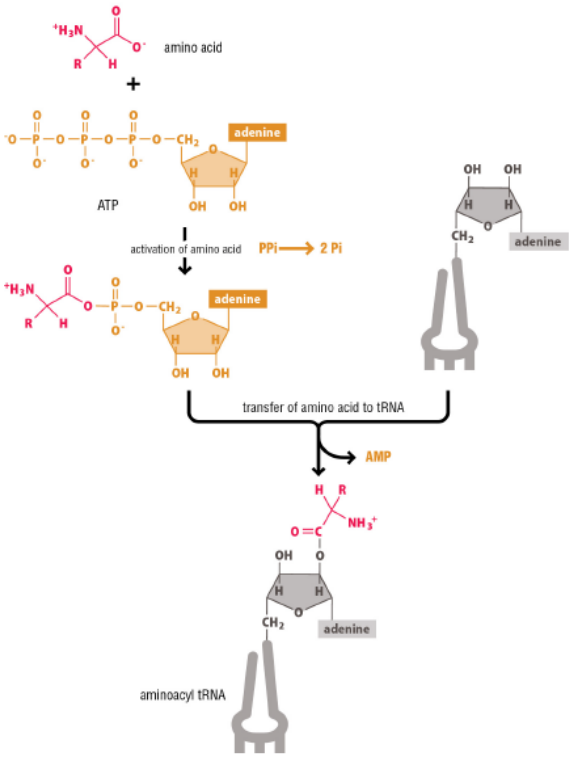
Codon usage

Some codons are used more infrequently than others, these are termed as rare codons, they tend to be decoded by rare tRNAs.

Codon usage is species-specific, the genetic code is almost the same is all organisms (with some exception) and in addition evolution has conserved codons so that single nucleotide mutations that change the encoded amino acid, result in a similar amino acid taking its place.

Aminoacyl-tRNA

Aminoacylation attaches amino acids to tRNAs, this is mediated by aminoacyl-tRNA transferases, in a two step process requiring ATP.



In the first step the amino acid is activated by attachment of AMP, this releases pyrophosphate and provides energy; the activated aminoacyl-adenylate remains attached to the enzyme. Secondly the enzyme transfers the amino acid to the tRNA 3' CCA tail, detaching AMP.

Each amino acid has its own aminoacyl-tRNA synthetase, the specific amino acid with which a tRNA is loaded is indicated with a three letter superscript such as $tRNA^{met}$ or with the notation aaRS; the correct amino acid for a tRNA is referred to as **cognate**.

Loading is an accurate process fewer than one error per 10000 aminoacylation events.

Aminoacyl-tRNA synthetases recognise tRNAs by sequence and structural feature, called **identity elements** that are checked during the two step process of loading.

Most aminoacyl-tRNA synthetases have an **aminoacylation site** and an **editing site**.

The **size exclusion** keeps non-cognate amino acids that are too big **out of the amino-acylation site** and the editing site can accommodate the activated amino acid (editing pre-transfer) or the amino acid after attachment to the tRNA (editing post-transfer).

If an **amino acid is rejected** pre-transfer the aminoacyl-adenylate (activated amino acid) is **hydrolysed**, while if rejection occurs post- transfer the amino acid is **cleaved from the tRNA**.

tRNA synthetases classes

There exist **two classes** of aa-tRNA synthetases, class I and II, each with about 10 members.

- Class I usually recognise the minor groove of the acceptor stem, class II the major groove.
- Class I attach amino acids to the 2' OH of the terminal ribose, class II to the 3' OH.

The protein structures of the two classes are very different but, apart from some exceptions, class I and class II synthetases typically recognise the same groups of amino acids.

Transamidases and desulfurases

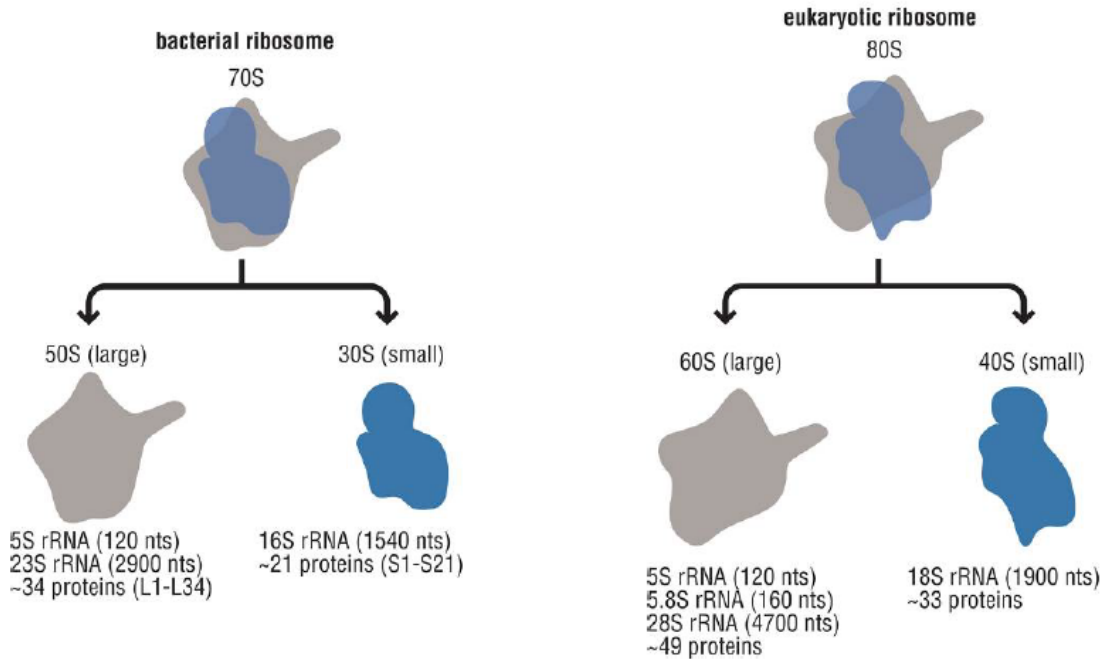
Some bacteria and archaea have fewer than 20 synthetases, those for attaching glutamine and asparagine are usually the ones missing.
In these cases, the aspartate and glutamate synthetases (AspRS and GluRS) have dual specificity; a transamidase reaction changes the side chain of the attached amino acid from an acid to an amide, producing asparagine and glutamine bound tRNAs

In other cases, CysRS is can be missing, here, a phosphorylated serine is loaded, and cysteine desulfurase converts the attached serine to cysteine.

Ribosome

Ribosomes catalyse the formation of peptide bonds between amino acids, they are large complexes, 2.5 MDa to 4 MDa, 2/3 of which is ribosomal RNA (rRNA) and 1/3 protein.

Ribosomes have a large and a small subunit, each containing ribosomal proteins (rproteins) and RNA; the small subunit mediates interactions between mRNA and tRNA while the large subunit catalyses bond formation and has an exit tunnel through which the growing polypeptide emerges.



Eukaryotic and bacterial ribosomes are generally conserved, but differ in their composition.

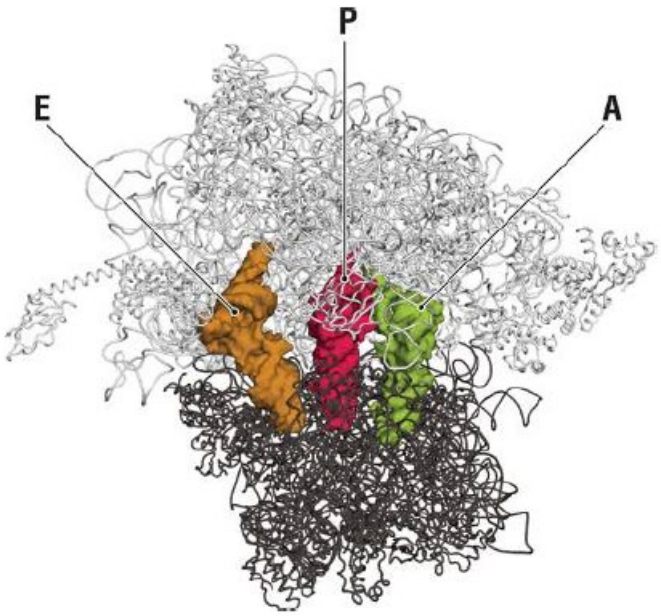
The interface between the subunits is important for movement of tRNAs and mRNA and it is rich in rRNA; while, within a subunit the proteins extends arms into the RNA regions, these arms are usually highly basic and are thought to help with packing the negatively charged rRNA phosphate backbones.

Additional protein and RNA layers are found with increasing organism complexity, the functions of these additional components are not clearly understood.

Ribosomal RNAs and proteins are extremely highly conserved across species, this suggests that the RNA components of ribosomes were present before the protein components, and that proteins were recruited to the ribosomes later in evolution together with DNA (RNA-world origin of life).

A-P-E
Each ribosome has 3 sites, between the small and large subunit:

- A site: aminoacyl site
- P site: peptidyl site
- E site: exit site



These 3 sites are fundamental for the function of the ribosome and they are occupied by tRNAs, the tails of the tRNA kiss each other (they are bound to the amino acids) and the anticodon loop points toward the small subunit, where the mRNA sits.

At the A site we have the proper decoding of the codon anticodon pairs, new tRNAs are inserted in this site.
Once an aminoacyl tRNA is inserted some changes associated with GTPase activity happened and allow the amino acids in site A and P to get closer.
In the peptidyl site the peptide bond is formed.
The exhaust of tRNA happens in the E site, after it has been separated from its amino acid.
In active elongation all the 3 sites are occupied and they shift.

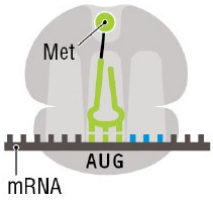
Phases

- **Initiation**: governed by initiation factors (IF), if there is a lowercase “e” we refer to eukaryotic factors (eIF)
- **Elongation**: governed by elongation factors (EF)
- **Termination**: governed by release factors (RF)
- **Ribosome recycling**: governed by recycling factors (RRF), it is important to achieve high levels of translation, each ribosome is a very bulky structure, the cell invested a lot of energy to produce them so it is necessary to reuse them

Initiation

The two subunits of the ribosome are produced separately and at initiation they need to be assemble at the site where translation starts.

The starting site of translation is identified by a signature in the RNA sequence, one very important element is the start codon; generally it is **AUG** (bacteria could also have GUG).



The initiation factors have to cast the two ribosomal units onto the translation start site, firstly they will load the **methionyl tRNA**, and the **small subunit** and after this happens the **large subunit** joins.

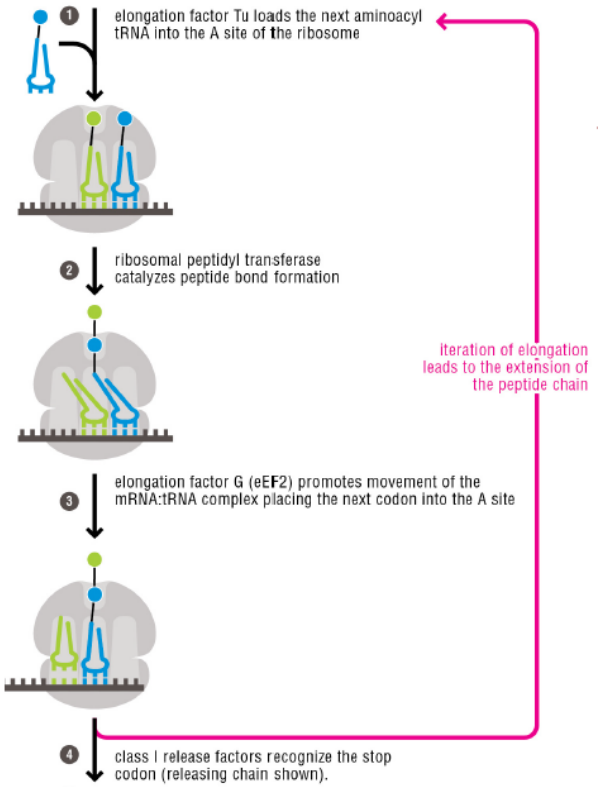
Notice that, since the entry tRNAs must enter the A site, the A site must be left unoccupied.

The initiator methionyl tRNA needs to be loaded in the P site, this leaves the complex in the right conformation to accept new entry tRNA in A site.

When the complex is set the ribosome can start its movement.

Elongation

An aminoacyl tRNA is added into the A site, this is promoted by **elongation factor**, in bacteria it is called EF-Tu while in eukaryotes eEF1A, these elongation factors have the responsibility to **carry the tRNA into the A site, efficiently**.



When this happens if the codon anticodon pairing is correct the A site undergoes a conformational change that will promote the formation of the peptide bond, catalysed between the amino acid in P site and the one in A stie.

This peptide bond formation involves the transfer of the growing amino acid chain onto the amino acid carried by tRNA in the A site, this drags the amino acid tail in the p site while anticodon loop is still in the A site.
The two subunits ratchet one with respect to the other.

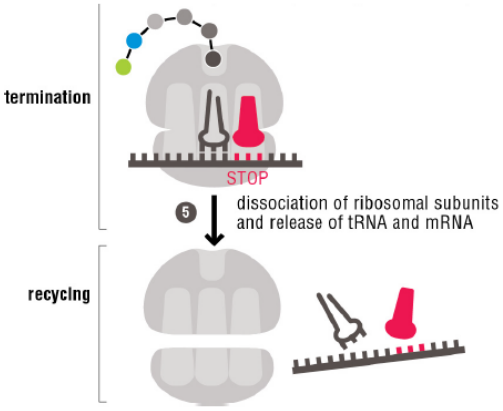
This is called **hybrid state**.

This is resolved with a **translocation** which is the moving one codon forward, promoted by another GTPase elongation factor called **EF-G**.

Now we have a liberated A site that is ready to accommodate a new aminoacyl tRNA.

Termination

Translation stops at a stop signal, the stop signal is universal and it is characterized by 3 possible codons: **UAG (amber)**, **UAA (ochre)**, **UGA (opal).**



At the stop signal elongation stops and ribosome dissociates, we need **release factors** for this.

The odd thing is that release factors are **not tRNA**, stop codons are coded by proteins that are tRNA mimics, they have the same shape of tRNA but they cannot accommodate the peptidyl bond transfer and this terminates translation.

There are 2 release factors in bacteria, **RF1** recognises UAA and UAG, while **RF2** recognises UAA and UGA (to remember, alphabetical order); in eukaryotes there is **eRF1** that recognises all three stop codons.

When RF are loaded in A site the translation stops, since the chain cannot be transferred on the release factor it falls and the two ribosomal subunits dissociate from mRNA.

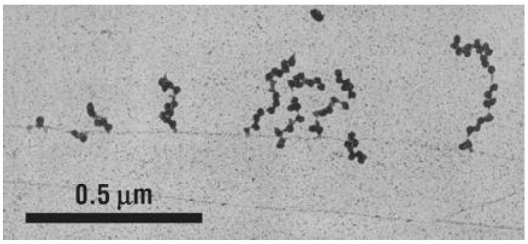
The ribosome can be used for a new cycle of translation thanks to recycling factors **RRF**.

Polysomes

As elongation proceeds the **initiator codon is freed** and it can be reached by IF to load another ribosome on the same transcript; when many ribosomes pile up in the same transcript this is called a polysome, this boosts translation.

Polysome is a mRNA translated by many ribosomes one after the other.

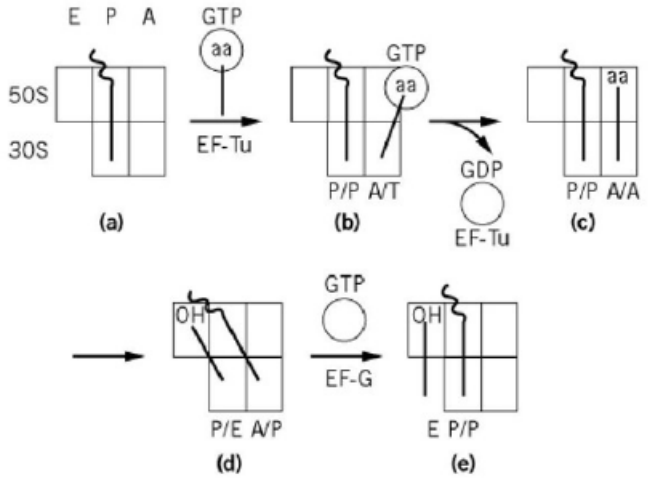
In bacteria translation and transcription happened simultaneously, DNA is transcribed with multiple RNA pol one after the other, at each transcript, as soon as the AUG codon emerges from the polymerase, the ribosome is assembled; eventually we have multiple polysomes linked to a DNA molecule.



Hybrid state

This particular phase involves the ratcheting of the two subunits one with respect to the other in a circular manner, this movement is allowed only in one direction, it can go forward but not backwards, and it is achieved through a conformational change (change is caused by the hydrolysis of GTP); for the resolution we need EF-G that is another GTPase.

The ratcheting is the movement of 50s with respect to the 30s which creates the hybrid state because anticodon loop of P and A site are in their position but the CCA tails with the amino acid and polypeptide chain are shifted of one position, respectively E and P site.



EF-Tu is the first GTPase that **loads the cognate tRNA** in the A site, if this is correct and codon anticodon minihelix is matched, the GTP activity of EF-Tu is activated and we have GTP hydrolysis. Then EF-Tu dissociates from the ribosome, the **non-hybrid state** is formed.

Next the peptidyl bond formation occurs, the polypeptide chain needs to be transferred on the amino acid in the A site, this transfer involves the moving of the tRNA tail of the A site toward the P site, forming the **hybrid state**.

Now the acceptor stem of the tRNA which has anticodon loop in the A site is in the P site and the same hybrid state is true for the tRNA in that has the acceptor stem in the E site and the anticodon loop in the P site.

To **solve** this state we need **EF-G** which is a GTPase that is able to move the small subunit to match the previous movement of the large subunit, this moves the whole complex of exactly one codon.

Translation factors

GTPases are the most important factors in translation but there are also other factors that act with different mechanism.
A common mechanism is to bind to ribosome and prevent unwanted interactions.

Examples:
IF1 and IF3 bind to the E and A site to prevent initiator tRNA to bind to A site.
Small and large subunits, don't assemble by chance thanks to IF6, only if the small subunit is associated with initial tRNA it can bind to the large subunit.

GTPases

GTPases Catalyse GTP hydrolysis, they can promote conformational changes and they are responsible for progressing the translation process.
In a way if we have to control the **right interaction before the hydrolysis** of GTP, if hydrolysis happens the molecules are bound and they can't go back.

GTPases act at several states in translation (EF-Tu, EF-G, eIF2), they interact with the long projection of protein arms of the ribosome in the protein-rich region, and they have a very conserved structure, since they all contain the **P-loop** which is a motive (Gly-X-X-X-X-Gly-Lys) important to bind GTP and it undergoes conformational change when GTP is hydrolysed.

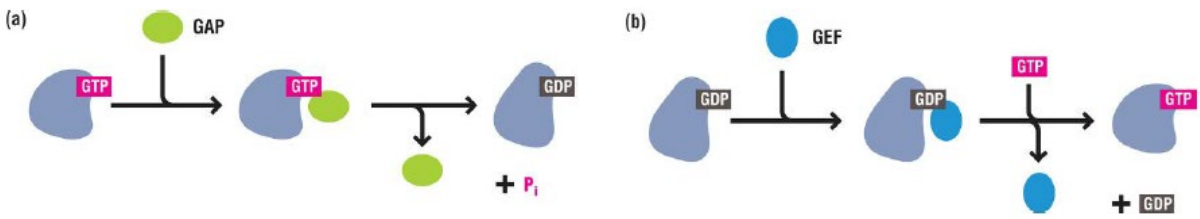
Usually the phosphorylation state of guanine nucleotide (GTP-GDP) dictates the conformation of the interacting ligand (in many phases of translation we have GTP bound to tRNA)

The ribosome is a complex network of conformational changes that drive the mechanism forward, for example, EFG-GTP recognises the pre-translocation state ribosomes (hybrid state), and this promotes hydrolysis to GDP that then dissociates from the ribosome (hydrolysis is anyhow energy given).

GAPs and GEFs

If a GTPase is bound with GTP and the GTP is hydrolysed into GDP this process is promoted by **GAPs**, **GTPase activating proteins**, GAPs stimulate GTPase activity and drive mechanism forward, they discharge the GTPase.

To recharge the GTPase are used **guanine-nucleotide exchange factors**, **GEFs**, they exchange GDP with GTP.



Some translation factors are GAPs other are GEFs (Ex. EF-TS it is a GEF for EF-Tu).

The **ribosome** itself can act as a **GAP**, when EF-Tu (GTPase) charged with a tRNA binds to the A site and the codon anticodon pair matches, a conformational change in the ribosome can promote the GAP activity of the ribosome toward EF-Tu activating GTPase activity.

Be aware that translation can proceed without these protein factors but with them it results to be much faster, moreover these proteins contribute also to the accuracy.

Molecular mimicry

Another very common theme in translation is the molecular mimicry.

The molecular mimicry consists in different proteins that adopt similar folds, a striking case is the mimicry of **release factors**, that mimic **tRNAs**, they have a similar conformation as tRNA and substitute it at the stop site.

EF-G (involved in the translocation, resolution of the hybrid state), has a very similar structure as EF-Tu (the factor that brings new aminoacyl tRNAs to the translation complex) and tRNA , this is because they fit in the same pocket, this makes sense because the two factor both bind in the A site.

Initiation

Initiation is extremely different in bacterial from eukaryotes, the initiation fact have to recognise beginning codon and they are helped by a RNA sequence near to it.

Bacteria

In bacteria the helping sequence is coded in the RNA, it is called the Shine-Dalgarno sequence, this sequence is about 6 nucleotides upstream of the initiator, AUG, and it is a purine rich sequence.

The Shine-Dalgarno sequence is recognised by sequence specific pairing, with a very conserved ribosomal RNA in the small subunit, the ribosome has an anti-Shine-Dalgarno sequence that allow by base pair to position the small subunit in the proper position if the initiator codon is present.

In bacteria the **small subunit searches** for the initiator signal, the initiator signal consists in the SD sequence preceding the initiator codon, the site that has this characteristic is called **ribosome binding site (RBS)**.

We can have long RNA, as soon as we have binding site the small subunit will bind and the initiator complex will assemble (if AUG present and initiator tRNA present), the initiator tRNA will bind the P site and when this happens the joining of the large subunit is promoted; the whole complex assembles at the RBS.

Eukaryotes

The initiator sites are not really conserved, there is just a **context** around the initiator codon, the Kozak sequence which is a sequence motive that promotes the start of translation.

In eukaryotes the initiator factors and the small subunit, do not load directly around the Kozak sequence, they bind to the cap at the 5’ end.

The **cap binding protein** binds to cap and **recruits all initiator factors** except of large subunit at the beginning of transcript and then this complex **scans** and moves along the transcript, until it finds Kozak sequence and a comfortable initiator codon.

Initiator tRNA

Initiator methionine tRNA must be different from a normal tRNA, they are modified and they usually bind AUG, with the anticodon UAC (opposite direction).

GTPases are responsible for driving the initiator tRNA into the P site and their GTPase activity is activated only if tRNA is actually loaded in the P site.

Bacteria

Bacterial initiator tRNA has a formyl group in the methionine, it is called formyl methionine tRNA ($tRNA^{fMET}$) and it has a CA wobble in the acceptor stem, but

Eukaryotes

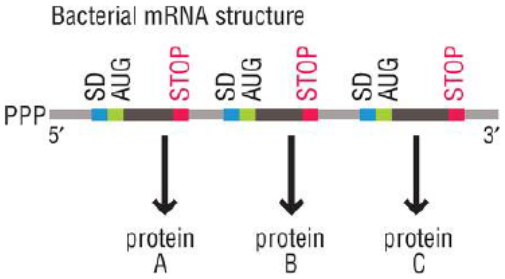
In eukaryotes the initiator tRNA has a normal methionine and also normal Watson-Crick base pairs in all the stems

Both initiator tRNAs have 3 **GC** pairs in **anticodon stem**, they avoid the binding of EF-Tu to the tRNA, EF-Tu is the GTPase responsible for loading the aminoacyl tRNAs into A site, the initiator tRNA doesn't have to be loaded into the A site, therefore these conserved GC pairs distinguish molecularly an initiator tRNA from normal methionine tRNA.

Start site

bacteria

Ex. Multicistronic operon that give rise to one mRNA:



Look how many proteins can be translated independently from the same transcript, each with its SD sequence and a stop codon, each of these will give rise to a different protein but they can be carried on the same RNA, sometimes different protein sequences can be overlapped.

The Shine-Dalgarno sequence is so important because it is recognised from anti-SD sequence which is present in 16s rRNA, in the 3' end, in a polypyrimidine region.

2 conditions need to be met to define a Shine-Dalgarno sequence:
- presence of AUG
- polypurine tract upstream 6-8 nucleotide from the starting codon.

The SD sequence is located in 5' UTR, just before the ORF (the ORF starts after the initiator codon).

Ex.

```
        U U C C U C C
T A C T A A G G A G G T T G T - - A T G
A C C T G A A G A T T A A A C - - A T G
  G T G G A G G A C T A A G A A - A T G
    T T A G A G G G A C A A T C G A T G
    G G G A G T A T G A A A A G T A T G
T A A C A G G T A G T G A A T - - A T G
    G T A A G G A A A T C C A T T A T G
  A C G A G G G G A A A T C T G - A T G
A A A T G A G G A G A G G T A - - A T G
```

As with bacteria promoter we have a consensus and the more the shine Dalgarno motive is similar to the optimum (AGGAGG or GGAGGA), the better the small subunit is recruited and the stronger translation start site will be.
In this case the recognition is not protein-RNA but it is RNA-RNA, the deviation from consensus controls strength of translation.

Initiation factors
At the start site we have 3 initiator factors involved on top of the small subunit of the ribosome and the initiator tRNA: IF1, IF2 and IF3; they help guiding the formyl methionyl tRNA into the P site.

IF1 binds the A site and **IF3 binds the E site**, these two factors occlude the wrong site in the small subunit.

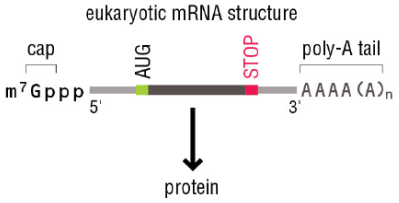
In the presence of mRNA with the SD sequence and with the small subunit loaded with IF 1 and 3, the mRNA can be recognised through the anti-SD and when this happens the initiator tRNA can be **loaded onto the P site** thanks to another GTPase which is **IF2**.

IF2 is able to sense that the P site is occupied by the initiator tRNA and that everything is set and ready, this activates its GTPase activity and when hydrolysis occurs then the large subunit is recruited to the small subunit.

IF2 is signal and **provides energy** to join large and small and remove all other factors.

Eukaryotes
We don't have the equivalence of the SD sequence and the small subunit doesn't need to recognise directly the initiator site.

The initiator complex is assembled at the cap and it proceeds by scanning; because cap is so important, this implies that eukaryotic RNAs are usually monocistronic (just one ORF), each ribosome will just read the first initiator sequence and the first stop, then it will detach with no possibility to link again to the transcript.

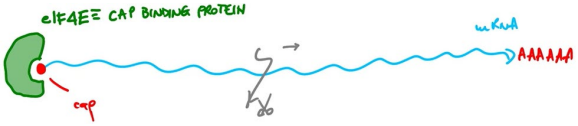


Usually since the scanning process starts from 5' end, the first AUG we find is the initiator, but the scanning can be aided by the **Kozak sequence**, which has the following consensus: RNN**AUG**G, this is a weak consensus but it promotes the recognition of the AUG inside of it as **initiator codon**.

The initiation process is also aided by the **poly-A tail**, 3' end, the eukaryotic RNA **circularise** because of initiation factors but the presence of the poly-A tail and the cap are both important to stimulate translation initiation.

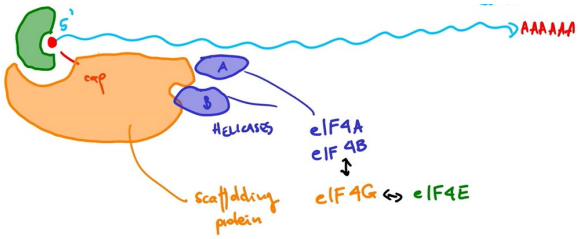
Pre-initiation complex
Cap can be recognised specifically by an initiation factor called **eIF4E** which is the **cap binding protein**.

If by chance the RNA is processed (cut) the processed part will not be translated because it doesn't have a cap, this mechanism allows a very nice proofreading mechanism, the cell will not translate a damaged RNA and it will not produce a truncated protein.

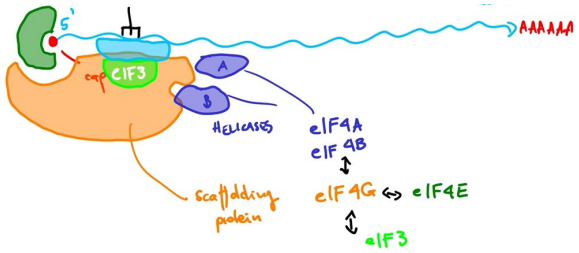


After the eIF4E the 5' end is reached by helicases (eIF4A and eIF4B) that are responsible for opening up secondary structures that may form in the mRNA.

Then **eIF4G** is recruited by eIF4E, it is a scaffolding protein and it interacts by protein-protein interaction with the cap binding protein and with the **helicases**.

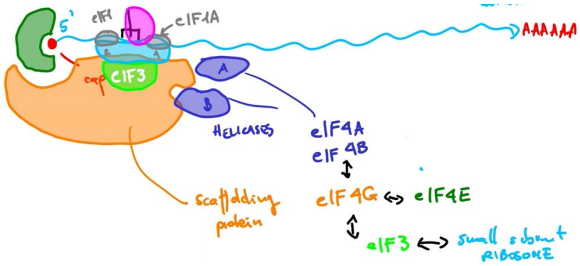


Another important factor is recruited by eIF4G, which is **eIF3**, it has the role to **bind the small subunit** of the ribosome



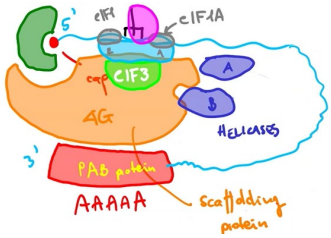
Through the protein-protein interactions mediated by the scaffold and eIF4E, the small subunit was able to be recruited.

Now, similarly to bacteria, we have 2 proteins that bind the A site and the E site, they are called **eIF1A** and **eIF1**; and, again similarly to prokaryotes, a **GTPase (eIF2)** goes along with the complex, and its hydrolysis will signal that the initiator site has been found and it will allow the assembly of the 2 subunits of the ribosome.

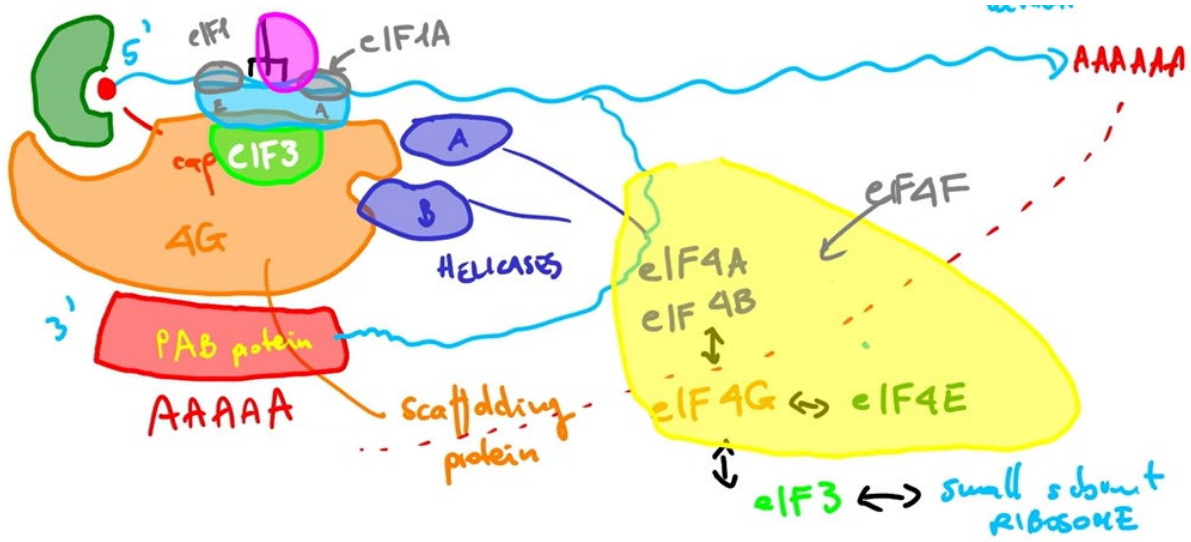


The scaffolding protein has another feature, it binds to the poly-A binding protein (**PAB protein**), this allows the transcript to circularise and the translation to start; translation could start without PAB

protein but it would be weaker.



Moreover, sometimes the complex formed by G(scaffold) E(cap binding protein) and A&B(helicases) is referred to as **eIF4F**.



All these interactions are fundamental to start properly translation.
The formation of the close loop acts as a **quality control**, because if we have a damaged DNA without a poly a tail the translation will be weak or off.
This huge complex associates at the beginning of the transcript and then it starts to scan, this complex is called the **pre-initiation complex (48s)**.

Initiation complex
The last step before the elongation is the scanning, the pre-initiation complex starts the scanning process, while the two ends of the transcript are still kept together by the appropriate factors.

The initiator tRNA starts looking for the initiator codon, usually it is either the first AUG codon and it is perfect if there is a Kozak sequence around it.

Once the **starting codon** has been **found** the large subunit must be joined, this process is mediated by another GTPase (eIF5B) that promotes the recruitment of the large subunit and that acts as a CAP toward eIF2 GTPase activity that fixes in place the initiator tRNA onto the AUG codon.

Hydrolysis of eIF5B causes also all the unnecessary factors to leave the ribosome that starts working alone, now we elongation starts and simultaneously another pre-initiation complex can be loaded on the cap.

Elongation

Elongation needs to be precise, otherwise all the information is lost, this is achieved through a lot of RNA-RNA interactions.

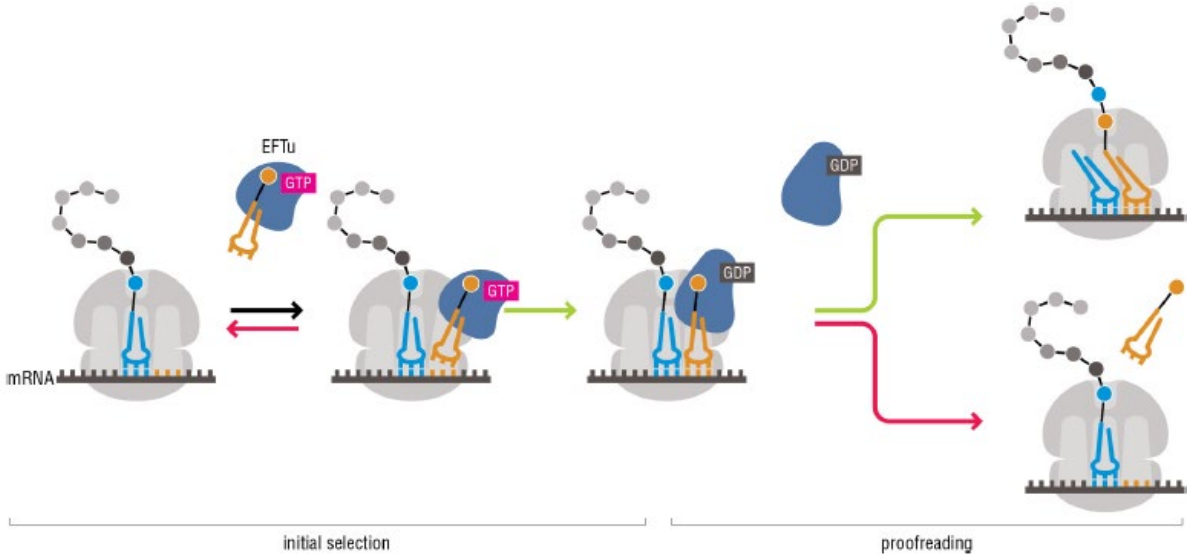
Bacteria

Cognate amino acid

A new aminoacyl tRNA is added in A site, the factor that mediates this step is the GTPase EF-Tu.

The decoding happens during the addition of each cognate amino acid.
Cognate means that base pair codon-anticodon is perfect, the wobble is obviously part of the match and it is tolerated.

Non-cognate and near-cognate need to be rejected while cognate need to be accepted, to do so the ribosome is helped by some different characteristics of the cognate and non-cognate binding:
a. In the cognate situation there is a **stronger binding**, with better affinity, the non-cognate may bind but it has a lower affinity, and ,most importantly, non-cognate amino acids do not form a p**erfect minihelix** in the A site through the anticodon-codon match.
b. when the cognate situation is met than there is the activation of **GTPase activity** of EF-Tu, the GTP hydrolysis drives the process further.



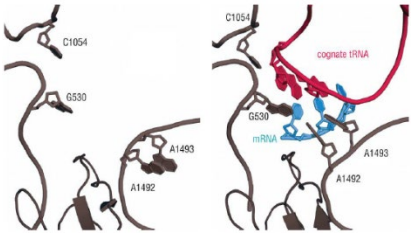
The formation of the perfect minihelix stimulates **conformational changes** in the ribosome, now the ribosome can recognise the formation of minihelix and the ribosome itself can act as **GAP** (this is stimulated by the conformational change), activating the GTPase activity of EF-Tu and accepting the tRNA in the A site.

The bind check acts on two levels: firstly, If GTP hydrolysis does not occur the tRNA is rejected and another aminoacyl tRNA can join; secondly, if the conformational changes in the ribosome are not correct, even if the GTPase activity was activated, the near-cognate tRNA doesn't fit and it can be rejected.

Minihelix
The ribosome controls the formation of the minihelix through an **A-minor triplex** (adenine forms a triple helix in the minor grove of the mini helix).

All **ribosomes have extremely conserved nucleotides** in the 16s (small subunit): A1492, A1493 and G530 present in all living organisms.
These 3 nucleotides can form an **hydrogen bonding network** with the **minihelix** in the A site of the ribosome.

Each of the 3 bases is flipped out at the core of the A site, when there is a cognate pair and the codon-anticodon matches perfectly the minihelix is formed, G530 and especially A1492 and A1493 can insert in the RNA helix (in the minor groove) and inspect the formation of the hydrogen bonds, especially in **position 1 and 2** of the codon (positions in which wobbling is not accepted).

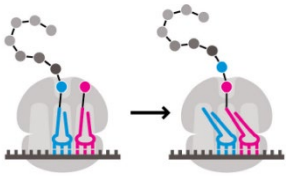


The small ribosome subunit has a reading head formed by adenine bases that flips inside the minor grove of the minihelix and inspect if the interaction is cognate.

The flipping of the adenine is the **conformational change**, that modifies the small ribosome and activates the gap activity.

Peptide bond

In elongation we have a second process that is the **formation of the peptide bond**, it involves the transfer of the elongating polypeptide chain onto the aminoacyl tRNA and it leads to the formation of the **hybrid state**.

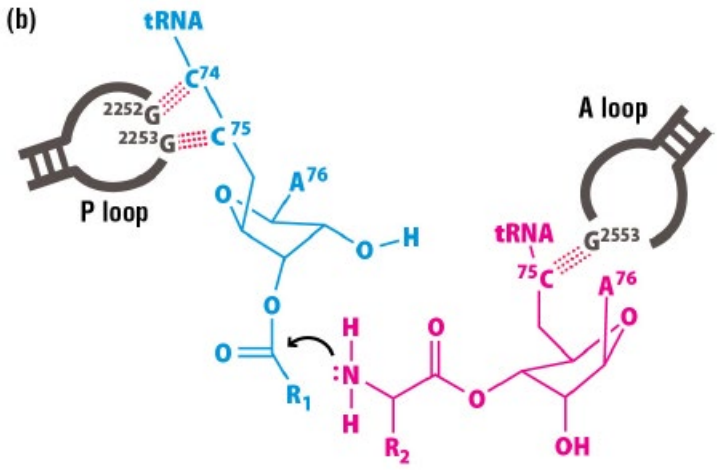


This phase is performed by some RNA components, and thanks to the 2'OH group of RNA.

The **CCA tail** of the acceptor stem of **P tRNA**, is kept in place in the P site by hydrogen bond networking by two **guanines** in the **P loop** of the 23s ribosomal RNA, **large subunit**.

A similar interaction occurs in the **A loop**, with the **acceptor stem** of the **A tRNA** again formed by RNA.

In this situation the **hydroxyl group** promotes the nu**cleophilic attach** and catalyses the transfer of the chain to the amino acid and thereby the formation of the peptide bond.



The hybrid state was formed.

Translocation

To solve the hybrid state the translocation of the mRNA-tRNA complex has to be promoted in order to empty the A site and proceed with the translation.

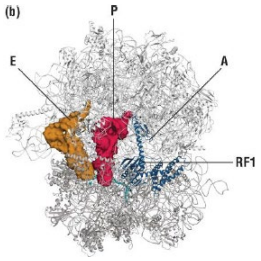
The anticodon loop of tRNA in the A site has to be moved to the P site and the tRNA in P site has to go to the E site; this series of movements is performed by GTPase **EF-G**, that binds and squeezes into A site, in order to push the aminoacyl tRNA anticodon loop one step ahead.

The accommodation of EF-G into the A site promotes structural rearrangements needed to translocate and solve the hybrid state (ratcheting).

Termination

Translation continues until the ribosome encounters a stop codon in the mRNA (UAA, UAG or UGA)

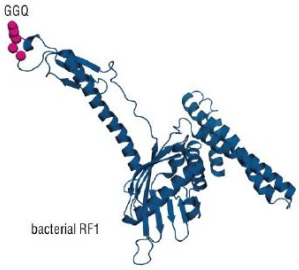
Class I release factors

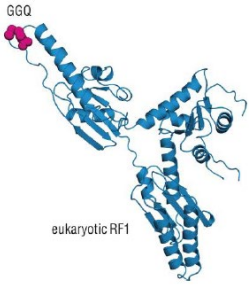


Class I release factors recognise stop codons, they are not tRNAs but they are protein that mimic the structure of tRNA because they need to dock into the A site; they also promote hydrolytic release of the finished peptide.

bacterial and eukaryotes have a conserved shape and a conserved GGQ motif that is needed for the catalysis

- Bacteria:
  - RF1: recognises AAG and UAA
  - RF2: recognises AGG and UGA
- Eukaryotes
  - eRF1: recognises all the stop codons

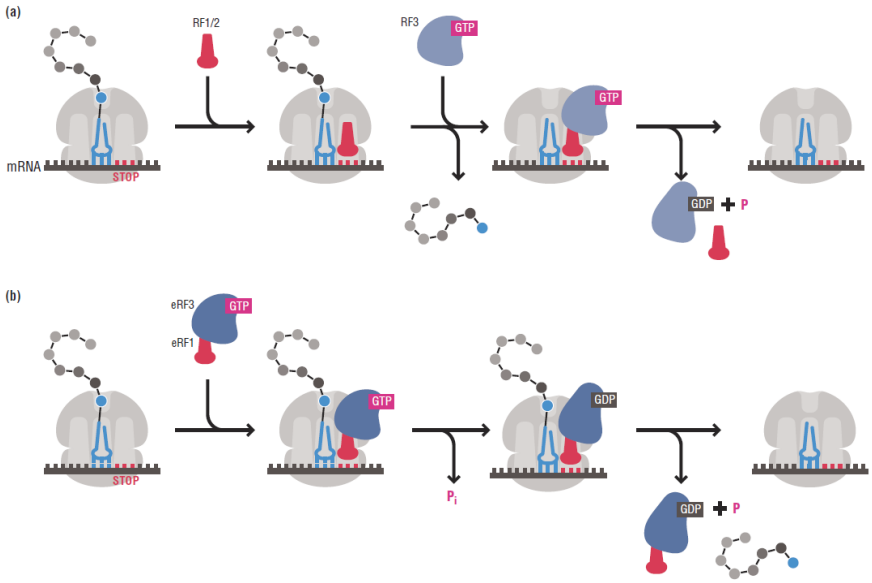




When the glutamine motif (GGQ) reaches A site it reacts with water and does not form the peptide bond.

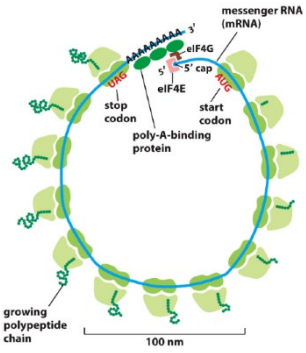
Class II release factors

Class II release factors promote the separation of class I factors from the A site after the releasing of the peptide; they are GTPases.
In bacteria RF3-GTP (derives from EF-G) promotes the dissociation of RF1 and RF2.



Circularisation

Polyribosomes are molecules of mRNA, they are folded up in a cyclic structure thanks to PAD protein and to the scaffolding protein.



The transcript is totally covered by ribosomes that perform translation and, once a ribosome has reached the termination, it can rapidly restart from the initiation, it is already in place, this is a super-efficient mechanism.

REGULATION OF TRANSLATION

As for transcription, there exist many ways in which the translation process can be regulated.

UTR regulation

UTR regions are not translated but they play an important role in the regulation.

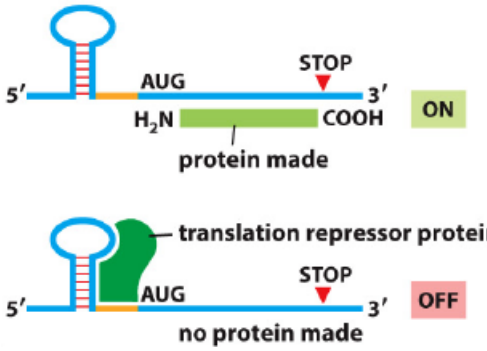
5' UTR

5' UTR can recruit factors that eat up the poly A tail, in this way there will be no protection onto the 3' end and the molecule will be unstable, also there will be no looping, so the translation will be less efficient.

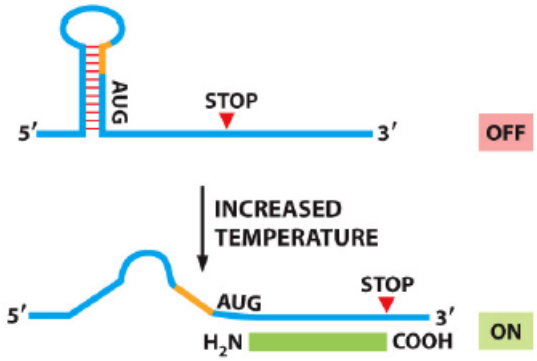
Most of the time translation **regulation in bacteria** is associated with the **accessibility of the Shine-Dalgarno sequence**.

Examples In bacteria (the yellow stretch is the Shine-Dalgarno sequence):

1. a **secondary structure** in the RNA can be **recognised** by a translation repressor protein that may bind and occlude the Shine-Dalgarno sequence, if it is occluded no translation is going to initiate.

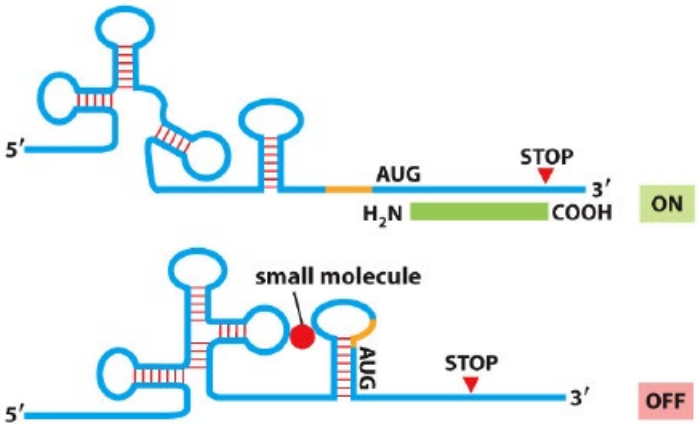


2. Riboswitch (thermometer): this mechanism works in cis, the regulator is part of the RNA itself, a **secondary structure** of the transcript can **involve the initiator** codon and the SD sequence, putting them into a stem; in this situation the initiation factors cannot access the SD sequence.
   The riboswitch can be activated by increasing temperature, because increasing temperature can open up the secondary structures and the binding site will become accessible to the initiation factors and translation will be turned on.



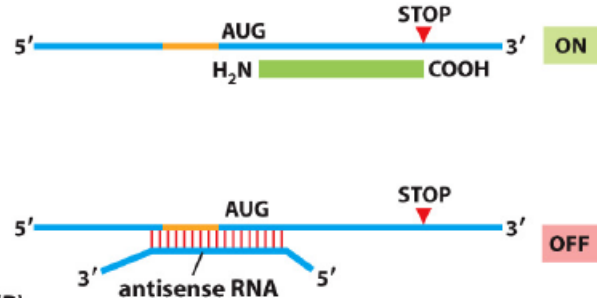
   by changing the bases in the stem the temperature to which the secondary structure melts changes too
3. Another example of cis riboswitch, in this case there can be an alternative secondary structure according to the presence of a **small molecule**.

aptamer domain effector platform



the aptamer domain binds the small molecule while the effector platform contain the secondary structure that hides or unhides the ribosome binding site according to the presence of the small molecule.

4. The ribosome binding site can be occluded by an **antisense or regulatory RNA**, these RNAs can pair by perfect base pairing but also there can be pairing with imperfect complementarity that causes the formation of secondary structure.
anyway if a duplex is formed with the ribosome binding site, the translation will be off.



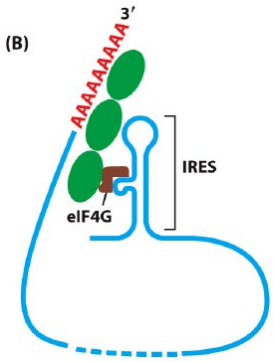
3’ UTR

The 3’ end especially in eukaryotes is rather targeted from RNA decay; the binding of particular RNA molecules or proteins at the 3’ end can stimulate the de-adenylation of the poly-A tail and lead to degradation.

Internal ribosome entry sites (IRES)

IRES are RNA sequences that build a secondary structure in mRNA that is able to bind to scaffolding protein (eIF4g), they allow the transcript to be translated even if it doesn’t have the cap, it is a bypass of cap recognition, it substitutes the cap.

(viruses use IRES sequences in order to hijack the translation machinery and translate their transcript)

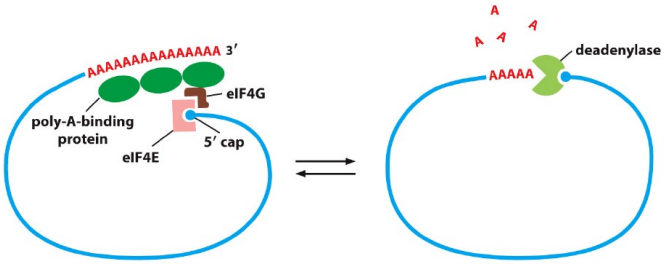


mRNA stability

mRNA stability is important for gene expression, if RNA is not translated, then it is degraded, we have an equilibrium, a **competition** between different machineries, translation machinery and degradation or decay machinery.

This mechanism is defined as a competition because if translation initiation factors are present they prevent the binding of decay factors; this competition allows for the checking that defective RNAs are not translated and they are degraded.
Even non-defective RNAs will be degraded, if not recognised, bound and protected by the initiation factors.

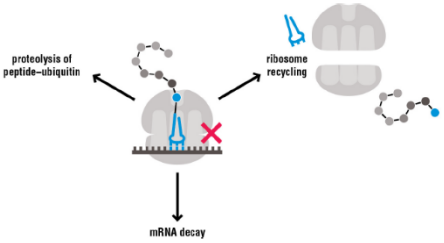
The initiation factors act as a protection because they loop RNA, preventing the starting of the decay from both ends, also the ribosomes physically occupy the central portion of the molecule and protect it.



Ribosome stall

Sometimes in bacteria where there is no capping and no poly A tail, there is the need of a mechanism to rescue ribosomes that are stuck.

When the ribosome can't proceed further it is a sign of danger for the cell, therefore the RNA must be degraded and it will be directed toward the RNA decay pathway, the polypeptide (which could have a loss of function) needs to be destroyed (proteasome) and the ribosome must be recycled.



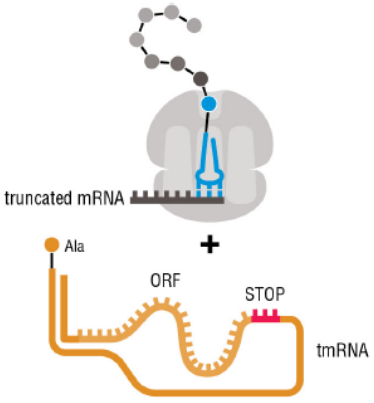
In eukaryotes the stalling is not a problem, because we have a temporal and physical separation between transcription and translation and we have capping and poly-A tail that control translation levels; while in bacteria transcription and translation are **coupled**, so translation can occur on **truncated RNA**, if this happens the ribosome will reach 3' end of the transcript **without** having encountered a **stop codon**.

A nice mechanism is able to solve this situation:

Translation will "end" with and empty A site if the transcript is truncated, in this case a super useful molecule can be recruited: tmRNA.

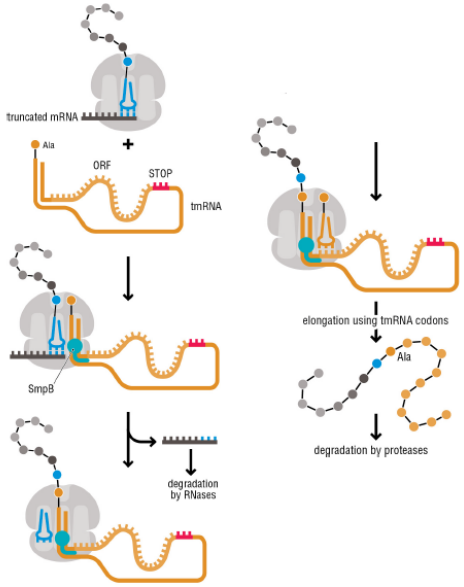
tmRNA is a bivalent mimic of a tRNA and a messenger RNA both in same molecule, it acts as a translation rescue system and a protein tagging system at once.

When there is an empty A site the tmRNA can act as a **tRNA** and it can load in an aminoacyl site, carrying an **alanine** along.
The real power of tRNA is that is not just an alanyl tRNA, it is a much longer molecule, it carries along, fused in cis, a shot **cistron**, a sequence that can be translated.
Once the tmRNA is loaded translation can continue on mRNA part of tmRNA while the original mRNA will be liberated and degraded.
tmRNA contains **11 codons** followed by a stop codon, the **stop codon** will allow recruitment of release factors that will terminate the process, while



the 11 codons correspond to **11 amino acids** appended to the polypeptide chain, these additional tag codes for a degradation signal for the truncated protein.

If any sequence has a mutation that eliminates a stop codon, then ribosome will reach the end of the transcript and stall, at that point the rescue mechanism will activate.



Reconding

With recoding we refer to the bypass of codon information, it happens rarely, and sometimes it is not an accidental mechanism, it happens because the sequence contains a motive, a signal and a conformation that promote this recoding.

Recoding is reinterpretation of a transcript.

There are two main types of recoding:

- **Nonsense suppression**: bypass of a stap signal
- **Programmed frameshift**: changing the reading frame
    - **-1**: ribosome backtracks 1 nucleotides and restarts.
    - **+1**: the ribosome reads 4 nucleotide, slipping 1 nucleotide (rare codons may be skipped.)

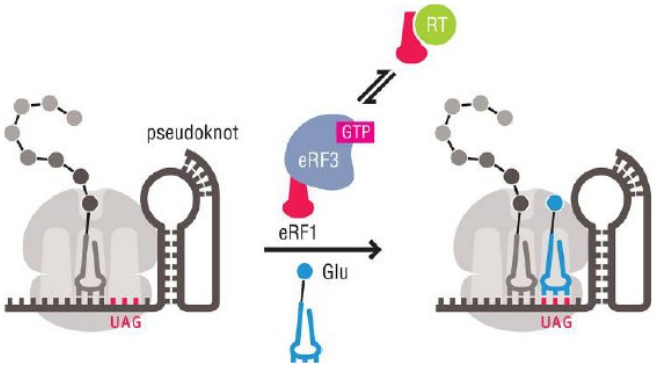
Nonsense suppression

Ex. Imagine we have a knockout in release factor 1 (RF1=UAA, UAG; RF2=UAA, UGA), if we encounter a UAG codon when the RF1 is knocked out we will not be able to stop the translation, it wont be read as stop codon.
In this case UAG will recruit another tRNA, in particular glutaminyl tRNA, this situation describes a mutation that causes an overriding of a codon, it is a bad instance.

In programmed nonsense suppression the **suppression** is **promoted** by the presence of a particular **secondary structure** ahead of the nonsense suppressor site.

If, downstream of an UAG codon, there is a pseudoknot, the ribosome will slow down.
In normal conditions, with plenty of release factor, there would be no problem and translation would terminate anyhow; however in some situations there can be factors that are able to recruit the release factor, kidnapping it; if RF is no more available, then the glutaminyl tRNA can compete for the binding in the A site and override the stop signal.

This mechanism consists in the balance of different factors: the likelihood of the decoding of UAG with glutaminyl tRNA has to compete against the stop signal, in this battle it is helped by the factors that switch off the RF1 and by the secondary structure that slows down the process, allowing a molecule with low affinity to bind to the codon.



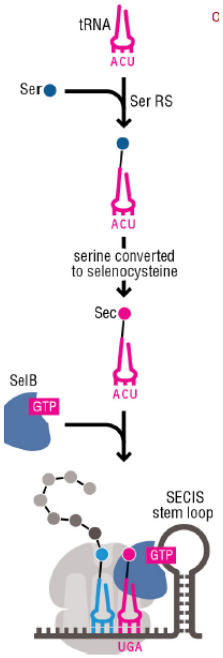
Incorporation of nonstandard amino acids

Stop codons also allow to incorporate nonstandard amino acids, this is used by biotechnologist to produce tagged protein.

Ex. (natural example) Selenocysteine is a particular amino acid, it is a cysteine that has selenium instead of sulphur, it is incorporated, in the catalytic site of several enzyme that are involved in redox reactions.

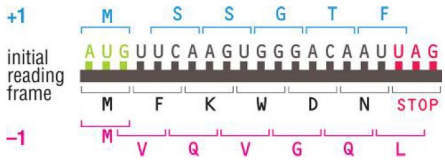
In order to code for selenocysteine, we need a tRNA charged with selenocysteine this particular tRNA derives from serine tRNA.
This means that the selenocysteine tRNA will match with the UGA codon.

Finally a particular secondary structure called **selenocysteine insertion sequences (SECIS)** can recruit particular proteins that aid the loading of the selenocysteine tRNA into the A site instead of the release factor.



Frameshifting

A frameshift is the change of the reading frame, it leads to a completely different RNA interpretation.



Frequently in frameshifting a premature stop codon occurs or the existing stop codons are overridden.

Sometimes frame shifting occurs because of errors, the ribosome may read incorrectly the sequence, but this is rare; frequently we have a secondary structure in RNA that promote frameshifting at specific site, this is programmed frameshifting.

Programmed frameshifting can have important roles in gene regulation (viruses use it to infect more efficiently).

+1 frameshift

Ex. Overriding of RF2 it reads UGA

Imagine we have **RF2** and then we have the **tRNA with leucine** that has the CUG anticodon, these are the two players.

If we have sufficient RF2 there is no problem, termination occurs; However, a specific sequence can trigger a frameshift, in particular, if the sequence is **CUU UGA C** the ribosome can slip one nucleotide ahead (the amount of RF drives the frame shifting).

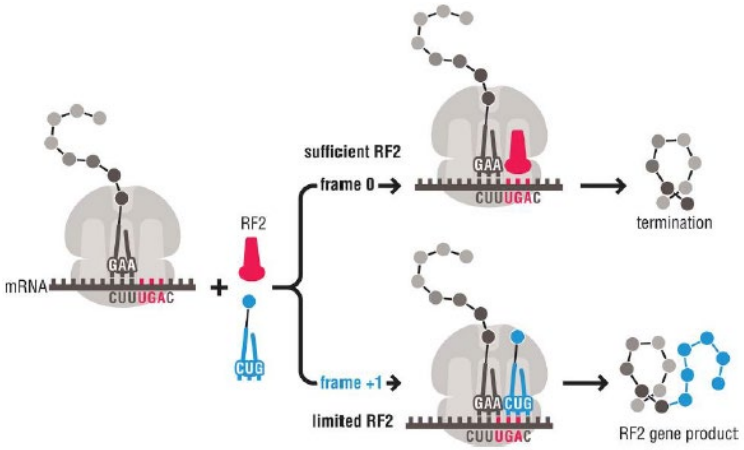
This is because CUU pairs with GAA, but GAA can also pair with UUU, there is a wobble involved but it is tolerated; because of this possibility, if the last U was part of a stop codon, after the frameshift it could form a GA- codon, in our example we yield a GAC frameshifted codon that can be read out by CUG anticodon, this allows to bypass the stop codon.

We have a bypass of a stop codon that involves a frameshift.

C U U U G A C	slip	C U U U G A C	produce	
G A A stop	→	G A A C U G	→	RF2
perfect match		decent match		

The product of this frameshifting is RF2 itself.

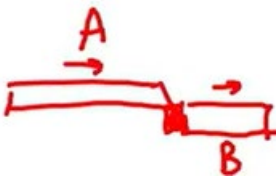
If there is enough RF it will stop its production if there is not enough it will promote a frameshifting responsible of its generation.



-1 frameshift

This example consists in the destabilisation of the correct advancement of the elongation process again with a pseudoknot that impedes progression of ribosome.

This is frequent in viruses: they can make 2 products out of same RNA sequence, these 2 products are from 2 coding sequences frameshifted by 1 base.



If there is no frameshift only polypeptide A is produced, with a frameshift the transcript results to be a **fusion of polypeptide A and B**.

Viruses use this to promote the proper expression of their polymerases, they can control whether if producing only 1 polypeptide or 2 fused polypeptide that have different functions

The **mechanism**:

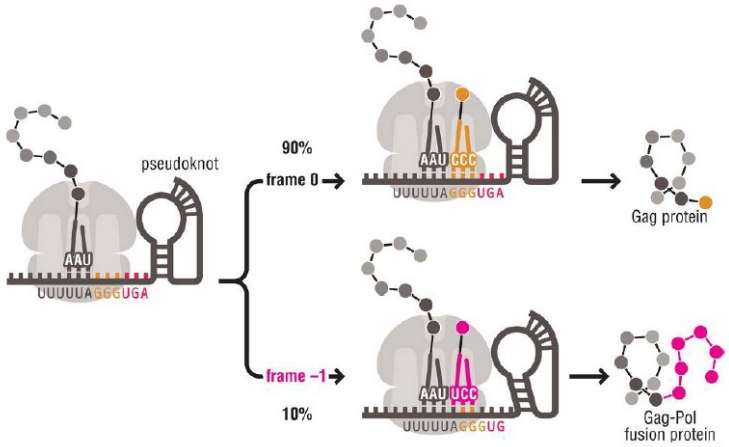
In most cases ribosome terminates at proper UGA stop codon, however in this programmed

frameshifting there is a very uracil rich sequence that allows the ribosome to backtrack by one position.

The backtracking is Induced by the pseudoknot that pushes the ribosome and by the RNA sequence, that is slippery (it is uracil).

U UUA GGG UGA	slip	UUU AGG GUG A
AAU CCC stop	→	AAU UCC CAC
stop codon is present		translation progresses

Now we have a near cognate situation that causes the bypass of the stop signal, translation progresses.



The mRNA sequence is said to be slippery because codons can be accommodated in a -1 shift.

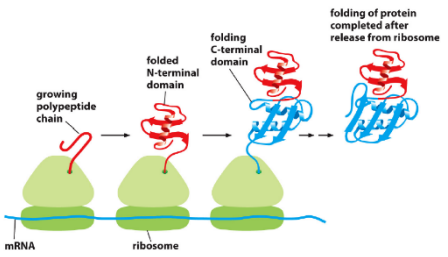
Antibiotics

A lot of antibiotics (used to control pathogenic infections) work specifically on translation mechanisms of bacteria, these antibiotics block growth of microorganisms.

Many of those antibiotics work at different steps of translation process, they can specifically target bacterial or fungal processes; the particularity is that there are differences in translation machinery between bacterial and eukaryotes, so antibiotics can target ribosomes at the different phases or provoking premature termination of translation

Co-translational protein folding

The polypeptide chain starts folding progressively as it exits from the ribosome, in modular proteins frequently the second domain is folded only if the first one has finished folding, this is very important, it is influenced by the rate and speed of translation.

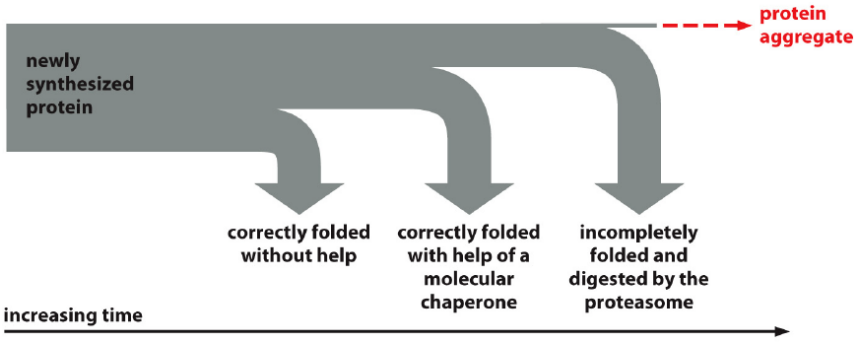


Sometimes there are pausing sites, encoded in the RNA (ex. Secondary structures) that slow down the translation in particular moments in order to give time to the first part of the polypeptide sequence to fold correctly before the rest of the polypeptide gets translated.

This gives time to the first part to fold properly, if we don't give time then the hydrophobic patches of one domain may interact with the patches of the other domain and the folding will be completely wrong.

This mechanism cause problems to scientists that are willing to express heterologous genes, for example mammalian gene in bacteria; this is because translation in bacteria is much faster and the mammalian polypeptide has not enough time to fold properly, the resulting molecule is called inclusion body or unfolded protein aggregate.
One trick to solve this is to grow the bacteria at a lower temperature so that replication and translation are slower and there is more time for folding, or introducing rare codons in particular positions in order to slow down ribosome and give time to peptide chain to fold properly

Post-translational regulation



Frequently translation yields proteins that are folded correctly but that need the help of a molecular chaperone, this process can be post translational and it controls the expression of a gene because it allows the correct folding.

Sometimes the protein or polypeptide is improperly folded, if this happens, to avoid protein aggregates that can be deleterious for the cell, it is better to degrade it; this is performed by the proteasome (we have seen 2 degradation machineries of the cell: the degradosome that is important in bacteria and it degrades RNA and the proteasome that digest proteins).

Molecular chaperone

If a problem of misfolding occurs, most of the time this misfolding is because of hydrophobic regions in polypeptide sequence that are not folded correctly.
Hydrophobic patches tend to interact and connect between each other, Usually they allow the correct folding.

Exposed **hydrophobic patches** signal that something went wrong, they can be used for protein **quality control**, in fact they are usually recognised by chaperones.

Some chaperones, like Hsp 60 and Hsp 70 (heat shock proteins) have affinity for hydrophobic patches and they allow the refolding of proteins that have exposed hydrophobic patches, but also they mask these hydrophobic patches and prevent further aggregation of misfolded proteins.

Chaperones are frequently **heat shock proteins** because when we have excessive heat it can denature the fold of proteins, if proteins are denatured hydrophobic patches are exposed (also other stresses can unfold proteins, so there can be other types of chaperons).

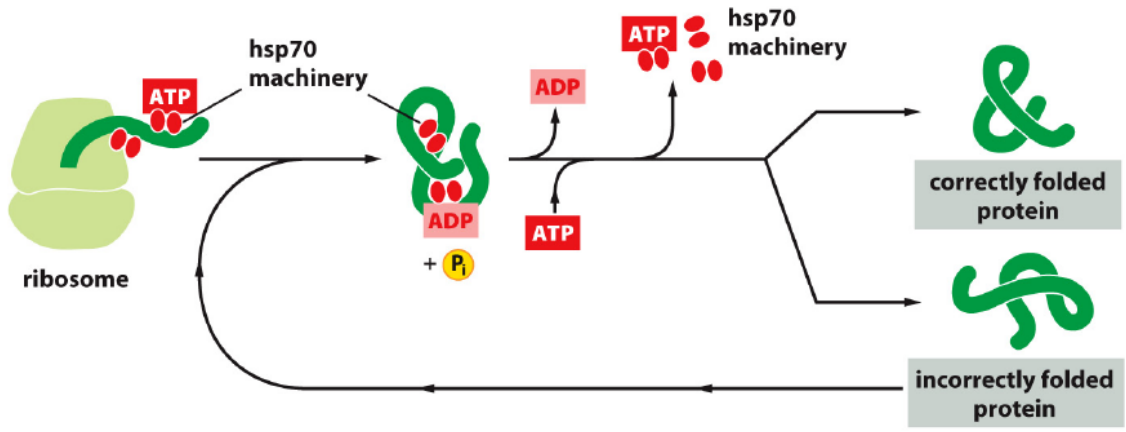
As post translational control all organisms have chaperones that aid when the temperature gets too high and it denatures the protein fold.

Hsp70

Hsp70 machinery acts early in life of a protein (actually it is co-translational), as polypeptide exits the ribosome it **binds to hydrophobic patches**, that may need to interact with membrane or other

similar post-interactions; Hsp70 binds to the protein thanks to its ATPase activity, after hydrolysis it associates tightly with the polypeptide chain.

When hsp70 is bound hydrophobic patches are not available, this gives time to have a pre-folding of the protein, which is kept in place by other types of interactions; when ATP binds again hsp70 dissociates from the pre-folded protein and then some selected hydrophobic patches interact with each other.



Sometimes this is not enough, the proteins could end up as misfolded.

Hsp60

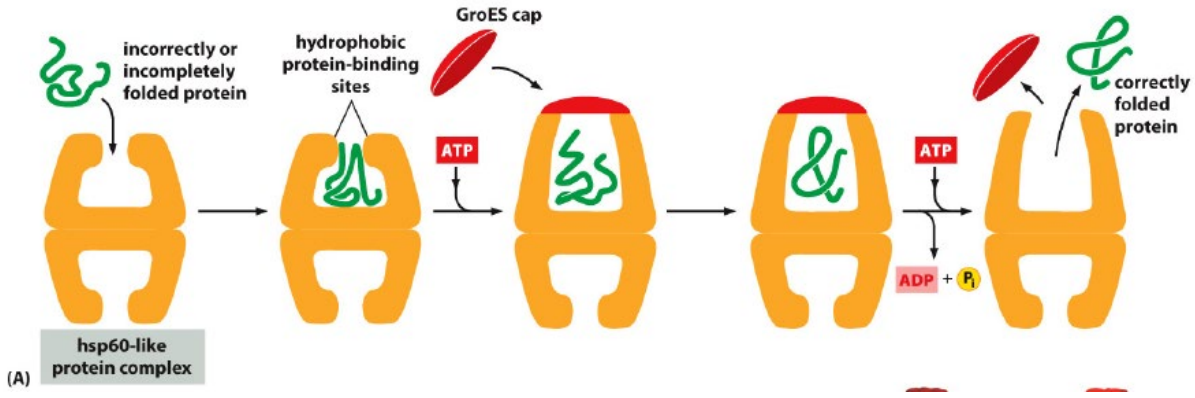
Hsp60 chaperon or GROEL-ES system is a complex that tries to fold back to the proper state the misfolded proteins.

It is a sort of barrel with a cap, where misfolded proteins get internalised, the lid is closed and then thanks to some biomolecular mechanisms a correctly folded protein comes out.

GROEL is the large yellow part, **barrel** like, it is an isolation chamber, covered with hydrophobic patches in the internal surface, that will interact with the hydrophobic patches of the incorrectly folded peptide.

We can spend ATP and by addition of the **cap** (**ES**) to have a conformational change that allows to unfold the incorrectly folded protein and provides an environment in which the protein can fold properly again.

When this happens we have to spend again ATP to open the lid and free the protein.



If this doesn't work we definitely have a incorrectly folded protein, but it could be deleterious, so it is better for the cell to degrade it through the proteasome

Proteasome

The proteasome is a huge bulky environment which has proteolytic activity, this activity aids out in the degradation of the misfolded protein.

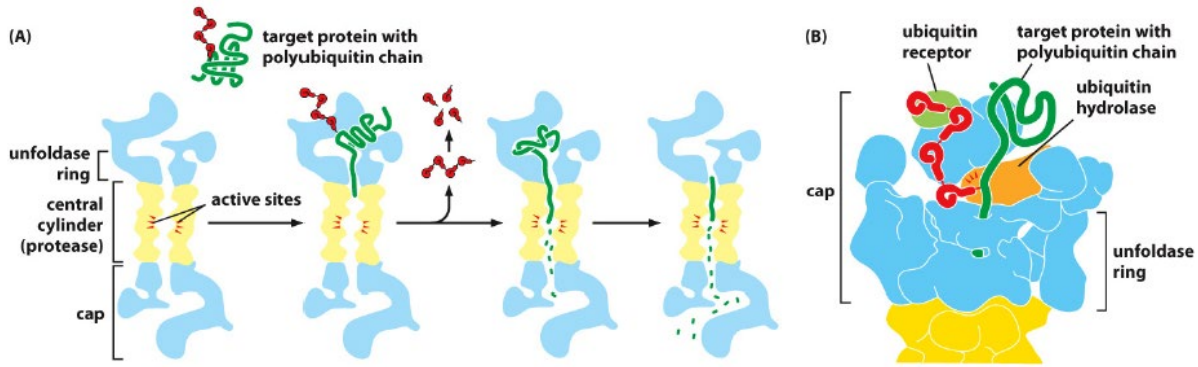
The structure of the proteasome resembles a sort of tube, with different domains (it is symmetrical):

- **Unfoldase** (top ring): responsible for unfolding the misfolded peptide.
- **Protease** (Core): active site, responsible for hydrolysing the polypeptide.
- **Cap** (terminal domain): the amino acids exit.

We heard about this before, the proteasome is able to degrade in eukaryotes especially ubiquitinated polypeptides; this is due to the fact that the proteasome has a **ubiquitin receptor** that allows the **docking** of the ubiquitinated polypeptide.

The recruitment of the misfolded occurs at the unfoldase domain: in this domain there is a **ubiquitin hydrolase,** responsible for removing ubiquitin tag and an **undolfase ring**, responsible for the unfolding of the protein in order to pass to the protease a linear polypeptide sequence.

We have to spend energy in order to solve the problem.



REGULATORY RNAs OR RIBOREGULATION

Riboregulation can happen both at transcriptional and post transcriptional level, it is a regulation performed by regulatory RNAs (short).

It has been discovered that this kind of regulation happens in all domains of life, and its discovery has created a new paradigm in molecular biology at the beginning of the millennium (Nobel prize gave to Craig and fire).

A lot of RNAs are non-coding for a protein product but their role is to regulate gene expression, at transcript or post transcriptional level.

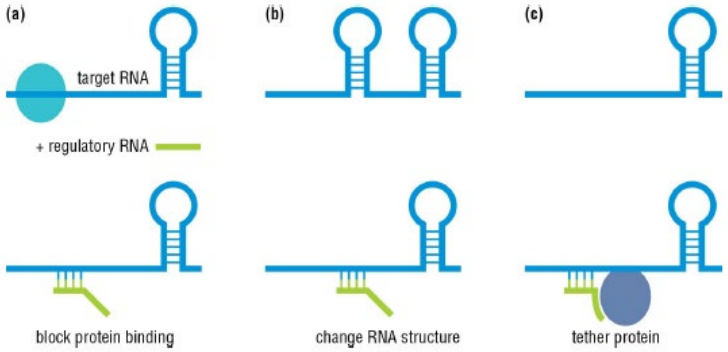
Regulatory RNAs can control RNA decay, they can change the rate of translation of a mRNA and much more, the astonishing characteristic is that they can do this in a sequence specific manner, thanks to base pair; this is extremely useful.

There exist 3 main **conserved themes** for these regulatory molecules:

1. Many transcripts of non-coding regulatory RNAs (ncRNA) are often **processed** to yield a functional product, usually a sort of maturation has to occur (especially in eukaryotes).
2. Regulatory RNAs use **base pairing** and complementarity to recognise their target (RNA or DNA), these regulators have specificity (they can also tolerate and pair with imperfect complementarity, maybe with the creation of some secondary structures).
3. Regulatory RNAs very often **interact with proteins** in order to boost their function, or they can recruit proteins in order to perform the degradation or the enhancement of the translation of a target.

Here following 3 examples of riboregualtion (target RNA in light blue, regulatory RNA in green):

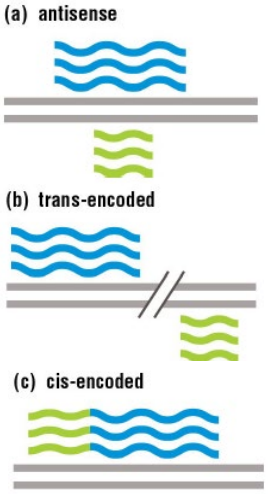
- Sometimes if we bind to a given site, we can prevent the binding of a protein, the RNA pairing could hide the initiation codon for translation or the Shine-Dalgarno sequence.
- The binding of a regulatory RNA can change the secondary structure of the target, this could resolve a secondary structure that was hiding the SD sequence.
- ncRNAs can bind with a single stranded domain to the target sequence while it is also bound through a secondary sequence to an enzyme that could be responsible for cleaving the mRNA and prevent translation (for example crispr-cas 9 or the mechanism of RNAse E).



Specificity of regulatory RNAs

Regulatory RNAs that base pair with the sequence that they regulate are produced many different and smart ways, usually they are in relation with the DNA region that they regulate exploiting it already during their transcription.

There exist 3 main types of riboregulators encoding:



a. **DNA strand antisense to the target (cis)**: the regulator RNA can be transcribed using the antisense sequence of the target, in order to produce a transcript that is already perfectly complementary.
b. **Trans-encoded**: the regulator RNA can be encoded from a completely different region of the genome, that is partially complementary by chance.
c. **Cis-encoded**: the regulator RNA can be part of the target, as in riboswitches, a change in the secondary structure of the regulatory RNA can hide or show the initiation sequence of the target RNA.

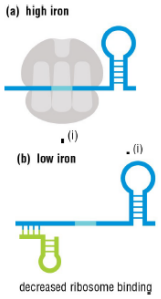
This mechanism allow a huge flexibility, RNA can bind DNA, RNA but also proteins and metabolites.

Trans-regulation

Bacterial regulatory RNAs are usually synthesized as independent 100-300 nt transcripts – small RNAs (sRNAs).

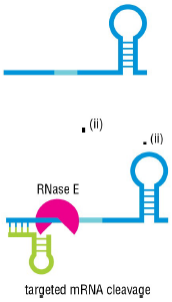
Examples of trans encoded RNAs in bacteria:

1. Regulatory RNA binds and occludes the Shine-Dalgarno sequence

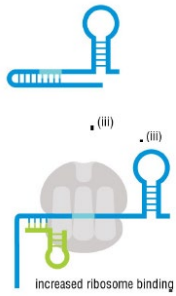


RyhB: present when there is low iron and absent in high iron

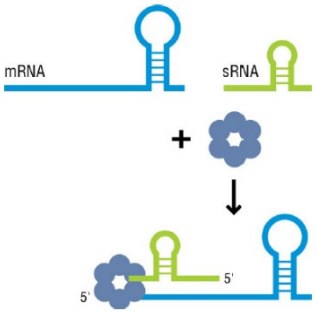
2. Small RNA can recruit RNAse E, it enhances the decay of that sequence



3. Removal of a secondary structure that was occluding the Shine Dalgarno sequence



Since in these cases the regulatory RNA pairs with low complementarity the ΔG is not high enough to have a stable association; the interaction of the regulatory RNA is therefore aided by chaperon Hfq.



Hfq is a homohexameric protein that helps the regulatory RNA to find the right sequence and bind tightly to it; it helps sRNA-mRNA interactions.

Cis-regulation

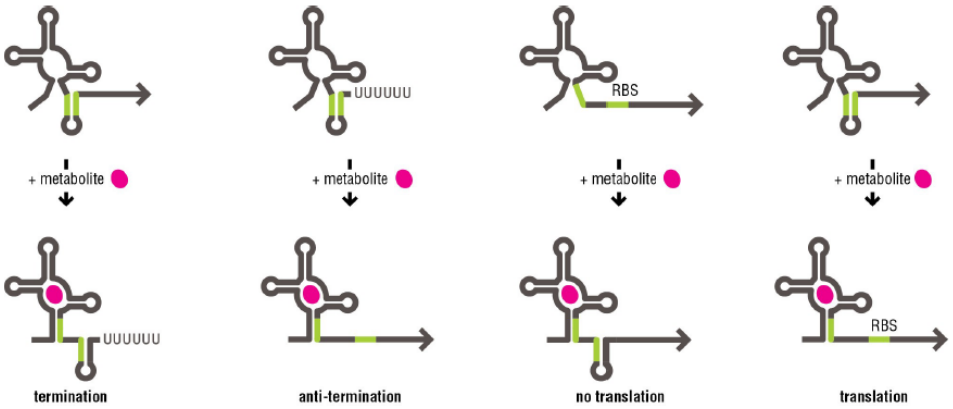
Riboswitches

Riboswitches are portions of a transcript that can directly bind a small molecule that controls the RNA secondary structure, they can regulate transcription and translation, negatively or positively.

Each riboswitch is mate of 2 portions:

4. Aptamer: binds the metabolite
5. Expression platform (effector): it controls the transcription or translation by changes in secondary structure that depend on the state of the aptamer.

Effects of riboswitches:
The effects of different riboswitches vary to promote or prevent transcription or translation. Metabolites also vary, and can include vitamins, ions, sugars, purines and others, and riboswitches can also recognise different states.



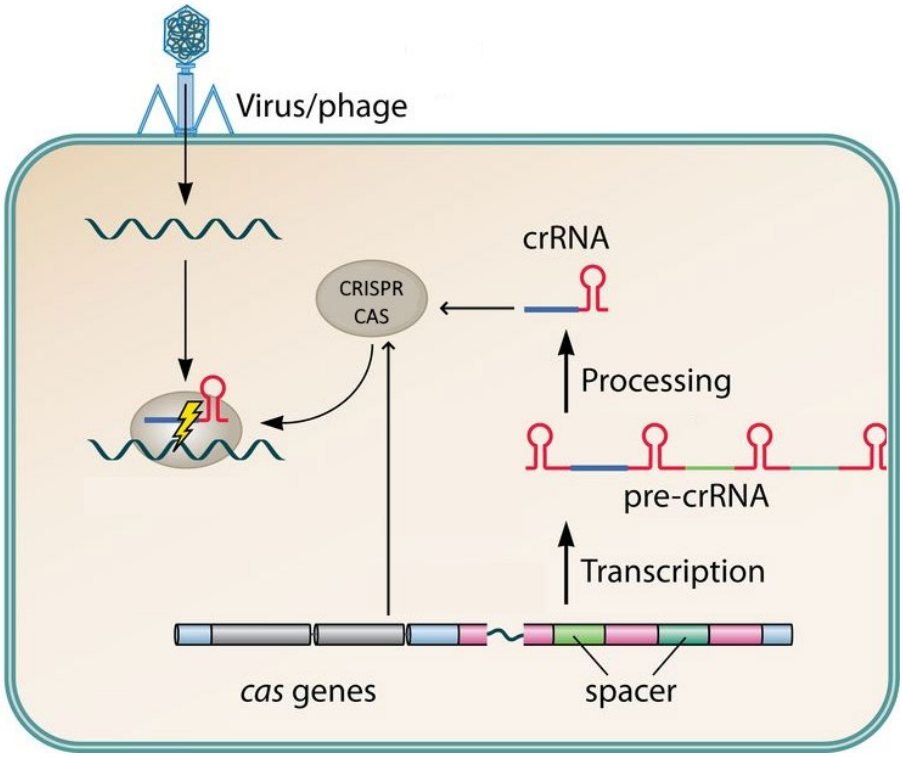
CRISPR

CRISPR is a molecular mechanism native of bacteria that uses small RNAs to drive the destruction of particular DNA sequences.

Bacteria use this mechanism as immune system, it is responsible for the degradation of viral DNA, the DNA is attacked by a complementary RNA strand (cr-RNA) that carries along an endonuclease that cleaves dsDNA.

The CRISPR RNA is encoded from CRISPR loci in the bacterial genome, these loci contain sequences of **viral genes** interspaced by **inverted repeat** sequences; associated to this locus there is an array of CAS genes.

A long transcript of CRISPR locus is synthetised (pre-crRNA), that is then maturated to form crRNA.



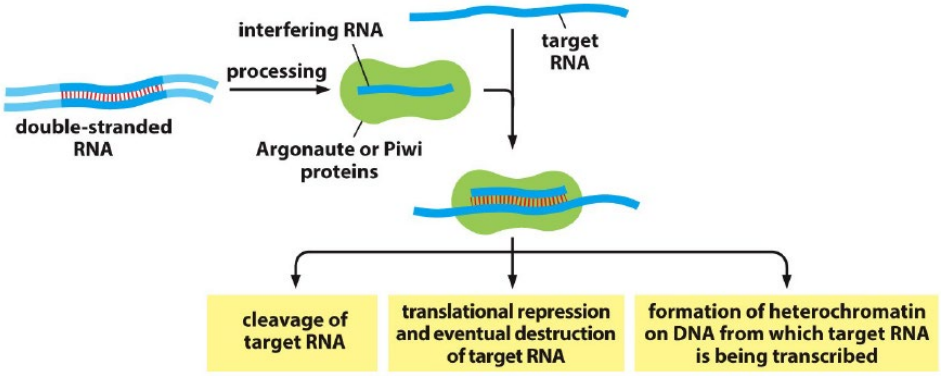
Thanks to the IR these RNA can be trimmed in specific points and each crRNA can be bound to CAS proteins, which are endonucleases that blindly cut double stranded DNA or RNA.

Once this mechanism is active whatever sequence will base pair to the CRISPR sequence it will be cut from the endonuclease associated to it.

This mechanism is adaptive and it can be expanded if new viruses attack the cell, in fact the CRISPR loci are just the history of the viral attacks that a cell has undergone.

Eukaryotic small RNAs

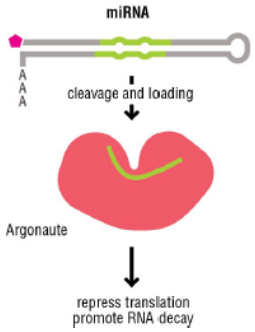
Small RNAs in eukaryotes are 20-30 nt in length and are derived from longer transcripts, as in bacteria they are used for many purposes, in which their specificity is exploited.



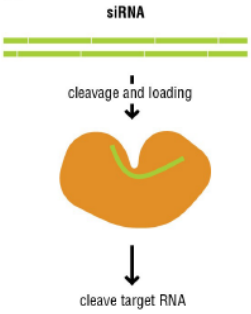
Eukaryotic sRNAs associate with Argonaute-family proteins, which facilitate interactions with the targets.

Eukaryotic sRNA are grouped into 2 main classes:

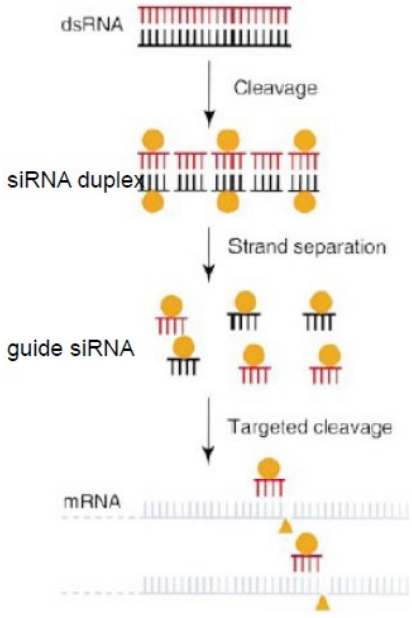
- **MicroRNAs** (miRNAs): derived from endogenous primary transcripts, generally they downregulate cytoplasmic RNAs through translational repression and mRNA decay



- **Small interfering RNAs** (siRNAs): derived from longer double stranded RNAs (such as viral or endogenous RNAs), they target RNAs for degradation and are thought to act as a cellular defence mechanism.



siRNA pathway



1. Exogenous dsRNA! Is a bad sign for the cell, it triggers the RNA degradation actuated through siRNA

2. Dicing (by dicer), it is a RNAse III type enzyme that cuts RNA and forms siRNA duplexes.
3. Association of siRNA duplexes to argounate.

4. Only one strand of each siRNA duplex is selected as a guide to slice RNA (the one with the lower stability at the 5' end)

5. Slicing, promoted by AGO protein

miRNA pathway

The endogenous precursors are produced in the nucleus and they are processed by microprocessors (RNAse III).

microRNA duplexes are formed, they can have a non-perfect complementarity, they interact with AGO, allowing base pairing with a target RNA (non-perfect complementarity).

In this case there is no slicing, they usually promote the deadenylation of mRNA to increase its decay rate.