

Molecular Biology of the Cell

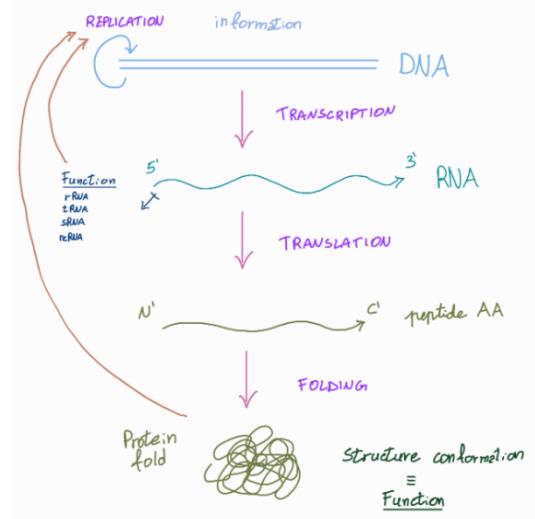
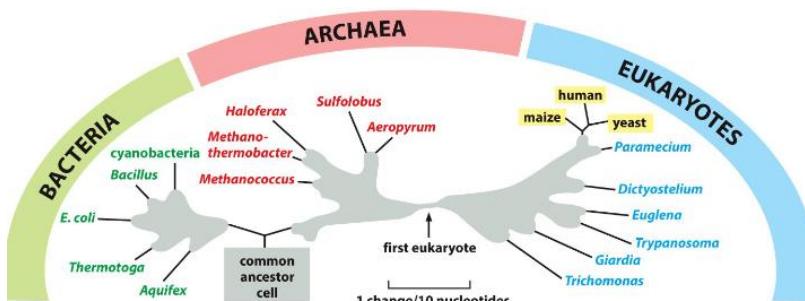
1 INTRODUCTION

Gene expression can be regulated at different levels:

- Transcriptional
- Post-transcriptional
- Translational
- Post-translational

The regulation of gene expression allows to silence or activate certain genes and to differentiate cell fate and behaviour.

Natural selection is when in a population with different versions (alleles) of a gene, all versions will persist unless, in certain circumstances, one version performs better than others.



The similarity of DNA sequences tells us how related organisms are – the more similar the sequence, the more closely related.

2 PROTEINS

2.1 SHAPE AND STRUCTURES OF PROTEINS

Amino acids are the building blocks of proteins.

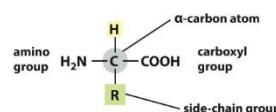
They're joined together by peptide bonds to form polypeptides.

Polypeptides are always written in the N-terminus toward the left.

The most common forms of secondary structure are alpha-helices (2 alpha helices → coiled coil) and beta-sheets.

Proteins fold in 3D shapes, it's conformation:

- Primary structure is the amino acid sequence
- Secondary structure to localized regions of repeated regular structures such as beta-strands and alpha-helices
- Tertiary structure refers to the final folded structure of a single polypeptide
- Quaternary structure refers to the combination of multiple polypeptides in a single protein (if monomeric protein, then 3' = 4')



AMINO ACID	SIDE CHAIN	AMINO ACID	SIDE CHAIN
Aspartic acid	Asp D negative	Alanine	Ala A nonpolar
Glutamic acid	Glu E negative	Glycine	Gly G nonpolar
Arginine	Arg R positive	Valine	Val V nonpolar
Lysine	Lys K positive	Leucine	Leu L nonpolar
Histidine	His H positive	Isoleucine	Ile I nonpolar
Asparagine	Asn N uncharged polar	Proline	Pro P nonpolar
Glutamine	Gln Q uncharged polar	Phenylalanine	Phe F nonpolar
Serine	Ser S uncharged polar	Methionine	Met M nonpolar
Threonine	Thr T uncharged polar	Tryptophan	Trp W nonpolar
Tyrosine	Tyr Y uncharged polar	Cysteine	Cys C nonpolar

— POLAR AMINO ACIDS —

— NONPOLAR AMINO ACIDS —

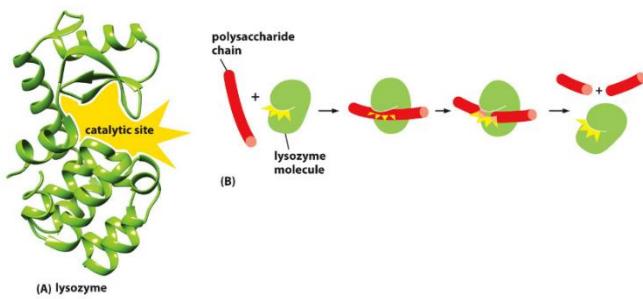
Proteins can be composed of multiple modular domains: compact regions of polypeptide that can fold on their own. These domains and folds have been conserved through evolution, so proteins can be classified into many families, each family member having an amino acid sequence and a 3D conformation resembling those of other family members. **Divergent evolution** happens when a protein has evolved a stable fold with useful properties and its structure could be modified during evolution to enable it to perform new functions.

Many examples show that proteins with more than 30% sequence identity usually share the same function. More deeply, when comparing protein sequences, it's possible to identify key amino acids vital for functioning, conservation of important amino acids is a strong indicator of similar function.

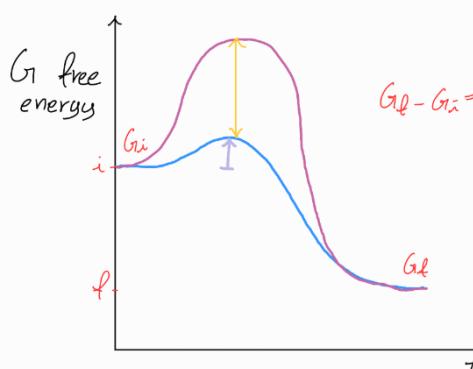
Convergence evolution instead is less common – proteins evolved not from a common ancestor that carry out similar function. Completely different folds but the catalytic site is identical: nature find the same solution.

2.2 PROTEIN FUNCTION

Proteins are the main cellular effectors; they act as enzymes to catalyse reactions that make (**anabolic**) or brake (**catabolic**) covalent bonds.



E.g. Lysozyme catalyses the cutting of polysaccharide chains in the cell walls of bacteria throughout hydrolysis between two sugar groups that causes the bond to break. A water molecule can do it only if the polysaccharide molecule is distorted into a particular shape called transition state and this is the first task accomplished by lysozyme that form a complex L + B, then it cleaves a bond throughout hydrolysis and rapidly dissociates.

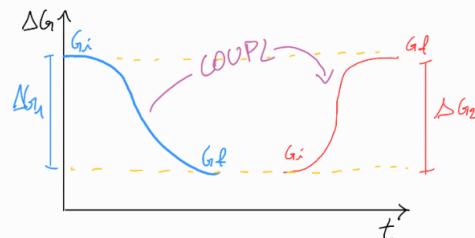


Reaction coupling says that it's possible to perform unfavourable reaction ($\Delta G > 0$) if you couple this reaction with another one with $\Delta G < 0$ and their sum is less than 0.

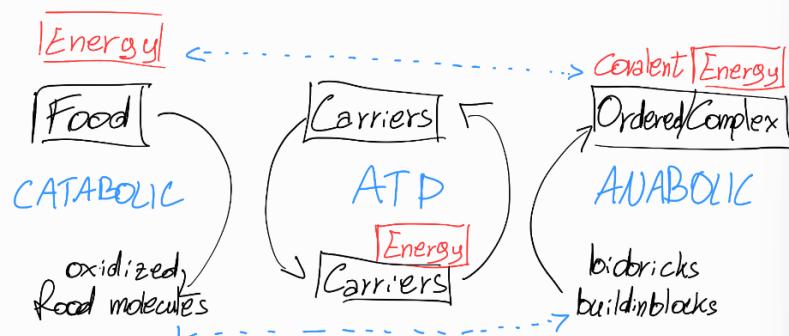
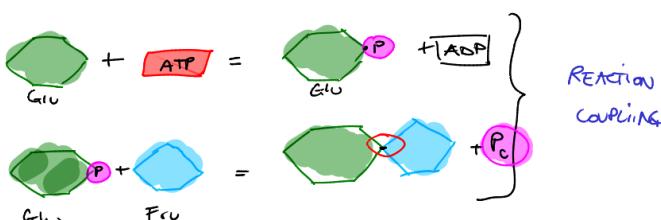
Reaction Coupling

two reactions are coupled
1. $\Delta G < 0$
2. $\Delta G > 0$

first reaction ($\Delta G > 0$) releases E needed for the second.

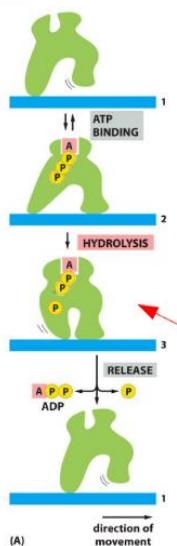
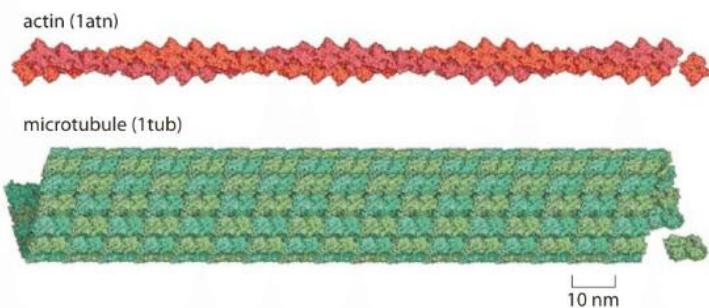


Reaction coupling uses **energy carriers**.



also have other functions:

- Maintaining cellular membrane structures



- Generating movements (molecular motors)

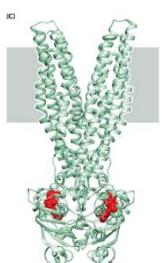
Generate the forces responsible for muscle contraction and the crawling and swimming of cells, but also power smaller-scale intracellular movements like dragging chromosomes to opposite ends. All these processes depend on proteins with moving parts, usually one is fixed to make any one of the changes in shape irreversible, this forces the entire cycle to proceed in one direction only.

Allosteric regulation:

ATP binding shifts a motor protein from conformation 1 to 2, the ATP bound is then hydrolysed to produce ADP and inorganic phosphate Pi (2 → 3), finally release of the bound ADP and Pi (3 → 1). This process allows myosin to walk along actin filaments, kinase to walk along microtubules. These movements can also be rapid like 1000 n/s. Note how protein activity is often accompanied by hydrolysis of activated carrier molecules providing energy (ATP, GTP, NADH, etc)

- Sensing, transducing, convening signals (receptors, regulators, hormones) (GAP, Ras, ...)
- Scavenging transporting binding scaffolding molecules

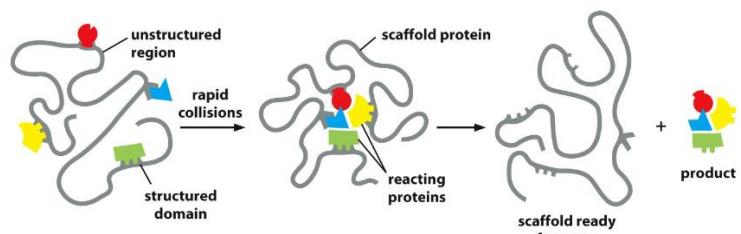
Allosteric reactions verify when a binder interact with the protein in the regulatory site (instead of the active site), inducing a conformational change.



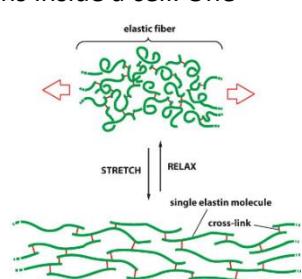
Allosteric proteins are also responsible for pumping specific ions or small molecules across a membrane.

ABC transporters (ATP-Binding Cassette transporters) are an important class of membrane-bound pump proteins, overproduction of proteins in this class contributes to the resistance of tumour cells to chemotherapeutic drugs. It's composed of 4 domains: a pair of membrane-spanning and a pair of ATP-binding domains.

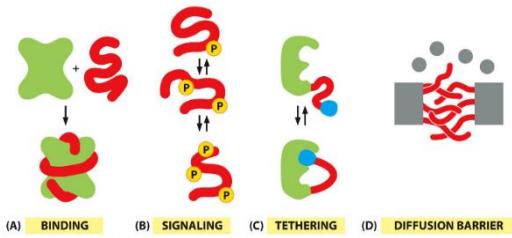
Proteins with activated receptors often recruit a set of other proteins to the inside surface of the plasma membrane to form a large protein complex that passes the signal on, these mechanisms frequently involve **scaffold proteins**. They have binding sites for multiple other proteins, and they serve both to link together specific sets of interacting proteins and to position them at specific locations inside a cell. One extreme are rigid scaffolds, and the other are large, flexible scaffold.



Like actin, many protein assemblies with elongated, fibrous shapes have structuring functions (keratin, microtubules,...).



Proteins often have loops of polypeptide chain, called **intrinsically disordered regions**, that protrude from the core region of a protein domain to bind to other molecules, they adopt a specific folded conformation only when the other molecule is bound. This characteristic of proteins is due to regions with low hydrophobicity with relatively high net charge and repeated sequences of amino acids. These proteins have one predominant function that is to form specific binding sites for other protein molecules that are of high specificity, but readily altered by protein phosphorylation, protein dephosphorylation, or any other covalent modifications that are triggered by cell signaling events.

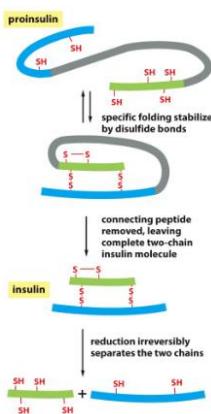


c) I. d. r. provide flexible linkers for the protein that can tether inhibitory domains.

d) A dense network of un. proteins can form a diffusion barrier (nuclear purse).

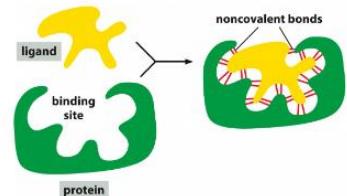
Many proteins are exposed to extracellular conditions and to help maintain their structures their polypeptide chains are often stabilized by covalent cross-linkages. These linkages can either be intraprotein or interprotein, and it is usually a **covalent sulfur-sulfur bond**. In the cytosol instead, these bonds are reduced to cysteine -SH groups.

It could be the case that part of the assembly information is provided by special enzymes and other proteins that perform the function of templates, called **assembly factors**, that guide construction but take no part in the final assembled structure. In insulin, proteolytic cleavage is an essential and irreversible step in the normal assembly process.



All proteins bind to other molecules. The substance bound to a protein is called **ligand** for that protein. In some cases this

binding is very tight in other cases it is weak and short-lived, usually made out of weak noncovalent bond and favorable hydrophobic interactions. Most of the time for a protein to work properly the binding must be **specific**: *is the property to bind selectively to one or few types of molecules out of thousand encountered in the crowded intracellular milieu or in the extracellular space*. Specificity can be determined by comparing binding constants of different interactions: $K_d^{\text{specific}} / K_d^{\text{non-specific}}$.



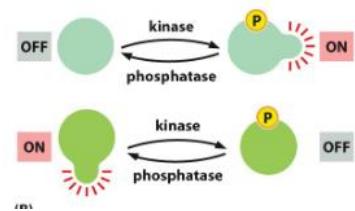
In binding reactions, specificity and affinity are related but not the same. **Affinity** is interpreted as the strength of the bond or the concentration of association (K_{on}) and dissociation (K_{off}):
binding affinity: $K_{\text{on}} / K_{\text{off}}$.

Surface conformation of a protein determines its chemistry. A ligand interacts with the protein binding site through a set of **weak non-covalent bonds** (hydrogen bonds, electrostatic interactions, Van der Waals attractions, hydrophobic interactions). We can think DNA and RNA as ligand of a protein, used to regulate gene expression through hydrophobic and polar interactions. This second ones are important because DNA is covered of minus charges because of the phospholipids at physiological pH (7.5) (under acid conditions it's protonated back again), so the protein that binds to DNA has to have proper shape, embrace the double stranded helix, has to have a positively charged cleft in which it accommodates the groove negatively charged. If the protein has to bind to a specific area in the DNA, it needs to interact with bases sequences.

This stuff is important to take in consideration to understand better (he's not gonna ask, just learn the concept of it)

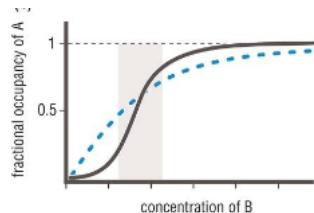
Phosphorilation in protein usually modifies its conformation that can change its affinity and specificity to ligant. It can be used to maintain homeostasis in the cell causing quick changes in protein conformation without requiring gene expression.

Phosphatase: brakes a phosphate-protein bond (also GTPase attaches the GTP molecule for conformational changes).



Proteins can have multiple signals that are integrated and converted in order to provide a protein logic processor, usually happens through allosteric regulation.

Src protein kinase gets activated only if both conditions are verified.



Cooperative binding verifies when a second molecule binds more favourably if a first molecule it's already bound (it can also be the opposite). Cooperativity is also a way to enhance specificity, in the sense that there are more steps that need to occur to get to the resulting cooperative bond and that creates higher affinity.

Proteins are often chemically modified after translation that affect the function of proteins:

- ❖ Lipid modification: needed for being recruited and attached on membranes
- ❖ Glycosylation: sugar chains attached (Ser, Thr or Asn) for protection from degradation and recognition by other proteins (by changing polarity of the protein)
- ❖ Phosphorylation: for regulation of interactions or activity
- ❖ Acetylation and Methylation: usually on histone chain that affect the chromatin
- ❖ Ubiquitination: formation of a covalent link with ubiquitin molecules that are killing signals

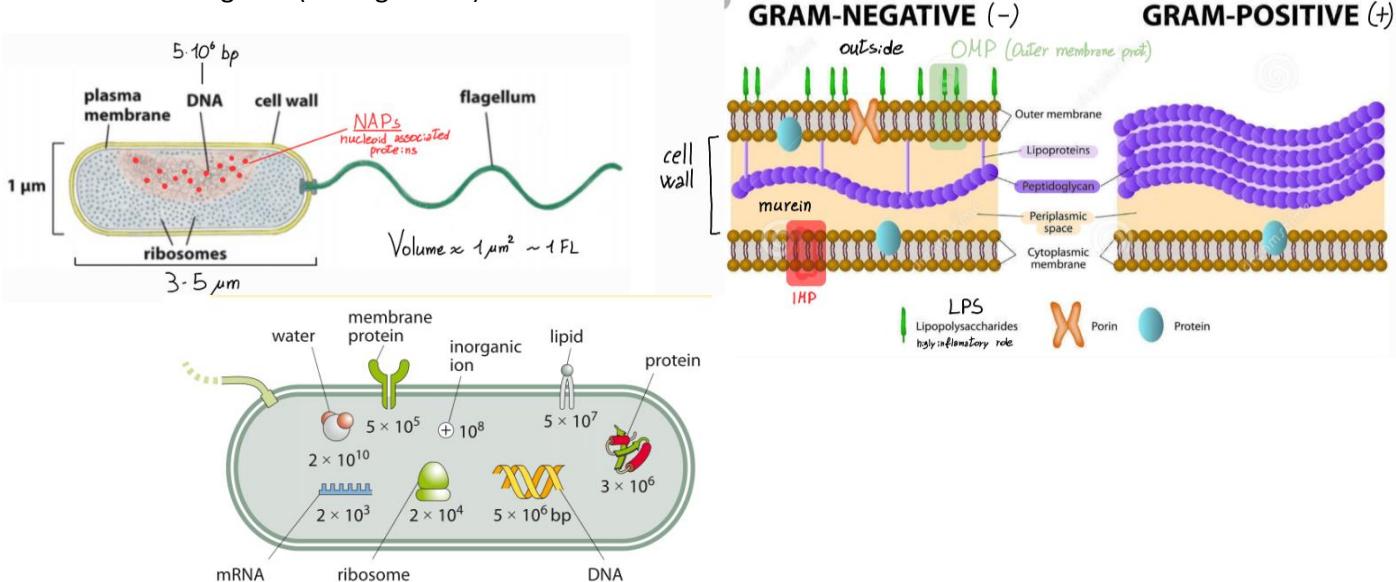
3 CELLULAR ORGANIZATION

All the living organisms in three different domains: bacteria, archaea and eukaryotes. They all behave as population.

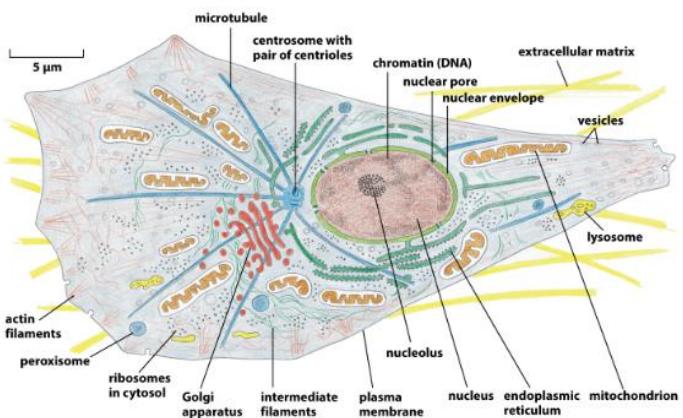
3.1 PROKARYOTES

Microbiology is the science that studies bacteria because they are in the micron size (μm).

Prokaryotes, by definition, have more internal organelles than eukaryotic cells, and their genome is about 1000 – 6000 genes (small genome).



3.2 EUKARYOTES

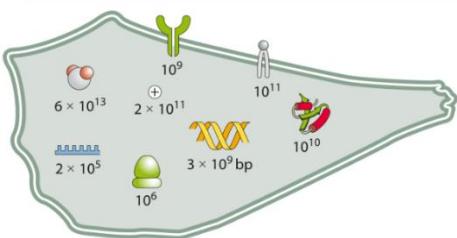


The main features of eukaryotic cells are:

- ❖ Compartmentalization is more complex than prokaryotes
- ❖ Chromosomes are found within nucleus
- ❖ Several additional membrane-bound regions (organelles) have distinct functions
- ❖ **Mitochondria** (derive energy) and **chloroplast** (harvest sunlight in plants) have their own genomes
- ❖ The ER and GA are involved in protein processing

Eukaryotic cells are separated into many distinct compartments interacting with the cytoskeleton, each of them is characterized by differences in molecular composition, ionic concentrations, membrane potential and pH. The cytoskeleton is a system of protein filaments crisscrossing the cytoplasm and forming a system of griders, ropes, and motors that gives the cell mechanical strength, controls its shape, and drives and guides its movements.

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

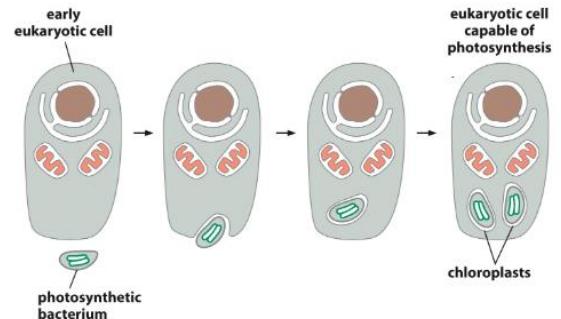


Eukaryotic cells contain **mitochondria**, small bodies in cytoplasm, enclosed by a double layer of membrane, take up oxygen and harness energy from oxidation of food molecules to produce ATP. They have similar size of bacteria, they have their own genome in the form of a circular DNA molecule, their own ribosomes and their own tRNAs. It's intuitive to understand that mitochondria originated from free-living oxygen-metabolizing (aerobic)

bacteria that were engulfed by an archaeal cell (**symbiosis**). Escaping digestion, these bacteria evolved in symbiosis with them and its progeny. Like mitochondria, chloroplasts have their own genome, they're present in plants and algae cells and they're responsible for photosynthesis.

Chloroplasts originated in a process of **endosymbiosis**, i.e. they were originally free-living cells – probably photosynthetic cyanobacteria – that were engulfed a billion years ago by cells that have become their new hosts. With time these originally distinct cells forged a tight collaboration in which most genes transferred from the engulfed cell to the host nucleus, in much the same way that the mitochondrial genome obtained its tiny size.

From genomes that probably originally contained over 3000 genes only about 130 genes remain in the chloroplasts of contemporary plants. These processes of engulfment followed by adaptation can still be observed today. Through a process known as kleptoplasty, different organisms ranging from dinoflagellates to sea slugs are able to digest algae while keeping the chloroplasts of these algae intact. These captured plastids are kept functional for months and are used to “solar power” these organisms.

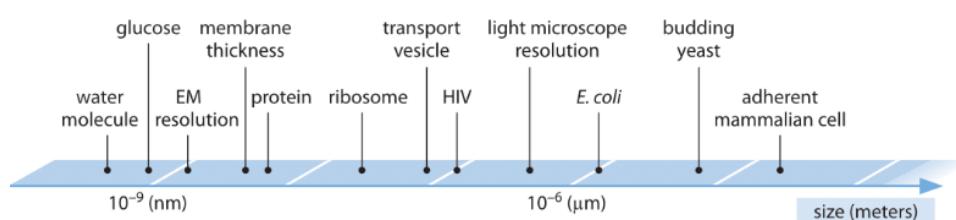


It can be seen that genes in the mitochondrial and chloroplast DNA had been moved from the symbiont genome into the DNA of the host cell nucleus.

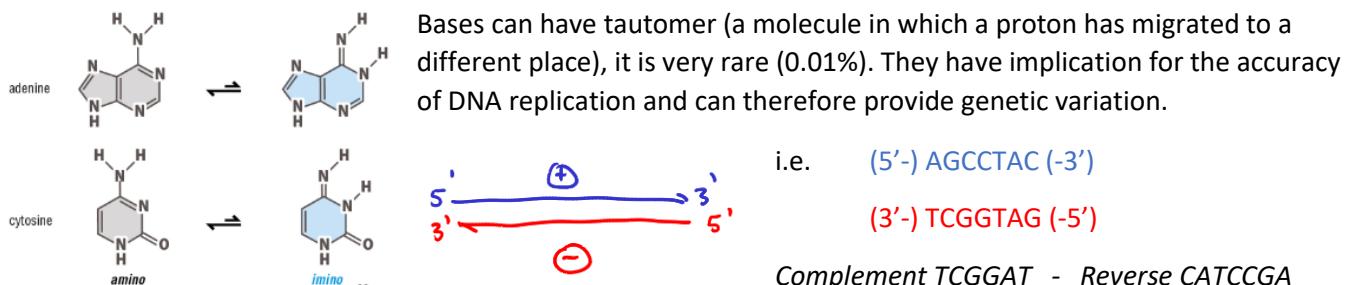
The **endoplasmic reticulum** is often the largest organelle in eukaryotic cells. Its structure is made up of a single, continuous membrane system, often spreading its cisternae and tubules across the entire cytoplasm.

- ❖ It serves as a vast processing unit for proteins, with about 20-30% of all cellular proteins passing through it as part of their maturation process.
- ❖ Produce most of the lipid that make up the cell's membranes.
- ❖ It's the main calcium deposit site in the cell, thus functioning as the crossroads for various intracellular signalling pathways.
- ❖ Its activity and size depend on the state of the cell.

A group of diverse eukaryotic, predominantly unicellular microscopic organisms, may share certain morphological and physiological characteristics with animals or plants or both. The term **protist** typically is used in reference to a eukaryote that is not a true animal, plant, or fungus or in reference to a eukaryote that lacks a multicellular stage.



4 DNA



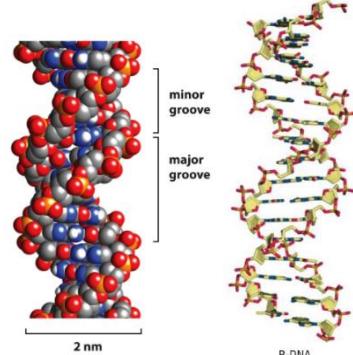
Nomenclature:

S = Strong G or C	R = Purine A or G	K = Keto G or T
W = Weak A or T (U)	Y = Pyrim. C or T (U)	M = Amino A or C

The DNA is composed of two filaments, and 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.

- ❖ Hydrophobic bases cluster in the centre
- ❖ Hydrophilic sugar phosphate backbone is on the outside
- ❖ Base pairs form a **stack** on the interior of the helix thanks to Van der Waals interactions

In **B-DNA** the helix forms two grooves of different sizes. A-T and G-C have similar widths. The B-DNA is the predominant configuration, and the helix

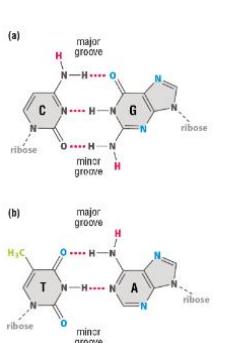


repeats every 10.5 bp that are 3.4 Å. Therefore, there are present a major groove and a minor groove.

- Major groove: is open wide and allows a easier reading

Furthermore, the sequences of bases in DNA can influence structure:

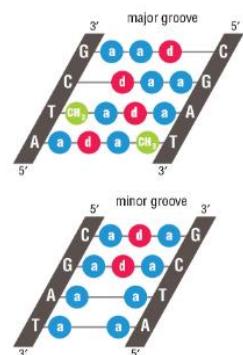
- ❖ Regions rich in A-T pairing tend to be more bendable
 - ❖ Certain base sequences have the tendency to narrow the minor groove



IMPORTANT READOUT

DNA-binding proteins most frequently bind in the **major groove**, forming non-covalent interactions with the exposed groups. Alpha helices and two-stranded beta sheets can fit in the major groove. This is **sequence readout** and it's specific.

The narrowing of the **minor groove** induced by certain base sequences can be recognized by Lys and Arg residues of certain DNA-binding proteins. This referred to as sequence-dependent **shape readout**, and it's not specific, it can only recognize if it's a A-T or G-C pair.



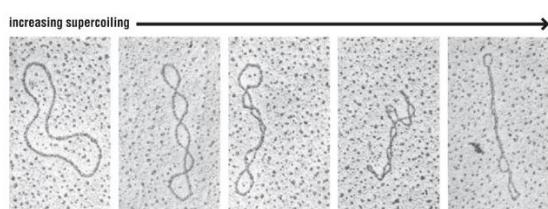
The **sequence-dependent shape readout** involves the minor groove. There are some short nucleotide sequences that narrow the minor groove and so get the major groove larger.

- ❖ This give the opportunity to the major groove to make certain interaction more likely to occur.
 - ❖ Secondly, the narrowed down minor groove accumulate electrostatic potential at the bottom of the groove and it becomes locally more negative. This negative charge can provide a specific docking environment with higher affinity for two aminoacids (Lys and Arg: long chains and positively charged).

Many DNA binding proteins that perform sequence readout in the major groove, very frequently they have Arg or Lys to interact better with the minor groove or the phosphate backbone.

A-DNA is right-handed helix, has 11 bp per turn and the grooves dimensions are almost the same. This conformation is due by DNA binding proteins. **Double stranded RNA has an A type helix.**

Supercoiling is when DNA is overwound or underwound, this allows the DNA to be compacted and stored in tiny spaces. Supercoiling occurs frequently when enzymatic reactions take place along the DNA.



(readout): duplication and transcription. To start these processes the denaturation of the DNA is needed and the process is favoured when you have underwound DNA because they're easier to separate. Some archaea living in very warm environments, have the overwound DNA so that the heat does not provide enough energy to denature the DNA.

Positive or negative supercoiling just depends on the direction in which they are wound.

When Polymerases pass on the DNA, they create conformational changes on the DNA because of the tension they create. These supercoils need to be balanced out by particular enzymes called **topoisomerases**:

- ❖ Type I Topoisomerases Generally relieve negative supercoil (unwind), usually monomers, they “nick” (cut) single stranded DNA.
 - ❖ Type II Gyrases Have 2 domains, and use ATP to cut DNA double strand. Induce negative sc DNA (overwind).

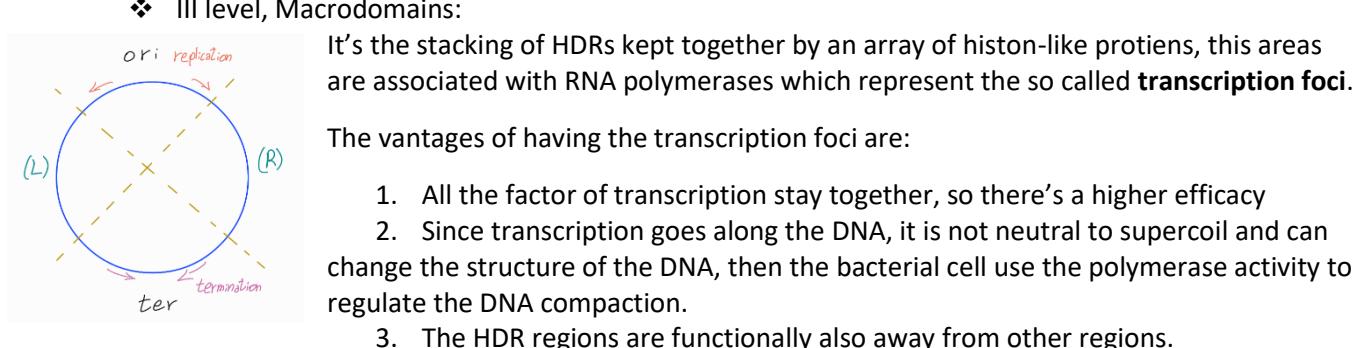
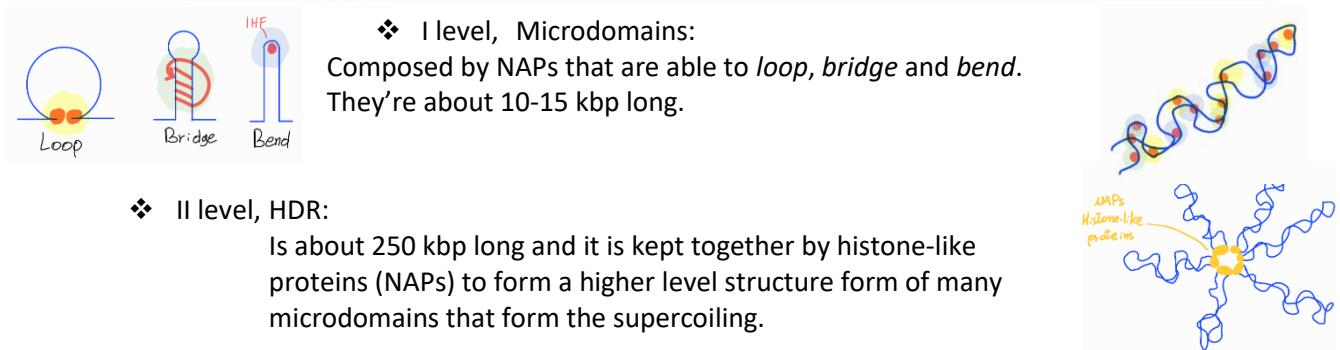
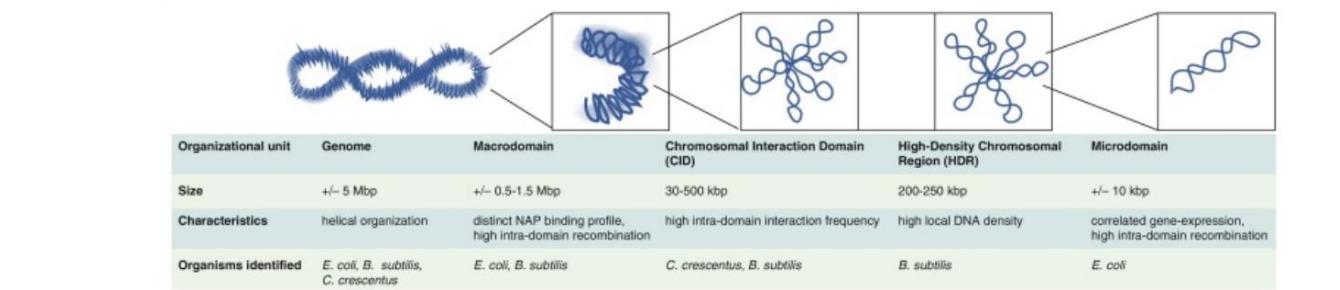
5 CHROMATIN AND CHROMOSOMAL DNA PACKAGING

The degree of compaction you need in order to fit the E. Coli genome inside its cell is calculated by considering the length of the cell ($3\mu\text{m}$), and the genome's length (calculating the bp: about 5 million, and the distance between them 0.34nm). $5 \cdot 10^6 \text{ bp} \times 0.34 \cdot 10^{-9} \text{ m/bp} \approx 1.7 \cdot 10^{-3} \text{ m}$

5.1 BACTERIA CHROMATIN

Chromatin is the packaged DNA and it's for both eukaryote and bacterial.

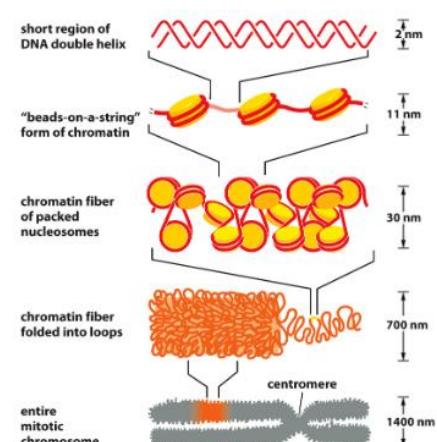
The heart of the bacterial chromatin is the nucleoid: DNA + Nucleoid associated proteins (NAPs). These proteins enable to form loops in the DNA to shape it and they participate in different levels of structuring the bacterial nucleoid.

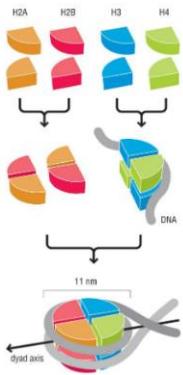


5.2 EUKARYOTIC CHROMATIN

Eukaryotic chromosomes are about the size of a bacterial cell.

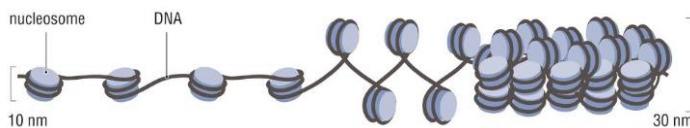
The compaction of the DNA starts from the nucleosome, which is the minimal repeated unit (**every 146 bp**). In the nucleosome is formed by an octamer of histones.





Starting from a dimer of dimers of H3 and H4, and then a dimer of dimers of H2A and H2B is bound to it. DNA exists in the same place as it got in.

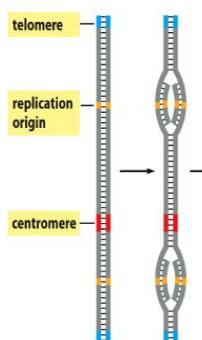
Histone tails project outwards from the nucleosome and are important for the regulation of eukaryotic chromatin. Every histone has a conformed fold that provides the core of the quaternary structure forming the nucleosome, they also have N-terminal unstructured regions that protrude out which role is to accept **post translational modifications** (acetylases – loose chromatin, deacetylases, methylated – recruiting promoters for heterochromatin, etc).



To compact the DNA in a 30nm conformation, there is the **histone H1** that binds to the linker DNA in between successive nucleosomes. These 30nm structures get to a more compacted form of 700nm that is the one visible

in the karyotype. Be aware that many times there are loops of chromatin coming out from this last compacted form.

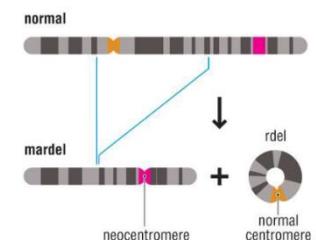
In the nucleolus there are many copies of rDNA genes to boost the production of rRNA transcript that are needed to generate the ribosomes. These sequences have many repetitions and therefore they could recombine. To avoid recombination to happen, these genes are in heterochromatic regions close to the telomeres. – Methylation is responsible for the packing of chromatin into heterochromatin, and some proteins host domains that allow the start of chain methylation reaction in adjacent nucleosomes (**spreading of heterochromatin**).



In chromosomes, telomeres and centromeres are heterochromatic, ORI are instead euchromatic. The region surrounding the centromere is called pericentromeric and it tends to be heterochromatic. However, centromere and pericentromeric regions have different histone completion (histone substitution), CENP-A is important to provide molecular marks that recruit the kinetochore proteins.

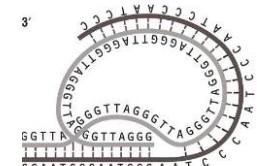
The centromere is defined by its chromatin structure, not by its DNA sequence. Alpha-satellite repetitions are usually the ones that became centromere regions when the centromere is removed.

There exist many variants of histones that are needed to carry out specific functions, like H2A.X substituting H2A, responsible for DNA damage signaling.



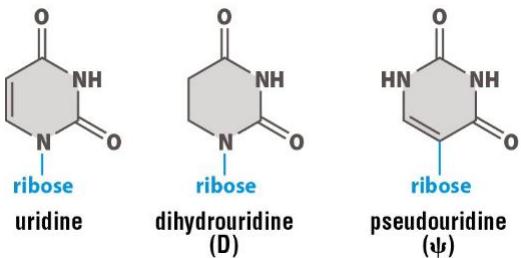
5.3 TELOMERS

Telomeric DNA in humans consist of repeats of TTAGGG and it goes up to 30 kbp long. These repeats are G-rich, and these repeats stretch out at the 3' end to allow the replication of the chromosome ends without losing chunks of the chromosome. These ends could also be dangerous for the cell because single stranded DNA is very reactive and could promote homologous recombination and very often, single stranded filaments are signs of stall replication processes. Therefore, telomeres bind proteins that protect the ends and help maintain length or bind to themselves forming a kind of loop (strand invasion).



Telomeric chromatin repeats are used as template for a particular enzyme called **telomerase**, a reverse transcriptase, that makes DNA into RNA. In this way they're able to extend the telomeres preventing the loss of chunks of DNA at the chromosome ends.

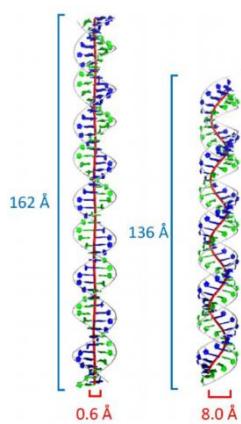
6 RNA



RNA chemical modification is frequent. Pseudouridine is very important because it is the one present in the Pfizer vaccine, making the RNA way more stable to get it translated.

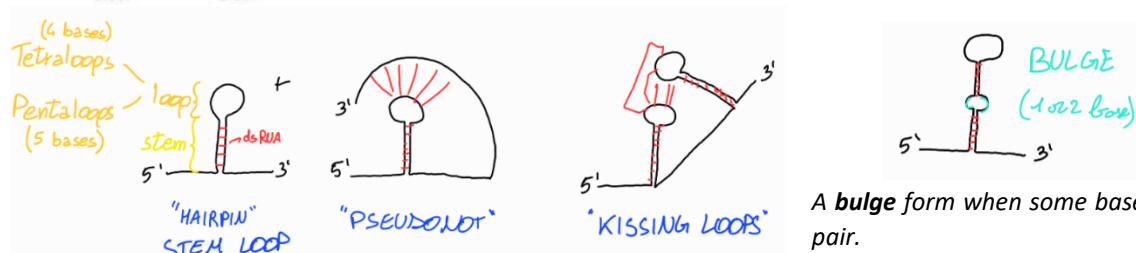
Chemical modifications are said to be permanent and not regulatory, but this is not true. There are found to be modifications in mRNA that are regulatory.

The 2' hydroxyl group in RNA favors the formation of A-type helices and double-stranded RNA, because the OH group tends to clash with the phosphate backbone. It is then very reactive and makes the RNA an unstable molecule.



The sugar pucker refers to the conformation of the ribose sugar ring that is different to the one in DNA because of the OH group. In ribose, the formation is called C3' endo, and is favored by the A-type helix. In the DNA instead there's a C2' endo. Because of the sugar pucker the dimension of the minor and major groove are more similar. This is important for the formation of a mini triple helix in the codon-anticodon base pair in the minor groove, in the presence of the ribosome. Life emerged from RNA.

Double stranded RNA could be a problem if it's longer than 40 bp, activating SOS responses especially in eukaryotic cells because many viral replication intermediates form dsRNA.

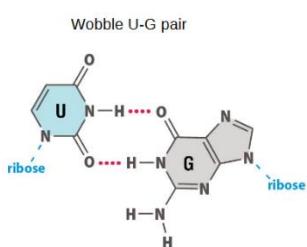


A **bulge** form when some base pairs in the stem do not pair.

RNA are complex folded molecules. Secondary structures are short double helical regions, and they can form stem-loops (require many C-G pairs in order to have enough strength to stabilize it), pseudoknot and kissing loops. These secondary structures are sensible to the laws of thermodynamics, so they can be formed or not depending on the temperature.

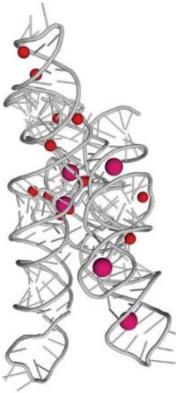
Furthermore, loops and bulges are hydrophobic and tend to stay towards inside, but modification of bases could force the structures to assume such a conformation that one or more bases are exposed to the outside of the molecule and form a specific sequence interaction (i.e., anticodon-codon interaction).

It is possible to predict the tertiary structure of RNA checking for potential base-pairing regions in an organism and looking for the same in another to verify the presence of that functional structure. But because these regions are short, this approach can falsely identify them. There could be different repeated sequences that give the same structure in different organisms, called **covariations**.



RNA could make non-canonical base pairs. Some of these base pairs are promoted by the modification of the bases and are cause of conformational stresses on the complementary strand. A **wobble base pair** is a G-U base pair in which there are 2 H bonds instead of 3, there's an offset with respect to the canonical Watson-Crick pair. Their strength is:

$$G-C > A-U > G-U$$



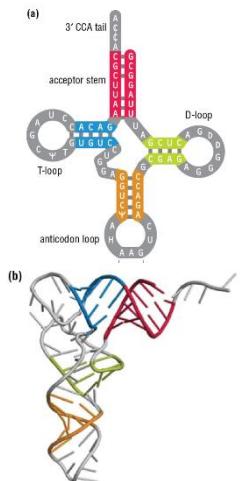
When a purine flip around its glycosidic bond it's called **Hoogsteen**. (*Do not to know the structure, just that it exists*)

Since the backbone phosphates in nucleic acids are negatively charged, they tend to repulse each other. To stabilize the molecule are then needed **Mg²⁺ cations** that bridge the negative charges of the two backbones.



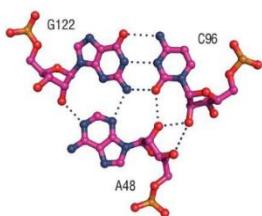
Coaxial stacking (base pair stacking like in DNA), fundamental in folding of tRNA since it's energetically favorable. This way of stacking can be seen in the interactions between the loops and a part of the filament, forming a pseudoknot.

In the loop of the orange stem, there's a H, standing for highly modified base. This modification is essential to constrict the 3 preceding nucleotides to be pointing outward from the molecule, allowing to pair the tRNA with the codon sequence of interest.



Ribose zipper is an RNA tertiary structure element in which two RNA chains are held together by hydrogen bonding interactions involving the 2' OH of ribose sugars on different strands. The 2' OH can behave as both hydrogen bond donor and acceptor, which allows formation of bifurcated hydrogen bonds with another 2' OH.

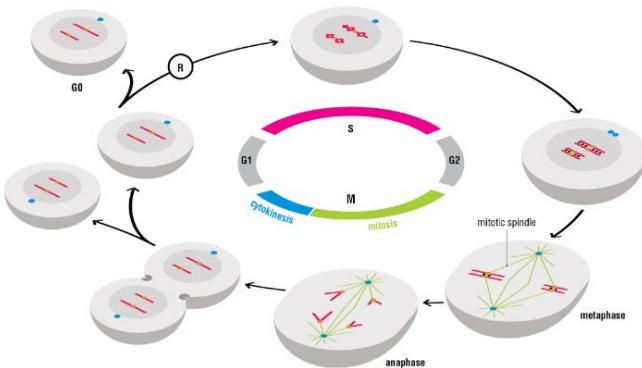
Triple and quadruple interaction are also possible in tertiary structures. Both major and minor groove are shallow and therefore another strand of RNA can enter the groove and form H bonds with the present bases. If many of these interaction happen, it could lead to



A-minor motif happens when an adenosine inserts in a double helical region, and it is stabilized by a network of H-bonds. In the decoding center of the ribosome there are two of these adenosine (very conserved), that enter in the minor groove of the codon exactly at nucleotide 1 and 2. This reading allows the ribosome to inspect the formation of the codon-anticodon interaction. If it binds correctly forming a mini triple-helix, it means that also the bases in the codon and anticodon are interacting correctly, so the ribosome can start working.

7 CELL CYCLE

Cell cycle is the process by which one cell divides to produce two daughter cells, and it's highly regulated.



Different cells have different cycles and different duration of the phases.

The most important checkpoint is after G1, the “start/restriction”, that requires enough nutrients and all the things needed to the cell for starting mitosis or meiosis. If the checkpoint is not passed, they enter in the G0 phase (often of unicellular organisms) that's a sort of resting face that can keep being until certain condition are matched.

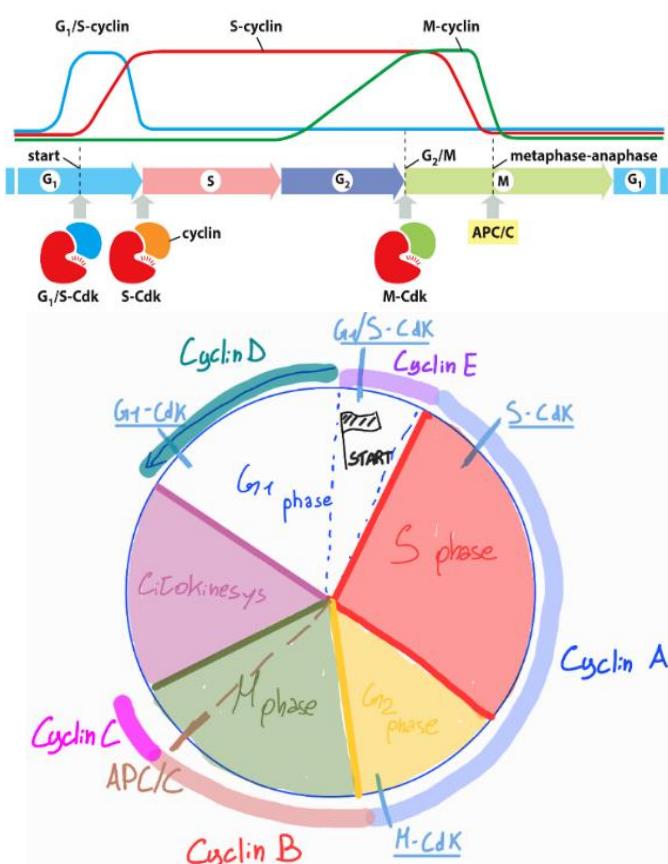
In S phase there's the synthesis of the DNA forming two sister chromatids held together by cohesion until mitosis. It's important that S phase terminate properly, in G2, there are mechanisms that check if the replication is terminated and the integrity of the chromosome.

7.1 CDKs

The cell cycle depends on effectors and enzymes called **Cdk**s (cyclin depending kinases). Different kinases pair up with different factors, which are the cyclin, to form the complex that it's characteristic for each cell phase.

The cell cycle is controlled by kinases that phosphorylate proteins that activate or deactivate interfering with one another allowing the cycle to go only in one direction (serially activation). Cyclins activate the enzymatic activity and the specificity of the kinases. The expression levels of the kinases are pretty much the same during the whole cycle, only kinases are the ones that vary.

There are four classes of cyclins, that trigger the activity of the kinase, progressing from a checkpoint:



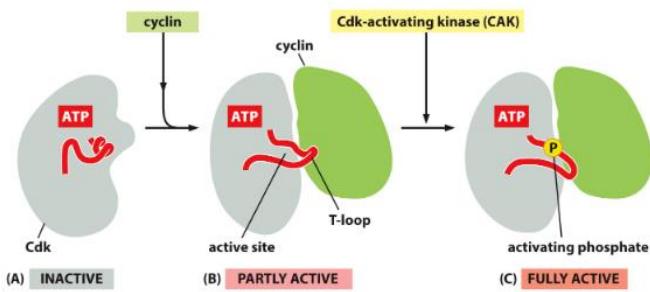
G1/S-cyclins activate the kinase in late G1 and help trigger progression through START resulting in commitment to cell-cycle entry

S-cyclins are present from S phase until the early mitotic events

M-cyclins stimulate the entry in mitosis at the G2/M transition

G1-cyclins are the ones that govern the activity of G1/S-cyclin.

Cyclin-Cdk complex	Cyclin	Cdk partner
G ₁ -Cdk	Cyclin D*	Cdk4, Cdk6
G ₁ /S-Cdk	Cyclin E	Cdk2
S-Cdk	Cyclin A	Cdk2, Cdk1**
M-Cdk	Cyclin B	Cdk1



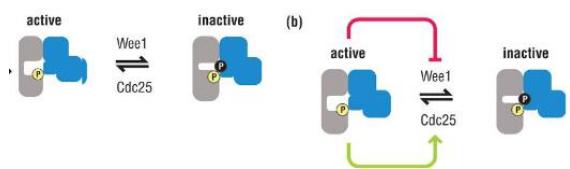
In absence of cyclin, the active site in the Cdk protein is partly obscured by a protein loop. Cyclin binding causes the loop (T-loop) to move away from the active site, partially activating the Cdk. Full activation of the cyclin-Cdk complex then occurs when a separate kinase, the Cdk-activating kinase (CAK), phosphorylates an amino acid near the entrance of the Cdk active site. This allows the kinase to phosphorylate its target proteins effectively

and bring to a cascade phosphorylation events, thereby producing specific cell-cycle events.

There is one specific case when the cell has to activate the M-Cdk (Cdk1 + M-cyclin) to enter the M phase. Once phosphorylated M-Cdk in a particular activation site, another kinase, Wee1, phosphorylate it back to inactivate it. A second enzyme involved called Cdc25, a phosphatase, remove the inhibitory phosphate activating again M-cyclin. Cdk1 is inhibitory towards Wee1 and stimulatory towards Cdc25.

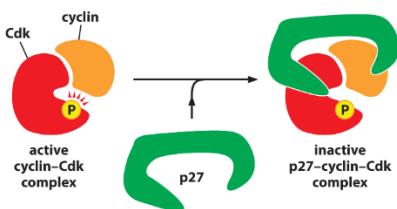
This circuit allow to control the dynamic of expression, filtering out errors and going abruptly to another phase. This can happen because Cdk1 and Cdc25 form a

positive loop, in which active Cdk1 stimulate the activation of Cdc25 that stimulate again Cdk1. On the other hand, Cdk1 has an inhibitory effect on Wee1, and Wee1 has a negative effect on Cdk1, forming another positive feedback loop.



7.2 ADDITIONAL REGULATORS

A negative loop is conducted by **CKIs**, that could bind and inactivate cyclin-Cdk complexes or just bind to the Cdks (inhibiting cycling to bond), inhibiting specific functions of the C-Cdk complex.



p27 is an inhibitor that binds to the Cyclin-Cdk complex, and it is responsible for entering the S phase. It binds to the whole complex and inhibits the ATP binding site so the molecule cannot activate.

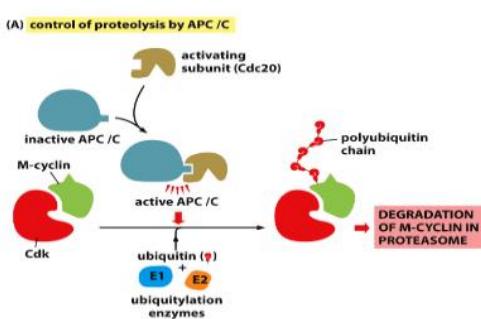
Frequently, these inhibitors need to be degraded (phosphorylation) in order to pass to the next phase. Post-translational modification (like phosphorylation and ubiquitination), control the activity and the stability of these proteins.

7.3 UBIQUITINATION

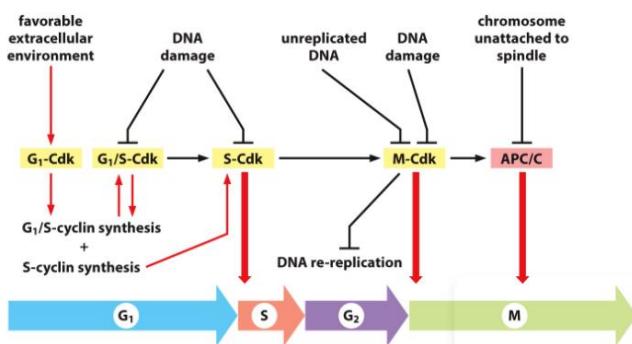
Ubiquitination is the post translational modification involving the ligation of one or more small ubiquitin molecules to the target protein. Ubiquitination promotes the degradation of the ubiquitinated protein, through the targeting proteasome.

One example of ubiquitin-dependent proteins is APC/C (factor important for the metaphase to anaphase transition). APC/C gets activated by a cofactor called Cdc20, and catalyzes the ubiquitination of:

- Securin protein important for keeping the two sister-chromatids together
- S- and M-cyclin allow the degradation of the proteins present in the previous phase



General name	Functions and comments
Protein kinases and protein phosphatases that modify Cdks	
Cdk-activating kinase (CAK)	Phosphorylates an activating site in Cdks
Wee1 kinase	Phosphorylates inhibitory sites in Cdks; primarily involved in suppressing Cdk1 activity before mitosis
Cdc25 phosphatase	Removes inhibitory phosphates from Cdks; three family members (Cdc25A, B, C) in mammals; primarily involved in controlling Cdk1 activation at the onset of mitosis
Sic1 (budding yeast)	Suppresses Cdk1 activity in G ₁ ; phosphorylation by Cdk1 at the end of G ₁ triggers its destruction
p27 (mammals)	Suppresses G ₁ /S-Cdk and S-Cdk activities in G ₁ ; helps cells withdraw from cell cycle when they terminally differentiate; phosphorylation by Cdk2 triggers its ubiquitylation by SCF
APC/C	Catalyzes ubiquitylation of regulatory proteins involved primarily in exit from mitosis, including securin and S- and M-cyclins; regulated by association with activating subunits Cdc20 or Cdh1
Cdc20	APC/C-activating subunit in all cells; triggers initial activation of APC/C at metaphase-to-anaphase transition; stimulated by M-Cdk activity

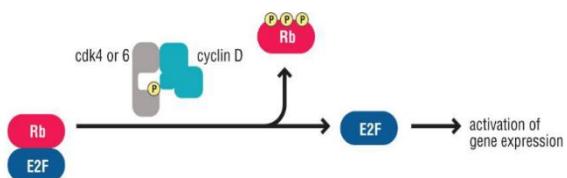


The cell cycle is a network of biochemical switching system that work in series.

At every step the cell proceeds only if certain conditions are met thanks to cell cycle checkpoints.

Cyclin-Cdks regulate the transcription of specific genes that control the cell cycle. They could either boost or alt the cell cycle or provoke suicide.

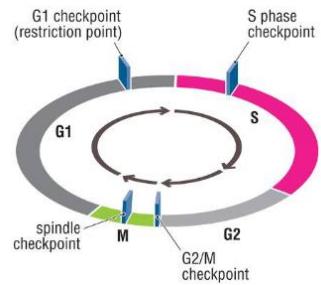
For example:



Rb is an oncosuppressor; when bound to E2F, that can promote transcription, it activates gene expression provoking the start of the cell cycle. Once G1-Cdk complex is active, it phosphorylates Rb leading to conformational change that release and activate E2F.

Rb (oncosuppressor) is a gene whose loss of function leads to cancer, so is a negative controller of the cell cycle. The opposite is an oncogene whose gain of function mutation promotes cancer (+ controller).

If a cycle is halted in a non-physiological way, the cell undergoes apoptosis.

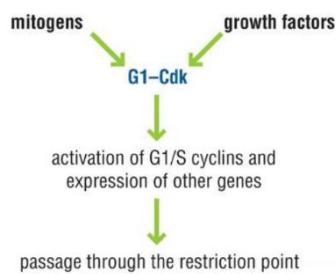


7.4 CHECKPOINTS

DNA damage checkpoint are fundamental in cell cycle (G₁/S and G₂/M). How can be check DNA damage: checking for ssDNA, that is bound to specific proteins that interact with a series of effectors that can halt the cell cycle. Also, when double stranded brakes there are some free short stretches of ssDNA that allow the recognition of the problem.

Spindle assembly checkpoint is at the climate of the cell cycle, where at the metaphase each pair of sister chromatids need to be contacted with microtubules coming from different poles at the centromere. If tension is sensed, so there are two opposite forces on the sis-chromatids, then it is possible to progress.

7.5 EXTRINSIC REGULATORS AND TOR



The first condition that must be fulfilled to start the cell cycle and progress through the restriction point involves **mitogens** and **growth factors**, and they are **extrinsic regulators** of cell cycle progression (external factors).

(Intrinsic factors i.e. nutrients enough, cell is big enough, ...)

Extrinsic regulators act on G1-Cdk, hormones stimulate the increase of cell size, and mitogens stimulate cell division.

Cell growth rate is regulated by **TOR** (kinase), increase in protein synthesis.

In high eukaryotes, TOR is activated by a signal cascade by bindings happening in receptors on the cell surface with outside stimuli.

In yeast amino acids are internalized in the cell and their presence stimulates the activation of TOR.

The growth factor receptors could be oncogene, a gain of function mutation could activate TOR without being bound by the growth factor.

EGFR (epidermal growth factor receptor, a mitogen) is usually overexpressed in cancer cell, mutation in EGFR are homologues in many cancers.

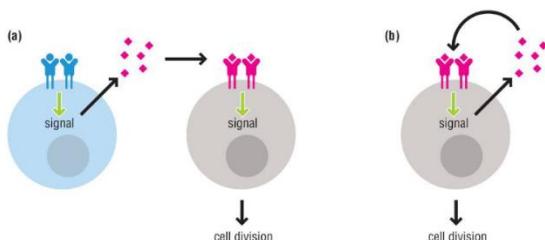
EGF is a small, skewed molecule that directly stimulate cell division by activating Cyclin-Cdk complexes. The pathway is activated by mitogen like **EGF** that binds to a cell-surface receptor that dimerize the receptor and this activate a phosphorylation cascade (signal transduction pathway).

There also exist **anti-proliferative** factors that can inhibit cell division.

TGF β is secreted by a variety of cells and binds the TGF β -receptor activating a signal transduction pathway that leads to:

- **inhibition of Rb phosphorylation** -> prevention of E2F activation -> transcriptional repression of promitotic factors
- activation of G1 and G1/s-Cdk inhibitors (CKIs)

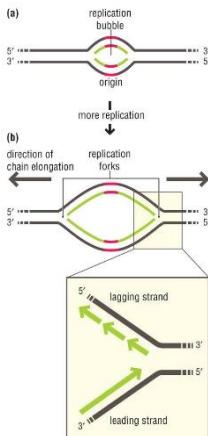
Cancer cells often lose the ability to respond to TGF β .



Oncogenes are aberrant versions of positive regulators of the cell cycle. Normally, cell proliferation would be stimulated when neighboring cells produce mitogens (a). But the mitogenic signaling pathway is often defective in cancer, which cells can aberrantly make a mitogen that stimulates their own division forming an **autocrine loop** (b).

Oncosuppressor genes are often negative regulators of the cell cycle. For example, when there's undamaged DNA, there's a MDM2 protein that can bind and ubiquitinate p53, so destroy it. If you have DNA damage, there's the activation of different factors that promote the dissociation of MDM2 and p53 by phosphorylation. P53 then is free to form multimers and interact with chromatin stimulating gene expression of genes that halt the cell cycle or promote cell death.

8 DNA REPLICATION



DNA replication is semiconservative and starts from discrete points called origin of replication (**ORI**). The process of replication moves away from the origin in both directions forming a **replication bubble**, and DNA double-helix is unwound to form **replication forks**. In the replication fork, single stranded DNA (ssDNA) is exposed, and DNA synthesis can occur in a **5' to 3'** direction. Therefore, one strand will be a **leading strand** and a **lagging strand**.

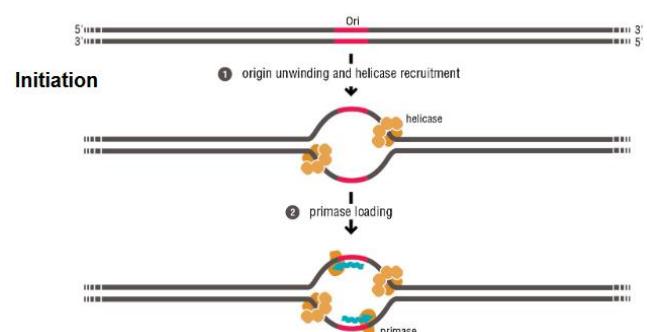
DNA replication has three phases: initiation, elongation and termination.

8.1 INITIATION

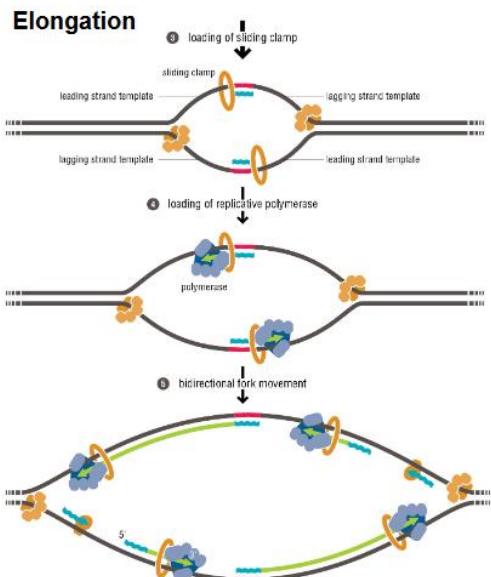
Occurs on ORIs and the first event that has to be promoted is the separation of the two strands priming the formation of the replication bubble. Secondly, the recruitment of **helicases** on top of ssDNA, that will move along with the replication fork ahead of the DNA polymerase unwinding the helix to expose ssDNA, which is coated with ssDNA binding proteins.

Replication needs to be promoted just once per cell cycle, so cells need to have a way to control the firing of origins (triggering); only some bacteria can fire ORI multiple times per cell cycle.

DNA polymerases need a **primer** to start replication, it's a short single strand of nucleotides generated in vivo by **primases**, special polymerases able to synthesize, without priming, an RNA primer. The presence of primers in the DNA filament it's a problem and it will have to be removed, so replicative DNA polymerases aren't able to start replication without a primer.



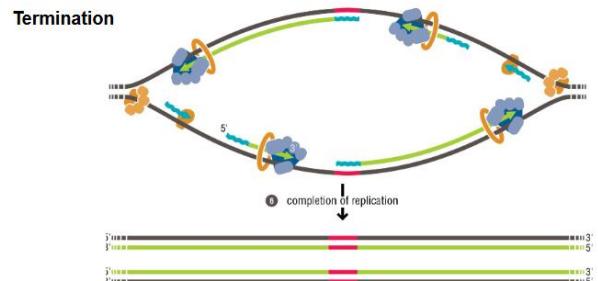
8.2 ELONGATION



Once synthesized the primer replication can start. In order to be processive (polymerase can synthesize long stretches of complementary DNA in 5' to 3' direction, processivity is the length of the fragment it can synthesize), polymerase needs to be tethered onto the DNA by the **sliding clamp** (very important and conserved in every organism). It recruits the polymerase forming a tight protein-protein interaction, so polymerase doesn't fall off the DNA.

To cast the ring of the sliding clamp around DNA (one on leading strand and one in lagging strand in each fork), the **clamp loader** cuts the ring, brings it to the right spot on the DNA and then closes it again.

Sliding clamp, helicase, polymerase primase for a complex called **replisome**.



8.3 TERMINATION

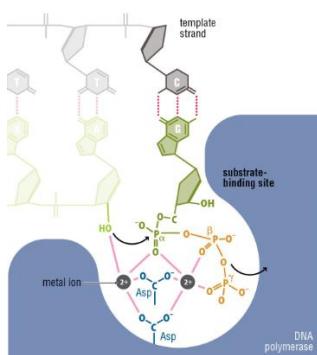
- RNA primers are removed and replaced with DNA
- DNA ligase connects adjacent strand

8.4 DNA POLYMERASE STRUCTURE AND FUNCTION

- ❖ DNA polymerases move along the DNA template strand, adding complementary bases and forming a new strand of DNA
 - ❖ Synthesis direction is $5' \rightarrow 3'$, adding nucleotides to the $3'$ OH of the previous one
 - ❖ Accuracy is fundamental, they have the ability of proofreading (checking for errors made)
 - ❖ Solve the errors, **exonucleases** activity of polymerase ($3' \rightarrow 5'$ direction)
 - ❖ They are very fast, especially in bacteria where polymerases go at a speed of 1000 nt/s (DNA pol III), in eukaryotes it's 50 nt/s (DNA pol δ and ξ)

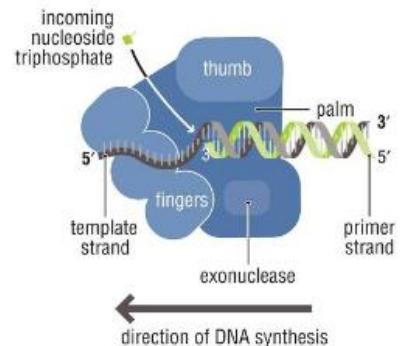
Not all of the polymerases are accurate, if the error is too big, they are not able to correct themselves but better to incorporate the wrong nucleotide if you cannot fix it rather than spending too much time on it.

- ❖ DNA polymerases have high processivity, thanks to the sliding clamp
 - ❖ It is formed by three domains: thumb, finger and palm (right hand)



8.4.1 Active site

The active site of DNA polymerase has carboxylate groups of two aspartate residues (negatively charged) and aid the coordination to Mg²⁺ ions. The active site catalyzes a phosphoryl transfer reaction linking the 5' phosphate of the incoming nucleotide to the 3' OH of the growing DNA to form a phosphodiester bond. The reaction consists of a nucleophilic attack by the 3' OH on the α-phosphate (the one closest to the 5' C) of the incoming nucleoside triphosphate, releasing two of the phosphates (β and γ) as pyrophosphate.



8.4.2 Accuracy

Proofreading activity: if the wrong nucleotide is incorporated, the speed of the polymerase decrease allowing the activation of its exonuclease activity. The mispaired end is transferred to exonuclease active site and mispaired base is excise (energetically unfavorable).

Mismatches: it does not require energy and occur during synthesis. Prevents the wrong nucleotide to be incorporated because the wrong pair form conformational stress, so there's a rejection caused by the disfavor nucleophilic attack.

8.4.3 Processivity

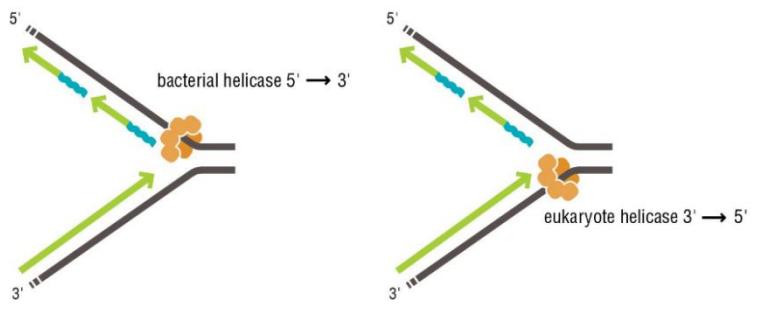
Processive synthesis means that the replicative polymerase walks long distances (thousand and thousands of nucleotides). If it would be non-processive, during synthesis it should recruit many times the polymerase with primers etc. Another example is DNA pol α which in eukaryotes synthesizes a short stretch of DNA onto the RNA primer made by the primase. Pol α is not very processive, adding only 30-100 nucleotides before dissociating from the template. RT uses RNA, not DNA

Family	Polymerase	Source	Function
RTA	Taq	<i>Thermus aquaticus</i>	replication (PCR)
B	pol α	eukaryotes	primase and repair
	pol δ	eukaryotes	replication
	pol ϵ	eukaryotes	replication
C	pol III	eukaryotes	replication
RT (RNA not DNA, important for telomeres)	reverse transcriptase	retrovirus	copy genome
		eukaryotes	copy retrotransposons
	telomerase	eukaryotes	elongate telomerase

8.5 MORE PROTEINS NEEDED FOR REPLICATION

Helicases

Unwinds the dsDNA helix for copying. Bacterial helicases slide on the lagging filament ($5' \rightarrow 3'$), eukaryotic helicases on leading strand ($3' \rightarrow 5'$).



Helicase's quaternary structure is a hexamer that form a ring. ssDNA is found at the center of it.

The helicase of *E. coli* is called **DnaB**, is a homo hexamer (composed by six identical subunits).

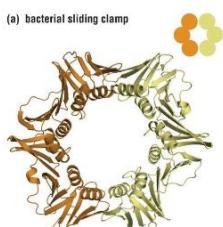
In eukaryotes and archaea, it is called the MCM complex and it's a hetero hexamer.

ssDNA-binding proteins

These proteins bind to ssDNA in order to prevent the formation of a DNA secondary structures.

Topoisomerases

Help to remove supercoils.



Sliding clamp

Bacterial one is a dimer composed by trimers.

Eukaryotic one is composed by a trimer of dimers.

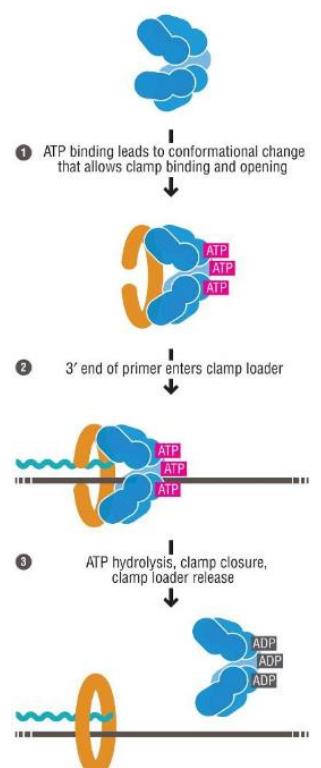


Diameter of the hole is about 35 Å that enclose the DNA and it's highly stable. It binds to the polymerase to give high processivity.

Clamp loader

Opens up the sliding clamp ring consuming energy (ATPase activity).

The mechanism starts by the activation of the clamp loader throughout an ATP being bound. This provides a conformational change that allows to bind the sliding clamp and open it. The complex now gain affinity for the 3' end of the RNA primer. This recruitment provokes ATP hydrolysis that provoke the closure of the ring and the clamp loader dissociate.

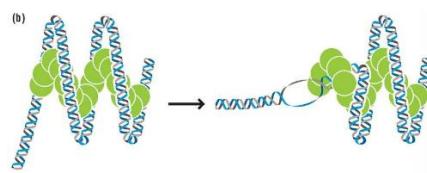
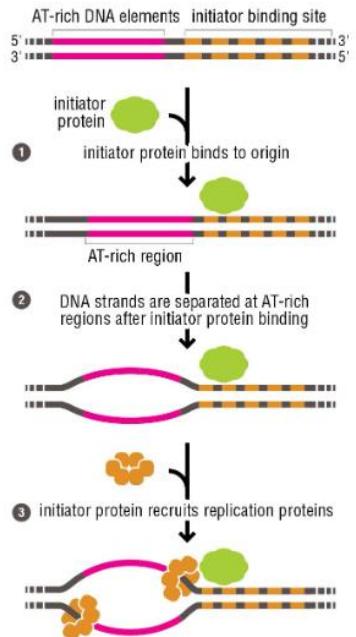


8.6 INITIATION PROCESS

Replication origins are the sites where DNA is initially unwound. The process starts with the **initiator protein** that bind to the initiator binding site, this allows helicases to bind and start unwinding the DNA. Initiator proteins are AAA+ ATP binding site proteins. (Some organisms have specific DNA sequences as origins, but it is the ability to bind initiator proteins that defines an origin).

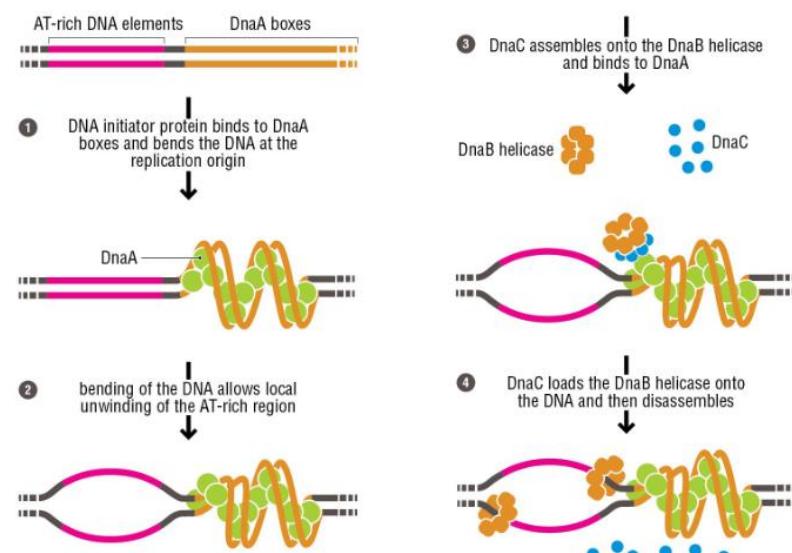
- In E. coli, the initiator is DnaA
- In eukaryotes, the Origin Recognition Complex (ORC) is the initiator
- In S. cerevisiae, ORC has 6 subunits, Orc1-6: ATP regulates initiator binding – ATP binding by Orc1 is required for ORC to bind DNA

Origins of replication sometimes have a defined DNA sequence with AT rich elements close by called **DNA unwinding element**. A and T form only two H bonds so it is easier to unwind.



8.6.1 Bacterial origins

In E. coli the origin (OriC) has a 245bp sequence, with seven 9bp DnaA boxes that bind DnaA. When bound to ATP, the AAA+ domains of DnaA multimerize into a spiral filament, which interact with DNA at the origin distorts the DNA facilitating the unwinding.

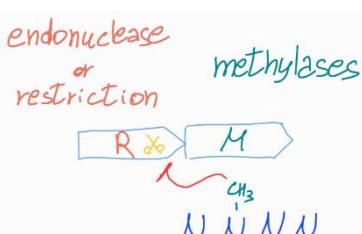


DnaA also recruits DnaB to the origin,

DnaB is a DNA helicase, loaded onto the DNA by DnaC.

After binding, DnaA unwinds DNA for replication initiation, so **Had** stimulates the hydrolysis of DnaA-ATP to ADP. This reaction makes DnaA-ADP dissociate and does not allow the re-initiation of origin firing (regulatory inactivation of DnaA, **RIDA**), preventing over-stimulation of replication initiation.

Another mechanism of control consists in DNA methylation (**dam**, DNA adenine methylase) that block re-initiation, by methylating the A residue in the sequence GATC (palindromic). This sequence is very important for bacteria and happens to be repeated 11 times at the OriC, but it is also present in the genome. It's fundamental against the phages because the methylation of DNA provides a system for the recognition of self-DNA from non-self-DNA in bacteria.



Very frequently, methyl transferases like dam, are accompanied in the same locus by a second enzyme that is a **restriction or endonuclease**, that cut the same specific sequence that interact with the methyl transferases present in the same locus. The methylation prevents the restriction enzyme from cutting that sequence, instead, the presence of non-methylated sequence is recognized by the restriction enzyme that cuts it (i.e. the virus one).

After the DNA transcription, the newly synthesized strand is not methylated (hemimethylation of DNA). Hemimethylation helps recognize the mother strand from the newly synthesized one, preventing the accidental firing of ORI: SecA binds hemimethylated GATC sites, protecting it from being fully methylated so they won't be immediately methylated by Dam and DnaA won't bind starting transcription because it can bind only to fully methylated DNA.

E. coli has nested replication OriC, this means that it's able to start another round of replication before ending the previous one.

8.6.2 Eukaryotic origins

The initiator complex is Orc, that binds to the origin and recruit factors that help the recruitment of the MCM helicase. The complex formed by these 4 proteins is called pre-replicative complex (pre-RC) that forms at the origin in G1.

Initiation of replication is achieved throughout phosphorylation of the pre-RC complex (2 helicases + Orc complex) by the S-Cdk complex.

Phosphorylated ORC stays on the origin up to the completion of DNA replication, occupying the ORI so no new ORC can bind and signal the beginning of a new replication.

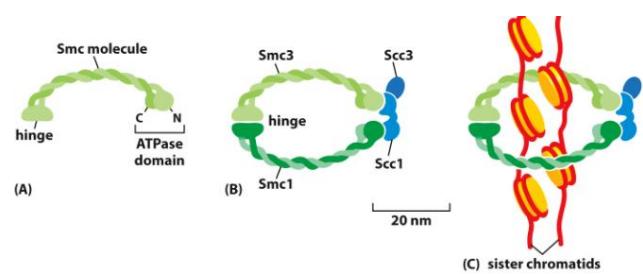
9 MITOSIS

Short memory refresher (Pinguini Migrano Al Tramonto)

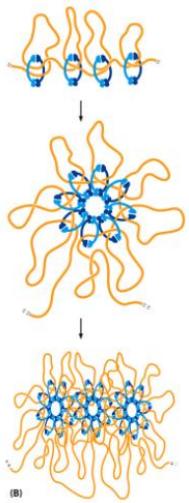
- ❖ **Prophase** Chromosomes are already replicated, they condense. Start formation of mitotic spindle and breaking down the nucleus.
- ❖ **Prometaphase** There's no nucleus and the mitotic spindle is evident. Interpolar microtubules become very visible and busy trying to align the chromosome at the metaphase plate.
- ❖ **Metaphase** It's the turning point and critical phase in the cell cycle in which chromosomes are aligned at the metaphase plate in order to segregate at the cell poles.
- ❖ **Anaphase** Chromosomes start to segregate to the spindle poles
- ❖ **Telophase** Where the chromosomes reach the spindle poles (very close to the centrioles)
- ❖ **Cytokinesis** A septum is created at the center of the cell to form the two daughters

9.1 COHESINS AND CONDENSIN

Cohesins hold sister chromatids together. They are formed of a complex of four proteins two of which are Smc proteins. These are large proteins formed each by a coiled coil with the terminal domain that has ATPase activity. Smc1 and Smc3 are bound together, and the final quaternary structure is a ring that embraces the two sister chromatids. ATPase domain is important for loading cohesin onto DNA.



The signal that triggers the entry in mitosis is M-Cdk (cyclin B + Cdk1).



Another important complex formed by Smc proteins is **condensin**, has to help configuring the mitotic chromosomes for separation. They start condensing the chromosomes, necessary for the separation of sister chromatids. Smc2 and Smc4 that form ring allowing to condense chromosome by interacting with each other by protein-protein interaction that form clusters of condense chromatin piled one onto the other.

M-Cdk has like target also condensin.

9.2 MITOTIC SPINDLE, CENTROSOME, CENTRIOLE AND KINESINS

Mitotic spindle protrudes out from the centrosome where a couple of centrioles are found. Microtubules projects out from the centrosome and are stabilized in one end by centrosome and the other reach out in the cytoplasm. Microtubules are very dynamic in M-phase.

- ❖ Interpolar microt. Projects out of the spindle pole towards the center and that eventually catch each other and interact
- ❖ Kinetochore microt. Make contact with the kinetochore
- ❖ Astral microt. They contact with the cell cortex helping the positioning of the spindle.

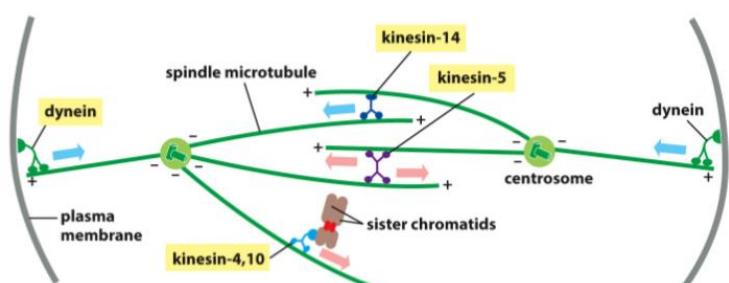
The **centrosome** is formed by a pair of **centrioles** surrounded by a pericentriolar matrix from where the microtubules point out. There are two couple of centrioles but usually cell have just one couple, so before spindle forms the cell needs to replicate the centrioles (starting from G1 ending in M).

Since tension is now created, there are different motor proteins that connect with the microtubules and help out the spindle assembly and focusing, called **kinesins**.

- ❖ Kinesin-5 has two motor domains, each of which associates with a different antiparallel interpolar microtubules. They have direction + end that pushes the two poles apart.
- ❖ Kinesin-14 has one motor domain, connects with antiparallel interpolar tubules with a – end direction movement. So it tends to pull the two poles together.
- ❖ Dynein connected with the cell cortex and have – end direction, so they push the centriole away from the cortex.
- ❖ Kinesin-4 and Kinesin-10 called also chromokinesins walk along the interpolar microtubules but has a motor domain that drives to the – end, and in the other side is attached to the chromosome.

The spindle has a self-organization in which by trials and errors they get to the result.

Completion of the mitotic spindle requires breakdown of the nuclear envelope, then microtubules dynamic instability greatly increase to explore the space and get in contact with their objective just by chance. Mitotic chromosomes are able to stabilize the microtubule, providing a favorable environment for microtubule nucleation.



Kinetochore attach sister chromatids to the spindle. It connects to the centromere (in replicated chromosomes there are two). It has a basket like structure which has a plate, and a complex of coiled coil molecules (Ndc80) that make contact with the microtubule stabilizing its + end.

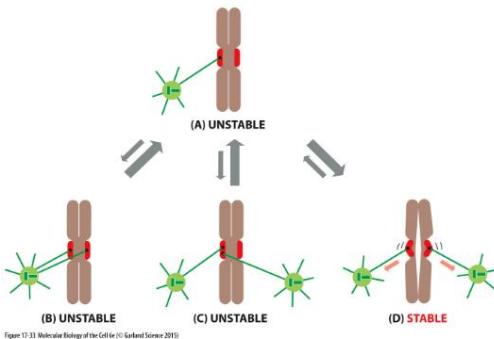


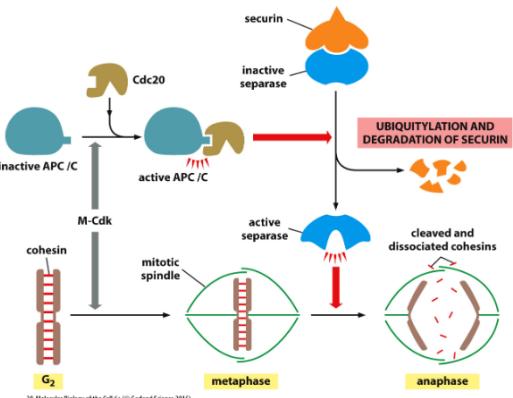
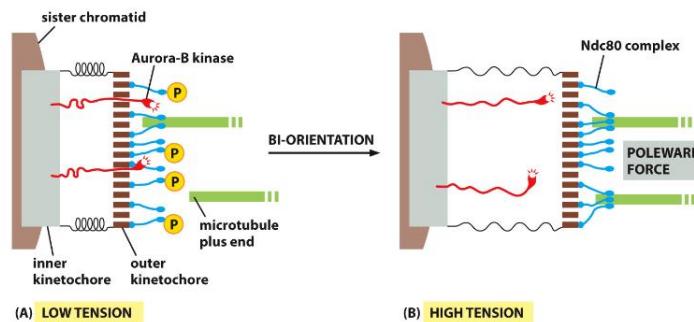
Figure 17-33 Molecular Biology of the Cell (© Garland Science 2015)

Before moving to anaphase, the cell has to check that all the kinetochores are attached by the microtubules (bi-orientation phase in metaphase), this is achieved because the kinetochores are organized in a back to back fashion in the centromere and this reduce the possibility of having the two contacted by the microtubule projecting out from one pole. The correct attachment is sensed by tension.

In order for the tension to be made cohesin is needed because it keeps the sister chromatids together. The kinetochore is formed by an inner plate connected elastically to an outer plate, a kinase called Aurora-B that has a long linker linking it to the inner plate but the enzymatic activity is carried at the outer plate. If there's high tension, the distance is higher between the inner and outer plate. If the two plates are close together the enzymatic activity of Aurora-B is outside the outer plate and phosphorylate Ndc80 not letting them bind to the microtubule.

It has been demonstrated that there are also forces applied on the chromosome arms.

The decision making for the metaphase to anaphase transition is APC/C. APC/C -(-)-| securin -(-)-| separase -(-)-| cohesin brings segregation of sister chromatids. This process is very fast because all the tension kept in equilibrium by cohesin. Mad2 together with other proteins, inhibit Cdc20 so consequently has a negative effect on APC/C, bind to unattached kinetochores because it's important to have all of them attached.



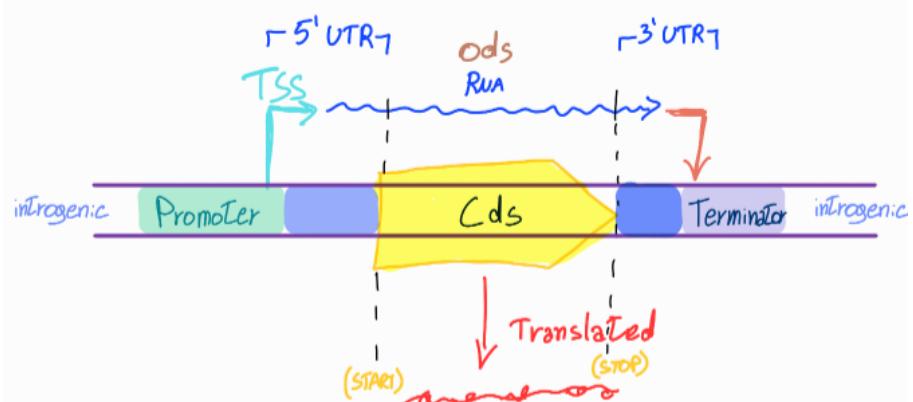
10 TRANSCRIPTION

It's the process that generates an RNA transcript from a DNA sequence.

The RNA polymerase as the DNA polymerase always synthesizes in 5' to 3' direction. The region to be transcribed have motive, specific sequences, that allow the cell to understand where to start and where to finish. The 5' Un-Translated Region and the 3' UTR are both transcribed but not translated.

The **promoter** is a nucleotide sequence in the DNA that enables the recognition of the region that must be transcribed, so it is bound by several factors including the RNA polymerase.

The first base that correspond to the incorporated first nucleotide in the transcript is called the TSS (Transcription Start Site).



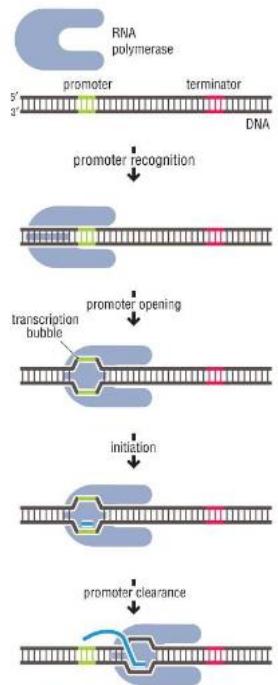
10.1 INITIATION

Initiation deal with the recognition of the promoter and the making of ssDNA template filament available. Initiation has two phases and two complexes formed by polymerase and DNA, that characterize them.

1. **Close promoter complex** Polymerase is associated with the promoter, but DNA is still ds
2. **Open promoter complex** now the **transcription bubble** is formed (14bp stretch of ssDNA that follow the RNA polymerase during the elongation process)

The open complex allows the initiation of anabolic activity of RNA polymerase. However, polymerase start generating short transcripts (2-9 nt) but it's not able to leave the promoter (**abortive initiation**). Suddenly if the RNA synthesised is sufficiently long then there's a conformational change in the polymerase and the association with the transcription bubble is stronger, increasing processivity. It's now able to leave the promoter (promotor clearance) and begin elongation.

The end of abortive initiation in bacteria is given by the sigma factor that has a loop that occludes partially the exit of the transcript from the reactive site. When the conformational change happens, the loops swing out and free the exit of the transcript, entering processive elongation.

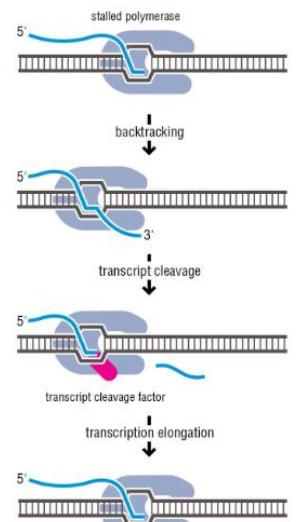


10.2 ELONGATION

Synthesis proceeds 5' to 3' (from upstream to downstream), as the polymerase move along the filament with the replication bubble. This means that there's constant unwinding of the DNA ahead and rewinding of the DNA behind it. The biochemical reaction catalysed by RNA pol is similar to DNA synthesis: successive nucleotides are added at the 3' end of the growing RNA molecule by nucleophilic attack, forming phosphodiester bond and releasing pyrophosphate (pol speed: eukaryotes: 25 nt/s, bacteria: 50 nt/s).

During the phase of processive elongation the **sigma factor is lost** because it's not needed anymore. In addition, RNA pol II in eukaryotes it's phosphorylated when it enters in elongation.

An error correction activity can be introduced when the RNA pol can pause or slow down. It can happen if there is the formation of secondary structures in the transcript that have consequence on the activity of the pol, or it could happen because of other factors that modulate the rate of elongation. For example, when a wrong nucleotide is inserted in the sequence, the pol would try to go back a couple of nucleotides, so the 3' end that was contained in the transcription bubble, protrudes out of the chore of the enzyme stimulating the endonuclease activity of the pol (cut the protruding 3' end and restart elongation).



The movement of the transcription bubble that unwinding the DNA, leads to changes in supercoiling (increase in + supercoiling ahead and - supercoiling behind). The changes in supercoiling could cause stalling of RNA pol and the tension must be relieved by topoisomerase. In E. coli, DNA gyrase removes the positive supercoils and DNA topoisomerase I removes the negative supercoils.

10.3 TERMINATION

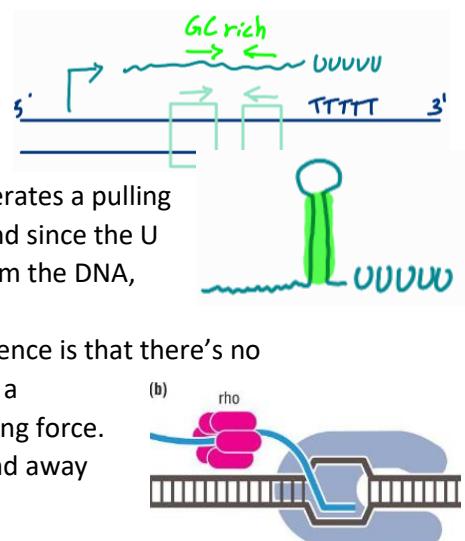
When the termination signal is encountered, and the polymerase dissociates, and the RNA filament is released from the polymerase (transcript is ready).

The two main classes of bacterial terminators are:

- ❖ **Intrinsic:** Are formed by a stem loop followed by U rich tract.

They're easily predictable because there need to be a G-C rich complementary sequence (form a stable loop, stronger pair) close by a T rich sequence. The formation of this secondary structure generates a pulling force on the RNA/DNA hybrid present in the transcription bubble, and since the U present in the hybrid has a low binding energy, it easily detaches from the DNA, terminating transcription.

- ❖ **Rho-dependent** (non-intrinsic): It's still about pulling, the only difference is that there's no formation of the hair pin but it uses the Rho terminator protein. It is a hexameric ATP-ase which bind to C-rich areas of RNA creating a pulling force. Through ATP hydrolysis drives pulling of the RNA through the ring and away from the RNA polymerase.



With respect to transcription, we will take in consideration only bacterial transcription because the eukaryotic is much more complex.

10.4 POLYMERASES

Because we have organized chromatin and nucleosomes, in eukaryotic DNA, there are many more constraints that lead to the freeing of DNA and allow transcription. Nucleosomes need to be uncased from ahead and reform them behind, along with RNA polymerase, there are nucleosome remodelling enzymes that reposition histones away from the DNA to be transcribed. Histone chaperone that captures the histone subunits and enzymes that modify the histone tail.

The prime effector in this process is RNA pol (composed by 12 domains in euk.). Archaea and bacteria have just one RNA polymerase.

Polymerases have very complex quaternary structure, in eukaryotes there are three different polymerases:

Functions of the eukaryotic RNA polymerases		Pol I and III are involved in transcription and translation
RNA polymerase	Genes transcribed	Pol II all messenger and regulatory RNA
RNA polymerase I	Ribosomal RNA (rRNA)	
RNA polymerase II	mRNA, small regulatory RNAs	
RNA polymerase III	tRNA, 5S RNA, sn RNA	

Bacterial polymerases

Are formed in general by five subunits $2\alpha \beta\beta' \omega$ (core enzyme) or $2\alpha \beta\beta' \omega \sigma$ (holoenzyme):

- 2α are both divided in two domains: N terminal (interacts with $\beta\beta'$) and C terminal is very important in transcriptional regulation because it can contact and bind to DNA, making the bind easier for a specific promoter.
- σ it's a factor that can be substituted (different sigma associated to the core). It helps the polymerase to find the specific promoter sequences. In the promoter, these sequences have motifs that are more or less conserved: the more it is conserved the tighter the pol binds (works better). Bacteria have different sigma factors: *housekeeping* (for genes that are always translated), "*heat-shock*" (housekeeping can be substituted by these to change the promoter to which pol has to bind. i.e. in order to allow blooming)

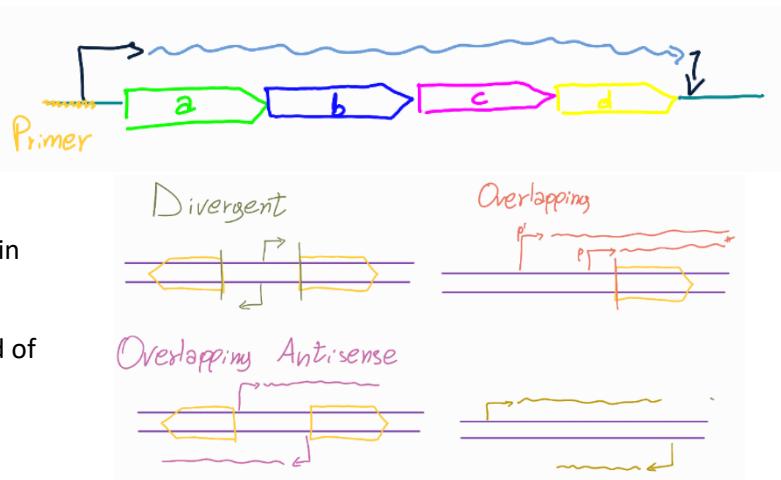
10.5 LOCUS ORGANIZATION IN BACTERIAL GENOME

Promoters and terminators are part of the information in finding the genes, so they also define the organization in the genome defining intergenic regions. From both DNA strands it's possible to generate transcripts and because bacterial genomes are small and the genetic information is tightly packed, frequently in bacteria you have transcriptional unit that contain more than one gene.

Such as multi-cistronic operon that allow more gene to be transcribed in one transcription, and the genes present are regulated at the same time because they're usually in the same pathway: *same function in same box*.

There are many situations, especially in bacteria, in which promoters can behave differently ->

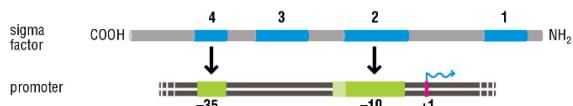
Overlapping promoters can cause dsRNA, method of regulation for the cell.



10.6 SIGMA FACTORS

It helps the polymerase to find a specific promoter sequences. In the promoter, these sequences have motifs that are more or less conserved: the more it is conserved the tighter the pol binds (works better).

Bacteria have different sigma factors: *housekeeping* (for genes that are always translated), "heat-shock" (housekeeping can be substituted by these to change the promoter to which pol has to bind. i.e. in order to allow blooming). So basically it helps switch the binding of pol from one promoter to another.



σ_v or σ^{70} is the most important, it is responsible for the recognition of two elements that define the core promoter in bacteria:
-35 box - TTGACA - and **-10 box - TATAAT -**

(Called -35 and -10 because transcription begins n nucleotides downstream from them)

Another important element in the core promoter is the distance between the two boxes and it is usually around 17bp (between 15 - 20 bp). If it's longer or shorter it's not a promoter.

The -10 element is fundamental, and deviation on the consensus are tolerated (if it's too much different they transcription rate will be low) so it is possible to define what is called the **strength of a promoter** (more similar to TATAAT -> stronger it is -> rate is higher).

This mechanism is also a "passive regulation" in which usually genes that need to be transcribed more often have stronger promoters than the others, so they can be "automatically" transcribed with a higher rate than the others.

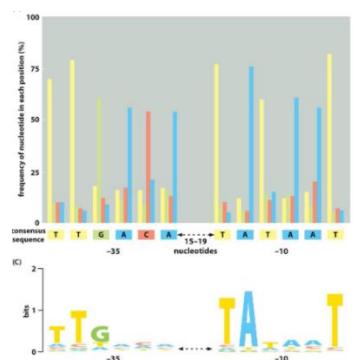
The bigger the letter the more important they are in the recognition of the boxes.

In -10 the most fundamental nucleotides are the first two and the last (**TATAAT**)

In -35 instead are the first three nucleotides (**TTGACA**)

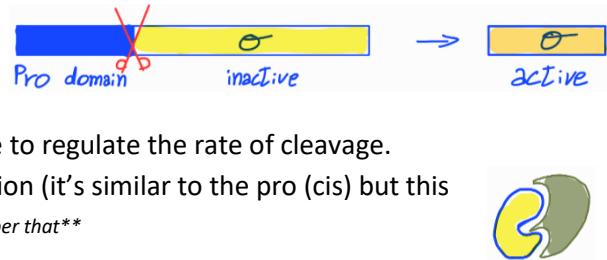
The other letters are still important but not that much, it's still easy for the sigma factor to recognize the boxes.

Sometimes, in some bacteria promoters there are extended -10 elements that present a **TGnTATAAT**, that provide further boost for transcription because it has a more strength bound.



One way of regulating gene expression is to regulate sigma. There are two regulatory path that act on sigma:

- **Pro- σ factor** it's a domain that needs to be cleaved off in order to make the σ active (post-translational regulation). In this case it's possible to regulate the rate of cleavage.
- **anti- σ factor** are proteins that bind σ inhibiting its action (it's similar to the pro (cis) but this one is in trans) ***idk what he meant exactly but just remember that***



An example of anti- σ is **FlgM**, that is involved in the regulation of the genes that make up the flagellum of the bacteria. In the process of making the flagellum there are early flagellar genes and late flagellar genes, that need to be activated in order. First the early flagellar genes that make a puzzle body that inserts in the cell wall of E. coli, and only after the late ones can start being transcribed.

Early genes are expressed by σ^70 , among the genes expressed there are also σ^F (transcribe late flagellar genes) and FlgM (inhibitor of σ^F). The puzzle body, when formed, form a sort of pore that allows FlgM to escape outside the membrane, freeing σ^F that is now allowed to bind to pol and start transcribing late flagellar genes.

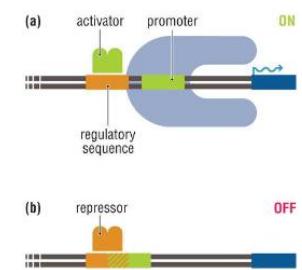
11 TRANSCRIPTION REGULATION

Transcriptional regulation is key to providing the cell with the correct amount of gene product at the correct time.

The regulatory machine is governed by activators and repressor proteins. The first level is constitutive, it's about the promoter strength. After that, there's at transcription regulation (most prevalent), at elongation or termination and at the level of the transcribed RNA itself.

Regulatory elements (operator sites in bacteria) are the sites are usually very close to the promoter (**cis**) or further away in some cases (**trans**), and they can be bound by activators or repressors.

In bacteria, the regulatory sequence is part of the promoter but not of the core promoter (-35 and -10 elements), the bind of the activator promote the recruitment of the pol. Instead the regulatory sequence that bind a repressor is usually in the promoter site, preventing the binding of the sigma factor.



If the conformation of the binding regulator changes, it also changes its DNA binding affinity and therefore regulate transcription.

Trans regulation can happen because of NAPs (nucleoid associated proteins) ability to induce looping in the DNA, so a sequence far away can be put in close contact with another distant region favoring protein-protein interaction from the **enhancer** (regulatory domain) to the promoter one. It's considered in trans when the distance between promoter and regulatory sequence is >1kb.

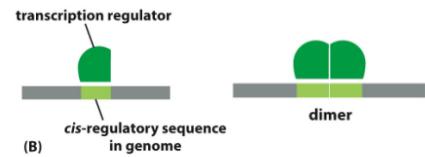
Enhancers are distal regulatory sequences that are more present in eukaryotic genomes, and usually have a transcription boost function. They can be either upstream or downstream.

In eukaryotes, instead of having activator or repressor, there are co-activators and co-repressors. They cannot bind DNA alone but are recruited by specific regulatory proteins.



To recruit regulators to the specific sites, they have sequence specific DNA readout or sequence dependent shape readout.

Since it would be very likely to have a DNA sequence present by chance in the genome if the consensus sequence is quite short, very frequently in bacteria, the regulator dimerizes. **Dimerization of transcription regulators** increase their affinity and specificity for DNA because the likelihood of having the two consensus sequence close together is lower.

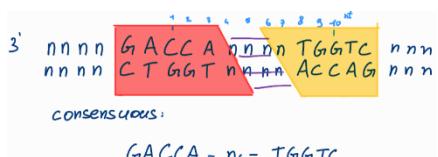


The cooperative binding gives an ALL-or-NONE response compared to the non-cooperative one.

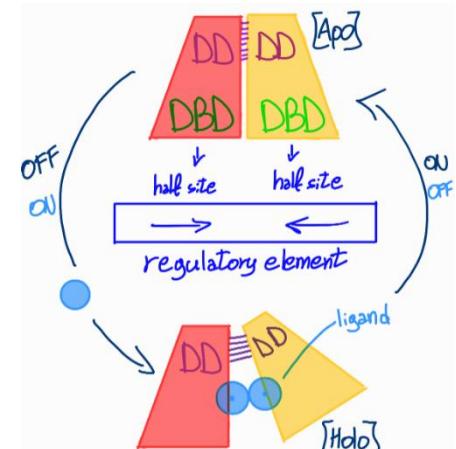
Allostery and cooperativity are very important in transcriptional regulation

The space between -35 and -10 is about 15/20 nucleotides, because it's more or less the dimension of a helix turn so the contact with the protein is made on the same side of the DNA.

Taking in consideration a regulator that is formed by two monomers forming a homodimer. Each monomer has two domains, one is the DNA binding domain (DBD) and the second is the dimerization domain (DD).

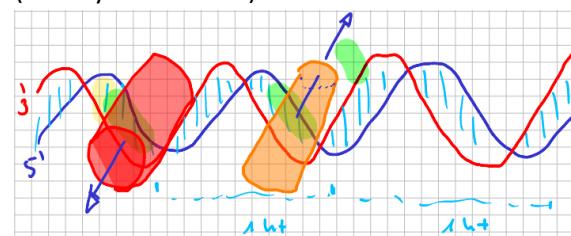


Allotropy permits to the regulator to respond to a given signal and bind with higher affinity if the given signal is present, this makes the regulation possible and gives the possibility to switch the regulator in the apo (without the ligand) / holo (with the ligand) conformation.



So, frequently these dimers bind half-sites that are separated by about 10 nt and the binding of one half-site favors the binding of the other (cooperativity).

The binding of the half-sites happens thanks to secondary structures (usually alpha-helices) that can bind on the major groove. The two half-sites are 10 nt distant from one another to allow an helical turn of the DNA. If this conformation is not met, the dimer cannot fit into the DNA structure.

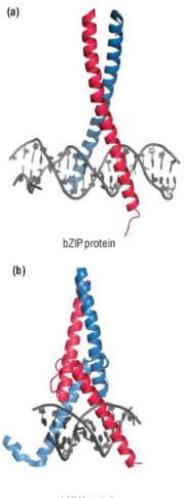


Different domains used to bind DNA happen to be very conserved.

One of the most common ones is **helix-turn-helix** (HTH) domain, formed by three alpha helices. One of them is the recognition helix, responsible for sequence readout since it has the perfect diameter to fit into the major groove. The rest of the domains form a structure to allow the recognition helices to be 3.4 nm apart from each other (major groove turn).

An example of DNA-binding domain is the **homeodomain**, very common in eukaryotes. It has the possibility to form H bonds and hydrophobic interaction with the bases and also ionic interaction with the backbone (non-specific). Homeodomains also have an N-terminal arm which has a long sidechain with a positively charged residue that enters deeply the minor groove.

Another common DNA binding class are the **zinc fingers**, small domains constituted by two beta strands antiparallel (like a mini beta sheets), which carry two cysteines, and an alpha helix. It's kept in place by a zinc atom coordinated by the two cysteines and two histidine found in the alpha helix. This domain provides a scaffold and it's possible to change the residues on the alpha helix then it's possible to readout different sequences of nucleotides.



Basic region-leucine zipper and helix-loop-helix are other two DNA binding domains that are frequently found in **eukaryotes**.

In the leucine zipper there are two coiled coils that form tweezers, they have a binding domain provided by alpha helices, which have **bZIP** region which is rich in leucine that interact with each other because are hydrophobic (dimerization domain), and another region positively charged that interact with the DNA backbone. It's easy to have different monomers binding together thanks to the hydrophobicity property, allowing different readout.

In some cases it's possible to have a **bHLH** (basic region helix-loop-helix), and the only difference is that there's a four-helix bundle that interrupt the two helix-binding domains.

In some rare cases alpha helices are not used. **Beta sheets** and loops can also mediate DNA recognition. It is not so common because loops and beta sheets aren't necessarily stable, and loops are not reliable to carry out the readout job.

Bacterial transcriptional regulator examples

Promoter occlusion

Trp Repressor (Tryptophan) found in *E. coli*. It works by **promoter occlusion** and it is a Helix-turn-helix protein (homodimer). Since the several genes needed to synthesize tryptophan are located in an **operon**, they are regulated and transcribed together.

If the tryptophan is available in the cell, the Trp repressor comes in action. From his apo-form (not able to bind and obstacle RNA pol), it binds to tryptophan (**co-repressor**) forming the holo-repressor that has a higher affinity to DNA and binds to the -10 element of the *chore* promoter.

Tryptophan \rightarrow *Trp repressor* \rightarrow *RNA pol*

Transcription activators

Catabolite activator protein (CAP) doesn't occlude the *chore* promoter, so the regulatory element doesn't overlap to the *chore* promoter. CAP regulates through two mechanisms: class I promoter and class II promoter.

CAP is able to bind to CRP (cAMP receptor protein, **co-activator**) and it does bind when carbon sources are low because if glucose is low then the level of cAMP increases.

glucose \rightarrow *cAMP* \rightarrow *CAP* \rightarrow *RNA pol*

So, one regulator binds overlapping the *chore* promoter (repressor), the other binds flanking the promoter and makes protein-protein interaction to recruit the RNA pol (activator).

In **class I promoter**, the CAP operator is upstream of the -35 box and can reach -100 nucleotides distance from it. It makes contact with the α CTD (α terminal domain of α that is part of the holoenzyme of the polymerase $2\alpha\beta\beta'\omega+\sigma$), leading the contact with the pol to the promoter.

In **class II promoter**, CAP box overlaps the -35 box. So σ domains bind weakly to the -35, but there are protein-protein interaction between pol and CAP that help out the recruitment of the pol onto the promoter.

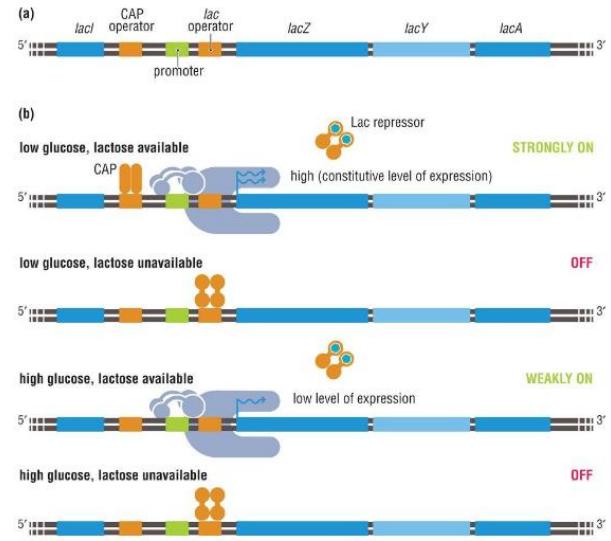
Two examples of repressor and activator that work if they are in holo-state

Both these mechanism act on the recognition of the chore promoter -10 and -35.



It works with the cooperation of both LacI and CAP. CAP help the RNA pol to bind with higher affinity when glucose is low, and LacI prevent the binding of RNA pol if the lactose is not available. When it is available, instead, the LacI repressor binds with allolactose in the holo-state so it leaves the DNA and let the pol bind.

Signal integration: different combination of signals changes the expression of a gene and it doesn't necessarily have to be only ON and OFF.



MerR regulators family (activators)

MerR regulators family influences the recruitment of polymerase because it makes easier the recognition of -10 and -35 boxes changing the topology of DNA. It binds to the DNA and bend it in a way in which the two boxes are like if they were 17-bp distant (perfect distance to be recognized by σ).

Regulation by looping

This example shows the effect of a bacteria enhancer.

NtrC is important for the nitrogen metabolism, and it regulates the RNA pol that uses σ^{54} (**alternative**), used to regulate other set of genes when there is the need. NtrC doesn't bind to DNA, however, when it gets phosphorylated the binding to DNA (on the enhancer) and multimerize. The σ^{54} already sits on the close promoter complex (DNA is still not open), so it needs the signal from NtrC, that reach it by looping and stimulates, by protein-protein interaction, the ATP hydrolysis (favorably delta G). The hydrolysis energy is used to unwind the dsDNA and start transcription.

araC is another example of looping (not to go too much in deep)

Signal transduction pathway

Process through which a signal or a stimulus is transduced into a response (from input to output). If the signal that needs to transpose it's a membrane permeant molecule, it can go through the membrane and bind the regulator. But very frequently the signal cannot go through the membrane, so they need either a transporter or other systems that transduce the signal across the membrane.

In bacteria a response regulator in the cytoplasm is composed of two domains: a transcriptional regulatory one and an effector domain. This protein changes its DNA binding affinity accordingly to its phosphorylation state.

A second component called sensor kinase and is usually a transmembrane associated protein and it's still composed by two domain: sensor domain (intracellular), kinase domain (extracellular). The first is

responsible for receiving the signal that will give a conformational change to it and kinase domain auto-phosphorylates. It then became active, and it can phosphorylate the regulatory domain of the response regulator.

Attenuation

Belongs only to bacteria because it also requires translation to occur. The levels of transcription are controlled by the rate of translation of that exact transcript.

Translation of a messenger that govern the transcription of the same messenger.

It is possible because in bacteria there's no separation between DNA and cytoplasm. So, transcription and translation in bacteria happen together and this process is called transcriptional attenuation.

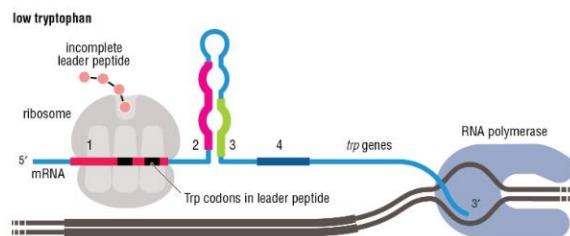
An example of attenuation is the regulation of the **Trp operon gene**, that involves translation and the formation of alternative secondary structures in mRNA.

The whole locus is transcribed, and the mRNA has a **leader sequence**, sequence that is translated and is found at the start of a transcript (but usually are not translated and are present only to confer some abilities). Its precise function is modulating or controlling the speed of translation.

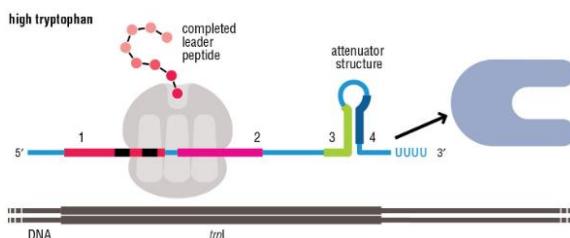
the leader sequence is able to slow down the translating ribosome: inside the leader sequence you find two codons one after the other for Trp, Trp is a very rare amino acid, so finding two codons for Trp in a row is rare. Because Trp is rare, there are few tRNAs needed to translate and form the protein.

On top of that there are three different regions in the mRNA that can form alternative stem loops, so in particular it is possible to have 2 situations:

- pairing 2-3 it prevents the formation of the hairpin (3-4 pairing) so there's not the presence of the terminator, so transcription can continue.



- **pairing 3-4** that form an intrinsic transcriptional terminator → transcription terminated

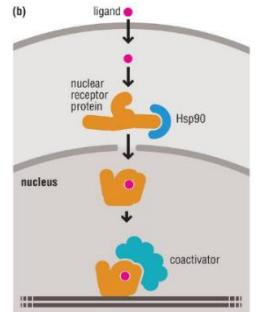


In case of low tryptophan level, the transcription of the Trp gene should be on. And the opposite. When tryptophan levels are high, the ribosome translating the mRNA goes faster and doesn't allow the formation of the stem-loop 2-3. In this case the formation of the hairpin 3-4 is too easy to make and the transcription terminator is formed (so less tryptophan synthesized).

Signal transduction in eukaryotes

- The ligand can pass through the membranes and reach the **nuclear receptor proteins**, there's a conformational change of the nuclear receptor allowing it to recruit of different co-repressor or co-activator (easy case and not very frequent).
- Usually the ligand passes through the cell membrane and then provokes some conformational change in the **cytosolic nuclear receptor protein** that allow its translocation inside the nucleus.

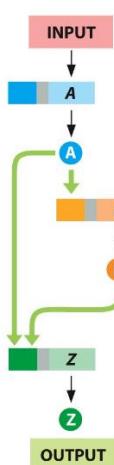
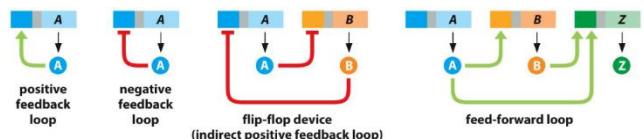
For example the NF- κ B pathway, important for the mammalian immune response. (not gonna be asked)



Transcription motifs that regulate transcription

These motifs are little element in which one regulator is hooked up one with the other. These circuits are serial or parallel.

Positive feedback loops, when are triggered have an explosion of gene expression and they don't go back. A negative one instead has a sinusoidal flow.

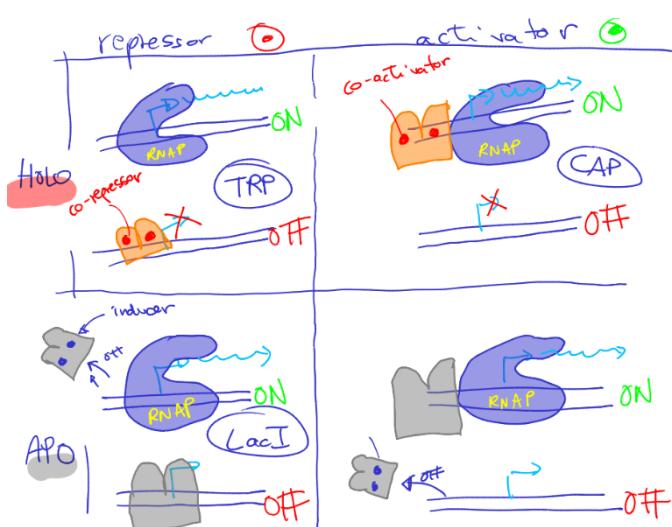
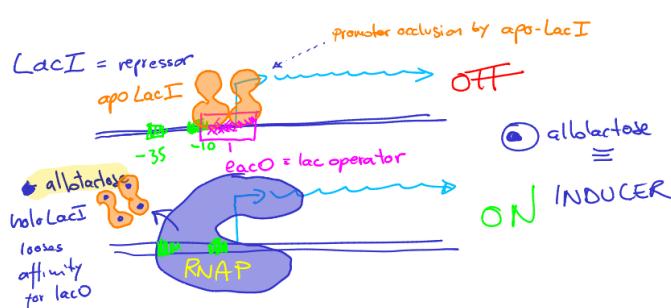


A feed forward loop can measure the duration of a signal

In this case Z is the output, however there is a direct way ($A \rightarrow C$) or indirect way ($A \rightarrow B \rightarrow C$).

There 8 possible combination of the circuit (4 coherent and 4 incoherent) we say coherent if the direct pathway has the same sine of the indirect one. Therefore the incoherent pathway is generator of impulses. The coherent one is instead noise cancellation.

Resume



12 RNA PROCESSING

Once a transcript is produced, it is not yet ready to attend its function.

- Cleavage: the premature mRNA is cut to get to its mature form
- Splicing: introns are taken out from the mRNA and the exons are fused back together
- **5' capping:** modification of the mRNA 5' end is important for stabilizing it and making it translatable
- **Polyadenylation:** 3' addition of a poly A tail, same function as 5' capping
- **Editing:** base insertion, base deletion, base modification

Processing is important because:

- it generates **diversity**, especially in alternative splicing which has the ability to choose from different alternative exons to form the mature mRNA.
- Contribute to regulation of gene activity, like 5' cap etc.
- It works as a quality control through which defective mRNAs are detected and degraded (post transcriptional)

This process involves many molecular complexes, and many of them contain both RNA and protein (ribonucleoproteins, RNPs). The RNA in RNPs can be structural but can also have catalytic activities (ribozymes). Other RNAs have the job to guide the RNP to the correct place for processing, called **guide RNAs**.

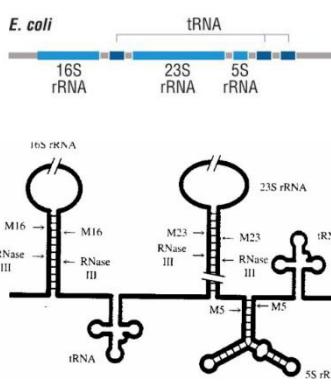
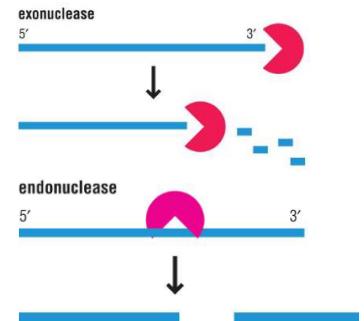


Cleavage

In *E. coli*, the rRNA genes provide the structural backbone for the Ribosome (small and large subunit), and are interspersed by non-translating sequences, moreover there are also tRNA genes present in the locus. So this locus is transcribed as a long precursor and needs to be processed to have the same amount of rRNA that the Ribosome needs.

This maturation process is achieved by ribonucleases:

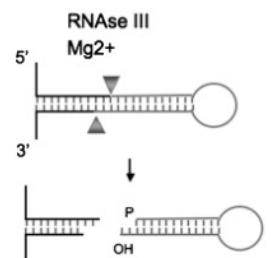
- **Exonucleases:** work 3' to 5' and basically "eat" the DNA without being sequence specific.
- **Endonucleases:** cleave the RNA within the strand and is more sequence specific
 - ssRNA cleaver
 - dsRNA cleaver



In E. coli, the process starts with **RNase III** that cleaves dsRNA.

The long precursor folds up in a very peculiar secondary structure where the ribosomal genes pile up in the loops that form on top of stems. So RNase III recognizes the stems and cut them out, freeing the rRNAs.

RNase III has a very peculiar signature: a 2nt overhang at the 3' end and a 5' monophosphate. And it's important for further processing these motifs.



RNase P instead cut ssRNA and has an RNA component. This RNase the RNA component itself is able to cleave mRNA.

In eukaryotic, archaeal, and mitochondrial RNase P instead, the RNA component cannot cut RNA alone.

Several tRNAs and some rRNAs have introns that require splicing, and it can be promoted by the RNA itself (self-splicing introns).

RNase E cleaves ssRNA and it can be guided by proteins and guide RNase to cut the target RNAs.

CCA addition at 3' end of tRNAs

The tRNAs have all a conserved CCA sequence that is the attachment site for the amino acid. This sequence is usually added later by a CCA adding enzyme that has a tRNA binding site and a nucleotide binding site. The addition of CCA is catalyzed by the nucleotide binding pocket that adds two CTP and one ATP. Sometimes the CCA adding enzyme can also degrade unstable tRNAs.

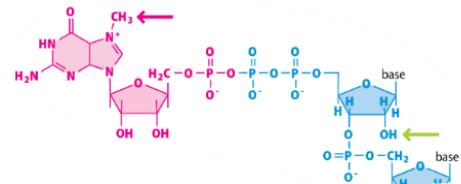
Modification of basis

In tRNAs more than 80 modification had been reported, used to generate diversity. Modification can be small (like methylation) or large (like addition of threonine). Also pseudo uridylation is an important modification that allows the RNA to be stable.

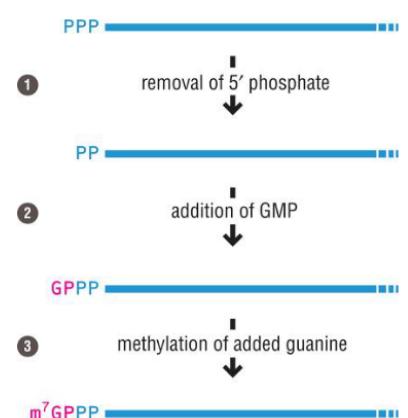
5' capping in eukaryotic mRNAs

The mRNA 5' end is triphosphate (not modified), so a modified nucleotide is linked (methylguanosine) and it is link via 5'-5' bond with the phosphate mRNA.

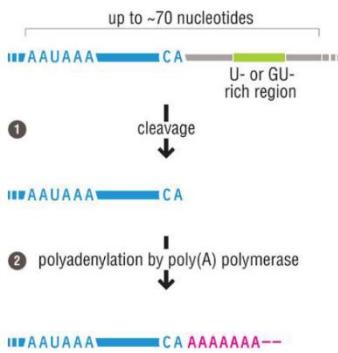
This prevents the RNA from being degraded 5' to 3' and it provides a tag to recognize the start of the RNA, for mRNA processing and export from the nucleus. The capping is also co-transcriptional since it is efficient for elongation and termination of the transcript. Sometimes the 2nd and 3rd bases are methylated at 2' C.



- It starts when the mRNA emerges from the RNA Pol II (about 20-30 nucleotides). *tRNAs and rRNAs do not have a cap.*
- Removal of a phosphate from the 5' end
- An enzyme called **guanyl transferase** attaches guanosine monophosphate (GMP) to the end in a 5'-5' triphosphate linkage.
- Finally the guanine is methylated



Polyadenylation



It is the addition of a poly-A tail. It stabilizes the 3' end and the longer the tail the better it is, because exonucleases degrade 3' to 5' so the 3' end of mRNAs are very sensitive. The tail is not genetically encoded.

mRNAs have polyadenylation sites where the pre-mRNA is cleaved and the poly-A is added.

This cleavage usually occurs after a CA that lies between a conserved AAUAAA hexamer and a U or GU-rich region. After cleavage, about 200 adenosines are added by poly-A polymerase.

The poly-A tail is also important because it allows circularization of the transcript, so the 3' end is able to be closed with the 5' end. This allows the ribosome to be more efficient since when it finishes transcription it already has the 5' end already ahead.

Capping and polyadenylation are coupled to RNA pol II activity

RNA pol II has a C terminal domain (CTD) that can be phosphorylated, and it is responsible for mediating mRNA processing.

- 1- When Pol II starts elongation, the CTD start being increasingly phosphorylated.
- 2- The partially phosphorylation recruits the capping enzyme on the polymerase. In this way as soon as the mRNA leaves the Pol, the capping enzyme is already present promoting capping.
- 3- Higher level of CTD phosphorylation promote the recruitment of splicing factor, so the transcript is co-transcriptionally spliced.
- 4- The 3' end processing complex is also recruited by the phosphorylated CTD.

RNA editing

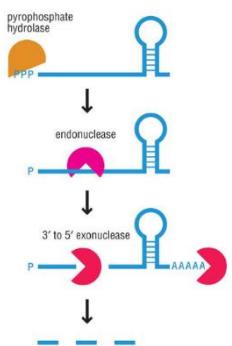
Deamination of Adenosine that give Inosine, the same if Cytidine is deaminated it gives Uridine. These editing can change the reading of the bases so recruit other tRNAs and finally give a different protein. And other things like that that he won't ask in the exam.

RNA degradation/decay in bacteria

Processing of defective or not useful RNAs. It is very important for the cell health. It is used in post-transcription regulation, and it can be very fast.

In bacterial mRNA, degradation is initiated by endonuclease, usually RNase E. It is able to cleave the bacterial transcript, then, the product of the first digestion are then degraded by exonucleases. Exonucleases start working after because the 3' end might be protected by a secondary structure and because it's missing the poly-A tail that in bacteria are needed to mark the transcript that need to be degraded.

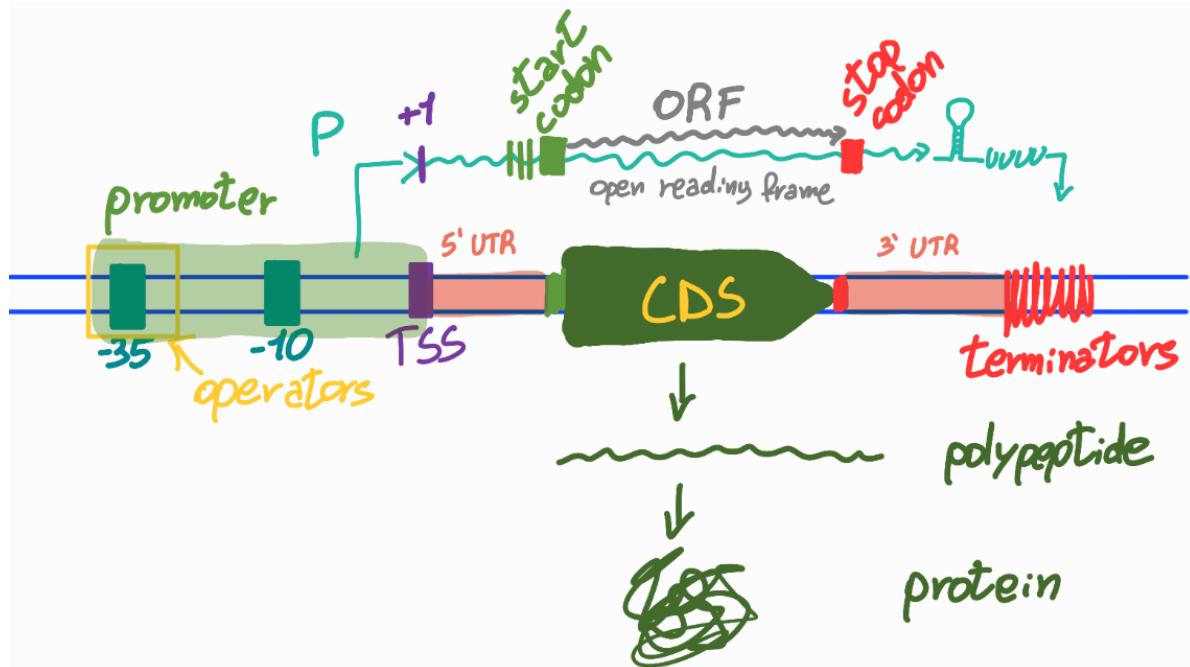
Intact RNAs have a triphosphate group at the 5', and the pyrophosphate hydrolase has the task to remove 2 phosphate when the transcript needs to be degraded.



Another way of starting degradation from RNase E, is having a small RNA that leads it to the mRNA. RNase E is part of the degradosome complex that is responsible for mRNA decay in bacteria.

RNA decay in eukaryotes, starts with the removal of the poly-A tail. Then digestion 3' to 5' by exonuclease can start, or the decapping enzyme acts on the mRNA and then the exonuclease XRN 1 can start degrading.

13 TRANSLATION

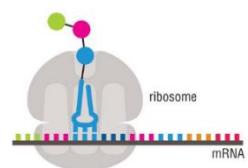


Translation is the generation of a polypeptide from a mRNA. Proteins are made out from 20 amino acids, and the point of translation is reading the codons and match its amino acid through tRNAs. These are able to pair their anticodon to the mature mRNA codon.

The linking of the right amino acid to the tRNA is fundamental and this job is performed by **aminoacyl-tRNA synthetases**.

Ribosomes carry out translation

In ribosomes the jobs are carried out by the RNA component and not the protein one. It has 2 subunits:



- **Small** responsible for the proper decoding. (Have I load the right tRNA? Has the codon-anticodon minihelix formed? If not reject)
- **Large** responsible for making the covalent bonds between one and the next amino acid, to form the polypeptide chain. This reaction is performed by the RNA components of the large subunits.

During synthesis polypeptide share information and the ribosome moves along the RNA 5' to 3' direction. The pace is similar of that of transcription in bacteria (15 aa/s → 45nt/s). It has a high error rate 1/1000 aa.

The ribosomes are assisted by a series of translation factor (often GTPases, so high energy consumption), translation is the most energy consuming process in the cell.

Translation has four phases: initiation, elongation, termination, and ribosome recycling.

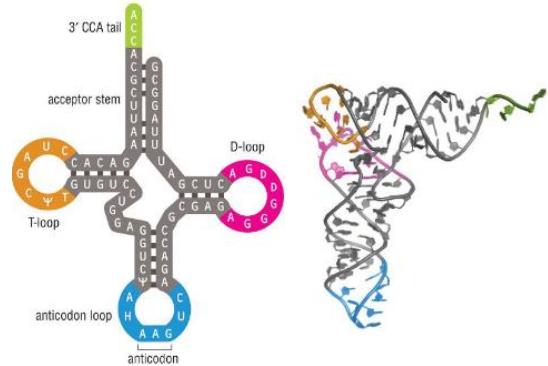
13.1 tRNA

The tRNA are bifunctional because it joins and link the aa world with the nucleotide information.

Amino acids are covalently linked to the CCA tail, while the codons in the mRNAs are recognized by the anticodon present in the anticodon loop.

They're 75-94 nts long, when they're bound to aa, they're called aminoacyl tRNA (important to call them like that in the exam).

the anticodon loop and the acceptor stem are in opposite side



The bases of tRNAs are highly modified.

- The **D-loop** is called after the **dihydrouridine** modification found in it.
 - The **T-loop** instead is after **ribothymidine**, and it also has **pseudouridine**.
 - In the **anticodon loop** there is present a **hypermodified purine (H)**. It is a conserved modification very useful because prevents the wrong base pairing with the codon delimiting the tRNA anticodon, secondly, it imposes a conformational change that flips out the basis of the anticodon making them available to make contact to the codon forming the minihelix.

Codons specify the genetic code

AGA									UUA									AGC				
AGG									UUG									AGU				
GCA	CAG								CUA									CCA	UCA	ACA		
GCC	CGC								CUC									CCC	UCC	ACC		
GCG	CGG	GAC	AAC	UGC	GAA	CAA	GGA	GGG	CAC	AUA	CUG	AAA	UUC	CCG	UCG	ACG		UAC	GUC	GUA		
GCU	CGU	GAU	AAU	UGU	GAG	CAG	GGU	CAU	AUC	AUU	CUU	AAG	UUU	CCU	UCU	ACU		UGG	UAU	GUG	GUU	UAA
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			

- STOP UAA, UAG, UGA
 - START AUG (Methionine)

There are cases of combination of degeneracy 4 and degeneracy 2, in the case of Arg Leu Ser that are degeneracy 6 (deg 4 + deg 2).

- In **degeneracy 4** the last nucleotide changes with the degeneracy. The reason for deg 4 is that sometimes **inosine** is present in the anticodon, and this can pair with U, C or A.
 - In **degeneracy 2** the last one could be either purine (R) or pyrimidine (Y). The reason of degeneracy 2 is the **wobble pairing** in the third base of the codon -> AAG – UUU instead of AAG – UUC.
Asp has GAC, GAU so GA(Y)

One tRNA doesn't recognize all codons for a specific amino acid. So tRNAs that carry the same amino acid are called **isoacceptors**.

There must be at least 20 different tRNAs and maximum is 64. There are actually **around 40 tRNAs** for the 61 codons, thanks to the inosine, wobbling and modifications.

The ribosome proceeds sequentially on the sequence, and it needs to readout codon per codon, implies having 3 different frames in the RNA sequence. Taking in consideration that the same mRNA sequence can specify three completely different amino acid sequences by changing the reading frame of one base, the ribosome needs a **punctuation sign** at the beginning of each coding sequence that sets the correct reading frame at the start of protein synthesis.

Some codons are used more frequently than others (rare codons), and these tend to be decoded by rarer tRNAs.

Codon usage is species-specific, and the genetic code is almost the same in all organisms (AUG methionine, three stop codons).

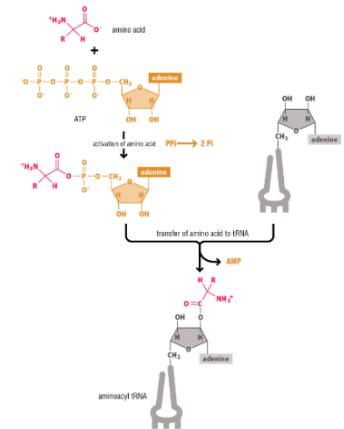
In addition, evolution has conserved codons so that mutations that change the amino acid that is encoded usually result in a similar amino acid taking its place (UUU phenylalanine to CUU leucine – both hydrophobic).

There are exceptions to the universal code, in mitochondria both AUG and AUA code for methionine. In mycoplasma, AUG encodes tryptophan instead of being a stop codon. Some organisms interpret stop codons as a signal to incorporate a non-standard amino acid.

Aminoacyl-tRNA synthetases carry out the first step in decoding

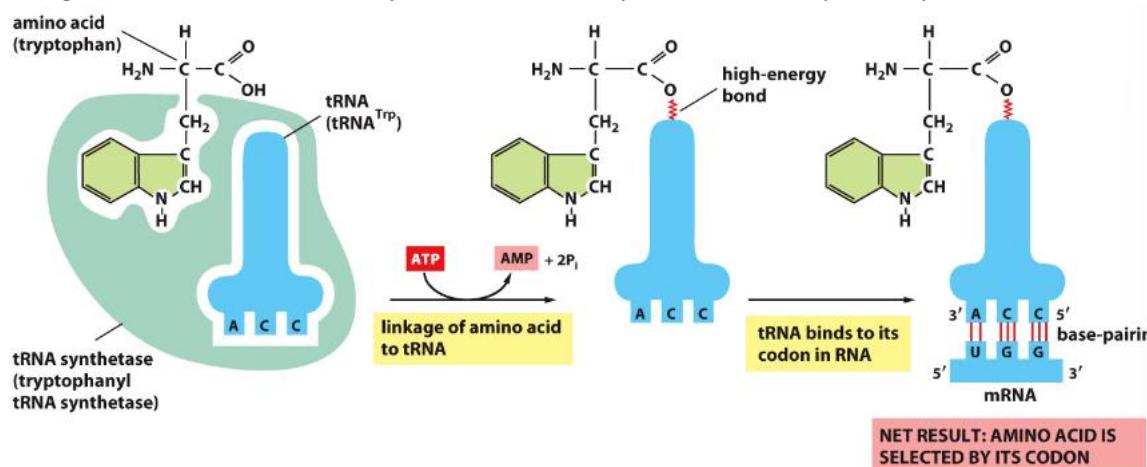
Aminoacylation attaches amino acids to tRNAs. This is done by **aminoacyl-tRNA transferases** in a two-step process requiring ATP. The enzyme for a certain amino acid is denoted **aaRS** e.g. GlyRS

- The amino acid is **activated** by attachment of AMP. This releases pyrophosphate and provides energy
- The activated **aminoacyl-adenylate** remains attached to the enzyme
- The enzyme then transfers the amino acid to the 2' or 3' OH of the ribose of the **terminal adenosine** on the tRNA 3' CCA tail.



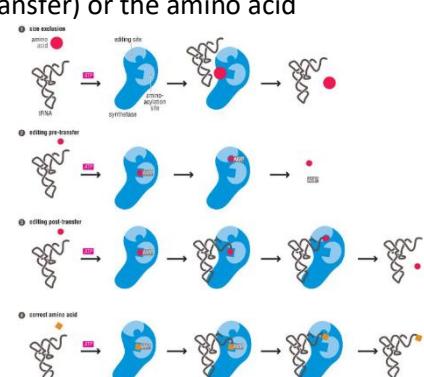
Each amino acid has its own aminoacyl-tRNA synthetases, which recognize tRNA by **identity elements** (sequence and structural features). The specific amino acid with which a tRNA is loaded is indicated with a three-letter superscript, such as tRNA^{Met}. Loading is accurate and the correct amino acids are chosen in a **two-step process**. Most aminoacyl-tRNA synthetases have an **aminoacylation site** and an **editing site**.

The genetic code is translated by means of two adaptors that act sequentially



Editing by aminoacyl-tRNA synthetases mechanisms:

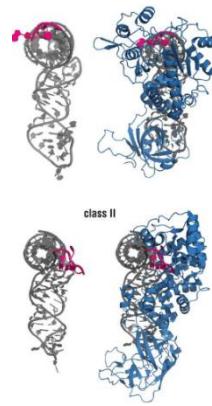
- Size exclusion keeps non-cognate amino acids that are too big out of the amino-acylation site
- The editing site can accommodate the activated amino acid (editing pre-transfer) or the amino acid after attachment to the tRNA (editing post-transfer)
- If the amino acid is rejected pre-transfer, the aminoacyl-adenylate (activated aa) is hydrolyzed. If the rejection occurs post-transfer, the aa is cleaved from the tRNA.



There are two classes of aminoacyl-tRNA synthetases, each with about 10 members:

- Class I usually recognize the minor groove of the acceptor stem.
- Class II recognize the major groove.

They have very different structures, and apart from some exceptions, they typically recognize the same groups of aa.



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 have dual specificity for $tRNA^{Asn}$ and $tRNA^{Asp}$, and $tRNA^{Glu}$ and $tRNA^{Gln}$.

A transamidase reaction changes the side chain of the attached amino acid from an acid to an amide, producing asparagine and glutamine bound tRNAs.

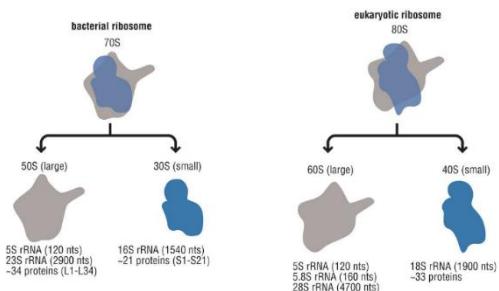
13.2 RIBOSOMES

Ribosomes catalyze the formation of peptide bonds between aa.

Are large 2.5 MDa to 4 MDa, about 2/3 of which is rRNA and 1/3 protein.

They're composed of a small and a large subunit, each containing ribosomal proteins and RNA.

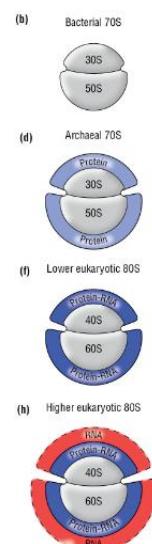
- The **small subunit** mediates interactions between mRNA and tRNA.



- The **large subunit** catalyzes bond formation and has an exit tunnel through which the growing polypeptide emerges (often target for antibiotics).

The interface between the subunits is important for movement of tRNAs and mRNA.

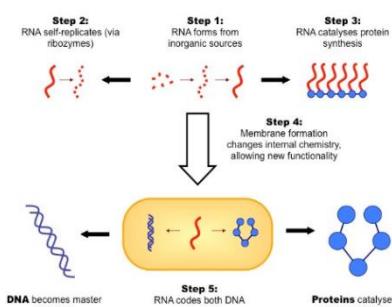
Eukaryotic and bacterial ribosomes are generally conserved but differ in their composition.



The interface between the subunits is rich in rRNA. Within a subunit, the proteins extend "arms" into the RNA regions, which are usually highly basic, and are thought to help with packing the negatively charged rRNA phosphate backbones.

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

The rRNAs in the ribosomal subunits are divided into domains and both rRNAs and proteins are extremely highly conserved across species.



RNA world origin hypothesis, says that the RNA components of ribosomes were present before the protein components, and that proteins were recruited to the ribosomes later in evolution. It is easy to say since the decoding region of 16S rRNA mediates the interaction between tRNA and mRNA, and the 23S rRNA in the peptidyl transferase center interacts with the tRNAs.

Ribosomes have three tRNA binding sites (A-P-E):

- **Aminoacyl** anticodon-codon recognition and loading of a new tRNA
- **Peptidyl** form the peptide bond with the A site aminoacyl-tRNA and then moves to the E site
- **Exit** just a site where the tRNA can be discharged

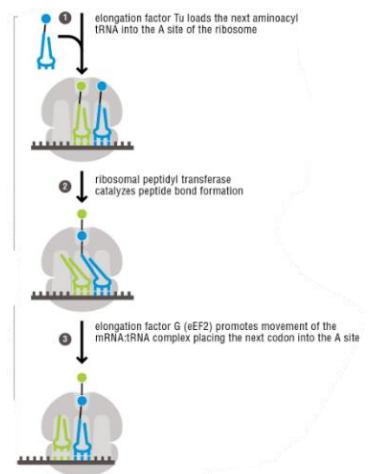
The steps of translation are accompanied by factors: Initiation Factors (IF), Elongation Factors (EF), Release Factors in termination (RF).

Initiation

The two ribosome's subunits need to be assembled at the site where translation starts, one important element for the recognition of the translation starting site is the start codon (AUG). The initiation factors have the responsibility for casting the two ribosomal subunits onto the translation start site. The IFs loads the methionyl-tRNA and the small subunit at the initiation site, and then the large subunit joins in the way in which the methionyl-tRNA is in the P site.

Elongation

An elongation factor (in bacteria Tu, in eukaryotes eEF1A) add an aminoacyl-tRNA to the A site. Then, if the codon-anticodon pairing is correct, then there is a conformational change that will promote the formation of the peptide bond. The bond is catalyzed between the A and P sites and involves the transfer of the amino acid chain onto the aa carried by the tRNA in the A site. This drags the aa into the P site while the tRNA is still in the A site (**hybrid state**) (2). The hybrid state is resolved by translation thanks to another GTPase elongation factor called **EF-G** in bacteria and EF2 in eukaryotes (3).



Termination

A stop signal interrupt elongation (UAA *ochre*, UAG *amber*, UGA *opal*), which is decoded by release factors RF that are **proteins** (not tRNAs !).

There are two types of RF in bacteria:

- RF1 decodes for UAA and UAG (ochre and amber)
- RF2 decodes for UAA and UGA (ochre and opal) *(helps remembering them in alphabetical order)*

In eukaryotes instead there's just one RF: eRF1 recognize all three of them.

When these release factors are loaded onto the A site, elongation stops, the polypeptide chain is released and the two subunits dissociate from the mRNA and **recycling factors** recycle the ribosome.

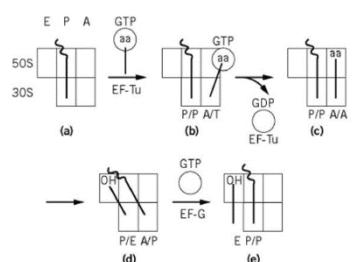
A **polysome** is a mRNA that is translated by many ribosomes simultaneously.

Hybrid state model

It's a model (not absolute true). Involves the **ratcheting** (movement of one element with respect to another in a circular way, through which you can go forward but not back) of the two ribosome subunits, thanks to conformational changes and hydrolysis of GTP. **EF-G** is the GTPases responsible for the translocation.

In ribosomes the ratcheting is the movement of the 50S with respect to the 30S.

This creates a situation of hybrid state because the anticodon loops of the tRNAs are still in the same position but the aa and the polypeptide are one position shifted.



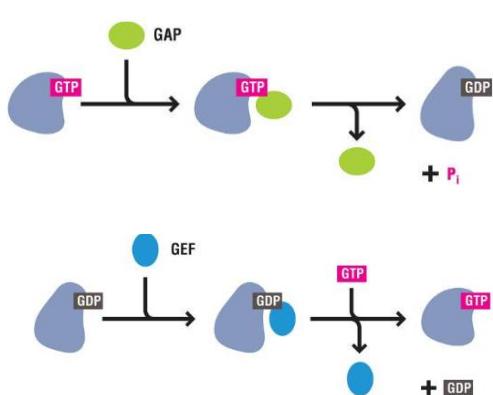
13.3 TRANSLATION FACTORS

Are very often **GTPase factors**, which can hydrolyze GTP providing the energy to drive the step forward (conformational changes). Before the hydrolysis of GTP the wrong tRNA in the A site can be rejected by after it's not possible to go back.

Other translation factors bind to the ribosome on the E site and the A site preventing the methionyl-tRNA to bind in the wrong site.

GTPase factors interact with the protein-rich ribosome flexible stalk region. They all contain P-loop which is a motif Gly-X-X-X-Gly-Lys, it is important to bind GTP. This motif after hydrolyzing GTP undergoes conformational change (allostery).

EFG-GTP recognizes pre-translocation state ribosomes, and the GDP form dissociates from the ribosome.



In presence of GTPase bound to GTP being hydrolyzed into GDP, the process is promoted by GTPase Activating Protein **GAP**.

From the GTPase with GDP, guanine-nucleotide exchange factors **GEF**, exchanges GDP with GTP, reforming the cycle.

Also the **ribosome itself can act as a GAP**, promoting hydrolysis of GTP by EFTu once an aminoacyl-tRNA is in the A site.

Translation can proceed without GTPases, but they contribute making it faster and more precise.

Furthermore, many translation factors mimic tRNA structures, and it is usually because they can fit in the same pocket. For example EFTu-tRNA that is similar to EF-G can both fit in the A site.

13.4 INITIATION

Initiation is very different in eukaryotes and bacteria.

- **Bacteria**

The key sequence coded in the RNA is the Shine-Dalgarno sequence, found about 6 nt upstream the AUG codon. It's a purine rich sequence and it is recognized by sequence specific pairing with a very conserved ribosomal RNA in the small subunit. The ribosome has an anti SD sequence that helps align it to the transcript.

In bacteria there is a ribosomal binding site.

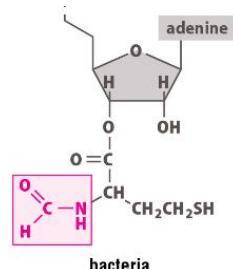
- **Eukaryotes**

It's present a Kozak sequence present around the AUG codon. The initiation factor and the small subunit doesn't load onto it, but what is recognized is the CAP at the 5' end, where it starts sliding until it finds the AUG codon (**scanning**).

The initiator tRNA must be different not to be exchanged by a normal RNA, so it is modified. In bacteria it has a modification in the methionine called formylmethionyl-tRNA (fMeth-tRNA).

There are GTPase that bind methionyl-tRNA to the P site. Their GTPase activity is activated only if the tRNA is loaded onto the P site.

In bacteria the initiator tRNA has a C-A wobble in the acceptor stem (normal in eukaryotes).



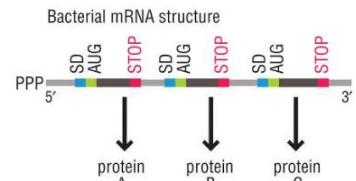
Both eukaryotes and bacteria have 3 G-C pairs in the anticodon stem. Which avoid the binding of EFTu to the tRNA, since the EFTu loads the tRNA onto the A site, not wanted with the initiator tRNA since it is different from the other Met-tRNA.

Locating the translation start site in bacteria

Starting from a multicistronic operon that gives rise to one mRNA. So many proteins can be made out of that mRNA, which have each a SD sequence. Sometimes they are even overlapping, there could be a stop on one frame and in the other a SD (frequent).

lambda cro	anti-Shine-Dalgarno	Shine-Dalgarno (in blue)
U U C C U C C		
A C T A A G G A G G T T G T - - A T G		
C T G A A G A T T A A C - - A T G	rpsA	
T G G A G G A C T A A G A A - A T G	rplA	
T T A G A G G G A C A A T C G A T G	tufB	
G G G A G T A T G A A A A G T A T G	araC	
A C A G G T A G T G A A T - - A T G	galS	
A T A A G G A A A T C C A T T A T G	lacY	

The **Shine-Dalgarno** sequence has the consensus AGGAGG (more or less, just remember a lot of purines), and it pairs with a polypyrimidine region in the 3' end of the bacterial 16S rRNA. The deviation from the consensus sequence controls the strength of translation.



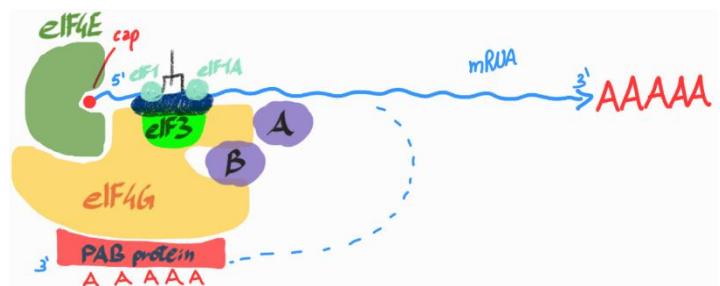
At the initiator start, there are 3 initiation factors: IF1, IF2, IF3. Which help guide the f-Met-tRNA to the P site. IF1 and IF3 bind to the A and E site to avoid errors. Then the small rRNA subunit is loaded onto the SD. IF2 loads the initiation tRNA and it also checks that everything is in place. So then IF2 hydrolyze GTP to provide energy for joining the large subunit. Finally, all three IF are displaced.

In eukaryotes

There are no equivalences of the SD, so scanning is required. And, because the 5' CAP is so important, **eukaryotic mRNAs are monocistronic**. Usually the first AUG in the mRNA is the start translation codon, but sometimes it is not and it is sensitive to the sequence context (Kozak sequence).

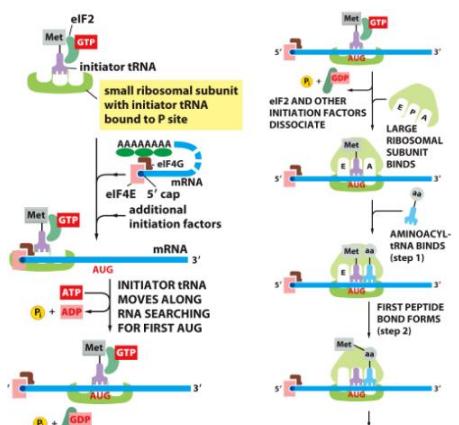
The recognition of the 5' CAP is aided by the 3' end of the RNA that circularize. The presence of both poly-A tail and CAP are both important to stimulate translation initiation. This is the reason why there are many IF factors at the 5' CAP end:

- **elf4E** that is a CAP binding protein
- **Helicases** (elf4A and elf4B not important to remember) needed to unfold the mRNA in case of secondary structures
- **elf4G** is a scaffolding protein to which all these factors bind
---- the upper four together form the **elf4F** ----
- **elf3** binds to the small ribosome subunit (40S)
- A GTPase like the IF2 in bacteria (same function)
- **PAB protein** (Poly-A Binding protein) binds the poly-A tail at the mRNA 3' end.



The formation of the close mRNA loop act as a quality control. And when it is achieved, 40S starts scanning until the initiator tRNA check for the AUG codon close to the Kozak sequence.

Then, to assemble the large subunit onto the small one, the GTPase called **elf5B**. The process is very similar to prokaryotes (with IF2):



13.5 ELONGATION

It has precise decoding involving a lot of RNA-RNA interactions.

Bacteria

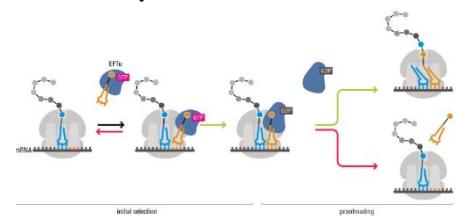
A new aminoacyl-tRNA in the A site, brought along by the EFTu. The proper addition of a **cognate** tRNA (3 nucleotide perfectly complementary), but it's possible to have also **near-cognate** (single mismatch) or **non-cognate** (more than one mismatch) that are rejected by the ribosome. These are recognized because the minihelix formed is not perfect.

When the cognate tRNA is matched, then there is the **activation of the GTPase activity** of EFTu.

The formation of the cognate match stimulates conformational changes in the ribosome that act like a GAP (GTPase activator protein) on EFTu. This promotes acceptance of the aminoacyl-tRNA.

How does the **ribosome control the formation of the minihelix**? Forming an **A-minor triplex**.

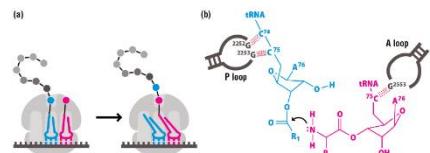
This happens because the extremely conserved nucleotides in the 16S RNA respond to cognate tRNA recognition: A1492, A1493 and G530. They can form a **H binding network** in the minihelix in the A site of the ribosome. These three nucleotides are usually outward oriented, but the **first two base pairs** binding of the minihelix formed by tRNA and mRNA. These are changes that promote **distant conformational changes** and stimulate GTPase activity of EFTu.



Formation of the peptide bond involves the transfer of the peptide chain onto the aminoacyl-tRNA onto the A site, that leads to the formation of the polypeptide chain.

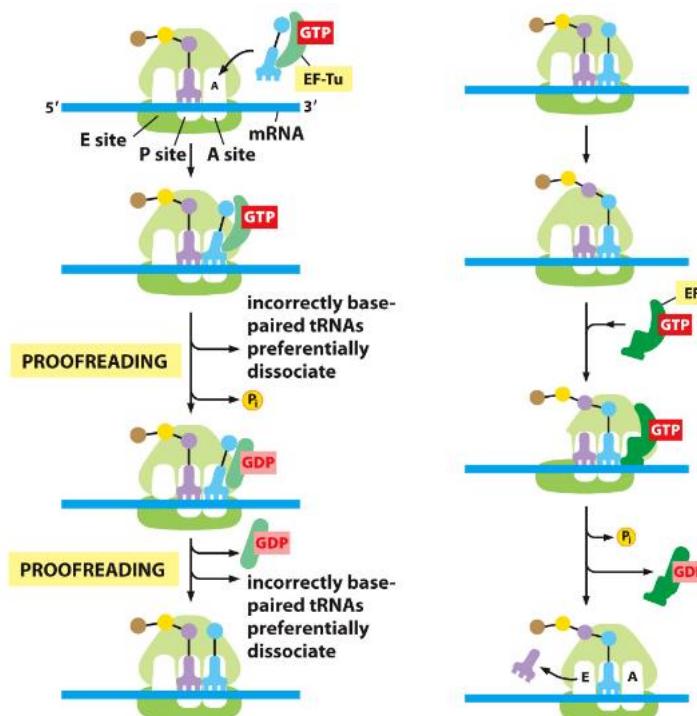
Peptide bond formation is performed by the RNA's 2' OH group.

It works in a complex way he tried to explain no one could follow I guess it's not in the exam. But basically after that the hybrid state is formed.



The hybrid state is solved promoting the translocation of the mRNA-tRNA complex. This is performed by the EF-G (GTPase) that can bind and squeeze into the A site pushing the aminoacyl-tRNA loop to the P site. The binding of EF-G also promotes the structural rearrangement that are needed to solve the hybrid state.

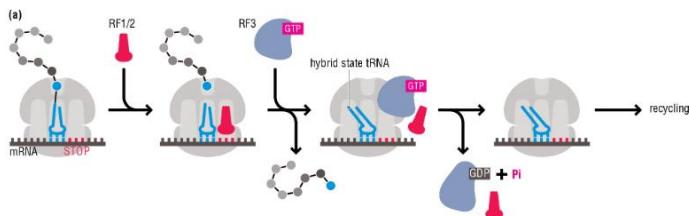
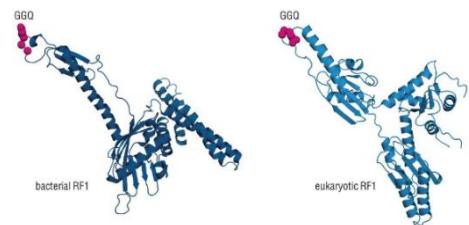
Bacterial elongation recap:



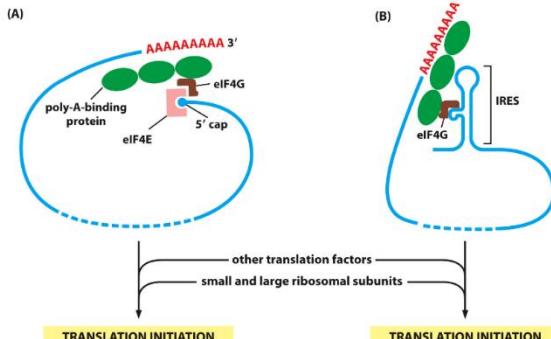
13.6 TERMINATION

Elongation continues until the ribosome encounters a stop codon in the mRNA. Class I release factors recognize stop codons and promotes hydrolytic release of the finished peptide.

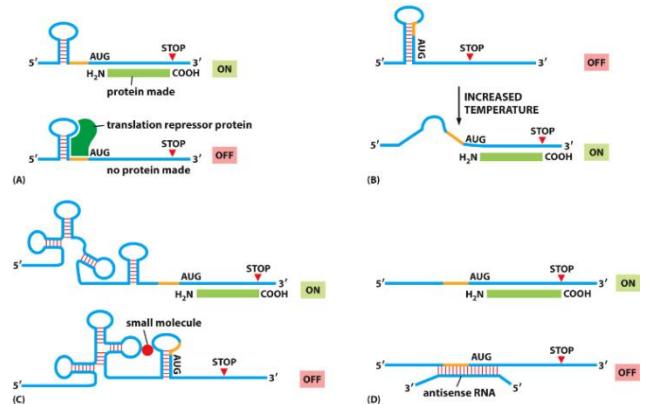
Class I release factors are two in bacteria (RF1, RF2), and one in eukaryotes (eRF1). Class I RFs in bacteria and eukaryotes are unrelated but have similar GGQ motif that is needed for catalysis.



5' (and 3') UTRs can control mRNA translation thanks to secondary structures. There could be stems that allow the binding of a repressor protein that covers the reading of the AUG codon. Or stems that comprehend the AUG codon (translation off), and that open when the temperature is high enough.



In bacteria, class II RFs are also needed for termination, RF3-GTP promotes dissociation of RF1/2 after peptide release. RF3 derives from EFG.



Another way of regulating translation initiation is by **Internal Ribosome Entry Sites** (IRES), viruses use it. In order to make the translation machinery translate their own transcripts. IRES is a **secondary structure in the mRNA that is able to bind to IF4G** (scaffolding protein), this allows the transcript to be translated even if it doesn't have a 5' cap (bypass of the cap recognition process).

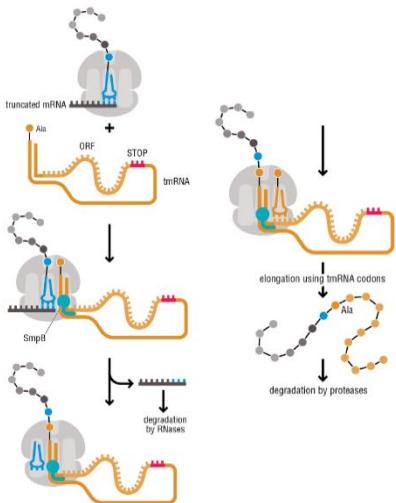
Translation initiation is always in competition with degradation factors.

Changes in mRNA stability can regulate gene expression. If the mRNA is not translated, then it is degraded by the mRNA decay. This balances with the translation initiation factors because they bind the ends of the mRNA and prevent the recognition of these by decay factors. Also ribosomes help protect the mRNA from endonucleases.

13.7 RIBOSOME STALL

Sometimes, particularly in bacteria that don't have the control of the 5' and 3' end, it happens that the ribosome stall. When the ribosome can't proceed forward, it's a sign of danger for the cell and the mRNA must be directed toward the RNA decay pathway, the ribosome is recycled and the polypeptide needs to be destroyed.

In eukaryotes, the stalling of the ribosome is not common because translation and transcription are separated and there's capping and polyadenylation that governs the rate of translation. However, in bacteria transcription and translation are coupled, so translation can occur also in truncated mRNAs which leads the ribosome to keep translating until the 3' end without having encountered the stop codon. The



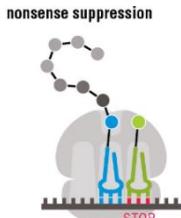
release factors won't work without it and the **A site would be left empty**. In bacteria, to overcome this problem, there's the **tmRNA** which is a bivalent mimic of the tRNA and the mRNA both in the same molecule. It is a translation rescue system and protein tagging system all at once. tmRNA acts as tRNA and loads into the A site carrying an Alanine along. However tmRNA is longer than a tRNA and **carries along short cistrons** (sequence that can be translated). This gives the possibility to the ribosome to keep going on with translation onto the tRNA and liberating the mRNA. tmRNA host the stop codon after **11 amino acid "tag"** that help recognize the protein and lead it to degradation.

13.8 RECODING

Recoding is the bypass of codon information. It happens very rarely, and it is not by accident but happens because in the sequence there are motifs that promotes the recoding.

- Nonsense suppression bypass of a stop signal
- Programmed frameshifting changing the reading frame by (-1) (very common in virus) or (+1)

Nonsense suppression

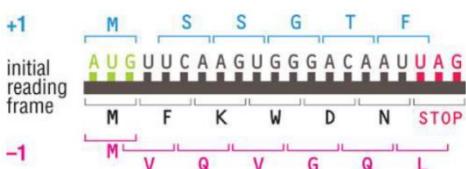


Programmed nonsense suppression is caused by knockout of class I and II suppressor. In which RF1 or 2 are substitute so the stop codon is not read like one. The glutamine-tRNA can substitute the stop signal.

In a programmed nonsense suppression (for example in eukaryotes) is induced by the presence of a mRNA secondary structure (pseudoknot) downstream the stop codon, which slows down the ribosome. In normal condition with plenty release factors, they would load and terminate translation, however, there are factors that able to recruit the RF so they won't be available. This gives the opportunity to the Gln-tRNA to compete in the overwriting the stop codon.

Stop codons can also allow incorporation of non-standard amino acids. For example selenocysteine (that has selenium instead of sulfur), that is incorporated in several enzymes in catalytic sites where it can act as a strong reducing agent.

The tRNA^{Sec} has an anticodon that matches UGA. In E. coli there are programmed sites where nonsense suppression can occur, there are particular secondary structure called SelenoCysteine Insertion Sequence, that can recruit particular proteins that aid the loading of the SelenoCysteinyl-tRNA instead of the RF.



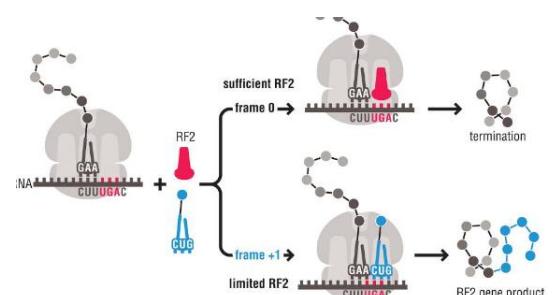
Frameshifting disruption

Usually in presence of frameshifting, there's the possibility of finding premature stop codon or not finding them at all. Programmed frameshifting is still induced by secondary structures.

+1 frameshifting

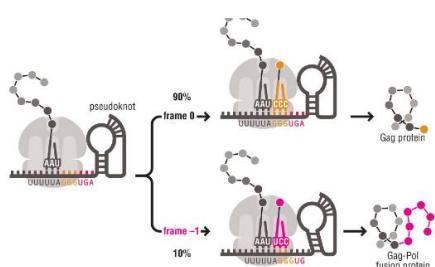
In a bacterial example, there could be an UGA stop codon, recognized by RF2, that if there would be no frameshifting it would simply end translation. But, Leucinyl-tRNA has CUG sequence. If there's sufficient amount of RF2 then it would normally promote translation termination.

However, the mRNA sequence is particular: CUU UGA C. So, the



ribosome can slip 1 nt ahead (+1) because the GAA anticodon binds with CUU, but it could also bind to UUU forming a wobble pairing with one shifting of 1 nt. Because of this possibility, if there's a C after the stop codon, it is possible to reform a GAC codon that can be readout by the CUG (Leucin) anticodon, bypassing the stop codon.

The part of the protein translated thanks to the shifting happens to be the RF2, this explains why when there are enough RF2 to interrupt the translation of themselves, they're not needed so they won't be translated. Instead, when the RF2 levels are low, the frameshifting allows to synthesize more of them.



-1 frameshifting

It occurs when the advancement of the translating ribosome with a pseudoknot. It happens very frequently in virus, which can make two products out of the same RNA by frameshifting. Sometimes viruses use this to promote the proper expression of the polymerase (like Sars-CoV-19).

In -1 frameshifting there usually is a very U-rich sequence upstream the terminator, which allow the ribosome to backtrack by one position. Basically the pseudoknot pushes back the ribosome and make it start translating -1 frameshifted.

BONUS: Antibiotics

Many antibiotics work selectively and specifically on the translation processes of bacteria or fungi. The best ones are the ones that target the ribosome at the different phases: premature arrest, initiation, elongation termination etc.

13.9 CO-TRANSLATIONAL PROTEIN FOLDING

Once the mRNA sequence is translated to a polypeptide sequence, there are some mechanisms that activate in order to fold it properly. Frequently co-translational protein means that once the polypeptide chain grows, it starts folding progressively as it exits the ribosome. In case of modular proteins (made out of more domains), it folds the first and then the second and so on.

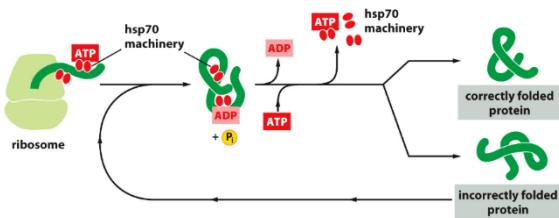
Folding is influenced by the rate and the speed of translation, indeed, there are sometimes particular **pausing sites** which slow the ribosome in order to give time to the polypeptide sequence to fold correctly before the second domain is translated itself. If it is not folded properly, the hydrophobic regions of one domain could interact with another domain and deform the quaternary structure giving **inclusion bodies** (unfold protein aggregates). For example expressing a mammalian gene in E. coli probably leads to this because translation in bacteria is much faster and the pause signals could not be respected.

Post-translational **molecular chaperone** is needed to fold proteins that are not able to fold on their own. Sometimes, this process is not possible because the protein remains unproperly folded. Then, these proteins are better be digested and degraded by the **proteasome**.

Chaperone

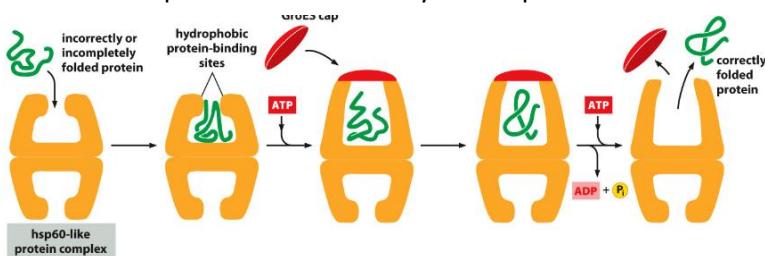
A problem with unfolded proteins is probably caused by hydrophobic patches interacting with each other, so exposed hydrophobic patches are used for protein quality control. Hsp60 (heat shock protein) and Hsp70 have affinity for these exposed patches, and their role is to allow the refolding of these proteins. They also have to mask these hydrophobic patches to prevent further aggregation of these patches.

Chaperones are heat shock protein because the heat can denature the proteins, so they intervene keeping proteins in place.



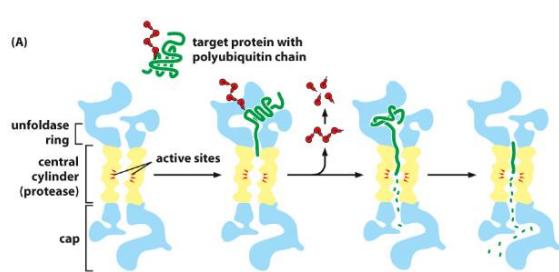
Most proteins are folded by chaperones, one example is Hsp70, which acts very early in the life of proteins. As the polypeptide exits the ribosome, Hsp70 binds to it where the hydrophobic patches are (patches that need to interact with cell membrane). The chaperone has ATPase activity, so when it binds tightly to the polypeptide chain it masks the patches, this gives time to have a prefold protein that is kept in place by other interaction. When ATP binds again with the chaperones, they leave the patches and give as result a perfectly folded molecule.

Sometimes it is not enough and there's need of **post-translational** actions, when the result is still an incorrectly folded protein, the Hsp60 chaperone takes action. It is a sort of barrel (GroEL) with a lid (GroES) that encapsulate the incorrectly folded protein in the isolation chamber which is covered by hydrophobic patches. Which interact with the ones in the protein opening its structure and forming an environment in which the protein can fold up properly again. The cap then open spending ATP, and the correct folded protein can exit the barrel.



Proteasome

It is responsible for the degradation of the proteins folded incorrectly. It is a bulky environment which has proteolytic activity. It looks like a tube with different domains:



- Unfoldase ring responsible for unfolding the misfolded peptide
- Protease domain is the core hosting the active site that hydrolyze the polypeptide exit for the amino acids
- Cap

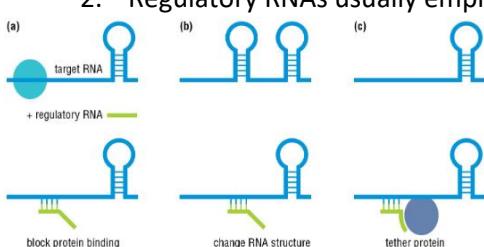
Usually the proteasome is able to degrade in eukaryotes especially, ubiquitinated proteins (with the ubiquitin tag)

14 RIBOREGULATION

Riboregulation is a regulation that can happen at both transcriptional and post-transcriptional level. It is a regulation performed by regulatory RNAs (ncRNA) which tend to be short.

It has been found out that many RNAs are non-coding but their role is to regulate gene expression, for example, controlling RNA decay modifying the life of a mRNA, or change the rate of translation of the mRNA. This can happen because these regulatory RNA can bind to their target mRNA thanks to their complementarity. *CRISPR is an endonuclease that is guided by a short RNA to the DNA*

1. Many transcripts of non-coding regulatory RNAs are often processed in order to yield a functional final product (maturation), especially in eukaryotes.
2. Regulatory RNAs usually employ base pairing and complementarity to recognize their target (RNA or DNA), so it has specificity.



3. They very often interact with protein in order to boost their function, or recruit proteins that perform the degradation or translation of the target.