Doctor 022

Biochemistry

Sheet no 1

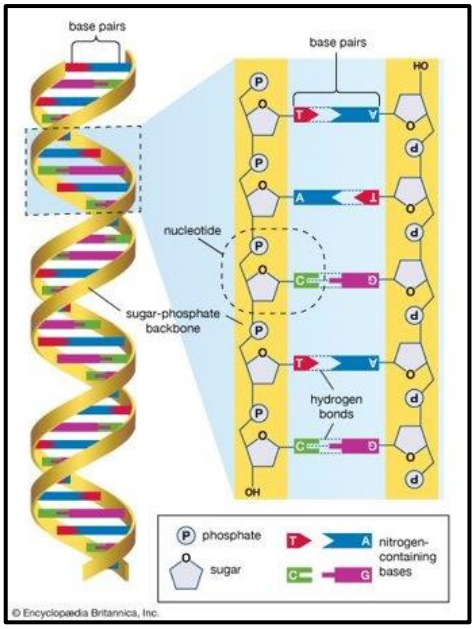
Molecular Biology part

Writer: Al-Razi Node

Corrector: Al-Razi Node

Doctor: Dr Mamoun

Important background information



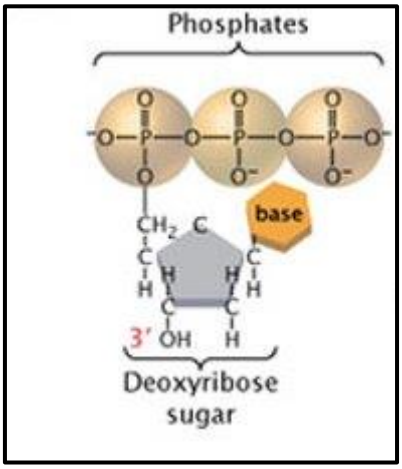
Double helical

two strands wind around each other in a helical pattern.

It's not a perfect helix, it contains grooves (major and minor)

Monomers and polymer

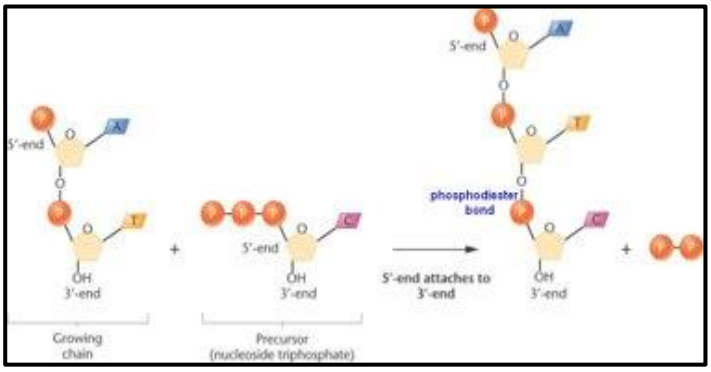
The DNA is the polymer, it is made of small monomers which are nucleotides (deoxyribonucleotides).



deoxyribonucleotides are composed of a deoxyribose which is the sugar (ribose) without its (OH) on C2, this sugar is connected to a nitrogenous base through a glycosidic linkage (glycosidic because it involves the anomeric carbon).

This base can be a purine (A/G, double ring structure) or pyrimidine (C/U/T, a single ring).

Thymine (T) is found in DNA while Uracil (U) is found in RNA. Adenine (A), Guanine (G) and cytosine (C) can be found in both.



The charge

Linked to this sugar are three phosphate groups that carry negative charges, resulting in the DNA molecule carrying an overall negative charge.

In order to link nucleotides, we need energy, this energy is derived from the phosphate groups (two of the three phosphate groups are released giving sufficient energy)

Complimentary

The two strand are complimentary to each other, adenine (A) always pairs with thymine (T), and cytosine (C) always pairs with guanine (G).

So, if you know the sequence of one strand you can predict the sequence of the other.

Anti-parallel

The two strands are complimentary meaning one of them goes in a 5 prime to 3 prime direction while the other goes in a 3 prime to 5 prime direction.

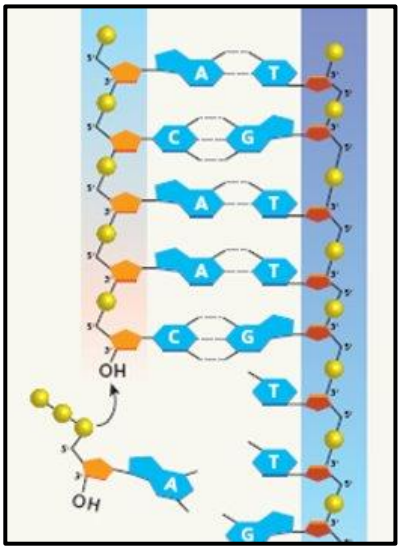
Phosphodiester bonds

The two nucleotides are connected to each other via Phosphodiester bonds.

Hydrogen bonds

The two strands are connected to each other via H-bonds between the nitrogenous bases. C is connected to G by 3 H-bonds, while A is connected to T by two H-bonds.

Enzymes that form a phosphodiester bond



In this lecture, we will focus on two important enzymes.

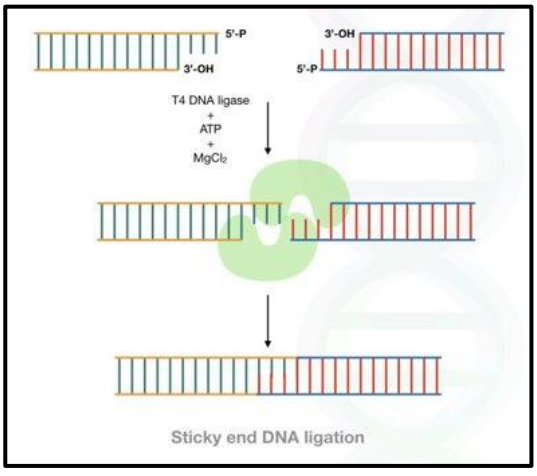
1.DNA polymerase: it is responsible for synthesizing the DNA strand by adding nucleotides then forming the phosphodiester bond between the two nucleotides depending on the base of the complimentary (or templet) strand.

For example, when it reads C on the complimentary strand it places G on the strand that is being synthesized.

It is very accurate.

The synthesis moves in 5 prime to 3 prime direction.

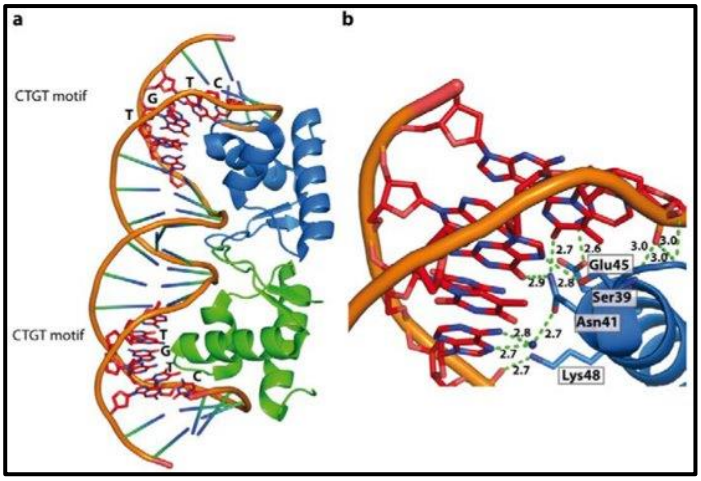
2. DNA ligase: Just like any other ligase it needs ATP to connect (ligate) two DNA fragments together via phosphodiester bond.



DNA-protein interaction

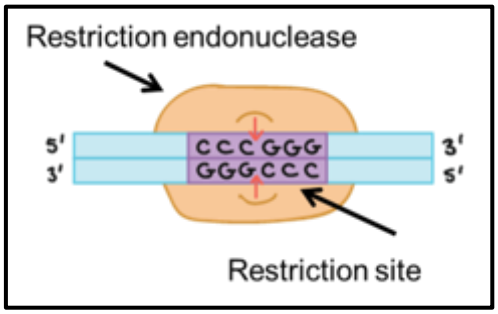
The amino acids of proteins can interact (non-covalent) with the bases of the DNA.

Usually, the proteins interact with the DNA through the major groove.

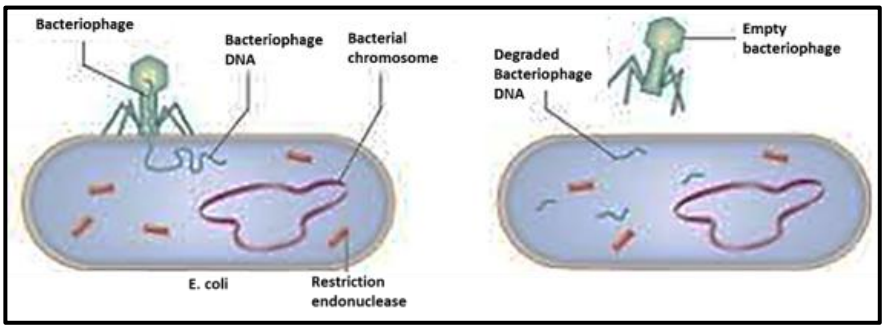


RESTRICTION ENDONUCLEASES

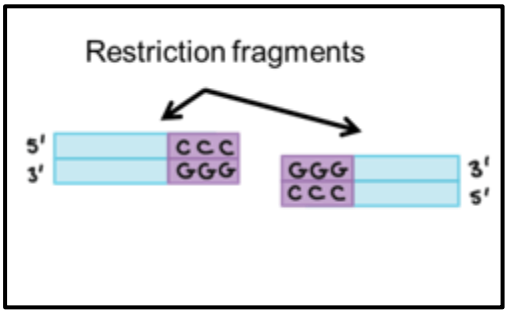
They are produced by bacteria to protect themselves from bacteriophages (a bacteriophage is a virus that inserts its DNA inside bacterial cells, and this DNA controls the genetic system inside the bacterial cell).



The tiny virus called bacteriophage takes control of the bacterial cell. It forces the cell to create many copies of itself, which are more tiny viruses. Eventually, the bacterial cell becomes full of these tiny viruses and bursts open, releasing them.



To protect themselves, bacteria make special tools called endonucleases. These tools cut the DNA of the tiny viruses. These tools are called "restriction endonucleases" because they limit the growth of the tiny viruses.



Endonucleases are enzymes that degrade DNA within the molecule. unlike exonucleases which degrades it from the ends.

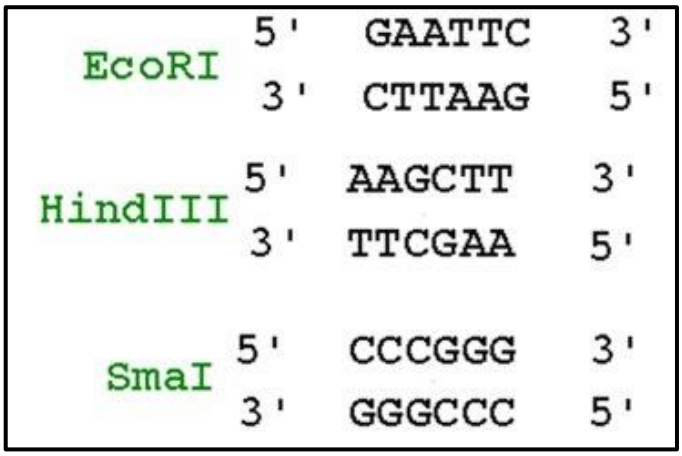
Restriction endonucleases: Bacterial enzymes that recognize and cut (break) the phosphodiester bond between nucleotides at specific sequences (4- to 8-bp restriction sites) generating restriction fragments. And within this restriction site the enzyme makes the cut.

Each bacterial species produces one or two types of endonucleases, and these restriction endonucleases recognize specific sequences, and these sequences are known as restriction sites.

The cleavage happens within these sites producing restriction fragments.

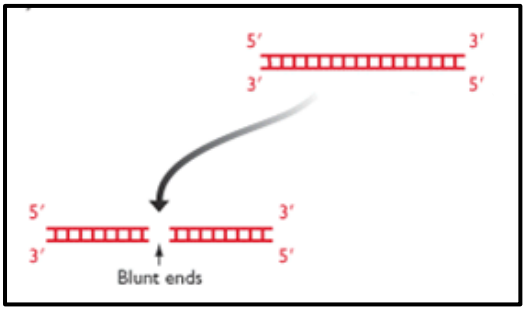
Palindromic sequences

The sequences recognized by restriction endonucleases—their sites of action—are read the same from left to right as they do from right to left (on the complementary strand). Note that we read the sequence from 5 to 3.



Types of cuts by restriction endonucleases

Restriction enzymes cut DNA in two different ways:



Blunt(sharp): enzymes cut at the same position on both strands giving blunt-ended fragments. So, there are no sticky ends.

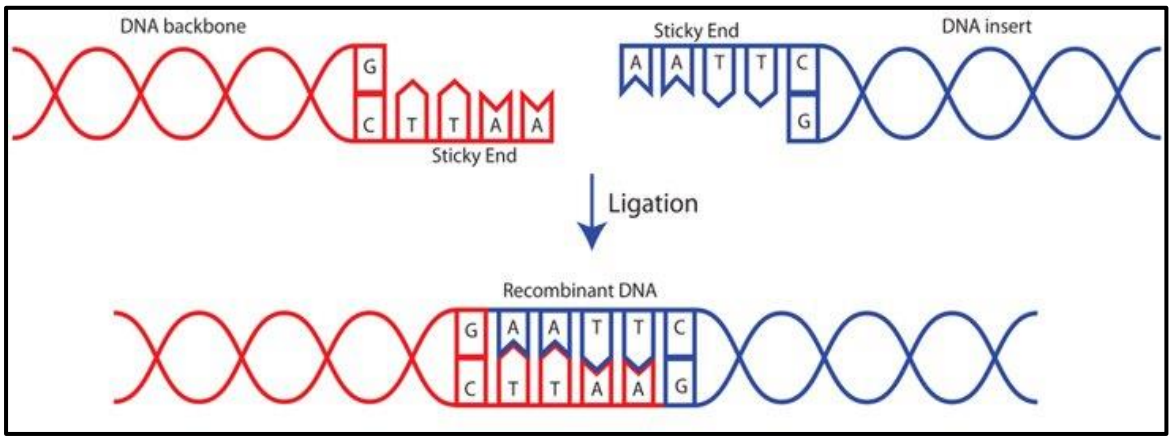
Staggered (off-center): enzymes cut the two DNA strands at different positions generating sticky or cohesive ends. (They are called sticky or cohesive because they can reform hydrogen bonds with each other).

The DNA restriction fragments would have short single-stranded overhangs at each end.

Remember the phosphodiester bond is a covalent bond (it is stable), while hydrogen bonds are not, they are reversible so the two fragments can come back and form hydrogen bonds with one another but they are not stable, so they can be dissociated or released from each other. And these ends are also known as overhangs.

Even if those sticky ends reform the H-bonds, they would still be unstable because we need to reform the covalent bond (phosphodiester). This can be accomplished by DNA ligase.

Zoom into the sticky ends

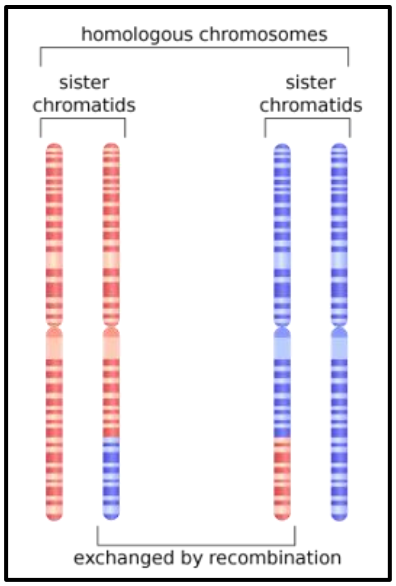


If we bring two different DNA strand from different sources and cut them by the same endonuclease, we can connect the two fragments using DNA ligase.

DNA ligase

It covalently joins DNA ends (example, restriction fragments) by catalyzing the ATP-dependent formation of phosphodiester bonds between the 3'-hydroxyl group of one strand and the 5'-phosphate end of another strand. producing a recombinant DNA.

Recombination and recombinant DNA



Recombination: Connecting or transferring a piece of DNA from whatever source (another chromosome, a short and synthetic piece of DNA, etc.) into another DNA molecule.

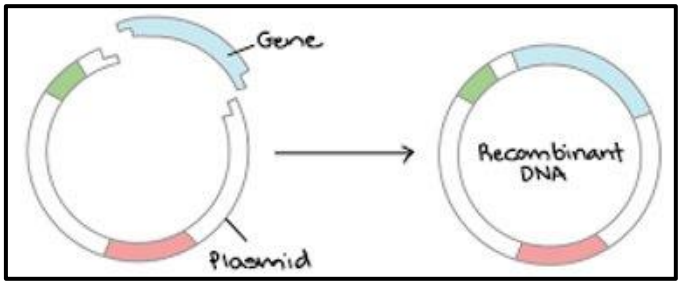
Recombinant DNA: a DNA that is made from two or more different sources.

What is DNA cloning?

Cloning: making another copy.

DNA cloning is a technique that allows for:

amplifying a DNA segment into many, many copies in a biological system.



expressing a gene inside a biological system such as bacteria, human cells grown in labs, animals, or even the human body as a whole.

It usually involves:

The formation of a recombinant DNA composed of a vector (a carrier of the gene or the DNA segment of interest; usually a bacterial plasmid) and a gene that encodes a protein or a non-coding RNA using restriction endonucleases.

We put the DNA that we want to clone inside this vector

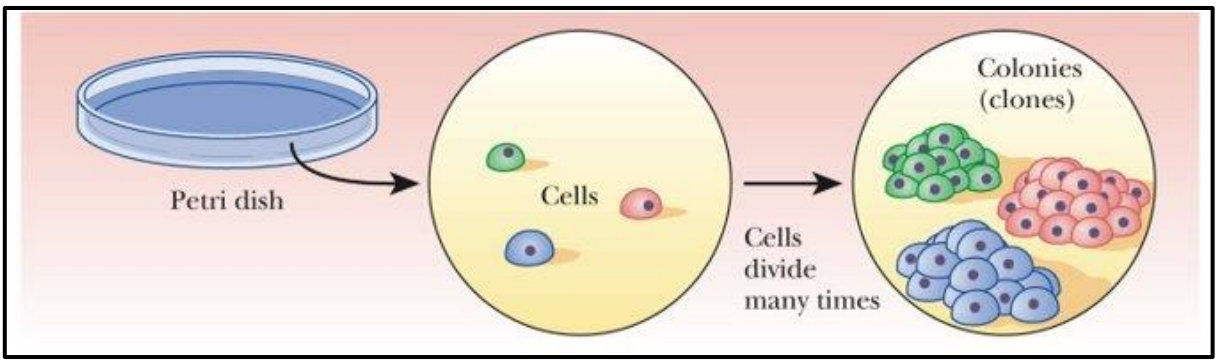
Insertion into the cell(s).

Cloning

Cloning means that you make several copies of one thing.

A clone is a genetically identical population, whether of organisms, cells, viruses, or DNA molecules.

Every member of the population is derived from a single cell, virus, or DNA molecule.



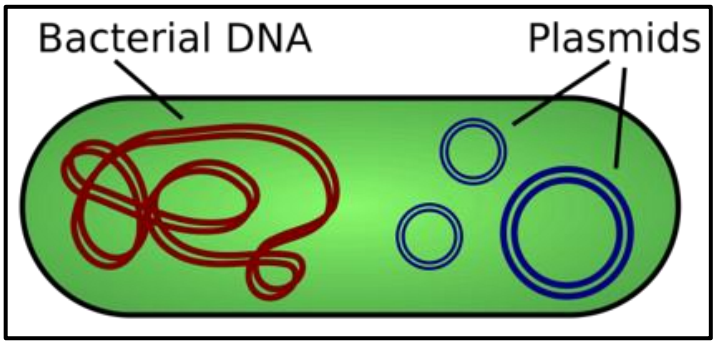
In the picture you can see single bacterial cells, that can grow into multiple cells by division, which results in a clone of cells (each group represents a clone that originates from one single cell). so human cloning means that we are making copies of the same human individual (I know it is weird to imagine it, but bear with me), they have the same genetic background. So, in the picture above, all of the cells in the group would have the same genetic background or DNA as the original one.

Using plasmids as vectors

Bacterial plasmids are natural bacterial circular DNA that is not part of the main circular DNA chromosome of the bacterium.

They are considered excellent vectors for cloning (cloning vectors).

A plasmid exists as a closed circle, and it replicates independently of the main bacterial genome.



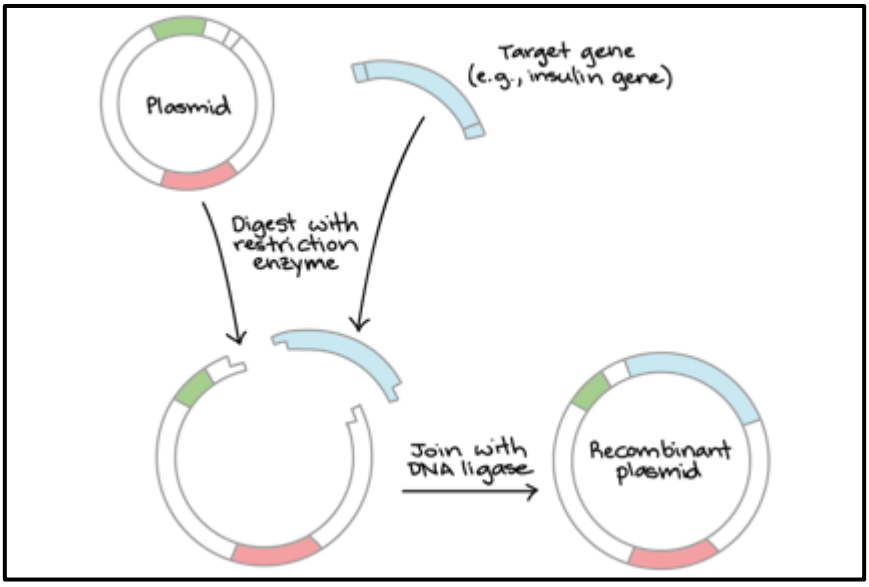
The plasmid can be transferred from one bacterial cell to another.

They contain genes that benefit bacterial cells.

They replicate independently of the bacterial chromosome, so bacterial cells can have multiple copies of the same plasmid.

How do we clone a DNA molecule?

A DNA fragment of interest (a human DNA in our example) is inserted into a plasmid that can be replicated.



it is usually a bacterial plasmid.

Firstly, we add the restriction endonuclease to the bacterial plasmid, it scans to find the restriction site and cuts the plasmid, producing sticky ends.

We add the same endonuclease to the target gene (a human gene in our example), and it produces sticky ends complementary to the ones in the plasmid.

So, they recognize each other and form hydrogen bonds with one another, but sill they are unstable, so we add DNA ligase, which reforms the phosphodiester bonds.

Let's say for example a bacterial plasmid replicates every 30 minutes, meaning that if I leave it for a day I would have millions of this plasmid carrying the DNA fragment of interest.

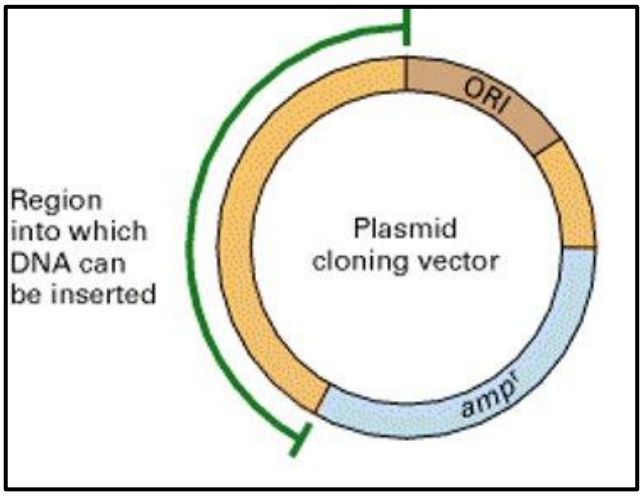
In order to extract the human DNA again from the cloned plasmids we use the same restriction endonuclease.

The resulting DNA molecule is what is known as a recombinant DNA molecule. or recombinant plasmid.

The procedure is known as recombinant DNA technology, which is part of genetic engineering.

Features of plasmid cloning vectors

Plasmid cloning vectors must have the following three components:



Their own origin of replication (OriC) that allows them to replicate independently of the bacterial chromosome.

An origin of replication is a sequence of DNA at which replication is initiated on a chromosome, plasmid or virus.

Additional info: human DNA has many origins of replication because it is larger.

A selectable gene such as an antibiotic resistance gene that makes resistant to an antibiotic and allows for selecting for the cells that have the plasmid.

it must have a selective gene that allows us to select the bacterial cells that have these plasmids, and bacterial cells that do not have these plasmids we don't want them ,An antibiotic resistance gene: is a gene that makes bacterial cells resistant to ampicillin for example, or any other antibiotic, so bacterial cells that have this plasmid would survive if we add ampicillin, and the ones that doesn't have this plasmid would die.

Bacterial cells that have the plasmid will be immune to this antibiotic.

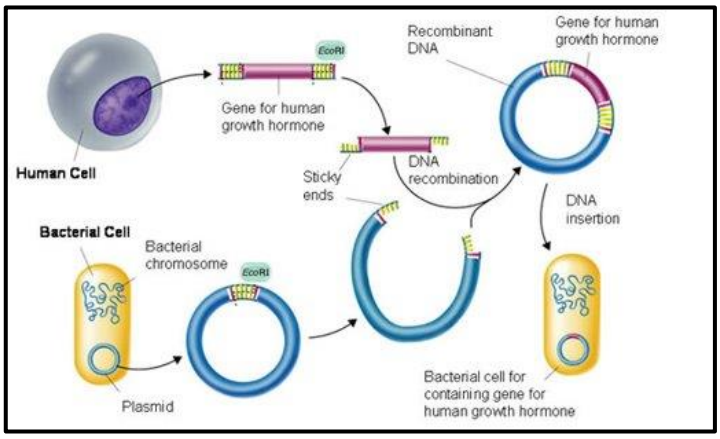
A restriction site that allows for insertion of the DNA segment of interest into the plasmid.

The endonuclease makes a single cut that opens the circular plasmid, giving us sticky ends, which allows the addition of a foreign gene.

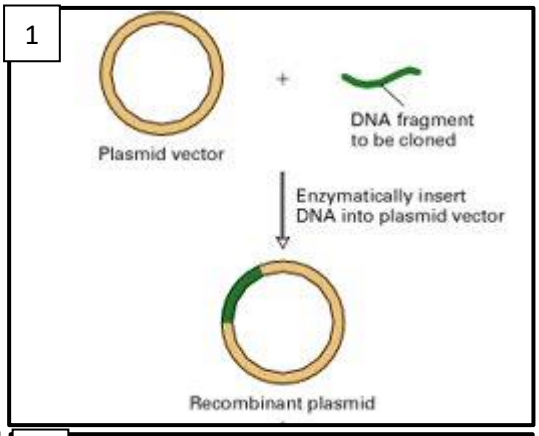
If the endonuclease makes 2 or 3 cuts, the plasmid will become fragmented, and we don't want that.

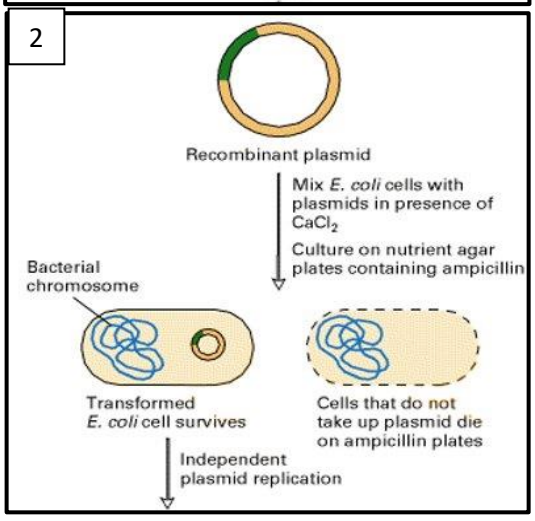
The making of a recombinant DNA

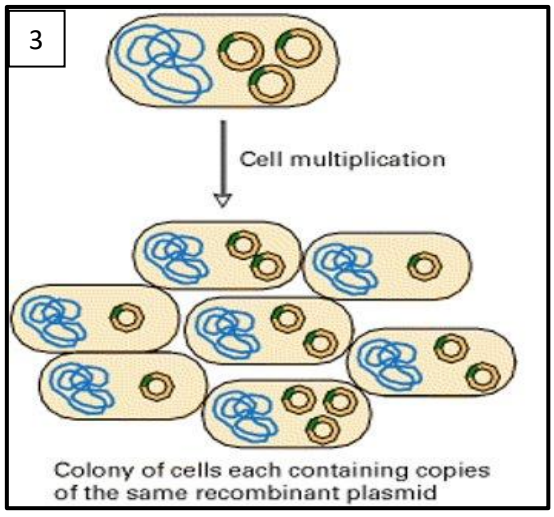
Both DNA fragments (the DNA to be cloned and a vector) are cut by the same restriction endonuclease that makes DNA fragments with same sticky-ends hybridize (anneal) to each other, when mixed.



When they are mixed, there will be a hybridization because the sticky ends of each DNA fragments (the host and the bacterial DNA) are complementary to each other.



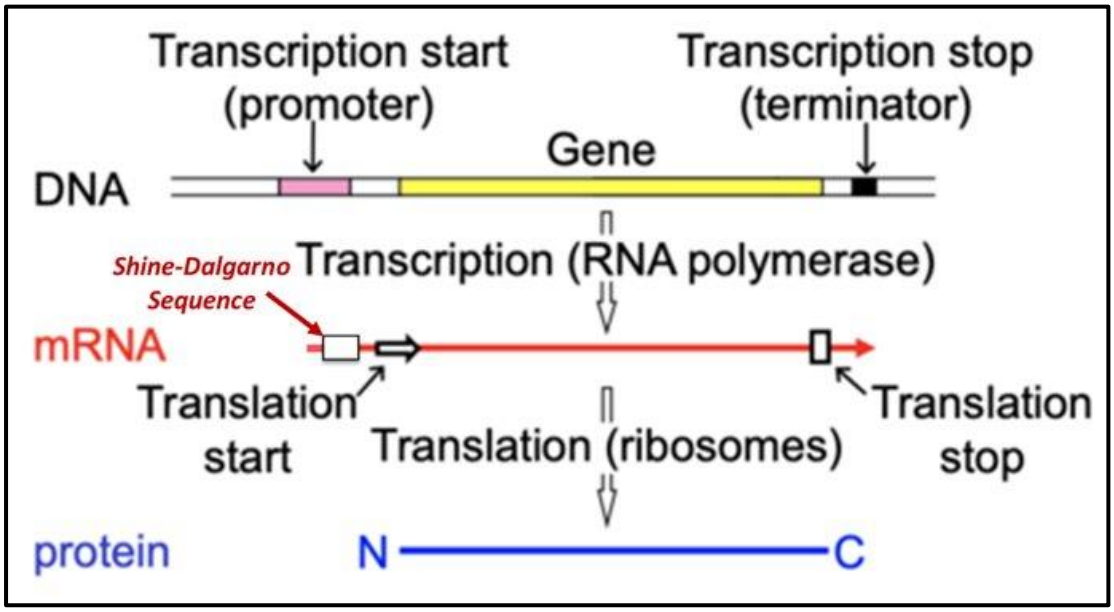




A DNA ligase is added to “close” the plasmid.

Again, we open up the plasmid by adding a restriction endonuclease and we add the same restriction endonuclease to the DNA fragment to be cloned then we add a DNA ligase to combine them, now we have the recombinant plasmid so we add it to the bacterial cell then this cell can make several copies of this plasmid (cloning)! When you add the antibiotic to the bacterial cell you will notice that the cells which have the plasmid would survive, whereas the ones which don’t have the plasmid would die because they are not resistant to the antibiotic.

Overview of gene expression



Gene: any piece of DNA that can be transcribed.

Remember that in order to express the genes you need to have a promoter.

A promoter is the binding site of **RNA polymerases** that will initiate the transcription producing mRNA.

You also need to have a termination sequence.

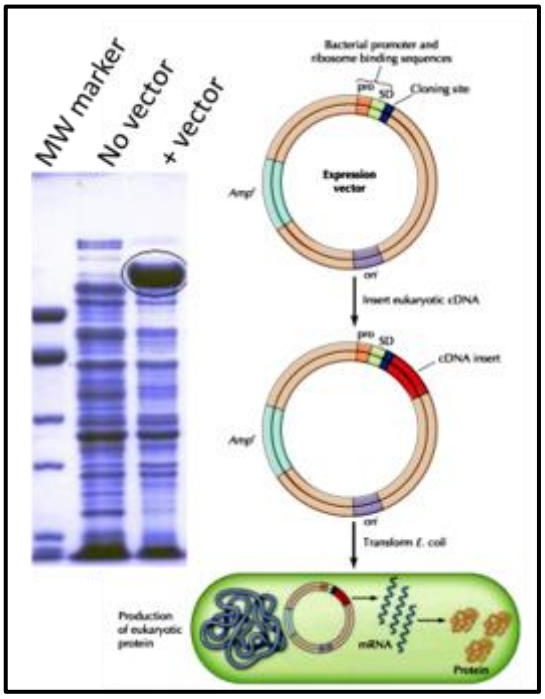
The mRNA is translated by ribosomes, again you need to have a translation start site which is AUG and a translation stop codon (UGA, UAG, UAA).

The translation moves in 5^{prime} to 3^{prime} direction.

Finally, you will get the polypeptide chain which folds forming 3D structure known as the protein.

Expression vectors

Researchers conceived the notion of employing bacterial plasmids to generate human proteins. This was achieved by incorporating the human gene into the plasmid, as detailed earlier. Subsequently, the plasmid would undertake the task of synthesizing human proteins, mimicking its intrinsic protein-production process, creating a small protein “factory”.



Expression vectors contain additional sequences:

So, in addition to the three sites discussed in the cloning vectors, in expression victors we need 3 more sequences.

Promoter sequences upstream of gene to be inserted

Ribosomal binding sequences (Shine Dalgarno [SD] sequences)

It will be transcribed but not translated, it is only used as a binding site for the ribosome and will not be a part of the peptide.

A transcription termination sequence.

The protein is expressed and purified.

Examples: insulin, growth hormone, plasminogen activator, erythropoietin.

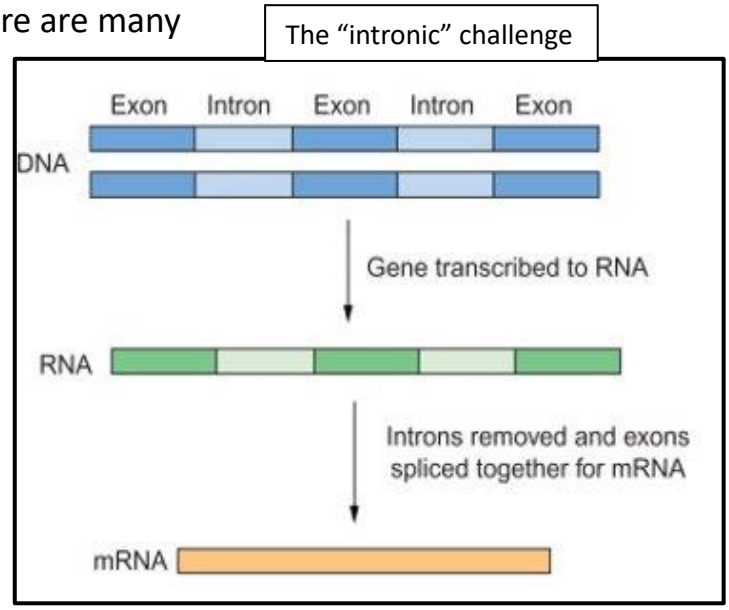
Look at the picture above, this is an SDS-PAGE, if you compare the protein content of the bacteria before and after the expression vector you will notice some new protein.

How do we select for human mRNA?

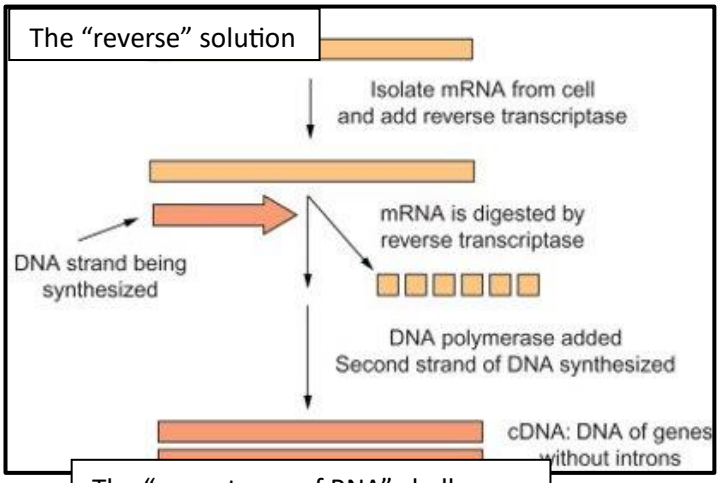
To use expression vectors, the process starts with RNA. This approach is necessary because the entire gene isn't used; rather, the focus is solely on the exons, excluding the introns. The usage of the entire gene wouldn't suffice for achieving this particular goal.

There are two challenges, first one is that there are introns inside the DNA, and bacteria can not deal with those introns.

And the second challenge is that there are many types of RNA molecules inside human cells, we have the coding RNA which is mRNA, but we also have noncoding RNA molecules, so how can we only select mRNA?

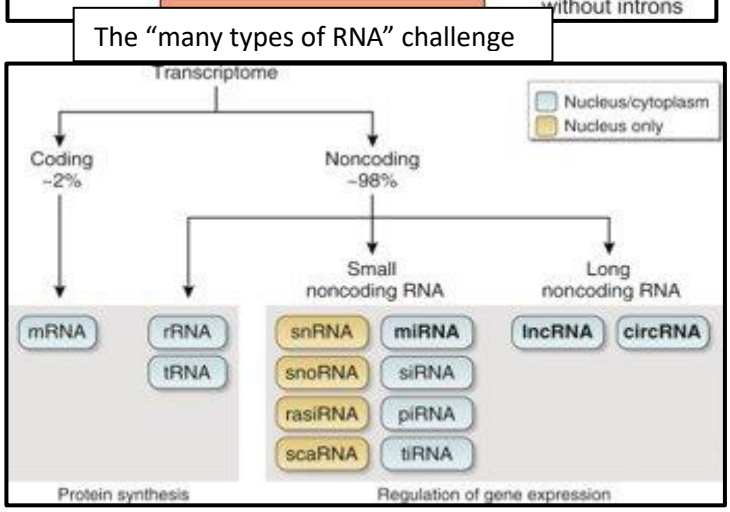


1.We use a reverse transcriptase which make a cDNA. This reverse transcriptase is found in viruses and facilitates the reverse transcription of RNA to a single DNA molecule. then by adding DNA polymerase we would have a double stranded DNA.



Now, the first challenge has been overcome through the generation of DNA from RNA. This resulting DNA exclusively contains the exons, devoid of the introns.

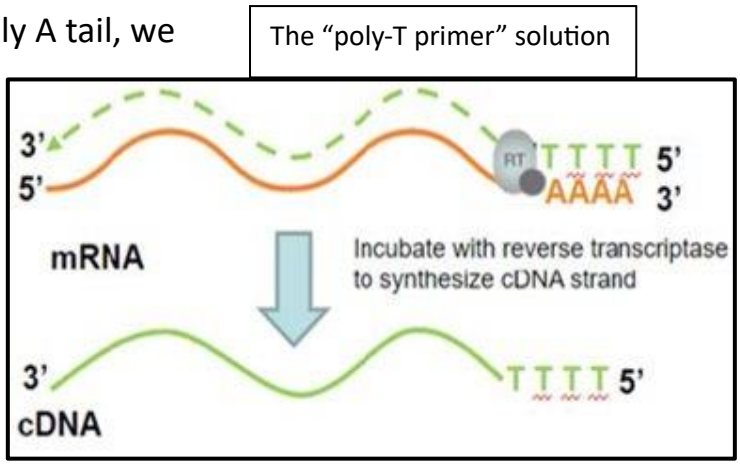
But now the second challenge arises, we have many RNA molecules, and we only want the mRNA to be reverse transcripted because it is the coding mRNA.



2.We can overcome this challenge by using the fact that mRNA contains a poly A tail.

The reverse transcriptase requires a primer, a short nucleotide sequence where the DNA polymerase starts the reverse transcription. We add a DNA primer which is a poly T primer, this poly T primer only binds to the poly A tail. So, it will bind to

the mRNA because it contains a poly A tail, we add the poly T primer and the reverse transcriptase then we get the cDNA.



how we distinguish the mRNA from other RNA types? Only mRNA contains poly A tail, so when poly T primer and reverse transcriptase are added only mRNA molecules will be reverse transcripted.

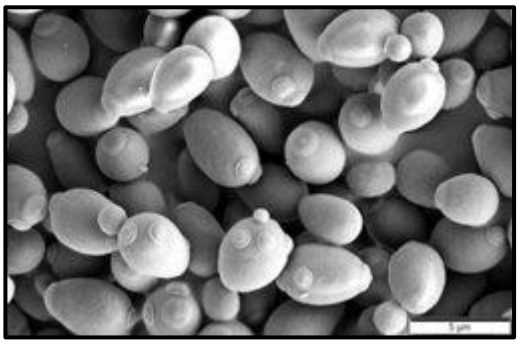
Challenges of protein expression in bacteria

No internal disulfide bonds

Bacterial cells can’t make disulfide bonds, and some proteins need disulfide bonds.

No post-translational modification (example: glycosylation)

It means adding sugars to the antibodies, in bacterial cells there is no glycosylation.



Protein misfolding

The three dimensional structure of the protein is not formed properly, some proteins need chaperons that aren’t found in bacteria.

Protein degradation

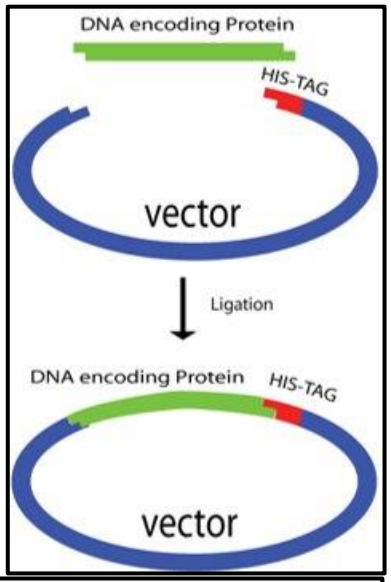
The proteins are not recognized as normal bacterial proteins, so the bacteria degrade these proteins.

Solution: use a eukaryotic system such as yeast

Yeasts are classified as single-celled eukaryotes, they can be manipulated easily. They possess the capability to establish disulfide bonds, facilitate protein folding, and exhibit rapid growth rates.

Protein tagging and creation of protein hybrids

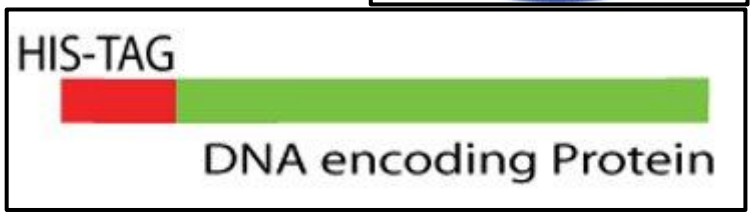
A protein-encoding gene is cloned in a special vector containing a tag gene producing a recombinant protein with an extra sequence of amino acids called tags. A tag is a small amino acid sequence that helps us recognise the protein.



These tags allow easy protein purification and detection.

This tag will not affect the folding and the structure of protein and thus will not affect its function.

In protein tagging, we use a vector containing an embedded tag segment. This rag is a short amino acid sequence.



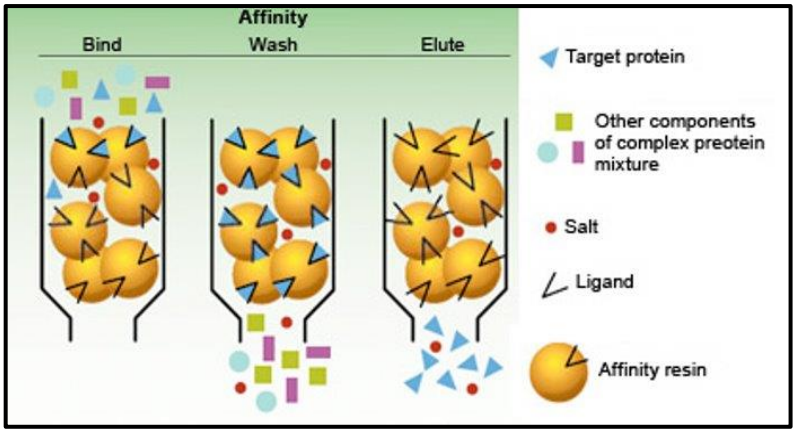
Then, we insert a gene precisely into the vector adjacent to the tag, it slots precisely into the tag's location within the vector, thereby integrating the tag into the gene sequence. Following this, the gene is expressed within bacterial cells, leading to the synthesis of both the tag and the protein. Consequently, the tag becomes a part of the protein, facilitating identifying, purifying, and isolating the target protein.

This can be used instead of having an antibody for each protein, we can have an antibody for the tag and insert the tag into the protein, and therefore it can be targeted.

Post-protein tagging

1.affinity chromatography

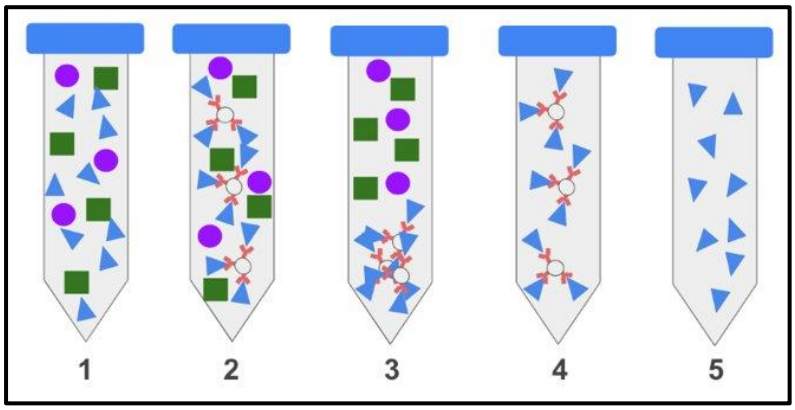
Here we have beads with antibodies attached to them, those antibodies target the tags and therefore target the tagged protein.



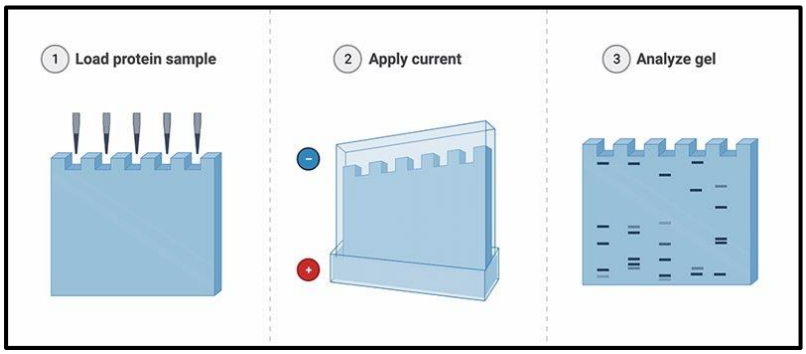
2.Immunoprecipitation

Here we are talking about antibodies which bind to proteins really specifically, in a similar way of the affinity chromatography.

we add beads which contain antibodies at the surface of them, then the antibodies will bind to the protein very specifically, these beads are heavy so they precipitate and go all the way down, we can then remove all other proteins, and simply release the proteins from the antibodies.

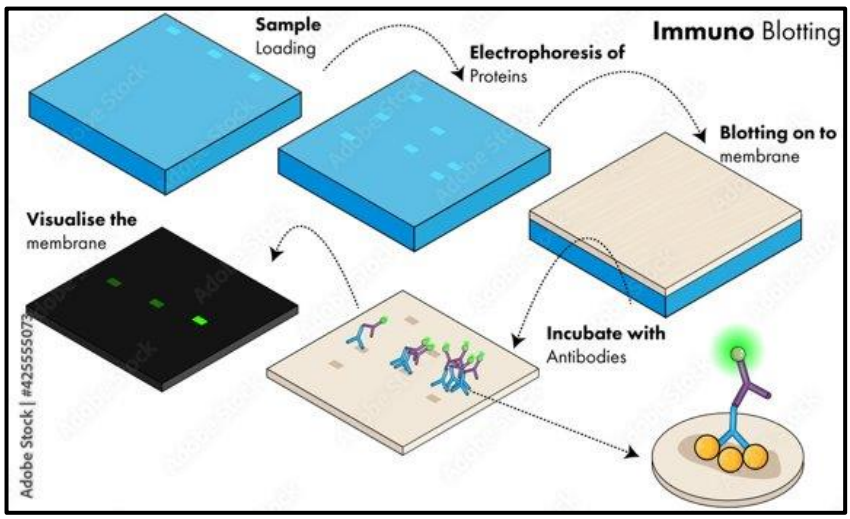


3) Gel electrophoresis (SDS-PAGE)



4) Immunoblotting

Just like southern blotting, we have our protein separated through a gel according to size, we transfer these proteins to a membrane, we then use antibodies to attach to the tagged protein and a secondary antibody which can be tracked (by its



enzymatic activity or radioactivity etc...)

Major protein tags:

Name	Amino acids	Detection	Purification
FLAG	DYKDDDDK	antibody	FLAG peptide
Green fluorescent proteins (GFP)	~220 aa protein	antibody or fluorescence	None
Glutathione S transferase (GST)	218 aa protein	antibody	glutathione
HA	YPYDVPDYA	antibody	HA peptide
Poly-His	HHHHHH	antibody	nickel, imidazole
Myc	EQKLISEED	antibody	Myc peptide
V5	GKPIPNPLLGLDST	antibody	V5 peptide

There are many tags that can be used, we will only focus on Poly-His and GFP tags.

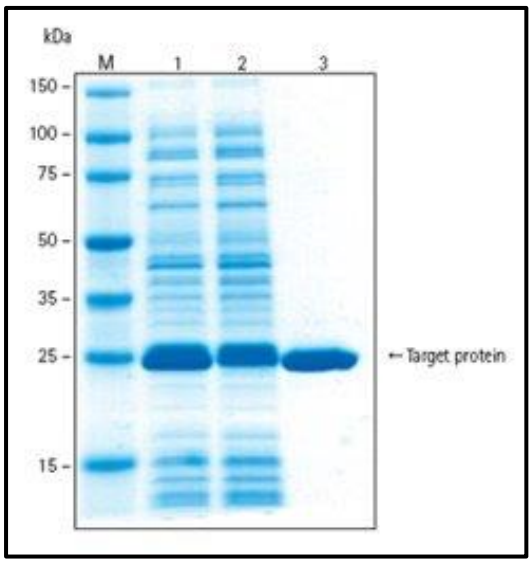
His tag

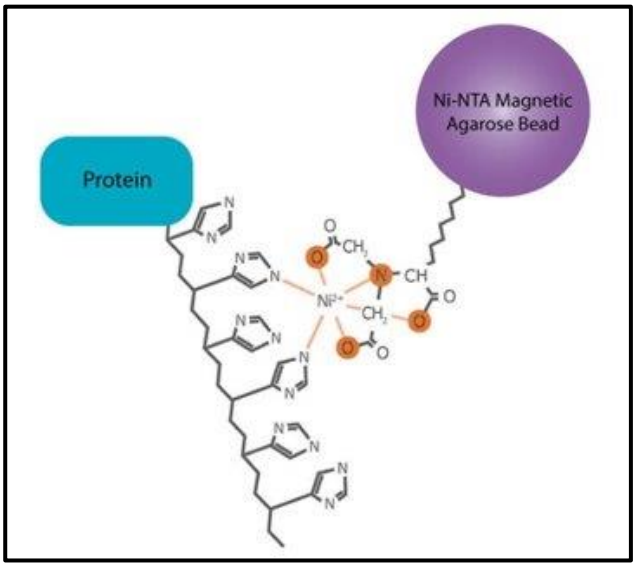
It is important because there is not a single bacterial or human protein that contain 6 Consecutive histidines. So, when I use it, it will be unique and recognizable.

The addition of six histidines to a protein would allow for purification using breads with bound nickel ions. They specifically bind to nickel ions.

So, when using affinity chromatography, we use beads covered by nickel and therefore, only proteins tagged with **poly-his** tag would bind.

Then after they bind, and we get rid of non-tagged proteins we can use nickel to wash out the bound proteins.





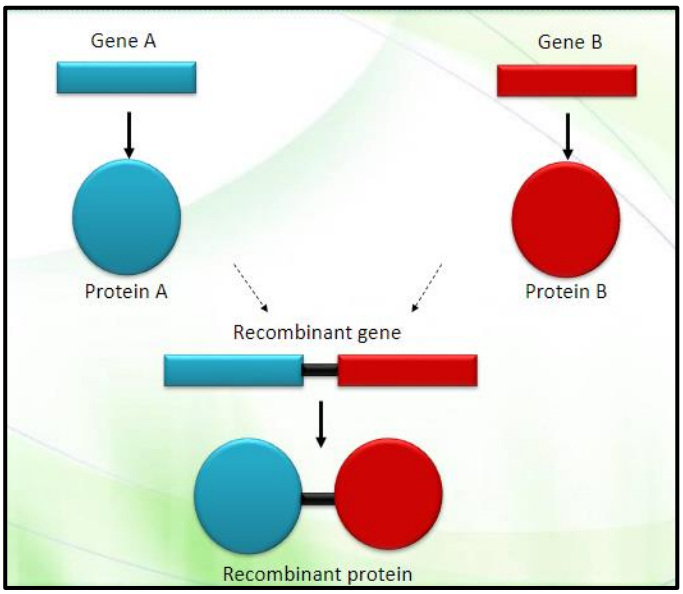
Notice the image above: we use expression vectors with tagged proteins, purify them by affinity chromatography and then use SDS-PAGE, as a result, we see the tagged protein.

Clone → Express → Purify → Analyse

Production of a recombinant protein

I bet you have heard about genetic engineering; genetic engineering is basically recombination of genes.

For example, we can recombine genes responsible for protein expression, let's say we have gene A and gene B. When these genes are transcribed within bacteria, they are transcribed as a unified mRNA, resulting in the synthesis of a singular protein. This protein is a fusion of both protein A and protein B, forming a recombinant protein product.

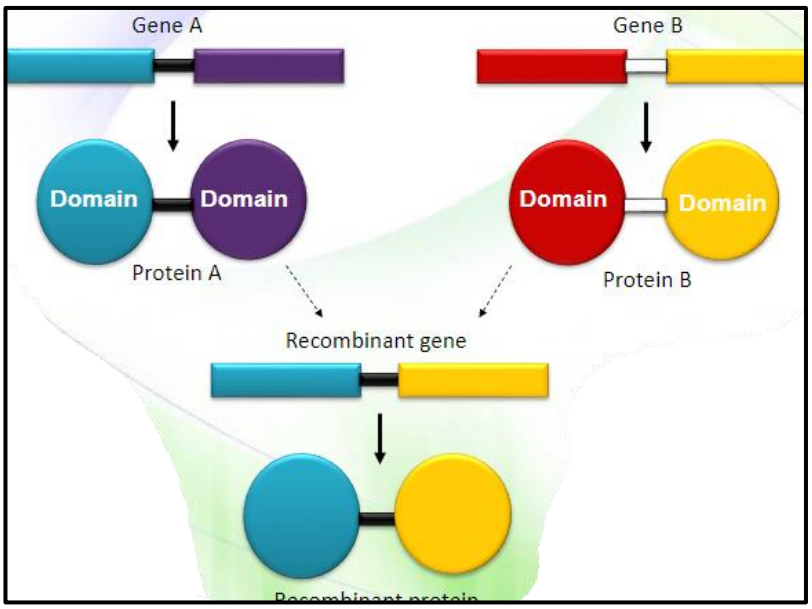


Production of a recombinant protein…The power of domains

A protein domain is a compact region (or part) of the protein's polypeptide chain that:

1- has a defined three-dimensional structure

2- is self-stabilizing



3- folds independently from the rest.

4- can be disconnected from the protein and yet maintains its structure and function.

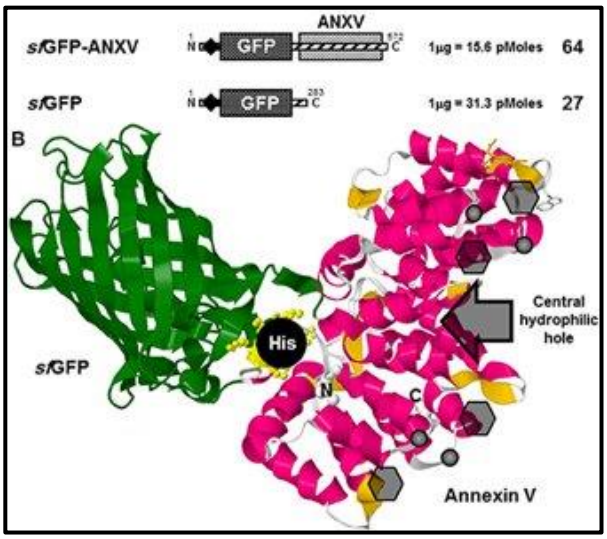
Proteins can have several domains.

So, I can recombine two domains from different proteins creating a new protein, each domain will still maintain its structure and function, this is due to the properties of domains.

GFP-tagged proteins

Green Fluorescent Protein (GFP) allows for protein detection rather than for purification purposes.

This protein is found in jellyfishes, so scientists took the gene that makes this protein and combined it with other genes creating a recombinant protein, but remember, each protein would still maintain its structure and function.



So, the GFP would still maintain its function and can be detected, and at the same time it wouldn't affect the function of the other protein.

Wherever the protein goes from now on, we can detect it through the GFP fluorescence.

It can be used for detection (by antibodies or fluorescence), but it wouldn't be beneficial for purification purposes.