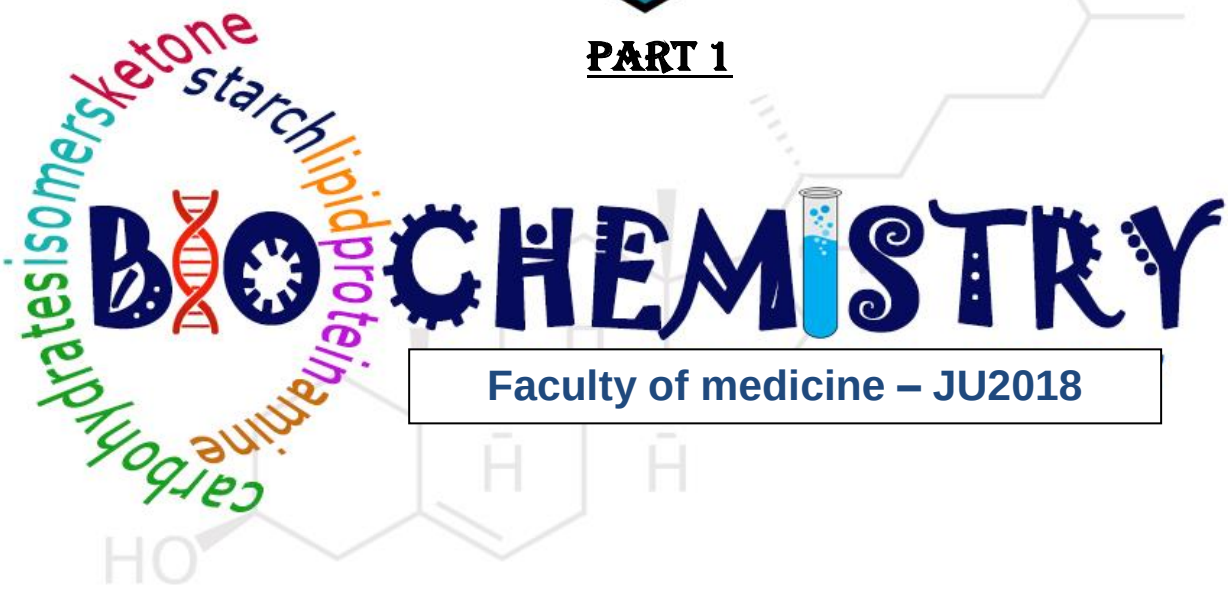
PART 1



◉Sheet ○Slides

DONE BY

Nabil G. Sweis

CONTRIBUTED IN THE SCIENTIFIC CORRECTION

Ibrahim Elhaj

CONTRIBUTED IN THE GRAMMATICAL CORRECTION

Ibrahim Elhaj

DOCTOR

Mamoun Ahram

❖ Quick Introduction:

- **Proteomics:** a field that allows us to study proteins in cells.

- **The objective** of this field is to identify and analyze a lot of different proteins present in living organisms, not just a single protein.
- One of the techniques used in this field is **"2-dimensional gel electrophoresis".**

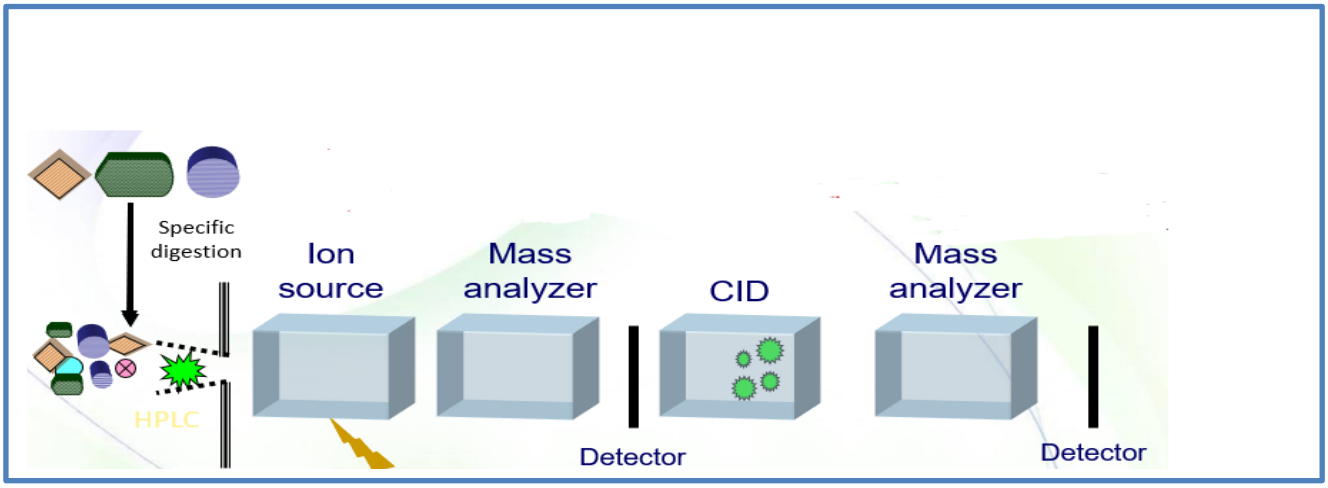
- **However, 2D gel electrophoresis has a prominent limitation:**

  - ✓ We have -on average- 6000 genes expressed in cells, many of the gene products (proteins) are modified (post-translational modifications such as glycosylation, phosphorylation, cleavage…etc). According to scientists' estimations, 100,000 different proteins emerge from these 6000 genes (notice that it is NOT a 1-to-1 ratio due to splicing and other modifications that diversify gene products).
  - ✓ From these 100,000 proteins, 2D gel electrophoresis enables us to identify around 1000 proteins which corresponds to merely 1% of the total number of different proteins we actually have.
  - ✓ This limitation called for the emergence of other proteomic techniques that would assist us in identifying a larger variety of proteins, which we will discuss next.

❖ Mass Spectrometry:

- ✓ **How does it work?**

  - **Proteins** are **digested/cleaved** into peptides (usually by using trypsin), separated by chromatography and **injected** into the mass spectrometer (a large instrument) where they get **ionized** and **travel** according to **size and charge** and reach a **detector.**
  - The **speed** of travel is measured, and then with the help of **bioinformatics**, we translate this piece of information into a protein sequence or peptide identity.



Please return to the animation which you can find in the "Lecture 31" video (starting at 3:10)

- Mass spectrometry was even further developed to what is known as **"tandem mass spectrometry".** (remember: tandem = one after the other).

- In **tandem mass spectrometry**, you have **two mass spectrometers** placed in succession (one after the other). After passing the first mass spectrometer, one of the peptides that reaches the detector in the first mass spectrometer **is further degraded** into smaller pieces and ionized again in the second mass spectrometer, and these new smaller pieces travel within the second mass spectrometer where they reach a new detector yielding new information. This new information uncovers even more detail about the protein being analyzed (you get more detail about the amino acid sequence).

- **Note: the peptides' speed of travel is dependent on two variables**:
  - ✓ **Charge** (the peptides get ionized and thus might be singly-charged, doubly-charged, triply-charged…etc).
  - ✓ **Mass.**

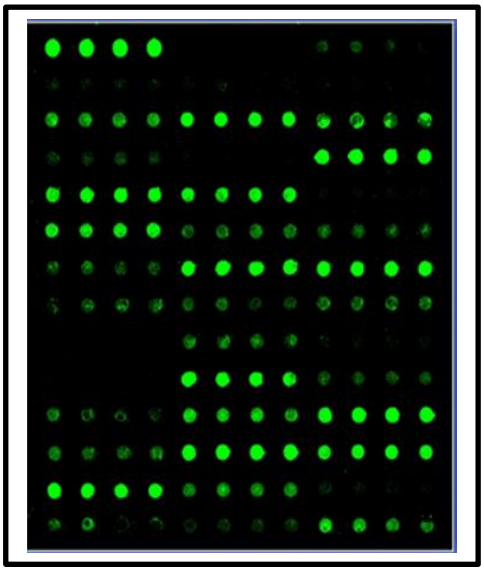
  (Each peptide travels according to its mass-to-charge ratio)

- **Note:** protein samples often contain mixtures of many different proteins which makes it harder to identify proteins. To enhance the results of mass spectrometry, we digest all of the proteins present in the sample (by trypsin) and use **chromatography techniques** to separate the resulting peptides. After that, we slowly inject them into the mass spectrometer. Some of the peptides are then selected to proceed to the second mass spectrometer (tandem mass spectrometry).

❖ Protein Arrays:

✓ Protein biochemists developed an instrument with a similar concept to DNA microarrays but specifically designed for proteins, called **"protein arrays"**.

✓ **However, protein microarrays are more complex than DNA microarrays because:**

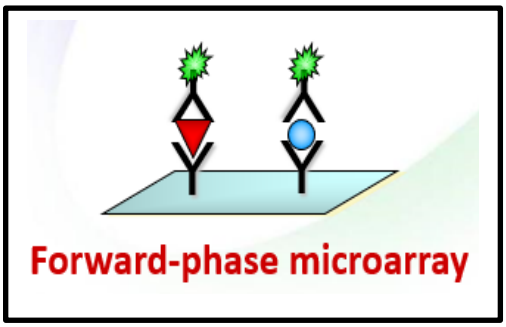


- Proteins are more **complex** than DNA (DNA is composed of only 4 nucleotides while proteins have more diverse structures as they consist of 20 amino acids and undergo various modifications like glycosylation)
- DNA does not change over time (your DNA in the morning is the same as your DNA at night), while protein expression differs with time depending on individual and environmental circumstances, and this affects their detection (high variability).

➢ Types of Protein Arrays:

A. **Expression Arrays:** these microarrays reveal whether a protein is present in a sample or not, and they include:

1. Forward-phase microarray

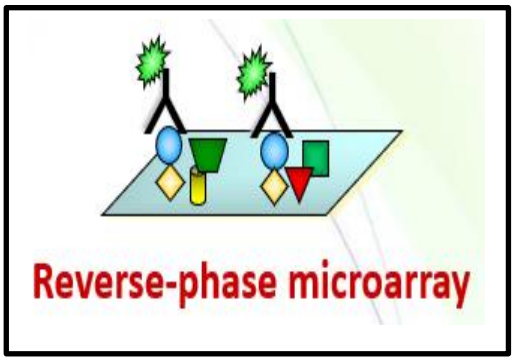


➢ In this array, we have a **solid surface** containing many spots, and each spot contains a **different** type of **antibody** that is specific to a certain protein (the same spot contains multiple copies of the same antibody). Then, we add our sample (which contains a mixture of proteins) to the microarray and we detect the proteins. The spots which emit a signal indicate that our sample contains the protein that corresponds to that spot.

➢ **Note:** The antibodies here are analogous to the probes in the DNA microarray.

2. Reverse-phase microarray

- The concept of this microarray is somehow the reverse of the previous type. Here, **each spot contains a different sample** (samples are probably taken from different patients or different tissues..etc). Then we add one type of labelled antibodies (specific to **one** protein) to the entire microarray. The spot that gives a signal indicates that the sample in that spot contains the specific protein targeted by the antibodies added.



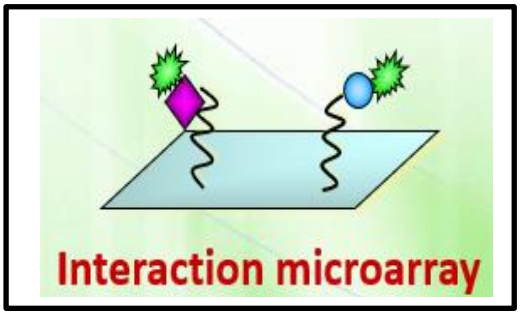
- **EXTRA: A comparison between forward-phase and reverse-phase microarrays to simplify them**

Forward-phase microarray	We are detecting the expression of **MULTIPLE proteins** in **ONE sample**	Each spot contains a **different type of antibody.**	We **add our one sample** to the entire microarray and then start detection.
Reverse-phase microarray	We are detecting the expression of **ONE protein** in **MULTIPE samples**	Each spot contains a **different sample**. (each sample contains a mixture of proteins)	We **add one type of antibodies** to the entire microarray and then start detection.

B. Functional microarrays: these microarrays serve to reveal the types of protein-protein, protein-DNA, protein-RNA, … etc interactions that a protein makes, or the enzymatic activity of a certain protein. They include:

1. Interaction microarray:

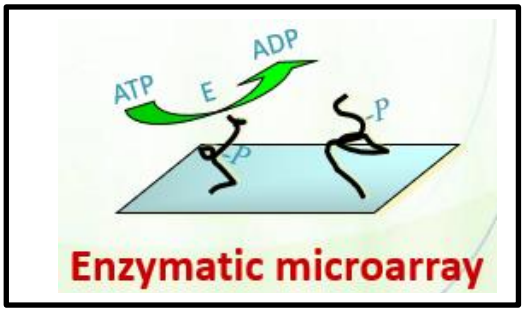
- Each spot in this microarray contains a **different and known DNA sequence**. Then we add our protein of interest to the microarray, and **we can determine which DNA sequences this protein interacts with**.



- Instead of DNA sequences, **we can have different proteins in each spot**, and then we add our protein of interest to see with which proteins it interacts.

2. Enzymatic Microarray

- Each spot in this microarray contains a **different substrate,** and then you add **one type of enzyme** to determine which substrate it acts on out of all the substrates spotted in the microarray.



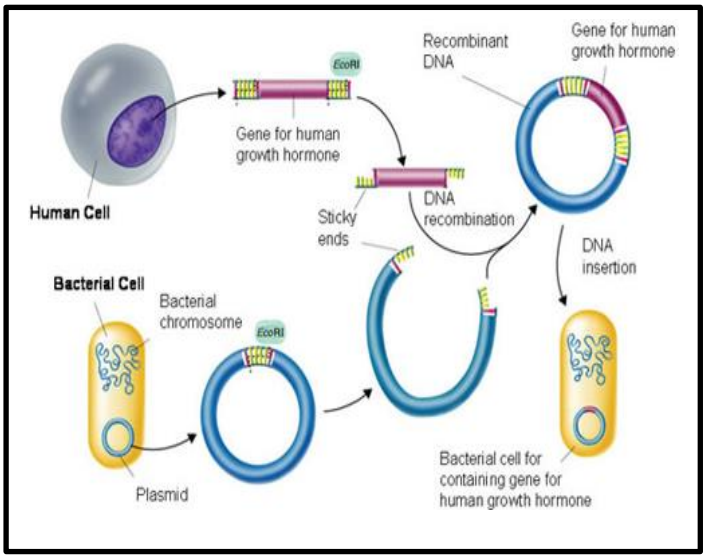
Recombinant DNA-based Molecular Techniques

- **Recombinant DNA:** a piece of DNA formed by joining DNA molecules from different sources using molecular techniques. **Example of recombinant DNA:** a plasmid (bacterial DNA) in which we inserted the insulin gene (human DNA).

❖ Cloning Vectors:

- **Vector:** a DNA molecule that can act as a carrier of foreign genetic material into another cell such as plasmids (plasmids are circular bacterial DNA molecules)

- **Cloning Vectors** are vectors that are inserted into cells in order to clone/amplify the number of copies of a certain piece of DNA. **An example is: Cloning plasmid vectors.**



➢ **Cloning vectors contain at least three essential parts for DNA cloning:**

1. **Origin of replication (they can replicate independently).**
2. **Selectable marker-antibiotic resistance gene** (it assists in **isolation/separation** of bacterial cells that contain the cloning vector from bacterial cells that do not)
   - ✓ **EXTRA: How does this separation occur?**
     We add the modified plasmids which contain the antibiotic-resistant gene to the bacterial cells. Not all bacterial cells will take up this new plasmid. In order to separate the cells that have this new plasmid from cells that do not, we add the antibiotic to all of the cells. Only bacterial cells that possess the recombinant plasmid would survive since they have the antibiotic-resistant gene. All other cells die, leaving us with the cells that have the plasmid.
3. **Insertion sites (parts that allow the insertion of a foreign DNA fragment)** (**EXTRA**: these parts are often restriction sites that allow restriction enzymes to cut and then we insert our DNA of interest).

❖ Expression Vectors:

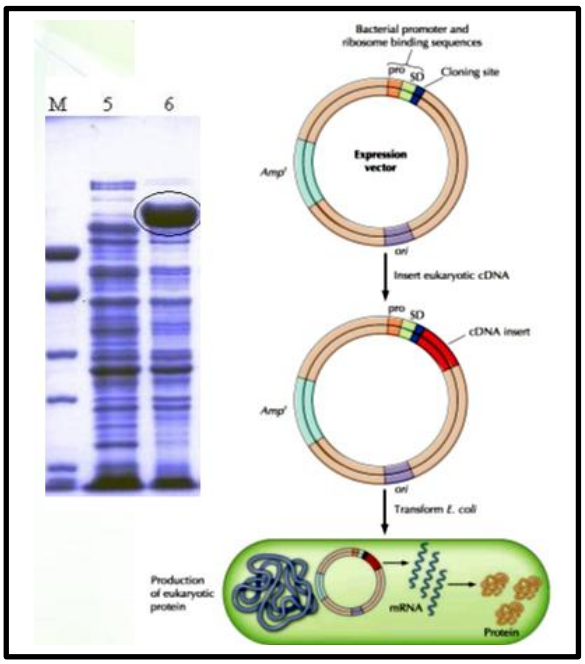
- **These vectors follow the same concept of cloning vectors, but the difference is that:**

✓ **Cloning vectors** are used to insert DNA fragments of interest into host cells so that they **make many copies** of this DNA fragment. (so that whenever we want to study this DNA fragment later on, we would have sufficient supply of it).

✓ **Expression vectors** are used to insert DNA fragments of interest into host cells so that they **express** this DNA fragment and produce gene products (proteins). Therefore, we find that expression vectors **contain additional DNA parts associated with transcription and translation** like promoters and the Shine-Dalgarno sequence, as they allow the host cell to express the gene and not merely replicate it.

- **Expression vectors contain:**

1. **Promoter sequences** upstream of the gene to be inserted.
   - ✓ The promoter does not have to be the same promoter for the gene to be inserted. It can be any promoter (We can trick cells into expressing whatever gene we want, as long as the promoter is inserted upstream of the gene. Cells do not know what gene is being expressed, they mainly care about the presence of the promoter).



2. **Ribosomal binding sequences (Shine-Dalgarno [SO] sequences)**
3. **An insertion (cloning) site.**
4. **A selectable marker (antibiotic resistance gene)**
5. **An origin of replication.**

- **Note:** notice that the expression vectors contain the same parts as the cloning vectors, but have additional parts like promoters and the SO sequence. Remember that while we want the DNA of interest to be expressed when using expression vectors, we still also want it to be replicated to further increase the copies and thus even more expression could take place.

- After our gene of interest is expressed, we can purify the proteins and use them to treat patients.

- **Examples of proteins produced using this method:** insulin, growth hormone, plasminogen activator (for blood clotting), erythropoietin (used for maturation of RBCs).

- **There are limitations to using bacterial cells as host cells for our expression vectors:**

1. Bacteria **cannot form disulfide bonds** (internal disulfide bonds in bacteria).
2. There is **no post-translational modification** of proteins like glycosylation.
3. **Misfolding** (bacterial cells do not have chaperones to fold proteins resulting in misfolded proteins).
4. **Degradation** of proteins due to misfolding (previous point).

- **Solution:** use a eukaryotic system such as yeast cells.

➢ **Yeast cells are:**

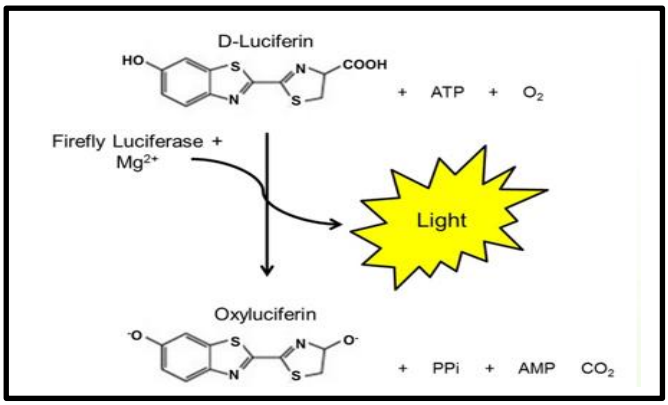
- ✓ Single-cell organisms
- ✓ Simple eukaryotes
- ✓ They can behave like bacteria (they grow fast)
- ✓ They have eukaryotic systems similar to those of human cells (they can do the things that our cells can do).

❖ Promoter analysis: Role of enzymes

- There are genes that are expressed differently in different conditions. If we want to know how the promoter of a particular gene is regulated, or what sequences are important in regulating a certain gene, we can use different systems. An important example is the use of **"fireflies"** (which are insects that emit light) to study the activity of a gene at certain conditions or elucidate the function of certain regions of the promoter.

➢ **Where does this light come from?**

✓ A reaction takes place within the bodies of fireflies, which involves the conversion of **D-Luciferin** into **Oxyluciferin** with the help of **the enzyme Luciferase**. Energy is released during the process in the form of light.



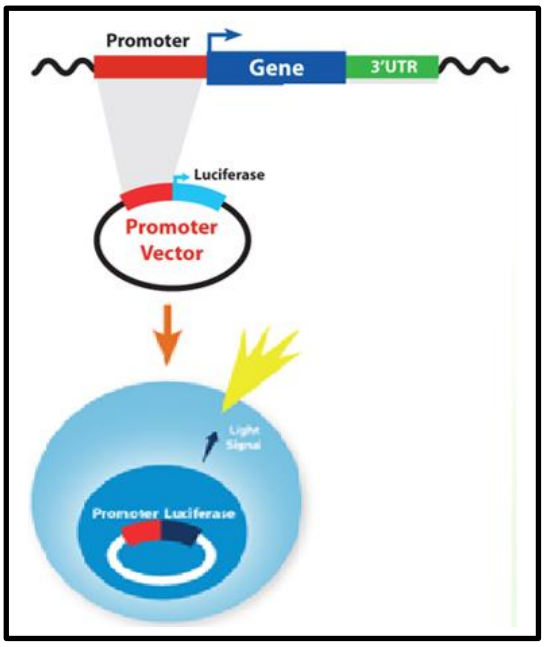
- Scientists took advantage of this phenomenon in fireflies: The **intensity of the light** emitted from a firefly can be used as an indicator of the **amount of expression** of the **Luciferase gene** (the more the expression, the higher the intensity of light). So, by introducing different promoters of interest upstream of the Luciferase gene, and cutting different sequences from the promoter, we can know which sequences affect expression and which do not by monitoring the intensity of light emitted after each experiment.

- Scientists designed an **expression vector**, where we insert the luciferase gene into the vector, and upstream of this gene we insert the promoter we wish to study.

✓ If this promoter is **inactive**, **no luciferase** is expressed, and thus no light would be detected.

✓ If the promoter is **slightly active**, **a little amount of luciferase** is produced, and **weak light** could be detected.

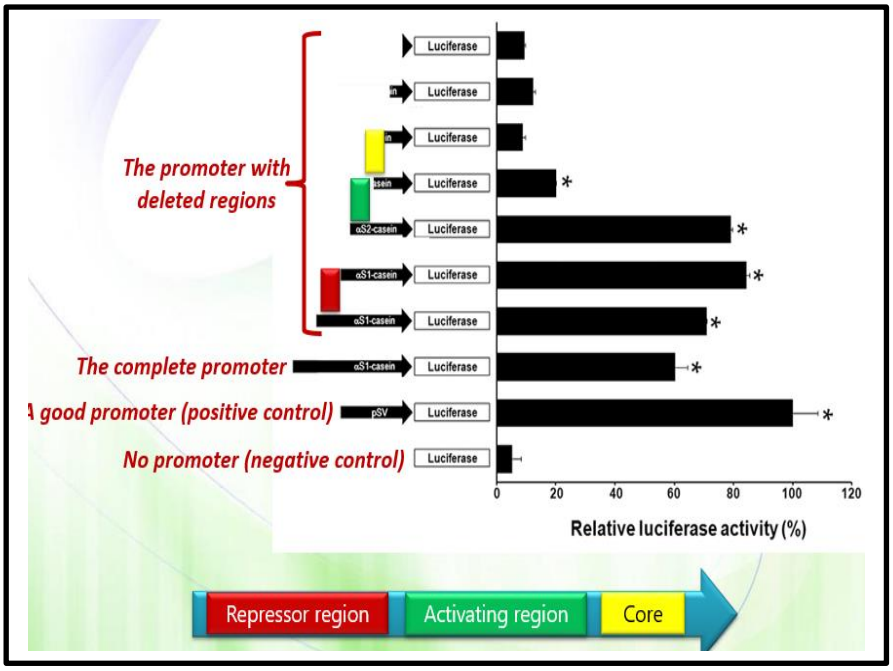
✓ If the promoter is **very active**, there will be **a lot of luciferase** and **a lot of light** would be detected.



- **Note: A reporter gene:** a gene that helps in reporting results about something. The luciferase gene is a **"reporter gene"**, because it "reports" the results of our experiment in a visible manner (light intensity) (In other words, it indicates how much expression occurred, this "report" is in the form of light intensity).
- In this experiment, we are not studying the luciferase gene, we are studying the promoter's activity in different circumstances.
- Extra note: D-Luciferin is added to the cells we are studying to guarantee that there is enough substrate for the reaction to occur and produce light.

Summary: The promoter of the gene only is placed upstream of a "reporter gene" such as the luciferase gene in a plasmid, the plasmid is transfected (inserted) into the cells, and the expression level of luciferase (instead of the gene itself) is measured.

➢ **The promoter has different regions:**



✓ The **core** promoter.
✓ The **activating** region.
✓ The **repressor** region.

➢ **Different experiments**

✓ **No promoter:** there should be almost no expression (we notice a very minimal amount of expression which is normal due to "leakage"). (negative control)

✓ **A good promoter:** we add a promoter that we know is active and functional**,** thus we notice that there is a good amount of expression. (**positive control**: meaning that we use this promoter -which we know is active and functional and will give positive results- just to make sure our experiment is being implemented correctly and that the genetic system is fine so that other experiments can take place)

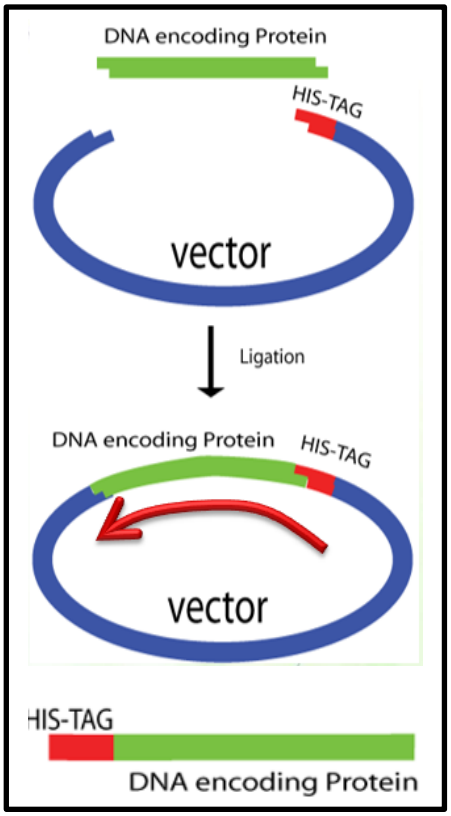
✓ **The complete promoter:** we add the entire promoter of interest containing all its original parts (core, activating and repressor regions). We notice that we have expression of the gene.

➢ **In the following experiments, we start to manipulate the DNA by chopping off parts of our promoter of interest:**

✓ **We start removing parts of the promoter in an orderly fashion:** at first, we notice that the more we remove from a certain part of the promoter, the greater the increase in the amount of expression. We deduce that the region from which we removed DNA must be a **repressor region**.
✓ **We remove even more of the promoter**: we notice a drop in the amount of expression. We conclude that we have now started to remove DNA from the **activating region**.
✓ **We further remove parts of the promoter (core promoter)** and we notice that there is even less expression (hardly any expression).

❖ Protein tagging or creation of protein hybrids:

- A **tag** is basically a **label** which allows us to **detect** or identify something.
- Tags can be introduced into proteins to facilitate their detection.

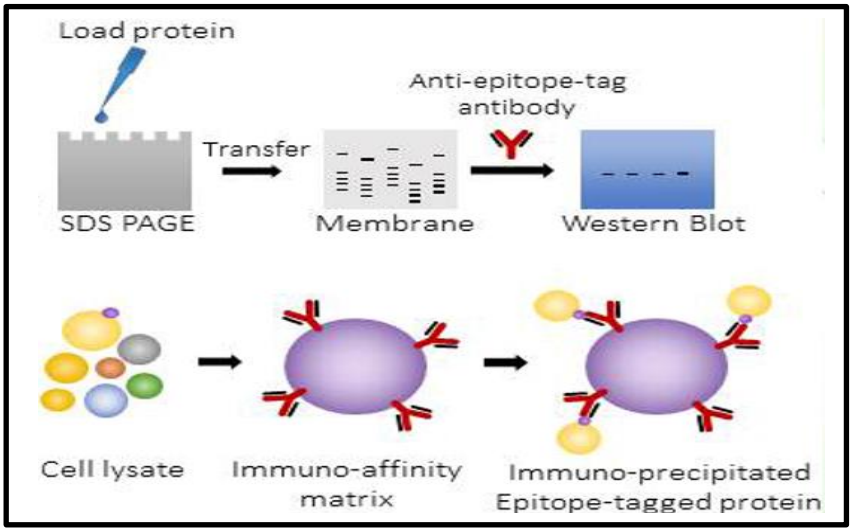


- **How do we tag proteins?**

✓ **A protein-encoding gene** (of interest) is inserted into a special vector containing a **tag gene**, so when the gene is expressed, the tag is expressed along with it. The result is the production of a **protein with an extra sequence of amino acids called tags**.

✓ These tags allow **easy protein purification and detection.**

- **How do we benefit from these tags?**
  - As mentioned before, it helps in detection as in **immunoblotting** for instance.
  - **Revision of immunoblotting**: SDS-PAGE ⟶ Transfer to a membrane ⟶ add antibodies ⟶ detection.
  - **BUT** here, our antibodies do not target the protein itself, but instead target the tag that is attached to it.
  - Additionally, we can use these tags to **purify** proteins using **affinity chromatography.** (the antibodies that target the tag are attached to the beads, and proteins that contain the tag would bind to the antibodies, while other proteins are washed away).



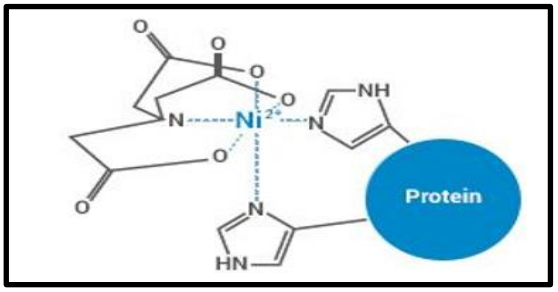
Name	Amino acids	Detection	Purification
FLAG	DYKDDDDK	antibody	FLAG peptide
Green fluorescent proteins (GFP)	~220 aa protein	antibody or fluorescence	None
Glutathione S transferase (GST)	218 aa protein	antibody	glutathione
HA	YPYDVPDYA	antibody	HA peptide
Poly-His	HHHHHH	antibody	nickel, imidazole
Myc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

➢ Comments on the previous table:

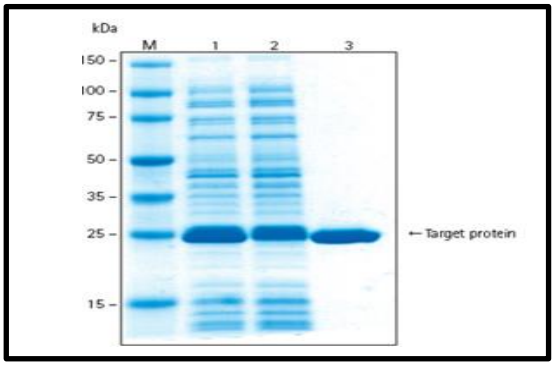
- ✓ The ones in red are the ones we will focus on **(GFP, GST, Poly-His)**
- ✓ There are different tags, some are **small** like the **Poly-His tag** (which consists of **6 Histidines**. So, the tagged protein would start with 6 Histidines, and this tag does not affect the rest of the structure or function of the protein)
- ✓ Some tags may be **large** like **GFP** and **GST**.

➢ His tag:

- ✓ We can use affinity chromatography to purify proteins that have His tags. (the addition of six histidines to a protein would allow for purification using **beads with bound nickel ions**).

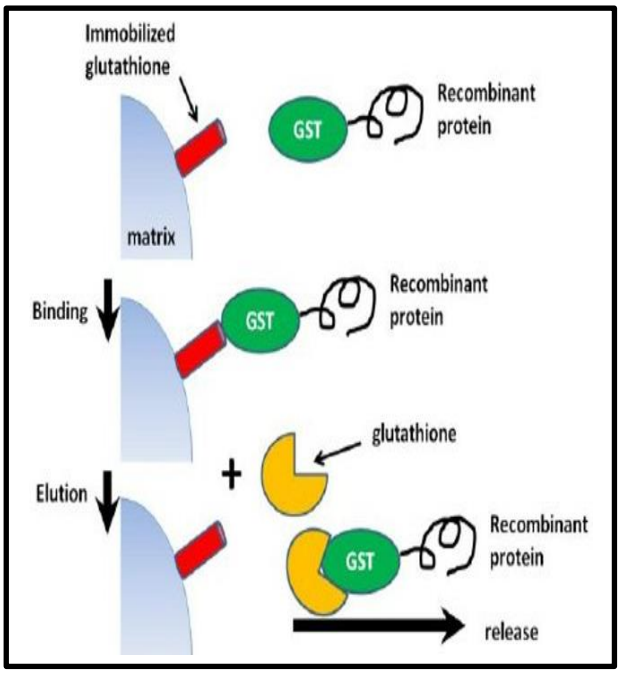


- ✓ Notice the image of gel electrophoresis to the right. Different types of proteins appear in samples 1 and 2. While in sample 3, only one protein band is shown due to purification of the protein of interest with His tags.



❖ Purification of GST tagged proteins:

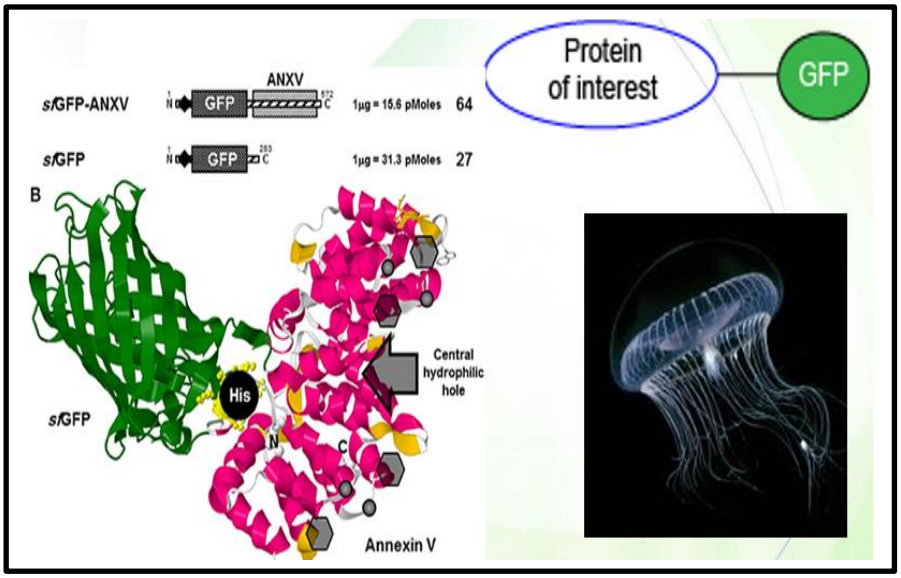
- **Protein hybrid:** a protein that consists of multiple different proteins.
- **Example:** GST-tagged proteins.
- **Glutathione S Transferase (GST)** is an enzyme that binds to a specific substrate which is **glutathione** (a tripeptide).
- We can design beads with attached glutathione residues, and when our sample passes through the beads, only GST-tagged proteins will attach to the beads. Other proteins are washed away.



- Then, we can release the GST-tagged proteins by adding free glutathione so the that the GST-tagged proteins bind to it and move along with it and are then collected.

❖ GFP-tagged proteins:

- **Another example of protein hybrids is:** GFP-tagged proteins.
- **GFP**: **G**reen **F**luorescent **P**rotein. It is a protein that comes from jellyfish (these living creatures give a fluorescent color due to production of GFP) and scientists took advantage of this protein in tagging. They were also able to separate proteins with other colors later on.
- The **GFP portion** of the GFP-tagged protein **folds independently** of the protein of interest, but is attached to it. (The same concept as **domains**). Both proteins maintain their function.



- **Remember: Domains** are regions that contain supersecondary structures (multiple secondary structures). They fold independently from the rest of the protein and retain their function.

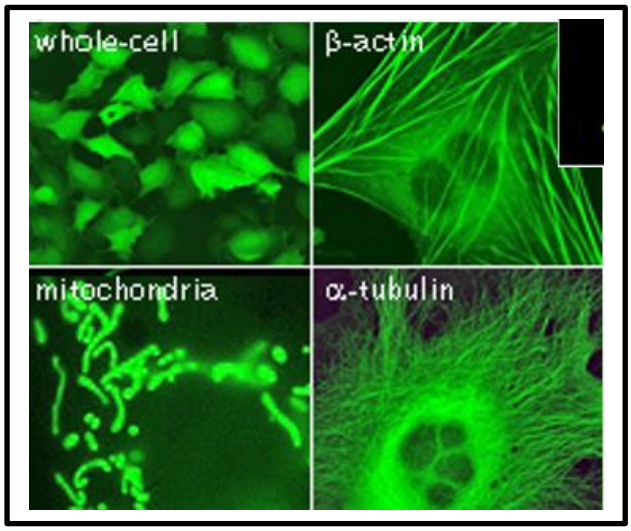
- GFP allows for protein **detection** rather than for purification purposes. (GFP is a fluorescent protein which gives a green color, so GFP-tagged proteins would obtain a green color which helps in labelling and detection).

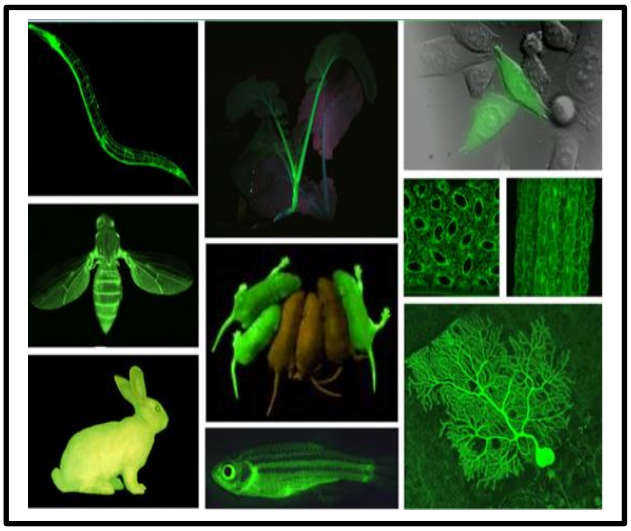
- **Examples of proteins that can be labelled with GFP:** actin, tubulin, mitochondrial proteins.

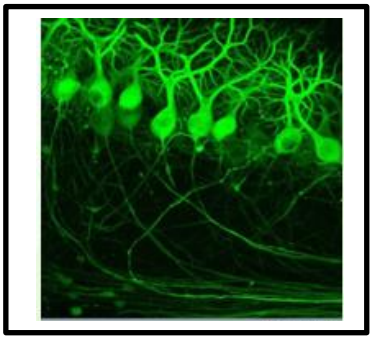
- **Whole cells** can also be labelled (like neurons, which allows us to see how they are connected to each other and study neural networks).

- **Whole animals** can be labelled as well (their health is not harmed by the process).

- So there is a world of possibilities.







MAKE SURE YOU CONTINUE TO PART 2 OF THIS SHEET.

Good Luck!