FACULTAD DE MEDICINA
UNIVERSIDAD DE CANTABRIA

GRADO EN MEDICINA

TRABAJO FIN DE GRADO

DISEÑO Y OPTIMIZACIÓN DE PROTEÍNAS TERAPÉUTICAS

DESIGN AND OPTIMIZATION OF THERAPEUTIC PROTEINS

Autor/a: PAULA ALONSO PRADO

Director/es: GABRIEL MONCALIÁN MONTES

Santander, MAYO 2023

DESIGN AND OPTIMIZATION OF THERAPEUTIC PROTEINS.

INDEX.

1.ABSTRACT.

Recombinant proteins have been in our daily lives for some years now, especially Therapeutic Recombinant Proteins that are being used for the treatment of diseases. Therapeutic Recombinant Proteins are very extended as medical treatments for various different illnesses in the context of very diverse medical specialties and they are here to stay. Nowadays, this use of therapeutic proteins is even booming, and day by day new molecules are modified and engineered to create plenty of new treatments of the so-called “personalized medicine”. Since the commercialization of Recombinant Insulin, which could be probably the most famous and used therapeutic recombinant molecule, another type of protein has become the queen of this age: Monoclonal Antibodies. Monoclonal Antibodies have reached the top of drug sales, having an important impact on current and future medicine, pharmacological industry and worldwide economy; being consequently a significant way for research and development in the area of design and optimization of therapeutic proteins.

1.RESUMEN.

Las proteínas recombinantes se encuentran presentes en nuestro día a día desde hace ya algunos años, especialmente las Proteínas Recombinantes Terapéuticas que se utilizan para el tratamiento de enfermedades. Las Proteínas Recombinantes Terapéuticas están muy extendidas como tratamiento médico para diferentes enfermedades en el contexto de especialidades médicas muy diversas, y han llegado para quedarse. Hoy en día, este uso de proteínas terapéuticas está incluso aumentando, y día a día se modifican y diseñan nuevas moléculas para crear multitud de nuevos tratamientos de la llamada "medicina personalizada". Desde la comercialización de la insulina recombinante, que podría ser probablemente la molécula recombinante terapéutica más famosa y utilizada, otro tipo de proteína se ha convertido en la reina de esta era: Los Anticuerpos Monoclonales. Los Anticuerpos Monoclonales han alcanzado la cima de las ventas de fármacos, teniendo un importante impacto en la medicina, la industria farmacológica y la economía mundial, tanto actuales como futuras; siendo en consecuencia una importante vía de investigación y desarrollo en el área del diseño y optimización de proteínas terapéuticas.

2.INTRODUCTION.

We live in the age of personalized medicine and some molecules allows us to make true this way to treat a lot of different diseases in each different patient. Proteins are polymers made up of amino acids linked by peptide bonds, forming long chains (polypeptides), which fold around themselves in several ways to form complex 3D structures. Proteins are the most abundant organic molecules present on earth, standing in every living cell, and its modification creating recombinant proteins has currently become the main approach to treat various diseases.(1) An important type of protein are antibodies, whose engineering and modification has created one of the most frequently used therapeutic proteins nowadays: Monoclonal Antibodies. As shown in **Table 1**(2), seven of the top twenty best selling drugs in 2022 are Antibodies, so this highly extended use of this molecules has had an important impact in such different spheres from medicine, like economy. As we can see in Table 1, some of the top drug sold in 2022 were COVID-19 vaccines, because of the prevention actuations to stop the pandemic. However, regarding the rest of the most current leading drug sales, a significant number of them are Monoclonal Antibodies, so this means that they are a new approach for treatment of different diseases and moreover an important progress and way of growing for pharmacological industry.

Drug name	**Manufacturers**	**2022 sales**	**Indications**
1. Comirnaty	Pfizer/BioNTech	$55,918,791,640	COVID-19 vaccine
2.Humira (adalimumab)	AbbVie	$21,237,000,000	Rheumatoid arthritis, Crohn's disease, ulcerative colitis
3.Keytruda (pembrolizumab)	Merck	$20,937,000,000	Various cancers
4. Paxlovid	Pfizer	$18,933,000,000	Prevention of severe COVID-19
5.Spikevax	Moderna	$18,435,000,000	COVID-19 vaccine
6. Eliquis (apixaban)	Bristol Myers Squibb and Pfizer	$18,269,000,000	Oral anticoagulant
7.Dupixent (dupilumab)	Sanofi Genzyme, Regeneron Pharmaceuticals	$17,417,046,200	Atopic dermatitis, asthma
8. Eylea (aflibercept)	Regeneron Pharmaceuticals, Bayer	$12,721,221,200	Age-related macular degeneration, diabetic retinopathy
9.Biktarvy	Gilead Sciences	$10,390,000,000	HIV
10.Revlimid (lenalidomide)	Bristol Myers Squibb	$9,978,000,000	Myelodysplastic syndrome, multiple myeloma
11.Stelara (ustekinumab)	Johnson & Johnson	$9,723,000,000	Ulcerative colitis and Crohn's disease

12.Imbruvica (ibrutinib)	AbbVie and Johnson & Johnson	$8,352,000,000	Chronic lymphocytic leukemia, Waldenström's macroglobulinemia
13.Opdivo (nivolumab)	Bristol Myers Squibb	$8,249,000,000	Various cancers
14.Darzalex (daratumumab)	Johnson & Johnson	$7,977,000,000	Multiple myeloma
15. Trikafta/Kaftrio	Vertex Pharmaceuticals	$7,686,800,000	Cystic fibrosis
16.Xarelto (rivaroxaban)	Johnson & Johnson/Bayer	$7,460,849,000	Oral anticoagulant
17.Trulicity (dulaglutide)	Eli Lilly	$7,439,700,000	Type 2 diabetes
18.Gardasil/Gardasil 9	Merck	$6,897,000,000	Human papillomavirus 9-valent vaccine
19.Prevnar family	Pfizer	$6,337,000,000	Pneumococcal vaccine
20.Ocrevus (ocrelizumab)	Roche	$5,764,380,000	Relapsing or primary progressive multiple sclerosis

Table 1(2)**. Best selling drugs in 2022.** Seven out of twenty are Monoclonal Antibodies (the ones whose names finish in -mab), like Humira (Adalimumab) and Keytruda (Pembrolizumab) being number 2 and 3 of the ranking, respectively.

As we can see for this year 2023, prediction of Monoclonal Antibodies sales continue growing exponentially (**Figure 1**(3)), unseating COVID-19 vaccines from the top of the podium, and becoming officially the most sold drugs in 2023.

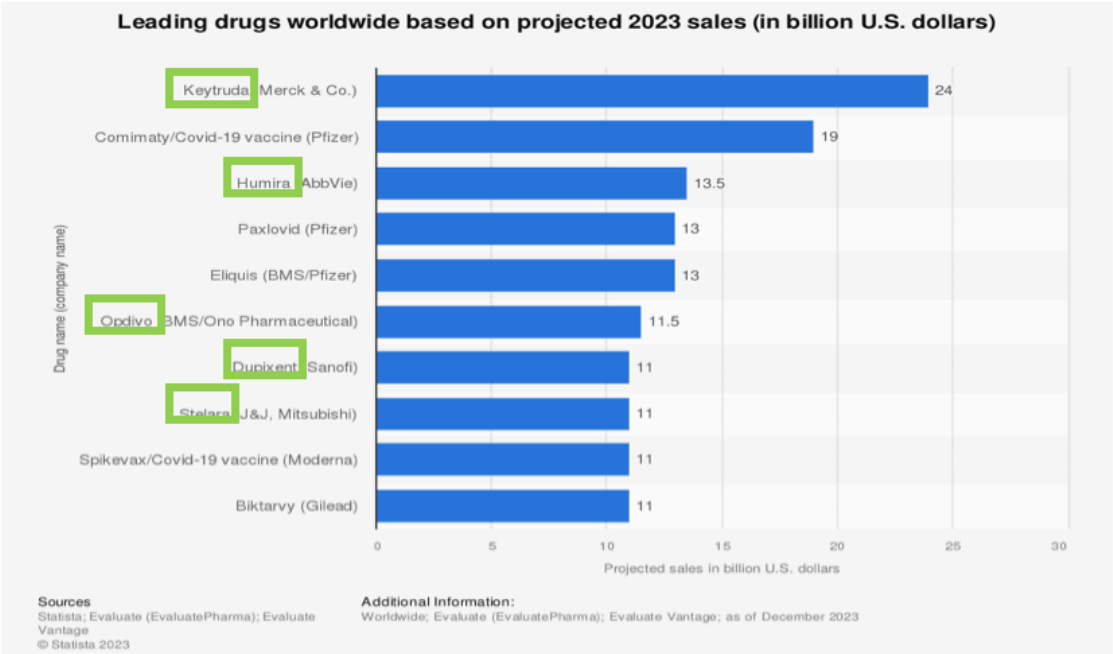

Figure 1. Prediction of best worldwide selling drugs in 2023. In the top of the ranking is projected to be the Keytruda (Pembrolizumab) an anti PD-1 Monoclonal Antibody to treat metastatic non-small cell lung cancer and melanoma, followed by other MABs like Humira (Adalimumab) and COVID-19 new oral drugs (Paxlovid, a combination of two antivirals: nirmatrelvir/ritonavir) and vaccines. Modified from (3).

This project is going to explain the importance of natural proteins, the different techniques to modify and enhance them to create the recombinant therapeutic proteins and some of their most important clinical applications, being focused mainly in the most day by day used Therapeutic Proteins in Medicine which apart from Insulin that has been extremally important and remains another one of the most used and sold drugs since long time ago are Monoclonal Antibodies, on the present and the early future of nowadays Medicine.

3.TYPES AND FUNCTIONS OF DIFFERENT PROTEINS.

Functions performed by proteins are multiple and diverse, involving almost all the processes which takes place in our body (some essential functions at the cellular level and others required for the better performance of the body as a whole). For all these reasons, they are so important molecules that let the life be. They can be divided into different categories attending to their respective function (**Figure 2**(1)):

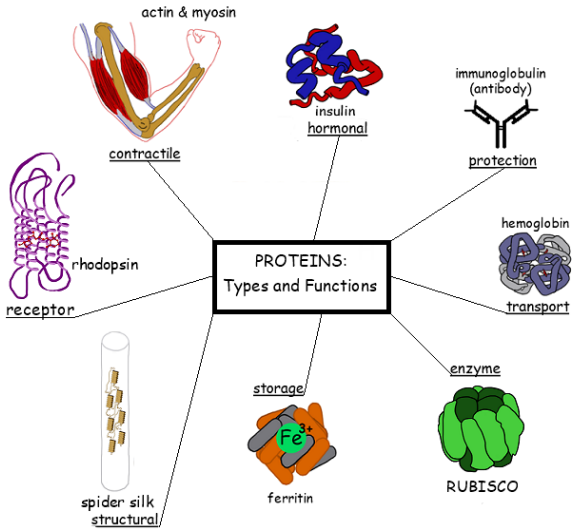

Figure 2. Types and Functions of different proteins. Main types of proteins are shown. Their respective functions are underlined. Taken from (1).

-Catalyzation of biochemical reactions: Enzymes.

Enzymes are proteins that act as biological catalysts by accelerating chemical reactions. The molecules upon which enzymes may act are called substrates, and the enzyme converts the substrates into different molecules known as products. Almost all metabolic processes in the cell need enzyme catalysis in order to occur at rates fast enough to sustain life. Metabolic pathways depend upon enzymes to catalyze individual steps. (4) We can see this in some examples such as **glycolysis** (which involves around 10 steps each requiring a particular enzyme to obtain energy from glucose) or **protein synthesis** (which involves transcription of DNA into mRN, requiring the action of **RNA**

polymerase, and then translation of RNA by ribosomes, made by Aminoacyl tRNA synthetase).(1)

-Regulation of process: Hormones.

Some hormones are also proteins, called protein hormones. Two of this protein hormones are pancreatic hormones Insulin and Glucagon, required for the regulation of blood glucose levels. **Insulin** is released when blood glucose levels are high. It promotes glucose uptake by the cells, its storage in the form of glycogen and also inhibits the gluconeogenesis. **Glucagon** is released when blood glucose levels are low. It promotes the breakdown of glycogen to release glucose and promotes gluconeogenesis too.(1)

-Protection: Antibodies.

Antibodies are the proteins that protect our bodies against harmful diseases. They are also called **immunoglobulins** and they are glycoproteins produced by plasma cells (white blood cells) in response to various disease-causing agents (antigens) entering our body. They are found in the plasma and fight against these pathogens and destroy them before it causes any disease, preventing us from their harmful effects, being an essential component of humoral immunity.(1)

-Muscle contraction.

Contractile proteins are present in muscle fibers and they are needed for contraction and relaxation of muscles. The most important are **Actin, Myosin and Troponin.**(1)

-Storage of substances.

Proteins also can be used as source to provide energy to the body in times of starvation. They act as storage substances that store thousands of amino acids, which are released from proteins when needed in the body. Some examples of this storage proteins are: **Casein** (present in milk) and **Albumin** (present in egg).(1)

-Controlling Genes' expression.

Gene expression is a process by which the information in a particular gene is copied in the form of mRNA and later, this mRNA is used by ribosomes to make the protein coded by that gene. This process is regulated by transcription factors, which are also proteins in nature. Thus, proteins regulate their own synthesis by regulating gene expression.(1)

-Transport of molecules within the body.

Proteins are necessary for the transport of various substances in the blood. Although blood acts as a transport medium, proteins are necessary to hold and transport some substances that cannot dissolve in blood. This function of proteins is also essential for the proper functioning of the body. Some examples of transport proteins present in blood are as follows:

Hemoglobin and myoglobin are the proteins required for **oxygen transport. Hemoglobin** is a protein present in the red blood cells and is responsible for the

transport of oxygen from the lungs to the tissues. Any deficiency or abnormality of hemoglobin impairs the oxygen transport by the blood, which means an interruption in oxygen supply that will result in cell death in the affected tissues. **Myoglobin** is a cytoplasmic protein having a higher affinity for oxygen molecules and its function is storing oxygen in tissues. **Albumin** is the major transport protein in blood (it is the most abundant plasma protein). It acts as a carrier for different substances as lipids, fatty acids, thyroid hormones, lipophilic drugs, heavy metals, calcium and bilirubin. **Prealbumin** is another transport protein in blood that carries steroid hormones, thyroxine, and vitamin A. **Haptoglobin** carries any free hemoglobin that is present in plasma. **Thyroxine binding protein** is specific for thyroid hormone. **Lipoproteins: HDL**, that transports cholesterol from tissues to the liver and **LDL**, that transports cholesterol from the liver to the tissues.(1)

-Components of cell membranes: receptors and transport channels.

Proteins are essential components of all the cell membranes and membranes of the organelles. One of the functions of these membrane proteins is that they act as **receptors**. Hormones, neurotransmitters, and other signaling molecules bind to these receptors and convey signals to cells, a role that is essential for the coordinated function of all our body cells. The role of proteins as receptors can be understood with the example of Insulin, that controls glucose levels in our blood, as we have seen previously, and performs its function by binding to its receptor that is a protein which sends signals for the opening of glucose channels to take up glucose from blood into the liver and muscle cells. If the **insulin receptors** are not present, insulin function cannot be made.

Proteins present in cell membranes also act as **transport channels**, needed for the entry of ions and larger-sized particles into the cells (that are not permeable directly through membranes). There are some examples of protein channels present in membranes, that allows the passage of different molecules within the cell: **Aquaporins** (water molecules), **GLUT** (glucose molecules), **Sodium channels** (sodium ions), **Potassium channels** (potassium ions), **Calcium channels** (calcium ions).

- Cytoskeleton components: Maintenance of cell shape, cell division and cytoplasmic transport.

Proteins clump together into interlinked filaments which form the cytoskeleton of the cell, being organized in the form of microtubules, microfilaments, and intermediate filaments. All these components are arranged in a particular fashion that **maintains the shape of the cell.** Important proteins that make cytoskeleton include **actin and tubulin.**

Spindle fibers are protein fibers that are involved in **cell division**, separating the chromosomes of a cell into two halves by pulling apart. Proteins are also required for the division of cytoplasm that occurs after the chromosomes have been divided.

Specific transport proteins, known as motor proteins, are needed in **intracellular transport** of different substances. These proteins use ATP to travel along the microtubules to transport various substances within the cytoplasm of a cell. An example of motor proteins is **kinesin protein**, which is involved in the transport of substances in axons of neurons. (1)

4.THERAPEUTIC RECOMBINANT PROTEINS.

4.1 WHAT MEANS THE TERM “THERAPEUTIC RECOMBINANT PROTEIN”?

Therapeutic recombinant proteins are exogenous proteins that are expressed in a production organism and used for the treatment or prevention of different diseases, produced by recombinant DNA technology.(5)

4.2 HOW TO CREATE A THERAPEUTIC RECOMBINANT PROTEIN: PROTEIN ENGINEERING. LABORATORY TECHNIQUES.

PROTEIN ENGINEERING.

The physical and chemical properties of natural proteins are sometimes not well suited to a medical, industrial, or other application. For this reason, in some cases, there are changes that can be made in a protein to better suit to a particular task. This may be obtained by using a natural gene from an organism that grows in an unusual and extreme environment, and also protein engineering techniques, as directed or random mutagenesis in which selection schemes can be used to create a mutant form of a gene that encodes a protein with the desired properties.

By using these mentioned two techniques, proteins with enhanced characteristics can be created for therapeutic or industrial applications. Some examples of these applications are going to be seen later as the improvement of the catalytic efficiency or substrate specificity of an enzyme (to decrease side effects); or the increase of thermal tolerance and pH stability of a protein (to enable the mutant protein to be used under extreme conditions that would inactivate the native version).(6)

LABORATORY TECHNIQUES.

4.2.1 Site Directed Mutagenesis.

The process for altering specific amino acids by changing the coding sequence at a targeted site in a gene is called directed mutagenesis.

In some cases, it is possible to predict in advance which specific amino acids or sequences of them contribute to a particular physical, kinetic, or chemical property of a protein, and they may be amino acids that are far apart from each other in the linear sequence, but juxtaposed as a result of the folding of the protein. Thus, this prediction of which amino acids of the protein should be changed generally requires an X-ray crystallographic analysis to characterize properly the three-dimensional structure of the protein or a similar one, but for many proteins such detailed information is often lacking, so directed mutagenesis becomes a trial-and-error strategy in which changes are made to those nucleotides that are most likely to yield the wanted particular change and then the protein encoded by each mutated gene is tested to ascertain whether the process has indeed generated the desired change. (6)

There are different options to carry out this technique:

4.2.1.1. Site-Directed Mutagenesis by Overlap Extension PCR.

This defined nucleotide substitutions into a gene can be made by overlap extension PCR. In the process are needed **four oligonucleotide primers** in the PCR; two external flanking primers anneal to the ends of the target cloned gene, and the other two overlapping, internal primers carry the mutation. It is so important that 3' end of a primer must be perfectly complementary to the annealing site on the template DNA to complete properly DNA synthesis, but mismatched nucleotides at the 5' end do not affect the reaction. The target gene is initially amplified in two separate reactions to generate overlapping fragments. After PCR amplification, the products are purified, and both fragments generated are combined. These mixed fragments are nextly denaturized and reannealed, producing DNA molecules that hybridize in the overlapping, complementary, mutated sequences. Then, DNA polymerase is added to extend the strands to form double-stranded DNA molecules. These DNA molecules are amplified by PCR with the flanking primers to enrich for full-length DNA molecules and are then cloned into a plasmid vector. This procedure (**Figure 3**(7)) results in the production of an altered gene that has mutated sites in the overlap region of internal oligonucleotides.(6)

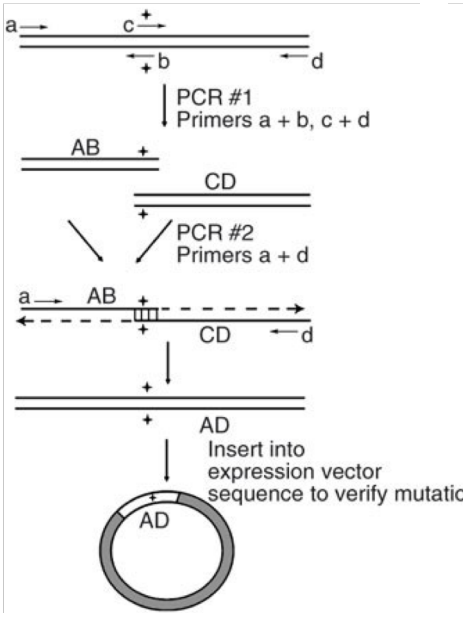

Figure 3. Scheme of Site-Directed Mutagenesis by Overlap Extension PCR. Four oligonucleotide primers (a, b, c, d) are used in this process to amplify the target gene in two separate reactions, generating the overlapping fragments (AB, CD), which are also amplified and combined (AD). This fragment (AD) is reannealed, producing DNA molecules, which are finally cloned into a plasmid vector. Taken from (7).

4.2.1.2. Site-Directed Mutagenesis by Inverse PCR.

In this case, **the entire plasmid is amplified**, which restricts the size of the plasmid to less than about 10 kb. The oligonucleotide primers used in the inverse PCR anneal in divergent orientation to adjacent sequences in the target gene, which means that their 3' ends are directed away from each other. For point mutations, nucleotide changes are introduced in the middle of one of the primers, and for mutations with long insertions, a stretch of mismatched nucleotides is added to the 5' end of one or both primers. To create deletion mutations, primers must flank the region of target DNA to be deleted and be perfectly matched to their annealing sites. PCR amplification yields linear double-

stranded DNA products that are circularized by ligation with T4 DNA ligase. This ligation requires the use of the enzyme polynucleotide kinase to phosphoryl 5' ends of the linear DNA molecules (either the primers or the PCR products), that have to be phosphorylated to produce this ligation **(Figure 4**(6)**)**.Finally, the recircularized plasmid DNA is transformed into E. coli. Since this protocol yields a very high frequency of plasmids with the desired mutation, screening three or four clones by sequencing the target DNA is usually sufficient to find one with the desired mutation. Given its simplicity and effectiveness, this procedure has come to be widely used. (6)

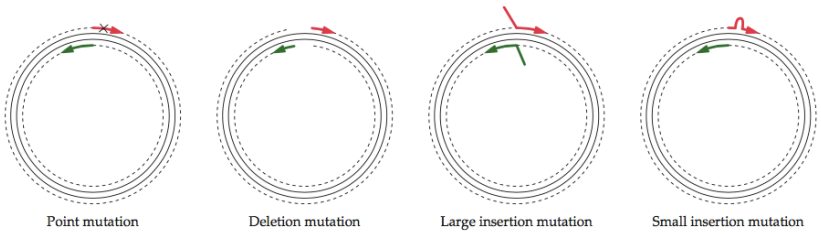

Figure 4. Site-Directed Mutagenesis by Inverse PCR. Techniques to introduce point mutations, deletions or insertions into DNA cloned into a plasmid. Forward and reverse primers are shown in red and green, respectively. Dotted lines represent newly synthesized DNA and the X refers to altered nucleotides. Taken from (6).

4.2.2 Random Mutagenesis.

In random mutagenesis, the amino acid changes that will result in the desired properties are unknown. A large number of mutant proteins, each with a different amino acid change, are generated by randomly altering individual nucleotides within a gene and then tested for the desired properties.

In those cases, a library of mutated sequences is generated, in which most of them will decrease the functioning of the encoded protein, and therefore an efficient screening process is required to identify proteins with the unusual mutations that result in beneficial changes. This approach, that often requires multiple rounds of random mutagenesis and selection to acquire a protein with the desired activity, is sometimes referred to as directed evolution.(6)

There are two experimental methods to carry out this random mutagenesis, as shown in **Figure 5**(6):

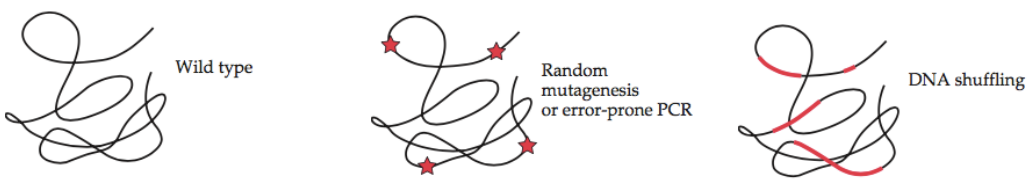

Figure 5. Different techniques for Random Mutagenesis. Wild type protein (left), its modification by error-prone PCR (center), and DNA shuffling (right). Modified from (6).

4.2.2.1 Error-Prone PCR (epPCR).

Some of the temperature-stable DNA polymerases that are used to amplify target DNA by PCR occasionally insert incorrect nucleotides during DNA replication. Then, this is a useful approach when the construction of a library of mutants of the target gene is the objective, so this is the base for the method used in Error-prone PCR random mutagenesis. Therefore, this is performed using DNA polymerases that lack proofreading activity, such as Taq DNA polymerase. The error rate may be increased by increasing the concentration of magnesium (Mg2+) to stabilize non-complementary base pairs, as well as the addition of manganese (Mnz+) and/or unequal amounts of the four deoxynucleoside triphosphates to the reaction buffer. The primer annealing sites on the template DNA define the region to be altered, and the number of nucleotide substitutions per template increases with the number of PCR cycles and the length of the template. When error-prone PCR is done, the randomly mutagenized DNA is cloned into an expression vector, introduced into a host cell, and screened for altered or improved protein activity. The DNA from those clones that encode the desired activity is isolated and sequenced to determine the relevant changes. (6)

4.2.2.2 DNA Shuffling.

Some biologically important proteins, such as α-interferon (IFN-α), are encoded by a family of several related genes, with each protein having slightly different biological activity. If all, or at least several, of the genes or cDNAs for a particular protein have been isolated, it is possible to recombine portions of these genes or DNAs to produce hybrid or chimeric forms. This DNA shuffling is done with the expectation that some of the hybrid proteins will have unique properties or activities that were not encoded in any of the original sequences and also, some of them may combine important attributes of two or more of the original proteins, like could be high activity and thermostability.

The simplest way to shuffle portions of similar genes is through the use of common restriction enzyme sites. Two or more DNAs that encode the native forms of similar proteins cut by one or more restriction enzymes in the same place, followed by ligation of the mixture of DNA fragments, can potentially generate a large number of hybrids.

Another way to shuffle DNA involves combining several members of a gene family, fragmenting the mixed DNA at random sites with deoxyribonuclease I (DNase I), and then performing a PCR. During PCR, gene fragments from different members of a gene family cross-prime each other after DNA fragments bind to one another by complementary base pairing in regions of high homology. The final products are amplified by PCR using terminal primers. After 20 to 30 PCR cycles, a large number of full-length, hybrid DNAs will be generated. The hybrid DNAs are then cloned to create a library that can be screened for the desired activity.

This technique works well with gene families or genes from different families that have a high degree of homology, but it is not especially useful when proteins have little or no homology. To remedy this situation and combine the genes of dissimilar proteins, several variations of the DNA-shuffling protocol have been described, as the so-called non-homologous random recombination. In this procedure, DNAs from different

sources (either defined or random DNA sequences, or a mixture of both) are combined and then partially digested with DNase I. These DNA fragments are made blunt ended by digestion with the enzyme T4 DNA polymerase, which both fills in 5' overhanging nucleotides and degrades 3' overhanging nucleotides. The DNA fragments are then mixed with a synthetic DNA molecule that forms a hairpin loop and contains a specific restriction enzyme site. The entire mixture is ligated by the addition of the enzyme T4 DNA ligase, which results in the formation of extended mosaic DNA molecules of variable lengths with a hairpin at each end, which is also ligated to prevent further addition of fragments to the molecules. Finally, restriction enzyme digestion removes the hairpin loops so that the resulting DNA fragments can be inserted into plasmid vectors and tested. Because this process, showed in **Figure 6**(8) randomly recombines DNA fragments, only a very small fraction of the recombined DNAs are likely to encode the desired activity.(6)

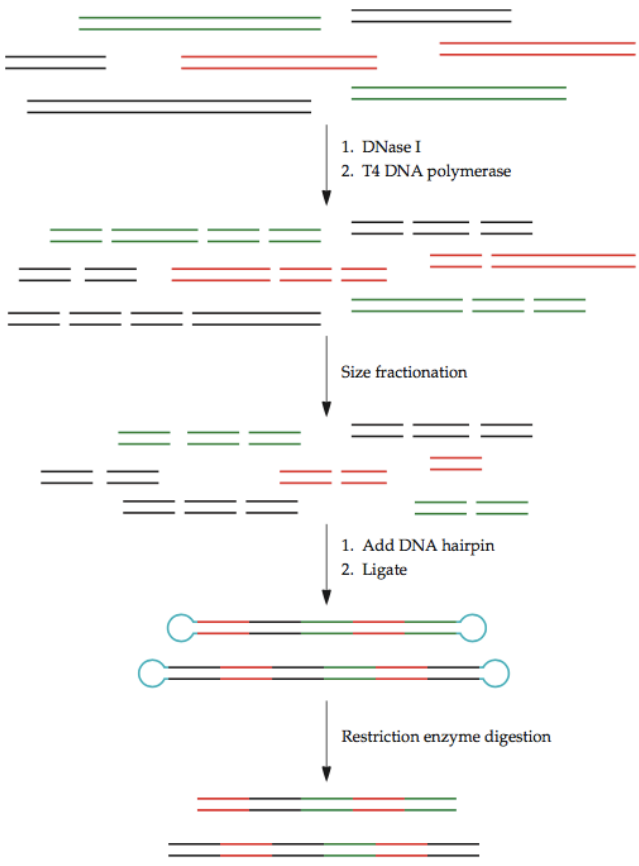

Figure 6. Schematic representation of DNA Shuffling techniques. Different DNAs shown in different colors are mixed together, digested, size fractionated, ligated with synthetic hairpin DNAs (which are later removed) and finally ligated into plasmid vectors. Taken from (8).

4.3 APPLICATIONS OF PROTEIN ENGINEERING.

Protein engineering has plenty of different applications. Some of the most used ones are described below:

-Increasing Protein Stability.

Proteins have evolved to perform a particular function for a microorganism, animal, or plant under natural conditions and are often not well suited for highly specialized biotechnology applications. Directed mutagenesis can be used to create a protein that will not readily unfold under the conditions in which it will be employed, such as high temperatures, changes in pH. **The addition of disulfide bonds**, through the introduction of specifically placed cysteines, can usually significantly increase the stability of a protein. But this addition of disulfide bond has to be made in the appropriate number, because extra disulfide bonds may perturb the normal functioning of a protein. For this reason, the activity as well as the stability of a modified protein must be tested. One example of increased stability of an engineered protein is **receptor-associated protein (RAP)**, used for the **treatment of hemophilia.** Low-density lipoprotein receptor-related protein 1 (LRP1) is a cell surface signaling protein that binds lipoproteins and other ligands and removes them from the bloodstream. The receptors also remove blood coagulation proteins, which leads to bleeding episodes in individuals with hemophilia. LRP1 is synthesized in the endoplasmic reticulum, and then escorted to the Golgi by a chaperone protein, receptor-associated protein (RAP). RAP is denatured in the acidic environment of the Golgi, thereby releasing the receptor proteins for subsequent processing and transport to the cell membrane. Exploiting its high affinity for LRP1, RAP could be administered exogenously to inhibit binding of blood coagulation proteins to the receptors and thereby prevent bleeding episodes in hemophiliacs. However, the acid sensitivity of RAP limits its potential as a therapeutic agent, because when RAP is administered, the LRP1-RAP complex that forms at the cell membrane is taken up by the cell in an endosome, which results in low pH-induced denaturation of RAP and recycling of LRP1 back to the cell surface. For this reason, researchers thought that introduction of a disulfide bond would increase the acid stability of RAP and thereby prevent its dissociation from LRP1. Computer modeling was used to predict optimal sites for introduction of cysteines, and in addition, four histidines were changed to phenylalanine to prevent histidine protonation at low pH that causes RAP to unfold and dissociate from LRP1. The two sets of mutations were introduced into RAP separately and in combination, using **oligonucleotide-directed mutagenesis.** At pH7.4, the wild-type and mutant RAP proteins were properly folded. At pH 5.5, wild-type RAP unfolded, and its affinity for LRP1 was reduced 181-told, however; mutant RAP with the disulfide bond and mutant RAP with histidines replaced with phenylalanine were substantially folded, as well as RAP with the combined mutations remained properly folded.(6)

-Modifying Protein Specificity

The enzyme **tissue plasminogen activator (tPA)** is a serine protease that is medically useful for the dissolution of blood clots. However, tPA is rapidly cleared from the circulation, so it must be administered by infusion. Therefore, to be effective with this form of delivery, high initial concentrations of tPA must be used. Unfortunately, under

these conditions, tPA can cause nonspecific internal bleeding. Thus, a long-lived tPA that has an increased specificity for fibrin in blood clots and is not prone to induce nonspecific bleeding would be desirable. These three properties were introduced by **directed mutagenesis** into the gene for the native form of tPA, creating an engineered protein being more persistent in plasma (10 times longer), much more specific for fibrin and with the same level of fibrinolytic activity than the native form. Since its approval in 2000 by the U.S. Food and Drug Administration and in 2001 by the European Medicines Agency, the triple-mutant tPA, known as **tenecteplase** and produced in CHO cells, has saved the lives of patients suffering from **acute myocardial infarctions**; it is currently being studied in phase III clinical trials for the treatment of ischemic strokes.(6)

-**Decreasing Protease Sensitivity**

The human serine protease mesotrypsin is overexpressed in many malignant tumors and is associated with invasiveness and metastasis of several cancers, including prostate, pancreatic, lung, colon, and breast cancers. It is therefore a potential target for inhibition in **cancer therapy.** One promising inhibitor is the **protease inhibitor domain from the human amyloid precursor protein (APPI).** However, similar to other potential protease inhibitors, APPI is susceptible to cleavage by mesotrypsin and does not bind to mesotrypsin with high affinity. Thus, APPI was subjected to **random mutagenesis** and screening to identify inhibitors with enhanced proteolytic stability and mesotrypsin affinity.

First, the 300-bp DNA sequence encoding APPI was assembled by PCR from six overlapping oligonucleotides that contained codons optimized for expression in yeasts (*S. cerevisiae* and *P. pastoris*) and cloned into a vector. This sequence was then used as a template for **error-prone PCR**, and the resulting PCR products were cloned into a vector and transformed into the yeast strains to generate a library of 9×10^6 clones, each carrying a variant of APPI with O to 3 mutations at random positions throughout the molecule. The mutated APPI sequences were cloned into the vector with the coding sequence for the Aga2p agglutinin cell adhesion protein so that they were expressed as fusion proteins directed to the yeast cell surface. The yeast library was first incubated with mesotrypsin to cleave susceptible APPI variants and then with a mutant form of mesotrypsin that binds to APPI but does not have proteolytic activity. Wild-type APPI was rapidly cleaved by mesotrypsin, even at low concentrations of the protease. However, the screen identified several clones displaying mutant APPIs that were resistant to all concentrations of mesotrypsin tested. DNA sequencing revealed that three amino acid changes were common among the proteolytically resistant APPIs: M17G, I18F, and F34V, where the number refers to the position of the amino acid and the letters before and after the number indicate the amino acids in the wild-type and mutants, respectively. M17G or I18F APPI mutations also showed enhanced binding to mesotrypsin. Combining these mutations resulted in an APPI triple mutant with 83-fold greater proteolytic stability and 1,500-fold greater affinity for mesotrypsin compared to wild-type APPI. The triple mutant APPI strongly inhibited invasiveness of a highly aggressive, metastatic prostate cancer cell line, in which mesotrypsin production is essential for invasive behavior. Invasiveness was not suppressed by the same concentration of wild-type APPI. These results suggest that mutant APPI may be a promising new treatment for mesotrypsin-dependent metastatic cancers.(6)

4.4 OPTIMIZATION OF THERAPEUTIC INSULIN: A SUCCESS STORY OF PROTEIN ENGINEERING.

There are some successful examples important to note, whose production was made by some of the optimization techniques described before. In this project are going to be mentioned the most used and sold recombinant proteins for medical applications. The majority of them are Monoclonal Antibodies (as shown at the beginning of the project in **Table 1**), for this reason these molecules are on the leading role (**Section 5. Currently most used therapeutic recombinant proteins: Monoclonal Antibodies**). But we cannot forget such **important therapeutic recombinant proteins as one of the classic ones, that is still used nowadays: insulin** (recombinant peptide hormone). In addition, it has to be known the use of some other therapeutic recombinant molecules as enzymes and vaccines, but we are not going to talk about them in this project.

INSULIN.
Recombinant insulin has been extremally used since it was created and nowadays it is still widely used all over the world to treat one of the most prevalent pathologies that causes a lot of morbidities if it is not well-controlled: diabetes.

Type I diabetes is a common metabolic disorder (it affects 40 to 45 million persons worldwide) in which the pancreas does not produce a sufficient amount of the hormone insulin to reduce blood sugar levels. Type I diabetics, who **must be treated with insulin**, account for approximately one-tenth of all diabetics worldwide, while type II diabetics, who do not usually require the provision of external insulin (except for those with **advanced type II diabetes**), make up the other 90% of diabetics. If left untreated, individuals with type I diabetes can expect to develop morbidities such as cardiovascular disease, stroke, kidney failure, foot ulcers, and damage to the eyes. (9) For the majority of individuals with type I diabetes, a T-cell-mediated autoimmune attack leads to the loss of insulin-producing beta cells in the islets of Langerhans in the pancreas. Since insulin is needed for the regulation of carbohydrate and fat metabolism, people who are unable to synthesize sufficient insulin require daily subcutaneous injections of **recombinant insulin.**

Oral administration of insulin would greatly simplify the lives of most insulin-dependent diabetics. However, this has not been possible due to insulin's poor stability during its passage through the gastrointestinal tract. As an alternative, researchers developed a single-chain insulin molecule that can be delivered to diabetic patients **using the bacterium L. lactis.** The peptide hormone insulin is produced in animal pancreatic cells as a single polypeptide, preproinsulin, that is processed to proinsulin following its secretion and then to insulin following proteolytic cleavage. The insulin A chain includes 21 amino acid residues, while the B chain consists of 30 amino acid residues. It was decided to synthesize a single-chain insulin in which the A and B chains are joined covalently by a 6-amino-acid-long linker peptide (because of the difficult to produce insulin in L. lactis, caused by the presence of two interchain and one intrachain disulfide bond) and four additional amino acid changes were made to the protein (introduced at the DNA level) to ensure that the single-chain insulin was as active as native insulin. The resulting single-chain insulin is more stable at high temperature and is less likely to aggregate than native insulin. (6)

-<u>Insulin Glargine (Lantus).</u>

Insulin Glargine is a long-acting modified form of medical insulin, used in the management of type I and type II diabetes. It is injected just under the skin and effects generally begin an hour after use.(10) Optimization of this therapeutic recombinant protein consists on promoting a **higher stability** (preventing Asn deamination by changing the isoelectric point of 5.4 to 6.7) and **higher solubility at acidic pH.** This is made by changing:

Its **mechanism of action (pharmacodynamics).** Insulin glargine have substitution of glycine for asparagine at A21 and two arginines added to the carboxy terminal of B chain **(Figure7**(11)**)**. The arginine amino acids shift the isoelectric point from a pH of 5.4 to 6.7, making the molecule more soluble at an acidic pH, allowing for the subcutaneous injection of a clear solution. The asparagine substitution prevents deamidization of the acid-sensitive glycine at acidic pH. In the neutral subcutaneous space, higher-order aggregates form, resulting in a slow, peakless dissolution and absorption of insulin from the site of injection. It can achieve a peakless level for at least 24 hours.

Its **acceptance and repartition in the body (pharmacokinetics).** Insulin glargine is formulated at an acidic pH 4, where it is completely water-soluble. After subcutaneous injection, when a physiologic pH (approximately 7.4) is achieved the increase in pH causes the insulin to come out of solution resulting in the formation of higher order aggregates of insulin hexamers. The higher order aggregation slows the dissociation of the hexamers into insulin monomers, the functional and physiologically active unit of insulin. This gradual process ensures that small amounts of insulin glargine are released into the body continuously, giving an almost peakless profile. (10)

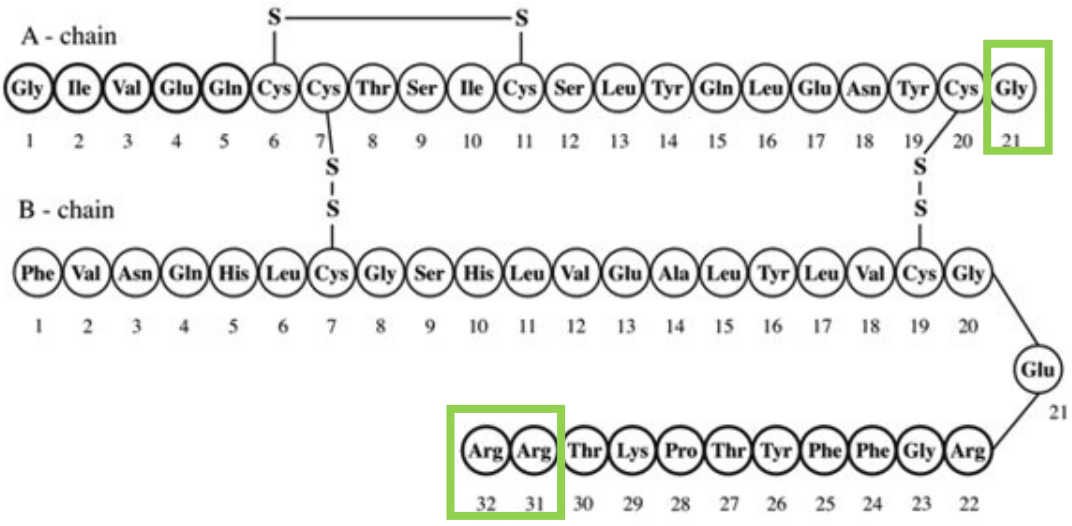

Figure 7. Insulin Glargine's structure. Optimization changes of the molecule are marked in green. Modified from(11).

-Insulin Lispro (Humilog).

Insulin lispro is a recombinant insulin analogue used to treat type 1 and type 2 diabetes. It is injected under the skin or within an insulin pump, and the onset of effects typically occurs within 30 minutes and lasts about 5 hours. It is a fast-acting insulin, which means that works faster than normal human insulin.(12)

Insulin lispro is identical to human insulin except for the **transposition of proline and lysine at positions 28 and 29 in the C-terminus of the B chain (Figure 8**(13)**).** The resultant reduced capacity for self-association in solution translates into **more rapid absorption of insulin lispro than human regular insulin from subcutaneous sites.** Maximum insulin concentrations are higher and are reached earlier with insulin lispro than with human regular insulin, and insulin concentrations return to baseline values more quickly with insulin lispro; consequently, insulin lispro has a more rapid onset and a shorter duration of glucose-lowering activity. (14)

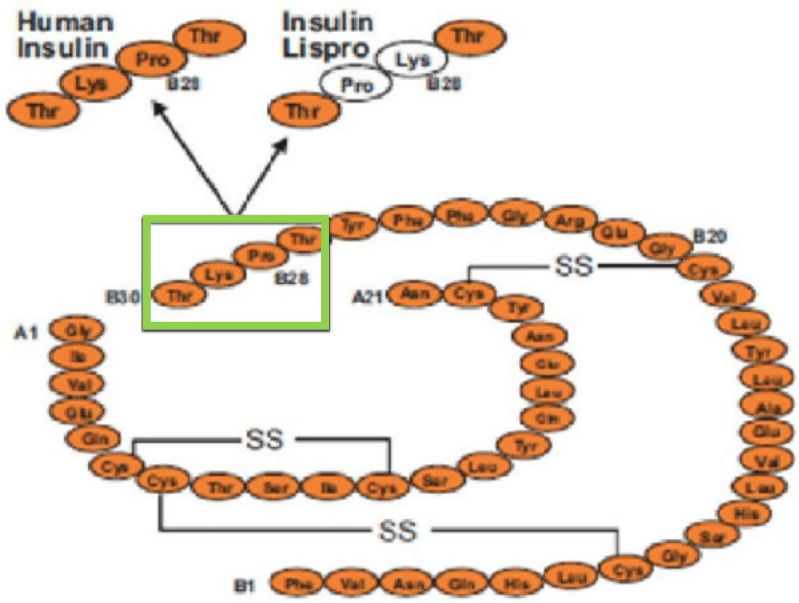

Figure 8. Insulin Lispro structure. Optimization changes of the molecule are signaled in green. Modified from(13).

5.CURRENTLY MOST USED THERAPEUTIC RECOMBINANT PROTEINS: MONOCLONAL ANTIBODIES.

Future and present of medicine has one name: personalized medicine, and this can be a reality made true by the use of some drugs like recombinant antibodies, which are one of the most sold and used drugs of XXI century and are here to stay.

5.1. STRUCTURE OF ANTIBODIES.

Antibodies, also known as immunoglobulins, are glycoproteins synthesized by B lymphocytes (plasma cells) that are activated by a specific antigen. They are secreted into the blood, lymphoid tissues, and mucosa of surface-exposed tissues such as the gastrointestinal, respiratory, and reproductive tracts, where they neutralize and eliminate extracellular microbes and toxins.

Regarding its structure, an antibody molecule has two regions with distinct functions:
- **The antigen-binding region**, that is highly variable in its amino acid sequence and for this reason is referred to as the variable (V) region. This region recognizes and binds to a specific molecule (antigen) of a pathogen that elicit the production of the antibody. The sequence variability endows an antibody with exquisite specificity of binding to a particular region of an antigen (epitope).
- **The effector region**, that is the sequence of the antibody that performs the effector functions and that does not vary among antibodies of a given class is therefore termed the constant (C) region. This region recruits immune cells such as phagocytes and molecules such as serum proteins (complement) that destroy the pathogen marked by the antibody.

An antibody molecule consists of four polypeptides, as shown in **Figure 9**(8): two identical heavy (H) protein chains (50 kDa each) and two identical light (L) protein chains (25 kDa each) held together by covalent disulfide bonds and noncovalent interactions (hydrogen bonds, electrostatic forces, van der Waals forces, and hydrophobic forces).
The V regions of the H and L chains (V_H, and V_L) together form the **antigen recognition (binding) site** of an antibody. The specific amino acids within this region that recognize and bind antigens are complementarity-determining regions (CDRs). The CDRs display the greatest variability in amino acid sequence of an antibody molecule. When the V_H and V_L domains are paired, the hypervariable regions (three in each chain, six in all) create a single hypervariable antigen-binding site. In contrast, the amino acids in the framework regions (FRs) flanking the CDRs are relatively conserved among antibodies. In addition to the V regions, each L chain contains one constant region (C_L), and each H chain contains three constant regions (C_{H1}, C_{H2}, and C_{H3}).
The effector functions of an antibody molecule are mediated by the **C regions of the H chain** and are not associated at all with the L chain. When antibodies are digested with a proteolytic enzyme, three fragments are released: two identical (Fab) fragments, each of which contains an intact L chain linked by a disulfide bond to the C_{H1} region of the H chain, and one Fc fragment, which consists of two H chain fragments, each containing the C_{H2} and C_{H3} domains and joined by a disulfide bond. **Five different classes of immunoglobulins (IgM, IgD, IgG, IgA, and IgE) are distinguishable by their Fc regions.** After antigen binding by the Fab regions (specifically, the N-terminal portions known as the Fv region) of an intact antibody molecule, the Fc portion of the molecule stimulates an immune response. Together with other immune system proteins that bind to the Fc portion of antibody molecules, foreign pathogens are inactivated.(6)

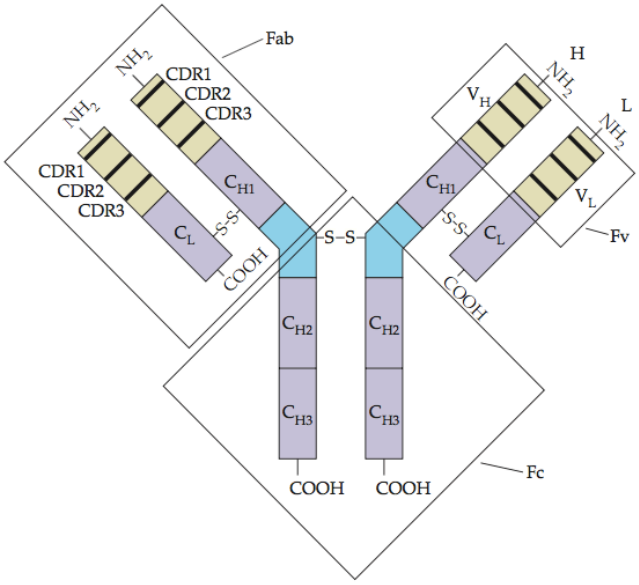

Figure 9. Structure of an antibody molecule. The H and L chains containing: their variable regions (V_L and V_H), with the CDRs (CDR_1, CDR_2, CDR_3) and their constant regions (C_L, C_{H1}, C_{H2} and C_{H3}). Taken from (8).

5.2. ENGINEERING DESIGN TECHNOLOGY OF MONOCLONAL ANTIBODIES.

5.2.1. Hybrid Human-Mouse Monoclonal Antibodies.

A Hybrid-Human Mouse Monoclonal Antibody is a mouse monoclonal antibody that has been converted into one that has some human segments but still retains its original antigen-binding specificity. Taking into account the portion of the mouse antibody that has been removed, as we can see in **Figure 10**(15), there are two types of this hybrid molecules:

- **Chimeric antibody** (approximately 70% human and 30% mouse DNA sequences). The **Fc fragment** of the mouse monoclonal antibody was replaced **with a human sequence.** It is selected to target this Fc fragment because it works poorly as an effector of immunological responses in humans, and it is also the most likely fragment to trigger the production of human antibodies against it. DNA coding sequences for the **Fv regions** of both the L and the H chains of a human immunoglobulin were substituted for the Fv DNA sequences for the L and H chains **from a specific mouse monoclonal antibody,** to diminish immunogenicity.

- **Humanized antibody** (approximately 95% human and 5% mouse DNA sequences). They are created by substituting into human antibodies only the **complementarity-determining regions (CDRs)** of the mouse monoclonal antibodies. These engineered humanized antibodies are more effective therapeutic agents and less likely to generate an undesirable immune response. This is because they have antigen-binding affinities similar to those of the original mouse monoclonal antibodies.(6)

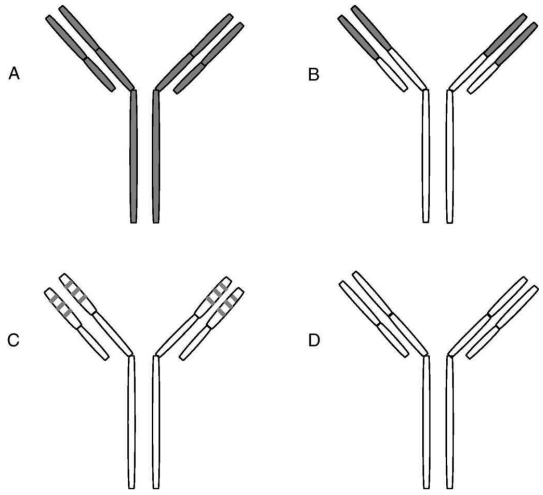

Figure 10. Engineered Antibodies. (A) Mouse antibody. (B) Chimeric antibody. It has Fv mouse fragments and Fc human fragments. (C) Humanized antibody. It is a human immunoglobulin with mouse CDRs. (D) Human antibody. Taken from (15).

Humanizing of mouse monoclonal antibodies is performed in some steps (**Figure 11**(6)):

Step 1: **cDNAs** for the vL and vH chains of a mouse hybridoma cell line **are isolated**. The variable regions of these cDNAs (involved in antigen recognition) are amplified by PCR. To delineate the limits of the CDRs, it is noticed that these regions are highly variable in sequence, while the sequences of the framework regions are relatively conserved.

Step 2: Six pairs of oligonucleotides **PCR primers** are synthesized. Each pair for each one of the **six mouse CDRs** (three from the L chain and three from the H chain).

Step 3: **Overlap extension PCR is used to replace**, one at a time, the complete DNA sequence for each of the **human CDRs with** the amplified DNA for the **mouse CDRs.** Thus, it is carried out in six cycles of overlap extension PCR, grafting the mouse CDRs onto the human antibody framework.

Step 4: The humanized variable-region cDNAs are then **cloned into expression vectors**, which are then introduced into appropriate host cells, usually Chinese hamster ovary (CHO) cells, for the production of antibodies.(6)

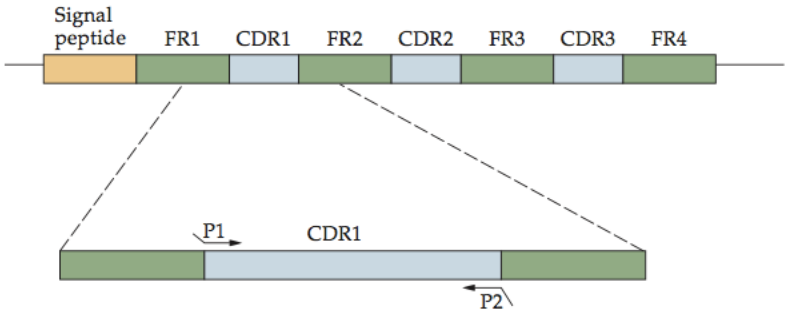

Figure 11. PCR amplification of CDR1 from a rodent monoclonal antibody L-chain DNA. PCR primers (P1 and P2), contain oligonucleotides complementary to the rodent CDR1 DNA and oligonucleotides (at its 5' ends) complementary to the human FRs L-chain DNA. Taken from (6).

To date, there are plenty different monoclonal antibodies that have been humanized. One of these examples was Obinutuzumab (Gazyva or GA101) that was approved by the FDA long time ago (on November 2013). **Obinutuzumab** is a **fully humanized monoclonal antibody** for the treatment of CD20-positive B-cell non-Hodgkin's lymphoma. It is directed against the CD20 protein found on B cells, like the famous **Rituximab** (a previously approved **chimeric monoclonal antibody** with the same target which was one of the first MABs commercialized and one of the most famous ones that is still used). For this reason, it is seen by **Genentech** (the manufacturer of Rituximab) as its commercial successor. When both molecules were compared directly, patients treated with Obinutuzumab had significantly longer periods of remission when they were disease free and did not require treatment.(6)

5.2.2. Human Monoclonal Antibodies.

Although most of the immunotherapeutic agents that have been developed have been effective, **there are drawbacks** to the use of monoclonal antibodies that contain **nonhuman sequences.** Some of the most important problems caused by these nonhuman sequences are **immunological cross-reactivity and sensitization of the patient** and these problems appears specially when multiple treatments are required, which is often the case. For this reason, it is desirable that the antibody contain no, or only a very limited amount of, nonhuman sequences.

Unfortunately, it is very difficult to create human monoclonal antibodies, for a number of **technical and ethical reasons.** Therefore, it has been necessary to devise other approaches for obtaining human monoclonal antibodies. To address this need, researchers constructed a "XenoMouse" (**Figure 12**(8)) in which the mouse antibody production genes are deleted and all of the human immunoglobulin loci are integrated into a mouse chromosome. The human H-chain and L-chain genes were cloned into a yeast artificial chromosome (YAC). The YAC with the human immunoglobulin genes (contained in yeast spheroplasts, which are cells from which the cell wall has been removed) were then introduced into mouse embryonic stem cells. These transfected cells were used to generate mice containing human immunoglobulin gene loci. Crossbreeding of two mouse lines, one carrying both mouse and human immunoglobulin genes and the other carrying only the deleted mouse immunoglobulin genes, produced a mouse strain (the so-called **XenoMouse) that expresses only human immunoglobulins.** (6) It is now possible, after immunization of a XenoMouse with a particular antigen, to produce a fully human immunoglobulin, many of which have been approved for human use (One of the first ones was Adalimumab (Humira), approved in 2002 to treat Rheumatoid Arthritis and that is still used, in fact, it is currently one of the most used (as shown at the beginning in section **2.INTRODUCTION** (**Table 1**(2))).

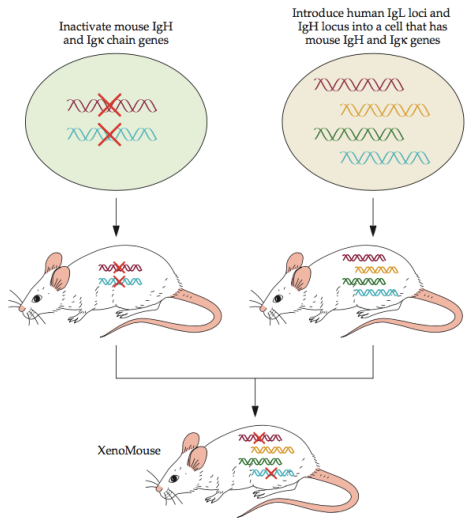

Figure 12. XenoMouse. It is a crossed-line strain of one deleted mouse immunoglobulin genes' mouse and one mouse and human immunoglobulin genes' mouse. This XenoMouse expresses only human immunoglobulins. Taken from (8).

5.2.3. Antibody Fragments.

Immunoglobulin G (IgG) is the main antibody found in mammalian serum, and it is the native form that is almost exclusively used in therapeutic antibodies. The fact that IgG molecules have two identical sites (Fv regions) that bind to two identical antigens (it means they are bivalent) generally increases their effectiveness in vivo. While the Fc portion of the IgG molecule is important in recruiting cytotoxic effector functions through complement (an amplifying cascade of proteins that facilitates pathogen destruction) or interaction with specific receptors, Fc-mediated effects are not necessary for all applications and may even sometimes be undesirable. For this reason, researchers concluded that Fc fragments (shown in **Figure 13**(6)) could be split off from the IgG and a variety of **IgG derivatives or fragments** could be constructed by **manipulating portions of the L- and H-chain cDNAs,** that may be used instead of whole antibody. Some of these molecules, because of their small size, bind more efficiently to targets that are inaccessible to conventional whole antibodies. Others have some different advantages like having multiple sites for binding to the same antigen, or binding specificities for two or more target antigens. (6)

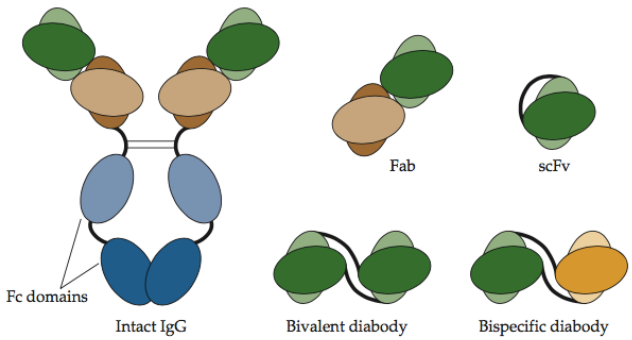

Figure 13. Schematic representation of antibody fragments. Taken from (6).

Antigen-binding single-protein chains (scFv), consisting of only VL and VH domains, may be used for a variety of **therapeutic and diagnostic applications** in which the Fc effector functions are not required and **when small size is an advantage.** These molecules have a molecular mass of approximately 27 kDa, compared with approximately 150 kDa for IgG molecules. For this reason, single-chain antibodies can penetrate and distribute more efficiently in large tumours than whole antibodies.

Drugs that are effective when tested in cell culture are often much less potent in a whole organism, because of the inability of the drug to reach the targeted site in the whole animal at a concentration sufficient to be effective. Increasing the dose of the drug often cause problematic side effects, so it is usually not the answer to this problem. However, **one successful strategy to enhance the delivery of a drug to its target site could be coupling the drug to a monoclonal antibody fragment that is specific for proteins found only on the surfaces of certain cells, e.g., tumour cells**. Alternatively, there are some other options like **the coupling of a specific enzyme** (that converts an inert prodrug to the active form of the drug) **to a monoclonal antibody** directed against a specific cell surface antigen. To ensure that the drug is released only in the vicinity of the target cells, the monoclonal antibody that is complexed with the prodrug-converting enzyme must bind to a protein that is highly specific to the target cell and be stable under physiological conditions but cleared rapidly from circulation, permitting the use of a much lower concentration than if it was administered directly, because only specifically targeted cells are exposed to the drug.

It is sometimes advantageous to use an antibody fragment (particularly one that is **conjugated to a toxin or radiochemical**) to target a tumour or pathogen cell, since the Fc portion of the full-length antibody may impede or prevent the rest of the molecule from binding to relatively inaccessible antigens. However, **despite the usefulness of antibody fragments in a variety of applications, their major limitation as therapeutic agents is that, since they lack the Fc portion of the molecule, they are unable to mediate a complete immune response.**(6)

Some examples of all of these applications are:

-**Antibody fragment scFv coupled to a toxin (Pseudomonas exotoxin A).** Pseudomonas exotoxin A is a 66-kDa protein with three separate domains; domain I is responsible for cell binding, domain II for translocation of the protein into the cell, and domain III for ADP-ribosylation, where the latter activity is responsible for the toxicity of the molecule. An immuno-toxin is generally synthesized by replacing the N-terminal target-binding domain of the toxin, (in Pseudomonas exotoxin A corresponds to domain I) with a single-chain antibody sequence, thereby creating a molecule very similar in size to the original toxin with the ability to bind, enter, and kill a specific cell. A number of immunotoxins that have antitumor activity in vitro and in animal models have been constructed. These include antibodies directed against some cancer cell surface proteins such as the p55 subunit of the interleukin-2 receptor, the transferrin receptor, carbohydrate antigens, and the epidermal growth factor receptor. Several different engineered immunotoxins are currently in clinical trials.

-**Whole Antibody coupled to a drug (HER2 receptor).** Sometimes, it is therapeutically advantageous. At the present time, most of the whole anti-body-drug conjugates that are in use are designed to enter into lysosomes (which are membrane-bound organelles that contain hydrolytic enzymes that act to break down various biomolecules) following their internalization into target cells. Importantly, human epidermal growth factor receptor 2 (HER2) monoclonal antibodies were already in clinical use (as **Pertuzumab**) for the treatment of HER2 breast cancers. To improve upon the outcomes with this monoclonal antibody, a monoclonal antibody-drug conjugate directed at HER2-positive breast cancer cells (cells that contain the protein HER2 on the tumour cells' surface), was modified by engineering the monoclonal antibody to reduce its affinity for its target at acidic endosomal pH relative to near neutral pH. The modified antibody retained its specificity for HER2 but has an altered isoelectric point, which resulted in the antibody-drug conjugate being able to more readily dissociate from its target, thereby facilitating entry of the target into the lysosome. It is thought that this type of modification should eventually increase the therapeutic efficiency of HER2-drug conjugates.

-**Smaller antibody-toxin complexes (Epstein-Barr virus. Colicin Ia).** It is possible to create peptides that are smaller than scFvs and still retain the ability to bind to a specific antigen. The rationale is that these smaller molecules are more likely to penetrate a tumour and thereby more completely stop tumour growth. It is well established that antibody-binding specificity resides within the six hypervariable loops called CDRs. Hence, it was speculated, at least for some antibodies, that the major portion of the antigen-binding site might reside primarily within **two CDRs**, one from the heavy chain and the other from the light chain, and not require significant contributions toward antigen binding from the four other CDRs. To test this, starting with genes for the variable portion of a monoclonal antibody against a surface protein from Epstein-Barr virus (thought to be the causative agent of Burkitt lymphoma and other cancers), eight different peptide combinations were synthesized. Each peptide contained at least one **CDR_3 loop** (known to be the major antigen-contacting segment), as well as one other CDR loop and an FR spacer (which acts as a linker peptide). One of the eight of these peptide combinations, **$VHCDR_1$-$VHFR_2$-$VLCDR_3$**, appeared to be promising. The short peptide (~3 kDa) apparently retained the binding specificity of the whole monoclonal antibody (~150 kDa) from which it was derived. This peptide was subsequently **coupled to a toxin molecule, colicin Ia** (~69 kDa), and when the combination was injected into mice, the peptide-colicin molecule efficiently travelled through the circulation and then found and killed the tumour cells expressing the target antigen. Importantly, without the antibody fragment, colicin Ia by itself does not affect these tumours to any significant extent, and also that the original monoclonal antibody is unable to penetrate into the tumour. This work is at an early stage of development. Nevertheless, the demonstration that a small peptide can mimic the binding specificity of an entire antibody molecule and successfully deliver a cellular toxin to targeted cells may provide the basis for a whole new approach for treating tumours. (6)

5.2.4. Affibody Molecules.

In an attempt to bypass the often tedious and time-consuming steps to isolate new monoclonal antibodies, researchers have constructed libraries of short peptides that can bind tightly to various target molecules. These peptides, known as affibodies, are selected for the ability to bind to selected target molecules. Currently, many literature reports of such high-affinity affibodies have been published. It is envisioned that, like monoclonal antibodies, affibodies may be used in a variety of both therapeutic and diagnostic applications. For example, affibodies' small size **facilitates their penetration into tumours, even coupled to radiolabelled chemicals or toxins**, often being much more efficient than similarly constructed monoclonal antibodies.

In one study, it was observed that, in a mouse model, labelled affibodies **were more efficient than labelled antibody fragments in penetrating the blood-brain barrier. For therapeutic applications** that employ monoclonal antibody fragments (scFv, Fv, or Fab), **affibodies are often equally or even more effective, and often have a low level of immunogenicity.** At this stage, it appears that the efficacy of affibody molecule-drug complexes compared to antibody and antibody fragment-drug complexes will need to be tested on a case-by-case basis. This fact notwithstanding, this technology provides scientists with an increased range of options for treating various diseases. (6)

5.3. MEDICAL USES OF MONOCLONAL ANTIBODIES.

Since the firsts Monoclonal Antibodies approved for medical uses in 1990s decade, (as the famous Rituximab approved in 1998), approved indications for new marketed Monoclonal Antibodies are just growing exponentially. (16) Nowadays, they are widely used on Medicine for the treatment of entirely different types of diseases.

5.3.1. Immune-mediated diseases.

Monoclonal antibodies have revolutionized the treatment of autoimmune diseases, being importantly used for this purpose (**Figure 14**(17)). Autoimmune diseases are characterized by the activation of autoreactive CD4+ lymphocytes in the peripheral lymph nodes, where naïve T cells interact with antigen-presenting cells and B cells. Activated T cells proliferate and migrate into the disease-targeted organ parenchyma, where the recognition of endogenous ligands leads to the production of cytokines and pro-inflammatory molecules, resulting in cell damage and disease progression. Monoclonal antibodies can target different components of the immune system to suppress the excessive responses that characterize autoimmune diseases. The most widely used approach to treat autoimmune disorders is **blockade of pro-inflammatory cytokines, especially TNF-α**, a cytokine with an essential role in autoimmunity that induces vasodilation and inflammation. These antibodies have been used for the therapy of rheumatoid arthritis for more than a decade, and also show efficacy in psoriatic arthritis, Crohn's disease, ulcerative colitis, psoriasis, and ankylosing spondylitis. **Rituximab** combined with chemotherapy plus Methotrexate is widely used to treat moderate to severe cases of Rheumatoid Arthritis.(17)

Antibody	Type	Target	Medical uses
Adalimumab	Human, mAb, IgG1	TNF-α	Rheumatoid arthritis, Crohn's disease, plaque psoriasis, psoriatic arthritis, ankylosing spondylitis, juvenile idiopathic arthritis
Alemtuzumab	Humanized, mAb, IgG1	CD52	Multiple sclerosis
Belimumab	Human, mAb, IgG1	BAFF	Systemic lupus erythematosus
Benralizumab	Humanized, mAb, IgG1	CD125	Asthma
Brodalumab	Human, mAb, IgG2	IL-17	Plaque psoriasis
Canakinumab	Human, mAb, IgG1	IL-1	Cryopyrin-associated periodic syndrome
Certolizumab pegol	Humanized, Fab', IgG1	TNF-α	Crohn's disease, rheumatoid arthritis, axial spondyloarthritis, psoriatic arthritis
Golimumab	Human, mAb, IgG1	TNF-α	Rheumatoid arthritis, psoriatic arthritis, ankylosing spondylitis
Guselkumab	Human, mAb, IgG1	IL23	Psoriasis
Infliximab	Chimeric, mAb, IgG1	TNF-α	Rheumatoid arthritis, ankylosing spondylitis, psoriatic arthritis, psoriasis, Crohn's disease, ulcerative colitis
Itolizumab	Humanized, mAb, IgG1	CD6	Psoriasis
Ixekizumab	Humanized, mAb, IgG4	IL-17A	Plaque psoriasis
Mepolizumab	Humanized, mAb, IgG1	IL-5	Asthma and white blood cell diseases
Natalizumab	Humanized, mAb, IgG4	Integrin α_4	Multiple sclerosis, Crohn's disease
Ocrelizumab	Humanized, mAb, IgG1	CD20	Rheumatoid arthritis, lupus erythematosus
Omalizumab	Humanized, mAb, IgG1	IgE Fc region	Allergic asthma
Reslizumab	Humanized, mAb, IgG4	IL-5	Inflammations of the airways, skin and gastrointestinal tract
Risankizumab	Humanized, mAb, IgG1	IL23A	Crohn's disease, psoriasis, psoriatic arthritis, asthma
Rituximab	Chimeric, mAb, IgG1	CD20	Rheumatoid arthritis
Ruplizumab	Humanized, mAb, IgG1	CD154	Rheumatic diseases
Sarilumab	Human, mAb, IgG1	IL6	Rheumatoid arthritis, ankylosing spondylitis
Secukinumab	Human, mAb, IgG1	IL17A	Uveitis, rheumatoid arthritis, psoriasis
Tildrakizumab	Humanized, mAb, IgG1	IL23	Immunologically mediated inflammatory disorders
Tocilizumab	Humanized, mAb, IgG1	IL-6 receptor	Rheumatoid arthritis
Ustekinumab	Human, mAb, IgG1	IL-12, IL-23	Multiple sclerosis, psoriasis, psoriatic arthritis
Vedolizumab	Humanized, mAb, IgG1	Integrin $\alpha_4\beta_7$	Crohn's disease, ulcerative colitis

Figure 14. Therapeutic Monoclonal Antibodies used for the treatment of autoimmune diseases. Taken from (17).

5.3.2. Oncology.

A significant number of monoclonal antibodies have been developed for the treatment of various neoplasias, including both hematologic and solid tumours (**Figure 15**(17)). The first approach is the use of monoclonal antibodies to target tumour antigens and kill cancer cells. The main targets are **growth factor receptors** that are overexpressed in tumour cells, such as members of the epidermal growth factor receptor (EGFR) family, including EGFR and HER2(explained in paragraph **5.2.3. Antibody Fragments**). Monoclonal antibodies block these receptors, in turn blocking ligand binding/signaling, which can decrease growth rate, induce apoptosis and sensitize tumours to chemotherapy. Other targets besides growth factors include **hematopoietic differentiation antigens (CD20, CD30, CD33, CD52)**, which are glycoproteins found on the surface of normal and tumour cells.(17) For instance, Rituximab targets CD20, like **Obinutuzumab**, which targets and destroys CD20-positive B cells by directly binding to the tumour cell surface-exposed CD20 molecules and initiating antibody-dependent cytotoxicity (ADCC), where the antibody recruits the immune system to attack the CD20-positive B cells. It is an example of glycol-engineered optimized monoclonal antibody, which means that its complement of surface carbohydrates has been modified, so that it does not contain any fucosylated sugar molecules. Fucose is a hexose deoxy sugar that is found on the surface of many eukaryotic cells (including mammals), and it is thought that the lack of fucose moieties may result in an increase in the ADCC activity of the antibody, being modified for this purpose. (18)

Monoclonal antibodies can also be used for other purposes as the **delivery of radioisotopes selectively to cancer cells** (like **Ibritumab Tiuxetan**, a monoclonal antibody labeled with Yttrium 90 or Indium 111, used for the treatment of non-Hodgkin's lymphoma), or the **inhibition of angiogenesis** (as **Bevacizumab**, a monoclonal antibody which blocks the binding of vascular endothelial growth factors (overexpressed in various cancers) to the receptor in the vascular endothelium).

Another approach for anticancer monoclonal antibodies-based therapies is the targeting of immune cells, the so-called **immune-checkpoint inhibitors**, that enhance antitumor immune responses. The main immune-check- point inhibitors target cytotoxic T-lymphocyte associated antigen 4 (CTLA-4) and programmed cell death protein 1 (PD-1)/ PD1 ligand 1 (PD-L1). CTLA-4 can be expressed by regulatory T cells infiltrating tumour lesions, and it mediates immunosuppression by inhibiting T-cell functions. CTLA-4 blockade restores T-cell function to kill malignant cells. One example of this is **Ipilumab** (an anti- CTLA4 approved for advanced melanoma in 2011).

The receptor-ligand pair PD-1/PD-L1 negatively regulates T cell-mediated immune responses and can be used by tumours as a mechanism of evasion of antigen-specific T-cell immunologic responses. These drugs are extremely important, because of its impact on current pharmacological applications (being the projected top 1 sold drug for this year 2023). We are referring to **Nivolumab** (Opdivo), a PD-1 inhibitor (shown at the introduction (**point 2.INTRODUCTION, in Figure 1**(2))), approved for the treatment of various cancers. (17)

Antibody	Type	Target	Medical uses
Alemtuzumab	Humanized, mAb, IgG1	CD52	B-cell chronic lymphocytic leukemia
Bevacizumab	Human, mAb, IgG2	VEGF	Colorectal cancer, non-squamous non-small cell lung cancer, glioblastoma, renal cell carcinoma
Gemtuzumab ozogamicin	Human, ADC, IgG4	CD33	Acute myelogenous leukemia
Trastuzumab-emtansine	Humanized, ADC, IgG1	HER2	Metastatic breast cancer
Brentuximab-vedotin	Chimeric, ADC, IgG1	CD30	Hodgkin's lymphoma
Trastuzumab	Humanized, mAb, IgG1	HER2	HER2-positive breast cancer, gastric/gastroesophageal junction carcinoma
Cetuximab	Chimeric, mAb, IgG1	EGFR	Squamous cell cancer of the head and neck, metastatic EGFR-positive colorectal cancer
Panitumumab	Human, mAb, IgG2	EGFR	EGFR-positive metastatic colorectal carcinoma
Ipilimumab	Human, mAb, IgG1	CTLA-4	Unresectable or metastatic melanoma
Rituximab	Chimeric, mAb, IgG1	CD20	CD20-positive B cell non-Hodgkin lymphoma and chronic lymphocytic leukemia
Ofatumumab	Human, mAb, IgG1	CD20	Refractory chronic lymphocytic leukemia
^{90}Y-Ibritumomab Tiuxetan	Murine, mAb, IgG1	CD20	Relapsed or refractory, low-grade or follicular B-cell non-Hodgkin's lymphoma
^{131}I-Tositumomab	Murine, mAb, IgG2	CD20	CD20-expressing relapsed or refractory low-grade, follicular or transformed non-Hodgkin's lymphoma
Atezolizumab	Humanized, mAb, IgG1	PD-L1	Triple-negative breast cancer
Avelumab	Human, mAb, IgG1	PD-L1	Merkel-cell carcinoma
Blinatumomab	Murine, mAb, IgG1	CD19	Acute lymphoblastic leukemia
Cemiplimab	Human, mAb, IgG1	PD-1	Metastatic cutaneous squamous cell carcinoma
Daratumumab	Human, mAb, IgG1	CD38	Multiple myeloma
Dinutixumab	Human, mAb, IgG1	GD2	Neuroblastoma
Elotuzumab	Humanized, mAb, IgG1	SLAMF7	Multiple myeloma
Necitumumab	Human, mAb, IgG1	EGFR	Non-small cell lung cancer
Obinutuzumab	Humanized, mAb, IgG1	CD20	Chronic lymphocytic leukemia
Pembrolizumab	Humanized, mAb, IgG1	PD-1	Melanoma and other cancers

Figure 15. Therapeutic Monoclonal Antibodies used for anticancer therapy. Taken from (17).

5.3.3. Infectious diseases.

Figure 16(17) shows monoclonal antibodies approved for prophylaxis and/or treatment of infectious diseases. The first effective treatment for infectious diseases was the administration of hyperimmune sera from immunized animals or human donors. Although this approach was widely replaced with antibiotic treatment, it still remains useful for the treatment of infectious diseases, including those caused by cytomegalovirus, hepatitis A and B viruses, among others. There are advantages of monoclonal antibodies for the treatment of infections, over immune sera-derived preparations, such as low risk of pathogen transmission, and no immunological complications associated with the use of heterologous sera. However, the development of monoclonal antibodies against infectious diseases has been slower in comparison to their development for oncology and immune/inflammatory diseases. (19)

The first monoclonal antibody approved for an infectious disease was **Palivizumab**, used for the prevention of severe respiratory disease due to respiratory syncytial virus in high risk populations, inhibiting virus replication and reducing the frequency of severe disease in premature infants. Another important example was **Ibalizumab**, approved in 2018 for the treatment of multidrug-resistant HIV-1 infection. This was the first new HIV treatment medication approved in over a decade, and it acts as a post-attachment inhibitor by binding CD4 receptors and blocking viral entry into the host CD4+ T cells.

Development of other potential monoclonal antibodies to treat infectious diseases is underway, including Ebola virus disease, hepatitis B and C and herpes simplex virus. (17)

Antibody	Type	Target	Medical uses
Bezlotoxumab	Human, mAb, IgG1	*Clostridium difficile*	*Clostridium difficile* colitis
Ibalizumab	Humanized, mAb, IgG4	CD4	Multidrug-resistant HIV infection
Oblitoxaximab	Chimeric, mAb, IgG1	*Bacillus anthracis* anthrax	Anthrax (prophylaxis and treatment)
Palivizumab	Human, mAb, IgG1	F protein of respiratory syncytial virus	Respiratory syncytial virus (prevention)
Raxibacumab	Human, mAb, IgG1	Anthrax toxin	Anthrax (prophylaxis and treatment)
Rmab	Human, mAb, IgG4	Rabies virus G glycoprotein	Post-exposure prophylaxis of rabies

Figure 16. Therapeutic Monoclonal Antibodies used for infectious diseases treatment. Taken from (17).

5.3.4. Other indications.

Various monoclonal antibodies have been developed for **antiplatelet therapy**, although only one, **Abciximab**, has been approved so far. It prevents integrin binding to fibrinogen and von Willebrand factor (a blood glycoprotein involved in haemostasis).

Another indication for monoclonal antibodies is the prophylaxis treatment of **migraines**, targeting calcitonin gene-related peptide (CGRP) for preventative migraine therapy. This peptide acts on the CGRP receptor and is involved in pain modulation, perception, and central sensitization. For this reason, monoclonal antibodies targeting this peptide like **Erenumab, Fremanezumab and Galcanezumab** have shown a benefit in these patients.

Finally, there are potential applications of monoclonal antibodies in the development of **immune complex vaccines**, as both preventive and therapeutic immunization approaches. Antigen-monoclonal antibody immune complex-based vaccines mimic natural immune complex functions and have been used for poultry for the prevention of infectious **bursa disease**. Following this success, several human infectious diseases are being targeted by this approach, including HIV-1, hepatitis B, and Ebola. (17)

6.CONCLUSION AND FUTURE PERSPECTIVES.

Recombinant proteins are the XXI century molecules, that has plenty of applications such as chemical industry (industrial applications), pharmacological industry (medical applications, which correspond to therapeutic recombinant proteins being the main subject in this project) and also in food industry. Therapeutic recombinant proteins, the ones which are used for treatment of diseases on medical world, have reached the top selling drugs and are here to stay, with Monoclonal Antibodies being the most used therapeutic recombinant proteins and the ones that occupy the highest position in this ranking. Genetic engineering can be used to modify and optimize Monoclonal Antibodies for the treatment of different diseases, being the most used drugs nowadays for some applications as Immune-mediated diseases or Anticancer Therapies. Monoclonal Antibodies are very present in medical daily life nowadays and they have become an important way of research and development that certainly will be increased in the future years, becoming the central axis of “personalized medicine”.

7.REFERENCES.

1. Functions Of Proteins | 5 Main Functions, What, Where & How [Internet]. 2020 [cited 2023 Apr 15]. Available from: https://alevelbiology.co.uk/notes/functions-of-proteins/
2. Buntz B. 50 best-selling pharmaceuticals of 2022: Market leaders and trends [Internet]. Drug Discovery and Development. 2023 [cited 2023 May 8]. Available from: https://www.drugdiscoverytrends.com/50-of-2022s-best-selling-pharmaceuticals/
3. Top pharmaceutical drugs by projected 2023 global sales [Internet]. Statista. [cited 2023 Apr 22]. Available from: https://www.statista.com/statistics/973523/top-drugs-by-year-on-year-sales-increase/
4. Enzyme. In: Wikipedia [Internet]. 2023 [cited 2023 May 8]. Available from: https://en.wikipedia.org/w/index.php?title=Enzyme&oldid=1132111677
5. Burnett MJB, Burnett AC. Therapeutic recombinant protein production in plants: Challenges and opportunities. PLANTS PEOPLE PLANET. 2020;2(2):121–32.
6. Glick BR. Molecular Biotechnology – Principles and Applications of Recombinant DNA, 6th Edition. Hoboken, NJ; 2022. 896 p.
7. Heckman KL, Pease LR. Gene splicing and mutagenesis by PCR-driven overlap extension. Nat Protoc. 2007 Apr;2(4):924–32.
8. Glick BR, Pasternak JJ, Patten CL. Molecular biotechnology: principles and applications of recombinant DNA. 4th ed. Washington, DC: ASM Press; 2010. 1000 p.
9. Deshpande AD, Harris-Hayes M, Schootman M. Epidemiology of Diabetes and Diabetes-Related Complications. Phys Ther. 2008 Nov;88(11):1254–64.
10. Insulin glargine. In: Wikipedia [Internet]. 2023 [cited 2023 May 10]. Available from: https://en.wikipedia.org/w/index.php?title=Insulin_glargine&oldid=1150020171
11. Insulin glargine (insulin glargine) injection, solution,Insulin glargine SOLOSTAR (insulin glargine) injection, solution [Internet]. Available from: http://dailymed.nlm.nih.gov/dailymed/lookup.cfm?setid=d5e07a0c-7e14-4756-9152-9fea485d654a
12. Insulin lispro. In: Wikipedia [Internet]. 2023 [cited 2023 May 11]. Available from: https://en.wikipedia.org/w/index.php?title=Insulin_lispro&oldid=1149841243
13. Wilde MI, McTavish D. Insulin Lispro. Drugs. 1997 Oct 1;54(4):597–614.
14. Islam N, Khanna NR, Zito PM. Insulin Lispro. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2023 [cited 2023 May 11]. Available from: http://www.ncbi.nlm.nih.gov/books/NBK507840/
15. García Merino A. Anticuerpos monoclonales. Aspectos básicos. Neurología. 2011 Jun 1;26(5):301–6.
16. Dimitrov DS. Therapeutic Proteins. Methods Mol Biol Clifton NJ. 2012;899:1–26.
17. Castelli MS, McGonigle P, Hornby PJ. The pharmacology and therapeutic applications of monoclonal antibodies. Pharmacol Res Perspect. 2019 Dec 20;7(6):e00535.
18. Wang B, Gallolu Kankanamalage S, Dong J, Liu Y. Optimization of therapeutic antibodies. Antib Ther. 2021 Jan 1;4(1):45–54.
19. Grace PS, Gunn BM, Lu LL. Engineering the supernatural: monoclonal antibodies for challenging infectious diseases. Curr Opin Biotechnol. 2022 Dec 1;78:102818.