NOVA SCHOOL OF SCIENCE & TECHNOLOGY

DEPARTMENT OF LIFE SCIENCES

HUGO ALEXANDRE LEITÃO MATEUS

BSc in Biochemistry

IDENTIFICATION OF PEPTIDES PRESENTED BY MAJOR HISTOCOMPATIBILITY COMPLEX IN MACROPHAGES THROUGH IMMUNOPEPTIDOMICS

MASTER IN MOLECULAR GENETICS AND BIOMEDICINE

NOVA University Lisbon

September, 2023

DEPARTAMENT OF LIFE SCIENCES

IDENTIFICATION OF PEPTIDES PRESENTED BY MAJOR HISTOCOMPATIBILITY COMPLEX IN MACROPHAGES THROUGH IMMUNOPEPTIDOMICS

HUGO ALEXANDRE LEITÃO MATEUS

Master/BSc in Biochemistry

Adviser: Prof. Dr. Paulo Bettencourt
Assistant Professor, Universidade Católica Portuguesa

Co-advisers: Prof. Dr. Hugo Santos
Assistant Researcher, LAQV-REQUIMTE, Department of Chemistry,
NOVA School of Science and Technology, NOVA University Lisbon

Examination Committee:

Chair: Name of the committee chairperson,
Full Professor, FCT-NOVA

Rapporteurs: Name of a rapporteur,
Associate Professor, Another University
Name of another rapporteur,
Assistant Professor, Another University

Adviser: Name of the adviser present in defense,
Associate Professor, University

Members: Yet another member of the committee,
Full Professor, Another University
Yet another member of the committee,
Assistant Professor, Another University

MASTER IN MOLECULAR GENETICS AND BIOMEDICINE

NOVA University Lisbon
September, 2023

Identification of Peptides Presented by Major Histocompatibility Complex In Macrophages through Immunopeptidomics

ACKNOWLEDGMENTS

I would really like to thank both my supervisors, Prof. Paulo Bettencourt and Prof. Hugo Santos for everything they did for me in the last year. They were really supportive, comprehensive and helped me a lot. I would also like to thank all the people of Católica Medical School and No

ABSTRACT

Immunopeptidomics is a mass spectrometry-based method for the quantitative and qualitative analysis of peptides presented by MHC-I and MHC-II molecules. Is a field of research that has evolved tremendously in the last years due to advances in data analysis software and mass spectrometry equipment, now with much more sensitivity, resolution and accuracy.

MHC class I molecules present peptides at the cell surface of every nucleated cell and MHC class II present peptides at the surface of antigen-presenting cells that are recognized by T cells and may lead to the initiation of the immune response. Identification of peptides through immunopeptidomics is a method for the understanding of the immune system.

This project aimed to establish a functional immunopeptidomic workflow starting with a macrophage culture, proceeding to MHC immunoprecipitation and consequent peptide purification before injection of peptide samples in the mass spectrometer for the obtention of MS spectra for posterior analysis. Applying this workflow, it was possible to identify a total of 2913 unique MHC class I peptides presented by macrophages and characterize them by peptide length distribution, MHC binding affinity and binding motifs. 288 out of the total peptides were predicted by as strong binders for five different MHC allotypes and 669 as weak binders for the same allotypes. The binding motifs presented for each allotype resemble the consensus binding motifs presented in the literature.

With this work, it was possible to establish an immunopeptidomic platform of success that can be used in the future for the identification of peptides presented by macrophages under physiological and pathological conditions to identify antigens of interest for vaccine design and development.

Keywords: Immunopeptidomics, mass spectrometry, macrophages, MHC-I, peptides

RESUMO

Imunopeptidómica é um método baseado em espetrometria massa para a análise qualitativa e quantitativa de péptidos apresentados por moléculas MHC de tipo I e do tipo II. É uma área de investigação que se desenvolveu de forma tremenda nos últimos anos devido a avanços em software de análise de dados e em equipamentos de espetrometria de massa, automate com muito maior sensibilidade, resolução e precisão.

Moléculas MHC do tipo I apresentam péptidos na superfície de todas as células nucleadas e MHC do tipo II apresentam péptidos à superfície de células apresentadoras de antigénios que são reconhecidos por células T e podem conduzir à iniciação da resposta imune. Identificação de péptidos através de immunopeptidómica é um método para a compreensão do sistema imune.

Este projeto pretendia estabelecer uma linha de trabalho funcional com imunopeptidómica começando pela cultura de macrófagos, passando para a imunoprecipitação de MHC e consequente purificação de péptidos antes da injeção das amostras de péptidos no espetómetro de massa para a obtenção de espetros para posterior análise.

Aplicando este protocolo, foi possível identificar um total de 2913 péptidos únicos de MHC I apresentados por macrófagos e caracterizá-los por tamanho, afinidade de ligação com MHC e pelos "motifs" de ligação. 288 do total de todos os péptidos foram previstos como péptidos com forte ligação a cinco tipos differentes de alótipos de MHC e 669 previstos como péptidos com fraca ligação para os mesmos alótipos.

Com o trabalho realizado, foi possível estabelecer uma plataforma de imunopeptidómica de sucesso que pode ser usada no futuro para a identificação de péptidos apresentados por macrófagos em condições fisiológicas ou patológicas para identificar antigénios de interesse para o desenvolvimento e design de vacinas.

Palavras chave: Imunopeptidómica, espetrometria de massa, macrófagos, MHC-I, péptidos

ABBREVIATIONS

ACN	Acetonitrile
AM	Alveolar macrophages
BCG	Bacille Calmette-Guérin
BCG-GFP	Bacille Calmette-Guérin expressing green fluorescent protein
CFU	Colony forming units
CID	Collision-induced spectra
CLIP	Class II-associated *invariant chain* peptide
DDA	Data-dependent acquisition
DIA	Data-independent acquisition
DMEM	Dulbecco's Modified Eagle Medium
DMP	Dimethyl pimelimidate dihydrochloride
DRiPs	Defective ribosomal products
EDTA	Ethylenediaminetetraacetic acid
ER	Endoplasmic reticulum
ERAP	Endoplasmic reticulum associated amminopeptidase
ESI	Electrospray ionization
FBS	Fetal bovine serum
FT	Fourier transform ion cyclotron
GILT	Gamma-interferon-inducible lysosomal thiol reductase
HIV	Human immunodeficiency virus
HLA	Human leukocyte antigen
HPLC	High-performance liquid chromatography
IGEPAL	Octylphenyl-polyethylene glycol
IL	Interleukin

IP	Immunoprecipitation
KCl	Potassium chloride
MALDI	Matrix-assisted laser desorption/ionization
MARCH	Membrane-associated RING-CH-type finger
MHC	Major histocompatibility complex
MIIC	Late endosomal MHC class II compartment
MOI	Multiplicity of infection
MS	Mass spectrometry
M.tb	*Mycobacterium tuberculosis*
NaCl	Sodium chloride
NaOH	Sodium hydroxide
NET	Neutrophil extracellular trap
OADC	Oleic acid albumin dextrose complex
OD	Optical density
PBS	Phosphate buffered saline
PDIM	Phthiocerol dimycocerosates
PenStrep	Penicilin-Streptomycine
PMA	Phorbol 12-Myristate Acetate
PMN	Polymorphonuclear neutrophils
ROS	Reactive oxygen species
RPMI	Roswell Park Memorial Institute
TAP	Transporter associated with antigen processing
TAPBPR	TAP binding protein-related
TB	Tuberculosis
TCR	T-cell receptor
TFA	Trifluoroacetic acid
TLR	Toll-like receptors
TNF	Tumor necrosis factor
TOF	Time-of-flight
TRIS	Tris(hydroxymethyl)aminomethane

CONTENTS

List of Figures

List of Tables

1

INTRODUCTION

1.1 Tuberculosis

Tuberculosis (TB) is an infectious disease caused by the pathogenic bacterium *Mycobacterium tuberculosis* and transmitted by aerosol droplets containing the bacillus expelled by an individual with active tuberculosis[1,2]. Every year, there are about 10 million new cases of tuberculosis worldwide and approximately 1.3 million deaths[1]. *M. Tuberculosis* (M.tb) is a gram-positive bacterium that has an extra layer rich in glycolipids, lipids and polysaccharides located over the peptidoglycan[3]. The cell wall components generated by its biosynthetic pathways play a role in the longevity of M.tb. Some of those components are lipoarabinomannan, arabinogalactan, mycolic acid, mycocerosic acid and phenolphthiocerol and may set off the inflammatory reaction[4].

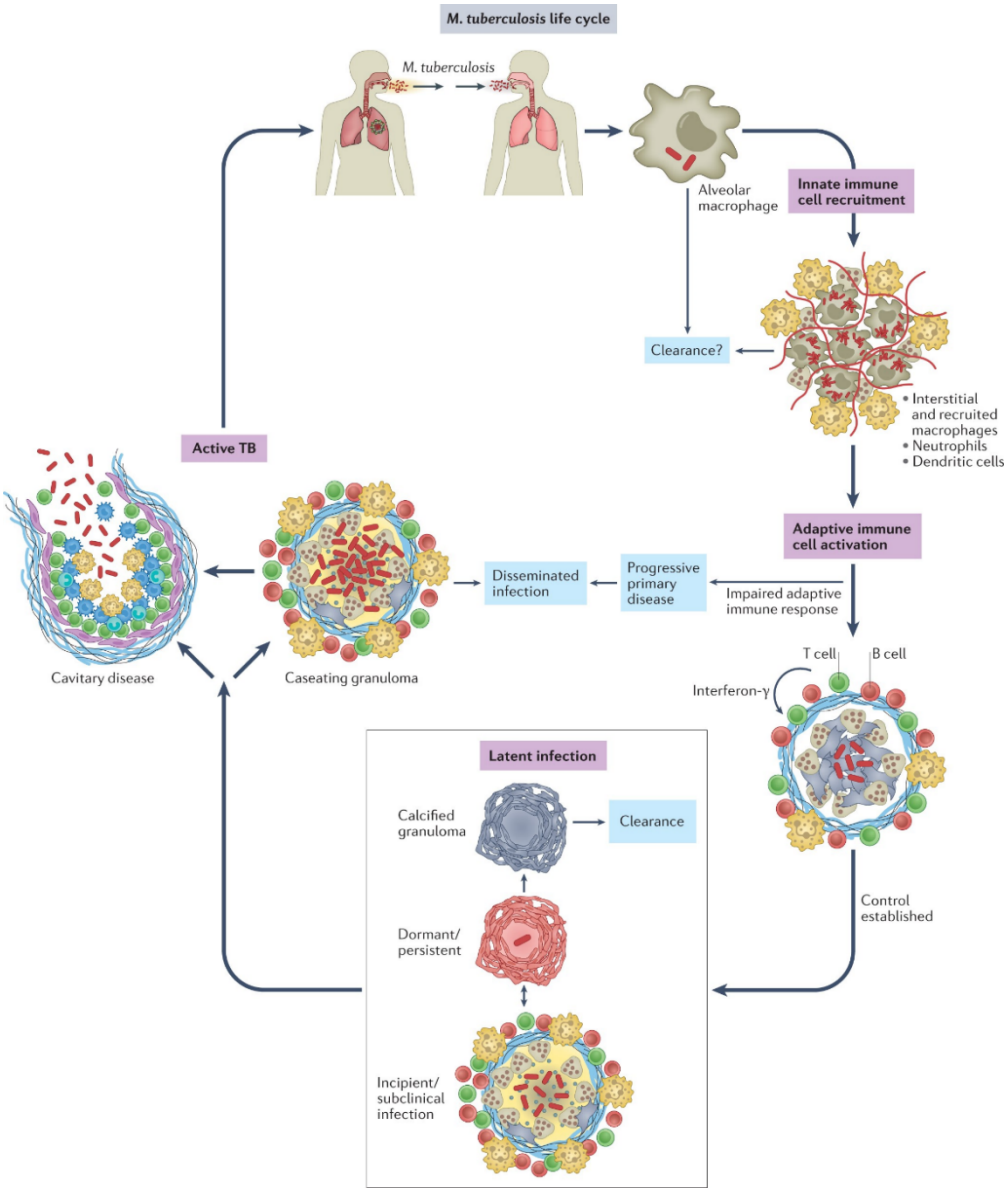

Figure 1.1.1 - Mycobacterium tuberculosis life cycle. Adapted from[6]

After inhalation, the droplets are deposited in the lung alveoli[5]. M.tb is an intracellular bacterium and can infect different types of cells, however, alveolar macrophages (AM) are the first cells infected and the most commonly infected (**Figure 1.1**). Beyond alveolar macrophages, M.tb also infects pulmonary epithelial cells and discharges some virulence lipids including sulfolipids and phthiocerol dimycocerosates (PDIM) into the epithelial cells[6].

The adaptive immune response starts when infected AM travels to the lung interstitium and M.tb infects monocyte-derived and tissue-resident macrophages, neutrophils and dendritic cells[7]. AM migration to the lung interstitium relies on the M.tb ESX-1 secretion system and IL-1β production by the host[6].

Neutrophils infected by M.tb produce reactive oxygen species (ROS) and neutrophil extracellular traps (NETs) (**Figure 1.2**). This response contributes more to increase the inflammatory response than to control bacterial propagation[8]. M.tb is able to obstruct the production of ROS by neutrophils and macrophages utilizing NuoG and the catalase/peroxidase KatG to detoxify reactive oxygen[9]. Macrophages use some mechanisms to eliminate M.tb and control infection like autophagy, phagolysosomal fusion and oxidative stress. To counter the attempt of elimination, M.tb produces virulence factors like PDIM, CpnT and EsxA which induce necrosis and this way allow the bacteria to spread and replicate extracellularly[10,11] (**Figure 1.2**). Besides the production of virulence factors, M.tb can also promote a foamy macrophage phenotype that aids bacterial survival by increased production of lipids by the macrophage that bacteria use for nutrition[12]. Inflammation of the tissues caused by cytokines including leukotrienes, tumor necrosis factor (TNF) and type I interferons contributes to the recruitment of more immune cells[13] (**Figure 1.2**).

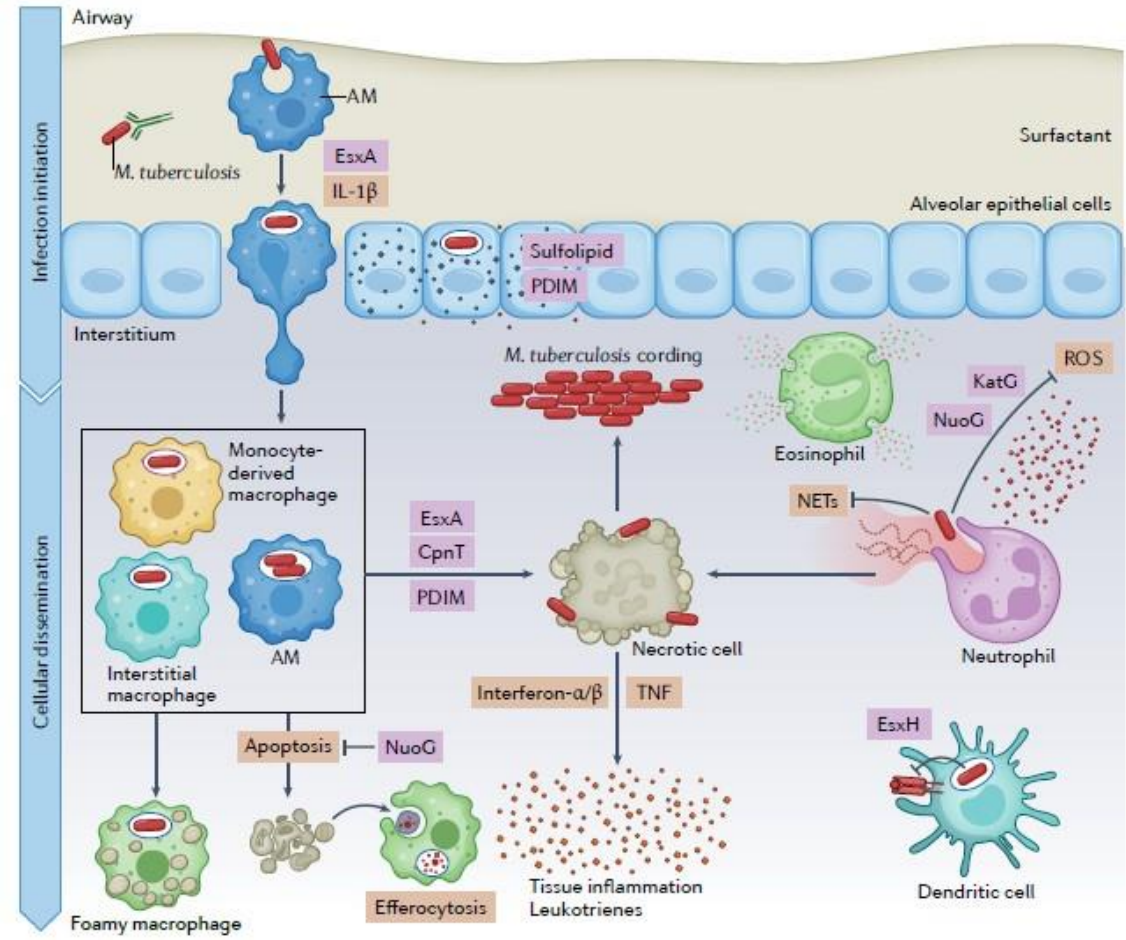

Figure 1.1.2 - M.tb immune evasion strategies. Adapted from[6]

Dendritic cells disseminate to the lymph nodes where the presentation of M.tb-derived pathogenic antigens by dendritic cells causes expansion and differentiation of antigen-specific naive T cells to effector T cells[14]. M.tb produces EsxH which represses the process of antigen presentation by dendritic cells slowing the beginning of the adaptive immune response and before that slowing down the migration of dendritic cells to the lungs through NuoG action by antigen uptake inhibition by dendritic cells and stopping infected polymorphonuclear neutrophils (PMNs) apoptosis[6] (**Figure 1.3**). M.tb also induces antigen degradation which impairs the ability of dendritic cells to prime $CD4^+$ T cells and interferes with T cell response using lipoglycans like lipoarabinonamman. $CD4^+$ T cell priming and differentiation is also repressed by phtiocerol dimycoserosate generated by M.tb and another pathogenic strategy applied is antigen decoy in which $CD4^+$ and $CD8^+$ T cells respond to decoy antigens with downregulated expression after T cell priming and their targeting by T cells is not protective[6] (**Figure 1.3**).

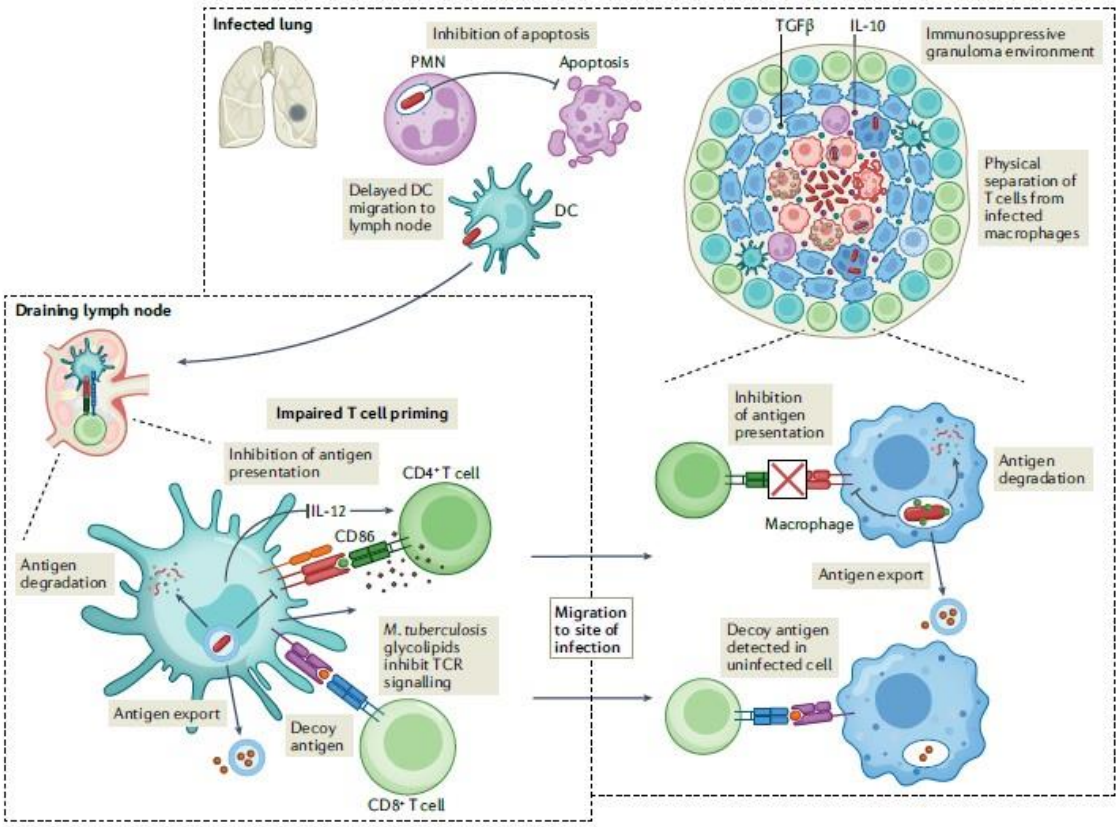

Figure 1.1.3 - Impaired immune response by infection. Adapted from[6]

The activated antigen-specific effector T cells then return to the site of infection and with other leukocytes give rise to organized structures denominated granulomas which comprise lymphocytes, macrophages and fibroblasts[15]. Inside the granulome activated macrophages limit bacterial replication and dissemination[2] (**Figure 1.3**). When the infection is controlled people have what is called latent TB,

without symptoms but that can develop into an active disease in a moment during their lives[15]. Is estimated that 5% to 10% of individuals with latent TB will develop the active disease at any moment of their lives and also, about one-third of the world population have latent TB[16].

Bacille Calmette-Guérin (BCG) is the vaccine used against TB and it was introduced about a hundred years ago. It is an attenuated form of *Mycobacterium bovis* but it was not successful in eliminating TB. Its efficacy against the active form of TB is highly variable in different populations[2]. For this reason, it is extremely important to develop an effective vaccine especially when considering the increased incidence of TB and its high lethality in people with human immunodeficiency virus (HIV)[4]. Several strategies for the development of effective vaccines can be attempted. Some of those strategies involve antibodies that can neutralize virulence factors or cell surface molecules which are needed by M.tb to establish infection or opsonizing antibodies that can auxiliate macrophages by assisting in the bacterial uptake or can even conduct the bacteria to an intracellular bactericidal pathway[17]. Epigenetic alterations in macrophages can create more protective macrophage phenotypes, for instance, macrophages able to identify antigens capable of causing a protective T-cell response instead of decoy antigens that dominate throughout infection. These types of strategies could give protection by allowing macrophage effector function reestablishment including the production of ROS, lysosomal trafficking, protective cell death pathways and optimal metabolic responses[7].

1.2 MHC

The major histocompatibility complex (MHC) locus is the most polymorphic region of the human genome and is this characteristic that allows the MHC molecules to bind and present a wide number of different antigen peptides[18]. MHC molecules are classified into two classes, MHC class I molecules and MHC class II molecules. The MHC system is responsible for the presentation of peptides and for the T-cell response that occurs afterward[19].

The ability of T-cells to discriminate self from nonself antigen peptides presented by MHC molecules is a property acquired during the development stage in the thymus. During that phase, T-cells gain the capacity for high selectivity even considering the large number of antigen peptides that are presented[19]. The activation of T cells is dependent on an interaction between a receptor and a ligand, being the receptor a self-MHC molecule and the ligand a peptide antigen. This mechanism is responsible for the two main consequences of the immune system function, the autoimmune response and the protection against infection[18]. T-cell receptors (TCR) are receptors present at the cell surface of T-cells, are unique

and recognize and bind different peptides bound to MHC molecules. A process of random recombination of numerous gene segments leads to the formation of diverse cell-specific receptors. These receptors recognize antigen peptides initiating the T cell clonal expansion which initiates the adaptive immune response[20]. During the immune response, T cells act to eliminate pathogens in case these can pass the physical and chemical barriers. T cells act by killing pathogens in infected cells but also aid B cells, so these can produce a more diversified and specific number of antibodies to protect against toxins produced by the pathogens and also to directly help eliminate the pathogen itself[19]. Hence, T cells need to contact with other cells of the immune system as B cells as well as with the infected cells. However, T cells can only interact with the outside of cells and for that reason need something to perceive if a cell ingested a bacteria for instance, or is synthesizing mutant or viral proteins[21]. To overcome this obstacle, the antigen presentation system at the cell surface has evolved so that T cells can be aware of what is occurring inside of the cell. In this way, the antigen presentation pathways pass information to T cells about the pathogens, and therefore T cells can act properly and recruit more immune cells[22].

In the cytosol of all nucleated cell peptides are generated inside of the cell and bind to MHC class I molecules to be displayed at the cell surface to present to $CD8^+$ T cells. This presentation, thereby allows $CD8^+$ cells to recognize and eliminate infected cells. When dendritic cells, B cells and macrophages phagocytose pathogens, antigens present in the endocytic compartments bind to MHC class II molecules at the cell surface and are presented to $CD4^+$ T cells[23].

MHC class I and class II molecules have chaperone-like structures with two Ig domains covered by two parallel alpha helixes above a platform of beta-pleated sheets[24]. This structure creates a peptide-biding groove[25,26] (**Figure 1.4**). MHC molecules are extremely polymorphic. For MHC class I, more than ten thousand different alleles have been identified already. This is what allows MHC restriction. TCRs are specific not only for the antigenic peptide but also for a specific allele of the MHC molecule that contains polymorphic residues on the alpha helixes located at the top of the MHC molecules. In MHC molecules, inside of the peptide binding groove exist different pockets in which the side chain of a complementary amino acid is going to be placed, called anchor residues. The polymorphic residues located in the peptide binding groove change the nature of these pockets so that MHC molecules with different polymorphic residues present different fragments of a certain antigen[27]. Apart from the anchor residues, most of the other amino acids of the peptide are in a free space in the peptide binding groove and can vary from each one of the twenty amino acids in most cases[28,29]. Due to this variability, a large number of peptides can be presented by each kind of MHC molecule[21].

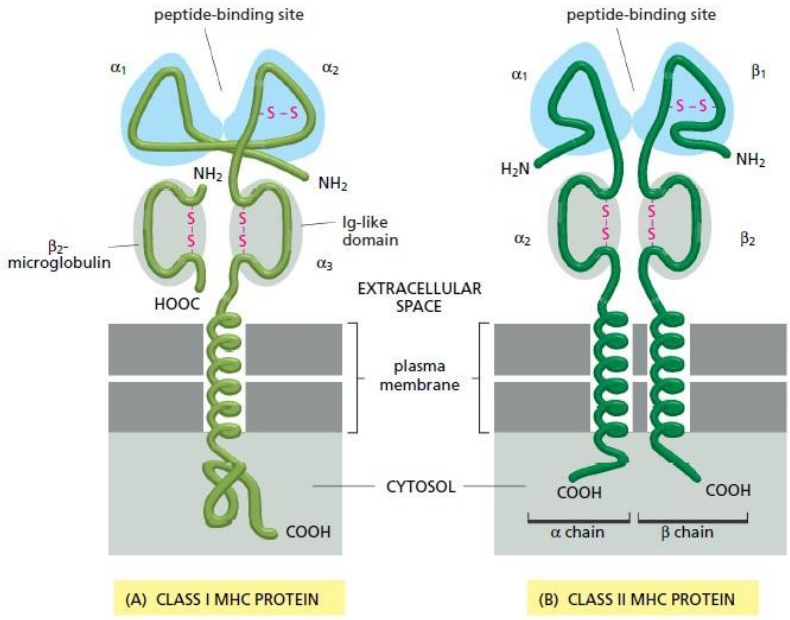

Figure 1.1.4 - MHC class I and class II molecules structure. Adapted from[24]

On the surface of B and T cells exist around two hundred thousand MHC class I molecules and around twenty thousand MHC class II molecules[30]. If a pathogen alters one of the anchor residues of the antigenic peptide it may prevent the presentation of that peptide in one person but not in another in which the MHC molecules present different antigens of the same pathogen. This is an example that illustrates why the MHC polymorphism can be beneficial for the survival of the population but not specifically of one individual[21].

3.2.1 MHC class I

MHC-I molecules are expressed at the surface of all nucleated cells and present normal and abnormal self-antigens, and also nonself-pathogen antigens. The non-self pathogen peptides are presented and recognized by $CD8^+$ effector T-cells[18]. MHC class I molecules present peptides derived from the hydrolysis of proteins synthesized by the cells. $CD8^+$ T cells are tolerant to these proteins, therefore in healthy cells, $CD8^+$ T cells do not act to eliminate them. When the cell starts expressing foreign genes such as microbial genes, the product of that expression will result in antigenic peptides that are recognized by $CD8^+$ T cells as non-self. Upon that recognition, $CD8^+$ T cells will take action to eliminate these abnormal cells[21].

Inside the cell, both normal and pathogenic proteins are degraded into peptide fragments by the cell proteasome[31]. The peptide fragments are then trimmed and a great part of them are destroyed by

peptidases in the cytosol[32] but some of the fragments can escape to the endoplasmic reticulum (ER) and survive. The peptide fragments enter the ER by a membrane peptide transporter called a transporter associated with antigen processing (TAP)[33]. Inside the ER there are empty MHC I molecules. Peptides enter the ER through TAP and become available to bind empty MHC I molecules with the TAP forming the center of the peptide loading complex. Aside from TAP, this complex includes three chaperones, calreticulin, tapasin, and protein disulfide isomerase ERp57 and also the empty MHC I molecule[34]. After peptide translocation through TAP, tapasin puts MHC I empty molecules in a peptide-receptive state and also promotes the binding of peptides with a slow-off rate. The MHC molecules bind the appropriate peptide with a low off-rate after testing and trying out the binding of several peptides, releasing all until it binds to one with a proper off-rate[35]. The peptides selected to bind with MHC I molecules generally have a length of eight to ten amino acids and specific anchor residues[21]. Endoplasmic reticulum-associated amminopeptidase 1 (ERAP1), an ER resident aminopeptidase can trim peptides longer than ten amino acids before they bind to MHC I molecules in the peptide-loading complex[36] (**Figure 1.5**). Empty MHC I molecules in the ER can also be associated with a chaperone called TAP binding protein-related (TAPBPR), a tapasin look-alike that also helps MHCI molecules to select the appropriate peptide to bind[37]. Binding peptides longer than eight or nine residues by ERAP1 causes a conformational change that results in the hydrolysis of the peptides. With this mechanism, peptide trimming to the optimal length for MHC binding is attained. Thus, the selected peptides for binding are the ones trimmed to the proper length and that escaped further hydrolysis and therefore can strongly bind to MHC I[38].

MHC I molecules are only stable when bound to peptides or the chaperones inside the ER. Therefore, peptides should be considered the third subunit of the MHC I[39]. The MHC I molecules can only be transported to the cell surface after release from the ER quality control system. This is only possible when they are bound to a peptide[40]. By selecting only the peptides that bind strongly to the MHC I molecules, the low off-rate peptide selection system prevents the easy replacement of endogenous peptides for antigenic exogenous peptides. If these replacements took place could lead to the destruction of the cell by $CD8^+$ due to the presentation of peptides that do not reflect what is occurring inside the cell[21].

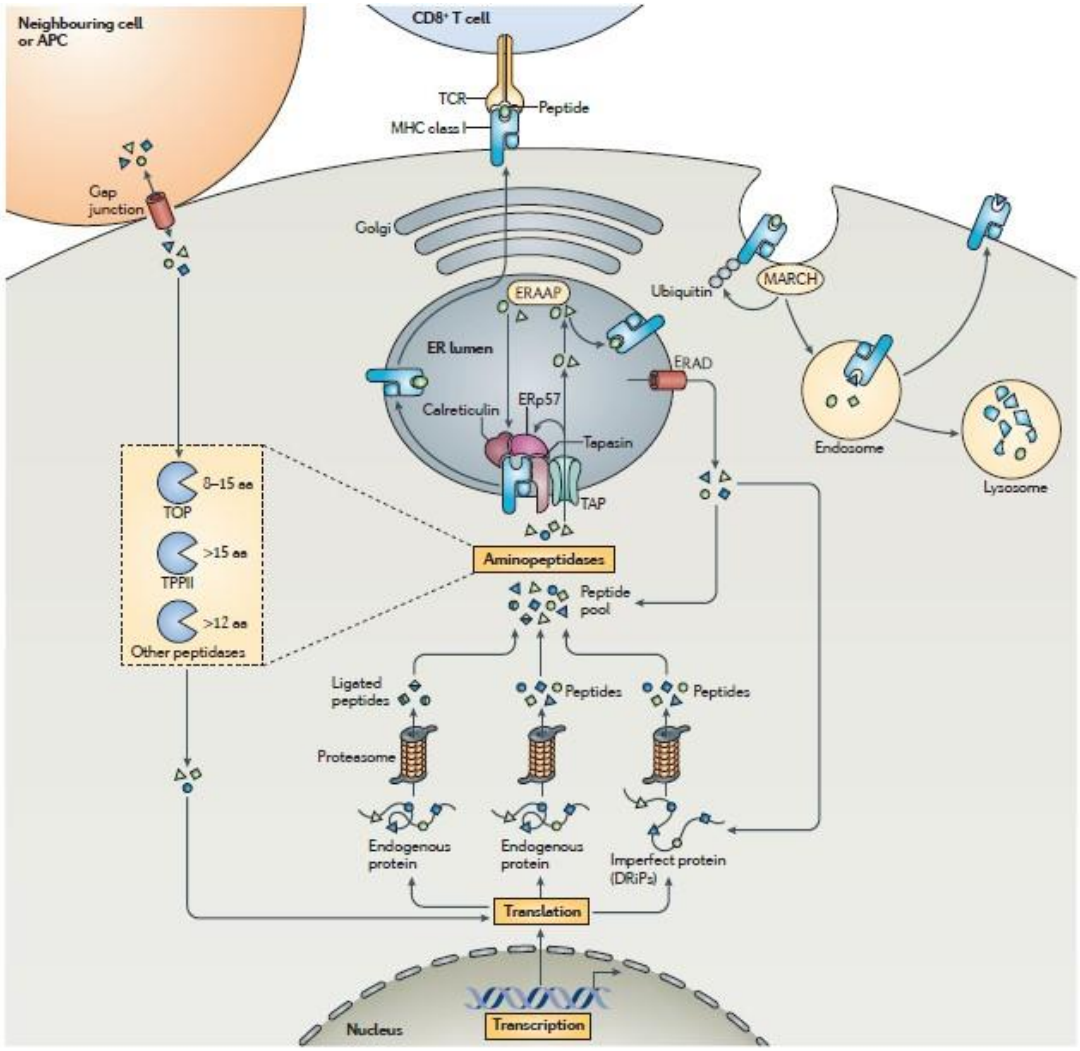

Figure 1.1.5 - MHC class I antigen presentation pathway. Adapted from[33]

The MHC I molecules in humans are named human leukocyte antigen (HLA) and are expressed in different loci of the human genome. Depending on the genomic locus in humans exist three types of HLA class I, HLA-A, HLA-B and HLA-C. Adding to this, the many polymorphic allelic forms contribute to the large diversity in MHC I molecules. The different regions of the loci responsible for MHC I have different patterns of expression, peptide binding and stability[21].

The peptides that will bind empty MHC I molecules that give rise to the MHC complexes are not only derived from the natural protein turnover in the cell but can also be derived from the degradation of abnormal proteins that arise from errors in several cellular processes such as protein translation, folding or pairing. To prevent the aggregation of these abnormal proteins, protein degradation is fundamental and likely supplies the antigen-presenting pathways with peptides from disabled proteins[41]. These antigens are called defective ribosomal products (DRiPs) and their production is increased during periods

of high protein synthesis for instance during viral infection. The degradation of these defective products after infection generates antigenic peptides quickly following the initial antigen translation allowing for a fast detection of infected cells[42] (**Figure 1.5**).

In addition to the proteasome, peptides can also be originated by degradation with another proteasome, an immunoproteasome that exists in cells of the immune system like dendritic cells and lymphocytes. This immunoproteasome is phylogenetically younger compared to the conventional proteasome found in other cells and its evolution is concurrent with the evolution of the immune system. In dendritic cells and lymphocytes, a set of subunits (β1i, β2i, β5i) is constitutively expressed and incorporates newly assembling particles to form the immunoproteasome[43]. The expression of these subunits can be induced in other cells by interferons during viral infections for instance. The conventional proteasome and the immunoproteasome generate different sets of peptides from protein degradation[44]. In cells, when the immunoproteasome is assembled by the expression of β1i, β2i and β5i subunits, the generation of MHC I peptides increases compared to when only the conventional proteasome is active. Many of those peptides are unique ones[45]. From all the peptides generated is estimated that only 0.02% survive for MHC I presentation to the immune system[46].

Through the peptides presented at the cell surface by MHC molecules, the immune system receives information on the state of the cell, since those peptides are generated by the cell-expressed genes at a given moment. In case of infection, the cells may change their immunoproteome in response to certain signals, resulting in several small post-translational modifications such as phosphorylation[47], acetylation, glycosylation[48] and others[49]. Peptides with post-translational modifications can enter the ER through the TAP transporter and then bind to the empty MHC I molecule and be presented to $CD8^+$ T cells. These peptides with these small modifications are not genetically encoded but are neo-epitopes that may not be tolerated by $CD8^+$ T cells[50] and in that case will allow $CD8^+$ T cells and subsequently the rest of the immune system to detect changes in the state of a cell and in some cases lead to the cell elimination[51].

Other peptides that are not genetically encoded are the ones that arise from peptide splicing by the proteasome. Usually, the proteasome promotes the proteolysis of the proteins in peptides but in this case performs the opposite reaction, forming a peptide by linking two smaller fragments[52]. These peptides can also be presented to $CD8^+$ T cells in MHC I molecules to stimulate an immune response. Whether the splicing process occurs just as the reverse of the proteolysis reaction or whether it is influenced by a specific cellular state remains unclear. Due to the post-translational modifications and the splicing reaction, the set of peptides presented, usually referred to as the presentome includes many more

peptides than just the genetically encoded. These characteristics of the antigen presentation pathways enable the immune system for the detection of abnormal cells[21].

Despite being a complex system, some pathogens as viruses have evolved some mechanisms to escape detection by the MHC system. Some of those mechanisms include code for proteins that inhibit the TAP transporter or the transport of MHC I complexes to the cell surface[53], decreasing the expression of MHC class I or inducing MHC I molecule degradation in the plasma membrane or ER. Any step of the antigen-presenting pathway can be manipulated by viruses to stop their presentation as long as does not compromise the cell viability[21]. Tumors can in certain conditions generate variants that escape the control of $CD8^+$ T cells by losing important components of the MHC I pathway that are not essential for the cell viability. This occurs because cancer cells are often genetically unstable[54]. Pathologists often detect in human tumors a decrease in MHC I molecule expression[55].

The MHC class I molecules present not only peptides generated by the natural protein turnover of the cell but also antigens from infected cells and tumor cells. This is called cross-presentation and it is a pathway that occurs in dendritic cells and phagocytes[56]. This pathway is of extreme importance for the correct functioning of the immune system[57]. In the periphery, dendritic cells acquire antigenic peptides from infected cells and then travel to the lymphoid organs where present those antigens to naive $CD8^+$ T cells, consequently activating them and initiating the next steps of the immune response. By their turn, phagocytes obtain those antigenic peptides by phagocytosis, ingesting the cell debris[58] or even taking part of the living cell[59], or through receptor-mediated endocytosis for instance antibody-bound antigens by Fc receptors or glycan-modified proteins across lectin receptors[60]. One mechanism of cross-presentation consists of the passage of the antigen from the phagosome to the cytosol where it can suffer proteolysis by the proteasome before entering the ER to bind to an empty MHC I molecules[61]. The fragments can reenter the phagosome and in there be loaded into phagosomal MHC I molecules[58]. Other mechanisms involve the degradation of antigens by lysosomal proteases. Then, the fragments will be loaded into recycling MHC I molecules[62]. Accumulation of MHC I molecules inside recycling endosomes can occur induced by TLR signaling, which therefore stimulates the cross-presentation pathway[63]. The mechanism of cross-presentation in MHC class I molecules is not completely understood yet but may share some features with the mechanism of antigen presentation in MHC class II molecules and presents antigenic peptides that usually are presented by the latter[21].

3.2.2 MHC class II

MHC class II molecules share some structural and functional similarities with MHC class I molecules. However, there are differences, especially regarding the mechanism of antigen presentation. While MHC I molecules are expressed in every nucleated cell, MHC II molecules are expressed in cells of the immune system as dendritic cells, monocytes, macrophages and B cells. Epithelial cells can also express MHC II molecules in response to inflammatory signals [21]. MHC-II molecules present abnormal and non-self pathogen antigens that activate $CD4^+$ T-cells[18]. The structure of the MHC II molecule is similar to the structure of MHC I molecules and both are highly polymorphic proteins[64].

Dendritic cells acquire antigens from the site of infection and then present those antigenic peptides on MHC II molecules to naive $CD4^+$ T cells, consequently activating them. Later in the immune response, MHC II molecules take part in the interaction between macrophages and B cells with effector $CD4^+$ T cells[51]. Individuals with low expression of MHC II molecules suffer from a condition called bare lymphocyte syndrome which highly increases the susceptibility to infections and may lead to death at a young age. This condition reveals how important is this mechanism of MHC class II antigenic presentation for the normal functioning of the human body[65].

While MHC I molecules present peptide fragments usually with eight to ten amino acids, MHC II molecules present larger peptide fragments because the peptide-binding groove of MHC class II molecules is open and this feature allows the fragments to stretch out of the peptide binding groove[66]. These peptides can vary largely in size, frequently from twelve up to twenty-six amino acid acids per peptide fragment consisting of a common binding core of nine amino acids flanked by varying residues. The common binding core interacts with the peptide-binding groove which consists of four main binding pockets. Is in these binding pockets that reside most of the allelic differences. The flanking residues also have a role in the interaction with T cells as they can directly contact the T cell receptors[67].

The peptides presented by MHC II molecules are obtained from endocytosed, vesicular and cytosolic proteins[68]. The peptides selected for presentation by MHC II molecules shape the basis of recognition by $CD4^+$ T cells leading to the immune response against pathogens, cancer cells or autoimmune antigens[67]. Preceding the binding with an antigen peptide, the class II MHC molecules inside the ER are associated with the invariant chain Ii that occupies the peptide-binding groove as a pseudopeptide[69]. This invariant chain has a cytosolic dileucine motif that targets the MHC II molecules to the endosomal pathway[70]. In a late endosomal compartment referred to as late endosomal MHC class II compartment (MIIC), MHC II molecules encounter the peptide fragments generated by MIIC resident proteases[71]. For the binding of MHC II molecules with the antigenic peptides, the invariant chain Ii needs to be degraded by proteases which include cathepsin S and L[64]. The degradation of the invariant chain results in an

invariant chain fragment named class II-associated invariant chain peptide (CLIP) that cannot be cleaved by the proteases and stays within the peptide-binding groove[72] until is exchanged for peptides with a higher affinity for the peptide-binding groove with the assistance of HLA-DM, a dedicated class II MHC-like chaperone[73]. When HLA-DM is associated with the MHC II molecules the groove opens locally to liberate peptides with low affinity like CLIP and HLA-DM determines the peptide with quality to bind to the MHC II molecules[74]. When HLA-DM is released from the peptide-binding groove the proper antigenic peptide binds to the MHC II molecule peptide-binding groove[66] (**Figure 1.6**). After all these steps that occur inside the MIIC compartment, the MHC II complexes travel by vesicular transport or in tubules to the plasma membrane[75]. After degradation of the invariant chain within the MIIC compartment, there is no targeting information and the MHC II molecules bind to the proper peptide fragment and can reside stably on the cell surface[51].

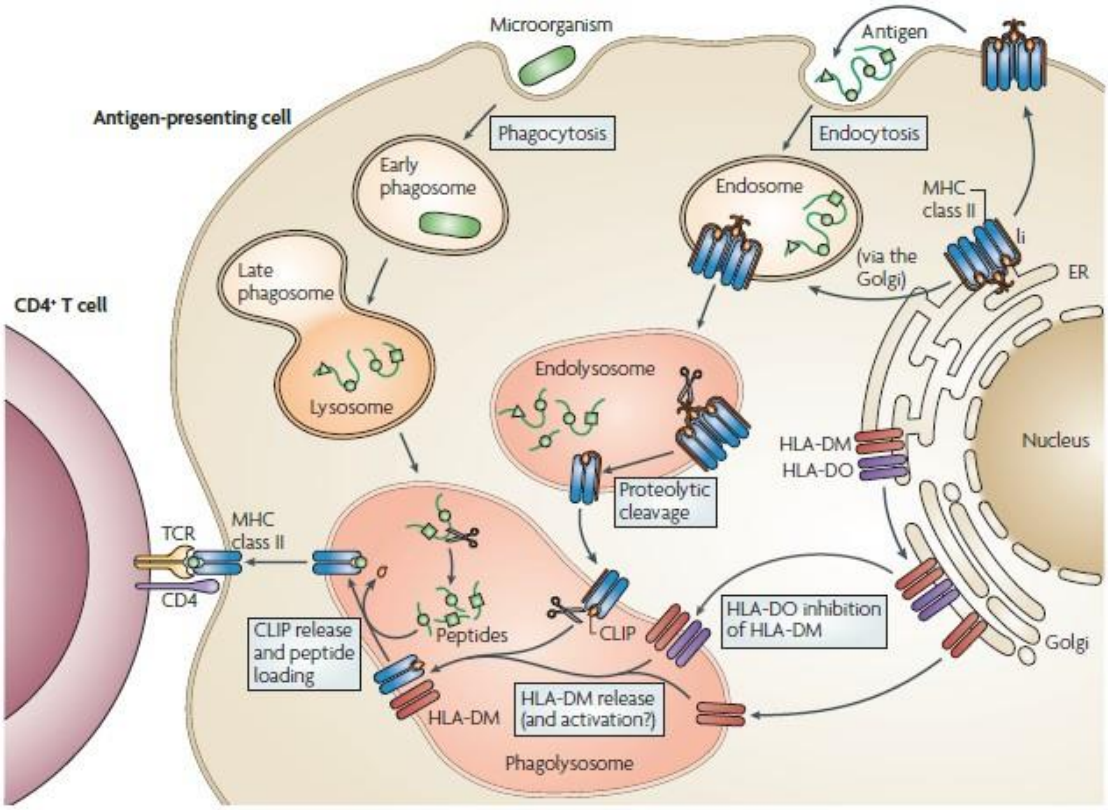

Figure 1.1.6 - MHC class II antigen presentation pathway. Adapted from[69]

As referred to MHC I molecules, in the human genome exist different loci for the MHC II system. In humans, there exist three different loci named HLA-DP, HLA-Q and HLA-DR. MHC II molecules are also highly polymorphic and more than three thousand different alleles are known. In the same manner as it occurs for the MHC I system, the polymorphisms shape the peptide-binding pockets[64]. MHC II alleles with different anchor residues bind different peptides[76] and this is an explanation for why certain

MHC II alleles are linked with defined auto-immune diseases[77]. When the MHC II peptide loading step takes place in the ER or in recycling endocytic compartments lacking HLA-DM[78] molecules, the MHC II molecules bind a peptide with a non-optimal conformation due to the absence of the HLA-DM. This may lead to the recognition of a self-peptide as a non-self peptide due to the unconventional conformation of the antigen peptide[79].

The invariant chain can also be a source of MHC class II variation. Despite the name invariant, several splice variants exist that alter the peptide presentation on MHC II molecules. Some of those variants can cause alterations in the activity of the proteases that are associated with antigen preparation for presentation. One of these invariant chain splice variants possesses one more protease inhibitor domain called cystatin. This leads to an alteration in the activity of proteases involved in the antigen presentation pathway. Amongst the cells that express MHC II molecules, the proteases also differ as well as their inhibitors. In the endosomal pathway, these proteases are cathepsins and their inhibitors cystatins[64].

In humans, the chaperone HLA-DM helps in the selection of peptides for presentation by MHC II molecules. Additionally, a dedicated co-chaperone controls the action of HLA-DM. This co-chaperone is called HLA-DO in humans and pairs with HLA-DM in a strong association[80]. The association of HLA-DM and HLA-DO takes place in the same interface in which MHC II molecules are located[66]. The association of HLA-DO with HLA-DM inhibits the process of MHC II peptide loading controlled by the second. To revert this inhibition the complex of HLA-DM and HLA-DO needs to enter alongside the MHC II molecules in extremely acidic endosomes. In this way, the action of HLA-DO prevents peptide binding in early vesicular compartments and therefore has an important role in the selection of peptides for MHC II binding[80].

Another factor in the MHC II system is the transport of MHC II molecules bound to peptides to the cell surface. This transportation process is not constitutively active in monocytes, dendritic cells and B cells. For the transport to occur dendritic cells need to be activated. The activation of dendritic cells, besides promoting the transport of MHC II complexes to the plasma membrane also increases the half-life of MHC II[81]. When activated dendritic cells may reach about two million MHC II molecules per cell that continue to present antigenic peptides on the cell surface for a long time[81]. MHC class II expression is also controlled by the action of cytokines. Interleucin-10 causes an increased expression in humans of membrane-associated RING-CH-type finger 1 (MARCH-1), a ubiquitin ligase membrane-associated. This protein provokes the internalization and subsequent destruction of MHC II molecules through tail ubiquitination of the MHC II molecules on the cell surface[82].

As it happens in the MHC I system, pathogens can also affect the expression of the MHC II system. The ones that occupy phagosomes containing MHC II molecules may inhibit peptide loading or MHC II molecule expression[83]. Several other ways by which pathogens can evade the immune response include pH level manipulation, modification of protein networks, MHC ubiquitination induction by MARCH homologs by manipulating the interaction between HLA-DM and MHC II molecules and other ways of avoiding antigen presentation by MHC II molecules[84].

1.3 Immunopeptidomics

The immunopeptidome is the group of all peptides antigens presented by MHC molecules, also known in humans as HLA[85]. Its composition cannot be deduced precisely by transcript or protein abundance and depends essentially on the analysis of peptides by mass spectrometry experiments, after the elution of the peptides from the MHC complex[18]. Immunopeptidomics is a mass spectrometry-based method in which the peptides that constitute the immunopeptidome of a biological sample are analysed quantitatively and qualitatively by high-performance liquid chromatography (HPLC) coupled with mass spectrometry (MS) to provide information about the set of peptides presented by a specific cell type or biological sample[85].

The first immunopeptidomic studies were only able to characterize peptides in the magnitude order of tens of peptides in samples with a high number of cells, sometimes even billions of cells[85]. Advances in MS instruments and analysis software in recent years made it possible to identify tens of thousands of peptides in much smaller samples with only a few million cells[85,86]. The success of mass spectrometry in immunopeptidomics is due to its specificity of identification and extreme sensitivity that can extend to a single ion[87].

In an immunopeptidomic workflow, several different methods can be used for each step of the workflow. Different mass spectrometry acquisition methods, different search engines and allele prediction algorithms[18]. Immunoaffinity purification or immunoprecipitation (IP) of MHC class I and II peptide complexes from cell lysates (**Figure 1.7**) has two main advantages. First, the process of peptide extraction has high specificity for peptides associated with MHC molecules and second, the range of purification. This method can be used for peptides that bind both MHC I and MHC II molecules and also from different sources such as primary tissues, cell lines and plasma[88]. From the peptides purified by IP up to 90% were reported to be MHC specific[89].

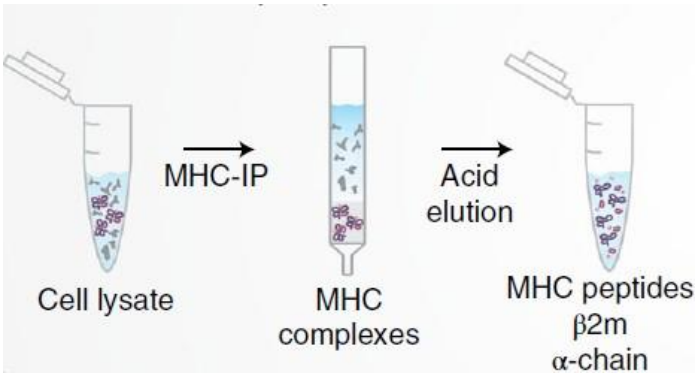

Figure 1.7 - MHC-peptide complexes immunoprecipitation. Adapted from[86]

In a mass spectrometry experiment, the analytes in the biological sample are ionized and volatilized before the mass analysis can occur. A mass spectrometer is a complex detection analytical instrument that incorporates several devices with distinctive functions. These include an ion source that ionizes the analytes in the sample, a mass analyser that measures m/z, the ratio of mass-to-charge of the analytes ionized previously, and lastly a detector. For every ion, a mass-to-charge value is measured by the detector[90]. Usually, two techniques are used to ionize and volatize the peptides in a biological sample and those are matrix-assisted laser desorption/ionization (MALDI) and electrospray ionization (ESI). MALDI sublimates and ionizes samples through laser pulses in a dry, crystalline matrix (**Figure 1.8**) and is frequently used for the analysis of more simple peptide samples[90] while ESI ionizes the analytes directly from a solution (**Figure 1.9**) and is easily compatible with liquid chromatography, thus being preferred to analyse more complex samples[91].

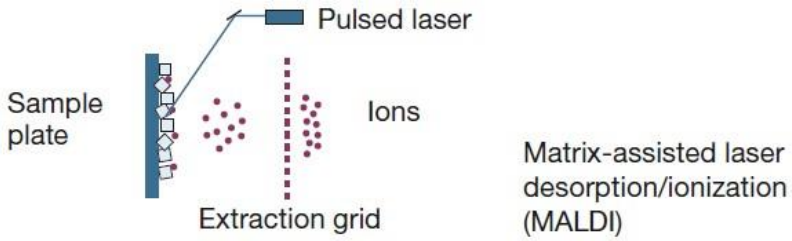

Figure 1.1.8 - Matrix-assisted laser desorption/ionization (MALDI). Adapted from[90]

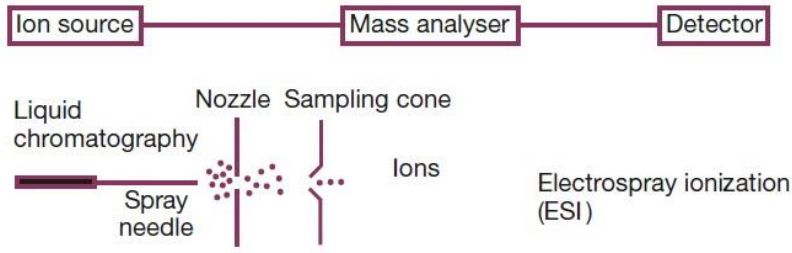

Figure 1.1.9 - Electrospray ionization (ESI). Adapted from[90]

Apart from the different ionization and volatilization techniques, a mass spectrometer can have one of four different types of mass analysers. Those are ion-trap, quadrupole, time-of-flight (TOF) and Fourier transform ion cyclotron (FT)[90]. Each one has differences regarding parameters such as mass accuracy, resolution, sensitivity and the capacity to generate mass spectra rich in information from the peptides present in a sample[92]. Sometimes, more than one mass analyser can be used so that combining them increases the performance of the mass spectrometer[90] (**Figure 1.10**). In an ion-trap mass analyser, the ionized fragments are trapped for a determined time period before MS analysis. Its main advantages are high sensitivity, robustness and not being as expensive when compared with other types of mass analysers. For this reason, a great part of the data generated in proteomics was produced using ion-trap analysers. The main disadvantage is low mass accuracy. A development that caused the increase in mass accuracy and also resolution and sensitivity was the linear ion trap in which the ions are trapped in a cylindrical volume larger than in the traditional ion trap[93]. In the Fourier transform ion cyclotron ions are also trapped but are captured in a strong magnetic field under a high vacuum. The main advantages of this mass analyser are resolution, mass accuracy and high sensitivity. However, the high cost and low peptide-fragmentation efficiency have limited their use[94]. Time-of-flight mass analysers are usually used coupled with MALDI and measure the mass of intact peptides. In mass spectrometers with this type of mass analyser the ionized peptides are accelerated to high speeds, turn around in a reflector and are separated according to their respective velocity. TOF-TOF mass spectrometers include two TOF segments separated by a collision cell where ions of a specific mass-to-charge ratio selected in the first TOF segment are fragmented before going into the second TOF section where the fragmented ions are separated by their mass[90]. Quadrupole mass spectrometers contain four rods with a time-varying electric field. In the first rod ions with a particular mass-to-charge ratio are selected and then fragmented in the next section, a collision cell. The fragments are then separated in the third segment. After separation ions are captured in a quadruple section and excited by a resonant electric field before being scanned to create the mass spectra[90]. ESI ion sources are mainly coupled with ion traps and quadrupole instruments measuring the mass of ionized peptide fragments of selected precursor ions generating collision-induced spectra (CID)[92]. Some MS equipment combine parts of different mass analysers. In a quadrupole-TOF mass spectrometer, the front part of the quadrupole has a TOF reflector segment to measure the mass of ions[90].

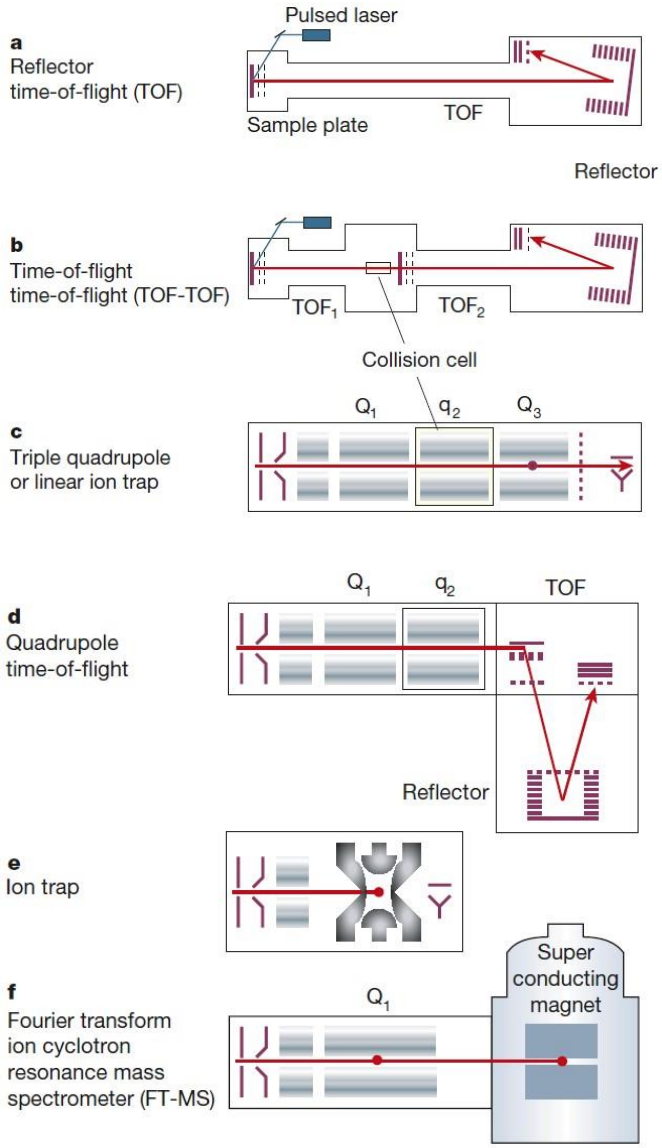

Figure 1.1.10 - Types of mass analysers. Adapted from[90]

There are two main data acquisition methods in mass spectrometry. Data-dependent acquisition (DDA) and data-independent acquisition (DIA). In DDA only certain peptide ions generated in the first cycle of MS with a selected mass-to-charge ratio are fragmented before MS/MS[95] (**Figure 1.11**). In DIA all the peptide ions generated in the first MS cycle are then fragmented and acquired. In this way, DIA gives a comprehensive map for a given sample while the selection of abundant fragment ions in DDA can lead to lower sensitivity and reproducibility between samples[96]. To solve the complexity of MS/MS spectra obtained by using DIA two different approaches can be applied, spectrum-centric or peptide-centric. In the spectrum-centric strategy the identifications are acquired after MS/MS spectra deconvolution with fragmentation coelution profiles and then using database search[97]. The peptide-centric

approach matches DIA spectra with mass-to-charge ratios and retention time (RT) values acquired from DDA spectral libraries[98].

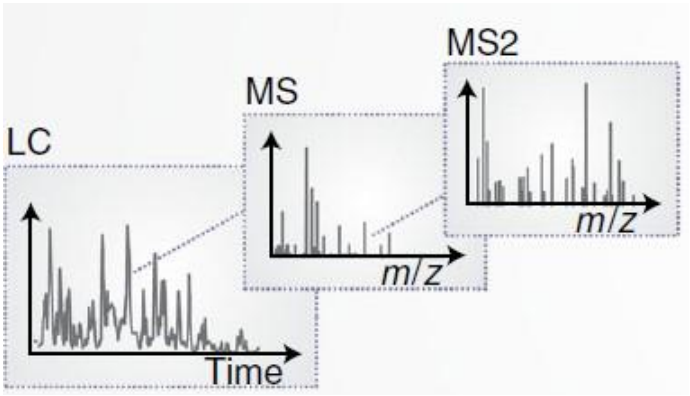

Figure 1.1.11 - MS/MS spectra. Adapted from[86]

The identification of peptides utilizing collision-induced spectra (CID) is more precise when compared with mass mapping since the peak pattern observed in CID spectra provides information not only about the peptide mass but also about the peptide sequence. The information can not be easily converted into an unambiguous peptide sequence[90]. That is the problem with de novo sequencing in mass spectrometry. Due to this limitation, proteome databases are used in several different software. The spectra obtained are scanned against the protein sequences. From the peak pattern, a short and precise amino acid sequence can be determined, usually referred to as a peptide sequence tag. Combining this peptide sequence tag, with mass information and using the protein sequence database is possible to determine the peptide[99]. The construction of theoretical mass spectra from the sequences in the database is another tool to improve peptide sequencing. The overlap of the theoretically predicted mass spectra with the measured mass spectra allows the determination of the best match and a more accurate determination of the correct peptide[100]. De novo sequencing uses collision-induced spectra of fragment ions to calculate an amino acid mass (**Figure 1.12**). The peptide ions are fragmented in the collision cell and MS/MS spectra are obtained for the amino acid residues ionized. The mass difference between two fragment ions allows the calculation of an amino acid mass of a given peptide and therefore determines the peptide sequence[101].

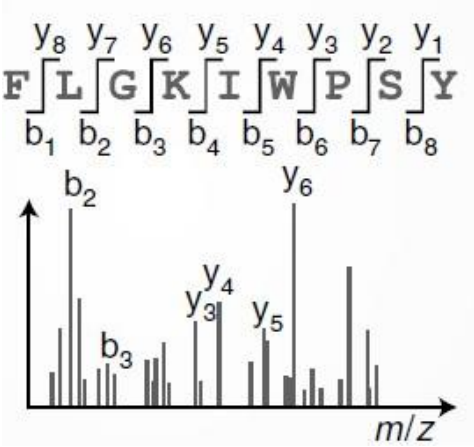

Figure 1.1.12 - Peptide identification. Adapted from[86]

1.4 **Peptide Vaccine**

Peptide-based vaccines utilize specific peptides or peptide combinations to initiate or intensify the T cell response against a pathogen or malignant cell and constitute a low side effect strategy of vaccination[102]. By making use of a peptide or a combination of peptides these vaccines can activate peptide-specific T cells. The peptides used in the formulation belong to the immunopeptidome and therefore are presented at the cell surface of HLA molecules in humans and recognized by TCR of $CD4^+$ and $CD8^+$ T cells[103].

The critical aspect in the formulation of a clinically effective peptide vaccine is the choice of the best antigen targets (**Figure 1.13**). Optimal antigens are those presented at the cell surface of the pathogen with higher frequency and for that reason allow an increased recognition by T cells[102]. At first glance may seem pretty simple but selecting the peptides that cause the best response is not that easy and several aspects have to be taken into consideration, for instance, the peptide length, the number of peptides used in the vaccine formulation and the choice between epitopes of $CD4^+$ or $CD8^+$ T cells. It has been shown in the last years that the combination of HLA class I presented antigens targets that stimulate $CD4^+$ T cell response with HLA class II presented antigens that activate $CD8^+$ T cells has a major benefit for the clinical outcome[103].

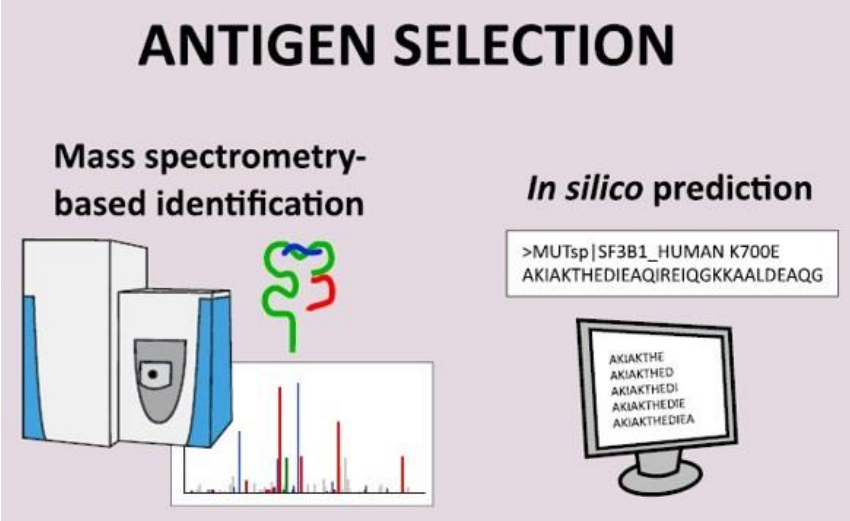

Figure 1.1.13 - Antigen selection for peptide-based vaccines. Adapted from[102]

Long peptides increase the T-cell response because as opposed to short peptides with an exact epitope that does not need any additional processing, long peptides require further priming and as a consequence are only presented by professional antigen-presenting cells[104]. Short antigens bind without any additional processing to HLA class I molecules of not only antigen-presenting cells but also other cells which can cause suboptimal T cell priming and lead to T cell tolerance[105]. Another advantage of using longer HLA II antigenic peptides is the fact that the same epitope can be presented to a $CD4^+$ T cell by different HLA II molecules. This promiscuity of HLA II target antigens distinguishes them from HLA I target antigens, in which HLA restriction of the peptides presented is much tighter. Hence, the presentation of the same epitope in different HLA class II molecules is beneficial when considering the individual differences in HLA allotypes. In this way, the vaccination of people with different HLA allotypes is possible by making use of the same peptide formulation[106]. Beyond being cost-effective, the peptide-based vaccine approach combining several different antigens in the same injection also attenuates the problem of antigen loss and immune escape risk[107]. The observation of patients exhibiting increased disease control supports the multi-peptide vaccine strategy[108].

Beyond the conjugation of different peptides, another way of increasing the efficacy of a peptide-based vaccine is the modification of the peptide sequence or its chemical modification[109]. By changing the amino acid sequence on the individual peptides it is possible to increase the peptide binding affinity for its respective HLA molecules and subsequently expand the T cell response[102]. To increase antigen uptake and as a result improve antigen presentation, the antigen peptides may be conjugated with ligands of pattern recognition receptors[110]. Other important factors for an optimal T-cell response generated by peptide vaccines are the use of strong adjuvants (**Figure 1.14**) and the choice of an appropriate route of delivery. These factors are important to make sure that the peptides delivered activate the immune

system properly and produce a cell-mediated immunity followed by activation and expansion of peptide-specific T cells[111]. Appropriate adjuvants have several roles, comprising peptide protection from degradation, peptide delivery to antigen-presenting cells and their activation. Toll-like receptors and their ligands generate a good activation of APCs and for that reason, strong adjuvants are often similar to TLR ligands[102].

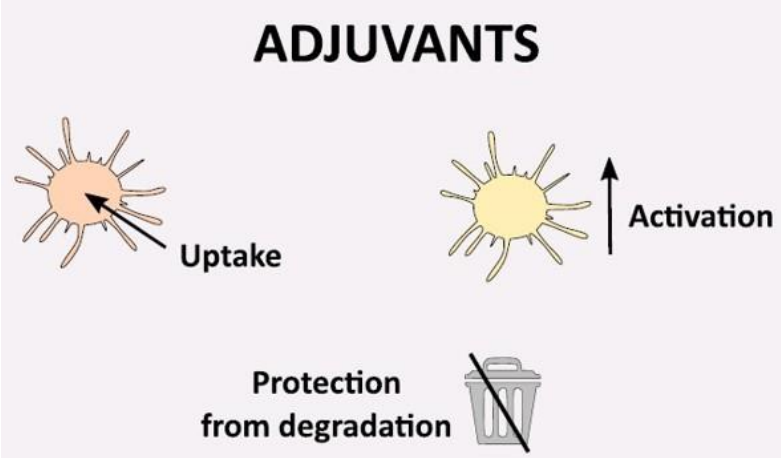

Figure 1.1.14 - Adjuvants main functions. Adapted from[102]

Until peptide-based vaccines can be widely used several aspects need to be improved such as the ones previously discussed ranging from the selection of the proper peptides to the use of potent adjuvants and delivery routes, chemical modifications and others[102].

2

OBJECTIVES

This project aimed to establish an immunopeptidomic platform for the identification of peptides presented by macrophages using mass spectrometry instrumentation and data analysis.

The first objective was the preparation of peptide samples. With that purpose, the work started with the growth of THP-1 monocytes and W/32 hybridoma antibody-producing cells. The antibody produced by W6/32 cells was purified and used for MHC immunoprecipitation of the THP-1 sample followed by purification of MHC class I peptides. The optimization of the protocols was also another important objective.

The second part of this work had as the main objective the identification and characterization of MHC class I peptides with mass spectrometry and using DeNovoGUI as the software of choice for the MS data analysis. Peptide length distribution, prediction of MHC binding affinity with NetMHCpan and binding motifs analysis with WebLogo and MHC Motif Atlas were done with the expectation of identifying and characterizing the most relevant peptides.

Finally, putting all these experimental protocols and data analysis steps working in conjunction for the validation of an immunopeptidomic platform was the ambition of this project. With a functional workflow, the possibility of uncovering peptides of interest for potential vaccine development would be available.

3

MATERIALS AND METHODS

3.1 BCG-GFP growth

Bacille Calmette-Guérin expressing green fluorescent protein (BCG-GFP) was cultured from BCG-GFP passage 19 in middlebrook 7H9 broth (Sigma-Aldrich) medium containing 10% middlebrook oleic acid albumin dextrose complex (OADC) enrichment (Thermo Fisher Scientific), 0.2% glycerol (Sigma-Aldrich) and 0.05% tylaxopol (Sigma-Aldrich) and incubated at 37ºC on a shaker at 200 rpm in aerobic conditions for eleven days.

3.1.1 OD

For eleven consecutive days, the optical density (OD) of BCG-GFP passage 20 was measured using a spectrophotometer at λ=600nm.

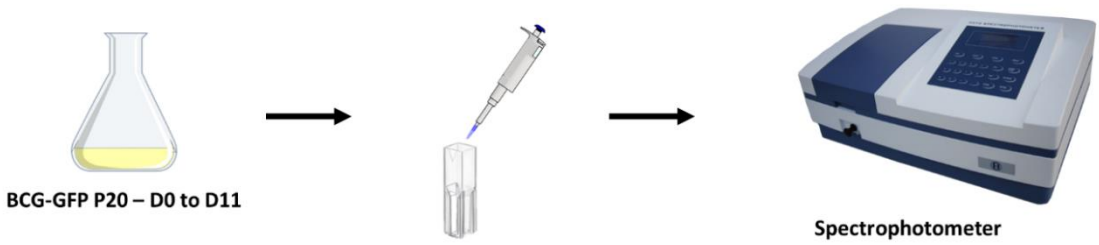

Figure 3.1 - OD measurement protocol overview

3.1.2 CFU

BCG-GFP P20 was cultured in middlebrook 7H10 agar medium (Sigma-Aldrich) for eleven consecutive days. A serial dilution was done with dilutions of 10^{-1}, 10^{-2} and 10^{-3} in triplicates. Then, the petri dishes were put in an incubator at 37º C in aerobic conditions for approximately a week. After a week, the number of colony-forming units (CFU) was then counted using the microscope and counting on the dots that belong to the dilution where the number of CFUs was comprised between 30 and 300 CFUs.

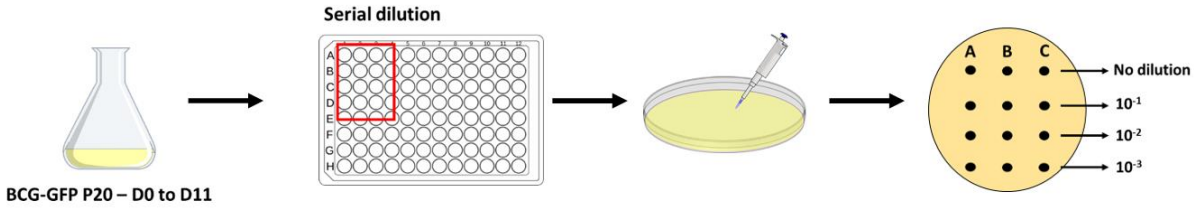

Figure 3.2 - CFU measurement protocol overview

3.1.3 Flow cytometry for % determination of BGG-GFP

Flow cytometry was used to determine the percentage of BCG-GFP that in fact are tagged with GFP in passage 20 compared to BCG-GFP recently defrosted. In a 24-well plate, wells were prepared as described in **Figure 3.3**. Wells A1, A2 and A3 were the blank, wells B1, B2 and B3 with BCG-GFP passage 0 and wells C1, C2 and C3 with BCG-GFP passage 20. Bacteria were resuspended in cytometry buffer (PBS + 2% FBS + 2 mM EDTA) and then the flow cytometry experiment was carried out in the BD Accuri™ C6 Plus Flow cytometer (**Figure 3.4**).

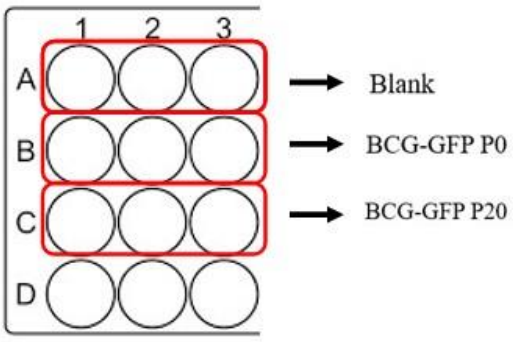

Figure 3.3 - Representation of the 24-well plate used in the experiment

Figure 3.4 - BD Accuri™ C6 Plus Flow Cytometer used for rate of infection determination

3.2 THP-1 growth

THP-1 monocytes (ATCC TIB-202) were defrosted from liquid nitrogen and then cultured in T-flasks with RPMI Medium 1640 (Gibco™ Chemicals Thermo Fisher Scientific) containing 10% FBS, 200mM L-glutamine (Gibco™ Chemicals Thermo Fisher Scientific) and 10 000 U/mL Penicilin-Streptomycine (PenStrep, Gibco™ Chemicals Thermo Fisher Scientific). Initially, 1 ml of cells were cultured in 9 mL of medium, a dilution of 1:10. From this dilution, a 1:100 dilution was prepared. Cells were transferred to new flasks with fresh medium every two to three days after centrifugation at 2000 rpm for 5 minutes. Before every transfer, cell counting was performed in a hemocytometer slide 10 with grids (KOVA™ Glasstic™) using 0.4% trypan blue solution (Sigma-Aldrich). The cells were incubated at 37º C with 5% CO_2 between every manipulation.

3.2.1 THP-1 sample preparation

After reaching the desired number of THP-1 monocytes in the T-flasks, cells were stimulated with 20 nm Phorbol 12-Myristate Acetate (PMA, Sigma-Aldrich) to into phagocytic macrophage-like cells and the cells were incubated overnight at 37º C with 5% CO_2. After 24 hours the culture medium was replaced and the cells incubated again overnight. 48 hours after PMA activation, the cells were washed with PBS. After this step, the cells from all T-flasks were removed with a cell scraper in a small volume of PBS to a 50 mL falcon. Then, the cells were centrifuged at 2000 rpm for 5 minutes, the supernatant was discarded, and the pellet stored at -80ºC.

3.3 W6/32 and L243 hybridoma growth

W6/32 (ATCC HN-95) and L243 (ATCC HB-55) hybridoma cells were defrosted from liquid nitrogen and then cultured in T-flasks with RPMI Medium 1640 containing 10% FBS, L-glutamine 200mM and PenStrep 10 000 U/ml. Initially, for each hybridoma type, 1 ml of cells were cultured in 9 mL of medium, a dilution of 1:10. From this dilution, a 1:100 dilution was prepared. Cells were transferred to new flasks with fresh medium every two to three days after centrifugation at 2000 rpm for 5 minutes. Before every transfer, cell counting was performed in a hemocytometer slide 10 with grids using 0.4% trypan blue solution. The cells were incubated at 37º C with 5% CO_2 between every manipulation. After W6/32 hybridoma cells growing in T-flasks reached 50 x 10^6 viable cells, they were transferred to two different bioreactors, 25 x 10^6 cells for each.

3.3.1 Corning CELLine CL1000 bioreactor

25 x 10^6 viable W6/32 hybridoma cells were centrifuged at 2000 rpm for 3 minutes and the pellet was resuspended with 15 mL of RPMI Medium 1640 containing 10% FBS, 200mM L-glutamine and 10 000 U/mL PenStrep. The 15 mL of medium with the hybridoma cells were added to the cell compartment (small lid) of the CELLine CL1000 Corning bioreactor (**Figure 3.5**) and 250 mL of RPMI Medium 1640 containing 200mM L-glutamine and 10 000 U/mL PenStrep were added to the medium compartment (big lid). Every three to four days the 15 mL of supernatant was collected and stored in 50 mL falcons at -20º C and the hybridoma cells were transferred again into the cell compartment after centrifugation at 2000 rpm for 5 minutes. The 250 mL of medium without FBS was changed as depicted in **Figure 3.6**. The hybridoma cells in the bioreactor were incubated at 37º C with 5% CO_2 between every manipulation.

Figure 3.5 - Corning CELLine CL1000 bioreactor

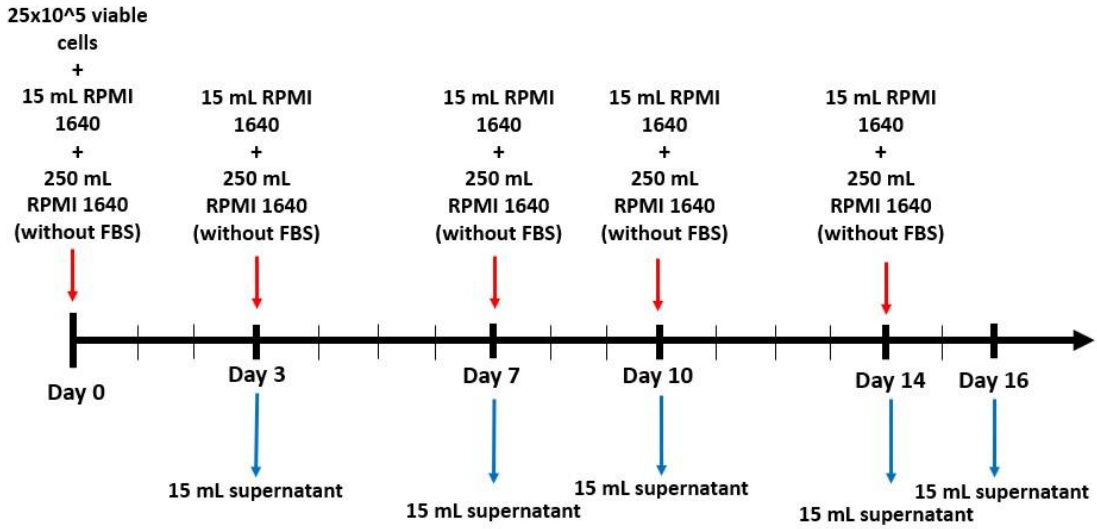

Figure 3.6 - Corning CELLine CL1000 bioreactor hybridoma cell growth workflow overview

3.3.2 KDBio EZ-Flask bioreactor

25 x 10^6 viable W6/32 hybridoma cells were added to the KDBio EZ-Flask bioreactor (**Figure 3.7**) and RPMI Medium 1640 containing 10% FBS, 200mM L-glutamine and 10 000 U/ml PenStrep was added until the final volume was 250 ml. After 4 days, another 250 mL of RPMI Medium 1640 containing 10% FBS, 200mM L-glutamine and 10 000 U/mL PenStrep were added to the bioreactor and on the seventh day more 500 mL were added for a total final volume of 1000 mL. After several days the medium containing W6/32 hybridoma cells was removed and stored in 50 mL falcons at -20º C (**Figure 3.8**). The hybridoma cells in the bioreactor were incubated at 37º C with 5% CO_2 between every manipulation.

Figure 3.7 - KDBio EZ-Flask bioreactor

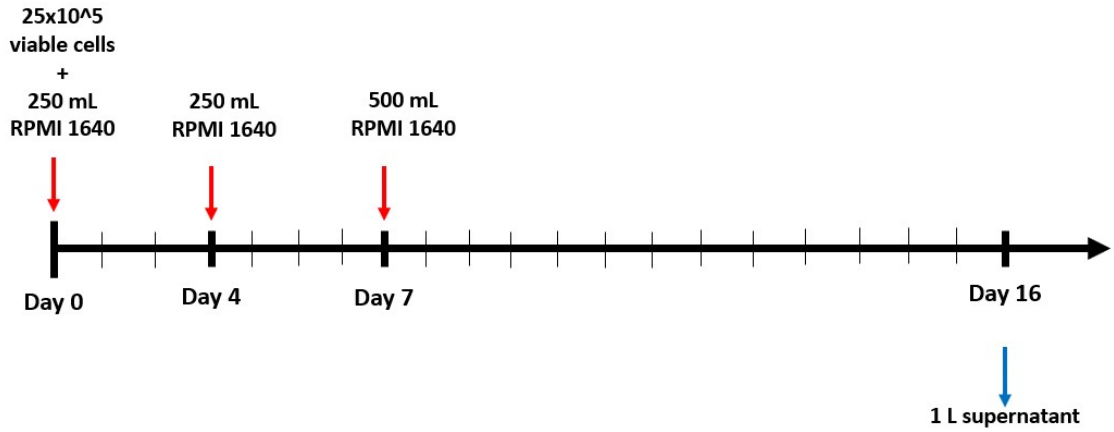

Figure 3.9 - KDBio EZ-Flask bioreactor hybridoma cell growth workflow scheme

3.3.3 Antibody purification

An Econo-Column® Chromatography Column, 1.5 x 10 cm (BIO-RAD) was washed with 10% acetic acid (Honeywell), tap water and milli-Q water. In an erlenmeyer flask Sepharose with Protein A (Bio-Vision) beads were resuspended on the hybridoma cell supernatant in a ratio of 500 mL of supernatant: 1 mL of beads and incubated for 1 hour, slowly mixing. After this step, the supernatant was passed into the column and washed with PBS (Nzytech). 100 Mm glycine (Thermo Scientific) at pH 2.5 was added on top of the beads to elute the antibody-bound protein A and collect to a falcon containing 1M tris(hydroxymethyl)aminomethane TRIS (VWR) at pH 9.2 in a proportion of Glycine: TRIS of 20:1. Concentration of purified antibody was quantified using NanoDrop™ One (Thermo Fisher Scientific) measuring A_{280nm}, and the purified antibody stored at -20ºC.

3.4 MHC-peptide complex immunoprecipitation

THP-1 cell sample prepared in **3.2.1** was defrosted and subjected to MHC-I immunoprecipitation for the MHC class I complexes. The protocols used are described in the next chapters (**3.4.1**, **3.4.2** and **3.6.3**).

3.4.1 Preparation of cross-linked beads

The first step of MHC immunoprecipitation is the preparation of cross-linked beads. Firstly, with Protein A Sepharose beads are resuspended by agitation and centrifuged for 1 minute at 1000g to pellet the beads. The supernatant containing ethanol is removed. Then, the antibody produced by W6/32 and L243 hybridoma cells is added in a sufficient volume to the Sepharose with Protein A beads and incubated slowly at room temperature for 30 minutes to promote the binding of the antibody to the beads. Approximately 1 mg of antibody was used for each 10^8 cells. An Econo-Column® Chromatography Column, 1.5 x 10 cm was used to promote the cross-linking reaction between the antibody and the protein A on the surface of the Sepharose beads. The column was washed with 10% acetic acid to remove polymeric substances and other contaminants and with PBS. The pH was measured to make sure it was neutral before advancing. The beads were loaded into the column and washed with 10 column volumes (1 column volume is the volume of beads used) of borate buffer prepared with 0.1 M boric acid (VWR), 0.1 M potassium chloride (KCl, VWR) and 0.1M sodium hydroxide (NaOH, VWR) to pack the column and remove free amines. After that, 10 column volumes of cross-link solution dimethyl pimelimidate dihydrochloride (DMP, Sigma-Aldrich) were added to the column and run almost completely. The flow was stopped when the meniscus was above the beads and the incubation took place for 30 minutes at room temperature. To terminate the cross-linking reaction, the beads were washed two times with 10 column volumes of ice-cold 0.2 M TRIS at pH 8.0. Then, the beads were washed with 10 column volumes of borate buffer to neutralize the column and the pH of the column was measured to verify that was neutral.

3.4.2 Preparation of reagents for HLA purification

For the HLA purification protocol, several solutions were required to wash the beads within the column. These were W1 – 50 mM TRIS pH 8.0, 150 mM sodium chloride (NaCl, VWR), 5 mM ethylenediaminetetraacetic acid (EDTA, Invitrogen™); W2 - 50 mM TRIS pH 8.0, 150 mM NaCl; W3 - 50 mM

TRIS pH 8.0, 450 mM NaCl; and W4 - 50 mM TRIS pH 8.0. For the lysate preparation 2x lysis buffer - 1% octylphenyl-polyethylene glycol (IGEPAL, Sigma-Aldrich), 300 mM NaCl, 100 Mm TRIS pH 8.0 with protease inhibitor cocktail tablet (UltraCruz®) - was prepared.

3.4.3 HLA purification

For each one of the THP-1 cell samples, 0.5 mL of 2x ice-cold lysis buffer was added per 10^8 cells to obtain a homogenous lysate solution. Before adding lysis buffer, the cells were defrosted on ice to avoid the release of DNA. Then, the lysate was incubated on ice for 30 minutes with agitation for the solubilization of the MHC complexes. The next step was to centrifuge the lysate at 500g for 10 minutes at 4º C to remove the nuclei and the supernatant was transferred into multiple 2 mL LoBind Eppendorf tubes. The pellet was preserved and stored for a possible future DNA isolation. The lysate was again centrifuged at 20 000 g for 45 minutes at 4º C to clear the lysate from other possible insoluble components. After this, the beads within the column were transferred to a falcon tube and the supernatant from the previous centrifugation was added directly to the beads with cross-linked antibody previously prepared. The beads were incubated overnight in the fridge with the lysate in the falcon with agitation. The day after, the beads were loaded in the Econo-Column® Chromatography Column, 1.5 x 10 cm. Again, the column was washed with 10% acetic acid to remove polymeric substances and other contaminants and with PBS. The pH was measured to make sure it was neutral before advancing to the next step. The flow-through lysate was collected and stored in a falcon for further analysis. Then, the beads were washed with 10 column volumes of W1 (50 mM TRIS pH 8.0, 150 mM NaCl, 5 mM EDTA) and 10 column volumes of W2 (50 mM TRIS pH 8.0, 150 mM NaCl) to remove detergent. Then, were washed with 10 column volumes of W3 (50 mM TRIS pH 8.0, 450 mM NaCl) to remove material non-specifically bound and lastly with 10 column volumes of W4 (50 mM TRIS pH 8.0) to remove salt. After this, the column was washed with PBS. The MHC complexes were eluted with 5 column volumes of 10% acetic acid and collected in multiple 2 mL LoBind Eppendorf tubes. The beads were recovered with PBS for a falcon. The MHC complexes samples on the several LoBind Eppendorf tubes were then dried down completely by vacuum centrifugation in the SpeedVac Concentrator plus Eppendorf overnight and then stored at -20º C until the peptide purification protocol was conducted.

3.5 Peptide purification

Following MHC immunoprecipitation, the peptides were purified from the MHC-peptide complexes. The MHC complexes samples previously purified and stored in LoBind Eppendorf tubes at -20º C were resuspended with 0.1% trifluoroacetic acid (TFA, Sigma-Aldrich) in 1% acetonitrile (ACN, VWR) (v/v) and then centrifuged at 15 000 rpm for 5 minutes. The C18 cartridge (Waters) column was used for the purification of peptides. Firstly, the column was conditioned with 80% ACN in 0.1% TFA (v/v). Then, the column was equilibrated with 0.1% TFA and the supernatant of the sample resuspended in 0.1% TFA in 1% ACN (v/v) was loaded in the cartridge. After loading the sample, the cartridge was washed with 0.1% TFA (v/v). The peptides were eluted for LoBind Eppendorf tubes with 30% ACN in 0.1% TFA (v/v). After elution, the purified peptide samples were dried down completely overnight by vacuum centrifugation using SpeedVac Concentrator plus Eppendorf and then stored at -20º C for posterior use in mass spectrometry experiments. The C18 cartridge columns were stored at -4º C degrees for a possible recovery of the HLA heavy chains and beta-2-microglobulin. These can be eluted using 80% ACN in 0.1% TFA (v/v).

3.6 Mass Spectrometry

MS analysis was performed using UltiMate 3000 ultra-high performance liquid chromatographer, from Thermo Scientific, coupled to UltraHigh Resolution Quadrupole Time-Of-Flight (UHR-QTOF) IMPACT HD mass spectrometer, from Bruker.

Peptides were resuspended in 20 µl of 1% ACN in 0,1% TFA, and 7 µl of peptide sample were loaded on-to a µPAC™ Trapping column and desalted for 2.7 min with 1% (v/v) Acetonitrile (ACN) in 0.1% aqueous formic acid (FAaq) at a flow rate of 15 µL min -1. Then the peptides were separated using an analytical column (200 cm µPACTM PharmaFluidics). With a linear gradient at 500 nL min^{-1} (mobile phase A: FAaq 0.1% (v/v); mobile phase B: 99.9% (v/v) ACN and 0.1% (v/v) FAaq) 0-2 min from 3% to 5% of mobile phase B, 5-76 min from 5% to 17% of mobile phase B, 76-104 17% to 25% B, 104-121 25% to 35% B. Chromatographic separation was carried out at 35 ºC. MS acquisition was set to MS (2 Hz) cycles, followed by MS/MS (8-32 Hz), cycle time 3.0 seconds, active exclusion, exclude after one spectrum, release after 2 min. The precursor was reconsidered if its current intensity was 3.0 higher than the previous intensity and intensity threshold for fragmentation of 2500 counts.

3.7 Data analysis

MS data were analysed with DeNovoGUI CompOmics for identification of peptide sequences. Spectra were matched to human SwissProt reviewed entries of UniProt proteomes. The searches were performed with no static or variable modifications, precursor mass tolerance of 25 ppm and fragment mass tolerance of 0,07 Da.

4

RESULTS AND DISCUSSION

4.1 BCG-GFP growth

BCG-GFP was cultured in middlebrook 7H9 broth medium for eleven days and plated in middlebrook 7H10 agar medium to measure the optical density and the number of colony-forming units in that period of eleven days and construct a plot showing the relationship between these two variables along the time the bacterial culture was growing. OD allows us to follow only three of the four phases of bacterial growth. This occurs because after the lag phase, exponential phase and stationary phase and upon entering the death phase it is not possible to detect the latter because optical density is determined by the absorbance and therefore, not only the viable bacteria are contributing to the OD value measured but also non-viable bacteria. Using the values of OD measured from day zero (D0) to day eleven (D11) the plot shown below was constructed (**Figure 4.1**).

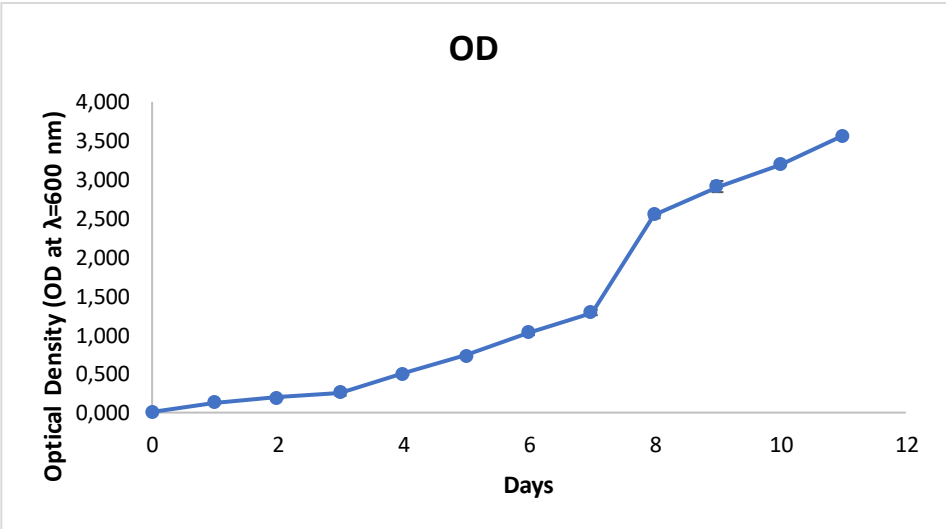

Figure 4.1 - BCG-GFP optical density from D0 to D11

As we can see, OD increases slightly until day three. From day three until day seven the increase is a bit more significant. From day seven to day eleven there is a great increase in the values measured, due to the entry in the exponential phase of bacterial growth. However, the bigger increase is from day seven to day nine so possibly the bacterial culture was already entering the stationary phase by day ten or eleven. For the reason mentioned before, it is not possible to observe the stationary phase precisely and also the death phase because the light transmission is blocked by non-viable bacteria and debris.

At the same time, BCG-GFP was plated in middlebrook 7H9 agar every day. After 1 week of incubation at 37º C, the CFU was quantified and the results are shown below (**Figure 4.2**). The number of CFUs indicates the presence of viable bacteria capable of forming small colonies. With this method, the problem created by non-viable bacteria is eliminated. Comparing the graph of OD with the graph for CFU per mL it is possible to see the similarity from day zero to day nine which shows us that in this period, mostly the viable bacteria were responsible for the OD value measured. However, there is a difference between the two. For the OD the biggest increase occurs between day seven and day eight while for the number of CFUs, the biggest increase occurs between day eight and day nine.

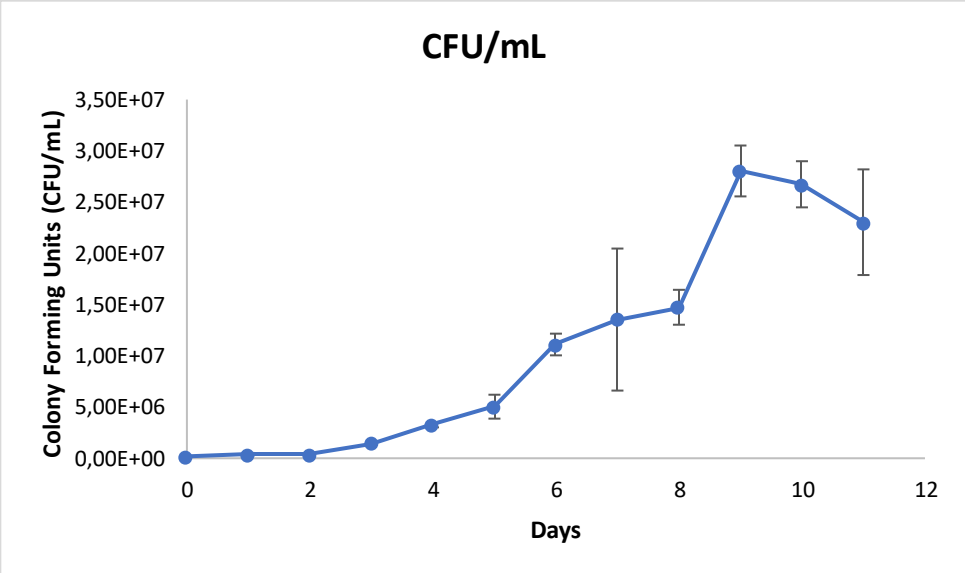

Figure 4.2 - BCG-GFP colony-forming units from D0 to D11

The results from optical density and CFU allow the construction of an OD vs. CFU/mL plot (**Figure 4.3**), establishing a correlation between these two bacterial growth indicators with a line function and

an R^2 coefficient. For the correlation between the two variable indicators to be considered strong, the R^2 coefficient should be higher than 0,9. In our case, the coefficient = 0,9139. The correlation is shown on **Figure 4.3**.

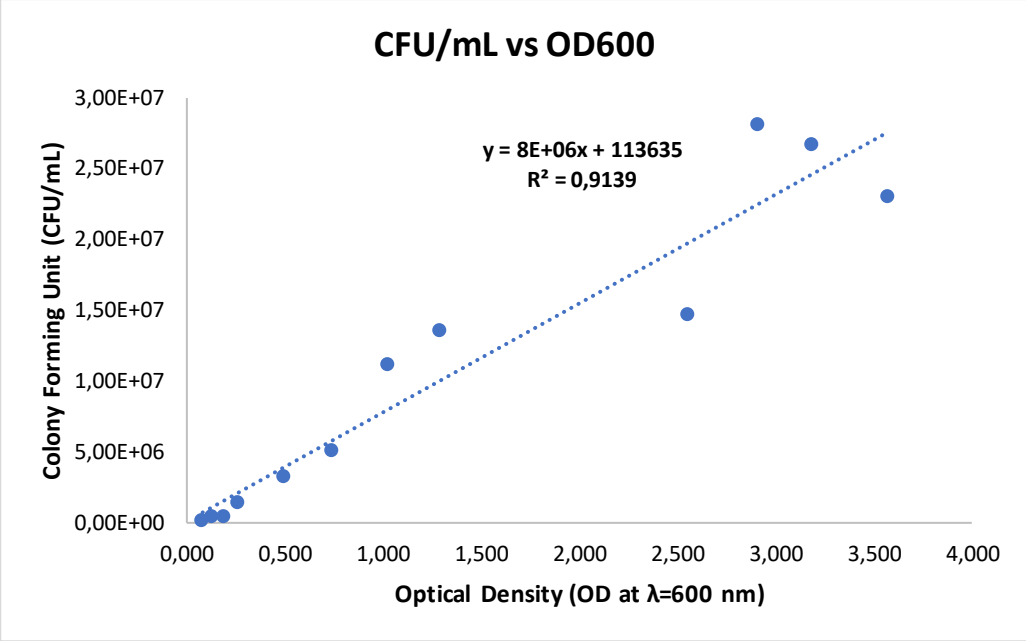

Figure 4.3 - CFU/mL vs OD600nm from D0 to D11

From this graph, it is possible to measure only the optical density and correlate to the bacterial cell count of a culture. The growth of BCG-GFP occurred as expected and a good line function and R^2 coefficient were obtained, making these plot useful for future applications of this bacterial strain.

4.1.1 Flow cytometry for percentage determination of BGG-GFP

To confirm that BCG-GFP are expressing the green fluorescent protein (GFP) the flow cytometry experiment comparing BCG-GFP passage 0 recently defrosted at the time and BCG-GFP passage 20 was performed. As BCG-GFP emits green fluorescence with GFP, the flow cytometry filter used was fluorescein isothiocyanate (FITC). In the first graph (**Figure 4.4**) is represented side scatter versus forward scatter as a way to distinguish the bacteria from debris. The forward scatter is a relative measure of the cell size and the side scatter a relative measure of the granularity and complexity of the cell. In that way, using the first graph (**Figure 4.4**) a gate was established that potentially includes the majority of the bacteria. From the first gate, a second graph was obtained but now a forward scatter area versus forward scatter height (**Figure 4.5**). This graph was used to select only the singular cells from non-single cells

or doublets or even clumps of bacteria or debris. The second gate was defined in this graph selecting only the dots that follow a linear relation between the forward scatter height and area and 79% of the events were selected for the construction of the next graph. For the third graph (**Figure 4.6**), it was now used a measure of green fluorescence by selecting the FITC channel filter to measure GFP fluorescence. As previously determined using a negative control, the fluorescence obtained in the third graph (**Figure 4.6**) is caused by the autofluorescence of the bacteria and not by GFP. In this case, we only consider GFP fluorescence when above 10^3. For the fourth graph (**Figure 4.7**), a plot of event count versus FITC fluorescence and a gap for the positive fluorescence events was defined, so when positive events appear in the gap a percentage from the total events is presented. These fourth graphs (**Figure 4.4**, **Figure 4.5**, **Figure 4.6** and **Figure 4.7**) were the results for the control sample and no fluorescence in the FITC filter was detected.

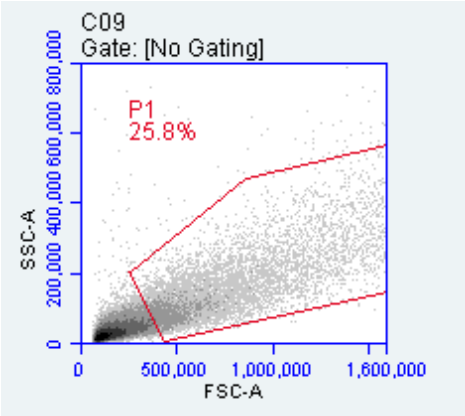

Figure 4.4 - Blank: side scatter vs forward scatter

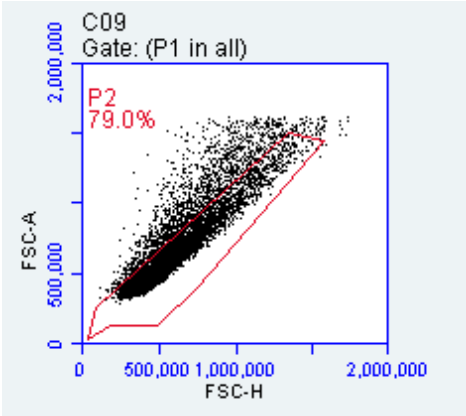

Figure 4.5 - Blank: forward scatter area vs forward scatter height

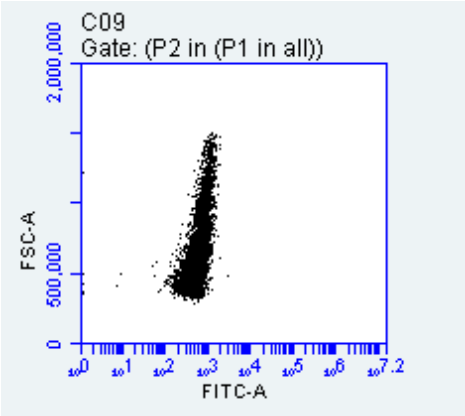

Figure 4.6 - Blank: event count vs FITC-A

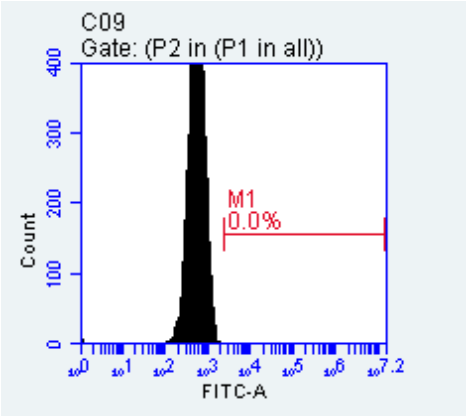

Figure 4.7 - Blank: forward scatter area vs FITC-A

With the gates defined in the blank, those gates were applied to the two samples. For these two samples, only the fourth graph with a plot of event count versus FITC-A fluorescence is here shown. First, for BCG-GFP passage 0 (**Figure 4.8**) and then for BCG-GFP passage 20 (**Figure 4.9**). The first three graphs for each one of these two samples are depicted in the annexes section (**Figure A1**, **Figure A2**, **Figure A3**, **Figure A4**, **Figure A5** and **Figure A6**). In the graph with BCG-GFP passage 0 is visible that approximately 37,7% of the vents are within the gap established previously for positive events, meaning that only that percentage of bacteria have emitted green fluorescence because of GFP. In the case of BCG-GFP passage 20, that percentage is even reduced to only approximately 14,2%, which is an approximated reduction of 25,2% along twenty bacterial passages to new Erlenmeyers and fresh medium.

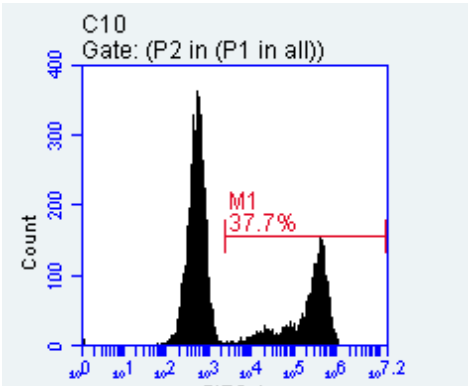

Figure 4.8 - BCG-GFP P0: event count vs FITC-A

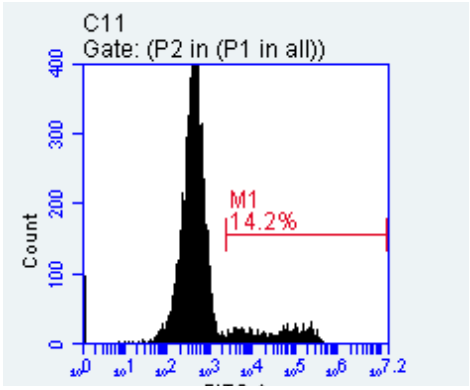

Figure 4.9 - BCG-GFP P20: event count vs FITC-A

The results obtained were then compared all together in one single figure, containing the comparison between the mean FITC fluorescence of the three samples (**Figure 4.10**). The results of BCG-GFP growth and the percentage of bacteria with GFP help the characterization of the bacteria and increase the expertise in bacterial manipulation.

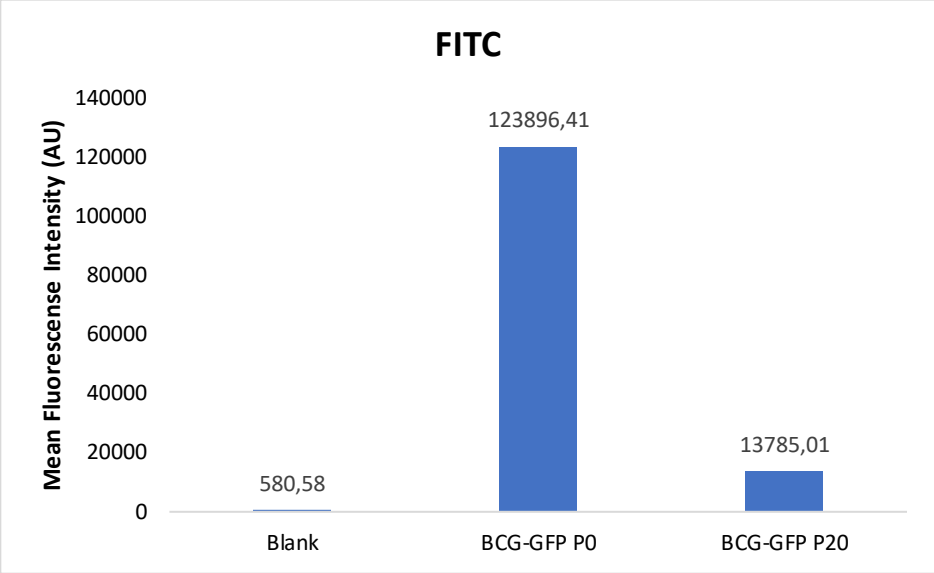

Figure 4.10 - FITC fluorescense for the three samples

4.2 Yield comparison between two bioreactors

For the MHC immunoprecipitation experiment, antibodies that bind to MHC class I molecules and to MHC class II molecules are required. W6/32 hybridoma cells produce antibodies that bind to MHC class I molecules while L243 hybridoma cells produce antibodies that bind to MHC class II molecules. W6/32 and L243 cells were initial grown in T-flasks. To produce a higher titer of antibody with a higher yield and reduce the waste, growing hybridoma cells were transferred to two different types of bioreactors, the Corning CELLine CL1000 bioreactor and the KDBio EZ-Flask bioreactor, thus eliminating the need of large number of T-flasks every time the cells need new medium. Despite the two bioreactors being a better alternative to the use of T-flasks a comparison between the two was made to improve the workflow. For a period of 16 days, both bioreactors were used to produce antibodies with W6/32 hybridoma cells using the same number of cells to start. At the end of the 16 days, the total amount of antibody produced was purified and compared to assess which bioreactor had the highest yield. The last step of the antibody purification protocol was the measurement of, at A_{280nm} using NanoDrop™ One (Thermo Fisher Scientific), to determine the antibody concentration. For each bioreactor, the amount of antibody present in the medium stored during the 16 days was purified twice in consecutive days and the results are shown in the table below (**Table 4.1**). From A_{280nm} measurement it was observed that the antibody purified originated in hybridoma cells that were growing in the Corning CELLine CL1000 bioreactor and has a higher concentration compared with the antibody with origin in the KDBio EZ-

Flask bioreactor (**Table 4.1**). However, within 16 days the volume of medium extracted from the CL1000 bioreactor was much lower compared to the amount extracted from the EZ-Flask bioreactor. This difference is also reflected in the final volume obtained at the end of the antibody purification protocol (**Table 4.1**). Considering the final concentration of the purified antibody and also the final volume obtained it is possible to calculate the mass of antibody that was possible to produce using each one of the bioreactors. The result of those calculations is shown in **Table 4.1**.

Table 4.1 - Total mass calculations of antibody produced

		Concentration (mg/mL)	Total Volume (mL)	Total mass (mg)
Corning CL1000	**1st purification**	1,132	2	2,264
	2nd purification	1,212	2	2,424
KDBio EZ-Flask	**1st purification**	0,641	20	12,82
	2nd purification	0,315	20	6,3

The combined final amount was 4,688 mg of W6/32 antibody produced by CL1000 bioreactor and 19,12 mg of W6/32 antibody produced by EZ-Flask bioreactor (**Table 4.2**).

Table 4.2 - Total antibody mass produced from two different bioreactors

	Total mass (mg)
Corning CL1000	4,688
KDBio EZ-Flask	19,12

Despite the bigger antibody concentration measured for CL1000 the higher volume produced by EZ-Flask makes it more advantageous. Not only the antibody mass produced is higher but the KDBio EZ-Flask is easier to manipulate, requires less labor and therefore less material and reagents. Consequently, it is a cheaper and a better option when compared with the Corning CL1000 bioreactor.

4.3 Peptide Identification

MS data from the sample with 5 x 10^8 THP-1 macrophage cells were analysed with DeNovoGUI CompOmics software for the identification of peptide sequences. For that sample, two injections in the mass spectrometer were performed and as a result, two sets of peptides were identified. The two different sets were named 4204 and 4205. The first step of the data analysis was to put together all peptides from the

two sets and eliminate all peptide duplicates. A number of 2913 unique MHC class I peptides were identified after the removal of duplicated peptides and then distributed in function of the peptide amino acid length, from eight to thirty amino acids (**Figure 4.12**). The overlap between the peptides identified in 4204 and 4205 is depicted in **Figure 4.11**. The percentage of peptides for each amino acid length was calculated and is also presented in **Figure 4.12**.

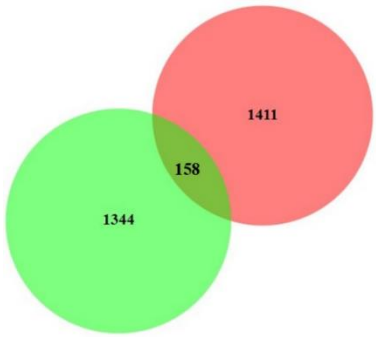

Figure 4.11 - Overlap between replicates of THP-1 peptides presented by MHC-I

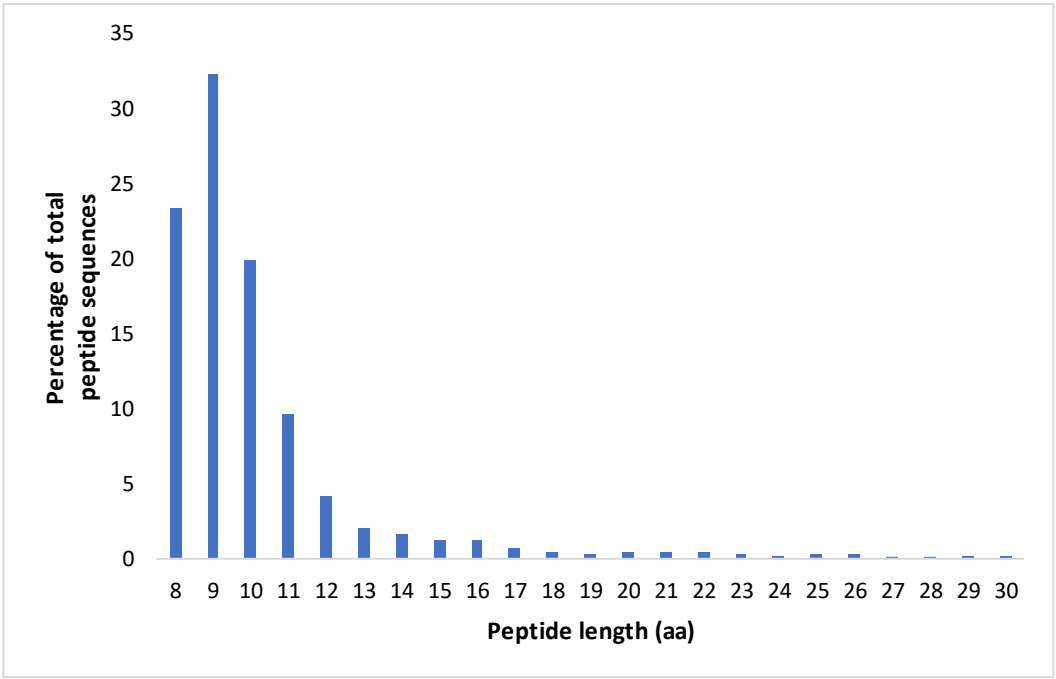

Figure 4.12 - Peptide distribution by length

The peptides identified showed a typical distribution of amino acid length with a predominant peak of 9-mers but also a high percentage of 8-mers and 10-mers as expected from the knowledge on MHC class I peptides[112]. 88% of the peptides had an amino acid length comprised between eight and eleven amino acids (**Figure 4.12**). The remaining 12% were perhaps non-specific peptides also eluted during MHC immunoprecipitation or peptide purification. 75% percent of the total unique MHC class I peptides were 8-mers, 9-mers and 10-mers.

4.4 HLA binding affinity

The 2913 unique MHC class I peptides were distributed by length and the peptides with eight, nine and ten amino acids were then ranked using NetMHCpan 4.1 as strong or weak binders for HLA class I molecules. NetMHCpan 4.1 predicts the binding affinity of peptides for different HLA allotypes. The MHC-I peptides with a NetMHCpan 4.1 rank lower than 0,5 were considered strong binders and the MHC-I peptides with a rank between 0,5 and 3 were considered weak binders. The peptides with a rank above 3 were excluded. The peptides were ranked for five different HLA allotypes and these were HLA-A*02:01, HLA-B*15:1, HLA-A*15:15, HLA-A*03:03 and HLA-A*03:13[113]. The total number of 8-mers, 9-mers and 10-mers that were strong and weak binders for each of the five allotypes is shown in **Figure 1.13**. The number of weak binders predicted by NetMHCpan 4.1 was higher compared to the number of strong binders for all the different five allotypes. The distribution of weak and strong binders for each allotype is represented in **Figure 1.13**. From all the 2913 unique MHC class I peptides identified with DeNovoGUI, 288 were predicted as strong binders which corresponds to 9,8% of the total number of peptides. By their turn, 669 peptides were predicted as weak binders which converts to 22,9%.

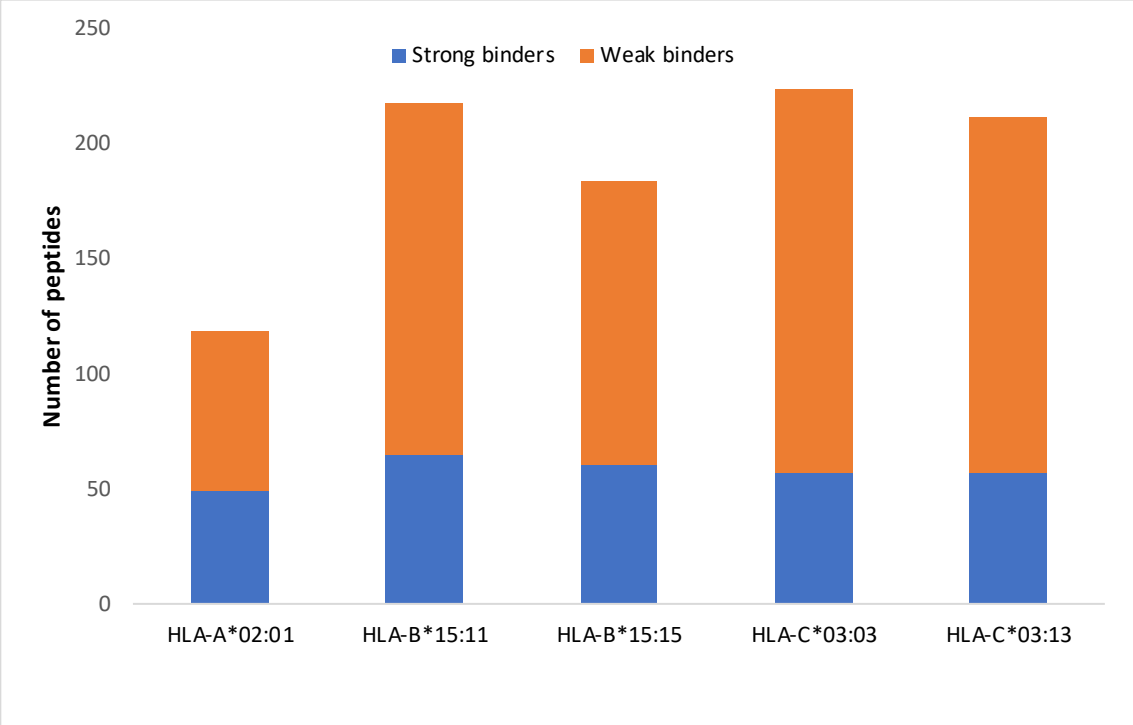

Figure 4.13 - Distribution of strong and weak binders by allotype

4.5 Peptide Binding motifs

Deconvolution of the binding motifs of the peptides predicted as strong and weak binders by NetMHCpan 4.1 was performed using WebLogo. For each allotype of the five previously referred the binding motifs of 8-mers (**Figure 4.14**, **Figure 4.16, Figure 4.18, Figure 4.20** and **Figure 4.22**), 9-mers (**Figure 4.24**, **Figure 4.26, Figure 4.28, Figure 4.30** and **Figure 4.32**) and 10-mers (**Figure 4.34**, **Figure 4.36, Figure 4.38, Figure 4.40** and **Figure 4.42**) are represented below, all on the left side. For all allotypes with the exception of HLA-A*02:01 for 8-mers, the results were expected as the most conserved amino acids are in the position of conserved pockets of MHC class I molecules, amino acid in position 2 and amino acid in position 9[114]. The result for HLA-A*02:01 with 8-mers is due to the low number of predicted strong and weak binders for that specific allotype. However, is still possible to see two conserved amino acids in position 8, in the amino acid of the C-terminal. Amino acids leucine and valine are well conserved for 8-mers, 9-mers and 10-mers in the last amino acid of the sequence, alanine is conserved in 9-mers and 10-mers and amino acids leucine and isoleucine are both well conserved in 9-mers and 10-mers as the second amino acid of the sequence, all for HLA-A*02:01 binders. Amino acids proline, valine and alanine are all conserved as the amino acid in position 2 for HLA-B*15:11 allotype binders, as well as tyrosine, phenylalanine and leucine in the last position of both HLA-B*15:11

and HLA-B*15:15 binders. Leucine, methionine and phenylalanine are amino acids conserved in the C-terminal of both HLA-C*03:03 and HLA-C*03:13 allotypes binders in 8-mers, 9-mers and 10-mers. Alanine and isoleucine are well conserved as the second amino acid of 8-mers and 9-mers and in the third residue position for 10-mers of HLA-C*03:03 and HLA-C*03:13 allotypes. Amino acids located more in the center of the peptide are not conserved as well as the first amino acid of 10-mers.

Figure 4.14 - HLA-A*02:01 obtained binding motif for 8-mers

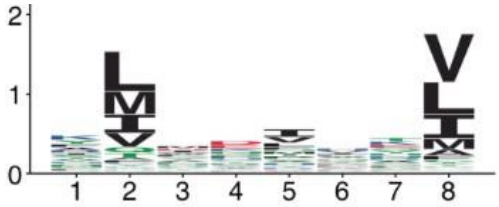

Figure 4.15 - HLA-A*02:01 consensus binding motif for 8-mers

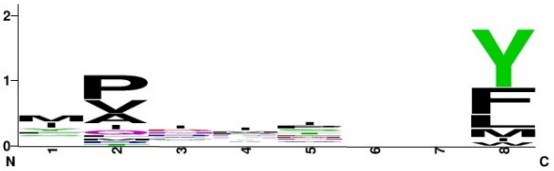

Figure 4.16 - HLA-B*15:11 obtained binding motif for 8-mers

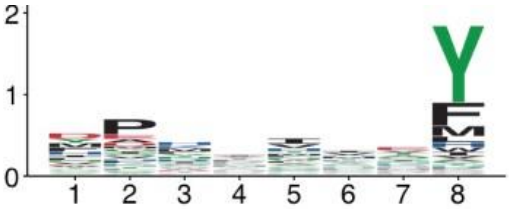

Figure 4.17 - HLA-B*15:11 consensus binding motif for 8-mers

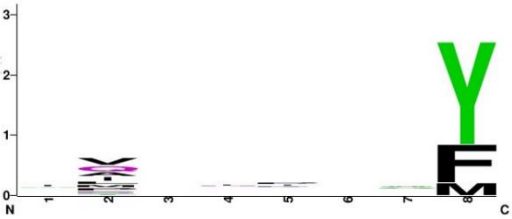

Figure 4.18 - HLA-B*15:15 obtained binding motif for 8-mers

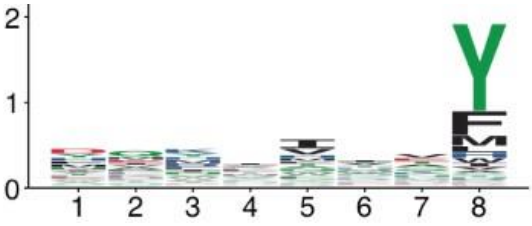

Figure 4.19 - HLA-B*15:15 consensus binding motif for 8-mers

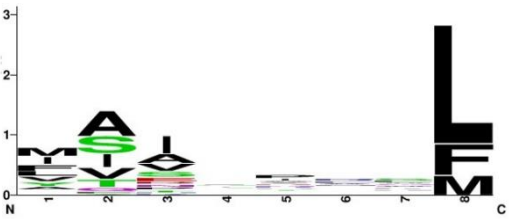

Figure 4.20 - HLA-C*03:03 obtained binding motif for 8-mers

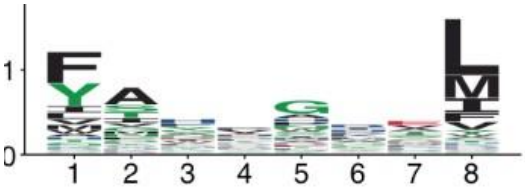

Figure 4.21 - HLA-C*03:03 consensus binding motif for 8-mers

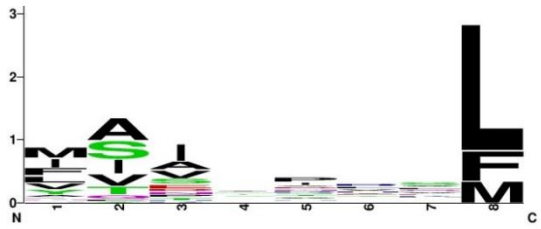

Figure 4.22 - HLA-C*03:13 obtained binding motif for 8-mers

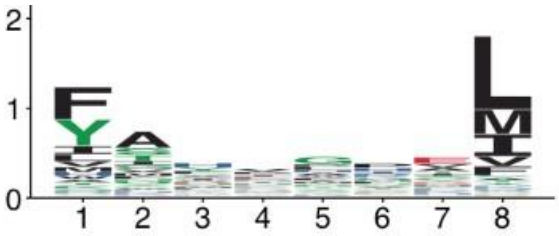

Figure 4.23 - HLA-C*03:13 consensus binding motif for 8-mers

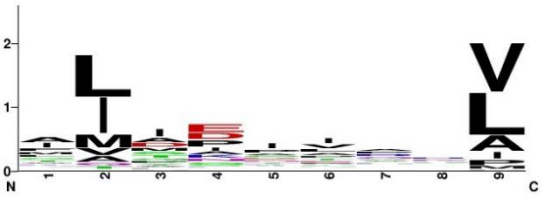

Figure 4.24 - HLA-A*02:01 obtained binding motif for 9-mers

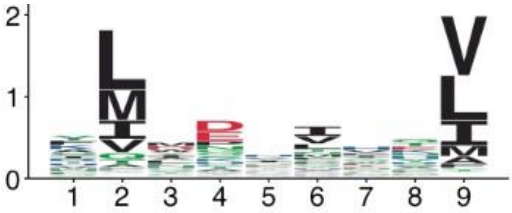

Figure 4.25 - HLA-A*02:01 consensus binding motif for 9-mers

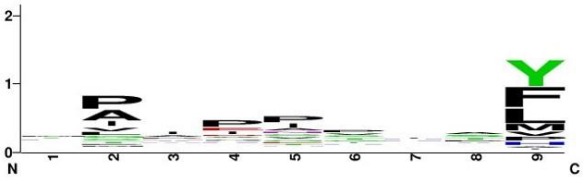

Figure 4.26 - HLA-B*15:11 obtained binding motif for 9-mers

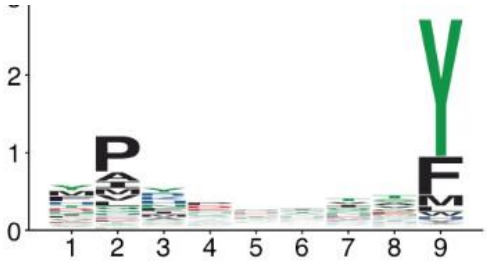

Figure 4.27 - HLA-B*15:11 consensus binding motif for 9-mers

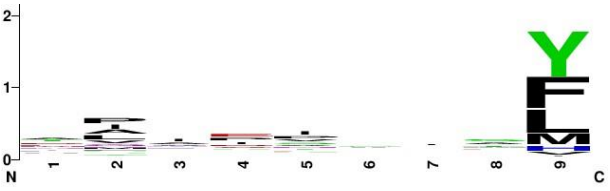

Figure 4.28 - HLA-B*15:15 obtained binding motif for 9-mers

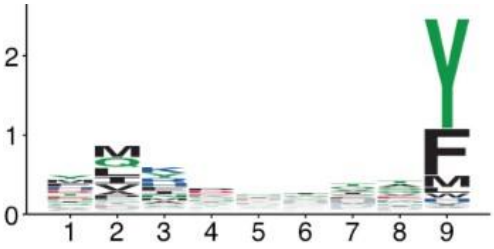

Figure 4.29 - HLA-B*15:15 consensus binding motif for 9-mers

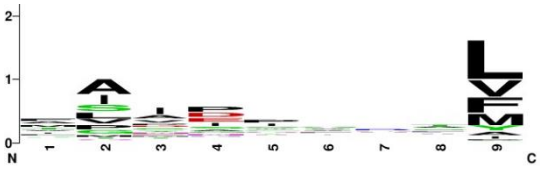

Figure 4.30 - HLA-C*03:03 obtained binding motif for 9-mers

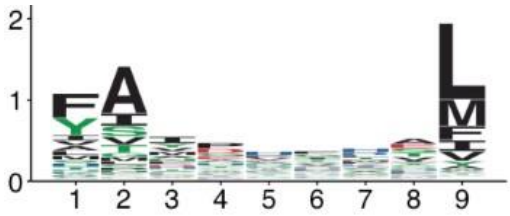

Figure 4.31 - HLA-C*03:03 consensus binding motif for 9-mers

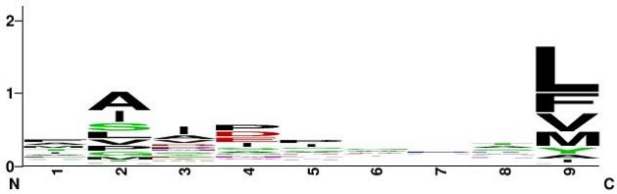

Figure 4.32 - HLA-C*03:13 obtained binding motif for 9-mers

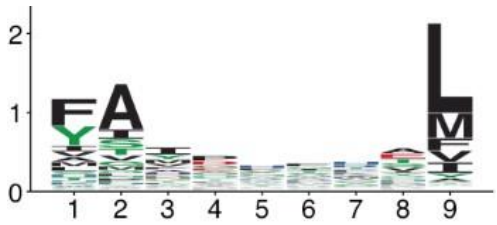

Figure 4.33 - HLA-C*03:13 consensus binding motif for 9-mers

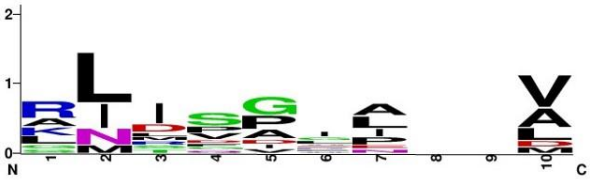

Figure 4.35 - HLA-A*02:01 obtained binding motif for 10-mers

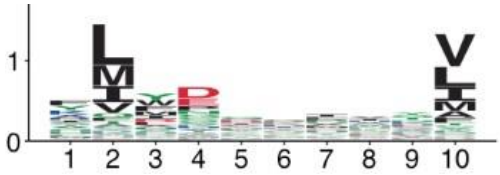

Figure 4.34 - HLA-A*02:01 consensus binding motif for 10-mers

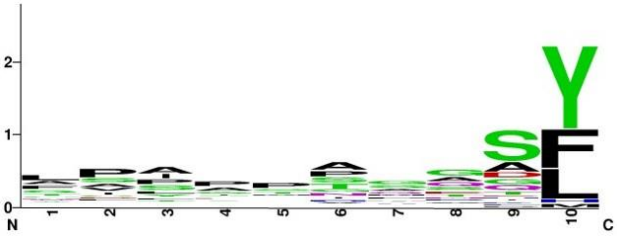

Figure 4.37 - HLA-B*15:11 obtained binding motif for 10-mers

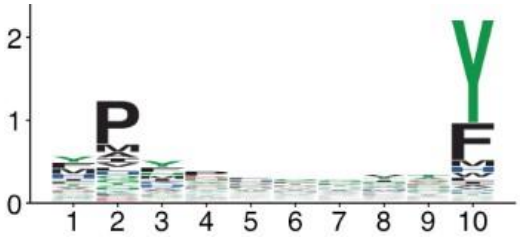

Figure 4.36 - HLA-B*15:11 consensus binding motif for 10-mers

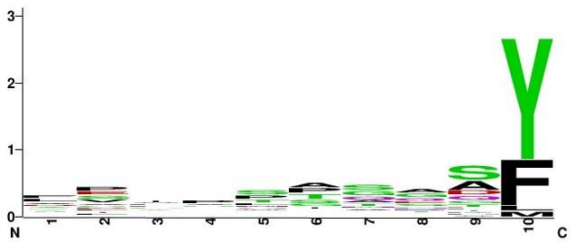

Figure 4.38 - HLA-B*15:15 obtained binding motif for 10-mers

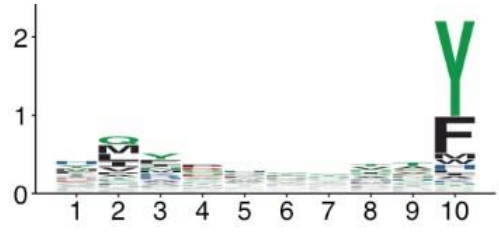

Figure 4.39 - HLA-B*15:15 consensus binding motif for 10-mers

Figure 4.40 - HLA-C*03:03 obtained binding motif for 10-mers

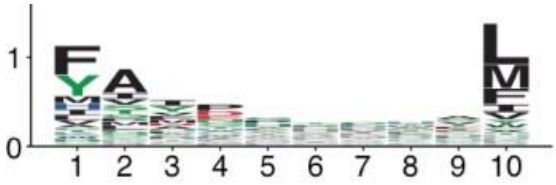

Figure 4.41 - HLA-C*03:03 consensus binding motif for 10-mers

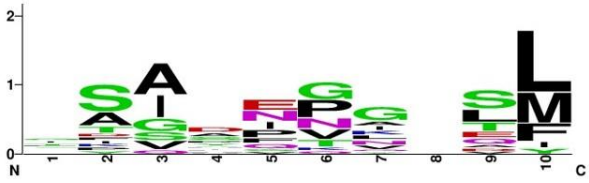

Figure 4.43 - HLA-C*03:13 obtained binding motif for 10-mers

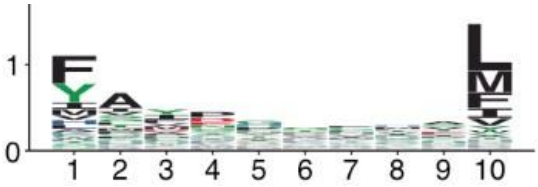

Figure 4.42 - HLA-C*03:13 consensus binding motif for 10-mers

On the right side are the consensus binding motifs for the five allotypes of 8-mers (**Figure 4.15**, **Figure 4.17, Figure 4.19, Figure 4.21** and **Figure 4.23**), 9-mers (**Figure 4.25**, **Figure 4.27, Figure 4.29, Figure 4.31** and **Figure 4.33**) and 10-mers (**Figure 4.35**, **Figure 4.37, Figure 4.39, Figure 4.41** and **Figure 4.43**) obtained from MHC Motif Atlas for comparison to the binding motifs obtained with WebLogo. There are a lot of similarities between the motifs, especially regarding the amino acid of the HLA pocket in position 2 and/or 3 and the amino acid in the C-terminal of the peptide.

5

CONCLUSIONS

W6/32 and L243 hybridoma cells were cultured to produce antibodies that bind to MHC class I and MHC class II, respectively. The antibody production suffered a scale-up after passing the cells from T-flasks to bioreactors. By comparing two different bioreactors and assessing the one with the highest yield, it was possible to optimize an important step in the workflow, the antibody production for use in the MHC immunoprecipitation protocol. With the findings observed, it was preferable to utilize the KDBio EZ-Flask bioreactor to produce higher amounts of antibodies. Nevertheless, for the production of highly concentrated antibody, the better option seems to be the CELLine CL1000 bioreactor.

DeNovoGUI CompOmics software runs several de novo sequencing algorithms and allowed the identification of 2913 unique MHC class I peptides from a sample with 5 x 10^8 THP-1 macrophage cells. From these identified peptides, 288 were predicted by NethMHCpan 4.1 as being strong binders for MHC class I molecules and 669 predicted as weak binders. Peptides with a NetMHCpan 4.1 rank below 0,5 were considered strong binders and ranked between 0,5 and 3 were considered weak binders. In combination, all predicted binders were used to analyse the binding motifs in the peptides identified for five different allotypes, HLA-A*02:01, HLA-B*15:1, HLA-A*15:15, HLA-A*03:03 and HLA-A*03:13. For 8-mers, 9-mers and 10-mers was possible to identify well-conserved amino acid residues in specific positions of the peptides such as the amino acid in the C-terminal of the peptide and the second amino acid for 8-mers and 9-mers and the third for 10-mers. Comparison showed that the binding motifs obtained are similar to the consensus binding motifs of those five allotypes.

With this work, it was possible to establish a functioning and viable immunopeptidomic platform, the first in Portugal. Starting with the growth of THP-1 macrophages, W6/32 and L243 hybridoma cells, then the purification of antibodies, MHC-peptide complexes immunoprecipitation, purification of peptides and finally the analysis of MS spectra obtained by mass spectrometry using a de novo sequencing

software. With more or less success in some of those steps, it was possible to carry out the immunopeptidomic workflow from start to finish and now it is possible to apply it in search of new findings.

REFERENCES

1. Flynn, J. A. L. & Chan, J. Immune cell interactions in tuberculosis. *Cell* **185**, 4682–4702 (2022).
2. Nunes-Alves, C. *et al.* In search of a new paradigm for protective immunity to TB. *Nat. Rev. Microbiol.* **12**, 289–299 (2014).
3. Kolattukudy, P. E., Fernandes, N. D., Azad, A. K., Fitzmaurice, A. M. & Sirakova, T. D. Biochemistry and molecular genetics of cell-wall lipid biosynthesis in mycobacteria. *Mol. Microbiol.* **24**, 263–270 (1997).
4. Dijl, V., Haven, N., Sweet, R. M., Mosimann, S. & Medical, B. C. Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. **396**, (1998).
5. Russell, D. G., Barry, C. E. & Flynn, J. L. Tuberculosis : What We Don ' t Know. *Science (80-.).* **328**, 852–856 (2010).
6. Chandra, P., Grigsby, S. J. & Philips, J. A. Immune evasion and provocation by Mycobacterium tuberculosis. *Nat. Rev. Microbiol.* **20**, 750–766 (2022).
7. Chandra, P., Grigsby, S. J. & Philips, J. A. Immune evasion and provocation by Mycobacterium tuberculosis. *Nat. Rev. Microbiol.* **20**, (2022).
8. Corleis, B. *et al.* Escape of Mycobacterium tuberculosis from oxidative killing by neutrophils. (2012) doi:10.1111/j.1462-5822.2012.01783.x.
9. Miller, J. L., Velmurugan, K., Cowan, M. J. & Briken, V. The Type I NADH Dehydrogenase of Mycobacterium tuberculosis Counters Phagosomal NOX2 Activity to Inhibit TNF- a -Mediated Host Cell Apoptosis. **6**, (2010).
10. Biology, H. *et al.* Mycobacterium tuberculosis Toxin CpnT Is an ESX-5 Substrate and Requires Three Type VII Secretion Systems for Intracellular. **12**, 1–16 (2021).
11. Cell, T., Lipid, W., Contributes, P., Escape, P. & Exit, H. C. The Cell Wall Lipid PDIM Contributes to Phagosomal Escape and Host Cell Exit of. **8**, 1–12 (2017).
12. Peyron, P. *et al.* Foamy Macrophages from Tuberculous Patients ' Granulomas Constitute a Nutrient-Rich Reservoir for M . tuberculosis Persistence. **4**, 1–14 (2008).
13. Zhang, L., Jiang, X., Pfau, D., Ling, Y. & Nathan, C. F. Type I interferon signaling mediates Mycobacterium tuberculosis – induced macrophage death. **218**, (2020).
14. Wolf, A. J. *et al.* Initiation of the adaptive immune response to Mycobacterium tuberculosis depends on antigen production in the local lymph node, not the lungs. *J. Exp. Med.* **205**, 105–

115 (2008).

15. Flynn, J. L., Chan, J. & Lin, P. L. Macrophages and control of granulomatous inflammation in tuberculosis. *Mucosal Immunol.* **4**, 271–278 (2011).
16. Barry, C. E. *et al.* The spectrum of latent tuberculosis: Rethinking the biology and intervention strategies. *Nat. Rev. Microbiol.* **7**, 845–855 (2009).
17. Khader, S. A. *et al.* Targeting innate immunity for tuberculosis vaccination. **129**, (2019).
18. Thibault, P. & Perreault, C. Immunopeptidomics: Reading the Immune Signal That Defines Self From Nonself. *Mol. Cell. Proteomics* **21**, 100234 (2022).
19. Goldberg, A. C. arl. & Rizzo, L. V. icent. MHC structure and function – antigen presentation. Part 1. *Einstein (Sao Paulo).* **13**, 153–156 (2015).
20. Ogretmen, B. Genetics and the causes of ankylosing spondylitis. *Physiol. Behav.* **176**, 139–148 (2019).
21. Rock, K. L., Reits, E. & Neefjes, J. Present Yourself! By MHC Class I and MHC Class II Molecules. **37**, 724–737 (2017).
22. Maenaka, K. & Jones, E. Y. MHC superfamily structure and the immune system. *Curr. Opin. Struct. Biol.* **9**, 745–753 (1999).
23. Potter, E. &. Origin and evolution of the adaptive immune system: genetic events and selective pressures. *Bone* **23**, 1–7 (2008).
24. Mukrimaa, S. S. *et al. Molecular Biology of the Cell 5th*. *Jurnal Penelitian Pendidikan Guru Sekolah Dasar* vol. 6 (2016).
25. Brown, J. H. *et al.* Three-dimensional structure of the human class II histocompatibility antigen HLA-DR1. *Nature* **364**, 33–39 (1993).
26. Bjorkman PJ, Saper MA, Samraoui B, Bennett WS, Strominger JL, Wiley, D. The foreign antigen binding site and T cell recognition regions. *Nature* **329**, 512–518 (1987).
27. Ezekowitz RA, Williams DJ, Koziel H, Armstrong MY, Warner A, Richards FF, R. R. Allele-specific motifs revealed by sequencing of self-peptides eluted from MHC molecules. *Nature* file:///D:/Thesis Project/Project/literature/most (1991).
28. Smith, K. J. *et al.* Bound water structure and polymorphic amino acids act together to allow the binding of different peptides to MHC class I HLA-B53. *Immunity* **4**, 215–228 (1996).
29. Madden, D. R., Garboczi, D. N. & Wiley, D. C. The antigenic identity of peptide-MHC complexes: A comparison of the conformations of five viral peptides presented by HLA-A2. *Cell* **75**, 693–708 (1993).
30. Walz, S. *et al.* The antigenic landscape of multiple myeloma: Mass spectrometry (re)defines targets for T-cell-based immunotherapy. *Blood* **126**, 1203–1213 (2015).
31. Rock, K. L. *et al.* Inhibitors of the proteasome block the degradation of most cell proteins and

the generation of peptides presented on MHC class I molecules. *Cell* **78**, 761–771 (1994).

32. Reits, E. *et al.* Peptide Diffusion, Protection, and Degradation in Nuclear and Cytoplasmic Compartments before Antigen Presentation by MHC Class I. *Immunity* **18**, 97–108 (2003).
33. Neefjes, J., Jongsma, M. L. M., Paul, P. & Bakke, O. Towards a systems understanding of MHC class i and MHC class II antigen presentation. *Nat. Rev. Immunol.* **11**, 823–836 (2011).
34. Cresswell, P., Bangia, N., Dick, T. & Diedrich, G. The nature of the MHC class I peptide loading complex. *Immunol. Rev.* **172**, 21–28 (1999).
35. Wearsch, P. A. & Cresswell, P. Selective loading of high-affinity peptides onto major histocompatibility complex class I molecules by the tapasin-ERp57 heterodimer. *Nat. Immunol.* **8**, 873–881 (2007).
36. Saveanu, L. *et al.* Concerted peptide trimming by human ERAP1 and ERAP2 aminopeptidase complexes in the endoplasmic reticulum. *Nat. Immunol.* **6**, 689–697 (2005).
37. Hermann, C. *et al.* TAPBPR alters MHC class I peptide presentation by functioning as a peptide exchange catalyst. *Elife* **4**, 1–22 (2015).
38. York, I. A. *et al.* The Er aminopeptidase ERAP I enhances or limits antigen presentation by trimming epitopes to 8-9 residues. *Nat. Immunol.* **3**, 1177–1184 (2002).
39. Ezekowitz RA, Williams DJ, Koziel H, Armstrong MY, Warner A, Richards FF, R. R. Peptide-induced conformational change of the class I heavy chain. *Nature* file:///D:/Thesis Project/Project/literature/most (1991).
40. Schumacher, T. N. M. *et al.* Direct binding of peptide to empty MHC class I molecules on intact cells and in vitro. *Cell* **62**, 563–567 (1990).
41. Princiotta, M. F. *et al.* Quantitating protein synthesis, degradation, and endogenous antigen processing. *Immunity* **18**, 343–354 (2003).
42. Yewdell, J. W. DRiPs solidify: Progress in understanding endogenous MHC class I antigen processing. *Trends Immunol.* **32**, 548–558 (2011).
43. Monaco, J. J. & Nandi, D. The genetics of proteasomes and antigen processing. *Annu. Rev. Genet.* **29**, 729–754 (1995).
44. Toes, R. E. M. *et al.* Discrete cleavage motifs of constitutive and immunoproteasomes revealed by quantitative analysis of cleavage products. *J. Exp. Med.* **194**, 1–12 (2001).
45. Kincaid, E. Z., Murata, S., Tanaka, K. & Rock, K. L. Specialized proteasome subunits have an essential role in the thymic selection of CD8+ T cells. *Nat. Immunol.* **17**, 938–945 (2016).
46. Yewdell, J. W., Reits, E. & Neefjes, J. Making sense of mass destruction: Quantitating MHC class I antigen presentation. *Nat. Rev. Immunol.* **3**, 952–961 (2003).
47. Meyer, V. S. *et al.* Identification of natural MHC class II presented phosphopeptides and tumor-derived MHC class I phospholigands. *J. Proteome Res.* **8**, 3666–3674 (2009).

48. Haurum, J. S. *et al.* Presentation of cytosolic glycosylated peptides by human class I major histocompatibility complex molecules in vivo. *J. Exp. Med.* **190**, 145–150 (1999).
49. Petersen, J., Purcell, A. W. & Rossjohn, J. Post-translationally modified T cell epitopes: Immune recognition and immunotherapy. *J. Mol. Med.* **87**, 1045–1051 (2009).
50. Zarling, A. L. *et al.* Phosphorylated peptides are naturally processed and presented by major histocompatibility complex class I molecules in vivo. *J. Exp. Med.* **192**, 1755–1762 (2000).
51. Goldberg, A. C. arl. & Rizzo, L. V. icent. MHC structure and function - antigen presentation. Part 2. *Einstein (Sao Paulo).* **13**, 157–162 (2015).
52. Berkers, C. R. *et al.* Definition of Proteasomal Peptide Splicing Rules for High-Efficiency Spliced Peptide Presentation by MHC Class I Molecules. *J. Immunol.* **195**, 4085–4095 (2015).
53. Ziegler, H., Muranyi, W., Burgert, H. G., Kremmer, E. & Koszinowski, U. H. The luminal part of the murine cytomegalovirus glycoprotein gp40 catalyzes the retention of MHC class I molecules. *EMBO J.* **19**, 870–881 (2000).
54. Mittal, D., Gubin, M. M., Schreiber, R. D. & Smyth, M. J. New insights into cancer immunoediting and its three component phases-elimination, equilibrium and escape. *Curr. Opin. Immunol.* **27**, 16–25 (2014).
55. Garcia-Lora, A., Algarra, I. & Garrido, F. MHC class I antigens, immune surveillance, and tumor immune escape. *J. Cell. Physiol.* **195**, 346–355 (2003).
56. Rock, K. L., Gamble, S. & Rothstein, L. Presentation of exogenous antigen with class I major histocompatibility complex molecules. *Science (80-.).* **249**, 918–921 (1990).
57. Kurts, C., Robinson, B. W. S. & Knolle, P. A. Cross-priming in health and disease. *Nat. Rev. Immunol.* **10**, 403–414 (2010).
58. Guermonprez, P. *et al.* ER-phagosome fusion defines an MHC class I cross-presentation compartment in dendritic cells. *Nature* **425**, 397–402 (2003).
59. Kovacsovics-bankowski, A. M., Clark, K., Benacerraf, B. & Kenneth, L. Efficient major histocompatibility complex class I presentation of exogenous antigen upon phagocytosis by macrophages. (2014).
60. Amigorena, S. Fc Receptors and Cross-Presentation in Dendritic Cells. **195**, 1–3 (2002).
61. Schuette, V. & Burgdorf, S. The ins-and-outs of endosomal antigens for cross-presentation. *Curr. Opin. Immunol.* **26**, 63–68 (2014).
62. Gromme, M. *et al.* Recycling MHC class I molecules and endosomal peptide loading. **96**, 10326–10331 (2014).
63. Shen, L., Sigal, L. J., Boes, M. & Rock, K. L. Important Role of Cathepsin S in Generating Peptides for TAP-Independent MHC Class I Crosspresentation In Vivo. **21**, 155–165 (2004).
64. Unanue, E. R., Turk, V. & Neefjes, J. Variations in MHC Class II Antigen Processing and

Presentation in Health and Disease. 1–33 (2016) doi:10.1146/annurev-immunol-041015-055420.

65. Muhlethaler-mottet, A. Lessons from the bare lymphocyte syndrome : molecular mechanisms regulating MHC class II expression. 148–165.
66. Pos, W. *et al.* Crystal Structure of the HLA-DM – HLA-DR1 Complex Defines Mechanisms for Rapid Peptide Selection. *Cell* **151**, 1557–1568 (2012).
67. Suri, A., Lovitch, S. B. & Unanue, E. R. The wide diversity and complexity of peptides bound to class II MHC molecules. 70–77 (2006) doi:10.1016/j.coi.2005.11.002.
68. Dongre, A. R. *et al.* In vivo MHC class II presentation of cytosolic proteins revealed by rapid automated tandem mass spectrometry and functional analyses. *Eur. J. Immunol.* **31**, 1485–1494 (2001).
69. Harding, C. V. & Boom, W. H. Regulation of antigen presentation by Mycobacterium tuberculosis: A role for Toll-like receptors. *Nat. Rev. Microbiol.* **8**, 296–307 (2010).
70. Cresswell, P. & Roche, P. A. Invariant chain-MHC class II complexes: Always odd and never invariant. *Immunol. Cell Biol.* **92**, 471–472 (2014).
71. Neefjes, J. CIIV, MIIC and other compartments for MHC class II loading. *Eur. J. Immunol.* **29**, 1421–1425 (1999).
72. The structure of an intermediate in class II MHC maturation: CLIP bound to HLA-DR3.
73. Denzin, L. K. & Cresswell, P. HLA-DM Induces CLIP Dissociation from MHC Class II ap Dimers and Facilitates Peptide Loading. **82**, 155–165 (1995).
74. Muntasell, A. *et al.* Dissection of the HLA-DR4 Peptide Repertoire in Endocrine Epithelial Cells: Strong Influence of Invariant Chain and HLA-DM Expression on the Nature of Ligands. *J. Immunol.* **173**, 1085–1093 (2004).
75. Wubbolts, R. *et al.* Direct Vesicular Transport of MHC Class II Molecules from Lysosomal Structures to the Cell Surface. **135**, 611–622 (2014).
76. Nielsen, M., Justesen, S., Lund, O., Lundegaard, C. & Buus, S. NetMHCIIpan-2.0 - Improved pan-specific HLA-DR predictions using a novel concurrent alignment and weight optimization training procedure. *Immunome Res.* **6**, 9 (2010).
77. Fernando, M. M. A. *et al.* Defining the role of the MHC in autoimmunity: A review and pooled analysis. *PLoS Genet.* **4**, (2008).
78. Sinnathamby, G. & Eisenlohr, L. C. Presentation by Recycling MHC Class II Molecules of an Influenza Hemagglutinin-Derived Epitope That Is Revealed in the Early Endosome by Acidification. *J. Immunol.* **170**, 3504–3513 (2003).
79. Wan, X., Thomas, J. W. & Unanue, E. R. Class-switched anti-insulin antibodies originate from unconventional antigen presentation in multiple lymphoid sites. 1–12 (2016)

doi:10.1084/jem.20151869.

80. Liljedahl, M., Kuwana, T., Jackson, M. R., Peterson, P. A. & Karlsson, L. HLA-DO is a lysosomal resident which requires association with HLA-DM for efficient intracellular transport. **15**, 4817–4824 (1996).
81. Cella, M., Engering, A. & Pinet, V. Inflammatory stimuli induce accumulation of MHC class II complexes on dendritic cells. **388**, 782–787 (1997).
82. Thibodeau, J. *et al.* Interleukin-10-induced MARCH1 mediates intracellular sequestration of MHC class II in monocytes. 1225–1230 (2008) doi:10.1002/eji.200737902.
83. Zwart, W. *et al.* Spatial Separation of HLA-DM / HLA-DR Interactions within MIIC and Phagosome-Induced Immune Escape. **22**, 221–233 (2005).
84. Niel, G. Van *et al.* Dendritic Cells Regulate Exposure of MHC Class II at Their Plasma Membrane by Oligoubiquitination. 885–894 (2006) doi:10.1016/j.immuni.2006.11.001.
85. Shapiro, I. E. & Bassani-sternberg, M. The impact of immunopeptidomics : From basic research to clinical implementation. *Semin. Immunol.* **66**, 101727 (2023).
86. Purcell, A. W., Ramarathinam, S. H. & Ternette, N. Mass spectrometry–based identification of MHC-bound peptides for immunopeptidomics. *Nat. Protoc.* **14**, 1687–1707 (2019).
87. Aebersold, R. & Mann, M. Mass-spectrometric exploration of proteome structure and function. *Nature* **537**, 347–355 (2016).
88. Caron, E. *et al.* Analysis of major histocompatibility complex (MHC) immunopeptidomes using mass spectrometry. *Mol. Cell. Proteomics* **14**, 3105–3117 (2015).
89. Mommen, G. P. M. *et al.* Expanding the detectable HLA peptide repertoire using electron-transfer/higher-energy collision dissociation (EThcD). *Proc. Natl. Acad. Sci. U. S. A.* **111**, 4507–4512 (2014).
90. Aebersold, R. & Mann, M. Mass spectrometry-based proteomics. *Nature* **422**, 198–207 (2003).
91. Wilm, M. Principles of electrospray ionization. *Mol. Cell. Proteomics* **10**, 1–8 (2011).
92. Vertes, A. Mass spectrometry in proteomics. *Med. Appl. Mass Spectrom.* 173–194 (2007) doi:10.1016/B978-044451980-1.50010-0.
93. Hager, J. W. A new linear ion trap mass spectrometer. *Rapid Commun. Mass Spectrom.* **16**, 512–526 (2002).
94. Marshall, A. G., Hendrickson, C. L. & Jackson, G. S. Fourier transform ion cyclotron resonance mass spectrometry: A primer. *Mass Spectrom. Rev.* **17**, 1–35 (1998).
95. Gatlin, C. L., Eng, J. K., Cross, S. T., Detter, J. C. & Yates, J. R. Automated identification of amino acid sequence variations in proteins by HPLC/microspray tandem mass spectrometry. *Anal. Chem.* **72**, 757–763 (2000).
96. Liu, H., Sadygov, R. G. & Yates, J. R. A model for random sampling and estimation of relative

protein abundance in shotgun proteomics. *Anal. Chem.* **76**, 4193–4201 (2004).

97. Pak, H. *et al.* Clustering and filtering tandem mass spectra acquired in data-independent mode. *J. Am. Soc. Mass Spectrom.* **24**, 1862–1871 (2013).
98. Li, L. *et al.* Targeted data extraction of the MS/MS spectra generated by data independent acquisition: a new concept for consistent and accurate proteome analysis. *Image (Rochester, N.Y.)* 1–45 (2012).
99. Mann, M. & Wilm, M. Error-Tolerant Identification of Peptides in Sequence Databases by Peptide Sequence Tags. *Anal. Chem.* **66**, 4390–4399 (1994).
100. Yates, J. R., Eng, J. K., McCormack, A. L. & Schieltz, D. Method to Correlate Tandem Mass Spectra of Modified Peptides to Amino Acid Sequences in the Protein Database. *Anal. Chem.* **67**, 1426–1436 (1995).
101. Standing, K. G. Peptide and protein de novo sequencing by mass spectrometry. 595–601 (2003) doi:10.1016/j.sbi.2003.09.005.
102. Nelde, A., Rammensee, H. G. & Walz, J. S. The peptide vaccine of the future. *Mol. Cell. Proteomics* **20**, 100022 (2021).
103. Malonis, R. J., Lai, J. R. & Vergnolle, O. *Peptide-Based Vaccines: Current Progress and Future Challenges*. *Chemical Reviews* vol. 120 (2020).
104. Bijker, M. S. *et al.* Superior induction of anti-tumor CTL immunity by extended peptide vaccines involves prolonged, DC-focused antigen presentation. *Eur. J. Immunol.* **38**, 1033–1042 (2008).
105. Aichele, P., Brduscha-Riem, K., Zinkernagel, R. M., Hengartner, H. & Pircher, H. T cell priming versus T cell tolerance induced by synthetic peptides. *J. Exp. Med.* **182**, 261–266 (1995).
106. Southwood, S. *et al.* Several Common HLA-DR Types Share Largely Overlapping Peptide Binding Repertoires . *J. Immunol.* **160**, 3363–3373 (1998).
107. Grupp, S. A. *et al.* Chimeric Antigen Receptor–Modified T Cells for Acute Lymphoid Leukemia. *N. Engl. J. Med.* **368**, 1509–1518 (2013).
108. Kowalewski, D. J. *et al.* HLA ligandome analysis identifies the underlying specificities of spontaneous antileukemia immune responses in chronic lymphocytic leukemia (CLL). *Proc. Natl. Acad. Sci. U. S. A.* **112**, E116–E175 (2015).
109. Candia, M., Kratzer, B. & Pickl, W. F. On peptides and altered peptide ligands: From origin, mode of action and design to clinical application (Immunotherapy). *Int. Arch. Allergy Immunol.* **170**, 211–233 (2016).
110. Zom, G. G. *et al.* TLR2 ligand-synthetic long peptide conjugates effectively stimulate tumor-draining lymph node T cells of cervical cancer patients. *Oncotarget* **7**, 67087–67100 (2016).
111. Vermaelen, K. Vaccine strategies to improve anticancer cellular immune responses. *Front. Immunol.* **10**, 1–17 (2019).

112. Bettencourt, P. *et al.* Identification of antigens presented by MHC for vaccines against tuberculosis. *npj Vaccines* **5**, (2020).
113. Bettencourt, P. *et al.* Identification of antigens presented by MHC for vaccines against tuberculosis. *npj Vaccines* **5**, (2020).
114. Tadros, D. M., Eggenschwiler, S., Racle, J. & Gfeller, D. The MHC Motif Atlas: a database of MHC binding specificities and ligands. *Nucleic Acids Res.* **51**, D428–D437 (2023).

A

ANNEXES

Table A1 - Mean Fluorescense Intensity on FITC channel filter

	Mean Fluorescense Intensity
	FITC
Blank	580,58
BCG-GFP P0	123896,41
BCG-GFP P20	13785,01

Table A2 - Antibody concentration

	[Ab] (A280nm)	
Corning CL1000	**1st purification**	1,132 mg/mL
	2nd purification	1,212 mg/mL
KDBio EZ-Flask	**1st purification**	0,641 mg/ml
	2nd purification	0,315 mg/mL

Table A3 - Number of peptide sequences by length

Peptide length (aa)	Number of peptide sequences	Percentage
8	679	23
9	939	32
10	581	20
11	279	10
12	121	4
13	61	2
14	48	2
15	36	1
16	35	1
17	20	1
18	12	0
19	9	0
20	14	0
21	14	0
22	13	0
23	10	0
24	7	0
25	9	0
26	8	0
27	2	0
28	3	0
29	7	0
30	6	0
Total	2913	

Table A4 - Number of strong and weak binders for each allotype and separated by 8-mers, 9-mers and 10-mers

	8mers		9mers		10mers	
	Strong binders	Weak binders	Strong binders	Weak binders	Strong binders	Weak binders
HLA-A*02:01	1	4	46	56	2	10
HLA-B*15:11	3	23	50	105	12	25
HLA-B*15:15	2	22	47	82	11	20
HLA-C*03:03	6	37	47	113	4	17
HLA-C*03:13	7	34	46	105	4	16

Table A5 - Number of strong and weak binders by allotype

	Strong binders	Weak binders
HLA-A*02:01	49	70
HLA-B*15:11	65	153
HLA-B*15:15	60	124
HLA-C*03:03	57	167
HLA-C*03:13	57	155