The mechanism of tapasin-mediated peptide exchange of MHC Class I molecules

Inaugural-Dissertation

to obtain the academic degree

Doctor rerum naturalium (Dr.rer.nat.)

submitted to the Department of Biology, Chemistry, Pharmacy

of Freie Universität Berlin

by

Huan Lan

2021

The work presented in this thesis was conducted from October 2016 to December 2020 under the supervision of Prof. Dr. Christian Freund at the Institute of Chemistry and Biochemistry, Freie Universität Berlin, Germany.

First reviewer: Prof. Dr. Christian Freund
Freie Universität Berlin
Institute for Chemistry and Biochemistry
Thielallee 63, 14195 Berlin, Germany

Second reviewer: Prof. Dr. Markus Wahl
Freie Universität Berlin
Institute for Chemistry and Biochemistry
Takustr. 6, 14195 Berlin, Germany

Date of defense: August 17th, 2021

Affidavit

I declare that my doctoral thesis "**The mechanism of tapasin-mediated peptide exchange of MHC class I molecules**" has been written independently and with no other sources or aid than those cited.

Huan Lan

Content

1 Introduction

1.1 The immune system

Vertebrates have evolved sophisticated defense mechanisms to protect themselves against dangerous environmental invasions, such as virus, bacteria, fungi, and parasites[1,2]. These defense mechanisms rely on the immune system, which consists of different organs, tissues, cells and soluble mediators (Fig. 1.1). Based on the defense mechanism, the immune system is divided into two types: the innate immunity and the adaptive immunity[3]. Innate immunity is the first line of the host defense. The adaptive immunity arose 450 million years ago in vertebrates and has evolved well to work in concert with the innate immune system[2].

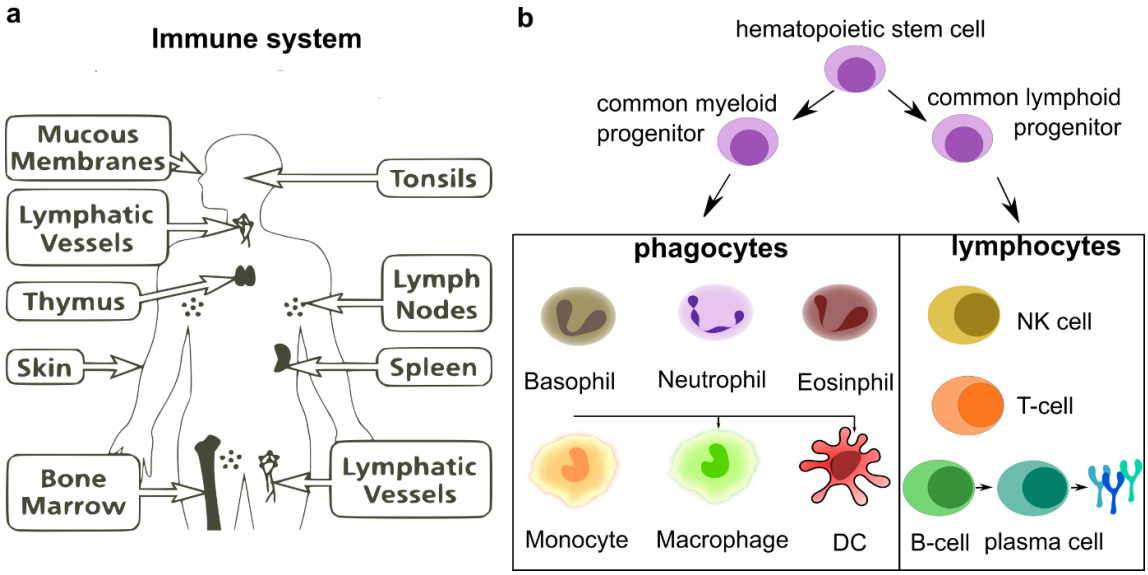


Figure 1.1 The human immune system. a) The organs and tissues of human immune system are indicated on the human body cartoon. This figure is modified from an online picture https://en.wikibooks.org/wiki/The_Organ_Systems/Immune. **b)** The maturation of white blood cells (leukocytes) and cell types involved in innate (left) and adaptive (right) immune response. Monocyte differentiates into macrophage and dendritic cells (DCs). Lymphocytes differentiate into natural killer cells (NK cells), T cells and B cells.

1.1.1 Innate immune system

The innate immune system is comprised of different components (Fig. 1.1), including anatomical barriers, white blood cells and the complement system. Anatomical barriers consist of physical (e.g., skin, mucous membranes of mouth, nasopharynx, gastrointestinal tract), chemical (e.g., mucus, saliva, lysozymes) and biological barriers. The innate immune responses are mediated by white blood cells (Fig. 1.2a), including granulocytes (e.g., eosinophils, mast cell, and basophils), phagocytes (e.g., neutrophils, monocytes, macrophages), dendritic cells (DCs) and natural killer cells (NK cells), as well as a wide range of inflammatory chemical factors and cytokines. The complement system is a "complement" of the immune system to improve the ability of antibodies to clear pathogens. It is made up of a large number of plasma proteins (e.g., C-reactive protein, mannose-binding lectin) that work together to opsonize pathogens and induce a series of inflammatory responses. The major functions of the vertebrate innate immune system are: (i) preventing infection, (ii) activating the complement cascade

to identify pathogens for destruction and clearance of antibody-complex or dead cells, (iii) recruiting white blood cells to the infected site by secreting chemical factors, (iv) eliminating pathogens by specialized white blood cells, and (v) stimulating the adaptive immune response through antigen presentation process.

The skin is impermeable to most infectious agents, acting as the first line of the innate immune system to keep pathogens at bay. Once pathogens escape the surveillance of the first line, innate immune responses will be activated. The complement system plays a major role in innate immune response and its activation involves a series of cleavage reactions that generate activated complement proteins (e.g. C3). The complement proteins can covalently bind to pathogens for engulfment by phagocytes, can act as chemoattractant that recruit other immune cells to the infected site, or can damage bacteria by creating pores in the bacterial membrane. The complement cascade can be initiated by three pathways: (i) molecules on the pathogen surface that can be recognized by C3b complement protein. (ii) mannose on pathogen surface recognized by MB-binding lectin. (iii) antigen-antibody complexes.

Inflammation is one of the first immune responses to infection. At the cellular level, white blood cells are localized at the infected area. Granulocytes, phagocytes, DCs and NK cells with other white blood cells work coordinately either by engulfing and eliminating pathogens, or secreting soluble immune-regulating mediators (e.g., pro-inflammatory cytokines and chemokines). Some of these chemical factors can prevent pathogen growth, but also cause damage to healthy cells. At the molecular level, white blood cells use the membrane-associated pattern recognition receptors (PRRs) expressed on their cell surface to recognize conserved pathogen-associated molecular patterns (PAMPs) on pathogens[4,5]. PAMPs are broadly shared by different pathogens and typically composed of sugar, nucleic acids or lipopolysaccharides, thereby allowing the innate immune cells to differentiate foreign pathogens from self-cells.

Macrophages are tissue resident phagocytes and bear Toll-like receptors (TLRs, one type of PRRs). Macrophages can secrete pro-inflammatory cytokines and chemokines to recruit the neutrophils to the infected region. They also secrete cytokines to enhance the adaptive immune response. DCs are responsible for the first recognition of the pathogens, followed by their phagocytosis, migration to the regional lymph node, priming of naïve T-cells and regulation of NK cells and B-cells' response. The destruction of pathogens and infected cells by DCs or macrophages lead to components of the pathogen presented to T cells, which further trigger the adaptive immune response (Fig. 1.2a). NK cells play an important role in the innate immune response. They recognize and eliminate infected or tumoral cells with their cell surface expressed killer-cell immunoglobulin-like receptors (KIRs)[6]. KIRs regulate the killing function of NK cells by interacting with their ligands, i.e., the major histocompatibility complex (MHC) class I molecules. NK cells initiate the immune response without previous activation by other cells, allowing for fast immune response.

1.1.2 Adaptive immune system

The adaptive immune response is mediated by lymphocyte cells (Fig. 1.1b). There are two major types of lymphocytes: B lymphocytes and T lymphocytes. Both B and T lymphocytes (T and B cells) are

derived from multipotent hematopoietic stem cells in the bone marrow. During their early developmental stages, different sets of genes encoding prototypic immunoglobulin (Ig) variable (V), diversity (D), and joining (J) regions within lymphocytes are rearranged to generate the antigen binding regions of B cell receptors (BCRs) and T cell receptors (TCRs). The random combination process results in the generation of a variety of specificities, expressed on individual cells derived from a clone carrying a distinct BCR or TCR. Cells expressing receptors that survive first the positive selection and then the negative selection become naïve mature B or T cells. Positive selection allows the selection of B cells expressing normal functional BCRs and T cells expressing receptors capable of binding self-peptide presented by MHC molecules with intermediate affinity and avidity. Cells expressing receptors that have not enough affinity or avidity to the self-peptide will be eliminated. In the negative selection, B cells and T cells expressing receptors that strongly recognize self-antigens may be clonally deleted in the bone marrow or thymus, respectively. The selected B and T cells then migrate into lymphoid tissues, where they may encounter invading pathogens, be primed and differentiate into effector cells. The immune response mediated by B cells and T cells is defined as humoral immunity and cell-mediated immunity, respectively. Both types of adaptive immunity are intimately connected, since B cells also serve as antigen-presenting-cells and both are engaged when the immune system is activated.

In humoral immune responses, BCRs expressed on the B cell surface recognize the exposed epitopes (antigens) of intact molecules, including surface proteins and carbohydrate moieties of invasive microbes. B cells can be activated in a T-cell dependent or independent manner. In the T-cell dependent mode, which is the predominant mode of activation, naïve B cells act as antigen presenting cells to present antigens to the helper T cells. After interaction with antigens either from free proteins or associated with an intact pathogen by BCRs, B cells extract antigens and then internalize them. Internalized antigens are processed and presented with MHC class II molecules on B-cell surface, which are then recognized by TCRs on the helper T cells. Activated helper T cells secrete cytokines that activate B cells and cause proliferation into clonal daughter cells (i.e., plasma cells) and differentiation into memory B cells (Fig. 1.2b). Plasma cells secrete different classes of antibodies, the soluble forms of the former BCR's, to neutralize or label the pathogens for future processing by the complement system or phagocytes. While memory B cells can generate more powerful immune response in future. In the T-cell independent mode, B cells are activated by the T-independent antigens (e.g., polysaccharide capsules, lipopolysaccharide). Once activated, B cells proliferate into plasma B cells that secrete lgM antibody. Unlike T-cell independent activation, no memory B cells are produced in the T-independent mode.

By contrast, TCRs expressed on the T cell surface recognize peptide fragments presented by the cell-surface of other MHC-expressing cells. T cells therefore typically recognize antigens that have been partially digested by the antigen-presenting cells, primarily DCs, phagocytic cells, and B lymphocytes. This recognition triggers T-cells to release cytokines which can help T cell activation (Fig. 1.2b). Activated T cells secrete cytokines to help T cell proliferation. A group of the T cells become helper T cells, which can secrete cytokines to attract more macrophages and neutrophils to the infected area. Another group of T-cells differentiate to cytotoxic T cells that can eliminate the infected cells, while

other T cells become memory T cells, which can last for many years and respond more rapidly when the same pathogen invades in further. According to the cell surface markers, T cells are divided into two types: $CD4^+$ and $CD8^+$ T cells (CD, cluster of differentiation). Helper $CD4^+$ T cells recognize the peptide-bound-MHC-Class II (pMHC-II) on the cell surface of antigen presenting cells (APCs) and regulate the local response as described in humoral immune response. While cytotoxic $CD8^+$ T cells recognize peptide-bound-MHC-Class I (pMHC-I) expressed on all nucleated cells and once are activated, induce death of the infected cells.

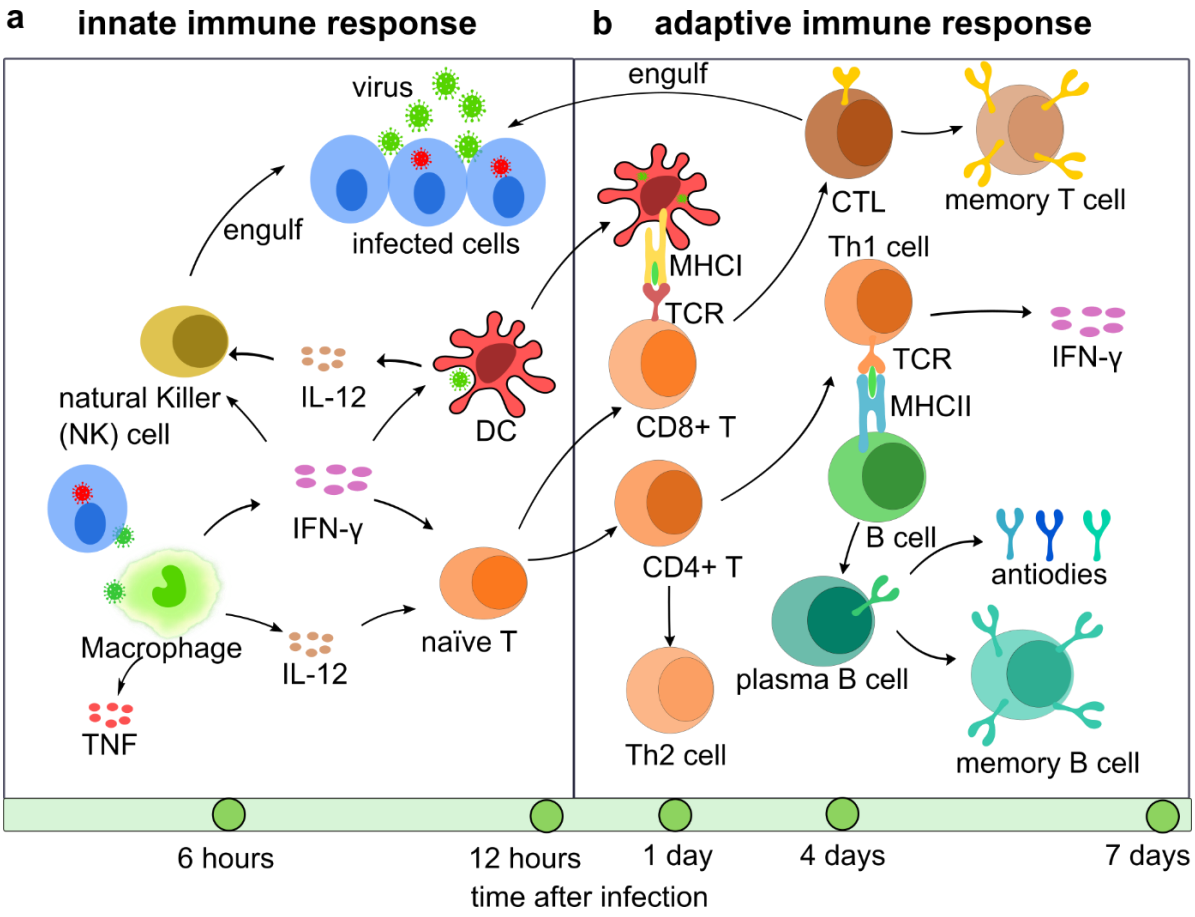


Figure 1.2 Comparison of innate immune response and adaptive immune response. a) Innate immune response: the major scavenger cells are macrophages, DCs and NK cells. **b)** Adaptive immune response: the proliferation and differentiation of T and B cells. Lower panel shows the time point of the response after infection. CTL stands for Cytotoxic T lymphocyte.

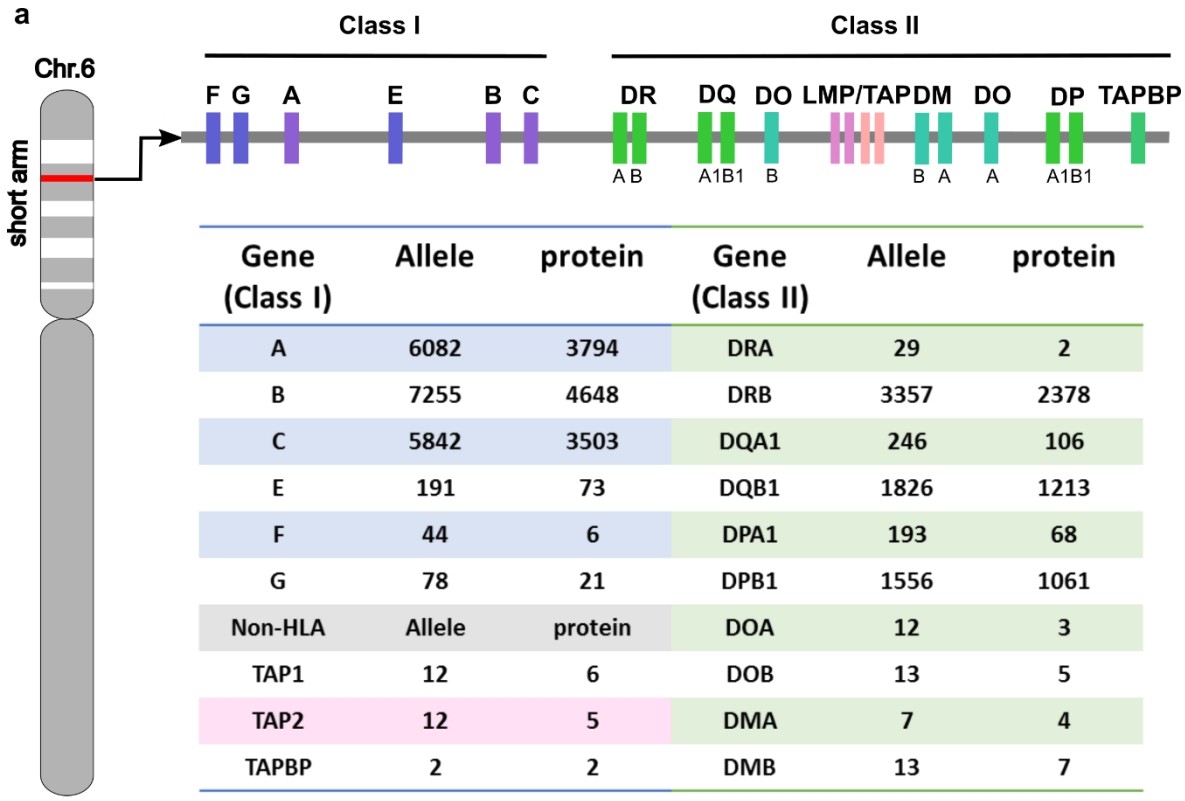
Unlike the innate immune response, the adaptive immune response is slow, but highly specific. The pathogens are variable and B cells and T cells have evolved a strategy to produce a matching diversity via their BCRs and TCRs. Antigen-presenting cells, which display fragments of pathogenic proteins on their cell surface, contribute to this variability though MHC proteins.

1.2 Major Histocompatibility Complex (MHC) molecules

1.2.1 Major Histocompatibility Complex (MHC)

The adaptive immune response relies on the recognition of antigens presented via the Major Histocompatibility Complex (MHC) proteins on the cell surface by TCRs. The human MHC proteins are encoded by the gene locus called human leukocyte antigen (HLA) which is located on the short arm

of chromosome 6 (6p21/22) (Fig. 1.3a)[7,8]. Most genes in the HLA locus encode immune-related proteins, covering all MHC class I (MHC-I), MHC class II (MHC-II), MHC class III (genes involved in inflammation), MHC class I-like and proteins involved in the antigen processing and cytokines[9]. Among these genes, MHC-I and MHC-II are the core components, which are also the most gene-dense and polymorphic regions in the human genome[10]. The nomenclature of HLA/MHC alleles, in brief, is that the first two digits of an allele describe its serological antigen (called an allotype) while the third and fourth pair of digits are used to represent the allele subtypes based on their sequences. More details of HLA/MHC nomenclature are found at the website http://hla.alleles.org/nomenclature/naming.html.



Gene (Class I)	Allele	protein	Gene (Class II)	Allele	protein
A	6082	3794	DRA	29	2
B	7255	4648	DRB	3357	2378
C	5842	3503	DQA1	246	106
E	191	73	DQB1	1826	1213
F	44	6	DPA1	193	68
G	78	21	DPB1	1556	1061
Non-HLA	Allele	protein	DOA	12	3
TAP1	12	6	DOB	13	5
TAP2	12	5	DMA	7	4
TAPBP	2	2	DMB	13	7

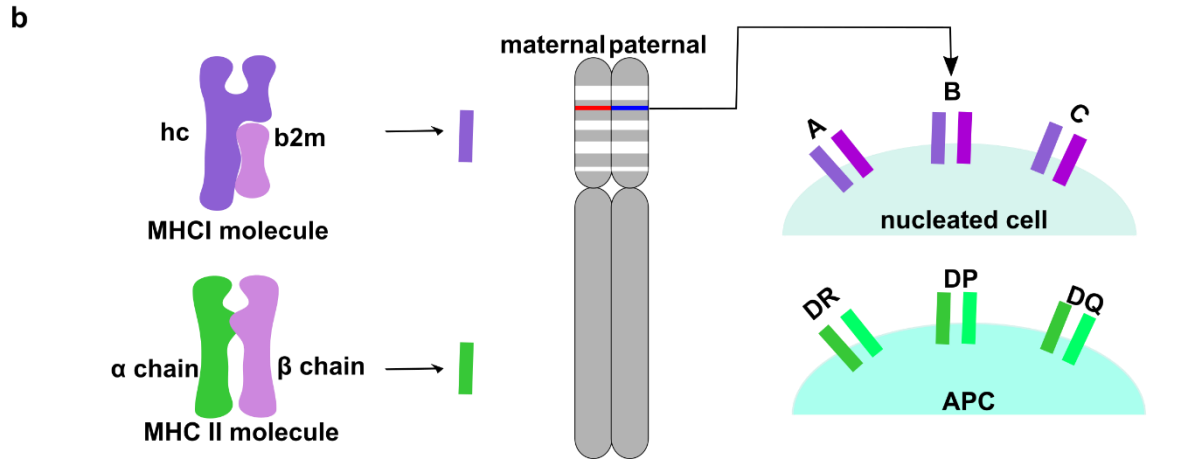


Figure 1.3 Polymorphism of MHC. a) Left: Carton representation of human chromosome 6 with red box highlighting the human leukocyte antigen (HLA) region. Upper: simplified organization map of MHC class I and MHC class II genes, DPA2/DPB2, DQA2/DQB2 is not shown because of its low polymorphism. TAP, LMP and TAPBP are antigen-processing related genes. TAPBP is located to the extended MHC-II region. Middle: Table of allele numbers of MHC-I and MHC-II, data source[11]: https://www.ebi.ac.uk/ipd/imgt/hla/stats.html (data updated Oct, 2020)**. b)** Left: Cartoon representation of MHC-I

and MHC-II molecules. Right: chromosome 6 of an individual comes from a maternal and a paternal chromatid, giving rise to up 6 classical MHC-I molecules and 6 classical MHC-II molecules expressed on one APC (for MHC-I, all nucleated cells). MHC protein expressed on the cell surface is simplified as a bar (indicated as in the left figure, MHC-I is the purple bar and MHC-II is the green bar).

The heavy chain of MHC-I molecules is encoded by highly polymorphic genes in the MHC-I loci. There are three classical MHC-I genes in the MHC-I loci: namely HLA-A, HLA-B and HLA-C. Besides classical MHC-I molecules, the MHC-I loci region also encodes non-classical MHC-I molecules, such as HLA-F, -G and E, but their expression abundance on the cell surface is low and shows limited polymorphism when compared to classical MHC-I molecules (Fig.1.3a)[10,12].

MHC-II molecules consist of a α chain and β chain. Both chains are encoded by genes in the MHC-II loci[8]. HLA-DR, HLA-DQ and HLA-DP genes encode the classical MHC-II molecules, and their polymorphisms are even higher than MHC-I molecules due to the polymorphism of both chains, which provides more potential combinations by chain pairing (Fig. 1.3a). For example, there are two proteins for HLA-DR α chain, meanwhile there are more than 2300 proteins for HLA-DR β chain (numbers are cited from www.ebi.ac.uk), which could form a combination of HLA-DR proteins theoretically reaching more than 2*2300. MHC-II-like genes on MHC-II region encode non-classical MHC-II molecules, including HLA-DM and HLA-DO. These genes show limited polymorphism. In addition to the MHC-II molecules, gene located to the MHC-II locus also encode proteins that are involved in the MHC-I antigen processing and presenting pathway (Fig. 1.3a).

Two features of HLA/MHC loci are the intensive degree of linkage disequilibrium (LD) and the high degree of polymorphism[13]. LD describes the phenomenon that the frequency of genes at neighboring ("frozen blocks") is higher than the frequency of other alleles of the same genes in population. For example, A*01:01~C*07:01~B*08:01~DRB3*01:01~DRB1*03:01~DQB1*02:01 is a very common haplotype (haplotype: combination of different MHC-I and MHC-II alleles on one chromosome) in the European population[14]. The polymorphism in MHC-I and MHC-II genes can be explained by the host-pathogen co-evolution[15]. Since MHC alleles of an individual come from a maternal haplotype and a paternal haplotype, every individual theoretically can express 6 allotypes of classical MHC-I molecules on the surface of nucleated cells and more than 6 allotypes of classical MHC-II molecules on antigen presenting cells (Fig. 1.3b). The MHC system provides the individual with a wide spectrum protection from potential infectious dangers. However, it could cause barriers to organ/cell transplantation in clinical practice. Typing MHC alleles of an individual is crucial in transplantation medicine, especially the HLA-A, HLA-B and HLA-DR types need to be matched, because mismatching of HLA alleles between a donor and a recipient would result in graft rejection and acute graft-versus-host disease (GVHD)[16,17].

1.2.2 MHC-I molecules

MHC-I molecules consists of three units: a heavy chain (HC), encoded by a gene in the MHC-I locus, a non-polymorphic light invariant chain which is encoded by the beta-2 microglobulin (β2m) gene on chromosome 15 (15q21.1)[18] and a peptide of 8-10 amino acids (aa). From a structural point of view, MHC-I molecules share typical structural features/architecture[19–21] (Fig. 1.4a): The α1 and α2 domains

of the heavy chain fold to form the binding groove, consisting of eight β strands at the bottom and flanked by two helices from the α1 and α2 domains; The α3 domain of the heavy chain interacts with β2m to form the lower scaffold, followed by the transmembrane domain and cytoplasmic tail. Peptides are bound in the binding groove typically in an extended conformation. Due to the narrow ends of the binding cleft, peptides that bind into the binding groove are normally 8-10 amino acids in length, with some cases up to 13 aa[22,23]. There are two types of interactions that determine the affinity of the peptide to MHC-I molecules: 1) H-bonds between side chains of residues in the heavy chain and the backbone of the bound peptide including the amide group of P1 (the first residue of the peptide) and carboxyl group of PΩ (Fig. 1.4b, the upper panel. PΩ: the last residue of the peptide). 2) The interactions between the side chain of the bound peptide (anchor residues) and residues of the pockets in the binding groove (Fig.1.4b, the lower panel). The first type of interactions is highly conserved: The amide group of the P1 forms H-bonds both with residue Y7 and Y171 of the heavy chain; The carboxyl group of the PΩ forms H-bonds with residue Y84, T143 and K146 of the heavy chain; The backbone carboxyl of the P1 forms an H-bond with residue Y159; The backbone carboxyl group of P8 forms an H-bond with residue W147; The backbone amide-group of PΩ forms an H-bond with residue D77 on the α1-helix[24]. In contrast, the second interaction is dependent ON the MHC-I allele and the peptide sequence[25]. For example, in the case of HLA-B*27:09/RL9 structure (PDB ID 1OF2)[26], residue Q155 forms H-bonds with the side chain of P5 and P7, and residue E163 forms an H-bond with the side chain of P1 (Lys here). The peptide binding groove can be divided into 6 pockets that accommodate the side chains of the bound peptide (Fig. 1.4c)[27–29]. Amongst these six pockets, the F-pocket was studied extensively as it has been shown to modulate the stability of MHC-I molecules and their chaperon dependence[30,31]. For most human MHC-I allotypes, the anchor residues are P2 and PΩ (Fig. 1.4d), while some allotypes also have auxiliary anchor (normally it is P3, with rare case is P8). For few allotypes, the anchor residues are P3, P5 and PΩ, such as HLA-B8 (Fig. 1.4d), which is similar to one murine MHC-I allotype (e.g., H2-D^b). Since the nature of the pocket of a certain MCH-I allotype is fixed, the bound peptide to it has conserved anchor residues[25,27]. Thus, the peptide motif of a certain MHC-I allotype can be predicted[32,33]. Side chains of residues that point out of the binding groove usually do not affect the affinity of the peptide with MHC-I but affect the recognition of pMHC-I by TCRs.

Classical MHC-I molecules are expressed in all nucleated cells and present protease-digested peptides to TCRs on $CD8^+$ T cells to trigger adaptive immune responses. There are three types of classic MHC-I molecules: HLA-A, HLA-B and HLA-C. The most comprehensive characterization of MHC-I molecules has been performed with HLA-A and HLA-B allotypes, as their cell surface abundance is higher than that of HLA-C[34]. MHC-I molecules not only mediate adaptive immune responses, but also play a role in innate immune responses as a ligand of inhibitory KIRs[35,36]. HLA-C bears a conserved sequence in the α1 helix, which acts as the binding/recognition site by KIRs[35,37]. Since all nucleated cell express MHC-I molecules, there are inhibitory and activating KIRs expressed by NK cells to maintain a balance between self-tolerance and the acquisition of NK cell functional competence[38]. Additionally, HLA-C plays a role in the implanting placenta, where HLA-C is the only classical MHC-I expressed, particularly on the cell surface of placental extravillous trophoblasts[37]. The non-classical HLA-G also plays role in pregnancy in maternal immune responses[39]. HLA-E is the ligand of the natural-killer group

2, member D (NKG2D) receptors and the binding groove of it is suitable for peptides derived from the leading sequence of other MHC-I molecules[40]. HLA-F is mainly expressed in the cellular content and barely travels to the cell surface[41]. The MHC-I chain-related protein A and B (i.e., MICA, MICB) share similar structure as classical MHC-I molecules but do not bind peptides. MICA and MICB are expressed in stress-induced cells and ligands of NKG2D receptors[42–44].

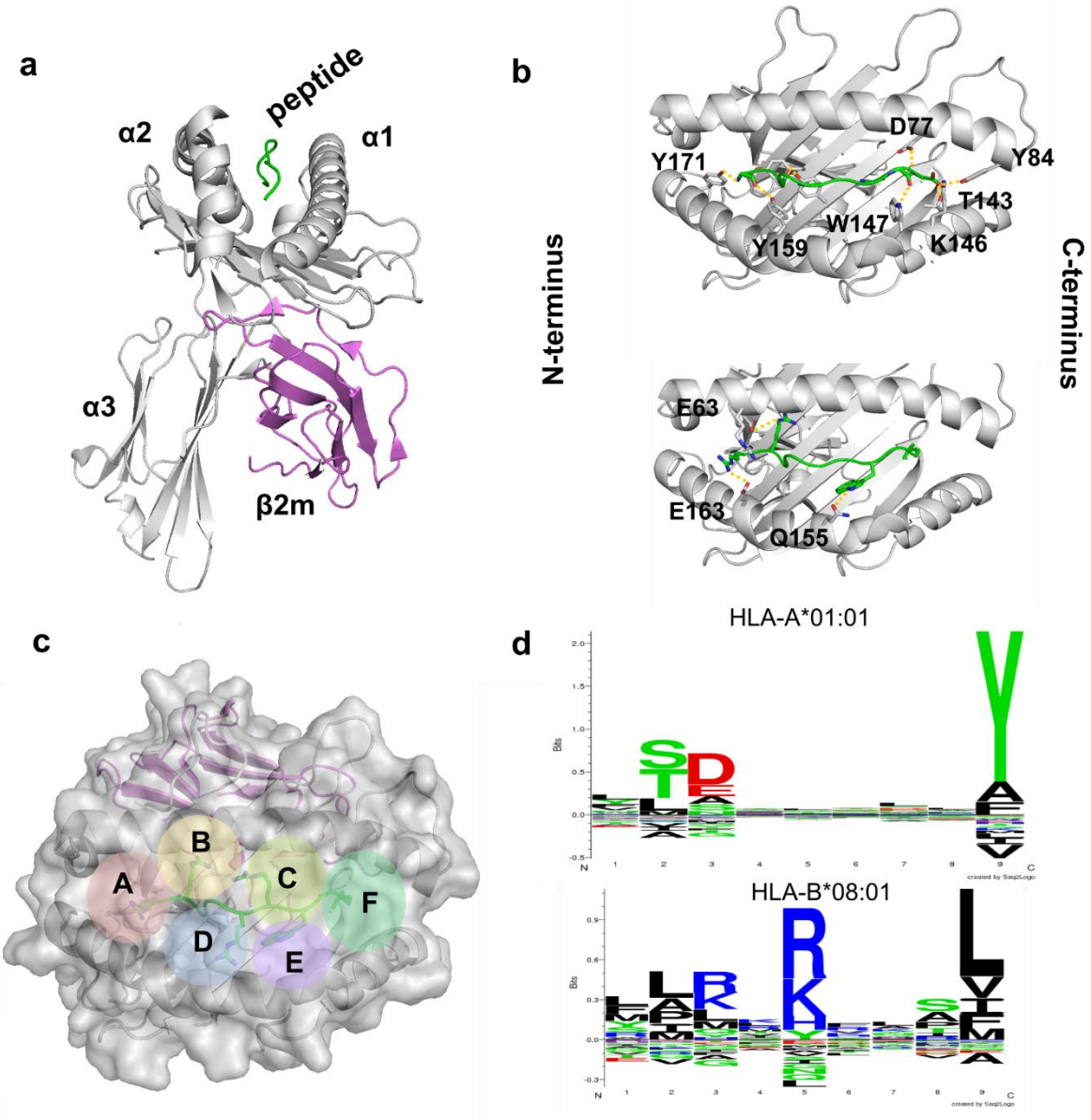
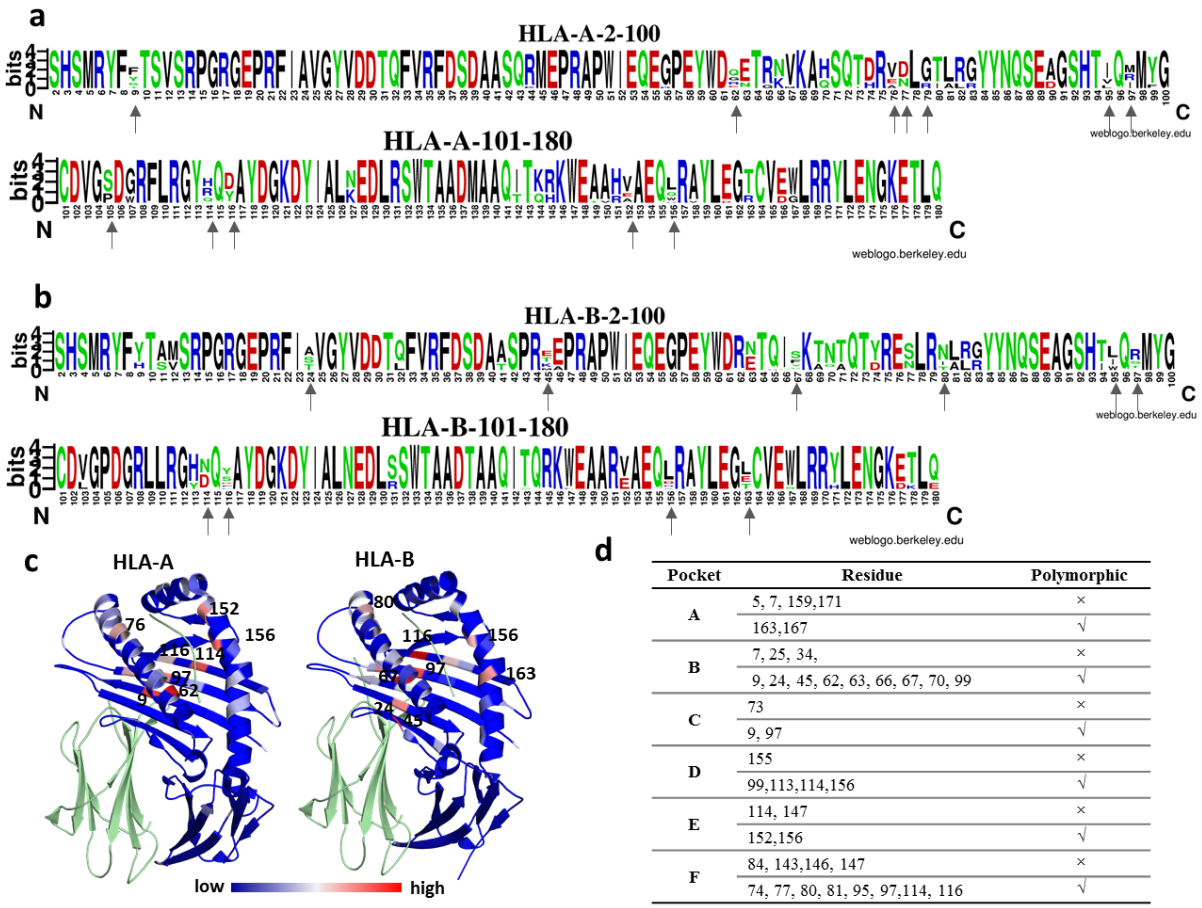


Figure 1.4 Structure of MHC-I molecules. a) Typical structure of a MHC-I molecule (PDB ID 1OF2) with peptide bound in the binding groove. **b)** Upper: H-bonds between side chain of residues in the heavy chain and the backbone of the bound peptide including the amide of P1 and C-terminal carboxy of PΩ (PΩ means the last amino acid of the peptide). Lower: H-bonds between side chains of the bound peptide and side chain of the heavy chain. **c)** Transparent surface representation of MHC-I molecule with highlighting A-F-pockets in colored circles. **d)** Examples of peptide sequence motif of A*01:01 and B*08:01, to show the preference of anchor residues.

1.2.3 Polymorphism of MHC-I molecules

As mentioned above, MHC-I molecules are composed of a highly polymorphic heavy chain and an invariant light chain β2m. There are more than 10,000 MHC-I molecule variants until now (data source: www.ebi.ac.uk), among them HLA-B is the most variable. Amino acid differences between different

allotypes are heavily concentrated in the binding groove that interacts with peptides (Fig. 1.5a, b and c), which results in different peptide repertoires for alleles[25]. For HLA-A allotypes, the most variable residues are 9, 62, 76, 95, 97, 114, 116, 152 and 156, which are mainly located in the B- and F-pocket. For HLA-B allotypes, the most variable residues are 24, 45, 67, 80, 75, 97, 116, 156 and 163, which are also mainly located in the B- and F-pocket. These variations are consistent with specific anchor residues (P2 and PΩ) for each allotype.



Pocket	Residue	Polymorphic
A	5, 7, 159,171	×
	163,167	√
B	7, 25, 34,	×
	9, 24, 45, 62, 63, 66, 67, 70, 99	√
C	73	×
	9, 97	√
D	155	×
	99,113,114,156	√
E	114, 147	×
	152,156	√
F	84, 143,146, 147	×
	74, 77, 80, 81, 95, 97,114, 116	√

Figure 1.5 Polymorphism of HLA-A and HLA-B molecules. a) Sequence logo of HLA-A binding groove (residue 2-180, n=5871) and **b)** HLA-B (residue 2-180, n=7180). Arrows indicates the strong polymorphic sites. Polar, acidic, basic and hydrophobic residues are colored in green, red, blue and black, respectively. Sequence data used for plotting were downloaded from IPD and alleles labelled with N* were left out. **c)** Entropic scores are mapped on the structure of HLA-A and HLA-B. High entropic values indicate high randomization of this position. Highly polymorphic sites are labelled with residue's number. This figure is modified from Dr. Esam T. Abualrous's review (still in preparation). **d)** Residues composing A-F-pockets and their polymorphism. × means conserved, √ means polymorphic.

1.2.4 MHC-II molecules

The MHC-II molecule consists of two polypeptides: α chain and β chain, both encoded by genes in the MHC-II locus, and a peptide. Both chains have two domains (α chain: α1 and α2 domain, β chain: β1 and β2 domain) and contain a C-terminal transmembrane domain followed by a short cytosolic tail. Unlike MHC-I molecules of which the binding groove is formed by one chain, the binding groove of MHC-II molecule is formed by the α1 and the β1 domains (Fig. 1.6a), while α2 and β2 domains fold to form the lower Ig-like domain[45–47]. The architecture of MHC-II molecule looks similar to MHC-I

molecule, but there are some differences with respect to the binding groove. The ends of the MHC-II binding groove are open, which allows longer peptides ranging from 12-25 aa in length to bind[48].

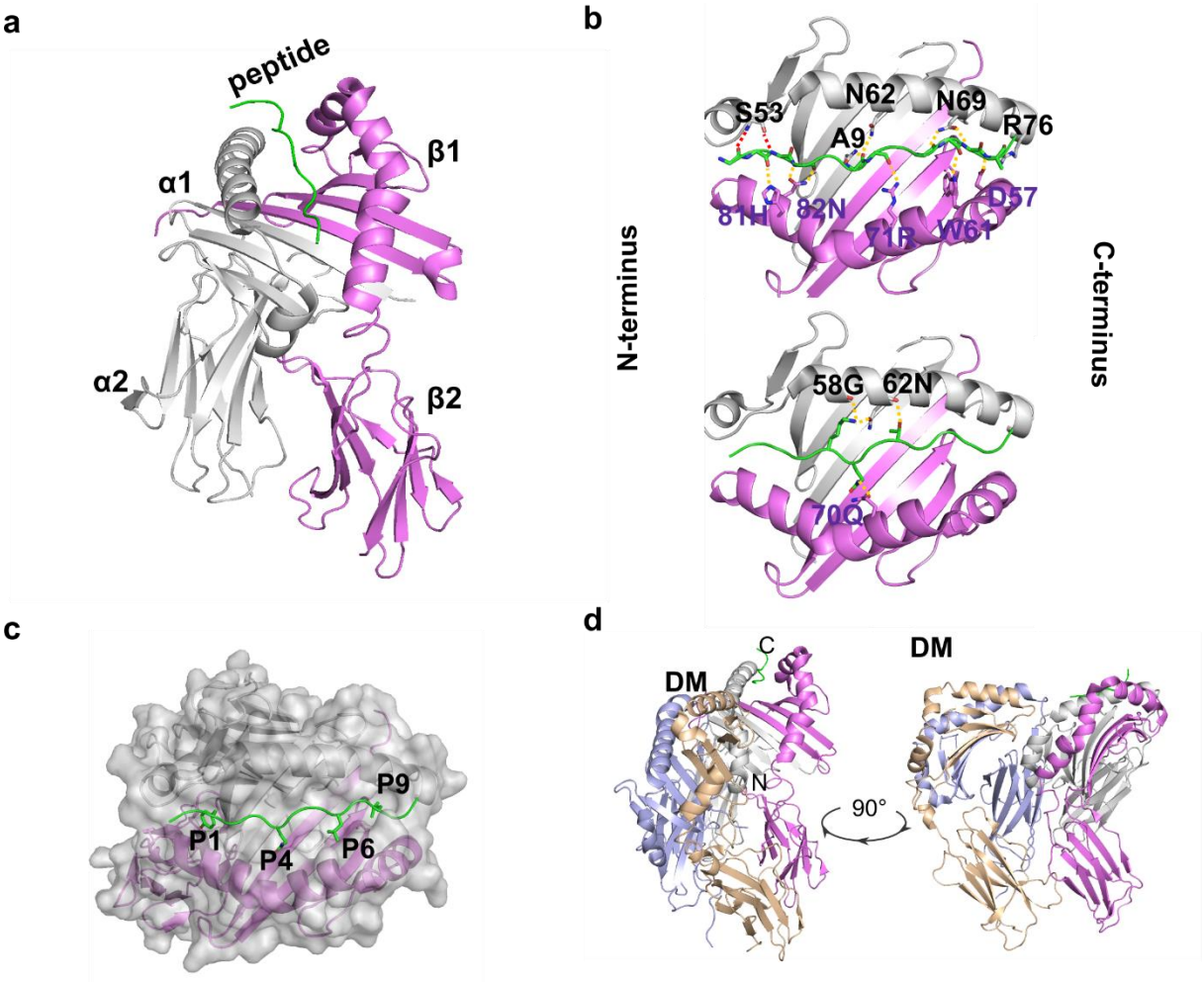


Figure 1.6 Structure of MHC-II molecules. a) Cartoon structure of a classical MHC-II molecule with peptide bound (HLA-DR1/HA 307-319 peptide, HA: influenza hemagglutinin) in the binding groove (PDB ID: IKG0)[49]. α chain is in gray and β chain is in purple. **b)** upper: H-bonds between side chain of residues of heavy chains and the backbone of the bound peptide are shown in yellow dash line. In addition to that, backbone of S53 forms H-bonds with backbone of peptide are shown in red. Lower: the interaction between side chains of the bound peptide and side chain of heavy chain. Residues from α chain are labelled in black and residues from β chain are labelled in blue. **c)** Transparent surface representation of MHC-II molecule with highlighting P1, P4, P6, P9 pockets. Anchor residues are shown as sticks. **d)** Crystal structure of DR1-DM complex. DM (α chain is in gray blue and β chain is in wheat) shows similar structure as MHC-II molecule and binds to MHC-II from the N-terminus of the binding groove. The binding groove of DM is narrower compared to classical MHC-II molecules. Thus, no peptides can bind into the binding groove of DM.

There are also two types of interactions (Fig. 1.6b) that contribute to the affinity of peptide: 1) H-bonds between the backbone of the peptide and side chains of residues along the α and β chains of the binding groove; 2) The interactions between side chains of the peptide and residues of the pockets in the binding groove. Similar to MHC-I, the first type of interactions in MHC-II molecules is highly conserved: residue α9, α62, α69 and residue β61, β71, β81, β82 form H bonds with the backbones of the peptide (Fig. 1.5b, the upper panel). In contrast to the first type of interactions, the second type of interactions is MHC-II and peptide-sequence dependent (Fig. 1.6b, the lower panel). The binding groove of MHC-II can be divided into 9 pockets, named as P1 - P9, starting from the N- to the C-terminus of canonically

bound peptides[50]. Among them P1, P4, P6 and P9 pockets accommodates the side chain of the bound peptide (Fig. 1.6c).

There are three classes of classical MHC-II molecules: HLA-DR, HLA-DP, HLA-DQ. Those proteins are usually only expressed in antigen presenting cells (APCs: such as B cells, macrophages and DCs) and present peptides derived from extracellular sources processed by endosome to $CD4^+$ T-cells. There are also MHC-II-like proteins, in particular, HLA-DM (DM) and HLA-DO (DO). These non-classical MHC-II molecules have a similar structure compared to classical MHC-II molecules, except that there are no peptides in the binding groove. Roles of non-classical MHC-II-like proteins are related to the antigen processing and presentation pathway[51,52]. For example, DM catalyzes peptide exchange of MHC-II molecules[53]. While DO, the expression of which is restricted to the thymic medulla, B cells and certain dendritic cell subsets, can form a stable complex with DM to inhibit the activity of DM[54]. However, the function of DO in antigen presentation is still poorly understood[55].

1.2.5 MHC associated diseases

The MHC locus is one of the most intensively studied regions in human genome because of its association with autoimmune, infectious and inflammatory diseases[56]. Due to the LD phenomenon and high polymorphism of the MHC locus, it is not easy to identify which single MHC allele is responsible for the cause of a specific disease. Over the past decades, large datasets containing more precise sequencing as well as advanced computational methods have been developed, which makes it possible to define the amino acid in MHC genes that is related to the associated disease[57].

Variants of MHC alleles have been described to be associated with autoimmune diseases. Different alleles that correlated with one specific disease normally show similar structure features, such as the peptide binding pockets or the interface with TCR, demonstrating the presented peptide pool is similar. Ankylosing spondylitis (AS) was considered to correlate with HLA-B27 alleles (HLA-B*27:05, B*27:02, B*27:04, and B*27:07, but not B*27:06 and B*27:09)[58,59], where the differences between these alleles are mainly in the F-pocket region. Multiple Sclerosis (MS) is a chronic inflammatory demyelinating disorder of the central nervous system and was reported to be associated with the HLA-DR2 haplotype (DRB5*01:0, HLA-DQB1*06:02, HLA-DRB1*15:01)[60]. Type 1 Diabetes (T1D) is a chronic autoimmune disease characterized by T cell-mediated destruction of pancreatic islet beta cells, resulting in irreversible insulin deficiency and long-term dysfunction of several organs and tissues. To date most evidence supports a role for variation at the HLA-DQ-DR haplotype as the major disease predisposing locus[61]. The predominant role of DRB1*04-DQA1*03:01-DQB1*03:02 and DRB1*03-DQA1*05:01-DQB1*02:01 haplotypes in susceptibility to T1D in European populations has been described[61]. Systematic Lupus Erythematosus (SLE) is a multisystem autoimmune disease primarily affecting young women. The susceptibility to SLE has been shown to correlate with the HLA-DR3 (DRB1*03:01), HLA-DR2 (DRB1*15:01) and HLA-DR9-containing haplotypes[62]. Ulcerative Colitis (UC) and Crohn's Disease (CD) are related inflammatory diseases of the gastrointestinal tract commonly known as inflammatory bowel diseases and have been described to be associated with HLA-DRB1*15:02 and HLA-DRB1*01:03. Rheumatoid arthritis (RA) is associated with variations at HLA-DRB1[63]. Studies on infectious diseases such as human immunodeficiency virus (HIV), human hepatitis

B virus (HBV), human hepatitis C virus (HCV), human papilloma virus (HPV) seropositivity, and tuberculosis have also shown that control of virus infection correlates with MHC variants[64–68].

1.3 Assembly of peptide-MHC complexes

Presentation of antigens by MHC-I and MHC-II molecules on the cell surface to T-cells is a highly coordinated process, coupled with peptide processing and maturation of MHC molecules. This process is accomplished by complex pathways, involving different regulatory proteins as well as trafficking of protein between different subcellular compartments. During the assembly process of MHC molecules, a number of chaperones and accessory proteins assist the MHC molecule folding and optimize the peptide repertoire presented.

1.3.1 The antigen processing and presentation of MHC-I

Antigen processing and presentation is the process of generating optimal peptides for MHC-I and MHC-II molecules for presentation on the cell surface. Peptides presented by MHC-I molecules are derived either from the degradation of ubiquitin-labelled endogenous proteins (from normal cellular proteins and intracellular bacteria and virus) or exogenous proteins that are transported to the cytosol from the endosome[69,70] (Fig. 1.7). In normal conditions, MHC-I molecules present self-peptides derived from endogenous proteins on every nucleated cell[71]. Peptides from self-proteins are normally ignored by TCRs. In viral infections or tumor formation conditions, peptides derived from the non-self- or mutated-self-protein are presented on the cell surface to the cytotoxic T cell. Only the peptides derived from pathogens or mutated self-proteins (like tumor cells) can trigger the immune response, which allows $CD8^+$ T cells to perform its surveillance role in eradicating viruses and other intracellular pathogens, and also exert potent antitumor activity. The recognition of viral peptides or neoantigen by TCR eventually leads to the death of the abnormal cells. The co-stimulatory receptor CD8 on the T cells binds to the conserved α3 domain to further enhance the response intensity[72].

Under infection or tumor formation conditions, subunits of the proteasome such as LMP2, LMP7 encoded in the MHC-II loci are upregulated by IFN-γ and replace several subunits of the standard proteasome to form the so-called immunoproteasome[73,74]. The exchange of subunits guarantees that peptides digested by the immunoproteasome bear hydrophobic or positively charged residues at the C-terminus, which is compatible with the peptide selectivity of transporter associated with antigen processing (TAP) and F-pocket anchor residue preference of MHC-I molecules. This reflects the coevolution or coordination of proteins involved in the antigen presentation pathway. Besides the effect of the immunoproteasome on peptide processing, it has been shown that the immunoproteasome influences the CD8+ T-cell responses by modifying the peptide repertoire of the corresponding T cells[75]. Peptides produced by the immunoproteasome have a length distribution of 4-30 residues with most ranging between 6-11 residues. The N-terminus of peptide unsuitable in length will be further trimmed by cytosol-resident aminopeptidases (Fig. 1.7), such as tripeptidyl peptidase II (TPPII)[76], to generate optimal peptides. Digested peptides then will be exported by TAP into the endoplasmic reticulum (ER) lumen[36] (Fig. 1.7). The TAP-mediated peptide translocation is peptide-sequence and -length dependent in a way compatible with the property of peptides isolated from MHC-I molecules[77–80]. The N-terminus

of transported longer peptides can be further trimmed by the ER aminopeptidase (ERAP1 and ERAP2) to optimize their size for binding to the binding groove of MHC-I molecules[81–84] (Fig. 1.7). The optimization of the N-terminus of the peptide by ERAP can even take place after its binding to MHC-I molecules, shown by a recent research[85].

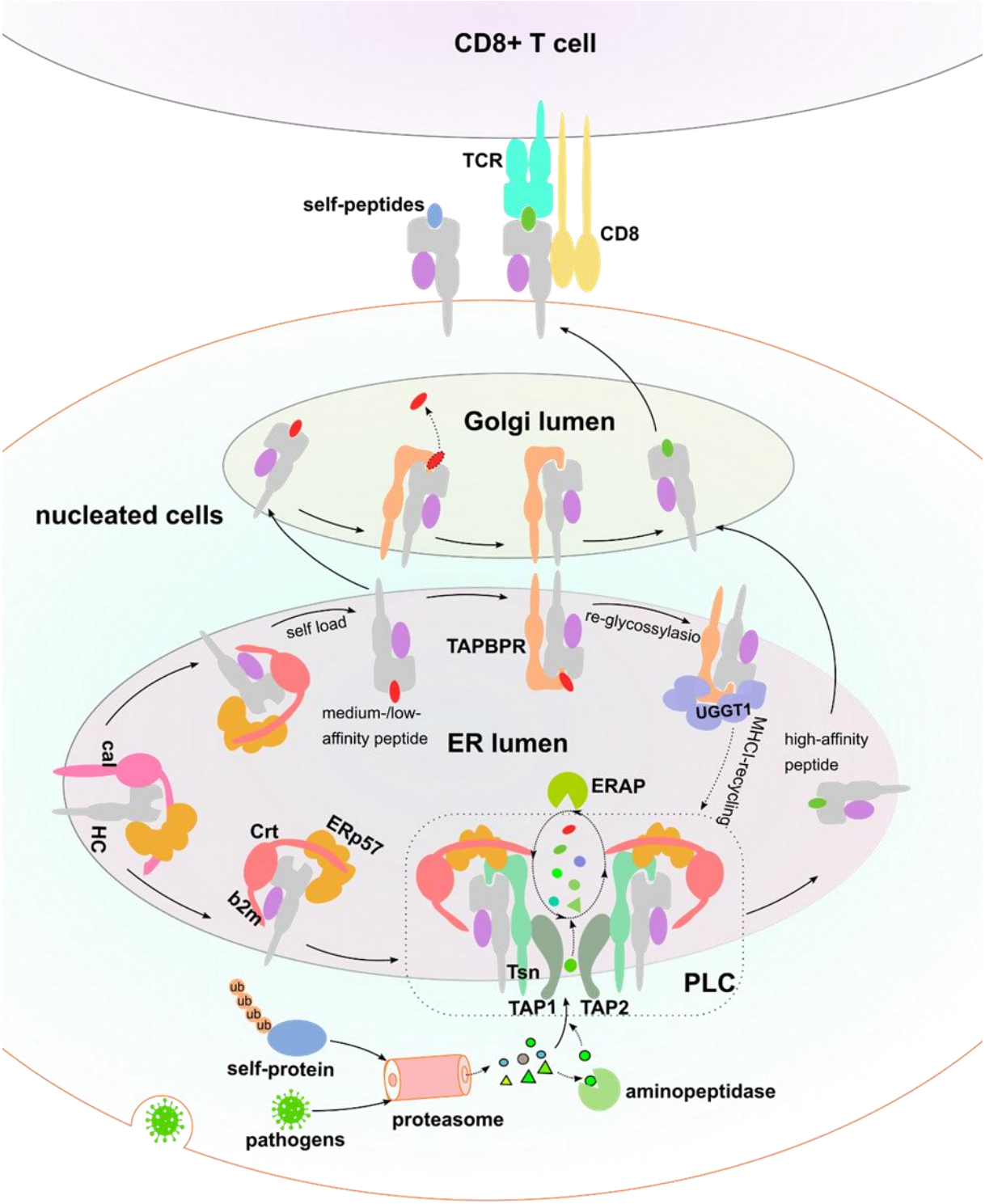


Figure 1.7 MHC-I antigen presentation pathway. Nucleated cells present peptides derived from self-protein labeled by ubiquitin or from pathogens under infection condition to $CD8^+$ T cells. The whole process can be divided into two flows: peptide processing flow and MHC-I assembly flow. The two flows meet in the PLC. Peptide processing flow: digested peptides by proteasome either first are trimmed by aminopeptidase for suitable length to be transferred into ER by TAP1/2 or directly transferred into ER by TAP1/2 if the length is optimal. Peptides transferred into ER will be further processed by ERAPs to

suitable length. The MHC-I assembly flow: Newly synthesized heavy chain folds with the help of calnexin via glyco-moiety and thiol oxidoreductase ERp57. Then the properly folded heavy chain associates with β2m with the help of calreticulin to form the MHC-I heterodimer. Tapasin recruits MHC-I heterodimers to the peptide loading complex to access peptide pool. From this step on, the optimal peptide loading to MHC-I heterodimers is the key issue. Two editors control the quality of MHC-I molecules: Tapasin, the first MHC-I editor and core subunit of the PLC, can stabilize the empty MHC-I heterodimer and promote peptide loading of MHC-I. TAPBPR, the second quality control and peptide editor outside of the PLC, promotes suboptimal-peptide-loaded MHC-I molecules to exchange with high affinity peptides or promote re-glycosylasion of MHC-I in the absence of suitable peptides via glucosyltransferase UGGT1 (this procedure can take place both in ER and Golgi). Re-glycosylated MHC-I can be recognized by calnexin and join the MHC-I assembly flow again. Only MHC-I molecules loaded with high-affinity peptides leave ER and traffic through Golgi to the cell surface.

Loading of the peptides in the ER to MHC-I molecules is mediated by the peptide loading complex (PLC), which comprises TAP, tapasin, ERp57, calreticulin and HC-β2m heterodimer (Fig. 1.7). After HC-β2m heterodimers load with high affinity peptides in the PLC, pMHC-I molecules will be released from the PLC and traffic through the ER and the Golgi apparatus within vesicles, and eventually to the cell surface. In addition to the PLC, the tapasin-related protein (TAPBPR) was found to catalyze peptide editing of MHC-I molecules as well[86–89]. Only stable pMHC-I molecules on the cell surface can be recognized by cognate TCRs on $CD8^+$ T cells and trigger efficient immune response. Thus, the assembly of MHC-I molecules with optimal peptides is important for the efficiency of the immune response.

1.3.2 The folding of MHC-I molecules

Several refolding steps take place before HC-β2m heterodimers reach to the PLC for peptide loading. Some glucosidases and ER chaperones participate in the folding steps of HC-β2m heterodimers [90–93]. If MHC-I molecules are not correctly folded they will be recognized by UDP-glucose glucosyltransferase (UGGT) and will be re-monoglucosylated[92,93]. In the early stage of MHC-I maturation, newly synthesized class I HC polypeptides associate with the ER chaperones after HCs are translocated into the ER, such as the immunoglobulin binding protein (BiP) and membrane-bound calnexin. BiP binds to human class I HC at an earlier stage or at the same time as calnexin. BiP is one of the most abundant chaperone molecules in the ER and stabilizes newly synthesized polypeptides during folding, prevents aggregation, mediates retention in the ER and suppresses formation of nonnative disulfide bond[92]. Calnexin is a so-called "sugar-binding" protein and interacts specifically with the monoglucosylated N-glycosylated protein. Calnexin interacts with class I HC by recognizing the $Glc_1Man_9GlcNAc2$ glycan moiety on the conserved residue Asn86 of the HC[94]. The interaction with calnexin stabilizes class I HCs and helps them properly fold[95]. The thiol oxidoreductase ERp57 then binds to the calnexin-HC to guarantee the correct disulfide bond formation of class I HCs[95]. The formation of the first disulfide bond in the α3 domain takes place soon after the calnexin binds to the class I HC and the association of calnexin-HC with ERp57. The second disulfide bond in the α2 domain remains unformed as long as the peptide binding groove of MHC-I is not occupied by a peptide. The conformation of class I HC at this time point can be recognized by β2m. After HC association with β2m, calnexin dissociates from the HC-β2m heterodimer, leaving a conformation that can be recognized by calreticulin (Crt)[96]. Then tapasin recruits HC-β2m heterodimer/Crt/ERp57 to TAPs for accessing the peptide pool[97]. Tapasin bridges TAP and HC-β2m heterodimer by interacting with TAP via its transmembrane domain and interacting with MHC-I via its N-terminal domain, respectively. Thus, HC-β2m heterodimer, Crt, tapasin, ERp57 and TAPs together form the PLC[95].

1.3.3 The components of the PLC

The PLC is important for the assembly of MHC-I molecules. The structures and functions of the PLC components TAP, calreticulin and ERp57 are descripted in this section. Tapasin, the focus of this thesis, will be covered in more detail in the following section.

TAP is a heterodimer formed by two homologous subunits, TAP1 and TAP2[98]. Both are encoded by genes located in the MHC-II locus. TAP1 and TAP2 are multi-transmembrane proteins with an ATP-binding cassette (ABC) and belong to the family of ABC transporters[99]. They interact with one another in the ER membrane to form TAP. TAP transmembrane domains (TMDs) contains two sets of four transmembrane helix bundles (TMD0) and one core part that contains the binding site for peptides (Fig. 1.8a), whereas the nucleotide-binding domains (NBDs) of TAP are responsible for ATP binding and hydrolysis (Fig. 1.8a). TMD0s provide tapasin the binding site[100,101], which is important for the assembly of the PLC. The translocation of peptides by TAP into ER is initiated by peptide association with TMDs, followed by steps involving allosteric conformational rearrangement of TAP and ATP hydrolysis[102,103]. In brief, peptide binding to the TMDs triggers the allosteric cross talk between TMDs and NBDs, which induces NBDs dimerization. The TMDs then switch from an inward-open conformation that seals the pathway of peptides to the ER lumen, to an outward-open conformation that allows for the peptide to cross the membrane (Fig. 1.8b). The tight NBD dimer binds to two ATP molecules. ATP hydrolysis is coupled to the peptide release into the ER lumen (Fig. 1.8b). TAP selectively transports peptides that are compatible with the peptide preference of MHC-I molecules[77–80]. The human TAP transport efficiency of peptides in the range of 8-11 residues with hydrophobic or basic C-termini are highest, although transportation of longer peptides can happen but in a reduced efficiency[104].

Calreticulin (Crt) is a homolog of the ER-chaperone calnexin. Unlike calnexin, which is a membrane-bound protein, Crt is a soluble protein that consists of three domains (Fig. 1.8c). The globular domain of Crt bears a calcium-binding site and glycan-binding sites that specifically interact with the monoglucosylated glycans of glycoproteins[91], including glycans on the MHC-I HCs and tapasin. The P-domain of Crt is flexible, arm-like and proline-rich. Different crystal structures of P domain that reveal the P domain orientation relative to the globular domain is unfixed[105]. The tip of the P domain is the binding site for ERp57[106] which is a co-chaperone of Crt[107]. The acidic domain in the C-terminal region of Crt can binds to calcium ion with low affinity, which is important for cellular calcium homeostasis[108]. This domain is absent in the calnexin. Crt contributes to the assembly of MHC-I molecules in several ways[109]. First, Crt binds to the class I HC in an earlier stage to stabilize the HC. Second, Crt functions in the folding and stabilization of tapasin. Tapasin is also a glycanprotein and has a single N-linked glycosylation site at asparagine 233 (N233). Crt can bind to this glycosylation site of tapasin, which is important for the stabilization of tapasin, thereby for that of the whole PLC[110]. Recruitment of Crt to the PLC requires its association with ERp57 and the glycan site of tapasin[111,112]. Furthermore, Crt binds to the sub-optimally loaded MHC-I proteins during their trafficking to the cell surface and recycle them to the ER for peptide optimazation[113].

ERp57 is a luminal protein of the ER and catalyzes the formation of disulfide bonds in newly synthesized glycoproteins during their folding process[114]. It is a member of protein disulfide isomerase

(PDI) family. It works as a component of calnexin/Crt chaperone system to promoting the folding and oxidation of the glycoprotein[107,114]. Like other disulfide isomerase, ERp57 has four domains (Fig. 1.8d): the *a* and *a′* domains contain redox active CXXC motifs (where 'X' is any amino acid) that are responsible for the oxidase function, whereas the *b* and *b′* domains assist in the substrate binding but do not have catalytical activity. ERp57 interacts directly with HCs before they become an integral of the PLC and its association with HCs at this stage promotes the second disulfide formation in HCs[115,116]. ERp57 forms a stable disulfide conjugate with tapasin (ERp57-Tsn) via its *a* domain with tapasin residue C95[117–119]. This conjugate is an integral part of the PLC[95]. Unexpectedly, the redox activity of ERp57 is not essential for its functions in peptide loading of MHC-I in the PLC[120,121]. No direct interaction between ERp57 active domains (*a* and *a′*) and class I HCs are observed. The ERp57-Tsn conjugate prevents the disulfide bond in the binding groove against reduction and maintain the MHC-I molecule in a peptide receptive state[122].

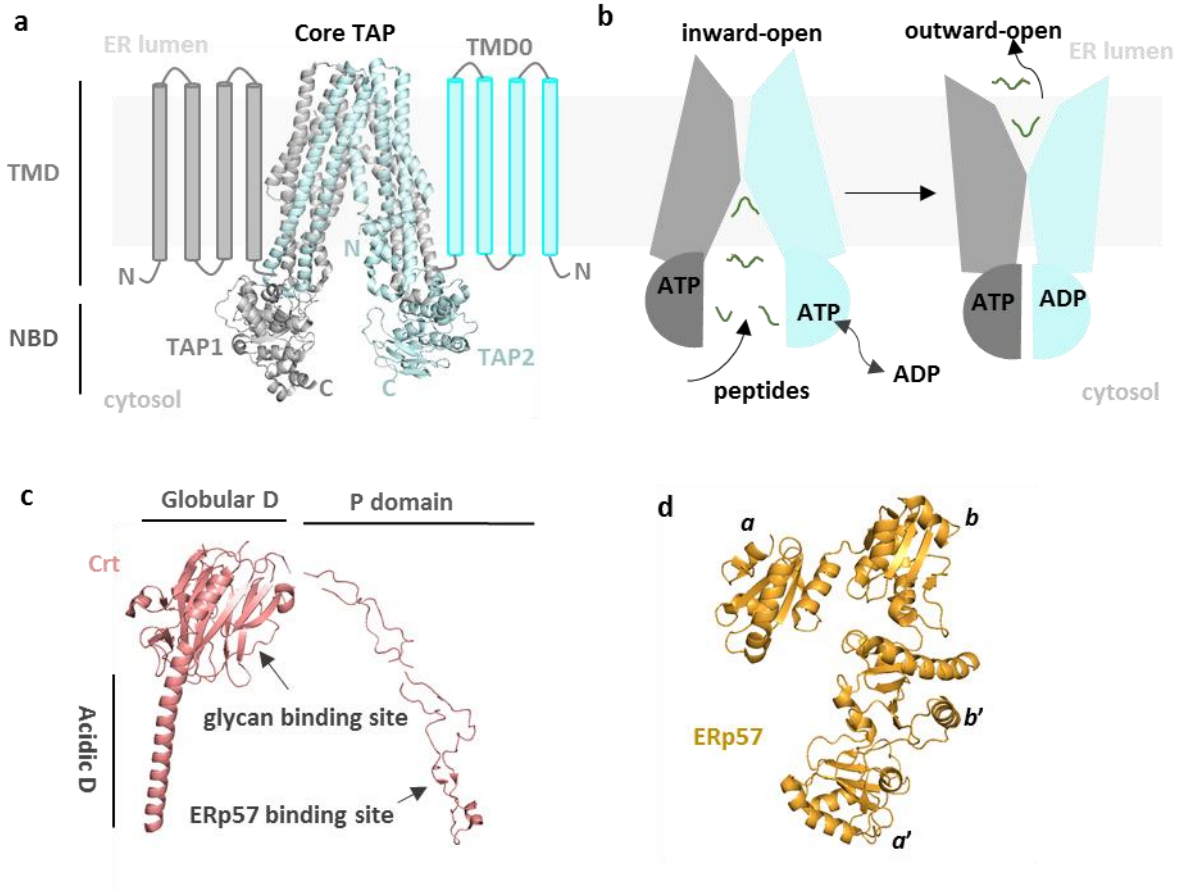


Figure 1.8 Chaperones in the ER that help MHC-I molecules assembly. a) The organization of TAP structure based on the cryo-electron microscopy (cryo-EM) of core TAP (PDB ID 5UID)[123]. **b)** The model of TAP conformational exchange during peptide translocation. Inward-open and outward-open conformation are depicted. **c)** The structural features of calreticulin, modified from PDB ID 6ENY. **d)** The structure of ERp57, modified from PBD ID 3F8U. Four domains are marked.

1.3.4 The assembly of MHC-II molecules

Since the antigen presentation pathway of MHC-II is not the focus of this thesis, a short description of this process is given here as a comparison with antigen presenting pathway of MHC-I. Peptides presented by MHC-II molecules are derived from extracellular proteins taken up by endocytosis and

degraded in endosomes/lysosomes[124,125]. The proteolysis activity is triggered by acidic pH (pH = 4-5). Newly synthesized MHC-II α/β chains are translocated into the ER and assemble as heterodimers. The MHC-II heterodimers first bind to a trimerized invariant chain (Ii) complex to form nonameric (MHC-II-Ii)$_3$ complex, in which the binding groove of MHC-II heterodimer is occupied by the unstructured motif of Ii[126,127]. The (MHC-II-Ii)$_3$ complex is transported into endosome/lysosomes, guided by the leucine-based targeting motif at its cytoplasmic trail through the trans-Golgi network. In the endosome/lysosome compartment, the Ii in the (MCH-II-Ii)$_3$ complex is hydrolyzed by aspartyl protease, the lysosomal endoprotease cathepsin S, L and F, leaving a core motif of Ii which is called class-II associated invariant chain peptide (CLIP) in the binding groove of MHC-II[128,129], binding in a conformation as an antigenic peptide[130]. CLIP is then exchanged for the processed peptides in the lysosome with the help of catalyst DM. This step is the central step of MHC-II antigen-presentation pathway[53,131]. DM promotes peptide exchange of CLIP/MHC-II with high affinity peptides by interacting with the part of the binding groove that usually accommodates the N-terminus of the peptide and affects the P1 pocket (Fig. 1.5d)[132]. It has been shown that DM promotes MHC-II peptide exchange by selecting the DM-catalyzed conformation sampled by MHC-II molecules[133,134]. The activity of DM is regulated by DO, another MHC-II-like protein[54,135]. After exchanged with high affinity peptide, pMHC-II molecules are then transported to the cell surface and are recognized by the cognate TCRs on the CD4$^+$ T cells, triggering the immune response. Furthermore, the co-stimulatory receptor CD4 on the T cells binds to the conserved β2 domain of MHC-II with its N-terminal Ig domain to enhance response intensity[72].

1.4 The MHC-I peptide-editing chaperones

1.4.1 Tapasin

Human tapasin (Tsn) is 428 amino acids in length (not including the signal peptide, Fig. 1.9c)[136]. It contains a soluble domain (amino acid 1-392), a transmembrane domain (393-417) and a short C-terminal cytosolic tail with an ER retention motif (418-428, Fig. 1.9b). It was first identified as a TAP-associated glycoprotein[96,136]. Experiments using deletion mutants of tapasin revealed that its N-terminal region is required for the interaction with MHC-I molecules, whereas the C-terminal region (including the transmembrane domain and the cytosolic tail) is critical for TAP binding[137,138]. It stabilizes TAP in a way that enhances the peptide translocation ability of TAP from cytosol into the ER[137,138]. It acts as a physical connection between MHC-I and TAP, thereby placing the MHC-I close to the peptide source[97,139]. This bridge function was thought to be only important for the assembly of the PLC and exportation of MHC-I molecules from the ER to the cell surface. However, more data led to the speculation of tapasin being involved in editing the peptide repertoire presented by MHC-I, because eluted peptide pool and immune response of MHC-I molecules are affected in tapasin-deficient cells[97,140–144]. Furthermore, re-expression of soluble tapasin which lacks the association site to TAP in tapasin-deficient cells restores the cell surface expression and function of MHC-I molecules, suggesting tapasin does not only function as a bridge between MHC-I and TAP, but also enhances peptide loading of MHC-I molecules[137,138]. The peptide-editing function was further demonstrated by its role in optimization of peptides for MHC-I molecules[119,140,141]. All these cellular studies illustrate that tapasin

affects the assembly of MHC-I molecules. However, how exactly tapasin performs its editing role was unclear. The crucial function of tapasin in MHC-I complex assembly was further highlighted by the finding that tapasin forms a conjugate with ERp57 through a disulfide bond (C95 of Tsn with C57 of ERp57)[117–119]. ERp57 catalyzes the formation of disulfide bonds in newly synthesized glycoproteins during folding process[114]. The covalent association between Tsn and ERp57 was thought important for catalyzing the isomerization of disulfide bonds in the α2 domain of MHC-I heavy chains[119]. Later it showed that ERp57-Tsn conjugate can protect the disulfide bond in the α2 domain from reduction and thus stabilizes the MHC-I[118,122].

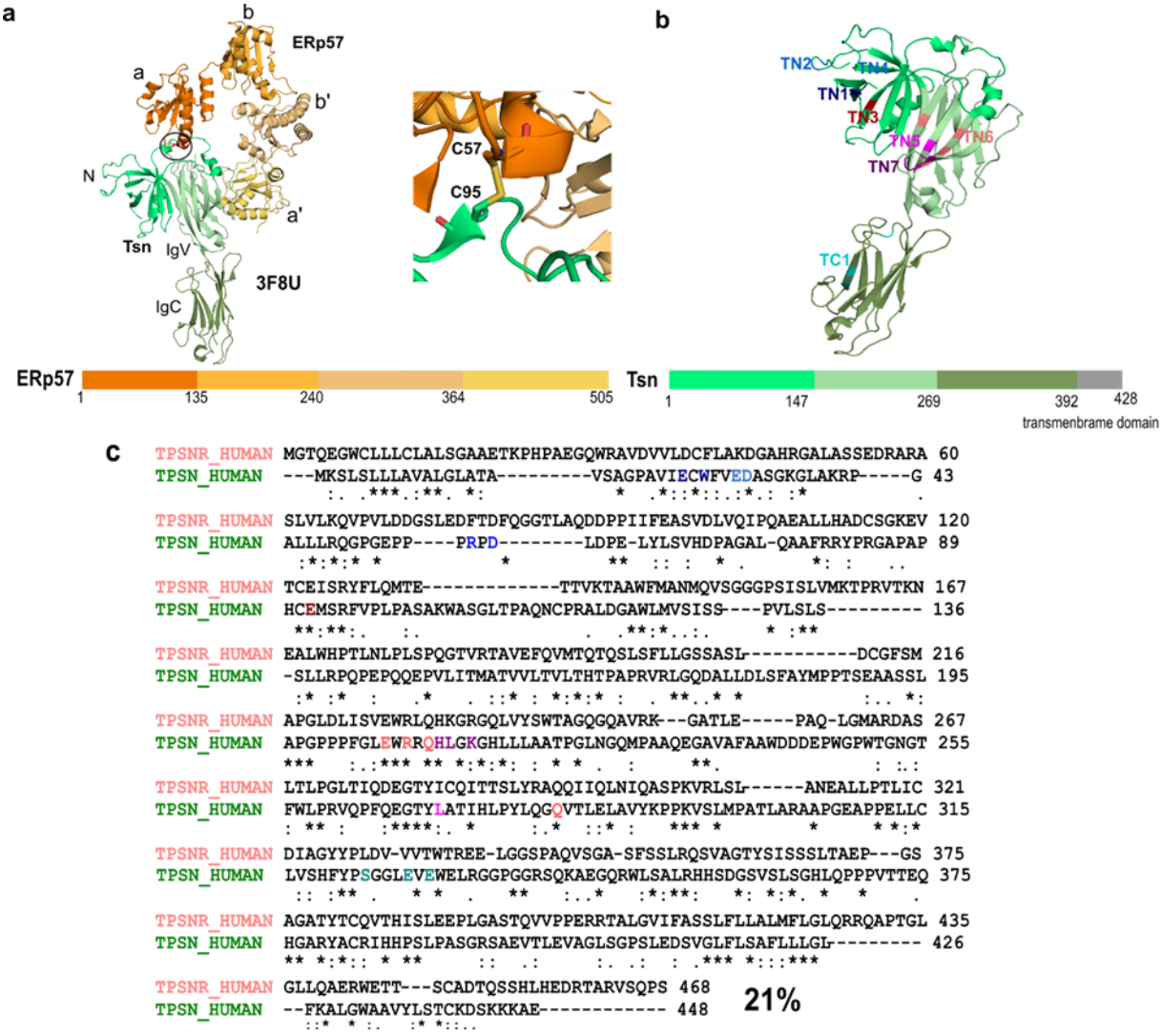


Figure 1.9 Structure of Tapasin and its sequence. **a)** Upper: Crystal structure of ERp57-Tsn (PDB ID 3F8U)[145]. Domains of ERp57 are colored in different yellow and domains of Tsn are colored in different green. A zoom in of the circled area illustrates the disulfide bond between ERp57 and Tsn. The linear sequence labelled with mumbers is indicated in the lower panel. **b)** Mutations of TN1(E6K/W8A), TN2(R37E/D39K), TN3(E72K), TN4(E11K/D12R), TN5(L250K), TN6(E185K/R187E/Q189S/Q261S), TN7(H190S/L191A/K193E), TC1(S303W/E307K/E309K) are mapped on the structure of Tsn[145]. **c)** Sequence alignment of tapasin (TPSN_HUMAN) and TAPBPR (TPSNR_HUMAN). Mutant TN1-TN6 are highlighted in colors. Color code is the same as in **b**.

Cell-free experiments have shown that the recombinantly expressed ERp57-Tsn conjugate enhances peptide loading and editing of MHC-I molecules extracted from tapasin-deficient cells *in vitro*[118], highlighting the ERp57-Tsn conjugate is the functional unit of the PLC. It also demonstrates that ERp57

is essential for maintaining Tsn's full function, because the ability of soluble Tsn alone to enhance peptide loading is lower than that of the conjugate. Shown by crystallography of the ERp57-Tsn conjugate, Tsn forms an L-shaped structure composed of an N-terminal domain and a C-terminal Ig-like domain. The N-terminal domain is a mix of a 7-stranded β-barrel (residues 1-147, β-strands 1-7) stabilized by a disulfide bond (between Cys7 and Cys71) and an Ig-like domain (residues 150-269, β-strands 8-14, Fig. 1.9b). The C-terminal domain of Tsn also forms an Ig-like fold stabilized by a disulfide bond between residue Cys295 and Cys362. It is worth noting that in the ERp57-Tsn crystal structure, residues 12-19, 28-33 and 171-176 of Tsn are poorly ordered, indicating flexibility of these regions. The flexibility of these regions might relate to the function of Tsn in facilitating peptide loading.

1.4.2 The interactions between tapasin and MHC-I molecules

As mentioned above, the molecular understanding of how tapasin catalyzes peptide editing was unclear. *In vitro* experiments using recombinantly expressed soluble Tsn linked to MHC-I by Jun-Fos linkage showed that tapasin exerts two critical functions in the MHC-I assembly[146]. One is that Tsn promotes peptide loading to MHC-I by acting as a chaperone to stabilize the peptide-receptive conformation of MHC-I. The other is that Tsn promotes peptide dissociation of MHC-I molecules by interrupting the conformation in the peptide binding groove[146]. It has been thought that tapasin contacts two regions of class I HCs in the PLC: one region close to the $\alpha_{2\text{-}1}$-helix that might interact with the N-terminal domain of Tsn and one region (a loop) in the α3 domain that probably interact with the Tsn Ig-like domain[147–149]. Mutations in these two regions showed either diminished interaction with tapasin or reduced peptide editing by tapasin. Mutational analysis of Tsn reveals that several conserved surfaces are important for its MHC-I binding ability and are critical for its peptide loading and editing function (Fig. 1.9b)[145]. Impaired MHC-I binding has been shown for Tsn mutants TN3 (E72K), TN4 (E11K/D12R), TN5 (L250K) and TN7 (H190S/L191A/K193E). The ability to catalyzing peptide loading of these mutants was reduced to ~50% of the wild-type Tsn[145]. Tsn mutant TN6 (E185K/R187E/Q189S/Q261S) completely loses its binding ability to MHC-I molecules and peptide loading catalyzing activity[145]. The molecular dynamic (MD) simulation model of Tsn/B*44:02 complex shows that Tsn contacts B*44:02 $\alpha_{2\text{-}1}$-helix with its K16, L18, and L79, the $\alpha_{2\text{-}1/2}$ hinge with its W85, and the underside of the β-strands with its R187 and Q189. From the side of B*44:02, residues in the $a_{2\text{-}1}$ helix (mainly I142, R145), R151 and residues in the β-strands (D129, S132, T134) are involved in the interaction with tapasin[150]. In combination with the crystal structure of ERp57-Tsn[145] (Fig.1.8a), the roles of interactions between Tsn with MHC-I become clearer. The interaction of the $\alpha_{2\text{-}1}$-helix of MHC-I with the conserved residues of tapasin (indicated by TN6, TN7) is likely to stabilize the binding groove of MHC-I in a peptide-receptive conformation.

1.4.3 Dependence of MHC-I molecules on tapasin

Allele-specific reduction in MHC-I cell surface expression was first observed in tapasin-negative cell lines (i.e., human B cell line .220)[151]: The cell surface expression of HLA-A1 and -B8 was strongly reduced, HLA-A11, -A24, and -B5 were moderately reduced, and HLA-A2, -A3, and -B7 were reduced slightly[151]. Similarly, a later study investigated the cell surface expression of MHC-I dependence on tapasin in 721.220 cell line (known to be tapasin-deficient) and showed cell surface expression of

B*27:05 was slightly affected, B8 intermediately, B*44:02 significantly reduced in the absence of tapasin[152]. Interestingly, the peptide repertoire of B*27:05 showed subtle difference in the presence and absence of tapasin[141]. B*44:02 and B*44:05 showed opposite requirement on tapasin for antigen presentation and the surface expression of B*44:05 showed no dependence on tapasin[153]. Another study analyzed the binding ability of B*15 subtypes to tapasin and observed that association with tapasin is higher for B*15:10 than B*15:01 and B*15:18[154]. A*24 subtypes A*24:06 and A*24:13 showed less dependence on tapasin for assembly than A*24:02, although their association level with tapasin were similar[155]. A very recently published paper reported the dependence on tapasin of 97 MHC-I allotypes (include HLA-A,-B and -C allotypes) for cell surface expression and observed expression levels of the highly tapasin-dependent allotypes (e.g., A*01:02, A*01:01, A*36:01, B*50:01, B*44:03, B*44:02, B*49:01, B*58:02, C*04:01) were increased more than 10-fold, whereas the cell surface expression of the highly tapasin-independent allotypes (e.g., A*68:01, A*02:02, A*02:01, B*81:01, B*35:01, B*42:01) did not change in the presence of tapasin[156].

Questions then arise what features determine the allele-specific dependence on tapasin for MHC-I assembly and antigen presentation and the different affinity and activity of tapasin towards MHC-I allotypes. Polymorphism in the heavy chain (e.g., residue 116 in the F-pocket) for sure makes a difference. A study investigated the impact of the polymorphism of HLA-B on their tapasin dependence and showed that the polymorphic residues in the F-pocket region of HLA-B allotypes are determinants of the tapasin-dependent assembly[157]. Residues 114 and 116 of the heavy chain have also been shown to determine the dynamics of MHC-I molecules, thereby affecting tapasin dependence of B allotypes[158]. Residue 156 was reported to influence tapasin dependence of MHC-I, although 156 is not observed to directly interact with tapasin, nor located in the F-pocket region[159]. This indicates that cross talk between pockets could affect the dynamics of the whole binding grove and therefore affects interaction between MHC-I and tapasin. Residue triad 97/114/156 was reported to determine the relative levels of tapasin-independence for HLA-A*24 subtypes[155]. The conformational dynamics of pMHC-I correlates with its tapasin dependence[160]. The stability of pMHC-I molecules affects their tapasin-dependence for peptide exchange[161–163]. Features including peptide affinity, peptide length which affect the pMHC-I stability, influence tapasin activity on MHC-I[163].

In summary, polymorphism concentrated on the MHC-I binding groove region not only affects the peptide repertoire presented to the T-cells, but also influences dynamics in the binding groove[164–167], thereby affects the stability of MHC-I molecules, their interactions with tapasin and their requirements on tapasin for antigen presentation[30,87,141,157,160,168–170].

1.4.4 Another MHC-I peptide editing chaperone: TAPBPR

As mentioned previously, TAPBPR also edits peptide loading of MHC-I molecules but outside of the PLC. In 2000, *TAPBPR* (TAP binding protein-related) gene was first discovered as a tapasin-like gene on chromosome 12p13.3 and it was predicted to encode a tapasin-related protein (*TAPBPR*, or tapasin-like, *TAPBP-L*) of 468 amino acids with 21% sequence similarity (Fig.1.8c)[171]. The potential function and the ligand of TAPBPR was not explored for the following decade. In 2013, TAPBPR was classified as the second peptide editor in the MHC-I presentation pathway, but not an integral component of the

PLC as tapasin nor does it form a conjugate with ERp57[86]. Like tapasin and other proteins involved in the antigen presentation pathway TAPBPR is IFN-γ inducible[86]. From the structural point of view, TAPBPR has three domains (the N-terminal domain, the Ig-like fold and C-terminal Ig domain) and a transmembrane region followed by a cytoplasmic tail. The orientation of TAPBPR binding to MHC-I was predicted in a similar way to that of tapasin[89]. From a functional point of view, TAPBPR associates with MHC-I molecules bearing an empty or low affinity-peptide-loaded binding groove, and enhance peptide loading and exchange with high affinity peptides, similarly to tapasin.

However, TAPBPR still shows some different features compared to tapasin. First, TAPBRR is not only limited to the ER, but also traffics to Golgi together with MHC-I molecules. The difference of subcellular environment might explain why the affinity of TAPBPR with MHC-I molecules is normally higher than that of tapasin with MHC-I since the peptide concentration outside of the PLC is lower than that in the PLC. Second, the effect of TAPBPR on peptide repertoire is more subtle compared to the effect of tapasin[87]. In TAPBPR-deficient cells, the cell surface expression level of MHC-I is similar to the level in the presence of TAPBPR, indicating the peptide loading of MHC-I is not grossly affected. But the eluted peptide spectrum is broader in the absence of TAPBPR, and a subgroup of these peptides is of low affinity, indicating that TAPBPR can refine the peptide repertoire and function as quality control of MHC-I molecules. Third, TAPBPR interacts with UDP-glucose:glycoprotein glucosyltransferase 1 (UGT1) and facilitates the re-glycosylation of MHC-I, controlling the recycling of MHC-I by promoting their recognition by calnexin[172]. Fourth, TAPBPR can form tight complex with some MHC-I allotypes alone, particularly some murine MHC-I, which potentially allowed to crystallize the MHC-I-TAPBPR complex [173,174].

1.4.5 The mechanism of TAPBPR-mediated peptide editing

The discovery of TAPBPR led to a renewed interest in the mechanisms of antigen processing and presentation of MHC-I and subsequently provided insights into the function and mechanism of tapasin as well. Due to the much higher affinity between TAPBPR and MHC-I, the assembly of MHC-I-TAPBPR complex by using purified soluble TAPBPR with MHC-I molecules (e.g., HLA-A*02, H2-D^b, H2-D^d) was successful, which permitted structural analysis and mechanism studies by X-ray and NMR. Two X-ray structure of human TAPBPR in complex with mouse MHC-I, one is H2-D^b [174], the other one is H2-D^d [173], have been reported (Fig. 1.10a). The crystal structures showed that the binding groove of MHC-I in complex with TAPBPR is widened, especially the F-pocket region that harbors the dominant anchor residue of the bound peptide (Fig. 1.10c). The $\alpha_{2\text{-}1}$ helix (residues 138-150) moves outwards by 4 Å and is held by the cradle formed by the N-terminus of TAPBPR (Fig. 1.10c). Particularly Y84 of the heavy chain, which is well known to form an H-bond with the C-terminus of the bound peptide, swings out from pointing towards the F-pocket to outside and forms an H-bond with residue E102 of TAPBPR.

Three key interactions are observed in the crystal structure of MHC-I-TAPBPR. The scoop $loop_{22\text{-}36}$ of TAPPBR interacts with the F-pocket region of the MHC-I binding groove. The hairpin $loop_{187\text{-}196}$ of TAPPBR interacts with the bottom of the MHC-I binding groove, especially the $loop_{126\text{-}136}$ of MHC-I. The Ig-like domain of TAPBPR interacts with the α3 domain of heavy chain. Later NMR analysis of

the TAPPBR-MHC-I complex confirmed the interaction surfaces on H2-D^{b}[169]. Since the scoop loop was not resolved in one crystal structure and showed low electron density in the other, speculations of the function of this scoop loop arose. Later a report identified that the scoop loop is critical for the role of TAPPBR in mediating peptide dissociation of MHC-I, highlighting L30 in the loop competing with the C-terminal residue of the bound peptide for the F-pocket of MHC-I[175]. Two other reports also studied the function of this scoop loop but came up with different interpretations: The Tampé group suggests that the scoop loop has dual functions, one as an internal peptide surrogate to stabilize the empty MHC-I binding groove and a second one as peptide filter to shape the peptide repertoire[176]; The Sgourakis group demonstrates that the side chain of the scoop loop is unlikely to stretch into the F-pocket by NMR and points out that the function of the scoop loop is rather as peptide trap to increase the peptide concentration for MHC-I peptide loading when peptide concentration is low in the case of medium-high affinity peptide[177].

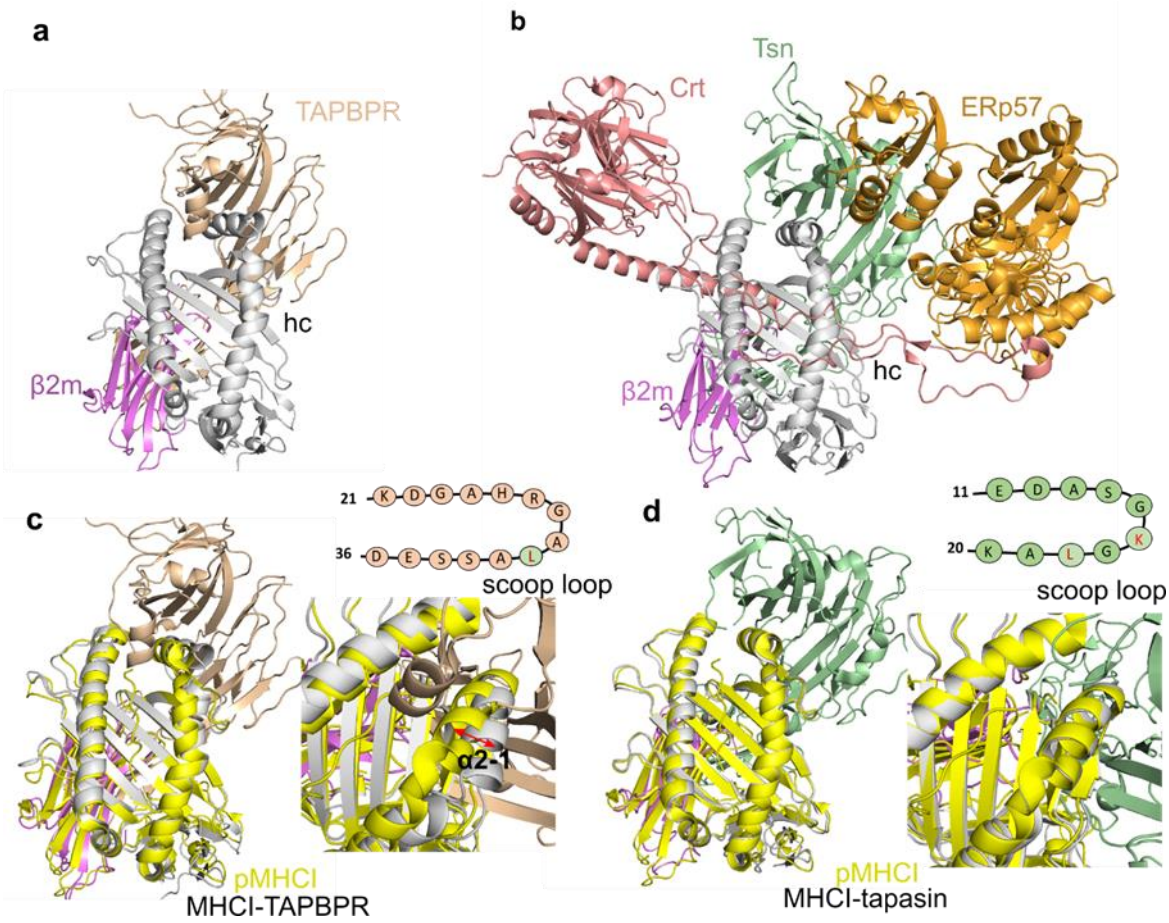


Figure 1.10 Structure of MHC-I with two editors: TAPBPR and tapasin. a) Cartoon crystal structure of MHC-I-TAPBPR (PDB ID 5WER)[173]. **b)** Cryo-EM structure of PLC (PDB ID 6ENY)[178]. **c)** Overlay of peptide-loaded MHC-I (yellow, PDB ID 2XPG) and MHC-I in complex (grey), the zoom view of F-pocket region of the binding groove showed the $\alpha_{2\text{-}1}$-helix in the complex swung out. **d)** Overlay of peptide-loaded MHC-I (yellow, PDB ID 2XPG) and MHC-I (grey) in the PLC, the zoom view of F-pocket region of the binding groove showed the $\alpha_{2\text{-}1}$-helix in the PLC was the same as in peptide-loaded MHC-I. The sequence of the scoop loop of both editors which contacts the F-pocket region is shown besides the pictures.

As for allele-specific recognition, cell-based studies have shown that TAPBPR can bind to a broad range of MHC-I allotypes, but has high preference for HLA-A, especially for the subtypes A2 and A24, over HLA-B and -C alleles[168]. Additionally, the propensity of HLA-A molecules to TAPPBR has been shown

to be related to the nature of the F-pocket[168]. A mutation analysis of HLA-A*02:01 revealed that TAPBPR can bind to a broad range of nascent MHC-I at an early stage during its folding and has a narrower specificity when MHC-I is properly folded with peptides[179]. Two NMR studies point out the correlation between recognition of MHC-I by TAPBPR and the allele-specific dynamics of MHC-I molecules and show TAPBPR selectively recognizes a minor conformational state sampled by MHC-I molecules[169,179].

1.4.6 The cryo-electron microscopy structure of the PLC

Almost at the same time when the crystal structures of MHC-I-TAPBPR complex were published, the structure of the native PLC purified from human Burkitt's lymphoma cells was solved by cryo-electron microscopy (cryo-EM)[178]. The cryo-EM analysis of the PLC reveals two sets of ERp57-Tsn editing modules and MHC-I-calreticulin arranging in a pseudo-symmetric fashion around the heterodimeric TAP. TAP serve as the central supporting point of the PLC. The architecture of the PLC visualizes how the components of PLC interplay and arrange with each other (Fig. 1.10b and d). Similar to MHC-I-TAPBPR complex, three interactions between tapasin and MHC-I were observed. Interestingly, the $\alpha_{2\text{-}1}$-helix of MHC-I was not observed to move out in the PLC, and the width of its binding groove is almost the same as in the peptide-loaded MHC-I molecule (Fig. 1.8d), indicating that the PLC helps to stabilize the MHC-I in a conformation that is ready for peptide binding without much allosteric conformational change. Since the resolution of the complex is 7.2 Å for the two editing modules and 5.8 Å for a single editing module, the atomic details of specific interaction in critical positions is not clear. Thus, detailed molecular analysis of the MHC-I-tapasin interaction is needed and will be instructive to the understanding of the mechanism.

It is worth noting that the individual PLC particles lack one or more components such as tapasin, MHC-I, ERp57 or calnexin were also found in cryo-EM single particles, indicating the PLC is a quite dynamic hub for peptide loading. Interestingly, not only the polymorphic HLA-A, HLA-B and HLA-C was identified in the PLC, but also the non-classical MHC-I heavy chains, such as HLA-E, HLA-G and HLA-F, were found in the cryo-EM single particle.

1.5 Aim of this work

The aims of this work are related to questions regarding the mechanism of tapasin-mediated peptide exchange of MHC-I molecules. MHC-I molecules are highly polymorphic and display conformational dynamics in their peptide binding groove[30,87,141,157,160,168–170], which are related to the differential dependence on tapasin or TAPBPR for efficient peptide loading, editing and cell surface expression. The recent PLC structure provides insights on how tapasin interplays with other components in the PLC, but still, the mechanism of how tapasin catalyzes peptide exchange is not clear. Furthermore, the structural elements that contribute to differential tapasin-dependence of MHC-I molecules remain elusive. Previous functional and mutational data on tapasin were mainly obtained from cellular experiments, whereas studies of the direct interaction of MHC-I and tapasin are not many (e.g., work from the Cresswell group[145]). Besides, comparison of MHC-I allotypes in turns of the peptide exchange catalyzed by tapasin *in vitro* is missing. Thus, it is necessary to investigate the direct interactions

between them as well as the effects of tapasin on peptide exchange of different MHC-I allotypes *in vitro* with biophysical and biochemical methods.

The "scoop loop" ($loop_{22-35}$ connecting β-strands 1-2) of TAPBPR, as mentioned above in section 1.4.5, one possible function of it is as a surrogate of the peptide's C-terminus competing for F-pocket binding, thereby promoting peptide dissociation and subsequently stabilizing the peptide receptive conformation of MHC-I[173–176]. However, the corresponding tapasin scoop $loop_{11-20}$, which might be responsible for peptide exchange in a similar manner, is neither resolved in the current cryo-EM structure of the native PLC nor in the crystal structure of tapasin in complex with ERp57[16,22]. Whether the scoop loop of tapasin has a similar function is unclear. However, the Leu side chain of the murine tapasin scoop-$loop_{11-20}$-derived peptide (EDAGGGGLK), which was not resolved in the PLC, is observed to bind in the F-pocket of MHC-I[180]. It is of note, however, that it is not shown whether the tapasin scoop loop can reach the F-pocket region in full length tapasin. Given that the tapasin scoop $loop_{11-20}$ is considerably shorter than that of TAPBPR, further investigation is needed to find out whether tapasin and TAPBPR engage similar or distinct molecular mechanisms to foster peptide exchange. In addition to that, a comprehensive biochemical comparison between tapasin and TAPBPR is still missing.

Based on the points mentioned above, NMR spectroscopy and X-ray crystallography in combination with other biochemical and biophysical methods were used to answer the questions listed below:

1) What are the features of MHC-I molecules that contribute to the observed allele-specific binding and catalysis by tapasin?
2) What are the critical elements for the interaction between MHC-I molecule and tapasin?
3) What is the role of the tapasin scoop loop in catalyzing peptide exchange?
4) What is the difference of tapasin and TAPBPR regarding MHC-I interactions and catalyzed exchange?

2 Materials and Methods

2.1 Materials

2.1.1 Chemicals

Chemicals (unless otherwise specified) used for buffer preparations were bought from Sigma-Aldrich or Carl-Roth (Karlsruhe, Germany). 100 bp and 1 kb DNA Latter were bought from NEB (Frankfurt am main, Germany). Marker12™ unstained standard and PageRuler™ prestained Protein Ladder were bought from ThemoFisher. MMTS (S-methyl mechanethiosulfonate) for tapasin purification and DSSO (disuccinimidyl sulfoxide) for crosslinking were bought from ThemoFisher. StageTip was home made with Empore™ Octadecyl C18 membrane (SUPELCO, USA). Detergents GDN (glycol-diogenin) and fluorinated FOS-choline-8 for cryo-EM sample preparation was from Avanti.

2.1.2 DNA synthesis and primers

DNA encoding desired proteins were ordered from ThermoFisher. Primers for subcloning were synthesized by MWG Biotech AG (Ebersberg, Germany) or integrated DNA technology (IDT, Germany).

2.1.3 Enzymes used for cloning

All restriction enzymes (*Bam*I, *Xho*l, *Sim*I, *Nde*l, *Sma*I, *Sph*I, *Xba*l, *Dpn*I) and T4 DNA ligase were ordered from NEB (Frankfurt am main, Germany). KOD hot start DNA polymerase (Novagen, Japan) and Pfu DNA Polymerase (EURx, Germany) was used to amplify desired DNA fragments from ordered constructs. OptiTaq DNA Polymerase (EURx, Germany) was used in colony PCR to select the right colony after DNA fragment ligation.

For site-directed mutation PCR, Pfu plus DNA polymerase (EURx, Germany) was used. This polymerase is very robust in cycle PCR compared to other DNA polymerases and no repeats of primers were detected to insert in the mutation sites.

2.1.4 Constructs

DNA encoding human β_2m (1-100, mature protein) was inserted into pET-3d vector. The construct of HLA- A*02:01 was a gift from Springer Lab. The luminal domain of HLA-A*02:01 (1-276, linked to a C-terminal Avi-Tag by a linker with the sequence GSGGSGGSAGG) was inserted into pET-3d vector. The luminal domain of HLA-A*03:01(1-274), HLA-B*27:05 (1-277) and HLA-B*27:09 (1-277) was inserted into pET21a vector between *Nde*l and *Xho*I. The heavy chain used for preparing NMR sample of HLA-B*27:05 and HLA-B*27:09 contains a N-terminal 6xHis tag followed by VDDDDK.

Since Tsn itself is not stable enough, it was co-expressed with ERp57 in pFastBac-Dual vector[117]. Tsn (1-398) fused with a C-terminal 6xHis tag was inserted into pFastBac-Dual vector under the polyhedrin promoter between *BamH*I and *Xba*l and ERp57(1-504, C60A was introduced to capture the conjugate state) was inserted under the p10 promotor between *Sma*I/*Sph*I. TABPBR (1-405) fused with a C-terminal 6xHis tag was inserted into pFastBac1 vector between *BamH*I and *Xba*l. C97A mutation was introduced to increase the expression yield of TAPBPR without affecting the activity of TAPBPR[172].

All mutants of β_2m, heavy chain, ERp57-Tsn or TAPBPR were generated by standard site-directed mutagenesis PCR using β_2m WT, heavy chain WT, ERp57-Tsn^{WT} or TAPBPR^{WT} as template, respectively. All constructs were sequencing confirmed in seqLab (Germany).

Table 2.1 Plasmids used in this thesis are listed below.

Name	Protein	Sequence and tag
PET3d- β_2m^{WT}	β2m^{WT}	1-100
PET3d-β2m^{T5C}	β2m^{T5C}	1-100, Thr5 is mutate to Cys
PET3d-HLA-A*02:01	HLA-A*02:01	1-276, connected with a C-terminal AviTag™ by a linker GSGGSGG-SAGG, Avi sequence is GLNDIF-EAQKIEWHE
PET21a-HLA-A*03:01	HLA-A*03:01	1-274
PET21a-HLA-B*27:05	HLA-B*27:05	1-277
PET21a-HLA-B*27:09	HLA-B*27:09	1-277
PET21a-HLA-B*27:05-his tag	HLA-B*27:05	1-277, N-terminal his tag
PET21a-HLA-B*27:09-his tag	HLA-B*27:09	1-277, N-terminal his tag
Pfastbac-1-TABBPR C97A	Human TAPBPR^{WT}	1-405, C-terminal his tag, C97A
Pfastbac-1-TAPBPR Tsn loop	Human TAPBPR$^{Tsn\ loop}$	1-405, C-terminal his tag, the loop region$_{21-36}$ is replaced by tapasin loop$_{10-21}$
pFastbacdual-ERp57-Tsn-WT	ERp57-Tsn^{WT}	ERp57, 1-504, under the Pp10 promoter, Tapasin has a C-terminal his tag, 1-398, under the Ph10 promotor
pFastbacdual-ERp57-Tsn-K16G	ERp57-Tsn^{K16G}	Lys16 is mutated to Gly
pFastbacdual-ERp57-Tsn-K16L	ERp57-Tsn^{K16L}	Lys16 is mutated to Leu
pFastbacdual-ERp57-Tsn-L18G	ERp57-Tsn^{L18G}	Leu18 is mutated to Gly
pFastbacdual-ERp57-Tsn-L18K	ERp57-Tsn^{L18K}	Leu16 is mutated to Lys
pFastbacdual-ERp57-Tsn-GGLGG	ERp57-Tsn^{GGLGG}	SGKGL is mutated to GGLGG
pFastbacdual-ERp57-Tsn-GGKGG	ERp57-Tsn^{GGKGG}	SGKGL is mutated to GGKGG
pFastbacdual-ERp57-Tsn-GGGGL	ERp57-Tsn^{GGGGL}	SGKGL is mutated to GGLGG
pFastbacdual-ERp57-Tsn-GGGGK	ERp57-Tsn^{GGGGK}	SGKGL is mutated to GGGGK
pFastbacdual-ERp57-Tsn-GGLGL	ERp57-Tsn^{GGLGL}	SGKGL is mutated to GGLGL
pFastbacdual-ERp57-Tsn-TAP-BPRloop	ERp57-Tsn$^{TAPBPR\ loop}$	SGK is replaced by TAPBPR loop AHRGAL
pFastbacdual-ERp57-Tsn-G304C	ERp57-Tsn^{G304C}	Gly304 is mutated to Cys
pFastbacdual-ERp57-Tsn-K16L-G304C	ERp57-Tsn$^{K16L-G304C}$	Gly304 is mutated to Cys, Lys16 is mutated to Leu
pFastbacdual-ERp57-Tsn-GGLGG-G304C	ERp57-Tsn$^{GGLGG-G304C}$	Gly304 is mutated to Cys, loop is mutated to GGLGG
pFastbacdual-ERp57-Tsn-GGKGG-G304C	ERp57-Tsn$^{GGKGG-G304C}$	Gly304 is mutated to Cys, loop is mutated to GGKGG

2.1.5 Peptides

Peptides were purchased either from GL Biochem (China) or DG peptides (China). In order to monitor peptide exchange by fluorescent polarization (FP) assay, the 5th residue of the peptide was replaced by

a lysine, of which side chain is modified with a fluorescein isothiocyanate function group (FITC). Photocleavable residue (indicated as J*: 3-amino-3-(2-nitro) phenyl-propionic acid) was introduced into 4th or 8th position of the peptide for generating peptide-receptive MHC-I molecules[181]. The KLIETYFSK peptide (named KK9 in the thesis) is a naturally processed epitope of human proteolipid protein[182]. GILGFVFTL (GV9) is a naturally processed epitope from influenza A MP58-66. G was mutated to K to increase the peptide solubility of photoKV9 during synthesis[181]. The RRKWRRWHL (RL9) is a self-peptide from vasoactive intestinal peptide type 1 receptor and presented by HLA-B*27:05[183]. IRAAPPPLF (IF9) is derived from Human Cathepsin A.

Table 2.2 Peptides used in this thesis are listed below.

Peptide name	Alleles	Sequence
photoKV9	HLA-A*02:01	KILGFVFJ*V
photoKK9	HLA-A*03:01	KLIETYFJ*K
photoRL9	HLA-B*27:05, HLA-B*27:09	RRKWRRWJ*L
photoRL9-4J*	HLA-B*27:05, HLA-B*27:09	RRKWJ*RWHL
PhotoIF9	HLA-B*27:05, HLA-B*27:09	IRAAPPPJ*F
FITC-GV9	HLA-A*02:01	GILGK^{FITC}VFTV
FITC-KK9	HLA-A*03:01	KLIEK^{FITC}-YFSK
FITC-RL9	HLA-B*27:05, HLA-B*27:09	RRKWK^{FITC}RWHL
FITC-IF9	HLA-B*27:05, HLA-B*27:09	IRAAK^{FITC}PPLF
FITC-RK9	HLA-B*27:05, HLA-B*27:09	RRKWK^{FITC}RWHK
KV9	HLA A*02:01	KILGFVFTV
KK9	HLA-A*03:01	KLIETYFSK
RL9	HLA-B*27:05, HLA-B*27:09	RRKWRRWHL
RA9	HLA-B*27:05, HLA-B*27:09	RRKWRRWHA
RK9	HLA-B*27:05, HLA-B*27:09	RRKWRRWHK
IF9	HLA-B*27:05, HLA-B*27:09	IRAAPPPLF

2.1.6 Cell lines

Table 2.3 Cell lines used in this thesis

Strain	Application
NEB® DH5α-alpha Competent *E. coli*	for subclone and plasmid amplification
BL21 (DE3) Competent *E. coli*	for protein expression (inclusion bodies of heavy chain and β2m)
DH10 Bac™(*E. coli*) (homemade)	for bacmid production with the pFastBac system
Sf9 (*S. frugiperda*)	for expanding viral titer and protein expression (ERp57-Tsn and TAPBPR mutants)

2.2 Molecular biological methods

2.2.1 Inclusion body expression and purification

Constructs of β2m and heavy chains of different allotypes were transformed into BL21 (DE3) competent cells and were cultured at 37 °C in LB medium or 2YT medium. Protein expression was induced with 1 mM IPTG when OD600 was around 0.8 and continued to culture at 37 °C overnight.

Induced cells were harvested after overnight expression by spin down at speed of 6000 g for 10 min. Cell pellets were resuspended with P1 lysis buffer and sonicated for 30 min on ice water. The sonicated

lysate was centrifuged at the speed of 12000 g for 10 min and the supernatant was trashed. The white pellet was the inclusion body. The inclusion body was first washed with lysis P2 buffer, sonicated for 10 mins and then centrifuged at the speed of 12000 g for 10 mins. This step was repeated twice with P2 buffer and 1 time with PBS buffer. In the last step, inclusion bodies were resuspended with P3 buffer, dissolved overnight while rotating, then centrifuged at the speed of 12000 g for 10 min to remove the unsolved particles. The supernatant is the desired denatured protein. The supernatant was then transferred into a new tube, shock froze and stored at -80 °C until use.

Table 2.4 Buffers used for inclusion body purification

Buffer	Composition
P1 buffer	50 mM Tris, 1% (w/v) sodium deoxycholate, 1% (w/v) TritonX-100, 100 mM NaCl, 0.1% (w/v) NaN3, 0.5 mM $MgCl_2$, 10 mM DTT*, 20 µg/mL DNase I*, pH 7.5
P2 buffer	50 mM Tris, 100 mM NaCl, 0.5% TritonX-100, 0.5 mM EDTA, 0.15 (w/v) NaN3, 10 mM DTT*, pH 8.0) 1X PBS pH 7.4
P3 buffer	20 mM Tris, 0.5 EDTA, 4M urea, 10 mM DTT*, pH 8.0

* These components need to be freshly added before use.

2.2.2 Expression of deuterated MHC-I heavy chains for NMR

The sequence of B*27:05/09 heavy chain for preparing NMR samples was the same as the BMRB Entry 25714/25713 (which has a His tag at the N-terminus). In order to confirm the assignments of the backbone amide of B*27:05/09, heavy chain of B*27:09/05 was isotope-labelled with ^{15}N and ^{13}C in a ^{2}H background. For 1 L deuterated M9 medium culture, 1 g $^{15}NH_4Cl$ and 1.5 g D-glucose-U-^{13}C6,1,2,3,4,5,6,6-d7 (Cambridge Isotope Laboratories, Inc. USA) were used. In order to investigate methyl-group dynamics of MHC-I molecules, Ile, Val and Leu methyl groups were ^{1}H-^{13}C labeled in $^{15}N/^{12}C/^{2}H$ background. For 1 L deuterated M9 medium culture, 100 mg precursor for Val/Leu (2-keto-(3-methyl-^{13}C)-butyric-4-^{13}C,3-D acid sodium salt, Santa Cruz Biotechnology, Inc), 50 mg precursor for Ile (2-ketobutyric acid-4-^{13}C,3,3-d2 sodium salt hydrate, Sigma-Aldrich, USA), 1.0 g $^{15}NH_4Cl$ and 1.5 g D-glucose-^{12}C6,1,2,3,4,5,6,6-d7 (Cambridge Isotope Laboratories, Inc. USA) were added[184].

For deuterated expression, freshly transformed BL21 (DE3) clones were first inoculated into 50%-D_2O M9 medium and cultured overnight. Cells were spun down at the speed of 3000 g for 5 min and washed twice each with 2 ml D_2O. Washed cells were resuspended and inoculated into 1 L deuterated M9 medium until the OD600 reaching 0.1 ~ 0.2. Expression was induced with 0.5 mM IPTG when the OD600 reached ~ 0.7, cultured in 30 °C overnight. Inclusion bodies were purified as described in section 2.2.1.

Table 2.5 1L deuterated M9 medium culture

Component	Amount
NH_4Cl	1g (either ^{15}N or ^{14}N)
Glucose	1.5g (either ^{12}C6, 1,2,3,4,5,6,6-d7 D-glucose or ^{13}C6, 1,2,3,4,5,6,6-d7 D-glucose)
100x trace element	10 ml
10x M9 salt buffer	100 ml

Thiamin	1.5 mg
Biotin	1.5 mg
$MgCl_2$	1 ml 1 M stock
$CaCl_2$	300 µl 1 M stock
Antibiotic	100 mg
D_2O	add to 1 L
Precursors	100 mg 2-keto-3-methyl-^{13}C-butyric-4-^{13}C-d acid sodium salt for Leu/Val, 50 mg 2-ketobutyric acid-4-^{13}C,3,3-d_2 sodium salt hydrate for Ile

All components for preparing deuterated M9 medium were dissolved in D_2O and filtered.

2.2.3 Refolding and purification of pMHC-I molecules

In order to obtain functional MHC-I molecules, heavy chain, β2m and the corresponding restricted high affinity peptide were refolded together in refolding buffer (0.4 M arginine, 0.1 M Tris, GSH:GSSG = 5 mM:0.5 mM, 2 mM EDTA, 0.5 mM PMSF, pH 8.0 for A*03:01 and A*02:01, pH 7.5 for B*27:05/09). For 1 L refolding, 10 mg peptide (photocleavable peptide photoKV9 for A*02:01 need first be dissolved in DMSO because it does not dissolve in water) was first added into pre-cooled refolding buffer, followed by dropwise adding 30 mg β2m. After stirring for 2 h at 4 °C, 30 mg heavy chain was dropwise added. The refolding bottle was then kept stirring at 4 °C (for photocleavable peptides, the refolding bottle need to be put in dark) until purification. The refolding time is different for different allotypes: 3 days for A*02:01, 5 days for A*03:01, 7 days for B*27:09 (with or without His-tag) and B*27:05 without His-tag, 14 days for B*27:05 with His-tag.

To purify the refolded MHC-I molecules, the refolding mixture was first filtered with 0.2 µm filter to remove precipitants/aggregations and then concentrated by viva flow device to a volume less than 100 ml. Concentration was continued with small concentrators until the volume less than 5 ml. The concentrated mixture was purified by size-exclusion chromatography (SEC) with a Superdex 200 10/300 increase column or HiLoad Superdex 200 pg preparative column pre-equilibrated with PBS buffer. Fractions were run on SDS-PAGE to check purity. Pure fractions were pool together. The T_m (melting temperature) of pooled protein was measured by Sypro-orange to evaluate the refolding quality. MHC-I refolded with photocleavable peptide is referred to as photo-pMHC-I.

2.2.4 Expression and purification of ERp57-Tsn/TAPBPR in insect cells

Constructs expressing ERp57-Tsn^{WT} and its mutants, TAPBPR and its mutants were transfected into Sf9 (*S. frugiperda*) insect cells according to the manufacturer's protocol (Bac-to-Bac® Baculovirus Expression System, Thermo Fisher) to generate virus stocks. Both ERp57-Tsn and TAPBPR proteins were expressed in cells but not secrete into the medium, which was verified by dot plot. 0.7 L Sf9 cells at the density of 1.5 - 2.0*10^6 cells/ml were infected with 40 ml P2 virus stock and harvested 4 days after transfection by spin down at the speed of 6000 g for 10 min. Cell pellets were washed with cold PBS buffer and stored at - 80 °C.

0.7 L cell pellets were resuspended with 40 ml cold lysis buffer on ice for 10 min and followed by sonication in ice water for 2 min. Cell debris was removed by centrifuging at the speed of 12000 g for 30 min. The lysate supernatant was incubated with 0.4 ml equilibrated nickel beads for 30 min (Ni-NTA

Agarose, Macherey-Nagel, Germany) while rotating at 4 °C. Nickel beads were spun down at the speed of 500 g for 5 min and then transferred into a 2 ml eppendorf tube. Nickel beads were washed with 1 ml equilibrium buffer 3 times and then with 1 ml wash buffer per time until the A280 absorption was lower than 0.2. Protein was eluted with 0.4 ml elution buffer each time until the A280 absorption below 0.1. Eluted fractions were pooled and concentrated to a volume less than 1 ml, then loaded on Biorad protein purification system which is installed with Superdex 200 10/300 size exclusion column, pre-equilibrated with PBS buffer. Fractions were run on non-reducing SDS-PAGE to check purity. Pure fractions were pooled, shock frozen and kept at - 80 °C until use.

Table 2.6 Buffers used for ERp57-Tsn conjugates proteins purification

Buffer	Components
Lysis buffer	100 mM KH_2PO_4, 600 mM KCl, 1% Triton, pH = 8.0 Freshly add protease inhibitor, DNase I and $MgCl_2$ to 0.5 mM *MMTS to final concentration 5 mM
Equilibrium buffer	100 mM KH_2PO_4, 600 mM KCl, pH=8.0
Wash buffer	20 mM imidazole ,100 mM KH_2PO_4, 600 mM KCl, pH=8.0
Elution buffer	150 mM imidazole ,100 mM KH_2PO_4, 600 mM KCl, pH=8.0

*MMTS: S-methyl mechanethiosulfonate (Sigma-Aldrich, Switzerland), only added when purify ERp57-Tsn protein and its mutants.

2.2.5 Fluorescence Polarization Assay of MHC-I to monitor peptide exchange

Fluorescence polarization (FP) measurements provide information of the molecular orientation and mobility, thus providing a way to measure the factors that modulate them, such as protein-protein interaction, protein-DNA interaction[185]. Most fluorescent molecules have the property of polarization. Experimentally, the degree of polarization is determined from measurements of fluorescence intensities parallel and perpendicular with respect to the plane of linearly polarized excitation light, and is expressed in terms of fluorescence polarization (P) or anisotropy (r): P= (I-I90°)/(I+I90°), where I is the intensity parallel to the excitation plane and the I90° is the intensity perpendicular to the excitation plane.

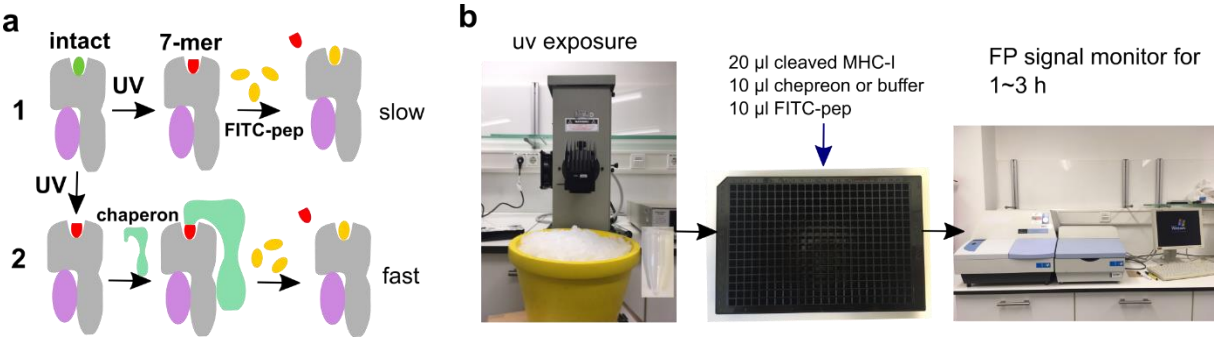


Figure 2.1 Principle of peptide exchange assays monitored by FP. a) Cartoons showing the peptide exchange of UV-cleaved MHC-I molecules in the presence or absence of the chaperone protein. **b)** Workflow of peptide exchange assays. The UV-exposure of photo-MHC-I molecules were done in the 1.5 ml epi on ice under the UV-lamp.

In this thesis, fluorescence polarization assays were performed with MHC-I molecules refolded with photocleavable peptide (or normal medium affinity peptide) with/without UV-exposure exchanging against FITC-labelled version of the same high affinity peptides. The principle of the assay is shown in Fig. 2.1a: UV-exposure cleaves photocleavable peptides into two fragments, 7-mer and dimer; the

affinity of both fragments to the binding groove is low, so that they can be replaced by the high-affinity FITC-peptide that can bind to the MHC-I molecules, resulting in an increase in FP signal. In the presence of chaperones, the FP signal (the peptide exchange of MHC-I) is expected to increase faster.

Peptide exchange assays of MHC-I were done with UV-exposed photo-MHC-I exchanging against FITC-labelled high affinity peptides in the absence or presence of ERp57-Tsn^{WT}/TAPBPR^{WT}/their mutants. Briefly (Fig. 2.1b): 20 µl of 1µM purified photo-MHC-I (diluted in PBS 7.0) was UV exposed (arc-UV lamp, Newport 6295NS, filter 345 nm, 1000 W) on ice for 5 min at a distance of 15 cm and then added to the 384-well assay plate (Corning® 384-well Low Flange Black Flat Bottom Polystyrene NBS Microplate 3575, USA). 10 µl * 2 µM ERp57-Tsn^{WT}/TAPBPR^{WT}/their mutants or 10µl PBS buffer were added and then 10 µl * 40 nM (400 nM for FITC-RL9) of corresponding FITC-peptide was added. Wells containing only FITC-peptides (a final concentration 10 nM for FITC-IF9, FITC-KK9, FITC-KV9, FITC-RK9, and 100 nM for FITC-RL9) were measured as baseline. Each experiment has a total volume of 40 µl and was performed at room temperature with three replicates. All proteins used in the assay were purified in the same buffer (PBS). FP signals were recorded on PerkinElmer VICTOR^{3}V with an excitation filter of λ_{ex} = 485 nm and an emission filter of λ_{em} = 535 nm. For kinetic measurements, the FP signals were monitored right away after adding FITC-peptide over a time range from 1 to 3 h. For statistic measurements, plates were incubated overnight at room temperature and then measured. FITC-labelled high affinity peptides used in this study are GILG-Lys(FITC)-VFTV for A*02:01, KLIE-Lys(FITC)-YFSK for A*03:01, RRKW-Lys(FITC)-RWHL for B*27:05/09, and IRAA-Lys(FITC)-PPLF for B*27:09/05.

Experimental data analysis was processed with Origin 2019 and *Kon* of association curve was obtained by using fitting function $Y = Y_0 - Ae^{-kt}$, where Y_0 is the maximum value and A is the increased FP range.

2.2.6 The thermofluor assay

The principle of the thermofluor assay is that the dye, such as Sypro Orange, binds to hydrophobic cores of protein when the protein unfolds. In this thesis, Sypro-orange dye was used to determine the melting temperature T_m of proteins. Sypro-orange nonspecifically binds to hydrophobic surfaces when proteins unfold, resulting in an increase in fluorescence signal, while water strongly quenches its fluorescence. 20 µl protein (0.2 ~ 0.4 mg/ml) was mixed with 5 µl Sypro-orange diluted buffer (243.5 µl buffer + 1.5 µl Sypro-orange, Invitrogen, USA), then added to the MicroAmpROptical 96-well Reaction Plate (Life Technologies), measured on the qPCR machine (Stratagene Mx3005PTM, ThemoFisher) using the protocol for heating at the speed of 2°C/min from 25 to 90 °C. Each sample had three replicates. Experimental data analysis was processed with Origin 2019 and the T_m was obtained by fitting with the Boltzmann function.

2.2.7 Nuclear magnetic resonance spectroscopy NMR

2.2.7.1 Basics of NMR

Nuclear magnetic resonance (NMR) is a powerful tool to study protein structure and dynamics in solution. It uses the magnetic spin properties of atomic nuclei within a protein to identify interactions

between atoms that are close in space (either because of covalent bonds or via folding). Only certain isotopes, such as ^{1}H, ^{13}C and ^{15}N or ^{31}P, have magnetic spin. Protein samples can be isotope-labelled by feeding bacteria with ^{15}N ($^{15}NH_4Cl$) and/or ^{13}C (^{13}C-glucose) enriched compounds in the culture. The isotope-labelled protein sample is then placed in a magnetic field B_0, which aligns with the z-axis. Spins of NMR-active nuclei will algin with or against the field B_0 (a lower energy α state and a higher energy β state) and precess around the z-axis at a constant angle with their Larmor frequency ω_0 (Fig. 2.2a), which is calculated by equation $\omega_0 = \gamma B_0$, where γ is the gyromagnetic ratio which is specific for every nucleus. The energy difference between α state and β state is calculated by the equation: $\Delta E = \hbar\omega_0$, where ħ is h/2π, h is the Planck constant. At equilibrium, the spins in α state is slightly more populated than the β state, leaving a net magnetization along +z axis (Fig. 2.2a). Due to the small population difference between α and β state, NMR measurements are insensitive and need high sample concentrations.

The ω_0 frequency is dependent on the type of nucleus and also, to a small extent, the chemical environment of the nuclei. This variation is expressed by the chemical shift (δ) in ppm (parts per million), which represents the detected frequency normalized to a reference frequency. The difference to the reference frequency can also be illustrated by the rotating frame, assuming that the x- and y-axes rotate about the z-axis with the reference frequency (ω_0). In order to measure the procession rate ω_0 (the Larmor frequency), the system needs be nudged away from equilibrium. A radiofrequency pulse (p1 pulse, B1-field) orthogonal to the z-axis is applied to turn the net magnetization of the sample by 90° to the xy-plane. The length of this p1 pulse needs to be determined for every sample. After the pulse, spins will return to the ground state by relaxation, thereby precessing around the z-axis with their ω_0 frequency and subsequent measurement of the precession rate and the chemical shift, to produce an NMR spectrum (Fig. 2.2b). This precession is detected by a receiver coil, giving rise to an exponentially decaying cos-signal (the FID, free induction decay). The relaxation occurs fast or slow depending on the size and tumbling of the protein in solution. Small proteins tumble fast and the signal decays slowly, leading to sharp and narrow signals after Fourier Transformation. Big proteins tumble slowly and produce broad peaks that often overlap with each other explaining the size limitation of NMR spectroscopy of proteins.

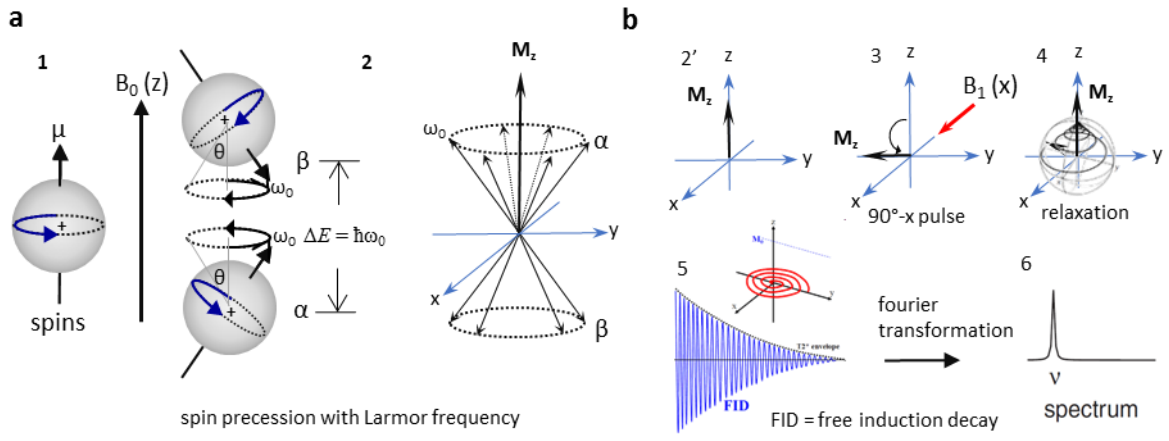


Figure 2.2 Principle of one-dimensional NMR. a) 1: a nucleus spins about an axis. Resulting circulation creates a magnetic moment μ. 2: Applying a B_0 filed aligned along +z-axis results in spins precessing about the z-axis both up and down at a certain angle θ with Lamor frequency ω_0. All xy-plane components cancel, leaving a net magnetization M_z aligned with +z

axis. **b)** 2'-3: Rotating the M_z to the xy-plane by applying a 90° B_1 pulse. 4-6: Detecting the signal after the pulse is removed and spins return to the ground state by relaxation. Pictures were modified from the book "NMR Primer: An HSQC-Based Approach"[186].

Once the B_1 pulse is removed, the net magnetization will begin to return to +z-axis. The return to the equilibrium is called relaxation, which can be divided into two categories. The rate of the net magnetization along +z recovering to M_{z0} (the equilibrium) is called R_1 and is characterized by a time constant T_1. Similarly, the rate of the xy-plane net magnetization decays to zero is called R_2 and is characterized by a time constant T_2. T1 and T2 are also described as the longitudinal and transverse relaxation times, respectively. Due to any factor that could cause signal decay, such as field inhomogeneities, the measurement of T2 is not as simple. The measuring of "true" T2 is initiated with rotating the equilibrium magnetization to the xy plane by 90°, followed by repeats of short delay τ-180-τ pulse to suppress the uniformity of the spectrometer. This τ-180-τ period is called a Carr-Purcell-Meiboom-Gill (CPMG) sequence. The rephasing at the end of each τ-180-τ is called a "spin echo".

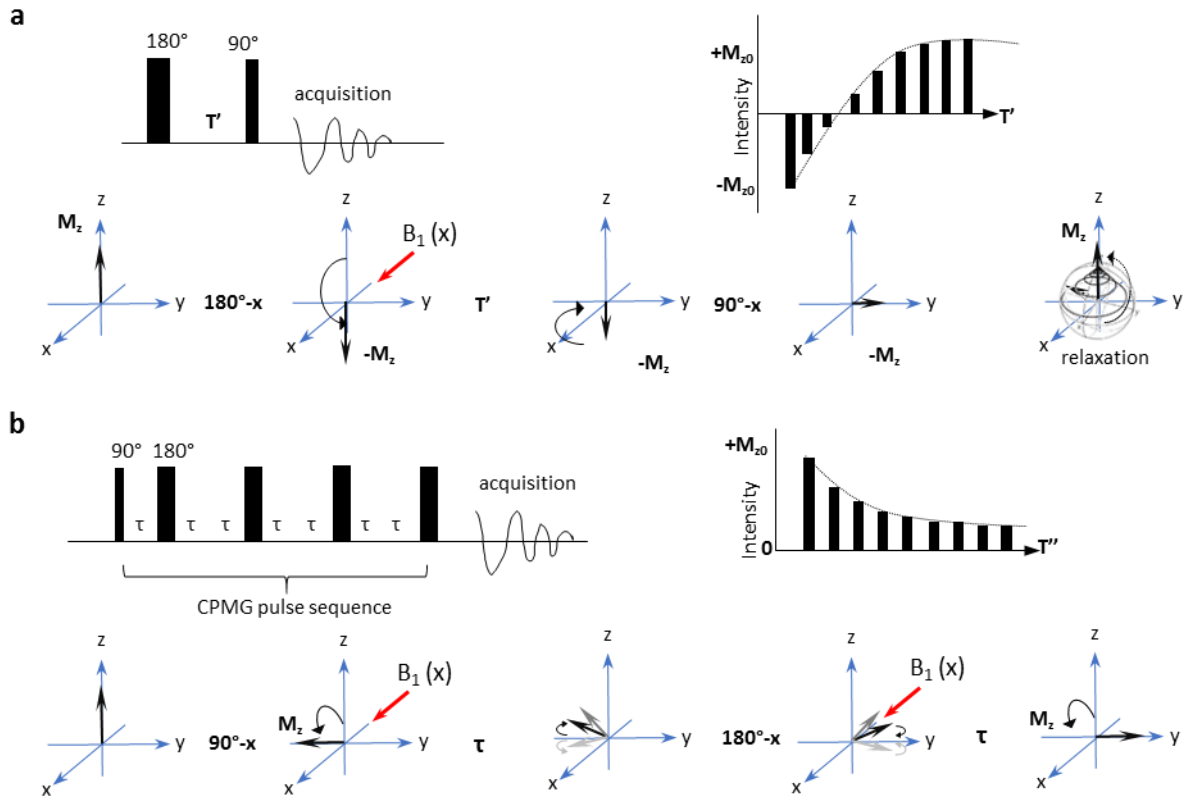


Figure 2.3 T1 and T2 relaxation. a) Measuring T1: Pulse sequence, with variable delay T', used to measure T1. The net magnetization is rotated to align with -z-axis with 180° pulse. After T', 90°x pulse is used to rotate the Mz to the xy-plane and detect the signal. **b)** Measuring T2 relaxation: Pulse sequence, with variable delay T'', used to measure T2. The net magnetization is rotated to xy-plane with a 90°x pulse, then followed by a series of τ-180-τ spin echo elements.

2.2.7.2 Multi-dimensional NMR

As proteins contain many protons, hydrogen spectra are usually very crowded and hard to interpret. The only useful information gained from it is to tell if the protein is folded. To overcome this problem, two-dimensional or more dimensional NMR are introduced, which allow plotting correlation of different nuclei, such ^{1}H-^{1}H, ^{1}H-^{15}N, ^{1}H-^{13}C. For example, by the transfer of magnetization from H- to N-atoms in the context of a complex sequence of different pulses and delay times, a two-dimensional ^{1}H-^{15}N-

HSQC (Heteronuclear Single Quantum Correlation) spectrum is recorded (Fig. 2.4a). In a standard HSQC spectrum, each peak specifically corresponds to a NH-group in the protein, which serves as a fingerprint of the protein and contains structural information, while the peak intensity or shape/width contains information about the molecular weight and dynamics. These spectra are suitable for the analysis of protein-ligand interactions, as peaks corresponding to amide-groups close to the ligand will shift or attenuate upon binding.

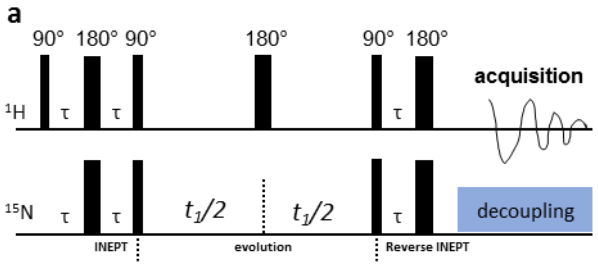


Figure 2.4 Principle of 2D NMR. a) decoupled INEPT-^{1}H-^{15}N-HSQC pulse sequence. INEPT: Insensitive Nucleus Enhanced by Polarization Transfer.

In this thesis we measured the ^{1}H-^{15}N HSQC or ^{1}H-^{13}C HSQC of heavy chain ^{15}N-labeled MHC-I molecules refolded with normal peptide and photocleavable peptide. TROSY (Transversal Relaxation Optimized Spectroscopy)-HSQC was used to measure deuterated samples to enhance the resolution of the spectra. All NMR spectra (unless otherwise stated) were acquired on a Bruker Avance III 700 MHz spectrometer equipped with a 5 mm triple resonance cryoprobe. Spectra were processed with TopSpin3.2 (Bruker, Billerica, USA) and analyzed with CcpNmr Analysis 2.4.2[41]. All NMR samples were measured in PBS 7.2 buffer with 10% D_2O at 300K (unless otherwise stated).

2.2.7.3 NMR Chemical Shift Assignments

A three-dimensional NMR measurement correlates three nuclei: combination includes ^{1}H-^{1}H-^{13}C;^{1}H-^{1}H-^{15}N or ^{1}H-^{13}C-^{15}N. Three-dimensional spectra correlating nuclei via covalent bonds (as for example in the HNCA, HN(CA)CO spectra) are usually used for assignments and are displayed as two-dimensional strips. Nuclear Overhauser Effect (NOE) is the dipole-dipole interaction of spins through space. The effect restraints within 5 Å (distance between two spins). The principle of NOESY (NOE Spectroscopy) is based on the NOE effects and NOESY-HSQC can also be used to confirm assignments or to retrieve distance information through space.

The backbone amide assignments of B*27:09 heavy chain were transferred from available assignments for B*27:09/RL9 (BMRB Entry 25713[42]) and were confirmed with HN(CO)CA, and ^{1}H-^{15}N-TROSY-HSQC spectra of ^{15}N-^{13}C-^{2}H B*27:09/RL9 (270 μM). ^{1}H-^{15}N-TROSY-HSQC spectrum of ^{2}H-^{15}N B*27:09/photoRL9 was recorded at a concentration of 250 μM and the assignments were transferred from the confirmed assignments of ^{15}N-^{13}C-^{2}H B*27:09/RL9 (Appendix Fig. 1). The backbone amide assignments of B*27:05 heavy chain were transferred from available assignments for B*27:05/RL9 (BMRB Entry 25714[42]) and were confirmed with HNCA, HNCO and ^{1}H-^{15}N-TROSY-HSQC spectra of 220 μM ^{15}N-^{13}C-^{2}H B*27:05/photoRL9 (Appendix Fig. 2).

ILV methyl assignments of B*27:09 were transferred from available assignments for B*27:09/RL9 (BMRB Entry 25713[42]). These assignments as well as those available for A*02:01 (BMRB Entry 27631) and A*01:01 (BMRB Entry 27632) were used to transfer assignments to other complexes used in this publication[30,31]. Transferred assignments were validated where possible by methyl-methyl and methyl-amide NOEs with 170 µm ^{1}H-^{13}C-ILV-^{15}N-^{2}H B*27:05/photoRL9 in combination with the crystal structure of B*27:05/RL9 (PDB ID 1OGT). The ^{1}H-^{13}C HSQC spectra of B*27:09/RL9, B*27:09/photoRL9 and B*27:05/photoRL9 were shown in Appendix Fig. 3.

ILV methyl assignments of A*03:01 was transferred from those available for A*02:01 (BMRB Entry 27631) and A*01:01 (BMRB Entry 27632) and confirmed by methyl-methyl NOEs with ^{1}H-^{13}C-ILV-^{15}N-^{2}H 170 µM A*03:01/photoKK9 (Appendix Fig. 4). Due to the long measuring time, several NMR experiments in a row is impossible because the sample starts degrading after long time at room temperature (Appendix Fig. 5a).

2.2.7.4 HSQC spectra of MHC-I before/after UV-exposure and in the presence of ERp57-Tsn or TAPBPR

In order to study how UV-exposure affects MHC-I molecules, ^{1}H-^{15}N-TROSY-HSQC spectra of 200-250 µM ^{1}H-^{13}C-ILV-^{15}N-^{2}H B*27:09 or 05/photoRL9 were recorded before and after UV-exposure (Appendix Fig. 6). ^{1}H-^{13}C-HSQC spectra of 200-250 µM ^{1}H-^{13}C-ILV-^{15}N-^{2}H B*27:09 or 05/photoRL9 were also recorded before and after UV-exposure (Appendix Fig. 7). Peak intensity ratios (I/I0) were calculated from spectra before (I0) and after (I) UV-exposure. Residues were considered as significant line broadening if its I/I0 was smaller than the average I/I0 minus standard deviation of all I/I0. Chemical shift changes were calculated according to the formula $\Delta\delta^{NH} = \sqrt{(\delta 1H)^2 + (0.15 \cdot \delta 15N)^2}$, chemical shift differences $\Delta\delta^{NH}$ were considered significant if it was larger than the average $\Delta\delta^{NH}$plus standard deviation of all $\Delta\delta^{NH}$.

In order to study how B*27:09 would be affected by tapasin/TAPBPR in solution, ^{1}H-^{15}N-TROSY-HSQC spectra of 50 µM ^{1}H-^{13}C-ILV-^{15}N-^{2}H B*27:09/photoRL9 in the absence (I0) and in the presence (I) of 50 µM un-labelled ERp57-Tsn variants/TAPPBR purified in matched buffer (PBS 7.2) were recorded (Appendix Fig. 8a) and peak intensity ratios (I/I0) were calculated.

In order to study how B*27:05 would be affected by tapasin in solution, ^{1}H-^{15}N-TROSY-HSQC spectra of 30 µM ^{1}H-^{13}C-ILV-^{15}N-^{2}H B*27:05/photoRL9 in the absence (I0) and in the presence (I) of 27 µM un-labelled ERp57-Tsn^{WT} purified in matched buffer (PBS 7.2) were recorded (Appendix Fig. 8b) and peak intensity ratios (I/I0) were calculated.

2.2.7.5 CPMG Relaxation Dispersion

The principle of Carr-Purcell-Meiboom-Gill (CPMG) relaxation dispersion is detect the contribution of conformational exchange to the R_2 relaxation rate. This is achieved by applying a train of τ-180- τ pulses (Fig. 2.3b). In contrast to the R_2 relaxation rate measurement, where a constant frequency of CMPG pulses is applied for a variable time, here it is applied with a variable frequency (ν_{CPMG}) for a constant time (40 ms). In the presence of conformational exchange R_{ex} contributes to $R_{2,0}$ and a reduced peak intensity is observed at low frequency of the CPMG pulse train.

Biofunctions such as enzyme catalysis, protein refolding, ligand binding and protein isomerization always go along with conformational exchange. CPMG relaxation dispersion is a powerful NMR experiment to study protein conformational exchange on millisecond to microsecond time scales. Normal HSQC NMR can only detect the major state of the protein or highly populated minor states, while CPMG can let us see the "invisible" state, as long as the minor state is populated to greater than approximately 0.5%. By applying CPMG pulses on a sample, the transverse relaxation rates (R_2^{eff}) of probes such as ^{15}N, ^{1}H or ^{13}C nuclei are monitored as a function of the frequency, υCPMG, at which refocusing 180° pulses are applied during a relaxation delay. Line broadening can be reduced by refocusing and at the same time quenches the intensity of signal if the conformational exchange is happening. A typical curve of a probe undergoing dynamics on the micro- to millisecond timescale is R_2^{eff} decays as υCPMG increases. In the absence of conformational exchange, a flat line is observed. If there is exchange between two states (major state and minor state, or also described as active state and inactive state when it comes to the enzyme activity) can be detected by CPMG. It provides thermodynamic, kinetic, structural, and mechanistic information on systems that undergo conformational exchange.

In this thesis, methyl groups of Ile, Val and Leu were labeled with $^{13}CH_3$. Methyl side chain relaxation rates were determined using ^{1}H-^{13}C multiple quantum CPMG relaxation dispersion experiments. CPMG relaxation dispersion experiments of ^{1}H-^{13}C-ILV-^{15}N-^{2}H B*27:09/RL9(170 µM), B*27:09/photoRL9 (165 µM), B*27:05/photoRL9 (200 µM), A*03:01/photoKK9 (170 µM) were recorded at 300 K using a Bruker Avance 700 MHz spectrometer. Spectra were recorded single-scan interleaved using an INEPT transfer for excitation[46] and a WATERGATRE element for water suppression in order to measure in 10% D_2O. CPMG pulse frequencies of 0, 950, 50, 300, 150, 800, 100, 250, 400, 600, 200, 50, 500, 300, 700 with a constant time delay of 40 ms were used. Peak intensities were converted to the R_2^{eff} transverse decay rates with the equation $R_2^{eff} = 1/T_{CPMG} \times \ln (I0/Iv_{CPMG})$. Only assigned non-overlapping peaks were analyzed. CPMG profiles of methyl groups displaying dispersion ($R_{ex} > 2\ s^{-1}$) were fitted with a two-state model using the program NESSY[47].

2.2.8 Crystallography

2.2.8.1 Crystallization

Sample (B*27:09/photoRL9) for crystallography was purified two times by SEC in 20 mM Tris/HCl pH 7.4, 150 mM NaCl, 0.01% sodium azide and concentrated to 13-15 mg/ml as measured by the absorbance at 280 nm. Crystals were obtained by the sitting-drop vapor-diffusion method at 18 °C with a reservoir solution composed of 20% (w/v) polyethylene glycol 3350, and 200 mM KSCN. Crystal growth and diffraction quality was improved by micro-seeding. Prior to flash cooling of the crystals in liquid nitrogen, glycerol was added to the mother liquor to a final concentration of 25% (w/v).

2.2.8.2 X-ray data collection, structure determination and refinement

Data collection and structure determination were performed by Dr. Bernhard Loll. Synchrotron diffraction data on crystals of B*27:09/photoRL9 were collected at the beamline 14.1 of the MX Joint Berlin laboratory at BESSY II (Berlin, Germany). X-ray data collection was performed at 100 K. Diffraction data were indexed and processed with the XDS. Data collection and refinement statistics are

given in Appendix Table 1. The phase problem was solved by molecular replacement using the program PHASER[187] with the structure of HLA-B*27:09/pVIPR as search model (PDB ID 5IB5[188]). The structure was initially refined by applying a simulated annealing protocol and in later refinement cycles by maximum-likelihood restrained refinement using in PHENIX[189,190]. Model building and water picking was performed with COOT[191]. Intermediate and final structures were evaluated with and MOLPROBITY[192] and the JCSG validation server (JCSG Quality Control Check v3.1). Figures were prepared with PyMOL (Version 1.8 Schrödinger, LLC). Atomic coordinates and structure factor amplitudes have been deposited in the Protein Data Bank under accession code 7ALO.

2.2.9 Mass-spectrometric analysis

To confirm the correct expression of desired proteins, gel bands containing the expected protein were cut and digested with trypsin. LC-MS of treated samples were measured by Dr. Benno Kuropka.

In order to detect fragments of the photo-cleavable peptide, UV-exposed photo-MHC-I was purified by SEC to remove any possible fragment and fractions were analysed by matrix-assisted laser desorption ionization-time of flight mass spectrometry (MALDI-TOF-MS) using α-cyano-4-hydroxycinnamic acid (HCCA) as matrix. The samples were applied on the MALDI target using the dried-droplet technique. Spectra were recorded on an Ultraflex-II TOF/TOF instrument (Bruker Daltonics, Bremen, Germany) in the positive reflector mode in the m/z range of 600-4,000. Samples of photo-cleavable peptide before and after UV-exposure served as a reference.

2.2.10 Modelling of the tapasin scoop loop and its mutants

The modelling was performed by Dr. Esam T. Abualrous. The EM structure of the PLC (PDB ID 6ENY) served as a scaffold to build the MHC-I-ERp57-Tsn models using the crystal structures of MHC-I allotypes HLA-A2:01 (PDB ID 1OGA), HLA-A3:01 (PDB ID 2XPG), HLA-B27:05 (PDB ID 1OGT), HLA-B27:09 (PDB ID 1OF2), and the ERp57-Tsn complex (PDB ID 3F8U). The unresolved atoms in the tapasin scoop loop were modelled and refined with the MODELLER software[193,194] using the corresponding sequence to generate an ensemble of 100 models for each allotype. With the Amber 12 simulation package, each MHC-I-ERp57-Tsn model was placed in an octahedral TiP3[195] water box and neutralized with counterions. Then, each complex was energy minimized, positionally restrained (25 kcal $mol^{-1}Å^{-2}$), and heated from 100 to 300 K. The restraints were resolved in five steps, and then each complex was equilibrated for 2 ns. Short-range nonbonded interactions were taken into account up to a cut-off value of 9 Å.

The structural modeling of the tapasin mutants was performed with PyRosetta using the “ref2015” energy function[196–198]. The desired scoop loop sequence was computationally introduced into tapasin, followed by Monte Carlo-based simulated annealing sidechain and backbone minimization steps. The energy scores were calculated with the “score_jd2” application and averaged for all generated models per mutant.

2.2.11 Sequence analysis

Tapasin sequences and TAPBPR sequences of different species were downloaded from Uniprot and were aligned in https://www.uniprot.org/align/. Peptides presented by different species were downloaded from IEDB by selecting linear epitope, MHC-I restriction for the corresponding host and further filtered by length between 8-11 aa in excel. Peptide motif sequence logo of an allotype was plotted from peptides by the corresponding allotype.

MHCI allele sequences of different species were downloaded from https://www.ebi.ac.uk/ipd/imgt/hla/. MHC-I sequences of mouse which is missing in this data base were collected from Uniprot. MHC-I sequence data were tidied in R and sequence logo of F-pocket were generated by https://weblogo.berkeley.edu/logo.cgi.

2.2.12 Cryo-EM sample preparation and single data acquisition.

In order to gain insight in the interaction between MHC-I and tapasin, the MHC-I-ERp57-Tsn complex was prepared by incubating ERp57-Tsn$^{G304C-X}$ (X stands for the mutation in the scoop loop) with 1.5 equimolar UV-cleaved photo-pMHC-I/β2m^{T5C} overnight in the redox buffer (GSH : GSSH = 1 mM : 10 mM, pH 7.5). The mixture was first purified by ion exchange (20 mM HEPEs, with gradient 0 - 0.5 M NaCl over four column volumes) and then by gel filtration with Superdex 200 10/300 increase column in HEPEs buffer (20 mM HEPEs, 150 mM NaCl, pH 7.4).

Cryo-EM data was measured and processed by Dr. Tarek Hilal at the BioSupraMol facility of the Freie Universität Berlin. Quality of purified complex samples was first checked by negative stain at the concentration of 20, 50, 100 µg/ml. For negative stain grids preparation, 3 µl of pure fraction sample was applied to the freshly glow-discharged carbon-coated grids (Quantifoil R1/4 Micromachined Holey Carbon Grids) and stained by 2% uranyl formate solution. After 40 s incubation, the remaining uranyl formate was removed by paper. Grids were dried at RT overnight. Images of samples were recorded on a FEI Talos L120C transmission electron microscope (TEM), operated at 120 kV, equipped with a FEI CETA 16 M CCD camera. Images were collected manually in low dose mode using TEM Imaging & Analysis (TIA) software. Images analysis was carried out with cryoSPARC. Crosslinked samples were also checked by negative stain. Sample was treated with 0.025% glutaraldehyde solution for 20 min at RT and the reaction was quenched by Tris buffer. The crosslinked sample was directly applied on grid for negative stain.

For cryo-EM grids preparation, 3µl of pure fraction sample at the concentration of 0.2 mg/ml (without detergent) or 2~5 mg/ml (with detergent, detergent was mixed freshly right before applying the sample to the grid) was applied to plasma-treated carbon grids and plunged into liquid ethane using Vitrobot Mark IV (Thermo Fisher, former FEI). Images were acquired on Titan Krios (300 kV) TEM with a Falcon 3EC camera. Acquired images were processed and analyzed by Dr. Tarek Hilal.

3 Results

3.1 Characterization of MHC-I in complex with a photocleavable peptide

To better understand the peptide-editing process of MHC-I, we first investigated the conformational and dynamic properties of peptide-bound MHC-I, and then followed structural elements undergoing change upon peptide cleavage. The routine method to generate peptide-free MHC-I molecules is to refold heavy chains and β2m with conditional ligands together, such as photocleavable high-affinity peptides which can be cleaved by UV-exposure[181]. In this chapter, different biophysical and biochemical methods were used to characterize the properties of MHC-I refolded with photocleavable peptides. The allotypes HLA-B*27:09 and HLA-B*27:05 were chosen as study models. They differ only by one residue in position 116 of the heavy chain (His in B*27:09, Asp in B*27:05). B*27:05 but not B*27:09 is associated with the autoimmune disease ankylosing spondylitis (AS)[152,199]. Their thermostability, peptide exchange and dynamics in solution were compared in this chapter.

3.1.1 Comparison of the crystal structure of MHC-I in complex with photocleavable and non-cleavable peptide

Photo-pMHC-I was refolded from *E.Coli*-expressed heavy chain, β2m and photocleavable peptide and purified as described in section 2.2.3. The photocleavable peptide used for refolding is RRKWRRWJ*L (named as photoRL9, photocleavable site was introduced at the 8th amino acid), which is derived from a self-peptide pVIPR (RRKWRRWHL, from vasoactive intestinal peptide type 1 receptor). Introducing photocleavable amino acids might reduce the affinity of the peptide with MHC-I molecules since the photocleavable amino acid is a β amino acid, the backbone of which might not fit as well into the binding groove as an α amino acid. Indeed, the refolding yield and the melting temperature (T_m) of photo-pMHC-I was lower than that of MHC-I in complex with the corresponding non-cleavable peptide (Fig. 3.1.1a , 54.55 ± 0.14 °C for B*27:09/photoRL9 vs 60.2 ± 0.15 °C for B*27:09/RL9).

In order to address the potential correlation between the structure and thermostability of MHC-I in complex with the photocleavable peptide and in complex with the non-cleavable peptide (i.e., B*27:09/photoRL9 vs B*27:09/RL9), the crystal structure of B*27:09/photoRL9 was solved (Appendix Table 1, resolution is 1.80 Å, the crystal structure was solved by Dr. Bernhard Loll). The overall structure of B*27:09/photoRL9 shows the typical hallmarks of a stable pMHC-I complex and is nearly identical to the previously published structure of B*27:09/RL9[26] (Fig. 3.1.1b). In both structures, H-bonds between residues in A, B and F-pockets and *p*R1 (*p*R1 stands for residue R1 of the peptide), *p*R2 and the C-terminal *p*L9 were observed, respectively (Table 3.1). However, the backbone of photoRL9 from P4-P8 is different from RL9 due to the introduction of the photocleavable amino acid at P8 (Fig. 3.1.1b and 3.1.1d), resulting in the absence of three H-bonds, the formation of one new H-bond (Fig. 3.1.1c, Table 3.1) and the re-orientation of several residues' side chain in the heavy chain (E76, K146, R151, Q156 and R157, Fig. 3.1.1d). All these differences were also reflected in chemical shift changes in the ^{1}H-^{15}N TROSY-HSQC spectrum of B*27:09/photoRL9 compared to B*27:09/RL9, whereas the resonances of the α3 domain of B*27:09/photoRL9 superimposed well with those of B*27:09/RL9

(Appendix Fig. 1). Less H-bonds between the bound peptide and the binding groove might explain why the refolding yield and the T_m of B*27:09/photoRL9 was lower than that of B*27:09/RL9.

It has been reported that the RL9 peptide adopts two conformations (i.e., canonical and non-canonical) in the crystal structure of B*27:05/RL9[26]. In the canonical conformation *p*R5 points up, while in the non-canonical conformation *p*R5 points towards the binding groove and forms an H-bond with D116 (Fig 3.1.1e)[26]. However, in the crystal structure of B*27:09/RL9, the RL9 peptide has been shown to adopt only the canonical conformation. It has been further shown that the RL9 peptide in B*27:05/RL9 is presented only in the non-canonical conformation in solution[200]. The binding of metal ions to *p*H8 allows the RL peptide to adopt both conformations in the binding groove of B*27:05[200]. In contrast, no similar conformational change was found for RL9 peptide in complex with B*27:09. Interestingly, in the crystal structure of B*27:09/photoRL9, the photoRL9 adopts a non-canonical conformation in the binding groove, with the side chain of *p*R5 forming an H-bond with D77 (Fig. 3.1.1f). This data suggests that the formation of the non-canonical conformation is not only dependent on the subtype but also on the bound peptide sequence. It is conceivable that mutation at P8 in the self-peptide RL9 might lead to the conformation which can be recognized by the TCR, thereby causing autoimmune related disease[201,202].

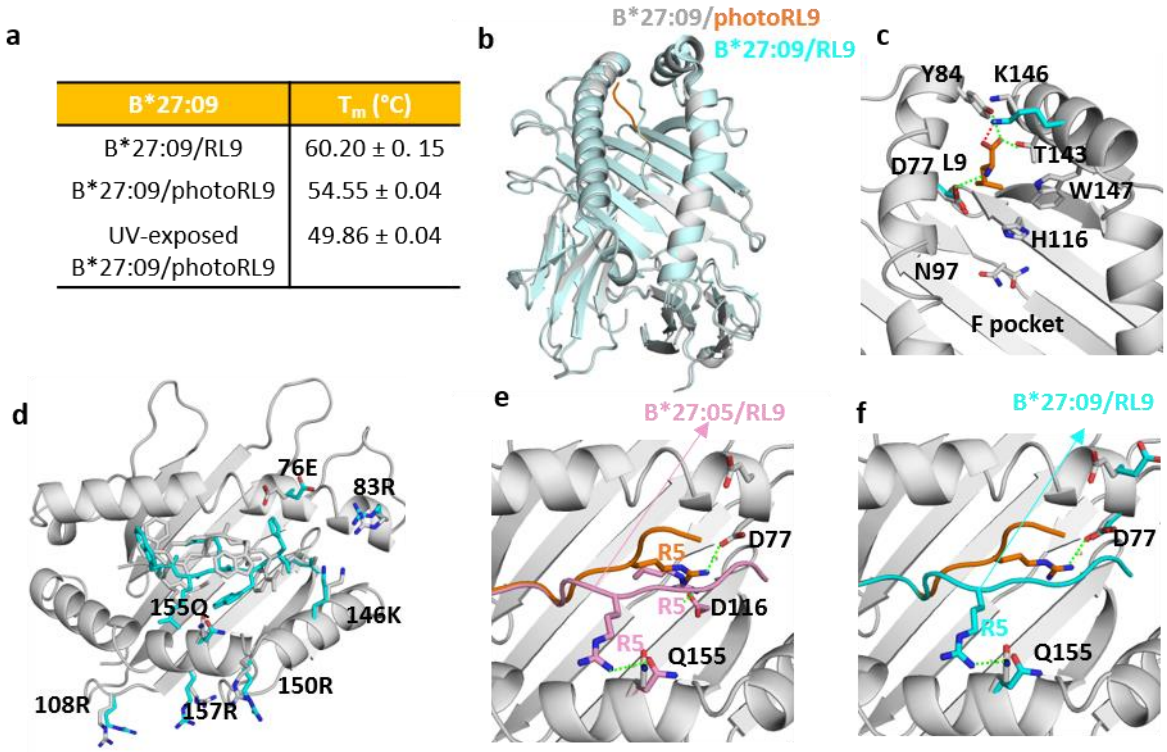


Figure 3.1.1 Comparison of crystal structure of B*27:09/photoRL9 with B*27:09/RL9. **a)** T_m of B*27:09/RL9, B*27:09/photoRL9 and UV-exposed B*27:09/photoRL9 measured by the thermofluor assay using Sypro-orange. **b)** Crystal structure of B*27:09/photoRL9 (gray, PDB 7ALO) and B*27:09/RL9 (cyan, PDB IOF2), superimposed based on the heavy chain. PhotoRL9 (orange) is shown to address the orientation different from RL9. **c)** Close view of the F-pocket region with H-bonds (green dashed line) in the structure of B*27:09/photoRL9. The missing H-bond in B*27:09/photoRL9 is colored red. Residues (K147, D77) in B*27:09/RL showing different side chain orientation from B*27:09/photoRL9 are colored cyan. *p*L9 is shown in orange. **d)** Top view of the binding groove with the bound peptide shown in stick in the structure of B*27:09/photoRL9 (gray) and B*27:09/RL9 (cyan). Residues showing different side chain orientation between two structures are highlighted as sticks. **e)** *p*R5 (pink) binds to B*27:05 in both canonical and non-canonical conformations (1GOT). The side

chain of *p*R5 forms an H-bond with D116 in B*27:05/RL9 in the non-canonical conformation. **f)** PhotoRL9 (orange) binds to B*27:09 in the non-canonical conformation with *p*R5 forming an H-bond with D77, while RL9 (cyan) binds to B*27:09 in the canonical conformation with *p*R5 forming an H-bond with Q155.

Table 3.1 Detailing polar contacts within 3.5 Å between residues of B*27:09 heavy chain and the bound peptide in the crystal structure of B*27:09/RL9 and B*27:09/photoRL9

Pocket	B*27:09/RL9 (IOF2)	B*27:09/photoRL9 (7ALO)
A pocket	Tyr7-OH with Arg1-N	Tyr7-OH with Arg1-N
	Tyr171-OH with Arg1-N	Tyr171-OH with Arg1-N
	Tyr159-OH with Arg1-O	Tyr159-OH with Arg1-O
	Glu163-Oε1 with Arg1-Nη1	Glu163-Oε1 with Arg1-Nη1
B pocket	Tyr24-OH with Arg2-Nη1	Tyr24-OH with Arg2-Nη1
	Glu45-Oε1 with Arg2-Nη1	Glu45-Oε1 with Arg2-Nη1
	Glu45-Oε1 with Arg2-Nε1	Glu45-Oε1 with Arg2-Nε1
	Arg62- Nη1 with Arg2-O	Arg62- Nη1 with Arg2-O
	Glu63-Oε1 with Arg2-N	Glu63-Oε1 with Arg2-N
	Glu63-Oε1 with Arg2-Nε	Glu63-Oε1 with Arg2-Nε
	Tyr99-OH with Lys3-O	Tyr99-OH with Lys3-O
D pocket	Gln155-Oε1 with Arg5-Nη1	**Not observed**
	Gln155-Oε1 with Trp7-Nε1	**Not observed**
F-pocket	Asp77-Oε1with Leu9-N1	Asp77-Oε1 with Leu9-N1
		Asp77-Oε1 with Arg5-Nη1
	Tyr84-OH with Leu9-O1	Tyr84-OH with Leu9-O1
	Thr143-OH with Leu9-O1	Thr143-OH with Leu9-O1
	Lys146-Nζ with Leu9-O2	**Not observed**
	Trp147-Nε1 with His8-O	Trp147-Nε1 with J*8-O

Analysis was performed with PyMOL using find polar contacts function. Missing and newly emerging H-bonds are labelled in bold fond.

3.1.2 Comparison of conformational dynamics of B*27:09/photoRL9 and B*27:09/RL9 in solution by NMR

To characterize the impact of the introduced photocleavable site in the peptide on the conformational dynamics of pMHC-I molecules, dynamic differences of these two pMHC-I complexes were investigated. CPMG relaxation dispersion NMR experiments of these two pMHC-I complexes were performed, in which methyl groups of Ile, Leu, Val of the heavy chain were labelled by ^{13}C (Fig. 3.1.2a). This method allows to detect the conformational exchange within time scales of micro- to milli-seconds (µs-ms), unveiling rare populated states for the labelled residues. The assignments of ^{13}C methyl resonances of B*27:09/photoRL9 were transferred from B*27:09/RL9 (BMRB Entry 25713) and confirmed by methyl-methyl and methyl-amide NOEs when possible. Assignments are summarized in Table 3.2 and the assigned spectrum is shown in Appendix Fig. 3a. Although the ^{1}H-^{15}N TROSY-HSQC spectrum of B*27:09/photoRL9 differed significantly from B*27:09/RL9 (Appendix Fig. 1), the ^{1}H-^{13}C

HSQC spectrum of B*27:09/photoRL9 overlapped well with that of B*27:09/RL9, with only few peaks showing chemical shift changes (Appendix Fig. 3a).

Table 3.2 Summary of Ile, Val, Leu assignments in the ^{13}C-^{1}H spectrum of B*27:09/RL9

	Leu	Val	Ile
Total Number	21	13	7
Assigned amount	19	12	7
Assigned residues	32L[1],78L[1],81L,82L[1],109L[1], 110L,126L,130L,160L,168L, 172L, 201L, 206L, 215L[1], 230L[1], 266L, 270L[1], 272L	12V, 25V, 28V, 34V, 103V, 152V, 165V,189V, 231V, 247V, 248V, 249V, 261V	23I, 52I, 66I[2], 124I, 142I, 194I, 213I
Not assigned	95L, 165L, 179L	152V[2]	

[1]note: Residue[1] means only one $^{13}CH_3$ was assigned. Theoretically there should be 76 peaks in the ^{1}H-^{13}C HSQC spectrum of B*27:09/photoRL9, however, only 64 peaks were observed, which means 12 peaks were not able to detect or they overlapped with other peaks in the spectrum. [2]note: 152V and 66I were assigned in B*27:09/RL9, but resonances of these two residues were not observed in the ^{1}H-^{13}C HSQC spectrum of B*27:09/photoRL9.

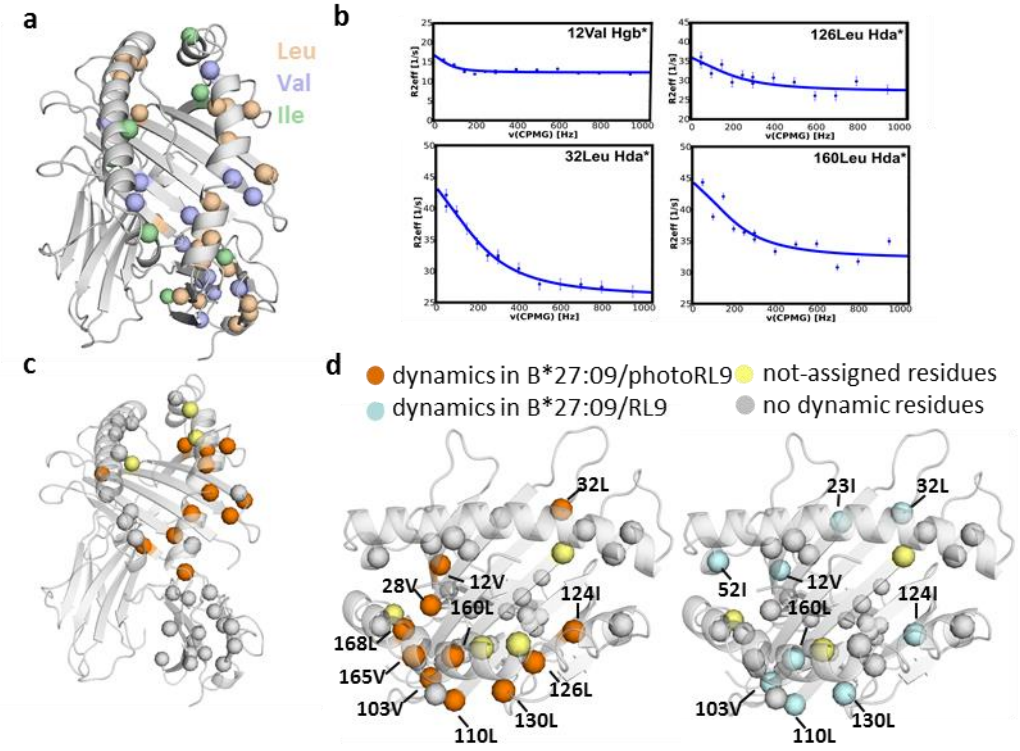


Figure 3.1.2 Comparison of dynamics of B*27:09/photoRL9 with B*27:09/RL9. **a)** ILV methyl groups are selectively labelled with $^{13}CH_3$ and are mapped as colored spheres on the structure of B*27:09/photoRL9. Labelling distributes across the whole protein. **b)** Representative relaxation dispersion curves of methyl groups of B*27:09/photoRL9 undergoing chemical exchange ($R_{ex} > 2\ s^{-1}$) in ^{1}H-^{13}C methyl CPMG experiments recorded at 27 °C at a ^{1}H field of 700 MHz. The effective transverse relaxation rate (R_2^{eff}) is shown as a function of the CPMG pulse frequency (ν_{CPMG}) for selected methyl groups. **c)** front view of the structure of B*27:09/photoRL9 with dynamic residues shown as orange spheres. No residues in the α3 domain showed dynamics. Other assigned methyl ILV probes which did not show dynamics are mapped as gray spheres. Unassigned residues are shown in light yellow spheres. **d)** Residues showing µs-ms dynamics derived from methyl-CPMG experiments of B27:09/photoRL9 (170 µM) are highlighted as orange spheres and in B*27:09/RL9 (220 µM) are highlighted as cyan spheres.

The curve of the effective relaxation rate (R_2^{eff}) of residues undergoing conformational exchange in the μs-ms timescale decays as a function of the frequency of the refocusing pulse (νCPMG), as representative examples show in Fig. 3.1.2b. A relatively large number of residues exhibit conformational exchange in both complexes, with more residues of B*27:09/photoRL9 showing dynamics (Fig. 3.1.2d). No residues in the α3 domain of the heavy chain in both complexes were observed to show this type of dynamics (Fig. 3.1.2c). It is worth to note that residues (I124, L126, L130) in B*27:09/photoRL9 located underneath the $\alpha_{2\text{-}1}$-helix display dynamics, which is not the case for L126 in B2709. It is unclear how additional dynamical residues in B*27:09/photoRL9 affect the physiological function of MHC-I in antigen presentation or its interaction with tapasin.

It has been reported that pMHC-I molecules exhibiting different thermostability show peptide-dependent conformational fluctuations (i.e., major and minor states)[203]. The peptide-dependent minor state population of pMHC-I molecules correlates well with the thermostability of pMHC-I molecules. Namely, the more stable a pMHC-I complex is, the higher population of minor state it has, as the minor state is more dehydrated and more tightly packed towards the peptide as they explained[203]. In our study, B*27:09/RL9 and B*27:09/photoRL9 did show different profiles of dynamics. According to their hypothesis, the population of minor state in B*27:09/RL9 is higher than that of B*27:09/photoRL9 since the T_m of the former is higher than that of the latter, although the latter bearing more residues showing conformational dynamics.

3.1.3 UV-exposure of photo-pMHC-I resulted in peptide-receptive state of the partially empty binding groove

UV-exposure of the photocleavable peptide is expected to result in a 7-mer and a dipeptide fragment[181]. MALDI-TOF mass spectrometry analysis showed that no signal of intact peptide (MW: 1,448.82 Da) was detected after UV-exposure while signal of 7-mer peptide (MW: 1,142.47 Da) was observed (Fig. 3.1.3b), indicating the cleavage reaction was as expected and the cleavage efficiency was nearly 100%. In order to check if the 7-mer fragment peptide would dissociate from the binding groove, we purified the UV-exposed B*27:09/photoRL9 complex by gel filtration to remove any possible peptide fragments and measured the MS data. Unexpectedly, the signal of the 7-mer peptide was still detected (Fig. 3.1.3b), indicating that the 7-mer fragment was still bound into the binding groove after UV-exposure. This is not in line with the previous research claiming that both fragments (7-mer and dipeptide) would fall off the binding groove after UV-exposure[169]. It is not clear whether the dipeptide is still occupying the F-pocket or not, due to the limitation of MALDI-TOF MS method, which can only detect molecular masses above 600 Da. However, the T_m of UV-cleaved B*27:09/photoRL9 was 5 °C lower than that of non-cleaved B*27:09/photoRL9 (Fig. 3.1.1). A reduction in T_m is most likely due to the loss of H-bonds in the F-pocket (e.g., Tyr84-OH with Leu9-O1, ~ 5.0 kcal/mol, Thr143-OH with Leu9-O1, ~ 5.0 kcal/mol and Trp147-Nε1 with J*8-O, ~1.9 kcal/mol, Fig. 3.1.1c) when the dipeptide falls off. Thus, a pMHC-I bearing a partially empty binding groove was generated by UV-exposure.

Subsequently, peptide exchange assays were performed by adding high affinity FITC-peptide to test if the partially empty MHC-I is peptide-receptive. Loading of FITC-RL9 was monitored by Fluorescence

Polarization (FP). As shown in Fig. 3.1.3c, peptide exchange of pMHC-I before UV-exposure was extremely slow (gray curve), while upon UV-exposure, peptide exchange happened immediately (orange curve), demonstrating that the UV-cleaved B*27:09/photoRL9 was peptide-receptive and the 7-mer fragment can be replaced by the newly added high affinity peptide.

Together, our data showed that after UV-exposure, the photocleavable peptide was successfully cleaved and the dimer dissociated from the F-pocket, leaving the binding groove of B*27:09/photoRL9 partially empty. The release of the C-terminal dipeptide fragment in the F-pocket upon UV-exposure shifts B*27:09/photoRL9 into a peptide-receptive state, which allows for the exchange of the peptide.

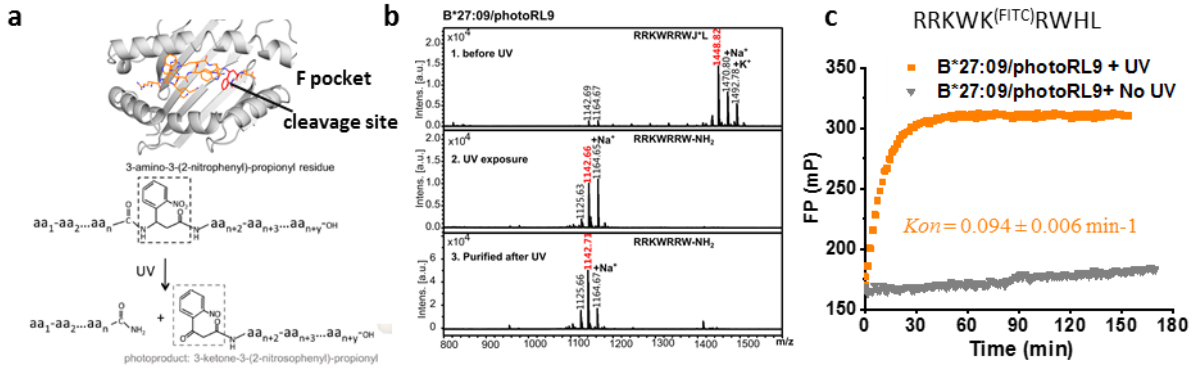


Figure 3.1.3 The 7-mer peptide fragment is still bound to the binding groove after UV-exposure. **a)** Upper: photoRL9 in the binding groove and the cleavage site is colored in red. Lower: UV-induced cleavage mechanism. **b)** The average relative abundance for the selected mass range of MALDI-TOF-MS analysis of B*27:09/photoRL9 before UV-exposure, after UV-exposure and purified after UV-exposure. The 7-mer fragment peptide was still bound to MHC-I after UV-exposure. The signal of the intact peptide and 7-mer fragment RRKWRRW-NH_2 is noted in red (observed mass 1,142.66 Da, expected mass 1,143.47 Da) as well as signal $+Na^+$ and $+K^+$. **c)** Peptide exchange profiles of 27:09/photoRL9 against RRKWK^{FITC}RWHL before (gray) and after (orange) UV-exposure. The observed association rate (*Kon*) after UV exposure is given.

3.1.4 Characterization of the peptide-receptive state of B*27:09/photoRL9 by NMR

Subsequently the dynamic features of B*27:09/photoRL9 after peptide cleavage were investigated by NMR spectroscopy. The ^{1}H-^{15}N TROSY-HSQC spectra of UV-cleaved ^{15}N-B*27:09/photoRL9 was recorded and compared to that of non-cleaved B*27:09/photoRL9 (Fig. 3.1.4).

Upon UV-exposure, resonances from residues in the binding groove underwent either chemical shift changes, or line broadening, or both (Fig. 3.1.4a, b, c). Interestingly, chemical shift changes, which indicate an altered chemical environment, for example, due to peptide dissociation (C-terminus) or rearrangement of residues that constitute the peptide/MHC-I interface, are observed throughout the binding groove (Fig. 3.1.4d). Moreover, residues in the β-strand 7, $loop_{126\text{-}133}$ and β-strand 8 showing chemical shift changes upon UV cleavage were also shown to undergo conformational dynamics in the intact complex (such as I124, L126, L130, Fig. 3.1.1d), suggesting the importance of this region for distinguishing peptide-receptive and peptide-bound state. Line-broadened resonances, which is indicative of a direct interaction and/or conformational dynamics in the µs-ms range, were observed throughout the binding groove as well. Particularly, residues which compose the architecture of the F-pocket, including T143 that forms an H-bond with the C-terminus of the peptide, residues M95, N97, and A117 at the bottom of the F-pocket of MHC-I as well as residues R75, E76, D77, and R79 from the

α_1-helix in the F-pocket region (Fig. 3.1.4a and b), showed strong line broadening, indicating altered dynamics of this region upon UV-exposure.

Besides the F-pocket region, MHC-I heavy chain residues 151-155 located close to the central part of the peptide were also strongly affected, indicating that the altered dynamics might propagate towards the middle of the binding groove due to the allosteric rearrangement of the 7-mer peptide. In contrast, the A-pocket residues Y7 and Y171 which form H-bonds with the N-terminus of the peptide, showed only slight line broadening, which corresponds to the MS data that the 7-mer peptide remains in the binding groove. No assigned residues in the α3 domain showed significant signal reduction or chemical shift changes (Fig. 3.1.4c and d).

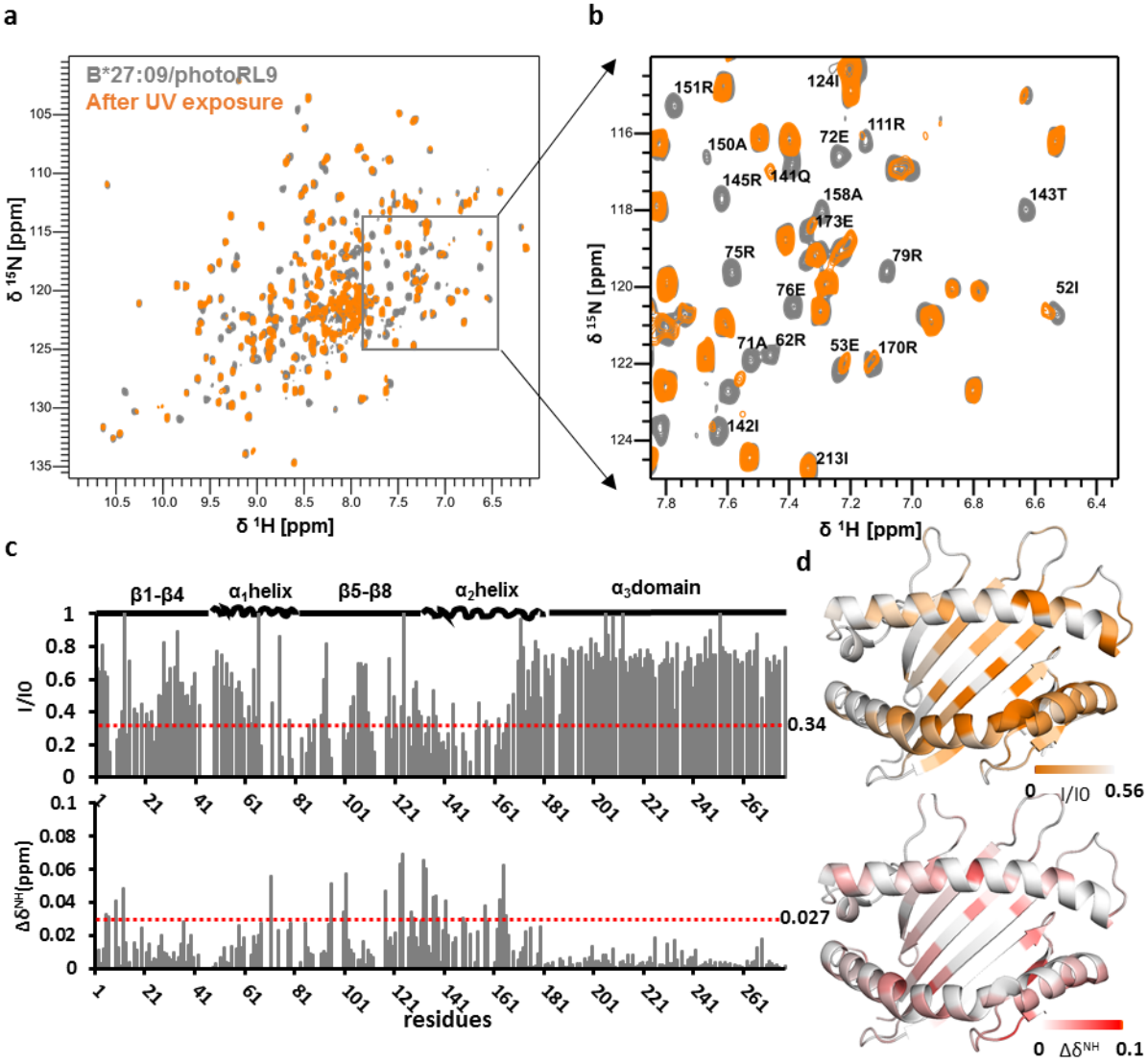


Figure 3.1.4 Characterization of peptide-receptive B*27:09/photoRL9 by NMR. a) Overlay of 1H-15N TROSY-HSQC spectrum of B*27:09/photoRL9 (gray) and UV-exposed B*27:09/photoRL9 (orange). **b)** A zoom view is shown for better illustration. Line-broadening peaks upon UV-exposure are noted. **c)** The upper chart is the peak intensity ratio analysis (I/I0) of UV-exposed B*27:09/photoRL9 compared to B*27:09/photoRL9. The dotted red line represents 1σ less than the average I/I0. Intensity ratio below the red dashed line is considered significantly reduced. The lower chart is the chemical shift deviation (CSD, ΔδNH, p.p.m.) of peaks between UV-exposed B*27:09/photoRL9 and non-exposed B*27:09/photoRL9. The dotted red line represents 1σ higher than the average CSD value. **d)** Intensity ratio (in orange) and CSD effects (in red) are mapped on the structure of B*27:09, respectively. The range of the intensity ratio is from 0 to the average intensity value. Intensity ratio of

line-broadened residues are marked as 0. Unassigned residues were filled with the value of the average intensity. The range of CSD is from 0 to 0.1.

Besides changes in ^{1}H-^{15}N TROSY-HSQC spectrum after UV-exposure, changes in ^{1}H-^{13}C HSQC spectrum upon UV-exposure were also analyzed. Unlike the ^{1}H-^{15}N TROSY-HSQC spectrum, where resonances are very sensitive to changes in backbone conformation, only few chemical shift changes were observed in the ^{1}H-^{13}C HSQC spectrum. Interestingly, residues V34, L81, L82, I124, L126, L130, I142 and L160 showed significant line broadening or chemical shift changes (Appendix Fig. 7a). Notably, most of them also showed conformational fluctuations in CPMG measurements (Fig. 3.1.2d). Particularly, the resonance of residue I124 split into two peaks, indicating that two conformations exist after UV-exposure.

Thus, we characterized the structural and dynamic changes of the pMHC-I molecule when it turns into the peptide-receptive state. These changes mainly refer to the binding groove, especially to the F-pocket region. They are also polarized toward the α_2-helix side including β-strand 7, $loop_{126\text{-}133}$ and β-strand 8 which are located beneath the $\alpha_{2\text{-}1}$-helix. Therefore, these changes can be used as references when studying dynamics of MHC-I.

3.1.5 Incubating UV-exposed B*27:09/photoRL9 with new peptide can recover line-broadened peaks

As expected, line-broadened peaks reappeared when the UV-exposed B*27:09/photoRL9 was incubated with the full length RL9 peptide and the overall ^{1}H-^{15}N HSQC spectrum after incubation was superimposed well to the ^{1}H-^{15}N HSQC spectrum of B*27:09/RL9, indicating the UV-induced-peptide-receptive state of MHC-I can turn back into the peptide-bound state by exchanging with high affinity peptide. Reloading of the peptide-receptive MHC-I with new peptide can recover the conformation of pMHC-I exactly as it was when MHC-I was initially refolded with the RL9 peptide. These results exclude the possibility that UV-exposure would cause functional damages to the backbone or side chains of the heavy chain.

It is worth noting that UV-cleaved B*27:09/photoRL9 cannot load peptide anymore after several days, which means the UV-cleaved B*27:09/photoRL9 gradually switched to the peptide-non-receptive state. This might be due to the empty F-pocket not being stabilized by peptides. In order to monitor conformational changes during the process of state switch (peptide-receptive state to peptide-non-receptive state), the spectrum of UV-cleaved B*27:09/photoRL9 was measured at different time points (0 h, 3 days, 8 days). As shown in Fig. 3.1.5a, the spectrum of 3 days, even 8 days was largely superimposable to the 0 h spectrum and no crowded peaks (signs of aggregation/degradation) appeared in the spectrum, indicating the peptide-non-receptive state was not due to the unfolding of the heavy chain. Interestingly, differences between these three spectra were small and mainly in the $loop_{126\text{-}133}$ as well as the F-pocket: L126, L95 and the side-chain of W133 showed chemical shift as time increases, while L130 and S132 split into two peaks and the new peak becomes stronger as time increases (Fig. 3.1.5b and c). These observations suggest that the conformation of $loop_{126\text{-}133}$ might report on the state of pMHC-I molecule: peptide-receptive or peptide-non-receptive conformations.

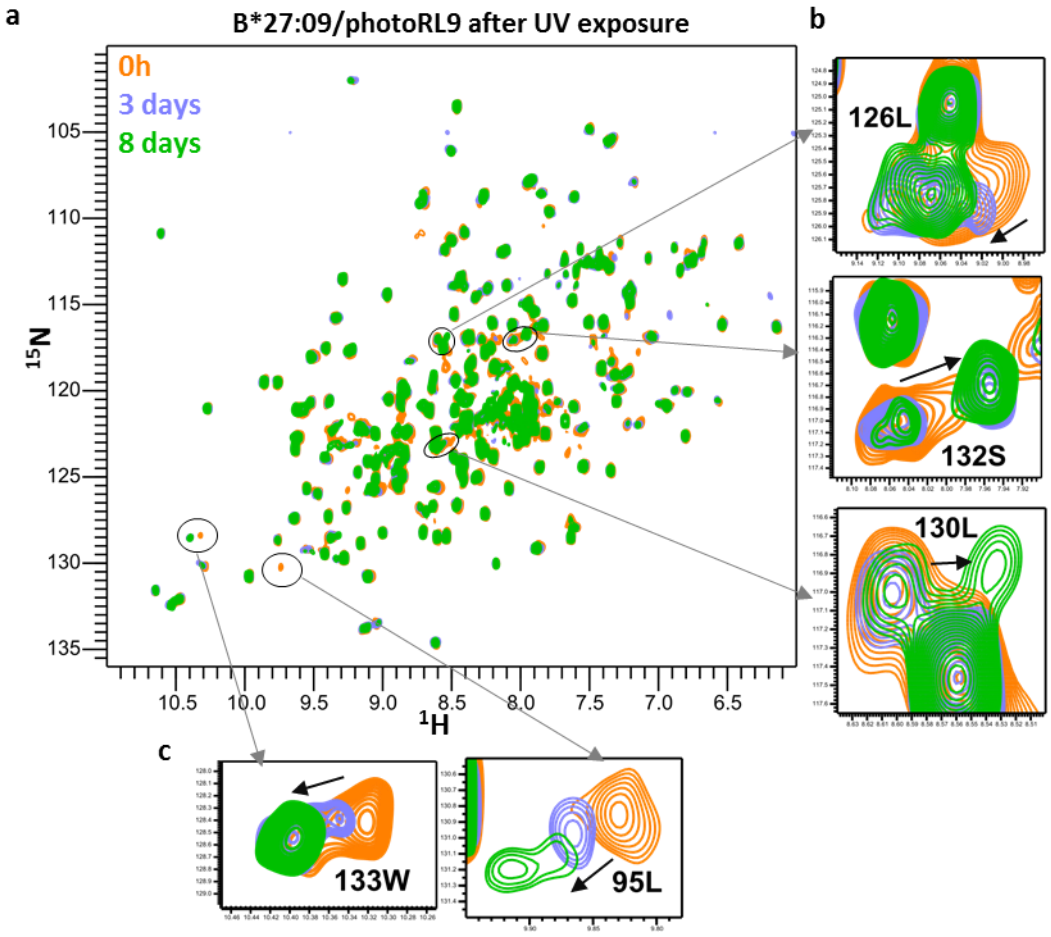


Figure 3.1.5. Monitor conformation changes of B*27:09/photoRL9 after UV-exposure as time increases. **a)** Overlay of ^{1}H-^{15}N TROSY-HSQC spectrum of B*27:09/photoRL9 at 0 h (orange), 3 days (light blue) and 8 days (green) after UV-exposure. When not measured, the sample was kept at 4 °C. Zoom in view of circled peaks are shown in **b)** and **c)**. **b)** Zoom in view of residues in the loop$_{123-133}$ at different time points. **c)** Peaks of L95 and the side chain of W133 at different time points.

3.1.6 Differences between B*27:05 and B*27:09

After characterizing the structure and the conformational dynamics of B*27:09, we expanded the study to another B*27 subtype, B*27:05, which differs from B*27:09 only by residue 116.

The refolding yield of B*27:05 with the same peptide was lower than that of B*27:09 (5%-10% for B*27:05/photoRL9 and 15%-30% for B*27:09/photoRL9), indicating B*27:05 is less stable than B*27:09 *in vitro*. However, the T_m of B*27:05/RL9 was higher than that of B*27:09/RL9 by ~ 2 °C (61.88 ± 1.29 °C for B*27:05 VS 60.20 ± 0.15 °C for B*27:09, Appendix Table 3) and the T_m of B*27:05/photoRL9 was higher than that of B*27:09/photoRL9 by ~ 3 °C (57.58 ± 0.08 °C for B*27:05 VS 54.55 ± 0.13 °C for B*27:09, Appendix Table 3), indicating there is no correlation between the refolding yield and thermostability of pMHC-I molecule. Interestingly, when refolded with another peptide IF9 (IRAAPPPLF, derived from cathepsin A, the binding affinity of it to B*27:09/05 is lower than RL9, Appendix Table 2), the T_m of B*27:05/photoIF9 was lower than that of B*27:09/photoIF9 by ~ 5 °C, whereas the T_m of the cleaved B*27:05/photoIF9 was higher than that of the cleaved B*27:09/photoIF9 by ~ 6 °C. The 7-mer fragment was still in the binding groove of B*27:05/photoRL9 after UV-exposure, as observed for in the cleaved B*27:09/photoRL9.

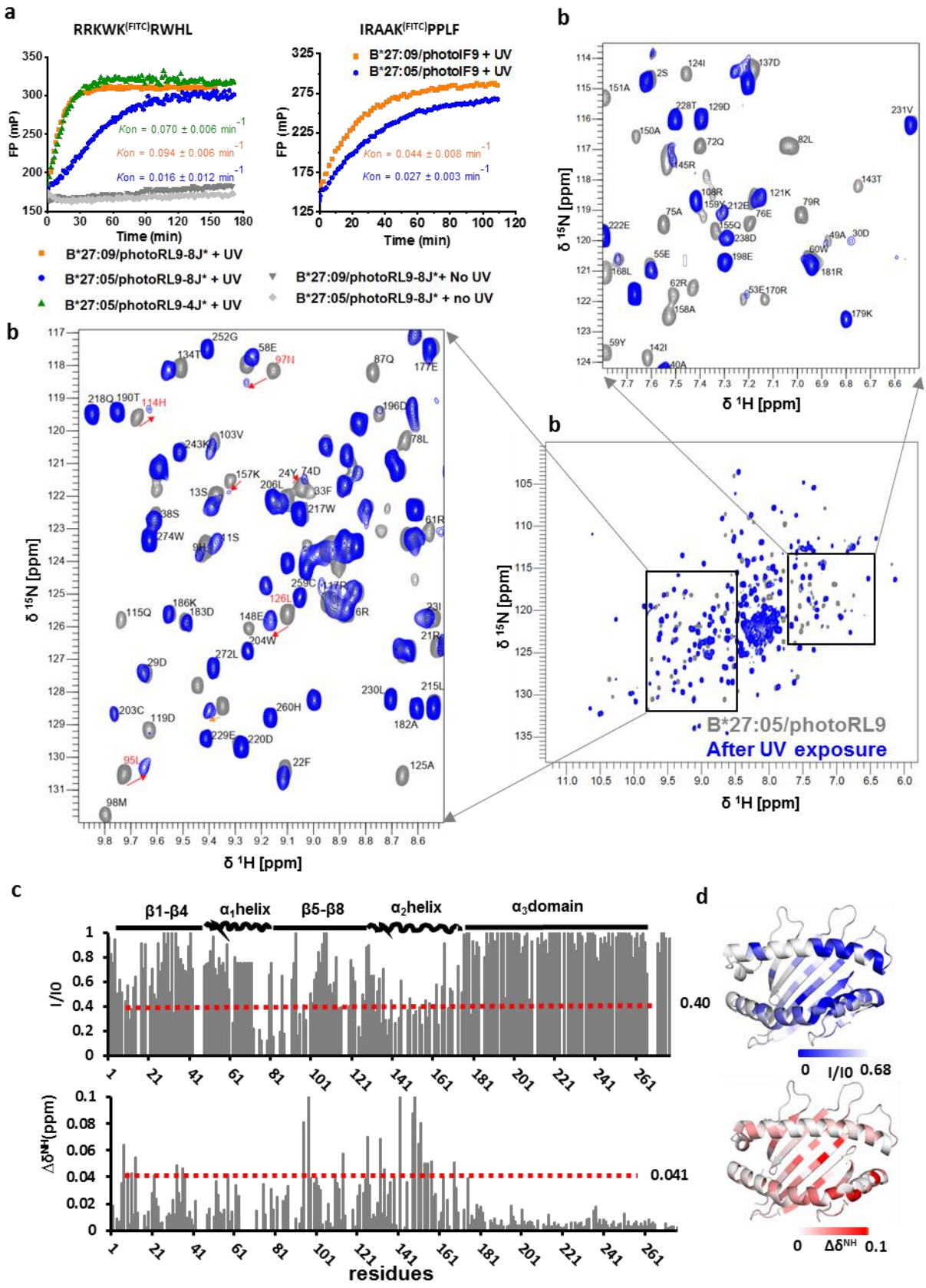


Figure 3.1.6 Characterization of peptide-receptive state of B*27:05/photoRL9 after UV exposure. **a)** Left panel: Peptide exchange profiles of UV-exposed or non-UV-exposed B*27:05 or 27:09/photoRL9 against 100 nM RRKWK$^{(FITC)}$RWHL. 8J* and 4J* indicates the cleavage site is at 8th and 4th position, respectively. Right panel: peptide exchange profiles of 500 nM UV-exposed B*27:09/photoIF9 and UV-exposed B*27:05/photoIF9 against 10 nM IRAAK$^{(FITC)}$PPLF. **b)** Overlay of ^{1}H-^{15}N HSQC spectrum of B*27:05/photoRL9 (gray) and UV-exposed B*27:05/photoRL9 (blue), two zoom views are chosen for better

illustration. Chemical shift direction upon UV-exposure is indicated by red arrow. **c)** The upper chart is the normalized peak intensity ratio analysis (I/I0) of the UV-exposed B*27:05/photoRL9 compared to intact B*27:05/photoRL9. Dotted red line represents 1σ less than the average I/I0. Intensity ratio below the red dashed line is considered significantly reduced. The lower chart is the chemical shift deviation (CSD, $\Delta\delta^{NH}$, p.p.m.) of peaks between UV-exposed B*27:05/photoRL9 and non-exposed B*27:05/photoRL9. The dotted red line represents 1σ higher than the average CSD value. **d)** Intensity ratio (in blue) and CSD effects (in red) are mapped on the structure of B*27:05/RL9 (1OGT)[26]. The range of the intensity ratio is from 0 to the average intensity value. Intensity ratio of vanished residues are marked as 0. The range of CSD is from 0 to 0.1.

Taken together, the comparison between B*27:05 and B*27:09 further confirmed that the thermostability of pMHC-I was dependent both on the sequence of the heavy chain and the bound peptide[204].

The peptide exchange rate of the cleaved B*27:09/photoRL9 was 6-fold faster than that of the cleaved B*27:05/photoRL9 (Fig. 3.1.6b left panel, compare the blue curve with the orange curve). The same trend was observed when B*27:05 and B*27:09 is refolded with another medium affinity peptide IF9. However, the difference was smaller as the *Kon* of B*27:09/photoIF9 with FITC-IF9 was only 1.6-fold faster than that of B*27:05/photoIF9 exchanging with FITC-IF9 (Fig. 3.1.6a, right panel). In order to check whether the position of cleavage site in the bound peptide would make a difference in peptide exchange rate, RRKJ*RRWHL was used for refolding (notated as photoRL9-4J*) with B*27:05/09. Strikingly, peptide exchange of the cleaved B*27:05/photoRL9-4J* with FITC-RL9 was increased by 4-fold compared to that of the cleaved B*27:05/photoRL9-8J* (Fig. 3.1.6a, olive curve). Nevertheless, peptide exchange of the cleaved B*27:09/photoRL9-4J* against FITC-RL9 was slightly increased compared to B*27:09/photoRL9-8J*. These results indicated that peptide exchange of B*27:05 is more sensitive to the position of the cleavage site of the bound peptide.

Less resonances in the ^{1}H-^{15}N TROSY-HSQC spectrum of B*27:05/photoRL9 were assigned (Appendix Fig. 2) due to B*27:05/photoRL9 being less stable during NMR data acquisition. After UV-exposure, similar patterns of affected residues were observed in the ^{1}H-^{15}N TROSY-HSQC between the two subtypes. Resonances of residues in the binding groove either showed chemical shifts or line-broadening (Fig. 3.1.5b and c). But it is worthy of note that many peaks of B*27:05/photoRL9 showed chemical shift changes rather than only line-broadening after UV-exposure. For example, residues 114H, 95L, 97N which line the F-pocket showed chemical shift changes as well as intensity reduction (Fig. 3.1.5b, left panel), indicating the cleaved B*27:05/photoRL9 went through more drastic conformational changes compared to the cleaved B*27:09/photoRL9. Strikingly, resonances of residues in the $loop_{126-133}$ of cleaved B*27:05/photoRL9 showed large chemical shift differences (Fig. 3.1.5b), indicating a change in conformation of the $loop_{126-133}$ when the F-pocket is empty. Interestingly, the side-chain of W133 in B*27:05/photoRL9 before UV-exposure already has two conformations (Appendix Fig. 6a) and the proportion of the non-peptide-receptive state after UV-exposure increased, which also explained why the peptide exchange of the cleaved B*27:05/photoRL9 was slower than that of the cleaved B*27:09/photoRL9. Similar to the ^{1}H-^{13}C HSQC spectra of B*27:09/photoRL9 after UV-exposure, not many changes were observed in the ^{1}H-^{13}C HSQC spectrum of B*27:05/photoRL9 after UV-exposure (Appendix Fig. 7b). Residues V34, L81, L82, I124, L126, I142 and L160 showed significant line broadening or chemical shift differences (Appendix Fig. 7b) as in B*27:09/photoRL9.

In contrast to B*27:09/photoRL9, I124 of B*27:05/photoRL9 showed line-broadening rather than chemical shift changes (I124 splits into two peaks in B*27:09/photoRL9 after UV exposure, Appendix Fig. 7a), while L81 showed large chemical shift difference rather than line-broadening (Appendix Fig. 7b). Collectively, the overall features were similar when B*27:05/photoRL9 turned into peptide-receptive state after UV-exposure compared to B*27:09/photoRL9, but pronounced differences were observed in the region of the F-pocket.

3.2 Structural features of MHC-I that contribute to tapasin dependence

Tapasin is a peptide editor and catalyzes peptide exchange of MHC-I molecules within the PLC. Studies showed that peptide exchange can be enhanced *in vitro* by ERp57-Tsn conjugate[145,150], or by soluble Tsn when covalently linked to the heavy chain by leucin zipper[146], otherwise only low enhancement of peptide exchange in the presence of soluble Tsn alone was observed due to the weak binding affinity between pMHC-I and tapasin[146]. For this reason, in this work, tapasin was co-expressed with ERp57 (bearing a C60A mutant) as a conjugate (referred to as ERp57-Tsn^{WT}) in insect cells and purified as described in section 2.2.4. In this chapter, the influence of tapasin on the peptide exchange kinetics of MHC-I was investigated by comparing the MHC-I subtypes B*27:09 and B*27:05 refolded with different photocleavable peptides for which affinities ranged from high to medium. Since MHC-I molecules are extremely polymorphic, this comparison was then expanded to more MHC-I allotypes to define the structural elements that determine the dependence of MHC-I molecules on tapasin-mediated peptide exchange. HLA-A*02:01 and HLA-A*03:01 were chosen because of their high frequency in the Caucasian population. In addition, the interactions between tapasin and pMHC-I were analyzed in detail by NMR measurements. As a comparison, TAPBPR, the second known peptide editor, was also included in these experiments.

3.2.1 Influence of ERp57-Tsn on peptide exchange kinetics

Although the peptide exchange of B*27:09/05 in complex with photoRL9 can be enhanced in the presence of ERp57-Tsn^{WT} without UV-exposure (Appendix Fig. 10a), the rate was extremely slow. We thus decided to perform all experiments under conditions of UV-exposure to explore how tapasin influences the peptide exchange kinetics of B*27:09/05 *in vitro*. Peptide exchange experiments were done as descripted in the following sentences: After UV-exposure of photo-MHC-I, equimolar amounts of ERp57-Tsn^{WT} were added and peptide exchange was monitored immediately after adding high affinity FITC-peptides. FP assays were all performed at pH 7.0. As shown in Fig. 3.2.1a, b and g, the peptide exchange rate (*Kon*) of the cleaved B*27:09/photoRL9 against FITC-RL9 was increased by 1.64 ± 0.08 -fold and the cleaved B*27:05/photoRL9 by 2.50 ± 0.34 -fold in the presence of ERp57-Tsn^{WT}, respectively. This indicates that both allotypes are susceptible to tapasin-mediated peptide exchange.

Furthermore, we investigated if the affinity of pre-bound peptides would lead to a different enhancing effect of tapasin on the peptide exchange of MHC-I molecules. B*27:09 refolded with photoIF9 (the affinity of it is significantly lower than photoRL9, affinity data is listed in Appendix Table 2) was used to exchange against FITC-RL9 in the absence or presence of ERp57-Tsn^{WT} (Fig. 3.2.1a). To indirectly obtain the affinities of the cleaved peptide to MHC-I molecules, we measured the thermostability of the cleaved B*27:09/photoRL9 and cleaved B*27:09/photoIF9 and observed a 7 °C higher stability of

B*27:09 in complex with photoRL9 compared to photoIF9 after UV-exposure (Fig. 3.2.1e). This indicates a correspondingly higher affinity of cleaved photoRL9 (7-mer) to the complex. Accordingly, peptide exchange rate of B*27:09/photoIF9 was enhanced 1.23 ± 0.09-fold by ERp57-Tsn^{WT} after UV-exposure (Fig. 3.2.1g), indicating the pre-bound peptide affects the susceptibility of MHC-I to tapasin-catalyzed peptide exchange (which probably influences the peptide dissociation rate). The uncatalyzed peptide exchange rate (the intrinsic peptide exchange rate) was higher for the less stable UV-cleaved photoIF9 complexes (Fig. 3.2.1f) while the factor of enhancement of *K*on by ERp57-Tsn^{WT} was larger in case the higher affinity peptide photoRL9 was used as pre-bound peptide (Fig. 3.2.1g). The same trend was observed for B*27:05 allotype (Fig. 3.2.1b and d, g).

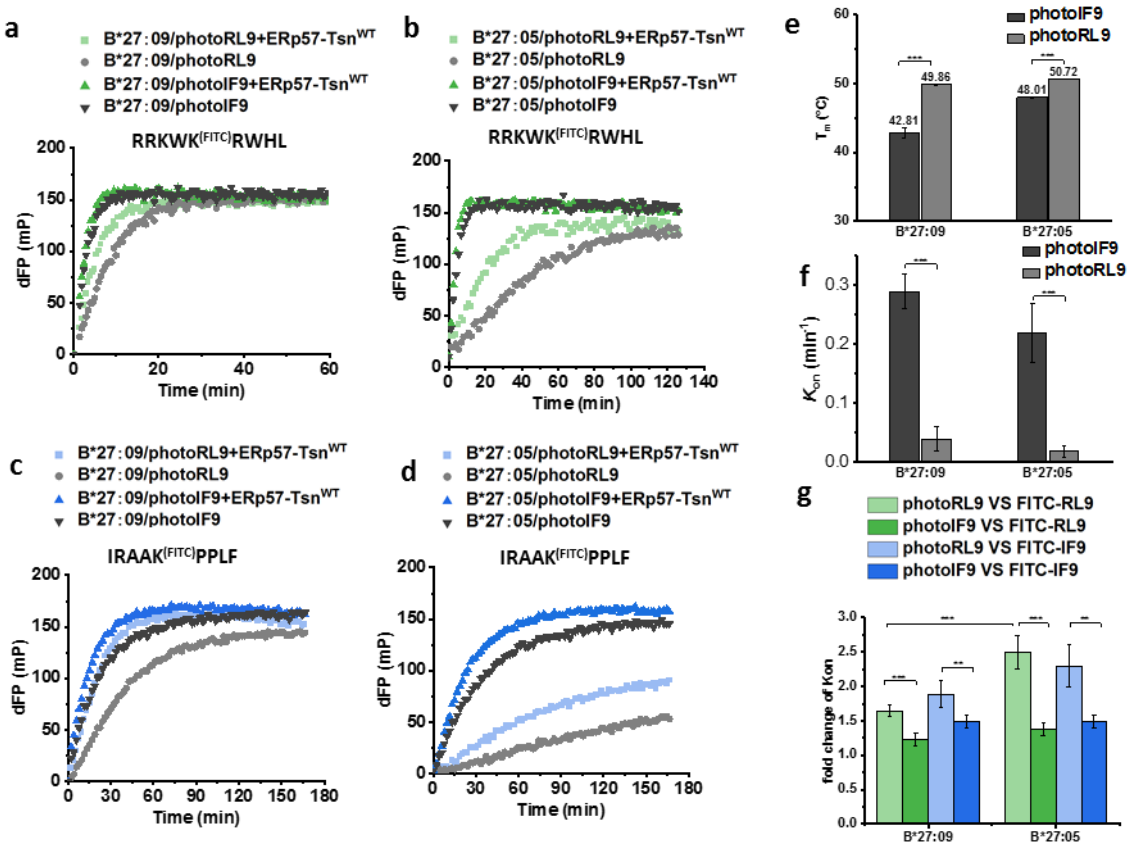


Figure 3.2.1 The peptide exchange of B*27:09/05 in the presence of tapasin. Peptide exchange profile of 500 nM B*27:09/photoRL9 (gray) and B*27:09/photoIF9 (dark) in the absence or presence of equimolar ERp57-Tsn (green or blue) after UV-exposure **a)** against FTTC-RL9 and **c)** against FITC-IF9. Peptide exchange profile of 500 nM B*27:05/photoRL9 (gray) and B*27:05/photoIF9 (dark) in the absence or presence of equimolar ERp57-Tsn (green) after UV-exposure **b)** against FTTC-RL9 and **d)** against FITC-IF9. dFP means the baseline was subtracted. **e)** The T_m values of B*27:09/05 in complex with photoIF9 (dark) and photoRL9 (gray) after UV-exposure. **f)** The uncatalyzed peptide exchange rate (*Kon*) of B*27:09/05 in complex with photoIF9 (dark) and photoRL9 (gray) exchange against FITC-RL9 after UV-exposure. **g)** Fold changes of *Kon* of B*27:09/05 in the presence of tapasin in complex with photoRL9, photoIF9 against FITC-RL9 or FITC-IF9. Errors bars represents SD and were calculated form three repeats. *p < 0.05, **p < 0.01, ***p < 0.001.

We also investigated if the affinity of the incoming peptide would affect the susceptibility of MHC-I to tapasin. The fold change of *Kon* of B*27:09 in the presence of tapasin against FITC-IF9 was slightly higher than that against FITC-RL9 when refolded with the same peptide (Fig. 3.2.1c, g, 1.89 ± 0.20 -fold against FITC-IF9 vs 1.64 ± 0.08 -fold against FITC-RL9, 1.49 ± 0.09 -fold against FITC-IF9 vs 1.23 ± 0.09-fold against FITC-RL9). The similar trend was observed for B*27:05 (Fig. 3.2.1b and d, g).

we conclude that the affinity of the incoming peptide has a more subtle effect on the enhancement of *K*on by ERp57-Tsn^{WT} when the allotype is in complex with the same photo-peptide (Fig. 3.2.1g). Collectively our data demonstrated that the degree of tapasin-catalyzed peptide exchange within an allotype is related to both the pre-bound peptide and incoming peptide affinity with the effect of pre-bound peptide affinity being more pronounced. This result is in line with the previous finding obtained with the B*44:02 allotype that show tapasin to catalyze thermodynamically favored but kinetically disfavored exchange of suboptimal peptide for high affinity peptides[150]. It is also in line with a previously report that tapasin enhancing peptide loading generally parallels pMHC-I complex stability[162].

It has been shown that the F-pocket region of pMHC-I is important for its tapasin dependence[30]. When we compared the fold change of *Kon* of B*27:05 catalyzed by tapasin, which differs from B*27:09 by only one residue 116 in the F-pocket (D instead of H), with B*27:09 in complex with the same peptide and measuring exchange rates against the same high affinity peptide, B*27:05 is found to be more susceptible to tapasin-mediated peptide exchange than B*27:09 (Fig. 3.2.1g, 2.50 ± 0.34 -fold vs 1.64 ± 0.08 -fold exchange against FITC-RL9). Our data is in line with previous results that B*27:05 is more tapasin-dependent than B*27:09[30]. It is of note that their difference in susceptibility to tapasin becomes smaller when the complex is less stable after UV exposure (Fig. 3.2.1g).

3.2.2 Allelic dependence of MHC-I on tapasin-mediated peptide exchange

Results from the last section showed that the dependence of B*27:05/09 on tapasin-mediated peptide exchange was different, even though the two allele differ only by one residue in the F-pocket. In order to further study the question regarding tapasin dependence of different allotypes, we then extended our analysis to the comparison of peptide exchange catalysis by tapasin for two HLA-A allotypes that contain distinct F-pockets. HLA-A*02:01 and HLA-A*03:01 were chosen because of their high frequency in Caucasian population. The allotype A*02:01 comprises a hydrophobic F-pocket with Y116 forming the floor. A*03:01 contains a D at this position, the same as in B*27:05. The charged surface area of the individual allotypes is displayed in Fig. 3.2.2a, highlighting the critical amino acids that compose the F-pocket as well as the peptide residue 9 (green) diving into the pocket in the respective crystal structures. Clearly, the F-pocket of these four allotypes shows a gradient in acidity: A*03:01 > B*27:05 > B*27:09 > A*02:01, in line with a differential preference for basic or hydrophobic residues at the C-terminus (PΩ) of bound peptides (Fig. 3.2.2b). Peptide exchange experiments were performed by exchanging against fluorescently-labeled versions of same high-affinity peptides used for refolding (high-affinity peptides are used to maximum the change factor, peptide affinity data and stability data of photo-MHC-I are listed in Appendix Table 2 and 3). The fluorescent amino acid was introduced at P5 for all FITC-peptides.

In the presence of ERp57-Tsn^{WT}, the peptide exchange rate was increased by 3.55 ± 0.37 -fold for A*03:01, 2.50 ± 0.20 -fold for B*27:05, 1.64 ± 0.10 -fold for B*27:09, 1.01 ± 0.01 -fold for A*02:01 (Fig. 3.2.2c). ERp57-Tsn^{WT} did not accelerate the peptide exchange of A*02:01 (Fig. 3.2.2c and d) under these conditions, which is consistent with a previous study that the cell surface expression of peptide GILGFVFTL (influenza A MP58-66, the same peptide as we used here) by A*02:01 was not affected in tapasin-deficient cells[205]. However, some other previous studies also found that refolding of A*02:01

can be facilitated by the N-terminal domain of tapasin[161]. B*27:05 was reported to be intermediate-dependence on tapasin, based on the fact that the spectrum of presented peptides by B*27:05 in tapsin-deficient cells was not significantly changed[158]. Thus, our data indicate that susceptibility to tapasin-mediated peptide exchange of these four allotypes is likely positively correlated with the gradient of the acidity of the F-pocket.

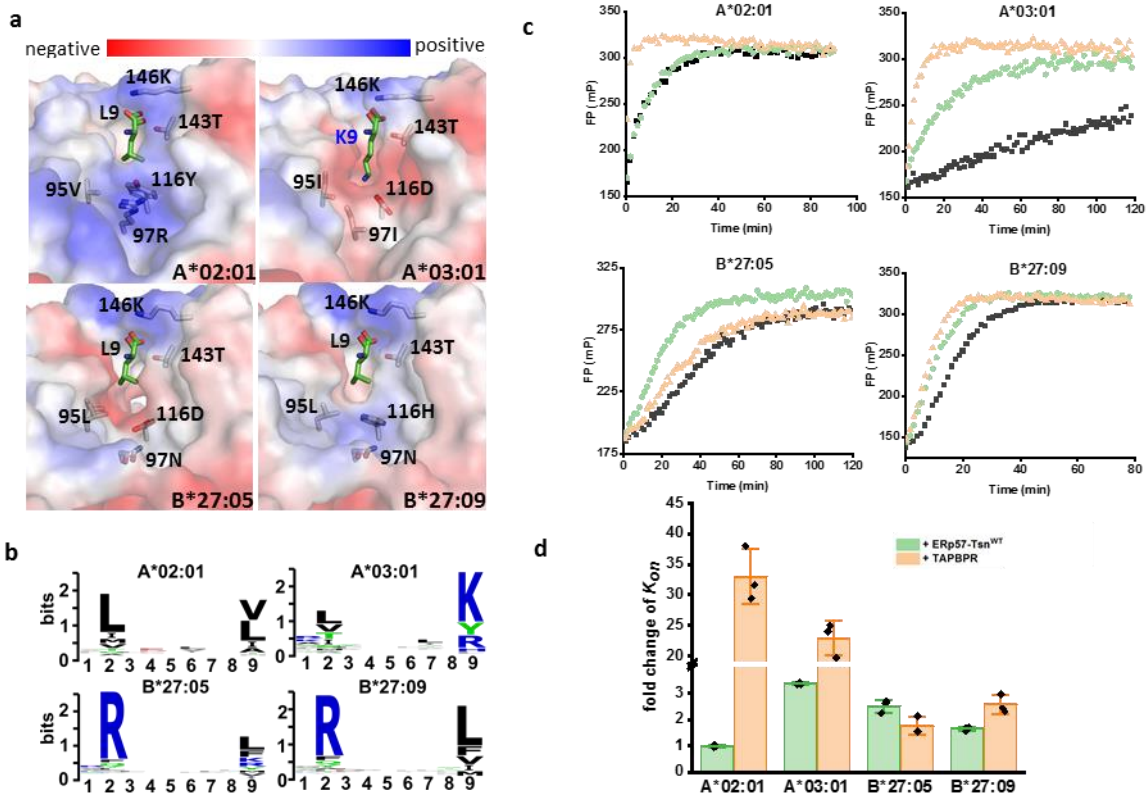


Figure 3.2.2 Comparison of catalytic activity of Tapasin and TAPBPR on different alleles. a) F-pocket of A*02:01(1OGA)[206], A*03:01(3RL1)[207], B*27:09 (1OF2)[26] and B*27:05 (1OGT)[26] are shown as surfaces colored by their vacuum electrostatics. The residues composing the F-pocket and P9 residue (green) of bound peptide are shown as labeled sticks. **b)** Sequence logo of peptides presented by A*02:01, A*03:01, B*27:09 and B*27:05. Peptides used for plotting were downloaded from IEBD and sequence logos were generated in https://weblogo.berkeley.edu. **c)** Representative peptide exchange profiles of 500 nM pMHC-I folded with photocleavable high affinity peptides in the absence (black) and presence of equimolar amounts of ERp57-Tsn^{WT} (green) or TAPBPR (wheat) after UV-exposure. Peptide exchange process was monitored after adding the FITC-version of the peptide used for refolding: 10 nM GILG-Lys$^{(FITC)}$-VFTV to A*02:01/photoGV9, 10 nM KLIE-Lys^{FITC}YFSK to A*03:01/photoKK9, 100 nM RRKW-Lys$^{(FITC)}$-RWHL to B*27:05/photoRL9 and B*27:09/photoRL9. The measuring time was different for each allotype. **d)** The fold change of *Kon* in the presence of tapasin or TAPBPR of four allotypes. The fold change (ratio) of *Kon* was calculated by comparing the catalyzed *Kon* with the uncatalyzed *Kon* for every allotype. Error bar (SD) was calculated from 3 independent experiments.

TAPBPR has been reported to have a strong preference to bind to HLA-A alleles over HLA-B, HLA-C alleles[168] and thus was included in this section to compare the difference of one allotype's dependence on tapasin and TAPBPR. TAPBPR^{WT} (carrying a C97A mutation) was recombinantly expressed in insect cells and purified as described in section 2.2.4. In the presence of TAPBPR, peptide exchange rates of A*02:01, A*03:01, B*27:09, and B*27:05 were increased by 32.99 ± 4.5, 22.88 ± 2.8, 2.56 ± 0.39, and 1.54 ± 0.33-fold, respectively (Fig. 3.2.2d), which is in line with previous reports that TAPBPR has a strong preference for HLA-A alleles, especially the HLA-A*02 and HLA-A*24 superfamily[168]. It

is worth noting that in our experimental set up our data showed B*27:09 is more TAPBPR susceptible than B*27:05, in accord with a previous report where in the presence of TAPBPR no significant enhancement of peptide binding for B*27:05 was detected, in contrast an enhancement of over 100% at the level of peptide exchange was observed for B*27:09[168]. Thus our data showed that the susceptibility to TAPBPR-mediated peptide exchange of human allotypes is likely negatively correlated with the gradient of the acidity of the F-pocket within HLA-A or HLA-B supertype, namely: 22.88 -fold for A*03:01 < 32.99 -fold for A*02:01, 1.74 fold for B*27:05 < 2.56 -fold for B*27:09. Indeed, it has been reported that TAPBPR prefers to promote peptide exchange of MHC-I allotype bearing a hydrophobic F-pocket[168].

Taken together, our data suggests that the F-pocket nature of MHC-I has an impact on its susceptibility to tapasin and TAPBPR. Futhermore, it also implies that the susceptibility of a certain MHC-I allotype to tapasin-mediated or TAPBPR-mediated peptide exchange is not necessarily correlated. For example, the catalytic activity of tapasin and TAPBPR promoting peptide exchange of B*27:09 and B*27:05 is of opposite magnitude.

3.2.3 Tapasin promotes pMHC-I to adopt a peptide-receptive state

In order to specifically characterize the interface between tapasin and pMHC-I, stoichiometric amounts of ERp57-Tsn^{WT} were added into ^{1}H-^{13}C-ILV-^{15}N-^{2}H MHC-I bearing the intact photocleavable peptide and subsequently its ^{1}H-^{15}N TROSY-HSQC spectrum was recorded. The reasons for using non-cleaved MHC-I are twofold. First, changes of peaks of interest (especially residues in the binding groove) could not be detected using the cleaved peptide since most of them are already showing line broadening after cleavage (Fig. 3.1.4). Second, ERp57-Tsn^{WT} can still catalyze peptide exchange of non-cleaved MHC-I with FITC-peptide, although slowly (Appendix Fig. 10a, compare the gray curve with the green curve), which mimics ERp57-Tsn^{WT} interacting with low-affinity-peptide-loaded-MHC-I within the PLC.

Upon adding ERp57-Tsn^{WT}, while almost no chemical shift changes could be seen (Appendix Fig. 8a), indicating transient interaction, line broadening for resonances assigned to residues of the peptide binding groove are observed. Line broadening can occur either because of a direct interaction with tapasin or altered dynamics induced by tapasin. Interestingly, many of these resonances (e.g., R79, N97, A117, D120, I124, T143, A151, Fig. 3.2.3c and d) overlapped with those affected by UV-exposure (compare Fig. 3.2.3a and b), although line broadening is not as strongly as in the UV-exposed spectrum. In particular, signal intensities of F-pocket residue T143 is significantly attenuated (Fig. 3.2.3c), together with other F-pocket residues N97, A117, indicating that tapasin induced altered F-pocket dynamics similar to the features seen for the MHC-I molecule devoid of the C-terminal dipeptide after UV-exposure. It has been shown that the F-pocket region of pMHC-I is important for its tapasin dependence[30]. In agreement with that we show here that residues lining the F-pocket were affected by the presence of tapasin, overlapping with many residues affected by an empty F-pocket. Thus, we suggest that tapasin promotes pMHC-I to adopt a peptide-receptive state by "loosening" the residues in the F-pocket, and eventually catalyzing peptide release in case that a competitor peptide is available.

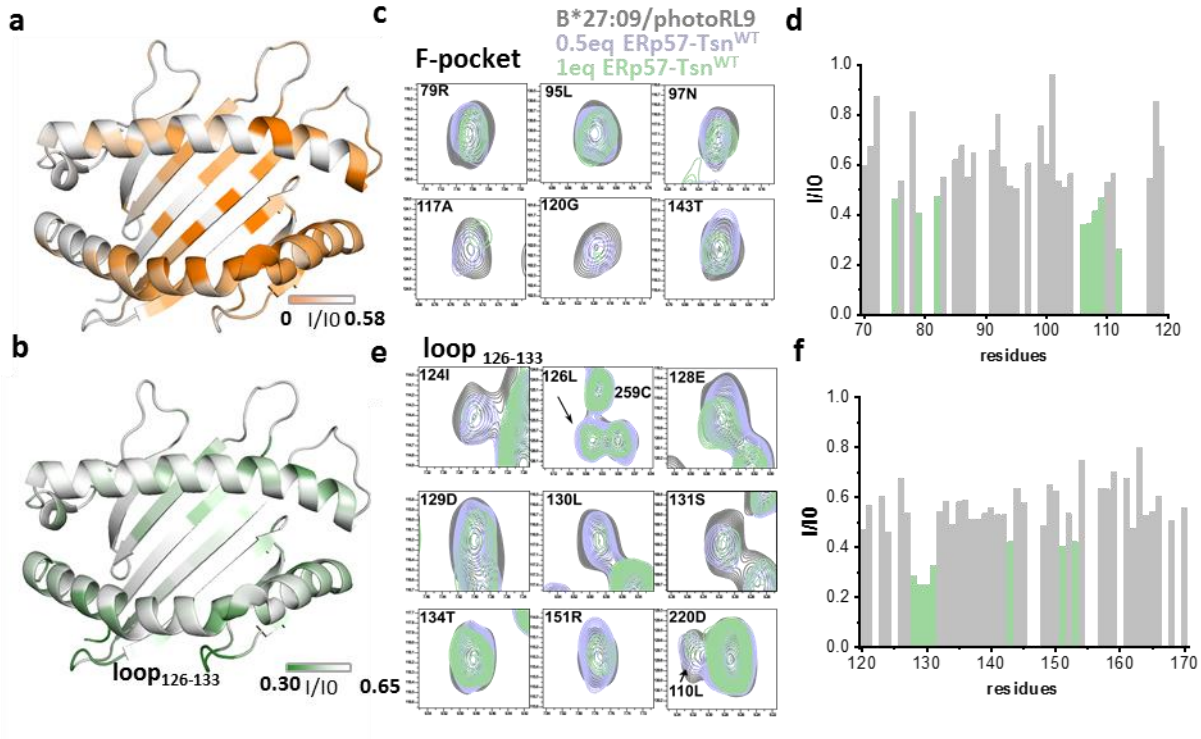


Figure 3.2.3 Tapasin "loosens" C-terminus of the bound peptide in the pMHC-I. a) Peak intensity ratio (I/I0) of cleaved B*27:09/photoRL9 (I) relative to non-cleaved B*27:09/photoRL9 (I0) or **b)** Peak intensity ratio (I/I0) of B*27:09/photoRL9 in the presence of ERp57-Tsn^{WT} (I) relative to B*27:09/photoRL9 alone (I0) is mapped as B-factor on the structure of B*27:09 (PDB ID IOF2). The range of the intensity ratio colored in **a)** is from 0 to the average I/I0 (0.58). The range of the intensity ratio in **b)** is from 0.3 to one of the average intensity I/I0 (0.65). Unassigned peaks were filled with its average I/I0 value. **c)** Representative peaks of residues in the F-pocket and **e)** in the loop$_{126-133}$ affected in the presence of ERp57-Tsn^{WT} (0.5eq in light blue, 1eq in green). Peak T134 and D220 was used as control to show non-affected residues. **d)** and **f)** Peak intensity ratio analysis (I/I0) in the presence of 1eq ERp57-Tsn^{WT} against B*27:09/photoRL9 (I0) are plotted against residue sequence. Unassigned peaks were left with empty value.

In addition to the F-pocket region, the loop$_{126-133}$ connecting strands β7-β8 was also significantly affected by addition of ERp57-Tsn^{WT} (Fig. 3.2.3d), which is in line with the observed interaction between this region with the hairpin loop$_{187-196}$ (loop connecting β-strands 9-10) of tapasin in the cryo-EM structure of the PLC(Fig. 1.10b)[178]. Notably, I124, L126 and L130, close or within the loop$_{126-133}$ that was affected by tapasin binding, also showed conformational exchange dynamics in the B*27:09/photoRL9 complex (Fig. 3.1.2d), suggesting that an alternative conformation of this loop might pre-exist in solution and could be recognized by tapasin. Indeed, mutations of residues (D128, W133) within the loop$_{126-133}$ have been previously shown to affect the MHC-I interaction with tapasin and it was suggested that the conformation of the loop$_{126-133}$ might be modulated by tapasin[148]. Thus, the loop$_{126-133}$ represents an important region for complex formation.

There are also some important differences between tapasin-induced and UV-exposure-induced changes in MHC-I: when ERp57-Tsn^{WT} was added, the entire $\alpha_{2\text{-}1}$-helix did not experience line-broadening as it is seen after UV exposure. While individual residues of the F-pocket seem to indicate "loosening", a full dissociation and destabilization of the $\alpha_{2\text{-}1}$-helix is not observed.

We also performed the experiment with ^{1}H-^{13}C-ILV-^{15}N-^{2}H B*27:05/photoRL9 in the presence of tapasin and observed more residues showing stronger signal reduction (Appendix Fig. 8b). In addition to T143, residue H114, N97 and L95 (these three residues are assigned in B*27:05/photoRL9, but not in B*27:09/photoRL9) located in the F-pocket were significantly affected, in line with the result that B*27:05 is more susceptible to the tapasin-mediated peptide exchange than B*27:09.

Thus, we provide the direct evidence of altered MHC-I dynamics in the presence of tapasin, which has been extensively studied by mutational and computational methods. Two regions of pMHC-I were identified to be affected by adding tapasin, which is consistent with contacts observed in the PLC structure[20] and a previous MD simulation of B*44:02/Tsn complex[150]. We speculate affected regions of pMHC-I are due to the interaction with two distinct regions of tapasin: One is the loop$_{187\text{-}196}$ connecting β-strands 9-10 of tapasin, that may recognize a dynamic region of MHC-I for binding and helps release the C-terminus of the peptide; And one, the so-called scoop loop$_{11\text{-}20}$, that may function in the second step, peptide exchange, by altering F-pocket dynamics. Since different MHC-I allotypes differ in their F-pocket, the latter structural element should explain the well-described allelic activity of the exchange catalyst.

3.2.4 Comparison between the interactions of pMHC-I-tapasin with pMHC-I-TAPBPR suggests different recognition mechanism

For comparison, the ^{1}H-^{15}N TROSY-HSQC spectrum of ^{1}H-^{13}C-ILV-^{15}N-^{2}H B*27:09/photoRL9 in the presence of TAPBPR was also recorded and the affected residues were analyzed. Similar to tapasin, no chemical shift changes but line-broadening was observed. A stronger intensity reduction in the binding groove of MHC-I upon TAPBPR addition was detected compared to tapasin binding. Especially residues at the bottom of the F-pocket were affected (Fig. 3.2.4c, compare the middle figure with the bottom figure), in agreement with the observed higher TAPBPR-mediated enhancement of peptide exchange on B*27:09 (Fig. 3.2.4a, Appendix Fig. 10b). Although the peak intensity of residues in the loop$_{126\text{-}133}$ were significantly reduced in the presence of TAPBPR, this was not the most affected region. Instead, the N-terminus of the α$_{2\text{-}1}$-helix, which was not affected in the presence of tapasin, was strongly affected by TAPBPR, indicating TAPBPR induces more dynamics in this region or interact differently with MHC-I. The same region was strongly affected upon peptide cleavage of B*27:09/photoRL9 (Fig. 3.1.4). It has been long suggested that the dynamics of the α$_{2\text{-}1}$-helix is important for peptide binding and exchange[150,170,208,209], and here we observed altered dynamics of α$_{2\text{-}1}$-helix in the presence of TAPBPR.

In contrast to tapasin, peak intensities of residues in the β-sheet$_{225\text{-}232}$ of the α3-domain of the heavy chain were strongly reduced in the presence of TAPBPR, which is in line with the observed interaction between Ig-like domain of TAPBPR and β-sheet$_{225\text{-}232}$ of the α3-domain in the crystal structure of the TAPBPR-MHC-I complex[173,174]. This implies the interaction between the IgG domain of TAPBPR with α3 domain of heavy chain was stronger than that of tapasin. Besides, residues in the kink$_{148\text{-}151}$ were not affected by TAPBPR although they were affected by tapasin. This might be explained by the contacts with tapasin helix$_{83\text{-}90}$ that is missing in TAPBPR (Fig. 3.2.4d).

To conclude, our data showed that different regions of MHC-I were affected by TABPPR in comparison to tapasin, suggesting a different recognition model compared to tapasin. Nevertheless, the dynamics of F-pocket region were affected by both, suggesting TAPBPR and tapasin share similar mechanism to promoting peptide exchange after binding to MHC-I: they are both likely to turn MHC-I to a peptide-receptive state by modulating the dynamics in the F-pocket region.

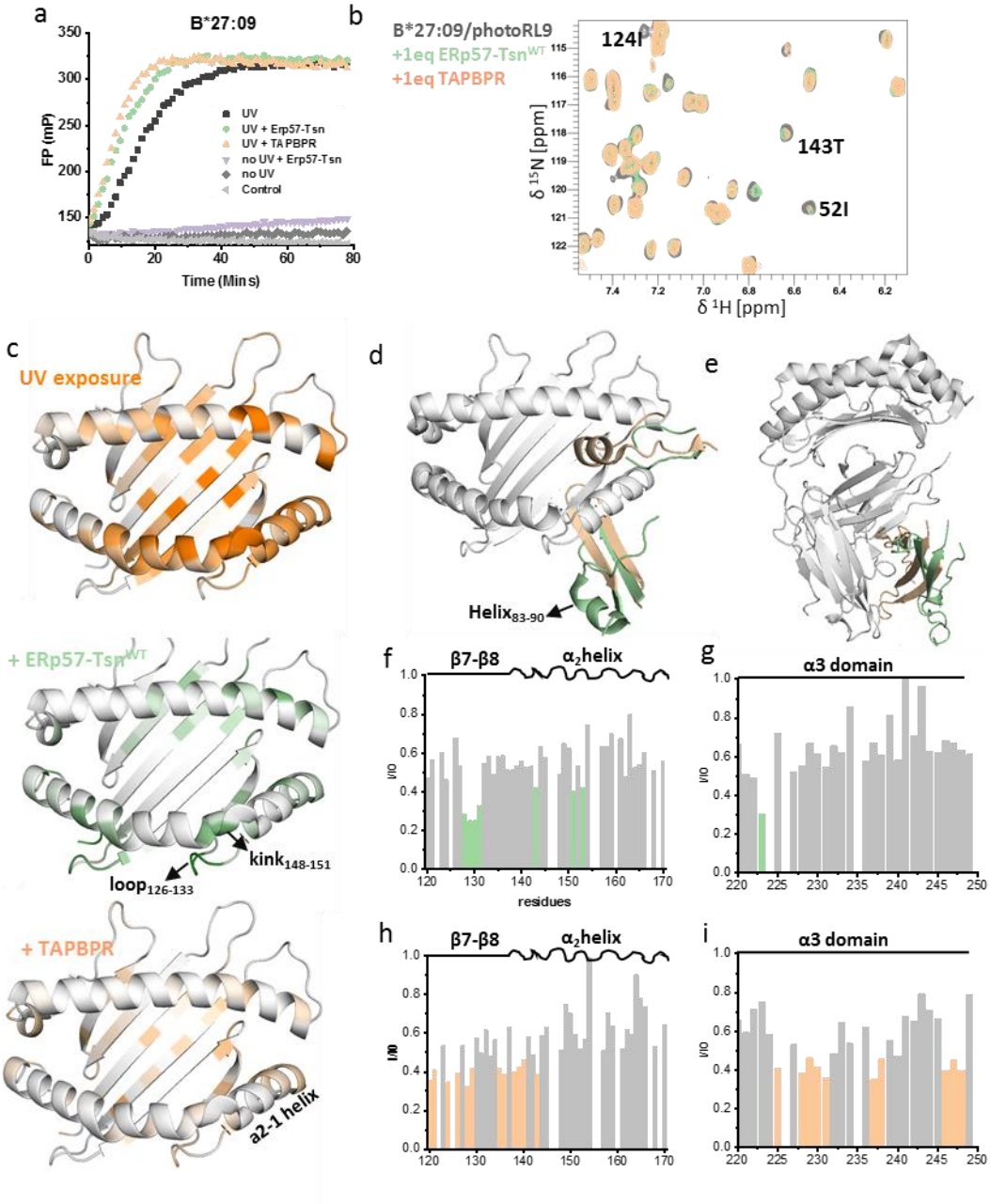


Figure 3.2.4 Comparison of the affected regions of pMHC-I by ERp57-Tsn^{WT} and TAPBPR binding. **a)** The peptide exchange profiles of 500 nM UV-cleaved B*27:09/photoRL9 in the absence (black) and presence of ERp57-Tsn^{WT} (green) or TAPBPR (wheat), as well as 500 nM non-UV-cleaved B*27:09/photoRL9 in the absence (grey) and presence of ERp57-Tsn^{WT} (purple). **b)** Representative zoom view of overlay of ^{1}H-^{15}N TROSY-HSQC spectrum of ^{1}H-^{13}C-ILV-^{15}N-^{2}H 50 μM B*27:09/photoRL9 (gray) and in the presence of equimolar ERp57-Tsn^{WT} (green) or equimolar TAPBPR (wheat). Peaks showing signal reduction are noted. **c)** Peak intensity ratio (I/I0, upper) of cleaved B*27:09/photoRL9 (I) against non-cleaved B*27:09/photoRL9 (I0), and peak intensity ratio (I/I0) in the presence of ERp57-Tsn^{WT} against B*27:09/photoRL9 alone (middle), and peak intensity ratio (I/I0) in the presence of TAPBPR against B*27:09/photoRL9 alone (bottom) are mapped as B-factor on the structure of B*27:09/RL9 (PDB ID IOF2). The strongly affected region of loop$_{126-133}$ was noted. The range of

the intensity ratio of UV-exposure is from 0 to the average I/I0. The range of the intensity ratio of the rest is from 0.5 to 1 of the average intensity ratio value. Unassigned peaks were filled with average I/I0 value. **d-e)** Overlay of MHC-I-tapasin (derived from PLC structure, PDB ID 6ENY) and MHC-I-TAPBPR complex (PDB ID 5OPI) to show interactions between tapasin or TABPPR and MHC-I. Only direct contact parts are displayed: **d)** The interaction between the N-terminal domain of tapasin (green) and TAPBPR (wheat) with MHC-I binding groove and **e)** The interaction between the IgG domain of chaperon and α3-domain of heavy chain. **f-i)** The peak intensity ratio analysis (I/I0) of corresponding region in the presence of tapasin and TAPBPR plotted against the sequence, significantly reduced intensity ratios are indicated by colored bars. Unassigned peaks are treated as 0.

3.2.5 The correlation between dependence on tapasin and MHC-I dynamics - dynamics analysis of four allotypes by CPMG

The above results raised another intriguing question: Which structural feature contributes to the different dependence of allotypes on tapasin or TAPBPR? A recent study pointed out the correlation of different MHC-I allotypes and their TAPBPR dependence by methyl-based NMR and suggests that TAPBPR exploits localized structural adaptations, both near and distant to the peptide-binding groove, to selectively recognize discrete conformational states sampled by MHC-I allotypes[169]. We speculated that tapasin might recognize a different conformation of MHC-I from the one recognized by TAPBPR considering one allotype showed different dependence on tapasin and TAPBPR. In order to verify the speculation, in addition to the already described CPMG analysis of B*27:09, we analyzed the dynamics of the remaining three allotypes by ^{1}H-^{13}C multiple-quantum CPMG NMR experiments as in section 3.1.2 (Fig. 3.2.5b). Unfortunately, the expression yield of the deuterated ^{13}C-ILV-A*02:01 heavy chain was very low and the sample quality of ^{1}H-^{13}C-ILV-^{15}N-^{2}H A*02:01/photoKV9 was low (indicated by crowded peaks in the middle of the ^{1}H-^{15}N HSQC spectrum, Appendix Fig. 5b). The dynamics data of A*02:01 was thus taken from the published data[179], in which methyl groups of the heavy chain ILAV residues were ^{13}C labelled. To make the cited dynamics data of A*02:01 more comparable with our data, we only considered the dynamics of ILV residues of A*02:01(Fig. 3.2.5b).

12 residues (Ala is not included) in A*02:01/NV9[179], 9 residues in A*03:01/photoKK9, 12 residues in B*27:09/photoRL9 and 5 residues in B*27:05/photoRL9 were observed to show conformational exchange dynamics. These residues are found mainly on the pleated β-strands of the groove, $\alpha_{2\text{-}2}$-helix and the loop$_{126\text{-}133}$ of heavy chain (Fig 3.2.5b). No residues in the α3 domain were found to show dynamics, which is different to the dynamics profile of the murine MHC-I[169,179].

Interestingly, when these allotypes are sorted into two groups based on the acidic or hydrophobic nature of the F-pocket, it seems there is a correlation between the allotypes' dynamics of β-strands 7-8 as well as the loop$_{126\text{-}133}$ connecting them and its tapasin dependence. For example, A*02:01 and B*27:09, both bear a hydrophobic F-pocket, three residues from this region (β-strands 7-8 and the loop$_{126\text{-}133}$) in B*27:09 show dynamics, while in A*02:01 only L130 from this region shows dynamics. Additionally, peptide exchange of B*27:09 is more susceptible to tapasin than that of A*02:01 (Fig. 3.2.3d). While for A*03:01 and B*27:05 that bear more acidic F-pocket, there are two residues in this region of A*03:01 showing dynamics while only one residue in B*27:05 showed dynamics. Thus, we speculate that the dynamics of the loop$_{126\text{-}133}$ and β-strands 7-8 are important for tapasin binding and plays a role in the allele-specific interaction with tapasin (Fig. 3.2.5c).

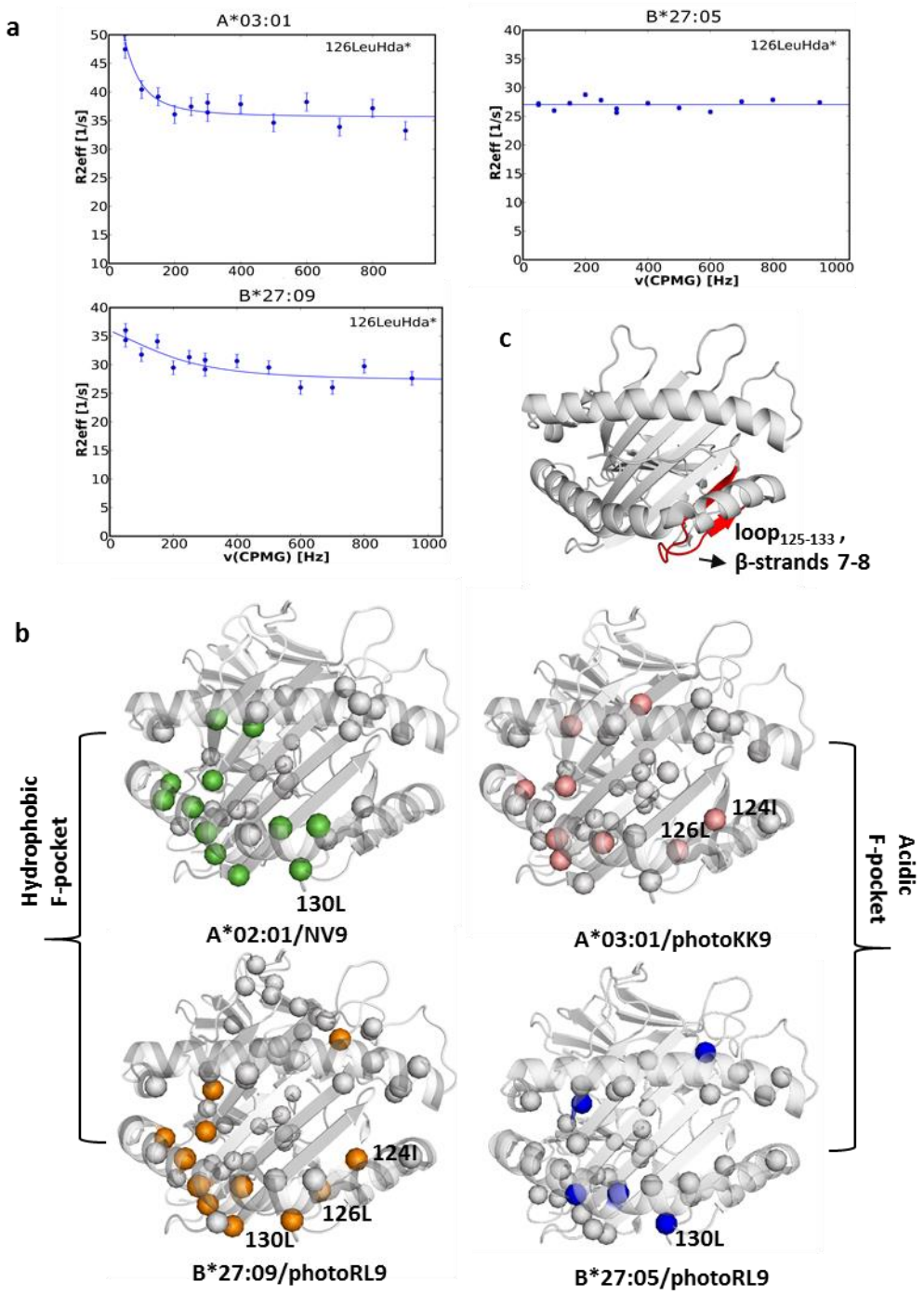


Figure 3.2.5 The dynamics data of pMHC-I obtained by methyl probe CPMG. a) Representative relaxation dispersion curves highlighting methyl groups undergoing chemical exchange ($R_2^{eff} > 2\ s^{-1}$) in ^{1}H-^{13}C multiple-quantum CPMG experiments recorded at 300K at a ^{1}H field of 700 MHz. The effective transverse relaxation rate (R_2^{eff}) is shown as a function of the CPMG pulse frequency (ν_{CPMG}) for selected methyl groups. **b)** Residues showing dynamics are highlighted as colored spheres on the crystal structure of the corresponding MHC-I. As published earlier, in A*02:01/NV9, residues V25, V28, V34, V103, L110, L130, V152, L156, L165, L168, L172, L179 show dynamics[179]. However, these data havebeen obtained at a different temperature, thus need careful interpretation. In A*03:01/photoKK9, residues I23, V28, V34, V103, I124, L126, L160, V165, L179 show dynamics. In B*27:09/photoRL9, residues V12, V25, V28, L32, I52, L82, V103, L110, I124, L126, L130, L160, V165, L168, L179 show dynamics. In B*27:05/photoRL9, residues V12, V28, L32, V34, L130, L160, V165 show dynamics. All other ILV methyl groups not showing dynamics are mapped on the structure as gray spheres. **c)** The region (β-strands 7-8 as well as the $loop_{126-133}$ connecting them) that might be important for tapasin binding is colored in red.

3.3 The function of the tapasin loop$_{11\text{-}20}$ in *in vitro* peptide exchange

The aligned structure of ERp57-Tsn conjugate (3U8F) with MHC-I crystal structure (2XPG) based on the recent cryo-EM structure of the PLC implies that the loop$_{11\text{-}20}$ of tapasin interacts with the F-pocket region of the binding groove (Fig. 3.3.1a). However, residues within the loop$_{11\text{-}20}$ are not resolved in the crystal structure of ERp57-Tsn conjugate[145], suggesting an increased flexibility of it. The corresponding loop$_{22\text{-}36}$ in TAPBPR, which is longer than the tapasin loop$_{11\text{-}20}$, was demonstrated to play an important role in promoting peptide dissociation of MHC-I molecules[175,176]. NMR results in section 3.2.3 showed altered dynamics of the F-pocket in the presence of tapasin. Thus, we speculated that the loop$_{11\text{-}20}$ of tapasin plays a role in mediating peptide exchange. Based on these speculations, different mutants (Fig. 3.3.1b and Fig. 3.3.2a) were designed to explore the function of the tapasin loop$_{11\text{-}20}$ in peptide exchange. The expression of all mutants was successful (Appendix Fig. 9). Since different MHC-I allotypes differ in their F-pocket, the interplay between loop$_{11\text{-}20}$ and the polymorphic F-pocket region should explain the observed allotype-specific catalytic activity of tapasin.

3.3.1 L18 is the key residue in the loop$_{11\text{-}20}$ that mediates peptide exchange

L30 in the loop$_{22\text{-}36}$ of TAPBPR was shown to be the key residue to promote peptide dissociation[175]. In the tapasin loop$_{11\text{-}20}$, the side chains of two residues, Lys16 and Leu18, are potentially long enough to reach the F-pocket of MHC-I molecules to promote peptide dissociation. Two mutants, ERp57-Tsn^{L18G} and ERp57-Tsn^{K16G}, were generated to identify which of these two residues is the key residue within the tapasin loop$_{11\text{-}20}$ (Fig. 3.3.1b). Glycine was chosen as a substitute to allow for maximum flexibility of the loop, potentially allowing the remaining key residue to reach into the F-pocket. The stability of mutants was similar to ERp57-Tsn^{WT} (Appendix Fig. 9). Since tapasin did not increase peptide exchange of A*02:01, the remaining three allotypes investigated here (A*03:01, B*27:05, B*27:09) were used to check the activity of Tsn mutants on catalyzing peptide exchange. Peptide exchange rates mediated by mutants were measured and the relative activity of the mutant to catalyze peptide exchange compared to wild-type tapasin were calculated by the formula $(Kon\_Tsn^{mutant} - Kon\_Tsn^{-})/(Kon\_Tsn^{WT} - Kon\_Tsn^{-})$.

As shown in Fig. 3.3.1c, the activity of ERp57-Tsn^{L18G} in catalyzing the peptide exchange of these three allotypes was strongly reduced (0.38 ± 0.05 for A*03:01, 0.07 ± 0.05 for B*27:05 and 0.09 ± 0.04 for B*27:09), indicating that L18 is critical for tapasin catalytic activity. In contrast to that, the catalytic activity of ERp57-Tsn^{K16G} on peptide-exchange was 0.73 ± 0.15 for A*03:01, 0.86 ± 0.05 for B*27:05 and 0.96 ± 0.10 for B*27:09 (Fig. 3.3.1c, blue bars), indicating that mutating K16 did not cause severe activity loss to tapasin. This also indicates a negative correlation with the gradient of acidity of the allotypes' F-pocket (Fig. 3.2.2a).

Taken together, for B*27:05 and B*27:09, the activity of ERp57-Tsn was nearly abolished by the L18G mutation, resulting in the peptide exchange rate nearly equal to that of uncatalyzed MHC-I (black curve, Fig. 3.3.1c), while the K16G mutation did not significantly impair the activity of ERp57-Tsn. Thus, for the B*27:05 and B*27:09 allotypes, L18 but not K16 appears to be important for tapasin catalytic activity. In contrast to B*27:05 and B*27:09, the peptide exchange of A*03:01 was still enhanced by ERp57-Tsn^{L18G} (Fig. 3.3.1c), indicating that K16 can rescue the activity of ERp57-Tsn^{L18G} for A*03:01

to a certain extent. Concomitantly, ERp57-Tsn^{K16G} showed significantly reduced activity (0.73 ± 0.15) for A*03:01, suggesting that K16 is also important for the catalytic activity of tapasin in case of the A*03:01 allotype.

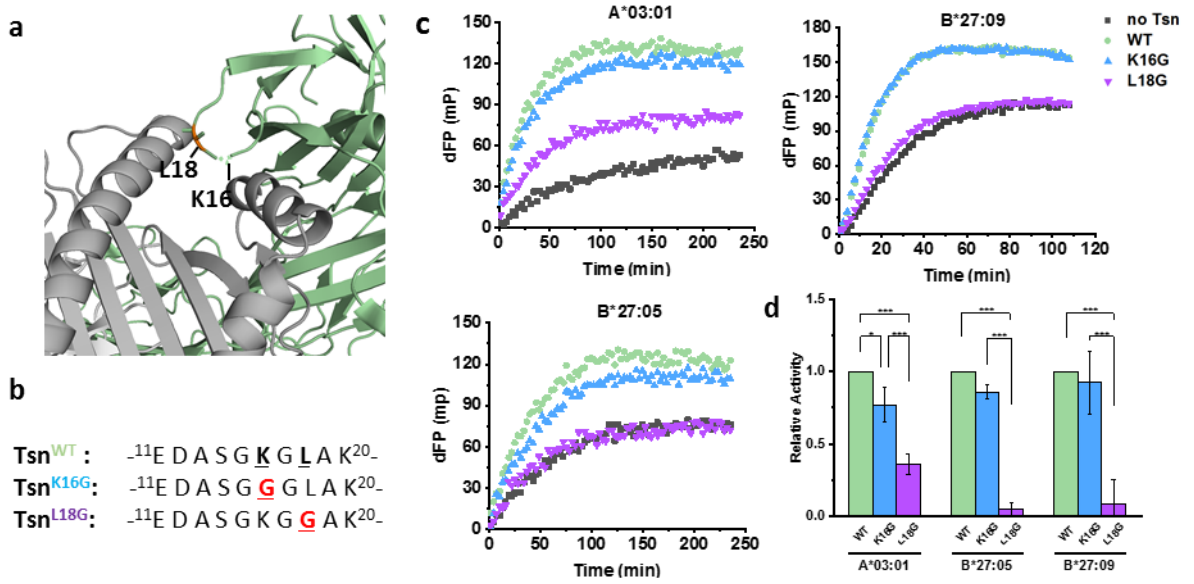


Figure 3.3.1. Design of loop$_{11-20}$ mutants and the importance of Leu18 in peptide exchange of MHC-I. a) Zoom-in view of the N-terminal domain of tapasin (light green) interacts with C-terminus of MHC-I binding groove (gray). Missing loop is drawn as dash line with K16 and L18 noted. **b)** The sequence of the loop$_{11-20}$ mutants. **c)** Peptide exchange profiles of 500 nM pMHC-I of three allotypes in the absence (black) and presence of equimolar of ERp57-Tsn^{WT} (green), ERp57- Tsn^{L18G} (blue), ERp57-Tsn^{K16G} (purple). dFP means baseline was subtracted. **d)** The *Kon* were determined for each mutant on the corresponding MHC-I allotype and relative activity was calculated by the formula of (*Kon_Tsn^{mutant}* - *Kon_Tsn^{-}*)/(*Kon_Tsn^{WT}* - *Kon_ Tsn^{-}*). Error bar (SD) was calculated from three independent experiments. For comparison of loop mutants to WT, the two-sample unequal variance Student's t-test was performed (*p < 0.05, **p < 0.01, ***p < 0.001).

3.3.2 K16 plays an auxiliary role for pMHC-I bearing an acidic F-pocket

Tapasin showed the highest efficiency in accelerating peptide exchange for A*03:01 which bears the most acidic F-pocket among the four allotypes tested, implying there is a correlation between the tapasin activity on a certain MHC allele with its F-pocket acidity. As mentioned above (Fig. 3.3.1d), the presence of K16 in the tapasin loop appears to play a role. Thus, we speculated K16 in the loop$_{11-20}$ contributes to accelerate peptide exchange of HLA allotype bearing a negative charged F-pocket. In order to test this hypothesis, more variants of tapasin with different combinations of leucine or lysine in position 16 and 18 on the loop$_{11-21}$ were designed: K16L (SGLGL), L18K(SGKGK), G^{14}GG^{16}GL, G^{14}GG^{16}GK^{18}, G^{14}GL^{16}GG^{18} and G^{14}GK^{16}GG^{18} (sequences are listed in Fig. 3.3.2a, stability data is listed in Appendix Fig. 9). The A*03:01 allotype bearing the most acidic F-pocket (preference for basic residues) and B*27:09 bearing the least acidic F-pocket (preference for hydrophobic residues) were chosen to test the catalytic activity of these tapasin mutants.

As can be seen in Fig. 3.3.2b, the ERp57-Tsn^{K16L} mutant showed no significantly different catalytic activity compared to ERp57-Tsn^{WT} for A*03:01 (0.93 ± 0.16 of WT) and increased catalytic activity for B*27:09 (1.26 ± 0.06 of WT). Whereas ERp57-Tsn^{L18K} displayed no activity for both allotypes. The relative activity of ERp57-Tsn^{GGGGL} was 0.45 ± 0.12 for A*03:01 and 0.77 ± 0.14 for B*27:09, while

the relative catalytic activity of ERp57-Tsn^{GGGGK} was 0.23 ± 0.05 for A*03:01 and showed no catalytic activity for B*27:09. However, ERp57-Tsn^{GGKGG} and ERp57-Tsn^{GGLGG} showed an opposite behaviour in catalysing peptide exchange of A*03:01 and B*27:09: ERp57-Tsn^{GGKGG} can enhance peptide exchange of A*03:01 as efficiently as ERp57-Tsn^{WT}, while ERp57-Tsn^{GGLGG} can only slightly enhance the peptide exchange of A*03:01 (with 0.36 ± 0.05, the activity was similar to that observed for ERp57-Tsn^{L18G}, 0.38 ± 0.05 of WT). In contrast, ERp57-Tsn^{GGKGG} was not able to promote peptide exchange of B*27:09 at all, while the activity of ERp57-Tsn^{GGLGG} on B*27:09 was the same as ERp57-Tsn^{WT}. These results indicate that the exact position of K16 and critical L18 residues is malleable to a certain extent and the contribution of K16 is distinct for the two allotypes.

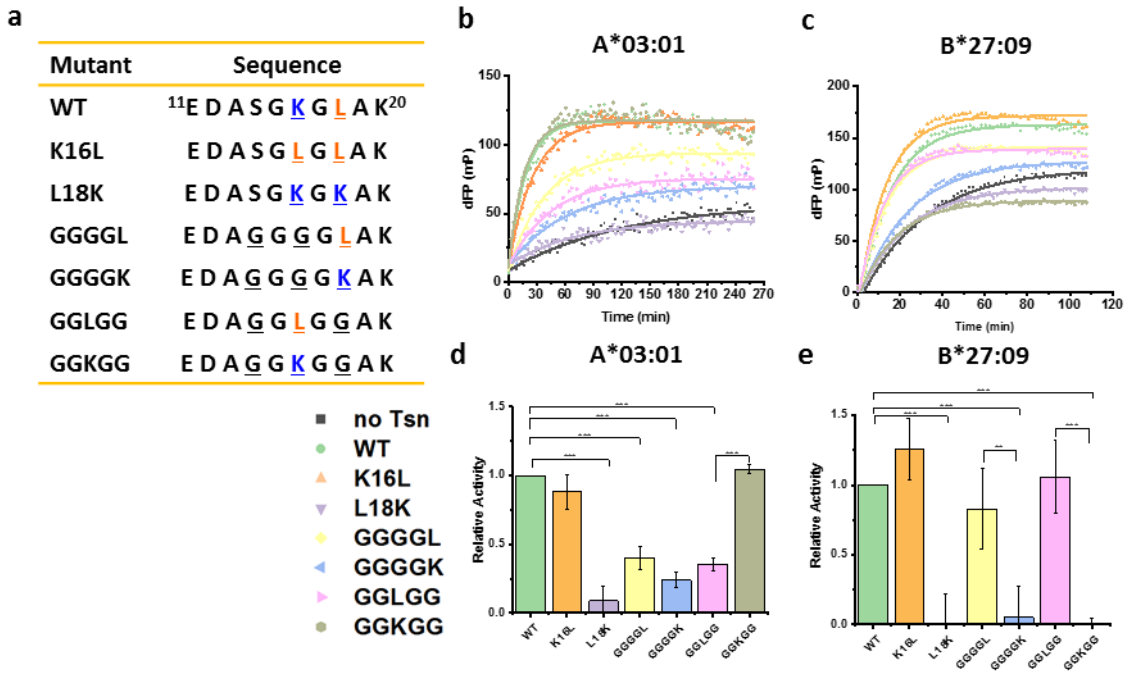


Figure 3.3.2. The role of K16 on tapasin loop$_{11-20}$ mutants on mediating peptide exchange of two allotypes. a) The sequence of the loop mutants. **b)** and **c)** Peptide exchange profiles of 500 nM pMHC-I of A*03:01 and B*27:09 in the absence (black) and in the presence of equimolar of ERp57-Tsn loop mutant. The y-axis (dFP) shows baseline was subtracted. **d)** and **e)** Relative activity of other loop mutants compared to ERp57-Tsn^{WT} towards A*03:01 and B*27:09 was shown in bar chart. Error bars (SD) were calculated from 3 independent experiments. For the comparison of loop mutants to WT, the two-sample unequal variance Student's t-test was performed (*p < 0.05, **p < 0.01, ***p < 0.001).

The catalytic activity of the ERp57-Tsn^{K16L} mutant, which has two leucine residues in the loop, was slightly higher than ERp57-Tsn^{WT} for B*27:09, suggesting position 16 can contribute to the activity of tapasin. This again demonstrates that the role of the residue at position 16 strongly depends on the polarity of the F-pocket. A similar dichotomy was observed for position 18, albeit not as pronounced as for position 16. The opposite activity of ERp57-Tsn^{GGKGG} and ERp57-Tsn^{GGLGG} to catalyse peptide exchange of A*03:01 and B*27:09 suggests that when there is no Leu at position 18, only one Leu or Lys at position 16 in the loop is sufficient for tapasin to maintain its peptide exchange function. The activity of ERp57-Tsn^{GGLGG} was higher than that of ERp57-Tsn^{GGGGL} for B*27:09, demonstrating the importance of the loop's mobility, indicating that the residue at position 16 can stretch into the F-pocket better than Leu in position 18 in the context of a flexible loop.

3.3.3 The interaction of ERp57-Tsn^{L18G} with MHC-I indicates critical structural elements

Next, in order to investigate if scoop loop mutants differently affect the interaction of tapasin with pMHC-I, the affected residues of B*27:09/photoRL9 upon addition of ERp57-Tsn^{GGLGG} and ERp57-Tsn^{L18G} mutants were analysed by NMR. Similar to ERp57-Tsn^{WT}, line-broadening was observed in the presence of both mutants and the affected residues in the binding groove by ERp57-Tsn^{GGLGG} and ERp57-Tsn^{L18G} overlapped largely with residues affected by ERp57-Tpn^{WT} (Fig 3.3.3b). Thus, both mutants still interact with p-MHC-I. Interestingly, when comparing the NMR line-broadening profiles of them, we observed enhanced dynamics comprising the N-terminal end of the $\alpha_{2\text{-}1}$-helix in the presence of the L18G mutant, but not in the presence of the other two (WT and GGLGG). WT and GGLGG showed the same catalytic activity on B*27:09. For WT tapasin and GGLGG mutant, it is the loop$_{126\text{-}133}$ of MHC-I that displays higher than average dynamics, but not the $\alpha_{2\text{-}1}$-helix. In contrast, for the L18G mutant, a larger region of the heavy chain binding pocket, extending into the $\alpha_{2\text{-}1}$-helix, was affected, implying that the Leu residue in the loop$_{11\text{-}20}$ stabilizes the F-pocket and prevents enhanced dynamics or local unfolding of the critical $\alpha_{2\text{-}1}$-helix structural element.

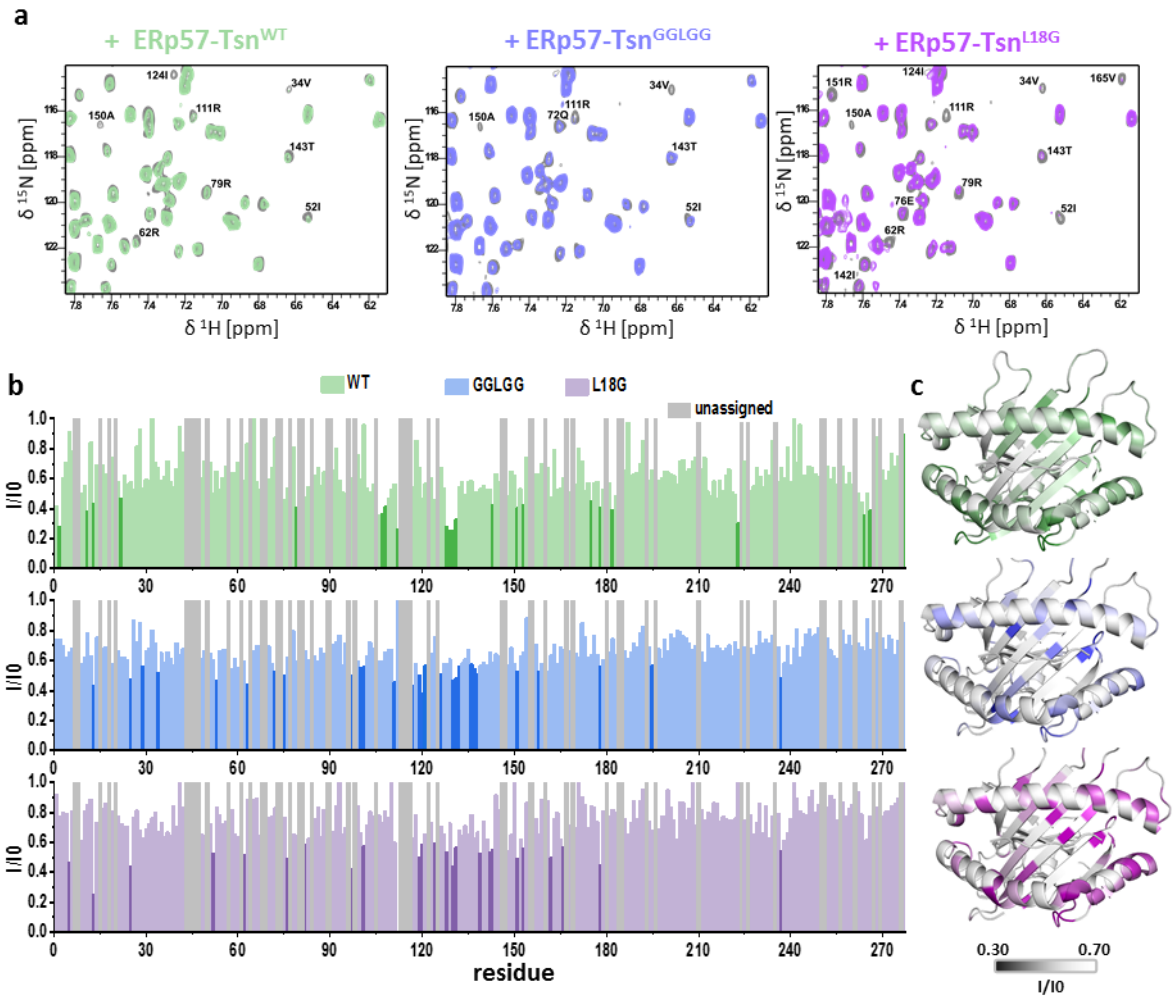


Figure 3.3.3 Analysis of interactions between ERp57-Tsn mutants and pMHC-I. a) Representative zoom view of overlay of ^{1}H-^{15}N TROSY-HSQC spectra of 50 µM B*27:09/photoRL9 (gray) and in the presence of equimolar ERp57-Tsn^{WT} (green), ERp57-Tsn^{GGLGG} (blue) and ERp57-Tsn^{L18G} (purple). Attenuated signal of residues is marked out. **b)** Peak intensity ratio

analysis (I/I0) of B*27:09/photoRL9 in the presence of ERp57-Tsn^{WT}, ERp57-Tsn^{L18G} and ERp57-Tsn^{GGLGG} against B*27:09/photoRL9 (I0) are plotted against residue sequence. Significantly reduced residues (below $I/I0_{average}$- δ) were highlighted in corresponding dark color. Unassigned residues are plotted as gray bars. **c)** Peak intensity ratio (I/I0) are mapped as b-factor on the structure of B*27:09. The range of the intensity ratio is from 0.3 to 0.7. The color code is the same as in **a**.

Unexpectedly, residue T143 was also affected by adding ERp57-Tsn^{L18G} (Fig. 3.3.3c), which suggests that a mutation in the $loop_{11\text{-}20}$ does not significantly affect tapasin interacting with pMHC-I's F-pocket region, even though L18 is critical to the catalytic activity of tapasin. It is well known that Y84 and T143, form essential H-bonds with the carboxyl group of the bound peptide in pMHC-I[210]. And we also observed this H-bond in the crystal structure of B*27:09/photoRL9 (Table 3.1). In the TAPBPR-MHC-I complex, the side chain of Y84 was seen to swing out of the binding groove, probably allowing the scoop loop to be accommodated instead, and stabilized by an H-bond to E102 of TAPBPR. Interestingly, mutating E72 of Tsn, which corresponds to E105 of TAPBPR, impaired its MHC-I binding and reduced catalytic activity (50% of wild-type)[145]. Thus, we suggest that the attenuation of signal T143 was due to the interaction of the Y84 side chain with E72 of tapasin, which interrupts the H-bond network in the F-pocket. It also indirectly implies the $loop_{11\text{-}20}$ functions as a peptide competitor for the F-pocket after the H-bond network in the F-pocket is interrupted by E72.

3.3.4 Rosetta energy calculations confirmed the important role of L18 and K16.

Additionally, we used the cryo-EM structure of the PLC (PDB ID 6ENY) as a scaffold to model an ensemble of MHC-I-ERp57-Tsn complex structures to illustrate potential conformations of $loop_{11\text{-}21}$ of tapasin (Fig. 3.3.4a) and calculated the Rosetta energy scores of the MHC-I-ERp57-Tsn complex models with the desired scoop loop mutation for each allotype[196–198].

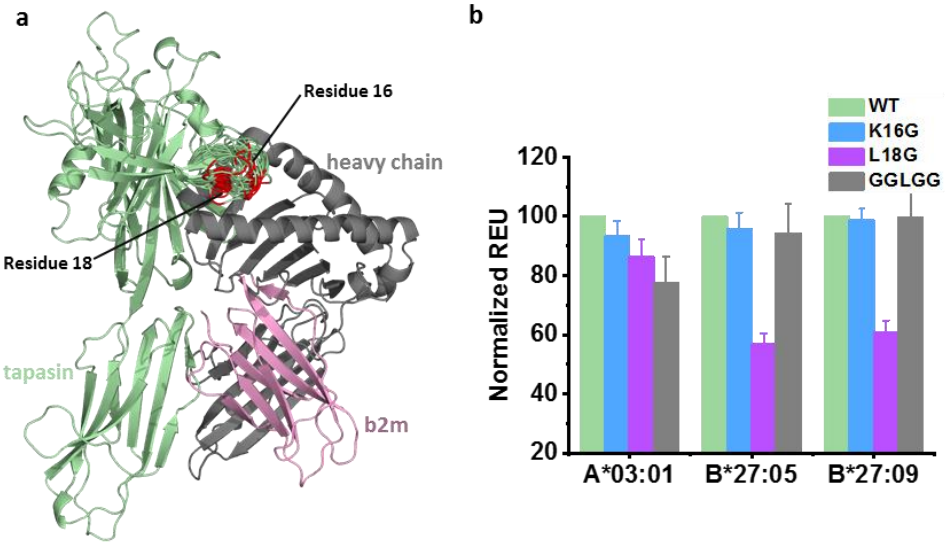


Figure 3.3.4 Modeling the interaction between tapasin mutants with pMHC-I by Rosetta. a) The $loop_{11\text{-}20}$ of tapasin was modeled in the MHC-I-ERp57-Tsn-ERp57 complex (ERp57 was not shown here). Residue 16 and 18 was highlighted in red to show their interaction with the F-pocket of MHC-I. **b)**. Normalized free energy of different tapasin loop mutants calculated by Rosetta free energy value.

The calculation was done by Dr. Esam T. Abualrous. The Rosetta energy scores were averaged for all generated models per mutant. As shown in Fig. 3.3.4b, the relative Rosetta energy scores (rREU) of all three MHC-I allotypes in complex with ERp57-Tsn^{K16G} was only slightly reduced (93% for A*03:01,

95% for B*27:05 and 98% for B*27:09 as compared to ERp57-Tsn^{WT}), suggesting a marginal impact of the K16G mutation. In contrast, rREU of the MHC-I-ERp57-Tsn^{L18G} complex was reduced to 56% for B*27:05 and 60% for B*27:09, indicating the important role of L18 to stabilize these complexes. Interestingly, for A*03:01 rREU of pMHC-I-ERp57-Tsn^{L18G} was still 86% that of wild type, this again indicates K16 is important for tapasin function. While the rREU of MHC-I-ERp57-Tsn^{GGLGG} complex was 78% for A*03:01, 95% for B*27:05 and almost 100% for B*27:09, which again prove that one Leu in position 16 is enough to stabilize the interaction for allotypes bearing hydrophobic F-pockets when the loop is mobile.

Taken together, our data highlight the critical role that L18 plays for tapasin interaction with pMHC-I while K16 plays an auxiliary role that is of importance for allotypes bearing an acidic F-pocket as exemplified here by A*03:01. Also, the importance of the flexibility the tapasin scoop loop is again illustrated here.

3.3.5 The role of K16 is indicative of co-evolution between tapasin and MHC-I molecules

Based on the above result that the tapasin loop composition and flexibility is important for catalysis, we then asked whether the diversification of MHC-I alleles in other organisms coincides with tapasin scoop loop adaptations. We compared the tapasin scoop loop sequence of different species and found that L18 is conserved among species (Fig. 3.3.5a), This is in agreement with our result that L18 is the key residue to mediate peptide exchange, and also in line with a previous study showing that mutating corresponding Leu in the mouse tapasin loop significantly reduced the surface expression level of H2-K^{b} [34]. In addition to the conserved Leu, we are surprised to observe that non-rodent mammal species always have a basic residue (Lys or Arg) in its tapasin scoop loop (Fig. 3.3.5a). Could the emergence of the basic residue in the scoop loop of other species implies a co-evolution between tapasin and MHC-I molecules?

In order to answer that question, we categorized available MHC-I allotypes (in database IPD-IMGT/HLA) of different species into five groups based on the F-pocket nature (predicted by the netMHCpan motif view function) and could find a match between the percentage of MHC-I allotypes bearing an acidic F-pocket and the presence of a basic residue in the tapasin scoop loop (Fig. 3.3.5b). For example, 15% of HLA, 17% of Patr (chimpanzee MHC-I), 5% of Mamu (monkey MHC-I), 5% of BoLA (bovine), 6% of Ovar (sheep), 28% of FLA (cat), 14% of DLA (dog) bear an acidic F-pocket, which coincides with the presence of a basic residue in the corresponding tapasin scoop loop. In contrast, for the pig MHC-I allotype (SLA), 65% of SLA can accommodate polar residues (mainly Tyr) in its F-pocket, which coincides with the presence of the polar Gln in position 17 of the pig tapasin scoop loop. Meanwhile, none of reported H2 allotypes (murine MHC-I allotype) bears an acidic F-pocket, which coincides with no basic residue in the mouse tapasin scoop loop.

We also analyzed the chemical nature of MHC-I heavy chain residue 116 (as well as other F-pocket residues), which locates to the bottom of the F-pocket and is one of the key determinants for the F-pocket nature, of MHC-I in different species. We found that the percentage of acidic amino acids in position 116 of MHC-I in non-rodent mammals matches with the presence of a basic residue in the corresponding scoop loop (Fig. 3.3.5c). For more information, we categorized a large data set of

available MHC-I presented peptides of different species into four groups based on the nature of the C-terminal residue and found a good correlation between the percentage of allotypes with an acidic F-pocket and the percentage of peptides bearing a C-terminal basic residue (Fig. 3.3.5d).

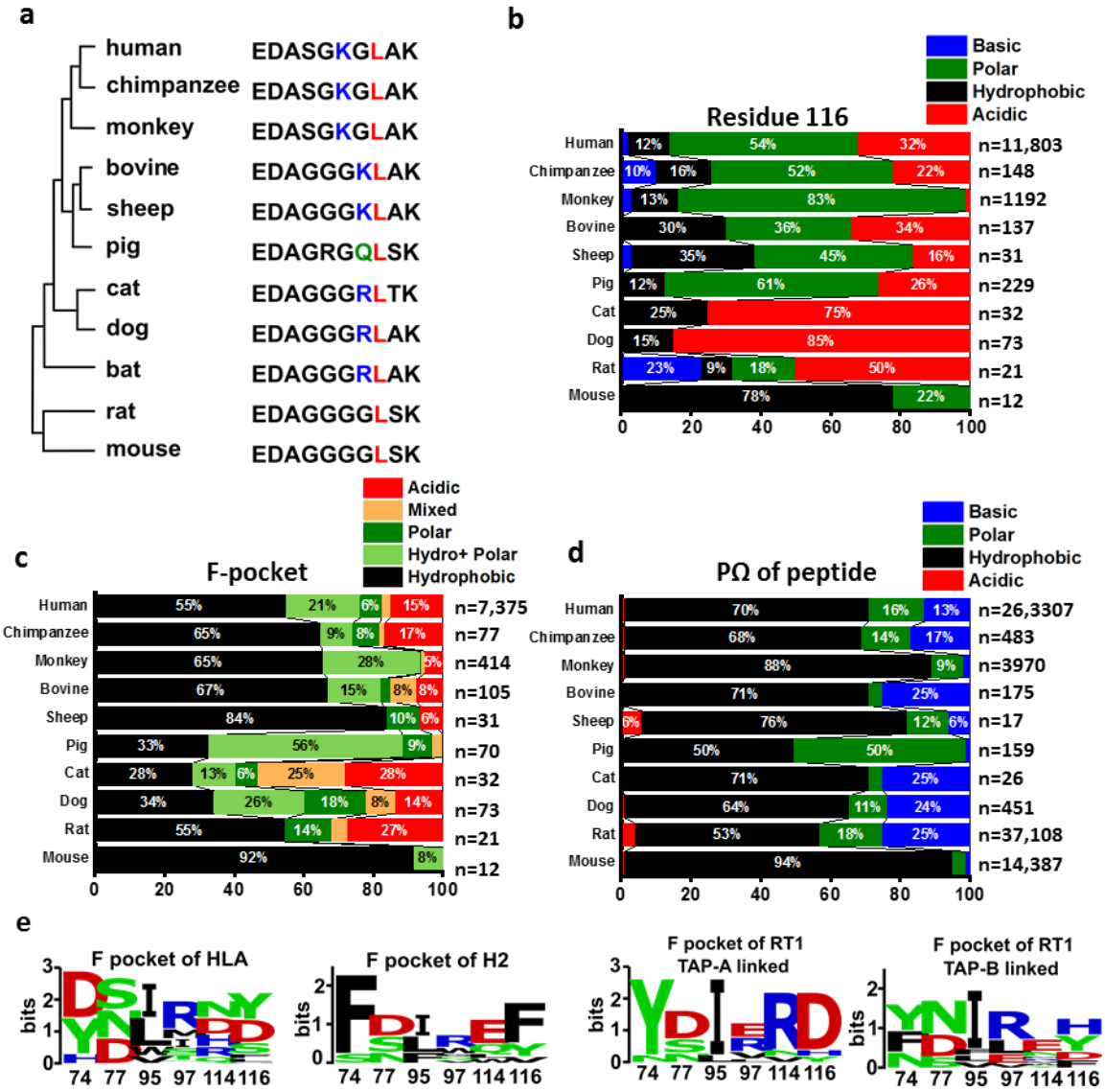


Figure 3.3.5 Co-evolution relationship between tapasin and MHC-I molecules. a) Evolution tree of tapasin from common species and the alignment of corresponding scoop loop sequence. Basic residue (K and R) is highlighted in blue, L is highlighted in red, and polar residue (Q) is highlighted in olive. **b)** Percentage of residue 116 chemical nature in MHC-I of different species. **c)** Percentage of the F-pocket chemical nature in MHC-I different species. The F-pocket's chemical nature are characterized into five groups based on its preferred binding residue. Hydrophobic: prefers hydrophobic residues; Hydro + Polar: prefers both hydrophobic and polar residues (mainly Y); Polar: prefers polar residue (mainly Y); Mixed: can accommodate hydrophobic, polar and basic residues; Acidic: prefers basic residues (R and K). Data used for statistics was collected from NetMHCPan-4.1 motif viewer function (when the allele is available in the website, http://www.cbs.dtu.dk/services/NetMHCpan-4.1/logos_ps.php) or based on the F-pocket residues composition. **d)** Percentage of presented peptide chemical nature in different species. Peptides were categorized into four groups based on the C-terminal amino acid (PΩ) chemical nature. Presented peptides used for calculation were downloaded from IEDB and filtered by length between 8-11 aa. In **b-d**: n stands for the sample number used for calculation. Percentage values below 5% are not labelled in the bar. **e)** Sequence logo of important residues in the F-pocket of human, mouse and rat. The rat MHC-I was divided into two groups by the linked TAP type.

The only exception is rat, which has no basic residue in the scoop loop while 27% of RT1 (rat MHC-I) have an acidic F-pocket and 24% of presented peptides bear a C-terminal basic residue. How to explain the exception of rat? It is already known that the peptide pool in the ER for MHC-I loading is shaped by the TAP transporter, which imports peptides into the ER based on the C-terminal residue of the peptide[211]. In mouse, the TAP transporter preferentially transports peptides bearing a hydrophobic residue and all murine MHC-I allotypes preferentially accommodate hydrophobic residue in their F-pocket. In human, the TAP transporter displays a broad specificity, including those with positively charged C-terminal residues. However, in rat, there are two allelic variants of TAP: TAP-A, which resembles human TAP, is genetically associated with MHC-I alleles bearing an acidic F-pocket (six out of eleven, Appendix Table 4). In contrast, TAP-B is genetically associated with MHC-I alleles with a hydrophobic F-pocket and transports peptides bearing a hydrophobic C-terminus, analogous to mouse TAP. Indeed, the two rat TAP transporters have been reported to have coevolved with rat MHC-I alleles[212,213]. It seems two routes of co-evolution have developed, with most species having adapted their tapasin $loop_{11\text{-}20}$ to efficiently catalyze peptide exchange for newly emerging MHC-I allotypes (Fig. 4.6b).

3.4 The Structure of MHC-I-ERp57-Tsn complex

Although the cryo-EM structure of the PLC provides insights into the overall architecture of the PLC, the molecular mechanism of tapasin-mediated peptide exchange is not clear due to the low resolution of the structure. Thus, a high-resolution structure of the PLC or MHC-I-ERp57-Tsn complex is needed for better understanding of the mechanism. TAPBPR can form a tight complex with murine MHC-I. However, although Morozov *et al.* showed that by the native SDS-PAGE that A*02:01 can form a complex with TAPBPR after UV-exposure[89,176], we cannot repeat their results when we analyzed a mixture of UV-exposed A*02:01/photoKV9 with TABPPR by gel filtration, as shown in Appendix Fig. 11h-i). Unlike TAPBPR, the affinity between ERp57-Tsn and MHC-I molecules is low, as indicated by the NMR results and FP assays. Trials to assemble MHC-I-ERp57-Tsn complex by incubating ERp57-Tsn with UV-exposed MHC-I failed (shown in Appendix Fig. 11a-d), even when A*03:01 was used, which showed the highest susceptibility to tapasin in this work. Considering the cleaved 7-mer peptide was still in the binding groove, which might hinder tapasin binding, MHC-I refolded with peptide with a cleavage site at the 4th position was used (photoRL9-4*, Table 2.2). Unfortunately, these conditions still yielded no complex (Appendix Fig. 11e-f). Ultimately, we considered introducing a disulfide bond to stabilize the weak interaction between MHC-I molecules and ERp57-Tsn.

3.4.1 Assembly of MHC-I-ERp57-Tsn complex by introducing a disulfide bond

Based on the observed contacts between Tsn and MHC-I in the cryo-EM structure of the PLC, three pairs of cysteine mutations (i.e., $\beta2m^{T5C}$~Tsn^{G304C}, MHC-I^{S132C}~Tsn^{S82C} and MHC-I^{Q226C}~Tsn^{H335C}, Fig. 3.4.1a) were designed to test the possibility of stabilizing the MHC-I-ERp57-Tsn complex via a disulfide bond. Mutant ERp57-Tsn^{G304C}, ERp57-Tsn^{S82C} and ERp57-Tsn^{H335C} were successfully expressed, but the expression yield of ERp57-Tsn^{G304C} was only 10% of the ERp57-Tsn^{WT}. The mutants pMHC-I/$\beta2m^{T5C}$ and pMHC-I^{Q226C} were successfully refolded, but the refolding of pMHC-I^{S132C} mutant

failed. Thus, two pairs of mutants (β2m^{T5C}~Tsn^{G304C}, MHC-I^{Q226C}~Tsn^{H335C}) were used to perform the complex assembly experiments.

In order to test the complex formation with the β2m^{T5C}~Tsn^{G304C} pair mutation, A*03:01 refolded with β2m^{T5C} and photocleavable peptide (e.g., A*03:01/β2m^{T5C}/photoKK9) was UV-exposed and incubated with equimolar ERp57-Tsn^{G304C} overnight. Incubated mixture was then analyzed by SDS-PAGE under non-reducing condition. As shown in Fig. 3.4.1b, a new band higher than the band of ERp57-Tsn^{G304C} showed up and the intensity of this band was stronger than that in the sample without UV-exposure (compare lane 3 and lane 6 in Fig. 3.4.1b). Based on the apparent molecular weight, we speculated that the new band is the complex between ERp57-Tsn^{G304C} and β2m^{T5C}. MS analysis of this band confirmed peptides from β2m, indicating that β2m^{T5C} in A*03:01/β2m^{T5C}/photoKK9 formed a disulfide bond with Tsn^{G304C}. Around 80% of ERp57-Tsn^{G304C} formed a complex with the UV-exposed A*03:01/β2m^{T5C}/photoKK9 (lane 6 in Fig. 3.4.1b). The reaction mixture was then analyzed by gel filtration, a peak eluted at 11.78 ml, which is earlier than the peak of ERp57-Tsn^{G304C} conjugate alone (12.75 ml). The fraction of the peak$^{11.78\ ml}$ showed three bands and four bands on non-reducing and reducing SDS-PAGE (Fig. 3.4.1d), respectively. This conformed that the peak$^{11.78ml}$ is the A*03:01/β2m^{T5C}-ERp57-Tsn^{G304C} complex. Other allotypes such as B*27:05, B*27:09 were observed to form similar complexes using the same mutant combination. Unexpectedly, even A*02:01/β2m^{T5C}/photoKV9 can form complex with ERp57-Tsn^{G304C}, although the peptide exchange of A*02:01 cannot be promoted by ERp57-Tsn^{WT} *in vitro* in our experimental set up.

For the MHC-I ^{Q226C}~Tsn^{H335C} pair mutation, B*27:09^{Q226C}/β2m^{WT}/photoRL9 was UV-exposed and incubated with equimolar ERp57-Tsn^{H335C} overnight. As shown in Fig. 3.4.1c, there was a new band higher than ERp57-Tsn^{H335C}. MS analysis showed that this band contains peptides from B*27:09 heavy chain, suggesting this band was ERp57-Tsn^{H335C}-heavy chain^{Q226C}. But the intensity of this band was very weak, indicating low efficiency of complex formation. Thus, we continued the following experiments with β2m^{T5C}~Tsn^{G304C} mutant combination.

The low expression yield (0.2mg/L) of ERp57-Tsn^{G304C} hindered the attempt to enlarge the complex amount for crystal screening. Since the mutant ERp57-Tsn^{GGLGG}, ERp57-Tsn^{K16L} showed almost the same activity as ERp57-Tsn^{WT} for B*27:09, mutant ERp57-Tsn^{GGKGG} showed the same activity as ERp57-Tsn^{WT} for A*03:01, mutation G304C was introduced into these mutants. The expression yield of ERp57-Tsn$^{GGLGG-G304C}$, ERp57-Tsn$^{K16L-G304C}$, ERp57-Tsn$^{GGKGG-G304C}$ was 6~8 mg/L, 1.5-2.5 mg/L and 2.5-3 mg/L, respectively, which was much higher than that of ERp57-Tsn^{G304C}. It is worth noting that a dimer peak of ERp57-Tsn$^{X-G304C}$ mutant (X stands for GGLGG, K16L or GGKGG) was observed in gel filtration (Fig. 3.4.1e) and the amount of the ERp57-Tsn$^{X-G304C}$ dimer increased with time after purification by gel filtration (Fig. 3.4.1e, lane 4 and 7). But the dimer can be reduced to monomer in the GSSG/GSH buffer while the disulfide bond between ERp57 and Tsn won't be reduced (Fig. 3.4.1f). The ability of the reduced monomer of ERp57-Tsn$^{X-G304C}$ to from a complex with pMHC-I-β2m^{T5C} was the same as the fresh monomer ERp57-Tsn$^{X-G304C}$ (Fig. 3.4.1g, lane 2-7). ERp57-Tsn$^{GGLGG-G304C}$, ERp57-Tsn$^{K16L-G304C}$, ERp57-Tsn$^{GGKGG-G304C}$ can form a complex with all four allotypes refolded with β2m^{T5C},

respectively, but with different efficiency (Appendix Fig.12). And ERp57-Tsn$^{K16L-G304C}$ showed the highest efficiency in forming complex with all four allotypes.

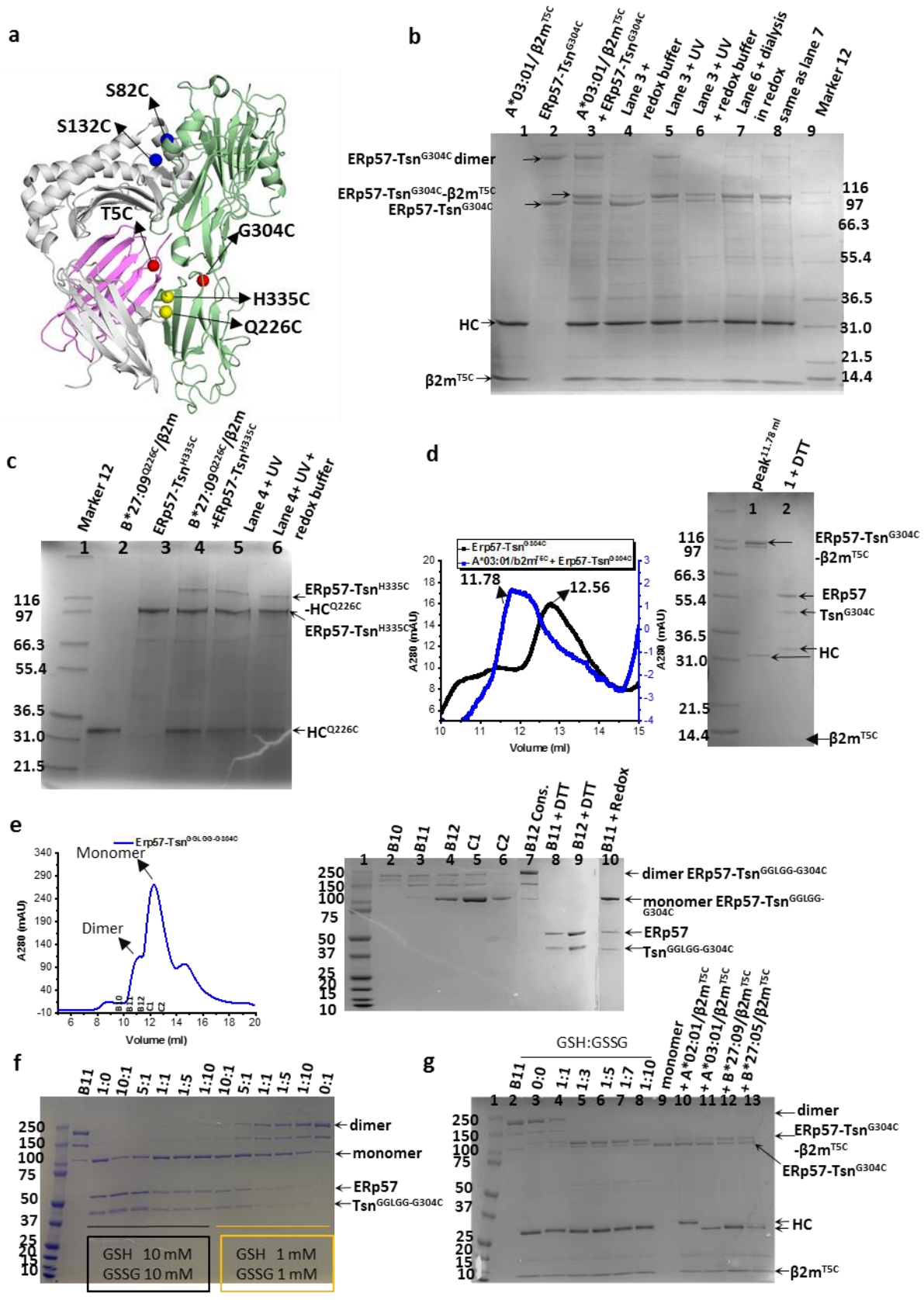


Figure 3.4.1 MHC-I-ERp57-Tsn complex formation by introducing a disulfide bond. a) Three pairs of cysteine mutations are indicated on the structure of MHC-I-ERp57-Tsn complex (modified from PLC, 6NEY) as spheres: $\beta2m^{T5C}$~Tsn^{G304C} in red, MHC-I^{S132C}~Tsn^{S82C} in blue, MHC-I^{Q226C}~Tsn^{H335C} in yellow. **b)** Analysis of A*03:01/$\beta2m^{T5C}$- ERp57-Tsn^{G304C} complex formation on non-reducing 4-12% gradient SDS-PAGE gel in MOPs running buffer. Lane 3: A*03:01/$\beta2m^{T5C}$/photoKK9 was incubated with equimolar of ERp57-Tsn^{G304C} without UV-exposure. **HC** strands for heavy chain. **c)** Analysis of B*27:09^{Q226C}/$\beta2m^{WT}$- ERp57-Tsn^{H335C} complex formation on non-reducing 4 -12% gradient SDS-PAGE gel in MOPs running buffer. β2m band ran out of the gel. Lane 4: B*27:09^{Q226C}/$\beta2m^{WT}$/photoRL9 was incubated with equimolar of ERp57-Tsn^{H335C} without UV-exposure. Lane 6: redox buffer (GSSG: GSH = 5 mM : 0.5 mM, pH = 8.0) was added to help disulfate bond forming. **d)** Chromatography profiles of mixture of 0.09 µmol UV-cleaved A*03:01/$\beta2m^{T5C}$/photoKK9 incubated with equimolar ERp57-Tsn^{G304C} overnight on superdex200 increase 300/10 column. The black curve is the chromatography profile of ERp57-Tsn^{G304C} alone as reference. Lane 1: peak fraction from 11.78ml was concentrated and loaded with non-reducing loading buffer. Lane2: lane 1 + DTT. **e-g)** Dimer of ERp57-$Tsn^{GGLGG\text{-}304C}$ can be reduced to monomer and form complex with pMHC-I-$\beta2m^{T5C}$ in a GSH/GSSG buffer**. e)** Chromatography profile of ERp57-$Tsn^{GGLGG\text{-}G304C}$ purified from 0.3 L insect cell pellets on Superdex 200 increase 10/300 column. Monomer peak eluted at 12.5ml, dimer peak eluted at 11.07ml. Fractions were analyzed on non-reducing SDS-PAGE. Lane 7: fraction B12 was concentrated, stored at 4 °C fridge three days and then ran on the gel. Redox buffer (GSSG: GSH = 5 mM : 0.5 mM, pH= 8.0) was added to dissociate the dimer. **f)** Screening for the optimal GSH:GSSG ratio to reduce the dimer. 5 µl of 1.5 µM ERp57-$Tsn^{GGLGG\text{-}G304C}$ from fraction B11 was incubated with GSH: GSSG buffer mixed in different ratio. Black line indicates the redox buffer was mixed from stock buffer (GSH 10 mM and GSSG 10 mM). Yellow line indicates the redox buffer was mixed from stock buffer (GHS 1 mM and GSSG 1 mM). The legend on the top of the gel is the ratio of GSH: GSSG. The optimal condition is the concentration of GSH is 100-fold of the protein concentration, while GSSG is 10-fold of GSH's concentration. **g)** Reduced monomer can form complex with MHC-I-$\beta2m^{T5C}$. Analysis of complex formation of B*27:09/$\beta2m^{T5C}$ with dimer ERp57-$Tsn^{GGLGG\text{-}304C}$ in different redox condition on non-reducing SDS-PAGE. Lane 3: B*27:09/$\beta2m^{T5C}$ was incubated with dimer ERp57-$Tsn^{GGLGG\text{-}G304C}$ after UV-exposure. Lane 4 - 8: the same as lane 3 but in different GSH: GSSG buffer. Lane 9: monomer ERp57-$Tsn^{GGLGG\text{-}G304C}$ from fresh fraction C2. Lane 10 - 13: monomer ERp57-$Tsn^{GGLGG\text{-}304C}$ was incubated with 4 equimolar A*02:01/$\beta2m^{T5C}$, A*03:01/$\beta2m^{T5C}$, B*27:09/$\beta2m^{T5C}$ and B*27:05/$\beta2m^{T5C}$, respectively, to compare the efficiency of complex formation.

3.4.2 Peptide exchange rate increased dramatically in the presence of disulfide bond

Since the transient interaction between tapasin and MHC-I is stabilized by the disulfide bond, it can be expected that the peptide exchange of MHC-I-ERp57-Tsn complex should be enhanced compared to that in the presence of the unlinked ERp57-Tsn.

As shown in Fig. 3.4.2a, peptide exchange of UV-exposed A*03:01/$\beta2m^{T5C}$/photoKK9 (but not UV-exposed A*03:01/$\beta2m^{WT}$/photoKK9) in the presence of the G304C mutant was increased tremendously. The fold change of *Kon* (compared to uncatalyzed one) is 50-fold by ERp57-$Tsn^{K16L\text{-}304C}$, 23-fold by ERp57-$Tsn^{GGKGG\text{-}304C}$, 22-fold by ERp57-$Tsn^{WT\text{-}304C}$, which is much higher than the corresponding mutant without G304C mutation (Fig. 3.4.2d). A similar trend was also observed for B*27:05/$\beta2m^{T5C}$/photoRL9. On top of that, when incubating the UV-exposed MHC-I/$\beta2m^{T5C}$ with ERp57-$Tsn^{X\text{-}G304C}$ longer to allow the disulfide bond formation and then adding the corresponding high affinity FITC-peptide, the peptide exchange became even faster and the curve reached the plateau within seconds, which was too fast to record the beginning of the association curve.

Surprisingly, even the peptide exchange of A*02:01/$\beta2m^{T5C}$/photoKV9 (but not A*02:01/$\beta2m^{WT}$/photoKV9), which cannot be promoted by the free ERp57-Tsn^{WT}, was increased by 5-fold by ERp57-$Tsn^{K16L\text{-}304C}$, 8.7-fold by ERp57-$Tsn^{GGLGG\text{-}304C}$, 1.7-fold by ERp57-$Tsn^{WT\text{-}304C}$, (Fig. 3.4.2c and e). This result indicates that the observation that the peptide exchange of A*02:01 was not susceptible to free ERp57-Tsn^{WT} *in vitro* in our experimental set-up may be due to the low affinity between tapasin and A*02:01. In the context of the PLC, the peptide exchange of A*02:01 should be promoted by tapasin. However, ERp57-$Tsn^{X\text{-}G304C}$ mutants showed similar activity as the corresponding

ERp57-Tsn^{X} mutant on B*27:09/β2m^{T5C}/photoRL9 (Fig. 3.4.2b), indicating that the high activity of tapasin on B*27:09 is independent of a stronger interaction. It indicates that the interaction between tapasin and B*27:09 is of an affinity where disulfide bond formation does not further contribute to catalytic enhancement.

Why did ERp57-Tsn$^{K16L-G304C}$ increase the peptide exchange of A*03:01 and A*02:01 dramatically compared to the unlinked ERp57-Tsn^{K16L}, while it did not enhance the peptide exchange of B*27:09 compared to the unlinked ERp57-Tsn^{K16L}? Possibly some structural elements are changing during the complex formation of A*03:01 and A*02:01 and these structural elements are critical for binding the incoming peptide. However, for B*27:09/β2m^{T5C}, this critical structural element does not change during the complex formation. The question arises which structural element might be responsible for the observed differences in peptide exchange in a formed complex.

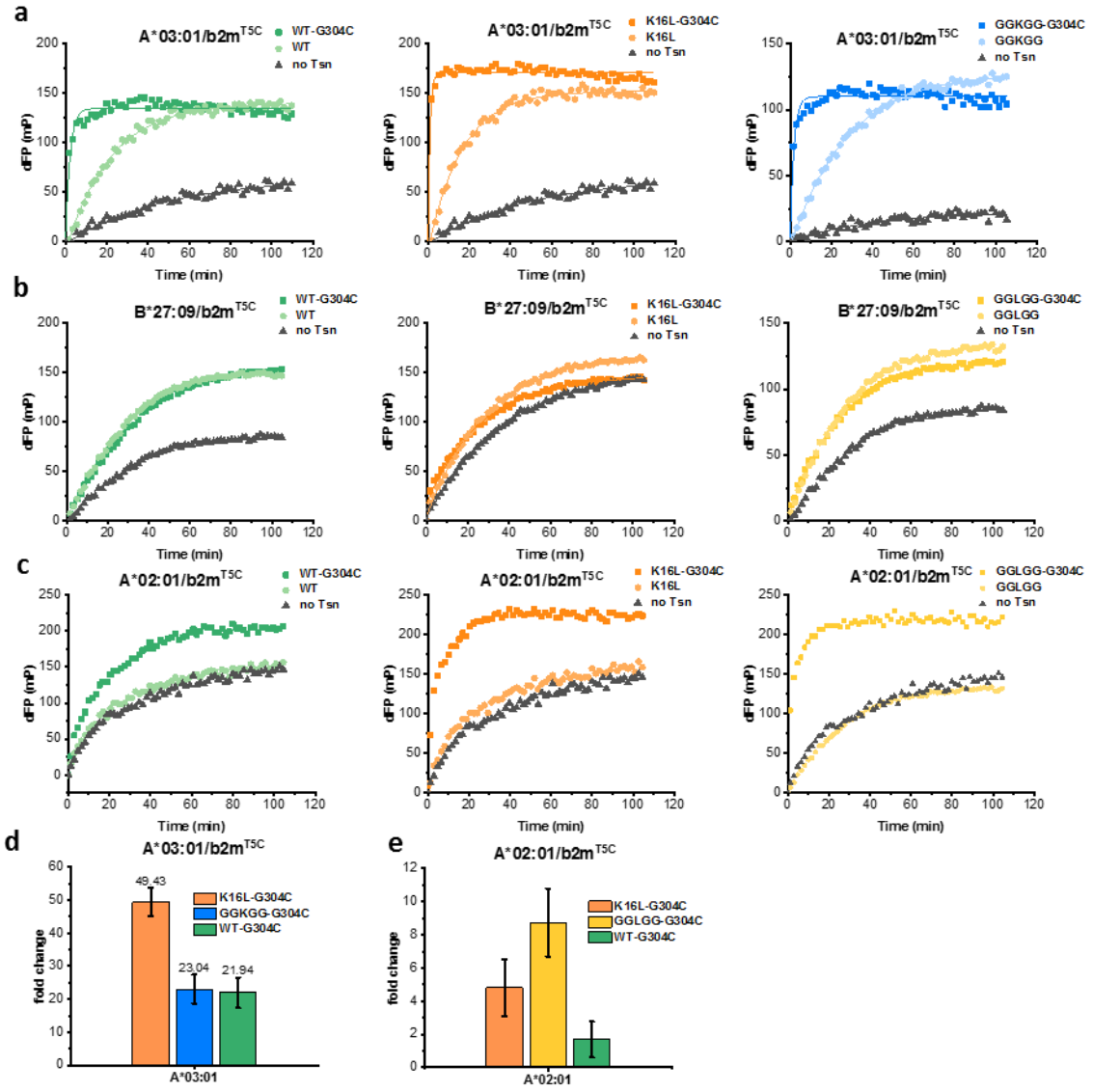


Figure 3.4.2. Peptide exchange of MHC-I-β2m^{T5C} when covalently linked to tapasin via a disulfide bond. a) Peptide exchange profiles of 500 nM UV-exposed A*03:01/β2m^{T5C}/photoKK9 in the absence (black) and in the presence of equimolar amounts of ERp57-Tsn^{WT} (green), ERp57-Tsn$^{WT-G304C}$ (dark green), ERp57-Tsn^{K16L} (orange), ERp57-Tsn$^{K16L-G304C}$ (dark orange), ERp57-Tsn^{GGKGG} (blue), ERp57-Tsn$^{GGKGG-G304C}$ (dark blue). **b)** Peptide exchange profiles of 500 nM UV-exposed B*27:09/β2m^{T5C}/photoRL9 and **c)** 500 nM non-UV-exposed A*02:01/β2m^{WT}/photoKV9 in the absence (black) and in the

presence of equimolar amounts of ERp57-Tsn^{WT} (green), ERp57-Tsn$^{WT-G304C}$(dark green), ERp57-Tsn^{K16L} (orange), ERp57-Tsn$^{K16L-G304C}$ (dark orange), ERp57-Tsn$^{GGLGG-G304C}$ (yellow), ERp57-Tsn$^{GGlGG-G304C}$ (dark yellow). The three black curves looked different because the experiments were performed with different batch of B*27:09/β2m^{T5C}/photoRL9, but the trend is the same. **a-c:** Peptide exchange was monitored immediately after adding ERp57-Tsn mutant into UV-exposed MHC-I/β2m^{T5C} or non-UV-exposed MHC-I/β2m^{T5C} (A*02:01) and the corresponding FITC-peptide without incubation. dFP means the baseline was subtracted. **d)** and **e)** Fold change of *Kon* in the presence of ERp57-Tsn$^{X-G304C}$ mutants compared to the uncatalyzed *Kon* of corresponding allotypes.

3.4.3 The dual functions of tapasin when ERp57-Tsn$^{X-G304C}$ forms complex with MHC-I

In order to answer the question raised in section 3.4.2, we checked if the cleaved peptide is still bound in the peptide binding groove when forming a complex with ERp57-Tsn mutant by MS.

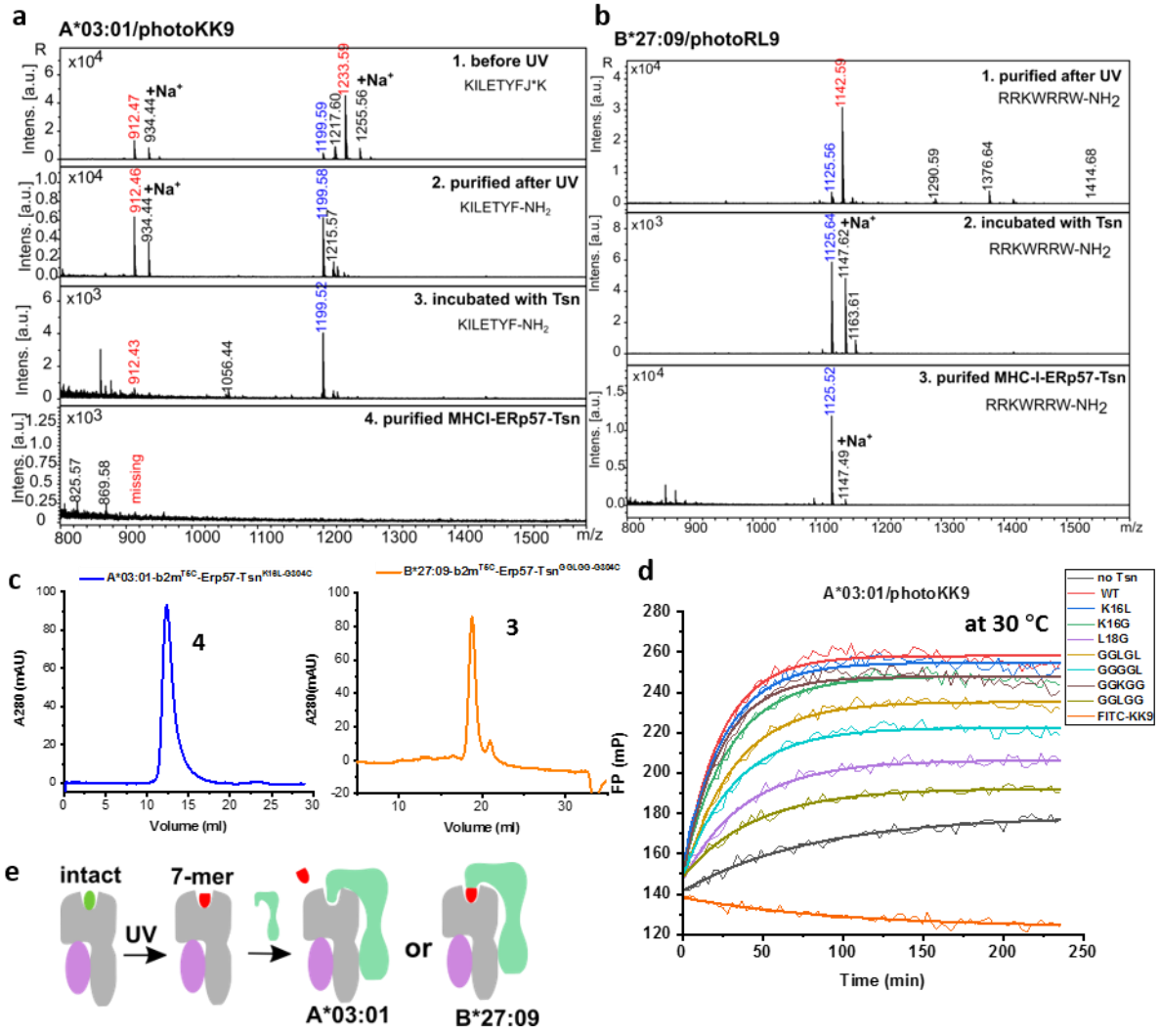


Figure 3.4.3 MS analysis of cleaved 7-mer peptide after MHC-I forming complex with tapasin. a) MALDI-TOF analysis of A*03:01/β2m^{T5C}/photoKK9 before UV-exposure (sample 1), purified by gel filtration after UV-exposure (sample 2), purified after incubated with ERp57-Tsn (sample 3), in the complex A*03:01/β2m^{T5C}-ERp57-Tsn$^{K16L-G304C}$ (sample 4). The signal of 7-mer fragment KILETYF-NH$_2$ was detected (observed mass 912.47 Da, expected mass 912.08 Da). Peak 1199.58 which is 34 Da less than the intact peptide was the intramolecular cyclization product of the reactive intermediate. The proposed reaction mechanism is explained in Appendix Fig. 13. **b)** MALDI-TOF analysis of B*27:09/β2m^{T5C}/photoRL9 purified by gel filtration after UV-exposure (sample 1), purified after incubated with ERp57-Tsn (sample 2) and in the complex B*27:09/β2m^{T5C}-ERp57-Tsn$^{GGLGG-G304C}$ (sample 3). The signal of 7-mer fragment RRKWRRW-NH$_2$ was detected (observed mass 1142.59 Da, expected mass 1143.37 Da). Peak 1125.52 which is 16 Da less than the 7-mer peptide was observed when the sample was kept for long times after UV-exposure. **c)** Chromatograph of A*03:01/β2m^{T5C}-ERp57-Tsn$^{K16L-G304C}$ complex in Superdex 200 increase column (sample 4) and Chromatograph of B*27:09/β2m^{T5C}-ERp57-Tsn$^{GGLGG-G304C}$ complex in Mono-

Q column (sample 3). **d)** Peptide exchange profiles of A*03:01/photoKK9 in the presence of ERp57-Tsn mutants at 30 °C. The plateau is different in the presence of ERp57-Tsn mutants. **e)** Cartoons showing tapasin promoting suboptimal peptide dissociation and stabilize the empty binding groove of MHC-I when forming complex. Gray is heavy chain, purple is β2m, green is tapasin.

For A*03:01/β2m^{T5C}/photoKK9, the cleaved 7-mer peptide (KILETYF-NH_2) was still bound to the binding groove of the purified UV-exposed A*03:01/β2m^{T5C}/photoKK9 (Fig. 3.4.3a), in line with the results in section 3.1.3. The signal for the 7-mer peptide was still detected after incubation with unlinked ERp57-Tsn^{K16L}. However, there was no peptide signal anymore when UV-exposed A*03:01/β2m^{T5C} formed the disulfide-linked complex with ERp57-Tsn$^{K16L-G304C}$ (Fig. 3.4.3a, sample 4), which explains why the peptide exchange of UV-exposed A*03:01/β2m^{T5C}/photoKK9 can be largely enhanced (~50-fold) by ERp57-Tsn$^{K16L-G304C}$, but only ~3-fold by ERp57-Tsn^{K16L}. Strikingly, for B*27:09/β2m^{T5C}/photoRL9, the 7-mer peptide (RRKWRRW-NH_2) was still in the binding groove of the UV-exposed B*27:09/β2m^{T5C}/photoRL9 no matter if it was incubated with ERp57-Tsn^{X} or formed tight disulfide-linked complex with ERp57-Tsn$^{X-G304C}$ (Fig. 3.4.3b). This explains why the ERp57-Tsn$^{X-G304C}$ mutant showed the same activity as the corresponding ERp57-Tsn^{X} mutant for B*27:09/β2m^{T5C}/photoRL9.

It has been suggested that one function of tapasin is to promote the suboptimal peptide leaving the binding groove of MHC-I in the PLC and then stabilizes the empty MHC-I. It is worthy of note that the plateau of FP signals varied in the presence of tapasin mutants when the peptide exchange assays of A*03:01 were performed at higher temperature (Fig. 3.4.3d), while for B*27:09, the plateaus remained unchanged, indicating tapasin could stabilize A*03:01 complex when the binding groove is half-empty. Moreover, the T_m of the A*03:01/β2m^{T5C}/ERp57-Tsn$^{X-G304C}$ complex (the binding groove was completely empty) was nearly the same as UV-exposed A*03:01/β2m^{T5C} MHC-I (where the 7-mer fragment was still bound) (Appendix Table 3), indicating that tapasin could stabilize the binding groove of the cleaved pMHC-I in the disulfide-linked complex. To conclude, we provided evidence that tapasin can promote suboptimal peptide dissociation and functions as a chaperone to stabilize the empty MHC-I (Fig. 3.4.3e).

In case the above hypothesis holds true, why then would the cleaved peptide remain bound in the disulfide-linked B*27:09/β2m^{T5C}/ERp57-Tsn$^{X-G304C}$ complex? The H-bond formed between *p*R5 and D77 as indicated by the crystal structure of B*27:09/photoRL9 (Fig. 3.1.1e) might keep the peptide in the binding groove. It is unclear whether this is related to the property of the peptide photoRL9 or is allotype-specific or both. Further experiments are needed to perform to answer this question.

3.4.4 Cryo-EM structure of MHC-I-ERp57-Tsn

After obtaining the MHC-I-ERp57-Tsn complex, trials to grow MHC-I-ERp57-Tsn complex crystals failed. Considering the size of the complex about 150 kDa, there is a chance to solve the structure of the complex by cryo-EM. Samples (B*27:09/β2m^{T5C}-ERp57-Tsn$^{GGLGG-G304C}$ or A*03:09/β2m^{T5C}-ERp57-Tsn$^{K16L-G304C}$) for cryo-EM trials were purified by gel filtration as the final purification step (Fig. 3.4.4a). Negative stains showed nice single particle dispersion (Fig. 3.4.4b). When the B*27:09 allotype was used for complex assembly, non-random orientation of particles (without crosslinking and no detergent) was observed in the single data acquisition of image on Titan TEM (Appendix Fig.14a). This hindered

obtaining higher resolution. Adding detergent (0.15% (w/v) *n*-octylglucoside) into B*27:09/β2m^{T5C}-ERp57-Tsn$^{GGLGG-G304C}$ did not overcome the problem but induced precipitation of the sample on the grid. When A*03:01 was used to assemble the complex, the orientation dispersion of particles improved (without crosslinking, no detergent) (Appendix Fig. 14b) and a resolution of ~7.5 Å was obtained (Appendix Fig. 14c). Most regions of the complex align nicely to the corresponding proteins in the already published cryo-EM structure of the PLC[178], with the only exception of the IgG domain of tapasin which moves away from the α3 domain of the heavy chain in our structure (Fig. 3.3.4b). Well-defined density was observed for the heavy chain, β2m and tapasin, while ERp57 showed low density or even was not visible in the complex due to its high flexibility, which hindered the acquisition of higher resolution density maps. Correspondingly, a recent all-atom molecular dynamics (MD) simulation based on PLC density found out that in the PLC both calreticulin and ERp57 showed higher root-mean-square deviation (RMDS) than the core part MHC-I-Tsn[214]. Low resolution impedes detailed modeling of the complex, thus no further insight about the loops involved in tapasin binding could be obtained.

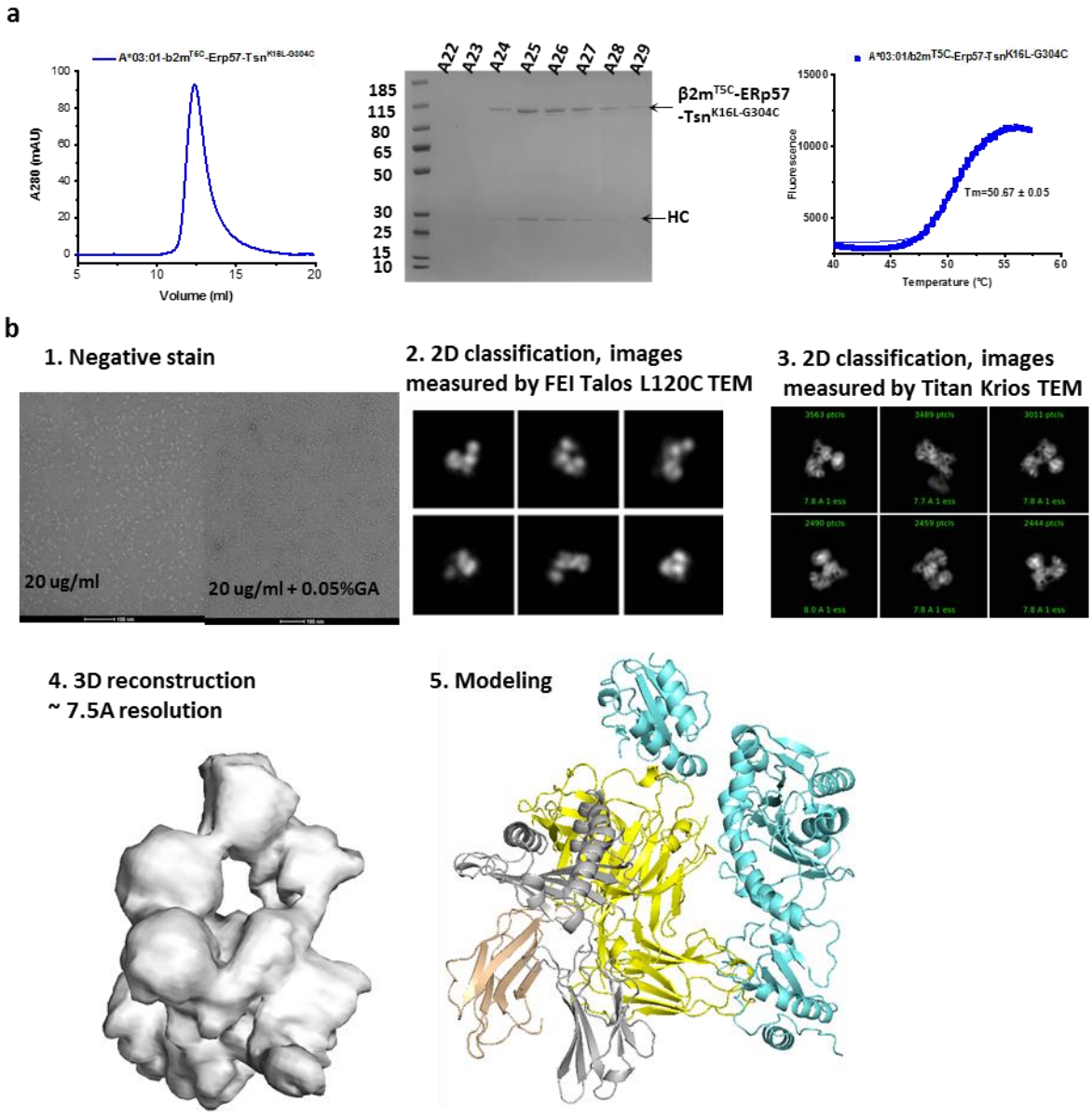


Figure 3.4.4. Sample preparation of MHC-I-ERp57-Tsn complex for cryo-EM. a) Characterization of the MHC-I-ERp57-Tsn complex**. Left:** Chromatography profile of A*03:01/β2m^{T5C}-ERp57-Tsn$^{K16L-G304C}$ complex in superdex200 increase column. Mixture of A*03:01/β2m^{T5C} with ERp57-Tsn$^{K16L-G304C}$ was first purified by Mono-Q column and then the

corresponding fraction was further purified by superdex200 increase column in HEPEs running buffer. **Middle:** fractions of monomer peak eluted at 12.05 ml were run on non-reducing SDS-PAGE to check purity. ERp57-Tsn$^{K16L-G304C}$ cannot be completely separated from the complex due to the low difference in molecular weight. **Right:** Melting curve of A*03:01/β2m^{T5C}-ERp57-Tsn$^{K16L-G304C}$ complex after gel filtration, T_m is 50.67 ± 0.05 °C. **b)** The workflow of cryo-EM sample preparation and data acquisition. **1:** The representative micrographs of negatively stained complex without any treatment (left) and fixed by 0.05% GA. Adding cross-linking procedure do not affect the organization of the complex. The scale bar is 100 nm in the micrograph. **2-3:** 2D Class averages measured by Talos TEM (2) and by Titan TEM (3). Class average reproducibly resemble similar. The scale bar is 25 nm in both. **4.** The full dataset was directly submitted to multimodel classification in 3D. **5**. The published PLC structure fits into the density of the complex here.

3.4.5 H408R of ERp57 or P96L of tapasin increase the PLC stability

In order to improve the complex stability and potentially reduce its flexibility for acquiring better Cryo-EM pictures, the introduction of other mutations was considered based on the literature. Padariya *et al.* analyzed how cancer-associated mutations in ERp57 or Tsn alter their binding to each other by MD simulation and found that the H408R mutation of ERp57 and P96L mutation in Tsn increased its structural stability by -5 kcal/mol and -3 kcal/mol, respectively[215]. Furthermore, the Tsn^{P96L} mutant forms a larger amount of H-bonds in the ERp57-Tsn complex and the C-terminal domain of it moves closer to the PLC in the MD simulation. In contrast, the ERp57^{H408R}-Tsn mutant, the C-terminal domain of Tsn moves away from the PLC. In the attempt to stabilize the complex, these two mutations were introduced in our system, resulting in complexes formed of ERp57-Tsn$^{G304C-P96L}$, ERp57-Tsn$^{K16L-G304C-P96L}$, and ERp57^{H408R}-Tsn$^{G304C-P96L}$, ERp57^{H408R}-Tsn$^{K16L-G304C-P96L}$ that showed increased T_m (work in progress).

3.4.6 The concept of exchange peptide on the cell surface with tapasin cysteine mutants

As our mutants lead to tight complexes and higher peptide exchange rate, they might be useful for the technology of preparing pMHC-I libraries. To probe diverse, polyclonal TCR repertoires, the routine method is to stain with pMHC multimers followed by flow cytometry to interrogate T cell responses, to characterize antigen-specific TCR repertoires and to identify immunodominant clones. The technologies to prepare pMHC multimers includes using conditional ligands, increased temperature, or using dipeptide to catalyze peptide exchange[181,216–218]. However, pMHC-I multimer libraries have limitations, as they cannot reflect the real situation when APCs are recognized by T cells with costimulatory signals missing. The solution to circumvent the need of preparing pMHC-I multimer is to prepare cell libraries by exchanging peptides on the cell surface. It has already been shown that exogenous TAPBPR can exchange peptides on the surface[219].

As shown above, the peptide exchange of MHC-I/β2m^{T5C} in the presence of ERp57-Tsn$^{G304C-X}$ was increased tremendously, even for the tapasin-independent allotypes (Fig. 3.4.2c), we followed the idea of peptide exchange on the cell surface by establishing an *in vitro* approach as a first step towards this goal. Previous researches have shown that β2m dissociates from pMHC-I at a rate depending on the peptide occupancy of the binding groove and exogenous β2m can stabilize the pMHC-I complex[220,221]. We performed experiments *in vitro* with A*02:01/β2m^{WT}/photoKV9. As shown in Fig. 3.4.6b, peptide exchange was increased by 2.6-fold and 2.4-fold in the presence of equimolar β2m^{WT} or β2m^{T5C}, respectively. This is in agreement with previous studies that the peptide loading of MHC-I was greatly facilitated in the presence of high concentration of exogenous β2m[222–224]. It has been shown that

exogenous β2m promotes peptide loading of the pMHC-I molecules by re-associating with pMHC-I to stabilize the peptide-receptive state[222–224]. In principle, we chould see $\beta2m^{WT}$ in the pMHC-I getting replaced by free $\beta2m^{T5C}$ when incubating MHC-I/$\beta2m^{WT}$ with $\beta2m^{T5C}$ and the re-associated $\beta2m^{T5C}$ in MHC-I could form a disulfide bond when later incubating with ERp57-$Tsn^{X\text{-}304C}$. As shown in Fig. 3.4.6b, peptide exchange was increased by 4.1-fold in the presence of ERp57-$Tsn^{G304C\text{-}K16L}$ together with $\beta2m^{T5C}$ (light blue curve), but not increased (if any, slightly) by ERp57-$Tsn^{G304C\text{-}K16L}$ alone (green curve), while still 2.5-fold in the presence of ERp57-$Tsn^{G304C\text{-}K16L}$ together with $\beta2m^{WT}$, indicating the disulfide bond was formed between the re-associated $\beta2m^{T5C}$ but not the re-associated $\beta2m^{WT}$ in A*02:01 with ERp57-$Tsn^{G304C\text{-}K16L}$. The fold change of *Kon* was even lager when A*02:01/$\beta2m^{WT}$/photoKV9 was incubated with $\beta2m^{T5C}$ and ERp57-$Tsn^{G304C\text{-}K16L}$ for longer time (pink curve, 6.5-fold) to allow the re-association of $\beta2m^{T5C}$ and disulfide bond forming before adding FITC-peptide or by increasing the amount of $\beta2m^{T5C}$ ($\beta2m^{T5C}$ is 10 times of A*02:01/$\beta2m^{WT}$/photoKV9, 7.9-fold) or by both (20.9-fold).

The results confirmed the concept that exogenous $\beta2m^{T5C}$ can replace the pre-bound $\beta2m^{WT}$ in the pMHC-I and that the re-associated $\beta2m^{T5C}$ can form a disulfide bond with the ERp57-$Tsn^{X\text{-}G304C}$ mutant.

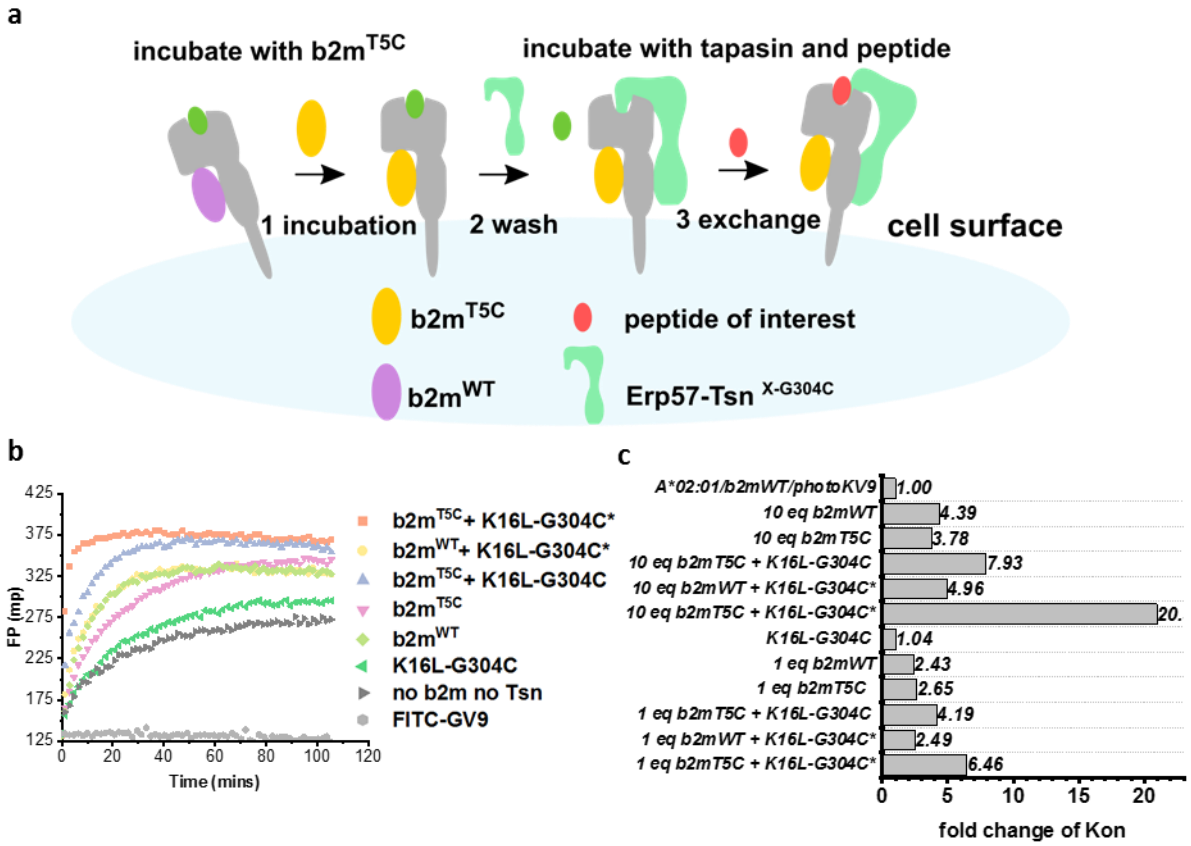


Figure 3.4.6 Exogenous $\beta2m^{T5C}$ and ERp57-$Tsn^{G304C\text{-}K16L}$ promote peptide exchange of A*02:01/$\beta2m^{WT}$. a) Concept of peptide exchange on the cell surface. Step1: incubate cells with $\beta2m^{T5C}$ to replace the refolded $\beta2m^{WT}$. Step2: wash away the extra $\beta2m^{T5C}$. Step3: incubate cells with ERp57-$Tsn^{G304C\text{-}K16L}$ and peptides of interest. **b)** Peptide exchange profile of A*02:01/$\beta2m^{WT}$ in the presence of different combination of incubating β2m and tapasin mutant. Peptide exchange experiments was done by incubating 500 nM A*02:01/$\beta2m^{WT}$ with equimolar $\beta2m^{T5C}$ or $\beta2m^{WT}$ for three hours to let the β2m to exchange. Then ERp57-$Tsn^{G304C\text{-}K16L}$ was added right before adding FITC-peptide to monitor the peptide exchange process. * indicates A*02:01/$\beta2m^{WT}$ was incubated with $\beta2m^{T5C}$ (or $\beta2m^{WT}$) and ERp57-$Tsn^{G304C\text{-}K16L}$ together for three hours. Experimental profiles of 10 equimolar of $\beta2m^{T5C}$ or $\beta2m^{WT}$ were not shown. **c)** Fold change of *Kon* in different conditions is shown as bar chart. * indicates the same as described in **b**.

4 Discussion

The repertoire of peptides presented by MHC-I molecules on the cell surface is tailored by the ER-resident PLC that contains the exchange catalyst tapasin. Tapasin promotes the replacement of suboptimal antigens by high-affinity peptides and thereby fosters the formation of thermodynamically stable complexes fit to serve as T cell antigens once transported to the cell surface. It has been shown that the conformational exchange dynamics of the MHC-I molecules over the whole binding groove and the F-pocket are associated with its stability/peptide exchange. In this work, biochemical and biophysical methods were used to investigate the dynamics of four natural MHC-I allotypes and conformational exchanges of one MHC-I allotype (B*27:09) in the presence of tapasin and its mutants. results revealed two essential features of MHC-I are exploited by tapasin: 1, A conserved allosteric site underneath the $\alpha_{2\text{-}1}$-helix of MHC-I is critical for tapasin binding. After binding, tapasin "loosens" the F-pocket region that binds to the C-terminus of the peptide. 2, The F-pocket of MHC-I is sampled by tapasin scoop $loop_{11\text{-}20}$, enabling competitive peptide binding.

Additionally, the differences in conformational changes and allele-specific peptide exchange activity between tapasin and TAPBPPR were compared. Furthermore, by comparing the catalytic activity of tapasin scoop $loop_{11\text{-}20}$ mutants in peptide exchange of MHC-I, L18 of the tapasin scoop $loop_{11\text{-}20}$ was identified as the critical residue mediating peptide exchange, while K16 plays an accessory role for MHC-I allotypes bearing an acidic F-pocket. The emergence of K16 in the $loop_{11\text{-}20}$ coincides with the occurrence of MHC-I alleles bearing an acidic F-pocket in human MHC-I allotypes. Finally, by introducing cysteine mutations, a cryo-EM structure of the MHC-I/ERp57-Tsn complex was obtained, which lays the foundation for further understanding the molecular mechanism of tapasin-mediated peptide exchange.

4.1 The features of MHC-I molecules and tapasin dependency

It has been shown that pMHC-I molecules display conformational exchange dynamics[160]. This type of dynamics is associated with its thermostability, affecting its peptide exchange and interactions with the peptide editors. There have been many cellular studies on the tapasin dependence of MHC-I[140,152,157–159,163]. It has been shown that tapasin displays allele-specific influences on the peptide repertoire of MHC-I at the cellular level. Previous studies investigated how tapasin affects the kinetics of *in vitro* peptide exchange of one MHC-I allotype[146]. No such data has been obtained to data for the allele-specific influences of tapasin on peptide exchange *in vitro*. In this thesis, the tapasin-catalyzed enhancement of the peptide association rate K_{on} of four allotypes bearing F-pockets of distinct nature were compared (section 3.2.2). In this section, the discussion will focus on the features (such as dynamics, thermostability and peptide affinity) of MHC-I molecules and how these features correlate with their tapasin dependence.

4.1.1 The dynamics of MHC-I molecule and its thermostability

Conformational exchange dynamics exerted by MHC-I molecules over the whole binding groove and F-pocket are associated with their thermostability. Yanaka *et al.* have studied the conformational

fluctuations of three HLA-B*35:01 in complex with three different peptides by ^{15}N-^{1}H CPMG relaxation dispersion and suggested the minor state of B*35:01 complex plays a role in the peptide-dependent stabilization of the B*35:01 complex[203]. In their hypothesis, the minor state of MHC-I is more dehydrated and packed and thus more thermostable, whereas the major state of MHC-I seems to accommodate a peptide more loosely compared to the packed minor conformation, as it allows the loading of structurally different peptides in a similar conformation. As this was shown only for B*35:01, it was unclear, if a similar correlation between the minor population and thermostability will generally exist for other MHC-I allotypes. In this study, the conformational fluctuations of another HLA allotype, B*27:09, in complex with two different peptides which show different thermostability, were analyzed (section 3.1.2). These two B*27:09 complexes showed different conformational fluctuations in the binding groove. As CPMG relaxation dispersion measurements were used, it would in principle be possible to calculate the population of different conformational states. However, CPMG experiments were only measured at a single magnetic field strength to get qualitative information about dynamics in this study, which prevents a reliable calculation of populations.

According to Yanaka's hypothesis, since the thermostability of B*27:09/photoRL9 is lower than that of B*27:09/RL9, the minor state population of B*27:09/photoRL9 is predicted to be lower than that of B*27:09/RL9. In other words, the major state population (peptide loosely bound state) of B*27:09/photoRL9 is higher than that of B*27:09/RL9. The major state should be more susceptible to peptide exchange because the peptide binds loosely in the binding groove. Indeed, intact B*27:09/photoRL9 can exchange with FITC-RL9 (Fig. 3.1.3c, the gray curve), although slowly, while B*27:09/RL9 is not able to exchange with FITC-RL9 at all. Additionally, the thermostability of the allotype B*27:05 in complex with the peptide photoRL9 is higher than that of B*27:09/photoRL9 (Appendix Table 3), and less dynamics were observed for B*27:05/photoRL9 compared to B*27:09/photoRL9 (Fig. 3.2.5), which means the major state population of B*27:05/photoRL9 is higher than that of B*27:09/photoRL9. Consequently, lower peptide exchange rate of B*27:05/photoRL9 was observed compared to B*27:09/photoRL9 against FITC-RL9 (Appendix Fig. 10a). These above observations suggest that the thermostability of pMHC-I not only reflects the dynamics of the pMHC-I complex, but also correlates with the intrinsic peptide exchange rate of pMHC-I molecules.

4.1.2 The different dynamic profiles of MHC-I allotypes

Natural polymorphic variants of MHC-I (A*02:01, A*03:01, B*27:05 and B*27:09) showed different dynamics profiles (section 3.2.5) in this study. Most residues showing conformation fluctuations are located mainly on the pleated β-strands of the groove, $\alpha_{2\text{-}2}$-helix and the $loop_{126\text{-}133}$ of the heavy chain (Fig 3.2.5b). Interestingly, residues of the $loop_{126\text{-}133}$ and β-strands 7-8 showing allele-specific dynamic patterns were affected in the presence of tapasin. Thus, one can speculate that the dynamics of the $loop_{125\text{-}133}$ and β-strands 7-8 are important for tapasin binding (Fig. 3.2.5a).

According to the previous finding that MHC-I regions of increased flexibility correlate with residues required for TAPBPR association[169], less residues in B*27:05/photoRL9 are expected to show dynamics since the dependence of B*27:05/photoRL9 on TAPBPR is lowest among these four allotypes. Indeed, fewer residues (only one) of B*27:05/photoRL9 in the binding interface with TAPBPR showed

dynamics (Fig. 3.2.5). In addition, B*27:05 and B*27:09 show an inverse dependence on tapasin and TAPBPR (section 3.2.2), which implies tapasin recognizes MHC-I in a state of reduced flexibility.

4.1.3 Peptide exchange kinetics of pMHC-I in the presence of tapasin

Peptide exchange of stable pMHC-I molecules was not or only slightly enhanced by tapasin (Appendix Fig. 10a), whereas significant effects of tapasin on peptide exchange were seen after UV-exposure of photo-pMHC-I molecules. This is in accordance with a previous result using mouse MHC-I H2-K^b demonstrated that tapasin mediates the binding of the high affinity peptide to MHC-I by accelerating the dissociation of the peptide from an unstable intermediate[225]. The peptide exchange rates of B*27:09 and B*27:05 with different photocleavable peptides with affinities ranging from high to medium in the presence of Erp57-Tsn^{WT} were compared (section 3.2.1). The results demonstrated that the enhancing effect of tapasin is more pronounced when the affinity of the remaining 7-mer peptide is relatively high (Fig. 3.2.1g). In this case the peptide-free F-pocket presumably allows the scoop loop of tapasin to stabilize the binding groove further as suggested by the peptide model of the scoop loop[180]. In case the 7-mer peptide affinity is low, dissociation will be fast (the requirement of tapasin for catalyzing peptide dissociation decreases) and the loading of the incoming peptide bearing high affinity (uncatalyzed *Kon*) is fast (Fig. 3.2.1f). The results are in line with statements obtained with a leucine zippered complex of HLA-B*08:01 and tapasin by Chen *et al.*[146] and with a study that claims tapasin catalyzes thermodynamically favored but kinetically disfavored exchange of suboptimal peptide for the HLA-B*44:02-ERp57-Tsn complex linked by dimeric NeutrAvidin[150]. This is also relevant to a previous study conducted by Howarth *et al.* that demonstrated enhancing effect of tapasin on peptide loading to H2-K^b molecules presented on the cell surface correlates with the peptide half-life (dissociation rate)[162].

As for the allelic differences in tapasin-catalyzed peptide exchange, the influence of tapasin on peptide association kinetics seems positively correlated with the acidity of the F-pocket of the MHC-I molecule. This is particularly indicated by the tapasin-catalyzed peptide exchange of B*27:05 and B*27:09, considering the two allotypes differ only by one residue in the F-pocket. The data is in consistent with the previous result that B*27:05 is more tapasin-dependent than B*27:09[30]. Interestingly, previously data showed B*44:02 which bears a more polar F-pocket is tapasin-dependent, while B*44:05 bears a hydrophobic F-pocket that is tapasin-independent[153].

Rizvi *et al.* studied the intrinsic assembly and refolding efficiencies of HLA-B allotypes and found an inverse correlation between low cell surface expression of MHC-I allotypes under tapasin-deficient conditions and the degree of tapasin-mediated enhancement on assembly [157]. They showed within the same supertype the intrinsic assembly of tapasin-independent MHC-I alleles (e.g., within B*44, B*18:01 and B*44:05) was more competent than the tapasin-dependent MHC-I alleles (B*44:02 and B*44:03) in the tapasin-deficient cell line (M553)[157]. Highly tapasin-independent allotype (e.g., within B*7, B*35:01 and B*35:03) bound to peptides more efficiently than the tapasin-dependent allotype (B*51:01) in the absence of tapasin[157]. Interestingly, a similar trend was also observed in this study that the uncatalyzed peptide association rate of the four tested allotypes after UV-exposure is inversely correlated with the degree of tapasin enhancement on peptide exchange (Fig. 4.1a). Although not absolute, there is a trend visible: the uncatalyzed *Kon* of A*02:01 was the highest while its peptide

exchange enhanced by tapasin was the lowest. Within the B*27 supertype, the uncatalyzed *Kon* of B*27:05 was lower than that of B*27:09 while its peptide exchange was more catalyzed by tapasin. This might correlate with the thermostability of the complex after UV-exposure (Fig. 4.1b). Correspondently, a previous report found that tapasin enhancing peptide loading generally parallels pMHC-I complex stability[162].

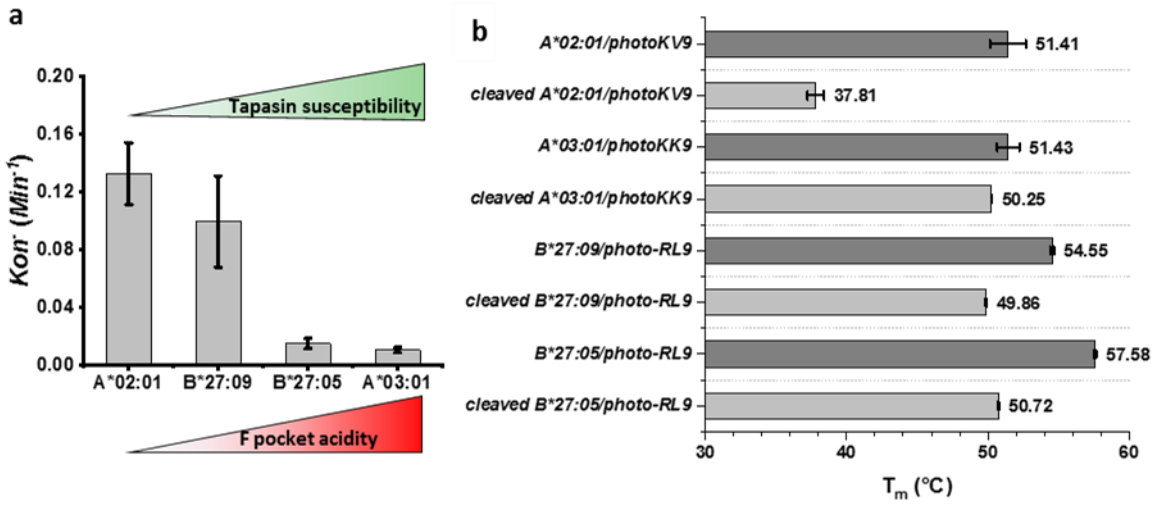


Figure 4.1 The F-pocket nature of MHC-I and its tapasin susceptibility. a) The Y axis is the uncatalyzed *Kon* of the tested allotypes. The green bar represents the degree of tapasin enhancement on peptide exchange and the red bar represents the gradient of the F-pocket acidity as the allotypes showed. **b)** T_m of the photo-MHC-I before and after UV-exposure.

There have been many cellular studies on the tapasin dependence of MHC-I[30,152,155–158,161,163]. In those cellular studies, tapasin dependence of an allotype is evaluated by the cell surface expression yield and stability of the presented MHC-I molecules in tapasin-deficient cells. In contrast, the concept of tapasin susceptibility/tapasin enhancing peptide exchange of MHC-I allotypes in this thesis is based on the change factor of peptide exchange rates, which is a kinetic parameter and not a composite equilibrium parameter such as the cell surface expression. This could be the reason why we might observe some inconsistency between our data and previous cellular data. The refolding of A*02:01 could be facilitated by the N-terminal domain of tapasin[205], while in our experimental set up peptide exchange of A*02:01/photoKV9 was not enhanced by free ERp57-Tsn^{WT}. When A*02:01/photoKV9 was linked with ERp57-Tsn by an artificial disulfide bond (three scoop loop mutants were tested in the linked form), the peptide exchange of A*02:01/photoKV9 was enhanced (Fig. 3.4.2c). This suggests that tapasin can catalyzes the peptide exchange of A*02:01 in the PLC context where the interaction between tapasin and A*02:01 is stabilized with the help of other PLC components. However, it does not necessarily mean that A*02:01 is tapasin-dependent. Similarly, although enhancement of tapasin on peptide exchange of A*03:01 was highest among the allotypes tested in this thesis, it does not necessarily mean that A*03:01 is a high tapasin-dependent allotype.

4.2 The interaction between MHC-I molecule and tapasin in detail

The cryo-EM structure of the PLC provides insight into how tapasin arranges in the PLC and interacts with MHC-I. However, due to the low resolution of this whole structure and some critical regions being

not resolved, the exact mode of interaction between tapasin and MHC-I is not clear. This section will discuss the interactions between MHC-I and tapasin. In this thesis, two regions (the loop$_{126-133}$ region and the F-pocket region) of MHC-I were identified that are affected by the presence of tapasin using NMR measurements (section 3.2.3). Combined with previous results from the literature[145], tapasin loop$_{187-196}$ (connecting β-strands 9-10) is proposed to recognize the dynamics of MHC-I loop$_{126-133}$ region for tapasin binding which is independent of the nature of the F-pocket and facilitates the release the C-terminus of the peptide. In this regard, tapasin's scoop loop$_{11-20}$ may exert its function during peptide exchange, interacting with and potentially altering the F-pocket dynamics.

4.2.1 The MHC-I loop$_{126-133}$ is important for tapasin binding

Evidence both from previous studies and from this work indicate that the loop connecting MHC-I β-strands 7-8 (loop$_{126-133}$) is important for tapasin binding[148,226]. Yu *et al*. have shown that introducing mutations within the loop$_{126-138}$ (they named it as loop$_{128-136}$) of H2-L^d affected its association to the peptide loading complex components[148]. They have speculated that the loop$_{128-136}$ reflects the state of the peptide binding groove and its conformation is probably monitored by tapasin, which only recognizes it when a "open" conformation (similar to the loosely bound state mentioned in section 4.1.1 or ground state) exists[148]. In this work, chemical shift changes of residues in the loop$_{126-133}$ were observed when the cleaved B*27:09/photoRL9 shifted to peptide-non-receptive state (Fig. 3.1.5). Besides, when comparing the crystal structure of B*27:09/RL and B*27:09/photoRL9, the side chain orientation of residue E128 is different. Together (NMR and crystal structure), these observations suggest that the conformation of the loop$_{126-133}$ might report on the receptiveness of the conformational state of the pMHC-I molecule. Moreover, NMR peak intensities of residues close or within the loop$_{126-133}$ region were strongly reduced in the presence of tapasin (Fig. 3.2.3), demonstrating that tapasin binding induces conformational changes to the loop$_{126-133}$. Furthermore, allele-specific dynamics of the loop$_{126-133}$ region were observed (Fig. 3.2.5) for allotypes showing different tapasin dependence. Interestingly, residues from this loop$_{126-133}$ region are highly conserved in human MHC-I allotypes (Fig. 4.2b), with only residue 127 and 131 showing few variabilities. The residue 127 of A*02:01 is K and the peptide exchange of A*02:01 was not significantly enhanced by tapasin (Fig. 3.2.2), while the other three allotypes tested in this work have N at this position. In conclusion, the dynamics of the loop$_{126-133}$ is important for tapasin recognition and could thereby provide a landing platform for tapasin binding. In addition, the loop$_{126-133}$ region of B*27:05/photoRL9 showed less dynamics and higher tapasin dependence compared to that of B*27:09/photoRL9. Thereby, it implies a more rigid loop$_{126-133}$ region is needed for tapasin to bind.

In the cryo-EM structure of the PLC, the loop$_{126-133}$ of MHC-I is close to the tapasin loop connecting β-strands 9-10 (loop$_{187-196}$, "hairpin" loop). Residue T134 of the heavy chain was predicted to point to the critical residue R187 of tapasin (Fig. 4.2a). However, in this work, residue T134 was not affected in the presence of tapasin in NMR experiments (Fig. 3.3.3b), suggesting no direct interaction of T134 with tapasin. One could speculate that tapasin loop$_{187-196}$ is responsible for recognizing the dynamics of MHC-I loop$_{126-133}$ rather than T134 for binding. Furthermore, this recognition mode might also be used by other species, since the loop$_{126-133}$ is also highly conserved in MHC-I allotypes of other species (Fig. 4.2b) as well as the hairpin loop$_{187-196}$ of tapasin is highly conserved among species (Fig. 4.2c).

Another interesting point is that residue 127 is N among all HLA allotypes except for HLA-A2, HLA-A24 and HLA-A68 supertype that have K in position 127. HLA-A2 (including A*02 and A*68) and HLA-A24 supertypes were reported as strong TAPBPR binder[168], which implies K127 is important for TAPBPR binding. It would be of interest to check whether the peptide exchange of A*02:01^{K127N} mutant will be enhanced by tapasin in the follow up studies.

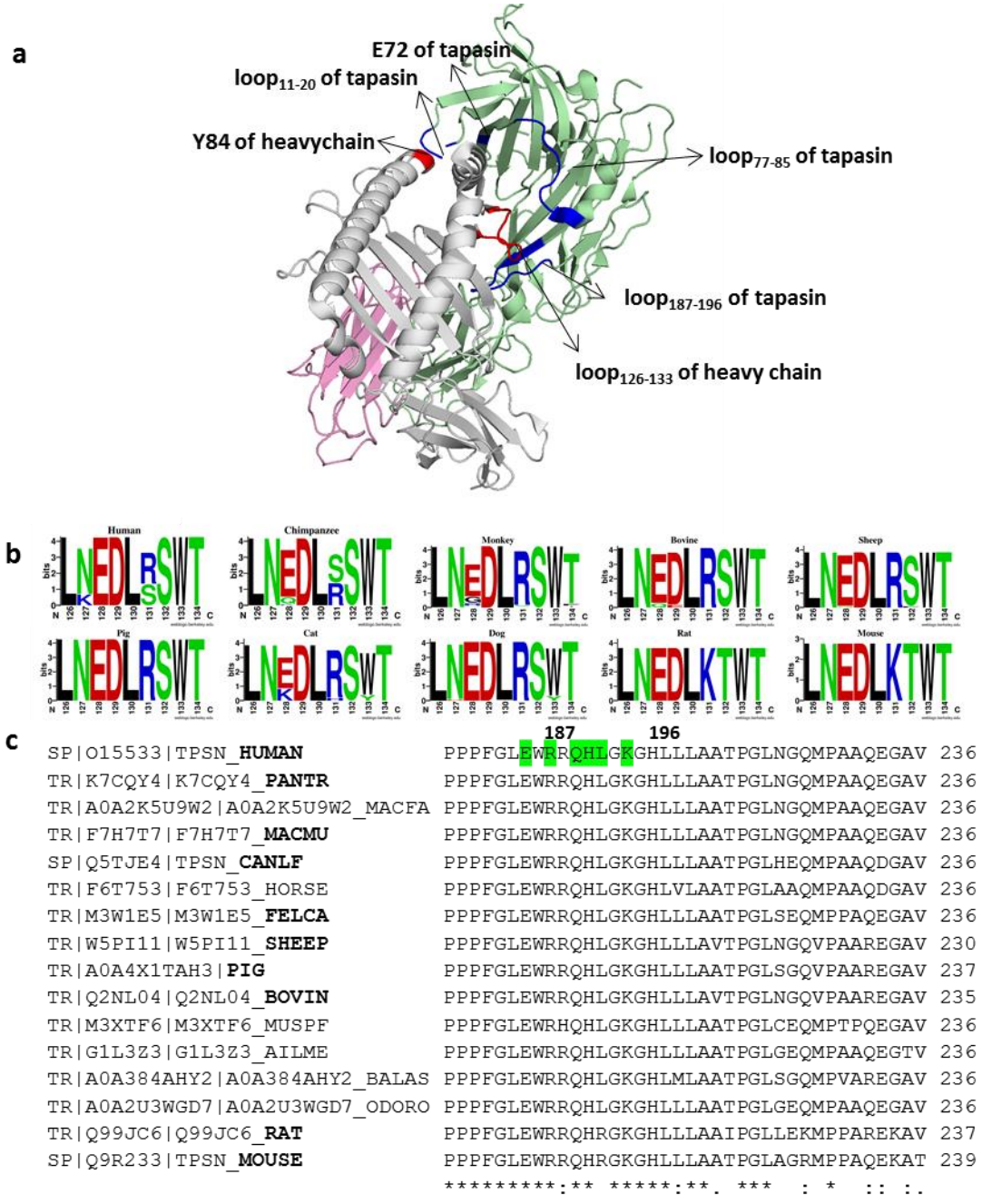


Figure 4.2 The MHC-I heavy chain loop$_{126-133}$ and the tapasin loop$_{187-196}$ are highly conserved across species. a) The interactions between tapasin and MHC-I based on the PLC (PDB ID 6ENY). The loop$_{126-133}$ of heavy chain samples conformations that are important for the binding to tapasin loop$_{187-196}$. After binding with tapasin, E72 is in proximity of the F-pocket region and can interact with Y84, which will affect the dynamics of the F-pocket and subsequently allow the scoop loop$_{11-20}$ to compete for the F-pocket. Loop$_{77-85}$ might interact with the hinge region of the α_{2-1}-helix and affects the dynamics of the binding groove, as indicated by the signal reduction of R151 in the presence of tapasin in NMR. **b)** Sequence logos of residue 126-134 of MHC-I of different species. The loop$_{126-133}$ is highly conserved within species. Sequence logos were

generated using https://weblogo.berkeley.edu. T134 was included because it was shown of importance for PLC complex assembly and next to the $loop_{126-133}$. **c)** Sequence alignment of residues 179-236 of tapasin from different species. Species in bold are the ones analyzed in Fig. 3.5.5c. Residues in human tapasin showing impaired MHC-I binding ability in a previous study are highlighted green[145]. Number 187 and 196 was noted to indicate the beginning and end of the $loop_{187-196}$ which interacts with the highly conserved $loop_{126-133}$ of MHC-I as shown in **a.**

4.2.2 Tapasin modulates the state of F-pocket of MHC-I

A recent publication reported that the switch between a peptide-free unlocked state and a peptide-bound locked state of the F-pocket of MHC-I involves the change of side chain orientations of the residue 97-114-116 triad in the F-pocket of A*02:01 and that the switch between these two states is important for peptide binding[227]. Indeed, in our study, line-broadening of residues N97 (both for B*27:05/photoRL9 and B*27:09/photoRL9) and H114 (in B*27:05/photoRL9, H114 is not assigned in B*27:09/photoRL9) was observed upon UV-exposure. After UV-exposure, MHC-I becomes peptide-receptive and the peptide exchange was faster than the one without UV-exposure (Fig. 3.1.4). In their study they also suggest that the transition between unlocked and locked states of the F-pocket can be modulated by the chaperones tapasin or TAPBPR[227]. Indeed, in our study, residues of the F-pocket (such as N97 and H114 of B*27:05/photoRL9, Appendix Fig. 8b) were affected by the presence of tapasin and overlapped largely with residues of MHC-I that underwent changes upon UV-induced peptide cleavage in the F-pocket. These results could be interpreted as tapasin promotes peptide exchange of pMHC-I by switching the F-pocket to the unlocked state. It is likely that tapasin unlocks the F-pocket state by interrupting the H-bonds of the F-pocket region, illustrated by the observed signal intensity reduction of T143 in the presence of tapasin (Fig. 3.3.3a). Weakening of peptide binding forces by tapasin has been suggested before in a MD simulation study comparing the residue-residue pairwise forces between B*44:02 and the truncated peptide (Δ1C) in tapasin-bound and -free MHC-I[209]. The question then arises which part of tapasin contributes to this loosening of the F-pocket region. Based on the interactions observed between MHC-I and tapasin in the PLC structure, it could be speculated that the tapasin $loop_{11-20}$ adopt a position able to unlock the F-pocket region, although this region is not resolved in the structure of the PLC. Indeed, a recent MD simulation study starting from an atomistic model based on the 9.8 Å PLC Cryo-EM density pointed out that the $loop_{11-20}$ fluctuates between MHC-I and the calreticulin glycan-binding domain in the absence of the Glycan-N86 of MHC-I and establishes directly contacts with the binding groove in the region of the peptide's C-terminus[214]. The authors further suggested that whether the specific interactions between F-pocket and the $loop_{11-20}$ might induce structural rearrangements that need to be further studied. In our NMR measurements, the intensity of heavy chain residue T143, which forms an H-bond with the C-terminus of the peptide, was altered in the presence of tapasin, implying that the $loop_{11-20}$ might modulate the state of the F-pocket.

4.2.3 Y84 of MHC-I might interact with E72 of tapasin

If T143 line broadening really reported on the interruption of H-bond in the F-pocket by tapasin scoop $loop_{11-20}$, one would expect to see no NMR signal reduction of T143 in B*27:09 in the presence of the ERp57-Tsn^{L18G} mutant since this mutant displayed no activity towards B*27:09. However, the affected residues in the F-pocket of B*27:09 in the presence of the ERp57-Tsn^{L18G} mutant are similar to those observed changes he changed residues observed in the presence of wild-type tapasin (although less

affected). This indicates that dynamic changesin the F-pocket are not only affected by the $loop_{11\text{-}20}$ itself or its exact composition. Based on a previous result in which tapasin E72K mutant (E72 corresponds to E105 in TAPBPR, which was showed to form an H-bond with Y84 in TAPBPR-MHC-I complex) showed impaired MHC-I binding and loading activity compared to wild-type tapasin[145], it could be speculated that similar to TAPBPR, E72 of tapasin, which is in direct proximity of the F-pocket region in the structure of the PLC, provides an alternative H-bond to the side-chain of Y84 and thus allows to pull out Y84 from its position in the locked conformation of the F-pocket (Fig. 4.2a). Indeed, in the same MD simulation study mentioned above direct contacts between Y84 of MHC-I and E72 of tapasin were observed[214]. A previous study has shown that the cell surface expression of H2-D^d mutant A139C/Y84C is independent of tapasin, although it can interact with the PLC[228], suggesting a stabilized F-pocket region by introducing an artificial Y84C-A139C disulfide bond is important for peptide binding and the stabilizing effect of Y84C-A139C disulfide bond is similar to the effect by forming an H-bond between Y84 and E72 of tapasin. Thus, we propose that unlocking of the F-pocket by interaction between E72 in the β3 strand (aa 70-76) of tapasin and Y84 of MHC-I allows for the $loop_{11\text{-}20}$ region to place its scoop loop into the F-pocket, thereby stabilizing the pocket and allowing for efficient competition with a bound or incoming peptide (Fig. 4.3).

It is worthy to note that the crystal structure of H-$2D^b$/FP7/TL10 complex (TL10 is a peptide derived from the murine tapasin scoop loop EDAGGGGLSK)[180] on the one hand shows that the L18 of tapasin loop can bind into the F-pocket; On the other hand, it shows that Y84 forms an H-bond with L18 of the loop peptide, not pointing out of the F-pocket[180], which is different from what we discussed above. But it has to be kept in mind that this crystal structure only crystalized a peptide corresponding to the tapasin $loop_{11\text{-}20}$ rather than the full length tapasin with MHC-I. Thus, it cannot be excluded that Y84 can interact with other tapasin residues in the vicinity, like E72, in the presence of the full length tapasin protein.

4.3 The structural elements important for tapasin catalytic activity

After tapasin binds to the MHC-I and turns it into an unlocked and peptide-receptive state, tapasin starts exerting its catalytic function. In this section discussion will be focused on structural elements from both MHC-I molecules and tapasin that contribute to the catalytic capacity of tapasin.

4.3.1 The F-pocket nature of pMHC-I affects its tapasin susceptibility

It has been long suggested that the F-pocket region of pMCH-I is important for its tapasin dependency[30]. Most polymorphic sites of MHC-I molecules are located to the F-pocket region (Fig. 1.5), thereby defining the chemical nature of the C-terminus of the peptide that can be best accommodated by this pocket. Based on the preferentially accommodated amino acid in the F-pocket, the F-pocket can be sorted into five types: hydrophobic, hydrophobic-to-polar, polar (mainly accommodates Y), broad (can accommodate polar, hydrophobic and basic residues) and acidic. The F-pocket nature of the four allotypes used in this thesis are hydrophobic (A*02:01 and B*27:09), broad (B*27:05) and acidic (A*03:01). The influence of tapasin on the peptide exchange kinetics of these four allotypes were investigated. It was found here that the degree of tapasin enhancing peptide exchange rate appears positively related to the acidity of the F-pocket. For example, A*03:01, which has the most acidic F-

pocket among the four allotypes, the peptide exchange rate of it was the most enhanced by tapasin. This is also line with the difference in tapasin dependence of the well-studied allotypes B*44:02 and B*44:05, which differ by only one residue at the bottom of the F-pocket: B*44:02 which has a D at position 116 is tapasin-dependent, while B*44:05 bearing a Y at position 116 is tapasin-independent. The difference in tapasin dependence seen between B*27:05 and B*27:09 (or B*44:02 and B*44:05) was explained by the hypothesis that that the acidic F-pocket in B*27:05 (or B*44:02) is more disordered due to the electrostatic repulsion between F-pocket residues. In contrast, in B*27:09 (or B*44:05) this kind of repulsion is neutralized by H116 (or Y116), thus promoting the higher conformational stability[30]. However, two types of MHC-I allotypes are missing in the experimental set up of this thesis. It would be interesting to examine if the tapasin-induced *Kon* enhancement of allotypes bearing a hydrophobic-to-polar (e.g., B*44:03) or polar (e.g., A*01:01 and A*01:02) F-pocket would follow the same rule.

How is the observation that higher catalytic activity of tapasin on peptide exchange of MHC-I allotypes bearing an acidic F-pocket relevant to the physiological function of tapasin or MHC-I in antigen presentation? It has been shown that tapasin-independent allotypes present broader peptide repertoire on the cell surface[229]. It was further shown that the presented peptides by tapasin-independent allotypes normally have hydrophobic amino acid at their C-termini. For instance, A*02:01 is a promiscuous allotype, accommodating hydrophobic amino acids that are common in proteins. In contrast, A*01:01, which is a fastidious MHC-I allotype and highly tapasin-dependent[156], presenting a narrow repertoire (the peptide motif is more stringent)[229]. Interestingly, a study investigated the effects of different proteasome inhibitors on the cell surface expression of MHC-I and their presented peptide pools[230]. Results showed that peptides with basic carboxyl termini (bind to HLA-B*27:05, HLA-A*68, and HLA-A*03) represent a significant component of the peptides generated by proteasome inhibitor-insensitive proteolytic activities[230]. This perhaps explain the higher activity of tapasin on these allotypes.

4.3.2 The mobility and composition of scoop loop$_{11\text{-}20}$ is important for tapasin activity

The function of tapasin loop$_{11\text{-}20}$ was investigated (section 3.3) and L18 of it was identified as the critical residue mediating peptide exchange (section 3.3.1), which is in line with published studies on the role of Leu both in the tapasin and TAPBPR loops[175,179]. In addition, K16 was found to contribute to the activity of tapasin for allotypes bearing a negatively charged F-pocket (section 3.3.2).

The mutational studies presented here indicate that the mobility of the loop tunes the catalytic activity of tapasin: (i) The mutants ERp57-Tsn^{GGLGG} and ERp57-Tsn^{GGKGG} showed opposite activity for A*03:01 and B*27:09 (Fig. 3.3.2), demonstrating the importance of charge complementarity between the tapasin loop and the F-pocket of MHC-I. (ii) The activity of ERp57-Tsn^{GGLGG} was higher than that of ERp57-Tsn^{GGGGL} towards B*27:09, or the activity of ERp57-Tsn^{GGKGG} was higher than that of ERp57-Tsn^{GGGGK} towards A*03:01, demonstrating the importance of the loop's mobility. (iii) The catalytic activity of ERp57-Tsn^{L18G} (SGKGG) was significantly lower than that of ERp57-Tsn^{GGKGG} towards A*03:01, although the only difference is the presence of a serine in position 14 of the tapasin loop in ERp57-Tsn^{L18G}, indicating that the flexibility of the loop$_{11\text{-}20}$ tunes tapasin activity. We expected to see introduction of an additional K in position 18 would increase the activity of tapasin for A*03:01 as K fit better with the acidic F-pocket, which was not the case here. The reduced activity might be explained

by the electrostatic repulsion between K16, K18, and K20 in the ERp57-Tsn^{L18K} loop. Based on this, we hypothesize that the position of the second important residue like Lys or Arg can be between position 14 and position 17. The observation that tapasin mutant L18G can still affect MHC-I molecules in a similar manner as wild-type tapasin as detected by NMR, suggests that the loop$_{11-20}$ of tapasin is important for catalytic function rather than contributing to binding affinity to MHC-I. Most likely, the scoop loop samples different conformations, leading to a transient occupancy of the F-pocket, and the rebinding of the same peptide might then be prevented while a higher affinity peptide could effectively compete with tapasin.

Additionally, the mutational data of the tapasin loop inspires innovative applications to tailor tapasin to specific MHC-I allotypes and/or alter the spectrum of the immunogenic peptides loaded by tapasin (ERp57-Tsn^{GGKGG} and ERp57-Tsn^{GGLGG}). For example, it would be interesting to re-express ERp57-Tsn^{GGKGG} in tapasin-defective cells to study how tapasin alters the peptide spectrum of MHC-I allotypes bearing a hydrophobic F-pocket while not affecting the peptide editing of MHC-I allotypes bearing an acidic F-pocket. Furthermore, it would also be interesting to test if it is possible to improve the catalytic activity of tapasin by engineering a longer flexible loop containing the critical residue (such as G_nLG_n or G_nKG_n) to load peptide of MHC-I allotypes that are not catalyzed by tapasin.

4.3.3 α3 domain of MHC-I contributes to tapasin catalytic activity by interacting with tapasin C-terminal domain

According to the PLC EM-structure[178], the C-terminal domain of tapasin interacts with the α3 domain of the MHC-I heavy chain. However, in our NMR experiments, none of the assigned residues in the α3 domain of MHC-I were significantly affected in the presence of wild-type or mutant tapasin. In contrast, significant signal reduction of α3 domain residues were observed in the presence of TAPBPR (Fig. 3.2.4i). One reason for this discrepancy could be due to the lack of calreticulin in our system. In the PLC context the C-terminal domain of tapasin is trapped between the long helix of calreticulin and the α3 domain of MHC-I, which limits the mobility of the C-terminal domain of tapasin. Indeed, in the PLC all-atom simulation model from Fisette *et al.*, the calreticulin C-terminal domain was observed to constrain Tsn motions to facilitate the docking of MHC-I molecules to the complex[214]. This might also explain why the peptide exchange enhancement by ERp57-Tsn *in vitro* was generally lower compared to the enhancement by TAPBPR in our experimental set-up (e.g., for A*03:01, 3.5-fold by tapasin vs 22-fold by TAPPBR). TAPBPR does not need other components for stabilizing the interaction with MHC-I. If the transient interaction between MHC-I and tapasin could be stabilized in our system, the peptide exchange should be enhanced more efficiently. Indeed, when the MHC-I molecule is linked to tapasin by a disulfide bond, the peptide exchange rate was increased 21.8-fold for A*03:01 (Fig. 3.4.2d), compared to the 3.5-fold increase caused by the unlinked form of tapasin (Fig. 3.2.2d). Surprisingly, even the peptide exchange of A*02:01, which cannot be promoted by the free form of wild-type tapasin, was increased by 2-fold when linked to tapasin (Fig. 3.4.2e). This suggests that the interaction between the α3 domain of MHC-I contributes to the tapasin catalytic activity. But still, if the peptide exchange of an allotype is not intrinsically susceptible to tapasin, the increase in exchange rate will not be as high as for the tapasin-dependent allotype.

4.4 The mechanism of tapasin-mediated peptide exchange

In summary, the thermostability and dynamics of pMHC-I (that report on peptide/MHC-I interactions) and the polymorphism of MHC-I influence pMHC-I-tapasin interactions, thereby influence the tapasin-mediated peptide exchange and ultimately determine the displayed peptide repertoire. Combining all points discussed above, the mechanism of tapasin-mediated peptide exchange of MHC-I is proposed (Fig. 4.3): MHC-I in complex with suboptimal peptides explores conformations that can be recognized by the hairpin loop ($loop_{187-196}$) of tapasin and thereby allowing for the initial complex encounter. Subsequently, tapasin binding induces the unlocking of the F-pocket region by the interaction between the F-pocket and the scoop $loop_{11-20}$ of tapasin probably, or the interaction between the heavy chain residue Y84 and tapasin residue E72, or both. The scoop $loop_{11-20}$ then competes with the C-terminus of the peptide for the F-pocket until the peptide is kicked out. The scoop $loop_{11-20}$ stabilizes the peptide-receptive state of MHC-I until the high affinity peptide binds to the F-pocket.

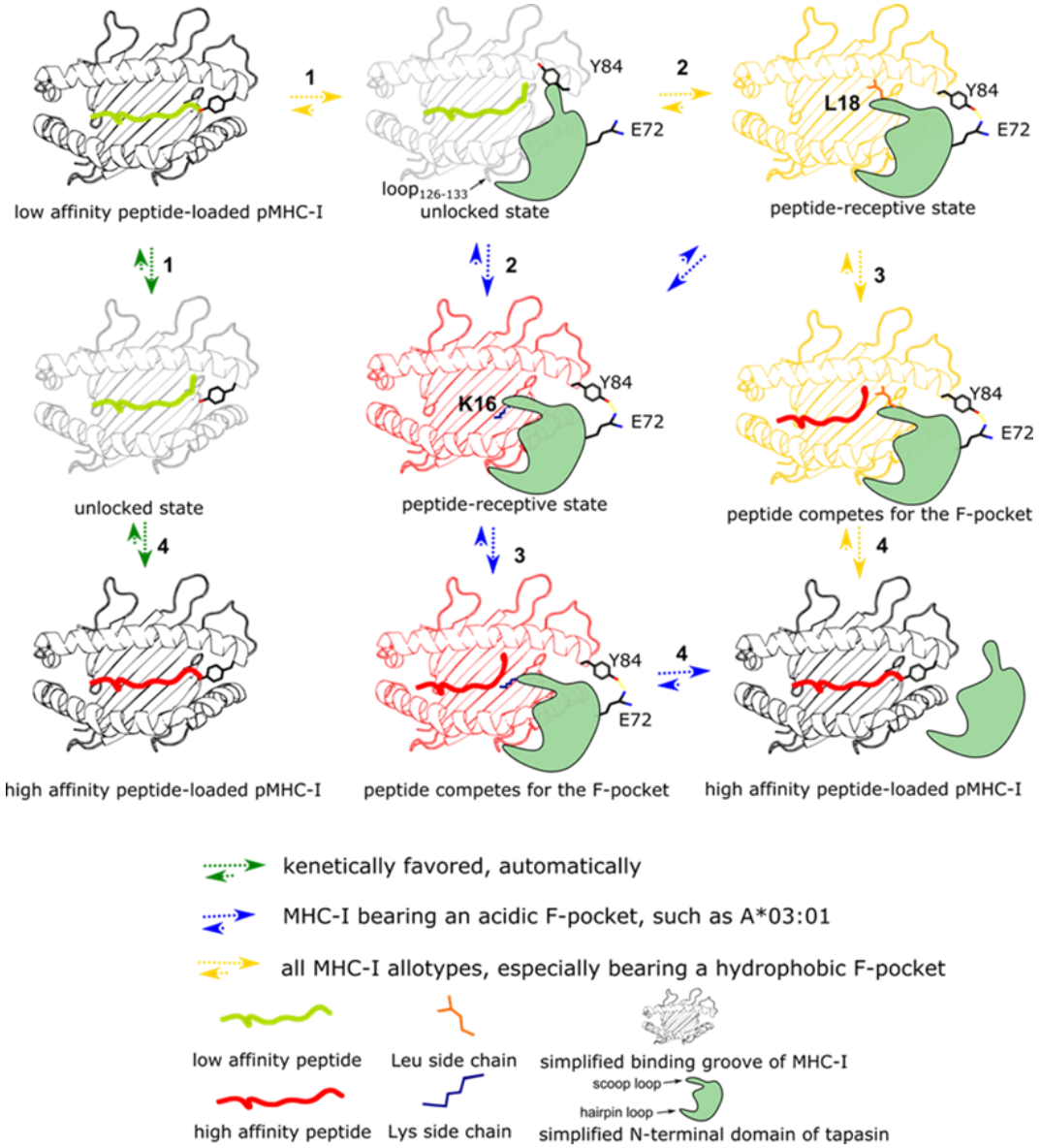


Figure 4.3 The mechanism of tapasin-mediated peptide exchange. Step1: pMHC-I molecules sample conformations, which determines peptide binding and tapapsin binding. Tapasin recognizes a certain state of pMHC-I by approaching pMHC-I from

the α_2-helix side of pMHC-I. Interaction between its $loop_{187-196}$ with the $loop_{126-133}$ of heavy chain provide initial binding which alters the dynamics of the peptide binding groove and subsequently together with the interaction between its E72 in strands β3 and Y84 of the heavy chain promote the C-terminus of peptide to adopt loose binding in the F-pocket accompanied. Now pMHC-I switches to the unlocked state. Allotypes which have low tapasin susceptibility like A*02:01 can switch to peptide unlocked state automatically by the motion of the bound peptide adopting loose binding state. **Step 2:** The scoop $loop_{11-20}$ competes for the F-pocket and stabilize the F-pocket with side chain of L18. Now pMHC-I switches to the peptide receptive state. For allotypes bearing a negatively charged F-pocket, tapasin can also use residue K16 in the $loop_{11-20}$ to compete for the charged F-pocket, which increases the efficiency of tapasin-mediated peptide exchange. **Step 3:** The incoming peptide competes for the F-pocket until the affinity is high enough to replace the side chain. **Step 4:** Once the high affinity peptide binds, tapasin is released from the pMHC-I.

4.5 Comparison of tapasin and TAPBPR

In this section tapasin and TAPBPR will be compared in terms of structure, function and how these differences between them are related to their different roles in the peptide presentation pathway.

4.5.1 Comparison between the functions of the scoop loop in tapasin and TAPBPR

Among the already published crystal structures containing TAPBPR (PDB ID: 5WER and 5OPI) or tapasin (PDB ID: 3F8U), only the scoop loop of TAPBPR in 5OPI was resolved, indicating the flexibility of the scoop loop in both editors. Interestingly, in the resolved scoop loop of TAPBPR, L30 was observed to form a carbonyl-mediated hydrogen bond with a tryptophan residue (W73) in the α1-helix of the heavy chain[174]. While in the structure of H2-D^d73C-TAPBPR complex (PDB ID: 5WER) from Margulies *et al.*, residue 73 was mutated to Cys to form a disulfide bond with the truncated peptide in the binding groove of H2-D^d[173]. This difference might explain why the scoop loop in H2-D^d73C-TAPBPR (PBD 5WER) was not resolved: the flexible scoop loop of TAPBPR cannot be stabilized by the H-bond interaction with α_1-helix of heavy chain. In human MHC-I allotypes, residue 73 is usually a T or I (Fig. 1.5), side chains of both cannot form an H-bond with residues of the TAPBPR scoop loop. This is potentially the reason for failed trials of crystalizing the human-MHC-I-TAPBPR complex[89].

Several groups recently have investigated the function of the TAPBPR scoop loop, and one of these studies found TAPBPR preferentially promoting peptide exchange of MHC-I allotypes bearing a hydrophobic F-pocket[168]. Ilca *et al.* proposed that L30 in the human TAPBPR loop is the key residue to mediate peptide dissociation[175]. Unlike L in the tapasin $loop_{11-20}$, which is evolutionary conserved, L in TAPBPR is not found across species. Most species have an aromatic P at this position of TAPBPR (Fig. 4.4b). Although P has a bulkier side chain and is more hydrophobic, MHC-I molecules with a hydrophobic F-pocket should similarly well accommodate L and P. Sagert *et al.* revealed that the TAPBPR scoop loop functions as an internal peptide-surrogate in peptide-depleted environments stabilizing empty MHC-I molecules and preventing low affinity peptide rebinding[176]. While the importance of dual functions (i.e., chaperone and catalysis functions) of the TAPBPR scoop loop has been well established, a recent study from McShan *et al.* indicates that the long scoop loop (they named the scoop loop as G24-R36) of TAPBPR does not reach into the F-pocket (no chemical shifts in the F-pocket were detected when TAPBPR binding to the MHC-I), but rather sits on top of the F-pocket thereby acting as a kinetic trap for medium-to-high affinity peptides under conditions where peptide concentrations are low[177]. However, in their experiments, they added TAPBPR into UV-exposed MHC-I, in which most of NMR signals corresponding to residues of interest especially in the F-pocket are

already showing line-broadening. Any changes in the F-pocket upon adding TAPBPR cannot be detected afterwards. Instead, in our NMR experiments, although no chemical shifts were observed when adding TAPBPR into pMHC-I, line-broadened peaks of residues in the F-pocket region were detected (Fig. 3.2.4). Thus, NMR data of this thesis support more the idea of peptide-surrogate function for TAPBPR scoop loop.

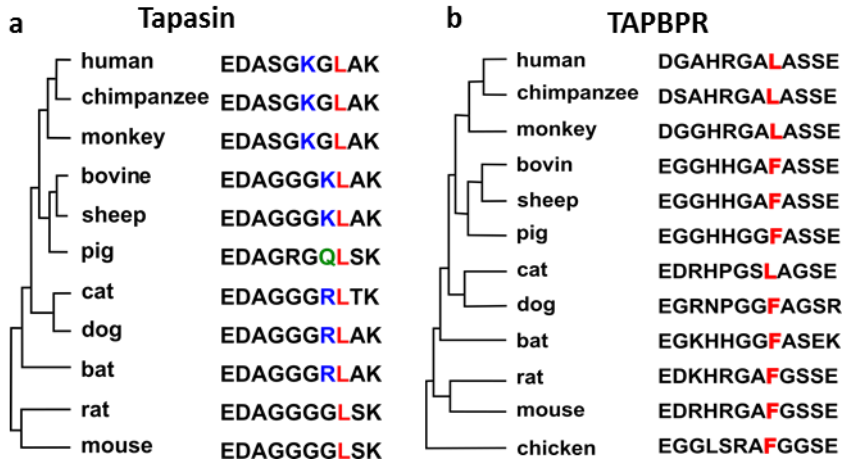


Figure 4.4 a) Evolution tree of tapasin from common species and the alignment of corresponding scoop loop sequence. Basic residue (K and R) is highlighted in blue, L is highlighted in red, and polar residue (Q) is highlighted in olive. **b)** Evolution tree of TAPBPR from common species and the alignment of corresponding scoop loop$_{23\text{-}35}$ sequence. Hydrophobic residue is highlighted in red.

The tapasin scoop loop contains both Leu and Lys residues with long enough side chains to reach into the F-pocket. Furthermore, the tapasin scoop loop is significantly shorter than the TAPBPR scoop loop, it is unlikely that the tapasin loop is long enough to sit on the F-pocket as a kinetic trap. The assumption is that the tapasin scoop loop functions as peptide C-terminus surrogate competing for F-pocket and thereby preventing low-affinity peptide rebinding, considering the peptide concentration is high in the PLC environment.

4.5.2 Comparison between the hairpin loop in tapasin and TAPBPR

Residues in the loop$_{126\text{-}133}$ of B*27:09 displayed line-broadening when adding TAPBPR, indicating that TAPBPR interacts with or/and alters the dynamics of this region. In the crystal structure of MHC-I-TAPBPR, this region is contacted by the "hairpin" loop of TAPBPR[173,174]. Direct interaction between TAPBPR R207 and heavy chain E128, N127, T134 (PDB ID: 5OPI, in other crystal structure PBD ID: 5WER, E128 interacts with Q215 of TAPBPR, N127 and T134 contacts with R207 of TAPBPR), TAPBPR Q209 and heavy chain D121, were observed in the crystal structure of MHC-I-TAPBPR[173,174]. Boyle *et al.* showed that mutating E128 of the MHC-I heavy chain (HLA-A*68) abrogated MHC-I interacting with TAPBPR, while mutating residues E205, R207, Q209 and H210 on TAPBPR affected HLA-A*68 cell surface expression yield[231]. Marozov *et al.* showed that TAPBPR TN6 (E205K, R207E, Q209S, Q272S) mutant completely lost binding ability to A*02:01[89]. In addition, McShan *et al.* showed that residue L126 in the loop$_{126\text{-}133}$ displayed chemical shift changes when H2-D^{d} forms a tight complex with TAPBPR[179]. Key residues are highly conserved in the hairpin sequence of TAPBPR and tapasin,

(Fig. 4.5), which implies a similar mode of interaction between the loop$_{126-133}$ of MHC-I with the respective hairpin loops of tapasin and TAPBPR.

```
  SP|O15533|TPSN_HUMAN           PPPFGLEWRRQHLGKGHLLLAATPGLNGQMPAAQEGAV 236
SP|Q9BX59|TPSNR_HUMAN          PGLDLISVEWRLQHKGRGQLVYSWT-AGQGQAVRKGAT 254
TR|H2Q5A3|H2Q5A3_PANTR         PGLDLISVEWRLQHKGRGQLVYSWT-AGQGQAVRKGAI 254
TR|A0A0D9RDY8|A0A0D9RDY8_CHLSB PGLDLISVEWRLQHKGRGQLVYSWT-TGQGQAVRKGAT 254
TR|F1N3G5|F1N3G5_BOVIN         PGLDVTSVEWRLQHKGSGRLVYHWT-MGQGQAKREGAT 250
TR|M3W6N0|M3W6N0_FELCA         PGLDLTGVEWRLQHKGHGQLVYSWT-TGQRQAKREGAS 239
TR|G1PPR0|G1PPR0_MYOLU         PGLDLARVEWRLQHKGSGQLVYSWT-TGQGQAKREGAT 258
TR|W5PEI3|W5PEI3_SHEEP         PGLDVTSVEWRLQHKGSGQLVYHWT-MGQGQAKREGAT 252
TR|E2RH54|E2RH54_CANLF         PGLVLTSVEWRRQHKGSGQMVYRWT-KGQQHAKREGAI 405
TR|F1SL31|F1SL31_PIG           PGLDLTSVEWRQQHGGSGHLVYSWTTMGQGQAQREGAT 251
TR|D4A6L1|D4A6L1_RAT           PGLDLTGVEWRLQHKGSGQLVYSWK-TGQGQAKRKGAT 255
SP|Q8VD31|TPSNR_MOUSE          PGLDLTGVEWRLQHKGSGQLVYSWK-TGQGQAKRKGAT 259
                               *** :  **** ** * *::** *.  ** :* *:**
                                        207      215
```

Figure 4.5 The sequence of the hairpin loop of tapasin and TAPBPR (206-216). The first row in bold is human tapasin hairpin loop and residues in green are critical residues for MHC-I binding. The rest rows are TAPBPR sequences of different species. The number 207 and 215 notes the start and end residue of the hairpin loop in TAPBPR.

4.5.3 Tapasin and TAPBPR might recognize different conformational states of MHC-I

McShan *et al.* analyzed the dynamics of four MHC-I molecules (A*01:01, A*02:01, H2-D^d, H2-L^d) by methyl CPMG relaxation dispersion and revealed allele-specific dynamics at the interface with TAPBPR, both near and distant to the peptide binding groove[169,179]. They suggested that allele-specific dynamics determine the susceptibility to TAPBPR catalysis through recognition of the minor excited state sampled by MHC-I molecules. They showed reduction in dynamics by introducing a disulfide bond in the peptide binding groove (Y84C-A139C) of H2-D^d that abrogated TAPPBR binding. In this thesis, the allele-specific catalytic activity of TAPBPR was analyzed by FP assays (section 3.2.2), and the order of peptide catalyzing activity was in line with a previous cellular result[168]. In addition, a positive relation between the number of residues showing dynamics and its susceptibility to TAPBPR-mediated peptide exchange was observed (section 3.2.5). Since the susceptibility to tapasin-mediated or TAPBPR-mediated peptide exchange of a certain allotype shows an inverse trend, it is possible that tapasin recognizes a different conformational state sampled by MHC-I based on the discussion above.

Fisette *et al.* compared the dynamics of peptide-loaded/peptide-deficient MHC-I and MHC-I-Tsn complexes by molecular dynamics (MD) simulations and found that tapasin modulates the width of the binding groove in the peptide-deficient conformation[232]. Tapasin is thought to lower the fluctuations of MHC-I, especially in the α1 and α_2 helices of the binding groove, to stablize the conformation. Recently they performed the all-atom MD simulations of the PLC based on the cryo-EM density on a microsecond timescale and found reduced dynamics of MHC-I in the environment of the PLC as compared to the sole or merely tapasin-bound MHC-I[214]. However, the peptide-deficient conformation is still observed in the full complex of the PLC[214]. The MHC-I in the PLC exerts higher fraction of peptide-loaded conformation which is ready for peptide binding. This indicates that in the PLC, tapasin recognizes the ground state (peptide-loaded state but loosely bound state as mentioned in section 4.1.1, the major state of pMHC-I) and then stabilize this conformation for peptide binding. Consequently, the more dynamic pMHC-I molecules show, the less the peptide exchange is susceptible to tapasin. Indeed, in section 3.2.2, we showed that B*27:05/photoRL9 bears fewer dynamic residues compared to B*27:09/photoRL9,

while the peptide-exchange of it was enhanced 2.5-fold by tapasin, higher than the 1.7-fold of B*27:09/photoRL9. We also examined the dynamics of other HLA allotypes and could explain the correlation between the thermostability and tapasin dependence with this hypothesis[203]. Namely, the lower the thermostability of a pMHC-I complex is, the higher population the major ground state it has, and the more tapasin-dependence it is. For example, the T_m of A*03:01/photoRL9 was the lowest (Appendix Table 3) and the tapasin dependence of it is the highest among four allotypes tested.

4.5.4 The role of the two editors in the antigen presenting pathway

Why are there two peptide editors in the antigen presenting pathway? Although experiments have shown that both TAPBPR and tapasin can chaperon and promote peptide exchange of MHC-I molecules, whether they display unique, overlapping or redundant function in the antigen presentation pathway is still unclear. Tapasin and TAPBPR have different effects on the presented peptide pool on the cell surface, as deletion of tapasin clearly alters the spectrum of presented peptides[140,141,143], while deleting TAPBPR has only a subtle effect[87]. Given the different locations of tapasin and TAPBPR in the ER, one could speculate TAPBPR being a complementary peptide editor as a second quality control of MHC-I molecules. In our study, the peptide exchange enhancement of one MHC-I allotype by TAPBPR or tapasin was inversely correlated, such as peptide exchange of A*02:01 is highly TAPBPR-susceptible while is tapasin-independent in our experimental system. On top of that, we showed different residues of pMHC-I were affected by adding tapasin and TAPBPR (section 3.2.4), indicating they might apply different mechanisms to recognize MHC-I molecules, thereby resulting in different catalytic activity for different MHC-I allotypes. Our results seem to support the supplementary relationship between them.

Tapasin and TAPBPR residents in different cellular compartments. TAPBPR is not only restricted in the ER, but it can also travel to the Golgi together with MHC-I molecules[86]. pH differs in cellular compartments (ER: pH ~7.2, cis-Golgi: pH ~6.7, trans-Golgi: pH ~6.0)[171]. It has been shown that tapasin works optimally in a near neutral pH, while TAPBPR functions in a relatively wide pH range as it moves through the secretory pathway. It is unclear in which subcellular environment TAPPBR functions as peptide editor[171]. As mentioned previously, in contrast to tapasin, TAPBPR prefers to bind to MHC-I molecules bearing a hydrophobic F-pocket[168]. If TAPBPR predominantly operates in slightly acidic environments, the preference towards a hydrophobic F-pocket can be explained by the lack of stability of the empty acidic F-pocket due to the electrostatic repulsion in MHC-I molecules after they leave the PLC.

The peptide sources also vary in different cellular environments. The peptide source in the ER but not in the PLC mainly consists of the signal peptides cleaved from proteins during their transport into the ER for refolding[233,234]. Most of them are hydrophobic residues, which is in accordance with TAPBPR preferring to act on MHC-I with a hydrophobic F-pocket. On the contrary, peptides loaded in the PLC are mainly TAP-transported peptides. As we mentioned in section 3.3.5, human TAPs transfer peptides containing a basic or hydrophobic C-terminus. It is still unclear whether the peptide in the PLC will diffuse to the other part of ER and where the peptide source in Golgi is.

Lastly, peptide concentration is also different in cellular environments. It can be assumed that peptide concentration is a critical parameter affecting the finally presented peptide repertoire. Tapasin primarily serves its peptide editing function within the PLC, an environment rich in peptides that are optimal for MHC class I binding. In contrast, the peptide concentration where TAPBPR operates is considerably lower compared to the PLC. Once MHC-I molecules are loaded with peptides, they leave the PLC and travel through other subcellular environments. It is likely that the suboptimal peptides can dissociate faster in the low peptide concentration environment. The primary function/task of TAPBPR in this regard would be to bind to suboptimal-loaded MHC-I in peptide limiting concentrations thereby preventing the dissociation of peptides. This function/task is indicated by experimental evidence that showed TAPBPR can form a complex with relatively low-affinity-peptide-loaded MHC-I[86,168,169]. In contrast, tapasin has not been observed to form complex with low-affinty-peptide-MHC-I. In case of high affinity peptides abundantly present, TAPBPR promotes the exchange of the relatively low affinity peptide for high affinity peptide and release the high-affinity-peptide-loaded MHC-I[89].

4.6 Co-evolution of proteins in antigen processing and presenting pathway

Research studies have descripted the evolution of the MHC genomic organization across different species[235–238]. MHC loci of mammals and non-mammals display different arrangements of genes involved in antigen processing, loading and presentation. The best characterized non-mammalian MHC-I is that of chicken, in which the *tapasin*, *TAP* and *MHC-I* loci form a minimal essential region that cannot be disrupted by recombination[239]. Co-evolution of the components is clearly seen in the chicken MHC-I system, as the specificity of peptide translocated by the *TAP* alleles correlates well with the peptide motif of chicken MHC-I molecule that comes from the same haplotype[240]. On top of that, there are natural chicken tapasin variants that show allele-specific functionality towards their associated chicken MHC-I[241]. Such co-evolution is not found in most mammalian MHC-I systems, because in human and other mammals *tapasin*, *TAP* (*TAP1/TAP2*) and *LMP* (absent in chicken) genes are located far away from the MHC-I locus (e.g., in the extended class II region, Fig 4.6a). Few sequence alleles of *TAP*, *tapasin* and *LMP* have been reported (e.g., for human shown in Fig. 1.3a) and they appear to be functionally monomorphic[242–244]. Thus, it seems likely that within most mammals, tapasin, TAP and LMP function universally towards whichever MHC-I molecules are expressed. The ability of TAP, tapasin and LMP to work on different MHC-I allotypes might have been acquired through co-evolution along MHC-I diversification.

Several studies showed that the human immunoproteasome that contains the INF-γ inducible LMP-2 and LMP-7 has different cleavage preferences compared to the usual proteasome[211,245]. These studies suggested that the proteasome containing LMP2/LMP7 chain generates peptides with carboxy-termini adapted for the preferential loading of the MHC-I molecules by enhancing cleavage after positively charged or hydrophobic residues. It has been known that the carboxy termini of peptides loaded on the MHC-I are either positive or hydrophobic. Accordingly, the INF-γ inducible immunoproteasome has co-evolved to generate loadable peptides for the MHC-I molecules. A study investigated the effects of different proteasome inhibitors on the expression of MHC-I and their presented peptide pools. Results showed that peptides with basic carboxyl termini (bind to HLA-B*27:05, HLA-A*68, and HLA-A*03)

represent a significant component of the peptides generated by proteasome inhibitor-insensitive proteolytic activities[230]. This demonstrated that the emergence of the inhibitor-insensitive immunoproteasome (which is absent in chicken) is to provide peptides for MHC-I alleles that have more stringent peptide motifs.

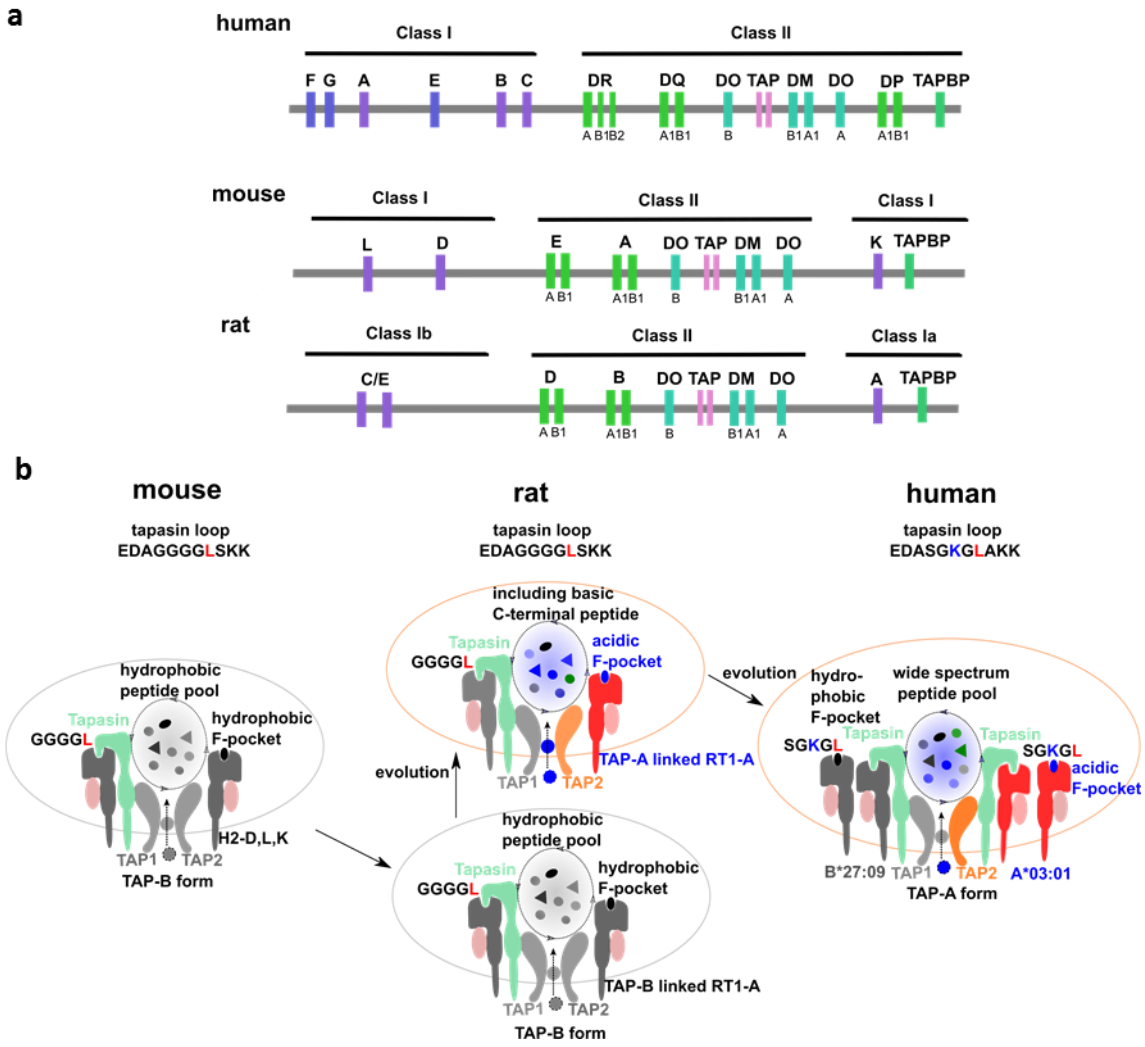


Figure 4.6 two routes of co-evolutions happened in antigen processing pathway. a) Simplified MHC locus of human, mouse and rat. Only important genes are displayed. **b)** co-evolution of TAP, MHC-I and tapasin in a simplified peptide loading complex. Rat has both TAP-A and TAP-B type transporters. The color background of peptide pool in the ER shaped by TAP represents the peptide nature. MHC-I bearing a hydrophobic F-pocket is shown in grey and bearing an acidic F-pocket in red. The $loop_{11\text{-}20}$ sequence of tapasin is indicated on the top. Two interwoven routes of evolution in order to foster the presentation of peptides of diverse composition comparison of mouse, rat and human Note: human TAP2 and rat TAP2 are independent evolutionary events[68].

Interestingly, we observed that the composition of the functional $loop_{11\text{-}20}$ matched well with the F-pocket nature of MHC-I allotypes of different species (mammals, Fig. 3.3.5c and 3.3.5e) with rat being the only exception. This exception can be explained by the presence of two allelic variants of TAPs (TAP-A and TAP-B) in rat, that deliver peptides based on the C-terminal sequence and were reported to co-evolve with rat MHC-I molecules[212,213,246]. We hypothesize that the TAP-A form is an evolutionary route realized in rat to additionally present peptides with basic residues, thus leading to higher immunological fitness (Fig. 4.6b). However, the older TAP-B gene is maintained in the rat population

and it was argued that tapasin might serve as a link helping to explain the maintenance of this variant[213]. Based on our data we suggest that rat strains containing TAP-A may have a functional disadvantage in terms of tapasin-mediated peptide exchange, since rat tapasin contains only the critical Leu within its loop (ERp57-Tsn^{GGGGL}). This loop is best suited to serve peptide exchange of TAP-B gene-coupled rat MHC-I allotypes with a hydrophobic F-pocket. For species other than rat we suggest that they have exploited another path to enhance antigenic diversity. According to our mutational data they rely on tapasin molecules that evolved a single loop element to efficiently catalyze loading of peptides containing either basic, polar or hydrophobic C-terminal residues. Pathogens might have provided the selective pressure for co-evolution of molecules involved in the peptide presentation pathway and it likely was driven by the requirement to present a diverse set of peptides with high efficiency to fight infections more effectively.

4.7 The limitations and outlook of this research

Given the high polymorphism in the F-pocket of MHC-I, the statement that tapasin predominantly catalyzes MHC-I bearing an acidic F-pocket is premature in this regard. It is unclear if the proposed mechanism of tapasin-mediated peptide exchange (especially about the role of the tapasin scoop loop) holds true for other MHC-I allotypes bearing an acidic (e.g., A*11:01 and A*30:01) or polar (e.g. A*01:01) F-pocket. Systematically generating F-pocket mutants of A*03:01 (D116Y) and A*02:01 (Y116D) to test the peptide exchange with several tapasin scoop loop mutants would also strengthen the conclusion.

The mobility and composition of scoop loop$_{11-20}$ is important for tapasin activity. The scoop loop mutant L18G and GGKGG behave differently on A*03:01 (but not B*27:09), although the only difference between them is the Ser at position 14. Thus, it is of importance to check the role of this residue in peptide exchange of tapasin by using the S14G mutant.

Similar patterns of altered dynamics in the F-pocket of pMHC-I in the presence of ERp57-Tsn^{WT} and the mutant ERp57-Tsn^{L18G} (no catalytic activity) were observed in NMR experiments (section 3.3.3). This result leads to the speculation that residue L18 in the scoop loop is critical for the catalytic activity, but not for the binding ability of tapasin. However, this speculation seems unconvincing due to the lack of the binding affinity data between tapasin mutants and MHC-I. It would be beneficial to measure the affinity data between pMHC-I and different tapasin scoop loop mutants by SPR or ITC to obtain more evidence in support of this hypothesis.

As discussed in section 4.1.1, the minor state population of pMHC-I molecules that exert conformational changes cannot be calculated from the CPMG-data measured only at one magnetic field strength. Thus, the speculation that tapasin recognizes a conformation different from the one (the rare state) recognized by TAPBPR is premature. It would be beneficial to measure at another field strength to calculate the minor state population.

The speculation of co-evolution existing between tapasin and MHC-I across species seems unconvincing without additional experimental data. In order to further elude the co-evolution between

tapasin and MHC-I molecules, experiments with mouse and rat MHC-I and their corresponding tapasin are needed.

As descripted in section 3.4.4, the low resolution of the MHC-I-ERp57-Tsn complex impedes modeling. Further experiments including intruding mutation to stabilize the complex, screening adding different detergents are needed to improve the resolution.

5 Summary

Tapasin, the central component of the peptide loading complex (PLC), acts as the peptide editor to shape the peptide pool presented by MHC-I on the cell surface for immune recognition. The previously published Cryo-EM structure of the PLC provides insight into how tapasin interacts with MHC-I as well as other PLC components. However, the molecular mechanism of tapasin-mediated peptide exchange of MHC-I remains elusive, given the inherent dynamics of critical regions within the complex. Here, dynamic changes of MHC-I in the presence of tapasin as detected by NMR spectroscopy were investigated and conformational dynamics affected by tapasin binding were observed in two regions of the MHC-I molecule. Based on that, I propose tapasin exploiting two essential structural features of MHC-I: First, tapasin samples a conserved allosteric site underneath the $\alpha_{2\text{-}1}$-helix of MHC-I, "loosening" the F-pocket region that binds with the C-terminus of the peptide. Second, tapasin scoop $loop_{11\text{-}20}$ transiently samples the F-pocket, enabling competitive peptide binding. By investigating the role of individual residues of the $loop_{11\text{-}20}$ for tapasin-catalyzed peptide exchange *in vitro*, tapasin residues L18 and K16 was defined to be critical for tapasin's allele-specific activity.

Furthermore, I compared dynamic profiles of four human MHC-I allotypes and their susceptibilities to tapasin-mediated or TAPBPR-mediated peptide exchange. The inverse dependence of a particular allotype on these two editors was observed, which suggested that tapasin and TAPBPR apply different recognition models. By comparing residues affected upon tapasin or TAPBPR binding, differently affected residues of MHC-I were identified. Based on that, it can be proposed that tapasin turns MHC-I into an "unlocked" conformation after binding to MHC-I instead of recognizing a minor state, a mechanism applied by TAPBPR. The study here provides a comprehensive comparison between tapasin and TAPBPR.

Moreover, I observed that the composition of the functional $loop_{11\text{-}20}$ matched well with the F-pocket nature of MHC-I allotypes of different species. The mechanistic findings unravel the co-evolution of the composition of the tapasin scoop loop and the F-pocket composition of mammalian MHC-I allotypes. Thus, regions of evolutionary conservation and allelic diversification cooperate to provide structurally malleable elements in MHC-I molecules tailored towards tapasin editing and the presentation of a broadened antigenic repertoire.

Lastly, we obtained the cryo-EM structure of the MHC-I-ERp57-Tsn complex by introducing cysteine mutations on β2m and Tsn for disulfide linkage. We observed that the peptide exchange rate was enhanced dramatically when using the disulfide-linked complex to perform peptide exchange assays for some allotypes (such as A*03:01 and A*02:01) but not for B*27:09. We revealed the difference in enhancement of peptide exchange was due to the chaperone function of tapasin. Thus, we provided the *in vitro* experimental evidence to show tapasin's dual function: catalyzing peptide exchange and stabilizing the empty MHC-I. We propose the feasibility to use the disulfide bond constructs for cell surface peptide exchange experiments *in vitro*, showing the potential in application of building cell libraries for characterizing antigen-specific TCR repertoires and identifying immunodominant clones.

6 Zusammenfassung

Tapasin, die zentrale Komponente des Peptid-Beladungs-Komplex (PLC), fungiert als Peptid-Editor, um den Peptidpool zu modulieren, der von MHC-I auf der Zelloberfläche zur Immunerkennung präsentiert wird. Die aktuell veröffentlichte Cryo-EM-Struktur des PLC gibt Aufschluss darüber, wie Tapasin mit MHC-I sowie anderen PLC-Komponenten interagiert. Der molekulare Mechanismus des Tapasin-vermittelten Peptidaustauschs von MHC-I bleibt jedoch aufgrund der inhärenten Dynamik der kritischen Regionen innerhalb des Komplexes schwer zu verstehen. Hier Hier habe ich die dynamischen Veränderungen von MHC-I in Gegenwart von Tapasin mittles NMR-Spektroskopie untersucht und konnte Konformationsdynamiken beobachten, die durch die Tapasin-Bindung in zwei Regionen des MHC-I-Moleküls beeinflusst werden. Die Ergebnisse zeigen, dass Tapasin zwei wesentliche strukturelle Merkmale von MHC-I ausnutzt: Erstens tastet Tapasin eine konservierte allosterische Stelle unterhalb der $\alpha_{2\text{-}1}$-Helix von MHC-I ab und "lockert" so die F-Taschenregion, die den C-Terminus des Peptids beherbergt. Zweitens bindet die Schaufel-artige Schleife$_{11\text{-}20}$ von Tapasin vorübergehend die F-Tasche, was eine kompetitive Peptidbindung ermöglicht. Wir untersuchten im Detail die Rolle einzelner Reste der Schleife$_{11\text{-}20}$ für den Tapasin-katalysierten Peptidaustausch *in vitro* und zeigten die kritische Rolle der Tapasin-Reste L18 und K16 für Tapasins Allel-spezifische Katalyse-Aktivität.

Darüber hinaus habe ich die dynamischen Profile von vier humanen MHC-I-Allotypen und deren Empfindlichkeit für den Tapasin- oder TAPBPR-vermittelten Peptidaustauschs verglichen. Es zeigte sich eine entgegengesetzte Abhängigkeit eines bestimmten Allotyps von den beiden Peptid-Editoren was nahe legt, dass Tapasin und TAPBPR durch unterschiedliche Erkennungsmodelle beschrieben werden können. Der Vergleich der die Reste von MHC-I, die bei der Bindung von Tapasin oder TAPBPR eine Rolle spielen, zeigte, dass unterschiedliche Reste betroffene sind. Das impliziert, dass Tapasin nach der Bindung eine "entriegelte" Konformation in MHC-I induziert, anstatt einen seltenen konformationellen Zustand zu erkennen, ein Mechanismus, der von TAPBPR angewendet wird. Meine Arbeit liefert einen umfassenden Vergleich zwischen Tapasin und TAPBPR.

Darüber hinaus konnt ich beobachten, dass die Zusammensetzung der Schleife$_{11\text{-}20}$ gut mit den F-Taschen-Eigenschaften von MHC-I-Allotypen in verschiedenen Spezies korreliert. Unsere mechanistischen Erkenntnisse entschlüsseln die Co-Evolution der Zusammensetzung der Tapasin-Funktionsschleife und der F-Pocket-Zusammensetzung der MHC-I-Allotypen von Säugetieren. Somit kooperieren Regionen evolutionärer Konservierung und allelischer Diversifizierung, um strukturell formbare Elemente in MHC-I-Molekülen bereitzustellen, die auf die Tapasin-Editierung und die Präsentation eines erweiterten Antigen-Repertoires zugeschnitten sind.

Schließlich erhielten wir die Cryo-EM-Struktur des MHC-I-ERp57-Tsn-Komplexes durch Einführung von Cystein-Mutationen auf β2m und Tsn für die Disulfid-Bindung. Wir beobachteten, dass die Peptidaustauschrate bei der Verwendung des disulfidverknüpften Komplexes zur Durchführung von Peptidaustauschassays für einige Allotypen (z. B. A*03:01 und A*02:01) drastisch erhöht war, nicht aber für B*27:09. Wir konnten zeigen, dass der Unterschied in der Erhöhung der Peptidaustauschrate auf die Chaperonfunktion von Tapasin zurückzuführen ist. Damit lieferten wir den experimentellen *in*

vitro Beweis für die Doppelfunktion von Tapasin: Katalysierung des Peptidaustauschs und Stabilisierung des MHC-I im Peptid ungebundenen Zustand. Außerdem haben wir das Konzept des Peptidaustauschs auf der Zelloberfläche unter Verwendung der Disulfid verknüpften Konstrukte dargerlegt, welches Anwendungspotenzial für den Aufbau von Zellbibliotheken zur Charakterisierung von Antigen-spezifischen TCR-Repertoires und zur Identifizierung von immundominanten T-Klonen bietet.

7 Appendix

Appendix Table 1. Data collection and refinement statistics of crystal structure of B*27:09/photoRL9 (PDB ID 7ALO).

Data collection	
Wavelength [Å] [a]	0.91841
Resolution [Å] [a]	50.00 - 1.80 (1.92 - 1.80)
Space group	*P*1
Unit cell a;b;c [Å]	45.2; 69.6; 82.6;
α; β, γ [°]	80.2; 88.3; 89.9
Total reflections [a]	333,000 (59,466)
Unique reflections [a]	89,225 (15,619)
Multiplicity [a]	3.7 (3.8)
Completeness [%] [a]	97.1 (96.8)
<I/□(I)> [a]	5.4 (1.06)
Wilson B-factor	32.2
R_{meas} [%] [a, b]	19.8 (161.9)
$CC_{1/2}$ [c]	99.0 (45.0)
Refinement	
Resolution [Å] [a]	50.00 - 1.80 (1.82 - 1.80)
Test set [%]	5.0
Non-hydrogen atoms	7346
R_{work} [%] [a, d]	22.3 (38.1)
R_{free} [%] [a, f]	26.5 (39.9)
Asymmetric unit	
Heavy chain (residues)	276 (chain A) / 276 (chain D)
β2m (residues)	99 (chain B) / 99 (chain E)
peptide (residues)	9 (chain C) / 9 (chain F)
Ligand molecules (PEG3350, glycerol, Tris)	17
water molecules	527
Average B-factor [Å^2]	34.3
heavy chain	32.0 (chain A) / 35.1 (chain D)
β2m	32.8 (chain B) / 33.4 (chain E)
peptide	44.9 (chain C) / 44.9 (chain F)
ligands (PEG3350, glycerol, Tris)	51.7
water molecules	36.4
r.m.s.d. [g] (bonds) [Å]	0.006
r.m.s.d. [g] (angles) [°]	1.032
Validation	
Ramachandran favoured [%]	96.8
Ramachandran allowed [%]	3.2

Ramachandran outliers [%]	0.0
Rotamer outliers [%]	1.27
Molprobity score [i]	1.55
Molprobity clash score [i]	5.47

[a] Values in parentheses refer to the highest resolution shells.

[b] $R_{meas}(I) = \sum_h [N/(N-1)]^{1/2} \sum_i \left| I_{ih} - \langle I_h \rangle \right| / \sum_h \sum_i I_{ih}$, in which $\langle I_h \rangle$ is the mean intensity of symmetry-equivalent reflections h, I_{ih} is the intensity of a particular observation of h and N is the number of redundant observations of reflection h[247].

[c] $CC_{1/2} = (\langle I^2 \rangle - \langle I \rangle^2) / (\langle I^2 \rangle - \langle I \rangle^2) + \sigma^2_{\varepsilon}$, in which σ^2_{ε} is the mean error within a half-dataset[248].

[e] $R_{work} = \sum_h \left| F_o - F_c \right| / \sum F_o$ (working set, no σ cut-off applied).

[f] R_{free} is the same as R_{work} but calculated on the test set of reflections excluded from refinement.

[g] r.m.s.d. – root-mean-square deviation.

[I] Calculated with MolProbity[192].

Appendix Table 2.

Table of affinity data of peptides used in this paper

Peptide name	Allotype	Sequence	* Predicted affinity (nM)
photoKV9	A*02:01	KILGFVFJ*V	2.08
photoKK9	A*03:01	KLIETYFJ*K	17.74
photoRL9	B*27:05, B*27:09	RRKWRRWJ*L	35.64 for B*27:05 196.44 for B*27:09
photoIF9	B*27:05, B*27:09	IRAAPPPJ*F	720.18 for B*27:05 1762.61 for B*27:09
photoRL9-4J*	B*27:05, B*27:09	RRKJ*RRWHL	49.46 for B*27:05 357.56 for B*27:09
FITC-GV9	A*02:01	GILGK^{FITC}VFTV	5.95
FITC-RL9	B*27:05, B*27:09	RRKWK^{FITC}RWHL	72.27 for B*27:05 476.51 for B*27:09
FITC-IF9	B*27:05, B*27:09	IRAAK^{FITC}PPLF	385.52 for B*27:05 2143.59 for B*27:09
FITC-KK9	A*03:01	KLIEK^{FITC}YFSK	18.04
KV9	A*02:01	KILGFVFTV	2.31
KK9	A*03:01	KLIETYFSK	21.68
RL9	B*27:05, B*27:09	RRKWRRWHL	50.24 for B*27:05 322.6 for B*27:09
IF9	B*27:05, B*27:09	IRAAPPPLF	479 for B*27:05 1589 for B*27:09

*affinity values were predicted by NetMHCpan4.1 Server.

J* was treated as F when performing the prediction in NetMHCpan4.1 Server

Appendix Table 3.

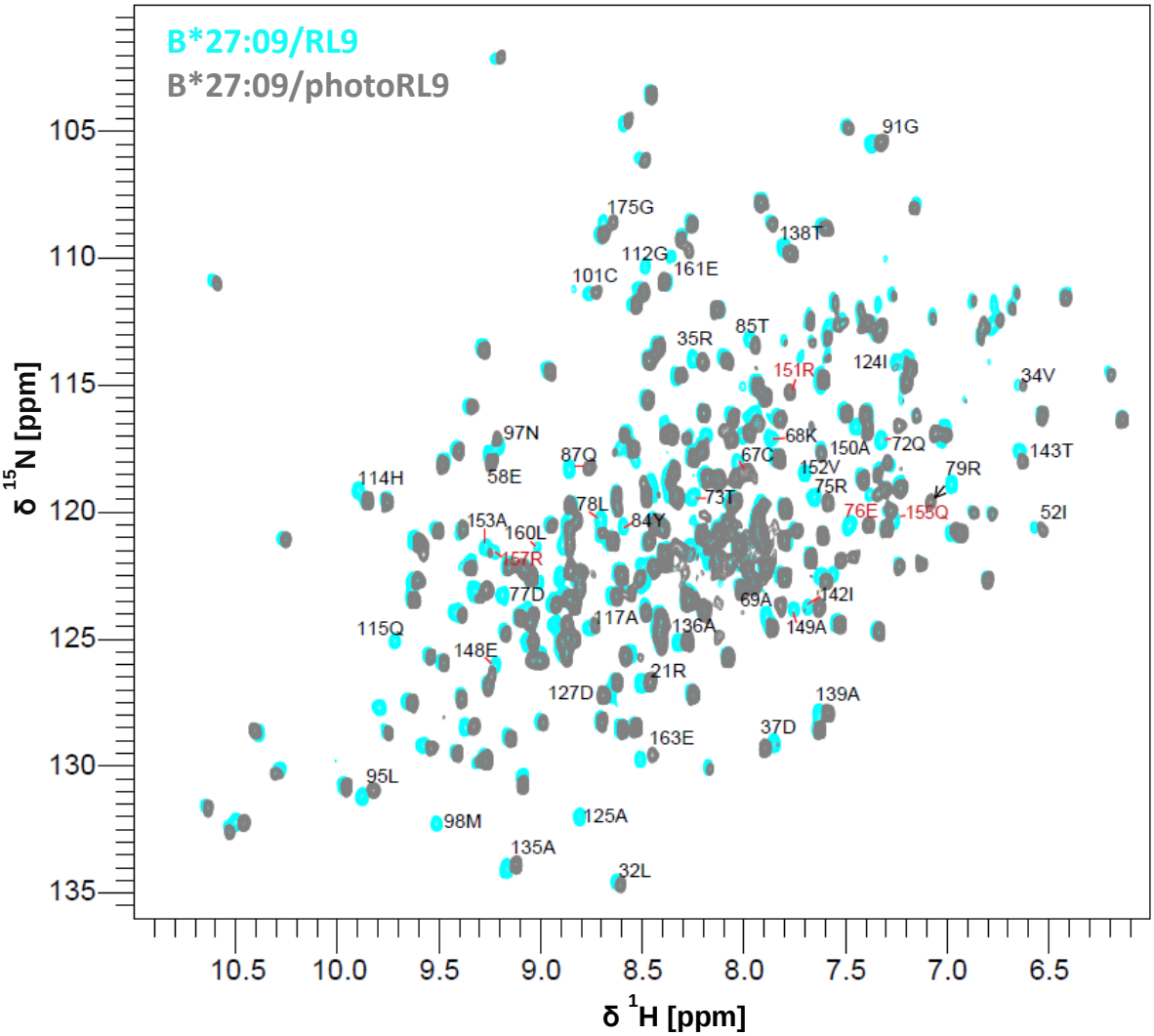
List of melting temperatures of pMHC-I complexes

Allotype	Refolded peptide	T_m before UV (°C)	T_m after UV (°C)
A*02:01	photoKV9	51.41 ± 1.27	37.81 ± 0.57
A*03:01	photoKK9	51.43 ± 0.79	51.39 ± 0.13
	$\beta2m^{T5C}$/photoKK9	51.48 ± 0.04	50.44 ± 0.03
	$\beta2m^{T5C}$/photoKK9/ERp57-$Tsn^{K16L-G304C}$		50.67 ± 0.13
B*27:09	photoRL9	54.55 ± 0.13	49.86 ± 0.04
	photoIF9	57.94 ± 0.18	42.80 ± 0.37
	photoRL9-4J*	49.10 ± 0.04	47.32 ± 0.15
	$\beta2m^{T5C}$/photoRL9/ERp57-$Tsn^{GGLGG-G304C}$		51.52 ± 0.15
	IF9	56.20 ± 0.13	
	RL9	60.20 ± 0.18	
B*27:05	photoRL9	57.58 ± 0.08	50.72 ± 0.04
	photoIF9	51.13 ± 0.21	48.01 ± 0.14
	photoRL9-4J*	56.40 ± 0.07	44.92 ± 0.25
	RL9	61.88 ± 1.29	

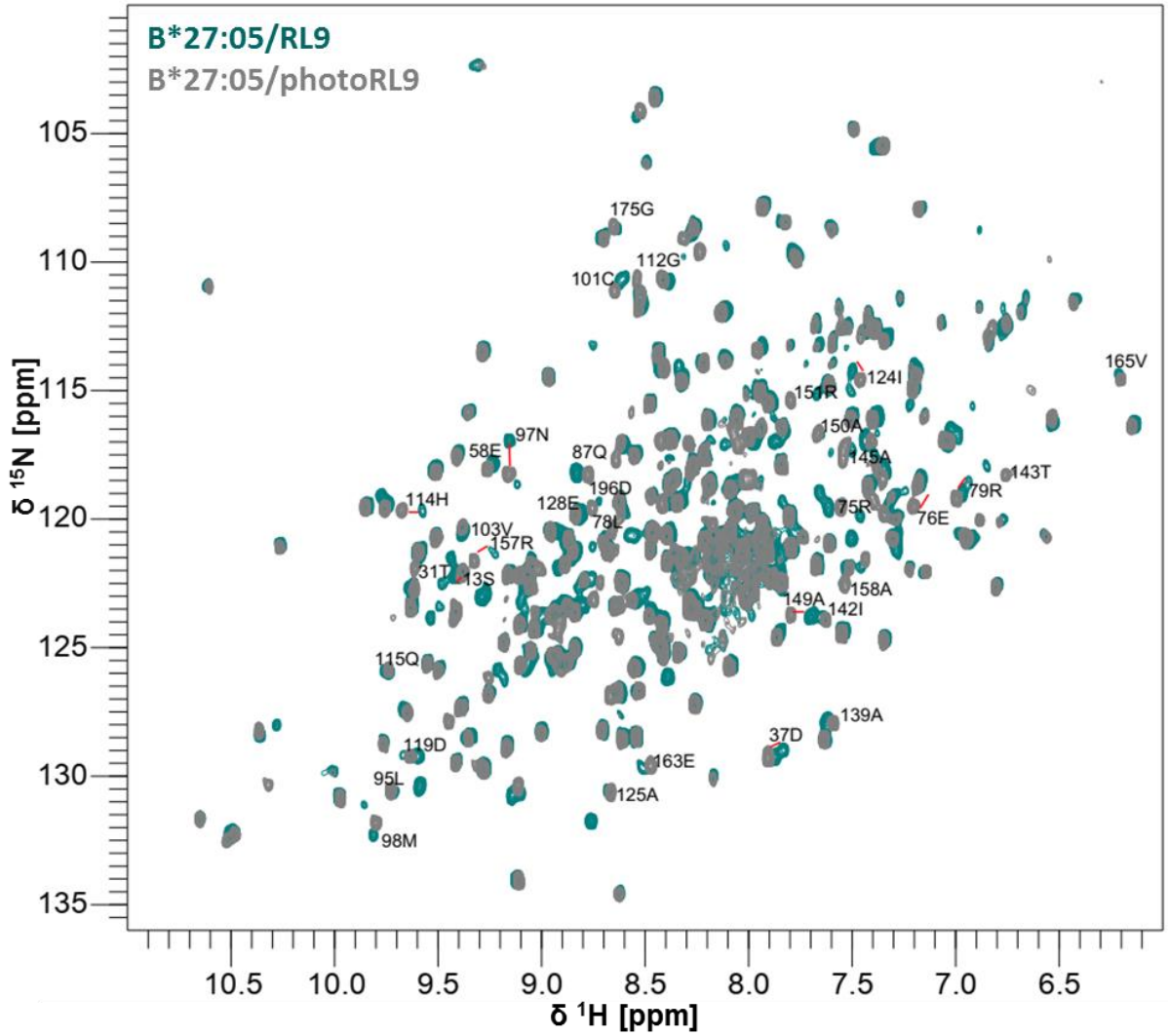
Appendix Table 4. The F-pocket residues of rat MHC-I (Rano) linked to TAP-A and TAP-B.

	Rat MHCI	74	77	80	81	95	97	99	114	116
TAP-A linked MHC-I	Rano-A*av1	Y	**D**	T	L	I	**E**	Y	R	**D**
	Rano-A*f	Y	**D**	T	L	I	**E**	Y	R	**D**
	Rano-A*g	N	**D**	T	L	I	V	F	R	**D**
	Rano-A1*o	Y	**D**	T	L	I	**E**	Y	R	**D**
	Rano-A2*f	Y	**D**	T	L	I	V	Y	R	**D**
	Rano-A2*o	Y	**D**	T	L	I	**E**	Y	R	**D**
	Rano-A*l	Y	S	N	L	I	R	Y	R	**D**
	Rano-A1*f	Y	S	N	L	I	R	Y	R	**D**
	Rano-A1*k	Y	N	T	L	I	V	S	R	**D**
	Rano-A1*q	Y	N	T	L	L	L	Y	V	H
	Rano-A2*q	Y	N	T	A	I	R	Y	N	Y
TAP-B linked MHC-I	Rano-A1*b	N	D	T	L	I	E	F	E	H
	Rano-A*u	N	N	T	L	I	V	F	R	D
	Rano-A1*c	F	N	T	L	I	E	Y	E	F
	Rano-A1*h	F	D	T	A	F	L	F	M	Y
	Rano-A1*n	F	N	T	A	I	R	S	E	Y
	Rano-A2*b	N	D	T	L	I	R	H	E	H
	Rano-A2*c	Y	S	T	A	I	R	Y	E	Y
	Rano-A2*h	Y	S	N	L	I	L	Y	W	F
	Rano-A2*n	Y	D	T	A	I	R	Y	Y	D
	Rano-A3*n	F	N	T	L	V	R	S	N	H

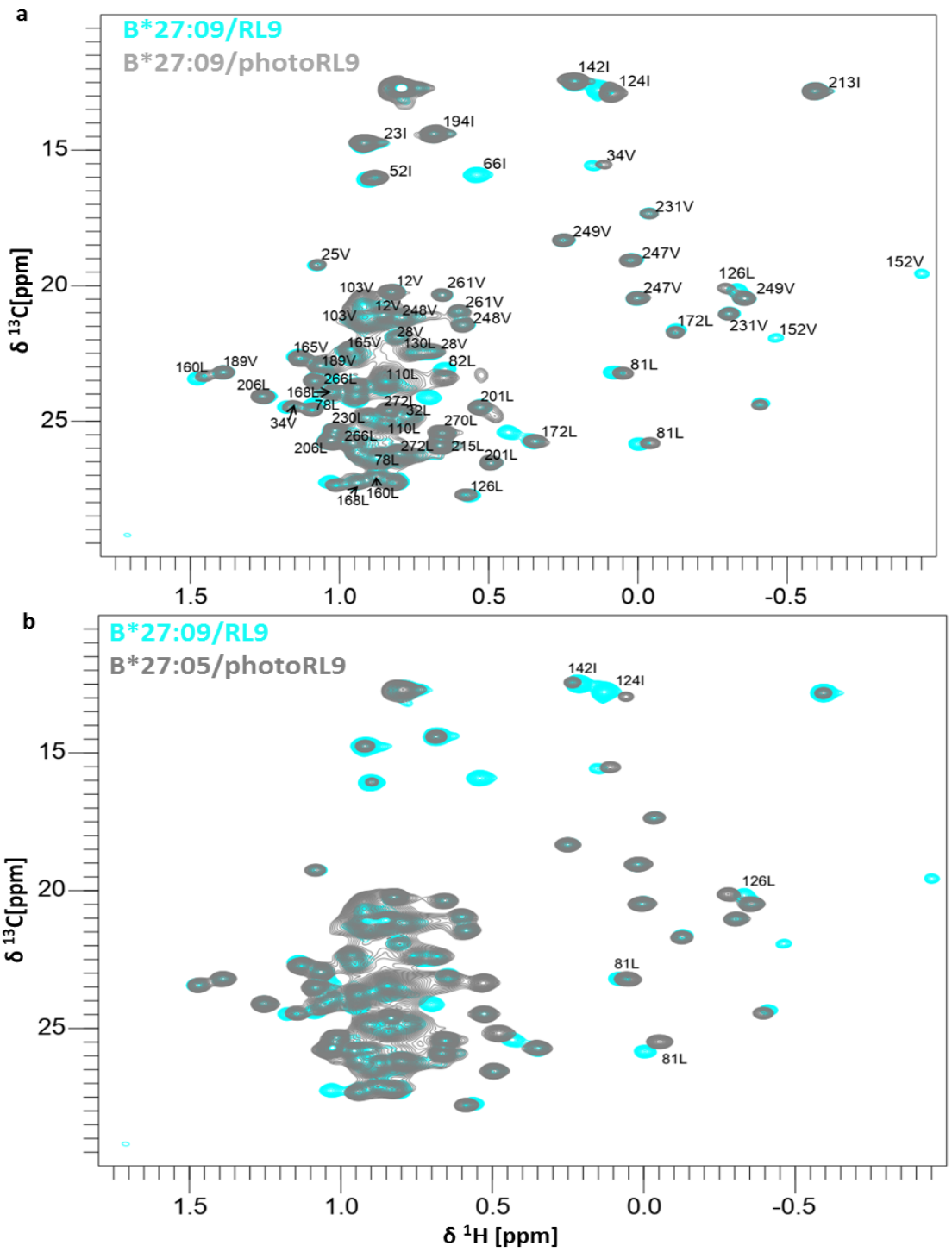
MHC-I bearing an acidic F-pocket was colored in pink background. f,g,o,l,k,q are linked to TAP-A, b,u,c,h,n are linked to TAP-B[249].



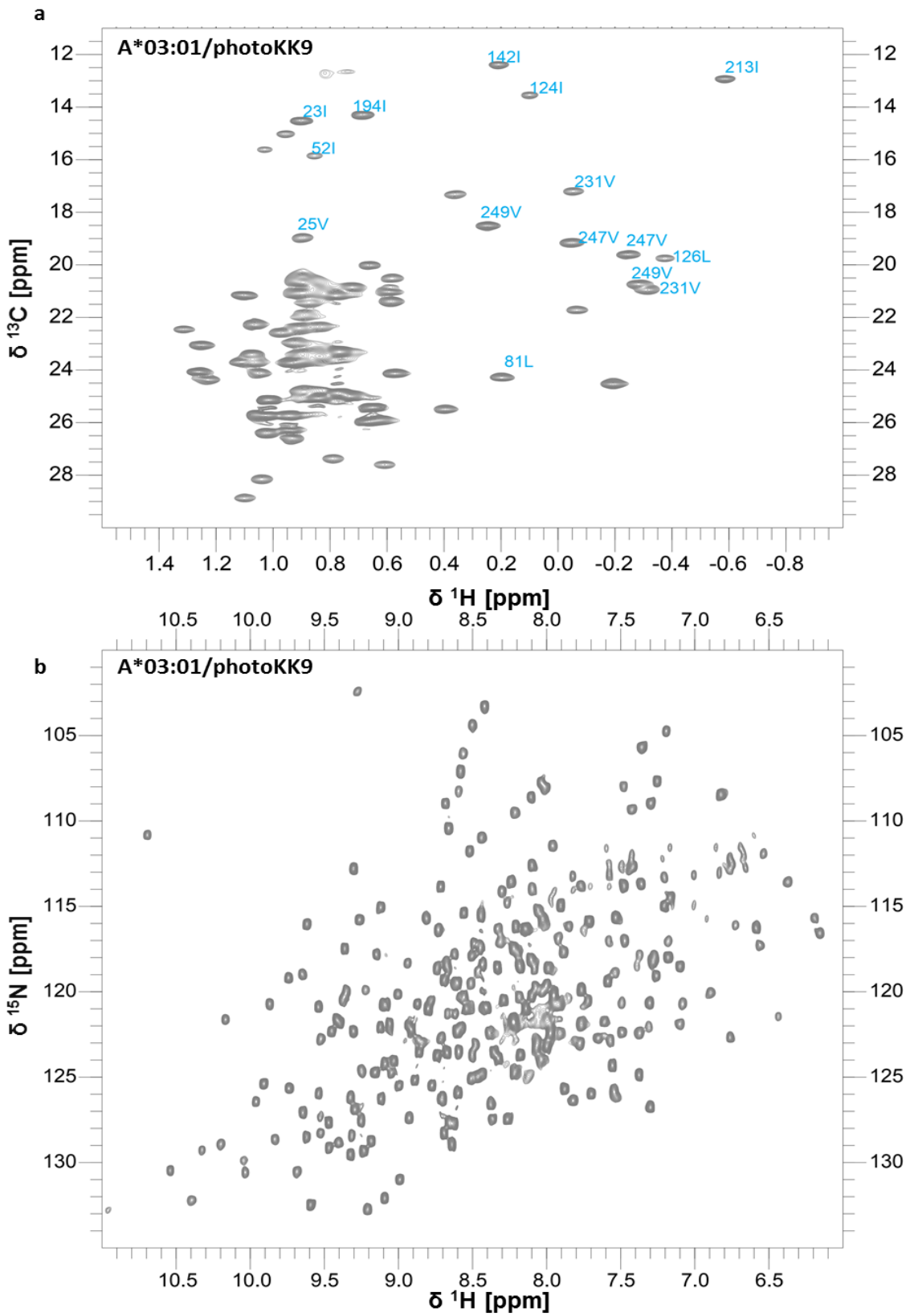
Appendix Figure 1. Overlay of ^{1}H-^{15}N TROSY-HSQC spectra of ^{15}N-^{13}C-^{2}H B*27:09/RL9 (280 μM, cyan) and ^{15}N-^{13}C-^{2}H B*27:09/photoRL9 (220 μM, gray) in PBS pH = 7.2 buffer measured on Bruker 700 MHz at 300 K. Residues showing chemical shift difference between the two spectra are marketed. Resonances of 7T, 8M, 68K, 69A, 70L, 72Q, 73T, 77D, 84Y, 98M, 110L, 114H, 115Q, 116H, 125A, 147W, 155Q, 160L are not observed in the spectrum of B*27:09/photoRL9.



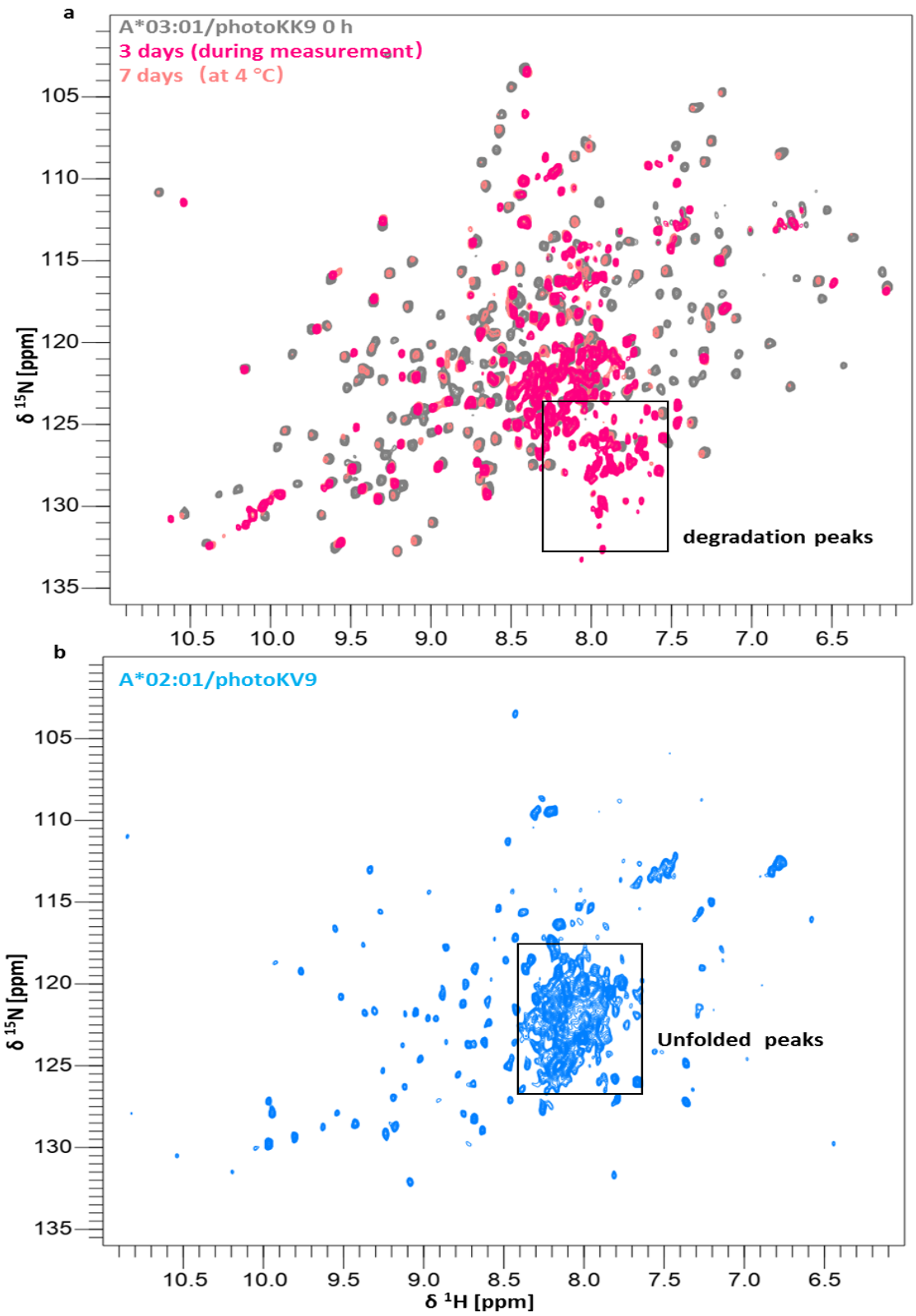
Appendix Figure 2. Overlay of ^{1}H-^{15}N TROSY-HSQC spectra of ^{15}N-^{13}C-^{2}H B*27:05/RL9 (250 µM, teal) and ^{15}N-^{13}C-^{2}H B*27:05/photoRL9 (220 µM, gray) in PBS 7.2 buffer measured on Bruker 700 MHz at 300K. Residues showing chemical shift difference between two spectra are labelled besides the peak.



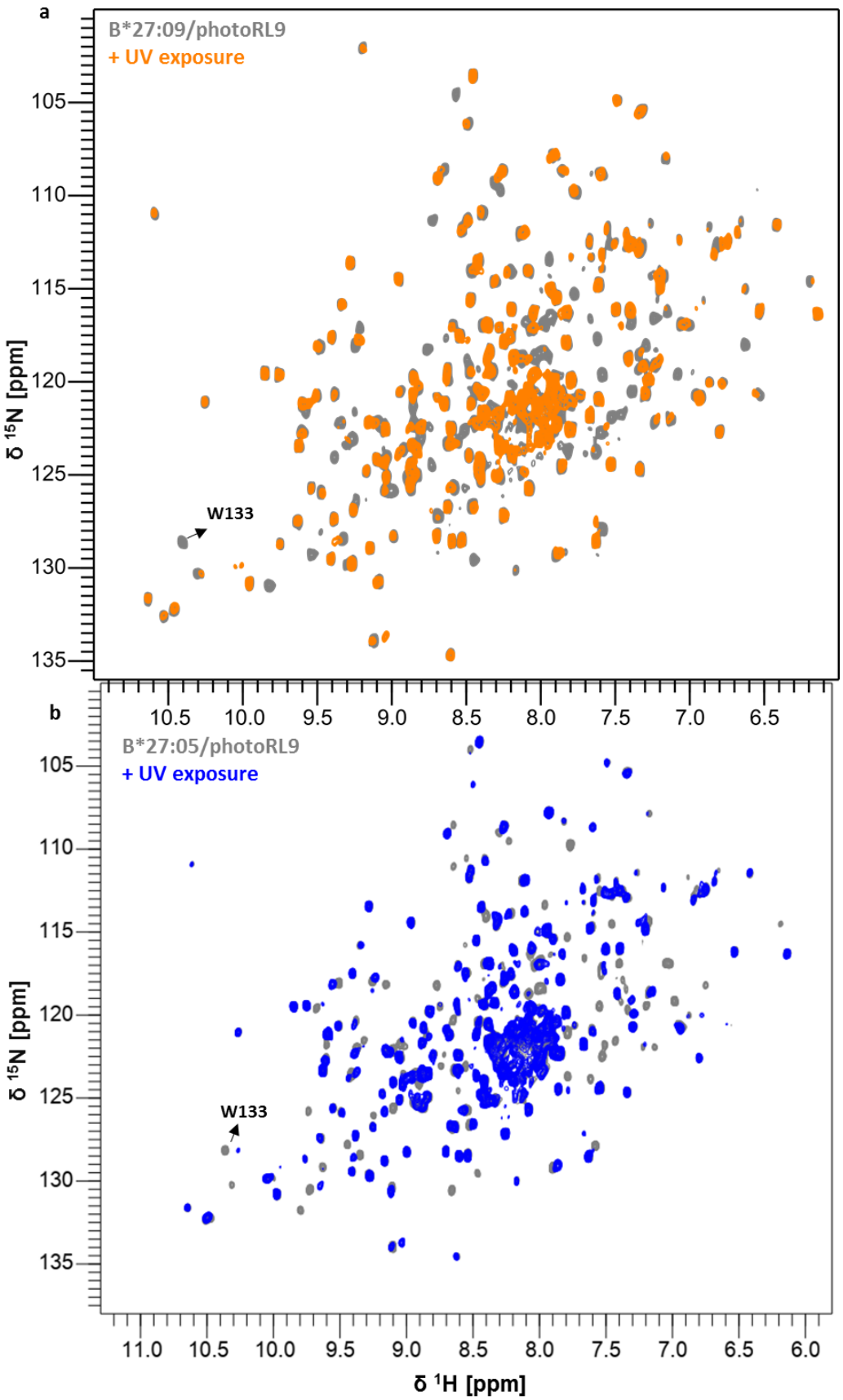
Appendix Figure 3. a) Overlay of ^{1}H-^{13}C HSQC spectra of ^{1}H-^{13}C-ILV-^{15}N-^{2}H B*27:09/RL9 (150 µM, cyan) and ^{1}H-^{13}C-ILV-^{15}N-^{2}H B*27:09/photoRL9 (170 µM, grey) with assignments, measured in PBS 7.2 buffer with 10% D_2O at 300K on a Bruker Avance III 700 MHz spectrometer. Resonances of 66I and 152V are not detected in the ^{1}H-^{13}C HSQC spectrum of B*27:09/photoRL9. **b)** Overlay of ^{1}H-^{13}C HSQC spectra of ^{1}H-^{13}C-ILV-^{15}N-^{2}H B*27:09/RL9 (150 µM, cyan) and ^{1}H-^{13}C-ILV-^{15}N-^{2}H B*27:05/photoRL9 (170 µM, grey) with assignments (only residues showing different chemical shift from B*27:09/photoRL9 are labelled), measured in PBS 7.2 buffer with 10% D_2O at 300K on a Bruker Avance III 700 MHz spectrometer. Resonances of 66I and 152V are not detected in the ^{1}H-^{13}C HSQC spectrum of B*27:05/photoRL9.



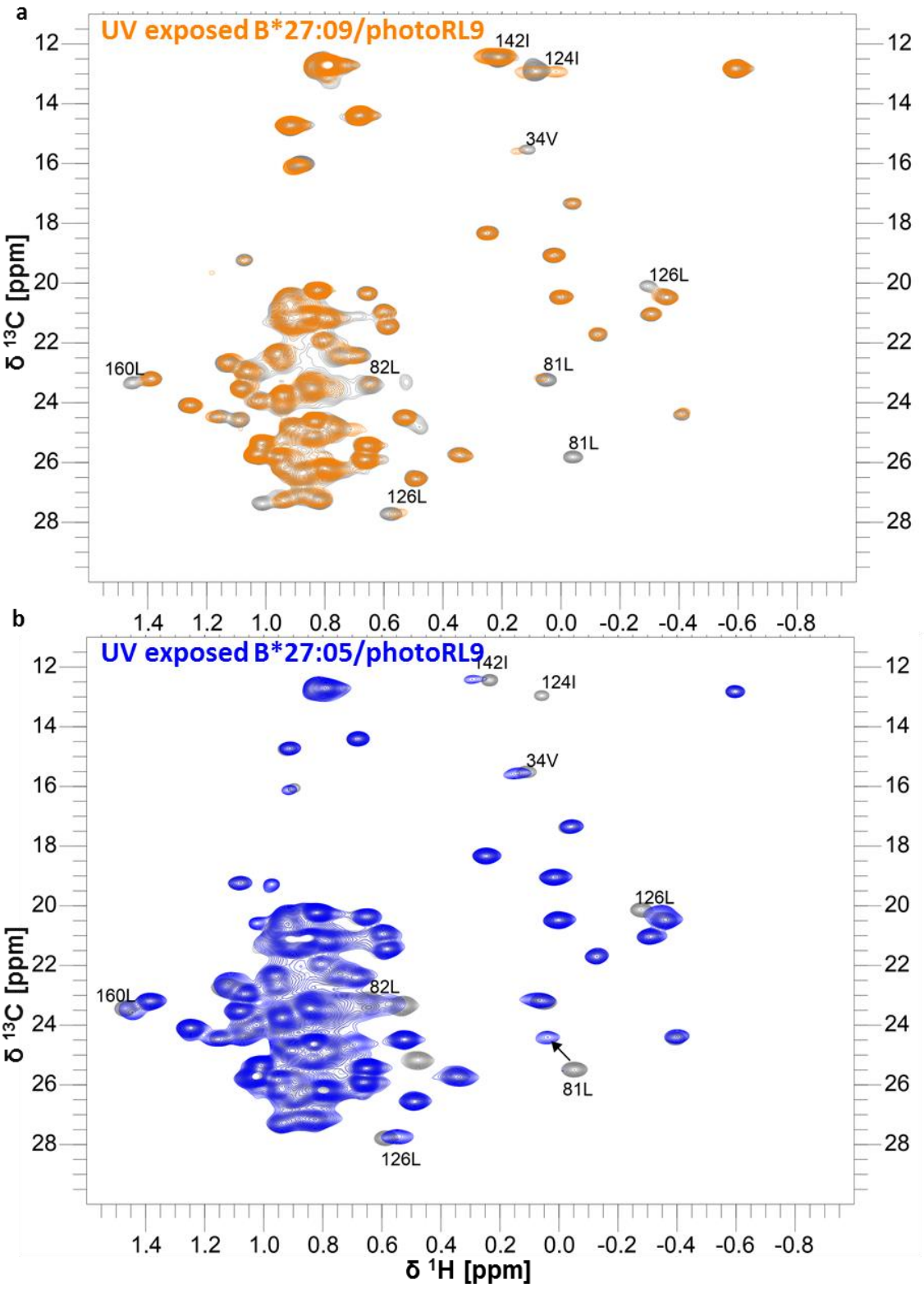
Appendix Figure 4. a) ^{1}H-^{13}C HSQC spectrum of ^{1}H-^{13}C-ILV-^{15}N-^{2}H A*03:01/photoKK9 (170 µM, gray) with available assignments, measured in PBS 7.2 buffer with 10% D_2O at 300K on a Bruker Avance III 700 MHz spectrometer. **b)** ^{1}H-^{15}N TROSY-HSQC spectra of ^{1}H-^{13}C-ILV-^{15}N-^{2}H A*03:01/photoKK9 (170 µM, gray) measured in PBS 7.2 buffer with 10% D_2O at 300K on a Bruker Avance III 700 MHz spectrometer.



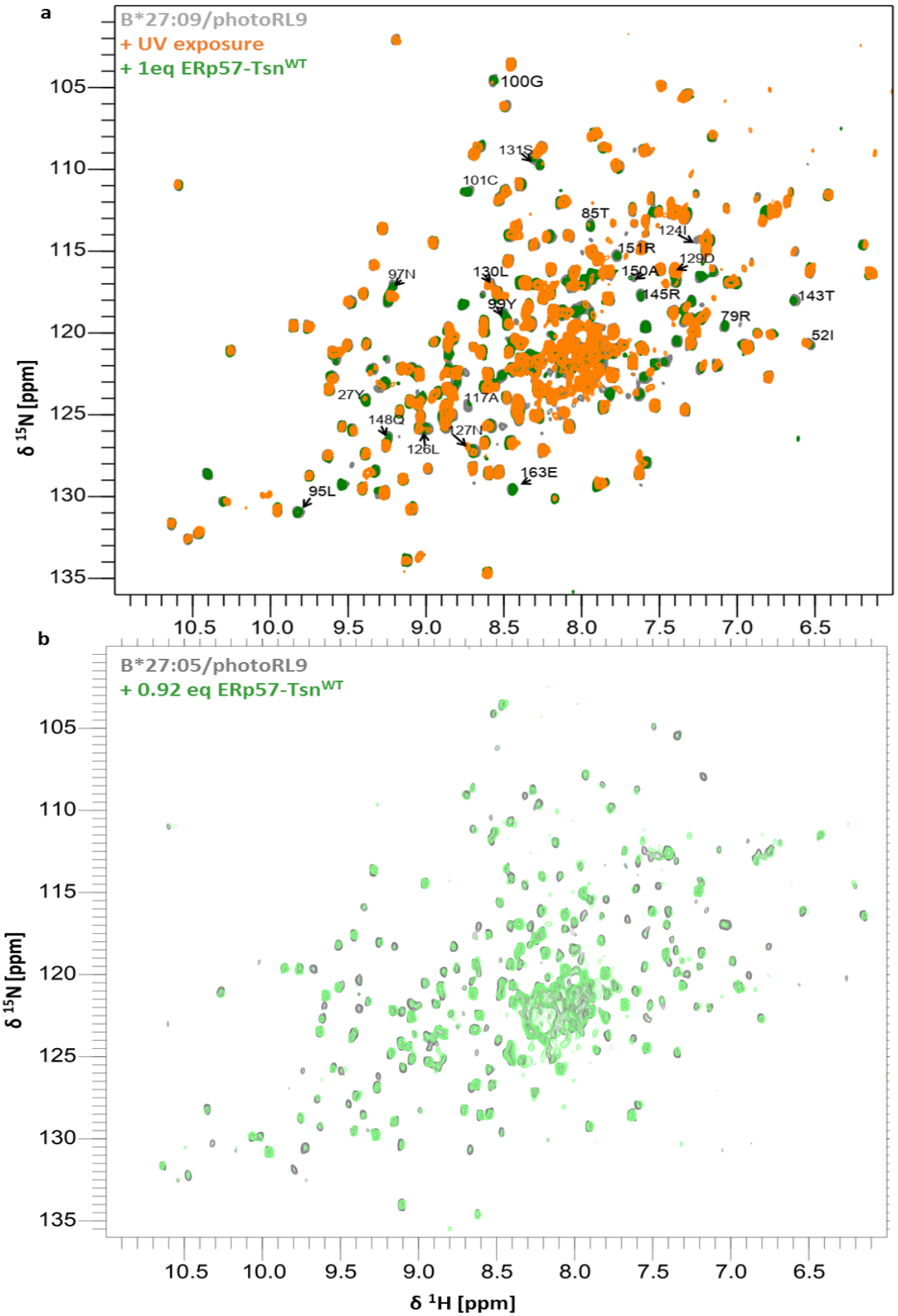
Appendix Figure 5. a) Overlay of ^{1}H-^{15}N TROSY-HSQC spectra of ^{1}H-^{13}C-ILV-^{15}N-^{2}H A*03:01/photoKK9 (70 µM) at different time point. Gray: fresh sample. Magenta: sampled measured for 3 days in the spectrometry at 300K. Light pink: sample measured after 7 days keeping at 4 °C. Spectra were measured in PBS 7.2 buffer with 10% D_2O at 300K on a Bruker Avance III 700 MHz spectrometer. Peaks in the black box is to indicate the degradation peaks. **b)** ^{1}H-^{15}N TROSY-HSQC spectra of ^{1}H-^{13}C-ILV-^{15}N-^{2}H A*02:01/photoKV9 (120 µM, sky blue) showing the refolding of A*02:01/photoKV9 failed.



Appendix Figure 6. a) Overlay of ^{1}H-^{15}N TROSY-HSQC spectra of ^{1}H-^{13}C-ILV-^{15}N-^{2}H B*27:09/photoRL9 before (gray) and after UV-exposure (orange). **b)** Overlay of ^{1}H-^{15}N TROSY-HSQC spectra of ^{1}H-^{13}C-ILV-^{15}N-^{2}H B*27:05/photoRL9 before (gray) and after UV-exposure (blue).

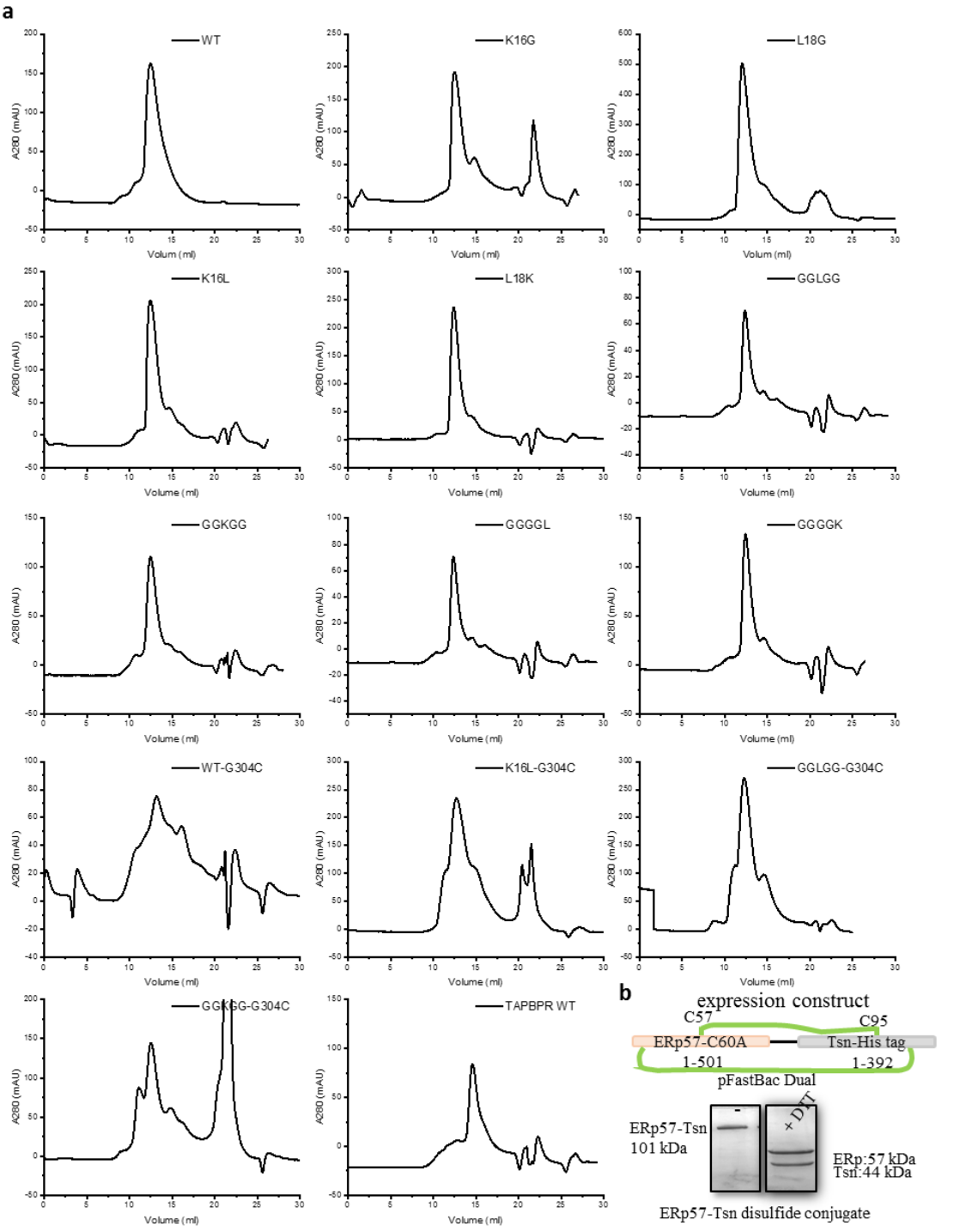


Appendix Figure 7. a) Overlay of ^{1}H-^{13}C HSQC spectra of ^{1}H-^{13}C-ILV-^{15}N-^{2}H B*27:09/photoRL9 before (gray) and after UV-exposure (orange). **b)** Overlay of ^{1}H-^{13}C HSQC spectra of ^{1}H-^{13}C-ILV-^{15}N-^{2}H B*27:05/photoRL9 before (gray) and after UV-exposure (blue). Significantly changed peaks are noted.



Appendix Figure 8. a) Overlay of ^{1}H-^{15}N TROSY-HSQC spectra of ^{1}H-^{13}C-ILV-^{15}N-^{2}H B*27:09/photoRL9 before (grey) and after UV exposure (orange) and in the presence of equimolar ERp57-Tsn^{WT} (forest). **b)** Overlay of -^{15}N TROSY-HSQC spectra of ^{1}H-^{13}C-ILV-^{15}N-^{2}H B*27:05/photoRL9 alone (gray) and in the presence of equimolar ERp57-Tsn^{WT} (green). Significantly changed peaks are noted.

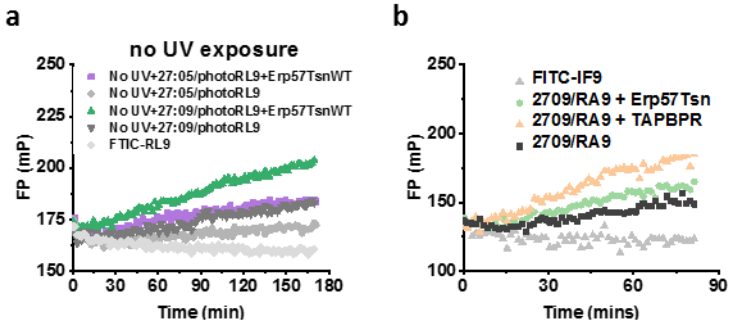
Appendix Figure 9



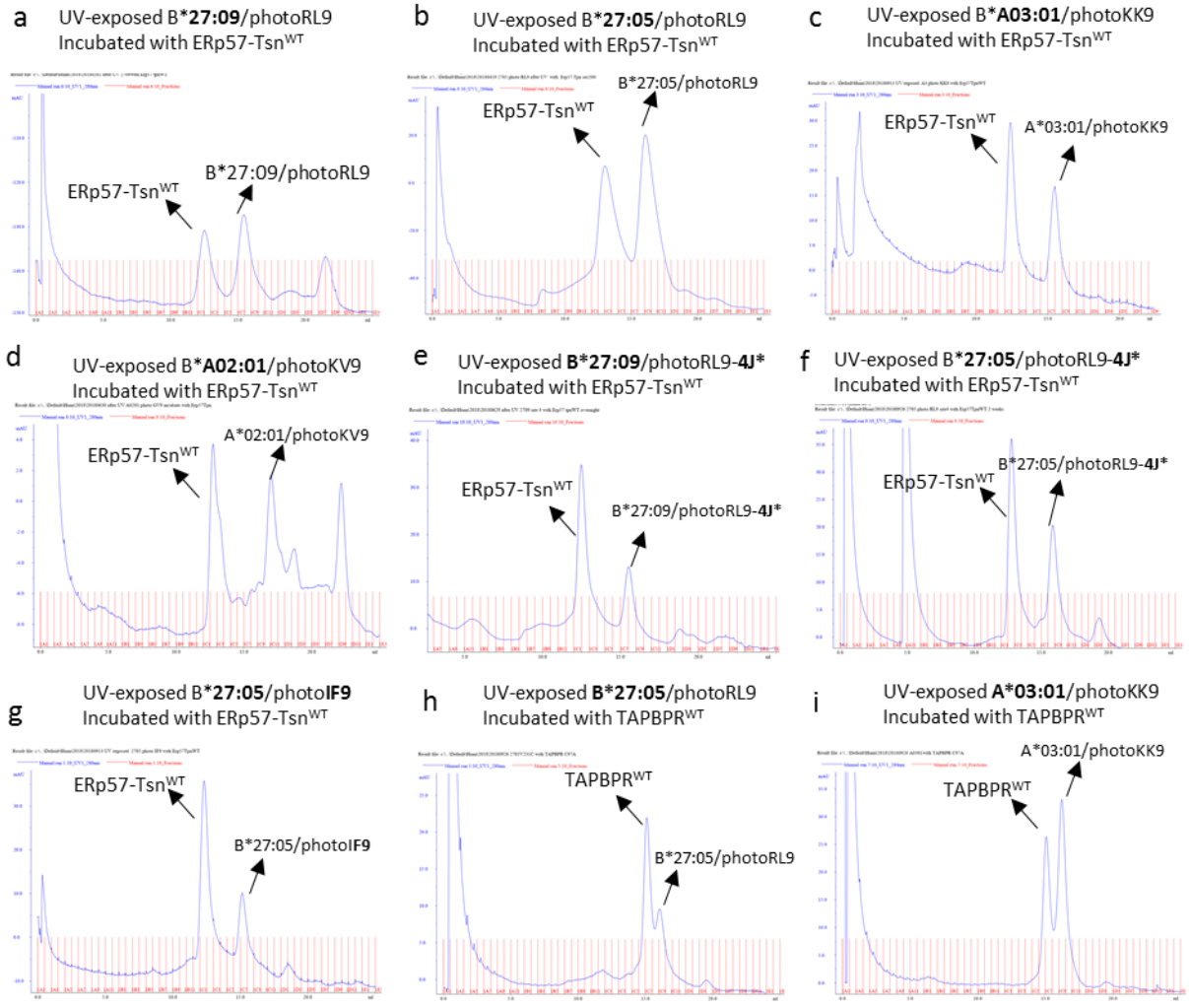
e

mutants	T_m
WT	49.33 ± 0.02
K16G	49.60 ± 0.03
L18G	50.85 ± 0.01
K16L	49.64 ± 0.02
L18K	49.36 ± 0.03
GGLGG	50.80 ± 0.02
GGKGG	48.85 ± 0.03
GGGGL	49.70 ± 0.03
GGGGK	49.52 ± 0.03
TAPBPR	52.69 ± 0.03

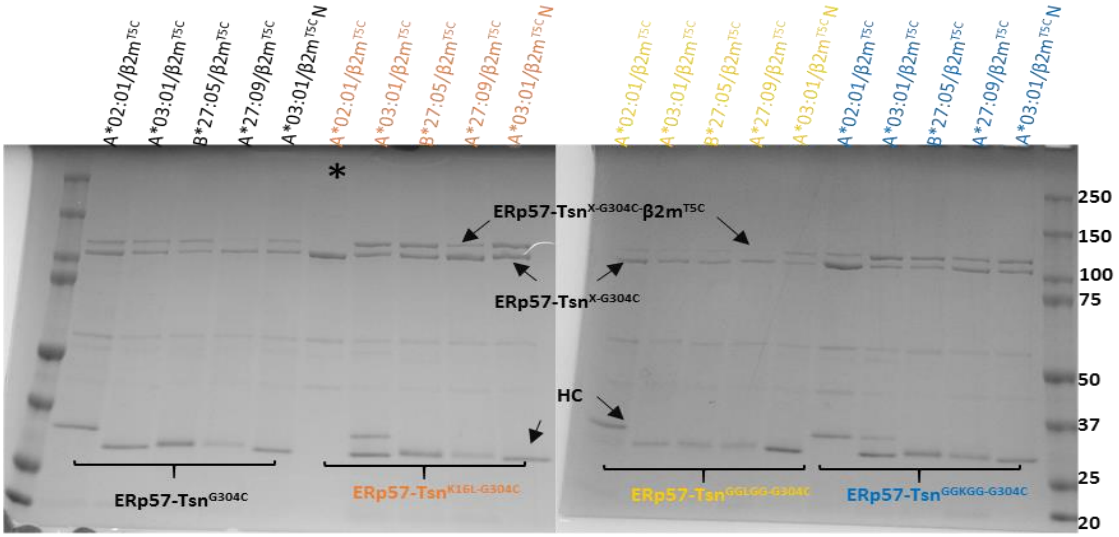
Appendix Figure 9. Characterization of ERp57-Tsn mutants and TAPBPR. a) Chromatography of ERp57-Tsn and TAPBPR mutants run in Superdex200 increase 10/300 GL. **b)** The expression construct and the validation of the ERp57-Tsn conjugate expression. **c)** Non-reducing SDS-PAGE analysis of purified tapasin mutants and TAPBPR. **d)**The melting curves of tapasin mutants. **e)** The melting temperature T_m values are listed in the table.



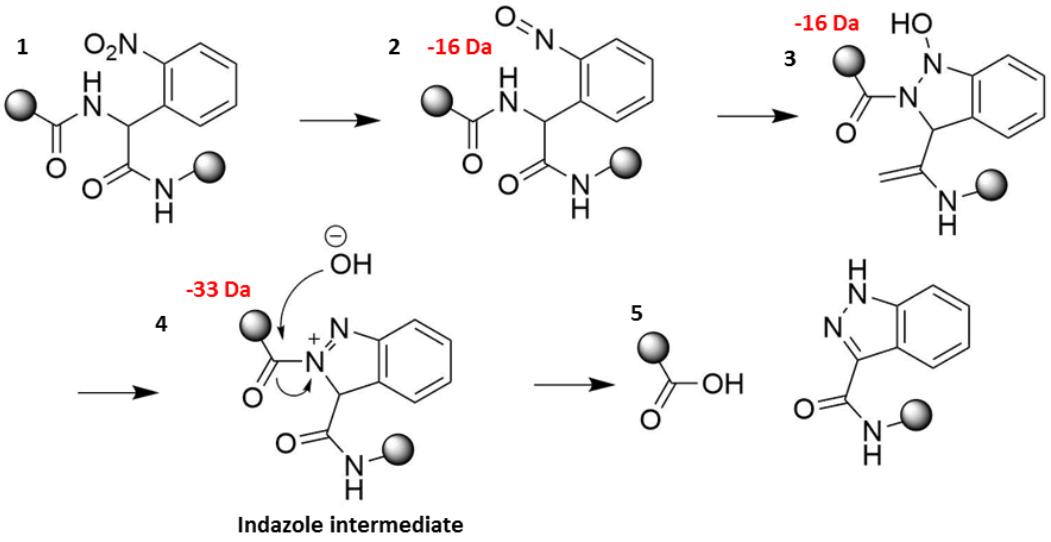
Appendix Figure 10. a) Peptide exchange profiles of 500 nM pMHC-I of B*27:09/photoRL9 and B*27:05/photoRL9 in the absence and in the presence of equimolar of ERp57-Tsn^{WT} without UV-exposure. **b)** Peptide exchange profiles of 500 nM pMHC-I of B*27:09/RA9 in the absence and in the presence of equimolar of ERp57-Tsn^{WT} and TAPBPR.



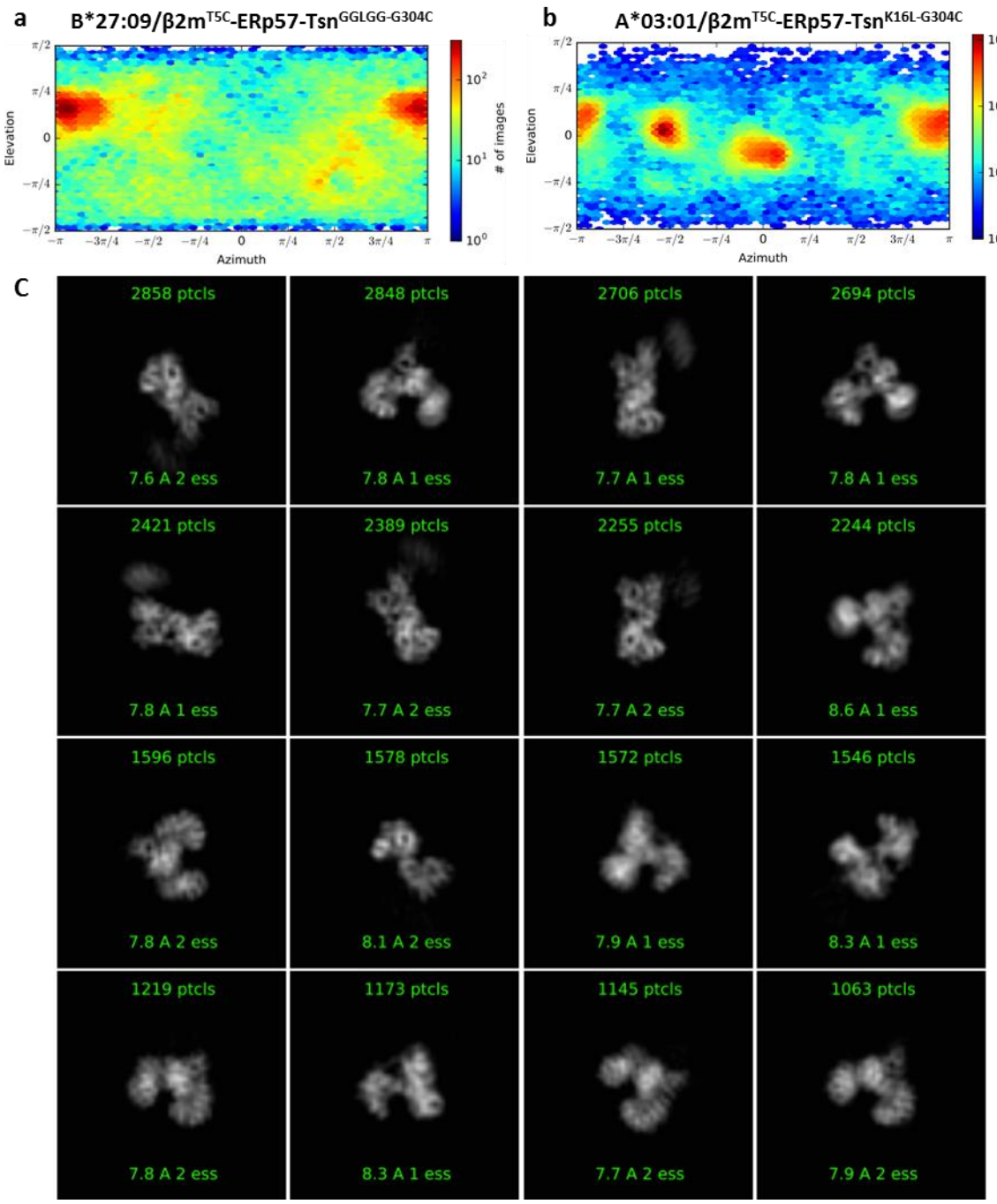
Appendix Figure 11. Resemble of MHC-I-ERp57-Tsn^{WT} complex and MHC-I-TAPBPR^{WT} complex. a-d) and g) Chromatography of UV-exposed photo-MHC-I incubated with ERp57-Tsn^{WT}. **e-f)** The photo cleavable site is introduced at the 4th residue of the peptide. Still no complex formed. **h-i)** Chromatography of UV-exposed photo-MHC-I incubated with TAPBPR^{WT}. no complex formed as expected. **a, b** was run on superdex200 10/300 GL column, **c-i** was run on Superdex200 increase 10/300 GL.



Appendix Figure 12. Complex formation of ERp57-Tsn$^{X\text{-}G304C}$ with all four allotypes refolded with β2m^{T5C}. UV-exposed MHC-I-β2m^{T5C} allotypes were incubated with corresponding C304C mutant and analyzed on non-reducing 4-12% gradient commercial SDS-PAGE gel in MOPs running buffer. β2m band ran out of the gel. Lane 2-7: ERp57-Tsn^{G304C} was incubated with A*02:01/β2m^{T5C}, A*03:01/β2m^{T5C}, B*27:09/β2m^{T5C} and B*27:05/β2m^{T5C}, respectively. Lane 8-12: ERp57-Tsn$^{K16L\text{-}G304C}$ (orange); Lane 1-5 on the right gel: ERp57-Tsn$^{GGLGG\text{-}G304C}$ (yellow); Lane 6-10 on the right gel: ERp57-Tsn$^{GGKGG\text{-}G304C}$ (blue). N stands for different batch of A*03:01/β2m^{T5C}/photoKK9. * means in this lane A*02:01/ β2m^{T5C}/photoKV9 was forgot to add by mistake.



Appendix Figure 13. The proposed mechanism of side reaction after UV-exposure. -34Da was due to the intramolecular cyclization products of reactive intermediates[250]. The resulting product is a peptide containing carboxyl group, not the amide group.



Appendix Figure 14. Cryo-EM analysis of the MHC-I-ERp57-Tsn complex. a) and b) Orientation distribution of two particles for two complexes on the cryo-EM grid. Particles of B*27:09/β2m^{T5C}-ERp57-Tsn$^{GGLGG-G304C}$ exhibited predominately views. The orientation distribution of A*03:01/β2m^{T5C}-ERp57-Tsn$^{K16L-G304C}$ particles improved. **c)** Selected 2D class averages of A*03:01/β2m^{T5C}-ERp57-Tsn$^{K16L-G304C}$.

8 Abbreviations

ABC	ATP-binding cassette
APCs	antigen presenting cells
AS	Ankylosing spondylitis
BCRs	B cell receptors
BiP	the immunoglobulin binding protein
Crt	calreticulin
CD	cluster of differentiation
CD	Crohn's Disease
CLIP	class-II associated invariant chain peptide
CPMG	Carr-Purcell-Meiboom-Gill
Cryo-EM	electron cryo-microscopy
DC	dendritic cell
ER	endoplasmic reticulum
ERAPs	ER aminopeptidases
ERP1/2	ER-resident animopeptidases
ERp57	thiol oxidoreductase Erp57
FITC	fluorescein isothiocyanate
FP	Fluorescence Polarization
GDN	glycol-diogenin
GVHD	graft-versus-host disease
HBV	Human hepatitis B virus
HC	Heavy chain
HCV	Human hepatitis C virus
HIV	human immunodeficiency virus
HLA	human leukocyte antigen
HPV	human papilloma virus
HSQC	Heteronuclear Single Quantum Correlation
Ii	invariant chain
INEPT	Insensitive Nucleus Enhanced by Polarization Transfer.
KIRs	immunoglobulin-like receptors
LD	linkage disequilibrium
LD	linkage disequilibrium
LMPs	low-molecular-weight proteins
MALDI-TOF-MS	matrix-assisted laser desorption ionization-time of flight mass spectrometry
MHC	Major Histocompatibility Complex
MHC-I	MHC class I
MHC-II	MHC class II
MMTS	S-methyl mechanethiosulfonate
MS	Multiple Sclerosis
NBD	nucleotide-binding domains
NKC	natural killer cell
NMR	Nuclear Magnetic Resonance
NOE	Nuclear Overhauser Effect
PAMP	pathogen-associated molecular patterns

PAMPs	pathogen-associated molecular patterns
PDI	protein disulfide isomerase
PLC	peptide loading complex
pMHC-I	peptide-bound MHC class I
pMHC-II	Peptide-bound MHC class II
PRRs	pattern recognition receptors
pVIPR	RRKWRRWHL
pVIPR	vasoactive intestinal peptide type 1 receptor peptide
PΩ	The last amino acid of the bound peptide in the MHC-I
SEC	Size Exclusion Chromatography
SLE	Systematic Lupus Erythematosus
T1D	Type 1 Diabetes
TAP	Transporter associated with Antigen Processing
Tapasin	TAP-associated Protein
TAPBPR	tapasin-related protein
TCR	T cell receptor
Tm	melting temperature
TMD	transmembrane domains
TPPII	tripeptidyl peptidase II
TROSY	Three-dimensional transverse relaxation optimized spectroscopy
Tsn	tapasin
UC	Ulcerative Colitis
UGT1	UDP-glucose:glycoprotein glucosyltransferase 1
Δδ	Chemical shift difference
ω0	the reference frequency

9 Acknowledgment

I would like to thank my supervisor **Prof. Christian Freund** to offer me the chance to work on such interesting topic. Almost five year ago, he offered me the chance to work in his lab, which made everything possible today. Over the past four years, he provided me with all kinds of guides to help me with the PhD study. His scientific insights into the project and hard-working motivated me to become a good researcher. He not only cares about the project progress, but also care about the individual development of us as PhD student. I am so grateful and also feel so lucky to have a such nice supervisor.

During the past four years, I really enjoyed working with so many wonderful scientists as well as colleagues who helped me with their experiences and knowledges. Special thanks to **Dr. Esam T. Abualrous** for the guide during the whole project, form the setup of the project to almost every experimental discussion. No matter what kind of questions or problems I have, he could always give me useful advices. He also performed the computational work of modelling tapasin scoop loop. Special thanks to **Dr. Jana Sticht** for the help with NMR data acquisition and analysis. She is a good teacher and always explained well all these complex concepts regarding NMR measurements. Thank **Dr. Bernhard Loll** for resolving the crystal structure of B*27:09 and **Dr. Tarek Hilal** for the cryo-EM sample preparation, data acquisition and structure analysis of the tapasin complex. Thank Prof. Peter Schmieder and his student Dr. Martin Ballaschk for providing NMR assignments and the pulse sequence for the CPMG measurements. Thank **Dr. Chris Weise** and **Dr. Benno Kuropka** for taking care of MS samples and nice discussions. Thank Dr. Miguel Alvaro-Benito for questions regarding cell culture. I also owe thanks to my students Laura Arroyo Fernandez and Tamina Werk, who contributed to this work in protein expression and purification.

Many thanks are to be owed to all other present and past members of the Freund lab for great scientific collaboration and the amazing atmosphere in the lab: Dr. Fabian Gerth, Dr. Eliot Morrison, Tatjana Wegner, Nirdosh Dadwal, Dr. Miriam Bertazzon, Steffen Dähn, Xiao Jakob Schmitt, Dr. Almudena Hurtado Pico, Frank Kuppler, Jan Licha. Thank **Dr. Marek Wieczorek**, who supervised me at the beginning when I joint the lab. Thank Tatjana and Nirdosh, we three always cheer each other up when met problems in experiments, which made the PhD life colorful. Thank Steffen and Almudena who could always order stuff in time and organized the lab well.

Finally, I would like to thank my family, my country and Germany for providing mental support, financial support and scientific working environment, respectively.

10 Curriculum Vitae

For reasons of data protection, the curriculum vitae is not published in the electronic version.

11 Own publications

Publications related with the thesis

1. **Huan Lan***, Esam T. Abualrous, Jana Sticht, Laura Maria Arroyo Fernandez, Tamina Werk, Christoph Weise, Martin Ballaschk, Peter Schmieder, Bernhard Loll & Christian Freund. Exchange catalysis by tapasin exploits conserved and allele-specific features of MHC-I molecules. Manuscript has been submitted into *Nat.Commun.*, 2021
2. Manuscript "Spatial and temporal plasticity of neoantigen-specific T-cell responses bases on characteristics associated to antigen and TCR" has been submitted to Clinical Cancer Research, 2019

Publications not related with the thesis

1. Yi-Tong Li*, Cong Yi, Chen-Chen Chen, **Huan Lan**…Lei Liu *et al*, A semisynthetic Atg3 reveals that acetylation promotes Atg3 membrane binding and Atg8 lipidation. *Nat. Commun.* 2017, 8, 14846

2. **Huan Lan*** et al, Total synthesis of mambalgin-1/2/3 by two-segment hydrazide-based native chemical ligation. *J. Pept. Sci.* 2016, 22: 320–326

3. Man Pan[#], Shuai Gao[#], Yong Zheng[#], Xiaodan Tan, **Huan Lan**… Lei Liu *et al*, Quasi-racemic X-ray structures of K27-Linked Ubiquitin chains prepared by total chemical synthesis, J. Am. Chem. Soc. 2016, 138, 7429−7435

4. Man Pan[#], Qingyun Zheng[#], Shuai Gao, Qian Qu, Yuanyuan Yu, Ming Wu, **Huan Lan**…Changlin Tian & Lei Liu *et al*, Chemical Synthesis of Structurally Defined Phosphorylated Ubiquitins Suggests Impaired Parkin Activation by Phosphorylated Ubiquitins with a Non-Phosphorylated Distal Unit, *CCS Chem*. 2019, 1, 476–489

*first author, [#] equal contribution

12 Bibliography

1. Cooper, M. D. & Alder, M. N. The evolution of adaptive immune systems. *Cell* **124**, 815–822 (2006).

2. Pasquier, L. D. Origin and evolution of the vertebrate immune system. *APMIS* **100**, 159–185 (2000).

3. Iwasaki, A. & Medzhitov, R. Regulation of adaptive immunity by the innate immune system. *Science* **327**, 291–295 (2010).

4. Beyaert*, J. S. and & Beyaert*, R. Role of Toll-like receptors in pathogen recognition. *Clin. Microbiol. Rev* **16**, 637–646 (2003).

5. Medzhitov, R. & Janeway, C. Jr. Innate immune recognition: mechanisms and pathways. *Immunol Rev* **173**, 89–97 (2000).

6. Vivier, E., Tomasello, E., Baratin, M., Walzer, T. & Ugolini, S. Functions of natural killer cells. *Nat Immunol* **9**, 503–510 (2008).

7. Horton, R. *et al.* Gene map of the extended human MHC. *Nat Rev Genet* **5**, 889–899 (2004).

8. Trowsdale, J. Genomic structure and function in the MHC. *Trends in Genetics* **9**, 117–122 (1993).

9. The MHC sequencing consortium. Complete sequence and gene map of a human major histocompatibility complex. *Nature* **401**, 921–923 (1999).

10. Norman, P. J. *et al.* Sequences of 95 human MHC haplotypes reveal extreme coding variation in genes other than highly polymorphic HLA class I and II. *Genome Res* **27**, 813–823 (2017).

11. Robinson, J. *et al.* IMGT/HLA and IMGT/MHC: sequence databases for the study of the major histocompatibility complex. *Nucleic Acids Res* **31**, 311–314 (2003).

12. Foroni, I. *et al.* HLA and associated important diseases: HLA-E, HLA-F and HLA-G - the non-classical side of the MHC cluster. (2014) doi:10.5772/57507.

13. Stenzel, A. *et al.* Patterns of linkage disequilibrium in the MHC region on human chromosome 6p. *Hum Genet* **114**, 377–385 (2004).

14. Parham, P. & Ohta, T. Population biology of antigen presentation by MHC class I molecules. *Science* **272**, 67–74 (1996).

15. Spurgin, L. G. & Richardson, D. S. How pathogens drive genetic diversity: MHC, mechanisms and misunderstandings. *Proc Royal Soc B Biological Sci* **277**, 979–988 (2010).

16. Kanda, J. Effect of HLA mismatch on acute graft-versus-host disease. *Int J Hematol* **98**, 300–308 (2013).

17. Petersdorf, E. W. Role of major histocompatibility complex variation in graft-versus-host disease after hematopoietic cell transplantation. *F1000research* **6**, 617 (2017).

18. GOODFELLOW, P. N. *et al.* The β2-microglobulin gene is on chromosome 15 and not in the HL-A region. *Nature* **254**, 267–269 (1975).

19. Bjorkman, P. J. *et al.* Structure of the human class I histocompatibility antigen, HLA-A2. *Nature* **329**, 506–512 (1987).

20. Madden, D. R. The three-dimensional structure of peptide-MHC complexes. *Annu Rev Immunol* **13**, 587–622 (1995).

21. Chelvanayagam, G., Apostolopoulos, V. & McKenzie, I. F. Milestones in the molecular structure of the major histocompatibility complex. *Protein Eng Des Sel* **10**, 471–474 (1997).

22. Collins, E. J., Garboczi, D. N. & Wiley, D. C. Three-dimensional structure of a peptide extending from one end of a class I MHC binding site. *Nature* **371**, 626–629 (1994).

23. Persson, K. & Schneider, G. Three-dimensional structures of MHC class I-peptide complexes: implications for peptide recognition. *Arch. Immunol. Ther. Exp.* **48**, 135–142 (2000).

24. Wilson, I. A. & Fremont, D. H. Structural analysis of MHC class I molecules with bound peptide antigens. *Seminars in Immunology* **5**, 75–80 (1993).

25. Falk, K., Rötzschke, O., Stevanović, S., Jung, G. & Rammensee, H.-G. Allele-specific motifs revealed by sequencing of self-peptides eluted from MHC molecules. *Nature* **351**, 290–296 (1991).

26. Hülsmeyer, M. *et al.* Dual, HLA-B27 Subtype-dependent Conformation of a Self-peptide. *JEM* **199**, 271–281 (2004).

27. Matsumura, M., Fremont, D., Peterson, P. & Wilson, I. Emerging principles for the recognition of peptide antigens by MHC class I molecules. *Science* **257**, 927–934 (1992).

28. Garrett, T. P. J., Saper, M. A., Bjorkman, P. J., Strominger, J. L. & Wiley, D. C. Specificity pockets for the side chains of peptide antigens in HLA-Aw68. *Nature* **342**, 692–696 (1989).

29. Huyton, T., Ladas, N., Schumacher, H., Blasczyk, R. & Bade-Doeding, C. Pocketcheck: updating the HLA class I peptide specificity roadmap. *Tissue Antigens* **80**, 239–248 (2012).

30. Abualrous, E. T. *et al.* F pocket flexibility influences the tapasin dependence of two differentially disease-associated MHC Class I proteins. *Eur. J. Immunol* **45**, 1248–1257 (2015).

31. Abualrous, E. *et al.* The Carboxy Terminus of the Ligand Peptide Determines the Stability of the MHC Class I Molecule H-2Kb: A Combined Molecular Dynamics and Experimental Study. *PLOS ONE* **10**, e0135421 (2015).

32. Rammensee, H.-G., Friede, T. & Stevanović, S. MHC ligands and peptide motifs: first listing. *Immunogenetics* **41**, 178–228 (1995).

33. Rammensee, H.-G., Bachmann, J., Emmerich, N. P. N., Bachor, O. A. & Stevanović, S. SYFPEITHI: database for MHC ligands and peptide motifs. *Immunogenetics* **50**, 213–219 (1999).

34. Zemmour, J. & Parham, P. Distinctive polymorphism at the HLA-C locus: implications for the expression of HLA-C. *J Exp Medicine* **176**, 937–950 (1992).

35. Parham, P. MHC class I molecules and kirs in human history, health and survival. *Nat Rev Immunol* **5**, 201–214 (2005).

36. Leone, P. *et al.* MHC Class I Antigen Processing and Presenting Machinery: Organization, Function, and Defects in Tumor Cells. *JNCI* **105**, 1172–1187 (2013).

37. Papúchová, H., Meissner, T. B., Li, Q., Strominger, J. L. & Tilburgs, T. The Dual Role of HLA-C in Tolerance and Immunity at the Maternal-Fetal Interface. *Front Immunol* **10**, 2730 (2019).

38. Anfossi, N. *et al.* Human NK Cell Education by Inhibitory Receptors for MHC Class I. *Immunity* **25**, 331–342 (2006).

39. Hackmon, R. *et al.* Definitive class I human leukocyte antigen expression in gestational placentation: HLA-F, HLA-E, HLA-C, and HLA-G in extravillous trophoblast invasion on placentation, pregnancy, and parturition. *Am J Reprod Immunol* **77**, e12643 (2017).

40. Joosten, S. A., Sullivan, L. C. & Ottenhoff, T. H. M. Characteristics of HLA-E Restricted T-Cell Responses and Their Role in Infectious Diseases. *J Immunol Res* **2016**, 1–11 (2016).

41. Lepin, E. J. M. *et al.* Functional characterization of HLA-F and binding of HLA-F tetramers to ILT2 and ILT4 receptors. *European Journal of Immunology* **30**, 3552–3561 (2000).

42. Andrade, L. F. de *et al.* Antibody-mediated inhibition of MICA and MICB shedding promotes NK cell–driven tumor immunity. *Science* **359**, 1537–1542 (2018).

43. Ghadially, H. *et al.* MHC class I chain-related protein A and B (MICA and MICB) are predominantly expressed intracellularly in tumour and normal tissue. *Brit J Cancer* **116**, 1208–1217 (2017).

44. Jinushi, M. *et al.* Expression and role of MICA and MICB in human hepatocellular carcinomas and their regulation by retinoic acid. *Int J Cancer* **104**, 354–361 (2003).

45. Murthy, V. L. & Stern, L. J. The class II MHC protein HLA-DR1 in complex with an endogenous peptide: implications for the structural basis of the specificity of peptide binding. *Structure* **5**, 1385–1396 (1997).

46. Brown, J. H. *et al.* Three-dimensional structure of the human class II histocompatibility antigen HLA-DR1. *Nature* **364**, 33–39 (1993).

47. Fremont, D. H., Hendrickson, W. A., Marrack, P. & Kappler, J. Structures of an MHC Class II Molecule with Covalently Bound Single Peptides. *Science* **272**, 1001–1004 (1996).

48. Rudensky, A. Y., Preston-Hurlburt, P., Hong, S.-C., Barlow, A. & Janeway, C. A. Sequence analysis of peptides bound to MHC class II molecules. *Nature* **353**, 622–627 (1991).

49. Mullen, M. M., Haan, K. M., Longnecker, R. & Jardetzky, T. S. Structure of the Epstein-Barr Virus gp42 Protein Bound to the MHC Class II Receptor HLA-DR1. *Mol Cell* **9**, 375–385 (2002).

50. Chicz, R. M. *et al.* Specificity and promiscuity among naturally processed peptides bound to HLA-DR alleles. *J Exp Medicine* **178**, 27–47 (1993).

51. Mellins, E. D. & Stern, L. J. HLA-DM and HLA-DO, key regulators of MHC-II processing and presentation. *Current Opinion in Immunology* **26**, 115–122 (2014).

52. The impact ofthe non-classical MHC proteins HLA-DM and HLA-DO on loading of MHC class II molecules. (n.d.).

53. Denzin, L. K. & Cresswell, P. HLA-DM induces clip dissociation from MHC class II αβ dimers and facilitates peptide loading. *Cell* **82**, 155–165 (1995).

54. Guce, A. I. *et al.* HLA-DO acts as a substrate mimic to inhibit HLA-DM by a competitive mechanism. *Nat Struct Mol Biol* **20**, 90–98 (2013).

55. Freund, C. & Höfer, T. A Missing Switch in Peptide Exchange for MHC Class II Molecules. *Front Immunol* **10**, 2513 (2019).

56. Fernando, M. M. *et al.* Defining the Role of the MHC in Autoimmunity: A Review and Pooled Analysis. *PLoS Genet.* **4**, e1000024 (2008).

57. Kennedy, A. E., Ozbek, U. & Dorak, M. T. What has GWAS done for HLA and disease associations? *Int J Immunogenet.* **44**, 195–211 (2017).

58. Gómez, P., Montserrat, V., Marcilla, M., Paradela, A. & de Castro, J. A. L. B*2707 differs in peptide specificity from B*2705 and B*2704 as much as from HLA-B27 subtypes not associated to spondyloarthritis. *European Journal of Immunology* **36**, 1867–1881 (2006).

59. Khan, M. A. HLA-B27 and its subtypes in world populations Editorial review. *Current Opinion in Rheumatology* **7**, (1995).

60. Ramagopalan, S. V. & Ebers, G. C. Multiple sclerosis: major histocompatibility complexity and antigen presentation. *Genome Med* **1**, 105 (2009).

61. Erlich, H. *et al.* HLA DR-DQ Haplotypes and Genotypes and Type 1 Diabetes Risk: Analysis of the Type 1 Diabetes Genetics Consortium Families. *Diabetes* **57**, 1084–1092 (2008).

62. Niu, Z., Zhang, P. & Tong, Y. Value of HLA-DR genotype in systemic lupus erythematosus and lupus nephritis: a meta-analysis. *International Journal of Rheumatic Diseases* **18**, 17–28 (2015).

63. Newton, J., Harney, S., Wordsworth, B. & Brown, M. A review of the MHC genetics of rheumatoid arthritis. *Genes Immun* **5**, 151–157 (2004).

64. Goulder, P. J. R. & Watkins, D. I. Impact of MHC class I diversity on immune control of immunodeficiency virus replication. *Nat Rev Immunol* **8**, 619–630 (2008).

65. Tardif, K. D. & Siddiqui, A. Cell Surface Expression of Major Histocompatibility Complex Class I Molecules Is Reduced in Hepatitis C Virus Subgenomic Replicon-Expressing Cells. *J Virol* **77**, 11644–11650 (2003).

66. Thomas, R. *et al.* A Novel Variant Marking HLA-DP Expression Levels Predicts Recovery from Hepatitis B Virus Infection. *J Virol* **86**, 6979–6985 (2012).

67. Torres, M. *et al.* Role of Phagosomes and Major Histocompatibility Complex Class II (MHC-II) Compartment in MHC-II Antigen Processing of Mycobacterium tuberculosis in Human Macrophages. *Infect Immun.* **74**, 1621 (2006).

68. Gameiro, S. F. *et al.* Analysis of Class I Major Histocompatibility Complex Gene Transcription in Human Tumors Caused by Human Papillomavirus Infection. *Viruses* **9**, 252 (2017).

69. York, I. A., Goldberg, A. L., Mo, X. & Rock, K. L. Proteolysis and class I major histocompatibility complex antigen presentation. *Immunological Reviews* **172**, 49–66 (1999).

70. Kloetzel, P.-M. The proteasome and MHC class I antigen processing. *Biochim Biophys Acta* **1695**, 225–233 (2004).

71. Rock, K. L. & Goldberg, A. L. Degradation of cell proteins and the generation of the MHC class I-presented peptides. *Annu Rev Immunol* **17**, 739–779 (1999).

72. Li, Y., Yin, Y. & Mariuzza, R. A. Structural and Biophysical Insights into the Role of CD4 and CD8 in T Cell Activation. *Front Immunol* **4**, 206 (2013).

73. Ferrington, D. A. & Gregerson, D. S. Chapter 3 Immunoproteasomes Structure, Function, and Antigen Presentation. *Prog Mol Biol Transl* **109**, 75–112 (2012).

74. Griffin, T. A. *et al.* Immunoproteasome Assembly: Cooperative Incorporation of Interferon γ (IFN-γ)–inducible Subunits. *J Exp Medicine* **187**, 97–104 (1998).

75. Chen, W., Norbury, C. C., Cho, Y., Yewdell, J. W. & Bennink, J. R. Immunoproteasomes Shape Immunodominance Hierarchies of Antiviral Cd8+ T Cells at the Levels of T Cell Repertoire and Presentation of Viral Antigens. *J Exp Medicine* **193**, 1319–1326 (2001).

76. Reits, E. *et al.* A Major Role for TPPII in Trimming Proteasomal Degradation Products for MHC Class I Antigen Presentation. *Immunity* **20**, 495–506 (2004).

77. Neefjes, J., Momburg, F. & Hammerling, G. Selective and ATP-dependent translocation of peptides by the MHC-encoded transporter. *Science* **261**, 769–771 (1993).

78. Shepherd, J. C. *et al.* TAP1-dependent peptide translocation in vitro is ATP dependent and peptide selective. *Cell* **74**, 577–584 (1993).

79. Momburg, F., Roelse, J., Hämmerling, G. & Neefjes, J. Peptide size selection by the major histocompatibility complex-encoded peptide transporter. *J Exp Medicine* **179**, 1613–1623 (1994).

80. Burgevin, A. *et al.* A Detailed Analysis of the Murine TAP Transporter Substrate Specificity. *Plos One* **3**, e2402 (2008).

81. Evnouchidou, I. & Endert, P. van. Peptide trimming by endoplasmic reticulum aminopeptidases: Role of MHC class I binding and ERAP dimerization. *Human Immunology* **80**, 290–295 (2019).

82. Saveanu, L. *et al.* Concerted peptide trimming by human ERAP1 and ERAP2 aminopeptidase complexes in the endoplasmic reticulum. *Nat Immunol* **6**, 689–697 (2005).

83. Evnouchidou, I., Weimershaus, M., Saveanu, L. & Endert, P. van. ERAP1–ERAP2 Dimerization Increases Peptide-Trimming Efficiency. *J Immunol* **193**, 901–908 (2014).

84. Hattori, A. & Tsujimoto, M. Endoplasmic reticulum aminopeptidases: biochemistry, physiology and pathology. *J Biochem* **154**, 219–228 (2013).

85. Chen, H. *et al.* ERAP1-ERAP2 dimers trim MHC I-bound precursor peptides; implications for understanding peptide editing. *Sci Rep-uk* **6**, 28902 (2016).

86. Boyle, L. H. *et al.* Tapasin-related protein TAPBPR is an additional component of the MHC class I presentation pathway. *PNAS* **110**, 3465–70 (2013).

87. Hermann, C. *et al.* TAPBPR alters MHC class I peptide presentation by functioning as a peptide exchange catalyst. *eLife* **4**, e09617 (2015).

88. Hermann, C., Trowsdale, J. & Boyle, L. TAPBPR: a new player in the MHC class I presentation pathway. *Tissue Antigens* **85**, 155–166 (2015).

89. Morozov, G. I. *et al.* Interaction of TAPBPR, a tapasin homolog, with MHC-I molecules promotes peptide editing. *PNAS* **113**, E1006-15 (2016).

90. Paulsson, K. & Wang, P. Chaperones and folding of MHC class I molecules in the endoplasmic reticulum. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **1641**, 1–12 (2003).

91. Rutkevich, L. A. & Williams, D. B. Participation of lectin chaperones and thiol oxidoreductases in protein folding within the endoplasmic reticulum. *Current Opinion in Cell Biology* **23**, 157–166 (2011).

92. Gething, M.-J. & Sambrook, J. Protein folding in the cell. *Nature* **355**, 33–45 (1992).

93. Parodi, A. J. Role of N-oligosaccharide endoplasmic reticulum processing reactions in glycoprotein folding and degradation. *The Biochemical journal* **348 Pt 1**, 1–13 (2000).

94. Hebert, D. N., Garman, S. C. & Molinari, M. The glycan code of the endoplasmic reticulum: asparagine-linked carbohydrates as protein maturation and quality-control tags. *Trends Cell Biol* **15**, 364–370 (2005).

95. Hughes, E. A. & Cresswell, P. The thiol oxidoreductase ERp57 is a component of the MHC class I peptide-loading complex. *Curr Biol* **8**, 709–713 (1998).

96. Sadasivan, B., Lehner, P. J., Ortmann, B., Spies, T. & Cresswell, P. Roles for calreticulin and a novel glycoprotein, Tapasin, in the interaction of MHC class I molecules with TAP. *Immunity* **5**, 103–114 (1996).

97. Ortmann, B. *et al.* A critical role for Tapasin in the assembly and function of multimeric MHC Class I-TAP complexes. *Science* **277**, 1306 (1997).

98. Ritz, U. & Seliger, B. The Transporter Associated with Antigen Processing (TAP): Structural Integrity, Expression, Function, and Its Clinical Relevance. *Mol Med* **7**, 149–158 (2001).

99. Schmitt, L. & Tampé, R. Structure and mechanism of ABC transporters. *Curr Opin Struc Biol* **12**, 754–760 (2002).

100. Blees, A. *et al.* Assembly of the MHC I peptide-loading complex determined by a conserved ionic lock-switch. *Sci Rep-uk* **5**, 17341 (2015).

101. Koch, J., Guntrum, R., Heintke, S., Kyritsis, C. & Tampé, R. Functional Dissection of the Transmembrane Domains of the Transporter Associated with Antigen Processing (TAP)*. *J Biol Chem* **279**, 10142–10147 (2004).

102. Neumann, L. & Tampé, R. Kinetic analysis of peptide binding to the TAP transport complex: evidence for structural rearrangements induced by substrate binding. *J Mol Biol.* **294**, 1203–1213 (1999).

103. Lehnert, E. & Tampé, R. Structure and Dynamics of Antigenic Peptides in Complex with TAP. *Front Immunol* **8**, 10 (2017).

104. Koopmann, J.-O., Post, M., Neefjes, J. J., Hämmerling, G. J. & Momburg, F. Translocation of long peptides by transporters associated with antigen processing (TAP). *European Journal of Immunology* **26**, 1720–1728 (1996).

105. Pocanschi, C. L. *et al.* Structural and Functional Relationships between the Lectin and Arm Domains of Calreticulin*. *J Biol Chem* **286**, 27266–27277 (2011).

106. Frickel, E.-M. *et al.* TROSY-NMR reveals interaction between ERp57 and the tip of the calreticulin P-domain. *Proc National Acad Sci* **99**, 1954 (2002).

107. Frickel, E.-M. *et al.* ERp57 is a multifunctional thiol-disulfide oxidoreductase*. *J Biol Chem* **279**, 18277–18287 (2004).

108. Nakamura, K. *et al.* Functional specialization of calreticulin domains. *J Cell Biology* **154**, 961–972 (2001).

109. Raghavan, M., Wijeyesakere, S. J., Peters, L. R. & Cid, N. D. Calreticulin in the immune system: ins and outs. *Trends Immunol* **34**, 13–21 (2013).

110. Rizvi, S. M., Cid, N. D., Lybarger, L. & Raghavan, M. Distinct functions for the glycans of Tapasin and heavy chains in the assembly of MHC class I molecules. *J Immunol* **186**, 2309–2320 (2011).

111. Cid, N. D. *et al.* Modes of Calreticulin Recruitment to the Major Histocompatibility Complex Class I Assembly Pathway2. *J Biol Chem* **285**, 4520–4535 (2010).

112. Wearsch, P. A., Peaper, D. R. & Cresswell, P. Essential glycan-dependent interactions optimize MHC class I peptide loading. *Proc National Acad Sci* **108**, 4950–4955 (2011).

113. Howe, C. *et al.* Calreticulin-dependent recycling in the early secretory pathway mediates optimal peptide loading of MHC class I molecules. *Embo J* **28**, 3730–3744 (2009).

114. Oliver, J., Roderick, H., Llewellyn, D. & High, S. ERp57 functions as a subunit of specific complexes formed with the ER lectins calreticulin and calnexin. *Mol Biol Cell* **10**, 2573–2582 (1999).

115. Morrice, N. A. & Powis, S. J. A role for the thiol-dependent reductase ERp57 in the assembly of MHC class I molecules. *Curr Biol* **8**, 713–716 (1998).

116. Lindquist, J. A., Jensen, O. N., Mann, M. & Hämmerling, G. J. ER-60, a chaperone with thiol-dependent reductase activity involved in MHC class I assembly. *Embo J* **17**, 2186–2195 (1998).

117. Peaper, D. R., Wearsch, P. A. & Cresswell, P. Tapasin and ERp57 form a stable disulfide-linked dimer within the MHC class I peptide-loading complex. *The EMBO Journal* **24**, 3613–3623 (2005).

118. Santos, S. G. *et al.* Major Histocompatibility Complex Class I-ERp57-Tapasin Interactions within the Peptide-loading Complex. *JBC* **282**, 17587–17593 (2007).

119. Dick, T. P., Bangia, N., Peaper, D. R. & Cresswell, P. Disulfide bond isomerization and the assembly of MHC class I-peptide complexes. *Immunity* **16**, 87–98 (2002).

120. Peaper, D. R. & Cresswell, P. The redox activity of ERp57 is not essential for its functions in MHC class I peptide loading. *PNAS* **105**, 10477 (2008).

121. Zhang, Y. *et al.* ERp57 Does Not Require Interactions with Calnexin and Calreticulin to Promote Assembly of Class I Histocompatibility Molecules, and It Enhances Peptide Loading Independently of Its Redox Activity*. *J Biol Chem* **284**, 10160–10173 (2009).

122. Kienast, A., Preuss, M., Winkler, M. & Dick, T. P. Redox regulation of peptide receptivity of major histocompatibility complex class I molecules by ERp57 and tapasin. *Nat Immunol* **8**, 864–872 (2007).

123. Oldham, M. L., Grigorieff, N. & Chen, J. Structure of the transporter associated with antigen processing trapped by herpes simplex virus. *Elife* **5**, e21829 (2016).

124. Huotari, J. & Helenius, A. Endosome maturation. *Embo J* **30**, 3481–3500 (2011).

125. Watts, C. The exogenous pathway for antigen presentation on major histocompatibility complex class II and CD1 molecules. *Nat Immunol* **5**, 685–692 (2004).

126. Marks, M., Blum, J. & Cresswell, P. Invariant chain trimers are sequestered in the rough endoplasmic reticulum in the absence of association with HLA class II antigens. *J Cell Biology* **111**, 839–855 (1990).

127. Neumann, J. & Koch, N. Assembly of major histocompatibility complex class II subunits with invariant chain. *Febs Lett* **579**, 6055–6059 (2005).

128. Landsverk, O., Bakke, O. & Gregers, T. MHC II and the Endocytic Pathway: Regulation by Invariant Chain. *Scand J Immunol* **70**, 184–193 (2009).

129. Watts, C. The endosome–lysosome pathway and information generation in the immune system. *Biochimica Et Biophysica Acta Bba - Proteins Proteom* **1824**, 14–21 (2012).

130. Riberdy, J. M., Newcomb, J. R., Surman, M. J., Barbosat, J. A. & Cresswell, P. HLA-DR molecules from an antigen-processing mutant cell line are associated with invariant chain peptides. *Nature* **360**, 474–477 (1992).

131. Morris, P. *et al.* An essential role for HLA–DM in antigen presentation by class II major histocompatibility molecules. *Nature* **368**, 551–554 (1994).

132. Yin, L. & Stern, L. J. HLA-DM Focuses on Conformational Flexibility Around P1 Pocket to Catalyze Peptide Exchange. *Front Immunol* **4**, 336 (2013).

133. Wieczorek, M. *et al.* MHC class II complexes sample intermediate states along the peptide exchange pathway. *Nat Commun* **7**, 13224 (2016).

134. Yin, L. *et al.* Susceptibility to HLA-DM Protein Is Determined by a Dynamic Conformation of Major Histocompatibility Complex Class II Molecule Bound with Peptide. *J Biol Chem* **289**, 23449–23464 (2014).

135. Denzin, L. K., Sant'Angelo, D. B., Hammond, C., Surman, M. J. & Cresswell, P. Negative Regulation by HLA-DO of MHC Class II-Restricted Antigen Processing. *Science* **278**, 106–109 (1997).

136. Herberg, J. A. *et al.* Genomic analysis of the Tapasin gene, located close to the TAP loci in the MHC. *European Journal of Immunology* **28**, 459–467 (1998).

137. Lehner, P. J., Surman, M. J. & Cresswell, P. Soluble Tapasin restores MHC class I expression and function in the tapasin-negative cell line .220. *Immunity* **8**, 221–231 (1998).

138. Bangia, N., Lehner, P. J., Hughes, E. A., Surman, M. & Cresswell, P. The N-terminal region of tapasin is required to stabilize the MHC class I loading complex. *European Journal of Immunology* **29**, 1858–1870 (1999).

139. III, A. G. G., Lehner, P. J., Cresswell, P. & Spies, T. Regulation of MHC class I heterodimer stability and interaction with TAP by tapasin. *Immunogenetics* **46**, 477–483 (1997).

140. Purcell, A. W. *et al.* Quantitative and qualitative influences of Tapasin on the class I peptide repertoire. *The Journal of Immunology* **166**, 1016–1027 (2001).

141. Williams, A. P., Peh, C., Purcell, A. W., McCluskey, J. & Elliott, T. Optimization of the MHC class I peptide cargo is dependent on Tapasin. *Immunity* **16**, 509–520 (2002).

142. Grandea, A. G. *et al.* Impaired Assembly yet Normal Trafficking of MHC Class I Molecules in Tapasin Mutant Mice. *Immunity* **13**, 213–222 (2000).

143. Garbi, N. *et al.* Impaired immune responses and altered peptide repertoire in tapasin-deficient mice. *Nat Immunol* **1**, 234–238 (2000).

144. Barnden, M. J., Purcell, A. W., Gorman, J. J. & McCluskey, J. Tapasin-mediated retention and optimization of peptide ligands during the assembly of class I molecules. *J Immunol* **165**, 322–330 (2000).

145. Dong, G., Wearsch, P. A., Peaper, D. R., Cresswell, P. & Reinisch, K. M. Insights into MHC Class I Peptide Loading from the Structure of the Tapasin-ERp57 Thiol Oxidoreductase Heterodimer. *Immunity* **30**, 21–32 (2009).

146. Chen, M. & Bouvier, M. Analysis of interactions in a tapasin/class I complex provides a mechanism for peptide selection. *The EMBO Journal* **26**, 1681–1690 (2007).

147. Wright, C. A., Kozik, P., Zacharias, M. & Springer, S. Tapasin and other chaperones: models of the MHC class I loading complex. *Biol Chem* **385**, 763–778 (2004).

148. Yu, Y. Y. *et al.* An Extensive Region of an MHC Class I α2 Domain Loop Influences Interaction with the Assembly Complex. *J. Immunol.* **163**, 4427 (1999).

149. Turnquist, H. R. *et al.* The Ig-like domain of tapasin influences intermolecular interactions. *J Immunol.* **172**, 2976–84 (2004).

150. Fleischmann, G. *et al.* Mechanistic Basis for Epitope Proofreading in the Peptide-Loading Complex. *J. Immunol.* **195**, 4503–4513 (2015).

151. Greenwood, R., Shimizu, Y., Sekhon, G. S. & DeMars, R. Novel allele-specific, post-translational reduction in HLA class I surface expression in a mutant human B cell line. *J. Immunol.* **153**, 5525 (1994).

152. Peh, C. *et al.* HLA-B27–restricted antigen presentation in the absence of Tapasin reveals polymorphism in mechanisms of HLA class I peptide loading. *Immunity* **8**, 531–542 (1998).

153. Zernich, D. *et al.* Natural HLA Class I Polymorphism Controls the Pathway of Antigen Presentation and Susceptibility to Viral Evasion. *J Exp Medicine* **200**, 13–24 (2004).

154. Turnquist, H. R. *et al.* HLA-B polymorphism affects interactions with multiple endoplasmic reticulum proteins. *European Journal of Immunology* **30**, 3021–3028 (2000).

155. Badrinath, S., Kunze-Schumacher, H., Blasczyk, R., Huyton, T. & Bade-Doeding, C. A Micropolymorphism Altering the Residue Triad 97/114/156 Determines the Relative Levels of Tapasin Independence and Distinct Peptide Profiles for HLA-A*24 Allotypes. *J Immunol Res* **2014**, 1–12 (2014).

156. Bashirova, A. A. *et al.* HLA tapasin independence: broader peptide repertoire and HIV control. *PNAS* **117**, 28232–28238 (2020).

157. Rizvi, S. *et al.* Distinct Assembly Profiles of HLA-B Molecules. *J Immunol.* **192**, 4967–4976 (2014).

158. Park, B., Lee, S., Kim, E. & Ahn, K. A Single Polymorphic Residue Within the Peptide-Binding Cleft of MHC Class I Molecules Determines Spectrum of Tapasin Dependence. *J Immunol* **170**, 961 (2003).

159. Badrinath, S. *et al.* Position 156 influences the peptide repertoire and tapasin dependency of human leukocyte antigen B*44 allotypes. *Haematologica* **97**, 98–106 (2012).

160. Garstka, M. *et al.* Tapasin dependence of major histocompatibility complex class I molecules correlates with their conformational flexibility. *FASEB J.* **25**, 3989–3998 (2011).

161. Roder, G. *et al.* Tapasin discriminates peptide-human leukocyte antigen-A*02:01 complexes formed with natural ligands. *J Biol Chem* **286**, 20547–20557 (2011).

162. Howarth, M., Williams, A., Tolstrup, A. B. & Elliott, T. Tapasin enhances MHC class I peptide presentation according to peptide half-life. *PNAS* **101**, 11737 (2004).

163. Geironson, L. *et al.* Tapasin facilitation of natural HLA-A and -B allomorphs Is strongly influenced by peptide length, depends on stability, and separates closely related allomorphs. *J Immunol.* **191**, 3939–3947 (2013).

164. Fabian, H. *et al.* HLA–B27 heavy chains distinguished by a micropolymorphism exhibit differential flexibility. *Arthritis Rheum.* **62**, 978–987 (2010).

165. Pöhlmann, T. *et al.* Differential Peptide Dynamics Is Linked to Major Histocompatibility Complex Polymorphism. *J Biol Chem* **279**, 28197–28201 (2004).

166. Mage, M. G. *et al.* The Peptide-Receptive Transition State of MHC Class I Molecules: Insight from Structure and Molecular Dynamics. *J Immunol* **189**, 1391–1399 (2012).

167. Bailey, A. *et al.* Selector function of MHC I molecules is determined by protein plasticity. *Sci Rep* **5**, 14928 (2015).

168. Ilca, T. F., Drexhage, L. Z., Brewin, G., Peacock, S. & Boyle, L. H. Distinct Polymorphisms in HLA Class I Molecules Govern Their Susceptibility to Peptide Editing by TAPBPR. *Cell Reports* **29**, 1621-1632.e3 (2019).

169. McShan, A. C. *et al.* Peptide exchange on MHC-I by TAPBPR is driven by a negative allostery release cycle. *Nature chemical biology* **14**, 811–820 (2018).

170. Sieker, F., Springer, S. & Zacharias, M. Comparative molecular dynamics analysis of tapasin-dependent and -independent MHC class I alleles. *Protein Sci* **16**, 299–308 (2007).

171. Neerincx, A. & Boyle, L. H. Properties of the tapasin homologue TAPBPR. *Current Opinion in Immunology* **46**, 97–102 (2017).

172. Neerincx, A. *et al.* TAPBPR bridges UDP-glucose:glycoprotein glucosyltransferase 1 onto MHC class I to provide quality control in the antigen presentation pathway. *Elife* **6**, e23049 (2017).

173. Jiang, J. *et al.* Crystal structure of a TAPBPR-MHC I complex reveals the mechanism of peptide editing in antigen presentation. *Science* **358**, 1064–1068 (2017).

174. Thomas, C. & Tampé, R. Structure of the TAPBPR–MHC I complex defines the mechanism of peptide loading and editing. *Science* **358**, 1060 (2017).

175. Ilca, F. T. *et al.* TAPBPR mediates peptide dissociation from MHC class I using a leucine lever. *eLife* **7**, (2018).

176. Sagert, L., Hennig, F., Thomas, C. & Tampé, R. A loop structure allows TAPBPR to exert its dual function as MHC I chaperone and peptide editor. *eLife* **9**, (2020).

177. McShan, A. C. *et al.* TAPBPR Promotes Antigen Loading on MHC-I Molecules Using a Peptide Trap. *bioRxiv* 2020.04.24.059634 (2020) doi:10.1101/2020.04.24.059634.

178. Blees, A. *et al.* Structure of the human MHC-I peptide-loading complex. *Nature* **551**, 525–528 (2017).

179. McShan, A. C. *et al.* Molecular determinants of chaperone interactions on MHC-I for folding and antigen repertoire selection. *PNAS* **116**, 25602–25613 (2019).

180. Hafstrand, I. *et al.* Successive crystal structure snapshots suggest the basis for MHC class I peptide loading and editing by tapasin. *PNAS* **116**, 5055–5060 (2019).

181. Rodenko, B. *et al.* Generation of peptide–MHC class I complexes through UV-mediated ligand exchange. *Nat. Protoc.* **1**, 1120–1132 (2006).

182. McMahon, R. M. *et al.* Structure of HLA-A*0301 in complex with a peptide of proteolipid protein: insights into the role of HLA-A alleles in susceptibility to multiple sclerosis. *Acta Crystallogr Sect D Biological Crystallogr* **67**, 447–454 (2011).

183. Rückert, C. *et al.* Conformational Dimorphism of Self-peptides and Molecular Mimicry in a Disease-associated HLA-B27 Subtype*. *J Biol Chem* **281**, 2306–2316 (2006).

184. Goto, N. K., Gardner, K. H., Mueller, G. A., Willis, R. C. & Kay, L. E. A robust and cost-effective method for the production of Val, Leu, Ile (δ1) methyl-protonated 15N-, 13C-, 2H-labeled proteins. *J. Biomol. NMR* **13**, 369–374 (1999).

185. Lea, W. A. & Simeonov, A. Fluorescence polarization assays in small molecule screening. *Expert Opin Drug Discov* **6**, 17–32 (2010).

186. Pascal, S. M. *NMR Primer: an HSQC-based approach*. (IM Publications LLP, 2008).

187. McCoy, A. J. *et al.* Phaser crystallographic software. *J Appl Crystallogr* **40**, 658–674 (2007).

188. Driller, R. *et al.* Crystal structure of HLA-B*27:09 complexed with the self-peptide pVIPR and Copper. (2017).

189. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr Sect D Biological Crystallogr* **66**, 213–221 (2010).

190. Afonine, P. V. *et al.* Towards automated crystallographic structure refinement with phenix.refine. *Acta Cryst.* **D68**, 352–367 (2012).

191. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr.* **66**, 486–501 (2010).

192. Williams, C. J. *et al.* MolProbity: More and better reference data for improved all-atom structure validation. *Protein Sci* **27**, 293–315 (2018).

193. Webb, B. & Sali, A. Comparative Protein Structure Modeling Using MODELLER. *Curr Protoc Bioinform* **54**, 5.6.1-5.6.37 (2016).

194. Fiser, A., Do, R. K. G. & Šali, A. Modeling of loops in protein structures. *Protein Sci* **9**, 1753–1773 (2000).

195. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J Chem Phys* **79**, 926–935 (1983).

196. Chaudhury, S., Lyskov, S. & Gray, J. J. PyRosetta: a script-based interface for implementing molecular modeling algorithms using Rosetta. *Bioinformatics* **26**, 689–691 (2010).

197. Park, H. *et al.* Simultaneous Optimization of Biomolecular Energy Functions on Features from Small Molecules and Macromolecules. *J. Chem. Theory Comput.* **12**, 6201–6212 (2016).

198. Alford, R. F. *et al.* The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design. *J Chem Theory Comput* **13**, 3031–3048 (2017).

199. Castro, J. A. L. de *et al.* HLA-B27: a registry of constitutive peptide ligands. *Tissue Antigens* **63**, 424–445 (2004).

200. Driller, R. *et al.* Metal-triggered conformational reorientation of a self-peptide bound to a disease-associated HLA-B*27 subtype. *J Biol Chem* **294**, 13269–13279 (2019).

201. Devlin, J. R. *et al.* Structural dissimilarity from self drives neoepitope escape from immune tolerance. *Nat Chem Biol* **16**, 1269–1276 (2020).

202. Wu, D., Gallagher, D. T., Gowthaman, R., Pierce, B. G. & Mariuzza, R. A. Structural basis for oligoclonal T cell recognition of a shared p53 cancer neoantigen. *Nat Commun* **11**, 2908 (2020).

203. Yanaka, S. *et al.* Peptide-dependent Conformational Fluctuation Determines the Stability of the Human Leukocyte Antigen Class I Complex. *J Biol. Chem.* **289**, 24680–24690 (2014).

204. Dédier, S., Reinelt, S., Reitinger, T., Folkers, G. & Rognan, D. Thermodynamic Stability of HLA-B*2705·Peptide Complexes: effect of peptide and major histocompatibility complex protein mutation *. *J Biol Chem* **275**, 27055–27061 (2000).

205. Lewis, J. W., Sewell, A., Price, D. & Elliott, T. HLA-A*0201 presents TAP-dependent peptide epitopes to cytotoxic T lymphocytes in the absence of tapasin. *Eur J Immunol* **28**, 3214–3220 (1998).

206. Stewart-Jones, G. B. E., McMichael, A. J., Bell, J. I., Stuart, D. I. & Jones, E. Y. A structural basis for immunodominant human T cell receptor recognition. *Nat Immunol* **4**, 657–663 (2003).

207. Zhang, S. *et al.* Structural basis of cross-allele presentation by HLA-A*0301 and HLA-A*1101 revealed by two HIV-derived peptide complexes. *Molecular Immunology* **49**, 395–401 (2011).

208. Zacharias, M. & Springer, S. Conformational Flexibility of the MHC Class I α1-α2 Domain in Peptide Bound and Free States: A Molecular Dynamics Simulation Study. *Biophys J* **87**, 2203–2214 (2004).

209. Fisette, O., Wingbermühle, S. & Schäfer, L. V. Partial Dissociation of Truncated Peptides Influences the Structural Dynamics of the MHCI Binding Groove. *Frontiers in immunology* **8**, 408 (2017).

210. Achour, A. *et al.* The Crystal Structure of H-2Dd MHC Class I Complexed with the HIV-1-Derived Peptide P18-I10 at 2.4 Å Resolution Implications for T Cell and NK Cell Recognition. *Immunity* **9**, 199–208 (1998).

211. Howard, J. C. Supply and transport of peptides presented by class I MHC molecules. *Curr Opin Immunol* **7**, 69–76 (1995).

212. Joly, E. *et al.* Co-evolution of rat TAP transporters and MHC class I RT1-A molecules. *Curr Biol* **8**, 169–180 (1998).

213. Joly, E. & Butcher, G. W. Why are there two rat TAPs? *Immunol Today* **19**, 580–585 (1998).

214. Fisette, O., Schröder, G. F. & Schäfer, L. V. Atomistic structure and dynamics of the human MHC-I peptide-loading complex. *PNAS* **117**, 20597–20606 (2025).

215. Padariya, M., Kalathiya, U., Houston, D. R. & Alfaro, J. A. Recognition Dynamics of Cancer Mutations on the ERp57-Tapasin Interface. *Cancers* **12**, (2020).

216. Saini, S. K. *et al.* Dipeptides promote folding and peptide binding of MHC class I molecules. *PNAS* **110**, 15383–15388 (2013).

217. Saini, S. K. *et al.* Dipeptides catalyze rapid peptide exchange on MHC class I molecules. *PNAS* **112**, 202–207 (2015).

218. Overall, S. A. *et al.* High throughput pMHC-I tetramer library production using chaperone-mediated peptide exchange. *Nat. Commun.* **11**, 1909 (2020).

219. Ilca, F. T., Neerincx, A., Wills, M. R., Roche, M. de la & Boyle, L. H. Utilizing TAPBPR to promote exogenous peptide loading onto cell surface MHC I molecules. *PNAS* **115**, E9353–E9361 (2018).

220. Parker, K. C., DiBrino, M., Hull, L. & Coligan, J. E. The beta 2-microglobulin dissociation rate is an accurate measure of the stability of MHC class I heterotrimers and depends on which peptide is bound. *J. Immunol.* **149**, 1896 (1992).

221. Berko, D. *et al.* Membrane-Anchored β₂-Microglobulin Stabilizes a Highly Receptive State of MHC Class I Molecules. *J Immunol* **174**, 2116 (2005).

222. Vitiello, A., Potter, T. & Sherman, L. The role of beta 2-microglobulin in peptide binding by class I molecules. *Science* **250**, 1423–1426 (1990).

223. Rock, K. L., Gamble, S., Rothstein, L. & Benacerraf, B. Reassociation with beta 2-microglobulin is necessary for Db class I major histocompatibility complex binding of an exogenous influenza peptide. *Proc National Acad Sci* **88**, 301–304 (1991).

224. Kozlowski, S. *et al.* Excess β2 microglobulin promoting functional peptide association with purified soluble class I MHC molecules. *Nature* **349**, 74–77 (1991).

225. Praveen, P. V. K., Yaneva, R., Kalbacher, H. & Springer, S. Tapasin edits peptides on MHC class I molecules by accelerating peptide exchange. *Eur J Immunol* **40**, 214–224 (2010).

226. Lewis, J. W., Neisig, A., Neefjes, J. & Elliott, T. Point mutations in the α2 domain of HLA-A2.1 define a functionally relevant interaction with TAP. *Current Biology* **6**, 873–883 (1996).

227. Anjanappa, R. *et al.* Structures of peptide-free and partially loaded MHC class I molecules reveal mechanisms of peptide selection. *Nat. Commun.* **11**, 1314 (2020).

228. Hein, Z. *et al.* Peptide-independent stabilization of MHC class I molecules breaches cellular quality control. *J Cell Sci* **127**, 2885–2897 (2014).

229. Paul, S. *et al.* HLA Class I Alleles Are Associated with Peptide-Binding Repertoires of Different Size, Affinity, and Immunogenicity. *J Immunol* **191**, 5831–5839 (2013).

230. Luckey, C. *et al.* Differences in the Expression of Human Class I MHC Alleles and Their Associated Peptides in the Presence of Proteasome Inhibitors. *J Immunol.* **167**, 1212–1221 (2001).

231. Hermann, C., Strittmatter, L. M., Deane, J. E. & Boyle, L. H. The Binding of TAPBPR and Tapasin to MHC Class I Is Mutually Exclusive. *J Immunol* **191**, 5743–5750 (2013).

232. Fisette, O., Wingbermühle, S., Tampé, R. & Schäfer, L. V. Molecular mechanism of peptide editing in the tapasin–MHC I complex. *Sci Rep* **6**, 19085 (2016).

233. Hunt, D. *et al.* Characterization of peptides bound to the class I MHC molecule HLA-A2.1 by mass spectrometry. *Science* **255**, 1261–1263 (1992).

234. Wei, M. L. & Cresswell, P. HLA-A2 molecules in an antigen-processing mutant cell contain signal sequence-derived peptides. *Nature* **356**, 443–446 (1992).

235. Flajnik, M. F. & Kasahara, M. Comparative Genomics of the MHC Glimpses into the Evolution of the Adaptive Immune System. *Immunity* **15**, 351–362 (2001).

236. Kaufman, J. Generalists and Specialists: A New View of How MHC Class I Molecules Fight Infectious Pathogens. *Trends Immunol* **39**, 367–379 (2018).

237. Abduriyim, S., Zou, D. & Zhao, H. Origin and evolution of the major histocompatibility complex class I region in eutherian mammals. *Ecol Evol* **9**, 7861–7874 (2019).

238. Piertney, S. B. & Oliver, M. K. The evolutionary ecology of the major histocompatibility complex. *Heredity* **96**, 7–21 (2006).

239. Kaufman, J. *et al.* The chicken B locus is a minimal essential major histocompatibility complex. *Nature* **401**, 923–925 (1999).

240. Walker, B. A. *et al.* The dominantly expressed class I molecule of the chicken MHC is explained by coevolution with the polymorphic peptide transporter (TAP) genes. *Proc National Acad Sci* **108**, 8396–8401 (2011).

241. Hateren, A. van *et al.* A Mechanistic Basis for the Co-evolution of Chicken Tapasin and Major Histocompatibility Complex Class I (MHC I) Proteins. *J Biol Chem* **288**, 32797–32808 (2013).

242. Furukawa, H. *et al.* Polymorphism of TAPASIN and its linkage disequilibria with HLA class II genes in the Japanese population. *Tissue Antigens* **52**, 279–281 (1998).

243. Williams, A. P. *et al.* Identification of novel Tapasin polymorphisms and linkage disequilibrium to MHC class I alleles. *Immunogenetics* **52**, 9–11 (2000).

244. Obst, R., Armandola, E. A., Nijenhuis, M., Momburg, F. & Hämmerling, G. J. TAP polymorphism does not influence transport of peptide variants in mice and humans. *European Journal of Immunology* **25**, 2170–2176 (1995).

245. Howard, J. C. & Seelig, A. Peptides and the proteasome. *Nature* **365**, 211–212 (1993).

246. Deverson, E. *et al.* Functional analysis by site-directed mutagenesis of the complex polymorphism in rat transporter associated with antigen processing. *J Immunol* **160**, 2767–79 (1998).

247. Diederichs, K. & Karplus, A. P. Improved R-factors for diffraction data analysis in macromolecular crystallography. *Nat Struct Biol* **4**, 269–275 (1997).

248. Karplus, A. P. & Diederichs, K. Linking Crystallographic Model and Data Quality. *Science* **336**, 1030 (2012).

249. Joly, E. *et al.* The distribution of Tap2 alleles among laboratory rat RT1 haplotypes. *Immunogenetics* **40**, 45–53 (1994).

250. Peters, F. B., Brock, A., Wang, J. & Schultz, P. G. Photocleavage of the Polypeptide Backbone by 2-Nitrophenylalanine. *Chem Biol* **16**, 148–152 (2009).