Structural Investigations of MHC Class II Conformers

Dissertation zur Erlangung des akademischen Grades des

Doktors der Naturwissenschaften (Dr. rer. nat.)

eingereicht im Fachbereich Biologie, Chemie, Pharmazie

der Freien Universität Berlin

vorgelegt von

Sebastian Günther

aus Oldenburg (Oldb.)

2011

Die vorliegende Arbeit wurde in der Zeit von Okt. 2006 bis Dez. 2009 in der Arbeitsgruppe „Zelluläre Immunologie“ am Max-Delbrück-Centrum für Molekulare Medizin in Berlin-Buch unter der Leitung von Dr. Kirsten Falk und Dr. Olaf Rötzschke und von Jan. 2010 bis Sept. 2011 in der Arbeitsgruppe „Protein Engineering“ am Leibniz-Institut für Molekulare Pharmakologie (FMP) in Berlin-Buch unter der Leitung von Prof. Dr. Christian Freund durchgeführt.

1. Gutachter: Prof. Dr. Udo Heinemann

2. Gutachter: Prof. Dr. Christian Freund

Disputation am 1.12.2011

Contents

1 Introduction

The immune system is formed by organs, cells and molecules that protect the body against invading microorganisms, contaminants, toxins and malignant cells. Of outermost importance hereby is the ability to discriminate between “foreign” and “self”, so that the immune system is not attacking the body’s own healthy cells but only potential pathogens or malignant cells. In order to cope with this, the vertebrate immune system has evolved two main, intimately connected arms of defense, the innate and the adaptive immune system.

1.1 Innate immunity

The innate immune system is the more ancient part of the body’s defense system. It is the first line of defense and responsible for the early detection of pathogens in a non-specific manner. The epithelial cells in the skin, gastrointestinal tract or the lungs represent a first barrier for a pathogen. In case it still invades the body, a set of plasma proteins called complement is activated, ultimately leading to removal of the pathogen by direct cell lysis, by phagocytes and by recruitment of inflammatory cells. Proteolytic activation of the complement cascade is initiated by antibody or sugar-receptor binding and also by direct binding of complement proteins to the pathogen surface. In addition, cells like granulocytes, macrophages and Natural Killer cells form the cellular component of the innate immune system. These cells possess pattern recognition receptors by which they can bind conserved structures of microbial pathogens (so called pathogen-associated molecular patterns; PAMPs), e.g. lipopolysaccharide, double stranded viral RNA or bacterial DNA. Upon recognition of PAMPs, these cells, especially macrophages and neutrophils but also dendritic cells (DC), can phagocytose the pathogen and use parts of it to alarm the specific arm of the immune system, the adaptive immunity.

In general, the innate immune system can directly attack the pathogen and, additionally, initiates a local inflammation leading to the infiltration of other immune cells.

1.2 Adaptive immunity

The central paradigm for the adaptive immunity is described by the clonal selection theory as formulated by Burnet (Burnet, 1957). Basically, it states that the adaptive immune response is

mediated by antigen-specific lymphocytes. These are activated through specific recognition of the antigen by receptors that initiate the cell to proliferate clonally into effector and also memory cells with receptor specificity identical to the parental cell. Self-reactive lymphocytes are removed at an early stage of the cell development. Later it was found that remaining self-reactive lymphocytes are suppressed by regulatory cells. The most distinguishing factor of the adaptive immune system is, first, the exquisite specificity by which pathogenic structures can be targeted and, second, the ability to form long-lived memory cells that remember these structures and thereby boost an immune response upon reencounter. In the innate immune system the specificity of the receptors used to detect a pathogen is somatically encoded and therefore only of limited diversity. In contrast, the receptors of the adaptive immune system can be assembled by somatic rearrangements, leading to an almost indefinite diversity and enabling the immune system to adapt to new, foreign antigens.

The adaptive immune system has two major routes to elicit a specific response against the pathogen. First, on the cellular side of the adaptive immune system, activated cytotoxic T lymphocytes can directly kill infected or malignant cells. Second, on the humoral arm of the adaptive immune system, T lymphocytes can provide stimulating cytokines for B lymphocytes which secrete highly specific antibodies that neutralize the pathogen or mark it for removal by other components of the immune system.

The two major antigen receptors of the adaptive immune system are the T cell receptor (TCR) and B cell receptor (BCR), respectively antibodies. Both receptors belong to the immunoglobulin superfamily. In case of T cells the TCR is membrane-bound, whereas B cells initially express a membrane-bound form (the BCR) that upon stimulation leads to the release of soluble antibodies. While antibodies can directly bind the antigen, T cell receptors only recognize the antigen when it is processed into short peptide fragments and bound to Major Histocompatibility Complex (MHC) molecules, a fact known as MHC-restriction of T cells (Zinkernagel & Doherty, 1974).

T cells can be further discriminated into two major classes based on the expression of either the CD4- or the CD8-coreceptor. $CD8^+$-T cells, also known as cytotoxic T lymphocytes (CTL), are only activated by MHC class I molecules in complex with a peptide that is mostly of endogenous origin. Upon activation these cells can directly kill the infected or malignant cell by inducing apoptosis using effector molecules like granzyme or perforin and activate further immune cells by the release of interferon (IFN) γ and other cytokines (Barry & Bleackley, 2002). Peptides presented by MHC class II are mostly of exogenous origin and

activate exclusively $CD4^+$-T cells (T helper or T_H cells). These cells do not directly attack the pathogen but secrete cytokines that steer a tailored immune response. Initially, upon activation two distinct lineages were described for T_H cells. They can be polarized to T_H1 cells secreting IFNγ, thereby promoting inflammation and activation of e.g. macrophages. Alternatively, they can develop to T_H2 cells secreting interleukin 4 (IL-4) and IL-10, promoting antibody secretion by B cells (Abbas *et al.*, 1996). Both polarizations are mutually exclusive with the cytokines promoting one developmental route at the same time inhibiting the other route. Recently, several new T_H subtypes have been proposed, among which T_H17 cells are the most notable, shown to secrete IL-17A, IL-17F and IL22 and being important for immunity against extracellular bacteria and fungi (Zhu *et al.*, 2010). Another very important $CD4^+$-T cell subset are regulatory T cells (T_{reg}), characterized by the expression of the transcription factor FoxP3 (Hori *et al.*, 2003). These cells have immunosuppressive functions by which they balance the proinflammatory function of other T cells and prevent overreaction during an immune response against pathogens. The route taken by a $CD4^+$-T cell depends on the cytokine milieu in which the T cell is activated. And this in turn depends very much on the antigen-presenting cell (APC). These cells, in particular dendritic cells (DC), steer their cytokine secretion based on different pattern recognition receptors that can be activated by molecular structures that are conserved over different pathogens and allow discrimination from the body's own cells. These structures are called pathogen-associated molecular patterns (PAMP). This differential activation enables the APC to modulate the type of immune response.

1.3 The Major Histocompatibility Complex

For activation all T cells need to bind with their antigen-specific TCR to a cognate peptide-MHC complex (pMHC). These glycoproteins as well as most of the molecules needed to generate and load the peptides onto the MHC are encoded within a contiguous region on chromosome 6 of the human genome. These genes were first identified because of their critical involvement in the immune response to transplanted tissues and were thus termed major histocompatibility complex (Snell, 1948).

The molecules encoded by the MHC can be discriminated into two classes. While the overall structure of class I and class II molecules is very similar, there are distinguishing features (Figure 1). Both proteins are type I transmembrane glycoproteins with one (MHC I) or two (MHC II) short cytoplasmic tails. In both cases the two membrane-distal domains make up the antigen-binding site with an extended β-sheet and two flanking α-helices. In between these

helices the antigenic peptide is bound. The two membrane-proximal domains are immunoglobulin-like. The domain organization is different between the two classes. In MHC I three of the four domains are formed by one polypeptide, the heavy chain, the fourth domain is the non-covalently bound β_2-microglobulin (β_2m) as invariant part of all MHC I. The MHC II is formed by two almost equally long chains (α and β), which both contain two domains and each of the two chains forms one half of the antigen-binding site (Figure 1). The most distinguishing factor is the architecture of the antigen-binding site. For MHC I the length of the bound peptide is generally restricted to 8-10 amino acids by the fact that both ends of the binding groove are closed by conserved amino acids (Fremont *et al.*, 1992; Madden *et al.*, 1993). These residues fix the antigen at both termini by interactions with the peptide main chain (Figure 1B). But they allow the peptide to bulge out in the middle, thereby enabling the binding of significantly longer peptides, even up to 16 amino acids (Bade-Doding *et al.*, 2011; Fremont *et al.*, 1992; Madden *et al.*, 1993). The terminal residues of the peptide are buried within the MHC but the remaining side chains are free to be engaged by the TCR. In contrast, in MHC II the ends of the peptide binding groove are not closed by conserved hydrophobic residues but are open. Bound peptides are commonly 13-18 amino acids long, but can also be significantly longer (Chicz *et al.*, 1992; Hunt *et al.*, 1992). The peptide is fixed by hydrogen bonds from side chains of conserved MHC II residues to the main-chain of the peptide along the whole binding site (Figure 1B), fixing the peptide in a polyproline type II helix in a sequence-independent manner (Jardetzky *et al.*, 1996).

The MHC is the most polymorphic gene locus in the human genome (Mungall *et al.*, 2003). Interestingly, for both MHC classes the polymorphisms cluster in the peptide-binding domains. The consequences thereof were discovered in the early 1990ies by sequencing of peptides eluted from different MHC alleles. It was shown that each MHC allelic variant has a particular binding motif (Falk *et al.*, 1991). Together with the crystal structures of MHC I and II it was clear that these polymorphisms generate different binding pockets to accommodate side chains of the bound peptide (Figure 1C). For class I the pockets are named from A to F, with pockets A and F being the most important. Pocket A accommodates the amino-terminus and pocket F the carboxy-terminus of the peptide. In MHC II the pockets are named from P1 to P9 according to which residue of the peptide ligand they accommodate counting from the first pocket-anchored residue at the peptide's amino-terminus towards its carboxy-terminus. Thus, in addition to the sequence-independent backbone interaction of the MHC with the peptide, there are always sequence-dependent interactions between the antigen-binding site and the peptide side chains.

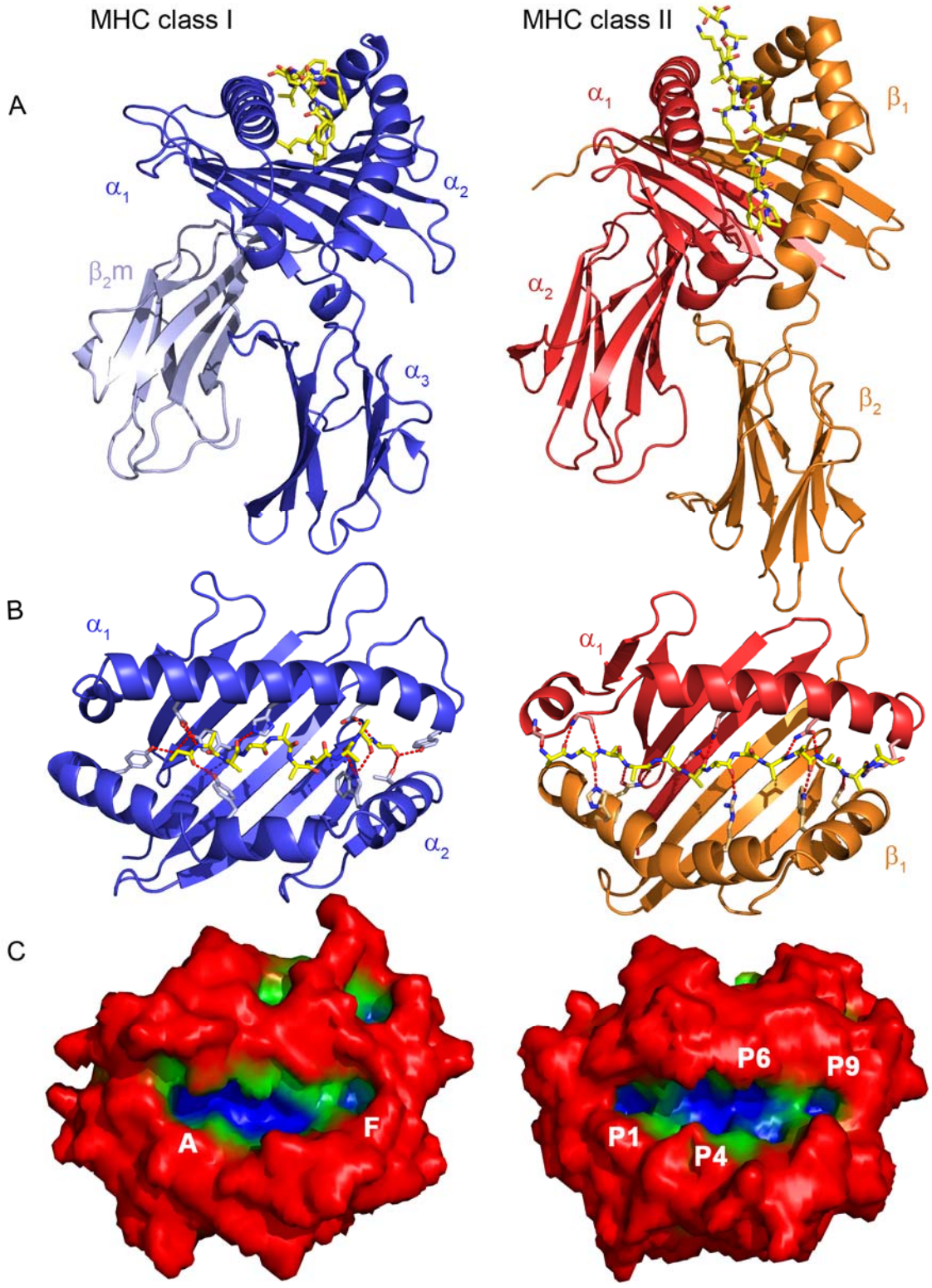


Figure 1: *Structural comparison of MHC class I and II complexes.* **A**, Cartoon representation of the extracellular domains for each class with the bound peptide as yellow sticks (MHC I: HLA-A2 with hepatitis B virus derived peptide, PDB: 1HHH; MHC II: HLA-DR1 with influenza virus derived peptide, PDB: 2G9H). Individual domains are indicated. **B**, View into the antigen-binding site with the bound peptide represented by a poly-alanine model. MHC-residues participating in the conserved hydrogen bonding network (red lines) are shown as sticks. **C**, Surface representation of MHC reveals pockets in the antigen-binding site (orientation as in B). The peptide was omitted for clarity. The most important pockets for both classes are indicated. The surface is colored from red to blue according to the pocket depth from the surface. Pocket depth was calculated using the Travel Distance Suite (Coleman & Sharp, 2006).

While MHC I are found on all nucleated cells, MHC II are only found on professional antigen-presenting cells like DCs, B cells and macrophages. For MHC class I there are three different α-chain genes (termed HLA-A, -B, -C) and for MHC class II three pairs of α- and β-chain genes (termed HLA-DR, -DP, -DQ), often including the expression of a second DR-β-chain. Thus, from each MHC three different MHC I and up to four different MHC II can be expressed. And due to the co-dominant expression of both parental haplotypes, altogether 14 different MHC molecules can be found on a single cell. This combinatorial diversity ensures that the immune system is able to present any given foreign (and self-) antigen at least by one of its MHC molecules.

1.4 Antigen processing

The generation of peptides for the presentation by MHC I is tightly linked to the default degradation pathway of proteins by the proteasome. Peptide fragments generated by the 20S-proteasome are transported via the heterodimeric transporter associated with antigen processing (TAP) from the cytosol into the endoplasmic reticulum (ER) lumen. Here, the processed peptides associate with the freshly synthesized empty MHC I, which is stabilized by the ER-resident chaperone calnexin. The loading itself is mediated by the peptide-loading complex (PLC), which consists of several integral proteins like the lectin calreticulin, the thiol oxidoreductase ERp57, the glycoprotein tapasin and also TAP. The covalently linked complex of tapasin and ERp57 can stabilize the empty MHC I and facilitate the editing of the bound peptide repertoire for those with the highest stability (Howarth *et al.*, 2004; Wearsch & Cresswell, 2007). Usually peptides arriving in the ER already have a carboxy-terminus with hydrophobic or positively charged residues suitable for binding to MHC I derived from their degradation in the proteasome (Mo *et al.*, 1999). Longer peptides can be further trimmed by proteases in the cytosol (York *et al.*, 2006). The final amino-terminal end is generated inside the ER by ER aminopeptidases (ERAP) 1&2 (Saveanu *et al.*, 2005). When a stable peptide complex is formed, the peptide-MHC complex (pMHC) is transported via the Golgi apparatus to the plasma membrane ready for scanning by $CD8^+$-T cells.

The generation of peptide-MHC II complexes is fundamentally different because the antigens are not of endogenous but exogenous origin. Here, the proteins are taken up from the extracellular milieu by endocytosis, phagocytosis or macropinocytosis (Figure 2). The antigens en-

counter the MHC II in a late endosome/lysosome which is termed MHC class II loading compartment (MIIC). These compartments contain different cysteine-, serine- and aspartate proteases that act as exo- and endopeptidases (cathepsins, Cat, and asparaginyl endopeptidase, AEP) and digest the antigen into small peptides. Additional enzymes like the thiolreductase GILT (Maric *et al.*, 2001) help to unfold the proteins. The MHC II, however, is not empty when it reaches the MIIC. The MHC α- and β-chain associate within minutes during biosynthesis in the ER (Kvist *et al.*, 1982), likely facilitated by interactions in the transmembrane regions. Then, early in biogenesis, the α/β-chain complex associates with the invariant chain (Ii) (Kvist *et al.*, 1982), which acts as a co-chaperone (Stumptner-Cuvelette & Benaroch, 2002). The Ii is a type II transmembrane glycoprotein of 33 kDa that contains a carboxy-terminal trimerization domain and thereby forms nonameric $(\alpha\beta Ii)_3$ complexes (Jasanoff *et al.*, 1998; Roche *et al.*, 1991). Ii interacts with the MHC mainly via an unfolded stretch known as the class II-associated invariant chain peptide (CLIP)-region, which binds into the antigen-binding site like a normal antigen, but also secondary interaction sites have been described (Stumptner-Cuvelette & Benaroch, 2002). An amino-terminal motif in Ii sorts the MHC/Ii complexes into the endocytic pathway (Bakke & Dobberstein, 1990; Lotteau *et al.*, 1990). At the same time binding of Ii to MHC αβ-dimers prevents binding of self-antigens in the ER to the MHC (Teyton *et al.*, 1990; Viville *et al.*, 1993). Leaving the ER through the Golgi apparatus and trans-Golgi network (TGN), there are two routes for MHC/Ii-complexes. They can take the normal pathway to the cell surface, where these complexes are endocytosed in an AP-2- and clathrin-dependent manner and then reach the MIIC (McCormick *et al.*, 2005). Alternatively, the complexes are directly sorted from the TGN into the MIIC (Bikoff *et al.*, 1993; Viville *et al.*, 1993). Either way, the Ii is degraded in a specific, sequential manner by cathepsins. First, from Iip33 (33 kDa fragment) Iip10 is formed, which lacks the trimerization domain. Then, cathepsins S and L make the final cuts and leave only the CLIP fragment in the antigen-binding site. These complexes lack the endosomal retention signal from Ii. Now CLIP can be exchanged for antigens. However, this process *per se* is rather inefficient and depends very much on the affinity of CLIP for the specific allelic variant of the MHC. Naturally, in the MIIC an MHC II-like molecule, HLA-DM, enhances the exchange of CLIP for other higher affine peptides. The exact mechanism by which HLA-DM is mediating this process is not understood yet. But once a stable peptide/MHC complex is formed it is transported to the cell surface. In addition to the classical intracellular antigen loading another pathway has been described where antigens are directly loaded onto MHC II at the cell surface (described in detail later, Figure 2).

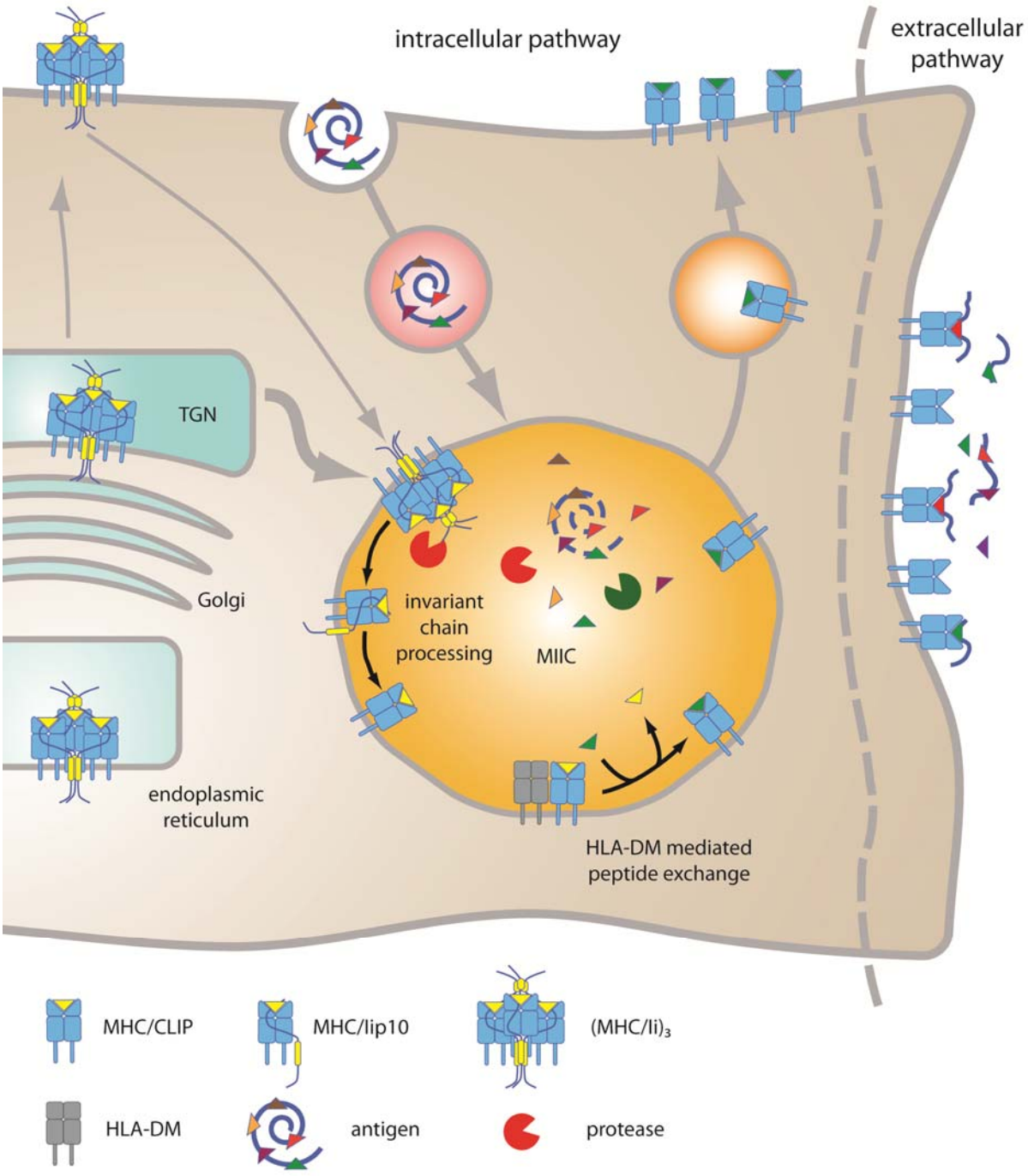


Figure 2: *The MHC II antigen processing and presentation pathway.* The classical, intracellular pathway for the generation of MHC II/peptide molecules is shown schematically. On the right, the additional pathway for direct generation of MHC II/peptide complexes on the cell surface of an antigen presenting cell is depicted. TGN, Trans-Golgi network; MIIC, MHC II loading compartment

As mentioned earlier, MHC II molecules are recycled from the cell surface by AP-2/clathrin-dependent endocytosis. In addition, recently a second ubiquitin-dependent pathway has been described. It was shown that in immature DC the cytosolic tail of the MHC β-chain was ubiquitinated leading to a high turnover of surface-presented pMHC (Ohmura-Hoshino *et al.*,

2006; Shin *et al.*, 2006; van Niel *et al.*, 2006). Upon induction of maturation, e.g. by bacterial lipopolysaccharid (LPS), the DCs down-regulated the ubiquitin ligase MARCH I (De Gassart *et al.*, 2008). This leads to an enhanced life-time of MHC II molecules on the cell surface. In contrast, in immature DCs ubiquitinated MHC are readily degraded upon internalization (Walseng *et al.*, 2010).

1.5 Conformational variants of MHC II

The structure of the peptide-loaded human and murine MHC II is well known (Figure 1). It has been determined by crystallography manifold for different alleles and in complex with different peptides as well as secondary proteins like TCRs, superantigens, viral proteins or its natural co-receptor CD4 (appendix). The overall structure of the MHC II is unchanged. However, there is ample evidence for conformational variants of this uniform structure observed by crystallography.

1.5.1 Ligand-independent conformers

Empty MHC II molecules that have just lost their peptide ligand can rebind new ligands with fast kinetics. However, in the absence of a suitable ligand these molecules turn into a peptide "non-receptive" state, characterized by very slow rebinding of ligands (Rabinowitz *et al.*, 1998). In the MIIC the empty MHC II is stabilized in its receptive form by HLA-DM (Denzin *et al.*, 1996; Grotenbreg *et al.*, 2007). Kinetically, the receptive as well as the non-receptive state are well characterized (Carven *et al.*, 2004; Joshi *et al.*, 2000; Sato *et al.*, 2000; Zarutskie *et al.*, 2001). But very limited information about the underlying structural features of these states is available. Structural studies have been hindered by the short lifetime of the receptive state and by the tendency for aggregation of the non-receptive state. Studies using monoclonal antibodies specific for the empty state of MHC II identified a region on the β-chain (β58-69) that had distinctively different conformations in the empty and peptide-loaded states (Carven *et al.*, 2004; Zarutskie *et al.*, 1999). Also circular dichroism (CD)-spectroscopy and hydrodynamic flow measurements suggested small conformational differences in the empty MHC II (Zarutskie *et al.*, 1999) compared to the peptide-loaded MHC.

1.5.2 Ligand-dependent conformers

In addition to the previously mentioned intracellular loading route, a second pathway has emerged for antigen loading on MHC II. Antigens can be loaded directly on cell surface-

presented MHC II (Figure 2). Especially immature DCs have a high number of empty MHC II on their cell surface that can be loaded extracellularly (Santambrogio *et al.*, 1999a; Santambrogio *et al.*, 1999b). By this route new antigens can be presented that escape the natural procession by proteases in the MIIC and thereby enhance the possibility for a pathogenic antigen to be presented by the APC. At the same time, this bears the risk of uncontrolled presentation of antigens captured from the extracellular milieu as not only pathogenic structures are present but also self-antigens. Importantly, the group of Unanue showed that the immune system can discriminate if an antigen is loaded in the endosomal pathway or directly on the cell surface (Viner *et al.*, 1996). They identified T cells that were reactive either against the extracellularly loaded peptide (type B T cells) or against the same peptide when derived from the antigen by endosomal processing (type A T cells). The corresponding pMHC complexes were termed isomers. These were chemically identical and could be edited by H2-M, the murine homologue of HLA-DM (Pu *et al.*, 2004). Recently, the group showed that both types of T cells can be found in diabetogenic mice in the Non-Obese Diabetic (NOD) mouse model (Mohan *et al.*, 2010). Moreover, type B T cells were able to induce diabetes upon adoptive transfer as was previously predicted (Pu *et al.*, 2002). Also others have found distinctive conformers of single peptide MHCs (Huang *et al.*, 2003; Viret *et al.*, 2003). In addition to the cell biological evidence, first structural data on isomerism was derived by NMR-spectroscopy of murine pMHC II (Schmitt *et al.*, 1999). The factors influencing the formation of these isomers were the location of the MHC/antigen loading, pH, the involvement of H2-M and also the amino-terminal peptide overhang (Lovitch *et al.*, 2006). Altogether, the pMHC recognition by the TCR and activation of the T cell does not only rely on the bound ligand but the T cell also senses the conformation of the MHC.

Strominger and colleagues also reported on the conformational heterogeneity of pMHC II induced by the flanking regions of the antigen (Rotzschke *et al.*, 1999). Amino- or carboxy-terminal extensions of known peptide ligands of HLA-DR1 induced conformational changes, whereby the side of attachment made a distinguishable difference. These conformational changes were reflected in drastically changed electrophoretic migration behavior and hydrodynamic radii. Similar results were also obtained by others using a different peptide ligand (Georges *et al.*, 2000). But nothing is known about the structural differences underlying these proposed conformational changes.

Yet another possible explanation for pMHC isomers derives from the analysis of the antigen-binding site. As described earlier there are no length restrictions for bound peptide ligands as the ends of the MHC II antigen-binding site are open. Therefore, one peptide ligand can bind

in different registers, which likely favors certain registers over others. In fact, Kappler and colleagues proved that the complex of murine I-A^{g7}/β9-23 can bind in three different registers (Stadinski *et al.*, 2010). This is the same epitope that is recognized by diabetogenic T cells in the non-obese diabetic (NOD) mouse model (Mohan *et al.*, 2010). The different registers were generated by rational engineering of disulfide traps into the MHC and peptide or by introducing favorable peptide anchors to enforce a certain register..

Additionally, the two alleles HLA-DR2a (DRB5*0101) and HLA-DR2b (DRB1*1501), both positively associated with multiple sclerosis, induce T cells that are autoreactive against the same peptide from the myelin basic protein (MBP 86-105), when they are in complex with this peptide (Vogt *et al.*, 1994; Wucherpfennig *et al.*, 1994). The crystal structures of both complexes revealed that the same peptide was used in two different registers (Li *et al.*, 2000; Smith *et al.*, 1998). However, the most promiscuous binding peptide is certainly CLIP, for which register shifting has also been described (Fallang *et al.*, 2008). Despite the polymorphism and the resulting heterogeneity of peptide binding motifs, all newly synthesized MHC II are associated with the invariant chain, and more specifically with the CLIP region (Kvist *et al.*, 1982). Although MHC II does not need the invariant chain to reach the cell surface (Bikoff *et al.*, 1993; Viville *et al.*, 1993), it has undoubtedly an important regulatory function in antigen processing. Two crystal structures of the human HLA-DR3 (Ghosh *et al.*, 1995) and the murine I-A^b (Zhu *et al.*, 2003), each in complex with the human CLIP, revealed a uniform pocket occupation in both MHC II. Interestingly, CLIP is also presented on the cell surface of APC and it has a direct immunoregulatory function. It influences the T cell polarization during activation with a bias towards T_H2 cells (Pezeshki *et al.*, 2011; Rohn *et al.*, 2004) when it is presented at the cell surface, here in complex with HLA-DR1.

1.6 Objectives

The pMHC II complex is central for the initiation of an adaptive immune response. Its loading and presentation need to be tightly regulated as a misguided immune response can lead to autoimmunity. Little is known of how the peptide loading state of an MHC II is controlled. Despite the considerable database of MHC II structures (appendix), the conformational dynamics in these protein complexes remain elusive and so far mechanistic evidence is mostly indirect. Here, I want to investigate the structural basis for three different aspects of the conformational flexibility of MHC II

1) Empty MHC II has the tendency to aggregate (Stern & Wiley, 1992). Therefore, the structure of the empty MHC II has not been solved yet by direct methods such as X-ray crystallography. Only recently, molecular dynamics (MD) simulations have emerged as a valuable tool to analyze the structure of MHC II (Gupta *et al.*, 2008; Painter *et al.*, 2008; Yaneva *et al.*, 2008). We therefore planned MD simulations (conducted by Bernd Rupp, FMP Berlin) to model empty MHC II. A first focus of my own work was then to experimentally probe the structural basis for the receptive and non-receptive state of the molecule and to compare *in silico* derived data by *in vitro* experiments using the human MHC II allele HLA-DR1.

2) pMHC isomerism predicts that the peptide loaded MHC II can acquire at least two distinct conformations. While the formation of these conformers seems to depend on the route by which the peptide was loaded on the MHC (Lovitch & Unanue, 2005), their biological function as well as the underlying structural basis are still unknown. Isomerism appears to depend on the size and the precise location of the ligand in the binding groove. An amino-terminal overhang of the peptide ligand was shown to influence the pMHC conformation (Lovitch *et al.*, 2006) and Rötzschke et al. provided evidence that the extension of known core peptide ligands on both termini induces a conformational change in the MHC II (Rotzschke *et al.*, 1999). Based on this work I set out to analyze the structural foundations underlying these observations. The goal was here to provide a direct proof for these conformational variants by solving their X-ray crystal structures.

3) The complex of MHC II with the placeholder peptide CLIP is an important intermediate in the generation of MHC/antigen complexes by the antigen presenting cell. The human HLA-DR1 is widely used to analyze the function of HLA-DM that catalyzes the exchange of CLIP for other antigens. However, the structure of this important intermediate in antigen processing is only inferred from predictions based on complexes with other MHC alleles. The naturally found CLIP is heterogeneous in size and uncertainty of predicted structures derives from possible register shifts or pMHC isomerism. Therefore, the aim was to determine the X-ray structure of HLA-DR1 in complex with two length variants of CLIP.

2 Materials and Methods

2.1 Materials

2.1.1 Chemicals

If not stated otherwise, all chemicals were obtained from Carl Roth (Karlsruhe, Germany) or Sigma-Aldrich (St. Louis, USA).

2.1.2 Plasmids

Plasmids used in the presented work are listed in Table 1.

Table 1*: Plasmids used for expression of HLA-DR1*

Plasmid	Description	Expression host
pET11a::DRa	expression of DRA*0101, aa 1-192	*E. coli*
pET24d::DRb	expression of DRB1*0101, aa 1-198	*E. coli*
pFastBac1::DRa	expression of DRA*0101, aa 1-191, GP64 leader	*S. frugiperda*
pFastBac1::DR1b	expression of DRB1*0101, aa 1-198, GP64 leader	*S. frugiperda*
pFastBac1::DRa:fos	expression of DRA*0101, aa 1-191, GP64 leader, C-terminal linker and fos-zipper for dimerization (plus 45 aa)	*S. frugiperda*
pFastBac1::DR1b:jun	expression of DRB1*0101, aa 1-198, GP64 leader, C-terminal linker and jun-zipper for dimerization (plus 46 aa)	*S. frugiperda*

2.1.3 Cell lines

Table 2 lists the cell lines used in the presented work

Table 2: Overview of cell lines used for cloning and expression of HLA-DR1

Cell line	Species	Description
Sf9	*S. frugiperda*	generation of baculovirus stocks
Sf21	*S. frugiperda*	expression of proteins by baculovirus expression vector system
S2-DM	*D. melanogaster*	cells stably transfected with HLA-DM under Cu-inducible promoter (Sloan *et al.*, 1995)
Top10	*E. coli*	used for cloning; genotype: *mcr*A, Δ(*mrr-hsd*RMS-*mcr*BC), Δ*lac*X74, *deo*R, *rec*A1, *ara*D139Δ(*ara-leu*)7697, *gal*K, *rps*L, *end*A1, *nup*G
DH10Bac	*E. coli*	used for generation of bacmids from pFastBac1 vectors; genotype:.F- *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) φ80*lac*ZΔM15 ΔlacX74 *rec*A1 *end*A1 *ara*D139 Δ(*ara*, *leu*)7697 *gal*U *gal*K λ- *rps*L *nup*G/bMON14272/pMON7124
BL21(DE3)	*E. coli*	used for expression; genotype: F- *ompT hsdS*B (rB-mB-) *gal dcm* (DE3)

2.1.4 Antibodies

Antibodies used in the presented work are listed in Table 3.

Table 3: Overview of antibodies used for purification and detection of MHC II.

Name	Epitope	Source	References
L243	HLA-DR α-chain	in house	(Lampson & Levy, 1980)
LB3.1	HLA-DR α-chain	in house	(Knudsen & Strominger, 1986)
M2	FLAG epitope (DYKDDDDK)	SIGMA Aldrich	(Brizzard *et al.*, 1994)
α-DRab	rabbit serum raised against α- and β-chain	provided by H.L. Ploegh	-
goat-α rabbit	HRP coupled to polyclonal antibody	DAKO	-

2.1.5 Peptides

The peptides used in the presented work are listed in Table 4. If not stated otherwise, all peptides were synthesized by EMC microcollections, Germany.

Table 4: Overview of peptides used for the study

Name	Sequence	MW [g/mol]
HA306-318	PKYVKQNTLKLAT	1504
HA289-318	PNDKPFQNVNKITYGACPKYVKQNTLKLAT	3395
HA306-322	PKYVKQNTLKLATGMRN	1962
HA306-326	PKYVKQNTLKLATGMRNVPEK	2416
HA306-330	PKYVKQNTLKLATGMRNVPEKQTRG	2858
HA-biot	Biot*-Ahx*-SGSG-PKYVKQNTLKLAT	2131
HA-FITC	PKYVK(CF)*QNTLKLAT	1862
CLIP106-120¶	KMRMATPLLMQALPM	1732
CLIP106-120, M107W	KWRMATPLLMQALPM	1787
CLIP106-120, M115W	KMRMATPLLWQALPM	1787
CLIP102-120	KPVSKMRMATPLLMQALPM	2144
CLIP102-120, M107W	KPVSKWRMATPLLMQALPM	2199
CLIP102-120, M115W	KPVSKMRMATPLLWQALPM	2199
CLIP106-120-biot	Biot-Ahx-Ahx-KMRMATPLLMQALPM	2185
TT830-843	QYIKANSKFIGITE	1612
TT1149-1159	DYMYLTNAPSYTN	1553
Bapa361-380	NANIRYVNTGTAPIYNVLPT	2191
Phlp5b68-86	DKFKTFEAAFTSSSKAAAA	1978
p4a293-307	SMRYQSLIPRLVEFF	1886

*) CF: carboxyfluorescein, Ahx: 6-aminohexanoic acid; biot: biotin
¶) numbering according to the longer p35 form of invariant chain (Strubin *et al.*, 1986)

2.1.6 Oligonucleotides

The oligonucleotides used in the presented work are listed in Table 5.

Table 5: Overview of oligonucleotides

Name	Sequence (5'-3' direction)	Description
DR1aQ9A(+)	GAAGAACATGTGATCATCGCGGCCGAGTTCTATCTGAATCC	mutation αQ9A, forward primer
DR1aQ9A(-)	GGATTCAGATAGAACTCGGCCGCGATGATCACATGTTCTTC	mutation αQ9A, reverse primer
DR1aQ9C(+)	GAACATGTGATCATCTGCGCCGAGTTCTATCTG	mutation αQ9C, forward primer
DR1aQ9C(-)	CAGATAGAACTCGGCGCAGATGATCACATGTTC	mutation αQ9C, reverse primer
DR1aE11A(+)	CATGTGATCATCCAGGCCGCCTTCTATCTGAATCCTGAC	mutation αE11A, forward primer
DR1aE11A(-)	GTCAGGATTCAGATAGAAGGCGGCCTGGATGATCACCTG	mutation αE11A, reverse primer
DR1bC15S(+)	GGCAGCTTAAGTTTGAAAGTCATTTCTTCAATGGG	mutation βC15S, forward primer
DR1bC15S(-)	CCCATTGAAGAAATGACTTTCAAACTTAAGCTGCC	mutation βC15S, reverse primer
DR1bR71A(+)	AAGGACCTCCTGGAGCAGGCCCGGGCCGCGGTGGACACC	mutation βR71A, forward primer
DR1bR71A(-)	GGTGTCCACCGCGGCCCGGGCCTGCTCCAGGAGGTCCTT	mutation βR71A, reverse primer
DR1bH81F(+)	GACACCTACTGCAGATTCAACTACGGGGTTGGTGAG	mutation βH81F, forward primer
DR1bH81F(-)	CTCACCAACCCCGTAGTTGAATCTGCAGTAGGTGTC	mutation βH81F, reverse primer
DR1bN82A(+)	GACACCTACTGCAGACACGCCTACGGGGTTGGTGAGAGC	mutation βN82A, forward primer
DR1bN82A(-)	GCTCTCACCAACCCCGTAGGCGTGTCTGCAGTAGGTGAC	mutation βN82A, reverse primer
pT7	TAATACGACTCACTATAGG	sequencing pET vectors, forward
pT7T	TATGCTAGTTATTGCTCAG	sequencing pET vectors, reverse
pH	TTAAAATGATAACCATCTCG	sequencing pFastBac1, forward
HSV Tk	GGTATTGTCTCCTTCCGTG	sequencing pFastBac1, reverse

2.2 General methods

2.2.1 Cloning

DNA-sequences coding for the ectodomains of the α- and β-chain of HLA-DR1 were sub-cloned either from eukaryotic expression vectors containing the full length sequences for DRA*0101 and DRB1*0101 or from insect cell expression vectors containing only the ecto-

domains of the respective chain. These ectodomains are formed by residues 1-191 of the α-chain, respectively residues 1-198 of the β-chain. Therefore, suitable oligonucleotides (section 2.1.6) were used to amplify the corresponding sequences by polymerase chain reaction including attachment of endonuclease restriction sites for later integration into either *E. coli*- or baculovirus-expression vectors. Digestion of amplified sequences and vectors with suitable endonucleases yielded overlapping ends that were used to integrate the sequences into vectors by ligating fragments with T4 DNA Ligase (New England Biolabs, USA).

2.2.2 Site-directed mutagenesis

Point mutations into HLA-DR1 were introduced by site-directed mutagenesis, basically following the QuikChange Site-Directed Mutagenesis protocol from Stratagene, LaJolla, CA, USA. Oligonucleotides were designed as complementary pairs with one or two base pair mismatches according to the desired mutation with a length between 25 and 45 nucleotides and a melting temperature between 70-80°C. The PCR mix with plasmid DNA, oligonucleotides, dNTPs, and DNA-polymerase was prepared according to manufacturer's advice. If the recommended standard protocol for the thermo cycler did not work, a "touchdown PCR" strategy was used (for details see Table 6).

Afterwards, the parental plasmid was digested by treatment with DpnI and the reaction mixture was transformed into *E. coli* strain Top10.

Table 6: Cycling parameters for site-directed mutagenesis PCR

	Standard protocol		**Touchdown protocol**		
	Temperature [°C]	**Time [s]**	**Temperature [°C]**	**Time [s]**	
Initial denaturation	95	30	95	30	
Denaturation	95	30	95	30	*)first cycle at T_m+5°C, thereafter 10 cycles with -1°C/cycle, finally 20 cycles at T_m
Annealing	55	60	T_m+5*	30	
Elongation	72	300	72	600	
Final extension			72	600	

2.2.3 DNA purification

DNA was purified using standard kits from Qiagen, Hilden, Germany, that are based on the rapid alkaline extraction method (Birnboim & Doly, 1979) in combination with anion-exchange purification of DNA. Bacmid-DNA was usually not purified with kits but with standard alkaline extraction followed by phenol-chloroform extraction and subsequent EtOH-precipitation of DNA.

2.2.4 Sequencing

All cloned vectors and introduced mutations were checked for errors by sequencing using appropriate primers (Table 5) by the company Seqlab, Göttingen, Germany.

2.2.5 Protein expression in baculovirus-infected insect cells

HLA-DR1 was initially expressed in insect cells with the baculovirus expression system, generally following the Bac-to-Bac system (Invitrogen, UK). The ectodomains of HLA-DR1 α- and β-chain (amino acids 1-191, resp. 1-198 of the mature protein) were cloned in separate pFastBac vectors both under control of the polyhedrin promoter and with the leader sequence of the viral envelope glycoprotein 64 (GP64). For the expression of the mutant forms of HLA-DR1 the complex formation was enhanced by addition of leucine zipper motifs to the C-terminus of the chains (fos-zipper to the DR1α-chain, jun-zipper to the DRβ-chain), as described (Fourneau *et al.*, 2004).

For virus stock generation Sf9 cells were infected at $1*10^6$ cells/mL with an MOI of ~10 and virus stocks were harvested after six to seven days. For protein expression Sf21 cells in the logarithmic growth phase at $2\text{-}3*10^6$ cells/mL were double-infected with viruses for the α- and β-chain at an MOI of 1-10. The secreted protein was harvested from the supernatant after 72-84 h by centrifugation followed by sterile filtration. Insect cells were grown in Sf900 medium containing penicillin and streptomycin in flasks or bottles in a shaking incubator at 100 rpm and 27°C.

2.2.6 Protein expression in *E. coli*

HLA-DR1 cannot be expressed as soluble heterodimer in *E. coli*. Therefore, a published refolding procedure (Frayser *et al.*, 1999) was adapted by Andreas Schlundt, FMP Berlin (Schlundt *et al.*, 2009). The ectodomains of the α-chain (HLA-DRA*0101), comprising residues 1-192, were cloned into the pET11a vector. The ectodomains of the β-chain (HLA-DRB1*0101, residues 1-198), were cloned into expression vector pET24d. The chains were then separately expressed in the *E. coli* strain BL21(DE3). Expression cultures in LB-medium were inoculated and grown at 37°C to an OD600 of 0.5-0.8 and expression was induced by

addition of 1 mM IPTG. Expression cultures were grown overnight at 30°C. Cells were then harvested by centrifugation.

2.2.7 Protein expression in drosophila cells

Drosophila melanogaster S2 cells stably expressing HLA-DM under the control of the metallothionein promoter (Bunch *et al.*, 1988; Sloan *et al.*, 1995) were grown in suspension culture in Sf900 medium (Invitrogen, UK), supplemented with 5 % FCS and penicillin (100 U/mL), streptomycin (100 µg/mL) and amphotericin B (0.25 µg/mL) (Invitrogen, UK) in spinner flasks at 28°C. Alternatively, cells were also cultivated in shaker flasks (28°C, 100 rpm). Cells were held at a density of 4 to $20*10^6$ cells/mL to allow for logarithmic cell growth. Protein expression was induced by addition of 1 mM $CuSO_4$ at a cell density of $5*10^6$ cells/mL. The cells were further cultivated before the secreted protein was harvested from the supernatant 96 h later.

2.2.8 Protein purification of HLA-DR1 from baculovirus-infected insect cells

HLA-DR1 expressed as secreted form in baculovirus-infected insect cells was purified by immunoaffinity chromatography. The HLA-DR1 containing supernatant was repeatedly passed over a column containing the HLA-DR1-specific antibody LB3.1 coupled to Sepharose-4FastFlow (GE Healthcare, USA). Contaminations that bind unspecifically to the column material were removed by incorporating a precolumn with inactivated Sepharose-4FastFlow with no coupled antibody. The column was first washed with at least 10 column volumes (CV) PBS, pH 7.4 and subsequently with 10 CV 10 mM Na_2HPO_4, pH 7.0. The protein was eluted using a basic buffer (50 mM CAPS, pH 11.5). Protein elution was either carried out with a Bio-Rad BioLogic FPLC system (Bio-Rad, USA) monitoring the absorption at 280nm or by collecting fractions and, later on, determining the protein containing aliquots by Bradford assay. The protein was immediately dialyzed against PBS, pH 7.4 and after sterile filtration stored at 4°C. The column was washed again with 10 CV PBS, pH 7.4 and 10 CV 10 mM Na_2HPO_4, pH 7.0 before reapplication of the remaining cell culture supernatant. Depending on the volume of the supernatant and the protein expression level, this purification cycle was repeated four times. For long term storage the column buffer was exchanged to PBS, pH 7.4 containing 0.1 % (w/v) NaN3.

2.2.9 Protein purification of HLA-DR1 from *E. coli*

Harvested cells from expression were resuspended in 50 mM Tris, 1mM EDTA, 735 mM sucrose, 10 mM DTT, 0.1 % NaN_3, pH 8.0. Cell lysis was initiated by addition of 2.5fold volume of 20 mM Tris, 1 % Triton, 1 % deoxycholat, 0.1 % NaN_3, 100 mM NaCl, 0.3 mg/mL lysozyme, 10 mM DTT, pH 7.5 and stirred for 30 min at RT. Addition of 20 µg/mL DNase1 and 0.5 mM $MgCl_2$ digested DNA. A subsequent freeze/thaw cycle enhanced cell lysis. Inclusion bodies were then separated by centrifugation (20 min, 10000g) and the pellet resuspended in 50 mM Tris, 0.5 % Triton, 100 mM NaCl, 0.1 % NaN_3, 1 mM EDTA, 10 mM DTT, pH 8.0. The pellet was resuspended by the help of a sonicator or dounce homogenizer. Centrifugation and resuspension cycles were repeated three times plus twice without Triton. Finally, inclusion bodies were solubilized in 20 mM Tris, 8 M urea, 0.5 mM EDTA and 10 mM DTT, pH 8.0 overnight. Remaining insoluble parts were removed by ultracentrifugation for 1 h at 30000 g and RT.

The individual chains were then purified by anion-exchange chromatography in a denatured state using a MonoQ column (GE Healthcare, USA) and 20 mM Tris, 8 M urea, 1 mM DTT, pH 8.0 as buffer system with a salt gradient up to 500 mM NaCl. Fractions containing purified chains were pooled, 0.5 mM EDTA was added and pH adjusted to 8.0. Samples were flash frozen and stored at -80°C or directly used for refolding.

For reconstitution of the heterodimer complex the individually purified and denatured chains were refolded together in degassed 20 mM Tris, 25 % (v/v) glycerol, 0.5 mM EDTA, 3 mM GSH, 0.3 mM GSSG, pH 8.0 at a concentration of 10 mg/L for each chain. For direct generation of peptide loaded MHC complexes a 10 fold molar excess of peptide was added to the refolding buffer. After 4 days at 4°C the refolded MHC was concentrated using an ultrafiltration unit (Vivascience/Sartorius AG, Göttingen) and dialyzed against PBS, pH 7.4. Subsequently, the protein was further purified by immunoaffinity chromatography using the α-HLA-DR antibody LB3.1 coupled to Sepharose-4FastFow (GE Healthcare, USA) as described for the insect cell-derived HLA-DR1, except that no precolumn was used. Elutions were repeated until no significant amount of protein eluted from the column.

2.2.10 Protein purification of HLA-DM from insect cells

The supernatant from the expression in S2-cells was centrifuged and sterile filtered before application to columns with the α-FLAG antibody M2 coupled to agarose beads (Sigma-Aldrich, St. Louis, MO, USA). The supernatant was recycled over the column at 4°C overnight and washed afterwards with 10-20 column volumes PBS, pH 7.4. The protein was eluted by either adding 0.15 mg/mL FLAG-peptide (sequence: DYKDDDDK) in PBS, pH 7.4 or by acidic elution with 0.1 M glycine, pH 3.5. Protein elution was either monitored by connecting the column to a Bio-Rad BioLogic FPLC system (Bio-Rad, USA) and online observation of the optical density at 280 nm or small fractions were collected unmonitored and subsequently checked for protein concentration. Individual fractions were directly neutralized by addition of 1 M Tris, pH 8.0. The protein containing fractions were pooled and dialyzed against PBS, pH 7.4 overnight and samples were finally sterile filtered and stored at 4°C.

2.2.11 Generation of antibody column for HLA-DR1 purification

The affinity chromatography column for the purification of HLA-DR1 was prepared by immobilizing the murine α-DR antibody LB3.1 on sepharose. N-hydroxylsuccinimde (NHS)-activated Sepharose 4 Fast Flow (GE healthcare, USA) was first activated by the addition of 1 mM HCl. After soft centrifugation the supernatant was removed and the antibody, in PBS, pH 7.4, was added. 4 to 5 mg LB3.1 was used per mL beads. The coupling reaction was incubated under constant moving overnight at 4°C. If necessary, the reaction was continued the next day at RT until the coupling efficiency was >95 %. After removal of unbound antibodies, the remaining activated NHS-groups were inactivated by addition of 200 mM ethanolamine, pH 8. Finally, the beads were washed with PBS, 0.1 % (m/v) NaN_3 was added for preservation and the beads were stored at 4°C until usage. For preclearing of proteins unspecifically interacting with the sepharose material, beads were prepared without antibody but direct inactivation by ethanolamine.

The antibody was derived from in house production using the B cell hybridoma cell line LB3.1. It was purified from the cell culture supernatant by precipitation with $(NH_4)_2SO_4$ and subsequent affinity-purification using protein G-sepharose (GE Healthcare, USA).

2.2.12 SDS-PAGE

For estimation of protein purity and MHC-peptide complex formation samples were run on a 12 or 15 % SDS-PAGE. SDS-sample buffer (with or without β-mercaptoethanol) was added and samples were boiled for 5 min unless peptide-complex formation was to be observed. In these cases samples were directly applied to the gel.

After electrophoresis the gel was stained either by silver or coomassie staining, depending on the applied protein amount.

2.2.13 Western Blot analysis

For cross-linking experiments of HLA-DR1 mutants Sf21 cells were double-infected with α- and β-chain variants. Cells were harvested 48 h postinfection and lysed in 50 mM Tris, pH 8.0, 150 mM NaCl, 1 % NP40 and 20 mM NEM. For reduction of samples 100 mM DTT was added. 1 µg total protein was then separated on a 12 % SDS-PAGE and subsequently transferred by wet blotting to a polyvinyldifluoride membrane (Immobilon P, Millipore, USA). Therefore, the membrane was briefly soaked in methanol and then placed on top of the SDS-PAGE gel. On each side two layers of Whatman paper and a sponge on the outer side were added. Everything was clamped together and placed in a wet tank blotting system (Bio-Rad, USA) filled with transfer buffer (25 mM Tris, 192 mM glycine, 20 % (v/v) methanol) for 60 min at 60V. Afterwards the membrane was blocked with 5% skim milk in PBS/0.1 % Tween20 (PBST) for 1 h at RT on a shaker. After two washing steps with PBST the primary antibody diluted in 5 % skim milk was added usually at 4°C overnight. MHC was detected with rabbit serum raised against the HLA-DRα- and β-chain as described (Rotzschke *et al.*, 1997). Unbound antibody was removed by washing three times with PBST. Finally, a goat-α-rabbit HRP-conjugate antibody (DAKO, Denmark) was used as secondary antibody. After washing three times with PBST chemiluminescent detection solution was added (ECL Plus, GE healthcare, USA) and the membrane was exposed to film (Kodak, USA) and developed.

2.2.14 Analytical size exclusion chromatography

For analysis of the hydrodynamic radius of peptide loaded HLA-DR1 empty MHC II was loaded with at least 50 fold molar excess of peptide ligand and the addition of 1 mM AdEtOH as loading enhancer. After overnight incubation at 37°C complexes were applied to a Super-

dex200 PC 3.2/30 column (GE Healthcare, USA) connected to a Pharmacia SMART HPLC System (GE Healthcare, USA) equilibrated with PBS, pH 7.4, at a flow rate of 40 µL/min. For size estimation the column was calibrated using standard proteins in the range of 12 to 200 kDa (Sigma-Aldrich, USA).

2.2.15 ELISA (loading and competition)

Peptide loading was analyzed by detection of bound biotinylated peptide ligands and soluble, empty HLA-DR1 either from insect cells or *E. coli* in an enzyme-linked immunosorbent assay (ELISA).

Maxisorb 96 or 384 well plates (Nunc, Denmark) were coated with 1 µg/mL α-HLA-DR antibody L243 (in house production) in 100 mM bicarbonate buffer, pH 8.3 overnight at 4°C or for 2 h at 37°C. Unbound antibody was removed by washing twice with PBS, 0.05 % Tween20 (PBST). Plates were then blocked with PBS/2 % BSA for 2h at 37°C and subsequently washed four times with PBST. The loading reaction of HLA-DR1 and biotinylated peptide was stopped by diluting with cold PBS/1 % BSA and, subsequently, split and distributed on the antibody-coated plate and incubated for 1.5 h at 4°C. Unbound protein was removed by washing six times with PBST and then 200 ng/mL Eu^{3+}-labeled streptavidin (Perkin Elmer, USA) in PBS/1 % BSA was added and incubated for 30 min at RT. The plate was washed eight times with PBST and finally, enhancer solution was added (15 µM β-naphtoyltrifluoroacetone, 50 µM tri-n-octylphosphine oxide, 6.8 mM potassium hydrogen phthalate, 100 mM acetic acid, 0.1 % Triton X-100). The plate was then read out in a Victor fluorescence reader (Perkin Elmer, USA) using the time-resolved mode with an excitation wavelength of 340 nm and an emission wavelength of 614 nm.

For competition experiments 100 nM HLA-DR1 was incubated with 100 nM biotinylated HA306-318 and increasing amounts of unlabeled competitor peptide. The reaction mix was incubated for 2 to 3 d at 37°C and then further processed as described above. The data was fitted to a four parameter logistic function to determine the effective concentration (EC50) at which the binding of the labeled peptide was inhibited by 50 % (the Hill slope gives the steepness of the curve):

$$y = \min + \frac{(\max - \min)}{1 + \left(\frac{x}{EC50}\right)^{-Hillslope}}$$

2.2.16 Fluorescence polarization

A FITC-tagged HA peptide was used to determine the real time binding of a model antigen to HLA-DR1 by fluorescence polarization. When a fluorophore is excited with polarized light, the light emitted from this fluorophore is polarized as well. The degree of polarization is measured by the fluorescence intensities parallel ($F_{\parallel}$) and perpendicular ($F_{\perp}$) to the plane of the linearly polarized excitation light. The polarization (P) is then calculated as:

$$P = \frac{F_{\parallel} - F_{\perp}}{F_{\parallel} + F_{\perp}}.$$

The degree of polarization is highly dependent on the mobility of the fluorophore during the excitation and emission of the polarized light, with a high mobility leading to enhanced polarization loss in the direction of the exciting light. Here, FITC coupled to HA306-318 was used as probe. When the peptide (and the fluorophore) binds to HLA-DR1, the tumbling of the whole complex is greatly reduced as compared to the free peptide/fluorophore and the loss of polarization is greatly reduced.

1 µM HLA-DR1 was incubated with 100 nM FITC labeled HA306-318 in a non-binding black microplate (Corning, USA). Measurements were conducted in a Victor 3V reader (Perkin Elmer, USA) at 37°C. In between measurements the plate was kept in the dark at 37°C. The binding rate was calculated as observed rate constant at the given concentration by fitting data points to a monomodal association model ($FP=FP_{max}(1-e^{-kt})$) (Buchli *et al.*, 2004).

2.2.17 Thermofluor assay

Thermal stability of MHC/peptide complexes was assessed by a differential scanning fluorimetry assay (Niesen *et al.*, 2007), more commonly known as thermofluor assay (Ericsson et al, 2006; Pantoliano et al, 2001). In this assay the protein is slowly heated in the presence of an environment-sensitive dye. Here, SyproOrange (Invitrogen, UK) was used. This dye has a low quantum yield in solvents with high dielectric constants, such as water. But in solvents with low dielectric constants, such as organic solvents, or in the environment of an unfolded protein this dye is highly fluorescent. The excitation lies at 490 nm and the emission maximum at 575 nm, this can be detected with a standard filter sets in real time thermocycler like the iCycler iQ Real Time Detection System (Biorad, USA) or the StepOnePlus real-time PCR system

(Applied Bioystems, Carlsbad, CA, USA). For *a posteriori* peptide loaded HLA-DR1 samples it was mandatory to remove aggregated protein by gel filtration prior to the experiment to reduce the background signal from the aggregated protein. The final protein concentration was 0.2-0.5 mg/mL and 5x SyproOrange in 25 µL total reaction volume. The assay was conducted in 96 multi well PCR plates. The samples were heated from 25-95°C with an increase of 1°C/min. The fluorescence intensity was plotted against the temperature and the data were fitted to a sigmoidal curve:

$$FI = FI_0 + \frac{FI_{\max} - FI_{\min}}{1 + \left(\frac{T}{T_m}\right)^a},$$

where FI is the fluorescence intensity, T_m the unfolding transition midpoint, and *a* a slope factor. The transition mid point was defined as melting point T_m.

2.3 X-ray crystallography

2.3.1 Introduction

Today X-ray crystallography is still the most widely used technique to gather information on the three-dimensional structure of molecules with 90 % of all structures in the Protein Data Bank (PDB) determined by this technique. The theoretical foundations for this technique will briefly be given in this chapter before the practical application is described in the following sections.

The prerequisite for X-ray crystallography is the growth of highly ordered protein crystals in which the protein of interest is arranged in an ordered and repetitive manner. The crystal is build up from a unit cell defined by three edges (a, b, c) and three angles (α, β, γ). In case the molecules within the unit cell possess internal symmetry, the cell can be further divided into the asymmetric unit. Vice versa, the whole content of the unit cell can be generated by applying symmetry operations given by the crystal's space group to the motif in the asymmetric unit. When the crystal is exposed to a monochromatic, electromagnetic wave (here: X-rays), the crystal behaves like a lattice and the incoming wave is scattered. In the diffraction pattern each spot corresponds to the reflection of the incident beam by one set of parallel lattice planes of the crystal. William L. Bragg developed the following formula that describes the conditions for the spacing (d_{hkl}) between the lattice planes, the diffraction angle (θ) and the

wavelength (λ) that has to be fulfilled for constructive interference (n denotes an integer and gives the order of reflection):

$$2d_{hkl} \sin\theta = n\lambda \, .$$

Each diffraction spot can be assigned to a specific set of lattice planes in the crystal at which it was reflected. The orientation of these planes relative to the unit cell and the interplanar spacing can precisely be described with the Miller indices *h, k, l*. An additional helpful concept is the reciprocal lattice derived from the real space lattice together with the Ewald sphere construction and Bragg's law as depicted in Figure 3. It follows that diffraction can be observed when a reciprocal lattice point *hkl* (here: $\bar{1}01$) intersects with the Ewald sphere with the radius of 1/λ. The reciprocal lattice vector $d^*_{\bar{1}01}$ is collinear and equal in magnitude with the scattering vector S. Rotation of the crystal and thus the reciprocal lattice enables the sampling of the reciprocal space as different reciprocal lattice points will intersect with the Ewald sphere. The diffraction pattern will give a picture of the reciprocal lattice. Assignment of the Miller indices to individual reflections, i.e. from which real space lattice plane the according diffracted beam was reflected, enables the determination of the geometry of the unit cell and symmetry by analysis of the symmetry and spacing of the reflections. Information about the structure within the unit cell can be derived from the distribution of intensities of the reflections.

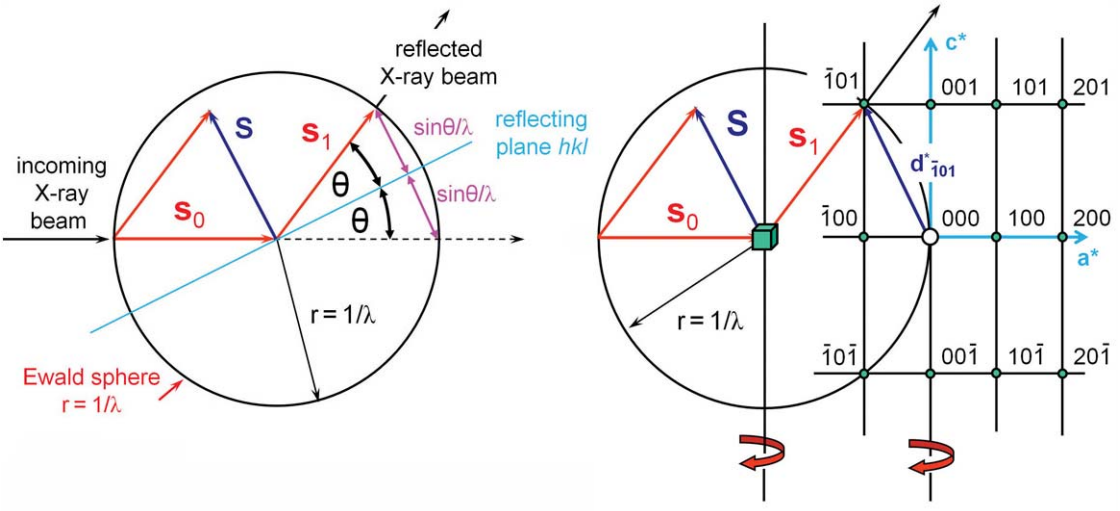


Figure 3: *Construction of the Ewald sphere and the reciprocal lattice.* Left, The Ewald sphere allows the graphical interpretation of diffraction in relation to the reciprocal lattice vector *hkl*. Right, The origin of the reciprocal lattice is set on the intersection of the Ewald sphere with the undiffracted X-ray beam. When a reciprocal lattice vector (here shown for $d^*_{\bar{1}01}$) lies on the Ewald sphere, the Bragg condition for diffraction is fulfilled (because $d^*_{hkl} = 1/d_{hkl}$. Rotation of the crystal (and thus the reciprocal lattice) brings more lattice points to intersection with the Ewald sphere leading to diffraction (picture adapted from (Rupp, 2010)).

The intensities are proportional to the square of the structure factor amplitudes:

$$I(hkl) \propto |F(hkl)|^2 .$$

The structure factor $F(hkl)$ is a complex function and the sum of all atomic scattering contributions in a unit cell in the direction defined by the Miller indices *hkl*. It can be written as

$$F_{hkl} = |F(hkl)| \cdot e^{i\alpha_{hkl}} ,$$

where $|F(hkl)|$ is the structure factor amplitude and α_{hkl} the phase of the wave function. The electron density in the unit cell may now be derived by an inverse Fourier summation over all combinations of h, k, and l:

$$\rho(x,y,z) = \frac{1}{V} \cdot \sum_{hkl} |F(hkl)| \cdot e^{ia(hkl)} \cdot e^{-2\pi(hx+ly+kz)} ,$$

x, y, z are the fractional coordinates of the atoms in the unit cell and V is its volume. As stated earlier the amplitude of this wave function ($|F(hkl)|$) can be measured in form of the intensities of the reflections, whereas the phase angle $\alpha(hkl)$ cannot. This is the so called "crystallographic phase problem". In the following section possible solutions to this problem will be introduced.

2.3.2 Solving the phase problem

There are several methods to obtain the phase information needed to calculate the electron density from the observed diffraction pattern. They will be briefly summarized and the method of molecular replacement will be discussed in more detail as it was used exclusively during this thesis.

For the first set of methods no prior structural information is needed. For these experimental phasing methods the specific incorporation of heavy atoms is needed and initially the substructure of these atoms is solved, whose complexity is significantly reduced as compared to the entire structure. Here, basically two methods can be discriminated.

Nowadays, the most popular direct method uses anomalous diffraction. Here, atoms are incorporated into the protein (e.g. by substituting selenomethionine for methionine) that show anomalous scattering at certain energies of the X-ray. These resonance effects add up to the

atomic scattering factor and the imaginary component leads to the breakdown of the reciprocal lattice centrosymmetry, i.e. the Friedel's law $I(hkl) = I(-h-k-l)$ is not true any more. These differences can be used to determine the coordinates of the heavy atom substructure. It can be distinguished between measurements at a single wavelength (single-wavelength anomalous diffraction, SAD) or multiple wavelengths (multi-wavelength anomalous diffraction, MAD). For these methods in principle only one well-diffracting crystal is needed.

Traditionally, isomorphous replacement was used to obtain initial phases for structure solutions. Here, heavy atoms are specifically bound to the protein and at least two datasets, one for the heavy atom derivative and one for the native protein, are recorded (single isomorphous replacement, SIR). Unambiguous solutions are obtained when more than one heavy atom derivative is generated (multiple isomorphous replacement, MIR). These methods can also be combined with the anomalous scattering (AS) effects of the heavy atoms (SIRAS/MIRAS). The prerequisite for these methods, however, is that the crystals are unchanged by the derivatization.

However, molecular replacement is the most frequently used technique to solve the phase problem. Here, the model of a structurally similar protein is needed to obtain initial phases. The task is to place the model in the correct orientation at the right position in the unit cell of the new crystal. This can be broken down in two separate problems, for which the Patterson function is used. This function can be directly derived from the measured intensities and does not depend on phase information. It gives the interatomic distances between all atoms in the unit cell. To obtain the correct orientation of the model, its Patterson map is first rotated to match the experimentally derived Patterson map as closely as possible (rotation search). In a second step, the best solutions are subjected to a translation search. Additionally, these solutions can be used to calculate structure factor amplitudes $|F_{calc}|$ and be compared to the experimentally derived $|F_{obs}|$ using the crystallographic R-factor (see below). The model of the best solution based on Patterson search and structure factor amplitudes is used to calculate initial phases α_{calc} and an initial electron density map.

2.3.3 Model building and refinement

Once initial phases are obtained by either of the above mentioned methods, an electron density map can be calculated and the initial model can be manipulated in real space using appro-

priate software. Usually, two different types of electron density are used. These difference maps are the $2F_o$-F_c map, calculated as

$$\rho(xyz) = \frac{1}{V}\sum_{hkl}\left\|2\cdot F_{obs}\right| - \left|F_{calc}\right\| e^{-2\pi i(hx+ky+lz)+i\alpha_{calc}}$$

and the F_o-F_c map, calculated as

$$\rho(xyz) = \frac{1}{V}\sum_{hkl}\left\|\cdot F_{obs}\right| - \left|F_{calc}\right\| e^{-2\pi i(hx+ky+lz)+i\alpha_{calc}} \ .$$

In the first map the bias introduced by the model is reduced by weighting the experimental data more strongly. In the latter map those parts of the model can be easily identified that still have to be built (positive difference density peaks) and those which are wrong (negative difference density peaks).

Once the model has manually been optimized based on the observations in the electron density maps (the real space), the coordinates of the model can be further improved by computational refinement in the reciprocal space. The refinement is an optimization task that aims at minimizing the difference between $\left|F_{obs}\right|$ and $\left|F_{calc}\right|$. Basically, two mathematical strategies are followed: the least-square minimization and the maximum-likelihood refinement. The measure for the fit of experimental to calculated data is the R factor:

$$R = \frac{\sum_{hkl}\left\|F_{obs}\right| - \left|F_{calc}\right\|}{\sum_{hkl}\left|F_{obs}\right|} \ .$$

One problem in macromolecular X-ray crystallography is the low observation to parameter ratio. To improve this, prior knowledge of the macromolecular stereochemistry in form of restraints for the refinement can be employed. In addition to the optimization of the coordinates also their dynamic mobility can be refined as positional displacement around a mean position. The B-factor describes the statistical motion of an atom with an isotropic sphere r:

$$B = 8\pi^2 r^2 \ .$$

To avoid overfitting of the parameters to the data, a second R-factor was introduced by A. Brünger (Brunger, 1992). For the R_{free} usually 5 % randomly selected reflections are not used in the refinement process. Overfitting of parameters can then be observed by an increase in the R_{free} factor, whereas the R factor calculated from the remaining reflections (R_{work}) is still reduced. Further criteria for the quality of the refined model are parameters that were not used

in the refinement process, e.g. the root mean square deviations (RMSD) from standard bond angle, lengths and planarity and the torsion angles between the peptide planes at the Cα atom of an amino acid. The latter are restricted to certain values, which is the basis for the Ramachandran plot.

2.3.4 Crystallization

For crystallization of individual HLA-DR1/peptide complexes empty MHC was loaded with a 10 to 20 fold molar excess of free peptide in the presence of 1 mM AdEtOH (Höpner *et al.*, 2006) or 2 mM Ac-FR-NH2 (Gupta *et al.*, 2008) as molecular loading enhancer overnight at 37°C. In some cases the peptide was already added during refolding of HLA-DR1. Unbound peptide was removed by size exclusion chromatography using a Superdex 200 column (GE Healthcare, USA) into 20 mM MES, 50 mM NaCl, pH 6.4. This buffer was chosen for the optimized stability of the peptide-MHC complex. A screen for the optimal stability of HLA-DR1/HA306-330 in different buffer systems and pH was conducted in a thermofluor-based assay as described (Ericsson *et al.*, 2006) (data not shown). The maximal stability was in the range of pH 5.6-6.8. The protein was concentrated to 10 mg/mL by ultrafiltration.

An analysis of the reported crystallization conditions for published MHC II crystal structures revealed that almost all conditions included a polyethylene glycol (PEG) variant as main precipitant. Therefore, for initial screening of crystallization conditions a focus was set on commercially available sparse matrix or systematic screens with a high number of PEG conditions, like the PACT, PEG, and PEGII screens (Qiagen, Germany). But depending on availability of protein also the sparse matrix screens Classics, Classics II, Protein Complex, and JCSG+ were tested. Initial screens were conducted in the vapor diffusion sitting drop set up. Drops with a size of 0.4-1 µL (containing an 1:1 mixture of protein and buffer) were prepared using a Hydra® II Plus One microdispensing system (Thermo Scientific, USA) in low profile 96 well CrystalQuick plates (Greiner Bio-One, Deutschland). The plates were stored in a RT (HomeBase, The Automation Partnership, USA) or 4°C (Rock Imager 500, Formulatrix, USA) storage system. Plate observation was automatically realized using incorporated CCD-camera systems in a defined time schedule. The resulting images were used for annotation with the online observation tool “Observation”, developed in the group of U. Heinemann, MDC Berlin, Germany.

Single crystals grown to a sufficient size were harvested with a nylon loop and generally cryo-preserved by transfer to a drop containing the corresponding precipitant and 20 % glycerol prior to flash freezing in liquid nitrogen.

2.3.5 Application

In the following section, an overview about the general steps leading from data collection to structure determination is given as it was employed throughout this thesis.

All data were collected at synchrotron source BESSY II at Helmholtz-Center, Berlin, Germany at beamlines 14.1 and 14.2 with the oscillation method at 100 K.

For initial assessment of crystal quality two frames lying 90° apart were collected. For crystals of sufficient quality the strategy for data collection yielding complete data sets was developed by determining the crystal symmetry using the program MOSFLM (Leslie, 2006). After data collection the program XDS (Kabsch, 2010) was used to first index all recorded reflections with their according Miller indices and then to integrate the detector derived pixel information to intensities. The quality of the data was then evaluated by the R_{sym}-factor calculated as:

$$R_{sym} = \frac{\sum_{hkl}\sum_{i}\left|I_i - \langle I \rangle\right|}{\sum_{hkl}\sum_{i}\left|I_i\right|},$$

where I_i is the intensity of an independent reflection and $\langle I \rangle$ the average of the independent reflections with the indices hkl. In the next step, the data was scaled with the program XSCALE (Kabsch, 2010), i.e. individual images that contained the same reflection *hkl* (and in case of a native data set its Friedel mate -*h*-*k*-*l*) were scaled so that this reflection has the same intensity on all images. Also Bijvoet pairs (reflections related by crystallographic symmetry) should have the same intensity. Moreover, scaling can also compensate for further errors introduced during measurement e.g. by X-ray absorption effects or radiation damage. The scaled data were further processed by the program XDSCONV (Kabsch, 2010) to obtain the suitable reflection input file format for later work with the CCP4 (Winn *et al.*, 2011) or PHENIX (Adams *et al.*, 2010) program suites.

The general structure of MHC II molecules and particular HLA-DR1 has been determined manifold. In the presented work the only variable in these structures was the peptide bound to

the MHC. Therefore, no experimental methods were needed to obtain initial phase information for structure solution. Instead, the molecular replacement program PHASER (McCoy *et al.*, 2007) was used with HLA-DR1 (PDB 2G9H (Fernandez *et al.*, 2006)) as search model, only using the coordinates for the MHC α- and β-chain and not the bound peptide. In all cases PHASER gave clear solutions and initial, interpretable electron density maps were readily available and clear difference electron density was observed for the missing peptide. The model was further refined with Refmac5 (Murshudov *et al.*, 1997) or PHENIX (Adams *et al.*, 2010). After each refinement round the electron density map was inspected with COOT (Emsley *et al.*, 2010) and the missing or wrong parts of the model were manually built or rebuilt. In later refinement rounds also the peptide was built into the clear difference electron density. During the end of the refinement process also missing water molecules and other solvent molecules were added. The quality of the model was checked during the refinement using the validation tools within COOT or the PHENIX suite, which are based on MolProbity (Davis *et al.*, 2007).

3 Results

3.1 Conformation of the non-receptive MHC II

As outlined before, the structure of the empty MHC II has not been resolved yet, neither by X-ray crystallography nor NMR spectroscopy. To gain insight into the structure and intrinsic dynamics of the empty, non-receptive MHC, molecular dynamics calculations (MD) were carried out by Bernd Rupp (FMP, Berlin). Based on this data a model for the transition of the empty MHC from the receptive to the non-receptive form was developed. This model was tested experimentally as described in the following chapter.

3.1.1 Model of the non-receptive MHC by MD

Three different starting models were used to conduct MD simulations using the GROMACS software suite (Van Der Spoel *et al.*, 2005). The model of the MHC was surrounded by a water box to simulate solvent. As peptide-loaded structure either HLA-DR1/HA306-318 (Stern *et al.*, 1994) or HLA-DR1/GAG166-181 (Zavala-Ruiz *et al.*, 2004) were used. The empty MHC was generated *in silico* by removing the peptide. As reference for the receptive state, a dipeptide-loaded HLA-DR1 was included ("P1-stabilized"), in which the phenyl side chain of the dipeptide FR was docked into the prominent binding pocket 1 (P1) of the empty molecule. It has been shown that occupation of P1 by these dipeptides renders the empty MHC II highly receptive for peptide loading (Gupta *et al.*, 2008). The different models were simulated for up to 30 ns to allow for conformational convergence. Simulations were also repeated to validate individual results (the empty state was calculated nine times, the two peptide loaded structures each twice and the P1-stabilized structure three times). Figure 4 summarizes the results of these MD simulations.

While the flexibility in the membrane-proximal α2- and β2-domains shows only negligible differences in mobility between the three models, the situation for the α1- and β1-domains forming the antigen-binding site is strikingly different (compare Figure 4A and B). As expected, the lowest flexibility was observed for the peptide-loaded MHC. The bound peptide significantly enhanced the stability of the antigen-binding site, so that only minor shifts in the structure were detected. The P1-stabilized molecule shows an intermediate behavior. While the antigen-binding site remains open (Figure 5B), it exhibits enhanced flexibility that is mostly confined to the ends of the helices flanking the binding site and loop regions. The most drastic transitions were observed for the empty MHC. Here a contiguous region in the β-chain α-helix (β64-77) was identified as a hot spot of flexibility. This region includes the link be-

tween two α-helical parts of the β-chain and thus might function as a hinge in the β-chain α-helix. A closer inspection of the MD simulations revealed that the flexibility in the hinge region allows the β-chain helix to acquire a straightened and presumably energetically favored conformation, which resulted in the closure of the antigen binding cleft (Figure 5A). In contrast to the differential mobility in the α1/β1-domains, the mobility in the membrane-proximal α2/β2-domains was comparable in all simulations, independent of the loading state of the MHC II. Moreover, the mobility of the β2-loop regions is in agreement with the flexibility observed in the different crystal structures of MHC II, as judged by higher B-factors or conformational variability between structures.

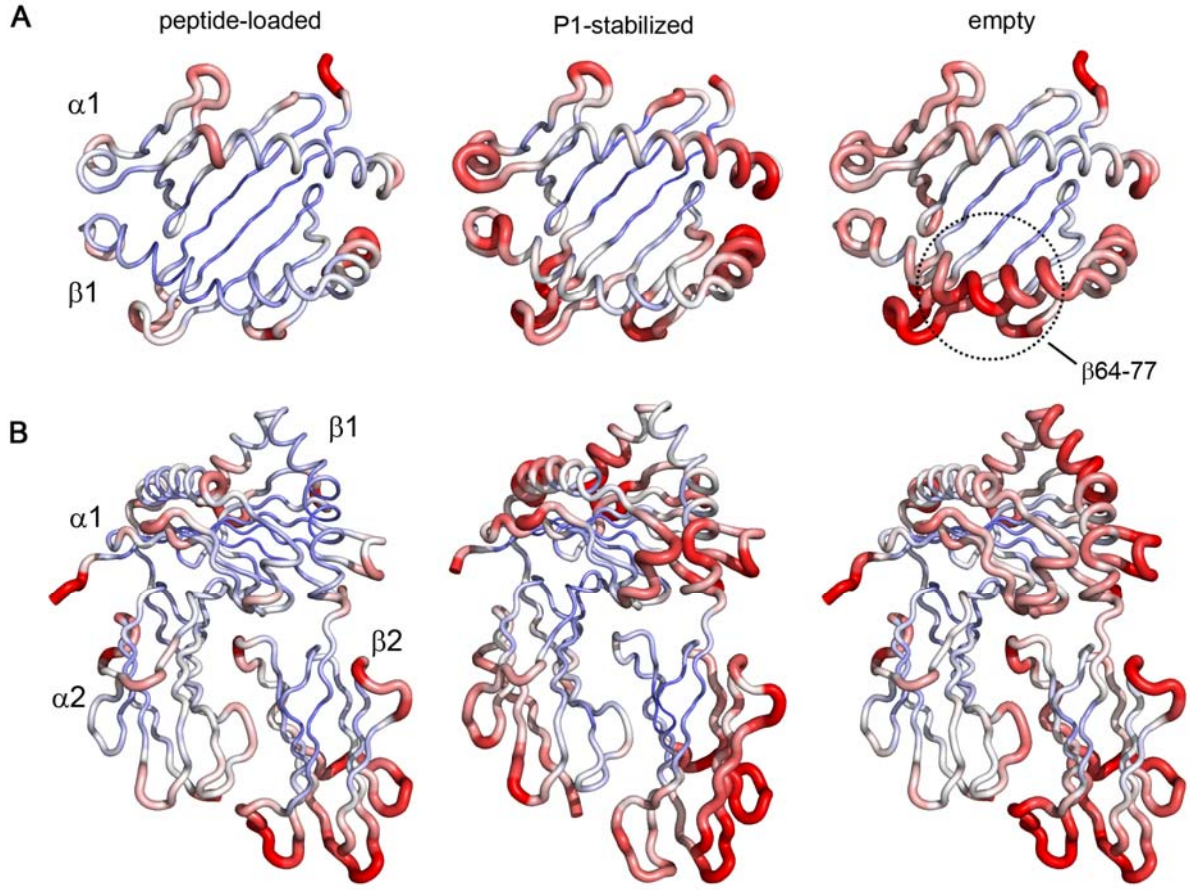


Figure 4: *MD simulations suggest a flexible hinge region as major point of conformational change in empty MHC II.* Three different starting conformations of HLA-DR1 were analyzed by MD simulations and compared: peptide loaded (two different peptides), P1-stablizied (with dipeptide) and empty HLA-DR1. The observed flexibility in the antigen-binding site in the different MD simulations for the three model systems was summarized by plotting the root mean square deviation (RMSD) of the Cα-atoms onto the HLA-DR1 structure. Peptide ligand or dipeptide (in the P1-stabilized structure) has been removed for clarity. The color code from blue to red represents RMSD values from 0 to ≤ 6 Å, the thickness varies accordingly. **A**, Only the α1/β1-domain is shown. The region with the highest RMSD values in the simulations of the empty HLA-DR1 is indicated. The bound peptide and dipeptide are not shown for clarity. **B**, shows the entire ectodomains of HLA-DR1 as used in the MD simulations. MD simulations were carried out by B. Rupp, FMP Berlin.

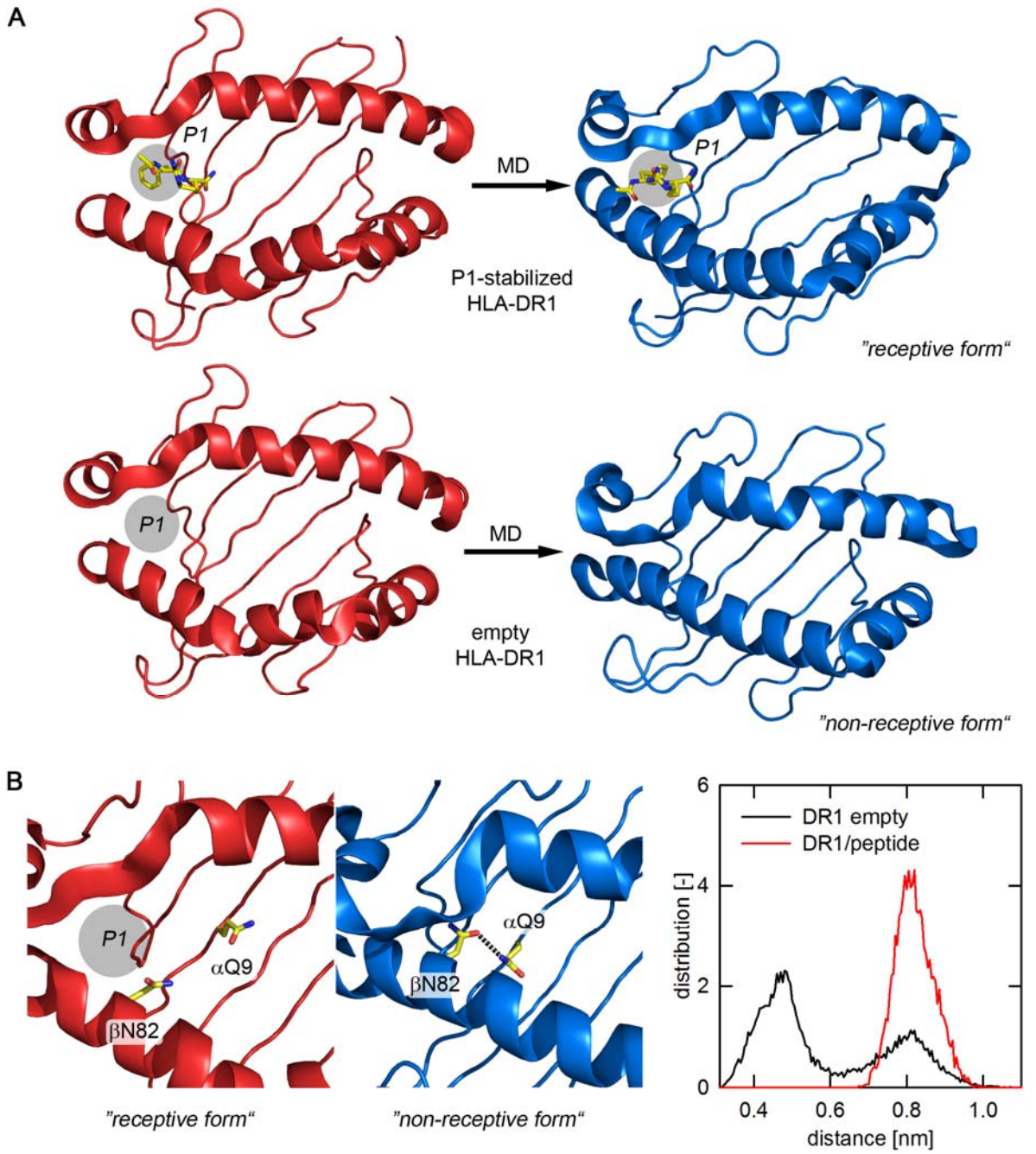


Figure 5: *Antigen-binding site closure upon loss of antigen is stabilized by newly established hydrogen bond.* **A**, In the P1-stabilized, receptive state of empty MHC the antigen-binding site remains open (upper panel), whereas in the empty, non-receptive state the binding site is closed mainly by movements in the β-chain α-helix (lower panel). Cartoon representations of the α1/β1-domain of HLA-DR1, starting structures for MD simulations are red, equilibrated end structures are blue. Representative frames are shown. **B**, A hydrogen bond is formed upon closure of the binding site. Left panels show the position of residues αQ9 and βN82 at the start and end of the MD simulations. A distribution plot for the distances between the center of the carboxamides of αQ9 and βN82 in the MD simulations demonstrates that both residues come only in the empty MHC close enough to form a hydrogen bond (right panel). The position of the P1 pocket is indicated. MD simulations were carried out by B. Rupp, FMP Berlin. Figure taken from (Rupp *et al.*, 2011).

In a next step, a more detailed analysis revealed some key residues that appeared crucial for the stabilization of the closed conformation. In the loaded complex, these residues are involved in peptide binding but are free for new intra-MHC interactions in the peptide-free

form. A new hydrogen bond between αQ9 and βN82 is formed upon closure of the antigen-binding site (Figure 5B). This new interaction links the β-chain α-helix to the floor of the antigen binding cleft. Concomitant with this interaction is the straightening of the β-chain α-helix, a probably energetically favored conformation. These movements result in a narrowing of the distance between the two helices flanking the P1 pocket, which now becomes inaccessible for other peptide ligands (Figure 5A). In the antigen-bound form the interaction between αQ9 and βN82 is prevented by the bound peptide ligand. In contrast, in the P1-stabilized model the occupation of the P1 pocket with the docked dipeptide prevents efficient formation of the αQ9/βN82-lock observed in the peptide-free simulations. The narrowing of the flanking helices is not observed and interestingly the whole binding site remains open (Figure 5A). These *in silico*-derived analyses of the mechanism behind the transition from the receptive to the non-receptive state of empty MHC II were the basis for functional studies *in vitro* as described in the following chapters.

3.1.2 Validation of helix flexibility by introduction of intramolecular disulfide bond

The MD simulations predicted a major movement of the β-chain α-helix in the region from β64-77 (Figure 4A and Figure 5A). To validate this model by experimental data, the flexibility of the helices flanking the antigen-binding site was tested. Based on the MD simulations a disulfide bond was engineered into HLA-DR1 that would cross-link both chains. The positions for the introduction of the cross-linking cysteines was chosen so that the residues were only close enough to form the disulfide bond if the predictions were correct (Figure 6A). Therefore, the existing disulfide bridge between βC15 and βC79 was broken by mutating βC15 to serine and a new cysteine was introduced on the α-chain by the mutation of αQ9 to cysteine. Only in the equilibrated, non-receptive model the Cα-distance between αQ9 and βC79 is sufficient for the formation of a disulfide bond (Sowdhamini *et al.*, 1989).

The mutated proteins were expressed in baculovirus-infected Sf21 cells. After separation by SDS-PAGE the proteins were detected by Western Blot using a polyclonal α-HLA-DR rabbit serum (Figure 6B). When the samples were not reduced prior to loading on the SDS gel, the α- and β-chain of wt HLA-DR1 ran as separate subunits, whereas the mutated HLA-DR1 (αQ9C/ βC15S) ran as single band with approximately double the size of the subunits (MW α/β-chain: 45 kDa, α-chain: 22 kDa, β-chain: 23 kDa). Importantly, when the samples were

reduced before loading onto the gel, also the cross-linked heterodimer separated into its subunits showing that the cross-linking was reversible. The higher molecular weight band observed in the reduced samples between 46 and 58 kDa is likely an endogenous insect cell protein because it was also present in samples of uninfected insect cells (data not shown). The multiple bands observed for the monomers especially in the non-reduced sample derived from heterogeneous glycosylation of the proteins. Although the formation of the disulfide bonds is not a very sensitive measure for the calculated distances in the simulations, it was a first indication of the validity of the model. However, further detailed analysis of the atomic model for the conformational transition was necessary.

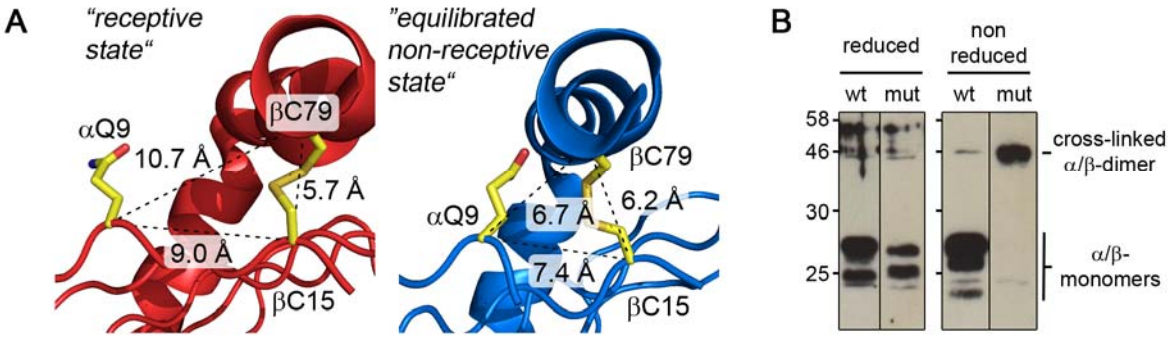


Figure 6: *Introduction of intramolecular disulfide bond for validation of postulated flexibility in MHC antigen-binding site.* **A**, Position of the residues chosen for engineering of the intramolecular disulfide bridge in the empty starting structure of the MD simulations (red) and in the equilibrated end structure (blue). The residues are shown as yellow sticks. The distances between the Cα-atom positions are indicated. **B**, Western blot analysis of the engineered proteins. A comparison between the wt and mutant (αQ9C/βC15S) HLA-DR1 as expressed in baculovirus-infected insect cells. Protein was detected with polyclonal rabbit anti-serum raised against HLA-DR1α/β-chain. A molecular weight marker is indicated in kDa. Figure taken from (Rupp *et al.*, 2011).

3.1.3 Analysis of MD simulations reveals individual amino acids essential for maintaining the non-receptive state

A closer analysis of the conformational changes observed in the MD simulations suggested that two amino acids are critically involved in the closure of the empty antigen-binding site (Figure 5). The residues αQ9 and βN82 form a hydrogen bond which stabilizes the inward movement of the β-chain α-helix. To test this hypothesis, these amino acids were mutated to alanine and the protein was expressed as soluble variant without transmembrane regions in the baculovirus expression system and subjected to functional analysis.

As starting point the two centrally involved residues αQ9 and βN82 were mutated to alanine. In addition to these residues, βH81 was mutated to phenylalanine as an isosteric form of histidine lacking its hydrogen-bonding capacity. βH81 has previously been reported to be crucial in the HLA-DM mediated ligand release (Narayan *et al.*, 2007). Moreover, also stabilization of the closed MHC II conformation by this residues was predicted by others (Painter *et al.*, 2008). In contrast to these reports, however, in our model this residue is constantly exposed to the solvent and therefore cannot participate in stabilization of the empty MHC II.

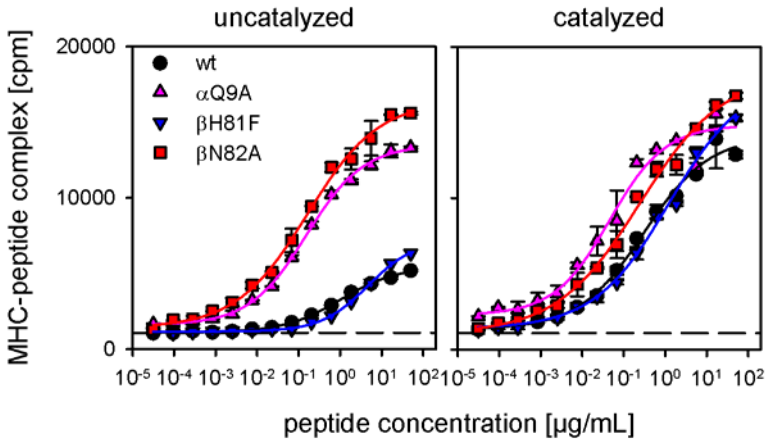


Figure 7: *Breakage of central hydrogen bond increases peptide loading capacity drastically.* In an ELISA-based loading assay empty HLA-DR1 wt and indicated mutants were loaded with biotinylated HA306-318. A clear difference between the uncatalyzed experiment (left panel) and that catalyzed with 1 mM AdEtOH (right panel) is observable. HLA-DR1 was derived from baculovirus infected insect cells. Figure taken from (Rupp *et al.*, 2011).

The ability of the various HLA-DR1 variants to bind peptide antigen was tested in an ELISA-based peptide loading assay using a biotinylated viral peptide antigen (HA306-318), known to bind HLA-DR1 with high affinity. Empty wt HLA-DR1 is expected to be in the thermodynamically stable non-receptive state, evident by the observed inefficient loading of peptide antigen without catalysis (Figure 7, left panel). Contrary to this, the removal of either αQ9 or βN82 drastically increases the amount of formed peptide MHC complexes during the 2 h incubation period. The concentration needed to reach half-maximal loading of antigen was reduced at least 10^2 to 10^3 times, confirming their putative function in stabilizing the closed conformation. No enhancing effect was observed for βH81F. Removing the hydrogen-bonding capacity of βH81 did not increase the peptide loading efficiency of this mutant, which was in contrast to the function suggested by a recent report (Narayan *et al.*, 2007). Importantly, the receptiveness of all proteins could be restored when the P1-targeting MLE AdEtOH (Höpner *et al.*, 2006) was used to catalyze the loading reaction. This included also the βH81F variant, indicating that all mutants are in principle able to load the peptide antigen.

The differences in the uncatalyzed loading reaction were largely overcome when using the MLE.

In the peptide-loaded complex αQ9 and βH81 form one, βN82 even two hydrogen bonds to the peptide ligand. An alternative explanation for the experimental observation described above may therefore be that the increase in peptide loading could be caused by differences in the release of prebound ligand, which may be more easily exchanged in the mutants than the wt form. Sadegh-Nasseri and colleagues found that at least a fraction of HLA-DR1 from baculovirus-infected insect cells is occupied with a peptide from the host cell (Hartman *et al.*, 2010). To exclude this possibility, the same experiment was repeated with HLA-DR1 molecules expressed in *E. coli*. Because these are expressed as inclusion bodies and both subchains are individually purified in a denatured state and only subsequently refolded together to reconstitute the whole MHC, these proteins are devoid of any copurified peptide ligands (Frayser *et al.*, 1999). Thus, by following this approach truly empty MHC molecules could be used in the following functional experiments.

The wt and all variants described above were recloned into bacterial expression vectors and expressed in *E. coli*. In addition, to further discriminate our model from previously reported MD simulations, two more mutations were included (αE11A and βR71A). βR71 has been reported by Painter *et al.* to be responsible for the stabilization of a segment that unfolds from the α-helix of the α-chain. In their model this segment is occupying the P1 pocket area (Painter *et al.*, 2008). In contrast, in our model this unfolding is not observed. In line with this the mutation of βR71 to alanine also did not affect the loading rate, suggesting that this residue plays no role in the stabilization of the non-receptive state (Figure 8A). αE11, instead, was not implicated in Painter's model, but in our model it is an integral part of a hydrogen bond network with αQ9 and αD66 (Figure 8A). Based on this, a mutation of αE11 is expected to have a profound influence on the electrostatics in this region of the antigen-binding site, which has already been shown for this network in the murine HLA-DR homologue I-E^k (Wilson *et al.*, 2001).

Using the *E. coli*-derived material, the results from the insect cell-derived material could be verified (compare Figure 8B to Figure 7), indicating that the release of contaminating ligands plays only a minor role. Mutations αQ9A and βN82A rendered the empty HLA-DR1 highly susceptible for peptide loading. Similar to the βH81F substitution, also the mutation βR71A had only little influence on the receptiveness of empty HLA-DR1. In contrast, the impact of αE11A on the peptide loading ability was comparable to mutants αQ9A and βN82A, further

supporting our model. Also here, the functionality of the proteins was shown by inducing peptide loading with AdEtOH.

Because the ELISA-based assay only allows for discontinuous assessment of peptide loading, the same mutants were used in a fluorescence polarization assay using FITC-tagged HA306-318 peptide as model antigen. This assay allows real time measurement of the loading reaction. Confirming the data obtained by ELISA, also in this setting mutant βN82A has the highest receptiveness (time for half-maximal loading ($t_{1/2}$) with about 0.2 h), followed by αE11A and αQ9A (0.7 and 1.3 h, respectively, see Figure 8C) compared to 12 h to reach $t_{1/2}$ for the wt. βH81F had a roughly two times faster loading rate ($t_{1/2}$=5.5 h), whereas βR71A was also in the wt range (15.5 h). More details are given in Table 7.

Table 7: Loading rate and half-maximal loading time of HLA-DR1 wt and mutants loaded with HA306-318 as determined by fluorescence polarization.

HLA-DR1	**Rate [1/h]**	**$t_{1/2}$ [h]**
wt	0.06 ± 0.01	12.2 ± 3.0
αQ9A	0.54 ± 0.05	1.3 ± 0.1
αE11A	1.05 ± 0.1	0.7 ± 0.1
βR71A	0.07 ± 0.01	15.5 ± 1.6
βH81F	0.13 ± 0.02	5.5 ±0.9
βN82A	3.83 ± 0.66	0.2 ± 0.03

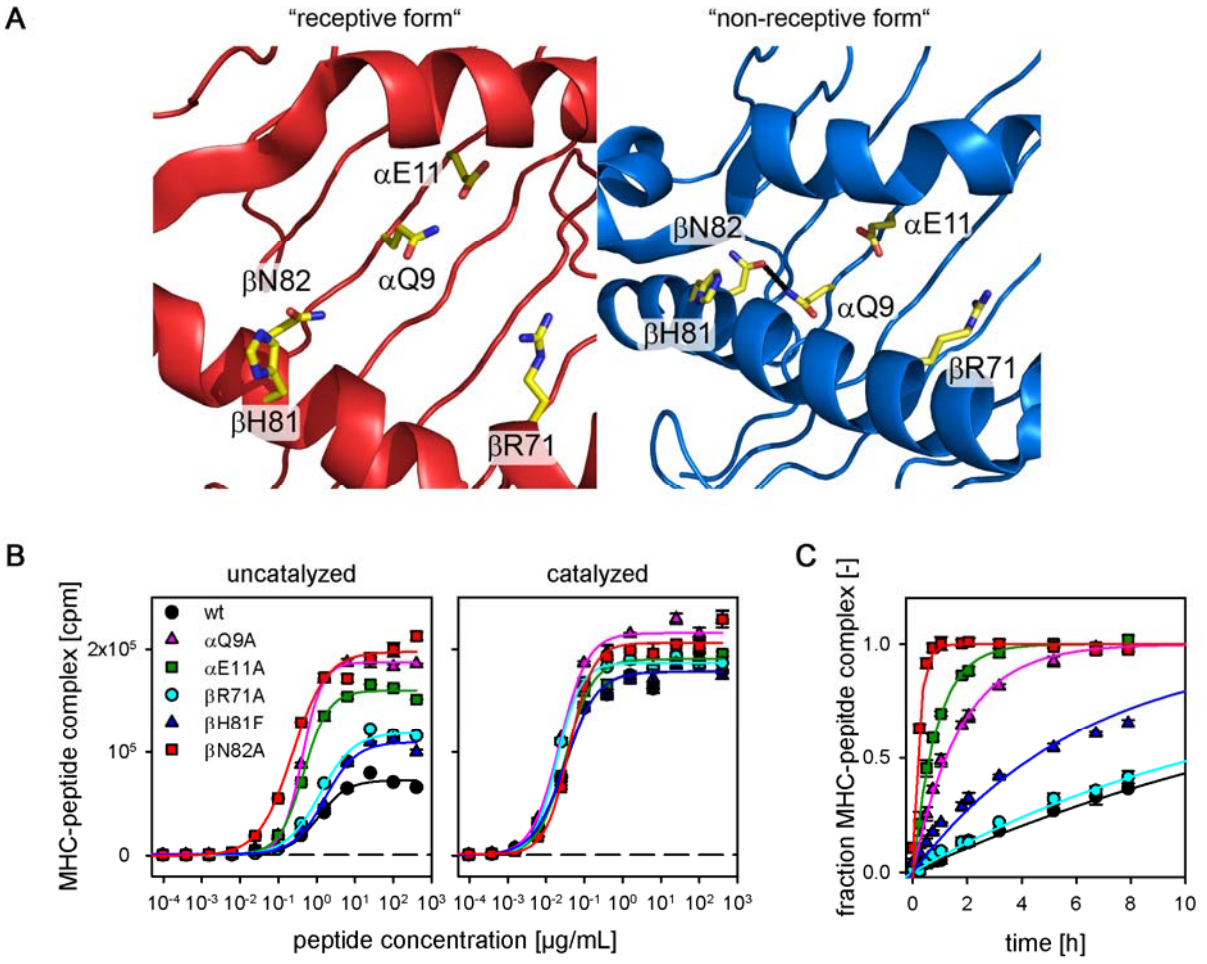


Figure 8: *Model of the residues critically involved in the stabilization of the non-receptive state of MHC II.* **A**, Representative frames from the MD simulations of the empty HLA-DR1 from the start (left panel, red) and an equilibrated end structure (right panel, blue) show the position of amino acids analyzed in *in vitro* experiments as stick model. The hydrogen bond between αQ9 and βN82 is indicated. **B**, ELISA-based loading assay using mutant versions of empty HLA-DR1 and biotinylated HA306-318. MHC II preparations from *E. coli* were used (left panel: uncatalyzed loading reaction, right panel: catalyzed with 1 mM AdEtOH to induce receptive state in empty HLA-DR1 wt and mutants. C, Fluorescence polarization (FP) assay allows continuous observation of loading reaction with HLA-DR1 and FITC tagged HA306-318. Data was corrected for background polarization and normalized to maximal signal for each protein. Figure taken from (Rupp *et al.*, 2011).

Finally, the mutants were characterized by their thermal stability in a thermofluor assay (Figure 9) (Ericsson *et al.*, 2006). In this assay HLA-DR1 is gradually heated in the presence of the environment-sensitive dye SyproOrange (Figure 9A). During thermal unfolding, hydrophobic regions become exposed and the fluorescence emission is increased allowing to determine the melting point of the protein. At higher temperatures the protein aggregates and precipitates, giving rise to a decreased signal. Generally, the empty MHC samples showed a higher background signal which is likely caused by the exposure of the hydrophobic binding cleft in the absence of peptide ligand. Removing hydrogen bonding capacity of any of the three amino acids centrally involved in the stabilization of the non-receptive form (αQ9,

αE11 and βN82) reduces the stability of the MHC by 14-15°C as compared to the wt HLA-DR1 (T_m = 66°C) (summary Figure 9B). The contribution of βR71 or βH81 to the overall stability of the empty MHC was much smaller (T_m = 63°C). The addition of HA306-318 as peptide ligand increased the stability for all variants. However, the gain in stability was not equal for all. The binding of HA306-318 to wt HLA-DR1 increased its stability by 17°C. The measured stability for the wt HLA-DR1 in its empty and also in its HA306-318 loaded form is in good correlation with values measured by other techniques, such as circular dichroism spectroscopy (Sato *et al.*, 2000). The gain in stability for the two mutants not directly affecting the locking-mechanism of the non-receptive state (βR71A and βH81F) was in a similar range (ΔT_m 13-15°C). αQ9A and βN82A showed the lowest overall as well as relative increase in stability (ΔT_m 7°C, resp. 3°C). The discrepancy between the contribution of individual amino acids to the overall stability of the peptide-loaded MHC II has been shown previously (McFarland *et al.*, 2001).

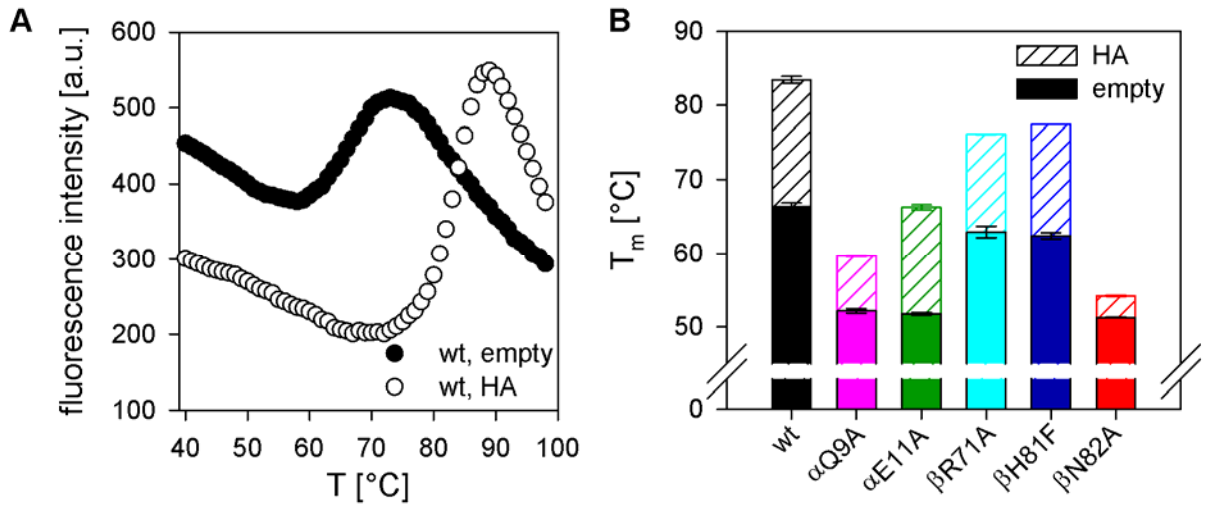


Figure 9: *Removal of critical hydrogen bonds destabilizes peptide loaded and empty MHC II.* **A**, Thermal shift assay is shown for HLA-DR1 wt empty (black dots) and loaded with HA306-318 (white dots). The stabilization by the bound peptide is seen by the delay in thermal unfolding. **B**, Summary of thermal shift assay measurements for HLA-DR1 wt and mutants reveals the distinct influence of single amino acid exchanges on the overall stability of empty (filled bars) and HA306-318 loaded (shaded bars) HLA-DR1. Experiment was conducted in PBS, pH 7.4, average of 4 (empty) and 2 (HA306-318 loaded) experiments ± S.D.

The analysis of the thermal stability indicated already different stability of peptide-loaded mutant MHC complexes. Another measure for the stability of these complexes is also the ability to exchange a preloaded peptide for another one, thereby simulating the situation inside the cell when CLIP is exchanged for another antigen. Figure 10 shows the exchange of a biotin-labeled CLIP bound to HLA-DR1 against an excess of unlabeled HA306-318. Over the course of the experiment (68 h) little or a comparatively low amount of peptide was exchanged for

the wt, the βR71A and βH81F mutants. In contrast, the three other mutants (αQ9A, αE11A and βN82A) showed a complete exchange at the end of the experiment even in the absence of HLA-DM as natural peptide exchange factor. This is in line with the low thermal stability of the latter three mutants. A previous publication implied that the H-bond formed by βH81 would be disrupted by HLA-DM and thereby renders the peptide-MHC instable. Ablation of this bond by mutagenesis would leave the MHC in a "post-DM" conformation (Narayan *et al.*, 2007). However, this could not be validated either in the presented work (Figure 10) or by others (Ferrante & Gorski, 2009; Zhou *et al.*, 2009). Moreover, in line with a published study (Zhou *et al.*, 2009) the HLA-DM mediated peptide exchange is not abrogated by any of the conserved hydrogen bonds analyzed here.

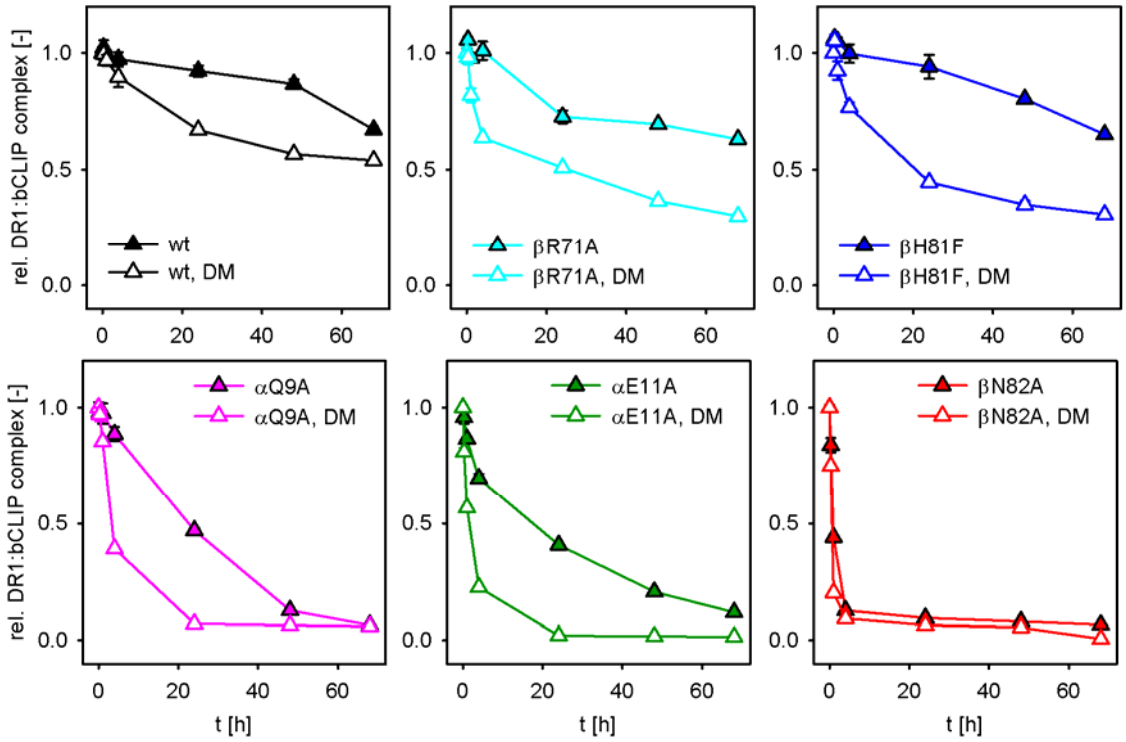


Figure 10: *Destabilized HLA-DR1 mutants are still susceptible to HLA-DM mediated ligand release.* HLA-DR1 wt and mutants were preloaded with biotinylated CLIP106-120 and subsequently exchanged for unlabeled HA306-318 in the presence or absence of the natural peptide exchange factor HLA-DM at pH 5.4. A representative experiment is shown. Values are the mean of duplicates and the experiment was repeated twice. Error bars indicate data range.

3.2 Peptide-induced conformational variants

MHC II molecules do not only display conformational flexibility in their empty state, but also in their peptide-loaded forms. Rötzschke et al. described the effect of peptide-ligand extensions on the conformation of MHC II, in particular on the human HLA-DR1 (Rotzschke *et al.*, 1999). Here, this effect was analyzed in more detail.

3.2.1 The effects of peptide ligand extensions on the conformation of HLA-DR1

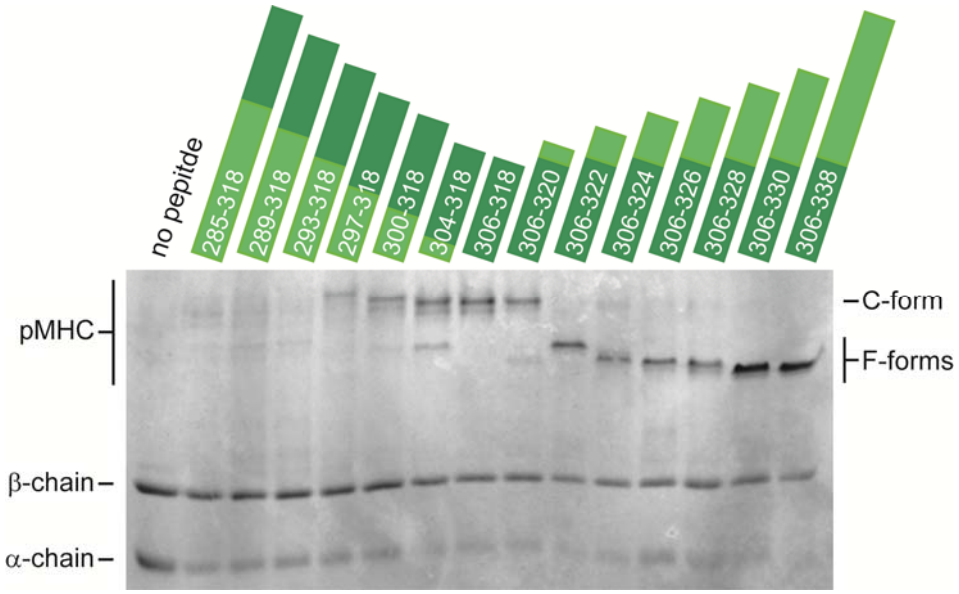


Figure 11: *SDS-PAGE of HLA-DR1 loaded with length variants of the HA306-318 core peptide.* Successive extension of either N- or C-terminal peptide overhang drastically influences the migration behavior and SDS-stability of peptide-MHC complexes. Empty HLA-DR1 produced in *E. coli* was loaded with the indicated HA-peptides. Samples were separated by SDS-PAGE (12 % gel). The lower two bands represent the subunits of HLA-DR1. The upper band corresponds to the peptide-MHC complex. Samples were not boiled prior to loading on the gel. The green bars are scaled representations of the extensions (light green) in relation to the core peptide (HA306-318, dark green). The gel was coomassie stained.

One indication for the conformational change is observable by the significantly disturbed SDS-PAGE migration behavior of peptide-MHC complexes. Certain peptides can form SDS-stable complexes when bound to MHC II, as long as they are not boiled. These complexes can be analyzed by SDS-PAGE. Optimal occupation of the large P1 pocket with an aromatic anchor has been described as one critical factor (Natarajan *et al.*, 1999). Rötzschke et al. showed that N- or C-terminal extension of known core peptide epitopes induces an MHC conformation which runs significantly faster on an SDS-PAGE, despite of the increased molecular weight. These complexes were termed F-conformer (for “floppy” or “fast migrating”) as op-

posed to the C-conformer (for "compact") of the core epitopes. In their study Rötzschke et al. used dimerized versions of known peptide ligands as well as extensions based on the natural sequence of these ligands, leading to two distinct F-conformers. Based on truncations of tetraethylene glycol (teg)-linked peptide dimers, the minimum extension necessary to induce the F-conformation was estimated to be 5-10 amino acids starting at a distance of 5-10 amino acids at either side of the 13mer peptide. Here, this requirement was tested within a more natural setting using ligands consisting of amino acids only and based on the natural sequence of the HA peptide antigen. Figure 11 illustrates the influence of successive extensions of the peptide overhang on the migration behavior of the peptide/MHC II complex. Empty HLA-DR1 was loaded with excessive amounts of different length variants of the original HA306-318 core epitope extending to the N- or C-terminal side. To avoid unwanted dimerization of peptide ligands, Cys305 in N-terminally extended peptides was replaced by serine. The general observation of Rötzschke and colleagues could be verified, the longer the overhang the faster the peptide-MHC complexes migrated in the gel. The highest band corresponds to the previously described C-form (Rotzschke *et al.*, 1999). In addition to the known F-forms, distinct bands at intermediate sizes were found for each peptide length variant. The first F-form was observed for HA306-322 (C-terminal overhang $+6_C$; in the following the overhang is counted starting from the residue before P1 at the N-terminus, Tyr308, or after P9 at the C-terminus, Leu316). The most extreme form was found for HA306-330 (overhang $+14_C$) with no further change for longer overhangs. Thus, the minimum required C-terminal overhang for the induction of the F-conformation was $+6_C$. In contrast, the behavior of the N-terminal extended peptide-MHC complexes was different. Initial extensions led to slight increase in apparent size up to HA297-318 (overhang $+11_N$). But beyond this point, the peptide-MHC complexes had a largely decreased SDS-stability. Moreover, a faster migrating complex was already found for HA304-318 (overhang $+4_N$) and longer peptides, which might correspond to the previously described F-form. But also this suffered from SDS-instability.

Table 8: Size exclusion chromatographic analysis of HLA-DR1/HA complexes. The elution volume was measured twice on an analytical Supderdex200 column in PBS, pH 7.4. HLA-DR1 is from *E. coli* preparations.

	Elution volume [mL]	
Peptide	**1. Run**	**2. Run**
HA306-318	1.577	1.579
HA306-322	1.577	1.577
HA306-326	1.565	1.563
HA306-330	1.559	1.561
HA289-318	1.567	1.567

As an additional indicator for the conformational change, the hydrodynamic radius of the pMHC complexes was measured by analytical gel filtration. In average the elution volume of the pMHC loaded with longer HA peptide variants was smaller than of the C-form, reflecting a larger apparent hydrodynamic radius (Table 8 and Figure 10A). This trend was in agreement with native PAGE and analytical gel filtration experiments reported by Rötzschke et al., although the effect on the hydrodynamic radius seemed to be more pronounced in their study (Rotzschke *et al.*, 1999).

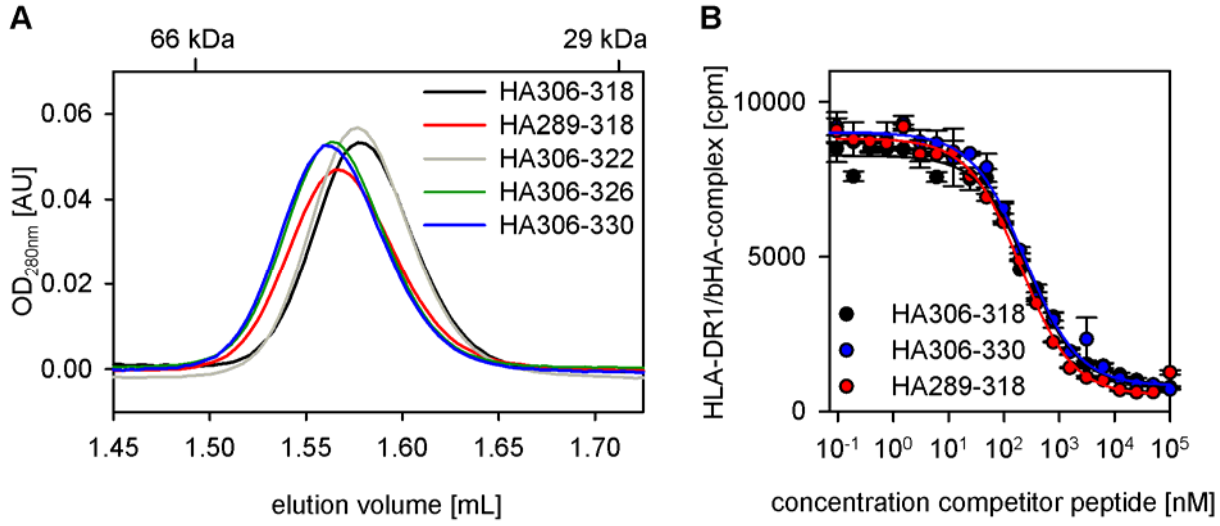


Figure 12: *Analysis of HLA-DR1 complexes with N- and C-terminal extended version of HA306-318.* **A**, Analytical gel filtration of peptide MHC complexes. The longer the peptide overhang, the earlier the peptide MHC complexes eluted from the size exclusion column (Superdex 200, GE Healthcare, USA) in PBS, pH 7.4. Thus, the apparent hydrodynamic radius increases with peptide ligand length. The elution volume of reference proteins with known molecular weight is indicated. Representative runs are shown. **B**, Relative affinity of extended peptides is similar to the core peptide. A competition-ELISA was conducted with one C- and one N-terminally extended HA peptide against biotinylated HA306-318.

The SDS-sensitivity of the N-terminal extended HLA-DR1/HA complexes could have been an indication for an altered binding register, which is less SDS-stable. To test for a potential difference in the binding mode of the peptide ligands, an ELISA-based competitive binding assay was used and the ability of HA289-318 and HA306-330 to compete with biotinylated HA306-318 for binding to HLA-DR1 was compared (Figure 12B). No difference between the relative affinities of the HA-based peptides could be observed. Therefore, it was not expected that the observed differences in SDS-migration and -stability were caused by a shifted binding register, because this would likely lead to a different affinity in the ELISA.

3.2.2 Crystal structure of HLA-DR1 in complex with HA-peptide length variants

So far, the major indication for a change in conformation came from SDS-PAGE analysis. As this is not a direct indicator, the conformation of the different MHC complexes on the atomic level was analyzed by X-ray crystallography. Because the C-terminal overhang showed a gradually change in the SDS-PAGE gel shift experiments and the N-terminal extended peptides lead to mainly SDS-instable complexes, the focus was set to the F-conformation induced by a C-terminal overhang.

Initial crystallization trials were carried out with HLA-DR1 purified from baculovirus-infected insect cells. To validate the quality of the preparations of the protein, crystals of HLA-DR1 in complex with the core peptide HA306-318 were reproduced. A fine screen based on published conditions (Stern *et al.*, 1994) indeed yielded several crystals (which were not further analyzed). Next, complexes with the C-terminally extended HA306-330 were prepared and subjected to crystallization. However, these crystallization trials were not successful until HLA-DR1 derived from refolding of *E. coli* expressed protein was used. The heterogeneous glycosylation of the MHC α- and β-chain in combination with the peptide overhang most likely hindered well ordered crystallization of the baculovirus-derived material.

Initial crystals of the *E. coli*-derived material grew only after a long incubation period of 3-10 months. Crystal growth after such extended incubation times often indicates proteolytic cleavage of flexible regions, which otherwise prevent crystal formation. To ensure the integrity of the peptide MHC-overhang, the complexes were tested by a simple SDS-PAGE-shift assay (Figure 13A). Crystal drops were directly diluted and completely dissolved with H_2O. Subsequent SDS-PAGE analysis revealed peptide MHC overhangs that were presumably between HA308-318 (C-conformer) and HA306-330 (F-conformer). Additionally, higher order aggregates were visible in all five tested conditions, probably derived from the aggregated protein in the crystallization drops. Because the exact length of the remaining peptide in these crystals could not be determined, refinement of the initial crystallization conditions was carried out with pMHC complexes with three different peptide-overhang extensions (HA306-322, -326, -330) and as a reference also with the core peptide (HA306-318). Diffraction-quality crystals could be obtained for all variants. Interestingly, crystals appeared in nearly identical conditions (Table 9). The crystals were flash-frozen and data sets were collected at beamline BL14.1 at BESSY II at Helmholtz-Center, Berlin, Germany. A summary of the statistics of the best crystals for each peptide/MHC complex is given in Table 10.

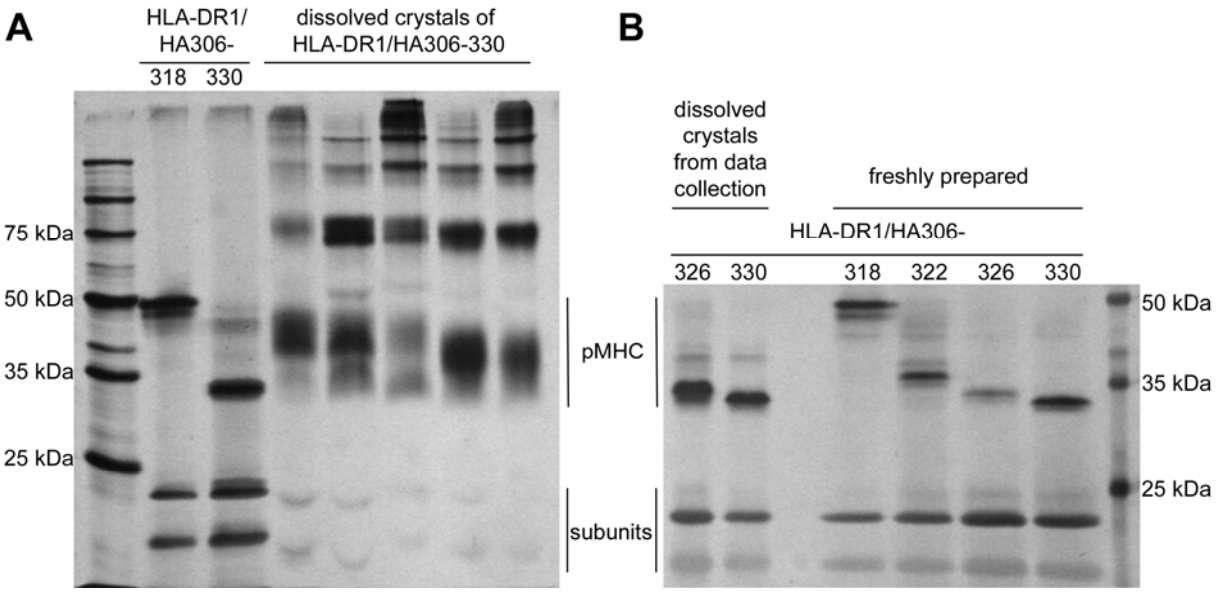


Figure 13: *SDS-PAGE analysis of dissolved crystals indicated integrity of peptide overhang.* **A**, Crystals of HLA-DR1/HA306-330 from the initial crystallization screen showed intermediate SDS-PAGE running behavior. To check for integrity of the longer HA306-330 peptide ligand, crystals of the initial crystallization screen were removed from the crystallization plate, washed, dissolved and run on a 12 % SDS-PAGE. The peptide MHC complexes showed an intermediate size between the C- and F-conformation. As reference, freshly prepared complexes for HLA-DR1/HA306-318 and HA306-330 were applied. **B**, Dissolving of crystals from X-ray diffraction experiments indicates integrity of full peptide overhang. Two crystals from diffraction experiments were thawed and dissolved in SDS-PAGE sample buffer and applied to a 12 % SDS gel. Two crystals of HLA-DR1/HA306-326 and -330 are presented in comparison with freshly prepared pMHC molecules with indicated peptides. Both gels were silver-stained.

The data were integrated and scaled using XDS and XSCALE (Kabsch, 2010). Data collection and processing was conducted under the guidance of Yvette Roske, MDC Berlin. The space groups as well as the unit cell dimensions are nearly identical in all cases. The resolution ranges from 1.95 Å for HLA-DR1/HA306-318 to 1.35 Å for HLA-DR1/HA306-330. The quality of the latter data set justified also the application of an anisotropic atomic displacement model. The crystal for HLA-DR1/HA306-318 was grown in a smaller crystallization drop than the other three complexes and therefore was also smaller, which is the likely cause for the reduced resolution of this data set. Only the data set for HLA-DR1/HA306-322 suffered from the presence of a second crystal lattice, apparently from a break in the crystal. However, the data could be integrated and scaled in XDS, even though the completeness of this data set was smaller than for the other data sets but still acceptable and useful for the generation of interpretable electron density maps and further structure refinement. Unfortunately, the raw data for the complex HLA-DR1/HA306-318 was later lost so that the redundancy for this data set could not be determined.

The phase problem for all four data sets was solved by molecular replacement using the coordinates of HLA-DR1 from PDB-entry 2G9H (Fernandez *et al.*, 2006). The peptide was re-

moved prior to molecular replacement using PHASER. The structure for HLA-DR1/HA306-318 has already been determined and published several times, either alone (Stern *et al.*, 1994) or in complex with other proteins like TCR (Hennecke *et al.*, 2000), superantigens (Fernandez *et al.*, 2006; Petersson *et al.*, 2001; Petersson *et al.*, 2002; Sundberg *et al.*, 2003; Zhao *et al.*, 2004), a combination of both (Saline *et al.*, 2010), or a viral protein (Mullen *et al.*, 2002). Nonetheless, also this data set was solved, because it was important to have the structure of the presumably C-form of pMHC determined from the same crystallization conditions as reference, to exclude that any potentially observed changes were due to crystallization artifacts.

Table 9: Statistics for data collection and refinement of HLA-DR1/HA peptide length variant structures

	HLA-DR1/ HA306-318	**HLA-DR1/ HA306-322**	**HLA-DR1/ HA306-326**	**HLA-DR/ HA306-330**
Data collection				
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Cell dimensions				
a, *b*, *c* (Å)	45.33, 97.61, 98.78	97.03, 100.35, 48.14	50.19, 97.15, 102.4	50.29, 96.96, 102.38
α, β, γ (°)	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90
mol/asu	1	1	1	1
Resolution (Å)*	31.6-1.95 (2.0-1.95)	44.6-1.59 (1.69-1.59)	30.9-1.48 (1.52-1.48)	45.3-1.35 (1.41-1.35)
R_{sym}*	6.8 (52.6)	4.4 (38.6)	4.0 (43.8)	3.7 (51.8)
$I / \sigma I$*	15.0 (3.3)	10.5 (1.6)	16.6 (2.1)	20.0 (2.5)
Completeness (%)*	99.8(100)	86.5 (73.3)	96.8 (85.3)	99.7 (99.8)
Redundancy*	n.a.[1]	3.2 (3.8)	3.0 (2.3)	3.6 (3.5)
Refinement				
Resolution (Å)	31.6-1.95	44.6-1.59	30.9-1.48	45.3-1.35
No. reflections	32673	54657	81450	110170
R_{work} / R_{free}	21.45/25.98	19.26/22.00	17.21/19.96	14.97/18.20
No. atoms				
Protein	3159	3227	3346	3432
Ligand/ion	-	-	10	14
Water	208	565	742	771
B-factors				
Protein	34.35	21.55	16.27	15.62
Ligand/ion	-	-	34.5	32.92
Water	39.37	31.76	30.99	31.48
R.m.s. deviations				
Bond lengths (Å)	0.009	0.006	0.006	0.005
Bond angles (°)	1.105	1.079	1.070	1.016
Ramachandran				
Favored (%)	98.4	98.5	98.5	98.8
Outlier (%)	0	0.3	0	0
Crystallization conditions	21% (w/v) PEG3350, 0.2M Na-malonate, 0.1M BisTrisPropane, pH 7.5	21% (w/v) PEG3350, 0.2M Na-malonate, 0.1M BisTrisPropane, pH 7.5	23% (w/v) PEG3350, 0.1M Na-citrate, 0.1M BisTrisPropane, pH 7.5	23% (w/v) PEG3350, 0.1M Na-citrate, 0.1M BisTrisPropane, pH 7.0

*) values in parentheses correspond to the highest resolution shell

1) not available; redundancy could not be calculated, because unmerged data was not available.

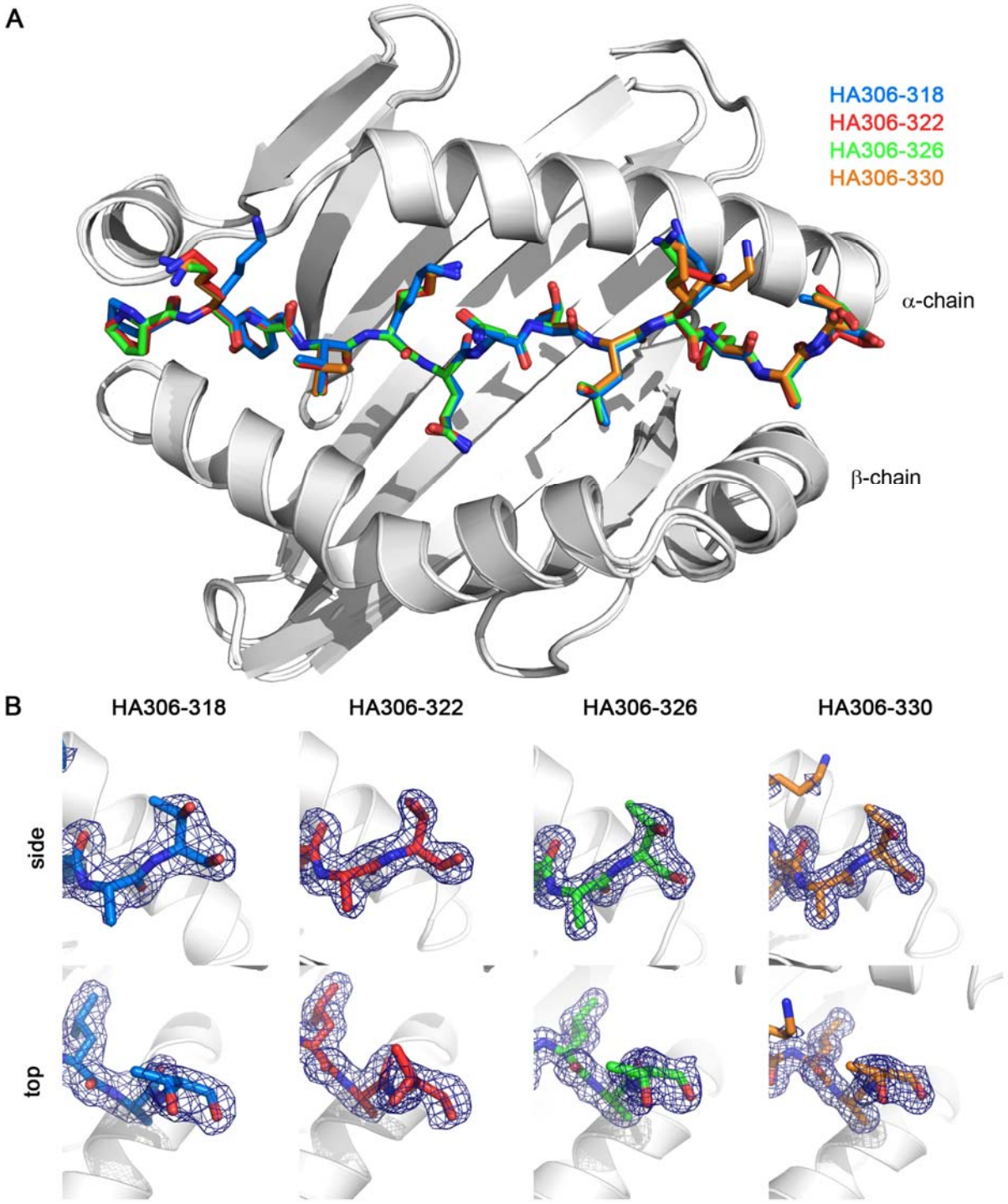


Figure 14: *Overview of crystal structures with HA peptide length variants.* **A**, View into the antigen-binding site of overlaid HLA-DR1 structures (only α1/β1-domain shown) in grey with the bound HA peptide variants as stick model (blue: HA306-318, red: HA306-322, green: HA306-326, orange: HA306-330). **B**, No clear electron density can be seen beyond Thr318 at the C-terminus of the HA peptides. The electron density from σ_A-weighted $2F_o$-F_c average kick maps (Praaenikar *et al.*, 2009) is contoured at 1σ and carved at 2Å around the peptide. The view is from the side (upper panel) with the α-chain α-helix in the background and from the top (lower panel) with part of the β-chain α-helix visible.

As expected in all four pMHC complexes (HA306-318, -322, -326, -330) the core binding motif remained the same (Figure 14A), no register shifts were observable. The overall structure of the MHC was also not changed. A superposition of the four structures showed an overall RMSD of 0.2 Å for the backbone of the α- and β-chain. Surprisingly, no interpretable electron density could be observed beyond residue 318 of the HA peptide in any of the three structures with a peptide ligand longer than this. Figure 14B presents the N-terminal part of the HA-peptide in all four structures with electron density from σ_A-weighted $2F_o$-F_c average kick maps. For calculation of these maps the coordinates of the model are displaced by random shifts of increasing RMSD and subsequently, all maps are averaged to yield a bias- and noise-reduced map (Praaenikar *et al.*, 2009).

As verification that the peptide overhang was still present, two crystals of HLA-DR1/HA306-326 and HA306-330 that have been used for data collection were also dissolved and applied to a gel shift experiment (Figure 13B). Therefore, the crystals, frozen in the nylon loop from data collection, were thawed and washed twice in drops containing the original crystallization buffer. Finally, these were dissolved in SDS-PAGE loading buffer by pipetting and applied to SDS-PAGE. Both complexes are in correspondence with the size of freshly prepared HLA-DR1/HA-complexes. Additionally, a faint second complex band above the expected band is visible. But in comparison to the freshly prepared samples the expected overhang in these complexes would also be longer than HA306-318. Thus, the peptides in the pMHC complexes used for data collection were still intact and did not suffer from proteolysis.

In summary, the analysis of the effect of the progressive extension of the HA-peptide ligand bound to HLA-DR1 yielded no structural changes in the conformation of the peptide-MHC complex. The only significant observed difference was the SDS-migration behavior. Finally, it was also tried to analyze the different complexes in solution by NMR-spectroscopy. A. Schlundt, FMP Berlin, recorded $^{1}H/^{15}N$ –heteronuclear single quantum coherence (HSQC) spectra of HLA-DR1 in complex with different C-terminally extended peptides (data not shown). However, these experiments did also not reveal any gross changes in the overall structure of HLA-DR1 when compared to the C-form complex with HA306-318. Taken together, the influence of peptide ligand extensions on the SDS-mobility of HLA-DR1/peptide complexes was confirmed, however, the structural basis underlying this effect could not be determined by X-ray crystallography.

3.3 New peptide binding mode of MHC II

3.3.1 The flipped binding mode

All newly synthesized MHC II are associated with an invariant peptide ligand to prevent uncontrolled peptide binding early in the ER. This peptide is derived from the invariant chain polypeptide, which is proteolytically degraded to the class II-associated invariant chain peptide (CLIP) during maturation of the MHC II. These CLIP fragments are heterogeneous in size, but share a common core sequence (Chicz *et al.*, 1992; Riberdy *et al.*, 1992). The structure of a CLIP complex has been solved for HLA-DR3 (Ghosh *et al.*, 1995) and also murine I-A^b (Zhu *et al.*, 2003), but no structural information was available for HLA-DR1. In this study three different HLA-DR1 structures with two length variants of CLIP were determined in order to analyze the influence of the length variations on the MHC II conformation.

To identify potential differences in the fine structure of the HLA-DR1/CLIP complexes caused by the loading process, two different strategies were used to prepare the complexes. The ectodomains of the individual subchains of HLA-DR1 were separately expressed as inclusion bodies in *E. coli* and subsequently purified under denaturing conditions. Then the purified chains were refolded together by rapid dilution either in absence (for empty HLA-DR1) or in the presence of peptide ligand (for peptide-loaded HLA-DR1). The empty MHC II was loaded posterior by incubating with excess free peptide and one of the two described molecular loading enhancers AdEtOH or Ac-FR-NH_2 (Gupta *et al.*, 2008; Höpner *et al.*, 2006). The complexes were subjected to crystallization by vapor-diffusion in sitting-drop wells.

Initially, the shortest reported CLIP-sequence was used for crystallization as this peptide was predicted to have only short overhangs pointing out of the antigen-binding site that could interfere with crystal packing. Furthermore, CLIP106-120 was also reported to have an affinity for HLA-DR1 that is at least as high as that of longer versions and was isolated from HLA-DR1 with the highest abundance (Chicz *et al.*, 1992). The peptide was loaded posterior on empty MHC II and then crystallized at RT. Diffraction quality crystals appeared after 5 days in several conditions. Data were collected from three different crystals at the synchrotron source BESSY II at Helmholtz-Center, Berlin, Germany beamline 14.1. Details are given in Table 10.

Table 10: Summary of data collection and refinement of HLA-DR1/CLIP complexes

	HLA-DR1/ $CLIP_{106-120}$, flipped			**HLA-DR1/ $CLIP_{106-120}$, canonical**	**HLA-DR1/ $CLIP_{102-120}$**
	crystal 1	crystal 2	crystal 3		
PDB	3PGC	-	-	3PGD	3PDO
Data collection					
Space group	$P3_121$	C222	$P2_1$	$P4_32_12$	$P2_12_12_1$
Cell dimensions					
a, *b*, *c* (Å)	138.15, 138.15, 104.52	67.34, 143.82, 108.73	70.61, 80.93, 80.61	94.45, 94.45, 275.53	45.45, 97.63, 99.05
α, β, γ (°)	90, 90, 120	90, 90, 90	90, 92.12, 90	90, 90, 90	90, 90, 90
mol/asu	2	1	2	2	1
Resolution (Å)*	34.54-2.66 (2.73-2.66)	34.14-2.68 (2.75-2.68)	32.76-2.49 (2.58-2.49)	34.44-2.72 (2.79-2.72)	33.26-1.95 (2.00-1.95)
R_{sym}*	8.1 (67.9)	7.5 (46.8)	6.9 (47.9)	13.1 (87.3)	7.1 (54.6)
$I / \sigma I$*	18.0 (3.0)	14.8 (2.6)	13.4 (2.6)	12.2 (2.2)	16.6 (3.4)
Completeness (%)*	99.9 (100)	97.4 (98.7)	97.2 (97.3)	99.9 (100)	99.6 (99.3)
Redundancy*	7.3 (7.5)	3.3 (3.3)	2.6 (2.6)	7.6 (6.8)	4.8 (4.9)
Refinement					
Resolution (Å)	34.54-2.66	32.17-2.68	32.76-2.49	33.39-2.72	33.26-1.95
No. reflections	33390	14845	31078	34451	31168
R_{work} / R_{free}	19.8/25.0	19.1/23.9	20.1/25.4	19.6/24.2	19.1/24.0
No. atoms					
Protein	6052	3027	6154	6151	3207
Ligand/ion	13	-	-	-	15
Water	115	95	226	173	320
B-factors					
Protein	53.4	42.4	37.5	44.7	27.6
Ligand/ion	66.7	-	-	-	32.7
Water	49.8	36.5	34.9	38.1	34.7
R.m.s. deviations					
Bond lengths (Å)	0.008	0.004	0.009	0.009	0.016
Bond angles (°)	1.13	0.83	1.27	1.13	1.54
Ramachandran					
Favored (%)	96.9	97.0	97.3	96.0	97.9
Allowed (%)	3.0	3.0	2.6	3.7	2.1
Outlier (%)	0.1	-	0.1	0.3	-
Crystallization conditions	25% (w/v) PEG 1500, 0.1 M MIB buffer, pH 6	0.2 M sodium malonate, 20% (w/v) PEG 3350	0.2 M ammonium acetate, 15% (w/v) PEG 4000, 0.1M sodium acetate, pH 4.0	0.2 M sodium citrate, 20% (w/v) PEG 3350, 0.1M Bis Tris propane, pH 6.5	0.2 M magnesium formate, 20 % (w/v) PEG 3350

*) values in parentheses correspond to the highest resolution shell

All three crystals belonged to different crystal space groups, but were in a similar resolution range (2.49-2.68 Å). The data was readily indexed, integrated and scaled with XDS and XSCALE (Kabsch, 2010). Initial phases were obtained by molecular replacement with PHASER (McCoy *et al.*, 2007) using PDB-ID 2G9H (Fernandez *et al.*, 2006) with the coordinates for the peptide removed. Initially, only the model for the α- and β-chain was built and only at later stages the peptide was modeled into unambiguous electron density (Figure 16 B). The overall structure of the HLA-DR1/CLIP106-120 complex resembles a typical MHC II (Figure 15). Notably however, the peptide binding mode is highly untypical. Contrary to the structure of the more than 30 unique peptide-MHC II complexes solved so far (appendix), the peptide lies in a reversed orientation in the antigen binding groove ("HLA-DR1/CLIP106-120,flipped"). The N-terminus of the peptide is accommodated near the P9 pocket and the C-terminus near the P1 pocket. The data sets from three different space groups enabled us to exclude that the observed binding mode was an artifact from crystallization. The comparison of five individual HLA-DR1/CLIP106-120,flipped structures from the three crystal forms had all different crystal contacts, yet the overall structure of the peptide bound to the MHC II remained the same (Figure 15 and Table 10). The only variances were observed at the very C-terminus of the peptide ligand that was not directly bound anymore and the arginine side chain at P8, which is pointing away from the binding site (Figure 15).

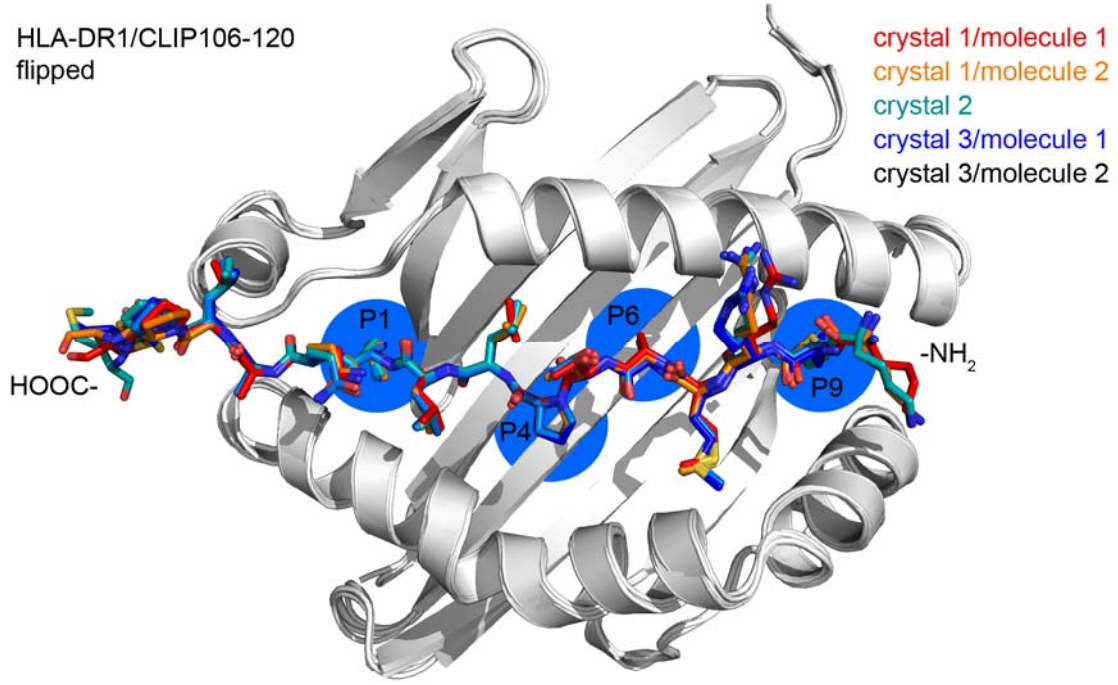


Figure 15: *Reversed peptide binding mode as observed in different crystal forms.* The five individual HLA-DR1/CLIP106-120,flipped structures as found in the three different crystal forms were superposed. Only the peptide binding site is shown for clarity. The α1/β1-domains are shown in grey cartoon representation, CLIP is shown as stick model with individual colors for each HLA-DR1/CLIP106-120,flipped structure. The termini as well as the major binding pockets in the antigen-binding site are indicated.

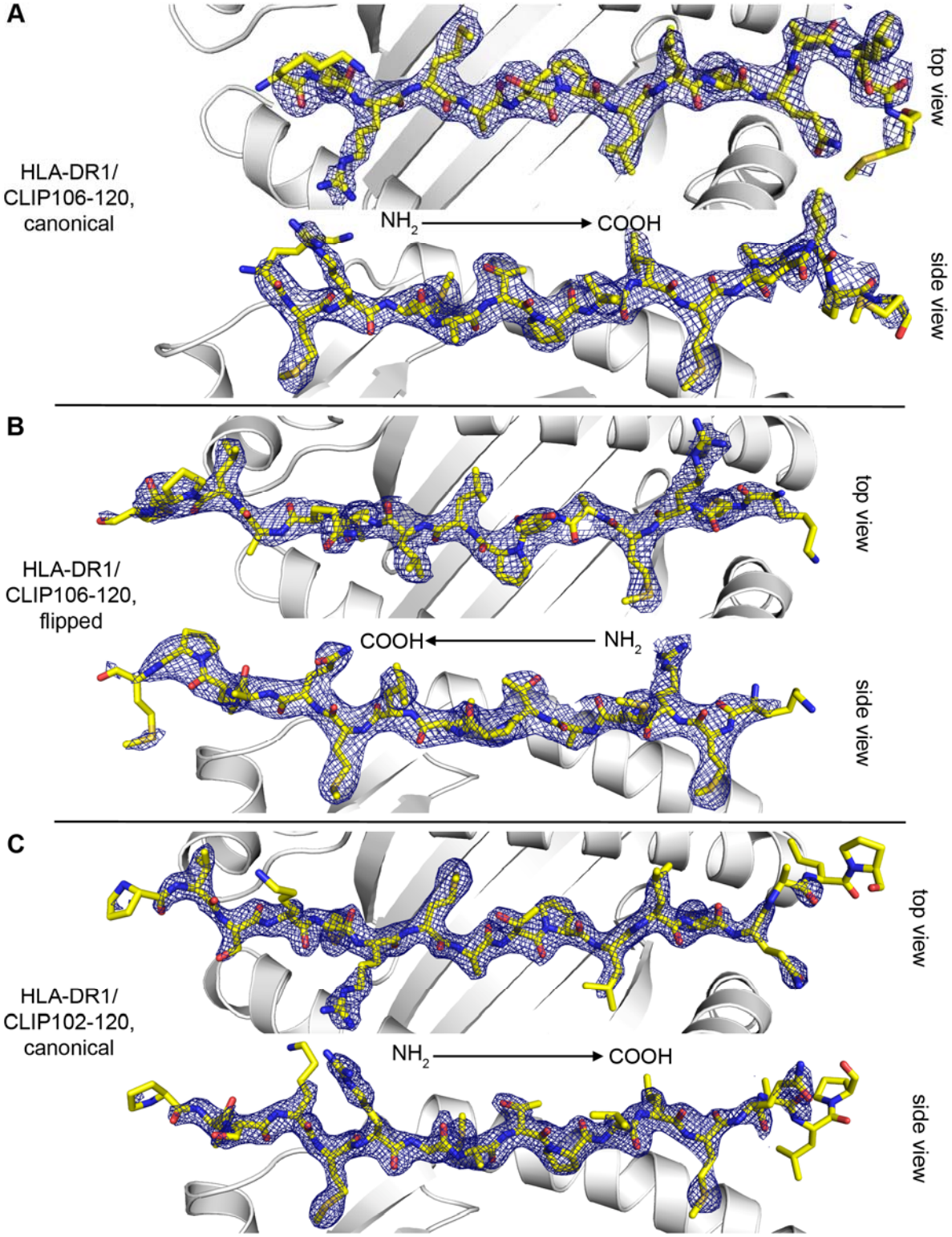


Figure 16: *Distinct peptide binding orientations observed in HLA-DR1 in complex with different CLIP peptides.* For all three complexes two views of the peptide in the antigen-binding site are given (top view, from the top down into the binding site; side view, from the front with the α-chain α-helix in the background, the β-chain is omitted for clarity). The direction of the peptide (yellow sticks) is indicated. The electron density from σ_A-weighted $2F_o$-F_c average omit kick maps (Praaenikar *et al.*, 2009) is contoured at 1σ and carved at 2Å around the peptide. The peptide coordinates of the final models were removed before electron density calculation.

A possible explanation for this flipped binding mode emerged when the potential conventional binding mode for CLIP106-120 was considered. Based on published binding motifs for HLA-DR1 (Murthy & Stern, 1997; Stern *et al.*, 1994) and comparison with two peptide-MHC II structures, using longer CLIP versions bound to different human (Ghosh *et al.*, 1995) or murine (Zhu *et al.*, 2003) MHC allelic variants, Met107 was predicted to be the P1 anchoring residue. Using CLIP106-120, this would lead to the loss of hydrogen bonds at position P-2, which would be formed in longer CLIP variants by residue 105. However, in the flipped peptide orientation these hydrogen bonds are reestablished (Figure 17). Therefore, we reasoned that the canonical peptide binding mode should be feasible if CLIP was extended N-terminally by sufficient amino acids to enable normal hydrogen bond formation at P-1/-2. We repeated the crystallization with CLIP102-120. Posterior loading of the peptide onto HLA-DR1 did not lead to diffraction-quality crystals. Only when the peptide was added already during refolding of the MHC II, high-quality crystals could be grown. Again, a data set from a single crystal was recorded at the synchrotron source BESSY II at Helmholtz-Center, Berlin, Germany at beamline 14.1 and the structure was solved following the same strategy as mentioned above (for details see Table 10). The crystal diffracted to 1.95 Å and the data gave a clear picture of the peptide binding mode (Figure 16C). Indeed, in this set up the peptide was now bound in the predicted binding register and with the canonical orientation (Figure 17).

The major difference in the process of sample preparation was the point at which the peptide was added, either during refolding of the MHC II or by later loading on already refolded, empty MHC II. Therefore, the way of complex preparation could have led to artifacts in peptide binding. Another technique that enabled us to gain information at the atomic level of the pMHC in solution was used to exclude this possibility. Andreas Schlundt (FMP, Berlin) employed nuclear magnetic resonance spectroscopy (NMR) to analyze the reversed peptide binding in solution (Gunther *et al.*, 2010). $^{1}H/^{15}N$-HSQC spectra give a fingerprint of the structure of the MHC in solution as each NH-group has a unique chemical shift that is influenced by its direct chemical environment. Changes in the three dimensional structure are reflected by chemical shift changes in a two dimensional HSQC spectrum. The assignment of the peaks in a spectrum to individual residues allows the localization of these changes in the structure when the three-dimensional structure is known. Initially, spectra were recorded for HLA-DR1 freshly co-refolded (Figure 18, blue spectrum) or *a posteriori* loaded with CLIP106-120 (Figure 18, green spectrum). Both spectra are largely overlapping, but also show clear differences.

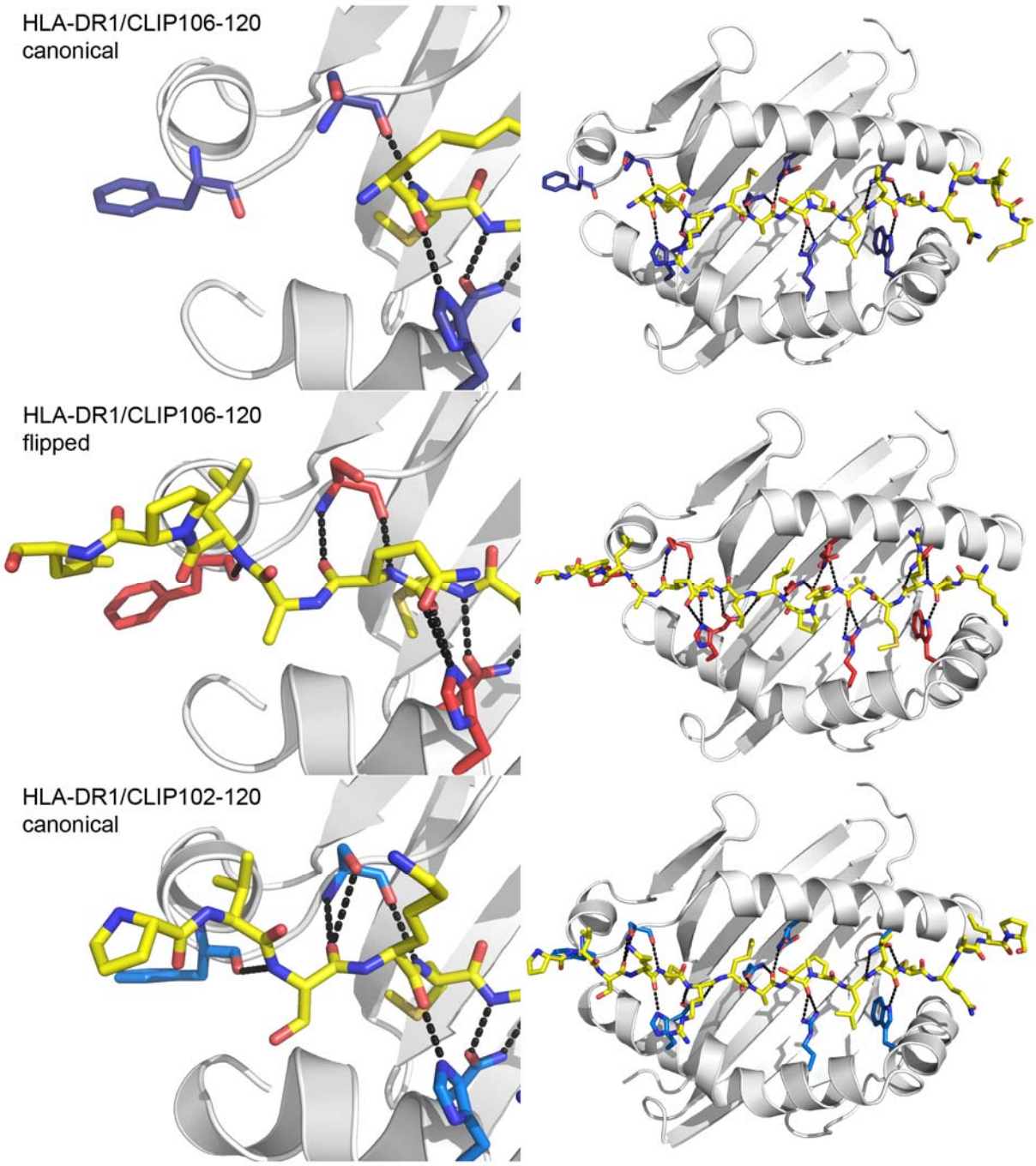


Figure 17: *N-terminal hydrogen bonds are critical for peptide MHC stability.* Overview of HLA-DR1 structures obtained in complex with two different CLIP length variants. Cartoon representation of MHC α1/β1 domains with view onto antigen binding groove (right panel) and a close-up view of the terminal region of the binding groove close to pocket 1 (left panel). Amino acids making conserved hydrogen bonds to the backbone of the peptide are highlighted as colored sticks. Conserved hydrogen bonds are indicated as black dashes.

The assignment reveals that these differences are mostly located in the β1-domain, i.e. the antigen-binding site (Gunther *et al.*, 2010). In the presented experiments only the β-chain was isotope labeled. The α-chain remains invisible. However, the same results were obtained for α-chain labeled samples (A. Schlundt, private communication). Surprisingly, repeated re-

cordings of HSQC spectra after incubating the co-refolded HLA-DR1/CLIP106-120 at 37°C led to changes in the spectrum. Again, these changes could be assigned to residues in the antigen-binding site. Since mass spectrometry confirmed the intactness of the bound peptide over time, it became clear that the differences in the spectra were likely caused by changes in the peptide binding mode. Intriguingly, the spectra of the equilibrated co-refolded sample were identical with the *a posteriori* loaded sample. Thus, it was proposed that in the initial structure after co-refolding MHC and peptide are in a kinetically-trapped conformation that changes after thermodynamical equilibration. This kinetically-trapped conformation could be the canonically bound CLIP106-120, which is lacking two N-terminal conserved hydrogen bonds.

Based on these assumptions, it was tried to crystallize this "trapped" conformation of HLA-DR1/CLIP106-120 by reducing the crystallization temperature to 4°C, prolonging the time for rearranging the peptide within the MHC. And, indeed, by this approach diffraction-quality crystals were obtained. The details of the crystal structure are found in Table 10. The structure was solved by molecular replacement following the same strategy as for the other HLA-DR1/CLIP complexes. In this structure the peptide was accommodated in the canonical orientation with Met107 occupying the P1 pocket in the same way as observed for the longer CLIP complex and, as expected, with the N-terminal H-bonds not formed (Figure 16A and Figure 17).

The NMR-experiments revealed also why the *a posteriori* loaded HLA-DR1/CLIP102-120 complex was not successfully crystallized. These samples showed a $^{1}H/^{15}N$-HSQC spectra that had features of both the canonical and flipped HLA-DR1/CLIP106-120. Moreover, the samples of co-refolded CLIP102-120 were also not stable. The spectra changed over time like the spectra for CLIP106-120, although on a much longer time-scale (Gunther *et al.*, 2010). Additionally, posterior loading of CLIP102-120 on HLA-DR1 led to a mixture of both conformations (Gunther *et al.*, 2010).

Finally, the usage of spin-labeled CLIP enabled the direct detection of the peptide-orientation in the NMR experiment. Therefore, a paramagnetic relaxation enhancing group was attached to the N-terminus of CLIP106-120. *A posteriori* loading of this peptide onto HLA-DR1 revealed HSQC spectra with signal loss for residues only in the region around the P9 pocket, reflecting the flipped binding mode of CLIP (A. Schlundt, FMP, Berlin) (Gunther *et al.*, 2010).

This set of three structures with two different CLIP length variants enabled a detailed analysis of the principles governing the reorientation of the CLIP peptides in the MHC binding groove.

First of all, the most astonishing observation was that no major reorganization of the MHC binding groove was necessary for the reversed binding mode (Figure 19). All amino acids in the MHC binding groove that participate in the conserved hydrogen bonding network are also an integral part of the flipped peptide complex and only minor adjustments of side chain conformations are observed. Only the carboxamide group of αN69 is tilted by 40 degrees to form two hydrogen bonds to the backbone of the peptide and also the benzyl ring of αF51 is reoriented, but here the hydrogen bonds to the peptide are formed by the backbone carboxyl group. All other hydrogen bond forming MHC residues are perfectly superimposable (Figure 19). Altogether in the complex of HLA-DR1/CLIP102-120 17 hydrogen bonds are formed (15 sequence-independent to the CLIP-backbone/2 to CLIP side chains). In the complex of HLA-DR1/CLIP106-120,canonical three N-terminal hydrogen bonds are missing. However, in the HLA-DR1/CLIP106-120,flipped complex these bonds are reestablished and in total 16 hydrogen bonds are formed (15/1). Interestingly, residues αQ9, αN62, αN69 and βN82 have been described to form 9-membered and 11-membered ring structures with the bound peptide and that these interactions might me important to induce the polyproline like fold of the peptide (Jardetzky *et al.*, 1996; Le Questel *et al.*, 1993). In both peptide orientations these bidentate hydrogen bonding structures are formed with the difference that all 9-membered rings turned into 11-membered rings and vice versa (Figure 19B). Only αN62 does not form two hydrogen bonds to the canonically bound peptide because the proline at position 6 of CLIP is lacking a free hydrogen at the backbone nitrogen.

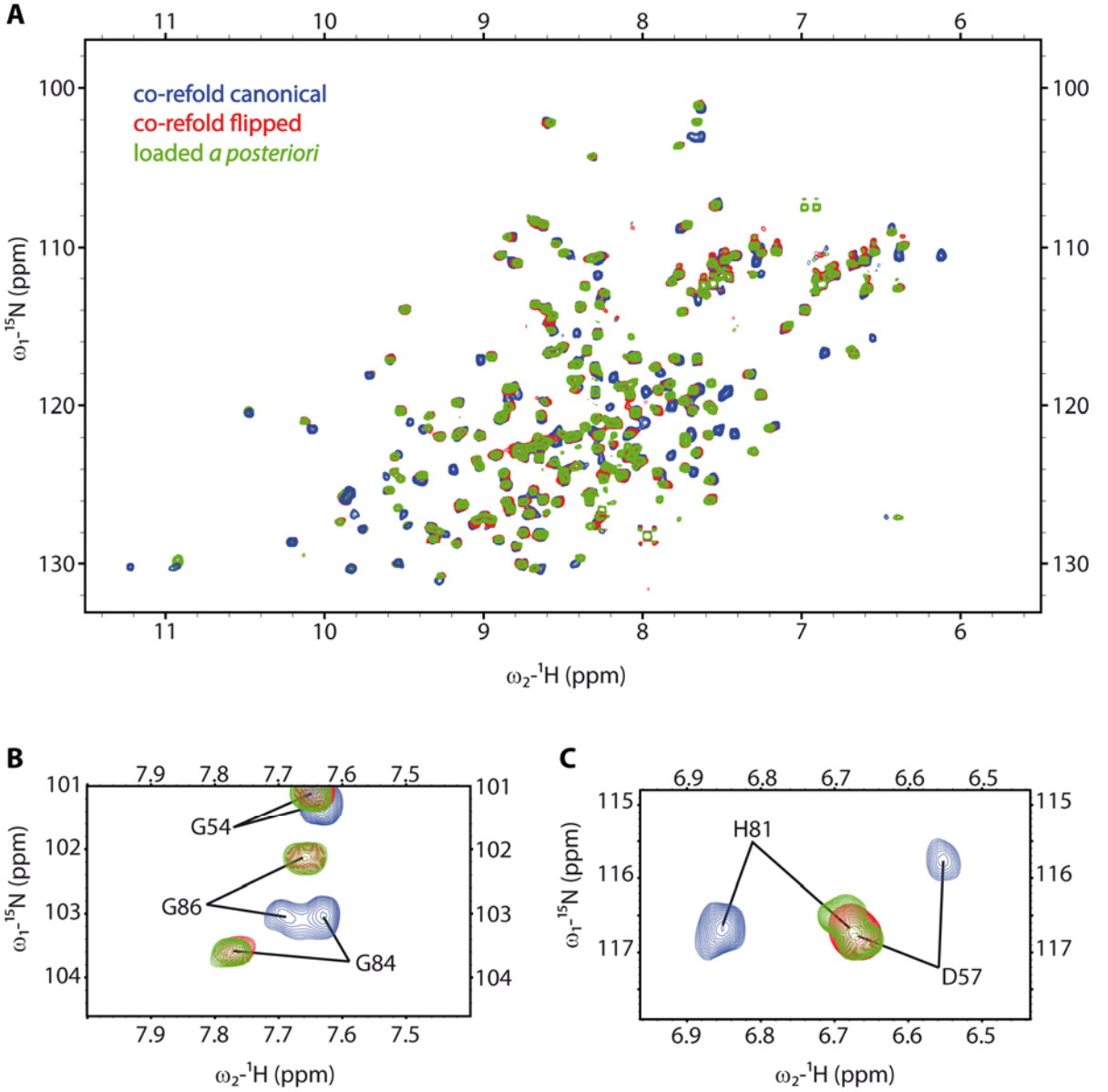


Figure 18: *NMR spectra reveal dynamics of co-refolded HLA-DR1/CLIP106-120 complex.* **A**, $^{1}H/^{15}N$- HSQC spectra of HLA-DR1 were recorded either directly after co-refolding (blue) or after equilibration for 5 days at 37°C (red). Both spectra show significant chemical shift differences. Comparison of the equilibrated HLA-DR1/CLIP106-120 complex (red) with the spectrum of HLA-DR1 loaded *a posteriori* with CLIP106-120 (green) implies it corresponds to the flipped state of CLIP. **B** and **C**, Excerpts from A, show details for selected residues of the β-chain. Only the β-chain was ^{15}N-labeled, the α-chain remains invisible. Spectra were kindly provided by A. Schlundt, FMP Berlin.

Analysis of the thermal stability of the different complexes by the thermofluor assay did not reveal a major difference between HLA-DR1/CLIP106-120 and CLIP102-120 complexes, neither with co-refolded nor with posterior loaded complexes (Table 11). Complexes with the longer CLIP102-120 were slightly more stable. This shows that not only the overall number of hydrogen bonds but also the favorable accommodation of peptide side chains in the MHC binding pockets has an impact on pMHC stability. Of course, the peptide binding direction is

not fixed during the heating, thus it cannot be excluded that the peptide is flipping and thereby obscuring any more pronounced differences between the two peptide binding directions.

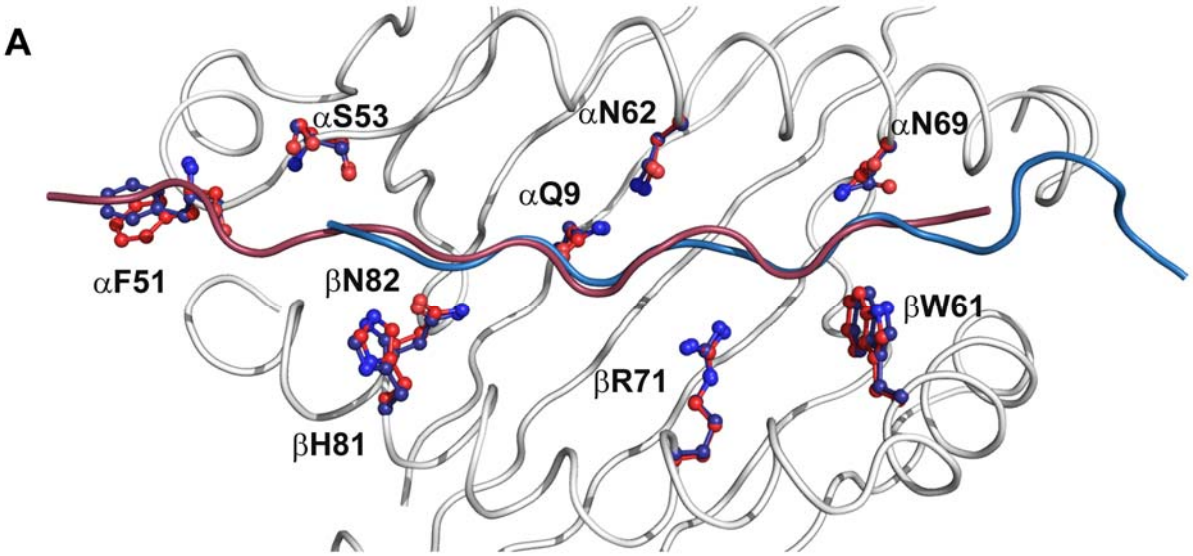
Table 11: Summary of thermal stability of different HLA-DR1/CLIP complexes as determined by thermofluor assay, in PBS, pH 5.8.

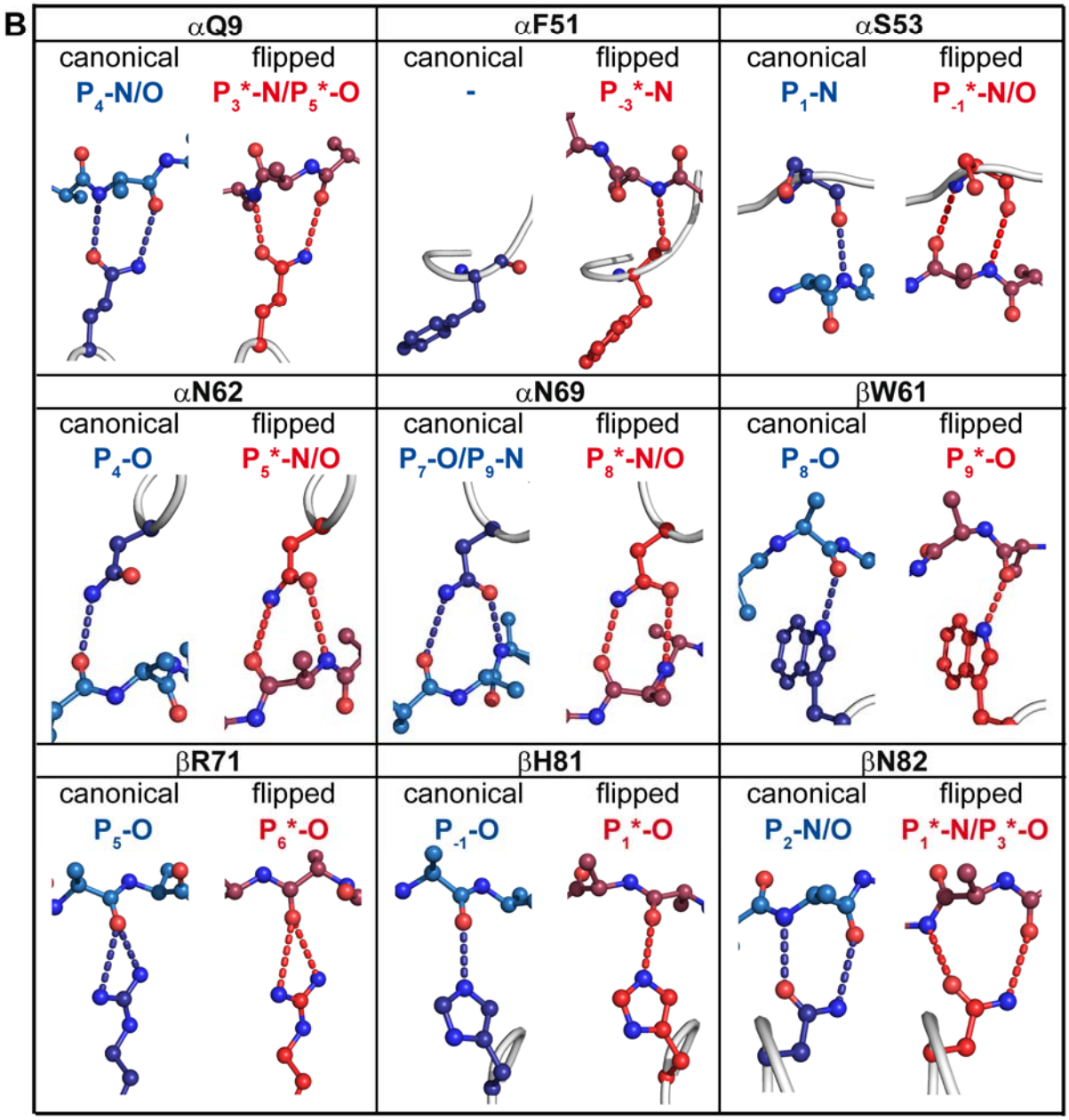
HLA-DR1 bound to:	**T_m [°C]**	**No. of experiments**
$CLIP_{106-120}$, loaded *a posteriori*	75.4 ± 0.2	8
$CLIP_{106-120}$, fresh co-refolded	74.2 ± 0.1	4
$CLIP_{102-120}$, loaded *a posteriori*	78.0 ± 0.2	4
$CLIP_{102-120}$, fresh co-refolded	77.4 ± 0.2	4

3.3.2 Design of orientation-specific CLIP ligands

Until now the phenomenon of a flipped binding mode could only be shown by X-ray crystallography and NMR-spectroscopy on soluble HLA-DR1. The analysis of this binding mode in an *in vivo* situation on APCs is impossible to assess by these methods. However, nature has evolved molecules with exquisite specificity. On the one hand there is the natural binding partner of pMHC, the T cell receptor (TCR), which detects even subtle changes in the MHC-presented peptide (Deng *et al.*, 2007). On the other hand also antibodies have the specificity that would be required to discriminate between two differently bound versions of one peptide. However, in order to raise TCRs or antibodies with specificity for either of the two orientations, the CLIP bound MHC II on cells has to be in a single defined orientation. As it is shown in Figure 18, posterior loading of CLIP106-120 leads only to the flipped conformation. Moreover, posterior loading of CLIP102-120 on HLA-DR1 results in an approximately even distribution between the canonical and flipped conformation (data not shown) and the co-refolded complex is also not stable. Therefore, a selection of antibodies or TCRs for the canonically bound CLIP seems impossible by wt CLIP/HLA-DR1 complexes.

Figure 19 (next page): *Detailed comparison of conserved hydrogen bonding network in canonical and reversed peptide binding mode*. **A**, Superposition of conserved amino acids making peptide sequence independent hydrogen bonds to the backbone of the bound peptide. The structures of HLA-DR1/CLIP106-120 in the canonical (blue) and flipped (red) orientation are shown as sticks. For a better overview, the α1/β1-domain is only depicted for the canonical structure (as loop structure, grey). The peptide is shown as loop structure in blue (canonical) and purple (flipped). **B,** Side-by-side comparison of hydrogen bonds from conserved MHC side chains to the peptide backbone. Color coding is as in A. The peptide is shown as stick model with the side chains only up to the Cβ-atom. Hydrogen bonds are indicated as dashed lines. The hydrogen bonds are named by the peptide's relative position to the MHC binding pockets. In the flipped binding mode the numbering refers to the canonical pocket numbering and is marked with an asterisk.


A
αS53
αN62
αN69
αQ9
αF51
βN82
βW61
βR71
βH81


B
αQ9
canonical
flipped
P4-N/O
P3*-N/P5*-O
αF51
canonical
flipped
-
P-3*-N
αS53
canonical
flipped
P1-N
P-1*-N/O
αN62
canonical
flipped
P4-O
P5*-N/O
αN69
canonical
flipped
P7-O/P9-N
P8*-N/O
βW61
canonical
flipped
P8-O
P9*-O
βR71
canonical
flipped
P5-O
P6*-O
βH81
canonical
flipped
P-1-O
P1*-O
βN82
canonical
flipped
P2-N/O
P1*-N/P3*-O

The flipping of CLIP is feasible because of the chemical identity of the P1/P9 residues (Met) and the hydrophobic nature of pocket 1 and 9. However, both pockets share the preference for hydrophobic side chains but the P1 pocket is much bigger and can even accommodate Trp side chains (PDB: 1AQD) (Murthy & Stern, 1997) or even bigger unnatural amino acids (PDB:2IPK) (Venkatraman *et al.*, 2007). In contrast, the P9 pocket would be too small to bind these amino acids. To generate orientation-specific CLIP peptides, one of each anchoring Met was mutated to Trp to favor one of both binding directions, with either the canonical bound peptide (M107W) or the flipped peptide (M115W). Although wt CLIP106-120 already binds exclusively in the reversed orientation when loaded *a posterior*i, the latter mutation M115W was also analyzed to show that the tryptophan can also be accommodated in the P1 pocket in the flipped orientation of the peptide. This might guide later mutations to fix other peptides than CLIP in either of both binding orientations.

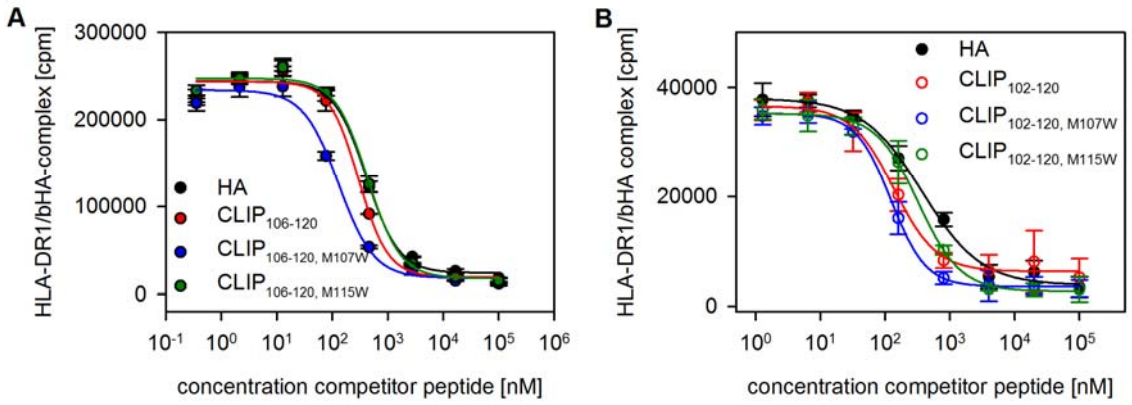


Figure 20: *Comparable peptide binding of P1-anchor-mutants.* The relative affinity of CLIP variants was determined by a competition ELISA using a constant concentration of biotinylated HA306-318 and varying amounts of unlabeled CLIP wt and Trp-mutants as competitor peptide. **A**, CLIP106-120 variants; **B**, CLIP102-120 variants. Data points represent the average of quadruplicates with SD. Representative experiments are shown.

First, the new peptide ligands were characterized by their binding behavior to HLA-DR1. Competitive binding studies with labeled HA306-318 showed similar results for both length variants of CLIP (Figure 20). All CLIP variants competed at least as good as unlabeled HA for binding to HLA-DR1. Details are given in Table 12. The wt sequences showed constantly higher affinity for HLA-DR1 than HA. This is in agreement with previous observations (Sette *et al.*, 1995; Zarutskie *et al.*, 1999). Mutation M107W enhanced the binding to HLA-DR1 compared to the wt, whereas M115W did not increase the peptide affinity significantly com-

pared to HA and had only a slightly lower affinity than the wt. Second, the thermal stability of the MHC complexes was measured, revealing similar effects. All complexes were generated by *a posteriori* loading of peptides on empty HLA-DR1. Free peptide was then removed by gel filtration. For this reason, it was expected that the complex of the shorter wt CLIP106-120 and HLA-DR1 was entirely in the flipped binding mode, whereas the complex with the longer CLIP102-120 would lead to a mixture of canonically and flipped bound peptide. The Trp-mutants were expected to obtain the desired binding modes (M107W canonical, M115W flipped). Therefore, a distinct contribution of the point mutations to the overall stability could not be drawn for all complexes. But in general, the longer CLIP variants had a slightly higher stability than the shorter versions. The effect was most pronounced for the M107W mutants (ΔT_m = 5°C), whereas the increase in stability for the M115W mutant was smaller (ΔT_m = 2°C). This difference was likely caused by the additional N-terminal H bonds formed by the peptide extension of the CLIP102-120, M107W peptide, whereas the additional amino acids in CLIP102-120, M115W were not expected to contribute to binding. As expected from the competitive binding experiments (Figure 20), the M107W variants were more stable than their wt counterparts. Surprisingly, the M115W variants were slightly less stable than the wt peptides although the tryptophan is expected to yield more hydrophobic interactions with the P1 pocket as compared to methionine.

Table 12: Summary of binding characteristics (determined by competition against biotinylated HA306-318 peptide) and thermal stability of HLA-DR1/peptide complexes loaded *a posteriori* (determined by thermofluor assay). Binding experiments were carried out in PBS, pH 7.4, thermal stability was determined in PBS, pH 5.8. Values are the mean of at least two independent experiments. For thermal stability measurements peptides were loaded *a posteriori*.

Peptide	**$EC50_{pep}/EC50_{HA}$**	**T_m [°C]**
HA	1	84.1 ± 0.1
CLIP106-120	0.6 ± 0.1	74.2 ± 0.2
CLIP106-120, M107W	0.3 ± 0.04	76.1 ± 0.1
CLIP106-120, M115W	0.9 ± 0.5	73.2 ± 0.1
CLIP102-120	0.5 ± 0.1	77.1 ± 0.1
CLIP102-120, M107W	0.4 ± 0.1	81.4 ± 0.1
CLIP102-120, M115W	0.9 ± 0.1	75.5 ± 0.1

Table 13: Data collection and refinement statistics for HLA-DR1/CLIP mutants

	HLA-DR1/CLIP102-120, M107W	**HLA-DR1/CLIP106-120, M115W**
Data collection		
Space group	$C222_1$	$P2_1$
Cell dimensions		
a, *b*, *c* (Å)	73.78, 90.85, 138.79	70.68, 81.04, 80.41
α, β, γ (°)	90, 90, 90	90, 91.83, 90
mol/asu	1	2
Resolution (Å)*	34.7-1.34 (1.37-1.34)	34.46-2.36 (2.42-2.36)
R_{sym}*	4.9 (50.5)	10.7 (69.3)
$I / \sigma I$*	15.1 (2.0)	12.9 (2.2)
Completeness (%)*	98.2 (83.8)	98.5 (98.5)
Redundancy*	4.0 (3.0)	4.1 (4.1)
Refinement		
Resolution (Å)	34.7-1.34	34.46-2.36
No. reflections	97450	
R_{work} / R_{free}	13.88/17.27	18.3/23.72
No. atoms	3920	6619
Protein	3357	6205
Ligand/ion	6	34
Water	557	380
B-factors		
Protein	16.8	28.5
Ligand/ion	29.4	41.6
Water	29.9	29.6
R.m.s. deviations		
Bond lengths (Å)	0.015	0.008
Bond angles (°)	1.523	1.085
Ramachandran		
Favored (%)	97.8	98.0
Outlier (%)	0.2	0
Crystallization conditions	0.2 M K/Na-tartrate, 20% (w/v) PEG3350, 0.1M BisTrisPropane, pH 8.5	19% (w/v) PEG3350, 0.1M MMT, pH 5.0

*) values in parentheses correspond to the highest resolution shell

All interpretation is based on the assumption that the binding register for the CLIP variants is as intended by the mutations. The only two methods to unambiguously determine the correct binding register of the mutated peptide ligands are either NMR spectroscopy or X-ray crystallography. Thus, it was attempted to crystallize both versions of the mutated HLA-DR1/CLIP mutants. To circumvent the possible problem of kinetically-trapped binding modes of CLIP as observed for the wt complex when co-refolded, the mutated CLIPs were loaded *a posteriori* on empty HLA-DR1. Diffraction-quality crystals could be grown for HLA-DR1/CLIP102-120, M107W and HLA-DR1/CLIP106-120, M115W in sitting-drop wells by vapor diffusion. In contrast, the other two complexes HLA-DR1/CLIP106-120, M107W and HLA-DR1/CLIP102-120, M115W yielded only crystals of insufficient quality, mostly tangly

grown. In parallel recorded HSQC spectra indicated that the complex formation was not homogenous for the latter two but contained a minor population of complexes with, putatively, oppositely bound peptides (A. Schlundt, private communication). However, for the first two complexes the combination of peptide length and P1-anchor mutation was sufficient to generate homogenous complexes by *a posteriori* loading of the peptide, which would also be the method of choice for peptide loading of cell surface presented HLA-DR1. Complete data sets were recorded at the synchrotron source BESSY II at Helmholtz-Center, Berlin, beamline 14.1 (Table 13) and both structures could be solved by molecular replacement. The complex of HLA-DR1/CLIP102-120, M107W showed excellent diffraction behavior and data could be used to a resolution of 1.34 Å. The high number of recorded unique reflections allowed the usage of an anisotropic displacement parameter model. For the M115W mutant the resolution limit was 2.36 Å. The twofold non crystallographic symmetry was included in the structure refinement.

Both structures showed the predicted binding mode for the mutated peptides (Figure 21), with the CLIP102-120, M107W in the canonical and the CLIP106-120, M115W in the flipped orientation. In both cases the newly introduced tryptophan resides in the P1 pocket and the rest of the CLIP resembles the corresponding HLA-DR1/CLIP wt structures. No rearrangements were necessary to accommodate the indole ring of the tryptophan in the P1 pocket. Only the benzyl ring of αF24 is tilted slightly in both structures as compared to the corresponding wt structures. The same adjustment has been observed in the structure of HLA-DR1/A2 (Murthy & Stern, 1997), in which the P1 pocket is also occupied by tryptophan.

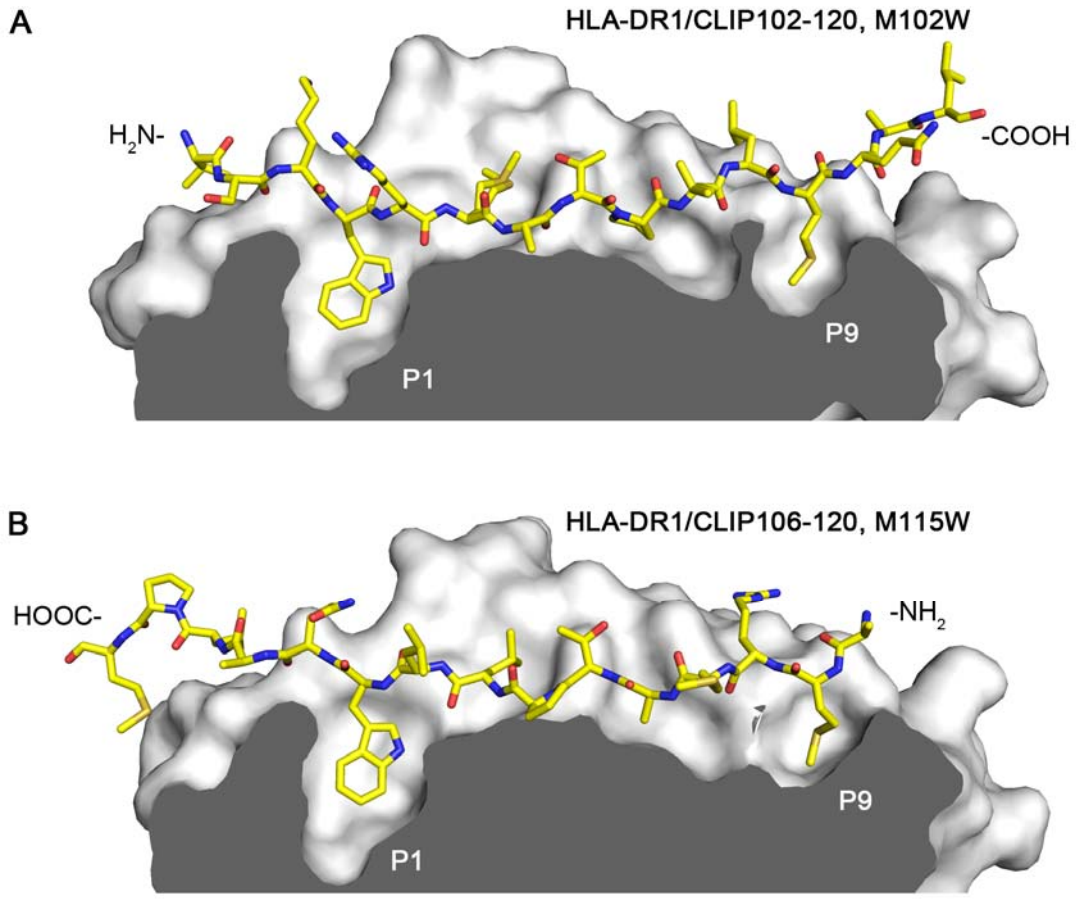


Figure 21: *Pocket P1 of HLA-DR1 allows accommodation of tryptophan in canonical and flipped binding orientation.* The lateral cut through the surface representation of HLA-DR1 reveals the asymmetrical size of the outermost peptide binding pockets P1 and P9. **A**, Crystal structure of CLIP102-120, M107W (yellow sticks) canonically bound to HLA-DR1. **B**, CLIP106-120, M115W reversely bound to HLA-DR1.

Important for the recognition of the pMHC by antibody or TCR are only those parts that are surface-exposed. If the binding-mode stabilized CLIP variants were to be used for the generation and selection of TCRs and antibodies, it is important that these complexes are identical to their wt counterparts except for the exchanged P1-anchor. For this reason, the surface-charge distribution of the mutated and wt complexes was compared as one main prerequisite to obtain antibodies and TCRs that are also specific for the two orientations of wt CLIP. Similar to the wt CLIP, this analysis illustrates the clear difference between the canonical and flipped binding orientation (Figure 22), but, as expected, corresponding wt and Trp-variants are identical. These mutant HLA-DR1 complexes could now be used as tools to generate antibodies and T cells specific for both directions.

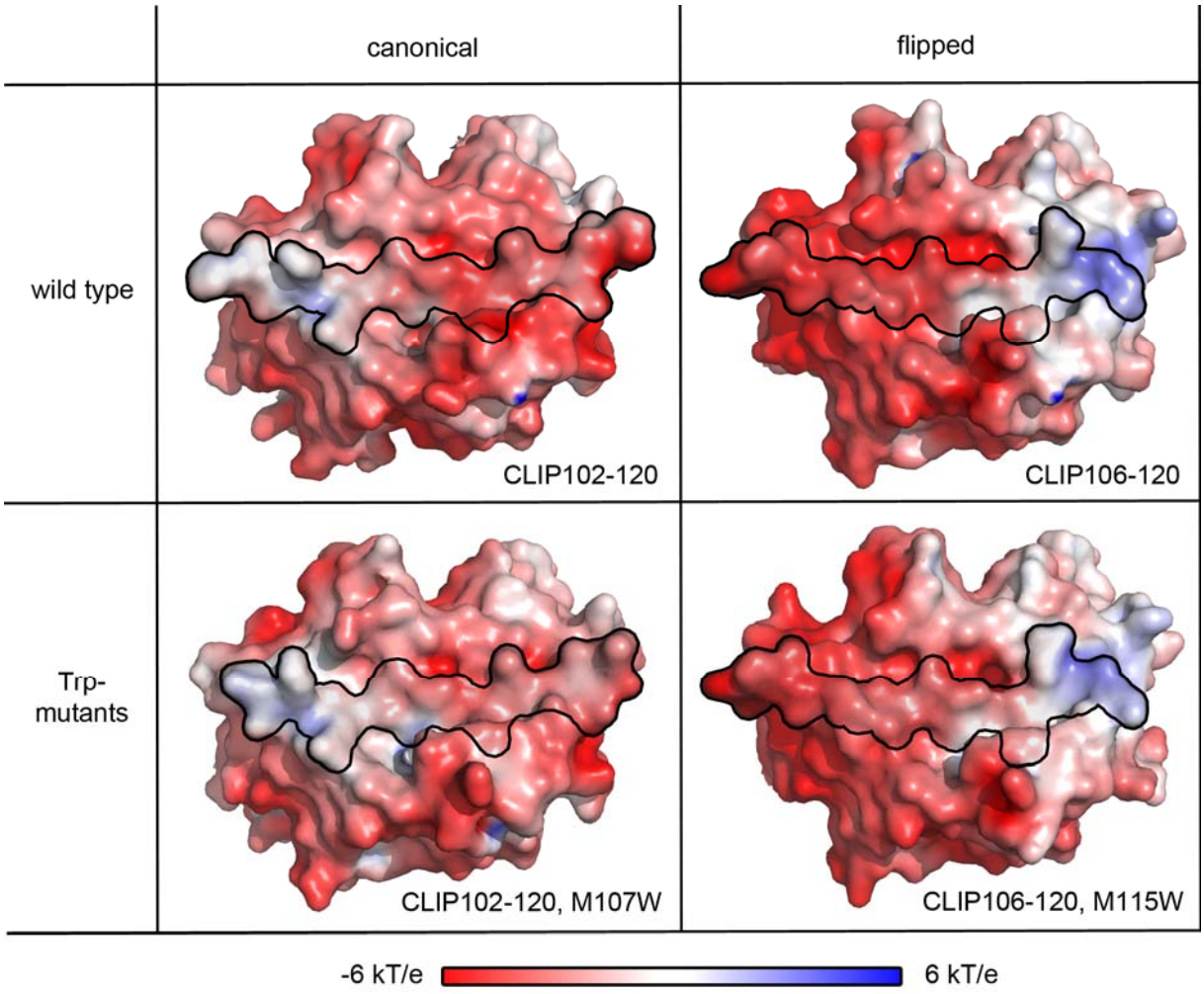


Figure 22: *Change of peptide direction results in reversed surface-charge distribution.* The charge distribution was calculated using APBS (Baker *et al.*, 2001) with red showing negatively- and blue positively-charged surface regions. Upper panel, HLA-DR1/CLIP102-120, canonical and HLA-DR1/CLIP106-120, flipped show differences in surface-charge caused by the differently oriented peptide. Lower panel, HLA-DR1/CLIP102-120, M107W and HLA-DR1/CLIP106-120, M115W show the same distribution as the corresponding wt structures. The view is down on the α1/β1-domains with the bound peptide indicated by a thin outline. Bottom, scale for surface-charge distribution from -6 to +6 kT/e.

3.3.3 Search for other reversely bound peptides

The structure of the reversely bound CLIP is the first proof for an MHC peptide ligand in this new binding mode. Although the interaction between peptides and MHC molecules has been investigated in many ways, reverse peptide binding has not been described before. Therefore, the question arises whether this binding mode is a common phenomenon. Up to now, the only methods to unequivocally determine the direction of the bound peptide are NMR-spectroscopy with spin-labeled peptides and structure determination by X-ray crystallography. As these approaches are not amenable to high throughput screening some peptide candidates were preselected for further analysis. Several databases are available with known MHC pep-

tide ligands. The Immune Epitope Database (IEDB) (Vita *et al.*, 2010) was chosen as starting point because it is the most comprehensive database available. This database lists known B and T cell epitopes and also contains information about MHC ligands that were defined e.g. by direct peptide elution from MHC or *in vitro* binding studies.

Table 14: Overview of peptides selected for closer analysis

No.	Peptide	Sequence	Antigen
I	tetX (TT 234-246)	ELIHVLHGLYGMQ	*Clostridium tetani* toxin
II	SMU.616 33-43	SILGGVATYGA	*Streptococcus mutans* uncharacterized protein
III	MAGEA3	GDNQIMPKAGLLIIV	Melanoma associated antigen 3
IV	pstS1 86-101	TGSGAGIAQAAAGTVNI	*Mycobacterium tubercolosis* posphate binding protein
V	HA1 255-270	RGYFKMRTGKSSIMRS	Influenza A virus hemagglutinin
VI	MBP 278-297	VDAQGTLSKIFKLGGRDSRS	Myelin basic protein
VII	P4a 293-307	SMRYQSLIPRLVEFF	vaccinia virus core protein 4a
VIII	Phlp5b 68-86	DKFKTFEAAFTSSSKAAAA	*Phleum pratense* pollen allergen
IX	lqpH 1-15	EHRVKRGLTVAVAGA	*Mycobacterium tubercolosis* 19 kDa lipoprotein
X	TT 830-843	QYIKANSKFIGITE	*Clostridium tetani* toxin
XI	TT 1147-1159	DYMYLTNAPSYTN	*Clostridium tetani* toxin
XII	Bapa 361-380	NANIRYVNTGTAPIYNVLPT	*Bacillus anthracis* protective antigen

A list of peptides was composed that had the following features: the peptides were HLA-DR1 (DRA*0101/DRB1*0101) restricted and T cells had been defined that are activated by the corresponding pMHC II complexes. This led to a list of 374 peptides (appendix). The list was extended by known T cell epitopes from the SYFPEITHI database (Rammensee *et al.*, 1999) not listed in the IEDB. In a first round five peptides were selected based on the known preferences for peptide binding to HLA-DR1 (Table 14, I-V), i.e. two hydrophobic amino acids at position 1 and 9 and a rather small amino acid (G, P, S, T) at the potential position 6. This reflects the known binding epitope of HLA-DR1 (Fleckenstein *et al.*, 1996; Hammer *et al.*, 1992). However, here the peptide sequences were scanned to a suitable motif from the C- to the N-terminus. Additionally, 41 peptides were also tested for reversed binding to HLA-DR1 by a SPOT assay (Kramer & Schneider-Mergener, 1998), where peptides were directly synthesized onto a cellulose membrane (data not shown, A. Schlundt private communication). Five additional peptides were then chosen based on their proven binding activity in the SPOT assay (Table 14, VI-X). Moreover, the known MHC II peptide binding prediction methods only consider the canonically bound peptides, therefore B. Rupp, FMP Berlin, developed a new HLA-DR1 binding algorithm. The scoring function behind this algorithm is based solely on the HLA-DR1 peptide binding motifs observed in crystal structures and does not account

for the direction of the peptide. The 374 previously mentioned peptides were scored by this algorithm and the top two peptides were selected for further testing (Table 14, XI and XII).

Complexes of all listed peptides with HLA-DR1 were generated by *a posteriori* loading onto empty HLA-DR1 overnight with the use of the MLE AdEtOH or the dipeptide Ac-FR-NH_2. Then, for all complexes $^{1}H/^{15}N$-HSQC spectra were recorded (A. Schlundt, FMP Berlin, data not shown). Subsequently, the unbound peptides were removed by gel filtration and the buffer was simultaneously exchanged from the NMR buffer (PBS, pH 5.8) to crystallization buffer (MES/Cl, pH 6.0). Depending on the available amount of protein at least one primary crystallization screen was carried out at RT. As in earlier crystallization of HLA-DR1/CLIP the focus was set to screens with a high number of PEG containing crystallization conditions (JCSG, PACT, PEG, PEG II, ProComplex from Qiagen, Germany). Only the complex formation with the peptides SMU.616 and pstS1 did not yield a sufficient amount of pMHC for screening, possibly because these peptides have a too low affinity for HLA-DR1. For peptides tetX, MAGEA3, HA1, MBP, and lqpH either crystals could not be grown or they did not diffract to sufficient quality. For the remaining five complexes crystals could be derived that were successfully used in diffraction experiments at the synchrotron source BESSY II at Helmholtz-Center, Berlin, beamline 14.1. The statistics for these five datasets are given in Table 15. All five structures could be readily solved by molecular replacement as described before. The resolution of the data sets ranged from 1.4 Å for HLA-DR1/TT1147-1159 to 2.4 Å for HLA-DR1/Phlp5b. The data set for HLA-DR1/p4a suffered from mild anisotropy (diffraction along a* 2.2 Å versus 1.9 Å along b* and c*). This is probably the cause for the slightly higher R-factors (R_{work} 21.4 %, R_{free} 26.3 %) as would be expected for this resolution and as compared to the other two data sets in this resolution range HLA-DR1/Bapa and HLA-DR1/TT830-843 (Table 15). The data set for HLA-DR1/TT1147-1156 was of high enough quality to allow anisotropic modeling of the atomic displacement parameter (ADP), which significantly improved both R-factors. All other models were refined with an isotropic ADP model. For the structure of HLA-DR1/Phlp5b68-86, the fourfold non-crystallographic symmetry was used as additional restraints.

During the initial rounds of refinement the peptide was not modeled. But in all five cases the peptide was modeled at later stages into unambiguous difference electron-density. In all cases the prominent P1 pocket is occupied by an aromatic side chain (4x Tyr, 1x Phe). Together with the electron density for other large side chains, e.g. Arg, Tyr, Phe, Lys, or also for smaller side chains, e.g. Gly, Ala, Pro, the direction of the peptides in the binding groove could be determined unambiguously (Figure 23). The quality of the electron density outside the core

binding region from P-2 to P9 is reduced so that not all amino acids of the peptide ligands can be observed. Especially at the C-terminal end, no amino acids can be observed beyond P11 (only in HLA-DR1/TT1147-1152 one more amino acids is observable). An overview of the individual binding registers can be found in Table 16.

Table 15: Data collection and refinement statistics for HLA-DR1/peptide complexes of potential reversely bound peptide candidates

	HLA-DR1/ TT147-1159	**HLA-DR1/ p4a293-307**	**HLA-DR1/ Bapa361-380**	**HLA-DR1/ TT830-843**	**HLA-DR1/ Phlp5b68-86**
Data collection					
Space group	R3	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	C2
Cell dimensions					
a, *b*, *c* (Å)	119.4, 119.4, 77.5	46.4, 97.9, 99.4	47.3, 97.5, 100.1	47.0, 97.8, 99.9	154.8, 131.8, 117.5
α, β, γ (°)	90, 90, 120	90, 90, 90	90, 90, 90	90, 90, 90	90, 115.1, 90
mol/asu	1	1	1	1	4
Resolution (Å)*	31.02-1.42 (1.46-1.42)	33.72-1.87 (1.92-1.87)	34.39-1.86 (1.91-1.86)	34.21-1.75 (1.80-1.75)	33.58-2.36 (2.42-2.36)
R_{sym}*	3.3 (48.0)	7.6 (46.0)	8.8 (79.7)	7.7 (67.0)	12.8 (75.2)
$I / \sigma I$*	16.9 (2.1)	11.5 (2.1)	12.1 (2.2)	15.0 (2.1)	9.5 (1.9)
Completeness (%)*	99.6 (99.6)	97.6 (89.9)	99.1 (99.9)	99.7 (99.4)	99.4 (99.2)
Redundancy*	2.8 (2.8)	3.9 (3.0)	4.5 (4.5)	4.4 (4.2)	4.1 (4.1)
Refinement					
Resolution (Å)	31.03-1.42	33.7-1.87	34.4-1.86	34.22-1.75	33.58-2.36
No. reflections	77512	37348	39366	47036	87206
R_{work} / R_{free}	13.97/17.96	21.37/26.32	18.40/22.47	17.18/20.72	19.33/23.43
No. atoms					
Protein	3199	3174	3168	3190	12192
Ligand/ion	18	-	6	12	-
Water	509	350	429	551	799
B-factors					
Protein	21.16	29.29	25.62	18.93	28.5
Ligand/ion	25.75	-	31.82	27.14	-
Water	33.07	34.39	34.83	30.98	30.52
R.m.s. deviations					
Bond lengths (Å)	0.005	0.007	0.007	0.006	0.008
Bond angles (°)	0.998	1.035	1.013	1.026	1.066
Ramachandran					
Favored (%)	98.4	97.9	98.7	98.7	97.4
Outlier (%)	0	0	0	0	0
Crystallization conditions	0.1 M lithium chloride, 25 % (w/v) PEG 6000, 0.1 M HEPES, pH 7.5	20.0 % (w/v) PEG 8000, 200 mM magnesium chloride, 100 mM Tris-HCl, pH 8.5	20.0 % (w/v) PEG 8000, 200 mM magnesium chloride, 100 mM Tris-HCl, pH 8.5	0.2M magnesium chloride, 20% (w/v) PEG 6000, 0.1M Tris, pH 8	2.00 M ammonium sulphate, 100 mM Bis-Tris, pH 5.5

*) values in parentheses correspond to the highest resolution shell

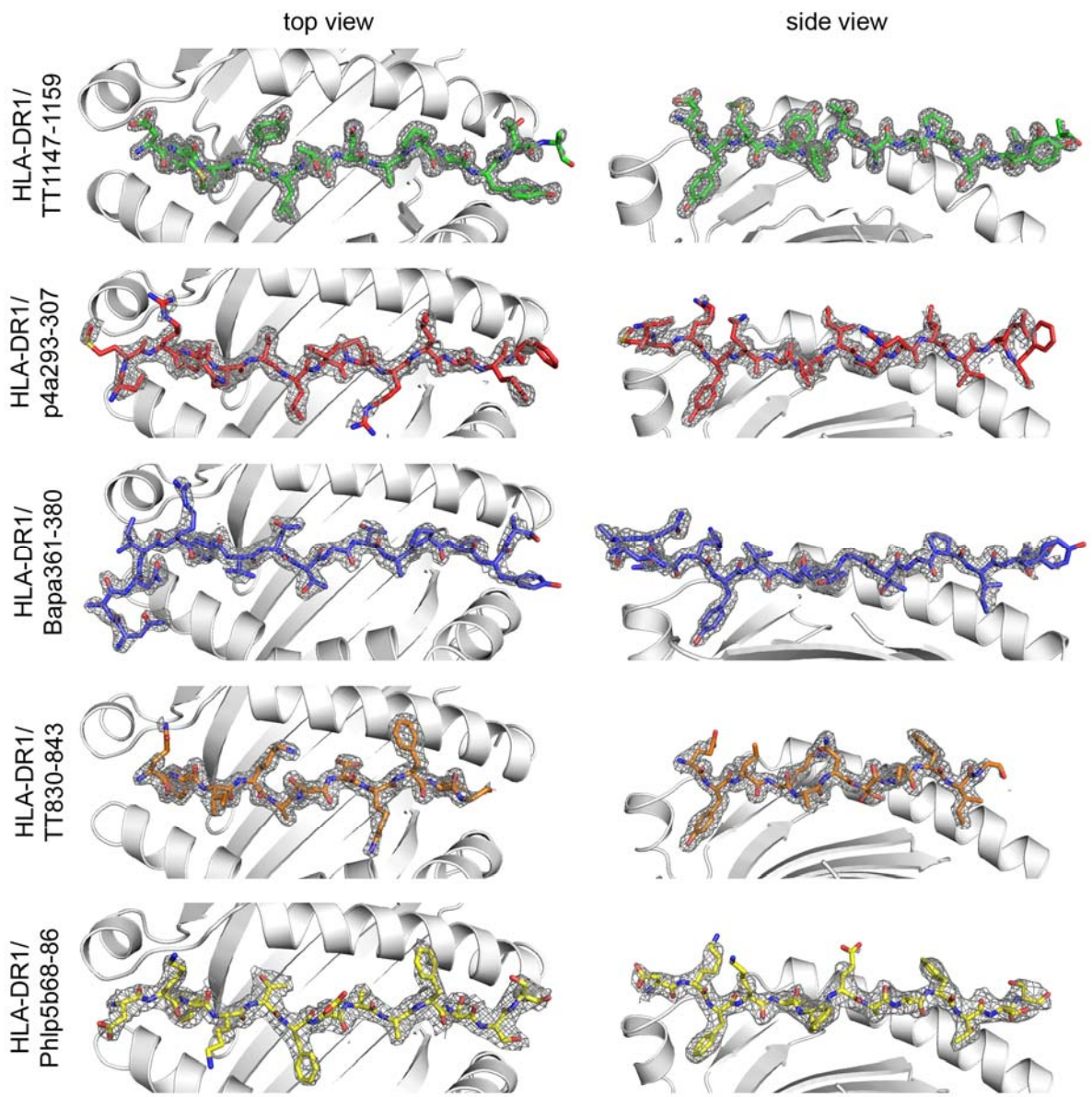


Figure 23: *X-ray structure of peptide MHC II complexes proves canonical binding mode for all five tested peptides.* Left panel, view top down into the antigen binding groove with only the α1/β1-domain of the MHC II depicted. Right panel, lateral view of binding groove with the α-chain α-helix in the background. The peptide is shown as colored stick together with the σ_A-weighted $2F_oF_c$-electron density, carved at 1.6 Å around the peptide and contoured at 0.9 σ. The electron density was calculated as averaged omit kick map (Praaenikar *et al.*, 2009) using the final model with the coordinates of the peptide removed beforehand.

While the accommodation of aromatic amino acids in the P1 pocket was not surprising, the asparagine in P6 of the TT1147-1159 peptide was not expected. In fact, based on the experimentally determined binding motif for HLA-DR1, it was known that small amino acids (G, A, S, P) are preferred at this position. And this was also the basis for the initial selection of peptides for further analysis of their binding mode. The plasticity of the antigen binding within the MHC peptide binding groove, however, had not been considered. Whereas the MHC side

chains are comparatively rigid and do not change significantly, instead the peptide is adapting to the binding site. Compared to other Cα-atoms at P6, the asparagine Cα shifts approximately 1 Å away from the pocket to compensate for the longer side chain. In contrast, the Cα of serine at P6 in the HLA-DR1/TT830-843 complex is pulled into the pocket to allow the deep insertion of the corresponding side chain into the P6 pocket (for comparison of Cα-positions see Figure 24A). The RMSD-plot of the five peptides observed in the presented structures reveals that regions within the antigen-binding site exist where the peptide is more restricted than at other positions (Figure 24A and B). Not surprisingly, the Cα-atoms at P-1 and P1 show very low variance. This is maybe caused in part by the nearly identical P1 anchors (Tyr and Phe). In addition, the P4 position shows low RMSD from the averaged Cα-position. And surprisingly, the P8 and also the P9 region show little variability. This is likely caused by the dense, conserved hydrogen bonding network around these regions (compare with overview in Figure 19B). Due to the hydrogen bonds to the peptide backbone, the position of these atoms is relatively fixed. The comparison of the RMSD for the MHC antigen-binding site (0.2 Å for the floor, 0.27 Å for the flanking helices, Cα only) with the average RMSD for the peptide (0.4 Å) indicates that the plasticity of antigen binding to the MHC II resides predominantely in the peptide itself, whereas the MHC is more rigid.

Noteworthy, the comparison of the five structures revealed also a new feature of the P7 pocket. Previous structures showed that this pocket can accommodate even large side chain like Tyr (Murthy & Stern, 1997). Here, we show that it can also fit basic residues like Lys and Arg. In the presented structures βL67 adopts two different conformations (Figure 24C). In case of a small amino acid like alanine at P7 (Bapa, Phlp5b, TT1147) the leucine points towards the antigen-binding site thereby closing the pocket 7. In contrast, in the two structures with arginine (p4a) and lysine (TT830) at P7, the leucine side chain points away from the groove and opens up the pocket 7. In these cases βL67 adopts a different rotamer and thereby enlarges the pocket 7 significantly (Figure 24C) so that both side chains fit in. Additionally, both head groups of the peptides form hydrogen bonds that stabilize the side chains (Figure 24C).

While all these ligands were found to bind in the canonical orientation, these structures underline the adaptability of the antigen-binding site whereby the major adjustments are found in the peptide itself.

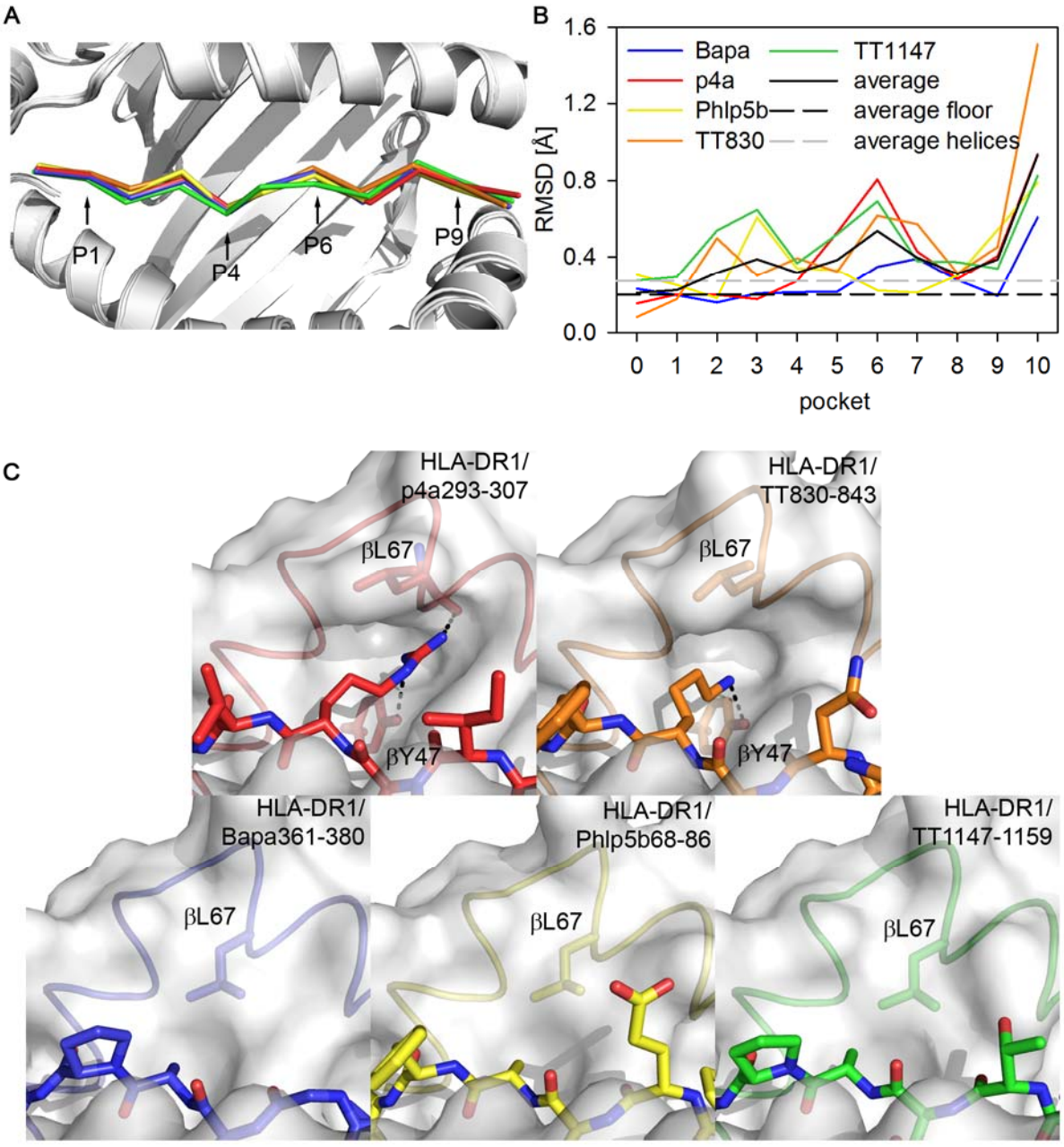


Figure 24: *Variability of peptide binding in the antigen-binding site.* **A**, The superposition of HLA-DR1/Bapa361-380 (blue), HLA-DR1/p4a293-307 (red), HLA-DR1/Phlp5b68-86 (yellow), HLA-DR1/TT830-843 (orange) and HLA-DR1/TT147-1159 (green) reveals the variability of the peptide in the binding groove. Only the α1/β1-domains are shown, with the bound peptide in the core region from P-1 to P10 (as Cα-trace). The major binding pockets are indicated. **B**, RMSD plot of individual peptide Cα-atoms along the binding site reveals points of higher flexibility in the peptide backbone. The RMSD was calculated after superposition of the β-sheets (α4-43, β9-50) and calculation of an average Cα-position. The average RMSD of the individual Cα-atoms is also given (black line). Additionally, the average Cα-RMSD for the floor (α4-43, β9-50) and helices (α51-80, β51-90) of the binding site are given as reference (black and grey dashed line). **C**, size of P7 pocket adapts to peptide side chain. Close-up view of the P7 pocket in the five determined structures with the peptide as stick model and part of the β-chain α-helix as loop structure and the transparent surface of the MHC II in white. The conformation of βL67 (as stick) determined the depth of the pocket. In the upper panel the rotamer of βL67 generates a deep pocket that accommodates lysine and arginine, side chain stabilizing hydrogen bonds are indicated as black, dashed lines. In the lower panel the pocket is closed by βL67.

Table 16: Alignment of the peptide ligands from the five newly determined HLA-DR1 structures with peptide ligands from all published X-ray structures of HLA-DR1/peptide complexes. In case of multiple structure determinations only the first published structure is given. The P1, P4, P6 and P9 anchors are written in bold from left to right.

Peptide	Sequence	PDB	Comment
TT 1147-1159	DY**M**Y**L**T**N**AP**S**YTN	this work	
p4a 293-307	SMR**Y**QS**L**I**P**RL**V**EFF	this work	
Bapa 360-381	NANIR**Y**VN**T**GT**A**P**I**YNVLPT	this work	
TT 830-843	Q**Y**IK**A**N**S**KF**I**GITE	this work	
Phl p 5b 68-86	DK**F**KT**F**E**A**AF**T**SSSKAAAA	this work	
HA 306-318	PK**Y**VK**Q**N**T**LK**L**AT	1DLH	
A2 103-117	VGSD**W**RF**L**R**G**YH**Q**YA	1AQD	
synthetic peptide	AA**Y**SD**Q**A**T**PL**L**LSPR	1T5W	
CII 259-273	AG**F**KGE**Q**G**G**PK**G**EPG	2FSE	
Gag[PP13]	PE**V**IPM**F**S**A**L**S**EG	1SJH	
Gag[PP16]	PE**V**IPM**F**S**A**L**S**EGATP	1SJE	
TPI 23-37, wt	GEL**I**GT**L**N**A**AK**V**PAD	1KLU	
TPI 23-37, T28I	GEL**I**GI**L**N**A**AK**V**PAD	1KLG	
pMART-1	APPA**Y**EK**L***S***A**EQ**S**PP	3L6F	*S* = phosphor-
synthetic peptide	Ac-**F**VK**Q**N**A***A***L**-NH2	1PYW	*A* = N-methyl
(4-Dapa)-HA 306-318	PK**X**VK**Q**N**T**LK**L**AT	2IPK	X = (4-Dapa)
CLIP 102-120	KPVSK**M**RM**A**T**P**LLMQALPM	3PDO, this work	
CLIP 106-120	K**M**RM**A**T**P**LLMQALPM	3PGD, this work	
CLIP 102-120, M107W	KPVSK**W**RM**A**T**P**LLMQALPM	this work	
CLIP 106-120, flipped	*MPLAQ**M**LL**P**TA**M**RM**K***	3PGC, this work	reversely bound
CLIP 106-120, M115W	*MPLAQ**W**LL**P**TA**M**RM**K***	this work	reversely bound

4 Discussion

The specific aim of this thesis was to identify and to characterize putative conformational variants of MHC II complexes. As one of the key results, the presented work indeed revealed important conformational flexibility in MHC II molecules, both ligand-dependent and -independent. Ligand-independent conformations of the empty MHC II were explored in detail by functional analysis, based on site-directed mutagenesis of residues suggested by an MD-derived model, whereas ligand-dependent conformations were investigated by X-ray crystallography of MHC II in complex with peptide length variants. While the first approach revealed indications of the structural basis for peptide-receptiveness of the MHC molecule, the second led to the unexpected discovery of an inverse peptide-binding mode.

4.1 Ligand-independent conformations

The current model for the empty state of HLA-DR1 suggests the existence of two clearly distinguishable conformations. Underlying is the observation that MHC II molecules that have just lost their ligand can be effectively loaded by free peptides (known as the receptive state). This feature is lost rather quickly rendering the empty molecule inaccessible for peptide binding (non-receptive state). The conformational transitions that correlate with the kinetically observed forms of empty MHC II are not known, but it has been speculated that these are caused by a closure or a collapse of the MHC peptide binding site (Joshi *et al.*, 2000; Rabinowitz *et al.*, 1998).

Using molecular dynamics (MD) simulations, we could derive a model of the underlying structural alterations in the empty MHC. Our model suggests that upon loss of the peptide the whole binding site is destabilized and the flanking helices exhibit higher flexibility. As a consequence, two of the three helical fragments that normally form the β-chain α-helix now form one consecutive helix (Figure 5A and Figure 25). Concomitantly, both flanking helices, especially the β-chain, move inward. This, in turn, basically closes the access to the P1 pocket. This now-non-receptive conformation is additionally stabilized by the formation of a new hydrogen bond between βN82 in the helix and αQ9 on the floor of the antigen binding groove (Figure 25). The presence of a peptide ligand in the binding groove prevents this closure, an effect also achieved with a short dipeptide targeting the P1 pocket of the MHC II molecule.

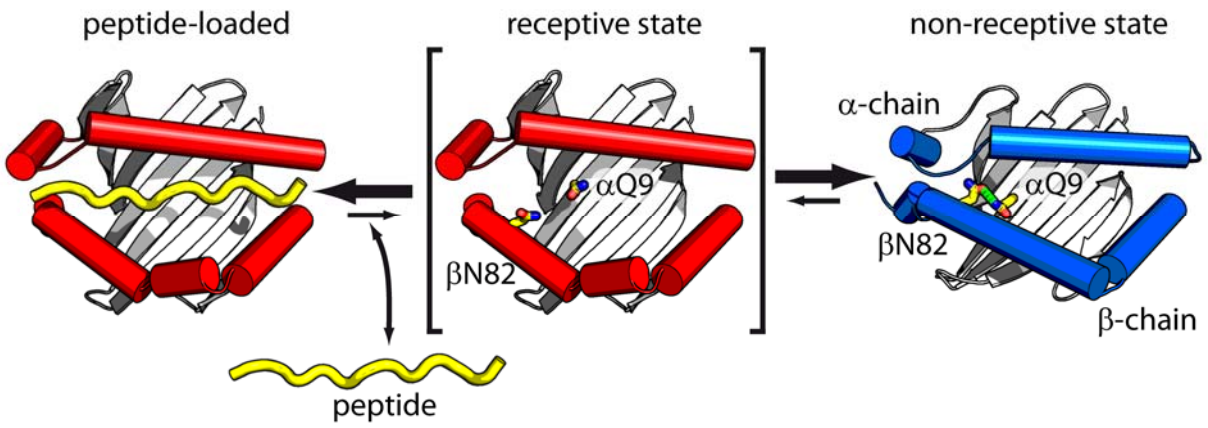


Figure 25: *Model for the non-receptive state of empty MHC II.* Upon loss of peptide ligand, empty MHC II is transiently in a state that is able to rapidly rebind peptides (receptive state). In the absence of suitable peptides, this state transforms into the non-receptive state, which only slowly binds peptides. Our *in silico*-derived model for this state suggests that the inability to bind peptides is caused by a closure of the antigen binding groove, mainly by straightening and inwards movement of the β-chain α-helix. This conformational change is accompanied by the formation of a new hydrogen bond (indicated in green) between βN82 on the helix and αQ9 on the floor of the antigen-binding site.

The model derived from the MD simulations could be validated experimentally by introducing point mutations at key residues involved in the formation and stabilization of the closed state. The introduction of an intramolecular disulfide bond that enabled the covalent cross-linking of the α- and β-chain supported the postulated conformational freedom allowing for the helix movement observed in the simulations. Even stronger evidence was provided for the suggested function of the hydrogen bond-"lock" formed between αQ9 and βN82 that is established after peptide ligand loss. Ablation of either αQ9 or βN82, which in the model of the closed state form a stabilizing hydrogen bond, indeed resulted in highly receptive MHC II species. In addition to these two directly involved residues, also residue αE11 was identified as critical component for the stabilization of the empty state. But instead of directly establishing a new hydrogen bond, this residue is involved in an elaborated hydrogen bonding cluster with αQ9, αN62 and αD66. The removal of the αE11 side chain is therefore expected to have a profound influence on the electrostatics in this region and on the strength of the αQ9/βN82 interaction, evident here in the increased receptiveness of the αE11A mutant. Notably, in the murine HLA-DR homolog I-E^k this cluster of hydrogen bonds was already discussed to be responsible for a conformational change from a peptide-exchange-susceptible to a more stable, insusceptible MHC conformation (Wilson *et al.*, 2001).

During the course of the thesis two similar MD studies were published by other groups (Painter *et al.*, 2008; Yaneva *et al.*, 2008). In the first study Painter and colleagues suggested that an extended portion of the α-chain α-helix, which flanks the antigen-binding site, par-

tially unfolds and occupies the P1 pocket mimicking a peptide ligand, an effect not observed in our simulations. Moreover, in their model the phenyl ring of αF54, which is part of the unfolded segment, binds into the P1 pocket and the conserved residues βR71, βH81 and βN82 form hydrogen bonds to the backbone of the unfolded segment like in a conventional pMHC II complex. Based on this, the mutation βR71A should have direct influence on the stabilization of the non-receptive state, whereas in our model this residue does not play an apparent role in stabilizing the non-receptive conformation. At least in our hands, this mutation had indeed no influence on peptide loading on empty MHC II, further supporting our model that involved less complex rearrangements of structural elements in the antigen binding groove (Figure 8). Zacharias and colleagues used a structure of HLA-DR3 to model the same three states (empty, peptide-loaded and dipeptide-stabilized) as in our study (Yaneva *et al.*, 2008). In initial simulations they observed a similar unfolding of the α-chain as Painter et al., but considered this as artifact because they could prevent this major loss of secondary structure elements by adjusting the protonation state of amino acids in this region. Instead, similar to our model, they observed an overall narrowing of the antigen-binding site flanking helices with mobility in both chains.

The closure of the P1 pocket by the flanking α-helices can be prevented by pocket occupation. In our simulations the P1-interaction with a dipeptide stabilizes the open conformation of the empty molecule and allows easy access to the binding groove, providing a rationale for the catalytic activity of peptide-MLE (Gupta *et al.*, 2008). This was also shown in the study by Yaneva *et al.*, where the docking of a dipeptide in the P1 pocket prevented the closed conformation much like in our model (Yaneva *et al.*, 2008). Recently, another MD study used a model of the empty HLA-DR1 docked with an adamantyl-derivative in the P1 pocket (Zaheer ul & Khan, 2011) that supported earlier experimental observations on the catalyzing effect of AdEtOH on HLA-DR1 peptide loading (Höpner *et al.*, 2006). Again, the occupation of the P1 pocket stabilized the entire binding site. In support of our work, Zaheer ul and Khan also simulated empty HLA-DR1, where, in contrast to the Painter study, no major loss of secondary structure in the P1 pocket flanking α-chain α-helix was reported.

While βN82 apparently plays a key role in stabilizing the closed form of the empty molecule, it is equally important in fixing the peptide ligand in the loaded complex. A number of prior studies already reported the contribution of single amino acids to the overall stability of pMHC II including the murine I-A^d (McFarland *et al.*, 1999; McFarland *et al.*, 2001; McFarland *et al.*, 2005) and I-E^k (Saito *et al.*, 2004) as well as the human HLA-DR1

(Anderson & Gorski, 2005; Narayan *et al.*, 2007). In all cases the investigated amino acids contributed to the conserved hydrogen bonding network between the MHC and the bound peptide. Especially βN82 was shown to have an important role, as the mutation βN82S in the murine MHC II I-A^d and I-E^d decreased the peptide/MHC complex stability significantly (Griffith *et al.*, 1988; Sant *et al.*, 1999). Concomitant defects in the transport of these mutants to the cell surface (Tan *et al.*, 1997) could be recovered by binding of high affine peptides (Arneson *et al.*, 2001). In all cases the observed effect was attributed to the interaction of the βN82 side chain with the peptide. Here, we identified another important function of βN82 in stabilizing the empty MHC II by interaction with residues from the floor of the antigen-binding site. While insect-cell derived HLA-DR1 can be preloaded with peptides from the expression host (Hartman *et al.*, 2010), the peptide-free HLA-DR1 from *E. coli* used here established the importance of single residues for the stability of the empty MHC II without any bias.

The residues αQ9, αE11 and βN82 that form the proposed hydrogen bonding lock of the non-receptive form are highly conserved in all human HLA-DR and DR-like MHC molecules in other species (mice, rat, swine, cattle, and sheep). In the other two human MHC II, HLA-DQ and HLA-DP, βN82 is also ubiquitously present but the corresponding α-chain residues are less well conserved. Thus, the overall mechanism controlling the receptiveness in these alleles might be similar to the one observed for HLA-DR, but differ in detail. Certainly, the hydrogen bond between αQ9 and βN82 will not be the only interaction stabilizing the non-receptive state. The movement of the antigen-binding site flanking helices will allow for further new interactions. Moreover, the presented non-receptive state of empty MHC may not be the end-point of conformational change. But larger rearrangements are unlikely to be observed within the time scale of the simulation. However, the repetition of individual simulations as conducted in our study is expected to cover a wider spectrum of the folding landscape than single simulations as was done in the other reports (Painter *et al.*, 2008; Yaneva *et al.*, 2008; Zaheer ul & Khan, 2011).

In addition to the conclusions derived from the data presented here with soluble HLA-DR1, the same mutants were also analyzed in full length HLA-DR1. When introduced into murine fibroblast cells and presented on the cell surface, the HLA-DR1 mutants had the same effect on the peptide-loading behavior as the soluble proteins (data not shown). Thus, the identified inactivation mechanism of soluble, empty MHC II may also be valid in a more natural setting. If empty MHC II molecules are presented at the cell surface of antigen-presenting cells, they

could bind any suitable antigen from the extracellular milieu, including potentially harmful self-antigens. This extracellular peptide loading pathway has actually been described for immature dendritic cells (DC) as an alternative pathway playing a particular role in tolerance induction (Clement *et al.*, 2011; Potolicchio *et al.*, 2005; Santambrogio *et al.*, 1999a; Santambrogio *et al.*, 1999b). Under certain circumstances, however, this might cause unwanted immune responses. If happening on immature DC, it circumvents the classical, controlled uptake mechanisms like Fc- or scavenger-receptor mediated endocytosis and is certainly different from the antigen-processing and presentation pathway in thymic epithelial cells during T cell selection. Thus, the rapid inactivation of empty MHC II on the cell surface could be seen here as safeguard-mechanism preventing this uncontrolled peptide binding.

However, this hypothesis has to be proven. Interestingly, the group of Unanue established a model where the route of antigen encounter decided which of two T cell subsets specific for the same peptide/MHC complex is activated. Type A T cells react against peptides that are processed intracellularly after antigen uptake whereas type B T cells are activated only when the peptide is loaded onto the MHC II extracellularly or in early endosomes (Lovitch & Unanue, 2005). Importantly, the group could show that in the Non-Obese Diabetic (NOD) mouse model both T cells can be found in diabetogenic mice and that type B T cells, recognizing the antigen only after "non-classical" extracellular loading of MHC molecules, induced diabetes after adoptive transfer (Mohan *et al.*, 2010).

Taken together, we have strong evidence from mutational analyses that support our MD-derived model for the empty, non-receptive conformation of MHC II, which displays a closed antigen binding groove with distinct structural features stabilized by a new hydrogen bond.

4.2 Ligand-dependent conformations

A number of studies suggested that the conformation of MHC II molecules is also affected by the bound peptide. In this thesis two different ways were analyzed by which the bound peptide ligand influences the overall conformation of MHC II. These ways were related, firstly, by the length of the peptide and, secondly, by the orientation of the peptide ligand.

4.2.1 The influence of peptide extensions on the MHC II conformation

The impact of the ligand length on the conformation was initially reported by Rötzschke et al. (Rotzschke *et al.*, 1999). Based on this report, the structural basis for conformational variants of HLA-DR1 in complex with long peptide ligands was investigated. Using a combination of tetraethylene glycol-linked peptide fragments, Rötzschke and colleagues estimated that a minimal peptide overhang of more than 5 but less than 10 amino acids at a distance of 7 to 12 amino acids counting from the P1 residues at the N- and the P9 residue at the C-terminus was required to induce two different "F-conformations". These conformations were initially characterized by a faster migration on SDS-PAGE. The faster migration suggested a more compact conformation resulting in smaller hydrodynamic radii of the complexes. In contrast to this, size exclusion chromatography indicated increasing apparent hydrodynamic radii with increasing ligand length. While in both cases the size of the peptide overhang was apparently responsible for the effect, no specific structural information on the nature of the F-conformations was available.

In this thesis the effect of the peptide overhang was closer analyzed using progressive amino acid extensions of the natural HA306-318 peptide ligand. Notably, the N-terminal extension of the peptide did not lead to a single F-form in SDS-PAGE, but instead, an additional faster migrating F-form was already induced after extension by only two amino acids, i.e. a four amino acid ($+4_N$) overhang before the P1-residue. Any further extension beyond $+8_N$ led to a less SDS-stable form. And extensions bigger than $+13_N$ finally caused a complete loss of SDS-stability. Thus, a clear singular F-form could not be reproduced with the N-terminal extended peptides. In contrast, the C-terminal extension revealed several F-forms that migrated gradually faster until a maximum was reached with a $+14_C$ extended HA (HA306-330). Interestingly, the first F-form was already induced by a $+6_C$ overhang (HA306-322). Notably, another report described the same effect for a different peptide bound to HLA-DR1. The complex with tetanus toxin (TT) 830-843 runs as C-form, but already the extension by two amino acids (TT830-845) increased the migration speed in an SDS-PAGE drastically (Georges *et al.*, 2000).

Because the exact binding register for the tetanus peptide was only predicted from the preferred binding motif of HLA-DR1, only the crystal structure of HLA-DR1/TT830-843, which was solved for this thesis to analyze the peptide's orientation, enabled the exact determination of the peptide overhang that induced the changed SDS-migration behavior. The comparison

with the structure of HLA-DR1/TT830-843 showed that the shortest necessary overhang of the tetanus peptide was also $+6_C$, exactly like for the HA peptide. In contrast, Rötzschke *et al.* postulated a minimal necessary overhang of 5-10 amino acids at a distance of 5-10 amino acids on either side of the peptide ligand, which is considerably longer than the length determined here. However, their ligands contained a teg-linker which probably does not have the same properties as an polypeptide.

Despite the changed SDS-PAGE migration, no other assays revealed any further surprising characteristics of the F-forms. As expected, the hydrodynamic radius determined by SEC increased slightly with the length of the peptide ligand. The comparable relative affinity of the different peptide-extensions did not indicate an altered binding behavior. This assured that the core binding motif that makes up for the affinity-determining interactions with the MHC II is preserved.

In order to directly analyze potential conformational changes in the MHC going along with the increasing ligand length, the crystal structures of four HLA-DR1/HA size variants were solved. However, the structures did not reveal any differences between complexes adopting the F- and C-conformations. Moreover, the peptide extensions were not visible in the electron density maps indicating high flexibility and no specific interaction with the MHC. Since it cannot be excluded that the crystal packing imposed a uniform conformation of the MHC II incompatible with conformational changes, it was also tried to cocrystallize HLA-DR1/HA306-330 together with the superantigen SEC3,3B2 (Sundberg *et al.*, 2003). This should have led to a different crystal packing. However, the crystallization attempts remained unsuccessful. Also a lack of electron density for overhang residues has been observed before. For example, for the first MHC II X-ray structure HLA-DR1 was used that was directly purified from a lymphoblastoid cell line, containing a mixture of endogenous peptide ligands (Jardetzky *et al.*, 1994). Because the electron density for the individual side chains was averaged out by the heterogeneity of the ligand, only a polyalanine model could be built for the peptide. Moreover, the electron-density for the C-terminus was bifurcated at position 11 indicating multiple conformers of the overhang that averaged out leading to no interpretable electron density (Jardetzky *et al.*, 1996).

Furthermore, the comparison of all available MHC II structures in the PDB shows that for crystallization mostly peptides were used that were not much longer than the core binding region (71% were between 12 and 15 amino acids long, i.e. an overhang of usually one to three amino acids; see also appendix). The overhangs of longer peptides are usually not re-

solved unless they are stabilized by contacts to other molecules. Tthe last residue of the peptide that is bound to the MHC is the P10 residues, which is hydrogen bonding to αR76 and βD57 in HLA-DR1. Beyond this point, no conserved interactions have been identified. However, in some structures a large overhang was resolved (appendix) when it was part of a linker covalently coupling the peptide to the β-chain of either MHC or TCR (Fremont *et al.*, 2002; Fremont *et al.*, 1996; Kersh *et al.*, 2001; Liu *et al.*, 2002; Wilson *et al.*, 2001). This is a common technique to stabilize peptide/MHC/TCR or peptide/MHC and, to a certain extend, mimics the situation of a C-terminally prolonged peptide ligand. However, these structured extensions did not induce any obvious structural changes in the MHC, either.

In summary, no major changes could be identified from the crystal structures solved with the HA-derivatives or any evidence be found by analyzing available structures in the PDB. This leaves the question of the cause for the significant changes in SDS-PAGE migration. A plausible explanation might be a disproportionally increasing SDS-binding capacity of the additional amino acids in the peptide-overhang under the assay condition (non-reduced, non-boiled samples). It is known that single point mutations can lead to distinct SDS-PAGE migration behavior as was shown for α-crystallin (de Jong *et al.*, 1978). This was mainly attributed to different hydropathy of amino acids and their ability to bind SDS molecules but charged amino acids might also directly influence the migration behavior.

While there was no structural evidence for conformational variants induced by peptide ligand overhangs, nevertheless the question remains if the MHC itself can “sense” the length of its bound peptide. Peptides presented by MHC II on the cell surface have an average length of 13-18 amino acids and core peptides with variations on both termini can be found (Rudensky *et al.*, 1992; Rudensky *et al.*, 1991). But it is known that MHC II ligands do not have a length restriction and even whole proteins can be bound (Lindner & Unanue, 1996). The final maturation signal of the MHC that enables transport to the cell surface is not known. It is conceivable that the observed average length of MHC II ligands is simply a consequence of the proteolysis by endo- and exopeptidases in the MIIC, when the core of the peptide is protected by binding to the MHC and the overhang is prone to degradation (Sercarz & Maverakis, 2003). Although purely speculative, another possible explanation might be that the MHC changes its conformation, once the peptide has reached the optimal length. Yet another potential role for ligand length dependent structural changes might be the quality control of pMHC II complexes on the cell surface. It could function as a safeguard mechanism in the case that an empty MHC II binds a ligand at the cell surface. As described earlier this cell surface loading

is independent of the intracellular antigen processing machinery (Santambrogio *et al.*, 1999a) and therefore, binding of non-trimmed ligands could be sensed by the APC via a conformational change in the MHC II. However, this is speculative and no evidence has been reported so far.

4.2.2 The influence of peptide orientation on the MHC II conformation

4.2.2.1 The discovery of a novel peptide binding mode for CLIP

Another aspect of ligand-dependent conformational variation of pMHC, that has not been considered yet, is peptide orientation. In the second part of the studies of ligand-induced MHC II conformational variants, the structures of HLA-DR1 associated with the invariant chain-derived CLIP revealed a surprising, new conformation of pMHC II complexes. Since the publication of the first MHC II crystal structure (Brown *et al.*, 1993), more than thirty unique peptide/MHC II complex structures have been reported (appendix). However, all exhibited the same canonical binding mode of the peptide ligand with the amino terminus near the P1 pocket and the carboxy terminus near the P9 pocket. The structure of HLA-DR1/CLIP106-120,flipped, presented here, is the first to indicate the existence of a completely new binding mode with the peptide lying in the binding cleft in a reversed orientation.

The canonical orientation of peptides in any MHC binding groove has never been questioned before. For MHC I there is no ambiguity in the peptide binding direction. Although in the first X-ray structures of MHC I the identity of the peptide could not be resolved (Bjorkman *et al.*, 1987), it was clear from later studies that the termini of the peptide are tightly bound to conserved residues of the MHC I via hydrogen bonds (Madden *et al.*, 1993), which allows only one direction for the bound peptide. In contrast, in MHC II the ends of the peptide binding groove are open and the peptide can extend significantly to both sides. In the first structure of an MHC II the peptide ligand was a mixture of self-peptides and their structure could not be exactly determined (Brown *et al.*, 1993). In a later structure in complex with the superantigen SEB, the resolution was improved and a polyalanine peptide model could be built into the averaged density of the heterogeneous ligands (Jardetzky *et al.*, 1994). At the same time, the first structure of an MHC II with a single peptide bound was solved in the same laboratory (Stern & Wiley, 1994) with the peptide lying clearly in the canonical orientation. This information helped to build the polyalanine peptide model for the HLA-DR1/SEB complex, certainly influencing the interpretation of the X-ray data. Later on, a lot of structures were also

solved by covalently linking the peptide to the MHC β-chain, which allows only the canonical binding mode (Fremont *et al.*, 1996; Kozono *et al.*, 1994).

Interestingly, the principle possibility of a reoriented peptide in the MHC II binding groove has already been discussed in an *in silico* modeling study. Here, HLA-DQ3 has been proposed to bind to a GAD-peptide in the reversed orientation (DeWeese *et al.*, 1996). However, direct experimental evidence has not been reported. The structure of HLA-DR1/CLIP106-120,flipped is the first proof for the existence of this reversed binding mode. Notably, no structural rearrangements of MHC II side chains are necessary to establish the conserved hydrogen bonding network also in the reversed binding orientation. The geometry of most hydrogen bonds is maintained, only the bidentate hydrogen bonds formed by αQ9, αN62, αN69 and βN82 adapt to the reoriented peptide backbone by switching between 9- and 11-membered ring structures (Le Questel *et al.*, 1993) (Figure 19B). Seemingly, only the favorable placement of side chains in the pockets of the MHC binding groove determines successful binding as long as the peptide ligand is long enough to satisfy all hydrogen bonds.

With the proof for the principle existence of a flipped binding mode *in vitro*, it remains to be shown that inverted MHC II/CLIP complexes also occur *in vivo*. During biosynthesis of MHC II, it might well be that the canonical binding orientation is the preferred because of interactions of the invariant chain with MHC II at a secondary site (Neumann & Koch, 2006; Wilson *et al.*, 1998) or due to steric hindrances by formation of the supertrimer of $(Ii/MHC\ II)_3$, whose overall architecture is not known in detail yet (Jasanoff *et al.*, 1998). Notably, for the structure determination of HLA-DR3/CLIP complexes were used that were directly cleaved from the cell surface of a cell line with a defect in peptide exchange of CLIP. Thus in principle, these complexes reflect the natural intermediate before CLIP is exchanged for other antigens. Interestingly, these naturally derived complexes showed the canonical binding mode, too (Ghosh *et al.*, 1995). There are two main reasons why no reversely bound CLIP was observed in this case and why these complexes reflect probably not a kinetically trapped folding intermediate, as in the case of HLA-DR1, but the thermodynamically stable complex. Firstly, in HLA-DR3 a reversed binding of the peptide is probably disfavored by the length of the bound peptides enabling critical H bond formation at the N-terminus and, secondly, also by the size of pocket 4. HLA-DR3 has an arginine at position β74 that would probably interfere with the proline in the flipped CLIP at this position, whereas HLA-DR1 has an alanine at β74. Although rather unlikely, it raises the question if the reversed binding orientation of CLIP might be a peculiarity of HLA-DR1.

As mentioned earlier, the flipping might also be a consequence of the pseudosymmetric motif in the CLIP sequence. As already proposed by Wiley and colleagues (Ghosh *et al.*, 1995), CLIP might have evolved to fit to any MHC II binding groove by a motif fulfilling the basic requirements for all MHC II variants. The amino acids in the core binding region are optimal for all MHC II motifs, i.e. the four large side chains Met107, Met109, Leu113 and Met115 are placed in those pockets that show the highest conservation in the different allelelic variants of MHC II, whereas the two pockets with the highest variability, P4 and P6, are occupied by Ala110 and Pro112, two residues with small side chains (Ghosh *et al.*, 1995). And exactly this degeneracy prevents the preference for one binding direction, at least in the case of HLA-DR1. The two methionines Met107 and Met115 can be both accommodated in either the P1 or the P9 pocket and the other two anchoring residues at P4 and P6, Ala112 and Pro114, are chemically similar enough to allow for both possible binding orientations. The only notable difference is the loss of a hydrogen bond in case of the canonically oriented proline in P6. At this point, it is not clear if this is just a consequence of the degenerated supermotif of CLIP or if CLIP has evolved to bind in any possible direction as soon as it is freed from Ii. This potential bidirectional binding enhances the chances to form a stable pMHC II when the empty MHC II encounters CLIP. This can be important to prevent MHC molecules from destruction because MHC II molecules devoid of any bound antigen are rapidly degraded in the MHC II loading compartment (Ceman *et al.*, 1998; Germain, 1995).

The general relevance of CLIP as immune-modulatory self-antigen *in vivo* has already been reported. During maturation of DCs, the level of cell surface-presented CLIP is increased and polarizes naïve T cells towards a T_H2 response (Pezeshki *et al.*, 2011; Rohn *et al.*, 2004). Analysis of the corresponding MHC-presented peptidome shows the prevalence of short CLIP sequences that, based on our findings, are preferably presented in an inversed state (Rohn *et al.*, 2004). Thus, it will be important to develop the necessary tools for the *in vivo* analysis of the CLIP binding direction to see f the flipped CLIP is indeed important for steering immune responses.

4.2.2.2 The relevance of an inverted peptide binding mode for other antigens

As the inverted peptide orientation so far has only been shown for CLIP, the question remains if it is restricted to CLIP or is of more general validity for other antigens as well. Searching for potentially inverted pMHC II ligands, we selected twelve candidate peptides. Only five of

these complexes yielded crystal structures. Thus, it cannot be excluded that some of the unsolved complexes do bind in a reversed orientation. However, the solved structures indicated that the antigen-binding groove allows more peptides to bind canonically than initially expected. The opening of the pocket 7 and the flexibility in the positioning of the peptide main chain in the groove are examples for this (Figure 24). Several studies proved the disproportionally high contribution that a favorable interaction at P1 has on the HLA-DR1/peptide stability (Ferrante & Gorski, 2007; Jardetzky *et al.*, 1990; Natarajan *et al.*, 1999). In our structures, the placement of an aromatic P1 anchor might counterbalance other less favorable interactions, like a bigger amino acid at pocket 6. With the current knowledge it is difficult to reliably predict the contribution of the conserved hydrogen-bonding network in the canonical or reversed orientation. Also, it cannot be excluded that the binding motif of a reversely bound antigen is different from the canonically bound one because some amino acids would have to adapt unfavorable rotamer conformations. However, the example of the reversely bound CLIP proves that a reversed binding mode is principally possible and that no obvious reasons exist that this binding mode is exploited by nature more commonly than currently recognized.

From an immunologic point of view, there are no reasons why a reversely bound peptide should not be functional. During thymic selection the T cell receptor is trained to have a basal affinity for MHC (positive selection) but not to react strongly against self-peptide/MHC complexes (negative selection). It has been proposed that the TCR docks onto the MHC guided by germline encoded amino acids thereby retaining a conserved, diagonal docking mode onto MHC (Garcia *et al.*, 2009; Scott-Browne *et al.*, 2009). In addition, the TCR interacts with the antigenic peptide via its variable regions. These interactions are often mediated by the side chains of the bound peptide but can also be mediated by main chain contacts (Rudolph *et al.*, 2006). In both binding orientations the same side chains are available for contact with the TCR (mainly P 2, 5, 8). Depending on the sequence of the peptide, a flipping of the antigen generates two largely different binding surfaces for the TCR (Figure 22. This results in two easily discernible pMHC complexes. However, conserved interactions of the TCR with the backbone of the bound peptide allowing direct discrimination of the binding orientation of the peptide relative to the MHC were not detected. Thus, for the immune system the possibility to place peptides in two orientations in the MHC binding groove enhances the possibility for any given antigen to contain a suitable motif for presentation by MHC II.

How could this reverse binding mode remain unnoticed? Up to now, the only direct method to observe the directionality of the peptide bound to the MHC II is X-ray crystallography. Here, on the one hand, the selection of pMHC II complexes might have been biased towards com-

plexes with a canonically expected binding motif, simply because these were predicted to have a higher stability. On the other hand, almost half of the crystal structures were solved with covalently attached peptides. In this case only one peptide orientation is possible. But some experimental hints for the reversed binding mode already existed in literature. For example, in an effort to enhance the stability of antigens in peptide-based vaccinations, antigenic sequences were replaced by D-amino acids and synthesized in a reversed order. These "retro-inverso" peptides were expected to occupy the same pockets with the same conformations of the respective side chains as the natural L-amino acid-based ligands. Only the direction of the peptide backbone would be reversed. In one example these peptides retained binding capacity to the MHC II, although with a reduced affinity (Bartnes *et al.*, 1997). In contrast, for another antigen no binding affinity was detected (Herve *et al.*, 1997). This might indicate that the hydrogen bonding network formed in both directions (as observed for HLA-DR1/CLIP) does not contribute equally to the overall stability.

Another highly interesting case is the earlier mentioned concept of peptide isomerism as reported by the group of Unanue. The same effects that enhance the reorientation of the canonically bound CLIP, i.e. N-terminal overhang and HLA-DM activity, influence as well the activation of type A and B T cells (Lovitch *et al.*, 2006). But reorientation of the peptide is not the only possible explanation for this effect. Another likely explanation is shifting of the peptide within the antigen-binding site, which was shown for exactly the same pMHC II for which peptide isomers have also been reported (Stadinski *et al.*, 2010). The HLA-DM-dependent generation of different isomers might also be attributed to pH-dependent motif changes of the described MHC II alleles (Munz *et al.*, 2002).

Finally, another field of study that will likely be affected by the results of this thesis is the computational prediction of MHC II peptide binders. For MHC I these algorithms have already reached a high success rate. In contrast, the predictive power of MHC II algorithms is still far from optimal (Lin *et al.*, 2008; Wang *et al.*, 2008). Certainly, a major problem is the uncertainty of binding modes due to translational invariance, i.e. shifting of the peptide (Fleckenstein *et al.*, 1999). The previously unrecognized reversed binding mode adds another level of uncertainty on top. However, once the rules of the reversed binding mode are better understood, its incorporation in the established algorithms should improve their success rate significantly.

In summary, the structures of HLA-DR1 determined during the course of this thesis elucidate the mechanisms that lead to enhanced stability of pMHC complexes. The ability of the MHC

II to bind different antigens is based less on conformational flexibility of the MHC itself but more on the adaptability of the same interaction strategies to antigens with different sequences and also orientations. This adaptability of the MHC II is likely the consequence of the its role in the adaptive immune system. It has to bind to as many peptides as possible, i.e. it needs to have a low specificity. But at the same time the peptide-MHC complexes need to be highly stable to be presented to T cells long enough to ensure a proper immune response. Failure in this process can not only lead to uncontrolled infections but is also believed to be a central element of autoimmune disease. Self-antigens of low affinity for the respective MHC are improperly presented in the thymus enabling self-reactive T cells to escape negative selection (Liu *et al.*, 1995). The identification of such low affinity peptides that can break the central tolerance is of pivotal importance for the study of these autoimmune diseases. Consideration of the orientational promiscuity of peptide ligands as described in this thesis is expected to have a significant contribution to these investigations.

5 Outlook

Because the antigen presentation via MHC molecules is of such pivotal importance for the initiation of an adaptive immune response, it is an attractive target for immune-modulatory therapies. Therefore, the insight into the different peptide-ligand dependent and independent conformations obtained in the presented work can have impact on different areas of immunological research.

Vaccinations have significantly improved the human fight against pathogens. Classically, whole pathogens, attenuated or dead, or complete protein antigens are administered. However, these strategies can also have unwanted side reactions caused for example by copurified pathogenic or suppressive antigens. Here, the application of the MHC peptide epitopes instead of the complete antigen has clear advantages. However, the bottleneck hereby is the availability of empty and peptide-receptive MHC molecules. A handful of small molecules have been identified that can enhance the extracellular loading of MHC molecules (so called molecular loading enhancer, MLE) and already been tested in proof of concept studies in mice (Call *et al.*, 2009; Dickhaut *et al.*, 2009). Unfortunately, their activity requires high dosages. Here, the knowledge of the underlying structural features of the receptive and non-receptive state, as presented in this work, can be used in structure-guided drug design with the aim to create more potent MLEs. The presented work based on *in silico* data combined with *in vitro* mutagenesis data gave good indications for the conformational space covered by the empty MHC II.

Following from the identification of a new peptide binding mode for MHC II in the second part of the thesis, one of the most important next steps will be the proof for a more general validity of this binding mode. The design of targeted orientation-specific CLIP ligands enables the development of the necessary tools to analyze the prevalence of the flipped HLA-DR1/CLIP complexes also *in vivo*. These complexes are currently used for T cell generation in HLA-DR1 transgenic mice and for phage-display generation of monoclonal antibodies. However, it will be equally important to identify further reversely bound peptides. The failed attempts to identify these in available databases indicated that our current knowledge of the prerequisites for the new binding mode is limited. For this reason, one of the next steps has to be to thoroughly decipher the contributions of sequence-independent hydrogen bonds in both binding directions and to analyze the possibility of a change in the preferred binding motif of a reversed epitope as compared to the canonical. The limiting step in the identification of such peptides is the final verification of their binding mode, which has to be done by X-ray crystal-

lography or NMR spectroscopy. Development of other methods for the identification of peptide directionality will facilitate broader screening of candidate peptides. Possible techniques might include for example Förster resonance energy transfer (FRET) spectroscopy.

The most interesting aspect, however, will be to investigate the potential contribution of flipped peptides to autoimmune diseases. It is conceivable that T cells during their selection in the thymus see only one of several possible binding register or binding modes of a self peptide. Later in the periphery, changed conditions for peptide loading (e.g. loading at different intracellular compartments, different pH), might lead to a changed binding mode, resulting in a different T cell epitope. Now, this new epitope can break the central tolerance and activate autoreactive T cells. Importantly, it could already be shown for one of the model antigens of murine diabetes that of three possible binding registers the one with the lowest affinity was the T cell activating epitope (Stadinski *et al.*, 2010). One of the obstacles in the study of these complexes is, of course, the low affinity of these peptides. However, epitope trapping by mutagenesis of anchoring residues or introduction of disulfide bridges are possible strategies to overcome this problem. In our laboratory yet another interesting approach, especially for stabilization of reversely bound peptides, is explored by covalently linking the peptide of interest to the MHC α- instead of the β-chain, which is commonly used. This allows the accommodation of the peptide only in the reversed orientation, as we could already show for CLIP. This enables us now to employ this strategy also for other disease associated MHC alleles with candidate self-antigens.

One aspect of conformational flexibility of MHC II has not been covered in the presented work yet, but that is of extraordinary importance for antigen processing and presentation, is the interaction of MHC II with the HLA-DM, the peptide-exchange catalyst. Both individual structures are known. However, their interaction has remained elusive and only indirect evidence has been gained (Doebele *et al.*, 2000; Pashine *et al.*, 2003; Stratikos *et al.*, 2002). Opposing models propose that either some of the conserved hydrogen bonds in the peptide binding groove are broken (Mosyak *et al.*, 1998) or that HLA-DM induces a more global conformational change (Belmares *et al.*, 2002). Recently, Wucherpfennig and colleagues reported an MHC II with N-terminal region of the binding site partially unoccupied as the target of HLA-DM (Anders *et al.*, 2011), but the exact conformation that is recognized by HLA-DM was not determined. It is well possible that this conformation is very low populated and therefore difficult to detect. However, the recent progress in NMR spectroscopy of MHC II (Gunther *et al.*, 2010; Schlundt *et al.*, 2009) in combination with the structures presented here

make the application of NMR-based methods feasible that are well suited to gain information on otherwise “invisible” states (Baldwin & Kay, 2009).

In summary, these efforts are not only of fundamental basic interest but have also the potential to improve the curability of immune related diseases either by harnessing the power of the immune system (in case of vaccinations) or by specifically modulating its overreaction (in case of autoimmune reactions).

6 Summary / Zusammenfassung

Summary

Molecules of the major histocompatibility complex (MHC) present antigenic peptides for surveillance by T cells and, thereby, are central for the initiation of an adaptive immune response. The structure of MHC class II (MHC II) has been known for almost 20 years and all structures solved so far show a single overall conformation. However, there is ample evidence for variations of this uniform picture. In this thesis I wanted to lay the structural basis for these conformers.

Empty MHC II molecules quickly lose their ability to rebind peptides, what is thought to be accompanied by a conformational change, but direct structural proof is missing. Here, a molecular dynamics simulations-derived model for the transition of empty MHC II from a peptide-receptive to a non-receptive state is presented. It predicts a closure of the binding site by straightening and inwards movement of a flanking helix, stabilized by the formation of a hydrogen bond between highly conserved residues, functioning as lock. This model was verified by *in vitro* mutagenesis studies. First, an intramolecular disulfide bridge was designed that proved the conformational flexibility predicted by the model. More importantly, disabling the locking mechanism by mutagenesis led to highly peptide receptive MHC II species. Thus, the data presented here strongly supports the postulated model for the non-receptive state of MHC II.

In addition, also ligand-dependent conformers of MHC II have been described. For the human leukocyte antigen (HLA)-DR1 it has been reported that terminal extension of peptide ligands beyond the core binding motif influences the MHC II conformation. Here, the structural basis for this effect was probed by X-ray crystallography. Structures of four HLA-DR1 molecules in complex with length variants of a viral antigen were solved, but no conformational difference was identified. In contrast, X-ray crystallography of HLA-DR1 in complex with two length variants of the class II-associated invariant chain derived peptide (CLIP), a natural intermediate in antigen-processing, surprisingly revealed a new conformation. The short version of CLIP was found in an inversed orientation as compared to the canonical binding mode. Remarkably, no rearrangements in the binding groove were necessary to establish the same conserved hydrogen bonding network for both directions. Accompanying NMR experiments proved that thermodynamic stability of the complex drives CLIP to reorient within the bind-

ing groove from the canonical to the reversed state. It was possible to trap the antigen in only one orientation by introduction of favorable anchors. Finally, after screening databases for other potentially reversed antigens, five new HLA-DR1 structures with different peptides were solved. However, none of these showed the non-canonical binding mode. Instead, the structures illustrated the high adaptability of the MHC to different peptide sequences, e.g. by a before unseen opening of the binding pocket 7. In conclusion, the structures of several pMHC II suggest a novel peptide binding mode and underscore the various strategies of MHC II to form highly stable peptide complexes.

Zusammenfassung

Moleküle des Haupthistokompatibilitätskomplexes (MHC) präsentieren Antigenpeptide für die Überwachung durch T-Zellen und sind somit ein zentraler Bestandteil der adaptiven Immunantwort. Die Struktur von MHC Klasse II (MHC II) Komplexen ist seit fast 20 Jahren bekannt und alle Strukturen zeigen eine einheitliche Konformation. Es gibt jedoch reichhaltige Beweise für weitere Konformationen. Mit der vorliegenden Arbeit möchte ich die strukturelle Grundlage hierfür legen.

Leere MHC II Moleküle verlieren schnell die Fähigkeit, erneut Peptide zu binden. Es wird angenommen, dass dies mit einer Änderung ihrer Konformation einhergeht, aber bislang fehlt ein direkter struktureller Beweis. In dieser Arbeit wird ein Modell für den Übergang von einem peptidrezeptiven zu einem nicht rezeptiven Zustand vorgestellt, das auf Simulation der Molekulardynamik des leeren MHC basiert. Dabei bewirkt die Streckung und Einwärtsbewegung einer flankierenden Helix das Verschließen der Antigenbindungsstelle, was durch die Ausbildung einer als Verschluss fungierenden Wasserstoffbrückenbindung zwischen hoch konservierten Aminosäuren stabilisiert wird. Dieses Modell konnte durch *in vitro* Mutagenesestudien verifiziert werden. Dafür wurde zunächst eine intramolekulare Schwefelbrücke eingebaut, die die konformationelle Flexibilität des Modells bewies. Von größerer Bedeutung war jedoch die Erkenntis, dass das Ausschalten des Verschlussmechanismus zu einer hoch peptidrezeptiven MHC II Art führte. Die vorgelegten Daten unterstützen somit das postulierte Modell des nicht-rezeptiven MHC II.

Weiterhin sind auch ligandabhängige MHC II Konformationen beschrieben worden. Die Konformation des humanen Leukozytenantigen (HLA)-DR1 wird beispielsweise durch Verlänge-

rungen der Peptidligandentermini über das Kernbindungsmotiv hinaus beeinflusst. Die strukturelle Basis hierfür wurde mittels Röntgenstrukturanalyse untersucht. Dabei wurden vier Strukturen von HLA-DR1 im Komplex mit Längenvarianten eines viralen Antigens gelöst, wobei jedoch keine Änderung in der Konformation zu beobachten war. Im Gegensatz dazu zeigte die Röntgenstrukturanalyse von HLA-DR1 im Komplex mit zwei Längenvarianten des „class II-associated invariant chain derived peptide“ (CLIP), ein natürliches Intermediat bei der Antigenprozessierung, überraschenderweise eine neuartige Konformation. Die kürzere CLIP Version lag, im Vergleich zur kanonischen Bindungsweise, umgedreht in der Antigenbindungstasche. Dabei waren bemerkenswerterweise keine Änderungen innerhalb der Bindungsfurche nötig, um dasselbe konservierte Wasserstoffbrückennetzwerk für beide Richtungen auszubilden. Begleitende NMR-Experimente bewiesen, dass die Neuorientierung des Peptides durch die thermodynamische Stabilität des Komplexes angetrieben wurde. Es war möglich, diese Neuorientierung durch den Einbau von Peptidankern zu beeinflussen. Zusätzlich wurden Datenbanken nach weiteren potentiell umgedrehten Peptiden durchsucht und fünf neue Strukturen von HLA-DR1 im Komplex mit unterschiedlichen Peptiden gelöst. Dabei zeigte jedoch keine die gewünschte Orientierung. Stattdessen betonten sie die hohe Anpassungsfähigkeit des MHC an verschiedene Peptidsequenzen, z.B. durch eine zuvor unbeobachtete Öffnung der Bindungstasche 7. Zusammenfassend deuten die verschiedenen Strukturen von Peptid-MHC II Komplexen einen neuartigen Peptidbindungsmodus an und unterstreichen die verschiedenen Möglichkeiten des MHC II, hochstabile Peptidkomplexe auszubilden.

7 References

Abbas AK, Murphy KM, Sher A (1996) Functional diversity of helper T lymphocytes. *Nature* **383:** 787-793

Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC, Zwart PH (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66:** 213-221

Anders AK, Call MJ, Schulze MS, Fowler KD, Schubert DA, Seth NP, Sundberg EJ, Wucherpfennig KW (2011) HLA-DM captures partially empty HLA-DR molecules for catalyzed removal of peptide. *Nat Immunol* **12:** 54-61

Anderson MW, Gorski J (2005) Cooperativity during the formation of peptide/MHC class II complexes. *Biochemistry* **44:** 5617--5624

Arneson LS, Katz JF, Liu M, Sant AJ (2001) Hydrogen Bond Integrity Between MHC Class II Molecules and Bound Peptide Determines the Intracellular Fate of MHC Class II Molecules. *The Journal of Immunology* **167:** 6939-6946

Bade-Doding C, Theodossis A, Gras S, Kjer-Nielsen L, Eiz-Vesper B, Seltsam A, Huyton T, Rossjohn J, McCluskey J, Blasczyk R (2011) The impact of human leukocyte antigen (HLA) micropolymorphism on ligand specificity within the HLA-B*41 allotypic family. *Haematologica* **96:** 110-118

Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA (2001) Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci U S A* **98:** 10037--10041

Bakke O, Dobberstein B (1990) MHC class II-associated invariant chain contains a sorting signal for endosomal compartments. *Cell* **63:** 707-716

Baldwin AJ, Kay LE (2009) NMR spectroscopy brings invisible protein states into focus. *Nat Chem Biol* **5:** 808-814

Barry M, Bleackley RC (2002) Cytotoxic T lymphocytes: all roads lead to death. *Nat Rev Immunol* **2:** 401-409

Bartnes K, Hannestad K, Guichard G, Briand JP (1997) A retro-inverso analog mimics the cognate peptide epitope of a CD4+ T cell clone. *Eur J Immunol* **27:** 1387-1391

Belmares MP, Busch R, Wucherpfennig KW, McConnell HM, Mellins ED (2002) Structural factors contributing to DM susceptibility of MHC class II/peptide complexes. *J Immunol* **169:** 5109-5117

Bikoff EK, Huang LY, Episkopou V, van Meerwijk J, Germain RN, Robertson EJ (1993) Defective major histocompatibility complex class II assembly, transport, peptide acquisition, and CD4+ T cell selection in mice lacking invariant chain expression. *J Exp Med* **177:** 1699-1712

Birnboim HC, Doly J (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res* **7:** 1513-1523

Bjorkman PJ, Saper MA, Samraoui B, Bennett WS, Strominger JL, Wiley DC (1987) Structure of the human class I histocompatibility antigen, HLA-A2. *Nature* **329:** 506-512

Brizzard BL, Chubet RG, Vizard DL (1994) Immunoaffinity purification of FLAG epitope-tagged bacterial alkaline phosphatase using a novel monoclonal antibody and peptide elution. *Biotechniques* **16:** 730-735

Brown JH, Jardetzky TS, Gorga JC, Stern LJ, Urban RG, Strominger JL, Wiley DC (1993) Three-dimensional structure of the human class II histocompatibility antigen HLA-DR1. *Nature* **364:** 33--39

Brunger AT (1992) Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* **355:** 472-475

Buchli R, VanGundy RS, Hickman-Miller HD, Giberson CF, Bardet W, Hildebrand WH (2004) Real-time measurement of in vitro peptide binding to soluble HLA-A*0201 by fluorescence polarization. *Biochemistry* **43:** 14852-14863

Bunch TA, Grinblat Y, Goldstein LS (1988) Characterization and use of the Drosophila metallothionein promoter in cultured Drosophila melanogaster cells. *Nucleic Acids Res* **16:** 1043-1061

Burnet FM (1957) A modification of Jerne's theory of antibody production using the concept of clonal selection. *Australian Journal of Science* **20:** 67-69

Call MJ, Xing X, Cuny GD, Seth NP, Altmann DM, Fugger L, Krogsgaard M, Stein RL, Wucherpfennig KW (2009) In vivo enhancement of peptide display by MHC class II molecules with small molecule catalysts of peptide exchange. *J Immunol* **182:** 6342--6352

Carven GJ, Chitta S, Hilgert I, Rushe MM, Baggio RF, Palmer M, Arenas JE, Strominger JL, Horejsi V, Santambrogio L, Stern LJ (2004) Monoclonal antibodies specific for the empty conformation of HLA-DR1 reveal aspects of the conformational change associated with peptide binding. *J Biol Chem* **279:** 16561--16570

Ceman S, Wu S, Jardetzky TS, Sant AJ (1998) Alteration of a single hydrogen bond between class II molecules and peptide results in rapid degradation of class II molecules after invariant chain removal. *J Exp Med* **188:** 2139-2149

Chicz RM, Urban RG, Lane WS, Gorga JC, Stern LJ, Vignali DA, Strominger JL (1992) Predominant naturally processed peptides bound to HLA-DR1 are derived from MHC-related molecules and are heterogeneous in size. *Nature* **358:** 764-768

Clement CC, Rotzschke O, Santambrogio L (2011) The lymph as a pool of self-antigens. *Trends Immunol* **32:** 6-11

Coleman RG, Sharp KA (2006) Travel depth, a new shape descriptor for macromolecules: application to ligand binding. *J Mol Biol* **362:** 441-458

Davis IW, Leaver-Fay A, Chen VB, Block JN, Kapral GJ, Wang X, Murray LW, Arendall WB, 3rd, Snoeyink J, Richardson JS, Richardson DC (2007) MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res* **35:** W375-383

De Gassart A, Camosseto V, Thibodeau J, Ceppi M, Catalan N, Pierre P, Gatti E (2008) MHC class II stabilization at the surface of human dendritic cells is the result of maturation-dependent MARCH I down-regulation. *Proc Natl Acad Sci U S A* **105:** 3491-3496

de Jong WW, Zweers A, Cohen LH (1978) Influence of single amino acid substitutions on electrophoretic mobility of sodium dodecyl sulfate-protein complexes. *Biochem Biophys Res Commun* **82:** 532-539

Deng L, Langley RJ, Brown PH, Xu G, Teng L, Wang Q, Gonzales MI, Callender GG, Nishimura MI, Topalian SL, Mariuzza RA (2007) Structural basis for the recognition of mutant self by a tumor-specific, MHC class II-restricted T cell receptor. *Nat Immunol* **8:** 398-408

Denzin LK, Hammond C, Cresswell P (1996) HLA-DM interactions with intermediates in HLA-DR maturation and a role for HLA-DM in stabilizing empty HLA-DR molecules. *J Exp Med* **184:** 2153-2165

DeWeese C, Kwok W, Nepom G, Lybrand T (1996) Characterization of a Novel Reverse-orientation Model for a Peptide/MHC Complex Putatively Associated with Type I Diabetes Mellitus. *Journal of Molecular Modeling* **2:** 205-216

Dickhaut K, Hoepner S, Eckhard J, Wiesmueller K-H, Schindler L, Jung G, Falk K, Roetzschke O (2009) Enhancement of tumour-specific immune responses in vivo by 'MHC loading-enhancer' (MLE). *PLoS ONE* **4:** e6811

Doebele RC, Busch R, Scott HM, Pashine A, Mellins ED (2000) Determination of the HLA-DM interaction site on HLA-DR molecules. *Immunity* **13:** 517-527

Emsley P, Lohkamp B, Scott WG, Cowtan K (2010) Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66:** 486-501

Ericsson UB, Hallberg BM, Detitta GT, Dekker N, Nordlund P (2006) Thermofluor-based high-throughput stability optimization of proteins for structural studies. *Anal Biochem* **357:** 289-298

Falk K, Rotzschke O, Stevanovic S, Jung G, Rammensee HG (1991) Allele-specific motifs revealed by sequencing of self-peptides eluted from MHC molecules. *Nature* **351:** 290-296

Fallang LE, Roh S, Holm A, Bergseng E, Yoon T, Fleckenstein B, Bandyopadhyay A, Mellins ED, Sollid LM (2008) Complexes of two cohorts of CLIP peptides and HLA-DQ2 of the autoimmune DR3-DQ2 haplotype are poor substrates for HLA-DM. *J Immunol* **181:** 5451-5461

Fernandez MM, Guan R, Swaminathan CP, Malchiodi EL, Mariuzza RA (2006) Crystal structure of staphylococcal enterotoxin I (SEI) in complex with a human major histocompatibility complex class II molecule. *J Biol Chem* **281:** 25356-25364

Ferrante A, Gorski J (2007) Cooperativity of hydrophobic anchor interactions: evidence for epitope selection by MHC class II as a folding process. *J Immunol* **178:** 7181-7189

Ferrante A, Gorski J (2009) Cutting Edge: HLA-DM-Mediated Peptide Exchange Functions Normally on MHC Class II-Peptide Complexes That Have Been Weakened by Elimination of a Conserved Hydrogen Bond. *J Immunol*

Fleckenstein B, Jung G, Wiesmuller KH (1999) Quantitative analysis of peptide-MHC class II interaction. *Semin Immunol* **11:** 405-416

Fleckenstein B, Kalbacher H, Muller CP, Stoll D, Halder T, Jung G, Wiesmuller KH (1996) New ligands binding to the human leukocyte antigen class II molecule DRB1*0101 based on the activity pattern of an undecapeptide library. *Eur J Biochem* **240:** 71-77

Fourneau J-M, Cohen H, van Endert PM (2004) A chaperone-assisted high yield system for the production of HLA-DR4 tetramers in insect cells. *J Immunol Methods* **285:** 253--264

Frayser M, Sato AK, Xu L, Stern LJ (1999) Empty and peptide-loaded class II major histocompatibility complex proteins produced by expression in Escherichia coli and folding in vitro. *Protein Expr Purif* **15:** 105--114

Fremont DH, Dai S, Chiang H, Crawford F, Marrack P, Kappler J (2002) Structural basis of cytochrome c presentation by IE(k). *J Exp Med* **195:** 1043-1052

Fremont DH, Hendrickson WA, Marrack P, Kappler J (1996) Structures of an MHC class II molecule with covalently bound single peptides. *Science* **272:** 1001-1004

Fremont DH, Matsumura M, Stura EA, Peterson PA, Wilson IA (1992) Crystal structures of two viral peptides in complex with murine MHC class I H-2Kb. *Science* **257:** 919-927

Garcia KC, Adams JJ, Feng D, Ely LK (2009) The molecular basis of TCR germline bias for MHC is surprisingly simple. *Nat Immunol* **10:** 143-147

Georges B, Loing E, Neveu R, Melnyk O, Gras-Masse H, Auriault C (2000) Structural diversity of human class II histocompatibility molecules induced by peptide ligands. *FEBS Lett* **481:** 249-254

Germain RN (1995) Binding domain regulation of MHC class II molecule assembly, trafficking, fate, and function. *Semin Immunol* **7:** 361-372

Ghosh P, Amaya M, Mellins E, Wiley DC (1995) The structure of an intermediate in class II MHC maturation: CLIP bound to HLA-DR3. *Nature* **378:** 457-462

Griffith IJ, Nabavi N, Ghogawala Z, Chase CG, Rodriguez M, McKean DJ, Glimcher LH (1988) Structural mutation affecting intracellular transport and cell surface expression of murine class II molecules. *J Exp Med* **167:** 541--555

Grotenbreg GM, Nicholson MJ, Fowler KD, Wilbuer K, Octavio L, Yang M, Chakraborty AK, Ploegh HL, Wucherpfennig KW (2007) Empty class II major histocompatibility complex created by peptide photolysis establishes the role of DM in peptide association. *J Biol Chem* **282:** 21425--21436

Gunther S, Schlundt A, Sticht J, Roske Y, Heinemann U, Wiesmuller KH, Jung G, Falk K, Rotzschke O, Freund C (2010) Bidirectional binding of invariant chain peptides to an MHC class II molecule. *Proc Natl Acad Sci U S A* **107:** 22219-22224

Gupta S, Höpner S, Rupp B, Günther S, Dickhaut K, Agarwal N, Cardoso MC, Kühne R, Wiesmüller K-H, Jung G, Falk K, Rötzschke O (2008) Anchor side chains of short peptide fragments trigger ligand-exchange of class II MHC molecules. *PLoS ONE* **3:** e1814

Hammer J, Takacs B, Sinigaglia F (1992) Identification of a motif for HLA-DR1 binding peptides using M13 display libraries. *J Exp Med* **176:** 1007-1013

Hartman IZ, Kim A, Cotter RJ, Walter K, Dalai SK, Boronina T, Griffith W, Lanar DE, Schwenk R, Krzych U, Cole RN, Sadegh-Nasseri S (2010) A reductionist cell-free major histocompatibility complex class II antigen processing system identifies immunodominant epitopes. *Nat Med* **16:** 1333-1340

Hennecke J, Carfi A, Wiley DC (2000) Structure of a covalently stabilized complex of a human alphabeta T-cell receptor, influenza HA peptide and MHC class II molecule, HLA-DR1. *EMBO J* **19:** 5611-5624

Herve M, Maillere B, Mourier G, Texier C, Leroy S, Menez A (1997) On the immunogenic properties of retro-inverso peptides. Total retro-inversion of T-cell epitopes causes a loss of binding to MHC II molecules. *Mol Immunol* **34:** 157-163

Höpner S, Dickhaut K, Hofstätter M, Krämer H, Rückerl D, Söderhäll JA, Gupta S, Marin-Esteban V, Kühne R, Freund C, Jung G, Falk K, Rötzschke O (2006) Small organic compounds enhance antigen loading of class II major histocompatibility complex proteins by targeting the polymorphic P1 pocket. *J Biol Chem* **281:** 38535--38542

Hori S, Nomura T, Sakaguchi S (2003) Control of regulatory T cell development by the transcription factor Foxp3. *Science* **299:** 1057-1061

Howarth M, Williams A, Tolstrup AB, Elliott T (2004) Tapasin enhances MHC class I peptide presentation according to peptide half-life. *Proc Natl Acad Sci U S A* **101:** 11737-11742

Huang JC, Han M, Minguela A, Pastor S, Qadri A, Ward ES (2003) T cell recognition of distinct peptide:I-Au conformers in murine experimental autoimmune encephalomyelitis. *J Immunol* **171:** 2467-2477

Hunt DF, Michel H, Dickinson TA, Shabanowitz J, Cox AL, Sakaguchi K, Appella E, Grey HM, Sette A (1992) Peptides presented to the immune system by the murine class II major histocompatibility complex molecule I-Ad. *Science* **256:** 1817-1820

Jardetzky TS, Brown JH, Gorga JC, Stern LJ, Urban RG, Chi YI, Stauffacher C, Strominger JL, Wiley DC (1994) Three-dimensional structure of a human class II histocompatibility molecule complexed with superantigen. *Nature* **368:** 711-718

Jardetzky TS, Brown JH, Gorga JC, Stern LJ, Urban RG, Strominger JL, Wiley DC (1996) Crystallographic analysis of endogenous peptides associated with HLA-DR1 suggests a common, polyproline II-like conformation for bound peptides. *Proc Natl Acad Sci U S A* **93:** 734-738

Jardetzky TS, Gorga JC, Busch R, Rothbard J, Strominger JL, Wiley DC (1990) Peptide binding to HLA-DR1: a peptide with most residues substituted to alanine retains MHC binding. *EMBO J* **9:** 1797-1803

Jasanoff A, Wagner G, Wiley DC (1998) Structure of a trimeric domain of the MHC class II-associated chaperonin and targeting protein Ii. *EMBO J* **17:** 6812-6818

Joshi RV, Zarutskie JA, Stern LJ (2000) A three-step kinetic mechanism for peptide binding to MHC class II proteins. *Biochemistry* **39:** 3751--3762

Kabsch W (2010) Xds. *Acta Crystallogr D Biol Crystallogr* **66:** 125-132

Kersh GJ, Miley MJ, Nelson CA, Grakoui A, Horvath S, Donermeyer DL, Kappler J, Allen PM, Fremont DH (2001) Structural and functional consequences of altering a peptide MHC anchor residue. *J Immunol* **166:** 3345-3354

Knudsen PJ, Strominger JL (1986) A monoclonal antibody that recognizes the alpha chain of HLA-DR antigens. *Hum Immunol* **15:** 150-163

Kozono H, White J, Clements J, Marrack P, Kappler J (1994) Production of soluble MHC class II proteins with covalently bound single peptides. *Nature* **369:** 151-154

Kramer A, Schneider-Mergener J (1998) Synthesis and screening of peptide libraries on continuous cellulose membrane supports. *Methods Mol Biol* **87:** 25-39

Kvist S, Wiman K, Claesson L, Peterson PA, Dobberstein B (1982) Membrane insertion and oligomeric assembly of HLA-DR histocompatibility antigens. *Cell* **29:** 61-69

Lampson LA, Levy R (1980) Two populations of Ia-like molecules on a human B cell line. *J Immunol* **125:** 293-299

Le Questel JY, Morris DG, Maccallum PH, Poet R, Milner-White EJ (1993) Common ring motifs in proteins involving asparagine or glutamine amide groups hydrogen-bonded to main-chain atoms. *J Mol Biol* **231:** 888-896

Leslie AG (2006) The integration of macromolecular diffraction data. *Acta Crystallogr D Biol Crystallogr* **62:** 48-57

Li Y, Li H, Martin R, Mariuzza RA (2000) Structural basis for the binding of an immunodominant peptide from myelin basic protein in different registers by two HLA-DR2 proteins. *J Mol Biol* **304:** 177-188

Lin HH, Zhang GL, Tongchusak S, Reinherz EL, Brusic V (2008) Evaluation of MHC-II peptide binding prediction servers: applications for vaccine research. *BMC Bioinformatics* **9 Suppl 12:** S22

Lindner R, Unanue ER (1996) Distinct antigen MHC class II complexes generated by separate processing pathways. *EMBO J* **15:** 6910-6920

Liu GY, Fairchild PJ, Smith RM, Prowle JR, Kioussis D, Wraith DC (1995) Low avidity recognition of self-antigen by T cells permits escape from central tolerance. *Immunity* **3:** 407-415

Liu X, Dai S, Crawford F, Fruge R, Marrack P, Kappler J (2002) Alternate interactions define the binding of peptides to the MHC molecule IA(b). *Proc Natl Acad Sci U S A* **99:** 8820-8825

Lotteau V, Teyton L, Peleraux A, Nilsson T, Karlsson L, Schmid SL, Quaranta V, Peterson PA (1990) Intracellular transport of class II MHC molecules directed by invariant chain. *Nature* **348:** 600-605

Lovitch SB, Pu Z, Unanue ER (2006) Amino-terminal flanking residues determine the conformation of a peptide-class II MHC complex. *J Immunol* **176:** 2958-2968

Lovitch SB, Unanue ER (2005) Conformational isomers of a peptide-class II major histocompatibility complex. *Immunol Rev* **207:** 293-313

Madden DR, Garboczi DN, Wiley DC (1993) The antigenic identity of peptide-MHC complexes: a comparison of the conformations of five viral peptides presented by HLA-A2. *Cell* **75:** 693-708

Maric M, Arunachalam B, Phan UT, Dong C, Garrett WS, Cannon KS, Alfonso C, Karlsson L, Flavell RA, Cresswell P (2001) Defective antigen processing in GILT-free mice. *Science* **294:** 1361-1365

McCormick PJ, Martina JA, Bonifacino JS (2005) Involvement of clathrin and AP-2 in the trafficking of MHC class II molecules to antigen-processing compartments. *Proc Natl Acad Sci U S A* **102:** 7910-7915

McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ (2007) Phaser crystallographic software. *J Appl Crystallogr* **40:** 658-674

McFarland BJ, Beeson C, Sant AJ (1999) Cutting edge: a single, essential hydrogen bond controls the stability of peptide-MHC class II complexes. *J Immunol* **163:** 3567--3571

McFarland BJ, Katz JF, Beeson C, Sant AJ (2001) Energetic asymmetry among hydrogen bonds in MHC class II*peptide complexes. *Proc Natl Acad Sci U S A* **98:** 9231--9236

McFarland BJ, Katz JF, Sant AJ, Beeson C (2005) Energetics and cooperativity of the hydrogen bonding and anchor interactions that bind peptides to MHC class II protein. *J Mol Biol* **350:** 170--183

Mo XY, Cascio P, Lemerise K, Goldberg AL, Rock K (1999) Distinct proteolytic processes generate the C and N termini of MHC class I-binding peptides. *J Immunol* **163:** 5851-5859

Mohan JF, Levisetti MG, Calderon B, Herzog JW, Petzold SJ, Unanue ER (2010) Unique autoreactive T cells recognize insulin peptides generated within the islets of Langerhans in autoimmune diabetes. *Nat Immunol* **11:** 350-354

Mosyak L, Zaller DM, Wiley DC (1998) The structure of HLA-DM, the peptide exchange catalyst that loads antigen onto class II MHC molecules during antigen presentation. *Immunity* **9:** 377-383

Mullen MM, Haan KM, Longnecker R, Jardetzky TS (2002) Structure of the Epstein-Barr virus gp42 protein bound to the MHC class II receptor HLA-DR1. *Mol Cell* **9:** 375-385

Mungall AJ, Palmer SA, Sims SK, Edwards CA, Ashurst JL, Wilming L, Jones MC, Horton R, Hunt SE, Scott CE, Gilbert JG, Clamp ME, Bethel G, Milne S, Ainscough R, Almeida JP, Ambrose KD, Andrews TD, Ashwell RI, Babbage AK, Bagguley CL, Bailey J, Banerjee R, Barker DJ, Barlow KF, Bates K, Beare DM, Beasley H, Beasley O, Bird CP, Blakey S, Bray-Allen S, Brook J, Brown AJ, Brown JY, Burford DC, Burrill W, Burton J, Carder C, Carter NP, Chapman JC, Clark SY, Clark G, Clee CM, Clegg S, Cobley V, Collier RE, Collins JE, Colman LK, Corby NR, Coville GJ, Culley KM, Dhami P, Davies J, Dunn M, Earthrowl ME, Ellington AE, Evans KA, Faulkner L, Francis MD, Frankish A, Frankland J, French L, Garner P, Garnett J, Ghori MJ, Gilby LM, Gillson CJ, Glithero RJ, Grafham DV, Grant M, Gribble S, Griffiths C, Griffiths M, Hall R, Halls KS, Hammond S, Harley JL, Hart EA, Heath PD, Heathcott R, Holmes SJ, Howden PJ, Howe KL, Howell GR, Huckle E, Humphray SJ, Humphries MD, Hunt AR, Johnson CM, Joy AA, Kay M, Keenan SJ, Kimberley AM, King A, Laird GK, Langford C,

Lawlor S, Leongamornlert DA, Leversha M, Lloyd CR, Lloyd DM, Loveland JE, Lovell J, Martin S, Mashreghi-Mohammadi M, Maslen GL, Matthews L, McCann OT, McLaren SJ, McLay K, McMurray A, Moore MJ, Mullikin JC, Niblett D, Nickerson T, Novik KL, Oliver K, Overton-Larty EK, Parker A, Patel R, Pearce AV, Peck AI, Phillimore B, Phillips S, Plumb RW, Porter KM, Ramsey Y, Ranby SA, Rice CM, Ross MT, Searle SM, Sehra HK, Sheridan E, Skuce CD, Smith S, Smith M, Spraggon L, Squares SL, Steward CA, Sycamore N, Tamlyn-Hall G, Tester J, Theaker AJ, Thomas DW, Thorpe A, Tracey A, Tromans A, Tubby B, Wall M, Wallis JM, West AP, White SS, Whitehead SL, Whittaker H, Wild A, Willey DJ, Wilmer TE, Wood JM, Wray PW, Wyatt JC, Young L, Younger RM, Bentley DR, Coulson A, Durbin R, Hubbard T, Sulston JE, Dunham I, Rogers J, Beck S (2003) The DNA sequence and analysis of human chromosome 6. *Nature* **425:** 805-811

Munz C, Hofmann M, Yoshida K, Moustakas AK, Kikutani H, Stevanovic S, Papadopoulos GK, Rammensee HG (2002) Peptide analysis, stability studies, and structural modeling explain contradictory peptide motifs and unique properties of the NOD mouse MHC class II molecule H2-A(g7). *Eur J Immunol* **32:** 2105-2116

Murshudov GN, Vagin AA, Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* **53:** 240-255

Murthy VL, Stern LJ (1997) 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

Narayan K, Chou C-L, Kim A, Hartman IZ, Dalai S, Khoruzhenko S, Sadegh-Nasseri S (2007) HLA-DM targets the hydrogen bond between the histidine at position beta81 and peptide to dissociate HLA-DR-peptide complexes. *Nat Immunol* **8:** 92--100

Natarajan SK, Stern LJ, Sadegh-Nasseri S (1999) Sodium dodecyl sulfate stability of HLA-DR1 complexes correlates with burial of hydrophobic residues in pocket 1. *J Immunol* **162:** 3463--3470

Neumann J, Koch N (2006) A novel domain on HLA-DRbeta chain regulates the chaperone role of the invariant chain. *J Cell Sci* **119:** 4207-4214

Niesen FH, Berglund H, Vedadi M (2007) The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nat Protoc* **2:** 2212-2221

Ohmura-Hoshino M, Matsuki Y, Aoki M, Goto E, Mito M, Uematsu M, Kakiuchi T, Hotta H, Ishido S (2006) Inhibition of MHC class II expression and immune responses by c-MIR. *J Immunol* **177:** 341-354

Painter CA, Cruz A, López GE, Stern LJ, Zavala-Ruiz Z (2008) Model for the peptide-free conformation of class II MHC proteins. *PLoS ONE* **3:** e2403

Pashine A, Busch R, Belmares MP, Munning JN, Doebele RC, Buckingham M, Nolan GP, Mellins ED (2003) Interaction of HLA-DR with an acidic face of HLA-DM disrupts sequence-dependent interactions with peptides. *Immunity* **19:** 183-192

Petersson K, Hakansson M, Nilsson H, Forsberg G, Svensson LA, Liljas A, Walse B (2001) Crystal structure of a superantigen bound to MHC class II displays zinc and peptide dependence. *EMBO J* **20:** 3306-3312

Petersson K, Thunnissen M, Forsberg G, Walse B (2002) Crystal structure of a SEA variant in complex with MHC class II reveals the ability of SEA to crosslink MHC molecules. *Structure* **10:** 1619-1626

Pezeshki AM, Cote MH, Azar GA, Routy JP, Boulassel MR, Thibodeau J (2011) Forced Expression of HLA-DM at the Surface of Dendritic Cells Increases Loading of Synthetic Peptides on MHC Class II Molecules and Modulates T Cell Responses. *J Immunol*

Potolicchio I, Chitta S, Xu X, Fonseca D, Crisi G, Horejsi V, Strominger JL, Stern LJ, Raposo G, Santambrogio L (2005) Conformational variation of surface class II MHC proteins during myeloid dendritic cell differentiation accompanies structural changes in lysosomal MIIC. *J Immunol* **175:** 4935--4947

Praaenikar J, Afonine PV, Guncar G, Adams PD, Turk D (2009) Averaged kick maps: less noise, more signal... and probably less bias. *Acta Crystallogr D Biol Crystallogr* **65:** 921-931

Pu Z, Carrero JA, Unanue ER (2002) Distinct recognition by two subsets of T cells of an MHC class II-peptide complex. *Proc Natl Acad Sci U S A* **99:** 8844-8849

Pu Z, Lovitch SB, Bikoff EK, Unanue ER (2004) T cells distinguish MHC-peptide complexes formed in separate vesicles and edited by H2-DM. *Immunity* **20:** 467-476

Rabinowitz JD, Vrljic M, Kasson PM, Liang MN, Busch R, Boniface JJ, Davis MM, McConnell HM (1998) Formation of a highly peptide-receptive state of class II MHC. *Immunity* **9:** 699--709

Rammensee H, Bachmann J, Emmerich NP, Bachor OA, Stevanovic S (1999) SYFPEITHI: database for MHC ligands and peptide motifs. *Immunogenetics* **50:** 213-219

Riberdy JM, Newcomb JR, Surman MJ, Barbosa JA, Cresswell P (1992) HLA-DR molecules from an antigen-processing mutant cell line are associated with invariant chain peptides. *Nature* **360:** 474-477

Roche PA, Marks MS, Cresswell P (1991) Formation of a nine-subunit complex by HLA class II glycoproteins and the invariant chain. *Nature* **354:** 392-394

Rohn TA, Boes M, Wolters D, Spindeldreher S, Muller B, Langen H, Ploegh H, Vogt AB, Kropshofer H (2004) Upregulation of the CLIP self peptide on mature dendritic cells antagonizes T helper type 1 polarization. *Nat Immunol* **5:** 909-918

Rotzschke O, Falk K, Mack J, Lau JM, Jung G, Strominger JL (1999) Conformational variants of class II MHC/peptide complexes induced by N- and C-terminal extensions of minimal peptide epitopes. *Proc Natl Acad Sci U S A* **96:** 7445-7450

Rotzschke O, Falk K, Strominger JL (1997) Superactivation of an immune response triggered by oligomerized T cell epitopes. *Proc Natl Acad Sci U S A* **94:** 14642-14647

Rudensky A, Preston-Hurlburt P, al-Ramadi BK, Rothbard J, Janeway CA, Jr. (1992) Truncation variants of peptides isolated from MHC class II molecules suggest sequence motifs. *Nature* **359:** 429-431

Rudensky A, Preston-Hurlburt P, Hong SC, Barlow A, Janeway CA, Jr. (1991) Sequence analysis of peptides bound to MHC class II molecules. *Nature* **353:** 622-627

Rudolph MG, Stanfield RL, Wilson IA (2006) How TCRs bind MHCs, peptides, and coreceptors. *Annu Rev Immunol* **24:** 419-466

Rupp B (2010) *Biomolecular crystallography*, New York: Garland Science.

Rupp B, Gunther S, Makhmoor T, Schlundt A, Dickhaut K, Gupta S, Choudhary I, Wiesmuller KH, Jung G, Freund C, Falk K, Rotzschke O, Kuhne R (2011) Characterization of Structural Features Controlling the Receptiveness of Empty Class II MHC Molecules. *PLoS ONE* **6:** e18662

Saito K, Oda M, Sarai A, Azuma T, Kozono H (2004) Bound peptide-dependent thermal stability of major histocompatibility complex class II molecule I-Ek. *Biochemistry* **43:** 10186--10191

Saline M, Rodstrom KE, Fischer G, Orekhov VY, Karlsson BG, Lindkvist-Petersson K (2010) The structure of superantigen complexed with TCR and MHC reveals novel insights into superantigenic T cell activation. *Nat Commun* **1:** 119

Sant AJ, Beeson C, McFarland B, Cao J, Ceman S, Bryant PW, Wu S (1999) Individual hydrogen bonds play a critical role in MHC class II: peptide interactions: implications for the dynamic aspects of class II trafficking and DM-mediated peptide exchange. *Immunol Rev* **172:** 239--253

Santambrogio L, Sato AK, Carven GJ, Belyanskaya SL, Strominger JL, Stern LJ (1999a) Extracellular antigen processing and presentation by immature dendritic cells. *Proc Natl Acad Sci U S A* **96:** 15056-15061

Santambrogio L, Sato AK, Fischer FR, Dorf ME, Stern LJ (1999b) Abundant empty class II MHC molecules on the surface of immature dendritic cells. *Proc Natl Acad Sci U S A* **96:** 15050-15055

Sato AK, Zarutskie JA, Rushe MM, Lomakin A, Natarajan SK, Sadegh-Nasseri S, Benedek GB, Stern LJ (2000) Determinants of the peptide-induced conformational change in the human class II major histocompatibility complex protein HLA-DR1. *J Biol Chem* **275:** 2165--2173

Saveanu L, Carroll O, Lindo V, Del Val M, Lopez D, Lepelletier Y, Greer F, Schomburg L, Fruci D, Niedermann G, van Endert PM (2005) Concerted peptide trimming by human ERAP1 and ERAP2 aminopeptidase complexes in the endoplasmic reticulum. *Nat Immunol* **6:** 689-697

Schlundt A, Kilian W, Beyermann M, Sticht J, Günther S, Höpner S, Falk K, Roetzschke O, Mitschang L, Freund C (2009) A xenon-129 biosensor for monitoring MHC-peptide interactions. *Angew Chem Int Ed Engl* **48:** 4142--4145

Schmitt L, Boniface JJ, Davis MM, McConnell HM (1999) Conformational isomers of a class II MHC-peptide complex in solution. *J Mol Biol* **286:** 207-218

Scott-Browne JP, White J, Kappler JW, Gapin L, Marrack P (2009) Germline-encoded amino acids in the alphabeta T-cell receptor control thymic selection. *Nature* **458:** 1043-1046

Sercarz EE, Maverakis E (2003) Mhc-guided processing: binding of large antigen fragments. *Nat Rev Immunol* **3:** 621-629

Sette A, Southwood S, Miller J, Appella E (1995) Binding of major histocompatibility complex class II to the invariant chain-derived peptide, CLIP, is regulated by allelic polymorphism in class II. *J Exp Med* **181:** 677-683

Shin JS, Ebersold M, Pypaert M, Delamarre L, Hartley A, Mellman I (2006) Surface expression of MHC class II in dendritic cells is controlled by regulated ubiquitination. *Nature* **444:** 115-118

Sloan VS, Cameron P, Porter G, Gammon M, Amaya M, Mellins E, Zaller DM (1995) Mediation by HLA-DM of dissociation of peptides from HLA-DR. *Nature* **375:** 802-806

Smith KJ, Pyrdol J, Gauthier L, Wiley DC, Wucherpfennig KW (1998) Crystal structure of HLA-DR2 (DRA*0101, DRB1*1501) complexed with a peptide from human myelin basic protein. *J Exp Med* **188:** 1511-1520

Snell GD (1948) Methods for the study of histocompatibility genes. *J Genet* **49:** 87-108

Sowdhamini R, Srinivasan N, Shoichet B, Santi DV, Ramakrishnan C, Balaram P (1989) Stereochemical modeling of disulfide bridges. Criteria for introduction into proteins by site-directed mutagenesis. *Protein Eng* **3:** 95--103

Stadinski BD, Zhang L, Crawford F, Marrack P, Eisenbarth GS, Kappler JW (2010) Diabetogenic T cells recognize insulin bound to IAg7 in an unexpected, weakly binding register. *Proc Natl Acad Sci U S A* **107:** 10978-10983

Stern LJ, Brown JH, Jardetzky TS, Gorga JC, Urban RG, Strominger JL, Wiley DC (1994) Crystal structure of the human class II MHC protein HLA-DR1 complexed with an influenza virus peptide. *Nature* **368:** 215--221

Stern LJ, Wiley DC (1992) The human class II MHC protein HLA-DR1 assembles as empty alpha beta heterodimers in the absence of antigenic peptide. *Cell* **68:** 465--477

Stern LJ, Wiley DC (1994) Antigenic peptide binding by class I and class II histocompatibility proteins. *Structure* **2:** 245-251

Stratikos E, Mosyak L, Zaller DM, Wiley DC (2002) Identification of the lateral interaction surfaces of human histocompatibility leukocyte antigen (HLA)-DM with HLA-DR1 by formation of tethered complexes that present enhanced HLA-DM catalysis. *J Exp Med* **196:** 173-183

Strubin M, Long EO, Mach B (1986) Two forms of the Ia antigen-associated invariant chain result from alternative initiations at two in-phase AUGs. *Cell* **47:** 619-625

Stumptner-Cuvelette P, Benaroch P (2002) Multiple roles of the invariant chain in MHC class II function. *Biochim Biophys Acta* **1542:** 1-13

Sundberg EJ, Andersen PS, Schlievert PM, Karjalainen K, Mariuzza RA (2003) Structural, energetic, and functional analysis of a protein-protein interface at distinct stages of affinity maturation. *Structure* **11:** 1151-1161

Tan LJ, Ceman S, Chervonsky A, Rodriguez-Paris J, Steck TL, Sant AJ (1997) Late events in the intracellular sorting of major histocompatibility complex class II molecules are regulated by the 80-82 segment of the class II beta chain. *Eur J Immunol* **27:** 1479--1488

Teyton L, O'Sullivan D, Dickson PW, Lotteau V, Sette A, Fink P, Peterson PA (1990) Invariant chain distinguishes between the exogenous and endogenous antigen presentation pathways. *Nature* **348:** 39-44

Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJ (2005) GROMACS: fast, flexible, and free. *J Comput Chem* **26:** 1701-1718

van Niel G, Wubbolts R, Ten Broeke T, Buschow SI, Ossendorp FA, Melief CJ, Raposo G, van Balkom BW, Stoorvogel W (2006) Dendritic cells regulate exposure of MHC class II at their plasma membrane by oligoubiquitination. *Immunity* **25:** 885-894

Venkatraman P, Nguyen TT, Sainlos M, Bilsel O, Chitta S, Imperiali B, Stern LJ (2007) Fluorogenic probes for monitoring peptide binding to class II MHC proteins in living cells. *Nat Chem Biol* **3:** 222-228

Viner NJ, Nelson CA, Deck B, Unanue ER (1996) Complexes generated by the binding of free peptides to class II MHC molecules are antigenically diverse compared with those generated by intracellular processing. *J Immunol* **156:** 2365-2368

Viret C, He X, Janeway CA, Jr. (2003) Altered positive selection due to corecognition of floppy peptide/MHC II conformers supports an integrative model of thymic selection. *Proc Natl Acad Sci U S A* **100:** 5354-5359

Vita R, Zarebski L, Greenbaum JA, Emami H, Hoof I, Salimi N, Damle R, Sette A, Peters B (2010) The immune epitope database 2.0. *Nucleic Acids Res* **38:** D854-862

Viville S, Neefjes J, Lotteau V, Dierich A, Lemeur M, Ploegh H, Benoist C, Mathis D (1993) Mice lacking the MHC class II-associated invariant chain. *Cell* **72:** 635-648

Vogt AB, Kropshofer H, Kalbacher H, Kalbus M, Rammensee HG, Coligan JE, Martin R (1994) Ligand motifs of HLA-DRB5*0101 and DRB1*1501 molecules delineated from self-peptides. *J Immunol* **153:** 1665-1673

Walseng E, Furuta K, Bosch B, Weih KA, Matsuki Y, Bakke O, Ishido S, Roche PA (2010) Ubiquitination regulates MHC class II-peptide complex retention and degradation in dendritic cells. *Proc Natl Acad Sci U S A* **107:** 20465-20470

Wang P, Sidney J, Dow C, Mothe B, Sette A, Peters B (2008) A systematic assessment of MHC class II peptide binding predictions and evaluation of a consensus approach. *PLoS Comput Biol* **4:** e1000048

Wearsch PA, Cresswell P (2007) Selective loading of high-affinity peptides onto major histocompatibility complex class I molecules by the tapasin-ERp57 heterodimer. *Nat Immunol* **8:** 873-881

Wilson N, Fremont D, Marrack P, Kappler J (2001) Mutations changing the kinetics of class II MHC peptide exchange. *Immunity* **14:** 513-522

Wilson NA, Wolf P, Ploegh H, Ignatowicz L, Kappler J, Marrack P (1998) Invariant chain can bind MHC class II at a site other than the peptide binding groove. *J Immunol* **161:** 4777-4784

Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AG, McCoy A, McNicholas SJ, Murshudov GN, Pannu NS, Potterton EA, Powell HR, Read RJ, Vagin A, Wilson KS (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* **67:** 235-242

Wucherpfennig KW, Sette A, Southwood S, Oseroff C, Matsui M, Strominger JL, Hafler DA (1994) Structural requirements for binding of an immunodominant myelin basic protein peptide to DR2 isotypes and for its recognition by human T cell clones. *J Exp Med* **179:** 279-290

Yaneva R, Springer S, Zacharias M (2008) Flexibility of the MHC class II peptide binding cleft in the bound, partially filled and empty states: A molecular dynamics simulation study. *Biopolymers*

York IA, Bhutani N, Zendzian S, Goldberg AL, Rock KL (2006) Tripeptidyl peptidase II is the major peptidase needed to trim long antigenic precursors, but is not required for most MHC class I antigen presentation. *J Immunol* **177:** 1434-1443

Zaheer ul H, Khan W (2011) Molecular and structural determinants of adamantyl susceptibility to HLA-DRs allelic variants: an in silico approach to understand the mechanism of MLEs. *J Comput Aided Mol Des* **25:** 81-101

Zarutskie JA, Busch R, Zavala-Ruiz Z, Rushe M, Mellins ED, Stern LJ (2001) The kinetic basis of peptide exchange catalysis by HLA-DM. *Proc Natl Acad Sci U S A* **98:** 12450--12455

Zarutskie JA, Sato AK, Rushe MM, Chan IC, Lomakin A, Benedek GB, Stern LJ (1999) A conformational change in the human major histocompatibility complex protein HLA-DR1 induced by peptide binding. *Biochemistry* **38:** 5878-5887

Zavala-Ruiz Z, Strug I, Walker BD, Norris PJ, Stern LJ (2004) A hairpin turn in a class II MHC-bound peptide orients residues outside the binding groove for T cell recognition. *Proc Natl Acad Sci U S A* **101:** 13279--13284

Zhao Y, Li Z, Drozd SJ, Guo Y, Mourad W, Li H (2004) Crystal structure of Mycoplasma arthritidis mitogen complexed with HLA-DR1 reveals a novel superantigen fold and a dimerized superantigen-MHC complex. *Structure* **12:** 277-288

Zhou Z, Callaway KA, Weber DA, Jensen PE (2009) Cutting edge: HLA-DM functions through a mechanism that does not require specific conserved hydrogen bonds in class II MHC-peptide complexes. *J Immunol* **183:** 4187--4191

Zhu J, Yamane H, Paul WE (2010) Differentiation of effector CD4 T cell populations (*). *Annu Rev Immunol* **28:** 445-489

Zhu Y, Rudensky AY, Corper AL, Teyton L, Wilson IA (2003) Crystal structure of MHC class II I-Ab in complex with a human CLIP peptide: prediction of an I-Ab peptide-binding motif. *J Mol Biol* **326:** 1157-1174

Zinkernagel RM, Doherty PC (1974) Restriction of in vitro T cell-mediated cytotoxicity in lymphocytic choriomeningitis within a syngeneic or semiallogeneic system. *Nature* **248:** 701-702

8 Own publications

Gupta S, Höpner S, Rupp B, **Günther S**, Dickhaut K, Agarwal N, Cardoso MC, Kühne R, Wiesmüller K-H, Jung G, Falk K, Rötzschke O (2008) Anchor side chains of short peptide fragments trigger ligand-exchange of class II MHC molecules. *PLoS ONE* **3:** e1814

Schlundt A, Kilian W, Beyermann M, Sticht J, **Günther S**, Höpner S, Falk K, Roetzschke O, Mitschang L, Freund C (2009) A xenon-129 biosensor for monitoring MHC-peptide interactions. *Angew Chem Int Ed Engl* **48:** 4142--4145

Günther S*, Schlundt A*, Sticht J, Roske Y, Heinemann U, Wiesmüller KH, Jung G, Falk K, Rötzschke O, Freund C (2010) Bidirectional binding of invariant chain peptides to an MHC class II molecule. *Proc Natl Acad Sci U S A* **107:** 22219-22224

Rupp B*, **Günther S***, Makhmoor T, Schlundt A, Dickhaut K, Gupta S, Choudhary I, Wiesmüller KH, Jung G, Freund C, Falk K, Rötzschke O, Kühne R (2011) Characterization of Structural Features Controlling the Receptiveness of Empty Class II MHC Molecules. *PLoS ONE* **6:** e18662

*) equal contribution

Acknowledgements

This success of this work was only possible with the constant support of a number of people.

First of all, I would like to thank Dr. Kirsten Falk und Dr. Olaf Rötzschke for giving me the opportunity to work on this interesting topic and for their continuous support.

I also want to thank Prof. Dr. Christian Freund for giving me the opportunity to continue and finish my thesis in his laboratory.

I am also grateful to Prof. Dr. Udo Heinemann for being my doctor father right from the start of my thesis and for providing all the necessary support for the crystallographic part of this thesis.

Especially, I want to thank Andreas Schlundt. He was a perfect co-fighter against the miseries of MHC biology and a continuous source for discussions. I am also very grateful to Yvette Roske for so much support and help from crystallization to data collection and beyond. Moreover, I want to thank Bernd Rupp for giving me so much insight into dynamic aspects of protein structures. I also want to thank Jana Sticht for her constant input, especially at the end of the thesis.

Furthermore, I want to thank various people that had an impact on this work at various stages: Ronald Kühne, Karl-Heinz Wiesmüller, Günther Jung, Christoph Wirblich and Talat Makhmoor. I want to thank all members of the former AGFR, especially Katha, Sabine, Shashank, Sabrina, Mireille, Jörg, Steph, Sonja and Petra and also the members of the AG Freund, especially, Dani, Bernhard, Marek, Kathrin and all the others.

Zum Schluss möchte ich meiner Mutter und Harald für das Verständnis und die durchgehende Unterstützung auf all meinen Wegen bedanken. Ohne Euch wäre ich jetzt nicht hier.

Und zu guter Letzt möchte ich mich bei Dir bedanken, Rina, für Deine Geduld, Dein Verständnis, Deine Motivation und Unterstützung und fürs einfach Dasein.

Auf in die Zukunft!

Abbreviations

Å	Ångström
a.u.	arbitrary unit
aa	amino acid
ADP	anisotropic displacement paramter
AEP	asparaginyl endopeptidase
ahx	6-aminohexanoic acid
APC	antigen presenting cell
Bapa	*Bacillus anthracis* protective antigen
BESSY	Berliner Elektronen-Speicherring - gesellschaft für Synchrotronstrahlung
biot	biotin
BSA	bovine serum albumin
Cat	cathepsin
CD	circular dichrosim
CF	carboxyfluorescin
CLIP	class-ii-associated invariant chain peptide
cpm	count per minute
C-terminal	carboxy-terminal
Da	dalton
DC	dendritic cell
e	charge of an electron
EC50	half maximal effective concentration
ELISA	enzyme-linked immunosorbent assay
ER	endoplasmic reticulum
FI	fluorescence intensity
FP	fluorescence polarization
GILT	gamma-ifn-inducible lysosomal thiol reductase
GP64	glycoprotein 64
HA	hemagglutinin
HLA	human leukocyte antigen
HSQC	heteronuclear single quantum coherence
IFN	interferon
IL	interleukin
k	Boltzmann constant
MD	molecular dynamics
MES	2-(*N*-morpholino)ethanesulfonic acid
MHC	major histocompatibility complex
MIB	malonate, imidazole and boric acid
MIIC	MHC class II loading compartment
MMT	L-Malic acid, MES, Tris
MW	molecular weight
NMR	nuclear magnetic resonance
NOD	non-obese diabetic
N-terminal	amino-terminal
OD	optical density
p4a	vaccinia virus core protein 4a

PAGE	polyacrylamid gel electrophoresis
PAMP	pathogen-associated molecular patterns
PDB	protein data bank
PEG	poly ethylene glycol
Phlp5b	*Phleum pretense* pollen allergen
pMHC	peptide-MHC
ppm	parts per million
RMSD	root mean square deviation
rpm	round per minute
RT	room temperature
s.d.	standard deviation
SA	streptavidin
SDS	sodium dodecyl sulfate
T	temperature
TAP	transporter associated with antigen processing
teg	tetra-ethylene glycol
Tris	2-Amino-2-hydroxymethyl-propane-1,3-diol
TSA	thermo shift assay
TT	tetanus toxin
w/v	weight per volume
wt	wild type

Amino Acids

Full name	Three letter code	One letter code
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic acid	Asp	D
Cysteine	Cys	C
Glutamic Acid	Glu	E
Glutamine	Gln	Q
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

Curriculum vitae

Due to data privacy, the curriculum vitae is not included in the online version.

Due to data privacy, the curriculum vitae is not included in the online version.

Appendix

Table 17: HLA-DR1 restricted antigens with proven T cell stimulating activity as listed in the Immune Epitope Database (IEDB; http://www.iedb.org)

Epitope ID	Peptide sequence	Antigen	Organism
603	ACTSGVMTRGRLKAE	65 kDa lower matrix phosphoprotein	Cytomegalovirus
1046	AELKIYSVIQAEINKHLSSS	Virulence-associated V antigen	*Yersinia pestis*
1197	AETPGCVAYIGISFLDQASQ	Phosphate-binding protein pstS 1	*Mycobacterium tuberculosis*
1489	AGEFGTLRAGRVANQ	Major outer membrane protein P.IA	*Neisseria meningitidis*
1509	AGFKGEQGPKGEP	alpha-1 type II collagen	*Homo sapiens*
2048	AIPPRGTQAVVLKVYQNAGG	Immunogenic protein MPT64	Mycobacterium tuberculosis
3720	APQIVRGASEDVRKQPYNLTIAWF RMGG	Glycoprotein D	Herpes simplex virus
3783	APRMEPAAPPVAPAP	Cellular tumor antigen p53	*Homo sapiens*
4070	AQSLVIKLMPNITLL	measles virus	
4312	ASAGQISVQPTFSVQRNLPF	nucleoprotein NP	Influenza A virus
4375	ASDVETAEGGEIHELLRLQ	phosphoprotein	Measles virus
4868	ATAAVCLLFIQGYSIYENYGN	putative 16.3k protein	Modified Vaccinia Ankara virus
5422	AVLEDPYILLVSSKV	60 kDa chaperonin 2	*Mycobacterium tuberculosis*
5425	AVLEEPYILLVSSKV	60 kDa chaperonin 2	*Mycobacterium leprae*
5851	AYKEFRVVELD	Staphylokinase	*Staphylococcus aureus*
6795	CPKYVKQNTLKLATG	Hemagglutinin	Influenza A virus
6797	CPKYVKQNTLKLATGMRNV	Hemagglutinin	Influenza A virus
6920	CRFPNITNSHVPILQERPPLENRVLT G	envelope glycoprotein	Human T-lymphotropic virus1
7494	DAEFRHDSGYEVHHQKLVFFAEDV GSNKGAIIGLMVGGVVIAT	Amyloid beta A4 protein	*Homo sapiens*
7887	DDVVLATGSQDFFVRSIGSK	F1 capsule antigen	*Yersinia pestis*
8268	DFFVRSIGSKGGKLAAGKYT	F1 capsule antigen	*Yersinia pestis*
8622	DIEKKIAKMEKASSV	circumsporozoite protein	*Plasmodium falciparum*
8623	DIEKKIAKMEKASSVF	circumsporozoite protein	*Plasmodium falciparum*
8624	DIEKKIAKMEKASSVFN	circumsporozoite protein	*Plasmodium falciparum*
8625	DIEKKIAKMEKASSVFNV	circumsporozoite protein	*Plasmodium falciparum*
8626	DIEKKIAKMEKASSVFNVV	circumsporozoite protein	*Plasmodium falciparum*
8627	DIEKKIAKMEKASSVFNVVN	circumsporozoite protein	*Plasmodium falciparum*
8628	DIEKKIAKMEKASSVFNVVNS	circumsporozoite protein	*Plasmodium falciparum*
8632	DIEKKICKMEKCSSVFNVVNS	Circumsporozoite protein	*Plasmodium falciparum*
8734	DINNDIISDISGFNSSVITY	Tetanus toxin	*Clostridium tetani*
8869	DKFKTFEAAFTSSSKAAAA	Pollen allergen Phl p 5b	*Phleum pratense*
9297	DLTFIAEKNSFSEEPFQDEI	Tetanus toxin	*Clostridium tetani*
9326	DLVSQKTTQLSDITSRFNSA	Virulence-associated V antigen	*Yersinia pestis*
9568	DNVLDHLTGR	pertussis toxin S1 subunit	*Bordetella pertussis*
9573	DNVLDHLTGRSCQ	Pertussis toxin subunit 1	*Bordetella pertussis*
10684	DVNKSKTHISVNGRKI	Diphtheria toxin	*Corynebacterium diphtheriae*
10730	DVSLYGEIKAGVEGRNIQ	porin, class I outer membrane protein	*Neisseria meningitidis*
11404	EDNLGFLMHAPAFETAGTYLRLVKI NDWTEITQF	Glycoprotein D	Herpes simplex virus
11434	EDSEYLFRIVSTVLPHLCLDY	major core protein P4a	Modified Vaccinia Ankara virus
11519	EEAYHACDIKD	Pollen allergen Amb a 3	*Ambrosia artemisiifolia var. elatior*
11996	EFFWDANDIYRIF	HCMVUL83	Human herpesvirus 5
12067	EFRVVELDPSA	Staphylokinase	*Staphylococcus aureus*
12366	EHRVKRGLTVAVAGA	lipoprotein lpqH	*Mycobacterium tuberculosis*
12733	EKKIAFMEKASSVFNVV	circumsporozoite protein	*Plasmodium falciparum*
12737	EKKIAKKEKASSVFNVV	circumsporozoite protein	*Plasmodium falciparum*
12747	EKKIAKMEKASSDFNVV	circumsporozoite protein	*Plasmodium falciparum*
12748	EKKIAKMEKASSGFNVV	circumsporozoite protein	*Plasmodium falciparum*
12749	EKKIAKMEKASSKFNVV	circumsporozoite protein	*Plasmodium falciparum*
12750	EKKIAKMEKASSVDNVV	circumsporozoite protein	*Plasmodium falciparum*
12751	EKKIAKMEKASSVFEVV	circumsporozoite protein	*Plasmodium falciparum*

Epitope ID	Peptide sequence	Antigen	Organism
12752	EKKIAKMEKASSVFFVV	circumsporozoite protein	*Plasmodium falciparum*
12753	EKKIAKMEKASSVFKVV	circumsporozoite protein	*Plasmodium falciparum*
12754	EKKIAKMEKASSVFNDV	circumsporozoite protein	*Plasmodium falciparum*
12755	EKKIAKMEKASSVFNGV	circumsporozoite protein	*Plasmodium falciparum*
12756	EKKIAKMEKASSVFNKV	circumsporozoite protein	*Plasmodium falciparum*
12757	EKKIAKMEKASSVFNVV	circumsporozoite protein	*Plasmodium falciparum*
12758	EKKIAKMEKASSVFNVVNS	circumsporozoite protein	*Plasmodium falciparum*
12759	EKKIAKMEKASSVGNVV	circumsporozoite protein	*Plasmodium falciparum*
12760	EKKIAKMEKASSVKNVV	circumsporozoite protein	*Plasmodium falciparum*
12766	EKKIAKMFKASSVFNVV	circumsporozoite protein	*Plasmodium falciparum*
12767	EKKIAKMGKASSVFNVV	circumsporozoite protein	*Plasmodium falciparum*
12768	EKKIAKMKKASSVFNVV	circumsporozoite protein	*Plasmodium falciparum*
12780	EKKLAKMEKASSVFNVV	circumsporozoite protein	*Plasmodium falciparum*
13172	ELLVLMENERTLDFHD	hemagglutinin	Influenza A virus
13596	ENTIKIQEGSGLSKEE	heat shock protein 70, hsp70A2	*Mycobacterium leprae*
14070	ERRNKYLEEHPSAGKDPKKT	Nucleoprotein	Influenza A virus
14071	ERRNRYLEEHPSAGKDPKKT	nucleoprotein NP	Influenza A virus
14464	ETLLRAVESYLLAHS	Major pollen allergen Bet v 1-A	*Betula pendula*
14562	ETTGVVLLLEYIPEITLPVI	Tetanus toxin	*Clostridium tetani*
14956	EWSYIVEKANPVND	hemagglutinin	Influenza A virus
15056	EYLNKIQNSLSTEWSPCSVT	circumsporozoite protein	*Plasmodium falciparum*
15116	EYWGAQLNAMKGDLQSSLGA	Antigen 85-B	*Mycobacterium tuberculosis*
15309	FAYGSFVRTVSLPVGA	heat shock protein	*Mycobacterium tuberculosis*
15507	FEDQLVFNSISARALKAYFTA	putative 35.9k protein	Modified Vaccinia Ankara virus
15955	FGAQFITWRGIPLIPSDKV	Major membrane protein I	*Mycobacterium leprae*
16035	FGPTQFITWHGIRLPSDKV	Major membrane protein 1	*Mycobacterium avium*
16182	FIEDLEKVRVEQLTGHGSSV	Virulence-associated V antigen	*Yersinia pestis*
16392	FKKIAKMEKASSVFNVV	circumsporozoite protein	*Plasmodium falciparum*
17166	FNILGGWVAAQL	polyprotein	Hepatitis C virus
17207	FNNFTVSFWLRVPKVSASHLE	Tetanus toxin	*Clostridium tetani*
17576	FRDYVDRFYKTLRAEQASQD	Gag polyprotein	Human immunodeficiency virus 1
17605	FRKEIGRML	anchored capsid (anchC) protein	Dengue virus 4
17660	FRYMSSEPII	mRNA capping enzyme large subunit	Vaccinia virus
18366	FWRGENGRKTRSAYERMCNILKGK	Nucleoprotein	Influenza A virus
19039	GDLSIPSSELENIPSENQYF	protective antigen	*Bacillus anthracis*
19501	GFATQRLTSLFALGPSQK	Genome polyprotein	Hepatitis C virus
19613	GFRLGFLHSGTAKSV	Cellular tumor antigen p53	*Homo sapiens*
19864	GGGVTLLQAAPALD	60 kDa chaperonin 2	*Mycobacterium leprae*
19899	GGKLAAGKYTDAVTVTVSNQ	F1 capsule antigen	*Yersinia pestis*
20211	GIAGFKGEQGPKGEP	collagen, type II, alpha 1	*Homo sapiens*
20282	GIGNYTQINAASVGLR	Major outer membrane protein P.IA	*Neisseria meningitidis*
20597	GKKIAKMEKASSVFNVV	circumsporozoite protein	*Plasmodium falciparum*
20705	GKTTQCMNIMESIPANTIKYL	thymidylate kinase	Vaccinia virus
21476	GNNDNVLDHLTGR	Pertussis toxin subunit 1	*Bordetella pertussis*
21686	GPLKAEIAQRLE	Matrix protein 1	unidentified influenza virus
21839	GPWVPEQWMFQGAPPSQGTD	EBNA3A nuclear protein	Human herpesvirus 4
22023	GRETVIEYLVSFGVW	core	Hepatitis B virus
22323	GSFGSAPSNGWLKLGLVEFG	mtp40 protein	*Mycobacterium tuberculosis*
22612	GTAYNALAPKGAPNP	Hexon protein	Human adenovirus 5
22990	GVGTMVMELIRMIKRGINDR	nucleoprotein NP	Influenza A virus
23155	GVSTANATVYMIDSVLMPP	Immunogenic protein MPT70	*Mycobacterium tuberculosis*
23565	HASLVIKLMPNITLL	Fusion glycoprotein F0	Measles virus strain Edmonston
24568	HQALVIKLMPNITLL	Fusion glycoprotein F0	Measles virus strain Edmonston
24610	HQSAVIKLMPNITLL	Fusion glycoprotein F0	Measles virus strain Edmonston
24615	HQSLAIKLMPNITLL	Fusion glycoprotein F0	Measles virus strain Edmonston
24618	HQSLVAKLMPNITLL	Fusion glycoprotein F0	Measles virus strain Edmonston
24620	HQSLVIALMPNITLL	Fusion glycoprotein F0	Measles virus strain Edmon-

Epitope ID	Peptide sequence	Antigen	Organism
			ston
24625	HQSLVIKAMPNITLL	Fusion glycoprotein F0	Measles virus strain Edmonston
24626	HQSLVIKLAPNITLL	Fusion glycoprotein F0	Measles virus strain Edmonston
24628	HQSLVIKLMANITLL	Fusion glycoprotein F0	Measles virus strain Edmonston
24630	HQSLVIKLMPAITLL	Fusion glycoprotein F0	Measles virus strain Edmonston
24631	HQSLVIKLMPNATLL	Fusion glycoprotein F0	Measles virus strain Edmonston
24633	HQSLVIKLMPNIALL	Fusion glycoprotein F0	Measles virus strain Edmonston
24634	HQSLVIKLMPNITAL	Fusion glycoprotein F0	Measles virus strain Edmonston
24637	HQSLVIKLMPNITLA	Fusion glycoprotein F0	Measles virus strain Edmonston
24640	HQSLVIKLMPNITLL	Fusion glycoprotein F0	Measles virus strain Edmonston
25269	IAFNSGLEPGVVAEK	60 kDa chaperonin 2	*Mycobacterium tuberculosis*
25272	IAFNSGMEPGVVAEK	60 kDa chaperonin 2	*Mycobacterium leprae*
25561	IDELKTNSSLLTSILTYHVV	Immunogenic protein MPT70	*Mycobacterium tuberculosis*
25594	IDKDIRKILSGYIVEIEDTE	protective antigen	*Bacillus anthracis*
25853	IEKKIAKMEKASSVFNVVNS	circumsporozoite protein	*Plasmodium falciparum*
26269	IGLSMAGSSAMILAAY	Antigen 85-B	*Mycobacterium tuberculosis*
26406	IHRGKPFQLEAVFE	Mite group 2 allergen Der p 2	*Dermatophagoides pteronyssinus*
26805	IKIFMLVTAVVLLCCSGVAT	Immunogenic protein MPT64	*Mycobacterium tuberculosis*
27941	IPLSEMVVKLTAVCMKCFKEA	thymidine kinase	Modified Vaccinia Ankara virus
27951	IPLYKKMEAVKLRDLKTYSV	Tetanus toxin	*Clostridium tetani*
28304	IRGSVTPAVSQFNARTADGI	MPT63	*Mycobacterium tuberculosis*
28327	IRLHTLLAVLTAAPLLLAAA	Phosphate-binding protein pstS 1	*Mycobacterium tuberculosis*
29056	ITQFILEHRAKGSCKYALPLRIPPSACLSPQ	Glycoprotein D	Herpes simplex virus
29712	IYTYRIIKSSFPVPT	putative 43.5k protein	Modified Vaccinia Ankara virus
29870	KALPVVLENARILKNCV	Major allergen I polypeptide chain 1	*Felis catus*
30059	KCIEWEKAQHGA	Major pollen allergen Art v 1	*Artemisia vulgaris*
30269	KDVQLKNITDYMYLTNAPSY	Tetanus toxin	*Clostridium tetani*
31385	KIPFLLLSGSPITNT	early gene transcription factor VETF 70 kDa small subunit	Modified Vaccinia Ankara virus
31484	KIYSYFPSVISKV	Tetanus toxin	*Clostridium tetani*
31562	KKIAKMEKASSVFNVVNS	circumsporozoite protein	*Plasmodium falciparum*
31596	KKKIAKMEKASSVFNVV	circumsporozoite protein	*Plasmodium falciparum*
33253	KRWIILGLNKIVRMYSPTSI	gag polyprotein	Human immunodeficiency virus 1
33866	KTSLYNLRRGTALA	EBNA-1 protein	Human herpesvirus 4
34301	KWIEVYKLVKAKWLGTVNTQ	Tetanus toxin	*Clostridium tetani*
34509	KYLKKIKNSLSTEWSPCSVT	circumsporozoite protein	*Plasmodium falciparum*
34577	KYQEFFWDAND	HCMVUL83	Human herpesvirus 5
34578	KYQEFFWDANDIYRI	65 kDa lower matrix phosphoprotein	Cytomegalovirus
34771	LAGEFGTLRAGRVANQ	Major outer membrane protein P.IA	*Neisseria meningitidis*
34772	LAGEFGTLRAGRVANQFD	Major outer membrane protein P.IA	*Neisseria meningitidis*
34823	LAIYHPQQFVYAGAMSGLLD	Secreted antigen Ag85A	*Mycobacterium tuberculosis*
35023	LAWTRQQNQWKEPDV	HCMVUL83	Human herpesvirus 5
35217	LDLLFWEQGGLCKALQEQCRFP	envelope glycoprotein	Human T-lymphotropic virus 1
35767	LEYIPEITLPVIA	Tetanus toxin	*Clostridium tetani*
35768	LEYIPEITLPVIAALSIAES	Tetanus toxin	*Clostridium tetani*
35919	LFNILGGWVA	polyprotein	Hepatitis C virus
36647	LINSTKIYSYFPSVISKVNQ	Tetanus toxin	*Clostridium tetani*
37178	LLDRLVRLIGNPDVSGPKLT	Nucleoprotein	Measles virus strain Edmonston

Epitope ID	Peptide sequence	Antigen	Organism
37930	LLVSSKVSTVKDLLP	60 kDa chaperonin 2	*Mycobacterium tuberculosis*
38574	LPLKMLNIPSINVH	65 kDa lower matrix phosphoprotein	Human herpesvirus 5
38763	LPVVLENARILKNCVDAK	Major allergen I polypeptide chain 1	*Felis catus*
39248	LRYDTEYYLIPVASSSKDVQ	Tetanus toxin	*Clostridium tetani*
40485	LVNLLPAILS	polyprotein	Hepatitis C virus
41001	MAAHKGLMNIALAISAQQVN	Phosphate-binding protein pstS 1	*Mycobacterium tuberculosis*
41851	MKKISSVIAIALFGTIATAN	F1 capsule antigen	*Yersinia pestis*
41893	MKVVIVTSVASLLDASIQFQK	putative 26.5k protein	Modified Vaccinia Ankara virus
42240	MNSLSIFFIVVATAAVCLLFI	putative 16.3k protein	Modified Vaccinia Ankara virus
42443	MQWNSTTFHQTLQDPRVRGLYFPAGG	Major surface antigen	Hepatitis B virus subtype ayw
43228	NANIRYVNTGTAPIYNVLPT	protective antigen	*Bacillus anthracis*
43697	NENIKELLDKIN	Merozoite surface protein 1	*Plasmodium falciparum K1*
43881	NFLLPDAQSIQAAAAGFASK	Phosphate-binding protein pstS 1	*Mycobacterium tuberculosis*
43952	NFWRGENGRKTR	Nucleoprotein	unidentified influenza virus
44200	NIDDNTIYQYLYAQKSPTTL	Tetanus toxin	*Clostridium tetani*
44267	NIKNQLAELNATNIYTVLDK	protective antigen	*Bacillus anthracis*
44284	NILMQYIKANSKFIGITELK	Tetanus toxin	*Clostridium tetani*
44558	NKNKIPFLLLSGSPITNTPNT	VETF-1, early transcription factor	Modified Vaccinia Ankara virus
44788	NLLPAIL	polyprotein	Hepatitis C virus
45121	NNHQFTTKVIGKDSRDFDIS	F1 capsule antigen	*Yersinia pestis*
45282	NNVLSPLPSQAMDDL	Cellular tumor antigen p53	*Homo sapiens*
45727	NRRGLDLLFWEQGGLCKALQE	envelope glycoprotein	Human T-lymphotropic virus 1
46139	NTKEGRYLVLKAVKVCDVRTV	11k DNA binding phosphoprotein	Modified Vaccinia Ankara virus
46273	NTSTREYLKLIGITAIMFATY	putative 25.9k protein	Modified Vaccinia Ankara virus
46789	PAFEWYYQSGLSIVMP	Antigen 85-B	*Mycobacterium tuberculosis*
46895	PAPQAPYQGYQEPPAPQAPY	Epstein-Barr nuclear antigen 6	Human herpesvirus 4 (strain B95-8)
47237	PEAKYDAYVATLSEALRIIAG	Pollen allergen Phl p 5a	*Phleum pratense*
47452	PEVIPMFSALSEGATP	gag polyprotein	Human immunodeficiency virus 1
47691	PGKLDVNKSKTHISVN	Diphtheria toxin	*Corynebacterium diphtheriae*
47945	PIPFSYSKNLDCWVDNEEDI	Tetanus toxin	*Clostridium tetani*
48231	PKYVKQNTLKLA	Hemagglutinin	Influenza A virus
48237	PKYVKQNTLKLAT	Hemagglutinin	Influenza A virus
48376	PLKAEIAQRLEDV	Matrix protein 1	Influenza A virus
48381	PLKMLNIPSINVHHY	65 kDa lower matrix phosphoprotein	Cytomegalovirus
48757	PPAYRPPNAPILSTL	precore/core protein	Hepatitis B virus
49220	PRMCSLMQGSTLPRRSGAAG	nucleoprotein NP	Influenza A virus
49283	PRYVKQNTLKLAT	Hemagglutinin	unidentified influenza virus
50262	QAGFFLLTRILTIPQS	Major surface antigen	Hepatitis B virus
50441	QDAYNAAGGHNAVFNF	Antigen 85-B	*Mycobacterium tuberculosis*
51026	QHRDVLQLYAPEAFNYMDKF	Lethal factor	*Bacillus anthracis*
51037	QHYREVAAAKSSE	Trans-activator protein BZLF1	Human herpesvirus 4
51732	QNLLKAEKGNKAAAQR	Histone H1-like protein Hc1	*Chlamydia trachomatis*
51891	QPPSLPITVYYAVLERACRSVLLNAPSEAPQIVR	Glycoprotein D	Herpes simplex virus (type 1 / strain 17)
52025	QQFIYAGSLSALLDPSQGM	Antigen 85-B	*Mycobacterium tuberculosis*
52471	QTANVVFRYMSSEPIIFGESS	mRNA capping enzyme, large subunit	Modified Vaccinia Ankara virus
52929	QYIKANSKFIGITE	Tetanus toxin	*Clostridium tetani*
52988	QYLKKIQNSLSTEWSPCSVT	circumsporozoite protein	*Plasmodium falciparum*
53706	RFGISNYCQIYPPNANKIREAL	Der p 1 allergen	*Dermatophagoides pteronyssinus*
53908	RGLDLLFWEQGGLCK	envelope glycoprotein	Human T-lymphotropic virus 1
54022	RGYFKMRTGKSSIMRS	hemagglutinin HA1	Influenza A virus

Epitope ID	Peptide sequence	Antigen	Organism
54955	RMVLSAFDERRNRYLEEHPS	nucleoprotein NP	Influenza A virus
55882	RSLFPEFSELFAAFPSFAGL	14 kDa antigen	Mycobacterium tuberculosis
56039	RTEIIRMMESARPEDVSFQG	nucleoprotein NP	Influenza A virus
56304	RVGYNAPGIPLYKKMEAVKL	Tetanus toxin	*Clostridium tetani*
56724	SACLSPQAYQQGVTVDSIGMLPRFI PENQRTVAVY	Glycoprotein D	Herpes simplex virus (type 1 / strain 17)
57415	SEFAYGSFVRTVSL	14 kDa antigen	*Mycobacterium tuberculosis*
57417	SEFAYGSFVRTVSLPVGADE	14 kDa antigen	*Mycobacterium tuberculosis*
57719	SFAAAGLAALAVAVSPPAA	Immunogenic protein MPT70	*Mycobacterium tuberculosis*
58175	SGPLKAEIAQKLEDVFAGKN	matrix protein 1	Influenza A virus
58177	SGPLKAEIAQRLEDV	matrix protein M1	Influenza A virus
58375	SHLECRTFFLTQGALLNDKH	neuraminidase	Influenza A virus
58602	SILGGVATYGA	hypothetical protein SMU.616	*Streptococcus mutans*
58968	SKYALVDASLKMADPNRFRGKDLP VLDQL	Glycoprotein D	Herpes simplex virus
59332	SLLVPFVQWFVGLSPTVWLSV	Major surface antigen	Hepatitis B virus subtype ayw
59548	SLVGIDPFRLLQNSQVFSLI	nucleoprotein NP	Influenza A virus
59747	SMRYQSLIPRLVEFF	core protein 4a	Modified Vaccinia Ankara virus
60989	SSFFRNVVWLIKK	hemagglutinin	Influenza A virus
62947	TAPIYNVLPTTSLVLGKNQT	protective antigen	*Bacillus anthracis*
63100	TCSDKSRPLNDLVSQKTTQL	Virulence-associated V antigen	*Yersinia pestis*
63277	TDYMYLTNAPSYTNGKLNIY	Tetanus toxin	*Clostridium tetani*
63960	TGSGAGIAQAAAGTVNI	Phosphate-binding protein pstS 1	*Mycobacterium tuberculosis*
64143	TIAWFRMGGNCAIPITVMEYTECS YNKS	Glycoprotein D	Herpes simplex virus
64592	TKIYSYFPSVISKV	Tetanus toxin	*Clostridium tetani*
65115	TLTYKEGAPITIMDNGNIDT	F1 capsule antigen	*Yersinia pestis*
65307	TNEFISFLLLTSIPIYNILFW	putative 31.9k protein	Modified Vaccinia Ankara virus
65995	TRANPNPYTSRRSVASIVGTLVRM	Pertussis toxin subunit 1	*Bordetella pertussis*
66086	TRQQNQWKEPDVYYT	65 kDa lower matrix phosphoprotein	Cytomegalovirus
66309	TSLNFLGGSPVCLGQ	Major surface antigen	Hepatitis B virus subtype adw2
66310	TSLNFLGGTTVCLGQ	Major surface antigen	Hepatitis B virus subtype ayw
66322	TSLYNLRRGTALA	Epstein-Barr nuclear antigen 1	Human herpesvirus 4
66323	TSLYNLRRGTALAI	Epstein-Barr nuclear antigen 1	Human herpesvirus 4
66503	TSYDQIIAGVDYDFSKRTSAIVSGA WLKRNTG	Major outer membrane protein P.IA	*Neisseria meningitidis*
66607	TTFHQTLQDPRVRGL	Major surface antigen	Hepatitis B virus subtype ayw
67007	TVFYNIPPMPL	Epstein-Barr nuclear antigen 2	Human herpesvirus
67907	VDAQGTLSKIFKLGGRDSRS	Myelin basic protein	*Homo sapiens*
67917	VDCYINLGARWSLDY	Hexon protein	Human adenovirus 5
68559	VFRYMSSEPI	mRNA capping enzyme large subunit	Vaccinia virus
69355	VLACAIATHAKIRD	Mite group 2 allergen Der p 2	*Dermatophagoides pteronyssinus*
69953	VMGDDGVLACAIAT	Mite group 2 allergen Der p 2	*Dermatophagoides pteronyssinus*
70067	VNFTDAAGDPMYLTFTSQDG	F1 capsule antigen	*Yersinia pestis*
70124	VNLVDTLNSGQYT	Immunogenic protein MPT70	*Mycobacterium tuberculosis*
70463	VPRISYAHGFDLIERGKKG	Major outer membrane protein P.IA	*Neisseria meningitidis*
70903	VSDLKSSTAVIPGYPVAGQV	MPT63	*Mycobacterium tuberculosis*
71155	VSTIVPYIGPALNI	Tetanus toxin	*Clostridium tetani*
72675	WKGIYTYRIIKSSFPVPTIKS	Putative 43.5k protein	Modified Vaccinia Ankara virus
72942	WPKFRVVKPNSFTF	RNA polymerase subunit rpo132	Vaccinia virus
73049	WRRRPLSSALLSFGLLLGGLPL	Uncharacterized protein	*Mycobacterium tuberculosis*
73378	YAIVNNRQKDAATAQTLQAF	Phosphate-binding protein pstS 1	*Mycobacterium tuberculosis*
73603	YDVPDYASLRSLVASS	Hemagglutinin	Influenza A virus
74220	YIGQFDMRFLNSLAIHEKFDA	topoisomerase II	Modified Vaccinia Ankara virus
75137	YNDKLPLYISNPNYKVNVYA	protective antigen	*Bacillus anthracis*
75172	YNINISLPSYYPDQKSLENY	Immunogenic protein MPT64	*Mycobacterium tuberculosis*
75477	YQEFFWDANDIYRIF	HCMVUL83	Human herpesvirus 5

Epitope ID	Peptide sequence	Antigen	Organism
76160	YTQINAASVGLRHKF	Major outer membrane protein P.IA	*Neisseria meningitidis*
76317	YVKQNTLKL	hemagglutinin HA1	Influenza A virus
76318	YVKQNTLKLA	hemagglutinin HA1	Influenza A virus
76444	YVYAKEGYEPVLVIQSSEDY	Lethal factor	*Bacillus anthracis*
77778	GDLLGILESRGIKAR	fusion protein	Measles virus
77804	TVPKYVATQGYLISN	fusion protein	Measles virus
79513	ALLLLVILAGPCILR	Envelope glycoprotein gp62	Human T-cell lymphotrophic virus type 1
79650	SQWAREALQTGITLV	Envelope glycoprotein gp62	Human T-cell lymphotrophic virus type 1
79668	VILAGPCILRQLRHL	Envelope glycoprotein gp62	Human T-cell lymphotrophic virus type 1
79784	DPFRLLQNSQVYS	Nucleoprotein	Influenza A virus
79792	DTIIFEANGNLIA	Hemagglutinin	Influenza A virus
79802	EIIRMMESARPED	Nucleoprotein	Influenza A virus
79808	ELIHVLHGLYGMQ	tetanus toxin text	*Clostridium tetani E88*
79809	ELLVLLENERTLD	Hemagglutinin	Influenza A virus
79816	EVYKLVKAKWLGT	tetanus toxin text	*Clostridium tetani E88*
79835	GELIGILNAAKVPAD	triosephosphate isomerase 1	*Homo sapiens*
79836	GELIGTLNAAKVPAD	triosephosphate isomerase 1	*Homo sapiens*
79844	GPLKAEIAQRLED	matrix protein 1	Influenza A virus
79846	GQISIQPTFSVQRN	Nucleoprotein	Influenza A virus
79881	KGFKGVDAQGTLSKI	MBP protein	*Homo sapiens*
79946	NKYLEEHPSAGKD	Nucleoprotein	Influenza A virus
79972	PLYKKMEAVKLRD	tetanus toxin text	*Clostridium tetani E88*
79984	QEIYMQHTYPISA	tetanus toxin text	*Clostridium tetani E88*
80042	SVIEKMNTQFTAV	Hemagglutinin	Influenza A virus
80302	AKAFAYYIEPQHRDVLQLYA	Lethal factor	*Bacillus anthracis*
80503	ASYISCTANSWNVIP	putative B5R	Vaccinia virus Copenhagen
81917	EYIRIDAKVVPKSKIDTKIQ	Lethal factor	*Bacillus anthracis*
83206	HQSIGSTLYNKIYLYENMNI	Lethal factor	*Bacillus anthracis*
83456	IGLEIKDVQIIKQSEKEYIRIDAKVVP	Lethal factor	*Bacillus anthracis*
84412	KIYLYENMNINNLTATLGAD	Lethal factor	*Bacillus anthracis*
84479	KLITFNVHNRYASNIVESAY	Lethal factor	*Bacillus anthracis*
85110	LENGKLILQRNIGLEIKDVQI	Lethal factor	*Bacillus anthracis*
85862	LSMITMSAFLIVRLN	putative A33R	Vaccinia virus Copenhagen
86334	MKTISVVTLLCVLPA	putative B5R	Vaccinia virus Copenhagen
86731	NGLYYQGSCYILHSD	putative A33R	Vaccinia virus Copenhagen
86808	NKALGLPKYTKLITFNVHNR	Lethal factor	*Bacillus anthracis*
87648	QNSNEVQEVFAKAFAYYIEP	Lethal factor	*Bacillus anthracis*
87771	QYVKQNTLKLATGMRNVPEK	Hemagglutinin	Influenza A virus
88294	RVIGLCIRISMVISL	putative A33R	Vaccinia virus Copenhagen
89938	VENTEKALNVYYEIGKILSR	Lethal factor	*Bacillus anthracis*
90174	VLVIQSSEDYVENTEKALNV	Lethal factor	*Bacillus anthracis*
94612	NKSLGACPIRTQPRWNYYDSFSAVSEDNLGF	Glycoprotein D	Herpes simplex virus
95169	YRILQRGLLGRSQ	nonstructural protein 3	Dengue virus 1
95245	ANGNLIAPWYAFALSRGF	hemagglutinin	Influenza A virus
95288	CQTPQGAINSSLPFQNVH	hemagglutinin	Influenza A virus
95353	ELLISKESWSYIVETPNP	hemagglutinin	Influenza A virus
95364	EQLSSVSSFERFEIFPKE	hemagglutinin	Influenza A virus
95390	FPKESSWPNHTVTGVSAS	hemagglutinin	Influenza A virus
95402	GDTIIFEANGNLIAPWYA	hemagglutinin	Influenza A virus
95450	HPPNIGNQRALYHTENAY	hemagglutinin	Influenza A virus
95454	IAPLQLGNCSVAGWILGN	hemagglutinin	Influenza A virus
95457	IGECPKYVRSAKLRMVTG	hemagglutinin	Influenza A virus
95505	KLCLLKGIAPLQLGNCSV	hemagglutinin	Influenza A virus
95580	LVLLENERTLDFHDSNVK	hemagglutinin	Influenza A virus
95588	MDECDAKCQTPQGAINSS	hemagglutinin	Influenza A virus
95609	NCSVAGWILGNPECELLI	hemagglutinin	Influenza A virus
95618	NGKSSFYRNLLWLTGKNG	hemagglutinin	Influenza A virus
95623	NKVNSVIEKMNTQFTAVG	hemagglutinin	Influenza A virus
95630	NNKEKEVLVLWGVHHPPN	hemagglutinin	Influenza A virus
95725	QNVHPVTIGECPKYVRSA	hemagglutinin	Influenza A virus
95738	QRALYHTENAYVSVVSSH	hemagglutinin	Influenza A virus

Epitope ID	Peptide sequence	Antigen	Organism
95767	RNLLWLTGKNGLYPNLSK	hemagglutinin	Influenza A virus
95826	SFERFEIFPKESSWPNHT	hemagglutinin	Influenza A virus
95893	SWSYIVETPNPENGTCYP	hemagglutinin	Influenza A virus
95896	TCYPGYFADYEELREQLS	hemagglutinin	Influenza A virus
95905	TGMVDGWYGYHHQNEQGS	hemagglutinin	Influenza A virus
95923	TPNPENGTCYPGYFADYE	hemagglutinin	Influenza A virus
95943	TYADTICIGYHANNSTDT	hemagglutinin	Influenza A virus
95969	VLEKNVTVTHSVNLLEDS	hemagglutinin	Influenza A virus
95982	VRSAKLRMVTGLRNIPSI	hemagglutinin	Influenza A virus
95988	VTHSVNLLEDSHNGKLCL	hemagglutinin	Influenza A virus
96007	WTYNAELLVLLENERTLD	hemagglutinin	Influenza A virus
96016	YHKCNNECMESVKNGTYD	hemagglutinin	Influenza A virus
96266	EQYTHQDEIYEQVHSKGLYV	Lethal factor	*Bacillus anthracis*
96883	SKKFIDIFKEEGSNLTSYGR	Lethal factor	*Bacillus anthracis*
99923	DAEFRHDSGYEVHHQKLVFFAEDV GSNK	Amyloid Beta- Peptide	*Homo sapiens*
101904	MHCQTTLKYAIKTGH	Glutamate decarboxylase 2	*Homo sapiens*
103218	GGVIHLSCKSGFILT	putative B5R	Vaccinia virus Copenhagen
103288	ISVVTLLCVLPAVVY	putative B5R	Vaccinia virus Copenhagen
103399	LWEIEFAKQL	mitochondrial ribosomal protein S31	*Homo sapiens*
103581	SAFLIVRLNQCMSAN	putative A33R	Vaccinia virus Copenhagen
103939	GLMQNCNQMHASYLFQQDK	Glutamate decarboxylase 2	*Homo sapiens*
103978	HTNVCFWYIPPSLRTLEDNE	Glutamate decarboxylase 2	*Homo sapiens*
104140	MYAMMIARFK	Glutamate decarboxylase 2	*Homo sapiens*
104141	MYAMMIARFKMF	Glutamate decarboxylase 2	*Homo sapiens*
104163	NYAFLHATDLLP	Glutamate decarboxylase 2	*Homo sapiens*
104268	SPSLWEIEFAKQLASV	28S ribosomal protein S31, mitochondrial	*Homo sapiens*
104275	SPSLWEIEIAKQLASV	28S ribosomal protein S31, mitochondrial	*Homo sapiens*
104914	NKKCDKKCIEWEKAQHGA	Major pollen allergen Art v 1	*Artemisia vulgaris*
105147	DSYEIMQGRAVV	synthetic peptide	synthetic peptide
105396	MYELLQPAAIV	synthetic peptide	synthetic peptide
105480	PSLWEIEFAKQLAS	28S ribosomal protein S31, mitochondrial	*Mus musculus*
105946	LVFFAEDVGSNKGAIIGLMVGGVVI	Amyloid Beta- Peptide	*Homo sapiens*
106420	FKGEQGPKGE	alpha-1 type II collagen	*Homo sapiens*
106454	GIAGFKGEQGPKGET	collagen, type II, alpha 1	*Homo sapiens*
106455	GIAGFKGEQGPKGET	collagen, type II, alpha 1	*Homo sapiens*
107133	GEPGIAGFKGDQGPKGEPGPA	Collagen alpha-1(II) chain	*Homo sapiens*
107138	GEPGIAGFKGEQGPKGEPGPA	Collagen alpha-1(II) chain	*Homo sapiens*
107149	GIAGFKGEQGPKGET	collagen, type II, alpha 1	*Rattus norvegicus*
107225	PYKVKQNTLKLAT	hemagglutinin	Influenza A virus
107398	GKPGIAGFKGEQGPKG	collagen, type II, alpha 1	*Homo sapiens*
107430	GRSFTLASSETGVGAP	Chitinase-3-like protein 1	*Homo sapiens*
107504	LGPKGQTGEPGIAGFKGEQGPKGE PGPAGPQGA	collagen, type II, alpha 1 isoform 1	*Homo sapiens*
109115	GIAGFKGEQGPKGEB	collagen, type II, alpha 1	*Homo sapiens*
109116	GIAGFKGEQGPKGEP	Collagen alpha-1(II) chain	*Homo sapiens*
111192	DENPVVHFFKNIVTPRTPP	MBP protein	*Homo sapiens*
111200	DKKQRFHNIRGRWTG	E6	Human papillomavirus type 16
111454	KIPKTITVSTGFMSI	E2 protein	Human papillomavirus type 16
111816	TAYSQQTRGLLG	polyprotein	Hepatitis C virus
113169	SKYLATASTMDHARHGFLPR	myelin basic protein	*Homo sapiens*
113267	AHKGFKGVDAQTLSK	myelin basic protein	*Homo sapiens*
114376	GTLSKIFKLGGRDSR	MBP protein	*Homo sapiens*
115821	KGFKGVDAQGTLSKIF	MBP protein	*Homo sapiens*
116104	HKGFKGVDAQGTLSK	MBP protein	*Homo sapiens*
116851	VFFAEDVGSNKGAIIGL	Amyloid beta A4 protein	*Homo sapiens*
117063	QDENPVVHFFKNIVTPRTPP	MBP protein	*Homo sapiens*

Table 18: Overview of MHC II structures deposited in the Protein Data Bank and determined in this thesis

Allele	Peptide	Other proteins	Length [in aa]	Overhang (visible/total)		Resolu-tion [Å]	PDB	Remarks
				N-terminal	C-terminal			
HLA-DP2	DR α-chain	-	15	2/2	4/4	3.25	3LQZ	peptide linked to MHC
HLA-DQ1	MBP (85-99)	TCR (Hy.1B11)	15	3/4	2/4	2.55	3PL6	peptide linked to TCR
HLA-DQ2	alpha1-gliadin (1-11)	-	11	2/2	0/0	2.22	1S9V	peptide linked to MHC
HLA-DQ6	hypocretin (1-13)	-	13	2/2	2/2	1.8	1UVQ	peptide linked to MHC, 6 aa from linker visible
HLA-DQ8	insulin beta chain (9-23)	-	14	2/2	3/3	2.4	1JK8	peptide linked to MHC
HLA-DQ8	alpha1-gliadin (223-240)	-	18	2/6	2/3	2.1	2NNA	peptide linked to MHC
HLA-DR1	(4-Dapa)-HA (306-318)	SEC3, 3B2	13	2/2	2/2	2.3	2IPK	modified P1 anchor
HLA-DR1	A2 (103-117)	-	15	2/4	2/3	2.45	1AQD	
HLA-DR1	Bapa (361-380)	-	20	5/5	2/6	1.9	unpublished	
HLA-DR1	CII (259-273)	-	15	2/4	2/2	3.1	2FSE	peptide linked to MHC
HLA-DR1	CLIP (102-120)	-	19	4/5	4/5	1.95	3PDO	
HLA-DR1	CLIP102-120, M107W	-	19	3/5	3/5	1.3	unpublished	
HLA-DR1	CLIP (106-120)	-	15	5/5	1/1	2.66	3PGC	flipped peptide
HLA-DR1	CLIP (106-120)	-	15	1/1	5/5	2.72	3PGD	
HLA-DR1	CLIP106-120, M115W	-	15	5/5	1/1	2.4	unpublished	flipped peptide
HLA-DR1	GAG (166-178)	SEC3, 3B2	13	2/2	2/2	2.25	1SJH	
HLA-DR1	GAG (166-181)	SEC3, 3B2	16	2/2	4/5	2.45	1SJE	
HLA-DR1	HA (306-318)	-	13	2/2	2/2	2.75	1DLH	
HLA-DR1	HA (306-318)	TCR (HA1.7)	13	2/2	2/2	2.6	1FYT	peptide linked to TCR
HLA-DR1	HA (306-318)	SEH	13	2/2	2/2	2.6	1HXY	
HLA-DR1	HA (306-318)	SEC3, wt	13	2/2	2/2	2.7	1JWM	
HLA-DR1	HA (306-318)	SEC3, 3B1	13	2/2	2/2	2.6	1JWS	
HLA-DR1	HA (306-318)	SEC3, 3B2	13	2/2	2/2	2.3	1JWU	
HLA-DR1	HA (306-318)	EBV, gp42	13	2/2	2/2	2.65	1KG0	
HLA-DR1	HA (306-318)	SEA, D227A	13	2/2	2/2	3.2	1LO5	

Allele	Peptide	Other proteins	Length [in aa]	Overhang (visible/total)		Resolu-tion [Å]	PDB	Remarks
				N-terminal	C-terminal			
HLA-DR1	HA (306-318)	MAM	13	2/2	2/2	2.6	1R5I	
HLA-DR1	HA (306-318)	SEI	13	2/2	2/2	2	2G9H	
HLA-DR1	HA (306-318)	MAM, TCR (TCR-T7)	13	2/2	2/2	2.4	2ICW	
HLA-DR1	HA (306-318)	MAM	13	2/2	2/2	3.0	2OJE	
HLA-DR1	HA (306-318)	-	13	2/2	2/2	2.3	2XN9	
HLA-DR1	mixture of peptides	SEB		4/4	2/2	2.7	1SEB	
HLA-DR1	P4a (293-307)	-	15	3/3	3/3	1.9	unpublished	
HLA-DR1	Phl p 5b (68-86)	-	19	2/2	2/8	2.4	unpublished	
HLA-DR1	pMART-1	-	15	3/3	2/2	2.1	3L6F	
HLA-DR1	synthetic peptide	-	15	2/2	4/4	2.4	1T5W	
HLA-DR1	synthetic peptide	SEC3, 3B2	15	2/2	4/4	2.4	1T5X	
HLA-DR1	synthetic peptide	SEC3, 3B2	9	0/0	0/0	2.1	1PYW	
HLA-DR1	TPI (23-37), T28I	SEC3, 3B2	15	3/3	3/3	2.4	1KLG	
HLA-DR1	TPI (23-37), T28I	TCR (E8)	15	3/3	3/3	2.8	2IAM	peptide linked to TCR
HLA-DR1	TPI (23-37), wt	SEC3, 3B2	15	3/3	3/3	1.93	1KLU	
HLA-DR1	TPI (23-37), wt	TCR (E8)	15	3/3	3/3	2.8	2IAN	peptide linked to TCR
HLA-DR1	TT (830-843)	-	14	1/1	1/4	1.8	unpublished	
HLA-DR1	TT (1147-1159)	-	13	1/1	3/3	1.4	unpublished	
HLA-DR2a	EBV (628-641)	-	14	3/3	2/2	3.1	1H15	
HLA-DR2a	MBP (86-105)	-	20	6/6	5/5	1.9	1FV1	
HLA-DR2a	MBP (89-101)	SpeC	13	2/2	2/5	3.2	1HQR	
HLA-DR2a	MBP (89-101)	TCR (3A6)	13	2/2	3/2	2.8	1ZGL	peptide linked to TCR
HLA-DR2b	EngA (346-360)	TCR (Ob.1A12)	>17/16	5/5	3/3	3.15	2WBJ	peptide linked to TCR
HLA-DR2b	MBP (85-99)	-	15	4/4	1/2	2.6	1BX2	peptide linked to MHC
HLA-DR2b	MBP (85-99)	TCR (Ob.1A12)	15	4/4	1/2	3.5	1YMM	peptide linked to TCR
HLA-DR3	CLIP (82-104)	-	23	4/9	2/5	2.75	1A6A	
HLA-DR4	synthetic peptide	SEB	6	0/0	0/0	2.45	1D5X	
HLA-DR4	synthetic peptide	SEB	7	0/0	0/0	2.45	1D6E	

Allele	Peptide	Other proteins	Length [in aa]	Overhang (visible/total)		Resolution [Å]	PDB	Remarks
				N-terminal	C-terminal			
HLA-DR4	synthetic peptide	SEB	7	0/0	0/0	2	1D5Z	
HLA-DR4	synthetic peptide	SEB	7	0/0	0/0	2	1D5M	
HLA-DR4	CII (1168–1180)	SEB	13	2/2	1/2	2.5	2SEB	
HLA-DR4	CII (1168–1180)	SEB	13				2SEB	
HLA-DR4	HA (306-318)	TCR (HA1.7)	13	2/2	2/2	2.4	1J8H	peptide linked to TCR
HLA-DR4	MBP (114-126)	TCR (MS2-3C8)	16	2/2	5/5	2.8	3O6F	peptide linked to TCR
HLA-DR52a	3pl6 (50-61)	-	12	1/4	2/2	2.25	2Q6W	
HLA-DR52c	Tu (343-355)	-	13	2/2	2/2	1.8	3C5J	peptide linked to TCR
I-A^b	p3K/Ea (54-66)	-	13	2/2	2/2*	2.5	1LNU	peptide linked to MHC, complete linker visible
I-A^b	CLIP (81-104)	-	24	4/10	2/5	2.15	1MUJ	peptide linked to MHC
I-A^b	p3K/Ea (54-66)	-	13	2/2	2/2	2.6	3C5Z	peptide linked to MHC, 1 aa from linker visible
I-A^b	p3K/Ea (54-66)	-	13	2/2	2/2	3.1	3C60	peptide linked to MHC
I-A^b	p3K/Ea (54-66)	-	13	2/2	1/2	3.4	3C6L	peptide linked to MHC
I-A^d	OVA (323-339)	-	17	3/3	0/5	2.6	1IAO	
I-A^d	HA (126-138)	-	13	2/2	2/2	2.4	2IAD	
I-A^{g7}	GAD65 (207-220)	-	14	2/2	3/3	2.6	1ES0	peptide linked to MHC
I-A^{g7}	HEL (11-24)	-	14	3/3	2/2	3.1	1F3J	peptide linked to MHC
I-A^{g7}	GAD (221-235)	-	15	2/3	3/3	3.1	3CUP	peptide linked to MHC, 1 aa from linker visible
I-A^{g7}	HEL (11-27)	TCR (21.30)	17	4/4	4/4	2.9	3MBE	peptide linked to MHC, 1 aa from linker visible
I-A^k	CA (134-146)	TCR (D10)	16	3/3	4/4	3.2	1D9K	peptide linked to MHC
I-A^k	HEL (47-62)	-	16	2/5	2/2	1.9	1IAK	peptide linked to MHC
I-A^k	CA (134-146)	CD4	16	3/3	4/4	4.3	1JL4	
I-A^u	MBP (Ac1-11)	-	12	3/3	1/1	2.2	1K2D	peptide linked to MHC
I-A^u	MBP (Ac1-11)	TCR (172.10)	12	2/3	1/1	2.42	1U3H	peptide linked to MHC
I-A^u	MBP (125-135)	-	11	4/4	1/1	2.15	2P24	peptide linked to MHC
I-A^u	MBP (Ac1-11, 4Y)	TCR (1934.4)	12	1/3	1/1	2.23	2PXY	peptide linked to MHC
I-A^u	MBP (Ac1-11, 4Y)	TCR (cl19)	12	1/3	1/1	2.7	2Z31	peptide linked to MHC

Allele	Peptide	Other proteins	Length [in aa]	Overhang (visible/total)		Resolu-tion [Å]	PDB	Remarks
				N-terminal	C-terminal			
I-E^k	Hb E73D (64-76)	-	14	3/4	0/0	1.9	1FNE	peptide linked to MHC, complete linker visible
I-E^k	Hb (64-76)	-	14	3/4	0/0	1.9	1FNG	peptide linked to MHC, complete linker visible
I-E^k	Hb (64-76)	-	14	4/4	0/0	2.3	1IEA	peptide linked to MHC, complete linker visible
I-E^k	HSP70	-	16	4/4	2/2	2.7	1IEB	peptide linked to MHC, complete linker visible
I-E^k	MCC (92-103)	-	12	3/3	0/0	2.8	1KT2	peptide linked to MHC, complete linker visible
I-E^k	PCC (92-104, T102S)	-	12	3/3	0/0	2.4	1KTD	peptide linked to MHC, complete linker visible
I-E^k	K3 (MCC variant)	-	14	3/4	1/1	2.5	1R5V	
I-E^k	K5 (MCC variant)	-	14	3/4	1/1	2.8	1R5W	
I-E^k	MCC (92-108)	TCR (2B4)	17	3/8	0/0	2.7	3QIB	peptide linked to TCR
I-E^k	MCC (92-108)	TCR (226)	17	3/8	0/0	2.7	3QIU	peptide linked to TCR
I-E^k	MCC (92-108, K104E)	TCR (226)	17	3/8/	0/0	3.3	3QIW	peptide linked to TCR
I-E^k, mut	Hb (64-76)	-	13	4/4	0/0	2.4	1I3R	peptide linked to MHC, complete linker visible