In Silico Biology 5 (2005) 505–528
IOS Press

T Cell Receptor/Peptide/MHC Molecular Characterization and Standardized pMHC Contact Sites in IMGT/3Dstructure-DB

Quentin Kaas and Marie-Paule Lefranc*

IMGT, the international ImMunoGeneTics information system®, Université Montpellier II, Laboratoire d'ImmunoGénétique Moléculaire LIGM, UPR CNRS 1142, Institut de Génétique Humaine IGH, 141 rue de la Cardonille, 34396 Montpellier Cedex 5, France
Tel.: +33 4 99 61 99 65; Fax: +33 4 99 61 99 01

Edited by E. Wingender; received 2 September 2005; revised and accepted 18 September 2005; published 20 October 2005

ABSTRACT: One of the key elements in the adaptive immune response is the presentation of peptides by the major histocompatibility complex (MHC) to the T cell receptors (TR) at the surface of T cells. The characterization of the TR/peptide/MHC trimolecular complexes (TR/pMHC) is crucial to the fields of immunology, vaccination and immunotherapy. In order to facilitate data comparison and cross-referencing between experiments from different laboratories whatever the receptor, the chain type, the domain, or the species, IMGT, the international ImMunoGeneTics information system® (http://imgt.cines.fr), has developed IMGT-ONTOLOGY, the first ontology in immunogenetics and immunoinformatics. In IMGT/3Dstructure-DB, the IMGT three-dimensional structure database, TR/pMHC molecular characterization and pMHC contact analysis are made according to the IMGT Scientific chart rules, based on the IMGT-ONTOLOGY concepts. IMGT/3Dstructure-DB provides the standardized IMGT gene and allele names (CLASSIFICATION), the standardized IMGT labels (DESCRIPTION) and the IMGT unique numbering (NUMEROTATION). As the IMGT structural unit is the domain, amino acids at conserved positions always have the same number in the IMGT databases, tools and Web resources. For the TR alpha and beta chains, the amino acids in contact with the peptide/MHC (pMHC) are defined according to the IMGT unique numbering for V-DOMAIN. The MHC cleft that binds the peptide is formed by two groove domains (G-DOMAIN), each one comprising four antiparallel beta strands and one alpha helix. The IMGT unique numbering for G-DOMAIN applies both to the first two domains (G-ALPHA1 and G-ALPHA2) of the MHC class I alpha chain, and to the first domain (G-ALPHA and G-BETA) of the two MHC class II chains, alpha and beta. Based on the IMGT unique numbering, we defined eleven contact sites for the analysis of the pMHC contacts. The TR/pMHC contact description, based on the IMGT numbering, can be queried in the IMGT/StucturalQuery tool, at http://imgt.cines.fr.

Availability: IMGT/3Dstructure-DB is freely available at http://imgt.cines.fr.

KEYWORDS: IMGT, T cell receptor, TR, major histocompatibility complex, MHC, pMHC, TR/peptide/MHC complex, TR/pMHC, three-dimensional structure, 3D structure, contact analysis, IMGT/3Dstructure-DB, IMGT/StructuralQuery, immunoinformatics, immunogenetics, immune system

INTRODUCTION

T cells are involved in the specific immune response against a stress of viral, bacterial, fungal or tumoral origin. They identify antigenic peptides presented by the major histocompatibility complex

*Corresponding author. E-mail: lefranc@ligm.igh.cnrs.fr.

Electronic publication can be found in *In Silico Biol.* **5**, 0046 <http://www.bioinfo.de/isb/2005/05/0046/>, 20 October 2005.

(MHC) cell surface glycoproteins. The recognition is carried out by the T cell receptor complex (TcR), a multisubunit transmembrane surface complex made up of a T cell receptor (TR) and of the CD3 chains, that is associated, in the immunological synapse, to the CD4 or CD8 coreceptors, to the CD28 and CTLA-4 costimulatory proteins, to the CD2 adhesion molecule and to intracellular kinases [1]. The TR directly binds the peptide/MHC complex (pMHC), and activates the T cell through interactions with the CD3 and other components of the TcR [2–4]. Three-dimensional (3D) structures of the TR, pMHC and TR/pMHC complexes provide an atomic description of their interactions [5,6].

Since 1989, IMGT, the international ImMunoGeneTics information system® [7–10], http://imgt.cines.fr, created by Marie-Paule Lefranc, Laboratoire d'ImmunoGénétique Moléculaire (LIGM) (Université Montpellier II and CNRS) at Montpellier, France, has offered standardized genetic and structural data on immunoglobulins (IG), TR and MHC, and on related proteins of the immune system (RPI) that belong to the immunoglobulin superfamily (IgSF) and to the MHC superfamily (MhcSF). In order to facilitate data comparison and cross-referencing between experiments from different laboratories whatever the receptor, the chain type, the domain, or the species, IMGT has developed IMGT-ONTOLOGY [11], the first ontology in immunogenetics and immunoinformatics.

Based on the IMGT-ONTOLOGY concepts, the IMGT Scientific chart provides the controlled vocabulary and the annotation rules necessary for the identification, the description, the classification and the numbering of the IG, TR, MHC and RPI [8]. The IDENTIFICATION concept refers to the IMGT standardized keywords indispensable for the sequence and 3D structure assignments. The DESCRIPTION concept provides the IMGT standardized labels used to describe structural and functional regions that compose IG, TR, MHC and RPI sequences and 3D structures. Standardized labels have also been defined to characterize the three-dimensional assembly of domains and chains. The CLASSIFICATION concept provides immunologists and geneticists with a standardized nomenclature per locus and per species. The human IG and TR gene nomenclature elaborated by IMGT was approved by the Human Genome Organisation (HUGO) Nomenclature Committee, HGNC [12], in 1999. The mouse IG and TR gene names with IMGT reference sequences were provided by IMGT to HGNC and to the Mouse Genome Database (MGD) [13], in July 2002. The NUMEROTATION concept provides the IMGT unique numbering for the IG and TR V-DOMAIN and V-LIKE-DOMAIN of the IgSF proteins other than IG or TR [14], and for the IG and TR C-DOMAIN and C-LIKE-DOMAIN of the IgSF proteins other than IG or TR [15]. An IMGT unique numbering has also been set up for the MHC G-DOMAIN and G-LIKE-DOMAIN of the MhcSF proteins other than MHC [16].

The IMGT standardization has allowed to build a unique frame for the comparison of the TR, peptides and MHC interactions in the different resources provided by the information system. IMGT/3Dstructure-DB [5], the IMGT structural database, is used with the IMGT sequence databases (IMGT/LIGM-DB [7, 8] and IMGT/MHC-DB [17]), the IMGT gene database (IMGT/GENE-DB [18]), the IMGT tools for sequence analysis (IMGT/V-QUEST [19], IMGT/JunctionAnalysis [20]) and the IMGT tool for 3D structure analysis (IMGT/StructuralQuery [5]), to explore the TR and MHC conserved structural features. In this paper, we describe the molecular characterization and standardized contact analysis of the TR/pMHC complexes in IMGT/3Dstructure-DB. Coordinate files are from IMGT/3Dstructure-DB [5], http://imgt.cines.fr, with original crystallographic data from the Protein Data Bank PDB [6]. Eleven IMGT pMHC contact sites were defined (C1 to C11) which can be used to compare pMHC interactions. We provide the description of the interactions of the TR V-ALPHA and TR V-BETA with MHC and the peptide using the IMGT unique numbering for V-DOMAIN [14] and the IMGT unique numbering for G-DOMAIN [16], which allows, for the first time, to compare interaction data, whatever the TR gene group (TRAV, TRBV), whatever the MHC class (MHC-I, MHC-II), and whatever the species (*Homo sapiens, Mus musculus*).

Table 1
IMGT standardized labels for the DESCRIPTION of the T cell receptors, chains, domains and regions

IMGT receptor labels	IMGT chain labels	IMGT domain labels	IMGT region labels
TR-ALPHA_BETA	TR-ALPHA	V-ALPHA	V-J-REGION
		C-ALPHA	Part of C-REGION (1)
	TR-BETA	V-BETA	V-D-J-REGION
		C-BETA	Part of C-REGION (1)
TR-GAMMA_DELTA	TR-GAMMA	V-GAMMA	V-J-REGION
		C-GAMMA	Part of C-REGION (1)
	TR-DELTA	V-DELTA	V-D-J-REGION
		C-DELTA	Part of C-REGION (1)

(1) The TR chain C-REGION also includes the CONNECTING-REGION, the TRANSMEMBRANE-REGION and the CYTOPLASMIC-REGION which are not present in the 3D structures (Correspondence between labels for IG and TR domains in IMGT/3Dstructure-DB and IMGT/LIGM-DB, IMGT Scientific chart).

Table 2
IMGT standardized labels for the DESCRIPTION of the MHC receptors, chains, domains and domain numbers

IMGT receptor labels	IMGT chain labels	IMGT domain labels	Domain numbers
MHC-I-ALPHA_B2M	I-ALPHA	G-ALPHA1	[D1]
		G-ALPHA2	[D2]
		C-LIKE	[D3] (1)
	B2M	C-LIKE	[D]
MHC-II-ALPHA_BETA	II-ALPHA	G-ALPHA	[D1]
		C-LIKE	[D2] (1)
	II-BETA	G-BETA	[D1]
		C-LIKE	[D2] (1)

(1) The I-ALPHA, II-ALPHA and II-BETA chains include at the C-terminal end of the C-LIKE-DOMAIN, the CONNECTING-REGION, the TRANSMEMBRANE-REGION and the CYTOPLASMIC-REGION which are not present in the 3D structures.

TR AND MHC CHAINS AND DOMAINS

The T cell receptor (TR) is made of two chains, an alpha chain (TR-ALPHA) and a beta chain (TR-BETA) for the TR-ALPHA_BETA receptor, a gamma chain (TR-GAMMA) and a delta chain (TR-DELTA) for the TR-GAMMA_DELTA receptor [1]. Each complete TR chain comprises an extracellular region made up of a variable domain V-DOMAIN (for instance, V-ALPHA for the alpha chain) and a constant domain C-DOMAIN (for instance, C-ALPHA for the alpha chain), a connecting region, a transmembrane region and a very short intracytoplasmic region (Table 1, Fig. 1).

The MHC-I is formed by the association of an heavy chain (I-ALPHA) and a light chain (beta-2-microglobulin B2M) (Table 2, Fig. 1). The MHC-II is an heterodimer formed by the association of an alpha chain (II-ALPHA) and a beta chain (II-BETA). The I-ALPHA chain of the MHC-I, and the II-ALPHA and II-BETA chains of the MHC-II comprise an extracellular region made of three domains for the MHC-I and of two domains for the MHC-II chains, a connecting region, a transmembrane region and an intracytoplasmic region.

The I-ALPHA chain comprises two groove domains (G-DOMAIN), G-ALPHA1 [D1] and G-ALPHA2 [D2], and one C-LIKE domain [D3]. The B2M corresponds to a single C-LIKE domain. The II-ALPHA chain and the II-BETA chain each comprises two domains, G-ALPHA [D1] and C-LIKE [D2], and G-BETA [D1] and C-LIKE [D2] (Table 2). Only the extracellular region that corresponds to these domains has been crystallized (Fig. 1).

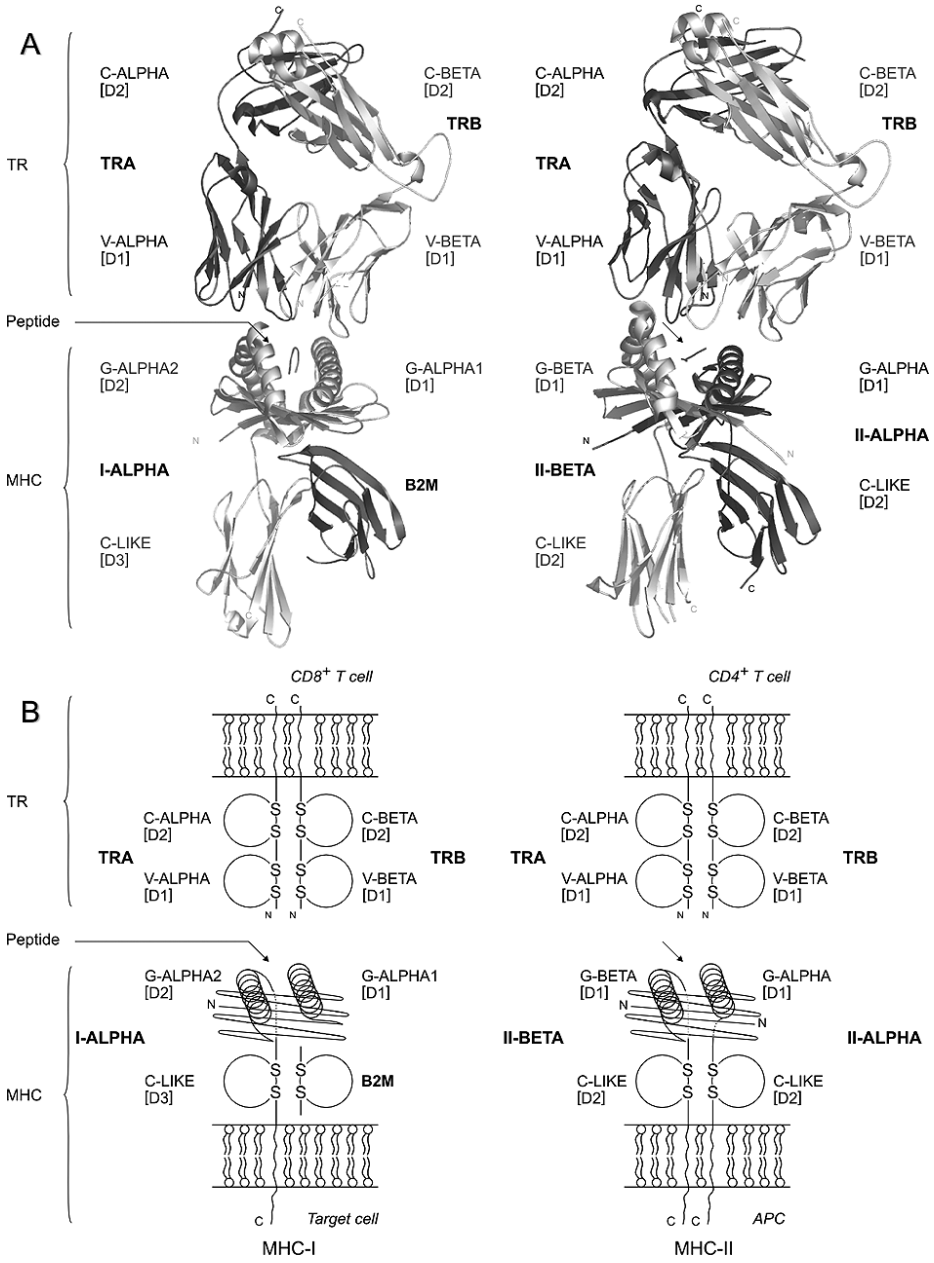

Fig. 1. T cell receptor/peptide/MHC complexes with MHC class I (TR/pMHC-I) and MHC class II (TR/pMHC-II). [D1], [D2] and [D3] indicate the domains. (A) 3D structures of TR/pMHC-I (1oga) [22] and TR/pMHC-II (1j8h) [23]. The figure was generated with Pymol, http://pymol.sourceforge.net. (B) Schematic representation of TR/pMHC-I and TR/pMHC-II. The TR (TR-ALPHA and TR-BETA chains), the MHC-I (I-ALPHA and beta-2-microglobulin B2M chains) and the MHC-II (II-ALPHA and II-BETA chains) are shown with the extracellular domains (V-ALPHA and C-ALPHA for the TR-ALPHA chain; V-BETA and C-BETA for the TR-BETA chain; G-ALPHA1, G-ALPHA2 and C-LIKE for the I-ALPHA chain; C-LIKE for B2M; G-ALPHA and C-LIKE for the II-ALPHA chain; II-BETA and C-LIKE for the II-BETA chain), and the connecting, transmembrane and cytoplasmic regions. Arrows indicate the peptide localization in the G-DOMAIN groove. The MHC G-DOMAINs and TR V-DOMAINs are likely to be in a diagonal rather than in a vertical position relative to the cell surface [24, 25].

The TR V-DOMAINs and MHC G-DOMAINs that are directly involved in the TR/pMHC interactions are described in the next sections.

TR V-DOMAINs

The V-DOMAINs have an immunoglobulin fold, that is an antiparallel beta sheet sandwich structure with 9 strands [14,21], the A, B, E and D strands being on one sheet, and the G, F, C, C' and C" strands on the other sheet. These strands are indicated in the IMGT Colliers de Perles (Fig. 2) which are IMGT 2D graphical representations based on the IMGT unique numbering for V-DOMAINs [14]. IMGT Colliers de Perles of the V-ALPHA and V-BETA domains from 1ao7 [26] are shown as examples in Fig. 2.

The V-ALPHA and V-BETA domains share main conserved characteristics of the V-DOMAIN which are the disulfide bridge between cysteine 23 (1st-CYS) and cysteine 104 (2nd-CYS), and the other hydrophobic core residues tryptophan 41 (CONSERVED-TRP) and leucine (or hydrophobic) 89 [14] (Fig. 2). The A strand comprises positions 1 to 15, B strand positions 16 to 26, C strand positions 39 to 46, C' strand positions 47 to 55, C" strand positions 66 to 74, D strand positions 75 to 84, E strand positions 85 to 96, F strand positions 97 to 104, and G strand positions 118 to 128 [14]. Compared to the general V-DOMAIN 3D structure, the V-ALPHA domains have shorter C" and D strands at the C'D turn (with 7 gaps at positions 71 to 77) and, in contrast, longer D and E strands at the DE turn (with additional positions at 84A, 84B and 84C).

The three hypervariable loops or complementarity determining regions (CDR-IMGT) of each V-DOMAIN are involved in the pMHC recognition. The CDR1-IMGT comprises positions 27 to 38, the CDR2-IMGT positions 56 to 65 and the CDR3-IMGT positions 105 to 117 [14]. The CDR3-IMGT corresponds to the junction resulting from the V-J and V-D-J rearrangement, and is more variable in sequence and length than the CDR1-IMGT and CDR2-IMGT that are encoded by the V-REGION only [1]. Lengths of the CDR1-IMGT are shown separated by dots between brackets [14]. For examples, 1ao7 [6.5.11] V-ALPHA means that in the V-ALPHA domain of 1ao7, CDR1-IMGT has a length of 6 amino acids, CDR2-IMGT a length of 5 amino acids and CDR3-IMGT a length of 11 amino acids, and 1ao7 [5.6.14] V-BETA means that in the V-BETA domains of 1ao7, CDR1-IMGT, CDR2-IMGT and CDR3-IMGT have a length of 5, 6 and 14 amino acids, respectively [14].

pMHC CONTACT ANALYSIS

Owing to its standardization, the IMGT unique numbering for G-DOMAIN [16] has allowed to graphically represent, in the IMGT Colliers de Perles for G-DOMAIN, the MHC amino acid positions that have contacts with the peptide side chains. Eleven IMGT pMHC contact sites were defined (C1 to C11) which can be used to compare pMHC interactions. Examples of contact sites for a MHC-I binding a 9-amino acid peptide (1ao7), for a MHC-I binding a 8-amino acid peptide (1jtr) and for a MHC-II binding 9 amino acids of the peptide in the groove (1j8h) are shown in Figs 4, 5 and 6, respectively.

In contrast to previous attempts to define pockets [28], structural data for defining the IMGT pMHC contact sites take into account the length of the peptides and are considered independently of the MHC class and sequence polymorphisms. The interactions between the peptide amino acid side chains and MHC amino acids were computed using an interaction scoring scheme based on true mean energy ratio. The score assigned to each contact is a constant value, independent on the distance between atoms (hydrogen bond 40, water mediated hydrogen bond 20, contact between polar atoms 20, contact between

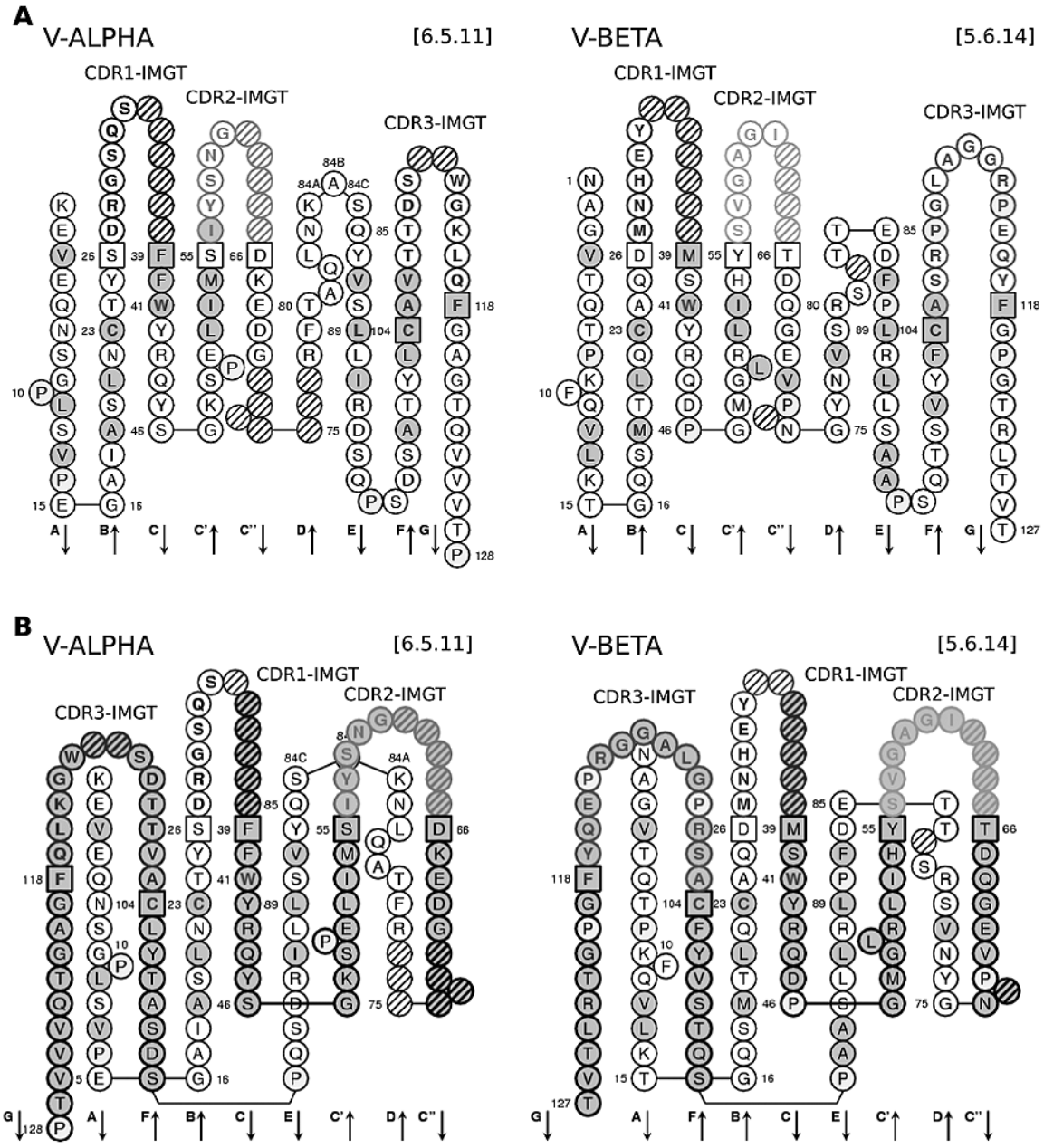

Fig. 2. IMGT Collier de Perles of the V-ALPHA and V-BETA domains from 1ao7 [26] (IMGT/3Dstructure-DB [5], http://imgt.cines.fr) (A) on one layer (B) on two layers. Amino acids are shown in the one-letter abbreviation. Hydrophobic amino acids (hydropathy index with positive value) and tryptophan (W) found at a given position in more than 50% of analyzed IG and TR sequences are shown. The CDR-IMGT are limited by amino acids shown in squares, which belong to the neighbouring FR-IMGT and represent anchor positions. The CDR3-IMGT extend from position 105 to 117 [14]. Hatched circles correspond to missing positions according to the IMGT unique numbering. Arrows indicate the direction of the beta sheets and their different designations in 3D structures.

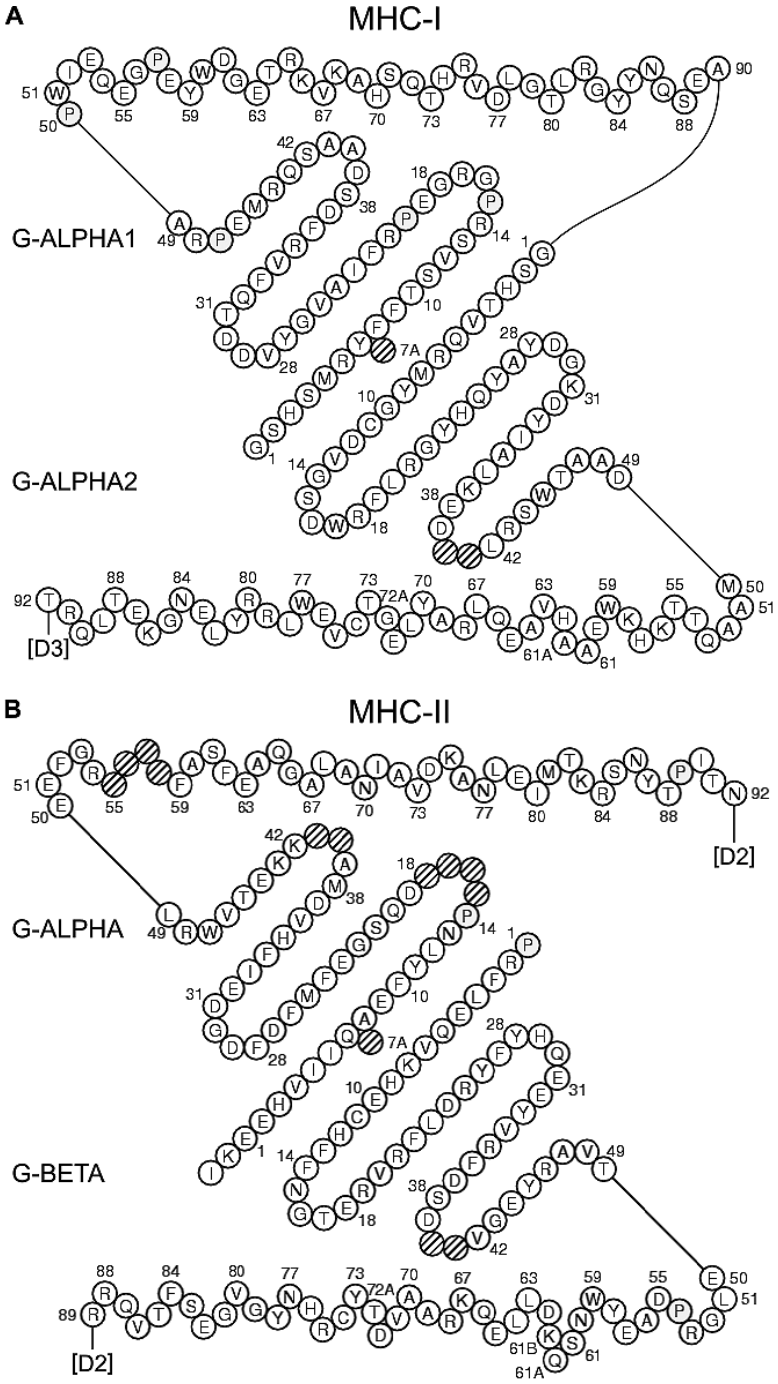

Fig. 3. IMGT Collier de Perles of MHC G-DOMAINs. (A) MHC-I G-ALPHA1 and G-ALPHA2 domains (B) MHC-II G-ALPHA and G-BETA domains. MHC-I G-DOMAINs are from 1ao7 [26] and MHC-II G-DOMAINs are from 1j8h [23] (IMGT/3Dstructure-DB [5], http://imgt.cines.fr). Amino acid positions are according to the IMGT unique numbering for G-DOMAIN [16]. Positions 61A, 61B and 72A are characteristic of the G-ALPHA2 and G-BETA domains (and are not reported in the G-ALPHA1 and G-ALPHA IMGT Collier de Perles).

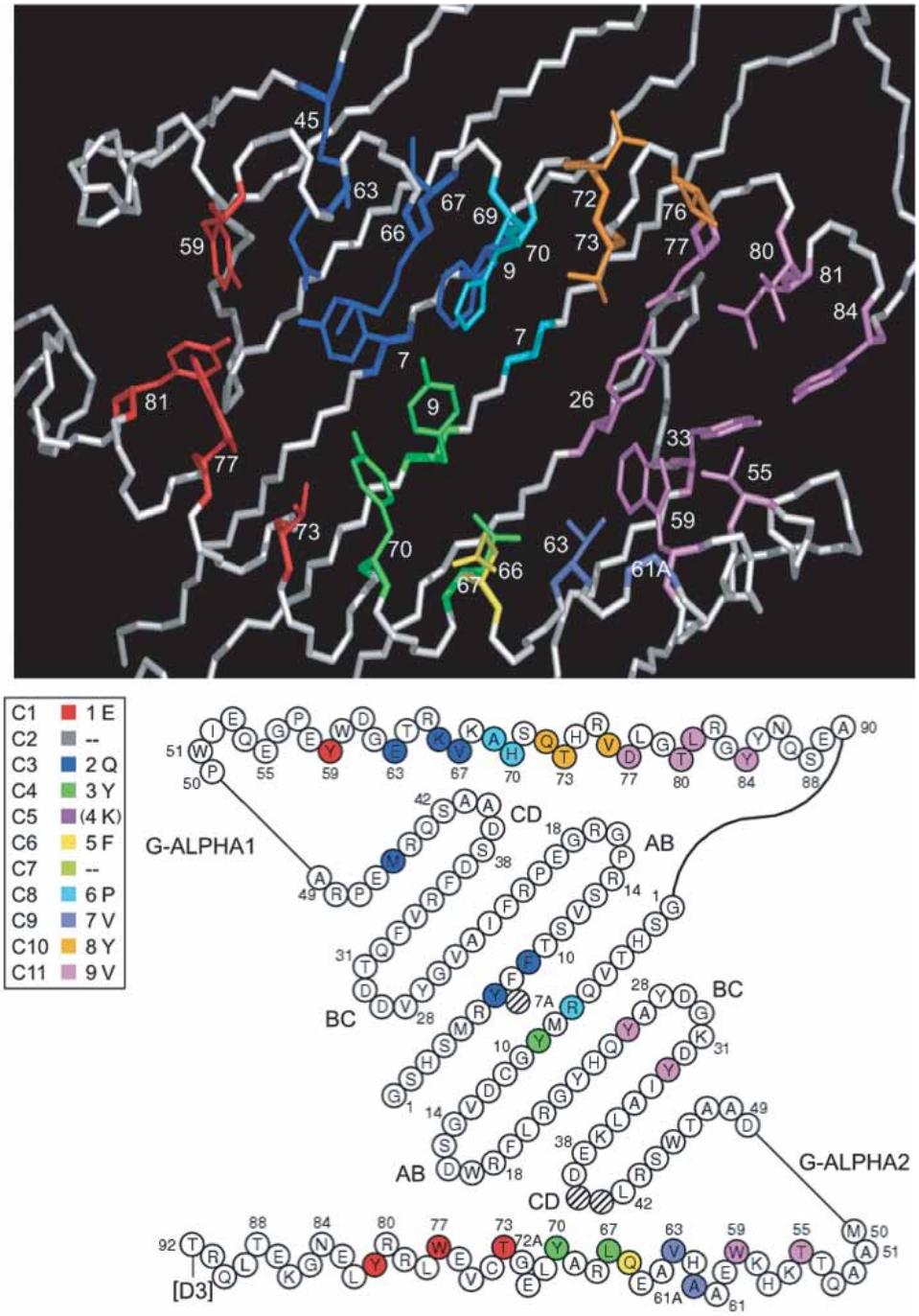

Fig. 4. IMGT pMHC contact sites of human HLA-A*0201 MHC-I and a 9-amino acid peptide side chains (1ao7) [26]. *Upper section:* 3D structure of the human HLA-A*0201 groove. *Lower section:* IMGT pMHC contact sites IMGT Collier de Perles. Both views are from above the cleft, with G-ALPHA1 on top and G-ALPHA2 on bottom. In the box, C1 to C11 refer to contact sites. 1 to 9 refer to the numbering of the peptide amino acids P1 to P9. There are no C2 and C7 in MHC-I 3D structures with 9-amino acid peptides. There is no C5 in this 3D structure as P4 does not contact MHC amino acids (4G is shown between parentheses in the box).

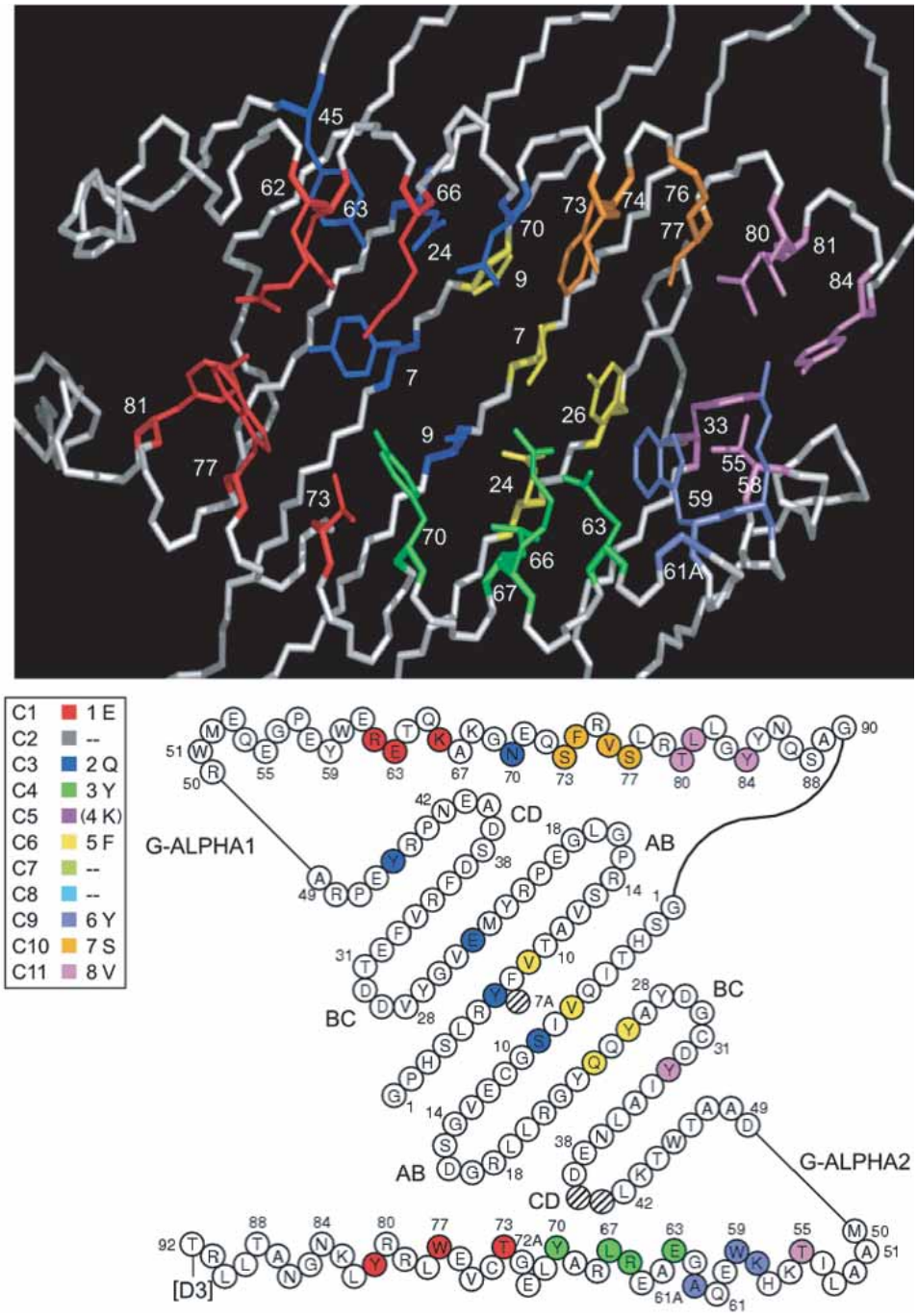

Fig. 5. IMGT pMHC contact sites of mouse H2-K1*01 MHC-I and a 8-amino acid peptide side chains (1jtr) [27]. Upper section: 3D structure of the mouse H2-K1*01 groove. Lower section: IMGT pMHC contact sites IMGT Collier de Perles. Both views are from above the cleft with G-ALPHA1 on top and G-ALPHA2 on bottom. In the box, C1 to C11 refer to contact sites, 1 to 8 refer to the numbering of the peptide amino acids P1 to P8. There are no C2, C7 and C8 in MHC-I 3D structures with 8-amino acid peptides. There is no C5 in this 3D structure as P4 does not contact MHC amino acids (4K is shown between parentheses in the box).

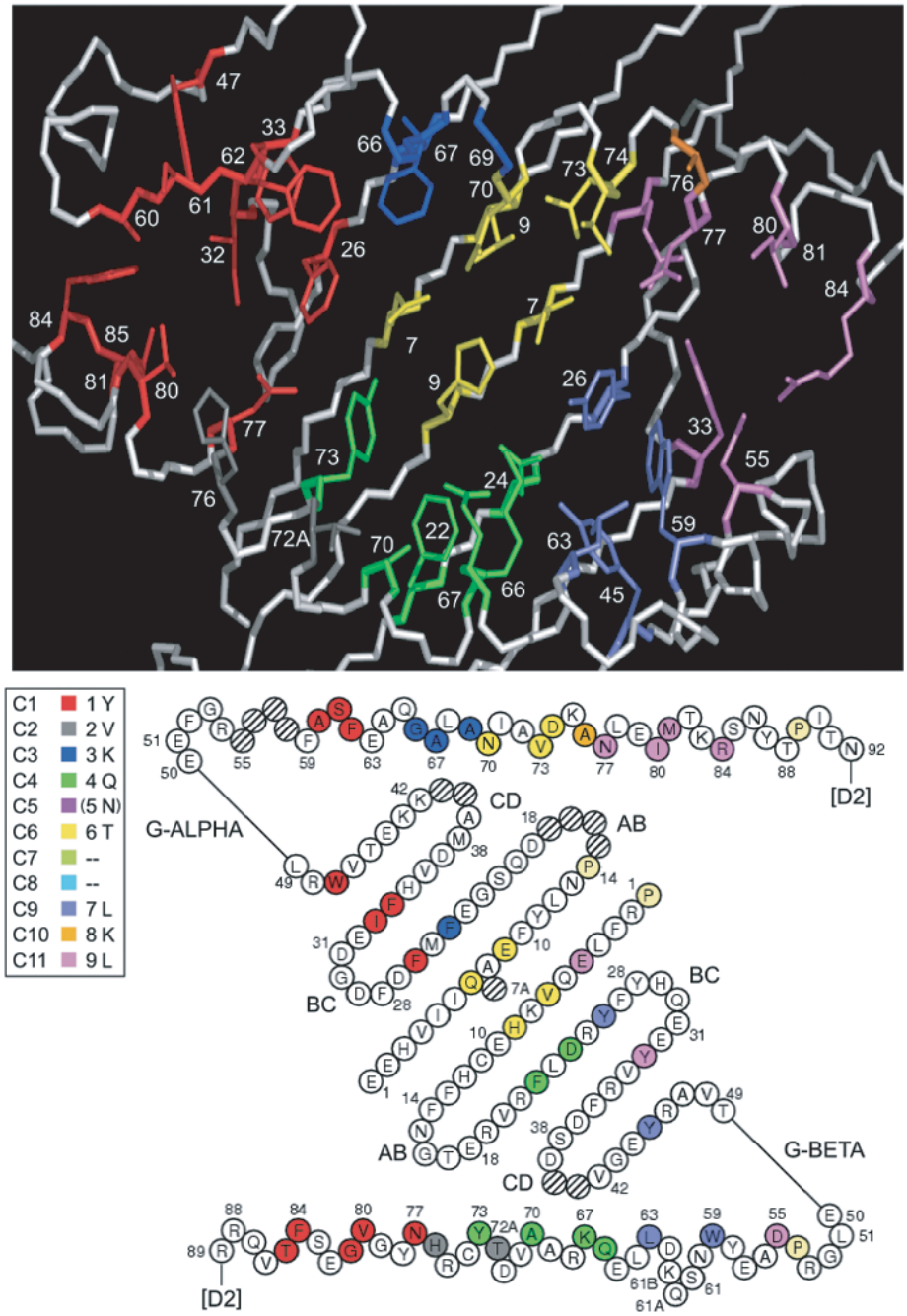

Fig. 6. IMGT pMHC contact sites of human HLA-DRA*0101 and HLA-DRB1*0401 MHC-II and the peptide side chains (9 amino acids located in the groove) (1j8h) [23]. Upper section: 3D structure of the human HLA-DRA*0101 and HLA-DRB1*0401 groove. Lower section: IMGT pMHC contact sites IMGT Collier de Perles. Both views are from above the cleft with G-ALPHA on top and G-BETA on bottom. In the box, C1 to C11 refer to contact sites. 1 to 9 refer to the numbering of the peptide amino acids 1 to 9 located in the groove. There are no C7 and C8 in MHC-II 3D structures with peptide of 9 amino acids located in the groove. There is no C5 in this 3D structure as 5 does not contact MHC amino acids (5N is shown between parentheses in the box).

Table 3
IMGT reference pMHC contact sites. (A) MHC-I, (B) MHC-II

(**A**) **MHC-I**			
		G-ALPHA1	G-ALPHA2
8-amino acid peptides			
C1	1	59 62 63 66	73 77 81
C3	2	7 24 45	9
C4	3		9 24 63 66 67 70
C5	4		
C6	5	7 9 22 70 74	7 9 24 26
C9	6		59 61A 63 66
C10	7	77 73 76	
C11	8	77 80 81 84	5 26 33 34 55 59
9-amino acid peptides			
C1	1	5 59 62 63 66	73 77 81
C3	2	7 9 22 24 34 45 63 66 67 70	
C4	3		7 9 24 66 67 70
C5	4	65 66	66
C6	5	70 73 74	7 26 66 67
C8	6	66 69 70 73 74	7 24 62 66
C9	7		7 24 59 61A 63 66
C10	8	72 73 76 80	58
C11	9	77 80 81 84	5 26 33 34 55 59
(**B**) **MHC-II**			
		G-ALPHA	G-BETA
C1	1	26 33 34 47 60 61 62	77 80 81 82 84 85
C2	2		72A 73 76
C3	3	7 24 62 63 66 67 69	
C4	4	7	9 11 22 24 66 67 70 73 74
C5	5		66
C6	6	9 69 70 73 74	7 26
C9	7		24 26 45 59 63 66
C10	8	73 76	
C11	9	77 80 81 84	5 33 55

(A) IMGT reference pMHC contact sites result from one hundred and four pMHC-I 3D structures (30 with 8-amino acid peptides and 74 with 9-amino acid peptides). (B) IMGT reference pMHC contact sites result from forty-four pMHC-II 3D structures with 9 amino acids in the groove.

non polar atoms 1). All direct contacts (defined with a cut off equal to the sum of the atom van der Waals radii and of the diameter of a water molecule) and water mediated hydrogen bonds were taken into account for the definition of the IMGT pMHC contact sites. The analysis was carried out for the pMHC available in IMGT/3Dstructure-DB [5], http://imgt.cines.fr. One hundred fourteen 3D structures with peptides of 8, 9 and 10 amino acids bound to MHC-I and forty-four 3D structures of pMHC-II were identified. The contact analysis was performed for the peptide amino acid side chains of the 9 amino acids located in the groove. Results for MHC-I with 8-amino acid peptides (30 pMHC-I 3D structures), MHC-I with 9-amino acid peptides (74 pMHC-I 3D structures), and MHC-II for the 9 amino acids located in the groove (44 pMHC-II 3D structures) are reported in Table 3 (the results for the ten pMHC-I with 10-amino acid peptides are not shown). These "IMGT reference pMHC contact sites" are also available as IMGT Colliers de Perles. They will be updated as the number of 3D structures increases.

IMGT Colliers de Perles for IMGT pMHC contact sites are provided for each individual pMHC and TR/pMHC entry in IMGT/3Dstructure-DB. They allow to easily identify the amino acid contacts between

the MHC and the peptide amino acid side chains and to compare them with the "IMGT reference pMHC contact sites".

C1 to C11 refers to the eleven contact sites. 1 to 9 refers to the numbering of the peptide amino acids in the groove. The peptide binding mode to MHC-I is characterized by the N and C peptide ends docked deeply with C1 and C11 contact sites that correspond to the two conserved pockets A and F, and by the peptide length that mechanically constrains the peptide conformation in the groove. There are no C2, C7 and C8 contact sites for MHC-I with 8-amino acid peptides, no C2 and C7 for MHC-I with 9-amino acid peptides. In contrast, for MHC-II, C2 is present but there is no C7 and C8. Whereas C1 and C11 correspond to the conserved pockets A and F, respectively, the correspondence between the other contact sites and the previously defined pockets is more approximative. For MHC-I with a peptide of 8-amino acids C3, C4, C6 and C9 correspond roughly to the B, D, C and E pockets, and for MHC-I with a peptide of 9-amino acids C3, C4 and C9 correspond to the B, D and E pockets.

TR/pMHC 3D STRUCTURES

Eighteen TR/pMHC 3D structures are available in IMGT/3Dstructure-DB [5] (Table 4). Fourteen 3D structures (twelve TR/pMHC-I and two TR/pMHC-II) comprise the complete extracellular region of the alpha-beta TR (TR-ALPHA_BETA) whereas four 3D structures comprise a Fv variable fragment (FV-ALPHA_BETA).

The IMGT Protein display (Fig. 7) shows the amino acid sequences of the different V-ALPHA and V-BETA domains found in the crystallized TR/pMHC. Lengths of the V-DOMAIN CDR-IMGT from available TR/pMHC 3D structures are reported in Table 4, together with the names of the V, D and J genes [1]. For examples, the 1ao7 V-ALPHA [6.5.11] results from the TRAV12-2-TRAJ24 rearrangement, and the 1ao7 V-BETA [5.6.14] results from the TRBV6-5-TRBD2-TRBJ2-7 rearrangement. The amino acid sequences of the different G-DOMAINs found in the crystallized TR/pMHC are shown in the IMGT Protein display (Fig. 8).

TR/MHC AND TR/PEPTIDE INTERFACES

The analysis of the pairwise contacts that occur at the TR/MHC and TR/peptide interfaces was carried out using the IMGT unique numbering for V-DOMAINs [14] for the TR, and the IMGT unique numbering for G-DOMAINs for the MHC. Table 5 shows the interactions of the TR V-ALPHA and TR V-BETA with MHC-I and the peptide, in nine TR/pMHC-I 3D structures. Table 6 shows the interactions of the TR V-ALPHA and TR V-BETA with MHC-II and the peptide, in two TR/pMHC-II 3D structures. These tables provide for the first time the contacts using the IMGT unique numbering for V-DOMAIN [14] and the IMGT unique numbering for G-DOMAIN [16], allowing to compare data whatever the gene group (TRAV, TRBV), the MHC class (MHC-I, MHC-II), and whatever the species (*Homo sapiens, Mus musculus*).

The results show that positions implicated in the binding are well conserved but not the pairwise interactions. The MHC contact positions belong to the G-DOMAIN helices. The TR positions that are involved in the contacts belong mostly to the CDR-IMGT and to anchor positions (shown by squares in Fig. 2). The FR-IMGT positions involved in the contacts are positions 84 and 84A that are located at the DE turn (designated as "hypervariable 4" or HV4).

Table 4
T cell receptor/peptide/MHC (TR/pMHC) complexes in IMGT/3Dstructure-DB [5], http://imgt.cines.fr

(A)TR/pMHC-I										
		T cellreceptor				**Peptide**		**MHC-II**		
Code	Ref	Name	Sp	V-DOMAIN genes	CDR-IMGT	Sequence	Length	Sp	Gene and allele	R(Å)
1ao7	26	A6	Hs	TRAV12-2-TRAJ24	[6.5.11]	LLFGYPVYV	9	Hs	HLA-A*0201	2.6
			Hs	TRBV6-5-TRBD2-TRBJ2-7	[5.6.14]					
1qrn	29	A6				LLFGYAVYV	9			2.80
1qse	29	A6				LLFGYPRYV	9			2.80
1qsf	29	A6				LLFGYPVAV	9			2.80
1bd2	30	B7	Hs	TRAV29/DV5-TRAJ54	[6.6.10]	LLFGYPVYV	9	Hs	HLA-A*0201	2.5
			Hs	TRBV6-5-(TRBD2)-TRBJ2-7	[5.6.13]					
1oga	22	JM22	Hs	TRAV27-TRAJ42	[5.6.10]	GILGFVFTL	9	Hs	HLA-A*0201	1.4
			Hs	TRBV19-(TRBD2)-TRBJ2-7	[5.6.11]					
1mi5	4	LC13	Hs	TRAV26-2-TRAJ52	[7.4.14]	FLRGRAYGL	9	Hs	HLA-B*0801	2.50
			Hs	TRBV7-8-(TRBD2)-TRBJ2-7	[5.6.11]					
1lp9	31	12.2	Mm	TRAV12D-2-TRAJ50	[6.6.13]	ALWGFFPVL	9	Hs	HLA-A*0201	2.00
			Mm	TRBV13-3-(TRBD2)-TRBJ2-7	[5.6.11]					
1g6r	32	2C	Mm	TRAV9-4-TRAJ35	[6.7.10]	SIYRYYGL	8	Mm	H2-K1*01	2.80
			Mm	TRBV13-2-(TRBD2)-TRBJ2-4	[5.6.9]					
1jtr	27	2C				EQYKFYSV	8			2.40
2ckb	33	2C				EQYKFYSV	8			3.2
1mwa	27	2C				EQYKFYSV	8			2.40
1fo0	34	BM3.3	Mm	TRAV16-TRAJ32	[7.7.14]	INFDFNTI	8	Mm	H2-K1*01	2.50
			Mm	TRBV1-TRBD1-TRBJ1-3	[6.6.12]					
1nam	35	BM3.3				RGYVYQGL	8			2.70

Table 4, continued

1kj2	36	KB5-C20	Mm	TRAV14-1-TRAJ15	[6.6.11]	KVITFIDL	8	Mm	H2-K1*01	2.71
			Mm	TRBV1-TRBD2-TRBJ2-3	[6.6.16]					
(B)TR/pMHC-II										
		T cellreceptor				**Peptide**		**MHC-II**		
Code	Ref	Name	Sp	V-DOMAIN genes	CDR-IMGT	Sequence	Length	Sp	Gene and allele	R(Å)
1fyt	37	HA1.7	Hs	TRAV8-4-TRAJ48	[6.7.13]	PKYVKQNTLKLAT	13	Hs	HLA-DRA*0101	2.60
			Hs	TRBV28-TRBD1-TRBJ1-2	[5.6.12]			Hs	HLA-DRB1*0101	
1j8h	23	HA1.7				PKYVKQNTLKLAT	13	Hs	HLA-DRA*0101	2.40
								Hs	HLA-DRB1*0401	
1d9k	38	D10	Mm	TRAV14D-2-TRAJ4	[6.6.10]	GNSHRGAIEWEGIESG	16	Mm	H2-AA*02	3.2
			Mm	TRBV13-2-TRBD2-TRBJ2-1	[5.6.11]			Mm	H2-AB*02	

Sp: species, Hs: *Homo sapiens*, Mm: *Mus musculus*, R(Å): Crystallographic resolution in angstrom. Fourteen 3D structures (12 TR/pMHC-I and 2 TR/pMHC-II) correspond to complete extracellular region of TR receptors (TR-ALPHA_BETA). Four 3D structures (1d9k, 1fo0, 1kj2 and 1nam) correspond to an Fv variable fragment (FV-ALPHA_BETA). Gene and allele names are according to IMGT/GENE-DB [18] for human and mouse TR, to IMGT/HLA-DB [17] for human MHC, and to MGD [13] and IMGT for mouse MHC. Amino acid sequences of the TR V-DOMAINs and MHC G-DOMAINs are reported in Figure 7 and Figure 8, respectively. H2-K1*01 encodes H2-K1b, H2-AB*02 and H2-AA*02 encode I-Abk and I-Aak, respectively. Between brackets, lengths of the CDR-IMGT are according to Lefranc et al. 2005 [14].

The contact analysis confirms that the V-ALPHA CDR2-IMGT seats on top of the G-ALPHA2 (MHC-I) or G-BETA (MHC-II) helices, and that the V-BETA CDR2-IMGT seats on top of the G-ALPHA1 (MHC-I) or G-ALPHA (MHC-II) helices (Tables 5 and 6). This agrees with data from Sim et al. 1996 [39] who showed that most of the TR/MHC specificity comes from the CDR1 and CDR2 because mutations in these CDRs are able to change specificity between MHC-I and MHC-II. V-ALPHA and V-BETA CDR3-IMGT usually follow the same G-DOMAIN contact preference as the CDR2-IMGT but they can also have contacts with the other G-DOMAINs. For example, in the 1oga 3D structure [22], position 66 of G-ALPHA2 is contacted by the V-ALPHA CDR3-IMGT but also by the V-BETA CDR3-IMGT.

The diagonal orientation of the TR/pMHC complex puts the TR in a globally conserved position to "read-out" the peptide [31]. V-ALPHA is on top of the peptide N terminus while V-BETA is on top of the peptide C terminus. TR positions implicated in the peptide recognition are in the CDR3-IMGT and generally to a lesser extent in the V-ALPHA CDR1-IMGT (Tables 5 and 6). Nearly every 3D structure shows different CDR3 conformations and binding mode. In the JM22/peptide/HLA-A complex (1oga) [22], the V-BETA CDR3-IMGT extensively contacts the peptide and G-ALPHA2 through hydrogen bonds (Table 5), by inserting itself between the peptide and the G-ALPHA2. In contrast, the 2C/peptide/H2-K1 complex (1jtr) [27] has comparatively fewer contacts between the V-BETA CDR3-IMGT and the peptide and the MHC, however the V-BETA CDR1-IMGT has more contacts

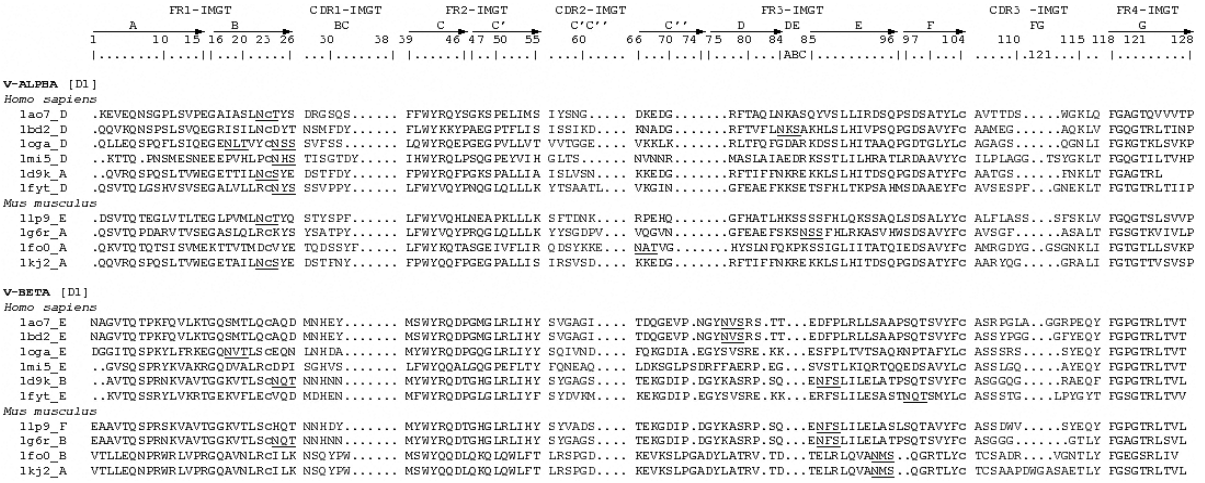

Fig. 7. IMGT Protein display of the TR V-ALPHA and V-BETA domains found in the TR/pMHC complexes in IMGT/3Dstructure-DB [5], http://imgt.cines.fr. Amino acid sequences and gaps (shown by dots) are according to the IMGT unique numbering for V-DOMAIN [14]. The three additional positions in the CDR3-IMGT are 111.1, 112.2 and 112.1. Potential N-glycosylation sites are underlined. Assignments of the V, D and J genes are shown in Table 1.

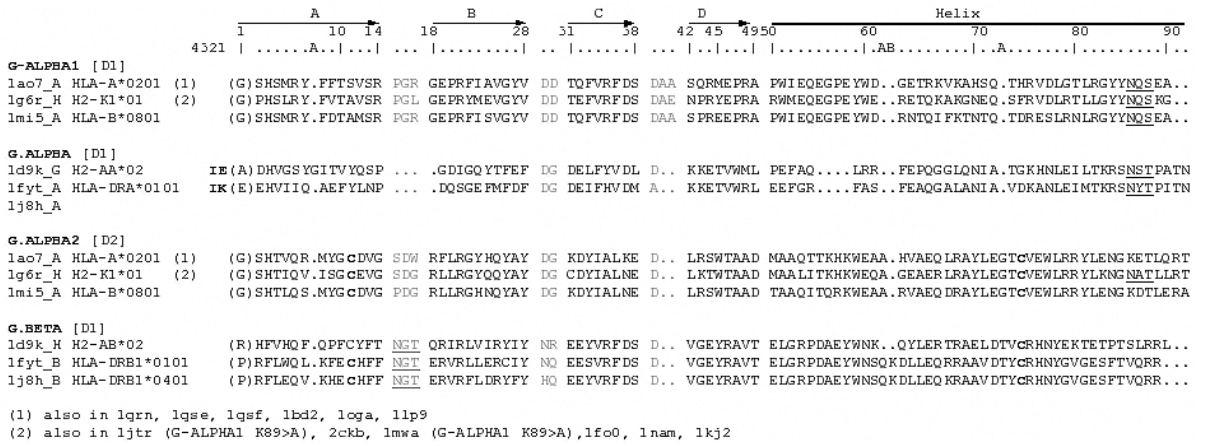

Fig. 8. Protein display of the G-DOMAINs found in the TR/pMHC complexes in IMGT/3Dstructure-DB [5], http://imgt.cines.fr. Amino acid sequences and gaps (shown by dots) are according to the IMGT unique numbering for G-DOMAIN. Amino acid resulting from the splicing with the preceding exon are shown within parentheses. Potential N-glycosylation sites are underlined. Positions 61A, 61B and 72A are characteristic of the G-ALPHA2 and G-BETA domains. The corresponding gaps in G-ALPHA1 and G-ALPHA, shown in this IMGT Protein display, are not reported in the IMGT Colliers de Perles as these gaps are shared by those two domains. H2-K1*01 encodes H2-K1b, H2-AB*02 and H2-AA*02 encode I-Abk and I-Aak, respectively.

and hydrogen bonds with the peptide and G-ALPHA2.

The TR LC13 and 2C were crystallized both alone and in complex with a pMHC. The structural superimposition of both V-DOMAIN scaffold alpha carbons reveals large movements of the CDR3 and of the CDR1, respectively. The V-ALPHA domains of LC13, in the 1mi5 and 1kgc 3D structures, have 3.5Å root mean square (RMS) between their CDR3. The V-ALPHA domains of 2C, in the 2ckb and 1tcr 3D structures, have 2.3Å RMS between their CDR1. The TR A6 was crystallized in complex with the same MHC but with different peptides. In these structures, the V-BETA CDR3 adopt different

Table 5
TR V-ALPHA and V-BETA CDR interactions with pMHC-I. (A) V-ALPHA CDR-IMGT interactions, (B) V-BETA CDR-IMGT interactions, (C) V-ALPHA and V-BETA FR-IMGT interactions

(A) V-ALPHA CDR-IMGT interactions

V-ALPHA CDR1-IMGT

	CDR1	G-ALPHA1	Peptide	G-ALPHA2
1ao7 [6.]	27_D	**58_E**		
	28_R	58_E		77_W 80_R
	29_G		1_L	77_W
	31_Q	66_K	1_L **2_L** 3_F 4_G 5Y	70_Y 73_T
	32_S		**5_Y**	
1bd2 [6.]	28_S		1_L	76_E **77_W**
	29_M	58_E 59_Y 62_G 63_E 66_K	1_L	77_W
	31_D	66_K	4_G **5_Y**	66_Q 73_T
	32_Y		5_Y	66_Q
1oga [5.]	30_S			65_E 66_Q
1mi5 [7.]	29_S	62_R		
	30_G			69_A
	31_T		4_G	**66_Q** 70_Y 73_T
	33_Y		7_Y	$61A_A$ 62_R 63_V 64_A 65_E 66_Q
1lp9 [6.]	28T			76_E
	29_Y			69_A $72A_G$ 73_T 76_E 77_W
	30_S			69_A
	32_F			65_E 66_Q 69_A
1g6r [6.]	27_Y	62_R		
	28_S	58_E **62_R**		
	29_A	62_R		
	30_T			76_E
	32_Y		3_Y 4_R	66_R
1jtr [6.]	27_Y	62_R		
	28_S	58_E **62_R**		
	29_A	62_R	1_E	77_W
	30_T			76_E
	32_Y		3_Y 4_K	66_R
1fo0 [7.]	28_Q	58_E 62_R		
	29_D	62_R		
	30_S			**73_T**

(B) V-BETA CDR-IMGT interactions

V-BETA CDR1-IMGT

	CDR1	G-ALPHA1	Peptide	G-ALPHA2
1ao7 [5.]	30_E		**8_Y**	
1oga [5.]	30_D		**8_T**	58_K
1mi5 [5.]	30_V	76_E 80_N		
	31_S	76_E		
1lp9 [5.]	30_D	**72_Q** 76_V		
	31_Y	69_A 73_T	6_F	
1g6r [5.]	28_N		6_Y	58_K
	29_H		6_Y	61_Q $61A_A$
	30_N		**6_Y** 7_G 8_L	58_K
	31_N		6_Y	
1jtr [5.]	27_N			61_Q
	28_N		6_Y	58_K **61_Q**
	29_H		6_Y	61_Q $61A_A$
	30_N		**6_Y** 7_S 8_V	58_K 59_W
	31_N		6_Y	
1fo0 [6.]	29_Q			61_Q
	32_W	76_V	7_T	
1kj2 [6.]	29_Q			58_K 59_W 61_Q $61A_A$
	30_Y			$61A_A$
	31_P		7_D	
	32_W	69_G 72_Q		

V-BETA CDR2-IMGT

	CDR2	G-ALPHA1	Peptide	G-ALPHA2
1bd2 [.6.]	61_I	72_Q		
1oga [.6.]	57_Q	69_A	**4_G** 5_F **6_V**	
	58_I	69_A 72_Q 73_T 76_V	6_V 8_T	
	59_V	72_Q 76_V		
	60_N	72_Q 75_R		
	61_D	65_R 68_K 69_A 72_Q		
1mi5	57_Q	72_Q 75_R 76_E		

Table 5, continued

	31_{S}			69_{A}
	33_{F}			66_{R}
1kj2 [6.]	27_{D}	58_{E} **62_{R}**		
	29_{T}	62_{R}	1_{K}	77_{W}
	31_{N}			73_{T}
V-ALPHA CDR2-IMGT				
	CDR2	**G-ALPHA1**	**Peptide**	**G-ALPHA2**
1ao7 [.5.]	57_{Y}			65_{E} 66_{Q} 69_{A}
	58_{S}			69_{A}
	59_{N}			76_{E}
1bd2 [.6.]	57_{S}			65_{E} 66_{Q} 69_{A}
	58_{S}			69_{A} 70_{Y} 73_{T}
	59_{I}			68_{R} 69_{A} 72_{E} $72A_{G}$
1oga [.6.]	57_{V}			62_{H} 65_{E}
1mi5 [.4.]	56_{G}			62_{R}
	57_{L}			65_{E} 66_{Q} 69_{A}
	58_{T}			65_{E}
	59_{S}			65_{E}
1lp9 [.6.]	57_{F}			$61A_{A}$ 62_{H} 65_{E} 66_{Q}
	58_{T}			62_{H} 65_{E}
	61_{K}			65_{E}
1g6r [.7.]	57_{Y}			66_{R} 69_{A}
	58_{S}			69_{A} $72A_{G}$ 73_{T} 76_{E}
1jtr [.7.]	57_{Y}			65_{E} 66_{R} 69_{A} 73_{T}
	58_{S}			**69_{A}** 70_{Y} $72A_{G}$ 73_{T}
1fo0 [.7.]	59_{Y}			62_{G} 65_{E} 66_{R} 69_{A}
	60_{K}			65_{E}
1kj2 [.6.]	57_{R}			69_{A} 72_{E}
	58_{S}			76_{E}
	59_{V}			72_{E} $72A_{G}$ 76_{E}
V-ALPHA CDR3-IMGT				
	CDR3	**G-ALPHA1**	**Peptide**	**G-ALPHA2**
1ao7 [.11]	108_{T}	**65_{R}** 66_{K}	4_{G} 5_{Y}	
	109_{D}	62_{G} **65_{R}** 66_{K}	4_{G} 5_{Y}	
	110_{S}		**4_{G}** 5_{Y} 6_{P}	

[.6.]	58_{N}	**79_{R}**		
	59_{E}	**79_{R}**		
1lp9 [.6.]	57_{Y}	65_{R} 68_{K} 69_{A} 72_{Q}		
	58_{V}	72_{Q}		
	61_{S}	68_{K}		
1g6r [.6.]	57_{Y}	69_{G} 70_{N} 72_{Q} 73_{S} 76_{V}		
	58_{G}	76_{V}		
	59_{A}	76_{V} 79_{R}		
	60_{G}	79_{R}		
	61_{S}	76_{V}		
1jtr [.6.]	57_{Y}	69_{G} 72_{Q} 73_{S} 76_{V}	7_{S}	
	58_{G}	76_{V}		
	59_{A}	79_{R}		
	60_{G}	79_{R}		
	61_{S}	76_{V} 79_{R}		
1fo0 [.6.]	57_{R}	76_{V} 79_{R} 80_{T}		
	58_{S}	76_{V} **79_{R}**		
	59P	79_{R}		
1kj2 [.6.]	57_{R}	72_{Q} 73_{S} 76_{V}	**7_{D}**	
	58_{S}	72_{Q}		
	61_{D}	72_{Q}		
V-BETA CDR3-IMGT				
	CDR3	**G-ALPHA1**	**Peptide**	**G-ALPHA2**
1ao7 [.14]	107_{R}		5_{Y}	
	109_{G}		6_{P}	
	110_{L}	69_{A} 72_{Q} 73_{T}	6_{P} 7_{V} **8_{Y}**	
	111_{A}		7_{V} 8_{Y}	$61A_{A}$
	112_{G}		5_{Y} 7_{V}	**$61A_{A}$** 62_{H} 63_{V} **66_{Q}**
	112.1_{G}		5_{Y} 7_{V}	$61A_{A}$
	113_{R}		5_{Y}	**61_{A}** $61A_{A}$ 62_{H} 66_{Q}
	114_{P}		5_{Y}	66_{Q}
1bd2 [.13]	108_{Y}		8_{Y}	
	109_{P}		6_{P} 7_{V}	
	110_{G}		6_{P} 7_{V} **8_{Y}**	
	111_{G}		7_{V} 8_{Y}	$61A_{A}$
	112_{G}		7_{V}	$61A_{A}$

Table 5, continued

	113_{W}	65_{R} 68_{K} 69_{A} 72_{Q}		
	114_{G}	65_{R}		
1bd2 [.10]	107_{M}		5_{Y}	
	108_{E}	58_{E} 62_{G} 65_{R} 66_{K}		
	109_{G}	65_{R} 66_{K}	4_{G} 5_{Y}	
	113_{A}		4_{G} 5_{Y}	
	114_{Q}	65_{R} 69_{A}		
	115_{K}	65_{R}		
1oga [.10]	107_{A}			66_{Q}
	108_{G}		5_{F}	66_{Q}
	109_{S}		4_{G} 5_{F}	66_{Q}
	113_{Q}	66_{K}	4_{G} 5_{F}	
1mi5 [.14]	108_{L}		6_{A} 7_{Y}	66_{Q}
	109_{A}	62_{R}		
	110_{G}	62_{R} 66_{I}		
	111_{G}	65_{Q} 66_{I} **69_{T}**	4_{G}	
	112_{S}	69_{T}	6_{A}	
	112.1_{T}	62_{R} 65_{Q} 66_{I} 69_{T}		
	113_{Y}	69_{T} 72_{Q}	6_{A}	
1lp9 [.13]	107_{F}		5_{F}	66_{Q}
	109_{A}		3_{W} 4_{G} 5_{F}	66_{Q}
	110_{S}		2_{L} 3_{W} 4_{G}	66_{Q} 69_{A} 70_{Y} 73_{T}
	111_{S}	63_{E} **66_{K}**	2_{L} 4_{G}	73_{T} 77_{W}
	112_{S}	66_{K}	4_{G} 5_{F}	
	113_{F}	65_{R} 66_{K} 69_{A}	4_{G} 6_{F}	
	114_{S}		**4_{G}** 5_{F} 6_{F}	
1g6r [.10]	107_{S}		**4_{R}**	
	108_{G}		**4_{R}**	
	109_{F}	62_{R} 65_{Q} 66_{K}	4_{R}	
	113_{A}		4_{R}	
	114_{S}		4_{R}	
1jtr [.10]	107_{S}		4_{K}	
	108_{G}	66_{K}	4_{K}	
	109_{F}	61_{E} 62_{R} 65_{Q} 66_{K}	4_{K}	
	113_{A}		**4_{K}**	
	114_{S}		4_{K}	
1fo0 [.14]	110_{Y}	65_{Q}		
	111_{G}	65_{Q}		

	114_{Y}		5_{Y} 7_{V}	$61A_{A}$ 63_{V} 66_{Q}
1oga [.11]	108_{S}			$61A_{A}$
	109_{R}		5_{F} 6_{V} 7_{F}	**$61A_{A}$** **62_{H}** 63_{V} **66_{Q}**
	110_{S}		5_{F} 6_{V}	66_{Q}
	113_{S}		5_{F}	66_{Q}
	114_{Y}			**61_{A}** $61A_{A}$ 62_{H}
1mi5 [.11]	108_{L}	76_{E}		58_{K}
	109_{G}	76_{E}		
	110_{Q}	69_{T} **72_{Q}** **73_{T}** **76_{E}**	5_{R} **6_{A}**	
	113_{A}		**6_{A}** 7_{Y}	
	114_{Y}	76_{E}	7_{Y} **8_{G}**	58_{K} 59_{W} $61A_{A}$
1lp9 [.11]	109_{W}		5_{F} 6_{F} 7_{P} 8_{V}	58_{K} 59_{W} $61A_{A}$ 63_{V}
	110_{V}		5_{F}	$61A_{A}$
	113_{S}		5_{F}	
	114_{Y}		5_{F}	61_{A} $61A_{A}$ 62_{H} 66_{Q}
1g6r [.9]	107_{G}		6_{Y}	
	108_{G}		6_{Y}	$61A_{A}$ 63_{E}
	109_{G}		4_{R} 6_{Y}	$61A_{A}$ 66_{R}
	114_{G}		4_{R}	66_{R}
	115_{T}			$61A_{A}$
1jtr [.9]	107_{G}		6_{Y}	
	108_{G}		6_{Y}	$61A_{A}$ 63_{E} 66_{R}
	109_{G}		6_{Y}	63_{E} 66_{R}
	114_{G}			66_{R}
1fo0 [.12]	108_{A}			58_{K}
	109_{D}		6_{N} **7_{T}**	58_{K} 59_{W}
	110_{R}	**69_{G}** 70_{N} 72_{Q} **73_{S}**	4_{D} **5_{F}** 6_{N}	
	112_{V}		4_{D} 5_{F} **6_{N}**	66_{R}
	113_{G}		6_{N}	
	114_{N}		6_{N}	$61A_{A}$
1kj2 [.16]	108_{A}		6_{I}	66_{R}
	109_{A}		4_{T} 6_{I}	66_{R}
	110_{P}		4_{T}	
	111_{D}		**4_{T}**	66_{R}
	111.1_{W}			62_{G} 65_{E} 66_{R} 69_{A}

Table 5, continued

	112.1$_{G}$	65$_{Q}$		
1kj2 [.11]	108$_{Y}$	62$_{R}$		
	109$_{Q}$	63$_{E}$ **66$_{K}$**	1$_{K}$ **2$_{V}$** 3$_{I}$ 4$_{T}$	70$_{Y}$ 73$_{T}$
	110$_{G}$	66$_{K}$	4$_{T}$	
	114$_{R}$	**65$_{Q}$** 68$_{K}$ 69$_{G}$ 72$_{Q}$		

	112$_{S}$			61$_{Q}$ 61A$_{A}$
	112.1$_{A}$			61A$_{A}$
	114$_{E}$			69$_{A}$

(C) V-ALPHA and V-BETA FR-IMGT interactions

V-ALPHA FR-IMGT				
	Position	**G-ALPHA1**	**Peptide**	**G-ALPHA2**
1ao7	2$_{K}$	58$_{E}$		
	26$_{S}$	58$_{E}$		
	84A$_{K}$			**73$_{T}$ 76$_{E}$**
1bd2	2$_{Q}$	58$_{E}$ 65$_{R}$		
	84A$_{K}$			72A$_{G}$ 73$_{T}$
1oga	84C$_{R}$			65$_{E}$
1mi5	40$_{H}$		7$_{Y}$	
	52$_{Y}$			62$_{R}$
	55$_{H}$		7$_{Y}$	**61A$_{A}$ 62$_{R}$**
	67$_{V}$			62$_{R}$
1lp9	84A$_{K}$			65$_{E}$
1g6r	2$_{Q}$		4$_{R}$	
	55$_{K}$			65$_{E}$
1jtr	55$_{K}$			65$_{E}$
	84A$_{K}$			76$_{E}$
1kj2	84A$_{K}$			76$_{E}$

V-BETA FR-IMGT				
	Position	**G-ALPHA1**	**Peptide**	**G-ALPHA2**
1bd2	55$_{Y}$	65$_{R}$		
	67$_{D}$	68$_{K}$		
1oga	67$_{Q}$	65$_{R}$		
1mi5	55$_{Y}$	72$_{Q}$ 76$_{E}$		
	66$_{L}$	72$_{Q}$ 75$_{R}$		
1lp9	55$_{Y}$	65$_{R}$		
	67$_{E}$	**65$_{R}$** 68$_{K}$		
1g6r	67$_{E}$	72$_{Q}$		
	84$_{Q}$			58$_{K}$
1jtr	67$_{E}$	72$_{Q}$		
	84$_{Q}$			58$_{K}$

TR positions in bold indicate hydrogen bonds. Three dimensional (3D) structures are from IMGT/3Dstructure-DB [5], http://imgt.cines.fr. Lengths of the CDR-IMGT are shown within brackets. Amino acids are shown in the one-letter code. Sequences of the peptides are reported in Table 4, sequences of the TR V-ALPHA and V-BETA domains in Figure 7 and sequences of the MHC-I G-ALPHA1 and G-ALPHA2 in Figure 8.

conformations to adapt to the different peptides [40]. The CDR3 conformational change does not increase the binding surface but gives a better shape complementarity to the interface [41].

CONCLUSION

The 3D structure of the MHC main chain is well conserved and the peptide binding groove specificity is due to side chains physicochemical characteristics [38]. Both MHC-I and MHC-II grooves have pockets where side chains of bound peptides may anchor [42], the specificity of a peptide to a given MHC being controlled by the physicochemical properties of the pockets. Conversely comparison of peptide sequence alignments and pMHC 3D structures have revealed that some anchored peptide positions with

Table 6
V-ALPHA and V-BETA CDR interactions with MHC-II. (A) V-ALPHA CDR-IMGT interactions, (B) V-BETA CDR-IMGT interactions, (C) V-ALPHA and V-BETA FR-IMGT interactions

(A) V-ALPHA CDR-IMGT interactions

V-ALPHA CDR1-IMGT

	Position	G-ALPHA	Peptide	G-BETA
1j8h [6.]	28_S		2_K	76_H
	29_V		2_K 4_V	76_H
	30_P		4_V	$72A_T$ 76_H
	32_Y			$72A_T$
1d9k [6.]	27_D		3_S	
	28_S			$72A_T$ 76_H
	29_T		3_S 4_H **5_R**	$72A_T$ 76_H
	30_F		5_R	$72A_T$
	31_D		**5_R** 8_I	**66_R** 69_A $72A_T$
	32_Y			66_R

V-ALPHA CDR2-IMGT

	Position	G-ALPHA	Peptide	G-BETA
1j8h [.7.]	57_T			**65_E**
	58_S			69_A $72A_T$
	59_A			65_E
1d9k [.6.]	57_S			65_E 66_R 69_A
	58L			69_A 72_D $72A_T$
	59_V			65_E 66_R 68_R 69_A
	60_S			65_E

V-ALPHA CDR3-IMGT

	Position	G-ALPHA	Peptide	G-BETA
1j8h [.13]	108_E	63_E	**2_K** 4_V	
	110_P		7_N	66_Q
	111_F		7_N 9_L	62_D 63_L 66_Q
	114_E	66_G 69_A 70_N	5_K	
1d9k [.10]	107_T			66_R
	108_G		5_R 8_I	66_R
	109_S	69_Q	8_I	66_R

(B) V-BETA CDR-IMGT interactions

V-BETA CDR-IMGT

	Position	G-ALPHA	Peptide	G-BETA
1j8h [5.]	27_M		10_K	
	28_D	76_A	**10_K**	
	29_H		10_K	
	30_E	72_A 73_V 76_A	**10_K**	
	31_N	69_A		
1d9k [5.]	30_N	76_H		
	31_N	**69_Q**		

V-BETA CDR2-IMGT

	Position	G-ALPHA	Peptide	G-BETA
1j8h [.6.]	57_Y	65_Q 66_G 68_L 69_A 72_A		
	58_D	68_L 72_A **75_K**		
	61_M	43_K 68_L		
1d9k [.6.]	57_Y	65_Q 66_G 68_L 69_Q 72_A		

V-BETA CDR3-IMGT

	Position	G-ALPHA	Peptide	G-BETA
1j8h [.12]	108_S	73_V	10_K	
	109_T	69_A 70_N 73_V	5_K 7_N 8_T	
	110_G	73_V	8_T 9_L **10_K**	
	112_L		10_K	58_Y
	113_P			$61A_Q$ 62_D 63_L
1d9k [.11]	108_G		**11_E**	
	109_Q		11_E	58_Y $61B_Y$
	110_G		10_W 11_E	$61B_Y$ 66_R
	113_R			61_K **$61A_Q$** $61B_Y$ 65_E 66_R
	114_A			66_R

Table 6, continued

	113F	**69Q** 73T	8I 9E 10W 11E	61BY 66R
	114N	69Q		66R
	115K	65Q		

(C) V-ALPHA and V-BETA FR-IMGT interactions				
V-ALPHA FR-IMGT				
	Position	**G-ALPHA**	**Peptide**	**G-BETA**
1j8h	55K			62D
1d9k	84AK			**72D**
V-BETA FR-IMGT				
	Position	**G-ALPHA**	**Peptide**	**G-BETA**
1j8h	55F	65Q		
	66K	43K		
	67E	**43K 65Q**		
	84K	72A 76A	10K	
1d9k	55Y	**65Q**		
	66T	**43K**		
	67E	**43K** 65Q 68L		
	68K	65Q		

TR positions in bold indicate hydrogen bonds. Three dimensional (3D) structures are from IMGT/3Dstructure-DB [5], http://imgt.cines.fr. Lengths of the CDR-IMGT are shown within brackets. Amino acids are shown in the one-letter code. Sequences of the peptides are reported in Table 4, sequences of the TR V-ALPHA and V-BETA domains in Figure 7, and sequences of the MHC-II G-ALPHA and G-BETA in Figure 8.

conserved properties were needed to bind a peculiar MHC allele. Several databases, SYFPEITHI [43], JenPep [44] and MHCpep [45], provide peptide sequences associated with MHC alleles together with anchor positions and experimental data on affinity. These observations have extensively been used in peptide/MHC binding prediction [46–48] (a list of prediction programs and servers is available at "The IMGT Immunoinformatics page", http://imgt.cines.fr). Nevertheless exceptions have been found [49–51] and it has been noted that only 30% of peptides with the expected pattern really bind whereas some peptides without the expected pattern do bind [52]. Peptide/MHC binding prediction and epitope prediction remain a big challenge. In order to compare interactions between MHC domains of classes I and II and with peptides of different lengths, we have defined eleven IMGT pMHC contact sites which are based on the IMGT unique numbering for G-DOMAIN and G-LIKE-DOMAIN [16]. IMGT contact sites allow comparison either with the IMGT reference pMHC contact sites, or with other IMGT contact sites. They also allow to underline the impact of mutations of altered peptides, such as the ones observed in altered Tax peptide in 1qsf and 1qse [29]. IMGT pMHC contact sites are available for all the pMHC and TR/pMHC in IMGT/3Dstructure-DB [5], http://imgt.cines.fr.

With only 18 TR/pMHC 3D structures, the atomic details of TR/pMHC interactions already show a

great deal of variability. IMGT standardization is a step towards a better understanding of the mechanisms ruling TR/pMHC recognition. It will help comparing new experimentally resolved 3D structures with published data. However the TR/pMHC interactions are far from being unravelled and the study of the TR/pMHC interactions with the other proteins of the immunological synapse will be crucial. For example, the interaction between a MHC and the CD4 considerably enhances the pMHC/TR sensibility [53,54]. The understanding of the T cell triggering early events is subject to active studies.

Although the TR/pMHC binding represents a necessary step for the TR recognition, many factors, the TR affinity for the pMHC, the relocation of surface proteins such as CD4 or CD8 in the immunological synapse are necessary for generating the T cell activation signal. Each of these steps needs to be described and characterized so that data from different experiments can be integrated. IMGT standardization will be further extended on the IMGT Web site at http://imgt.cines.fr as new parameters will become available.

CITING IMGT/3DSTRUCTURE-DB

Users are requested to cite reference 5 and this article, and to quote the IMGT home page URL, http://imgt.cines.fr.

ACKNOWLEDGEMENTS

We are grateful to the IMGT team for helpful discussion. K. Q. was the recipient of a doctoral grant from the Ministère de l'Education Nationale, de l'Enseignement Supérieur et de la Recherche (MENESR) and is currently supported by a grant from the Association pour la Recherche sur le Cancer (ARC). IMGT is a registered Centre National de la Recherche Scientifique (CNRS) mark. IMGT is a National RIO Bioinformatics Platform since 2001 (CNRS, INSERM, CEA, INRA). IMGT was funded in part by the BIOMED1 (BIOCT930038), Biotechnology BIOTECH2 (BIO4CT960037) and 5th PCRDT Quality of Life and Management of Living Resources (QLG2-2000-01287) programmes of the European Union and received subventions from ARC and from the Génopole-Montpellier-Languedoc-Roussillon. IMGT is currently supported by the CNRS, the MENESR (Université Montpellier II Plan Pluri-Formation, BIOSTIC-LR2004 Région Languedoc-Roussillon and ACI-IMPBIO IMP82-2004), and GIS-AGENAE. Part of this work was carried out in the frame of the European Science Foundation Scientific Network Myelin Structure and its role in autoimmunity (MARIE).

REFERENCES

[1] Lefranc, M.-P. and Lefranc, G. (2001). The T cell receptor FactsBook. Academic Press, London, UK, 398 pages. http://imgt.cines.fr/textes/IMGTindex/factsbook.html

[2] Vasmatzis, G., Cornette, J., Sezerman, U. and DeLisi, C. (1996). TcR recognition of the MHC-peptide dimer: structural properties of a ternary Complex. J. Mol. Biol. **261**, 72-89.

[3] Sim, B. C., Zerva, L., Greene, M. I. and Gascoigne, N. R. (1996). Control of MHC restriction by TCR Valpha CDR1 and CDR2. Science **273**, 963-966.

[4] Kjer-Nielsen, L., Clements, C. S., Purcell, A. W., Brooks, A. G., Whisstock, J. C., Burrows, S. R., McCluskey, J. and Rossjohn, J. (2003). A structural basis for the selection of dominant alphabeta T cell receptors in antiviral immunity. Immunity **18**, 53-64.

[5] Kaas, Q., Ruiz, M. and Lefranc, M.-P. (2004). IMGT/3Dstructure-DB and IMGT/StructuralQuery, a database and a tool for immunoglobulin, T cell receptor and MHC structural data. Nucleic Acids Res. **32**, D208-D210.

[6] Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. and Bourne, P. E. (2000). The Protein Data Bank. Nucleic Acids Res. **28**, 235-242.
[7] Lefranc, M.-P., Giudicelli, V., Kaas, Q., Duprat, E., Jabado-Michaloud, J., Scaviner, D., Ginestoux, C., Clément, O., Chaume, D. and Lefranc, G. (2005). IMGT, the international ImMunoGeneTics information system®. Nucleic Acids Res. **33**, D593-D597.
[8] Lefranc, M.-P., Clément, O., Kaas, Q., Duprat, E., Chastellan, P., Coelho, I., Combre, K., Ginestoux, C., Giudicelli, V., Chaume, D. and Lefranc, G. (2004). IMGT-Choreography for Immunogenetics and Immunoinformatics. In Silico Biol. **5**, 0006.
[9] Lefranc, M.-P., Giudicelli, V., Ginestoux, C., Bosc, N., Folch, G., Guiraudou, D., Jabado-Michaloud, J., Magris, S., Scaviner, D., Thouvenin, V., Combres, K., Girod, D., Jeanjean, S., Protat, C., Yousfi Monod, M., Duprat, E., Kaas, Q., Pommié, C., Chaume, D. and Lefranc, G. (2003). IMGT-ONTOLOGY for immunogenetics and immunoinformatics. In Silico Biol. **4**, 0004.
[10] Lefranc, M.-P. (2004). IMGT-ONTOLOGY and IMGT databases, tools and Web resources for immunogenetics and immunoinformatics. Mol. Immunol. **40**, 647-660.
[11] Giudicelli, V. and Lefranc, M.-P. (1999). Ontology for immunogenetics: the IMGT-ONTOLOGY. Bioinformatics **15**, 1047-1054.
[12] Wain, H. M., Bruford, E. A., Lovering, R. C., Lush, M. J., Wright, M. W. and Povey, S. (2002). Guidelines for human gene nomenclature. Genomics **79**, 464-470.
[13] Blake, J. A., Richardson, J. E., Bult, C. J., Kadin, J. A. and Eppig, J. T. (2003). MGD: the Mouse Genome Database. Nucleic Acids Res. **31**, 193-195.
[14] Lefranc, M.-P., Pommié, C., Ruiz, M., Giudicelli, V., Foulquier, E., Truong, L., Thouvenin-Contet, V. and Lefranc, G. (2003). IMGT unique numbering for immunoglobulin and T cell receptor variable domains and Ig superfamily V-like domains. Dev. Comp. Immunol. **27**, 55-77.
[15] Lefranc, M.-P., Pommié, C., Kaas, Q., Duprat, E., Bosc, N., Guiraudou, D., Jean, C., Ruiz, M., Da Piedade, L., Rouard, M., Foulquier, E., Thouvenin, V. and Lefranc, G. (2005). IMGT unique numbering for immunoglobulin and T cell receptor constant domains and Ig superfamily C-like domains. Dev. Comp. Immunol. **29**, 185-203.
[16] Lefranc, M.-P., Duprat, E., Kaas, Q., Tranne, M., Thiriot, A. and Lefranc, G. (2005). IMGT unique numbering for MHC groove G-DOMAIN and MHC superfamily (MhcSF) G-LIKE-DOMAIN. Dev. Comp. Immunol. **29**, 917-938.
[17] Robinson, J., Waller, M. J., Parham, P., de Groot, N., Bontrop, R., Kennedy, L. J., Stoehr, P. and Marsh, S. G. (2003). IMGT/HLA and IMGT/MHC: sequence databases for the study of the major histocompatibility complex. Nucleic Acids Res. **31**, 311-314.
[18] Giudicelli, V., Chaume, D., Lefranc, M.-P. (2005). IMGT/GENE-DB: a comprehensive database for human and mouse immunoglobulin and T cell receptor genes. Nucleic Acids Res. **33**, D256-D261.
[19] Giudicelli, V., Chaume, D. and Lefranc, M.-P. (2004). IMGT/V-QUEST, an integrated software program for immunoglobulin and T cell receptor V-J and V-D-J rearrangement analysis. Nucleic Acids Res. **32**, W435-W440.
[20] Yousfi Monod, M., Giudicelli, V., Chaume, D. and Lefranc, M.-P. (2004). IMGT/JunctionAnalysis: the first tool for the analysis of the immunoglobulin and T cell receptor complex V-J and V-D-J JUNCTIONs. Bioinformatics **20**, I379-I385.
[21] Lesk, A. M. and Chothia, C. (1982). Evolution of proteins formed by beta-sheets. II. The core of the immunoglobulin domains. J. Mol. Biol. **160**, 325-342.
[22] Stewart-Jones, G. B. E., McMichael, A. J., Bell, J. I., Stuart, D. I. and Jones, E. Y. (2003). A structural basis for immunodominant human T cell receptor recognition. Nat. Immunol. **4**, 657-663.
[23] Hennecke, J. and Wiley, D. C. (2002). Structure of a complex of the human alpha/beta T cell receptor (TCR) HA1.7, influenza hemagglutinin peptide, and major histocompatibility complex class II molecule, HLA-DR4 (DRA*0101 and DRB1*0401): insight into TCR cross-restriction and alloreactivity. J. Exp. Med. **195**, 571-581.
[24] Wang, J. H., Meijers, R., Xiong, Y., Liu, J. H., Sakihama, T., Zhang, R., Joachimiak, A. and Reinherz, E. L. (2001). Crystal structure of the human CD4 N-terminal two-domain fragment complexed to a class II MHC molecule. Proc. Natl. Acad. Sci. USA **98**, 10799-10804.
[25] Wang, J. H. and Reinherz, E. L. (2002). Structural basis of T cell recognition of peptides bound to MHC molecules. Mol. Immunol. **38**, 1039-1049.
[26] Garboczi, D. N., Ghosh, P., Utz, U., Fan, Q. R., Biddison, W. E. and Wiley, D. C. (1996). Structure of the complex between human T-cell receptor, viral peptide and HLA-A2. Nature **384**, 134-141.
[27] Luz, J. G., Huang, M., Garcia, K. C., Rudolph, M. G., Apostolopoulos, V., Teyton, L. and Wilson, I. A. (2002). Structural comparison of allogeneic and syngeneic T cell receptor-peptide-major histocompatibility complex complexes: a buried alloreactive mutation subtly alters peptide presentation substantially increasing V(beta) Interactions. J. Exp. Med. **195**, 1175-1186.
[28] Zhang, C., Anderson, A. and DeLisi, C. (1998). Structural principles that govern the peptide-binding motifs of class I MHC molecules. J. Mol. Biol. **281**, 929-947.

[29] Ding, Y. H., Baker, B. M., Garboczi, D. N., Biddison, W. E. and Wiley, D. C. (1999). Four A6-TCR/peptide/HLA-A2 structures that generate very different T cell signals are nearly identical. Immunity **11**, 45-56.
[30] Ding, Y. H., Smith, K. J., Garboczi, D. N., Utz, U., Biddison, W. E. and Wiley, D. C. (1998). Two human T cell receptors bind in a similar diagonal mode to the HLA-A2/Tax peptide complex using different TCR amino acids. Immunity **8**, 403-411.
[31] Buslepp, J., Wang, H., Biddison, W. E., Appella, E. and Collins, E. J. (2003). A correlation between TCR Valpha docking on MHC and CD8 dependence: implications for T cell selection. Immunity **19**, 595-606.
[32] Degano, M., Garcia, K. C., Apostolopoulos, V., Rudolph, M. G., Teyton, L. and Wilson, I. A. (2000). A functional hot spot for antigen recognition in a superagonist TCR/MHC complex. Immunity **12**, 251-261.
[33] Garcia, K. C., Degano, M., Pease, L. R., Huang, M., Peterson, P. A., Teyton, L. and Wilson, I. A. (1998). Structural basis of plasticity in T cell receptor recognition of a self peptide-MHC antigen. Science **279**, 1166-1172.
[34] Reiser, J. B., Darnault, C., Guimezanes, A., Gregoire, C., Mosser, T., Schmitt-Verhulst, A. M., Fontecilla-Camps, J. C., Malissen, B., Housset, D. and Mazza, G. (2000). Crystal structure of a T cell receptor bound to an allogeneic MHC molecule. Nat. Immunol. **1**, 291-297.
[35] Reiser, J. B., Darnault, C., Gregoire, C., Mosser, T., Mazza, G., Kearney, A., van der Merwe, P. A., Fontecilla-Camps, J. C., Housset, D. and Malissen, B. (2003). CDR3 loop flexibility contributes to the degeneracy of TCR recognition. Nat. Immunol. **4**, 241-247.
[36] Reiser, J. B., Gregoire, C., Darnault, C., Mosser, T., Guimezanes, A., Schmitt-Verhulst, A. M., Fontecilla-Camps, J. C., Mazza, G., Malissen, B. and Housset, D. (2002). A T cell receptor CDR3beta loop undergoes conformational changes of unprecedented magnitude upon binding to a peptide/MHC class I complex. Immunity **16**, 345-354.
[37] Hennecke, J., Carfi, A. and Wiley, D. C. (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.
[38] Reinherz, E. L., Tan, K., Tang, L., Kern, P., Liu, J., Xiong, Y., Hussey, R. E., Smolyar, A., Hare, B., Zhang, R., Joachimiak, A., Chang, H. C., Wagner, G. and Wang, J. (1999). The crystal structure of a T cell receptor in complex with peptide and MHC class II. Science **286**, 1913-1921.
[39] Sim B.C., Zerva L., Grene M.I. and Gascoigne N.R. (1996), Control of MHC restriction by TCR Valpha CDR1 and CDR2, Science **273**, 963-964.
[40] Rudolph, M. G., Luz, J. G. and Wilson, I. A. (2002). Structural and thermodynamic correlates of T cell signaling. Annu. Rev. Biophys. Biomol. Struct. **31**, 121-149.
[41] Lawrence, M. C. and Colman, P. M. (1993). Shape complementarity at protein/protein interfaces. J. Mol. Biol. **234**, 946-950.
[42] Falk, K., Rotzschke, O., Stevanovic, S., Jung, G. and Rammensee, H.-G. (1991). Allele-specific motifs revealed by sequencing of self-peptides eluted from MHC molecules. Nature **351**, 290-296.
[43] Rammensee, H.-G., Bachmann, J., Emmerich, N. P. N., Bachor, O. A. and Stevanovic, S. (1999). SYFPEITHI: database for MHC ligands and peptide motifs. Immunogenetics **50**, 213-219.
[44] Blythe, M. J., Doytchinova, I. A. and Flower, D. R. (2002). JenPep: a database of quantitative functional peptide data for immunology. Bioinformatics **18**, 434-439.
[45] Brusic, V., Rudy, G. and Harrison, L. C. (1998). MHCPEP, a database of MHC-binding peptides: update 1997. Nucleic Acids Res. **26**, 368-371.
[46] Singh, H. and Raghava, G. P. (2003). ProPred1: prediction of promiscuous MHC Class-I binding sites. Bioinformatics **19**, 1009-1014.
[47] Adams, H. P. and Koziol, J. A. (1995). Prediction of binding to MHC class I molecules. J. Immunol. Methods **185**, 181-190.
[48] Vasmatzis, G., Zhang, C., Cornette, J. L. and DeLisi, C. (1996). Computational determination of side chain specificity for pockets in class I MHC molecules. Mol. Immunol. **33**, 1231-1239.
[49] Mandelboim, O., Bar-Haim, E., Vadai, E., Fridkin, M. and Eisenbach, L. (1997). Identification of shared tumor-associated antigen peptides between two spontaneous lung carcinomas. J. Immunol. **159**, 6030-6036.
[50] Apostolopoulos, V., Yu, M., Corper, A. L., Teyton, L., Pieters, G. A., McKenzie, I. F. C. and Wilson, I. A. (2002). Crystal structure of a non-canonical Low-affinity Peptide Complexed with MHC Class I: A New approach for vaccine design. J. Mol. Biol. **318**, 1293-1305.
[51] Scott, C. A., Peterson, P. A., Teyton, L. and Wilson, I. A. (1998). Crystal structures of two I-Ad-peptide complexes reveal that high affinity can be achieved without large anchor residues. Immunity **8**, 319-329.
[52] Gulukota, K., Sidney, J., Sette, A. and DeLisi, C. (1997). Two complementary methods for predicting peptides binding major histocompatibility complex molecules. J. Mol. Biol. **267**, 1258-1267.
[53] Irvine, D. J., Purbhoo, M. A., Krosgaard, M. and Davis, M. M. (2002). Direct observation of ligand recognition by T cells. Nature **419**, 845-849.
[54] Davis, M. M. (2002). A new trigger for T cells. Cell **110**, 285-287.