ImmunoGrid
The European Virtual Human Immune System Project

D1.4 New concepts of "INTERACTION" and "LOCALIZATION" and deduced IMGT Scientific chart rules report

Project Acronym: IMMUNOGRID

Contract no: IST-2004-028069

Date: 2 March 2009

Due Date: 14 March 2009

Principal Authors: Marie-Paule Lefranc, François Ehrenmann, Véronique Giudicelli and Patrice Duroux (CNRS)

Revision: 1.0

Dissemination Level: PU

Content

Section 1. Introduction 3
Section 2. Overview with previous deliverables 3
2.1. D1.1- New and enhanced concepts and rules 3
2.2. D1.2- Scientific chart rules and ontologies report 4
2.3. D1.3- IMGT unique numbering and controlled vocabularies report 6
2.4. D1.4- New concepts do "INTERACTION" and LOCALIZATION" and deduced IMGT Scientific chart rules 8
Section 3. The TR/peptide/MHC trimolecular complex 8
Section 4. pMHC interactions 10
4.1. MHC G-DOMAIN 10
4.2. Peptide/MHC interactions: Concept of pMHC contact sites 12
Section 5. TR/pMHC interactions 18
5.1. TR V-DOMAIN 18
5.2. TR/pMHC interactions: Concept of IMGT Residue@Position 20
Section 6. Implementation of the WP1 concepts 21
Section 6. ImmunoGrid concepts and modelling of the immune system 23
Section 7. References 24

Section 1. Introduction

The focus of WP1 ***"Immune system standardized concepts"*** is the setting up of the standardized rules and concepts which are part of the identification, description and classification of the biological components and processes, in the modelling of the "Virtual Immune System" (VIS).

The aim of this deliverable D1.4 *New concepts of "INTERACTION" and "LOCALIZATION" and deduced IMGT Scientific chart rules* is to define the concepts of the INTERACTION axiom that are necessary for a standardized description of the interactions between structural domains, and the concepts of the LOCALIZATION axiom that are needed for establishing correlations between multi-scale models (molecule, cell, organ, organism) in the VIS modelling. The concepts have been detailed for IG, TR and MHC. The domains include the groove (G) domains of the MHC proteins, the variable (V) domains and the constant (C) domains of the IG and TR which have an immunoglobulin fold. The characterization of domain interactions and antigen/receptor interactions is based on the standardization of both sequences and 3D structures. This is achieved by the use of the standardized amino acid positions according to the IMGT unique numbering. In this report we focus on the interactions of the peptide/MHC (pMHC) and TR/pMHC trimolecular complexes as examples as they are key components of the current models in the simulator and they trigger the signalling cascade of the T cells involved in the immune responses.

Section 2 provides an overview of the previous deliverables in relation with the current deliverable D1.4. Section 3 briefly specifies the components of the TR/peptide/MHC trimolecular complexes. Section 4 defines the pMHC interactions based on the concepts of pMHC contact sites and IMGT unique numbering for G-DOMAIN, for the analysis and modelling of pMHC interactions. Section 5 defines the TR/pMHC interactions based on the concept of IMGT Residue@Position, for the analysis and modelling of interactions between TR and pMHC. Section 6 specifies the implementation of the WP1 concepts. Section 7 describes the impact of the ImmunoGrid concepts in the modelling of the immune system.

Section 2. Overview with previous deliverables

2.1. D1.1– New and enhanced concepts and rules

The immunogenetics knowledge is particularly complex. IMGT-ONTOLOGY [1] is the first and so far unique ontology in immunogenetics and immunoinformatics. IMGT-ONTOLOGY provides a semantic specification of the terms to be used in immunogenetics and immunoinformatics and manages the related knowledge, thus allowing the standardization for immunogenetics data from genome, proteome, genetics and three-dimensional (3D) structures [2-5]. IMGT-ONTOLOGY results from a deep expertise in the

domain and an extensive effort of conceptualization and is crucial to make immunogenetics knowledge amenable to modelling.

IMGT-ONTOLOGY manages the immunogenetics knowledge through diverse facets relying on seven axioms, "IDENTIFICATION", "CLASSIFICATION", "DESCRIPTION", "LOCALIZATION", "NUMEROTATION", "ORIENTATION" and "OBTENTION". These axioms postulate that objects, processes and relations have to be identified, described, classified, numerotated, localized, orientated, and that the way they are obtained has to be determined (Fig. 1). The axioms constitute the Formal IMGT-ONTOLOGY, also designated as IMGT-Kaleidoscope [4].

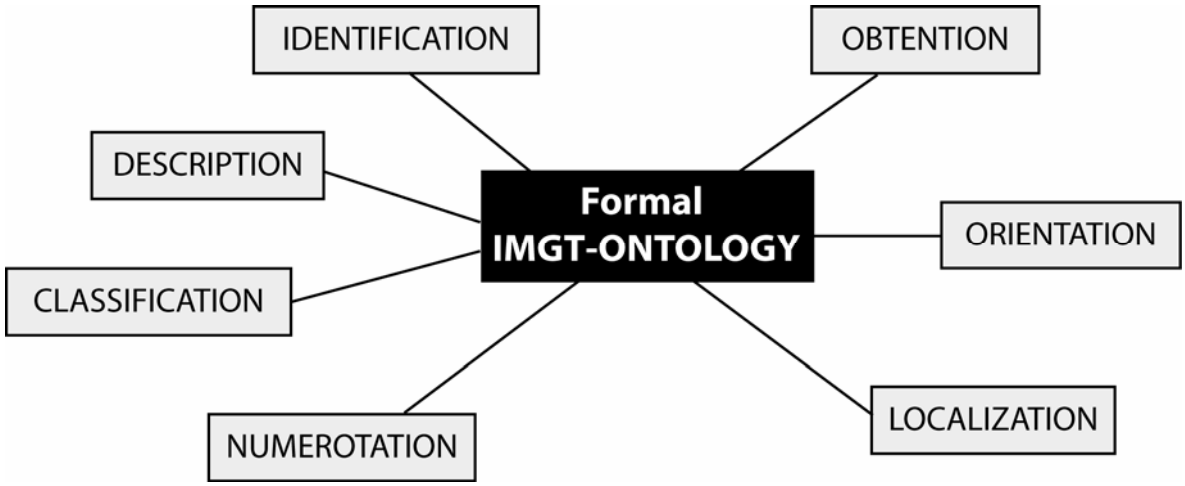


Figure 1. The axioms of the Formal IMGT-ONTOLOGY or IMGT-Kaleidoscope.

The deliverable D1.1 *New and enhanced concepts and rules* (internal ImmunoGrid report) provided to WP2 (Molecular level modelling), WP3 (System level modelling) and WP4 (Simulator design) the semantic classification and standardization of the knowledge necessary for the "Virtual Immune System" (VIS) modelling.

2.2. D1.2– Scientific chart rules and ontologies report

The deliverable D1.2 *Scientific chart rules and ontologies report* formalized the new and enhanced concepts and rules for the identification, description and classification of the antigen receptors (IG, TR) and the MHC, that are major molecular components of the "Virtual Immune System" modelling. The IG, TR and MHC proteins are 450 to 500 million years "old" and are characteristic of the adaptive immune responses in vertebrates. They allow a very fine specific recognition of the "non self" represented by infectious pathogens, viruses, bacteria, parasites and their products (toxins...), and by vaccine and tumoral antigens. These complex and heterogenous data are managed in IMGT® (http://www.imgt.org), the flagship of Europe in Immunogenetics and immunoinformatics (BIOMED, BIOTECH, 5th PCRDT) [6, 7] and a key component of the ImmunoGrid project (http://www.immunogrid.org/).

The standardization in the deliverable D1.2 is based on IMGT-ONTOLOGY [1]. Novelty resides in the emergence, identification and characterization of new standards and concepts in IMGT-ONTOLOGY that are required for a systemic approach of the adaptive immune responses and that can represent the corresponding knowledge in other fields of biology. Three axioms, "IDENTIFICATION", "DESCRIPTION" and "CLASSIFICATION", of the Formal IMGT-ONTOLOGY or IMGT-Kaleidoscope, were defined in D1.2, with the concepts that have been essential for the conceptualization of the molecular immunogenetics knowledge. The IMGT-ONTOLOGY axioms, the related IMGT Scientific chart rules, ImmunoGrid deliverables and ImmunoGrid examples which were defined and formalized in D1.2 are listed in Table 1.

Table 1. IDENTIFICATION, DESCRIPTION, CLASSIFICATION and NUMEROTATION axioms of the Formal IMGT-ONTOLOGY.

IMGT-ONTOLOGY axioms	IMGT Scientific chart rules[1]	ImmunoGrid deliverables	Examples in ImmunoGrid
IDENTIFICATION	Keywords	D1.2 *Scientific chart rules and ontologies report (PU)*	IG, TR, MHC and RPI nucleotide and amino acid sequence, and 3D structure identification
DESCRIPTION	Labels		IG, TR, MHC and RPI nucleotide and amino acid sequence, and 3D structure description
CLASSIFICATION	Nomenclature		IG, TR, MHC and RPI gene and allele names
NUMEROTATION	Unique numbering	D1.3 *IMGT unique numbering and controlled vocabularies report (PU)*	IMGT Colliers de Perles for V-DOMAIN and V-LIKE-DOMAIN
			IMGT Colliers de Perles for C-DOMAIN and C-LIKE-DOMAIN
			IMGT Colliers de Perles for G-DOMAIN and G-LIKE-DOMAIN

[1] The corresponding controlled vocabulary and rules are available in the IMGT Scientific Chart at http://www.imgt.org

The examples in ImmunoGrid include standardized IMGT keywords (IDENTIFICATION axiom), standardized IMGT labels[1] for the receptors, chains, domains and regions (DESCRIPTION axiom) and standardized IMGT gene and allele names (CLASSIFICATION axiom) [8-12]. The deliverable D1.2 represents the state of the art for the IDENTIFICATION, DESCRIPTION and CLASSIFICATION axioms and concepts. It was provided to WP2 (Molecular level modelling), WP3 (System level modelling) and WP4 (Simulator design) in September 2007 and made publicly available on the ImmunoGrid Web site at http://www.immunogrid.org/immunogrid/publications.

2.3. D1.3– IMGT unique numbering and controlled vocabularies report

The deliverable D1.3 *IMGT unique numbering and controlled vocabularies report* defines the concepts which are necessary for a standardized description of the structural domains in the Virtual Immune System modelling. They are detailed for the antigen receptors - immunoglobulins (IG) and T cell receptors (TR) - and for the major histocompatibility complex (MHC). These proteins are key actors of the adaptive immune response and are characterized, for the antigen receptors, by an incredible diversity (10^{12} IG or antibodies and 10^{12} TR per individual) due to molecular mechanisms such as DNA rearrangements, N-diversity, and for the IG, somatic mutations and, for the MHC (designated as HLA in humans), by an extensive allelic polymorphism. The concepts of numerotation (IMGT unique numbering of the V and C domains of the IG and TR, G domains of the MHC [13-15]) have been extended to the V-like and C-like domains of the IgSF (other than IG and TR) and to the G-like domain of the MhcSF (other than MHC) [16-17]. Examples of concepts of numerotation for G and V type domains are shown in Tables 2 and 3.

Table 2. Examples of concepts of numerotation for a G type domain (G-DOMAIN and G-LIKE-DOMAIN).

Strands, turns and helix	IMGT positions[a]	Lengths[b]	Characteristic positions[c]
A-STRAND	1-14	14	7A, CYS-11
AB-TURN	15-17	3 (or 0)	
B-STRAND	18-28	11	
BC-TURN	29-30	2	
C-STRAND	31-38	8	
CD-TURN	39-41	3 (or 1)	
D-STRAND	42-49	8	
HELIX	50-92	43-48	54A, 61A, 61B, 72A, CYS-74, 92A

[a] based on the IMGT unique numbering for G-DOMAIN and G-LIKE-DOMAIN [15].
[b] in number of amino acids (or codons).
[c] for more details, see [15].

[1] IMGT labels of the IMGT-ONTOLOGY DESCRIPTION concept are written in capital letters. Definitions of IMGT labels are available in the IMGT Scientific chart at http://www.imgt.org

Table 3. Examples of concepts of numerotation for a V type domain (V-DOMAIN and V-LIKE-DOMAIN).

Strands and loops	IMGT positions[a]	Lengths[b]	Characteristic positions	FR-IMGT and CDR-IMGT in V-DOMAIN
A-STRAND	1-15	15 (14 if gap at 10)		FR1-IMGT
B-STRAND	16-26	11	1st-CYS 23	
BC-LOOP	27-38	12 (or less)		CDR1-IMGT
C-STRAND	39-46	8	CONSERVED-TRP 41	FR2-IMGT
C'-STRAND	47-55	9		
C'C"-LOOP	56-65	10 (or less)		CDR2-IMGT
C"-STRAND	66-74	9 (or 8 if gap at 73)		FR3-IMGT
D-STRAND	75-84	10 (or 8 if gaps at 81,82)		
E-STRAND	85-96	12	hydrophobic 89	
F-STRAND	97-104	8	2nd-CYS 104	
FG-LOOP	105-117	13 (or less, or more)		CDR3-IMGT
G-STRAND	118-128	11	(1)	FR4-IMGT

[a] based on the IMGT unique numbering for V-DOMAIN and V-LIKE-DOMAIN [13].
[b] in number of amino acids (or codons).
(1) In the IG and TR V-DOMAINs, the G-STRAND is the C-terminal part of the J-REGION, with J-PHE or J-TRP 118 and the canonical motif F/W-G-X-G at positions 118-121.

The deliverable D1.3 represents the state of the art for the NUMEROTATION axiom and for the standardization of both sequences and 3D structures by providing standardized amino acid positions according to the IMGT unique numbering (NUMEROTATION axiom). It was provided to WP2 (Molecular level modelling), WP3 (System level modelling) and WP4 (Simulator design) in February 2008.
The D1.3 deliverable is publicly available on the ImmunoGrid Web site at http://www.immunogrid.org/immunogrid/publications.

2.4. D2.4– New concepts of “INTERACTION” and “LOCALIZATION” and deduced IMGT Scientific chart rules

The deliverable D1.4 defines the concepts of the INTERACTION axiom that are necessary for a standardized description of the interactions between structural domains, and the concepts of the LOCALIZATION axiom that are needed for establishing correlations between multi-scale models (molecule, cell, organ, organism) in the VIS modeling. The concepts have been detailed for IG, TR and MHC. The domains include the groove (G) domains of the MHC proteins, the variable (V) domains and the constant (C) domains of the IG and TR which have an immunoglobulin fold. The deliverable D1.4 represents the state of the art for domain interactions and antigen/receptor interactions as their characterization is based on the standardization of both sequences and 3D structures. This is achieved by the use of the standardized amino acid positions according to the IMGT unique numbering (NUMEROTATION axiom). In this report we focus on the interactions of the peptide/MHC (pMHC) and TR/pMHC trimolecular complexes which are major components of the current models in the simulator. Indeed these concepts are fundamental for the analysis and modelling of the interactions between pMHC and TR (pMHC-I and TR expressed by $CD8^+$ cytotoxic T cells, and pMHC-II and TR expressed by $CD4^+$ helper T cells) which trigger the signalling cascade of the T cells involved in the immune responses.
The D1.4 report is publicly available on the ImmunoGrid Web site at http://www.immunogrid.org/immunogrid/publications.

Section 3. The TR/peptide/MHC trimolecular complex

T cells are implicated in the adaptative (specific) immune response against a stress of viral, bacterial, fungal or tumoral origin. They identify antigenic peptides presented by the major histocompatibility complex (MHC, HLA in humans) at the cell surface of antigen presenting cells. The recognition is carried out by the T cell receptor complex (TcR), a multisubunit transmembrane surface complex made up of the 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. One of the key elements in the adaptive immune response is therefore the presentation of peptides by the MHC to the TR at the surface of T cells. As a consequence, the characterization of the peptide/MHC (pMHC) complexes and of the TR/pMHC trimolecular complexes is crucial to the fields of immunology, vaccination and immunotherapy and is at the core of the ImmunoGrid models.
Previous deliverables based on IMGT-ONTOLOGY have provided the concepts that are necessary for a standardized characterization of sequences and three-dimensional (3D) structures of the TR and MHC involved in a TR/pMHC trimolecular complex (Fig. 2). Deliverable D1.2 has provided the concepts of classification (standardized IMGT gene and allele names, CLASSIFICATION axiom) and the concepts of description (standardized IMGT labels, DESCRIPTION axiom). Deliverable D1.3 has provided the concepts of numerotation (IMGT unique numbering, NUMEROTATION axiom).

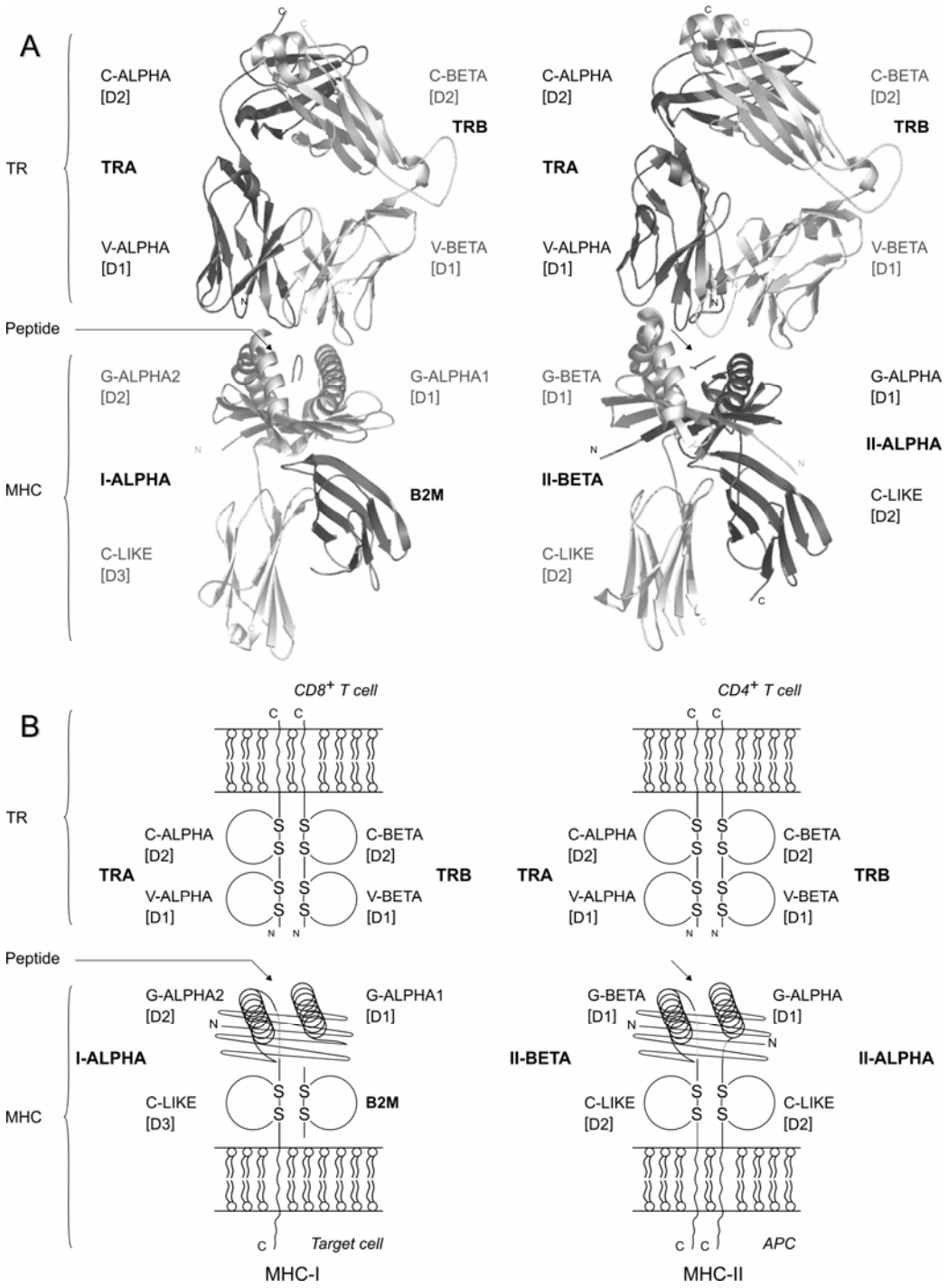


Figure 2. T cell receptor/peptide/MHC complexes with MHC class I (TR/pMHC-I) and MHC class II (TR/pMHC-II). [D1], [D2] and [D3] indicate the domains. (A) 3D structures of TR/pMHC-I (1oga) and TR/pMHC-II (1j8h) [18, 19]. (B) Schematic representation of TR/pMHC-I and TR/pMHC-II. The TR (TR-ALPHA and TR-BETA) chains, the MHC-I (I-ALPHA and β-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.

The MHC cleft that binds the peptide is formed by two groove domains (G-DOMAIN), each one comprising four antiparallel β strands and one α helix. The IMGT unique numbering for G-DOMAIN applies both to the first two domains (G-ALPHA1 and G-ALPHA2) of the MHC-I α chain (I-ALPHA), and to the first domain (G-ALPHA and G-BETA) of the MHC-II α chain (II-ALPHA) and β chain (II-BETA), respectively.
The variable domains of the TR alpha and beta chains that bind the pMHC complex comprise three complementarity determining regions (CDR-IMGT) supported by four framework regions (FR-IMGT). The IMGT unique domain for V-DOMAIN applies to the V-ALPHA domain of the α chain and to the V-BETA domain of the β chain. Coordinate files that were used to define the D1.4 concepts are from IMGT/3Dstructure-DB [18, 19] (http://www.imgt.org) with original crystallographic data from the RCSB Protein Data Bank PDB(http://www.rcsb.org/pdb/home/home.do).

Section 4. pMHC interactions

4.1. MHC G-DOMAIN

The four G-DOMAIN, G-ALPHA1 and G-ALPHA2 of the MHC-I, and G-ALPHA and G-BETA of the MHC-II have a similar groove 3D structure that consists of one sheet of four antiparallel β strands ("floor" of the groove or platform) and one long helical region ("wall" of the groove). In order to compare data from different MHC sequences and 3D structures, the IMGT unique numbering for G-DOMAIN has been set up (Deliverable 1.3). This has allowed to provide the frame to the MHC domain Protein display (Fig. 3) and to the IMGT Colliers de Perles for G-DOMAIN (Fig. 4).
Based on the standardized criteria defined in Deliverable D1.3, the IMGT Colliers de Perles and the following information were provided to WP2, WP3 and WP4 for the simulator:

For each G-DOMAIN, the A strand comprises positions 1 to 14, B strand positions 18 to 28, C strand positions 31 to 38, and D strand positions 42 to 49. The helix (positions 50 to 92) seats on the β sheet and its axis forms an angle of about 40 degrees with the β strands. The helix is split into two parts separated by a kink, positions 58 of G-ALPHA1, 61 of G-ALPHA2, 63 of G-ALPHA, and 62 of G-BETA being the "highest" points on the floor groove. The G-ALPHA2 and G-BETA domains have a disulfide bridge between positions 11 and 74. The G-ALPHA1 and G-ALPHA domains have a conserved N-glycosylation site at position 86 (N-X-S/T, where N is asparagine, X any amino acid except proline, S is serine and T is threonine), except for HLA-DMB and H2-DMB1. Asparagine 15 of the G-BETA domains also belongs to a conserved N-glycosylation site.

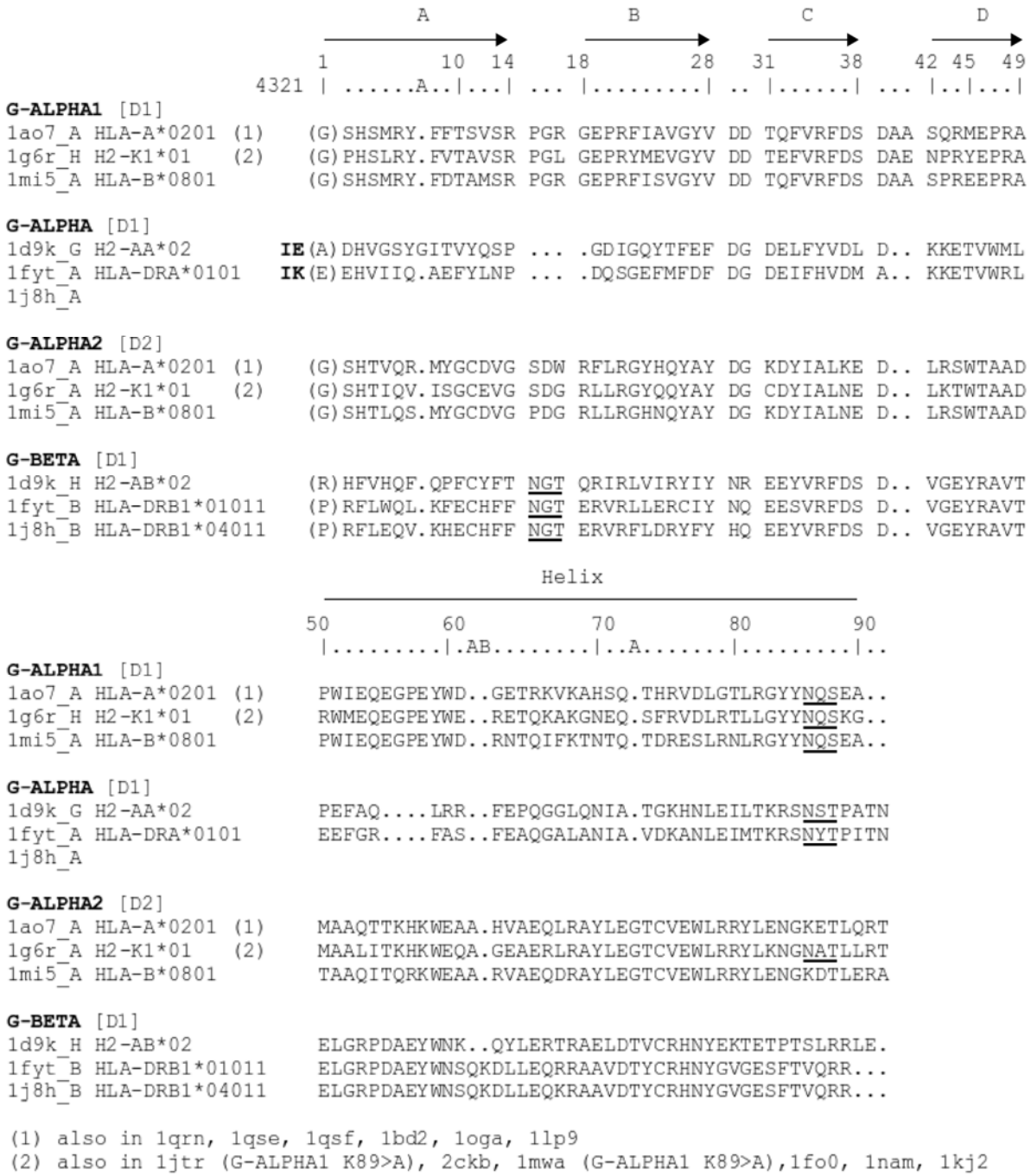


Figure 3. Protein display of the G-DOMAINs found in the TR/pMHC complexes in IMGT/3Dstructure-DB [19], http://www.imgt.org. Amino acid sequences and gaps (shown by dots) are according to the IMGT unique numbering for G-DOMAIN [15]. Amino acids resulting from the splicing with the preceding exon are shown 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.

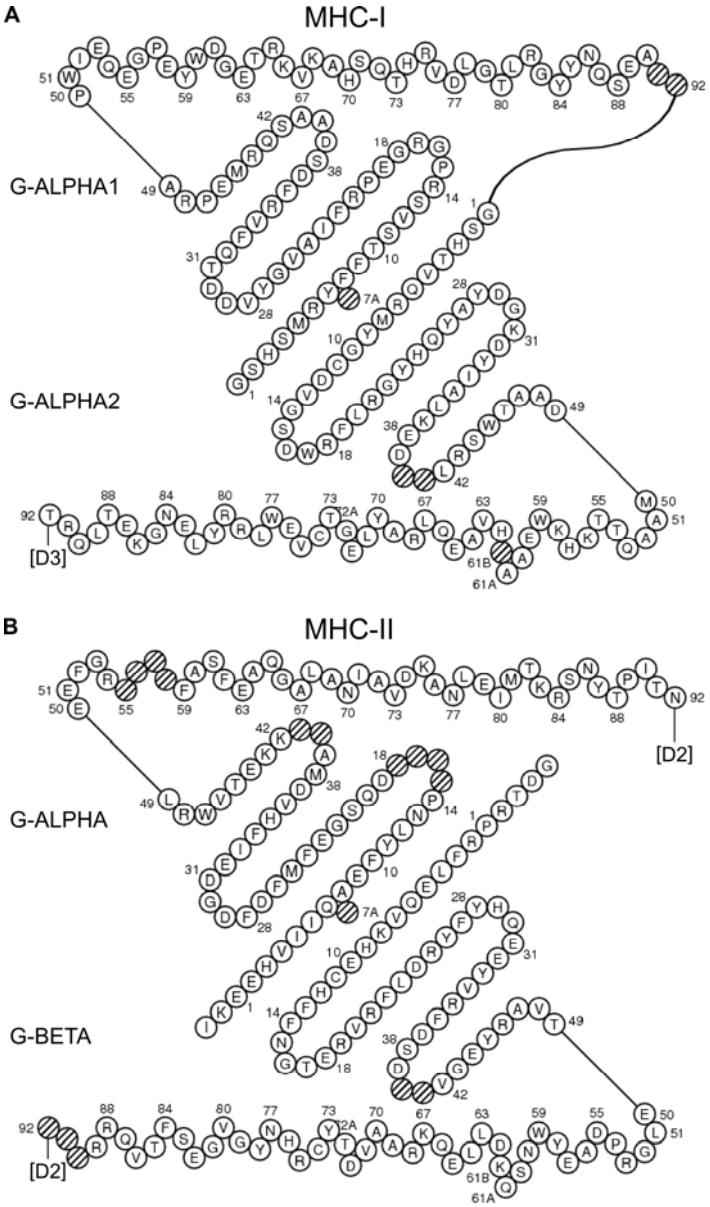


Figure 4. IMGT Collier de Perles of MHC G-DOMAINs. (A) MHC-I G-ALPHA1 and G-ALPHA2 domains from 1ao7 (B) MHC-II G-ALPHA and G-BETA domains from 1j8h (IMGT/3Dstructure-DB [19], http://www.imgt.org). Amino acids positions are according to the IMGT unique numbering for G-DOMAIN [15]. Positions 61A, 61B 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).

4.2. Peptide/MHC interactions: Concept of pMHC contact sites

The 3D structure of the MHC main chain is well conserved and the peptide binding groove specificity is due to side chains physicochemical characteristics. Both MHC-I and MHC-II grooves have pockets where side chains of bound peptides may anchor, 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 conserved properties were needed to bind a peculiar MHC allele (WP2). Peptide/MHC binding prediction and

epitope prediction remain a big challenge that has justified the development of the molecular models in the simulator.

In contrast to previous attempts to define pockets, structural data for defining the IMGT pMHC contact sites (MHC amino acid positions that have contacts with the peptide side chains) take into account the length of the peptides but are considered independently of the MHC class and sequence polymorphisms. This approach was selected as it provides for the first time the possibility to analyse simultaneously pMHC-I and pMHC-II. The goal in D1.4 was to demonstrate the proof of concept by analysing existing data from IMGT/3Dstructure-DB [19] (http://www.imgt.org) and by defining the criteria that were necessary for modelling the pMHC interactions. 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 as relevant to the analysis. 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 [18]. 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. For MHC-II, the contact analysis was performed for the peptide amino acid side chains of the 9 amino acids that were identified as located in the groove.

Eleven IMGT pMHC contact sites were defined (C1 to C11) which can be used to compare pMHC interactions. 1-9 refer 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. Examples of contact sites for a MHC-I binding an 8-mer peptide (1jtr), for a MHC-I binding a 9-mer peptide (1ao7) and for an MHC-II binding the 9 amino acids of a peptide (1j8h) are shown in Fig. 5, 6 and 7, respectively. There are no C2, C7 and C8 contact sites for MHC-I with 8-amino acid peptides (Fig. 5) and no C2 and C7 contact sites for MHC-I with 9-amino acid peptides (Fig. 6). In contrast, for MHC-II, C2 is present but there is no C7 and C8 (Fig. 7).

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.

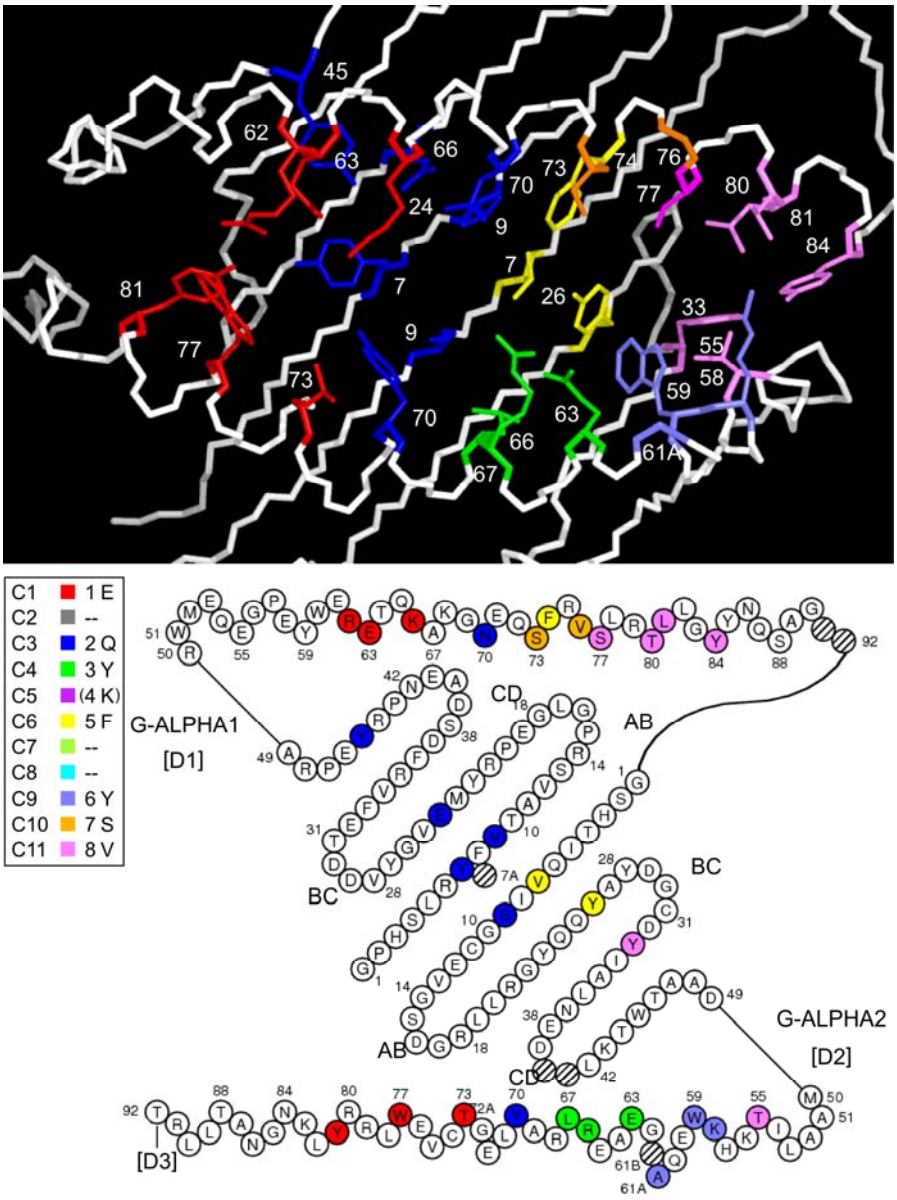


Figure 5. IMGT pMHC contact sites of mouse H2-K1 MHC-I and a 8-amino acid peptide (1jtr). (A) 3D structure of the mouse H2-K1*01 groove. (B) IMGT pMHC contact sites IMGT Collier de Perles. Both views are from above the cleft with G-ALPHA1 on top and G-ALPHA2 on bottom. In the box, C1 to C11 refer to contact sites [18], 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).

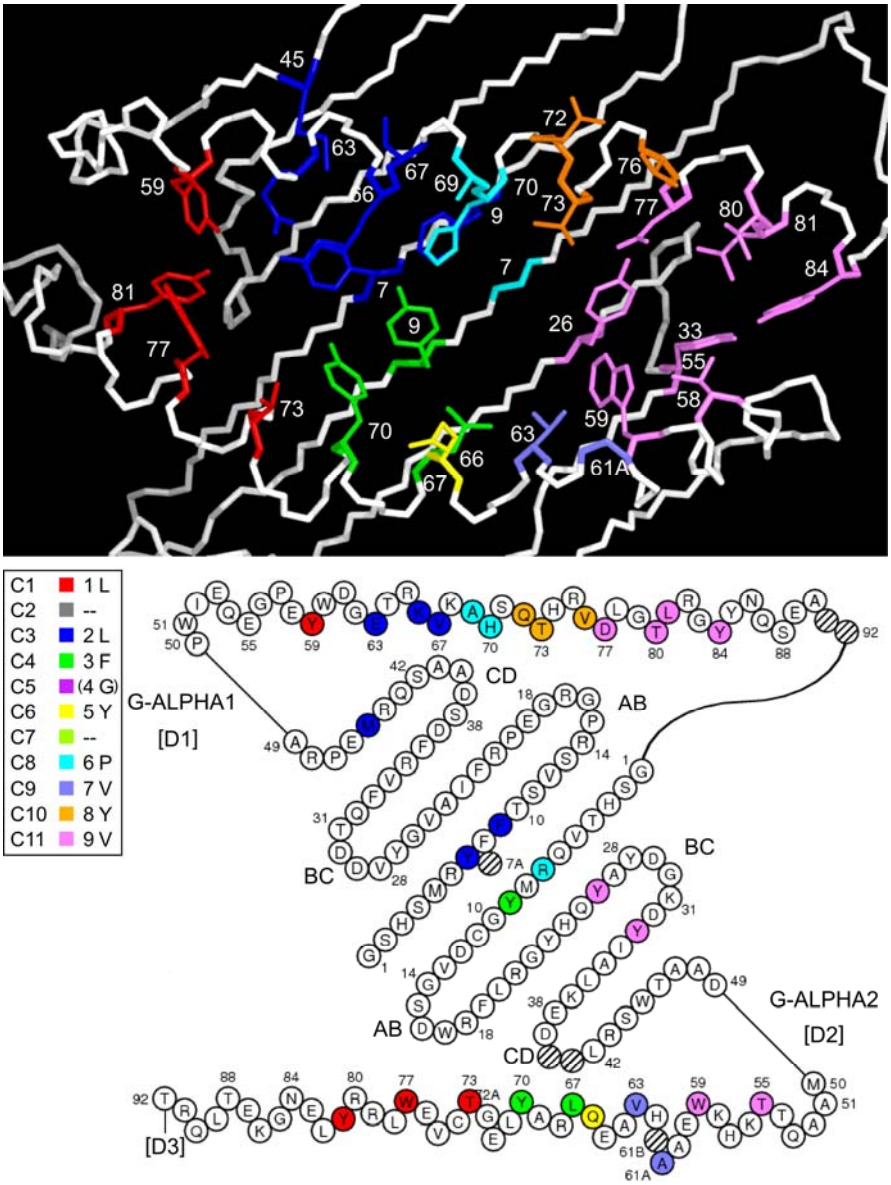


Figure 6. IMGT pMHC contact sites of human HLA-A*0201 MHC-I and a 9-amino acid peptide (1ao7). (A) 3D structure of the human HLA-A*0201 groove. (B) IMGT pMHC contact sites IMGT Collier de Perles. Both views are from above the cleft with G-ALPHA1 on top and G-ALPHA2 on bottom. In the box, C1 to C11 refer to contact sites [18]. 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).

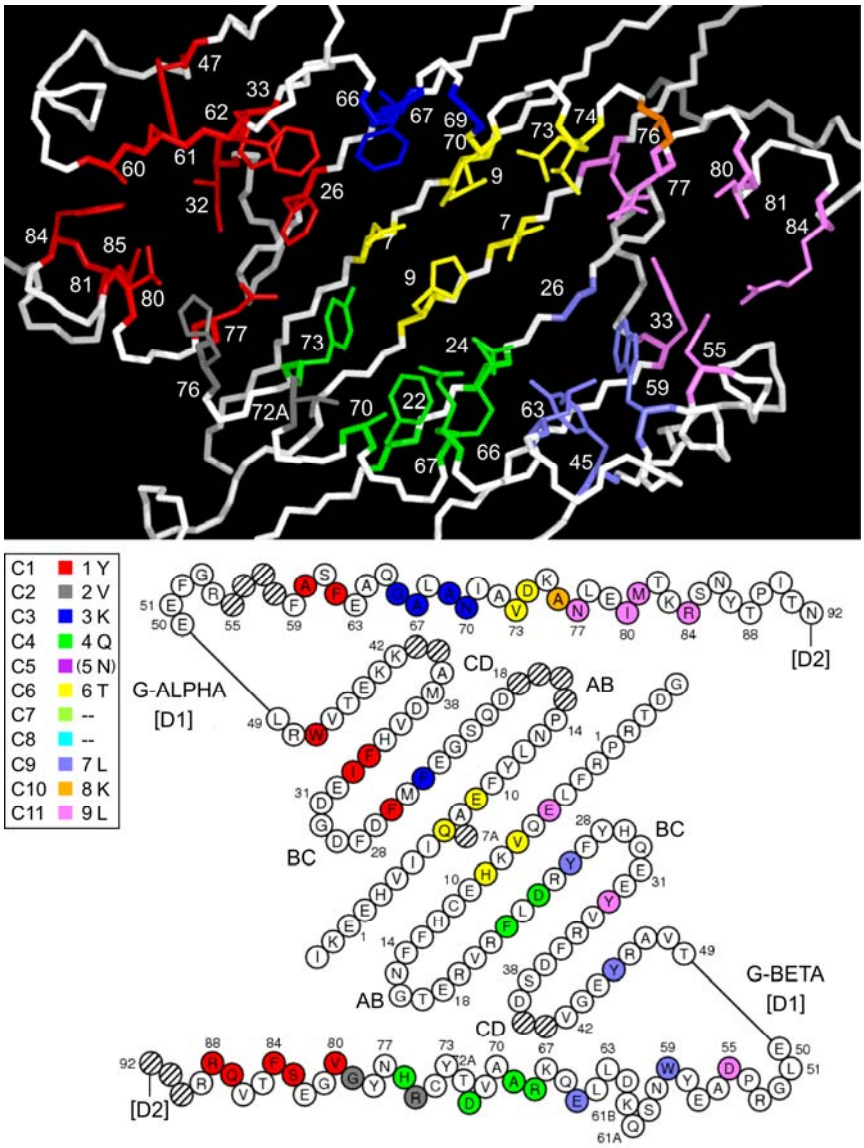


Figure 7. IMGT pMHC contact sites of the human HLA-DRA*0101 and HLA-DRB1*0401 MHC-II and the peptide side chains (9-amino acids located in the groove). (A) 3D structure of the human HLA-DRA*0101 and HLA-DRB1*0401 groove (1j8h). (B) IMGT pMHC contact sites IMGT Collier de Perles. Both views are from above 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 is no C5 and C7 in MHC-I 3D structures with 9-amino acid peptides. There is no C5 in this 3D structure as 5 does not contact MHC amino acids (5N is shown between parentheses in the box).

In order to make these results more widely available for modelling, "IMGT reference pMHC contact sites" were constructed with the analysed data (Table 4). The current "IMGT reference pMHC contact sites" includes results from:

- 30 pMHC-I 3D structures for MHC-I with 8-amino acid peptides
- 74 pMHC-I 3D structures for MHC-I with 9-amino acid peptides
- 10 pMHC-I 3D structures for MHC-I with 10-amino acid peptides
- 44 pMHC-II 3D structures with the 9 amino acids identified as being in the groove.

Table 4. IMGT reference pMHC contact sites. (A) MHC-I. Results for 104 pMHC-I 3D structures (30 with 8-amino acid peptides and 74 with 9-amino acid peptides). (B) MHC-II. Results from 44 pMHC-II 3D structures with 9 amino acids in the groove.

(A) MHC-I

8-amino acid peptides

		G-ALPHA1	G-ALPHA2
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

		G-ALPHA1	G-ALPHA2
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

The “IMGT reference pMHC contact sites” for MHC-I 3D structures with 10-amino acid peptides are not reported in Table 4 as the number of structures available is limited, reflecting the fact that this type of pMHC interaction is found less frequently in vivo. However it fully demonstrates that the concept could also be applied to that peptide length.

“IMGT reference pMHC contact sites” are also available as IMGT Colliers de Perles that will be updated as the number of 3D structures increases. Thus, “IMGT reference pMHC contact sites” demonstrate that the proof of concept is applicable whatever the peptide length, whatever the MHC class (MHC-I or MHC-II), and whatever the species. This is particularly important when modelling involve data sets both from humans and from animal models.

The results for MHC-II go beyond what was expected. Indeed the length of the peptides binding to MHC-II is generally longer (13-18 amino acids) than that of the peptides binding to MHC-I (8-10 amino acids) and prediction tools based on sequences are performing poorly. By helping to predict the nine amino acids of the peptides which are localized in the groove, the “IMGT reference pMHC contact sites” provide structural criteria which allow to improve MHC-II peptide binding prediction. This is a major outcome of D1.4.

Section 5. TR/pMHC interactions

5.1. TR V-DOMAIN

The variable domain (V-DOMAIN) of the TR alpha and beta chains has an immunoglobulin fold, that is an antiparallel β sheet sandwich structure with 9 strands, the A, B, E and D strands being on one sheet, and the G, F, C, C' and C" strands on the other sheet [13]. In order to compare data from different IG and TR variable domain sequences and 3D structures, the IMGT unique numbering for V-DOMAIN has been set up (Deliverable 1.3). This has allowed to provide the frame to the IMGT Protein display and to the IMGT Colliers de Perles for V-DOMAIN. Figure 8 shows the IMGT Protein display of the different V-ALPHA and V-BETA domains found in TR/pMHC used for setting the proof of concept in D1.4. Figure 9 shows, as examples, the IMGT Colliers de Perles of the V-ALPHA and V-BETA domains from 1ao7. 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 [13]. 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 [13]. Compared to the general V-DOMAIN 3D structure, the V-ALPHA domains have a shorter C” strand.

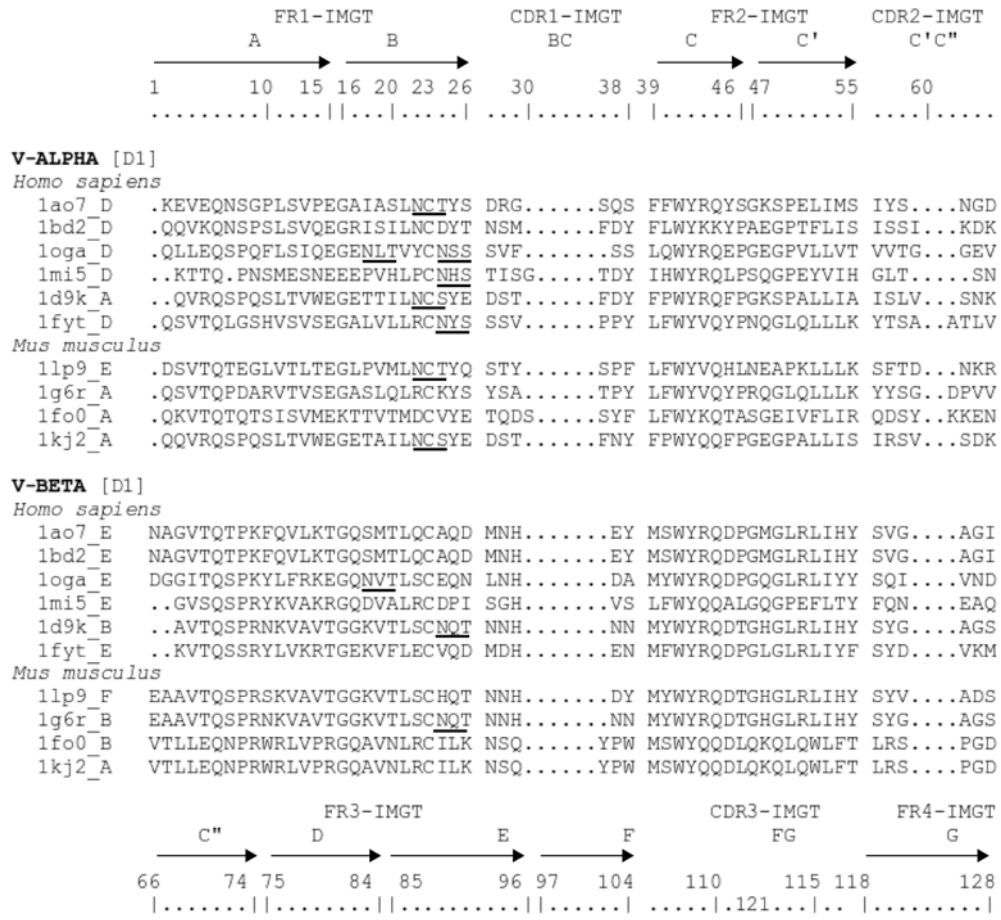


Figure 8. Protein display of the TR V-ALPHA and V-BETA domains found in the TR/pMHC complexes in IMGT/3Dstructure-DB [19], http://www.imgt.org. Amino acid sequences and gaps (shown by dots) are according to the IMGT unique numbering for V-DOMAIN [13]. The three additional positions in the CDR3-IMGT are 111.1, 112.2 and 112.1. Potential N-glycosylation sites are underlined.

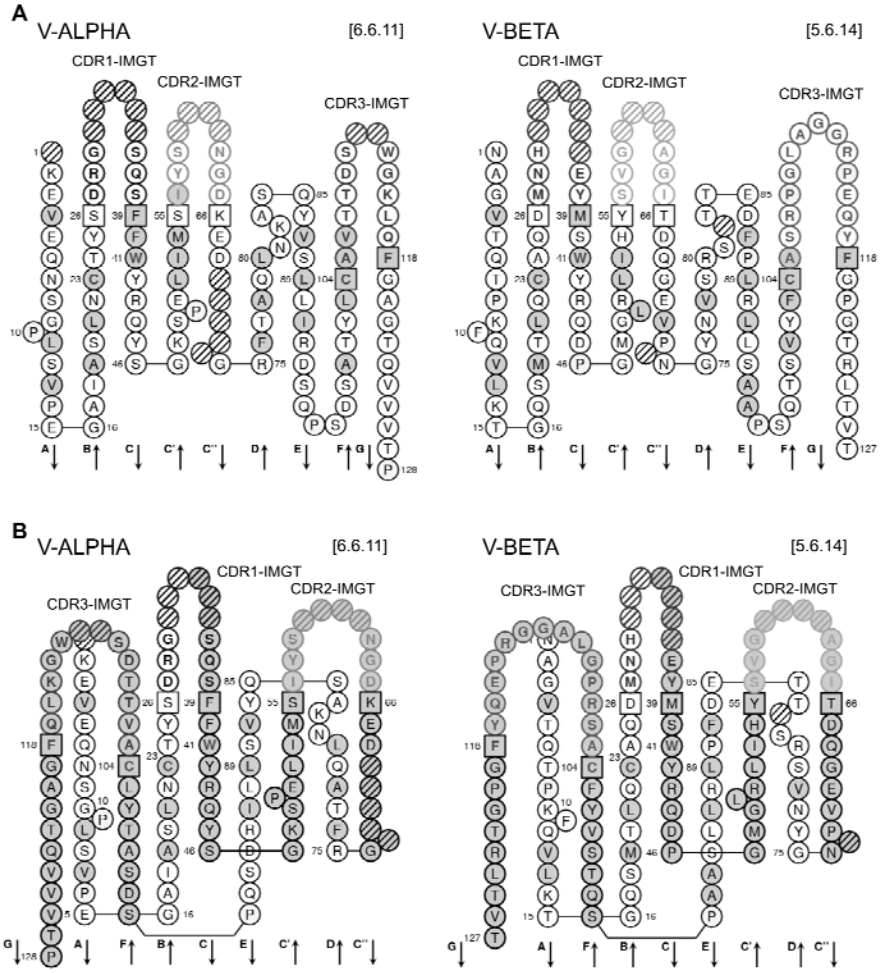


Figure 9. IMGT Collier de Perles of the V-ALPHA and V-BETA domains from 1ao7 (IMGT/ 3Dstructure-DB [19], http://www.imgt.org) (A) on one layer (B) on two layers. Amino acids are shown in the one-letter abbreviation. Hydrophobic amino acids (hydropathy index with positive value) and tryptophan (W) found at a given position in more than 50% of analysed IG and TR sequences are shown. The CDR-IMGTs are limited by amino acids shown in squares, which belong to the neighbouring FR-IMGT and represent anchor positions. Hatched circles correspond to missing positions according to the IMGT unique numbering [13]. Arrows indicate the direction of the β sheets.

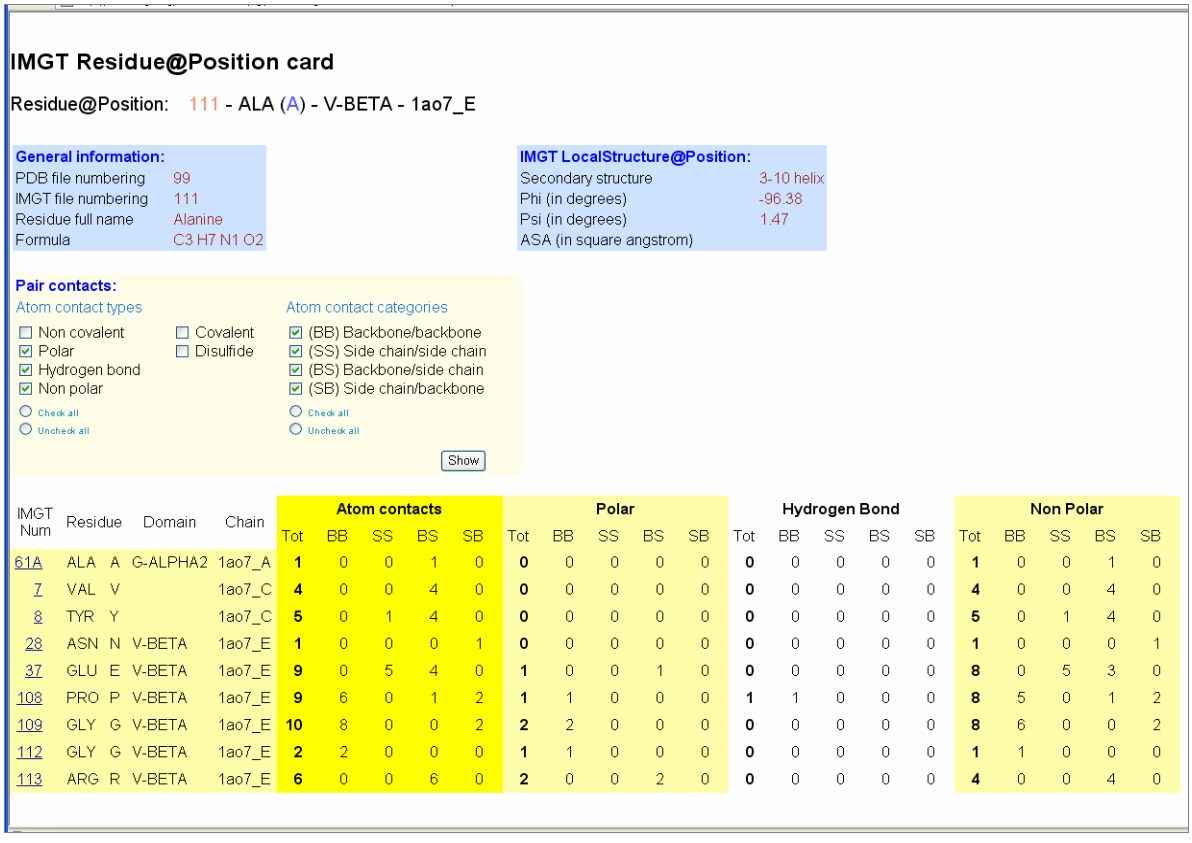
5.2. TR/pMHC interactions: Concept of IMGT Residue@Position

The objective was to set up the criteria for a standardized representation of the interactions between T cell receptor and pMHC. In order to set up the proof of concept, eighteen TR/pMHC structures were analysed. Fourteen structures, 12 TR/ pMHC-I and two TR/pMHC-II, have complete extracellular regions of the α-β TR (TR-ALPHA_BETA comprising variable and constant domains) whereas four structures are Fv variable fragment (FV-ALPHA_BETA comprising only variable domains). The 18 TR/pMHC 3D structures are: 1ao7 (Garboczi et al. 1996) [20], 1qrn, 1qse, 1qsf (Ding et al. 1999) [21], 1bd2 (Ding et al. 1998) [22], 1oga (Stewart-Jones et al. 2003) [23], 1mi5 (Kjer-Nielsen et al. 2003) [24],

1lp9 (Buslepp et al. 2003) [25], 1g6r (Degano et al. 2000) [26], 1jtr, 1mwa (Luz et al. 2002) [27], 2ckb (Garcia et al. 1998) [28], 1fo0 (Reiser, et al. 2000) [29], 1nam (Reiser et al. 2003) [30], 1kj2 (Reiser et al. 2002) [31], 1fyt (Hennecke et al. 2000) [32], 1j8h (Hennecke and Wiley 2002) [33], 1d9k (Reinherz et al. 1999) [34].

As for the antibody/antigen interactions, the TR/pMHC interactions involve primarily the hypervariable loops or complementarity determining regions (CDR) of the V-ALPHA and V-BETA domains. Based on the IMGT unique numbering, the CDR1-IMGT comprises positions 27 to 38, the CDR2-IMGT positions 56 to 65 and the CDR3-IMGT positions 105 to 117 [13]. 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 [9]. Lengths of the CDR-IMGT are shown separated by dots between brackets [13]. For example, for the V-ALPHA of 1ao7, [6.5.11] means that in this domain, 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. The V-ALPHA CDR3-IMGT results from the TRAV12-2–TRAJ24 rearrangement. In the same way, for the V-BETA of 1ao7, [5.6.14] means that in that domain, CDR1-IMGT, CDR2-IMGT and CDR3-IMGT have a length of 5, 6 and 14 amino acids, respectively [13]. The V-BETA CDR3-IMGT results from the TRBV6-5–TRBD2–TRBJ2-7 rearrangement.

In order to analyse the interactions at the amino acid level and to take into account their physicochemical characteristics, the concept of IMGT Residue@Position has been defined in IMGT-ONTOLOGY [18, 19, 40]. It takes into account the position according to the IMGT unique numbering, the amino acid, the domain and the IMGT/3Dstructure-DB chain, for example 111 – ALA(A) – V-BETA – 1ao7_E (Fig. 10). It allows to characterize the interactions (Pair contacts) between domains, and between domains and ligands characterized per atom contact types and atom contact categories.



IMGT Residue@Position card

Residue@Position: 111 - ALA (A) - V-BETA - 1ao7_E

General information:

PDB file numbering	99
IMGT file numbering	111
Residue full name	Alanine
Formula	C3 H7 N1 O2

IMGT LocalStructure@Position:

Secondary structure	3-10 helix
Phi (in degrees)	-96.38
Psi (in degrees)	1.47
ASA (in square angstrom)	

Pair contacts:

Atom contact types: ☐ Non covalent ☐ Covalent ☑ Polar ☐ Disulfide ☑ Hydrogen bond ☑ Non polar — Check all / Uncheck all

Atom contact categories: ☑ (BB) Backbone/backbone ☑ (SS) Side chain/side chain ☑ (BS) Backbone/side chain ☑ (SB) Side chain/backbone — Check all / Uncheck all

Show

IMGT Num	Residue		Domain	Chain	Atom contacts					Polar					Hydrogen Bond					Non Polar				
					Tot	BB	SS	BS	SB	Tot	BB	SS	BS	SB	Tot	BB	SS	BS	SB	Tot	BB	SS	BS	SB
61A	ALA	A	G-ALPHA2	1ao7_A	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0
7	VAL	V		1ao7_C	4	0	0	4	0	0	0	0	0	0	0	0	0	0	0	4	0	0	4	0
8	TYR	Y		1ao7_C	5	0	1	4	0	0	0	0	0	0	0	0	0	0	0	5	0	1	4	0
28	ASN	N	V-BETA	1ao7_E	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1
37	GLU	E	V-BETA	1ao7_E	9	0	5	4	0	1	0	0	1	0	0	0	0	0	0	8	0	5	3	0
108	PRO	P	V-BETA	1ao7_E	9	6	0	1	2	1	1	0	0	0	1	1	0	0	0	8	5	0	1	2
109	GLY	G	V-BETA	1ao7_E	10	8	0	0	2	2	2	0	0	0	0	0	0	0	0	8	6	0	0	2
112	GLY	G	V-BETA	1ao7_E	2	2	0	0	0	1	1	0	0	0	0	0	0	0	0	1	1	0	0	0
113	ARG	R	V-BETA	1ao7_E	6	0	0	6	0	2	0	0	2	0	0	0	0	0	0	4	0	0	4	0

Figure 10. IMGT Residue@Position card based on the IMGT Residue@Position concept.

Section 6. Implementation of the WP1 concepts.

Any domain represented by an IMGT Collier de Perles is characterized by the length of its strands, loops and turns and, for the G type, by the length of its helix [13-15]. The strand, loop, turn or helix lengths (the number of amino acids or codons, that is the number of occupied positions) become crucial information which characterizes the domains. This first feature of the IMGT standardization based on the IMGT unique numbering allowed, for instance, to show that the distinction between the C1, C2, I1 and I2 domain types found in the literature and in the databases to describe the IgSF C type domains is unnecessary and moreover unapplicable when dealing with sequences for which no structural data are known (discussed in [14]).

A second feature of the IMGT standardization is the comparison of cDNA and/or amino acid sequences with genomic sequences, and the identification of the splicing sites, to delimit precisely the domains: a V-LIKE-DOMAIN, a C-DOMAIN, a C-LIKE-DOMAIN, a G-DOMAIN or a G-LIKE-DOMAIN is frequently encoded by a unique exon [13-15]. This IMGT standardization for the domain delimitation explains the discrepancies observed with the generalist UniProt/Swiss-Prot database which identifies domains based on amino acid sequences and does not take into account the genomic information. The IMGT Colliers de Perles also put the question of the leader region. Indeed, the N-terminal end of the first domain of an IgSF or MhcSF chain depends on the proteolytic cleavage site of the leader region (peptide signal) which is rarely determined experimentally. When this site is not known, the IMGT Colliers de Perles start with the first amino acid resulting from the splicing ("Splicing sites" in IMGT Aide-mémoire, http://www.imgt.org). For a IG and TR V-DOMAIN the leader proteolytic site is known (or is extrapolated) and the IMGT Colliers de Perles start with the first amino acid of the V-REGION [8, 9].

The IMGT Colliers de Perles allow a precise visualization of the inter-species differences for the IgSF V and C type domain strands and loops, and MhcSF G type domain strands and helix, even in the absence of 3D structures. This has been applied to the teleost CD28 family members and their B7 family ligands and to the BTLA protein, which belong to the IgSF by their V type and/or C type domains [35, 36]. The IMGT Colliers de Perles are particularly useful in molecular engineering and antibody humanization design based on CDR grafting. Indeed they allow to precisely define the CDR-IMGT and to easily compare the amino acid sequences of the four FR-IMGT (FR1-IMGT: positions 1 to 26, FR2-IMGT: 39 to 55, FR3-IMGT: 66 to 104, and FR4-IMGT: 118 to 128) between the murine and the closest human V-DOMAINs. A recent analysis performed on humanized antibodies used in oncology underlines the importance of a correct delimitation of the CDR regions to be grafted [37].

The IMGT Colliers de Perles also allow a comparison with the IMGT Collier de Perles statistical profiles for the human expressed IGHV, IGKV and IGLV repertoires [38]. These statistical profiles are based on the definition of eleven IMGT amino acid physicochemical characteristics classes which take into account the hydropathy, volume and chemical characteristics of the 20 common amino acids [39] ("Amino acids" in IMGT Aide-mémoire, http://www.imgt.org). The statistical profiles identified positions which are conserved for the physicochemical characteristics: 41 FR-IMGT positions for the human IGHV and 59 FR-IMGT positions for the human IGKV and IGLV at >80% threshold (see Plate 3 in [38]). After assignment of the IMGT Collier de Perles amino acids to the IMGT amino acid physicochemical classes, comparison can be made with the statistical profiles of the human expressed repertoires. This comparison is useful to identify potential immunogenic residues

at given positions in chimeric or humanized antibodies [37] or to evaluate immunogenicity of primate antibodies [39].

IMGT Colliers de Perles are also of interest when 3D structures are available. In IMGT/3Dstructure-DB [19], "IMGT Collier de Perles on 2 layers" are displayed with hydrogen bonds for V type and C type domains. Clicking on a residue in 'IMGT Collier de Perles on one layer' gives access to the corresponding IMGT Residue@Position card which provides the atom contact types and atom contact categories for that amino acid. IMGT Colliers de Perles display the IMGT pMHC contact sites for 3D structures with peptide/MHC (pMHC) complexes [18], which can be compared with the "IMGT reference pMHC contact sites" available in IMGT/3Dstructure-DB [40]. The Residue@Position concept has been crucial for the characterization of the antibody/antigen and TR/pMHC interactions.

The IMGT Colliers de Perles for the V type, C type and G type, based on the IMGT unique numbering, represent therefore a major step forward for the comparative analysis of the sequences and structures of the IgSF and MhcSF domains, for the study of their evolution and for the applications in antibody engineering [37], IG and TR repertoires in autoimmune diseases and leukemias [41], pMHC contact analysis [40], and more generally ligand-receptor interactions involving V type, C type and/or G type domains.

Section 7. ImmunoGrid concepts and modelling of the immune system

The inherent difficulties due to the complexity and diversity of immunogenetics knowledge gave rise to a conceptualization in IMGT-ONTOLOGY which has been developed on an original and unprecedented approach. The axioms of the Formal IMGT-ONTOLOGY or IMGT-Kaleidoscope postulate that the approach to manage biological data and to represent knowledge in biology comprises various facets. The IMGT-ONTOLOGY concepts generated from these axioms have allowed the representation, at the molecular level, of knowledge related to the genome, transcriptome, proteome, genetics and 3D structures. This multi-faceted approach has great potential for multi-scale system biology. Indeed, the IDENTIFICATION, DESCRIPTION and CLASSIFICATION axioms defined in the deliverable D1.2, the NUMEROTATION axiom defined in the deliverable D1.3, the INTERACTION and LOCALIZATION axioms defined in this deliverable D1.4 are valid, not only for molecules, but also for cells, tissues, organs, organisms or populations. In addition, the ORIENTATION and OBTENTION axioms (in development) will allow the integration of the time and space concepts and the follow-up of the components and their changes of states and properties, as well as the definition and characterization of processes, functions and activities. Thus, IMGT-ONTOLOGY represents, by its 8 axioms (including the INTERACTION axiom developed in this ImmunoGrid project) and the concepts generated from them, a paradigm for the elaboration of ontologies in system biology which requires to identify, to describe, to classify, to numerotate, to localize, to orientate and to determine the interaction, obtaining and evolution of biological knowledge from molecule to population, in time and space.

The concepts of IMGT-ONTOLOGY are available, for the users of the ImmunoGrid simulator and for the biologists in general, in natural language in IMGT Scientific chart

(http://www.imgt.org), and have been formalized for programming purpose in IMGT-ML (XML Schema). IMGT-ONTOLOGY is being implemented in Protégé and OBO-Edit to facilitate the export in formats such as OWL, and to link, whenever possible, the concepts of IMGT-ONTOLOGY to those of other ontologies in biology such as the Gene Ontology (GO) (http://www.geneontology.org/), and in immunology, such as the Immunome Epitope database and Analysis Resource (IEDB) (http://www.immuneepitope.org/) and other Open Biomedical Ontologies (OBO) (http://obo.sourceforge.net).

The concepts of IMGT-ONTOLOGY are currently used for the exchange and the sharing of knowledge in very diverse fields of research at the molecular level: (i) fundamental and medical research (repertoire analysis of the IG antibody sites and of the TR recognition sites in normal and pathological situations such as autoimmune diseases, infectious diseases, AIDS, leukemias, lymphomas, myelomas), (ii) veterinary research (IG and TR repertoires in farm and wild life species), (iii) genome diversity and genome evolution studies of the adaptive immune responses, (iv) structural evolution of the IgSF and MhcSF proteins, (v) biotechnology related to antibody engineering (scFv, phage displays, combinatorial libraries, chimeric, humanized and human antibodies), (vi) diagnostics (clonalities, detection and follow-up of residual diseases) and (vii) therapeutical approaches (grafts, immunotherapy, vaccinology).

IMGT-ONTOLOGY has represented a key component in the elaboration and setting up of standards of the European ImmunoGrid project (http://www.immunogrid.org/) whose aim is to define the essential concepts for modelling of the immune system. IMGT-ONTOLOGY will allow interactions with IMGT®, the global reference in immunogenetics and immunoinformatics. This will further strengthen the importance of standardization in pharmaceutical and clinical research, as demonstrated by companies which, in Europe (SANOFI-AVENTIS, Institut Pierre Fabre...), in Japan (ASTELLAS, AGENSYS...) and in the USA (CENTOCOR Johnson and Johnson, MERCK, AMGEN...), have IMGT® licences and contracts.

As the same axioms can be used to generate concepts for multi-scale level approaches, the Formal IMGT-ONTOLOGY represents a paradigm for system biology ontologies, which need to identify, to describe and to classify objects, processes and relations at the molecule, cell, tissue, organ, organism or population levels.

These axioms are particularly important as they represent the crucial step of the ImmunoGrid approach, linked to the specificity of the immune response (antigen recognition, specificity antigen-receptor, B cell epitope and T cell epitope characterization, peptides used in vaccinology and immunotherapy, humanized antibodies used in cancerology, etc). Immune responses at the cellular level and organism level depend on this molecular level, whose component interactions trigger the whole cascade of events.

Section 8. References

1. Giudicelli, V. and Lefranc, M.-P. (1999) Ontology for Immunogenetics: IMGT-ONTOLOGY. Bioinformatics 15:1047-1054.
2. 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., Monod, Y.M., Duprat, E., Kaas, Q., Pommié, C., Chaume, D. and Lefranc, G. (2004) IMGT-ONTOLOGY for immunogenetics and immunoinformatics. *In Silico* Biol. 4:17-29.

3. Lefranc, M.-P., Clément, O., Kaas, Q., Duprat, E., Chastellan, P., Coelho, I., Combres, K., Ginestoux, C., Giudicelli, V., Chaume, D. and Lefranc, G. (2005) IMGT-Choreography for Immunogenetics and Immunoinformatics. *In Silico* Biol. 29:185-203.
4. Duroux, P., Kaas, Q., Brochet, X., Lane, J., Ginestoux, C., Lefranc, M.-P. and Giudicelli, V. (2008) IMGT-Kaleidoscope, the formal IMGT-ONTOLOGY paradigm. Biochimie 90:570-583.
5. Lefranc, M.-P. (2008) IMGT-ONTOLOGY, IMGT® databases, tools and Web resources for Immunoinformatics. In: Immunoinformatics (Schoenbach C., Ranganathan S. and Brusic V. eds.), Immunomics Reviews, Springer, New York, USA, pp.1-18.
6. Lefranc, M.-P. (2005) IMGT, the international ImMunoGeneTics information system®: a standardized approach for immunogenetics and immunoinformatics. Immunome Res. 2005 Sep 20;1:3.
7. Lefranc, M.-P., Giudicelli, V., Ginestoux, C., Jabado-Michaloud, J., Folch, G., Bellahcene, F., Wu, Y., Gemrot, E., Brochet, X., Lane, J., Regnier, L., Ehrenmann, F., Lefranc, G. and Duroux, P. (2009) IMGT®, the international ImMunoGeneTics information system®. Nucl. Acids Res. 37:D1006-D1012.
8. Lefranc, M.-P. and Lefranc, G. (2001) The Immunoglobulin FactsBook. London: Academic Press, 458 pages.
9. Lefranc, M.-P. and Lefranc, G. (2001) The T cell receptor FactsBook. London: Academic Press, 398 pages.
10. Giudicelli, V., Chaume, D. and Lefranc, M.-P. (2005) IMGT/GENE-DB: a comprehensive database for human and mouse immunoglobulin and T cell receptor genes. Nucl. Acids Res. 33:D256-D261.
11. Lefranc, M.-P. (2008) WHO-IUIS Nomenclature Subcommittee for Immunoglobulins and T cell receptors report August 2007, 13th International Congress of Immunology, Rio de Janeiro, Brazil. Dev. Comp. Immunol. 32:461-463.
12. Lefranc, M.-P. (2007) WHO-IUIS Nomenclature Subcommittee for Immunoglobulins and T cell receptors report. Immunogenetics 59:899-902.
13. 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.
14. 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.
15. 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.
16. Lefranc, M.-P., Giudicelli, V., Regnier, L. and Duroux, P. (2008) IMGT®, a system and an ontology that bridge biological and computational spheres in bioinformatics. Brief. Bioinform. 9:263-275.
17. Kaas, Q. and Lefranc, M.-P. (2007) IMGT Colliers de Perles: standardized sequence-structure representations of the IgSF and MhcSF superfamily domains. Curr. Bioinformatics 2:21-30.

18. Kaas, Q. and Lefranc, M.-P. (2005) T cell receptor/peptide/MHC molecular characterization and standardized pMHC contact sites in IMGT/3Dstructure-DB. *In Silico* Biol. 5:505-528.
19. 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. Nucl. Acids Res. 32:D208-D210.
20. 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.
21. 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.
22. 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.
23. 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.
24. 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 αβ T cell receptors in antiviral immunity. Immunity 18:53-64.
25. Buslepp, J., Wang, H., Biddison, W.E., Appella, E. and Collins, E.J. (2003) A correlation between TCR Vα docking on MHC and CD8 dependence: implications for T cell selection. Immunity 19:595-606.
26. 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.
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(β) Interactions. J. Exp. Med. 195:1175-1186.
28. 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.
29. 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.
30. 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.
31. 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 CDR3β loop undergoes conformational changes of unprecedented magnitude upon binding to a peptide/MHC class I complex. Immunity 16:345-354.
32. Hennecke, J., Carfi, A. and Wiley, D.C. (2000) Structure of a covalently stabilized complex of a human αβ T-cell receptor, influenza HA peptide and MHC class II molecule, HLA-DR1. EMBO J. 19:5611-5624.

33. 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.
34. 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.
35. Garapati, V.P. and Lefranc, M.-P. (2007) IMGT Colliers de Perles and IgSF domain standardization for T cell costimulatory activatory (CD28, ICOS) and inhibitory (CTLA4, PDCD1 and BTLA) receptors. Dev. Comp. Immunol. 31:1050-1072.
36. Hansen, J.D., Pasquier, L.D., Lefranc, M.-P., Lopez, V., Benmansour, A. and Boudinot, P (2009). The B7 family of immunoregulatory receptors: A comparative and evolutionary perspective. Mol. Immunol. 46:457-72.
37. Magdelaine-Beuzelin, C., Kaas, Q., Wehbi, V., Ohresser M., Jefferis R., Lefranc M.-P. and Watier H. (2007) Structure-function relationships of the variable domains of monoclonal antibodies approved for cancer treatment. Crit. Rev. Oncol./Hematol. 64:210-225.
38. Pommié, C., Levadoux, S., Sabatier, R., Lefranc, G. and Lefranc, M.-P. (2004) IMGT standardized criteria for statistical analysis of immunoglobulin V-REGION amino acid properties. J. Mol. Recognit. 17:17-32.
39. Laffly, E., Danjou, L., Condemine F., Vidal D., Drouet E., Lefranc M.-P., Bottex C. and Thullier P. (2005) Selection of a macaque Fab with human-like framework regions, high affinity, and that neutralizes the protective antigen (PA) of *Bacillus anthracis*. Antimicrob. Agents Chemother. 49:3414-3420.
40. Kaas, Q., Duprat, E., Tourneur, G. and Lefranc M.-P. (2008) IMGT standardization for molecular characterization of the T cell receptor/peptide/MHC complexes. In: Immunoinformatics (Schoenbach C., Ranganathan S. and Brusic V. eds.), Immunomics Reviews, Springer, New York, USA, chap. 2, pp.19-49.
41. Belessi, C.J., Davi, F.B., Stamatopoulos, K.E. et al. (2006) IGHV gene insertions and deletions in chronic lymphocytic leukemia: “CLL-biased” deletions in a subset of cases with stereotyped receptors. Eur. J. Immunol. 36:1963-1974.