NEW STRATEGIES IN IDENTIFICATION & QUANTIFICATION OF MINOR HISTOCOMPATIBILITY ANTIGENS

CHOPIE HASSAN

New Strategies in Identification & Quantification of Minor Histocompatibility Antigens

ISBN: 978-94-6259-752-5

COVER: Chopie Hassan

LAYOUT: Chopie Hassan

PRINT: IpsKamp Drukkers

FINANCIAL SUPPORT

The work described in this thesis was financially supported by the Landsteiner Foundation for Blood Transfusion Research.

New Strategies in Identification & Quantification of Minor Histocompatibility Antigens

Proefschrift

ter verkrijging van

de graad van doctor aan de Universiteit Leiden

Op gezag van Rector Magnificus prof. mr. C.J.J.M. Stolker

volgens besluit van het College voor Promoties

te verdedigen op 16 september 2015

Klokke 15:00 uur

door

Chopie Hassan

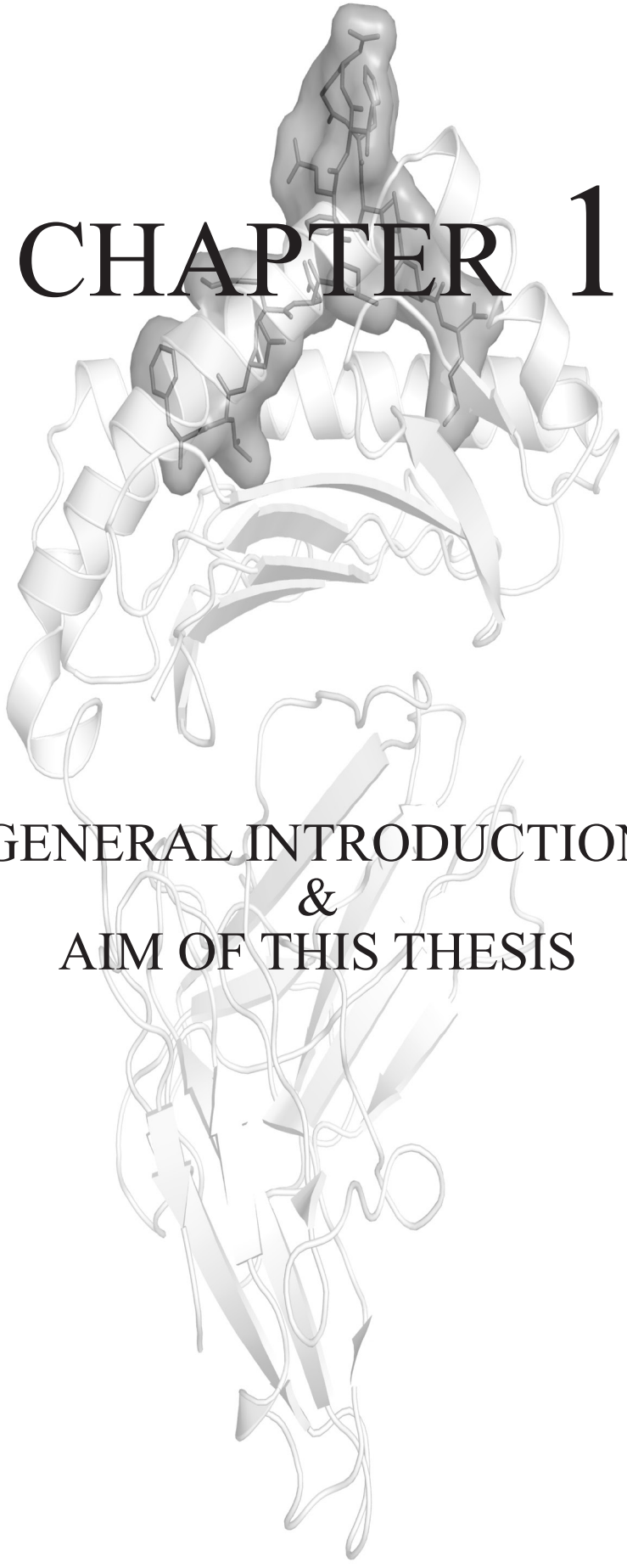
geboren te Suleimania (Irak) in 1971

PROMOTIECOMMISSIE

Promotoren:	Prof. Dr. J. H. F. Falkenburg Prof. Dr. F. Koning
Co-promotor:	Dr. P. A. van Veelen
Overige leden:	Prof. Dr. F. H. J. Claas Prof. Dr. S. H. van der Burg Dr. C. A. C. M. van Els (RIVM)

Aan Rebwar & Rawen

CONTENTS



CHAPTER 1

GENERAL INTRODUCTION
&
AIM OF THIS THESIS

Based on:

Chopie Hassan
Peter A. van Veelen

1

GENERAL INTRODUCTION
&
AIM OF THIS THESIS

GENERAL INTRODUCTION

Immunotherapy and the role of the minor histocompatibility antigens

Allogeneic hematopoietic stem cell transplantation (HSCT)

Allogeneic hematopoietic stem cell transplantation (HSCT) has emerged as immunotherapy and is the main curative treatment of hematologic malignancies and disorders [1-3]. It is also applied as a rescue procedure to replace the damaged hematopoietic stem cells from the patient, after chemotherapy and radiotherapy, and generate a new healthy hematopoietic system with the donor stem cells in the patient. The donor T cells, transplanted with the stem cells into the patient, mediate immune reactions in the patient and induce graft versus leukemia (GVL) [4]. There is evidence that genetic disparity between donor and the patient is required for a significant GVL effect [5]. The main complication of this immunotherapy is the induction of graft versus host disease (GVHD) [6,7]. It is a major source of morbidity and mortality following HSC administration. Reducing its incidence and severity is a major challenge [8]. T cell depleted stem cell transplantation has been applied as an option to reduce the severity of GVHD while the incidence of relapse is increased in these patients [9,10] and is reviewed in [11]. Therefore, donor lymphocyte infusion (DLI) has been administrated several months after HSCT to restore the GVL response. The lymphocytes are derived from the same donor and infused into the patient who received the T cell depleted stem cell transplantation [12-14]. DLI can produce complete remissions in a high number of chronic myeloid leukemia (CML) patients while in other cases additional chemotherapy or antibody administration may be required to achieve complete remission [5]. Improving the outcome of HSCT remains one of the highest priorities in immunotherapy and this can be achieved by finding a strategy to increase the rate of GVL while preventing GVHD [5,11].

Graft versus Leukemia (GVL) and Graft versus host disease(GVHD)

The key to GVL and GVHD in an HLA matched donor/patient pair is minor histocompatibility antigens (MiHA). MiHA are polymorphic peptides, derived from endogenously processed cellular polymorphic proteins and presented on the cell surface by human leukocyte antigen (HLA)- class I and class II molecules. HLA-class I molecules are expressed on the surface of all eukaryotic cells, presenting peptides for $CD8^+$ T cells. After allogeneic HSCT of an HLA matched donor/patient pair the donor T cells recognize MiHA presented in HLA molecules on the surface of malignant cells of the patient and induce GVL (Figure 1). In 1976 for the first time the involvement of minor histocompatibility antigens (MiHA) was reported in hematopoietic stem cell transplantation [15]. Since, the recognition of MiHA by T cells has been described by different

groups. The distribution of MiHA determines the effect of this therapy. If the MiHA is presented only on the hematopoietic cells, including the malignant cells, the donor T cells recognize the malignant cells inducing GVL, while, if the MiHA are also presented on the surface of non-hematopoietic cells of the patient the donor T cells attack the healthy cells of the patient , which is called GVHD (Figure 1). GVHD is still the major life-threatening complication of this therapy. The cells of liver, skin, gut, and lungs are most prone to GVHD [16].

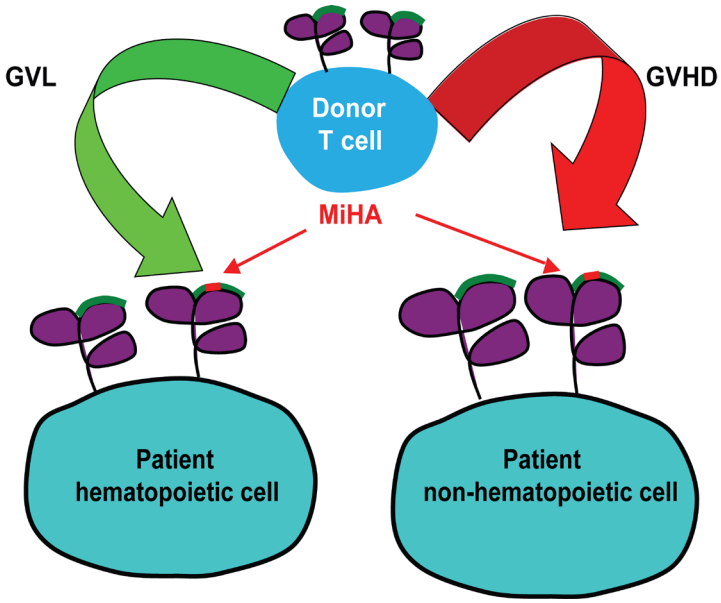


Figure 1. Schematic overview of graft versus leukemia (GVL) and graft versus host disease (GVHD). The donor T cells recognize MiHA on the patient cells. Note; the small difference (usually just one amino acid, indicated in red in some of the presented peptides (in green).

Comprehensive knowledge of MiHA would greatly improve allogeneic HSCT. In addition, the identification of MiHA presented only on hematopoietic cells may allow infusion of T cells recognizing only hematopoietic cells from the patient thereby increasing the likelihood of specific GVL response and decreasing the incidence of GVHD. Clearly, antigen processing plays a crucial role in the presentation of MiHA.

HLA molecules and antigen processing

Human leukocyte antigen (HLA) molecules

HLA molecules are divided in two classes: HLA-class I and HLA-class II. HLA-class I has six different isotypes, namely HLA-A, HLA-B, HLA-C, HLA-E, HLA-F and HLA-G. HLA-class II molecules are termed HLA-DR, HLA-DQ and HLA-DP. The HLA-class I molecule consists of an α chain, comprising domains α 1, α 2 and α 3, and β2¬-microglobulin. The latter is necessary for the stability of the HLA-class I molecule,

while the α3 domain is membrane-spanning. HLA-A, HLA-B, and HLA-C are highly polymorphic and the variability is most pronounced in the α1 and α2 domains, which form the peptide binding groove [17,18]. It consists of two α helices forming a wall on each side and eight β-pleated sheets forming a floor (Figure 2A and 2B). The most polymorphic residues in the groove interact with amino acids in certain positions within the peptide sequence. Most of these peptides that bind a specific isoform of an HLA-class I molecule have the same amino acid or chemically similar amino acids at particular positions. These amino acids are called anchor residues, and the pattern of these preferred amino acids in the peptides is called the peptide binding motif [19,20]. Not all theoretically possible peptide stretches with the correct binding motif will eventually bind to HLA molecules, because of the complexity of the intracellular processing. It has been reported that only 0.1% of the peptides survive intracellular processing to eventually bind to the available HLA molecules and become presented on the cell surface [21].

HLA-class I molecules can now be produced by in vitro folding of recombinant HLA α chain and β2m in

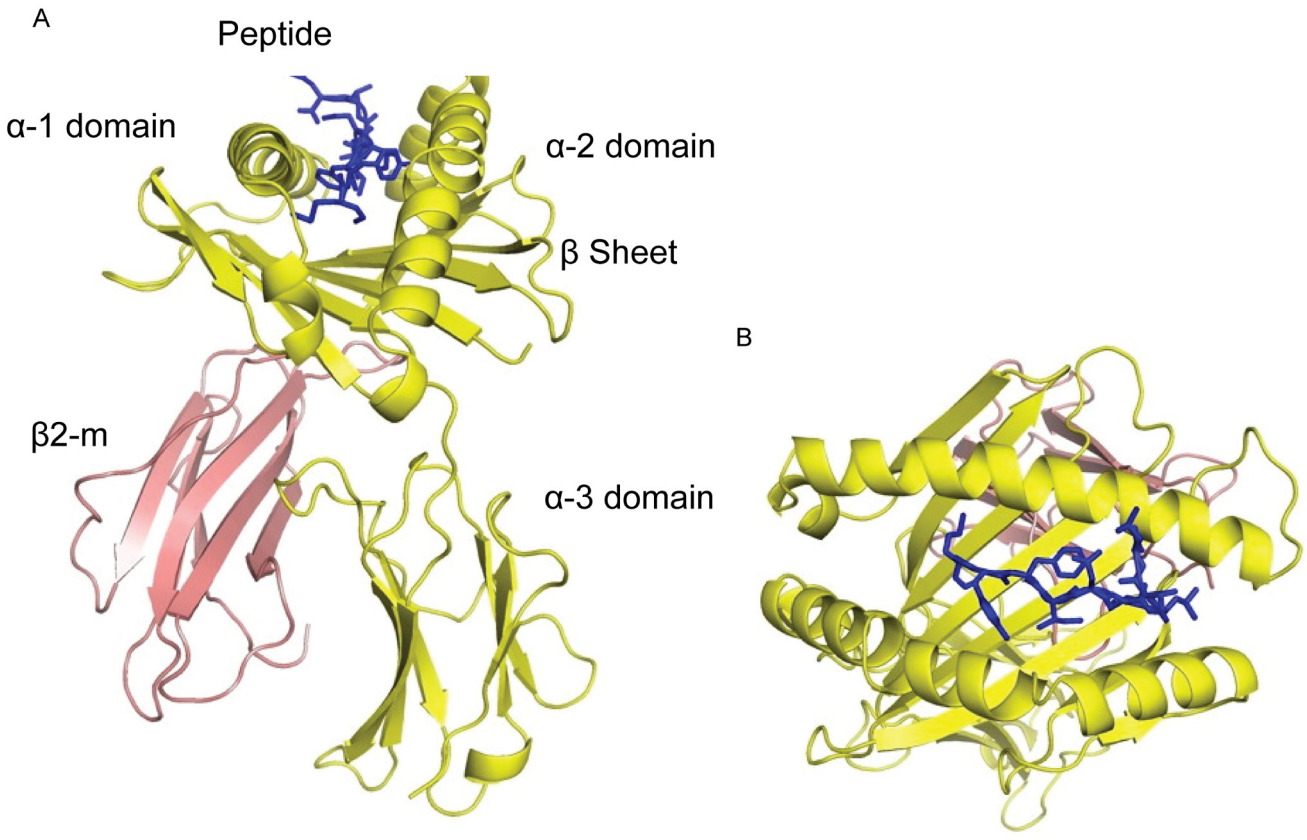


Figure 2. HLA class I molecule consist of alpha (α) chains and β2-microgobulin (β2-m). The trans membrane domain (α-3domain) bound non-covalently to the β2-m. (A) is the side view of the HLA- class I molecule and (B) is the top view which shows the binding grove of the peptides in HLA-class I molecules. The figure adapted from the links below.
http://nfs.unipv.it/nfs/minf/dispense/immunology/lectures/files/structure_mhc.html

combination with the peptide of interest [22]. The assembled molecules are called monomers. The monomers can be utilized in different applications, for instance production of multimers (tetramers), which serve as a platform for peptide specific T cell recognition [23,24] and (Chapter 4). Tetramers are indispensable tools to fish out selected peptide-specific T cells from donor blood. Monomers can also be used for absolute quantification of HLA-presented peptides, and determination of the yield of peptide elution experiments (Chapter 5), by addition of isotopically labeled peptides during the monomer assembly.

Antigen processing and presentation

Classically, HLA-class I associated peptides (antigens) are generated inside the cell by multi step process from endogenous proteins, or e.g. of viral proteins. Proteins can be degraded by various enzymes, the most prominent being the proteasome. In addition, nardilysin and/or thimet oligopeptidase (TOP), insulin degrading enzyme (IDE) can also play roles, in combination with other endo or exoproteases, to produce peptides from 4 to 20 amino acids long [21,25]. The peptides are subsequently transported to the ER via an ATP dependent transporter for antigen processing, TAP, comprising TAP1 and TAP2 [26,27]. Longer peptides will be further trimmed by endoplasmic reticulum aminopeptidase (ERAP) to a size that can fit the binding groove of HLA-class I molecules [28]. In addition to TAP, other molecules as Tapasin [29], Calreticulin (CRT) [30] and ERp57 [31] are involved in the loading of the peptides into HLA-class I molecules [32,33]. In the ER, peptides that have survived the intracellular processing and that have the correct binding motif can be loaded in HLA-class I molecules (Figure 3). The complex of the HLA-class I molecule and the peptide is

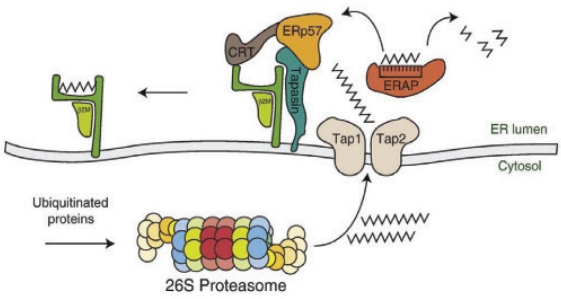


Figure 3. Schematic overview of antigen processing pathway for HLA-class I molecules from endogenous proteins. Endogenous proteins are degraded by the proteasome into smaller units called peptides and transported via ATP dependent antigen transporter (TAP) to endoplasmic reticulum (ER). In ER the peptides bind to HLA molecules and transported via Golgi apparatus to the cell membrane to be presented to $CD8^+$ T cells. Adopted from [34].

transported to the cell surface through the secretory pathway via the Golgi apparatus to be presented on the cell surface for $CD8^+$ T cells.

T cell recognition of a peptide associated with HLA-class I molecule

The $CD8^+$ T cell receptor (TCR) consists of two different polypeptide chains: α chain and β chain. These two chains have a variable region (V region) and a conserved region (C region). The V region is involved in antigen recognition of the HLA-class I- peptide complex [35]. All T cells undergo negative and positive selection in thymus. T cells with intermediate or low avidity against self-antigens are tolerated. High avidity T cells against self-antigens will be deleted while the high avidity T cells against non self (foreign) antigens will be tolerated. These selection steps are crucial to avoid autoimmunity and maintain responsiveness to infectious diseases, respectively [36-39]. T cells are generally tolerant only to HLA-restricted peptides (ligands) presented during T-cell development in the thymus [40]. Presentation of an altered repertoire of HLA class I binding peptides will be perceived as foreign and will trigger $CD8^+$ T-cell responses. In this respect an interesting disturbance of the immune balance was observed in some patients, carrying HLA-

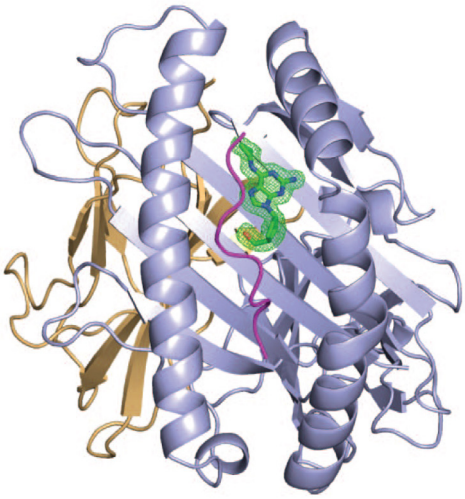


Figure 4. Top View of the HLA-B*5701-abacavir (green) and peptide (purple) complex structure. adapted from [42].

B*5701, who displayed hypersensitivity to treatment with the antiviral drug Abacavir. The hypersensitivity was found to be related to binding of Abacavir in the binding groove of the HLA-molecule, thereby altering the presented peptide repertoire [41-43], see Figure 4.

In summary, $CD8^+$ T cells can recognize previously not presented (new) HLA-presented peptides, such as derived from viruses, post translationally modified peptides, endogenous peptides bound to the altered HLA binding grooves due to a small reactive molecules from drug metabolites [42] and polymorphic peptides derived from endogenous proteins (MiHA) after allogeneic HSCT [6,43-45].

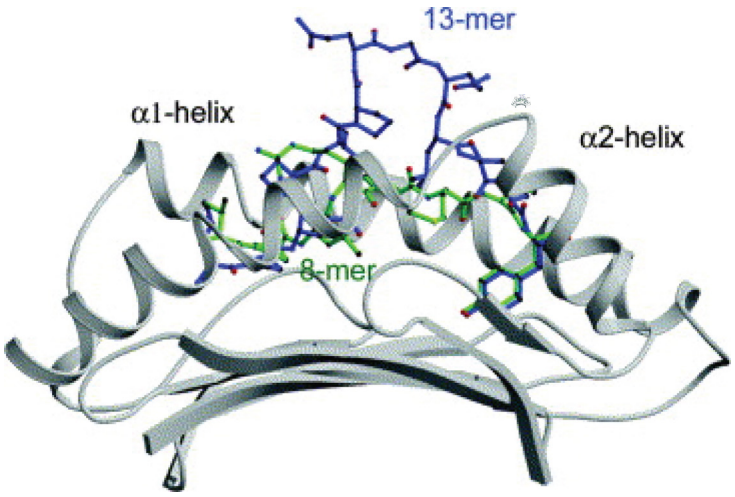


Figure 5. A super-bulged 13mer EBV antigen BZLF1 (52LPEPLPQGQLTAY6) (blue) in the antigen-binding cleft of HLA B35 in comparison with the 8 mer HIV1 Nef (75VPLRPMTY82) peptide (green). Adopted from [49].

Peptides associated with HLA- class I molecules

Generally HLA-class I associated peptides have a length of 8-11 amino acids, typically 9 [46]. However, longer peptides have also been reported. Because of the longitudinal restriction of the binding groove, longer peptides have to bulge out of the groove, see e.g. the 13 mer (viral epitope) bound to HLA-B*3508 molecule, see Figure 5 [47,48].

In addition, the synthetic elongation of known T cell epitopes between the P2 and PΩ anchors demonstrated binding up to at least 25 mers to some HLA-class I molecules, by so-called superbulging [50]. There is also evidence that N-terminally extended peptides can bind to HLA-B*35 [51]. Until now, naturally processed and presented peptides up to 14 amino acids have been published to bind HLA-A*0201 molecules [52]. From analysis of the eluted peptides we could demonstrate that also longer naturally processed peptides can be presented by HLA-A2 molecules (Chapter 7). In addition to peptide length, the binding motif is another criterion for binding to HLA-class I molecules. The two most important positions are position 2 (P2) and the carboxyl terminal (PΩ) in the peptide sequence. At P2 and PΩ one (or a few) amino acid (anchor residue) is dominant per HLA-class I molecule [53-55]. Furthermore, auxiliary anchor residues have been described to enhance the binding of the peptides to HLA-class I molecule as P3 and P5 [55,56].

The peptides (ligands) associated with HLA-class I molecules derived from endogenous proteins originate from different compartments in the cell (e.g. nucleus, cytoplasm, plasma membrane, etc.) representing every protein [57,58]. The presented peptides can be derived either from mature proteins, degrading due to age, due to unfolding, or from defective ribosomal products (DRiPs). The DRiPs/ SLiPs or rapidly degraded polypeptides (RDPs) are considered important sources of peptides. DRiPs consist of prematurely terminated polypeptides and misfolded polypeptides produced from translation of mRNAs in the proper reading

frame [59-61]. The turnover kinetics of the MHC peptides confirm the presence of DRiP and SLiP-derived peptides in MHC molecules [62]. Recently, it has been reported that specific groups of proteins, such as ribosomal and T complex protein 1, contributed DRiPs to the HLA ligandome [61]. In addition the alteration in cellular metabolism resulted in changes in peptides presented in HLA-class I molecules [63]. It has been reported that the peptides presented in HLA-class I molecules are preferentially derived from transcripts bearing miRNA response elements [64]. Furthermore, also post-translationally modified peptides can be presented in HLA-class I molecules. These modified peptides can induce T cell responses. The common example for modifications of the peptides is phosphorylation [65-69]. Phosphorylated peptides can be associated with deregulated signalling in cancer cells, and could be used as targets for immunotherapy in the

Table 1. Summary of the number of HLA-class I ligands identified and published by different groups regardless of the selection criteria used in each study.

HLA type	Nr. of Ligands	Cell source	Ref.
HLA-A2 & B7	3000	human cancer cell line	72
HLA-B*2705	1268	human chondrosarcoma cells	73
HLA-B*2705	200	human lymphoid cell line transfected with B27-C1R	74
HLA-A*6801	816	human lymphoid cell line	75
HLA-A2	3686	32 x GBM (stage VI glioma) tumor cell line	76
HLA-A*0201	1266	B-lymphoblastoid cell line	68
HLA-A & B	2375	B-lymphoblastoid cell line	64
HLA-B44	402	B-lymphoblastoid cell line	77
HLA-class I	12199	B-lymphoblastoid cell line	71
HLA-class I	7137	JY cell line	61
HLA-class I	2359	RPMI8226	61
HLA-class I	1651	U937	61
HLA-A*0201, B*0702, B*4402 & HLA-C	10867	HHC-B-lymphoblastoid cell line	Chapter 2
HLA-A*0201, B*0702 & HLA-C	6493	JY-B-lymphoblastoid cell line	Chapter 2

Table 2. Summary of the number of HLA-class I phosphorylated ligands reported by different groups.

HLA type	Nr. of Ligands	Cell source	Ref.
HLA-A & B	11	RCC tumor tissue	67
HLA-A & B	71	JY-B-lymphoblastoid cell line	65
HLA-A & B	103	Ruppen	65
HLA-A & B	36	ABB	65
HLA-A & B	9	VBT2	65
HLA-A2 & B7	99	Primary hematological malignant tissue	69
HLA-A*0201	35	JY-B-lymphoblastoid cell line	66
HLA-class I	59	B-lymphoblastoid cell line	71
HLA-A*0201, B*0702, B*4402 & HLA-C	221	B-lymphoblastoid cell line	Chapter 2

treatment of cancer. In fact, PhosImmune Inc, a spin off company of the University of Virginia, by Hunt and co-worker, develops vaccines bases on aberrantly expressed phosphorylated antigens. Their studies are currently in clinical phase I. In addition to phosphorylated ligands, cysteinylated HLA-class I ligands are often observed, which can be crucial for T cell recognition [44,45,70,71] (Chapter 2).

The number of reports on identified HLA-class I peptides is increasing, as is the number of identified ligands. However, due to both biochemical and technical limitations, most publications only identify a glimpse of the ligandome. A summary of the identified number of HLA-class I ligands for a number of studies is listed in Table 1 and the phosphorylated ligands in Table 2.

Polymorphic peptides/MiHA

HLA molecules present, next to monomorphic peptides, also polymorphic peptides on the cell surface to T cells. Polymorphic peptide is derived from a protein translated from a gene containing Single Nucleotide Polymorphism (SNP). Although at least 90,000 non-synonymous SNPs exist in the human genome, which provides for a potentially large number of polymorphic proteins [72], only 10% of the HLA-class I ligands are derived from polymorphic proteins (Chapter 2). Within an individual, polymorphic peptides have un-

dergone the thymic selection process and do not elicit T cell responses. However, polymorphic peptides can elicit T cell responses after allogeneic HLA-matched HSCT, and are called minor histocompatibility antigen (MiHA) [6,7,72]. The first human MiHA described is HA-1 that arises as a consequence of a SNP in the HMHA1 gene (73). MiHA can also arise from homozygous gene deletion of one gene of a polymorphic gene family in the donor (UGT2B17). So far, 49 MiHA (from HLA-class I and II) have been identified from different genes [18]. Different biochemical and molecular approaches have been used to identify MiHA, and mass spectrometry based approaches have played an important role.

Mass spectrometry based approaches used in identification of MiHA

To identify novel MiHA for immunotherapy, forward and reverse approaches have been used for a decade. Utilization of mass spectrometry in both approaches has resulted in the identification of several MiHA. The identification of MiHA was pioneered by Hunt and co-worker [74]. HA-1, HA-2, H-Y minor histocompatibility antigens were identified using mass spectrometry [44,65,75]. Other groups have also used mass spectrometry in the identification of HLA ligands [34,67,76]. Nowadays, mass spectrometric sensitivity allows detection of low attomole to high zeptomole amounts of peptide depending on the physicochemical properties of the peptide [77,78].

The forward approach

The forward approach to identify MiHA is powerful because of its sensitivity and the availability of relevant T-cells, which have been isolated from patients after allogeneic HSCT. In the forward biochemical approach, B- cells from the patient are used to generate Epstein Barr virus transformed B lymphoblastoid (EBV-LCL) cell lines, from which the MiHA can be eluted. This is done by isolation of the HLA molecules by immunopurification using an anti HLA-antibody, either specific for a certain type of HLA (like BB7.1 (anti HLA-A2), BB7.2 (anti HLA-B7)), or a pan HLA-class I antibody, like W6/32. After acid elution and isolation of the peptide pool and fractionation by HPLC [73], the fractions are screened with the T cells isolated from the same patient. The fraction containing the MiHA, the active fraction, elicits a T cell response. The active fraction is further fractionated using a preferably orthogonal liquid chromatography method and is finally analyzed by mass spectrometry. This process, although having been successful regularly, is still challenging because of the complexity of the sample in combination with sensitivity needed. An important limitation of the forward approach is the need of establishing the T cell response after allogeneic HSCT to start looking for MiHA [34].

The reverse approach

The reverse approach classically does not need a T cell as starting point, but starts with the prediction of MiHA. Algorithms are used to find potential MiHA in e.g. selected genes present in the target cell. Different software packages have been developed to predict the binding affinity and presentation of peptides on the

cell surface [79-81]. These include SYFPEITHI, which is the first web-based epitope prediction program [82], NetMHC [81], BIMAS [83]. The prediction programs use the experimental information from previous sequencing efforts, Edman sequencing and tandem mass spectrometry [46,84]. In addition, a wealth of information is compiled in the immunoepitope database (IEDB) [85,86].The main limitation of reverse approaches is that most potential ligands do not reach the HLA-groove. Less than 0.1% of the specific peptides survive the intracellular processing [21]. In addition, post translation modified peptides, peptides derived from alternative open reading frames, splice variants, gene insertion/deletions cannot be predicted by the available algorithms. Alternative reverse approaches are cDNA libraries screening [87,88], whole genome association scanning [52,89], and synthetic peptide library based approach [90-93].

AIM OF THIS THESIS

MiHA play an essential role in immunotherapy for the treatment of hematological malignancies. Due to the high frequencies of polymorphisms of the population, a large panel of possible target MiHA is essential to develop new immunotherapeutic to cure the majority of patients. Therefore, the aim of this thesis was the large-scale identification of relevant MiHA to improve the chances of success in therapy of hematological malignancies by a peptidomics approach. **In chapter 2**, therefore, we attempted to develop an approach to identify as many HLA ligands, and thus MiHA, as possible. To improve the identification of MiHA from the raw data, i.e. the huge number of tandem MS spectra of a majority of monomorphic and a minority of polymorphic HLA-ligands eluted from the cells, we developed a specific database described in **chapter 3** which is called the human short peptide variation (HSPV) database. **In chapter 4** we applied in our group developed database and algorithm to the data obtained from the peptidomics approach described in chapter 2 to identify novel MiHA. Obviously, the abundance of the presented peptides plays an important role in T cell recognition. Therefore, **in chapter 5** we developed an SRM/PRM-based method to accurately quantitate the ligand copy number. Remarkably, in chapter 2 several MiHA were identified together with their allelic counterparts. Generally, however, the T cell response appears to be unidirectional. **In chapter 6**, we determine the relative copy number of some MiHA and their AC on the surface of heterozygous cell lines, using the hpMHC-method developed in chapter 5 . The large number of ligands identified in chapter 2 allowed an in-depth view of many aspects of the HLA-ligandome, amongst others the peptide length distribution. The ligandome appeared to contain a substantial number of relatively long (non-canonical) peptides. **In chapter 7** we studied a selection of 15-mer HLA-A*0201-binding peptides in great detail, by immunological and crystallographic techniques.

REFERENCES

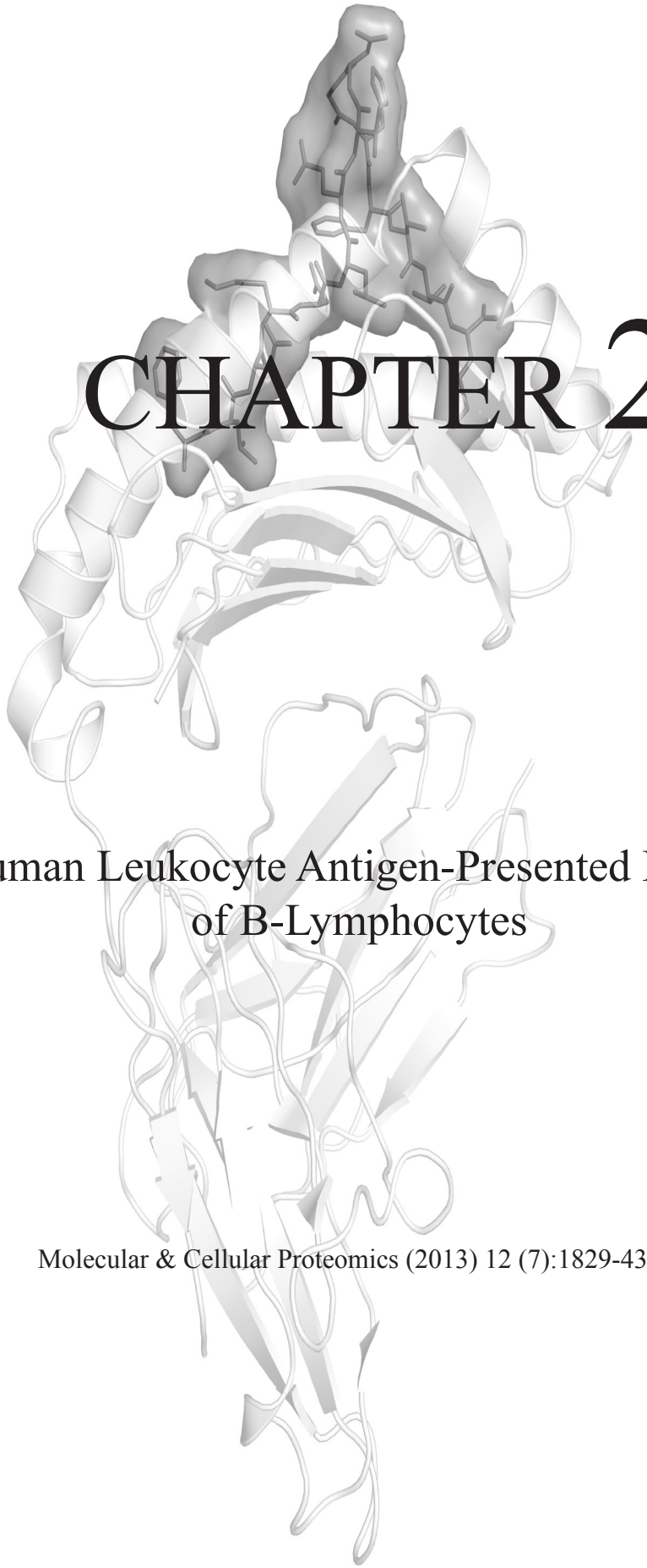
1. Bleakley, M. and Riddell, S.R. (2004) Molecules and mechanisms of the graft-versus-leukaemia effect. Nat Rev Cancer, 4, 371-380.
2. Welniak, L.A., Blazar, B.R. and Murphy, W.J. (2007) Immunobiology of allogeneic hematopoietic stem cell transplantation. Annu Rev Immunol, 25, 139-170.
3. Hansen, J.A., Petersdorf, E., Martin, P.J. and Anasetti, C. (1997) Hematopoietic stem cell transplants from unrelated donors. Immunol Rev, 157, 141-151.
4. Porter, D.L. (2011) Allogeneic immunotherapy to optimize the graft-versus-tumor effect: concepts and controversies. Hematology Am Soc Hematol Educ Program, 2011, 292-298.
5. Falkenburg, J.H. and Warren, E.H. (2011) Graft versus leukemia reactivity after allogeneic stem cell transplantation. Biol Blood Marrow Transplant, 17, S33-38.
6. Falkenburg, J.H., Marijt, W.A., Heemskerk, M.H. and Willemze, R. (2002) Minor histocompatibility antigens as targets of graft-versus-leukemia reactions. Curr Opin Hematol, 9, 497-502.
7. Goulmy, E. (1996) Human minor histocompatibility antigens. Curr Opin Immunol, 8, 75-81.
8. Socie, G., Stone, J.V., Wingard, J.R., Weisdorf, D., Henslee-Downey, P.J., Bredeson, C., Cahn, J.Y., Passweg, J.R., Rowlings, P.A., Schouten, H.C. et al. (1999) Long-term survival and late deaths after allogeneic bone marrow transplantation. Late Effects Working Committee of the International Bone Marrow Transplant Registry. N Engl J Med, 341, 14-21.
9. Marmont, A.M., Horowitz, M.M., Gale, R.P., Sobocinski, K., Ash, R.C., van Bekkum, D.W., Champlin, R.E., Dicke, K.A., Goldman, J.M., Good, R.A. et al. (1991) T-cell depletion of HLA-identical transplants in leukemia. Blood, 78, 2120-2130.
10. Horowitz, M.M., Gale, R.P., Sondel, P.M., Goldman, J.M., Kersey, J., Kolb, H.J., Rimm, A.A., Ringden, O., Rozman, C., Speck, B. et al. (1990) Graft-versus-leukemia reactions after bone marrow transplantation. Blood, 75, 555-562.
11. Warren, E.H. and Deeg, H.J. (2013) Dissecting graft-versus-leukemia from graft-versus-host-disease using novel strategies. Tissue Antigens, 81, 183-193.
12. Hale, G., Cobbold, S. and Waldmann, H. (1988) T cell depletion with CAMPATH-1 in allogeneic bone marrow transplantation. Transplantation, 45, 753-759.
13. Martin, P.J., Hansen, J.A., Buckner, C.D., Sanders, J.E., Deeg, H.J., Stewart, P., Appelbaum, F.R., Clift, R., Fefer, A., Witherspoon, R.P. et al. (1985) Effects of in vitro depletion of T cells in HLA-identical allogeneic marrow grafts. Blood, 66, 664-672.
14. Mitsuyasu, R.T., Champlin, R.E., Gale, R.P., Ho, W.G., Lenarsky, C., Winston, D., Selch, M., Elashoff, R., Giorgi, J.V., Wells, J. et al. (1986) Treatment of donor bone marrow with monoclonal anti-T-cell antibody and complement for the prevention of graft-versus-host disease. A prospective, randomized, double-blind trial. Ann Intern Med, 105, 20-26.
15. Goulmy, E., Termijtelen, A., Bradley, B.A. and van Rood, J.J. (1976) Alloimmunity to human H-Y. Lancet, 2, 1206.
16. Blazar, B.R., Murphy, W.J. and Abedi, M. (2012) Advances in graft-versus-host disease biology and therapy. Nat Rev Immunol, 12, 443-458.
17. Parham, P., Lomen, C.E., Lawlor, D.A., Ways, J.P., Holmes, N., Coppin, H.L., Salter, R.D., Wan, A.M. and Ennis, P.D. (1988) Nature of polymorphism in HLA-A, -B, and -C molecules. Proc Natl Acad Sci U S A, 85, 4005-4009.
18. Warren, E.H., Zhang, X.C., Li, S., Fan, W., Storer, B.E., Chien, J.W., Boeckh, M.J., Zhao, L.P., Martin, P.J. and Hansen, J.A. (2012) Effect of MHC and non-MHC donor/recipient genetic disparity on the outcome of allogeneic HCT. Blood.
19. DiBrino, M., Parker, K.C., Margulies, D.H., Shiloach, J., Turner, R.V., Biddison, W.E. and Coligan, J.E. (1995) Identification of the peptide binding motif for HLA-B44, one of the most common HLA-B alleles in the Caucasian population. Biochemistry, 34, 10130-10138.
20. Parker, K.C., Shields, M., DiBrino, M., Brooks, A. and Coligan, J.E. (1995) Peptide binding to MHC class I molecules: implications for antigenic peptide prediction. Immunol Res, 14, 34-57.
21. Yewdell, J.W., Reits, E. and Neefjes, J. (2003) Making sense of mass destruction: quantitating MHC class I antigen presentation. Nat Rev Immunol, 3, 952-961.

22. Rodenko, B., Toebes, M., Hadrup, S.R., van Esch, W.J., Molenaar, A.M., Schumacher, T.N. and Ovaa, H. (2006) Generation of peptide-MHC class I complexes through UV-mediated ligand exchange. Nat Protoc, 1, 1120-1132.
23. Altman, J.D., Moss, P.A., Goulder, P.J., Barouch, D.H., McHeyzer-Williams, M.G., Bell, J.I., McMichael, A.J. and Davis, M.M. (1996) Phenotypic analysis of antigen-specific T lymphocytes. Science, 274, 94-96.
24. van Buuren, M.M., Dijkgraaf, F.E., Linnemann, C., Toebes, M., Chang, C.X., Mok, J.Y., Nguyen, M., van Esch, W.J., Kvistborg, P., Grotenbreg, G.M. et al. (2014) HLA micropolymorphisms strongly affect peptide-MHC multimer-based monitoring of antigen-specific CD8+ T cell responses. J Immunol, 192, 641-648.
25. Kessler, J.H., Khan, S., Seifert, U., Le Gall, S., Chow, K.M., Paschen, A., Bres-Vloemans, S.A., de Ru, A., van Montfoort, N., Franken, K.L. et al. (2011) Antigen processing by nardilysin and thimet oligopeptidase generates cytotoxic T cell epitopes. Nat Immunol, 12, 45-53.
26. Lehner, P.J. and Cresswell, P. (1996) Processing and delivery of peptides presented by MHC class I molecules. Curr Opin Immunol, 8, 59-67.
27. Koopmann, J.O., Hammerling, G.J. and Momburg, F. (1997) Generation, intracellular transport and loading of peptides associated with MHC class I molecules. Curr Opin Immunol, 9, 80-88.
28. Serwold, T., Gaw, S. and Shastri, N. (2001) ER aminopeptidases generate a unique pool of peptides for MHC class I molecules. Nat Immunol, 2, 644-651.
29. Grandea, A.G., 3rd and Van Kaer, L. (2001) Tapasin: an ER chaperone that controls MHC class I assembly with peptide. Trends Immunol, 22, 194-199.
30. Turnquist, H.R., Vargas, S.E., McIlhaney, M.M., Li, S., Wang, P. and Solheim, J.C. (2002) Calreticulin binds to the alpha1 domain of MHC class I independently of tapasin. Tissue Antigens, 59, 18-24.
31. Zhang, Y., Baig, E. and Williams, D.B. (2006) Functions of ERp57 in the folding and assembly of major histocompatibility complex class I molecules. J Biol Chem, 281, 14622-14631.
32. Vyas, J.M., Van der Veen, A.G. and Ploegh, H.L. (2008) The known unknowns of antigen processing and presentation. Nat Rev Immunol, 8, 607-618.
33. Wearsch, P.A. and Cresswell, P. (2008) The quality control of MHC class I peptide loading. Curr Opin Cell Biol, 20, 624-631.
34. Hoppes, R., Ekkebus, R., Schumacher, T.N. and Ovaa, H. (2010) Technologies for MHC class I immunoproteomics. J Proteomics, 73, 1945-1953.
35. Marrack, P. and Kappler, J. (1987) The T cell receptor. Science, 238, 1073-1079.
36. Jiang, H. and Chess, L. (2009) How the immune system achieves self-nonself discrimination during adaptive immunity. Adv Immunol, 102, 95-133.
37. Hengartner, H., Odermatt, B., Schneider, R., Schreyer, M., Walle, G., MacDonald, H.R. and Zinkernagel, R.M. (1988) Deletion of self-reactive T cells before entry into the thymus medulla. Nature, 336, 388-390.
38. Kappler, J.W., Roehm, N. and Marrack, P. (1987) T cell tolerance by clonal elimination in the thymus. Cell, 49, 273-280.
39. Pircher, H., Rohrer, U.H., Moskophidis, D., Zinkernagel, R.M. and Hengartner, H. (1991) Lower receptor avidity required for thymic clonal deletion than for effector T-cell function. Nature, 351, 482-485.
40. 40. Starr, T.K., Jameson, S.C. and Hogquist, K.A. (2003) Positive and negative selection of T cells. Annu Rev Immunol, 21, 139-176.
41. Ostrov, D.A., Grant, B.J., Pompeu, Y.A., Sidney, J., Harndahl, M., Southwood, S., Oseroff, C., Lu, S., Jakoncic, J., de Oliveira, C.A. et al. (2012) Drug hypersensitivity caused by alteration of the MHC-presented self-peptide repertoire. Proc Natl Acad Sci U S A, 109, 9959-9964.
42. Illing, P.T., Vivian, J.P., Dudek, N.L., Kostenko, L., Chen, Z., Bharadwaj, M., Miles, J.J., Kjer-Nielsen, L., Gras, S., Williamson, N.A. et al. (2012) Immune self-reactivity triggered by drug-modified HLA-peptide repertoire. Nature, 486, 554-558.
43. Bharadwaj, M., Illing, P., Theodossis, A., Purcell, A.W., Rossjohn, J. and McCluskey, J. (2012) Drug hypersensitivity and human leukocyte antigens of the major histocompatibility complex. Annu Rev Pharmacol Toxicol, 52, 401-431.
44. Meadows, L., Wang, W., den Haan, J.M., Blokland, E., Reinhardus, C., Drijfhout, J.W., Shabanowitz, J., Pierce, R., Agulnik, A.I., Bishop, C.E. et al. (1997) The HLA-A*0201-restricted H-Y antigen contains a post-translationally modified cysteine that significantly affects T cell recognition. Immunity, 6, 273-281.
45. Petersen, J., Purcell, A.W. and Rossjohn, J. (2009) Post-translationally modified T cell epitopes: immune recognition and immunotherapy. J Mol Med (Berl), 87, 1045-1051.

46. Rammensee, H.G., Falk, K. and Rotzschke, O. (1993) Peptides naturally presented by MHC class I molecules. Annu Rev Immunol, 11, 213-244.
47. Tynan, F.E., Reid, H.H., Kjer-Nielsen, L., Miles, J.J., Wilce, M.C., Kostenko, L., Borg, N.A., Williamson, N.A., Beddoe, T., Purcell, A.W. et al. (2007) A T cell receptor flattens a bulged antigenic peptide presented by a major histocompatibility complex class I molecule. Nat Immunol, 8, 268-276.
48. Liu, Y.C., Chen, Z., Burrows, S.R., Purcell, A.W., McCluskey, J., Rossjohn, J. and Gras, S. (2012) The energetic basis underpinning T-cell teceptor tecognition of a super-bulged peptide bound to a major histocompatibility complex class I molecule. J Biol Chem.
49. Burrows, S.R., Rossjohn, J. and McCluskey, J. (2006) Have we cut ourselves too short in mapping CTL epitopes? Trends Immunol, 27, 11-16.
50. Bell, M.J., Burrows, J.M., Brennan, R., Miles, J.J., Tellam, J., McCluskey, J., Rossjohn, J., Khanna, R. and Burrows, S.R. (2009) The peptide length specificity of some HLA class I alleles is very broad and includes peptides of up to 25 amino acids in length. Mol Immunol, 46, 1911-1917.
51. Escobar, H., Crockett, D.K., Reyes-Vargas, E., Baena, A., Rockwood, A.L., Jensen, P.E. and Delgado, J.C. (2008) Large scale mass spectrometric profiling of peptides eluted from HLA molecules reveals N-terminal-extended peptide motifs. J Immunol, 181, 4874-4882.
52. Van Bergen, C.A., Rutten, C.E., Van Der Meijden, E.D., Van Luxemburg-Heijs, S.A., Lurvink, E.G., Houwing-Duistermaat, J.J., Kester, M.G., Mulder, A., Willemze, R., Falkenburg, J.H. et al. (2010) High-throughput characterization of 10 new minor histocompatibility antigens by whole genome association scanning. Cancer Res, 70, 9073-9083.
53. Falk, K., Rotzschke, O., Stevanovic, S., Jung, G. and Rammensee, H.G. (2006) Allele-specific motifs revealed by sequencing of self-peptides eluted from MHC molecules. 1991. J Immunol, 177, 2741-2747.
54. Maier, R., Falk, K., Rotzschke, O., Maier, B., Gnau, V., Stevanovic, S., Jung, G., Rammensee, H.G. and Meyerhans, A. (1994) Peptide motifs of HLA-A3, -A24, and -B7 molecules as determined by pool sequencing. Immunogenetics, 40, 306-308.
55. Engelhard, V.H. (1994) Structure of peptides associated with class I and class II MHC molecules. Annu Rev Immunol, 12, 181-207.
56. Huczko, E.L., Bodnar, W.M., Benjamin, D., Sakaguchi, K., Zhu, N.Z., Shabanowitz, J., Henderson, R.A., Appella, E., Hunt, D.F. and Engelhard, V.H. (1993) Characteristics of endogenous peptides eluted from the class I MHC molecule HLA-B7 determined by mass spectrometry and computer modeling. J Immunol, 151, 2572-2587.
57. de Verteuil, D., Granados, D.P., Thibault, P. and Perreault, C. (2012) Origin and plasticity of MHC I-associated self peptides. Autoimmun Rev, 11, 627-635.
58. Mester, G., Hoffmann, V. and Stevanovic, S. (2011) Insights into MHC class I antigen processing gained from large-scale analysis of class I ligands. Cellular and molecular life sciences : CMLS, 68, 1521-1532.
59. Yewdell, J.W., Anton, L.C. and Bennink, J.R. (1996) Defective ribosomal products (DRiPs): a major source of antigenic peptides for MHC class I molecules? J Immunol, 157, 1823-1826.
60. Yewdell, J.W. (2011) DRiPs solidify: progress in understanding endogenous MHC class I antigen processing. Trends Immunol, 32, 548-558.
61. Bourdetsky, D., Schmelzer, C.E. and Admon, A. (2014) The nature and extent of contributions by defective ribosome products to the HLA peptidome. Proc Natl Acad Sci U S A, 111, E1591-1599.
62. Milner, E., Barnea, E., Beer, I. and Admon, A. (2006) The turnover kinetics of major histocompatibility complex peptides of human cancer cells. Mol Cell Proteomics, 5, 357-365.
63. Caron, E., Vincent, K., Fortier, M.H., Laverdure, J.P., Bramoulle, A., Hardy, M.P., Voisin, G., Roux, P.P., Lemieux, S., Thibault, P. et al. (2011) The MHC I immunopeptidome conveys to the cell surface an integrative view of cellular regulation. Mol Syst Biol, 7, 533.
64. Granados, D.P., Yahyaoui, W., Laumont, C.M., Daouda, T., Muratore-Schroeder, T.L., Cote, C., Laverdure, J.P., Lemieux, S., Thibault, P. and Perreault, C. (2012) MHC I-associated peptides preferentially derive from transcripts bearing miRNA response elements. Blood, 119, e181-191.
65. Zarling, A.L., Ficarro, S.B., White, F.M., Shabanowitz, J., Hunt, D.F. and Engelhard, V.H. (2000) Phosphorylated peptides are naturally processed and presented by major histocompatibility complex class I molecules in vivo. J Exp Med, 192, 1755-1762.

66. Zarling, A.L., Polefrone, J.M., Evans, A.M., Mikesh, L.M., Shabanowitz, J., Lewis, S.T., Engelhard, V.H. and Hunt, D.F. (2006) Identification of class I MHC-associated phosphopeptides as targets for cancer immunotherapy. Proc Natl Acad Sci U S A, 103, 14889-14894.
67. Meyer, V.S., Drews, O., Gunder, M., Hennenlotter, J., Rammensee, H.G. and Stevanovic, S. (2009) Identification of natural MHC class II presented phosphopeptides and tumor-derived MHC class I phospholigands. J Proteome Res, 8, 3666-3674.
68. Scull, K.E., Dudek, N.L., Corbett, A.J., Ramarathinam, S.H., Gorasia, D.G., Williamson, N.A. and Purcell, A.W. (2012) Secreted HLA recapitulates the immunopeptidome and allows in-depth coverage of HLA A*02:01 ligands. Mol Immunol, 51, 136-142.
69. Cobbold, M., De La Pena, H., Norris, A., Polefrone, J.M., Qian, J., English, A.M., Cummings, K.L., Penny, S., Turner, J.E., Cottine, J. et al. (2013) MHC class I-associated phosphopeptides are the targets of memory-like immunity in leukemia. Sci Transl Med, 5, 203ra125.
70. Hassan, C., Kester, M.G., Ru, A.H., Hombrink, P., Drijfhout, J.W., Nijveen, H., Leunissen, J.A., Heemskerk, M.H., Falkenburg, J.H. and Veelen, P.A. (2013) The human leukocyte antigen-presented ligandome of B lymphocytes. Molecular & Cellular Proteomics.
71. Mommen, G.P., Frese, C.K., Meiring, H.D., van Gaans-van den Brink, J., de Jong, A.P., van Els, C.A. and Heck, A.J. (2014) Expanding the detectable HLA peptide repertoire using electron-transfer/higher-energy collision dissociation (EThcD). Proc Natl Acad Sci U S A.
72. Bleakley, M. and Riddell, S.R. (2011) Exploiting T cells specific for human minor histocompatibility antigens for therapy of leukemia. Immunol Cell Biol, 89, 396-407.
73. den Haan, J.M.M., Meadows, L.M., Wang, W., Pool, J., Blokland, E., Bishop, T.L., Reinhardus, C., Shabanowitz, J., Offringa, R., Hunt, D.F. et al. (1998) The minor histocompatibility antigen HA-1: A diallelic gene with a single amino acid polymorphism. Science, 279, 1054-1057.
74. Hunt, D.F., Henderson, R.A., Shabanowitz, J., Sakaguchi, K., Michel, H., Sevilir, N., Cox, A.L., Appella, E. and Engelhard, V.H. (1992) Characterization of peptides bound to the class I MHC molecule HLA-A2.1 by mass spectrometry. Science, 255, 1261-1263.
75. den Haan, J.M., Sherman, N.E., Blokland, E., Huczko, E., Koning, F., Drijfhout, J.W., Skipper, J., Shabanowitz, J., Hunt, D.F., Engelhard, V.H. et al. (1995) Identification of a graft versus host disease-associated human minor histocompatibility antigen. Science, 268, 1476-1480.
76. Meiring, H.D., van der Heeft, E., ten Hove, G.J. and de Jong, A.P.J.M. (2002) Nanoscale LC–MS(n): technical design and applications to peptide and protein analysis. Journal of Separation Science, 25, 557-568.
77. Nilsson, T., Mann, M., Aebersold, R., Yates, J.R., 3rd, Bairoch, A. and Bergeron, J.J. (2010) Mass spectrometry in high-throughput proteomics: ready for the big time. Nat Methods, 7, 681-685.
78. Engelhard, V.H. (2007) The contributions of mass spectrometry to understanding of immune recognition by T lymphocytes. Int J Mass Spectrom, 259, 32-39.
79. Tenzer, S., Peters, B., Bulik, S., Schoor, O., Lemmel, C., Schatz, M.M., Kloetzel, P.M., Rammensee, H.G., Schild, H. and Holzhutter, H.G. (2005) Modeling the MHC class I pathway by combining predictions of proteasomal cleavage, TAP transport and MHC class I binding. Cell Mol Life Sci, 62, 1025-1037.
80. Peters, B., Bulik, S., Tampe, R., Van Endert, P.M. and Holzhutter, H.G. (2003) Identifying MHC class I epitopes by predicting the TAP transport efficiency of epitope precursors. J Immunol, 171, 1741-1749.
81. Lundegaard, C., Lamberth, K., Harndahl, M., Buus, S., Lund, O. and Nielsen, M. (2008) NetMHC-3.0: accurate web accessible predictions of human, mouse and monkey MHC class I affinities for peptides of length 8-11. Nucleic Acids Res, 36, W509-512.
82. Rammensee, H., Bachmann, J., Emmerich, N.P., Bachor, O.A. and Stevanovic, S. (1999) SYFPEITHI: database for MHC ligands and peptide motifs. Immunogenetics, 50, 213-219.
83. Parker, K.C., Bednarek, M.A. and Coligan, J.E. (1994) Scheme for ranking potential HLA-A2 binding peptides based on independent binding of individual peptide side-chains. J Immunol, 152, 163-175.
84. 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.
85. Schisler, N.J. and Palmer, J.D. (2000) The IDB and IEDB: intron sequence and evolution databases. Nucleic Acids Res, 28, 181-184.
86. Vita, R., Zarebski, L., Greenbaum, J.A., Emami, H., Hoof, I., Salimi, N., Damle, R., Sette, A. and Peters, B. (2010) The immune epitope database 2.0. Nucleic Acids Res, 38, D854-862.

87. Stumpf, A.N., van der Meijden, E.D., van Bergen, C.A., Willemze, R., Falkenburg, J.H. and Griffioen, M. (2009) Identification of 4 new HLA-DR-restricted minor histocompatibility antigens as hematopoietic targets in antitumor immunity. Blood, 114, 3684-3692.
88. van de Corput, L., Chaux, P., van der Meijden, E.D., De Plaen, E., Frederik Falkenburg, J.H. and van der Bruggen, P. (2005) A novel approach to identify antigens recognized by CD4 T cells using complement-opsonized bacteria expressing a cDNA library. Leukemia, 19, 279-285.
89. Kawase, T., Nannya, Y., Torikai, H., Yamamoto, G., Onizuka, M., Morishima, S., Tsujimura, K., Miyamura, K., Kodera, Y., Morishima, Y. et al. (2008) Identification of human minor histocompatibility antigens based on genetic association with highly parallel genotyping of pooled DNA. Blood, 111, 3286-3294.
90. Gavin, M.A., Dere, B., Grandea, A.G., 3rd, Hogquist, K.A. and Bevan, M.J. (1994) Major histocompatibility complex class I allele-specific peptide libraries: identification of peptides that mimic an H-Y T cell epitope. Eur J Immunol, 24, 2124-2133.
91. Hiemstra, H.S., Duinkerken, G., Benckhuijsen, W.E., Amons, R., de Vries, R.R., Roep, B.O. and Drijfhout, J.W. (1997) The identification of CD4+ T cell epitopes with dedicated synthetic peptide libraries. Proc Natl Acad Sci U S A, 94, 10313-10318.
92. Hiemstra, H.S., van Veelen, P.A., Willemen, S.J., Benckhuijsen, W.E., Geluk, A., de Vries, R.R., Roep, B.O. and Drijfhout, J.W. (1999) Quantitative determination of TCR cross-reactivity using peptide libraries and protein databases. Eur J Immunol, 29, 2385-2391.
93. Zhao, Y., Gran, B., Pinilla, C., Markovic-Plese, S., Hemmer, B., Tzou, A., Whitney, L.W., Biddison, W.E., Martin, R. and Simon, R. (2001) Combinatorial peptide libraries and biometric score matrices permit the quantitative analysis of specific and degenerate interactions between clonotypic TCR and MHC peptide ligands. J Immunol, 167, 2130-2141.



CHAPTER 2

The Human Leukocyte Antigen-Presented Ligandome of B-Lymphocytes

Molecular & Cellular Proteomics (2013) 12 (7):1829-43

Based on:

Chopie Hassan*
Michel G.D. Kester*
Arnoud H. de Ru
Pleun Hombrink
Jan Wouter Drijfhout
Harm Nijveen
Jack A.M. Leunissen
Mirjam, H.M. Heemskerk
J.H. Frederik Falkenburg
Peter A. van Veelen

* equal contribution

2

The Human Leukocyte Antigen-Presented Ligandome of B-Lymphocytes

CHAPTER 2

ABSTRACT

Peptides presented by human leukocyte antigen (HLA) molecules on the cell surface play a crucial role in adaptive immunology, mediating the communication between T cells and antigen presenting cells. Knowledge of these peptides is of pivotal importance in fundamental studies on T cell action, and in cellular immunotherapy and transplantation. In this study we present the in-depth identification and relative quantification of 14,500 peptide ligands constituting the HLA-ligandome of B-cells. This large number of identified ligands provides a general insight in the presented peptide repertoire and antigen presentation. Our uniquely large set of HLA-ligands allowed us to characterize in detail the peptides constituting the ligandome in terms of relative abundance, peptide length distribution, physicochemical properties, binding affinity to the HLA molecule and presence of post-translational modifications. The presented B-lymphocyte ligandome is shown to be a rich source of information by the presence of minor histocompatibility antigens, virus-derived epitopes and post-translationally modified HLA ligands and can be a good starting point to solve a wealth of specific immunological questions. These HLA ligands can form the basis for reversed immunology approaches to identify T cell epitopes, not based on *in silico* predictions, but based on the *bona fide* eluted HLA-ligandome.

INTRODUCTION

Peptides presented by human leukocyte antigen (HLA) molecules on the cell surface play a crucial role in immunology, and mediate the communication between T cells and antigen presenting cells. Knowledge of these peptides is of pivotal importance in fundamental studies on T cell action and in design of T cell-mediated therapy, like tumor immunotherapy [1], and treatment of hematological malignancies by a combination of hematopoietic stem cell transplantation (HSCT) and donor lymphocyte infusion (DLI) [2]. In addition, T cells can play an important role in organ rejection following transplantation.

The presented HLA class I ligands are the products of the intracellular processing machinery with its continuous cycle of protein synthesis and degradation [3]. Much is known about the proteins involved in antigen processing, but high fidelity ligand/epitope predictions are at present not possible. The complexity of antigen processing has been shown to be even more complex by the discovery of new enzymes involved in antigen processing [3, 4] and the exciting discovery of peptide splicing [5]. Moreover, gene expression studies have shown many non-standard, unexpected protein products, including the production of antigens derived from aberrant protein fragments as a result of expression in alternative reading frames [6]. Several studies report on the identification of HLA ligands [7-10]. Many results have been collected and discussed in a recent review on large-scale analysis of HLA class I ligands [11]. Collectively, these reports illustrate the need for in-depth elucidation of the HLA-ligandome.

Elucidation of T cell epitopes has traditionally been achieved by a forward immunological approach, as pioneered by Hunt and coworkers [12, 13]. In this approach, the cognate peptide of T cells with the appropriate activity profile is elucidated by repeated rounds of chromatographic separation in combination with T cell recognition assays. Since T cells are not always available from the start, reverse immunological approaches [14-17] have been developed to predict T cell epitopes by a combination of bioinformatics and e.g. *in vitro* proteasome digests. Predicted epitopes are synthesized and tested for their capability to activate T cells. The main disadvantage of this approach is that less than 0.1% of the peptides that survive intracellular processing, are presented on HLA- class I molecules [3].

Therefore, we developed a large scale peptidomics approach, which is a reverse immunology approach, not based on algorithms, but based on the *bona fide* eluted ligandome, which means that the identified peptides are known to have survived processing and are *bona fide* HLA-ligands. Following the identification of the ligandome as comprehensively as possible, T cells can subsequently be selected on the basis of the immunological question at hand, as will be illustrated in a separate paper. The development of MHC exchange tetramers for finding relevant T cell epitopes is instrumental to this approach [18, 19].

To improve ligandome coverage we applied and compared three off-line first dimension separation techniques,

followed by on-line nanoHPLC-tandem MS.

The tandem mass spectra were interrogated by matching against the IPIhuman database [20]. In a second step post translation modifications (phosphorylation, cysteinylation) were allowed in the database search. In a third step the tandem mass spectra were matched against a newly in-house developed database for optimal finding of polymorphic ligands, for finding potential minor histocompatibility antigens [21]. This led to the identification of approximately 14,000 HLA class I ligands, the majority of which was also relatively quantitated. Next, we analyzed the peptides constituting our ligandome as detailed as possible to confirm the correct identification of the vast majority of the ligands. We achieved this by a combination of several physicochemical and biological checks and comparison with existing ligand and epitope databases.

Finally, as an additional quality check, we illustrate the functional relevance of the ligandome by the identification of previously known and new minor histocompatibility antigens, virus-derived epitopes and post-translationally modified HLA ligands (phosphorylated ligands, and cysteinylated ligands) [22-24]. Since this is the largest ligandome reported to date, it allows a general insight in the presented peptide repertoire. This study supports the building of the 'immunopeptidome' as has recently been suggested [25]. A proteomics approach as a starting point to contribute to immunology by providing a peptidome landscape can be used in many immunological studies, both fundamental and applied.

EXPERIMENTAL PROCEDURES

Sample preparation

The Epstein-Barr virus (EBV) transformed B lymphoblastic cell lines B-LCL-HHC (typing: HLA-A*0201, B*0702, B*4402, Cw*0501 & Cw*0702) and B-LCL-JY pp65 (typing: HLA-A*0201, B*0702 & Cw*0702) were used as source of HLA-class I molecules. The CMV-derived pp65 transduced cell line was used to introduce an internal control for the ligandome since the CMVpp65-derived T cell epitopes are known [26]. Cells were expanded in roller bottles using IMDM supplemented with 10% heat-inactivated fetal bovine serum (FBS), penicillin/streptomycin and L-glutamine, were collected, washed with ice cold PBS, and stored at -80^{o}C until use.

The hybridoma cell line was expanded in roller bottles to obtain W6/32 (anti HLA-Class I) antibody using protein free hybridoma medium supplemented with penicillin/streptomycin, and L-glutamine. Antibodies produced by the hybridoma cell lines were purified from the supernatant using Prot-A sepharose beads, and eluted from the Prot-A beads with Glycine pH 2.5. The eluted antibodies were used to produce immunoaffinity column (W6/32- Prot-A sepharose 2.5 mg/ml). The W6/32 antibodies were covalently bound to Prot-A sepharose beads using dimethylpimelimidate (DMP). The columns were stored in PBS pH 8.0 and 0.02% NaN_3 at 4 ^{o}C.

Isolation of HLA class I-presented peptides

Extraction of peptides associated with HLA-class I molecules was performed as described previously [13, 27]. Briefly, pellets from 60 x10^9 B-LCL-JYpp65 cells and 40 x10^9 B-LCL-HHC cells were lysed in 50 mM Tris-HCl, 150 mM NaCl, 5 mM EDTA, and 0.5% Nonidet-P40, (pH 8.0) and supplemented with Complete® protease inhibitor (Sigma Aldrich). The total concentration of the cells in the lysis buffer was 0.1x 10^9 cells/ml.

After 2 hours incubation with tumbling of the cells in the lysis buffer at 4 °C, the preparation was centrifuged for 10 minutes at 2,500 rpm and 4 °C. The supernatant was transferred to a new tube and centrifuged for 35 minutes at 11,000 rpm and 4°C. The supernatant was pre-cleared with CL4B beads and subjected to the immunoaffinity column with a flow rate of 2.5 ml/min. After washing, bound HLA-class I/peptide complexes were eluted from the column, and dissociated with 10% acetic acid. Peptides were separated from the HLA-class I molecules by passage through a 10 kDa membrane (Pall macrosep centrifuge devices). The filtrate was freeze dried. If an oily sample remained after freeze drying, the sample was dissolved and the peptides were further purified by solid phase extraction (C18 Oasis, 100 µl bed volume, Waters). The peptides were eluted from the C18 Oasis column with 500 µl 50/50/0.1 water/ACN/FA, v/v/v. The eluted peptides from B-LCL-HHC cell lines were divided into two equal portions, freeze dried and dissolved in 95/3/0.1 water/ACN/FA, v/v/v. The eluted peptides from B-LCL-JYpp65 cell lines were divided into three equal portions, freeze dried, and dissolved in 95/3/0.1 water/ACN/FA, v/v/v.

Peptide separation

For peptide IEF separations, the OFFGEL Agilent 3100 fractionator (Agilent Technologies, Waldbronn, Germany) was used. A modified method was applied by addition of 1 M urea to the buffer sample and rehydration buffer, instead of 5% glycerol only. The commercially available 13 cm IPG dry strips with a linear pH gradient ranging from 3-10 (GE-Health care) were used. The strips were rehydrated with 40 µL/well rehydration solution in the assembled device for 30 min. 150 µL of the prepared samples were loaded on each well, the cover fluid (mineral oil, Agilent technologies) was added onto both ends of the gel strip. The focusing methods, OG12PE01, as supplied by the manufacturer was applied for 12 well fractionations. The performance of the 3100 OFFGEL fractionator was checked under similar conditions in a separate run by determination of the pH using a pH indicator pH 3-10 (Fluka Analytical, Germany). Fractions were recovered and desalted by solid phase extraction, to desalt the fractions and to avoid the presence of oil and gel pieces in the samples, using C18 Oasis columns. The column was prewetted with 10/90 water/ACN v/v and equilibrated with 95/3/0.1 water/ACN/FA v/v/v. The samples were eluted with 50/50/0.1 water/ACN/FA v/v/v, freeze dried and dissolved in 100 µL 95/3/0.1 water/ACN/FA v/v/v. The twelve fractions generated from IEF were analyzed in triplicate

and duplicate for B-LCL-HHC and B-LCL-JYpp65 cell lines, respectively, with nanoLC-MS/MS.

Next to the peptide IEF separation, two chromatographic separation techniques were applied, strong cation chromatography (SCX) and RP-C18 chromatography. For SCX separations, one portion of the eluted peptides from the W6/32 column was fractionated with a home-made SCX column (320 µm ID, 15 cm, polysulfoethyl A 3 um, Poly LC), run at 4 ul/min. Gradients were run for 10 min at 100% solvent A (100% water/0.1% TFA), after which a linear gradient started to reach 100% solvent B (250 mM KCl, 35% ACN/0.1% TFA) in 15 min, followed by 100 % solvent C (500 mM KCl, 35% ACN/0.1% TFA) in the next 15 min and remained at 100 % solvent C for 5 min, then switched again to 100 % solvent A. Twenty 4 ul-fractions were collected in vials prefilled with 100 µl 95/3/0.1 water/ACN/FA v/v/v.

One portion of the eluted peptides from B-LCL-JY pp65 cell lines was fractionated on a home-made RP Reprosil-Pur C18-AQ column (200 µm ID, 3 µm x 15 cm) (Dr. Maisch, GmbH, Ammerbuch, Germany). The sample was loaded in solvent A (10/90/0.1 water/ACN/FA v/v/v), and the gradient was run from 0-50% B (10/90/0.1 water/ACN/FA v/v/v) in 30 min at a flow rate of 3 ul/min. The samples were taken up in a make-up flow of 50/50/0.1 water/ACN/FA at 100 ul/min supplied via a T-piece through the annular space between the separation capillary and an auxillary capillary. In this way 45 half a minute wide fractions were collected, subsequently freeze dried, and dissolved in solvent A for analysis by nanoLC-MS/MS.

LC-MS/MS analysis

The dissolved fractions were analyzed by on-line nano-HPLC mass spectrometry with a system, consisting of a conventional Agilent 1100 gradient HPLC system (Agilent, Waldbronn, Germany), described by Meiring et al [28], and a LTQ-FT Ultra mass spectrometer (Thermo, Bremen, Germany). Fractions were injected onto a home-made pre-column (100 um×15 mm; Reprosil-Pur C18-AQ 3 µm, Dr. Maisch, Ammerbuch, Germany) and eluted via in home-made analytical nano-HPLC column (15 cm×50 µm; Reprosil-Pur C18-AQ 3 um). The gradient was run from 0% to 50% solvent B (10/90/0.1 water/ACN/FA v/v/v) in 90 min. The nano-HPLC column was drawn to a tip of approximately 5 µm and acted as the electrospray needle of the MS source. The mass spectrometer was operated in data dependent mode, automatically switching between MS and MS/MS acquisition. Full scan mass spectra were acquired in the FT-ICR with a resolution of 25,000 at a target value of 5,000,000. The two most intense ions were then isolated for accurate mass measurements by a selected ion monitoring scan in FT-ICR with a resolution of 50,000 at a target accumulation value of 50,000. The selected ions were then fragmented in the linear ion trap using collision-induced dissociation at a target value of 10,000. In a post analysis process, raw data were converted to peak lists using Bioworks Browser software, Version 3.2.0.

Data analysis

The tandem mass spectra were matched against the International Protein Index (IPI) human database version 3.87, using the mascot search engine version 2.2.04 (Matrix Science, London, UK)), with a precursor mass tolerance of 2 ppm, with methionine oxidation as a variable modification, and a product ion tolerance of 0.5 Da. For finding post-translationally modified HLA-ligands phosphorylation on serine, threonine and tyrosine were allowed, and cysteinylation of cysteine in separate searches. Scaffold software version 3 (www.proteomesoftware.com) was subsequently used to process the mascot output files and generate spectrum reports. Duplicates were removed, and peptides with a best mascot ion score >35 and 8-11 amino acids long, were selected for the production of supplemental Table 1. For the immunological examples a best mascot score of >20 was selected, and the length was restricted to 8-18 amino acids. Next to the above mentioned procedure, Proteome Discoverer 1.3 (Thermo, Bremen, Germany) was used to extract all identified peptides from the input *.RAW-files, using the mascot server mentioned above, and calculate their intensity, as reported in supplemental Table 1. False discovery rates were as determined by Proteome Discoverer, for the Homo Sapiens-extracted Uniprot/SWISSProt database [29] (release 2010_11) (www.ebi.ac.uk) containing 20,259 protein sequences, and the IPIhuman 3.87 database (www.ebi.ac.uk/IPI/IPIhuman.html) see also supplemental Table 2. Icelogo version 1.2 was used to generate the binding motifs as presented in Figure 5 and supplemental Figure 1 [30]. The GRAVY index was calculated using http://www.bioinformatics.org/sms2/ protein_gravy.html. The pIs of the identified peptides were calculated using http://www.expasy.org/tools/pi_tool.html. NetMHC 3.2 (http://www.cbs.dtu.dk/services/NetMHC) was used to predict the binding affinity (nM) of the identified peptides to HLA-A*0201, HLA-B*0702 and HLA-B*4402. NetMHCpan 2.4 (http://www.cbs.dtu.dk/services/NetMHCpan) was used to predict the binding affinity of the identified peptides to HLA-C*0501 and HLA-C*0702. Overall protein turnover values were taken from Cambridge et al. [31]. Phosphosite.org [32] (www.phosphosite.org) and Phospho.ELM 8.3 [33] (http://phospho.elm.eu.org) were used to find known phosphosites in the identified peptides. NetPhos 2.0 [34] (http://www.cbs.dtu.dk/services/NetPhos) was used to predict phosphosites in identified phosphorylated ligands. For the identification of polymorphic peptides, the tandem mass spectra were matched against the HSPVdb, a database optimized for finding polymorphic peptides [21]. For searching CMV pp65-derived ligands/epitopes in B-LCL-JYpp65, a separate database was constructed only containing the DNA sequence of the pp65 protein (NCBI; pp65_AD169_seq with intron Human herpesvirus 5, complete genome). For searching EBV-derived epitopes, a separate database was constructed containing the DNA sequence of EBV selected from RefSeq database [35] (>gi|82503188|ref|NC_007605.1| Human herpesvirus 4 type 1, complete genome).

Peptide synthesis

Peptides were synthesized according to standard fluorenylmethoxycarbonyl (F-moc) chemistry using a SyroII peptide synthesizer (MultiSynTech, Witten, Germany). The integrity of the peptides was checked using RP-HPLC and MS. Phospho peptides were synthesized using the building blocks Fmoc-Ser(PO(OBzl)OH)-OH or Fmoc-Thr(PO(OBzl)OH)-OH. Couplings of these building blocks were performed identical to normal couplings with one exception. A 3-fold excess of N-methylmorpholine was used instead of the routinely applied 2-fold excess.

RESULTS

The peptides constituting the HLA-ligandome

The pool of HLA-peptides eluted from two Epstein-Barr virus-transformed B lymphoblastoid cell lines (EBV-B-LCL), either B-LCL-JYpp65 or B-LCL-HHC, is quite complex. From previous experiments a simple count of the total number of peaks in the MS1 spectra in all fractions following a first dimension C18-separation resulted in a number of approximately 30,000. Because of this complexity, multidimensional separations were performed to reduce the complexity before mass spectrometric analysis. Three first dimension separations were chosen, reverse phase C18 (RP-C18) chromatography, strong cation exchange (SCX) chromatography, and peptide isoelectric focusing (IEF), since these are based on a different separation mechanism, and most commonly used. The second dimension separation was RP-C18 chromatography coupled on-line to the mass spectrometer in all cases. Below the results of the three first dimension separations in combination with the second dimension on-line HPLC-MS are compared and discussed.

Peptide isoelectric focusing as the 1st dimension

Before starting our IEF-experiments with HLA-eluted peptides, the separation performance of the OFFGEL 3100 system was tested using a trypsin-digested cell lysate, using 13 cm IPG strips, pH range of 3-10 (GE-Healthcare). Separation efficiency was studied as a function of loading. The results were as reported in [36], with approximately 83% of peptides present in a unique fraction and 96% present in one or two unique fractions , with peptide amounts up to100 ug, confirming the high separation efficiency of the IEF process. Peptide loading below 10 μg results in considerable sample losses, which precludes use of IEF for HLA-ligandome studies of smaller cell amounts. Application of a pH indicator (pH 3 - pH 10) after IEF fractionation showed that the low pH side of the strip was *not* pH 3, but was actually close to pH 4.

The 12 fractions obtained from the IEF fractionation were analyzed by on-line nanoHPLC tandem mass spectrometry, and the tandem mass spectra were matched against the IPI human protein database. The results for both B-LCLs are summarized as Venn diagrams in Figure. 1(A & B). As can be seen, analysis of the

B-LCL-HHC ligandome yielded 6878 unique 8-11 mer peptides with BMI>35, and the B-LCL-JYpp65 ligandome yielded 3531 unique peptides. The large difference in the number of peptides found for B-LCL-JYpp65 and B-LCL-HHC can be largely explained by the additional HLA-molecules expressed on B-LCL-HHC compared to B-LCL-JY cells. B-LCL-JY is homozygous for the B-allele (HLA-B7*0702) and C allele (HLA-Cw*0702), while B-LCL-HHC is heterozygous for the B-allele (HLA-B*0702 and HLA-B*4402) and C-allele (HLA-Cw*0501 and HLA-Cw*0702). The additional HLA-molecules can present two additional sets of peptides.

Strong cation exchange (SCX) chromatography as the 1st dimension

Twenty fractions were collected after separation on the SCX-column. These fractions were analyzed, like the IEF-fractions, by on-line nanoHPLC tandem mass spectrometry and they were matched against the IPI human protein database. The results for both B-LCLs are summarized as Venn diagrams in Figure. 1 (C& D) . Analysis of the B-LCL-HHC ligandome yielded 7766 unique peptides (BMI>35) and the B-LCL-JYpp65 ligandome yielded 3410 unique peptides (BMI>35).

Reverse Phase-C18 Chromatography as the 1st dimension

Forty five fractions were collected after separation on the RP-C18-column. These fractions were analyzed, like the IEF- and SCX-fractions, by on-line nanoHPLC tandem mass spectrometry and the tandem mass spectra were again matched against the IPI human protein database. Analysis of the B-LCL-JYpp65 ligandome yielded 2725 unique peptides (BMI>35). The results for B-LCL-JY are summarized as Venn diagrams in Figure. 1 (E). The B-LCL-HHC ligandome was not measured with the combination of RP-C18 and on-line nanoHPLC.

Combination of the results of IEF, SCX and RP-C18

To get an overview, a non-redundant list of all peptides obtained via the above-mentioned first dimension separation techniques was produced. The results are shown as Venn diagrams in Figure. 2. The analysis of the B-LCL-HHC ligandome yielded 10,867, Figure. 2 (A) unique peptides (BMI>35) and the B-LCL-JYpp65 ligandome yielded 6,493, Figure. 2 (B) unique peptides (BMI>35). Combination of the two B-LCL-derived peptide lists yields a non-redundant peptide list containing 14,065 members, Figure. 2 (C), of which 11,511 could be relatively quantified. This is by far the most comprehensive list of HLA-ligands reported to date. A complete listing of the HLA-presented peptides, sorted by intensity, is given in supplemental Table S1, including the protein name and IPI accession number they are derived from and their intensity in B-LCL-JY and/or B-LCL-HHC, the best mascot ion score, the predicted binding affinity by NetMHC, the GRAVY hydrophobicity value, the protein half-life, the copy number per cell calculated from the intensity and work-up yield, the peptide length and if peptides are present in the IEDB [37] and/or SYFPEITHI database [38], and the calculated FDR using the Uniprot/SWISSProt database and IPIhuman database. In addition, we compared

our data with those reported by [7-10].

The NetMHC algorithm was used to predict the HLA-binding affinity of the ligands. It is clear from the results that the majority of peptides is a binder in its particular HLA-molecule, see supplemental Table S1. In the table a NetMHC cut-off score of 1000 was used.

The number of peptides found in every experiment, independent of the work up procedure, is 835 of which 770 could be quantified, Figure. 2 (B). Because of the shotgun nature of the experimental approach these 835 peptides are expected to be the most abundant on average, the peptides detected in either overlap of 2 experiments to be less abundant, and the peptides detected only in a single first dimension experiment to be the lowest. This was checked for the peptides presented on B-LCL-JY and the result displayed in Figure. 3.

Peptide length distribution

Our large set of peptides allows a closer look at the length distribution of peptides constituting the HLA-ligandome, as presented in Figure. 4. As can be seen, most ligands have a length of 9 amino acids, while a substantial part is either 10 or 11 amino acids long. The 8-mers, and 12-14 mers are all below 5%. However, still 708 long (>11 amino acids) peptides (amounting to 5% of the ligandome) are present, having the correct anchors for binding to the HLA-A and B alleles. Figure. 5 shows the HLA-A2 binding motif for the 9, 12 and 14-mers. All different peptide lengths display the same binding motif, and a binding motif can still be discerned for 15-mers.

The dynamic range of the HLA-ligandome

To get an impression of the dynamic range of the presented peptide repertoire, the intensities of the presented peptides in the ligandome were determined. To this end, the *.RAW-files were processed in Proteome Discoverer.

A complete listing of all peptides, including their intensity, presentation on the particular B-LCL, and its best mascot ion score is given in supplemental Table S1. Intensity extremes, expressed as peak area values in PD, range from 4E9 to 1E5, but the majority ranges from 1E9 to 1E6, a range of a factor of 1000. So the dynamic range of the peptide repertoire is quite high.

The summed intensity (integration value) of all peptides eluted from 2E10 B-LCL-JY or B-LCL-HHC cells is approximately 3E11, as reported by Proteome Discoverer (PD), see supplemental Table S1. To determine to what amount of peptide this number corresponds, we followed the same analytical procedure with a known amount of a mix of 20 synthetic peptides. An integration value of 1E8 as reported by PD corresponds to 1 pmol. From this we determine the total eluted peptide amount to be 3 nmol (3E11/1E8).

The composition of the ligandome can also be represented in terms of relative intensity of the peptides, as shown in Figure. 6 (A). The intensities have been translated to copy number in Figure. 6 (B).

The 14065 HLA-ligands are derived from 7059 different proteins, a graphical representation is shown Figure. 7 (A). In Figure. 7 (B) the average protein length is plotted against the number of peptides/protein. The higher the protein mass, the higher the number of derived HLA-ligands. 61 proteins are even represented by 10 peptides or more on the cell surface, with the exceptional case of 29 and 41 peptides/protein. The intensity of the peptides derived from the same protein varies greatly, as presented in Table 2.

Physicochemical properties of the peptides comprising the HLA-ligandome

To check if peptide detection depended on the peptide physicochemical properties, with either one of the employed first dimension separation techniques, we studied the influence of isoelectric point and hydrophobicity. To study the influence of the isoelectric point of the peptide we calculated the theoretical pIs of all identified ligands. The calculated pIs ranged from 3.2 to 10.4. When the peptides were assigned to pI-bins as shown in Figure. 8 (A), it was immediately clear that ligands with pI extremes migrate out of the pI-strip (since these were not detected with IEF as the 1st dimension separation technique, but in contrast were detected using SCX and C18 as the 1st dimension separation technique) and are, therefore, lost for detection using IEF. In addition, the actual pH at the low pH side of the strip is not 3, but almost 4, which accounts for extra losses experienced by application of IEF. Since peptides binding to HLA-B44 are on average more acidic, they are more prone to be lost in IEF. Next, we studied the influence of peptide hydrophobicity on the peptide's chance to be detected. To this end we calculated the gravy index of all identified peptides and plotted them in hydrophobicity bins as shown in Figure. 8 (B). As can be seen, the distribution of the peptides is evenly spread independent of the 1st dimension separation technique employed. The more hydrophilic peptides are slightly disfavored by RP-C18 as the first dimension separation technique. In summary, peptides with pI extremes might be lost during IEF, but, generally the majority of peptides was detected independent of the 1st dimension separation technique used.

The overall quality of the presented HLA-ligandome data

To illustrate that the majority of the peptides reported here was indeed derived from the expected HLA-molecules, we showed above that these peptides carry the correct anchor residues required for binding. Another way to verify the overall quality of the ligandome reported, is to check the predicted binding affinity for the HLA-molecule. In addition, the overall physicochemical characteristics of the peptides bound depend on the particular HLA-molecule.

The HLA-binding affinity can be approximated theoretically by the NetMHC algorithm. In Figure. 9, the hydrophobicity distribution for HLA-A2, HLA-B7 and HLA-B44 predicted peptides is plotted. The HLA-A2 presented peptides are centered around a GRAVY value of +1, those from HLA-B7 around 0 and those from HLA-B44 around -1.

In addition to the overall quality check as described above, we compared our data on the individual peptide level with the two main public sources of HLA-ligands, which are the SYFPEITHI-database and the Immune Epitope Database (IEDB). These databases are collections of ligands and/or epitopes that result from different immunological experiments on a variety of cell types, and generally not fed by proteomics-type of experiments but by individual immunological reports. To compare our peptide lists with the SYFPEITHI-database we selected the reported HLA-A2, HLA-B7 and HLA-B44 ligands and compared it to our eluted peptide sequences. In addition, we compared our results with the peptides presented in the papers by K.E. Scull et al. on B-cells [7], N. Hillen et al. on HLA-B44 supertypes on B-cells [8], D. Granados et al. on B-cells [9] and C. Bade-Doeding et al. on B-cells [10] The results of this comparison are shown in supplemental Table S1.

We found 253 of the 784 ligands described in the SYFPEITHI database, and 72 out of 263 ligands described in the IEDB. Out of cumulative 1179 (8-11 mer) selected HLA-A2 ligands recently published by Scull et al. 889 were found in our dataset [7].

Immunological and biological values

Below a number of immunologically relevant ligands are described. These include minor histocompatibility antigens and virus-derived epitopes.

MiHA

Minor histocompatibility antigens (MiHA) are peptides derived from polymorphic proteins and are relevant in allogeneic hematopoietic stem cell transplantation (HSCT) as targets for immunotherapy for the treatment of hematological malignancies [2, 13, 39-42].

To find and identify polymorphic peptides in the eluted peptides from B-LCL-HHC and B-LCL-JYpp65 cell lines the tandem mass spectra were matched with a dedicated database, HSPVdb [21]. We found 1439 polymorphic peptides of 8-11 mer length and mascot ion score >35, accounting for 10% of the number of peptides in the ligandome. 6 out of 16 MiHA and 4 allelic counterparts of known minor antigens, which could potentially be present on our cells, considering the SNP-typing of the B-LCL were found in our eluted dataset which was confirmed by matching the MS/MS spectra of the synthetic peptides with the eluted peptides. These peptides are listed in Table 1. It should be noted that 3 out 10 of the peptides listed have a BMI<35, but were correctly identified.

Virus-derived peptides

Since EBV-transformed cell lines were used in this study, our tandem mass spectra were matched with a specific database containing the genomic EBV information. Four EBV peptides were found in our data, see supplemental Table S3. The binding affinity of the newly identified peptides was predicted using NetMHC. Two peptides displayed high binding affinity for HLA-A2 and the other two showed a high binding affinity for

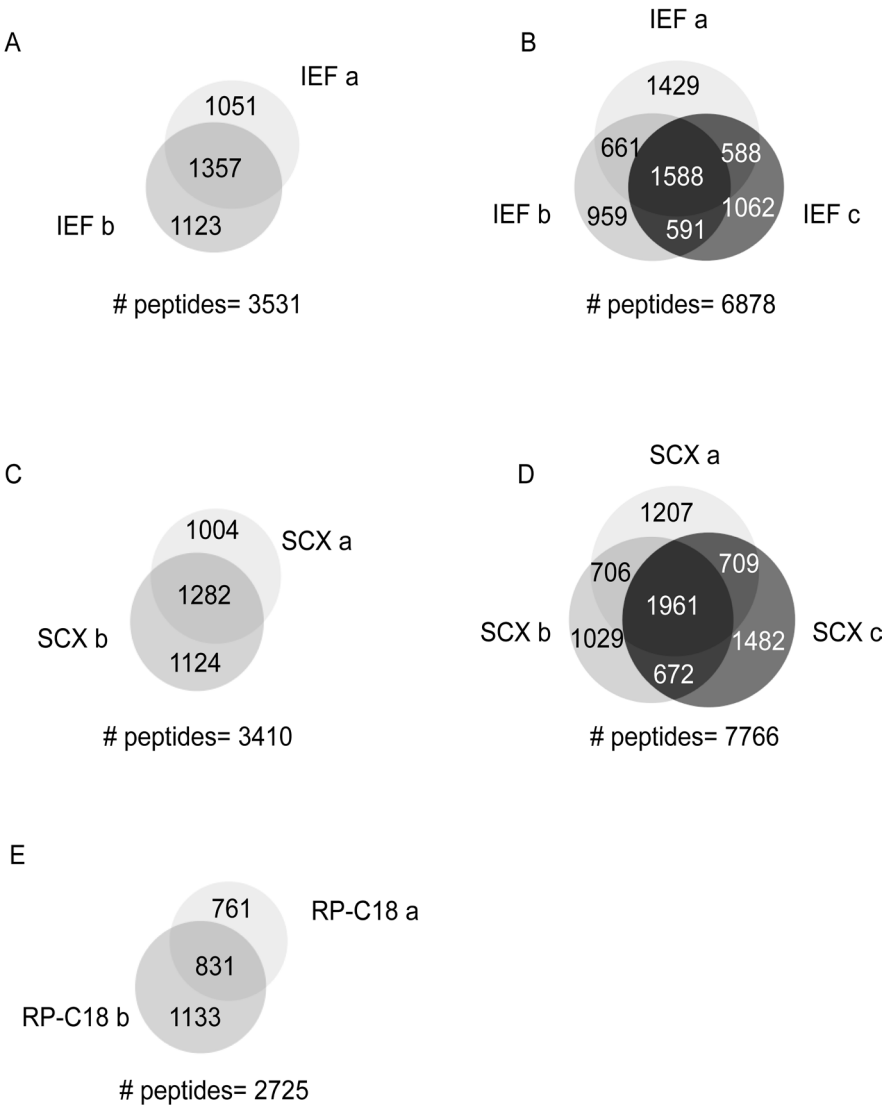


Figure 1. Venn diagrams of identified HLA-ligands per separation. The number of identified unique peptides per mass spectrometric replicate for IEF (A), SCX (C) and RP-C18 (E) for cell line B-LCL JYpp65. IEF (B) and SCX (D) for cell line B-LCL HHC. The area of the circles corresponds with the number of peptides. Total, fractional and overlapping numbers of unique peptides are indicated.

HLA-B7. The correct identification of these four peptides was confirmed by matching the tandem mass spectra of the eluted peptides with their synthetic counterparts. These four EBV-derived peptides have not been reported before.

The B-LCL-JY cell line used in this study was transduced with the CMVpp65 protein, which gave us the opportunity to study the presentation of CMVpp65-derived ligands. To identify pp65-derived ligands, we matched the tandem mass spectra with a specific database containing the pp65 sequence. We could find two known pp65-derived epitopes, see supplemental Table S3. One epitope is presented in HLA-A2, the other

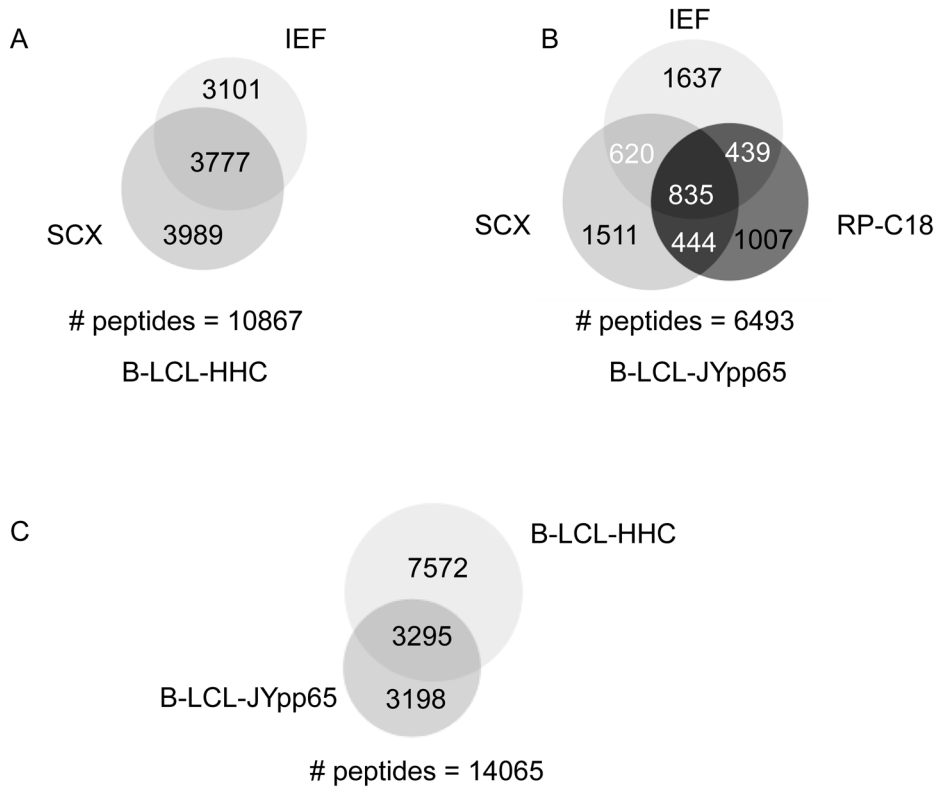


Figure 2. Venn diagrams of cumulative number of identified HLA-ligands. Cumulative number of identified unique peptides per cell line for B-LCL HHC (A) and cell line B-LCL JYpp65 (B). Cumulative number of identified unique peptides for B-LCLs JYpp65 and HHC (C). The area of the circles corresponds with the number of peptides. Total, fractional and overlapping numbers of unique peptides are indicated

epitope is presented in HLA-B7 [26]. Again the identification of these two epitopes, with BMIs of 28 and 12, was confirmed by matching the tandem mass spectra of the eluted peptides with their synthetic counterparts. In summary, we identified four new EBV-derived ligands peptides and two known epitopes derived from the transduced CMVpp65 protein. The EBV peptides might be candidates for new EBV epitopes. It is important to note that low scoring peptides should not be discarded lightly in this field of application, without checking their functional relevance.

Independent immunological testing in combination with the synthesis of good candidates provides a good strategy to exclude false positive identifications without losing the immunologically relevant false negative identifications.

Together, the above clearly illustrates that our list of ligands harbours immunologically relevant peptides, suggesting that many more relevant peptides are in this list.

Post-translationally modified HLA ligands

The presentation of post-translationally modified peptides has been reported before [22, 24, 43], although not in large numbers. Therefore, we investigated our tandem mass spectra for the presence of phosphorylation of serine, threonine and tyrosine, or cysteinylation of cysteine residues, since both modifications have been

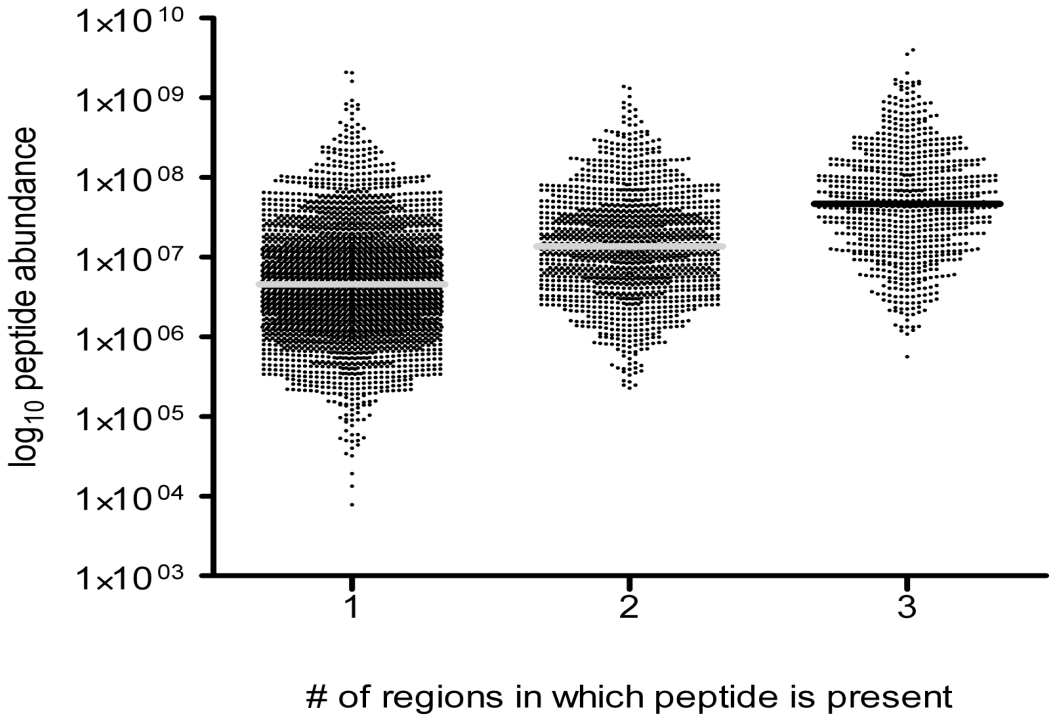


Figure 3. Shotgun nature of the experiment. The weighted intensity distribution of the peptides found in the various regions in the Venn diagrams of Fig. 2B. Three regions were defined (X-axis): (1) peptides found in only a unique first dimension separation technique (i.e. the 'outer' regions), (2) peptides found in either of two overlapping regions, and (3) peptides always found irrespective of the first dimension separation technique. Average peptide intensity increases from region 1 to 3, as expected in a shotgun-type of proteomics experiment. The median peptide intensity is indicated.

reported to influence T cell recognition [23, 43].

Aberrant phosphorylation has been implied in cancer. Inclusion of phosphorylation in our database matching process yielded 451 phosphopeptide hits of lengths 8 to 11 amino acids with a BMI>35 (phosphorylation on anchor positions was not allowed), of which 221 were estimated to be correctly identified, see supplemental Table S4 and discussion. Similarly, a cysteinylated cysteine as PTM was set as a modification in the database matching, which resulted in 1221 identified peptides with a cysteinylated cysteine (cysteinylation on anchor positions was not allowed). The peptides are listed in supplemental Table S5.

Phosphorylated HLA ligands

Inclusion of phosphorylation in the database matching process yielded 451 phosphopeptide hits of length 8 to 11 amino acids with a best mascot score of greater or equal to 35, as listed in supplemental Table S4. The distribution was as expected: 267 on serine, 154 on threonine and 30 on tyrosine. Most peptides were singly phosphorylated, 13 peptides were doubly phosphorylated.

To evaluate if these phosphopeptides were properly assigned we compared our results with previously reported phosphopeptides at Phosphosite.org (http://www.phosphosite.org). This yielded 72 hits. Phospho.ELM 8.3 (http://phospho.elm.eu.org) yielded 38 hits, all of which were also present in phosphosite.org. Next we used

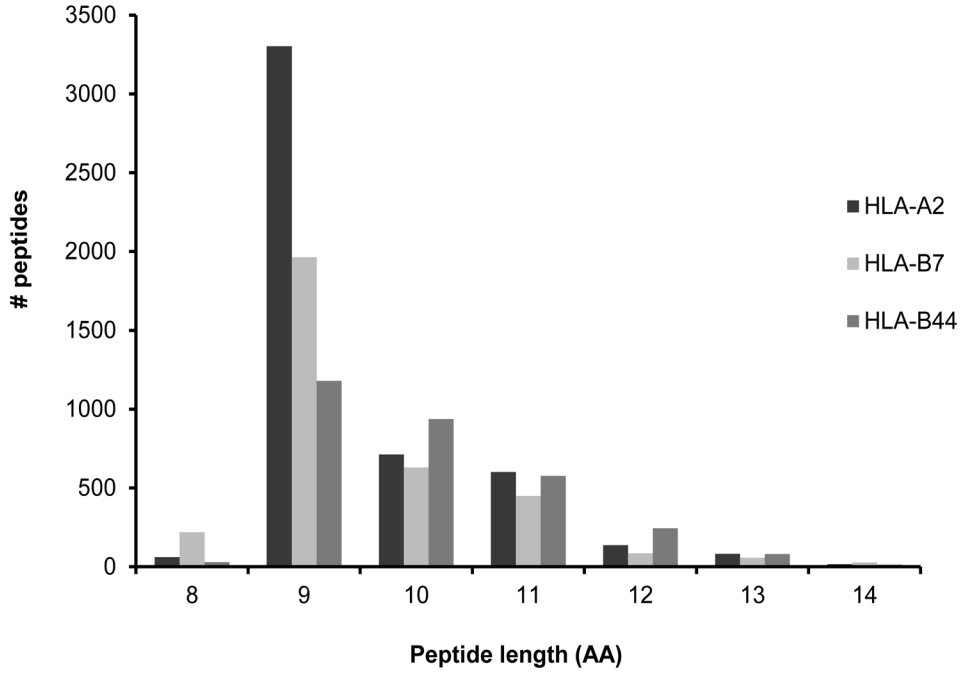


Figure 4. Peptide length distribution of the HLA-ligandome. Of note are the longer peptides and the relative high number of HLA-B7 binding peptides of 8-mer length, and the relatively wide spreading of HLA-B44 binding peptides over 9-11 amino acids length. Number of 13-mers approximately 80 per allele and 14-mers approximately 20 per allele.

NetPhos 2.0 to predict phosphosites in the peptides, which resulted in 83 phosphosites, of which 28 overlapped with the 72 phosphosites detected with the two other sites mentioned above. In total 127 out of 451 phosphopeptides with either a known or predicted phosphosite were found.

As shown in Figure. 7 proteins can be represented by several peptides in the HLA-ligandome. Here three proteins were chosen that are represented by 7, 5 and 5 peptides in HLA in B-LCL HHC. As can be seen these peptides can be presented by different alleles, have greatly differing hydrophobicity. Importantly, peptides derived from that same protein have greatly differing abundance. So, the abundance of a HLA-ligand derived from a certain protein does not simply correlate with its protein intensity.

To confirm the identification of the phosphorylated peptides in our dataset we chose to synthesize a selection of peptides and compare the synthetic phosphopeptides with their eluted counterparts. 26 peptides were selected, of which 9 peptides with a known phosphosite or predicted phosphosite and 17 peptides without a known or predicted phosphosite with varying phosphorylation position in the peptide sequence and different mascot ion scores. These were synthesized and the MS/MS spectra of these peptides were matched with their eluted counterparts. The 9 peptides with a known or predicted phosphosite appeared to be correctly identified.

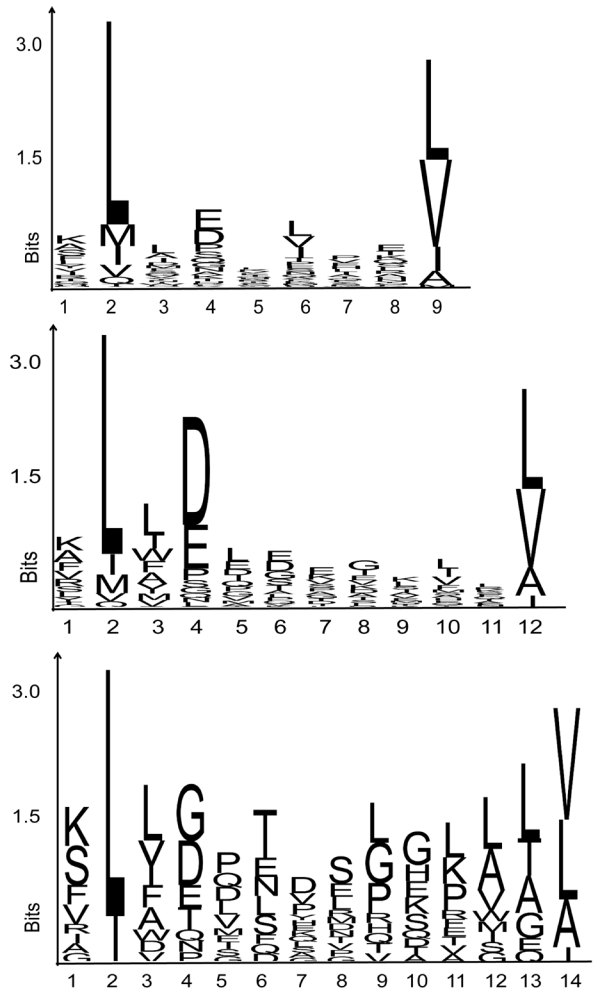


Figure 5. Binding motif of various HLA-A2 length variants. Icelogo plots of HLA-A2 binding peptides of 9-mer (N=3302), 12-mer (N=137), and 14-mer (N=17) length. Peptides of different length clearly display a similar binding motif.

Of the other 17 peptides without known of predicted phosphosite, 29 % (5 out of 17) were found to be correctly identified.

First the 127 hits found in our study and which were also either known from previously identified proteomics studies or predicted, were extrapolated to be correctly identified, leaving (451-127)= 324 hits. Of these 324 hits 29% (=94 hits) was extrapolated to be correct. Therefore, we estimate that 221 out of 451 (49%) have been correctly identified.

Cysteinylation of HLA ligands

To find peptides with a cysteinylated cysteine as PTM in our eluted dataset we searched the dataset obtained from B-LCL-HHC and B-LCL-JYpp65 cell lines with variable modification on cysteinylation (+119 Da). We found 1221 identified peptides with cysteinylation, see supplemental Table S5. Since the binding affinity of the peptides probably does not change due to cysteinylation we used NetMHC server to check the binding affinity of these peptides. The vast majority was predicted to bind to either HLA-A2 (594), HLA-B7 (143) or HLA-B44 (294).

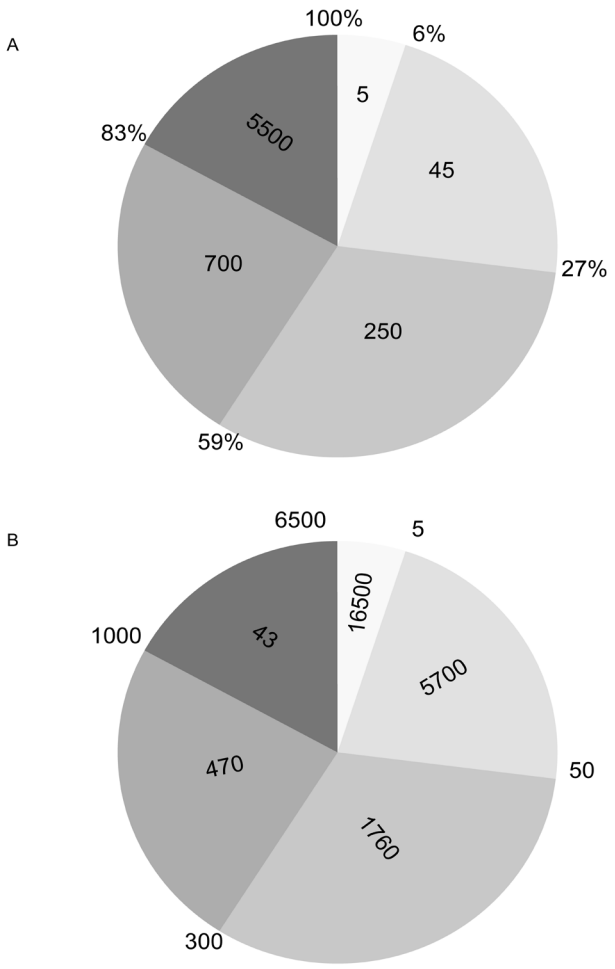


Figure 6. Occupation of the ligandome. The intensity distribution of peptides constituting the ligandome is shown for B-LCL JYpp65. (A) The outer rim of the pie chart displays the percentage occupation of the HLA-ligandome and the number within each pie segment is the number of unique peptides in that segment Of the total number of 6,500 identified peptides, the top 5 abundant proteins occupy 6% of all HLA molecules. The top 50 abundant peptides occupy 27% of all HLA-molecules, etc. Note that the lowest abundant 5,500 peptides occupy only 17% of all HLA-molecules. (B) The numbers in (A) have been converted to copy number, i.e. the number of ligands per cell. The outer rim of the pie chart displays the number of unique peptides and the number within each pie segment is the average copy number for the peptides in that segment The top 5 abundant peptides are on average present with 16500 copies on the cell surface. Peptides 5-50 are on average present with 5700 copies, etc. The average copy number of the lower abundant peptides is 43, the very lower abundant peptides are present at copy numbers close to 1 or even less.

DISCUSSION

The quality of the data

The quality of the elucidated ligandome was checked on three independent levels. First, by application of false discovery rates. Second, by consideration of chromatographic criteria in combination with the correct distribution over physical parameters. Third by the predicted binding strength. Additionally, the presence of biologically relevant peptides illustrates the good quality of our dataset, but these peptides will be discussed separately.

Mass spectrometry in HLA-ligandomics and choice of false discovery rates

HLA-presented peptides arise from the complex antigen processing process, involving many enzyme specificities. Therefore, enzyme restriction during the database matching process is not possible, leading to a large increase in database search space. The best way to reduce the number of falsely identified peptides is to apply a strict precursor mass tolerance of 2 ppm, which was achieved by including a SIM scan in the FTMS-measurements. Strict application of a tight FDR of 1%, or even 5%, as in trypsin-based proteomics experiments, leads to loss of a wealth of valuable peptides (false negatives) as we will illustrate below. To our opinion this is unacceptable in immunological workflows. We have chosen for a BMI>35, which we know from experience

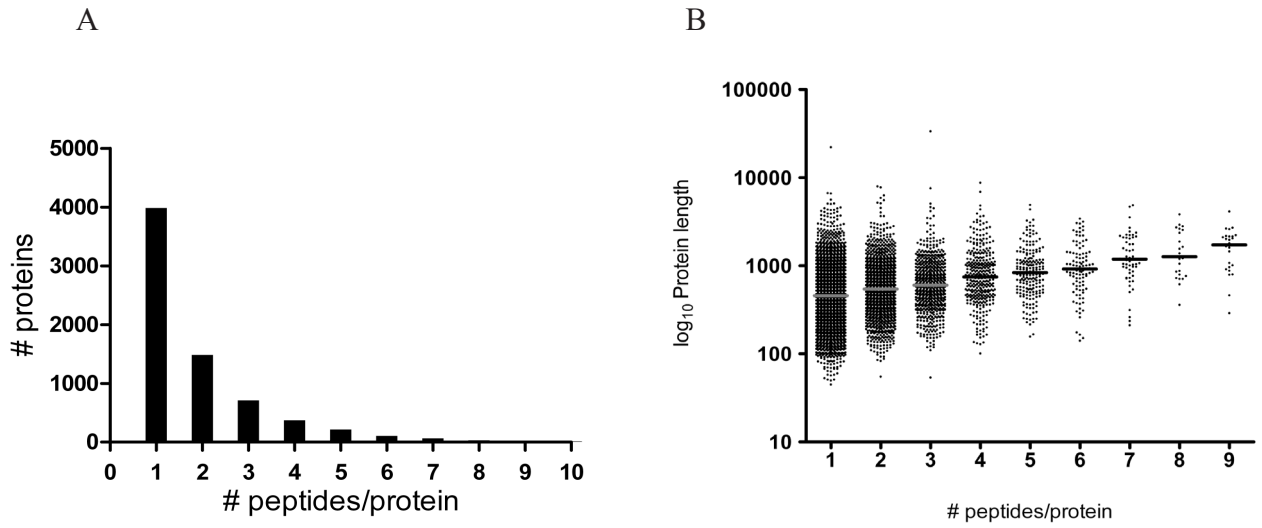


Figure 7. Representation of proteins in the HLA-ligandome. (A)This figure depicts the number of proteins as a function of the number of peptides identified per protein. HLA-ligands derived from 7059 unique proteins are in our dataset. In our dataset about 50% of the proteins is represented by 1 peptide, 20% is represented by 2 peptides, 10% is represented by 3 peptides, and so on. As can be seen there is a gradual decrease in the number of peptides by which a protein is represented. Up to 41 peptides/protein are present in our dataset. (B) The protein length is depicted as a function of the number of peptides by which the protein is represented in HLA. The number of peptides in HLA of a protein correlates well with its protein mass.

is a decent starting cut-off. Many known immunologically relevant peptide ligands are still found far below this limit. In supplemental Table S1, we listed peptides with a BMI>35, but also indicated below what FDR percentage the peptides are, after searching either the smaller SWISSProt (homo sapiens) or the larger IPIhuman database. We included 1%, 5% and 10% FDR. The results of these searches and the effect of the different search conditions is summarized in supplemental Table S2. A 10% FDR has e.g. been applied by Dudek et al. in a HLA class I-peptidomics-based study [44]. An FDR of even 37% has been used in a HLA-class II-based peptidomics study by Chornoguz et al. [45]. Importantly, in targeted immunological approaches, i.e. searching for ligands/epitopes or post-translationally modified epitopes derived from a specific antigen, we advise not to use a strict statistical cut-off score, be it a BMI or a false discovery rate during the first discovery phase. In the next step first round 'hits' should be tested independently, be it by MS/MS analysis of the synthetic candidate epitope or by a biological assay, e.g. a T cell activity assay. As illustrated by the BMI scores < 35 for 3 out of 10 validated MiHA (and 6 out of 10 with BMI<39), and both CMV peptides, with one CMV peptide having a BMI of 12 (!), low scoring peptides should be prevented from being assigned as false negatives. It is important to note that initially wrongly assigned peptides will immediately be filtered out by immunological follow-up experiments. Next to the MS-methodology available, additional non-MS checkpoints were applied to estimate the value of a peptide, see below.

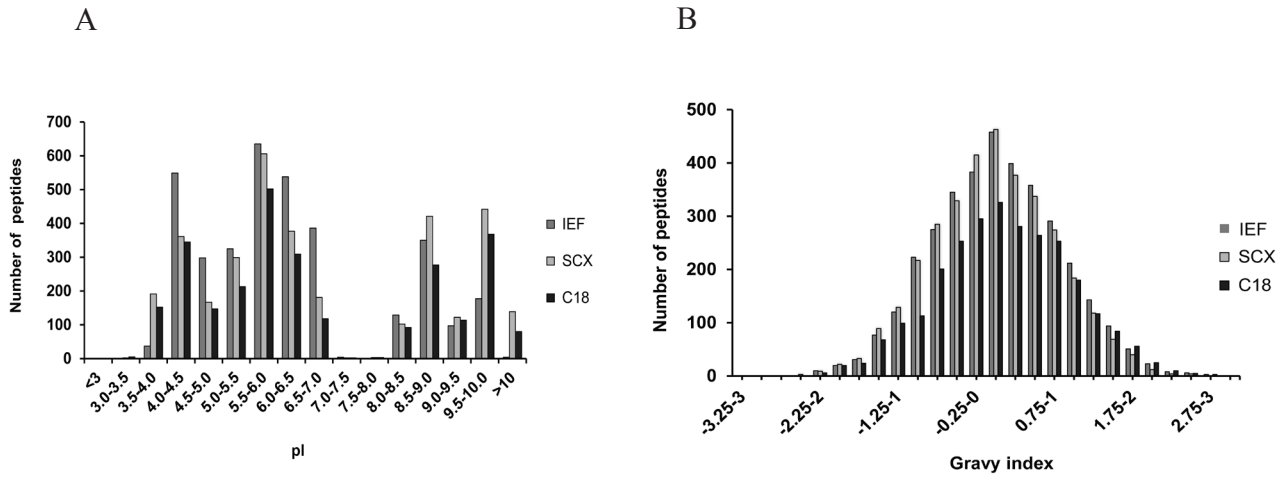


Figure 8. HLA-ligandome physicochemical properties. (A) Distribution of HLA-peptides over the pI range as a function of the first dimension separation technique used. As can be seen IEF, SCX en RP-C18 do not discriminate for peptides based on their theoretical pI. The only exception is that peptides with low pI are clearly underrepresented in the IEF process. This is caused by the actual pI in the pI 3-10 strips, which in fact appear to be pI 4-10, as also independently checked with a pH-indicator after isoelectric focusing. (B) Distribution of HLA-peptides over the hydrophobicity range depending on the first dimension separation technique used. As can be seen IEF, SCX en RP-C18 do not discriminate for peptides based on their hydrophobicity. The above plots illustrate our HLA-ligandome is a good representation of the real HLA-ligandome.

Chromatography and physicochemical properties

Sample (pre)treatment might inadvertently favor peptides with certain physicochemical properties. Therefore, we checked the distribution of our peptides over the hydrophobicity scale (GRAVY index), see Figure. 8 (B) and pI scale, see Figure. 8 (A). The peptides were evenly distributed, so sample pretreatment has not favored peptides on the basis of their physicochemical properties. However, peptides at the pH extremes of the IEF strips were lost. It appeared that the real pH range of the strips is not 3-10, but instead are closer to pH 4-9, which explains their loss. The hydrophobicity distribution of the peptides is shown in Figure. 9 for peptides eluted from each of the three alleles used in this study. The centre of the hydrophobicity distribution is clearly shifted to the hydrophobic side for HLA-A2, and shifted to the hydrophilic side for HLA-B44. The HLA-B7 peptides are at intermediate hydrophobicity. This is in perfect agreement with the general notion that HLA-A2 presented peptides are more hydrophobic [46], while HLA-B44 derived peptides are hydrophilic and HLA-B7 peptides have intermediate hydrophobicity [47-50]. The relative position of the hydrophobicity distributions supports the correct overall assignment of the ligandome.

In addition to the identification of the peptides we determined their relative intensity. From the shotgun nature of the experiments, peptides found in three experiments (the 835 peptides found Figure. 2 (B) are expected to have a higher intensity than those peptides present in only one experiment. The results as shown in Figure. 3 clearly illustrate that this is the case in our experiments.

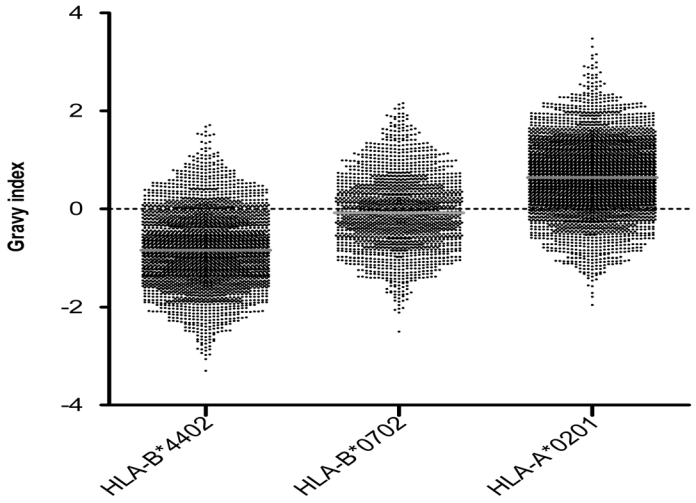


Figure 9. Allele-specific distribution of HLA-peptides over the hydrophobicity range. These plots show a clear average hydrophobicity shift going from HLA-B44 (most hydrophilic, median gravy-index of -0.84) via HLA-B7 (median gravy index of -0.08) to HLA-A2 (median gravy index of 0.64). The positions of the three distributions are as expected and illustrate the overall quality of our dataset.

Binding prediction

Because of the large number of peptides found, peptide identities cannot be verified by MS/MS of their synthetic counterpart. Alternatively, the NetMHC algorithm was used to predict binding affinity. As can be seen in supplemental Table S1 the large majority of our reported peptides has a low NetMHC score, i.e. is predicted to be a (strong) binder. Although at the NetMHC-website a score <50 is considered a strong binder and between 50 and 500 a weak binder, from our experience with known epitopes, we know this value is too strict.

From the mass spectral quality, the distribution of physicochemical parameters, the expected shotgun distribution, in which the overlapping regions in the Venn diagram represent the most abundant peaks and the excellent binding affinity we are convinced that the sample pretreatment and mass spectrometric analysis have not resulted in skewing of the identified ligandome as a whole. The fact that many known biologically relevant peptides are found is another strong indication of the high quality of our ligandome.

Ligandome characteristics

Our large dataset opens up opportunities to discuss a number of properties of the HLA-ligandome. It is important to realize that the peptides we report here are the survivors of the intracellular processing. Therefore, this listing is valuable for studies on antigen processing, to all *in silico* studies on ligand/epitope definition and the definition of antigen processing rules. In addition, it can be used to look up if peptides from specific protein antigens can survive the intracellular processing and are presented on the cell surface. Thus, it provides an indirect view of cellular processes. Many more peptides are probably presented in HLA-molecules, but may escape detection because of low abundance and/or a relatively low sensitivity in the electrospray ionization process.

The Venn diagrams of Figure. 2 still display relatively large regions unique to either 1st dimension separation technique, which shows that many peptides are just detected, or may still escape detection. Although the

Table 1. MiHA. Listing of known MiHA and allelic counterparts of known MiHA identified in our data set by searching the HSPV database (16), illustrating that immunological relevance of our dataset. In addition several potential MiHA are present in our dataset, which form the basis for finding new clinically relevant MiHA.

Gene name	HLA type	Peptide sequence	Eluted cell line	BMI
HA-2V	A*02:01	YIGEVLVSV	B-LCL-JY, B-LCL-HHC	50
HA-8R	A*02:01	RTLDKVLEV	B-LCL-JY	34
LB-SSR1-1S	A*02:01	VLFRGGPRGSLAVA	B-LCL-HHC	29
SSR1-1L*	A*02:01	VLFRGGPRGLLAVA	B-LCL-JY	37
LB-ERAP1-1R	B*07:02	HPRQEQIAL	B-LCL-JY, B-LCL-HHC	36
LB-GEMIN4-1V	B*07:02	FPALRFVEV	B-LCL-JY	43
SMCY-B7	B*07:02	SPSVDKARAEL	B-LCL-JY	58
EBI3-1V*	B*07:02	RPRARYYVQV	B-LCL-JY, B-LCL-HHC	26
PDCD11-1L *	B*07:02	GPDSSKTLLCL	B-LCL-JY	39
ACC-2G *	B*44	KEFEDGIINW	B-LCL-HHC	81

* allelic counterpart of known MiHA

ligandome presented here is the most comprehensive report to date, we interpret the occurrence of many unique peptides in a single type of 'precursor dimension' as a strong indication that the HLA-ligandome must contain many more peptides than we report. This notion is supported by our experience with 3- and 4-dimensional separations before MS-analysis in the classical forward approach to find T cell epitopes. Every additional chromatographic dimension yields peptides, not detected in 'precursor'-dimensions, indicating that a significant part of the presented HLA-ligandome still goes undetected. Based on the current listing of 14,000 peptides, the many false negatives, i.e. correct peptides with a BMI<35 and the peptides going undetected, we estimate the HLA-ligandome to comprise at least 3 times more peptides, i.e. at least 50,000 members.

Binding motif and peptide length

The ligandome contains a sizeable fraction of unusually long peptides for HLA class I binding. We found 5% of the peptides to be longer than 11 amino acids. This is in line with the results of Scull et al. on a smaller set of HLA-A2 peptides [7].

In supplemental Figure 1 the Icelogo graphs show the HLA groove can accommodate at least up to 14-mers. Peptides need to bulge out of the groove considerably to accommodate the extra amino acids between the fixed anchors at position 2 and the C-terminus, as beautifully shown by crystallographic data [51]. Our

data on the peptide length distribution, Figure. 4, underline the need to take into account longer peptides in the search for epitopes, both in practical MS-based strategies and *in silico* prediction approaches.

Intensity, dynamic range and number of peptides of the ligandome

The total peptide intensity (sum of all identified peptide intensities) of both cell lines was similar, and amounts to 3 nmol of total eluted peptide), which corresponds to approximately 3 ug of peptides. Our experience from previous elutions is that the amount of isolated HLA after immunoaffinity precipitation (IP) is 150-200 ng HLA/1E6 B-LCL-JY cells, which is in line with the independently reported amount of 150 ng/1E6 cells [52]. Therefore, assuming 150 ng/1E6 cells, a yield of 3 mg HLA/2E10 cells, corresponding to 60 ug of peptides was expected. Therefore, the overall work-up efficiency is about 5%. From the above the copy number (ligands per cell) can easily be calculated. These numbers are shown in supplemental Table 1. Intensity extremes, expressed as peak area values in PD, range from 4E9 to 1E5, so the dynamic range of the peptide repertoire is quite high. Considering that the lowest intensity ligands are most prone to go undetected, the actual dynamic range is probably even higher. This translates to a range from 24,000 copies/cell to (<)1 copy/cell. Clearly the chance of a protein being represented by a peptide in the ligandome varies enormously.

Of cells available in limited amount, e.g. DCs, the ligandomes have been studied of only (!) a few tens of millions of cells [53, 54].The number of identified ligands, however, is low (a few hundreds). Considering the large dynamic range as discussed above, it is intrinsically impossible to find the relatively low intensity ligands using small cell populations for full ligandomics experiments. An inventory of the possible ligandome on a large scale followed by MRM [55] on the selected small population might be a fruitful approach. However, 'more is better' definitely seems to apply to HLA-ligandomics.

The composition of the ligandome in terms of relative intensity of the peptides is shown in Figure. 6. Interestingly, it appears that in B-LCL JYpp65 the 5 most abundant peptides already 'fill' 6% of the HLA molecules. The top 50 abundant peptides 'fill' 27%, the top 300 'fill' 59% and the top 1000 'fill' 83%. Importantly, this leads to the observation that all other ligands, i.e. approximately 5,500 peptides, fill the final 17%, as shown in Figure. 6 (A). The intensities have been translated to copy number in Figure. 6 (B).

The 14,065 HLA-ligands are derived from 7,059 different proteins, a graphical representation is shown in Figure. 7 (A). Half the proteins is represented by 1 peptide while the other proteins are represented by more peptides. A gradual decline in peptide number/protein is seen. In Figure. 7 (B) the average protein length is plotted against the number of peptides/protein. 61 proteins are even represented by 10 peptides or more on the cell surface, with the exceptional case of 29 and 41 peptides/protein. The longer the protein, the more ligands are derived from it. It also appeared that the position of the ligands in their 'precursor' proteins is evenly distributed (data not shown). In addition, the intensities of the ligands derived from the same protein can differ

greatly, see Table 2 for a few examples, which makes it impossible to predict ligand copy number from overall protein copy number. This is in line with previous work on correlation between mRNA/protein expression and ligand density on the cell surface [56]. Furthermore, no correlation could be found between peptide intensity and binding strength to HLA. The former observations may seem logical, but these imply that there is no clear selection on the basis of protein properties during antigen processing. In fact, every part of the proteome seems equally suitable for sampling by the HLA-molecules. Selection of particular ligands takes place at another level, namely the binding affinity to the available HLA-molecules

From our peptide listing it is tantalizing to speculate on the origin of HLA-presented ligands and the cellular processing machinery and distill a glimpse of the cellular protein processing landscape. A courageous effort based on a rather limited data set has been made to describe the presence of peptides in relation to their corresponding protein turnover in terms of protein stability, DRiPs and SLIPs [57]. Studies on the scale as presented here might be required to provide a better view at the origin of HLA-presented peptides.

Protein turnover times, as in supplemental Table S1 cannot be correlated with HLA-ligand intensity. This lack of correlation might be seen as reinforcement of the DRiPs hypothesis [58]. Pulse chase experiments have shown that some HLA-presented peptides make it to the cell surface within 30 minutes of the pulse [57]. A general predictive description of the route from protein to its presented peptides in the HLA-ligandome is disturbed by the many underlying processes leading to antigen presentation, (transient) protein abundance, HLA-peptide complex stability and re-internalization of HLA-peptide complexes. Compartmentalization and multiple sources of peptides, and a subset of rapidly degraded polypeptides (RDPs) have been suggested to be needed to get a grip on the intricate antigen processing pathways [58]. However, the final result of the complex cellular protein machinery is the representation of virtually all proteins on the cell membrane. In this way all cellular proteins are under the scrutiny of T cell surveillance.

The biological value

Our reversed immunology approach based on the *bona fide* eluted ligandome was started to explore its possibilities for finding T cell epitopes. The above were global observations and considerations concerning the ligandome. Below the focus is on the immunological value of the data contained in our data set.

Comparison with sources of known epitopes

First we looked for known epitopes contained in the International Immuno Epitope Database, and the SYFPEITHI database for HLA-A2, HLA-B7 and HLA-B44. Comparison showed that 50% and 29% listed in the IEDB and SYFPEITHI database, respectively, were found in our ligandome, as indicated in supplemental Table S1. Considering that both the IEDB and the SYFPEITHI databases are a collection of ligands from many studies on a variety of cells, the biological value of our data set is clearly demonstrated. In addition,

there are probably many more epitopes in our list, which are to be explored in the field of immunology.

MiHA

The first category of biologically important T cell epitopes are the minor histocompatibility antigens, important in the treatment of hematological malignancies. Analysis of the data set, using our in house developed human short peptide variation database (HSPVdb) [21], led to the identification of approximately 1,400 polymorphic ligands. Among these were known MiHA, underlining the possibilities of our proteomics approach to find MiHA. From the set of 1400 polymorphic peptides 80 were selected for further study to investigate their potential for medical application. The details of the selection method and the immunological follow up will be described in detail elsewhere. The total length of all proteins in Refseq release 46, expressed in base pairs (bp), is 56,113,216 bp. The total number of non-synonymous SNPs and in/del within a coding sequence (CDS) is 296,901. Thus the chance that a nucleotide is polymorphous is 0.5%. Therefore, for a 10-mer peptide (30 nucleotides), the chance of not containing a SNP is $(0.995)^{30}$ is 0.86. Therefore, the chance of a 10-mer peptide containing at least one polymorphic amino acid is 14%. The number we experimentally found is 10%, so of the same order as the theoretical number. From our data there is no reason to suspect that polymorphic peptides represent a 'special case'. The same conclusion was drawn in the recent paper by D. Granados et al. on B-cells [9].

Virus-derived peptides

The antigen presenting cells in this study were EBV-transformed, and B-LCL-JY was also CMVpp65 transduced. Therefore, the presence of EBV- and CMVpp65-derived peptides was checked in our ligandome, and the results listed in supplemental Table S3. Four EBV-derived ligands and 2 CMVpp65 derived epitopes were found. All identifications were checked by comparison of the tandem mass spectra to those of their synthetic counterparts. All four EBV-derived ligands were newly found.

We expected that most of the previously reported EBV epitopes would not be detected, since HLA-A2 and HLA-B7 restricted T cell clones isolated from EBV seropositive individuals and specific for the different EBV proteins BZLF-1, BRLF-1, and EBNA-3A are not reactive or low reactive against HLA-A2 and HLA-B7 positive EBV-LCLs, whereas the T cell clones are reactive when the genes encoding these proteins are additionally introduced into EBV-LCLs. These results indicate that the published EBV epitopes are not or hardly presented by EBV transformed B cells, and therefore it is very well possible that these peptides were not detected in our elutions from HLA of the EBV-LCLs. The reason why these epitopes are not efficiently processed and presented is most likely due to the fact that only a minority of the cells in the EBV transformed cell lines is in a lytic cell cycle state. The four epitopes described in the manuscript are derived from other EBV proteins: BGLF4, RPMS1 (derived from an alternative reading frame) and EBNA3B/C and HS4ENVGP

(the latter two both derived from the UTR), and are indicated in supplemental Table S3. CMVpp65 specific HLA-A2 restricted T cells (peptide NLVPMVATV (found)) and CMV-pp65 specific HLA-B7 restricted T cells (peptides RPHERNGFTV (found) and TRPVTGGGAM (not found)) efficiently recognize pp65 transduced EBV-LCLs, indicating that the three pp65 peptides are processed and presented in HLA-A2 and HLA-B7. Two out of the three peptides were detected in our study, whereas one peptide was not. The exact reason why we did not detect this particular peptide is unknown, but this peptide may have a low expression or may have been missed due to the shotgun nature of the experiment.

Together, the presence of the IEDB matching peptides, the MiHA, and the virus-derived peptides in our list proves the relevance of our peptide listing for immunological research. The fact that not all previously reported epitopes were found in our analysis illustrates that still a considerable fraction of the ligandome goes undetected.

Post-translationally modified ligands

Inclusion of phosphorylation and cysteinylation as a PTM in the database matching process yielded approximately 221 phosphorylated ligands and 1221 cysteinylated ligands. Previously 36 and 18 HLA-A2 and 8 HLA-B7 phosphorylated peptides were reported by Zarling et al. [22, 43] and Meyer et al., [24] respectively. 15 peptides of these overlap with our dataset. In summary, 2 out of 150 ligands in our study are phosphorylated. Many cysteinylated ligands were identified. The total number of cysteine-containing (modified and unmodified) peptides is 1386, which represents 9% of all peptides (15286). The frequency of cysteine in the Swissprot human database is 2%. Therefore, we calculate the chance of a peptide ligand containing a cysteine residue to be 15% and 13% for a 10-mer ($1-(0.98)^8=0.15$) or 9-mer ($1-(0.98)^7=0.13$) ligand respectively (we excluded the two anchor positions for this calculation because cysteine is not favorable at anchor positions). 9% of our peptides contain a cysteine residue. Scull et al. find a cysteine residue in 4% (50 out of 1179 unique 8-11 mer peptides) of their identified peptides [7]. The numbers indicate that we are close to the theoretically expected number of cysteine-containing HLA-ligands. Whether the cysteinylation has occurred *in vivo* or *in vitro*, the identified cysteine-containing peptides are true identifications. Since there is no suitable criterion for (de) selecting biological relevance of these cysteinylated peptides we do not wish to discard these true identifications. Both phosphorylation and cysteinylation have been shown to play a role in T cell recognition [23, 24, 43]. The number of phosphorylated HLA-ligands described here is by far the largest reported to date. Considering we did not apply a specific phosphopeptide-directed work up procedure, such as described in [59], we expect a considerably larger number of phosphorylated ligands to be presented. Other PTM, not included in this study, will certainly raise the number of post-translationally modified ligands.

In summary, our in-depth HLA-ligandome study enabled a detailed look at antigen processing and provided relevant ligands. From the presented data we estimate the ligandome to comprise at least 50,000 ligands.

Considering the above number and the average number of peptides per protein being between 2 to 3, the intracellular peptide pool generated by protein breakdown, represents most, if not all, proteins, which seems the perfect way to present the cellular state on the outside of the cell for immune surveillance. Overall, every part of the proteome seems equally suited for sampling by the HLA-molecules, although even the abundance of peptides from the same protein varies greatly. The final composition of the ligandome is determined by the enzymes involved, the transporter associated with antigen presentation (TAP), the location of a particular protein, and the relative contribution of DriPs for a given protein. The available HLA-molecules select "what fits them best". Overall this process leads to 'delegates' from many proteins (5,000 found for B-LCL-HHC) in our set.

A rich and diverse repertoire of ligands is presented to the immune system, including a considerable number of post-translationally modified ligands, an additional 10% is derived from polymorphic peptides and 5% is >11 amino acids.

We have shown that the ligandome presented here can be a good starting point to solve a wealth of specific immunological questions. Our large list of peptides presented here fits in the efforts towards a human immunopeptidome project [25].

ACKNOWLEDGEMENT

We would like to dedicate this paper to Jack Leunissen, one of the first Dutch bioinformaticians, whom we had to say goodbye to in the course of this work. N. Dolezal and R. Cordfunke are thanked for providing the synthetic peptides. Dr. G. Janssen and F. Koning are thanked for critical reading of the manuscript. This research was made possible by the financial assistance of the Landsteiner Foundation for Blood Transfusion Research (LSBR0713).

Table 2. Peptide intensities derived from the same protein. Intensity of different HLA-presented peptides from some selected proteins.

Protein name	Peptide sequence	Intensity	HLA allele	BMI	NetMHC (nM)	Gravy
Isoform 1 of Clathrin heavy chain 1	NENSLFKSL	7.6E+7	B44	59	359	-0.622
	NPASKVIAL	2.6E+7	B7	51	13	0.7
	AEELFARKF	2.3E+7	B44	39	351	-0.267
	EEAFAIFRKF	6.4E+6	B44	47	17	0.11
	AETGQVQKI	5.3E+6	B44	41	671	-0.556
	FVHDLVLYL	3.3E+6	A2	41	13	1.622
	LLMVLSPRL	1.1E+6	A2	43	16	1.6
Mediator of RNA polymerase II transcription subunit 24	SLVEQLTMV	2.0E+7	A2	70	12	1.044
	SPTDRSMSSSL	1.9E+7	B7	56	10	-0.782
	KERWSDYQW	6.4E+6	B44	36	123	-2.533
	AEQMHKTGF	3.6E+6	B44	53	594	-0.967
	SLMDPPGTAL	1.3E+6	A2	35	31	0.27
Isoform 1 of Dedicator of cytokinesis protein 2	KMWEEAISL	3.4E+8	A2	58	9	-0.067
	SPRLSQTFLQL	5.3E+7	B7	52	21	-0.109
	IPGMVGPIL	5.2E+7	B7	36	89	1.656
	LEAYIQQHF	4.3E+6	B44	39	145	-0.233
	ILLQVAALKYI	1.7E+6	A2	49	63	1.773

REFERENCES

1. Kenter, G.G., Welters, M.J., Valentijn, A.R., Lowik, M.J., Berends-van der Meer, D.M., Vloon, A.P., Essahsah, F., Fathers, L.M., Offringa, R., Drijfhout, J.W. et al. (2009) Vaccination against HPV-16 oncoproteins for vulvar intraepithelial neoplasia. N Engl J Med, 361, 1838-1847.
2. Falkenburg, J.H., Marijt, W.A., Heemskerk, M.H. and Willemze, R. (2002) Minor histocompatibility antigens as targets of graft-versus-leukemia reactions. Curr Opin Hematol, 9, 497-502.
3. Yewdell, J.W., Reits, E. and Neefjes, J. (2003) Making sense of mass destruction: quantitating MHC class I antigen presentation. Nat Rev Immunol, 3, 952-961.
4. Kessler, J.H., Khan, S., Seifert, U., Le Gall, S., Chow, K.M., Paschen, A., Bres-Vloemans, S.A., de Ru, A., van Montfoort, N., Franken, K.L. et al. (2011) Antigen processing by nardilysin and thimet oligopeptidase generates cytotoxic T cell epitopes. Nat Immunol, 12, 45-53.
5. Vigneron, N., Stroobant, V., Chapiro, J., Ooms, A., Degiovanni, G., Morel, S., van der Bruggen, P., Boon, T. and Van den Eynde, B.J. (2004) An antigenic peptide produced by peptide splicing in the proteasome. Science, 304, 587-590.
6. van Bergen, C.A.M., Kester, M.G.D., Jedema, I., Heemskerk, M.H.M., van Luxemburg-Heijs, S.A.P., Kloosterboer, F.M., Marijt, W.A.E., de Ru, A.H., Schaafsma, M.R., Willemze, R. et al. (2007) Multiple myeloma-reactive T cells recognize an activation-induced minor histocompatibility antigen encoded by the ATP-dependent interferon-responsive (ADIR) gene. Blood, 109, 4089-4096.
7. Scull, K.E., Dudek, N.L., Corbett, A.J., Ramarathinam, S.H., Gorasia, D.G., Williamson, N.A. and Purcell, A.W. (2012) Secreted HLA recapitulates the immunopeptidome and allows in-depth coverage of HLA A*02:01 ligands. Mol Immunol, 51, 136-142.
8. Hillen, N., Mester, G., Lemmel, C., Weinzierl, A.O., Muller, M., Wernet, D., Hennenlotter, J., Stenzl, A., Rammensee, H.G. and Stevanovic, S. (2008) Essential differences in ligand presentation and T cell epitope recognition among HLA molecules of the HLA-B44 supertype. Eur J Immunol, 38, 2993-3003.
9. Granados, D.P., Yahyaoui, W., Laumont, C.M., Daouda, T., Muratore-Schroeder, T.L., Cote, C., Laverdure, J.P., Lemieux, S., Thibault, P. and Perreault, C. (2012) MHC I-associated peptides preferentially derive from transcripts bearing miRNA response elements. Blood, 119, e181-191.
10. Bade-Doeding, C., Cano, P., Huyton, T., Badrinath, S., Eiz-Vesper, B., Hiller, O. and Blasczyk, R. (2011) Mismatches outside exons 2 and 3 do not alter the peptide motif of the allele group B*44:02P. Hum Immunol, 72, 1039-1044.
11. Mester, G., Hoffmann, V. and Stevanovic, S. (2011) Insights into MHC class I antigen processing gained from large-scale analysis of class I ligands. Cell Mol Life Sci, 68, 1521-1532.
12. den Haan, J.M.M., Meadows, L.M., Wang, W., Pool, J., Blokland, E., Bishop, T.L., Reinhardus, C., Shabanowitz, J., Offringa, R., Hunt, D.F. et al. (1998) The minor histocompatibility antigen HA-1: A diallelic gene with a single amino acid polymorphism. Science, 279, 1054-1057.
13. den Haan, J.M., Sherman, N.E., Blokland, E., Huczko, E., Koning, F., Drijfhout, J.W., Skipper, J., Shabanowitz, J., Hunt, D.F., Engelhard, V.H. et al. (1995) Identification of a graft versus host disease-associated human minor histocompatibility antigen. Science, 268, 1476-1480.
14. Warren, E.H., Vigneron, N.J., Gavin, M.A., Coulie, P.G., Stroobant, V., Dalet, A., Tykodi, S.S., Xuereb, S.M., Mito, J.K., Riddell, S.R. et al. (2006) An antigen produced by splicing of noncontiguous peptides in the reverse order. Science, 313, 1444-1447.
15. Bansal, A., Carlson, J., Yan, J.Y., Akinsiku, O.T., Schaefer, M., Sabbaj, S., Bet, A., Levy, D.N., Heath, S., Tang, J.M. et al. (2010) CD8 T cell response and evolutionary pressure to HIV-1 cryptic epitopes derived from antisense transcription. J Exp Med, 207, 51-59.
16. Berger, C.T., Carlson, J.M., Brumme, C.J., Hartman, K.L., Brumme, Z.L., Henry, L.M., Rosato, P.C., Piechocka-Trocha, A., Brockman, M.A., Harrigan, P.R. et al. (2010) Viral adaptation to immune selection pressure by HLA class I-restricted CTL responses targeting epitopes in HIV frameshift sequences. J Exp Med, 207, 61-75.
17. Hombrink, P., Hadrup, S.R., Bakker, A., Kester, M.G., Falkenburg, J.H., von dem Borne, P.A., Schumacher, T.N. and Heemskerk, M.H. (2011) High-throughput identification of potential minor histocompatibility antigens by MHC tetramer-based screening: feasibility and limitations. PLoS One, 6, e22523.
18. Hadrup, S.R., Bakker, A.H., Shu, C.J., Andersen, R.S., van Veluw, J., Hombrink, P., Castermans, E., Thor Straten, P., Blank, C., Haanen, J.B. et al. (2009) Parallel detection of antigen-specific T-cell responses by multidimensional encoding of MHC multimers. Nat Methods, 6, 520-526.

19. Hadrup, S.R. and Schumacher, T.N. (2010) MHC-based detection of antigen-specific CD8+ T cell responses. Cancer Immunol Immunother, 59, 1425-1433.
20. Kersey, P.J., Duarte, J., Williams, A., Karavidopoulou, Y., Birney, E. and Apweiler, R. (2004) The International Protein Index: an integrated database for proteomics experiments. Proteomics, 4, 1985-1988.
21. Nijveen, H., Kester, M.G., Hassan, C., Viars, A., de Ru, A.H., de Jager, M., Falkenburg, J.H., Leunissen, J.A. and van Veelen, P.A. (2011) HSPVdb--the Human Short Peptide Variation Database for improved mass spectrometry-based detection of polymorphic HLA-ligands. Immunogenetics, 63, 143-153.
22. Zarling, A.L., Ficarro, S.B., White, F.M., Shabanowitz, J., Hunt, D.F. and Engelhard, V.H. (2000) Phosphorylated peptides are naturally processed and presented by major histocompatibility complex class I molecules in vivo. J Exp Med, 192, 1755-1762.
23. Meadows, L., Wang, W., den Haan, J.M., Blokland, E., Reinhardus, C., Drijfhout, J.W., Shabanowitz, J., Pierce, R., Agulnik, A.I., Bishop, C.E. et al. (1997) The HLA-A*0201-restricted H-Y antigen contains a posttranslationally modified cysteine that significantly affects T cell recognition. Immunity, 6, 273-281.
24. Meyer, V.S., Drews, O., Gunder, M., Hennenlotter, J., Rammensee, H.G. and Stevanovic, S. (2009) Identification of natural MHC class II presented phosphopeptides and tumor-derived MHC class I phospholigands. J Proteome Res, 8, 3666-3674.
25. Admon, A. and Bassani-Sternberg, M. (2011) The Human Immunopeptidome Project, a Suggestion for yet another Postgenome Next Big Thing. Mol Cell Proteomics, 10.
26. Kondo, E., Akatsuka, Y., Kuzushima, K., Tsujimura, K., Asakura, S., Tajima, K., Kagami, Y., Kodera, Y., Tanimoto, M., Morishima, Y. et al. (2004) Identification of novel CTL epitopes of CMV-pp65 presented by a variety of HLA alleles. Blood, 103, 630-638.
27. Heemskerk, M.H., de Paus, R.A., Lurvink, E.G., Koning, F., Mulder, A., Willemze, R., van Rood, J.J. and Falkenburg, J.H. (2001) Dual HLA class I and class II restricted recognition of alloreactive T lymphocytes mediated by a single T cell receptor complex. Proc Natl Acad Sci U S A, 98, 6806-6811.
28. Meiring, H.D., van der Heeft, E., ten Hove, G.J. and de Jong, A.P.J.M. (2002) Nanoscale LC–MS(n): technical design and applications to peptide and protein analysis. Journal of Separation Science, 25, 557-568.
29. (2012) Reorganizing the protein space at the Universal Protein Resource (UniProt). Nucleic Acids Res, 40, D71-75.
30. Colaert, N., Helsens, K., Martens, L., Vandekerckhove, J. and Gevaert, K. (2009) Improved visualization of protein consensus sequences by iceLogo. Nat Methods, 6, 786-787.
31. Cambridge, S.B., Gnad, F., Nguyen, C., Bermejo, J.L., Kruger, M. and Mann, M. (2011) Systems-wide proteomic analysis in mammalian cells reveals conserved, functional protein turnover. J Proteome Res, 10, 5275-5284.
32. Hornbeck, P.V., Kornhauser, J.M., Tkachev, S., Zhang, B., Skrzypek, E., Murray, B., Latham, V. and Sullivan, M. (2012) PhosphoSitePlus: a comprehensive resource for investigating the structure and function of experimentally determined post-translational modifications in man and mouse. Nucleic Acids Res, 40, D261-270.
33. Dinkel, H., Chica, C., Via, A., Gould, C.M., Jensen, L.J., Gibson, T.J. and Diella, F. (2011) Phospho.ELM: a database of phosphorylation sites--update 2011. Nucleic Acids Res, 39, D261-267.
34. Blom, N., Gammeltoft, S. and Brunak, S. (1999) Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. J Mol Biol, 294, 1351-1362.
35. Pruitt, K.D., Tatusova, T. and Maglott, D.R. (2005) NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Res, 33, D501-504.
36. Hubner, N.C., Ren, S. and Mann, M. (2008) Peptide separation with immobilized pI strips is an attractive alternative to in-gel protein digestion for proteome analysis. Proteomics, 8, 4862-4872.
37. Vita, R., Zarebski, L., Greenbaum, J.A., Emami, H., Hoof, I., Salimi, N., Damle, R., Sette, A. and Peters, B. (2010) The immune epitope database 2.0. Nucleic Acids Res, 38, D854-862.
38. Rammensee, H., Bachmann, J., Emmerich, N.P., Bachor, O.A. and Stevanovic, S. (1999) SYFPEITHI: database for MHC ligands and peptide motifs. Immunogenetics, 50, 213-219.
39. Bleakley, M. and Riddell, S.R. (2004) Molecules and mechanisms of the graft-versus-leukaemia effect. Nat Rev Cancer, 4, 371-380.
40. Goulmy, E. (1996) Human minor histocompatibility antigens. Curr Opin Immunol, 8, 75-81.
41. Mutis, T. and Goulmy, E. (2002) Hematopoietic system-specific antigens as targets for cellular immunotherapy of hematological malignancies. Semin Hematol, 39, 23-31.
42. Laurin, D., Hannani, D., Pernollet, M., Moine, A., Plumas, J., Bensa, J.C., Cahn, J.Y. and Garban, F. (2010)

Immunomonitoring of graft-versus-host minor histocompatibility antigen correlates with graft-versus-host disease and absence of relapse after graft. Transfusion, 50, 418-428.

43. Zarling, A.L., Polefrone, J.M., Evans, A.M., Mikesh, L.M., Shabanowitz, J., Lewis, S.T., Engelhard, V.H. and Hunt, D.F. (2006) Identification of class I MHC-associated phosphopeptides as targets for cancer immunotherapy. Proc Natl Acad Sci U S A, 103, 14889-14894.
44. Dudek, N.L., Tan, C.T., Gorasia, D.G., Croft, N.P., Illing, P.T. and Purcell, A.W. (2012) Constitutive and Inflammatory Immunopeptidome of Pancreatic beta-Cells. Diabetes.
45. Chornoguz, O., Gapeev, A., O'Neill, M.C. and Ostrand-Rosenberg, S. (2012) Major Histocompatibility Complex Class II+ Invariant Chain Negative Breast Cancer Cells Present Unique Peptides That Activate Tumor-specific T Cells From Breast Cancer Patients. Mol Cell Proteomics.
46. Hunt, D.F., Henderson, R.A., Shabanowitz, J., Sakaguchi, K., Michel, H., Sevilir, N., Cox, A.L., Appella, E. and Engelhard, V.H. (1992) Characterization of peptides bound to the class I MHC molecule HLA-A2.1 by mass spectrometry. Science, 255, 1261-1263.
47. Maier, R., Falk, K., Rotzschke, O., Maier, B., Gnau, V., Stevanovic, S., Jung, G., Rammensee, H.G. and Meyerhans, A. (1994) Peptide motifs of HLA-A3, -A24, and -B7 molecules as determined by pool sequencing. Immunogenetics, 40, 306-308.
48. Huczko, E.L., Bodnar, W.M., Benjamin, D., Sakaguchi, K., Zhu, N.Z., Shabanowitz, J., Henderson, R.A., Appella, E., Hunt, D.F. and Engelhard, V.H. (1993) Characteristics of endogenous peptides eluted from the class I MHC molecule HLA-B7 determined by mass spectrometry and computer modeling. J Immunol, 151, 2572-2587.
49. Macdonald, W., Williams, D.S., Clements, C.S., Gorman, J.J., Kjer-Nielsen, L., Brooks, A.G., McCluskey, J., Rossjohn, J. and Purcell, A.W. (2002) Identification of a dominant self-ligand bound to three HLA B44 alleles and the preliminary crystallographic analysis of recombinant forms of each complex. FEBS Lett, 527, 27-32.
50. DiBrino, M., Parker, K.C., Margulies, D.H., Shiloach, J., Turner, R.V., Biddison, W.E. and Coligan, J.E. (1995) Identification of the peptide binding motif for HLA-B44, one of the most common HLA-B alleles in the Caucasian population. Biochemistry, 34, 10130-10138.
51. Liu, Y.C., Chen, Z., Burrows, S.R., Purcell, A.W., McCluskey, J., Rossjohn, J. and Gras, S. (2012) The energetic basis underpinning T-cell teceptor tecognition of a super-bulged peptide bound to a major histocompatibility complex class I molecule. J Biol Chem.
52. Engelhard, V.H., Strominger, J.L., Mescher, M. and Burakoff, S. (1978) Induction of secondary cytotoxic T lymphocytes by purified HLA-A and HLA-B antigens reconstituted into phospholipid vesicles. Proc Natl Acad Sci U S A, 75, 5688-5691.
53. Meiring, H.D., Kuipers, B., van Gaans-van den Brink, J.A.M., Poelen, M.C.M., Timmermans, H., Baart, G., Brugghe, H., van Schie, J., Boog, C.J.P., de Jong, A.P.J.M. et al. (2005) Mass Tag-Assisted Identification of Naturally Processed HLA Class II-Presented Meningococcal Peptides Recognized by CD4+ T Lymphocytes. The Journal of Immunology, 174, 5636-5643.
54. van Haren, S.D., Herczenik, E., ten Brinke, A., Mertens, K., Voorberg, J. and Meijer, A.B. (2011) HLA-DR-presented peptide repertoires derived from human monocyte-derived dendritic cells pulsed with blood coagulation factor VIII. Mol Cell Proteomics, 10, M110 002246.
55. Tan, C.T., Croft, N.P., Dudek, N.L., Williamson, N.A. and Purcell, A.W. (2011) Direct quantitation of MHC-bound peptide epitopes by selected reaction monitoring. Proteomics, 11, 2336-2340.
56. Weinzierl, A.O., Lemmel, C., Schoor, O., Muller, M., Kruger, T., Wernet, D., Hennenlotter, J., Stenzl, A., Klingel, K., Rammensee, H.G. et al. (2007) Distorted relation between mRNA copy number and corresponding major histocompatibility complex ligand density on the cell surface. Mol Cell Proteomics, 6, 102-113.
57. Milner, E., Barnea, E., Beer, I. and Admon, A. (2006) The turnover kinetics of major histocompatibility complex peptides of human cancer cells. Mol Cell Proteomics, 5, 357-365.
58. Yewdell, J.W. (2011) DRiPs solidify: progress in understanding endogenous MHC class I antigen processing. Trends Immunol, 32, 548-558.
59. Pinkse, M.W., Mohammed, S., Gouw, J.W., van Breukelen, B., Vos, H.R. and Heck, A.J. (2008) Highly robust, automated, and sensitive online TiO2-based phosphoproteomics applied to study endogenous phosphorylation in Drosophila melanogaster. J Proteome Res, 7, 687-697.

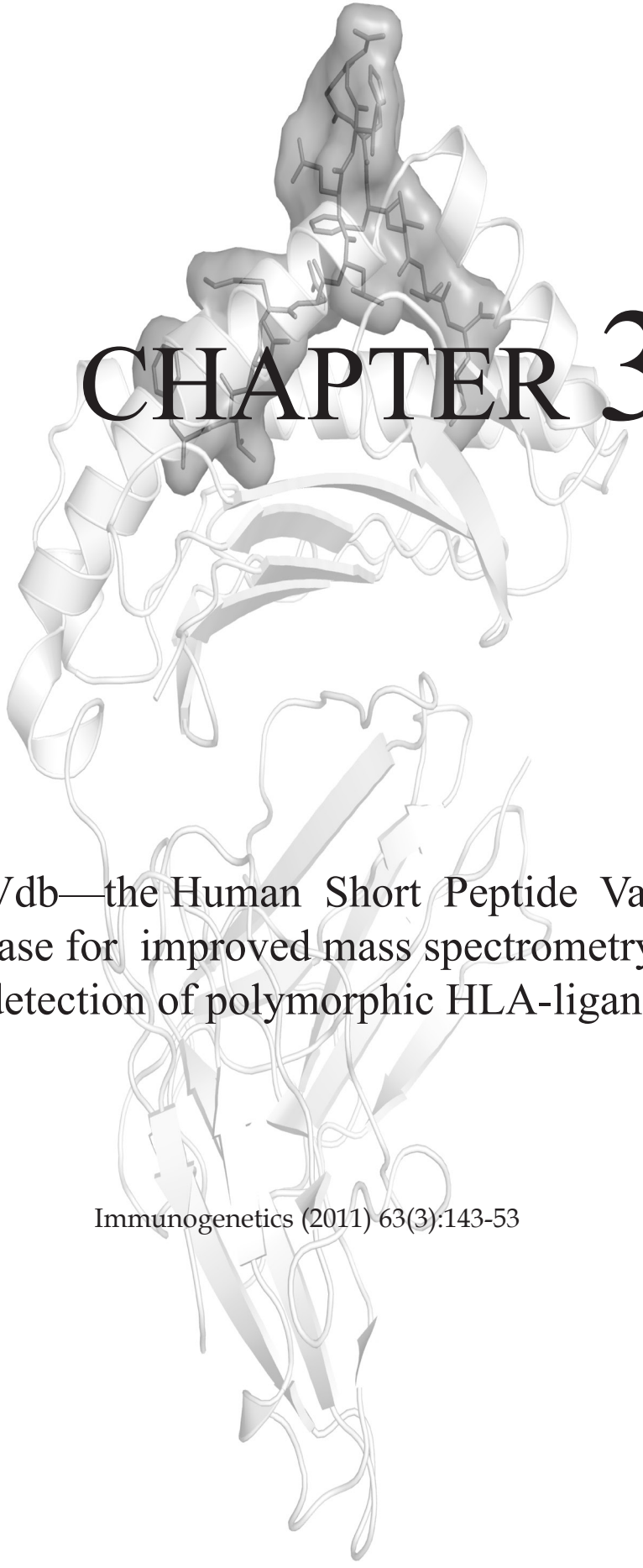
SUPPLEMENTAL DATA

Supplemental Table 1: Summary of the number of identified unique 8-11 mer peptides after searching the indicated database with the indicated FDR and/or best mascot ion score >35, and the number of overlapping peptides with the listing based on the use of a best mascot ion score of >35.

Databases	FDR	8-11 unique peptides	overlap
Swissprot homo	1%	2480	2314
	5%	7540	6483
	10%	12918	9770
IPI human 3.87	1%	0	0
	5%	3790	3453
	10%	7044	6053
mascot ion score ≥35		14065	

Supplemental Table 2: Virus-derived peptides identified, and confirmed by MS/MS of their synthetic counterpart, in our dataset. UTR=untranslated region. ARF=alternative reading frame.

Source	Gene name	Peptide sequence	BMI	Allele
EBV	EBNA3B/C (UTR)	ITAPLLPAV	54	HLA-A2
EBV	BGLF4	GLKDAVYFL	46	HLA-A2
EBV	HS4ENVGP (UTR)	SPLSPmARL	40	HLA-B7
EBV	RPMS1 (ARF)	APGGGERAL	45	HLA-B7
CMV	PP65	NLVPMVATV	28	HLA-A2
CMV	PP65	RPHERNGFTV	12	HLA-B7



CHAPTER 3

HSPVdb—the Human Short Peptide Variation Database for improved mass spectrometry-based detection of polymorphic HLA-ligands

Immunogenetics (2011) 63(3):143-53

Based on:

Harm Nijveen

Michel G. D. Kester

Chopie Hassan

Aurélie Viars

Arnoud H. de Ru

Machiel de Jager

Frederik Falkenburg

Jack A. M. Leunissen

Peter A. van Veelen

3

HSPVdb—the Human Short Peptide Variation Database for improved mass spectrometry-based detection of polymorphic HLA-ligands

ABSTRACT

T cell epitopes derived from polymorphic proteins or from proteins encoded by alternative reading frames (ARFs) play an important role in (tumor) immunology. Identification of these peptides is successfully performed with mass spectrometry. In a mass spectrometry-based approach, the recorded tandem mass spectra are matched against hypothetical spectra generated from known protein sequence databases. Commonly used protein databases contain a minimal level of redundancy, and thus, are not suitable data sources for searching polymorphic T cell epitopes, either in normal or ARFs. At the same time, however, these databases contain much non-polymorphic sequence information, thereby complicating the matching of recorded and theoretical spectra, and increasing the potential for finding false positives. Therefore, we created a database with peptides from ARFs and peptide variation arising from single nucleotide polymorphisms (SNPs). It is based on the human mRNA sequences from the well-annotated reference sequence (RefSeq) database and associated variation information derived from the Single Nucleotide Polymorphism Database (dbSNP). In this process, we removed all non-polymorphic information. Investigation of the frequency of SNPs in the dbSNP revealed that many SNPs are non-polymorphic "SNPs". Therefore, we removed those from our dedicated database, and this resulted in a comprehensive high quality database, which we coined the Human Short Peptide Variation Database (HSPVdb). The value of our HSPVdb is shown by identification of the majority of published polymorphic SNP and/or ARF-derived epitopes from a mass spectrometry-based proteomics workflow, and by a large variety of polymorphic peptides identified as potential T cell epitopes in the HLA-ligandome presented by the Epstein–Barr virus cells.

INTRODUCTION

T cell-mediated immunotherapy is an attractive treatment of cancer as it exploits the potential of cytolytic T cells to specifically recognize antigens that are selectively expressed on tumor cells [1-6]. The enormous specificity of T cells involved in killing tumor cells makes this kind of treatment very attractive. An excellent example is the powerful graft-versus-leukemia (GVL) effect witnessed after allogeneic hematopoietic stem cell transplantation. GVL is characterized by remission of a hematological malignancy coinciding with the in vivo expansion of tumor-specific T cells. These T cells react to a patient-specific epitope presented in human leukocyte antigen (HLA) molecules on tumor cells [7, 8]. T cell epitopes are peptides with a length of generally 8–11 amino acids. T cells are capable of distinguishing epitopes differing by only one amino acid, caused by a single nucleotide difference between patient and donor [9]. T cell epitopes, identified to play a role in (tumor) immunology, may arise from regular reading frames, but can also be encoded by alternative reading frames (ARFs) [10]. Given the need for therapeutically useful T cell epitopes, the identification of new epitopes is of unceasing importance. The identification of T cell epitopes has been achieved with an array of methods, among which mass spectrometry is one of the most prominent techniques [11-13]. Peptide identification by tandem mass spectrometry is most successfully applied in an ever increasing number of proteomics studies. In a typical high throughput proteomics/ligandomics setting [14], the experimentally determined tandem mass spectra are matched against a database of hypothetical spectra generated from known peptide sequences using search engines like Mascot [15] and Sequest [16].

For mass spectrometry-based identification of epitopes from polymorphic proteins, like minor histocompatibility antigens (MiHA) and peptides arising from ARFs, the commonly used protein databases like UniProt [17], IPI [18] and RefSeq [19] are unsuitable data sources, since these display very incomplete information about polymorphisms. Most of the published polymorphic MiHA are, therefore, not present in the standard protein databases, used in mass spectrometry-based workflows. Several strategies have been employed to address this problem (MSIPI [20], PepHum [21]), each with its own merits and limitations, trying to find the right balance between database size and completeness. In addition, there is a wealth of ligand and/or epitope information databases [22], but these are not applicable in mass spectrometry (MS)-based workflows. Knowing that customized search databases that provide detailed control over the search space can vastly outperform standard strategies [23], we designed a database dedicated to MiHA, thereby improving the chance of their identification in a proteomics type of experimental set up.

Our approach is based on the coding potential of the human genome, including its documented variations, as described in the RefSeq database. We chose RefSeq because it contains minimal redundancy, while still retaining splice variants, incorporates single nucleotide polymorphism (SNP) data from Single Nucleotide

Polymorphism Database (dbSNP) [24], which are richly annotated. We have created a database that contains all possible short peptides in different reading frames from a non-redundant mRNA set, combined with the known and annotated variations/SNPs. In this process, we removed all non-polymorphic information. Investigation of the frequency of SNPs in the dbSNP revealed that many of these SNPs are non-polymorphic "SNPs". Therefore, we removed those from our dedicated database as well, and this resulted in a high quality comprehensive polymorphic peptide database. Centered on the amino acid polymorphisms of non-synonymous SNPs, our dedicated Human Short Peptide Variation Database (HSPVdb) outperforms existing databases in MS/MS-based T cell epitope identification.

The value of our HSPV database is shown by identification of the majority of published polymorphic SNP and/or ARF-derived epitopes from a mass spectrometry-based proteomics workflow, as well as by a large variety of polymorphic peptides identified as potential T cell epitopes in the HLA-ligandome presented by EBV cells.

MATERIALS AND METHODS

Database preparation

The HSPVdb consists of peptides derived from genomic sequence variations. The database only contains peptides of seven amino acids or longer. The RefSeq database release 32 was downloaded from the NCBI FTP site and indexed using our local SRS installation [25], (http:// srs.bioinformatics.nl). The human mRNA subsection of RefSeq was extracted by selecting records with molecule type "mRNA" and organism source "Homo sapiens". The resulting list of RefSeq records was subsequently processed using a series of Perl scripts. To create the peptides derived from genomic sequence variations, we made use of the variation annotations that were added to RefSeq by the dbSNP staff. Variations found in the 5′ and 3′ UTRs were purposely included to allow detection of T cell epitopes derived from ARFs. For each annotated variation, the nucleotide sequences corresponding to the different alleles were generated. Instead of duplicating the complete mRNA sequence for each allele, we took a fragment starting from 30 nucleotides upstream and ending 32 nucleotides down- stream of the variation. The three forward reading frames of each allele were translated to amino acid sequences. This typically results in three peptide sequences of 20 amino acids. Translation ignored the presence or absence of start codons. Codons that could not be translated to a single amino acid due to ambiguous nucleotides were translated to a stop codon. The amino acid translation was split on stop codons to get peptides derived from a continuous reading frame. Only the peptides, including the variation were kept in the database. To minimize redundancy, a translation for an allele was only included when the variation gives rise to a change in amino acid sequence (non-synonymous SNPs). This part of the database is optimized for

finding peptides in the size range between 8 and 11 amino acids, but databases containing other peptide lengths can be produced at will. The database presented here consists of 20-mer peptides.

Each peptide sequence that was created, was stored as a separate database record and annotated with the ID of the originating mRNA sequence and the location of its encoding reading frame. If the RefSeq entry contains a coding sequence (CDS), the protein identifier and the position of that CDS on the mRNA with corresponding protein identifier, were added as annotation to the database record. For variations, we included the corresponding dbSNP identifiers, the positions of the variations, the nature of the amino acid changes and the percentage heterozygosity. If a variation causes an amino acid substitution, a SAP (single amino acid polymorphism), the possible amino acids were listed. Insertions or deletions were annotated as "in/del". The resulting database was stored as a flat file in FASTA format for mass spectrometry-based proteomics purposes. This HSPVdb is fully dedicated to finding polymorphic epitopes. To reduce the size of this database, all duplicate amino acid sequences were deleted. These peptides contain both polymorphisms for each position, thereby describing all possible SNP information.

Subsets of the HSPV database were created based on reported heterozygosity. Three heterozygosity categories were defined: 0/1, unknown, all others. Additionally, for all categories ARFs were either included or left out. Peptides for which the encoding DNA sequence is not part of the RefSeq-annotated open reading frame are labeled as alternative reading frame or ARF peptides. These include CDS that are in a different reading frame and sequences that are located up- or downstream of the annotated open reading frame.

SNP genotyping assays

Genomic DNA was isolated from 192 HLA A*0201- positive patient and donor samples (peripheral blood mononuclear or bone marrow cells) by the Gentra Systems PUREGENE genomic isolation kit (Biocompare, San Francisco, CA). SNPs rs4848158, rs61378134, rs3602 31 50, rs1 1 540 52 6, rs1 1 5 5 4 279 , rs 359 58 18 9, rs56013141, rs11541290, rs34422048, rs11541416, rs28 6599 89, rs2 0701 59, rs4 2610 80, rs11 5 571 42, rs11555631, rs11479605, rs11541519, rs5030742, rs11548263 were analyzed using a KASPar assay with allele-specific primers labeled with VIC and FAM dyes, (KBioScience, Hoddesdon, UK). Genotyping was per- formed according to manufacturer's instructions.

Ilumina custom array was used for genotyping rs10960, rs1143 138 , r s1 298 600 2, rs346 691 46, rs10 478 44, rs11266765, rs11539866, rs11541416, rs11541519, rs11542419, rs11542836, rs11544489, rs11545551, rs11548082, rs11553285, rs11553982, rs11554156, rs11554279, rs11555631, rs11557142, rs11558570, rs13202878, rs17848351, rs17851857, rs17853301, rs17853718, rs1803 181, rs2070159, rs2261324, rs2 893 488 7, r s289 351 71, rs28 940 302 , rs3 180 96 1, rs3 413 69 99 , r s34 4 1 8 7 12, rs39

62 697 , r s4 848 15 8, rs5030742, rs6112008, rs6686209, rs6794514. Genotyping was performed according to manufacturer's instructions.

Sample preparation for test set

Peptide synthesis

Peptides were synthesized by standard Fmoc chemistry on a Syro II peptide synthesizer as described previously [26]. The integrity of the peptides was checked by reversed-phase high-performance liquid chromatography (HPLC) and mass spectrometry. Liquid chromatography–mass spectrometry

The peptides studied are listed in Table 1. These are minor histocompatibility antigens as identified by different research groups around the world. A more complete listing of MiHA can be found at http://www.lumc.nl/dbminor. To perfectly mimic the conditions used in a normal mass spectrometry-based HLA-ligand identification process, all peptides included in Table 1 were measured by on-line chromatography/mass spectrometry (see below), and tandem mass spectra were recorded of their singly, doubly, and triply charged form. Subsequently, a selection of relevant charge states was made for each peptide, and charge states with a substantial contribution to the overall intensity only were used to construct a Mascot generic file (MGF) containing 31 tandem mass spectra, see Table 2.

Sample preparation for determination of the EBV-LCL ligandome

Cell collection, preparation, and HLA elutions

Peripheral blood samples were obtained from healthy donors after approval by the Leiden University Medical Center Institutional Review Board and informed consent according to the Declaration of Helsinki. Mononuclear cells (MNC) were isolated by Ficoll-Isopaque separation and cryopreserved. Stable Epstein–Barr virus (EBV)-transformed B cell lines (EBV-LCL) were generated using standard procedures. EBV-LCL and HeLa cells were cultured in Iscove's Modified Dulbecco's Medium (IMDM, BioWhittaker, Verviers, Belgium) supplemented with 10% bovine fetal serum (FBS, BioWhittaker).

Peptide isolation

Peptide isolation was performed with protein A beads (GE healthcare) covalently linked to the major histocompatibility complex (MHC) class I mAb W6/32 (3 mg W6/32 on 1 ml of ProtA sepharose) using dimethyl pimelimidate according to the standard protocol [27].The complex MHC-peptide pool was prefractionated on a C18 RP-HPLC system (2 mm × 15 cm; Reprosil-C18-AQ 3 um; Dr. Maisch GmbH, Ammerbuch, Germany), using a gradient 0–60% A to B. A: water, 5% Acetonitrile (ACN),0.1% TFA, B: ACN, 0.1% TFA.

Table 1 Overview of known MiHA used as a test set in this study. It displays the epitope name and the HLA-molecule it is presented in. In addition, its immunogenicity is indicated together with the gene name and the polymorphisms are indicated. *Names according to http://www.lumc.nl/dbminor

Epitope name*/ HLA	Sequence	Charge	remarks	Gene	polymorphic AA	dnSNP entry
HA1 / A2	VLHDDLLEA	1, 2	immunogenic	HMHA1	VL[R/H]DDLLEA	rs1801284
HA2 / A2	YIGEVLVSV	1	immunogenic	MYO1G	YIGEVLVS[V/M]	rs61739531
HA3 / A1	VTEPGTAQY	1, 2	immunogenic	AKAP13	V[M/T]EPGTAQY	rs2061821
HA8 / A2	RTLDKVLEV	1, 2, 3	immunogenic	KIAA0020	[R/P]TLDKVLE[V/I]	rs2173904; rs2270891
HA1 / B60	KECVLHDDL	1, 2	immunogenic	HMHA1	KECVL[R/H]DDL	rs1801284
LB-ADIR-1F / A2	SVAPALALFPA	1, 2	immunogenic; ARF in 5'UTR	TOR3A (ADIR)	SVAPALAL[F/S]PA	rs2296377
LB-ADIR-1S / A2	SVAPALALSPA	1, 2	allelic counterpart			
CTSHr / A31	ATLPLLCAR	1, 2	immunogenic	CTSH	ATLPLLCA[G/R]	rs2289702
CTSHr / A33	WATLPLLCAR	2	immunogenic	CTSH	WATLPLLCA[G/R]	rs2289702
ACC1y / A24	DYLQYVLQI	1, 2	immunogenic	BCL2A1	DYLQ[C/Y]VLQI	rs1138357
ACC1c / A24	DYLQCVLQI	1, 2	immunogenic			
ACC1c +cystinylated	DYLQCVLQI	1, 2	immunogenic			
HB1h / B44	EEKRGSLHVW	2, 3	immunogenic	HMHB1	EEKRGSL[H/Y]VW	rs161557
ACC2d / B44	KEFEDDIINW	1, 2	immunogenic	BCL2A1	KEFED[G/D]IINW	rs3826007
ACC2g / B44	KEFEDGIINW	1, 2	allelic counterpart			
LB-ECGF1-1H/ B7#	RPHAIRRPLAL	2, 3	immunogenic; ARF	TYMP (ECGF1)	RP[H/R]AI[R/C]RPLAL	no entry; rs1061205

Liquid chromatography–mass spectrometry

Peptide fractions were reduced to near dryness and resuspended in 95/3/0.1 v/v/v water/acetonitrile/formic acid. These resuspended fractions were analyzed by on-line nano-HPLC mass spectrometry with a system described by Meiring et al [28]. Fractions were injected onto a precolumn (100 um × 15 mm; Reprosil-Pur C18-AQ 3 um, 5 um, Phenomenex) and eluted via an analytical nano-HPLC column (15 cm × 50 um; Reprosil- Pur C18-AQ 3 um). The gradient was run from 0% to 50% Solvent B (10/90/0.1 v/v/v water/ acetonitrile/formic acid) in 90 min. The nano-HPLC column was drawn to a tip of approximately 5 um and

Table 2. Summary of the searches with the test set of known MiHA against the IPI, MSIPI, PepHum, and HSPV database. The peptide names and sequences are given together with the charge of the precursor, submitted to tandem mass spectrometry. For each database, three columns are displayed: (1) whether the peptide is present in the database (Pr?), followed by (2) the mascot ion score assigned to the tandem mass spectrum (black filling if the mascot ion score is above the threshold of the search), and (3) the evaluation, i.e., was the tandem mass spectrum matched to the correct peptide (black filling and (Y) if correct, and above the mascot threshold (cut-off score), gray filling if correct and below (ye) the mascot threshold. In short, the blacker the better. Wr wrong interpretation of MS2 spectrum; np no matching/no proposal from mascot search. The HSPVdb scores very well, due to its reduced format in combination with a high density of relevant SNP information. 4+ was the most abundant in the charge distribution of peptide LB- ECGF-1H, but its MS2 spectrum was of such poor quality that it was not included for database searching. LB-ADIR peptides are from an ARF. ACC1+ Cys represents a special case in which the cysteine residue in the epitope can be modified by formation of an S–S bridge with free cysteines. This is relevant for both in vivo recognition and mass spectrometric interpretation.

Database				**IPI369**			**MSIPI367**			**PepHum**			**HSPV20**	
Mass accuracy (ppm)				1			1			1			1	
Mascot cut-off score				**37**	**37**		**38**	**38**		**44**	**44**		**28**	**28**
Peptide name	**Sequence**	Charge	**Pr?**	**sco**	**int**	**Pr?**	**sco**	**int**	**Pr?**	**sco**	**int**	**Pr?**	**sco**	**int**
CTSHr A31	ATLPLLCAR	2		10	wr	Y	42	Y	Y	42	ye	Y	42	Y
CTSHr A33	ATLPLLCAR	1			np	Y		np	Y	8	ye	Y		np
HA3t A1	VTEPGTAQY	2		12	wr	Y	34	ye	Y	34	ye	Y	34	Y
HA3t A1	VTEPGTAQY	1		9	wr	Y	18	ye	Y	18	wr	Y	18	ye
HA2v A2	YIGEVLVSV	1	Y	18	wr	Y	18	wr	Y	28	wr	Y	16	ye
LB-ADIR-1S A2	SVAPALALSP/	2		22	wr		22	wr	Y	48	Y	Y	48	Y
LB-ADIR-1S A2	SVAPALALSP/	1		13	wr		13	wr	Y	34	wr		8	wr
HA1h A2	VLHDDLLEA	2		28	wr	Y	28	wr	Y	28	wr	Y	15	ye
HA1h A2	VLHDDLLEA	1		26	wr	Y	40	Y	Y	40	ye	Y	40	Y
LB-ADIR-1F A2	SVAPALALFP/	2		17	wr		17	wr		28	wr	Y	66	Y
LB-ADIR-1F A2	SVAPALALFP/	1		25	wr		25	wr		29	wr		4	wr
HA1h B60	KECVLHDDL	2		14	wr	Y	36	ye	Y	36	ye	Y	36	Y
HA1h B60	KECVLHDDL	1		5	wr	Y	29	ye	Y	29	ye	Y	29	Y
HA8rv A2	RTLDKVLEV	3	Y	37	Y	Y	37	ye	Y	37	ye	Y	37	Y
HA8rv A2	RTLDKVLEV	2	Y	34	wr	Y	34	wr	Y	34	wr	Y	30	Y
HA8rv A2	RTLDKVLEV	1	Y	32	ye	Y	32	ye	Y	32	ye	Y	32	Y
ACC1c	DYLQCVLQI	2	Y	50	Y	Y	50	Y	Y	50	Y	Y	50	Y
ACC1c	DYLQCVLQI	1	Y	36	wr	Y	36	wr	Y	40	wr	Y	15	ye
CTSHr A33	WATLPLLCAR	2		8	wr	Y	37	ye	Y	37	ye	Y	37	Y
ACC1y BCL2A1-A24	DYLQYVLQI	2		27	wr	Y	58	Y	Y	58	Y	Y	58	Y
ACC1y BCL2A1-A24	DYLQYVLQI	1		22	wr	Y	25	ye	Y	25	ye	Y	25	ye
ACC1c +cys	DYLQCVLQI	2	Y	42	Y	Y	42	Y	Y	42	ye	Y	42	Y
ACC1c +cys	DYLQCVLQI	1	Y	36	ye	Y	36	ye	Y	36	ye	Y	36	Y
HB1h B44	EEKRGSLHVW	3	Y	10	wr	Y	10	wr	Y	13	wr	Y	6	ye
HB1h B44	EEKRGSLHVW	2	Y	16	ye	Y	16	ye	Y	16	wr	Y	16	ye
ACC2g BCL2A1-B44	KEFEDGIINW	2	Y	48	Y	Y	48	Y	Y	48	Y	Y	48	Y
ACC2g BCL2A1-B44	KEFEDGIINW	1	Y	34	ye	Y	34	ye	Y	34	ye	Y	34	Y
LB-ECGF-1H B7	RPHAIRRPLAI	3		5	wr		5	wr		16	wr		3	wr
LB-ECGF-1H B7	RPHAIRRPLAI	2			np			np		4	wr			np
ACC2d BCL2A1-B44	KEFEDDIINW	2		17	wr	Y	39	Y	Y	39	ye	Y	39	Y
ACC2d BCL2A1-B44	KEFEDDIINW	1		15	wr	Y	45	Y	Y	45	Y	Y	45	Y

acted as the electrospray needle of the MS source.

The mass spectrometer was an LTQ-FT Ultra (Thermo, Bremen, Germany) and was operated in data-dependent mode, automatically switching between MS and MS/MS acquisition. Full scan mass spectra were acquired in the FT-ICR with a resolution of 25,000 at a target value of 5,000,000. The two most intense ions were then isolated for accurate mass measurements by a selected ion monitoring scan in FT-ICR with a resolution of 50,000 at a target accumulation value of 50,000. The selected ions were then fragmented in the linear ion trap using collision-induced dissociation at a target value of 10,000. In a post analysis process, raw data were converted to peak lists using Bioworks Browser software, Version 3.1. For peptide identification, MS/MS data were submitted to the human IPI database using Mascot Version 2.2.04 (Matrix Science) with the following settings: 2 ppm and 0.8-Da deviation for precursor and fragment masses, respectively; no enzyme was specified. The Mascot output files were loaded into Scaffold (http://www.proteomesoftware.com) and exported to Excel as peptide reports and duplicates were removed.

RESULTS

To investigate the value of our database, we studied two sets of samples. First, a test set comprising approximately 30% of all MiHA known today, as listed in Table 1, and second, a set of peptides eluted from HLA from an EBV- cell line.

Validation of HSPVdb with a test set of known MiHA

Our test set of known polymorphic peptides and allelic counterparts were synthesized and measured in standard on-line nanoHPLC/MS experiments, as in our normal proteomics workflow on HLA-ligands [29]. Of all significantly occurring charge states, tandem spectra were recorded. Tandem mass spectra of varying quality are present in this data set, reflecting a “real-world” situation, where the spectral quality depends on intrinsic peptide properties. A combined peak list was constructed from these spectra for searching the databases used in this work. This led to a set of 31 experimental tandem MS derived from 15 peptides (Table 2).

For validation of our HSPVdb, we compared it to the MSIPI and PepHum databases that were specifically constructed to address the lack of peptide variation in common databases like IPI. A summary of the databases used in this study is shown in Table 3.

The HSPVdb is similar to the size of the IPI and MSIPI databases, but it includes all SNP information in all forward and ARFs (MSIPI: 170.242 SNPs; HSPVdb: 380.182

SNPs). When leaving out the alternative reading frame information (i.e., HSPVdb subset 1, see Table

3), the size of our HSPVdb is reduced to only 25% of the size of IPI and MSIPI, which is of great importance when searching databases.

The test set containing the tandem mass spectra of known MiHA was searched against the IPI, MSIPI, PepHum, and our HSPVdb. Searches were performed using the Mascot search engine (Matrix science), with various settings for mass accuracy (1, 2, 5, 10, and 50 ppm) representing the mass accuracy of various MS and/or experimental setups. The enzyme setting was "none". It is important to note that in the elucidation of HLA-ligands, the peptide termini are unknown in contrast to the vast majority of cases in standard proteomics experiments, in which peptide matching against databases can be done with an additional and very stringent condition, namely an enzyme cleavage site (in most cases, trypsin). In the standard proteomics approach, the enzyme restriction has an enormous positive impact on specificity and search time. For the sequencing of T cell epitopes, enzyme restriction is not applicable. However, for binding to the presenting HLA molecule, HLA-ligands have to satisfy certain conditions imposed by the HLA molecule, the binding motif. This binding motif can be used as additional help to some extent to assess the value of the matched sequence by the search engine. In addition, netMHC, http://www.cbs.dtu.dk/ services/NetMHC/, could be applied to some extent, but neither of the two can be directly applied in the database search as a fixed condition. The best proof of a correct peptide assignment, in spite of improvements in peptide matching algorithms, is still the comparison of the tandem spectrum of the proposed eluted epitope with its synthetic counterpart.

All output of the Mascot search engine was assessed manually, and a summary of the results for a 1-ppm mass accuracy is shown in Table 2, and a full report of the searches is given in Supplemental Table 1.

Table 2 shows a selection of the searches in the four databases with a 1-ppm mass measurement accuracy. For every individual tandem mass spectrum, the Mascot ion score is reported. The results from the database search were classified by the following criteria: (1) was the tandem mass spectrum correctly identified by the search engine 30 (indicated by black and gray filling in the first column) for each database? and (2) was the identification score above (indicated by black filling in the second column for each database) or below the Mascot significance threshold (cut-off score)? Therefore, "the blacker the better". The presence ("Pr") of each peptide in the particular database is indicated by "Y" in the appropriate column. Supplemental Table 1 shows the results of all searches performed with the test set of 31 tandem mass spectra to the IPI 3.69, MSIPI 3.67, PepHum, and HSPVdb.

From Table 2, it is immediately clear that the IPI database is not useful for finding MiHA, since it lacks essential variation information. The PepHum database, based on expressed sequence tags (ESTs) information, including ARFs, is relatively large, by which relevant information for finding our polymorphic epitopes is "diluted", and consequently, a serious amount of "noise" is generated, increasing the chance of

finding false positives. The consequence of this is reflected in the outcome of the database search for PepHum. The number of significantly scoring peptides is only 5 as compared to the 19 peptides identified by our HSPVdb, see also Figure 1a. This low score is only partially rescued by the number of correctly assigned peptides with a score below the Mascot significance threshold. In addition, ESTs may be more prone to experimental sequencing errors, leading to occurrence of false SNPs.

The elegantly produced MSIPI does quite well, but also here, most correct peptide hits are below the statistical significance threshold score, which makes it hard to decide if a hit is true or a false positive in a "non-test set" setting. In addition, the MSIPI does not contain information from ARFs and UTRs.

For the HSPVdb, out of 31 MS/MS spectra, 19 are identified correctly above the Mascot significance threshold, while another 7 are also correctly identified, but below the significance threshold. Only three tandem mass spectra were wrongly assigned (false positives).

These wrong assignments are caused by the poor quality of the tandem mass spectra of these peptides, due to intrinsic peptide properties. To two tandem mass spectra, no match was assigned. These tandem mass spectra represent two peptides, "YIGEVLVSV", which yields a bad mass spectrum and "RPHAIRRPLAL", which is not present in the HSPVdb subset, because it is derived from a SNP not found in the dbSNP database. The HSPVdb, designed to reduce non-informative sequence information, outperforms the other databases.

Next to the size of the database, relieving the accuracy condition from 1 to 50 ppm (Figure 1b) has a detrimental effect on both the number of correctly assigned peptides above and below the Mascot significance threshold. This effect can even lead to a false-positive score, as illustrated by a high and significant Mascot score of 63 (!) for MS/MS/query #6 (in HSPVdb, 50 ppm), see supplemental Table 1a. This result emphasizes the value of high mass accuracy.

So far, the good performance in the MS/MS-based identification of T cell epitopes of HSPVdb can be attributed to the compact nature and the special focus on polymorphic peptides. A reduced database size directly translates to a lower noise level in the database search, which is especially important in high-throughput T cell epitope elucidation, where search space limiting constraints like an enzyme cleavage site cannot be used. Another parameter affecting search quality is mass accuracy, which is also proven to be a prominent factor in avoiding false positives.

To further improve the quality of our HSPVdb, we focused on the quality of the SNPs in dbSNP, since we noted that the reported frequency of a substantial number of SNPs in dbSNP is "0" or "1" or "unknown". This made us decide to study a random set of 52 SNPs with no frequency reported in dbSNP. We developed a SNP assay to screen a random HLA-A*02-positive Dutch donor population using the KASPar assay (92 DNA

samples) and a SNP array (192 DNA samples). In our test population, 46 out of the 52 SNPs (90%) were not polymorphic, having an allele frequency of 1 or 0 in the SNP assays. Two SNPs (4%) were very rare (allele frequencies of 0.97, and 0.99), and 4 SNPs (8%) had a reasonable distribution in our population (0.77; 0.70; 0.20; 0.13).

A large number of reported "SNPs" in dbSNP is apparently not polymorphic, thereby contaminating our proteomics approach and the chance of finding suitable patient/donor MiHA pairs. Therefore, since reduction of the search space greatly enhances the chance of finding true positives in database searches, we decided to test our HSPVdb after removal of either "unknowns" or "0" and "1", or both. The results are shown in supplemental Table 1b. Subset 3, the leanest form of HSPVdb with both "0" and "1" and "unknown frequency" SNPs removed and without ARFs, is reduced to only one fourth of its original size. Therefore, the significance threshold is clearly lowered (from 28 to 22 for 1-ppm mass accuracy), increasing the chance of finding true positives. In particular, those derived from tandom mass spectra of relatively poor quality with accompanying intrinsic low Mascot scores. Only one true positive is lost, because its frequency is not reported in the dbSNP. Similarly, the other subsets (subsets 1–8) of HSPVdb have reduced significance thresholds (data not shown). The application of these various forms of the database can be adapted to the needs of the user.

So far, we have shown that the selective reduction of the database size by removal of both the non-polymorphic peptide stretches and the SNPs of limited value, leads to a comprehensive high quality database file dedicated to improving the elucidation of MiHA.

Database quality and inconsistencies

During this work, we discovered inconsistencies in the number of SNPs included in several RefSeq and MSIPI versions, see Figure 2a and b. The number of reported human SNPs dropped by 50% going from RefSeq release 28 to release 30, and by more than 50% in MSIPI going from version 37 to version 38. We reported this in October 2008 to the respective database producers who acknowledged there were problems and improved their efforts. Recently, we encountered a problem with the SNPs reported by 1000 genomes. org in dbSNP which is being solved. Therefore, we continued using version 3.32 (on our website the HSPVdb version based on either Refseq release 32 or release 40 can be chosen). We would like to warn users for the status of the RefSeq with respect to this. MSIPI, also being a secondary database, suffered from the same errors during several versions, but this has been repaired, starting from version 49, although a strong decrease can be seen in version 3.67 (Figure 2b). In general, as a user of these databases, it is very hard to judge the value of the database, so caution should be taken: newer versions are not always better.

Application of HSPVdb to finding potential MiHA presented in HLA on EBV-cells

To investigate the effects of application of our database to a representative HLA-ligand elution experiment, we eluted peptides from an EBV-LCL cell line (EBV-JY). After lysis, affinity purification was performed with

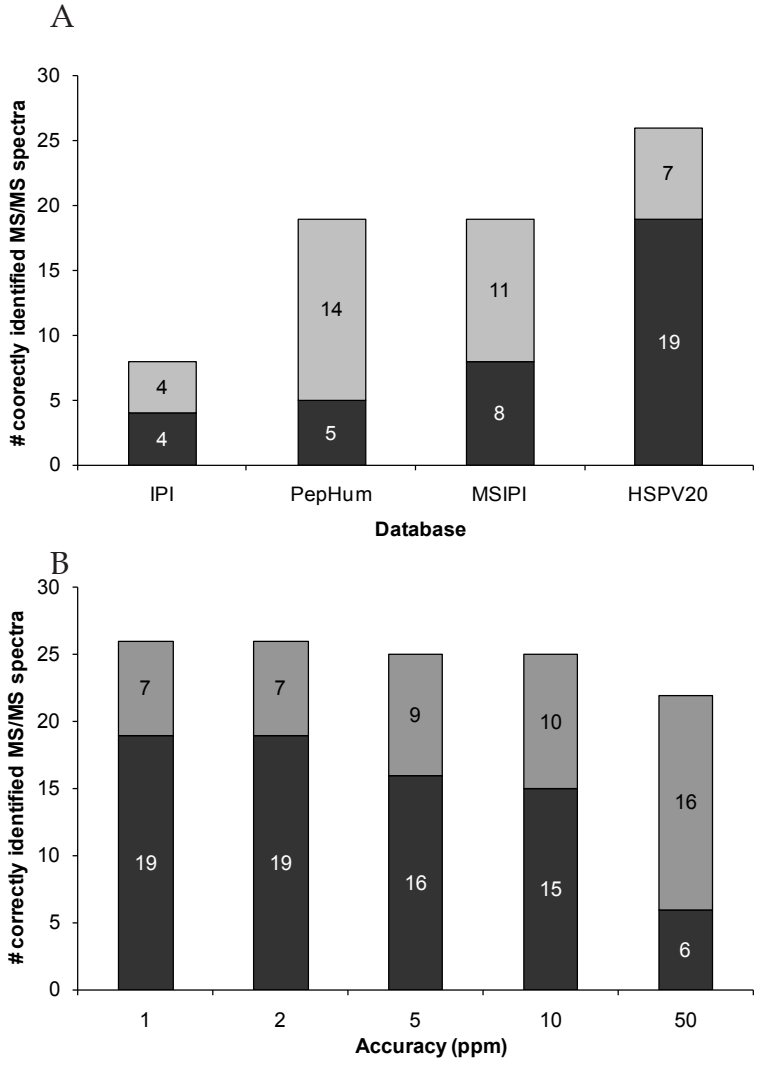


Figure 1. (A) Summary of the searches with 1-ppm accuracy against the IPI, MSIPI, PepHum, and HSPV databases. The color coding is as follows: black correct hit and above the MASCOT significance threshold; gray correct hit, but below the significance threshold. (B) Summary of the searches against HSPVdb with various mass measurement accuracies. The color coding is as above

BB7.2 antibody for HLA-A2, followed by separation of HLA and peptides. Subsequently, the complex peptide pool was analyzed by on-line nanoHPLC-tandem MS. The tandem mass spectra were matched against several databases for comparison, in particular, MSIPI and various subsets of our HSPV database. Here, MSIPI is compared to the smallest subset of our HSPV database without ARFs (subset 3) and with ARFs included (subset 7), the advantages of which have been illustrated for the test set described above. These trimmed subsets do not include SNPs of which the frequencies in dbSNP are reported to be 0/1 or unknown. By searching against the smaller compact database containing all relevant SNPs, intermediate scoring peptides appear in the database search that would otherwise fall below the significance threshold when matching tandem mass spectra against larger databases.

This is illustrated by the number of intermediate scoring peptides, i.e., those peptides that score below the Mascot significance threshold when matching against MSIPI, and are, therefore, peptides not found otherwise. An additional 130 peptides were found for subset 3 and an additional 400 for subset 7.

Table 3. Overview of the databases used in this study, listing the number of entries and the number of amino acid residues present in each database. In addition, the presence of ARFs and the (type of) SNP information in the various databases is indicated. The number of residues of each database relative to the IPI database and the relative size of the HSPV subsets is given. The number of SNPs in MSIPI 3.67 is 170.242; the number of SNPs in HSPVdb (subsets 1 and 5) is 380.182

Database	Nr. of sequences	Nr. of residues	Size relative to IPI 3.69	ARFs?	0/1?	Unk?
IPI (HUMAN v3.69)	87130	35200044	1			
MSIPI (HUMAN v3.67)	87040	42553286	1.209		√	√
PepSeq	75237	176019757	5.001	√	√	√
HSPV20	1575171	45422884	1.290	√	√	√
			Rel. to set 5			
HSPV subset 1	423015	8344552	0.184		√	√
HSPV subset 2	377269	7440614	0.164			√
HSPV subset 3	106379	2108989	0.046			
HSPV subset 4	152125	3012927	0.066		√	
HSPV subset 5	2634086	45422884	1	√	√	√
HSPV subset 6	2378073	41106669	0.905	√		√
HSPV subset 7	729721	12444311	0.274	√		
HSPV subset 8	985734	16760526	0.369	√	√	

These extra peptides need to be checked for false positives (peptides with tandem mass spectra that match better with non- SNP containing peptides), and for the presence of a SNP. The extra peptides found can, e.g., be evaluated by application of netMHC. This approach, starting from our small experimental elution experiment, yielded eight peptides from subset 7 (including ARFs), and five peptides from subset 3 with a netMHC score below 50 (i.e., a stringent condition for strong binding). These peptides, shown in Table 4, are currently evaluated as potential MiHA.

All peptides found only by searching against the dedicated HSPV database increase the chance of finding relevant MiHA. The excellent annotation of the SNPs reported in our HSPV database enables the user to directly jump to the relevant information about the polymorphism, a feature that was largely lacking so far. The HSPV database described here is an integral part of a complete peptidomics pipeline for finding therapeutically useful MiHA, a strategy that is currently under development.

Availability and web interface

A flat file with the content of the HSPV database can be requested by sending an e-mail to hspv@ bioinformatics.nl. A simple interactive query interface is available at: http:// srs.bioinformatics.nl/hspv/.

This web interface allows the biologist to query the database for peptide sequences. It returns a list of RefSeq mRNA entries that contain a continuous reading frame encoding the query peptide, the start position of that reading frame, the position of the encoding nucleotide Sequence with respect to any

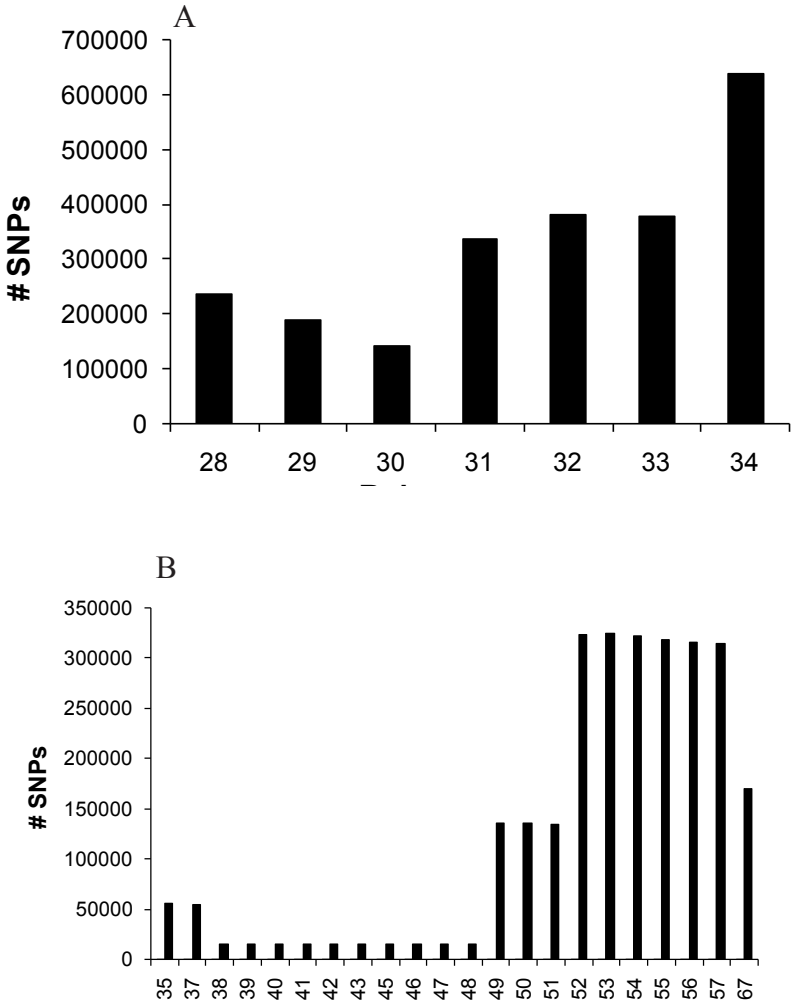


Figure 2. Number of incorporated SNPs per release of RefSeq (A) and of MSIPI (B)

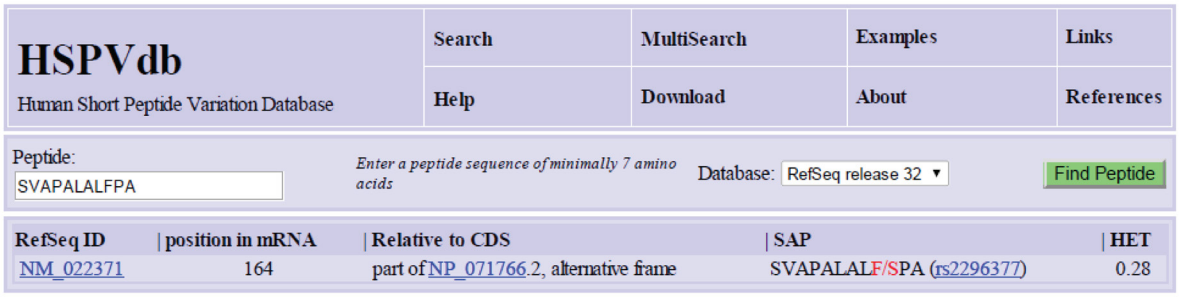
annotated CDS, and a description of the variations if the peptide contains any, see Figure 3a. This is a great feature for the initial assessment of the quality and potential usefulness of the output of our database searches. The richness of SNP information of our database is shown in Figure 3b, for the peptide "TLSELHCD" displaying SAPs at every position in the peptide.

Improvements in the quality and quantity of dbSNP entries, among others by the 1000 genomes project (http:// www.1000genomes.org), if well controlled, will greatly enhance the use of our database by reporting useful frequencies and removal of spurious frequencies in the current dbSNP releases.

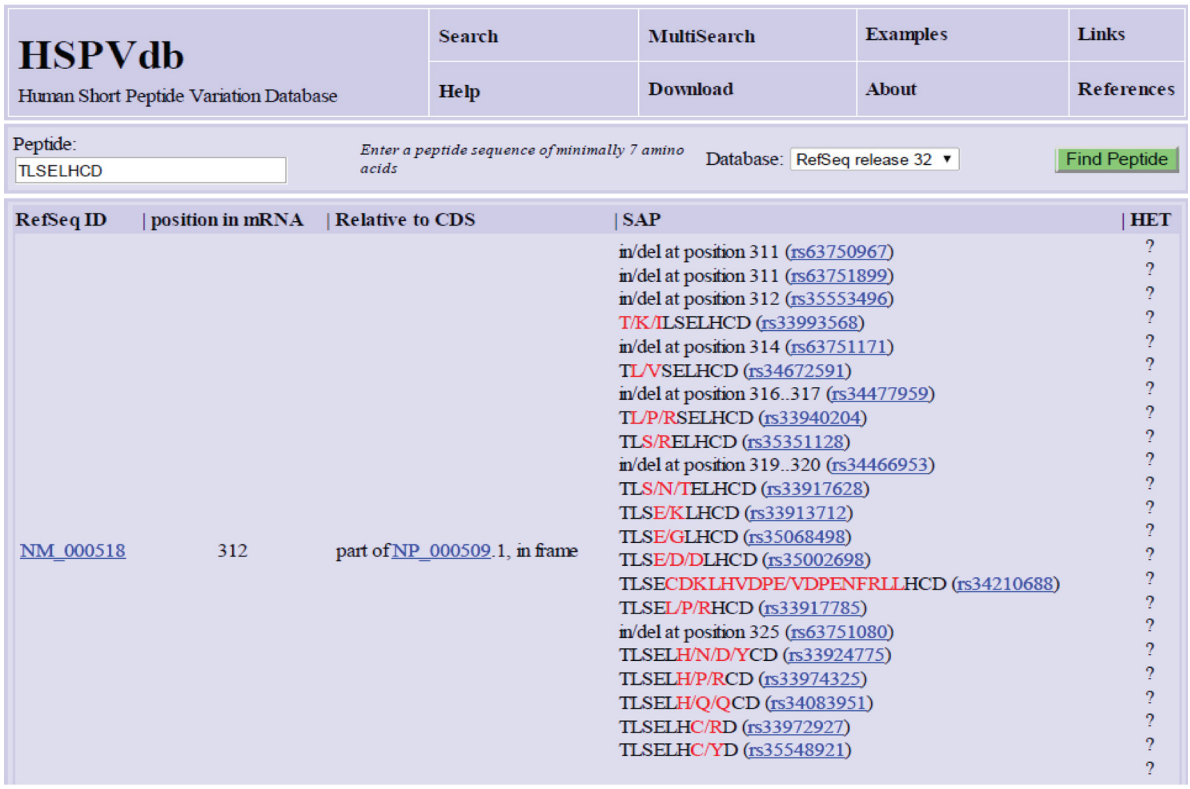
The website (http://srs.bioinformatics.nl/hspv/) provides easy access to relevant information about SNPs by its good annotation and hyperlinks incorporated in the HSPVdb.

Figure 3. Screen shots show the output of a query for the peptides (A) SVAPALALFPA and (B) TLSELHCD.It clearly illustrates the effect of the large number of annotated variations at the amino acid level.

A



B



CONCLUSION

We have shown that selective reduction of the database size by removal of both the non-polymorphic peptide stretches and the non-polymorphic “SNPs” leads to a comprehensive high quality database file dedicated to improving the elucidation of MiHA.

Table 4. Exclusive peptides with selected info from the HSPVdb. Peptides are either in frame (y) or in an ARF (n). The position of a SNP is indicated in the column SNP. In addition, the heterozygosity and NetMHC score is given

Peptide	mRNA	Gene	Protein	rel2cds	in frame	dbSNP	SNP	Het	NetMHC
FLIPKTLVGV	NM_017700	FLJ20184	NP_060170.1	down stream	y	rs2121558	FLIPKTLVG[E/V]	0,47	9
SLSDLIYAL	NM_001080837	SEBOX	NP_001074306.2	inside	y	rs9910163	SLSDLIYA[L/S]	0,13	7
GLWEQENHL	NM_024713	C15orf29	NP_078989.1	inside	y	rs34998154	GLW[E/K]QENHL	0,05	41
FIVTVIHTI	NM_024607	PPP1R3B	NP_078883.2	down stream	n	rs330915	FIVTVIHT[I/F]	0,49	30
FLSEHPNVTL	NM_145298	APOBEC3F	NP_660341.2	inside	y	rs17000697	FL[A/S]EHPNVTL	0,28	19
FLNQRSIML	NM_030956	TLR10	NP_112218.2	up stream	n	rs9998678	FLNQ[R/W]SIML	0,05	29
LLQSLVSI	NM_198889	ANKRD17	NP_942592.1	inside	n	rs6855349	LLQS[S/L]VSI	0,46	46
TLLDPNEKYLL	NM_016243	CYB5R1	NP_057327.2	inside	y	rs2232842	TLLDP[N/S]EKYLL	0,31	31

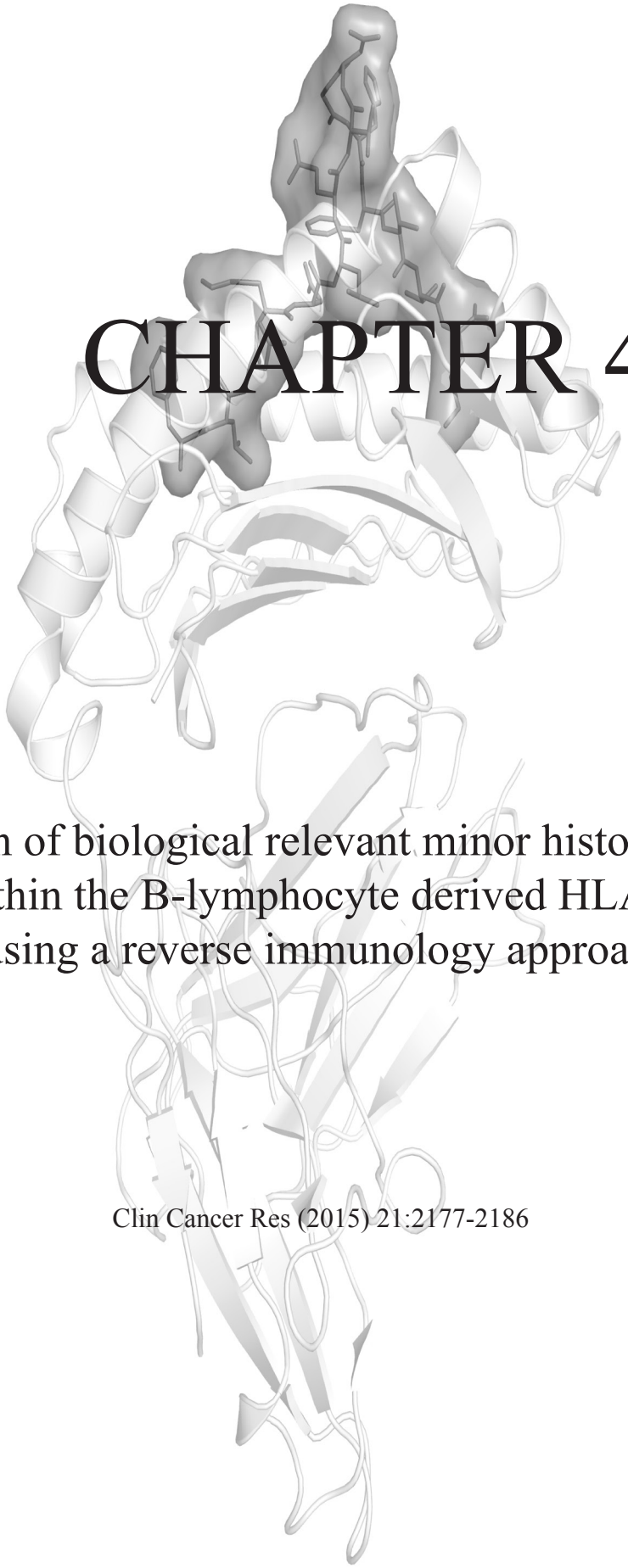
ACKNOWLEDGEMENT

The authors would like to thank David Kloet for the initial work on the project. Peter de Koning and Antoinette Teixeira are thanked for peptide synthesis. H.N. was supported by the BioAssist program of the Netherlands Bioinformatics Centre. This research was made possible by the financial assistance of the Landsteiner Foundation for Blood Transfusion Research. The authors declare that they have no conflict of interest.

REFERENCES

1. Storb, R., Allogeneic hematopoietic stem cell transplantation--yesterday, today, and tomorrow. Exp Hematol, 2003. 31(1): p. 1-10.
2. Hambach, L. and E. Goulmy, Immunotherapy of cancer through targeting of minor histocompatibility antigens. Curr Opin Immunol, 2005. 17(2): p. 202-10.
3. Kessler, J.H. and C.J. Melief, Identification of T-cell epitopes for cancer immunotherapy. Leukemia, 2007. 21(9): p. 1859-74.
4. Falkenburg, J.H., et al., Minor histocompatibility antigens in human stem cell transplantation. Exp Hematol, 2003. 31(9): p. 743-51.
5. Bleakley, M. and S.R. Riddell, Molecules and mechanisms of the graft-versus- leukaemia effect. Nat Rev Cancer, 2004. 4(5): p. 371-80.
6. Eisenlohr, L.C., L. Huang, and T.N. Golovina, Rethinking peptide supply to MHC class I molecules. Nat Rev Immunol, 2007. 7(5): p. 403-10.
7. Marijt, W.A., et al., Hematopoiesis-restricted minor histocompatibility antigens HA- 1- or HA-2-specific T cells can induce complete remissions of relapsed leukemia. Proc Natl Acad Sci U S A, 2003. 100(5): p. 2742-7.
8. van Bergen, C.A., et al., Multiple myeloma-reactive T cells recognize an activation- induced minor histocompatibility antigen encoded by the ATP-dependent interferon-responsive (ADIR) gene. Blood, 2007. 109(9): p. 4089-96.
9. Spierings, E., et al., Phenotype frequencies of autosomal minor histocompatibility antigens display significant differences among populations. PLoS Genet, 2007. 3(6): p. e103.
10. Ho, O. and W.R. Green, Alternative translational products and cryptic T cell epitopes: expecting the unexpected. J Immunol, 2006. 177(12): p. 8283-9.
11. Engelhard, V.H., The contributions of mass spectrometry to understanding of immune recognition by T lymphocytes. International Journal of Mass Spectrometry, 2007. 259(1-3): p. 32-39.
12. Hillen, N. and S. Stevanovic, Contribution of mass spectrometry-based proteomics to immunology. Expert Rev Proteomics, 2006. 3(6): p. 653-64.
13. Nesvizhskii, A.I., O. Vitek, and R. Aebersold, Analysis and validation of proteomic data generated by tandem mass spectrometry. Nat Methods, 2007. 4(10): p. 787-97.
14. Oliveira, C.C., et al., The nonpolymorphic MHC Qa-1b mediates CD8+ T cell surveillance of antigen-processing defects. Journal of Experimental Medicine, 2010. 207(1): p. 207-21.
15. Perkins, D.N., et al., Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis, 1999. 20(18): p. 3551-67.
16. Eng, J.K., A.L. McCormack, and J.R. Yates, An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. J Am Soc Mass Spectrom, 1994. 5(11): p. 976-89.
17. The universal protein resource (UniProt). Nucleic Acids Res, 2008. 36(Database issue): p. D190-5.
18. Kersey, P.J., et al., The International Protein Index: an integrated database for proteomics experiments. Proteomics, 2004. 4(7): p. 1985-8.
19. Pruitt, K.D., T. Tatusova, and D.R. Maglott, NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Res, 2007. 35(Database issue): p. D61-5.
20. Schandorff, S., et al., A mass spectrometry-friendly database for cSNP identification. Nat Methods, 2007. 4(6): p. 465-6.
21. Edwards, N.J., Novel peptide identification from tandem mass spectra using ESTs and sequence database compression. Mol Syst Biol, 2007. 3: p. 102.
22. Salimi, N., et al., Design and utilization of epitope-based databases and predictive tools. Immunogenetics, 2010. 62(4): p. 185-96.
23. Reisinger, F. and L. Martens, Database on Demand - an online tool for the custom generation of FASTA-formatted sequence databases. Proteomics, 2009. 9(18): p. 4421-4.
24. Sherry, S.T., et al., dbSNP: the NCBI database of genetic variation. Nucleic Acids Res, 2001. 29(1): p. 308-11.
25. Etzold, T., A. Ulyanov, and P. Argos, SRS: information retrieval system for molecular biology data banks. Methods Enzymol, 1996. 266: p. 114-28.

26. Hiemstra, H.S., et al., The identification of CD4+ T cell epitopes with dedicated synthetic peptide libraries. Proc Natl Acad Sci U S A, 1997. 94(19): p. 10313-8.
27. Stepniak, D., et al., Large-scale characterization of natural ligands explains the unique gluten-binding properties of HLA-DQ2. J Immunol, 2008. 180(5): p. 3268-78.
28. Meiring, M.S., et al., In vitro effect of a thrombin inhibition peptide selected by phage display technology. Thromb Res, 2002. 107(6): p. 365-71.
29. Oliveira, C.C., et al., The nonpolymorphic MHC Qa-1b mediates CD8+ T cell surveillance of antigen-processing defects. J Exp Med, 2010. 207(1): p. 207-21.



CHAPTER 4

Identification of biological relevant minor histocompatibility antigens within the B-lymphocyte derived HLA-ligandome using a reverse immunology approach

Clin Cancer Res (2015) 21:2177-2186

Based on:

Pleun Hombrink*
Chopie Hassan*
Michel G.D. Kester
Lorenz Jahn
Margot J. Pont
Arnoud H. de Ru
Cornelis A.M. van Bergen
Marieke Griffioen
J.H. Frederik Falkenburg
Peter van Veelen#
Mirjam H.M. Heemskerk#

*equal contribution
Shared Senior Authorship

4

Identification of biological relevant minor histocompatibility antigens within the B-lymphocyte derived HLA-ligandome using a reverse immunology approach

ABSTRACT

T-cell recognition of minor histocompatibility antigens (MiHA) plays an important role in the beneficial graft-versus-leukemia (GVL) effect of allogeneic stem cell transplantation (allo-SCT) but also mediates serious graft versus host disease (GVHD) complications associated with allo-SCT. Using a reverse immunology approach we aim to develop a method enabling the identification of T-cell responses directed against predefined antigens, with the goal to select those MiHAs that can be used clinically in combination with allo-SCT. In this study, we used a recently developed MiHA selection algorithm to select candidate MiHAs within the HLA-presented ligandome of transformed B-cells. From the HLA-presented ligandome,that predominantly consisted of monomorphic peptides, 25 polymorphic peptides with a clinically relevant allele frequency were selected. By high-throughput screening the availability of high-avidity T-cells specific for these MiHA-candidates in different healthy donors was analyzed. With the use of MHC-multimer enrichment, analyses of expanded T-cells by combinatorial coding MHC-multimer flow cytometry, and subsequent single cell cloning, T-cell clones directed to 2 new MiHA: LB-CLYBL-1Y and LB-TEP1-1S could be demonstrated, indicating the immunogenicity of these 2 MiHAs. The biological relevance of MiHA LB-CLYBL-1Y was demonstrated by the detection of LB-CLYBL-1Y specific T cells in a patient suffering from acute myeloid leukemia (AML) that experienced an anti-leukemic response after treatment with allo-SCT.

INTRODUCTION

Allogeneic HLA-matched hematopoietic stem cell transplantation (allo-SCT) and subsequent donor lymphocyte infusion (DLI) to eradicate persistent or relapsed malignant cells are considered an effective curative treatment for patients with high-risk hematological malignancies or patients who fail to respond to chemotherapy [1]. The curative potential of this therapy has been attributed to the recognition of malignant cells by donor T-cells. Detailed analyses of T-cell immune responses in patients responding to DLI have demonstrated that the donor T-cells recognize minor histocompatibility antigens (MiHA) presented in the context of HLA-molecules on malignant cells. MiHA are peptides derived from polymorphic proteins that differ between donor and recipient due to single nucleotide polymorphisms (SNP) [2-5]. Previously it has been demonstrated that T-cells directed against MiHA that are ubiquitously expressed can mediate life-threatening graft-versus-host-disease (GVHD) [6], whereas T-cells directed against MiHA with hematopoietic restriction may mediate graft-versus-leukemia (GVL) response in absence of GVHD [7].

Although the introduction of various molecular and genomic technologies resulted in an increased number of HLA-class I restricted MiHA identified by forward immunological approaches, the number of MiHA with therapeutic relevance is still limited [8,9]. A major limitation for clinical implementation of MiHA is the small number of identified MiHA derived from genes that are exclusively expressed in hematopoietic cells. To allow the selective analysis of hematopoietic restricted MiHA, we and others have used reverse immunological approaches in which predicted polymorphic peptides are the starting point and peptide candidates are subsequently screened for their capacity to induce a specific T-cell response. This approach has the potential to specifically screen for T-cells recognizing MiHA that are exclusively expressed by hematopoietic cells. However, it has been reported that when such peptide predictions are solely based on computer algorithms that predict peptide-HLA binding affinity and proteolytic cleavage, the detected T-cell responses are often directed against epitopes that are not naturally processed and presented and therefore fail to lyse malignant target cells [10-12]. To circumvent this peptide selection problem we and others previously introduced mass spectrometry (MS) based HLA-ligandomes as a reliable source for naturally processed and presented peptides and developed an algorithm that can be exploited to identify T-cells specific for potential clinically relevant MiHA[13,14].

The emerging availability of high quality SNP data in combination with the implementation of HLA-ligandomes provides a large supply of MiHA-candidates with guaranteed processing and HLA-restricted presentation at the cell surface [15]. We used our recently established algorithm [13] to select the top 25 MiHA candidates from our newly established large set of B-lymphocyte derived HLA-class I eluted peptides [16].

To validate these MiHA candidates is essential to demonstrate their immunogenic potential by isolation of high avidity specific T-cells. We therefore screened the TCR repertoire of 16 unrelated donors for the presence of T-cells specific for these MiHA-candidates. MHC-multimer positive T-cells lines were generated from peripheral blood mononuclear cells (PBMC) by magnetic activated cell sorting (MACS). Flow cytometric analysis of antigen-specific T-cells, followed by functional testing of MHC-multimer sorted and expanded T-cell clones using both IFN-γ and GM-CSF as readout demonstrated the immunogenic potential of LB-CLYBL-1Y, LB-TEP1-1S and two previously described MiHA. For one of the newly defined MiHA we were able to confirm the biological relevance by demonstrating MHC-multimer positive T-cells in a patient suffering from AML after treatment with allo-SCT and subsequent DLI. Our data illustrate that with this reverse immunology approach biologically relevant MiHA can be identified as well as MiHA that are not frequently induced in vivo but can potentially be used for immunotherapeutic strategies.

EXPERIMENTAL PROCEDURE

Cell collection and culture conditions

Peripheral blood was obtained from different individuals after informed consent (Sanquin Reagents, Amsterdam and Leiden University Medical Center, Leiden, The Netherlands). All experiments were approved by the local medical ethics committees. Blood samples were HLA-typed by high-resolution genomic DNA-typing. Peripheral blood mononuclear cells (PBMC) were isolated by Ficoll gradient separation and cryopreserved for further use. T-cells were cultured in T-cell medium consisting of Iscove's Modified Dulbecco's Media (IMDM: Lonza) supplemented with 5% human serum (HS), 5% fetal bovine serum (FBS), 100 IU/ml IL-2 (Proleukin), 2 mM L-glutamine, 100 U/ml penicillin, and 100 μg/ml streptomycin (Invitrogen). Epstein-Barr virus-transformed lymphoblastoid B-cell lines (EBV-LCL) and phytohaemagglutinin (PHA; Murex Diagnostics) blasts were generated using standard procedures. The T2 cell line was cultured in IMDM with penicillin/streptomycin and 5% FCS.

Peptide library of HLA-class I eluted peptides

The peptides used for this study are derived from a newly established large peptide library that has been recently described by Hassan et al [16]. In short: Peptide elution, reverse phase high-performance liquid chromatography (RP-HPLC), and tandem mass spectrometry (MS/MS) were carried out as previously described [17,18]. Briefly, 60 x 109 HLA-A*0201 and HLA-B*0702 positive EBV-LCL were washed with ice cold PBS and pellets were stored at -80°C until use. Peptide-HLA-class I complexes were purified from cell lysate by affinity chromatography. Subsequently, peptides were eluted from HLA-molecules, and separated from the

HLA heavy chain fragments and β2-microglobulin by size filtration. The peptide mixture was separated by various first dimension separation techniques, after which the peptides were measured by on-line nanoHPLC-MS/MS.

Selection of MiHA-candidates from a set of eluted peptides

MiHA-candidates were selected based on our recently developed MiHA selection algorithm from the recently established peptide elution library [13]. Briefly, the tandem mass spectra of eluted peptides were submitted to the HSPVdb [19], a database optimized for finding polymorphic peptides. To select for MiHA-candidates within this set we evaluated the polymorphic peptides using strict threshold scores for mass spectrometry defined sequence reliability (BMI≥30, ppm≤2.0), peptide length [8-11 amino acids), predicted peptide-HLA affinity (<500 nM), allele frequencies of the SNP encoding polymorphism (0.05-0.7%) and specific gene exclusion criteria (no extreme polymorphic genes). After confirming their integrity by comparing the tandem mass spectra of the synthetic peptides with that of the eluted counterparts the top 25 MiHA-candidates, with the highest threshold scores, were selected for further analysis.

Generation of peptide-MHC complexes

All peptides were synthesized in-house using standard Fmoc chemistry. Recombinant HLA-A*0201 and HLA-B*0702 heavy chain and β2m were in-house produced in Escherichia coli. MHC-class I refolding was performed as previously described with minor modifications [20]. MHC-class I complexes were purified by gel-filtration HPLC in PBS and stored at 4 ºC. The peptide HLA-A*0201 or HLA-B*0702 binding affinity was assessed by subjecting prefolded UV-sensitive peptide-MHC complexes (100 μg/ml) to 366 nm UV light (Camag) for 1 hr. in presence of the peptide of interest (200 μM) [21]. As controls the CMV PP65 NLVPMVATV, CMV PP65 TPRVTGGAM, modified MART1 ELAGIGILTV and wild-type MART1 AAGIGILTV peptide were used. After exchange, samples were spun at 16,000g for 5 min and supernatants were used to assess HLA-monomer rescue using a bead assay as previously described [22].

Isolation of MHC-multimer positive T-cells by MHC-multimer pull down

Prior to isolation PBMC samples were stained with PE-coupled MHC-multimers for 1 hour at 4 ºC. Subsequently, cells were washed and incubated with anti-PE beads (MiltenyiBiotec, BergischGladbach, Germany). PE-positive T-cells were isolated by magnetic activated cell sorting (MACS), according to the manufacturer's protocol. Positive fraction was cultured with irradiated autologous feeder cells in T-cell medium supplemented with 5 ng/ml IL-15 (Biosource) and anti-CD3/CD28 Dynabeads (Invitrogen). After 2 weeks, T-cell cultures

were analyzed by MHC-multimer combinatorial coding flow cytometry. Data acquisition was performed on an LSR-II flow cytometer (BD Biosciences, San Diego, USA) and MHC-multimer positive T-cell populations were single cell sorted on a FACSAria (BD) into 96-round bottomed well plates containing 100 µl T-cell medium supplemented with 800 ng/ml PHA and 1*105 irradiated feeder cells. T-cell receptor-β variable chain (TCR-Vβ) usage was investigated by flow cytometry using specific monoclonal antibodies as included in the TCR-Vβ repertoire kit (Beckman Coulter).

Cytokine secretion assay

For analysis of IFN-γ and GM-CSF production, 5000 T-cells were stimulated with 20,000 EBV-LCL or 10,000 fibroblasts loaded with different concentrations of peptides in 96-round bottomed well plates. Before stimulation, fibroblasts were either pre-treated with IFN-γ (100 IU/ml) or not for 5 days. For recognition of primary malignant cells, 1000 T-cells were stimulated with 5,000 malignant cells in a 384-well plate. After 18 hours, supernatants were harvested, and the concentration of IFN-γ and GM-CSF were measured by an enzyme-linked immunosorbent assay (ELISA; Sanquin Reagents). An arbitrary detection limit was set to 100 pg/ml for both cytokines.

SNP genotyping and microarray gene expression analysis

A panel of 100 HLA-typed EBV-LCL was selected for SNP screening as previously described [9]. Briefly, genomic DNA was isolated of 5×106 EBV-LCL cells using GentraPuregene Cell Kit (Qiagen) and PCR-free whole genome amplification (WGA) was performed. The DNA samples were hybridized to Illumina Human1M-duo arrays. For gene expression analysis, total RNA was isolated using small and micro scale RNAqueous isolation kits (Ambion), and amplified using the TotalPrep RNA amplification kit (Ambion).
After preparation using the whole-genome gene expression direct hybridization assay (Illumina), cRNA samples were dispensed onto Human HT-12 v3 Expression BeadChips (Illumina) according to the manufacturer's protocol. Sample collection was performed as previously described [23-25]. Microarray gene expression data were analyzed using R 2.15.

RESULTS

Selection of high affinity HLA-A*0201 and B*0702 binding MiHA-candidates from a set of HLA-class I eluted peptides

We have recently reported a MiHA selection algorithm to be able to select MiHA candidates from a library of HLA-class I eluted peptides [13]. This MiHA selection algorithm comprises several evaluation steps that are summarized in the material and methods section. This algorithm was used to screen our newly established library of eluted HLA-class I peptides derived from multiple HLA-A*0201 and B*0702 positive EBV-LCLs, to select for potential MiHA candidates [16].To validate this newly established library of approximately 16.000 eluted HLA class I peptides comprising of mainly monomorphic peptides, we first screened for the presence of known MiHA. Peptide sequences or their relevant length variants were identified for 10 out of 13 MiHA that

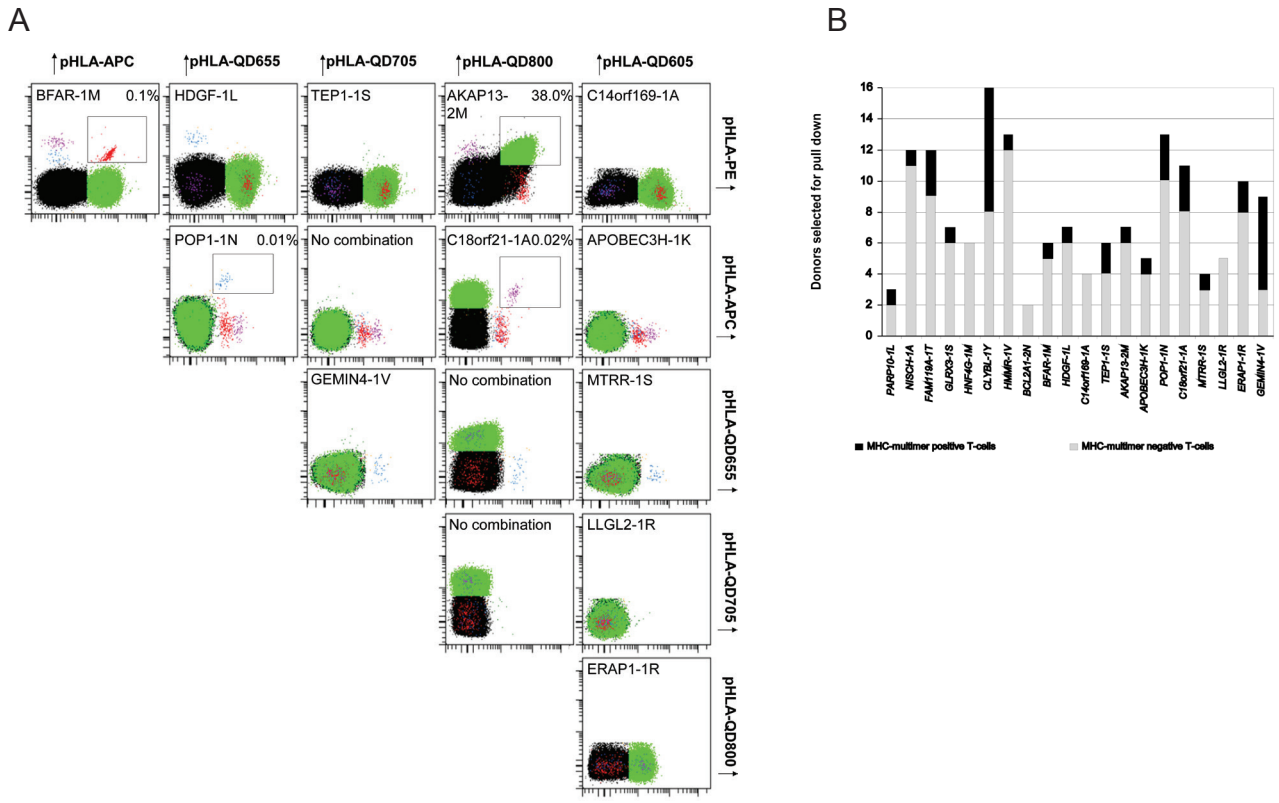


Figure 1. Detection of MiHA-specific T-cells after MHC-multimer enrichment and expansion. FACS analyses were performed to detect MiHA-specific T-cells through combinatorial coding MHC-multimer screening after second pull down and in vitro expansion. Each peptide-MHC complex was encoded by a unique combination of fluorochromes. Two HLA-restricted sets of MHC-multimers were used to screen for detection of all MiHA-candidates specific T-cells. (A) Representative analysis of the two times enriched T-cell line derived from donor OMH with the HLA-B*0702 restricted MHC-multimer set. Total $CD8^+$ T-cells are shown and MHC-multimer negative cells or cells staining with more than 2 fluorochromes are indicated in black. All dot plots are shown with bi-exponential axes and display fluorescence intensity for the fluorochromes indicated on each axis. Frequencies indicate MHC-multimer positive T-cells of total $CD8^+$ cells. (B) Total MHC-multimer positive T-cell populations detected in 16 tested PBMC donors after second pull down by combinatorial coding MHC-multimer screen. Bars indicate the number of donors homozygous negative for the indicated SNP encoding MiHA and applicable for MHC-multimer pull down. The number of MHC-multimer positive T-cell populations detected per MiHA candidate is indicated in black. No homozygous negative donors were found for SCRIB-1L, and therefore SCRIB-1L is not indicated in the Figure. At least two homozygous negative donors were included for the other 20 MiHA-candidates.

were expressed by the EBV-LCLs as revealed by SNP genotyping (Supplemental Table 1) [3,8,9], illustrating the high quality of this peptide elution library.

Table I. Validated MiHA candidates

HLA	Peptide	Sequenceα	Gene	RS number	SNP	Allele Frequency
A*02:01	P1	GL<u>L</u>GQEGLVEI	PARP10	rs11136343	L/P	0.66%
	P2	ALAPAP<u>A</u>EV	NISCHβ	rs887515	A/V	0.17%
	P3	AMLERQF<u>T</u>V	FAM119A	rs2551949	T/I	0.19%
	P4	FL<u>S</u>SANEHL	GLRX3	rs2274217	S/P	0.25%
	P5	M<u>M</u>YKDILLL	HNF4G	rs1805098	M/I	0.40%
	P6	SLAA<u>Y</u>IPRL	CLYBL	rs17577293	Y/D	0.05%
	P7	SLQEK<u>V</u>AKA	HMMR	rs299295	V/A	0.20%
	P8	VLQ<u>N</u>VAFSV	BCL2A1	rs1138358	N/K	0.69%
B*07:02	P9	APNTGRANQQ<u>M</u>	BFAR	rs11546303	M/R	0.57%
	P10	LPMEVEKNST<u>L</u>	HDGF	rs4399146	L/P	0.40%
	P11	RPR<u>A</u>PTEELAL	C14orf169	rs3813563	A/V	0.40%
	P12	APDGAKVA<u>S</u>L	TEP1	rs1760904	S/P	0.49%
	P13	APAGVREV<u>M</u>	AKAP13	rs7162168	M/T	0.37%
	P14	KPQQ<u>K</u>GLRL	APOBEC3H	rs139298	K/E	0.52%
	P15	LPQKKS<u>N</u>AL	POP1	rs17184326	N/K	0.10%
	P16	LPQQPP<u>L</u>SL	SCRIB	rs6558394	L/P	0.64%
	P17	NPATP<u>A</u>SKL	C18orf21	rs2276314	A/T	0.21%
	P18	SPAS<u>S</u>RTDL	MTRR	rs1532268	S/L	0.68%
	P19	SPSL<u>R</u>ILAI	LLGL2	rs1671036	R/H	0.50%
	P20	HP<u>R</u>QEQIAL	ERAP1β	rs34753	R/P	0.31%
	P21	FPALRFVE<u>V</u>	GEMIN4β	rs1045481	V/E	0.20%

α SNP underlined

β Published MiHA epitope or length variant

Non– and Immunogenic Allele indicated by amino acid code, allele frequencies are calculated by quantification in a panel of 100 Dutch individuals using SNP genotyping array.

In the next step we selected a set of 25 MiHA-candidates using the MiHA selection algorithm, including 22 novel MiHA-candidates as well as the previously reported LB-NISCH-1A, LB-ERAP1-1R and LB-GEMIN4-1V MiHA (Supplemental Table 2). We first analyzed the capacity of the 25 MiHA-candidates to stabilize a peptide-HLA complex in a HLA-binding assay that is based on UV-induced conditional ligand cleavage as described previously [10,22,26]. After UV-exchange, the HLA-binding affinity of the tested peptides was normalized to the high-affinity control peptides CMV-PP65-NLV and CMV-PP65-TPR for HLA-A*0201 and B*0702, respectively (Supplemental Figure 1). MiHA-candidates were selected when their HLA-binding affinity exceeded that of the MART1-WT-AAG peptide, which binds with low affinity to HLA-A*0201. HLA-

binding affinity as measured with HLA-rescue scores exceeded that of the MART1-WT-AAG control for 8 HLA-A*0201 and 13 B*0702 binding MiHA-candidates. Peptide sequences and allele frequencies of the MiHA are shown in Table 1.

Isolation of peripheral blood derived MHC-multimerpositive T-cells by MHC-multimer pull down

To validate the 21 peptides as MiHA with immunogenic potential,we generated MHC-multimers and analyzed the T-cell repertoire of 16 healthy HLA-A*0201 and B*0702 positive donors for MHC-multimer reactive T-cells. MiHA-specific T-cell lines were generated by incubating 100*106 peripheral blood mononuclear cells (PBMC) with a specific set of MHC-multimers, followed by enrichment of MHC-multimer positive cells on a magnetic column. To allow the isolation of high avidity T-cell populations, the set of MHC-multimers was specifically adjusted for each PBMC sample to cover only those MiHA for which the encoding SNP was

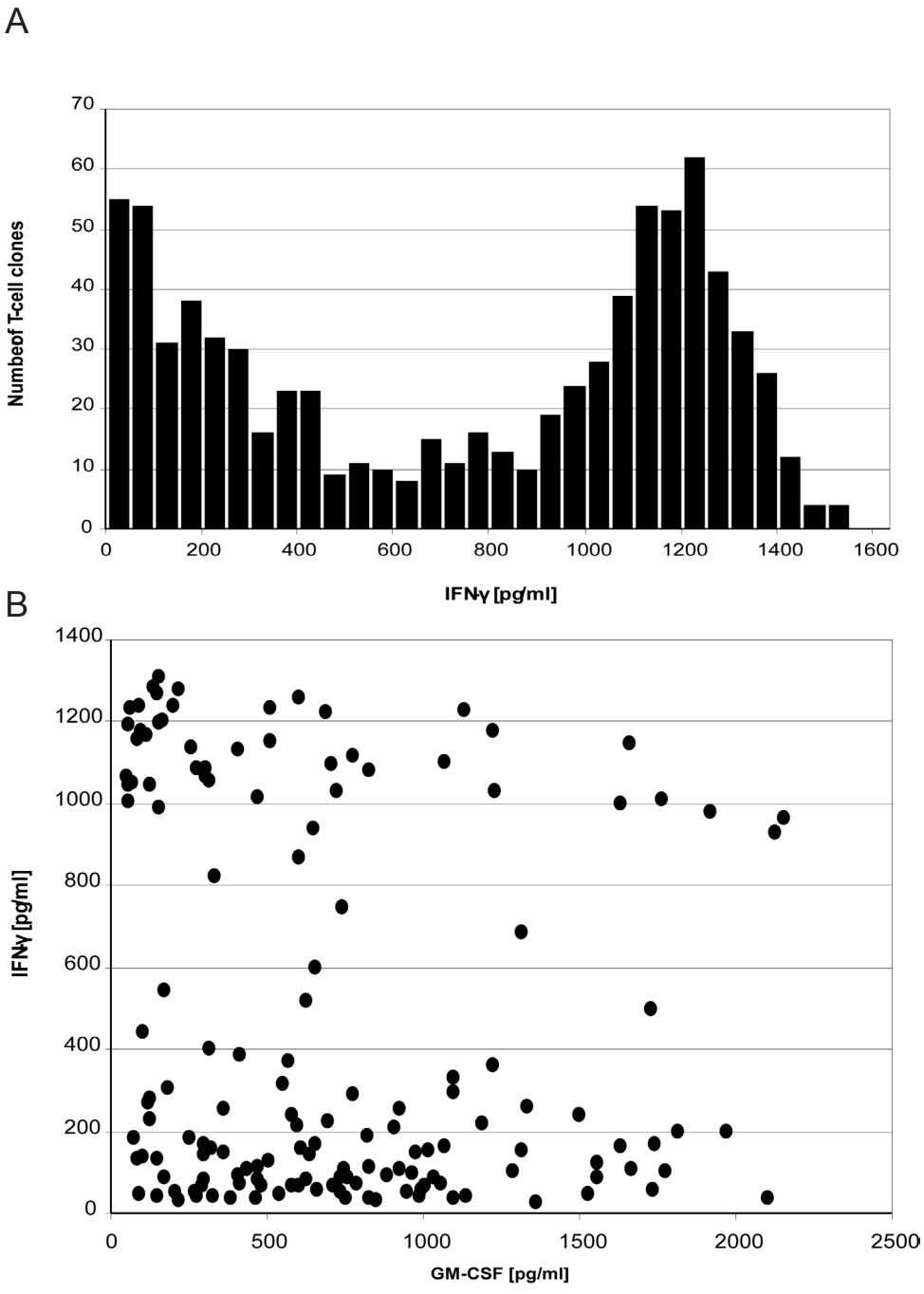


Figure 2. Isolated T-cell clones vary in production of IFN-γ and GM-CSF upon αCD3/28 stimulation (A) Distribution of the intrinsic IFN-γ secretion capacity of 806 MiHA-specific MHC-multimer positive T-cell clones stimulated with αCD3/28 beads. (B) Production of GM-CSF for 150 selected T-cell clones with variable IFN-γ secretion capacity. MiHA specific MHC-multimer positive T-cell clones were stimulated for 18 hours in presence of αCD3/28 beads and GM-CSF and IFN-γ secretion was measured by standard ELISA.

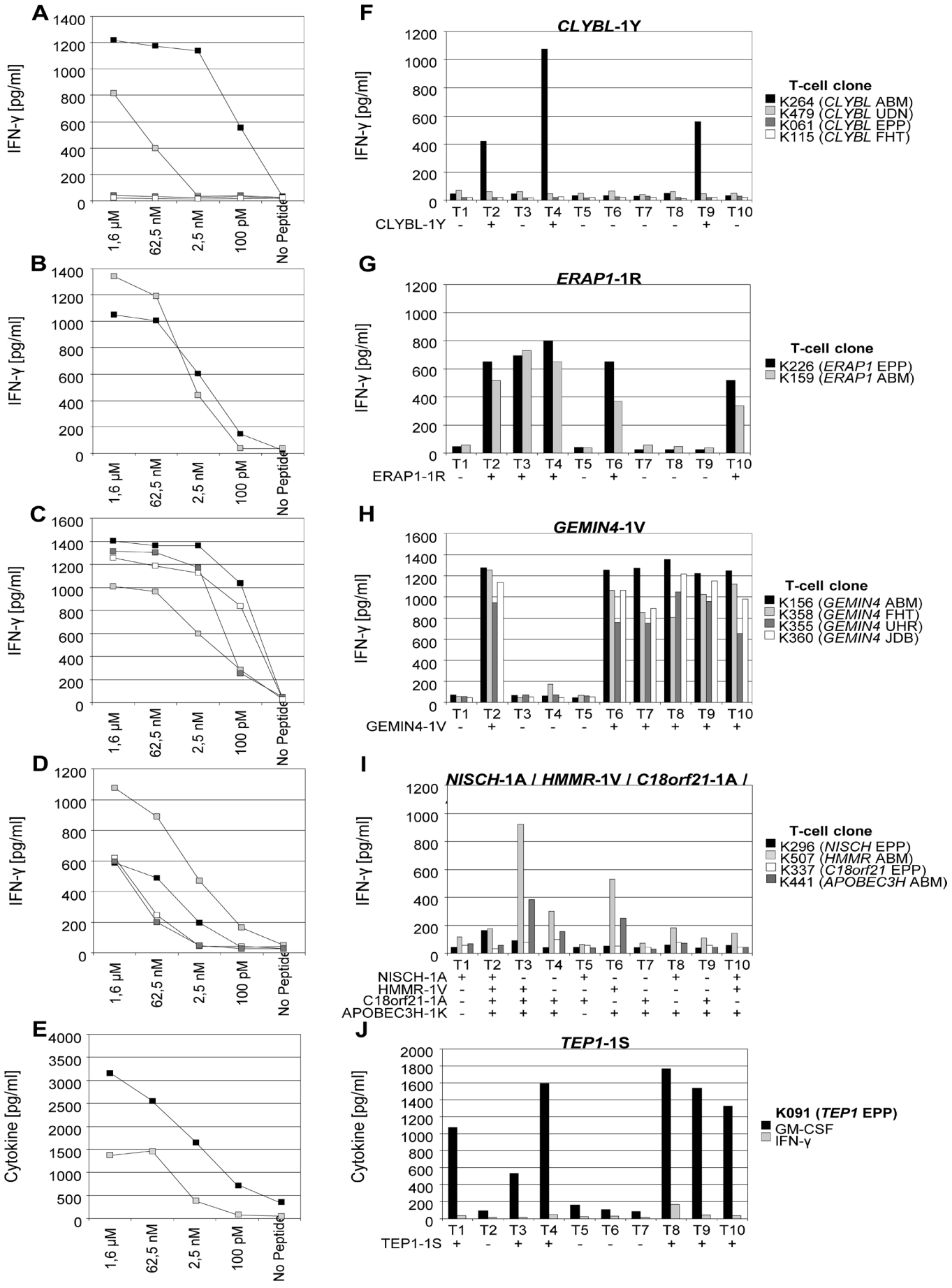

A
B
C
D
E
IFN-γ [pg/ml]
Cytokine [pg/ml]
1,6 µM
62,5 nM
2,5 nM
100 pM
No Peptide
F
CLYBL-1Y
CLYBL-1Y - + - + - - - - + -
G
ERAP1-1R
ERAP1-1R - + + + - + - - - +
H
GEMIN4-1V
GEMIN4-1V - + - - - + + + + +
I
NISCH-1A / HMMR-1V / C18orf21-1A /
NISCH-1A + + - - + - - + - +
HMMR-1V - + + - - + - - - +
C18orf21-1A - + + + + - + - + -
APOBEC3H-1K - + + + - + + + + +
J
TEP1-1S
TEP1-1S + - + + - - - + + +
T1 T2 T3 T4 T5 T6 T7 T8 T9 T10
T-cell clone
K264 (CLYBL ABM)
K479 (CLYBL UDN)
K061 (CLYBL EPP)
K115 (CLYBL FHT)
K226 (ERAP1 EPP)
K159 (ERAP1 ABM)
K156 (GEMIN4 ABM)
K358 (GEMIN4 FHT)
K355 (GEMIN4 UHR)
K360 (GEMIN4 JDB)
K296 (NISCH EPP)
K507 (HMMR ABM)
K337 (C18orf21 EPP)
K441 (APOBEC3H ABM)
K091 (TEP1 EPP)
GM-CSF
IFN-γ

Figure 3. MHC-multimer positive T-cell clones demonstrate MiHA specific reactivity. Isolated MHC-multimer positive T-cell clones were stimulated with HLA-A*0201 and B*0702 positive T2 cells loaded with titrated concentrations of peptides for 18 hours [E:T] 5,000:25,000. For each donor the MiHA specific T-cell clone that demonstrated the highest peptide avidity is shown for the respective cytokine. MiHA specific IFN-γ production was observed for T-cell clones specific for (A) CLYBL-1Y, (B) ERAP1-1R, (C) GEMIN4-1V (D) NISCH-1A, HMMR-1V, C18orf21-1A, APOBEC3H-1K,and (E) TEP1-1S MiHA-candidates. For the TEP1-1S specific T-cell clone K091 derived from donor EPP both the peptide induced IFN-γ and GM-CSF secretion are shown. (F,G,H,I,J) To measure the reactivity of the MiHA specific T-cell clones against endogenously processed and presented antigens, all MiHA specific T-cell clones that demonstrated IFN-γ or GM-CSF production in response to specific peptide were stimulated with a panel of 10 HLA-A*0201 and B*0702 positive EBV-LCL target cells positive (+) or negative (-) for the indicated MiHA for 18 hours. Cytokine secretion was measured by standard ELISA.

screened homozygous negative in the respective donor, since in this individual's T-cell repertoire high avidity MiHA specific T-cells will not have been deleted due to negative selection. The set of PBMC donors was specifically adjusted to cover as many applicable donors per MiHA candidate. Unfortunately, no homozygous negative donors were found for the SCRIB-1L MiHA candidate. The MHC-multimer enriched T-cells were expanded for 14 days in presence of αCD3/28 beads, IL-2 and IL-15. To increase the frequency of MHC-multimer positive T-cells, a second pull down was performed at day 14 using the identical initial set of MiHA candidate specific MHC-multimers. After both rounds of MHC-multimer enrichment we analyzed the expanding T-cell lines for the presence of MHC-multimer positive T-cells by FACS. A representative FACS analysis after two rounds of MHC-multimer enrichment is demonstrated in Figure 1A in which 4 MHC-multimer positive T-cell populations specific for HLA-B*0702 binding MiHA-candidates are detected in donor OMH. After the first MHC-multimer pull down, MHC-multimer positive T-cell populations were detected specific for 11 of the 20 tested MiHA-candidates in one or more T-cell cultures. MHC-multimer positive T-cell frequencies varied between 0.01% and 5.0% of total $CD8^+$ T-cells (Supplemental Table 3). These low frequencies are most likely due to the very low frequency of MHC-multimer positive T-cells in the naïve repertoire. After the second MHC-multimer pull down, MHC-multimer positive T-cell populations were detected specific for an increased number of 16 of the 20 tested MiHA-candidates with frequencies up to 85% of total $CD8^+$ T-cells (Supplemental Table 3). As demonstrated in Figure 1B, T-cells reactive with the CLYBL-1Y and LB-GEMIN4-1V MHC-multimer were frequently detected in 8 of the 16 and 6 of the 9 enriched T-cell lines, respectively. In contrast, T-cells specific for the other MiHA appeared more restricted to a few donors.

Detection of high avidity T-cell clones by screening for MiHA-specific IFN-γ and GM-CSF production

Next, a total number of 806 MHC-multimer positive T-cell clones representing all detected T-cell populations was generated by FACS sorting. To demonstrate recognition of potential MiHA-candidates, all 806 T-cell clones were stimulated with specific antigen and cytokine production was determined as a measure for anti-

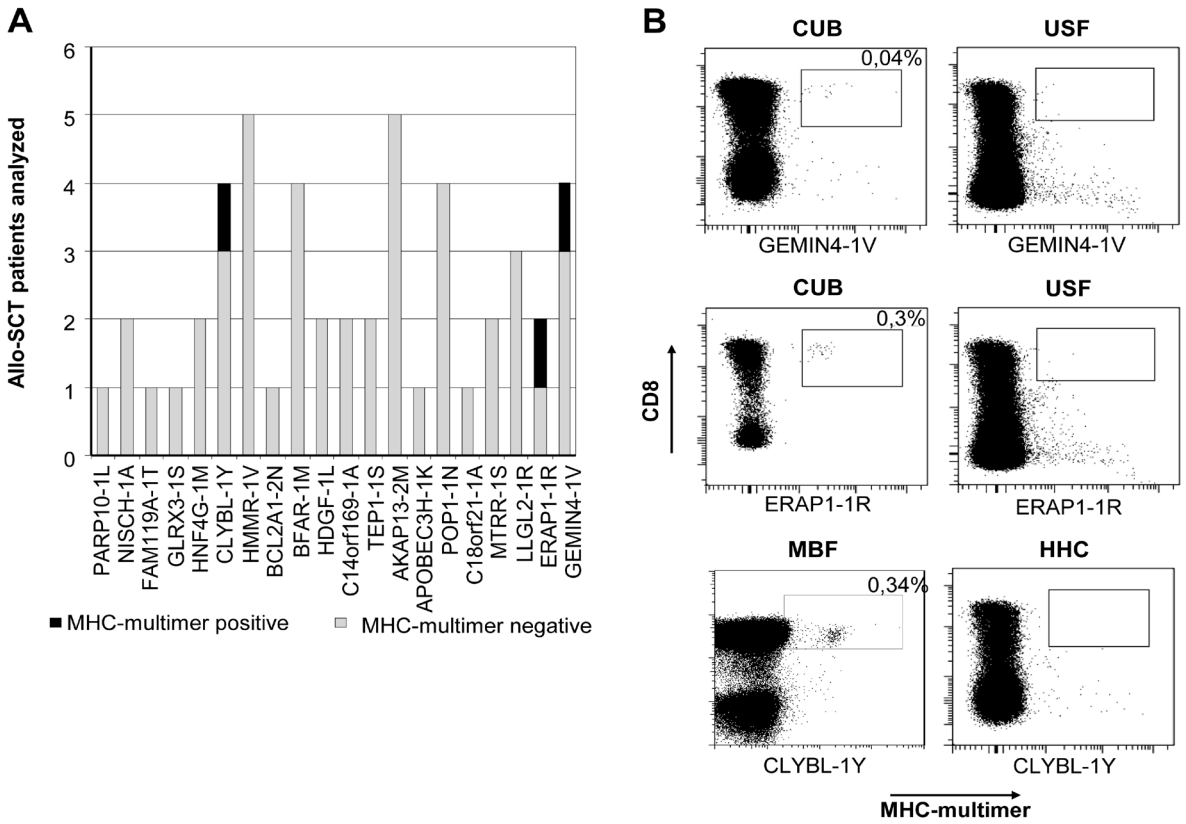


Figure 4. MHC-multimer positive T-cell populations detected in patients after allo-SCT. (A) Number of MHC-multimer positive T-cell populations detected by combinatorial coding MHC-multimer analysis in peripheral blood samples of 16 patients with various hematologic malignancies that received an allo-SCT and demonstrated a clinical response to DLI. Patients were only screened when they were positive for the SNP encoding MiHA and received a DLI from a donor that was homozygous negative for the SNP encoding MiHA. Bars indicate the number of patients applicable for MiHA-specific MHC-multimer screening. The number of MHC-multimer positive T-cell populations detected per MiHA candidate is indicated in black. (B) Representative FACS analysis of MHC-multimer positive T-cell populations detected in the indicated patients. All dot plots display fluorescence intensity for CD8 and specific MHC-multimer staining. Total lymphocytes are shown. Frequencies indicate MiHA specific T-cells of total CD8+ T-cells.

gen specific reactivity. Because the T-cell clones were most likely derived from the naive T-cell repertoire and have presumably not all acquired the capacity to produce IFN-γ upon antigen encounter, we first screened all T-cell clones for their potential to secrete IFN-γ after αCD3/28 stimulation. As demonstrated in Figure 2A, the generated T-cell clones demonstrated a broad range of IFN-γ secretion. To investigate whether GM-CSF was of additional value to improve the screening efficiency, the GM-CSF production of part of the T-cell clones with variable IFN-γ secretion potential was measured after αCD3/28 stimulation. The results demonstrate that a substantial number of T-cell clones with poor intrinsic IFN-γ production was able to produce pronounced GM-CSF levels (Figure 2B). Therefore by using both IFN-γ and GM-CSF as a readout we could increase the number of MHC-multimer positive T-cell clones that could be screened. By setting an arbitrary detection limit to 100 pg/ml for both cytokines, 98% of generated T-cell clones could subsequently be screened for MiHA specific reactivity.

To demonstrate the capacity of the isolated MHC-multimer positive T-cell clones to recognize specific MiHA

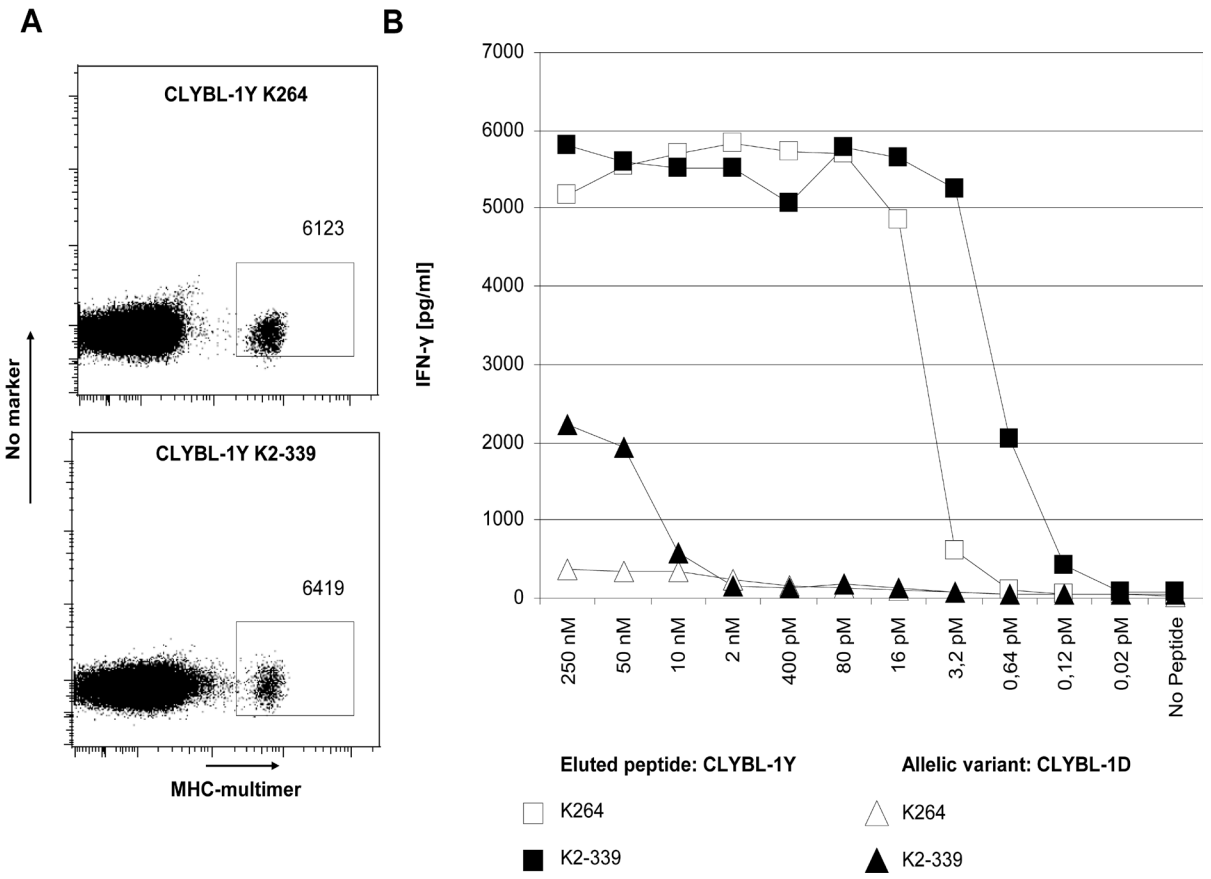


Figure 5. Functional avidity of patient and donor derived CLYBL-1Y specific T-cell clones. (A) FACS analysis of MHC-multimer positive CLYBL-1Y specific T-cell clones. To prevent doublet formation, CLYBL-1Y specific T-cell clones were mixed with CD4 $^{+}$ T-cell blasts. Dot plots display fluorescence intensity for MHC-multimer staining. Total lymphocytes are shown. (B) CLYBL-1Y specific T-cell clones derived from donor ABM (K264; white) or from patient MBF (K2-339; black) were stimulated with HLA-A*0201 positive T2 target cells loaded with titrated concentrations of CLYBL-1Y peptide (squares) or allelic variant (triangles)for 18 hours [E:T] 5,000:25,000.IFN-γ secretion was measured by standard ELISA.

peptides we stimulated all T-cell clones with HLA-A*0201 and B*0702 positive T2 target cells loaded with titrated concentrations of specific peptide. MiHA specific T-cell reactivity was observed for 8 out of 16 tested MiHA-candidates and either IFN-γ or GM-CSF was detected after stimulation with CLYBL-1Y, ERAP1-1R, GEMIN4-1V, NISCH-1A, HMMR-1V, C18orf21-1A, APOBEC3H-1K and TEP1-1S MiHA peptides. For each donor the MiHA specific T-cell clone that demonstrated the highest peptide avidity is shown (Figure 3A, B, C, D and E). MiHA specific T-cell clones demonstrated variable peptide avidity and half maximum cytokine production (IC50) varied between an IC50 of ±100 pM for the high avidity T-cell clone K156 specific for GEMIN4 (Figure 3C) and an IC50 of ±62.5 nM for the low avidity T-cell clone K337 specific for C18orf21 (Figure 3E). For the CLYBL-1Y MiHA candidate, MHC-multimer positive T-cell clones were successfully isolated from 4 different donors. High avidity CLYBL-1Y specific T-cell clones were isolated from donor ABM, whereas only low avidity or non-reactive T-cells were isolated from donor UDN, EPP and FHT respectively (Figure 3A). For the previously described MiHA ERAP1-1R, both high and low avidity T-cell clones were isolated from the 2 donors that demonstrated MHC-multimer positive T-cell populations after pull down.

For the previously described MiHA GEMIN4-1V, high avidity T-cell clones were only isolated from 4 out of 6 donors that demonstrated MHC-multimer positive T-cell populations after pull down (Figure 3B,C; low-avidity T-cell clones not shown). The NISCH-1A, HMMR-1V, C18orf21-1A and APOBEC3H-1K specific T-cell clones were each successfully isolated from one donor and exhibited high (HMMR-1V) to low peptide avidity (NISCH-1A, C18orf21-1A and APOBEC3H-1K) (Figure 3D). For the TEP1-1S MiHA candidate high avidity T-cell clones were isolated from donor EPP. T-cell clone K091 demonstrated high peptide specific GM-CSF production and low IFN-γ production (Figure 3E). No peptide specific T-cell clones specific for the other MiHA-candidates were observed.

All T-cell clones that demonstrated peptide specific cytokine production were stimulated with a panel of 10 SNP genotyped HLA-A*0201 and B*0702 positive EBV-LCL target cells to screen for reactivity against

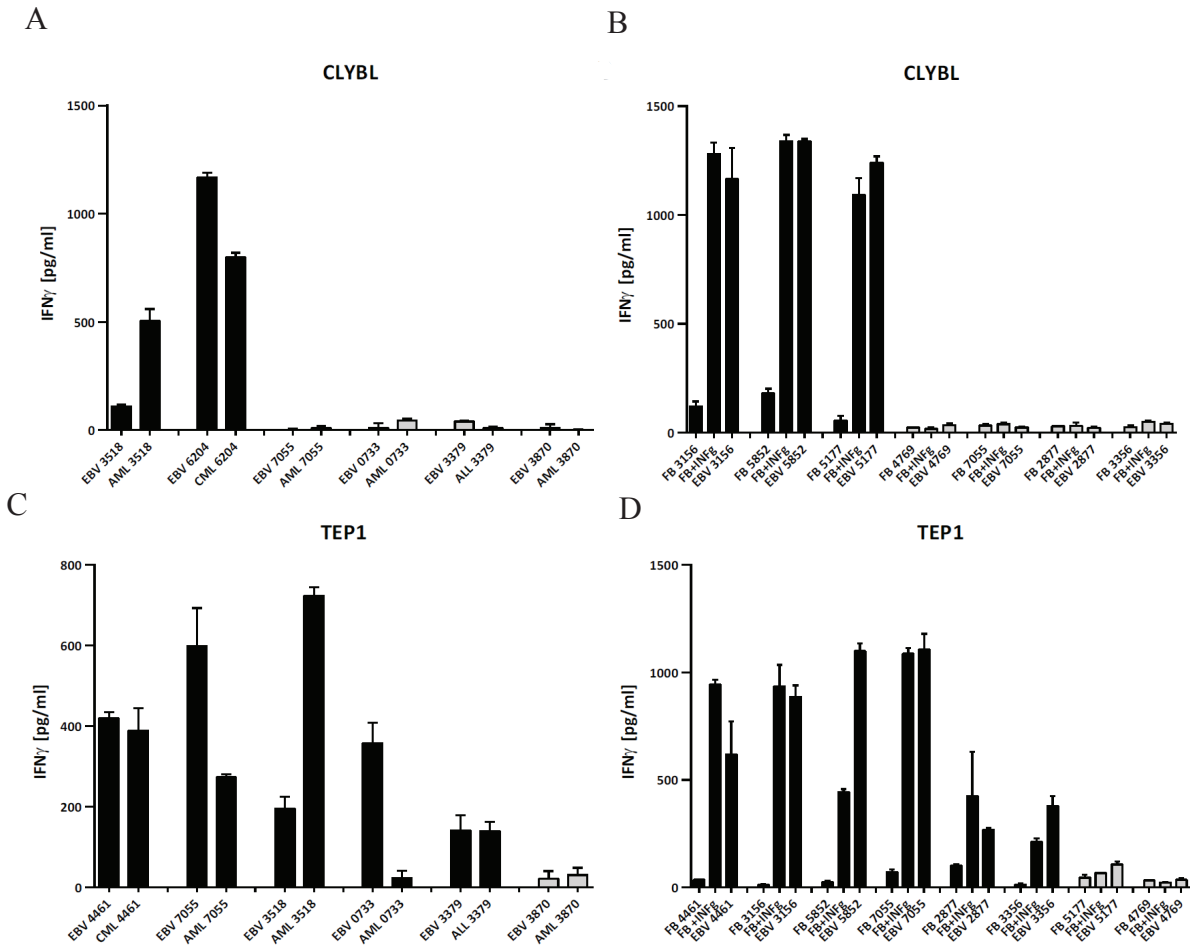


Figure 6. MiHA specific T-cell reactivity towards hematopoietic malignancies and fibroblasts. Isolated MHC-multimer positive T-cell clones were screened for recognition of various primary hematopoietic malignancies and non-hematopoietic fibroblasts isolated from different individuals. As a control for T-cell recognition, all T-cell clones were also stimulated with EBV-LCL generated from the same individuals. (A) CLYBL-1Y clone K264 and (C) TEP1-1S clone K091 were stimulated with HLA-A*0201 or B*0702 positive primary AML, CML and ALL cells either positive (black bars) or negative (gray bars) for the specific MiHA, directly after isolation of malignant cells for 18 hours [E:T] 1,000:5,000. To measure the recognition of non-hematopoietic cells, (B) CLYBL-1Y clone K264 and (D) TEP1-1S clone K091 were stimulated with fibroblasts either pre-treated with IFN-γ (100 IU/ml) or not. MiHA specific recognition of target cells was measured by standard IFN-γ ELISA.

endogenously processed and presented peptide. As demonstrated in Figure 3F, G, H and J all high avidity T-cell clones specific for CLYBL-1Y, TEP1-1S, ERAP1-1R and GEMIN4-1V demonstrated recognition of all target cells that endogenously process and present their respective MiHA, whereas targets that were negative for the MiHA were not recognized. Surprisingly, the high avidity T-cell clone K508 specific for HMMR-1V did not show a recognition pattern that correlated with MiHA expression. This may be caused by the absence of endogenously presented MiHA peptide by some of the SNP positive EBV-LCL or by the recognition of allo-HLA molecules expressed by EBV-LCL. No or only marginal target cell recognition was observed for the NISCH-1A, C18orf21-1A and APOBEC3H-1K specific T-cell clones (Figure 3I). These results indicate that the MiHA CLYBL-1Y and TEP1-1S represent potentially immunological relevant MiHA-candidates.

Detection of MiHA specific T-cell responses in patients after allo-SCT

To validate the biological relevance of the MiHA-candidates we analyzed the peripheral blood of patients suffering from various hematopoietic malignancies that received an allo-SCT and DLI and demonstrated a clinical response revealed by declining patient chimerism, for the detection of MiHA specific T-cells. Patients were only screened with MiHA-specific MHC-multimers when they were positive for the MiHA and received a DLI from a donor who was homozygous negative for the SNP encoding MiHA (Supplemental Table 4). Tested patient samples were obtained during or after the peak response, 5 to 8 weeks after treatment with DLI .For the newly identified CLYBL-1Y MiHA, MHC-multimer positive T-cells were detected in 1 of the 3 screened patients (Figure 4A) with a frequency of 0.34% of total $CD8^+$ T-cells at day 41 after DLI (AML patient MBF, Figure 4B). For the previously described ERAP1-1R and GEMIN4-1V MiHA, MHC-multimer positive T-cells were detected in 1 of the 2 and 1 of the 4 screened patients (both multiple myeloma (MM) patient CUB at day 86), respectively (Figure 4A,B). Detected frequencies of circulating MHC-multimer positive T-cells ranged between 0.04% and 0.34% of total $CD8^+$ T-cells. For the other 17 MiHA-candidates with validated HLA-binding affinity and SNP occurrence, including the newly identified MiHA TEP1-1S, no MHC-multimer positive T-cells were detected in the peripheral blood of 1 to 5 screened patients (Figure 4 A,B, Supplemental Table 4).

To investigate whether the MHC-multimer enriched T-cell clones exerted comparable peptide specific avidity as the in vivo generated patient derived MiHA specific T-cells, we generated CLYBL-1Y specific T-cell clones by single cell sorting of CLYBL-1Y MHC-multimer positive T-cells from patient MBF. After expansion the T-cell clones were stained with the MHC-multimer and TCR-Vβ mAbs. The T-cell clone K264, which was generated by MHC-multimer enrichment from donor ABM, demonstrated similar MHC-multimer staining intensity as the patient derived K2-339 clone (Figure 5A), but a difference in TCR Vβ usage; clones derived

from donor ABM expressed TCR Vβ22 and patient MBF derived clones were TCR-Vβ1 positive. Clone K264 and K2-339 were stimulated with T2 cells loaded with titrated concentrations of either the specific or the allelic variant peptide and IFN-γ production was measured (Figure 5B). Both T-cell clones demonstrated high CLYBL-1Y specific peptide reactivity, with IC50 varying between 1-4 pM, whereas the allelic CLYBL-1D variant was not or hardly recognized by both T-cell clones, demonstrating that T-cell clones derived from an unprimed setting can be equally potent as T-cells derived from an in vivo primed setting.

CLYBL-1Y and TEP1-1S specific T-cell recognition of hematopoietic malignant cells

To investigate the expression pattern of the CLYBL and TEP1 genes we performed a microarray gene expression array using a panel of primary and cultured malignant (and non-malignant) hematopoietic and non-hematopoietic cells (Supplemental Figure 2). The data showed that the CLYBL gene is broadly expressed in hematopoietic and non-hematopoietic cells. Expression of the TEP1 gene was not significantly measured in the majority of the samples. To investigate whether the CLYBL-1Y and TEP1-1S specific T-cell clones were able to recognize hematopoietic malignant cells they were stimulated with primary CML, AML and ALL cells derived from different MiHA positive and negative patients that were positive for the restricting HLA-molecule. As a control, T-cell clones were also tested for recognition of EBV-LCL generated from the same individuals. Both the high avidity T-cell clones CLYBLK264 and TEP1K091 demonstrated MiHA specific recognition of primary hematopoietic malignant cells (Figure 6 A&C). No reactivity was observed against MiHA negative target cells. These data indicate that the tested MiHA can be presented in the context of HLA at the surface of leukemic cells and may therefore serve as direct targets of $CD8^+$ T-cells involved in a GVL response. In addition, the potential of the MiHA to serve as target in GVHD was estimated and the MiHA specific T-cell clones were tested for recognition of non-hematopoietic fibroblasts. To mimic the pro-inflammatory cytokine milieu, early after transplant or during potent GVHD responses, fibroblasts were pretreated with IFN-γ. Although both the CLYBLK264 and TEP1K091 T-cell clone poorly recognized non treated fibroblasts they clearly recognized those that were IFN-γ pretreated (Figure 6 B&D). These data indicate that both CLYBL-1Y and TEP1-1S may be considered as MiHA with potential therapeutic value under non-inflammatory conditions, but they may participate in toxic GVHD responses in a pro-inflammatory environment.

DISCUSSION

In this study we demonstrate the identification of two biologically relevant MiHA by a reverse immunology approach. We started our approach by selecting MiHA-candidates from our recently generated B-lymphocyte derived HLA-class I eluted peptide library [13,16]. By MiHA candidate specific MHC-multimer enrichments and subsequent single cell sorting of MHC-multimer positive T-cells, high avidity MiHA specific T-cell clones directed against the MiHA TEP1-1S and CLYBL-1Y could be identified that recognize primary hematological malignancies expressing the respective MiHA, indicating the immunogenicity of the two MiHA. In addition, we were able to demonstrate an in vivo induced immune response against the new CLYBL-1Y MiHA in a patient suffering from AML that experienced an anti-leukemic response after allo-SCT and subsequent DLI treatment, demonstrating that by using this reversed immunology approach, biologically relevant MiHA can be identified.

For the CLYBL-1Y MiHA, MHC-multimer analysis revealed the presence of T-cells specific for this MiHA in one of the 3 patients screened. These in vivo primed T-cells demonstrated to be high avidity T-cells specific for the CLYBL-1Y allelic MiHA variant. For the TEP1-1S MiHA only two patients could be screened for the presence of TEP1-1S specific MHC-multimer positive cells. Therefore, the absence of an in vivo induced immune response against TEP1-1S could either be due to the low number of patients that could be screened or due to subdominant of TEP1-1S in the immune response. Our reverse immunology approach has the advantage to allow identification of subdominant MiHA as the T-cell repertoire of patients that are screened in forward immunology approaches is skewed towards highly immunodominant MiHA specific responses. Subdominant MiHA, however, may be of therapeutic interest as they can be exploited in potential peptide vaccination or adoptive T-cell therapies when they demonstrate promising gene expression patterns.

A major limitation for the identification of large numbers of MiHA using a reverse immunology approach may be the low frequency of high avidity MiHA specific T-cells within an individual's T-cell repertoire. By performing two rounds of MHC-multimer enrichments each followed by a 10 days expansion step we increased the number of MHC-multimer positive T-cells that was isolated, since we observed that after the second enrichment round previously undetected MHC-multimer positive T-cell populations were found. In addition, by measuring GM-CSF in addition to IFN-γ as readout for T-cell reactivity, we were able to increase the number of MHC-multimer positive T-cell clones that could be screened for functional activity. However, the failure to isolate high-avidity T-cells for the previously identified LB-NISCH-1A MiHA indicates that isolation of high avidity T-cells may still be a matter of chance [13]. Alternatively, T-cells may be tolerant for some MiHA due to molecular mimicry with non-polymorphic epitopes or due to their failure to discriminate between both

allelic variants of a SNP encoding MiHA [27], explaining why not all MiHA-candidates will be of clinical relevance.

For 8 of the 20 MiHA-candidates MHC-multimer positive T-cell clones were isolated that demonstrated MiHA specific peptide reactivity. The different MHC-multimer positive T-cell clones however demonstrated functional heterogeneity. We have recently demonstrated that CMV-specific MHC-multimer positive T-cells isolated from CMV negative individuals by MHC-multimer enrichments also demonstrated a large variation in functional avidity [28]. This heterogeneity in functional activity of the MHC-multimer positive T-cell populations may be specific for T-cells isolated from the naïve repertoire as memory T-cells are skewed toward a high-avidity range as a result of antigen encounter in vivo. Non-responsiveness of MHC-multimer positive T-cell clones has also been reported by others for T-cells derived from the naive repertoire [29,30].The discrepancy between MHC-multimer reactivity and T-cell functionality can most likely be explained by the staining with multimerized MHC-peptide complexes. Multimerization of MHC-peptide complexes alters the TCR-MHC-peptide dissociation on- and off-rate kinetics and will result in increased binding avidity of the multimerized MHC-peptide complex to surface TCR. By measuring the strength of TCR binding to monomeric peptide-MHC complexes using the StreptamerKoff rate assay, we recently demonstrated that the dissociation kinetics correlated with the observed functional avidity of different MHC-multimer positive T-cell clones, and that T-cells demonstrating lower dissociation rates confer significantly better antigen specific reactivity than those with fast dissociation rates [28,31].

For the high throughput isolation of antigen specific T-cells using recombinant MHC-peptide complexes the use of MHC-multimer complexes is necessary,since MHC-monomers do not stably bind to TCRs. However, multimerization of MHC-peptide complexes may false positively increase the binding avidity of the multimerized MHC-peptide complex to T-cells. The analysis of conventional MHC-multimer based TCR dissociation kinetics is complicated by multivalent binding and dissociation rates, even when competitor reagents or cold MHC-multimer staining are used [32].The controlled disassembly of MHC-multimers to monomers directly after MHC-multimer based T-cell isolation and subsequent FACS sorting of T-cells expressing a TCR with a low MHC-peptide dissociation rate may further increase the efficiency of future reverse immunology approaches [33].

In conclusion, with this reverse immunology approach we have demonstrated the immunogenic potential of two newly identified MiHA; LB-CLYBL-1Y and LB-TEP1-1S. The biologically relevance of LB-CLYBL-1Y was demonstrated by the detection of MHC-multimer positive T-cells with equally high peptide avidity in a patient suffering from AML who experienced an anti-leukemic response after treatment with allo-SCT and subsequent DLI. The identification of LB-TEP1-1S demonstrated the feasibility of our reverse immunology

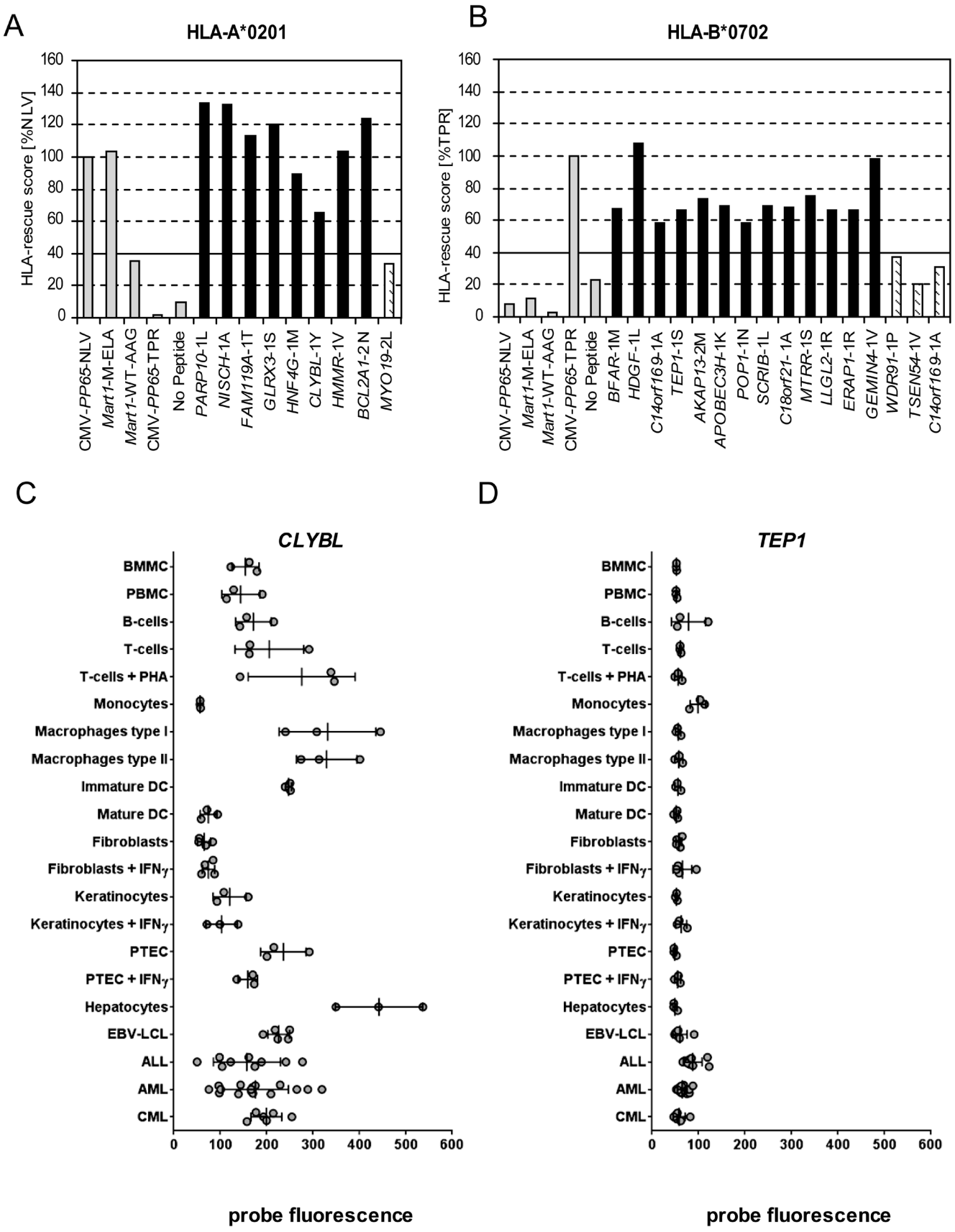
approach to identify MiHA that are not frequently induced in vivo but can potentially be used for immuno-therapeutic strategies. In addition, these results demonstrate that this approach will enable the identification of clinically relevant MiHA when MiHA candidates are selected based on a hematopoietic restricted gene expression pattern.

REFERENCES

1. Appelbaum, F. R. 2001. Haematopoietic cell transplantation as immunotherapy. Nature 411:385-389. Falkenburg, J. H., van de Corp, E. W. Marijt, and R. Willemze. 2003. Minor histocompatibility antigens in human stem cell transplantation. Exp. Hematol. 31: 743-751.
2. Feng, X., K. M. Hui, H. M. Younes, and A. G. Brickner. 2008. Targeting minor histocompatibility antigens in graft versus tumor or graft versus leukemia responses. Trends Immunol. 29: 624-632.
3. Goulmy, E. 1997. Human minor histocompatibility antigens: new concepts for marrow transplantation and adoptive immunotherapy. Immunol Rev 157: 125-140.
4. Mullally, A., and J. Ritz. 2007. Beyond HLA: the significance of genomic variation for allogeneic hematopoietic stem cell transplantation. Blood 109: 1355-1362.
5. Ferrara, J. L., J. E. Levine, P. Reddy, and E. Holler. 2009. Graft-versus-host disease. Lancet 373: 1550-1561.
6. Warren, E. H., P. D. Greenberg, and S. R. Riddell. 1998. Cytotoxic T-lymphocyte-defined human minor histocompatibility antigens with a restricted tissue distribution. Blood 91: 2197-2207
7. Bleakley, M., and S. R. Riddell. 2011. Exploiting T cells specific for human minor histocompatibility antigens for therapy of leukemia. Immunol. Cell Biol. 89: 396-407.
8. van Bergen, C. A., C. E. Rutten, E. D. van der Meijden, S. A. van Luxemburg-Heijs, E. G. Lurvink, J. J. Houwing-Duistermaat, M. G. Kester, A. Mulder, R. Willemze, J. H. Falkenburg, and M. Griffioen. 2010. High-throughput characterization of 10 new minor histocompatibility antigens by whole genome association scanning. Cancer Res. 70: 9073-9083.
9. Hombrink, P., S. R. Hadrup, A. Bakker, M. G. Kester, J. H. Falkenburg, P. A. von dem Borne, T. N. Schumacher, and M. H. Heemskerk. 2011. High-throughput identification of potential minor histocompatibility antigens by MHC tetramer-based screening: feasibility and limitations. PLoS. One. 6: e22523.
10. Ofran, Y., H. T. Kim, V. Brusic, L. Blake, M. Mandrell, C. J. Wu, S. Sarantopoulos, R. Bellucci, D. B. Keskin, R. J. Soiffer, J. H. Antin, and J. Ritz. 2010. Diverse patterns of T-cell response against multiple newly identified human Y chromosome-encoded minor histocompatibility epitopes. Clin. Cancer Res. 16: 1642-1651.
11. Popovic, J., L. P. Li, P. M. Kloetzel, M. Leisegang, W. Uckert, and T. Blankenstein. 2011. The only proposed T-cell epitope derived from the TEL-AML1 translocation is not naturally processed. Blood 118: 946-954.
12. Hombrink, P., C. Hassan, M. G. Kester, A. H. de Ru, C. A. Van Bergen, H. Nijveen, J. W. Drijfhout, J. H. Falkenburg, M. H. Heemskerk, and P. A. van Veelen. 2013. Discovery of T cell epitopes implementing HLA-peptidomics into a reverse immunology approach. J Immunol 190: 3869-3877.
13. Gilchuk, P., C. T. Spencer, S. B. Conant, T. Hill, J. J. Gray, X. Niu, M. Zheng, J. J. Erickson, K. L. Boyd, K. J. McAfee, C. Oseroff, S. R. Hadrup, J. R. Bennink, W. Hildebrand, K. M. Edwards, J. E. Crowe, Jr., J. V. Williams, S. Buus, A. Sette, T. N. Schumacher, A. J. Link, and S. Joyce. 2013. Discovering naturally processed antigenic determinants that confer protective T cell immunity. J Clin. Invest 123: 1976-1987.
14. The 1000 Genomes Project Consortium. 2010. A map of human genome variation from population scale sequencing. Nature 467: 1061-1073.
15. Hassan, C., M. G. Kester, A. H. de Ru, P. Hombrink, J. W. Drijfhout, H. Nijveen, J. A. Leunissen, M. H. Heemskerk, J. H. Falkenburg, and P. A. van Veelen. 2013. The human leukocyte antigen-presented ligandome of B lymphocytes. Mol. Cell Proteomics. 12: 1829-1843.
16. den Haan, J. M., N. E. Sherman, E. Blokland, E. Huczko, F. Koning, J. W. Drijfhout, J. Skipper, J. Shabanowitz, D. F. Hunt, V. H. Engelhard, and . 1995. Identification of a graft versus host disease-associated human minor histocompatibility antigen. Science 268: 1476-1480.
17. van Bergen, C. A., M. G. Kester, I. Jedema, M. H. Heemskerk, S. A. van Luxemburg-Heijs, F. M. Kloosterboer, W. A. Marijt, A. H. de Ru, M. R. Schaafsma, R. Willemze, P. A. van Veelen, and J. H. Falkenburg. 2007. Multiple myeloma-reactive T cells recognize an activation-induced minor histocompatibility antigen encoded by the ATP-dependent interferon-responsive (ADIR) gene. Blood 109: 4089-4096.
18. Nijveen, H., M. G. Kester, C. Hassan, A. Viars, A. H. de Ru, J. M. de, J. H. Falkenburg, J. A. Leunissen, and P. A. van Veelen. 2011. HSPVdb--the Human Short Peptide Variation Database for improved mass spectrometry-based detection of polymorphic HLA-ligands.Immunogenetics 63: 143-153.
19. Burrows, S. R., N. Kienzle, A. Winterhalter, M. Bharadwaj, J. D. Altman, and A. Brooks. 2000. Peptide-MHC Class I Tetrameric Complexes Display Exquisite Ligand Specificity. J Immunol 165: 6229-6234.
20. Rodenko, B., M. Toebes, S. R. Hadrup, W. J. van Esch, A. M. Molenaar, T. N. Schumacher, and H. Ovaa. 2006. Generation of peptide-MHC class I complexes through UV-mediated ligand exchange. Nat. Protoc. 1: 1120-

1132.
21. Eijsink, C., M. G. Kester, M. E. Franke, K. L. Franken, M. H. Heemskerk, F. H. Claas, and A.Mulder. 2006. Rapid assessment of the antigenic integrity of tetrameric HLA complexes by human monoclonal HLA antibodies. J Immunol Methods 315: 153-161.
22. Nauta, A. J., H. S. de, B. Bottazzi, A. Mantovani, M. C. Borrias, J. Aten, M. P. Rastaldi, M. R. Daha, K. C. van, and A. Roos. 2005. Human renal epithelial cells produce the long pentraxin PTX3. Kidney Int. 67: 543-553.
23. Griffioen, M., M. W. Honders, E. D. van der Meijden, S. A. van Luxemburg-Heijs, E. G. Lurvink, M. G. Kester, C. A. van Bergen, and J. H. Falkenburg. 2012. Identification of 4 novel HLA-B*40:01 restricted minor histocompatibility antigens and their potential as targets for graft-versus-leukemia reactivity. Haematologica 97: 1196-1204.
24. Kremer, A. N., E. D. Van Der Meijden, M. W. Honders, J. J. Goeman, E. J. Wiertz, J. H. Falkenburg, and M. Griffioen. 2012. Endogenous HLA class II epitopes that are immunogenic in vivo show distinct behavior toward HLA-DM and its natural inhibitor HLA-DO. Blood 120:3246-3255.
25. Toebes, M., M. Coccoris, A. Bins, B. Rodenko, R. Gomez, N. J. Nieuwkoop, W. van de Kasteele, G. F. Rimmelzwaan, J. B. Haanen, H. Ovaa, and T. N. Schumacher. 2006. Design and use of conditional MHC class I ligands. Nat. Med. 12: 246-251.
26. Macdonald, W. A., Z. Chen, S. Gras, J. K. Archbold, F. E. Tynan, C. S. Clements, M. Bharadwaj, L. Kjer-Nielsen, P. M. Saunders, M. C. Wilce, F. Crawford, B. Stadinsky, D. Jackson, A. G. Brooks, A.W. Purcell, J. W. Kappler, S. R. Burrows, J. Rossjohn, and J. McCluskey. 2009. T cell allorecognition via molecular mimicry. Immunity. 31: 897-908.
27. Hombrink, P., Y. Raz, M. G. Kester, B. R. de, B. Weissbrich, P. A. von dem Borne, D. H. Busch, T. N. Schumacher, J. H. Falkenburg, and M. H. Heemskerk. 2013. Mixed functional characteristics correlating with TCR-ligand k -rate of MHC-tetramer reactive T cells within the naive T-cell repertoire. Eur. J Immunol. Ford, M. L., B. H. Koehn, M. E. Wagener, W. Jiang, S. Gangappa, T. C. Pearson, and C. P. Larsen. 2007. Antigen-specific precursor frequency impacts T cell proliferation, differentiation, and require ment for costimulation. J. Exp. Med. 204: 299-309.
28. Moon, J. J., H. H. Chu, M. Pepper, S. J. McSorley, S. C. Jameson, R. M. Kedl, and M. K. Jenkins. 2007. Naive CD4(+) T cell frequency varies for different epitopes and predicts repertoire diversity and response magnitude. Immunity. 27: 203-213.
29. Nauerth, M., B. Weissbrich, R. Knall, T. Franz, G. Dossinger, J. Bet, P. J. Paszkiewicz, L. Pfeifer, M. Bunse, W. Uckert, R. Holtappels, D. Gillert-Marien, M. Neuenhahn, A. Krackhardt, M. J. Red dehase, S. R. Riddell, and D. H. Busch. 2013. TCR-Ligand koff Rate Correlates with the Protective Capacity of Antigen-Specific CD8+ T Cells for Adoptive Transfer. Sci Transl Med 5: 192ra87.
30. Wang, X. L., and J. D. Altman. 2003. Caveats in the design of MHC class I tetramer/antigen-specific T lymphocytes dissociation assays. J Immunol Methods 280: 25-35.
31. Knabel, M., T. J. Franz, M. Schiemann, A. Wulf, B. Villmow, B. Schmidt, H. Bernhard, H. Wagner, and D. H. Busch. 2002. Reversible MHC multimer staining for functional isolation of T-cell populations and effective adoptive transfer. Nat. Med. 8: 631-637.

SUPPLEMENTAL DATA



Supplemental Figure 1. The HLA-binding affinity of 25 eluted highly potential MiHA candidates was analyzed by a binding assay that is based on HLA-recovery after UV-exchange monomer technology. After UV-exchange the HLA-monomer rescue score was normalized to the peptides CMV PP65-NLV and CMV PP65-TPR that exhibit a high binding affinity for HLA-A*0201 or B*0702, respectively. Peptides with a higher HLA-recovery score (>40%) compared to the low affinity HLA-A*0201 binding MART1-WT-AAG were selected for further analysis. As negative control no rescue peptide was added. All control peptides are indicated in grey. The percentage of HLA-recovery as a measure for HLA-binding affinity for 9 HLA-A*0201 (a) and 16 HLA-B*0702 (b) restricted MiHA candidates was shown. MiHA candidates with HLA-recovery score above the selection threshold of 40% are indicated in black. MiHA candidates that demonstrated a HLA-recovery below 40% were removed from the study and are indicated with stripes. (C and D) Gene expression of CLYBL and TEP1 were analyzed. Normal hematopoietic T-cells, monocytes, B-cells and hematopoietic stem cells (HSC) were isolated from (G-CSF mobilized) peripheral blood and bone marrow mononuclear cells (PBMC and BMMC) from donors by flowcytometry cell sorting based on expression of CD3, CD14, CD19 and CD34, respectively. Skin derived fibroblasts (FB), keratinocytes (KC) and PTEC were cultured with and without IFN-γ (100 IU/ml) and hepatocytes were freshly isolated from liver specimen. AML, ALL and CML cells were isolated from peripheral blood and bone marrow samples from patients by flowcytometry cell sorting based on expression of CD33, CD19 and CD34, respectively. EBV-LCL and PHA blasts were generated using standard procedures. Total RNA was isolated and whole-genome gene expression assay was performed as previously described (26). The probe fluorescence for the (a) CLYBL and (b)TEP1 genes is shown for each sample, and the mean with SD is indicated for each cell type.

Supplemental Table 1. Eluted published MiHA epitopes.

HLA	Genotyped MiHA	Eluted Sequenceα	MiHA positive EBV-LCL	Eluted and matched	BMI*
A*02:01	HA-2V	YIGEVLVSV	JY, ALY, HHC	JY, HHC	50.4
	HA-8R	RTLDKVLEV	JY, ALY	JY, ALY	33.8
	LB-SSR1-1S	VLFRGGPRGSLAVA	HHC	HHC	28.6
	LB-PRCP-1D	MWDVAEDLKAβ	HHC	HHC	14.3
	LB-WNK1-1I	TLSPEIITVβ	JY, HHC	HHC	28.4
	LB-NISCH-1A	ALAPAPAEV	JY	JY	39.0
	SMCY-A2	FIDSYICQV	JY	NDγ	-
B*07:02	LB-EBI3-1I	RPRARYYIQV	ALY	ALY	14.3
	LB-ERAP1-1R	HPRQEQIALβ	JY, HHC	JY, HHC	36.2
	LB-GEMIN4-1V	FPALRFVEV	JY	JY	42.5
	LB-PDCD11-1F	GPDSSKTFLCL	ALY, HHC	ND	-
	SMCY-B7	SPSVDKARAEL	JY	JY	57.8
	LB-APOBEC3B-1K	KPQYHAEMCF	ALY, HHC	NDγ	-

Total detected MiHA

α Sequence confirmed by matching the MS spectra of synthetic and eluted peptide

β Length variant of published MiHA epitope

γ Not Detected (ND)

* Highest detected Best Mascot Ion Score (BMI)

Supplemental Table 2. Validation scores MiHA candidates.

HLA	Peptide	Sequence[α]	Gene	PPM[β]	BMI[γ]	netMHC[μ] (nM)
A*02:01	P1	GLLGQEGLVEI	PARP10	0.7	56	96
	P2	ALAPAPAEV	NISCH	0.9	40	12
	P3	AMLERQFTV	FAM119A	0.1	35	3.4
	P4	FLSSANEHL	GLRX3	0.6	36	10.6
	P5	MMYKDILLL	HNF4G	0.1	63	6.9
	P6	SLAAYIPRL	CLYBL	0.8	40	4.1
	P7	SLQEKVAKA	HMMR	0.5	46	356.4
	P8	VLQNVAFSV	BCL2A1	1.1	40	12.7
	#	RLLEAIIRL	MYO19	0.1	50	10
B*07:02	P9	APNTGRANQQM	BFAR	0.6	44	55.8
	P10	LPMEVEKNSTL	HDGF	0.1	50	49.7
	P11	RPRAPTEELAL	C14orf169	0.1	38	3.3
	P12	APDGAKVASL	TEP1	0.3	45	73.1
	P13	APAGVREVM	AKAP13	0.1	72	20.5
	P14	KPQQKGLRL	APOBEC3H	0.3	36	19.6
	P15	LPQKKSNAL	POP1	0.3	38	6.2
	P16	LPQQPPLSL	SCRIB	0.4	54	14.2
	P17	NPATPASKL	C18orf21	0.1	43	71.6
	P18	SPASSRTDL	MTRR	0.8	45	11.1
	P19	SPSLRILAI	LLGL2	0.1	48	20.4
	P20	HPRQEQIAL	ERAP1	1.3	36	9
	P21	FPALRFVEV	GEMIN4	0.5	43	33
	#	SPRVGFLSSL	WDR91	0.6	56	17
	#	LPDGGVRLL	TSEN54	0.2	46	135
	#	RPRAPTEEL	C14orf169	0.1	36	12

Low HLA-affinity peptides

α SNP underlined

β Parts-Per-Million: Difference between observed and exact ion mass

γ Best-Mascot-Ion-Score: Match between observed MS spectrum and stated peptide

μ Predicted netMHC affinity (nM), Strong binder ≤50; Weak binder ≤500

Supplemental Table 3.

Expanded T-cell lines after first MHC-multimer pull down

Donor ID	*PARP10*-1L	*NISCH*-1A	*FAM119A*-1T	*GLRX3*-1S	*HNF4G*-1M	*CLYBL*-1Y	*HMMR*-1V	*BCL2A1*-2N	*BFAR*-1M	*HDGF*-1L	*C14orf169*-1A	*TEP1*-1S	*AKAP13*-2M	*APOBEC3H*-1K	*POP1*-1N	*C18orf21*-1A	*MTRR*-1S	*LLGL2*-1R	*ERAP1*-1R	*GEMIN4*-1V
FHT	-/-	-/-				0,5	-/-								-/-		-/-	-/-	-/-	0,2
EPP		0,01	-/-	-/-		0,2	-/-		-/-			0,01			-/-	0,01			0,01	
UBE		-/-	-/-			0,2	-/-			-/-		-/-	-/-	-/-	-/-	-/-				-/-
AKH		-/-	-/-	-/-		-/-				-/-		-/-							-/-	
JSB		-/-	-/-		-/-	-/-	-/-		-/-						-/-	-/-			-/-	2,8
DSP		-/-	-/-	-/-	-/-	-/-	-/-		-/-	-/-					-/-					-/-
USQ	-/-	-/-	-/-	-/-		-/-	-/-		-/-	-/-				-/-		-/-				
ABM			-/-			5,0	0,01				-/-			-/-	-/-				0,01	2,0
UDN		-/-	-/-			0,2	-/-						-/-	-/-	-/-	-/-		-/-		
OMH		-/-				-/-	-/-		-/-				0,01		-/-	-/-			-/-	
UBF		-/-			-/-	0,1		-/-	-/-	0,1	-/-	0,1			-/-	-/-	-/-			
OGV			-/-			0,01	-/-	-/-					-/-		0,01	-/-	-/-	-/-	-/-	
MMY		-/-	-/-	-/-	-/-	-/-					-/-		-/-			-/-	-/-		-/-	0,01
CLO		-/-	-/-	-/-	-/-	-/-						-/-	-/-		-/-	-/-			-/-	-/-
UHR				-/-		-/-	-/-			-/-					-/-	-/-		-/-		-/-
ADB	0,2		-/-		-/-	-/-	-/-			-/-	-/-	-/-	-/-	-/-	-/-			-/-	-/-	0,2

Expanded T-cell lines after second MHC-multimer pull down

Donor ID	*PARP10*-1L	*NISCH*-1A	*FAM119A*-1T	*GLRX3*-1S	*HNF4G*-1M	*CLYBL*-1Y	*HMMR*-1V	*BCL2A1*-2N	*BFAR*-1M	*HDGF*-1L	*C14orf169*-1A	*TEP1*-1S	*AKAP13*-2M	*APOBEC3H*-1K	*POP1*-1N	*C18orf21*-1A	*MTRR*-1S	*LLGL2*-1R	*ERAP1*-1R	*GEMIN4*-1V
FHT	-/-	-/-				0,8	-/-								-/-		0,1	-/-	-/-	65
EPP		28	0,3	0,01		8,4	-/-		-/-			2,0			-/-	6,7			3,6	
UBE		-/-	-/-			0,5	-/-			-/-		-/-	-/-	-/-	-/-	-/-				-/-
AKH		-/-	-/-	-/-		-/-				-/-		-/-							-/-	
JSB		-/-	-/-		-/-	-/-	-/-		-/-						-/-	-/-			-/-	2,8
DSP		-/-	-/-	-/-	-/-	-/-	-/-		-/-	-/-					-/-					-/-
USQ	-/-	-/-	-/-	-/-		-/-	-/-		-/-	-/-				-/-		-/-				
ABM			-/-			28,3	0,2				-/-			0,3	-/-				0,2	5,6
UDN		-/-	-/-			85	-/-						-/-	-/-	-/-	-/-		-/-		
OMH		-/-				-/-	-/-		0,1				38		0,01	0,02			-/-	
UBF		-/-			-/-	0,1		-/-	-/-	0,1	-/-	0,1			0,01	0,02	-/-			
OGV			0,2			54	-/-	-/-					-/-		0,01	-/-	-/-	-/-	-/-	
MMY		-/-	-/-	-/-	-/-	-/-					-/-		-/-			-/-	-/-		-/-	0,04
CLO		-/-	0,1	-/-	-/-	-/-						-/-	-/-		-/-	-/-			-/-	-/-
UHR				-/-		-/-	-/-			-/-					-/-	-/-		-/-		85
ADB	0,2		-/-		-/-	-/-	-/-			-/-	-/-	-/-	-/-	-/-	-/-			-/-	-/-	0,2

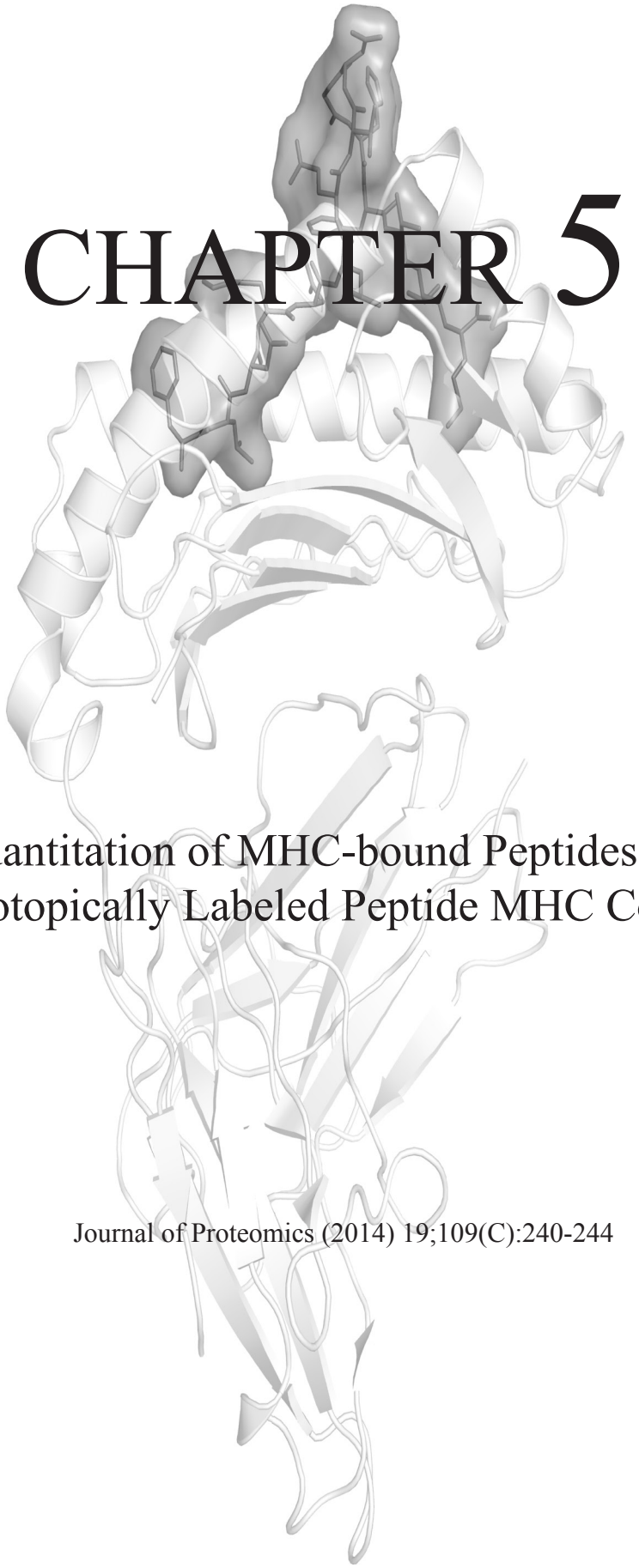
-/- Indicates T-cell population screened for in PBMC donor homozygous negative for the indicated SNP encoding MiHA. Detected T-cell populations are indicated in gray. Frequencies indicate MHC-multimer positive T-cells out of total $CD8^+$ T-cells as revealed by FACS analysis

Supplemental Table 4.

Circulating MHC-multimer positive T-cell frequencies detected in allo-SCT patients

Allo-SCT Patient	Dis-ease	Days After DLI	PARP10-1L	NISCH-1A	FAM119A-1S	GLRX3-1S	HNF4G-1M	CLYBL-1Y	HMMR-1V	BCL2A1-2N	BFAR-1M	HDGF-1L	C14orf169-1A	TEP1-1S	AKAP13-2M	APOBEC3H	POP1-1N	C18orf21-1S	MTRR-1S	LLGL2-1R	ERAP1-1R	GEMIN4-1V
BWB	CML	83																				
AZP	CML	62																				
BDV	MM	49																				
MDD	RCC	67																				
HHC	MDS	49																				
CUB	MM	86																			0,3	0,04
MBS	CMMOL	54																				
USF	PCL	95																				
JBZ	CML	59																				
OBB	CML	172																				
ASQ	AA	35																				
RBH	CML	106																				
MBF	AML	41						0,34														
YKV	CML	50																				
JLW	CML	90																				
MHU	MM	100																				

Screening for MHC-tetramer positive T-cell populations in 16 MiHA positive patient that received a DLI of a donor that was homozygous negative for the SNP-encoding MiHA. All T-cell populations screened for are indicated in gray
Frequencies indicate detected MHC-multimer positive T-cells out of total $CD8^+$ T-cells as revealed by FACS analysis



CHAPTER 5

Accurate Quantitation of MHC-bound Peptides by Application of Isotopically Labeled Peptide MHC Complexes

Journal of Proteomics (2014) 19;109(C):240-244

Based on:

Chopie Hassan
Michel G.D. Kester
Gideon Oudgenoeg
Arnoud H. de Ru
George M.C. Janssen
Jan W. Drijfhout
Robbert, M. Spaapen
Connie R. Jiménez
Mirjam H. M. Heemskerk
J. H. Frederik Falkenburg
Peter A. van Veelen

5

Accurate Quantitation of MHC-bound Peptides by Application of Isotopically Labeled Peptide MHC Complexes

ABSTRACT

Knowledge of the accurate copy number of HLA class I presented ligands is important in fundamental and clinical immunology. Currently, the best copy number determinations are based on mass spectrometry, employing single reaction monitoring (SRM) in combination with a known amount of isotopically labeled peptide. The major drawback of this approach is that the losses during sample pretreatment, i.e.immunopurification and filtration steps, are not well defined and must, therefore, be estimated. In addition, such losses can vary for individual peptides. Therefore, we developed a new approach in which isotopically labeled peptide-MHC monomers (hpMHC) are prepared and added directly after cell lysis, i.e. before the usual sample processing. Using this approach, all losses during sample processing can be accounted for and allows accurate determination of specific MHC class I-presented ligands. Our study pinpoints the immunopurification step as the origin of the rather extreme losses during sample pretreatment and offers a solution to account for these losses. Obviously, this has important implications for accurate HLA-ligand quantitation. The strategy presented here can be used to obtain a reliable view of epitope copy number and thus allows improvement of vaccine design and strategies for immunotherapy.

INTRODUCTION

Peptides presented by human leukocyte antigen (HLA) molecules on the cell surface play a crucial role in immunology, and mediate the communication between T cells and antigen presenting cells. Both the identity and abundance of these peptides are of great importance in fundamental studies on T cell action and in design of T cell-mediated therapy. These include tumor immunotherapy [1] and treatment of hematological malignancies by a combination of hematopoietic stem cell transplantation (HSCT) and treatment with donor T cells specific for minor histocompatibility antigens (MiHA) expressed on malignant cells from the patient [2]. Knowledge of accurate copy number of MHC class I presented epitopes can for example be used to study the relationship between differing epitope abundance and T cell immunodominance hierarchies [3, 4], and thereby help to improve vaccination strategies to elicit more effective immune responses. In addition, knowledge of ligand copy number is important in fundamental studies on antigen processing. Currently, the best copy number determinations are based on mass spectrometry, employing SRM in combination with mixing in a known amount of isotopically labeled peptide, known as the AQUA approach [3-6]. The AQUA standard can be mixed in at different stages during sample pretreatment, but preferably as early as possible. The complexity of the ligandome [7] demands various sample pretreatment steps before the actual peptide quantitation by LC-MS, involving cell lysis, immunoaffinity purification (IP), filtration and usually chromatographic fractionation. The major uncertainty in the accurate quantitation of ligand copy number is not caused by the final mass spectrometric experiment, but introduced by the variable yields of the sample pretreatment steps, which in addition can be peptide dependent. However, the overall yield of the sample pretreatment was not determined in previous studies, but rather estimated.

Therefore, we developed a new approach extending the AQUA approach in which isotopically labeled peptide-MHC monomers, i.e. peptide-MHC monomers loaded with a peptide in which two heavy amino acids have been incorporated (heavy peptide-MHC, or hpMHC), were prepared as a spike in standard and added directly following cell lysis, after which the usual sample processing was applied. Using this approach all losses during sample processing, including those during immunopurification, can be accounted for and it allows accurate determination of specific MHC class I-presented ligands, see Figure 1. We added a medium labeled peptide just before final LC-SRM analysis to enable determination of the overall yield of the sample pretreatment. Our hpMHC approach (addition of hpMHC before IP) was compared with the AQUA approach (addition of heavy peptide after IP).

T cells specific for MiHA have been shown to induce graft versus leukemia reactivity after allogeneic stem cell transplantation [2, 8, 9], and a number of MiHA has been used to design strategies for cellular immunotherapy but with limited success [10, 11]. An accurate quantitation of the presented peptides involved

may help to understand the efficacy of this treatment. Here, we illustrate our approach by analysis of two recently discovered minor histocompatibility antigens (MiHA), presented in the HLA-A*0201 molecule called (LB-NISCH-1A, ALAPAPAEV) [12], and LB-SSR1-1S, VLFRGGPRGSLAVA [13]. To assess the copy number of the endogenous (light) MiHA, we added the HLA-A*0201-monomer loaded with stable heavy isotope labeled peptide A**L**A**P**APAEV or V**L**FRGG**P**RGSLAVA (ΔM=+13 Da) as internal reference (heavy isotope labeled amino acids are in bold and underlined). The hpMHC were prepared by folding of recombinant HLA alpha chain and β2m in the presence of the heavy isotope labeled peptide. We combined the hpMHC with the AQUA strategy by using the medium labeled peptide A**L**APAPAEV or V**L**FRGGPRGSLAVA (ΔM=+7 Da), which was added to the chromatographic fractions just before final LC-MS quantitation. The medium labeled peptide for AQUA was added to each fraction prior to LC-SRM

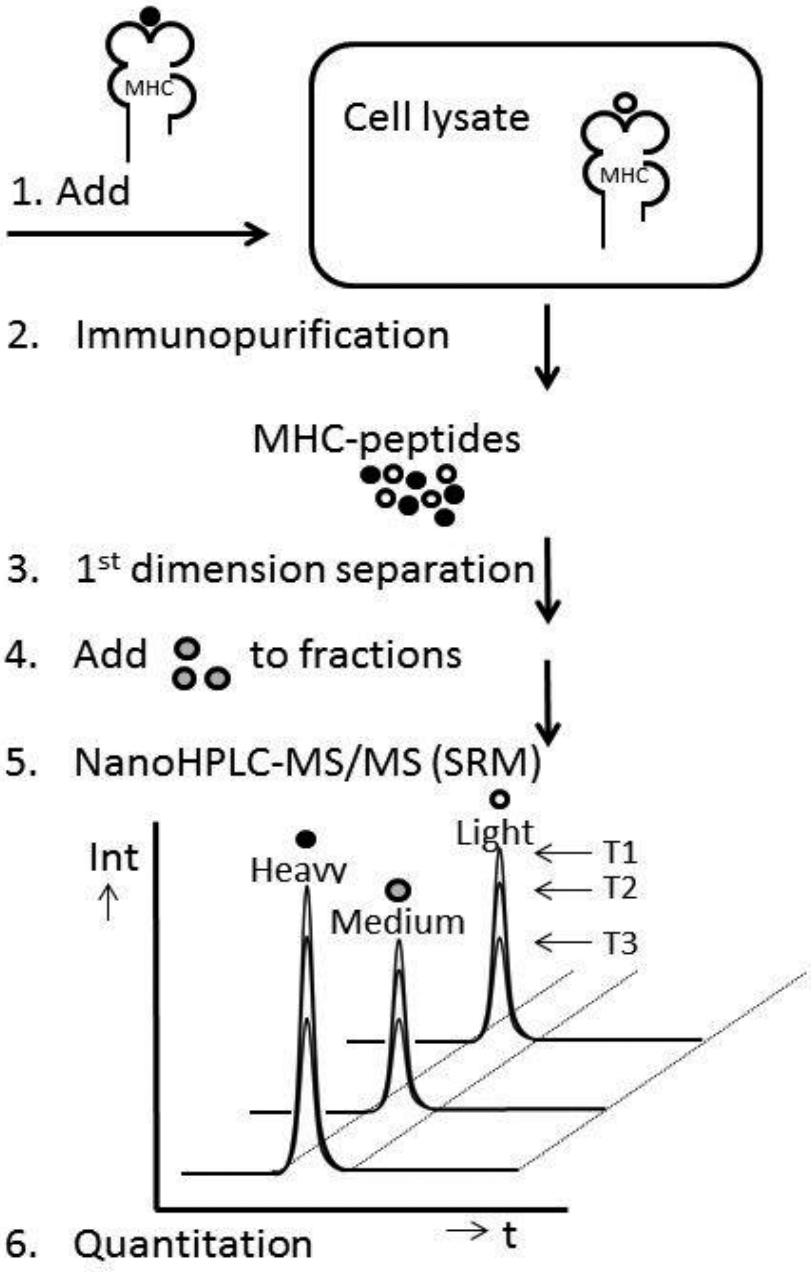


Figure 1. Strategy for the accurate quantitation of MHC-class I-presented peptides. An MHC-monomer loaded with heavy labeled peptide (hpMHC) is added to the cell lysate (step 1) in which the endogenous (wt; unlabeled) peptide is present. Subsequently, the hpMHC and the wt-pMHC undergo the same further sample pretreatment. Just before LC-MS analysis a medium labeled variant of the peptide is added (step 4) for determination of the yield of the procedure. Black dots: heavy peptide, grey dots: medium peptide and open dots: wt (unlabeled) peptide. The three different forms of the peptide, light, medium and heavy, coelute and these three forms are all quantitated by 3 SRM-transitions (T1-T3).

(or PRM) analysis to determine the overall yield of the sample pretreatment procedure. In the AQUA approach the labeled peptide is preferably added as early as possible during sample pretreatment . In contrast, we have added the medium peptide just before MS analysis to be able to calculate the overall yield of the hpMHC procedure, i.e. the ratio of the medium and heavy labeled peptide.

The quality of the hpMHC was assessed by mixing equimolar amounts of hpMHC and the medium labeled peptide. As can be seen in Supplemental Figure 1, the medium and heavy peptides display the same peak height in the MS spectrum, from which we conclude that the purified hpMHC has been correctly loaded with a single peptide, and the measured hpMHC concentration, as determined by Bradford, is in line with the pure synthetic peptide concentration. Therefore, the intensities of the heavy peptide and the medium peptide can be used for their relative quantitation.

Table 1. Quantitation summary of the native LB-NISCH-1A and LB-SSR1-1S peptides for the 3 biological replicates. Total peptide amounts, recovery and copy numbers are listed. For the AQUA measurements no copy number is listed, since the yield is not known. For the AQUA experiments the amount of heavy peptide and recovery is listed.(n.a.:not applicable)

hpMHC-approach			
LB-NISCH-1A (ALAPAPAEV)	Native peptide (pmol)	Recovery	Copy number
Sample 1 (6.7E9 cells)	8,8 ±0,1%	1.50%	789 ±7
Sample 2 (2E9 cells)	17.6±7.1%	0.90%	5273 ±374
Sample 3 (2E9 cells)	10.0±3.6%	1.10%	3003 ±108
LB-SSR1-1S (VLFRGGPRGSLAVA)	Native peptide (pmol)	Recovery	Copy number*
Sample 1 (6.7E9 cells)	33 ±0,4%	2.80%	2943 ±36
Sample 2 (2E9 cells)	33.2±6.7%	0.80%	9960 ±667
Sample 3 (2E9 cells)	40.1±3.2%	0.50%	12030±385
AQUA-approach			
LB-NISCH-1A (ALAPAPAEV)	Heavy peptide (pmol)	Recovery	Copy number*
Sample 1 (6.7E9 cells)	n.a.	n.a.	n.a.
Sample 2 (2E9 cells)	1.5±7.4%	29.40%	n.a
Sample 3 (2E9 cells)	1.3±3.9%	25.80%	n.a
LB-SSR1-1S (VLFRGGPRGSLAVA)	Heavy peptide (pmol)	Recovery	Copy number*
Sample 1 (6.7E9 cells)	n.a.	n.a.	n.a.
Sample 2 (2E9 cells)	2.2±11.7%	43.50%	n.a.
Sample 3 (2E9 cells)	1.2±2.8%	23.30%	n.a.

Initial LC-SRM measurements (bioreplicate 1) were performed on a Q-TRAP instrument. Similar SRM transitions were used per peptide for the light, medium and heavy labeled peptides (three transitions per peptide). After collision energy optimization, the most abundant transitions were chosen, as displayed in Supplemental Table 1, and Supplemental Figure 2 for the MS2 spectra. Bioreplicates 2 and 3 were recorded by PRM on a Q-Exactive.

Next we prepared a calibration curve, in which the linearity of the mass spectrometric determination of the medium labeled AQUA peptides A**L**APAPAEV and V**L**FRGGPRGSLAVA were verified with the pure synthetic peptides dissolved in buffer A (data not shown) and verified with the pure synthetic peptides dissolved in matrix . The matrix was composed of a pool of neighboring first dimension C18 HPLC fractions, not containing the peptide of interest. The response was found to be linear between injections of 100 amol to100 fmol absolute on column, see Supplemental Figure 3. Following the completion of the calibration curve, the MiHA LB-NISCH-1A and LB-SSR1-1S were quantitated. To this end, the cell lysate of 6.7 x 109 (6.7E9) EBV-LCL JYpp65 cells was spiked with 10 pmol of each hpMHC for bioreplicate 1. For bioreplicates 2 and 3, 2.0 x 109 (2E9) cells were spiked with 5 pmol of each hpMHC. After immunopurification of HLA class-I molecules, the HLA-class I α-chain and ß2m proteins were separated from the peptides by 10 kDa filtration and the peptides were fractionated using C18 chromatography and freeze dried. Just before mass spectrometric quantitation, fractions were redissolved in solvent A, supplemented with the medium labeled peptides to a final concentration of 2.5 and 3.5 fmol/µl for ALAPAPAEV and VLFRGGPRGSLAVA, respectively.

The results are listed in Table 1 and shown in supplemental Figure 4. The spiked-in medium labeled peptide is present in all fractions. The ratio between the three SRM transitions of each MiHA is constant, so there is no interference from other potentially contaminating background peptides. The heavy labeled peptide and the light (endogenous) peptide, which both underwent the whole sample pretreatment procedure, are coeluting as expected and are only present in two neighboring fractions. The overall yield of the sample pretreatment was calculated from the ratio of the medium labeled peptide and the heavy labeled peptide. As an example for bioreplicate 1, for LB-NISCH-1A the sum of the heavy labeled peptide peaks in fractions 16 and 17 (panel C) should correspond to 330 fmol of heavy peptide (i.e. 1/30th of 10 pmol spike in of hpMHC in lysate, since 1/30th of the peptide pool was injected on column) at 100% peptide recovery from the lysate. However, the actual response corresponds to approximately 4.9 fmol (the sum of the medium labeled peptides in fractions 16 and 17 in panel A). Therefore, the yield of the heavy labeled LB-NISCH-1A peptide, derived from the hpMHC, is determined to be 1.5% (4.9/330). The amount of endogenous LB-NISCH-1A (panel B) can easily be calculated from the ratio of the heavy peptide and light peptide, and was

determined to be 8.8 pmol. The absolute amount of 8.8 pmol on 6.7E9 cells corresponds to a copy number of 789, see Table 1. Similarly for LB-SSR1-1S the yield of the heavy labeled peptide (panel F) was determined to be 2.8%.The amount of endogenous LB-SSR1-1S (panel E) was calculated to be 33 pmol, which corresponds to a copy number of 2943, see Table 1. Similarly the numbers for the other bioreplicates were determined. These numbers represent the cell surface copy number, since the majority of the pMHC is on the cell surface, rather than in an internal compartment, see supplemental Figure 5.

Striking differences in recovery were observed between the hpMHC and AQUA approach. The steps following the IP, comprising filtration, first dimension chromatographic fractionation and the final LC-MS run, have an overall recovery of 23 to 43% (which amounts to 3 steps of approximately 70% on average per step), as determined by the AQUA approach. In sharp contrast with the former the yield of the hpMHC approach was determined to be 0.5-3%. Therefore, rather extreme losses, i.e. 90-99%, occur during the immunopurification step.

Quite some variation in native peptide copy number between biological replicates can be observed in Table 1. It is important to note that the ratio between the medium peptide and the heavy peptide (added as hpMHC) was constant throughout all experiments, as it should since these were taken from the same stock solution for all experiments. The former rules out any issues with the added labeled peptide and hpMHC. If the measured variation can be fully attributed to biological variation is not clear. The stability of the native pMHC complexes during cell storage might play a role.

The hpMHC approach was used to account for all losses during MHC ligand isolation and purification. Our results show that the use of hpMHC allows accurate quantitation of the HLA class I-presented peptides, by accounting for all losses during sample pretreatment. These losses can be very serious, as shown for the MiHA presented in this study. Since our first elution experiments in the mid 1990's we have found these low yields. The fact is that the average yields are indeed low, which we illustrated to be approximately 5% on the basis of total HLA measured versus total peptide intensity in the mass spectrometer [7].Based on the current measurements, with a lower starting number of cells, the average yield seems closer to 1-2%. In addition, a simple calculation in combination with final mass spectrometric responses shows the same. Starting from 1 billion cells (1E9) and assuming a conservative number of 300.000 HLA I molecules on the cell surface of an EBV-LCL, we calculate a total number of 3E14 HLA I molecules. Using Avogadro's number this yields (3E14/6E23 = 0.5E-9 mol), i.e. 500 pmol HLA. Assuming 1,000 different ligands representing the majority of the total HLA-ligandome, the amount per peptide totals 500 fmol. These peptide levels, based on 100% yield, are not observed in our nanoHPLC-experiments. We detect much lower amounts, compared to known peptide standard responses. Furthermore, we have also

performed MHC-elutions, according to Illing et al.[14] of as low as 25 million cells, still identifying over 1100 HLA ligands using the same stringent criteria used in Hassan et al. [7]. These 1100 identified ligands represent a very competitive number, so the losses we report and compensate for, appear to be real and seem a general phenomenon in peptidome analysis. Summarizing the above, we conclude that sample loss issues have never been fully addressed analytically. We are confident that the pMHC monomers are stable, since we can fold stable pMHC monomers and subsequently stable tetramers for T cell staining. All of the above, to our opinion, clearly illustrates the need for an internal standard from the beginning of the sample processing, i.e. the need for a hpMHC. Naturally, for every peptide of interest the corresponding hpMHC must be prepared, similar to LC-SRM drug/metabolite monitoring experiments of small compounds in which specific heavy isotopes are introduced by chemical synthesis. Similarly, this strategy is applicable to MHC class II complexes, although these cannot yet as easily be produced as class I pMHC complexes. The strategy presented here can be used to get a better view of epitope copy number and as such improve vaccine design and immunotherapy.

ACKNOWLEDGEMENTS

W. Benckhuijsen, N. Dolezal and R. Cordfunke and J. de Keijzer are thanked for technical assistance. CH and MGDK were supported by the Landsteiner Foundation for Blood Transfusion Research (LSBR0713). Dr. S. Piersma is thanked for help with Skyline. Dr A. Thompson is thanked for careful reading of the manuscript. The MS data gathered for this study are MIAPE compliant.

The authors have declared no conflict of interest.

REFERENCES

1. Kenter, G.G., Welters, M.J., Valentijn, A.R., Lowik, M.J., Berends-van der Meer, D.M., Vloon, A.P., Essahsah, F., Fathers, L.M., Offringa, R., Drijfhout, J.W. et al. (2009) Vaccination against HPV-16 oncoproteins for vulvar intraepithelial neoplasia. N Engl J Med, 361, 1838-1847.
2. Miller, J.S., Warren, E.H., van den Brink, M.R., Ritz, J., Shlomchik, W.D., Murphy, W.J., Barrett, A.J., Kolb, H.J., Giralt, S., Bishop, M.R. et al. (2010) NCI First International Workshop on The Biology, Prevention, and Treatment of Relapse After Allogeneic Hematopoietic Stem Cell Transplantation: Report from the Committee on the Biology Underlying Recurrence of Malignant Disease following Allogeneic HSCT: Graft-versus-Tumor/Leukemia Reaction. Biol Blood Marrow Transplant, 16, 565-586.
3. Croft, N.P., Smith, S.A., Wong, Y.C., Tan, C.T., Dudek, N.L., Flesch, I.E., Lin, L.C., Tscharke, D.C. and Purcell, A.W. (2013) Kinetics of antigen expression and epitope presentation during virus infection. PLoS Pathog, 9, e1003129.
4. Tan, C.T., Croft, N.P., Dudek, N.L., Williamson, N.A. and Purcell, A.W. (2011) Direct quantitation of MHC-bound peptide epitopes by selected reaction monitoring. Proteomics, 11, 2336-2340.
5. Gerber, S.A., Rush, J., Stemman, O., Kirschner, M.W. and Gygi, S.P. (2003) Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. Proc Natl Acad Sci U S A, 100, 6940-6945.
6. Hogan, K.T., Sutton, J.N., Chu, K.U., Busby, J.A., Shabanowitz, J., Hunt, D.F. and Slingluff, C.L., Jr. (2005) Use of selected reaction monitoring mass spectrometry for the detection of specific MHC class I peptide antigens on A3 supertype family members. Cancer immunology, immunotherapy : CII, 54, 359-371.
7. Hassan, C., Kester, M.G., Ru, A.H., Hombrink, P., Drijfhout, J.W., Nijveen, H., Leunissen, J.A., Heemskerk, M.H., Falkenburg, J.H. and Veelen, P.A. (2013) The human leukocyte antigen-presented ligandome of B lymphocytes. Molecular & Cellular Proteomics.
8. Falkenburg, J.H. and Warren, E.H. (2011) Graft versus leukemia reactivity after allogeneic stem cell transplantation. Biol Blood Marrow Transplant, 17, S33-38.
9. Warren, E.H., Zhang, X.C., Li, S., Fan, W., Storer, B.E., Chien, J.W., Boeckh, M.J., Zhao, L.P., Martin, P.J. and Hansen, J.A. (2012) Effect of MHC and non-MHC donor/recipient genetic disparity on the outcome of allogeneic HCT. Blood, 120, 2796-2806.
10. Meij, P., Jedema, I., van der Hoorn, M.A., Bongaerts, R., Cox, L., Wafelman, A.R., Marijt, E.W., Willemze, R. and Falkenburg, J.H. (2012) Generation and administration of HA-1-specific T-cell lines for the treatment of patients with relapsed leukemia after allogeneic stem cell transplantation: a pilot study. Haematologica, 97, 1205-1208.
11. Warren, E.H., Fujii, N., Akatsuka, Y., Chaney, C.N., Mito, J.K., Loeb, K.R., Gooley, T.A., Brown, M.L., Koo, K.K., Rosinski, K.V. et al. (2010) Therapy of relapsed leukemia after allogeneic hematopoietic cell transplantation with T cells specific for minor histocompatibility antigens. Blood, 115, 3869-3878.
12. Hombrink, P., Hassan, C., Kester, M.G., de Ru, A.H., van Bergen, C.A., Nijveen, H., Drijfhout, J.W., Falkenburg, J.H., Heemskerk, M.H. and van Veelen, P.A. (2013) Discovery of T cell epitopes implementing HLA-peptidomics into a reverse immunology approach. J Immunol, 190, 3869-3877.
13. Van Bergen, C.A., Rutten, C.E., Van Der Meijden, E.D., Van Luxemburg-Heijs, S.A., Lurvink, E.G., Houwing-Duistermaat, J.J., Kester, M.G., Mulder, A., Willemze, R., Falkenburg, J.H. et al. (2010) High-throughput characterization of 10 new minor histocompatibility antigens by whole genome association scanning. Cancer Res, 70, 9073-9083.
14. Illing, P.T., Vivian, J.P., Dudek, N.L., Kostenko, L., Chen, Z., Bharadwaj, M., Miles, J.J., Kjer-Nielsen, L., Gras, S., Williamson, N.A. et al. (2012) Immune self-reactivity triggered by drug-modified HLA-peptide repertoire. Nature, 486, 554-558.

SUPPLEMENTAL

MATERIAL & METHOD

Preparation of hpMHC monomers

Recombinant HLA-A2 heavy chain and human β2m light chain were in-house produced in Escherichia coli. The refolding was performed by adding 1.8 mg of HLA-A*0201 heavy chain dissolved in 8M urea, 1.2 mg of β2m dialyzed to PBS and 2 mg of heavy labeled peptide (ALAPAPAEV or VLFRGGPRGSLAVA) dissolved in DMSO, to 50 ml of cold refolding buffer; (400 mM L-arginine HCl, 100 mM Tris-HCl pH 8; 5 mM reduced glutathione, 0.5 mM oxidized glutatione-Na, 2 mM EDTA, 5% glycerol, Complete protease inhibitors (Roche)), and vigorously mixed after each step. The mixture was incubated for 72 hr. at 10 °C [1]. The refolded protein mixture was concentrated to a volume of 0.5 ml with an Amicon concentrator (membrane cutoff, 30 kDa), then purified by gel filtration using fast protein liquid chromatography on a Superdex 75 column (Amersham Biosciences) and PBS as eluent. Complexes were stored at -80°C. The quality of the monomers was tested by analysis of a known amount of hpMHC complex (as determined in a Bradford assay) relative to pure medium peptide. hpMHC was washed on a microconspin filter with a cutoff of 10 kDa to remove residual unbound peptide and small molecule impurities. An equimolar amount of medium labeled peptide was added to the washed hpMHC. Subsequently 10% acetic acid was added to the sample to dissociate the heavy labeled peptides from the heavy chain and β2m proteins. The medium and heavy labeled peptides were spun through the filter and analysed using nano-LC-LTQ-FT, see supplemental Figure 1.

Determination of cell surface pMHC vs non-cell surface pMHC

To distinguish pMHC on the cell surface and pMHC still in the ER or cis-golgi, cells were lysed and total class I pMHC was isolated by IP with W6/32 mAb [2]. After washing, the samples were split in two. One was digested with Endoglycosidase H (EndoH), while the other was left untreated. Subsequently, the samples were Western blotted and the heavy chain of MHC class I, was probed using HC10 mAb. As can be seen in the left hand side of the supplemental Figure 4, there is hardly any shift in the MHC class I position, while the positive control MHC class II fully shifts as expected (right hand side of the picture). Only immature MHC class I molecules are endo-H sensitive and are still in the ER/golgi, i.e. not on the surface. The fraction of immature MHC class I is very low (<10%), and will hardly change the cell surface copy number. The alpha chain of MHC class II was used as a control in these experiments, with mAb L243 (IP) and mAb 1B5 (blot).

Peptide quantitation

Q-TRAP SRM; Biorepilicate 1 was analyzed in triplicate on an Ultimate 3000 RSCL Nanosystem (Dionex), hyphenated to a QTRAP® 5500 instrument (AB SCIEX, Foster City,CA) operated in positive SRM mode and equipped with a nano-electrospray source with applied voltage of 2.2 kV and a capillary heater temperature of 175 °C. The Nanoflow LC system and QTRAP® 5500 system were both controlled using Analyst 1.5.1 software. The scheduled SRM mode comprised the following parameters: SRM detection window of 600 sec, centered around the retention time (at 34.5 min for ALAPAPAEV and 37,3 min for VLFRGGPRGSLAVA), target scan time of 3.0 s, curtain gas of 15, ion source gas 1 of 15, declustering potential of 100, entrance potential of 10. Q1 and Q3 resolution were set to unit. Pause between mass ranges was set to 1 ms. Collision cell exit potential (CXP) was set to 36 for all transitions. Peak integration was performed using MultiQuant™ software version 2.1 (AB SCIEX, Foster City, CA) software and manually reviewed. Chromatographic separation of peptides was performed by a 68 min gradient at 300 nL/min. Solvent A (0.05% formic acid water) and solvent B (0.05% formic acid, 80% acetonitrile) were mixed at 2% B from 0-3 min, 15% B at 4 min, 36% B at 49 min, 99% B from 50-54 min, 2% B at 55-68 min. The 20 cm x 75 µm ID fused silica nano-LC columns, custom packed with 3 µm 100 Å ReproSil Pur C18 aqua (Dr Maisch GMBH, Ammerbuch-Entringen, Germany), were made in house. After injection, peptides were trapped at 6 µl/min at 2% B, before elution to the analytical column at 300 nL/min. The mass spectrometer operated in SRM mode to detect endogenous (wild type; light), medium and heavy labeled peptides using the optimized parameters (supplemental Table 1). The selected and optimized fragment of the medium labeled peptides were used to select the fragments of endogenous (light) and heavy labeled versions of peptide. All other parameters, CE, CXP, EP, DP and gasses were chosen identical to those for each corresponding transition of the medium labeled peptide.

Q-Exactive PRM; biological samples 2 and 3 were analyzed in triplicate with an on-line C18-nano-HPLC-MS system, consisting of an Easy nLC 1000 gradient HPLC (Thermo, Bremen, Germany), and a Q-Exactive mass spectrometer (Thermo). Fractions were injected onto a homemade precolumn (100 µm × 15 mm; Reprosil-Pur C18-AQ 3 µm, Dr. Maisch, Ammerbuch, Germany) and eluted via a homemade analytical nano-HPLC column (15 cm × 50 µm; Reprosil-Pur C18-AQ 3 um). The gradient was run from 0% to 30% solvent B (10/90/0.1 water/ACN/FA v/v/v) in 45 min. The nano-HPLC column was drawn to a tip of approximately 5 µm and acted as the electrospray needle of the MS source. The Q-Exactive mass spectrometer was recording in parallel reaction monitoring (PRM) mode. One cycle comprised one full scan and nine transitions. Full scan parameters were; resolution 70,000 at an AGC target value of 3,000,000, maximum fill time of 100 ms (full scan), and resolution 35,000 at an AGC target value of 1,000,000/maximum fill time of 100 ms

Supplemental Table 1. Parameters used to analyse the endogenous, wild type (light), medium and heavy labeled peptides by SRM. z= ion charge, CE=collision voltage.

Labeling NISCH-1A	Peptide sequence	z(Q1)	Q1 mass	Q3 mass	CE
wt (light; no label)	ALAPAPAEV	2	419.8	415.22 (y4)	12
	ALAPAPAEV	1	838.5	583.31 (y6)	14
	ALAPAPAEV	1	838.5	415.22 (y4)	16
Medium	A**L**APAPAEV	2	423.3	415.22 (y4)	12
	A**L**APAPAEV	1	845.5	583.31 (y6)	14
	A**L**APAPAEV	1	845.5	415.22 (y4)	16
Heavy	A**L**A**P**APAEV	2	426.3	415.22 (y4)	12
	A**L**A**P**APAEV	1	851.5	589.31 (y6)	14
	A**L**A**P**APAEV	1	851.5	415.22 (y4)	16
Labeling SSR1-1S	**Peptide sequence**	**z(Q1)**	**Q1 mass**	**Q3 mass**	**CE**
	VLFRGGPRGSLAVA	3	467.3	655.9 (b13 (2+))	16
wt (light; no label)	VLFRGGPRGSLAVA	3	467.3	606.4 (b12 (2+))	18
	VLFRGGPRGSLAVA	3	467.3	570.8 (b11 (2+))	27
	VLFRGGPRGSLAVA	3	469.6	659.4 (b13 (2+))	16
Medium	V**L**FRGGPRGSLAVA	3	469.6	609.9 (b12 (2+))	18
	V**L**FRGGPRGSLAVA	3	469.6	574.3 (b11 (2+))	27
	V**L**FRGGPRGSLAVA	3	471.6	662.4 (b13 (2+))	16
Heavy	V**L**FRGG**P**RGSLAVA	3	471.6	612.9 (b12 (2+))	18
	V**L**FRGG**P**RGSLAVA	3	471.6	577.3 (b11 (2+))	27
	V**L**FRGG**P**RGSLAVA	3	471.6	577.3 (b11 (2+))	27

for MS/MS. PRM was performed at a normalized collision energy (NCE) of 27. The isolation window was 2 m/z.

Data analysis

Q-TRAP SRM: the LC-SRM data were analysed using MultiQuant™ software version 2.1 (AB SCIEX, Foster City, CA). Peak integration was performed using MultiQuant™ software and manually reviewed. The retention times and ratio of the three transitions were checked for each peptide. The summed peak area of the three transitions was used for the calculation of the amount of endogenous (light) peptide. The recovery of the enrichment method was assessed using the summed peak area of the three transitions of the heavy labeled peptide compared to the peak area of medium labeled peptide. The amount of the MiHA peptides was determined from the ratio of the endogenous (light) peptide and the heavy peptide , for which the average of three measurements of the summed peak areas of three transitions were used. Q-Exactive PRM; for quantitation, raw files were processed in Skyline v. 2.5.0.6079 [3]. In the transition settings dialog box monoisotopic was chosen for precursor and product ion, and resolution was set to 70,000 at m/z 200. The results from Skyline were exported as csv-files and further processed with MS Excel, and visualized with Graphpad.

Sample preparation

The Epstein-Barr virus transformed B lymphoblastic cell line (EBV-LCL) LCL-JYpp65 (typing: HLA-A*0201, B*0702 & Cw*0702) was used as a source of HLA-class I molecules. The MiHA in this study have previously been reported by Hombrink et al. [4] and van Bergen et al. [5]. Pellets of LCL-JYpp65 cells (6.7E9) were lysed in 50 mM Tris-HCl, 150 mM NaCl, 5 mM EDTA, and 0.5% Zwittergent 3-12, (pH 8.0) and supplemented with protease inhibitor (Complete, Sigma Aldrich). The final concentration of the cells in the lysis buffer was 1E8 cells/ml. After 2 hours incubation with tumbling at 4 °C, the preparation was centrifuged for 10 minutes at 2500 rpm and 4 °C. The supernatant was transferred to a new tube and centrifuged for 35 minutes at 11000 rpm and 4 °C. The supernatant of the cell lysate was spiked with 10 pmol monomer (recombinant HLA-A*0201 associated with heavy labeled MiHA (hpMHC) and pre-cleared on a 7 ml CL4-B column and subjected to the 7 ml immunoaffinity column (W6/32 anti HLA class-I antibody (2.5 mg Ab /ml protein-A-Sepharose)) with a flow rate of 2.5 ml/min. For bioreplicates 2 and 3 2 ml CL4-B and IP columns were used. After washing, the bound HLA-class I/peptide complexes were dissociated and eluted from the column with 10% acetic acid. The proteins and the peptides in the eluate were separated by ultrafiltration using a 10 kDa membrane filter (Pall Macrosep centrifuge devices). The filtrate was lyophilized, redissolved in solvent A and the peptides were purified by solid phase extraction (C18 Oasis, 100 µl bed volume, Waters). The purified peptides were fractionated on a homemade RP Reprosil-Pur C18-AQ column (200 µm ID, 3 um x 15 cm) (Dr. Maisch, GmbH, Ammerbuch, Germany). 40 half minute wide fractions were collected and subsequently freeze dried, redissolved in buffer A, containing 2.5 fmol/ul medium labeled peptide (ALAPAPAEV) and 3.5 fmol/ul of the medium labeled peptide VLFRGGPRGSLAVA and analyzed by LC-SRM.

Peptide synthesis

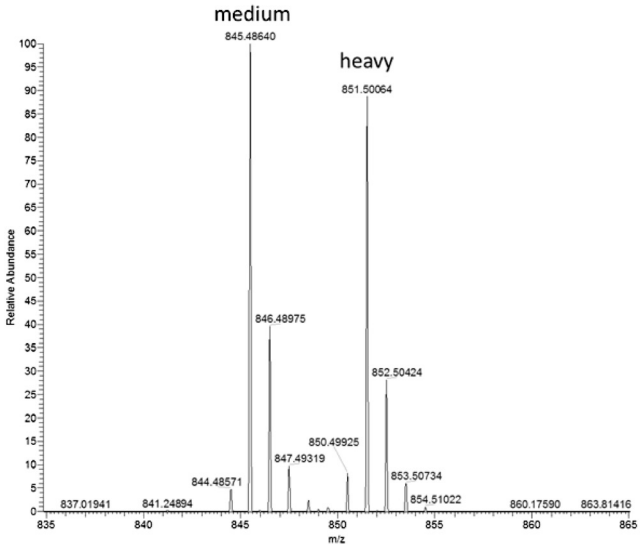
The medium and heavy MiHAs LB-NISCH-1A and LB-SSR1-1S were synthesized by incorporation of stable isotope labeled amino acids. The 13C, 15N-labeled amino acids leucine (L) and proline (P) (Sigma-Aldrich, Netherlands), were used to synthesize the LB-NISCH-1A and LB-SSR1-1S peptides using standard Fmoc chemistry on a SyroII peptide synthesizer (MultiSynTech, Witten, Germany). The peptides were labeled either with only leucine (medium) or leucine and proline (heavy) isotope labeled amino acids, see supplemental Table 1. The integrity of the peptides was confirmed using reverse phase UHPLC and MS.

Amino acid analysis

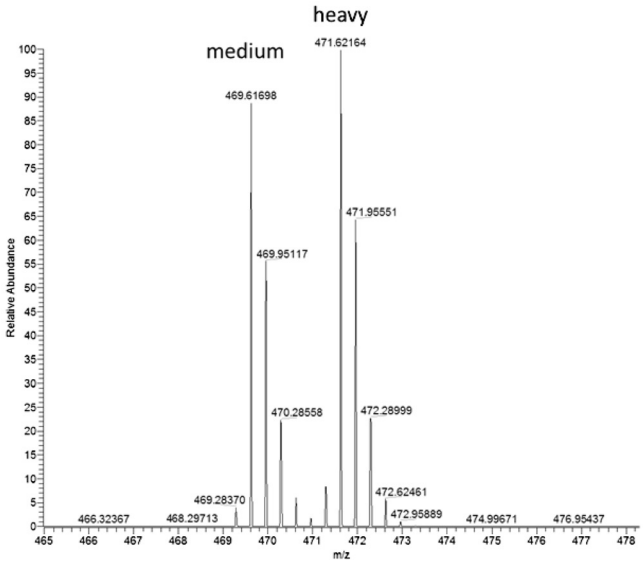
Amino Acid Analysis was performed by Eurosequence (Groningen, The Netherlands), the method of which is partly proprietary. The Aminoquant method was executed as described by Schuster [6]. Because of the

Supplemental Figure 1. The quality of the hpMHC. Heavy peptide dissociated from the hpMHC was mixed with an equimolar amount of synthetic medium labeled peptide. Equal peak heights are observed in the mass spectrometer, which indicates that the hpMHC has the correct stoichiometry. (A) LB-NISCH-1A, (B) LB-SSR1-1S

A



B

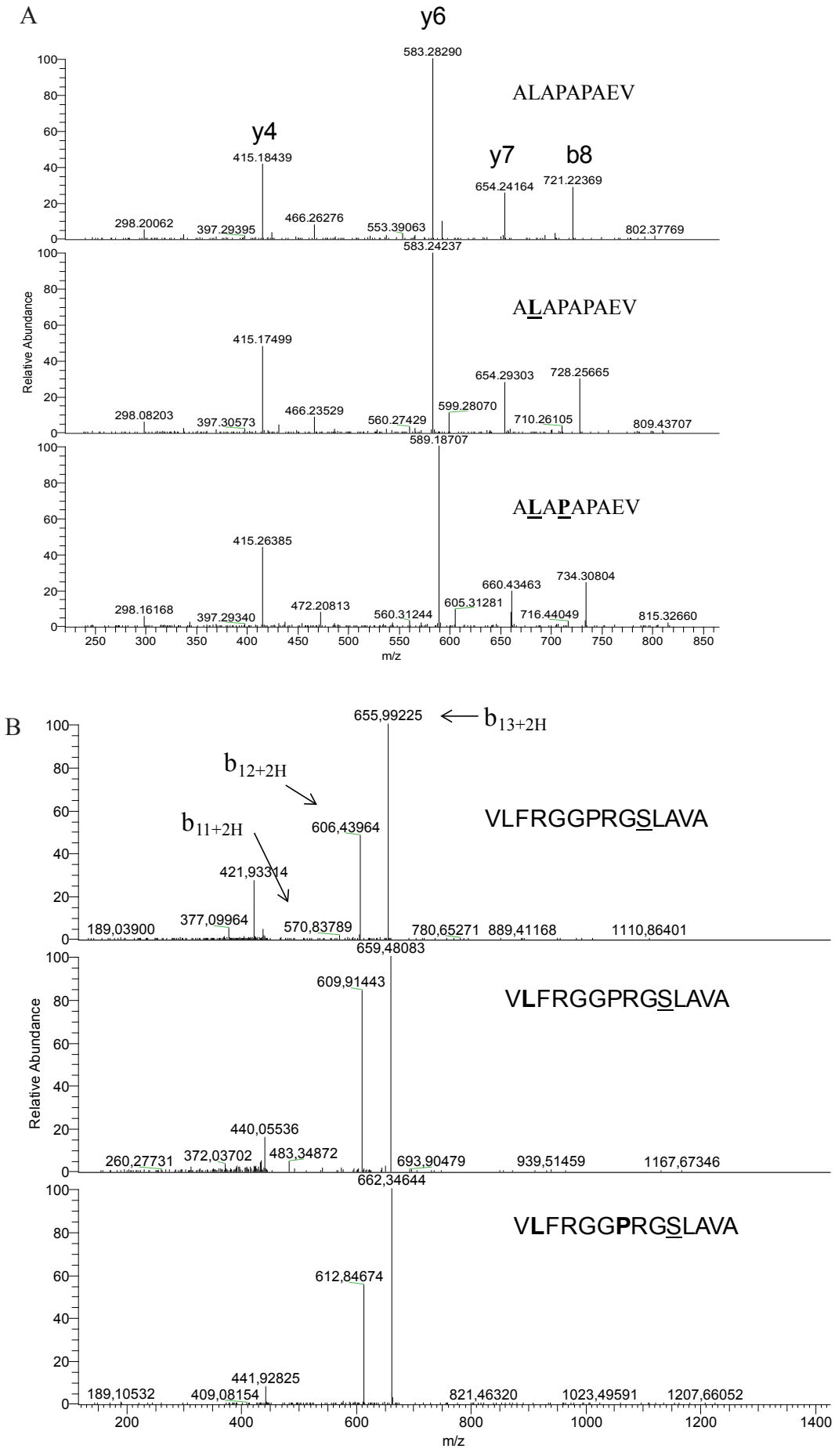


presence of DMSO in the sample, no gas-phase hydrolysis, but liquid phase hydrolysis was used. The sample was hydrolyzed in 5.7 M HCl for 24 hours at 110 °C. The hydrolysate was dissolved in 100 µl of 0.1 N borate buffer containing 100 pmol nVal/µl, and 1 µl was used for the analysis. Columns and reagents as described in literature were used [7].

Spectroscopic analysis of medium peptide concentration

Spectroscopic peptide concentration determination was performed on a Ultrospec 2100 pro UV/VIS spectrophotometer (Amersham/Biosciences) using absorption at either 280 nm fortyrosine (Y) (ε= 1400 mol-1cm-1), or 355 nm for the o-phthaldialdehyde (OPA)-derivatives. OPA is a particularly favorable reagent because the fluorescence and long wavelength absorption only develops due to the reaction with the peptide amino terminus, so the reagent itself causes no background, and the reaction is fast (2 to 3 minutes at room temperature). Lysine residues are also labeled, but these are not present in our peptide. Since fluorescence is mostly quenched in peptides, as we experienced and has been reported by others [6], and references therein, we used the absorbance characteristics of the OPA derivatives. First the peptide concentration of a Y-containing peptide was determined at 280 nm (with ε(Y) = 1400 mol-1cm-1), and subsequently the concentration was used to determine the ε (OPA) at 355 nm. The ε(OPA) at 355 nm was determined to be 2040 mol-1cm-1. This value for the OPA-derivatives was checked by determination of the concentrations of 4 other peptides at 280 nm and at 355 nm after reaction with OPA and the number of 2040 mol-1cm-1 yielded the same results. So use of the absorbance characteristic of the OPA reagent is just as valuable as

Supplemental Figure 2. Tandem mass spectra of the light, medium and heavy form of the MiHAs LB-NISCH-1A (top) and LB-SSR1-1S (bottom). Bold and underlined letters refer to the heavy labeled amino acids. Notice the shifts in the tandem mass spectra of the differently labeled species.



measurement of Y at 280 nm. This method was subsequently applied to the MiHA LB-NISCH-1A and LB-SSR1-1S. The concentration was determined to be 0.50 mg/ml, which is in agreement with the AAA value of 0.48 mg/ml. The results are in the supplemental Table 2.

Formally from the AQUA approach the copy numbers cannot be calculated since the yield of the immuno-purification step is not known. Note however that the measured peak areas for the native peptides are very similar for the hpMHC and AQUA approaches for bioreplicates 2 and 3.

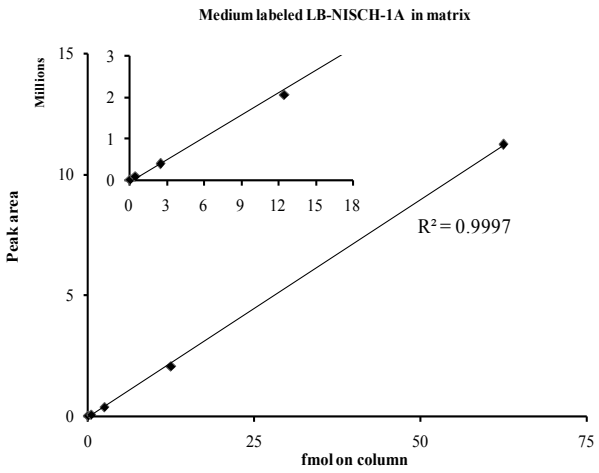
Supplemental Table 2. Concentration of medium labeled peptides, numbers in the table are in mg/ml. *Average of 3 independent measurements.

Sequence	Y280*	OPA*	AAA
SLAAYIPRL	1.5	1.5	-
GVPGADIFYEANPR	0.96	1.12	-
DRVYIHPF	0.77	0.72	-
FLPSDYFPSV	1.04	0.92	-
YVLDHLIVV	0.95	1.15	-
ALAPAPAEV	n.a.	0.5	0.48
VLFRGGPRGSLAVA	n.a.	0.69	-

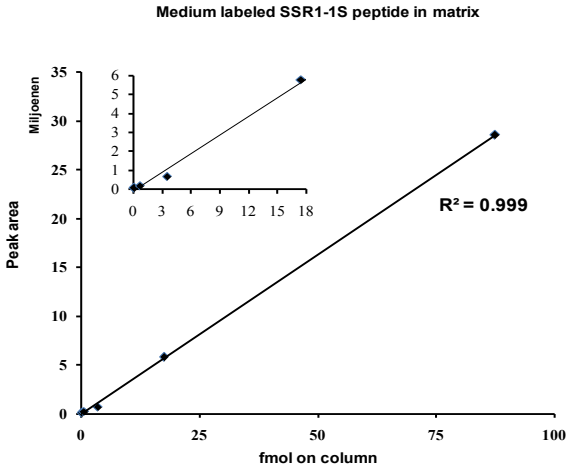
Supplemental Figure 3. (A) calibration curve of medium labeled peptide ALAPAPAEV in matrix, (B) calibration curve of medium labeled peptide VLFRGGPRGSLAVA in matrix.

Bioreplicates 1

A

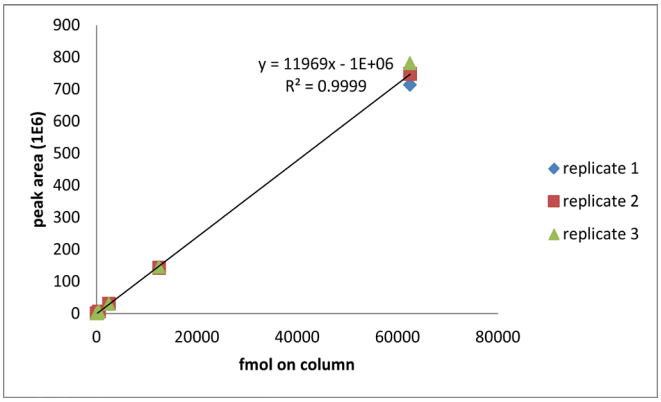


B



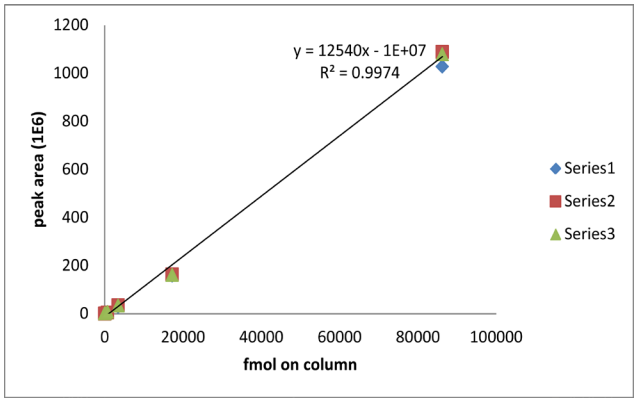
Bioreplicate 2&3; LB-NISCH-1A

A



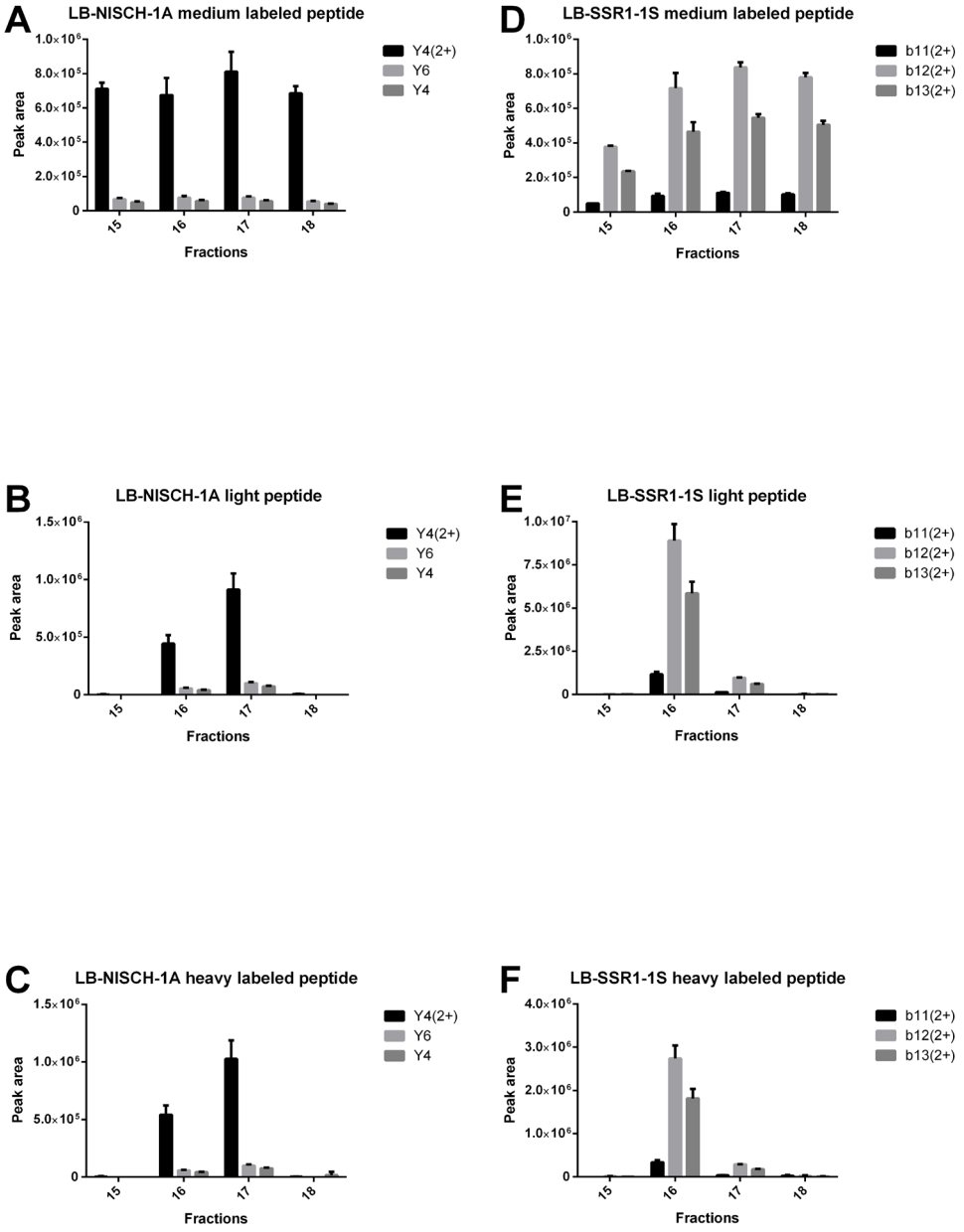
Bioreplicate 2&3; LB-SSR1-1S

B

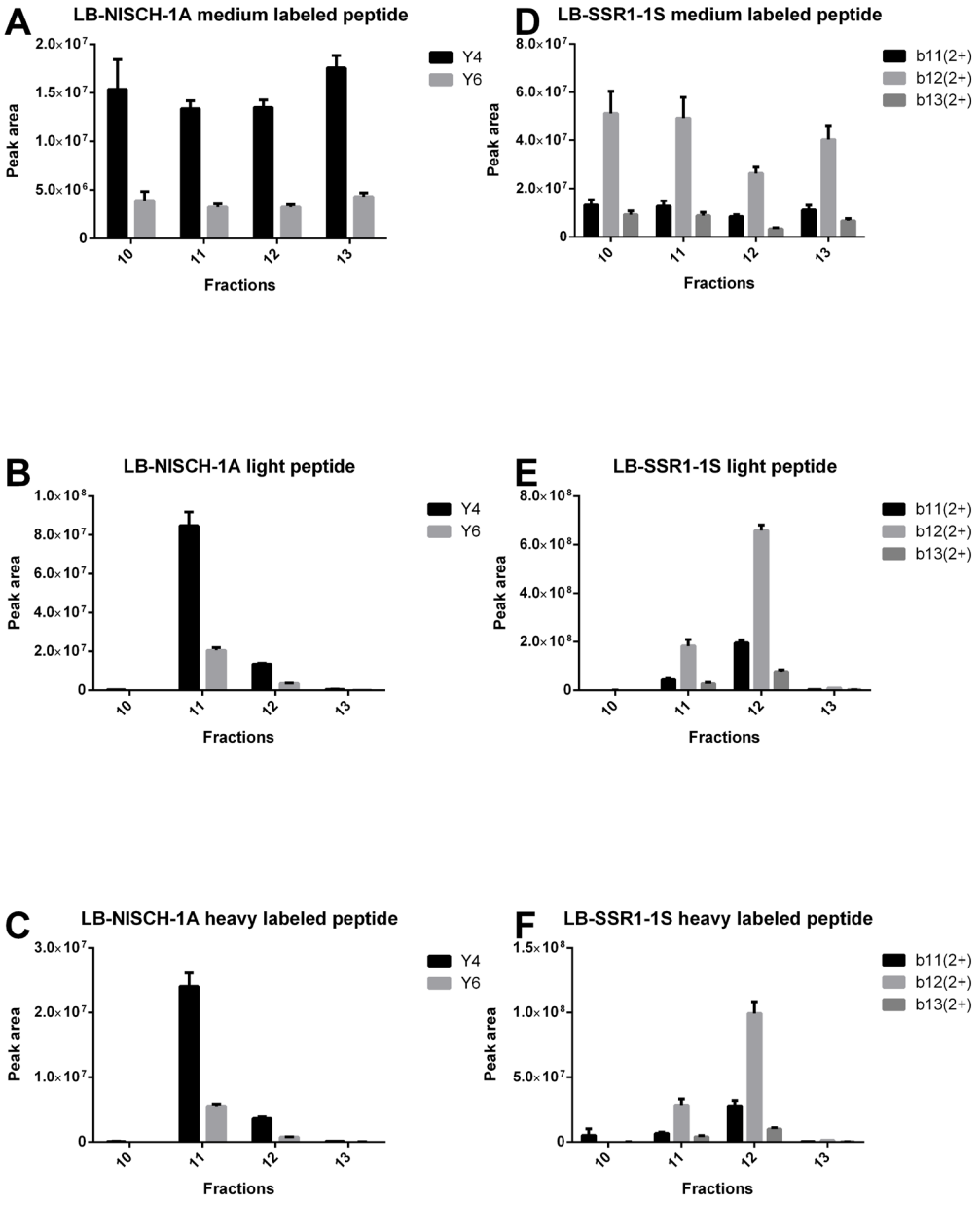


Supplemental Figure 4 (see next 5 pages). (A) Average peak area of the (A&D) medium labeled, (B&E) endogenous (light) peptide and (C&F) heavy labeled peptide in the chromatographic fractions. Average of three measurements. The annotations in the figures indicate the transitions from the doubly charged precursor to fragment y4 (white), singly charged precursor to fragment y6 (gray) and singly charged precursor to y4 (black) for LB-NISCH-1A peptides, and the transitions from triply charged precursor to doubly charged fragments b13 (white), b12 (gray) and b11 (black) for LB-SSR1-1S peptide.

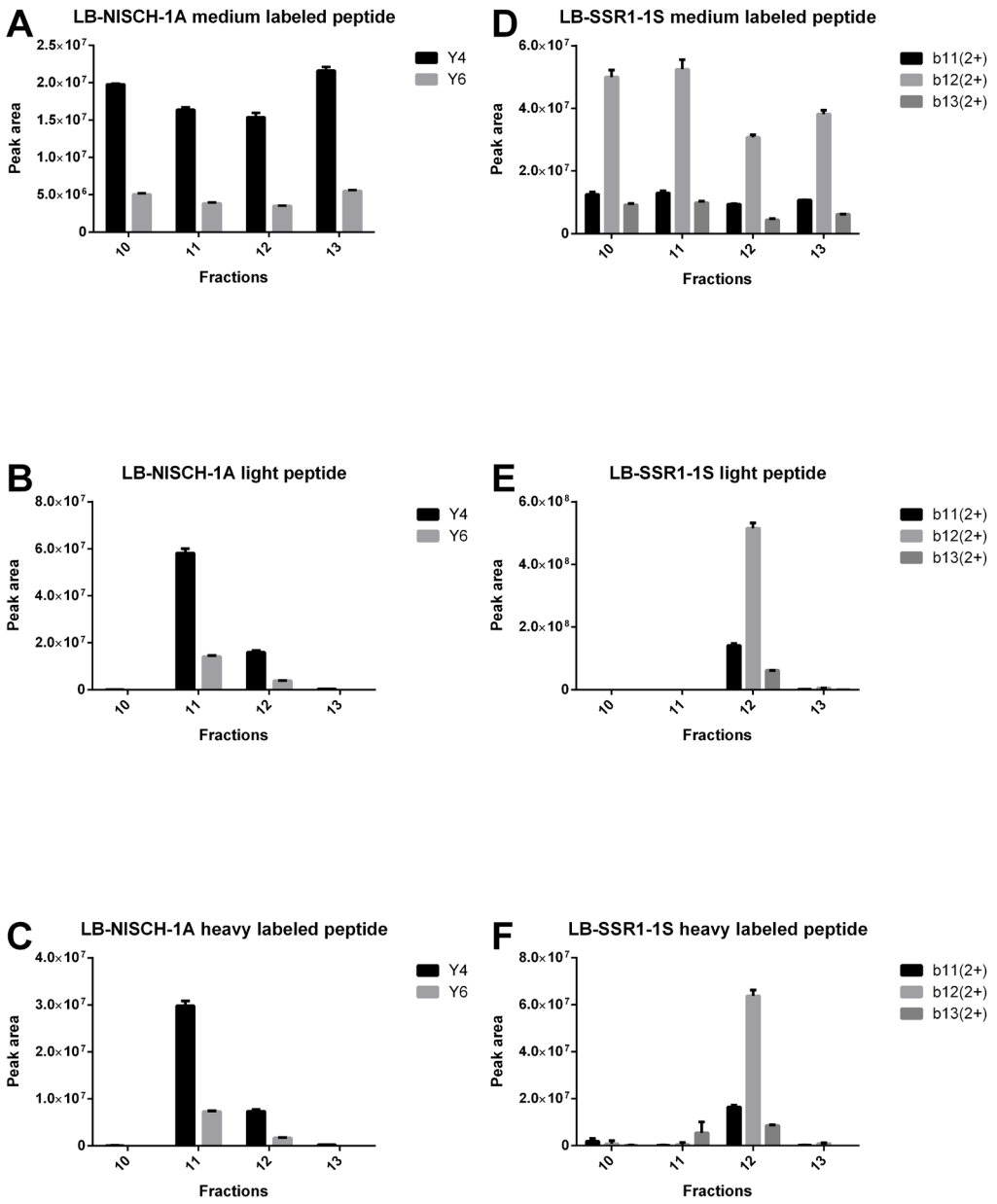
Bioreplicate 1 (hpMHC approach, SRM Q-TRAP)



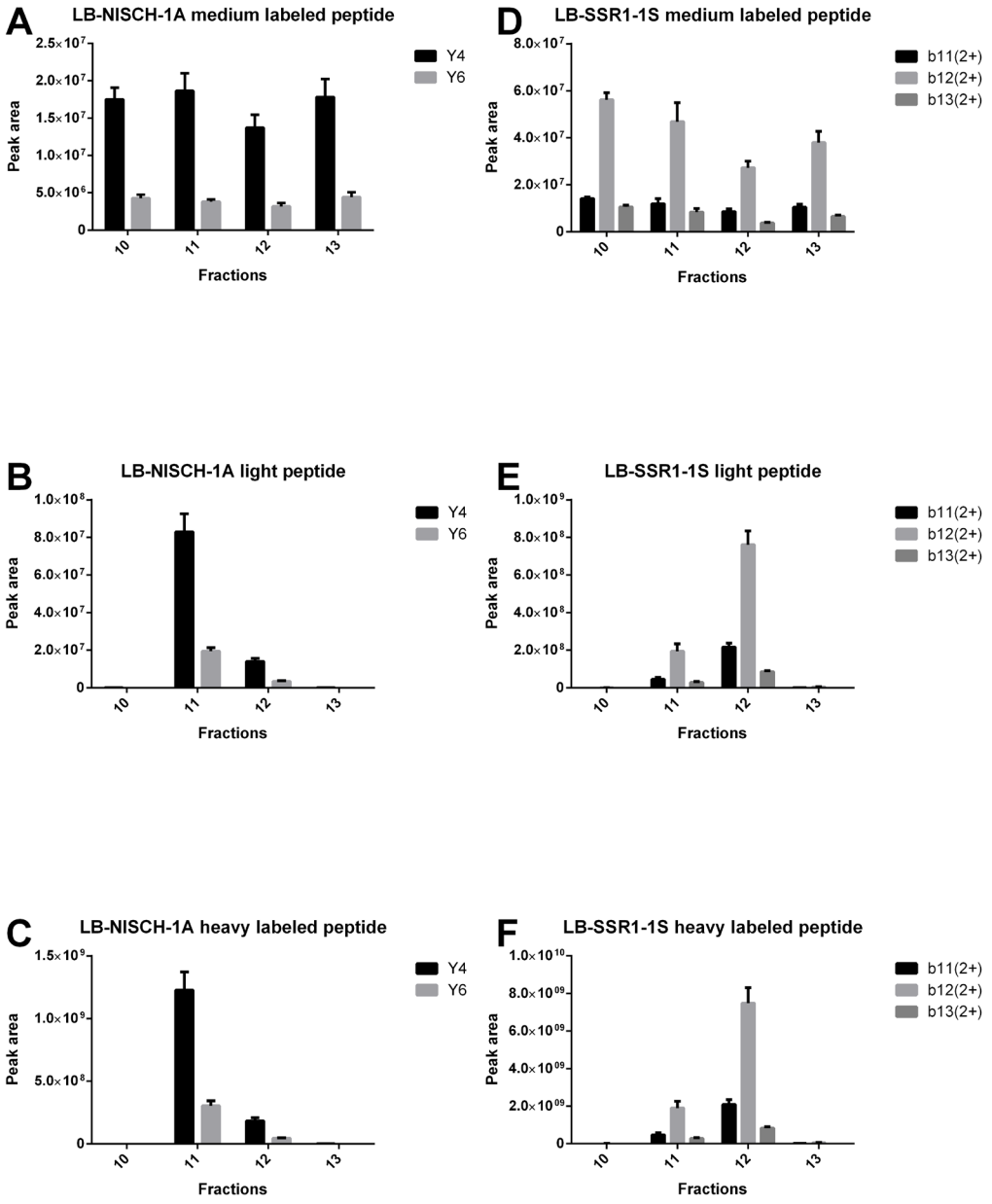
Bioreplicate 2 (hpMHC approach, PRM Q-Exactive)



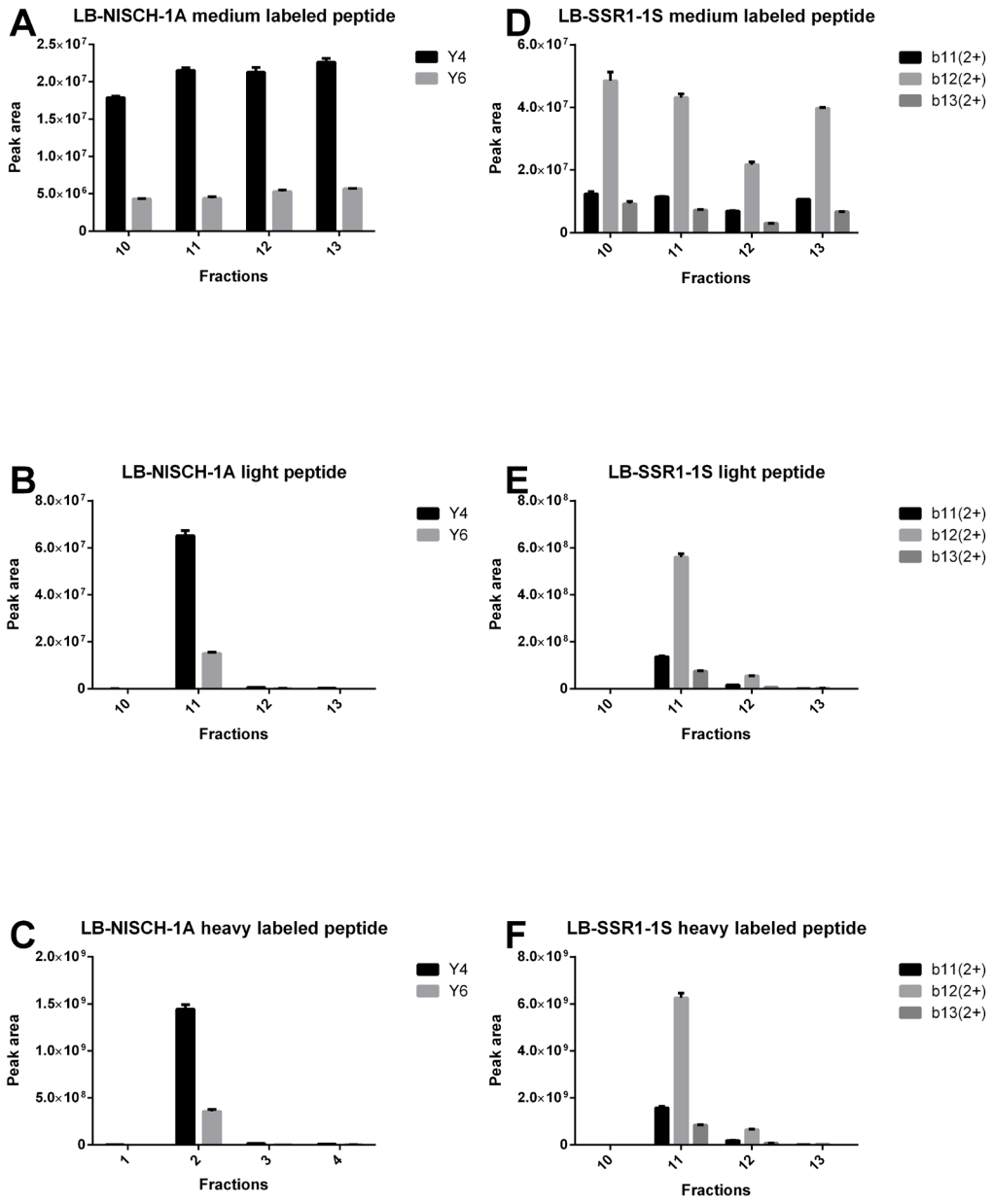
Bioreplicate 3 (hpMHC approach, PRM Q-Exactive)



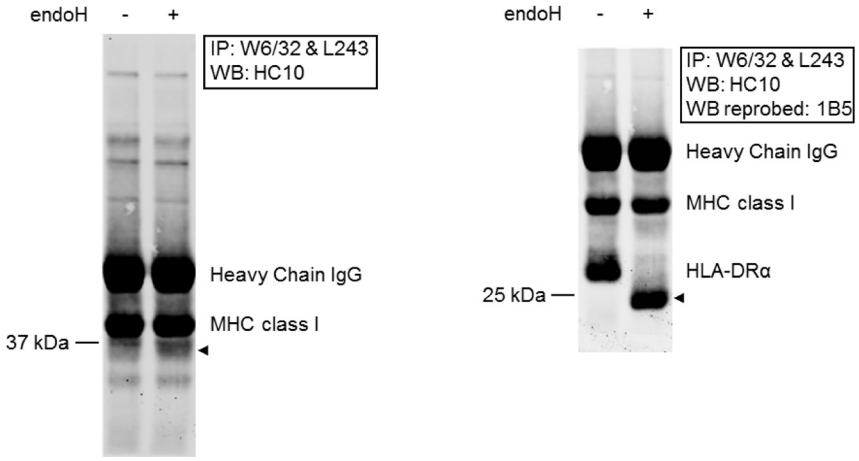
Bioreplicate 2 (AQUA approach, PRM Q-Exactive)



Bioreplicate 3 (AQUA approach, PRM Q-Exactive)



Supplemental Figure 5: Distribution of pMHC in EBV-JY LCL. As can be seen in the left panel the fraction of endoH-sensitive HLA class-I is very small (<10%), marked with the arrow, which indicates that the vast majority of pMHC in EBV-JY LCL is on the cell surface (i.e. endoH-insensitive).

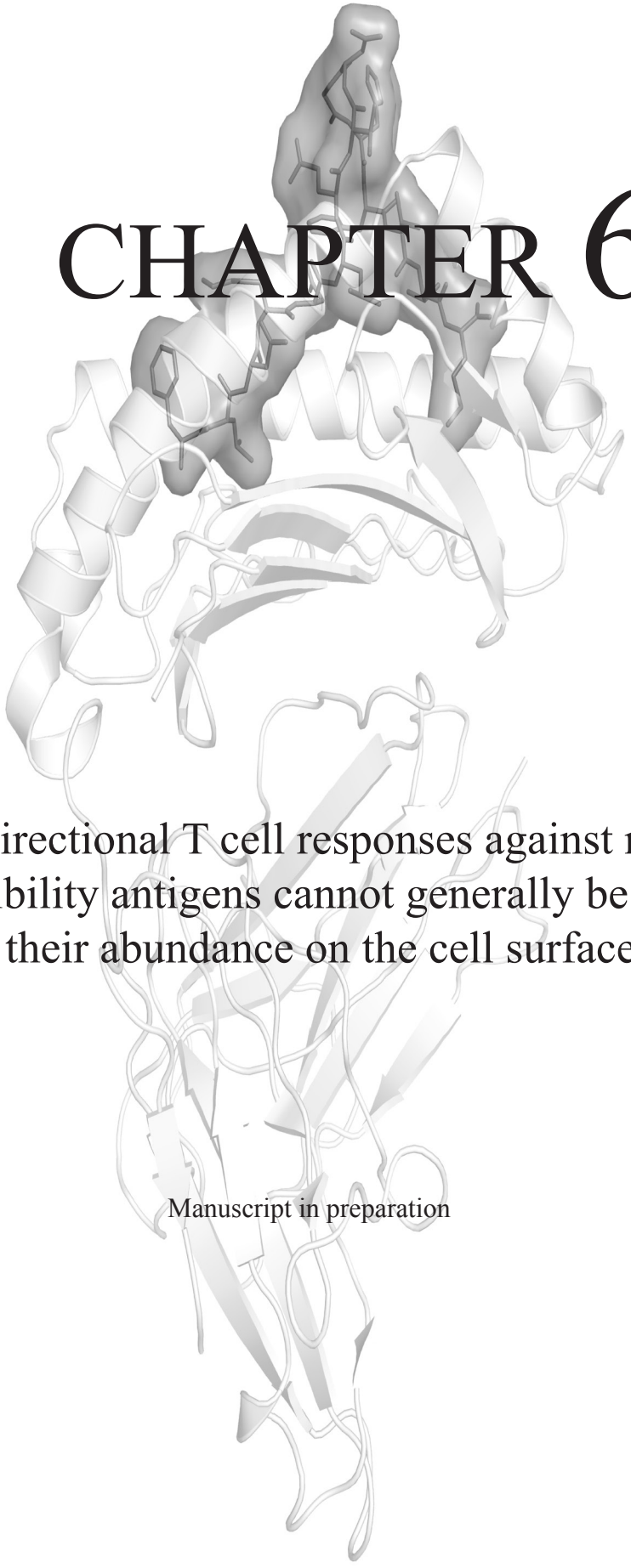


REFERENCES

1. Altman JD, Moss PA, Goulder PJ et al.Phenotypic analysis of antigen-specific T lymphocytes. Science 1996; 274:94-6.
2. Bennett EM, Bennink JR, Yewdell JW, Brodsky FM. Cutting edge: adenovirus E19 has two mechanisms for affecting class I MHC expression. J Immunol 1999; 162:5049-52.
3. Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. MacLean B, Tomazela DM, Shulman N, Chambers M, Finney GL, Frewen B, Kern R, Tabb DL, Liebler DC, MacCoss MJ. Bioinformatics. 2010 Apr 1;26(7):966-8
4. Hombrink P, Hassan C, Kester MG et al. Discovery of T cell epitopes implementing HLA-peptidomics into a reverse immunology approach. J Immunol 2013; 190:3869-77.
5. Van Bergen CA, Rutten CE, Van Der Meijden ED et al.High-throughput characterization of 10 new minor histocompatibility antigens by whole genome association scanning. Cancer Res 2010; 70:9073-83.
6. Schuster R. Determination of amino acids in biological, pharmaceutical, plant and food samples by automated precolumn derivatization and high-performance liquid chromatography. J. Chromatogr. 1988; 431:271-284.
7. Carlson R.G. SK, Givens R.S., Matuszewski B.K. New derivatizing agents for amino acids and peptides. 1. Facile synthesis of N-substituted 1-cyanobenz[f]isoindoles and their spectroscopic properties. J.Org.Chem 1986; 51:3978-3983.

NOTE:

Finally, it is important to note that there are two issues that influence the accuracy of the determined pMHC cell surface copy number. First, it is generally assumed that the large majority of pMHC is present on the cell surface. Therefore, we determined the fraction of cell surface pMHC and non-cell surface pMHC, according to Bennett et al. [2]. From supplemental Figure 5 it can be seen that the large majority of the pMHC in our EBV-JY cells is present on the cell surface. Second, the accuracy of the described approach depends on the accuracy of the determination of the peptide and protein standards used in the study, i.e. on the amino acid analysis and the Bradford assay used. The intrinsic inaccuracy of these methods is generally estimated to be 10-20%, which obviously impacts the overall accuracy of the cell surface peptide copy number.



CHAPTER 6

Unidirectional T cell responses against minor histocompatibility antigens cannot generally be explained by their abundance on the cell surface

Manuscript in preparation

Based on:

Chopie Hassan
Gideon Oudgenoeg
Michel G.D. Kester
Pleun Hombrink
Arnoud H. de Ru
Jan Wouter Drijfhout
Hugo D. Meiring
Ad P. de Jong
J.H. Frederik Falkenburg
Connie R. Jiménez
Mirjam H. M. Heemskerk
Peter A. van Veelen

6

Unidirectional T cell responses against minor histocompatibility antigens cannot generally be explained by their abundance on the cell surface

ABSTRACT

Minor histocompatibility antigens (MiHA) are peptides encoded by polymorphic genes that are presented by Human leukocyte antigen (HLA) molecules and are recognized by T cells in recipients of allogeneic hematopoietic cell transplants. The allelic counterpart (AC) of the MiHA is derived from the same gene and is generally considered to be non-immunogenic. The polymorphic peptide response is thought to be unidirectional, and the lack of immunogenicity of the AC has generally been attributed to the lack of peptide presentation at the cell surface. Here we report on the occurrence of both polymorphic forms on the cell surface in the context of the same presenting HLA-molecule. This study, based on quantitative mass spectrometric determination of the specific MiHA/AC pairs, describes three cases in which the abundance of the AC is similar to the MiHA. Therefore, the unidirectionality of MiHA specific immune responses can in these cases not be explained by absence of the AC, from which we conclude that an underlying immunological selection mechanism must be operative.

INTRODUCTION

Minor histocompatibility antigens (MiHA) are HLA-presented peptides, derived from polymorphic proteins that differ between individuals due to genomic single nucleotide polymorphisms (SNP). T cells directed against these MiHAs can induce beneficial graft versus leukemia reactivity after allogeneic hematopoietic stem cell transplantation (allo-SCT) and subsequent donor lymphocyte infusion (DLI) [1-4]. The non-immunogenic polymorphic peptide derived from the same gene is called the allelic counterpart (AC). AC differ from MiHA by at least one amino acid in the peptide sequence. The lack of immunogenicity of the AC has generally been attributed to the lack of peptide presentation on the cell surface, due to changes in intracellular processing [5, 6], low HLA binding affinity [7, 8], gene deletion [9], frame shifts due to nucleotide insertion [10], alternative splicing [11, 12], or from the introduction of the translation of termination codon [13]. However, bidirectional MiHA immune responses have been described, indicating that the AC of certain MiHAs can be presented (BCL2-A1; [14] & HB-1; [15]). The above examples are summarized in Table 1, as immunological situation A and B. In situation A the lack of an immunological response towards the AC could be explained by absence of the AC on the cell surface. In situation B an immunological response against the AC is found, indicating that the AC is at the cell surface.

The basis of this study is to investigate whether in those cases, where no immune response against AC has been documented (i.e. situation C), the MiHA as well as the AC are present at the cell surface.

To investigate whether differences between the immunogenicity of the MiHA and the AC are due to differences in the cell surface expression level of the two peptide variants, we quantified on the peptide level the cell surface expression of the MiHA and the AC. To quantify the absolute copy number of the MiHAs

Table 1: Overview of the three immunological situations in which a MiHA is present, in combination with the absence or presence of a T cell response to the allelic counterpart (AC), and the absence or presence of the AC.

Imm. Situation	MiHA	AC	AC Present?
A	+	-	-
B	+	+	+
C	+	-	+

and their ACs presented in the context of HLA on the cell surface, we employed the newly developed heavy peptide-MHC (hp-MHC) method using stable isotope labeling in combination with single reaction monitoring (SRM) [16]. SRM is a targeted method with high sensitivity for the detection of peptides isolated from biological samples, and enables the accurate quantification of peptides in a complex biological background. SRM has been widely used in the field of proteomics for quantitation of in vitro digested proteins (peptides) [17-19]. In 2005 Hogan and coworkers [20] used SRM to detect and quantify peptides associated with the major histocompatibility complex (pMHC), recently SRM was used for the quantitation of the SIINFEKL epitope presented by ovalbumin-transfected cells [21] and presentation of eight vaccinia virus peptide-MHC complexes (epitopes) on infected cells [22]. In this study we describe the quantitation of three MiHA and their ACs, which have previously been identified in our lab [23, 24] and chapter 5 of this thesis, on the cell surface of heterozygous cells. We show, for three individual MiHA/AC pairs, similar abundance of MiHA and AC. The results clearly show that the lack of immunogenicity of the AC cannot be explained by differences in the relative abundance on the cell surface.

MATERIALS AND METHODS

Sample preparation

Epstein-Barr virus transformed B lymphoblastic cell lines (B-LCL) LCL-HHC (typing: HLA-A*0201, B*0702, B*4402, Cw*0501 & Cw*0702) and LCL-JY (typing: HLA-A*0201, B*0702 & Cw*0702) were used as source of HLA-A*0201 molecules. These two cell lines were selected according to the genotype of the SNPs encoding for the three MiHAs that are the subject of this study. The LCL-HHC is heterozygous for the SNP genotype of CLYBL (citrate lyase subunit beta-like), while the LCL-JY cell line is heterozygous for the SNP genotype of NISCH (Imidazoline receptor 1) and SSR1 (Signal sequence receptor subunit alpha) (Table 1). B-LCLs were expanded in roller bottles, and cultured in IMDM supplemented with 10% heat-inactivated fetal bovine serum (FBS), penicillin/streptomycin and L-glutamine. B-LCLs were collected, washed with ice cold PBS and stored at -80 ^{o}C until use. Hybridoma cell lines were expanded in roller bottles to obtain W6/32 (anti HLA-class I) antibody using protein free hybridoma medium supplemented with penicillin/streptomycin, and L-glutamine. Antibodies produced by the hybridoma cell lines were purified from the supernatant using Prot-A sepharose beads and eluted from the Prot-A beads with Glycine pH 2.5. The eluted antibodies were used to produce immunoaffinity columns (W6/32- Prot-A sepharose 2.5 mg/ml). The W6/32 antibodies were covalently bound to Prot-A sepharose beads using dimethylpimelimidate (DMP). The columns were stored in PBS pH 8.0 and 0.02% NaN3 at 4 ^{o}C.

Preparation of the heavy peptide loaded monomers (hpMHC)

Recombinant HLA-A2 heavy chain and human β2m light chain were in-house produced in Escherichia coli. The refolding was performed by adding 1.8 mg of HLA-A*0201 heavy chain dissolved in 8M urea, 1,2 mg of β2m dialyzed with PBS and 2 mg of heavy labeled peptides dissolved in DMSO, to 50 ml of cold refolding buffer; (400 mM L-arginine HCl, 100 mM Tris-HCl pH 8; 5 mM reduced glutathione, 0.5 mM oxidized glutatione-Na, 2 mM EDTA, 5% glycerol, complete protease inhibitors (Roche)), vigorously mixed after each step, the mixture was incubated for 72 hr. at 10 °C [25]. The refolded mixture was concentrated to a volume of less than 0.5 ml with an Amicon concentrator (membrane cutoff, 30 kDa), then purified by fast protein liquid chromatography on a Superdex 75 column (Amersham Biosciences) with PBS. Complexes were stored at -80°C. The quality of the monomers was confirmed by analysis of a known amount of hpMHC, as determined by a Bradford assay, added to a spin filter with a cutoff of 10 kDa. An equimolar amount of the corresponding medium labeled peptide was added to the sample and subsequently 10% acetic acid was added to the sample to dissociate the hpMHC. The medium and heavy labeled peptides were spun down and analyzed by MS, displaying an equal intensity, see Supplemental Figure 1 and [16].

Peptide synthesis

Isotope labeled peptides were synthesized by incorporation of the stable isotope labeled (13C, 15N) amino acids leucine (L) and proline (P) (Sigma-Aldrich, The Netherlands) using standard Fmoc (N-(9-fluorenyl) methoxycarbonyl) chemistry with a SyroII peptide synthesizer (MultiSynTech, Witten, Germany). Peptides were labeled either with isotope labeled leucine (medium peptides) or with isotope labeled leucine and isotope labeled proline (heavy peptides), see Table 2. The integrity of the peptides was checked using reverse phase HPLC and MS.

Isolation of HLA class I-presented peptides

The peptides associated with HLA-class I molecules were extracted from the pellets of 7 x 109 LCL-JY cells and 7 x 109 LCL-HHC cells. The pellets were lysed in 50 mM Tris-HCl, 150 mM NaCl, 5 mM EDTA, and 0.1% zwittergent 3-12, (pH 8.0) and supplemented with complete protease inhibitors (Roche). The total concentration of cells in the lysis buffer was 1 x 108 cells/ml. 10 pmol of HLA-class I monomer loaded with a heavy isotope labeled peptide (hpMHC) [16] of each polymorphic variant was added to the lysate to account for all sample loss during the chromatographic process used in extraction of the peptides from the cell lysate [26]. After 2 hours incubation with tumbling of the cells in the lysis buffer at 4 °C, the preparation was centrifuged for 10 minutes at 2500 rpm and 4°C. The supernatant was transferred to a new tube and centrifuged for 35 minutes at 11000 rpm and 4 °C. The supernatant was pre-cleared with CL4B beads and subjected to the im-

munoaffinity column with a flow rate of 2 ml/min. After washing, bound pMHC complexes were eluted from the column, and dissociated, with 10% acetic acid. Peptides were separated from the HLA-class I molecules by passage through a 10 kDa membrane (Pall macrosep centrifuge devices). The filtrate was freeze dried, redissolved and the peptides were further purified by solid phase extraction (C18 Oasis,100 µl bed volume, Waters), freeze dried and resuspended in 95/3/0.1 water/ACN/FA, v/v/v.

Peptide separation

To reduce the complexity of the eluted peptide pool, we fractionated the peptide pool using an home-made RP Reprosil-Pur C18-AQ column (200 µm ID, 3 um x 15 cm, Dr. Maisch, GmbH, Ammerbuch, Germany). The sample was loaded in solvent A (100/0.1 water/FA v/v), and the gradient was run from 0-50% B (10/90/0.1 water/ACN/FA v/v/v) in 30 min at a flow rate of 4 ul/min. The samples were taken up in a make-up flow of 50/50/0.1 water/ACN/FA at 100 ul/min supplied via a T-piece through the annular space between the separation capillary and an auxiliary capillary. In this way 45 half-a-minute wide fractions were collected, subsequently freeze dried and stored at -80 ^{o}C pending LC-SRM analyses. An IFN-γ release test was performed of all fractions to detect the fractions in which the MiHA is present using MiHA specific T cell clones. In addition to the MiHA positive fractions, neighboring fractions were also selected to be analyzed with LC-SRM for the eluted (light) MiHA peptides. Since the fractions containing the AC could not be detected with the T cell clones, fractions were selected on the basis of their known elution behavior relative to the MiHA. The selected fractions were analyzed for quantification of the eluted (light) AC peptides.

LC-SRM-analyses

The selected HPLC fractions were lyophilized and resuspended in 30 ul mix of medium labeled peptides. The concentrations of the peptides were 2.5 fmol/ul (ALAPAPAEV; LB-NISCH-1A), 2.5 fmol/ul (ALAPAPVEV; NISCH-1V), 4.5 fmol/ul (SLAAYIPRL; LB-CLYBL-1Y), 4.5 fmol/ul (SLAADIPRL; CLYBL-1D, 3.5 fmol/ul (VLFRGGPRGSLAVA; LB-SSR1-1S), 3.5 fmol/ul (VLFRGGPRGLLAVA; SSR1-1L). These samples were analysed in triplicate on an Ultimate 3000 RSCL Nanosystem (Dionex) that was hyphenated to an QTRAP® 5500 instrument (AB SCIEX, Foster City,CA) operated in positive SRM mode and equipped with a nano-electrospray source with applied voltage of 2.200 kV and a capillary heater temperature of 175 °C. The Nanoflow LC system and QTRAP® 5500 system were both controlled using Analyst 1.5.1 Software. The scheduled SRM mode comprised the following parameters: SRM detection window of 600 sec, target scan time of 3.0 s, curtain gas of 15, ion source gas 1 of 15, declustering potential of 100, entrance potential of 10. Q1 and Q3 resolution were set to unit. Pause between mass ranges was set to 1 ms. Collision cell exit

potentials (CXP) was set to 36 for all transitions. The retention times for the different peptides were 54.3 min (SLAAYIPRL; LB-CLYBL-1Y), 46.7 min (SLAADIPRL; CLYBL-1D), 34.6 min (ALAPAPAEV; LB-NISCH-1A), 41.8 min (ALAPAPVEV; NISCH-1V), and 37.7 min (VLFRGGPRGSLAVA; LB-SSR1-1S), 47.7 min (VLFRGGPRGLLAVA; SSR1-1L). Peak integration was performed using MultiQuant™ software version 2.1 (AB SCIEX, Foster City, CA) software and manually reviewed. Chromatographic separation of peptides was performed by a 68 min gradient at 300 nL/min. Solvent A (0.05% formic acid water) and solvent B (0.05% formic acid, 80% acetonitrile) were mixed at 2% B from 0-3 min, 15% B at 4 min, 36% B at 49 min, 99% B from 50-54 min, 2% B at 55-68 min. The nano-LC columns were made in house and consisted of 20 cm x 75 µm ID fused silica custom packed with 3 µm 100 Å ReproSil Pur C18 aqua (Dr Maisch GMBH, Ammerbuch-Entringen, Germany). After injection, peptides were trapped at 6 µl/min at 2% buffer B, before elution to the analytical column at 300 nL/min.

SRM assay development

An SRM assay for the target proteins was developed using the MRMPilot™ software version 2.1 from AB SCIEX (Concord, ON, Canada). Assay development included verification of the peptide transitions, and collision energy (CE) optimisation of these transitions. During verification the highest responding transitions and optimum CE-energy were determined, see Supplemental Table 1. During CE-optimisation the transitions selected after verification are optimised during the chromatographic elution of the peptide. All data of CE-optimisation were uploaded to the MRM Pilot and for each peptide three transitions at the experimentally found optimum and the experimentally found retention time were included in the final assay. The corresponding fragment ions were selected for light, medium and heavy versions of peptides, using the same parameters.

Data analysis

The LC-SRM data were analyzed using MultiQuant™ software version 2.1 (AB SCIEX, Foster City, CA) software Peak integration was performed using MultiQuant™ software and manually reviewed. The retention time and the ratio of the three transitions per peptide were checked to exclude interferences from other peptides.

Quantitation using isotope labeled peptides

Peptide abundances were calculated by comparing the peak area of the eluted (light) and heavy peptides with the peak area of corresponding medium labeled peptide spiked into the fractions prior to LC-SRM analysis, as previously described [16].

Next we prepared a calibration curve, in which the linearity of the mass spectrometric determination of the medium labeled AQUA peptides SLAAYIPRL, SLAADIPRL, ALAPAPAEV, ALAPAPVEV, VLFRGGPRG-

Table 2. Overview of the gene name, protein name, and cell lines from which the MiHA and AC were derived from. The polymorphism is underlined and the heavy isotope labeled amino acids are indicated in bold

Gene name	Protein name	Cell lines	Light	Medium	Heavy
LB-CLYBL-1Y*	Citrate lyase subunit beta-like	HHC	SLAAYIPRL	SLAAYIPRL	SLAAYIPRL
CLYBL-1D†			SLAADIPRL	SLAADIPRL	SLAADIPRL
LB-NISCH-1A*	Imidazoline receptor 1	JYpp65	ALAPAPAEV	ALAPAPAEV	ALAPAPAEV
NISCH-1V†			ALAPAPVEV	ALAPAPVEV	ALAPAPVEV
LB-SSR1-1S*	Signal sequence receptor subunit alpha	JYpp65	VLFRGGPRGSLAVA	VLFRGGPRGSLAVA	VLFRGGPRGSLAVA
SSR1-1L†			VLFRGGPRGLLAVA	VLFRGGPRGLLAVA	VLFRGGPRGLLAVA

*MiHA, and † Allelic counterpart (AC)

SLAVA and VLFRGGPRGLLAVA were verified with the pure synthetic peptides dissolved in matrix . The matrix was composed of a pool of neighboring first dimension C18 HPLC fractions, not containing the peptide of interest. The response was found to be linear between injections of 100 amol to 100 fmol absolute on column, see Supplemental Figure 2.

IFNγ release assay

MiHA specific T-cell clones (2×103) were stimulated with 1 x 104 T2 or HLA-A*0201 expressing B-LCLs cells in 384 well plates for 18 hr. at 37°C and 5% CO2. Peptide pulsing was performed by incubating stimulator cells for 1 hr. with the eluted peptides (C18-fractions) or with different concentrations (0.0256 - 250 pM) of synthetic MiHA and AC peptides in IMDM containing 2% FCS. The stimulator cells were washed twice before addition to the T cells. Cytokine release was measured by IFN γ ELISA (Sanquin, CLB, Amsterdam, The Netherlands) according to the manufacturer's instructions.

Competition refolding assay

The competition refolding assay has been developed in our lab by Tan et al. [27]. Briefly, this assay employs unfolded recombinant heavy chain molecules in combination with β2m and the commercially available fluorescent standard peptide (FLPSDCFlFPSV a modified HBV epitope), and relies on protein folding during the assay. The peptide of interest competes with the labeled standard peptide for binding. After 24 hr. of incubation protein complexes and free peptide are separated by HPLC size-exclusion chromatography, during which the fluorescence of protein and peptide fractions is continuously monitored. Following peak integration of the fluorescent signals the ratio of label in the protein and peptide fractions is calculated. The affinities of the peptides are expressed as IC50 , the peptide concentration at which binding of the standard peptide is reduced

to 50%. In this assay we measured the 3 MiHA/AC pairs, CLYBL-1Y/D, NISCH-1A/V and SSR1-1S/L.

RESULTS

Differential recognition of MiHA and AC peptide by the MiHA specific T cell clones

To validate the MiHA specificity of the T cell clones, the three MiHA specific T cell clones were stimulated with titrated concentrations of each MiHA/AC peptide pair, and the T cell response was determined by measuring IFNγ release overnight, see Figure 1. For all three MiHA specific T cell clones a clear differential recognition pattern of the respective MiHA and AC peptide pair was demonstrated. For both the LB-CLYBL-1Y specific T cell-clone K264 and the LB-SSR1-1S specific clone 11-5 even very high concentrations of the AC CLYBL-1D and SSR1-1L, respectively, were not able to elicit a T cell response. For the LB-NISCH-1A specific clone 10H5 we were able to demonstrate a 1500-fold lower sensitivity for the AC peptide compared to the MiHA peptide. In addition, B-LCL cell lines homozygous positive or negative for the MiHA encoding

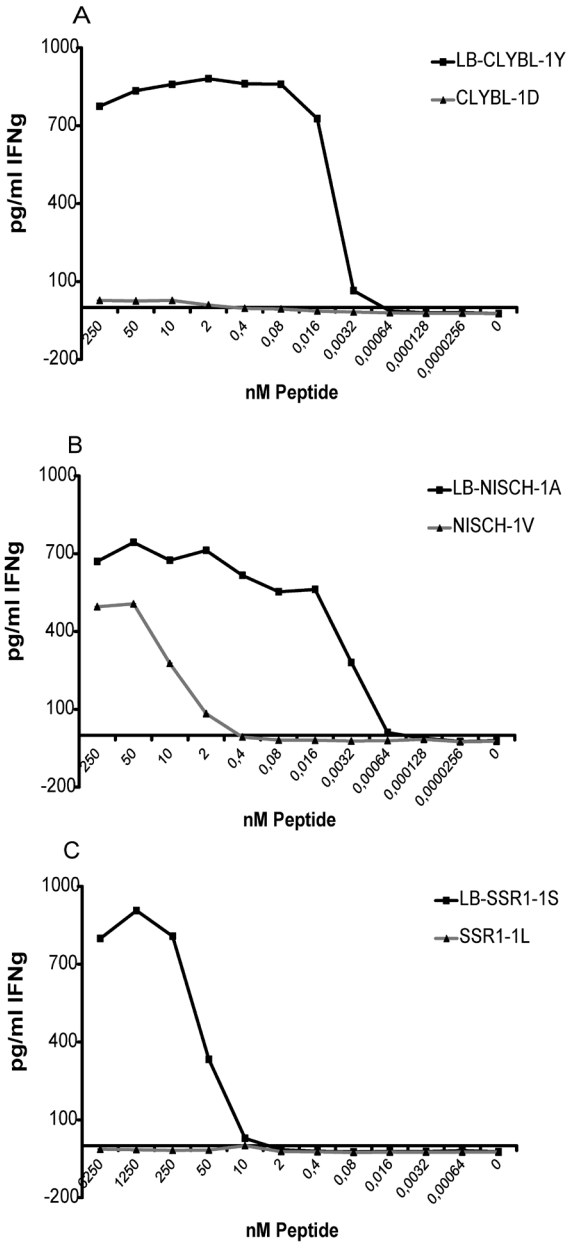


Figure 1. Differential reactivity of MiHA and AC by the MiHA specificT cell clones (A) MiHA specific T cell clone K264, was tested against titrated concentrations of MiHA (LB-CLYBL-1Y) and respective AC (CLYBL-1D) peptide loaded on T2 cells. (B) peptide clone 10H5 was tested against MiHA (LB-NISCH-1A) and respective AC (NISCH-1V) peptide and (C) clone11-5 was tested against MiHA (LB-SSR1-1S) and respective AC (SSR1-1L). IFN-γ was measured after overnight stimulation using standard ELISA.

SNPs confirm the specificity of the T cell clones.

MiHA and AC have similar binding affinities

To exclude that the differential reactivity directed to MiHA and AC were due to differences in binding affinity of the MiHA and respective AC, the binding affinities of the MiHA/AC pairs were determined by a competition assay [27]. The IC50s measured in this assay of the 3 MiHA/ AC pairs were demonstrated to be comparable and are listed in Table 3. The predicted binding affinities were calculated by NetMHC version 3.2. server, and these appeared to similar for each MiHA/AC pair. In addition, the determined affinities and the predicted binding affinities for each MiHA/AC pair were comparable, see Table 3

Table 3. Predicted and tested binding affinity of MiHA and AC peptides.

Gene name	Peptide sequence	Length (AA)	IC50 (nM)	NetMHC prediction (nM)
LB-CLYBL-1Y	SLAAYIPRL*	9	22	8
CLYBL-1D	SLAADIPRL†	9	16	13
LB-NISCH-1A	ALAPAPAEV*	9	52	12
NISCH-1V	ALAPAPVEV†	9	46	12
LB-SSR1-1S	VLFRGGPRGSLAVA*	14	78	250
SSR1-1L	VLFRGGPRGLLAVA†	14	94	221

*MiHA and †Allelic counterpart (AC)

MiHA and AC display similar abundance at the cell surface

To investigate whether the MiHA and AC display differences in abundance at the cell surface that could explain the differential immunogenicity of the MiHA and AC, the pMHC of MiHAs and corresponding ACs were measured by quantitative mass spectrometry. To quantify the amount of MiHA and AC peptides, we incorporated as internal standard the different peptides containing one stable isotope labeled heavy amino acid (medium peptides), as described previously [20-22, 28, 29]. In addition, to correct for sample losses during sample work-up, hpMHC complexes were added, that were loaded with peptides containing two stable isotope

labeled heavy amino acids (heavy peptide) [16]. All three forms of the peptide, i.e. the native form to be determined, and the two quantitation standards, are measured in the same LC/MS run and their abundance ratios determined for accurate quantitation. As an example the results for LB-CLYBL-1Y and CLYBL-1D are shown in Figure 2, in which the intensities are shown of the light eluted peptide, the medium peptide, which was added to all fractions just before MS measurement, and the heavy peptide, derived from the added hpMHC. The results of LB-SSR1-1S, SSR1-1L, LB-NISCH-1A and NISCH-1V, are shown in the supplemental Figures 3 and 4. The medium AQUA peptide is obviously present in all fractions, while the eluted peptide and the heavy peptide are present in only one or two fractions, as they should. Since the amounts of added medium

Table 4. Summary of the quantitation data of MiHA and AC: the amount of eluted (light) peptides, percentage of the recovery of heavy peptides, calculated number of copy per cell and the ratio of the MiHA to allelic counterpart.

Gene name	Peptides sequence	pmol	Copy number	Ratio MiHA:AC	Heavy labeled
LB-CLYBL-1Y	SLAAYIPRL*	18 ± 1	1616 ± 111		11%
				6:1	
CLYBL-1D	SLAADIPRL†	3 ± 0,04	273 ± 4		10%
LB-NISCH-1A	ALAPAPAEV*	8,8 ± 0,08	789 ± 7		1.5%
				1:5	
NISCH-1V	ALAPAPVEV†	40 ± 0,22	3591 ± 20		3%
LB-SSR1-1S	VLFRGGPRGSLAVA*	33 ± 0,4	2943 ± 36		2.8%
				2:1	
SSR1-1L	VLFRGGPRGLLAVA†	13 ± 0,2	1242 ± 18		5%

*MiHA and †Allelic counterpart (AC)

peptide and heavy peptide were known, the amount of eluted peptide and the recovery could be easily calculated. The results are listed in Table 4.

To demonstrate that the similar abundance of the 3 MiHA and their respective AC is not an exception, we mined the ligandome presented in chapter 2 and found several MiHA/AC pairs that are similarly expressed at the cell surface, which are shown in Table 5. Although not quantitated by use of heavy isotope labelled internal standards the results are indicative of more common occurrence of both allelic variants than previously thought.

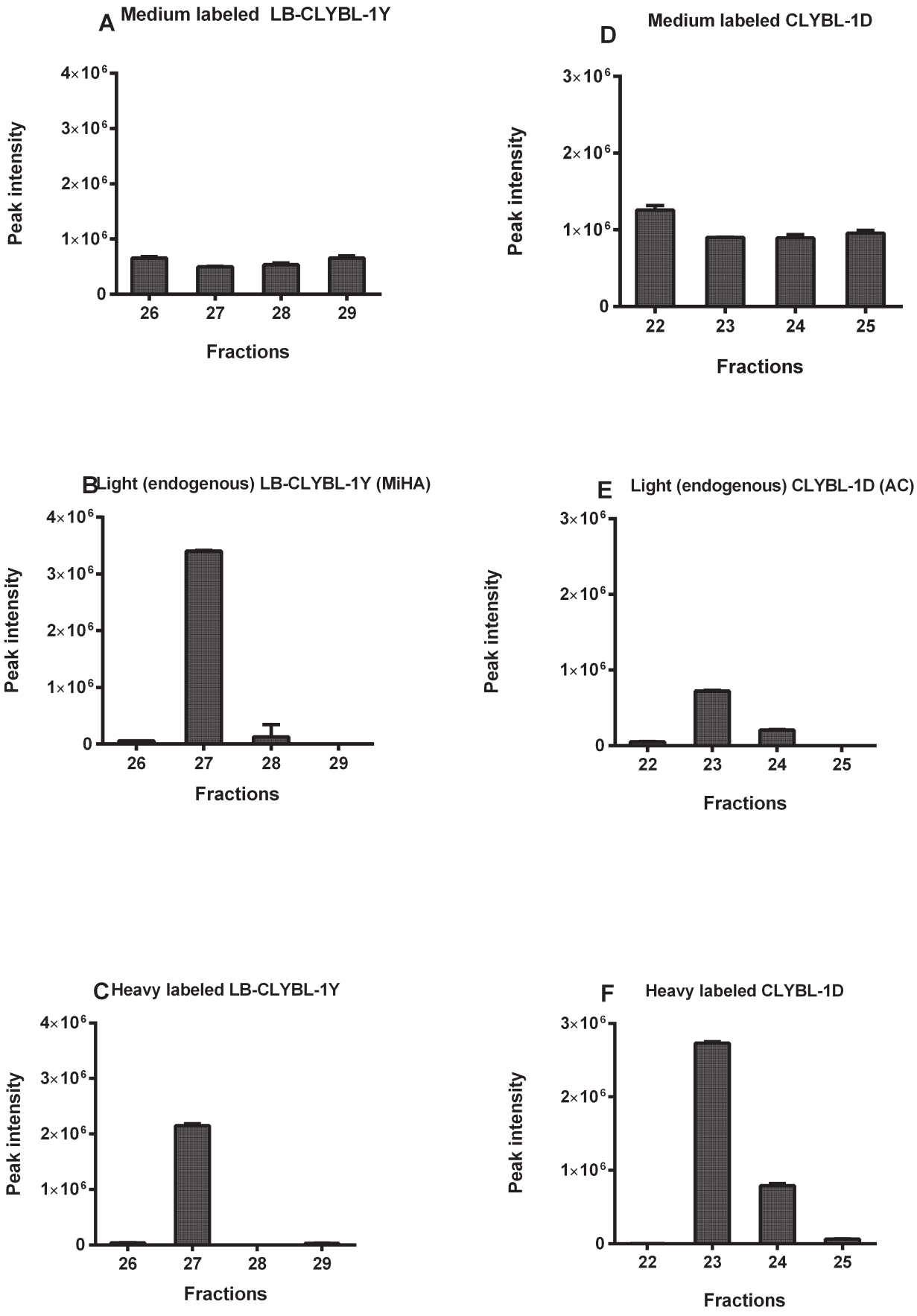


Figure 2. Combined data of peak area of the selected SRM transitions of MiHA (LB-CLYBL-1Y) and AC (CLYBL-1D) in LCL-HHC fractions.

DISCUSSION

In this study, we demonstrate by quantitative MS that the immunogenicity of 3 different MiHAs is not determined by differential surface expression of the immunogenic MiHA and the non-immunogenic AC. The accurate mass spectrometric determination of peptide copy number was achieved by a combination of a recently developed procedures in which stable isotope labeled variants of the peptide are spiked in at different stages of the procedure, see Figure 2 and supplemental Figure 3 and 4. Known quantities of hpMHC are added to the cell lysate, to account for losses during peptide elution, HPLC fractionation and sample preparation [16]. Thus, the wild type peptide to be determined and the hpMHC-derived heavy peptide variant undergo all subsequent steps of the procedure together, including affinity chromatography, further chromatographic purification and separations steps and final MS measurement. The ratio of the known amount of the heavy peptide and the wild type peptide allows a direct determination of the amount of wild type peptide. Just before MS measurement a known amount of medium peptide was added, which allows to determine the overall yield of the sample pretreatment procedure.

MiHAs differ from their AC by their ability to elicit T cell responses. This unidirectional reactivity of the T cell response is generally explained by the absence or very low abundance of the allelic counterpart on the cell surface. The low abundance of non-immunogenic AC has previously been described to be due to differences in intracellular processing between the MiHA and the AC [5, 6] or due to low HLA binding affinity of the AC [7, 8]. Furthermore, gene deletion [9], frame shifts due to nucleotide insertion [10], alternative splicing [11, 12], and introduction of a termination codon [13] could potentially result in the absence or low abundant expression of the AC at the cell surface. In previous studies bidirectional T cell responses against both the MiHA and AC have been described [14, 15], indicating that some allelic counterparts are presented at the cell-surface. We have recently shown that the ACs of some MiHA could be detected on the cell surface [23, 26] and in chapter 5 of this thesis, which seems to imply that the immunogenicity of the MiHA cannot always be explained by the absence of the AC.

We hypothesize that for a surface expressed AC not to be recognized by a T cell, the AC should be clearly less abundant than the surface expression of the MiHA, at least by a factor of 10. Probably this factor is even greater.

As can be seen in this study the cell surface expression of MiHA LB-SSR1-1S is approximately 3 times more abundant than its corresponding AC (SSR1-1L) and the cell surface expression of MiHA LB-CLYBL-1Y is approximately 5 times more abundant than the corresponding AC (CLYBL-1D). However, for the NISCH peptide pair, the AC is approximately 4 times more abundant on the cell surface than the corresponding MiHA. So in the three cases reported here, the expression of the MiHA and the corresponding AC have a comparable

cell surface expression. Therefore the unidirectional reactivity of the three MiHA specific T cell responses cannot be explained by a large differential cell surface expression of the MiHA and the AC.

The heterozygous B-LCLs used in this study presented both polymorphic variants and therefore could theoretically be recognized by T cells directed against the MiHA, as well as by T cells directed to its polymorphic counterpart. Based on the clear presence and similar abundance of the polymorphic variants we hypothesize that immune responses directed towards both variants could be induced. However, thus far we have only found immune responses to the MiHAs LB-SSR1-1S, LB-NISCH-1A and LB-CLYBL-1Y, and not against their respective ACs. In short, we screened patient samples after allogeneic stem cell transplantation (allo-SCT) and subsequent donor lymphocyte infusion for pMHC multimer positive T cells ([23, 24] and in chapter 5 of this thesis), and we enriched from the naïve T-cell repertoire MiHA and AC specific T cells using pMHC multimers [30]. Although T cells were identified specific for the MiHA LB-SSR1-1S and the LB-CLYBL-1Y in leukemia patients who entered complete remission after allo-SCT, no T cells directed against the ACs were observed in this cohort of leukemia patients. T cells directed against LB-NISCH-1A and LB-CLYBL-1Y were identified from the naïve T-cell repertoire, also here no T cells were observed directed against the two ACs. The absence of an immune response to the ACs might be due to the absence of a high affinity TCR repertoire due to tolerance against the AC. Tolerance induction against the AC might be based on homology with other self-peptides presented in the same HLA or alternative peptide-HLA complexes. Alternatively, absence of an immune response directed against the AC variant might be due to a numbers game, meaning that by carefully screening of a large cohort of patients experiencing an alloreactive immune responses after allo-SCT and subsequent donor lymphocyte infusion an AC specific immune response might be identified.

Above we have thoroughly quantified three MiHAs and their ACs. Besides these MiHA/AC pairs we identified 6 additional MiHA/AC pairs, see Table 5, during the discovery phase of our previously reported reverse immunology approach [23, 26] and chapter 5 of this thesis, from which we conclude that the phenomenon of clear presence of the AC on the cell surface does not seem to be exception, contrary to the general opinion.

In this study we accurately quantitated three MiHA and their ACs expressed in the context of HLA at the cell surface of cells heterogeneous for the two polymorphic variants. The results clearly demonstrate that both variants are expressed to a similar extent at the cell surface, demonstrating that the lack of immunogenicity towards the AC variant cannot be explained by differences in the relative abundance on the cell surface of the two polymorphic variants.

Table 5. Identification of additional MiHA and AC peptides.

Peptide sequence	BMI	Cell line	Length	NetMHC	Allele	Polymorphic AA
EEFGQAFSF	52	HHC	9	117	B44	Q
EEFGRAFSF	57	HHC	9	129	B44	R
ISISALQSL	40	HHC	9	>1000	?	L
ISISAMQSL	37	HHC	9	>1000	?	M
SPNAQILAL	42	JYpp65	9	6	B7	I
SPNAQVLAL	37	JYpp65	9	6	B7	V
YLIDGAHKI	35	HHC	9	3	A2	A
YLIDGTHKI	43	HHC	9	3	A2	T
MILPVGAAKF	65	HHC	10	>1000	?	K
MILPVGAANF	66	HHC	10	>1000	?	N
SEESAVPERSW	75	HHC	11	35	B44	E
SEESAVPKRSW	51	HHC	11	36	B44	K

ACKNOWLEDGEMENT

W. Benckhuijsen, N. Dolezal and R. Cordfunke are thanked for technical assistance. CH and MGDK were supported by the Landsteiner Foundation for Blood Transfusion Research (LSBR0713).

REFERENCE

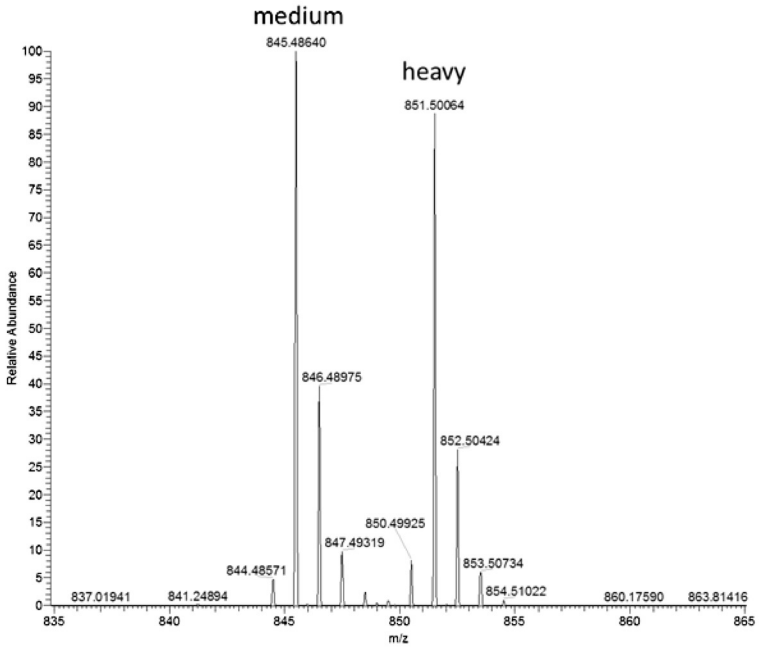
1. Kenter, G.G., Welters, M.J., Valentijn, A.R., Lowik, M.J., Berends-van der Meer, D.M., Vloon, A.P., Essahsah, F., Fathers, L.M., Offringa, R., Drijfhout, J.W. et al. (2009) Vaccination against HPV-16 oncoproteins for vulvar intraepithelial neoplasia. N Engl J Med, 361, 1838-1847.
2. Miller, J.S., Warren, E.H., van den Brink, M.R., Ritz, J., Shlomchik, W.D., Murphy, W.J., Barrett, A.J., Kolb, H.J., Giralt, S., Bishop, M.R. et al. (2010) NCI First International Workshop on The Biology, Prevention, and Treatment of Relapse After Allogeneic Hematopoietic Stem Cell Transplantation: Report from the Committee on the Biology Underlying Recurrence of Malignant Disease following Allogeneic HSCT: Graft-versus-Tumor/Leukemia Reaction. Biology of blood and marrow transplantation : journal of the American Society for Blood and Marrow Transplantation, 16, 565-586.
3. Ringden, O., Karlsson, H., Olsson, R., Omazic, B. and Uhlin, M. (2009) The allogeneic graft-versus-cancer effect. British journal of haematology, 147, 614-633.
4. Falkenburg, J.H. and Warren, E.H. (2011) Graft versus leukemia reactivity after allogeneic stem cell transplantation. Biology of blood and marrow transplantation : journal of the American Society for Blood and Marrow Transplantation, 17, S33-38.
5. Spierings, E., Brickner, A.G., Caldwell, J.A., Zegveld, S., Tatsis, N., Blokland, E., Pool, J., Pierce, R.A., Mollah, S., Shabanowitz, J. et al. (2003) The minor histocompatibility antigen HA-3 arises from differential proteasome-mediated cleavage of the lymphoid blast crisis (Lbc) oncoprotein. Blood, 102, 621-629.
6. Brickner, A.G., Warren, E.H., Caldwell, J.A., Akatsuka, Y., Golovina, T.N., Zarling, A.L., Shabanowitz, J., Eisenlohr, L.C., Hunt, D.F., Engelhard, V.H. et al. (2001) The immunogenicity of a new human minor histocompatibility antigen results from differential antigen processing. J Exp Med, 193, 195-206.
7. Spierings, E., Gras, S., Reiser, J.B., Mommaas, B., Almekinders, M., Kester, M.G., Chouquet, A., Le Gorrec, M., Drijfhout, J.W., Ossendorp, F. et al. (2009) Steric hindrance and fast dissociation explain the lack of immunogenicity of the minor histocompatibility HA-1Arg Null allele. J Immunol, 182, 4809-4816.
8. Nicholls, S., Piper, K.P., Mohammed, F., Dafforn, T.R., Tenzer, S., Salim, M., Mahendra, P., Craddock, C., van Endert, P., Schild, H. et al. (2009) Secondary anchor polymorphism in the HA-1 minor histocompatibility antigen critically affects MHC stability and TCR recognition. Proc Natl Acad Sci U S A, 106, 3889-3894.
9. Murata, M., Warren, E.H. and Riddell, S.R. (2003) A human minor histocompatibility antigen resulting from differential expression due to a gene deletion. J Exp Med, 197, 1279-1289.
10. de Rijke, B., van Horssen-Zoetbrood, A., Beekman, J.M., Otterud, B., Maas, F., Woestenenk, R., Kester, M., Leppert, M., Schattenberg, A.V., de Witte, T. et al. (2005) A frameshift polymorphism in P2X5 elicits an allogeneic cytotoxic T lymphocyte response associated with remission of chronic myeloid leukemia. J Clin Invest, 115, 3506-3516.
11. Kawase, T., Akatsuka, Y., Torikai, H., Morishima, S., Oka, A., Tsujimura, A., Miyazaki, M., Tsujimura, K., Miyamura, K., Ogawa, S. et al. (2007) Alternative splicing due to an intronic SNP in HMSD generates a novel minor histocompatibility antigen. Blood, 110, 1055-1063.
12. Broen, K., Levenga, H., Vos, J., van Bergen, K., Fredrix, H., Greupink-Draaisma, A., Kester, M., Falkenburg, J.H., de Mulder, P., de Witte, T. et al. (2011) A polymorphism in the splice donor site of ZNF419 results in the novel renal cell carcinoma-associated minor histocompatibility antigen ZAPHIR. PLoS One, 6, e21699.
13. Brickner, A.G., Evans, A.M., Mito, J.K., Xuereb, S.M., Feng, X., Nishida, T., Fairfull, L., Ferrell, R.E., Foon, K.A., Hunt, D.F. et al. (2006) The PANE1 gene encodes a novel human minor histocompatibility antigen that is selectively expressed in B-lymphoid cells and B-CLL. Blood, 107, 3779-3786.
14. Akatsuka, Y., Nishida, T., Kondo, E., Miyazaki, M., Taji, H., Iida, H., Tsujimura, K., Yazaki, M., Naoe, T., Morishima, Y. et al. (2003) Identification of a polymorphic gene, BCL2A1, encoding two novel hematopoietic lineage-specific minor histocompatibility antigens. J Exp Med, 197, 1489-1500.
15. Dolstra, H., de Rijke, B., Fredrix, H., Balas, A., Maas, F., Scherpen, F., Aviles, M.J., Vicario, J.L., Beekman, N.J., Ossendorp, F. et al. (2002) Bi-directional allelic recognition of the human minor histocompatibility antigen HB-1 by cytotoxic T lymphocytes. Eur J Immunol, 32, 2748-2758.
16. Hassan, C., Kester, M.G., Oudgenoeg, G., de Ru, A.H., Janssen, G.M., Drijfhout, J.W., Spaapen, R.M., Jimenez, C.R., Heemskerk, M.H., Falkenburg, J.H. et al. (2014) Accurate quantitation of MHC-bound peptides by application of isotopically labeled peptide MHC complexes. Journal of proteomics, 109C, 240-244.
17. Selevsek, N., Matondo, M., Sanchez Carbayo, M., Aebersold, R. and Domon, B. (2011) Systematic quantification of peptides/proteins in urine using selected reaction monitoring. Proteomics, 11, 1135-1147.
18. Kitteringham, N.R., Jenkins, R.E., Lane, C.S., Elliott, V.L. and Park, B.K. (2009) Multiple reaction monitoring

for quantitative biomarker analysis in proteomics and metabolomics. J Chromatogr B Analyt Technol Biomed Life Sci, 877, 1229-1239.

19. Whiteaker, J.R., Zhao, L., Lin, C., Yan, P., Wang, P. and Paulovich, A.G. (2012) Sequential multiplexed analyte quantification using Peptide immunoaffinity enrichment coupled to mass spectrometry. Mol Cell Proteomics, 11, M111 015347.
20. Hogan, K.T., Sutton, J.N., Chu, K.U., Busby, J.A., Shabanowitz, J., Hunt, D.F. and Slingluff, C.L., Jr. (2005) Use of selected reaction monitoring mass spectrometry for the detection of specific MHC class I peptide antigens on A3 supertype family members. Cancer immunology, immunotherapy : CII, 54, 359-371.
21. Tan, C.T., Croft, N.P., Dudek, N.L., Williamson, N.A. and Purcell, A.W. (2011) Direct quantitation of MHC-bound peptide epitopes by selected reaction monitoring. Proteomics, 11, 2336-2340.
22. Croft, N.P., Smith, S.A., Wong, Y.C., Tan, C.T., Dudek, N.L., Flesch, I.E., Lin, L.C., Tscharke, D.C. and Purcell, A.W. (2013) Kinetics of antigen expression and epitope presentation during virus infection. PLoS Pathog, 9, e1003129.
23. Hombrink, P., Hassan, C., Kester, M.G., de Ru, A.H., van Bergen, C.A., Nijveen, H., Drijfhout, J.W., Falkenburg, J.H., Heemskerk, M.H. and van Veelen, P.A. (2013) Discovery of T cell epitopes implementing HLA-peptidomics into a reverse immunology approach. J Immunol, 190, 3869-3877.
24. Van Bergen, C.A., Rutten, C.E., Van Der Meijden, E.D., Van Luxemburg-Heijs, S.A., Lurvink, E.G., Houwing-Duistermaat, J.J., Kester, M.G., Mulder, A., Willemze, R., Falkenburg, J.H. et al. (2010) High-throughput characterization of 10 new minor histocompatibility antigens by whole genome association scanning. Cancer Res, 70, 9073-9083.
25. Altman, J.D., Moss, P.A., Goulder, P.J., Barouch, D.H., McHeyzer-Williams, M.G., Bell, J.I., McMichael, A.J. and Davis, M.M. (1996) Phenotypic analysis of antigen-specific T lymphocytes. Science, 274, 94-96.
26. Hassan, C., Kester, M.G., Ru, A.H., Hombrink, P., Drijfhout, J.W., Nijveen, H., Leunissen, J.A., Heemskerk, M.H., Falkenburg, J.H. and Veelen, P.A. (2013) The human leukocyte antigen-presented ligandome of B lymphocytes. Mol Cell Proteomics.
27. Tan, T.L., Geluk, A., Toebes, M., Ottenhoff, T.H. and Drijfhout, J.W. (1997) A novel, highly efficient peptide-HLA class I binding assay using unfolded heavy chain molecules: identification of HIV-1 derived peptides that bind to HLA-A*0201 and HLA-A*0301. J Immunol Methods, 205, 201-209.
28. Sturm, R., Sheynkman, G., Booth, C., Smith, L.M., Pedersen, J.A. and Li, L. (2012) Absolute Quantification of Prion Protein (90-231) Using Stable Isotope-Labeled Chymotryptic Peptide Standards in a LC-MRM AQUA Workflow. J Am Soc Mass Spectrom, 23, 1522-1533.
29. Gerber, S.A., Rush, J., Stemman, O., Kirschner, M.W. and Gygi, S.P. (2003) Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. Proc Natl Acad Sci U S A, 100, 6940-6945.
30. Hombrink, P., Hadrup, S.R., Bakker, A., Kester, M.G., Falkenburg, J.H., von dem Borne, P.A., Schumacher, T.N. and Heemskerk, M.H. (2011) High-throughput identification of potential minor histocompatibility antigens by MHC tetramer-based screening: feasibility and limitations. PLoS One, 6, e22523.

SUPPLEMENTAL DATA

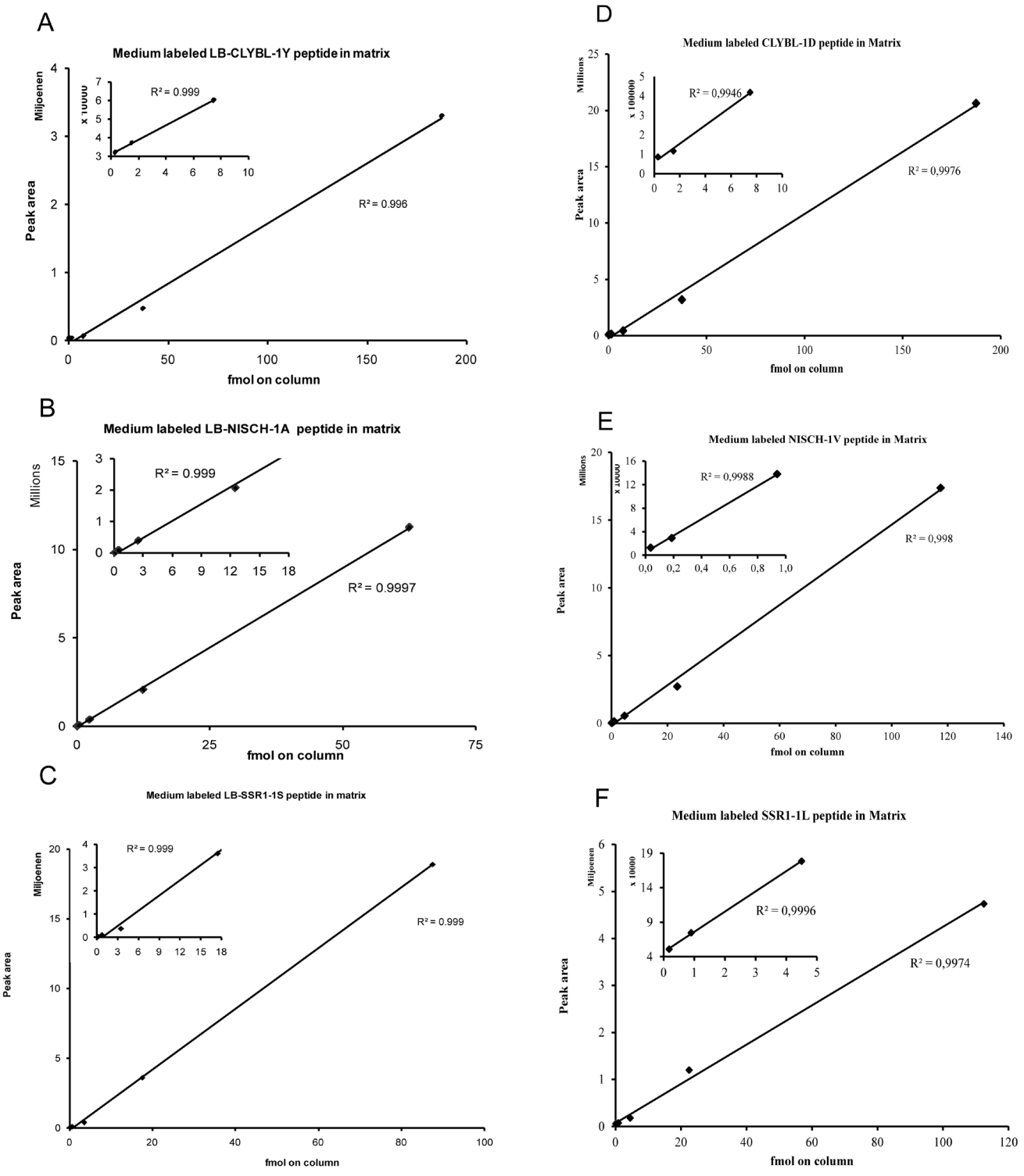
Supplemental Figure 1. The quality of the hpMHC was assessed by mixing equimolar amounts of hpMHC and the medium labeled AQUA peptide. Both peptides display the same peak height in the MS spectrum, from which we conclude that the hpMHC has been correctly loaded and purified and the measured hpMHC concentration as determined by Bradford is in line with the pure synthetic peptide concentration. Therefore the intensities of the heavy peptide and the medium peptide can be used for their relative quantitation.



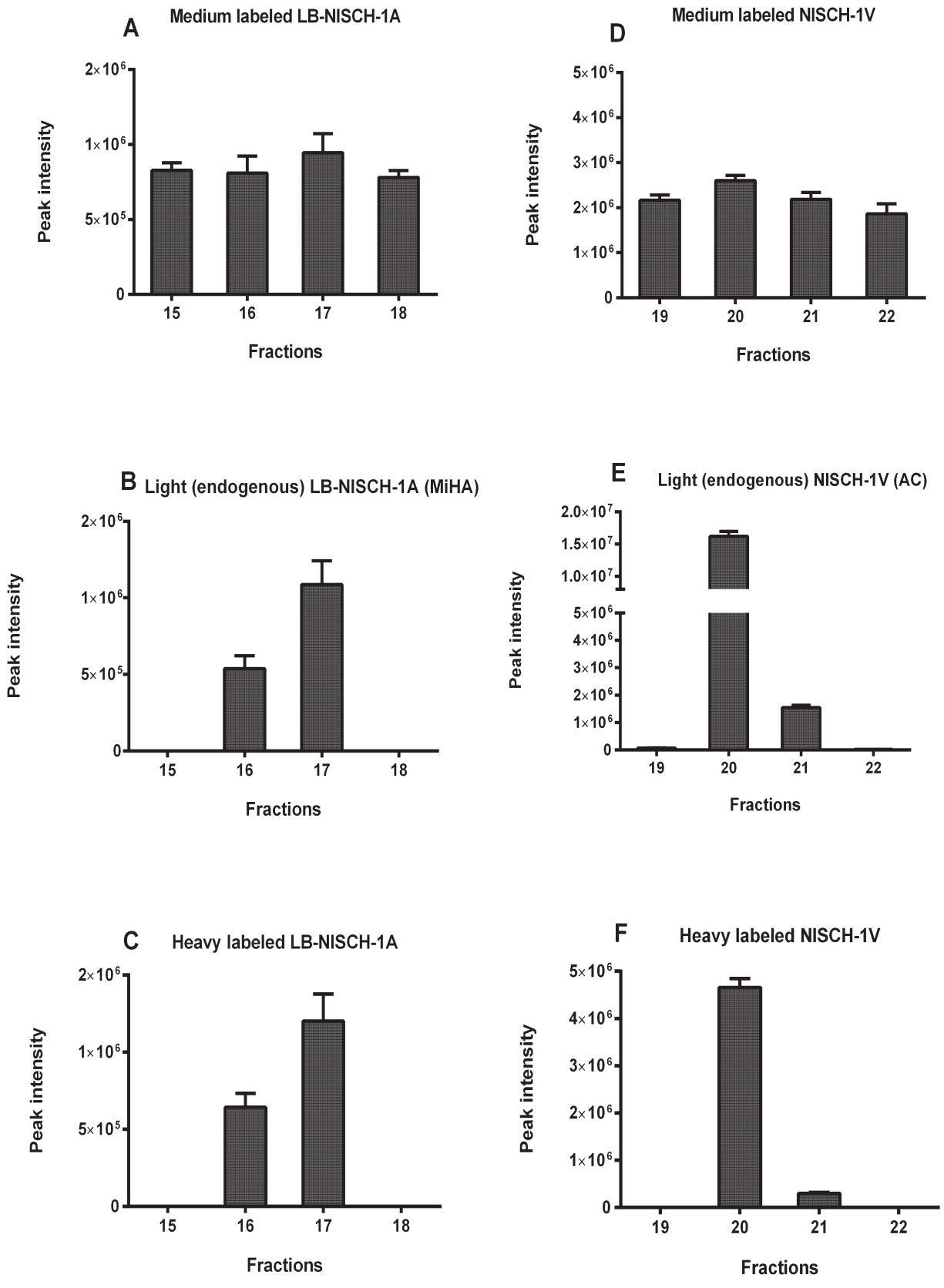
Supplemental Table 1. SRM parameters used for the acquisitions of MiHA(*) and AC(†) for each eluted (light), medium and heavy labeled peptides. The highest responding transitions and optimum CE-energy were used in the sample analysis.

Gene name	Peptide sequence	z (Q1)	Q1	Q3	CE
LB-CLYBL-1Y*	SLAAYIPRL	2	502,3	803,48 (y7)	18
		2	502,3	661,4 (y5)	21
		2	502,3	506,28 (b5)	18
	S**L**AAYIPRL	2	505,81	803,48 (y7)	18
		2	505,81	661,4 (y5)	21
		2	505,81	513,28 (b5)	18
	S**L**AAYI**P**RL	2	508,81	809,48 (y7)	18
		2	508,81	667,4 (y5)	21
		2	508,81	513,28 (b5)	18
CLYBL-1D†	SLAADIPRL	2	478,3	498,34 (y4)	32
		2	478,3	755,44 (y7)	20
		2	478,3	571,33 (b6)	17
	S**L**AADIPRL	2	481,79	498,34 (y4)	32
		2	481,79	755,44 (y7)	20
		2	481,79	578,33 (b6)	17
	S**L**AADI**P**RL	2	484,79	504,34 (y4)	32
		2	484,79	761,44 (y7)	20
		2	484,79	578,33 (b6)	17
LB-NISCH-1A*	ALAPAPAEV	2	419,75	415,22 (Y4)	12
		1	838,5	583,31 (Y6)	14
		1	838,5	415,22 (Y4)	16
	A**L**APAPAEV	2	423,25	415,22 (Y4)	12
		1	845,5	583,31 (y6)	14
		1	845,5	415,22 (y4)	16
	A**L**APA**P**AEV	2	426,25	415,22 (y4)	12
		1	851,5	589,31 (y6)	14
		1	851,5	415,22 (y4)	16
NISCH-1V†	ALAPAPVEV	1	866,53	611,34 (y6)	29
		1	866,53	443,25 (y4)	31
		1	866,53	620,39 (b7)	36
	A**L**APAPVEV	1	873,53	611,34 (y6)	29
		1	873,53	443,25 (y4)	31
		1	873,53	627,39 (b7)	36
	A**L**APA**P**VEV	1	879,53	617,34 (y6)	29
		1	879,53	443,25 (y4)	31
		1	879,53	633,39 (b7)	36
LB-SSR1-1S*	VLFRGGPRGSLAVA	3	467,32	655,9 (b13 (2+))	16
		3	467,32	606,36 (b12(2+))	18
		3	467,32	570,84 (b11(2+))	27
	V**L**FRGGPRGSLAVA	3	469,62	659,4 (b13(2+))	16
		3	469,62	609,86 (b12(2+))	18
		3	469,62	574,34 (b11(2+))	27
	V**L**FRGG**P**RGSLAVA	3	471,62	662,4 (b13(2+))	16
		3	471,62	612,86 (b12(2+))	18
		3	471,62	577,34 (b11(2+))	27
SSR1-1L†	VLFRGGPRGLLAVA	3	476	669,1 (b13(2+))	16
		3	476	619,39 (b12(2+))	19
		3	476	583,87 (b11(2+))	28
	V**L**FRGGPRGLLAVA	3	478,3	672,42 (b13(2+))	16
		3	478,3	622,89 (b12(2+))	19
		3	478,3	587,37 (b11(2+))	28
	V**L**FRGG**P**RGLLAVA	3	480,3	675,42 (b13(2+))	16
		3	480,3	625,89 (b12(2+))	19
		3	480,3	590,37 (b11(2+))	28

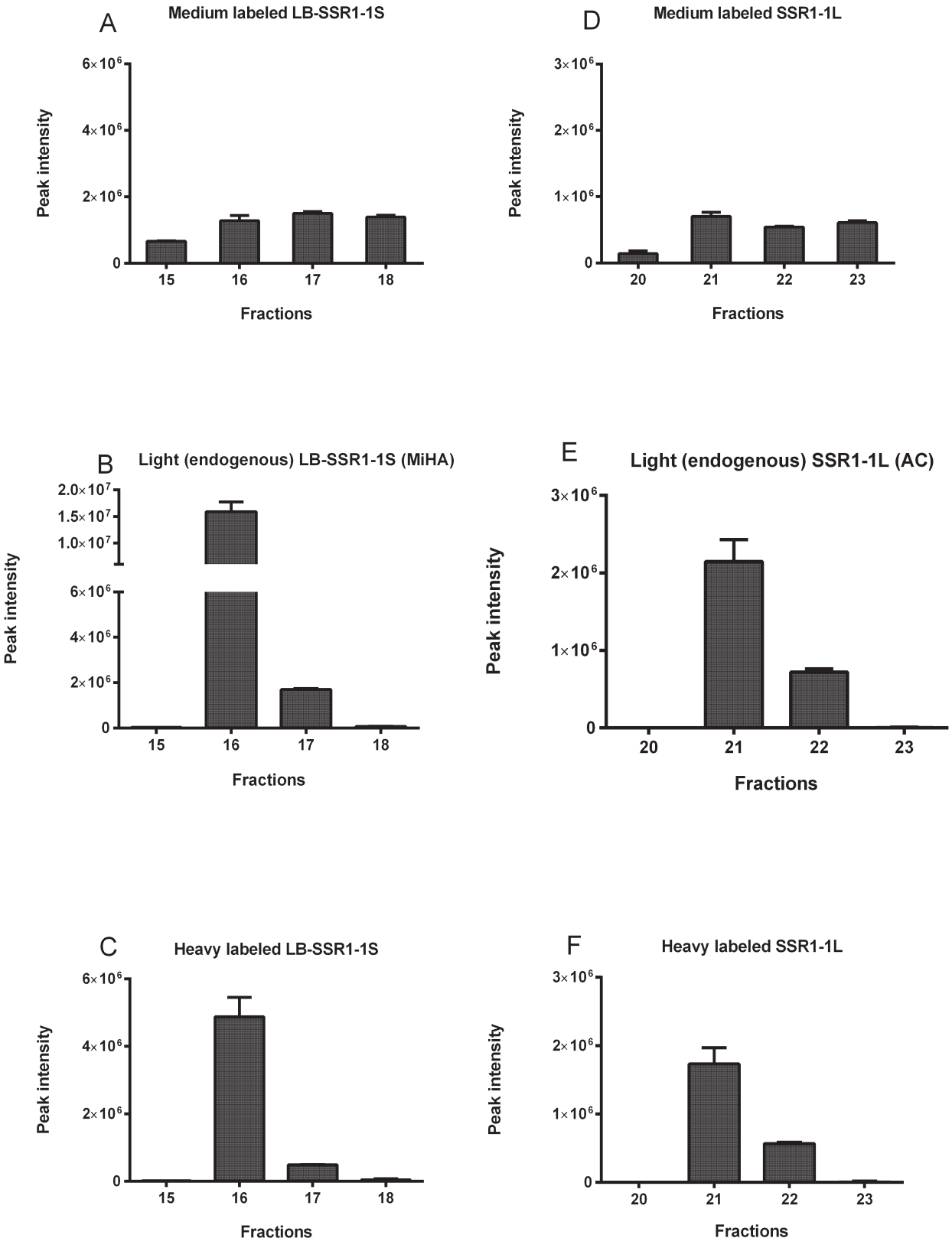
Supplemental Figure 2. Calibration curve of medium labeled peptide (A) LB-CLYBL-1Y, (B) LB-NISCH-1A, (C) LB-SSR1-1S, (D) CLYBL-1D, (E) NISCH-1V and (F) SSR1-1L in matrix.

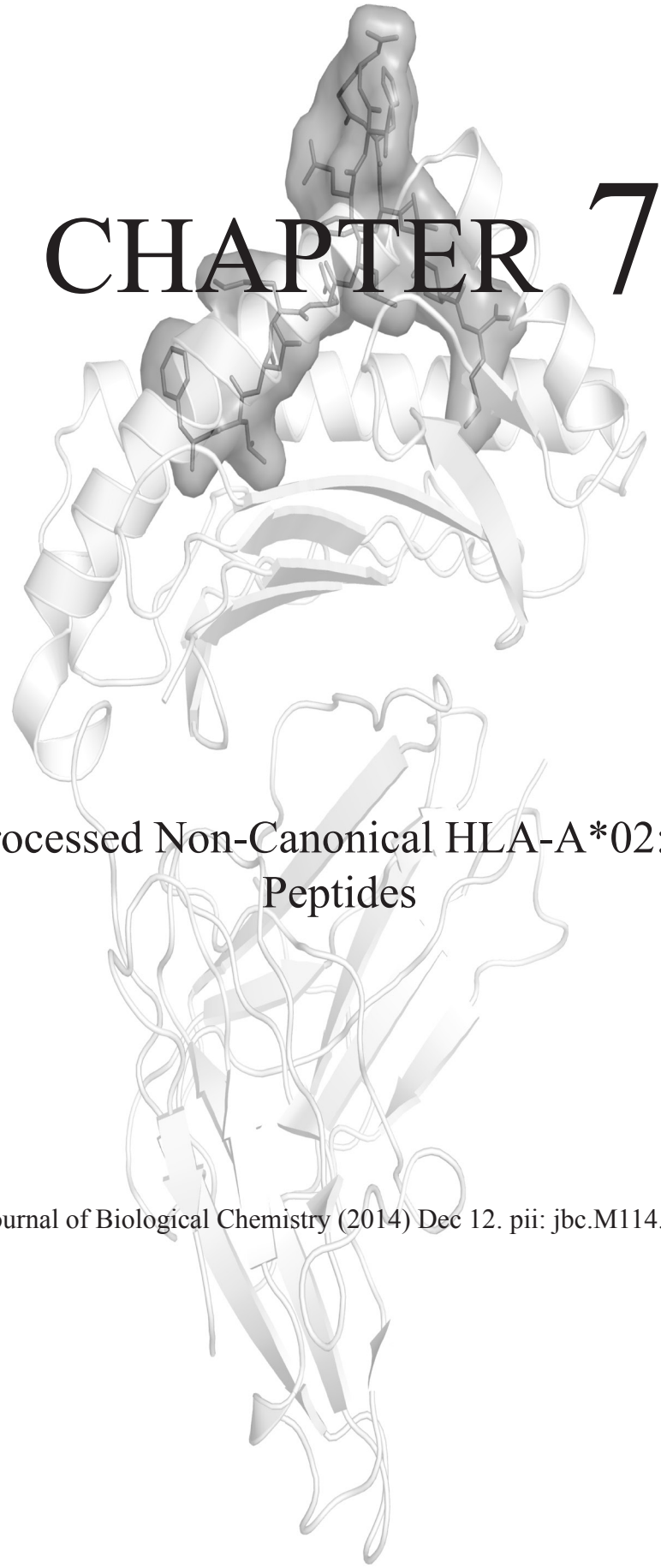


Supplementary Figure 3. Combined data of peak area of the selected SRM transitions of MiHA (LB-NISCH-1A) and AC (NISCH-1V) in LCL-JY fractions



Supplemental Figure 4: . Combined data of peak area of the selected SRM transitions of MiHA (LB-SSR1-1S) and AC (SSR1-1L) in LCL-JY fractions.





CHAPTER 7

Naturally Processed Non-Canonical HLA-A*02:01 Presented Peptides

The Journal of Biological Chemistry (2014) Dec 12. pii: jbc.M114.607028

Based on:

Chopie Hassan*
Eric Chabrol*
Lorenz Jahn
Michel G.D. Kester
Arnoud H. de Ru
Jan W. Drijfhout
Jamie Rossjohn
J.H. Frederik Falkenburg
Mirjam H. M. Heemskerk
Stephanie Gras#
Peter A. van Veelen#

* equal contribution
# shared senior authorship

7

Naturally Processed Non-Canonical HLA-A*02:01 Presented Peptides

ABSTRACT

Human Leukocyte Antigen (HLA) class I molecules generally present peptides (p) of 8 to 11 amino acids (aa) in length. Although an increasing number of examples with lengthy (>11 aa) peptides, presented mostly by HLA-B alleles, have been reported. Here we characterise HLA-A*02:01 restricted, in addition to the HLA-B*0702 and HLA-B*4402 restricted, lengthy peptides (>11 aa) arising from the B-cell ligandome. We analysed a number of 15-mer peptides presented by HLA-A*02:01, and confirmed pHLA-I-formation by HLA-folding and thermal stability assays. Surprisingly the binding affinity and stability of the 15-mer epitopes in complex with HLA-A*02:01 were comparable to the values observed for canonical length (8 to 11 aa) HLA-A*02:01-restricted peptides. We solved the structures of two 15-mer epitopes in complex with HLA-A*02:01, within which the peptides adopted distinct super-bulged conformations. Moreover, we demonstrate that T-cells can recognize the 15-mer peptides in the context of HLA-A*02:01, indicating that these 15-mer peptides represent immunogenic ligands. Collectively, our data expand our understanding of longer epitopes in the context of HLA-I, highlighting that they are not limited to HLA-B family, but can bind the ubiquitous HLA-A*02:01 molecule, and play an important role in T-cell immunity.

INTRODUCTION

Human Leukocyte Antigen (HLA) class I molecules are expressed on the surface of all nucleated cells presenting peptides for $CD8^+$ T-cell recognition. The peptides presented in HLA class I molecules are protein fragments of intracellular origin, which are degraded by an array of proteases, the most prominent of which is the proteasome. The protein fragments are truncated to smaller peptides and translocated into the endoplasmic reticulum (ER). In the ER, the peptide-HLA class I molecule (pHLA) is assembled from a peptide, a polymorphic heavy chain and the monomorphic light chain called β2-microglobulin (β2m). Both β2m and the peptide are required for the stability of the HLA class I molecule. A peptide with adequate binding motif residues will bind into the peptide-binding groove of the HLA class I molecule, allowing the assembled molecule to leave the ER and be transported via the Golgi complex to the cell surface to display the peptides to $CD8^+$ T-cells [1].

It has long been reported that HLA class I molecules can accommodate 8-11 mer peptides, typically 9-mers [1-3]. Over the last few years, different groups have reported the binding of 12-mer, 13-mer, 14-mer, and even a 16-mer peptides to HLA class I molecules [4-13]. Crystallographic and biophysical studies showed the binding of a 13-mer viral epitope to the HLA-B*3508 molecule and T-cell recognition of the bulged peptide [12, 14-16]. The synthetic elongation of previously defined T-cell epitopes by central amino acid insertion revealed binding of 8-25 mer peptides to HLA-B*3508, although central amino acid insertion was not generally tolerated well for all peptides [17].

Since some longer peptides are recognized by T-cells, such peptides may play an important role in T-cell mediated therapies for cancer, and in vaccine design. So far, a rather limited number of naturally processed and presented longer peptides have been reported, and notably the majorities involve HLA-B alleles. Generally, previous reports on longer peptides have focused on a single or a few isolated peptides. A more general view on the contribution of longer peptides to the HLA-ligandome, in-depth analysis is required. One of our previous studies [8] provided an in-depth analysis, and allowed the selection of longer peptides for follow-up studies. Therefore, in the present study, we report on these longer peptides, i.e 14-23-mers, binding to the HLA-B family members, namely HLA-B*4402 and HLA-B*0702, and more surprisingly to the HLA-A family molecule, HLA-A*02:01. Our analysis was focused on the common HLA-A*02:01 allele and its ability to bind 15 amino acid long epitopes. After elution and sequencing of the 15-mer peptides, bound to HLA class I molecules, we analysed the pHLA-A*02:01 stability. We compare the binding affinity and stability of 15-mer-HLA-A*02:01 complexes with the canonical length 9 and 10-mer peptides bound to the same HLA molecule. We subsequently solved the structures of two distinct 15-mer epitopes in complex with the HLA-A*02:01 molecule, and show that they exhibited contrasting conformations of their central

bulged region. Finally we formally establish that HLA-A*02:01 loaded with 15-mer peptides are antigenic targets for the T-cells, using tetramers loaded with the 15-mer epitopes to isolate reactive T-cells.

EXPERIMENTAL PROCEDURES

Cellular sample preparation

Sample preparation was as described in [8]. Briefly, Epstein-Barr virus (EBV) transformed B lymphoblastic cell lines (EBV-LCL) LCL-HHC (typing: HLA-A*02:01, B*0702, B*4402, Cw*0501 and Cw*0702) and LCL-JYpp65 (typing: HLA-A*02:01, B*0702 and Cw*0702) were expanded in roller bottles, using IMDM, supplemented with 10% heat-inactivated fetal bovine serum (FBS), penicillin/streptomycin and L-glutamine, were collected, washed with ice cold PBS and stored at -80 ^{o}C until use. Antibodies were produced, purified and stored as described in [8].

Isolation of HLA class I-presented peptides

Pellets of LCL-JYpp65 and LCL-HHC cell lines were lysed in 50 mM Tris-HCl, 150 mM NaCl, 5 mM EDTA, and 0.5% Nonidet-P40 (pH 8.0) and supplemented with Complete protease inhibitors (Roche). The total concentration of the cells in the lysis buffer was 0.1x 109 cells/ml. After 2 hours incubation with tumbling of the cells in the lysis buffer at 4 ^{o}C, the preparation was centrifuged at 4 ^{o}C for 10 minutes at 2070 xg. The supernatant was transferred to a new tube and centrifuged at 4 ^{o}C for 35 minutes at 19,000 xg. The supernatant was pre-cleared with CL4B beads and subjected to the W6/32 immunoaffinity column with a flow rate of 2 ml/min. After washing, bound peptide-HLA class I complexes were eluted from the column, and dissociated, with 10% acetic acid. Peptides were separated from the HLA class I molecules by passage through a 10 kDa membrane (Pall macrosep centrifuge devices), and further purified by solid phase extraction (C18 Oasis, 100 µl bed volume, Waters), freeze dried and resuspended in 95/3/0.1 water/ACN/FA, v/v/v.

Peptide separation

The pools of peptides eluted form the two EBV-LCL lines were divided in three portions (LCL-JYpp65) or two portions (LCL-HHC). The LCL-JYpp65 pools were separated by peptide IEF, SCX and C18 chromatography and the LCL-HHC pools were separated by peptide IEF and SCX chromatography, as described in [8] to achieve a high number of identified peptides. The fractions obtained from the three off line separation techniques were further fractionated and analyzed by nano-LC-MS/MS.

Mass Spectrometry data analysis

The tandem mass spectra were matched against the International Protein Index (IPI) human database version 3.87, using the mascot search engine version 2.2.04 (Matrix Science, London, UK)), with a precursor

mass tolerance of 2 ppm, with methionine oxidation as a variable modification, and a product ion tolerance of 0.5 Da. Scaffold software version 3 (www.proteomesoftware.com) was subsequently used to process the mascot output files and generate spectrum reports. Duplicates were removed, and peptides longer than 11 amino acids with mascot ion score ≥35 were selected (Supplemental Table 1). The selection of a mascot ion score >35 has been thoroughly discussed in Hassan et al.[8].

Peptide synthesis

Peptides were synthesized using standard fluorenylmethoxycarbonyl (Fmoc) chemistry using a SyroII peptide synthesizer (MultiSynTech, Witten, Germany) (Table 1). The integrity of the peptides was checked using RP-HPLC and MS. The purity of the peptides was higher than 95%.

Refolding of pHLA monomers

Recombinant HLA-A*02:01 heavy chain and human β2m light chain were in-house produced in Escherichia coli. The refolding was performed by adding 1.8 mg of HLA-A*02:01 heavy chain solubilised in 8 M urea, 1.2 mg of β2m dialyzed to PBS and 2 mg of peptide dissolved in DMSO, to 50 ml of cold refolding buffer; (400 mM L-arginine HCl, 100 mM Tris-HCl pH 8; 5 mM reduced glutathione, 0.5 mM oxidized glutatione-Na, 2 mM EDTA, 5% glycerol, Complete protease inhibitors (Roche)), and vigorously mixed after each step. The mixture was incubated for 72 hr. at 10 °C. The refolded protein mixture was concentrated to a volume of 0.5 ml with an Amicon concentrator (membrane cutoff, 30 kDa), then purified by gel filtration using fast protein liquid chromatography on a Superdex 75 column (Amersham Biosciences) and PBS as eluent. Complexes were stored at -80°C. The efficiency of the refolding (recovery) is determined by protein concentration measurement of the formed pHLA by the Bradford protein assay.

Preparation of pHLA tetramers

Biotinylated pHLA complexes containing the FLNKDLEVDGHFVTM (FLNKD) or the ALQDAGDSSRKEYFI (ALQDA) peptide bound to HLA-A*02:01 (RAB9AFLNKD:HLA-A*02:01 or GYPCALQDA:HLA-A*02:01, respectively) were conjugated to streptavidin-coupled phycoerythrin (SA-PE, Invitrogen) or allophycocyanin (SA-APC, Invitrogen) to form pHLA-tetramers. Thereto, RAB9AFLNKD:HLA-A*02:01 and GYPCALQDA:HLA-A*02:01 complexes were incubated with SA-PE or SA-APC at empirically determined ratios of 12:1 and 10:1, respectively, based on biotinylation efficiency. Concentration of pHLA-tetramers was adjusted to 0.2 μg/μl with PBS. pHLA-tetramers were stored at 4 °C.

Isolation of peptide-specific T-cell clones

After having obtained informed consent, peripheral blood mononuclear cells (PBMCs) of HLA-A*02:01-negative healthy individuals were isolated by Ficoll-density gradient. To isolate RAB9A or GYPC-reactive T cells by enrichment with pHLA-tetramers, a previously described protocol was used with mi-

nor modifications (18). PBMCs were incubated with PE-labelled RAB9AFLNKD:HLA-A*02:01 and GYPCALQDA:HLA-A*02:01-tetramers for 1 hour at 4 °C. Cells were washed twice and incubated with anti-PE magnetic microbeads (Miltenyi Biotec, Bergisch Gladbach, Germany). PE-labelled cells were enriched via magnetic associated cell sorting (MACS) on a LS column (Miltenyi Biotec) according to manufacturer's instructions (Miltenyi Biotec). Subsequently, positive fractions were incubated with an antibody against CD8 (Invitrogen/Caltag, Buckingham, UK) in combination with antibodies against CD4, CD14, and CD19 (BD Pharmingen, San Jose, CA, USA) for 15 min at 4 °C. Cells were washed twice. pHLA-binding $CD8^+$ T-cells were single-cell sorted on a FACSAria (Becton Dickinson Bioscience) into 96-well round-bottom culture plates containing 50,000 irradiated (35 Gy) autologous PBMCs in 100 µl culture medium composed of IMDM (Lonza, Basel, Switzerland) supplemented with 100 IU/ml IL-2 (Proleukine; Novartis Pharma, Arnhem, The Netherlands), 5 % FBS (Gibco, Life Technologies, Carlsbad, CA, USA), 5 % human serum, and 0.8 µg/ml phytohemagglutinin (PHA; Remel, Lenexa, KS, USA).

FACS analysis of isolated T-cell clones

20,000 T-cells of a particular clone were stained with 10 µl of pHLA-tetramers in a final concentration of 2 µg/ml per pHLA-tetramer for 15 min at 37 °C. Cells were washed once and analysed on a LSRII (Becton Dickinson Biosciences) using Diva software (Becton Dickson Biosciences).

Functional analysis of T-cell clones

2,000 T-cell of a particular clone were co-incubated with 30.000 T2 cells or EBV-transformed B lymphoblastic cell lines (B-LCLs). T2 cells were loaded with different concentrations of peptide for 30 min at 37 °C prior to co-incubation with T-cell clones. Following 18 hr. of co-culture, supernatant was harvested and GM-CSF secretion was assessed using standard enzyme-linked immunosorbent assay (ELISA; R&D systems) following manufacturer's instructions.

HLA competition refolding assay

The competition refolding assay has been developed previously [19]. Briefly, this assay employs unfolded recombinant HLA-A*02:01 heavy chain in combination with folded β2m and the commercially available fluorescent standard peptide (FLPSDCFlFPSV, a modified HBV epitope), and relies on protein folding during the assay. The peptide of interest competes with the labelled standard peptide for binding. After 24 hr. of incubation, protein complexes and free peptide are separated by size-exclusion chromatography, during which the fluorescence of protein and peptide fractions are monitored. Following peak integration of the fluorescent signals, the ratio of label in the protein and peptide fraction is calculated. The affinities of the peptides are expressed as IC50, the peptide concentration at which binding of the standard peptide is reduced to 50% (Table 1). In this assay we used three epitopes with high binding affinity to the HLA-A*02:01 molecules;

(LB-NISCH-1A (ALAPAPAEV), CMV-pp65-NLV (NLVPMVATV) and MART1-M-ELA (ELAGIGILTV) [18]. We used MART1-WT-AAG (AAGIGILTV) as a low affinity binder to the HLA-A*02:01 molecule [18].

Thermal stability assay

To assess the stability of each peptide in complex with the pHLA-A*02:01, a thermal shift assay was per-

Table 1. HLA-A*02:01 specific peptides

Gene name	Protein name	Peptide sequence	Length	BMI	IC50 (nM)
GYPC	Isoform Glycophorin-C of Glycophorin-C	ALQDAGDSSRKEYFI	15	40	414
GNB3	Guanine nucleotide-binding protein G(I)/G(S)/G(T) subunit beta-3	ALWDIETGQQKTVFV	15	39	15
RAB9A	Ras-related protein Rab-9A	FLNKDLEVDGHFVTM	15	76	83
NUDCD2	NudC domain-containing protein 2	KLFDSTIADEGTWTL	15	75	10
NFKB1	Isoform 2 of Nuclear factor NF-kappa-B p105 subunit	KLLEIPDPDKNWATL	15	36	24
ZNF828	Zinc finger protein 828	KLMEALEPPLEEQQI	15	55	1366
EEF2	Elongation factor 2	LLYEGPPDDEAAMGI	15	92	55
ARPC3	Actin-related protein 2/3 complex subunit 3	SLMDPDTKLIGNM*AL	15	57	51
LB-NiSCH-1A	Imidazoline receptor antisera-selected protein2	ALAPAPAEV	9	42	52
CMV	PP65	NLVPMVATV	9	28	45
MART1-M-ELA	Melanoma antigen modified analogue	ELAGIGILTV	10	n.a.	46
MART1-WT-AAG	Melanoma antigen wild type	AAGIGILTV	9	n.a.	6955

AA, amino acid length; BMI, best mascot ion score; IC50, and binding affinity of peptides; *oxidized methionine residue; n.a., not applicable.

formed. The fluorescent dye Sypro orange was used to monitor the protein unfolding. The thermal stability assay was performed in the Real Time Detection system (Corbett RotorGene 3000), originally designed for PCR. Each pHLA-A*02:01 complex in 10 mM Tris-HCl pH 8, 150 mM NaCl, at two concentrations (5 and 10 μM) in duplicate, was heated from 25 to 95°C with a heating rate of 1°C/min. The fluorescence intensity was measured with excitation at 530 nm and emission at 555 nm. The Tm, or thermal melt point, represents the temperature required to unfold 50% of the protein [20] (Table 2).

Crystallisation, data collection and structure determination

Crystals of the HLA-A*02:01-FLNKD and HLA-A*02:01-ALQDA complexes were grown by the hanging-drop, vapour-diffusion method at 20°C with a protein/reservoir drop ratio of 1:1, at a concentration of 10 mg/mL of protein using 18-22% PEG 3350; 0.1 M HEPES pH 7.5 and 0.1 M MgCl2. Crystals were soaked in a cryoprotectant solution containing mother liquor solution with the PEG concentration increase to 35%(w/v) and then flash frozen in liquid nitrogen. The data were collected on the MX1 beamline at the Australian Synchrotron (Clayton) using an ADSC-Quantum 210 CCD detector (at 100K), processed using the XDS software [21] and scaled using SCALA software [22] from the CCP4 suite [23]. The structures were determined by molecular replacement using the PHASER [24] program with the HLA-A*02:01 minus the peptide as the search model for the MHC (Protein Data Bank accession number, 3GSO [25]). Manual model building was conducted using the Coot software [26] followed by maximum-likelihood refinement with the PHENIX program [27]. The final models have been validated using the Protein Data Base validation web site and the final refinement statistics are summarized in Table 3. Coordinates submitted to PDB database, HLA-A*02:01-FLNKD code 4U6X and HLA-A*02:01-ALQDA code 4U6Y. All molecular graphics representations were created using PyMol [28]. The interactions between the peptides and the HLA have been calculated using CONTACT in the CCP4 software suite (23).

RESULTS

Non-canonical peptides presented in HLA class I molecules

The list of eluted peptides from the two EBV-LCLs comprised 15,882 peptides, based on a length of 8-23 amino acids and a mascot ion score >35. The list contained 1,568 peptides of 12-23 length, of which 1,145 were 12-14 mers and 423 peptides are longer than 14 amino acids (Supplemental Table 1 and Figure 1).

The 8-11 mer peptides have been reported by Hassan et al. [8] (Figure 1A), and so we concentrated our study on the peptides of non-canonical length (> 11 aa). It is important to note that in large scale proteomics experiments a certain false discovery rate (FDR) is acceptable. For HLA-presented peptides 5% is accepted as FDR (8, 29). Therefore, it cannot be excluded that a few peptides might have been incorrectly assigned, but the large majority will have been correctly assigned. In addition, we performed our immunopurification

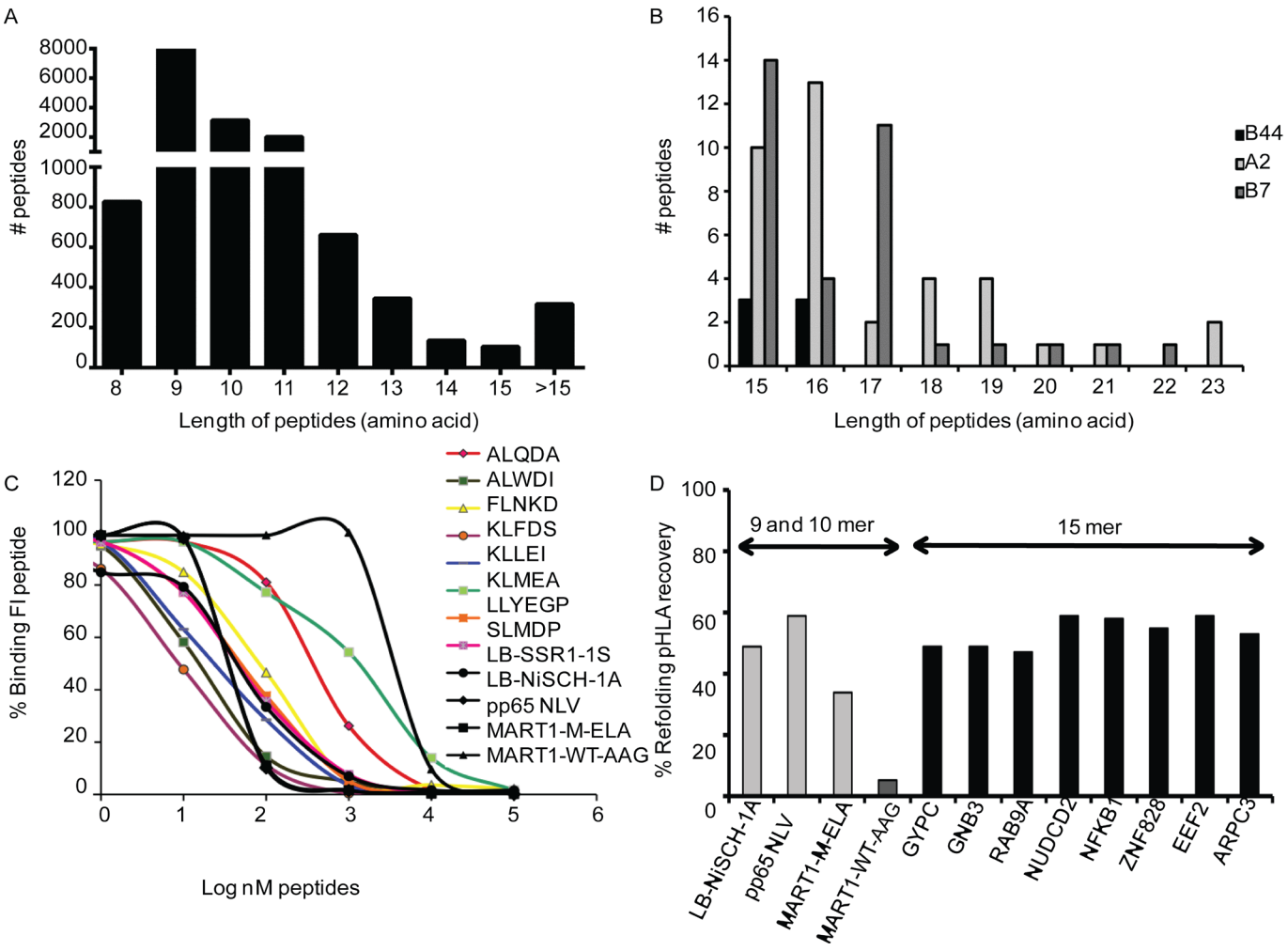


Figure 1. Peptide length distribution of the HLA-ligandome and binding affinity. Distribution of all peptides eluted from two EBV-LCL cell lines (A), and a focus on 15-23 mer peptides (B). A substantial percentage of HLA-ligands are longer than of canonical 8-11 mers. Binding affinity as determined in an HLA competition refolding assay (C), and an HLA refolding efficiency assay (D).

experiments with a pan class I-antibody, w6/32, which might complicate assignment of peptides to a particular allele. However, in this study the A alleles and B alleles have clearly distinct motifs. The known HLA C allele present in our cells, as known from the SYFPEITHI database, do not fulfill our A and B motifs.

To estimate the number of potentially relevant non-canonical binders to the HLA-molecules we used NetMHC, and initially used a simple definition of binders by definition of the P2 anchor; HLA-A*02:01 (P2: LMV), HLA-B*0702 (P2: P), HLA-B*4402 (P2: E). 922 out of the 1,145 12-14 mers (81%) fulfilled this P2 anchor criterion, which compares well with the 75% of binders as found using NetMHC (with a score <1000 nM).

Of the listed 15-23 mer peptides 214 out of 423 (51%) fulfilled the P2 anchor criterion (Figure 1B). Since NetMHC does not allow calculation of binding affinities beyond 14-mers, we defined potential binders more stringently, based on the following mandatory anchor residues, including the PΩ position; HLA-A*02:01 (P2: LMV and PΩ: LMIVA), HLA-B*0702 (P2: P and PΩ: L), HLA-B*4402 (P2: E and PΩ: F,Y). This additional constraint resulted in the selection of 77 peptides with a length of 15-23 amino acids (Supplemental Table 2).

Surprisingly, our result indicated that while previous studies on longer epitopes were based on HLA-B molecules, it was evident from the elution assay and mass spectrometry analysis that HLA-A molecules, including HLA-A*02:01, have also the ability to bind longer epitopes (Figure 1B).

Competition refolding assay

From the ten 15-mer peptides found fulfilling the P2 and PΩ criteria for HLA-A*02:01, eight with P2=L were synthesized for further characterisation (Table 1). To show the binding efficacy of these naturally processed 15-mer peptides, we performed a refolding and competition assays (Table 1) [19]. The two assays are complementary. The competition refolding assay shows the ability of the peptide to bind. The refolding assay shows the efficiency of formation (i.e. the yield of the HLA-monomer folding process). The yield is an additional important parameter for pHLA stability and a predictor of efficient tetramer formation. Several other peptides with known binding affinities were included in the assay to evaluate the relative binding affinity of the 15-mer peptides. LB-NISCH-1A (ALAPAPAEV), MART1-M-ELA (ELAGIGILTV), and CMV-pp65-NLV (NLVPMVATV) peptides are known high affinity binders to the HLA-A*02:01 molecule, and were included as control [18]. The MART1-WT-AAG (AAGIGILTV) epitope was included as a low affinity binder to the HLA-A*02:01 molecule. In the competition assay the fluorescein (F1)-labeled reference peptide (FLPSDCFl FPSV), known to bind efficiently to the HLA-A*02:01 molecule, and the peptide of interest compete for binding in the HLA class I groove during folding. The affinities of the peptides are expressed as IC50 (Table 1). The calculated percentage of bound fluorescent reference peptide after competition with

the 15-mer peptides, and the high and low affinity reference peptides are listed in Table 1, and are plotted in Figure 1C.

The results showed that all eight synthesized 15-mer peptides, fulfilling the HLA A*0201-motif, have an IC50 between 10 nM to 1366 nM, most of which are in the high binding affinity range [19]. For comparison, the low binding affinity peptide MART1-WT-AAG (AAGIGILTV) has a higher IC50 of approximately 7,000 nM, while the high binding affinity peptide pp65-NLV has an IC50 of 45 nM. These results illustrate that the 15-mer peptides bind to the HLA-A*02:01 molecule with similar affinity as 8-11-mer peptides, and some even with higher affinity such as the KLFDS (IC50 of 10 nM, Table 1).

In summary the 15-mer epitopes exhibited affinities comparable to that of 9-10 mers bound to HLA-A*02:01, showing that the length was not an obstacle for peptides to bind the common HLA-A*02:01

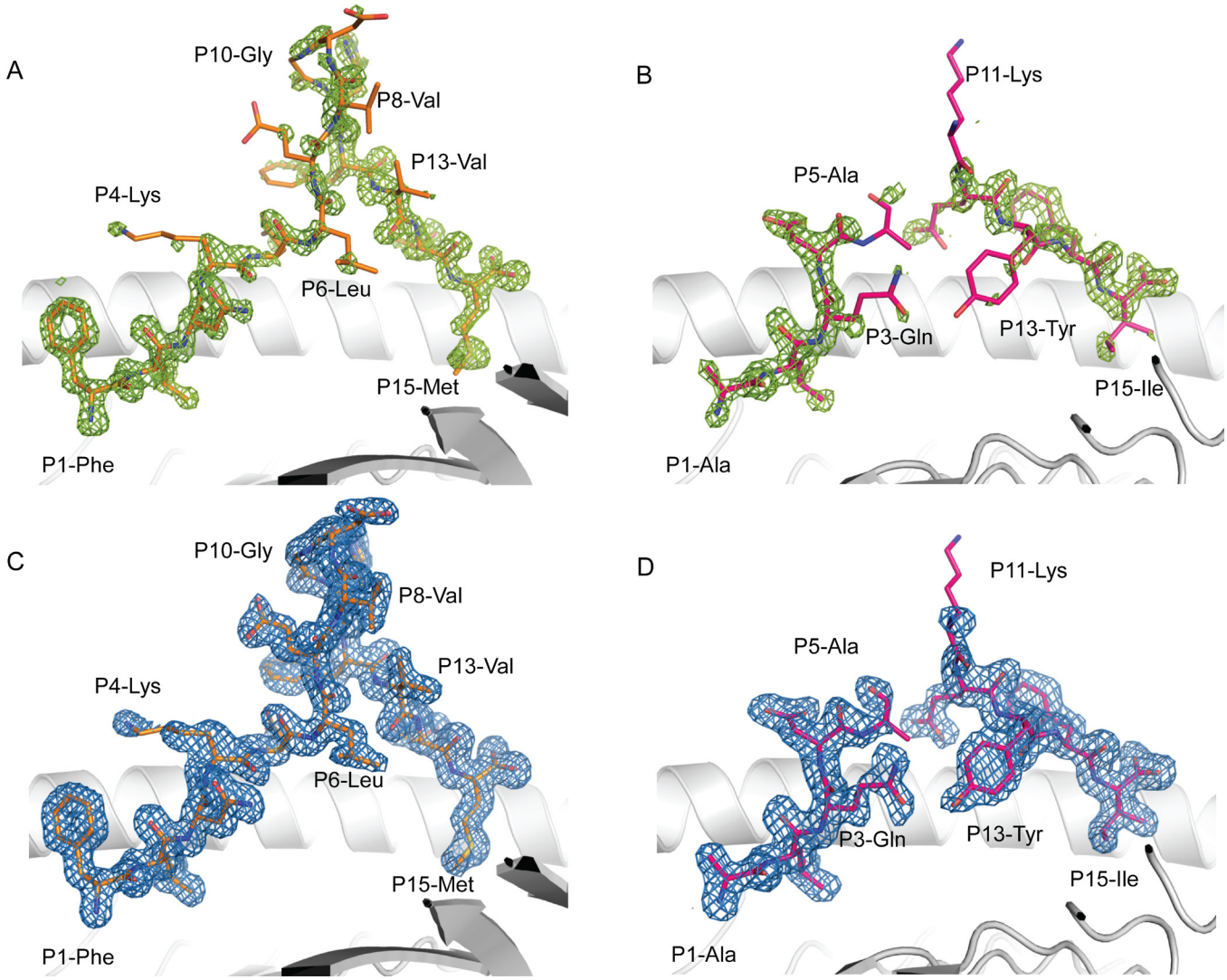


Figure 2. Electron density map of 15-mer epitopes bound to HLA-A*02:01. The A and B panels show the omit map (Fo-Fc) at 3σ in green for the FLNKD and ALQDA peptides in complex with the HLA-A*02:01 respectively. The C and D panels show the 2Fo-Fc map contoured at 1σ in blue after final refinement for the FLNKD and ALQDA peptides respectively. The HLA-A*02:01 is represented as white cartoon; the peptides are represented in stick and coloured in orange for the FLNKD and pink for the ALQDA.

allele.

pHLA complexes refolding efficiency assay

We next applied an HLA refolding efficiency assay to assess the binding of 15-mer peptides to HLA-A*02:01 by measurement of the yield of formation of pHLA. This assay determined the capacity of the peptides to support stable refolding of the heavy chain and β2m recombinant subunits of the HLA-A*02:01 complex. The yield of folded pHLA-A*02:01 was determined for the classical length (9-mer) and longer peptides (15-mer) under the same refolding conditions. HLA recovery levels of 47%-59% were obtained for the eight 15-mer peptides. The yields of the three known high affinity binders LB-NISCH-1A, CMV-pp65-NLV and MART1-M-ELA were 49 %, 58.8 % and 34% respectively. The weak binder MART-1-WT-AAG showed a pHLA recovery yield of 5.4% (Figure 1D). These results indicate that the 15-mer peptides have a similar binding efficiency as the classical high affinity 9-mer peptides, and so are able to stabilize the formation of the HLA-A*02:01-ß2m complex as well as canonical peptides.

Stability of the 15-mer-HLA-A*02:01 complexes

We then assessed the thermal stability, after refolding, of HLA-A*02:01 bound to four distinct 15-mer peptides and compared these values to HLA-A*02:01 bound to a canonical 9-mer epitope CMV-pp65-NLV [25]. The thermal melt point, or Tm, observed for HLA-A*02:01 in complex with the CMV-pp65-NLV peptide was 63.9°C (Table 2). We then performed the same assay with the four HLA-A*02:01-15-mer complexes, along with the HLA-A*02:01-NLV complex. The FLNKD and ALQDA peptides exhibited the lowest Tm, with a value of ~ 48°C, which was notably lower than the HLA-A*02:01-NLV complex. In contrast the ALWDI and KLLEI Tm were 58°C and 66.5°C respectively (Table 2). Interestingly the two 15-mer peptides with the lowest Tm have non-optimal HLA-A*02:01 anchor residue at PΩ, namely a valine residue (Table

Table 2. Thermal stability of peptide-HLA-A*02:01 complexes. Tm represents the temperature required to unfold 50% of the protein.

Peptide-HLA-A*02:01	Tm (°C)
FLNKDLEVDGHFVTM	47.9 ± 0.5
ALQDAGDSSRKEYFI	48.0 ± 1.0
KLLEIPDPDKNWATL	66.5 ± 1.8
ALWDIETGQQKTVFV	58.0 ± 1.0
NLVPMVATV	63.9 ± 0.5

1). In summary the HLA-A*02:01-restricted 15-mer epitopes can exhibit a range of Tm, with some of them highly stable in the cleft of HLA-A*02:01.

Crystal structures of 15-mer-HLA-A*02:01 complexes

To date only seven structures of HLA in complex with long epitopes (> 11 aa) are available [4, 7, 12, 13, 29, 30] as well as one structure of a rat MHC in complex with a 13-mer peptide [31]. The seven pHLA structures include: two 12-mer EBV epitopes bound to HLA-B*4403 [30] and to HLA-B*3508 [13]; a 13-mer EBV epitope in complex with closely related allomorphs HLA-B*3501 and HLA-B*3508 [12]; a 13-mer epitope bound to HLA-B*0702 [7], a self 14-mer peptide in complex with HLA-B*3501 [7, 29] and a self 16-mer peptide bound to HLA-B*4102 [4]. These structures solved to date reveal that the N and C termini of the peptides bind in similar fashion to the one observed for the classical length peptides, and that the

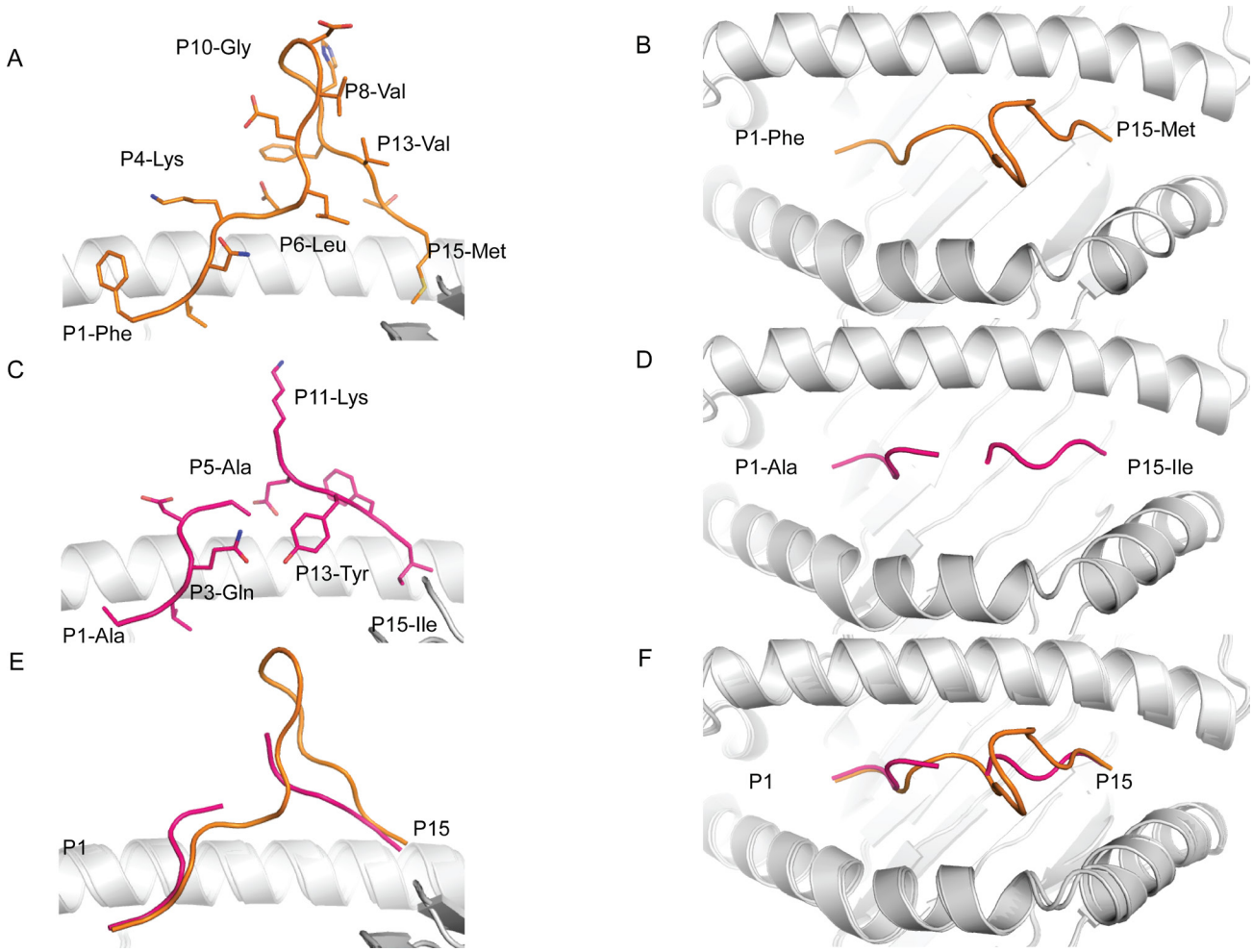


Figure 3. Crystal structures of 15-mer peptides in complex with HLA-A*02:01 molecule. Side view (A, C and E panels) and top-view (B, D and F panels) of the HLA-A*02:01 cleft (white cartoon) bind to the FLNKD peptide (orange stick and loop) or the ALQDA peptide (pink stick and loop). The bottom panels (E and F) show a superimposition of the two peptides in the cleft of HLA-A*02:01 molecule in the same orientation as the above panels.

central part of the peptide bulges out of the binding cleft.

Interestingly of the long epitopes characterized none of them were in complex with HLA-A molecule. In order to understand how the HLA-A*02:01 molecule can present long epitopes of 15 residues in length, we determined the structure of the HLA-A*02:01-FLNKD and HLA-A*02:01-ALQDA complexes at high resolution (Table 3). The two peptide-HLA complexes were crystallized in the same space group with the same unit cell dimension. Therefore, the difference in peptide structures was attributable to the peptide sequence. The two peptides bind with the canonical P2-Leu into the B pocket and with non-canonical PΩ-Val residues

Table 3. Data collection and refinement statistics

Data Collection Statistics	**HLA-A*02:01-FLNKD**	**HLA-A*02:01-ALQDA**
Resolution range (Å)	29.44 - 1.46 (1.51 - 1.46)	39.53 - 1.67 (1.73 - 1.67)
Space group	P 21	P 21
Cell Dimensions (a,b,c) (Å)	50.88, 79.76, 54.84 ß=111.75°	51.04, 79.06, 54.90 ß=111.11°
Total number of observations	392206 (37550)	264156 (24945)
Number of unique observations	69403 (6859)	46434 (4518)
Multiplicity	5.6 (5.5)	5.7 (5.5)
Data completeness (%)	99.79 (98.69)	99.71 (97.75)
I/σ(I)	19.3 (2.5)	17.8 (2.4)
aRpim (%)	2.5 (29.7)	3.1 (31.8)
Refinement Statistics	**HLA-A*02:01-FLNKD**	**HLA-A*02:01-ALQDA**
bRfactor (%)	16.12 (23.54)	15.49 (21.89)
bRfree (%)	19.04 (27.17)	19.78 (28.18)
Number of non-hydrogen atoms	4037	3859
Macromolecules	3472	3460
Ligands	15	27
Water	550	372
Rms deviations from ideality		
Bond lengths (Å)	0.009	0.008
Bond angles (°)	1.26	1.13
Ramachandran plot (%)		
Allowed region (%)	98	98
Generously allowed region (%)	2	2
Ramachandran outliers (%)	0	0
Average B-factor	21.5	23.7
Macromolecules	19.4	22.2
Ligands	42.4	59.7
Water	33.8	34.8

[a] $R_{p.i.m} = \Sigma_{hkl} [1/(N-1)]^{1/2} \Sigma_i | I_{hkl,i} - \langle I_{hkl} \rangle | / \Sigma_{hkl} \langle I_{hkl} \rangle$

[b] $R_{factor} = \Sigma_{hkl} || F_o | - | F_c || / \Sigma_{hkl} | F_o |$ for all data except ≈ 5% which were used for R_{free} calculation.

Values in parentheses are for the highest resolution shell.

in the F pocket for HLA-A*02:01, a methionine for FNLKD peptide and an isoleucine for the ALQDA peptide (Figures 2 A & B). The FNLKD peptide density was clear and unambiguous in the cleft of the HLA-A*02:01 molecule (Figures 2 A & C), while the central part of the ALQDA was poorly defined (Figures 2B & D). Despite the two 15-mer peptides exhibiting a similar Tm value to the same HLA-A*02:01 (Table 2), the conformation of the two peptides were notably different (Figure 3).

The ALQDA was mobile in the cleft of the HLA-A*02:01 molecule (Figure 2D), and as a result the central region from P6 to P10 was not built in the final model of the pHLA complex. Flexibility is often associated with long peptide presentation by HLA class I molecules, as exemplified by the 16 mer AEMY self-peptide presented by the HLA-B*41:03 molecule [4]. The ALQDA binds the HLA-A*02:01 molecule via 9 of its residues and forms 188 contacts with the HLA (9 salt-bridges and 16 hydrogen bonds). The number of bonds formed by the 15-mer ALQDA was similar to the 9-mer NLV peptide (185 contacts, 2 salt bridges and 14 hydrogen bonds), despite the extra 6 residues. A small amino acid such as valine is optimal at the C-terminal part of the peptide sequence as it fits well in the F pocket of the HLA-A*0201 cleft. As observed in the NLV peptide structure (PDB code: 3GSO(25)), whereby the P9-Val sat on the top of Tyrosine 116 of the HLA-A*0201 molecule. The change to larger amino acids, such as methionine or isoleucine, at the C-terminal position of the peptide leads to rotation of the Tyrosine 116 to avoid steric clashes that pushes the Arginine 97. This rearrangement of buried amino acids within the antigen binding cleft appears to be less favourable to the overall stability of the pHLA-A*0201 complex (Table 2).

Contrasting the flexible ALQDA 15-mer, the FNLKD peptide was well defined in the cleft of the HLA-A*02:01 molecule (Figure 2C), and is the longest well-defined epitope observed in complex with a HLA class I molecule to date. The FNLKD bulges out of the HLA-A*02:01 cleft forming a ß-sheet hairpin structure from P7 to P12 residues (Figure 3A). The secondary structure formation in the bulged part of the peptide made intra-molecular contacts, constraining and rigidifying the peptide [33] and probably explains how the FNLKD can be such a long peptide and being so rigid in HLA-A*02:01 cleft. Interestingly the ß-sheet hairpin formation is higher than the hinge of the α2-helix of the HLA-A*02:01 and would represent an immediate contact point for T cell receptor. The stable conformation of the FLNKD was also associated with a higher number of contacts with the HLA-A*02:01 molecule, with the peptide engaging 11 of its residues to interact with the HLA, and making a total of 215 contacts (6 salt bridges and 14 hydrogen bonds). This β-sheet hairpin structure is the first reported for an epitope bound to class I HLA. An α-helix has been previously reported in the 12-mer CPS bound to the HLA-B*3508 complex [13].

The crystal structures of the two 15-mers in complex with HLA-A*02:01 show that,like the HLA-B molecule, HLA-A can present long peptides in a diverse array of conformations from mobile to highly stable,

and could represent some new antigen for T-cells.

15-mer epitopes presented by HLA-A*02:01 can activate CD8$^+$ T-cells

We demonstrate that HLA-A*02:01 can bind longer peptides with high binding affinity, forming stable pHLA complexes, and determined how HLA-A*02:01 can present 15-mer peptides. In order to establish if 15-mer-HLA*0201 complexes were antigenic and recognized by T-cells, we used pHLA-tetramers with the two structurally characterized 15-mer peptides FLNKD and ALQDA (RAB9AFLNKD:HLA-A*02:01 and GYPCALQDA:HLA-A*02:01, respectively). Since negative selection during thymic development depletes

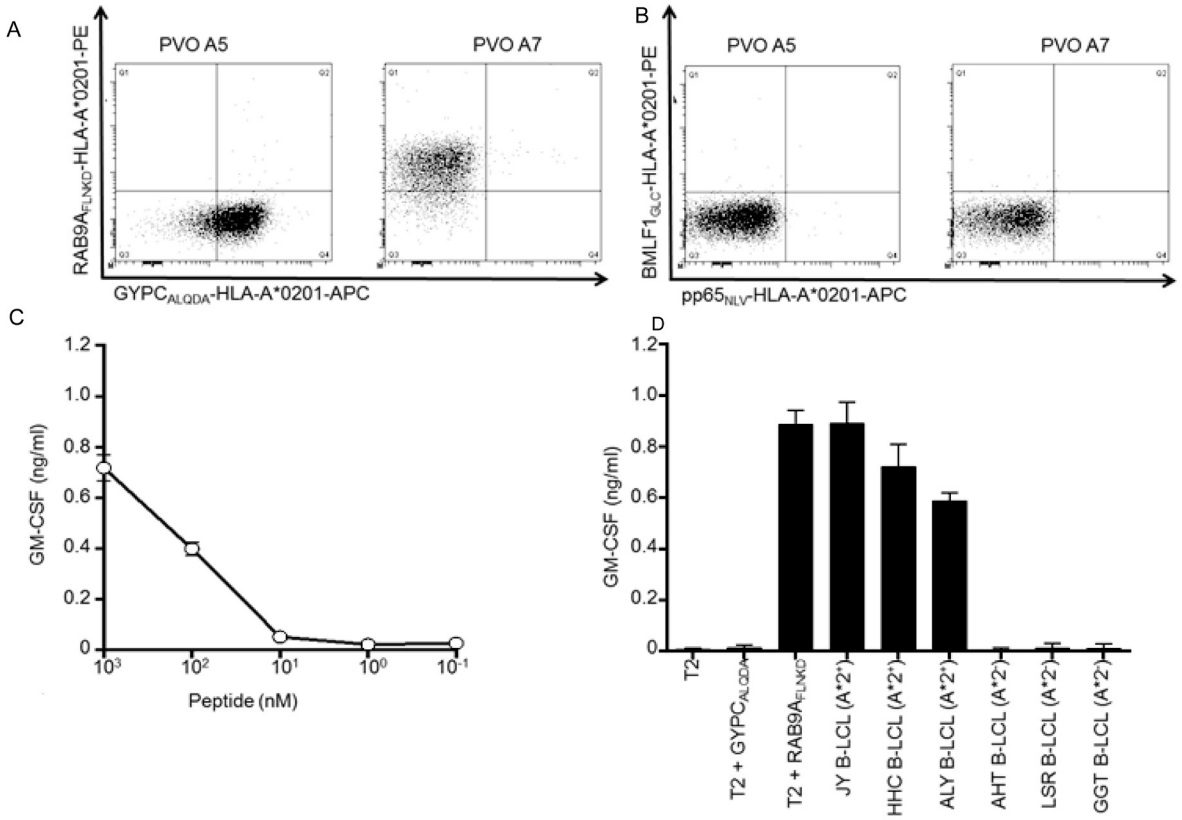


Figure 4. Specific T-cell clone recognition of 15-mer peptide FLNKD presented by HLA-A*02:01
T cell clone PVO A5 and PVO A7 were isolated using pHLA-tetramers composed of GYPC-derived peptide ALQDA or RAB9A-derived peptide FLNKD bound to HLA-A*02:01 (GYPCALQDA:HLA-A*02:01 or RAB9AFLNKD:HLA-A*02:01)from an HLA- A*02:01-negative individual. (A) T cell clone PVO A5 (left) and PVO A7 (right) specifically bound to GYPCALQDA:HLA-A*02:01-tetramer andRAB9AFLNKD:HLA-A*02:01-tetramer, respectively, after staining with PE-labelled RAB9AFLNKD:HLA-A*02:01 and APC-labelled GYPCALQDA:HLA-A*02:01-tetramer. (B) The panel shows the lack of staining with two control tetramers composed of HLA-A*02:01 displaying two virus-derived epitopes for T cell clone PVO A5 (left) and PVO A7 (right). All plots are shown with bi-exponential axis. Numbers in corners indicate percent cells in each quadrant. (C) GM-CSF production was measured after co-culturing T-cell clone PVO A7 with HLA-A*02:01-positive T2 cells loaded with decreasing concentration of FLNKD peptide. (D) GM-CSF production was measured after co-culturing T cell clone PVO A7 with T2 cells, T2 cells loaded with 500 nM peptide (+GYPCALQDA or +RAB9AFLNKD), three HLA-A*02:01-positive B-LCLs (A2+) which naturally express and present RAB9A or three HLA-A*02:01-negative B-LCLs (A2-). Experiments were performed in triplicate. Shown is one representative experiment of two independent experiments. Error bars indicate standard deviation.

T-cells recognizing such self-antigens bound to self-HLA, T-cells were isolated from HLA-A*02:01-negative healthy individuals which contain a naïve T-cell repertoire capable of recognizing such self-antigens presented in HLA-A*02:01. From PBMCs of HLA-A*02:01-negative individuals, $CD8^+$ T-cells clones were expanded that bind pHLA-tetramers RAB9AFLNKD:HLA-A*02:01 and GYPCALQDA:HLA-A*02:01 by first enriching pHLA-tetramer binding cells by MACS followed by immediate single-cell FACS sorting. Among the isolated T-cells, clone PVO A5 showed specific binding of GYPCALQDA:HLA-A*02:01-tetramer while binding to RAB9AFLNKD:HLA-A*02:01-tetramer was absent (Figure 4A). In contrast, T-cell clone PVO A7 specifically bound to tetramer RAB9AFLNKD:HLA-A*02:01 while binding to GYPCALQDA:HLA-A*02:01 was absent (Figure 4A). In addition, both T-cell clones did not bind to two control tetramers composed of HLA-A*02:01 displaying either EBV-derived peptide GLCTLVAML or human cytomegalovirus (CMV)-derived peptide NLVPMVATV, further indicating specific binding of both T-cell clones to their respective pHLA-tetramer displaying a 15-mer peptide (Figure 4B). Next, peptide-dependent recognition for both T-cell clones was assessed by pulsing HLA-A*02:01-positive T2 cells with the two 15-mer peptides. GYPCALQDA-specific T-cell clone PVO A5 did not recognize peptide pulsed T2 cells indicating insufficient sensitivity for HLA-bound ALQDA (data not shown). In contrast, T-cell clone PVO A7 recognized T2 cells pulsed with peptide FLNKD (Figure 4C). This recognition was peptide-specific since no recognition of T2 cells pulsed with ALQDA was observed (Figure 4D). In addition, T-cell clone PVO A7 recognized three HLA-A*02:01-positive B-LCLs which naturally express RAB9A and were used to elute the 15-mer peptide FLNKD (Figure 4D). Lack of recognition of three HLA-A*02:01-negative B-LCLs indicates that the observed reactivity of T-cell clone PVO A7 was HLA-A*02:01-dependent (Figure 4D). These data indicate that 15-mer peptide FLNKD presented in HLA-A*02:01 on the cell surface can be recognized by T-cells in a peptide-dependent manner.

DISCUSSION

Classically, HLA class I molecules present 8-11-mer peptides, although, an expanding list of lengthy (>11 aa) HLA-restricted peptides have emerged [6]. Crystallographic studies have reported on seven pHLAs structures involving a 12-mer to 16-mer epitopes [4, 7, 12, 13, 29, 30]. These previous studies were all focused on HLA-B molecules, and here we describe the ability of HLA-A*02:01 molecule to bind long epitopes too, with 538 12-14-mers being defined. Further, 77 peptides are listed of 15-23 amino acids long that fulfill both the P2 and PΩ anchors criteria in either HLA-A*02:01, HLA-B*0702 or HLA-B*44. A comparable percentage of longer peptides was found in the reprocessed data of Mommen et al [9] , in particular in HLA-A*0301 and HLA-B*0702, and to a lesser extent in HLA-A*0101 and HLA-B*2705. The listing of peptides shows that HLA-A molecules appear to be just as suitable for presenting longer peptides as the HLA-B alleles. Both the intensity and the hydrophobicity of the longer peptides resemble that of the canonical 8 to 11-mer peptides. There was a steady decline in the number of longer peptides for every additional amino acid, which probably represents the probability of a peptide to survive in the cellular proteolytic environment. Longer peptides have an increasing chance of being cut by a protease. Of note, the amino acids between the P2 and PΩ anchors are not generally enriched for specific amino acid residues, so, the amino acid stretch between the anchors does not seem to be specifically resistant to proteolytic degradation on the basis of its primary structure. Longer peptides can be translocated by TAP into the ER, although generally somewhat less efficiently [34]. In the ER, peptides are protected from being trimmed to short peptides for presentation in HLA by the nature of ERAAP [35]. The fact that there seems to be no clear-cut length limitation (on the long side) imposed by the HLA class I binding groove, can be explained by the phenomenon of (super)bulging of the peptide, with the P2 and PΩ anchors residues position fixed in the peptide binding groove, but freedom to ‘leave’ the binding groove for (part) of the peptide between these anchors residues. The two refolding assays we employed, both showed that the behavior of the 15-mer peptides resembled that of the canonical 8-11-mer peptides, i.e the 15-mer peptides just as easily formed pHLA complexes and competed to the same degree as known good binders of 8-11-mer length. From the two 15-mer peptides solved in complex with the HLA-A*02:01, we observed two different conformations of the long epitopes in the cleft of HLA-A*02:01. Firstly the ALQDA peptide was highly mobile, and its central region was poorly defined, reminiscent of the 16-mer self-peptide observed in complex with the HLA-B*4103 [4]. Contrasting with the high flexibility of the ALQDA, the FLNKD was well defined and adopted one single rigid conformation when bound to HLA-A*02:01, similar to the 13-mer EBV epitope in complex with HLA-B*3508 [12]. The FLNKD peptide central region formed a ß-hairpin secondary structure that bulged out of the HLA-A*02:01 cleft, and could be a potential contact point for the FLNKD-specific T-cells, and it will be of high interest to know how T-cells can engage a highly rigid bulge

peptide like the FLNKD epitope. The TCR could potentially "struggle" to bind it or it will mostly focus on the peptide (like SB27,(15)) or might bind on the side of the peptide.

T-cells could be isolated from HLA-A*02:01-negative healthy individuals that contain a naïve T-cell repertoire capable of recognizing self-antigens presented in HLA-A*02:01. T-cell clones demonstrated specific binding of pHLA-tetramer and furthermore, peptide-dependent recognition was observed for selected T-cell clones with HLA-A*02:01-positive T2 cells pulsed with the peptides as well as recognition of endogenously processed peptide on HLA-A*02:01-positive B-LCLs. Clone PVO A5 lacks functional reactivity against peptide-loaded target cells although there is specific staining of that clone with GYPCALQDA:A2 pHLA-tetramer. We have previously demonstrated that pHLA-tetramer staining alone is a poor indicator of functional avidity of a T-cell clone (18). Therefore, it is most likely that clone PVO A5's avidity for HLA bound peptide GYPCALQDA is insufficient to trigger T-cell activity while binding of pHLA-tetramer is possible. To circumvent the depletion of high avidity T cells targeting self-peptides presented in self-HLA during thymic development, pHLA-tetramer binding T cells were isolated from a healthy HLA-A2-negative individual. Based on, previous results we estimated to isolate both high as well as low affinity T cells (18). The results demonstrate that PVO A5 represents a low avidity T cell clone for GYPCALQDA:A2 while clone PVO A7 represents a high avidity T cell clone specific for RAB9AFLNKD:A2.

Clone PVO A7 demonstrated peptide-dependent recognition of peptide-loaded T2 stimulator cells in the nanomolar range. Virus-specific T cells demonstrate peptide sensitivity as low as in the picomolar range. However, caution must be exerted when comparing sensitivity between T cell clones recognizing different epitopes based solely on recognition of peptide-loaded stimulator cells. Not only affinity of the TCR for its cognate peptide is important but also binding properties of the peptide to its respective HLA-molecule is critical, since exogenously loaded peptide need to compete with already HLA-bound peptide. These properties can differ between peptides. Furthermore, PVO A7 is able to recognize endogenously processed and presented peptide on three HLA-A2-positive B-LCL indicating high functional avidity comparable to virus-specific T cells. From these findings, T-cells appear to be capable to specifically recognize longer peptides. So far, there seems to be no clear limitation on peptide length for T-cell recognition of HLA class I presented peptides.

Altogether our data show that HLA class I restricted presentation and recognition is less restrictive than previously anticipated. Our data expand our understanding of HLA class I ligand presentation, and show that longer peptides are regular members of the HLA-ligandome, and should not be discarded in epitope discovery experiments, since these peptides might be useful in immunotherapy. Furthermore, the non-canonical peptide sequences presented here provide insight in antigen presentation and antigen processing.

ACKNOWLEDGEMENTS

This research was made possible by the financial assistance of the Landsteiner Foundation for Blood Transfusion Research (LSBR0713). We thank N. Dolezal, R. Cordfunke and W. Benckhuijsen for peptide synthesis. We thank Kristy Campbell and Nathan Croft for their technical assistance; the staff at the Australian Synchrotron for assistance with data collection. JR by a NHMRC Australia Fellowship and SG is supported by an ARC Future Fellowship (FT120100416). This work was supported by the Australian Research Council and the National Health and Medical Research Council of Australia.

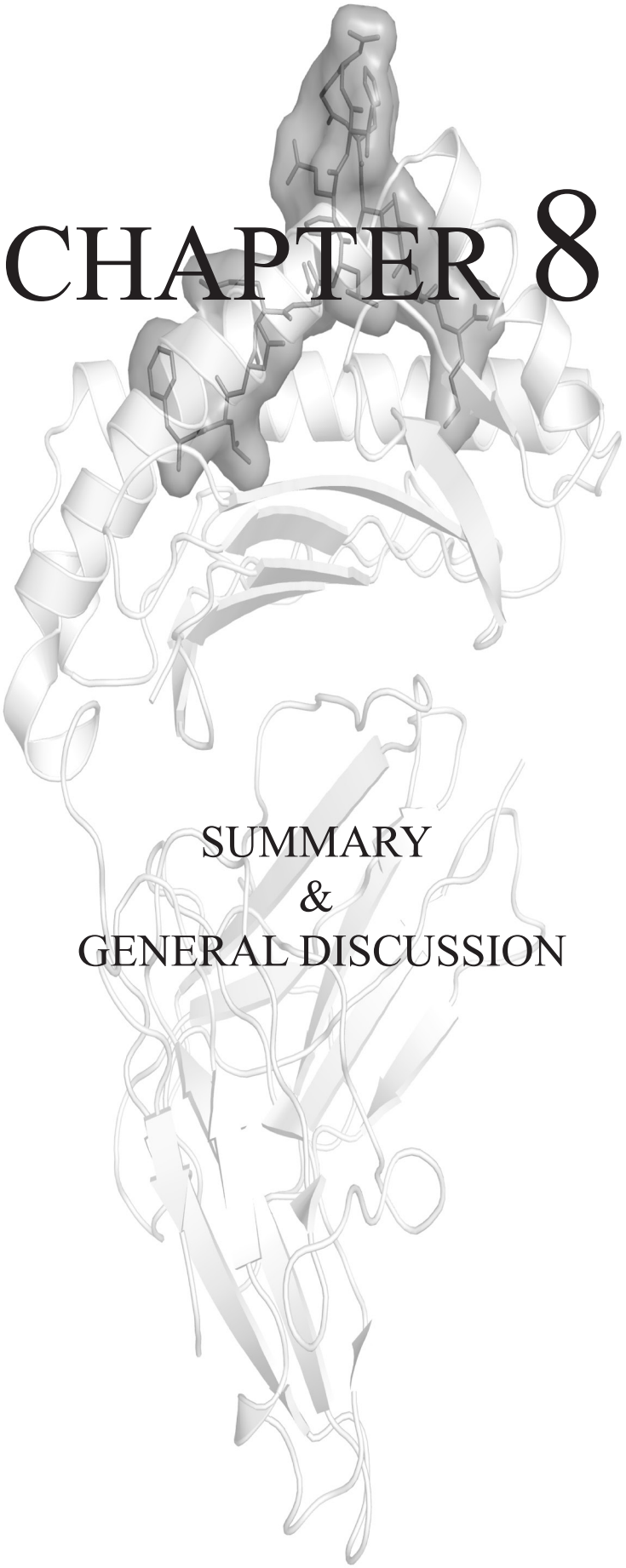
PDB deposition. The coordinates have been deposited in the PDB: HLA-A*02:01-FLNKD: 4U6Y and HLA-A*02:01-ALQDA: 4U6X.

Abbreviations: pHLA, peptide-Human Leukocyte Antigen; pMHC, peptide-Major Histocompatibility Complex; aa, amino acid.

REFERENCES

1. Neefjes, J., Jongsma, M.L., Paul, P. and Bakke, O. (2011) Towards a systems understanding of MHC class I and MHC class II antigen presentation. Nat Rev Immunol, 11, 823-836.
2. Rammensee, H.G., Falk, K. and Rotzschke, O. (1993) Peptides naturally presented by MHC class I molecules. Annu Rev Immunol, 11, 213-244.
3. Yewdell, J.W., Reits, E. and Neefjes, J. (2003) Making sense of mass destruction: quantitating MHC class I antigen presentation. Nat Rev Immunol, 3, 952-961.
4. Bade-Doding, C., Theodossis, A., Gras, S., Kjer-Nielsen, L., Eiz-Vesper, B., Seltsam, A., Huyton, T., Rossjohn, J., McCluskey, J. and Blasczyk, R. (2011) The impact of human leukocyte antigen (HLA) micropolymorphism on ligand specificity within the HLA-B*41 allotypic family. Haematologica, 96, 110-118.
5. Ben Dror, L., Barnea, E., Beer, I., Mann, M. and Admon, A. (2010) The HLA-B*2705 peptidome. Arthritis Rheum, 62, 420-429.
6. Burrows, S.R., Rossjohn, J. and McCluskey, J. (2006) Have we cut ourselves too short in mapping CTL epitopes? Trends Immunol, 27, 11-16.
7. Ebert, L.M., Liu, Y.C., Clements, C.S., Robson, N.C., Jackson, H.M., Markby, J.L., Dimopoulos, N., Tan, B.S., Luescher, I.F., Davis, I.D. et al. (2009) A long, naturally presented immunodominant epitope from NY-ESO-1 tumor antigen: implications for cancer vaccine design. Cancer research, 69, 1046-1054.
8. Hassan, C., Kester, M.G., Ru, A.H., Hombrink, P., Drijfhout, J.W., Nijveen, H., Leunissen, J.A., Heemskerk, M.H., Falkenburg, J.H. and Veelen, P.A. (2013) The human leukocyte antigen-presented ligandome of B lymphocytes. Molecular & Cellular Proteomics, 12, 1829-1843.
9. Mommen, G.P., Frese, C.K., Meiring, H.D., van Gaans-van den Brink, J., de Jong, A.P., van Els, C.A. and Heck, A.J. (2014) Expanding the detectable HLA peptide repertoire using electron-transfer/higher-energy collision dissociation (EThcD). Proceedings of the National Academy of Sciences of the United States of America, 111, 4507-4512.
10. Probst-Kepper, M., Stroobant, V., Kridel, R., Gaugler, B., Landry, C., Brasseur, F., Cosyns, J.P., Weynand, B., Boon, T. and Van Den Eynde, B.J. (2001) An alternative open reading frame of the human macrophage colony-stimulating factor gene is independently translated and codes for an antigenic peptide of 14 amino acids recognized by tumor-infiltrating CD8 T lymphocytes. The Journal of experimental medicine, 193, 1189-1198.
11. Scull, K.E., Dudek, N.L., Corbett, A.J., Ramarathinam, S.H., Gorasia, D.G., Williamson, N.A. and Purcell, A.W. (2012) Secreted HLA recapitulates the immunopeptidome and allows in-depth coverage of HLA A*02:01 ligands. Mol Immunol, 51, 136-142.
12. Tynan, F.E., Borg, N.A., Miles, J.J., Beddoe, T., El-Hassen, D., Silins, S.L., van Zuylen, W.J., Purcell, A.W., Kjer-Nielsen, L., McCluskey, J. et al. (2005) High resolution structures of highly bulged viral epitopes bound to major histocompatibility complex class I. Implications for T-cell receptor engagement and T-cell immunodominance. J Biol Chem, 280, 23900-23909.
13. Wynn, K.K., Fulton, Z., Cooper, L., Silins, S.L., Gras, S., Archbold, J.K., Tynan, F.E., Miles, J.J., McCluskey, J., Burrows, S.R. et al. (2008) Impact of clonal competition for peptide-MHC complexes on the CD8+ T-cell repertoire selection in a persistent viral infection. Blood, 111, 4283-4292.
14. Liu, Y.C., Miles, J.J., Neller, M.A., Gostick, E., Price, D.A., Purcell, A.W., McCluskey, J., Burrows, S.R., Rossjohn, J. and Gras, S. (2013) Highly divergent T-cell receptor binding modes underlie specific recognition of a bulged viral peptide bound to a human leukocyte antigen class I molecule. J Biol Chem, 288, 15442-15454.
15. Tynan, F.E., Burrows, S.R., Buckle, A.M., Clements, C.S., Borg, N.A., Miles, J.J., Beddoe, T., Whisstock, J.C., Wilce, M.C., Silins, S.L. et al. (2005) T cell receptor recognition of a 'super-bulged' major histocompatibility complex class I-bound peptide. Nat Immunol, 6, 1114-1122.
16. Liu, Y.C., Chen, Z., Burrows, S.R., Purcell, A.W., McCluskey, J., Rossjohn, J. and Gras, S. (2012) The energetic basis underpinning T-cell teceptor tecognition of a super-bulged peptide bound to a major histocompatibility complex class I molecule. J Biol Chem, 287, 12267-12276.
17. Bell, M.J., Burrows, J.M., Brennan, R., Miles, J.J., Tellam, J., McCluskey, J., Rossjohn, J., Khanna, R. and Burrows, S.R. (2009) The peptide length specificity of some HLA class I alleles is very broad and includes peptides of up to 25 amino acids in length. Mol Immunol, 46, 1911-1917.

18. Hombrink, P., Hassan, C., Kester, M.G., de Ru, A.H., van Bergen, C.A., Nijveen, H., Drijfhout, J.W., Falkenburg, J.H., Heemskerk, M.H. and van Veelen, P.A. (2013) Discovery of T cell epitopes implementing HLA-peptidomics into a reverse immunology approach. J.Immunol., 190, 3869-3877.
19. Tan, T.L., Geluk, A., Toebes, M., Ottenhoff, T.H. and Drijfhout, J.W. (1997) A novel, highly efficient peptide-HLA class I binding assay using unfolded heavy chain molecules: identification of HIV-1 derived peptides that bind to HLA-A*0201 and HLA-A*0301. J Immunol Methods, 205, 201-209.
20. Gras, S., Wilmann, P.G., Chen, Z., Halim, H., Liu, Y.C., Kjer-Nielsen, L., Purcell, A.W., Burrows, S.R., McCluskey, J. and Rossjohn, J. (2012) A structural basis for varied alphabeta TCR usage against an immunodominant EBV antigen restricted to a HLA-B8 molecule. J Immunol, 188, 311-321.
21. Kabsch, W. (2010) Xds. Acta Crystallogr D Biol Crystallogr, 66, 125-132.
22. Evans, P. (2006) Scaling and assessment of data quality. Acta Crystallogr D Biol Crystallogr, 62, 72-82.
23. (1994) The CCP4 suite: programs for protein crystallography. Acta crystallographica. Section D, Biological crystallography, 50, 760-763.
24. Read, R.J. (2001) Pushing the boundaries of molecular replacement with maximum likelihood. Acta Crystallogr D Biol Crystallogr, 57, 1373-1382.
25. Gras, S., Saulquin, X., Reiser, J.B., Debeaupuis, E., Echasserieau, K., Kissenpfennig, A., Legoux, F., Chouquet, A., Le Gorrec, M., Machillot, P. et al. (2009) Structural bases for the affinity-driven selection of a public TCR against a dominant human cytomegalovirus epitope. Journal of immunology (Baltimore, Md. : 1950), 183, 430-437.
26. Emsley, P., Lohkamp, B., Scott, W.G. and Cowtan, K. (2010) Features and development of Coot. Acta Crystallogr D Biol Crystallogr, 66, 486-501.
27. Adams, P.D., Afonine, P.V., Bunkoczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W. et al. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr D Biol Crystallogr, 66, 213-221.
28. DeLano, W.L. (2002) The PyMOL Molecular Graphic System. DeLano Scientific.
29. Bourdetsky, D., Schmelzer, C. E., and Admon, A. (2014) The nature and extent of contributions by defective ribosome products to the HLA peptidome. Proc Natl Acad Sci U S A111, E1591-1599
30. Probst-Kepper, M., Hecht, H.J., Herrmann, H., Janke, V., Ocklenburg, F., Klempnauer, J., van den Eynde, B.J. and Weiss, S. (2004) Conformational restraints and flexibility of 14-meric peptides in complex with HLA-B*3501. Journal of immunology (Baltimore, Md. : 1950), 173, 5610-5616.
31. Rist, M.J., Theodossis, A., Croft, N.P., Neller, M.A., Welland, A., Chen, Z., Sullivan, L.C., Burrows, J.M., Miles, J.J., Brennan, R.M. et al. (2013) HLA peptide length preferences control CD8+ T cell responses. J Immunol, 191, 561-571.
32. Speir, J.A., Stevens, J., Joly, E., Butcher, G.W. and Wilson, I.A. (2001) Two different, highly exposed, bulged structures for an unusually long peptide bound to rat MHC class I RT1-Aa. Immunity, 14, 81-92.
33. Theodossis, A., Guillonneau, C., Welland, A., Ely, L.K., Clements, C.S., Williamson, N.A., Webb, A.I., Wilce, J.A., Mulder, R.J., Dunstone, M.A. et al. (2010) Constraints within major histocompatibility complex class I restricted peptides: presentation and consequences for T-cell recognition. Proc Natl Acad Sci U S A, 107, 5534-5539.
34. Koopmann, J.O., Post, M., Neefjes, J.J., Hammerling, G.J. and Momburg, F. (1996) Translocation of long peptides by transporters associated with antigen processing (TAP). Eur J Immunol, 26, 1720-1728.
35. Serwold, T., Gaw, S. and Shastri, N. (2001) ER aminopeptidases generate a unique pool of peptides for MHC class I molecules. Nat Immunol, 2, 644-651.



CHAPTER 8

SUMMARY & GENERAL DISCUSSION

SUMMARY

Minor histocompatibility antigens (MiHA) are naturally processed polymorphic peptides originated from endogenous proteins which are presented in human leukocyte antigen (HLA) molecules on the cell surface to be recognized by T cells from HLA-matched individuals. The involvement of MiHA in the graft versus tumor is described about three decades ago. Nowadays, there are many evidences that, after HLA-matched allogeneic hematopoietic stem cell transplantation (HSCT) MiHA specific T cells mediate both graft versus leukemia (GVL) reactivity and graft versus host disease (GVHD), which is a major cause of morbidity and mortality. To allow skewing of a donor derived T cell response against MiHA towards GVL while preventing GVHD, the identification of many MiHA is essential. So far, approximately 50 MiHA have been identified using different approaches. The two main approaches used for identification of MiHA are forward and reverse approaches. The forward approach can only be explored if there is indication of a GVL and/or GVHD response in a patient received allogeneic HSCT and the relevant T cells can be isolated and characterized. A reverse approach uses algorithms and available databases to predict novel MiHA. The major limitation of reverse approach is that most predicted MiHA might not be presented on the cell surface and that post-translationally modified peptides cannot be predicted. To avoid these limitations and to improve the outcome of the reverse approach we decided to establish a new strategy based on the *bone fide* eluted ligandome. **In Chapter 2** we applied our new strategy by identifying the peptide content of HLA molecules from two B-lymphoblastoid cell lines (B-LCL). We built a HLA-class I ligandome library from the obtained data. Since the dynamic range of the peptides eluted from HLA-class I molecules is very high, it is not possible to identify as many ligands as possible using single dimension separation techniques. Therefore, we applied multidimensional chromatography. Three different first dimension separation techniques were used including peptide isoelectric focusing (IEF), strong cation exchange chromatography (SCX) and RP-C18 chromatography. The fractions obtained from these three first dimension separation techniques were analysed using nano-RP-C18 liquid chromatography coupled to mass spectrometry (nano-LC-MS/MS). The results have shown that application of different first dimension separation techniques leads to increased numbers of identified HLA-ligands. Our new strategy applied to reverse immunology resulted in the identification of at least 16,000 ligands with high mascot ion score (BMI>35) from HLA-A2, HLA-B7 and HLA-B44 including the polymorphic and post translationally modified ligands. This huge number of identified HLA-ligands formed the basis for the next steps of our study. To validate the identification of these peptides we used different methods including synthesis of selected peptides and comparison of the MS/MS spectra of the synthetic peptides with the counterpart eluted peptides. In addition, we checked different physicochemical properties of the peptides like the binding affinity to HLA

molecules using the NetMHC algorithm, length and hydrophobicity. After validation of these peptides we looked in depth into the characteristics of HLA-ligands and antigen processing. To improve the identification of polymorphic peptides/MiHA, in **Chapter 3** we developed a database dedicated for identification of polymorphic peptides called human short peptide variation database (HSPVdb). This database is based on the human mRNA sequences from the well-annotated reference sequence (RefSeq) database and associated variation information derived from the Single Nucleotide Polymorphism Database (dbSNP). Investigation of the frequencies of SNPs in the dbSNP revealed that many SNPs are non-polymorphic "SNPs". Therefore, we removed those from our dedicated database, and this resulted in a comprehensive high quality database. We identified 1500 polymorphic peptides with high mascot ion score (BMI>35) using the developed database. To identify potential MiHA among the huge number of polymorphic peptides eluted from HLA-A*0201 or HLA-B*0702 in **Chapter 4** we selected the top 25 MiHA candidates identified in chapter 2. From this set of data we could validate 2 novel potential MiHA being (LB-CLYBL-1Y & LB-TEP1-1S). Since we knew that we did not obtain a comprehensive view of the ligandome yet, although the number of ligands we found was already impressive, we decided to study all aspects/steps of the work-up process, to find out at which stages the greatest losses occurred. **In Chapter 5** we developed a method to be able to quantitate MiHA on the cell surface. We used stable isotope labeled peptides coupled to recombinant HLA-class I molecule (hpMHC), spiked in the first step of the purification method to account for all losses in the procedure. In addition, for determination of the recovery of the purification method, we added medium isotope labeled peptide to the fractions prior LC-SRM or LC-PRM analysis. The previously identified MiHA LB-NISCH-1A and LB-SSR1-1S were used to determine the recovery of the purification method and to determine the copy number of MiHA. Major losses, from 90-99%, were determined to be in the immunopurification step. Since the developed method is very accurate in quantitation of MiHA on the cell surface, **in Chapter 6** we studied the phenomenon of the unidirectional T cell responses against MiHA. The unidirectional phenomenon means that T cells can recognize only the MiHA, but not their allelic counterpart (AC). The absence of immunogenicity against AC may be explained by the absence of AC on the cell surface. However in the heterozygous cells presenting both MiHA and AC this explanation is not valid anymore, as we have found a couple of MiHA and their AC in the data obtained in chapter 2. Therefore, we hypothesized that the expression of MiHA should be a factor 10 higher on the cell surface in comparison with their AC to be able to elicit T cell response. To investigate our hypothesis we selected three MiHA and their AC: LB-NISCH-1 (MiHA/AC; A/V), LB-SSR1-1 (MiHA/AC; S/L) and LB-CLYBL-1 (MiHA/AC; Y/D). We used the approach developed in chapter 5 to determine the copy number of both the MiHA and their AC on the cell surface. The results showed that the expressions of both MiHA and their AC were similar. Therefore, the unidirectional response of T cells against MiHA cannot be explained by

the higher copy number of MiHA on the cell surface than of the AC. Other immunological reactions might be involved in eliciting T cell responses that need to be investigated. The high number of identified HLA-class I ligands from chapter 2 include a high number of non- canonical long peptides. Therefore, **in Chapter 7** we studied in depth the presentation of non-canonical long peptides in HLA-A*0201 molecules. From the elution data described in chapter 2 we could find 1,568 HLA-ligands with lengths of 12-23 amino acids and BMI ≥ 35, of which 423 peptides had a length of 15-23 amino acids. To study in depth the binding affinity of these peptides to HLA-A*0201 we selected eight 15 mer peptides, having binding motifs for HLA-A*0201 molecule. We used two binding assays: a competition refolding assay,and a pHLA refolding efficiency. Both assays showed high binding affinity of the 15 mer similar to the binding affinity of the 9 mer peptides to HLA-A*0201 molecule. Furthermore, we determined the thermal stability of the 15-mer pHLA complexes, and also in this assay the 15-mer and 9-mer peptides showed similar thermal stability. Moreover, we showed that the 15-mer epitopes presented by HLA-A*02:01 can activate $CD8^+$ T-cells. Finally, we used crystallography technique to determine the structure of the pHLA complexes and the data showed that the 15 mer peptides are bulged out from the HLA-A*02:01 molecules.

GENERAL DISCUSSION& PERSPECTIVES

The identification of HLA class I ligands and MiHA

MiHA and other tumor-specific T cell epitopes are important candidates for immunotherapy. In the introduction of this thesis we described the main strategies that have been used in the identification of these peptides. In this thesis we developed and used an alternative strategy to identify MiHA which are presented in HLA class I molecules. The data obtained from this new strategy were the basis of our ligandome database as identified by peptidomics. The developed strategies in this thesis can be applied to other tumor-specific T cell epitopes.

Previously, the number of different HLA class I presented peptides per cell has been estimated at over 10,000 different molecules per cell (reviewed in [1]). Although different groups have used different approaches to identify these peptides, the number of the reported peptides was very limited [2-12]. Since the last two years the large scale analysis of HLA-ligandome increased the number of reported HLA-class I associated peptides to about 50,000 in three different studies including the data presented in chapter 2 of this thesis [13-15]. The number of identified post translationally modified peptides and polymorphic peptides were also limited. The number of MiHA identified so far is approximately 50 [16-46] and recently reviewed in [47]. MiHA can be elucidated by forward and reverse immunology approaches. The forward approach is based on the isolation of a T cell clone after induction of an immunological reaction in a patient who underwent allogeneic hematopoietic stem cell transplantation (HSCT). Reverse immunology approaches were based on algorithms, e.g. for HLA-binding and proteasome processing, to predict potential MiHA and subsequently the search for and the selection of T cells. The advantage of this approach is that many MiHA candidates can be selected, while the major limitation of this strategy is that most candidate MiHA will not be presented on the cell surface, as most candidates will not 'survive' antigen processing [48]. In addition, the available algorithms do not predict post-translationally modified antigens.

To avoid the major limitation of the classical reverse approach in chapter 2 we designed a strategy which is based on the *bona fide* eluted ligandome, as identified by mass spectrometry, which means that peptide presentation in HLA is guaranteed, before subsequent experiments are performed. Furthermore, post-translationally modified peptides and polymorphic peptides can be identified with this ligandomics approach. The data can also be used to learn more about antigen processing and general features of the ligandome, which can be used to build an immunopeptidome atlas [49], and to improve the software packages used in the prediction of HLA-class I ligands (e.g. binding affinity, length, hydrophobicity..etc.) [50-54]. The post translationally modified ligands are getting more interest, especially phosphorylated ligands, that have been associated with deregulated signalling and cancer [55-58]. Furthermore, the binding of non-canonical long

peptides to HLA-class I molecules has emerged from the eluted peptide data using the peptidomics approach (chapter 7).

Furthermore, using the peptides library to select peptides specific for the B cell receptor (BCR) CD79b gene to study the benefits and limitations of these peptides in the treatment of B-cell malignancies or hematological malignancies has illustrated the benefit of this approach [59]. Altogether, the results and the wide application of the peptidomics approach show the benefits of this new strategy in the identification of relevant HLA-ligands.

Despite the high number of HLA-ligands identified (>10,000-20,000) using our peptidomics approach, the peptide recovery was found to be remarkably low, ranging from <1 to 5%, estimated from the whole ligandome study (chapter 2) and the heavy peptide-MHC approach (chapter 5). The low recovery presents a serious limitation of the peptidomics approach and should be addressed. Our results show that the main loss of the peptides occurs in the immunoaffinity step (chapter 5).Therefore, more work has to be done to improve the recovery of this step. For instance, utilization of CNBr activated beads instead of protein-A coupled to CL4B sepharose beads has been reported to reduce the loss of peptides during immunoaffinity isolation of HLA associated ligands [58,60], although our experiments on lower cell numbers do not confirm these findings. Additional issues are the dimensions of the immunoaffinity column and leaving out the ultra-filtration step, and perform a C18-chromatography separation of proteins and peptides instead, as was applied by Illing et al., [61] might increase the recovery of the method, depending on the number of cells used during sample pretreatment. The validation of all HLA-ligands obtained from large scale peptidomics experiments, by comparison of the MS/MS spectra of the eluted peptides with the synthetic counterparts, is not easy due to the huge number of identified peptides. In addition, selection of peptides according to the low false positive percentage (1% FDR), as it is commonly applied in shotgun proteomics [62,63], was found impossible in combination with a low mass accuracy MS2 instrument, such as ion traps. The enormously increased database search space, due to the undefined peptide termini of HLA-ligands and the limited MS2 accuracy, led us to synthesize interesting candidate peptides, such as e.g. CMV epitopes and previously identified MiHA, and compare the MS2 spectra of the eluted peptides with the MS2 spectra of their synthetic counterparts. The results of this comparison yielded a mascot ion score cut-off value of 35 as a good working number, which was further applied. Further comparison of our peptide listings with very recently produced lists obtained on a high mass accuracy instrument [14], confirmed that a cut-off value of 35 is a good number. However, for immunological applications it is advisable to include also low scoring peptides and validate these by peptide synthesis and MS2, since we found in our data set many known relevant peptides, which showed low matching scores, due to low quality MS2 spectra.

At present, quantitation of HLA class I ligands is largely lacking. The best reported mass spectrometry-based strategy was the AQUA approach. In this approach a heavy isotope labeled counterpart of the epitope of interest was added to compensate for losses during sample pre-treatment [64]. Since we showed that the major losses occurred in the IP step the AQUA approach as such is not sufficient to accurately determine HLA-peptide copy number. To this end, we developed the hpMHC approach (chapter 5), which takes the AQUA approach to the next level by spiking in HLA-monomers loaded with heavy isotope labeled peptides, which are added to the cell lysate, i.e. before immunopurification step. For this approach hpHLA complexes need to be prepared for every peptide of interest. The number of peptides that can be measured in parallel with this approach would be mainly limited by the preparation of all the different pHLA molecules.

The results of our peptidomics approach opens the possibility for immunologist to identify HLA-ligands and epitopes presented in different disease states. A recent paper on peptides eluted from leukemia cells, using the strategy of our reversed immunology approach, identified several leukemia-specific peptides, which might become candidates for immunotherapy [13]. The peptidomics strategy described in this thesis, and the knowledge gained from it, may help in the improvement of vaccination strategies and immunotherapy by studying the different stages of malignant transformation in tumors, studying autoimmune diseases to identify 'trigger proteins/peptides', studying rejection of transplanted organs after transplantation.

HSPV database and MiHA

To identify polymorphic peptides/MiHA eluted with the large number of HLA-ligands in chapter 2, we developed in chapter 3 a specific database dedicated for identification of polymorphic peptides, called the Human short peptide variation (HSPV) database [65], which is based on the SNP information in dbSNP. The main advantage of this database is the small size of the database which reduces the false positive identified polymorphic peptides and increases the chance of identification of polymorphic peptides with a relatively poor database matching score. Other advantages of the HSPV database are the incorporation of SNP information by its good annotation and hyperlinks to dbSNP databases to have the genome (SNP) information directly available with the polymorphic peptide sequence. To further improve HSPVdb some issues has to be addressed: cleaning up the database from all non-polymorphic peptides, which have frequencies with unknown or (?) quantifications. Incorporation of correct SNP information from dbSNP and HSPVdb has to be updated with the new identified polymorphisms. Although, the HSPVdb is very useful in immunology area, a drawback of this database is the limited utilization of the database by other biologist due to the specificity of the database for a specific field in biology as immunology. From the HSPV database we identified 1500 polymorphic peptides using the data of chapter 2 [15]. As it is known, not all polymorphic peptides can be MiHA and therefore, we

used the polymorphic peptides list to select MiHA candidates using a homemade and developed algorithm by Hombrink et al.[39].

Selection and validation of MiHA among a list of polymorphic peptides

Selection of MiHA among the large number of the eluted polymorphic peptides needs more effort. In order to identify a clinically relevant MiHA restricted to hematopoietic cells in chapter 4 we used the data obtained in chapter 2, matched with HSPV database in combination with in our group developed algorithm (used in the selection of MiHA and T cell recognizing the selected MiHA exclusively) [39].

The selection steps in the previously in our group developed algorithm include: (1) mass spectrometry based validation of eluted peptides, (2) frequency of MiHA allele, (3) binding affinity of peptides to HLA molecule, (4) generation of MHC-tetramer with the selected MiHA candidate, (5) pull down experiment by selection of homozygous negative donors for the SNP encoding MiHA, (6) detection of MHC tetramer positive T cells using FACS analysis, (7) generation of T cell clones, and functional test to detect high avidity T cells.

In chapter 4 we could identify a new clinically relevant MiHA called LB-CLYBL-1Y. In addition another MiHA (LB-TEP1-1S) has been identified as well. Although, we could not prove the clinical relevancy of the LB-TEP1-1S, still it can be used in immunotherapy strategies and in vaccine production against hematological malignancy or any hematological disorder.

Since the number of identified MiHA is low, we studied in depth the developed algorithm to improve the selection steps. Most of the polymorphic peptides (10 from 40 candidates) were excluded in the first step of the algorithm using mascot ion score (BMI)>30 and mass deviation of 10 ppm. Selection of polymorphic peptides using the combination of high (BMI) >35% and mass deviation <2 ppm might increase the selection efficiency. Furthermore, setting the allele frequency between 0.05-0.07% might increase the efficiency of the selection of homozygous negative donor. In addition, utilization of multimer in the fishing of T cells from the peripheral blood of homozygous negative donors is very specific and sensitive but the clear limitation of this method is that the MiHA candidate with low binding affinity has to be excluded due to the folding efficiency of the multimer, which may explain the low number of identified MiHA.

hpMHC in combination with AQUA in immunology

Quantitation of MiHA on the cell surface can be useful for improving vaccination strategy but it is still challenging. Two groups have reported the quantitation of T cell epitopes using AQUA approach by addition of isotope labeled peptides to the fractions prior to LC-MS/MS or LC-SRM analysis [64,67]. Since the isolation of HLA-ligands/epitopes (MiHA) goes through a number of steps (Immunoaffinity purification (IP), filtration and separation) prior to LC-MS/MS or LC-SRM [15], addition of labeled peptides to the last step prior to LC-

MS/MS or LC-SRM cannot cover the losses in all preparation steps as performed by Tan et al. [64] and Croft et al. [66]. Therefore, in chapter 5 we developed a new method to account for all losses during sample preparation using heavy peptide MHC (hpMHC). hpMHC is based on stable isotope labeled peptide associated with recombinant MHC. Our methodology has the benefit of adding the labeled peptide associated with recombinant MHC (hpMHC) to the first step (IP) of the preparation method which enables monitoring all losses during sample preparation. In addition, we add a medium labeled peptide (AQUA) to the sample prior to the filtration step or the LC-SRM step to determine the recovery of different steps of the purification method. Our data show that the losses of MiHA are mainly in the IP step, with losses of 90-99%. The low recovery of the peptides by IP step highlights the need for hpMHC in the quantitation of the HLA-ligands/epitopes (MiHA). To improve the recovery of the isolation method especially immunoaffinity column CNBr activated beads can be used instead of Protein-A coupled to CL4B beads and reducing the dimension of the column might be helpful.

The production of hpMHC can be seen as a drawback of this method if higher number of peptides has to be analysed. From our experience up to 20 peptides is very well possible in an immunological environment.

The hpMHC method has been applied in chapter 7 to six candidates, to study the unidirectional phenomenon of the T cell recognition of MiHA by quantitation of the copy number of MiHA and their allelic counterparts on the cell surface.

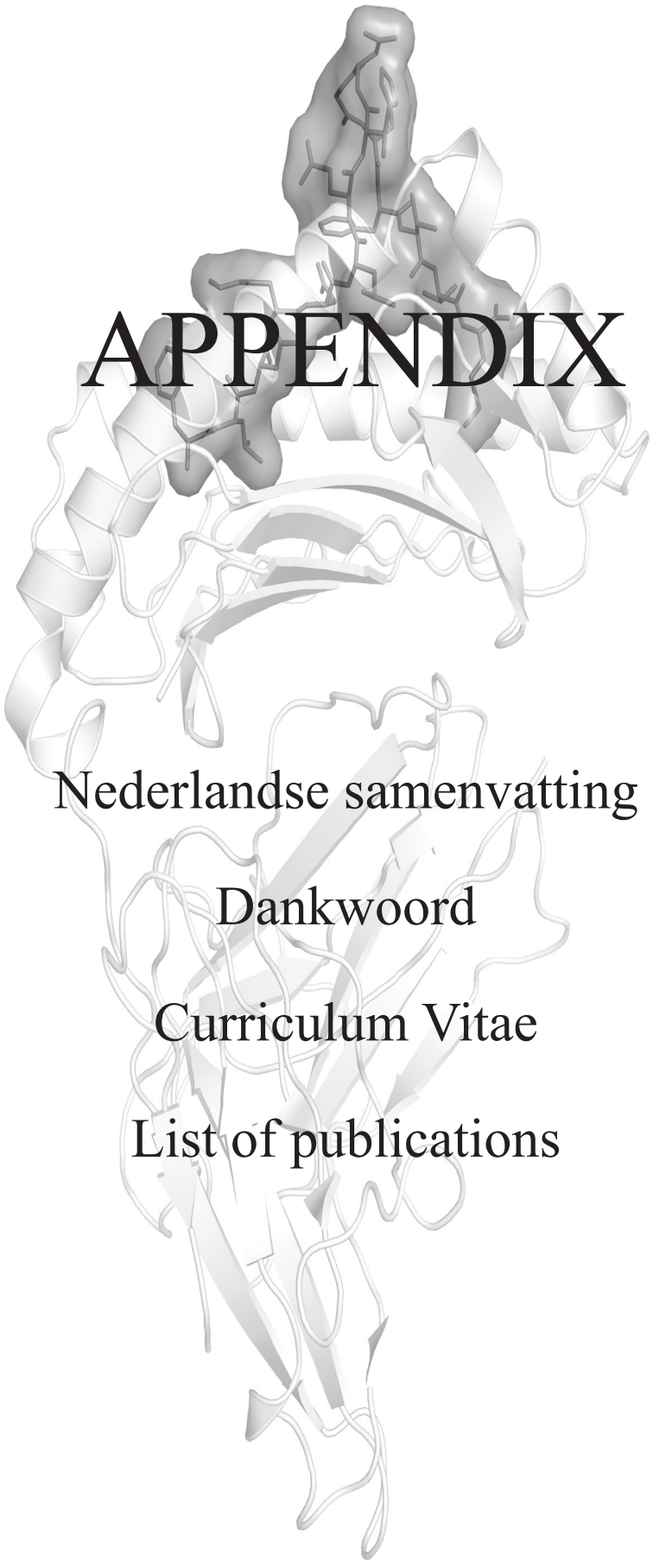
REFERENCES

1. Mester, G., Hoffmann, V. and Stevanovic, S. (2011) Insights into MHC class I antigen processing gained from large-scale analysis of class I ligands. Cell Mol Life Sci, 68, 1521-1532.
2. 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.
3. Hofmann, S., Gluckmann, M., Kausche, S., Schmidt, A., Corvey, C., Lichtenfels, R., Huber, C., Albrecht, C., Karas, M. and Herr, W. (2005) Rapid and sensitive identification of major histocompatibility complex class I-associated tumor peptides by Nano-LC MALDI MS/MS. Molecular & cellular proteomics : MCP, 4, 1888-1897.
4. Hickman, H.D., Batson, C.L., Prilliman, K.R., Crawford, D.L., Jackson, K.L. and Hildebrand, W.H. (2000) C-terminal epitope tagging facilitates comparative ligand mapping from MHC class I positive cells. Human immunology, 61, 1339-1346.
5. Prilliman, K., Lindsey, M., Zuo, Y., Jackson, K.W., Zhang, Y. and Hildebrand, W. (1997) Large-scale production of class I bound peptides: assigning a signature to HLA-B*1501. Immunogenetics, 45, 379-385.
6. Hickman, H.D., Luis, A.D., Buchli, R., Few, S.R., Sathiamurthy, M., VanGundy, R.S., Giberson, C.F. and Hildebrand, W.H. (2004) Toward a definition of self: proteomic evaluation of the class I peptide repertoire. J Immunol, 172, 2944-2952.
7. Barnea, E., Beer, I., Patoka, R., Ziv, T., Kessler, O., Tzehoval, E., Eisenbach, L., Zavazava, N. and Admon, A. (2002) Analysis of endogenous peptides bound by soluble MHC class I molecules: a novel approach for identifying tumor-specific antigens. Eur J Immunol, 32, 213-222.
8. Ben Dror, L., Barnea, E., Beer, I., Mann, M. and Admon, A. (2010) The HLA-B*2705 peptidome. Arthritis Rheum, 62, 420-429.
9. Buchsbaum, S., Barnea, E., Dassau, L., Beer, I., Milner, E. and Admon, A. (2003) Large-scale analysis of HLA peptides presented by HLA-Cw4. Immunogenetics, 55, 172-176.
10. Shoshan, S.H. and Admon, A. (2004) MHC-bound antigens and proteomics for novel target discovery. Pharmacogenomics, 5, 845-859.
11. Maier, R., Falk, K., Rotzschke, O., Maier, B., Gnau, V., Stevanovic, S., Jung, G., Rammensee, H.G. and Meyerhans, A. (1994) Peptide motifs of HLA-A3, -A24, and -B7 molecules as determined by pool sequencing. Immunogenetics, 40, 306-308.
12. Hunt, D.F., Henderson, R.A., Shabanowitz, J., Sakaguchi, K., Michel, H., Sevilir, N., Cox, A.L., Appella, E. and Engelhard, V.H. (1992) Characterization of peptides bound to the class I MHC molecule HLA-A2.1 by mass spectrometry. Science, 255, 1261-1263.
13. Berlin, C., Kowalewski, D.J., Schuster, H., Mirza, N., Walz, S., Handel, M., Schmid-Horch, B., Salih, H.R., Kanz, L., Rammensee, H.G. et al. (2014) Mapping the HLA ligandome landscape of acute myeloid leukemia: a targeted approach toward peptide-based immunotherapy. Leukemia.
14. Mommen, G.P., Frese, C.K., Meiring, H.D., van Gaans-van den Brink, J., de Jong, A.P., van Els, C.A. and Heck, A.J. (2014) Expanding the detectable HLA peptide repertoire using electron-transfer/higher-energy collision dissociation (EThcD). Proc Natl Acad Sci U S A.
15. Hassan, C., Kester, M.G., Ru, A.H., Hombrink, P., Drijfhout, J.W., Nijveen, H., Leunissen, J.A., Heemskerk, M.H., Falkenburg, J.H. and Veelen, P.A. (2013) The human leukocyte antigen-presented ligandome of B lymphocytes. Mol Cell Proteomics.
16. den Haan, J.M.M., Meadows, L.M., Wang, W., Pool, J., Blokland, E., Bishop, T.L., Reinhardus, C., Shabanowitz, J., Offringa, R., Hunt, D.F. et al. (1998) The minor histocompatibility antigen HA-1: A diallelic gene with a single amino acid polymorphism. Science, 279, 1054-1057.
17. Goulmy, E., Termijtelen, A., Bradley, B.A. and van Rood, J.J. (1976) Alloimmunity to human H-Y. Lancet, 2, 1206.
18. Laurin, D., Spierings, E., van der Veken, L.T., Hamrouni, A., Falkenburg, J.H., Souillet, G., Vermeulen, C., Farre, A., Galambrun, C., Rigal, D. et al.(2006) Minor histocompatibility antigen DDX3Y induces HLA-DQ5-restricted T cell responses with limited TCR-Vbeta usage both in vivo and in vitro. Biol Blood Marrow Transplant, 12, 1114-1124.
19. van der Harst, D., Goulmy, E., Falkenburg, J.H., Kooij-Winkelaar, Y.M., van Luxemburg-Heijs, S.A., Goselink, H.M. and Brand, A. (1994) Recognition of minor histocompatibility antigens on lymphocytic and myeloid leukemic cells by cytotoxic T-cell clones. Blood, 83, 1060-1066.

20. den Haan, J.M., Sherman, N.E., Blokland, E., Huczko, E., Koning, F., Drijfhout, J.W., Skipper, J., Shabanowitz, J., Hunt, D.F., Engelhard, V.H. et al. (1995) Identification of a graft versus host disease-associated human minor histocompatibility antigen. Science, 268, 1476-1480.
21. Spierings, E., Brickner, A.G., Caldwell, J.A., Zegveld, S., Tatsis, N., Blokland, E., Pool, J., Pierce, R.A., Mollah, S., Shabanowitz, J. et al. (2003) The minor histocompatibility antigen HA-3 arises from differential proteasome-mediated cleavage of the lymphoid blast crisis (Lbc) oncoprotein. Blood, 102, 621-629.
22. Brickner, A.G., Warren, E.H., Caldwell, J.A., Akatsuka, Y., Golovina, T.N., Zarling, A.L., Shabanowitz, J., Eisenlohr, L.C., Hunt, D.F., Engelhard, V.H. et al. (2001) The immunogenicity of a new human minor histocompatibility antigen results from differential antigen processing. J Exp Med, 193, 195-206.
23. Dolstra, H., Fredrix, H., Maas, F., Coulie, P.G., Brasseur, F., Mensink, E., Adema, G.J., de Witte, T.M., Figdor, C.G. and van de Wiel-van Kemenade, E. (1999) A human minor histocompatibility antigen specific for B cell acute lymphoblastic leukemia. J Exp Med, 189, 301-308.
24. Dolstra, H., de Rijke, B., Fredrix, H., Balas, A., Maas, F., Scherpen, F., Aviles, M.J., Vicario, J.L., Beekman, N.J., Ossendorp, F. et al.(2002) Bi-directional allelic recognition of the human minor histocompatibility antigen HB-1 by cytotoxic T lymphocytes. Eur J Immunol, 32, 2748-2758.
25. van Bergen, C.A., Kester, M.G., Jedema, I., Heemskerk, M.H., van Luxemburg-Heijs, S.A., Kloosterboer, F.M., Marijt, W.A., de Ru, A.H., Schaafsma, M.R., Willemze, R. et al.(2007) Multiple myeloma-reactive T cells recognize an activation-induced minor histocompatibility antigen encoded by the ATP-dependent interferon-responsive (ADIR) gene. Blood, 109, 4089-4096.
26. Akatsuka, Y., Nishida, T., Kondo, E., Miyazaki, M., Taji, H., Iida, H., Tsujimura, K., Yazaki, M., Naoe, T., Morishima, Y. et al. (2003) Identification of a polymorphic gene, BCL2A1, encoding two novel hematopoietic lineage-specific minor histocompatibility antigens. J Exp Med, 197, 1489-1500.
27. Tykodi, S.S., Fujii, N., Vigneron, N., Lu, S.M., Mito, J.K., Miranda, M.X., Chou, J., Voong, L.N., Thompson, J.A., Sandmaier, B.M. et al. (2008) C19orf48 encodes a minor histocompatibility antigen recognized by CD8+ cytotoxic T cells from renal cell carcinoma patients. Clinical cancer research : an official journal of the American Association for Cancer Research, 14, 5260-5269.
28. Torikai, H., Akatsuka, Y., Miyazaki, M., Tsujimura, A., Yatabe, Y., Kawase, T., Nakao, Y., Tsujimura, K., Motoyoshi, K., Morishima, Y. et al. (2006) The human cathepsin H gene encodes two novel minor histocompatibility antigen epitopes restricted by HLA-A*3101 and -A*3303. British journal of haematology, 134, 406-416.
29. Slager, E.H., Honders, M.W., van der Meijden, E.D., van Luxemburg-Heijs, S.A., Kloosterboer, F.M., Kester, M.G., Jedema, I., Marijt, W.A., Schaafsma, M.R., Willemze, R. et al.(2006) Identification of the angiogenic endothelial-cell growth factor-1/thymidine phosphorylase as a potential target for immunotherapy of cancer. Blood, 107, 4954-4960.
30. Kawase, T., Akatsuka, Y., Torikai, H., Morishima, S., Oka, A., Tsujimura, A., Miyazaki, M., Tsujimura, K., Miyamura, K., Ogawa, S. et al. (2007) Alternative splicing due to an intronic SNP in HMSD generates a novel minor histocompatibility antigen. Blood, 110, 1055-1063.
31. Stumpf, A.N., van der Meijden, E.D., van Bergen, C.A., Willemze, R., Falkenburg, J.H. and Griffioen, M. (2009) Identification of 4 new HLA-DR-restricted minor histocompatibility antigens as hematopoietic targets in antitumor immunity. Blood, 114, 3684-3692.
32. Van Bergen, C.A., Rutten, C.E., Van Der Meijden, E.D., Van Luxemburg-Heijs, S.A., Lurvink, E.G., Houwing-Duistermaat, J.J., Kester, M.G., Mulder, A., Willemze, R., Falkenburg, J.H. et al.(2010) High-throughput characterization of 10 new minor histocompatibility antigens by whole genome association scanning. Cancer Res, 70, 9073-9083.
33. Griffioen, M., van der Meijden, E.D., Slager, E.H., Honders, M.W., Rutten, C.E., van Luxemburg-Heijs, S.A., von dem Borne, P.A., van Rood, J.J., Willemze, R. and Falkenburg, J.H. (2008) Identification of phosphatidylinositol 4-kinase type II beta as HLA class II-restricted target in graft versus leukemia reactivity. Proceedings of the National Academy of Sciences of the United States of America, 105, 3837-3842.
34. Brickner, A.G., Evans, A.M., Mito, J.K., Xuereb, S.M., Feng, X., Nishida, T., Fairfull, L., Ferrell, R.E., Foon, K.A., Hunt, D.F. et al. (2006) The PANE1 gene encodes a novel human minor histocompatibility antigen that is selectively expressed in B-lymphoid cells and B-CLL. Blood, 107, 3779-3786.
35. de Rijke, B., van Horssen-Zoetbrood, A., Beekman, J.M., Otterud, B., Maas, F., Woestenenk, R., Kester, M., Leppert, M., Schattenberg, A.V., de Witte, T. et al.(2005) A frameshift polymorphism in P2X5 elicits an allogeneic cytotoxic T lymphocyte response associated with remission of chronic myeloid leukemia. J Clin Invest, 115, 3506-3516.

36. Warren, E.H., Vigneron, N.J., Gavin, M.A., Coulie, P.G., Stroobant, V., Dalet, A., Tykodi, S.S., Xuereb, S.M., Mito, J.K., Riddell, S.R. et al. (2006) An antigen produced by splicing of noncontiguous peptides in the reverse order. Science, 313, 1444-1447.
37. Murata, M., Warren, E.H. and Riddell, S.R. (2003) A human minor histocompatibility antigen resulting from differential expression due to a gene deletion. J Exp Med, 197, 1279-1289.
38. Rosinski, K.V., Fujii, N., Mito, J.K., Koo, K.K., Xuereb, S.M., Sala-Torra, O., Gibbs, J.S., Radich, J.P., Akatsuka, Y., Van den Eynde, B.J. et al. (2008) DDX3Y encodes a class I MHC-restricted H-Y antigen that is expressed in leukemic stem cells. Blood, 111, 4817-4826.
39. Hombrink, P., Hassan, C., Kester, M.G., de Ru, A.H., van Bergen, C.A., Nijveen, H., Drijfhout, J.W., Falkenburg, J.H., Heemskerk, M.H. and van Veelen, P.A. (2013) Discovery of T cell epitopes implementing HLA-peptidomics into a reverse immunology approach. J Immunol, 190, 3869-3877.
40. Pierce, R.A., Field, E.D., den Haan, J.M., Caldwell, J.A., White, F.M., Marto, J.A., Wang, W., Frost, L.M., Blokland, E., Reinhardus, C. et al. (1999) Cutting edge: the HLA-A*0101-restricted HY minor histocompatibility antigen originates from DFFRY and contains a cysteinylated cysteine residue as identified by a novel mass spectrometric technique. Journal of immunology, 163, 6360-6364.
41. Ivanov, R., Aarts, T., Hol, S., Doornenbal, A., Hagenbeek, A., Petersen, E. and Ebeling, S. (2005) Identification of a 40S ribosomal protein S4-derived H-Y epitope able to elicit a lymphoblast-specific cytotoxic T lymphocyte response. Clinical cancer research : an official journal of the American Association for Cancer Research, 11, 1694-1703.
42. Wang, W., Meadows, L.R., den Haan, J.M., Sherman, N.E., Chen, Y., Blokland, E., Shabanowitz, J., Agulnik, A.I., Hendrickson, R.C., Bishop, C.E. et al. (1995) Human H-Y: a male-specific histocompatibility antigen derived from the SMCY protein. Science, 269, 1588-1590.
43. Meadows, L., Wang, W., den Haan, J.M., Blokland, E., Reinhardus, C., Drijfhout, J.W., Shabanowitz, J., Pierce, R., Agulnik, A.I., Bishop, C.E. et al. (1997) The HLA-A*0201-restricted H-Y antigen contains a post-translationally modified cysteine that significantly affects T cell recognition. Immunity, 6, 273-281.
44. Torikai, H., Akatsuka, Y., Miyazaki, M., Warren, E.H., 3rd, Oba, T., Tsujimura, K., Motoyoshi, K., Morishima, Y., Kodera, Y., Kuzushima, K. et al. (2004) A novel HLA-A*3303-restricted minor histocompatibility antigen encoded by an unconventional open reading frame of human TMSB4Y gene. Journal of immunology, 173, 7046-7054.
45. Warren, E.H., Gavin, M.A., Simpson, E., Chandler, P., Page, D.C., Disteche, C., Stankey, K.A., Greenberg, P.D. and Riddell, S.R. (2000) The human UTY gene encodes a novel HLA-B8-restricted H-Y antigen. Journal of immunology, 164, 2807-2814.
46. Vogt, M.H., Goulmy, E., Kloosterboer, F.M., Blokland, E., de Paus, R.A., Willemze, R. and Falkenburg, J.H. (2000) UTY gene codes for an HLA-B60-restricted human male-specific minor histocompatibility antigen involved in stem cell graft rejection: characterization of the critical polymorphic amino acid residues for T-cell recognition. Blood, 96, 3126-3132.
47. Spierings, E. (2014) Minor histocompatibility antigens: past, present, and future. Tissue Antigens, 84, 374-360.
48. Yewdell, J.W., Reits, E. and Neefjes, J. (2003) Making sense of mass destruction: quantitating MHC class I antigen presentation. Nat Rev Immunol, 3, 952-961.
49. Admon, A. and Bassani-Sternberg, M. (2011) The Human Immunopeptidome Project, a Suggestion for yet another Postgenome Next Big Thing. Mol Cell Proteomics, 10.
50. Lundegaard, C., Lamberth, K., Harndahl, M., Buus, S., Lund, O. and Nielsen, M. (2008) NetMHC-3.0: accurate web accessible predictions of human, mouse and monkey MHC class I affinities for peptides of length 8-11. Nucleic Acids Res, 36, W509-512.
51. Ponomarenko, J., Papangelopoulos, N., Zajonc, D.M., Peters, B., Sette, A. and Bourne, P.E. (2011) IEDB-3D: structural data within the immune epitope database. Nucleic Acids Res, 39, D1164-1170.
52. Rammensee, H., Bachmann, J., Emmerich, N.P., Bachor, O.A. and Stevanovic, S. (1999) SYFPEITHI: database for MHC ligands and peptide motifs. Immunogenetics, 50, 213-219.
53. Salimi, N., Fleri, W., Peters, B. and Sette, A. (2010) Design and utilization of epitope-based databases and predictive tools. Immunogenetics, 62, 185-196.
54. Vita, R., Zarebski, L., Greenbaum, J.A., Emami, H., Hoof, I., Salimi, N., Damle, R., Sette, A. and Peters, B. (2010) The immune epitope database 2.0. Nucleic Acids Res, 38, D854-862.

55. Cobbold, M., De La Pena, H., Norris, A., Polefrone, J.M., Qian, J., English, A.M., Cummings, K.L., Penny, S., Turner, J.E., Cottine, J. et al. (2013) MHC class I-associated phosphopeptides are the targets of memory-like immunity in leukemia. Sci Transl Med, 5, 203ra125.
56. Zarling, A.L., Ficarro, S.B., White, F.M., Shabanowitz, J., Hunt, D.F. and Engelhard, V.H. (2000) Phosphorylated peptides are naturally processed and presented by major histocompatibility complex class I molecules in vivo. J Exp Med, 192, 1755-1762.
57. Zarling, A.L., Polefrone, J.M., Evans, A.M., Mikesh, L.M., Shabanowitz, J., Lewis, S.T., Engelhard, V.H. and Hunt, D.F. (2006) Identification of class I MHC-associated phosphopeptides as targets for cancer immunotherapy. Proc Natl Acad Sci U S A, 103, 14889-14894.
58. Meyer, V.S., Drews, O., Gunder, M., Hennenlotter, J., Rammensee, H.G. and Stevanovic, S. (2009) Identification of natural MHC class II presented phosphopeptides and tumor-derived MHC class I phospholigands. J Proteome Res, 8, 3666-3674.
59. Jahn, L., Hombrink, P., Hassan, C., Kester, M.G., van der Steen, D.M., Hagedoorn, R.S., Falkenburg, J.H., van Veelen, P.A. and Heemskerk, M.H. (2014) Therapeutic targeting of the BCR associated protein CD79b in a TCR-based approach is hampered by aberrant expression of CD79b. Blood.
60. Weinzierl, A.O., Lemmel, C., Schoor, O., Muller, M., Kruger, T., Wernet, D., Hennenlotter, J., Stenzl, A., Klingel, K., Rammensee, H.G. et al.(2007) Distorted relation between mRNA copy number and corresponding major histocompatibility complex ligand density on the cell surface. Mol Cell Proteomics, 6, 102-113.
61. Illing, P.T., Vivian, J.P., Dudek, N.L., Kostenko, L., Chen, Z., Bharadwaj, M., Miles, J.J., Kjer-Nielsen, L., Gras, S., Williamson, N.A. et al. (2012) Immune self-reactivity triggered by drug-modified HLA-peptide repertoire. Nature, 486, 554-558.
62. Perkins, D.N., Pappin, D.J., Creasy, D.M. and Cottrell, J.S. (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis, 20, 3551-3567.
63. Peng, J., Elias, J.E., Thoreen, C.C., Licklider, L.J. and Gygi, S.P. (2003) Evaluation of multidimensional chromatography coupled with tandem mass spectrometry (LC/LC-MS/MS) for large-scale protein analysis: the yeast proteome. J Proteome Res, 2, 43-50.
64. Tan, C.T., Croft, N.P., Dudek, N.L., Williamson, N.A. and Purcell, A.W. (2011) Direct quantitation of MHC-bound peptide epitopes by selected reaction monitoring. Proteomics, 11, 2336-2340.
65. Nijveen, H., Kester, M.G., Hassan, C., Viars, A., de Ru, A.H., de Jager, M., Falkenburg, J.H., Leunissen, J.A. and van Veelen, P.A. (2011) HSPVdb--the Human Short Peptide Variation Database for improved mass spectrometry-based detection of polymorphic HLA-ligands. Immunogenetics, 63, 143-153.
66. Croft, N.P., Smith, S.A., Wong, Y.C., Tan, C.T., Dudek, N.L., Flesch, I.E., Lin, L.C., Tscharke, D.C. and Purcell, A.W. (2013) Kinetics of antigen expression and epitope presentation during virus infection. PLoS Pathog, 9, e1003129.



APPENDIX

Nederlandse samenvatting

Dankwoord

Curriculum Vitae

List of publications

NEDERLANDSE SAMENVATTING

Minor histocompatibiliteits antigenen (of minor antigenen) zijn natuurlijk gevormde peptiden die gepresenteerd worden in HLA-moleculen op het celmembraan, waar ze herkend kunnen worden door T cellen van individuen met hetzelfde HLA type. Reeds dertig jaar geleden is de rol van minor antigenen beschreven bij het graft versus tumor effect. Inmiddels is er een overvloed aan bewijs dat in allogene stamceltransplantatie (alloSCT) minor antigenen zowel het graft versus tumor effect als het graft versus host effect mediëren. Het graft versus host effect is een belangrijke, soms fatale, complicatie tijdens deze behandeling. Om alloSCT succesvol te maken is het belangrijk om minor antigenen te vinden die wel het graft versus tumor effect vertonen, maar niet, of nauwelijks, aanleiding geven tot het graft versus host effect. Op dit ogenblik zijn er ongeveer 50 minor antigenen geïdentificeerd. Deze antigenen zijn peptiden, die gevormd zijn door enzymatische afbraak in de cel. Een deel van deze peptiden wordt geladen in HLA-moleculen, en de HLA-peptide complexen worden vervolgens naar de membraan getransporteerd, waar enkele honderdduizenden complexen worden gepresenteerd. Om te identificeren welke peptiden worden gepresenteerd, worden de HLA-peptide complexen vrijgemaakt uit de membraan en opgezuiverd, en wordt de pool van enkele tienduizenden peptiden vervolgens geïsoleerd en geïdentificeerd. Er zijn twee hoofdstrategieën te onderscheiden om minor antigenen te identificeren, de voorwaartse benadering en de terugwaartse benadering. De voorwaartse benadering kan worden toegepast als er reactieve T cellen zijn gevonden na alloSCT. De terugwaartse benadering gebruikt algoritmes en beschikbare databanken om nieuwe minor antigenen te voorspellen. Een belangrijke beperking van de terugwaartse benadering is dat veel van de voorspelde antigenen niet op het celmembraan worden gepresenteerd. Daarnaast kunnen post-translationeel gemodificeerde peptiden niet voorspeld worden. Daarom verkennen we in deze thesis de mogelijkheden om op basis van het geëlueerde peptidenrepertoire, het ligandoom, nieuwe minor antigenen te identificeren.

In hoofdstuk 2 passen we deze nieuwe strategie toe, door het ligandoom van twee EBV-getransformeerde B cellen zo goed volledig mogelijk te bepalen. Vanwege de grote complexiteit en het dynamisch bereik van het ligandoom hebben we multidimensionele chromatografie toegepast, met in de eerste dimensie een scheiding met isoelectrische focussering, ionenuitwisseling of omkeerfase chromatografie gevolgd door een tweede scheiding met omkeerfase chromatografie gekoppeld met een massaspectrometer. Met onze strategie werden ongeveer 16.000 liganden, waaronder minor antigenen en post-translationeel gemodificeerde peptiden, geïdentificeerd met hoge betrouwbaarheid uit de HLA typen A2, B7 en B44. De identificatie van de peptiden werd gevalideerd door synthese van geselecteerde peptiden, het controleren op physicochemische eigenschappen, bindingsaffiniteit op basis van het NetMHC algoritme, lengte en hydrofobiciteit. Deze grote set peptiden vormt de basis voor de volgende stappen van deze studie. Om minor antigenen beter te kunnen identificeren hebben we in hoofdstuk 3 een databank ontwikkeld, de Human Short Peptide Variation databank (HSPVdb), die geconstrueerd is door combinatie van de humane mRNA sequenties in de RefSeq databank en de geassocieerde informatie uit de Single Nucleotide Polymorphism Databank (dbSNP). Het merendeel van de irrelevante SNPs werden verwijderd, wat resulteerde in een compacte databank van hoge kwaliteit om minor antigenen met verhoogde kans op te sporen. Door matching van onze data met de HSPVdb konden we 1500 polymorfe peptiden met grote betrouwbaarheid identificeren. Om potentiele minor antigenen te identificeren uit de set van 1500 polymorfe peptiden hebben we in hoofdstuk 4 de 25 meest veelbelovende kandidaten geselecteerd. In deze selectie van 25 peptiden werden 2 nieuwe minor antigenen geïdentificeerd, te weten LB-CLYBL-1Y en LB-TEP1-1S.
Omdat we weten dat we het ligandoom nog niet volledig kunnen identificeren hebben we verschillende stap-

pen in het identificatieproces, en met name de opwerking, in detail bestudeerd. In hoofdstuk 5 hebben we daarom een methode ontwikkeld om HLA gepresenteerde peptiden nauwkeurig te kwantificeren. Dit werd gedaan door stabiel isotoop gelabelde peptiden te binden in HLA moleculen. Een bekende hoeveelheid van deze peptide/HLA complexen werd helemaal aan het begin van de opwerking toegevoegd, waardoor verliezen in het opwerkingsproces verrekend konden worden. Om ook de opbrengst van de opwerking te kunnen bepalen werd een andere stabiel isotoop gelabelde variant van dezelfde peptiden toegevoegd voor de finale massaspectrometrische metingen. Van de minor antigenen LB-NISCH1-1A en LB-SSR1-1S werd de opbrengst van de opwerking bepaald op 1-5%. De verliezen bleken bijna volledig toe te schrijven aan de eerste verrijkingsstap waarin HLA complexen worden opgezuiverd met behulp van een antilichaam gericht tegen HLA moleculen.
Deze kwantificeringsmethode werd in hoofdstuk 6 toegepast om de unidirectionele aard van de T cel respons tegen minor antigenen te bestuderen. De unidirectionele aard van de T cel respons betekent dat er slechts een respons is tegen het minor antigeen maar niet tegen de allelische variant. Dat laatste zou verklaard kunnen worden door afwezigheid van voornoemde variant op het celmembraan. We hebben echter in hoofdstuk 2 op verschillende heterozygote cellen van meerdere peptiden beide allelische varianten geïdentificeerd. Van 3 peptiden werden beide allelische varianten in detail bestudeerd: LB-NISCH1 (minor antigeen/allelische variant; A/V), LB-SSR1 (S/L) en LB-CLYBL (Y/D). Uit kwantificering van deze peptiden bleek dat het minor antigeen en de allelische variant in vergelijkbare aantallen op het celmembraan aanwezig waren. Daarmee werd de werkhypothese, dat de antigeendichtheid op het celmembraan minstens 10x lager moet zijn om het verschil in T cel respons te kunnen verklaren, verworpen. De unidirectionele aard van de T cel respons tegen minor antigenen kan dus niet algemeen verklaard worden door de afwezigheid van de allelische variant op het celmembraan. Andere immunologische mechanismen spelen dus een rol.
In hoofdstuk 7 bestuderen we in detail een aantal, voor HLA klasse I, ongewoon lange peptiden, welke werden gevonden in de studie beschreven in hoofdstuk 2. De data uit hoofdstuk 2 bevatte 1.568 peptiden met een lengte van 12-23 aminozuren, waarvan 423 met een lengte van 15-23 aminozuren. Acht 15-meer peptiden werden geselecteerd voor diepgaande studie, waarin de efficiëntie van peptide/HLA monomeervorming werd bestudeerd, en hun thermische stabiliteit. Hieruit bleek dat de 15-meren zich gedroegen als ‘normale’ 9-meer peptiden. Vervolgens werd van twee 15-meer peptide/HLA complexen de kristalstructuur bepaald, waaruit bleek dat de peptiden een zogenaamde ‘super bulged’ structuur aannamen, wat wil zeggen dat de peptiden de normale ankerposities gebruiken, en ver uit de HLA groeve steken.

DANKWOORD

In 2008 ben ik begonnen met mijn promotieonderzoek wat uiteindelijk geresulteerd heeft in dit proefschrift. Ik ben trots op het resultaat, maar dit proefschrift zou nooit tot stand zijn gekomen zonder de hulp van een heleboel mensen.

Als eerste wil ik Michel bedanken. Michel zonder jou zou dit proefschrift niet tot stand gekomen zijn. Vanaf de eerste dag van mijn promotieonderzoek was jij al betrokken bij alle details van het onderzoek. Wij hebben de eerste twee jaren veel experimenten samen uitgevoerd. Ik zal nooit je hulp bij de celkweek, eiwitzuivering en de eluties vergeten. Pleun en Lorenz heel erg bedankt voor jullie bijdrage aan het onderzoek dat is beschreven in dit boek, vooral alle biologische experimenten met de tetrameren en T-cellen. Natuurlijk zal ik de tijd die ik doorgebracht heb bij het Laboratorium voor Experimentele Hematologie niet vergeten. Hierbij wil ik graag mijn collega's en alle medewerkers van de Experimentele Hematologie bedanken voor hun hulp tijdens mijn werk bij de afdeling.

George, bedankt voor je hulp bij de praktische uitdagingen bij mijn experimenten. Ik waardeer ontzettend de gesprekken waarin je jouw ervaringen met anderen deelt. Emmely, Jeroen en Kees, bedankt voor de leuke wetenschappelijke en niet wetenschappelijke discussies. Het was heel gezellig om samen met jullie in de kamer te zitten en te werken. Veronica, Frederike, Yvonne, Annemieke, Lianne, Allan en Jeroen, bedankt voor de gezellige dagen en jullie steun tijdens mijn promotie. Willemien, Robert en Natasja, bedankt voor de fijne samenwerking. Ik wil ook mijn collega's van de Afdeling Immunohematologie en Bloedtransfusie bedanken voor hun steun en hulp tijdens mijn promotie.

Dan de paranimfen: Kayi, bedankt omdat je mij altijd hebt gesteund vanaf het begin van de BFW-studie tot nu toe. Arnoud, ik wil je bedanken voor alle fractioneringen en metingen van de monsters met massaspectrometer die je hebt uitgevoerd. Zonder jou zou ik dit moment niet hebben kunnen bereiken. Het is een eer voor mij om naast jullie te staan tijdens mijn verdediging.

Ik wil hiermee mijn familie en vrienden in het Koerdisch bedanken voor hun steun en belangstelling in mijn promotie:

لیرەوە پێم خۆشە زۆر سوپاسی باوک و دایکی زۆر خۆشەویستم بکەم بۆبەخێو کردنی من بەشێوازێکی زانستیانە و

مۆدێرنانە. ئێوە هەمیشە پاڵپشتی من بوون لە هەموو هەنگاوێکی ژیانمدا خوا بۆمنتان بهێلێت و تەمەن درێژ بن.

هەروەها ئەمەوێت سوپاسی هاوسەری خۆشەویستم ڕێبوار، کوڕی شیرینم ڕەوێن گیان، براکانم و هاوڕێ

خۆشەویستەکانم بکەم کە هەمیشە پاڵپشت بوون بۆم و کارەکانیان بۆ ئاسان کردوم.

Rawen, door mijn studie en mijn promotie heb ik je vaak alleen moeten laten. Bedankt voor je geduld, steun en begrip in alle opzichten. Rebwar, zonder jou kon ik mijn promotie niet afronden. Bedankt voor je zorg, voor alle momenten die je mij gerust kon stellen en voor je oneindige begrip en geduld met mij.

CURRICULUM VITAE

Chopie Hassan werd geboren op 12 januari 1971 te Sulaimania, Koerdistan (Irak). In het jaar 1987 behaalde zij haar VWO-diploma in Sulaimania, Irak. In datzelfde jaar begon zij met de studie chemie aan de Universiteit van Salahadin in de stad Erbil (Irak). Na de eerste Golfoorlog heeft zij het bachelor diploma chemie behaald (1991). Door de moeilijke situatie in Irak mocht zij niet doorstuderen na die oorlog. Sinds 2000 woont zij in Nederland. In 2001 begon zij aan de vijfjarige studie Bio-farmaceutische wetenschappen aan de Universiteit Leiden. De masterstage werd uitgevoerd op de afdeling Farmacologie van het Leiden Amsterdam Center for Drug Research (LACDR) onder begeleiding van dr. Liesbeth de Lange en de prof. M . Danhof). De tweede stage vond plaats bij de Human Genetica-afdeling van het Leids Universitaire Medisch Centrum (LUMC) onder begeleiding van prof. Arn van den Maagdenberg. In 2006 werd de mastertitel behaald in de bio-farmaceutische wetenschappen. In hetzelfde jaar is zij begonnen als studieleider in vitro ADME te werken bij NOTOX in ‘s-Hertogenbosch. Daarna heeft zij als researchanalist gewerkt op de afdeling van Analytical Bio-Sciencesvan de Leidse Universiteit tot 2008. In hetzelfde jaar is zij begonnen aan het in dit proefschrift beschreven promotieonderzoek bij de afdeling van Immunohematologie bloedtransfusie’ en het ‘Laboratorium voor experimentele hematologie’van het LUMC onder begeleiding van dr. P.A.van Veelen en prof. dr. J.H.F. Falkenburg. Na de promotie periode begon zij werk als onderzoeker bij prof. Thomas Hankemeier op de Analytical Bio-Sciences afdeling van de Universiteit Leiden. Op dit moment werkt zij als Study Manager bij MicroSafe Laboratories in Leiden.

LIST OF PUBLICATION

Identification of Biological Relevant Minor Histocompatibility Antigens within the B-lymphocyte-Derived HLA-Ligandome Using a Reverse Immunology Approach.
Hombrink P, **Chopie Hassan**, Kester MG, Lorenz Jahn, Pont MJ, de Ru AH, van Bergen CA, Griffioen M, Falkenburg JH, van Veelen PA, Heemskerk MH. Clin Cancer Res. (2015) 1;21 (9):2177-2186.

Naturally processed non-canonical HLA-A*02:01 presented peptides.
Chopie Hassan, Eric Chabrol, Lorenz Jahn, Michel G.D. Kester, Arnoud H. de Ru, Jan W. Drijfhout, Jamie Rossjohn, J.H. FrederikFalkenburg, Mirjam H. M. Heemskerk, Stephanie Gras & Peter A. van Veelen. J. Biol. Chem. (2014),M114.607028.

Therapeutic targeting of the BCR associated protein CD79b in a TCR-based approach is hampered by aberrant expression of CD79b.
Lorenz Jahn, Pleun Hombrink, **ChopieHassan**, Michel G.D. Kester, Dirk M. van der Steen, Renate S. Hagedoorn, J.H. FrederikFalkenburg, Peter A. van Veelen, andMirjam H.M. Heemskerk. Blood. (2014)-07-587840.

Accurate quantitation of MHC-bound peptides by application of isotopically labeled peptide MHC complexes.
Chopie Hassan, Kester MG, Oudgenoeg G, de Ru AH, Janssen GM, Drijfhout JW, SpaapenRM, Jiménez CR, Heemskerk MH, Falkenburg JH, van Veelen PA. J Proteomics. (2014) 19 (109):240-244.

Unique peptide-binding motif for Mamu-B*037:01: an MHC class I allele common to Indian and Chinese rhesus macaques.
de Groot NG, Heijmans CM, de Ru AH, **Chopie Hassan,** Otting N, Doxiadis GG, Koning F, van Veelen PA, Bontrop RE. Immunogenetics. (2013)65(12):897-900.

The human leukocyte antigen-presented ligandome of B lymphocytes.
Chopie Hassan, Kester MG, de Ru AH, Hombrink P, Drijfhout JW, Nijveen H, Leunissen JA, Heemskerk MH, Falkenburg JH, van Veelen PA. Mol Cell Proteomics(2013) 12(7):1829-43.

Discovery of T cell epitopes implementing HLA-peptidomics into a reverse immunology approach.
Hombrink P, **Chopie Hassan**, Kester MG, de Ru AH, van Bergen CA, Nijveen H, Drijfhout JW, Falkenburg JH, Heemskerk MH, van Veelen PA. J Immunol. (2013)15 (190):3869-3877.

Alternative peptide repertoire of HLA-E reveals a binding motif that is strikingly similar to HLA-A2. Lampen MH, **Chopie Hassan**, Sluijter M, Geluk A, Dijkman K, Tjon JM, de Ru AH, vander Burg SH, van Veelen PA, van Hall T. Mol Immunol. (2013)53(1-2):126-31.

Comparison of peptide and protein fractionation methods in proteomics.
Ekaterina Mostovenko, **Chopie Hassan,** Janine Rattke, André M. Deelder, Peter A. van Veelen, Magnus Palmblad. EuPA Open Proteomics, (2013) 1: 30 37.

HSPVdb--the Human Short Peptide Variation Database for improved mass spectrometry-based detection of polymorphic HLA-ligands.
Nijveen H, Kester MG, **Chopie Hassan**, Viars A, de Ru AH, de Jager M, Falkenburg JH, Leunissen JA, van Veelen PA. Immunogenetics. (2011) 63(3):143-53.