



## Signal peptides and trans-membrane regions are broadly immunogenic and have high CD8+ T cell epitope densities: Implications for vaccine development

Riva Kovjazin<sup>a</sup>, Ilan Volovitz<sup>c</sup>, Yair Daon<sup>a</sup>, Tal Vider-Shalit<sup>b</sup>, Roy Azran<sup>b</sup>, Lea Tsaban<sup>b</sup>, Lior Carmon<sup>a,\*</sup>, Yoram Louzoun<sup>b,\*\*</sup>

<sup>a</sup> Vaxil BioTherapeutics Ltd, 13A, WIS Science Park, Nes-Ziona 74036, Israel

<sup>b</sup> Department of Mathematics, Gonda Brain Research Center, Bar-Ilan University, Ramat Gan 52900, Israel

<sup>c</sup> The Cancer Immunotherapy Lab, The Neurosurgery Dept, The Tel-Aviv Sourasky Medical Center, Israel

### ARTICLE INFO

#### Article history:

Received 1 November 2010

Received in revised form

28 December 2010

Accepted 9 January 2011

Available online 12 February 2011

#### Keywords:

Epitopes

CD8+ T cells

Vaccine

Signal peptide

SIR

### ABSTRACT

Cell mediated immune response has a major role in controlling the elimination of infectious agents. The rational design of sub-unit peptide vaccines against intracellular pathogens or cancer requires the use of antigenic sequence/s that can induce highly potent, long lasting and antigen-specific responses in the majority of the population. A promising peptide selection strategy is the detection of multi-epitope peptide sequences with an ability to bind multiple MHC alleles. While past research sought the best epitopes based on their specific antigenicity, we ask whether specific defined domains have high epitope densities. Signal peptides and trans-membrane domains were found to have exceptionally high epitope densities. The improved MHC binding of these domains relies on their hydrophobic nature and, in signal peptides, also on their specific sequence. The high epitope density of SP was computed using *in-silico* methods and corroborated by the high percentage of identified SP epitope in the IEDB (immune epitope database). The enhanced immunogenicity of SP was then experimentally confirmed using a panel of nine peptides derived from Mycobacterium tuberculosis (MTb) proteins used in human PBMC proliferation assays and T cell lines functional assays. Our results show the exceptionally high antigen specific response rates and population coverage to SP sequences compared with non-SP peptide antigens derived from the same proteins. The results suggest a novel scheme for the rational design of T cell vaccines using a domain based rather than an epitope based approach.

© 2011 Elsevier Ltd. All rights reserved.

### 1. Introduction

Traditional vaccines are based on dead or attenuated pathogens (Artenstein, 2008), which can be potentially dangerous or difficult to maintain in ambient conditions. An alternative to whole-pathogen based vaccines are subunit peptide vaccines (Brossart et al., 2000). These vaccines are composed of small peptides presented to T lymphocytes in the context of MHC molecule (Agudelo and Patarroyo, 2010). These peptides are fully synthetic, non-hazardous and usually easy to maintain in ambient conditions, enabling the vaccination of populations in countries with lower medical standards.

An important issue in the design of such vaccines is the choice of optimal epitopes that can induce the maximal immune activation in the largest fraction of the population. Numerous methodologies

have been introduced to design optimal multi-epitope vaccines that could cater to a large proportion of the population. These methodologies were usually based on combining peptides that bind highly frequent MHC alleles from a single antigen (Ossendorp et al., 1998; Porgador et al., 1996) or from several antigens (Odunsi et al., 2007; Surman et al., 2000; Toussaint et al., 2008; van Mierlo et al., 2004; Vider-Shalit et al., 2007b; Welters et al., 2008). In this study we test the possibility of a novel alternative of selecting highly immunogenic regions containing multiple epitopes, based on their protein domain identity.

MHC class I molecules, found on all nucleated cells and platelets, are loaded with peptides which are representative of the protein repertoire of these cells. The antigen presentation machinery in eukaryotic cells breaks down all intracellular proteins into short peptides by the proteasome. The proteasome has three subunits: LMP2, LMP7 and LMP10, with different protease or peptidase activities. Some of the generated peptides have lengths enabling their binding to MHC class I. The transport of proteasome-degraded peptides into the endoplasmic reticulum (ER) involves an ATP-dependent transporter designated TAP (Lindquist et al., 1998; Lyko et al., 1995; Martoglio and Dobberstein, 1998). The TAP is

\* Corresponding author. Tel.: +972 8 9396948.

\*\* Corresponding author. Tel.: +972 3 5317610; fax: +972 3 5317165.

E-mail addresses: [lior.carmon@vaxilbio.com](mailto:lior.carmon@vaxilbio.com) (L. Carmon), [louzouy@math.biu.ac.il](mailto:louzouy@math.biu.ac.il) (Y. Louzoun).

a membrane-spanning heterodimer consisting of TAP1 and TAP2 subunits. TAP from different organisms have different preferences for transported peptides based on the hydrophobicity or the charge of the C terminus and of other parts of the transported peptide. In order for a peptide to serve as a good CD8+ Cytotoxic T Lymphocyte (CTL) epitope, it must be cleaved, transferred to the ER and then bind with sufficient affinity to MHC-I molecules.

Based on these three requirements (cleavage, TAP binding/transfer to ER and MHC binding), we have developed a set of highly precise algorithms (Ginodi et al., 2008; Louzoun et al., 2006; Vider-Shalit et al., 2007a) to test the epitope density in different sequences. We have previously shown the precision of these algorithms and their applicability to several key questions in immunology (Almani et al., 2009; Ginodi et al., 2008; Louzoun et al., 2006; Vider-Shalit et al., 2007a,b, 2009a,b). These tools were used in this study to systematically analyze the epitope density of different protein domains. Proteins can be targeted for ubiquitination for a variety of reasons; a major source of such proteins is defective ribosomal products (DRiPs) (Dolan et al., 2010). In the current analysis, the source of the proteins is irrelevant to the analysis, as we computed only the fraction of ninemers that can become epitopes given the degradation of the protein.

A genome-wide scan of epitopes in protein domains was performed, showing that certain protein domains, such as signal peptides (SP) and trans-membrane (TM) domains, encode for the highest number of MHC-I epitopes per sequence length. Our results suggest that while the TM sequences are preferred purely based on their hydrophobicity, SP seems to contain sequences, that are not only hydrophobic, but are also well adapted, sequence-wise to presentation on MHC-I molecules. Our *in-silico* results with SP domain coincided with the large number of identified immune epitopes found in SP domains in the IEDB database.

Following up on these bioinformatics results, we went on to experimentally compare SP epitopes to epitopes in other regions of the same proteins. In particular, we asked whether the *in-silico* estimations regarding the high immunogenicity of SP domains can actually be translated into better vaccine targets with higher immunodominant properties. To this end, we tested the immunogenicity of nine proteins derived from the Mycobacterium tuberculosis (MTb).

MTb is a pathogenic gram positive bacteria and the causative agent of most cases of tuberculosis. MTb emerged 40,000 years ago from its African progenitor, coinciding with the expansion of “modern” human populations out of Africa (Wirth et al., 2008). MTb is a facultative intracellular bacterium that has evolved sophisticated mechanisms to survive and replicate inside host mononuclear phagocytes (Welsh and Mason, 2001). The pathogenicity of MTb, its long history of co-existence with humans and its high incidence of infection (approximately 1/3 of the human population is infected) together with the involvement of CTL in anti-MTb immunity (Cho et al., 2000; Hervas-Stubbs et al., 2006), make the MTb an excellent model to search for co-evolution footprints of the adaptive immune system antigen presentation machinery and proteins expressed by this pathogen. Indeed, we found that SP induced stronger immune responses in a larger proportion of the tested population than non-SP peptides in the same proteins.

## 2. Materials and methods

### 2.1. Human and mouse sequences

Human and mouse gene sequences were used for this analysis. The sequences were obtained from the Ensembl database (Hubbard et al., 2009). The sequences represent the last build of the human and mouse genomes where only known proteins were included.

Some domains have extremely high or low average epitope densities, which are due to a low frequency, often of highly homologous sequences. In order to avoid such domains, we did not use any domain that had less than 50 appearances in the genome in the human domains. In the mouse, fewer domains were defined, so we used even domain groups with 15 occurrences. The resulting list has 44,274 and 11,308 human and mouse defined domains, respectively. A single sequence was analyzed in each gene, even if multiple sequences of the same gene were available.

### 2.2. Domain definition

The domains in each gene were defined according to the EXPASY nomenclature (<http://www.expasy.ch/>). For each gene, we extracted from the EXPASY website (Gasteiger et al., 2003) the domains defined for this gene. The sequence of this specific domain was downloaded and the number of epitopes in the domains was computed. Some of the domains overlapped, while some regions were not defined in any domain. We merged the results of similar domains with slightly varying names. The list of domains used can be found in the supplementary material (Supplementary material Table 1).

### 2.3. Relative epitope density computation

We have analyzed the ratio between the epitope density in given domains and the epitope density expected in a random sequence. The epitope number was computed using three algorithms: a homemade cleavage algorithm (Ginodi et al., 2008), a TAP binding algorithm developed by Peters et al. (2003) and the BIMAS MHC binding algorithms. We have used 31 HLA alleles and weighted the results according to the allele frequency in the global human population. The algorithms accuracy was systematically validated using epitope databases. A detailed description of the algorithms and their validation can be found in previous works (Almani et al., 2009; Louzoun et al., 2006; Vider-Shalit et al., 2007a). Briefly, for each domain we computed the number of epitopes in the domain divided by the number of ninemer peptides that are fully in this domain and defined this ratio as the epitope density. We did not take into account epitopes spanning two domains.

### 2.4. MHC binding predictions for MTb epitopes

The MTb binding prediction was performed for the most frequent HLA class I alleles in the world populations. The binding strength of 9 ninemers to the Class I alleles was predicted again using the BIMAS software (<http://bimas.cit.nih.gov/>). Binding strength for Class I were defined as: strong, peptide score of 100+; medium, 10–100; weak, 5–10.

MHC class I epitopes prediction from each MTb antigen was performed on the entire functional SP domain as defined by SignalP 3.0 server (Bendtsen et al., 2004). In order to select control peptides for MHC binding, in non-SP domain, we used the following selection criteria:

1. In antigens with no previously known immunodominant properties e.g. ‘Rv0476/MTO4941 precursor’ or ‘Uncharacterized protein Rv1334/MT1376 precursor’, sequences with a similar length to that of the entire SP domain were analyzed in the entire mature protein and were categorized as high (a), moderate (b), or low (c) MHC binding affinities.
2. In proteins that were previously reported to encode immunogenic MHC-I epitopes like ‘Lipoprotein lpqH’ and ‘ATP dependent helicase putative’ (De Groot et al., 2005; Mustafa et al., 2000; Hohn et al., 2003; McMurphy et al., 2005), the length of these published control epitopes was increased using the native sequence

to match the length of the SP sequence of the same target antigen. One exception to the above rules was applied in the case of 'Antigen 85B'. Since the 'Antigen 85B's published epitopes were much shorter than the antigen's SP domain (19 mer vs. 40 mer), we kept their original length as is.

## 2.5. Statistics

All pair-wise comparisons were performed using *T* tests with a Bonferroni correction. When comparing multiple groups, first an ANOVA was performed. Then each domain was compared to the mean of all others using a two sided *T* test with a Bonferroni correction.

## 2.6. Peptide synthesis

MTb SP-derived peptides were chemically synthesized (EMC microcollections, Tübingen, Germany). Other peptides were chemically synthesized (GL Biochem, Shanghai, China) by fully automated solid-phase peptide synthesis using Fmoc/tBu-strategy and Rink-amide-polystyrene resin. The purity and identity of all the peptides were determined by HPLC-Mass Spectra analysis and were >70% or >90% for the SP domains and >90% for any other control non-SP peptides.

## 2.7. Proliferation induced by MTb peptides

PBMC from whole blood of naïve donors collected by the Israeli National Blood Bank was separated on ficoll gradient (Histopaque, Sigma, Israel) at 1800 rpm for 30 min. PBMC was then suspended at a final concentration of  $2 \times 10^6$  cells/ml in RPMI medium (Biological Industries, Beit haemek, IL) supplemented with 5% Human AB serum (Sigma Israel, IL), L-glutamine, sodium pyruvate, non-essential amino-acid, HEPES and Gentamycin (Biological Industries, Beit haemek, IL). Next, 100  $\mu$ l of the PBMC suspension was placed in 96-well flat bottom plates (Costar, Corning, US) together with 100  $\mu$ l of RPMI medium containing an evaluated synthetic peptide at the final concentrations of 10  $\mu$ g/ml. To evaluate the overall proliferation capacity of the PBMC, PHA was added at a final concentration of 2  $\mu$ g/ml. As a negative control, we used PBMC with no stimulant or with 10  $\mu$ g/ml of human Fab (Jackson ImmunoResearch Suffolk, UK). Plates were cultured in 37 °C, 5% CO<sub>2</sub> and after 5–6 days 0.5  $\mu$ Ci of <sup>3</sup>[H] Thymidine (Amersham, Little Chalfont, Buckinghamshire, UK) was added to each well for an incubation of an additional 18 h. Plates were harvested on UniFilter 96-well plates (PerkinElmer, Waltham, MA, USA), and counted using a  $\beta$  counter (Packard Matrix 96 direct beta-counter, Downers Grove, IL USA).

## 2.8. Development of SP induced T cell lines

Thawed PBMCs obtained from naïve donors underwent initial stimulation for 7 days with peptides-pulsed autologous DC at a ratio of 20:1 in RPMI medium supplemented with 50 IU/ml of human recombinant IL-7 (PeproTech Asia, IL). Next, PBMCs underwent a second stimulation for 5 days with peptide-pulsed adherent autologous monocytes in RPMI medium supplemented with 50 IU/ml of Human IL-7. After 5 days, the cell growth medium was replaced with a fresh medium containing 1  $\mu$ g/ml of peptides and 50 IU/ml of human recombinant IL-2 (PeproTech Asia, IL) and the cells were re-stimulated for an additional 48 h. The enriched T cell lines developed using this protocol were used for phenotype and functional assays.

## 2.9. Phenotypic analysis of SP-induced T cell lines

Phenotype evaluation of SP-induced T cell lines was performed by FACS analysis. Briefly, T cells were suspended at a final concentration of  $20 \times 10^6$  cells/ml in a blocking solution of PBS containing 3% FCS and 0.1% sodium azide; 50  $\mu$ l from this cell suspension ( $1 \times 10^6$  cells) were transferred into a 5 ml FACS tube and incubated for 30 min on RT in the dark with the following conjugated antibodies; anti-CD3-PE, anti-CD4 or anti-CD8-PerCPCy5.5, anti-CD44-APC and anti-CD62L-FITC (Ebiosciences, San Diego, CA, USA). Next, the cells were washed with 2 ml of the blocking buffer and re-suspended in 0.5 ml of PBS. Samples were then read analyzed by LSR II Flow cytometer (BD Biosciences San Jose, CA, USA).

## 2.10. Intracellular staining (ICS) flow cytometry using SP-induced T cell lines

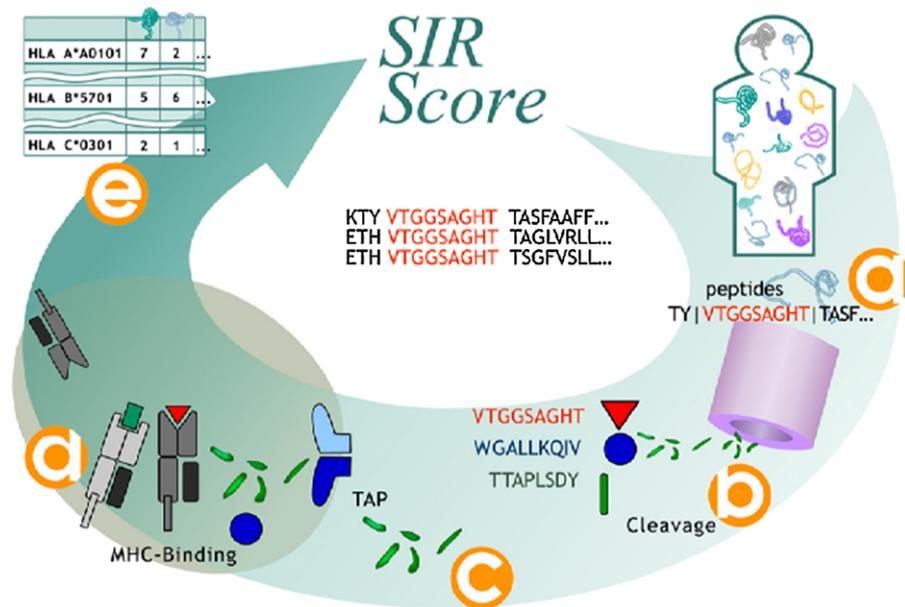
For ICS analysis, T cells were re-stimulated for 6 h using autologous monocytes loaded with the evaluated SP peptides. Two hours after the stimulation was initiated, 1  $\mu$ l/ml of BFA (GolgiPlug, BioLegend, San Diego, CA, USA) was added to each sample for additional 4 h incubation. At the end of the stimulation, T cells were washed twice with PBS and underwent cell surface staining with anti-CD3-PE and anti-CD4 or anti-CD8-PerCPCy5.5. For ICS, T cells were fixed and permeabilized using the LeucoPerm Kit (AbD Serotec, Kidlington, Oxford, UK). The cells were stained using APC conjugated IFN-gamma antibody or FITC conjugated IL-4 antibody (both from eBioscience, San Diego, CA, USA). For determination of the maximal cytokine secretion, T cells were stimulated with 50 ng/ml of PMA and 750 ng/ml of ionomycin. Samples were read analyzed by LSR II Flow cytometer.

## 3. Results

### 3.1. SIR scores

We define the "Size of Immune Repertoire" (*SIR*) score of an amino acid sequence as the ratio between the predicted CTL epitope density found in this sequence and the epitope density expected in a random sequence. The amino acid composition of the random sequence follows the single amino acids and amino acid couples distribution of human proteins (Vider-Shalit et al., 2007a). For instance, assume a 308-amino acid sequence from a human protein having 300 overlapping ninemers. If a set of 300 random nine-mers with a typical human amino acid distribution is expected to have 10 HLA A\*0201 epitopes and the original sequence is computed to have only 4 HLA A\*0201 epitopes, then the *SIR* score of this sequence for HLA A\*0201 would be 0.4. The *SIR* score of a protein sequence in a population is defined as the weighted average *SIR* score for all HLA, weighted by the HLA allele frequency in that population. An average *SIR* score of less than 1 typically represents an immune under-representation of the epitopes in the sequence; conversely, an average *SIR* score of more than 1 represents an immune over-representation of those epitopes. The current analysis used the HLA frequencies of the entire human population and not of any specific sub-population.

Effective presentation of CTL epitopes requires fulfilling three criteria: (1) production of the peptide following proteasomal cleavage; the peptide should be produced by efficient cleavage at its flanking sides and not be cleaved in its center (Ginodi et al., 2008); (2) Efficient transport of the peptide through the TAP machinery into the ER; (3) High affinity binding to an MHC-I molecule (Fig. 1). In the case of SP, the transition through TAP to the ER may not be always required since they can enter the ER via a TAP-independent mechanism (Lyko et al., 1995; Martoglio and Dobberstein, 1998).



**Fig. 1.** Algorithm for SIR score computation. Each human protein is divided into all ninemers and the appropriate flanking regions (a). For each ninemer a cleavage score is computed (b). For all ninemers with a positive cleavage score, a TAP binding was computed and supra-threshold peptides were chosen (c). The MHC binding score of all TAP binding and cleaved ninemers was computed (d). Ninemers passing all these selection stages are defined as epitopes. (e) The number of epitopes per protein per HLA allele was then calculated.

Incorporating all the three requirements has resulted in high quality prediction of naturally processed epitopes (Ginodi et al., 2008; Vider-Shalit et al., 2007a,b, 2009a). All the algorithms used here were successfully validated using various quality assurance processes on seven different databases of experimentally determined epitopes to ensure that the error levels are low enough to allow a thorough and reliable systematic analysis of the domain repertoires (see <http://peptibase.cs.biu.ac.il/peptibase/validation.htm>).

### 3.2. SIR scores of different protein domains

The functional protein domains of all human proteins were extracted from ExPasy and gathered into groups of similar domain name/activity. The human proteins used were those defined as known proteins appearing in the Ensembl database (Hubbard et al., 2009). Similar domains were grouped together (see Supplementary materials Table 1 for a list of domains used in the analysis). To allow for reliable statistics, only domains appearing in at least 50 proteins in the human genome were used.

SIR scores were calculated using the most common HLA-A (9 alleles), HLA-B (19 alleles) and HLA-C (3 alleles) with well-defined binding motifs. These combined HLA alleles should statistically cover 80–90% of the human population. Since a comparative analysis was performed, octamers and decamers were ignored under the assumption that their statistics would resemble those of nonamers. We have checked the epitope density either with or without the TAP transition algorithm as SP can enter the ER also in a TAP independent manner (Yewdell et al., 1998).

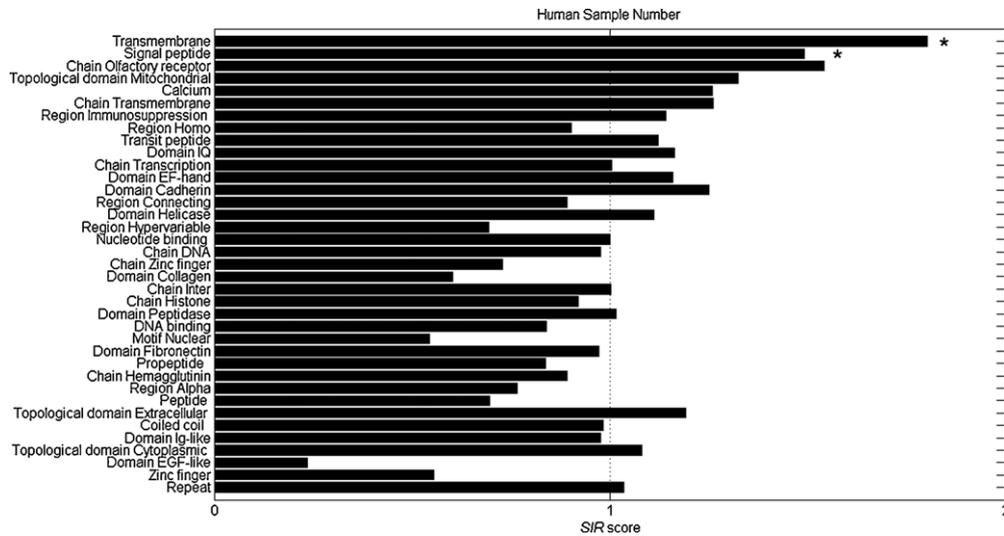
### 3.3. TM and SP domain sequences exhibit high SIR scores in human and mouse proteins

When comparing the SIR scores of all human protein domains without using the TAP module, two domains emerge as having a significantly high SIR score ( $p < 1.e-8$ ,  $T$ -test between each group and the overall mean of all sampled domains with Bonferroni correction): The TM and the SP regions (Fig. 2). When incorporating

the TAP module, no domain has a significantly higher epitope density than any of the other domains (same statistics, Supplementary material Fig. 1). SP probably does not necessarily require TAP binding in order to enter to the ER (Yewdell et al., 1998) and thus they remain as the best candidates for high epitope density regions. Past publications support our analysis. Weinzierl et al. used the TAP-deficient cell line LCL721.174 and its TAP-expressing progenitor cell line LCL721.45 to identify more than 160 HLA ligands, 50 of which were presented through a TAP-independent pathway. About half of the identified TAP-independent presented peptides were derived from SP. The rest were not derived from SP and may have partly been generated by the proteasome. The HLA of the identified SP ligands were mainly HLA-A\*0201 and HLA-B\*5101 (Weinzierl et al., 2008). Note that in the current analysis, all epitopes not fully included within a given domain (e.g. epitopes starting in a domain but ending outside the domain) were ignored.

In order to test the robustness of these results, a similar analysis was performed on mouse proteins using six mouse class I MHC alleles ( $D^b$ ,  $D^d$ ,  $K^b$ ,  $K^d$ ,  $K^k$ ,  $L^d$ ), weighting equally each allele. In the mouse analysis, we compared the expected epitope density in SP and TM domains to the average epitope density in all mouse domains. Again the SP and TM domains had a significantly higher SIR score compared to other domains when the TAP module was not used (Supplementary material Fig. 2) ( $p < 0.05$ ,  $T$ -test between each group and mean with Bonferroni correction). When the TAP module was used, the difference between the SIR score of SP and TM domains and all other domains was not significant (data not shown). Note that although the mouse results echoed those of the human, the frequencies of the alleles used does not necessarily represent their evolved frequency in the natural house-mouse population.

To cross validate these results, we also checked the human SP on mouse MHC alleles and vice versa (Fig. 3). Human SP exhibit higher epitope density on human HLA alleles than on mouse MHC alleles and vice versa. The results indicate that the preference for SP domains is species-dependent and not solely dependent on attributes shared by these two domain types, such as their hydrophobicity.



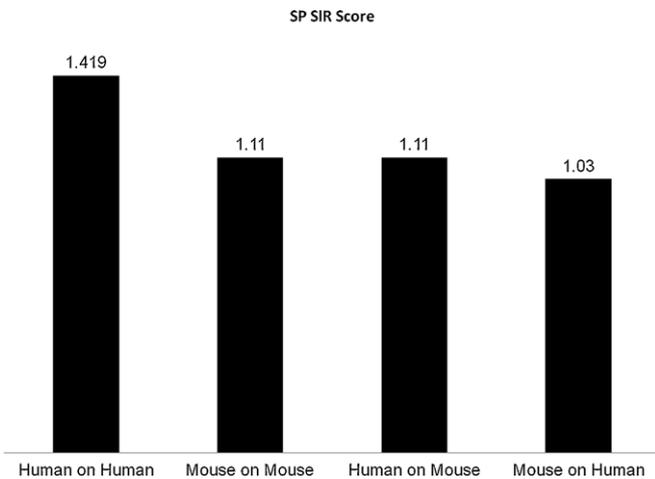
**Fig. 2.** SIR score (without TAP) of all domains in the Expsy domain list with more than 50 segments. The domains were arranged by the significance of their deviation from an SIR score of 1. Only the SP and the TM domains have a significantly high epitope density ( $p < 0.05$ ). For all other domains there is no significant difference ( $p > 0.05$ ).

3.4. Published epitopes found in sp domains

In order to validate our bioinformatic results, we repeated the analysis on published epitopes. All ninemere epitopes from the IEDB database (Sathiamurthy et al., 2005), discovered by MHC ligand elution assays from a human source presented on a human MHC, were downloaded (1012 epitopes ignoring repetitions). The epitopes were not filtered for their HLA allele or any other properties. We only required these epitopes to be presented on human MHC molecules in natural consequences (e.g. excluding peptides inserted into cells). The source protein of each epitope was extracted from the IEDB database. We then downloaded from the 'Uniprot' database the distribution of domains in each protein containing an IEDB epitope. The frequency per amino acid of epitopes (i.e. the number of epitopes divided by the number of candidate ninemers) in SP was three time higher than expected from the distribution of SP in the proteins containing the epitopes (Chi square test,  $p < 1.e-6$ ). TM domains were not significantly over-represented in the published epitopes.

3.5. Tapasin-independent MHC alleles exhibit enhanced immunogenicity of SP sequences

TAP is involved in the transfer of peptides to the ER, but also in the loading of peptides to the MHC molecule. The loading requires the help of the Tapasin molecule (Wright et al., 2004). While the need for Tapasin is shared by most HLA alleles, there is a limited number of alleles that exhibit TAP and Tapasin independent peptide loading. If SP based epitopes are transferred to the ER in a TAP independent pathway, they will probably not be loaded by Tapasin into the MHC-I molecule binding cleft. We repeated the analysis of the epitope prediction for all human proteins (as previously described) for each HLA allele separately. We checked the ratio between the epitope density in SP versus all other domains in the TAP and Tapasin-independent alleles for which there were available MHC binding prediction algorithms (A\*0201, B\*0702, B\*2702, B\*2705, B\*5102, B\*5103 (Brusic et al., 1998)). These alleles have a cumulative frequency of 45% in the Caucasian population. Fig. 4 shows that the epitope density in SP compared to other proteins is even higher in Tapasin independent alleles compared to all other alleles. These results suggest that if evolutionary selection had occurred for the preferential presentation of SP, it mainly occurred on the level of peptides that enter the ER through the TAP-independent pathway and bind the MHC molecules without the aid of Tapasin.

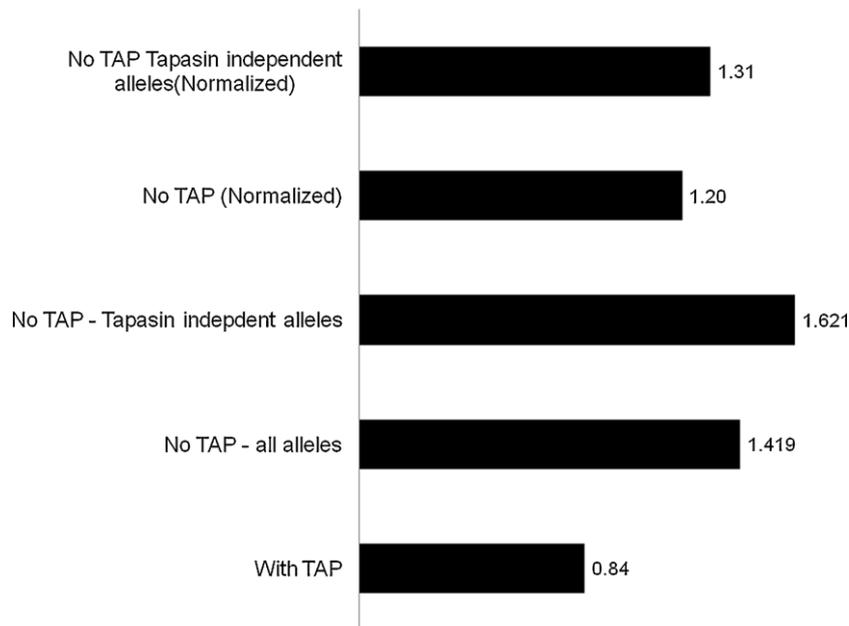


**Fig. 3.** SP SIR scores. The first two columns are the SP SIR scores in human alleles (using human proteins) and in mouse alleles (using mouse proteins). The last two columns are the cross validation (human on mouse and mouse on human). All measurements were performed without the TAP computation. The first three columns are significant at a  $p < 0.05$  level.

3.6. Hydrophobicity enhances the SIR scores of TM and SP domains

A plausible way to explain the high epitope density of the SP and TM regions may be their inherent high hydrophobicity. Hydrophobic regions were previously reported to have high epitope densities (Lucchiari-Hartz et al., 2003). To check the effect of hydrophobicity on the epitope repertoire and whether this effect can explain the high epitope density in TM and SP domains, we measured the correlation between the epitope density and hydrophobicity on all human protein domains, as defined above. The SIR score with or without TAP binding was compared to the hydrophobicity of each domain calculated using three different hydrophobicity scales (Eisenberg, 1984; Janin, 1979; Kyte and Doolittle, 1982).

The SP and TM domains, being highly hydrophobic regions, exhibited a high epitope density. Moreover, the three tested different hydrophobicity scales showed a clear positive association



**Fig. 4.** SP SIR scores. The first column is the full SIR score with TAP. The second column is the SIR without TAP. The third column is the SIR when only Tapasin independent alleles are incorporated. The last two columns are similar to the second and third columns, but divided by the SIR score of the same regions scrambled. These measures are used to estimate the effect of the hydrophobicity on the SIR score. Columns 2, 3, and 5 are significantly higher than 1 ( $p < 0.05$ ). The SIR score was normalized to have an average SIR score of 1 on all human domains measured.

appeared between the SIR score and the observed hydrophobicity of TM or SP domains either with or without the TAP binding (see, for example, Fig. 5 using the Kyte and Doolittle scale hydrophobicity scale (Kyte and Doolittle, 1982)).

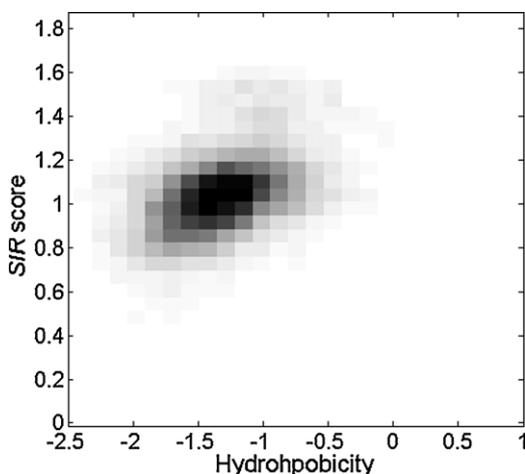
### 3.7. The amino acid sequence influences the SIR scores of SP domains, but not of TM domains

To test whether the high SIR score for these two domains (SP and TM) was due only to their higher hydrophobicity or to their specific sequence, we compared the original sequences of the SP and TM domains found in all human protein domains, to a scrambled version of these sequences. TM domains had similar SIR scores

as compared to their scrambled counterparts ( $p = \text{NS}$ ). The original SP domains had a significantly higher SIR score compared to their scrambled counterparts ( $p < 0.05$ ), both when checking all alleles or Tapasin-independent alleles (Fig. 4). Thus, it seems that while in TM regions the only element affecting the high epitope density is their amino acid composition, in SP, the detailed amino acid sequence also significantly affects their unique immunogenicity. This effect can be either due to some arbitrary correlation between the SP sequence and many MHC binding motifs, to some proteasomal preferential cleavage, or to an evolutionary selection for preferable presentation of SP. The current analysis cannot determine whether structural or evolutionary-immunological reasons underlie the differences in immunogenicity between SP and other domains. However, even if the high epitope density of SP domains is the result of arbitrary correlation between the properties of these domains and the binding motifs, such domains may still be optimal candidates for vaccines.

### 3.8. SP domains of *Mycobacterium tuberculosis* proteins are more immunogenic than other domains in the same proteins

To check whether the *in-silico* estimations regarding the high immunogenicity of SP domains can actually be translated into better vaccine targets with enhanced immunodominant properties, we used proteins from the MTb. For an initial screening, we selected 9 known and novel MTb proteins. 19–40mer peptides representing the entire SP domains in these targets were synthesized, partially purified (>70% purity) and evaluated in a proliferation assay, using PBMCs obtained from eleven independent naïve healthy donors representing a large pool of HLA alleles. All nine SP-peptides (domains) exhibited a positive  $\geq 2$  stimulation index (SI) in some donors. The average SI of each SP peptide was significantly higher ( $T$ -test  $p < 0.05$ ) than that of HuFab used in this and other studies (Morgan and Weigle, 1981) as the negative control (data not shown). In order to exclude past exposure to MTb, proliferation was performed in all evaluated donors to MTb purified protein derivative (PPD). No correlation was found between the proliferation to PPD and to SP Domains (data not



**Fig. 5.** Combined histogram of SIR score and hydrophobicity (based on the Kyte-Doolittle measure). The SIR score and hydrophobicity of each domain were computed and the joint distribution of their value is represented as a two-dimensional histogram. The two measures are correlated ( $p < 1.e-10$ ). Similar results were obtained for other hydrophobicity scales. A high hydrophobicity leads to a high SIR score. This hydrophobicity-related effect explains the large part of the observed relative increase in SP and TM epitope densities.

**Table 1**  
MTb epitopes and SP domains used in this study and the proliferative response they induce. Various peptides were assayed for proliferation using PBMC of 13 healthy donor. SI represents an average of 8 different experiments using 13 different naïve donors. Percent positive represents patients with SI > 2. Bolded numbers represent groups significantly different from other groups responding to peptides in the same protein (Fisher exact  $p < 0.05$ ).

Peptide details							Proliferation			
MTb peptide name	Antigen	Sequence	Location in protein	Length (aa)	MTb SP	Hum SP <sup>a</sup>	Donors <sup>b</sup>	Percent positive	SI	
VXL201 Total	Antigen 85B A5U3Q3	MTDVS RKIRA WGRRLMIGTA AAVVLPGLVG LAGGAATAGA	1–40	40	Yes	Yes	13	85%	2.2	
VXL201a		QQFIYAGSLSALLDPSQGM	181–199	19	No	No	6	50%	1.97	
VXL201b		YLQVPSPSMGRDIKVFQ	50–67	18	No	No	6	17%	1.4	
VXL201c		FSRPGLPVEYLQVPSPSM	41–58	18	No	No	6	0%	1.3	
VXL203 Total	Lipoprotein IpqH P0A5J0	MKRGLTVAVA	1–42	24	Yes	Yes	13	54%	2	
VXL203 Pure		GAAILVAGLS GCSS	1–42	24	Yes	Yes	9	56%	5.4	
VXL203a		AATGIAAVLT DGNPPEVKSV GLGN	81–104	24	No	No	6	50%	2	
VXL208 Total	ATP dependent helicase putative CDC1551	MRFAQPSALS RFSALTRDWF	1–32	32	Yes	Yes	13	77%	2.6	
VXL208 Pure		TSTFAAPTAA QA	1–32	32	Yes	Yes	9	66%	5.4	
VXL208a		EVLRLIRRRS LAALRAQAEP VSTAAYGRFL PA	1172–1203	32	No	No	6	0%	1.7	
VXL211 Total	Un char protein Rv0476/MTO4941 prec	MLVLLVAVLVTAVYAFVHA	1–19	19	Yes	Yes	13	85%	2.71	
VXL211 Pure			1–19	19	Yes	Yes	9	78%	5.3	
VXL211a		P64695	MSACASGV YLVDVVRPKLL E	63–81	19	No	No	6	0%	1.6
VXL211b			AMSACASGV YLVDVVRPKLL	62–80	19	No	No	6	0%	1.3
VXL212 Total	Un char protein Rv1334/MT1376 prec	MLLRKGTVYVVLVIRADLVNAMVAHA	1–25	25	Yes	Yes	13	54%	1.94	
VXL212 Pure			1–25	25	Yes	Yes	9	78%	3.4	
VXL212a		P64813	TEAYPSRT DVKLATEPDA HYVLVST	93–117	25	No	No	6	0%	1.3
VXL212b			P DEACGVLGAP EGSDRPERHI PMTN	30–54	25	No	No	6	0%	1.1
	Total SP sequences			140 (tot)	Yes	Yes	65	56%		
	Total non-SP sequences			199 (tot)	No	No	90	13%		

UnChar= Uncharacterized; Prec= Precursor.

<sup>a</sup> Vaccine's sequence represent the human SP. MTb SP is usually longer.

<sup>b</sup> Bold represent significant differences between observed and expected SP and non-SP sequences ( $p < 0.015$  or lower).

**Table 2**  
Phenotype and Functional properties of SP-induced T cell lines: T cell lines derived from naïve healthy donor PBMCs were assayed using flow cytometry for CD4, CD8, CD44, CD62 co-expression and for IFN-gamma and IL-4 secretion. Results are represented as percent positive cells following the 3rd stimulation.

T cell line	Phenotype (FACS analysis)				Function (ICS)			
	CD4+ T Cells		CD8+ T cells		CD4+ T Cells		CD8+ T Cells	
	CD44 <sup>high</sup>	CD62L <sup>+</sup>	CD44 <sup>high</sup>	CD62L <sup>+</sup>	IFN-gamma	IL-4	IFN-gamma	IL-4
Anti-VXL201	40.9%	17	79.4	11.8	ND	ND	ND	ND
Anti-VXL203	40.5–49.2	10.5–39.4	53–79.7	8.2–17.8	1.7–3.4	0	2.6–8.2	0
Anti-VXL208	52.6	22	50.2	32.1	0	0	0.3	0
Anti-VXL211	38–41	6.5–28.7	47.6–80	4.4–40.9	1.4–3.4	0	3.4–10.6	0
Anti-VXL212	ND <sup>a</sup>	ND	ND	ND	ND	ND	ND	ND
Naïve PBMC	0	2.1–4	1.3–1.5	5.9–6	0	0	0	0

<sup>a</sup> ND: Not done.

shown). The percent of responders to SP sequences in the PPD positive donors was not different from that of the PPD negative donors. From these nine evaluated peptides, five (namely, VXL201, VXL203, VXL208, VXL211 and VXL212) manifested a stronger SI and a higher fraction of responding donors and were subject to further evaluation (Table 1). To specifically check immunogenic properties of the SP-derived peptides/domains, all peptides, excluding VXL201, were further purified (>90% purity) and checked for their ability to induce proliferation on the previous and/or additional naïve healthy donors compared to proliferation to control peptides derived from non-SP domains in the same five antigens (see Section 2)

Table 1 summarizes the proliferation results for the five tested proteins on PBMC obtained from eleven naïve healthy donors tested on cruder SP (11) peptides (purity >70%) or PBMC from six naïve donors using only purer (purity >90%) SP and non-SP peptides (excluding VXL201). In total, while on average a SP sequence induced an immune response in 56% of healthy donors, only 13% of the same donors responded to an equal-length non-SP related control sequence ( $p < 0.000005$ , Fisher exact). These results were systematically observed in 4/5 of the specific MTb proteins, where more donors responded to SP than to non-SP ( $*p < 0.05$  or  $**p < 0.1$ , Fisher exact). Systematically higher stimulation Indices were recorded to SP than to non-SP peptides when cruder peptides were used. When purer peptides of SP or non-SP peptides were used, the differences were even further enhanced.

To confirm MHC class I preference of the SP peptides, we have characterized the phenotype and function of T cell lines that we have developed to the five MTb-specific SP peptide/domains referred to above. The emerging cell lines from 2 individual donors were found to express the memory-cell activation marker CD44<sup>high</sup> and the central memory marker CD62L<sup>+</sup> in both CD4+ and CD8+ subpopulations. Results, in CD4+ T cells, ranged from 38 to 52% and 17 to 39.4% for CD44<sup>high</sup> and CD62L<sup>+</sup>, respectively. Results in CD8+ T cells were higher and ranged from 47 to 80% and 4.5 to 41% for CD44<sup>high</sup> and CD62L<sup>+</sup>, respectively (Table 2). The same T cell lines presented high and specific IFN-gamma production at the range of 1.5–3.5% and 2.5–10.5% for both CD4+ and CD8+ cells, respectively, following stimulation with the SP domains in the VXL203 and VXL211 peptides. In these donors, cytokine production was Th1 associated, since the same peptide-stimulated T cell lines did not produce the Th2 cytokine IL-4. The same cell lines efficiently lysed MHC-restricted autologous target cells infected with the MTb bacteria (data not shown). A detailed analysis of these cell lines properties is beyond the scope of the current work.

Taken together, the in vitro results suggest that, on average, more than 50% (range 39–69%) of the target evaluated population responded to SP domains from MTb antigens, a response frequency that is significantly higher than that observed to non-SP inner parts of the same MTb antigens.

#### 4. Discussion

The rational design of sub-unit peptide vaccines relies on the smart choice of epitopes. However, the focus on specific allele-dependent epitopes such as HLA\*A0201 limits the applicability of such vaccines only to these HLA alleles. An alternative method to this is to use protein domains with an inherent broad immunogenicity.

Here we show that in humans and mice, two domains are inherently immunogenic: the TM and SP domains. These domains had consistently high epitope densities when the peptide cleavage and MHC binding were included in the analysis. When TAP was included in the analysis, the differences were not significant; however, at least in the case of SP, the effect of TAP is less crucial, since SP can actively enter the ER even in the absence of functional TAP machinery (Lyko et al., 1995; Martoglio and Dobberstein, 1998). The hydrophobicity of the TM and SP domains contributes to their inherent immunogenicity. In SP the specific position of key amino acids also supports the immunogenicity especially in regard to alleles that do not need tapasin to aid in peptide loading onto the MHC.

SP are found in both prokaryotes and eukaryotes and target proteins to different cellular compartments (e.g. ER, nucleus, mitochondria) (Lyko et al., 1995). SPs are usually found in the N-terminus of protein and share an organelle-related common motif; notwithstanding, different SP of different antigens exhibit high sequence variability with no particular sequence identity while conforming to the motif needed to maintain their functionality (Martoglio and Dobberstein, 1998). This allows the induction of antigen specific response using merely the antigen's SP domain. In this study, each of the evaluated SP domains was verified by BLAST analysis to have no identity with any other human sequence. The antigen specific properties of SPs were previously shown by other groups (Martoglio, 2003). The bioinformatically suggested inherent immunogenicity of the SP domain was checked in vitro, showing that SP of MTb proteins are more immunogenic than similar length, non-SP inner-peptides that in few of the cases were reported to encode highly immunogenic epitopes.

A study by Jiang et al. appears to support our findings. Results in this study showed that vaccination of mice with a 18 mer SP derived from the bacterial antigen Ag2/PRA, either as a gene or as a synthetic peptide, could induce protective immunity against the *Coccidioidomycosis* microbial pathogen from which the SP was derived. The protective response was superior to that of the mature protein without the SP. Moreover, the protection was highly specific as a frame-shift mutation in Ag2/PRA SP abolished the specific immune activity (Jiang et al., 2002). Similar advantages of using SP epitopes were also shown when the epitopes were derived from MUC1, a highly prevalent tumor associated antigen (TAA) (Brossart et al., 1999; Carmon et al., 2000; Correa et al., 2005).

Using entire functional protein domains in general, and SP domains specifically, has a number of advantages over the use of full length protein antigens or single epitope peptides. Specifically:

- (a) *Promiscuous binding to MHC alleles*: entire functional domains can induce a more robust response in the majority of the population. Unlike single epitope peptide vaccines, using entire functional domains does not require individuals/patients to be selected according to their MHC haplotype. Our results with five SP domains for MTb proteins showed promiscuous activation of multiple T cell clones both CD4<sup>+</sup> and CD8<sup>+</sup> T cells and the expression of a key Th1 cytokine: IFN- $\gamma$ .
- (b) *Combination of CD4<sup>+</sup> and CD8<sup>+</sup> T cell epitopes*. Vaccines fully based on CD8<sup>+</sup> T cell epitopes are not expected to induce a CD4<sup>+</sup> T cell response. Vaccines based on entire functional domains can induce these two types of response.
- (c) *Non hazardous*. In contrast to attenuated viruses/bacteria, peptides composed of a single domain have no functional cytopathic effect. They represent a good balance between the need to present multiple epitopes and the required safety of the vaccine.
- (d) *Stand-alone vaccines*: unlike most single epitope peptide vaccines which are usually poor immunogens (Marchand et al., 1999), SP domains have lipophilic sequences that enable prompt delivery across the cell membrane. This allows enhanced immunogenicity in vitro and their administration without any additional delivery systems. Past studies demonstrated augmented immune responses to MHC class I epitopes from MART or from Ovalbumin antigens via linking to a SP sequence derived from the E3/19 protein of adenovirus type 2. In the case of Ovalbumin, the immune response produced by the SP linked epitopes was comparable to that achieved via IFA (Minev et al., 1994, 2000).
- (e) *TAP-independent presentation*: we and others have shown that SP domains/epitopes have the unique ability to avoid a frequent immune escape mechanism - TAP deficiency, adopted by cancer cells and intracellular pathogens like MTb (Dorfel et al., 2005; Lyko et al., 1995).

One caveat that can be suggested against the basic concept of the epitope density as a meaningful measure is that given the immunodominance of a limited number of epitopes (Irvine and Bennink, 2006), the total amount of epitopes is of no importance. This caveat however does not apply to regions of a few tens of amino acids, since such regions rarely contain more than one epitope and usually have none. Thus, in short sequences, epitope density represents the presence or absence of HLA-related immunogenic epitopes. A high epitope density, in these cases, represents higher HLA allele coverage.

An intriguing question that emerges from these results is the underlying mechanism explaining the high epitope density in SP domains. This high epitope density may be the result of an arbitrary overlap between the typical patterns of SP and frequent MHC binding motifs. Alternatively, the high epitope density in such domains could be explained on the basis of an evolutionary advantage caused by preferential presentation of hydrophobic sequences and especially SPs. Host-pathogen co-evolutionary pressures can change the relative frequency of the different HLA alleles in a population (Kiepiela et al., 2004). It is thus possible that HLA alleles preferentially presenting SP epitopes were selected.

The highly conserved SP motif is almost similar in Eukaryotes and Prokaryotes (Nielsen et al., 1997). Taking together the absolute need for pathogens to use secreted proteins to manipulate their hosts (Dietrich and Doherty, 2009; Mazzaccaro et al., 1996; Teitelbaum et al., 1999), and the host mechanisms that evolved in order to present bacterial peptides on MHC-I molecules

(Mazzaccaro et al., 1996), these two notions might both substantiate the evolutionary/immunological incentive to over represent SP peptides on host MHC-I and a plausible way by which to execute this incentive. A similar logic might explain why evolutionarily highly-conserved proteins, such as heat-shock proteins, are known to exhibit cross-species immunogenicity (Zugel and Kaufmann, 1999a; Zugel and Kaufmann, 1999b). More experimental proof for the generality of the domain-based immunogenicity, of SP and other domains, is yet to be pursued.

## Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.molimm.2011.01.006.

## References

- Agudelo, W.A., Patarroyo, M.E., 2010. Quantum chemical analysis of MHC-peptide interactions for vaccine design. *Mini Rev. Med. Chem.* 10, 746–758.
- Almani, M., Raffaelli, S., Vider-Shalit, T., Tsaban, L., Fishbain, V., Louzoun, Y., 2009. Human self-protein CD8+ T-cell epitopes are both positively and negatively selected. *Eur. J. Immunol.* 39, 1056–1065.
- Artenstein, A.W., 2008. New generation smallpox vaccines: a review of preclinical and clinical data. *Rev. Med. Virol.* 18, 217–231.
- Bendtsen, J.D., Nielsen, H., von Heijne, G., Brunak, S., 2004. Improved prediction of signal peptides: SignalP 3.0. *J. Mol. Biol.* 340, 783–795.
- Brossart, P., Heinrich, K.S., Stuhler, G., Behnke, L., Reichardt, V.L., Stevanovic, S., Muhm, A., Rammensee, H.G., Kanz, L., Brugger, W., 1999. Identification of HLA-A2-restricted T-cell epitopes derived from the MUC1 tumor antigen for broadly applicable vaccine therapies. *Blood* 93, 4309–4317.
- Brossart, P., Wirths, S., Stuhler, G., Reichardt, V.L., Kanz, L., Brugger, W., 2000. Induction of cytotoxic T-lymphocyte responses in vivo after vaccinations with peptide-pulsed dendritic cells. *Blood* 96, 3102–3108.
- Brusic, V., Rudy, G., Harrison, L.C., 1998. MHCPEP, a database of MHC-binding peptides: update 1997. *Nucleic Acids Res.* 26, 368–371.
- Carmon, L., El-Shami, K.M., Paz, A., Pascolo, S., Tzehoval, E., Tirosh, B., Koren, R., Feldman, M., Fridkin, M., Lemonnier, F.A., Eisenbach, L., 2000. Novel breast-tumor-associated MUC1-derived peptides: characterization in Db-/- x beta2 microglobulin (beta2m) null mice transgenic for a chimeric HLA-A2.1/Db-beta2 microglobulin single chain. *Int. J. Cancer* 85, 391–397.
- Cho, S., Mehra, V., Thoma-Uszynski, S., Stenger, S., Serbina, N., Mazzaccaro, R.J., Flynn, J.L., Barnes, P.F., Southwood, S., Celis, E., Bloom, B.R., Modlin, R.L., Sette, A., 2000. Antimicrobial activity of MHC class I-restricted CD8+ T cells in human tuberculosis. *Proc. Natl. Acad. Sci. U.S.A.* 97, 12210–12215.
- Correa, I., Plunkett, T., Coleman, J., Galani, E., Windmill, E., Burchell, J.M., Taylor-Papadimitriou, J., 2005. Responses of human T cells to peptides flanking the tandem repeat and overlapping the signal sequence of MUC1. *Int. J. Cancer* 115, 760–768.
- De Groot, A.S., McMurry, J., Marcon, L., Franco, J., Rivera, D., Kutzler, M., Weiner, D., Martin, B., 2005. Developing an epitope-driven tuberculosis (TB) vaccine. *Vaccine* 23, 2121–2131.
- Dietrich, J., Doherty, T.M., 2009. Interaction of Mycobacterium tuberculosis with the host: consequences for vaccine development. *APMIS* 117, 440–457.
- Dolan, B.P., Li, L., Takeda, K., Bennink, J.R., Yewdell, J.W., 2010. Defective ribosomal products are the major source of antigenic peptides endogenously generated from influenza A virus neuraminidase. *J. Immunol.* 184, 1419–1424.
- Dorfel, D., Appel, S., Grunebach, F., Weck, M.M., Muller, M.R., Heine, A., Brossart, P., 2005. Processing and presentation of HLA class I and II epitopes by dendritic cells after transfection with in vitro-transcribed MUC1 RNA. *Blood* 105, 3199–3205.
- Eisenberg, D., 1984. Three-dimensional structure of membrane and surface proteins. *Annu. Rev. Biochem.* 53, 595–623.
- Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R.D., Bairoch, A., 2003. ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res.* 31, 3784–3788.
- Ginodi, I., Vider-Shalit, T., Tsaban, L., Louzoun, Y., 2008. Precise score for the prediction of peptides cleaved by the proteasome. *Bioinformatics* 24, 477–483.
- Hervas-Stubbis, S., Majlessi, L., Simsova, M., Morova, J., Rojas, M.J., Nouze, C., Brodin, P., Sebo, P., Leclerc, C., 2006. High frequency of CD4+ T cells specific for the TB10.4 protein correlates with protection against Mycobacterium tuberculosis infection. *Infect. Immun.* 74, 3396–3407.
- Hohn, H., Kortsik, C., Tully, G., Nilges, K., Necker, A., Freitag, K., Neukirch, C., Galle, P., Lohr, H., Maeurer, M.J., 2003. Longitudinal analysis of Mycobacterium tuberculosis 19-kDa antigen-specific T cells in patients with pulmonary tuberculosis: association with disease activity and cross-reactivity to a peptide from HIVenv gp120. *Eur. J. Immunol.* 33, 1613–1623, <http://bimas.cit.nih.gov/>.
- Hubbard, T.J.P., Aken, B.L., Ayling, S., Ballester, B., Beal, K., Bragin, E., Brent, S., Chen, Y., Clapham, P., Clarke, L., Coates, G., Fairley, S., Fitzgerald, S., Fernandez-Banet, J., Gordon, L., Graf, S., Haider, S., Hammond, M., Holland, R., Howe, K., Jenkinson, A., Johnson, N., Kahari, A., Keefe, D., Keenan, S., Kinsella, R., Kokocinski, F., Kulesha, E., Lawson, D., Longden, I., Megy, K., Meidl, P., Overduin, B., Parker, A., Pritchard, B., Rios, D., Schuster, M., Slater, G., Smedley, D., Spooner, W., Spudich, G., Trevanion,

- S., Vilella, A., Vogel, J., White, S., Wilder, S., Zadissa, A., Birney, E., Cunningham, F., Curwen, V., Durbin, R., Fernandez-Suarez, X.M., Herrero, J., Kasprzyk, A., Proctor, G., Smith, J., Searle, S., Flicke, P., 2009. Ensembl 2009. *Nucleic Acids Res.* 37, D690–D697.
- Irvine, K., Binnink, J., 2006. Factors influencing immunodominance hierarchies in TCD8+ mediated antiviral responses. *Expert Rev. Clin. Immunol.* 2, 135–147.
- Janin, J., 1979. Surface and inside volumes in globular proteins. *Nature* 277, 491–492.
- Jiang, C., Magee, D.M., Ivey, F.D., Cox, R.A., 2002. Role of signal sequence in vaccine-induced protection against experimental coccidioidomycosis. *Infect. Immun.* 70, 3539–3545.
- Kiepiela, P., Leslie, A.J., Honeyborne, I., Ramduth, D., Thobakgale, C., Chetty, S., Rathnavalu, P., Moore, C., Pfafferott, K.J., Hilton, L., Zimbwa, P., Moore, S., Allen, T., Brander, C., Addo, M.M., Altfeld, M., James, I., Mallal, S., Bunce, M., Barber, L.D., Szigler, J., Day, C., Klenerman, P., Mullins, J., Korber, B., Coovadia, H.M., Walker, B.D., Goulder, P.J., 2004. Dominant influence of HLA-B in mediating the potential co-evolution of HIV and HLA. *Nature* 432, 769–775.
- Kyte, J., Doolittle, R.F., 1982. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* 157, 105–132.
- Lindquist, J.A., Jensen, O.N., Mann, M., Hammerling, G.J., 1998. ER-60, a chaperone with thiol-dependent reductase activity involved in MHC class I assembly. *EMBO J.* 17, 2186–2195.
- Louzoun, Y., Vider, T., Weigert, M., 2006. T-cell epitope repertoire as predicted from human and viral genomes. *Mol. Immunol.* 43, 559–569.
- Lucchiari-Hartz, M., Lindo, V., Hitziger, N., Gaedicke, S., Saveanu, L., van Endert, P.M., Greer, F., Eichmann, K., Niedermann, G., 2003. Differential proteasomal processing of hydrophobic and hydrophilic protein regions: contribution to cytotoxic T lymphocyte epitope clustering in HIV-1-Nef. *Proc. Natl. Acad. Sci. U.S.A.* 100, 7755–7760.
- Lyko, F., Martoglio, B., Jungnickel, B., Rapoport, T.A., Dobberstein, B., 1995. Signal sequence processing in rough microsomes. *J. Biol. Chem.* 270, 19873–19878.
- Marchand, M., van Baren, N., Weynants, P., Brichard, V., Dreno, B., Tessier, M.H., Rankin, E., Parmiani, G., Arienti, F., Humblet, Y., Bourlond, A., Vanwijck, R., Lienard, D., Beauduin, M., Dietrich, P.Y., Russo, V., Kerger, J., Masucci, G., Jager, E., De Greve, J., Atzpodien, J., Brasseur, F., Coulie, P.G., van der Bruggen, P., Boon, T., 1999. Tumor regressions observed in patients with metastatic melanoma treated with an antigenic peptide encoded by gene MAGE-3 and presented by HLA-A1. *Int. J. Cancer* 80, 219–230.
- Martoglio, B., 2003. Intramembrane proteolysis and post-targeting functions of signal peptides. *Biochem. Soc. Trans.* 31, 1243–1247.
- Martoglio, B., Dobberstein, B., 1998. Signal sequences: more than just greasy peptides. *Trends Cell Biol.* 8, 410–415.
- Mazzaccaro, R.J., Gedde, M., Jensen, E.R., van Santen, H.M., Ploegh, H.L., Rock, K.L., Bloom, B.R., 1996. Major histocompatibility class I presentation of soluble antigen facilitated by Mycobacterium tuberculosis infection. *Proc. Natl. Acad. Sci. U.S.A.* 93, 11786–11791.
- McMurry, J., Sbai, H., Gennaro, M.L., Carter, E.J., Martin, W., De Groot, A.S., 2005. Analyzing mycobacterium tuberculosis proteomes for candidate vaccine epitopes. *Tuberculosis (Edinb)* 85, 95–105.
- Minev, B.R., Chavez, F.L., Dudouet, B.M., Mitchell, M.S., 2000. Synthetic insertion signal sequences enhance MHC class I presentation of a peptide from the melanoma antigen MART-1. *Eur. J. Immunol.* 30, 2115–2124.
- Minev, B.R., McFarland, B.J., Spiess, P.J., Rosenberg, S.A., Restifo, N.P., 1994. Insertion signal sequence fused to minimal peptides elicits specific CD8+ T-cell responses and prolongs survival of thymoma-bearing mice. *Cancer Res.* 54, 4155–4161.
- Morgan, E.L., Weigle, W.O., 1981. Polyclonal activation of human B lymphocytes by Fc fragments I. Characterization of the cellular requirements for Fc fragment-mediated polyclonal antibody secretion by human peripheral blood B lymphocytes. *J. Exp. Med.* 154, 778–790.
- Mustafa, A.S., Shaban, F.A., Abal, A.T., Al-Attayah, R., Wiker, H.G., Lundin, K.E., Oftung, F., Huygen, K., 2000. Identification and HLA restriction of naturally derived Th1-cell epitopes from the secreted Mycobacterium tuberculosis antigen 85B recognized by antigen-specific human CD4(+) T-cell lines. *Infect. Immun.* 68, 3933–3940.
- Nielsen, H., Engelbrecht, J., Brunak, S., von Heijne, G., 1997. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng.* 10, 1–6.
- Odunsi, K., Qian, F., Matsuzaki, J., Mhawech-Fauceglia, P., Andrews, C., Hoffman, E.W., Pan, L., Ritter, G., Vilella, J., Thomas, B., Rodabaugh, K., Lele, S., Shrikant, P., Old, L.J., Gnajtic, S., 2007. Vaccination with an NY-ESO-1 peptide of HLA class I/II specificities induces integrated humoral and T cell responses in ovarian cancer. *Proc. Natl. Acad. Sci. U.S.A.* 104, 12837–12842.
- Ossendorp, F., Mengede, E., Camps, M., Filius, R., Melief, C.J., 1998. Specific T helper cell requirement for optimal induction of cytotoxic T lymphocytes against major histocompatibility complex class II negative tumors. *J. Exp. Med.* 187, 693–702.
- Peters, B., Bulik, S., Tampe, R., Endert, P.M.V., Holzhtutter, H.G., 2003. Identifying MHC class I epitopes by predicting the TAP transport efficiency of epitope precursors. *J. Immunol.* 171, 1741–1749.
- Porgador, A., Snyder, D., Gilboa, E., 1996. Induction of antitumor immunity using bone marrow-generated dendritic cells. *J. Immunol.* 156, 2918–2926.
- Sathiamurthy, M., Peters, B., Bui, H.H., Sidney, J., Mokili, J., Wilson, S.S., Fleri, W., McGuinness, D.L., Bourne, P.E., Sette, A., 2005. An ontology for immune epitopes: application to the design of a broad scope database of immune reactivities. *Immunome Res.* 1, 2.
- Surman, D.R., Dudley, M.E., Overwijk, W.W., Restifo, N.P., 2000. Cutting edge: CD4+ T cell control of CD8+ T cell reactivity to a model tumor antigen. *J. Immunol.* 164, 562–565.
- Teitelbaum, R., Cammer, M., Maitland, M.L., Freitag, N.E., Condeelis, J., Bloom, B.R., 1999. Mycobacterial infection of macrophages results in membrane-permeable phagosomes. *Proc. Natl. Acad. Sci. U.S.A.* 96, 15190–15195.
- Toussaint, N.C., Donnes, P., Kohlbacher, O., 2008. A mathematical framework for the selection of an optimal set of peptides for epitope-based vaccines. *PLoS Comput. Biol.* 4, e1000246.
- van Mierlo, G.J., Boonman, Z.F., Dumortier, H.M., den Boer, A.T., Franssen, M.F., Nouta, J., van der Voort, E.L., Offringa, R., Toes, R.E., Melief, C.J., 2004. Activation of dendritic cells that cross-present tumor-derived antigen licenses CD8+ CTL to cause tumor eradication. *J. Immunol.* 173, 6753–6759.
- Vider-Shalit, T., Almami, M., Sarid, R., Louzoun, Y., 2009a. The HIV hide and seek game: an immunogenomic analysis of the HIV epitope repertoire. *AIDS* 23, 1311–1318.
- Vider-Shalit, T., Fishbain, V., Raffaelli, S., Louzoun, Y., 2007a. Phase-dependent immune evasion of herpesviruses. *J. Virol.* 81, 9536–9545.
- Vider-Shalit, T., Raffaelli, S., Louzoun, Y., 2007b. Virus-epitope vaccine design: informatic matching the HLA-I polymorphism to the virus genome. *Mol. Immunol.* 44, 1253–1261.
- Vider-Shalit, T., Sarid, R., Maman, K., Tsaban, L., Levi, R., Louzoun, Y., 2009b. Viruses selectively mutate their CD8+ T-cell epitopes—a large-scale immunomic analysis. *Bioinformatics* 25, i39–i44.
- Weinzierl, A.O., Rudolf, D., Hillen, N., Tenzer, S., van Endert, P., Schild, H., Rammensee, H.G., Stevanovic, S., 2008. Features of TAP-independent MHC class I ligands revealed by quantitative mass spectrometry. *Eur. J. Immunol.* 38, 1503–1510.
- Welsh, D.A., Mason, C.M., 2001. Host defense in respiratory infections. *Med. Clin. North Am.* 85, 1329–1347.
- Welters, M.J., Kenter, G.G., Piersma, S.J., Vloon, A.P., Lowik, M.J., Berends-van der Meer, D.M., Drijfhout, J.W., Valentijn, A.R., Wafelman, A.R., Oostendorp, J., Fleuren, G.J., Offringa, R., Melief, C.J., van der Burg, S.H., 2008. Induction of tumor-specific CD4+ and CD8+ T-cell immunity in cervical cancer patients by a human papillomavirus type 16 E6 and E7 long peptides vaccine. *Clin. Cancer Res.* 14, 178–187.
- Wirth, T., Hildebrand, F., Allix-Beguec, C., Wolbeling, F., Kubica, T., Kremer, K., van Soolingen, D., Rusch-Gerdes, S., Locht, C., Brisse, S., Meyer, A., Supply, P., Niemann, S., 2008. Origin, spread and demography of the Mycobacterium tuberculosis complex. *PLoS Pathog.* 4, e1000160.
- Wright, C.A., Kozik, P., Zacharias, M., Springer, S., 2004. Tapasin and other chaperones: models of the MHC class I loading complex. *Biol. Chem.* 385, 763–778.
- Yewdell, J.W., Snyder, H.L., Bacik, I., Anton, L.C., Deng, Y., Behrens, T.W., Bachi, T., Binnink, J.R., 1998. TAP-independent delivery of antigenic peptides to the endoplasmic reticulum: therapeutic potential and insights into TAP-dependent antigen processing. *J. Immunother.* (1997) 21, 127–131.
- Zugel, U., Kaufmann, S.H., 1999a. Immune response against heat shock proteins in infectious diseases. *Immunobiology* 201, 22–35.
- Zugel, U., Kaufmann, S.H., 1999b. Role of heat shock proteins in protection from and pathogenesis of infectious diseases. *Clin. Microbiol. Rev.* 12, 19–39.