PROCESSING SITES ARE DIFFERENT IN THE GENERATION OF HLA-A2.1-RESTRICTED, T CELL REACTIVE TUMOR ANTIGEN EPITOPES AND VIRAL EPITOPES

X.F. Yang[1], **D. Mirkovic**[2], **S. Zhang**[1], **Q.E. Zhang**[3], **Y. Yan**[1], **Z. Xiong**[1], **Fan Yang**[1], **Irene H. Chen**[1], **L. Li**[4], and **H. Wang**[1]

[1]Department of Pharmacology, Temple University School of Medicine, Philadelphia, PA 19140, U.S.A.

[2]Department of Radiation Physics, University of Texas M. D. Anderson Cancer Center, Houston, TX 77030, U.S.A.

[3]Department of Medicine, Baylor College of Medicine, Houston, TX 77030, U.S.A.

[4]Department of Mathematics, University of New Orleans, New Orleans, LA 70148, U.S.A.

SUMMARY

In order to improve the processing efficiency of T cell tumor antigen epitopes, this bioinformatic study compared proteolytic sites in the generation of 47 experimentally identified HLA-A2.1-restricted immunodominant tumor antigen epitopes to those of 52 documented HLA-A2.1-restricted immunodominant viral antigen epitopes. Our results showed that the amino acid frequencies in the C-terminal cleavage sites of the tumor antigen epitopes, as well as several positions within the 10 amino acid (aa) flanking regions, were significantly different from those of the viral antigen epitopes. In the 9 amino acid epitope region, frequencies differed somewhat in the secondary-anchored amino acid residues on E3 (the third aa of the epitope), E4, E6, E7, and E8; however, frequencies in the primary-anchored positions, on E2 and E9, for binding in the HLA-A2.1 groove remained nearly identical. The most frequently occurring amino acid pairs in both N-terminal and C-terminal cleavage sites in the generation of tumor antigen epitopes were different from those of the viral antigen epitopes. Our findings demonstrate, for the first time, that these two groups of epitopes may be cleaved by distinct sets of proteasomes and peptidases or similar enzymes with lower efficiencies for tumor epitopes. In the future, in order to more effectively generate tumor antigen epitopes, targeted activation of the immunoproteasomes and peptidases could be achieved that mediate the cleavage of viral epitopes; thus, enhancing our potential for antigen-specific tumor immunotherapy.

Correspondence should be addressed to: Xiao-Feng Yang, M.D., Ph.D., Department of Pharmacology, Temple University School of Medicine, 3420 North Broad Street, MRB, Suite 300, Philadelphia, PA 19140, U.S.A. Telephone: 215-707-5985; Fax: 215-707-7068; xfyang@temple.edu.

INTRODUCTION

Vaccines capable of eliciting T cell immune responses have been successfully developed for prevention of 26 viral and bacterial infectious diseases(1). In contrast, despite significant progress(2), effective vaccines for most types of tumor are still lacking(3). Since most tumor antigens reported are nonmutated self-antigens(2), peripheral T cell repertoire may be tolerized to self-antigens via thymic negative selection of autoreactive T cells but reacted to viral (foreign) antigens. This model of self-tolerance via thymic selection is often considered as a mechanism of underlying the efficiency differences between the vaccines against viral infections and that against tumors(4). However, self-tolerance, based on the avidity of T cells for self-MHC (major histocompatibility complex)/self-peptide complexes in the thymic selection process, is far from absolute(4). T cells with low avidity for ubiquitously expressed self-antigens or low level expressed self-antigens can escape clonal deletion in thymus and enter the periphery(4). Thus, thymic tolerance is one of the important factors but not the only factor in determining T cell immune responses to tumors and viral infection. T cell responses are also regulated by the process of antigen processing. Improving antigen processing of tumor antigens has been proposed to be a very important direction in development of novel vaccination strategies against tumors(5). Recent reports demonstrated that interferon (IFN)-γ, which is secreted in large amounts during viral infections(6), alters enzymatic processing and proteolytic specificities in generation of T cell antigen epitopes via induction of immunoproteasomes(7) and novel aminopeptides(8). In other words, these results suggested that proteasomes and other enzymes in generation of T cell antigen epitopes versus viral antigen epitopes could be different due to differential expression of IFN-γ during viral infections and tumor growth(7). Notwithstanding, the comprehensive features of proteolytic cleavage sites involved in differential generation of tumor antigen epitopes and viral antigen epitopes remain unknown.

For generation of MHC class I-restricted antigen epitopes, several requirements have been identified, including the following: 1) cleavage sites and favorable flanking sequence around cleavage sites can be effectively recognized by ubiquitin-proteasome complex(9,10) - although nonproteasomal mechanisms also seem to be involved in antigen processing(11,12); 2) transported antigen epitopes have high affinity for binding to transporter associated protein (TAP)(13,14) and for being transported to MHC class I complex (15); and 3) antigen epitopes have high affinity for binding and stabilizing HLA (human leukocyte antigen, human MHC) class I complex on the cell surface(16). It is well accepted that a small subset of peptides generated by proteasomes and processing peptidases, transported by TAP, loaded on MHC class I(17) are potent in elicitation of T cell immune responses and become immunodominant epitopes(18,19), which are desirable for development of novel immunotherapy.

Recently developed serological analyses of tumor antigens by recombinant expression cDNA cloning (SEREX)(20,21) have led to the identification of a large number of tumor antigens (22), which hold great promise as targets for novel antigen-specific tumor immunotherapy (23,24). Previously, we identified broadly immunogenic SEREX tumor antigens(25), CML66L (26,27) and CML28(28), with which specific high-titer IgG antibody responses were associated in the remission of chronic myelogenous leukemia (CML)(25,26,28). Recently, our findings indicated that the overexpression of CML66L in tumor cells, mediated by alternative splicing, is the mechanism of the immunogenicity of this antigen, suggesting that overexpression of SEREX-identified tumor antigens by vaccination could generate anti-tumor immune responses (29). Immunization using dominant antigenic peptides has been most effective in patients with tumors (30) and has generated surprisingly high levels of circulating T cells directed against tumor antigens with a therapeutic outcome(31). Thus, immunodominant epitopes capable of eliciting remarkable CD8+ T cell responses would contribute decisively to the improvement of peptide-based immunization protocols for patients with tumors(32). However, due to our

incomplete understanding of the mechanism underlying the generation of immunodominant (measurable T cell reactive) epitopes, as well as technical limitations, the identification of immunodominant T cell antigen epitopes from SEREX antigens has been accomplished at a slow pace; nonetheless, a few SEREX antigens (i.e., MAGE-1, tyrosinase, NY-ESO-1, coactosin-like protein and CML66) are reported to have the ability to elicit both cellular and humoral immune responses to tumor cells(29,33,34). To facilitate the identification of T cell reactive epitopes encoded by a large number of the SEREX antigens, two important questions must be addressed: 1) whether the structures around the cleavage sites generating the T cell reactive tumor antigen epitopes are different from those of identified immunodominant viral antigen epitopes; and 2) if bioinformatic features of the cleavage sites generating the dominant tumor antigen epitopes can be extracted using a statistical approach, whether the processing efficiency of immunodominant tumor epitopes can be improved thereby in the future.

In this study, we hypothesize that proteolytic cleavage sites generating the identified immunodominant tumor antigen epitopes may be statistically different from those generating documented immunodominant viral epitopes. To test this hypothesis, we focused on the statistical analysis of HLA-A2.1-restricted tumor antigen nonapeptide epitopes and viral antigen nonapeptide epitopes that were previously identified to be T cell reactive through the experimental approaches of others(18,19) (also, see the excellent web database**:** http://www.cancerimmunity.org/peptidedatabase/Tcellepitopes.htm). The purpose in establishment of the public database(s) of T cell antigen epitopes is for database mining to reveal novel information. The statistical approach that we applied in this study has the advantage of revealing important information on structural features through biochemical analysis of individual antigen epitopes, as we demonstrated previously(35). Of note, in contrast to the biochemical analyses with epitope peptides digestible by proteasomes or peptidases, we focused on analyzing the experimentally identified HLA-A2.1 restricted T cell reactive epitopes, which are desirable for future development of antigen specific immunotherapy(32). In contrast to the traditional research processes of designing experiments to test a hypothesis, we used the published experimental data to examine a hypothesis, which was successfully applied in our recent findings(35). Defining the features shared by experimentally identified tumor antigen epitope cleavage sites in a statistical approach would be a very important key to understanding the generation of tumor antigen epitopes versus that of the viral antigen epitopes. We found that experimentally identified, HLA-A2.1-restricted T cell reactive tumor antigen epitopes share structural features around the cleavage sites, but that these structural features were not identical to those used in the generation of viral antigen epitopes. Our new discoveries through the panoramic analysis, in return, have justified this bioinformatic approach. With such knowledge, we could perfect our ability to make processing of tumor antigen epitopes more efficiently, and improve tumor immunotherapy.

MATERIALS AND METHODS

Experimentally identified HLA-A2.1-restricted tumor antigen epitopes and viral antigen epitopes

The 47 HLA-A2.1-restricted tumor antigen epitopes previously identified by the experimental approaches of others (18,19) are listed in Table I (also, see the web database at: http://www.cancerimmunity.org/peptidedatabase/Tcellepitopes.htm). The experimentally confirmed HLA-A2.1-restricted viral antigen epitopes are as follows: the 24 immunodeficiency virus (HIV) viral antigen epitopes, including those encoded by five HIV viral proteins collected from the HIV Molecular Immunology Database (http://hivweb.lanl.gov/content/immunology/maps/ctl/p17.html), as well as published data (36); the 10 hepatitis B virus (HBV) epitopes encoded by two proteins of HBV(37); the four hepatitis C virus (HCV) epitopes encoded by four proteins(38); and the 14 influenza A virus

epitopes encoded by six proteins(39). To achieve parity, only nonapeptide tumor antigen epitopes and nonapeptide viral antigen epitopes—and not epitopes of other lengths—were analyzed in this study.

Nomination of amino acid positions in the antigen epitopes and flanking regions

In accordance with the enzymatic cleavage nomenclature of Schechter and Berger(40), our analyses (Fig. 1) included the ten amino acid residues flanking the N-terminal cleavage site of the nonapeptide epitopes and the ten residues flanking the C-terminal cleavage site of the antigen epitopes(41). The protein-protein BLAST search for short exact matches was performed on the NCBI website (http://www.ncbi.nlm.nih.gov/BLAST/) to retrieve both the N-terminal and C-terminal flanking regions of HLA-A2.1-restricted T cell antigen epitopes. During the final phase, epitopes are generated by two cleavages on both the N-terminus and C-terminus; thus, in addition to enzymatic cleavage nomenclature(40), the amino acid positions in the epitope (E1 to E9), the N-terminal (N10 to N1), and C-terminal (C1 to C10) flanking regions were further nominated (Fig. 1). Substrate specificities of the enzymes were retrieved from the Comprehensive Enzyme Information System, BRENDA (http://www.brenda.unikoeln.de/index.php4).

Algorithms for antigenic epitope prediction

The four algorithms, the MAPPP (http://www.mpiib-berlin.mpg.de/MAPPP/cleavage.html) (42), the MHC-Pathway (http://www.mhc-pathway.net/)(43), the MHC-Pathway immunoproteasome, and the NetChop3.0 neural network predictor (http://www.cbs.dtu.dk/services/NetChop/) (44), were used to predict the proteasome cleavage sites on antigens. In addition, a prediction algorithm (TAPPred) (http://www.imtech.res.in/raghava/tappred/) for the transporter associated protein (TAP) binding was used to predict the TAP binding potential of antigen epitopes. Furthermore, binding of the antigen peptide epitopes on the HLA-A2.1 molecule was predicted by using two different web-based algorithms, including one on the BIMAS/NIH (BIMAS) website (http://bimas.dcrt.nih.gov/molbio/hla_bind/) and another on the SYFPEITHI (SYF) website (http://syfpeithi.bmi-heidelberg.com/scripts/MHCServer.dll/home.htm).

Probability for the overlaps of amino acid pairs at the N- and C-terminal cleavage sites

The 400 potential pairs of amino acids in nature were numbered from 1 to 400. It was assumed that the amino acid pairs were represented in the cleavage sites of tumor antigens and virus antigens, which were assigned as the pair set k_1, and the pair set k_2, respectively, allowing for repetitions. The probability was calculated that there is an overlap between the sets such that the r_1 pairs from the first set are the same as the r_2 pairs in the second set (see Fig. 2A). Such an overlap as an ordered pair (r_1, r_2) was denoted. A simple counting method, as described (45), was used to compute event probabilities (the probability mass function) for different overlaps for $N = 400$, $k_1 = 47$ and $k_2 = 52$.

Statistical methods

Using a one-sample test for binomial proportion(46), the frequencies of amino acid residues in the N-terminal and C-terminal positions of the HLA-A2.1-restricted tumor antigen epitope cleavage sites were calculated and compared to the general occurrence frequencies of each amino acid in any position of the proteins(47). Similarly, the frequencies of amino acid residues in the N-terminal and C-terminal positions of both cleavage sites of the HLA-A2.1-restricted viral antigen epitopes were calculated and compared to the general occurrence frequencies of each amino acid in any position of the proteins(47). The Wilcoxon rank-sum test was used to compare proteasome cleavage probabilities between the tumor epitopes and the viral epitopes,

as predicted with the MAPPP algorithm
(http://www.mpiib-berlin.mpg.de/MAPPP/cleavage.html).

RESULTS

The preference of amino acids in the flanking positions of tumor antigen epitopes was different from that of the viral antigen epitopes

In order to determine whether cleavage sites in the generation of tumor antigen epitopes were different from those of viral epitopes, we analyzed all of the 47 HLA-A2.1-restricted immunodominant tumor antigen epitopes experimentally identified so far(18,19) and identified immunodominant 52 HLA-A2.1-restricted viral antigen epitopes, along with nine amino acid epitope residues, and the ten amino acids in the N-terminal and C-terminal flanking regions. These 47 epitopes were derived from 24 tumor antigens (see Table I), which included representatives from the four tumor antigen groups characterized thus far: the group of differentiation antigens, including tyrosinase, gp100; the group of amplified/oncogenic antigens, including HER-2/neu, WT1; the group of mutational antigens, including p53; and the group of cancer-testis antigens, including MAGE and NY-ESO-1. This collection allowed us to analyze the common structural features of the cleavage sites and the epitopes shared by various tumor antigens. For the purpose of comparison, we also collected 52 HLA-A2.1-restricted viral antigen epitopes encoded by HIV, HBV, HCV, and influenza A virus, as the reference epitopes. These viral antigen epitopes were suitable for comparison, due to the inclusion of both DNA and RNA viruses, which were categorized into several virus families, including the retroviridae (HIV), the hepadnaviridae (HBV), the flaviviridae (HCV), and the orthomyxoviridae (influenza virus A)(48).

In our preliminary studies, the mean and variance of the processing probabilities of tumor epitopes and viral epitopes were calculated by using the MAPPP algorithm. Based on these results, an estimation of sample size, in comparing the means of the tumor epitope group and the viral epitope group, was calculated according to previously published statistical methods (49). The results indicated that, in each group, 33 epitopes or more must be included in order to obtain 80% power (not shown); thus, the number of epitopes in the tumor epitope group (47 epitopes) and the viral epitope group (52 epitopes) exceeded the calculated power requirement (49). Of note, the statistical estimation of sample size suggested that the conclusion achieved in this study, with more than the sufficient sample size to gain the >80% high power levels is statistically significant, but not biased(49).

We analyzed the statistical differences in the frequency of each amino acid in 29 positions of the tumor antigen epitopes and the viral antigen epitopes, and the flanking regions in comparison to the general occurrence frequencies(47). Since both groups of antigen epitopes were compared against the same control amino acid occurrence frequencies(47), the results from these two groups were comparable. The amino acid frequencies that are statistically higher than the background are listed in Table II. Several findings were reported: 1) at 17 out of the 29 positions, amino acid distributions in the tumor antigen and viral antigen epitopes deviated significantly from the background (Table II) ($p<0.05$); 2) in the flanking region positions, on N9, N6, N5, and C8, only the tumor antigen—and not the viral antigen—epitopes deviated from the background; and (3) in the flanking region positions, on N10, C1, C3, C7, and C9, only the viral antigen—and not the tumor antigen—epitopes deviated from the background. Previous studies showed some amino acid preferences in the N-termini (Pn1′) and the C-termini (Pc1′) of the proteasome cleaved epitopes. These studies also showed that the C-terminal cleaved position (Pc1′) prefers K, R, A, and S, but does not favor F, D, and E(17). In contrast, our data did not find statistically significant differences in the amino acid occurrence frequencies in position Pn1′ of either the tumor antigen or viral antigen epitopes ($p>0.05$). In addition, our data on viral antigen epitopes showed that this C-terminal cleavage position (Pc1′)

favored T. Our analyses indicated that both the N-terminal and the C-terminal cleavage sites of the tumor antigen epitopes were different from those of the viral antigen epitopes.

Cumulatively, these results suggest the following. First, the epitope flanking regions of tumor and viral epitopes have amino acid preferences that are statistically different from the general amino acid frequency background. The following reports support our design in using general amino acid frequency as a background: (1) In contrast to bacteria, human viruses do not have their own protein translation machinery, and need to use human cell protein translation machinery for synthesis of viral proteins(48); (2) Since viruses can be efficiently replicated in human cells, human cell protein translation system must be capable of efficient translation of viral proteins(48); (3) Codon usage and amino acid frequencies of human proteins and viral proteins synthesized in human cells are generally determined by the expression levels of tRNAs with appropriate anticodons in human cells(50,51). Second, differences in amino acid preferences on positions in the flanking regions can be observed between the viral and tumor epitopes. Although the epitope positions were considered in the flanking regions for both the N-terminal and C-terminal enzyme cleavage sites, it is possible that amino acid restriction in the epitope positions for HLA(52-54) and TAP binding(13,14) may override enzymatic cleavage influences at both ends. Third, and most important, there are significant differences between these two groups of epitopes in positions N1 (Pn1), E1 (Pn1′), and C1 (Pc1′), suggesting that there are differences in the proteolytic enzymes involved in the generation of these two groups of epitopes.

Usage of amino acid residues in the tumor antigen epitopes was less diversified than that of the viral epitopes

As shown in Table II, in positions E2, E3, and E9, amino acid preference was conserved in the two groups of epitopes. The results of this conservation in the primary HLA-A anchor residues, on E2 and E9, and the secondary anchor residue, on E3, corresponded to prior reports emphasizing the dominant structural requirement for HLA-A2.1 binding, which corresponded to the previous findings (e.g., Leu or Met at position E2, and Val, Leu, or Ile at position E9 of the epitope regions)(52-54). It should be noted that an auxiliary anchor at E3 usually fine-tunes peptide recognition(55,56). In our study, the high restriction in both tumor antigen and viral antigen epitopes served as an appropriate positive control for the quality of our analyses. The significant differences between the two groups of epitopes in positions E4, E6, E7, and E8 suggest that HLA-A2.1 binding(52-54) and TAP binding(13,14) do not have high restriction in these positions. The results also suggest that the differences in these positions between the two groups of epitopes may reflect variances in enzyme recognition in the flanking regions, since the epitope region serves as the C-terminal flanking region for N-terminal cleavage, as does the N-terminal flanking region for C-terminal cleavage. Future work will also need to examine whether the structural features in the auxiliary anchor positions contribute to lower binding avidity between interaction of MHC/self-tumor antigen peptides and T cell antigen receptor (TCR) and higher binding avidity between MHC/viral peptides and TCR(4).

We observed that there are no sequence structures for proteasome cleavage sites inside the epitopes, as defined by two hydrophobic residues at the E2 and E9 positions. This finding suggests that the epitope candidates having both the right HLA-A2.1 anchor residues on E2 and E9 and the internal cleavage sites should have been degraded, and there is no opportunity for these epitopes to be presented. Similarly, there are no sequence structures for HLA-A2.1 anchor residues and proteasome cleavage sites within the 10 amino acid residues in the epitope N-terminal and C-terminal flanking regions, suggesting that a special feature of the flanking regions is to enable the efficient processing of epitopes.

The amino acid pairs covering the cleavage sites in the generation of tumor antigen epitopes were different from those of the viral antigen epitopes

Protein sequences encode more structural and functional information than amino acid occurrence frequencies. In order to further explore this difference, we compared occurrence frequencies of the amino acid pairs(17) in the Pn1 (N1)-Pn1′ (E1) and the Pc1 (E9)-Pc1′ (C1) of the tumor and viral epitopes. These positions were selected to compare amino acid pairs because they are primary structural features for enzyme recognition and cleavage(40). We argued that if the proteasomes and peptidases that process the immunodominant tumor epitopes are the same as or similar to that processing the immunodominant viral epitopes, the amino acid pairs that are identical in these two groups of epitopes would be in high percentages in these positions. As shown in Figs. 2B and 2C, the occurrence of amino acid pairs in N-terminal cleavage sites of the tumor epitopes was radically different from that of the viral antigen epitopes. In the N-terminal cleavage site (the Pn1-Pn1′ pair, depicted in Fig. 2B), out of 400 possible pairs, we found that 81% of the total pairs in the tumor epitopes and 79% of the total pairs in the viral epitopes were different. In Fig. 2C, the most frequently occurring 13 pairs consisted of 34.0% of tumor epitopes. The most frequently occurring 12 pairs in the viral epitopes covered 32.7% of the epitope group. In addition, the frequency of pairs with the basic amino acid at the Pn1 position of the pairs was significantly increased, from 10.6% of the most frequently occurring tumor epitopes, to 17.3% of the most frequently occurring viral epitopes, suggesting that increased trypsin-like activity mediates the processing of viral epitopes(10). Moreover, the frequency of pairs with hydrophobic amino acid at the Pn1 position of the pairs was significantly decreased, from 17.0% of the most frequently occurring tumor epitopes, to 7.7% of the most frequently occurring viral epitopes, suggesting that an increase in chymotrypsin-like activities is responsible for tumor epitope processing(10). Finally, the frequency of pairs with the basic amino acid at the C-position of the pairs was significantly decreased, from 21.3% of the most frequently occurring tumor epitopes, to 3.9% of the most frequently occurring viral epitopes. Again, these results suggest that the N-terminal cleavages of both the tumor and viral epitopes are mediated by two different groups of enzymes.

Similarly, the occurrence of amino acid pairs in the C-terminal cleavage sites of the tumor antigen epitopes (the Pc1-Pc1' pair) was different from that of the viral antigen epitopes. In Fig. 2B, 53% of the total pairs in the tumor epitopes did not share with 50% of the total pairs in the viral epitopes. The most frequently occurring 12 pairs covered 51.1% of tumor epitopes, and the most frequently occurring 14 pairs covered 51.9% of viral epitopes. In Fig. 2C, among those most frequently occurring pairs, three pairs were conserved between the tumor and viral epitope groups, comprising only 11.5% of the C-terminal cleavage sites. Moreover, differences in the physical features of amino acid pairs in the tumor epitopes versus the viral epitopes were less obvious in the C-terminal cleavage sites, in comparison with the N-terminal cleavage sites. Hydrophobic residues were present in most Pc1 positions of the most frequently occurring pairs, suggesting that chymotrypsin-like activity may be dominant in the processing of C-terminal cleavages of tumor and viral epitopes(10) in addition to the HLA-A2.1 binding preference at the position Pc1/E9 (52-54) and the TAP binding preference at this position(13, 14). These results suggest that pairs in the C-terminal sites are less diversified than those in the N-terminal sites. Our findings show that the proteolytic enzymes generating C-terminal cleavage sites of tumor and viral epitopes are also different.

Furthermore, in order to determine whether the percentages of amino acid pair overlapped in the N-terminal site and C-terminal site in the set of 47 tumor antigen epitopes and the set of 52 viral epitopes are statistically significant, we performed the computation for the probability of amino acid pairs overlapped in two steps. First, the probability mass function (pmf) for all the random overlaps of amino acid pairs in two sets of epitopes was analyzed with the surface plot (Fig. 2D), which had the visual demonstration of the probability distribution of random

overlaps of amino acid pairs in two sets of epitopes The results in Fig. 2D showed that the probability for the 95% of all the random overlaps of amino acid pairs (the "mountain area" in the surface plot) in the two sets of epitopes was ≥ 0.004; and the overlaps of amino acid pairs in two sets of epitopes with the probability < 0.004 were not random. The results suggested that higher percentages of amino acid pairs overlapped in two sets, having the probability < 0.004, reflected the shared specificities between the enzymes of cleaving tumor epitopes and the enzymes of cleaving viral epitopes to some extent. Second, the same analysis on the probability mass function was performed using the contour plot (Fig. 2E). The results showed that the amino acid pairs overlapped in the two sets of epitopes along the same probability contour line had the same probability. The 95% of all the random overlaps of amino acid pairs between tumor epitopes and viral epitopes was within (≥) the 0.004 probability contour loop, which was designated as 95% confidential interval (CI) loop. Once again, the overlaps of amino acid pairs in two sets of epitopes with the probability < 0.004, outside of the 95% CI loop were not random. Interestingly, the probability for the overlapped amino acid pairs in the N-terminal cleavage sites (Fig. 2B) was 3.3×10^{-3}; the probability for the overlapped amino acid pairs in the C-terminal sites (Fig. 2B) was 3.4×10^{-12}. Therefore, both probabilities were smaller than 0.004. In biochemical terms, if no enzymatic specificity is shared in cleavage of the N-terminal sites or C-terminal sites for two sets (in totally random conditions), the probability to have amino acid pairs overlapped in these cleavage sites of two sets of epitopes should be ≥ 0.004 (Fig. 2E). As shown in Fig. 2E, since the probabilities to have amino acid pairs overlapped at the N-terminal site and at the C-terminal site were not random, the enzymatic specificities for cleaving the amino acid pairs were required to share about 20 % to 50% amino acid pair overlaps (Fig. 2B) in two sets of epitopes. The enzyme(s) for cleaving the amino acid pairs in the C-terminal site (50% amino acid overlaps) shared more specificities than that in the N-terminal site (about 20 % amino acid pair overlaps) (Fig. 2B). Since the probability for overlaps of amino acid pairs in the N-terminal sites in two sets was near the 95% confidential interval loop for random overlaps, the shared specificities for the N-terminal sites for two sets were minimal. If the percentages for overlaps of amino acid pairs in the N-terminal sites and C-terminal sites are 100% or close to 100%, the proteolytic enzymes mediating cleavage in these two sets of epitopes should be the same. Since the percentages for overlaps of amino acid pairs in the N-terminal sites and C-terminal sites were far less than 100%, the cleavages in both sites are mediated by two different groups of enzymes.

Statistically significant differences were found in the proteasome cleavage probability of tumor versus viral epitopes

We further examined whether there are any differences between the tumor antigen and viral antigen epitopes by employing the commonly adopted algorithms for the prediction of processing probability by proteasome and immunoproteasome. Our reason for choosing the following four algorithms, the MAPPP (http://www.mpiib-berlin.mpg.de/MAPPP/cleavage.html)(42), the MHC-Pathway constitutive proteasome (http://www.mhc-pathway.net/), the MHC-Pathway immunoproteasome, and the NetChop3.0 neural network predictor (http://www.cbs.dtu.dk/services/NetChop/) (44) rather than other algorithms, such as PAProC (www.paproc.de)(57), is that the former four algorithms allow quantitative prediction. Of note, experimental methods chosen for experiments are not due to their perfection. Likewise, the chosen algorithms may not have 100% prediction efficiency. For any given tumor antigens, these four algorithms may predict some common cleavage sites as well as the different sites. However, using the same sets of algorithms to analyze tumor epitopes and viral epitopes, the results from both sets of epitopes are statistically comparable; the common sites predicted by all four algorithms may reflect the common features of proteasome cleavage revealed from different angles. As shown in Figs. 3A and B, predicted by the algorithm MAPPP, for example, the mean ± 1.96 standard error (SE) (95% confidence interval, CI) of the proteasome cleavage

scores, for tumor antigen epitopes were between 0.62 and 0.75. In contrast, the 95% CI of the scores for viral antigen epitopes predicted with the same algorithm was between 0.53 and 0.61. To further consolidate this finding, we applied the similar analyses with three other algorithms for proteasome peptide cleavage, including the NetChop 3.0 and MHC-Pathway constitutive proteasome and MHC-Pathway immunoproteasome. As shown in Figs. 3A and B, the results obtained were similar to that achieved with the algorithm MAPPP. Of note, the scores of both sets of antigen epitopes predicted with the algorithm MHC-Pathway immunoproteasome were higher than that predicted with the algorithm MHC-Pathway constitutive proteasome. With all four algorithms, there were no overlaps in the 95% CI of the predictive scores for these two groups of antigen epitopes, and a two-sided $p<0.05$, based on the Wilcoxon rank-sum test, suggested that tumor antigen epitopes are different from viral epitopes regarding their probability of being processed by proteasomes and immunoproteasomes ($p<0.05$). The results suggested that potential tumor antigen epitopes could not be processed efficiently. The lower probability that viral epitopes will be processed by these proteasomes suggests that the "threshold" for processing tumor antigen epitopes by proteasomes and immunoproteasomes is higher than that for viral antigen epitopes.

In Fig. 3B, the Wilcoxon rank-sum test comparison of the TAP-binding potential (predicted by the algorithm TAPPred)(15) of tumor antigen epitopes and viral antigen epitopes showed that the viral epitopes had a slightly higher potential than the tumor epitopes to bind to TAP for transfer into ER—even though there were no statistical differences ($p>0.05$). Furthermore, a comparison of the HLA-A2.1-binding potential of the tumor antigen and viral epitopes— predicted by the algorithms SYF [SYFPEITHI](58) and BIMAS [BIMAS/NIH](54)— demonstrated a similarity in the two groups of epitopes, suggesting that HLA-A2.1 restriction had overcome the differences in amino acid occurrences and physical characteristics. These results correlated with the fact there are only one type of HLA-A2.1 and limited human TAP polymorphism(13,59) for binding of all the epitopes regardless of tumor antigens or viral antigens. Based on analyses of the experimentally identified 47 tumor antigen epitopes and 52 viral antigen epitopes, generation of a 95% CI for the predictive scores by using HLA binding algorithms and the TAP algorithm, demonstrated—for the first time—predictive score ranges, with statistical confidence. A recent study reported that prediction with these algorithms could not always be verified with experimental data(16,60,61). However, our analytic methods and results on the 95% CIs for the experimentally identified epitopes have proven useful in the selection of predicted epitopes for further experimental verification.

DISCUSSION

In this study, our results have demonstrated, for the first time, that: 1) the major difference between immunodominant tumor antigen epitopes and viral antigen epitopes lies in the structural features of proteolytic sites, but not in that for TAP binding and HLA binding; and 2) the proteasomes and related peptidases, at least the cleaving efficiencies of those enzymes, in the generation of HLA-A2.1-restricted tumor antigen epitopes may be different from those of viral antigen epitopes. Our future work may further expand this study to compare these two groups of epitopes presented by other MHC alleles when more documented epitopes are reported. Previous studies showed that viral antigen epitopes are preferentially processed by immunoproteasomes, while most tumor antigen epitopes are processed by constitutive proteasomes(62). Since our results emphasize the critical role of proteasomes and related peptidases in the regulation of tumor antigen epitope processing and anti-tumor immune responses, the mechanisms underlying novel therapy of proteasome inhibition(63) in the modulation of tumor antigen epitope processing and anti-tumor immune responses remain an interesting topic in this field of research.

Our results diverged from the previously published analysis of 274 epitope flanking regions (17), which did not identify differences in amino acid occurrence in epitope flanking regions. The discrepancy is likely due to the fact that the previous study only collected naturally processed peptides. The non-immunodominant epitopes that can be eluted from HLA may not join anti-tumor immune responses but may be functional in maintaining T cell repertoire(64, 65). Therefore, the summarization of all diverted antigen peptides with potential immunodominance and that without immunodominance might average-out potential differences.

Of note, the noted differences could be caused by viral interference with antigen processing (66). However, recent reports on proteasomes and related proteases of antigen processing also support and explain our findings(62). In viral infections, expression of IFN-γ is induced by cytokines IL-2, IL-18, and IFN-α/β, or by stimulation through TCR or natural killer (NK) cell receptors(6). Consequently, IFN-γ alters proteasome activity quantitatively, by incorporating three immunosubunits, LMP2 (iβ1), LMP7 (iβ5), and MECL-1 (iβ2), to replace the constitutive β1 (δ), β2 (MB1), and β5 (Z) subunits in 20S core proteasome. Thus, two types of proteasome exist, "constitutive proteasomes," which are present in all somatic cells, and "immunoproteasomes," which are expressed under the influence of cytokines such as IFN-γ (7). In addition, IFN-γ also upregulates the expression of two other proteins, PA28α and PA28β, which form the heptameric proteasome activator complex PA28(67). In contrast with virally infected cells, in a large number of tumors the expression of IFN-γ-induced proteasome subunits LMP2 and LMP7 is downregulated(68), suggesting that processing of tumor antigen epitopes may be different from those of viral antigen epitopes(7). In conjunction with these findings, our results, via analyzing the substrates for proteolytic enzymes with unique bioinformatic approach, indicate that half of C-terminal cleavage sites in the tumor antigen epitopes are not shared by the viral epitopes, suggesting the possibility that more proteasomes than immunoproteasomes mediate cleavage of the C-terminus of tumor antigen epitopes; whereas, more immunoproteasomes are involved in production of the C-terminus of viral antigen epitopes. In addition to proteasomes, the parallel system potentially contributing to differences in the generation of tumor antigen and viral antigen epitopes is tripeptidyl peptidase II (TRPII; EC3.4.14.10), which is able to generate the HLA-restricted HIV Nef epitope independently of proteasomes(69). This parallel system may also contribute to differences between the two groups in the C-terminal cleavage sites_although we were unable to quantify the contributory percentages of processing by proteasomes/immunoproteasomes and other peptidases.

Our results indicate that, in comparison with the C-terminal cleavage sites of the epitopes, the differences in N-terminal cleavage sites between the two groups of epitopes were even larger. In addition, we found that the most frequently occurring amino acid pairs in both N-terminal (Pn1-Pn1′) and C-terminal cleavage sites (Pc1-Pc1′) in the generation of tumor antigen epitopes are different from those of the viral antigen epitopes. It has been reported that many antigenic peptides are generated as amino-terminal extended precursor peptides, and that these require amino-terminal trimming by aminopeptidases located either in the cytosol or in the endoplasmic reticulum (ER)(8,70-72). It should be noted that the expression of both cytosolic leucine aminopeptidase (EC3.4.11.1)(8) and ER aminopeptidase I (ERAP1)(73) is IFN-γ induced, suggesting that such expression may be involved more in the generation of viral antigen rather than tumor antigen epitopes, which is similar to the function of immunoproteasomes(7). In conjunction with these reports, our results showed that only one fifth of the N-terminal cleavage sites of the tumor antigen epitopes overlap with those of the viral antigen epitopes. In sum, we speculate that the differences between the two groups of epitopes may result from different enzymatic activities; it is more likely that the N-terminal cleavage sites of viral epitopes would be generated by IFN-γ-induced aminopeptidases; whereas, the N-terminus of the tumor antigen epitopes is most likely cleaved by IFN-γ

insensitive aminopeptidases. It is well documented that IFN-γ plays a critical role in mounting anti-tumor immune responses(74). Along with this finding, our bioinformatic results with statistical significance suggest that future tumor antigen-specific immunotherapy complemented by IFN-γ may enhance the processing and presentation of MHC class I-restricted tumor antigen epitopes via upregulation of immunoproteasomes and IFN-γ-induced peptidases.

Acknowledgments

REFERENCES

1. Finn OJ. Cancer vaccines: Between the idea and the reality. Nat. Rev. Immunol 2003;3:630. [PubMed: 12974478]

2. Yang F, Yang XF. New concepts in tumor antigens: Their significance in future immunotherapies for tumors. Cell Mol. Immunol 2005;2:331. [PubMed: 16368059]

3. Lollini PL, Cavallo F, Nanni P, Forni G. Vaccines for tumour prevention. Nat. Rev. Cancer 2006;6:204. [PubMed: 16498443]

4. De Visser KE, Schumacher TN, Kruisbeek AM. Cd8+ t cell tolerance and cancer immunotherapy. J. Immunother 2003;26:1. [PubMed: 12514424]

5. Van Der Bruggen P, Van Den Eynde BJ. Processing and presentation of tumor antigens and vaccination strategies. Curr. Opin. Immunol 2006;18:98. [PubMed: 16343880]

6. Malmgaard L. Induction and regulation of ifns during viral infections. J. Interferon Cytokine Res 2004;24:439. [PubMed: 15320958]

7. Kloetzel PM, Ossendorp F. Proteasome and peptidase function in mhc-class-i-mediated antigen presentation. Curr. Opin. Immunol 2004;16:76. [PubMed: 14734113]

8. Beninga J, Rock KL, Goldberg AL. Interferon-gamma can stimulate post-proteasomal trimming of the n terminus of an antigenic peptide by inducing leucine aminopeptidase. J. Biol. Chem 1998;273:18734. [PubMed: 9668046]

9. Rock KL, York IA, Goldberg AL. Post-proteasomal antigen processing for major histocompatibility complex class i presentation. Nat. Immunol 2004;5:670. [PubMed: 15224092]

10. Kloetzel PM. Generation of major histocompatibility complex class i antigens: Functional interplay between proteasomes and tppii. Nat. Immunol 2004;5:661. [PubMed: 15224091]

11. Geier E, Pfeifer G, Wilm M, Lucchiari-Hartz M, Baumeister W, Eichmann K, Niedermann G. A giant protease with potential to substitute for some functions of the proteasome. Science 1999;283:978. [PubMed: 9974389]

12. Luckey CJ, King GM, Marto JA, Venketeswaran S, Maier BF, Crotzer VL, Colella TA, Shabanowitz J, Hunt DF, Engelhard VH. Proteasomes can either generate or destroy mhc class i epitopes: Evidence for nonproteasomal epitope generation in the cytosol. J. Immunol 1998;161:112. [PubMed: 9647214]

13. Uebel S, Kraas W, Kienle S, Wiesmuller KH, Jung G, Tampe R. Recognition principle of the tap transporter disclosed by combinatorial peptide libraries. Proc. Natl. Acad. Sci. U. S. A 1997;94:8976. [PubMed: 9256420]

14. Daniel S, Brusic V, Caillat-Zucman S, Petrovsky N, Harrison L, Riganelli D, Sinigaglia F, Gallazzi F, Hammer J, Van Endert PM. Relationship between peptide selectivities of human transporters associated with antigen processing and hla class i molecules. J. Immunol 1998;161:617. [PubMed: 9670935]

15. Bhasin M, Raghava GP. Analysis and prediction of affinity of tap binding peptides using cascade svm. Protein Sci 2004;13:596. [PubMed: 14978300]

16. Pelte C, Cherepnev G, Wang Y, Schoenemann C, Volk HD, Kern F. Random screening of proteins for hla-a*0201-binding nine-amino acid peptides is not sufficient for identifying cd8 t cell epitopes recognized in the context of hla-a*0201. J. Immunol 2004;172:6783. [PubMed: 15153496]

17. Altuvia Y, Margalit H. Sequence signals for generation of antigenic peptides by the proteasome: Implications for proteasomal cleavage mechanism. J. Mol. Biol 2000;295:879. [PubMed: 10656797]

18. Renkvist N, Castelli C, Robbins PF, Parmiani G. A listing of human tumor antigens recognized by t cells. Cancer Immunol. Immunother 2001;50:3. [PubMed: 11315507]

19. Novellino L, Castelli C, Parmiani G. A listing of human tumor antigens recognized by t cells: March 2004 update. Cancer Immunol. Immunother. 2004

20. Sahin U, Tureci O, Schmitt H, Cochlovius B, Johannes T, Schmits R, Stenner F, Luo G, Schobert I, Pfreundschuh M. Human neoplasms elicit multiple specific immune responses in the autologous host. Proc. Natl. Acad. Sci. U. S. A 1995;92:11810. [PubMed: 8524854]

21. Chen YT. Cancer vaccine: Identification of human tumor antigens by serex [in process citation]. Cancer J. Sci. Am 2000;6(Suppl. 3):S208.

22. Scanlan MJ, Jager D. Challenges to the development of antigen-specific breast cancer vaccines. Breast Cancer Res 2001;3:95. [PubMed: 11250753]

23. Rosenberg SA. Shedding light on immunotherapy for cancer. N. Engl. J. Med 2004;350:1461. [PubMed: 15070799]

24. Chen, Y. Serex review. Cancer Immunity. 2004. http://www.cancerimmunity.org/SEREX/

25. Wu CJ, Yang XF, Mclaughlin S, Neuberg D, Canning C, Stein B, Alyea EP, Soiffer RJ, Dranoff G, Ritz J. Detection of a potent humoral response associated with immune-induced remission of chronic myelogenous leukemia. J. Clin. Invest 2000;106:705. [PubMed: 10974024]

26. Yang XF, Wu CJ, Mclaughlin S, Chillemi A, Wang KS, Canning C, Alyea EP, Kantoff P, Soiffer RJ, Dranoff G, Ritz J. Cml66, a broadly immunogenic tumor antigen, elicits a humoral immune response associated with remission of chronic myelogenous leukemia. Proc. Natl. Acad. Sci. U. S. A 2001;98:7492. [PubMed: 11416219]

27. Yan Y, Phan L, Yang F, Talpaz M, Yang Y, Xiong Z, Ng B, Timchenko NA, Wu CJ, Ritz J, Wang H, Yang XF. A novel mechanism of alternative promoter and splicing regulates the epitope generation of tumor antigen cml66-l. J. Immunol 2004;172:651. [PubMed: 14688378]

28. Yang XF, Wu CJ, Chen L, Alyea EP, Canning C, Kantoff P, Soiffer RJ, Dranoff G, Ritz J. Cml28 is a broadly immunogenic antigen, which is overexpressed in tumor cells. Cancer Res 2002;62:5517. [PubMed: 12359762]

29. Yan Y, Xiong Z, Yang F, Yang Y, Ng B, Lachman L, Wang H, Yang X-F. Broadly immunogenic serex antigen cml66-l elicits humoral and cellular immune responses. FASEB J 2004;18(4):A58.

30. Rosenberg SA. Identification of cancer antigens: Impact on development of cancer immunotherapies. Cancer J. Sci. Am 2000;6(Suppl. 3):S200.

31. Rosenberg SA. Progress in human tumour immunology and immunotherapy. Nature 2001;411:380. [PubMed: 11357146]

32. Manici S, Sturniolo T, Imro MA, Hammer J, Sinigaglia F, Noppen C, Spagnoli G, Mazzi B, Bellone M, Dellabona P, Protti MP. Melanoma cells present a mage-3 epitope to cd4(+) cytotoxic t cells in association with histocompatibility leukocyte antigen dr11. J. Exp. Med 1999;189:871. [PubMed: 10049951]

33. Van Der Bruggen P, Zhang Y, Chaux P, Stroobant V, Panichelli C, Schultz ES, Chapiro J, Van Den Eynde BJ, Brasseur F, Boon T. Tumor-specific shared antigenic peptides recognized by human t cells. Immunol. Rev 2002;188:51. [PubMed: 12445281]

34. Nakatsura T, Senju S, Ito M, Nishimura Y, Itoh K. Cellular and humoral immune responses to a human pancreatic cancer antigen, coactosin-like protein, originally defined by the serex method. Eur. J. Immunol 2002;32:826. [PubMed: 11870627]

35. Ng B, Yang F, Huston DP, Yan Y, Yang Y, Xiong Z, Peterson LE, Wang H, Yang XF. Increased noncanonical splicing of autoantigen transcripts provides the structural basis for expression of untolerized epitopes. J. Allergy Clin. Immunol 2004;114:1463. [PubMed: 15577853]

36. Corbet S, Nielsen HV, Vinner L, Lauemoller S, Therrien D, Tang S, Kronborg G, Mathiesen L, Chaplin P, Brunak S, Buus S, Fomsgaard A. Optimization and immune recognition of multiple novel conserved hla-a2, human immunodeficiency virus type 1-specific ctl epitopes. J. Gen. Virol 2003;84:2409. [PubMed: 12917462]

37. Logean A, Rognan D. Recovery of known t-cell epitopes by computational scanning of a viral genome. J. Comput. Aided Mol. Des 2002;16:229. [PubMed: 12400854]

38. Shirai M, Arichi T, Nishioka M, Nomura T, Ikeda K, Kawanishi K, Engelhard VH, Feinstone SM, Berzofsky JA. Ctl responses of hla-a2.1-transgenic mice specific for hepatitis c viral peptides predict epitopes for ctl of humans carrying hla-a2.1. J. Immunol 1995;154:2733. [PubMed: 7533182]

39. Gianfrani C, Oseroff C, Sidney J, Chesnut RW, Sette A. Human memory ctl response specific for influenza a virus is broad and multispecific. Hum. Immunol 2000;61:438. [PubMed: 10773346]

40. Schechter I, Berger A. On the size of the active site in proteases. I. Papain. Biochem. Biophys. Res. Commun 1967;27:157. [PubMed: 6035483]

41. Lohmuller T, Wenzler D, Hagemann S, Kiess W, Peters C, Dandekar T, Reinheckel T. Toward computer-based cleavage site prediction of cysteine endopeptidases. Biol. Chem 2003;384:899. [PubMed: 12887057]

42. Holzhutter HG, Frommel C, Kloetzel PM. A theoretical approach towards the identification of cleavage-determining amino acid motifs of the 20 s proteasome. J. Mol. Biol 1999;286:1251. [PubMed: 10047495]

43. Tenzer S, Peters B, Bulik S, Schoor O, Lemmel C, Schatz MM, Kloetzel PM, Rammensee HG, Schild H, Holzhutter HG. Modeling the mhc class i pathway by combining predictions of proteasomal cleavage, tap transport and mhc class i binding. Cell Mol. Life Sci 2005;62:1025. [PubMed: 15868101]

44. Kesmir C, Nussbaum AK, Schild H, Detours V, Brunak S. Prediction of proteasome cleavage motifs by neural networks. Protein Eng 2002;15:287. [PubMed: 11983929]

45. Bertsekas, DP.; John N., Tsitsiklis. Introduction to probability. Athena Scientific Publishing Corp.; Nashua: 2002.

46. Rosner, B. Hypothesis testing: One-sample inference. Rosner, B., editor. Duxbury Publishing Corp.; Australia, Canada, Mexico, Singapore, Spain, United Kingdom, United States: 2000. p. 211

47. Ellington, AACJ. Characteristics of amino acids. In: Ausubel, FM., et al., editors. Current Protocols in Molecular Biology. John Wiley & Sons, Inc.; Indianapolis: 1997. p. A.1C.1

48. Knipe, D.; Howley, PM. Fundamental virology. Lippincott Williams & Wilkins Publishing Corp.; Philadelphia, Baltimore, New York, London, Buenos Aires, Hong Kong, Sydney, Tokyo: 2001.

49. Rosner, B. Estimation of sample size and power for comparing two means. In: Rosner, B., editor. Fundamentals of Biostatistics. Duxbury Publishing Corp.; Australia, Canada, Mexico, Singapore, Spain, United Kingdom, United States: 2000. p. 307

50. Akashi H. Gene expression and molecular evolution. Curr. Opin. Genet. Dev 2001;11:660. [PubMed: 11682310]

51. Agris PF. Decoding the genome: A modified view. Nucleic Acids Res 2004;32:223. [PubMed: 14715921]

52. Falk K, Rotzschke O, Stevanovic S, Jung G, Rammensee HG. Allele-specific motifs revealed by sequencing of self-peptides eluted from mhc molecules. Nature 1991;351:290. [PubMed: 1709722]

53. Hunt DF, Henderson RA, Shabanowitz J, Sakaguchi K, Michel H, Sevilir N, Cox AL, Appella E, Engelhard VH. Characterization of peptides bound to the class i mhc molecule hla-a2.1 by mass spectrometry. Science 1992;255:1261. [PubMed: 1546328]

54. Parker KC, Bednarek MA, Coligan JE. Scheme for ranking potential hla-a2 binding peptides based on independent binding of individual peptide side-chains. J. Immunol 1994;152:163. [PubMed: 8254189]

55. Ruppert J, Sidney J, Celis E, Kubo RT, Grey HM, Sette A. Prominent role of secondary anchor residues in peptide binding to hla-a2.1 molecules. Cell 1993;74:929. [PubMed: 8104103]

56. Madden DR, Garboczi DN, Wiley DC. The antigenic identity of peptide-mhc complexes: A comparison of the conformations of five viral peptides presented by hla-a2. Cell 1993;75:693. [PubMed: 7694806]

57. Nussbaum AK, Kuttler C, Hadeler KP, Rammensee HG, Schild H. Paproc: A prediction algorithm for proteasomal cleavages available on the www. Immunogenetics 2001;53:87. [PubMed: 11345595]

58. Rammensee H, Bachmann J, Emmerich NP, Bachor OA, Stevanovic S. Syfpeithi: Database for mhc ligands and peptide motifs. Immunogenetics 1999;50:213. [PubMed: 10602881]

59. Daniel S, Caillat-Zucman S, Hammer J, Bach JF, Van Endert PM. Absence of functional relevance of human transporter associated with antigen processing polymorphism for peptide selection. J. Immunol 1997;159:2350. [PubMed: 9278325]

60. Doytchinova IA, Walshe VA, Jones NA, Gloster SE, Borrow P, Flower DR. Coupling in silico and in vitro analysis of peptide-mhc binding: A bioinformatic approach enabling prediction of superbinding peptides and anchorless epitopes. J. Immunol 2004;172:7495. [PubMed: 15187128]

61. Doytchinova IA, Flower DR. Quantitative approaches to computational vaccinology. Immunol. Cell Biol 2002;80:270. [PubMed: 12067414]

62. Van Den Eynde BJ, Morel S. Differential processing of class-i-restricted epitopes by the standard proteasome and the immunoproteasome. Curr. Opin. Immunol 2001;13:147. [PubMed: 11228406]

63. Rajkumar SV, Richardson PG, Hideshima T, Anderson KC. Proteasome inhibition as a novel therapeutic target in human cancer. J. Clin. Oncol 2005;23:630. [PubMed: 15659509]

64. Goldrath AW, Bevan MJ. Selecting and maintaining a diverse t-cell repertoire. Nature 1999;402:255. [PubMed: 10580495]

65. Surh CD, Sprent J. Homeostatic t cell proliferation: How far can t cells be activated to self-ligands? J. Exp. Med 2000;192:F9. [PubMed: 10952731]

66. Ehrlich R. Modulation of antigen processing and presentation by persistent virus infections and in tumors. Hum. Immunol 1997;54:104. [PubMed: 9297529]

67. Whitby FG, Masters EI, Kramer L, Knowlton JR, Yao Y, Wang CC, Hill CP. Structural basis for the activation of 20s proteasomes by 11s regulators. Nature 2000;408:115. [PubMed: 11081519]

68. Seliger B, Wollscheid U, Momburg F, Blankenstein T, Huber C. Characterization of the major histocompatibility complex class i deficiencies in b16 melanoma cells. Cancer Res 2001;61:1095. [PubMed: 11221838]

69. Tomkinson B. Tripeptidyl peptidases: Enzymes that count. Trends Biochem. Sci 1999;24:355. [PubMed: 10470035]

70. Fruci D, Lauvau G, Saveanu L, Amicosante M, Butler RH, Polack A, Ginhoux F, Lemonnier F, Firat H, Van Endert PM. Quantifying recruitment of cytosolic peptides for hla class i presentation: Impact of tap transport. J. Immunol 2003;170:2977. [PubMed: 12626550]

71. Knuehl C, Spee P, Ruppert T, Kuckelkorn U, Henklein P, Neefjes J, Kloetzel PM. The murine cytomegalovirus pp89 immunodominant h-2ld epitope is generated and translocated into the endoplasmic reticulum as an 11-mer precursor peptide. J. Immunol 2001;167:1515. [PubMed: 11466372]

72. Fruci D, Niedermann G, Butler RH, Van Endert PM. Efficient mhc class i-independent amino-terminal trimming of epitope precursor peptides in the endoplasmic reticulum. Immunity 2001;15:467. [PubMed: 11567636]

73. Saric T, Chang SC, Hattori A, York IA, Markant S, Rock KL, Tsujimoto M, Goldberg AL. An ifn-gamma-induced aminopeptidase in the er, erap1, trims precursors to mhc class i-presented peptides. Nat. Immunol 2002;3:1169. [PubMed: 12436109]

74. Bonehill A, Heirman C, Thielemans K. Genetic approaches for the induction of a cd4+ t cell response in cancer immunotherapy. J. Gene Med 2005;7:686. [PubMed: 15693037]

N - Cleavage Site   C - Cleavage Site

| N-Terminal Flanking Region | Epitope | C-Terminal Flanking Region |

A.   Pn10 ----------- Pn2 Pn1   Pn1' Pn2' ------ Pc2 Pc1   Pc1' Pc2' ----------- Pc10'
B.   N10 ------------- N2 N1   E1 E2 ----------- E9   C1 C2 ------------- C10
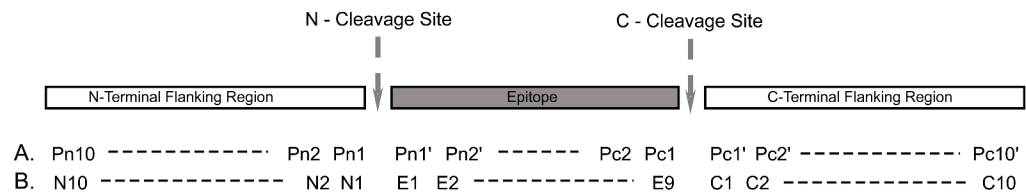
Fig. 1. Position nomenclature of the epitopes and the flanking regions
A. Schechter and Berger's enzymatic cleavage nomenclature is used in reference to the amino acid positions in the N-terminal and C-terminal flanking regions, relative to their respective cleavage sites. **B.** Two cleavage sites are required during the final step of antigen epitope generation. Since there is an overlap between the C-terminal flanking region of the epitope's N-terminal cleavage site and the N-terminal flanking region of the epitope's C-terminal cleavage site, in order to avoid potential confusion, the amino acids in the epitope (E1 to E9), the N-terminal flanking region (N10-N1), and the C-terminal flanking region (C1-C10) are uniquely defined.

A.

| r_1 | $k_1 - r_1$ | **Set 1 (tumor antigen epitopes)** |

Overlap probabilities

| r_2 | $k_2 - r_2$ | **Set 2 (viral antigen epitopes)** |

B.

N-terminal cleavage site
(amino acid pair: Pn1-Pn1')

81%   19%   Tumor epitopes

21%   79%   Virus epitopes

Overlapped amino acid pairs

C-terminal cleavage site
(amino acid pair: Pc1-Pc1')

53%   47%   Tumor epitopes

50%   50%   Virus epitopes
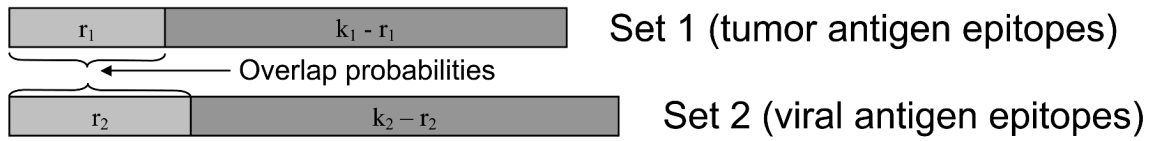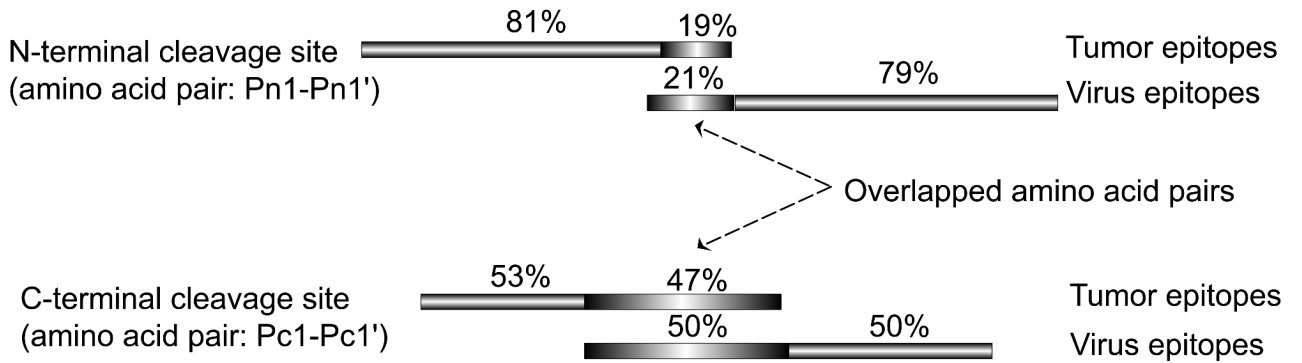
C.

Pn1-Pn1' (N1 - E1)[#]				Pc1-Pc1' (E9 - C1)			
Amino acid		Tumor	Virus	Amino acid		Tumor	Virus
* W	K	2.13		L	C	4.26	
R	K	4.26		V	T	6.38	5.77
R	R	4.26		L	H	4.26	
C	S	2.13		L	P	6.38	
R	H	2.13		L	L	8.51	
M	K	2.13		C	F	2.13	
L	R	4.26		V	Q	4.26	
V	H	2.13		C	P	2.13	
Q	F	2.13		V	C	2.13	1.92
Q	Q	2.13		V	E	4.26	3.85
I	Y	2.13		V	V	4.26	
M	G	2.13		Y	T	2.13	
T	Y	2.13		V	P		7.69
H	N		3.85	V	S		5.77
C	F		1.92	V	G		5.77
K	W		1.92	L	Y		3.85
Q	G		3.85	V	W		1.92
W	V		1.92	M	T		1.92
K	K		3.85	A	C		1.92
C	I		1.92	V	H		1.92
S	W		1.92	L	T		3.85
R	V		3.85	V	M		1.92
K	S		3.85	V	L		3.85
N	F		1.92				
F	N		1.92				

#The frequency of amino acid pair in protease cut site (%)
* Physical characters of amino acid in colors:
      Acidic: Underline; Basic: Italic; Hydrophobic: Bold; Neutral: Regular

D

95% Confidential Interval
for non-specific overlaps

E.
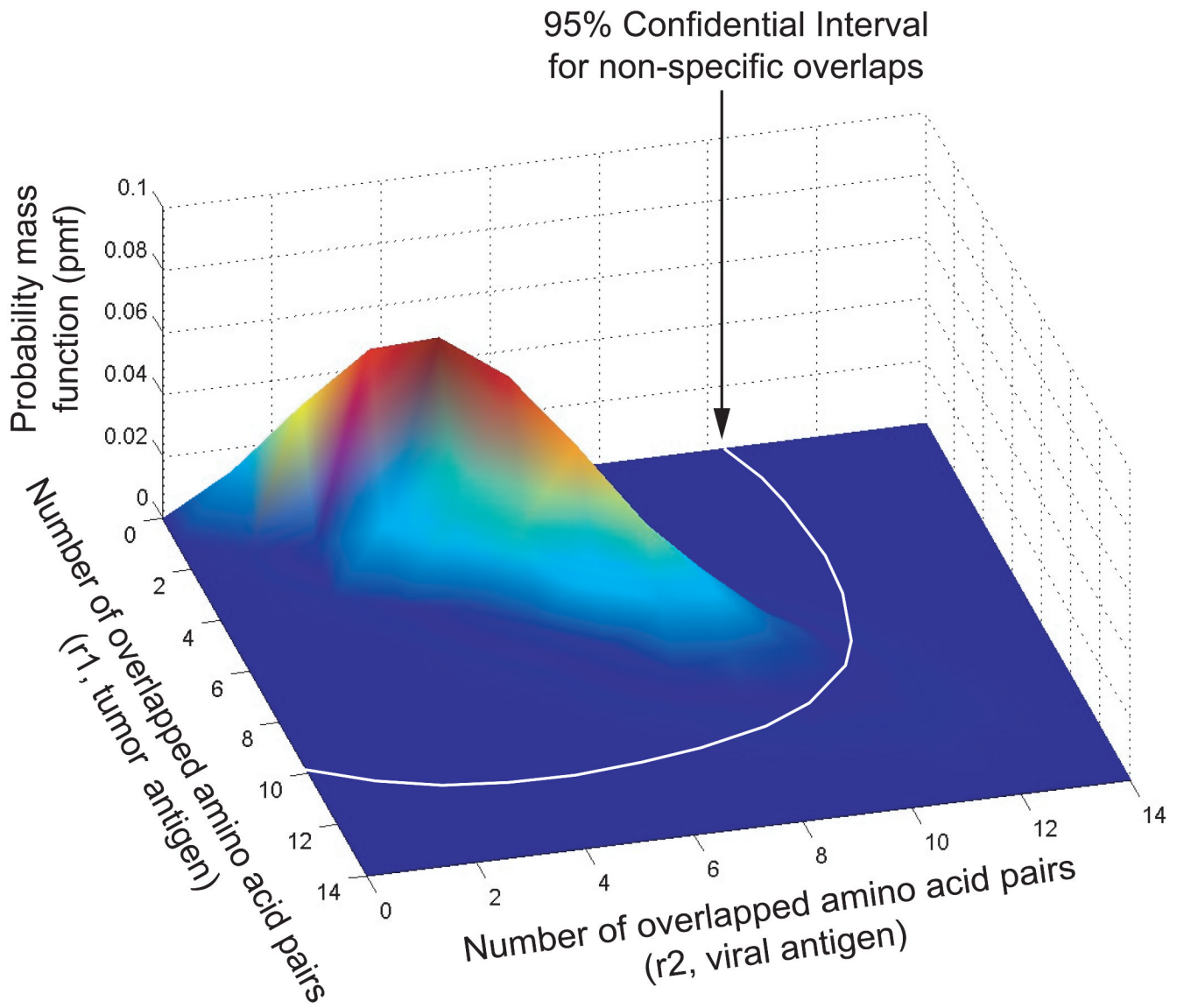
Fig. 2. Comparison of the amino acid pairs located in the N-terminal cleavage sites (Pn1-Pn1′) and the C-terminal cleavage sites (Pc1-Pc1′) in the tumor antigen and viral antigen epitopes

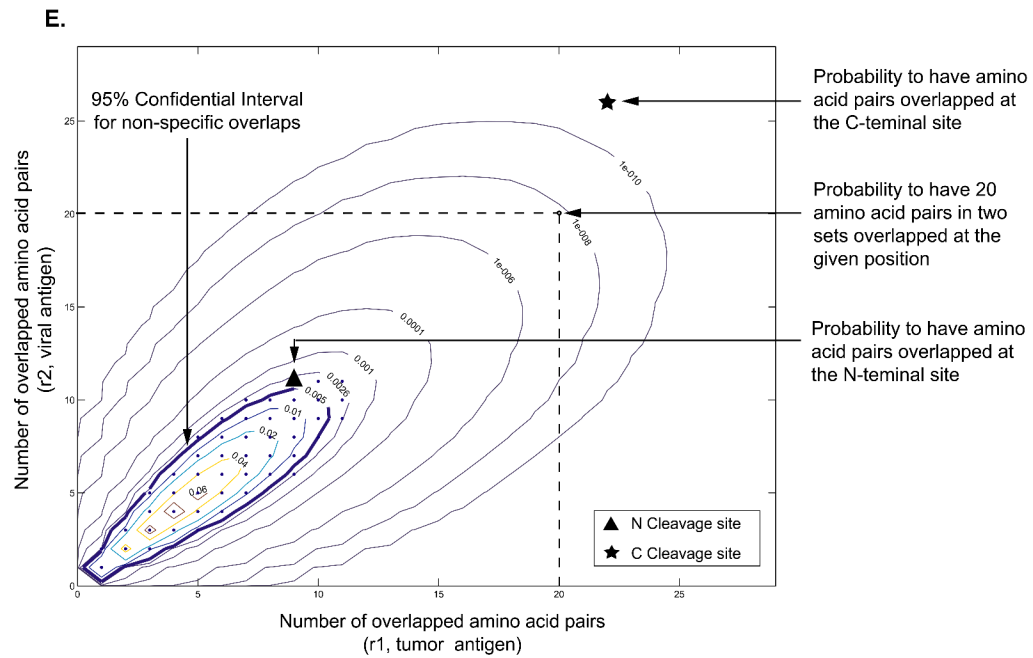**A.** Schematic representation of the notation for the two overlapping sets of amino acid pairs. For the computation of probability mass function, the overlapped amino acid pairs in set 1 (tumor epitopes) are assigned as r_1, the corresponding remainders (nonoverlapping) in the same set are assigned as k_1-r_1. Similarly, the overlapped amino acid pairs (nonoverlapping) in set 2 (viral epitopes) are assigned as r_2, the corresponding remainders in the same set are assigned as k_2-r_2. **B.** Schematic representation of the percentages of identical amino acid pairs versus the different amino acid pairs in the tumor antigen and viral antigen epitopes. Since identical amino acid pairs can be present more than once, percentages for the two groups are slightly different. Note that there are twenty amino acids; thus, theoretically, there are 400 possible amino acid pairs. **C.** The frequently occurring amino acid pairs (i.e., those occurring more than once) located in the N-terminal cleavage sites (Pn1-Pn1′) and the C-terminal cleavage sites (Pc1-Pc1′) in the tumor antigen and viral antigen epitopes. Percentages of the amino acid pairs in the tumor epitope or the viral epitope group are presented as frequencies. The physical characteristics of the amino acids are indicated through the use of different font formats: acidic, underline; basic, italic; hydrophobic, bold; and neutral, grey. **D.** The surface plot of probability mass function. The ordinate indicates the number of overlapped amino acid pairs (r2) in the viral epitopes, the abscissa indicates the number of overlapped amino acid pairs (r1) in the tumor epitopes, and the Z-axis indicates the probability mass function. The probability mass function demonstrates the distribution of the probability for any given overlaps of amino acid pairs in the N-terminal cleavage sites and C-terminal sites in two sets. The curve line indicates the probability for the 95% of all the random overlaps of amino acid pairs (95% confidential interval). **E.** The contour plot of probability mass function. The events sharing the probability are linked by a contour curve. The ordinate indicates the number of overlapped amino acid pairs (r1) in the tumor epitopes, and the abscissa indicates the number of overlapped amino acid pairs (r2) in the viral epitopes. A loop with the probability of 0.005 indicates the probability for the 95% of all the random overlaps of amino acid pairs (95% confidential interval), which corresponds to the curve line presented in Fig. 2D. The probabilities to have the amino acid pairs overlapped at the N-terminal sites and C-terminal sites in two sets of epitopes (Fig. 2B) are marked in the plot by a triangle and a star, respectively.
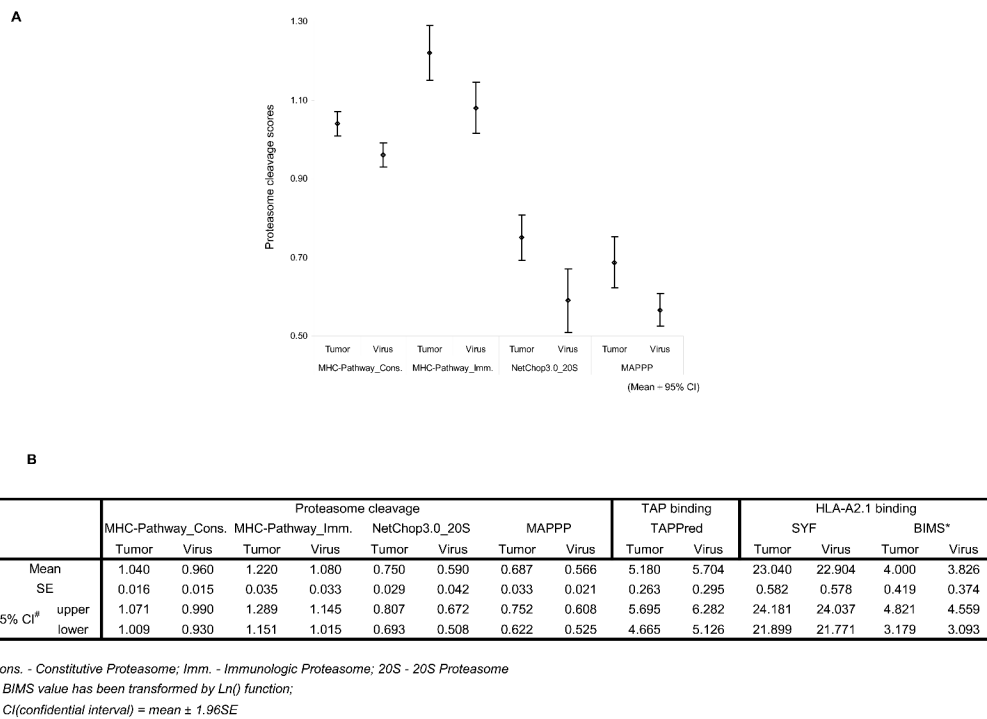
A

B

		Proteasome cleavage								TAP binding		HLA-A2.1 binding			
		MHC-Pathway_Cons.		MHC-Pathway_Imm.		NetChop3.0_20S		MAPPP		TAPPred		SYF		BIMS*	
		Tumor	Virus	Tumor	Virus	Tumor	Virus	Tumor	Virus	Tumor	Virus	Tumor	Virus	Tumor	Virus
Mean		1.040	0.960	1.220	1.080	0.750	0.590	0.687	0.566	5.180	5.704	23.040	22.904	4.000	3.826
SE		0.016	0.015	0.035	0.033	0.029	0.042	0.033	0.021	0.263	0.295	0.582	0.578	0.419	0.374
95% CI#	upper	1.071	0.990	1.289	1.145	0.807	0.672	0.752	0.608	5.695	6.282	24.181	24.037	4.821	4.559
	lower	1.009	0.930	1.151	1.015	0.693	0.508	0.622	0.525	4.665	5.126	21.899	21.771	3.179	3.093

Cons. - Constitutive Proteasome; Imm. - Immunologic Proteasome; 20S - 20S Proteasome

* BIMS value has been transformed by Ln() function;

 # CI(confidential interval) = mean ± 1.96SE

Fig. 3. Comparison of proteasome cleavage scores, and the binding potential of TAP and HLA-A2.1 in the tumor versus viral epitopes

A. Schematic representation of the differences in predicted proteasome cleavage scores of the tumor and viral epitopes. Prediction of proteasome cleavage scores of the epitope-containing antigen sequences was performed by using the algorithms MHC-pathway constitutive proteasomes, MHC-Pathway immunoproteasome, NetChop3.0, and MAPPP. The mean ± 1.96 SE of the proteasome cleavage scores of the antigen sequences was calculated for a 95% confidence interval (95% CI). The resulting lack of overlap in the 95% CI indicates that there is a difference in proteasome cleavage scores for the tumor versus viral epitopes. **B.** Statistical comparison of proteasome cleavage scores, TAP binding potential, and HLA-A2.1 binding potential of the tumor versus viral epitopes, predicted through commonly used algorithms. The BIMS value of antigen epitopes was transformed by Ln function before the statistical analyses.

Table I

Experimentally identified HLA-A2.1-restricted tumor and virus antigen epitopes

	Antigen	Epitope Starting Position	Accession No.
Tumor	AFP	542	NP_001125
	CEA	605	NP_004354
	CPSF	250	NP_037423
	G250	254	NP_001207
	HSP70-2M	286	NP_005336
	LAGE-1	157	NP_758965
	MAGE-10	254	NP_066386
	MAGE-3	112, 271	NP_005353
	MAGE-A2	112	NP_005352
	MAGE-B2	273	NP_002355
	MC1R	244, 283	NP_002377
	Melan-A	27, 32	NP_005502
	MUC1	12, 210	NP_002447
	MUC2	467, 645	NP_002448
	NY-ESO-1	155	NP_001318
	p53	65, 264	NP_000537
	PRAME	100, 425	NP_006106
	SSX-2	41	NP_783629
	Telomerase	540, 865	NP_937986
	TRP-2	180, 360	NP_001913
	Tyrosinase	369	NP_000363
	WT1	126	NP_077743
	GP100	154, 178, 209, 280, 619, 639	NP_008859
	HER-2	48, 369, 391, 402, 435, 466, 650, 654, 665, 689	NP_004439
HCV	C7A2	132	
	SK1	279	P17398[*]
	SK5	363	
	SK7	723	
HBV	POL	61, 422, 453, 502, 573, 814	P17394[*]
	ENV	177, 249, 250, 335	P17398[*]
Influenza A	NA	18	NP_775534
	NS1	76	NP_775538
	M1	58	NP_775536
	PB2	49, 630	NP_775529
	PB1	94, 162, 410, 413	NP_775530
	PA	46, 70, 225, 284, 594	NP_775531
HIV	Gag	77, 80, 150, 242, 341, 345, 362(9V), 362(9A)	NP_057850

Antigen	Epitope Starting Position	Accession No.
POL	326, 456, 598, 679	NP_789740
Vif	23 (9V), 23 (9I), 149, 158 (2I), 158 (2T)	NP_057851
Env	36, 67(2V), 67(2I), 558, 565	NP_057856
Rev	96(2M), 96(2V)	NP_057854

[*] Data from Swiss-Prot (Expert Protein Analysis System) Database

Table II

The frequencies of amino acids in the epitopes and flanking regions were statistically higher than the background (*p*<0.05)

Position	Tumor Epitope			Virus Epitope		
N10				L		
N9	*S					
N8						
N7						
N6	S					
N5	S					
N4						
N3						
N2						
N1						
E1						
E2	L			L	M	
E3	W			Y		
E4	G					
E5						
E6	I			*H*	M	V
E7	F	*H*	W	F		
E8				C		
E9	L	V		V	I	L
C1				T		
C2						
C3				P		
C4						
C5						
C6	S			I		
C7				M		
C8	L					
C9				M		

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Position	Tumor Epitope	Virus Epitope
C10		

*Physical characteristics of amino acid in font formats: Basic: Italic; Hydrophobic: Bold; Neutral: Regular