Computational characterization of the peptidome in transporter associated with antigen processing (TAP)-deficient cells

Antonio J. Martín-Galiano¹*, Daniel López²*

¹ Unidad de Bioinformática, Centro Nacional de Microbiología, Instituto de Salud Carlos III, Majadahonda, Madrid, Spain, ² Unidad de Procesamiento Antigénico, Centro Nacional de Microbiología, Instituto de Salud Carlos III, Majadahonda, Madrid, Spain

* mgaliano@isciii.es (AJMG); dlopez@isciii.es (DL)

Abstract

The transporter associated with antigen processing (TAP) is a key element of the major histocompatibility complex (MHC) class I antigen processing and presentation pathway. Nonfunctional TAP complexes impair the translocation of cytosol-derived proteolytic peptides to the endoplasmic reticulum lumen. This drastic reduction in the available peptide repertoire leads to a significant decrease in MHC class I cell surface expression. Using mass spectrometry, different studies have analyzed the cellular MHC class I ligandome from TAP-deficient cells, but the analysis of the parental proteins, the source of these ligands, still deserves an in-depth analysis. In the present report, several bioinformatics protocols were applied to investigate the nature of parental proteins for the previously identified TAP-independent MHC class I ligands. Antigen processing in TAP-deficient cells mainly focused on small, abundant or highly integral transmembrane proteins of the cellular proteome. This process involved abundant proteins of the central RNA metabolism. In addition, TAP-independent ligands were preferentially cleaved from the N- and C-terminal ends with respect to the central regions of the parental proteins. The abundance of glycine, proline and aromatic residues in the C-terminal sequences from TAP-independently processed proteins allows the accessibility and specificity required for the proteolytic activities that generates the TAP-independent ligandome. This limited proteolytic activity towards a set of preferred proteins in a TAP-negative environment would therefore suffice to promote the survival of TAP-deficient individuals.

Introduction

The proteasome, as well as other cytosolic proteases, continuously degrades misfolded or prematurely terminated proteins, also named defective ribosomal products (DRiPs), and mature proteins with normal turnover kinetics. This proteolysis generates short peptides that are transported into the endoplasmic reticulum (ER) by the transporter associated with antigen processing (TAP) [1]. In the ER lumen, the multisubunit peptide-loading complex assembles
nascent MHC class I heavy chain, β2-microglobulin and peptides to generate trimolecular stable MHC/peptide complexes that, after export to the cell surface, are recognized by cytolytic CD8+ T lymphocytes (reviewed in [2]). This antigen presentation pathway is the key element in the immune response against viruses and tumors.

Mutations in the TAP genes might generate nonfunctional TAP complexes that subsequently impair the transport of cytosolic peptides to the ER, as described both in mice [3] and humans [4]. Animals and patients with this MHC class I immunodeficiency present a very limited functional CD8+ T cell population. Remarkably, these individuals have a limited predisposition to suffer chronic respiratory bacterial, but not viral, infections or neoplasms and they are asymptomatic for long periods. As cytotoxic CD8+ T cells are required to control and eliminate both malignant and virus-infected cells, their ability to recognize TAP-independent peptide antigens seems to help protect against tumor and viral infections in immunocompromised individuals.

Although TAP-independent viral epitopes were identified decades ago (reviewed in [5–7]), very few studies have analyzed the cellular TAP-independent MHC class I peptidome [8–12]. In these articles, the properties of cellular TAP-independent ligands have been defined using extensive analysis by mass spectrometry analyses. However, the nature of the parental proteins of TAP-independent ligands has remained largely unaddressed. Thus, in the present report, we applied several algorithms to perform an in-depth analysis of the features of the parental proteins for TAP-independent MHC class I ligands identified by mass spectrometry.

**Materials and methods**

**Data acquisition and management**

Protein descriptors and peptide coordinates were obtained from the original references. Sequences were collected using these descriptors from current versions of NCBI [13], IPI (http://ftp.ebi.ac.uk/pub/databases/IPI/last_release/) and UniProt databases [14]. Peptides assigned to proteins that not were found in any of these databases (<0.5%), due to discontinuation or deletion, were not considered in the analysis. Redundant sequences present in different databases, were unified under one single descriptor using a BLAST all-against-all search of the original data and applying the threshold of 100% for both identity and alignment length. The correct positioning of the peptides in their respective protein sequences was verified in all cases, using the original manuscripts.

**Functional bioinformatics procedures**

The isoelectric point and GRAVY index were calculated using the ProtParam tool of EXPASY [15]. The GRAVY index was calculated using the Kyte and Doolittle hydrophobicity scale [16]. Transmembrane helices were predicted with TMHMM 2.0 [17]. Signal peptides were predicted with SignalP 4.1 by selecting eukaryotes as the organism group [18]. Gene Ontology data were downloaded from GO web [19]. Human gene associations were downloaded from http://www.ebi.ac.uk/GOA/human_release. RNA-Seq data from the spleen were downloaded from RNA-Seq Atlas [20]. Proteins in the TAP dataset were compared to gene products in the RNA-Seq Atlas using BLAST [21], and proteins showing >95% identity and >95% length were considered identical. The human proteome was downloaded from the UniProt database (http://www.uniprot.org/proteomes/UP000005640). Proteins were associated with a given dataset (TAP+, TAP C, TAP NC or HLA-II) when they were strictly contained in the dataset or the ratio to the observed to expected counts, considering the dataset size with respect to the total, was greater than 2. Cleavage statistics and residue composition were assessed using in-house Perl scripts.
Statistics
The significance of countable variables was estimated using the Chi-square test by comparing observed counts to those expected according to data from the control, TAP-sufficient cells. Likewise, differences in distributions were assessed using two-tailed Student’s t-test by comparing the non-TAP+ datasets to the control TAP-sufficient dataset.

Results
Data description
Several datasets were examined to investigate the study of the TAP-independent antigen processing pathways in TAP-deficient cells (Table 1). First, six datasets from four different studies, including 1051 MHC class I ligands from 727 parental proteins, were collected (Table 1). Individual studies contributed 42–543 peptides from 34–479 proteins. The TAP-deficient antigen presentation was also split into classical (TAP-C) versus nonclassical (TAP-NC) MHC, the latter was composed of the human HLA-E allele and its murine counterpart H-2 Qa-1b. As controls, two studies analyzing HLA class I ligands from TAP-sufficient (TAP+ ) cells (2125 peptides from 1557 proteins), including two of the alleles that were also present in the TAP-independent set (HLA-A2 and HLA-B27), were selected (Table 1). The initial data collection process was completed with six studies analyzing the HLA class II (HLA-II) peptidome (2799 peptides from 1027 proteins). Variable sensitivities and specificities should be expected due to the use of different techniques, the MHC alleles studied and the unavoidable variability of the laboratories involved. Nevertheless, the global analysis of these studies must reflect the general rules that govern the TAP-independent antigen presentation. However, the allele and study of origin were maintained throughout the manuscript to monitor such biological and technical biases. Because one of the two TAP-NC datasets was subjected to artificial length bias, we analyzed both datasets separately as well.

Some proteins recurrently appear in several studies, providing different peptides that bind to different alleles and are even represented in TAP-dependent, TAP-independent and HLA class II antigen processing pathways. Thus, these proteins are efficiently processed in different biochemical contexts. A relatively high overlap between TAP-independent and TAP-dependent datasets at the protein level was found (26% for the TAP-independent dataset, Fig 1). However, parental proteins from TAP-C and TAP-NC peptidomes only exhibited slight overlap (4%). Based on these data, the molecular rules governing antigen processing between TAP-C and TAP-NC should differ.

Small and highly integral transmembrane proteins are a relevant source of TAP-C ligands
Different features of the parental proteins from the four datasets were analyzed. The parental proteins for TAP+, TAP-NC, and HLA-II ligands had similar lengths (Fig 2A). In contrast, parental proteins for TAP-C ligands were significantly smaller than parental proteins for TAP+ (604 ± 667 vs 803 ± 877 average residues, P < 0.01; two-tailed Student’s t-test) and HLA class II (692 ± 708 average residues, P < 0.01) ligands. However, no relevant differences were observed in the pi and the hydrophobicity (Fig 2B and 2C).

In addition, the TAP+, TAP-C and TAP-NC datasets contained a comparable fraction of transmembrane proteins (Fig 2D). Nevertheless, parental proteins for TAP-C ligands were slightly enriched in membrane proteins with a highly integral character (>40% residues in transmembrane helices, P < 0.05; Chi-square analysis) compared with the TAP+ dataset, with
a percentage similar to parental proteins for HLA-II ligands (Fig 2D). The HLA-II dataset was strongly enriched in membrane proteins, although the low percentage (<10%) of residues in transmembrane helices of most of these proteins indicates that they are peripheral rather than integral membrane proteins.

Table 1. Summary of datasets included in this study.

<table>
<thead>
<tr>
<th>MHC</th>
<th>TAP</th>
<th>Number of peptides</th>
<th>MHC restriction</th>
<th>Number of proteins</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class I, classical</td>
<td>Independent</td>
<td>42</td>
<td>H-2Kd</td>
<td>34</td>
<td>[9]</td>
</tr>
<tr>
<td>Class I, classical</td>
<td>Independent</td>
<td>112</td>
<td>HLA-A2</td>
<td>74</td>
<td>[10]</td>
</tr>
<tr>
<td>Class I, classical</td>
<td>Independent</td>
<td>78</td>
<td>HLA-B27</td>
<td>40</td>
<td>[10]</td>
</tr>
<tr>
<td>Class I, classical</td>
<td>Independent</td>
<td>196</td>
<td>HLA-B51 or -C1</td>
<td>122</td>
<td>[10]</td>
</tr>
<tr>
<td>Class I, non-classical</td>
<td>Independent</td>
<td>80</td>
<td>H-2 Qa-1b</td>
<td>70</td>
<td>[11]</td>
</tr>
<tr>
<td>Class I, non-classical</td>
<td>Independent</td>
<td>543</td>
<td>HLA-E</td>
<td>479</td>
<td>[12]</td>
</tr>
<tr>
<td>TOTAL</td>
<td></td>
<td>1051</td>
<td></td>
<td>819</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>MHC</th>
<th>TAP</th>
<th>Number of peptides</th>
<th>MHC restriction</th>
<th>Number of proteins</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class I, classical</td>
<td>Dependent</td>
<td>1545</td>
<td>HLA-A2</td>
<td>1140</td>
<td>[22]</td>
</tr>
<tr>
<td>Class I, classical</td>
<td>Dependent</td>
<td>580</td>
<td>HLA-B27</td>
<td>487</td>
<td>[23]</td>
</tr>
<tr>
<td>TOTAL</td>
<td></td>
<td>2125</td>
<td></td>
<td>1557</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>MHC</th>
<th>TAP</th>
<th>Number of peptides</th>
<th>MHC restriction</th>
<th>Number of proteins</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class II</td>
<td>N/A</td>
<td>214</td>
<td>HLA class II</td>
<td>77</td>
<td>[24]</td>
</tr>
<tr>
<td>Class II</td>
<td>N/A</td>
<td>612</td>
<td>HLA class II</td>
<td>321</td>
<td>[25]</td>
</tr>
<tr>
<td>Class II</td>
<td>N/A</td>
<td>378</td>
<td>HLA class II</td>
<td>145</td>
<td>[26]</td>
</tr>
<tr>
<td>Class II</td>
<td>N/A</td>
<td>448</td>
<td>HLA class II</td>
<td>143</td>
<td>[27]</td>
</tr>
<tr>
<td>Class II</td>
<td>N/A</td>
<td>185</td>
<td>HLA class II</td>
<td>77</td>
<td>[28]</td>
</tr>
<tr>
<td>Class II</td>
<td>N/A</td>
<td>962</td>
<td>HLA class II</td>
<td>264</td>
<td>[29]</td>
</tr>
<tr>
<td>TOTAL</td>
<td></td>
<td>2799</td>
<td></td>
<td>1027</td>
<td></td>
</tr>
</tbody>
</table>

https://doi.org/10.1371/journal.pone.0210583.t001

Fig 1. Overlap of parental proteins between the datasets included in this study. Venn diagram showing absolute numbers of proteins shared between the different datasets analyzed: TAP-sufficient MHC class I (TAP+), TAP-independent classical (TAP-C) and nonclassical (TAP-NC) MHC class I and MHC class II (HLA-II).

https://doi.org/10.1371/journal.pone.0210583.g001
Protein abundance is a determining factor for TAP-C presentation

The cellular proteins encoded by mRNAs present at high concentrations are presumably abundant. Previous studies have shown that abundant proteins are likely the main source of endogenous TAP\(^{+}\) ligands, although proteins synthesized from poorly transcribed mRNAs also globally generate a significant fraction of these ligands \[30–32\]. Thus, considering the RNA levels of a lymphoid tissue, i.e., the spleen, as an indicator of protein levels, we confirmed that parental proteins for the TAP\(^{+}\) dataset were indeed more abundant than the average transcripts from the whole human proteome. Moreover, the link between protein abundance and MHC class I presentation was significantly stronger for the TAP-C dataset (Fig 3). In particular, proteins whose mRNAs are more abundant, measured as Reads Per Kilobase per Million mapped reads (RPKMs) > 8 were represented at a 1.94-fold higher level in the TAP-C dataset and its peptide contribution was 2.06-fold greater \(P < 0.001; \) Chi-square test) than in the TAP\(^{+}\) dataset.
Preferential processes and location of parental proteins for TAP-independent MHC class I ligands

An assay comparing protein functions between the four datasets was conducted using gene ontology enrichment. No biological processes were significantly associated with TAP+ ligands consistent with the fact that the proteasome evenly samples the whole proteome. However, the source of HLA-II ligands was highly enriched in genes involved in "immune and inflammatory responses", and "extracellular" or "endocytosis" processes as previously described [33] (Table 2). In addition, several central biological processes associated with mRNA production, translation and expression were linked to the TAP-C ligandome, data that correlate with their higher parental protein expression detected using the RPKM measurements (Fig 3). No biological processes were differentially associated with the generation of TAP-NC ligands. This finding, together with the low overlap between the TAP-C and TAP-NC sets showed in Fig 1, suggest that the source of these ligands seems to be different from TAP-C peptides (Table 2).

In addition, cellular components of the four datasets were analyzed (Table 3). No significant cellular components were associated with the TAP+ ligandome, which was also consistent with the ubiquitous protein degradation by the proteasome mentioned above. On the other hand, HLA-II ligands were associated with proteins involved in the endocytic pathway (endoplasmic reticulum, Golgi, vesicles, cell membrane, and extracellular matrix), as described in previous studies [34]. The sources of TAP-C ligands were enriched in clusters associated with RNA primary transcription, ribosome or aggregation of proteins and RNAs that appear when the cell is under stress, such as cytoplasmic stress granules (Table 3). Moreover, the cellular components associated with the TAP-NC ligandome were enriched in DNA-related structures, such as the kinetochore and chromatin. Altogether, these data support the existence of newly differentiated MHC class I antigen processing pathways and compartmentalization to yield classical and nonclassical TAP-independent ligands.

![Fig 3. Distribution of MHC ligand-containing proteins relative to their mRNA expression levels.](https://doi.org/10.1371/journal.pone.0210583.g003)
TAP-independent ligands were predominantly derived from the N- and, to a greater extent, C-terminal regions of parental proteins

TAP+ ligands were evenly distributed over the sequence of respective parental proteins (Fig 4A), indicating that the antigen processing by proteasomes mainly occurs on fully denaturalized proteins, as was classically described. Reversely, when the position of the TAP-C ligandome was mapped on the respective parental proteins whose lengths have been split into deciles, a “smile-shaped” graph was obtained. Specifically, the two N-terminal and the C-terminal deciles were overrepresented with respect the findings expected for random distribution, with an even more predominant contribution of the later category (Fig 4A). Although the TAP-NC ligandome exhibited a similar random distribution to TAP+ ligands (Fig 4A), when mouse and human HLA non classical alleles were separated, H-2 Qa-1b (but not HLA-E) also showed a “smile-shaped” graph (Fig 4B). This finding likely indicates a different antigen processing pathway for these HLA-E ligands compared with other TAP-independent ligands bound to different classical human and nonclassical mouse MHC class I alleles, although an artificial bias due to the exclusion of longer peptides in the original study cannot be excluded.

In addition, 21% of TAP-independent peptides were located exactly at the C-terminal position of their respective proteins. In these cases, only one endoproteolytic cleavage event was needed.
to release these particular ligands. In contrast, exact N-terminal ligands generally were not identified (less of 1% of peptides), indicating that most ligands required two different endoproteolytic cleavage events.

**The C-terminal sequences of parental proteins for HLA C ligands are enriched in Gly, Pro and aromatic residues**

The residue composition of the last 20 positions from the C-terminus of protein was analyzed and compared with the TAP” dataset to rationalize the abundance of TAP-C peptides in the...
C-terminal decile (Fig 4). No differences were observed with the exception of aromatic amino acids, Gly and Pro residues, which were statistically significant different despite the inherent noise (Fig 5). The two latter residues are common at the corner of turns in protein folds, which were explicitly overrepresented in several of these positions in the TAP-C dataset (Fig 5A). This effect was not observed in the extreme N-terminal, except for a Gly enrichment between positions 12 and 34 (average of 10.2% vs 8.7% in the TAP+ control; P-value: 0.002; Student’s t-test).

**Cleavage specificity of peptidases involved in the generation of TAP-independent ligands**

Finally, the specificity of peptidases involved in generating ligands in TAP-defective cells was addressed by determining the distribution of individual amino acid at the three flanking
residues (P$_3$, P$_2$, P$_1$, P$_1^{\prime}$, P$_2^{\prime}$ and P$_3^{\prime}$) on both sides of the hydrolyzed bonds of MHC ligands. Absolute residue counts at each position were normalized using the natural residue occurrence in the selected proteins. Fig 6 shows the distribution of amino acids, clustered by biochemical groups, found in the different flanking positions of scissile bonds.

The four datasets showed amino acids enrichment at several positions of flanking residues. The specificity of peptidases involved in the generation of TAP-independent ligands was more similar to peptidases that generate the TAP$^+$ peptides than enzymes involved in generating HLA-II antigens. However, as shown in Fig 6, differential enrichment of aromatic residues at the P3 position of the hydrolyzed bonds in the TAP-C ligandome and aliphatic residues enrichment at the P1 and P2 positions in the TAP-NC peptidome indicates that proteases with different specificity are involved in classical MHC class I antigen presentation generating the
classical and nonclassical peptide pools. Additionally, mouse and human HLA non classical alleles were separately analyzed, and differential enrichments of hydroxyl and amide (N/Q) residues was observed at the P3, P2 and P2’ positions of parental murine proteins.

Discussion

In this study, we provide information on the features of parental proteins, the source of TAP-independent ligands presented by alternative antigen processing pathways to the classical proteasome, TAP-dependent pathway. These results may explain why TAP mutations do not result in a lethal phenotype [3,4]. A systemic computational approach was used in the present study.

Several thousand ligands bound to specific MHC class I alleles were identified in an immunoproteomic analysis of TAP-sufficient cells in the previous studies [22,23]. In contrast, in the small number of similar studies examining TAP-deficient cells only tens [8,9] [11] or hundreds [10,12] of TAP-independent MHC peptides have been described. This finding is consistent with the very low expression of MHC class I molecules on the surface of TAP-deficient cells compared with normal, TAP-sufficient cells. The main difference between TAP-dependent and -independent MHC ligandomes analyzed using mass spectrometry in previous studies was the increased peptide lengths and the absence of strict binding motifs in the latter. In the later years, spliced peptides derived from a transpeptidation reaction mediates by proteasome activity between fragments distant in the parental protein have been described [35,36]. Although these these peptides are also longer than the TAP-dependent ligands, their contributions to antigen processing in TAP-deficient cells are very limited [37]. These differences between TAP-dependent and TAP-independent MHC ligandomes applied to all MHC class I alleles studied. The exception was the study by Weinzierl et al. [8], in which, in contrast to other TAP-independent peptidomes analyzed, the ligands were theoretically assigned to the

![Diagram of residues involved in the generation of naturally processed MHC ligands by peptidase cleavage.](https://doi.org/10.1371/journal.pone.0210583.g006)
respective HLA class I molecules using two in silico web tools: BIMAS (https://www-bimas.cit.nih.gov/) and SYFPEITHI (http://www.syfpeithi.de/) that only predict 8, 9 or 10 mer high affinity ligands. Thus, longer peptides or ligands with a low predicted HLA binding score (the majority of TAP-independent ligands in other mass spectrometry analyses performed [9,10]) were not detected by these algorithms and thus, the conclusions might be biased towards a rather limited nonrepresentative subset of the total TAP-independent HLA peptidome. Therefore, this study was not included in the present report. On the other hand, the data reported by Lampen et al. present an intermediate situation, since no bioinformatics strategy was used but only 8–13 mer peptides were manually included in the analysis of the nonclassical HLA-E peptidome [12]. Thus, in this mass spectrometry analysis, low affinity ligands were detected, but the most striking TAP-independent ligands (whose length is even greater than 20 residues) were excluded. This study was included in the current report, but independently analyzed to detect possible sources of bias as these peptides were more hydrophobic. In addition, as the same laboratory had previously analyzed the TAP-independent ligandome of murine orthologue (H-2 Qa-1B) of HLA-E with the same mass spectrometer and experimental procedures without length restriction [11], the effect of length selection was studied. The analysis of features of the proteins and peptides included in this report indicated that murine H-2 Qa-1B and human HLA-E peptidomes differed, as TAP-independent H-2 Qa-1B ligands exhibited very similar characteristics to TAP-C ligandomes from either murine or human TAP-deficient cells. Therefore, dramatic differences do not exist between the antigen processing pathways for classical or nonclassical TAP-independent ligands. Thus, the MHC peptidomes should be preferentially analyzed without any methodological restriction in order to avoid reaching spurious conclusions.

The prevalence of signal sequence-derived ligands was described in early mass spectrometry analyses [38] [39] and in other studies, focused on the most abundant cellular peptides bound to some HLA class I molecules [8] from TAP-deficient cells. This fact correlates with the binding specificity of the MHC class I molecules analyzed, HLA-A2 or -B51, which bind relatively hydrophobic peptides. In contrast, for other MHC class I molecules with positively or negatively charged residues as anchor motifs, the contribution of this pathway must be minor or even residual because signal sequences do not tolerate these residues. This finding is consistent with the origin of the viral TAP-independent epitopes identified by analyzing T cell immune responses. In a study of several MHC class I molecules, only 19% of TAP-independent viral epitopes were derived from signal sequences, a value that was less than those derived from other sources: luminal, transmembrane or even cytosolic proteins [40]. For example, only 2 of 13 TAP-independent ligands naturally presented by six different HLA class I molecules from vaccinia virus-infected TAP-deficient cells were derived from signal sequences [41]. Moreover, these low percentages may be overestimated, since HLA-A2 is the most frequently studied allele in antiviral immune responses.

Cellular proteins derived from highly abundant mRNA are much more common as a source of the TAP-independent ligandome than TAP-dependent HLA class I ligands. In addition, biological processes associated with mRNA production, translation and expression were also linked to the TAP-C ligandome. Both correlate with the poor performance of alternative antigen processing pathways compared with the high efficiency of classical proteasome, -TAP-dependent pathway. Thus, a reasonable hypothesis is that the protein abundance simply increases the possibility of interaction with the different "nonprofessional" antigen-producing proteases residing in the cellular compartments associated with TAP-independent antigen processing pathways in a TAP-free environment. Although the peptide and mRNA data were acquired from different sources, i.e., proteomic peptides from lymphoblastoid cell lines and RNA-Seq data from the spleen, respectively, these are more biologically relevant in the natural
environment of the later. Nevertheless, B cells are, by far, the predominant cellular type in the mouse spleen. Therefore, a worst case explanation is our results are an underestimation of the protein abundance as a key factor in TAP-deficient cells.

A strong tendency of TAP-C and H-2 Qa-1 to bind peptides located at the ends of the parental protein was observed. Notably, 30% and 35% of the TAP-C ligandome were present in the first two N- and the C-terminal sequence deciles, respectively. Similarly, 55% of H-2 Qa-1 ligands were located in these regions. Unlike the proteasome, when the protein substrates are unfolded by the 19S regulatory particle prior to their degradation, the peptidases that operate in the TAP-independent pathways must act on folded proteins. Only the surface of these proteins would be exposed, even transiently, and susceptible to protease activity, which reduces the diversity of potential ligands. These TAP-independent ligands were usually located at the exact C-terminal position of their respective proteins. The underlying explanation for this finding may be that peptides are released by a single cleavage event within the fully translated, likely folded protein which would favor the generation of ligands in a proteolysis-poor context. In addition, we identified a relative over-occurrence of Gly and Pro residues in these C-terminal regions. As Gly provides high flexibility to polypeptide chains, this amino acid is frequently found in protein regions without secondary structures, such as loops or coils. Pro is a structure-disrupting residue that is often found in loop regions. Altogether, this combination might explain, in terms of accessibility to proteolytic targets, the preference of TAP-independent antigen processing pathways for these C-terminal regions.

On the other hand, TAP-C or H-2 Qa-1 N-terminal peptides rarely include the first residue itself and only Gly residue biases were observed in this terminal region. A plausible explanation for the high N-terminal preference is the existence of nascent proteins that have not reached the critical number of translated residues to trigger the folding process, and thus proteases have a greater chance of targeting these sequences in this temporal window. In support of this hypothesis, two of the three most significant enriched cellular component GO terms (P-value < 10^-4 compared to TAP^+ data) for unique TAP-C proteins with peptides located in the first 50 positions of the N-terminus were "extracellular exosome" and "membrane". These proteins are initially cotranslationally inserted to the ER and thus, cleavages of their nascent unfolded DRiPs would generate TAP-independent ligands for MHC binding.

A relative abundance of aromatic or aliphatic residues in different flanking positions of scissile bonds and, for the Phe/Trp/Tyr amino acids, in the neighborhood of C-terminal regions of parental proteins, was observed in regions that are preferential source of TAP-independent ligands. Based in this finding, endoproteolytic proteases with specificity for these amino acids are likely relevant in the antigen processing pathway in TAP-deficient cells. Among these enzymes, cathepsins (MEROPS database: http://merops.sanger.ac.uk [42]), which were previously shown to involved in HLA class II antigen processing [43], show aromatic or aliphatic cleavage specificities and would be relevant candidates for generation of the TAP-independent ligands with adequate flanking regions.

Fundamental cellular processes, such as peptide presentation, rely on the robustness of proteins networks and therefore they must be explored with the lens of systems biology. Thus, antigen processing and presentation tolerates the absence of TAP by recruiting nonspecialized proteases that act as a safety net when the involvement of the proteasome is limited. These secondary alternative pathways contribute residually to total antigen processing in TAP-sufficient cells as assessed using TAP knockout-cell lines. The safety range for peptide presentation is substantially reduced but still significantly active in these cells. This observation represents the sum of marginal advantages that altogether counterbalance the lack of a dedicated proteolytic machinery such as the proteasome. This collective alternative appears to be sufficient to sustain the immunological clearance of most infections in TAP-deficient individuals.
In summary, the global picture emerging from the current report suggests that the TAP-independent antigen processing pathways are preferentially focused on small, abundant proteins with numerous Gly, Pro and aromatic residues in their C-termini, favoring peptide generation through a single cleavage event. Both the still-unfolded N-terminal and unfolded C-terminal sequences of parental proteins allow the protease to access and specifically perform the proteolytic cleavages that mainly generate the TAP-independent peptidome.

Author Contributions

Conceptualization: Daniel López.

Data curation: Antonio J. Martín-Galiano.

Formal analysis: Antonio J. Martín-Galiano.

Funding acquisition: Daniel López.

Investigation: Antonio J. Martín-Galiano, Daniel López.

Methodology: Antonio J. Martín-Galiano.

Project administration: Daniel López.

Software: Antonio J. Martín-Galiano.

Supervision: Daniel López.

Writing – original draft: Antonio J. Martín-Galiano, Daniel López.

Writing – review & editing: Antonio J. Martín-Galiano, Daniel López.

References


