Characterizing the N-Terminal Processing Motif of MHC Class I Ligands

Mark M. Schatz, Björn Peters, Nadja Akkad, Nina Ullrich, Alejandra Nacarino Martinez, Oliver Carroll, Sascha Bulik, Hans-Georg Rammensee, Peter van Endert, Hermann-Georg Holzhütter, Stefan Tenzer and Hansjörg Schild

*J Immunol* 2008; 180:3210-3217; doi: 10.4049/jimmunol.180.5.3210
http://www.jimmunol.org/content/180/5/3210

**References**
This article cites 42 articles, 14 of which you can access for free at:
http://www.jimmunol.org/content/180/5/3210.full#ref-list-1

**Subscription**
Information about subscribing to *The Journal of Immunology* is online at:
http://jimmunol.org/subscription

**Permissions**
Submit copyright permission requests at:
http://www.aai.org/About/Publications/JI/copyright.html

**Email Alerts**
Receive free email-alerts when new articles cite this article. Sign up at:
http://jimmunol.org/alerts
Characterizing the N-Terminal Processing Motif of MHC Class I Ligands

Mark M. Schatz,2* Björn Peters,2§ Nadja Akkad,* Nina Ullrich,* Alejandra Nacarino Martinez,* Oliver Carroll,¶¶ Sascha Bulik,‡ Hans-Georg Rammensee,‡ Peter van Endert,¶¶ Hermann-Georg Holzhütter,‡ Stefan Tenzer,3* and Hansjörg Schild3*

Most peptide ligands presented by MHC class I molecules are the product of an intracellular pathway comprising protein breakdown in the cytosol, transport into the endoplasmic reticulum, and successive N-terminal trimming events. The efficiency of each of these processes depends on the amino acid sequence of the presented ligand and its precursors. Thus, relating the amino acid composition N-terminal of presented ligands to the sequence specificity of processes in the pathway gives insight into the usage of ligand precursors in vivo. Examining the amino acid composition upstream the true N terminus of MHC class I ligands, we demonstrate the existence of a distinct N-terminal processing motif comprising approximately seven residues and matching the known preferences of proteasome and TAP, two key players in ligand processing. Furthermore, we find that some residues, which are preferred by both TAP and the proteasome, are underrepresented at positions immediately preceding the N terminus of MHC class I ligands. Based on experimentally determined aminopeptidase activities, this pattern suggests trimming next to the final N terminus to take place predominantly in the endoplasmic reticulum.


Major histocompatibility complex (MHC) class I molecules present peptide derived from intracellular proteins on the cell surface, allowing the immune system to detect pathogen infections and malignant mutations inside cells. These peptide ligands or their precursors are created through a multistep pathway starting with the breakdown of proteins by the proteasome alone (1). There are cases where the proteasome can already generate the final peptide ligand for MHC class I molecules (2). There are, however, also cases where the proteasome produces peptide precursors that possess the correct C terminus (3–5) while the N terminus carries extensions (6–12). Such precursor peptides are subject to proteolytic attack by endopeptidases and aminopeptidases in the cytosol (6–10), but some can escape complete degradation through transport into the endoplasmic reticulum (ER)4 via the TAP.

In the ER, the peptides are again subject to proteolytic attack by aminopeptidases (11–16), which synergize with MHC class I molecules to produce the final ligand (17). Two trimming enzymes, ERAP1 and ERAP2, have been identified in the ER of human cells with complementary trimming specificities (15), while mice express a homologue of ERAP1 but not ERAP2. Remarkably, ERAP1 prefers substrates 9–16 residues long (8, 16), which corresponds to the length of peptides efficiently transported by TAP. The recent availability of ERAP1 knockout mice has made it possible to study the impact of ERAP1 on MHC ligand generation and their recognition by T cells in vivo (18–22). These studies have shown changes in the epitope repertoire recognized in knockout mice, although with some discrepancies as to the magnitude and extent of the knockout impact. Overall, key enzymes of the ER trimming machinery are now well defined and their selective role in the generation of MHC class I ligands is clearly established.

Any selective process in MHC class I ligand generation is likely to leave a mark in the sequence composition of the ligands themselves or in their flanking regions. Previous studies have quantitatively characterized the sequence patterns preferred by proteasomal cleavage (23–28), TAP transport (29–34), and MHC binding (overviews in Refs. 35–37), and these patterns are indeed frequently found for MHC class I ligands. Most of these analyses focused on the C terminus of presented ligands, as N-terminal sequence patterns are complicated by the existence of different N-terminally prolonged ligand precursors. Although complicated, an analysis of the N-terminal processing motif can provide insights into the length distribution of ligand precursors generated during protein breakdown, the relative influence of cytosolic and endoplasmic trimming events, and the impact of the amino acid composition in the N-terminal extensions of ligand precursors for efficient ligand generation along the presentation pathway.

To address these questions, we here present a combined biostatistical and experimental approach to elucidate the role of amino acids in the sequence region preceding the N terminus of known MHC class I ligands. We demonstrate that the occurrence of these amino acids is not random but on the average forms a distinct
average scores calculated for random positions in the same set of proteins. These average scores at fixed positions were compared with root(MHC class I ligands. We refer to f(d) as the score associated with sequence look for patterns in the residue occurrence upstream the N terminus of trices described below. For each of these processes, we use its scores to port and aminopeptidase trimming can be evaluated using the scoring ma-

Our analysis was restricted to those ligands being flanked on both sides by at least 20 residues to avoid border effects, such as the possibilities that endopeptidases could not find a binding motif. Excluding those cases leaves us with that 1565 human MHC class I ligands having a size between 8 and 11 residues be epitopes capable of eliciting a T cell response. We extracted a set of the cell surface bound to MHC class I molecules, which may or may not be epitopes capable of eliciting a T cell response. We extracted a set of 1565 human MHC class I ligands having a size between 8 and 11 residues with known protein source sequences from the SYFPPEITHI database (38). Our analysis was restricted to those ligands being flanked on both sides by at least 20 residues to avoid border effects, such as the possibilities that aminopeptidase activity alone could liberate the ligand from its source or that endopeptidases could not find a binding motif. Excluding those cases leaves us with n = 1543 ligands.

Materials and Methods

MHC class I ligand data sets

We refer to MHC class I ligands as peptides that are naturally presented on the cell surface bound to MHC class I molecules, which may or may not be epitopes capable of eliciting a T cell response. We extracted a set of 1565 human MHC class I ligands having a size between 8 and 11 residues with known protein source sequences from the SYFPPEITHI database (38). Our analysis was restricted to those ligands being flanked on both sides by at least 20 residues to avoid border effects, such as the possibilities that aminopeptidase activity alone could liberate the ligand from its source or that endopeptidases could not find a binding motif. Excluding those cases leaves us with n = 1543 ligands.

Algorithm for proteasomal cleavage, TAP transport, and aminopeptidase trimming

FIGURE 1. Predicted efficiency of proteasomal generation and TAP transport of MHC class I ligand precursors. Average proteasomal cleavage scores at sequence positions pos = −15, −14, . . . , −1, and 0 upstream the N terminus of MHC class I ligands, and average TAP transport scores for the corresponding precursor peptides were calculated as described in the Algorithm section. Note that these prediction algorithms are not only based on the residue at the indicated sequence position, but include residues within a sequence window around this sequence position. The scores are given in units of SDs above or below predictions at random positions in the same set of proteins.

processing motif, which in length and residue composition reflects the influence of proteasomal cleavage and TAP transport on Ag processing. However, for the two residues directly preceding the N terminus of MHC class I ligands, we found systematic deviations from the proteasome and TAP preference which we suspected to reflect the influence of trimming processes by aminopeptidases. To test this hypothesis, we determined the cytotoxic and endoplasmic aminopeptidase activities toward all 20 amino acids. Relating the residue composition preceding MHC class I ligands to these measured peptidase activities provides evidence that removal of the two residues next to the N terminus of the final ligand is favored to take place inside the ER.

The Journal of Immunology

average score at a fixed position S(d) and the average score at random positions <S> in units of the SEM serr(S(d)).

The scores f_proteasome(d) for proteasomal cleavage between positions d-1 and d liberating a fragment with the residue at position d as its N terminus were calculated using the matrix described in Ref. 28. The matrices take into account a sequence region of six residues upstream and four residues downstream of a potential cleavage site. The matrices are based on proteasomal in vitro digests of whole proteins. Two matrices exist, one each for constitutive and immune proteasomes.

The scores f_TAP(d) for the TAP transport of fragments possessing the residue at position d as the N terminus were calculated using the matrix described in Ref. 31. Here, only the contribution of the three N-terminal residues of a peptide on TAP transport is taken into account.

For the calculation of aminopeptidase scores f_A pep(d) associated with position d, the turnover rates of the residues at positions d and d+1, and d+2 were taken from the experimental data described below. Logarithmic values of the turnover rates were used. This is coherent with the proteasomal cleavage prediction scores, which essentially give log(amount produced) values, and the TAP transport predictions, which give log(IC50) values.

Purification of the cytosolic and microsomal fraction

The purification of cytosol and microsomes is based on earlier protocols (39). The seven cell lines HeLa (human cervical adenocarcinoma), MC57 (mouse fibrosarcoma), RMA (mouse thymoma), J774 (mouse macrophage cell line), EL4 (mouse T cell lymphoma), D2SC1 (mouse melanoma cell line), and RAW309 (mouse macrophage cell line) were each cultured in 2 l RPMI 1640 (Biowhittaker) with 10% FCS (SeraPlus) at 37°C. Adherent cells were detached by a solution of 0.3 mM EDTA in PBS (pH 7.2). After pelleting (1,000 × g, 5 min, 4°C), the cells were washed once with PBS and twice with protease assay (PA)-buffer (20 mM HEPES/KOH (pH 7.6), 150 mM KCl, 5 mM MgCl2, and 0.5 mM DTT). The following step was conducted at 0–4°C. All buffers were filtered through a 0.45 μm membrane filter except buffers containing sucrose, which were filtered through a 1.2 μm membrane filter. A minimal volume of PA-buffer was added to the pellet before homogenization with an Elvehjem-potter (20×, 10–15 s up and down, 2000 rpm). The microsomes and the cytosol were purified by differential centrifugation, first for 10 min at 1,000 × g, then for 10 min at 10,000 × g, and last for 2.5 h at 140,000 × g. After incubation for 15 min at 0–4°C, the sample was centrifuged for 1 h at 140,000 × g through a sucrose cushion (0.5 M sucrose in PA-buffer; ratio sample to cushion 3:1) to separate cytosol and microsomes. Cushion and membrane-fractions were entirely removed and discarded. The supernatant, consisting of purified cytosol, was frozen in aliquots at −80°C before further use. The pellet, consisting of the microsomal fraction, was resuspended in PA-buffer and the same volume of stripping-buffer (50 mM EDTA and 1 M KCl in PA-buffer) was added to remove attached proteins. After incubation for 15 min at 0–4°C, the sample was centrifuged for 1 h at 140,000 × g through a sucrose cushion (0.5 M sucrose in PA-buffer; ratio sample to cushion 3:1). Cushion and supernatant were removed quantitatively and the pellet was washed with PA-buffer without disturbing it. The purified microsomes were resuspended with PA-buffer (1% Triton X-100 in PA-buffer) and stored in aliquots at −80°C.

Measuring aminopeptidase activity

The activity measurements with single-residue fluorogenic substrates were made in a 96-well flat-bottom transparent plate (Greiner Bioscience). For each amino acid, one well was filled with the particular cytosolic or microsomal fraction, the respective buffer (PA-buffer for the cytosol or PA-Lysis buffer for the microsomes), and BSA with a final concentration of 1 mg/ml. The resulting mixture was incubated with 20 μM of the respective substrate for 0, 10 min, 20 min, 30 min, 60 min, 90 min, or 120 min, the reaction stopped with 1% triton X-100 in PA-buffer and stored in aliquots at −80°C.

For the activity measurements with peptide substrates, ten peptides with the sequence H-XRGYYVYQGL-OH (X being either E, P, S, Q, I, R, L, F, M, W) or O were ordered from Schaefer-N at >80% purity. A total of 10 × 2 μl of each peptide [1 μg/μl in water] and 2 μl of purified cytosol or 20 μl of purified microsomes (see above) were mixed with PA-buffer (20 mM HEPES (pH 7.6), 150 mM KCl, and 5 mM MgCl2) to a final volume of 300 μl and aliquoted at 50 μl. Aliquots were incubated at 37°C for 0, 10, 20, 30 min, 60 min, 90 min, or 120 min, the reaction stopped with 1% formic acid (F.A.), and the aliquots frozen at −20°C. Capillary liquid chromatography was performed with a Waters NanoAcquity UPLC system.
Mobile phase A contained 0.1% F.A. in H2O and mobile phase B contained 0.1% F.A. in acetonitrile. Mass spectrometry analysis was performed using a Waters Q-Tof Premier in positive V-mode ESI. The mass spectrometer was calibrated with a [Glu1]-fibrinopeptide solution (500 fmol/l at 300 nl/min) delivered through the reference sprayer of the NanoLockSpray source.

Determination of initial turnover rates

The initial velocity (v0) of the turnover for each amino acid was calculated by determining the slope of the linear initial part of the curve in the RFU over time diagram. This value (RFU/min/l) was first normalized to the used volume (RFU/min/l) and then, to be able to compare different sets, displayed as a percentage of the overall activity, consisting of all 20 summarized single activities of the set: % of overall turnover = (v0 [RFUs/min/l] of the respective amino acid)/(Σ v0 [RFUs/min/l] of all 20 amino acids of the set)/100).

Results

Discovery of an N-terminal processing motif in MHC class I ligand precursor sequences

We compared the amino acid frequencies at up to 15 sequence positions upstream the N-terminal residue of 1543 known MHC class I ligands (taken out of the SYFPEITHI database (38)) with average frequencies found within the whole source proteins (Table I). A statistical analysis was performed across a set of 1543 MHC ligands with known source proteins. The left column depicts the average (expected) frequencies (Fexp) within the total set of source proteins. Observed amino acid frequencies differing from the expected frequency by more than two standard deviations are highlighted in light gray (above expected) or dark gray (below expected). The bottom row shows the p values of Pearson’s χ² test, which was performed to examine whether the frequency distribution of the 20 amino acids at a given sequence position (given by the values in one column of the matrix) is different from the overall distribution found in the source proteins (given by the values in the left column). Sequence positions are considered significantly different with pdiff ≤ 0.001.

Mobile phase A contained 0.1% F.A. in H2O and mobile phase B contained 0.1% F.A. in acetonitrile. Mass spectrometry analysis was performed using a Waters Q-Tof Premier in positive V-mode ESI. The mass spectrometer was calibrated with a [Glu1]-fibrinopeptide solution (500 fmol/µl at 300 nl/min) delivered through the reference sprayer of the NanoLockSpray source.

Determination of initial turnover rates

The initial velocity (v0) of the turnover for each amino acid was calculated by determining the slope of the linear initial part of the curve in the RFU over time diagram. This value (RFU/min¹/l) was first normalized to the used volume (RFU/min¹/l) and then, to be able to compare different sets, displayed as a percentage of the overall activity, consisting of all 20 summarized single activities of the set: % of overall turnover = (v0 [RFUs/min¹/l] of the respective amino acid)/(Σ v0 [RFUs/min¹/l] of all 20 amino acids of the set)/100).

Results

Discovery of an N-terminal processing motif in MHC class I ligand precursor sequences

We compared the amino acid frequencies at up to 15 sequence positions upstream the true N terminus of 1543 known MHC ligands (taken out of the SYFPEITHI database (38)) with average frequencies found within the whole source proteins (Table I). A χ² analysis shows that positions −5, −4, −3, −2, and −1 (numbers indicate the distance to the N-terminal residue of the final ligand)
exhibit sequence profiles that deviate strongly from the randomly expected ones ($p \leq 0.001$, $\chi^2$ test). Such significant deviations were not detected in a previous analysis (40), most likely due to the more limited amount of data available at the time.

Next, we compared the frequency deviations observed at the various sequence positions by performing a correlation analysis (Table II). This analysis reveals that neighboring positions between $-8$ and $-1$ tend to have a similar profile of frequency deviations, which means that in this sequence window neighboring positions have a similar amino acid profile. This also includes positions $-6$ to $-8$ for which the amino acid compositions do not differ significantly from those in the source proteins. On the contrary, more distant positions tend to have low or even negative correlations. In summary, there is a sequence window of approximately eight residues preceding an MHC ligand that clearly exhibits a nonrandom usage of amino acids that we call the N-terminal processing motif. This sequence motif is hypothesized to reflect residue preferences of various processes involved in the generation of epitope precursors, whereby these preferences are similar for neighboring sequence positions, but change with increasing distance.

The N-terminal processing motif matches the sequence specificities of the proteasome and TAP

Fig. 1 compares how the processing motif matches the known sequence preferences of the constitutive proteasome, the immuno
proteasome, and the TAP transporter. For this analysis, we applied sequence-based prediction algorithms to calculate proteasomal cleavage scores (28) (describing the specificity of the proteasome) for the various peptide bonds within the N-terminal extension and scores for the transport of the resulting peptide into the ER by TAP (31) (describing the specificity of TAP). As explained in the Materials and Methods section, the mean of these predicted scores was compared with the mean of reference scores calculated at random sequence positions of the source proteins. At each position up to five residues distal of the ligands N terminus, the mean cleavage scores for the constitutive and immuno proteasome are at least one SEM above those found at randomly chosen sequence positions. For TAP, the average transport score of precursors carrying an extension of up to eight residues is at least one SD above its random counterpart. Thus, amino acids preceding the N terminus of MHC class I ligands exhibit an accumulation of residues favoring proteasomal cleavage sites in this sequence region as well as the transport of the Nterminally extended peptides by TAP. This finding indicates that the N-terminal processing motif found in MHC class I ligand precursors results at least partly from the sequence selectivity of the proteasome and TAP.

Processing by the proteasome and TAP alone do not explain the observed N-terminal residue distribution

It is unlikely, however, that the processing motif is only influenced by the specificity of proteasomal cleavage and TAP transport. If N-terminally prolonged precursors are generated by the proteasome and transported by TAP, they have to be trimmed to their final size by aminopeptidases before presentation. It is expected that further upstream from the N terminus, at positions −7 to −4, these residues do occur at frequencies above average. This preliminary analysis suggests that aminopeptidases are involved in MHC class I ligand processing and those responsible for cleaving precursors with the one residue extensions have a low activity toward W, I, and V.

Determination of cytosolic and endoplasmic aminopeptidase specificity

To test this hypothesis, we used fluorogenic single-residue substrates for all 20 amino acids to monitor aminopeptidase activities in purified cytosol or purified solubilized microsomes, the latter reflecting the activity in the ER. The usage of single-residue substrates does not provide information about length preferences of aminopeptidases and they are not physiological substrates. However, they have the advantage that they cannot be cleaved by endopeptidases, which are incapable of digesting single-residue substrates, and thus cannot interfere with the assay. In particular, our amino acid probes cannot be cleaved by the two major cytosolic endopeptidases 20S proteasome and TPPII (Fig. 2(a)).

In our assay, we observed large differences in the turnover rates for different amino acids in the cytosol and the ER (Fig. 2, (c) and (d)). This preference for certain amino acids seems to be a general rule as can be concluded from the fact that similar substrate-dependent activity patterns were found for all cell types, including those of the immune system and from epithelia (Fig. 2(b)).

As expected, there were relevant differences between the cytosolic and the endoplasmic turnover rates (Fig. 2, (c) and (d)). One of the most prominent examples for such a difference is proline, which is one of the preferred amino acids for aminopeptidases in the cytosol, whereas there is almost a complete lack of proline-specific peptidase activity in the ER. The absence of proline activity in the microsomal preparation is a good measure for the purity of this fraction, as there is a strong activity for this residue in the cytosol. Similar to proline, the amino acid tryptophane is also trimmed much more efficiently in the cytosol. The opposite tendency is observed for the amino acids glycine, leucine, and methionine, all of which are degraded more efficiently in the ER. These differences in the specificity of cytosolic and endoplasmic aminopeptidases should also be reflected by the processing motif.

Next, we evaluated whether the digestion rates determined with nonphysiological single-residue substrates are comparable with those of oligomeric peptides. Ten 9-mer peptides differing only in...
their N-terminal residue were incubated with the purified cytosol (Fig. 3(a)) and microsomal fraction (Fig. 3(b)). Time courses of peptide degradation were recorded (Fig. 3) with the rates of degradation being R > W, M > L, F > I, Q > S, P > E in the cytosol and L, M > F, R, Q > I, W > S > E, P in the microsomal fraction. This hierarchy is very similar with the one determined with the single-residue substrates, which supports their use in this study.

The cytosolic and endoplasmic aminopeptidase specificities influence the N-terminal processing motif

To clarify the influence of aminopeptidase activities on MHC class I ligand generation and, thus, on the processing motif, we compared the residue-specific turnover rates found in the cytosol and the ER (see Fig. 2, (c) and (d)) with the amino acid frequencies observed in the N-terminal extensions of the 1543 MHC class I ligands (Fig. 4). A significantly higher turnover rate is found in the ER compared with the cytosol (Fig. 2, (c) and (d)). Time courses of peptide degradation were recorded (Fig. 3) with the rates of degradation being R > W, M > L, F > I, Q > S, P > E in the cytosol and L, M > F, R, Q > I, W > S > E, P in the microsomal fraction. This hierarchy is very similar with the one determined with the single-residue substrates, which supports their use in this study.

The cytosolic and endoplasmic aminopeptidase specificities influence the N-terminal processing motif

To clarify the influence of aminopeptidase activities on MHC class I ligand generation and, thus, on the processing motif, we compared the residue-specific turnover rates found in the cytosol and the ER (see Fig. 2, (c) and (d)) with the amino acid frequencies observed in the N-terminal extensions of the 1543 MHC class I ligands (Fig. 4). A significantly higher turnover rate is found in the ER compared with the cytosol for positions −2 and −1 (p < 0.001). A high turnover in the ER means efficient liberation of the ER (see Fig. 2, (c) and (d)). Time courses of peptide degradation were recorded (Fig. 3) with the rates of degradation being R > W, M > L, F > I, Q > S, P > E in the cytosol and L, M > F, R, Q > I, W > S > E, P in the microsomal fraction. This hierarchy is very similar with the one determined with the single-residue substrates, which supports their use in this study.

The cytosolic and endoplasmic aminopeptidase specificities influence the N-terminal processing motif

To clarify the influence of aminopeptidase activities on MHC class I ligand generation and, thus, on the processing motif, we compared the residue-specific turnover rates found in the cytosol and the ER (see Fig. 2, (c) and (d)) with the amino acid frequencies observed in the N-terminal extensions of the 1543 MHC class I ligands (Fig. 4). A significantly higher turnover rate is found in the ER compared with the cytosol for positions −2 and −1 (p < 0.001). A high turnover in the ER means efficient liberation of the final N-terminal extension of the 1543 MHC class I ligand, whereas a low turnover in the cytosol increases the probability of a peptide to be transported by TAP. This suggests that the final trimming of MHC class I ligand precursors at sequence positions −2 and −1 should predominantly take place in the ER. This conclusion is supported by the above mentioned finding that the residues W, I, and V are not preferred in the ER compared with the cytosol (Fig. 2, (c) and (d)).

Discussion

To examine the effect of N-terminal processing on MHC class I ligand selection, we analyzed the sequences of N-terminal extensions from 1543 MHC class I ligands. We identified an N-terminal processing motif with a length of approximately eight amino acids. This matches the size range for peptide products generated during protein breakdown as inferred from previous experimental publications. In a study by Cascio et al. (41), the OVA epitope SIINFEKL was mainly detected as an N-extended precursor with 1–7 additional residues after digestion of OVA by the proteasome. Reits et al. (6) showed that TPPII, which could play a role in precursor generation in combination with the proteasome, has a specificity for oligopeptides over 15 residues, creating epitopes or precursors of <16 amino acids, and thus, leaving N-terminal extensions of up to seven residues (given the fact that the shortest MHC class I epitope is an 8mer).

We found a significant increase in predicted proteasomal cleavage sites up to five residues distal from the MHC class I ligand N terminus and a significantly better predicted TAP transport rate of precursors carrying an extension of up to eight residues. The latter finding is in agreement with the reported maximum length of peptides efficiently transported by TAP (42). Determination of aminopeptidase activities led to the conclusion that the aminopeptidases of the ER are much more efficient than their cytosolic counterparts in removing the first two residues next to the true N terminus of the MHC class I ligand. This suggests that the final steps for trimming of most ligands take place in the ER. We did not find a significantly increased susceptibility toward aminopeptidase activity for the residues beyond sequence position −2. This can be due to equal participation of aminopeptidases in the ER and in the cytosol in the removal of these residues or it can be due to primary cleavage of longer precursors by endopeptidases.

Combining our results with data gathered in numerous previous studies, we propose the following model for Ag processing (Fig. 5). The first step is the cleavage of proteins by the proteasome into oligopeptides of up to 25 residues carrying already the correct C terminus for most MHC class I ligands. Proteasomal cuts preferentially lie within, but may also lie outside, the N-terminal processing motif formed by the sequence region −8 to −1 next to the true N terminus of the MHC class I ligand. Oligopeptides with >15 residues are likely to be further cleaved by TPPII. Either way, the proteasome alone or the proteasome in concert with TPPII generate MHC class I ligand precursors with N-terminal extensions of up to eight residues. Precursors of this length and below can be efficiently transported into the ER by TAP. Trimming at positions −8 to −3 can occur in either of the two compartments by cytosolic or endoplasmic aminopeptidases. Residues at positions −2 and −1 are predominantly removed in the ER. This final trimming event is followed immediately or is even accompanied by MHC class I binding of the ligand to protect it from further degradation (17, 43).

At positions −3 and beyond, no influence of aminopeptidase preference on the observed residues can be seen (Fig. 4). Two effects may account for this. First, aminopeptidases in both compartments could be equally responsible for the trimming of longer precursors. This would be consistent with our finding that the TAP specificity covers the whole processing motif. A second possible explanation is the involvement of endopeptidases, like the proteasome TPPII and others, that are able to remove longer peptide fragments. Such a processing mode could not be tested with the single-residue substrates of our assay.

As shown above, preferred cleavage sites for the proteasome can be found anywhere in various positions within the N-terminal processing motif and the same is likely to be true for TPPII, based on its peptide length preference. Hence, cleavage at residues further away from the ligand N terminus will be less dominated by the action of aminopeptidases, which lowers the probability to find clear aminopeptidase preferences at these positions. This again suggests that the significant values at positions −2 and −1 reflect a strong preference for endoplasmic trimming in this sequence.

A study by Nielsen et al. (27) was the first to derive a length distribution of MHC class I ligand precursors from proteasomal cleavage predictions. This was done by translating cleavage scores from the NetChop algorithm into cleavage probabilities using in vitro digest data. These probabilities were then used to describe MHC class I ligand precursor generation as a stochastic process. The resulting precursor length distribution showed a maximum
The knowledge gained by our analysis is of importance for the further development of the in silico identification of MHC class I ligands in arbitrary protein sequences. Although sequence-based predictions for proteasomal cleavage, TAP transport, and MHC binding are available, a mathematical model that includes epitope trimming is missing so far. To fill this gap, our data could help to create new algorithms incorporating the processing of the N terminus of MHC class I ligands. Our data on the specificities of trimming proteases could help to optimize the design of linkers between different epitopes to promote optimal processing and therefore optimal epitope presentation in multiepitope constructs.

**Acknowledgments**

We thank Jasmin Forsch and Agnes Malotta for expert technical assistance.

**Disclosures**

The authors have no financial conflict of interest.

**References**


