Probing proteasome activity and function : cancer diagnostics and mechanism of antigen processing

Berkers, C.R.

Chapter 3.2

How the proteasome cleaves and religates peptides to expand the antigen repertoire

How the proteasome cleaves and religates peptides to expand the antigen repertoire

Celia R. Berkers,[a] Annemieke de Jong,[a] Karianne G. Schuurman,[a] Carsten Linnemann,[b] Hugo Meiring,[c] Ton Schumacher,[b] Boris Rodenko[a] & Huib Ovaa[a,*]

Divisions of [a]Cell Biology II and [b]Immunology, The Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands. [c]The Netherlands Vaccine Institute, Unit Research and Development, P.O. Box 457, 3720 AL Bilthoven, The Netherlands.

Recently, three HLA class I epitopes have been reported, in which two distant parts of a protein are excised and ligated together to form a novel peptide. Such non-contiguous antigens can be produced by the proteasome via a transpeptidation mechanism, suggesting that splicing rules are likely to exist. In the present study we use small libraries of short peptides to identify amino acid sequences that are likely to promote this transpeptidation process and to deduce splicing rules for both the primed and non-primed proteasomal sites. Subsequently, we use a long viral polypeptide from the Influenza A neuraminidase 1 protein as a model to validate these rules. We quantify peptide ligation using a model peptide and demonstrate that it can occur with up to 30% ligation efficiency if structural requirements for ligation are met by both ligating partners. In addition, we show that protein splicing preferentially occurs via intrastrand recombination (recombination of peptides from the same protein) and that many spliced products may be formed from a single protein, some of which are likely to be immunogenic. Our data suggest that protein splicing occurs frequently and may occur at least 1,000 times more efficient as assumed previously. Therefore, protein splicing may play a much more significant role in immunity than previously assumed. The splicing rules presented here will facilitate the prediction and detection of spliced antigens to study their roles in immunity.

INTRODUCTION

Following malignant transformation or viral infection, a broad repertoire of antigenic peptides is presented on the cell surface by MHC class I complexes, enabling CD8+ T cells to sense the change in intracellular protein content, which eventually leads to a kill of the antigen presenting cell.[1] The 26S proteasome, a large threonine protease complex, is largely responsible for the generation of these antigenic peptides *in vivo*. The 26S proteasome consists of a 20S core, in which the catalytic activity resides, and 19S regulatory caps, which regulate the entry of substrates into the 20S core. The 20S core is formed by four stacked rings consisting of 7 subunits each, with an overall architecture of $\alpha(1\text{-}7)\beta(1\text{-}7)\beta(1\text{-}7)\alpha(1\text{-}7)$. Catalytic activity is provided by three β subunits, termed $\beta1$, $\beta2$ and $\beta5$, which are responsible for the caspase-like, tryptic and chymotryptic activ-

[*]Correspondence should be addressed to H.O. (h.ovaa@nki.nl).

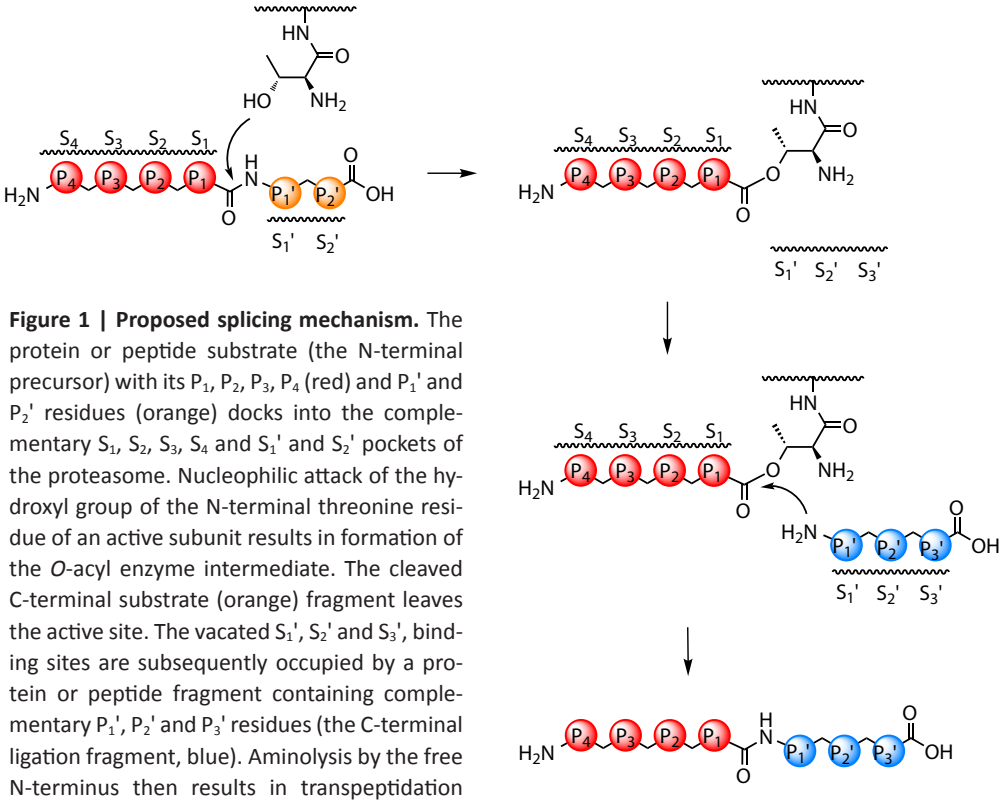


Figure 1 | Proposed splicing mechanism. The protein or peptide substrate (the N-terminal precursor) with its P_1, P_2, P_3, P_4 (red) and P_1' and P_2' residues (orange) docks into the complementary S_1, S_2, S_3, S_4 and S_1' and S_2' pockets of the proteasome. Nucleophilic attack of the hydroxyl group of the N-terminal threonine residue of an active subunit results in formation of the *O*-acyl enzyme intermediate. The cleaved C-terminal substrate (orange) fragment leaves the active site. The vacated S_1', S_2' and S_3', binding sites are subsequently occupied by a protein or peptide fragment containing complementary P_1', P_2' and P_3' residues (the C-terminal ligation fragment, blue). Aminolysis by the free N-terminus then results in transpeptidation and the formation of a novel peptide.

ity of the proteasome, respectively, and all have different cleavage preferences. The β1 subunit cleaves behind acidic amino acid residues, the β2 subunit prefers to cleave after basic residues, and the β5 subunit cleaves behind hydrophobic or aromatic residues. In immune tissues or under influence of the cytokine interferon-γ, these subunits are replaced by their immunoproteasomal counterparts, termed β1i, β2i and β5i to form the immunoproteasome,[2] while the existence of mixed-type hybrid proteasomes has also been reported.[3-5]

Recently, it has become apparent that not only contiguous antigens, but also non-contiguous antigens, that consist of two post-translationally fused peptides, can be presented to the immune system.[6-8] To date, three immunogenic, spliced antigens have been described, RTK-QLYPEW,[7] NTYAS-PRFK[6] and SLPRGT-STPK[8] (where the dash indicates the site of ligation). To form RTK-QLYPEW and NTYAS-PRFK, ligation occurred in-line (*i.e.* ligation of peptide fragments in the order in which they occurred in the parent protein), whereas SLPRGT-STPK was formed by religation of peptide fragments in the reverse order to that in which they occur in the parent protein. The ligation that produces these epitopes is thought to be catalyzed by the proteasome *in vivo* via a transpeptidation mechanism.[7-9] During hydrolysis, the N-terminal threonine residue of a catalytically active subunit reacts with the scissile peptide bond to form an *O*-acyl enzyme intermediate, resulting in the release of the C-terminal part of the peptide. In a second

step, water reacts with the intermediate ester resulting in the release of the N-terminal part of the peptide. During a transpeptidation reaction, the amino group of a second peptide competes with water for reaction with the intermediate ester, resulting in the formation of a new peptide bond (Figure 1).[10]

Proteasomal splicing results in a genuine post-translational modification, which can increase the diversity of the antigenic peptides presented, but which may also serve other yet unknown functions. A greater diversity of epitopes increases the chance that one or more of these epitopes are recognized by CD8+ T cells, which ultimately may result in more efficient elimination of infected or malignant cells.[10-12] It is currently not known whether proteasome splicing is a random event, or governed by explicit rules. As transpeptidation will only be able to compete with the hydrolysis reaction when both the N-terminal and C-terminal ligation fragments fit the active site pockets such that splicing is favored (Figure 1), it is not unlikely that splicing rules exist.[10] In addition, in barrel-shaped proteases such as the proteasome, reaction partners are generated *in situ* and remain confined to a small volume, which likely facilitates peptide ligation.[10] Because the various proteasomal active sites have different architectures, such rules, if existent, are likely site-specific.

In this study, we use short peptides to identify amino acid sequences that are likely to promote transpeptidation reactions, enabling the deduction of splicing rules. We validate these rules using a long viral polypeptide to mimic protein degradation in cells. In addition, we quantify the absolute amount of peptide ligation and demonstrate that it can occur efficiently with up to 30% ligation efficiency. Finally, we show that protein splicing preferentially occurs via intrastrand recombination (recombination of peptides from the same protein) and not via random recombination of peptides derived from different proteins. The elucidation of these splicing rules will facilitate the prediction and detection of spliced antigens to study their roles in immunity.

RESULTS

Rules for proteasomal antigen splicing exist

Two types of fragments participate in proteasomal ligation reactions: 1) N-terminal precursors, that fit into the proteasomal substrate binding non-primed pockets, so called S-pockets (S_1, S_2, S_3…S_n, depending on their proximity to the scissile peptide bond) and 2) C-terminal ligation partners, that fit into the proteasomal primed pockets (S_1', S_2', S_3'…S_n').[11] Residues in the substrate that interact with these proteasomal specificity pockets are referred to as P_1, P_2, P_3…P_n and P_1', P_2', P_3'…P_n', accordingly (Figure 1).[11] For both types of ligation partners, rules predicting proteasomal peptide ligation were deduced by comparing small libraries of potential ligation partners in digestion assays. To investigate which C-terminal ligation partners were likely to participate in ligation reactions, the N-terminal precursor YLDW|SY (where the vertical line indicates the site of cleavage) was incubated with proteasome in the presence of a 5-fold molar excess of the C-terminal ligation partner XXXSV in separate reactions (in which X is any natural amino acid). Transpeptidation led to the accumulation of YLDW-XXXSV, which has affinity for MHC class I (where the dash indicates the newly formed peptide bond). The extent of formation of MHC binders through transpeptidation in each sample could be measured using a fluorescence polarization (FP) assay[13,14] scoring for ligation efficiency (See Supplementary Information and Supplementary Figure 1 for details). By incubating YLDW|SY with proteasome in the presence of XLPSV (where

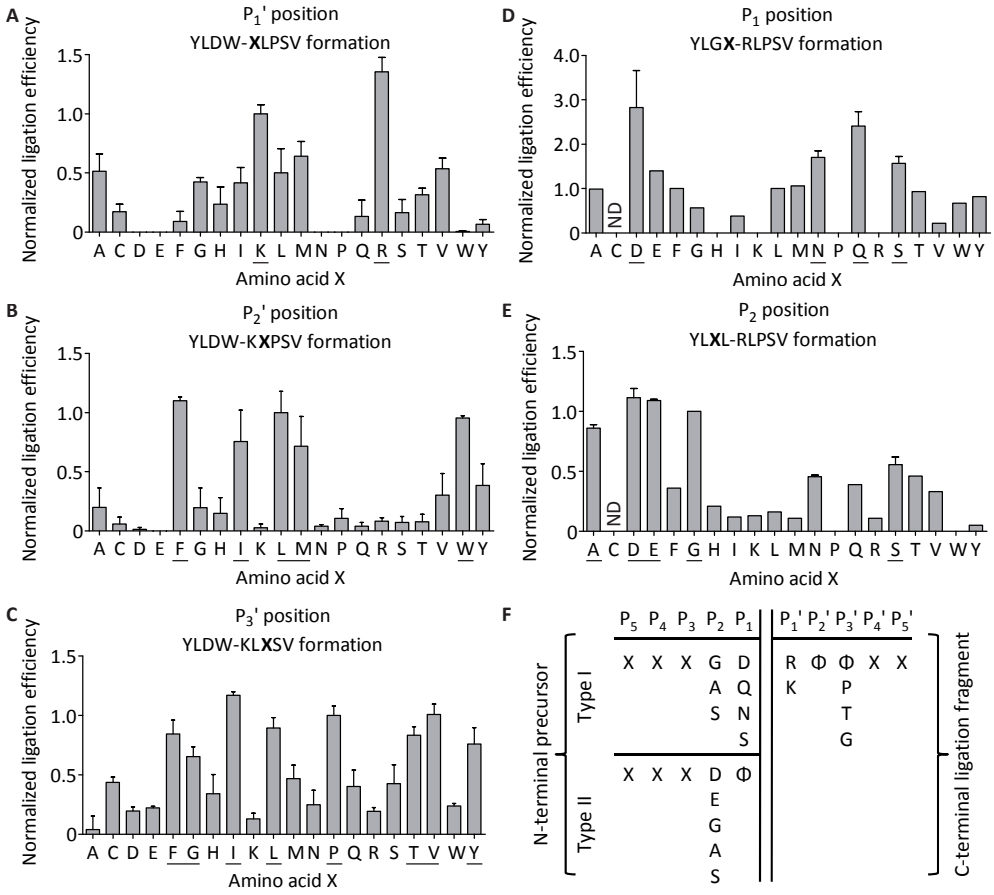


Figure 2 | Splicing rules to predict peptide ligation in the proteasome. A-C) Relative ligation efficiencies of different C-terminal ligation fragments measured in fluorescence polarization assays, indicating splicing preferences for the P_1' **(A)**, P_2' **(B)** and P_3' **(C)** positions. Efficiencies were normalized to the efficiency of KLPSV, which was included in all experiments and set to 1. Values are averaged over at least two independent experiments; error bars represent SEM. Residues that favored ligation on a particular position are underlined. **D,E)** Relative ligation efficiencies of different N-terminal ligation precursors measured in LC-MS assays, indicating splicing preferences for the P_1 **(D)** and P_2 **(E)** positions. Efficiencies were normalized to the efficiency of YLGL|SY, which was included in all experiments and set to 1. Residues that favored ligation on a particular position are underlined. Error bars represent SEM. ND: Not determined. **F)** Summary of the antigen splicing rules determined in these assays. Φ = hydrophobic residue; X = any natural amino acid.

X is any natural amino acid) in 20 separate reactions, we established which amino acid resides on P_1' promoted ligation (Figure 2A). In subsequent experiments, YLDW|SY was incubated with proteasome in the presence of the peptides KXPSV (Figure 2B) or KLXSV (Figure 2C) *in vitro* to define transpeptidation rules for the P_2' and P_3' position. All ligation efficiencies were related to the efficiency of YLDW-KLPSV formation, which was included in all experimental runs. To investigate which N-terminal precursors were likely to promote

transpeptidation reactions, the N-terminal precursor YLXX|SY was incubated with proteasome in the presence of one equivalent of the C-terminal ligation partner RLPSV for different time periods in separate reactions, in which transpeptidation led to the formation of YLXX-RLPSV. In contrast to experiments described above, hydrolysis rates differed between precursors (Supplementary Figure 2), hampering analysis of ligation efficiencies using the FP assay, which requires a constant hydrolysis rate for quantitative comparison of transpeptidation efficiencies. Therefore, we switched to LC-MS analysis of digestion mixtures. Ligation efficiency was defined as the percentage of YLXX|SY cleavages that resulted in ligation. Although MS is not generally used to quantify peptide levels, these peptides differ by a single amino acid per series and we compared efficiencies semi-quantitatively. By incubating YLGX|SY or YLXL|SY (where the vertical line is the site of cleavage and X is any natural amino acid) with proteasome in the presence of RLPSV in separate reactions, splicing rules for the P_1 and P_2 positions were determined (Figures 2D and 2E). Ligation efficiencies were determined for each time point and were normalized to the ligation efficiency of YLGL|SY, which was included in all experimental runs. Maximum ligation efficiencies were plotted for each residue position in Figures 2D and 2E.

From these data, distinct peptide motifs that are likely to either participate in or to abolish antigen splicing could be deduced for both the primed and the non-primed sites (Figure 2F). In general, rules seemed most strict for positions close to the actual site of transpeptidation (P_1, P_1', P_2 and P_2'), whereas multiple residues on P_3' favor transpeptidation. N-terminal precursors that favored ligation fell into the following categories: Type I precursors, which contained aspartate or a polar uncharged residue on P_1 combined with a small or polar (but not charged) residue on P_2 showed high ligation efficiencies but low cleavage rates (Figure 2F and Supplementary Figure 2A). Type II precursors, which combined a hydrophobic residue on P_1 with a basic or, to a lesser extent, polar residue on P_2, showed lower ligation efficiencies, but high cleavage rates (Figure 2F and Supplementary Figure 2A). The higher ligation efficiency of Type I precursors and the higher turnover rate of Type II precursors suggest that in proteins substantial amounts of ligation products may be produced from both types of precursor. C-terminal ligation partners that promoted ligation mostly contained a basic residue (K/R) on P_1' combined with hydrophobic residue on the P_2'. An additional hydrophobic residue or P/G on P_3' enhanced ligation further. Peptides containing a second charged residue on P_2' or P_3' in addition to K or R or P_1' did not serve as C-terminal ligation partners (Figure 2F).

We conclude from Supplementary Figure 2 that hydrolysis rates of N-terminal precursors were highest when a hydrophobic residue was present on P_1, suggesting that cleavage and thus ligation reactions were performed by the β5 subunit, for which the non-primed sites are well characterized.[15,16] Apart from the preference for a basic residue on P_1', the primed sites on the other hand have so far been only poorly characterized.[11,15] The amino acid preferences for ligation described above suggest that ligation efficiencies are highest with C-terminal ligation partners that have a good fit in the primed sites, judging from the requirement of K or R on P_1' for successful ligation. In addition, they further define the architecture of the β5 primed sites and suggest that hydrophobic residues are preferred on P_2' and to a lesser extent on P_3'. N-terminal ligation precursors on the other hand seemed more likely to be involved in ligation if the fit of the side chains of residues on the P_1 and P_2 positions in the corresponding S pockets was

sub-optimal, as inferred from the lower turnover rates for non hydrophobic residues (type I precursors).[15,16]

Splicing of optimized peptide sequences occurs with high efficiency

Next, we set out to characterize the efficiency and mechanism of splicing using an optimized model peptide. We combined the optimized N-terminal precursor YLGD (**a**) with the optimized C-terminal ligation fragment RLPSV (**c**), separated by a two amino acid sequence (SY; **b**) and we incubated this peptide YLGD|SY|RLPSV (**a**|**b**|**c**) with proteasome (where the vertical lines indicate proteasomal cleavage sites). The resulting reaction mixtures were analyzed by HPLC by monitoring absorption at 280 nm at various time points. At 280 nm, only the tyrosine residues in this peptide give rise to absorption, enabling quantification of hydrolysis and transpeptidation events. After a 2-h incubation period, the precursor peptide **abc** was still the main component in the digestion mixture (Supplementary Figure 3, red trace), while smaller amounts of the hydrolysis products **a**, **ab** and **bc** were also visible. In addition, various ligation products were formed (Peaks I to IV; Supplementary Figure 3) and identified by LC-MS as **a**-**a**, **a**-**c**, **a**-**ab** and **a**-**abc** (where the dash indicates the site of ligation). This suggests that peptide **a** is a very good precursor for ligation, as cleavage of the peptide bond between **a** and **b** (**ab** cleavage) resulted in the recombination of peptide **a** with other hydrolysis products or the complete precursor peptide to form a variety of ligation products. When the reaction was allowed to continue, the peak ratio changed. Precursor peptide **abc** almost completely disappeared, while hydrolysis product **a** became the main component in the digestion mixture (Supplementary Figure 3, black trace).

Ligation efficiency was defined as the percentage of **ab** cleavages that resulted in a ligation event. At 2h, the total ligation efficiency was 14% (Figure 3A), suggesting that one ligation product was formed every 7 times the **ab** bond was processed by the proteasome. Total ligation efficiency decreased gradually over time, while the ratio of different ligation products shifted from **a**-**abc** to shorter products. It is likely that the decrease in the amount of precursor **abc** over time led to increased processing of other peptides, including ligation products. Therefore, these shorter ligation products were probably predominantly formed by secondary processing of **a**-**abc**, the most abundant ligation product. When the concentration of precursor peptide **abc** was increased, hydrolysis rates increased, but the relative ratio of hydrolysis products did not shift (data not shown), indicating that the proteasomal cleavage preference was not altered. However, a 4-fold increase in peptide concentration doubled the ligation efficiency at 2 h from 15% to 30% (Figure 3B), suggesting that *in vitro* as many as 1 in 3 **ab** cleavages can result in a ligation event. Especially **a**-**abc**, resulting from the recombination of peptide **a** with the precursor **abc** was formed more efficiently (Figure 3B), suggesting that not only the sequence but also the concentration of C-terminal ligation partner in the proteasome is an important determinant for splicing. When reactions were allowed to continue, total ligation efficiencies decreased (data not shown), in accordance with data shown in Figure 3A.

The ligation efficiencies of splicing reactions for two reported spliced epitopes, RTK-QLYPEW[7] and NTYAS-PRFK[6] have been estimated. One spliced molecule of RTK-QLYPEW was estimated to be formed from ~ 1 x 10^4 precursor molecules RTK|AWNR|QLEPW.[7] The estimated efficiency of NTYAS-PRFK formation was even lower, with roughly one spliced peptide formed from ~ 5 x 10^5 precursor molecules.[9] Notably, these low splicing ef-

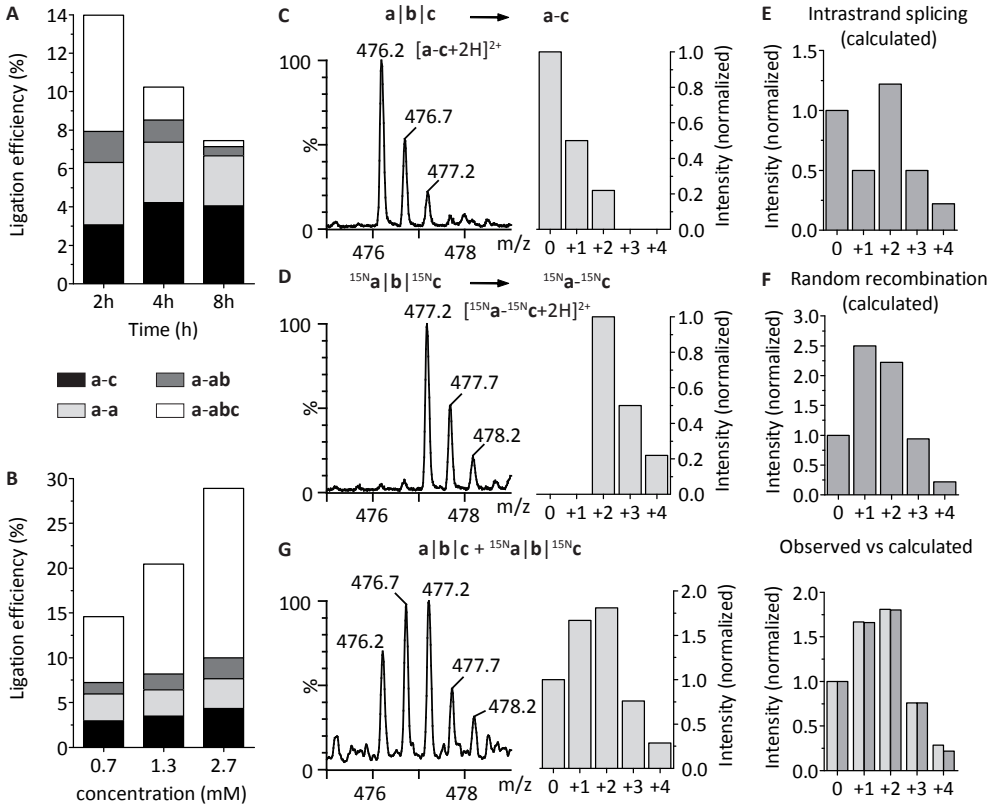


Figure 3 | Splicing of optimized peptide sequences occurs with high efficiency and predominantly via intrastrand splicing. **A)** Ligation efficiencies of the formation of different ligation products resulting from **a**|**b**|**c** digestion, measured at various time points. **B)** Ligation efficiencies of the formation of different ligation products resulting from digestion of increasing concentrations of **a**|**b**|**c**. **C)** Mass spectrum of **a**-**c**, formed during digestion of **a**|**b**|**c** (left panel). Isotope pattern derived from the mass spectrum of **a**-**c** by integration of mass peaks (right panel). **D)** Mass spectrum of ^{15N}**a**-^{15N}**c**, formed during digestion of ^{15N}**a**|**b**|^{15N}**c** (left panel). Isotope pattern derived from the mass spectrum of ^{15N}**a**-^{15N}**c** by integration of mass peaks (right panel). **E)** Calculated isotope distribution pattern for intrastrand splicing of a 1:1 digestion mixture of **a**|**b**|**c** and ^{15N}**a**|**b**|^{15N}**c**. **F)** Calculated isotope distribution pattern for random recombination of a 1:1 digestion mixture of **a**|**b**|**c** and ^{15N}**a**|**b**|^{15N}**c**. **G)** Mass spectrum resulting from the digestion of a 1:1 mixture of **a**|**b**|**c** and ^{15N}**a**|**b**|^{15N}**c** (left panel), showing a mixture of **a**-**c** and ^{15N}**a**-^{15N}**c**. Observed isotope pattern of a 1:1 digestion mixture of **a**|**b**|**c** and ^{15N}**a**|**b**|^{15N}**c** (middle panel). Comparison of the observed isotope distribution pattern with the calculated isotope pattern of 56% random recombination and 44% intrastrand splicing (right panel).

ficiencies of 0.0002% – 0.01% were sufficient to produce immunogenic epitopes that were presented at the cell surface by MHC class I to evoke an immune response. The splicing reactions creating these epitopes require two cleavage steps and for the model peptide described above the equivalent reaction would be the formation of **a**-**c** from **abc**. Splicing efficiencies for **a**-**c** formation were between 1.5% and 3% depending on concentration and reaction time. This indicates that splicing efficiencies may be several orders of magni-

tude higher than previously described if optimized N- and C-terminal ligation fragments are recombined. Therefore, spliced epitopes may be presented on the surface of cells to a larger extent than previously assumed.

Splicing of a model peptide occurs mainly via intrastrand splicing

Next, the mechanism of **abc** splicing was studied in more detail. In theory, **a**-**c** can be formed from **abc** via random recombination or via intrastrand splicing of **a** and **c**. In the latter case, **abc** is first be processed to **ab** and **c**. The liberated **ab** fragment then serves as precursor for ligation of the liberated **c** fragment derived from the same polypeptide chain. In contrast, a random **ab** or **abc** peptide can serve as precursor for **c** ligation, in which case recombination is said to occur randomly. To distinguish between random recombination and intrastrand splicing, a precursor peptide was used in which a ^{15}N-labeled amino acid was incorporated in part **a** and part **c** (^{15N}**a**|**b**|^{15N}**c**). First, either **abc** or ^{15N}**ab**^{15N}**c** were incubated with proteasome separately, and the resulting digestion mixtures were analyzed by LC-MS for the presence of **a**-**c** or ^{15N}**a**-^{15N}**c** (Figures 3C,D). In the digestion mixture of **abc**, a peak with an m/z value of 476.2 (M_0), corresponding to [**a**-**c**+2H]$^{2+}$ could be detected (Figure 3C, left panel), which showed a characteristic isotope pattern (Figure 3C, right panel). In the digestion mixture of ^{15N}**ab**^{15N}**c**, a peak with an m/z value of 477.2 and thus with a mass of +2 Dalton (M_{+2}), corresponding to [^{15N}**a**-^{15N}**c**+2H]$^{2+}$, was detected (Figure 3D, left panel), displaying a similar isotope pattern as **a**-**c** (Compare Figures 3C and 3D, right panels). Next, a 1:1 mixture of **abc** and ^{15N}**ab**^{15N}**c** was incubated with proteasome. In this experiment, intrastrand splicing would yield equal amounts of two products (**a**-**c**, and ^{15N}**a**-^{15N}**c**) with masses M_0 and M_{+2}. The resulting isotope pattern would therefore resemble the sum of Figures 3C and 3D, as shown in Figure 3E. When recombination of **a** and **c** in the mixture would be completely random on the other hand, indicating that intrastrand splicing is not preferred, four different peptides would be formed (**a**-**c**, ^{15N}**a**-**c**, **a**-^{15N}**c**, and ^{15N}**a**-^{15N}**c**) with masses M_0, M_{+1} and M_{+2} in a 1:2:1 ratio, resulting in a isotope distribution pattern as shown in Figure 3F. The actual MS spectrum and isotope distribution pattern of the 1:1 incubation mixture of **abc** and ^{15N}**ab**^{15N}**c** are shown in Figure 3G. The observed isotope pattern is an intermediate between Figures 3E and 3F, suggesting that both random and intrastrand recombination occur. The ratio between these two processes could be estimated by calculating the isotope distribution patterns of different ratios of random recombination and intrastrand splicing. The observed isotope pattern resembled an experiment in which 56% random recombination and 44% intrastrand splicing occurred (Figure 3G, right panel).

To interpret these results, the hydrolysis pattern of **abc** has to be taken into account. HPLC spectra indicate that **abc** is predominantly cleaved between **a** and **b** (66%), and not between **b** and **c** (34%) (data not shown). As intrastrand splicing can only occur when the **bc** bond is cleaved first, it can only occur in 34% of **abc** cleavages. Therefore, we expected that the isotope pattern would be predominantly determined by random recombination and that some enrichment in the M_0 and M_{+2} signals would occur as a consequence of possible intrastrand splicing of **a**-**c**. Remarkably, we observed 44% enrichment of M_0 and M_{+2} signals as a consequence of intrastrand splicing (see above). Such a high degree of enrichment is only possible if intrastrand splicing is highly preferred over random recombination. The proteasome creates N-terminal and C-terminal ligation partners almost simultaneously and as the two fragments are kept

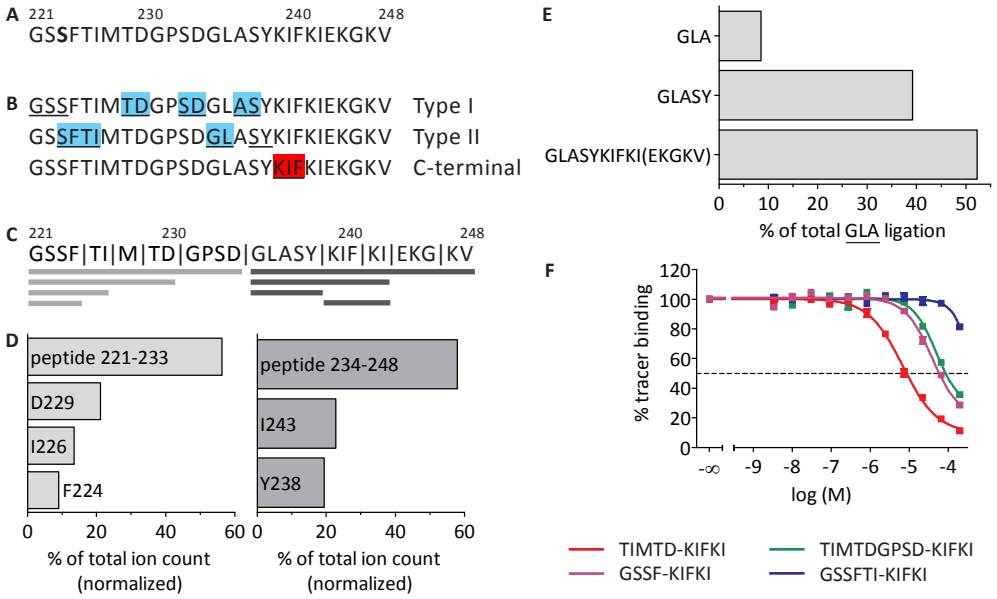


Figure 4 | Validating splicing rules: investigating influenza N1. A) Peptide 221-248/C223S (N1-1) derived from the neuraminidase 1 protein of Influenza A virus A/Puerto Rico/8/34 **B)** N-terminal (blue and underlined) and C-terminal ligation motifs (red and underlined) predicted by the splicing rules. Sequences that can participate in ligation but that are not likely to do so in this particular peptide are underlined only. **C)** Cleavage sites in peptide 221-248 (indicated by vertical lines) and the most predominant hydrolysis products formed (indicated by horizontal lines), as measured by LC-MS. **D)** Relative occurrence of secondary cleavages in peptides 221-233 (GSSFTIMTDGPSD) and 234-248 (GLASYKIFKIEKGKV). **E)** Relative contribution of different GLA-containing C-terminal fragments to total GLA ligation. **F)** HLA-A2.1 binding curves of peptides spliced from peptide 221-248/C223S (N1-1).

together inside the confined space of the catalytic chamber, ligation is promoted. This likely explains why intrastrand splicing occurs preferentially and more efficiently than random recombination.

Validating splicing rules: investigating influenza N1

Next, we validated the determined splicing rules by application to a polypeptide sequence of the neuraminidase 1 (N1) protein of Influenza A virus A/Puerto Rico/8/34 to mimic degradation of a real protein *in vivo*. *In vitro*, only peptides of limited lengths can be digested by 20S proteasome and therefore the 28-mer GSCFTIMTDGPSDGLASYKIFKIEKGKV (positions 221-248), which contained several N- and C-terminal splicing motifs, was selected for further analysis. Peptide 221-248 contained a cysteine residue on position 223, which complicates LC-MS analysis due to poor ionization efficiencies and/or the formation of disulfide bonds. Therefore, C223 was replaced by a serine residue (Figure 4A). Peptide 221-248/C223S (N1-1) contained several potential N- and C-terminal splicing precursors in accordance with our splicing rules. The sequences GS, SS, TD, SD and AS are all type I N-terminal transpeptidation motifs, while SF, TI, GL and SY are in line with Type II rules (Figure 4B, underlined). Only KIF (Figure 4B, underlined) was identified as a suitable

Table 1 | Ligation products identified in the digestion mixture of peptide 221-248.

Ligation product	Ion count	IC_{50} value
GSSF-KIFKI	1304	39 μM
GSSF-GLASY	477	> 1 mM
GSSF-GLASYKIFKIEKGKV	400	ND
GSSFTI-KIFKI	334	1 mM
GSSFTI-KIF	333	ND
GSSFTIMTDGPSD-KIFKI	328	ND
GSSF-KIF	319	ND
GSSF-KIFKIEKGKV	193	ND
GPSD-KIFKI	183	ND
GSSFTIMTDGPSD-KIFKIEKGKV	164	ND
GSSFTIMTD-GLASYKIFKIEKGKV	124	ND
GSSFTI-GLA	112	ND
GSSFTIMTD-KIFKI	109	ND
GSSF-GLASYKIFKI	104	ND
GSSFTI-KIFKIEKGKV	102	ND
GSSF-KIFKIEKG	79	ND
TIMTD-KIFKI	64	6.8 μM
TIMTDGPSD-KIFKI	56	49 μM
GSSFTIMTD-GLASY	40	ND
GSSFTI-GLASYKIFKI	38	ND
GSSFTIMTD-GLASYKIFKI	23	ND

ND: Not determined; a dash indicates the site of ligation.

C-terminal ligation fragment, (KIE contains a second charged residue, which was shown to be detrimental for ligation in experiments described above). In all our digestions, we have not observed cleavage within the first four residues from the N-terminus, which excludes GS and SS as suitable splicing precursors. Also SY must be rejected as splicing precursor, as SY directly precedes KIF and can therefore not participate in splicing with this C-terminal fragment. These restrictions leave six possible N-terminal precursors (indicated in blue) and one C-terminal ligation partner (indicated in red) in peptide N1-1 (Figure 4B) as predicted splicing sites.

Peptide N1-1 was synthesized and incubated with proteasome and the resulting digestion mixtures were analyzed by LC-MS. N1-1 contained nine main cleavage sites as indicated by the vertical lines (Figure 4C). Upon entering the proteasome, 90% of peptides were cleaved after D233, resulting in the accumulation of GSSFTIMTDGPSD (221-233) and GLASYKIFKIEKGKV (234-248) (Figures 4C,D). After this initial cleavage step, peptides 221-233 and 234-248 were both further processed. Fragment 234-248 was cleaved with equal preference after Y238 or I243, and to a lesser extent after F241 or G246. This yielded a variety of secondary cleavage products, of which GLASYKIFKI, GLASY and KIFKI were the most prominent. Fragment 221-233 was further processed after D229, I226 or F224, resulting in the accumulation of GSSFTIMTD, GSSFTI and GSSF (Figures 4C,D).

Table 1 shows all ligation products that could be identified in the digestion mixture. The identity of ligation products was confirmed by comparison with the corresponding synthetic peptide, as exemplified in Supplementary

Figure 4. Ligation was observed onto 4 different precursors of both type I (GSSFTIMTD and GSSFTIMTDGPSD) and type II (GSSF and GSSFTI). Peptides starting with either KIF or GLA served as C-terminal ligation partners, whereas only KIF was predicted by the model. GLA ligations mainly occurred with longer fragments (GLASYKIFKI and GLASYKIFKIEKGKV) (Figure 4E), which are rapidly formed during the initial cleavage step. In addition, GLA ligations were particularly favored over KIF ligations on position D229, which is separated from GLA by only four amino acids and is often the site where the next cleavage occurs. This suggests that GLA ligations occurred more efficiently than predicted because fragments were highly abundant and their generation directly preceded N-terminal precursor processing.

The analysis of ligation products indicates that the N-terminal ligation partner is the dominant factor determining ligation efficiency. Ligation will only take place if the N-terminal ligation precursor has a sequence that is suitable for ligation. Importantly, our splicing model accurately predicted the ability of peptides to serve as N-terminal ligation precursor for all cleavage sites. Ligation will preferentially occur with a suitable C-terminal ligation partner, but seems more dependent on the availability of peptides than on precise sequences. Whereas a hydrophobic amino acid on P_2' seemed required, the P_1' position was not limited to basic residues, but likely also allows for hydrophobic or small residues. Length was not a determining factor, as both short and long peptides were equally suited to serve as C-terminal ligation partners.

Splicing rules predict the formation of potential antigens from influenza N1

To investigate the affinity of spliced products for HLA-A2.1, binding curves of all products that consisted of 14 residues or less and contained at least one anchor residue (L/M/I on P2 and/or V/L/I on P9) were measured using an FP assay (Figure 4F, see Supplementary Information for details). Affinities (expressed as IC_{50} values) were defined as the concentration of peptide that inhibited 50% of tracer peptide binding and were determined for all binders (Table 1). Although N1-1 is a small protein fragment containing only 27 residues, N1-1 processing resulted in the splicing of four peptides that displayed affinity for HLA-A2.1. The peptide TIMTD-KIFKI, which contains both anchor residues and has the proposed optimal length for binding to HLA-A2, had the highest affinity for HLA-A2.1 (Figure 4F), with an IC_{50} value in the low micromolar range (Table 1). The splicing products GSSF-KIFKI and TIMTDGPSD-KIFKI, which had either the optimal length or both anchor residues, displayed intermediate affinity and showed IC_{50} values in the high micromolar range. The peptide GSSFTI-KIFKI had only low affinity for HLA-A2.1, while the other splicing products that were studied showed no affinity (Table 1). These data suggest that many potential HLA-A binding epitopes may be spliced from an average protein that contains 10 to 20 times as many residues as N1-1. Peptides that are spliced by the proteasome may be subjected to further processing by peptidases that reside in the cytosol and the endoplasmic reticulum. Therefore, peptides that do not display the right binding characteristics directly after proteasomal splicing, may be further processed to fit HLA-A molecules and be presented at the cell surface.

Validating splicing mechanism: investigating influenza N1

Next, we set out to compare the efficiency and mechanism of ligation of the influenza A derived N1-1 peptide to the **abc** model peptide described above. Ligation efficiency was defined as the percentage of precursor clea-

Table 2 | Ligation efficiencies of N-terminal precursors identified in peptide 221-248.

Ligation precusor	Efficiency (%)
GSSFTIMTDGPSD-	5.7
GSSFTIMTD-	7.3
GSSFTI-	23.9
GSSF-	59.5
Total	10.9

vages that resulted in a ligation event, and calculated from MS peak intensities. Total ligation efficiencies ranged from 5.7% to 59.5% for different N-terminal precursors (Table 2). The overall splicing efficiency was 10.9%, suggesting that one transpeptidation product was formed every 9 times peptide 221-248 was processed by the proteasome. These efficiencies are in the same range as ligation efficiencies determined for the **abc** peptide, suggesting that protein-derived peptides can be spliced efficiently by the proteasome.

To determine whether ligation would occur randomly or via intrastrand splicing, we replaced G221 and G234 in peptide N1-1 with ^{15}N Glycine (Figure 5A, indicated in red). N1-1 and ^{15N}N1-1 were incubated with proteasome and the formation of $^{(15N)}$GSSFTIMTD-$^{(15N)}$GLASYKIFKIEKGKV (**x**-**z** and ^{15N}**x**-^{15N}**z**; Figure 5A) by the proteasome was studied by LC-MS. In the digestion mixture of N1-1, a peak with an m/z value of 655.9 (M_0), corresponding to [**x**-**z**+4H]$^{4+}$ could be detected (Figure 5B), whereas in the digestion mixture of ^{15N}N1-1, a peak with a mass of +2 Dalton, corresponding to [^{15N}**x**-^{15N}**z**+4H]$^{4+}$, was detected (Figure 5C). A 1:1 mixture of these digestion samples, in which **x**-**z** and ^{15N}**x**-^{15N}**z** are present in equal amounts, resulted in the MS pattern shown in Figure 5D. The observed isotope pattern (Figure 5D, grey bars) resembled the sum of Figures 5B and 5C (Figure 5D, black bars) and is the expected isotope pattern for complete intrastrand splicing.

Next, a 1:1 mixture of N1-1 and ^{15N}N1-1 was incubated with proteasome (Figure 5E). The observed isotope pattern (grey bars) is an intermediate between complete intrastrand splicing (black bars) and complete random recombination (white bars), suggesting that both random and intrastrand recombination occur. A high peptide concentration within the proteasomal catalytic chamber likely increases the chance that two polypeptide strands are degraded simultaneously, favoring random recombination. We hypothesized that decreasing the precursor peptide concentration within the proteasome would shift the ratio between random recombination and intrastrand splicing towards intrastrand splicing. We expect that this shift is accompanied with only a minor decrease in ligation efficiency, as intrastrand splicing should be less effected by the peptide concentration as compared to random recombination. In addition, reducing the peptide concentration within the proteolytic chamber provides a better model for protein ligation *in vivo*, as the catalytic cavity is able to accommodate polypeptides between 25 and 30 kDa in size[17] and will therefore fit only one copy of most proteins. We incubated a 1:1 mixture of N1-1 and ^{15N}N1-1 with proteasome for longer time periods or with lower peptide concentrations to decrease the concentration of N1-1 within the proteasomal catalytic chamber (Figure 5F). Upon increasing the incubation time from 4 h to 16 h, the MS patterns shifted towards more intrastrand splicing. When the m/z peaks of M_0 and M_{+1} were integrated, the ratio shifted towards less M_{+1}, as less random recombination occurred and therefore fewer products with a mass of M_{+1} were formed. After a 16 h incubation period, the resulting MS

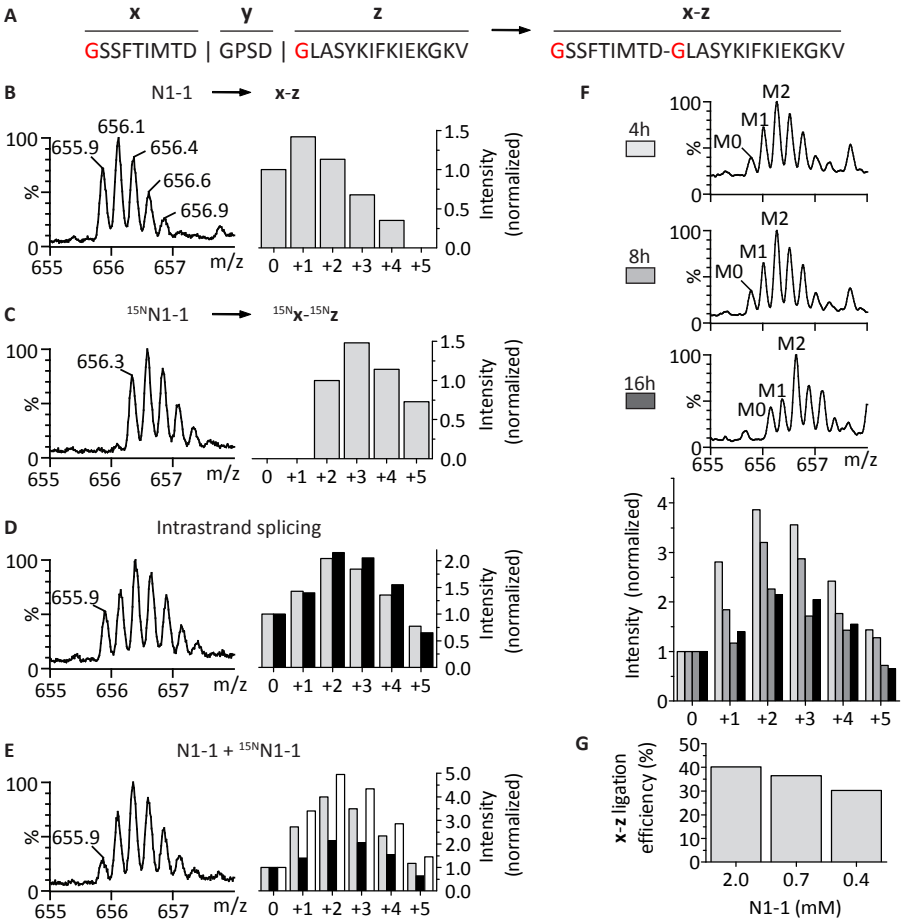


Figure 5 | Validating splicing mechanism: investigating influenza N1. A) Formation of $^{(15N)}$**x**-$^{(15N)}$**z** from $^{(15N)}$N1-1 (peptide 221-248/C223S from the neuraminidase 1 protein of Influenza A virus A/Puerto Rico/8/34) by the proteasome. ^{15}N glycine residues are indicated in red. **B)** Mass spectrum of **x**-**z**, formed during digestion of N1-1 (left panel). Isotope pattern derived from the mass spectrum of **x**-**z** by integration of mass peaks (right panel). **C)** Mass spectrum of ^{15N}**x**-^{15N}**z**, formed during digestion of ^{15N}N1-1 (left panel). Isotope pattern derived from the mass spectrum of ^{15N}**x**-^{15N}**z** by integration of mass peaks (right panel). **D)** Mass spectrum of a 1:1 mixture of digestion samples, in which **x**-**z** and ^{15N}**x**-^{15N}**z** are present in equal amounts (left panel). Observed and calculated isotope distribution patterns for intrastrand splicing (right panel). Grey bars represent the observed pattern; black bars represent the calculated isotope distribution pattern for intrastrand splicing of a 1:1 digestion mixture of N1-1 and ^{15N}N1-1. **E)** Mass spectrum resulting from the digestion of a 1:1 mixture of N1-1 and ^{15N}N1-1 (left panel). Observed isotope pattern of a 1:1 digestion mixture of N1-1 and ^{15N}N1-1 (right panel, grey bars), compared to the calculated isotope distribution patterns for random recombination (white bars) and intrastrand splicing (black bars). **F)** Mass spectra of samples resulting from the digestion of a 1:1 mixture of N1-1 and ^{15N}N1-1 measured at different time points during the digestion (top panel). Observed isotope patterns of a 1:1 digestion mixture of N1-1 and ^{15N}N1-1 measured at different time points during the digestion (4h, 8h, and 16h), compared to the calculated isotope distribution pattern for intrastrand splicing (black bars) (bottom panel). **G)** Ligation efficiencies of **x**-**z** formation from N1-1 at the indicated concentrations of N1-1.

spectrum closely resembled the calculated MS isotope pattern for complete intrastrand splicing (Figure 5F). Upon decreasing the concentration of N1-1 precursor in the sample, we observed only a minor decrease in ligation efficiencies for the formation of **x**-**z** (Figure 5G). Using **abc** as a model peptide, the efficiency of **a**-**abc** formation, which can only be formed via random recombination of two separate peptides, decreased from 19% to 7% (>60% decrease) when the concentration of **abc** was decreased with a factor 4 (Figure 3B). The efficiency of **x**-**z** formation on the other hand only decreased with 25% upon lowering the N1-1 concentration 5-fold (Figure 5G). Collectively, the observed shift in isotope pattern and the relative unchanged ligation efficiencies suggest that protein splicing occurs predominantly via intrastrand recombination.

DISCUSSION

Recently, it has become apparent that non-contiguous antigens, that consist of two post-translationally fused peptides are formed in cells and are presented to the immune system by MHC I.[6-8] These epitopes are thought to be produced by the proteasome via a transpeptidation mechanism.[7-9] As spliced antigens may mediate T cell responses to cancer, transplants and viral infection, knowledge of their identity is important for vaccine development, T cell based therapeutic strategies as well as pharmacological immunosuppression.[10] Novel epitopes are usually identified by cell surface elution, MS analysis and matching against protein databases.[18,19] Alternatively, epitopes may be predicted from protein sequences[20,21] and tested for T cell receptor recognition, e.g. by incorporation into fluorescent MHC class I tetramers, which are subsequently used to stain T cells in fluorescence assisted cell sorting (FACS) assays.[22,23] However, as spliced antigens are not contiguous and their formation requires post-translational splicing, they neither match existing databases nor can they be predicted from contiguous protein sequences. In the present study, we aimed at deciphering splicing rules that govern the production of non-contiguous antigens. Such rules will enable prediction and detection of spliced antigens to elucidate their role in immunity.

Splicing rules were determined for both the N- and C-terminal ligation fragments by analyzing proteasomal incubation samples from small libraries of both ligation partners using fluorescence polarization assays and LC-MS assays. Subsequently, these splicing rules were experimentally confirmed by application to a 27-mer polypeptide derived from the neuraminidase 1 protein from Influenza A virus A/Puerto Rico/8/34. Previous reports hypothesized that the transpeptidation mechanism would implicate that splicing is not determined by a particular sequence motif, but could occur at any major cleavage site.[7,9] Our data are however in strong disagreement with this notion and indicate that splicing rules exist, with ligation efficiencies being primarily determined by the N-terminal ligation partner. In general, N-terminal precursors promoted ligation when fitting sub-optimally in the non-primed S_1 and S_2 pockets. Two types of N-terminal precursors showed highest ligation efficiencies: precursors that I) contained aspartate or a polar uncharged residue on P_1 combined with a small or polar uncharged residue on P_2 or II) combined a hydrophobic residue on P_1 with a basic or small or, to a lesser extent, a polar uncharged residue on P_2. Although the highest ligation efficiency was observed with C-terminal fragments that had a good fit in the primed S_1', S_2' and S_3' pockets, structural requirements on the C-terminal ligation partner are less stringent, and ligation efficiency is rather determined by the presence than by the sequence of C-terminal ligation partners.

Ligation can only take place if the C-terminal ligation partner is able to compete with water molecules that are constantly present in the active site.[10,11] It has been hypothesized that splicing is most likely to occur with substrates that form intermediate esters with relatively long half-lives.[11]. Borrisenko and Groll suggested that this relatively long half-life could result from a high affinity of the N-terminal precursor for the non-primed sites.[11] Our data disagree with this hypothesis and suggest that the half-life of the *O*-acyl enzyme intermediate is prolonged if the fit of the N-terminal precursor in the non-primed pockets is suboptimal. Therefore, we propose a model in which a perfect fit may orientate the *O*-acyl enzyme intermediate such that the attack of a conserved water molecule on the intermediate ester is facilitated. A suboptimal fit of the N-terminal precursor, on the other hand, may prolong the half-life of the intermediate ester by influencing the orientation of the *O*-acyl enzyme intermediate, resulting in an intermediate complex that cannot be hydrolyzed easily and is more prone to ligation.

Our data indicate that a spliced product may be formed as often as once every 3 times a precursor peptide is processed (30% ligation efficiency). In previous studies, the ligation efficiencies of the splicing reactions for two described spliced epitopes were estimated to be 0.0002% to 0.01%.[7,9] Notably, these low efficiencies were sufficient to produce epitopes that were presented at the cell surface by MHC class I leading to an immune response. As our data suggest that splicing can occur with high efficiency *in vivo,* many spliced epitopes are likely to be present on the cell surface of all antigen presenting cells and may contribute to immunity. This splicing mechanism would ensure that a more diverse peptide repertoire reaches the cell surface, which increases the chance of recognition by CD8+ T cells and ultimately of elimination of the malignant or infected cell by the immune system.

The high splicing efficiencies observed in the present study imply that splicing rules must exist. If splicing would be a random process, high splicing efficiencies would result in the formation of high amounts of random (self)-antigens. Such random formation of self-antigens would hamper negative selection by cells in the thymus and cause auto-immunity.[10] Similarly, random recombination of different proteins would result in the formation of randomly formed spliced products, which would also severely hamper negative selection. We therefore expected protein ligation in the proteasome to occur via intrastrand splicing. Our data suggest that intrastrand splicing of peptides is strongly preferred over random ligation. In addition, only intrastrand splicing was observed when the peptide concentration within the proteasomal catalytic chamber was lowered. A reduced peptide concentration within the proteasomal chamber would provide a better model system for protein ligation *in vivo*. Only one copy of most proteins can physically fit into the confined space of the proteasomal catalytic cavity.[17] Experiments with ClpXP complexes, a proteasome-like protease complex from *E. coli,* have shown that two substrates could bind ClpXP simultaneously, but that simultaneous translocation from both ends only rarely occurred.[24] The 20S proteasome from *Thermoplasma acidophilum* can bind two substrates. However, the apparent rate-limiting step in protein degradation is the threading of the substrate through the pore followed by translocation, but not proteolytic cleavage.[25] It is likely that similar mechanisms underlie translocation in mammalian 26S proteasomes, which suggests that simultaneous proteolysis of two different proteins is a rare event. In accord with this, Dalet et al. have shown in a recent study that random recombination occurs readily *in vitro*, but that in cells random re-

combination occurs with much lower efficacy compared to intrastrand splicing.[9] Together, these data suggest that protein ligation occurs via intrastrand splicing *in vivo*, whereas the random recombination of proteins will rarely occur *in vivo* and is predominantly an *in vitro* artifact.

In conclusion, we determined splicing rules for both the primed and the non-primed sites. Such rules will enable the prediction and detection of spliced antigens to elucidate their role in immunity. In addition, we show that splicing occurs via intrastrand recombination with high efficiency and that many spliced products may be formed from a single protein, some of which are likely to be immunogenic. Therefore, the data presented here suggest that protein splicing may play a much more significant role in immunity than previously assumed.

MATERIALS AND METHODS

Peptide building blocks were purchased from Novabiochem and appropriately functionalized resins from Applied Biosystems. ^{15}N Glycine-N-Fmoc was purchased from Cambridge Isotope Laboratories. All solvents were purchased from Biosolve at the highest grade available. All other chemicals were purchased from Aldrich at the highest available purity. All solvents and chemicals were used as received. LC-MS analyses were carried out on a WATERS LCT mass spectrometer in line with a WATERS 2795 HPLC system and a WATERS 2996 photodiode array detector. HPLC purifications and analyses were performed on a Shimadzu LC-20AT prominence liquid chromatography system, coupled to a Shimadzu SPD-20A prominence UV/Vis detector and a Shimadzu CTO-20A prominence column oven.

Peptide synthesis

Peptides were synthesized using standard Fmoc-based solid-phase peptide synthesis protocols and appropriately functionalized PEG-polystyrene Wang resins. Functionalized resins were subjected to coupling cycles, in which deprotection of the Fmoc-group with piperidine/NMP (1/4 v/v), was followed by coupling with 4 equivalents each of Fmoc protected amino acid, di-isopropyl ethylamine (dipea) and (Benzotriazol-1-yloxy) tripyrrolidinophosphonium hexafluorophosphate (PyBop). Reactions were carried out in NMP at a volume of 1 mL/0.1 g of resin. After the final coupling step, the Fmoc group was removed and peptides were released from the resin and fully deprotected by treating the resin with TFA/H_2O/triisopropylsilane (93/5/2 v/v) for 2.5 h. Peptides were precipitated with cold diethylether/pentane (3/1 v/v) and lyophilized from H_2O/ACN/HOAc (65/25/10 v/v). Peptides were purified by HPLC over an Atlantis® Preparative T3 column (5 μm; 20 x 150 mm; Waters) to > 99.5% purity using a linear gradient (5% to 50% acetonitrile in H_2O containing 0.05% TFA). Peptide-containg fractions were pooled and lyophilized.

Proteasome purification

Proteasome was purified from bovine liver as described.[5] After each step, proteasome purity was monitored by incubating fractions with a fluorescent proteasome activity probe, followed by SDS-PAGE and scanning of the resulting gel for fluorescence emission, as described.[26] Briefly, bovine liver was homogenized in phosphate buffered saline (PBS), followed by precipitation using 40% saturated ammonium sulphate to remove unwanted proteins. The proteasome was subsequently precipitated by increasing the concentration of saturated ammonium sulphate to 60%. Following dialysis, the proteasome was further purified using a 10-40% sucrose gradient and anion exchange column chromatography, using DEAE Sephadex A25 resin. Proteasome containing fractions were pooled, concentrated and protein concentrations were determined using the Bradford assay (Biorad).

FP ligation assay to determine transpeptidation rules for the primed sites

50 nmol of the peptide YLDWSY was incubated with 5 μg proteasome in the presence of 250 nmol of XLPSV, KXPSV or KLXSV in ligation buffer (50 mM Tris-HCl, pH 8.5) in a total volume of 75 μL for 16 h at 37°C. As a control, 50 nmol of the peptide YLDWSY was incubated with 5 μg proteasome only. The DMSO concentration in all incubation mixtures was 8%. Following incuba-

tion, ligation mixtures were snap-frozen in liquid N_2, lyophilized, dissolved in 50 μL DMSO and the volume was adjusted to 1 mL with H_2O. Samples were purified over Strata™-X 33 μm polymeric reversed phase disposable columns (Phenomenex). To ensure equal input in all samples, 250 nmol of XLPSV, KXPSV or KLXSV was added to control samples prior to purification. Ligation products were eluted with ACN/H_2O (2/3 v/v), lyophilized and dissolved in FP buffer (20 mM Tris pH 7.4, 150 mM NaCl, 0.5 mg/ml BGG). Fluorescence polarization (FP) HLA-A2.1 binding assays with UV-mediated peptide exchange were performed as described previously[14] with minor changes. The peptide KILGFVFJ_1V (where J_1 is UV-cleavable 3-amino-3-(2-nitrophenyl)propionic acid) was used as conditional ligand and fluorescently labeled FLPSDC^{TMR}FPSV (where TMR is maleimido-linked tetramethylrhodamine) was used as tracer peptide. The ligation samples were used as competitor samples and binding curves of serial 2-fold dilutions of the expected ligation products were determined as reference. Assays were performed in 384-well low-volume black nonbinding surface assay plates (Corning 3820). Wells were loaded with 30μL total volume containing 0.5 μM HLA-A2.1-peptide complex, 1 nM tracer peptide and either 25% of each ligation or control sample or serial dilutions of the expected ligation products in FP buffer. The plate was spun for 1 min at 1000g at room temperature to ensure proper mixing of all components. To start UV-mediated cleavage of the conditional ligand and peptide exchange, the plate was placed under a 365-nm UV lamp at 10 cm distance (366 nm UV lamp, 2 × 15 W blacklight blue tubes, l×w×h 505 × 140 × 117 mm, Uvitec, UK) located in a cold room (4 °C). After 30 min irradiation, the plate was sealed with thermowell sealing tape (Corning) and incubated at room temperature for 24 hours, allowing cleaved peptide fragments to dissociate and to be exchanged for competitor and/or tracer peptide. Subsequently, fluorescence polarization measurements were performed, as described.[13] The relative ligation efficiency was calculated by comparing each sample to the corresponding control sample and ligation efficiencies were related to the efficiency of YLDW-KLPSV formation, which was included in all experimental runs. Data were analyzed using GraphPad Prism software (GraphPad).

LC-MS assay to determine transpeptidation rules for the non-primed sites

50 nmol of the peptide YLGXSY or YLXLSY was incubated with 5 μg proteasome in the presence of 50 nmol RLPSV in ligation buffer (50 mM Tris-HCl, pH 8.5) in a total volume of 75 μL for 2h, 4h, 8h, 16h or 24h at 37°C. The DMSO concentration in all incubation mixtures was 2.6%. Following incubation, ligation mixtures were snap-frozen in liquid N_2, lyophilized and dissolved in ACN/H_2O (1/1 v/v). Samples were filtered over Strata™-X 33 μm polymeric reversed phase disposable columns (Phenomenex) to remove proteasomal proteins and eluted directly. Eluted fractions were analyzed by LC-MS using an XBridge BEH300 C18 column (3.5 μm; 2.1 x 100 mm; Waters) using a linear gradient (5% to 50% acetonitrile in H_2O containing 0.1% Formic acid). Ligation efficiencies were calculated using the following formula: Efficiency = $I_L/(I_L+I_H)$ × 100%, where I_L is the peak intensity of the ligation product YLXX-RLPSV and I_H is the peak intensity of the hydrolysis product YLXX. Ligation efficiencies were related to the efficiency of YLGL-RLPSV formation, which was included in all experimental runs.

Investigating splicing efficiencies using a model peptide

The peptide YLGD|SY|RLPSV (**a**|**b**|**c**) (50 nmol, 100 nmol or 200 nmol) was incubated with proteasome (5 μg/ 50 nmol of peptide) in ligation buffer in a total volume of 75 μL for 2h, 4h or 8h. The DMSO concentration was 5.3% in all incubation mixtures. Incubation mixtures were snap-frozen, lyophilized and dissolved in ACN/H_2O (1/1 v/v). Samples were filtered over Strata™-X 33 μm polymeric reversed phase disposable columns (Phenomenex) to remove proteasomal proteins and eluted directly. Eluted fractions were diluted 1:1 with H_2O and analyzed by HPLC using a Symmetry® C18 column (5 μm, 4.6 x 150 mm, Waters) using a linear gradient (5% to 50% acetonitrile in H_2O containing 0.05% TFA). Spectra were analyzed using LCsolution software (Shimadzu). Integrated peaks were normalized to the number of tyrosine residues the corresponding peptide contained. Ligation efficiencies for the formation of distinct ligation products were calculated using the following formula: Efficiency = $I_L/(I_{LT}+I_H)$ × 100%, where I_L is the integral of ligation product (**a**-**a**, **a**-**c**, **a**-**ab** or **a**-**abc**), I_{LT} is the sum of the in-

tegrals of all ligation products (**a**-**a** + **a**-**c** + **a**-**ab** + **a**-**abc**) and I_H is the integral of the hydrolysis product **a**. Overall splicing efficiencies were calculated using the following formula: Efficiency = $I_{a\text{-}c}/(I_T\text{-}I_{abc}) \times 100\%$, where $I_{a\text{-}c}$ is the integral of the ligation product **a**-**c**, and I_T is the sum of all integrals and I_{abc} is the integral of the precursor peptide **abc**).

Investigating splicing mechanism using a model peptide

50 nmol YLGDSYRLGSV, 50 nmol YL^{15N}GDSYRL^{15N}GSV, or 50 nmol of a 1:1 mixture of both peptides was incubated with 5 μg proteasome in ligation buffer in a total volume of 75 μL for 2 h. The DMSO concentration was 1.3% in all incubation mixtures. Incubation mixtures were snap-frozen, lyophilized and dissolved in ACN/H_2O (1/1 v/v). Samples were filtered over Strata™-X 33 μm polymeric reversed phase disposable columns (Phenomenex) to remove proteasomal proteins and eluted directly with ACN/H_2O (1/1 v/v). Eluted fractions were analyzed by LC-MS using an XBridge BEH300 C18 column (3.5 μm; 2.1 x 100 mm; Waters) using a linear gradient (5% to 50% acetonitrile in H_2O containing 0.1% Formic acid).

Investigating influenza N1

Synthetic GSSFTIMTDGPSDGLASYKIFKIEKGKV (150 nmol) was incubated with 5 μg proteasome in ligation buffer in a total volume of 75 μL for 16 h. To investigate the splicing mechanism, different amounts of GSSFTIMTDGPSDGLASYKIFKIEKGKV, ^{15N}GSSFTIMTDGPSD^{15N}GLASYKIFKIEKGKV, or a 1:1 mixture of both peptides were incubated with 5 μg proteasome in ligation buffer in a total volume of 75 μL for 4 h, 8 h or 16 h. Incubation mixtures were snap-frozen, lyophilized and dissolved in ACN/H_2O (1/1 v/v). Samples were filtered over Strata™-X 33 μm polymeric reversed phase disposable columns (Phenomenex) to remove proteasomal proteins and eluted directly with ACN/H_2O (1/1 v/v). Eluted fractions were analyzed by LC-MS using an XBridge BEH300 C18 column (3.5 μm; 2.1 x 100 mm; Waters) using a linear gradient (5% to 50% acetonitrile in H_2O containing 0.1% Formic acid). To determine the affinities of spliced products for HLA-A2.1, an FP assay was performed as described above, with minor changes. 3-fold serial dilutions of synthetic products were used as competitor peptide and fluorescence polarization measurements were performed after a 4 h incubated period at room temperature. The binding affinity (IC_{50} value) of each spliced product was defined as the concentration that inhibited 50% of tracer peptide binding. Data were analyzed using GraphPad Prism software (GraphPad).

ACKNOWLEDGMENTS

The authors thank Henk Hilkmann for peptide synthesis and Jacques Neefjes for help with proteasome purification. This work was funded by the Dutch Cancer Society (grant NKI 2005-3368) and NWO/ALW.

REFERENCES

1. Shastri, N., Schwab, S. & Serwold, T. Producing nature's gene-chips: the generation of peptides for display by MHC class I molecules. *Annu. Rev. Immunol.* **20**, 463-493 (2002).
2. Rock, K.L. & Goldberg, A.L. Degradation of cell proteins and the generation of MHC class I-presented peptides. *Annu. Rev. Immunol.* **17**, 739-779 (1999).
3. Dahlmann, B., Ruppert, T., Kuehn, L., Merforth, S. & Kloetzel, P.M. Different proteasome subtypes in a single tissue exhibit different enzymatic properties. *J Mol. Biol.* **303**, 643-653 (2000).
4. Dahlmann, B., Ruppert, T., Kloetzel, P.M. & Kuehn, L. Subtypes of 20S proteasomes from skeletal muscle. *Biochimie* **83**, 295-299 (2001).
5. Raijmakers, R. et al. Automated online sequential isotope labeling for protein quantitation applied to proteasome tissue-specific diversity. *Mol. Cell. Proteomics* **7**, 1755-1762 (2008).
6. Hanada, K., Yewdell, J.W. & Yang, J.C. Immune recognition of a human renal cancer antigen through post-translational protein splicing. *Nature* **427**, 252-256 (2004).
7. Vigneron, N. et al. An antigenic peptide produced by peptide splicing in the proteasome. *Science* **304**, 587-590 (2004).
8. Warren, E.H. et al. An antigen produced by splicing of noncontiguous peptides in the reverse order. *Science* **313**, 1444-1447 (2006).
9. Dalet, A., Vigneron, N., Stroobant, V., Hanada, K. & Van den Eynde, B.J. Splicing of distant Peptide fragments occurs in the proteasome by transpeptidation and produces the spliced antigenic peptide derived from fibroblast growth factor-5. *J. Immunol.* **184**, 3016-3024 (2010).
10. Berkers, C.R., de Jong, A., Ovaa, H. & Rodenko, B.

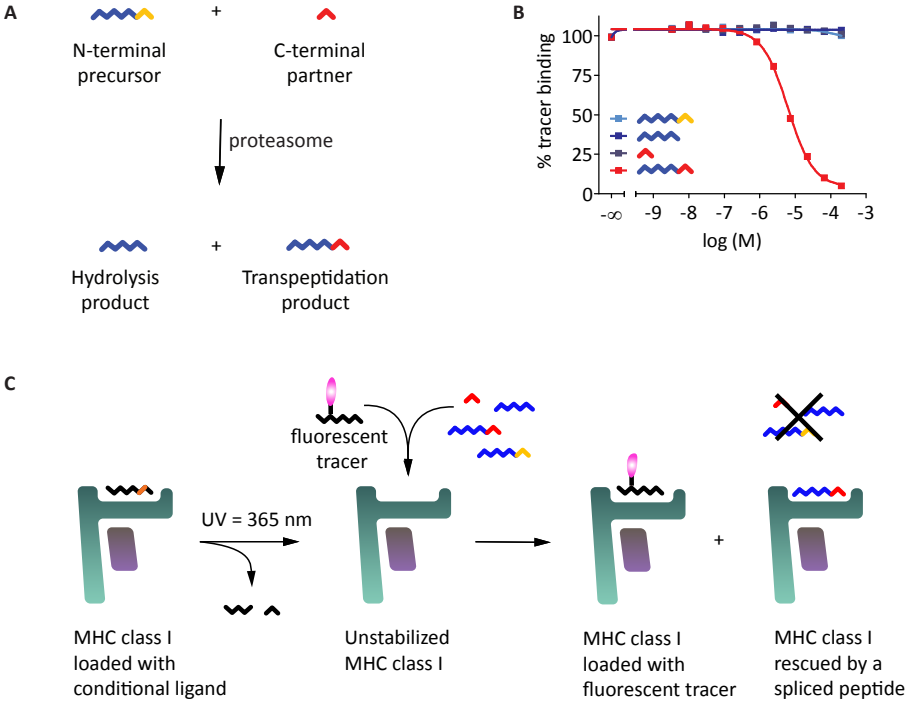
Transpeptidation and reverse proteolysis and their consequences for immunity. *Int J. Biochem. Cell Biol.* **41**, 66-71 (2009).
11. Borissenko, L. & Groll, M. Diversity of proteasomal missions: fine tuning of the immune response. *Biol. Chem.* **388**, 947-955 (2007).
12. Shastri, N. Cell biology. Peptides, scrambled and stitched. *Science* **313**, 1398-1399 (2006).
13. Bakker, A. H. et al. Conditional MHC class I ligands and peptide exchange technology for the human MHC gene products HLA-A1, -A3, -A11, and -B7. *Proc. Natl. Acad. Sci. U S A* **105**, 3825-3830 (2008).
14. Rodenko, B. et al. Class I major histocompatibility complexes loaded by a periodate trigger. *J. Am. Chem. Soc.* **131**, 12305-12313 (2009).
15. Nussbaum, A.K. et al. Cleavage motifs of the yeast 20S proteasome beta subunits deduced from digests of enolase 1. *Proc. Natl. Acad. Sci. U S A* **95**, 12504-12509 (1998).
16. Toes, R.E. et al. Discrete cleavage motifs of constitutive and immunoproteasomes revealed by quantitative analysis of cleavage products. *J. Exp. Med.* **194**, 1-12 (2001).
17. Sharon, M. et al. 20S proteasomes have the potential to keep substrates in store for continual degradation. *J. Biol. Chem.* **281**, 9569-9575 (2006).
18. Henderson, R.A. et al. HLA-A2.1-associated peptides from a mutant cell line: a second pathway of antigen presentation. *Science* **255**, 1264-1266 (1992).
19. Hunt, D.F. et al. Characterization of peptides bound to the class I MHC molecule HLA-A2.1 by mass spectrometry. *Science* **255**, 1261-1263 (1992).
20. Rammensee, H., Bachmann, J., Emmerich, N.P., Bachor, O.A. & Stevanovic, S. SYFPEITHI: database for MHC ligands and peptide motifs. *Immunogenetics* **50**, 213-219 (1999).
21. Nielsen, M. et al. Reliable prediction of T-cell epitopes using neural networks with novel sequence representations. *Protein Sci.* **12**, 1007-1017 (2003).
22. Toebes, M. et al. Design and use of conditional MHC class I ligands. *Nat. Med.* **12**, 246-251 (2006).
23. Altman, J.D. et al. Phenotypic analysis of antigen-specific T lymphocytes. *Science* **274**, 94-96 (1996).
24. Ortega, J., Lee, H.S., Maurizi, M.R. & Steven, A.C. Alternating translocation of protein substrates from both ends of ClpXP protease. *EMBO J.* **21**, 4938-4949 (2002).
25. Hutschenreiter, S., Tinazli, A., Model, K. & Tampe, R. Two-substrate association with the 20S proteasome at single-molecule level. *EMBO J.* **23**, 2488-2497 (2004).
26. Berkers, C.R. et al. Profiling proteasome activity in tissue with fluorescent probes. *Mol. Pharm.* **4**, 739-748 (2007).

Supplementary information

SUPPLEMENTARY RESULTS

Development of a fluorescence polarization assay to measure peptide ligation.

In order to screen small peptide libraries for their ability to participate in proteasomal transpeptidation reactions, we used a UV-mediated MHC exchange fluorescence polarization (FP) assay[1,2] scoring for ligation efficiency. In this assay, HLA-A complexes loaded with a UV-cleavable peptide ligand are subjected to UV light, which cleaves the ligand, resulting in the generation of empty HLA-A.[3] These empty HLA-A complexes are unstable, unless they are rescued by the addition of a 'rescue' peptide. When rescue is performed by adding both a fluorescent tracer peptide and increasing concentrations of a 'competitor' peptide, the affinity of this competitor peptide can be determined.[1,2] This assay was adapted so it could be used to determine the extent to which peptide ligation had taken place in a digestion mixture, resulting in potential HLA-A2.1 epitopes. First, proteasome was incubated with an N-terminal ligation precursor and a C-terminal ligation partner, resulting in the formation of both hydrolysis and transpeptidation products (Supplementary Figure



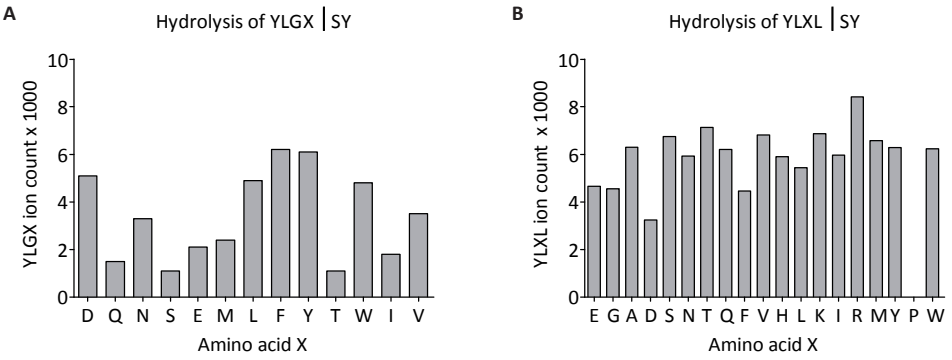
Supplementary Figure 1 | Development of a fluorescence polarization ligation assay. A) Schematic overview of hydrolysis and transpeptidation reactions in the digestion mixture. **B)** Affinities of components in the digestion mixture for MHC cass I. **C)** Schematic overview showing the fluorescence polarization assay used to determine ligation efficiencies in proteasomal digestion mixtures.

1A). This digestion mixture was used to rescue MHC class I. Peptides require two anchor residues to bind HLA-A2.1, one on the second position and one on the last position. Both N- and C-terminal precursors were designed such that they contained only one anchor residue and did not bind MHC class I (Supplementary Figure 1B). Upon ligation, peptides with both anchor residues were formed, resulting in a much-increased affinity for MHC class I (Supplementary Figure 1B). Therefore, competition with fluorescent tracer occurred only when transpeptidation products had been formed (Supplementary Figure 1C) and the relative amount of bound tracer inversely correlated to the efficiency of ligation in the corresponding digestion mixture. Samples were normalized to their respective control samples, in which ligation products could not be formed.

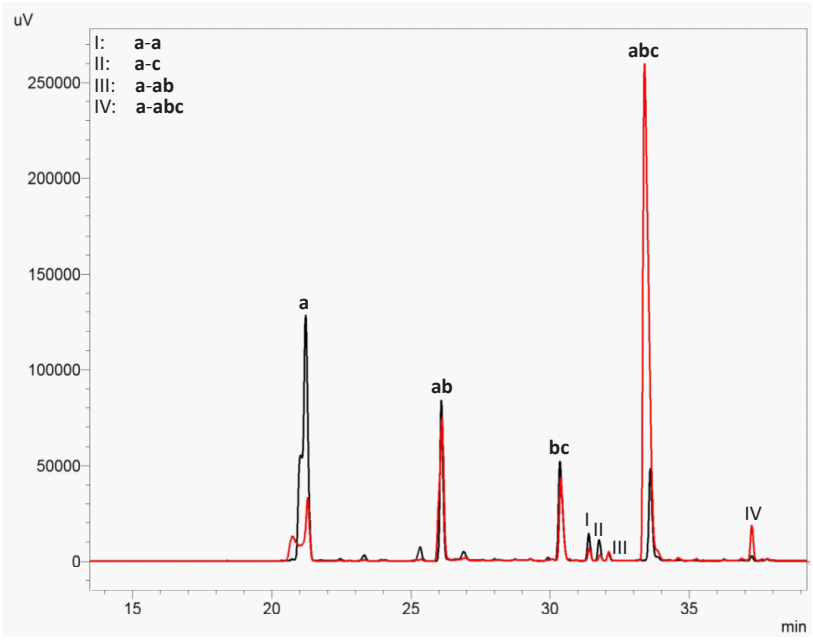
SUPPLEMENTARY REFERENCES

1. Bakker, A. H. et al. Conditional MHC class I ligands and peptide exchange technology for the human MHC gene products HLA-A1, -A3, -A11, and -B7. *Proc. Natl. Acad. Sci. U S A* **105**, 3825-3830 (2008).
2. Rodenko, B. et al. Class I major histocompatibility complexes loaded by a periodate trigger. *J. Am. Chem. Soc.* **131**, 12305-12313 (2009).
3. Toebes, M. et al. Design and use of conditional MHC class I ligands. *Nat. Med.* **12**, 246-251 (2006).

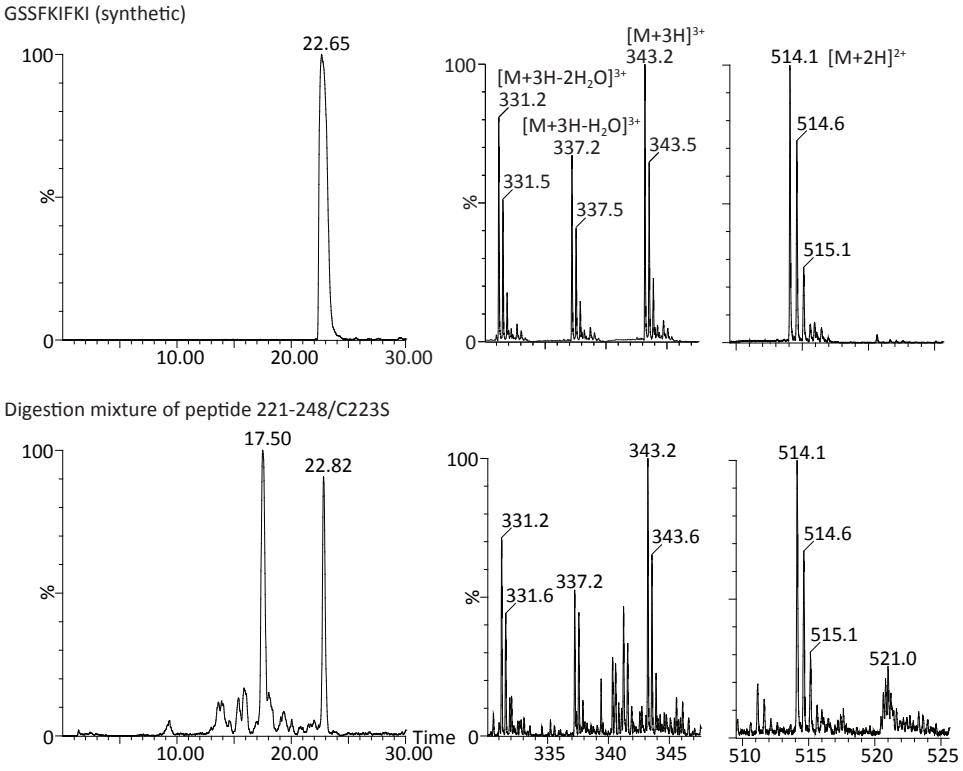
Supplementary figures



Supplementary Figure 2 | Hydrolysis rates of different N-terminal precursors. A) Amount of the hydrolysis product YLGX in proteasomal digestion mixtures of YLGX|SY after a 2h-incubation as measured by LC-MS. **B)** Amount of the hydrolysis product YLXL in proteasomal digestion mixtures of YLXL|SY after a 2h-incubation as measured by LC-MS. Cleavage sites are indicated with a vertical line.



Supplementary Figure 3 | Quantification of transpeptidation and splicing efficiencies using HPLC. HPLC spectra of proteasomal digestion mixtures of **abc** recorded at 280 nm after 2 h (red) and 8 h (black) incubations. Peaks were assigned to hydrolysis and ligation products using LC-MS. A dash indicates the site of ligation.



Supplementary Figure 4 | Identification of splicing products in peptide 221-248/C223S. Comparison of synthetic GSSFKIFKI and the digestion mixture of peptide 221-248/C223S. Ion traces detected at 342.7 m/z are compared (left panels) as well as in-line MS spectra of the peaks eluting at 22 to 23 minutes. Spectra taken from these peaks were identical, confirming that GSSF-KIFKI was formed in the digestion mixture by proteasomal splicing.