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

Berkers, C.R.

Chapter 3.3

Proteasomal splicing creates a novel type of antigen containing an isopeptide linkage

Proteasomal splicing creates a novel type of antigen containing an isopeptide linkage

Celia R. Berkers,[a] Annemieke de Jong,[a] Karianne G. Schuurman,[a] Jan A. J. Geenevasen,[b] Boris Rodenko[a,*] & Huib Ovaa[a,*]

[a]Department of Cell Biology II, The Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands. [b]Synthetic Organic Chemistry Group, Van 't Hoff Institute For Molecular Sciences, University of Amsterdam, Nieuwe Achtergracht 129, 1018 WS Amsterdam, The Netherlands.

Proteasomal transpeptidation reactions can occur particularly efficient if the C-terminal ligation partner has lysine or arginine at the site of ligation. As lysine has two amino groups that can theoretically both react with the *O*-acyl enzyme intermediate, this implies that the proteasome may be able to from isopeptide linkages. In the present study we use NMR to show for the first time that the proteasome can use both the α-amino group and the ε-amino group of lysine in transpeptidation reactions. The overall efficiency of $^{\varepsilon}$K-ligation reactions is 10-fold lower as compared to α-ligation and should produce sufficient amounts of peptide to evoke a T cell response, suggesting that the proteasome can create a novel type of antigen that may play a role in immunity in vivo. In addition, we show that isopeptides have unique properties that discern them from normal epitopes. Isopeptides of various lengths can bind to HLA-A2.1 and HLA-A3 proteins with high affinity. In addition, isopeptides are more stable towards further proteasomal processing as compared to normal peptides. These properties are likely to increase the fraction of ε-ligated peptides that enters the endoplasmic reticulum, and that is loaded onto MHC class I and transported to the cell surface for CD8+ T cell surveillance. We postulate that the formation of epsilon linkages is a genuine post-translational modification resulting from transpeptidation mechanisms.

INTRODUCTION

The eukaryotic 26S proteasome is responsible for the degradation of redundant and misfolded proteins and for the turnover of regulatory proteins involved in a wide range of cellular processes, including cell proliferation and survival, cell-cycle control and cellular stress responses.[1,2] In addition, the proteasome is responsible for the generation of peptides that are presented on the cell surface by Major Histocompatability Complex (MHC) class I proteins, known in humans as human leukocyte antigen (HLA) class I.[3] The proteasome generates peptides both from self and foreign proteins, which are subsequently transported into the endoplasmic reticulum, loaded onto MHC class I and transported to the cell surface. CD8+ T cells continuously scan the MHC I-peptide complexes on the cell surface, thereby receiving a blueprint of the intracellular protein content. This enables CD8+ T cells

*Correspondence may be addressed to H.O. or B.R. (h.ovaa@nki.nl or b.rodenko@nki.nl).

to sense viral infection or malignant transformation, which ultimately results in killing of the antigen presenting cell.

The 26S proteasome consists of a barrel-shaped 20S core, complexed at one or both ends with 19S regulatory particles. The 19S caps deubiquitinate and unfold protein substrates and regulate the entrance of substrates into the 20S core, where the proteolytic activity resides. The 20S catalytic chamber consists of four stacked heptameric rings and has an overall architecture of α(1-7) β(1-7)β(1-7)α(1-7). Whereas the two outer α-rings regulate entry into the complex and provide anchor points for the attachment of 19S regulatory caps, the actual catalytic activity resides within the inner β-rings. Once entered into the 20S core, a protein is degraded via the action of three catalytically active subunits, termed β1, β2 and β5, which all have different cleavage specificities and which are responsible for the proteasomal caspase-like, tryptic and chymotryptic activities, respectively. In lymphoid tissues these subunits are largely replaced by their immunoproteasome counterparts, termed β1i, β2i and β5i, to form the immunoproteasome.[4] The latter has been hypothesized to favor the production of antigenic peptides,[5] while the existence of hybrid 'mixed-type' proteasomes is now also well established.[6,7]

Recently, it has become apparent that the proteasome does not only produce contiguous peptides for presentation on MHC class I. Non-contiguous antigens, in which two distant parts of a protein are excised and ligated together to form a novel peptide, are presented on the cell surface and can evoke immune responses.[8-10] These non-contiguous antigens can be produced by the proteasome via a transpeptidation mechanism.[9-11] During the transpeptidation event, an active site N-terminal threonine hydroxyl group attacks the scissile peptide bond, resulting in the formation of an *O*-acyl enzyme intermediate and the release of the C-terminal part of the peptide. In a second step, this intermediate ester is captured by an amino group of a second peptide (the 'C-terminal ligation fragment'). This leads to the formation of a novel peptide bond and a spliced peptide in which two separate peptide fragments are combined (See Chapter 3.1, Figure 1; Chapter 3.2, Figure 1).[12] Transpeptidation competes with normal hydrolysis, in which the intermediate ester is hydrolyzed by water molecules present in excess. Hydrolysis results in the release of the N-terminal part of the peptide, which can be further processed in subsequent degradation rounds.

All peptides produced in the proteasomal catalytic core have a free amino terminus and can therefore theoretically participate in ligation reactions as C-terminal ligation partner. In a previous study (Chapter 3.2), we observed the highest ligation efficiencies with peptides that had a basic amino acid (arginine or lysine) as their N-terminal residue. As lysine has a free amino group on the C^{ε} position in addition to its free amino terminus, both the α-amine and the ε-amine are theoretically able to participate in ligation reactions. Whereas ligation with the α-amine results in the formation of a normal peptide bond, ligation with the ε-amine results in the formation of an isopeptide. Isopeptide linkage-containing epitopes may have unique properties and may form a novel class of post-translationally modified peptides that can be presented on MHC class I. Isopeptide bonds are notoriously difficult to identify and have not been observed in epitopes so far. In this study, we show that both the α- and ε-amino groups of lysine can participate in ligation reactions and we show how the unique properties of these isopeptides may contribute to an increased MHC class I surface expression, strongly suggesting that ε-ligation may contribute to immunity *in*

vivo, even though sensitive methods to detect such linkages directly do not exist.

RESULTS

The proteasome can form both normal and isopeptide bonds during splicing

To investigate whether both the α- and the ε-amino group of lysine could react with the *O*-acyl enzyme intermediate during splicing reactions, unprotected and mono-acetylated lysine were compared in splicing experiments. Only the unprotected α-amino group of N^{ε}-acetylated lysine ($K^{\varepsilon(Ac)}$) can participate in ligation reactions, resulting in the formation of normal peptide bonds only. On the other hand, N^{α}-acetylated lysine ($K^{\alpha(Ac)}$) can only react with the acyl-enzym intermediate through its unprotected ε-amino group, resulting in the formation of isopeptide bond containing peptides. The peptide C^{TMR}SLPRGTASSR is an N-terminally extended and fluorescently labeled version of the known splicing precursor peptide SLPRGTASSR.[10] This precursor was subjected to proteasomal degradation in the presence of proteasome inhibitors or lysine, followed by MALDI analysis. Peak heights in MALDI spectra could be semi-quantitatively compared, as all C-terminal hydrolysis and ligation products formed during these digestions contained a positively charged TMR (tetramethylrhodamine) label, which functioned as an ionization enhancer acting as the main determinant of signal intensities. The precursor C^{TMR}SLPRGTASSR (m/z 1616, Figure 1A, top panel) disappeared upon incubation with purified 20S proteasome, while two peaks appeared at m/z 1372 and 1214, which could be assigned to the hydrolysis products C^{TMR}SLPRGTAS and C^{TMR}SLPRGT, respectively (Figure 1A, second panel). Hydrolysis was completely abolished by the addition of the pan-proteasome inhibitor MG132, while the addition of the β5/β1 inhibitor bortezomib[13] only hampered the formation of C^{TMR}SLPRGTAS (m/z 1372), but not of C^{TMR}SLPRGT (m/z 1214) (Figure 1A, third and fourth panel). This suggests that C^{TMR}SLPRGTAS is formed by the β5 subunit, while C^{TMR}SLPRGT is a β2 hydrolysis product. When excess lysine was added to the reaction mixture, direct ligation of lysine onto the precursor peptide occurred, as evidenced by the appearance of additional peaks at m/z 1500 (Figure 1B, second panel) and m/z 1342 (data not shown). These peaks could be assigned to the β5 ligation product C^{TMR}SLPRGTAS-K and the β2 ligation product C^{TMR}SLPRGT-K (where a dash indicates the newly formed (iso)peptide bond). Ligation efficiencies, defined as the percentage of cleavages that resulted in ligation as opposed to hydrolysis, were 26% and 13% for the β5 and β2 active sites, respectively (Figure 1C). When either $K^{\alpha(Ac)}$ or $K^{\varepsilon(Ac)}$ was added to the digestion mixture, both MALDI spectra showed an additional peak at m/z 1542 (Figure 1B, third and fourth panel), which could be assigned to the β5 ligation product C^{TMR}SLPRGTAS-K^{Ac}. Ligation efficiencies were estimated at 29% for peptide bond formation with $K^{\varepsilon(Ac)}$ and 33% for isopeptide bond formation with $K^{\alpha(Ac)}$ (Figure 1C). Formation of the β2 ligation product C^{TMR}SLPRGT-K^{Ac} was only observed upon addition of $K^{\varepsilon(Ac)}$, with a ligation efficiency of 7% (Figure 1C). These data indicate that the α- and ε-amino groups of lysine are equally capable of participating in proteasomal splicing reactions in the β5 active site, whereas the β2 active site only accepts α-amino groups. This strongly suggests that both peptide and isopeptide bonds can be formed during proteasomal splicing reactions.

The β5 site accepts α-acetylated peptides as C-terminal ligation partner

Having established that isopeptide bonds can be formed by the proteasome when single

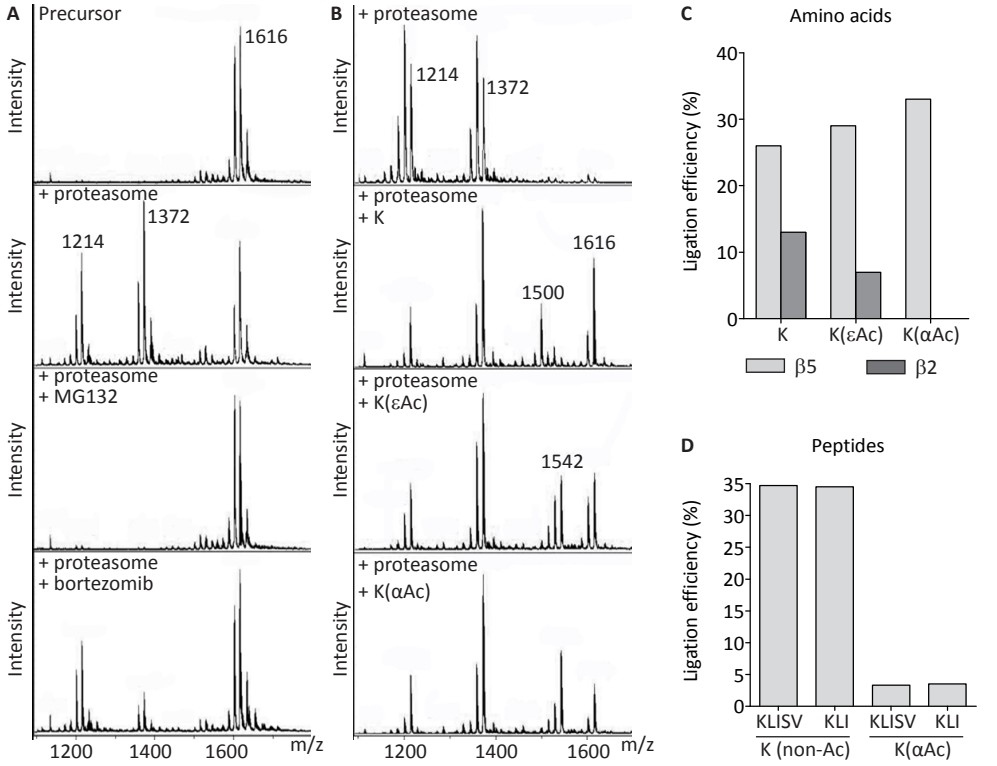


Figure 1 | The proteasome accepts α-acetylated amino acids and peptides as C-terminal ligation partner during splicing reactions. A) MALDI spectrum of C^{TMR}SLPRGTASSR (top panel). MALDI spectra of proteasomal digestion mixtures of C^{TMR}SLPRGTASSR only (second panel), C^{TMR}SLPRGTASSR and MG132 (third panel), and C^{TMR}SLPRGTASSR and bortezomib (bottom panel). **B)** MALDI spectra of proteasomal digestion mixtures of C^{TMR}SLPRGTASSR only (top panel), C^{TMR}SLPRGTASSR and lysine (second panel), C^{TMR}SLPRGTASSR and N^{ε}-acetylated lysine (third panel), and C^{TMR}SLPRGTASSR and N^{α}-acetylated lysine (bottom panel). **C)** Ligation efficiencies for the formation of β2 and β5 ligation products by lysine and acetylated lysine, calculated from MALDI experiments. **D)** Ligation efficiencies for the formation of YLGD-KLISV and YLGD-KLI using non-acetylated or N^{α}-acetylated C-terminal ligation fragments. Ligation efficiencies were calculated from LC-MS experiments. Ligation efficiency was defined as the percentage of cleavage that resulted in ligation as opposed to hydrolysis.

amino acids are added to digestion mixtures, we investigated whether N^{α}-acetylated peptides could serve as C-terminal ligation partners. The N-terminal precursor YLGD|SY has been established as a β5 splicing precursor (where the vertical line indicates the site of cleavage; Chapter 3.2). YLGD|SY was incubated with proteasome and a 5-fold excess of either unprotected or N^{α}-acetylated KLI or KLISV to form the hydrolysis product YLGD and the transpeptidation products YLGD-KLI/YLGD-$K^{\alpha(Ac)}$LI or YLGD-KLISV/YLGD-$K^{\alpha(Ac)}$LISV (where a dash indicates the newly formed (iso)peptide bond). The resulting digestion mixtures were analyzed by LC-MS. Ligation with non-acetylated KLI and KLISV occurred with equal efficiency (Figure 1D), suggesting that peptide bond formation is independent

of the length of the C-terminal ligation fragment, as has been observed before (Chapter 3.2). Ligation with N^{ε}-acetylated peptides gave similar results compared to reactions using non-acetylated peptides (data not shown). Ligation of N^{α}-acetylated peptides on the other hand occurred with 10-fold lower efficiency compared to non-acetylated peptides (Figure 1D), in contrast with results found with single amino acids. The length of the N^{α}-acetylated C-terminal ligation fragment did not influence ligation efficiency (Figure 1D), as was also observed with non-acetylated peptides. Together these data suggest that the proteasome can form isopeptide bonds between two peptide fragments, but that it favors α-ligation over ε-ligation.

ε-ligation can compete with α-ligation during proteasomal splicing events

Next, we studied whether the ε-amino group of an N-terminal lysine residue can compete with the α-amino group for ligation when both amines are unprotected. To this end, a ligation experiment was performed using a C-terminal ligation fragment with a ^{15}N-enriched ε-amine. A ligation product containing a ^{15}N-enriched amide will result from ε-ligation, whereas α-ligation will result in a ligation product containing a ^{15}N-enriched amino group (Figure 2A). As amides and amines are distinguishable by NMR spectroscopy, this enabled us to study to which extent ε-ligation occurred. ^{15}N-ε-Fmoc-Lys(Boc)-OH was synthesized and incorporated into $K^{\varepsilon 15N}$LI by standard solid phase peptide synthesis. Subsequently, a 5-fold excess of the N-terminal precursor YLGD|SY was incubated with proteasome and $K^{\varepsilon 15N}$LI to form YLGD-$K^{\varepsilon 15N}$LI, where the dash indicates the newly formed (iso)peptide linkage. The resulting digestion mixture was analyzed using LC-MS. Figure 2B shows absorption spectra recorded at 280 nm of the reaction mixture before incubation (left panel) and after incubation (second panel). In addition to the N-terminal precursor YLGD|SY (eluting at 11.9 min), the hydrolysis product YLGD (eluting at 10.0 min) and different ligation products could be detected in the digestion mixture after incubation (Figure 2B, second panel). MS analysis of the peak eluting at 13.7 min showed that this peak contained two peptides with masses of 821 Da (M_1) and 913.8 Da (M_2) (Figure 2B, third panel), which corresponded to the ligation products YLGD-$K^{\varepsilon 15N}$LI and YLGD-YLGD. The peak eluting at 14.6 min. contained a third ligation product with a mass of 1163.6 Da (M_3) (Figure 2B, fourth panel), corresponding to YLGD-YLGDSY.

Next, the ligation mixture was purified using HPLC and as YLGD-$K^{\varepsilon 15N}$LI and YLGD-YLGD were co-eluting and thus not separated using HPLC, NMR spectra were recorded of the fraction containing both peptides. First, we investigated whether both α-linked and ε-linked YLGD-$K^{\varepsilon 15N}$LI were formed during the ligation reaction. If both types of ligation would occur, a ^{15}N NMR spectrum should show two distinct peaks, one corresponding to the amine, and one to the amide (Figure 2A). As direct detection of ^{15}N atoms suffers from low sensitivity and long relaxation times,[14] ^{15}N signals are usually inverse-detected via the proton signal in a ^{1}H-^{15}N heteronuclear single quantum coherence (HSQC) experiment using a fixed $J_{H,N}$ coupling constant. Whereas amide groups can be readily detected in an HSQC experiment via the $^{1}J_{H,N}$ coupling, NMR characterization of lysine NH_3 groups via the $^{1}J_{H,N}$ coupling is challenging, due to rapid hydrogen exchange with water (Supplementary Information and Supplementary Figure 1).[15] In accordance, the ^{15}N amine signal could not be detected in HSQC experiments using $J_{H,N}$ coupling constant ranging from 55 to 90 Hz (data not shown), which covered the whole range of reported amine $^{1}J_{H,N}$ coupling constants (55-80 Hz).[14] When

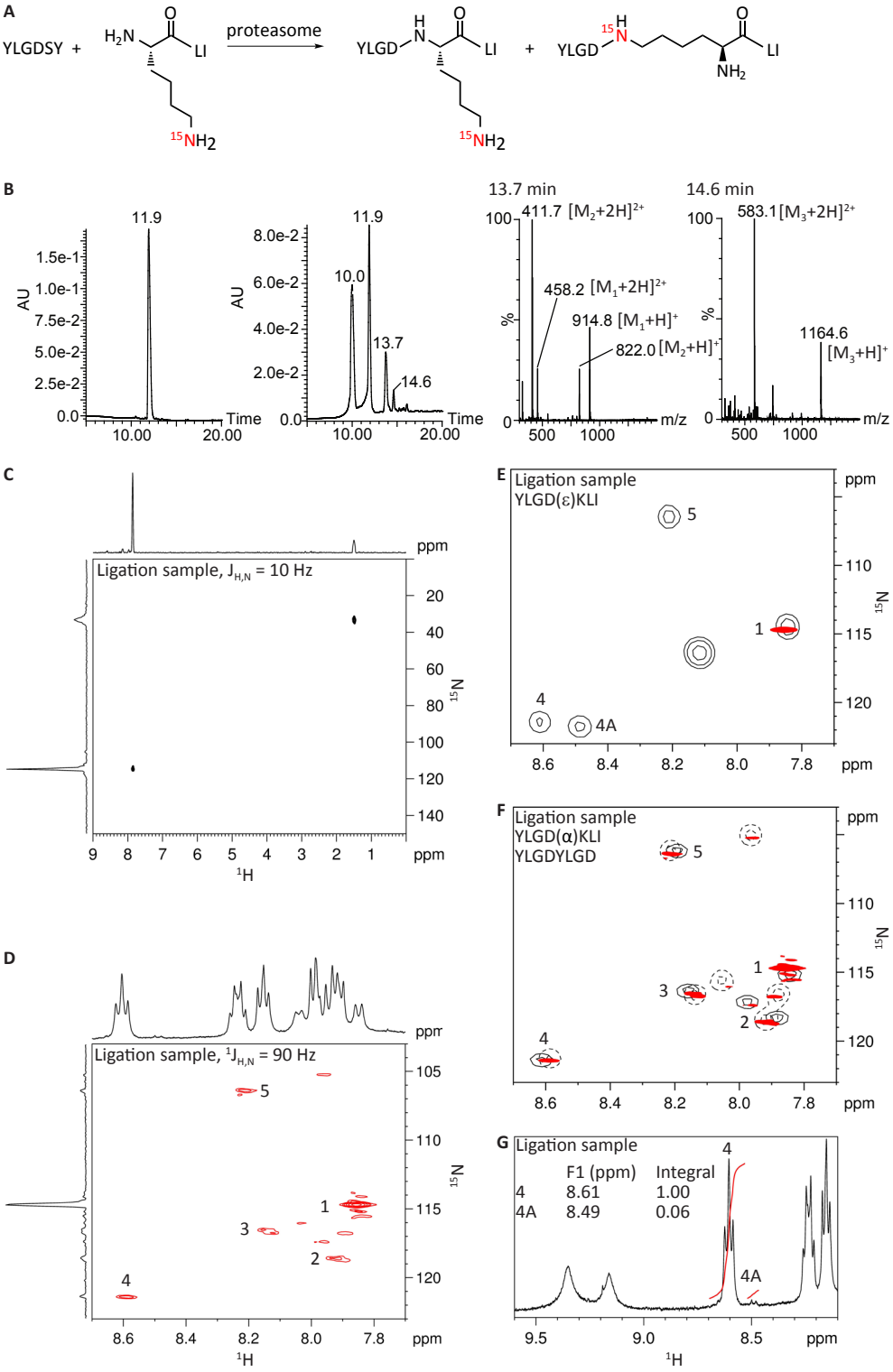

A
YLGDSY +
proteasome
YLGD
LI
+
YLGD
LI
B
11.9
10.0
13.7
14.6
AU
Time
13.7 min
411.7 [M2+2H]2+
458.2 [M1+2H]2+
914.8 [M1+H]+
822.0 [M2+H]+
14.6 min
583.1 [M3+2H]2+
1164.6
[M3+H]+
m/z
C
Ligation sample, JH,N = 10 Hz
ppm
15N
1H
D
Ligation sample, 1JH,N = 90 Hz
E
Ligation sample
YLGD(ε)KLI
F
Ligation sample
YLGD(α)KLI
YLGDYLGD
G
Ligation sample
F1 (ppm) Integral
4 8.61 1.00
4A 8.49 0.06

Figure 2 | ε-ligation can compete with α-ligation during proteasomal splicing reactions. **A)** Scheme showing the proteasomal ligation reaction of YLGDSY with $K^{\varepsilon 15N}$LI. **B)** HPLC profile (UV detection at 280 nm) of the ligation mixture before incubation (left panel) and after incubation (second panel). In-line MS spectra of the peaks eluting at 13.7 min (third panel) and 14.6 min (right panel). M_1: YLGD-KLI; M_2: YLGD-YLGD; M_3: YLGD-YLGDSY. **C)** ^{1}H-^{15}N HMQC spectrum of the ligation sample, recorded using $J_{H,N}$ = 10 Hz. 1D projections of the ^{1}H and ^{15}N signals are shown on top and left axes, respectively. **D)** ^{1}H-^{15}N HSQC spectrum of the ligation sample, recorded using $^1J_{H,N}$ = 90 Hz. The 1D projection of the ^{15}N signal is shown on the left axis, the ^{1}H spectrum of the ligation sample is shown on the top axis. Peaks were assigned as follows, 1: $^{\varepsilon}$K5, 2: not assigned, 3: L2/D4, 4: L2/D4, 4A: not assigned, 5: G3 (See Supplementary Information and Supplementary Figure 2 for details). **E)** Overlay of the ^{1}H-^{15}N HSQC spectra of the $^{15}N^{\varepsilon}$-enriched ligation sample (red) and synthetic, non ^{15}N enriched YLGD$^{\varepsilon}$KLI (black) **F)** Overlay of the ^{1}H-^{15}N HSQC spectra of the $^{15}N^{\varepsilon}$ enriched ligation sample (red), synthetic non ^{15}N-enriched YLGD$^{\alpha}$KLI (black, solid lines), and synthetic non ^{15}N enriched YLGDYLGD (black, dotted lines). **G)** ^{1}H spectrum of the ligation sample and integrals of peaks 4 and 4A. Peak 4 is attributed to L2 or D4 and is present in the ^{1}H spectra all products, whereas peak 4A is present in the ^{1}H spectrum of YLGD$^{\varepsilon}$KLI only (Supplementary Figures 2 and 3)

^{15}N signals are on the other hand inverse-detected via long-range $^2J_{H,N}$ and $^3J_{H,N}$ couplings ($J_{H,N}$ = 0.3 to 16 Hz),[14] both amide and amine resonances can be measured simultaneously in a single experiment (Supplementary Information and Supplementary Figure 1). Therefore, we performed a ^{1}H-^{15}N HMQC measurement, using a $J_{H,N}$ coupling constant of 10 Hz. The resulting HMQC spectrum showed two peaks, with shifts of 7.85; 114.7 and 1.45; 33.2 ppm (Figure 2C). Amines display a chemical shift of 0 to 70 ppm, while amides have a chemical shift of 80 to 170 ppm, compared to a $^{15}NH_3$ reference signal.[14] This indicates that during proteasomal digestion of YLGDSY in the presence of $K^{\varepsilon 15N}$LI, ligation products containing ^{15}N-enriched amines and ^{15}N-enriched amides were both formed. From these data we conclude that during proteasomal splicing reactions ε-ligation competes with α-ligation.

ε-ligation occurs in 1 out of 10 ligation events

Next, we set out to quantify the relative amount of ε-ligation product formed. It is however not possible to directly compare the ^{15}N amide and ^{15}N amine signals in a single HSQC measurement. ^{1}H-^{15}N correlations of the lysine NH_3 group are often difficult to observe in HSQC experiments optimized for backbone amides, even when the water exchange rate is slow enough to permit their detection by ^{1}H NMR.[15] This is due to two factors. First, the ^{15}N chemical shift of lysine NH_3 groups resonates ~ 90 ppm upfield from the backbone amide signals. As the rf strength that can be used for ^{15}N pulses is limited, application of a ^{15}N 180° pulse at 115 ppm (close to the amide resonance frequency) leads to an imperfect pulse at 30 ppm (close to the lysine NH_3 resonance frequency) and therefore to loss of signal intensity, and vice versa.[15] With standard HSQC pulse programs it is therefore not possible to obtain maximal signal intensities for ^{15}N amide and ^{15}N amine groups simultaneously. Second, the ^{15}N transverse relaxation of lysine NH_3 groups has been shown to be highly affected by water exchange, resulting in broadening of ^{15}N line shapes and a decrease in resolution and sensitivity,[15] as visible in Figure 2C. Therefore, the relative amount of ε-ligation product compared to α-ligation product was estimated by quantifying the signal intensity of the newly formed ^{15}N enriched amide bond

using amide resonances within the peptide due to naturally abundant ^{15}N as an internal reference. To ensure maximum signal intensities of amide resonances, the YLGD-K$^{\varepsilon 15N}$LI containing fraction was analyzed again in an HSQC experiment using a $^{1}J_{H,N}$ coupling constant of 90 Hz.[16] In the resulting spectrum, one intense cross peak was visible at 7.86; 114.7 ppm, which originated from the ^{15}N enriched isopeptide amide in YLGD-$^{\varepsilon}$KLI (Figure 2D, peak 1). In addition, several weaker signals were detected, which originate from naturally abundant ^{15}N amide signals of different ligation products.

The ligation sample contained the following ligation products: YLGD-$^{\alpha}$KLI, YLGD-$^{\varepsilon}$KLI and YLGD-YLGD as indicated by LC-MS (where a dash indicates the newly formed pepide bond and α/ε the type of linkage) (Figures 2B and 2C). We synthesized these three peptides and recorded ^{1}H and ^{1}H-^{15}N HSQC spectra as a reference to assign ^{1}H-^{15}N cross-peaks to one or more peptides (Supplementary Information and Supplementary Figure 2). An overlay of the ^{1}H-^{15}N HSQC spectra of the ^{15}N enriched ligation sample and synthetic YLGD$^{\varepsilon}$KLI confirmed that peak 1 resulted from the ^{15}N enriched $^{\varepsilon}$K5 resonance (Figure 2E). In addition, the overlay of HSQC spectra of the ligation sample, synthetic YLGD$^{\alpha}$KLI, and synthetic YLGDYLGD showed that all weaker signals could be attributed to naturally abundant ^{15}N amide resonances in one or more ligation products (Figure 2F). In the HSQC spectrum of the ligation sample, 56% of the signal of overlapping cross-peaks originated from YLGD-YLGD, while 44% originated from YLGD-$^{\alpha/\varepsilon}$KLI (Supplementary Information and Supplementary Figure 3). We integrated the signals of the $^{\varepsilon}$K5 cross peak (peak 1) and the 4 most intense naturally abundant cross-peaks (peaks 2 to 5) and calculated the signal intensity originating from YLGD-$^{\alpha/\varepsilon}$KLI resonances only (Table 1; 44% of total integrals). Taking the abundance of ^{15}N signals into account, the relative percentage of ε-ligation (YLGD-$^{\varepsilon}$KLI) versus α-ligation (YLGD-$^{\alpha}$KLI) was estimated to be 10.3 ± 1.7% (average ± SD of peaks 2 to 5). This relative percentage of ε-ligation could also be derived from the ^{1}H spectra of the ligation sample (Figure 2G). In this spectrum, the triplet at 8.61 ppm originates from naturally abundant peak 4, which is present in all peptides, while the doublet at 8.49 ppm originates from naturally abundant peak 4A, which is only present in YLGD$^{\varepsilon}$KLI (Supplementary Figure 2D). From the integral of these resonances, we derived that YLGD-$^{\varepsilon}$KLI constituted 6% of total ligation products (YLGD-YLGD + YLGD-$^{\alpha}$KLI + YLGD-$^{\varepsilon}$KLI) and estimated the relative percentage of ε-ligation versus α-ligation to be 14%. Both percentages also correlate with the LC-MS experiments described above, in which relative ε-ligation efficiencies of 10% were observed. Together, these data indicate that in one out of 8 to 10 ligation reactions ε-ligation occurred as opposed to α-ligation.

Table 1 | Integrals of ^{15}N enriched and naturally abundant amide cross-peaks.

Peak	Residue	f2 (ppm)	f1 (ppm)	Total Integral	YLGD-KLI only	^{15}N abundance
1	$^{\varepsilon}$K5	7.8552	114.6679	209654420	209654420	98%
2	-	7.9136	118.6336	20745315	9127938	0.37%
3	L2/D4	8.1374	116.6248	20142896	8862874	0.37%
4	L2/D4	8.5903	121.4045	15826456	6963640	0.37%
5	G3	8.2106	106.3924	14876321	6545581	0.37%

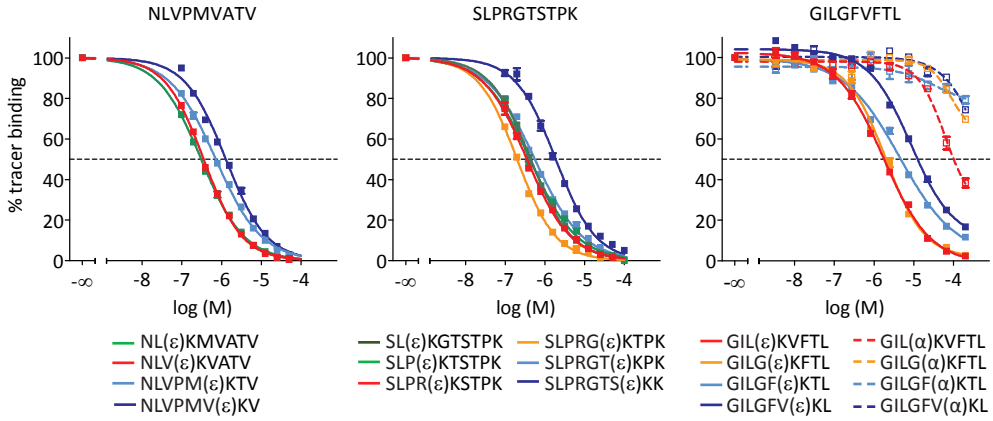


Figure 3 | Isopeptides have increased affinity for HLA-A complexes. Binding curves of (iso)peptides based on the epitopes NLVPMVATV (specific for HLA-A2.1), SLPRGTSTPK (specific for HLA-A3) and GILGFVFTL (specific for HLA-A2.1). (ε)K indicates the location of the isopeptide linkage. (α)K indicates a normally linked lysine residue.

Isopeptides have increased affinity for HLA-A complexes

It is likely that epitopes containing an isopeptide linkage have binding affinities for HLA-A molecules that differ from their non-isopeptide counterparts. To test whether epitopes containing an isopeptide bond could still bind HLA-A complexes, we measured the affinity of a set of isopeptide bond-containing peptides for HLA-A2.1 and HLA-A3. Isopeptides were based on three known epitopes, GILGFVFTL (influenza matrix protein 1 58-66),[17] NLVPMVATV (human cytomegalovirus pp65 495-503),[18] which are both presented on the cell surface by HLA-A2.1, and SLPRGTSTPK (spliced together from SP110 nuclear body protein 296-301 and 286-289),[10] which is presented by HLA-A3. As the length of an N^{ε}-linked lysine residue in a peptide backbone is twice that of an N^{α}-linked lysine residue (See Figure 2A), we replaced two adjacent amino acid residues at different positions in the epitope by one N^{ε}-linked K (Table 2). To ensure binding, the anchor residues on P2 and P9 (P10 in case of SLPRGTSTPK) were not modified.

To investigate binding to HLA-A2.1 or HLA-A3, a UV-mediated MHC exchange fluorescence polarization (FP) assay was used.[19,20] In this assay, HLA-A2.1/HLA-A3 complexes loaded with a UV-cleavable peptide ligand are subjected to UV light, which cleaves the ligand, resulting in the generation of "empty" HLA-A2.1/HLA-A3.[21] These "empty" HLA-A2.1/HLA-A3 complexes quickly disintegrate, 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 measured.[19,20] Binding curves of the NLVPMVATV-, SLPRGTSTPK- and GILGFVFTL-based isopeptides were measured (Figure 3, solid lines) and the affinities (expressed as IC_{50} values), defined as the concentration of peptide that inhibited 50% of tracer peptide binding, were determined (Table 2). For all potential epitopes, the presence of an isopeptide linkage did not abrogate HLA-A2.1/HLA-A3 binding, as all peptides were able to bind to HLA-A2.1 or HLA-A3 complexes with similar

Table 2 | IC_{50} values of different (iso)peptides for binding to HLA-A2.1 and HLA-A3.

Parent peptide	Replaced positions	Isopeptide	IC_{50} (μM)	Peptide	IC_{50} (μM)
GILGFVFTL	-	-	1.7	-	-
	3-4	GI(ε)KFVFTL	ND	GI(α)KFVFTL	ND
	4-5	GIL(ε)KVFTL	1.8	GIL(α)KVFTL	111
	5-6	GILG(ε)KFTL	2.0	GILG(α)KFTL	509
	6-7	GILGF(ε)KTL	4.8	GILGF(α)KTL	577
	7-8	GILGFV(ε)KL	14.5	GILGFV(α)KL	601
NLVPMVATV	-	-	4.7		
	3-4	NL(ε)KMVATV	0.31		
	4-5	NLV(ε)KVATV	0.35		
	5-6	NLVP(ε)KATV	ND		
	6-7	NLVPM(ε)KTV	0.71		
	7-8	NLVPMV(ε)KV	1.3		
SLPRGTSTPK	-	-	0.19		
	3-4	SL(ε)KGTSTPK	0.44		
	4-5	SLP(ε)KTSTPK	0.42		
	5-6	SLPR(ε)KSTPK	0.35		
	6-7	SLPRG(ε)KTPK	0.20		
	7-8	SLPRGT(ε)KPK	0.55		
	8-9	SLPRGTS(ε)KK	1.9		

(ε)K indicates the location of the isopeptide linkage. (α)K indicates a normally linked lysine residue. ND: not determined

affinities as the parent peptides (Table 2). In general, the isotope linkage was tolerated in the middle of the epitope (Figure 3, red and orange lines) or towards the N-terminus of the epitope (Figure 3, green lines) without loss of HLA class I affinity. When the isopeptide linkage was shifted towards the C-terminal side of the epitope on the other hand, affinity decreased. For comparison, the affinities of GILGFVFTL-based epitopes with identical sequences but a normal peptide bond were determined. As these peptides are 8 amino acids in length, they are expected to have low affinity to HLA-A2.1, which only binds peptides of 9-10 amino acids in length with high affinity. As can be judged from Figure 3 (dotted lines) and Table 2, the binding affinity of these GILGFVFTL-based peptides was decreased by a factor 40 to 200 compared to their isopeptide counterparts. Together, these data suggest that isopeptides are not only formed by the proteasome, but isopeptides of different lengths can also be presented at the cell surface by different HLA class I alleles. In addition, the presence of an isopeptide linkage enables 8-mer isopeptides to bind HLA-A2.1/HLA-A3 complexes with much increased affinity as compared to normal 8-mer peptides that do not contain an isopeptide linkage.

Isopeptides have enhanced stability towards proteasomal degradation

Intramolecular isopeptide bonds have been shown to provide proteolytic stability to pili on the surface of Gram-positive bacteria,[22-24] suggesting that epitopes containing an isopeptide linkage may be less susceptible to processing by proteases. Therefore, we investigated whether isopeptides showed enhanced stability to proteasomal processing

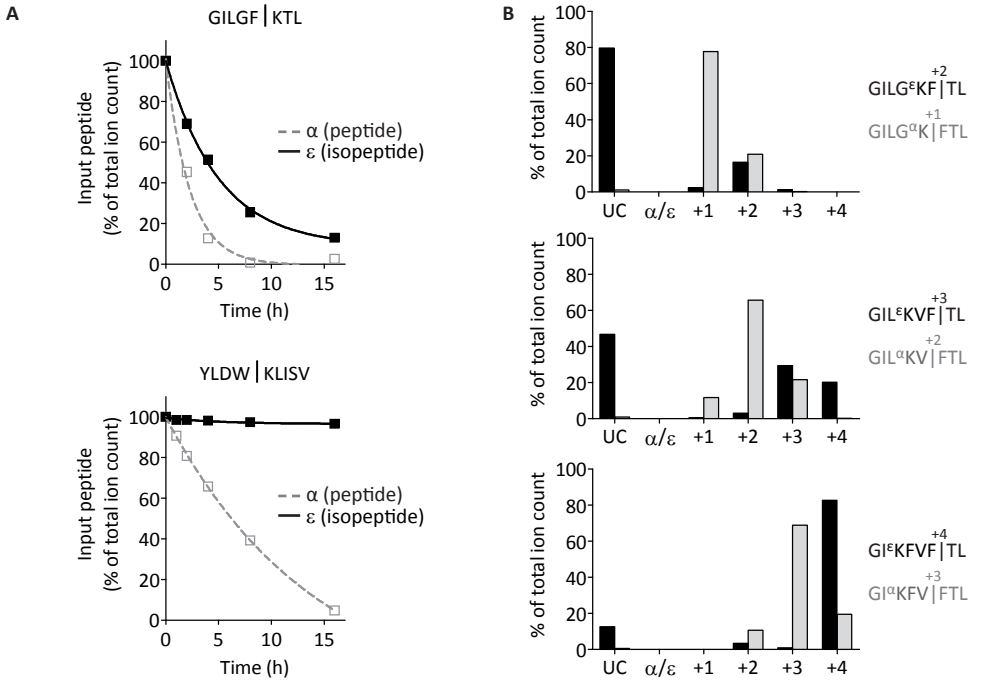


Figure 4 | Isopeptides show enhanced stability towards proteasomal degradation. A) Degradation of GILGF$^{\alpha/\varepsilon}$KTL and YLDW$^{\alpha/\varepsilon}$KLISV by the proteasome as a function of time. The amounts of input peptides and degradation products were quantified using LC-MS. The site of cleavage and the location of the lysine N$^{\alpha}$- or N$^{\varepsilon}$-linkage is indicated with a vertical line. **B)** Proteasomal degradation patterns of the indicated peptides. The amounts of input peptides and degradation products were quantified using LC-MS after incubation with proteasome for 16 h. Black bars indicate isopeptides and corresponding degradation products, grey bars indicate normal peptides and corresponding degradation products. The main site of cleavage is indicated with a vertical line. The location of the lysine N$^{\alpha}$- or N$^{\varepsilon}$-linkage is indicated with α or ε. Cleavage products are indicated by the site of cleavage relative to the lysine N$^{\alpha}$- or N$^{\varepsilon}$-linkage. UC: uncleaved input peptide.

over normal peptides. First, we investigated peptides in which cleavage of the (iso)peptide bond was the only proteasomal cleavage occurring. To this end, YLDW$^{\alpha/\varepsilon}$KLISV and GILGF$^{\alpha/\varepsilon}$KTL were incubated with proteasome for time periods up to 16 h, followed by LC-MS analysis of the degradation mixtures (where α/ε indicates the location of the lysine N$^{\alpha}$- or N$^{\varepsilon}$-linkage). The amounts of the different input peptides were plotted as the percentage of total ion count versus time (Figure 4A). GILGF$^{\alpha}$KTL (dotted lines) was cleaved readily ($t_{1/2} \sim 2$ h), while GILGF$^{\varepsilon}$KTL (solid lines) was cleaved at a slower rate ($t_{1/2} \sim 4$ h). The difference in hydrolysis rate between N$^{\alpha}$-linked and N$^{\varepsilon}$-linked lysine residues was even more apparent in YLDWKLISV. Whereas YLDW$^{\alpha}$KLISV was cleaved with a half-life of 6 h, less then 5% of YLDW$^{\varepsilon}$KLISV was cleaved within 16 h.

Next, we questioned whether isopeptides that were cleaved at positions other then at the isopeptide bond were also more stable to proteasomal processing compared to normal peptides. We incubated GI$^{\alpha/\varepsilon}$KFVFTL, GIL$^{\alpha/}$

$^{\varepsilon}$KVFTL, GILG$^{\alpha/\varepsilon}$KFTL and GILGFV$^{\alpha/\varepsilon}$KL for 16 h with proteasome and analyzed the resulting digestion mixtures with LC-MS to assess the extent as well as the site of peptide cleavage. The amounts of input peptide and cleavage products were quantified by LC-MS and plotted as the percentage of total ion count, which is a measure of total peptide content (Figure 4B). Cleavage products were indicated by the site of cleavage relative to the lysine N$^{\alpha}$- or N$^{\varepsilon}$-linkage (e.g. +1 cleavage products result from cleavage 1 residue towards the C-terminus from $^{\alpha/\varepsilon}$K). For all peptides in which the isopeptide linkage was located towards the N-terminus from the site of cleavage, differences were found both in the rate and the site of hydrolysis. The difference between peptide and isopeptide processing was most apparent when the isopeptide linkage was located within 1 to 2 residues of the site of cleavage (Figure 4B, top panel). The peptide GILG$^{\alpha}$KFTL was processed almost completely within 16 h and was cleaved predominantly at the KF bond, which is located one residue towards the C-terminus from $^{\alpha}$K (+1 cleavage products). Only 20% of GILG$^{\varepsilon}$KFTL was processed within 16 h on the other hand, and the main cleavage site in GILG$^{\varepsilon}$KFTL was the FT bond located two residues (corresponding to a length of three normal residues) from the isopeptide linkage (Figure 4B, top panel). Similar patterns were found when GIL$^{\alpha/\varepsilon}$KVFTL (Figure 4B, middle panel) and GI$^{\alpha/\varepsilon}$KFVFTL (Figure 4B, lower panel) were compared. Both GIL$^{\alpha}$KVFTL and GI$^{\alpha}$KFVFTL were completely processed within 16 h, while only 50% and 90% of GIL$^{\varepsilon}$KVFTL and GI$^{\varepsilon}$KFVFTL, respectively, were cleaved within the same time frame (Figure 4B). In line with results described above, the site of cleavage in isopeptides was located further away from the lysine residue compared to the normal peptides. Whereas GIL$^{\alpha}$KVFTL and GI$^{\alpha}$KFVFTL were cleaved preferably 2 and 3 residues towards the C-terminus from the N$^{\alpha}$-linked lysine, GIL$^{\varepsilon}$KVFTL and GI$^{\varepsilon}$KFVFTL were cleaved 3 and 4 residues away from the N$^{\varepsilon}$-linked lysine residue (Figure 4B). No differences in peptide composition were found between digestion mixtures of GILGFV$^{\alpha}$KL and GILGFV$^{\varepsilon}$KL, which are the only peptides in this series in which the $^{\alpha/\varepsilon}$K was located towards the C-terminus from the site of cleavage (data not shown). Together, these data suggest that peptides containing an isopeptide linkage are less susceptible to proteasomal degradation. This holds true both if the isopeptide linkage is the preferred site of cleavage and if the isopeptide linkage is located close to the preferred cleavage site. In addition, when proteasomal processing of isopeptides did occur, the actual cleavage site was located further away from the $^{\varepsilon}$K residue compared to an $^{\alpha}$K resiude.

DISCUSSION

Recent studies have shown that proteasomal splicing reactions can create a novel type of antigen via a transpeptidation mechanism.[8-11] During proteasomal transpeptidation reactions, a protein amide backbone linkage is attacked by one of the proteasomes catalytic N-terminal threonine residues, resulting in the formation of an *O*-acyl enzyme intermediate and the release of the C-terminal part of the protein. Subsequently, this intermediate ester reacts with an amino group, usually the amino terminus, from a second peptide (the 'C-terminal ligation partner'), resulting in the formation of a novel peptide linkage and a spliced product.[9,10] This transpeptidation reaction can apparently compete with normal hydrolysis, in which water reacts with the intermediate ester,[12] in such a way that sufficiently new peptide can be formed to invoke an immune response against transpeptidation products. Transpeptidation reactions can efficiently compete with hydrolysis if specific

structural requirements on the two ligating partners are met, as is discussed in Chapter 3.2. Ligation occurs particularly efficient if the C-terminal ligation partner has a basic amino acid residue (lysine or arginine) at the N-terminus (Chapter 3.2). As lysine has two amino groups that can theoretically both react with the *O*-acyl enzyme intermediate, this implies that the proteasome may be able to from isopeptide linkages.

In the present study we show for the first time that the proteasome can use both the α-amino group and the ε-amino group of lysine as a C-terminal ligation partner, suggesting that also *in vivo* the proteasome can create an isopeptide linkage and form a novel type of antigen. When single amino acids are ligated onto a precursor peptide, both amino groups were equally reactive and capable of participating in ligation reactions. However, when peptide ligation was monitored by NMR, ε-ligation was found to be 10-fold less efficient as compared to α-ligation. Likely, a single lysine amino acid is capable of fitting into the S_1' pocket in multiple orientations, thereby facilitating both α- and ε-ligation. A peptide on the other hand will adopt an orientation whereby its P_2' and P_3' side chains fit in the S_2' and S_3' pockets. This may favor orientation of the lysine residue in the S_1' pocket in such a way that the α-amino group is positioned correctly for a nucleophilic attack on the intermediate ester bond, resulting in normal peptide bond formation, disfavoring alternative ε-ligation and isopeptide bond formation.

The data presented here also show that isopeptides have unique properties that discern them from normal epitopes. First, isopeptides of various lengths can bind different HLA alleles with high affinity. The isopeptide linkage is tolerated in the middle and towards the N-terminus of different epitopes without loss of affinity for HLA-A2/A3 complexes. Only location of the isopeptide linkage towards the C-terminus of an epitope leads to loss of HLA-A2/A3 affinity. In addition, the presence of an isopeptide linkage allows peptides that are comprised of only 8 amino acids to bind HLA-A2/A3, whereas normally only peptides of 9 to 11 residues in length bind HLA-A2/A3 complexes with high affinity. Second, isopeptides are more stable towards further proteasomal processing compared to normal peptides. Degradation of isopeptides is slow compared to common peptides, both when the isopeptide bond itself is the preferred site of cleavage, and when the preferred site of cleavage is located close to the isopeptide linkage. In addition, the proteasome cleaves most isopeptides at a site that is located further away from the lysine residue compared to normal peptides. This indicates that proteasomal processing of isopeptides results in the generation of a peptide fragment in which the isopeptide linkage is still present and located towards the N-terminus of the peptide, favoring affinity for HLA-A2/A3. Together, these properties are likely to increase the fraction of ε-ligated peptides that enters the endoplasmic reticulum, and that is loaded onto MHC class I and transported to the cell surface for CD8+ T cell surveillance. As isopeptides are more stable towards proteasomal degradation, we consider it likely that the isopeptide linkage also hampers hydrolysis by other (cytosolic) proteases. Importantly, isopeptides that are formed and subsequently processed by the proteasome are likely to retain their isopeptide linkage at a location (towards the N-terminus) that does not affect HLA-A2/A3 affinity. Therefore, we anticipate an increased chance that an isopeptide containing epitope is efficiently loaded onto MHC allowing it to reach the cell surface once it is formed.

It has been proposed that proteasomal antigen splicing increases the diversity of the peptide repertoire available for presenta-

tion on MHC class I.[12,25] A greater diversity of epitopes increases the potential epitope repertoire that may be recognized by CD8+ T cells, which ultimately results in the elimination of infected or malignant cells.[12] Increasing diversity of T cell receptors by V(D)J recombination is essential for protection from a constantly changing bacterial and viral repertoire.[25] In addition, proteasomal splicing may increase diversity on the protein level. The data presented here suggest that also proteasomal $^{\varepsilon}$K-ligation reactions may increase immune diversity. To date, three spliced epitopes have been described *in vivo*,[8-10] suggesting that proteasomal splicing does occur in cells and with sufficient efficiency to evoke an immune response against spliced epitopes. The splicing efficiencies of two of the three described epitopes have been calculated and one spliced epitope was estimated to be produced from $\sim 5 \times 10^5$ to 1×10^4 molecules of precursor peptide (0.0002% to 0.01% efficiency).[9,11] This indicates that even very low splicing efficiencies lead to sufficient amounts of peptide for detection by CD8+ T cells. In a previous study, we have shown that splicing efficiencies can be several orders of magnitude higher if specific structural requirements for the two ligating partners are met (Chapter 4.2). We estimated that ligations reaction involving C-terminal fragments with a lysine residue at the site of ligation occurred with efficiencies of 1 to 10% (Chapter 4.2). This indicates that the overall efficiency of $^{\varepsilon}$K-ligation reactions, which is 10-fold lower as compared to α-ligation, can be as high as 0.1 to 1%, depending on the sequence of N- and C-terminal ligation partners, suggesting that one isopeptide may be formed from 100 to 1000 precursor molecules. Since a large fraction of spliced isopeptide epitopes may be able to reach the cell surface as a result of their increased proteolytic stability, ε-ligation should produce sufficient amounts of peptide to evoke a T cell response, suggesting that isopeptides may play a role in immunity *in vivo*. Thus although the epsilon linkage has not been evidenced before in MHC epitopes yet, which is easily explained by its difficult (and so far unanticipated) detection, we consider it more than likely that they will be detected, based on our *in vitro* experiments. With evidence of proteolytic stability and therefore accumulation of this modification we postulate it as a genuine post-translational modification resulting from transpeptidation mechanisms. We consider it likely that our *in vitro* findings will be validated *in vivo* soon.

MATERIALS AND METHODS

Peptide building blocks were purchased from Novabiochem and appropriately functionalized resins from Applied Biosystems. ^{15}N-ε-L-Lysine HCl was purchased from C/D/N Isotopes Inc (Pointe-Claire, QC, Canada). 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. All ^{1}H and ^{1}H-^{15}N HSQC (Heteronuclear Single Quantum Coherence) experiments were carried out in DMSO-d6 using a Bruker Avance 300 (^{1}H: 300 MHz, ^{15}N: 30.4 MHz) spectrometer. HMQC (Heteronuclear Multiple Quantum Coherence) experiments were performed on a Bruker ARX 400 (^{1}H: 400 MHz, ^{15}N: 40.5 MHz) spectrometer. NMR data were processed and analyzed using Bruker v2.1 Topspin software.

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 either directly fully deprotected and released from the resin directly by treating the resin with TFA/H_2O/triisopropylsilane (93/5/2 v/v) for 2.5 h or peptides were treated with 4 equivalents each of acetic anhydride and dipea for 40 minutes to acetylate the N-terminus prior to deprotection and release from the resin. Peptides containing N^{ε}-linked lysine residues were synthesized using Boc-L-lysine(Fmoc)-OH. Incorporation of this building blocked followed by removal of the Fmoc group ensures that peptide synthesis continues at the Lysine ε-amine group, resulting in isopeptide linkage containing peptides. Peptides were precipitated with cold diethylether/pentane (3/1 v/v) and lyophilized from H_2O/ACN/HOAc (65/25/10 v/v).

Synthesis of ^{15}N-ε-Fmoc-Lys(Boc)-OH

N^{α}-(9-Fluorenyl)methoxycarbonyl-*$^{15}N^{\varepsilon}$-tert*-butyloxycarbonyl-L-lysine was synthesized as described.[26] ^{1}H NMR (400 MHz, *d_6*-dmso) δ = 12.53 (bs, COOH), 7.91 (d, *J*=7.5 Hz, 2H, H_{Ar}), 7.73 (d, *J*=7.5 Hz, 2H, H_{Ar}), 7.62 (d, *J*=7.6 Hz, 1H, NH), 7.42 (t, *J*=7.5 Hz, 2H, H_{Ar}), 7.33 (t, *J*=7.5 Hz, 2H, H_{Ar}), 6.79 (dt, J_{NH}=92 Hz, J_{HH}=6.8 Hz, 1H, $^{15}N^{\varepsilon}H$), 4.29-4.26 (m, 2H, OCH_2), 4.24-4.20 (m, 1H, $OCH_2C\underline{H}$), 3.93-3.88 (m, 1H, H^{α}), 2.91-2.78 (m, 2H, CH_2), 1.70-1.52 (m, 2H, CH_2), 1.36 (s, 9H, $C(CH_3)_3$), 1.33-1.27 (m, 4H, 2xCH_2). ^{13}C NMR APT (300 MHz, $CDCl_3$) δ = 175.51 (COOH), 156.29 (2x CO), 143.74 (2x$C_{q\text{-}Ar}$), 141.30 (2x$C_{q\text{-}Ar}$), 127.69 (2xCH_{Ar}), 127.08 (2xCH_{Ar}), 125.14 (2xCH_{Ar}), 119.95 (2xCH_{Ar}), 79.57 ($\underline{C}CH_3$), 67.04 (CH_2O), 53.71 ($C^{\alpha}H$) 47.16 ($\underline{C}HCH_2O$), 40.15 ($C^{\varepsilon}H_2$, $^1J_{CN}$=11 Hz), 31.76 ($C^{\beta}H_2$), 29.54 ($C^{\delta}H_2$), 28.41 ($(CH_3)_3$), 22.31 ($C^{\gamma}H_2$). $^{15}N^1H$-HMQC (6.79 ppm, 83.7 ppm). MS (ESI): $[M+H]^+_{calc}$=470.2, $[M+H]^+_{found}$=470.1.

Proteasome purification

Proteasome was purified from bovine liver as described.[4] 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.[27] 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).

MALDI experiments

78 pmol C^{TMR}SLPRGTASSR was incubated with 10 μg proteasome in ligation buffer (50 mM Tris-HCl, pH 8.5) containing 6% DMSO for 16 h at 45°C. In different experiments, 100 μM MG132, 100 nM bortezomib, or 0.5 M lysine, 0.5 M N^{α}-acetylated lysine or 0.5 M N^{ε}-acetylated lysine were added to the reaction mixture. Following incubation, reaction mixtures were freeze-dried and dissolved in 50 μL acetonitrile/H_2O (5:95 v/v) containing 0.1 % TFA. Of this solution 10 μL was used for purification and desalting using reversed-phase ZipTip®C18 tips (Millipore, C18, spherical silica, 15 um, 200 A pore size). Peptide samples were bound to the ZipTip pipette tip by aspirating and dispensing 10 μL sample for approximately 20 cycles until the solution appeared nearly colorless and the ZipTip pipette tip had turned pink due to retained labeled peptides. The ZipTip pipette tip was washed twice with 10 μL of 0.1% TFA. Retained peptides were eluted by aspirating and dispensing 4 μL acetonitrile through the ZipTip pipette tip at least five times (the ZipTip pipette tip turned colorless). Eluates were analyzed by MALDI mass spectrometry. MALDI-TOF experiments were carried out on an Autoflex, linear MALDI-TOF-MS (Bruker Daltonik GmbH, Bremen, Germany). Droplets of 0.5 μL 10 mg/ml 2,5-dihydroxybenzoic acid (DHB, Bruker Daltonik) matrix solution in 0.1% TFA were spotted onto a MALDI target plate and 0.5 μL of the purified and desalted samples were pipetted into the DHB matrix droplets and left drying. For calibration, Peptide Calibration Standard (M 1046-3147 Da, Bruker Daltonik) was used. Spectra were analyzed with Bruker Daltonics FlexAnalysis Software. 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 and I_H is the peak intensity of the hydrolysis product.

Peptide ligation experiments

To compare acetylated and non-acetylated peptides, 50 nmol of the peptide YLGDSY was incubated with 250 nmol of (acetylated) C-terminal ligation fragment and 5 μg proteasome in a total volume of 75 μL ligation buffer (50 mM Tris-HCl, pH 8.5) containing 2% DMSO for 16 h at 37 °C. Incubation mixtures were snap-frozen, lyophilized, dissolved in DMSO/H_2O/ACN (1/2/1 v/v) and filtered over Strata™-X 33 μm polymeric reversed phase disposable columns (Phenomenex). Eluted fractions were analyzed by LC-MS.

For NMR experiments, 16 μmol $^{\varepsilon(15N)}$KLI was incubated with 80 μmol YLGDSY and 2.8 mg proteasome in 23.2 mL ligation buffer containing 4.8% DMSO for 24 h at 37°C. The reaction mixture was snap-frozen, lyophilized, dissolved in 12 mL acetonitrile/H_2O (1:1) and filtered over a Strata™-X 33 μm polymeric reversed phase disposable column (Phenomenex) to remove proteasomal proteins. The filtrate was lyophilized again, dissolved in 7.5 mL DMSO/H_2O (1:4) and diluted and purified by HPLC over an Atlantis® Preparative T3 column (5 μm; 20 x 150 mm; Waters) using a linear gradient (5% to 50% acetonitrile in H_2O containing 0.05% TFA). HPLC purifications 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. The YLGD-$^{\varepsilon(15N)}$KLI containing fraction was lyophilized and dissolved in 0.4 mL DMSO-d6 for NMR spectroscopy analysis. Reference spectra of synthetic YLGD$^{\alpha}$KLI, YLGD$^{\varepsilon}$KLI and YLGDYLGD were recorded using a 50 mM peptide solution in DMSO-d6. HSQC experiments were performed using a Bruker preprogrammed gradient enhanced echo-antiecho HSQC pulse program[28,29] (hsqcetgp, avance-version, 2D H-1/X correlation via double inept transfer, phase sensitive using Echo/Antiecho-TPPI gradient selection with decoupling during acquisition, using trim pulses in inept transfer) optimized for the detection of ^{15}N signals and ^{15}N signals were compared to a $^{15}NH_3$ reference signal. The relative percentage of ε-ligation versus α-ligation was calculated using the following formula: Efficiency of ε-ligation = I_1/(0.44×I_N) × 0.37/98 × 100%, where I_1 is the integral of the $^{15}N^{\varepsilon}$-enriched peak 1, I_N is the integral of peak N, resulting from a naturally abundant amide resonance, 0.44 is the fraction YLGD-$^{\alpha/\varepsilon}$KLI in the ligation sample as compared to total peptide content, 0.37 is the percentage of naturally abundant ^{15}N and 98 is the percentage of ^{15}N enrichment of $^{15}N^{\varepsilon}$-enriched peak 1. The efficiency of ε-ligation was calculated for peaks 2 to 5 and averaged.

Fluorescence polarization HLA-A2.1 and HLA-A3 binding assay

Fluorescence polarization (FP) HLA-A2.1 and HLA-A3 binding assays with UV-mediated peptide exchange were performed as described previously[19,20] with minor changes. For the HLA-A2.1 binding assay, 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. For the HLA-A3 binding assay, RIYRJ_1GATR was used as conditional ligand and KVPC^{TMR}ALINK was used as tracer peptide. Serial 2- or 3-fold dilutions of different (iso)peptides were used as competitor peptides. Assays were performed in 384-well low-volume black nonbinding surface assay plates (Corning 3820). Wells were loaded with 15 or 30μL total volume containing 0.5 μM HLA-A-peptide complex, 1 nM tracer peptide and serial dilutions of competitor peptide in PBS containing 0.5 mg/mL bovine γ-globulin. 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 4 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.[19] The binding affinity (IC_{50} value) of each competitor peptide was defined as the concentration that inhibited 50% of tracer peptide binding. Data were analyzed using GraphPad Prism software

(GraphPad).

(Iso)peptide stability assays

50 nmol YLDW$^{\alpha}$KLISV or YLDW$^{\varepsilon}$KLISV was incubated with 5 μg proteasome in ligation buffer containing 1-2% DMSO at 37 °C for different time periods up to 26 h. Incubation mixtures were snap-frozen, lyophilized and filtered over Strata™-X 33 μm polymeric reversed phase disposable columns (Phenomenex). Filtered samples were analyzed by LC-MS.

ACKNOWLEDGMENTS

The authors thank Olaf van Tellingen and Dorothe Linders for help with MALDI measurements, Henk Hilkmann for peptide synthesis, Jacques Neefjes for help with proteasome purification and Ton Schumacher for advice. This work was funded by the Dutch Cancer Society (grant NKI 2005-3368) and NWO/ALW.

REFERENCES

1. Adams, J. The development of proteasome inhibitors as anticancer drugs. *Cancer Cell* **5**, 417-421 (2004).
2. Nalepa, G., Rolfe, M. & Harper, J.W. Drug discovery in the ubiquitin-proteasome system. *Nat. Rev. Drug Discov.* **5**, 596-613 (2006).
3. 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).
4. 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).
5. Strehl, B. et al. Interferon-gamma, the functional plasticity of the ubiquitin-proteasome system, and MHC class I antigen processing. *Immunol. Rev.* **207**, 19-30 (2005).
6. 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).
7. Dahlmann, B., Ruppert, T., Kloetzel, P.M. & Kuehn, L. Subtypes of 20S proteasomes from skeletal muscle. *Biochimie* **83**, 295-299 (2001).
8. 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).
9. Vigneron, N. et al. An antigenic peptide produced by peptide splicing in the proteasome. *Science* **304**, 587-590 (2004).
10. Warren, E.H. et al. An antigen produced by splicing of noncontiguous peptides in the reverse order. *Science* **313**, 1444-1447 (2006).
11. 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).
12. 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).
13. Berkers, C.R. et al. Activity probe for in vivo profiling of the specificity of proteasome inhibitor bortezomib. *Nat. Methods* **2**, 357-362 (2005).
14. Marek, R. & Lycka, A. N-15 NMR spectroscopy in structural analysis. *Current Organic Chemistry* **6**, 35-66 (2002).
15. Iwahara, J., Jung, Y.S. & Clore, G.M. Heteronuclear NMR spectroscopy for lysine NH(3) groups in proteins: unique effect of water exchange on (15)N transverse relaxation. *J. Am. Chem. Soc.* **129**, 2971-2980 (2007).
16. Tjandra, N., Grzesiek, S. & Bax, A. Magnetic field dependence of nitrogen-proton J splittings in N-15-enriched human ubiquitin resulting from relaxation interference and residual dipolar coupling. *J. Am. Chem. Soc.* **118**, 6264-6272 (1996).
17. Falk, K., Rotzschke, O., Stevanovic, S., Jung, G. & Rammensee, H.G. Allele-specific motifs revealed by sequencing of self-peptides eluted from MHC molecules. *Nature* **351**, 290-296 (1991).
18. Solache, A. et al. Identification of three HLA-A*0201-restricted cytotoxic T cell epitopes in the cytomegalovirus protein pp65 that are conserved between eight strains of the virus. *J. Immunol.* **163**, 5512-5518 (1999).
19. 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).
20. Rodenko, B. et al. Class I major histocompatibility complexes loaded by a periodate trigger. *J. Am. Chem. Soc.* **131**, 12305-12313 (2009).
21. Toebes, M. et al. Design and use of conditional MHC class I ligands. *Nat. Med.* **12**, 246-251 (2006).
22. Budzik, J.M. et al. Intramolecular amide bonds stabilize pili on the surface of bacilli. *Proc. Natl. Acad. Sci. U S A* **106**, 19992-19997 (2009).
23. Kang, H.J., Coulibaly, F., Clow, F., Proft, T. & Baker, E.N. Stabilizing isopeptide bonds revealed in gram-positive bacterial pilus structure. *Science* **318**, 1625-1628 (2007).

24. Kang, H.J. & Baker, E.N. Intramolecular isopeptide bonds give thermodynamic and proteolytic stability to the major pilin protein of Streptococcus pyogenes. *J. Biol. Chem.* **284**, 20729-20737 (2009).
25. Shastri, N. Cell biology. Peptides, scrambled and stitched. *Science* **313**, 1398-1399 (2006).
26. Wiejak, S., Masiukiewicz, E. & Rzeszotarska, B. A large scale synthesis of mono- and di-urethane derivatives of lysine. *Chemical & Pharmaceutical Bulletin* **47**, 1489-1490 (1999).
27. Berkers, C.R. et al. Profiling proteasome activity in tissue with fluorescent probes. *Mol. Pharm.* **4**, 739-748 (2007).
28. Boyd, J., Soffe, N., John, B., Plant, D. & Hurd, R. The Generation of Phase-Sensitive 2D N-15-H-1 Spectra Using Gradient Pulses for Coherence-Transfer-Pathway Selection. *Journal of Magnetic Resonance* **98**, 660-664 (1992).
29. Davis, A.L., Keeler, J., Laue, E.D. & Moskau, D. Experiments for Recording Pure-Absorption Heteronuclear Correlation Spectra Using Pulsed Field Gradients. *Journal of Magnetic Resonance* **98**, 207-216 (1992).

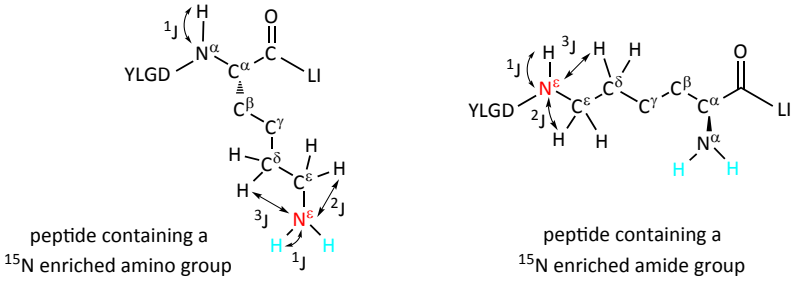
Supplementary information

SUPPLEMENTARY RESULTS

^{1}H-^{15}N HSQC and HMQC experiments

As direct detection of ^{15}N by NMR is difficult due to low naturally abundance of ^{15}N, low relative sensitivity and low relaxation times, the ^{15}N signal is frequently inverse-detected in ^{1}H-^{15}N HSQC (Heteronuclear Single Quantum Correlation) or HMQC (Heteronuclear Multiple Quantum Correlation) experiments. The correlation data provided by either HSQC or HMQC experiments are essentially equivalent and provide the same information. In inverse-detected experiments, ^{1}H nuclei that couple to ^{15}N nuclei are used for both the excitation and the detection of the ^{15}N signal. The ^{15}N signal is therefore detected via the ^{1}H signal. The resulting 2D spectrum contains one cross-speak for each proton that couples to a specific ^{15}N nucleus. Depending on the coupling constant J that is used in the experiment, ^{1}H-^{15}N coupling through one bond ($^{1}J_{H,N}$ ~ 90 Hz)[1] or through bonds (two or three bonds; $J_{H,N}$ = 0.3 to 16 Hz)[1] is utilized (Supplementary Figure 1). When coupling through one bond is utilized, the resulting 2D spectrum contains one cross-peak for each proton that is directly attached to a nitrogen. When coupling over multiple-bonds (^{2}J or ^{3}J) is utilized on the other hand, cross peaks are between protons and nitrogen atoms that can be either two or three bonds away.

In protein NMR, ^{1}H-^{15}N HSQC experiments are routinely performed to detect backbone amide resonances. Each residue in a protein, with the exception of proline, has an amide proton that is directly attached to a nitrogen and can therefore be used to inversely detect the ^{15}N signal. We performed ^{1}H-^{15}N HSQC experiments (J = 90 Hz) to detect backbone amide resonances in small peptides. In these experiments, both naturally abundant back-



Supplementary Figure 1 | Detection of amine and amide groups in HSQC experiments. Scheme showing ^{1}H-^{15}N couplings (J) that can be used to detect lysine's amide and amine groups in ^{1}H-^{15}N HSQC and HMQC experiments. Protons that exchange with the solvent are indicated in blue, other relevant protons are indicated in black, ^{15}N enriched nitrogen is indicated in red. Naturally abundant backbone amide groups (indicated in black) can be detected via the ^{1}J coupling (via the directly attached proton), while the amide group of a N$^{\varepsilon}$-linked lysine residue can be detected via ^{1}H-^{15}N coupling through one bond (^{1}J) or through multiple bonds (^{2}J or ^{3}J). As amino groups readily exchange protons with the solvent, the ^{15}N amine resonance cannot be inversely detected via the directly attached protons (^{1}J coupling). Lysine's free amino ^{15}N$^{\varepsilon}$ signal can be detected by utilizing ^{1}H-^{15}N coupling through multiple bonds.

bone amide resonances and ^{15}N enriched N^{ε}-linked lysine amide resonances can be detected (Supplementary Figure 1). By comparing the signal intensities of the naturally abundant backbone amide resonances to the intensity of the ^{15}N enriched N^{ε}-linked lysine signal, the relative amounts of different peptides in the sample can be determined.

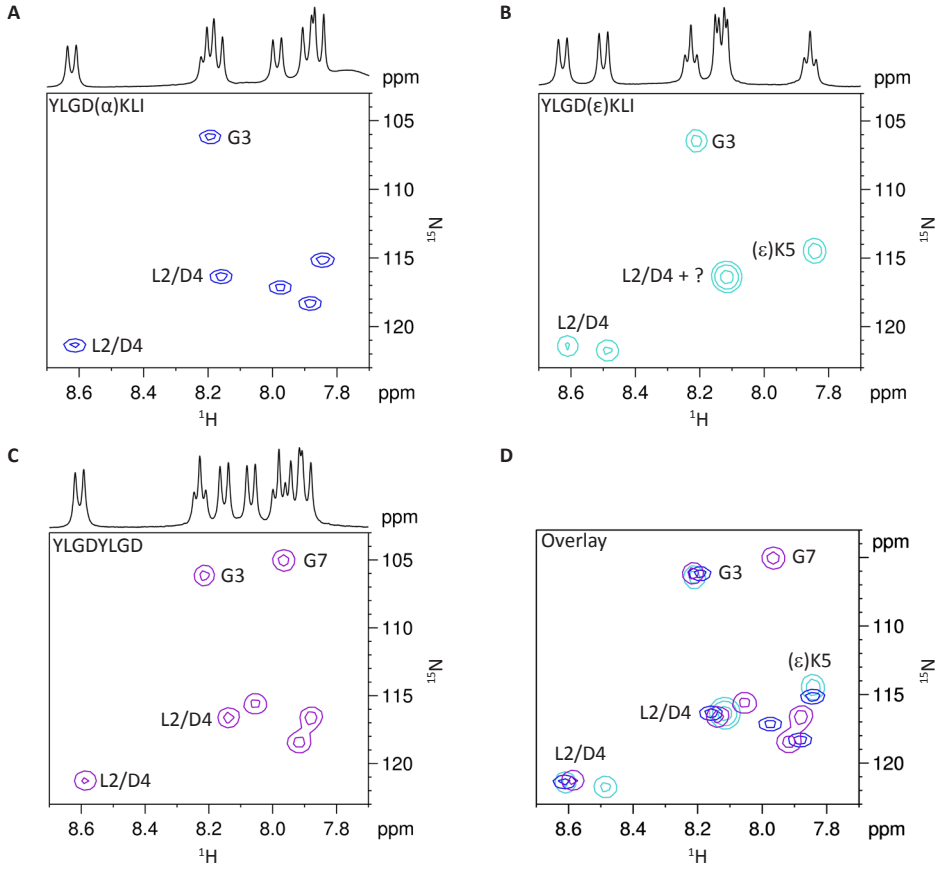
Whereas amide groups can be readily detected in an HSQC experiment via the $^{1}J_{H,N}$ coupling (J = 90 Hz), NMR characterization of lysine's free amino groups via the $^{1}J_{H,N}$ coupling is challenging. In aqueous solution, amino groups rapidly exchange protons with the solvent (Supplementary Figure 1), which hampers the detection of the amino protons in ^{1}H experiments and therefore the detection of amino $^{15}N^{\varepsilon}$ signals via these protons. Lysine's free amino $^{15}N^{\varepsilon}$ signal can be detected via protons that do not exchange with the solvent, e.g. the protons attached to lysine's C^{δ} or C^{ε}. Also the amide resonance of $^{15}N^{\varepsilon}$-linked lysine can be detected via these protons (Supplementary Figure 1). To visualize the ^{15}N enriched amide and amine resonances simultaneously, we therefore performed an HMQC experiment with a coupling constant of 10 Hz. In this experiment, the ^{15}N signal is detected via long range $^{2}J_{H,N}$ and $^{3}J_{H,N}$ couplings and the cross peaks are between the ^{15}N enriched lysine N^{ε} and the protons attached to the lysine C^{ε} and/or C^{δ}. Therefore, the resulting 2D spectrum of this experiment will show both the ^{15}N enriched amide and amine resonances. The amide and amine signal intensities measured in this experiment can however not be compared quantitatively, as explained in the main text.

Assignment of ^{1}H-^{15}N cross-peaks

To assign ^{1}H-^{15}N cross-peaks to one or more peptides, we synthesized YLGD$^{\alpha}$KLI, YLGD$^{\varepsilon}$KLI and YLGDYLGD and recorded ^{1}H and ^{1}H-^{15}N HSQC spectra as a reference (Supplementary Figure 2). In the HSQC reference spectra of both YLGD$^{\alpha}$KLI and YLGDYLGD all backbone amides are visible as separate cross-peaks, while in the spectrum of YLGD$^{\varepsilon}$KLI two backbone amide peaks overlap (Supplementary Figure 2). In all spectra several ^{1}H-^{15}N cross-correlations could be attributed to specific residues. Three cross-correlations were visible in all three spectra (8.2; 106, 8.1; 116, and 8.6; 121 ppm, Supplementary Figure 2D). Glycine ^{15}N signals resonate somewhat upfield (at 106 ppm) compared to other amino acids.[2] In addition, glycine amide protons are expected to show triplets in ^{1}H spectra due to the presence of two protons on C^{α}. The cross-peaks at 8.2; 106 ppm showed this upfield shift and characteristic splitting pattern (see ^{1}H spectra in Supplementary Figures 2B and 2C) and we therefore attributed this signal to G3, which is present in all peptides. As all peptides have YLGD as their first four residues, the other overlapping signals likely also originated from this part of the peptides (L2 and D4). The ^{1}H spectrum of YLGDYLGD (Supplementary Figure 2C) showed a second triplet at 8.0 ppm, and therefore the corresponding cross-peak in the ^{1}H-^{15}N HSQC spectrum (8.0; 105 ppm) was attributed to G7. The ^{1}H spectrum of YLGD$^{\varepsilon}$KLI (Supplementary Figure 2B) showed a second triplet at 7.86 ppm. As this splitting is expected for ε-linked K amide protons (due to the presence of two protons on the adjacent C^{δ}), the corresponding cross-peak in the ^{1}H-^{15}N HSQC spectrum (7.8; 114.6 ppm) was attributed to $^{\varepsilon}$K5.

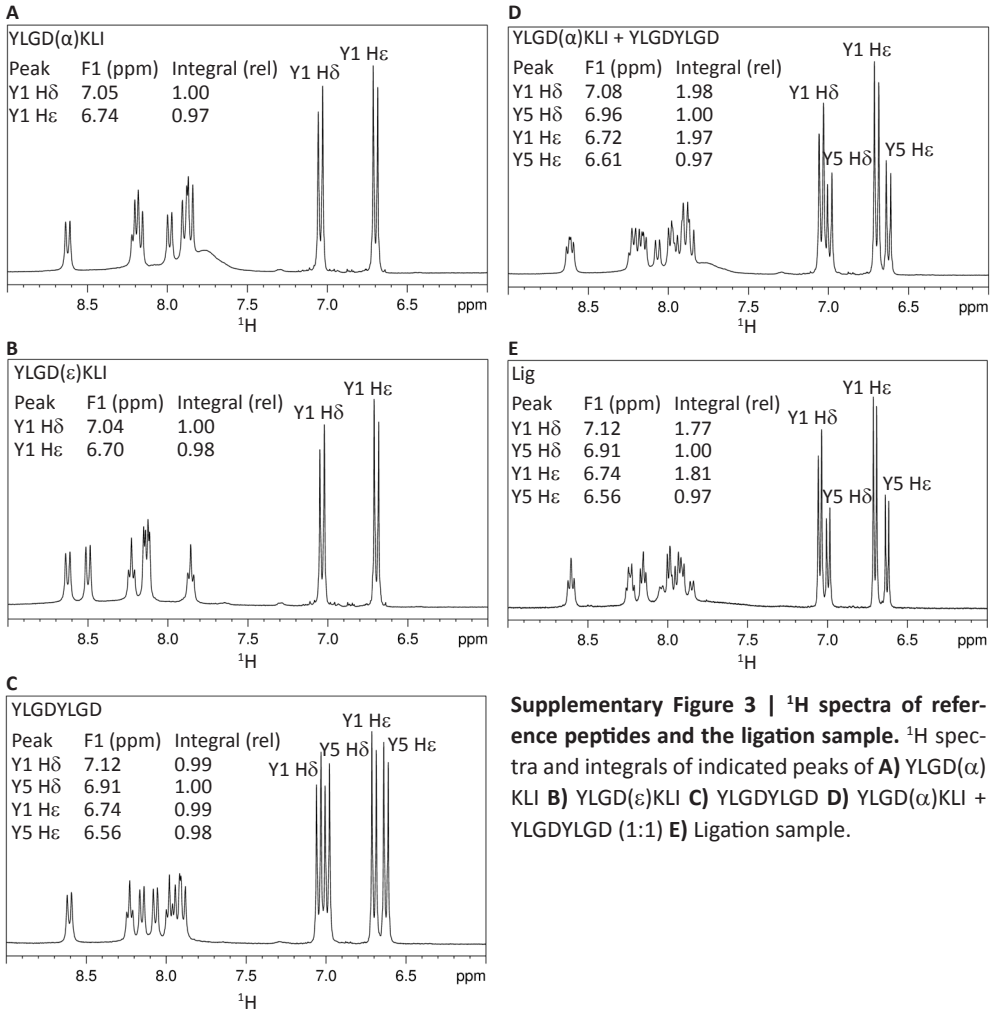
Estimation of the relative abundance of different ligation products in the ligation sample

As the most intense cross peaks due to naturally abundant amide resonances originate from all three peptides present in the ligation sample (Figure 2F), an estimation of the relative amounts of products in this sample is needed



Supplementary Figure 2 | HSQC spectra of reference peptides. A-C) ^{1}H-^{15}N HSQC spectra of YLGD(α)KLI **(A)** YLGD(ε)KLI (**B)** YLGDYLGD **(C)**. ^{1}H spectra of the corresponding peptides are shown on the top axis. **D)** Overlay of ^{1}H-^{15}N HSQC spectra of YLGD(α)KLI, YLGD(ε)KLI, and YLGDYLGD.

to estimate the efficiency of ε-ligation. To this end, the ^{1}H reference spectra of YLGD$^{\alpha}$KLI, YLGD$^{\varepsilon}$KLI and YLGDYLGD were compared to the ^{1}H spectrum of the ligation mixture (Supplementary Figure 3). All peptides contained one or two tyrosine residues, of which the Hδ and Hε protons are resonating at 7.0 and 6.6 ppm, respectively. Whereas YLGDKLI and YLGD$^{\varepsilon}$KLI show near identical resonances for Y1 Hδ and Hε (Supplementary Figures 3A and 3B), the ^{1}H spectrum of YLGDYLGD shows additional resonances for Y5, which are shifted somewhat downfield compared to Y1 and have equal intensities (Supplementary Figure 3C). The sum of the ^{1}H spectra of YLGDYLGD and YLGD$^{\alpha}$KLI (Supplementary Figure 3D) or YLGD$^{\varepsilon}$KLI (data not shown) indicates that the integrals of the Y1 signals are twice as high as the integrals of the Y5 signals if equal amounts of both peptides are used, as expected. In the ^{1}H spectrum of the ligation sample, the integrals of Y1 are only 1.8 times as intense as the integrals from Y5 (Supplementary Figure 3E). Therefore, we concluded that in the HSQC spectrum of the ligation sample, 56% of the signal of overlapping cross-peaks originated from YLGD-YLGD, while 44% originated from YLGD-$^{\alpha/\varepsilon}$KLI. In addition to the amount of pep-



Supplementary Figure 3 | ^{1}H spectra of reference peptides and the ligation sample. ^{1}H spectra and integrals of indicated peaks of **A)** YLGD(α)KLI **B)** YLGD(ε)KLI **C)** YLGDYLGD **D)** YLGD(α)KLI + YLGDYLGD (1:1) **E)** Ligation sample.

tide, the intensity of any ^{1}H-^{15}N cross-peak depends on the abundance of ^{15}N nuclei in these peptides. The naturally abundance of ^{15}N and thus of any ^{15}N amide is 0.37%, whereas the newly formed Lys N$^{\varepsilon}$-linkage is 98% ^{15}N enriched.

SUPPLEMENTARY REFERENCES

1. Marek, R. & Lycka, A. N-15 NMR spectroscopy in structural analysis. *Current Organic Chemistry* **6**, 35-66 (2002).
2. Wishart, D. S., Bigam, C. G., Holm, A., Hodges, R. S. & Sykes, B. D. H-1, C-13 and N-15 Random Coil Nmr Chemical-Shifts of the Common Amino-Acids .1. Investigations of Nearest-Neighbor Effects. *Journal of Biomolecular Nmr* **5**, 67-81 (1995).