# Driving Forces of Proteasome-catalyzed Peptide Splicing in Yeast and Humans*[S]

**Michele Mishto‡§‖, Andrean Goede‡, Kathrin Textoris Taube‡, Christin Keller‡, Katharina Janek‡, Petra Henklein‡, Agathe Niewienda‡, Alexander Kloss‡, Sabrina Gohlke‡, Burkhardt Dahlmann‡, Cordula Enenkel¶, and Peter Michael Kloetzel‡**

**Proteasome-catalyzed peptide splicing (PCPS) represents an additional activity of mammalian 20S proteasomes recently identified in connection with antigen presentation. We show here that PCPS is not restricted to mammalians but that it is also a feature of yeast 20S proteasomes catalyzed by all three active site β subunits. No major differences in splicing efficiency exist between human 20S standard- and immuno-proteasome or yeast 20S proteasome. Using $H_2{}^{18}O$ to monitor the splicing reaction we also demonstrate that PCPS occurs *via* direct transpeptidation that slightly favors the generation of peptides spliced in *cis* over peptides spliced in *trans*. Splicing efficiency itself is shown to be controlled by proteasomal cleavage site preference as well as by the sequence characteristics of the spliced peptides. By use of kinetic data and quantitative analyses of PCPS obtained by mass spectrometry we developed a structural model with two PCPS binding sites in the neighborhood of the active Thr1.** ***Molecular & Cellular Proteomics 11: 10.1074/mcp.M112.020164, 1008–1023, 2012.***

The 20S proteasome with its proteolytically active site β-subunits (β1, β2, and β5) is a N-terminal nucleophilic hydrolase, widely conserved during evolution from yeast to mammals. It is the central proteolytic machinery of the ubiquitin proteasome system and the catalytic core of the 26S proteasome that is built by the association of 19S regulator complexes with the 20S proteasome. As part of the 26S proteasome the 20S core degrades poly-ubiquitylated proteins to peptides of 3 to 20 residues in length (1). A small percentage of these peptides is transported to the endoplasmic reticulum, bound by major histocompatibility complex (MHC)[1] class I molecules, and presented at the cell surface to CD8+ cytotoxic T lymphocyte for immune recognition. This antigen presentation pathway is usually restricted to the proteasome-dependent processing of self- and viral-proteins (2). Antigen presentation is generally increased after IFN-γ stimuli because it induces, among others, the synthesis of alternative catalytic subunits (β1i, β2i, and β5i) and the concomitant formation of immunoproteasomes (i-proteasomes) (2).

All active β subunits carry an N-terminal threonine residue as reactive nucleophile. Therefore, their distinct cleavage preferences are determined by the structural features of the substrate binding pockets. In particular, the nonprimed substrate binding site of the active site β subunits binds the residues of the peptide substrate that are located at the N-terminal side of the cleaved residue. The residues of the peptide located C-terminally of the cleavage site are bound by the primed substrate binding site. The binding to both substrate binding sites of the active site β subunit provides the stability and the orientation of the substrate, which is mandatory to carry out the proteolytic cleavage (3).

Peptides can be produced by proteasomes during the degradation of proteins or polypeptides by conventional peptide bond hydrolysis or by proteasome-catalyzed peptide splicing (PCPS). The latter has been demonstrated *in vivo* so far only for four MHC class I-restricted epitopes (4–8), leading to the assumption that PCPS is most likely a rare event that lacks any wider functional importance (9). PCPS was suggested to occur in a direct transpeptidation reaction, in either *cis* or *trans,* by linking two proteasomal cleavage products (PCPs) derived either from the same or from two different substrate molecules, respectively (6, 10). Accordingly, during the cleavage of the peptide bond, the active site Thr of the proteolytic β-subunits transiently binds the C terminus of the N-terminal peptide fragment forming an acyl-enzyme intermediate. During normal proteolysis, this intermediate is rapidly hydrolyzed

From the ‡Institut für Biochemie, Charité - Universitätsmedizin Berlin, Oudenarder Straβe 16, 13347 Berlin, Germany; §Department of Experimental Pathology, University of Bologna, via S. Giacomo 12, 40126 Bologna, Italy; ¶Department of Biochemistry, University of Toronto, Medical Sciences Building MSB, 1 King's College Circle, Toronto, Ontario, M5S 1A8, Canada

Received April 30, 2012, and in revised form, July 9, 2012
Published, MCP Papers in Press, July 20, 2012, DOI 10.1074/mcp.M112.020164

[1] The abbreviations used are: MHC, major histocompatibility complex; *QME*, quantification with minimum effort; PCPs, proteasome-generated cleaved peptides; PCPS, proteasome–catalyzed peptide splicing; PSPs, proteasome-generated spliced peptides; Σ PCP/PSP, total proteasomal cleavage/splicing products; SCS, site-specific cleavage strength; SD, standard deviation.

and the fragment is liberated. In PCPS, the N terminus of another peptide fragment performs a nucleophilic attack on the acyl-enzyme intermediate, leading to "direct transpeptidation" and the final generation of the proteasome-generated spliced product (PSP). Although the transpeptidation model is widely accepted, direct experimental evidence for its validity during normal substrate processing is still missing. An implication of this model is that the reaction mechanism is not regulated by a particular sequence motif but can occur at any substrate cleavage site and that splicing does not occur in case the initial peptide cleavage is followed by hydrolysis and PCP release. Thus, PCPS would only depends on the proteasomal site-specific cleavage strength (SCS), which determines how frequently proteasomes cleaves specific peptide bond (5, 6). Interestingly, Dalet *et al.* identified a single PSP to be generated in absence of proteolysis, albeit with an extremely low efficiency, in a reaction that they defined as "condensation" (8) and that will be reported here as hydrolysis + transpeptidation to discriminate it from the direct transpeptidation reaction.

Because of the novelty of PCPS as part of our understanding of the ubiquitin proteasome system function and some inherent technical difficulties, the biochemical models as well as the comprehension of the relevance of PCPS were so far only partially investigated. Therefore, we carried out an *in vitro* study by investigating the mechanism of PCPS and its driving forces, thereby obtaining sufficient elements to design a novel model of the PCPS process with specific structural features. Furthermore, the quantification of PSPs, by an innovative method named *QME* (quantification with minimum effort), and the comparison of PCPS activity of different human and yeast proteasome iso-forms provided hints with regard to the physiological relevance of PCPS.

## EXPERIMENTAL PROCEDURES

*Peptides and Peptide Synthesis*—The sequence enumeration for the polypeptides $gp100_{40-52}$ (RTKAWNRQLYPEW), $gp100_{35-57}$ (VSRQLRTKAWNRQLYPEWTEAQR) and $gp100_{201-229}$ (AHSSSAFTITDQVPFSVSVSQLRALDGGNK) is referred to the human protein $gp100^{PMEL17}$, for the peptide $pp89_{16-40}$ (RLMYDMYPHFMPTNLGPSEKRVWMS) to the murine cytomegalovirus pp89 protein and for the peptide $LLO_{291-317}$ (AYISSVAYGRQVYLKLSTNSHSTKVKA) to the murine Listeria monocytogenes's Listeriolysin O protein. Peptide sequences of the 14 previously described PSPs (10) as well as the 25 new PSPs identified in the proteasomal processing of the four synthetic substrates are reported in supplemental Table S1. All peptides were synthesized using Fmoc solid phase chemistry as previously described (11). The purity of synthetic peptides was tested by amino acid analysis.

*Cell Cultures*—Lymphoblastoid cell lines (LcLs) are human B lymphocytes immortalized with Epstein-Barr virus (EBV), which mainly express i-proteasomes (12). T2 cell line is a human T cell leukemia/B cell line hybrid defective in TAP1/TAP2 (transporter associated with antigen presentation) and β1i/β5i subunits. T2.27 is a cell line originating from T2 cells and transfected with murine β1i and β5i subunits (13).

All yeast strains (supplemental Table S2) are isogenic and derivatives of WCGa (*MATa leu2–3,112 ura3 his3–11,15 Can$^S$ GAL2*), which were kindly provided by Wolfgang Heinemeyer (LMU München).

*20S Proteasome Purification*—The 20S proteasomes were purified from LcLs, T2, and T2.27 cells as previously described (14). The 20S proteasome purified from human spleen and erythrocytes (purchased from BioMol, Plymouth Meeting, PA) were in glycerol 50%. Yeast 20S proteasomes (supplemental Table S2) were purified by IgG Sepharose affinity protocol followed by a gel-filtration on Superose 6 column (PC 3.2/30, Amersham Biosciences, Uppsala, Sweden) (15).

*In Vitro Digestion of Synthetic Peptide Substrates*—Synthetic peptides at different concentration (30–40 μM) were digested by 1–3 μg 20S proteasomes in 100 μl TEAD buffer (Tris 20 mM, EDTA 1 mM, $NaN_3$ 1 mM, dithiothreitol 1 mM, pH 7.2) over time at 37 °C. For the experiments performed in $H_2{}^{18}O$-TEAD buffer we used water with 97% $^{18}O$ (Campro Scientific GmbH, Germany). To minimize undesired side reactions like the acidic-catalyzed $^{18}O$ labeling of carboxyl groups (*i.e.* at the C terminus or at acidic amino acids) (16), we performed the analyses of the samples by nano-liquid chromatography-matrix-assisted laser desorption ionization/time of flight/TOF-MS (LC-MALDI-TOF/TOF-MS) immediately after stopping the reaction by TFA acidification (0.3% final concentration). The relative quantification of the ratio direct transpeptidation/(hydrolysis + transpeptidation) has been based on the isotopic pattern of the PSPs [RTK]-[QLYPEW] ($gp100_{40-42/47-52}$) and [VSRQL] [VSRQL] ($gp100_{35-39/35-39}$) from the digestions, in $H_2{}^{18}O$-TEAD buffer and by LcL and yeast wild type 20S proteasomes, of the polypeptides $gp100_{40-52}$ and $gp100_{35-57}$, respectively. The isotope pattern of these two PSPs generated in $H_2{}^{16}O$-TEAD buffer and of the PCPs [QLYPEW] ($gp100_{47-52}$) and [RTKAWNR] ($gp100_{40-46}$) were used as reference controls. The congruence of the isotope patterns of the PSPs and PCPs with the theoretical isotope patterns evaluated according to the their elemental composition (17) was computed as reported in supplemental material. In summary, the congruence of the isotope patterns of PSPs generated in $H_2{}^{18}O$-TEAD buffer with the theoretical isotopic patterns represents the prevalence of direct transpeptidation whereas the congruence of the isotope patterns of PSPs generated in $H_2{}^{16}O$-TEAD buffer and of the PCPs with the theoretical isotopic patterns was used, on the contrary, to estimate the accuracy of our measurements.

All experiments reported in this study were repeated and measured at least twice.

*LC-ESI MS and Nano-LC-MALDI-TOF/TOF-MS*—LC-MS analyses were performed as previously described (10) with the electrospray ionization (ESI)-ion trap instrument DECA XP MAX (ThermoFisher Scientific) and the MALDI-TOF/TOF mass spectrometer 4700 Proteomics Analyzer (Applied Biosystems, Framingham, MA).

Analysis of ESI/MS/MS data was accomplished using Bioworks version 3.3 (ThermoFisher Scientific, USA). Database searching was performed using the SpliceMet's ProteaJ database version 1.0 released in 2010 (10) and the following parameters: no enzyme, mass tolerance for precursor ions 0.5 Da and for fragment ions 1amu. Oxidations of methionine and tryptophan were considered and ruled out as artificial. We rejected the following masses for the MS/MS analysis: 370.9, 371.9, 372.9, 391.1, 392.1, 393.1. These masses belong to plasticizer material derived from the MS instrument. In time-dependent processing experiments (signal intensity *versus* time of digestion) we analyzed the kinetics of the identified peaks by using LCquan software version 2.5 (Thermo Fisher).

Analysis of MALDI-TOF/TOF-MS data was accomplished by the peaklist-generating software 4000 Series Explorer version 3.6 (Applied Biosystem) and by using MASCOT version 2.1 (Matrixscience, London, UK). Database search was performed using SpliceMet's ProteaJ database (10) and the following parameters: no enzyme, mass tolerance for precursors, ± 80 ppm and for MS/MS fragment ions, ± 0.3 Da.

The number of entries in the searched database varied between different substrates because of their different sequence lengths. In particular, for the polypeptides $gp100_{40-52}$, $gp100_{35-57}$, $gp100_{201-229}$, $pp89_{16-40}$, and $LLO_{291-317}$ the number of database entries were 5810, 57982, 173355, 84255, and 112339, respectively.

MALDI-TOF/TOF-MS/MS spectra, ESI-MS/MS spectra and extracted ion chromatograms of the identified PSPs are reported in supplemental Figs. S1–S4.

*QME, Titration and Raw MS Methods*—In order to estimate the absolute amount of the total proteasomal cleavage/splicing products (Σ PCP/PSP) within the proteasomal digestion of the substrates we developed *QME* and compared it with the *titration* and the *raw MS* methods in the representative proteasomal digestion of the substrate $gp100_{40-52}$. *QME* estimates the absolute content of ∑ PCP/PSP based on their ESI-MS signal measured in the digestion probe. *QME* is based on the law of the mass conservation and MS instrument features. The parameters and parameters' values of the *QME* algorithm were empirically computed (supplemental Fig. S5–S9). From the quantitation of ∑ PCP/PSP we could compute the site-specific cleavage strength (SCS) by applying the SCS algorithm, which computes the frequency of proteasome cleavage after any given residue of the synthetic polypeptide substrate by analyzing the amount of any digestion product (18, 19).

*Estimation of the MHC Class I-restricted Potential Epitopes*—The list of the 9–12mer ∑ PCP/PSP, detected in the processing of all four synthetic substrates by 20S proteasomes, were screened by two MHC class I epitope prediction algorithms, *i.e.* SYFPHEITY (20) and IEDB (21), available on the Web. We adopted as threshold to identify the best candidates the score of 20 for SYFPHEITY and IC50 = 500 nM for IEDB. In the Results and Discussion sections we discussed mainly the results obtained adopting the IEDB prediction because of its increasing database, prediction power and the recently reported superior performances (22, 23).

*Statistics*—Statistical analyses of *cis/trans* PCPS (Table II*B*) and the relative amount of the direct transpeptidation (Table III) were performed using the *t*-Student test for independent tests adjusted using Bonferroni correction. $p < 0.05$ was considered statistically significant. In each data set, homogeneity of variance was checked by Levene's test. All analyses were implemented using SPSS software. The means and S.D. reported in Table II*A* and Table IV*B* represent the means, for each 20S proteasome, obtained from the sum of the four substrates degradation and the S.D. over time. This type of statistical analysis is supposed to better mimic the *in vivo* situation where proteasomes are processing different substrates at the same time producing a unique pool of peptides. The maximum and minimum frequency values of PSPs, 9–12mer PSPs and potential MHC class I PSP epitopes reported in the text refer to the time course means computed for each proteasome type and substrates.

A complete description of the methods can be found in supplemental material.

## RESULTS

To determine proteasomal cleavage and splicing preferences, and to investigate the quantitative relevance of PCPS and its underlying biochemical mechanisms it was mandatory to compute the absolute amount of reactant peptides that are available for peptide splicing, *i.e.* the proteasome-generated cleaved peptides (PCPs) and the amount of proteasome-generated spliced peptides (PSPs) produced during the PCPS reaction.

Therefore, we compared three different methods for quantifying the absolute amount of all proteasomal cleavage/splicing products (Σ PCP/PSP) generated over time by *in vitro* digestion of the human melanoma-derived synthetic 13mer polypeptide $gp100_{40-52}$ (RTKAWNRQLYPEW) by standard (s-proteasomes) and i-proteasomes. The *raw MS* method, which because of its immediacy has been used in the past (*e.g.* by Cardinaud *et al.* (24)), assumes that the MS signal of each peptide directly corresponds to its amount, thereby setting the conversion factor between MS signal and the absolute amount for any peptide equal to that of the substrate $gp100_{40-52}$. The *titration* method computes the conversion factor between the MS signal and the peptide's absolute amount by titrating the synthetic peptides corresponding to all products of digestion. By applying this method to all PCPs and PSPs identified in the digestion of $gp100_{40-52}$ we observed that the conversion factors differed from peptide to peptide, *de facto* invalidating the assumption of the *raw MS* method (supplemental Fig. S5). The *QME* method, an algorithm-based method developed by us, computes the conversion factor between MS signal and the peptide's absolute amount by combining the MS signal, MS instrument features and biochemical principles such as the mass conservation's law (see Experimental Procedures and supplemental material). Of the three methods tested, *QME* provided the best mass conservation in the over time reaction, whereas with the *titration* method a notable gain of mass was observed (Fig. 1*A*). In general, the *QME* and the *titration* methods resulted in a similar estimation of the amount of the generated Σ PCP and Σ PSP (Figs. 1*B*, 1*C* and supplemental Fig. S6). By focusing on the production (per nmol of substrate cleaved) of the two major PCPs, *i.e.* $gp100_{40-46}$ (RTKAWNR) and $gp100_{47-52}$ (QLYPEW) (Figs. 1*D*, 1*E*) we noted that applying the *titration* method, the values exceeded the limit of 1 nmol (that is the theoretical maximum obtained when the substrate is cleaved always in the same position thus generating only two PCPs), indicating that this method could in some cases overestimate the real Σ PCP/PSP amount. Application of *raw MS* on the other hand resulted in an underestimation of the amount of specific PCPs and PSPs, a phenomenon that was even more pronounced for shorter peptides (supplemental Fig. S6). We therefore considered *QME* as the method most suited and we applied *QME* to all subsequent analyses to determine the Σ PCP/PSP amount and the site-specific cleavage strength (SCS).

*PCPS is Catalyzed by All Active Sites of 20S Proteasomes*—So far PCPS was described only for human proteasomes. Therefore, we initially asked whether proteasomes of the yeast *Saccharomyces cerevesiae* were also able to catalyze peptide splicing or whether this was a function peculiar of human proteasomes designed to broaden the peptide repertoire for MHC class I antigen presentation. Yeast proteasomes were also chosen because of the availability of well-characterized mutant strains lacking one or more active sites and thus they could provide information regarding the involvement of the different active sites in PCPS. To allow a more

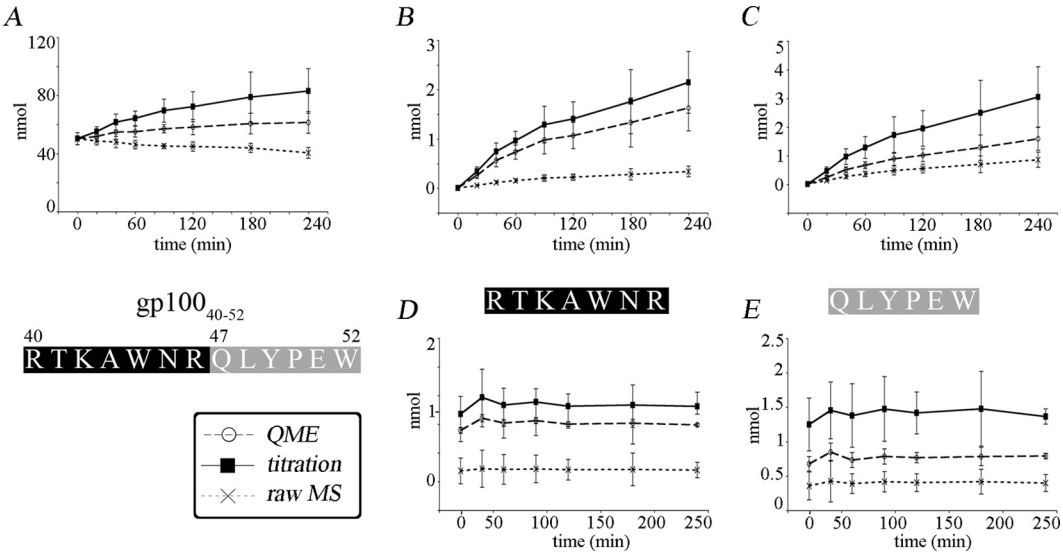



FIG. 1. **Comparison on the quantitative kinetics of Σ PCP/PSP estimated by *QME, titration* and *raw MS* methods.** Representative examples of the substrate $gp100_{40-52}$ digest kinetics and Σ PCP/PSP generation, computed by applying *QME, titration* and *raw MS* methods, are here reported. *A*, Total amino acid amount (substrate + Σ PCP/PSP) over time. *B*, *C*, Kinetics of the PCP $gp100_{40-46}$ (RTKAWNR) (*B*) and $gp100_{47-52}$ (QLYPEW) (*C*). *D*, *E*, nmol of $gp100_{40-46}$ (*D*) and of $gp100_{47-52}$ (*E*) produced per nmol of substrate cleaved. Theoretically this value should be lower than 1 nmol. Data reported in charts derive from the digestion by 3 μg human erythrocyte 20S proteasome of the synthetic $gp100_{40-52}$ (40 μM) in 100 μl solution. Similar results were obtained with human spleen 20S proteasome (data not shown). The data are means of 2 experiments measured three times by DECA MAXI XP in TFA solution. Bars represent S.D.

general conclusion we needed to collect data on a sizeable number of PSPs generated during the degradation of different synthetic polypeptides. We therefore searched 20S proteasomal degradation products for PSPs generated from the Lysteriolysin delineated peptide substrate $LLO_{291-317}$, the human melanoma gp100 delineated $gp100_{35-57}$ and $gp100_{201-230}$ polypeptides and the MCMV IE pp89 delineated $pp89_{16-40}$ polypeptide by applying the previously developed PSP search method SpliceMet (10). Using this approach we identified 39 PSPs, which were generated by both human and yeast 20S proteasomes thereby demonstrating that PCPS is not a peculiar function of human 20S proteasomes only (supplemental Table S1 and supplemental Fig. S1–S4).

To identify the catalytic sites responsible for the splicing reaction we analyzed the Σ PCP/PSP generated from the four synthetic substrates by 20S proteasomes purified from yeast wild type and mutant strains, which harbor β subunit mutants with inactive (β1 T1A, β2 T1A, β1 and β2 T1A) or affected (β5 K33A) cleavage sites (supplemental Table S2). In the β1 and β2 T1A mutants the active Thr1 is replaced by Ala thereby rendering the corresponding β subunits proteolytically inactive. The T1A replacement in the β5 subunit is lethal (25). Because Lys33 is required for autocatalytic β5 propeptide processing and for the subunit's peptide cleavage activity, Lys33 within the active site pocket is replaced by Ala. In consequence, K33A mutation abolishes or impairs β5 subunit maturation leading to the formation of a proteolytic intermediate with inactive β5 subunit (25, 26).

All types of yeast 20S proteasomes (wild type and the four active site mutants) produced either all or at least 50% of the PSP generated by human 20S proteasomes, providing first evidence that all three different catalytic sites can carry out PCPS. For more detailed investigation we next calculated the specific cleavage strength (SCS) for each polypeptide substrate. By comparing the SCS of the different yeast 20S proteasome mutants we determined which of the active sites were mainly responsible for cleavage after a given residue. We considered an active site mainly responsible for a specific cleavage when the β-subunit mutant proteasome showed a significant decrease in the frequency of cleavage after a given residue as compared with the wild type (see Experimental Procedures; Fig. 2 and supplemental Fig. S10). This information was compared with the amino acid sequences of all generated PSPs, thereby allowing the identification of β subunits that were responsible for those cleavages that generate the N- and C-terminal residues of the PCPs, *i.e.* the splice-reactants, used in the formation of a new spliced peptide. We named them as following: Pn = N-terminal residue of the N-terminal splice-reactant; P1 = C-terminal residue of the N-terminal splice-reactant; P1′ = N-terminal residue of the C-terminal splice-reactant; Pc- = C-terminal residue of the C-terminal splice-reactant (Fig. 2*A*). Importantly, the P1 residues suggested to form the acyl-enzyme intermediate with the active site threonine, were found to be generated by the β1, β2 and β5 subunits (Table I) establishing that the PCPS reaction can be carried out by all three active site β subunits.

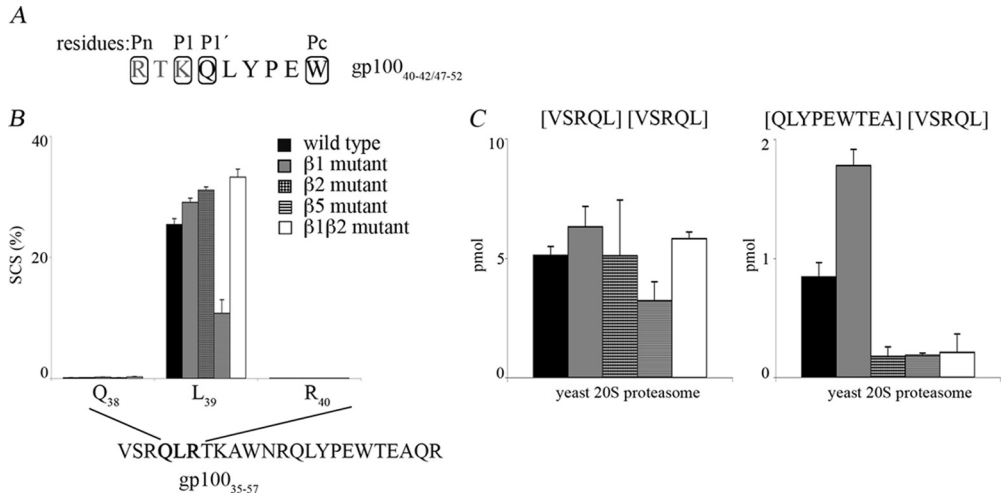



FIG. 2. **Cleavage and splicing in a representative specific substrate sequence by yeast active site β subunits.** *A*, Nomenclature of the N- and C-terminal residues of the splice-reactants. We named them as following: Pn = N-terminal residue of the C-terminal splice-reactant; P1 = C-terminal residue of the N-terminal splice-reactant; P1′ = N-terminal residue of the C-terminal splice-reactant; Pc = C-terminal residue of the C-terminal splice-reactant. *B*, Representative SCS of the amino acid sequence $QLR_{38-40}$ within the substrate $gp100_{35-57}$ digested by 20S proteasomes purified from yeast wild type and β1-, β2-, β5-, and β1β2-mutant strains. In the chart, only the SCS of $QLR_{38-40}$ is shown. The complete $gp100_{35-57}$ SCS is shown in supplemental Fig. S10. To note is the significant decrease of SCS at residue $L_{39}$ by the β5-mutant, indicating that β5 subunit is mainly responsible of the peptide bond hydrolysis between $L_{39}$ and $R_{40}$. *C*, Average content (pmol) of the PSPs [VSRQL][VSRQL] ($gp100_{35-39/35-39}$) and [QLYPEWTEA][VSRQL] ($gp100_{47-55/35-39}$) per nmol of substrate degraded after 0.5–2 h of digestion. This time frame (0.5–2 h of digestion; 4 time points), when 30–50% of the substrate was still present, was chosen to minimize the re-entry of Σ PCP/PSP into the catalytic chamber of the 20S proteasome. *B*, *C*, Bars are the S.D. of repeated measurements of a representative cleavage reaction of 100 μl in which 3 nmol $gp100_{35-57}$ were incubated with 1 μg of 20S proteasome for 0–4 h at 37 °C. The digestion products were detected by LC-ESI/MS and their absolute amount computed by *QME*.

TABLE I

*Proteolytic β-subunits responsible for the generation of the N- and C-termini of the splice-reactants*

The number of cleavage sites, ascribed to different active site β subunits, that produced the N- and C-termini of the splice-reactants of the four substrate polypeptides is here reported. We named them as following: Pn = N-terminal residue of the N-terminal splice-reactant; P1 = C-terminal residue of the C-terminal splice-reactant; P1′ = N-terminal residue of the C-terminal splice-reactant; Pc- = C-terminal residue of the C-terminal splice-reactant (Fig. 2a). The cleavages that generated the terminal residues, of the splice-reactants, which were afterwards located at the PSP position Pn, P1, P1′, and Pc could be ascribed to all proteolytically active β subunit. We defined "unknown" those cleavages that, based on the yeast 20S proteasome assays, were produced by all proteolytic β subunits to a similar extend (supplemental Fig. S10).

| | Position within the PSPs | | | | |
|---|---|---|---|---|---|
| Catalytic subunit | Pn | P1 | P1′ | Pc | Sum |
| β1 | 8 | 5 | 7 | 11 | 31 |
| β2 | 7 | 19 | 8 | 31 | 65 |
| β5 | 4 | 15 | 9 | 2 | 30 |
| Unknown | 4 | 9 | 2 | 12 | 27 |

A quantitative comparison of the SCS and the amount of correlating PSPs provided also first insights regarding the driving forces of PCPS. For example, the cleavage after the residue $Leu_{39}$ of the polypeptide $gp100_{35-57}$ was impaired in the 20S proteasome β5 mutant, suggesting that this subunit was mainly responsible of this cleavage. This is further confirmed by the enhanced site-specific proteolysis of the 20S proteasome β1-β2 mutant possessing only β5 subunits as active sites (Fig. 2*B*). Intriguingly, the PSP [VSRQL][VSRQL] ($gp100_{35-39/35-39}$), which had the $Leu_{39}$ as PSP residue P1, was generated with a lower efficiency by the β5 mutant proteasome (Fig. 2*C*), indicating a correlation between the amount of each produced PSP and the availability of associated splice-reactant peptides. The inactivation of the active site β subunits in no instance completely abolished the usage of the identified substrate cleavage sites. Therefore, it was essential to quantify and compare the amount of each PSP produced by different yeast mutant proteasomes because a simple qualitative (yes/no) evaluation was not possible.

*PCPS is Not a Rare Event*—The data reported above gave first insight into the driving forces of PCPS but also suggest that PCPS *in vitro*, at least for specific PSPs, was not a rare event. For an overall evaluation of the relative number of PSPs within Σ PCP/PSP, we analyzed the degradation of the substrates $gp100_{35-57}$, $gp100_{201-230}$, $pp89_{16-40}$ and $LLO_{291-317}$ by 20S proteasomes purified from human erythrocytes and T2 cell lines (s-proteasomes) as well as from human spleen, T2.27 and LcL cell lines (mainly i-proteasomes). Overall, the average Σ PSP amount produced by all 20S proteasomes turned out to be 1.89% of Σ PCP/PSP with notable differences between substrates (Σ PSP relative amount = 0.55 - 5.49%). Although the amount of specific PSPs largely varied between s- and i-proteasomes as shown in Fig. 3–Fig. 4, no

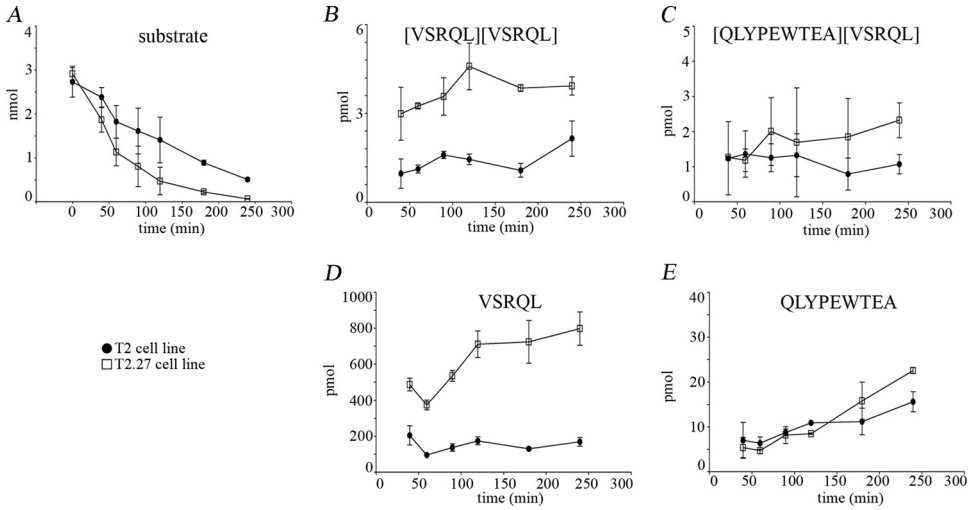



FIG. 3. **Correlation between the amounts of PCPs and PSPs generated from gp100$_{35–57}$ by 20S s- and i-proteasomes.** The PCPS activity of 20S proteasomes purified from T2 (s-proteasomes) and T2.27 (mainly i-proteasomes) cell lines was monitored by the kinetics of the generation of PSPs and of the correlated reactant peptides during the digestion of the substrate gp100$_{35–57}$. 3 nmol gp100$_{35–57}$ were cleaved in 100 $\mu$l reactions by 1 $\mu$g of 20S proteasome for 0–4 h at 37 °C. The digestion products were detected by LC-ESI/MS and their absolute amount computed by *QME*. *A*, T2.27 20S proteasomes revealed a higher gp100$_{35–57}$ degradation rate than T2 20S proteasomes. *B*, T2.27 20S proteasomes exhibited higher efficiency for the PSP [VSRQL][VSRQL] (gp100$_{35–39/35–39}$) production than T2 20S proteasomes. *C*, There exists no significant difference in the production of the PSP [QLYPEWTEA][VSRQL] (gp100$_{47–55/35–39}$) between T2 and T2.27 20S proteasomes. *D*, The PCP [VSRQL], shared by the two PSPs shown in *A* and *B*, was produced more efficiently by T2.27 20S proteasomes. *E*, The T2 and T2.27 proteasomes generate the PCP gp100$_{47–55}$ [QLYPEWTEA] with similar efficiency. In *B–E* the pmol of peptides produced per nmol of degraded substrate are reported. Bars represent the S.D. of independent experiments.

significant differences in the relative amount of Σ PSPs was observed when we compared s- and i-proteasomes (Table II*A*).

Because of the high number of *trans* PSPs identified so far (supplemental Table S1) we asked whether *trans* PCPS was more frequent process than *cis* PCPS. To quantify the frequency of *trans* PCPS reaction we performed *in vitro* digests in which the unmodified 13mer gp100$_{40–52}$ peptide was applied to proteasomal processing in the presence of the same peptide but with the heavy amino acid residues $^{13}C_6$-Lys, $^{13}C_6$-$^{15}$N-Leu, and $^{13}C_5$-$^{15}$N-Glu (RTK$^{+6}$AWNRQL$^{+7}$-YPE$^{+6}$W). We detected PSP variants being the results of *cis* (variants -A and -D) or of *trans* (variants -B and -C) PCPS and we computed their relative amount by comparing the MALDI-MS signals (Fig. 5 and supplemental Fig. S11). Interestingly, in experiments carried out either by s- and i-proteasomes *cis* PSPs prevailed over *trans* PSPs with a small but statistically significant difference (Table II*B*).

*Human and Yeast 20S Proteasomes Catalyze Peptide Splicing with Similar Rates and with no Active Site Preference*—To quantitatively analyze yeast 20S proteasome-catalyzed splicing reaction and to verify whether any of the $\beta$ subunits was prevalently involved in it, we measured the Σ PSP amount produced by yeast wild type and mutant 20S proteasomes by processing the substrate gp100$_{35–57}$. The relative amount of Σ PSP within the Σ PCP/PSP generated by yeast wild type proteasome, *i.e.*1.05% (± 0.38), was in the range of what had been measured for human s- and i-proteasomes, which was 1.83% (± 0.16) and 1.15% (± 0.42), respectively. Considering the quantitative relevance of active site $\beta$ subunits for PCPS by comparing the Σ PSP within the Σ PCP/PSP generated by yeast wild type or mutant proteasomes, no significant differences emerged (Table II*C*), although the amount of specific PSPs was affected (Fig. 2*C*).

*Sequence Preferences Regulate PSP Generation*—We next asked whether PCPS was driven by the amount of the splice-reactants, which is a corollary of the transpeptidation model or whether other factors were also involved. Therefore, we compared the amount of PSP pairs that shared one of the splice-reactants and the amount of the corresponding splice-reactants. This approach was mandatory, because by comparing only the amount of a single PSP with the amount of the corresponding splice-reactants we could not discriminate between possible sequence dependences of cleavage site usage and of the PCPS reaction. Among the Σ PCP/PSP generated by proteasomal digestion of the substrates gp100$_{35–57}$, gp100$_{201–230}$, and LLO$_{291–317}$ we identified several PSP pairs sharing one of the splice-reactants. To reduce the complexity of the comparison we focused on PSP pairs where one PSP was generated by the ligation of two identical PCPs, *e.g.* [VSRQL][VSRQL], whereas the other PSP was produced by ligation of the shared splice-reactant and a second different splice-reactant, *e.g.* [QLYPEWTEA][VSRQL]. Furthermore, we computed for each example reported in Fig. 3–Fig. 4 the amount of PCPs and PSPs per nmol of substrate cleaved to obtain data independent to the substrate degradation rate, which often differs between s- and i-proteasomes. For instance, for the 23mer

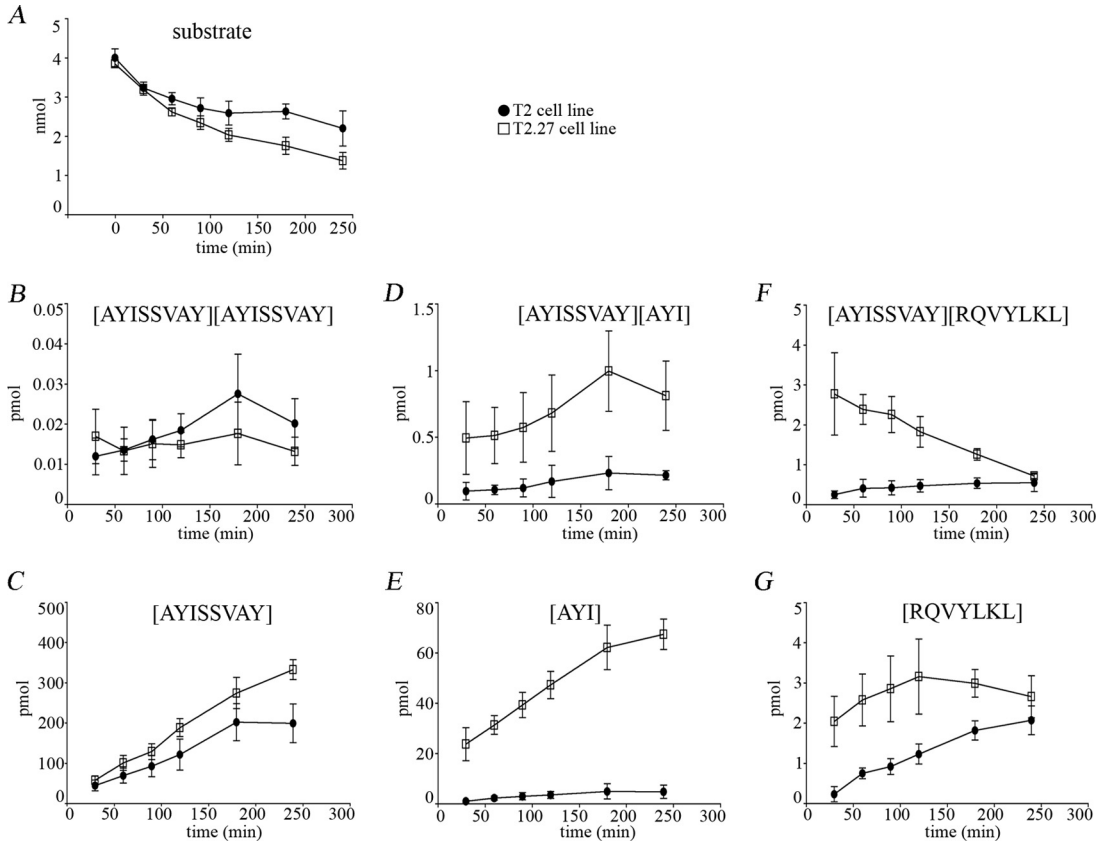



FIG. 4. **Correlation between the amounts of PCPs and PSPs generated from $LLO_{291–317}$ by 20S s- and i-proteasomes.** The PCPS activity of 20S proteasomes purified from T2 and T2.27 cell lines was monitored by the kinetics of PSPs and of the correlated splice-reactants during the digestion of the substrate $LLO_{291–317}$. 4 nmol $LLO_{291–317}$ were cleaved in 100 $\mu$l reactions by 2 $\mu$g of 20S proteasomes for 0–4 h at 37 °C. The digestion products were detected by LC-ESI/MS and their absolute amount computed by *QME*. *A*, T2.27 20S proteasomes revealed a slightly higher $LLO_{291–317}$ degradation rate than T2 20S proteasomes. *B*, T2 and T2.27 20S proteasomes exhibit a similar efficiency for the PSP [AYISSVAY][AYISSVAY] ($LLO_{291–298/291–298}$) production. *C*, The production of the PCP [AYISSVAY] ($LLO_{291–298}$) is only slightly higher when carried out by T2.27 proteasomes compared with the T2 proteasomes. *D*, T2.27 proteasomes generate more efficiently the PSP [AYISSVAY][AYI] ($LLO_{291–298/291–293}$) as well as the PCP [AYI] ($LLO_{291–293}$) than T2 proteasomes (*E*). *F*, T2.27 proteasomes produce more efficiently than T2 proteasomes the PSP [AYISSVAY][RQVYLKL] ($LLO_{291–298/300–306}$) as well as the PCP [RQVYLKL] ($LLO_{300–306}$) (*G*). In *B*-*G* the pmol of peptides produced per nmol of degraded substrate are reported. Bars represent the S.D. of repeated experiments.

$gp100_{35–57}$, we observed a considerably faster degradation by T2.27 i-proteasomes than by T2 s-proteasomes (Fig. 3*A*). Both proteasome isoforms produced the PSP [VSRQL][VSRQL] ($gp100_{35–39/35–39}$) (Fig. 3*B*) and [QLYPEWTEA][VSRQL] ($gp100_{47–55/35–39}$) (Fig. 3*C*). The shared reactant peptide [VSRQL] was generated more efficiently by T2.27 i-proteasomes than by T2 s-proteasome (Fig. 3*D*), evidence that directly correlated with increased efficiency of PSP $gp100_{35–39/35–39}$ formation (Fig. 3*B*). In contrast, T2 and T2.27 20S proteasomes generated the peptide [QLYPEWTEA] ($gp100_{47–55}$) with similar efficiencies and there was no difference in the generation of the PSP $gp100_{47–55/35–39}$ suggesting that, in this case, the reactant peptide which was driving the reaction was $gp100_{47–55}$ (Fig. 3*E*). The PCP $gp100_{47–55}$ was also produced at lower amounts than the other reactant peptide [VSRQL] by both the 20S proteasome isoforms, suggesting that the reactant peptide produced at lower amount was the rate-limiting factor of the splicing reaction.

Similar conclusions emerged from the analysis of the $LLO_{291–317}$ polypeptide substrate, which was degraded only slightly faster by T2.27 20S proteasomes (Fig. 4*A*). The amount of the PSP [AYISSVAY][AYISSVAY] ($LLO_{291–298/291–298}$) generated per nmol of substrate cleaved was similar for the two 20S proteasome isoforms and T2.27 proteasomes produced only a slightly higher amount of the PCP [AYISSVAY] ($LLO_{291–298}$) (Figs. 4*B*–4*C*). In contrast, the ligation of the PCPs [AYISSVAY] and [AYI], leading to the generation of the PSP [AYISSVAY][AYI] ($LLO_{291–298/291–293}$), was remarkably higher in the digestion by T2.27 proteasomes and correlated with the efficiency of the PCP [AYI] ($LLO_{291–293}$) generation (Figs. 4*D*–4*E*). Similarly, T2.27 proteasomes produced more efficiently than T2 proteasomes the PSP [AYISSVAY][RQVYLKL] ($LLO_{291–298/300–306}$) as well as the correlated PCP [RQVYLKL] ($LLO_{300–306}$) (Figs. 4*F*–4*G*). Also in these latter examples there appeared to be a correlation between the amount of the PSPs and their splice-reactants, with the PCPs present in lower amount, *i.e.*

TABLE II

*Amount of PSPs generated by in vitro proteasome processing*

(*A*) The relative amount of Σ PSPs among Σ PCP/PSP generated during the degradation of the peptides $gp100_{35-57}$, $gp100_{201-230}$, $pp89_{16-40}$, and $LLO_{291-317}$ by 20S c- (from T2 cell line and human erythrocyte) and i-proteasomes (from human spleen, T2.27 and LcL cell lines). Data were computed by applying *QME* on the MS analysis of the proteasomal degradations of the synthetic peptide substrates. Data are expressed as mean Σ PSPs / Σ PCP/PSP +/− S.D. between different time points of the reaction by summing up the data of all substrates. (*B*) The relative amount (%) of the *cis* and *trans* variants of the PSP $gp100_{40-42/47-52}$ produced by different 20S proteasomes during the processing of the 13mer $gp100_{40-52}$ and of its heavy $^{+19}$ analog is reported as well as the S.D. of independent experiments measured by MALDI-TOF-MS. Relative amount of *cis* PSPs significantly prevailed on *trans* PSPs both considering all s-proteasomes (from T2 cell line and human erythrocytes; *t*-student $p = 0.004$) and i-proteasomes (from T2.27 and LcL cell lines as well as human spleen; *t*-student $p = 0.007$). For quantification procedures see supplemental material. (*C*) The relative amount of Σ PSPs amongst Σ PCP/PSP generated during the degradation of the peptides $gp100_{35-57}$ by 20S proteasomes purified from yeast wild type or β1, β2, β5, and β1β2 mutants is reported. Data were computed by applying *QME* on the MS analysis of the proteasomal degradations of the synthetic peptide substrate. Data are expressed as mean Σ PSPs / Σ PCP/PSP +/− S.D. between different time points of the reaction.

A

| 20S proteasomes from: | Σ PSPs/Σ PCP/PSP (%) |
|---|---|
| T2 cell line | 2.27+/−0.35 |
| Human erythrocytes | 2.19+/−0.21 |
| T2.27 cell line | 1.50+/−0.26 |
| LcL | 1.86+/−0.48 |
| Human spleen | 1.27+/−0.48 |
| s-proteasomes (mean) | 2.23+/−0.25 |
| i-proteasomes (mean) | 1.54+/−0.34 |

B

| 20S proteasomes from: | *cis* variants (A + D) | *trans* variants (B + C) |
|---|---|---|
| T2 cell line | 51+/−8 | 47+/−3 |
| Human erythrocytes | 53+/−8 | 47+/−1 |
| T2.27 cell line | 50+/−4 | 50+/−1 |
| LcL | 53+/−7 | 47+/−3 |
| Human spleen | 55+/−6 | 45+/−2 |
| s-proteasomes (mean) | 52+/−2 | 48+/−2 |
| i-proteasomes (mean) | 53+/−4 | 47+/−4 |

C

| Yeast 20S proteasomes from: | Σ PSPs/Σ PCP/PSP (%) |
|---|---|
| Wild type | 1.05+/−0.38 |
| β1 mutant | 1.11+/−0.40 |
| β2 mutant | 1.48+/−0.77 |
| β5 mutant | 1.31+/−0.55 |
| β1-β2 mutant | 1.18+/−0.53 |

$LLO_{291-293}$ and $LLO_{300-306}$, being the rate-limiting compounds. Notably, this dependence of the PCPS reaction on the splice-reactant present in lower amounts seemed to be independent of the position of the splice-reactant in the nascent PSP, because we observed this phenomenon for both N- ($gp100_{47-55}$) and C- ($LLO_{291-293}$ and $LLO_{300-306}$) terminal splice-reactants.

Although we found a correlation between the amount of the reactant peptides (*i.e.* the PCPs that will be spliced) and the products (*i.e.* the PSPs), we asked whether that was the only driving factor of the PCPS or whether also the sequence of the splice-reactants would affect the PCPS efficiency. Therefore, we computed the SCS for each substrate and the frequency of the N- and C-terminal residues of the splice-reactants. We focused our attention on PSP P1 residues, because they are thought to be directly linked to the active site involved in the splicing reaction, and on PSP P1′ residues (Fig. 2*A*), which are supposed to perform the nucleophilic attack on the acyl-enzyme intermediate (5). In case PCPS was driven only by the amount of splice-reactants, the general frequency of cleavage (*i.e.* SCS) and the frequency of cleavage generating the PSP P1 and P1′ residues would be expected to be similar. Quite in contrast, for all substrates and in all the digestions carried out by 20S proteasome isoforms of any origin, we observed a substantial difference between the SCS and the frequency of cleavage generating the PSP P1 and P1′ residues (Fig. 6). In fact, both the PSP P1 and P1′ residues often derived from minor cleavage sites, thereby suggesting also a sequence-dependence of the PCPS process that is independent of the overall cleavage preferences.

*Cis and Trans PCPS Occur by Direct Transpeptidation*—Unexpectedly, the results shown above were partially in contradiction with the transpeptidation model as proposed by Vigneron *et al.*, which had been deduced from the observation that the 20S proteasome did not catalyze the ligation of the peptides RTK and QLYPEW (5). Therefore we set out to experimentally study the transpeptidation model by performing the digestion of the substrates $gp100_{40-52}$ (RTKAWNRQLYPEW) and $gp100_{35-57}$ (VSRQLRTKAWNRQLYPEWTEAQR) by LcL and wild-type yeast 20S proteasome in $H_2{}^{18}O$-TEAD buffer.

In the direct transpeptidation reaction, which is the core of the transpeptidation model (5), the formed acyl-enzyme intermediate, bound to the proteasomal Thr1, must be attacked by the C-terminal splice-reactant thereby preventing the N-terminal fragment from being released by hydrolysis (Fig. 7*A*). In opposite, in the hydrolysis + transpeptidation reaction, the N-terminal splice-reactant is released from the active site β subunit Thr1 residue by hydrolysis and subsequently forms a new acyl-enzyme intermediate followed by its ligation to the other splice-reactant and the formation of the final PSP. Therefore, the direct transpeptidation and hydrolysis + transpeptidation reactions can be discriminated by performing the digestions in $H_2{}^{18}O$-TEAD buffer, because only during the hydrolysis, peculiar of the hydrolysis + transpeptidation, the N-terminal peptide incorporates $^{18}O$. During the subsequent formation of the new acyl-enzyme intermediate, however, both $^{16}O$ and $^{18}O$ can be released with a ratio 1:1 because of the mesomery at the C terminus of the bound

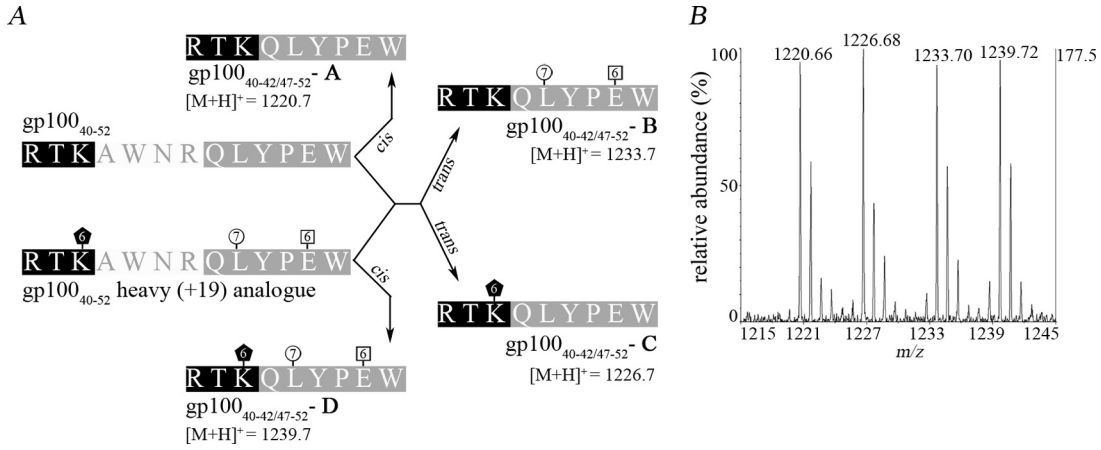



FIG. 5. **Relative quantification of *cis* and *trans* PSP gp100$_{40-42/47-52}$ variants.** *A*, To estimate the relative *in vitro* efficiency of *cis* and *trans* PCPS, the peptide gp100$_{40-52}$ and its heavy analog with amino acids $^{13}C_6$-Lys, $^{13}C_{6-15}$N-Leu, and $^{13}C_{5-15}$N-Glu (RT**K**$^{+6}$AWNRQ**L**$^{+7}$YP**E**$^{+6}$W) were digested for 24 h by different 20S proteasomes and the gp100$_{40-42/47-52}$ variants generated by *cis* (-A, -D) and *trans* (-B, -C) PCPS were investigated. *B*, All four peptides were detected by LC-MALDI-TOF-MS. As representative the mass spectrum of the gp100$_{40-42/47-52}$ variants produced by T2.27 20S proteasomes are reported (for MS/MS spectra see supplemental Fig. S11). Upper right, the MS intensity is noted.

fragment. Consequently, the PSP molecules produced at the last step of this reaction will possess to the 50% the monoisotopic mass and to 50% the monoisotopic mass + 2Da because of the $^{18}$O-labeling. Other condensation processes (not represented in Fig. 7) may occur, involving other proteasome sites instead of the active site $\beta$ subunit Thr1, but producing the same final result.

This experimental set up has the advantage that proteasome can cleave the substrate along its sequence during *in vitro* digestion and direct transpeptidation as well as hydrolysis + transpeptidation can occur within the same sample, thereby allowing a relative quantification of the two reactions.

We performed the experiments in $H_2{}^{16}O$- or $H_2{}^{18}O$-TEAD buffer with human (LcL) or yeast (wild type) 20S proteasomes and the substrates gp100$_{40-52}$ and gp100$_{35-57}$, we analyzed the isotopic pattern of the PSPs [RTK][QLYPEW] (gp100$_{40-42/47-52}$) or [VSRQL][VSRQL] (gp100$_{35-39/35-39}$), which are examples for *cis* and *trans* PCPS, respectively. We analyzed also the isotopic pattern of two PCPs, *i.e.* [QLYPEW] (gp100$_{47-52}$) and [RTKAWNR] (gp100$_{40-46}$), as additional controls of the reactions in $H_2{}^{18}O$-TEAD buffer and we calculated the congruence between the isotopic pattern of both PCPs and PSPs and the theoretical isotopic patterns (17). The congruence of the isotope patterns of PSPs generated in $H_2{}^{18}O$-TEAD buffer with the theoretical isotopic patterns represents the prevalence of direct transpeptidation. The congruence of the isotope patterns of PSPs generated in $H_2{}^{16}O$-TEAD buffer and of the PCPs with the theoretical isotopic patterns can, in addition to estimate the accuracy of our measurements (see supplemental material).

The PCP [QLYPEW] (gp100$_{47-52}$), which has the C terminus that is not produced by cleavage, revealed a similar isotopic pattern in the digestions carried out by LcL proteasome in $H_2{}^{16}O$- or $H_2{}^{18}O$-TEAD buffer (Fig. 7*B*). In the same digestions, the PSP gp100$_{40-42/47-52}$ showed a similar isotopic pattern, too (Fig. 7*C*). This PSP has a calculated *m/z* 1220.6 and because the C terminus is not produced by cleavage, the increased amount of the isotopes with *m/z* of 1222.6, 1223.6, 1224.6, and 1225.6 Da would be due to the incorporation of one $^{18}$O after the K residue during hydrolysis + transpeptidation. In the same digestions, the PCP [RTKAWNR] (gp100$_{40-46}$) showed a clear shift of all isotopes of +2 Da (Fig. 7*D*). This mass shift was due to the incorporation of one $^{18}$O at the C terminus during the hydrolysis as confirmed by MS/MS analysis (data not shown). A similar and expected shift of +2 Da of all isotopes in the digestion by LcL 20S proteasome of the substrate gp100$_{35-57}$ was detected also for the PSP gp100$_{35-39/35-39}$ (Fig. 7*E*) because of the incorporation of one $^{18}$O at the C terminus during the hydrolysis as confirmed by MS/MS analysis (supplemental Fig. S12). As shown in Table III, the congruence of the measured isotope patterns with the theoretical isotope patterns of the *trans* PSP was not significantly different in the digestions carried out by LcL proteasome in $H_2{}^{16}O$- or $H_2{}^{18}O$-TEAD buffer and also not different to the control PCP gp100$_{47-52}$, which had a congruence of the measured isotope patterns with the theoretical isotope patterns of 102.7 ± 8.0% and 96.9 ± 8.4% in $H_2{}^{16}O$- or $H_2{}^{18}O$-TEAD buffer digestions, respectively. A similar congruence of the measured isotope patterns with the theoretical isotope patterns was obtained for the *cis* PSP gp100$_{40-42/47-52}$ (Table III) leading to the conclusion that these two PSPs are both produced by direct transpeptidation because no significant incorporation of $^{18}$O at the splice sites where detected. Similar results were obtained in digestions performed by yeast wt 20S proteasome (data not shown).

*PSPs are More Often Potential MHC Class I Epitopes than PCPs*—Because of the possible implication of PCPS in MHC

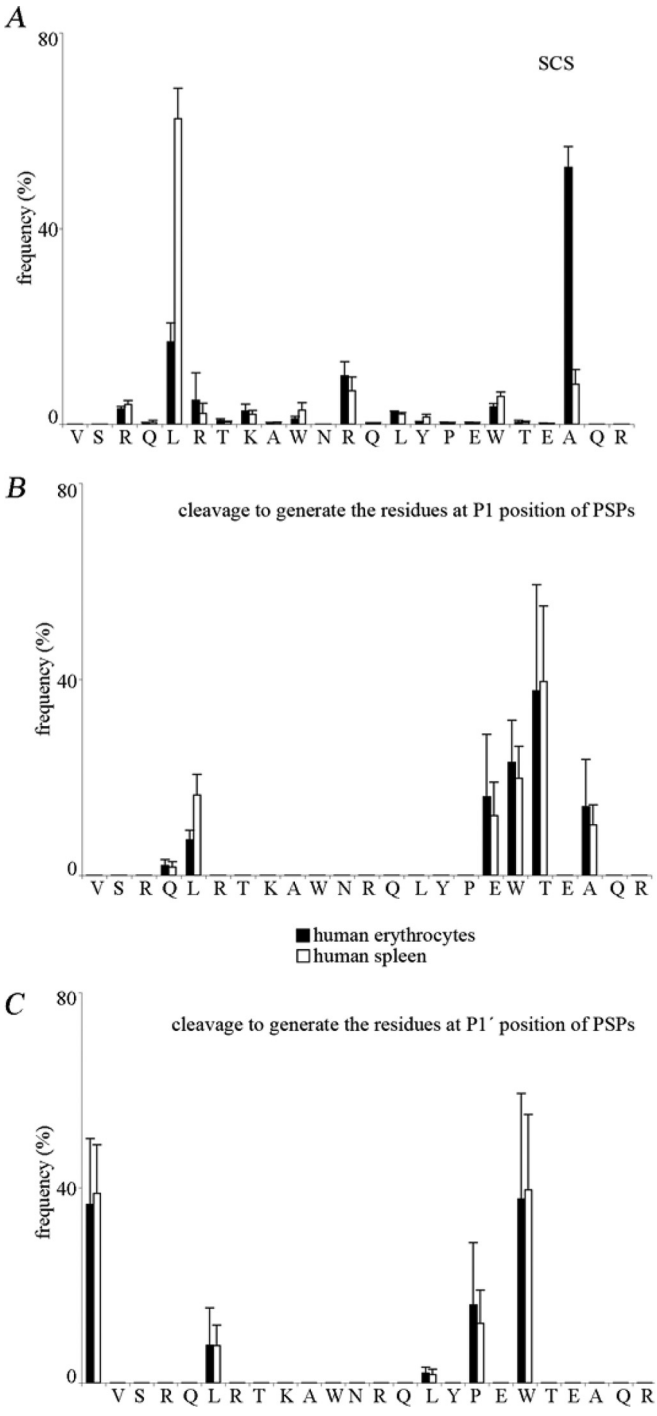



FIG. 6. **The frequencies of site-specific cleavage (SCS) and of cleavages generating the PSP P1 and P1′ positions remarkably differ.** The SCS frequency (*A*) and frequency of cleavages generating the residues at the P1 (*B*) and P1′ (*C*) positions (considering all PSPs of the representative substrate gp100$_{35-57}$) are reported and substantially differ. For example, in the digestions by spleen 20S proteasomes the relative amount of PSPs with $T_{53}$ at P1′ position is 39.6% (± 15.4). This information is here referred as the frequency of cleavage after the residue $W_{52}$, which generates the $T_{53}$ at P1′ position, considering in the computation only the PSPs and not the Σ PCP/PSP. Conversely, the frequency of the cleavage (*i.e.* SCS) after the residue $W_{52}$ is 5.7% (± 0.9) if we consider Σ PCP/PSP. The symbol "/" in the *x* axis refers to all splice-reactants whose N termini have been spliced without previous cleavage, like, for example, [VSRQL] or [VSRQLRT]. This event is, of course, possible only for the P1′ positions. Digestions of 3 nmol of synthetic substrate in 100 μl reactions were carried out by 1.5 μg 20S proteasome purified from human erythrocytes or spleen. The frequencies were computed by SCS algorithm from the *QME* calculation of the ESI-MS data. Relative frequencies are reported in % and the bars represents S.D. of three experiments measured three times each. Erythrocyte and spleen 20S proteasome data are reported in black or white histograms, respectively.

class I antigen presentation, we also investigated whether PSPs might have a different prevalence than PCPs with respect to potential MHC class I epitopes. We therefore computed the frequency and absolute amount of Σ PCP/PSP with a length between 9 and 12 amino acids (roughly the size of standard MHC class I epitopes and precursors) within the Σ PCP/PSP produced by proteasomal processing of the substrates gp100$_{35-57}$, gp100$_{201-230}$, pp89$_{16-40}$ and LLO$_{291-317}$. Surprisingly, the frequency of 9–12mers was higher among PSPs than PCPs in both s- and i-proteasome reactions. Indeed, 7.56% of the 9–12mers generated by all 20S proteasomes were PSPs, with a considerable difference between different substrates, *i.e.* 0.62% (gp100$_{201-230}$ cleaved by T2.27 proteasomes) and 29.69% (LLO$_{291-317}$ cleaved by T2 proteasomes) (Table III*A*). Notably, PCPs and PSPs produced during the degradation of the four substrates exhibited a similar average length although the frequency of 9–12mer was higher among PSPs because of the short length of the splice-reactants (Table IV*B*). PSPs also harbored a relatively higher percentage of potential MHC class I-restricted epitopes. Considering the epitope list selected by the prediction algorithm IEDB (21), we calculated that PSPs amounted to 11.61% of the potential MHC class I epitopes generated by all 20S proteasomes, corresponding to 15.58 pmol per nmol of substrate cleaved. Again, we observed a strong variation between substrates, with the frequency of Σ PSP among potential MHC class I epitopes varying between 0.25% (gp100$_{201-230}$ cleaved by T2.27 proteasomes) and 47.42% (LLO$_{291-317}$ cleaved by T2 proteasomes). Among the potential MHC class I epitopes, specific PSPs were better produced by proteasomes iso-forms. For example, the PSP [VSRQL][VSRQL], which was predicted to be precursor of a binder of the HLA-B*2705 by both MHC class I prediction programs, was more efficiently produced by i- than s-proteasomes (Fig. 3*B*). Nevertheless, although s- and i-proteasomes differed significantly in their production of specific PSPs that were predicted to be MHC class I binders, no significant difference emerged when we considered the Σ PSP's amount among potential MHC class I epitopes (Table IV*A*).

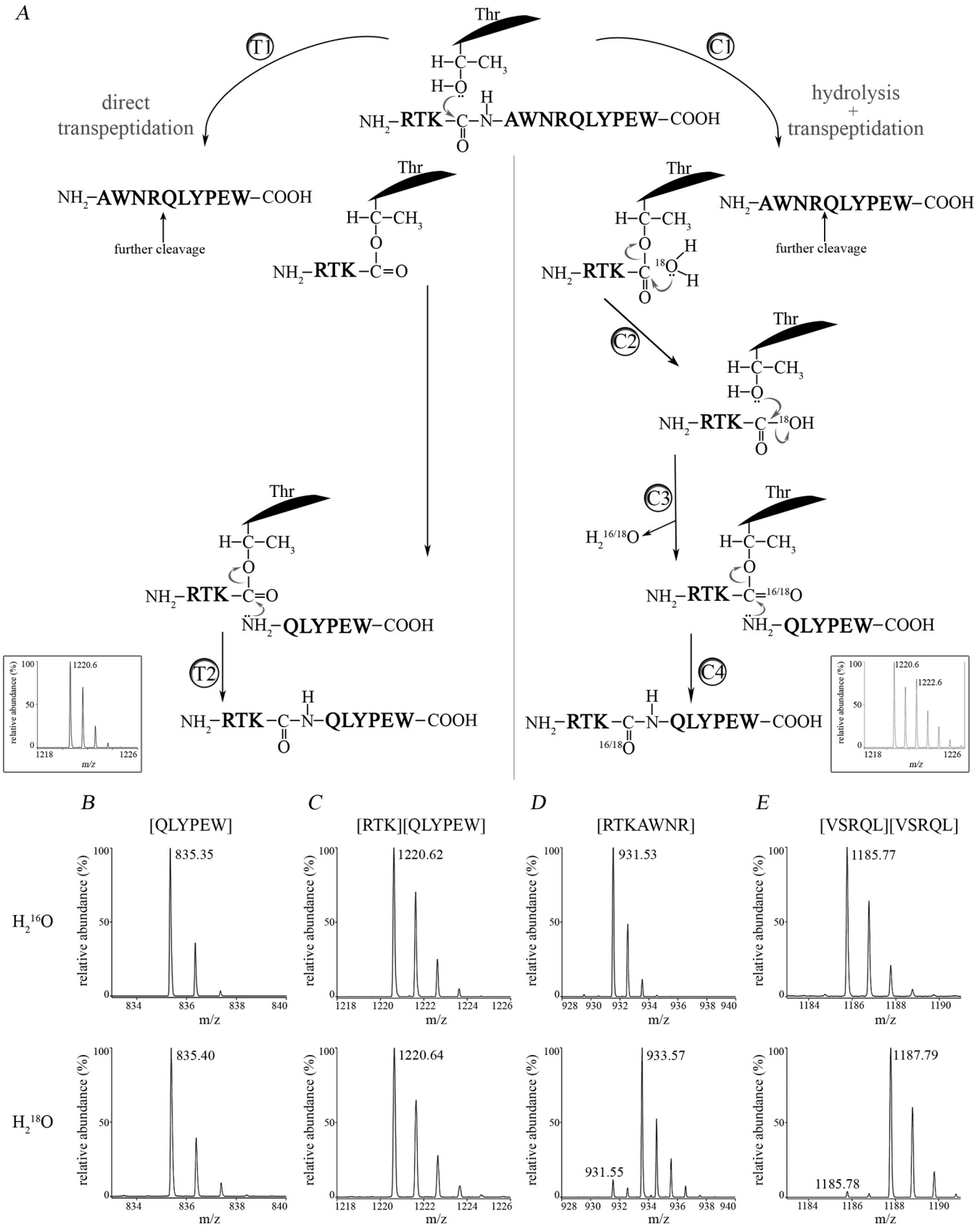


TABLE III
*cis* and *trans* PSPs occur by direct transpeptidation

The congruence of the isotope patterns of the PSPs $gp100_{40-42/47-52}$ and $gp100_{35-39/35-39}$ with the theoretical isotope patterns for each type of experiment is reported as percentage. The congruence of the isotope patterns represents for PSPs generated in $H_2{}^{18}O$-TEAD buffer the prevalence of direct transpeptidation, whereas in $H_2{}^{16}O$-TEAD buffer the accuracy of measurements. Indeed, in the experiments performed in $H_2{}^{16}O$-TEAD buffer, the congruence of the measured isotope pattern with the theoretically estimated distribution of the isotopic peaks (17) of the PSPs is supposed to be 100% and thereby the deviation from this value is ought to be due to technical issues. In opposite, in the experiments performed in $H_2{}^{18}O$-TEAD buffer the congruence of the measured isotope patterns of the PSPs with the theoretical isotope patterns can be due also to the incorporation of $^{18}O$ at the splice-site, whether PCPS occurs *via* hydrolysis + transpeptidation, and therefore it represents also the prevalence of direct transpeptidation. Congruence with the theoretical isotope patterns was computed as described in the supplemental material by elaborating the area of the isotopic peaks of PSPs $gp100_{40-42/47-52}$ $gp100_{35-39/35-39}$. The congruence's values here reported derive from two independent digestions carried out in $H_2{}^{16}O$- or $H_2{}^{18}O$-TEAD buffer and measured twice by LC-MALDI-TOF-MS. Corresponding S.D. is reported. Student *t*-test, adjusted using Bonferroni correction, revealed no significant differences of the congruence with the theoretical isotope pattern of the PSPs $gp100_{40-42/47-52}$ generated in $H_2{}^{16}O$- vs $H_2{}^{18}O$-TEAD buffer ($p = 0.064$), $gp100_{40-42/47-52}$ vs $gp100_{47-52}$ generated in $H_2{}^{18}O$-TEAD buffer ($p = 0.076$), $gp100_{35-39/35-39}$ generated in $H_2{}^{16}O$- vs $H_2{}^{18}O$-TEAD buffer ($p = 0.543$) and $gp100_{35-39/35-39}$ vs $gp100_{47-52}$ generated in $H_2{}^{18}O$-TEAD buffer ($p = 0.203$).

| PSP | PSP type | Buffer | Congruence with theoretical isotope pattern (%) |
|---|---|---|---|
| $gp100_{40-42/47-52}$ | *cis* | $H_2{}^{16}O$TEAD | 92.7+/−6.3 |
| $gp100_{40-42/47-52}$ | *cis* | $H_2{}^{18}O$TEAD | 86.1+/−2.0 |
| $gp100_{35-39/35-39}$ | *trans* | $H_2{}^{16}O$TEAD | 90.5+/−1.8 |
| $gp100_{35-39/35-39}$ | *trans* | $H_2{}^{18}O$TEAD | 91.0+/−8.3 |

DISCUSSION

In the present study, we identified a large number of new PSPs, which were generated by 20S proteasomes from four different polypeptide substrates. The large number of new PSPs and the number of different polypeptide substrates in combination with *QME,* a new algorithm-based method facilitating the computation of the absolute amount of all proteasomal processing products, enabled us to define the proteasomal active sites involved in PCPS, to identify mechanisms and driving factors of the reaction as well as to estimate the potential physiological relevance of PCPS.

It is worth mentioning that PCPS does not simply represent a reverse reaction of substrate proteolysis because it combines noncontiguous fragments to generate novel peptides (Fig. 8). The 20S proteasome and its isoforms could therefore differently catalyze reactions, which differ in the equilibrium between proteolysis and PCPS, thereby resulting in different total amounts of PSP (Σ PSP). In consequence, the comparison of the Σ PSP generated from a larger number of substrate polypeptides by different 20S proteasome isoforms, as performed here, allowed further insight into the PCPS mechanism and into the structural features of proteasome that support PCPS.

For example, our experiments demonstrate that also 20S proteasomes of the yeast *Saccharomyces cerevesiae* catalyze peptide splicing, suggesting that PCPS represents an evolutionary conserved intrinsic property of proteasomes, rather than a specific activity of mammalian 20S proteasomes developed to extend the diversity of MHC class I epitopes. Taking advantage of the different yeast 20S proteasome active site mutants our experiments also revealed that, without any obvious preference for a given active site, all proteolytic β subunits are able to catalyze the PCPS. Because of the struc-

FIG. 7. **Cis and *trans* PCPS occur through the formation of an acyl-enzyme intermediate without hydrolysis.** To verify if both *cis* and *trans* PCPS occur by forming an acyl-enzyme intermediate, which is not further hydrolyzed but is directly bound to the N terminus of another peptide, we performed proteasomal digestion experiments in buffer containing $H_2{}^{18}O$ or $H_2{}^{16}O$; as substrate we used the polypeptides $gp100_{40-52}$ (RTKAWNRQLYPEW) or $gp100_{35-57}$ (VSRQLRTKAWNRQLYPEWTEAQR), whose digestion produced the PSP [RTK][QLYPEW] ($gp100_{40-42/47-52}$) or [VSRQL][VSRQL] ($gp100_{35-39/35-39}$), respectively. *A*, Theoretical considerations. Using the example of PSP [RTK]-[QLYPEW] derived from the polypeptide $gp100_{40-52}$ two cases during a proteasomal digestion in $H_2{}^{18}O$-buffer are described: direct transpeptidation and hydrolysis + transpeptidation. *Left side (direct transpeptidation):* Step T1. Formation of the acyl-enzyme intermediate between the C terminus of the fragment [RTK] and the catalytic Thr1 of proteasome. Step T2. The N terminus of the fragment [QLYPEW], previously cleaved from the original substrate, forms an amide bond with the fragment [RTK], followed by the release of the PSP [RTK][Q-LYPEW] with a monoisotopic *m/z* 1220.7. Because no hydrolysis occurred during the reaction, no $^{18}O$ was incorporated. A theoretical MS spectrum of the PSP is reported. *Right side (hydrolysis + transpeptidation):* Step C1. Formation of the acyl-enzyme intermediate between the C terminus of the fragment [RTK] and the catalytic Thr1 of proteasome. Step C2. Release of the fragment [RTK] by hydrolysis of $H_2{}^{18}O$. This fragment is labeled by $^{18}O$. Step C3. The fragment [RTK] forms a new acyl-enzyme intermediate with the proteasomal Thr1. Because of the mesomery of the carboxyl group $H_2{}^{16}O$ or $H_2{}^{18}O$ is released. Step C4. The terminal amino group of the fragment [QLYPEW] forms an amide bond, followed by the release of the PSP [RTK][QLYPEW] with a Lys-$^{18}O$-labeling of 50%. The $^{18}O$-labeled PSP have a monoisotopic *m/z* 1222.7. Other condensation processes (not here described) may occur, involving other proteasome sites instead of the active site β subunit Thr1, though producing the same final result. A theoretical MS spectrum of the PSP is reported. Depicted are the LC-MALDI-TOF mass spectra of PCP [QLYPEW] ($gp100_{47-52}$; *m/z* 835.4) (*B*), of PSP [RTK][QLYPEW] ($gp100_{40-42/47-52}$; *m/z* 1220.6) (*C*), of PCP [RTKAWNR] ($gp100_{40-46}$; *m/z* 931.5) (*D*) and of PSP [VSRQL][VSRQL] ($gp100_{35-39/35-39}$; *m/z* 1185.8) (*E*). The PSPs [RTK][QLYPEW] and [VSRQL][VSRQL] are *cis* and *trans* PSPs, respectively. The upper spectra belong to digestions performed in 100 μl $H_2{}^{16}O$ buffer and the lower spectra to reactions carried out in 100 μl $H_2{}^{18}O$ buffer. Digestions of $gp100_{40-52}$ (40 μM) (*B–D*) or of $gp100_{35-57}$ (30 μM) (*E*) were carried out by LcL 20S proteasomes (3 μg or 1 μg, respectively) for 3h in TEAD buffer. All charts are representative examples of repeated experiments.

TABLE IV
*PSPs are relevant potential MHC class I epitopes*

(*A*) The content of PCP and PSP 9–12mers generated by different 20S proteasomes during the processing of the four substrates is reported in the second and third column. The amount of the PCP and PSP potential MHC class I epitopes is shown in the fourth and fifth column (as predicted by SYFPHEITY (20)) as well as in the sixth and seventh column (as predicted by IEDB (21)). The content is expressed as pmol Σ PCP or Σ PSP per nmol cleaved substrate (+/− S.D.) and it is the mean of PCPs or PSPs of all four substrates ($gp100_{35-57}$, $gp100_{201-230}$, $pp89_{16-40}$, $LLO_{291-317}$) as computed by *QME* for each time point of the reactions. S.D. is the standard deviation amongst different time points of the means of all experiments performed with all substrates. Notably, PCPs and PSPs produced during the degradation of the four substrates exhibited a similar average length but different statistical variance, with a higher number of PSPs close to the standard length of MHC class I epitopes and precursors. Such a phenomenon could explain the higher frequency of 9–12mers amongst PSPs and might be due to the significantly shorter length of the splice-reactants in comparison to the entire pool of PCPs. (*B*) The average length of PCPs and PSPs identified amongst the Σ PCP/PSP of the substrates $gp100_{35-57}$, $gp100_{201-230}$, $pp89_{16-40}$, $LLO_{291-317}$ is here reported and shall be compared to the average length of the PCPs that are spliced by proteasomes thereby generating the PSPs. This calculation is based on the number of PSPs and PCPs identified and not on their amount.

A

| 20S proteasomes | 9–12mers | | MHC class I potential epitopes (SYFPHEITY) | | MHC class I potential epitopes (IEDB) | |
|---|---|---|---|---|---|---|
| | Σ PCP | Σ PSP | Σ PCP | Σ PSP | Σ PCP | Σ PSP |
| T2 cell line | 192.35+/−37.68 | 21.61+/−1.53 | 56.60+/−15.23 | 2.77+/−0.94 | 74.89+/−21.06 | 16.19+/−3.07 |
| Human erythrocytes | 254.87+/−96.18 | 18.51+/−2.93 | 136.58+/−54.93 | 3.25+/−0.65 | 149.43+/−55.82 | 15.02+/−2.65 |
| T2.27 cell line | 245.86+/−85.64 | 16.53+/−3.02 | 91.50+/−52.22 | 3.90+/−0.76 | 139.10+/−82.34 | 11.94+/−1.17 |
| LcL | 473.16+/−89.31 | 26.84+/−13.11 | 110.60+/−18.64 | 4.26+/−1.67 | 186.56+/−21.98 | 21.52+/−9.77 |
| Human spleen | 195.41+/−29.28 | 15.02+/−4.04 | 110.97+/−54.18 | 3.55+/−0.87 | 148.18+/−52.77 | 13.18+/−6.02 |
| s-proteasomes (mean) | 223.61+/−35.16 | 20.06+/−1.60 | 96.59+/−23.41 | 3.01+/−0.48 | 112.16+/−22.99 | 15.60+/−1.61 |
| i-proteasomes (mean) | 304.81+/−26.06 | 19.46+/−6.34 | 104.36+/−19.01 | 3.91+/−0.80 | 157.95+/−27.33 | 15.55+/−4.77 |
| Proteasomes (mean) | 264.21+/−19.09 | 19.76+/−3.61 | 100.47+/−20.56 | 3.46+/−0.61 | 135.05+/−20.41 | 15.58+/−2.37 |

B

| | PCPs ($n$ = 316) | PSPs ($n$ = 39) |
|---|---|---|
| Average length | 10.8 | 12.7 |
| Length variance | 26.5 | 10.1 |
| Length S.D. | 5.1 | 2.9 |
| N-terminal splice-reactant's length mean | | 5.7 |
| N-terminal splice-reactant's length variance | | 6.2 |
| N-terminal splice-reactant's length S.D. | | 2.5 |
| C-terminal splice-reactant's length mean | | 6.2 |
| C-terminal splice-reactant's length variance | | 10.6 |
| C-terminal splice-reactant's length S.D. | | 3.3 |

tural similarity between yeast and mammalian 20S proteasomes this observation most likely also applies to the different active sites of mammalian 20S proteasomes (27). Interestingly, irrespective whether yeast wild type or β subunit mutant 20S proteasomes were studied, similar Σ PSP per nmol of cleaved substrate were generated. This on first sight surprising result may be explained by the fact that the major splicing sites such as the residues $Q_{38}$, $L_{48}$, $P_{50}$, $E_{51}$ and $T_{53}$ in the polypeptide $gp100_{35-57}$ were generated by all three proteolytic β subunits to a similar extent (supplemental Figs. S10 and S13).

Confirming the model proposed by Vigneron *et al.* (5), our experiments in $H_2{}^{18}O$-TEAD buffer demonstrate for the first time that PCPS indeed occurs via direct transpeptidation in *cis* and in *trans*. This occurs through the formation of an acyl-enzyme intermediate and concomitant breakdown of the peptide bond, followed by the nucleophilic attack of the N terminus of one PCP. Our experiments also confirmed the dependence of PSP generation on cleavage preferences (Figs. 3, 4) and demonstrated that the amount of the less abundant spliced reactant peptide is one of the rate-limiting factors of PCPS, independent of whether the splice-reactant is located at the N- or C-terminus of the generated PSP.

However, in striking contrast to Vigneron's model, which also implied that PCPS is not restricted by a particular sequence motif and can occur at any major substrate cleavage site used by 20S proteasomes, our experiments demonstrate that the main cleavage sites within the substrate sequence are often not the main ligation sites of the PSPs. The degradation analysis of $gp100_{35-57}$ by spleen 20S proteasome may serve as representative example (Fig. 6). Indeed, whereas the cleavages after $Leu_{39}$ represented 62.5% of the total proteasomal cuts, only 16.4% and 7.7% of cleavages behind $Leu_{39}$ produced the PSP's P1 and P1′ residues, respectively. Conversely, $Glu_{51}$, $Trp_{52}$, and $Thr_{53}$ were often P1 and P1′ splice-sites, despite the fact that they represented only minor cleavage sites. Hence, the sequence-specificity that determines cleavage site usage and the sequence specificity that determines ligation efficiency both affect PCPS although they do not necessarily overlap.

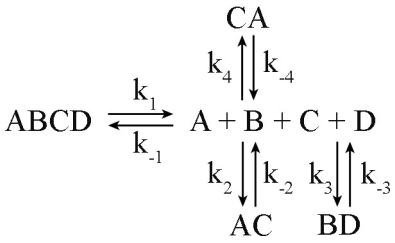
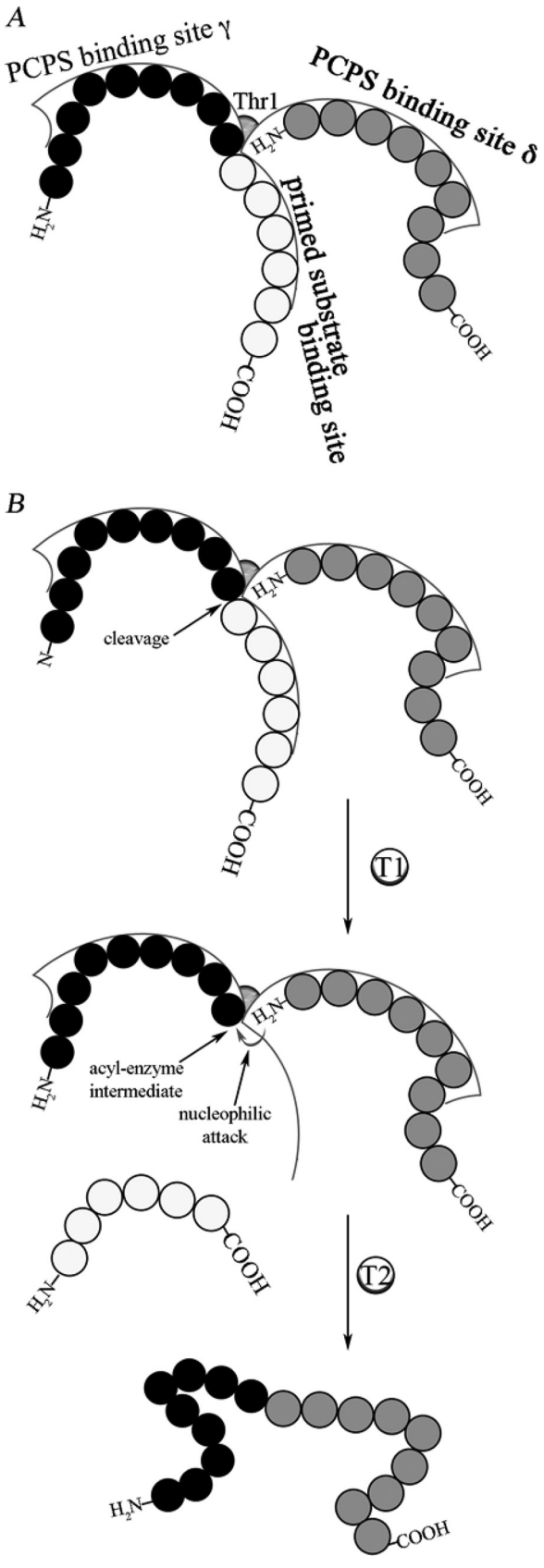



FIG. 8. **Biochemical representation of the cleavage and splicing reactions.** A biochemical model of proteasomal cleavage and splicing reactions of a hypothetical substrate ABCD. By cleavage after each residue (reaction $k_1$) the substrate ABCD is transformed in the products A, B, C, D. A reverse reaction ligates the product A to B to C to D thereby reconstituting the original substrate ABCD (reaction $k_{-1}$). Ligation reactions between the products of $k_1$ generate new compounds: AC (reaction $k_2$), BD (reaction $k_3$), CA (reaction $k_4$). For each of these new compounds a proteolytic reaction can reconstitute the product pool A + B + C + D, *i.e.* the reaction $k_{-2}$, reaction $k_{-3}$, reaction $k_{-4}$, respectively. Although the reactions $k_2$, $k_3$, $k_4$ are ligations as $k_{-1}$ they are all different reactions because they produce different products. Proteasomes, as any enzyme, are catalyzing a reaction which has its own equilibrium and they can only accelerate the reaction based on enzyme-substrate affinity. Nevertheless, proteasomes can accelerate one reaction more than others. Which reaction is favored by proteasome depends on its active site conformation and the affinity to the reactants. Therefore, different proteasome isoforms might accelerate reactions, which have different equilibrium between reactants and products. For example, in our experiments we sometimes observed that the amount of specific PSPs was higher than the amount of the less abundant reactant peptides. This was the case for the PSP [AYISSVAY][RQVYLKL] ($LLO_{291-298/300-306}$) and the PCP [RQVYLKL] ($LLO_{300-306}$) derived from the digestion of the polypeptide $LLO_{291-317}$ (Figs. 4*F*–4*G*). This PCPS reaction was an example of a reaction where the equilibrium between proteolysis and PCPS did not strongly privilege hydrolysis in contrast to the majority of the proteolysis/PCPS reactions like the PSP [AYISSVAY][AYISSVAY] ($LLO_{291-298/291-298}$), which was produced in a clear lower amount than its reactant [AYISSVAY] ($LLO_{291-298}$) in the same digestion (Figs. 4*B*-4*C*). Thus, one may legitimately say that, in the biochemical model represented here, the higher catalysis of the reactions $k_2/k_{-2}$ rather than the reactions $k_1/k_{-1}$ or the reactions $k_3/k_{-3}$ by one proteasome iso-form, because of a higher affinity for the peptides A and C, might lead to a substantially different amount of Σ PSP as well as of the spliced peptide AC.

The novel structural model of the PCPS catalytic pocket as described in Fig. 9 may explain the majority of our results. According to what is shown in Fig. 9*A*, the active Thr1 is localized between two PCPS binding sites. The PCPS binding site γ binds the N-terminal splice-reactant and the PCPS binding site δ binds the C-terminal splice-reactant. Because the stability of the acyl-enzyme intermediate is a key factor in PCPS and, at least in part, depends on the binding of the P1-P2-P3 residues of the N-terminal splice-reactant to the nonprimed substrate binding site of the catalytic site (27), the PCPS binding site γ and the nonprimed substrate binding site very likely coincide. In opposite, the PCPS binding site δ and the primed substrate binding site might be distinct. Indeed, one might hypothesize that the PCPS binding site δ and the primed substrate binding site are both allocated within the proteolytic pocket with their grooves ending at the active Thr1. The proteolytic pocket of the proteasome could have indeed sufficient room to allocate both the substrate and the C-terminal splice-reactant as suggested by the crystal structure of mouse proteasomes (27). A three dimensional representation of such an hypothesis is described for the chymotryptic-like pocket of the mouse i-proteasome in supplemental Fig. S14. Therefore, according to our model, although the substrate is entering the catalytic pocket and allocating its sequence at the nonprimed and primed substrate binding sites, the C-terminal splice-reactant could already be bound at the PCPS binding site δ and perform the nucleophilic attack on the acyl-enzyme intermediate as soon as (or while) the C-terminal substrate fragment leaves the primed substrate binding site (Fig. 9*B*). Being occupied by the splice-reactants, the catalytic pocket could therefore create an closer proximity of the splice-reactants thereby providing together with the surrounding proteasome surface a molecular crowding environment to facilitate PCPS (9). In addition, the retention time of the N-terminal splice-reactant as acyl-enzyme intermediate could be extremely short with the C-terminal splice-reactant already being in close proximity of the active Thr 1. Furthermore, we found that the average size of the N- and C-terminal splice-reactants is around 6 amino acids (Table IV*B*). This data fits with the hypothesis that the PCPS binding site γ and the nonprimed substrate binding site coincide. Indeed, amino acids located up to five residues before and after a cleavage position have been shown to determine specific cleavage usage suggesting that the substrate binding grooves might allocate around five substrate residues (3, 28).

Nevertheless, the frequencies of cleavages generating Σ PCP and the Σ N- and C-terminal splice-reactants substantially diverge (Fig. 6 and supplemental Fig. S13). This apparent contradiction could be explained by hypothesizing that: i. PCPS binding site δ and the primed substrate binding site do not coincide and therefore can exhibit distinct peptide specificities (Fig. 9); ii. the retention time of the N-terminal splice-reactant may have an opposite effect on cleavage and PCPS rates. Indeed, a short retention time of a given substrate sequence, which reflects its low affinity for the proteasomal substrate binding sites, leads to a null or low cleavage rate after the P1 peptide bond (3). Hence, a longer retention time of this sequence at the substrate binding sites produces a high cleavage of the peptide bond. Albeit a prolonged stabilization of the acyl-enzyme intermediate might slow down the activity of the proteolytic β subunit leading to an overall reduced cleavage of this peptide bond. Conversely, the endergonic nature of peptide ligation renders the process energetically unfavorable (9). A longer retention time of the splice-reactants in proximity of the proteasome catalytic site (Thr 1) might hence be expected to be compulsory for the reaction. Therefore, an increase of the life span of the peptide bound to the nonprimed substrate binding site




could lead to an overall reduced cleavage and remarkable ligation of the peptide bond. Our experimental data supporting this hypothesis and relative interpretation are reported in supplementary material.

Such a model would explain the correlation between the amount of splice reactants and PSP (Figs. 3, 4), the discrepancy between the SCS and the frequency of cleavages generating PSP P1 and P1′ residues (Fig. 6) and also why the ligation sites are often represented by minor cleavage sites (Fig. 6 and supplemental Fig. S13). Nevertheless, the structural model reported in Fig. 9 and supplemental Fig. S14 commands further structural and mechanistic elucidations to better understand the biochemical process and to verify the location of the PCPS binding site δ.

The structural model of PCPS with two binding sites that allocate preferentially small peptides might support the hypothesis of an evolution of the MHC class I pockets according to the features of the PCP and PSP produced by proteasome. This hypothesis would be in agreement with the idea that, by exploiting a preexisting process, PCPS might have contributed to maximizing the diversity of antigenic peptides at low energy cost for cells during evolution (9, 29). Indeed, both human and yeast 20S proteasomes generated a relative high amount of PSPs suitable for being presented on MHC class I molecules, because the limited length of the splice-reactants rendered the PSPs better MHC class I-restricted epitope candidates than the conventional PCPs (Table IV*A*).

This observation also implies that an important part of the MHC class I-restricted antigenic pool, produced by PCPS, may have been ignored so far.

FIG. 9. **Model of PCPS binding sites.** *A*, Illustration of the PCPS binding sites γ and δ and the primed substrate binding site convergent to the active Thr1 of the proteasome. The PCPS binding site γ most likely coincides with the nonprimed substrate binding site (27), whereas the PCPS binding site δ could be different than the primed substrate binding site as it is illustrated here. Both PCPS binding sites γ and δ have a pocket that can accommodate 5–6 residue peptides, which can have a N- or C-terminal extension of an undefined length, respectively. The substrate is here represented by black (N-terminal to the cleavage) and white (C-terminal to the cleavage) circles, whereas gray circles represent the C-terminal splice-reactant. Each circle symbolizes an amino acid. *B*, Substrate binds with its N terminus the PCPS binding sites γ/nonprimed substrate binding site and with its C terminus the primed substrate binding site. The C-terminal splice-reactant binds the PCPS binding site δ with its N terminus in proximity of the active Thr1. Substrate and C-terminal splice-reactant might bind at the same time the binding grooves of a catalytic β subunit. During the cleavage by Thr1 of one of the substrate peptide bond the acyl-enzyme intermediate is formed and the C-terminal fragment of the substrate is released (step T1). Consequently, the C-terminal splice-reactant, which is already in proximity of the Thr1, might perform the nucleophilic attack with its N terminus to the acyl-enzyme intermediate leading to the formation of the new PSP (step T2). A 3D representation of the proposed model is shown in supplemental Fig. S14.

*Acknowledgments*—We thank Dr. Elena Bellavista (University of Bologna), Elke Giessmann (Universitäetsmedizin Berlin Charitè) and Marion Weberruβ (University of Toronto) for the technical assistance, Juliane Liepe and Dr. Suhail Islam (Imperial College London) for the development of the 3D model of interaction between proteasome β5i subunit and splice-reactants.

* This work was financed in part by grants of the Deutsche Forschungsgemeinschaft KL427/15-1, SFB TR19/B3 as well as Berliner Krebsgesellschaft (KÖFF201102) to P. M. K. M.M. benefited from the A.V. Humboldt PostDoc fellowship.

[S] This article contains supplemental Tables S1 and S2 and Figs. S1 to S14.

‖ To whom correspondence should be addressed email: Institut für Biochemie, Charité - Universitätsmedizin Berlin, Oudenarder Straβe 16, 13347 Berlin, Germany. E-mail: michele.mishto@charite.de; p-m.kloetzel@charite.de.

REFERENCES

1. Schwartz, A. L., and Ciechanover, A. (2009) Targeting proteins for destruction by the ubiquitin system: implications for human pathobiology. *Annu. Rev. Pharmacol. Toxicol.* **49,** 73–96
2. Kloetzel, P. M. (2001) Antigen processing by the proteasome. *Nat. Rev. Mol. Cell Biol.* **2,** 179–187
3. Borissenko, L., and Groll, M. (2007) Diversity of proteasomal missions: fine tuning of the immune response. *Biol. Chem.* **388,** 947–955
4. Warren, E. H., Vigneron, N. J., Gavin, M. A., Coulie, P. G., Stroobant, V., Dalet, A., Tykodi, S. S., Xuereb, S. M., Mito, J. K., Riddell, S. R., and Van den Eynde, B. J. (2006) An antigen produced by splicing of noncontiguous peptides in the reverse order. *Science* **313,** 1444–1447
5. Vigneron, N., Stroobant, V., Chapiro, J., Ooms, A., Degiovanni, G., Morel, S., van der Bruggen, P., Boon, T., and Van den Eynde, B. J. (2004) An antigenic peptide produced by peptide splicing in the proteasome. *Science* **304,** 587–590
6. Dalet, A., Vigneron, N., Stroobant, V., Hanada, K., and Van den Eynde, B. J. (2010) 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
7. Hanada, K., Yewdell, J. W., and Yang, J. C. (2004) Immune recognition of a human renal cancer antigen through post-translational protein splicing. *Nature* **427,** 252–256
8. Dalet, A., Robbins, P. F., Stroobant, V., Vigneron, N., Li, Y. F., El-Gamil, M., Hanada, K., Yang, J. C., Rosenberg, S. A., and Van den Eynde, B. J. (2011) An antigenic peptide produced by reverse splicing and double asparagine deamidation. *Proc. Natl. Acad. Sci. U.S.A.* **108,** E323-E331
9. Berkers, C. R., de Jong, A., Ovaa, H., and Rodenko, B. (2009) Transpeptidation and reverse proteolysis and their consequences for immunity. *Int. J. Biochem. Cell Biol.* **41,** 66–71
10. Liepe, J., Mishto, M., Textoris-Taube, K., Janek, K., Keller, C., Henklein, P., Kloetzel, P. M., and Zaikin, A. (2010) The 20S Proteasome Splicing Activity Discovered by SpliceMet. *PLOS Computational Biol.* **6**, e1000830
11. Textoris-Taube, K., Henklein, P., Pollmann, S., Bergann, T., Weisshoff, H., Seifert, U., Drung, I., Mügge, C., Sijts, A., Kloetzel, P. M., and Kuckelkorn, U. (2007) The N-terminal flanking region of the TRP2360–368 melanoma antigen determines proteasome activator PA28 requirement for epitope liberation. *J. Biol. Chem.* **282,** 12749–12754
12. Mishto, M., Santoro, A., Bellavista, E., Sessions, R., Textoris-Taube, K., Dal Piaz, F., Carrard, G., Forti, K., Salvioli, S., Friguet, B., Kloetzel, P. M., Rivett, A. J., and Franceschi, C. (2006) A structural model of 20S immunoproteasomes: effect of LMP2 codon 60 polymorphism on expression, activity, intracellular localisation and insight into the regulatory mechanisms. *Biol. Chem.* **387,** 417–429
13. Kuckelkorn, U., Frentzel, S., Kraft, R., Kostka, S., Groettrup, M., and Kloetzel, P. M. (1995) Incorporation of major histocompatibility complex–encoded subunits LMP2 and LMP7 changes the quality of the 20S proteasome polypeptide processing products independent of interferon-gamma. *Eur. J. Immunol.* **25,** 2605–2611
14. Mishto, M., Bellavista, E., Ligorio, C., Textoris-Taube, K., Santoro, A., Giordano, M., D'Alfonso, S., Listi, F., Nacmias, B., Cellini, E., Leone, M., Grimaldi, L. M., Fenoglio, C., Esposito, F., Martinelli-Boneschi, F., Galimberti, D., Scarpini, E., Seifert, U., Amato, M. P., Caruso, C., Foschini, M. P., Kloetzel, P. M., and Franceschi, C. (2010) Immunoproteasome LMP2 60HH variant alters MBP epitope generation and reduces the risk to develop multiple sclerosis in Italian female population. *PLoS One* **5,** e9287
15. Bech-Otschir, D., Helfrich, A., Enenkel, C., Consiglieri, G., Seeger, M., Holzhütter, H. G., Dahlmann, B., and Kloetzel, P. M. (2009) Polyubiquitin substrates allosterically activate their own degradation by the 26S proteasome. *Nat. Struct. Mol. Biol.* **16,** 219–225
16. Niles, R., Witkowska, H. E., Allen, S., Hall, S. C., Fisher, S. J., and Hardt, M. (2009) Acid-catalyzed oxygen-18 labeling of peptides. *Anal. Chem.* **81,** 2804–2809
17. Li, L., Kresh, J. A., Karabacak, N. M., Cobb, J. S., Agar, J. N., and Hong, P. (2008) A hierarchical algorithm for calculating the isotopic fine structures of molecules. *J. Am. Soc. Mass Spectrom.* **19,** 1867–1874
18. Mishto, M., Luciani, F., Holzhütter, H. G., Bellavista, E., Santoro, A., Textoris-Taube, K., Franceschi, C., Kloetzel, P. M., and Zaikin, A. (2008) Modeling the in vitro 20S proteasome activity: the effect of PA28-alpha-beta and of the sequence and length of polypeptides on the degradation kinetics. *J. Mol. Biol.* **377,** 1607–1617
19. Peters, B., Janek, K., Kuckelkorn, U., and Holzhütter, H. G. (2002) Assessment of proteasomal cleavage probabilities from kinetic analysis of time-dependent product formation. *J. Mol. Biol.* **318,** 847–862
20. Rammensee, H., Bachmann, J., Emmerich, N. P., Bachor, O. A., and Stevanović, S. (1999) SYFPEITHI: database for MHC ligands and peptide motifs. *Immunogenetics* **50,** 213–219
21. Peters, B., and Sette, A. (2007) Integrating epitope data into the emerging web of biomedical knowledge resources. *Nat. Rev. Immunol.* **7,** 485–490
22. Ishizuka, J., Grebe, K., Shenderov, E., Peters, B., Chen, Q., Peng, Y., Wang, L., Dong, T., Pasquetto, V., Oseroff, C., Sidney, J., Hickman, H., Cerundolo, V., Sette, A., Bennink, J. R., McMichael, A., and Yewdell, J. W. (2009) Quantitating T cell cross-reactivity for unrelated peptide antigens. *J. Immunol.* **183,** 4337–4345
23. Salimi, N., Fleri, W., Peters, B., and Sette, A. (2010) Design and utilization of epitope-based databases and predictive tools. *Immunogenetics* **62,** 185–196
24. Cardinaud, S., Consiglieri, G., Bouziat, R., Urrutia, A., Graff-Dubois, S., Fourati, S., Malet, I., Guergnon, J., Guihot, A., Katlama, C., Autran, B., van Endert, P., Lemonnier, F. A., Appay, V., Schwartz, O., Kloetzel, P. M., and Moris, A. (2011) CTL escape mediated by proteasomal destruction of an HIV-1 cryptic epitope. *PLoS Pathog.* **7**, e1002049
25. Chen, P., and Hochstrasser, M. (1996) Autocatalytic subunit processing couples active site formation in the 20S proteasome to completion of assembly. *Cell* **86,** 961–972
26. Heinemeyer, W., Fischer, M., Krimmer, T., Stachon, U., and Wolf, D. H. (1997) The active sites of the eukaryotic 20 S proteasome and their involvement in subunit precursor processing. *J. Biol. Chem.* **272,** 25200–25209
27. Huber, E. M., Basler, M., Schwab, R., Heinemeyer, W., Kirk, C. J., Groettrup, M., and Groll, M. (2012) Immuno- and constitutive proteasome crystal structures reveal differences in substrate and inhibitor specificity. *Cell* **148,** 727–738
28. Nussbaum, A. K., Dick, T. P., Keilholz, W., Schirle, M., Stevanović, S., Dietz, K., Heinemeyer, W., Groll, M., Wolf, D. H., Huber, R., Rammensee, H. G., and Schild, H. (1998) 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
29. Vigneron, N., and Van den Eynde, B. J. (2011) Proteasome subtypes and the processing of tumor antigens: increasing antigenic diversity. *Curr. Opin. Immunol.* **24,** 84–91