The 20S Proteasome Splicing Activity Discovered by SpliceMet

Juliane Liepe1,2,*, Michele Mishota1,3,*, Kathrin Textoris-Taube1, Katharina Janek1, Christin Keller1, Petra Henklein1, Peter Michael Kloetzel1,*, Alexey Zaikin4

1 Institut für Biochemie, Charité, Universitätsmedizin Berlin, Berlin, Germany, 2 Centre for Bioinformatics, Division of Molecular Biosciences, Imperial College London, London, United Kingdom, 3 Interdepartmental Center for Studies on Biophysics, Bioinformatics and Biocomplexity ‘L. Galvani’ (CIG), University of Bologna, Bologna, Italy, 4 Institute for Women’s Health and Department of Mathematics, University College London, London, United Kingdom

Abstract

The identification of proteasome-generated spliced peptides (PSP) revealed a new unpredicted activity of the major cellular protease. However, so far characterization of PSP was entirely dependent on the availability of patient-derived cytotoxic CD8+ T lymphocytes (CTL) thus preventing a systematic investigation of proteasome-catalyzed peptide splicing (PCPS). For an unrestricted PSP identification we here developed SpliceMet, combining the computer-based algorithm ProteaJ with in vitro proteasomal degradation assays and mass spectrometry. By applying SpliceMet for the analysis of proteasomal processing products of four different substrate polypeptides, derived from human tumor as well as viral antigens, we identified fifteen new spliced peptides generated by PCPS either by cis or from two separate substrate molecules, i.e., by trans splicing. Our data suggest that 20S proteasomes represent a molecular machine that, due to its catalytic and structural properties, facilitates the generation of spliced peptides, thereby providing a pool of qualitatively new peptides from which functionally relevant products may be selected.

Introduction

The multiple subunit 20S proteasome is the central catalytic unit of the ubiquitin proteasome system (UPS) and catalytic core of the 26S proteasome that is built by the association of the two 19S regulator complexes with the catalytic 20S core [19S-20S-19S]. With its N-terminal threonine residues as the single active site of the protease, however, so far characterization of PSP was entirely dependent on the availability of patient-derived cytotoxic CD8+ T lymphocytes (CTL) thus preventing a systematic investigation of proteasome-catalyzed peptide splicing (PCPS). For an unrestricted PSP identification we here developed SpliceMet, combining the computer-based algorithm ProteaJ with in vitro proteasomal degradation assays and mass spectrometry. By applying SpliceMet for the analysis of proteasomal processing products of four different substrate polypeptides, derived from human tumor as well as viral antigens, we identified fifteen new spliced peptides generated by PCPS either by cis or from two separate substrate molecules, i.e., by trans splicing. Our data suggest that 20S proteasomes represent a molecular machine that, due to its catalytic and structural properties, facilitates the generation of spliced peptides, thereby providing a pool of qualitatively new peptides from which functionally relevant products may be selected.


Editor: Rob J. De Boer, Utrecht University, The Netherlands

Received December 17, 2009; Accepted May 24, 2010; Published June 24, 2010

Copyright: © 2010 Liepe et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was financed in part by grants of the Deutsche Forschungsgemeinschaft Sonderforschungsbereich (Kl421/15, SFB 740, TR19) to PMK and by VW foundation to AZ and JL and the UCLH/UCL NIHR Comprehensive Biomedical Research Centre to AZ. MM received funding from the AV Humboldt PostDoc fellowship. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: p-m.kloetzel@charite.de

† These authors contributed equally to this work.
Author Summary

MHC class I molecules present antigenic peptides derived from endogenously expressed foreign or aberrant protein molecules to the outside world so that they can be specifically recognised by cytotoxic T lymphocytes (CTLs) at the cell surface. Responsible for the generation of these peptides is the 20S proteasome, which is the major proteolytic enzyme of the cell. These peptides were so far believed to exhibit a linear sequence identical to that found in the unprocessed parental protein. Using patient derived CTL it was previously shown that by proteasome catalyzed peptide splicing, i.e., by fusion of two protea-

proteasomal splicing reaction, as such, must be a considerably more frequent event than so far assumed. But even if peptide proteasomal splicing is an intrinsic additional catalytic property from two distinct substrate molecules. Our data reveal that peptide splicing is an intrinsic additional catalytic property of the proteasome, which may provide a qualitatively new peptide pool for immune selection.

Proteasomal Splicing Discovered by SpliceMet

1) Calculation of all combinatorially possible PSP and setting of the ProteaJ database. The digestion of the substrate of length $L$ with a sequence of amino acids $a_i, i = 1...L$ may result in $SCP = \frac{1}{2} (L-L_{ext}+1)(L-L_{ext}+2)$ cleavage products (PCP) each of which can be denoted as $PCP_i$, where the product starts at the position $i$, $i = 1...L-L_{ext}+1$ (C-terminus) and ends at the position $j = i+L_{ext}-1...[N]$ (N-terminus). $L_{ext}$ describes the minimal length of a PCP that can produce a PSP (here $L_{ext} = 2$). Any two PCP, and $PP_{i,k}$ may be spliced into $PP_{i,k}^{cis}$ or $PP_{i,k}^{trans}$. For the total amount of generated products $S_{all}$, including PCP and PSP, we have \begin{equation}
S_{all} = \frac{1}{2} \sum_{i=1}^{L-2} \sum_{j=i+L_{ext}-1}^{L} L_{ext} - \sum_{k=1}^{L_{ext}} \sum_{l=k+1}^{L_{ext}} \sum_{n=L_{ext}+1}^{L}. \end{equation}

2) Matching with the LC-ESI/MS full spectra. The presence of the theoretical $m/z$ values was detected among MS signals of the digestion products of the investigated peptide of choice.

3) Peak detection of all the computed $m/z$ values. In the LC-ESI mass chromatogram we identified the significant peaks for each theoretical $m/z$ value. For each theoretical $m/z$ value either no peak or several peaks could be detected and defined by their $m/z$ and retention time (RT).

4) Analysis of $m/z$ time-dependent kinetics and establishment of an inclusion list for the LC-ESI/MS measurements. In time-dependent processing experiments (signal intensity versus time of digestion) identified peaks that did not fulfill the following criteria were eliminated from the candidate list: i. initial intensity ($t = 0$) smaller than MAX (e.g. here $= 10^7$ for measurements by DECA XP MAX instrument); ii. monotonously ascending signal intensity towards a maximum followed by a monotonous decline in case assay condition allowed re-entry of the PSP. It was assumed that the monotonous increase resulted from the continuous production of PSP and the decrease from the “re-entry” event.

Next, we defined $t_{max}$ as the digestion time when the highest amount of generated PSP was observed and sorted all pairs ($m/z$, RT) with respect to $t_{max}$ into groups indexed as $g$ of the size $D_g$. If $D_g > D_{max}$ (here 15 dependent on MS resolution) then the corresponding group was split into subgroups $g_i$ of size smaller than $D_{max}$. The number of groups determined the number of additional up-scaled processing assays in which the absolute concentration of substrate and proteasome were increased keeping the relative substrate/proteasome ratio constant, whereas the total

Before a systematic, CTL-independent investigation of PSP we therefore developed SpliceMet: a method that combines combinatorial computations (ProteaJ) with mass spectrometric (MS) analyses of proteasome-generated peptides. Based on a given protein or peptide sequence, ProteaJ produces a data set with the $m/z$ value of all theoretically possible PSP that may be generated by the proteasome through the combination of any two fragments (greater than one amino acid in length) generated from the same substrate molecule (in $cis$) or from separate substrate molecules (in $trans$) and ligated in a normal or reverse order. This is followed by MS analysis of in vitro digests of the synthetic peptide substrate and by comparison of the MS signals obtained with the theoretical ProteaJ-computed $m/z$ values. By matching the theoretical values with the experimentally obtained $m/z$ values and verifying the peptide generation kinetics, a restricted list of candidate PSP is generated. Their presence in 20S proteasome digests of substrates is then investigated by LC-ESI-MS/MS and LC-MALDI-TOF/TOF-MS/MS leading to the final identification of the PSP (Fig. 1).

Results

SpliceMet

The SpliceMet method is organized into two main experimental blocks characterized by 7 main steps (Figure 1). To reduce the number of possible proteasome generated spliced peptides (PSP) the first block utilizes the following 4 main steps that are subsequently investigated in the second block. The first experimental block combines the computational algorithm ProteaJ with proteasome in vitro digests of a synthetic peptide of choice and mass spectrometric (MS) analyses as follows:

1. Calculation of all combinatorially possible PSP and setting of the ProteaJ database.
2. Matching with the LC-ESI/MS full spectra.
3. Peak detection of all the computed $m/z$ values.
4. Analysis of $m/z$ time-dependent kinetics and establishment of an inclusion list for the LC-ESI/MS measurements.
number of subgroups represented the number of requested new MS runs. The resulting m/z, RT, t\text{max} established the inclusion list.

The second block consists of the following 3 steps:

5) **LC-ESI-MS/MS analysis with inclusion list.** Precursor ion selection for MS/MS analysis was performed using the established inclusion list enabling the fragmentation analysis of even low-abundance peptides. MS/MS spectra were analyzed with Bioworks software version 3.3 (Thermo Fisher) using the ProteaJ database. Significant hits which were annotated as PSP showed a peptide probability p < 0.00005.

---

**Figure 1. SpliceMet.** Applying the computer program ProteaJ on a peptide sequence of choice, m/z values of all theoretically possible proteasomal cleavage (PCP) and splicing (PSP) products are calculated (1st step). This is followed by an in vitro digest of the synthetic substrate and the comparison of the obtained MS signals with the theoretical m/z values (2nd). Matching of the signals and verification of peptide generation kinetics results in an inclusion list for LC-ESI-MS/MS analysis required for identification of the PSP (3rd, 4th). For final confirmation, the MS/MS spectra (5th) and the HPLC-RT of proposed PSP (6th) are compared with those of the analogous synthetic peptides. For the identification of those PSP candidates that do not fully satisfy these requisites, the generation of PSP is up-scaled followed by HPLC fractionation with an extended gradient and the fractions are analyzed by nano-LC-MALDI-TOF/TOF-MS (7th).

doi:10.1371/journal.pcbi.1000830.g001
Proteasomal Splicing Discovered by SpliceMet

<table>
<thead>
<tr>
<th>Table 1. Computation of cleavage and splicing products.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>products</strong></td>
</tr>
<tr>
<td>all fragments</td>
</tr>
<tr>
<td>PCP</td>
</tr>
<tr>
<td>cis - normal</td>
</tr>
<tr>
<td>cis - reverse</td>
</tr>
</tbody>
</table>

Described are the conditions to compute all products of a specific type (PCP, cis-normal PSP and -reverse PSP). The indices $i$, $j$, $k$ and $n$ are the amino acid positions of the product, e.g. $PSP_{13\text{mer}}$. $L$ is the length of the substrate, $L_{ext}$ is the minimal length of a PCP that can produce a PSP.

doi:10.1371/journal.pcbi.1000830.t001

6) Comparison with synthetic peptides. All identified PSP resulting from step 5 were manually confirmed by comparison with synthetic peptides of the same sequence. The candidate PSP and their synthetic analogues had to exhibit a similar RT (delta RT < 0.5 min) and fragmentation pattern in the LC-ESI-MS/MS analysis.

7) Validation of PSP sequences by MALDI-TOF. In some experiments the requirements outlined in step 5 and 6 were not fully met requesting further MS identification. In this case, we proceeded by fractionating the digestion products by reverse phase (RP)-HPLC and by analyzing each fraction by LC-ESI-MS/MS using an inclusion list with the $m/z$ values of the PSP candidates. Their RT in the HPLC run was also compared with that of the corresponding synthetic peptides. Those fractions with MS/MS and RT that matched the PSP were lyophilized and fractionated again using a more focused HPLC method to decrease the number of peptides in each fraction. The up-scaled fractions were subsequently compared with the RT of the synthetic PSP and analyzed by nano-LC-MALDI-TOF/TOF-MS/MS.

Validation of SpliceMet

For proof of principle we initially investigated 20S proteasome catalyzed peptide splicing during proteasomal degradation of the synthetic 13mer peptide $gp100_{40-52}$, RTKAWNRLQYPEW, previously shown to serve as substrate for PSP generation [6]. For the experiments we used 20S proteasomes of Lymphoblastoid cell Lines (LcL), which possess splicing activity [7] and predominantly resemble the immunoproteasome subtype [13,14]. Following each step of SpliceMet we obtained a progressive decrease of the number of candidate PSP leading to the identification of the previously described PSP $gp100_{40-42/47-52}$ by LC-ESI/MS/MS at the 6th step of SpliceMet (Figure 2). The substantial reduction of PSP in the candidate list (Table 2) and the final identification of the PSP $gp100_{40-42/47-52}$ validated our analysis method.

To verify the hypothesis of the occurrence of a proteasome-dependent trans splicing reaction we performed in vitro digestions 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 and $^{15}N$-Leu (RTK*AWNRLQ*YPEW). As shown in Figure 2, we indeed detected PSP variants as being the results of cis (variants $\&\&$) or of trans (variants $\&\&\&$) splicing, demonstrating that PCPS can occur not only in cis but also in trans (see also Figure S1).

Identification of new PSP in the proteasomal digestion of $gp100_{35-57}$

By applying SpliceMet we investigated the generation of new PSP derived from the proteasomal degradation products of the 23mer peptide $gp100_{40-57}$, which is a N- and C-terminally extended version of $gp100_{40-52}$ by LcL. 20S proteasome (Figure 4). In these experiments we identified eight new PSP, four of which were identified at step 6 (Figure 4) and four at step 7 of SpliceMet (Table 3 & Figure S2). We also identified a ninth PSP

![Figure 2. By applying SpliceMet we identified the known PSP produced by digestion of the synthetic 13mer $gp100_{40-52}$ by 20S proteasomes. Sequence of the substrate $gp100_{40-52}$ and of the PSP $gp100_{40-42/47-52}$ and its ESI-MS/MS spectrum (double protonated with $m/z$ 610.8) are shown. In the spectra B- and Y-ions are reported. Ions' loss of water is symbolized by '$'. In the experiments (100 µl of reaction) 4 nmol of $gp100_{40-52}$ were cleaved for 36 hours by 1 µg 20S proteasome purified from LcL.](image-url)
with the sequence [VSRQ][VSRQJ] derived from splicing of two distinct molecules of the PCP gp10035–39 (Figure 5). The identification of this PSP was of particular relevance because it was the first example of PSP derived from an in vitro proteasomal digestion of a single peptide sequence.

PSP formation is a general phenomenon not restricted to the gp10035–57 sequence

Since the sequence requirement for PCPS are not yet known in a rigorous manner one might argue that the observed frequent PSP generation when gp10035–57 was used as substrate was due to the effect of the larger gp10035–57 sequence specificities. To test this we applied SpliceMet for the analysis of PSP derived from another polypeptide sequence of the same protein, i.e. gp100201–229. Among the proteasome-generated degradation products of this 25mer we identified three PSP (Table 4 and Figure S3). Since peptide fragments with overlapping sequences were spliced together these PSP were generated by a trans splicing event.

In order to exclude a peculiarity and rare tendency of the entire gp100 sequence to be spliced by PCPS we investigated the in vitro digestion products of two other peptides, i.e. the 30mer HIV-derived gag-pol32–52 and the murine cytomegalovirus (MCMV)-derived 25mer polypeptide peptide gp8935–40. The in vitro processing of gag-pol32–52 by proteasomes produced at least one PSP (Table 4 & Figure S4), whereas two PSP were detected after the digestion of the MCMV derived gp8935–57 peptide (Table 4 & Figure S5).

Discussion

SpliceMet

The aim of our study was to develop a method for the identification of spliced peptides which would allow the identification of any theoretically possible PSP and which was independent of adventitiously available CD8+ T cells and T-cell recognition assays permitting the detection of only a single spliced epitope peptide. The availability of such a method would greatly facilitate systematic studies required to elucidate the molecular mechanism of PCPS. Therefore we have developed and applied a method - SpliceMet - that, by combining computational and experimental methods, facilitates the identification of proteasome-generated spliced peptides.

Although in this investigation we have considered only polypeptide substrates up to a length of 30 amino acid residues, SpliceMet could also be applied to longer peptides or proteins to further our understanding of the mechanisms that govern PCPS and, in particular, trans-splicing. It has to be pointed out however that an increase in substrate length will lead to an exponential expansion of the ProteaJ data base to the extent that it represents a substantial fraction of the proteasomal splicing activity. Since the sequence requirement for PCPS is not yet known in a rigorous manner one might argue that the observed frequent PSP generation when gp10035–57 was used as substrate was due to the effect of the larger gp10035–57 sequence specificities. To test this we applied SpliceMet for the analysis of PSP derived from another polypeptide sequence of the same protein, i.e. gp100201–229. Among the proteasome-generated degradation products of this 25mer we identified three PSP (Table 4 and Figure S3). Since peptide fragments with overlapping sequences were spliced together these PSP were generated by a trans splicing event.

In order to exclude a peculiarity and rare tendency of the entire gp100 sequence to be spliced by PCPS we investigated the in vitro digestion products of two other peptides, i.e. the 30mer HIV-derived gag-pol32–52 and the murine cytomegalovirus (MCMV)-derived 25mer polypeptide peptide gp8935–40. The in vitro processing of gag-pol32–52 by proteasomes produced at least one PSP (Table 4 & Figure S4), whereas two PSP were detected after the digestion of the MCMV derived gp8935–57 peptide (Table 4 & Figure S5).
peptides are known to closely resemble the in vivo situation [3] our data reveal that 20S proteasomes represent a molecular machine that facilitates the generation of spliced peptides from its own cleavage products. Therefore, our data may have considerable biological implications in that they provide evidence that proteasome-dependent protein degradation results in the generation of a second, so far undetected pool of spliced peptides, from which novel potentially functionally relevant peptides can be selected. Indeed, the two previously identified PSP were shown to be MHC class I epitopes recognized by CTL of human patients [6,7]. This and the relatively high number of PSP that we identified raises the possibility that peptide splicing in general may lead to an increase in the peptide pool available for epitope selection. For example, from the melanocytic gp100PMEL17 tumor antigen (661 amino acids) 1,786,862 9mers with a unique sequence could be theoretically produced. Of these, a maximum of 652 are unspliced proteasomal cleavage products while the rest (99.96%) represent theoretical PSP. At the moment we do not have any sufficient information to judge on how many of these PSP (as well as normal PCP) are really produced and which percentage of them may efficiently bind MHC class I molecules. Based on our preliminary data we are tempted to speculate that specific PCP are generated more efficiently than PSP even if the MS signal of some PSP (e.g. gp10047–55/35–39) was as high as that of many PCP (data not shown). Nevertheless, if, for example PCP were produced 1000-fold more efficiently than any given PSP, spliced peptides generated from gp100PMEL17 would still represent a significant peptide pool (i.e. the 73.26% of the 9mers derived)

Figure 3. Generation of PSP by proteasomal trans splicing. (A) To demonstrate the generation of a PSP, by the binding of two fragments originated from two distinct molecules of substrate, 5 nmol gp10040–52 and its heavy analogue with amino acids [13C6-Lys and [15N-Leu (RTK)AWNRQLYPEW] were digested together for 36 hours by 1.5 μg LcL 20S proteasomes in 100 μl buffer. Theoretically four different PSP could be generated from the cis or trans ligation of the proteasomal fragments [RTK] and [QLYPEW] with sequences [RTK][QLYPEW]: gp10040–42/47–52-a, [M]+H = 1220.7; gp10040–42/47–52-b, [M]+H = 1221.7; gp10040–42/47–52-c, [M]+H = 1226.7; gp10040–42/47–52-d, [M]+H = 1227.7. (B) LC-MALDI-TOF/TOF-MS spectra at RT = 41.3 min show peaks which can be assigned to all four possible PSP of gp10040–42/47–52 (for MS/MS spectra see Figure S1). doi:10.1371/journal.pcbi.1000830.g003
This basic computational analysis assumes that the splicing of proteasomal cleavage products can occur also in vitro. Our observation that the in vitro splicing reaction not only occurs in cis but also in trans indirectly supports such an assumption. The existence of the trans PSP implies the likely situation that two or more substrate molecules are present at the same time within the proteasomal cavity as suggested by some excellent previous studies [17-19] or that the cleavage products of a first substrate molecule remain within the catalytic chamber while a second molecule of substrate is cleaved. Very recently, Dalet and co-workers investigated trans proteasome splicing in vitro, providing some very interesting albeit not final insights. They showed that PSPtrans were generated in vitro when the precursor peptides of FGF-5 and gp100 were electroporated into COS cells, whereas only the FGF-5-derived PSPtrans (and in a very small amount) could be detected by CTL assay when COS cells were transfected with FGF-5 or gp100 plasmid [8]. Taking into account the high number of PSPtrans we identified within in vitro digestion products of four peptides, we are led to conclude that further studies in vitro and in vivo on different cellular and proteasome models are required to clarify this phenomenon.

An extensive application of SpliceMet on a wide range of polypeptide substrates would also help to identify putative peptide sequence motifs that facilitate the proteasomal splicing reaction. For example, in seven of the nine gp10035–57–derived PSP, the sequence VSR represents the N-terminus of those PCP, which according to the transpeptidation model [6,20] perform a nucleophilic attack on the acyl-enzyme intermediate, thereby forming the detected PSP. Likewise, for four PSP the sequence YPEW represents the C-terminus, which forms the acyl-enzyme intermediate that is subsequently attacked by the second PCP generating the new PSP. From these observations one might infer a higher affinity of these two peptide sequences for a transpeptidation reaction. However, only a more extensive investigation of this specific issue with SpliceMet, covering a large number of different polypeptides would allow to validate such a hypothesis.

For this and other aims, studies performed with the help of SpliceMet could be powered if coupled with algorithms for the prediction of proteasomal cleavages, mathematical modeling of degradation kinetics as well as of the MHC class I antigen presentation [21-26]. Such an approach would also facilitate the reduction of the theoretical PSP number, which might represent a limitation of SpliceMet application to very long proteins such as gp100PMEL17. By combining the SpliceMet results with the estimation of these and other algorithms it would be theoretically possible to restrict the PSP identification to a group of PSP possessing features of interest (e.g. epitope-specific for a defined HLA I haplotype) and to predict their altered expression upon proteasome modification [21].

### Methods

I. Peptides and peptide synthesis

All peptides were synthesized using Fmoc solid phase chemistry as previously described [27]. Exception had to be made for heavy analogues of gp10035–52. The isotope-labeled amino acids 15N-Fmoc-L-Leucine (3eq. amino acid, 3eq. HBTU, 6eq. DIEA in DMF) and L-Lysine-$\gamma$-N-Fmoc, 6-N-T-BOc, 13C6 (1.92eq. amino acid, 1.92eq. HBTU, 3.84eq. DIEA in DMF) were coupled over night.

The sequence enumeration for the peptides gp10040–52, gp10035–57 and gp10031–229 is referred to the human gp100PMEL17 sequence described by Adema and colleagues [28], for the peptide pp8916–40 is referred to the murine cytomegalovirus pp89 protein described by Lyons et al. [29]. The peptide sequence here named gag-pol29–58 is a modified version of the sequence 29–57 of the HIV gap-pol protein as described by Reitz et al. [30], where a Valin was inserted before the Threonin 53. All peptide sequences were extrapolated on the web site http://www.uniprot.org/.

### Table 3. PSP identified in the proteasomal digestion of the polypeptide gp10035–57.

<table>
<thead>
<tr>
<th>Peptide (gp100)</th>
<th>Sequence</th>
<th>Mr, calc</th>
<th>PSP type</th>
<th>Identification step of SpliceMet</th>
</tr>
</thead>
<tbody>
<tr>
<td>49–52/35–39</td>
<td>[YPEW][VSRQL]</td>
<td>1176.59</td>
<td>cis, reverse</td>
<td>6</td>
</tr>
<tr>
<td>49–50/35–37</td>
<td>[YPEW][VSR]</td>
<td>935.45</td>
<td>cis, reverse</td>
<td>6</td>
</tr>
<tr>
<td>47–55/40–42</td>
<td>[QLYPEWTEA][RTK]</td>
<td>1520.76</td>
<td>cis, reverse</td>
<td>6</td>
</tr>
<tr>
<td>47–52/35–37</td>
<td>[OL][YPEW][VSR]</td>
<td>1176.59</td>
<td>cis, reverse</td>
<td>7</td>
</tr>
<tr>
<td>47–48/35–39</td>
<td>[OL][VSRQL]</td>
<td>842.50</td>
<td>cis, reverse</td>
<td>7</td>
</tr>
<tr>
<td>45–52/35–37</td>
<td>[NRQL][YPEW][VSR]</td>
<td>1446.74</td>
<td>cis, reverse</td>
<td>7</td>
</tr>
<tr>
<td>37–38/49–57</td>
<td>[RR][YPEWTEAQR]</td>
<td>1462.70</td>
<td>cis, normal</td>
<td>7</td>
</tr>
<tr>
<td>35–39/35–39</td>
<td>[VSRQL][VSRQL]</td>
<td>1184.70</td>
<td>trans</td>
<td>7</td>
</tr>
</tbody>
</table>

The PSP identified by the application of SpliceMet on the proteasome-mediated digestion of the substrate gp10035–57 are here described. PSPnormal or PSPreverse result from splicing in the same order as the substrate or in reverse order to the substrate, respectively. PSPtrans are derived from the splicing of two non-overlapping sequences of the original substrate. In contrast, PSPtrans necessary originate from two distinct substrate molecules because of the overlapping sequences of the two peptides spliced together.

doi:10.1371/journal.pcbi.1000830.t003
II. Cell cultures

Lymphoblastoid cell lines (LcLs) are human B lymphocytes immortalized with Epstein Barr virus (EBV) which mainly express active immunoproteasomes [13,14]. LcLs were cultured in RPMI1640 medium supplemented with 10% FCS.

III. 20S proteasome purification

20S proteasomes were purified from 3*10^9 LcLs as previously reported [31]. The purity of 20S proteasome preparation was verified by SDS-PAGE electrophoresis (12, 5% poly-acrylamide gel stained with Coomassie dye) (Figure S6). Furthermore, a non-proteasome proteolytic activity of the preparation was tested and excluded (data not shown) by the digestion of 40 μM gp100_35–52 for 24 hours by 1 μg of LcL 20S proteasomes in presence of 400 μM Lactacystin (previously incubated with 20S proteasomes at room temperature for 10 min).

IV. In vitro digestion of synthetic peptide substrates

Synthetic peptides at different concentrations (from 40 to 100 μM) were digested by 0.25–1.5 μg 20S proteasomes in 50–100 μl Hepes buffer (Hepes 20 mM, KCl 1 mM, MgCl2 0.5 mM, DTT 1 mM, NaN3 1 mM, pH 7.3) for different time periods (from 20 min to 48 hours) at 37°C. Digestions were stopped by acidic inactivation and frozen. Digestions were performed also in TEAD buffer (Tris 20 mM, EDTA 1 mM, NaN3 1 mM, DTT 1 mM, pH 7.2) and no remarkable differences compared to Hepes buffer emerged (data not shown). In contrast, for SpliceMet step 7, 1.1 μmol of the peptides (at the final concentration of 100 μM) were digested for 24 hours by 62 μg of LcL 20S proteasomes in 10 ml Hepes buffer and the products up-scaled by RP-HPLC separation. All experiments reported in this study were repeated at least twice and each set of experiments was measured by each MS instrument at least twice.

V. LC-ESI MS

In LC-runs the peptide separation was carried out on a 2.1 mm x 10 cm C2/C18, 3 μm, 120 Å, Amersham) and a 1 mm RP column (Beta Basic-18, 100 mm x 1 mm, 3 μm, 150 Å, ThermoFisher) using a Surveyor system (ThermoFisher Scientific, USA). The mobile phase (A) was 100% water containing 0.05% (v/v) TFA and (B) was 70:30 (v/v) acetonitrile/water containing 0.045% (v/v) TFA or 0.1% acetic acid for the PSP identifications reported in Figure 3. Online MS analysis was performed by DECA XP MAX iontrap instrument (ThermoFisher Scientific, USA) and by LCQ-classic iontrap (ThermoFisher Scientific, USA) after HPLC separation (HP1100, Agilent). MS data were acquired with a triple scan method in positive ion mode (MS - mass range 250–2000 m/z, zoom scan, MS/MS).

To verify that the relatively high PSP number was not peculiar to the sequence gp100_35–57 we extended our investigation to three additional peptides. Six new PSP were identified within their products of in vitro proteasomal digest by applying SpliceMet. Three of them derived from the digestion of the sequence gp100_201–230, one from HIV gag-pol_29–58 and two from MCMV pp89_16–40. All of them were produced by a trans splicing reaction.

Table 4. PSP identified in proteasomal digestions of three additional polypeptides.

<table>
<thead>
<tr>
<th>Substrate</th>
<th>Sequence</th>
<th>Mr, calc</th>
<th>PSP Type</th>
<th>Identification step of SpliceMet</th>
</tr>
</thead>
<tbody>
<tr>
<td>gp100_201–230</td>
<td>[AHSSSAFTITQVPFSVSQRALDGGNK]</td>
<td>1301.60</td>
<td>trans</td>
<td>6</td>
</tr>
<tr>
<td>gag-pol_29–58</td>
<td>[YKLKIVVWASRELFARVNPGLLEVTSEG]</td>
<td>1439.58</td>
<td>trans</td>
<td>7</td>
</tr>
<tr>
<td>pp89_16–40</td>
<td>[RLMYDMYPHPMTNLGPEKR VMS]</td>
<td>1439.58</td>
<td>trans</td>
<td>6</td>
</tr>
</tbody>
</table>

To verify that the relatively high PSP number was not peculiar to the sequence gp100_35–57 we extended our investigation to three additional peptides. Six new PSP were identified within their products of in vitro proteasomal digest by applying SpliceMet. Three of them derived from the digestion of the sequence gp100_201–230, one from HIV gag-pol_29–58 and two from MCMV pp89_16–40. All of them were produced by a trans splicing reaction.

doi:10.1371/journal.pcbi.1000830.t004
of ESI/MS data was accomplished using Bioworks version 3.3 (ThermoFisher Scientific, USA). Database searching was performed using the ProteaJ database and the following parameters: no enzyme, mass tolerance for fragment ions 1 amu. 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). At step 3 of SpliceMet the significant peaks for each theoretical m/z value in the LC-ESI mass chromatogram were identified by Bioworks peak detection algorithm with a signal-to-noise ratio larger than 2 (here = 2).

VI. Digestion product up-scaling by RP-HPLC

Further identification of the PSP at step 7 of SpliceMet was performed by MALDI-TOF/TOF-MS analysis of the gp100 35–57 digestion products separated by two distinct rounds of RP-HPLC. In the first round 57 fractions were collected, lyophilized and analyzed by LC-ESI/MS to identify PSP candidates. The fractions containing the PSP candidates were then separated with more focused gradients (different for each selected fraction of the first round of HPLC separation) on the same column obtaining 47 fractions, which were lyophilized and investigated by MALDI-TOF/TOF-MS analysis. Each round was obtained by collecting the eluted fractions of the 5-15 runs (5–20 μl each) to maintain a good separation of the digestion products on the chromatogram. The runs were carried out on the column C18 (33 x 4.6 mm; ODS1 1.5 μm) by the HPLC Beckman SytemGold and different gradients of acetonitrile.

VII. Nano-LC-MALDI-TOF/TOF-MS

Peptide separation was carried out using an Ultimate HPLC system (Dionex, Idstein, Germany). Samples were concentrated on a trap column (PepMap C18, 5 mm x 300 μm x 5 μm, 100 Å, Dionex) and eluted onto an analytical column (PepMap C18, 150 mm x 75 μm x 3 μm, 100 Å, Dionex). The mobile phase (A) was 2.98 (v/v) acetonitrile/water containing 0.05% (v/v) TFA and (B) was 80:20 (v/v) acetonitrile/water containing 0.045% (v/v) TFA. Runs were performed at a flow rate of 200 nl/min using a binary gradient 0–15% B in 4 min, 15–60% B in 45 min, 60–100% B in 5 min. Column effluent was mixed with MALDI matrix (5 mg/ml α-cyano-4-hydroxy-cinnamic acid in 70:30 (v/v) acetonitrile/water containing 0.1% (v/v) TFA, 1 μl/min) and spotted at ten second intervals on MALDI steel targets using a Probot fractionation device (Dionex). MS analysis was performed on a 4700 Proteomics Analyzer (Applied Biosystems, Framingham, MA, USA). MS data were acquired in positive ion mode in the mass range 800–4000 m/z by accumulation of 1200 laser shots per spot and processed with default calibration. MS/MS spectra were generated by 1 keV collisions and accumulation of 2500 to 10000 laser shots. Analysis of MALDI MS data was accomplished using MASCOT version 2.1 (Matrixscince, London, UK). Database search was performed using ProteaJ database and the following parameters: no enzyme, mass tolerance for precursors, +/- 0.3 Da. Spectral images for manual validation were prepared with Data Explorer Software version 4.8 (Applied Biosystems).

Supporting Information

Figure S1 Verification of the PSP gp100 40–42/47–52 with sequence RTKQLYP EW generated by cis and trans splicing. Found at: doi:10.1371/journal.pcbi.1000830.s001 (0.97 MB TIF)

Figure S2 MS/MS identification of four gp100 35–37 PSP at step 7 of SpliceMet. Found at: doi:10.1371/journal.pcbi.1000830.s002 (0.55 MB TIF)

Figure S3 Identification of three PSP originated from the synthetic substrate gp100 201–290. Found at: doi:10.1371/journal.pcbi.1000830.s003 (0.90 MB TIF)

Figure S4 Identification of the PSP gag-pol 45–57/48–49. Found at: doi:10.1371/journal.pcbi.1000830.s004 (0.34 MB TIF)

Figure S5 Identification of two PSP originated from the synthetic substrate pp89 10–40. Found at: doi:10.1371/journal.pcbi.1000830.s005 (0.63 MB TIF)

Figure S6 SDS-PAGE Electrophoresis with 20S proteasome purified from LcLs. Found at: doi:10.1371/journal.pcbi.1000830.s006 (0.95 MB TIF)

Acknowledgments

We thank Agathe Nievienda, Elena Bellavista, Eberhard Krause and Heike Stephanowitz for their excellent technical assistance and supervision, Sascha Bulik for the estimation of the PSP number and Hermann-Georg Holzhuetter for inspiring discussions.

Author Contributions

Conceived and designed the experiments: JL MM PMK AZ. Performed the experiments: JL MM. Analyzed the data: JL MM KTT KJ CK. Contributed reagents/materials/analysis tools: PH. Wrote the paper: JL MM PK AZ.

References


Copyright of PLoS Computational Biology is the property of Public Library of Science and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.