Immunogenetics (2003) 55:437–449
DOI 10.1007/s00251-003-0585-6

ORIGINAL PAPER

Can Kesmir · Vera van Noort · Rob J. de Boer · Paulien Hogeweg

Bioinformatic analysis of functional differences between the immunoproteasome and the constitutive proteasome

Received: 27 March 2003 / Revised: 16 June 2003 / Published online: 30 August 2003

Abstract Intracellular proteins are degraded largely by proteasomes. In cells stimulated with gamma interferon , the active proteasome subunits are replaced by "immuno" subunits that form immunoproteasomes. Phylogenetic analysis of the immunosubunits has revealed that they evolve faster than their constitutive counterparts. This suggests that the immunoproteasome has evolved a function that differs from that of the constitutive proteasome. Accumulating experimental degradation data demonstrate, indeed, that the specificity of the immunoproteasome and the constitutive proteasome differs. However, it has not yet been quantified how different the specificity of two forms of the proteasome are. The main question, which still lacks direct evidence, is whether the immunoproteasome generates more MHC ligands. Here we use bioinformatics tools to quantify these differences and show that the immunoproteasome is a more specific enzyme than the constitutive proteasome. Additionally, we predict the degradation of pathogen proteomes and find that the immunoproteasome generates peptides that are better ligands for MHC binding than peptides generated by the constitutive proteasome. Thus, our analysis provides evidence that the immunoproteasome has co-evolved with the major histocompatibility complex to optimize antigen presentation in vertebrate cells.

C. Kesmir (✉) · V. van Noort · R. J. de Boer · P. Hogeweg
Theoretical Biology/Bioinformatics,
Utrecht University,
Padualaan 8, 3584 CH Utrecht, The Netherlands
e-mail: C.Kesmir@bio.uu.nl
Tel.: +31-30-2533 637
Fax: +31-30-2513655

C. Kesmir
Center for Biological Sequence Analysis, BioCentrum-DTU,
Technical University of Denmark Lyngby, Denmark

V. van Noort
Center for Molecular and Biomolecular Informatics,
Nijmegen Center for Molecular Life Sciences,
Nijmegen, The Netherlands

Keywords Proteasome · MHC class I epitopes · Antigen-processing, presentation · Specificity · Co-evolution

Introduction

The proteasome, an ATP-dependent, multi-subunit protease, plays the central role in intracellular protein degradation (Stoltze et al. 2000a; Kloetzel 2001; Yewdell and Bennink 2001). After protein substrates have been degraded, the resulting peptide fragments are translocated from the cytoplasm to the endoplasmic reticulum and then loaded onto major histocompatibility complex (MHC) class I molecules (Rock and Goldberg 1999). The recognition of the peptide-MHC complex on the cell surface by a cytotoxic T cell (CTL) causes lysis of the cell. This cellular response plays a crucial role in eliminating intracellular pathogens.

The eukaryotic proteasome is a complex formed by regulatory units and one cylindrical enzymatic chamber, the 20S proteasome. The 20S proteasome consists of 14 different protein subunits (Groll et al.1997), of which only three have an active site (Groll et al. 1997, 1999; Heinemeyer et al. 1997; Tanaka and Kasahara 1998). The activity of the proteasome in inflammatory sites is altered via induction of the regulatory units and replacement of the constitutive active subunits (β-1[δ,Y], β-2[MC14, LMP9,Z] and β-5[MB1,X]) by their immuno (β-1i[LMP2], β– 2i[MECL-1] and β-5i[LMP7]) counterparts (Groettrup et al. 1996; Tanaka and Kasahara 1998). Thus two forms of proteasome exist: the "immunoproteasome", which is expressed in cells stimulated by gamma interferon (IFN-γ) or tumor necrosis factor alpha (TNF-α), and in primary and secondary lymphoid organs, and the "constitutive proteasome", which is expressed in healthy, normal tissues and in immune-privileged organs such as the brain (Dahlmann et al. 2000; Noda et al. 2000; Kuckelkorn et al. 2002). During an antiviral or antibacterial immune response, immunoproteasomes largely replace constitutive proteasomes (Khan et al. 2001). This

replacement has a positive effect on MHC class I restricted antigen presentation, as has been demonstrated in several systems (see, for example, Kuckelkorn et al. 1995; Ehring et al. 1996; Morel et al. 2000; Sijts et al. 2000b; Van Hall et al. 2000; Chen et al. 2001; Khan et al. 2001; Schultz et al. 2002). The immunoproteasomes are not absolutely necessary to generate immunogenic epitopes, but immunodominant epitopes are mainly generated by the immunoproteasomes (Van Hall et al. 2000).

Phylogenetic analysis of the active subunits reveals that the immunosubunits evolve faster than their constitutive counterparts (see Hughes 1997 and our results below). This suggests a possible functional differentiation between the constitutive subunits and their immuno counterparts. Although different forms of proteasomes have been recognized for some time and have been studied extensively in vitro, the degradation data available does not give a clear picture of the specificity of different forms of the proteasome. The immunoproteasome has a greater capacity to cleave after hydrophobic and basic residues and a lower capacity to cleave after acidic residues. Consequently, the immunoproteasome should generate more MHC ligands. However, it appears that the induction of the immunoproteasome can also abrogate the presentation of antigens (Morel et al. 2000) (for a review on this subject see Van den Eynde and Morel (2001). Some epitopes are generated by both proteasomes (see, for example, Sun et al. 2002). Thus, the limited number of epitopes/proteins that are studied do not provide (1) a clear and quantitative picture of the specificity of the two proteasomes, (2) the effect of the specificity differences on a genome level (both for self and non-self) and (3) direct evidence that the immunoproteasome actually generates more antigenic peptides. Here we use a bioinformatics analysis to quantify the differences between the two forms of proteasomes based on the degradation of yeast enolase protein by human constitutive proteasome and immunoproteasome (Toes et al. 2001). By using well-defined measures for diversity and information content, we are able to account for the small sample size and the biases that might rise from the choice of epitopes/proteins. Comparing the specificity of the two types of proteasome, we are able to quantify that the immunoproteasome is more specific than the constitutive proteasome. Such an analysis allows us to predict the consequences of the specificity divergence at the pathogen proteome level, instead of being limited to one or a few epitopes (as is the case in experimental systems). The predictions of degraded peptides on a large data set of pathogen proteins suggest that the specificity of the immunoproteasome has co-evolved with the specificity of the MHC molecules at the C-terminal, but not at the N-terminal.

Materials and methods

Phylogenetic analysis

Complete protein sequences of the subunits were used in the phylogenetic study. The protein sequences were obtained from NCBI (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi) using a search based on annotation. To complete the data set a BLAST search was performed (http://www.ncbi.nlm.nih.gov/blast/, Altschul et al. 1990), because the annotation is not very consistent among organisms. The accession numbers of the proteasome sequences are given in Table 1.

Multiple alignments were made using ClustalW, and the Gonnet series was used to compute distances between sequences. Phylogenetic trees were constructed using the Neighbor Joining method. The numbers of synonymous and non-synonymous substitutions between two DNA sequences were calculated using the method of Yang and Nielsen (2000). This is an approximate method, involving three steps: counting synonymous and non-synonymous sites in the two sequences, counting synonymous and non-synonymous differences between the two sequences, and correcting for multiple substitutions at the same site. The method takes into account two major features of DNA sequence evolution: the transition/transversion rate bias and the base/codon frequency bias.

Information and diversity measures

The Shannon information content (Shannon1948) at position i, $I(i)$, in an alignment is defined as $I(i) = \log_2 20 + \sum_{L=1}^{20} p_i^L \log_2 p_i^L$, where p_i^L is the probability that a particular amino acid L occurs at position i in the alignment. The unit of $I(i)$ is bits per amino acid. The maximum information content is $log_2 20$=4.3 and it is obtained if at position i always the same amino acid is observed. The minimum Shannon information is zero and it is obtained if all amino acids occur with same frequency at position i. The Kullback-Leibler information content, $K(i) = \sum_{L=1}^{20} p_i^L \log_2 \left(p_i^L / q_i^L\right)$, where p_i^L is the probability that the amino acid L will occur at position i, q_i^L is the probability that a particular amino acid L will occur e.g., in natural proteins (background distribution). In this way the information content is corrected for the distribution of amino acids in natural proteins. To measure the diversity at position i in an alignment we define $D(i)$=1/$SI(i)$, where $SI(i)$ is the Simpson index, $SI(i) = \sum_{L=1}^{20} \left(p_i^L\right)^2$. If position i is fully conserved, e.g., if it is always a valine, then $D(i)$=1, if position i is fully degenerate where each amino acid occurs with same frequency, $D(i)$=20 . In all cases the probabilities satisfy $\sum_{L=1}^{20} p_i^L = 1$.

Viral and human proteins

We compiled SwissProt version 39 [http://www.expasy.org/sprot/sprot-top.html, Bairoch and Apweiler (2000)] to build a representative set of human proteins and human viruses. The final data sets consisted of 6,041 human proteins and 1,304 viral proteins (excluding HIV). Neither of the sets was homology reduced. The total number of amino acids in the human set was 2.9×10^6, giving a maximum of 2.85×10^6 9mers. The total number of amino acids in the pathogen set was 5.3×10^5, giving a maximum of 5.2×10^5 9mers.

Prediction of cleavage sites

To predict cleavage sites in proteins we use amino acid frequencies occurring at and around the cleavage sites during degradation of yeast enolase by the immunoproteasome and the constitutive proteasome, as given in Toes et al. (2001). In this paper we adapted the data for the immunoproteasome cleavages in Table 2. For each position i in a protein we define the probability that a position will be used by the immunoproteasome as a cleavage site (i.e., as a $P1$

Table 1 Accession numbers of proteasome subunit sequences used in phylogenetic analysis. Individual subunits are named using homologous subunit names in human. We use the nomenclature suggested by Baumeister et al. (1998); for some subunits the old nomenclature or original annotation is given in parenthesis. Abbreviations used in the phylogenetic trees are given after the organism names

Protein sequences

Bos taurus (Bota)	
β-5i (LMP7)	AAF22654
β-1i (LMP2)	AAD28184
Botryllus schlosseri (Bsc)	
β-5 (ε)	JC5074
Danio rerio (Dr)	
β-5 (β-5)	AAD53518
β-5A (X)	AAB87680
β-2 (β-7)	AAD53521
β-1iA (β-9A)	AAD53519
β-1iB (β-9B)	AAD53520
β-1iC (β-11)	AAD53516
β-2i (β-12)	AAD53517
β-1 (Y)	AAB87681
β-5i (LMP7)	AAB87679
Dictyostelium discoideum (Didi)	
β-1	JE0101
Drosophila melanogaster (Dm)	
β-2	AAB82570
β-5	AAF26416
β-1	AAF58077
Gallus gallus (Gaga)	
β-5	P34065
Geodia cydonium (Gcy)	
β-5 (ε)	JC5073
Ginglymostoma cirratum (Gici)	
β-5 (LMPX)	BAA10935
β-5iA (LMP7)	BAA10933
β-5iB (LMP7)	BAA10934
Homo sapiens (Hs)	
β-1 (Y)	P28072
β-2 (Z)	Q99436
β-5 (X)	P28074
β-1i (LMP2)	P28065
β-2i (MECL-1)	P40306
β-5i (LMP7)	P28062
Lampetra japonica (Laja)	
β-1 (Y)	BAA19761
Mus musculus (Mm)	
β-1 (β-6)	NP_032972
β-2 (β-7)	NP_035317
β-5	NP_035316
β-1i (β-9)	O35522
β-2i (β-10)	NP_038668
β-5i (β-8)	NP_034854
Myxine glutinosa (Mygl)	
β-5 (LMPX)	BAA10931
Oncorhynchus mykiss (Om)	
β-1 (δ)	AAD53036
β-1i (LMP2)	AAD53038
β-1i-like (LMP2)	AAD53037
Petromyzon marinus (Pema)	
β-5 (LMPX)	BAA10932
Rattus norvegicus (Rn)	
β-5 (ε)	P28075
β-1 (δ)	JX0228
β-2 (Z)	AAF97811
β-5i (C1)	S21126
β-1i (LMP2)	NP_036840
Saccharomyces cerevisiae (Sc)	
β-5 (Pre2)	P30656
β-1 (Pre3)	P38624
β-2 (Pup1)	P25043
Schizosaccharomyces pombe (Sp)	
β-5 (Pts1)	JS0753
β-1	T40487
β-2 (β-7)	Q09841
Sus scrofa (Susc)	
β-5i (LMP7)	AAD22390
Trypanosoma brucei (Tb)	
β-5	AAF37285
β-1	CAA10283
β-2	CAA10208
Trypanosoma cruzi (Tc)	
β-5	AAC97957
Xenopus laevis (Xl)	
β-1 (Y)	BAA19760
β-1i (LMP2)	BAA19759
β-5iA (LMPA)	I51536
β-5iB (LMPB)	I51537

Nucleotide sequences

Homo sapiens (Hs)	
α-1	X61972
α-2	NM_002787
α-3	NM_002789
α-4	NM_002792
α-5	NM_002790
α-6	NM_002786
α-7	NM_002788
β-1	BC000835
β-2	BC000509
β-3	NM_002795
β-4	NM_002794
β-5	NM_002797
β-6	NM_002793
β-7	NM_002796
β-1i	NM_002800
β-2i	NM_002801
β-5i	NM_148919
Mus musculus (Mm)	
α-1	NM_011968
α-2	XM_135563
α-3	NM_011966
α-4	NM_011969
α-5	NM_011967
α-6	AF060088
α-7	NM_011184
β-1	NM_008946
β-2	NM_011187
β-3	NM_011971
β-4	NM_011970
β-5	NM_011186
β-6	NM_011185
β-7	U65636
β-1i	NM_013585
β-2i	NM_013640
β-5i	NM_010724

Table 2 The frequency of amino acids found in positions P6 to P1 and P1′ to P6′ of peptides generated by the immunoproteasome. Adapted from Toes et al. (2001). Similar data are available for the constitutive proteasome (not shown). Each column adds up to one

	P6	P5	P4	P3	P2	P1	P1′	P2′	P3′	P4′	P5′	P6′
A	0.10	0.09	0.14	0.06	0.09	0.11	0.22	0.10	0.08	0.12	0.07	0.12
C	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
D	0.11	0.05	0.02	0.03	0.02	0.02	0.05	0.18	0.04	0.05	0.03	0.05
E	0.04	0.05	0.08	0.13	0.09	0.02	0.04	0.05	0.06	0.04	0.06	0.08
F	0.00	0.07	0.04	0.01	0.10	0.08	0.00	0.01	0.01	0.04	0.02	0.03
G	0.07	0.08	0.07	0.02	0.05	0.02	0.13	0.07	0.15	0.04	0.07	0.10
H	0.04	0.00	0.01	0.04	0.03	0.02	0.02	0.00	0.00	0.00	0.05	0.04
I	0.06	0.05	0.04	0.07	0.02	0.09	0.00	0.12	0.04	0.06	0.11	0.06
K	0.01	0.07	0.06	0.16	0.09	0.02	0.07	0.04	0.04	0.06	0.09	0.13
L	0.08	0.15	0.07	0.03	0.02	0.36	0.01	0.05	0.16	0.06	0.01	0.08
M	0.00	0.00	0.03	0.00	0.01	0.00	0.04	0.00	0.00	0.01	0.00	0.00
N	0.03	0.04	0.04	0.02	0.03	0.01	0.08	0.04	0.08	0.09	0.02	0.04
P	0.05	0.05	0.04	0.05	0.00	0.00	0.05	0.02	0.04	0.02	0.06	0.01
Q	0.00	0.04	0.03	0.01	0.02	0.00	0.02	0.01	0.03	0.03	0.01	0.00
R	0.05	0.01	0.01	0.03	0.06	0.03	0.07	0.00	0.02	0.04	0.05	0.01
S	0.08	0.03	0.09	0.03	0.11	0.02	0.05	0.05	0.04	0.10	0.05	0.08
T	0.13	0.03	0.01	0.03	0.08	0.03	0.04	0.07	0.04	0.03	0.04	0.00
V	0.04	0.07	0.09	0.14	0.10	0.05	0.01	0.08	0.07	0.08	0.14	0.09
W	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.01	0.00	0.00
Y	0.01	0.02	0.06	0.03	0.00	0.04	0.02	0.02	0.01	0.04	0.03	0.00

residue) as: $P_i(P1) = \prod_{k=P6}^{P6'} I_k$, where I_k is the frequency of the amino acid at position k in enolase digests by the immunoproteasome and is given in Table 2. Similarly, $P_c(P1) = \prod_{k=P6}^{P6'} C_k$, indicates the probability that a site will be used by the constitutive proteasome. A similar table to Table 2 is given in Toes et al. (2001) for the C_k values.These definitions are based on the assumption that each amino acid in the peptide contributes independently to the cleavage probability. We corrected the I_k and C_k values for the amino acid distribution in yeast enolase, and then ranked each position in 6,041 human and 1,304 pathogen proteins (compiled from Swissport version 39, see above) according to their P_i and P_c values. In the enolase protein 28.1% and 32.6% of positions are used by the immunoproteasome and the constitutive proteasome, respectively. Thus, it is predicted that the positions with P_i values falling in the top 28% will be used by the immunoproteasome as cleavage sites and that the positions with P_c values falling within the top 33% will be possible constitutive proteasome cleavage sites.

MHC binding predictions

The amino acid weight matrices of *HLA-A*0201*, *HLA-A*0205*, and *HLA-A*3101* were downloaded from http://bimas.dcrt.nih.gov/molbio/hla_bind/ (April 2002). These three alleles were chosen because their weight matrices were sufficiently detailed to allow for an almost continuous ranking of the peptides. In the weight matrices, dominant anchor residues have the highest weights, followed by auxiliary anchor residues, and favorable amino acids. Unfavorable amino acids have a weight of less than one. The binding score of each MHC–peptide combination is defined as $\prod_{i=1}^{9} w_i$, where w_i is the weight index of the amino acid at position i in the peptide. This definition is based on the assumption that each amino acid in the peptide contributes independently to the binding score.

Results

The immunosubunits evolve faster than their constitutive counterparts

The first gene duplication giving rise to the general structure of the proteasome in eukaryotes probably occurred prior to the divergence of archaebacteria and eukaryotes (Hughes 1997; Wollenberg and Swaffield 2001). Thereafter, gene duplications occurring prior to the divergence of hagfish and lamprey from jawed vertebrates resulted in immunosubunits (Hughes1997). We performed a phylogenetic analysis to compare evolutionary traits of different eukaryotic proteasome subunits. The results of this analysis are shown in Fig. 1. The immunosubunits accumulated more mutations than the constitutive counterparts, in agreement with an earlier study by Hughes (1997). There are two possible explanations: (1) the immunosubunits are encoded in a region that has higher mutation rates, i.e., they are encoded in a mutational "hotspot", or (2) the immunosubunits change faster than their constitutive counterparts. Next, we calculated the rates of synonymous (ds) and non-synonymous (dn) nucleotide substitution per site using the method of Yang and Nielsen (2000). These values are given in Table 3. Even though the immunosubunits are encoded in mutational hotspots, the ds value is not significantly different between the constitutive subunits and the immunosubunits ($P>0.09$ on Welch Two Sample t-test). However, the dn value for immunosubunits is significantly higher than it is for constitutive active ($P=0.02$) and inactive ($P=0.009$) β subunits and for α subunits ($P=0.01$). This results in higher dn/ds ratios for the immunosubunits (Table 3) and suggests that the immunosubunits are changing faster. The subunits that do not have enzymatic activity have lower dn/ds ratios. The immunosubunits are nevertheless not the fastest evolving

Table 3 dn/ds ratio of different proteasome subunits (*SU*) from mice and men. Only the subunits β-1,β-2 and β-5 and their immuno counterparts are active. For comparison, the dn/ds ratio is also calculated for a house-keeping gene, phosphatase subunit six, accession numbers NM_002721 and BC002223, and two MHC class I molecules, *HLA-A*02011* with accession number HLA00005 and *HLA-A*01011* with accession number HLA00001 from the IMGT/HLA (imgt.cines.fr) [Robinson et al. 2001] database. To test whether dn and ds values are significantly different, we grouped the subunits as the following: all α subunits, active α subunits (constitutive), inactive β subunits (constitutive) and immuno subunits. A Welch Two Sample t-test between all groups show that none of the groups have significantly different ds values ($P>0.09$). Immuno subunits have higher dn values from active constitutive subunits ($P=0.023$) and from inactive constitutive subunits ($P=0.009$). It is remarkable that given the small sample size (3 and 4), the P-values are very low, which suggests that the difference in the evolutionary traits of constitutive and immunosubunits is very convincing

SU	dn	ds	dn/ds	SU	dn	ds	dn/ds	SU	dn	ds	dn/ds
α-1	0.002	0.287	0.006	β-1	0.033	0.400	0.083	β-1i	0.091	0.489	0.186
α-2	0.004	0.351	0.012	β-2	0.039	0.373	0.105	β-2i	0.091	0.660	0.138
α-3	0.005	0.447	0.011	β-3	0.019	0.391	0.048	-	-	-	-
α-4	0.009	0.406	0.021	β-4	0.025	0.257	0.098	-	-	-	-
α-5	0.007	0.252	0.028	β-5	0.047	0.398	0.119	β-5i	0.130	0.750	0.170
α-6	0.03	0.327	0.090	β-6	0.062	0.581	0.106	-	-	-	-
α-7	0.014	0.362	0.039	β-7	0.032	0.432	0.075	-	-	-	-
Average	0.010	0.350	-	-	0.037	0.404	-	-	0.103	0.630	-
	0.009	0.060	-	-	0.012	0.110	-	-	0.017	0.110	-
MHC	0.069	0.049	1.418	-	-	-	-	-	-	-	-
Phosphatase 6	0.001	0.359	0.004	-	-	-	-	-	-	-	-

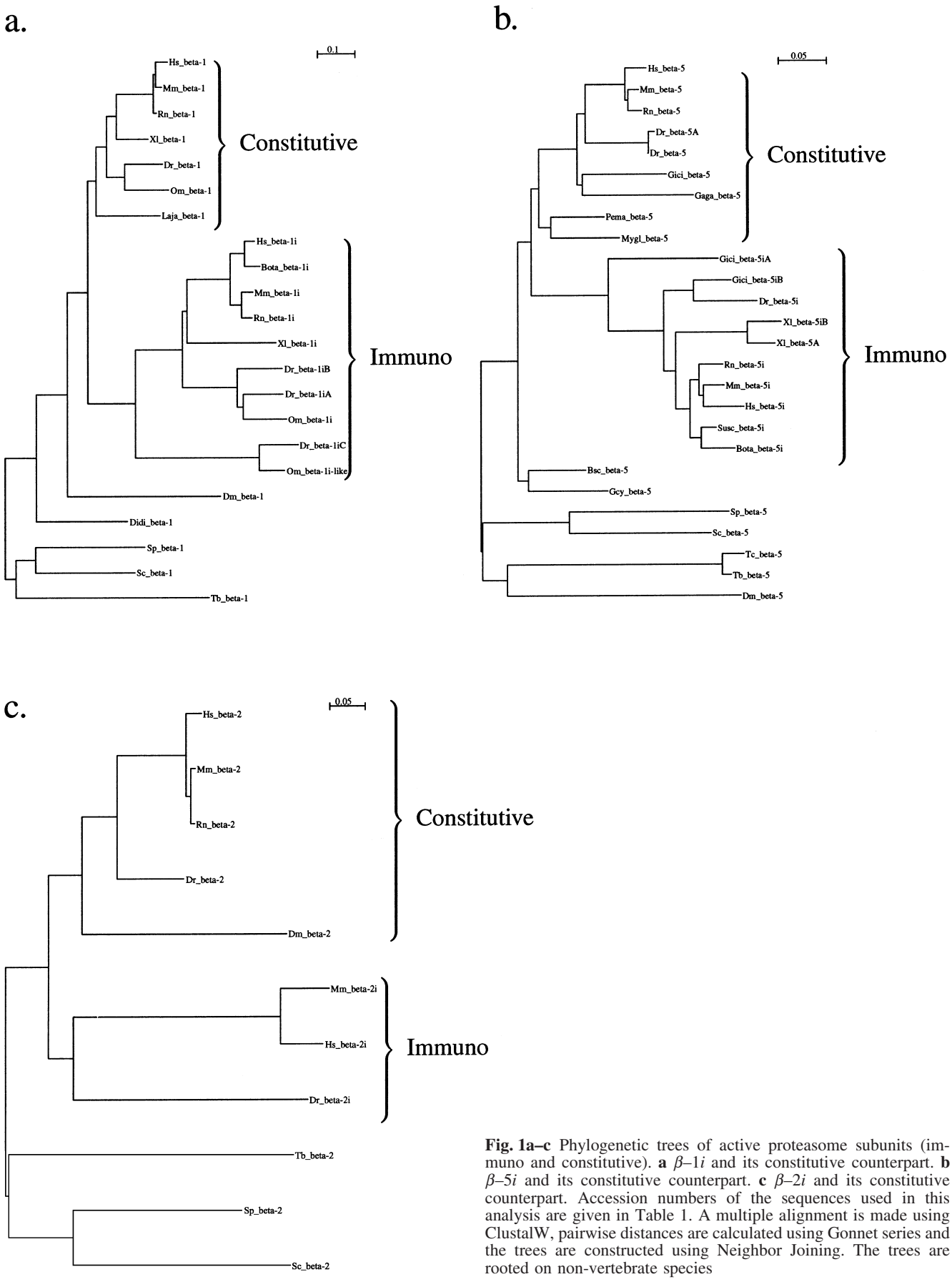

Fig. 1a–c Phylogenetic trees of active proteasome subunits (immuno and constitutive). **a** *β*–1*i* and its constitutive counterpart. **b** *β*–5*i* and its constitutive counterpart. **c** *β*–2*i* and its constitutive counterpart. Accession numbers of the sequences used in this analysis are given in Table 1. A multiple alignment is made using ClustalW, pairwise distances are calculated using Gonnet series and the trees are constructed using Neighbor Joining. The trees are rooted on non-vertebrate species

molecules involved in antigen processing and presentation; MHC molecules evolve even faster (Hughes and Nei 1988, 1989; Tanaka and Nei 1989; and Table 3]. In summary, this phylogenetic analysis points to a functional differentiation between the immunoproteasome and the constitutive proteasome.

The immunoproteasome is more specific than the constitutive proteasome

To pinpoint functional differences between immunoproteasomes and constitutive proteasomes, we use the degradation data of Toes et al. (2001). In short, proteasomes were purified from EBV-transformed B cells. Constitutive proteasomes were isolated from a cell line lacking LMP-2 (β-1i) and LMP-7 (β-5i). Lack of LMP-2 results in a defect to incorporate MECL-1 (β-2i) in the proteasomes; thus this cell line does not have any immunoproteasomes. The immunoproteasomes were isolated from cell lines that were capable of expressing and incorporating all three immunosubunits. The yeast enolase protein was incubated separately with immunoproteasome and the constitutive proteasome at a molar ratio of 150:1. The digestion reaction was stopped when 50% of the substrate was degraded. To calculate how often a possible site was actually used, the absolute amount of each peptide detected was determined by Edman sequencing. The details of the experimental procedure are described in Toes et al. (2001).

Toes et al. (2001) calculated the frequency distributions of amino acids at cleavage sites and their flanking regions (Table 2). The cleavage occurs between the P1 and the P1′ position (Berger and Schechter 1970). It has been suggested that the P1 position is the most important position determining cleavage (Altuvia and Margalit 2000; Cascio et al. 2001), although the flanking region may also be important (Ossendorp et al. 1996; Beekman et al. 2000; Mo et al. 2000). One way of analyzing such data is to calculate the information content at the cleavage site and in the region flanking the cleavage site. We use two different information measures. The first one is the Shannon information, which is a measure of to what extent a position in a sequence is conserved (Shannon 1948). The second information measure is the Kullback-Leibler information, which identifies by how much the observed distribution differs from the background distribution. In other words this measure corrects the information content for the distribution of amino acids in natural proteins. Both information measures are defined in Materials and methods.

The information content for the P1 position in the immunoproteasome digests is much higher than in the constitutive proteasome digests (Fig. 2a, b). In other words cleavage by the immunoproteasome is restricted to fewer amino acids, whereas the constitutive proteasome is more degenerate, i.e., many different amino acids can be used as potential cleavage sites. Thus, these results suggest that the immunoproteasome is more specific than the constitutive proteasome. Another measure for comparing the specificity of the two forms of the proteasome is the diversity of amino acids present at the P1 position, as defined in terms of the Simpson index (see Materials and methods). This diversity measure yields a value between 1 and 20; the higher the diversity the more degenerate the proteasome (i.e., 20 means that all amino acids are used with equal frequency at a position, and 1

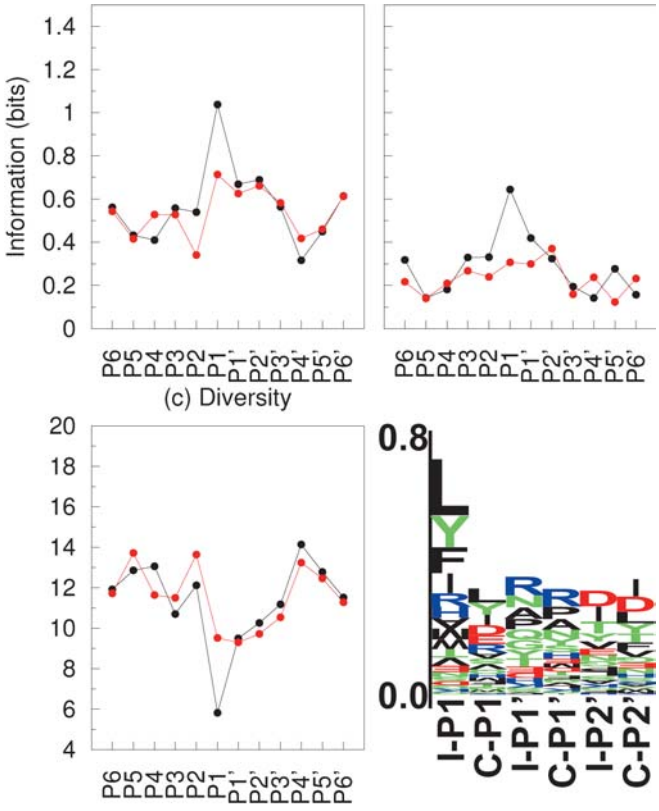

Fig. 2 Information content and diversity at and around the cleavage sites of the immunoproteasome (*black circles*), and the constitutive proteasome (*red circles*). The figure is generated using the frequency data of Toes et al. (2001). *Upper left panel* represents the Shannon information, *upper right panel* represents the Kullback-Leibler information, and *lower left panel* represents diversity (all defined in Materials and methods). *Lower right panel* is a sequence logo for the P1 position. This sequence logo has been corrected for the fact that some amino acids are found frequently or rarely in enolase protein, i.e., it is based on Kullback-Leibler information. As a result some amino acids are represented as upside down letters indicating that these occur less frequently than in the enolase. *I-P1* is the P1 position of the immunoproteasome digests, and *C-P1* stands for the P1 position of the constitutive proteasome digests (similarly for P1′ and P2′ positions). Amino acids are color coded according to their physicochemical characteristics. Neutral and polar, *green*; basic, *blue*; acidic, *red*; neutral and hydrophobic, *black*. One can also define a distance measure to quantify the difference in specificity of the immunoproteasome at different positions. Let $D(i,j) = \sqrt{\sum_{L=1}^{20} \left(p_i^L/q_i^L - p_j^L/q_i^L\right)^2}$ be the distance between position i and j, where p_i^L and q_i^L are observed and background amino acid frequencies for position i, respectively as defined in Materials and methods. The distance between the immunoproteasome and constitutive proteasome specificity is 3.4 at P1 position and 0.9 at P1′ and P2′ positions

means that only a single amino acid is found at a position). The diversity of the P1 position of the immunoproteasome digests is 5.83, and that of the constitutive proteasome is 9.53 (Fig. 2c), again suggesting that the immunoproteasome is more specific. Note that the diversity of the enolase protein is 18.45 averaged over each position.

The two proteasomes are also differentiated on the basis of the preferred amino acids at the P1 position. In Fig. 2d we show the distribution of amino acids at the P1 position in the form of a sequence logo. A sequence logo (Schneider and Stephens 1990) is a clear visualization of (1) to what extent a position in a sequence is conserved (given by the height of a bar, the information content) and (2) which amino acids are most frequently found at a particular position (the height of each amino acid in the logo is proportional to the frequency of occurrence at that position). The constitutive proteasome seems to use the acidic amino acids D and E more than one might expect from their distributions in the enolase protein. In the sequence logo of the P1 position of the immunoproteasome (Fig. 2d), D and E are drawn upside down, indicating that these amino acids are used less frequently than expected according to their occurrence in the enolase protein. Thus, the immunoproteasome makes considerably less use of the acidic amino acids than does the constitutive proteasome. This is in agreement with earlier experimental data (Eleuteri et al. 1997; Cardozo and Kohanski 1998). At positions P1′ and P2′, however, the two forms of proteasomes have hardly differentiated, see Fig. 2d; distances are reported in the figure caption. Since immunosubunits are the result of gene duplications occurring prior to the divergence of hagfish and lamprey from jawed vertebrates, this analysis suggests that the selection pressure that caused the differentiation of immuno and constitutive subunits was in operation only for P1 position.

We can repeat the analysis without taking into account the frequency with which fragments are produced, i.e., we can only look at the observed cleavage sites. According to the cleavage maps reported in Toes et al. (2001), out of 436 residues in enolase, 55 sites are used by the immunoproteasome only, 73 by the constitutive proteasome only and 69 by both proteasomes. In other words, the fraction of sites used by the immunoproteasome is (55+69)/436=0.28 and by the constitutive proteasome (73+69)/436=0.33. The expected value of the overlap then becomes 0.28×0.33×436=40 sites, i.e., the observed overlap is larger than the expected value. The large overlap was also observed in an independent study (Peters et al. 2002). The 55 sites used exclusively by the immunoproteasome are clearly more specific than the 73 sites used only by the constitutive proteasome as demonstrated by the high information content, and by the low diversity at the P1, P2 and P5′ positions used by the immunoproteasome, see Fig. 3a–c. In Fig. 3a, b, the Kullback-Leibler information is much lower than the Shannon information for the cleavage sites that are used by both proteasomes. This means that the overlapping

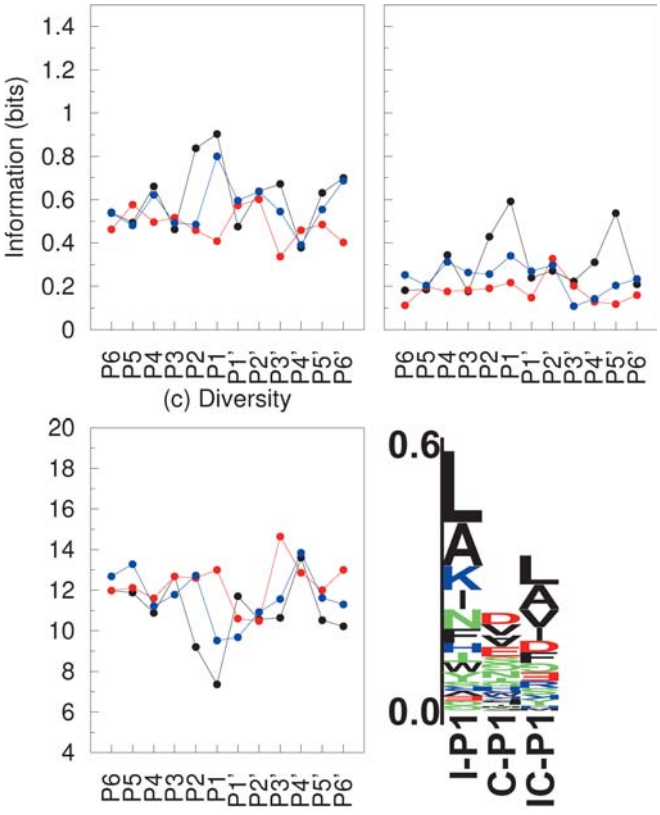

Fig. 3 Information content and diversity at and around the observed cleavage sites used only by the immunoproteasome (*black circles*), or only by the constitutive proteasome (*red circles*) or by both (*blue circles).* The figure is generated using the cleavage maps reported in Toes et al. (2001). *Upper left panel* represents the Shannon information, *upper right panel* represents the Kullback-Leibler information and *lower left panel* represents diversity, all defined in Materials and methods. In addition to the P1 position, the P2 and P5′ positions of the immunoproteasome digests are specific. *Lower right panel* is a sequence logo for the P1 position made using the Kullback-Leibler information. I-P1 are the P1 positions used only by the immunoproteasome, C-P1 are the P1 positions used only by the constitutive proteasome and IC-P1 are the P1 positions used by both forms of proteasome. For clarity P1′ and P2′ positions are not shown, but they follow the same trend as in Fig. 2, i.e., the immunoproteasome and the constitutive proteasome have similar specificities at P1′ and P2′ positions. The color coding of amino acids is as in Fig. 2

sites consist mainly of frequent amino acids that do not carry much information, i.e., the overlapping sites are not very specific. The sites used only by the constitutive proteasome contain hardly any motifs, which is consistent with the low information content and the high diversity of amino acid usage. This suggests that the constitutive proteasome uses semi-specific and degenerate sequence signals to cleave a protein. Again, the acidic amino acids are hardly used by the immunoproteasome (Fig. 3d) and the different amino acid preference is observed only for P1 position (results not shown), i.e., at P1′ and P2′ positions the specificity of the immunoproteasome and the constitutive proteasome is similar. The frequency of amino acids at the 239 sites that are used by neither of the proteasomes hardly deviates from the average amino acid

frequency in enolase (results not shown). Thus, the sites that are not used for cleavage do not contain information about sequence motifs inhibiting proteasomal cleavage.

Since all these results suggest that immunoproteasome is a more specific enzyme, one would expect the immunoproteasome to use fewer cleavage sites in a given protein. Therefore, it is remarkable that when enolase was degraded with immunoproteasomes and constitutive proteasomes, approximately the same number of cleavage sites were observed (similar results were obtained for degradation of ovalalbumin Cascio et al. 2001). One reason for this is that the immunoproteasome uses leucine, which is a very abundant amino acid, much more frequently than the constitutive proteasome. Therefore, the immunoproteasome seems to be able to degrade proteins efficiently despite its increased specificity and as a result of this the replacement of the constitutive subunits by immuno subunits does not inhibit cellular growth or viability (Groettrup et al. 2001a). This is important since the cells expressing the immunoproteasome transiently, or for longer periods, need to maintain the necessary housekeeping operations.

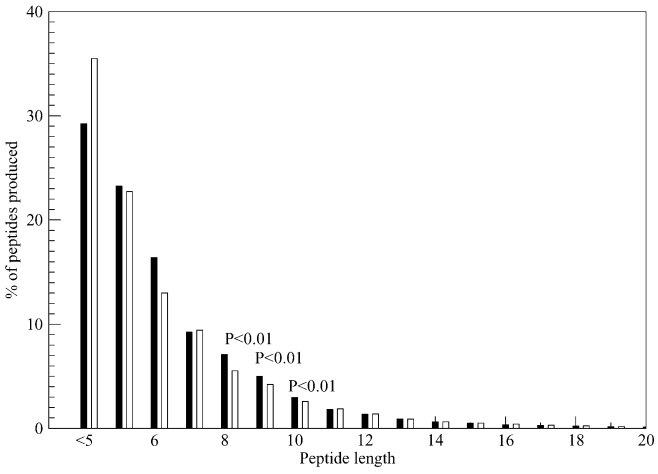

Fig. 4 Fragment distribution of 1,304 viral proteins from SwissProt (version 39) according to the predicted degradation by the immunoproteasome (*filled bars*) and the constitutive proteasome (*open bars*). The prediction method is explained in Materials and methods. The length of a predicted peptide is taken as the distance between two consecutive predicted cleavage sites. A chi-square test was performed to find out if the number of peptides predicted to be generated by the two forms of the proteasomes are different. The individual confidence intervals are given in the figure as *P* values for peptides of length eight, nine, and ten amino acids

The immunoproteasome generates more MHC ligands

The above data analysis gives a clear indication of functional differences between the two forms of proteasome. Going one step further, one can analyze the general pattern of the peptides that are expected to be generated from human and/or pathogen proteomes by each form of proteasome. We predict the products of the constitutive proteasome and the immunoproteasome using a simple algorithm based on the frequency data of Table 2 (see Materials and methods).

Throughout the paper we take the length of a predicted peptide as the distance between two consecutive predicted cleavage sites. Fig. 4 depicts the distribution of lengths of predicted fragments from viral proteins. Similar distributions are obtained for human proteins. Often peptides are produced by the proteasome with an extension on their amino-termini (Cascio et al. 2001). Aminopeptidases in the cytosol or in the endoplasmic reticulum trim the N-terminal of these peptides (Stoltze et al. 2000b). Therefore, a possible MHC ligand can be defined as fragments generated by two consecutive predicted cleavages, which are eight to 15 residues apart. According to this definition, 20.28% of the predicted products of the immunoproteasome can become MHC ligands, whereas among the predicted products of the constitutive proteasome, only 16.58% can become MHC ligands. Thus, the immunoproteasome generates significantly more ($P<0.001$ in a chi-square test) potential MHC Class I ligands than the constitutive proteasome. The constitutive proteasome, on the other hand, generates more short (less than five amino acid) peptides. The lengths of the predicted peptides resemble lengths recorded in the earlier experimental data where both proteasomes generate peptides of length 3–22 amino acids, and the average length is 7–8 amino acids (Kisselev et al. 1999; Toes et al. 2001). These experimental data were always interpreted as evidence that the immunoproteasome and the constitutive proteasome generate similar size peptides. In our predictions the mean size of the predicted peptides is very similar, too (7.22 for the immunoproteasome, and 7.06 for the constitutive proteasome). However, we do find a significant difference between the percentages of MHC ligands that can be produced by the immunoproteasome and the constitutive proteasome. (Cascio et al. 2001) also found that peptides containing the dominant ovalalbumin epitope (SIINFEKL) are generated significantly more often by the immunoproteasomes than the constitutive proteasomes despite the fact that the average length of ovalalbumin degradation products was the same for the immunoproteasome and the constitutive proteasome.

The above calculation assumes that each predicted site will be used by the proteasomes, i.e., no overlapping fragments can be generated by the proteasomes. However, it is known that the cleavage process is highly stochastic: overlapping fragments are very often found in the experimental systems, see e.g., Nussbaum et al. (1998). Thus, each predicted cleavage site will be used with a certain probability by the proteasome and some fragments may overlap. To include this effect, we repeat 50 independent cleavage "simulations" of all viral proteins in our data set, allowing each cleavage to occur with a 50% probability. Including this stochasticity allows for longer fragments, i.e., both immunoproteasome and constitutive proteasome fragment distribution shifts to the right compared to Fig. 4 (results not shown). Analyzing all the fragments generated in these simula-

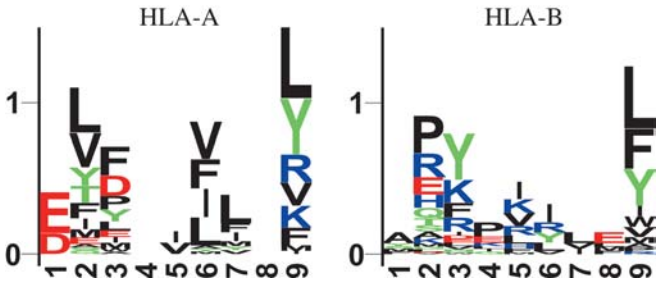

Fig. 5 Sequence logos for the binding motifs of 25 HLA-A molecules, 57 HLA-B molecules. The height of the each bar gives the Shannon information at that position. The color code of amino acids is given in Fig. 2

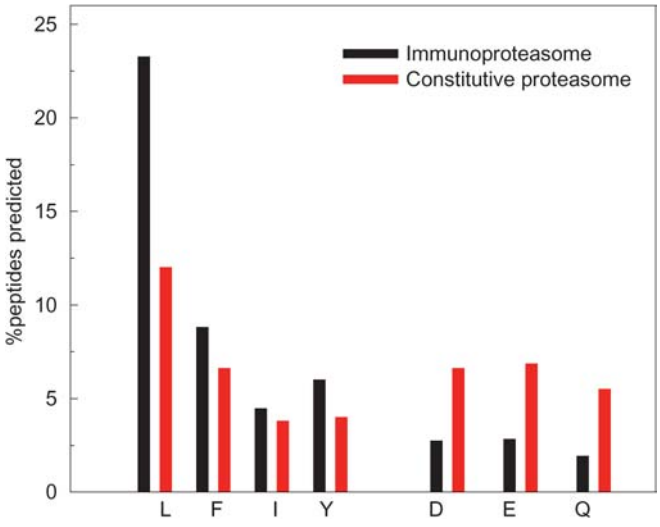

Fig. 6 Amino acid usage at the P1 position of the viral 9mers predicted to be generated by the immunoproteasome (*black bars*) and the constitutive proteasome (*red bars*). Only the amino acids which are used significantly differently by the two forms of the proteasome ($P<0.01$ in a chi-square test) are shown. Similar results were obtained for potential MHC ligands, i.e., predicted peptides of length 8–15 (not shown)

tions, we found that 25.63% of immunoproteasome products are potential ligands for MHC molecules, i.e., they are eight to 15 amino acids long, whereas for the constitutive proteasome only 22.4% of the products are potential MHC ligands. These results suggest that independent of ways of defining fragments, the immunoproteasome generates significantly more ($P<0.01$ in a chi-square test) MHC ligands than the constitutive proteasome.

Co-evolution of specificities: Immunoproteasome generates peptides that fit better to MHC molecules

Given the different sequence motifs used by the immunoproteasome and the constitutive proteasome, the peptides generated by these enzymes might interact differently with MHC molecules. To analyze this, we study binding motifs from human MHC class I (HLA) molecules from the A and B loci. At the moment binding motifs are available for 25 HLA-A alleles and 57 HLA-B alleles alleles in the SYFPEITHI database [http://www.uni-tuebingen.de/uni/kxi/, (Rammensee et al. 1999)]. The sequence logos of MHC binding motifs are given in Fig. 5. Although there are some exceptions, almost all of these MHC molecules use position two and nine as an anchor residue (see Fig. 5). Clearly, the most conserved position is position nine, and all MHC molecules included in this analysis prefer neutral and hydrophobic amino acids at position nine. The HLA-C locus is not included in our analysis because the binding motifs of only a very limited number of HLA-C alleles are known.

We compare these motifs with the cleavage motifs of the immunoproteasome and the constitutive proteasome. Note that when a peptide is generated by the proteasome, the last position of the peptide, e.g., position nine in a 9mer , corresponds to the P1 residue. The distribution of the amino acids at the C-terminal of the predicted viral 9mers is shown in Fig. 6. The figure shows only the amino acids that are used significantly differently by the constitutive proteasome and the immunoproteasome. The amino acids that tend to be used as anchor residues by human MHC molecules (i.e., L, F, I, Y) are used more frequently by the immunoproteasome. The amino acids that are known to inhibit MHC binding (i.e., D, E and Q) are used less frequently by the immunoproteasome than by the constitutive proteasome. A similar distribution is obtained for possible MHC ligands, i.e., predicted peptides that are 8–15 amino acids long (results not shown). A limited number of HLA alleles use positively charged amino acids at position nine, e.g., HLA-A3 and HLA-A11. However, it is shown that inhibition of the proteasome activity does not affect the antigen presentation by these molecules (Benham et al.1998), suggesting that the ligands for these alleles are generated via other proteolytic processes. Similar results have been obtained for predicted human peptides (results not shown).

One can use a publicly available MHC-peptide binding predictor to quantify how well the predicted peptides bind to different human MHC molecules. In the viral data set there are approximately 5.2×10^5 possible 9mers. We

Table 4 Average generation probability of good MHC binders. The binding affinity of all possible viral 9mers (5.2×10^5) is predicted for three human MHC molecules. Then, for the 1,000 9mers with highest binding capacities per MHC molecule we calculated the probability of being generated (i.e., being precisely cleaved in the C-terminus) by the immunoproteasome and the constitutive proteasome (as explained in Materials and methods). *IP* gives the average probability of being generated by the immunoproteasome and *CP* gives the similar probability of being generated by the constitutive proteasome

MHC molecule	IP	CP
*HLA*A0201*	0.26	0.15
*HLA*A0205*	0.29	0.15
*HLA*A3101*	0.17	0.12

predicted the binding of all these 9mers for HLA*0201 molecule (see Materials and methods) and ranked the 9mers according to their binding affinities. Then we calculated the probability that the first 1,000 high affinity 9mers would be generated (i.e., would be cleaved precisely at the C-terminal) by the immunoproteasome and the constitutive proteasome (as explained in Materials and methods). We repeated the same calculation for HLA*0205 and HLA*3101 molecules. The results are given in Table 4. In all cases the high affinity 9mers are significantly more likely to be generated by the immunoproteasome than by the constitutive proteasome ($P<0.01$ in a chi-square test). Similar results are obtained with the simulations where we add stochasticity to the predicted cleavage sites. As mentioned above, this allows for generating overlapping fragments, however, the fragments generated by the immunoproteasome remain to be binders of human MHC molecules (results not shown).

Taken together, these results support the hypothesis that the specificity of the MHC has co-evolved with the specificity of the immunoproteasome, but not with that of the constitutive proteasome (Groettrup et al. 2001a, 2001b). This co-evolution is limited to the specificity of C-terminal of MHC ligands and the specificity of the P1 position in the immunoproteasome digests. The P1′ and P2′ positions do not become the N-terminal of MHC ligands, due to additional trimming that takes place in the cytoplasm and the endoplasmic reticulum (Cascio et al. 2001; Stoltze et al. 2000b). Therefore, the specificity of the immunoproteasome and the constitutive proteasome has not diverged at these positions (Fig. 2). Comparison of Fig. 2 with Fig. 5 also suggests that the specificity of MHC at the N-terminal does not resemble the specificity of the proteasomes at the P1″ and P2″ positions.

Discussion

Using a simple bioinformatic analysis we show that the immunoproteasome has evolved to become a more specific enzyme than its constitutive counterpart and that the evolved specificity fits well with the specificity of human MHC molecules at the C-terminal. Such a co-evolutionary process increases the efficiency of the antigen presentation of the host (Groettrup et al. 2001a, 2001b). Both types of proteasome seem to have a similar efficiency in degrading proteins, despite the fact that the constitutive proteasome uses more degenerate sequence motifs for cleavage.

These findings have important consequences for self-tolerance. The immunoproteasome and the constitutive proteasome generate different sets of self peptides. Out of more than 2.85 million possible self 9mers included in this study, less than 4% were predicted to be generated either by the immunoproteasome or the constitutive proteasome and only approximately 0.3% of all possible self 9mers were predicted to be generated by both. Activated antigen-presenting cells expressing the immunoproteasome are expected to present a different subset of self peptides than the constitutive proteasome (Groettrup et al. 2001a). Thus, in order to prevent autoimmune reactions, the self has to be learned in the context of the immunoproteasome, i.e., the negative selection of T cells should be performed by antigen presenting cells which express immunoproteasomes. Possible cross-reactivity between self and non-self is further reduced if the majority of self and non-self are degraded by different forms of proteasome, because the two forms of proteasome use different cleavage motifs.

The immunoproteasome is a more specific enzyme, and thus it generates a pool of peptides that are less diverse than those generated by the constitutive proteasome: the diversity of the predicted pathogenic 9mers generated by the immunoproteasome is 16.3 on average per position, whereas the diversity generated by the constitutive proteasome is 18.4. Having a less diverse peptide pool enhances antigen presentation because large quantities of MHC ligands become available. Obviously, the specificity of the immunoproteasome could be even higher; however, this creates an easy immune evasion pathway for the pathogens. We recently showed that rapidly evolving viruses, such as HIV, adapt to evade degradation by the immunoproteasome in parts of their genome, but fail to do so in more conserved regions that appear as clusters of epitopes (Yusim et al. 2002). Only in rare cases can a virus evolve a protein that can escape proteasomal degradation, like the Epstein-Barr virus (EBV) encoded nuclear antigen (EBNA) 1 (Levitskaya et al. 1997). This is partly due to the fact that the immunoproteasome prefers to cleave after leucine, the most frequent amino acid and therefore the hardest to avoid. Thus, the level of specificity of the immunoproteasome seems to be a good compromise between preventing immune evasion and generating enough MHC ligands.

Next, one can wonder why the immunoproteasome does not seem to be polymorphic (Van Endert et al. 1992), since the polymorphism of MHC molecules is crucial for preventing evasion of the host immune response by pathogens. Polymorphism could be very difficult to achieve on two levels. At the individual level, since each cell expresses up to six different MHC class I molecules, an immunoproteasome allele has to co-evolve with different sets of MHC molecules in different individuals. Alternatively if different subunits are encoded in different loci, i.e., if each cell has more than one immunoproteasome molecule, very few MHC ligands would be available. At the population level, the polymorphism of the immunoproteasome requires a very close coupling with the corresponding MHC allele, which is again an evolutionarily difficult strategy. Thus, the immunoproteasome remains non-polymorphic, and MHC molecules cannot diversify their anchor residue usage at position nine because they all operate on the same peptide pool generated by immunoproteasome (Fig. 5). However, for position two, anchor residue HLA molecules clearly have different preferences.

It has been suggested that the human and mouse immunosubunits function similarly (Niedermann et al. 1997; Sijts et al. 2000a). If the specificity of the human immunosubunits has co-evolved with human MHC class I molecules, the functional identity between human and mouse immunosubunits implies similarities between the binding motifs of human and mouse MHC class I molecules. For the seven class I mouse MHC binding motifs available in the SYFPEITHI database [http://www.uni-tuebingen.de/uni/kxi/, (Rammensee et al.1999), K, L, and D loci], L, I, V, F, and M are preferred at the position nine (or eight) anchor residue, as is the case for human MHC binding motifs (see Fig. 5). This suggests that the coupling of the specificities of MHC and the immunoproteasome is conserved across species. This is remarkable since MHC molecules (and to a lesser extent the immunosubunits) evolve very fast.

A possible weak point in our method of predicting which fragments are generated by the immunoproteasome and the constitutive proteasome is the assumption that flanking regions and the P1 position all contribute independently to the cleavage process. This is not necessarily the case: since amino acid side chains have very different physico-chemical properties, the combination of two or more residues can be more cleavage inhibiting or enhancing than each residue on its own. In order to detect such interdependencies, we trained artificial neural networks [as explained in Kesmir et al. (2002)] using the observed cleavage sites. With these networks we find the fragment length distribution and the amino acid distribution of the C-terminal residues to be similar to the distributions given in Fig. 4 and in Fig. 6 (results not shown). Thus, our results are fairly independent of the prediction method. These networks were trained using the C-terminal of naturally processed MHC ligands (Kesmir et al.2002). When tested on Toes et al. (2001) data, the networks predict much more accurately the sites that are used by the immunoproteasome than the sites that are used by the constitutive proteasome (results not shown). This provides a separate line of evidence for the co-evolution of the specificity of the MHC at the C-terminal and the immunoproteasome.

Our analysis and the analysis performed by Toes et al. (2001) and Cascio et al. (2001) suggest that the immunoproteasome and the constitutive proteasome clearly have different cleavage preferences. However, a recent study by Peters et al. (2002) failed to find these differences and suggests that the differences between two forms of the proteasome can be an artifact of the re-processing of substrates in in vitro degradation experiments. Although this possibility cannot be ruled out, Toes et al. (2001) tried to minimize the chance of re-processing by carrying out the degradation experiments until approximately half of the substrates were degraded. Moreover, the analysis performed by Peters et al. (2002) is based on limited data, i.e., the degradation of two relatively short peptides (one 25-mer and one 27-mer), and therefore, it may well be inconclusive with regard to the degradation of proteins.

Differences between the immunoproteasome and the constitutive proteasome are not limited to the subunit composition. Immunoproteasomes contain an IFN-γ inducible regulatory unit called the proteasome activator PA28 (Rock and Goldberg 1999; Kloetzel 2001). PA28 on its own can influence antigen presentation drastically, mainly by changing the kinetics of the degradation process without altering the specificity (Preckel et al. 1999; Stohwasser et al. 2000; Van Hall et al. 2000; Kloetzel 2001; Sijts et al. 2002; Sun et al. 2002). Moreover, the immunoproteasomes, unlike constitutive proteasomes, may co-localize with TAP and thus increase MHC class I antigen presentation (Brooks et al. 2000). Using the data of Toes et al. (2001) we are obviously only able to study the effect of immunosubunits on the specificity of the immunoproteasome, and are obliged to omit other important factors which might further influence the peptide repertoire generated by the immunoproteasomes.

Acknowledgements We thank Alexander K. Nussbaum for discussing the analysis. The diversity measure used here was derived by Nigel Burroughs. We are grateful to S.M. McNab for linguistic advice.

References

Altschul S, Gish W, Miller W, Myers E, Lipman D (1990) Basic local alignment search tool. J Mol Biol 215:403–410

Altuvia Y, Margalit H (2000) Sequence signals for generation of antigenic peptides by the proteasome: implications for proteasomal cleavage mechanism. J Mol Biol 295:879–890

Bairoch A, Apweiler R (2000) The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000 Nucleic Acids Res 28:45–48

Baumeister W, Walz J, Zuhl F, Seemuller E (1998) The proteasome: paradigm of a self-compartmentalizing protease. Cell 92:367–380

Beekman NJ, Van Veelen PA, Van Hall T, Neisig A, Sijts A, Camps M, Kloetzel PM, Neefjes JJ, Melief CJ, Ossendorp F (2000) Abrogation of CTL epitope processing by single amino acid substitution flanking the C-terminal proteasome cleavage site. J Immunol 164:1898–1905

Benham A M, Gromme M, Neefjes J (1998) Allelic differences in the relationship between proteasome activity and MHC class I peptide loading. J Immunol 161:83–89

Berger A, Schechter I (1970) Mapping the active site of papain with the aid of peptide substrates and inhibitors. Philos Trans R Soc Lond B Biol Sci 257:249–264

Brooks P, Murray RZ, Mason GG, Hendil KB, Rivett AJ (2000) Association of immunoproteasomes with the endoplasmic reticulum. Biochem J 352:611–615

Cardozo C, Kohanski RA (1998) Altered properties of the branched chain amino acid-preferring activity contribute to increased cleavages after branched chain residues by the "immunoproteasome". J Biol Chem 273:16764–16770

Cascio P, Hilton C, Kisselev AF, Rock KL, Goldberg AL (2001) 26 s proteasomes and immunoproteasomes produce mainly N-extended versions of an antigenic peptide. EMBO J 20:2357–2366

Chen W, Norbury CC, Cho Y, Yewdell JW, Bennink JR (2001) Immunoproteasomes shape immunodominance hierarchies of antiviral CD8($^+$) T cells at the levels of T cell repertoire and presentation of viral antigens. J Exp Med 193:1319–1326

Dahlmann B, Ruppert T, Kuehn L, Merforth S, Kloetzel PM (2000) Different proteasome subtypes in a single tissue exhibit different enzymatic properties. J Mol Biol 303:643–653

Ehring B, Meyer TH, Eckerskorn C, Lottspeich F, Tampe R (1996) Effects of major-histocompatibility-complex-encoded subunits on the peptidase and proteolytic activities of human 20S proteasomes. Cleavage of proteins and antigenic peptides. Eur J Biochem 235:404–415

Eleuteri AM, Kohanski RA, Cardozo C, Orlowski M (1997) Bovine spleen multicatalytic proteinase complex (proteasome). Replacement of X, Y, and Z subunits by LMP7, LMP2, and MECL1 and changes in properties and specificity. J Biol Chem 272:11824–11831

Groettrup M, Soza A, Kuckelkorn U, Kloetzel PM (1996) Peptide antigen production by the proteasome: complexity provides efficiency. Immunol Today 17:429–435

Groettrup M, Khan S, Schwarz K, Schmidtke G (2001a) Interferon-gamma inducible exchanges of 20S proteasome active site subunits: why? Biochimie 83:367–372

Groettrup M, Van den Broek M, Schwarz K, Macagno A, Khan S, De Giuli R, Schmidtke G (2001b) Structural plasticity of the proteasome and its function in antigen processing. Crit Rev Immunol 21:339–358

Groll M, Ditzel L, Lowe J, Stock D, Bochtler M, Bartunik H D, Huber R (1997) Structure of 20S proteasome from yeast at 2.4 A resolution. Nature 386:463–471

Groll M, Heinemeyer W, Jager S, Ullrich T, Bochtler M, Wolf DH, Huber R (1999) The catalytic sites of 20S proteasomes and their role in subunit maturation: a mutational and crystallographic study. Proc Natl Acad Sci USA 96:10976–10983

Heinemeyer W, Fischer M, Krimmer T, Stachon U, Wolf DH (1997) The active sites of the eukaryotic 20S proteasome and their involvement in subunit precursor processing. J Biol Chem 272:25200–25209

Hughes AL (1997) Evolution of the proteasome components. Immunogenetics 46:82–92

Hughes AL, Nei M (1988) Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. Nature 335:167–170

Hughes AL, Nei M (1989) Nucleotide substitution at major histocompatibility complex class II loci: evidence for overdominant selection. Proc Natl Acad Sci USA 86:958–962

Kesmir C, Nussbaum AK, Schild H, Detours V, Brunak S (2002) Prediction of proteasome cleavage motifs by neural networks. Protein Eng 15:287–296

Khan S, Van den Broek M, Schwarz K, De Giuli R, Diener PA, Groettrup M (2001) Immunoproteasomes largely replace constitutive proteasomes during an antiviral and antibacterial immune response in the liver. J Immunol 167:6859–6868

Kisselev AF, Akopian TN, Woo KM, Goldberg AL (1999) The sizes of peptides generated from protein by mammalian 26 and 20S proteasomes. Implications for understanding the degradative mechanism and antigen presentation. J Biol Chem 274:3363–3371

Kloetzel PM (2001) Antigen processing by the proteasome. Nat Rev Mol Cell Biol 2:179–187

Kuckelkorn U, Frentzel S, Kraft R, Kostka S, Groettrup M, Kloetzel PM (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

Kuckelkorn U, Ruppert T, Strehl B, Jungblut PR, Zimny-Arndt U, Lamer S, Prinz I, Drung I, Kloetzel PM, Kaufmann SH, Steinhoff U (2002) Link between organ-specific antigen processing by 20S proteasomes and CD8($^+$) T cell-mediated autoimmunity. J Exp Med 195:983–990

Levitskaya J, Sharipo A, Leonchiks A, Ciechanover A, Masucci MG (1997) Inhibition of ubiquitin/proteasome-dependent protein degradation by the Gly-Ala repeat domain of the Epstein-Barr virus nuclear antigen 1. Proc Natl Acad Sci USA 94:12616–12621

Mo AX, Van Lelyveld SF, Craiu A, Rock KL (2000) Sequences that flank subdominant and cryptic epitopes influence the proteolytic generation of MHC class I-presented peptides. J Immunol 164:4003–4010

Morel S, Levy F, Burlet-Schiltz O, Brasseur F, Probst-Kepper M, Peitrequin AL, Monsarrat B, Van Velthoven R, Cerottini JC, Boon T, Gairin JE, Van den Eynde BJ (2000) Processing of some antigens by the standard proteasome but not by the immunoproteasome results in poor presentation by dendritic cells. Immunity 12:107–117

Niedermann G, Grimm R, Geier E, Maurer M, Realini C, Gartmann C, Soll J, Omura S, Rechsteiner MC, Baumeister W, Eichmann K (1997) Potential immunocompetence of proteolytic fragments produced by proteasomes before evolution of the vertebrate immune system. J Exp Med 186:209–220

Noda C, Tanahashi N, Shimbara N, Hendil KB, Tanaka K (2000) Tissue distribution of constitutive proteasomes, immunoproteasomes, and PA28 in rats. Biochem Biophys Res Commun 277:348–354

Nussbaum AK, Dick TP, Keilholz W, Schirle M, Stevanovic S, Dietz K, Heinemeyer W, Groll M, Wolf DH, Huber R, Rammensee HG, Schild H (1998) Cleavage motifs of the yeast 20S proteasome β subunits deduced from digests of enolase 1. Proc Natl Acad Sci USA 95:12504–12509

Ossendorp F, Eggers M, Neisig A, Ruppert T, Groettrup M, Sijts A, Mengede E, Kloetzel PM, Neefjes J, Koszinowski U, Melief C (1996) A single residue exchange within a viral CTL epitope alters proteasome-mediated degradation resulting in lack of antigen presentation. Immunity 5:115–124

Peters B, Janek K, Kuckelkorn U, Holzhutter HG (2002) Assessment of proteasomal cleavage probabilities from kinetic analysis of time-dependent product formation. J Mol Biol 318:847–862

Preckel T, Fung-Leung WP, Cai Z, Vitiello A, Salter-Cid L, Winqvist O, Wolfe TG, Von Herrath M, Angulo A, Ghazal P, Lee JD, Fourie AM, Wu Y, Pang J, Ngo K, Peterson PA, Fruh K, Yang Y (1999) Impaired immunoproteasome assembly and immune responses in PA28-/- mice. Science 286:2162–2165

Rammensee H, Bachmann J, Emmerich NP, Bachor O A, Stevanovic S (1999) SYFPEITHI: database for MHC ligands and peptide motifs. Immunogenetics 50:213–219

Robinson J, Waller M, Parham P, Bodmer J, Marsh S (2001) IMGT/HLA Database— a sequence database for the human major histocompatibility complex. Nucleic Acids Res 29:210–213

Rock KL, Goldberg AL (1999) Degradation of cell proteins and the generation of MHC class I-presented peptides. Annu Rev Immunol 17:739–779

Schneider TD, Stephens RM (1990) Sequence logos: a new way to display consensus sequences. Nucleic Acids Res 18:6097–6100

Schultz ES, Chapiro J, Lurquin C, Claverol S, Burlet-Schiltz O, Warnier G, Russo V, Morel S, Levy F, Boon T, Van den Eynde BJ, Van der Bruggen P (2002) The production of a new MAGE-3 peptide presented to cytolytic T lymphocytes by HLA-B40 requires the immunoproteasome. J Exp Med 195:391–399

Shannon CE (1948) A mathematical theory of communication. Bell System Tech J 27:379–423, 623–656

Sijts AJ, Ruppert T, Rehermann B, Schmidt M, Koszinowski U, Kloetzel PM (2000a) Efficient generation of a hepatitis B virus cytotoxic T lymphocyte epitope requires the structural features of immunoproteasomes. J Exp Med 191:503–514

Sijts AJ, Standera S, Toes RE, Ruppert T, Beekman NJ, Van Veelen PA, Ossendorp FA, Melief CJ, Kloetzel PM (2000b) MHC class I antigen processing of an adenovirus CTL epitope is linked to the levels of immunoproteasomes in infected cells. J Immunol 164:4500–4506

Sijts A, Sun Y, Janek K, Kral S, Paschen A, Schadendorf D, Kloetzel P (2002) The role of the proteasome activator PA28 in MHC class I antigen processing. Mol Immunol 39:165

Stohwasser R, Salzmann U, Giesebrecht J, Kloetzel PM, Holzhutter HG (2000) Kinetic evidences for facilitation of peptide

channelling by the proteasome activator PA28. Eur J Biochem 267:6221–6230

Stoltze L, Nussbaum AK, Sijts A, Emmerich NP, Kloetzel PM, Schild H (2000a) The function of the proteasome system in MHC class I antigen processing. Immunol Today 21:317–319

Stoltze L, Schirle M, Schwarz G, Schroter C, Thompson M W, Hersh L B, Kalbacher H, Stevanovic S, Rammensee HG, Schild H (2000b) Two new proteases in the MHC class I processing pathway. Nat Immunol 1:413–418

Sun Y, Sijts AJ, Song M, Janek K, Nussbaum AK, Kral S, Schirle M, Stevanovic S, Paschen A, Schild H, Kloetzel PM, Schadendorf D (2002) Expression of the proteasome activator PA28 rescues the presentation of a cytotoxic T lymphocyte epitope on melanoma cells. Cancer Res 62:2875–2882

Tanaka K, Kasahara M (1998) The MHC class I ligand-generating system: roles of immunoproteasomes and the interferon-gamma-inducible proteasome activator PA28. Immunol Rev 163:161–176

Tanaka T, Nei M (1989) Positive Darwinian selection observed at the variable-region genes of immunoglobulins. Mol Biol Evol 6:447–459

Toes RE, Nussbaum AK, Degermann S, Schirle M, Emmerich NP, Kraft M, Laplace C, Zwinderman A, Dick TP, Muller J, Schonfisch B, Schmid C, Fehling HJ, Stevanovic S, Rammensee HG, Schild H (2001) Discrete cleavage motifs of constitutive and immunoproteasomes revealed by quantitative analysis of cleavage products. J Exp Med 194:1–12

Van den Eynde BJ, Morel S (2001) Differential processing of class-I-restricted epitopes by the standard proteasome and the immunoproteasome. Curr Opin Immunol 13:147–153

Van Endert PM, Lopez MT, Patel SD, Monaco J, McDevitt HO (1992) Genomic polymorphism, recombination, and linkage disequilibrium in human major histocompatibility complex-encoded antigen-processing genes. Proc Natl Acad Sci USA 89:11594–11597

Van Hall T, Sijts A, Camps M, Offringa R, Melief C, Kloetzel PM, Ossendorp F (2000) Differential influence on cytotoxic T lymphocyte epitope presentation by controlled expression of either proteasome immunosubunits or PA28. J Exp Med 192:483–494

Wollenberg K, Swaffield JC (2001) Evolution of proteasomal ATPases. Mol Biol Evol 18:962–974

Yang Z, Nielsen R (2000) Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. Mol Biol Evol 17:32–43

Yewdell JW, Bennink JR (2001) Cut and trim: generating MHC class I peptide ligands. Curr Opin Immunol 13:13–18

Yusim K, Kesmir C, Gaschen B, Addo MM, Altfeld M, Brunak S, Chigaev A, Detours V, Korber BT (2002) Clustering patterns of cytotoxic T-lymphocyte epitopes in human immunodeficiency virus type 1 (HIV-1) proteins reveal imprints of immune evasion on HIV-1 global variation. J Virol 76:8757–8768