

RESEARCH ARTICLE

Open Access

Computational analysis and modeling of cleavage by the immunoproteasome and the constitutive proteasome

Carmen M Diez-Rivero^{1,2}, Esther M Lafuente², Pedro A Reche^{1,2*}

Abstract

Background: Proteasomes play a central role in the major histocompatibility class I (MHCI) antigen processing pathway. They conduct the proteolytic degradation of proteins in the cytosol, generating the C-terminus of CD8 T cell epitopes and MHCI-peptide ligands (*P1* residue of cleavage site). There are two types of proteasomes, the constitutive form, expressed in most cell types, and the immunoproteasome, which is constitutively expressed in mature dendritic cells. Protective CD8 T cell epitopes are likely generated by the immunoproteasome and the constitutive proteasome, and here we have modeled and analyzed the cleavage by these two proteases.

Results: We have modeled the immunoproteasome and proteasome cleavage sites upon two non-overlapping sets of peptides consisting of 553 CD8 T cell epitopes, naturally processed and restricted by human MHCI molecules, and 382 peptides eluted from human MHCI molecules, respectively, using *N-grams*. Cleavage models were generated considering different epitope and MHCI-eluted fragment lengths and the same number of C-terminal flanking residues. Models were evaluated in 5-fold cross-validation. Judging by the Mathew's Correlation Coefficient (*MCC*), optimal cleavage models for the proteasome ($MCC = 0.43 \pm 0.07$) and the immunoproteasome ($MCC = 0.36 \pm 0.06$) were obtained from 12-residue peptide fragments. Using an independent dataset consisting of 137 HIV1-specific CD8 T cell epitopes, the immunoproteasome and proteasome cleavage models achieved *MCC* values of 0.30 and 0.18, respectively, comparatively better than those achieved by related methods. Using ROC analyses, we have also shown that, combined with MHCI-peptide binding predictions, cleavage predictions by the immunoproteasome and proteasome models significantly increase the discovery rate of CD8 T cell epitopes restricted by different MHCI molecules, including A*0201, A*0301, A*2402, B*0702, B*2705.

Conclusions: We have developed models that are specific to predict cleavage by the proteasome and the immunoproteasome. These models ought to be instrumental to identify protective CD8 T cell epitopes and are readily available for free public use at <http://imed.med.ucm.es/Tools/PCPS/>.

Background

CD8 cytotoxic T cells play a key role fighting intracellular pathogens, eliminating infected cells that display on their cell surface foreign peptides bound to major histocompatibility complex class I (MHCI) molecules [1-3]. CD8 T cell epitopes and, in general, peptides presented by MHCI molecules, derive from protein fragments produced in the cytosol by the proteolytic action of the

proteasome [4,5]. Briefly, the proteasome generates protein fragments between 7 and 15 amino acids. Some of these peptides can be transported from the cytosol into the endoplasmic reticulum (ER) by the transporter associated with antigen processing (TAP), where they can be loaded onto nascent MHCI molecules. Interestingly, whereas different peptidases and proteases in the cytosol and the endoplasmic reticulum shape the N-terminus of the peptides presented by MHCI molecules [6], their C-terminus generally corresponds to the *P1* residue of the proteasome cleavage site [7,8].

The proteasome is a multisubunit ATP-dependent protease and it is primarily responsible for the degradation

* Correspondence: parecheg@med.ucm.es

¹Laboratory of Immunomedicine, Department of Microbiology I-Immunology, Facultad de Medicina, Universidad Complutense de Madrid, Ave Complutense S/N, Madrid 28040, Spain

Full list of author information is available at the end of the article

of cytosolic proteins [9]. The most common form of the proteasome is known as the 26 S proteasome, which is composed by a catalytic core (20S) and two regulatory complexes (19S), located one at each side of the core [5]. The catalytic activity of the proteasome is located at the subunits $\beta 5$ (X, LMP7), $\beta 2$ (Z, MECL-1) and $\beta 1$ (Y, LMP2) of the 20 S core, which cut after the C-terminus of hydrophobic (chymotrypsin-like activity), basic (trypsin-like activity) or acidic (caspase-like activity) amino acids, respectively [10]. Upon IFN- γ exposure, the three catalytic subunits of the constitutive 20 S core can be replaced by three new catalytic subunits: $\beta 5i$ (LMP2), $\beta 2i$ (MECL-1), and $\beta 1i$ (LMP2) [11]. This new form of proteasome is called immunoproteasome, as opposed to the constitutively expressed proteasome. The immunoproteasome is the constitutive form of proteasome presented in dendritic cells [12]. The immunoproteasome produces different but overlapping cleavage patterns with regard to those of the proteasome [13]; chiefly, the immunoproteasome does not cut after acidic residues [13,14]. Because the antigen-specific cytotoxic function of CD8 T cells is generally acquired upon the recognition of MHCI-bound peptide antigens displayed on the cell surface of dendritic cells (priming), it is likely that protective epitopes are those generated by the proteasome and the immunoproteasome [15].

Prediction of proteasome cleavage sites is relevant for CD8 T cell epitope identification and, subsequently, for the design of epitope-based vaccines eliciting CD8 T cell responses. Therefore, different methods to predict proteasome cleavage sites have been reported. Proteasome cleavage prediction methods were first developed using enolase and β -casein protein fragments generated *in vitro* by human constitutive proteasomes [16-18]. Likewise, a kinetic model of the proteasome proteolytic activity was also developed using peptide fragments from *in vitro* digestions [19,20]. Those models are specific for the constitutive 20 S proteasome that was used to generate the peptide fragments. Proteasome cleavages take place between the C-terminus of MHCI-restricted peptides ($P1$ residue of cleavage site) and their most proximal C-terminal flanking residue ($P1'$ residue of cleavage site). Therefore, proteasome cleavage prediction methods have also been developed using MHCI-restricted peptide ligands and their C-terminal flanking regions [21-23]. These latter methods appear to out-compete the former methods that were trained on actual proteolytic digestion data on the task of predicting cleavage sites defined by MHC I restricted peptides [24]. However, methods trained on experimental cleavage data can be more suitable for identifying protein fragments produced by the proteasome [18].

The problem of predicting proteasome cleavage sites resembles that of modeling grammatical rules. Therefore,

in this manuscript, we have applied statistical language models [25] to analyze and model the cleavage sites of the constitutive proteasome and the immunoproteasome. Proteasome cleavage sites were obtained from MHCI-eluted peptides and their C-terminal flanking regions, whereas immunoproteasome cleavage sites were rendered from naturally processed CD8 T cell epitopes and their C-terminal flanking regions. In cross-validation, optimal proteasome and immunoproteasome cleavage models achieved an *MCC* of 0.43 ± 0.07 and 0.36 ± 0.06 , respectively. These models were trained using 12-residue fragments, consisting of the C-terminal end of MHCI-restricted peptides ($P6 - P1$ residues of cleavage site) followed by the 6 most-proximal C-terminal flanking residues ($P1' - P6'$ residues of cleavage site). The fact that optimal models were trained using peptide fragments consisting of 6 amino acids at each side of the cleavage site is consistent with the activity exhibited by the proteasome [26]. Here, we have also shown that combining cleavage predictions by the constitutive and the immunoproteasome with MHCI-binding predictions serve to improve the prediction rate of CD8 T cell epitopes. Cleavage predictions using our models are available at <http://imed.med.ucm.es/Tools/PCPS/>.

Methods

Datasets and sequences

We assembled three non-overlapping datasets consisting of distinct MHCI-restricted peptides and their protein sources. The peptide content in these datasets was as follows. The first dataset encompassed 553 CD8 T cell epitopes from different sources but from Human Immunodeficiency Virus (HIV1) and were all restricted by human MHCI molecules. Immune responses against these epitopes have been verified experimentally using T cells from infected humans. Because CD8 T cell immune responses against these epitopes are elicited in the course of an infection, we assume that they are naturally processed. The second dataset included 382 peptides that were eluted from human MHCI molecules, and the third dataset encompassed 137 HIV1-specific CD8 T cell epitopes restricted by human MHCI molecules and naturally processed. MHCI-restricted peptides in these datasets were collected from the EPIMHC [27], Immunepitope [28] and Los Alamos databases [29], and consisted of unique nonapeptides (9-mers) that were subjected to a sequence similarity reduction schema using the *purge* utility implemented in the Gibbs Sampler [30]. As a result, peptides in these three datasets do not share more than 4 identical residues (global sequence similarity in the first, second and third datasets is 3.1 ± 11.7 , 3.9 ± 12.8 , and 3.5 ± 11.7 , respectively). Moreover, in all datasets the same MHCI molecule restricts less than 18% of all peptides. In additional file 1, we show the

distribution of commonly expressed MHC I alleles in each of the three datasets. The corresponding author will also provide these datasets upon written request.

Model building and evaluation

Cleavage models were trained and evaluated on datasets consisting of peptide fragments of the same length derived from MHC I-eluted peptides (proteasome models) and CD8 T cell epitopes (immunoproteasome models) and their C-terminal flanking regions, using the NGRAM-COUNT utility implemented by the SRLIM package [25]. Peptide fragments encompassed two portions with the same number of residues, one fraction consisting of the C-terminal end of MHC I-eluted peptides or CD8 T cell epitopes, and the other one of their C-terminal flanking region. Cleavage sites -defined between the C-terminus of MHC I-restricted peptides (*PI* residue of cleavage site) and the most proximal C-terminal flanking residue (*PI'* residue)- were indicated by a "|" symbol. Cleavage models were generated considering peptide fragments ranging from 4 to 18 residues. Representative peptide fragments of 6 and 12 amino acids are C T L | T I G and P S C C T L | T I G V S S, respectively, where C T L and P S C C T L are two C-terminal portions of the peptide and T I G and T I G V S S are C-terminal flanking residues drawn from the protein source. Cleavage models were tested and evaluated at different thresholds using the SRLIM HIDDEN-NGRAM utility. HIDDEN-NGRAM is a word boundary program that uses *N-gram* models [25] produced by NGRAM-COUNT to predict the probability of hidden tags -cleavage sites- in any peptide fragment. The evaluation of the models was carried out through 5-fold cross-validation experiments that were repeated 5 times, obtaining mean estimations and standard deviations of the measures of performance indicated below.

Measures of performance

Cleavage predictions were examined in each residue at different probability thresholds (*th*) and were judged following the schema proposed in previous works [22,24]. It is assumed that cleavage sites should preferentially occur after the C-terminus of MHC I-restricted peptides (*PI* residue of cleavage site) than over any other position within the peptide. Under such schema, any given test peptide was classified as follows:

- TP (True positive): Cleavage score at the C-terminus (*PI* residue of cleavage site) is above the *th*.
- FN (False negative): Cleavage score of *PI* residue is below the *th*.
- TN (True negative): All the residues within the test fragment have a cleavage score below the *th*. Alternatively, if there are residues with cleavage scores above the *th*, but smaller than that of the *PI* residue.

- FP (False positive): There is at least one residue within the peptide with a cleavage score that is both, above the *th* and above that of the *PI* residue.

Upon this classification approach, we computed the Sensitivity (*SE*), Specificity (*SP*) and Matthews correlation coefficient (*MCC*) [31] of the predictions using Equations 1, 2 and 3, respectively,

$$SE = \frac{TP}{TP+FN} \quad (1)$$

$$SP = \frac{TN}{TN+FP} \quad (2)$$

$$MCC = \frac{(TP*TN)-(FN*FP)}{\sqrt{(TN+FN)(TP+FN)(TN+FP)(TP+FP)}} \quad (3)$$

In addition, we also computed the parameter *BTR* (Better Than Random) which was first introduced by Reche et al. [32] to compare the *SE* of a given model and that of a random model producing the same number of cleavage sites (Equation 4).

$$BTR = SE - ECS \quad (4)$$

ECS (Expected Cleavage Sites) represents the ratio of cleavage sites correctly predicted by a model that distributes cleavage sites randomly and is given by Equation 5.

$$ECS = \frac{C}{F*N} \quad (5)$$

Where *C* is the total number of cleavage sites (above the *th*) predicted by a given cleavage model in a test set of peptide fragments -specifically, within the MHC I-restricted peptide portion of the peptide fragment-; *F* is the number of MHC I-restricted peptide residues included in the peptide fragments used for training and testing; and *N* is the total number of peptide fragments in the dataset. Note that peptide fragments used for model building and evaluation encompassed two portions with the same number of residues, one consisting of the C-terminal end of MHC I-restricted peptides and the other of their C-terminal flanking region (details elsewhere in Methods). *ECS* is somewhat equivalent to the *SE* of a model that distributes all the cleavage sites randomly. Thus, the bigger the difference between *SE* and *ECS* the better the predictions produced by the model.

Prediction of peptide binding to MHC I

We used Position Specific Scoring Matrices (PSSMs) to compute binding scores of peptides to the relevant

MHCI molecules [33]. Actual binding of peptides to a particular MHC I molecule was assessed relating its binding score to those of 10000 reference peptides, 9-mers randomly obtained from SwissProt, computed using the same relevant PSSM. Thus, a given peptide was considered to bind a specific MHC I molecule when its binding score ranked among the X percentile (threshold) of top binding scores. The same peptide was considered not to bind to that MHC I if it ranked below the X percentile of top binding scores. PSSMs are derived from alignments of peptides of the same size known to bind to a given MHC I molecule [32,34,35]. Given that MHC I-bound peptides are usually of 9 residues of length, in this study we used PSSMs specific for the prediction of peptide binders of that length (9mers).

ROC analysis

We used 5 different sets of CD8 T cell epitopes consisting of 316, 50, 70, 47 and 30 peptides restricted by A*0201, A*0301, A*2402, B*0702, and B*2705, respectively, to evaluate the discovery rate of CD8 T cell epitopes using MHC I peptide-binding predictions alone, or in combination with proteasome cleavage predictions. Receiver operating characteristic (ROC) curves [36] were used to analyze the predictions. In the ROC analysis, we represented the SE (Equation 1) versus $1-SP$ (Equation 2) of the T cell epitope predictions obtained over a continuous range of percentile thresholds of MHC I binding (detail elsewhere in Methods). Non-T cell epitopes, required to compute the SP of the predictions, consisted of peptides of 9 residues randomly selected from the SwissProt database. A 1:3 ratio of T cell epitopes to non-T cell epitopes data was used. When evaluating the combination of MHC I binding and proteasome cleavage predictions, we applied a filtering approach such as that used by Dönnes and Kohlbacher [37]. Under this approach, peptides that are not predicted to be cleaved by the proteasome are discarded prior to the ROC analysis.

The area under ROC curves (AUC) was used as a global threshold-independent measure of performance. The maximum accuracy corresponds to an $AUC = 1$ while an $AUC = 0.5$ is indicative of a random prediction. Predictions are poor for values of $AUC > 0.7$, good for values of $AUC > 0.8$ and excellent for values of $AUC > 0.9$. ROC analyses were repeated 10 times, using the same T cell epitopes but different non-T cell epitopes. Thus, we obtained confident values of AUC (mean and standard deviation). Statistical significance of the differences between AUC values was evaluated using standard one-side two sample Student t - tests ($p < 0.05$).

Web implementation

Immunoproteasome and proteasome cleavage models were implemented for free public use on the Web using a

PERL CGI (Common Gateway Interface) script that executes the predictions on user-provided input data and returns the results to the browser. In addition, we used JavaScript for handling and verification of the input data before submission. Proteasome and immunoproteasome cleavage models exhibited optimal predictions at different model-specific cleavage scores. Therefore, cleavage scores by the different models were normalized and standardized so that cleavage sites are predicted at a score ≥ 0.5 .

Results

Proteasome and immunoproteasome cleavage models

Cleavage models were generated from two types of MHC I-restricted peptides and their flanking regions using N -grams. N -gram models are frequently applied to speech recognition and natural language tagging [38], but they have also been applied to sequence analysis and motif identification [32,39-41]. We built two types of cleavage models. Immunoproteasome cleavage models were built upon a dataset encompassing 553 CD8 T cell epitopes that have been reported to be recognized by humans during the course of an infection. Epitope-specific CD8 T cell responses are generally primed by dendritic cells which express the immunoproteasome. Therefore, naturally processed CD8 T cell epitopes can be used to reproduced the cleavage by the immunoproteasome. In contrast, proteasome cleavage models were based on a set of 382 peptides that were eluted from human MHC I molecules. Peptide elution experiments are generally carried out using various types of cells (virtually never dendritic cells) and under conditions that do not induce the expression of the immunoproteasome. Therefore, we considered that MHC I-eluted peptides are produced by the proteasome. A detailed description of these datasets is elsewhere in Methods.

Numerous immunoproteasome and proteasome cleavage models were obtained from different training sets consisting of peptide fragments varying from 4 to 18 residues -in a given training set, all the peptides have the same size. Peptide fragments used for training included the C-terminus ($P1$ residue of cleavage site) of MHC I-restricted peptides (CD8 T cell epitopes and MHC I-eluted peptides) and comprised two distinct portions with the same number of residues: one consisting of the C-terminal end of MHC I-restricted peptides and the other one of their C-terminal flanking region (see Methods section for more details). Cleavage models were evaluated in 5-fold cross-validation experiments, considering a continuous range of cleavage thresholds. As measures of performance we computed SE , SP , MCC and BTR (see Methods section for details), but trusted BTR as the key measure of the goodness of the predictions. In Figure 1A we show the optimal BTR achieved by the cleavage models with regard to the size of the peptide fragments used

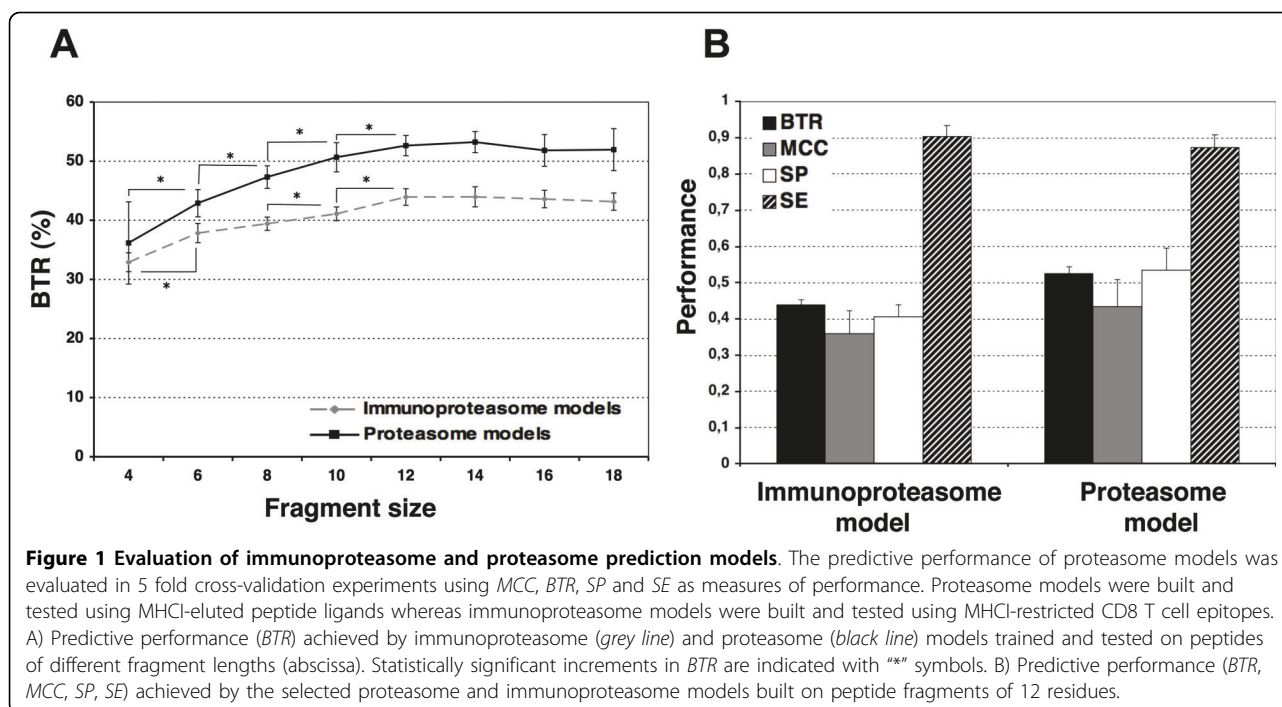


Figure 1 Evaluation of immunoproteasome and proteasome prediction models. The predictive performance of proteasome models was evaluated in 5 fold cross-validation experiments using *MCC*, *BTR*, *SP* and *SE* as measures of performance. Proteasome models were built and tested using MHCII-eluted peptide ligands whereas immunoproteasome models were built and tested using MHCII-restricted CD8 T cell epitopes. A) Predictive performance (*BTR*) achieved by immunoproteasome (grey line) and proteasome (black line) models trained and tested on peptides of different fragment lengths (abscissa). Statistically significant increments in *BTR* are indicated with "*" symbols. B) Predictive performance (*BTR*, *MCC*, *SP*, *SE*) achieved by the selected proteasome and immunoproteasome models built on peptide fragments of 12 residues.

for training. A complete summary of the performance of the cleavage models, which also includes the *MCC*, *SE*, *SP* of the predictions, is shown in Table 1.

The predictive performance of the cleavage models significantly increased ($p < 0.05$) with the length of the peptide fragments used for training, picking at a fragment size of 12-14 residues (Figure 1A); $BTR = 0.44 \pm 0.02$ for the immunoproteasome model and $BTR = 0.53 \pm 0.02$ for the proteasome model. In general, the predictive performance of proteasome cleavage models built upon MHCII-eluted peptides was higher than that achieved by immunoproteasome cleavage models, regardless of the length the peptides fragments used for training (Figure 1A). Increasing the size of the peptide fragments beyond 14 residues did not improve the predictive performance of the cleavage models (Figure 1A). Judging the predictions by the *MCC*, the immunoproteasome and proteasome models that were built on peptide fragments of 12 residues (Table 1) achieved the best results. Because no statistical difference was observed between the *BTR* achieved by the models trained on 12 and 14 residues, for further analysis, we used the models trained on 12-residue peptide fragments. The performance of the selected proteasome and immunoproteasome models is summarized in Figure 1B.

Comparison of the immunoproteasome and proteasome cleavage models

For further comparisons, we evaluated the immunoproteasome and proteasome cleavage models in an independent test set built from 137 HIV1-specific CD8 T cell epitopes

and their flanking regions (Figure 2). The immunoproteasome model achieved better results than the proteasome model, as judged by both, the *BTR* (0.45 for the immunoproteasome model and 0.39 for the proteasome model) and the *MCC* (0.30 for the immunoproteasome model and 0.18 for the proteasome model). These results indicate that the immunoproteasome model appears to be more suitable than the proteasome model to predict the cleavage sites defined by CD8 T cell epitopes.

Using the immunoproteasome and proteasome cleavage models, we analyzed the fragmentation patterns resulted from 100 proteins randomly selected from the SwissProt database (Figure 3). The immunoproteasome cleavage model generated fragments with a mean size of 2.23 ± 1.61 residues, whereas the proteasome cleavage model generated fragments with a mean size of 3.02 ± 2.33 residues. Using a Wilcoxon test, we observed no significant difference between the sizes of the fragments generated with the proteasome and immunoproteasome models (Figure 3A). This analysis also revealed that 36% of the peptide fragments generated by the proteasome and immunoproteasome are identical, and 67% of the cleavage sites are shared (Figure 3B).

Comparison with NetChop

We also used the 137 HIV1-specific CD8 T cell epitopes and their flanking regions to compare the cleavage predictions obtained with our *N-gram* cleavage models and those obtained using the NetChop web sever. The NetChop system uses an artificial neural-network model

Table 1 Predictive performance of immunoproteasome and proteasome cleavage models

Immunoproteasome					
Size	SE	SP	ECS	MCC	BTR
4	0.807 ± 0.030	0.851 ± 0.039	47.828 ± 2.001	0.660 ± 0.038	0.329 ± 0.016
6	0.763 ± 0.036	0.708 ± 0.042	38.495 ± 0.614	0.472 ± 0.069	0.378 ± 0.016
8	0.906 ± 0.023	0.545 ± 0.038	51.219 ± 1.008	0.484 ± 0.059	0.394 ± 0.011
10	0.802 ± 0.024	0.462 ± 0.019	39.083 ± 1.003	0.281 ± 0.045	0.411 ± 0.012
12	0.903 ± 0.031	0.407 ± 0.031	46.339 ± 0.481	0.357 ± 0.062	0.439 ± 0.014
14	0.872 ± 0.035	0.374 ± 0.023	43.190 ± 1.498	0.284 ± 0.056	0.434 ± 0.017
16	0.855 ± 0.030	0.306 ± 0.041	41.908 ± 1.406	0.193 ± 0.047	0.436 ± 0.015
18	0.857 ± 0.031	0.290 ± 0.028	42.536 ± 1.081	0.179 ± 0.039	0.432 ± 0.015
Proteasome					
Size	SE	SP	ECS	MCC	BTR
4	0.803 ± 0.125	0.871 ± 0.052	44.110 ± 9.249	0.681 ± 0.089	0.362 ± 0.069
6	0.792 ± 0.048	0.723 ± 0.037	36.274 ± 1.943	0.516 ± 0.082	0.429 ± 0.023
8	0.855 ± 0.037	0.603 ± 0.047	38.160 ± 2.112	0.473 ± 0.072	0.473 ± 0.019
10	0.885 ± 0.050	0.537 ± 0.046	37.839 ± 2.355	0.452 ± 0.069	0.506 ± 0.025
12	0.874 ± 0.034	0.534 ± 0.062	34.970 ± 1.704	0.434 ± 0.075	0.526 ± 0.017
14	0.871 ± 0.037	0.468 ± 0.065	33.699 ± 1.432	0.371 ± 0.085	0.532 ± 0.018
16	0.844 ± 0.058	0.403 ± 0.065	32.657 ± 1.692	0.276 ± 0.096	0.518 ± 0.027
18	0.794 ± 0.077	0.392 ± 0.060	27.510 ± 1.978	0.206 ± 0.126	0.519 ± 0.035

Cleavage models were built on peptide fragments of a given size encompassing the C-terminal end of MHCII-eluted ligands (proteasome model) or CD8 T cell epitopes (immunoproteasome) and the corresponding C-terminal flanking residues. Predictive performance was evaluated in 5-fold cross-validation experiments. SE: sensitivity; SP: specificity; MCC: Matthew's correlation coefficient; BTR: better than random (Eq. 4).

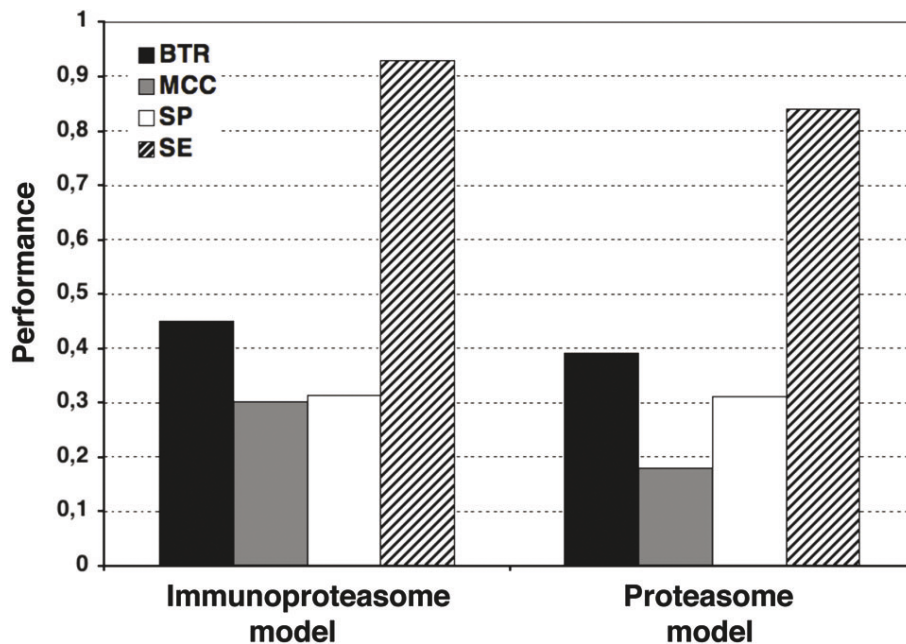
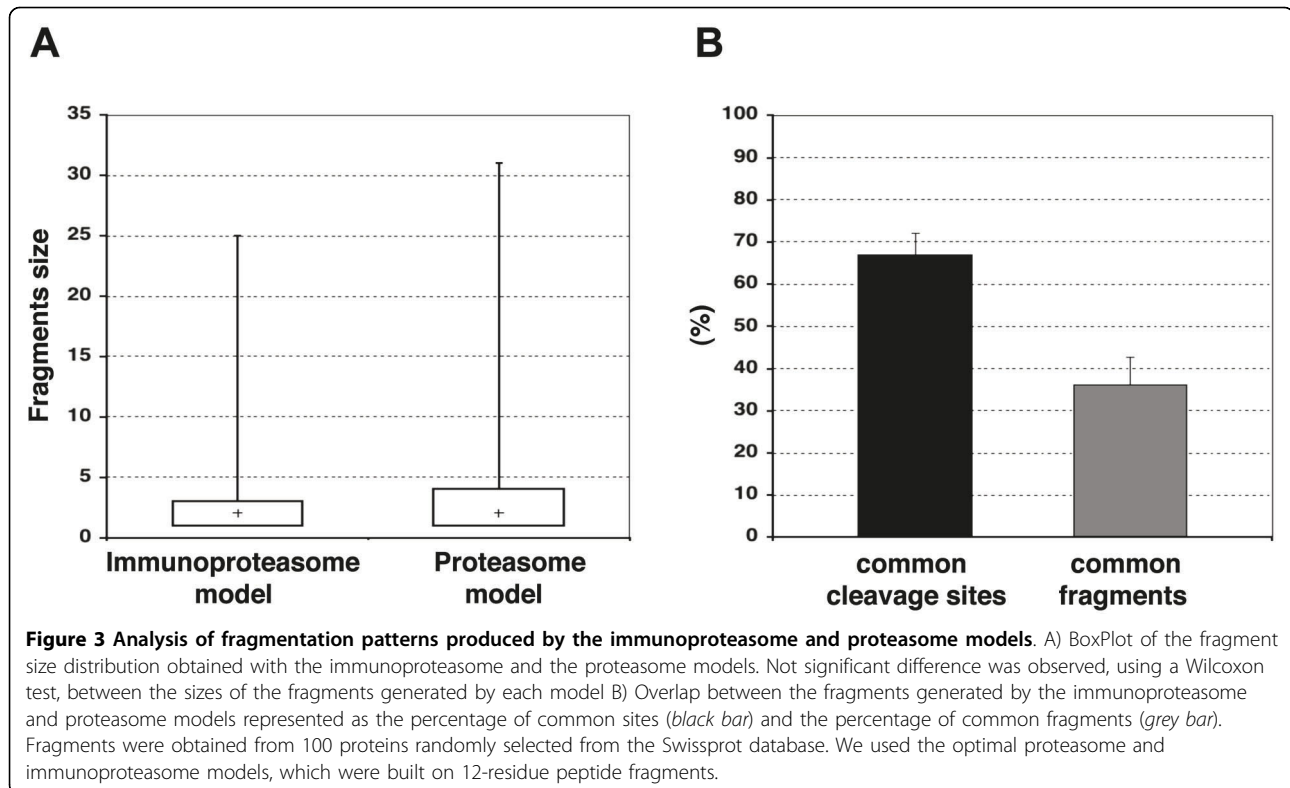


Figure 2 Model evaluation using an independent test dataset. The proteasome and immunoproteasome models were evaluated using an independent test consisting of HIV1-specific CD8 T cell epitopes. The predictive performance was evaluated using BTR (black bars), MCC (grey bars), SP (white bars) and SE (pattern bars).



that was built upon MHC-I-restricted peptides [21]. For this comparison, we used NetChop default settings (cleavage sites occur after residues having a probability of 0.5 or higher) in computing the *SE*, *SP*, and *MCC* of the predictions following the same schema reported by the NetChop developers [22] (see Methods section for details). In addition, we computed the *BTR* parameter defined in this study. Because NetChop models were trained on 18-residue peptide fragments consisting of full-length MHC-I-restricted peptides (9 residues) and the most proximal 9 residues flanking the C-terminus, in this comparison we evaluated *SE*, *SP*, *MCC* and *BTR* on peptide fragments consisting of the full-length HIV1-specific CD8 T cell epitopes. Note that in previous analyses these parameters were evaluated on the portion of the peptide fragments corresponding to the MHC-I-restricted peptides. The results of this analysis are depicted in Figure 4. The immunoproteasome and proteasome *N-gram* models achieved *MCC* values (0.20 and 0.19, respectively) similar to those obtained using NetChop (0.18). Likewise, NetChop and our *N-gram* models achieved similar *BTR* values around 0.44 (Figure 4).

Combination of MHC-I-peptide binding and cleavage predictions

We also evaluated the impact of combining cleavage and MHC-I-peptide binding predictions on T cell epitope

identification. Specifically, using a ROC analysis (see Methods section for details), we analyzed the result of such combination to discriminate CD8 T cell epitopes restricted by 5 different MHC-I molecules (A*0201, A*0301, A*2402, B*0702 and B*2705) from random peptides. We combined MHC-I-peptide binding predictions with cleavage predictions by the immunoproteasome and proteasome models, individually or together, and used *AUC* values (computed after the ROC analyses, see Methods for details) as a measure of the goodness of the predictions (Figure 5).

MHC-I-peptide binding predictions alone achieved high *AUC* values above 0.9 -regardless of the MHC-I molecule-, that did not leave much margin to observe any large improvements on CD8 T cell epitope predictions. Nevertheless, combining the proteasome and immunoproteasome models separately or together with MHC-I-peptide binding predictions resulted in increased *AUC* values (Figure 5B). Moreover, such increases were statistically significant ($p < 0.05$) in all cases. The major increment in *AUC* was observed for A*0301-restricted epitopes. Alone, MHC-I-peptide binding predictions reached an $AUC = 0.9063 \pm 0.0141$ for A*0301, whereas in combination with the immunoproteasome and proteasome cleavage predictions achieved *AUC* values of 0.9416 ± 0.017 , and 0.9411 ± 0.0095 , respectively.

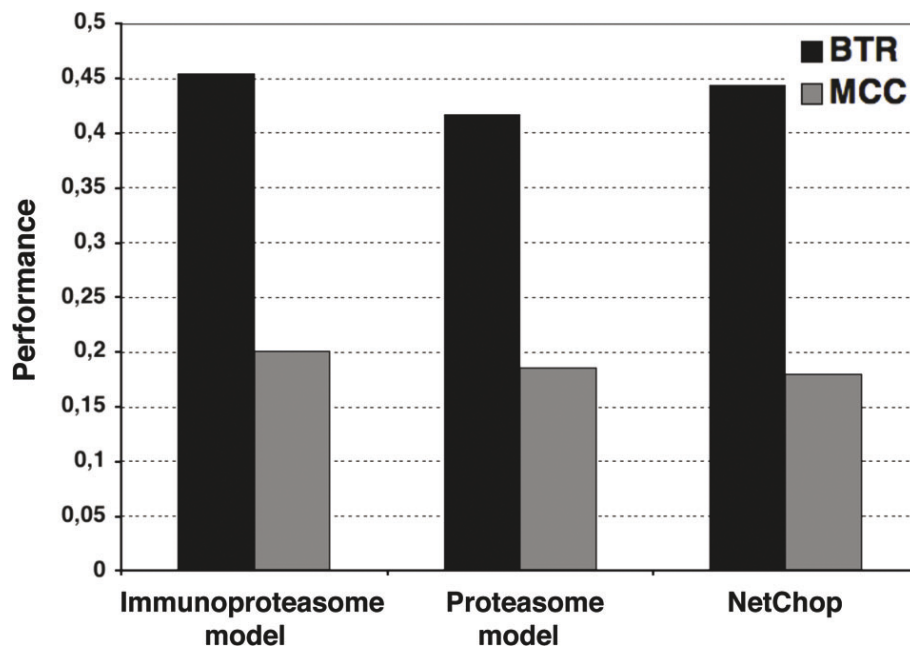


Figure 4 Comparative analysis of cleavage predictions. The figure depicts the MCC (black bars) and the BTR (grey bars) achieved by our immunoproteasome and proteasome models and NetChop on an independent test set of 137 HIV1-specific CD8 T cell epitopes. Because NetChop was built using complete nonameric MHC1-restricted peptides, in this analysis we have evaluated the cleavage predictions by the three models over the entire length of the T cell epitopes being tested.

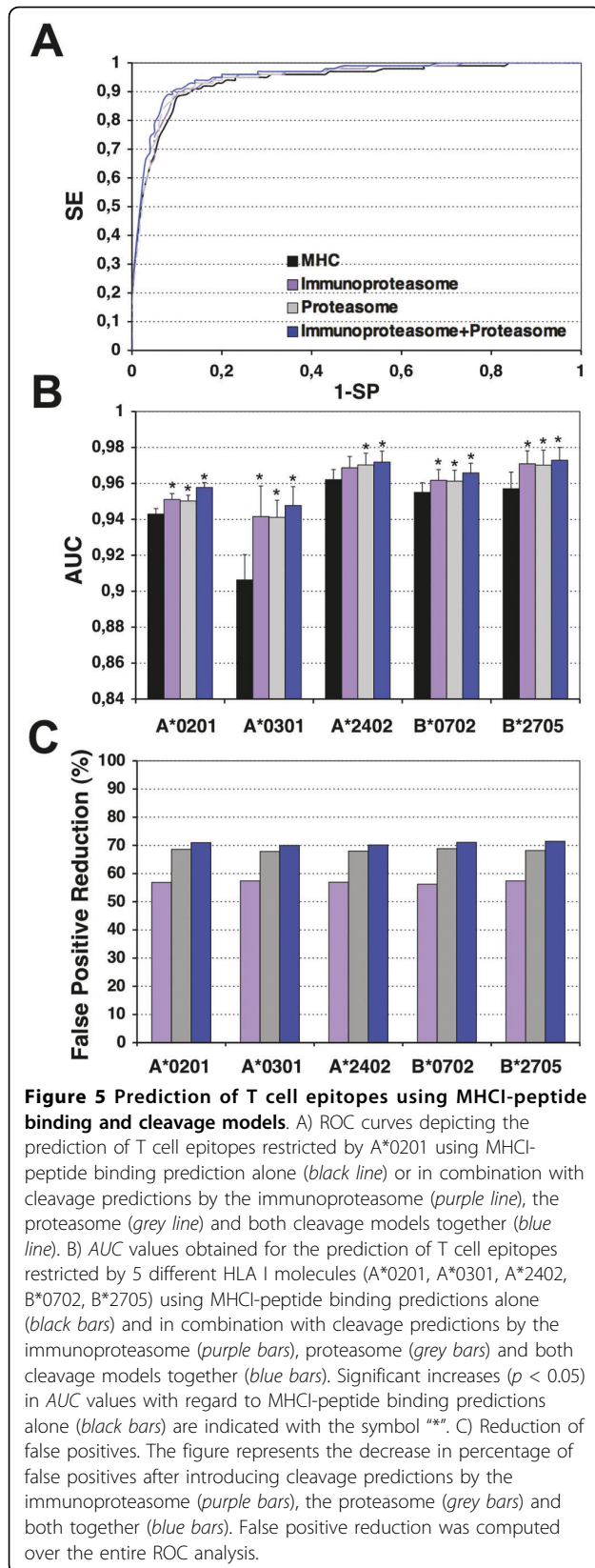
There were no differences between the results obtained combining MHC1-peptide binding and the cleavage predictions by the immunoproteasome model or the proteasome model, but the joint combination of both cleavage models (immunoproteasome and proteasome) with the MHC1-peptide binding resulted in *AUC* values larger than those obtained using single cleavage models (Figure 5B). Nevertheless, with the exception of A*0201 ($p < 0.05$), these increases in *AUC* were not statistically significant with regard to those *AUC* obtained using solely either cleavage model (Figure 5B).

Enhanced *AUC* values obtained upon combining the cleavage models with MHC1-peptide binding predictions are due to the reduction of the number of false positives detected with regard to the MHC1-peptide binding predictions alone (Figure 5C). Taking MHC1-peptide binding predictions alone as reference, we observed a ~56% decrease of false positives (computed over the entire range of thresholds used in the *ROC* analysis) when using the immunoproteasome model. The reduction of false positives was even larger (68%) when using the proteasome model and increased slightly when both models were combined (70%).

Proteasome Cleavage Prediction Server (PCPS)

We developed PCPS (Proteasome Cleavage Prediction Server) to allow the prediction of proteasome and immunoproteasome cleavage through our *N-gram*

models. PCPS is available for free public use at <http://imed.med.ucm.es/Tools/PCPS/>. PCPS was designed to be intuitive and user friendly (Figure 6A). The main input data for PCPS is one or several protein sequences that can be pasted or uploaded to the server in multiple formats, including FASTA, IG, GenBank, EMBL, Phylip, NBRF, GCG, DNASTrider, PIR, MSF, ASN and PAUP. The sequences provided to the server are subjected to a cleavage analysis using *N-gram* models that are selected by the user from the CLEAVAGE MODELS section. There are several models available for both proteasomes, constitutive and immunoproteasome, which differ in sensitivity and specificity, and users can combine different proteasome and immunoproteasome models. Cleavage models in PCPS were trained on peptide fragments of 12 (*models 1*), 8 (*models 2*) and 6 (*models 3*) residues. The models trained on 12 residues exhibited the best performance ($MCC = 0.43 \pm 0.07$ for the proteasome cleavage model and $MCC = 0.36 \pm 0.06$ for the immunoproteasome cleavage model) (Table 1). The output of PCPS consists of a table indicating the cleavage score of each residue in the protein queries (Figure 6B). Computed scores reflect the likelihood that the proteasome/immunoproteasome would cleave the protein after such residue (*PI* residue of cleavage site). Whenever the cleavage score is higher than 0.5, a tick marks the corresponding residue. The different models actually differ in the sensitivity, specificity, and BTR of the predictions.



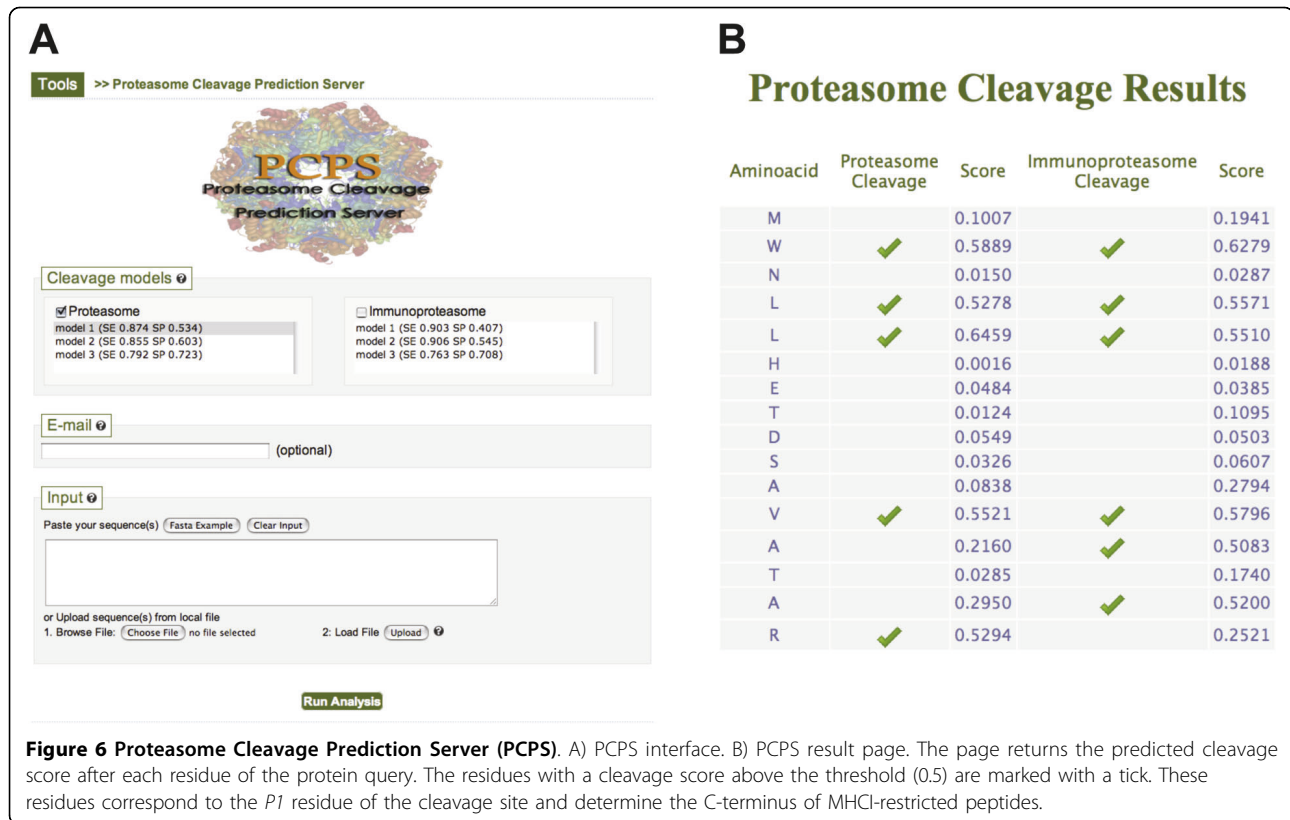
In PCPS, the indicated specificity and sensitivity of the models were achieved at cleavage thresholds of 0.5, but users can experiment with the server and decide different cleavage thresholds.

Discussion

It is generally believed that the C-terminus of most CD8 T cell epitopes, and in general that of most MHC I-restricted peptides, results from the proteolytic cleavage by the proteasome [4,9]. Some other proteases, chiefly tripeptidyl peptidase II (TPP II), also play some role generating the C-terminus of some MHC I-restricted peptides [42-44], specifically through the degradation of some proteolytic products generated by the proteasome that are longer than 15 residues [43]. However, because the majority of the peptide fragments generated by the proteasome are shorter than 15 residues [13], the proteasome is still the principal source of the C-terminus of peptides that are bound to MHC I molecules. As a result, proteasome cleavage models can be derived using cleavage sites recreated from MHC I-restricted peptides and their C-terminal flanking regions [21-23].

There are two types of proteasomes, the immunoproteasome and the constitutive proteasome, which differ in their cleavage patterns [14]. The constitutive proteasome is the form expressed in most nucleated cells, whereas the immunoproteasome is constitutively expressed in mature dendritic cells. Antigen presentation by dendritic cells is generally required to prime and instruct naïve CD8 T cells in an antigen specific manner. Subsequently, the effector function of CD8 T cells is executed upon recognizing the same antigenic peptides on target cells [45]. Consequently, the immunoproteasome is responsible for the generation of the C-terminus of the peptides that elicit the CD8 T cell response, whereas the constitutive proteasome determines the C-terminus of the MHC I-peptide ligands that can be the targets of such response. Protective CD8 T cell epitopes are likely those generated by both, the constitutive proteasome and the immunoproteasome [15].

In this work, we have assumed that MHC I-eluted peptides reflect protein degradation by the proteasome, whereas bona fide identified CD8 T cell epitopes elicited in patients during the course of an infection reflect protein degradation by the immunoproteasome, but not necessarily by the proteasome. The latter is due to the fact that epitope verification is generally carried out by measuring the response of T cells to synthetic peptides loaded onto antigen presenting cells, which are seldom dendritic, thus bypassing antigen processing by the proteasome in the test target cells. Subsequently, using *N-grams*, we have modeled the proteasome and immunoproteasome cleavage



from datasets of peptide fragments of different length built upon MHC-I-eluted peptides (proteasome model) and CD8 T cell epitopes (immunoproteasome model) and their C-terminal flanking regions. These models predict whether the C-terminus of a given peptide, in the context of its flanking residues, is likely to result from the proteolytic activity of the proteasome and/or the immunoproteasome (*P1* residue of cleavage site).

The best cleavage predictions for both proteasomes, constitutive and immunoproteasome, were obtained using *N-grams* trained on 12-residue peptide fragments, encompassing the 6 most proximal flanking residues to the C-terminus of the MHC-I-restricted peptides preceded by 6 residues from the C-terminal end of the MHC-I-restricted peptides (Figure 1). These results are consistent with reports indicating that proteasomes and immunoproteasomes scrutinize between 10 and 12 residues [10,26]. In contrast, related methods for the prediction of proteasome cleavage that are based on MHC-I-restricted peptides have been trained using 18 to 20 residue peptide fragments [19,21,23], which, makes these models, regardless of the results, somewhat artificial.

In cross-validation, the predictive performance of proteasome models exceeded that of immunoproteasome models; the best proteasome cleavage model achieved a

$BTR = 0.53 \pm 0.02$ and an $MCC = 0.43 \pm 0.07$, whereas the best immunoproteasome model achieved a $BTR = 0.44 \pm 0.01$ and an $MCC = 0.36 \pm 0.06$. Despite that both sets of peptides were subjected to the same sequence reduction procedure (See Methods), these results likely reflect that the set of CD8 T cell epitopes is more numerous and arguably more diverse than the set of MHC-I-eluted peptide (see Results). Dendritic cells exhibit non-classical pathways on antigen presentation and some can be immunoproteasome independent [45,46], which could actually account for a higher diversity in the epitope dataset. Nonetheless, the best immunoproteasome model achieved better results than the corresponding proteasome model when predicting the cleavage sites encompassed by an independent set consisting of HIV1-specific CD8 T cell epitopes (Figure 2). Taking into account all the above, the immunoproteasome model appears to be the most suitable to predict the C-terminus of CD8 T cell epitopes.

Our constitutive proteasome and immunoproteasome models produced different but overlapping fragmentation patterns that mirror those observed experimentally [13]; 68% of the cleavage sites (*P1* residues) and 36% of the fragments generated were identical (Figure 3B). However, the fragments yielded by the immunoproteasome and proteasome models were much smaller (2-3 residues)

than those determined experimentally (7-9 residues) [13,47]. The smaller fragment sizes produced by our models may reproduce the clustering and overlapping of epitopes found in protein regions [48]. On the other hand, it is important to note that our models are not meant, and are not suitable, to predict proteolytic fragments, but to indicate whether the C-terminus of a peptide can result from the cleavage produced by the proteasome and/or the immunoproteasome. Proteasome fragmentation patterns (the size of fragments) may be better reproduced by methods trained on actual cleavage data such as that by Tenzer *et al* [18].

Using a test set of HIV1-specific CD8 T cell epitopes, we found that the predictive performance of our optimal proteasome and immunoproteasome cleavage models was comparable to that of NetChop [22]; a reference method to predict proteasome cleavage sites [24] that it was also developed from MHCI-restricted peptides. The immunoproteasome and proteasome cleavage models achieved *MCC* values of 0.20 and 0.19, respectively, while NetChop achieved an *MCC* = 0.18. It is worth nothing that these results were obtained under conditions that were optimal for NetChop. First, NetChop was trained on peptide fragments encompassing full length MHCI-restricted peptides [22], and here we have evaluated and compared the cleavage predictions over the entire epitope sequences. Note that we only used a portion of the MHCI-restricted peptides for training (6 residues). Second, the HIV1-specific CD8 T cell epitopes used for testing were not used for training our *N-gram* models but were likely included in the NetChop training dataset. It is also important to mention that NetChop has been described as an immunoproteasome cleavage prediction method, but in fact it was trained on a dataset consisting of both, MHCI-eluted ligands and CD8 T cell epitopes. As we have discussed here, CD8 T cell epitopes can be considered as generated by the immunoproteasome. However, it is more appropriated to consider MHCI-eluted peptides as generated by the constitutive proteasome because they are obtained from different type of cells but seldom from dendritic cells. In sum, we have dealt with the prediction of proteasome and immunoproteasome cleavage sites from MHC-restricted peptides in a manner that is consistent with the mechanism of antigen presentation and recognition, and achieved a notorious performance.

Prediction of proteasome and immunoproteasome cleavage sites using our models is available at <http://imed.med.ucm.es/Tools/PCPS/>. In addition, there are several other online servers to predict proteasome cleavage, which differ in the data and approach used for generating the models [16,22,49]. Nonetheless, the problem of identifying proteasome cleavage sites with high precision is

still far from being solved. A simple manner to improve the prediction of proteasome cleavage sites could likely be achieved through a meta-server that would arrive to a consensus prediction from the available proteasome cleavage predictors. Such a consensus approach has resulted successful in the also difficult task of predicting peptide binding to MHC class II molecules [50].

It has been reported that proteasome prediction models can improve T cell epitope identification when combined with MHCI-peptide binding predictions [18,22,37,51,52]. Likewise, our proteasome and immunoproteasome models, separately or together, also served to improve CD8 T cell epitope discrimination when combined with MHCI-binding predictions (Figure 5). The improvements, judged by increases in *AUC*, could appear minor but were statistically significant (Figure 5), and were linked to a large reduction of the number of false positives detected (up to 70%). Therefore, combining proteasome cleavage and MHCI-peptide binding predictions would serve to decrease the experimental toll involved in epitope identification; there will be less peptides to be tested. The proteasome cleavage model alone or juxtaposed with the immunoproteasome model resulted in a significant loss of true positives (up to 20%). Therefore, the proteasome cleavage model will be more useful on large-scale epitope identification scenarios (e.g. predicting CD8 T cell epitopes from a large number of antigens). Finally, combining cleavage predictions by both proteasomes, constitutive and immunoproteasome, with MHCI-binding predictions ought to help defining protective CD8 T cell epitopes. Overall, these results call for the integration of our proteasome models with others taking into account TAP transport and MHC binding, as already pioneered by other authors [18,22,37,51,52].

Conclusion

We have derived *N-gram* models specific for the proteasome and the immunoproteasome that are consistent with the known biology of antigen presentation. The proteasome models were built upon MHCI-eluted peptides whereas the immunoproteasome models were built upon CD8 T cell epitopes. The *N-gram* models that exhibited the best performance were trained on 12-residue peptides, 6 residues at each side of the cleavage site, defined by the C-terminus of MHCI-restricted peptides and the most proximal C-terminal flanking residue. Finally, we have shown that combining cleavage predictions by the proteasome and immunoproteasome models with MHCI-binding predictions improves CD8 T cell epitope prediction. Cleavage predictions using our *N-gram* models are available for free public use at the PCPS site <http://imed.med.ucm.es/Tools/PCPS/>.

Additional material

Additional file 1: MHC I allele distribution in peptide datasets. The figure depicts the percentage of peptides restricted by 7 commonly expressed human MHC I alleles (A*0201, A*0301, A*1101, A*2402, B*0702, B*0801, B*2705) in the three datasets used in this study.

Abbreviations used

MHC: I molecules, major histocompatibility class I molecules; N-terminus: amino-terminus; C-terminus: carboxy-terminus.

Authors' contributions

CMDR did the work and wrote paper. EML interpreted results and wrote paper. PAR designed the work, interpreted results and rendered the final paper. All authors read and approved the final manuscript.

Acknowledgements

We wish to thank Dr Elena Rodriguez-Garcia for corrections and thoughtful comments. This work was supported by Grants SAF2006-07879 and SAF2009-08103 from Ministerio de Ciencia e Innovación of Spain, and by Grant CCG08-UCM/BIO-3769 from Comunidad Autonoma de Madrid to PAR.

Author details

¹Laboratory of Immunomedicine, Department of Microbiology I-Immunology, Facultad de Medicina, Universidad Complutense de Madrid, Ave Complutense S/N, Madrid 28040, Spain. ²Department of Microbiology I-Immunology, Facultad de Medicina, Universidad Complutense de Madrid, Ave Complutense S/N, Madrid 28040, Spain.

Received: 6 May 2010 Accepted: 23 September 2010

Published: 23 September 2010

References

- Garcia KC, Teyton L, Wilson IA: Structural basis of T cell recognition. *Annu Rev Immunol* 1999, **17**:369-397.
- Margulies DH: Interactions of TCRs with MHC-peptide complexes: a quantitative basis for mechanistic models. *Curr Opin Immunol* 1997, **9**(3):390-395.
- Wang J-H, Reinherz E: Structural basis of T cell recognition of peptides bound to MHC molecules. *Molecular Immunology* 2001, **38**:1039-1049.
- Pamer E, Cresswell P: Mechanisms of MHC class I-restricted antigen processing. *Annu Rev Immunol* 1998, **16**:323-358.
- Kloetzel PM: Antigen processing by the proteasome. *Nat Rev Mol Cell Biol* 2001, **2**(3):179-187.
- Serwold T, Gonzalez F, Kim J, Jacob S, Shastri N: ERAAP customizes peptides for MHC class I molecules in the endoplasmic reticulum. *Nature* 2002, **419**(6906):480-483.
- Craiu A, Akopian T, Goldberg A, Rock KL: Two distinct proteolytic processes in the generation of a major histocompatibility complex class I-presented peptide. *Proc Natl Acad Sci USA* 1997, **94**(20):10850-10855.
- Rock KL, Gramm C, Rothstein L, Clark K, Stein R, Dick L, Hwang D, Goldberg AL: Inhibitors of the proteasome block the degradation of most cell proteins and the generation of peptides presented on MHC class I molecules. *Cell* 1994, **78**(5):761-771.
- Rock KL, Goldberg AL: Degradation of cell proteins and the generation of MHC class I-presented peptides. *Annu Rev Immunol* 1999, **17**:739-779.
- Nussbaum AK, Dick TP, Keilholz W, Schirle M, Stevanović S, Dietz K, Heinemeyer W, Groll M, Wolf DH, Huber R, et al: Cleavage motifs of the yeast 20 S proteasome β subunits deduced from digests of enolase 1. *Proc Natl Acad Sci* 1998, **95**:12504-12509.
- Groettrup M, Standera S, Stohwasser R, Kloetzel PM: The subunits MECL-1 and LMP2 are mutually required for incorporation into the 20 S proteasome. *Proc Natl Acad Sci USA* 1997, **94**(17):8970-8975.
- Morel S, Levy F, Burlet-Schiltz O, Brasseur F, Probst-Kepper M, Peitrequin AL, Monsarrat B, Van Velthoven J, Cerottini JC, Boon T, et al: Processing of some antigens by the standard proteasome but not by the immunoproteasome results in poor presentation by dendritic cells. *Immunity* 2000, **12**(1):107-117.
- Toes E, Nussbaum AK, Degermann S, Schirle M, Emmerich NP, Kraft M, Laplace C, Zwinderman A, Dick TP, Muller J, et al: Discrete cleavage motifs of constitutive and immunoproteasomes revealed by quantitative analysis of cleavage products. *J Exp Med* 2001, **194**(1):1-12.
- Gaczynska M, Rock K, Spies T, Goldberg A: Peptidase activities of proteasomes are differentially regulated by the major histocompatibility complex-encoded genes for LMP2 and LMP7. *Proc Natl Acad Sci USA* 1994, **91**(20):9213-9217.
- Chapiro J, Claverol S, Piette F, Ma W, Stroobant V, Guillaume B, Gairin JE, Morel S, Burlet-Schiltz O, Monsarrat B, et al: Destructive cleavage of antigenic peptides either by the immunoproteasome or by the standard proteasome results in differential antigen presentation. *J Immunol* 2006, **176**(2):1053-1061.
- Nussbaum A, Kuttler C, Hadel K, Rammensee H, Schild H: PProC: a prediction algorithm for proteasomal cleavages available on the WWW. *Immunogenetics* 2001, **53**(2):87-94.
- Kuttler C, Nussbaum AK, Dick TP, Rammensee HG, Schild H, Hadel KP: An algorithm for the prediction of proteasomal cleavages. *J Mol Biol* 2000, **298**(3):417-429.
- Tenzen S, Peters B, Bulik S, Schoor O, Lemmel C, Schatz MM, Kloetzel PM, Rammensee HG, Schild H, Holzhtutter HG: Modeling the MHC class I pathway by combining predictions of proteasomal cleavage, TAP transport and MHC class I binding. *Cell Mol Life Sci* 2005, **62**(9):1025-1037.
- Holzhtutter H, Frömmel C, Kloetzel P: A theoretical approach towards the identification of cleavage-determining amino acid motifs of the 20 S proteasome. *J Mol Biol* 1999, **286**(4):1251-1265.
- Holzhtutter HG, Kloetzel PM: A kinetic model of vertebrate 20 S proteasome accounting for the generation of major proteolytic fragments from oligomeric peptide substrates. *Biophys J* 2000, **79**(3):1196-1205.
- Kesmir C, Nussbaum AK, Schild H, Detours V, Brunak S: Prediction of proteasome cleavage motifs by neural networks. *Protein Eng* 2002, **15**(4):287-296.
- Nielsen M, Lundegaard C, Lund O, Kesmir C: The role of the proteasome in generating cytotoxic T-cell epitopes: insights obtained from improved predictions of proteasomal cleavage. *Immunogenetics* 2005, **57**(1-2):33-41.
- Bhasin M, Raghava GPS: Pcleavage: an SVM based method for prediction of constitutive proteasome and immunoproteasome cleavage sites in antigenic sequences. *Nucleic Acids Research* 2005, **33**:W202-W207.
- Saxová P, Buus S, Brunak S, Kesmir C: Predicting proteasomal cleavage sites: a comparison of available methods. *International Immunology* 2003, **15**(7):781-787.
- Stolcke A: SRILM - An Extensible Language Modeling Toolkit. In *Proceedings of the International Conference of Spoken Language Processing*. Edited by: JJ Ohala TMN, BL Derwing M, Hodge M, Wiebe GE. Boulder, CO: Center for Spoken Language Research; 2002:2901-904.
- Altuvia Y, Margalit H: Sequence signals for generation of antigenic peptides by the proteasome: implications for proteasomal cleavage mechanism. *J Mol Biol* 2000, **295**(4):879-890.
- Reche PA, Zhang H, Glutting JP, Reinherz EL: EPIMHC: a curated database of MHC-binding peptides for customized computational vaccinology. *Bioinformatics* 2005, **21**(9):2140-2141, Epub 2005 Jan 21 18.
- Peters B, Sidney J, Bourne P, Bui H, Buus S, Doh G, Fleri W, Kronenberg M, Kubo R, Lund O, et al: The immune epitope database and analysis resource: from vision to blueprint. *PLoS Biol* 2005, **3**(3):e91.
- HIV Molecular Immunology 2006/2007. Los Alamos, New Mexico: Los Alamos National Laboratory, Theoretical Biology and Biophysics 2007.
- Neuwald AF, Liu JS, Lawrence CE: Gibbs motif sampling detection of bacterial outer membrane protein repeats. *Prot Sci* 1995, **4**:1618-1632.
- Matthews B: Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochim Biophys Acta* 1975, **405**:442-451.
- Reche PA, Glutting J-P, Reinherz EL: Enhancement to the RANKPEP resource for the prediction of peptide binding to MHC molecules using profiles. *Immunogenetics* 2004, **56**:405-419.
- Lafuente EM, Reche PA: Prediction of MHC-peptide binding: a systematic and comprehensive overview. *Curr Pharm Des* 2009, **15**(28):3209-3220.
- Reche PA, Glutting JP, Reinherz EL: Prediction of MHC class I binding peptides using profile motifs. *Hum Immunol* 2002, **63**(9):701-709.

35. Reche PA, Reinherz EL: **Prediction of peptide-MHC binding using profiles.** *Methods Mol Biol* 2007, **409**:185-200.
36. Swets JA: **Measuring the accuracy of diagnostic systems.** *Science* 1988, **240(4857)**:1285-1293.
37. Dönnes P, Kohlbacher O: **Integrated modeling of the major events in the MHC class I antigen processing pathway.** *Protein Science* 2005, **14(8)**:2132-2140.
38. Rosenfeld : **Two decades of statistical language modeling: Where do we go from here?** *Proceedings of the IEEE* 2000, **88(8)**:1-11.
39. Jimenez-Montano MA, Ebeling W, Pohl T, Rapp PE: **Entropy and complexity of finite sequences as fluctuating quantities.** *Biosystems* 2002, **64(1-3)**:23-32.
40. Wu C, Shivakumar S: **Back-propagation and counter-propagation neural networks for phylogenetic classification of ribosomal RNA sequences.** *Nucleic Acids Res* 1994, **22(20)**:4291-4299.
41. Wu CH, Zhao S, Chen HL, Lo CJ, McLarty J: **Motif identification neural design for rapid and sensitive protein family search.** *Comput Appl Biosci* 1996, **12(2)**:109-118.
42. Kloetzel PM: **Generation of major histocompatibility complex class I antigens: functional interplay between proteasomes and TPII.** *Nat Immunol* 2004, **5(7)**:661-669.
43. Reits E, Neijssen J, Herberts C, Benckhuijsen W, Janssen L, Drijfhout JW, Neefjes J: **A major role for TPII in trimming proteasomal degradation products for MHC class I antigen presentation.** *Immunity* 2004, **20(4)**:495-506.
44. Yewdell JW, Princiotta MF: **Proteasomes get by with lots of help from their friends.** *Immunity* 2004, **20(4)**:362-363.
45. Heath W, Belz T, Behrens G, Smith C, Forehan S, Parish I, Davey G, Wilson N, Carbone F, Villadangos J: **Cross-presentation, dendritic cell subsets, and the generation of immunity to cellular antigens.** *Immunological Reviews* 2004, **199(1)**:9-26.
46. Banchereau J, Briere F, Caux C, Davoust J, Lebecque S, Liu Y, Pulendran B, Palucka K: **Immunobiology of dendritic cells.** *Annu Rev Immunol* 2000, **18**:767-811.
47. Kisselev AF, Akopian TN, Woo KM, Goldberg AL: **The sizes of peptides generated from protein by mammalian 26 and 20 S proteasomes. Implications for understanding the degradative mechanism and antigen presentation.** *J Biol Chem* 1999, **274(6)**:3363-3371.
48. Meister GE, Roberts CG, Berzofsky JA, De Groot AS: **Two novel T cell epitope prediction algorithms based on MHC-binding motifs; comparison of predicted and published epitopes from Mycobacterium tuberculosis and HIV protein sequences.** *Vaccine* 1995, **13(6)**:581-591.
49. Ginodi I, Vider-Shalit T, Tsaban L, Louzoun Y: **Precise score for the prediction of peptides cleaved by the proteasome.** *Bioinformatics* 2008, **24(4)**:477-483, Epub 2008 Jan 2023.
50. Wang P, Sidney J, Dow C, Mothe B, Sette A, Peters B: **A systematic assessment of MHC class II peptide binding predictions and evaluation of a consensus approach.** *PLoS Comput Biol* 2008, **4(4)**:e1000048.
51. Doytchinova IA, Flower CR: **Class I T-cell epitope prediction: Improvements using a combination of proteasome cleavage, TAP affinity, and MHC binding.** *Molecular Immunology* 2006, **43(13)**:2037-2044.
52. Larsen MV, Lundegaard C, Lamberth K, Buus S, Brunak S, Lund O, Nielsen M: **An integrative approach to CTL epitope prediction: a combined algorithm integrating MHC class I binding, TAP transport efficiency, and proteasomal cleavage predictions.** *Eur J Immunol* 2005, **35(8)**:2295-2303.

doi:10.1186/1471-2105-11-479

Cite this article as: Diez-Rivero et al.: Computational analysis and modeling of cleavage by the immunoproteasome and the constitutive proteasome. *BMC Bioinformatics* 2010 **11**:479.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

