

Proteasomal cleavage site prediction of protein antigen using BP neural network based on a new set of amino acid descriptor

Yuanqiang Wang · Yong Lin · Mao Shu · Rui Wang · Yong Hu · Zhihua Lin

Received: 20 November 2012 / Accepted: 18 March 2013 / Published online: 13 April 2013
© Springer-Verlag Berlin Heidelberg 2013

Abstract The accurate identification of cytotoxic T lymphocyte epitopes is becoming increasingly important in peptide vaccine design. The ubiquitin–proteasome system plays a key role in processing and presenting major histocompatibility complex class I restricted epitopes by degrading the antigenic protein. To enhance the specificity and efficiency of epitope prediction and identification, the recognition mode between the ubiquitin–proteasome complex and the protein antigen must be considered. Hence, a model that accurately predicts proteasomal cleavage must be established. This study proposes a new set of parameters to characterize the cleavage window and uses a backpropagation neural network algorithm to build a model that accurately predicts proteasomal cleavage. The accuracy of the prediction model, which depends on the window sizes of the cleavage, reaches 95.454 % for the *N*-terminus and 95.011 % for the *C*-terminus. The results show that the identification of proteasomal cleavage sites depends on the sequence next to it and that the prediction performance of the *C*-terminus is better than that of the *N*-terminus on average. Thus, models based on the properties of amino acids

can be highly reliable and reflect the structural features of interactions between proteasomes and peptide sequences.

Keywords Backpropagation (BP) neural network · Cleavage site · Epitope · Proteasome

Introduction

With the development of tumor immunology, several studies have demonstrated that cytotoxic T lymphocytes (CTLs) play an important role in antitumor immunity. Given the tumor-specific killing ability of CTLs, they are used to treat malignant tumors [1]. As an integrity antigen, the peptide vaccine for CTL epitope poses no risk for infection and can be massively synthesized and oriented for antitumor and antiviral activities. Thus, the accurate identification of CTL epitopes has become increasingly important, and many studies are focused on developing advanced methods for CTL epitope screening [2–4].

The prediction of proteasomal cleavage sites and peptides transported by transporter associated protein (TAP) has not yet been considered. Thus, the current methods for CTL epitope prediction are not perfect, and most of them are constructed based only on the affinity between peptides of CTL epitopes and major histocompatibility complex (MHC) class I (MHC-I) molecules. Accordingly, an effective CTL epitope must undergo the following translocation mechanism in living cells. First, the CTL epitope is produced by proteasomes from the protein antigen and transported by TAP to the endoplasmic reticulum to assemble with the MHC-I complexes. Second, the product is excreted by the Golgi apparatus to the surface of the cell membrane and made available for interaction with T cell receptors. In the CTL response, these intermolecular recognition processes are connected like a chain, and any

Electronic supplementary material The online version of this article (doi:10.1007/s00894-013-1827-7) contains supplementary material, which is available to authorized users.

Y. Wang · Y. Lin · M. Shu · R. Wang · Y. Hu · Z. Lin (✉)
College of Pharmacy and Bioengineering, Chongqing University of Technology, Chongqing 400050, People's Republic of China
e-mail: zhlin@cqut.edu.cn

Y. Wang
State Key Laboratory of Silkworm Genome Biology, Southwest University, Chongqing 400715, People's Republic of China

Z. Lin
School of Chemistry and Chemical Engineering, Chongqing University, Chongqing 400044, People's Republic of China

errors in them invalidate the whole response. Several studies have also found that some peptides with high binding affinities for MHC-I molecules do not elicit a strong CTL response [5]. Therefore, predicting the binding affinity between peptides and MHC-I molecules is insufficient; a prediction method that combines the prediction of proteasomal cleavage sites, TAP transport peptides, and binding affinity of epitope peptides must be established to improve the accuracy of epitope prediction and identification.

Proteasomes are intracellular multi-subunit proteases that catalyze selective proteolytic protein processing within various cell signal-transduction pathways, such as cell cycle control, transcriptional regulation, and antigen presentation [6–8]. Proteasomes exist in several forms. For example, the proteolytically active core complexes or 20S proteasomes, association with ATP-dependent 19S cap complexes yields the larger 26S proteasomes, which can recognize proteins and are marked by ubiquitin for proteasomal degradation [9, 10]. Eukaryotic 20S proteasomes consist of four stacked rings (overall stoichiometry $\alpha_7\beta_7\beta_7\alpha_7$) each consisting of seven different subunits [11]. Each of the two inner β -rings carries three catalytically active sites on its inner surface, and proteolytic specificities of active site have been described as chymotrypsin like [cleaving after large, hydrophobic amino acids (AAs)], trypsin like (cleaving after basic AAs), and peptidyl-glutamyl-peptide hydrolyzing (cleaving after acidic AAs) [12]. Proteasomes cleave ubiquitin-protein conjugates into smaller peptides and generate the correct C-terminus of MHC-I-binding peptides [13, 14].

Several computational approaches for elucidating the cleavage specificity of proteasomes have been presented. These approaches are all based on experimentally verified cleavage sites within protein substrates and the analysis of the flanking region of such sites. EpiJen is an integrated approach for predicting proteasomal cleavage sites, transporting TAP, and binding MHC-I based on quantitative matrices [15]. Pierre *et al.* [16] used a new matrix-based method to establish a

proteasomal cleavage site prediction method based on experimentally verified proteasomal cleavage sites. Holzthutter *et al.* [17] used a statistical method based on the MAPPP prediction server to analyze cleavage sites found in a set of peptide substrates with lengths ranging from 22 AAs to 30 AAs. PAPProC [18, 19] is a method based on the proteasomal degradation of enolase protein. Up to 10 flanking AAs around the verified cleavage sites are used by an evolutionary algorithm to create the network-based model used for prediction. Another method for proteasomal cleavage prediction is NetChop, which uses a neural network for prediction [20]. Two different sets of training data are used by NetChop: verified cleavage sites within proteins and naturally processed MHC ligands. Altuvia and Margalit [21] used MHC-I ligands to study proteasomal cleavage.

This study proposes a new set of parameters to characterize the structure of AAs and uses a backpropagation (BP) neural network (BP-NN) to build a model for accurately predicting proteasomal cleavage. The accuracy of the prediction model reaches 95.454 % for the N-terminus and 95.011 % for the C-terminus, which depends on the cleavage window sizes. The results show that the proteasomal cleavage site identification depends on the sequence next to it and that the predictive performance of the C-terminus is better than that of the N-terminus on average.

Materials and methods

Experimental data

A total of 175 epitopes associated with HLA-A*0201 molecules with 8 AAs to 11 AAs were collected from the AntiJen database and used as the positive samples of proteasomal cleavage. The negative samples were constructed according to literature [20]. Briefly, pseudo-epitopes with a N- or C-termini at the middle of the epitopes are randomly produced, which assume the presence of non-cleavage sites within the epitopes.

Table 1 Three descriptor scales for natural amino acids

AA	V ₁	V ₂	V ₃	AA.	V ₁	V ₂	V ₃
Ala (A)	88.3	0.0373	-6.7	Gln (Q)	148.7	0.0761	17.2
Leu (L)	168.5	0.0000	-11.7	Ser (S)	88.7	0.0829	-2.5
Arg (R)	181.2	0.0959	51.5	Glu (E)	140.5	0.0058	34.3
Lys (K)	175.6	0.0371	36.8	Thr (T)	118.2	0.0941	-5.0
Asn (N)	125.1	0.0036	20.1	Gly (G)	60.0	0.005	-4.2
Met (M)	162.2	0.0823	-14.2	Trp (W)	227.0	0.0548	-7.9
Asp (D)	110.8	0.1263	38.5	His (H)	152.6	0.0242	12.6
Phe (F)	189.0	0.0946	-15.5	Tyr (Y)	193.0	0.0516	2.9
Cys (C)	112.4	0.0829	-8.4	Ile (I)	168.5	0.0000	-13.0
Pro (P)	122.2	0.0198	0.8	Val (V)	141.4	0.0057	-10.9

V₁ residue volume, V₂ electron-ion interaction potential values, V₃ hydrophobicity

Table 2 Correlation matrix of parameters

	V ₁	V ₂	V ₃
V ₁	1.000	0.050	0.045
V ₂	0.050	1.000	0.180
V ₃	0.045	0.180	1.000

The cleavage sites that produce the epitopes and pseudo-epitopes are defined as positive (+1) and negative (−1) samples, respectively. These samples are included in the appendix.

Structural descriptors and characterization

Three properties of AAs, including residue volume [22], electron-ion interaction potential values [23], and hydrophobicity [24], are selected to characterize the peptide. The values of the parameters are listed in Table 1, and their correlation matrix is shown in Table 2. Three descriptors are needed for each AA to characterize the peptide structures, and 3×N parameters are needed to characterize the peptide with N AAs. If a cleavage window contains N AAs, the window can be characterized by a set of 3×N descriptors. For example, a window constituting 4 AAs should be represented by 12 descriptors.

Neural network algorithm

Artificial neural networks have been developed as generalizations of mathematical models of biological nervous systems [25, 26]. The present study uses a BP-NN model with three layers. The nodes of the input layer are equal to the descriptors of AAs, and the node of the output layer is the ideal value, *i.e.*, a negative or positive sample represented by −1 or +1, respectively. The nodes of the hidden layer are calculated by the following formula:

$$n_l = \sqrt{n_i + n_o} + 10,$$

where n_l , n_i , and n_o are the nodes of the hidden, input, and output layers, respectively. The weights connected to the input, hidden, and output layers are randomly generated and adjusted using standard BP (a gradient descent type) algorithm with the circular samples presented. Finally, the process is terminated when the minimal error or the maximal iteration is reached. For example, a cleavage window sequence containing 4 AAs should be represented by 12 descriptors, which should all be placed in the input layer. Thus, the BP-NN for the 4 AA cleavage window must have 12 neurons in the input layer.

Table 3 Statistical results of N-terminus

Win.	SE (%)	SP (%)	AC (%)	CC	Win.	SE (%)	SP (%)	AC (%)	CC
0–1	69.280	52.983	61.131	0.500	3–4	90.480	90.657	90.569	0.833
0–2	89.194	91.697	90.446	0.827	3–5	90.703	91.400	91.051	0.840
0–3	91.863	93.046	92.454	0.862	3–6	91.891	91.794	91.843	0.852
0–4	91.954	92.166	92.060	0.856	4–0	60.931	59.823	60.377	0.503
0–5	91.737	91.200	91.469	0.846	4–1	67.857	61.554	64.706	0.526
0–6	91.657	92.131	91.894	0.852	4–2	89.749	89.920	89.834	0.819
1–0	52.474	58.131	55.303	0.462	4–3	90.829	91.566	91.197	0.841
1–1	61.971	58.634	60.303	0.504	4–4	91.697	91.794	91.746	0.850
1–2	89.640	91.509	90.574	0.830	4–5	92.343	91.589	91.966	0.854
1–3	91.331	91.531	91.431	0.846	4–6	92.183	91.869	92.026	0.855
1–4	90.863	91.354	91.109	0.840	5–0	61.800	61.137	61.469	0.502
1–5	91.389	90.949	91.169	0.843	5–1	65.891	62.560	64.226	0.514
1–6	91.560	91.143	91.351	0.844	5–2	89.126	89.531	89.329	0.811
2–0	60.423	59.411	59.917	0.490	5–3	91.451	91.309	91.380	0.844
2–1	66.600	61.549	64.074	0.530	5–4	91.309	91.526	91.417	0.845
2–2	90.354	89.983	90.169	0.826	5–5	92.554	91.417	91.986	0.853
2–3	91.091	92.223	91.657	0.849	5–6	92.960	92.183	92.571	0.863
2–4	90.383	91.606	90.994	0.838	6–0	64.240	61.320	62.780	0.511
2–5	91.189	90.771	90.980	0.838	6–1	65.863	61.897	63.880	0.529
2–6	90.663	91.131	90.897	0.840	6–2	88.720	89.234	88.977	0.806
3–0	57.474	61.857	59.666	0.490	6–3	92.109	91.086	91.597	0.847
3–1	62.069	60.880	61.474	0.500	6–4	91.669	91.503	91.586	0.847
3–2	89.383	90.011	89.697	0.819	6–5	92.051	92.029	92.040	0.855
3–3	90.377	91.080	90.729	0.834	6–6	92.891	92.223	92.557	0.863

Win window sizes, SE sensitivity, SP specificity, AC accuracy, CC coefficient of correlation

Optimization of cleavage windows

The cleavage window is formed with several AAs at the cleavage site and defined as “P_m...P₂P₁|P₁'P₂'...P_n”, where “|” is the cleavage site and *m* and *n* are 1 to 6, respectively. For example, the cleavage window has 5 AAs when *m*=2 and *n*=3, which means that the previous 2 and the next 3 AAs belong to the cleavage window. For each cleavage window, the network is trained for 100 times with all samples, and the statistical results are the average values. The shortest cleavage window, also called the optimized window, is formed after training for every window.

Evaluation of network performance

The network performance is evaluated using the correlation coefficient (CC) [27] as follows:

$$CC = \frac{TP \times TN - FN \times FP}{\sqrt{(TN + FN)(FN + TP)(TP + FP)(FP + TN)}}$$

where *TP* is the number of true positives (experimentally verified cleavage sites also the predicted cleavage sites), *TN* is the number of true negatives (experimentally verified

non-cleavage sites also the predicted non-cleavage sites), *FP* is the number of false positives (experimentally verified non-cleavage sites predicted as cleavage sites), and *FN* is the number of false negatives (experimentally verified cleavage sites predicted as non-cleavage sites). The additional performance measurements used are as follows:

$$SE = \frac{TP}{TP+FN} \times 100 \quad SP = \frac{TN}{TN+FP} \times 100$$

$$AC = \frac{TP+TN}{TP+FP+TN+FN} \times 100.$$

Results and discussion

The proteasomal cleavage window is important for the recognition of cleavage sites. Several studies have selected different windows [18, 20], and the maximum length of window is composed of 12 AAs in the present study. To determine the optimized window size, BP-NN models are built with different cleavage windows. The procedure is similar to window optimization and the final results are the average values of 50 times of training and testing.

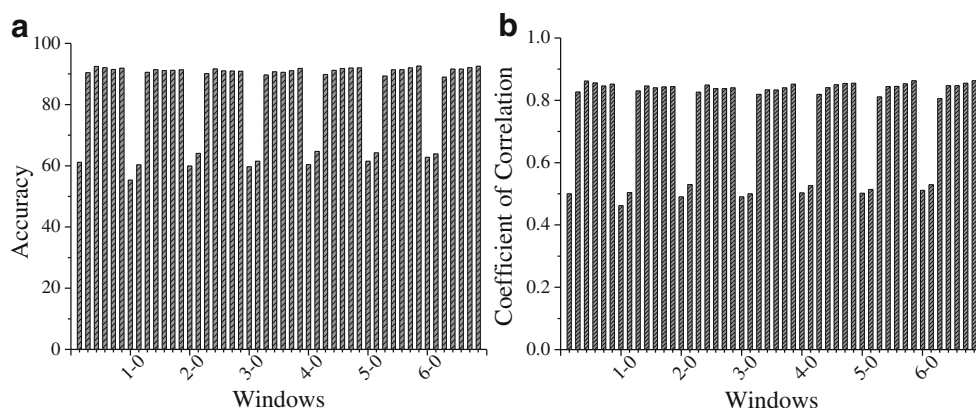
To identify the motif recognized by proteasomes, 48 windows are constructed according to the combination of

Table 4 Statistical results of *C*-terminus

Win.	SE (%)	SP (%)	AC (%)	CC	Win.	SE (%)	SP (%)	AC (%)	CC
0–1	50.851	53.154	52.003	0.448	3–4	95.017	94.200	94.609	0.899
0–2	57.257	58.366	57.811	0.490	3–5	95.097	94.514	94.806	0.902
0–3	56.429	61.800	59.114	0.487	3–6	94.211	94.183	94.197	0.891
0–4	58.463	60.777	59.620	0.483	4–0	93.937	94.817	94.377	0.896
0–5	62.086	58.589	60.337	0.486	4–1	94.966	94.817	94.891	0.904
0–6	63.703	62.240	62.971	0.515	4–2	94.509	93.971	94.240	0.893
1–0	94.234	97.589	95.911	0.922	4–3	94.674	94.029	94.351	0.894
1–1	95.337	97.777	96.557	0.934	4–4	95.143	94.389	94.766	0.901
1–2	95.040	96.691	95.866	0.921	4–5	93.983	94.086	94.034	0.889
1–3	95.063	96.297	95.680	0.918	4–6	94.411	93.771	94.091	0.889
1–4	94.777	95.297	95.037	0.906	5–0	94.789	95.234	95.011	0.906
1–5	94.554	93.869	94.211	0.893	5–1	94.829	94.577	94.703	0.901
1–6	94.543	94.114	94.329	0.894	5–2	94.240	94.400	94.320	0.894
2–0	94.577	97.080	95.829	0.921	5–3	94.400	94.634	94.517	0.897
2–1	95.503	97.023	96.263	0.928	5–4	94.846	93.874	94.360	0.894
2–2	94.606	95.857	95.231	0.910	5–5	94.166	93.389	93.777	0.885
2–3	94.371	94.886	94.629	0.900	5–6	94.091	93.389	93.740	0.883
2–4	94.211	94.971	94.591	0.898	6–0	94.589	94.406	94.497	0.897
2–5	93.623	93.280	93.451	0.879	6–1	95.074	94.657	94.866	0.903
2–6	94.200	93.960	94.080	0.889	6–2	95.149	94.594	94.871	0.903
3–0	95.486	96.286	95.886	0.921	6–3	94.526	94.166	94.346	0.894
3–1	95.400	95.846	95.623	0.917	6–4	94.646	94.006	94.326	0.894
3–2	94.994	94.840	94.917	0.904	6–5	94.674	93.554	94.114	0.890
3–3	94.429	93.766	94.097	0.890	6–6	94.554	93.577	94.066	0.889

Win window sizes, *SE* sensitivity, *SP* specificity, *AC* accuracy, *CC* coefficient of correlation

Fig. 1 Predicted results of *N*-terminus, **a** accuracy; **b** coefficient of correlation



AAs before and/or after at *N*- and *C*-termini cleavage sites. The sequence for every window is characterized by the parameters and modeled by BP-NN. The BP-NN predictive model is then evaluated and validated with the following four statistical parameters: sensitivity (SN), specificity (SP), accuracy (AC), and CC. The AC and CC are comprehensive parameters, including SN and SP information.

A total of 96 sets of evaluation results are obtained through calculation of the BP-NN models. These evaluation results are divided into two groups, namely the *N*- and *C*-termini. As shown in Table 3, the windows for the *N*-terminus containing 0 or 1 AA at both sides of the cleavage site have unsatisfactory reliability, whereas the other windows have high predictive ability (AC of about 90 % and CC above 0.80). As shown in Table 4, the windows containing no AA before the cleavage site have low predictive ability, whereas the windows containing one or more AAs before the cleavage site have high reliability (AC of about 95 % and CC of about 0.90). In general, the predictive ability of the models based on the *C*-terminus is superior to that of the models based on the *N*-terminus.

To compare the performance of the predictive models clearly, the AC and CC statistical parameters are shown with histograms for the *N*-terminus (Fig. 1) and *C*-

terminus (Fig. 2). As shown in Fig. 1, the performance of the model based on the *N*-terminus is not high, but the AAs after the cleavage site are less than 3, which indicate that the proteasome can at least recognize the cleavage site through the 3 AAs. Figure 1 shows the optimized window consisting of 3 AAs expressed as “ $P_1|P_2|P_3$ ”. The windows show the 3 AAs motif at the *N*-terminus of epitope. Similarly, Fig. 2 shows the optimized window size consisting of 2 AAs expressed as “ $P_1|P_1$ ”. The BP-NN models show high performance based on one or more AAs before the cleavage site. The windows expressed as “ $P_1|$ ” can be regarded as the most suitable motif for proteasome recognition. The optimized windows show that the epitope can be derived from antigen by proteasomes, which are recognized through AAs at the *N*-terminus and the *C*-terminus with sizes of 3 and 1, respectively. The motif located in the terminus is similar to the anchor residues of epitope and recognized by proteasomes instead of MHC.

Knowledge of the cleavage patterns is essential for understanding how the proteasome contributes to the generation of peptides presented by MHC class I molecules on the surface of vertebrate cells [27]. Compared with eukaryotic proteasomes, immunosubunits replace the constitutive proteolytically active subunits and significantly improve

Fig. 2 Predictive results of *C*-terminus, **a** accuracy; **b** coefficient of correlation

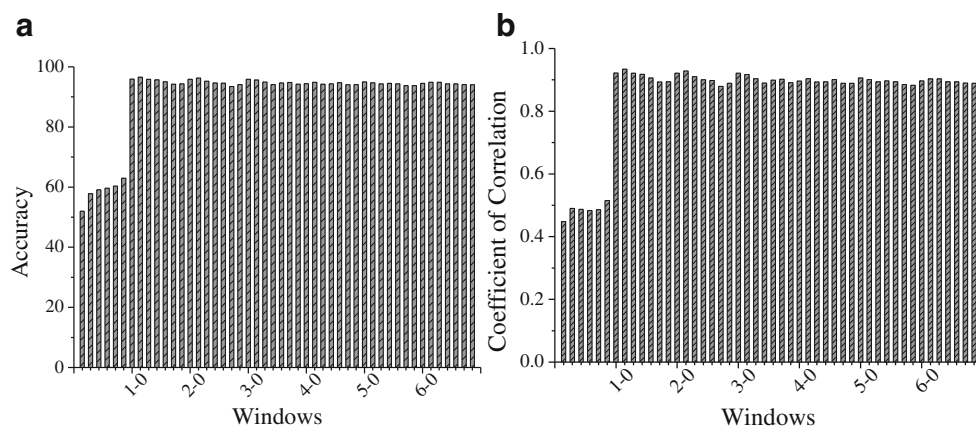
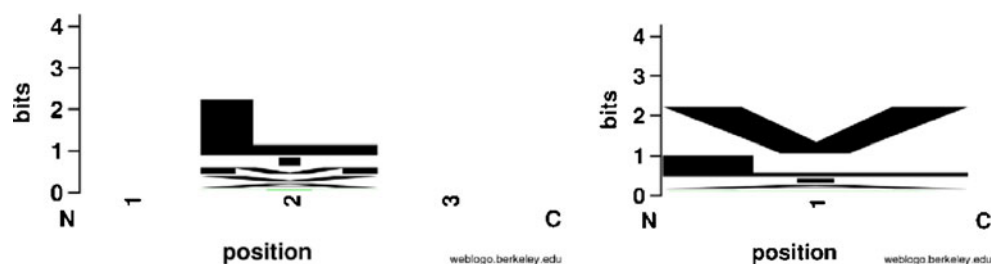


Fig. 3 Sequence logo of motif for *N*- and *C*-terminus



proteasome-dependent antigen processing. Groll *et al.* [28] reported that the conformation of MHC-I bound peptides is similar to that of the propeptides of proteasome active subunits, which suggests the co-evolution of MHC molecules and proteasomal ligand-binding sites. Conserved AAs should be present near the cleavage for recognition of proteasomes, which are analogous to the anchor residues of epitopes bound to MHC. In this study, the conserved analysis is performed by weblogo [29], and the sequence logo for the *N*- and *C*-termini is shown in Fig. 3. The motif for the *N*-terminus contains the conservative site at position 2 and the corresponding residues are Leu, Ile, and Met. Similarly, the motif for the *C*-terminus contains one residue and the conservative AAs are Val, Leu, and Ile. More than

75 % of peptides are nonamers and the conservative sites are consistent with nonamer epitopes, which are located at positions 2 and 9.

In eukaryotic cells, the ubiquitin-proteasome system plays a critical role in the cleavage of antigen protein in epitope processing and presentation in the MHC-I pathway. Ubiquitination occurs by the covalent attachment of multiple ubiquitin molecules onto the substrate, which is then transported and degraded in the 26S proteasome. Consequently, the weak interactions between the AAs closing the cleavage site and the proteasome depend on the AA properties, such as stericity, electrostatic properties, and hydrophobicity. The proteasome ligand-binding site is similar to MHC; thus, the interaction between the optimized

Table 5 Comparison with literature (*C*-terminal)

Method	Dataset	Web server	Number	SE	SP	AC	CC	Literature
BP-NN	Training	/	350	94.234	97.589	95.911	0.922	/
ANN	MHC I ligands	PAPROC	217	45.6	30.0		±0.25	[27]
		FragPredict	231	83.5	16.5		0.00	
		NetChop1.0	231	39.8	46.3		±0.14	
		NetChop2.0	231	73.6	42.4		0.16	
ANN	MHC ligands + MHC ligands	/	229	80	88		0.53	[20]
	MHC ligands and 20S + MHC ligands	/	413	72	92		0.56	
	Enlarged MHC Ligands + HIV protein	/	1110	66	74		0.37	
	20S + Peptide	/	184	68	84		0.53	
SVM (RBF)	In vitro data	Pcleavage	184	86.4	50.7	68.6	0.42	[28]
SVM (POLY)				84.6	55.6	70.0	0.43	
PEBLS				57.9	62.9	60.5	0.21	
Weke (Naïve Bayes)				51.3	70.9	61.6	0.23	
Weka (J48.PART)				50.8	69.2	60.0	0.20	
Weka (Logistic)				51.9	65.7	58.8	0.18	
SVM (RBF)				84.3	69.0	76.7	0.54	
SVM (POLY)				86.2	65.4	75.8	0.53	
PEBLS				25.3	96.2	88.5	0.28	
Weke (Naïve Bayes)				51.4	91.7	87.3	0.39	
Weka (J48.PART)	41.1	88.8	83.6	0.27				
Weka (Logistic)	54.9	88.3	84.6	0.36				

motif and proteasome depends on that between MHC and epitope. The properties, residue volume, electron-ion interaction potential values, and hydrophobicity can reflect the interaction between motif and proteasome, and the BP-NN models show high reliability and predictive ability.

Numerous predictive models have been constructed with the machine learning algorithm [20, 30, 31], including the support vector machine and ANN. Table 5 lists the statistical parameters of the predictive model built for the C-terminus from the present and previous studies. The present predictive model is superior to those in literature. Although the datasets, characterization of sequence, and modeling method differ among the listed models in Table 5, the BP-NN model has high predictive ability because of the properties representing the profile of the interaction between proteasome and motif. Therefore, the modeling based on the properties of AAs can provide a good scheme predicting proteasomal cleavage sites.

Conclusions

Sequences at both sides of a cleavage site are selected as windows for proteasome recognition. The optimized windows for the N- and C-termini are determined by BP-NN, wherein the optimization standard is the window with higher reliability and the shortest sequence. The prediction model from the BP-NN has high reliability that is also superior to those of previous models. The predicted CCs of BP-NN are 0.862 for N-terminus and 0.922 for C-terminus for the test dataset. The properties used in this work can build a good predictive model but cannot reflect the comprehensive structural features of cleavage sites. Therefore, more suitable parameters for characterizing cleavage sites are favorable for proteasome recognition at corresponding positions. The predicted BP-NN model can well calculate the cleavage site of the determinate and predict the novel cleavage site of the protein sequence.

Acknowledgments This study was supported by grants from the National Basic Research Program of China (No. 2012CB11460), National Natural Science Foundation of China (No. 81171508, 31170747) and Natural Science Foundation Project of CQ CSTC (No. cstc2013jjb10004).

References

- Frahm N, Korber BT, Adams CM et al. (2004) Consistent cytotoxic-T-lymphocyte targeting of immuno-dominant regions in human immunodeficiency virus across multiple ethnicities. *Virology* 78(5):2187–2200
- Takedatsu H, Shichijo S, Katagiri K et al. (2004) Identification of Peptide Vaccine Candidates Sharing among HLA-A3(+), -A11(+), -A31(+), and -A33(+). *J Cancer Patients Clin Cancer Res* 10(3):1112–1120
- Wu YZ, Shi TD (1998) EBVD: a new approach to molecular design of vaccine. *Immunologist* 8(s):1487–1489
- Shi TD, Wu Y (1998) Molecular design and immunological investigations on the therapeutic peptide vaccines against viral hepatitis B. *Immunologist* 8(s):330–334
- Rosenberg SA, Yang JC, Schwartzentruber DJ et al. (1998) Immunologic and therapeutic evaluation of a synthetic peptide vaccine for the treatment of patients with metastatic melanoma. *Nat Med* 4(3):321–327
- Ciechanover A (1994) The ubiquitin-proteasome proteolytic pathway. *Cell* 79(1):13–21
- Kisselev AF, Tatos N et al. (1999) The sizes of peptides generated from protein by mammalian 26 and 20 S proteasomes. *J Biol Chem* 274(6):3363–3371
- Coux O, Tanaka K, Goldberg AL (1996) Structure and functions of the 20S and 26S proteasomes. *Annu Rev Biochem* 65:801–847
- Jentsch S, Schlenker S (1995) Selective protein degradation: a journey's end within the proteasome. *Cell* 82(6):881–884
- Hershko A, Ciechanover A (1998) The ubiquitin system. *Annu Rev Biochem* 67:425–479
- Groll M, Ditzel L, Lowe J et al. (1997) Structure of 20S proteasome from yeast at 2.4 Å resolution. *Nature* 386:463–471
- Uebel S, Tampe R (1999) Specificity of the proteasome and the TAP transporter. *Current Opinion in Immunology. J Curr Opin Immunol* 11(2):203–208
- Craiu A, Akopian T, Goldberg A, Rock KL (1997) Two distinct proteolytic processes in the generation of a major histocompatibility complex class I-presented peptide. *Proc Natl Acad Sci* 94:10850–10855
- Niedermann G, King G, Butz S et al. (1996) The proteolytic fragments generated by vertebrate proteasomes: structural relationships to major histocompatibility complex class I binding peptides. *Proc Natl Acad Sci* 93:8572–8577
- Doytchinova IA, Guan PP, Flower DR (2006) EpiJen: a server for multistep T cell epitope prediction. *BMC Bioinforma* 7:131–141
- Pierre D, Oliver K (2005) Integrated modeling of the major events in the MHC class I antigen processing pathway. *Protein Sci* 14:2132–2140
- Holzthutter HG, Frommel C, Kloetzel PM (1999) A theoretical approach towards the identification of cleavage-determining amino acid motifs of the 20S proteasome. *J Mol Biol* 286:1251–1265
- Kuttler C, Nussbaum AK, Dick TP et al. (2000) An algorithm for the prediction of proteasomal cleavages. *J Mol Biol* 298:417–429
- Nussbaum AK, Kuttler C, Haderer KP et al. (2001) PProC: a prediction algorithm for proteasomal cleavages available on the WWW. *Immunogenetics* 53:87–94
- Kesmir C, Nussbaum AK, Schild H et al. (2002) Prediction of proteasome cleavage motifs by neural networks. *Protein Eng* 15:287–296
- 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
- Goldsack DE, Chalifoux RC (1973) Contribution of the free energy of mixing of hydrophobic side chains to the stability of the tertiary structure. *J Theor Biol* 39:645–651
- Cosic I (1994) Macromolecular bioactivity: is it resonant interaction between macromolecules?—Theory and applications. *J IEEE Trans Biomed Eng* 41:1101–1114
- Prabhakaran M (1990) The distribution of physical, chemical and conformational properties in signal and nascent peptides. *J Biochem* 269:691–696
- Shihab K (2006) A backpropagation neural network for computer network security. *J Comput Sci* 2(9):710–715

26. Abraham A (2005) Artificial neural networks. In: Sydenham PH, Thorn R (eds) Handbook of measuring system design. Wiley, New York, pp 901–908
27. Marques AJ, Palanimurugan R, Matias AC et al. (2009) Catalytic mechanism and assembly of the proteasome. *Chem Rev* 109(4):1509–1536
28. Groll M, Larionov OV, Huber R et al. (2006) Inhibitor-binding mode of homobelactosin C to proteasomes: new insights into class I MHC ligand generation. *Proc Natl Acad Sci USA* 103(12):4576–4579
29. Crooks GE, Hon G, Chandonia JM et al. (2004) WebLogo: a sequence logo generator. *Genome Res* 14(6):1188–1190
30. Saxova P, Buus S, Brunak S, Kesmir C (2003) Predicting proteasomal cleavage sites: a comparison of available methods. *Int Immunol* 15(7):781–787
31. Bhasin M, Raghava GP (2005) Pcleavage: an SVM based method for prediction of constitutive proteasome and immunoproteasome cleavage sites in antigenic sequences. *Nucleic Acids Res* 33(Web Server issue):W202–W207