Prediction of B-cell linear epitopes with a combination of support vector machine classification and amino acid propensity identification

Hsin-Wei Wang[1], Ya-Chi Lin[1], Tun-Wen Pai[1,2§], Hao-Teng Chang[3,4,5§]

[1]Department of Computer Science and Engineering, [2]Center of Excellence for Marine Bioenvironment and Biotechnology, National Taiwan Ocean University, Keelung, Taiwan, R.O.C.

[3]Graduate Institute of Molecular Systems Biomedicine, [4]Graduate Institute of Clinical Medical Science, [5]Graduate Institute of Basic Medical Science & Ph.D. Program for Aging, China Medical University, Taichung, Taiwan, R.O.C.

[§]Corresponding author: Dr. Hao-Teng Chang, Graduate Institute of Molecular Systems Biomedicine, College of Medicine, China Medical University, No. 91, Hsueh-Shih Road, Taichung, 40402, Taiwan, R.O.C. TEL:+886-4-22052121 ext 7721, FAX:+886-4-22333641, E-mail: htchang@mail.cmu.edu.tw

[§]Corresponding author: Dr. Tun-Wen Pai, Department of Computer Science and Engineering & Center of Excellence for Marine Bioenvironment and Biotechnology, National Taiwan Ocean University, No. 2, Peining Road, Keelung, 20224, Taiwan, R.O.C. TEL: +886-2-24622192 ex. 6618, FAX: +886-2-24623249, E-mail: twp@mail.ntou.edu.tw

Keywords: Linear epitope, antigenicity, support vector machine, machine learning, immunology

Abstract

Epitopes are antigenic determinants that are useful because they induce B cell antibody production and stimulate T cell activation. Bioinformatics can enable rapid, efficient prediction of potential epitopes. Here, we designed a novel B-cell linear epitope prediction system called LEPS, **L**inear **E**pitope Prediction by **P**ropensities and **S**upport Vector Machine, that combined physico-chemical propensity identification and support vector machine (SVM) classification. We tested the LEPS on four datasets: AntiJen, HIV, a newly generated PC, and AHP, a combination of these three datasets. Peptides with globally or locally high physico-chemical propensities were first identified as primitive linear epitope (LE) candidates. Then, candidates were classified with the SVM based on the unique features of amino acid segments. This reduced the number of predicted epitopes and enhanced the positive prediction value (PPV). Compared to four other well-known LE prediction systems, the LEPS achieved the highest accuracy (72.52%), specificity (84.22%), PPV (32.07%), and Matthews correlation coefficient (10.36%). The LEPS is freely available for academic use at http://LEPS.cs.ntou.edu.tw.

Introduction

Epitopes, also called antigenic determinants, are clusters of amino acid segments located on the surfaces of an antigen. Epitopes can elicit the immune response and are recognized by specific antibodies [1]. Basically, B-cell epitopes are categorized into two types: linear and conformational. Linear epitopes (LEs) are composed of contiguous amino acid residues within a continuous stretch of a primary protein sequence. Conformational epitopes (CEs) consist of amino acids that are dispersed among discontinuous regions, but become aggregated on the protein surface [2, 3]. In general, over 90% of B-cell epitopes are discontinuous [4, 5]; thus, CEs play critical roles in biological and biomedical applications, including the prevention and neutralization of pathogen infections, and the design of therapeutic drugs. However, the prediction and identification of CEs within a protein depend on resolved three-dimensional structural information. One major, generally accepted concept is that conformational epitopes cannot be properly formed without binding to a corresponding antibody [6]. Therefore, antigen-antibody co-crystallographic information is a major concern in CE prediction. On the other hand, because CEs are discontinuous epitopes, it is difficult to design a peptide that forms the same conformation as the predicted CE. Thus, CEs that are predicted by computational analysis may not be verifiable in biochemical experiments, except with the co-crystallographic approach. Although B-cell LEs occupy a small part of the entire epitope group, they are important in biochemistry [7], virology [8], immunology [9], and vaccine research [10]. Therefore, research and development of accurate computational approaches for LE prediction remains a critical challenge in bioinformatics and computational biology [6]. Most published B-cell LE predictors have been based on the characteristics of amino acids, like hydrophobicity, surface accessibility, mobility, protrusion area, physico-chemical properties, antigenicity, and pocket characteristics [1, 3, 11-16]. For example, BcePred [16], BEPITOPE [17], PEOPLE [11], VaxiJen [18], and LEP [12] are bioinformatics tool that use various mathematical approaches to predict LEs according to the physico-chemical propensities of amino acids. Nevertheless, in 2005, Blythe and Flower led a group that evaluated the physico-chemical propensities of amino acids to predict LEs in proteins; they reported that even the best physico-chemical propensity scales available performed only slightly better than a random model [19]. Hence, it was proposed that,

instead of using the antigenicity scale alone, LE prediction may be improved by integration with other computational approaches.

Several machine learning computational methods have been applied to improve the accuracy of LE prediction. For example, BepiPred combined a hydrophilicity scale with a hidden Markov model [20]; BCPred [21] and FBCPred [22] employed SVM with a subsequent kernel; Söllner and Mayer utilized a molecular operating environment with the decision tree and nearest neighbour approaches [6]. However, these machine learning approaches were mostly set to predict peptides of fixed lengths. It is difficult to analyze true LEs, because they generally range from 8-20 amino acid residues in length [11, 23-25]. Epitopes with fixed lengths are not typically sufficient to represent the whole region of antigenic determinants. To overcome the drawbacks of training and/or predicting fixed length epitopes, ABCPred used two artificial neural network methods, the feed-forward network and the recurrent neural network, for the prediction of B-cell LEs [26]. Both networks were used with different window lengths from 10 to 20 amino acids and a two-residue interval.

Although bioinformatists have expended great effort on developing LE predictors, there remains much room for improvement. Theoretically, an epitope identified by experimental immunological or biochemical methods must possess biological antigenicity that can induce antibody production in animals. However, when computational skills are used for the prediction, some experimentally identified epitopes could be missed or ignored. This generated the interesting study of how to retrieve the unpredictable epitopes and enhance their antigenicity score *in silico*.

In 2008, LEP was developed for predicting LEs based on physico-chemical propensities combined with a mathematical morphology approach. LEP could retrieve some of the LEs that were locally embedded in the noise signals of the antigenic index [12]. We reasoned that prediction accuracies could be further improved, and retain the advantage of variable length conditions, by combining the LEP with machine learning technologies.

As mentioned above, the machine learning methods used in previous LE prediction methods were often trained to predict epitopes with fixed lengths. Chen's study showed that the frequencies of occurrence for some amino acid pairs in the epitope dataset were significantly higher than in non-epitope datasets, or vice versa [23]. We noticed this important statistical feature and applied it to enhance the performance of

LE prediction systems. Hence, in order to explore the statistical advantages of verified epitopes and retain the antigenic characteristics of candidate peptides, we decided to extend the concept of amino acid pairs from Chen's study, which only considered peptides with 2 residues.

In this study, we developed a novel B-cell LE prediction system called LEPS (**L**inear **E**pitope Prediction by **P**ropensities and **S**upport Vector Machine). We adopted the library for SVM (LIBSVM) tool and trained it to recognize features of amino acid segments (AASs) with lengths from 2 to 4 residues. Then, SVM was used to characterize those patterns as epitope and non-epitope clusters [27]. Accordingly, the LEPS approach first performed physico-chemical propensities and mathematical morphology approaches, and then used the AAS features to cluster the predicted LE candidates and remove the less probable LEs.

Materials and Methods

Testing datasets and Predictors

Four datasets were used in this study. The AntiJen dataset was recommended at an international meeting sponsored by the National Institute for Allergy and Infectious Disease [6] and contained 171 protein sequences with 691 verified, non-overlapping epitopes [19]. The HIV dataset was a collection of the antigenic determinants located on 10 HIV proteins with 54 non-overlapping, verified epitopes [28]. The PC dataset, generated in this study, was a collection of 12 protein sequences with 98 non-overlapping, verified epitopes (Table 1). In order to balance out the variation of each dataset in quantity and antigen diversity, these three datasets were merged into one, comprehensive dataset called the "AHP dataset". These datasets were analyzed with different LE predictors, including the BepiPred [20], ABCPred [26], BCPred [21], and FBCPred [22], to compare performances with that of the LEPS developed here.

System flow

The proposed system was divided into three main steps (Fig. 1a). The first step retrieved primitive epitope candidates from a query protein sequence with LEP [12], which was developed in our previous work and was used with the default settings. Then, a SVM classifier was applied to remove less probable epitope candidates and

improve prediction accuracies. In the final step, the predicted epitope residues were highlighted in the query sequence and visualized in a predicted structure. The virtual structure was generated from Modeller 9.9, based on homologous protein structure modeling approaches [29].

Training datasets and SVM model

The process of training the SVM model comprised two major steps (Fig. 1b). The first step (step 1b) evaluated the statistical characteristics that determined the frequencies of occurrence of AASs with various lengths from an independent B-cell epitope dataset (Bcipep [30]) and a non-epitope dataset (Chen [23]). The second step (step 2b) produced a SVM model that recognized the epitopes and non-epitopes of the Chen dataset based on the statistical features derived from step 1b.

The Bcipep dataset comprised 1230 experimentally verified, B-cell, and non-redundant LEs with lengths that ranged from 3 to 56 residues that were identified in over 1000 antigen proteins. This dataset was used in step 1b to analyze the statistical characteristics associated with the frequencies of occurrence of AASs of 2 to 4 residues in length that represented epitopes.

The Chen dataset contained 872 epitopes and 872 non-epitopes. All epitopes and non-epitopes within this dataset were restricted to a length of 20 residues. These verified epitopes were retrieved from the Bcipep dataset by applying a ‘‘truncation-extension treatment’’. That is, when the length of an LE was longer than 20 residues, an equal number of superfluous residues were truncated from both the *N*- and *C*- termini to preserve the central 20 residues. Conversely, when the length of an LE was shorter than 20 residues, an equal number of residues were added to both the *N*- and *C*-termini until the epitope comprised 20 residues. On the other hand, the 872 non-epitopes were generated by randomly selecting peptide segments from the Swiss-Prot database [31], with the stipulation that none was the same as any of the 872 epitopes. The 872 non-epitopes were used to analyze the statistical characteristics of AASs for non-epitopes in step 1b. After determining the statistical features that were associated with frequencies of occurrence, the proposed system applied these features (step 2b) to produce a SVM model in a 5-fold cross-validation on the Chen dataset.

Statistical analysis of AASs and epitope indexes

For LE verification, we considered the statistical features to be AASs of 2 (AAS^2), 3 (AAS^3), and 4 (AAS^4) residues in length for both epitopes and non-epitopes. For AAS^2 , 400 possible combinations of residue pairs were analyzed for occurence frequencies within both the epitope and non-epitope datasets. The epitope index ($Epidex_i^2$) of the i[th] pattern (AAS_i^2) was calculated by taking logrithm value of the ratio of the number of AAS_i^2 among all epitopes $AASs^2$ compared to the same ratio in the non-epitope $AASs^2$ group with the following equation:

$$Epidex_i^2 = \log\left(\frac{f_i^{2^+} / \sum_i f_i^{2^+}}{f_i^{2^-} / \sum_i f_i^{2^-}}\right) \ (i = 1, 2, ...400)$$

where $f_i^{2^+}$ and $f_i^{2^-}$ were the numbers of AAS_i^2 in the epitope and non-epitope datasets, respectively, and $\sum_i f_i^{2^+}$ and $\sum_i f_i^{2^-}$ denoted the total number of AAS_i^2 in the corresponding dataset. Finally, the values of $Epidex_i^2$ were normalized to the range of [0, 1] to avoid dominance of any individual $Epidex_i^2$ in the classifier learning processes.

There were a total of 8000 and 160,000 possible combinations for AAS^3 and AAS^4, respectively. A large portion of AAS^3 or AAS^4 did not appear in the non-epitope dataset; this would cause a problem, because it could lead to a zero in the denominator. Hence, the definitions of $Epidex_i^3$ and $Epidex_i^4$ were modified from the definition for $Epidex_i^2$, and the corresponding epitope indexes for AAS^3 and AAS^4 were defined as follows:

$$Epidex_i^l = f_i^{l^+} / \sum_i f_i^{l^+} ,$$

where *l* was equal to 3 or 4. Again, the values of $Epidex_i^3$ and $Epidex_i^4$ were normalized to the range of [0, 1].

SVM features and model selection

In this study, we adopted the SVM as a learning method to classify epitope and non-epitope peptides. We employed the open source LIBSVM toolbox for executing this

classification. In LIBSVM, each instance in the training set possessed one target value (class label) and several features (attributes). In the testing set, only the features were required for each instance. The objective of SVM was to generate a model from the training set that facilitated the prediction of the target value of each instance in the testing set. In this study, a peptide corresponded to an instance and the target value (1 or -1) represented whether that peptide was an epitope. Each peptide contained three feature values based on $Epidex_i^2$, $Epidex_i^3$, and $Epidex_i^4$. For example, a 20-mer peptide was decomposed into 19 AAS_i^2 subsegments, and the corresponding epitope index of this peptide was obtained by taking the average of 19 $Epidex_i^2$ from the corresponding AAS_i^2. Similarly, the feature values of $Epidex_i^3$ and $Epidex_i^4$ could be obtained by calculating the averages of 18 $Epidex_i^3$ and 17 $Epidex_i^4$ subsegments, respectively.

The Chen dataset was used to construct a SVM model based on three feature values and the target values of each epitope and non-epitope. There were four common kernel functions provided by LIBSVM, including linear, polynomial, radial basis function (RBF), and sigmoid. We examined these four kernel functions with a 5-fold cross-validation. The training dataset was equally divided into 5 different subsets; four of the subsets were used for training the model and the last one was used for testing the model. These processes were repeated five times with each individual subset used as the testing subset. Here, the RBF kernel was selected as the default kernel function, because it provided the best cross-validation accuracy with the training data. Subsequently, the RBF kernel function was applied to train the whole testing dataset for constructing the final SVM classifier in the LEPS.

Performance measurement

To evaluate the performance of the LEPS at the level of the amino acid residue, five indicators were used to measure effectiveness at the default settings. These indicators were: (1) *sensitivity (SEN)*, defined as the percentage of epitopes that were correctly predicted as epitopes; (2) *specificity (SPE)*, defined as the percentage of non-epitopes that were correctly predicted as non-epitopes; (3) *positive predictive value* (*PPV*), defined as the probability that a predicted epitope was, in fact, an epitope; (4) *accuracy (ACC)*, defined as the proportion of correctly predicted peptides;

and (5) *Matthews correlation coefficient (MCC)*, which was a measure of the predictive performance that incorporated both SEN and SPE into a single value between -1 and +1 [26]. These parameters were calculated with the following equations:

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

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

(3) $$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$

(4) $$PPV = \frac{TP}{TP + FP}$$

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

where *TP* represented the true positive; *TN*, the true negative; *FP*, the false positive; and *FN*, the false negative.

Results and Discussion

A new linear epitope dataset: PC

The new dataset, called the PC dataset (collected by **P**ai and **C**hang), contained 12 sequences that did not overlap with other datasets. It was generated and analyzed in this study. The experimental epitopes in the PC dataset were identified with the peptide scan methodology, a conventional method for epitope determination. The average length of the identified epitopes in the PC dataset was 18.9 residues. This was considered a practical length for an epitope to be used in peptide vaccine development or antibody generation. The average epitope lengths in the HIV and AntiJen datasets were 26.4 and 16.3 residues, respectively. All sequences in the PC dataset were analyzed with the LEPS, and the predicted and experimentally verified epitopes are listed in Table 1.

The performance of LEPS

The epitope information collected from the PC, AntiJen, and HIV datasets were utilized to verify the performance of LEPS. The PC dataset was described in the previous section. The original AntiJen dataset comprised 3619 epitopes, of which

3168 were found in the Swiss-Port database. As in our previous report, we regenerated the original AntiJen dataset by removing the repeated epitopes [12]. The HIV dataset focused on one infectious pathogen and was recognized as a useful tool in the field of HIV immunology [28]. The AHP dataset combined these three datasets to balance the variations in each dataset including variations in epitope length and the physico-chemical properties of antigens. With these 4 datasets, we compared the performance of five LE predictors, including LEPS, BepiPred [20], ABCPred [26], BCPred [21], and FBCPred [22].

As expected, LEPS provided favorable results in all four datasets (Fig. 2). Table 2 shows that LEPS displayed the best specificity (SPE), with values of 88.33%, 84.48%, 74.84%, and 84.22% in the PC, AntiJen, HIV, and AHP datasets, respectively. Moreover, LEPS showed the best PPVs, with values of 45.12%, 28.85%, 71.44%, and 32.07% in the PC, AntiJen, HIV, and AHP datasets, respectively. The PPV indicated the rate of identifying real epitopes among all positive predicted candidates. It is one of the most important factors in conducting vaccine development. Reduction of the false positive candidates can improve the effectiveness and efficiency of identifying the real epitopes. Therefore, the LEPS will outperform the other predictors in terms of biological experiment cost-effectiveness. In the field of computational science, prediction accuracy is one of the most concerned factors for system evaluation. Except in the HIV dataset, LEPS displayed the best ACCs, with values of 61.66%, 73.81%, and 72.52% for the PC, AntiJen, and AHP datasets, respectively. These results showed that LEPS displayed excellent performance for LE prediction. The LEPS also showed the best performance in the MCC for the AntiJen and AHP datasets (10.10% and 10.36%), and the MCC was only a little lower (22.76%) than BCPred (29.80%) and FBCPred (27.81%) for the HIV dataset. Taken together, LEPS displayed excellent performance in SPE and PPVs for all four datasets; it also showed the best or equivalent ACCs for all datasets. However, it showed relatively low SEN compared to the other predictors, mainly due to less number of predicted LEs.

The LEPS platform

The LEPS provides a user-friendly interface for biologists to predict linear epitope candidates (Fig. 3a). LEPS will accept either FASTA format or text, and the default parameters were set as indicated. In this system, several physicochemical propensities can be dynamically modified by users, including secondary structures, hydropathy,

surface accessibility, flexibility, polarity, and other factors. The scanning window size for each parameter is also adjustable. After executing the prediction, the overall antigenicity of the query protein and the predicted LE candidates are displayed. For example, Fig. 3b shows the LEs in HIV integrase predicted by LEPS. Seventeen candidates were initially predicted by LEP based on the global and local distributions of antigenicity. These candidates were further filtered by SVM selection, with only 9 remaining candidates. Within these 9 epitope candidates, number 1 (residue 5-19), number 2 (residue 41-50), numbers 7 and 8 (residue 227-239, and residue 243-247), and number 9 (residue 261-266) overlapped with the experimental epitopes at residues 1-16, residues 42-55, residues 228-252, and residues 262-271, respectively. To verify the surface conditions of the predicted LEs within the query protein sequence, a protein structure was simulated based on homologous modeling approaches. This structure can be viewed and analyzed by clicking on the button labeled 'predicted structure'.

Visualization of the predicted LEs on 3D structures

Predicted structures of the query sequences can be rendered by Jmol (http://www.jmol.org/) in LEPS, and the corresponding PDBs and PyMOL script files (http://www.pymol.org/) are downloadable by request. For example, Figure 4 shows the simulated structure of HIV integrase as predicted by Modeller, with the predicted epitope segments displayed in yellow solid spheres. Because there is a high probability that true epitopes will be exposed on the protein surfaces for binding with antibodies, visualization of the predicted LEs on 3D structures can facilitate the selection of suitable epitopes from predicted candidates according to their surface distributions. Figure 5 shows an example of the experimentally verified epitopes and predicted epitopes for the 10 kDa chaperonin protein in the AntiJen dataset. The yellow spheres in both Fig. 5a and 5b show the true and predicted epitope atoms, respectively. The position of the remaining protein is shown in red and blue solid balls in the two simulated structures. In both cases, most of the epitope residues are located on the protein surface.

Acceptability of low sensitivities

Although LEPS can provide a highly accurate prediction of LEs, the low sensitivity is an issue that remains to be investigated. In general, epitope datasets confront a

challenge that biological experiments would not cover all the true epitopes within an individual antigen. Peptide scanning data could only identify potential epitopes that were recognized by a specific antibody. However, different antibodies to the same antigen might recognize different epitopes. These biological variations caused low coverage of epitopes within an antigen [32]. This situation implies that the sensitivities of a LE predictor should generally be low. Alternatively, a LE predictor might ubiquitously predict more epitopes to regain the sensitivities accompanying with the reduction of specificities. This will definitely lead to higher experimental costs in general. Nevertheless, to persuade biologists to conduct *in vitro* experiments on the predicted potential LEs, the accuracy and MCC values could provide balanced statistics for evaluating the performance of a prediction system.

In this study, LEPS displayed high accuracy, MCC, specificity, and PPV, although the sensitivity was a little low. However, the reduced sensitivity was offset by the high PPV. Therefore, the LEPS provides a high probability of success for molecular biologists in predicting and selecting functional epitopes effectively and efficiently.

Acknowledgments

This work was supported by National Science Council, Taiwan (NSC-98-2311-B-039-003-MY3 and NSC-99-2627-B-039-002 to H.T. Chang, and NSC 99-2627-B-019-007 and NSC98-2221-E-019-031-MY2 to T.W. Pai), and by Taiwan Department of Health Clinical Trial and Research Center of Excellence (DOH100-TD-B-111-004).

References

[1] D. R. Davies and G. H. Cohen, "Interactions of protein antigens with antibodies," *Proc Natl Acad Sci U S A,* vol. 93, pp. 7-12, Jan 9 1996.

[2] M. H. Van Regenmortel, "Immunoinformatics may lead to a reappraisal of the nature of B cell epitopes and of the feasibility of synthetic peptide vaccines," *J Mol Recognit,* vol. 19, pp. 183-7, May-Jun 2006.

[3] D. J. Barlow*, et al.*, "Continuous and discontinuous protein antigenic determinants," *Nature,* vol. 322, pp. 747-8, Aug 21-27 1986.

[4] D. C. Benjamin, "B-cell epitopes: fact and fiction," *Adv Exp Med Biol,* vol. 386, pp. 95-108, 1995.

[5] A. D. Vinion-Dubiel*, et al.*, "Antigenic diversity among Helicobacter pylori vacuolating toxins," *Infect Immun,* vol. 69, pp. 4329-36, Jul 2001.

[6] J. A. Greenbaum*, et al.*, "Towards a consensus on datasets and evaluation metrics for developing B-cell epitope prediction tools," *J Mol Recognit,* vol. 20, pp. 75-82, Mar-Apr 2007.

[7] O. Serup Andersen*, et al.*, "Identification of a linear epitope in sortilin that partakes in pro-neurotrophin binding," *J Biol Chem,* vol. 285, pp. 12210-22, Apr 16 2010.

[8] J. Xiang*, et al.*, "Expression and characterization of recombinant VP19c protein and N-terminal from duck enteritis virus," *Virol J,* vol. 8, p. 82, 2011.

[9] A. Lanza*, et al.*, "Controversial role of antibodies against linear epitopes of desmoglein 3 in pemphigus vulgaris, as revealed by semiquantitative living cell immunofluorescence microscopy and in-cell ELISA," *Int J Immunopathol Pharmacol,* vol. 23, pp. 1047-55, Oct-Dec 2010.

[10] M. Yadav*, et al.*, "Identification of major antigenic peptide of filarial glutathione-S-transferase," *Vaccine,* vol. 29, pp. 1297-303, Feb 1 2011.

[11] A. J. Alix, "Predictive estimation of protein linear epitopes by using the program PEOPLE," *Vaccine,* vol. 18, pp. 311-4, Sep 1999.

[12] H. T. Chang*, et al.*, "Estimation and extraction of B-cell linear epitopes predicted by mathematical morphology approaches," *J Mol Recognit,* vol. 21, pp. 431-41, Nov-Dec 2008.

[13] H. T. Chang*, et al.*, "A reinforced merging methodology for mapping unique peptide motifs in members of protein families," *BMC Bioinformatics,* vol. 7, p. 38, 2006.

[14] P. Haste Andersen*, et al.*, "Prediction of residues in discontinuous B-cell epitopes using protein 3D structures," *Protein Sci,* vol. 15, pp. 2558-67, Nov 2006.

[15] T. W. Pai*, et al.*, "REMUS: a tool for identification of unique peptide segments as epitopes," *Nucleic Acids Res,* vol. 34, pp. W198-201, Jul 1 2006.

[16] S. Saha and G. P. S. Raghhava, "BcePred: Prediction of Continuous B-Cell Epitopes in Antigenic Sequences Using Physico-chemical Properties.," *LNCS,* vol. 3239, pp. 197-204, 2004.

[17] M. Odorico and J. L. Pellequer, "BEPITOPE: predicting the location of continuous epitopes and patterns in proteins," *J Mol Recognit,* vol. 16, pp. 20-2, Jan-Feb 2003.

[18] I. A. Doytchinova and D. R. Flower, "VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines," *BMC Bioinformatics,* vol. 8, p. 4, 2007.

[19] C. P. Toseland, *et al.*, "AntiJen: a quantitative immunology database integrating functional, thermodynamic, kinetic, biophysical, and cellular data," *Immunome Res,* vol. 1, p. 4, Oct 6 2005.
[20] J. E. Larsen, *et al.*, "Improved method for predicting linear B-cell epitopes," *Immunome Res,* vol. 2, p. 2, 2006.
[21] Y. El-Manzalawy, *et al.*, "Predicting linear B-cell epitopes using string kernels," *J Mol Recognit,* vol. 21, pp. 243-55, Jul-Aug 2008.
[22] Y. El-Manzalawy, *et al.*, "Predicting flexible length linear B-cell epitopes," *Comput Syst Bioinformatics Conf,* vol. 7, pp. 121-32, 2008.
[23] J. Chen, *et al.*, "Prediction of linear B-cell epitopes using amino acid pair antigenicity scale," *Amino Acids,* vol. 33, pp. 423-8, Sep 2007.
[24] L. Florea, *et al.*, "Epitope prediction algorithms for peptide-based vaccine design," *Proc IEEE Comput Soc Bioinform Conf,* vol. 2, pp. 17-26, 2003.
[25] C. G. Roberts, *et al.*, "Prediction of HIV peptide epitopes by a novel algorithm," *AIDS Res Hum Retroviruses,* vol. 12, pp. 593-610, May 1 1996.
[26] S. Saha and G. P. Raghava, "Prediction of continuous B-cell epitopes in an antigen using recurrent neural network," *Proteins,* vol. 65, pp. 40-8, Oct 1 2006.
[27] C. C. Chang and C. J. Lin, *LIBSVM: a library for support vector machine*, 2001.
[28] B. T. M. Korber, *et al.*, "HIV Immunology and HIV/SIV Vaccine Databases," Los Alamos National Laboratory, Theoretical Biology and Biophysics, Los Alamos, New Mexico. LA-UR 04-8162., 2003.
[29] N. Eswar, *et al.*, "Comparative protein structure modeling using Modeller," *Curr Protoc Bioinformatics,* vol. Chapter 5, p. Unit 5 6, Oct 2006.
[30] S. Saha, *et al.*, "Bcipep: a database of B-cell epitopes," *BMC Genomics,* vol. 6, p. 79, 2005.
[31] B. Boeckmann, *et al.*, "The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003," *Nucleic Acids Res,* vol. 31, pp. 365-70, Jan 1 2003.
[32] S. E. Caoili, "Benchmarking B-cell epitope prediction for the design of peptide-based vaccines: problems and prospects," *J Biomed Biotechnol,* vol. 2010, p. 910524.
[33] M. Sachsamanoglou, *et al.*, "Antigenic profile of human recombinant PrP: generation and characterization of a versatile polyclonal antiserum," *J Neuroimmunol,* vol. 146, pp. 22-32, Jan 2004.
[34] L. L. Argiro, *et al.*, "Identification of a candidate vaccine peptide on the 37 kDa Schistosoma mansoni GAPDH," *Vaccine,* vol. 18, pp. 2039-48, Apr 3 2000.
[35] A. W. Burks, *et al.*, "Mapping and mutational analysis of the IgE-binding epitopes on Ara h 1, a legume vicilin protein and a major allergen in peanut hypersensitivity," *Eur J Biochem,* vol. 245, pp. 334-9, Apr 15 1997.
[36] S. J. Liu, *et al.*, "Immunological characterizations of the nucleocapsid protein based SARS vaccine candidates," *Vaccine,* vol. 24, pp. 3100-8, Apr 12 2006.
[37] X. Cui, *et al.*, "Identification and evaluation of an infertility-associated ZP3 epitope from the marsupial brushtail possum (Trichosurus vulpecula)," *Vaccine,* vol. 28, pp. 1499-505, Feb 10 2010.
[38] M. Mueller, *et al.*, "Antigenic characterization of recombinant hemagglutinin proteins derived from different avian influenza virus subtypes," *PLoS One,* vol. 5, p. e9097, 2010.

[39] A. N. da Silva*, et al.*, "Identification of continuous human B-cell epitopes in the envelope glycoprotein of dengue virus type 3 (DENV-3)," *PLoS One,* vol. 4, p. e7425, 2009.
[40] R. A. Stetler*, et al.*, "HSP27: mechanisms of cellular protection against neuronal injury," *Curr Mol Med,* vol. 9, pp. 863-72, Sep 2009.
[41] T. Senger*, et al.*, "Identification of B-cell epitopes on virus-like particles of cutaneous alpha-human papillomaviruses," *J Virol,* vol. 83, pp. 12692-701, Dec 2009.
[42] C. D. Kelly-Cirino and N. J. Mantis, "Neutralizing monoclonal antibodies directed against defined linear epitopes on domain 4 of anthrax protective antigen," *Infect Immun,* vol. 77, pp. 4859-67, Nov 2009.
[43] T. U. Consortium, "The Universal Protein Resource (UniProt) in 2010," *Nucleic Acids Res,* vol. 38, pp. D142-8, Jan 2010.

Table 1: Epitopes predicted in the PC dataset after analysis with LEPS

Antigen:Length (UniProt ID[a])	LEPS predicted Epitopes	Experimental Epitopes	Ref.
PrP:253 (P04156)	M_1ANLGCWML$_9$		
		R_{37}YPGQG$_{42}$	[33]
		Q_{52}GG$_{54}$	[33]
		Q_{91}GGGT$_{95}$	[33]
		N_{100}KPSKPKTNMKHMA$_{113}$	[33]
		G_{123}GLGGYMLG$_{131}$	[33]
	S_{143}DYEDRYYRENMHRYPN$_{159}$	H_{140}FG**SDY**$_{145}$	[33]
		Q_{160}VYYRPMD$_{167}$	[33]
		F_{198}TETD$_{202}$	[33]
	Y_{218}ERESQAYYQRGS$_{230}$		
GAPDH:338 (P20287)	A_4KVGING$_{10}$		
	A_{21}AFLKNTVDV$_{30}$		
	V_{31}SVNDPFIDL$_{40}$	**V_{31}SVNDPFIDL**EYM$_{43}$	[34]
	K$_{48}$RDSTHGTFP**GEVSTENGKLKVNG KL$_{73}$**	**G_{58}EVSTENGKLKVNGKL**ISVH**CERDP$_{82}$**	[34]
	C_{78}ERDPANIPWDKDGA$_{92}$		
	A_{108}QAHIKNNRAK$_{118}$	G$_{100}$VFTTIDK**AQAHIKN$_{114}$**	[34]
	S$_{123}$APSADAPM$_{131}$		
	V$_{136}$NENSYEKS$_{144}$		
	V$_{148}$SNASCTTN$_{156}$		
	K_{163}VIHDKFEIV$_{172}$	**K_{163}VIHDKFEIV**E$_{173}$	[34]
	V$_{188}$VDGPSSKLWRDGRGAM$_{204}$		
	A$_{210}$STGAAKAVG$_{219}$		
	L$_{225}$NGKLT$_{230}$		
	R$_{235}$VPTPDVSV$_{243}$		
	R$_{249}$LGKGASYEE$_{258}$		
	F_{287}VGSTSSS$_{294}$	S$_{268}$GPLKGILEYTEDEVVSSD**FVG$_{289}$**	[34]
	I$_{302}$SLNNNF$_{308}$		
	Y$_{315}$DNEFGY$_{321}$		
	I$_{329}$THMHKVDHA$_{338}$		
Ara h 1:626 (P43238)	**K_{26}SSPYQKK**TENPC$_{38}$	**K_{26}SSPYQKK$_{33}$**	[35]
	Q$_{47}$**QEPDDLK$_{54}$**	**Q_{48}EPDDLK**QKA$_{57}$	[35]
		E$_{66}$YDPRCVY$_{73}$	[35]
	P$_{75}$RGHTGTTNQRSPPG**ERTRGRQPG**	**E_{90}RTRGRQPGDYDDDRR$_{105}$**	[35]
	DYDDDRRQP**RREEGGRW**GPAGPRE	**R_{108}REEGGRW$_{115}$**	[35]
	R**EREEDWRQ**PR**EDWRRPSHQQPR**	**E_{124}REEDWRQ$_{131}$**	[35]
	KIRPEGREGEQEWGTPGSHVREETSR	E$_{134}$DWRRPSHQQPR**KIRPEG$_{151}$**	[35]
	NN$_{173}$	P$_{295}$GQFEDFF$_{302}$	[35]
		Y$_{312}$LQGFSRN$_{319}$	[35]
		F$_{325}$NAEFNEIRR$_{334}$	[35]
		Q$_{345}$EERGQRR$_{352}$	[35]
	K$_{381}$SVSKKGSEEEG**DI$_{394}$**	**D$_{393}$I**TNPINLRE$_{402}$	[35]
		N$_{409}$NFGKLFEVK$_{418}$	[35]
		G$_{463}$NLELV$_{468}$	[35]
	K$_{472}$EQQQRGRREEEEDEDEEEEGSNR EV$_{497}$		
		R$_{498}$RYTARLKEG$_{507}$	[35]
		E$_{525}$LHLLGFGIN$_{534}$	[35]
		H$_{539}$RIFLAGDKD$_{548}$	[35]
		I$_{551}$DQIEKQAKDLAFPGSGE$_{568}$	[35]
	P$_{587}$QSQSQSPSSPEKESPEKEDQEEEN QGGKGP$_{617}$		

SARS N:422 (Q19QW0)	H_{60}GKEEL$_{65}$ T_{77}NSGPDDQ$_{84}$ L_{140}NTPKDHIGTRNPNNN$_{155}$	A_{36}RPKQRRPQGLPNNTASWFT$_{55}$	[36]
		A_{156}ATVLQLPQGTTLPKGFYAEGSRGG$_{180}$	[36]
		T_{266}KQYNVTQAFGRRGP$_{280}$	[36]
		N_{286}FGDQDLIRQGTDYK$_{300}$	[36]
		K_{356}HIDAYKTFPPTEPKKDKKK$_{375}$	[36]
		R_{386}QKKQPTVTLLPAADMDDFSRQLQN$_{410}$	[36]
ZP3:399 (O77685, residue 24-422)	**T_{31}QSPAPGSSFSP$_{42}$**	**T_{31}QSPAPGSSFSP**PPVVA$_{47}$	[37]
		Q_{71}AAELTLGPSACAPVPAEPLSK$_{92}$	[37]
		H_{101}ECGSELQMTPDSLIYSTVLHY$_{122}$	[37]
	P_{124}N**LSQ$_{128}$**	**L_{126}SQ**SPLVLRSSP$_{137}$	[37]
		G_{156}IQPTWVPFHSTLSREQ$_{172}$	[37]
		D_{251}SSSIFISPRPG$_{262}$	[37]
		V_{291}TATDQAPSPLN$_{302}$	[37]
		A_{311}DEWLPVEGPRD$_{322}$	[37]
		Q_{346}EPGNPSEFEADLMLGPLVLSEAENGP$_{372}$	[37]
AIV-H4:511 (A3KF09, residue17-527)	Q_{17}NYTGNPVIC$_{26}$	D_{107}TCYPFDVPEYQSLR$_{121}$	[38]
		F_{137}QWNTVKQNGKSGACKRANVNDFFNRLNWLVK	[38]
	S_{169}DGNAYP$_{175}$	**SDGNAYP**LQNLTKINNGDYARLYIWGVHHPSTDT$_{202}$	
		N_{206}LYKNNPGRVTVSTK$_{220}$	[38]
		T_{224}SVVPNIGSGPLVRGGQSGRVSXYWTIV$_{250}$	[38]
		V_{257}FNTIGNLIAPRGHYKLNNQKKSTILNTAIPIGSCV SKCHTDKGSLSTTKPFQNISRIAVGDCPRYVKQGSL KLATGMRNIPEKASRGLFGAI$_{349}$	[38]
		D_{455}SEMNKLFERVRRQL$_{469}$	[38]
		A_{473}EDKGNGCFEIFHKCDNN$_{490}$	[38]
		N_{512}RFQIQGVKLTQGYM$_{526}$	[38]
AIV-H5:568 (A5HNY9)		A_{25}NNSTEQVDTIMEKNVTVTHAQDILEKTHNGKL$_{57}$	[38]
		E_{85}FLNVPEWSYIVEKINPANDLCYP$_{108}$	[38]
		C_{151}PYQGRSSFFRNVVW$_{165}$	[38]
		D_{199}AAEQTRLYQNPTTY$_{213}$	[38]
		R_{223}SKVNGQSGRMEFFWTILKPNDAINFESNGNFIAP ENAYKIV$_{273}$	[38]
		L_{472}RDNAKELGNGCFEFYHR$_{489}$	[38]
	E_{284}LEYGNCNTKC$_{294}$		
AIV-H12:527 (C7FPM3, residue 1-527)	**T_{35}LIEQNVPVT$_{44}$**	D_{31}TVN**TLIEQNVPVT**QVEELVH$_{51}$	[38]
		K_{127}YERVKMFDFTKWNVTYTGTSKACNNTSNQGSF YRSMRWLTLKSGQFPVQTDEY$_{180}$	[38]
		F_{190}TWAIHHPPTSDEQVKLYKNPNSLSSVTTDEINRS FRPNIGPRPL$_{234}$	[38]
		Q_{238}QGRMDYYWAVLKPGQTV$_{255}$	[38]
		T_{259}NGNLIAPEYGHLITGKSHGRILKNDLPIGQCTTEC $_{294}$	[38]
		T_{310}SKHYIGKCPKYIPS$_{324}$	[38]
		R_{334}NVPQAQDRGLFGAIAGFIEG$_{354}$	[38]
		I_{430}TDIWAYNAELLVLLENQKTLDEHDANVRNLHDR VR$_{465}$	[38]
		G_{478}CFEILHKCDDGCMDTIKNGT$_{498}$	[38]
		Q_{502}DYEEESKLERQRINGVKLEENSTYK$_{527}$	[38]
DEN-3 E-glycoprotein:493 (D2JWZ8, residue 281-773)		T_{331}QLATLRKLCIEGKI$_{345}$	[39]
		D_{351}SRCPTQGEAVLPEEQDPNY$_{370}$	[39]
		Q_{411}YENLKYTVIITVHTGDQHQVGNETQGVTAEITP QASTTE$_{450}$	[39]
		L_{476}LTMKNKAWMVHRQW$_{490}$	[39]
	S_{533}QEGA$_{537}$ W_{669}YKKGSSI$_{676}$ L_{707}NSLG$_{711}$	Q_{526}EVVVLG**SQEGA**MHT$_{540}$	[39]

O. tsutsugamushi 47-kDa antigen:466 (Q53246)		H_{21}SKSLLNQKAVLPQQKSDMHIN$_{42}$	[40]
		T_{65}NIGISLNNKVSKYQQEV$_{82}$	[40]
		V_{97}TNENVIAGR$_{106}$	[40]
		Y_{145}ATFGDSNQS$_{154}$	[40]
		V_{173}TNGIISSKGRDMG$_{186}$	[40]
		F_{193}IQTNAAIHM$_{202}$	[40]
		H_{201}MGSFGGPMF$_{210}$	[40]
		I_{233}PSNTVLEAV$_{242}$	[40]
	L_{245}KKGEKIR$_{252}$	**L_{245}KKGEKIR**RG$_{254}$	[40]
		L_{333}LRNGKSMTLKCKIIANK$_{350}$	[40]
		Q_{357}SNDQSLVVN$_{366}$	[40]
		L_{373}TPDLVKKYNITSA$_{386}$	[40]
HPV L1 protein:510 (A8BQ01)		D_{41}VYVTRTNVYYHGGSSRLLTVGHPYYSIKKSNNK VAVPKV$_{80}$	[41]
	V_{122}GRGQPL$_{128}$	V_{90}KLPDPNKFGLPDADLYDPDTQRLLWACVGVE**VG RGQPL**GV$_{130}$	[41]
		T_{205}TIEDGDMVET$_{215}$	[41]
		D_{219}ICTNTCKYPDYLKMAAEPY$_{238}$	[41]
		G_{235}DSMFFSLRREQMFTRHFFNRGGKMGDTIPD$_{285}$	[41]
	R_{326}AQGHNNGMCW$_{336}$		
		S_{350}TNVSLCATEA$_{360}$	[41]
		F_{370}KEYLRHMEEYDLQFIFQLCKITLTPEIMAY$_{400}$	[41]
	V_{416}PPPPSASL$_{424}$		
	K_{440}PTPPKTPTD**P_{450}**	**P_{450}**YASLTFWDVDLSESFSMDLD$_{470}$	[41]
	G_{497}TPPPTSKRKRV$_{508}$		
Bacillus anthracis, PA domain III and IV:248 (P13423, residue 488-735)	**N_{538}PSDPLETTKPDMT$_{551}$**	R_{532}RIAAV**NPSDPLETTKPDMT$_{551}$**	[42]
		A_{596}ELNATNIYTVL$_{607}$	[42]
		I_{620}RDKRFHYDRNNIAVGADES$_{639}$	[42]
		L_{692}NISSLRQDGKT$_{703}$	[42]
	N_{720}PNYK$_{724}$	L_{716}YIS**NPNYK**VNVYAVTKENT$_{735}$	[42]

[a]Because some of the epitopes in the PC dataset were partial antigen fragments, the serial numbers for the residues in each epitope were assigned according to the sequence information retrieved from the UniProt database [43]. The overlapping amino acids between the experimentally verified and predicted epitopes are shown in bold.

Table 2. Comparison of the performances of LEPS, BepiPred, ABCPred, BCPred, and FBCPred systems.

Systems	SEN[a]	SPE[a]	ACC[a]	PPV[a]	MCC[a]
PC dataset					
LEPS	12.78	**88.33**	**61.66**	**45.12**	3.65
BepiPred	48.23	59.72	55.33	38.19	**7.49**
$ABCPred_{0.8}$[b]	**65.46**	40.26	48.89	36.21	5.13
BCPred	50.92	59.35	52.83	36.07	4.43
FBCPred	51.03	52.55	52.20	35.26	3.17
AntiJen dataset					
LEPS	26.72	**84.48**	**73.81**	**28.85**	**10.10**
BepiPred	51.79	57.61	55.52	22.02	6.04
$ABCPred_{0.8}$	**67.33**	40.40	44.70	21.83	5.46
BCPred	58.84	54.87	53.92	23.34	8.93
FBCPred	60.31	51.21	51.45	22.33	6.73
HIV dataset					
LEPS	48.33	**74.84**	63.45	**71.44**	22.76
BepiPred	50.16	60.85	56.72	61.22	9.72
$ABCPred_{0.7}$	**87.97**	14.65	56.59	56.33	5.64
BCPred	80.18	54.57	66.57	65.55	**29.80**
FBCPred	73.20	58.20	**67.13**	65.56	27.81
AHP dataset[c]					
LEPS	26.97	**84.22**	**72.52**	**32.07**	**10.36**
BepiPred	51.48	57.91	55.57	25.06	6.32
$ABCPred_{0.8}$	**68.28**	39.06	45.58	24.51	5.45
BCPred	59.45	54.80	54.50	26.32	9.73
FBCPred	60.40	51.66	52.31	25.38	7.60

[a] SEN, sensitivity; SPE, specificity; PPV, positive prediction value; ACC, accuracy; MCC, Matthews correlation coefficient, unit, %

[b] The subscripts of ABCPred denote threshold values according to the highest accuracy.

[c] This dataset is a merge of the other 3 datasets.

Fig. 1 The design of LEPS. (**a**) Step 1a: Primitive epitope candidates with globally and locally high antigenicity were extracted by calculating weighting coefficients for various physic-chemical propensities of each amino acid. After the filtering process with the SVM classifier (step 2a), predicted epitopes were highlighted (step 3a) in the query sequence and the simulated structure. (**b**) Step 1b: 1230 experimentally verified epitopes and 872 non-epitopes were analyzed to determine the statistical characteristics of AASs. Step 2b: Subsequently, epitope indexes of 872 epitopes and 872 non-epitopes were used to train the SVM model to predict candidate epitopes based on the statistical characteristics defined in step 1b.

Fig. 2 Comparison of the performances of LEPS, BepiPred, ABCPred, BCPred, and FBCPred systems. The best performance for each indicator is marked with a star.

Fig. 3 The LEPS server. (a) Users can input a query sequence and manually adjust the weight and window size of each propensity. (b) The output information of HIV integrase predicted by LEPS shows 17 candidates, and only 9 candidates were retained after SVM filtration. The final predicted epitope segments are labeled in yellow at the bottom.

Fig. 4 The predicted LEs of HIV integrase mapped onto a simulated 3D structure. The predicted epitopes are labeled in yellow and the selected epitopes (number 1 and number 3) are shown in yellow spheres.

Fig. 5 The experimental and predicted epitopes of 10 kDa chaperonin. The structural surfaces display the true epitopes (a) and predicted epitopes (b) in yellow spheres. The red and blue spheres represent the remainder of the protein. Both figures were created with PyMOL.

Figure 1

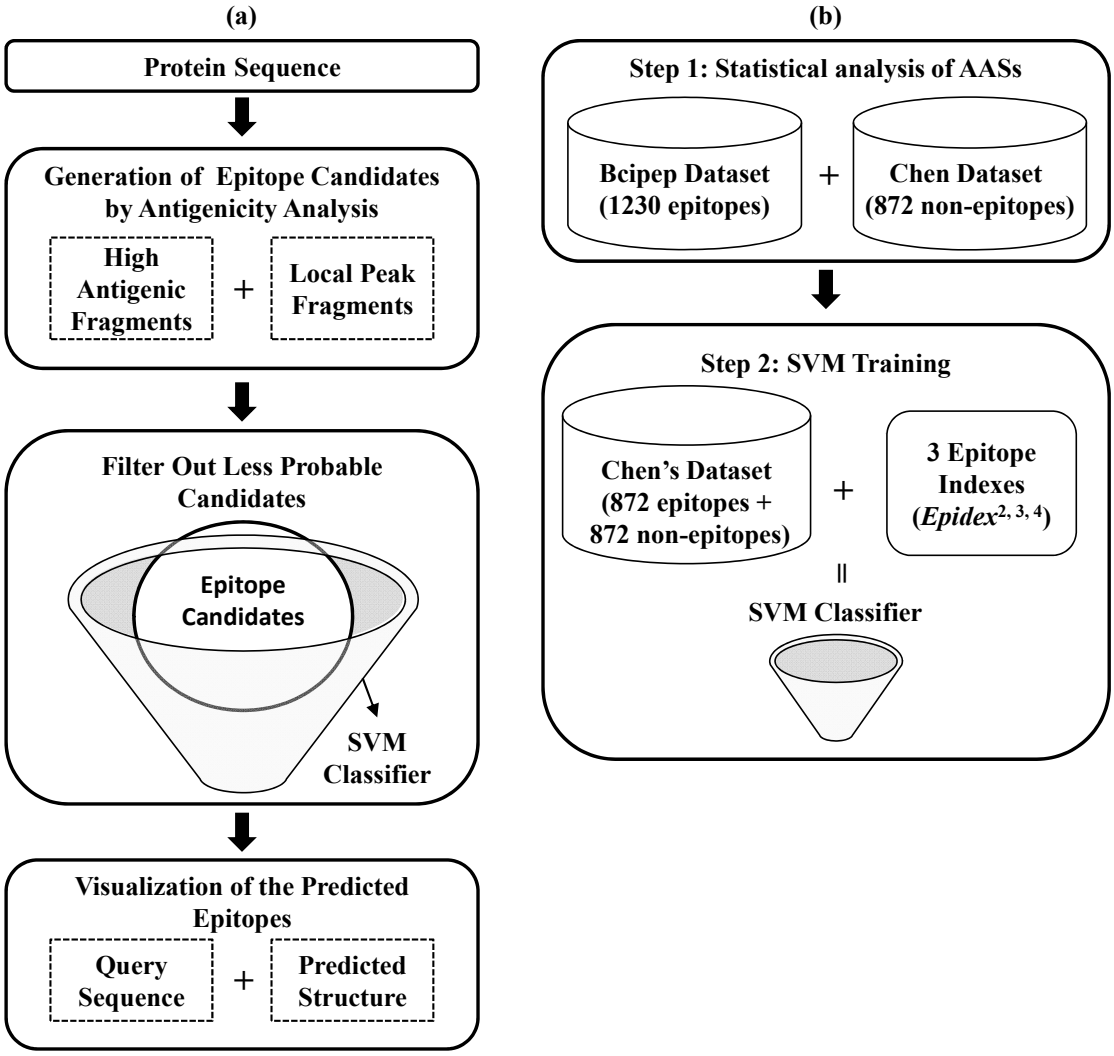

(a)
Protein Sequence
Generation of Epitope Candidates by Antigenicity Analysis
High Antigenic Fragments
+
Local Peak Fragments
Filter Out Less Probable Candidates
Epitope Candidates
SVM Classifier
Visualization of the Predicted Epitopes
Query Sequence
+
Predicted Structure
(b)
Step 1: Statistical analysis of AASs
Bcipep Dataset (1230 epitopes)
+
Chen Dataset (872 non-epitopes)
Step 2: SVM Training
Chen's Dataset (872 epitopes + 872 non-epitopes)
+
3 Epitope Indexes (Epidex2, 3, 4)
=
SVM Classifier

Figure 2

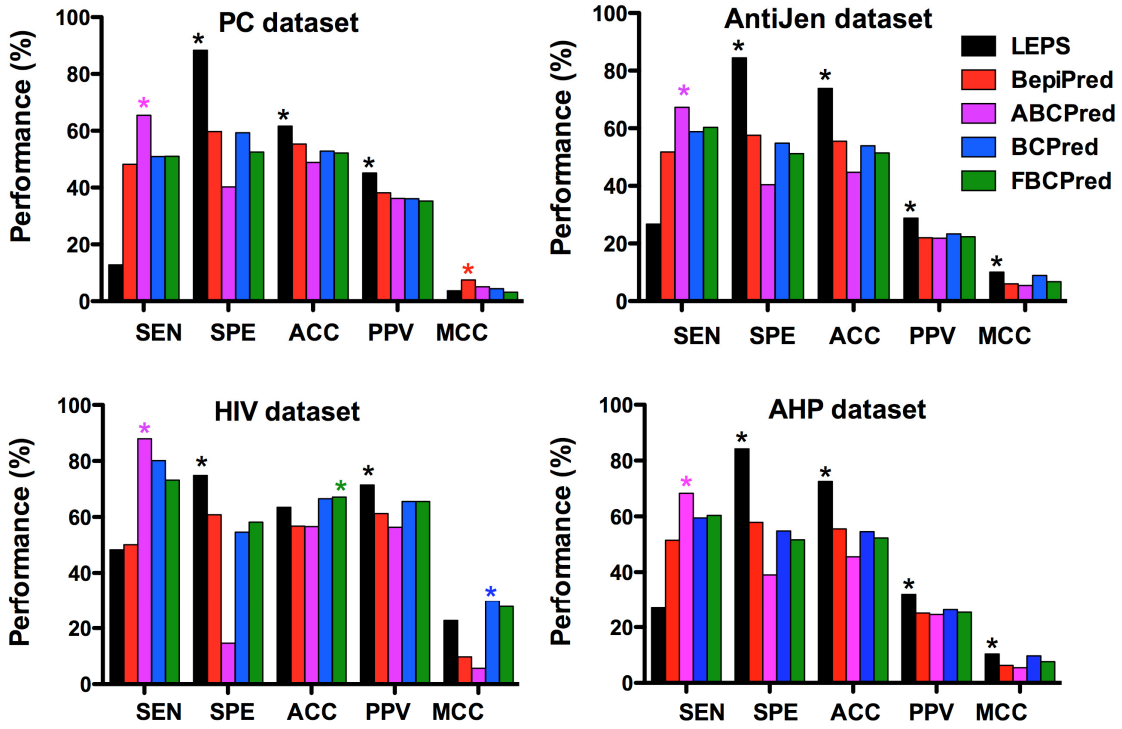

PC dataset
AntiJen dataset
HIV dataset
AHP dataset
Performance (%)
SEN
SPE
ACC
PPV
MCC
LEPS
BepiPred
ABCPred
BCPred
FBCPred

Figure 3

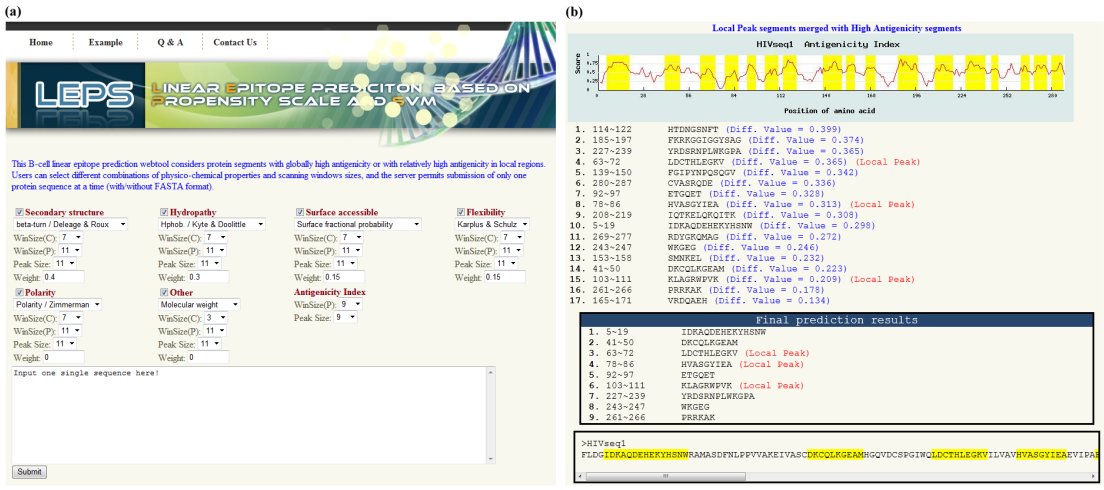
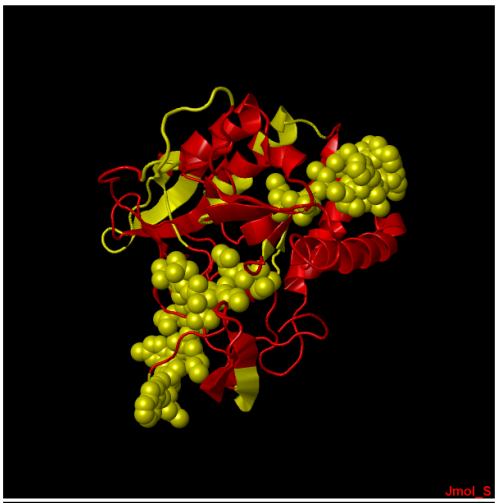

(a)
Home
Example
Q & A
Contact Us
LEPS
LINEAR EPITOPE PREDICITON BASED ON PROPENSITY SCALE AND SVM
This B-cell linear epitope prediction webtool considers protein segments with globally high antigenicity or with relatively high antigenicity in local regions. Users can select different combinations of physico-chemical properties and scanning windows sizes, and the server permits submission of only one protein sequence at a time (with/without FASTA format).
Secondary structure
beta-turn / Deleage & Roux
Hydropathy
Hphob. / Kyte & Doolittle
Surface accessible
Surface fractional probability
Flexibility
Karplus & Schulz
Polarity
Polarity / Zimmerman
Other
Molecular weight
Antigenicity Index
Input one single sequence here!
Submit
(b)
Local Peak segments merged with High Antigenicity segments
HIVseq1 Antigenicity Index
Position of amino acid
1. 114~122 HTDNGSNFT (Diff. Value = 0.399)
2. 185~197 FKRKGGIGGYSAG (Diff. Value = 0.374)
3. 227~239 YRDSRNPLWKGPA (Diff. Value = 0.365)
4. 63~72 LDCTHLEGKV (Diff. Value = 0.365) (Local Peak)
5. 139~150 FGIPYNPQSQGV (Diff. Value = 0.342)
6. 280~287 CVASRQDE (Diff. Value = 0.336)
7. 92~97 ETGQET (Diff. Value = 0.328)
8. 78~86 HVASGYIEA (Diff. Value = 0.313) (Local Peak)
9. 208~219 IQTKELQKQITK (Diff. Value = 0.308)
10. 5~19 IDKAQDEHEKYHSNW (Diff. Value = 0.298)
11. 269~277 RDYGKQMAG (Diff. Value = 0.272)
12. 243~247 WKGEG (Diff. Value = 0.246)
13. 153~158 SNNKEL (Diff. Value = 0.232)
14. 41~50 DKCQLKGEAM (Diff. Value = 0.223)
15. 103~111 KLAGRWPVK (Diff. Value = 0.209) (Local Peak)
16. 261~266 PRRKAK (Diff. Value = 0.178)
17. 165~171 VRDQAEH (Diff. Value = 0.134)
Final prediction results
1. 5~19 IDKAQDEHEKYHSNW
2. 41~50 DKCQLKGEAM
3. 63~72 LDCTHLEGKV (Local Peak)
4. 78~86 HVASGYIEA (Local Peak)
5. 92~97 ETGQET
6. 103~111 KLAGRWPVK (Local Peak)
7. 227~239 YRDSRNPLWKGPA
8. 243~247 WKGEG
9. 261~266 PRRKAK
>HIVseq1

Figure 4



Spin High Quality Reset Position Label CA
Display Style cartoon
Render Style style3
Create Surface Delete Surface Translucent Opaque

Final prediction results

1.	5-19	IDKAQDEHEKYHSNW
2.	41-50	DKCQLKGEAM
3.	63-72	LDCTHLEGKV
4.	78-86	HVASGYIEA
5.	92-97	ETGQET
6.	103-111	KLAGRWPVK
7.	227-239	YRDSRNPLWKGPA
8.	243-247	WKGEG
9.	261-266	PRRKAK

Download PDB file Download PyMOL script

Figure 5