The Relationship Between B-Cell Epitope and Mimotope Sequences

Chunhua Zhang[1,3], Yunyun Li[1,4], Weina Tang[1,3], Zhiguo Zhou[1], Pingping Sun[1,2,3]* and Zhiqiang Ma[1,3,]*

Pingping Sun

[1]School of Computer Science and Information Technology, Northeast Normal University, Changchun 130117, China; [2]National Engineering Laboratory for Druggable Gene and Protein Screening, Northeast Normal University, Changchun 130024, China; [3]Key Laboratory of Intelligent Information Processing of Jilin Universities, Northeast Normal University, Changchun 130117, China; [4]Ganjingzi District Dalian City Hengyuan Primary School, Dalian 116000, China

Zhiqiang Ma

Abstract: B-cell epitope is a group of residues which is on the surface of an antigen. It invokes humoral responses. Locating B-cell epitope is important for effective vaccine design, and the development of diagnostic reagents. Mimotope-based B-cell epitope prediction method is a kind of conformational B-cell epitope prediction, and the core idea of the method is mapping the mimotope sequences which are obtained from a random phage display library. However, current mimotope-based B-cell epitope prediction methods cannot maintain a high degree of satisfaction in the circumstances of employing only mimotope sequences. In this study, we did a multi-perspective analysis on parameters for conformational B-cell epitopes and characteristics between epitope and mimotope on a benchmark datasets which contains 67 mimotope sets, corresponding to 40 unique complex structures. In these 67 cases, there are 25 antigen-antibody complexes and 42 protein-protein interactions. We analyzed the two parts separately. The results showed the mimotope sequences do have some epitope features, but there are also some epitope properties that mimotope sequences do not contain. In addition, the numbers of epitope segments with different lengths were obviously different between the antigen-antibody complexes and the protein-protein interactions. This study reflects how similar do mimotope sequence and genuine epitopes have; and evaluates existing mimotope-based B-cell epitope prediction methods from a novel viewpoint.

Keywords: Epitope, mimotope, sequence, statistic.

1. INTRODUCTION

A B-cell epitope is a group of residues on the surface of an antigen which recognized by either a particular B-cell receptor (BCR) or a particular antibody molecule of the immune system [1]. Predicting epitopes is crucial for disease diagnosis, vaccine design, antibody design and immunological therapy [2]. The most reliable methods for locating B-cell epitopes are X-ray crystallography and Nuclear Magnetic Resonance (NMR) techniques [3,4]; however, these techniques are expensive in terms of cost and time-consuming, and they need a large number of professionals to operate. Under this circumstance, various computational methods featured for the low cost and high speed have been applied to predict B-cell epitope residues [5].

A B-cell epitope can be classified into two categories by its spatial structure: linear epitope or conformational epitope. A linear epitope (also called continuous epitopes) is composed of residues that are sequentially successive, whereas a conformational epitope (also known as discontinuous epitope) consists of sequential segments that are brought together in spatial proximity when the corresponding antigen is folded [6]. It has been testified that more than 90 % of B-cell epitopes are discontinuous B-cell epitopes [7]. Linear B-cell epitope prediction research has acquired many achievements, however, since most B-cell epitopes are conformational ones, and linear B-cell epitope can be considered as a special kind of conformational epitope which is composed by one sequential segment. So the methods of conformational B-cell epitope prediction are more comprehensive.

The current conformational B-cell epitope prediction methods are mainly divided into two types depending on the input information: antigen structure-based prediction and mimotope-based prediction.

Antigen structure-based B-cell epitope prediction algorithms need antigen 3D structure as input. The core idea of the prediction methods is the 3D structure of antigen and epitope-related propensity scales, including geometric attributes and specific physicochemical properties. The representative methods include CEP [8], DiscoTope [9], ElliPro [10], PEPITO [11], SEPPA [12] and Epitopia [13] etc. However, structure-based B-cell epitope prediction methods do not improved a lot in these years, because it is more complicated

*Address correspondence to these authors at the School of Computer Science and Information Technology, Northeast Normal University, Changchun 130117, China; E-mails: sunpp567@nenu.edu.cn; mazq@nenu.edu.cn

1875-5305/16 $58.00+.00

to extract features from the 3D structure of an antigen. Features mentioned in published papers do not have enough ability to distinguish the epitope residues from the rest residues [7].

Mimotope-based B-cell epitope prediction is a combinatorial method which requires both antibody affinity-selected peptides and the 3D structure of antigen as input. To attain affinity-selected peptides, random peptides are initially displayed on the surface of filamentous phages. Random peptides are screened, eluted and amplified, after 3-5 rounds, we can get higher affinity peptides. These affinity-selected peptides are considered mimotopes. Mimotopes and genuine epitopes can combine the same paratopes of monoclonal antibody and cause immune response, so they have the similar functionality with the genuine epitopes [14,15]. Besides, the selected mimotopes commonly share high sequential similarity with epitopes. Therefore these mimotopes can help finding the genuine epitopes more accurately.

In 1995, Pizzi released a B-cell epitope prediction software for the first time in this thought, and the software is known as MEPS [16]. MEPS uses the fixed-length short peptides represent the surface of antigen and then alignment with the motif to get the optimal matching short peptides. Soon afterwards, some other methods and software were proposed, such as Mapitope [17,18], PepSurf [19], Pep-3D-Search [20], EpiSearch [21] and MimoPro [22]. In 2011, we constructed a benchmark dataset for conformational B-cell epitope prediction and evaluated five mimotope-based prediction softwares [23]. The result showed that in no method did the performance exceed a 0.42 of precision and 0.37 of sensitivity.

Existing mimotope-based B-cell epitope prediction methods use only mimotope sequences to predict the epitopes; however despite of the continuous efforts, the performance of these algorithms is less than satisfactory and it seems to meet a performance bottleneck. Then several questions related to conformational epitopes and mimotopes are in urgent need of answering. How many epitope features does mimotopes contain? Does mimotopes contain other features except sequences? Or are there some effective features which mimotopes do not contain but contribute to epitope prediction?

To answer these questions, Benchmark 2.0 [24] of mimotope-based B-cell epitope prediction datasets was firstly chosen. The datasets contain 67 complexes structures, including 25 antigen-antibody complexes and 42 protein-protein interactions. Then a series of analysis were examined on this dataset from two aspects: (1) general characteristics of conformational B-cell epitope, such as sequence segment of epitopes, statistical distance, epitope patch and minimum surface path length; (2) comparison between epitopes and mimotopes from residue levels. This work is aimed to systematically detect the relationship of mimotopes and conformational B-cell epitopes, and try to evaluate existing mimotope-based B-cell epitope prediction methods from the view of the relationship between B-cell epitope and mimotope sequences.

2. THE CHARACTERISTICS BETWEEN EPITOPE AND MIMOTOPE

A lot of parameters have been applied to describe conformational B-cell epitope, including size, sequence continuity, residue accessibility, preference of amino acids, preference of residue-neighbor sets and evolutionary conservation etc. B-cell epitopes are conformational, while mimotopes are continuous, so parameters which can be used to make comparing between epitopes and mimotopes are limited. In this work, we try to make a systematic analysis between them and choose a part of parameters for our research. These parameters mainly reflect the structural characteristics and tendency of epitopes, and they have important contributions to the determination of the conformational B-cell epitope. These parameters are classified into two perspectives, including the general characteristics of epitopes and residue characteristics of epitopes and mimotopes. Details are shown in Table **1**.

Table1. Schematic table of parameters used in analysis. Epitope residues and mimotope residues have been investigated from two main perspectives in our work.

The characteristics between epitope and mimotope	General characteristis	Sequence segment of epitopes
		Statistical distances between every two epitope residues
		Epitope patch
		Minimum length of epitope surface path
	Residue characteristis	Preference of amino acids in epitope, mimotopes along with the surface
		Preference of amino acids according different classes
		Preference of residue-neighbor sets in epitope and mimotopes
		Residue accessibility

2.1. General Characteristics of Conformational B-cell Epitope

In this work, we did calculations of following general characteristics for conformational B-cell epitopes: sequence segment of epitopes, distances between every two epitope residues, epitope patch and minimum length of epitope surface path.

2.1.1. Statistic of Sequence Segment of Epitopes

While it has been suggested quite a while ago that 90% of all epitopes are discontinuous and it assumes that an epitope can be only one or the other. In fact, linear B-cell epitope can be considered as a special kind of conformational epitope which is composed by one sequential segment. In this work, we did statistic of sequence segment of epitopes for giving a general distribution of continuous segment of conformational B-cell epitopes. Our benchmark datasets contain antigen-antibody complexes and protein-protein interactions, so we analyzed the two parts separately.

From Figure **1**, we can see most of conformational B-cell epitopes have continuous sequence segments for both kinds of cases. The number of epitope segment whose length is longer than two is 55.77% (324 in 581). The result may lead to think whether conformational epitopes contain dominant continuous segments, and these continuous segments may directly accessible to biotechnology as immunogen for diagnostics and antibody production.

2.1.2. Statistical Distances between Every Two Epitope Residues

The core idea of mimotope-based epitope prediction methods is mapping the mimotope sequences back to the original antigen surface, and there are two parameters that usually be used, the first is distance. Distance is commonly used to judge whether two amino acids are neighbored , and define amino acids pairs and path which mimotope travelled on the antigen surface. In the research, we do statistic of Euclidean distances between every two epitope residues in each case. We calculate the number of epitope residues in each distance interval with the step size of 1Å. The results are showed in Figure **2**.

From Figure **2**, it can be seen clearly that the quantity of epitope residues in each distance interval. The cases of antigen-antibody complexes and protein-protein interactions have similar trends as Figure **1** shows. For antigen-antibody complexes, the distances between 83.68% epitope residues are from 2 Å to 22 Å (3780 in 4517); 6.20% epitope residues are belong to the distance interval 9 Å -10 Å (280 in 4517). Similarly, for protein-protein interactions, the distances between 83.65% epitope residues are between 4 Å to 26 Å (6471 in 7736); 5.44% epitope residues are belong to the distance interval 13 Å -14 Å (421 in 7736). The figure may provide a direction for future conformational B-cell epitope prediction.

2.1.3. Statistic for Epitope Patch

Another parameter usually used in mimotope-based epitope prediction methods is in patch dividing. Dividing patch is usually a crucial point in the conformational B-cell epitope prediction. Methods first divided the antigen surface into overlapped patches which one of them could contain most epitope residues, then design effective algorithm to find this certain patch. Though there are commonly used thresholds

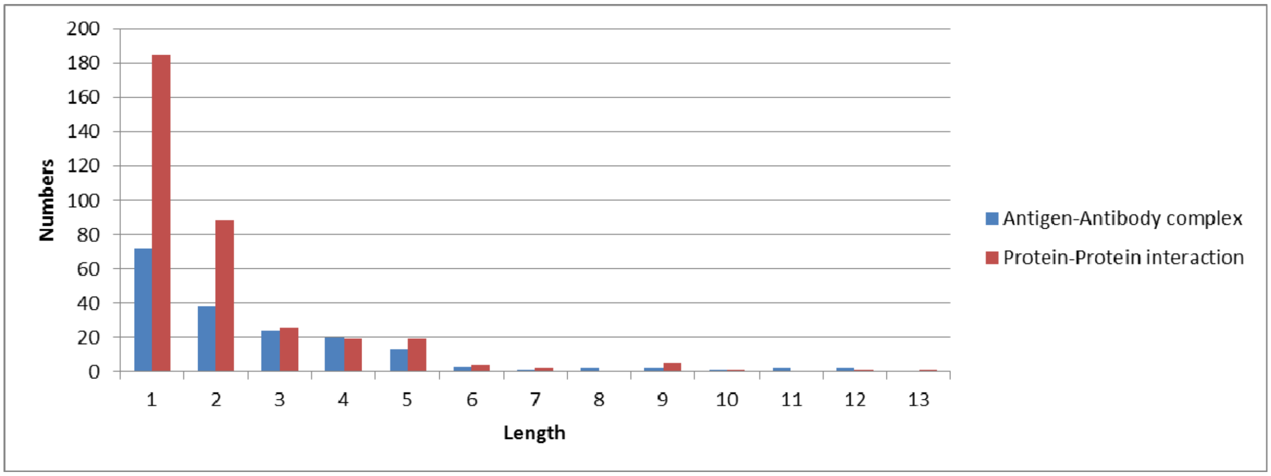


Figure 1. The number of epitope segments with different length. The X-axis indicates the length of segments; the Y-axis indicates the number of corresponding epitope segments with different length in our datasets.

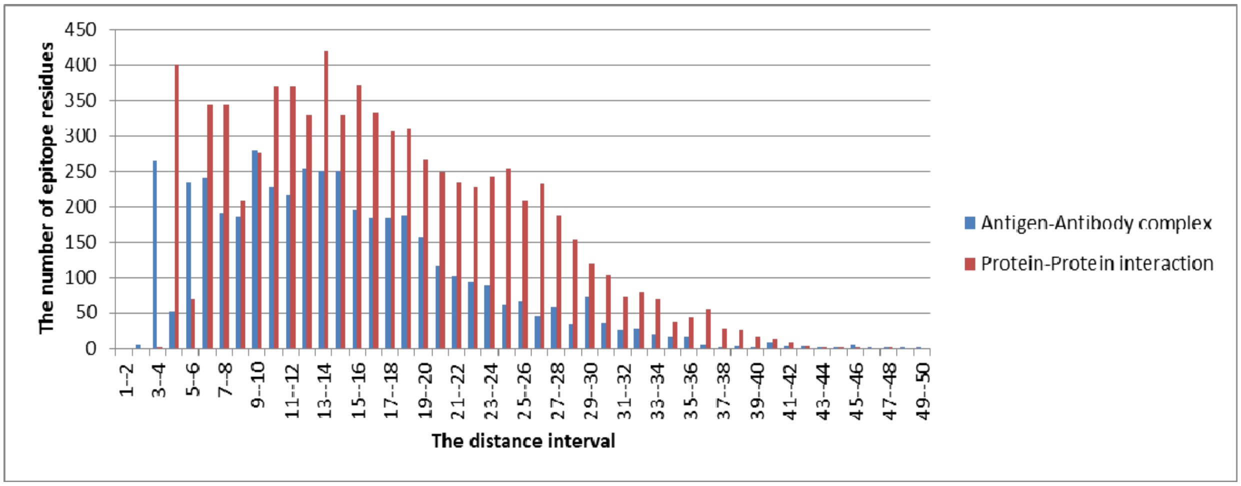


Figure 2. The number of epitope residues in each distance interval with the step size of 1Å. The X-axis indicates the distance intervals; the Y-axis indicates the number of epitope residues in each distance interval in our datasets.

for defining patch, there is no statistic on which extent these thresholds are accurate in the field of B-cell epitope prediction.

In this research, we make statistics about the relationship between the size of epitope patch which defined by means of number or radius and the proportion of epitope residues on the patch (which contains the most epitopes of each antigen structure surface). From the perspective of the number, the method takes the C_α (alpha-carbon atom) of every antigen surface residue as the center, and calculates some surface residues to become a patch, while the number of residues is from 0 to 400. From the perspective of the radius, the method also takes the C_α of every antigen surface residue as the center, but it divides antigen surface into overlapping patches with a radius, while the assignment of radius is from 0 Å to 50 Å. The proportion of epitopes in the patch which defined by the two means are shown in Figure **3** and Figure **4**.

From Figure **3**, it can be seen that for both antigen-antibody complexes and protein-protein interactions, dividing antigen surface patch from the perspective of number has little difference. When the number of the patch is 17, the proportion of epitopes in the patch reached 61.51%; and when the number of the patch is 31, the proportion of epitopes in the patch reached 80.22%. It is obvious that the proportion of epitopes in the patch is nearly 100% with the number of the patch is 96 (the proportion of epitopes in the patch is 99.09%) for both kinds of datasets.

From Figure **4**, it can be seen that for both antigen-antibody complexes and protein-protein interactions, dividing antigen surface patch from the perspective of radius has little difference. In terms of antigen-antibody complexes, when the radius of the patch is 13 Å, the proportion of epitopes in the patch is more than 80% (it reached 81.06%). Similarly, in the case of protein-protein interactions, when the radius of the patch is 17 Å, the proportion of epitopes in the patch is more than 80% (it reached 82.62%). When the radius is 26 Å, the proportion of epitopes contained in the patch is 99.59%. These two Figures show the proportions of epitopes are contained in the patch, which is defined by two ways. The results may provide help for future research.

2.1.4. Minimum Length of Epitope Surface Path

Some mimotope-based conformational B-cell epitope prediction methods use the idea of graph mapping. These prediction methods map the mimotope sequences to the original antigen surface. In this work, we calculate the minimal surface path length for each case in the datasets. The results are shown in Figure **5**.

Figure **5** shows the minimal length of epitope surface path in each testing case. Since antigenicity is generally considered as a surface effect on an approximated surface shape,

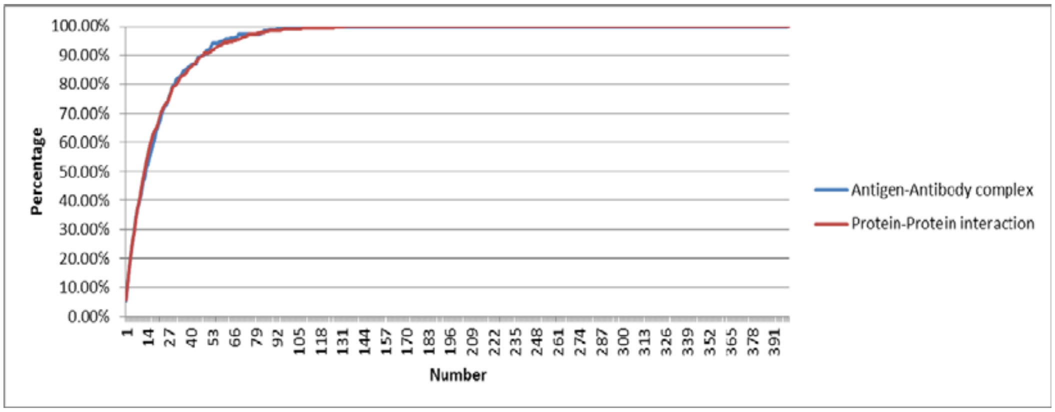


Figure 3. The proportions of epitopes contained in the patch which defined by the number. The X-axis indicates the numbers of surface amino acids within the patch; the Y-axis indicates the proportion of epitopes is contained in the patch.

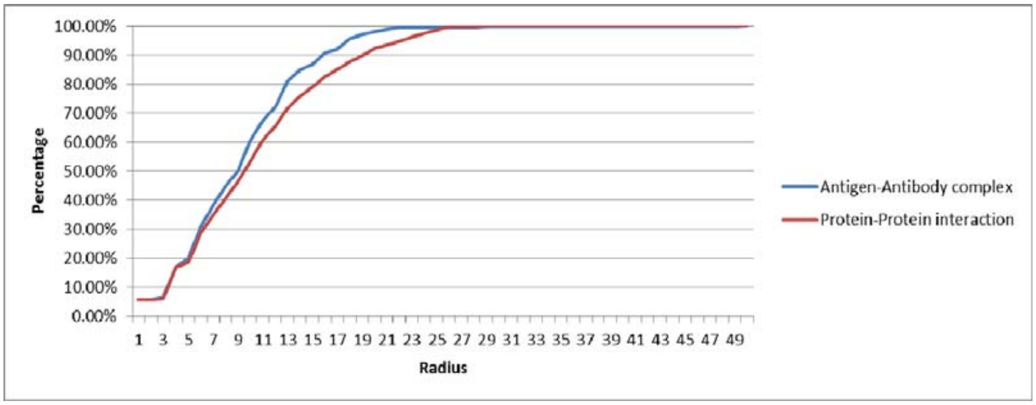


Figure 4. The proportions of epitopes contained in the patch which defined by the radius. The X-axis indicates the radiuses; the Y-axis indicates the proportion of epitopes is contained in the patch.

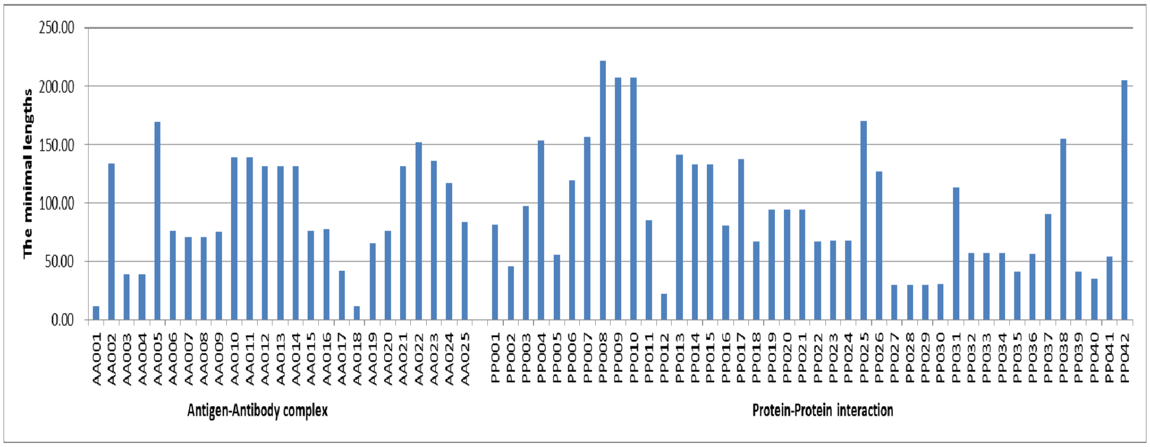


Figure 5. The minimal lengths of epitope surface path in each testing case. The X-axis indicates the testing cases in the datasets; the Y-axis indicates the minimal lengths of epitope surface path in each testing case.

the path which obtained by walking over the surface can reflect the geometry of the antigen. This backbone geometry of real epitopes may have more significance in conformational B-cell epitope prediction. The minimal surface paths for each testing case are given as Additional Materials.

2.2. Residual Characteristics of Conformational B-cell Epitope and Mimotopes

To know the relationship between conformational B-cell epitope and mimotopes, we choose the following parameters to make statistic: amino acid preference, residue-pair preference, residue accessibility.

2.2.1. Preference of Amino Acid in Epitope, Mimotopes and on the Surface

We calculate the proportion of amino acids in epitopes, mimotopes and on the surface respectively in the datasets to verify whether there are different preference of amino acid residues between epitopes and mimotopes. The proportion of amino acids in the antigen surface is calculated along to see whether epitopes and mimotopes divert from this, or one of them is closer to baseline. The results are shown in Figure **6**.

From Figure **6**, it can be seen obviously that the preferences are different between antigen-antibody complexes and protein-protein interactions. In terms of antigen-antibody complexes, the appearance frequencies of P, D and G(their frequencies are more than 7.5%) are higher in epitopes, while W and C (their frequencies are less than 2.0%) are lower in epitopes**;** the appearance frequencies of S, L and R (their frequencies are more than 7.5%) are higher in mimotopes, while the appearance frequency of M (the frequency is 1.89%) is lowest in mimotopes**;** the appearance frequencies of S (the frequency is 9.46%) are highest in surface residues, while W and M (their frequencies are less than 2.0%) are lower in surface residues. Similarly, in the case of protein-protein interactions, the appearance frequencies of L and E (their frequencies are more than 7.5%) are higher in epitopes, while C and M (their frequencies are less than 2.0%) are lower in epitopes; the appearance frequencies of L, P and R (their frequencies are more than 7.5%) are higher in mimotopes, while the appearance frequency of C (the frequency is 2.08%) is lowest in mimotopes; the appearance frequencies of L, E and D (their frequencies are more than 7.5%) are highest in surface residues, while W, C, M and H (their frequencies are less than 2.0%) are lower in surface residues. The graph indicates the preference of amino acids has difference in epitopes and mimotopes, even for both antigen-antibody complex and protein-protein interaction. Overall, the distribution of epitope residues is nearest to baseline.

Different amino acids were usually grouped into classes according to the properties, including acidic, polarity, hydrophobic，aromatic and so on. In this research, we grouped the twenty types of amino acids into twelve kinds as frequently-used [25], and the details are given in Table **2**.

Table 2. 12 groups of amino acids.

Properties	Amino acids
Acidic	ASP, GLU
Basic	ARG, HIS, LYS
Neutral	ALA, GLY, ILE, LEU, MET, PHE, PRO, TRP, TYR, VAL
Aliphatic	ALA, GLY, ILE, LEU, VAL
Aromatic	HIS, PHE, TRP, TYR
Cyclic	HIS, PHE, PRO, TRP, TYR
Acyclic	ALA, ARG, ASN, ASP, CYS, GLU, GLN, GLY, ILE, LEU, LYS, MET, SER, THR, VAL
Hydrophobic	ALA, GLY, ILE, LEU, MET, PHE, PRO, TRP, TYR, VAL
Polar	ARG, ASN, ASP, CYS, GLU, GLN, HIS, LYS, SER, THR
Small	ALA, GLY, SER
Medium	ASN, ASP, CYS, PRO, THR, VAL
Large	ARG, GLU, GLN, HIS, ILE, LEU, LYS, MET, PHE, TRP, TYR

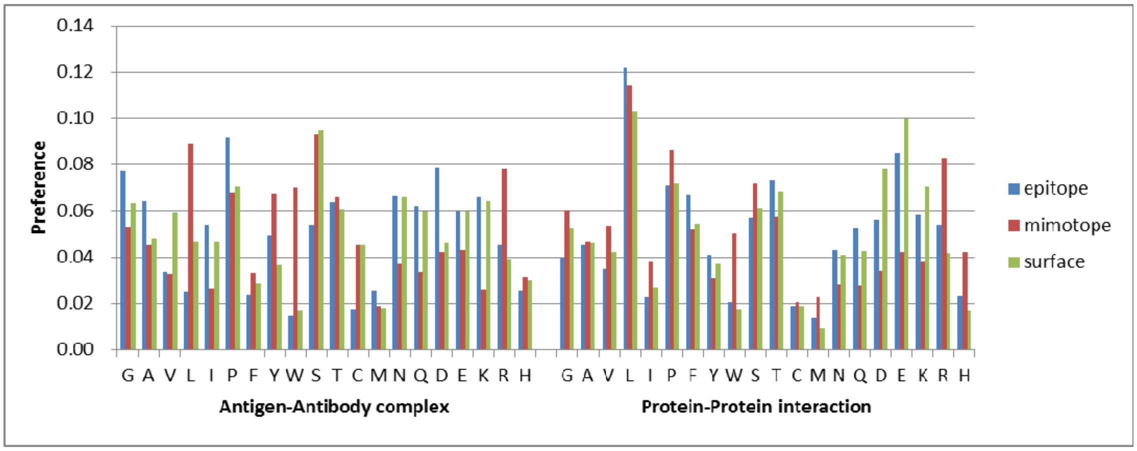


Figure 6. The preference of amino acid in epitopes, mimotopes and on the surface. The X-axis indicates twenty types of amino acid and their preference in epitopes, the Y-axis indicates mimotopes and on the surface. The bars in blue represent for the residue preference in epitopes, the red are represent for the residue preference in mimotopes and the green ones represent for the residue preference of antigen surface. (The color version of the figure is available in the electronic copy of the article).

Then, we calculate the proportion of 12 groups of amino acids which are in the epitopes and mimotopes by the datasets. The results are relatively smooth and stable, and it can be seen that the appearance frequency of all 12 groups is nearly the same in epitopes and mimotopes (Figure **7**). Compared with the proportion of these 12 groups of amino acids in the antigen surface, the results were more or less the consistent. That is, taking the 20 types of amino acids as 12 groups in the mimotope-based epitope predictions is a better choice.

2.2.2. Preference of Residue-neighbor Sets in Epitope and Mimotopes

Amino acid pair is an important measure in conformational B-cell epitope prediction; therefore, we calculate the preference of residue-neighbor sets in both epitopes and mimotopes in the dataset. The results for preference of residue-neighbor sets in epitopes and mimotopes are displayed in Figure **8**.

In Figure **8**, the color palette from ochre to blue indicates a growing preference of residue-neighbor sets in epitopes and mimotopes. The blue color indicates the higher probability of appearance for residue-neighbor sets in epitope or mimotope areas, while the ochre color means less appearance frequency. For antigen-antibody complex cases, compared with the appearance of residue-neighbor pairs in mimotopes, G-D, M-D and D-D residue-neighbor sets (their frequencies are more than 5.5%) were preferred in epitopes. While compared with epitope region, P-W, S-P, S-D, D-L and W-R residue-neighbor sets (their frequencies are more than 1.2%) were preferred in mimotopes. In addition, there are 54 kinds of residue-neighbor sets not appear in epitopes and only 7 kinds of residue-neighbor sets not appear in mimotopes. For protein-protein interaction cases, compared with the appearance of residue-neighbor pairs in mimotopes, G-D, M-D, A-M and F-R residue-neighbor sets (their frequencies are more than 5.5%) were preferred in epitopes. While compared with epitope region, P-L and L-P residue-neighbor sets (their frequencies are more than 1.2%) were preferred in mimotopes.

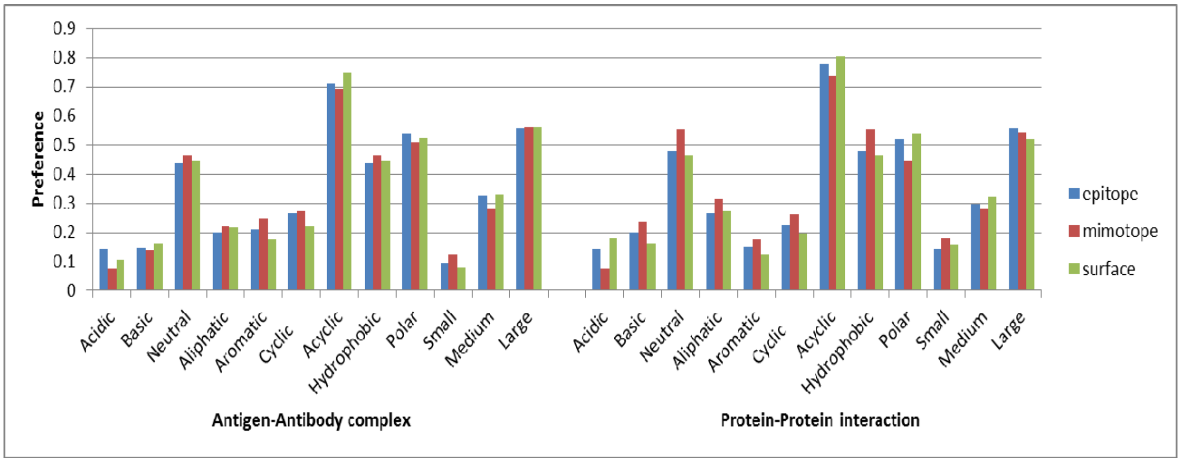
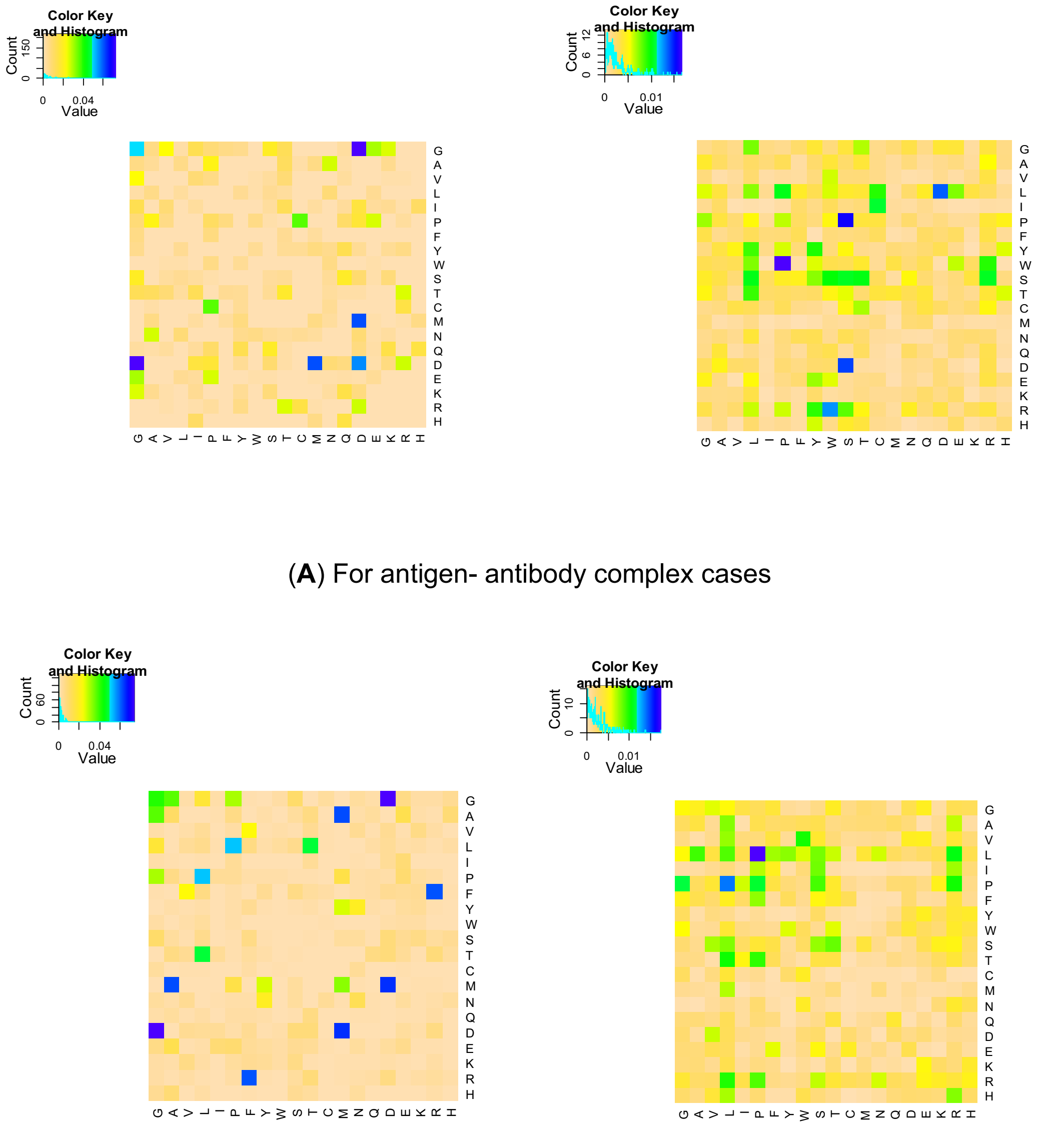


Figure 7. The preference of amino acid in epitopes, mimotopes and on the surface (amino acids are grouped into different classes according to the properties). The X-axis indicates twelve groups of amino acids and their preferences in epitopes, the Y-axis indicates mimotopes and on the surface.



(**A**) For antigen- antibody complex cases

(**B**) For protein-protein interaction cases

Figure 8. The preference of residue-neighbor sets in epitope and mimotopes. The figures both (**A**) and (**B**) in left side are the heat maps for epitopes, and the figures in right side are the heat maps for mimotopes. The color palette from ochre to blue indicates a growing preference for residue-neighbor sets in epitopes or mimotopes. (The color version of the figure is available in the electronic copy of the article).

In addition, there are 33 kinds of residue-neighbor sets not appear in epitopes and 15 kinds of residue-neighbor sets not appear in mimotopes. These results may also verify the point that antigen-antibody interaction is a kind of special protein-protein interaction, and has its own features [26]. From this point, the protein-protein interaction cases seem not suitable for effective verification of epitope prediction method.

2.2.3. Statistic Based on Residue Accessibility

In the process of protein binding, interacting residues are generally consider to have relatively higher accessibility, so that it can promote the contact with interacting counterpart [27]. The related research work has demonstrated that except for CYS, all other kinds of amino acids are more accessible in epitope areas than in non-epitope surface areas, and such difference was statistical significance [25].

In the research, we compared the solvent accessible surface areas (ASA) of epitope residues and mimotopes. We compare the relative ASA (RSA) of epitopes and mimotopes of the datasets by mean. The average values are calculated using formula (1).

Average value:

$$RSA_{ave}(i) = \frac{\sum_{j=1}^{n} x_{ij}}{n} \quad (i = 1, 2, \dots, 20; j = 1, 2, \dots, n) \quad (1)$$

In the formula above, x_i indicates a type of amino acid, x_{ij} indicates the RSA value of the jth amino acid x_i ; n indicates the number of amino acid x_i appeared in the datasets of epitopes or mimotopes. The results are shown in Figure **9**.

From figure **9**, it can be seen that almost all results, the average values of RSA in epitopes are higher than in mimotopes for both kinds of cases (except amino acid D, K and R in antigen- antibody complex cases). The results indicated that the RSA of epitopes is higher than mimotopes in average; and this may imply that RSA is another effective feature in mimotope-based conformational B-cell epitope prediction.

3. METHODS

3.1. Datasets

A reliable dataset should meet the requirement of non-redundant antigen structures, well-defined B-cell epitopes, and the mimotope sequences. Non-redundant and abundant datasets could avoid the performance of B-cell epitope prediction methods overly optimistic. Well-defined B-cell epitope is the premise of epitope relevant feature extraction, and directly impacts the prediction performance. Mimotopes sequence is especially important for the mimotope-based conformational B-cell epitope prediction. Furthermore, large and reliable datasets is important for statistic.

We use the new version of the benchmark datasets which we called Benchmark 2.0 for conformational B-cell epitope prediction using random peptide library screening. The datasets were derived from MimoDB and PDB. The Benchmark 2.0 consists of 40 complex structures with 67 mimotope sets; and the 67 complex cases contain 25 antigen-antibody complexes and 42 protein-protein interactions structures.

3.2. Definitions

3.2.1. Epitope Definition

During our research work, the functional epitopes for antigen proteins were first selected; while the functional epitopes can be obtain in CED [28] and IEDB [29]. For the one which has no functional epitopes, the structural epitopes which are defined as the residue of antigen which has a contact area above 4 Å^2 upon interaction with the antibody are used.

3.2.2. Residue-neighbor

Epitopes are always binding to several residues in the vicinity of antibody during the process of immune response. Residue-neighbor for each epitope residue was proposed to reflect such relation for further analysis. For epitopes, if the distance between any two residues in an epitope is less than 4Å, we called it a residue-neighbor. The method takes the C$_\alpha$ of every antigen surface residue as the center. For mimotopes, residue-neighbor was defined as the neighbor amino acids in primary mimotope sequences.

3.2.3. Residue Accessibility

In this paper, we calculate the relative ASA (RSA) for each epitope residue using formula (2):

$$RSA = \frac{ASA}{Max_{ASA}} \tag{2}$$

In the formula above, the ASA can be calculated through Surface Racer 4.0[30] with the radius of the probe is 1.4 Å, and Max$_{\text{ASA}}$ for twenty types of amino acids was the ASA value of residue X in tri-peptide ALA-X-ALA is residue maximum solution accessibility surface area. We adopt the calculation results (the values of MaxASA for twenty types of amino acids) of Ahmad S [31]. For mimotopes, we get the results from the SANN [32]. SANN is a web server for prediction of protein solvent accessibility by nearest neighbor method. SANN server can be accesses through http://lee.kias.re.kr/~newton/sann/.

RESULTS AND DISCUSSION

B-cell epitope prediction is important for vaccine design, and development of diagnostic reagents. It is also indispensable for elucidate the interactions between antigen and antibody. However, the improvement of prediction accuracy and efficiency for conformational epitope hinders its progress. In previous studies, the existing mimotope-based methods use only the mimotopes sequence information, and the perform-

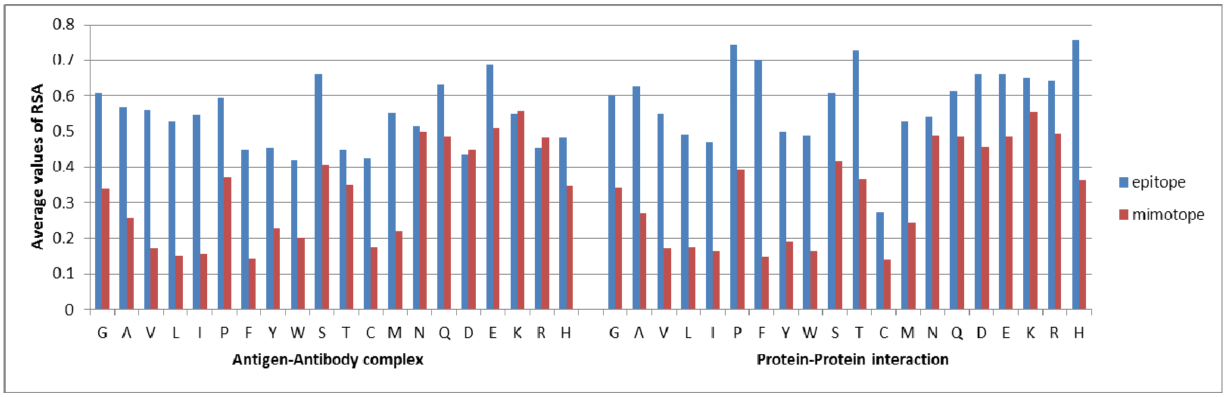


Figure 9. Average values of RSA for twenty types of amino acid. The X-axis indicates twenty types of amino acid and the Y-axis indicates their preference in epitopes and in mimotopes. The bars in blue represent for the preference of epitope residues and the red are represent for the preference of mimotope residues. (The color version of the figure is available in the electronic copy of the article).

ance of these algorithms did not improve more in a long time. That may be able to say, they encountered the bottleneck of development.

In this paper, we do a series of research work in order to systematically detect the relationship of mimotopes and conformational B-cell epitopes, and try to evaluate existing mimotope-based B-cell epitope prediction methods from the view of the relationship between B-cell epitope and mimotope sequences, and finally do our best to help evaluate the performance of the prediction algorithms based on mimotopes. From several aspects, this paper analyses the relationship between the epitopes and mimotopes, and tries to dig out of their deep relationship besides sequences.

We choose some parameters used to describe conformational epitopes from B-cell protein antigens, meanwhile we choose relatively few parameters to compare epitopes and mimotopes because of structural constraints. These parameters are classified into two perspectives, including general characteristics and residue characteristics.

The general characteristics including sequence segment of epitopes, distance between every two epitope residues, epitope patch and the minimum length of epitope surface path. We have calculated the numbers of sequence segments with different lengths and made statistics about the Euclidean distances between every two epitope residues by the datasets. Then we divide the antigen surface into patches from the perspective of both distance and radius, and calculate the proportion of epitopes on the patch under each parameter. Last we construct surface complete weighting graph for epitopes and give the minimum length of epitope path. The results may give guidelines to future B-cell epitope prediction.

The parameters of residue characteristics contain the amino acid preference, residue-pair preference, residue accessibility. For the amino acid preference, we calculate the proportion of 20 kinds of amino acids in the epitopes and mimotopes by the datasets, and along with the proportions on surface. Then we group 20 kinds of amino acids into 12 classes by their properties, and calculate the proportion of 12 classes of amino acids on the datasets. The result shows that grouping 20 kinds of amino acids into 12 classes is reliably in the field of conformational B-cell epitope prediction. For the residue-pair preference, we calculate the preference of residue-neighbor sets in epitopes and mimotopes by the datasets. The result shows that the preference of residue-neighbor sets is discrepant, and combining this specific feature of epitope-paratope into mimotope-based conformational B-cell epitope prediction may improve the performance of the method. For residue accessibility, we compare the relative solvent accessible surface areas of epitope residues and mimotopes and get that mimotopes contain little accessible feature, and add this solvent accessible feature into mimotope-based conformational B-cell epitope prediction may be a better choice.

For both general characters and residue characters, all the statistics are made on antigen-antibody complex cases and protein-protein interaction cases respectively. The results shows there are differences between these two kinds of interactions. In other words, protein-protein interactions are not suitable for the both training and testing in B-cell epitope predictions.

CONCLUSIONS

According to the analysis results, most of conformational B-cell epitopes are consist of continuous sequence segment. The distance between majority epitope residues is less than 20 Å. As well as the distance standard, most of the epitope residues are on the surface patch with the number of 31 residues or the radius of 15Å (more than 80% in the dataset). Besides, statistical differences are found between epitope and mimotopes with parameters in residual and sequence levels. The preference of 20 types of amino acids in the epitopes and in the mimotopes is different, while this difference decreases when classify the amino acids into 12 kinds according to properties. Amino acid enrichment and preference for specific types of residue-neighbor sets on epitopes and mimotopes have also been observed. The occurrences of residue-neighbor sets on epitopes and mimotopes show some patterns. In addition, the residue accessibility for epitopes and mimotopes is investigated, and the epitope ones are more accessible when comparing to mimotopes for both antigen-antibody complex and protein-protein interaction cases.

The results show some rules exist in conformational epitopes in segment, distance and the size of patch. This difference can give guidance for future conformational B-cell epitope prediction. On the other hand, the preference of amino acids in epitopes and mimotopes show classify amino acids into different kinds according to properties is reliable for mimotope-based B-cell epitope predictions. The residue-neighbor and residue accessibility characters indicate there is statistic difference between epitopes and mimotopes. This observation further gives new hope for conformational B-cell epitope predictions. In that case, it is a more effective and accurate approach to predict the conformational epitopes based on mimotopes.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

This work was supported by National Natural Science Funds of China (No. 61402098), the Science Foundation for Young Teachers of Northeast Normal University (NO. 14QNJJ030, 14QNJJ030), the China Postdoctoral Science Foundation (2013M541267, 2014M561273, 2015T80285), the 2012 postdoctoral research projects of Jilin Province (111900166), the Jilin Scientific and Technological Development Program (No. 20140520072JH), the Research Fund for the Doctoral Program of Higher Education of China (No. 20130043110016), the Natural Science Foundation of Jilin Province (No. 20140101179JC). CHZ conceived the idea and designed whole work. YYL, WNT designed experiments, gathered test data, and ZGZ was in charge of the statistics. ZQM suggested extension and modifications to the research. PPS supervised the whole research and revised the manuscript critically. All authors have read and approved the final manuscript.

REFERENCES

[1] Abbas, A.K.; Lichtman, A.H.; Pillai, S. *Cellular and Molecular Immunology*, Saunders Elsevier, 6 Ed.; W.B. Saunders Company. **2009**.

[2] Flower, D.R. Towards in silico prediction of immunogenic epitopes. *Trends Immunol.*, **2003,** *24*, 667-674.

[3] Rux, J.J.; Burnett, R.M. Type-specic epitope locations revealed by X-ray crystallographic study of adenovirus type 5 hexon. *Mol. Ther.*, **2000**, *1*, 18-30.

[4] Mayer, M.; Meyer, B. Group epitope mapping by saturation transfer difference NMR to identify segments of a ligand indirect contact with a protein receptor. *J. Am. Chem. Soc.*, **2001**, *123*, 6108-6117.

[5] Greenbaum, J.A.; Andersen, P.H.; Blythe, M.; Bui, H.H.; Cachau, R.E.; Crowe, J.; Davies, M.; Kolaskar, A.; Lund, O.; Morrison, S. Towards a consensus on datasets and evaluation metrics for developing B-cell epitope prediction tools. *J. Mol. Recog.*, **2007**, *20*, 75-82.

[6] Korber, B.; LaBute, M. and Yusim, K. Immunoinformatics comes of age. *Plos Comput. Biol.*, **2006**, *2*, 484-492.

[7] Sun, P.; Ju, H.; Liu, Z.; Ning, Q.; Zhang, J.; Zhao, X.; Huang, Y.; Ma, Z.; Li, Y. Bioinformatics Resources and Tools for Conformational B-Cell Epitope Prediction. *Comput. Math. Methods Med.*, **2013**, *20*, 75-82.

[8] Kulkarni-Kale, U.; Bhosle, S.; Kolaskar, A.S. CEP: a conformational epitope prediction server. *Nucleic Acids Res.*, **2005**, *33*, W168-W171.

[9] Andersen, P.H.; Nielsen, M.; Lund, O. Prediction of residues in discontinuous B-cell epitopes using protein 3D structures. *Protein Sci.*, **2006**, *15*, 2558-2567.

[10] Ponomarenko, J.; Bui, H.; Li, W. ElliPro: a new structure-based tool for the prediction of antibody epitopes. *BMC Bioinformatics*, **2008**, *9*, 514.

[11] Sweredoski, M.J.; Baldi, P. PEPITO: improved discontinuous B-cell epitope prediction using multiple distance thresholds and half sphere exposure. *Bioinformatics*, **2008**, *24*, 1459-1460.

[12] Sun, J.; Wu, D.; Xu, T. SEPPA: a computational server for spatial epitope prediction of protein antigens. *Nucleic Acids Res.*, **2009**, *37*, W 612-W 616.

[13] Rubinstein, N.D.; Mayrose, I.; Martz, E.; Pupko, T. Epitopia: a web-server for predicting B-cell epitopes. *BMC Bioinformatics*, **2009**, *10*, 6-11.

[14] Geysen, H.M.; Rodda, S.J.; Mason, T.J. A priori delineation of a peptide which mimics a discontinuous antigenic determinant. *Mol. Immunol.*, **1986**, *23*, 709-715.

[15] Moreau, V.; Granier, C.; Villard, S.; Laune, D.; Molina, F. Discontinuous epitope prediction based on mimotope analysis. *Bioinformatics*, **2006**, *22*, 1088-1095.

[16] Castrignanò, T.; De Meo, P.D.; Carrabino, D. The MEPS server for identifying protein conformational epitopes. *BMC Bioinformatics*, **2007**, *8*, supplement 1.

[17] Enshell-Seijffers, D.; Denisov, D.; Groisman, B. The mapping and reconstitution of a conformational discontinuous B-cell epitope of HIV-1. *J. Mol. Biol.*, **2003**, *334*, 87-101.

[18] Bublil, E.M.; Freund, N.T.; Mayroseet, I. Stepwise prediction of conformational discontinuous B-cell epitopes using the Mapitope algorithm. *Proteins*, **2007**, *68*, 294-304.

[19] Mayrose, I.; Shlomi, T.; Rubinstein, N.D. Epitope mapping using combinatorial phage-display libraries: a graph-based algorithm. *Nucleic Acids Res.*, **2007**, *35*, 69-78.

[20] Huang, Y.; Bao, Y.; Guo, S.; Wang, Y.; Zhou, C.; Li, Y. Pep-3D-Search: a method for B-cell epitope prediction based on mimotope analysis. *BMC Bioinformatics*, **2008**, *9*, 538.

[21] Negi, S.S.; Braun, W. Automated detection of conformational epitopes using phage display peptide sequences. *Bioinform. Biol. Insights*, **2009**, *3*, 71-81.

[22] Chen, W.; Sun, P.; Lu, Y. MimoPro: a more efficient web-based tool for epitope prediction using phage display libraries. *BMC Bioinformatics*, **2011**, *12*, 199.

[23] Sun, P.; Chen, W.; Huang, Y.; Wang, H.; Ma, Z. Lv, Yi. Epitope Prediction Based on Random Peptide Library Screening: Benchmark Dataset and Prediction Tools Evaluation. *Molecules*, **2011**, *16*, 4971-4993.

[24] http://cs.nenu.edu.cn:8080/bioinformatics/benchmark%20datasets/index.jsp

[25] Sun, J.; Xu, T.; Wang, S.; Li, G.; Wu, D.; Cao, Z. Does difference exist between epitope and non-epitope residues? Analysis of the physicochemical and structural properties on conformational epitopes from B-cell protein antigens. *Immun. Res.*, **2011**, *7*, 1-11.

[26] Liang, S.; Zheng, D.; Zhang, C. Prediction of antigenic epitopes on protein surfaces by consensus scoring. *BMC Bioinformatics*, **2009**, *10*, 302.

[27] Rubinstein, N.D.; Mayrose, I.; Halperin, D.; Yekutieli, D.; Gershoni, J.M.; Pupk, T. Computational characterization of B-cell epitopes. *Mol. Immunol.*, **2008**, *45*, 3477-3489.

[28] Huang, J.; Honda, W. CED: a conformational epitope database. *BMC Immunol.*, **2006**, *7*, 7.

[29] Vita, R.; Zarebski, L.; Greenbaum, J.A.; Emami, H. The Immune Epitope Database 2.0. *Nucleic Acids Res.*, **2010**, *38*, 1.

[30] Tsodikov, O.V.; Record, M.T.Jr.; Sergeev, Y.V. Novel computer program for fast exact calculation of accessible and molecular surface areas and average surface curvature. *J. Comput. Chem.*, **2002**, *23*, 600-609.

[31] Ahmad, S.; Gromiha, M.; Fawareh, H. ASAView: Database and tool for solvent accessibility representation in proteins. *BMC Bioinformatics*, **2004**, *5*, 51.

[32] Keehyoung, J.; Sung, J.L.; Jooyoung, L. SANN: Solvent accessibility prediction of proteins by nearest neighbor method. *Proteins*, **2012**, *80*, 1791-1797.

Received: July 3, 2015 **Revised: October 30, 2015** **Accepted: November 16, 2015**