Int J Clin Exp Med 2016;9(2):2748-2757
www.ijcem.com /ISSN:1940-5901/IJCEM0017209

Original Article

Prediction of conformational B-cell epitope binding with individual antibodies using phage display peptides

Chunhua Zhang[1,2], Bojian Sun[1,2], Weina Tang[1,2], Pingping Sun[1,2,3], Zhiqiang Ma[1,2]

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

Received October 1, 2015; Accepted December 18, 2015; Epub February 15, 2016; Published February 29, 2016

Abstract: B-cell epitopes are a group of residues on the surface of an antigen which invoke strong humoral responses. Predicting epitopes is crucial for disease diagnosis, vaccine design, antibody design and immunological therapy. It is also indispensable for elucidating interactions between antigen and antibody on a molecular level. Since reliable experimental methods are expensive and time consuming, the prediction of B-cell epitope has developed quickly in recent years, but the performance of current prediction methods is far from satisfactory. The results show that our method achieves a sensitivity of 0.59, specificity of 0.64, positive predictive value of 0.45, accuracy of 0.70 and F-measure of 0.49. Compared with the antibody-specific epitope propensity method, the Mapitope method and four other methods, our method could provide reliable, rational results and showed comparable sensitivity, precision and accuracy. In this work, we propose a new conformational B-cell epitope prediction method by binding with individual antibodies using phage display peptides. It is more effective when used to predict the conformational B-cell epitope binding with an individual antibody.

Keywords: Antibody, epitope prediction, mimotope

Introduction

A B-cell epitope [1] is a special group of residues on the surface of an antigen which invokes strong humoral responses. It is recognized by either a particular B-cell receptor (BCR) or a particular antibody molecule of the immune system [2]. Locating B-cell epitopes is important for vaccine design [3], development of diagnostic reagents and interpretation of the antigen-antibody interactions on a molecular level [4]. The experimental approaches are reliable, but they are always laborious and complicated. With the development of information technology and the increasing verification number of epitopes, the B-cell epitope prediction by computational method arises at a historic moment [5].

B-cell epitopes can be classified into two categories by their spatial structure: linear epitopes and conformational epitopes [6]. The linear epitopes are composed of short contiguous stretches of amino acids in the antigen sequence. The conformational epitopes are composed of sequential segments that are brought together in spatial proximity when the corresponding antigen is folded. It has been demonstrated that more than 90% of B-cell epitopes were conformational [7]. Therefore, conformational epitope predictions are more meaningful [8].

The conformational B-cell epitope prediction methods can be classified into two kinds: the method based on the 3D structure of the antigen and the method based on random peptide library screening [9]. In recent years, both kinds of methods have developed quickly.

Antigen structure-based conformational B-cell epitope prediction methods took the 3D structure of the antigen as input. These methods used the 3D structure of antigen, epitope-related propensity scales, geometric attributes and specific physicochemical properties. The first structure-based conformational B-cell epitope prediction method was CEP [10]. Then other methods based on different structure charac-

Table 1. The testing dataset of the work

PDB_ID	Template chain	Target	Mimotope size	Reference
1D4V	B	receptor superfamily member 10B	13*9	20156289
1EER	A	EPO-R	1*8	8662529
1FLT	X	Vascular endothelial growth factor A	4*7	17401149
1G1S	D	P -selectin	5*17	12393589
1G9M	G	Anti -gp120 monoclonal antibody 17b	10*12,1*10	14596802
1HX1	B	Heat shock cognate 71 kDa protein	8*15	7649995
1II4	A	Fibroblast growth factor receptor 2	30*7	12032665
1JRH	I	A6	59*5	11123892
1K4U	S	Neutrophil cytosol factor 2	28*9,2*10,4*12,2*6,1*8	8663333
1MQ8	B	Intercellular adhesion molecule 1	12*9,1*8	11532073
1OC0	B	Plasminogen activator inhibitor 1	8*13,1*7,1*11	16813566
1SHY	A	Hepatocyte growth factor receptor	2*12,1*13	17947467
1SQ0	A	Platelet glycoprotein Ib alpha chain	3*11	18363340
1TET	P	TE33	10*9	16273596
1WLP	A	Neutrophil cytosol factor 1	4*5,3*9,1*10,1*8	7624379
1WLP	B	Cytochrome b-245	2*8,31*9	7592831
1YCR	B	MDM2	20*16	21423613
1ZTX	E	E16	14*22	16813566
2ADF	A	82D6A3	3*8	12855771
2C9F	T	Penton protein	6*3	7588601
2C9F	A	Fiber protein	6*38	7588601
2DSQ	I	IGFBP-1	1*20	9636028
2GHW	A	80R	9*16,9*15,19*14,4*13	16630634
2NY7	G	Anti-gp120 monoclonal antibody b12	1*10,1*13,17*14	16940148
2OSL	P	Rituximab	13*9	16705086
3BT1	A	Urokinase plasminogen activator surface receptor	19*15	8041758
3DAB	F	Cellular tumor antigen p53	9*12	19255450
3DOW	B	CRT peptide	5*12	17916189
3EZE	B	Phosphoenolpyruvate-protein phosphotransferase	11*6	9350871
3IU3	I	Basiliximab	6*9	17440057
4HTC	I	Alpha-thrombin	1*7	12565725

ters were subsequently proposed. The representative methods include DiscoTope [11], ElliPro [12], EPCES [13] and CeePre [14]. These methods have made some progress, but they are still at an infancy stage.

The random peptide library screening-based B-cell epitope prediction method was based on both the 3D structure of antigen and the peptide sequences obtained from phage peptide library screening [15]. To attain these peptide sequences, firstly the random peptides were displayed on the surface of filamentous phages; then the ones that bound to a monoclonal antibody with a certain degree of affinity were screened, eluted and amplified. After several rounds of this operation, there were fewer but higher affinity resulting peptides [16]. The selected affinity peptides were called mimotopes [17], and this kind of conformational B-cell epitope prediction was called the mimotope-based conformational B-cell epitope prediction. Mimotopes commonly share high sequential similarity which implies that certain key binding motifs and physicochemical preferences exist during the interactions [18].

The core idea of mimotope-based conformational B-cell epitope prediction was mapping these mimotopes back to the source antigen;

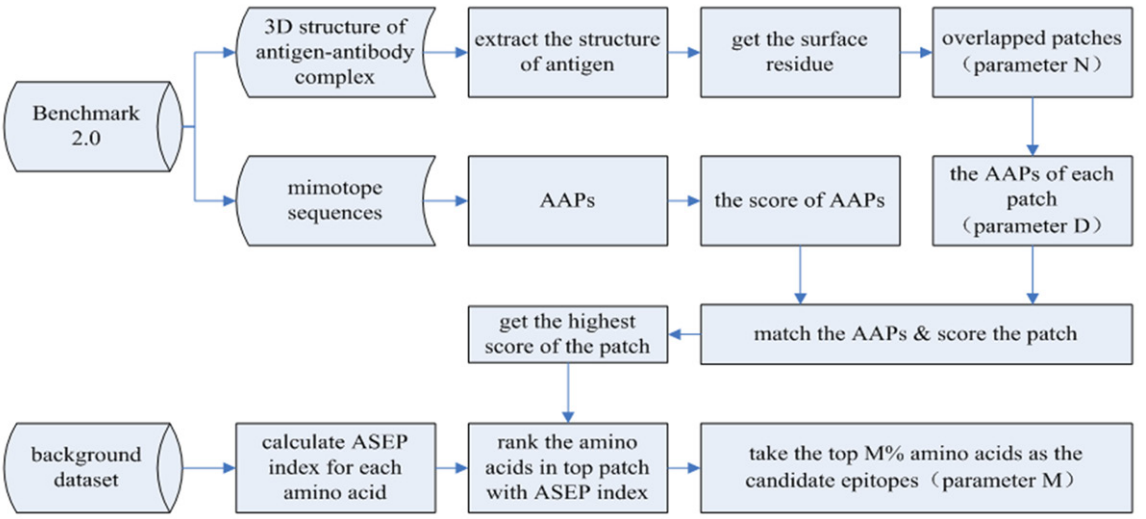

Figure 1. The algorithm flow diagram of the method.

this could help find the genuine epitopes more accurately. Since these affinity-selected peptides and the genuine epitopes both can combine the same paratope of monoclonal antibody and cause immune responses, it was inferred that similar function may have a similar sequence; so again, the selected mimotopes may share high sequential similarity.

In 1995, Pizzi et al [19] designed the first mimotope-based conformational B-cell epitope prediction method, MEPS [20]. Since then, some other algorithms and tools have been proposed, such as: FINDMAP [21]; Mimox [22]; PepSurf [23]; Pep-3D-Search [24]; EpiSearch [25] and MimoPro [26]. In 2011, our team constructed a benchmark dataset for mimotope-based conformational B-cell epitope prediction and evaluated five prediction tools. The results showed that in no method did the performance exceed a 0.42 precision and 0.37 sensitivity [27].

An antigen can be combined with multiple antibodies. The interaction sites of different antibodies with an antigen are not completely the same. The mimotope-based B-cell epitope prediction methods are based on affinity peptides which form a monoclonal antibody, therefore these kinds of prediction methods are mainly in connection with certain antibodies. Under this circumstance, we tried to add the information of the combination specificity of antigen and antibody to the conformational B-cell epitope prediction. In this work, we proposed a new mimotope-based conformational B-cell epitope prediction method. This new method is based on the antibody-specific epitope propensity (ASEP) index [28] and affinity-peptide sequences. Firstly, the ASEP index was calculated by a background database. Secondly, the method divided the surface of antigen into overlapping patches. Thirdly, the affinity peptide sequences were transformed into amino acids pairs (AAPs) [29] and the statistical significance of each AAP was calculated. Fourthly, every patch was scored according to the statistical significance of AAPs; and for the patch with the highest score; the amino acid in the patch was ranked by the ASEP index. Finally, the top 75% amino acids were taken as the candidate epitopes. We selected a testing dataset from Benchmark 2.0 [30] to evaluate the performance of our new method. On the testing dataset, the mean sensitivity of our method achieved 0.59, the mean specificity achieved 0.64 and the mean accuracy (ACC) achieved 0.70. We compared the performance of our method with some other commonly used methods, our method was a little better than the others on the testing dataset.

Materials and methods

Background dataset

We collected a relative comprehensive dataset which contains 161 antigen-antibody complexes for the conformational B-cell epitope prediction. The dataset was from the representative antigen-antibody complexes in Ponomarenko

Table 2. The results on the testing dataset of this algorithm

MimoDB	Complex	Sensitivity	Specificity	PPV	ACC	F-measure
PP002	1OC0_B	0.64	0.27	0.30	0.39	0.41
PP003	1HX1_B	0.45	0.69	0.28	0.64	0.35
PP004	1WLP_B	0.59	0.56	0.30	0.57	0.40
PP005	1WLP_A	0.87	0.40	0.68	0.68	0.76
PP006	1K4U_S	0.56	0.50	0.54	0.53	0.55
PP011	1FLT_X	0.41	0.62	0.23	0.58	0.30
PP012	3DOW_B	1.00	0.00	1.00	1.00	1.00
PP013	1SHY_A	0.10	0.81	0.07	0.72	0.08
PP015	1SQ0_A	0.26	0.78	0.17	0.70	0.21
PP016	1D4V_B	0.15	0.71	0.06	0.65	0.08
PP017	3BT1_A	0.26	0.73	0.25	0.61	0.25
PP019	3EZE_B	0.70	0.65	0.45	0.66	0.55
PP022	1MQ8_B	0.69	0.79	0.28	0.78	0.40
PP025	1II4_A	0.43	0.78	0.45	0.68	0.44
PP027	1G1S_D	0.83	0.40	0.63	0.64	0.71
PP030	2C9F_T	0.89	1.00	1.00	0.90	0.94
PP031	2C9F_A	0.13	0.92	0.08	0.88	0.10
PP034	2DSQ_I	0.75	0.52	0.38	0.58	0.50
PP035	1YCR_B	0.73	0.00	0.80	0.62	0.76
PP038	4HTC_I	0.57	0.75	0.70	0.66	0.63
PP040	3DAB_F	1.00	0.00	0.75	0.75	0.75
PP042	1EER_A	0.33	0.77	0.33	0.66	0.33
AA001	2OSL_P	0.89	0.64	0.57	0.72	0.69
AA002	3IU3_I	0.65	0.73	0.41	0.71	0.50
AA003	1TET_P	0.82	1.00	1.00	0.83	0.90
AA008	2ADF_A	0.50	0.51	0.17	0.51	0.25
AA010	2GHW_A	0.33	0.86	0.36	0.77	0.35
AA014	2NY7_G	0.64	0.94	0.55	0.90	0.59
AA016	1G9M_G	0.46	0.92	0.25	0.89	0.32
AA019	1ZTX_E	0.64	0.74	0.35	0.73	0.45
AA025	1JRH_I	0.89	0.86	0.64	0.86	0.74
AVG		0.59	0.64	0.45	0.70	0.49

and Bourne [31], the protein docking Benchmark 2.0 and the testing datasets in the relevant papers [12, 13, 26, 32, 33]. We selected all antigen-antibody complexes and excluded the redundant structures and the ones which have more than one antigen chain. The 3D structures of these complexes were obtained from Protein Data Bank (PDB) [34]. We calculated the ASEP index for each amino acid from this background dataset. This large and reliable dataset could meet the requirements for statistical analyses.

In order to calculate the ASEP index, we needed to know the epitope residues in advance. There are methods to define epitope according to the inference of the antigen-antibody complex 3D structure by X-ray diffraction. In this work, we define the epitope as the antigen residue where the distance from any antibody residue is less than 4 Å.

ASEP index

ASEP was first proposed by Shinji Soga in the structure-based conformational B-cell epitope prediction. Shinji used the ASEP index to judge the results of DiscoTope further. The structure-based method predicted epitopes through the structure of antigen and epitope-related propensity scales, and the method is used to predict all potential epitope residues. Comparatively, the mimotope-based conformational B-cell epitope predictions kept the focus on the epitope residues of certain antibodies. Therefore, the ASEP index may benefit the mimotope-based conformational B-cell epitope prediction. In this work, we used a background dataset to calculate the ASEP index.

Testing dataset

A testing dataset should meet the requirement of non-redundant antigen structures, the known B-cell epitopes and the mimotope sequences. A non-redundant and abundant dataset should ensure that the performance of B-cell epitope prediction methods is not overly optimistic. The mimotope sequences were especially important for mimotope-based conformational B-cell epitope prediction. In this work, we collected a dataset which contains 31 representative values from the Benchmark 2.0. Each value included the complex structure, the template chain, the mimotopes obtained from the corresponding phage display experiment and the epitope information. **Table 1** shows the testing dataset of this work.

Algorithm flow diagram

The algorithm flow diagram of this work is shown in **Figure 1**. Firstly, we calculated the ASEP index for each amino acid according to

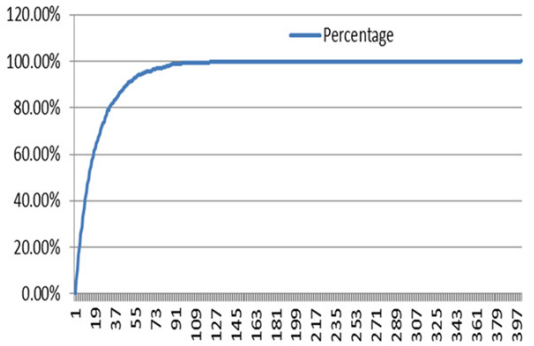

Figure 2. The proportion of epitopes in the patch on the background dataset.

the background dataset. Secondly, we transformed mimotope sequences into AAPs and scored each pair by the frequency of the pairs; meanwhile, we divided the antigen surface into overlapping patches and defined the AAPs in each surface patch. Thirdly, we scored each patch by AAPs. Fourthly, we ranked the amino acids in the patch which got the highest score by the ASEP index. Finally, the top 75% amino acids in that patch were taken as the candidate epitopes.

AAPs from mimotopes

In most fields of biology, it is considered that the fragment is more meaningful than the single amino acid. In this work, we took two amino acids as a fragment; these were then known as AAP. The concept of AAP was first used in Mapitope [35]. Mapitope first defined AAP with a predefined distance threshold between the central carbon atoms of two neighbor residues. After that, it defined statistically significant pairs (SSPs) by calculating the probabilities of each AAP, and then the SSPs were mapped to the 3D structure of an antigen to locate epitopes.

Our work transformed the mimotope sequence into overlapped AAPs. For example, a simple mimotope "CYAKS...." can be divided into AAPs as "CY" "YA" "AK" "KS" and so forth. To distinguish the AAPs in a patch, we named these types of AAPs as mAAPs. For an antigen-antibody complex, we got mAAPs of the whole mimotope sequence of the complex, and we calculated the frequency of their occurrence as the score of corresponding mAAPs.

Scoring antigen surface patch

Scoring antigen surface patch was an important algorithm in this method. It involves the following steps: 1) We extracted the unbound structure of antigen in the antigen-antibody complex structures. There are tools which can extract the unbound structure of the complexes. In this work, we employed the extract function of Pep-3D-Search. 2) For each antigen, we calculated the surface residues of the antigen. As epitope residues are always located on the surface of the antigen, the surface residues were first extracted from an antigen structure using solvent exposure. Solvent exposure was commonly measured by relative solvent accessibility (RSA) of the residue as the surface residue, with the formula: $RSA = \frac{ASA}{\max ASA}$. In the formula, ASA of a residue was calculated as the sum of exposed areas of atoms using a "rolling ball" algorithm which was developed by Shrake & Rupley. The algorithm was known as Surface Racer [36]. The maximum exposed area was measured as the exposed area of any type of amino acid in an ALA-X-ALA tripeptide [37]. In this work, firstly, ASA of each residue was calculated using Surface Racer 5.0 with a probe radius of 1.4 Å; secondly, the sum of ASA of all member atoms was calculated automatically; finally, any residue with an RSA larger than a predefined threshold of 0.05 $Å^2$ was determined as a surface residue. 3) We divided the antigen surface into overlapping patches. In this work, the antigen surface was divided into overlapped surface patches by a number N of 30. We took each surface residue as a center, and then we calculated the nearest 30 surface residues with the center to define the surface patch. 4) The antigen surface patch was transformed into AAPs. For the target antigen, the AAPs defined by this method were named pAAPs. For each patch, if the distance between any two amino acids was less than 9 Å, the two amino acids were taken to be a pAAP, and the patch was converted into a collection of pAAPs. 5) We used the scores of mAAPs to compute the score for the antigen surface patch. Finally, we matched the pAAPs in each patch and mAAPs. Then we assigned the score of mAAPs which were obtained from the mimotope sequences of the antigen to pAAPs of the patches in this antigen, and took the sum score of pAAPs as the score of the patch.

Table 3. The average prediction performance with different N

N	Sensitivity	Specificity	PPV	ACC	F-measure
20	0.50	0.61	0.37	0.70	0.37
21	0.52	0.62	0.38	0.71	0.38
22	0.51	0.62	0.39	0.70	0.39
23	0.51	0.63	0.38	0.70	0.38
24	0.53	0.65	0.44	0.71	0.39
25	0.51	0.63	0.37	0.70	0.39
26	0.54	0.69	0.35	0.69	0.41
27	0.53	0.62	0.42	0.72	0.45
28	0.56	0.66	0.43	0.71	0.46
29	0.53	0.60	0.37	0.70	0.48
30	0.59	0.64	0.45	0.70	0.49
31	0.56	0.66	0.44	0.69	0.49
32	0.51	0.69	0.44	0.70	0.42
33	0.53	0.61	0.42	0.70	0.46
34	0.54	0.70	0.42	0.68	0.43
35	0.51	0.65	0.41	0.68	0.45

Epitope prediction using ASEP index

After above steps, we took the highest scoring patch as the resulting patch, and ranked the amino acids in the patch by the ASEP index. In the end, the top 75% amino acids in the patch were predicted as the candidate epitopes. Specifically, so that the number of epitope residues to the number of antigen surface residues was no less than 0.75, the algorithm took the whole patch as candidate epitopes.

Results and discussion

Evaluation parameters

$$Sen(sensitivity\ or\ true\ positive\ rate) = \frac{TP}{TP+FN} \quad (1)$$

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

$$PPV(positive\ predictive\ value) = \frac{TP}{TP+FP} \quad (3)$$

$$ACC(accuracy) = \frac{TP+TN}{TP+TN+FP+FN} \quad (4)$$

$$F(F\text{-}measure) = \frac{2*PPV*Sen}{PPV+Sen} \quad (5)$$

In the above equations, TP was the number of predicted epitope residues proven to be the true epitope residues. FP was the number of predicted epitope residues proven not to be the true epitope residues. TN was the predicted non-epitope residues proven not to be the true epitope residues. FN was the number of predicted non-epitope residues proven to be the true epitope residues. Sensitivity and specificity were two commonly used statistical measures of the performance of binary classification tests. ACC and F-measure were two synthetic measures. We took these measures and tried to give a complete and fair evaluation of the prediction performance of this work.

Prediction results on the testing dataset

Table 2 shows the results on the testing dataset of this method. Sensitivity, specificity, positive predictive value (PPV), ACC, F-measure and average value (AVG) are listed.

In the case of 3DOW_B, the number of antigen surface residues was seven, and the entire antigen surface residue was epitopes; for 3DAB_F, the number of antigen surface residues was 12, and the number of epitope residues was nine. For these two cases in the dataset, the number of epitopes to the number of antigen surface residues is greater than 0.75, then the algorithm predicted the whole residues in the highest scoring patch as candidate epitopes. As seen in **Table 2**, the average predicted performance of our method on the testing dataset achieved a sensitivity of 0.59; specificity of 0.64; precision of 0.45 and ACC of 0.70.

Effects of algorithm parameters

The work calculated the ASEP index for every 20 amino acids, scored each antigen surface patch through AAPs, got the highest scoring patch, ranked the amino acids in the patch by ASEP index and took the top 75% as the candidate epitopes. There were three parameters in this work: the first was the parameter N which was used to divide the antigen surface patch; the second was the parameter D which was used to define patch-AAPs (pAAPs) in the antigen surface patch; the last was the parameter M which was used to determine the epitope residues in the highest scoring patch.

For the parameter D, we assigned the value 9 Å in the same way as the Mapitope did. For the parameter N, we first calculated the relationship between the size of the patch and the proportion of epitope residues on the patch using the background dataset and then we tested the new method with a different N using the testing dataset. The results are shown in **Figure 2** and **Table 3**.

Table 4. The average performances on the testing dataset with different M

Percent	Sensitivity	Specificity	PPV	ACC	F-measure
5%	0.03	0.93	0.23	0.69	0.05
10%	0.05	0.91	0.28	0.69	0.08
15%	0.11	0.89	0.33	0.69	0.19
20%	0.15	0.88	0.32	0.68	0.25
25%	0.19	0.85	0.35	0.68	0.29
30%	0.21	0.83	0.38	0.68	0.31
35%	0.24	0.80	0.38	0.68	0.33
40%	0.28	0.79	0.41	0.68	0.36
45%	0.30	0.77	0.41	0.68	0.38
50%	0.32	0.74	0.41	0.68	0.40
55%	0.35	0.73	0.42	0.68	0.42
60%	0.39	0.71	0.42	0.68	0.44
65%	0.41	0.70	0.43	0.68	0.45
70%	0.49	0.63	0.45	0.68	0.46
75%	0.59	0.64	0.45	0.70	0.49
80%	0.48	0.64	0.45	0.69	0.49
85%	0.50	0.63	0.45	0.68	0.51
90%	0.54	0.60	0.46	0.68	0.51
95%	0.56	0.56	0.46	0.68	0.51
100%	0.59	0.52	0.46	0.67	0.51

Table 5. The predictive performance of our method, ASEP and Mapitope

	Sensitivity	Specificity	PPV	ACC	F-measure
ASEP	0.54	0.43	0.33	0.46	0.35
Mapitope	0.32	0.5	0.33	0.65	0.37
Our method	0.59	0.64	0.45	0.7	0.49

We took Cα of every antigen surface residue as the center, and calculated the nearest N surface residues to form a patch, while the assignment of N was from 0 to 400. From **Figure 2**, it can be seen that when the number of the patch was 17, the proportion of epitopes in the patch reached 61.51%; when the number of the patch was 30, the proportion of epitopes in the patch reached 80.02%. Results showed that when N equaled 30, the proportion of epitopes was statistically significant (Mann-Whitney test, $P< 0.05$).

Conversely, we tested the method with a different N using the testing dataset. We also took the Cα of every antigen surface residue as the center, and took N from 20 to 35.

To make a fair comparison, we changed the parameter M from 5% to 100% for each N, and chose the best result as the predicting result of parameter N. That is to say, **Table 3** shows the testing result of different N with different M. As displayed in **Table 3**, the average sensitivity, precision and F-measure were all highest when N equaled 30. The values for specificity and the ACC were not the highest, but the sum of sensitivity and specificity was the highest. The whole performance was best when N equaled 30, so 30 was assigned to N in our method.

Then, for the parameter M, we tested the M from 5% to 100%, and we took 5% as the step size. The average performances of the new method on the testing dataset are shown in **Table 4**.

Table 4 shows the average predicted performances on the testing dataset with different M. The sensitivity is higher with the increased M, while the specificity becomes lower with the increased M. When M equaled 75%, the sum of sensitivity and specificity reached the highest value. For ACC, the highest value was 0.70 with M equaling 75%. The precision and F-measure was comparatively higher when M was assigned 75%. Since the sum of sensitivity and specificity and the ACC were relatively comprehensive measures, according to this analysis, 75% was assigned to M in our method.

Comparison with other methods

To validate the effectiveness of the methods in this work, we compared the predictive performance of our method to others. For the ASEP-based method, we took the ASEP index as the only feature for prediction. We calculated the ASEP index for every antigen surface amino acid and ranked them by ASEP index, and then we took the top 75% (75% was obtained as the parameter M in this work) as the predictive epitopes. For the AAP-based method, we chose to use Mapitope. **Table 5** shows the comparison results.

For the average performance, the results show that our new method is effective in epitope prediction. Especially for PPV and F-measure, the values of our method were much higher than both ASEP and Mapitope. The results showed that our new method was effective in epitope prediction.

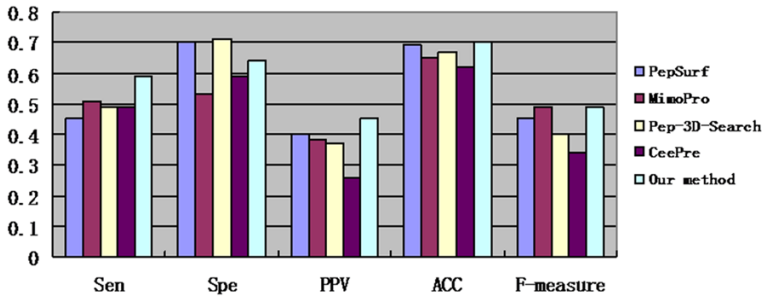

Figure 3. The predictive performance of our method and four other methods.

To further describe the performance of our new method, we chose three graph-based methods and one structure-based method for comparison. Graph-based methods are typical in the field of epitope prediction. The main idea of the method is to model the amino acids from an antigen as a graph structure and then use graph search methods to locate potential epitopes. PepSurf, Pep-3D-Search and MimoPro are similar methods. PepSurf searched the best-matched paths from the graph built from the antigen with mimotope sequences using a color-coding and a dynamic programming algorithm. Pep-3D-Search searched for the matched paths on the antigen surface with an ant colony optimization (ACO) algorithm. Candidate epitopes were then formed by clustering the resulting paths with a high *p*-value score by the Depth-First Search algorithm. MimoPro divided the antigen surface into overlapping patches by an adaptable distance threshold (ADT) regulated by compactness factor (CF), a novel parameter proposed in this method. Then on each single patch, the search was conducted to guarantee the best alignment for each mimotope sequence. Dynamic programming and branch-bound methods were also adopted for both, avoiding repetition in searches and further narrowing the search space. CeePre is a structure-based prediction method of conformational B-cell epitopes. It uses a new feature B factor (obtained from X-ray crystallography), combined with other basic physicochemical, statistical, evolutionary and structural features of each residue. **Figure 3** shows the predictive performance of our method and four other methods.

Figure 3 shows the overall performance of the four methods and our method on testing datasets. From the figure, it can be seen that the Sen, PPV, ACC and F-measure of our new method is the highest among the five methods. However the specificity is lower than for the PepSurf and Pep-3D-Search. For 3DOW_B, 1YCR_B and 3DAB_F, the number of antigen surface residues is a lot fewer, and the number of epitopes to the number of the three antigens is no less than 0.75. The algorithm then predicted the whole residues in the highest score patch as the candidate epitopes, this step had a high sensitivity but a specificity of 0. This is the main reason for the lower specificity of our new method. If we do not count these three cases, specificity will reach 0.71. So specificity of our method will be the highest of the five methods. However neither sensitivity nor specificity could evaluate the performance of a method individually. According to the prediction results, the sum of sensitivity and specificity of our new method is the highest when compared with the other four methods. This indicates that our new method is effective in conformational B-cell epitope prediction.

Conclusions

B-cell epitope prediction is important for vaccine design, development of diagnostic reagents and interpretation of the antigen-antibody interactions on a molecular level. Localizing epitopes by experimental methods is expensive in terms of time, cost and effort; therefore computational methods featured for low cost and high speed are employed to predict B-cell epitopes. In this work, we propose a new conformational B-cell epitope prediction method of binding with individual antibodies using phage display peptides. With this method, we first calculated the patch that potentially includes the most epitopes by dividing the mimotopes and antigen surface patch into AAPs, and then predicted the candidate epitopes using the ASEP index. This method is more effective than the single AAP and ASEP methods. The performance of the new method was measured with an average sensitivity of 59%, a specificity of 64%, a precision of 45%, an accuracy of 70%, and an F-measure of 0.49. When compared with other existing methods on the testing dataset, our method provides better predictive performance in sensitivity, PPV, accuracy and F-measure.

Our research is a new attempt to combine the idea of the AAP-based method and the ASEP index. It is more meaningful when it is used to predict the conformational B-cell epitope binding with individual antibodies. As seen from the results, for extremely difficult cases where amino acids forming the epitope include both consecutive segments and isolated amino acids, such as 3BT1 and 2HYM, the method failed in producing useful mappings. This indicates that we should make the prediction more efficient in future work.

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).

Disclosure of conflict of interest

The authors declare no conflict of interest.

Authors' contribution

CHZ conceived the idea and designed all the work. BJS optimized the algorithm and participated in the development. WNT designed experiments, gathered test data, and was in charge of the experiments. ZQM suggested extensions and modifications to the research. PPS supervised the whole research and revised the manuscript critically. All authors have read and approved the final manuscript.

Address correspondence to: Zhiqiang Ma, Pingping Sun, School of Computer Science and Information Technology, Northeast Normal University, Changchun 130117, Jilin, China. Tel: +86 431 8453 6338; Fax: +86 431 8453 6338; E-mail: zhiqiangmadoc@163.com (ZQM); sunpp567@nenu.edu.cn (PPS)

References

[1] Van Regenmortel MH. What is a B-cell epitope? Methods Mol Biol 2009; 524: 3-20.

[2] Abbas AK, Lichtman AH and Pillai S. Cellular and Molecular Immunology. Saunders Elsevier, 6th edition. W.B. Saunders Company; 2009.

[3] Denisova GF, Denisov DA and Bramson JL. Applying bioinformatics for antibody epitope prediction using affinity-selected mimotopes-relevance for vaccine design. Immunome Res 2010; 6 Suppl 2: S6.

[4] Flower DR. Towards in silico prediction of immunogenic epitopes. Trends Immunol 2003; 24: 667-674.

[5] Greenbaum JA, Andersen PH, Blythe M, Bui HH, Cachau RE, Crowe J, Davies M, Kolaskar AS, Lund O, Morrison S, Mumey B, Ofran Y, Pellequer JL, Pinilla C, Ponomarenko JV, Raghava GP, van Regenmortel MH, Roggen EL, Sette A, Schlessinger A, Sollner J, Zand M and Peters B. Towards a consensus on datasets and evaluation metrics for developing B-cell epitope prediction tools. J Mol Recognit 2007; 20: 75-82.

[6] Goldsby Richard, Kindt TJ, Osborne BA and Janis Kuby. Antigens. Immunology. 5th edition. New York: W. H. Freeman and Company; 2003.

[7] Barlow DJ, Edwards MS and Thornton JM. Continuous and discontinuous protein antigenic determinants. Nature 1986; 322: 747-748.

[8] Korber B, LaBute M and Yusim K. Immunoinformatics comes of age. PLoS Comput Biol 2006; 2: e71.

[9] Denisova GF, Denisov DA, Yeung J, Loeb MB, Diamond MS and Bramson JL. A novel computer algorithm improves antibody epitope prediction using affinity-selected mimotopes: a case study using monoclonal antibodies against the West Nile virus E protein. Mol Immunol 2008; 46: 125-134.

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

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

[12] Ponomarenko J, Bui HH, Li W, Fusseder N, Bourne PE, Sette A and Peters B. ElliPro: a new structure-based tool for the prediction of antibody epitopes. BMC Bioinformatics 2008; 9: 514.

[13] Liang S, Zheng D, Zhang C and Zacharias M. Prediction of antigenic epitopes on protein surfaces by consensus scoring. BMC Bioinformatics 2009; 10: 302.

[14] Ren J, Liu Q, Ellis J and Li J. Tertiary structure-based prediction of conformational B-cell epitopes through B factors. BMC Bioinformatics 2014; 30: i264-i273.

[15] Batori V, Friis EP, Nielsen H and Roggen EL. An in silico method using an epitope motif database for predicting the location of antigenic determinants on proteins in a structural context. J Mol Recognit 2006; 19: 21-29.

[16] Smith GP and Petrenko VA. Phage Display. Chem Rev 1997; 97: 391-410.
[17] Geysen HM, Rodda SJ and Mason TJ. A priori delineation of a peptide which mimics a discontinuous antigenic determinant. Mol Immunol 1986; 23: 709-715.
[18] Moreau V, Granier C, Villard S, Laune D and Molina F. Discontinuous epitope prediction based on mimotope analysis. Bioinformatics 2006; 22: 1088-1095.
[19] Pizzi E, Cortese R and Tramontano A. Mapping epitopes on protein surfaces. Biopolymers 1995; 36: 675-680.
[20] CastrignanòT, De Meo PD, Carrabino D, Orsini M, Floris M and Tramontano A. The MEPS server for identifying protein conformational epitopes. BMC Bioinformatics 2007; 8: S6.
[21] Mumey BM, Bailey BW, Kirkpatrick B, Jesaitis AJ, Angel T and Dratz EA. A New Method for Mapping Discontinuous Antibody Epitopes to Reveal Structural Features of Proteins. J Comput Biol 2003; 10: 555-567.
[22] Huang J, Gutteridge A, Honda W and Kanehisa M. MIMOX: a web tool for phage display based epitope mapping. BMC Bioinformatics 2006; 7: 451.
[23] Mayrose I, Shlomi T, Rubinstein ND, Gershoni JM, Ruppin E, Sharan R and Pupko T. Epitope mapping using combinatorial phage-display libraries: a graph-based algorithm. Nucleic Acids Res 2007; 35: 69-78.
[24] Huang YX, Bao YL, Guo SY, Wang Y, Zhou CG and Li YX. Pep-3D-Search: a method for B-cell epitope prediction based on mimotope analysis. BMC Bioinformatics 2008; 9: 538.
[25] Negi SS and Braun W. Automated detection of conformational epitopes using phage display peptide sequences. Bioinform Biol Insights 2009; 3: 71-81.
[26] Chen WH, Sun PP, Lu Y, Guo WW, Huang YX and Ma ZQ. MimoPro: a more efficient web-based tool for epitope prediction using phage display libraries. BMC Bioinformatics 2011; 12: 199.
[27] Sun P, Chen W, Huang Y, Wang H, Ma Z and Lv Y. Epitope Prediction Based on Random Peptide Library Screening: Benchmark Dataset and Prediction Tools Evaluation. Molecules 2011; 16: 4971-4993.
[28] Soga S, Kuroda D, Shirai H, Kobori M and Hirayama N. Use of amino acid composition to predict epitope residues of individual antibodies. Protein Eng Des Sel 2010; 23: 441-448.
[29] Enshell-Seijffers D, Denisov D, Groisman B, Smelyanski L, Meyuhas R, Gross G, Denisova G and Gershoni JM. The mapping and reconstitution of a conformational discontinuous B-cell epitope of HIV-1. J Mol Biol 2003; 334: 87-101.
[30] Mintseris J, Wiehe K, Pierce B, Anderson R, Chen R, Janin J and Weng Z. Protein-Protein Docking Benchmark 2.0: an update. Proteins 2005; 60: 214-216.
[31] Ponomarenko JV and Bourne PE. Antibody-protein interactions: benchmark datasets and prediction tools evaluation. BMC Struct Biol 2007; 7: 64.
[32] Sweredoski MJ and Baldi P. PEPITO: improved discontinuous B-cell epitope prediction using multiple distance thresholds and half sphere exposure. Bioinformatics 2008; 24: 1459-1460.
[33] Rubinstein ND, Mayrose I, Martz E and Pupko T. Epitopia: a web-server for predicting B-cell epitopes. BMC Bioinformatics 2009; 10: 287.
[34] Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN and Bourne PE. The Protein Data Bank. Nucleic Acids Res 2000; 28: 235-242.
[35] Bublil EM, Freund NT, Mayrose I, Penn O, Roitburd-Berman A, Rubinstein ND, Pupko T and Gershoni JM. Stepwise prediction of conformational discontinuous B-cell epitopes using the mapitope algorithm. Proteins 2007; 68: 294-304.
[36] Tsodikov OV, Record MT Jr and Sergeev YV. Novel computer program for fast exact calculation of accessible and molecular surface areas and average surface curvature. J Comput Chem 2002; 23: 600-609.
[37] Ahmad S, Gromiha M, Fawareh H and Sarai A. ASAView: Database and tool for solvent accessibility representation in proteins. BMC Bioinformatics 2004; 5: 51.