PREDICTING B CELL EPITOPE RESIDUES WITH NETWORK TOPOLOGY BASED AMINO ACID INDICES

JIAN HUANG[1, 2] hjian@kuicr.kyoto-u.ac.jp

WATARU HONDA[1] honda@kuicr.kyoto-u.ac.jp

MINORU KANEHISA[1, 3] kanehisa@kuicr.kyoto-u.ac.jp

[1] *Bioinformatics Center, Institute for Chemical Research, Kyoto University, Gokasho Uji, Kyoto 611-0011, Japan*

[2] *School of Life Science and Technology, University of Electronic Science and Technology of China, Chengdu 610054, China*

[3] *Human Genome Center, Institute of Medical Science, University of Tokyo, 4-6-1 Shirokanedai, Minato-ku, Tokyo 108-8639, Japan*

We evaluate the performance of six amino acid indices in B cell epitope residue prediction using the classical sliding window method on five data sets. Four of the indices: *i.e.* relative connectivity, clustering coefficient, closeness and betweenness are newly derived from the topological parameters of residue networks. The other two are Parker's hydrophilicity and Levitt's index, known as the best indices so far for B cell epitope prediction. On four of the data sets, the performance of all the indices was comparable and poor in general. When applied to one well-annotated data set, the performances improved and the 4 network based indices showed better performance than that of Parker's hydrophilicity and Levitt's index. When using the relative connectivity index on this data set, the prediction accuracy, sensitivity and specificity reached 73.6%, 73.0% and 75.0% respectively, with an area under the curve about 0.796. Thus, we suggested that this index is a good choice for B cell epitope prediction. It also indicates that the low performance of B cell epitope prediction is not only due to the methods and amino acid indices used, but also the data set as well. Interestingly, on the well-annotated data set, the performance of B cell epitope residue prediction is very similar to that of protein surface residue prediction, especially at the 10 and 20 $Å^2$ cutoffs. It is suggested that the performance in surface residue prediction might form a theoretical upper limit for the performance of B cell epitope residue prediction methods.

Keywords: B cell, epitope prediction, amino acid index, network topology

1. Introduction

The B cell epitopes of proteins are special regions on proteins that can be recognized by the antigen binding sites of antibodies or B cell receptors. Identified B cell epitopes are very useful because they can further be developed into diagnostics [1], therapeutics and vaccines [2, 3]. Therefore, it's only natural that B cell epitope mapping has been a major field of immunology research.

As identifying B cell epitopes experimentally is time-consuming and expensive, techniques to predict B cell epitopes have been developed for almost 30 years [4-17]. Most of these techniques are sliding window based sequence profiling methods. In brief, a window slides from the N-terminal to C-terminal of the query protein sequence. The

mean propensity value of the residues in the window is then assigned to the residue in the middle according on the amino acid index [18-20] (also known as the propensity scale) used in the prediction. By combining such predictions with experimental verification, many successful cases have been reported. However, the performance of this kind of B cell epitope predictions has been disputed [21-23]. In a recent report, Blythe *et al* assessed 484 amino acid indices in the AAindex database [20] with sequence profiling methods. They found that even the best set of amino acid indices performed only marginally better than random [24], indicating that better methods or new amino acid indices are needed for B cell epitope prediction. A very recent study has confirmed that Parker's hydrophilicity (*Ph*) [8] and Levitt's index (*Li*) [12] are the best two indices so far for sequence profiling based B cell epitope prediction. However, even the performance of *Ph* and *Li* are unsatisfactory [25].

In a previous study, we built four new amino acid indices, termed relative connectivity (*Rk*), relative clustering coefficient (*Rc*), relative closeness (*Ro*) and relative betweenness (*Rb*), based on residue networks constructed from 640 representative PDB structures [26]. Compared with *Ph* and *Li*, these network topology based indices have shown better performance in protein surface residue prediction [26]. Surface residue prediction is related to B cell epitope prediction, due to the requirement for epitopes to be surface accessible to interact with an antibody [27, 28]. Since the network topology based indices have significantly better performance than *Ph* and *Li* in protein surface residue prediction [26], will they perform better in B cell epitope residue prediction too?

To answer the above question, the performance of *Ph*, *Li* and the 4 residue network topology derived amino acid indices in B cell epitope residue prediction are evaluated and compared in this study.

2. Methods and Data Sets

2.1. *Data sets*

Five data sets of proteins with annotated B cell epitope residues are used in the current study. The first data set, originally composed by Pellequer *et al*, contains 14 protein sequences and 82 epitopes [15]. We took the recreated electronic form of this dataset from the Lund group [25]. The second and the third data sets, denoted as AntiJen data set and HIV data set respectively, are composed by the Lund group [25]. The fourth data set is the DiscoTope data set, which has 75 antigens with B cell epitope residue annotation [29].

We compiled a fifth data set from the CED database [30]. Taking the well-studied hen egg white lysozyme (HEL) as the model antigen, 19 epitopes were found and used as annotation. We call this data set the HEL data set. For the purpose of comparison, the solvent accessible area of hen egg white lysozyme is computed from its PDB structure (1HEL) with the NACCESS program [31] using default parameters. Surface residues are

assigned based on the solvent accessible area at different cutoff values of 1, 10, 20, 50 and 100 $Å^2$.

2.2. Amino acid indices

The 4 network topology based amino acid indices are taken from our pervious study [26]. The best two indices (Parker's hydrophilicity and Levitt's index) known for B cell epitope prediction are taken from references [12, 25]. They are listed in Table 1.

Table 1. Amino acid indices used in prediction.

	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y
Rk	1.05	1.17	0.88	0.85	1.07	0.99	0.99	1.11	0.88	1.07	1.04	0.93	0.92	0.93	0.94	0.96	0.99	1.12	1.05	1.05
Rc	0.99	0.89	1.11	1.13	0.92	1.08	1.00	0.89	1.10	0.92	0.95	1.07	1.01	1.06	1.04	1.05	1.01	0.90	0.93	0.94
Ro	1.00	1.13	0.95	0.95	1.03	0.99	1.01	1.04	0.96	1.02	1.02	0.96	0.96	0.97	0.98	0.98	0.99	1.04	1.01	1.02
Ro	0.96	1.60	0.63	0.61	1.31	0.77	1.03	1.43	0.61	1.30	1.24	0.72	0.83	0.73	0.82	0.80	0.90	1.35	1.20	1.16
Ph	0.03	0.11	2.46	1.86	-2.78	1.28	0.30	-2.45	1.26	-2.87	-1.41	1.64	0.30	1.37	0.87	1.50	1.15	-1.27	-3.00	-0.78
Li	-0.56	-0.44	1.43	0.11	-1.13	2.15	-0.85	-1.38	0.02	-1.16	-1.69	1.02	3.00	0.08	-0.22	1.15	0.27	-1.50	-0.60	0.30

2.3. Sequence profiling

Sequence profiling is completed with the classical sliding window method. Briefly, a window slides from the N-terminal to C-terminal of the query protein sequence. The mean propensity value of the residues in the window is then assigned to the residue in the middle. At the N- and C- termini, we use asymmetric windows to avoid omitting prediction examples. Different window sizes of 1, 3, 5, 7, 9 and 11 are tested. If an index correlates negatively to the B cell epitope residues, it is then multiplied by -1 when used in prediction; this process makes the index have a positive predictive power.

2.4. *Performance measures*

Receiver operating characteristics (ROC) curves are constructed by varying the prediction threshold and plot the false-positive proportion (1-specificity) on the x-axis against the true positive proportion (sensitivity) on the y-axis [32]. The area under the ROC curve (*Aroc*) is used as the performance measure. For a random prediction, *Aroc* equals 0.5; for a perfect method, *Aroc* equals 1. Empirically, a prediction with an *Aroc* between 0.9 and 1 would be considered as "excellent"; 0.8-0.9, "good"; 0.7-0.8, "fair"; 0.6-0.7, "poor"; 0.5-0.6, "fail". More practically, an *Aroc* value higher than 0.7 indicates a useful prediction performance [33]. In this study, ROC curves and related performance measures are constructed, visualized and calculated with the ROCR package [34].

2.5. *Programming and Statistics*

All analyses are implemented in Perl or R; the latter is a language and environment for statistical computing and graphics. For predictions of special interest, they are further bootstrapped 1000 times. Random predictions are simulated through permuting the prediction results 1000 times. The *Aroc* differences of different indices are evaluated with *t*-tests.

3. Results

3.1. Performances of indices on different data sets

For each index and for each data set, sliding window sizes of 1, 3, 5, 7, 9 and 11 were used for B cell epitope residue prediction. The performance of the indices on each data set is shown in Table 2. The results for the window size at which the majority of indices reach maximum performance are shown. For example, "Pellequer(7)" means the *Aroc* values are from testing the Pellequer dataset using a sliding window size of 7. As shown in Table 2, all indices performed poorly on the Pellequer, AntiJen, HIV, and DiscoTope data sets. However, their performance improved significantly when applied to the well-annotated HEL data set.

Table 2. Index performance on 5 data sets.

	Pellequer(7)	AntiJen(11)	HIV (11)	DiscoTope(9)	HEL (1)
Rk	0.627	0.561	0.588	0.591	0.794
Rc	0.637	0.564	0.586	0.608	0.752
Ro	0.609	0.566	0.575	0.583	0.787
Rb	0.633	0.567	0.589	0.610	0.772
Ph	0.655	0.565	0.586	0.622	0.733
Li	0.620	0.572	0.567	0.613	0.711

3.2. ROC curves and statistical analysis

According to Table 2, each index showed its best performance on the HEL data set at a sliding window size of 1. The prediction using the relative connectivity, Parker's hydrophilicity and Levitt's index were further bootstrapped 1000 times. Random predictions were simulated by permuting the prediction results 1000 times (see Table 3). Statistical tests showed that the relative connectivity index performed significantly better than the Parker's hydrophilicity and Levitt's index and the performance of all 3 were significantly better than random ($P < 2.2 \times 10^{-16}$). Their ROC curves were shown in Figure 1. On the HEL data set at a sliding window size of 1, the prediction accuracy, sensitivity and specificity of relative connectivity reached 73.6%, 73.0% and 75.0% respectively, with an area under the curve of 0.796.

Table 3. Comparing *Rk* performance with *Ph*, *Li* and Random on the HEL data set.

	Mean *Aroc*±Standard Error
Rk	0.796±0.001
Ph	0.733±0.001
Li	0.708±0.002
Random	0.497±0.002

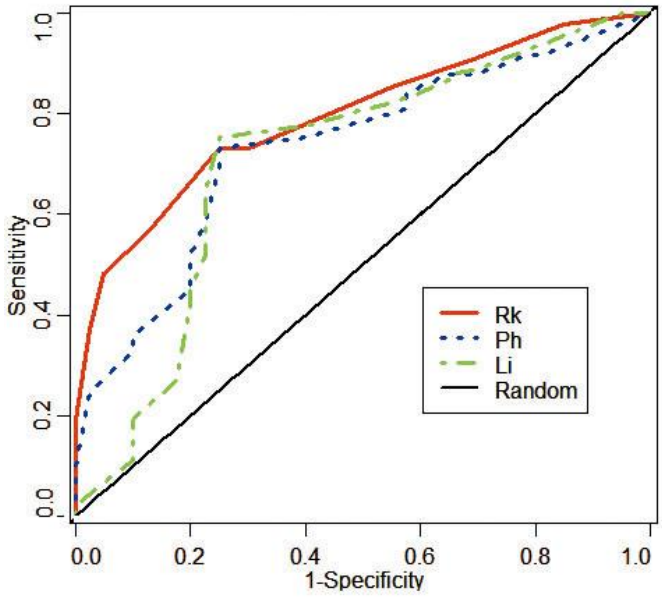

Figure 1: ROC curves for *Rk*, *Ph*, *Li* and Random. Based on the HEL data set at a window size of 1.

3.3. Similar performance between surface and B cell epitope residue prediction

The solvent accessible area of hen egg white lysozyme is computed from its PDB structure (1HEL) with the NACCESS program using default parameters. Surface residues are assigned based on the solvent accessible area at different cutoff values of 1, 10, 20, 50 and 100 $Å^2$. Prediction of surface residues was carried out using a sliding window size of 1. *Aroc* values are shown in Table 4, where "Bepi" means B cell epitope residue prediction and "Surf1" means surface residue prediction at the 1 $Å^2$ cutoff and so on. As shown in Table 4, the performance of B cell epitope residue prediction is very similar to that of protein surface residue prediction, especially at the 10 and 20 $Å^2$ cutoffs.

Table 4. Performance similarity between surface and B cell epitope residue prediction.

	Bepi	Surf1	**Surf10**	**Surf20**	Surf50	Surf100
Rk	**0.794**	0.758	**0.794**	**0.791**	0.778	0.731
Rc	**0.752**	0.727	**0.773**	**0.767**	0.732	0.594
Ro	**0.787**	0.753	**0.787**	**0.784**	0.766	0.707
Rb	**0.772**	0.744	**0.786**	**0.777**	0.744	0.624
Ph	**0.733**	0.713	**0.709**	**0.734**	0.673	0.574
Li	**0.711**	0.722	**0.711**	**0.729**	0.666	0.527

4. Discussions

4.1. *Relative connectivity can be a good choice for B cell epitope prediction*

The most common method for B cell epitope prediction is the sliding window and amino acid index based sequence profiling method [4-17]. However, its performance has been disputed [21-23]. In a recent report, 484 amino acid indices in the AAindex database were assessed with sequence profiling methods and the results showed that even the best set of amino acid indices performed only marginally better than random [24]. This indicates that better methods or new amino acid indices are needed for B cell epitope prediction. New methods such as neural networks, hidden Markov models and support vector machines have been applied to B cell epitope prediction very recently [25, 35-37]. However, the performance improvements are still limited.

In a previous study, we built four new amino acid indices based on the topological parameters of residue networks constructed from 640 representative PDB structures [26]. The Parker's hydrophilicity index and the Levitt's index have been confirmed to be the best two indices so far for B cell epitope prediction [12, 25]. Compared with the two indices, the 4 network topology based indices showed better performance in protein surface residue prediction [26]. Since surface accessibility implies antibody accessibility [27, 28], we wondered if the new indices would also show better performance in B cell epitope residue prediction.

Indeed, the results of this study show that the network based indices, especially relative connectivity, perform better than the Parker's hydrophilicity and Levitt's index on the well-annotated HEL data set. For other data sets, the performances of all indices are comparable. On the HEL data set at the sliding window size of 1, the relative connectivity performed significantly better than the Parker's hydrophilicity and Levitt's index and the performance of all 3 was significantly better than random prediction. The prediction accuracy, sensitivity and specificity of relative connectivity reached 73.6%, 73.0% and 75.0% respectively, with an area under the curve of 0.796. In fact, even on a poor performing data set (*e.g.* AntiJen data set), all the predictions were still significantly

better than random (data not shown). We concluded that the network topology based indices, especially the relative connectivity, are useful indices for B cell epitope residue prediction.

4.2. *Performance depends on the data sets*

A trend that better performance depends on better-annotated data set was observed. According to Table 2, all indices performed poorly on the Pellequer, AntiJen, HIV, and DiscoTope data sets. However, their performance improved significantly when applied to the HEL data set. The HEL data set is based on the well-studied model antigen hen egg white lysozyme. The protein was densely annotated with 19 epitopes, most of them derived from high quality structures of antigen-antibody complexes. The Discotope data set is also derived from crystals of antigen-antibody complexes. However, in this data set, each antigen sequence is only annotated with one epitope. Most data in the left 3 data sets is annotated with information from overlapping peptide experiments, which might have errors because a peptide can bind an antibody even if some residues of the peptide are not interacting with the antibody [29]. Even so, the trend in these 3 data sets is also obvious. The performance on the fully annotated Pellequer data set is better than the less annotated AntiJen and HIV data sets. One can also expect that some false positive predictions are actually undiscovered B cell epitope residues.

Thus, we concluded that the low performance of B cell epitope prediction is not only due to the methodology used, but the data set as well. The limited performance improvements observed with new methods might also be due to the data set itself. The importance of data set for B cell epitope prediction has also been addressed in a workshop very recently [38]. Another interesting phenomenon was also observed, that is the better the data set is annotated, the smaller the optimum sliding window size is (see Table 2 and data not shown).

4.3. *The relationship between surface residues and B cell epitope prediction*

The problem of surface residue prediction is related to that of B cell epitope prediction, due to the requirement for epitopes to be surface accessible to interact with an antibody [27, 28]. In fact, most amino acid index based B cell epitope prediction methods, if not all, utilize this correlation. The Parker's hydrophilicty index [8] and β turn scale [15] are two good examples. In our previous study, the 4 network topology based amino acid indices showed a useful performance in surface residue prediction and they also correlated with hydrophobicity (or hydrophilicty) and β propensity [26].

In the current study, we found that the performance of B cell epitope residue prediction is very similar to that of protein surface residue prediction on the well-annotated HEL data set, especially at the at the 10 and 20 $Å^2$ cutoffs (see Table 4). It is proposed that any part of the accessible surface of a globular protein antigen can be recognized by antibodies, and the entire exposed surface represents a "continuum" of overlapping potential epitopes [39]. Therefore, we suggest that the performance in

surface residue prediction might form a theoretical upper limit for the performance in B cell epitope residue prediction. As a B cell epitope is a context dependent immunological entity [38, 40, 41], a new paradigm for B cell epitope prediction may emerge, shifted from an "all B cell epitopes model" to a "single B cell epitope model", and from an "only antigen sequence based model" to a "multiple information based model." Besides the antigen sequence information, other information such as antigen structure [29], antibody sequence or mimotopes [42] are needed in the new generation of B cell epitope prediction methods.

Acknowledgments

We thank Dr Alex Gutteridge for copyediting the manuscript, giving valuable comments and helping us on R. This work was supported by grants from the Ministry of Education, Culture, Sports, Science and Technology and the Japan Science and Technology Agency. The computational resource was provided by the Bioinformatics Center, Institute for Chemical Research, Kyoto University. The support from NSFC project (30600138) is also gratefully acknowledged.

References

[1] Meloen, R.H., Puijk, W.C., Langeveld, J.P., Langedijk, J.P. and Timmerman, P., Design of synthetic peptides for diagnostics, *Curr Protein Pept Sci*, 4(4):253-260, 2003.

[2] Tanabe, S., Epitope peptides and immunotherapy, *Curr Protein Pept Sci*, 8(1):109-118, 2007.

[3] Naz, R.K. and Dabir, P., Peptide vaccines against cancer, infectious diseases, and conception, *Front Biosci*, 12:1833-1844, 2007.

[4] Kazim, A.L. and Atassi, M.Z., Prediction and conformation by synthesis of two antigenic sites in human haemoglobin by extrapolation from the known antigenic structure of sperm-whale myoglobin, *Biochem J*, 167(1):275-278, 1977.

[5] Hopp, T.P. and Woods, K.R., Prediction of protein antigenic determinants from amino acid sequences, *Proc Natl Acad Sci U S A*, 78(6):3824-3828, 1981.

[6] Hopp, T.P. and Woods, K.R., A computer program for predicting protein antigenic determinants, *Mol Immunol*, 20(4):483-489, 1983.

[7] Welling, G.W., Weijer, W.J., van der Zee, R. and Welling-Wester, S., Prediction of sequential antigenic regions in proteins, *FEBS Lett*, 188(2):215-218, 1985.

[8] Parker, J.M., Guo, D. and Hodges, R.S., New hydrophilicity scale derived from high-performance liquid chromatography peptide retention data: correlation of predicted surface residues with antigenicity and X-ray-derived accessible sites, *Biochemistry*, 25(19):5425-5432, 1986.

[9] Jameson, B.A. and Wolf, H., The antigenic index: a novel algorithm for predicting antigenic determinants, *Comput Appl Biosci*, 4(1):181-186, 1988.

[10] Hopp, T.P., Use of hydrophilicity plotting procedures to identify protein antigenic segments and other interaction sites, *Methods Enzymol*, 178:571-585, 1989.
[11] Menendez-Arias, L. and Rodriguez, R., A BASIC microcomputer program for prediction of B and T cell epitopes in proteins, *Comput Appl Biosci*, 6(2):101-105, 1990.
[12] Pellequer, J.L., Westhof, E. and Van Regenmortel, M.H., Predicting location of continuous epitopes in proteins from their primary structures, *Methods Enzymol*, 203:176-201, 1991.
[13] Maksyutov, A.Z. and Zagrebelnaya, E.S., ADEPT: a computer program for prediction of protein antigenic determinants, *Comput Appl Biosci*, 9(3):291-297, 1993.
[14] Pellequer, J.L. and Westhof, E., PREDITOP: a program for antigenicity prediction, *J Mol Graph*, 11(3):204-210, 191-202, 1993.
[15] Pellequer, J.L., Westhof, E. and Van Regenmortel, M.H., Correlation between the location of antigenic sites and the prediction of turns in proteins, *Immunol Lett*, 36(1):83-99, 1993.
[16] Alix, A.J., Predictive estimation of protein linear epitopes by using the program PEOPLE, *Vaccine*, 18(3-4):311-314, 1999.
[17] Odorico, M. and Pellequer, J.L., BEPITOPE: predicting the location of continuous epitopes and patterns in proteins, *J Mol Recognit*, 16(1):20-22, 2003.
[18] Nakai, K., Kidera, A. and Kanehisa, M., Cluster analysis of amino acid indices for prediction of protein structure and function, *Protein Eng*, 2(2):93-100, 1988.
[19] Tomii, K. and Kanehisa, M., Analysis of amino acid indices and mutation matrices for sequence comparison and structure prediction of proteins, *Protein Eng*, 9(1):27-36, 1996.
[20] Kawashima, S. and Kanehisa, M., AAindex: amino acid index database, *Nucleic Acids Res*, 28(1):374, 2000.
[21] Hopp, T.P., Retrospective: 12 years of antigenic determinant predictions, and more, *Pept Res*, 6(4):183-190, 1993.
[22] Van Regenmortel, M.H. and Pellequer, J.L., Predicting antigenic determinants in proteins: looking for unidimensional solutions to a three-dimensional problem?, *Pept Res*, 7(4):224-228, 1994.
[23] Hopp, T.P., Different views of protein antigenicity, *Pept Res*, 7(4):229-231, 1994.
[24] Blythe, M.J. and Flower, D.R., Benchmarking B cell epitope prediction: underperformance of existing methods, *Protein Sci*, 14(1):246-248, 2005.
[25] Larsen, J.E., Lund, O. and Nielsen, M., Improved method for predicting linear B-cell epitopes, *Immunome Res*, 2(2, 2006.
[26] Huang, J., Kawashima, S. and Kanehisa, M., New amino acid indices based on residue network topology, *Genome Informatics*, 18(1):in press, 2007.
[27] Novotny, J., Handschumacher, M., Haber, E., Bruccoleri, R.E., Carlson, W.B., Fanning, D.W., Smith, J.A. and Rose, G.D., Antigenic determinants in proteins coincide with surface regions accessible to large probes (antibody domains), *Proc Natl Acad Sci U S A*, 83(2):226-230, 1986.
[28] Thornton, J.M., Edwards, M.S., Taylor, W.R. and Barlow, D.J., Location of 'continuous' antigenic determinants in the protruding regions of proteins, *Embo J*, 5(2):409-413, 1986.

[29] Haste Andersen, P., Nielsen, M. and Lund, O., Prediction of residues in discontinuous B-cell epitopes using protein 3D structures, *Protein Sci*, 15(11):2558-2567, 2006.
[30] Huang, J. and Honda, W., CED: a conformational epitope database, *BMC Immunol*, 7:7, 2006.
[31] Hubbard, S.J. and Thornton, J.M., NACCESS, *Department of Biochemistry and Molecular Biology, University College London,* 1993.
[32] Swets, J.A., Measuring the accuracy of diagnostic systems, *Science*, 240(4857):1285-1293, 1988.
[33] Lund, O., Nielsen, M., Lundegaard, C., Kesmir, C. and Brunak, S., Immunological Bioinformatics, *The MIT Press*, 2005.
[34] Sing, T., Sander, O., Beerenwinkel, N. and Lengauer, T., ROCR: visualizing classifier performance in R, *Bioinformatics*, 21(20):3940-3941, 2005.
[35] Saha, S. and Raghava, G.P., Prediction of continuous B-cell epitopes in an antigen using recurrent neural network, *Proteins*, 65(1):40-48, 2006.
[36] Sollner, J. and Mayer, B., Machine learning approaches for prediction of linear B-cell epitopes on proteins, *J Mol Recognit*, 19(3):200-208, 2006.
[37] Chen, J., Liu, H., Yang, J. and Chou, K.C., Prediction of linear B-cell epitopes using amino acid pair antigenicity scale, *Amino Acids*, 2007.
[38] Greenbaum, J.A., Andersen, P.H., Blythe, M., Bui, H.H., Cachau, R.E., Crowe, J., Davies, M., Kolaskar, A.S., Lund, O., Morrison, S., Mumey, B., Ofran, Y., Pellequer, J.L., Pinilla, C., Ponomarenko, J.V., Raghava, G.P., van Regenmortel, M.H., Roggen, E.L., 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*, 20(2):75-82, 2007.
[39] Benjamin, D.C., Berzofsky, J.A., East, I.J., Gurd, F.R., Hannum, C., Leach, S.J., Margoliash, E., Michael, J.G., Miller, A., Prager, E.M. and et al., The antigenic structure of proteins: a reappraisal, *Annu Rev Immunol*, 2:67-101, 1984.
[40] Van Regenmortel, M.H.V., Mapping Epitope Structure and Activity: From One-Dimensional Prediction to Four-Dimensional Description of Antigenic Specificity, *Methods*, 9(3):465-472, 1996.
[41] Van Regenmortel, M.H., Immunoinformatics may lead to a reappraisal of the nature of B cell epitopes and of the feasibility of synthetic peptide vaccines, *J Mol Recognit*, 19(3):183-187, 2006.
[42] Huang, J., Gutteridge, A., Honda, W. and Kanehisa, M., MIMOX: a web tool for phage display based epitope mapping, *BMC Bioinformatics*, 7:451, 2006.