

B-cell conformational epitope prediction: current status and future direction

Jing Sun, Zhiwei Cao

As an essential step of adaptive immune response, the recognition between antigen and antibody triggers a series of self-protection mechanisms. Thus, the identification of antibody-binding sites (B-cell epitope) on protein antigens is crucial for biomedical application.

Existing methods for B-cell epitope prediction can be divided into two groups: linear epitope prediction and conformational epitope prediction. More than 90% of B-cell epitopes are discontinuous in sequence[1], which makes conformational B-cell epitope prediction more meaningful. For instance, the first server CEP[2] was erected in 2005 by introducing 'accessibility of residues' based on the 3D nature of the antigen proteins. Subsequently, DiscoTope[1] prediction method was designed by combining propensity scale matrixes with the spatial proximity and surface exposure. BEpro[3] improved DiscoTope method by introducing spatial attribute of half sphere exposure. After that, ElliPro[4], a web-tool that implements Thornton's method and, together with a residue clustering algorithm, the MODELLER program and the Jmol viewer, allows the prediction and visualization of antibody epitopes in a given protein sequence or structure. Comparing with peers, better performance was achieved by SEPPA[5]. More conformational factors have been taken into consideration in its algorithm, such as topological features and residue-triangle units' occurrences. On the base of these studies, EPSVR and EPMeta[6] have been developed. The first one uses a Support Vector Regression (SVR) method to integrate six scoring terms. Furthermore, it has been combined with five existing epitope prediction servers to construct EPMeta.

However, existing prediction methods are still underperformance[7]. Can B-cell

epitopes be distinguished from other surface regions with specific features? And is it applicable by using these features to predict antigenic regions on protein antigens? How much will the antigenicity be affected with the variation of these features?

To answer these questions, a multi-perspective bioinformatics analysis was carried on with a comprehensive B-cell epitope dataset[8], which contains immunoglobulin complex structures collected from PDB[9]. These B-cell epitopes were described with features from different perspectives. According to the analysis results, B-cell epitopes were relatively constant both in the number of composing residues and the accessible surface area. Though composed of spatially clustering residues, there were sequentially linear segments exist in these epitopes. Besides, statistical differences were found between epitope and non-epitope surface residues with residual features, compared to non-epitope surface residues, epitope ones were more accessible. Amino acid enrichment and preference for specific types of residue-pair set on epitope areas have also been observed: more polar and aromatic residues were involved. Additionally, epitope residues tended to be less conservative under the environmental pressure. Measured by topological features, epitope residues were surrounded with fewer residues but in a more compact way. The formation of epitope regions tended to be protruding from the calculation results of protruding index and planarity index. The occurrences of residue-pair sets between epitope and paratope also showed some patterns.

Hence, it can be concluded that features that can be used to distinguish B-cell epitope from other surface regions exist, and how to use spatial features is a key factor in epitope prediction. Besides, B-cell epitope are highly context dependent and cannot exist without a corresponding antibody[10]. Future epitope prediction algorithm should also take this into consideration.

Keyword: B-cell epitope, conformational epitope prediction, bioinformatics

1. Haste Andersen, P., M. Nielsen, and O. Lund, *Prediction of residues in discontinuous B-cell epitopes using protein 3D structures*. Protein Sci, 2006. **15**(11): p. 2558-67.
2. Kulkarni-Kale, U., S. Bhosle, and A.S. Kolaskar, *CEP: a conformational epitope prediction server*. Nucleic Acids Res, 2005. **33**(Web Server issue): p. W168-71.
3. Sweredoski, M.J. and P. Baldi, *PEPITO: improved discontinuous B-cell epitope prediction using multiple distance thresholds and half sphere exposure*. Bioinformatics, 2008. **24**(12): p. 1459-60.
4. Ponomarenko, J., et al., *ElliPro: a new structure-based tool for the prediction of antibody epitopes*. BMC Bioinformatics, 2008. **9**: p. 514.
5. Sun, J., et al., *SEPPA: a computational server for spatial epitope prediction of protein antigens*. Nucleic Acids Res, 2009. **37**(Web Server issue): p. W612-6.
6. Liang, S., et al., *EPSVR and EPMeta: prediction of antigenic epitopes using support vector regression and multiple server results*. BMC Bioinformatics, 2010. **11**: p. 381.
7. Blythe, M.J. and D.R. Flower, *Benchmarking B cell epitope prediction: underperformance of existing methods*. Protein Sci, 2005. **14**(1): p. 246-8.
8. Sun, J., et al., *Does difference exist between epitope and non-epitope residues? Analysis of the physicochemical and structural properties on conformational epitopes from B-cell protein antigens*. Immunome Res, 2011. **7**(3): p. 1-11.
9. Berman, H.M., et al., *The Protein Data Bank*. Nucleic Acids Res, 2000. **28**(1): p. 235-42.
10. Reitmaier, R., *Review of immunoinformatic approaches to in-silico B-cell epitope prediction*. Nature Precedings, 2007.