FOR THE RECORD

Benchmarking B cell epitope prediction: Underperformance of existing methods

MARTIN J. BLYTHE AND DARREN R. FLOWER

Edward Jenner Institute for Vaccine Research, Compton, Berkshire RG20 7NN, United Kingdom

(RECEIVED August 20, 2004; FINAL REVISION September 27, 2004; ACCEPTED September 27, 2004)

Abstract

Sequence profiling is used routinely to predict the location of B-cell epitopes. In the postgenomic era, the need for reliable epitope prediction is clear. We assessed 484 amino acid propensity scales in combination with ranges of plotting parameters to examine exhaustively the correlation of peaks and epitope location within 50 proteins mapped for polyclonal responses. After examining more than 10^6 combinations, we found that even the best set of scales and parameters performed only marginally better than random. Our results confirm the null hypothesis: Single-scale amino acid propensity profiles cannot be used to predict epitope location reliably. The implication for studies using such methods is obvious.

Keywords: active site/binding site/epitope mapping; proteins of the immune system; immunological methods; epitope prediction

Supplemental material: see www.proteinscience.org

Historically, vaccines have been attenuated whole pathogens such as BCG or Sabin's Polio vaccine. Safety concerns have led to other strategies for vaccine development that focus on antigen and epitope vaccines. The latter use peptides able to elicit antibodies that cross-react with a parent protein (Langeveld et al. 1994). Peptides recognized by protein-specific antibodies have diagnostic applications (Orbach and Shoenfeld 2003). B-cell epitopes are classified as either linear or discontinuous epitopes. Linear epitopes comprise a single continuous stretch of amino acids within a protein sequence, while an epitope whose residues are distantly separated in the sequence and are brought into physical proximity by protein folding is called a discontinuous epitope. Although most epitopes are thought to be discontinuous (Barlow et al. 1986), experimental epitope detection has focused on linear epitopes. The accurate in silico prediction of linear epitopes from protein sequences is highly desirable and is thought to be achievable by using amino acid scales to create sequence profiles (Castelletti et al. 2004; Hua et al. 2004). Though the predictive accuracy of such methods has been examined previously (Hopp 1986; Van Regenmortel and Daney de Marcillac 1988; Pellequer et al. 1994), the extensive and objective analysis we present here is the most comprehensive to date. It focuses on a thorough dissection of the underlying hypothesis: Peaks in single amino acid–scale propensity profiles are significantly associated with known linear epitope locations. We exhaustively compared known epitope locations with propensity profile peaks generated using 484 amino acid scales combined with ranges of profile parameters, and we could find no evidence to support this hypothesis.

Materials and methods

Four hundred eighty-four amino acid scales were taken from the AAindex database (http://www.genome.ad.jp/aaindex) (Kawashima and Kanehisa 2000). Each scale assigns a value to each of the 20 normal amino acids. Scale values were normalized. We tested a set of 50 epitope-mapped

Reprint requests to: Martin J. Blythe, Edward Jenner Institute for Vaccine Research, High Street, Compton, Berkshire, RG20 7NN, UK; e-mail: martin.blythe@jenner.ac.uk; fax: +44-1635-577908.

Article and publication are at http://www.proteinscience.org/cgi/doi/10.1110/ps.041059505.

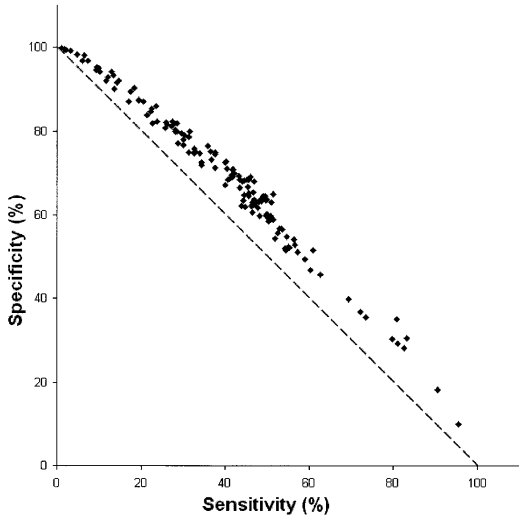

Figure 1. Specificity vs. sensitivity scores of the 150 optimized scales from the three amino acid sequence profiling methods. The broken line indicates random prediction.

proteins. Linear epitopes for each protein were extracted from the Antijen database (formally JenPep; http://www.jenner.ac.uk/AntiJen) (McSparron et al. 2003). Each experiment used polyclonal antibodies raised against the whole protein and resolved epitopes to one to three amino acids. We then verified the exact identity of published and test-set sequences. Mapping information from exclusively autoimmune studies was not included. Predicted and determined epitope and nonepitope residues were compared for all test-set protein residues using counts of true positive (TP), true negative (TN), false positive (FP), and false negative (FN). Prediction accuracy was characterized using four criteria (Baldi et al. 2000):

1. Specificity: $100 \times TN/(FP+TN)$
2. Sensitivity: $100 \times TP/(FN+TP)$
3. Matthews correlation coefficient (MCC): the overall accuracy of the prediction with a value between 1 and −1, where 0 represents random correlation (Baldi et al. 2000)
4. Mutual information coefficient (IC): a value between 0 (random correlation) and 1 (either complete agreement or disagreement between the observed and predicted) (Baldi et al. 2000).

Three scale implementation methods were used. For Method 1, each scale was implemented using the "sliding-window" method: a running average of amino acid properties for a defined number of residues along a protein sequence. The mean of the resulting profile was used to determine a cutoff value, enabling identification of predicted epitope residues. The 484 scales were initially tested using a reduced range of window sizes and cutoff values. Negative correlation was examined by inverting each scale. The 50 most accurate scales were then examined more fully using an increased range of variables including six other equations for determining the value for each window.

Method 2 was implemented as for Method 1, except that profiles resulting from different window sizes were examined collectively. Predicted epitopes were identified by selecting the highest scoring nonoverlapping windows after ranking individual window values. The top 50 scales were then examined using a wider range of variables. For Method 3, profile filtering was implemented in order to assess correlation between major profile peaks and epitope locations. Two filtering algorithms were applied independently to individual scales; both were based on an iterative reaveraging of sequence profile values. Major profile peaks were identified as a trend over 7 profile points. Scales were implemented in accordance with Method 1. The 50 most accurate scales identified in the initial round were then examined under a greater range of variables.

Results and Discussion

The literature indicates that amino acid propensity profiling is used routinely to predict or retrospectively identify epitope locations. Several papers have sought to quantify the relationship between the location of peaks and epitopes (Hopp 1986; Van Regenmortel and Daney de Marcillac 1988; Pellequer et al. 1994). The limited number of mapped proteins and the method of epitope identification used may have introduced bias into these results. Moreover, the scoring methods used were not sufficient to characterize fully their predictive performance. Our test set is significantly larger: We used 50 epitope-mapped proteins, as defined by polyclonal antibodies, which were mapped to a three–amino acid resolution. This represents the best impartial nonredundant test set currently available. Each propensity scale was initially examined using 17,100 combinations of algorithm parameters, with 228,900 combinations used subsequently for the 150 most accurate scales. The prediction accuracy calculations used here indicate clearly the performance of each scale.

MCC and IC scores suggest that for even the most accurate predictions the correlation between predicted and determined epitope residues is only marginally greater than random, as illustrated in Figure 1. This suggests that there is no significant correlation between the sequence profiles generated and the location of known linear epitopes. The hypothesis that sequence profiles generated with a single scale can be used to predict effectively linear epitopes from the primary sequence of proteins is thus not supported by the evidence. Given the complexity of antibody–antigen interaction and the relative simplicity of sequence profiling methods, this finding is not unsurprising, and the implications of our study are clear.

Though anecdotal evidence suggests that profiles can accurately predict linear epitopes in some instances (Castelletti et al. 2004; Hua et al. 2004), our results support the null hypothesis: Amino acid propensity profiles cannot be used effectively to predict linear epitopes. It is possible, although unlikely, that as-yet-undiscovered scales may perform adequately. Given the weak predictive performance of the profiling method demonstrated here, the development of more sophisticated approaches, perhaps using artificial intelligence techniques, is required to address this need and should be an obtainable goal.

Electronic supplemental material

The supplemental material contains a detailed description of the methods with additional references. Figure S1 demonstrates a sequence profile, and Tables S1 and S2 detail the five most accurate scales for each of the three methods.

Acknowledgments

We thank Paul Taylor and Pingping Guan for their technical advice.

References

Baldi, P., Brunak, S., Chauvin, Y., Andersen, C.A., and Nielsen, H. 2000. Assessing the accuracy of prediction algorithms for classification: An overview. *Bioinformatics* **16:** 412–424.

Barlow, D.J., Edwards, M.S., and Thornton, J.M. 1986. Continuous and discontinuous protein antigenic determinants. *Nature* **322:** 747–748.

Castelletti, D., Fracasso, G., Righetti, S., Tridente, G., Schnell, R., Engert, A., and Colombatti, M. 2004. A dominant linear B-cell epitope of ricin A-chain is the target of a neutralizing antibody response in Hodgkin's lymphoma patients treated with an anti-CD25 immunotoxin. *Clin. Exp. Immunol.* **136:** 365–372.

Hopp, T.P. 1986. Protein surface analysis. Methods for identifying antigenic determinants and other interaction sites. *J. Immunol. Methods* **88:** 1–18.

Hua, R., Zhou, Y., Wang, Y., Hua, Y., and Tong, G. 2004. Identification of two antigenic epitopes on SARS-CoV spike protein. *Biochem. Biophys. Res. Commun.* **319:** 929–935.

Kawashima, S. and Kanehisa, M. 2000. AAindex: Amino acid index database. *Nucleic Acids Res.* **28:** 374.

Langeveld, J.P., Casal, J.I., Osterhaus, A.D., Cortes, E., de Swart, R., Vela, C., Dalsgaard, K., Puijk, W.C., Schaaper, W.M., and Meloen, R.H. 1994. First peptide vaccine providing protection against viral infection in the target animal: Studies of canine parvovirus in dogs. *J. Virol.* **68:** 4506–4513.

McSparron, H., Blythe, M.J., Zygouri, C., Doytchinova, I.A., and Flower, D.R. 2003. JenPep: A novel computational information resource for immunobiology and vaccinology. *J. Chem. Inf. Comput. Sci.* **43:** 1276–1287.

Orbach, H. and Shoenfeld, Y. 2003. Anti-cyclic citrullinated peptide antibodies as a diagnostic test for rheumatoid arthritis. *Harefuah* **142:** 182–185, 239.

Pellequer, J.L., Westhof, E., and Van Regenmortel, M.H.V. 1994. Epitope predictions from the primary structure of proteins. In *Peptide antigens: A practical approach* (ed. G.B. Wisdom), pp. 7–25. Oxford University Press, Oxford.

Van Regenmortel, M.H. and Daney de Marcillac, G. 1988. An assessment of prediction methods for locating continuous epitopes in proteins. *Immunol. Lett.* **17:** 95–107.