

CEP: a conformational epitope prediction server

Urmila Kulkarni-Kale*, Shriram Bhosle and A. S. Kolaskar¹

Bioinformatics Centre and ¹Vice-Chancellor's office, University of Pune, Pune 411 007, India

Received February 14, 2005; Revised and Accepted April 13, 2005

ABSTRACT

CEP server (<http://bioinfo.ernet.in/cep.htm>) provides a web interface to the conformational epitope prediction algorithm developed in-house. The algorithm, apart from predicting conformational epitopes, also predicts antigenic determinants and sequential epitopes. The epitopes are predicted using 3D structure data of protein antigens, which can be visualized graphically. The algorithm employs structure-based Bioinformatics approach and solvent accessibility of amino acids in an explicit manner. Accuracy of the algorithm was found to be 75% when evaluated using X-ray crystal structures of Ag–Ab complexes available in the PDB. This is the first and the only method available for the prediction of conformational epitopes, which is an attempt to map probable antibody-binding sites of protein antigens.

INTRODUCTION

Antigen–antibody (Ag–Ab) complexes are non-obligatory heterocomplexes that are made and broken according to the environment and involve proteins that also exist independently. The most remarkable features of this special class of protein–protein interactions are high affinity and strict specificity of antibodies for their respective antigens. It is known that antibodies recognize the unique conformations and spatial locations on the surface of antigens. Therefore, epitopes are defined as the portions of the antigen molecules that interact with the antigen-binding site of antibody (paratope) to which they are complementary (1). The number of epitopes of every protein is equivalent to the number of monoclonal antibodies that can be generated against the protein. Delineation of epitopes for any protein antigen corresponds to the summation of the immune repertoire specific for the antigen in various hosts.

Epitopes are of two types, namely, sequential (when Ab binds to a contiguous stretch of amino acid residues that are linked by peptide bond) and conformational (when Ab binds to non-contiguous residues, brought together by folding of polypeptide chain). The specificity of sequential epitopes (SEs) is determined by the sequence and conformation of constituent

amino acids. However, specificity of conformational epitopes (CEs) depends on the spatial folding and the conformation of the contributing individual SEs (2).

Various algorithms have been developed to predict SEs given a protein sequence (3–7). Most of these algorithms use the propensity values of amino acid properties, such as hydrophilicity, antigenicity, segmental mobility, flexibility and accessibility to predict antigenicity. The accuracy of these algorithms lies in the range of 35–75%. However, no algorithm is available to predict the CE or antibody-binding sites of antigenic proteins.

It is known from the analyses of the crystal structures of Ag–Ab complexes that in order to be recognized by the antibodies, the residues must be accessible for interactions and thus be present on the surface of antigens. Therefore, an algorithm has been developed to predict SE and CE of the protein antigens with known 3D structure using accessibility in an explicit fashion (8,9) as against the algorithms mentioned above, a few of which use accessibility implicitly.

The predicted epitopes have applications in designing experiments for characterizing the antibody-binding sites of protein antigens. The method can be applied to engineer the 'designer binding sites' that mimic the interacting surface of the antigen, which is of immense use in the field of immunodiagnosics. Similarly, predicted epitopes can also be used to identify the candidate peptides for the development of peptide vaccines.

DESIGN AND IMPLEMENTATION

The CEP is implemented on Apache server with Linux 9.2 as an operating system. The web interface is designed using CGI Perl and JAVA scripts. The visualization package, Jmol, which is an open source software suite (<http://jmol.sourceforge.net/>), has been plugged in to facilitate visualization of the predicted conformational and sequential epitopes. Results are displayed in html format.

Algorithm

The algorithm predicts epitopes of protein antigens with known structures. It uses accessibility of residues and spatial distance cut-off to predict antigenic determinants (ADs),

*To whom correspondence should be addressed. Tel: +91 20 2569 0195/2569 2039; Fax: +91 20 2569 0087; Email: urmila@bioinfo.ernet.in

CEs and SEs [(8,9); U. Kulkarni-Kale and A. S. Kolaskar, unpublished data].

The steps are as follows:

- (i) Calculation of percentage accessibility of residues using an implementation of Voronoi polyhedron (10).
- (ii) Identification of antigenic residues with percentage accessible surface area ≥ 25 .
- (iii) Delineation of ADs, if at least three contiguous accessible residues are present.
- (iv) Extension of AD towards N'- and C'-termini by allowing the grace of one inaccessible residue.
- (v) Prediction of CE by collapsing ADs that are within the spatial proximity of 6 Å.
- (vi) Identification of SE that are not part of any CE.
- (vii) Inclusion of individual accessible residues that are part of CE and SE.
- (viii) Listing of AD, CE and SE.
- (ix) Define subsets for representing AD, CE and SE graphically.

Evaluation of algorithm

Accuracy of the algorithm has been critically evaluated using 21 Ag-Ab co-crystal structures from PDB (11). Evaluation dataset comprises two categories, namely, antigens characterized using multiple antibodies and those characterized using a

single antibody. Antibody-binding site (BS) for every Ag-Ab complex was defined as the sum of the residues that interact with antibody (IR) and those that are buried under antibody (BR). The list of IR was generated using the CONTACTSYM program (12) and the van der Waal's contact distances (13). The BRs were defined as those which loose at least 1 Å² accessible surface area upon the formation of complex with antibody. CE/antibody-binding sites of antigens were predicted for evaluation dataset. A prediction is said to be correct, if $\geq 60\%$ of the BS residues are part of a predicted CE. Of the 21 BS analysed, CEP algorithm correctly predicted 16, giving an overall accuracy of 75% [(9); U. Kulkarni-Kale and A. S. Kolaskar, unpublished data].

How to use the server?

A snapshot of CEP server is shown in Figure 1. User can predict CEs by either entering the PDB ID or uploading the coordinate file in PDB format. In the case of proteins with more than one chain, server prompts for selection of individual chains and/or oligomer. The server takes ~2 min to compute and display the results for a protein of size 250 residues.

Input data

CEP server requires the 3D coordinate data in PDB format. A sample input file is provided for reference. It is recommended

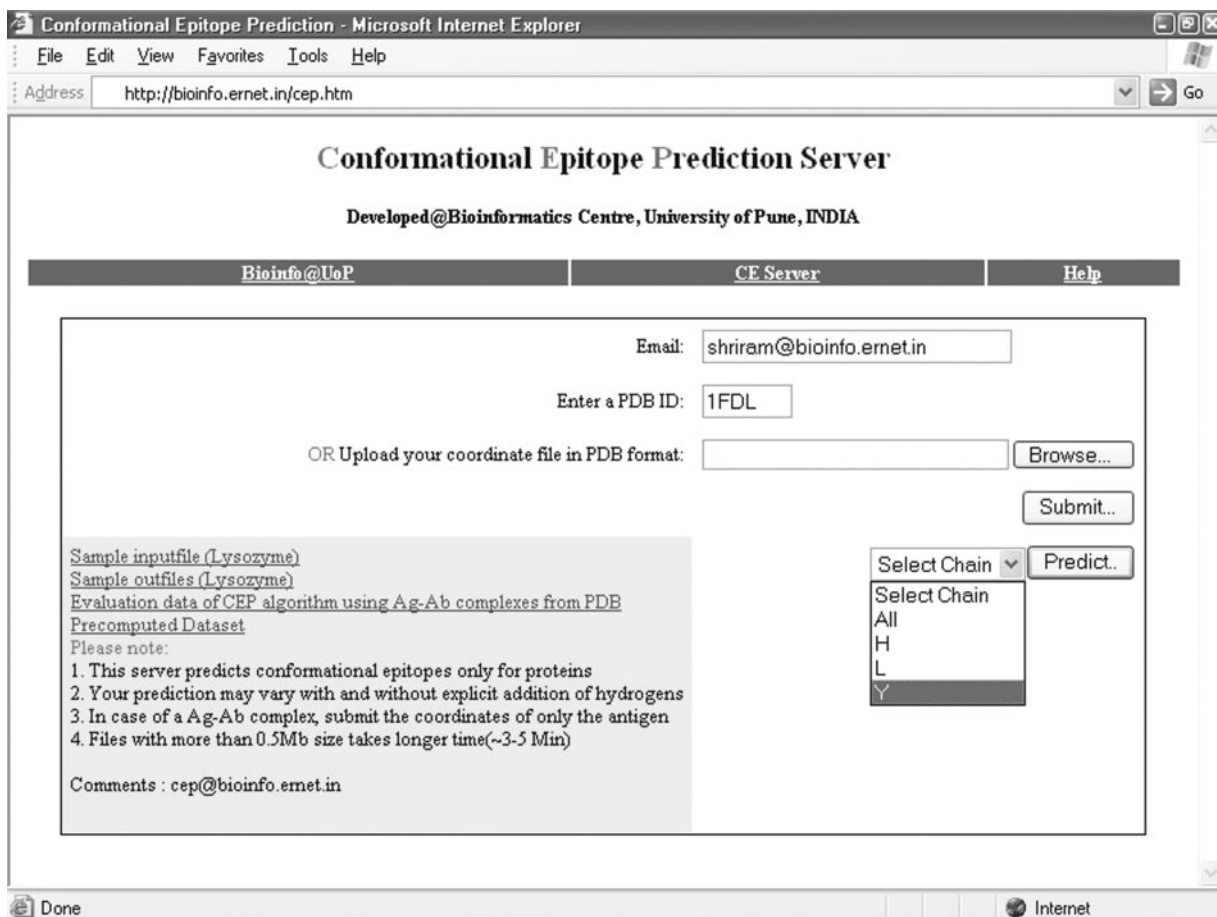


Figure 1. The CEP server home.

to submit structures for prediction with hydrogen atoms added; however, this is not mandatory.

Output data

Output is generated in html format. Predicted AD and CE are listed as separate tables. Predicted epitopes of lysozyme, PDB ID: 1FDL (14) are shown as Figure 2. As shown in Figure 2, AD number is followed by the chain ID and amino acid sequence of predicted AD along with the start and end positions. The residues that satisfy the percentage accessibility criterion are shown in uppercase, whereas those that do not satisfy this criterion are shown in lowercase. They are part of AD as they fulfil the criteria for extension (see Algorithm). The reference AD for every CE is shown in green and individual residues that are part of CE and SE are also listed (Figure 2). The predicted AD, CE and SE can be visualized by clicking the respective radio buttons. The experimentally characterized BS for the evaluation dataset can be visualized by clicking the BS radio button.

Precomputed data

In order to facilitate faster access, automated predictions have been carried out. Predictions for ~1800 proteins with resolution better than 1.5 Å (PDB release dated April 5, 2005) are available on the server and can be browsed using PDB ID.

DISCUSSION

A method to predict SEs and CEs is described above. The algorithm predicts epitopes using the 3D structure data of protein antigens. Assignment of protein function requires both, structure and interaction data. The algorithm described above is a step towards the new paradigm 'binding-determines function'.

Currently, no computational approaches are available to predict antibody-binding sites of protein antigens. Solution to the problem of identification of antibody-binding sites, even if approximate, will help in designing experiments to map the residues involved at the Ag-Ab interface. The algorithm reported here predicts probable antibody-binding sites of protein antigens.

The CEP algorithm has been evaluated using a curated dataset consisting of 21 co-crystal complexes available in PDB. The detailed evaluation of one of the PDB entries, 1FDL, is discussed. The binding site of 1FDL (14) is made up of 18 residues, consisting of 4 SEs, namely 18-19, 21-27, 116-121 and 124-126. The predicted CE5 of 1FDL best represents the BS and contains AD: 112-122 and AD: 13-24 along with individual accessible residues numbered 33, 34, 109 and 125. Thus, CEP server correctly predicts 13 out of the 18 BS residues, i.e. 72%. Since, 1FDL satisfies the objective criterion of prediction of $\geq 60\%$ of BS residues for a given

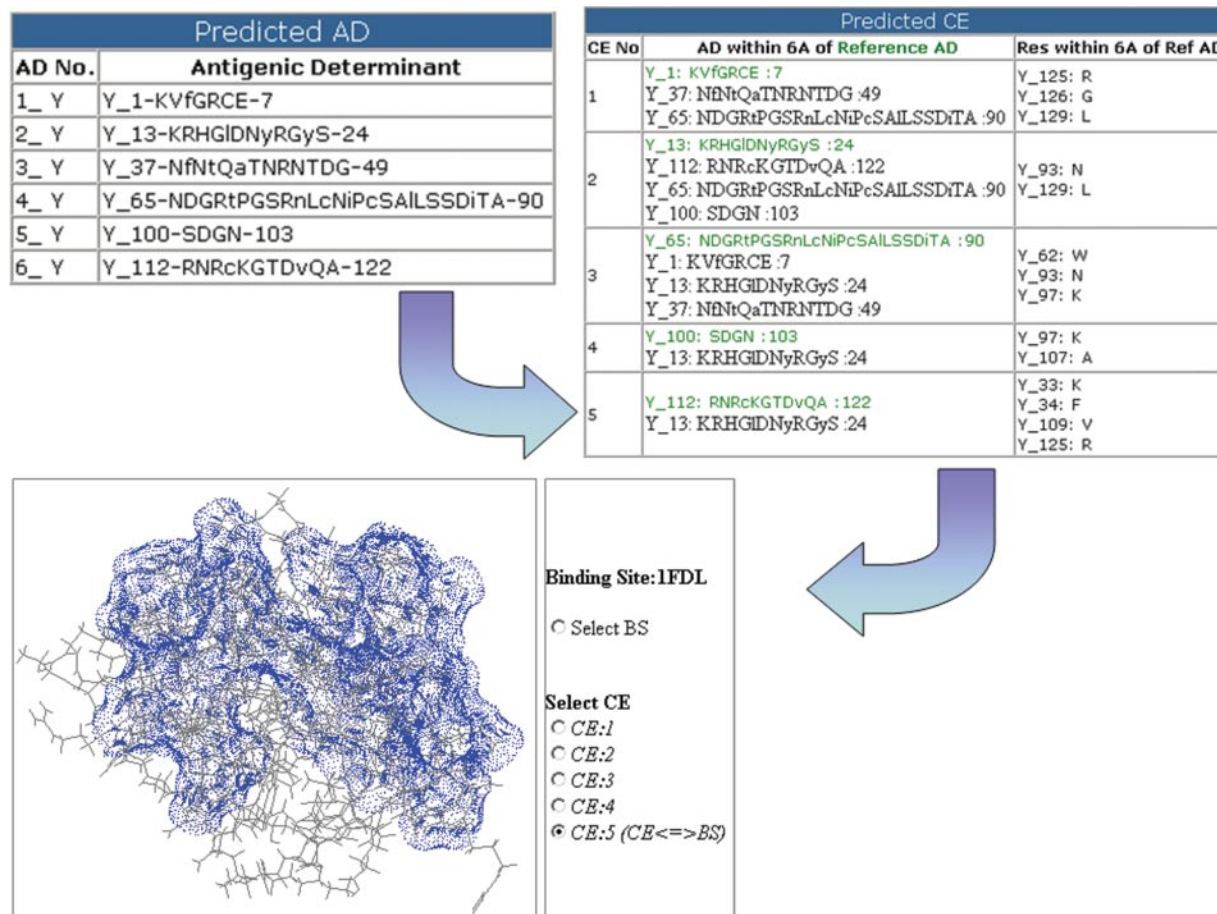


Figure 2. A snapshot of predicted AD, CE and graphical display of CE5, which corresponds to the binding site characterized by 1FDL, the complex of lysozyme with Fab D1.3. Note: The amino acid residues with percentage accessibility less than or equal to cut-off are shown in the lower case.

structure, it contributes positively to the calculation of the overall accuracy. Similar analyses were carried out for the remaining 20 structures and the evaluation results for 21 PDB files are available on the CEP server (<http://bioinfo.ernet.in/cep.htm>).

In the process of evaluation, it was observed that the algorithm predicts relatively larger binding sites for a few antigens, which may appear as false-positive predictions. An explanation for this can be drawn from the analyses of structures of antigens with multiple antibodies, such as lysozyme and neuraminidase. Complexes of lysozyme with various Abs have shown that the Abs of lysozyme have overlapping binding sites. Furthermore, it has also been shown that the same residue of an antigen may be a part of different epitopes and interact in a unique manner with respective paratopes (15). Thus, the residues that may appear 'additional' in the predicted CE need not be referred to as false positives, since the algorithm predicts all possible binding sites of the given antigens. Hence, the predicted CE/Ag-binding site is the sum of the binding sites of the individual antibodies.

Visualization and mapping of the predicted ADs and CEs on a given 3D structure enhances utility of the server. It must be mentioned that the usability of this server is limited by the availability of the 3D structure data of protein antigens. However, in the realm of structural genomics (16), we believe that structural information of proteins will no longer be a rate-limiting factor.

ACKNOWLEDGEMENTS

U.K.K. acknowledges Dr John Wootton, NCBI, USA for useful suggestions and comments. The project is funded by the Department of Biotechnology, Government of India. Contributions of Ms Sunitha Manjari in the evaluation of the server are acknowledged. The Open Access publication charges for this article have been waived by Oxford University Press.

Conflict of interest statement. None declared.

REFERENCES

1. Van Regenmortel, M.H. (1998) From absolute to exquisite specificity. Reflections on the fuzzy nature of species, specificity and antigenic sites. *J. Immunol. Methods*, **216**, 37–48.
2. Van Regenmortel, M.H. (1998) Mimotopes, continuous paratopes and hydrophobic complementarity: novel approximations in the description of immunochemical specificity. *J. Dispersion Sci. Technol.*, **19**, 1199–1219.
3. Hopp, T.P. and Woods, K.R. (1981) Prediction of protein antigenic determinants from amino acid sequences. *Proc. Natl Acad. Sci. USA*, **78**, 3824–3828.
4. Hopp, T.P. (1993) Retrospective: 12 years of antigenic determinant predictions, and more. *Pept. Res.*, **6**, 183–190.
5. Parker, J.M., Guo, D. and Hodges, R.S. (1986) 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**, 5425–5432.
6. Kolaskar, A.S. and Tongaonkar, P.C. (1990) A semi empirical method for prediction of antigenic determinants on protein antigens. *FEBS Lett.*, **276**, 172–174.
7. Alix, A.J. (1999) Predictive estimation of protein linear epitopes by using the program PEOPLE. *Vaccine*, **18**, 311–314.
8. Kolaskar, A.S. and Kulkarni-Kale, U. (1999) Prediction of three-dimensional structure and mapping of conformational antigenic determinants of envelope glycoprotein of Japanese encephalitis virus. *Virology*, **261**, 31–42.
9. Kulkarni-Kale, U. (2003) Prediction of 3D structure and function of proteins and peptides. PhD Thesis, University of Pune, India.
10. McConkey, B.J., Sobolev, V. and Edelman, M. (2002) Quantification of protein surfaces, volumes and atom-atom contacts using a constrained Voronoi procedure. *Bioinformatics*, **18**, 1365–1373.
11. Deshpande, N., Adress, K.J., Bluhm, W.F., Merino-Ott, J.C., Townsend-Merino, W., Zhang, Q., Knezevich, C., Xie, L., Chen, L., Feng, Z. *et al.* (2005) The RCSB Protein Data Bank: a redesigned query system and relational database based on the mmCIF schema. *Nucleic Acids Res.*, **33**, D233–D237.
12. Sheriff, S., Hendrickson, W.A. and Smith, J.L. (1987) Structure of myohemerythrin in the azidomet state at 1.7/1.3 Å resolution. *J. Mol. Biol.*, **197**, 273–296.
13. Ramachandran, G.N. and Sasisekharan, V. (1968) Conformation of polypeptides and proteins. *Adv. Protein Chem.*, **23**, 283–438.
14. Fischmann, T.O., Bentley, G.A., Bhat, T.N., Boulot, G., Mariuzza, R.A., Phillips, S.E., Tello, D. and Poljak, R.J. (1991) Crystallographic refinement of the three-dimensional structure of the FabD1.3-lysozyme complex at 2.5 Å resolution. *J. Biol. Chem.*, **266**, 12915–12920.
15. Malby, R.L., Tulip, W.R., Harley, V.R., McKimm-Breschkin, J.L., Laver, W.G., Webster, R.G. and Colman, P.M. (1994) The structure of a complex between the NC10 antibody and influenza virus neuraminidase and comparison with the overlapping binding site of the NC41 antibody. *Structure*, **2**, 733–746.
16. Burley, S.K. and Bonanno, J.B. (2002) Structuring the universe of proteins. *Annu Rev. Genomics Hum. Genet.*, **3**, 243–262.