Epitopes Prediction

For *Potato leaf roll virus* and development of polyclonal antibodies for viral detection

Youssef M. Bakr[3], Waiel F. Abd EL-Wahed[2] and Alaa A. Hemeida[1]

ABSTRACT

Epitopes prediction plays a vital role in the development of antibodies production and immunodiagnostic tests. This work is focused on building models for predicting linear B-cell epitopes based on support vector machine (SVM) and subsequence string kernel. The obtained models were tested by 10 fold cross validation method. We applied the obtained models to *Potato leaf roll virus* (PLRV) as a case study using Epitopes Model Applier Software (EMAS) which was developed as open source software and released under general public license (GPL) to predict immunogenic peptides suitable for antibodies production. The thirty amino acids peptide which start from position 163 to position 192 got high score and match the previous laboratory studies which make it one of the best candidates to be immunogenic and capable of producing antibodies that cross react with PLRV. The peptide was chemically synthesized and injected into animal (mouse). The obtained antibodies were tested by using TAS-ELISA and Immuno dot-blot assay. The obtained antibodies were positively reacted against PLRV infected potato tissues.

Keyword: Bioinformatics, Immunoinformatics, Support vector machine, String kernel.

Introduction

Epitope refer to any region of an antigen biomacromolecule which is recognized, or bound, by another biomacromolecule. The meaning is more restricted and refers to particular structures recognized by the immune system in particular ways. Epitope can be defined as the minimal structure necessary to invoke an immune response (Flower, 2008).

Epitopes prediction plays an important role in enhancing immunodiagnostic tests, reverse vaccinology, predicting allergenicity and antibodies production. Epitope prediction can be fairly described as both the high frontier of immunoinformatic investigation and a grand scientific challenge (Flower, 2007).

B-cell epitopes are regions of a protein recognized by antibody molecules. B-cell epitopes divided in two categories conformational epitopes and continuous epitopes. Conformational epitopes are discontinuous determinants on a protein antigen formed from several separate regions in the primary sequence of a protein brought together by protein folding. Continuous epitopes are linear antigenic determinants on proteins that are contiguous in amino acid sequence and do not require folding of a protein into its native conformation for antibody to bind with it (Cruse and Lewis, 2003). One of most important applications of predicting B-cell epitopes is computational design of immunogenic peptides to produce specific antibodies for specific protein.

The purposes of this work aimed to (1) Building datasets to train epitopes prediction models on it; (2) Building **B**-cell prediction **M**odels (BM), (3) Develop tool to apply the models to any protein sequence; (4) Predicting the epitopes in the case study (*Potato leaf roll virus*); (5) Selecting the immunogenic peptide to be injected into animal to produce antibodies that cross react with the *Potato leaf roll virus* and (6) Testing the obtained antibodies.

Materials and Methods

Waikato environment for knowledge analysis (WEKA)

The Waikato environment for knowledge analysis (WEKA) is the leading open-source project in machine learning. WEKA is a comprehensive collection of algorithms for data mining tasks written in Java and released under the GPL, containing tools for data pre-processing, classification, regression, clustering, association rules and visualization (Gewehr *et al.*, 2007). WEKA is developed in University of Waikato in New Zealand and it consists of WEKA Explorer, WEKA Experimenter, WEKA Knowledge Flow and WEKA simple command line interface. WEKA Explorer was used in this work for applying machine learning algorithm to datasets of epitopes to build **B**-cell prediction **M**odels (BM).

Datasets

The datasets used for building the epitopes prediction models are a set of the epitopes and non-epitopes peptides obtained from IEDB (Peters *et al.*, 2005) and datasets used in other work (El-Manzalawy *et al.*, 2008a). The datasets were built in ARFF format containing two class attributes (1) for positive peptides and (0) for negative peptide. We built four datasets LB01-dataset, LB02-dataset, LB03-dataset and LB04-dataset.

BM Models building

Support vector machine and subsequence string kernel were used to build models for predicting linear B-cell epitopes as described in El-Manzalawy *et al.* (2008 a and b). Table (1) shows the BM models decay factor (λ) parameter that used for building BM models.

Table 1: BM models parameters.

Model Name	Datasets	Decay factor (λ)	Sub-sequence length (SSL)
BM01	LB01	0.5	4
BM02	LB02	0.5	4
BM03	LB03	0.5	4
BM04	LB04	0.5	4
BM17	FlexLenBCPred.nr80	0.3	4
BM22	FlexLenBCPred.nr80	0.4	4
BM32	FlexLenBCPred.nr80	0.6	4
BM37	FlexLenBCPred.nr80	0.7	4

Epitopes Model Applier Software (EMAS)

Epitopes Model Applier Software (EMAS) was built on the top of Weka machine learning workbench (Frank *et al.*, 2004), Epitopes Toolkit (EpiT) and BioJava (Holland *et al.*, 2008). EMAS is available through this link:

(https://sites.google.com/site/epitopesprediction).

Case of study: *Potato leaf roll virus*

The case study was the coat protein sequence of the Egyptian isolates of *Potato leaf roll virus* (El-Attar *et al.*, 2010) where obtained from NCBI with accessions no. ACU80557 (Figure 1). The EMAS and BM models were used to predict most immunogenic peptide in this amino acids sequence of the coat protein of the Egyptian isolates of PLRV.

>gi|256387113|gb|ACU80557.1| coat protein [Potato leaf roll virus]
MSTVVVKGNVNGGVQQPRRRRRQSLRRRANRVQPVVMVTAPGQPRRRRRRRGGNRRSR
RTGVPRGRGSSETFVFTKDNLMGNSQGSFTFGPSLSDCPAFKDGIFKAYHEYKITSILLQFVS
EASSTSSGSIAYELDPHCKVRSFQSYVNKFQITKGGAKTYQARMINGVEWHDSSEDQCRIL
WKGNGKSSDTAGSFRVTIRVALQNPK

Figure 1: Amino acid sequence for coat protein of PLRV in fasta format.

Synthetic peptide and Immunization

Mice were injected five times with 50, 70, 150, 200 and 250 µg with one week interval between every injection with equal volume of complete Freund's adjuvant in first two injections and incomplete Freund's adjuvant for the rest injections. The blood was collected after one week from last injection then the antiserum was separated from blood and tested using immuno dot-blot analysis and Triple Antibody Sandwich ELISA (TAS-ELISA).

Serological detection of PLRV

PLRV-antiserum was tested using immuno dot-blot and TAS-ELISA according to the procedures described by Hans-L. Weidemann, (1988), D'Arey *et al.*,(1989) and El-Araby *et al.*, (2009).

RESULTS

Epitope prediction models

We built eight models and they are available through the link (https://sites.google.com/site/epitopesprediction). Performance evaluation of BM models done by 10 fold cross validation test and area under the Receiver Operation Characteristic (ROC) curve was calculated to all BM models (Table 2).

Table 2: BM models sorted by Area under ROC curve.

Model	ROC area	TP Rate	FP Rate	Precision	Recall	F-Measure
BM04	0.888	0.813	0.186	0.815	0.813	0.814
BM03	0.837	0.768	0.231	0.77	0.768	0.767
BM22	0.773	0.698	0.302	0.698	0.698	0.697
BM17	0.761	0.704	0.296	0.704	0.704	0.704
BM32	0.759	0.678	0.322	0.679	0.678	0.677
BM02	0.739	0.684	0.314	0.686	0.684	0.684
BM37	0.733	0.664	0.336	0.666	0.664	0.664
BM01	0.699	0.668	0.332	0.668	0.668	0.668

Epitopes Model Applier Software (EMAS)

EMAS which was developed as open source software and released under general public license (GPL) is a tool to apply models to any protein sequence. After downloading EMAS from (https://sites.google.com/site/epitopesprediction) EMAS can be run as in Figure (2) and steps to perform the prediction can be as follow:

1-Upload model file.
2-Upload test data.
3-Adjust peptide or window length.
4- Choose peptide based.
5-Choose input format as fasta sequence.
6-Make output file.
7-Click predict button to start the prediction process.

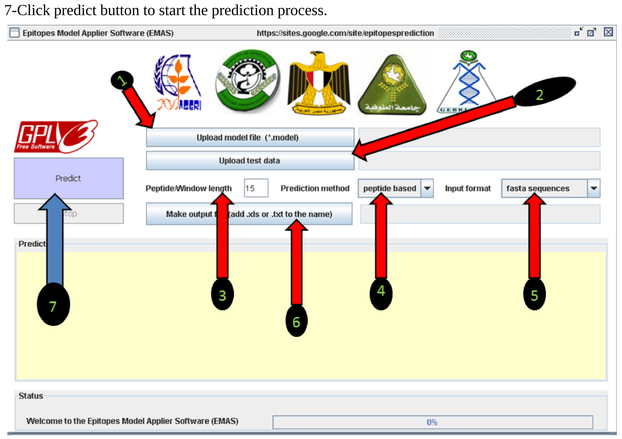

Figure 2: EMAS how to run.

Epitopes prediction to *Potato leaf roll virus* coat protein using EMAS

The *Potato leaf roll virus* coat protein sequence was retrieved from GenBank (accession no. ACU80557), then EMAS run seven times with each BM models using the PLRV coat protein sequence. The thirty amino acids peptide which starts from position 163 to position 192 got high score with most BM models. The results were match with Torrance (1992) and Terradot *et al.* (2001) studies, which make it one of the best candidates to be immunogenic and capable of producing antibodies that cross react with PLRV.

Antiserum production against PLRV-predicted epitopes

The PLRV coat protein peptide (ARMINGVEWHDSSEDQCRILWKGNGKSSDT) from position (163) to (192) was ordered from GenScript Corporation, NJ 08854, USA.

PLRV-antiserum raised against this synthetic peptide was produced using mice for immunization and was serologically tested as described below.

Serological detection of PLRV

The produced PLRV-antiserum was tested using immuno dot-blot and TAS- ELISA. Two more antisera were used for comparison: Antiserum raised against PLRV virus particles (viral antiserum) and antiserum raised against PLRV coat protein (CP antiserum).

Immuno dot-blot test

Viral, CP and synthetic peptide antisera were strongly reacted against PLRV-infected potato samples (Figure 2, samples 1, 3 and 5, respectively). However, the reaction against the synthetic peptide using the synthetic peptide-antiserum was higher than that of the viral and CP antisera (Figure 2, samples 6, 2 and 4, respectively). No reaction was detected against the healthy potato sample using synthetic peptide-antiserum (Samples 7 and 8).

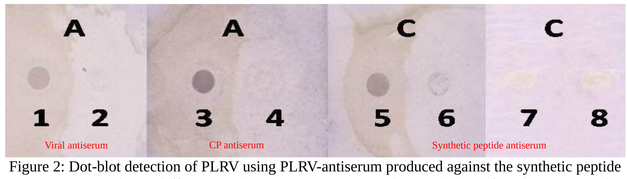

Figure 2: Dot-blot detection of PLRV using PLRV-antiserum produced against the synthetic peptide in comparison with two different PLRV-antisera. Dots 1, 3 and 5 are PLRV-infected potato sample. Dots 2, 4 and 6 are the synthetic peptide. Dot 7 and 8 are negative control. A, B and C are PLRV-antisera raised against the viral particles, the coat protein and the synthetic peptide, respectively.

TAS-ELISA

PLRV was specifically detected using the synthetic peptide, viral and CP antisera No reaction was appeared with the negative control (Table 3).

Table 3: ELISA detection of PLRV-infected potato leaves using PLRV antiserum in comparison with two different PLRV antisera

PLRV-antiserum	Healthy potato leaves	PLRV-Infected potato leaves[1]	PLRV- Infected potato leaves[2]
Virion antiserum	0.31	0.833	0.71
coat protein antiserum	0.23	0.75	0.69
Synthetic peptide antiserum	0.21	0.69	0.60

[1] and [2]: Two different samples of PLRV- potato leaves
[3] O.D. reading equal or greater than twice absorbance value of healthy controls was considered positive.

Discussion

Our approach for using computational methods for producing antibodies by predicting most immunogenic peptide in viral antigen agree with Saravanan *et al.* (2009), although they used another algorithm for epitopes prediction. They used antigenic index (residue-based predictors) which is calculated on a weighted scale by considering the presence of characters such as surface probability, hydrophilicity and flexibility of a given set of amino acids in the range of seven to eleven amino acids in a protein. So Saravanan *et al.* (2009) method depend on physical and chemical properties only which was reported for its low performance according to Blythe and Flower (2005) but Saravanan *et al.* (2009) overcome the low performance of prediction method by immunization with multiple peptides.

Our method belongs to epitope-based predictors. We used machine learning algorithms (SVM and Subsequence string Kernel) which was reported for its high performance in predicting linear B-cell epitopes by El-Manzalawy *et al.*, (2008 a and b) which enable us to use single peptide in immunization but in more length (30 instead of 11) to produce more specific antibodies.

As a conclusion, results indicate that our bioinformatics strategy is a powerful tool for antibodies production. The use of epitopes prediction by computational methods has eliminated the need to obtain large amounts of viral expressed proteins or purified virus. Also, results indicate that using BM models with EMAS in the designing and choosing of immunogenic peptide are reliable and have advantages like: (1) Producing antibodies faster and cheaper; (2) Producing antibodies for any protein we have information about its sequence even we don't have the protein itself physically. And (3) Commercialization of the produced antibodies faster and easier than antibodies produced by viral expressed proteins and cloning methods because of intellectual property rights issues related to cloning vectors.

REFERENCES

Blythe,M.J. and Flower,D.R. (2005). Benchmarking B cell epitope prediction: underperformance of existing methods. Protein Sci. 14, 246-248.

Cruse,J.M. and Lewis,R.E. (2003). Illustrated dictionary of immunology. (Boca Raton, Fla: CRC Press) pages 46, 166, 169, 221 and 236.

D'Arey, C.j., Torrance, L., and Martin,R. R. (1989).Discrimination among luteoviruses and their strain by monoclonal antibodies and identification of common Epitopes. Phytopathology 79:869-873.

El-Araby S., Ibrahim I. A., Hemeida A. A., Mahmoud Amal Soliman, A. M. El-Attar A. K. and Mazyad H. M. (2009) Biological, serological and molecular diagnosis of three major potato viruses in Egypt. International journal of virology 5[2], 77-88.

El Attar,A.K., Mohamed,A.S., Soliman,A.M. Riad,B.Y. and Mazyad,H.M.(2010). Experssion of the coat protein gene of *Potato leaf roll virus* in *Escherichia coli* and development of polyclonal antibodies. Arab J. Biotech., Vol. 13, No. (1):85-98.

El-Manzalawy,Y., Dobbs,D., and Honavar,V. (2008a). Predicting flexible length linear B-cell epitopes. Comput. Syst. Bioinformatics. Conf. 7, 121-132.

El-Manzalawy,Y., Dobbs,D., and Honavar,V. (2008b). Predicting linear B-cell epitopes using string kernels. J. Mol. Recognit. 21, 243-255.

Flower,D.R. (2007). Immunoinformatics: Predicting immunogenicity in silico. (Humana) page 21.

Flower,D.R. (2008). Bioinformatics for vaccinology. (Chichester: John Wiley) page 82.

Frank,E., Hall,M., Trigg,L., Holmes,G., and Witten,I.H. (2004). Data mining in bioinformatics using Weka. Bioinformatics. 20, 2479-2481.

Gewehr,J.E., Szugat,M., and Zimmer,R. (2007). BioWeka--extending the Weka framework for bioinformatics. Bioinformatics. 23, 651-653.

Hans-L. Weidemann (1988). Rapid detection of potato viruses by dot-ELISA. Potato Research 31, 485-492.

Holland,R.C., Down,T.A., Pocock,M., Prlic,A., Huen,D., James,K., Foisy,S., Drager,A., Yates,A., Heuer,M., and Schreiber,M.J. (2008). BioJava: an open-source framework for bioinformatics. Bioinformatics. 24, 2096-2097.

Peters,B., Sidney,J., Bourne,P., Bui,H.-H., Buus,S., Doh,G., Fleri,W., Kronenberg,M., Kubo,R., Lund,O., Nemazee,D., Ponomarenko,J.V., Sathiamurthy,M., Schoenberger,S., Stewart,S., Surko,P., Way,S., Wilson,S., and Sette,A. (2005). The Immune Epitope Database and Analysis Resource: From Vision to Blueprint. PLoS Biol 3.

Saha,S. and Raghava,G.P. (2006). Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. Proteins 65, 40-48.

Saravanan,P., Shrivastava,S., and Kumar,S. (2009). Synthesis of Highly Immunogenic Multiple Antigenic Peptides for Epitopes of Viral Antigen to Use in ELISA. International Journal of Peptide Research and Therapeutics 15, 313-321.

Terradot,L., Souchet,M., Tran,V., and Giblot Ducray-Bourdin,D. (2001). Analysis of a three-dimensional structure of Potato leafroll virus coat protein obtained by homology modeling. Virology 286, 72-82.

Torrance,L. (1992). Analysis of epitopes on potato leafroll virus capsid protein. Virology 191, 485-489.

ACKNOWLEDGEMENT

I would like to express my appreciation to my supervisors Prof. Dr. Alaa Eldin Abd Allah Hemeida and Prof. Dr. Waiel Fathi Abd El-Wahed for their scientific guidance and support that made the study scientifically oriented and the work successful. My thanks are also extended to all the staff members of Genetic Engineering and Biotechnology Research Institute (GEBRI) and Agricultural Genetic Engineering Research institute (AGERI). Special thanks to Dr. Ahmed Ashoub, Dr. Ahmed Shokry, Dr. Amal Mahmoud Hussien, Dr Hanan Nour El-Din and Dr. Nasser El-Din Abd El-Razik for their encouragement and supporting. Finally I would to express my appreciation to Dr. Yasser EL-Manzalawy for inspiring and helping me.

1. Genetic Engineering and Biotechnology Research Institute,
Minoufiya University,
Sadat City, Egypt.
2. Faculty of Computers and Information,
Minoufiya University,
Shebin El-Kom, Egypt.
3. Agricultural Genetic Engineering Research Institute,
Agricultural Research Center,
Giza, Egypt.

EMAS beside all models and datasets are available through this link: https://sites.google.com/site/epitopesprediction