

Immunoinformatics, antigenicity epitopes prediction in the solute carrier family 11 of the natural resistance associated macrophage protein 1 (nramp) related with brucellosis in cattle

Abstract

Brucellosis is a highly contagious zoonotic disease occurring in humans and various species of domesticated and wild animals including Cattle. As a common knowledge, the antigen is a heavyweight molecule that induces an immune system to produce antibodies against it. This study focused on antigenic epitopes binding prediction of antibodies that produced from the B-cells in cattle against Brucellosis disease, by depending on two major methods schools and compared between them. Firstly, the theory assumed that antigenic sites on proteins have the hydrophobic residues which coined.¹ Secondly, the assumption that came in contrast form the other which mentioned, that antigenic regions are primarily hydrophilic regions at the surface of the protein molecule, claimed.² The results of the antigenicity epitopes prediction in the solute carrier family 11 of the natural resistance-associated macrophage protein 1 (NRAMP) related with Brucellosis in cattle. The multiple sequence alignment of six breeds that show resistance against the diseases a highly similarities in proteins sequences by providing identity score 542/548 (98.9%), similarity score 547/548 (99.8%) and gaps score 0/548 (0.0%). Additionally, the higher summation frequencies score for the amino acids which have Hydrophobic properties (A, F, G, I, L, M, P, V, and W) that count 340/548 a.a. and the frequency scored 0.620 and the lower for the amino acids which have Hydrophilic (C, N, Q, S, T, and Y) count 133/548 a.a. with frequency score 0.243, with the remaining 75/548 a.a. and frequency 0.137 which have not hydrophobic or hydrophilic physiochemical properties. Moreover, a significant superiority results in favor of the first theory.¹

Keywords: cattle, brucellosis, antigenicity, epitopes predicted, emboss, emma, gnuPPSP and perl language

Volume 7 Issue 4 - 2018

Omar Esmail Hamad¹, Mehmet Sait Ekinci¹,
Emin Ozkose¹, Ismail Akyol²

¹Department of Animal Science, Kahramanmaraş Sutcu Imam University, Turkey

²Biotechnology and Gene Engineering Laboratory, Kahramanmaraş Sutcu Imam University Turkey

Correspondence: Mehmet Sait Ekinci, Department of Animal Science, Kahramanmaraş Sutcu Imam University, Turkey, Tel +90 344 2802100, Email sekinci@ksu.edu.tr

Received: June 26, 2018 | **Published:** August 07, 2018

Introduction

Brucellosis is responsible for enormous economic losses as well as considerable human morbidity in endemic areas. The bacteria infects animals such as swine, cattle, goat, sheep, and dogs. Humans can become infected indirectly through contact with infected animals or by animal products consumption. Brucellosis occurs worldwide, but it is well controlled in most developed countries. The disease is rare in industrialized nations because of routine screening of domestic livestock and animal vaccination programmers.⁴ Clinical disease is still common in the Middle East, Asia, Africa, South and Central America. This review article aims to describe the prevalence of brucellosis in some countries these data are available around different regions of the world, and risk factors associated infections according to regression models.^{3,4} There are two species of bacterial pathogens that recorded cause brucellosis disease in cattle. Firstly, the (*Brucella abortus*) with genome size 3,264,306 base pairs divided into two unequal size chromosomes ([https://www.ncbi.nlm.nih.gov/genome/?term=Brucella+ abortus](https://www.ncbi.nlm.nih.gov/genome/?term=Brucella+abortus)). The whole genome that downloaded from Gen Bank of the National Center of Biotechnology and Information (NCBI), within accession numbers NC_007618.1 and NC_007624.1 for chromosome I and II respectively.⁵ Secondly, the (*Brucella melitensis*) and its genome size 3,294,931 base pairs also divide into two unequal size chromosomes (<https://www.ncbi.nlm.nih.gov/genome/?term=Brucella+melitensis>). The NCBI accession numbers of whole-genome chromosome I and II are NC_003317.1 and NC_003318.1 respectively.⁶ Predicting the antigenic sites on proteins is of major importance for the production of synthetic artificial peptide vaccines and peptide probes of antibody structure. Many predictive methods, based on various assumptions about the nature of the antigenic response have been proposed and tested. This review will discuss the principles underlying the different approaches to predicting antigenic sites and will attempt to answer the question of how well they work. As a review of Kolaskar & Tongaonkar method which coined in 1990. Analysis of data from experimentally determined antigenic sites on proteins has revealed that the hydrophobic residues Cys, Z_XU, and Val if they occur on the surface of a protein, are more likely to be a part of antigenic sites. A semi-empirical method which makes use of physiochemical properties of amino acid residues and their frequencies of occurrence in experimentally known segmental epitopes was developed to predict antigenic determinants on proteins. Application of this method to a large number of proteins has shown that our method can predict antigenic determinants with about 75% accuracy which is better than most of the known.^{7,8} In another hand, the welling method for antigenicity prediction in 1985 came in contrast. Prediction of antigenic regions in a protein will be helpful for a rational approach to the synthesis of peptides which may elicit antibodies reactive with the intact protein. Earlier methods are based

Nih.gov/genome/?term=Brucella+melitensis). The NCBI accession numbers of whole-genome chromosome I and II are NC_003317.1 and NC_003318.1 respectively.⁶ Predicting the antigenic sites on proteins is of major importance for the production of synthetic artificial peptide vaccines and peptide probes of antibody structure. Many predictive methods, based on various assumptions about the nature of the antigenic response have been proposed and tested. This review will discuss the principles underlying the different approaches to predicting antigenic sites and will attempt to answer the question of how well they work. As a review of Kolaskar & Tongaonkar method which coined in 1990. Analysis of data from experimentally determined antigenic sites on proteins has revealed that the hydrophobic residues Cys, Z_XU, and Val if they occur on the surface of a protein, are more likely to be a part of antigenic sites. A semi-empirical method which makes use of physiochemical properties of amino acid residues and their frequencies of occurrence in experimentally known segmental epitopes was developed to predict antigenic determinants on proteins. Application of this method to a large number of proteins has shown that our method can predict antigenic determinants with about 75% accuracy which is better than most of the known.^{7,8} In another hand, the welling method for antigenicity prediction in 1985 came in contrast. Prediction of antigenic regions in a protein will be helpful for a rational approach to the synthesis of peptides which may elicit antibodies reactive with the intact protein. Earlier methods are based

on the assumption that antigenic regions are primarily hydrophilic regions at the surface of the protein molecule.⁹ The method presented here is based on the amino acid composition of known antigenic regions in 20 proteins which is compared with that of 314 proteins Sequences and Structure. Antigenicity values were derived from the differences between the two datasets. The method was applied to bovine ribonuclease, the B-subunit of cholera toxin and herpes simplex virus type 1 glycoprotein D. There was a good correlation between the predicted regions and previously determined antigenic regions.^{10,11} The most important point of this study is starting from scratch depending on a row data from the trusty and proved sources like NCBI. All of sequences and programs that mentioned before were examined by EMBOSS web servers then decided to choose the protein sequences that produced as antibodies by the B-cell which associated with Brucellosis resistance in cattle.

Material and methods

In this research depended mainly on the computation approach to predict the epitopes that bind with antigens depending on the results of previous researches. The six protein sequences that produced by the six genes as indicated in Table 1, have same sequences with difference in one or two amino acids that have not affected to make any change in functionality, which appeared in superfamily breeds of cattle that show a significant resistance against Brucellosis disease.^{12–15}

Table 1 Database sources of SLC11A1 gene with the NRAMP proteins that produce

Protein's accession number	DNA's accession number	DNA's accession number	DNA's references	Protein's references
1	NP_777077	*AC_000159	(Hedges, Kimmel, Snyder, Jerome & Jutila, 2013)	
2	ABF61463	DQ493965	** (Martinez et al, 2008)	* (Martinez et al, 2006)
3	ABM81484	DQ848779	** (Schutta et al, 2006)	* (Schutta et al, 2006)
4	ALC78257	KR002419	** (Zhang et al, 2015)	* (Zhang et al, 2015)
5	ALC78258	KR002420	** (Zhang et al, 2015)	* (Zhang et al, 2015)
6	ALC78259	KR002421	** (Shi et al, 2015)	* (Shi et al, 2015)

The sources of database

As a Bioinformatics investigation in genomics and proteomics that associated with Brucellosis in cattle must be looking for both inside the pathogen which causes brucellosis disease and the animal that infected with. Also searching inside the database of the genes that related with producing the antibody to provide the disease's resistance against brucellosis in cattle as demonstrated in Table 1. The most proved name and symbol is Solute Carrier Family 11 (SLC11A1) ([https:// www.uniprot.org/uniprot/](https://www.uniprot.org/uniprot/); <https://www.Omim.org/entry/>), also known as Natural Resistance-Associated Macrophage Protein (NRAMP).^{16,17} *The actual location of the gene is (107114929..107125854) within this accession number which refers to the chromosome II in Bos taurus breed Hereford. **unpublished as a research paper. Equally important, get downloaded the natural

resistance-associated macrophage protein 1, which translated from the SLC11A1 gene. As indicated in Table 1, the accession numbers resources of researchers that worked on this protein and proved in the laboratory, also worth to mention that all these sources of protein share same sequences and size 548 amino acids.^{18–20}

Computation approach

All packages of applications that used in this study are freely available for academic users and were compiled and/or under the Gnu/Linux operating system. The core applications used for deriving the antigenicity are in the European Molecular Biology Open Software Suite EMBOSS Stable released version 6.6.0, which is a free open source software analysis package specially developed for the needs of the molecular biology and Bioinformatics user community ([https:// emboss.sourceforge.net/apps/](https://emboss.sourceforge.net/apps/)). Mainly, there are three programs that used under EMBOSS. Additionally, the codes of these applications are tested, converted and developed to serve this research specifically like adding the welling method 1985 that have not before with other functions by compiling applying all codes by using Perl programming language (version 5.22.1).^{20,21} Firstly, the EMMA program, which designed for Multiple sequence alignment (Clustal W wrapper) by calculate the multiple alignment of nucleic acid or protein sequences according to the method of the usage of program by the command line order (% Emma) then paste the sequence in FASTA format.²² This is an interface to the Clustal W distribution. Secondly, the antigenic program for epitopes binding prediction and used by (% antigenic, and % antigenic-reformat off) in the command line (Terminal) to find antigenic sites in proteins sequences. The algorithm of data analysis from experimentally determined antigenic sites on proteins has revealed that the hydrophobic residues Cys, Leu and Val if they occur on the surface of a protein, are more likely to be a part of antigenic sites (<http://emboss.open-bio.org/rel/dev/apps/antigenic.html>). The method of Kolaskar and Tongaonkar also the welling method are applied to predict antigenic determinants in proteins is semi-empirical and makes use of physicochemical properties of amino acid residues and their frequencies of occurrence in experimentally known segmental epitopes.²³

Thirdly, the PLOTCON program Sequence conservation is calculated for windows of a specified length over the alignment. Within a window, the similarity of any one position is taken to be the average of all the possible pairwise substitution scores of the bases or residues at that position. The pairwise substitution scores are taken from the specified similarity matrix. The average of the position similarities within the window is plotted. Which depended on applying some results from ClustalW with a mathematical equation like (sequence weighting, matrix comparison table, number of sequences in the alignment and window size) within blocks substitution matrix. The application usage by main code of command line is (% plotcon-format MSF globins.msf-graph cps).²⁴ Furthermore, using the (GNU PSPP) and Gnumeric Spreadsheet 1.12.9 programs for statistical analysis of sampled data and display plots. Also, they are a Free replacement for the proprietary program SPSS and appears very similar to it with a few exceptions. To display the plots of antigenicity chart also to apply the basic calculations of amino acids physicochemical properties for protein's Hydrophobic and Hydrophilic through the sequence ([https:// www.gnu.org/ software /pspp/](https://www.gnu.org/software/pspp/)). These sequences stored as XML files which contain the FASTA format to use as an input data. The terminal in Linux help to apply for the programs though it by typing the codes

related with running the programs then paste the sequences to get the results saved specific folder within special format.²⁵

Result and discussion

The multiple sequence alignment

The results of multiple sequence alignment by ClustalW wrapper between the six protein sequences which have same amino acid size number (548 a.a.) with substitution in one or two amino acids for each sequence. Provide identity score 542/548 (98.9%), similarity score 547/548 (99.8%) and gaps score 0/548 (0.0%). The reason behind these high scores is the similar number of amino acids that cause no gaps because the gaps designed to produce same strings length after alignment, also the low value of substitutions to generate these values. Furthermore, the sequence alignment leads to similarity plot as demonstrated in Figure 1, the plot Presents highest Similarity scores between the six proteins sequences for each residue of 10→20 amino acids. Moreover the highest scores recorded for 2.0 and more of the string positions (120→130; 170→180; 220→240; 410→430 and 520→550) also the highest results ever was in position 220→260; on the other hand the lowest similarity that scores recorded was 1.0 in the position 210→218 but shown higher than the main of expect. The Hydrophobic and Hydrophilic. The importance of hydrophobic and hydrophilic estimation is related to the two different schools in Immunoinformatics. Firstly, the method of declared the antigenic sites on proteins has revealed that the hydrophobic residues.^{1,26,27} Secondly, welling method in 1985, assumed that antigenic regions are primarily hydrophilic regions at the surface of the protein molecule.²⁸ The results of hydrophobic and hydrophilic estimation depend on to calculate the amino acids physiochemical properties. Figure 2, evince he amino acids frequencies distributions, shows the higher summation frequencies score for the amino acids which have Hydrophobic properties (A, F, G, I, L, M, P, V, and W) that count 340/548 a.a. and the frequency scored 0.620 and the lower for the amino acids which have Hydrophilic (C, N, Q, S, T, and Y) count 133/548 a.a. with frequency score 0.243, with the remaining 75/548 a.a. and frequency 0.137 which have not hydrophobic or hydrophilic physiochemical properties. The antigenic epitopes binding prediction Because of the high similarity of the six protein sequences, the results of the antigenic epitopes binding prediction came in same for all over the six sequences. Thus, Table 2 illustrates the results with avoiding the repetition and indicated in sort from higher to lower score, also declare the position of an amino acid within a maximum score. The sequences of residues represented the peptides of epitopes that recognize and bind with the antigen of the bacterial pathogen which cause the Brucellosis in cattle as in the Figure 3, shown as a linearly interpolate between multi-dimensional points display as a comparison between method of versus welling method in 1985, The antigenicity plot starts with 0.0 score in the middle of vertical line and the score raise in two directions, up positively and down negatively; the positive results indicate the present to be a part of antigenic sites. by applying the converted amino acids sequence to numerical scores represent the antigenic prediction for every single position within GNU PSPP program to show a significant supremacy and confidence with the method of which display in black line which stays in the top score for all residues except one position 260→270.¹ Similarly, many types of research are a concert with these results.²⁹ The blue linear represent the plot results with the.²⁷ The black linear the plot results.²⁹

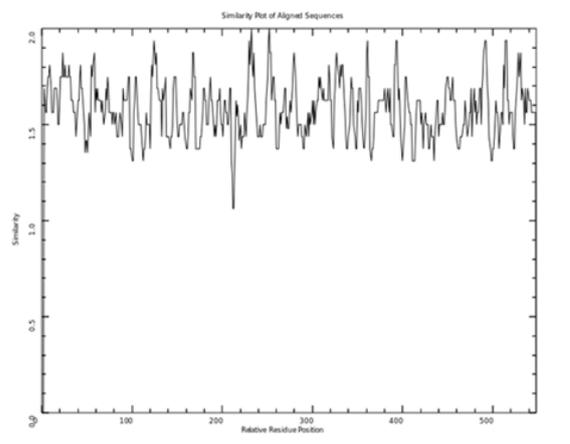


Figure 1 The plot presents high similarity between the six proteins sequences.

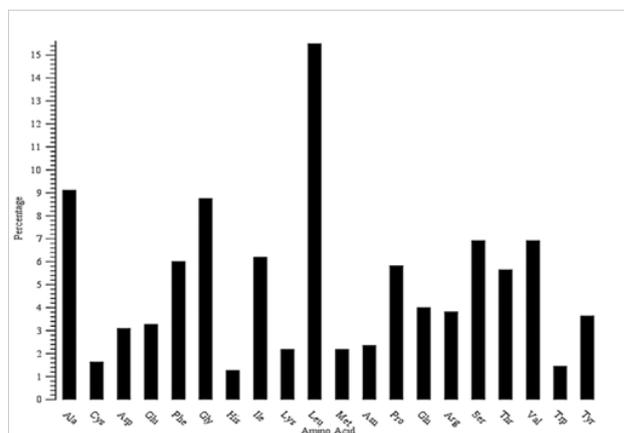


Figure 2 The amino acids frequencies distribution.

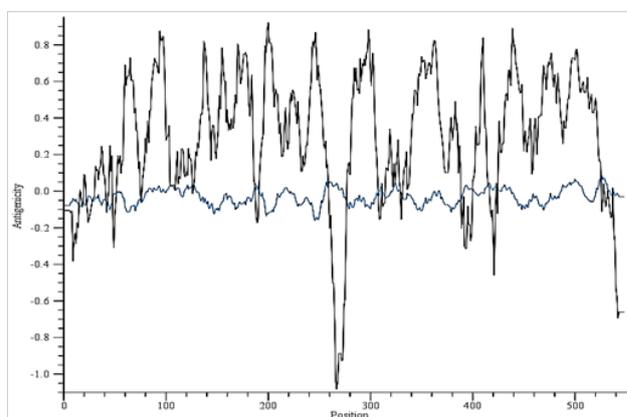


Figure 3 The antigenicity plot for epitopes binding prediction.

Table 2 The antigenic epitopes binding prediction

S. No	Score	Length	Maximum score position	Residue	Sequence
1	1.243	79	473	458→536	NGLVSKVITSSIMVLVCAVNLYFVISYLPSLPHPAYFSLVALLAAAYL-GLTTYLVWVTCITQGATLLAHSSHQRFLYGL
2	1.219	28	124	118→145	LGEVCHLYPKVPRILLWLTIELAIVGS
3	1.205	37	101	80→116	QAGAVAGFKLLWVLLWATVVLGLLCQRLAARLGVVTGK
4	1.202	35	350	335→369	NLTVAVDIYQGGVILGCLFGPPALYIWAUWVGLLAAG
5	1.201	24	433	424→447	LNDLLNVLQSLLLPFAVLPILTFT
6	1.191	20	173	168→187	WGGVLITVVDTFFFLFDNY
7	1.19	24	412	395→418	FARVLLTRSCAILPTVLLAVFRDL
8	1.168	76	258	189→264	LRKLEAFFGFLITIMALTFGYEVVVAQPAQGALLQGLFLPSCPGCG-QPELLQAVGIIGAIIMPHNIYLHSSLVKSR
9	1.164	30	292	279→308	MYFLIEATIALSVSFLINLVMMAVFGQAFY
10	1.126	16	157	150→165	VIGTAIAFSLLSAGRI
11	1.11	16	321	316→331	FNICADSSLHDYAPIF
12	1.092	7	67	64→70	LMSIAFL
13	1.06	7	38	37→43	SEKIPIP
14	1.054	13	21	17→29	SISSPPSPEPQQA
15	1.052	8	385	384→391	MEGFLKLR
16	1.05	7	55	52→58	LRKLWAF

Conclusion

The main aim behind prediction of antigenic epitopes is to find the maximum probability of potential peptides residues could recognize and bind with the antigen, which is become very handy to design drugs and looking for increasing the number of animals that have the ability to produce this protein of the natural resistance-associated macrophage protein 1 (NRAMP) related with Brucellosis in Cattle.

Acknowledgments

None.

Conflict of interests

The author declares no conflict of interest.

References

- Kolaskar AS, Tongaonkar PC. A semi-empirical method for prediction of antigenic determinants on protein antigens. *FEBS Letters*. 1990;276(2):172–174.
- Welling GW, Weijer WJ, Welling-Wester S, et al. Prediction of sequential antigenic regions in proteins. *FEBS Letters*. 1985;188(2):215–218.
- Robinson A. Guidelines for coordinated human and animal brucellosis surveillance. *FAO Rome*. 2003.
- Alton GG, Forsyth JRL, Brucella. *Medical Microbiology*. In: S Baron, editor. 4th ed. Galveston (TX): University of Texas Medical Branch at Galveston, USA; 1996.
- Lopes B, Nicolino L, Haddad PA, et al. Brucellosis-risk factors and prevalence: A review. *The Open Veterinary Science Journal*. 2010;4(1).
- De RK, Tomar N. *Immunoinformatics*. Springer; 2014.
- Xue J, Bai Y, Chen Y. Expression of Helicobacter pylori Alpi protein and its immunogenicity. *World Journal of Gastroenterology*. 2005;11(15):2260–2263.
- Del Vecchio VG, Kapatral V, Redkar RJ, et al. The genome sequence of the facultative intracellular pathogen Brucella melitensis. *National Academy of Sciences of the United States of America*. 2002;99(1):443–448.
- Sun C, Zhao J, Li, W, et al. Antigenicity and physicochemical properties of two human NR1a polypeptides related to activation of NMDA receptor. *USA*. 2002;82(1):50–53.
- Chain PSG, Commerce DJ, Tolmasky ME, et al. Whole-genome analyses of speciation events in pathogenic Brucellae. *Infection and Immunity*. 2005;73(12):8353–8361.
- Gedezha MP, Mphahlele MJ, Selabe SG. Characterization of HCV Genotype 5a Envelope Proteins: Implications for Vaccine Development and Therapeutic Entry Target. *Hepatitis Monthly*. 2014;14(11).
- Paixão TA, Poester FP, Carvalho Neta AV, et al. NRAMP1 3' untranslated region polymorphisms are not associated with natural resistance to Brucella abortus in cattle. *Infection and Immunity*. 2007;75(5):2493–2499.
- Zhang C, Wang Y, Chen H, et al. SLC11A1 gene polymorphisms are not associated with somatic cell score and milk yield in Chinese Holstein. *Veterinary Immunology and Immunopathology*. 2009;127(4):389–392.
- Feng J, Li Y, Hashad M, et al. Bovine natural resistance-associated macrophage protein 1 (Nramp1) gene. *Genome Research*. 1996;6(10):956–964.

15. Hedges JF, Kimmel E, Snyder DT, et al. Solute carrier 11A1 is expressed by innate lymphocytes and augments their activation. *Journal of Immunology*. 2013;190(8):4263–4273.
16. Horin P, Rychlik I, Templeton JW, et al. A complex pattern of microsatellite polymorphism within the bovine NRAMP1 gene. *Eur J Immunogenet*. 1999;26(4):311–313.
17. Flower DR. *Immunoinformatics: predicting immunogenicity in silico*. Totowa, NJ: Humana Press. 2007.
18. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties, and weight matrix choice. *Nucleic Acids Research*. 1994;22(22):4673–4680.
19. Wall L, Schwartz RL, Christiansen T. *Programming Perl*. 2nd ed. Sebastopol, CA: O'Reilly & Associates, USA; 1996.
20. Amat-ur-Rasool H, Saghir A, Idrees M. Computational Prediction and Analysis of Envelop Glycoprotein Epitopes of DENV-2 and DENV-3 Pakistani Isolates: A First Step towards Dengue Vaccine Development. *Plos One*. 2015;10(3):e0119854.
21. Cuadra D. Esperanza Albacete Javier. *Calle Elena Castro*; 2012.
22. Ison J, Rice P, Bleasby A. *EMBOSS developer's guide: bioinformatics programming*. New York; 2011.
23. Cai Y, Bao J, Lao X, et al. Computational design, functional analysis and antigenic epitope estimation of a novel hybrid of 12 peptides of hirudin and reteplase. *Journal of Molecular Modeling*. 2015;21(9):229.
24. Sealey KL, Kirk RS, Walker AJ, et al. Adaptive radiation within the vaccine target tetraspanin-23 across nine *Schistosoma* species from Africa. *International Journal for Parasitology*. 2013;43(1):95–103.
25. Parker JM, Guo D, Hodges RS. 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*. 1986;25(19):5425–5432.
26. Larsen JEP, Lund O, Nielsen M. An improved method for predicting linear B-cell epitopes. *Immunome Research*. 2006;2(1):1.
27. Emini EA, Hughes JV, Perlow D, et al. Induction of hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide. *Journal of Virology*. 1985;55(3):836–839.
28. Sales CM, Wakker PP, Alves PC, et al. MF Calculator: A web-based application for analyzing similarity. USA; 2012.
29. Schönbach C, Ranganathan S, Brusci V. *Immunoinformatics*. New York: Springer; 2008.
30. Stern PS. Predicting antigenic sites on proteins. *Trends in Biotechnology*. 1991;9(5):163–169.