Immunoinformatic prediction about potential novel vaccine in surface antigen fragment protein of *Toxoplasma gondii*

Seyed Sajjad Hasheminasab *[1], Hossein Maghsood [2], Sara Khalili [3]

Department of Parasitology, Faculty of Veterinary Medicine, University of Tehran, Tehran, Iran
*Corresponding author E-mail: s.sajjad.hn@gmail.com

Abstract

Toxoplasmosis is one of the most widespread infections in animals and humans. The *Toxoplasma gondii* major surface antigen, called SAG1 or p30, is a highly immunogenic protein which has generated great interest as a diagnostic reagent, as a potential subunit vaccine, and for its role in invasion. In this study, the epitopes of *Toxoplasma gondii* SAG1 were identified using bioinformatics. Through the analysis of the out¬put of both NetCTL and CTLPred, and B-cell epitope prediction, the position of all the epitopes were found and combined in four sequences. The different tasks including, T-cell and B-cell prediction, Antigenicity determination of the conserved peptides, Homology modeling, Allergenicity and epitope conservancy analysis were done on the conserved peptides. We predict that our proposed epitopes would also trigger an immune response in vitro.

Keywords: *Immunoinformatic; SAG1; Toxoplasma Gondii.*

1. Introduction

Toxoplasma gondii is an obligate, intracellular, parasitic protozoan that causes the disease toxoplasmosis (Hill & Dubey 2002). Found worldwide, T. gondii is capable of infecting virtually all warm-blooded animals(Tenter et al. 2000), although felids such as domestic cats are the only known definitive hosts in which the parasite can undergo sexual reproduction(Elmore et al. 2010). serological studies estimate that 30–50% of the global population has been exposed to and may be chronically infected with T. gondii, although infection rates differ significantly from country to country(Flegr et al. 2014). For example, previous estimates have shown the highest prevalence of persons infected to be in France, at 84%. Although mild, flu-like symptoms occasionally occur during the first few weeks following exposure, infection with T. gondii produces no readily observable symptoms in healthy human adults(Berdoy et al. 2000, Flegr et al 2014). In the present study, we have used an immunoinformatics approach to identify potential epitope that could form the basis of a vaccine against the TOXGO Surface antigen protein segment protein of T. gondii. Both T-cell and B-cell epitopes were assessed, and four epitopes in combining T-cell and B-cell predicted epitopes were found to be the most promising. The epitopes were also found to be no allergenic as well. In vivo study of our proposed peptide is advised for novel universal vaccine production, which might be an effective path to prevent T. gondii disease.

2. Methodology

Sequence retrieval and conserved region identification UniProtKB (Apweiler et al. 2004) database was used for the retrieval of the sequences of Surface antigen protein Fragment protein of Toxoplasma gondii the FASTA protein format. CLUSTAL OMEGA sequence alignment was used for the identification of the conserved region among the sequences through multiple-sequence alignment (MSA) (Clustal 1994). Mega6 software was used for further sequence analysis.

2.1. Antigenicity determination of the conserved peptides

VaxiJen v2.0 Web-based server (Doytchinova & Flower 2007), was used for the determination of the antigenicity of the conserved sequences. Herein, we used the default parameters with a threshold value of 0.4 for the prediction.

Table 1: Antigenicity Determination of the Conserved Peptides by Vaxijen Server

Peptide Number	Peptide	Vaxijen Score (Threshold: 0.4)
1	PTENHFTLKCPKTALTEPPTLAYSPNRQICPAGTTSSCTSKAVTLSSLIPE	0.5627
2	SASLDTAGIKLTVPIEKFPVTTQTFVVGCI	0.5826
3	DAQSCMVTVTVQARASSVVNNVARCSYGAD	0.8842
4	TLGPVKLSAEGPTTMTLVCGKDGVKVPQDNNQYCSGT	0.8333

Fig. 1: Multiple-Sequence Alignment of 8 Surface Antigen Protein Sequences by CLUSTAL OMEGA

2.2. T-cell epitope prediction

Two online servers were used for the prediction of the T-cell epitope. First of all, the NetCTL v1.2 server was used for the identification of the potential T-cell epitope(Larsen et al. 2007). We used the default approach to predict the epitopes. The threshold for epitope identification was set at 0.5 to maintain sensitivity and specificity of 0.89 and 0.94, respectively. Finally, CTLPred(Bhasin & Raghava 2004) was implemented additionally for further confirmation about the prediction with default parameters.

Table 2: Prediction of the T-Cell Epitope by Ctlpred Server

Peptide Rank	Start Position	Sequence	Score(ANN/SVM)
1	114	RASSVVNNV	0.98/1.3782908
2	29	AYSPNRQIC	0.79/1.0925069
3	101	DAQSCMVTV	0.97/0.64089241
4	70	SLDTAGIKL	0.63/0.73284585
5	9	TENHFTLKC	0.88/0.41939895

Abbreviations: ANN/SVM, artificial neural networks/support vector machines.

Table 3: Prediction of the T-Cell Epitope by Netctlserver on the Basis of Combined Score

Number	Epitope	Combined score (nM)
1	LTEPPTLAY	3.4348
2	KVPQDNNQY	0.8830
3	PTENHFTLK	0.8511
4	SLDTAGIKL	0.8056
5	DTAGIKLTV	0.6559
6	CTSKAVTLS	0.6202

2.3. B-cell epitope prediction

B-cell epitope initiates immunoresponse through the interaction with B lymphocytes and causes the differentiation of B lympho-cytes into plasma and memory cells(Nair et al. 2002). IEDB-AR hosts a number of Web-based tools for the prediction of B-cell epitope. Multiple tools, including the Kolaskar and Tongaonkar antigenicity scale(Kolaskar & Tongaonkar 1990).

Table 4: Kolaskar and Tongaonkar Antigenicity Analysis

No.	Start Position	End Position	Peptide	Peptide Length
1	14	20	TLKCPKT	7
2	23	31	TEPPTLAYS	9
3	33	40	NRQICPAG	8
4	42	58	TSSCTSKAVTLSSLIPE	17
5	71	97	LDTAGIKLTVPIEKFPVTTQTFVVGCI	27
6	102	128	AQSCMVTVTVQARASSVVNNVARCSYG	27
7	132	141	TLGPVKLSAE	10
8	146	158	MTLVCGKDGVKVP	13
9	162	168	NQYCSGT	7

2.4. Homology modeling

Homology model of the conserved region was obtained by SWISS MODEL(Šali et al. 1995), and the predicted model was assessed by PROCHECK server(Laskowski et al. 1996).
Allergenicity and epitope conservancy analysis
In order to predict the allergenicity of the proposed epitopes with high accuracy, a Web-based server AlgPred(Saha & Raghava 2006) was used. We used a hybrid prediction (SVMc + IgEepitope + ARPs BLAST + MAST) approach to predict the allergenicity with an accuracy of about 87% at a threshold value of −0.4. A Web-based tool from IEDB-AR(Bui et al. 2007) was used in order to identify the specific conservancy of the proposed epitopes.

3. Results

Analysis of the retrieved sequences and their divergence

A total of 8 Surface antigen protein Fragment protein of Toxoplasma gondii from different variants of the CCHFV were retrieved from the UniProt database. Conserved regions were found for the antigenic property analysis. Mega6 software was used to construct phylograms in order to analyze the divergence among the retrieved sequences and clustal omega was used for alignment.

3.1. Antigenic peptide identification

Initially, the conserved sequence of Surface antigen protein Fragment of Toxoplasma gondii was selected for further analysis. Then the VaxiJen v2.0 server was used to predict the antigenicity of the conserved peptides (Table 1). In this study the epitopes for both T-cell and B-cell were combined and finally we had 4 combined epitope peptides shown in table 1.

3.2. B-cell and t-cell epitope identification

T-cell epitopes were predicted using two online servers of CTLPred server and NetCTL v1.2 server. The results are shown in table 2 and table 3, respectively. Amino acid–based methods were used for the prediction of potential B-cell epitope. Antigenic property of the peptides was assessed using the Kolaskar and Tongaonkar antigenicity scale (table 4). The average antigenic propensity score of the protein was 1.044, with a minimum of 0.900 and a maximum of 1.175. The threshold value for antigenic determination of the peptide was 1.0, where all epitopes with values >1.0 were potential antigenic determinants. We found that 9 epitopes satisfied the threshold value set prior to the analysis, and they had the potential to express the B-cell response. The results are summarized in Table 4 and Figure 2.
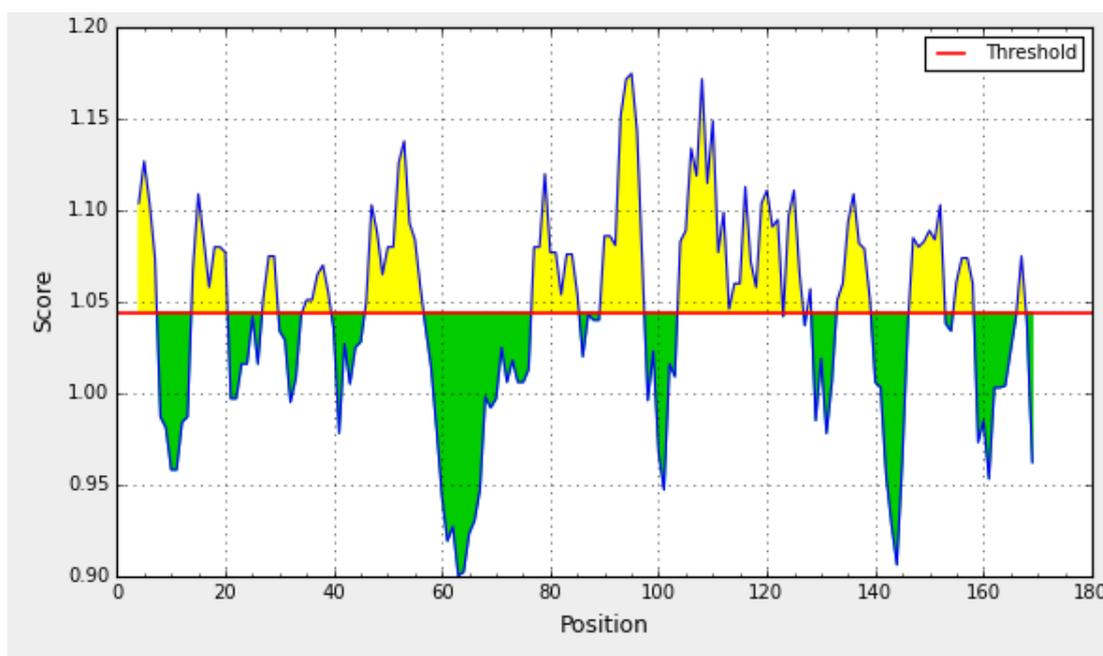
Average: 1.044   Minimum: 0.900   Maximum: 1.175

Fig. 2: Kolaskar and Tongaonkar Antigenicity Prediction of the Main Conserved Sequences.

Notes: The X- and Y-axes represent the sequence position and antigenic propensity score, respectively. The threshold value is 1.0. The regions above the threshold are antigenic, shown in yellow.

3.3. Structure analysis and protein variability determination

Homology model of the conserved region was obtained by the SWISS MODEL, which is shown in Figure 3. PROCHECK server validated the stereochemical quality of the model through Ramachandran Plot (Figure 4).

3.4. Epitope conservancy and allergenicity analysis

IEDB conservancy analysis tool analyzed the epitope conservancy of the proposed epitopes that all for combined peptides conservancy were 100%. AlgPred predicted the allergenicity of the epitopes based on amino acid composition. The prediction score of AlgPred for the two epitopes in combination was 0.60310896 at threshold of –0.4.
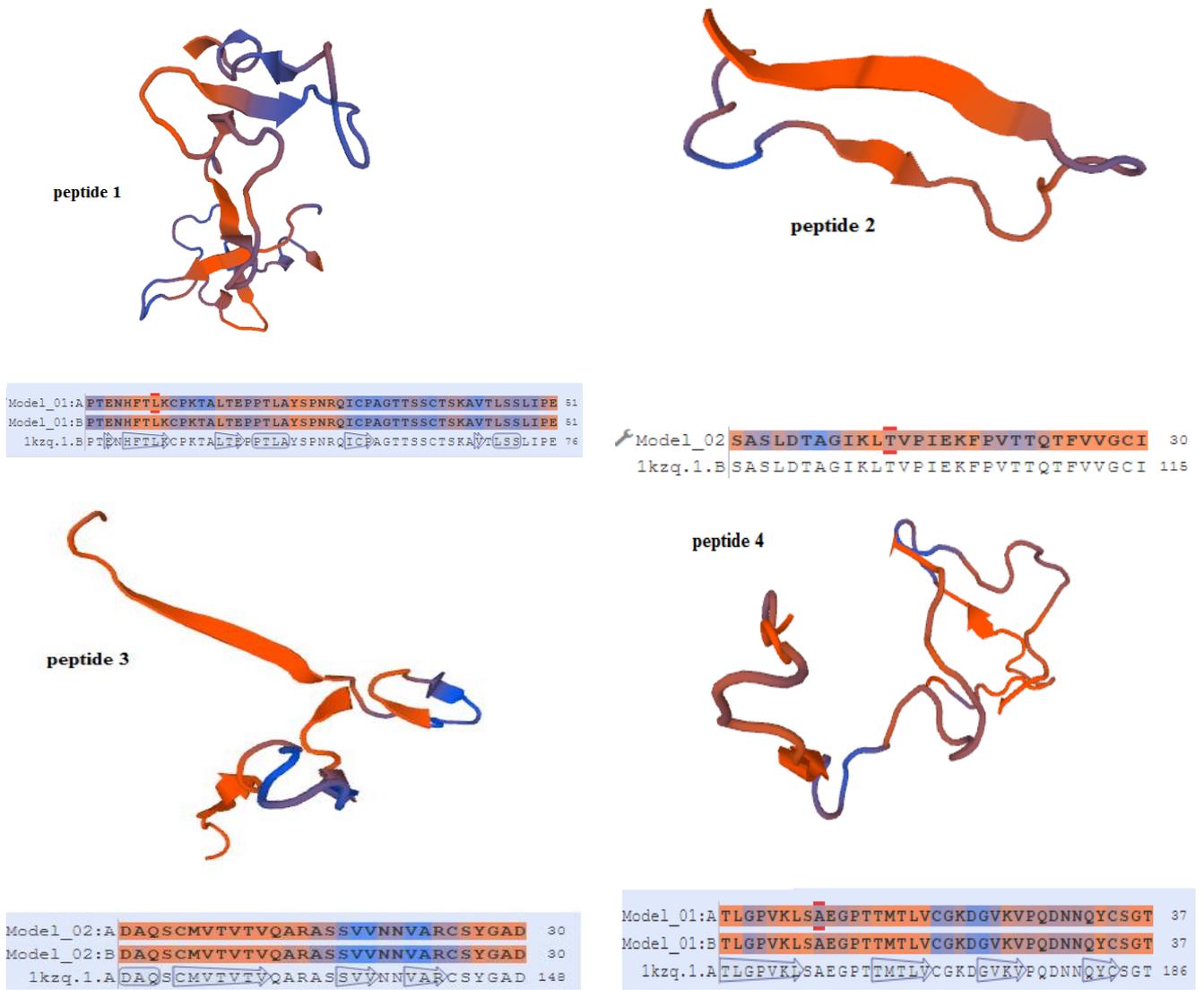
peptide 1

peptide 2

```
Model_01:A PTENHFTLKCPKTALTEPPTLAYSPNRQICPAGTTSSCTSKAVTLSSLIPE 51
Model_01:B PTENHFTLKCPKTALTEPPTLAYSPNRQICPAGTTSSCTSKAVTLSSLIPE 51
1kzq.1.B PTENHFTLKCPKTALTEPPTLAYSPNRQICPAGTTSSCTSKAVTLSSLIPE 76
```

```
Model_02 SASLDTAGIKLTVPIEKFPVTTQTFVVGCI 30
1kzq.1.B SASLDTAGIKLTVPIEKFPVTTQTFVVGCI 115
```

peptide 4

peptide 3

```
Model_02:A DAQSCMVTVTVQARASSVVNNVARCSYGAD 30
Model_02:B DAQSCMVTVTVQARASSVVNNVARCSYGAD 30
1kzq.1.A DAQSCMVTVTVQARASSVVNNVARCSYGAD 148
```

```
Model_01:A TLGPVKLSAEGPTTMTLVCGKDGVKVPQDNNQYCSGT 37
Model_01:B TLGPVKLSAEGPTTMTLVCGKDGVKVPQDNNQYCSGT 37
1kzq.1.A TLGPVKLSAEGPTTMTLVCGKDGVKVPQDNNQYCSGT 186
```

Fig. 3: Three-Dimensional Structure Prediction and Validation.

Notes: Three-dimensional model of four combined epitopes including the predicted epitopes for 4 peptides including T-cell and B-cell epitopes.
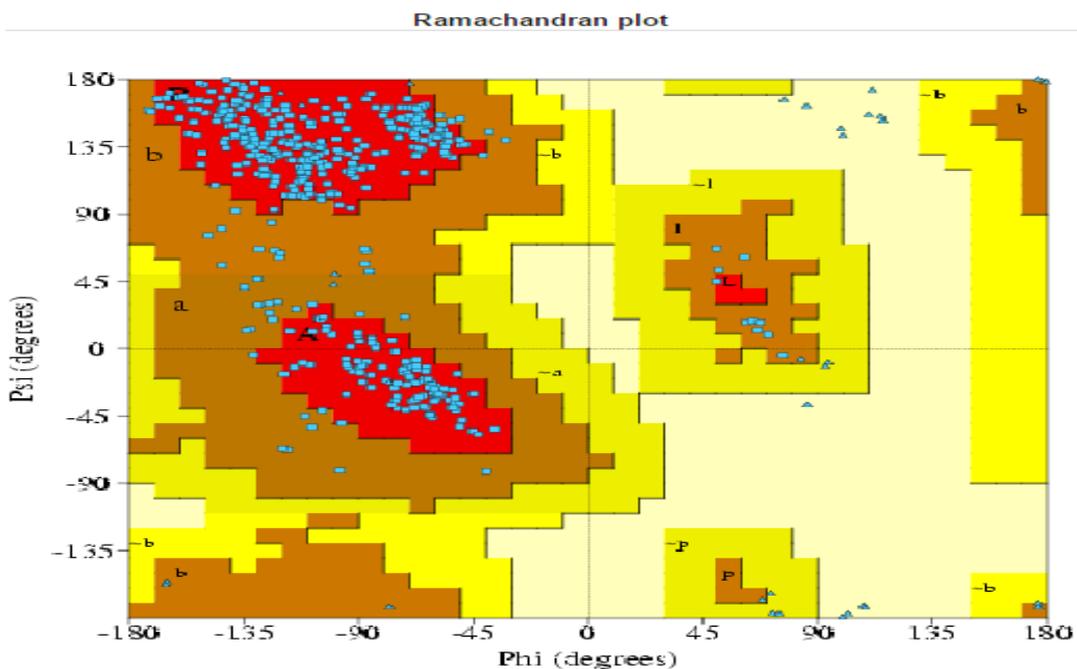

Fig. 4: Ramachandran Plot of the Predicted Model Shows that Most of the Residues are in the Allowed Region of the Plot, Proving the Validity of the Model. 87.6% Amino Acid Residues were Found within the Favored Region.

4. Discussion

Toxoplasmosis, caused by the protozoan parasite Toxoplasma gondii, is one of the most common parasitic infections of man and other warm-blooded animals(Hill & Dubey 2002). The ability of an epitope-based vaccine to stimulate an effective specific immune response with a minute structure and without any unexpected side effects has made it a good choice for vaccine development(Shrestha & Diamond 2004). In this present study, we started with the preferable target, the Surface antigen protein Fragment, to design a peptide vaccine against it. The alignment of the sequences are shown in figure1. SAG1, is considered an important antigen for the development of effective diagnostic tests or subunit vaccines(Hill & Dubey 2002). Antigenicity of the conserved peptides indicated their ability to provoke potential immune response and they were used for further analysis involving T-cell epitope prediction. Through the analysis of the output of both NetCTL and CTLPred, and B-cell epitope prediction, the position of all the epitopes were found and combined in four sequences. In order to see the exact location of the proposed epitope of the protein (shown in Figure 3), the three-dimensional structure of the conserved peptides were modeled. These predicted models were validated with Ramachandran Plot (Figure 4). The most important feature of an epitope enabling its use as a vaccine is its conservancy. Conservancy analysis of the proposed epitopes found 100% conservancy among all the available sequences. Another important criterion of the peptide vaccine is its allergenicity(McKeever et al. 2004); our proposed epitope was examined in silico and found to be nonallergenic in nature. As our proposed epitope is shown to be 100% conserved among examined toxoplasma surface antigen proteins, we suggest that it will be a good possible candidate for vaccine designing. Epitope-based vaccine designing is now becoming more popular and already has been established for rhinovirus(Lapelosa et al. 2009), dengue virus(Chakraborty et al. 2010), human corona virus(Oany et al. 2014), and some others. This type of work has also been proven in vitro(Khan et al. 2014). So, considering the above analysis, we predict that our proposed epitopes would also trigger an immune response in vitro.

5. Conclusion

The achieved results from this study denote that integrated computational approaches are very much effective for designing vaccine candidates against some deadly parasites such as Toxoplasma gondii, with the formally delineated experimental procedure. Therefore, computational studies save both time and cost for researchers and can lead the wet laboratory work with higher possibilities of getting the desired outcome.

References

[1] Apweiler R, Bairoch A, Wu C H, Barker W C, Boeckmann B, Ferro S, Gasteiger E, Huang H, Lopez R, Magrane M (2004) UniProt: the universal protein knowledgebase. Nucleic Acids Res. 32: D115-D119. http://dx.doi.org/10.1093/nar/gkh131.

[2] Berdoy M, Webster J P, Macdonald D (2000) fatal attraction in rats infected with Toxoplasma gondii. Proceedings of the Royal Society of London B: Biological Sciences. 267: 1591-1594. http://dx.doi.org/10.1098/rspb.2000.1182.

[3] Bhasin M, Raghava G (2004) Prediction of CTL epitopes using QM, SVM and ANN techniques. Vaccine. 22: 3195-3204. http://dx.doi.org/10.1016/j.vaccine.2004.02.005.

[4] Bui H-H, Sidney J, Li W, Fusseder N, Sette A (2007) Development of an epitope conservancy analysis tool to facilitate the design of epitope-based diagnostics and vaccines. BMC Bioinformatics. 8: 361. http://dx.doi.org/10.1186/1471-2105-8-361.

[5] Chakraborty S, Chakravorty R, Ahmed M, Rahman A, Waise T Z, Hassan F, Rahman M, Shamsuzzaman S (2010) A computational approach for identification of epitopes in dengue virus envelope protein: a step towards designing a universal dengue vaccine targeting endemic regions. In Silico Biol. 10: 235-246.

[6] Clustal W (1994) improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice Thompson, Julie D.; Higgins, Desmond G.; Gibson, Toby J. Nucleic Acids Res. 22: 4673-4680. http://dx.doi.org/10.1093/nar/22.22.4673.

[7] Doytchinova I A, Flower D R (2007) VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. BMC Bioinformatics. 8: 4. http://dx.doi.org/10.1186/1471-2105-8-4.

[8] Elmore S A, Jones J L, Conrad P A, Patton S, Lindsay D S, Dubey J (2010) Toxoplasma gondii: epidemiology, feline clinical aspects, and prevention. Trends in parasitology. 26: 190-196. http://dx.doi.org/10.1016/j.pt.2010.01.009.

[9] Flegr J, Prandota J, Sovičková M, Israili Z H (2014) Toxoplasmosis—a global threat. Correlation of latent toxoplasmosis with specific disease burden in a set of 88 countries. PLoS One. 9: e90203. http://dx.doi.org/10.1371/journal.pone.0090203.

[10] Hill D, Dubey J (2002) Toxoplasma gondii: transmission, diagnosis and prevention. Clin. Microbiol. Infect. 8: 634-640. http://dx.doi.org/10.1046/j.1469-0691.2002.00485.x.

[11] Khan M K, Zaman S, Chakraborty S, Chakravorty R, Alam M M, Bhuiyan T R, Rahman M J, Fernández C, Qadri F, Seraj Z I (2014) in silico predicted mycobacterial epitope elicits in vitro T-cell responses. Mol. Immunol. 61: 16-22. http://dx.doi.org/10.1016/j.molimm.2014.04.009.

[12] Kolaskar A, Tongaonkar P C (1990) A semi-empirical method for prediction of antigenic determinants on protein antigens. FEBS Lett. 276: 172-174. http://dx.doi.org/10.1016/0014-5793(90)80535-Q.

[13] Lapelosa M, Gallicchio E, Arnold G F, Arnold E, Levy R M (2009) in silico vaccine design based on molecular simulations of rhinovirus chimeras presenting HIV-1 gp41 epitopes. J. Mol. Biol. 385: 675-691. http://dx.doi.org/10.1016/j.jmb.2008.10.089.

[14] Larsen M V, Lundegaard C, Lamberth K, Buus S, Lund O, Nielsen M (2007) Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction. BMC Bioinformatics. 8: 424. http://dx.doi.org/10.1186/1471-2105-8-424.

[15] Laskowski R A, Rullmann J A C, MacArthur M W, Kaptein R, Thornton J M (1996) AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. J. Biomol. NMR. 8: 477-486. http://dx.doi.org/10.1007/BF00228148.

[16] McKeever T M, Lewis S A, Smith C, Hubbard R (2004) Vaccination and allergic disease: a birth cohort study. Am. J. Public Health. 94: 985-989. http://dx.doi.org/10.2105/ajph.94.6.985.

[17] Nair D T, Singh K, Siddiqui Z, Nayak B P, Rao K V, Salunke D M (2002) Epitope recognition by diverse antibodies suggests conformational convergence in an antibody response. The Journal of Immunology. 168: 2371-2382. http://dx.doi.org/10.4049/jimmunol.168.5.2371.

[18] Oany A R, Emran A-A, Jyoti T P (2014) Design of an epitope-based peptide vaccine against spike protein of human coronavirus: an in silico approach. Drug Des. Devel. Ther. 8: 1139. http://dx.doi.org/10.2147/dddt.s67861.

[19] Saha S, Raghava G (2006) AlgPred: prediction of allergenic proteins and mapping of IgE epitopes. Nucleic Acids Res. 34: W202-W209. http://dx.doi.org/10.1093/nar/gkl343.

[20] Šali A, Potterton L, Yuan F, van Vlijmen H, Karplus M (1995) Evaluation of comparative protein modeling by MODELLER. Proteins: Structure, Function, and Bioinformatics. 23: 318-326. http://dx.doi.org/10.1002/prot.340230306.

[21] Shrestha B, Diamond M S (2004) Role of CD8+ T cells in control of West Nile virus infection. J. Virol. 78: 8312-8321. http://dx.doi.org/10.1128/JVI.78.15.8312-8321.2004.

[22] Tenter A M, Heckeroth A R, Weiss L M (2000) Toxoplasma gondii: from animals to humans. Int. J. Parasitol. 30: 1217-1258. http://dx.doi.org/10.1016/S0020-7519(00)00124-7.