Original Article

Asian Pacific Journal of Tropical Biomedicine

journal homepage: www.apjtb.org

doi: 10.4103/2221−1691.269526 Impact factor:1.59

# Prediction of T cell and B cell epitopes of the 22−, 47−, 56−, and 58−kDa proteins of *Orientia tsutsugamushi*

Li−Na Niu[1,5#], Ting−Ting Fu[1#], Man−Ling Chen[3#], Yu−Ying Dong[1#], Jin−Chun Tu[4], Zi−Hao Wang[1], Si−Qi Wang[1], Xuan Zhao[1], Nai−Xu Hou[1], Qian Chen[2✉], Qiang Wu[1,6✉]

[1]*Laboratory of Tropical Biomedicine and Biotechnology, School of Tropical Medicine and Laboratory Medicine, Hainan Medical University, Haikou Hainan 571199, China*

[2]*Department of Laboratory, the First Affiliated Hospital of Hainan Medical University, Haikou Hainan 570102, China*

[3]*Department of Gynecology, the First Affiliated Hospital of Hainan Medical University, Haikou Hainan 570102, China*

[4]*State Key Laboratory of Marine Resource Utilization in South China Sea, College of Material science and Engineering, Hainan University, Haikou Hainan 570228, China*

[5]*Department of Pathogen Biology, School of Basic Medicine and Life Science, Hainan Medical University; Hainan Medical University−University of Hong Kong Joint Laboratory of Tropical Infectious Diseases, Haikou Hainan 571199, China*

[6]*Key Laboratory of Emergency and Trauma of Ministry of Education & Research Unit of Island Emergency Medicine of Chinese Academy of Medical Sciences, Hainan Medical University, Haikou Hainan 571199, China*

ARTICLE INFO

*Article history*:
Received 13 July 2019
Revision 23 July 2019
Accepted 7 October 2019
Available online 24 October 2019

*Keywords*:
*Orientia tsutsugamushi*
B cell epitopes
T cell epitopes
Bioinformatic prediction

ABSTRACT

**Objective:** To predict B cell and T cell epitopes of 22-kDa, 47-kDa, 56-kDa and 58-kDa proteins.

**Methods:** The sequences of 22-kDa, 47-kDa, 56-kDa and 58-kDa proteins which were derived from *Orientia tsutsugamushi* were analyzed by SOPMA, DNAstar, Bcepred, ABCpred, NetMHC, NetMHC Ⅱ and IEDB. The 58-kDa tertiary structure model was built by MODELLER9.17.

**Results:** The 22-kDa B-cell epitopes were located at positions 194-200, 20-26 and 143-154, whereas the T-cell epitopes were located at positions 154-174, 95-107, 17-25 and 57-65. The 47-kDa protein B-cell epitopes were at positions 413-434, 150-161 and 283-322, whereas the T-cell epitopes were located at positions 129-147, 259-267, 412-420 and 80-88. The 56-kDa protein B-cell epitopes were at positions 167-173, 410-419 and 101-108, whereas the T-cell epitopes were located at positions 88-104, 429-439, 232-240 and 194-202. The 58-kDa protein B-cell epitopes were at positions 312-317, 540-548 and 35-55, whereas the T-cell epitopes were located at positions 415-434, 66-84 and 214-230.

**Conclusions:** We identified candidate epitopes of 22-kDa, 47-kDa, 56-kDa and 58-kDa proteins from *Orientia tsutsugamushi*. In the case of 58-kDa, the dominant antigen is displayed on tertiary structure by homology modeling. Our findings will help target additional recombinant antigens with strong specificity, high sensitivity, and stable expression and will aid in their isolation and purification.

## 1. Introduction

Scrub typhus, whose main features include high fever, superficial shower, swelling, eschar and rash, is widely distributed in the Asia-Pacific region, and is a common cause of tropical epidemic disease. Due to the lack of specific early diagnostic methods, effective

#These authors contributed equally to this work.
✉Corresponding author: Qian Chen, Department of Laboratory, the First Affiliated Hospital of Hainan Medical University, Haikou Hainan 570102, China.
Tel: +86 13637645277
E-mail: 251100736@qq.com
Qiang Wu, Laboratory of Tropical Biomedicine and Biotechnology, School of Tropical Medicine and Laboratory Medicine, Hainan Medical University, Haikou Hainan 571199, China.
Tel: 86-898-66984599
Fax: 86-898-66984599
E-mail: wuqiang001001@aliyun.com



**How to cite this article:** Niu LN, Fu TT, Chen ML, Dong YY, Tu JC, Wang ZH, et al. Prediction of T cell and B cell epitopes of the 22-, 47-, 56-, and 58-kDa proteins of *Orientia tsutsugamushi*. Asian Pac J Trop Biomed 2019; 9(10): 443-448.

treatment is often delayed[1-4].

*Orientia tsutsugamushi* (*O. tsutsugamushi*) autolyzes rapidly in the environment; therefore, it is difficult to preserve, culture and purify. These limitations restrict the development of serological diagnosis of tsutsugamushi disease. Recently, the emergence of recombinant antigens has created new opportunities for serological diagnosis. An effective method for *O. tsutsugamushi* is to clone structural genes of its immunodominant antigens and express them in *Escherichia coli* to obtain numerous inexpensive recombinant antigens. Purified antigens can be used for immunological diagnosis[5-7]. Several antigenic proteins of *O. tsutsugamushi* 22-, 47-, 56-, and 58-kDa are the most favorable candidates for diagnosis because they are readily recognized by the host immune system and render significant immunity[8,9].

To obtain fusion antigens, bioinformatics software has been used to analyze and predict antigenic epitopes of these proteins. Most studies use a single bioinformatics software platform for predictive analysis; however, their results have limitations[10-12]. To obtain a more comprehensive and usable set of antigenic epitopes, seven different bioinformatics software programs were used to predict B-cell and T-cell epitopes of the 22-, 47-, 56-, and 58-kDa proteins of *O. tsutusgamushi*. The results of these epitope predictions provide targets for synthesis of recombinant antigens of *O. tsutsugamushi*.

## 2. Materials and methods

### 2.1. Sequence retrieval

Amino acid sequences of the 22-, 47-, 56-, and 58-kDa proteins were selected from GenBank (GenBank accession numbers M63076.1, L31934.1, AY956315.1, and M31887.1, respectively).

### 2.2. Prediction of secondary structure

Secondary structures of proteins were predicted using the SOPMA secondary structure prediction method (http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_sopma.html)[13,14]. Four conformational states (helices, sheets, turns, and coils) were analyzed. Default values were used for other parameters.

### 2.3. Prediction of physicochemical properties

Physicochemical properties of the 22-, 47-, 56-, and 58-kDa proteins were predicted using the protean module of DNAstar[15,16]. Amino acid sequences of proteins were input along with four properties (hydrophilicity, surface probability, antigenicity, and flexible regions). The hydrophilicity, flexibility, surface probability, and antigen indices of *O. tsutsugamushi* proteins were analyzed according to the Kyte-Doolittle, Karplus-Schulz, Emini, and Jameson-Wolf methods, respectively.

### 2.4. Prediction of B cell epitopes

B-cell epitopes were predicted using Bcepred (http://www.imtech.res.in/raghava/bcepred/bcepred_submission.html)[17,18], and ABCpred (http://www.imtech.res.in/raghava/abcpred/) online software[19,20]. Selection of length parameter was 16 amino acids for Bcepred and ABCpred.

### 2.5. Prediction of T cell epitopes

T-cell epitopes were predicted using the Immune Epitope Database (IEDB; http://tools.immuneepitope.org/main/index.html)[21,22], NetMHC 4.0 server (http://www.cbs.dtu.dk/services/NetMHC/)[23,24], and NetMHC Ⅱ 2.3 server (http://www.cbs.dtu.dk/services/NetMHC Ⅱ/)[24,25]. Parameter selection length for MHC-Ⅰ was 9 amino acids, and 15 amino acids for MHC-Ⅱ.

For MHC-Ⅰ, MHC alleles were set at HLA-A02*01; HLA-A24*02; HLA-B*40:01; HLA-B*58:01. For MHC-Ⅱ, MHC-Ⅱ alleles were selected at HLA-DRB1*01:01; HLA-DRB1*04:01; HLA-DRB1*07:01; HLA-DRB1*09:01.

### 2.6. Epitope display in tertiary structure

The homologous template of 58-kDa proteins was selected by BLAST server (http://ncbi.nlm.nih.gov/). The crystal structure of template was retrieved from PDB for subsequent homologous modeling. According to the BLAST results, protein models were generated by Modeller9.17 and evaluated by Discrete Optimized Protein Energy (DOPE) to select the most suitable model structure. Stereochemical quality of the selected model was evaluated by SAVES (https://servicesn.mbi.ucla.edu/SAVES/) server. In SAVES, ERRAT, PROCHECK and Verify 3D module were selected for evaluation.

## 3. Results

### 3.1. Prediction of the secondary structure

Using set conditions (default), the SOMPA online software predicted secondary structures for four kinds of the target protein. β-turn and random coil regions of these structures were typical binding sites for antibodies. When these structural elements are present on protein surfaces, a protein is more likely to be an antigenic epitope. Predicted secondary structures of target proteins are shown in Figure 1.

### 3.2. Prediction of physicochemical properties

Predicted physicochemical properties of the four proteins were analyzed using the DNAStar Protean software (Figure 2). Regions with a value >1 based on the Emini method were screened out, and those regions whose values were >1 based on the Kyte-Doolittle and Jameson-Wolf methods were selected, and the region based on the Karplus-Schulz method were retained squares. These physicochemical properties were further screened for prediction of B-cell epitopes.

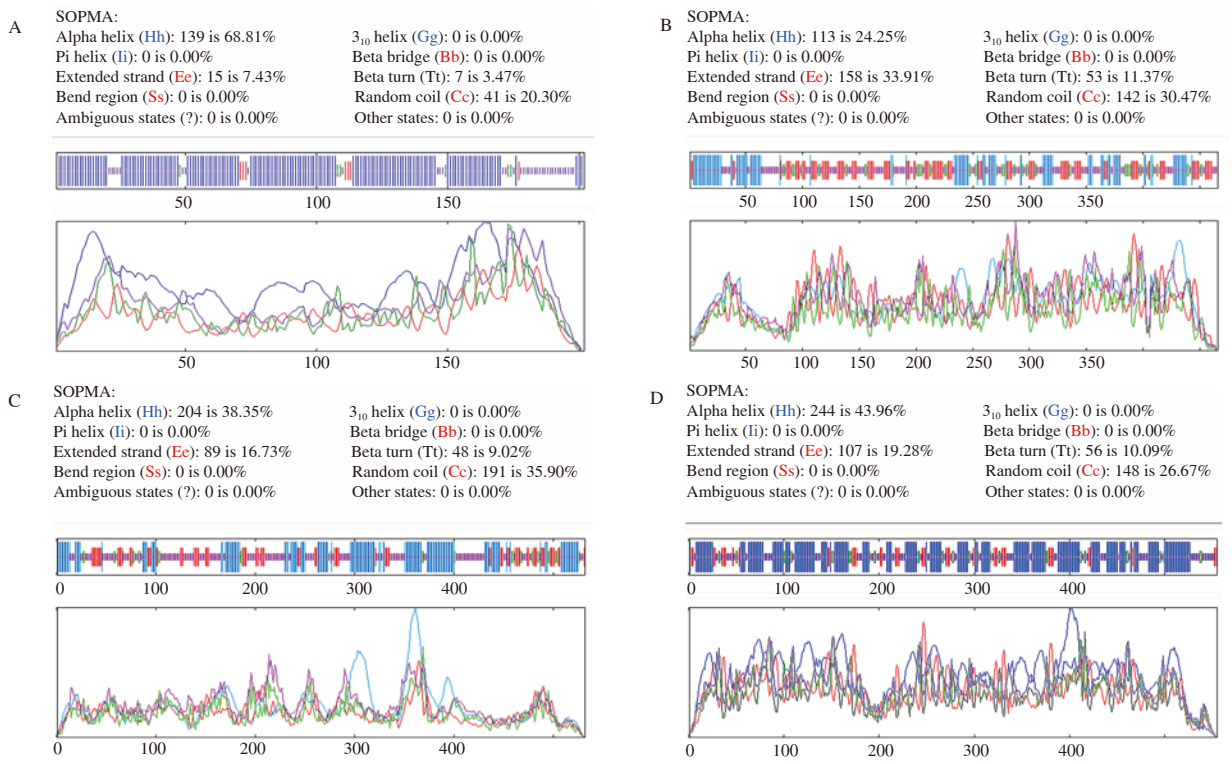


**Figure 1.** Secondary structure of target proteins using SOPMA. A: 22-kDa; B: 47-kDa; C: 56-kDa; D: 58-kDa.

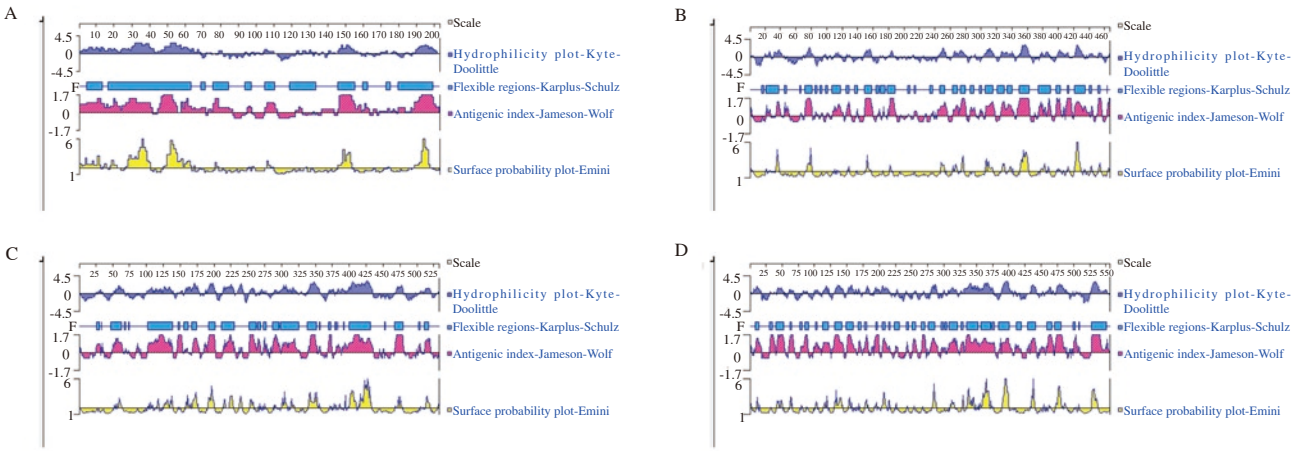


**Figure 2.** Physiochemical property predictions for four target proteins using DNAstar. Hydrophilicity plot in blue; flexible regions in cyan; antigenic index in magenta; surface probability plot in yellow. A: 22-kDa; B: 47-kDa; C: 56-kDa; D: 58-kDa.

### *3.3. Prediction of B cell epitopes*

For linear antigen epitopes of the four target proteins, prediction analysis was carried out using the ABCpred and Bcepred network servers. From these analyses, common predicted antigenic epitopes were selected. The number of amino acids in an epitope should be between 6 and 20. To identify reliable candidate antigen epitopes for subsequent experiments, antigenic regions with a greater number of amino acid residues were filtered out. For the screening process, a scoring method that considered the secondary structure, hydrophilicity, flexibility, surface probability, and antigenic index of the protein was used. A score of 1 was assigned to the antigen epitope region that was predicted from both B cell antigen prediction Programs. Moreover, it also scores 1 for favorable predictions of hydrophilicity, flexibility, surface probability, or antigenic index. Subsequently, the ratio of the number of amino acid residues in potential antigen epitopes to the total number of amino acid residues in the entire region was calculated. Comprehensive analysis of the four target proteins calculates the average score of the above three aspects. Finally, the Top 10 high average was divided into dominant antigen epitopes (Supplementary Table 1). As a result, a series of candidate epitopes were obtained, and region information are shown in Table 1.

**Table 1.** Key regions of the 22-kDa, 47-kDa, 56-kDa and 58-kDa proteins of B-cell epitopes.

| Name | Position | Epitopes region | Score |
|---|---|---|---|
| 22 kDa | 194-200 | QQKDSSI | 0.88 |
| | 20-26 | SQNSSIS | 0.84 |
| | 143-154 | VKHFSSPRDKIK | 0.71 |
| | 28-37 | EQRSQLEKEK | 0.67 |
| | 40-49 | LQGQIGDITG | 0.65 |
| | 61-87 | KLKEWMLKIKDFLISDDFSKLVDSAVK | 0.54 |
| | 108-114 | EKGIMGV | 0.54 |
| | 117-130 | GIQTVTSGFQNITQ | 0.45 |
| 47 kDa | 413-434 | PRDIILSVKRDDNKKDISVKTL | 0.70 |
| | 150-161 | DSNQSRVGDQVI | 0.65 |
| | 283-322 | TEVIKEGSAAQCGIAPGDVITKFHDKEIKTG RDLQVAVSS | 0.64 |
| | 163-174 | GSPFGLRGTVTN | 0.61 |
| | 447-459 | FFTVQRGDRMLYI | 0.61 |
| | 31-45 | LLPQQKSDMHINVNS | 0.58 |
| | 357-368 | QSNDQSLVVNGV | 0.58 |
| 56 kDa | 167-173 | PQLNDEQ | 0.71 |
| | 410-419 | EGDCKQQQGT | 0.68 |
| | 101-108 | QVEEGKVK | 0.67 |
| | 217-225 | NPVGNPPQ | 0.66 |
| | 110-116 | DSVGETK | 0.65 |
| | 187-199 | GIDYRVKNPNDPN | 0.62 |
| | 466-472 | YTSGKID | 0.61 |
| | 304-324 | MQELNDLLEELRESFDGYLGG | 0.57 |
| | 148-161 | RDFGIDIPNIPQQQ | 0.56 |
| 58 kDa | 312-317 | NDTSKL | 0.75 |
| | 540-548 | GGVGGGHHG | 0.75 |
| | 35-55 | RCVAIEQSYGPPKITKDGVSV | 0.63 |
| | 129-141 | DVRKNSSPVKNEE | 0.63 |
| | 148-154 | TVSSNGD | 0.63 |
| | 323-338 | VIVTKDHTTIVHDKNN | 0.62 |
| | 473-478 | SKSTDK | 0.61 |
| | 241-253 | HTGKPLVLIADDV | 0.59 |
| | 169-184 | GQEGVITVEDSKNFNF | 0.59 |

### *3.4. Prediction of T cell epitopes*

Analysis of the four proteins for MHC-Ⅰ T-cell epitopes used the IEDB and NetMHC online prediction softwares. For the IEDB software, peptide percentile rank ≤ 1.0 was considered as having high affinity. For the NetMHC software, peptides with a rank threshold for strong-binding peptides of 0.5 were considered as having high affinity. Combining prediction results of the two antigen prediction software programs identified the dominant epitopes (Supplementary Table 2).

The 22-kDa, 47-kDa, 56-kDa and 58-kDa proteins for MHC-Ⅱ T-cell epitopes were analyzed by IEDB and NetMHC Ⅱ online prediction software. IEDB selection criteria were the same as for MHC-Ⅰ. For NetMHC Ⅱ, a binding threshold of 50.00 was used to define high affinity. The antigen epitopes with higher frequency are selected as dominant epitopes (Supplementary Table 2). As a result, a series of candidate epitopes were obtained, and region information are shown in Table 2.

**Table 2.** Key regions of the 22-kDa, 47-kDa, 56-kDa and 58-kDa proteins of T-cell epitopes.

| Name | Position | Epitopes region | Sort |
|---|---|---|---|
| 22 kDa | 154-174 | KEALGAEGLAKLQAASAGLQN | 1 |
| | 95-107 | VSTEMMQAFTGMK | 2 |
| | 17-25 | KSASQNSSI | 3 |
| | 57-65 | TTMNKLKEW | 4 |
| 47 kDa | 129-147 | KINIALLKINSPAALSYAT | 1 |
| | 259-267 | MLNELTPEL | 2 |
| | 412-420 | RPRDIILSV | 2 |
| | 80-88 | QEVFLGSGV | 4 |
| | 210-233 | FNLEGKIIGINSIHVSYSGISFAI | 5 |
| | 6-25 | YLHLIVFALQGISNVHSKSL | 6 |
| | 165-181 | SPFGLRGTVTNGIISSK | 7 |
| | 338-346 | KSMTLKCKI | 8 |
| | 45-53 | SLSDIVEPL | 9 |
| | 308-316 | KEIKTGRDL | 10 |
| | 456-464 | MLYIALPNI | 10 |
| | 69-77 | ISFNNKVSK | 12 |
| 56 kDa | 88-104 | AEIGVMYLTNITAQVEE | 1 |
| | 429-439 | KEAEFDLSMIV | 2 |
| | 232-240 | FAIHNHEQW | 3 |
| | 194-202 | NPNDPNGPM | 4 |
| | 1-17 | MKKIMLIASAMSALSLP | 5 |
| | 504-513 | GSYMYSFSKI | 6 |
| | 383-391 | KLQRHAGIK | 7 |
| | 125-133 | APIRKRFKL | 8 |
| | 357-365 | QEAVAAAAV | 9 |
| | 211-219 | IPQGNPNPV | 10 |
| | 456-464 | IYAGVGAGL | 10 |
| | 327-335 | FANQIQLNF | 12 |
| | 64-72 | LSLTNGLPF | 13 |
| | 157-165 | IPQQQAQAA | 14 |
| 58 kDa | 415-434 | VPGGGVALFYASRVLDSLKF | 1 |
| | 66-84 | LNVGAQFVISVASKTADVA | 2 |
| | 214-230 | FENPYILLLDQKVSTVQ | 3 |
| | 514-528 | VASLVIATSAMITDH | 4 |
| | 14-22 | KIIEGINVV | 5 |
| | 31-39 | GPKGRCVAI | 6 |

### *3.5. Epitope display in tertiary structure*

Represented by 58-kDa, homologous modeling template was selected from NCBI blast. There were 95% query coverage and 54.8% sequence percent identity between 58-kDa protein and the protein in PDB database. The proteins were generated using MODELLER9.17. Model with the lowest DOPE score and GA341 score was selected for further analysis. Quality assessment of 58-kDa, protein modeling structure was reasonable. In ERRAT, the overall quality factor of the structure was 96.5%. In PROCHECK, most of the residues (94.5%) in Ramachandra plot were located in the most favorable region, 4.6% in other permissible region, 0.6% in generally permissible region, and 0.2% in non-permissible region. In Verify 3D, the compatibility of 3D and 1D structures was 94.7%. The 58-kDa protein tertiary structure epitope was displayed in Figure 3.

According to the predicted results of B cell and T cell epitopes, we summarize the regions ranked earlier as follows. In the B cell epitopes, the 22-kDa protein region is 194-200, 20-26 and 143-154, the 47-kDa protein region is 413-434, 150-161 and 283-322, the 56-kDa protein region is 167-173, 410-419 and 101-108, the 58-kDa

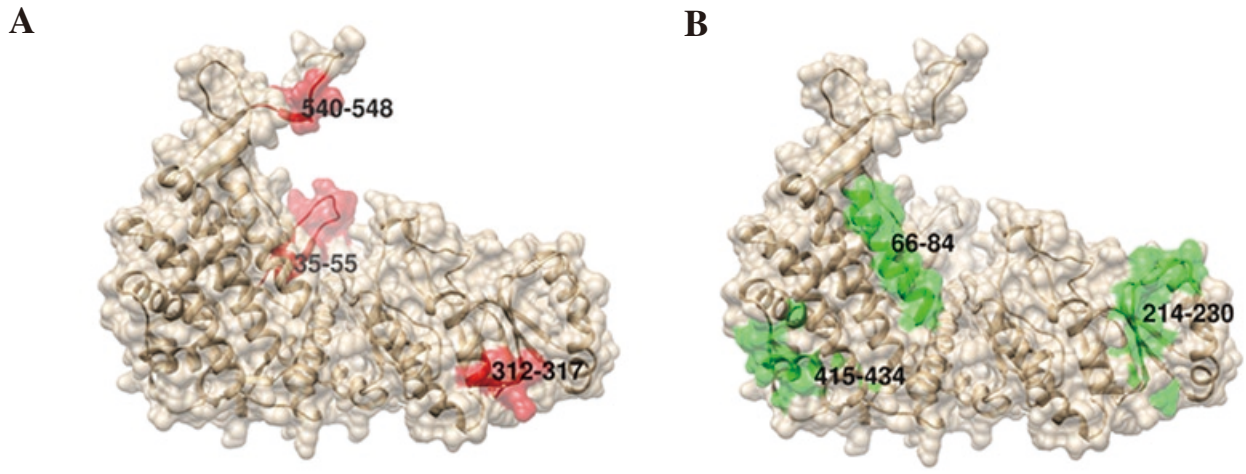


**Figure 3.** Epitope display in 58-kDa protein tertiary structure. A: predicted B cell epitopes; B: predicted T cell epitopes. Red: B cell epitopes, Green: T cell epitopes.

protein region is 312-317, 540-548 and 35-55. In the T cell epitopes, the 22-kDa protein region is 154-174, 95-107, 17-25 and 57-65, the 47-kDa protein region is 129-147, 259-267, 412-420 and 80-88, the 56-kDa protein region is 88-104, 429-439, 232-240 and 194-202, the 58-kDa protein region is 415-434, 66-84 and 214-230.

## 4. Discussion

*O. tsutsugamushi* is a pathogen causing scrub typhus, and taxonomically belongs to the oriental family of *Rickettsia tsutsugamushi*. *O. tsutsugamushi* is the only species in this genus, which comprises seven strains: Kato, Gilliam, Karp, Yonchon, Shimokoshi, Kawasaki, and Kuroki. The Karp strain is dominant in tropic areas[26]. Therefore, proteins of the Karp strain were selected for this study. Protein antigen epitope studies using traditional experimental methods are time-consuming and laborious. Bioinformatics methods will reduce the number of epitopes and focus on the epitopes which are most likely to be antigenic. In our study, seven prediction software programs that use a range of prediction principles were used to predict the B and T epitopes of four target proteins, respectively.

For prediction of B cell epitopes, the ABCpred server with an artificial neural network and the Bcepred server with amino acid pair antigenicity scale were used[27]. For T-cell epitope prediction, IEDB was used to obtain a consensus based on NN-align, SMM-align, and a combinatorial peptide library. NetMHC 4.0 and NetMHC Ⅱ 2.3 servers based on artificial neuron networks were also used. In addition, the physicochemical properties of B cell epitopes, including hydrophilicity, surface probability, antigenicity, flexible regions, and secondary structure were evaluated. Results from the above analyses were scored to establish the reliability of predicted antigenic epitopes.

This study focused on identifying adjacent or overlapping regions of B cell and T cell epitopes. Hickman found that the 81-100 region of the 47-kDa protein sequence elicits an antigenic response[28]. Chen found that the *C*-terminal region 333-430 can elicit a protective immune response[10]. The two regions are also included in the present study, although the former region rank is lower than others due to less frequent appearances and low epitope scores. Seong found that the three regions of the 56-kDa protein, 19-113, 142-203, and 243-328, were strongly antigenic, and Choi found that the 393-432 region of the sequence also showed an antigenic response[29,30]. The present study is consistent with these two studies. Overlapping areas might serve as a focus for future research using the 56-kDa recombinant antigens. With regard to the antigenic epitopes of the 22-kDa and 58-kDa proteins, no reports have been documented thus far. The antigenic epitopes identified in this study may assist the subsequent study of these two proteins.

According to homology modeling, the three-dimensional structure of 58-kDa is obtained, and the accuracy of our prediction can be further explained by the display of the epitope on the tertiary structure. Govindaraj D *et al* used the same method to predict B and T cell epitopes and three-dimensional structure of Per a 10 Allergen of *Periplaneta americana*, followed by *in vitro* validation[31]. The results of *in vitro* experiments showed certain correctness. This further illustrates the high accuracy of the prediction method.

In summary, bioinformatics methods are used to obtain detailed predictions of the four epitopes of *O. tsutsugamushi*, the Karp strain. This study provides experimental data for the identification and screening of epitopes.

## Conflict of interest statement

Authors declare that there are no competing interests.

## Funding

This work was supported by the Finance Science and Technology Project of Hainan Province (ZDYF2018106, ZDXM2014069), the National Natural Science Foundation of China (81860373, 51762012, 81760376, 81460306 and 31160030), the Education

Department of Hainan Province (Hnky2019ZD-27), the National Innovation and Entrepreneurship Training Program for College Students (201511810007, 201811810024), the Innovation and Entrepreneurship Training Program for College Students of Hainan Province (S201911810034) and Innovation and Entrepreneurship Training Program for College Students of Hainan Medical University (HYCX2014013, HYCX2018024), Research Unit of Island Emergency Medicine of Chinese Academy of Medical Sciences (2019RU013).

## References

[1] Wu YC, Qian Q, Soares Magalhaes RJ, Han ZH, Hu WB, Haque U, et al. Spatiotemporal dynamics of scrub typhus transmission in mainland china, 2006-2014. *Plos Negl Trop Dis* 2016; **10**(8): e0004875

[2] Varghese GM, Janardhanan J, Mahajan SK, David T, Paul T, John AJP, et al. Molecular epidemiology and genetic diversity of *Orientia tsutsugamushi*, from patients with scrub typhus in 3 regions of India. *Emerg Infect Dis* 2015; **21**(1): 64-69.

[3] Wei Y, Ma Y, Luo L, Wu X, Huang Y, Li X, et al. Differences in clinical and laboratory features for different genotypes of *Orientia tsutsugamushi* in Guangzhou, Southern China. *Vector-Borne Zoonot* 2017; **17**(4): 260-267.

[4] Ana B, Yoel L, Newton PN, Crump JA, Paris DH, Janet F. Estimating the burden of scrub typhus: A systematic review. *Plos Negl Trop Dis* 2017; **11**(9): e0005838.

[5] Chao CC, Zhang Z, Tatyana B, Wilawan T, Wiwit T, Saowaluk S, et al. An ELISA assay using a combination of recombinant proteins from multiple strains of *Orientia tsutsugamushi* offers an accurate diagnosis for scrub typhus. *BMC Bioinform* 2017; **17**(1): 413-423.

[6] Choi S, Jeong HJ, Ju YR, Gill B, Hwang KJ, Lee J. Protective immunity of 56-kDa type-specific antigen of *Orientia tsutsugamushi* causing scrub Ttphus. *J Microb Biot* 2014; **24**(12): 1728-1735.

[7] Le VN, Laroche M, Thi PHL, Viet NL, Mediannikov O, Raoult D, et al. Use of eschar swabbing for the molecular diagnosis and genotyping of *Orientia tsutsugamushi* causing scrub typhus in Quang Nam province, Vietnam. *Plos Negl Trop Dis* 2017; **11**(2): e0005397.

[8] Wu J, Wang SL, Jin YM, Ma Y, Zeng XJ, Jia PB, et al. Detection of *groEL* gene in spotted fever group rickettsia of patients with fever in Hainan province. *Mod Prev Med* 2014; **41**(13): 2417-2419.

[9] Park JW, Kim SH, Park DW, Jung SH, Park HJ, Seo MH, et al. Molecular epidemiology of an *Orientia tsutsugamushi* gene encoding a 56-kDa type-specific antigen in chiggers, small mammals, and patients from the Southwest Region of Korea. *Am J Trop Med Hyg* 2018; **98**(2): 616-624.

[10]Chen HW, Zhang Z, Huber E, Chao CC, Wang H, Dasch GA, et al. Identification of cross-reactive epitopes on the conserved 47-kilodalton antigen of *Orientia tsutsugamushi* and human serine protease. *Infect Immun* 2009; **77**(6): 2311-2319.

[11]Ramaiah A, Koralur MC, Dasch GA. Complexity of type-specific 56 kDa antigen CD4 T-cell epitopes of *Orientia tsutsugamushi* strains causing scrub typhus in India. *Plos One* 2018; **13**(4): e0196240.

[12]Rattanaphone P, Sonthayanon P, Pukrittayakamee S, Pairs DH, Newton PN, Feil EJ, et al. The diversity and geographical structure of *Orientia tsutsugamushi* strains from scrub typhus patients in Laos. *Plos Negl Trop Dis* 2015; **9**(8): e0004024.

[13]Geourjon C, Deléage G. SOPMA: Significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. *Cabios* 1995; **11**(6): 681-684.

[14]Narmadha G, Yenugu S. Molecular modeling of the human sperm associated antigen 11 B (SPAG11B) proteins. *Sys Biol Reprod Med* 2015; **61**(2): 78-88.

[15]Burland TG. DNASTAR's Lasergene sequence analysis software. *Methods Mol Biol* 2000; **132**: 71-91.

[16]Jing LI, Liu JH, Qiang FU, Song H, Fan B, Hu Y, et al. Prediction and identification of linear B-cell epitopes in gB glycoprotein of equine herpesvirus type 1. *China Anim Husb Vet Med* 2017; **44**(5): 1484-1490.

[17]Saha S, Raghava GPS. BcePred: Prediction of continuous B-cell epitopes in antigenic sequences using physico-chemical properties. *Lect Notes Comput Sci* 2004; **3239**: 197-204.

[18]Zhang J, Cui Y, Wu Y, Wang H, Ke J. Prediction and identification of B-cell epitopes for tumor necrosis factor- α . *Mol Med Rep* 2017; **16**(3): 3439-3444.

[19]Saha S, Raghava GP. Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. *Proteins* 2010; **65**(1): 40-48.

[20]Madrigal M, Alape-Girón A, Barboza-Arguedas E, Aquilar-Ulloa W, Flores-Diaz M. Identification of B cell recognized linear epitopes in a snake venom serine proteinase from the central American bushmaster, *Lachesis stenophrys*. *Toxicon* 2017; **15**(140): 72-82.

[21]Bui HH, Sidney J, Dinh K, Southwood S, Newman MJ, Sette A. Predicting population coverage of T-cell epitope-based diagnostics and vaccines. *BMC Bioinform* 2006; **7**(1): 153-160.

[22]Shi J, Sun J, Wu M, Hu N, Li J, Li Y, et al. Inferring protective CD8+ T-cell epitopes for NS5 protein of four serotypes of dengue virus Chinese isolates based on HLA-A, -B and -C allelic distribution: Implications for epitope-based universal vaccine design. *Plos One* 2015; **10**(9): e0138729.

[23]Andreatta M, Nielsen M. Gapped sequence alignment using artificial neural networks: Application to the MHC class I system. *Bioinformatics* 2016; **32**(4): 511-517.

[24]Cecilia RR, Bourdin E, García MI, Veaute CMI. *In silico*, prediction of T- and B-cell epitopes in PmpD: First step towards to the design of a *Chlamydia trachomatis* vaccine. *Biomed J* 2018; **41**(2): 109-117.

[25]Nielsen M, Lundegaard C, Lund O. Prediction of MHC class II binding affinity using SMM-align, a novel stabilization matrix alignment method. *BMC Bioinform* 2007; **8**(1): 238-240.

[26]Paris DH, Dumler J. Stephen. State of the art of diagnosis of rickettsial diseases. *Curr Opin Infect Dis* 2016; **29**(5): 433-439.

[27]Bayrami V, Keyhanfar M, Mohabatkar H, Mahdavi M, Moreau V. *In silico* prediction of B cell epitopes of the extracellular domain of insulin-like growth factor-1 receptor. *Mol Biol Res Commun* 2016; **5**(4): 201-214.

[28]Hickman CJ, Stover CK, Joseph SW, Oaks EV. Murine T-cell response to native and recombinant protein antigens of *Rickettsia tsutsugamushi*. *Infect Immun* 1993; **61**(5): 1674-1681.

[29]Seong SY, Park SG, Huh SH, Jang WJ, Kim HR, Han TH, et al. Mapping of antigenic determinant regions of the Bor56 protein of *Orientia tsutsugamushi*. *Infect Immun* 1997; **65**(12): 5250-5256.

[30]Choi MS, Seong SY, Kang JS, Kim YW, Huh SH, Kim IS. Homotypic and heterotypicantibody responses to a 56-kilodalton protein of *Orientia tsutsugamushi*. *Infect Immun* 1999; **67**(11): 6194-6197.

[31]Govindaraj D, Sharma S, Gaur SN, Lavasa S, Prasad N, Arora N. Immunogenic peptides: B & T Cell Epitopes of Per a 10 Allergen of *Periplaneta americana*. *Mol Immunol* 2016; **80**: 24-32.

Supplementary Table 1. Top 10 result of linear antigenic prediction of the 22-kDa, 47-kDa, 56-kDa and 58-kDa proteins.

| Name | pos | peptide | score | Length | Name | pos | peptide | score | Length |
|---|---|---|---|---|---|---|---|---|---|
| 22kDa | 194-200 | QQKDSSI | 0.88 | 7 | 56kDa | 167-173 | PQLNDEQ | 0.71 | 7 |
| | 20-26 | SQNSSIS | 0.84 | 7 | | 410-419 | EGDCKQQQGT | 0.68 | 10 |
| | 143-154 | VKHFSSPRDKIK | 0.71 | 12 | | 101-108 | QVEEGKVK | 0.67 | 8 |
| | 28-37 | EQRSQLEKEK | 0.67 | 10 | | 217-225 | NPVGNPPQ | 0.66 | 8 |
| | 40-49 | LQGQIGDITG | 0.65 | 10 | | 110-116 | DSVGETK | 0.65 | 7 |
| | 66-81 | MLKIKDFLISDDFSKL | 0.58 | 16 | | 187-199 | GIDYRVKNPNDPN | 0.62 | 13 |
| | 61-72 | KLKEWMLKIKDF | 0.55 | 12 | | 466-472 | YTSGKID | 0.61 | 7 |
| | 108-114 | EKGIMGV | 0.54 | 7 | | 304-319 | MQELNDLLEELRESFD | 0.57 | 16 |
| | 79-87 | SKLVDSAVK | 0.48 | 9 | | 148-161 | RDFGIDIPNIPQQQ | 0.56 | 14 |
| | 117-130 | GIQTVTSGFQNITQ | 0.45 | 14 | | 315-324 | RESFDGYLGG | 0.56 | 10 |
| 47kDa | 419-434 | SVKRDDNKKDISVKTL | 0.72 | 16 | 58kDa | 312-317 | NDTSKL | 0.75 | 6 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 297-312 | APGDVITKFHDKEIKT | 0.70 | 16 | 540-548 | GGVGGGHHG | 0.75 | 9 |
| 413-428 | PRDIILSVKRDDNKKD | 0.67 | 16 | 45-55 | PPKITKDGVSV | 0.64 | 11 |
| 150-161 | DSNQSRVGDQVI | 0.65 | 12 | 148-154 | TVSSNGD | 0.63 | 7 |
| 283-298 | TEVIKEGSAAQCGIAP | 0.65 | 16 | 129-141 | DVRKNSSPVKNEE | 0.63 | 13 |
| 163-174 | GSPFGLRGTVTN | 0.61 | 13 | 35-50 | RCVAIEQSYGPPKITK | 0.62 | 16 |
| 447-459 | FFTVQRGDRMLYI | 0.61 | 13 | 323-338 | VIVTKDHTTIVHDKNN | 0.62 | 16 |
| 31-45 | LLPQQKSDMHINVNS | 0.58 | 15 | 473-478 | SKSTDK | 0.61 | 6 |
| 357-368 | QSNDQSLVVNGV | 0.58 | 12 | 169-184 | GQEGVITVEDSKNFNF | 0.59 | 16 |
| 307-322 | DKEIKTGRDLQVAVSS | 0.57 | 16 | 241-253 | HTGKPLVLIADDV | 0.59 | 13 |

Supplementary Table 2. Analysis of the 22-kDa, 47-kDa, 56-kDa and 58-kDa proteins MHC-I and MHC-II T-cell epitopes using IEDB; NETMHC and NETMHCII online prediction software.

| Name | pos | peptide | MHC-I(Allel) | | | | pos | peptide | MHC-II(Allel) | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | IEDB | rank | NETMHC | rank | | | IEDB | rank | NETMHCII | rank |
| 22kDa | 17 | KSASQNSSI | HLA-B5801 | 0.7 | HLA-B5801 | 0.53 | 158 | GAEGLAKLQAASAGL | HLA-DRB10101 | 0.96 | HLA-DRB10101 | 0.8 |
| | | | | | | | | | | | HLA-DRB10701 | 8 |
| | 57 | TTMNKLKEW | HLA-B5801 | 0.8 | HLA-B5801 | 0.55 | | | | | HLA-DRB10901 | 0.4 |
| | 95 | VSTEMMQAF | HLA-B5801 | 0.25 | HLA-A4601 | 0.54 | 159 | AEGLAKLQAASAGLQ | HLA-DRB10901 | 0.58 | HLA-DRB10101 | 0.5 |
| | 99 | MMQAFTGMK | HLA-A0301 | 0.25 | HLA-A0301 | 0.042 | | | HLA-DRB10101 | 0.96 | HLA-DRB10701 | 8 |
| | | | | | | | | | | | HLA-DRB10901 | 0.2 |
| | 154 | KEALGAEGL | HLA-B4001 | 0.2 | HLA-B4001 | 0.031 | 160 | EGLAKLQAASAGLQN | HLA-DRB10901 | 0.81 | HLA-DRB10101 | 0.9 |
| | | | | | | | | | HLA-DRB10101 | 0.96 | HLA-DRB10701 | 8 |
| | | | | | | | | | | | HLA-DRB10901 | 0.2 |
| 47kDa | 6 | YLHLIVFAL | HLA-A0301 | 0.6 | HLA-A0201 | 0.25 | 11 | VFALQGISNVHSKSL | HLA-DRB10401 | 0.87 | HLA-DRB10101 | 8 |
| | | | | | | | | | HLA-DRB10701 | 0.94 | HLA-DRB10401 | 4 |
| | | | | | | | | | | | HLA-DRB10701 | 15 |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 45 | SLSDIVEPL | HLA-A0201 | 0.5 | HLA-A0201 | 0.07 | 129 | KINIALLKINSPAAL | HLA-DRB10901 | 0.08 | HLA-DRB10101 | 4 |
| | | | | | | | | | | HLA-DRB10401 | 4 |
| | | | | | | | | | | HLA-DRB10901 | 0.5 |
| 69 | ISFNNKVSK | HLA-A0301 | 0.5 | HLA-A0301 | 0.4 | 130 | INIALLKINSPAALS | HLA-DRB10901 | 0.08 | HLA-DRB10101 | 2 |
| | | | | | | | | | | HLA-DRB10401 | 2 |
| | | | | | | | | | | HLA-DRB10901 | 0.3 |
| 80 | QEVFLGSGV | HLA-B4001 | 0.2 | HLA-B4001 | 0.4 | 131 | NIALLKINSPAALSY | HLA-DRB10101 | 0.62 | HLA-DRB10101 | 0.7 |
| | | | | | | | | HLA-DRB10901 | 0.07 | HLA-DRB10401 | 8 |
| | | | | | | | | | | HLA-DRB10901 | 2 |
| 139 | SPAALSYAT | HLA-B0702 | 0.7 | HLA-B0702 | 0.4 | | | | | HLA-DRB11501 | 0.15 |
| 165 | SPFGLRGTV | HLA-B0702 | 0.5 | HLA-B0702 | 0.12 | 132 | IALLKINSPAALSYA | HLA-DRB10901 | 0.07 | HLA-DRB10101 | 0.7 |
| | | | | | | | | | | HLA-DRB10401 | 4 |
| | | | | | | | | | | HLA-DRB10901 | 0.15 |
| 173 | VTNGIISSK | HLA-A0301 | 0.35 | HLA-A0301 | 0.2 | 210 | FNLEGKIIGINSIHV | HLA-DRB11501 | 0.9 | HLA-DRB10101 | 16 |
| | | | | | | | | | | HLA-DRB10701 | 4 |
| | | | | | | | | | | HLA-DRB11501 | 4 |
| 225 | SYSGISFAI | HLA-A2402 | 0.3 | HLA-A2402 | 0.04 | 211 | NLEGKIIGINSIHVS | HLA-DRB11501 | 0.9 | HLA-DRB10101 | 16 |
| | | | | | | | | | | HLA-DRB10701 | 8 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | | HLA-DRB11501 | 4 |
| | 259 | MLNELTPEL | HLA-A0201 | 0.3 | HLA-A0201 | 0.03 | 212 | LEGKIIGINSIHVSY | HLA-DRB11501 | 0.9 | HLA-DRB10101 | 4 |
| | | | | | | | | | | | HLA-DRB10701 | 8 |
| | | | | | | | | | | | HLA-DRB11501 | 2 |
| | 308 | KEIKTGRDL | HLA-B4001 | 0.55 | HLA-B4001 | 0.15 | 213 | EGKIIGINSIHVSYS | HLA-DRB11501 | 0.9 | HLA-DRB10101 | 4 |
| | | | | | | | | | | | HLA-DRB10701 | 2 |
| | | | | | | | | | | | HLA-DRB11501 | 8 |
| | 338 | KSMTLKCKI | HLA-B5801 | 0.3 | HLA-B5801 | 0.15 | 214 | GKIIGINSIHVSYSG | HLA-DRB11501 | 0.9 | HLA-DRB10101 | 4 |
| | | | | | | | | | | | HLA-DRB10701 | 4 |
| | | | | | | | | | | | HLA-DRB11501 | 16 |
| | 412 | RPRDIILSV | HLA-B0702 | 0.3 | HLA-B0702 | 0.03 | 215 | KIIGINSIHVSYSGI | HLA-DRB11501 | 0.9 | HLA-DRB10101 | 8 |
| | | | | | | | | | | | HLA-DRB10701 | 4 |
| | | | | | | | | | | | HLA-DRB11501 | 16 |
| | 456 | MLYIALPNI | HLA-A0201 | 0.6 | HLA-A0201 | 0.1 | 216 | IIGINSIHVSYSGIS | HLA-DRB11501 | 0.9 | HLA-DRB10101 | 16 |
| | | | | | | | | | | | HLA-DRB10701 | 4 |
| | | | | | | | | | | | HLA-DRB11501 | 16 |
| 56kDa | 5 | MLIASAMSA | HLA-A0201 | 1 | HLA-A0201 | 0.25 | 1 | MKKIMLIASAMSALS | HLA-DRB10901 | 0.13 | HLA-DRB10101 | 0.7 |
| | | | | | | | | | | | HLA-DRB10401 | 1.25 |
| | 64 | LSLTNGLPF | HLA-B5801 | 0.8 | HLA-B5801 | 0.25 | | | | | HLA-DRB10701 | 8 |
| | | | | | | | | | | | HLA-DRB11501 | 4 |
| | 94 | YLTNITAQV | HLA-A0201 | 0.4 | HLA-A0201 | 0.04 | 2 | KKIMLIASAMSALSL | HLA-DRB10901 | 0.14 | HLA-DRB10101 | 0.4 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 125 | APIRKRFKL | HLA-B0702 | 0.5 | HLA-B0702 | 0.08 | | | | | HLA-DRB10401 | 1.1 |
| | | | | | | | | | | | HLA-DRB10701 | 4 |
| | 157 | IPQQQAQAA | HLA-B0702 | 0.9 | HLA-B0702 | 0.5 | | | | | HLA-DRB10901 | 1.5 |
| | | | | | | | | | | | HLA-DRB11501 | 4 |
| | 194 | NPNDPNGPM | HLA-B0702 | 0.3 | HLA-B0702 | 0.08 | 3 | KIMLIASAMSALSLP | HLA-DRB10901 | 0.23 | HLA-DRB10101 | 1.25 |
| | | | | | | | | | | | HLA-DRB10401 | 2 |
| | 211 | IPQGNPNPV | HLA-B0702 | 0.6 | HLA-B0702 | 0.3 | | | | | HLA-DRB10701 | 8 |
| | | | | | | | | | | | HLA-DRB10901 | 1.5 |
| | 232 | FAIHNHEQW | HLA-B5801 | 0.2 | HLA-B5801 | 0.03 | | | | | HLA-DRB11501 | 4 |
| | 327 | FANQIQLNF | HLA-B5801 | 0.7 | HLA-B5801 | 0.25 | 88 | AEIGVMYLTNITAQV | HLA-DRB10901 | 0.42 | HLA-DRB10101 | 4 |
| | | | | | | | | | | | HLA-DRB10401 | 4 |
| | 357 | QEAVAAAAV | HLA-B4001 | 0.65 | HLA-B4001 | 0.15 | | | | | HLA-DRB11501 | 4 |
| | 383 | KLQRHAGIK | HLA-A0301 | 0.35 | HLA-A0301 | 0.15 | 89 | EIGVMYLTNITAQVE | HLA-DRB10401 | 0.77 | HLA-DRB10101 | 2 |
| | | | | | | | | | | | HLA-DRB10401 | 4 |
| | 429 | KEAEFDLSM | HLA-B1301 | 0.4 | HLA-B4001 | 0.2 | | | | | HLA-DRB10901 | 4 |
| | 431 | AEFDLSMIV | HLA-B4001 | 0.45 | HLA-B4001 | 0.07 | | | | | HLA-DRB11501 | 4 |
| | | | HLA-B1301 | 0.4 | | | | | | | | |
| | 456 | IYAGVGAGL | HLA-A2402 | 0.5 | HLA-A2402 | 0.4 | 90 | IGVMYLTNITAQVEE | HLA-DRB10401 | 0.77 | HLA-DRB10101 | 4 |
| | | | | | | | | | | | HLA-DRB10401 | 4 |
| | 504 | GSYMYSFSK | HLA-A0301 | 0.3 | HLA-A0301 | 0.06 | | | | | HLA-DRB10901 | 1.5 |
| | 505 | SYMYSFSKI | HLA-A2402 | 0.25 | HLA-A2402 | 0.03 | | | | | HLA-DRB11501 | 4 |
| 58kDa | 14 | KIIEGINVV | HLA-A0201 | 1 | HLA-A0201 | 0.17 | 66 | LNVGAQFVISVASKT | HLA-DRB10901 | 0.5 | HLA-DRB10101 | 4 |
| | | | | | | | | | | | HLA-DRB10701 | 4 |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | HLA-DRB10901 | 0.9 |
| | | | | | | 67 | NVGAQFVISVASKTA | HLA-DRB10901 | 0.31 | HLA-DRB10101 | 2 |
| | | | | | | | | | | HLA-DRB10701 | 4 |
| | | | | | | | | | | HLA-DRB10901 | 0.4 |
| | | | | | | 68 | VGAQFVISVASKTAD | HLA-DRB10901 | 0.29 | HLA-DRB10101 | 2 |
| | | | | | | | | | | HLA-DRB10701 | 2 |
| | | | | | | | | | | HLA-DRB10901 | 0.2 |
| | | | | | | 69 | GAQFVISVASKTADV | HLA-DRB10901 | 0.24 | HLA-DRB10101 | 0.9 |
| | | | | | | | | | | HLA-DRB10401 | 4 |
| | | | | | | | | | | HLA-DRB10701 | 1.5 |
| | | | | | | | | | | HLA-DRB10901 | 0.15 |
| | | | | | | 70 | AQFVISVASKTADVA | HLA-DRB10901 | 0.26 | HLA-DRB10101 | 2 |
| | | | | | | | | | | HLA-DRB10701 | 1.5 |
| | | | | | | | | | | HLA-DRB10901 | 0.3 |
| | | | | | | 214 | FENPYILLLDQKVST | HLA-DRB10301 | 0.13 | HLA-DRB10101 | 4 |
| | | | | | | | | | | HLA-DRB10301 | 0.9 |
| | | | | | | | | | | HLA-DRB11501 | 4 |
| | | | | | | 215 | ENPYILLLDQKVSTV | HLA-DRB10301 | 0.13 | HLA-DRB10101 | 1.5 |
| | | | | | | | | | | HLA-DRB10301 | 0.3 |
| | | | | | | | | | | HLA-DRB10401 | 4 |
| | | | | | | | | | | HLA-DRB11501 | 4 |
| 31 | GPKGRCVAI | HLA-B0702 | 0.9 | HLA-B5801 | 0.5 | 216 | NPYILLLDQKVSTVQ | HLA-DRB10301 | 0.1 | HLA-DRB10101 | 4 |
| | | | | | | | | | | HLA-DRB10301 | 0.15 |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | | HLA-DRB10401 | 4 |
| | | | | | HLA-DRB11501 | 4 |
| | 415 | VPGGGVALFYASRVL | HLA-DRB11501 | 0.3 | HLA-DRB10101 | 8 |
| | | | | | HLA-DRB10701 | 8 |
| | | | | | HLA-DRB11501 | 0.5 |
| | 417 | GGGVALFYASRVLDS | HLA-DRB11501 | 0.17 | HLA-DRB10101 | 8 |
| | | | | | HLA-DRB10701 | 8 |
| | | | | | HLA-DRB11501 | 0.3 |
| | 418 | GGVALFYASRVLDSL | HLA-DRB11501 | 0.18 | HLA-DRB10101 | 8 |
| | | | | | HLA-DRB10701 | 8 |
| | | | | | HLA-DRB11501 | 0.4 |
| | 419 | GVALFYASRVLDSLK | HLA-DRB11501 | 0.3 | HLA-DRB10101 | 8 |
| | | | | | HLA-DRB10301 | 4 |
| | | | | | HLA-DRB10701 | 16 |
| | | | | | HLA-DRB11501 | 0.8 |
| | 420 | VALFYASRVLDSLKF | HLA-DRB11501 | 0.92 | HLA-DRB10101 | 8 |
| | | | | | HLA-DRB10301 | 4 |
| | | | | | HLA-DRB10701 | 8 |
| | | | | | HLA-DRB11501 | 0.25 |
| | 514 | VASLVIATSAMITDH | HLA-DRB10301 | 0.78 | HLA-DRB10101 | 4 |
| | | | | | HLA-DRB10701 | 8 |
| | | | | | HLA-DRB10901 | 4 |