*Bulgarian Chemical Communications, Volume 52, Issue 1 (pp. 114-120) 2020 DOI: 10.34049/bcc.52.1.5185*

# Determination of an effective immunogenic peptide against *Acinetobacter baumannii*

E. Valipour*

*Department of Molecular Biology and Genetics, Faculty of Arts and Sciences, Zonguldak Bulent Ecevit University, Zonguldak, Turkey*

Received July 15, 2019; Accepted October 7, 2019

Nosocomial infections caused by *Acinetobacter baumannii* are a growing clinical problem. *A. baumannii* biofilm-associated protein (Bap) is necessary for mature biofilm formation on medically relevant surfaces and pathogenicity. In this research, conserved regions of the Bap were determined using various programs. The epitope rich segments of the conserved regions were detected. Two fragments (151AA and 172AA) were selected as peptide fragments that were rich in the aspect of leaner and structural B-cell epitopes, CD4 T-cell epitopes and MHC binding sites. The two peptides were linked to each other by a linker peptide and were named as AC peptide. The codon usage and GC contents of the nucleotide sequences coding the AC peptide were optimized. MHC-II binding predictions were done by consensus method that showed the average of immunogenicity score, combined score, and median percentile to be 92.63, 46.18 and 15.22, respectively. Antigenic peptide prediction by Kolaskar and Tongaonkar method and by VAXIGENE program showed the average antigenic propensity for this peptide as 1.0281 and 0.8656, respectively. This immunogenic peptide can be used as a valuable tool to detect the *A. baumannii* and to defend against nosocomial infection caused by the strain.

**Keywords:** *Acinetobacter baumannii,* immunogenicity, biofilm, nosocomial, design

## INTRODUCTION

*Acinetobacter baumannii* is a non-motile aerobic gram-negative bacillus strain [1]. Nosocomial infections caused by *Acinetobacter baumannii* are serious problems in intensive care units. Urinary tract, soft-tissue, skin and wound infection, secondary meningitis, bloodstream infections are caused by *A. baumannii* [2]. *A. baumannii* strains are perilous because they are resistant to many antibiotics such as trimethoprim-sulfamethoxazole, tetracyclines, betalactams (including carbapenems), fluoroquinolones and aminoglycosides. Therefore, that is a serious problem that needs to be solved promptly by the international health care community [3]. Up to now, biofilm associated protein of the strains, auto-transporter, outer membrane proteins, capsular polysaccharide, and whole cell was used to raise antiserum against *A. baumannii*. All of the antigens showed to be a potential vaccine against the strains. The methods have some disadvantages including protection limitation and high cost. Up to now, there is no vaccine against *A. baumannii*, therefore, there is a serious need for more research to find a better vaccine candidate having the least disadvantage [4, 5]. Yet more research and works are needed to discover all factors involved in epidemicity and virulence of *A. baumannii*. To make infection in a host, the bacterial strain essentially needs to attach to host's epithelial cells and colonize them. *Acinetobacter* species usually adhere and colonize on human mucosal membrane and skin even for some weeks which shows that the adherence of the strain is an essential factor in pathogenecity [6]. In addition, many researchers discovered that the strain with high capability to make biofilm has high ability to adhere to epithelial cells [7]. *A. baumannii* to be able to make a biofilm on epithelial cells requires a biofilm-associated protein (Bap) [8, 9]. Inactivation of the biofilm-associated protein caused a decrease in biofilm formation and a decrease in adherence ability of the strain to human epithelial cells. The antibody against biofilm-associated protein inhibits the strain from binding to the epithelial cells. The proteins are expressed at the cell surface of Bap [10]. Therefore this study aimed to design new and effective antigenic recombinant peptide from Bap against *A. baumannii*.

## MATERIAL AND METHODS

### *Primary sequence analysis*

Sequences related to Bap were selected from clusters in Uniprot database at http://www.uniprot.org. To find the Bap family proteins a cluster with 50% sequence identity levels was selected. Some members of the cluster were subjected to multiple sequence alignments by MUSCLE at www.ebi.ac.uk/Tools/psa/emboss_needle/ using [11]. The alignment was visualized

* To whom all correspondence should be sent:
E-mail: Va.cell@yahoo.com

and edited by Jalview program (http://csbg.cnb.csic.es/PB/E1120) [12]. The program of Compared at http://bioinf4.cs.ucl.ac.uk:3000/dompred was used to annotate disorders and detect homolog domains. InterProScan online software at http://www.ebi.ac.uk/Tools/pfa/iprscan/ was used to detect homology within the PROSITE database [13]. The conserved regions of the protein were subjected to further analysis.

*Antigenicity prediction of the selected region*

T-Cell Epitope was predicted by NetTepi1 program (http://www.cbs.dtu.dk/services/NetTepi/ ) SYFPEITHI program (http://www.syfpeithi.de), Immune Epitope Database and Analysis Resource (IEDB) (http://www.immuneepitope.org) and http://imed.med.ucm.es/Tools/antigenic.pl [14, 15]. DNAStar program (V5.0) (http://www.dnastar.com), SYFPEITHI program (http://www.syfpeithi.de), ProPred (http://www.imtech.res.in/raghava/propred/index. html) and IEDB (http://tools.immuneepitope.org) were used to determine MHCII binding regions [16]. B-Cell Epitopes were predicted by BCPreds (http://ailab.cs.iastate.edu/bcpreds/index.html), ABCpred(http://www.imtech.res.in/raghava/abcpred/), DNAStar (V5.0) (http://www.dnastar.com) and conformational epitopes by Seppa program (http://lifecenter.sgst.cn/seppa), BEpro (http://pepito.proteomics.ics.uci.edu/cgi-bin/BEpro.cgi) and Discotope Server (http://www.cbs.dtu.dk/services/-DiscoTope) [17, 18]. Regions of proteins that are rich in the aspect of linear and conformational epitopes can be used as vaccine candidates. Therefore two segments of the Bap protein were selected to join together to make an immunogenic protein.

*Further scrutiny*

The gene sequence encoding the selected amino acid sequences was retrieved from the NCBI website; GC content and codon usage were evaluated according to *E. coli* expression system by the programs at http://eu.idtdna.com, http://www.genscript.com and http://www.jcat.de/ websites [19]. One of the most important factors for protein production is mRNA structure and stability. Therefore, the structure and stability of mRNA to be transcripted from the gene fragment was analyzed and controlled by http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi [20]. Protein stability and solubility were analyzed by http://web.expasy.org/protscale/ and https://protein-sol.manchester.ac.uk/, respectively. After that, the construct was subjected to antigenic evaluation by the programs mentioned above.

## RESULTS

*Primary sequence analysis*

The proteins having protein ID (UniProt) of G2JKC0, A0A3B0GRS0, B0VEF9, A0A3B0GRS0, UPI000A394F5F, B0LHN4 were selected from a cluster with 50% sequence identity. They are large proteins with tandem repeat modules. Alignment of these proteins revealed the conservation of several locations in sequences, mainly repeat modules. Within these conserved regions, two specific regions were selected based on PSI-BLAST search against a non-redundant protein database.

*Immunogenic regions selection*

Two regions (151AA and 172AA) were selected as peptide fragments that were rich in the aspect of leaner and structural B-cell epitopes, CD4 T-cell epitopes and MHC binding sites. The two peptides were linked to each other by a linker peptide (EAAAKEAAAKEAAAK) (Table 1) and were named AC peptide. The AC peptide coding gene sequence was extracted from *Acinetobacter baumannii* biofilm-associated protein (Bap) (GenBank: EU117203.1) and subjected to codon usage optimization and RNA structure analysis. The distribution of codon usage frequency along the length of the construct is very important. After codon usage optimization the initial CAI value (0.284) of the construct increased to 1.0 and initial GC-content (44.27) changed to 54.27 (Fig. 1).

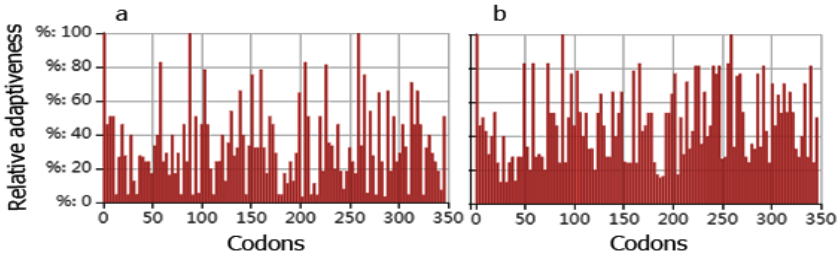



**Fig. 1.** Codon usage optimization. **a** shows the relative adaptiveness of the AC sequence before optimization; **b** represents that after optimization. The codon usage of this sequence was optimized due to the codon usage of *E. coli* k12 strain.

**Table 1**. A: AS peptide sequence that was made from regions of BAP having high antigenicity; B: The gene sequences coding AS peptide.

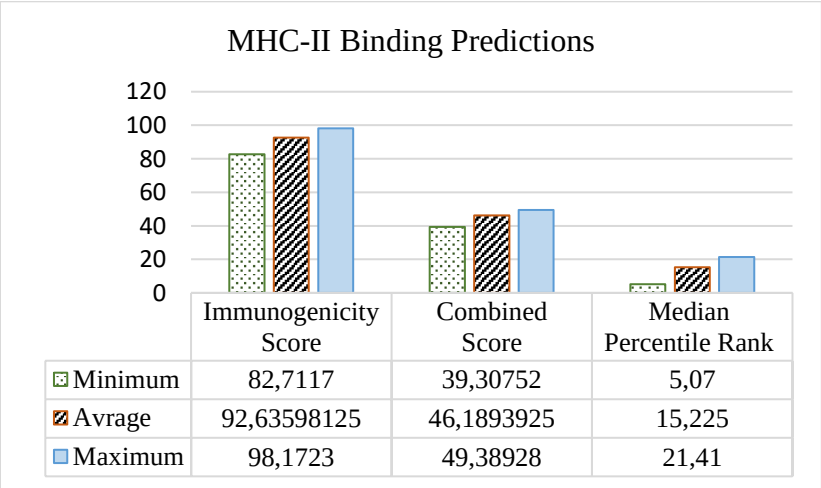
>A

MTATDAAGNVGTDTGVVTVDTAAPNTAGVTFTIDSVTADNVINASEAAGNVTITGVLKNI
PADATNTAVTVVINGVTYNATVDKTAGTWTVSVPGSGLVADADKTIDAKVTFTDAAGNSS
TVNDTQIYTLDTAAPAAPVIDPVNGTDPITGEAAAKEAAAKEAAAKNDAGEVDVVTPTTV
ISEVNGQPVVADGTSITGTYGTLVINLDGSYTYTPTASAAGVGQTDQFTYTLTDPVTGDT
AQANLNIQLSSVKAVDNVVTAEINPEPLLVADDVALGSSTYLAAVSLAGLDLQLLGNDAI
EFTVDPNREGTATFTFDAVITADLLSDYAIVVQKFDEATGQWVSIGGT

>B

ATGACTGCAACAGATGCAGCAGGTAACGTAGGTACAGATACAGGTGTTGTGACAGTTGAT
ACAGCTGCTCCAAATACAGCTGGCGTTACCTTTACGATTGACTCAGTAACAGCTGACAAT
GTGATTAATGCATCAGAAGCAGCAGGTAATGTCACGATTACTGGTGTTCTGAAAAACATT
CCAGCAGATGCGACTAACACGGCCGTTACAGTTGTTATTAATGGGGTAACTTATAATGCA
ACGGTAGATAAAACAGCAGGCACGTGGACAGTAAGTGTGCCGGGTAGTGGTTTGGTTGCT
GATGCAGATAAAACGATTGATGCTAAAGTAACGTTTACAGATGCAGCAGGTAATAGCAGC
ACTGTTAACGATACGCAAATTTATACATTAGACACAGCTGCTCCTGCAGCGCCAGTAATC
GACCCAGTTAACGGGACAGACCCAATTACAGGTGAAGCAGCAGCAAAAGAAGCAGCAGCA
AAAGAAGCAGCAGCAAAAAATGATGCAGGTGAGGTAGATGTTGTAACTCCAACAACAGTT
ATAAGTGAGGTTAATGGTCAGCCTGTTGTAGCAGATGGCACGAGCATCACTGGTACTTAC
GGTACATTAGTGATCAACCTGGATGGTTCATACACTTATACGCCTACTGCTAGTGCGGCA
GGTGTAGGACAAACAGATCAGTTTACCTATACCTTAACTGATCCTGTAACTGGTGATACG
GCTCAAGCAAACCTCAATATTCAATTGAGCTCTGTGAAAGCTGTGGATAATGTTGTAACT
GCAGAAATCAACCCAGAACCGTTGCTAGTTGCAGATGATGTTGCTCTAGGCAGTTCAACT
TACCTTGCAGCAGTATCGTTAGCTGGTCTGGACTTACAATTGCTTGGTAACGACGCAATT
GAGTTTACTGTTGATCCAAACCGCGAAGGTACAGCAACGTTCACATTCGATGCGGTAATC
ACAGCAGATTTACTCAGTGATTATGCGATTGTTGTTCAGAAATTTGATGAAGCAACAGGC
CAATGGGTATCGATCGGTGGTACT

**Table 2**. The MHC-II binding predictions done by the ElliPro program.




| | Immunogenicity Score | Combined Score | Median Percentile Rank |
|---|---|---|---|
| Minimum | 82,7117 | 39,30752 | 5,07 |
| Avrage | 92,63598125 | 46,1893925 | 15,225 |
| Maximum | 98,1723 | 49,38928 | 21,41 |

The free energy of the thermodynamic ensemble of the RNA from the construct was -329.77 kcal/mol. The centroid secondary structure had minimum free energy of -276.70 kcal/mol.

Instability index (II) of the protein was computed to be 8.57 and predicted scaled solubility was 0.761.

The AC peptide was converted to PDB format using the LOMETS program [21]. The model showed Cov Norm value of 0.95 and a Z score of 1.71. The model was also analyzed by I-TASSER that showed Tm-score of 0.914, RMSD value of 1.42, IDEN value of 0.281 and Cov value of 0.950. In this analysis, the protein having PDB accession number of 4p99A was reference protein.

### *Sequence-based epitope prediction of the AC peptide*

VAXIGENE program showed the Overall antigenicity prediction for the target peptide to be 0.8656. In this analysis, the threshold was 0.4. Antigenic Peptide Prediction at http://imed.med.ucm.es/Tools/antigenic.pl showed the average antigenic propensity for this peptide to be 1.0281. MHC-II Binding Predictions were done by consensus methods that showed the average of immunogenicity score, combined score, and median percentile to be 92.63, 46.18 and 15.22 respectively (Table 2). CD4 T cell immunogenicity prediction of the target peptide by Predicting Antigenic Peptides in http://imed.med.ucm.es/Tools/antigenic.pl showed the average antigenic propensity for this protein as 1.0292 (Fig. 2).

The antigenicity was determined using the method of Kolaskar and Tongaonkar [22]. Analysis of Linear B Cell Epitope by BepiPred-2.0 showed an average score of 0.477, a minimum score of -0.007 and a maximum score of 1.803 (Fig. 3).

### *Structure-based b-cell epitope prediction of the AC peptide*

In linear B cell epitopes predicted by ElliPro, the highest score is related to the first 29 amino acids 'TATDAAGNVGTDTGVVTVDTAAPNTAGVT' with a score of 0.848. The region starting from 282 to 320 'QLLGNDAIEFTVDPNREGTATFTFDAVITA-DLLSDYAIV' showed a score of 0.764 and the minimum score (0.583) was related to the region starting from 93 to 105 (PGSGLVADADKTI). Discontinuous B cell epitopes predicted from the 3D structure of a protein by ElliPro are shown in Table 3. The server showed 6 regions that involve in conformational epitopes. The region made from residues (1-29, 54-88 and 112-117) showed the maximum score (0.732) and residues of Q:P137, Q:A162, Q:A163, Q:A164, Q:K165, Q:D167, Q:A168, Q:G169, Q:E170, Q:V171, Q:D172, Q:V173, Q:V174 showed the minimum score (0539).

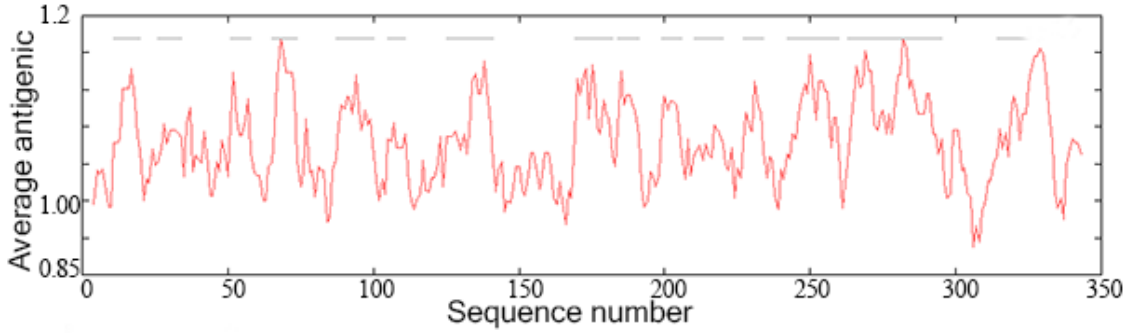



**Fig. 2.** CD4 T cell immunogenicity prediction of the target peptide by the IEDB program that showed a reasonable immunogenicity

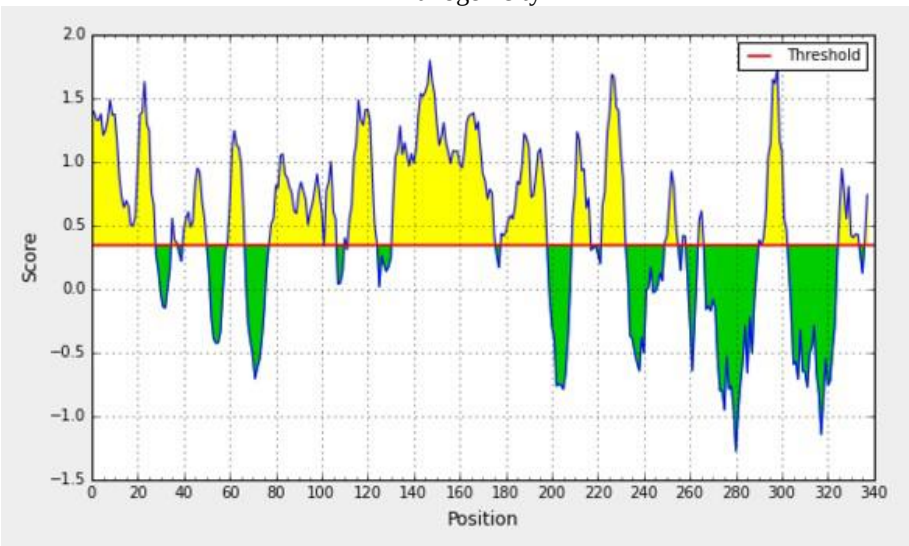
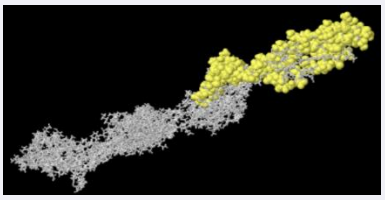
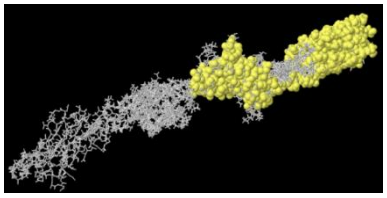
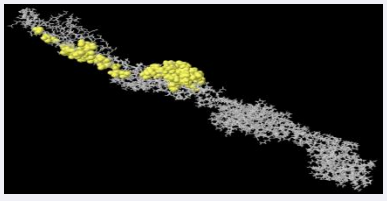
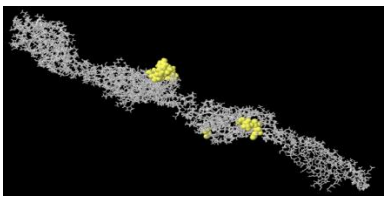


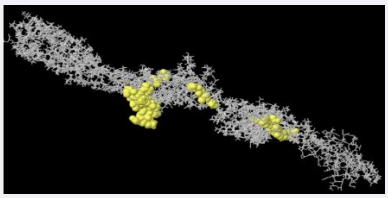
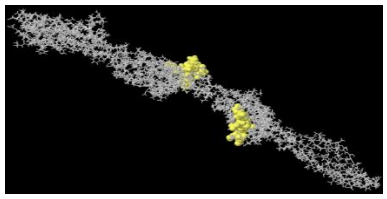

**Fig. 3**. Analysis of Linear B Cell Epitope by BepiPred-2.0 that showed an average score of 0.477, a minimum score of -0.007 and a maximum score of 1.803

## DISCUSSION

The UniRef (UniProt Reference Clusters) provides clustered sets of sequences from the UniProt Knowledge base (UniProtKB) and selected UniProt Archive records to obtain complete coverage of sequence space at several resolutions while hiding redundant sequences [23]. Therefore, in this research the cluster with 50% sequence identity level was used to find the members of Bap family. There is a significant correlation between high protein expression level and codon adaptation index (CAI). If the CAI value is high the expression level is high. The CAI value of 1.0 was considered as an ideal value and the CAI value of >0.8 was considered as a good value for protein expression in a host cell [19]. In addition, the GC present has a significant effect on the expression. The ideal percentage range of GC content is between 30% and 70%. In this research, after codon usage optimization the CAI value (0.284) of the construct increased to 1.0 and GC content (44.27) changed to 54.27. The free energy of the thermodynamic ensemble of the RNA from the construct was -329.77 kcal/mol. The instability index (II) was computed to be 8.57. A protein whose instability index is smaller than 40 is predicted as stable, a value above 40 predicts that the protein may be unstable [24]. Predicted scaled solubility of the protein was 0.761. The scaled solubility value (QuerySol) is the predicted solubility. The population average for the experimental dataset (PopAvrSol) is 0.45, and therefore any scaled solubility value greater than 0.45 is predicted to have a higher solubility than the average soluble *E.coli* proteins and any protein with a lower scaled solubility value is predicted to be less soluble [25]. The centroid secondary structure had minimum free energy of -276.70 kcal/mol. The AC peptide was converted to PDB format using the LOMETS program [21]. The model showed a Cov Norm value of 0.95 and a Z score of 1.71. The root-mean-square deviation (RMSD) value indicates how closely two structures (target protein and reference protein) are related. The value less than 3A° shows that our protein is a novel fusion protein. Also, the RMSD value more than 3A° indicates a lower accuracy and quality of the modeling. In that condition, the TM-score is more significant. Template modeling (TM) score value of more than 0.5 shows a model having almost acceptable topology [26]. Confidence score (C-score) is usually in the range of (-5, 2). The higher value shows the confidence of the model. The C-value of our protein was 0.950 showing high confidence. These results showed that our protein 3D structure was correctly predicted [4]. If a pathogen has sufficient binding domain to a major histocompatibility complex (MHC) it can elicit a high-level response of the immune system [27]. For MHC-II binding predictions, the mean of the area under the curve (AUC) value was usually ranged from 0.697 to 0.813. The consensus method having AUC value of (0.896, - 0.05) is a good method for prediction [28]. In this research MHC-II binding predictions were done by consensus methods. In this part, the consensus method of IEDB was used. The method is a combination of Sturniolo, CombLib, SMM-align, and NN-align. The server is a good tool for prediction of T- helper cell (allele independent) immunogenicity of the population [29]. In this method, the lower percentile value shows a high affinity [29, 30]. CD4 T cell immunogenicity prediction of the target peptide by Predicting Antigenic Peptides in http://imed.med.ucm.es/Tools/antigenic.pl showed the average antigenic propensity for this protein to be 1.0292. This tool uses the method of Tongaonkar and Kolaskar to predict antigenic parts of the peptide [22]. The accuracy of the data reported by this tool is almost 75%. Analysis of Linear B Cell Epitope by BepiPred-2.0 showed an average score of 0.477, a minimum score of -0.007 and a maximum score of 1.803. The BepiPred-2.0 tool uses a random forest algorithm to predict B-cell epitopes from crystal structures [31] In this research, the BepiPred-2.0 was used to predict the B-cell epitopes from antigen sequences. ElliPro predicts linear and discontinuous antibody epitopes based on a protein antigen's 3D structure. The prediction of epitopes by this tool was determined by the PI (Protrusion Index) value. The residue that has a large value of PI is more accessible to solvent. Discontinuous epitopes are defined by PI values and are clustered based on the distance R (in Å) between residue's centers of mass. The larger R is associated with larger discontinuous epitopes being predicted [32].

**Table 3.** Discontinuous B cell epitopes predicted from the 3D structure of the protein by ElliPro.

| No. | Residues | Number of residues | Score | 3D structure |
|---|---|---|---|---|
| 1 | Q:T1, Q:A2, Q:T3, Q:D4, Q:A5, Q:A6, Q:G7, Q:N8, Q:V9, Q:G10, Q:T11, Q:D12, Q:T13, Q:G14, Q:V15, Q:V16, Q:T17, Q:V18, Q:D19, Q:T20, Q:A21, Q:A22, Q:P23, Q:N24, Q:T25, Q:A26, Q:G27, Q:V28, Q:T29, Q:G54, Q:V55, Q:L56, Q:K57, Q:N58, Q:I59, Q:P60, Q:A61, Q:D62, Q:A63, Q:T64, Q:N65, Q:A79, Q:T80, Q:D82, Q:K83, Q:T84, Q:A85, Q:G86, Q:T87, Q:W88, Q:T112, Q:D113, Q:A114, Q:A115, Q:G116, Q:N117 | 56 | 0.738 |  |
| 2 | Q:V244, Q:D245, Q:N246, Q:V247, Q:V248, Q:T249, Q:A250, Q:E251, Q:I252, Q:N253, Q:P254, Q:E255, Q:P256, Q:L257, Q:L258, Q:V259, Q:A260, Q:D261, Q:D262, Q:V263, Q:A264, Q:L265, Q:G266, Q:S267, Q:S268, Q:T269, Q:Y270, Q:L271, Q:A272, Q:A273, Q:V274, Q:S275, Q:L276, Q:A277, Q:G278, Q:Q282, Q:L283, Q:L284, Q:G285, Q:N286, Q:D287, Q:A288, Q:I289, Q:E290, Q:F291, Q:T292, Q:V293, Q:D294, Q:P295, Q:N296, Q:R297, Q:E298, Q:G299, Q:T300, Q:A301, Q:T302, Q:F303, Q:T304, Q:F305, Q:D306, Q:A307, Q:V308, Q:I309, Q:T310, Q:A311, Q:D312, Q:L313, Q:L314, Q:S315, Q:D316, Q:Y317, Q:A318, Q:I319, Q:V320, Q:V321, Q:Q322, Q:G335, Q:G336, Q:T337 | 79 | 0.716 |  |
| 3 | Q:V70, Q:V71, Q:I72, Q:N73, Q:G74, Q:V75, Q:T76, Q:Y77, Q:P93, Q:S95, Q:G96, Q:V98, Q:A99, Q:D100, Q:A101, Q:D102, Q:K103, Q:T104, Q:I105, Q:D106, Q:A107, Q:K108, Q:N122, Q:I126 | 24 | 0.584 |  |
| 4 | Q:T175, Q:G197, Q:T198, Q:Y199, Q:G200, Q:T201 | 6 | 0.573 |  |

| No. | Residues | Number of residues | Score | 3D structure |
|---|---|---|---|---|
| 5 | Q:E182, Q:V183, Q:N184, Q:G185, Q:Q186, Q:P187, Q:V188, Q:V189, Q:A190, Q:G210, Q:G212, Q:Q213, Q:T214, Q:Q237, Q:L238, Q:S239 | 16 | 0.54 |  |
| 6 | Q:P137, Q:A162, Q:A163, Q:A164, Q:K165, Q:D167, Q:A168, Q:G169, Q:E170, Q:V171, Q:D172, Q:V173, Q:V174 | 13 | 0.539 |  |

***Conflict of interest.*** *There is no conflict of interest associated with the present manuscript.*

REFERENCES


1. A. Howard, M. O'Donoghue, A. Feeney, R. D. Sleator, *Virulence*, **3**, 243 (2012).
2. D. G. Moriel, S. A. Beatson, D. J. Wurpel, J. Lipman, G. R. Nimmo, D. L. Paterson, M. A. Schembri, *PLoS ONE*, **8**, e77631 (2013).
3. A. Y. Peleg, H. Seifert, D. L. Paterson, *Clin. Microbiol. Rev.*, **21**, 538 (2008).
4. E. Z. Hosseingholi, I. Rasooli S. L. M. Gargari, *Acta Biotheor.*, **62**, 455 (2014).
5. L. V. Bentancor, A. Routray, C. Bozkurt-Guzel, A. Camacho-Peiro, G. B. Pier, T. Maira-Litrán, *Infect. Immun.*, IAI, 06096 (2012).
6. J. C. Lee, H. Koerten, P. Van den Broek, H. Beekhuizen, R. Wolterbeek, M. Van den Barselaar, T. Van der Reijden, J. Van der Meer, J. Van de Gevel, L. Dijkshoorn, *Res. Microbiol.*, **157**, 360 (2006).
7. H. W. Lee, Y. Koh, J. Kim, J. C. Lee, Y. C. Lee, S. Y. Seol, D. T. Cho, J. Kim, *Clin. Microbiol. Infect.*, **14**, 49 (2008).
8. K. A. Brossard, A. A. Campagnari, *Infect. Immun.*, **80**, 228 (2012).
9. Y. Fattahian, I. Rasooli, S.L.M. Gargari, M.R. Rahbar, S.D.A. Astaneh, J. Amani, *Microb. Pathog.*, **51**, 402 (2011).
10. E. De Gregorio, M. Del Franco, M. Martinucci, E. Roscetto, R. Zarrilli, P. P. Di Nocera, *BMC Genomics*, **16**, 933 (2015).
11. R. C. Edgar, *Nucleic Acids Res.*, **32**, 1792 (2004).
12. A. M. Waterhouse, J. B. Procter, D. M. Martin, M. Clamp, G. J. Barton, *Bioinformatics*, **25**, 1189 (2009).
13. F. Sefid, I. Rasooli, A. Jahangiri and H. Bazmara, *Acta Biotheor.*, **63**, 129 (2015).
14. R. E. Soria-Guerra, R. Nieto-Gomez, D. O. Govea-Alonso, S. Rosales-Mendoza, *J. Biomed. Inform.*, **53**, 405 (2015).
15. Y. Yang, W. Sun, J. Guo, G. Zhao, S. Sun, H. Yu, Y. Guo, J. Li, X. Jin, L. Du, *Human Vaccines & Immunotherapeutics*, **11**, 795 (2015).
16. S. Pahil, N. Taneja, H. R. Ansari, G. Raghava, *PLoS ONE*, **12**, e0180505 (2017).
17. R. L. M. Guedes, C. M. F. Rodrigues, N. Coatnoan, A. Cosson, F. A. Cadioli, H. A. Garcia, A. L. Gerber, R. Z. Machado, P. M. C. Minoprio, M. M. G. Teixeira, *Genomics*, **111**(3),407 (2018).
18. R. C. Russi, E. Bourdin, M. I. García, C. M. I. Veaute, *Biomedical Journal*, **41**, 109 (2018).
19. E. Valipour, M.-L. Moosavi, J. Amani, S. Nazarian, *World J. Microbiol. Biotechnol.*, **30**, 1861 (2014).
20. B. Hamidi, F. Ebrahimi, A. Hajizadeh, H. K. Alikhani, *Eur. J. Exper. Biol.*, **2**, 1154 (2012).
21. S. Wu, Y. Zhang, *Nucleic Acids Res.*, **35**, 3375 (2007).
22. A. Kolaskar, P. C. Tongaonkar, *FEBS Lett.*, **276**, 172 (1990).
23. B. E. Suzek, H. Huang, P. McGarvey, R. Mazumder, C. H. Wu, *Bioinformatics*, **23**, 1282 (2007).
24. E. Gasteiger, C. Hoogland, A. Gattiker, M. R. Wilkins, R. D. Appel, A. Bairoch, in: The Proteomics Protocols Handbook, Springer, 2005, p. 571.
25. T. Niwa, B.-W. Ying, K. Saito, W. Jin, S. Takada, T. Ueda, H. Taguchi, *Proceedings of the National Academy of Sciences*, **106**, 4201 (2009).
26. J. Xu, Y. Zhang, *Bioinformatics*, **26**, 889 (2010).
27. M. N. Davies, D. R. Flower, *Drug Discov. Today*, **12**, 389 (2007).
28. M. Marmion, M. Parviainen, M. Luoto, R. K. Heikkinen, W. Thuiller, *Divers. Distrib.*, **15**, 59 (2009).
29. S. Paul, C. S. L. Arlehamn, T. J. Scriba, M. B. Dillon, C. Oseroff, D. Hinz, D. M. McKinney, S. C. Pro, J. Sidney, B. Peters, *J. Immunol. Methods*, **422**, 28 (2015).
30. S. K. Dhanda, E. Karosiene, L. Edwards, A. Grifoni, S. Paul, M. Andreatta, D. Weiskopf, J. Sidney, M. Nielsen, B. Peters, *Frontiers in Immunology*, **9** (2018).
31. J. Garnier, *J. Mol. Biol.*, **120**, 97 (1978).
32. J. Ponomarenko, H.-H. Bui, W. Li, N. Fusseder, P. E. Bourne, A. Sette, B. Peters, *BMC Bioinformatics*, **9**, 514 (2008).