FUDMA Journal of Sciences (FJS)
ISSN online: 2616-1370
ISSN print: 2645 - 2944
Vol. 8 No. 3, June, (Special Issue) 2024, pp 487 - 500
DOI: https://doi.org/10.33003/fjs-2024-0803-2530

STRUCTURAL PREDICTION AND ANTIGENIC ANALYSIS OF A HYPOTHETICAL *PLASMODIUM FALCIPARUM* PROTEIN USING BIOINFORMATICS TOOLS

***[1,4]Amlabu, E., [1]Collins, M. O., [1,4]Omale, J., [5]Adepoju, O. A., [1]Omeiza, A. S., [2,4]Christian, A. N., [1,4]Osiekafore, A. A., [1,4]Idih, F. M., [1,4]Tijani, M., [1,4]Akoji, T. F., [1,4]Tolulope, V. O., [1,4]Adegoke, A. J. and [3,4]Egu, S. A.**

[1]Department of Biochemistry, Prince Abubakar Audu University, Anyigba, Nigeria.
[2]Department of Microbiology, Prince Abubakar Audu University, Anyigba, Nigeria.
[3]Department of Chemistry, Prince Abubakar Audu University, Anyigba, Nigeria.
[4]Genomics and Molecular Biotechnology Research and Training Laboratory, Prince Abubakar Audu University, Anyigba, Nigeria.
[5]Department of Biochemistry, Ahmadu Bello University, Zaria, Nigeria.

*Corresponding authors' email: amlabu.e@ksu.edu.ng Phone: +2348124622228

ABSTRACT

Malaria is caused by *Plasmodium falciparum* which remains a major global health problem and there is no effective vaccine with broad operational impact. The genome of *P. falciparum* has been sequenced by others indicating 5,600 genes in the genome. Presently, about 60 % of the genes encoding proteins in the parasite have no-known function. Thus, identifying potential drug/vaccine candidates and biomarkers for evaluating malaria transmission intensity such as the uncharacterized protein Q8IJ58 is a crucial step towards effective malaria intervention. Computer-based approach was used to analyze an uncharacterized *Plasmodium falciparum* protein - Q8IJ58 (Gene ID: PF10_0342, PlasmoDB ID: PF3D7_1035100), for its basic and theoretical information such as the physicochemical properties, probable B- and T- cell epitopes, secondary and tertiary structures, cellular localization, and other criteria important for further *in vivo* study, for an efficacious vaccine candidate against malaria. The evaluation of the antigenicity and allergenicity showed that this protein was immunogenic and non-allergenic, also, several potential B- and T-cell epitopes were detected. A total of 60 potential post-translational modification sites were found in the sequence, with 56 phosphorylation sites and 4 acylation sites. The secondary structure of Q8IJ58 is made up of 28.34% alpha-helix, 53.83% random coil, and 17.83% extended strand. Iterative Threading ASSEmbly Refinement (I-TASSER) was used for the three-dimensional structure prediction of Q8IJ58, the Ramachandran plot showed that 96.7% residues were in the most favored region, 2.1% in the allowed regions, and 0.2% residues in the allowed regions, with an overall quality factor of 98.47%. Consensus gene ontology terms prediction for Q8IJ58 predicted the biological functions of the target protein. Our results suggest that Q8IJ58 is a potential vaccine candidate, and further studies are needed for *in vivo* evaluation.

Keywords: *Plasmodium falciparum*, Bioinformatics, *Pf*3D7_1035100, Vaccine, Q8IJ58

INTRODUCTION

Human malaria is a deadly tropical disease caused by infection with the protozoan Plasmodium, and it continues to be the most significant human parasitic illness in the world. The infection is transmitted by the bites of female Anopheles mosquitoes infected with plasmodium parasites (Arora *et al.*, 2021). Globally, the reported cases of malaria was estimated to be around 241 million with about 627,000 mortality in 2020, which represents an increase in morbidity of about 14 million, and 69,000 mortality compared to the scenario in 2019 (WHO, 2021). Malaria has remained a persistent global health problem, hence, it is no longer news that there is a pressing need for novel therapeutic approaches and interventions against the disease (Burns *et al.*, 2019). The asexual and sexual phases make up the developmental stages of the parasite. The asexual phase occurs primarily in the liver parenchymal of humans after the vector has deposited the sporozoites and then the erythrocytes. At infrequent rates during stress, the sexual phase occurs in the female Anopheles mosquito (vector) (Phillips *et al.*, 2017). The continuous ingress, intraerythrocytic development, and egress of merozoites during the blood stage represent a dogma related to the clinical symptoms of malaria and the development of parasite forms required for transmission to the mosquito vector (Cowman *et al.*, 2017; Phillips *et al.*, 2017; Varo *et al.*, 2020).

To subdue malaria and provide long term protective immunity against the disease, researchers have sequenced the genome of *P. falciparum* (Gardner et al., 2002) which provides a starting point for profiling vaccine candidates since vaccination as a therapeutic option cannot be overemphasized. However, the success of malaria vaccines have been impinged by several bottlenecks including; limited biomarkers, low efficacy, short half-life, and limited vaccine candidates, among others (Birkett *et al.*, 2013). We are hopeful that the recently approved RTS,S malaria vaccine will greatly help in malaria eradication (Egbewande, 2022), however, given the track record of the parasite, more efforts have to be put in place to identify novel vaccine candidates for the development of vaccines targeting different stages of the parasite to prevent cases of resurgences. The cost and time-saving approach of employing bioinformatics tools over traditional methods for antigen characterization has expedited the design and development of vaccines (Romano *et al.*, 2011). State-of-the-art online protein informatics tools have proven to be important toolkits for down-selection of malaria parasite proteins as therapeutic targets or biomarkers. Therefore, this study is targeted at using bioinformatics tools for the structural prediction, and antigenic analysis of Q8IJ58, to unravel potential epitopes for designing highly effective vaccines against *Plasmodium falciparum*.

MATERIALS AND METHODS

Protein selection and primary sequence retrieval

Plasmodium falciparum has about 5300 genes that have different spatio-temporal expressions in the parasite. Merozoite egress- and invasion-required genes are expressed during the late stages of development. The proteins encoded by these genes should possess a signal peptide for secretion and/or a membrane anchoring motif which can be a transmembrane domain or a GPI attachment signal. More importantly, we assessed the disruptability status of these genes based on the Piggybac transposon insertion mutagenesis screens which were used as an indication of their essentiality in the parasite prior to downstream immunoinformatics analysis. The details of the selection processes for the seven (7) hypothetical proteins evaluated in this study have been described in a previous report (Amlabu *et al.*, 2018). The amino acid sequence of Q8IJ58 was obtained using PlasmoDB (Aurrecoechea *et al.*, 2009).

Physicochemical parameters assessment

The properties of Q8IJ58 were evaluated with ProtParam tool (Gasteiger *et al.*, 2005). The evaluation includes amino acid composition, acidic and basic amino acids, isoelectric point (pI), instability index, protein size (MW), aliphatic index, extinction coefficients, estimated *in vitro* and *in vivo* half-life, and grand average of hydropathicity (GRAVY) were evaluated.

Identification of post-translational modification (PTM) sites

To predict the phosphorylation and acylation sites inherent in Q8IJ58 protein, NetPhos-3.1 (Blom *et al.*, 2004) and CSS-Palm Online servers (Ren *et al.*, 2008) were applied, correspondingly.

Prediction of transmembrane domains and subcellular localization

The TMHMM -2.0 online prediction server (Krogh *et al.*, 2001) was utilized to predict the transmembrane domains. Furthermore, the subcellular location of the Q8IJ58 was predicted using the PSORT II server (Nakai and Horton, 1999).

Secondary structure analysis and prediction of 3D models

GOR4 (Deléage, 2017) and PsiPred (McGuffin *et al.*, 2000) were utilized to predict the secondary structures of Q8IJ58. The 3D model structure is a critical component for vaccine design and protein function determination. The Iterative Threading ASSEmbly Refinement (I-TASSER) server was used for the 3D model prediction of the Q8IJ58 sequence. I-TASSER predicts protein structure and structure-based function annotation in a hierarchical manner (Roy *et al.*, 2010; Yang *et al.*, 2015). Furthermore, DiANNA online program was used for cysteine and disulfide bond predictions (Ferre and Clote, 2005).

Refinement and validation of the predicted 3D model structures

To generate predicted models, I-TASSER server was used and the models were refined using the GalaxyRefine program (Heo *et al.*, 2013), and SWISS-MODEL server was used to validate the refined models by using the Ramachandran plot at UCLA-DOE LAB — SAVES v6.0 (Colovos and Yeates, 1993). ProSA-web program was applied to evaluate the refined 3D models at ProSA-web (Wiederstein and Sippl, 2007).

B-cell epitopes prediction

Six physicochemical properties - accessibility, hydrophilicity, polarity, flexibility/mobility, exposed surface, and turns were employed to predict the Linear B-cell epitopes, using the Bcepred program at BCEPRED. This server predicts B-cell epitopes at a maximum accuracy of 58.7% and threshold of 2.38, when four amino acid properties: flexibility, hydrophilicity, polarity, and surface properties are combined (Saha and Raghava, 2004). Default parameters were applied during prediction using the Bcepred server. The artificial neural network (ANN) based B-cell epitope prediction server - ABCpred was also applied to predict the B-cell epitope (Saha and Raghava, 2006). This server can predict epitopes with an accuracy of 65.93% using a recurrent neural network (Saha and Raghava, 2006). The window length used for B-cell epitopes prediction was 16, with a threshold was 0.51%. Also, hydrophilicity, Bepipred linear epitope prediction (Larsen *et al.*, 2006), antigenicity (Kolaskar and Tongaonkar, 1990), surface accessibility (Emini *et al.*, 1985), beta-turn (Chou and Fasman, 1979), and flexibility (Karplus and Schulz, 1985), were predicted using the immune epitope database (IEDB).

MHC-I and MHC-II epitope prediction

The major histocompatibility complex class I and II (MHC-I and MHC-II) epitopes were predicted using the IEDB program at MHC-I Binding Predictions (Vita *et al.*, 2019), and NetMHCIIpan 4.0 server at NetMHCpan - 4.0 (Reynisson *et al.*, 2020). Ten alleles each, for MHC-I and MHC-II frequently occurring human alleles were selected: HLA–A*0101, HLA–A*0201, HLA–A*30:01, HLA–A*30:02, HLA–B*15:03, HLA–B*42:01, HLA–B*53:01, HLA–B*57:03, HLA–B*58:01, and HLA–B*58:02 for MHC-I epitope binding prediction, also, HLA–DRB1*0101, HLA–DRB1*0301, HLA–DRB1*0401, HLA–DRB1*0701 HLA–DRB1*1302, HLA–DRB3*0101, HLA–DRB5*0101, DPA10201- DPB10101, DQA10301- DQB10302, and DQA10501-DQB10301 for MHC-II. The window lengths used for MHC-1 and MHC-II epitope prediction were 9 and 15, respectively.

Cytotoxic T lymphocyte (CTL) epitopes prediction

Antigen presentation on the MHC-I surface is important for activating the immune system. CTLPred web server was used to predict CTL epitopes (Bhasin and Raghava, 2004). The default parameter was used to make the prediction, based on a consensus method (Nezafat *et al.*, 2014).

Antigenicity, allergenicity, and solubility evaluation

Prediction of the antigenicity and allergenicity of predicted epitopes are critical, as they are important criteria that contribute to the overall efficacy of a vaccine. The antigenicity and allergenic properties of predicted epitopes were evaluated using Vaxijen (Doytchinova and Flower, 2007), and AllerTop (Dimitrov *et al.*, 2013).

RESULTS AND DISCUSSION

Primary analysis of Q8IJ58

The amino acid sequence of the protein was obtained from the PlasmoDB server in FASTA format and used for further downstream analyses. The ProtParam server revealed that Q8IJ58 had a total of 561 amino acids, with molecular weight and isoelectric point of 64 kDa and 5.13 respectively. The result for the physicochemical properties is presented in Table 1.

Table 1: Physiochemical properties of Q8IJ58 characterized by using ProtParam server

Physicochemical Properties	
Number of amino acids	561
Molecular weight	64.879 KDa
Theoretical Pi	5.13
Number of negatively charged residues (Asp + Glu)	111
Number of positively charged residues (Arg+Lys)	78
Total number of atoms	9001
Extinction co-efficient	42080 M^{-1} cm^{-1}, assuming all pairs of Cys residues form cystines 41830 M^{-1} cm^{-1}, assuming all Cys residues are reduced
Half-life	Mammalian reticulocytes (in vitro): 1 Hour Yeast (*in vivo*): 30 Min *E. coli* (*in vivo*): >10 hours
Stability index	48.21
Aliphatic index	71.00
GRAVY	-1.138

Identification of PTM sites

The phosphorylation and acylation sites predicted using NetPhos 3.1 and CSS-Palm online servers, correspondingly, show 56 phosphorylation sites (Ser: 29, Thr: 11, Tyr: 16) (Figure 1) and 4 acylation sites (Table 2), indicating that the Q8IJ58 sequence likely consists of 60 total post-translational modification sites.

```
      MNKIWNITFCLLFQYLCIFQNKVSTNDITKEDKRNLRHRLGVDNFNIKAE   #     50
      NNKIEGKDGKKGSEASEKETKEKNSSNYHESSIDHNDNDNDNLLQSEEQT   #    100
      DTNFGVRDKTNNKKLIESTITNEKTQGKNSYHDDENTIECSHPEALVGGE   #    150
      KNRQNSFALRSNSIVTNKPITSKCGKNSSDDSNHDEIIDHHENLQIISLN   #    200
      APPSPIFDQKDVSIDVVSEHERPSELPSDSIQVNKDNRIHLDGHQSNIGS   #    250
      FQLVYEGLHHGNDGVHGATEDRKKARPQKIKGPNQERVERIHELTMDKFE   #    300
      AQKENTVSGDGDKRGKVIYDENDDSLISNSLVPETSYNQDYSLYTINILN   #    350
      GLPDKKSNMKNIDLEKFKDTSIHTKEEGPIVSKIHSVKEEQKDYHKYEKN   #    400
      ILHGTKKKSEKVYDLYDYILNWVFGPMKITETNVSEQNDSLELKQRTSWD   #    450
      RTNIISENEEEDDINNVEQLNEIQGEITEDEVEEEKEEIIEEEEYVDEQK   #    500
      MVNKKKIHDLEKINKNEEKNVITKDIETTKYTVTHLEKYNNIKKDVPNFF   #    550
      KTIINLFSDYK                                          #    600
%1    .......T..............S.........................      #     50
%1    ............S..S...T....SS.Y..SS.............S....   #    100
%1    .........T.......S...........SY.........S.........   #    150
%1    .....S......S..T....T......S...S..................   #    200
%1    ...S.............S.....S.....S.....................   #    250
%1    ...................T.........................T.....   #    300
%1    .......S...........Y.....S...........Y...YS........   #    350
%1    ......S............TS..T...........S.......Y......   #    400
%1    ....T...S...Y..Y..............T..S............S..   #    450
%1    .....S.......................T..................Y.....   #    500
%1    ......................T....TT.Y..T....Y...........   #    550
%1    ...........
```

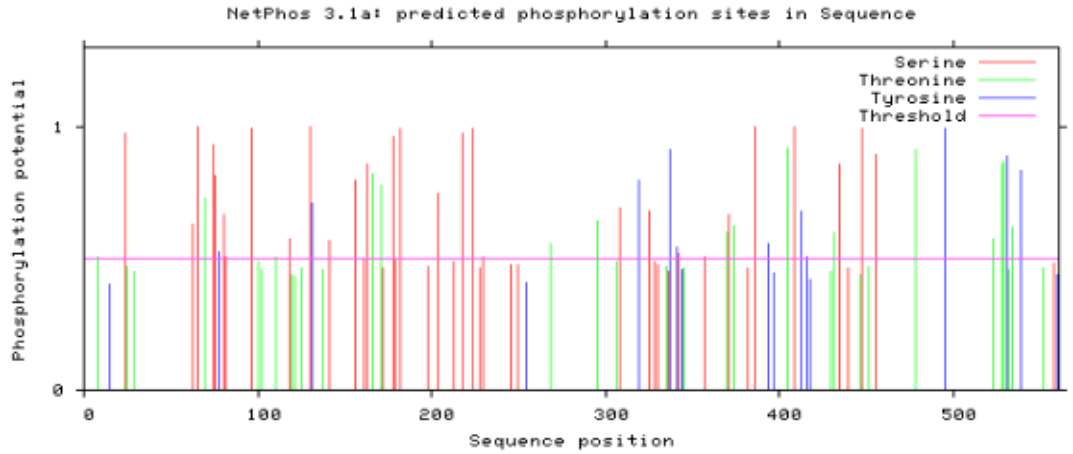


Figure 1: The output from NetPhos server for phosphorylation sites of Q8IJ58. (A) The number of predicted S/T/Y phosphorylation sites; residues having a prediction score above the threshold are indicated by 'S', 'T' or 'Y', respectively. (B) Diagram showing the prediction of phosphorylation sites

Table 2: The predicted acylation sites of Q8IJ58 sequence

ID	Position	Peptide	Score
PF3D7_1035100	10	KIWNITFCLLFQYLC	36.817
PF3D7_1035100	17	CLLFQYLCIFQNKVS	26.919
PF3D7_1035100	140	DDENTIECSHPEALV	3.991
PF3D7_1035100	174	NKPITSKCGKNSSDD	5.199

Prediction of transmembrane domains and subcellular localization

The TMHMM server v. 2.0 revealed the absence of transmembrane domain (Figure 2) in the Q8IJ58 sequence. Furthermore, the prediction of the subcellular localization of Q8IJ58 using PSORT II revealed: 43.5% nuclear, 34.8% cytoplasmic, 13.0 % mitochondrial, and 8.7% cytoskeletal.

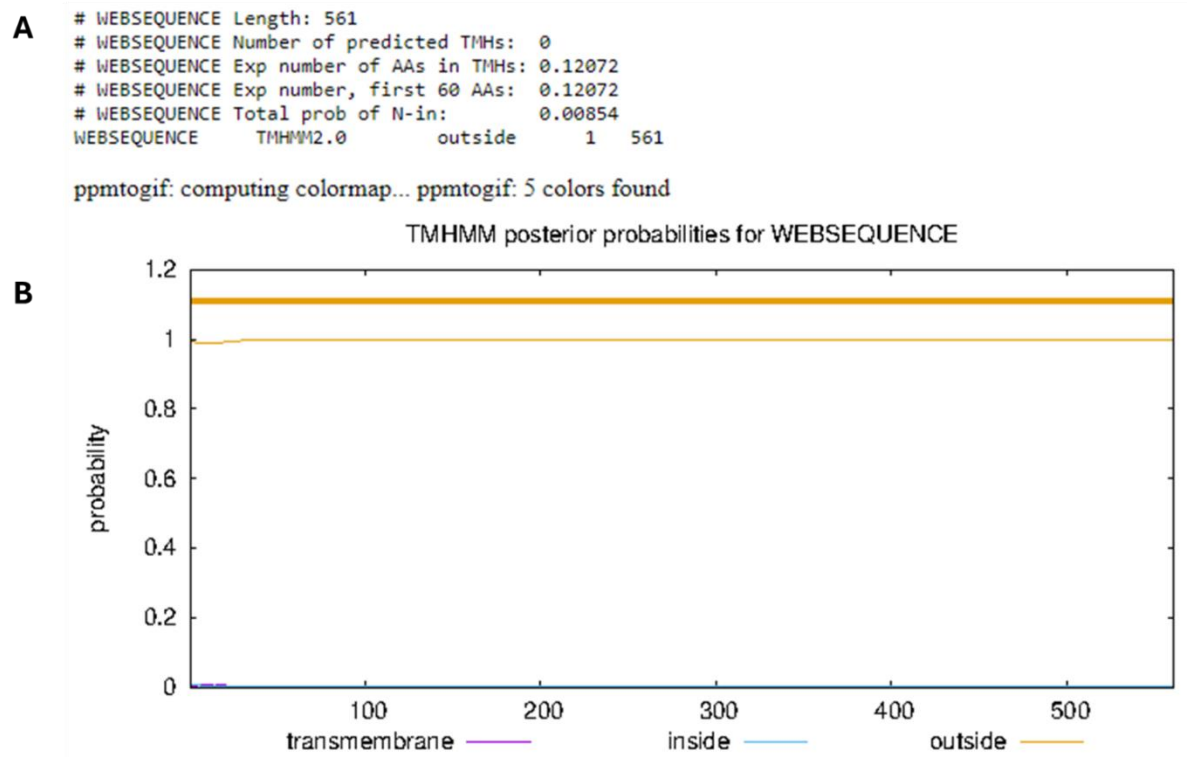


Figure 2: Prediction of transmembrane helices in Q8IJ58 protein. (A) list of predicted transmembrane helices location, and the predicted location of the intervening loop regions. If the Exp number of AAs in TMHs is greater than 18 it is likely to be a transmembrane protein or have a signal peptide; If the Exp number, first 60 AAs is more than a few, could indicate that there could be a signal peptide in the predicted transmembrane helix in the N-term; Total prob of N-in: This is the total probability that the Nterm is on the cytoplasmic side of the membrane. (B) Analysis of the transmembrane helices of Q8IJ58

Prediction and analysis of secondary structure

GOR4 and PSIPRED servers show protein structural moieties in the secondary structure of the Q8IJ58 protein (Figures 3 & 4). The data obtained showed that the proportions of the alpha-helix, random coil, and extended strand in the sequence were 28.34% (159/561), 17.83% (100/561), and 53.83% (302/561), correspondingly. DiANNA server is a tool for cysteine and disulfide bond prediction. A total of 4 cysteines were shown in our sequence and should form disulfide bonds at positions: 10–17, and 140-174 (Table 3).

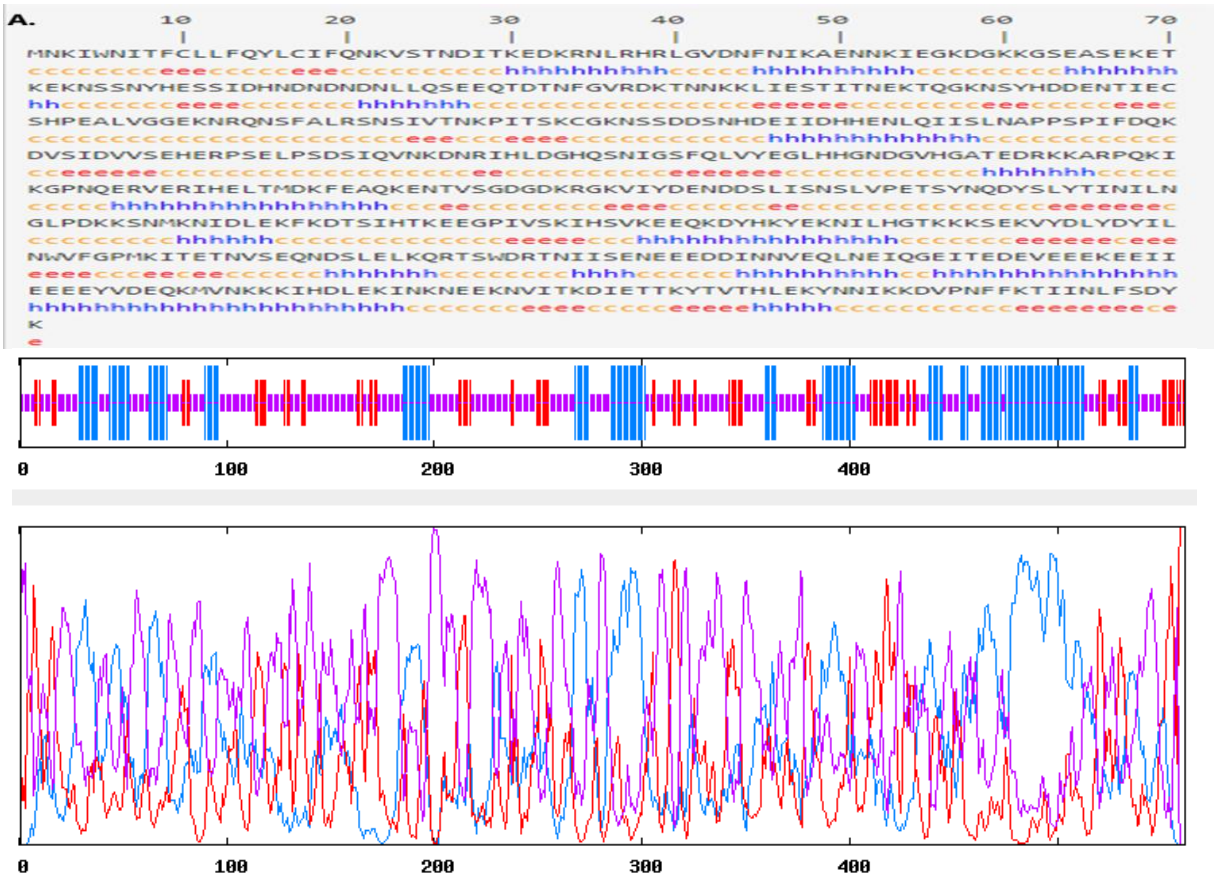
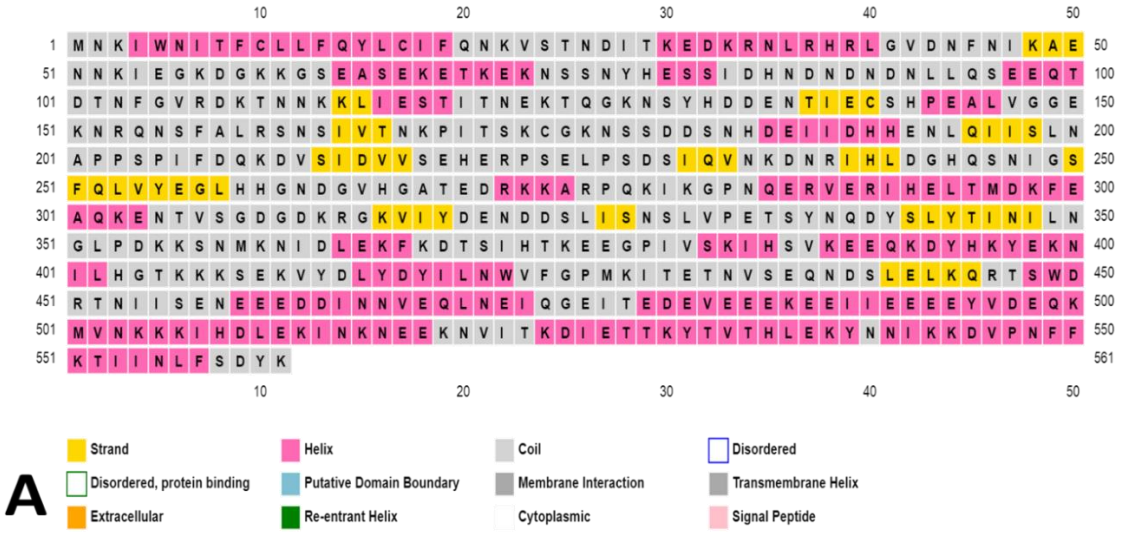


Figure 3: Analysis of the secondary structure of Q8IJ58 using GOR 4. A: Predicted secondary structure (h: helix, e: extended strand, c: coil). B: Graphical frame of secondary structure prediction



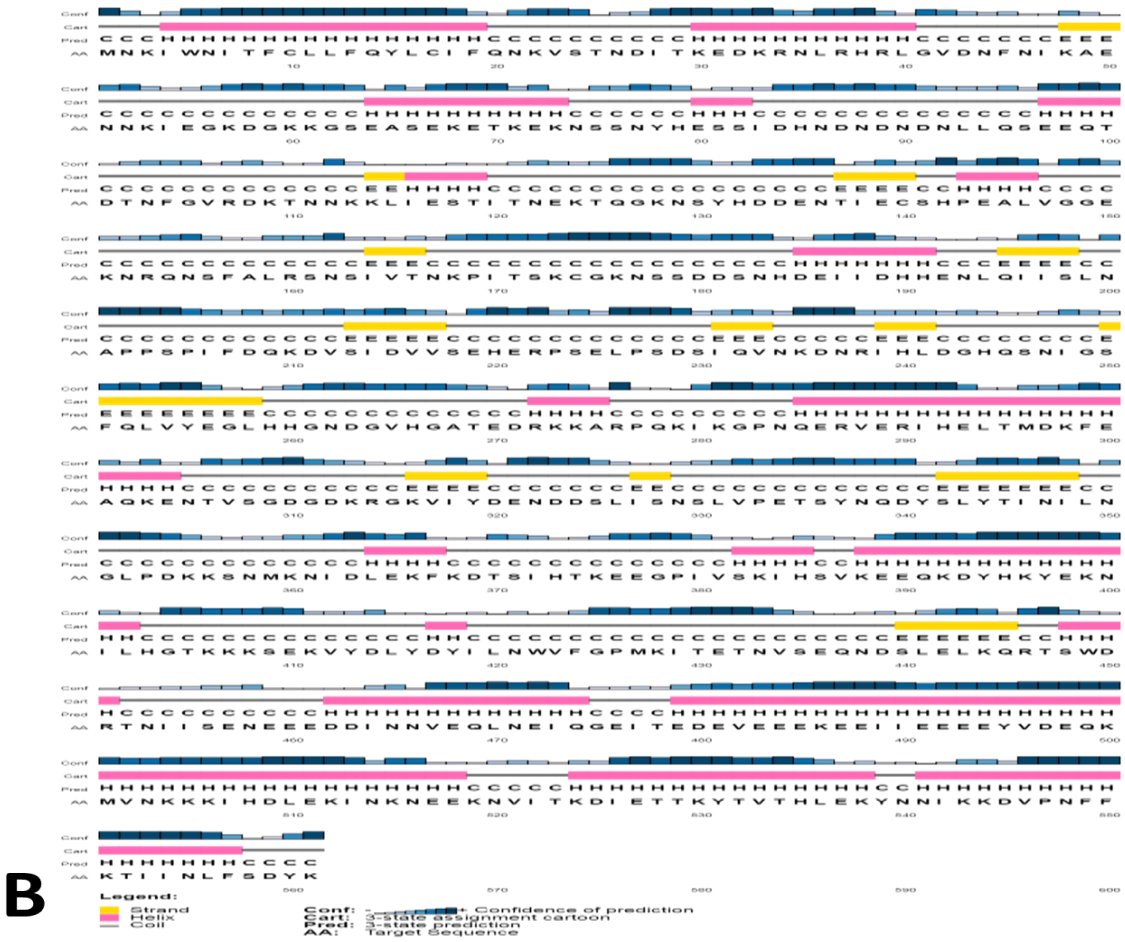


Figure 4: Graphical result from secondary structure prediction using PSIPRED

Table 3: Predicted disulfide bonds using a trained neural network (higher the score, higher the prediction reliability)

Cysteine sequence position	Distance	Bond	Score
10 – 17	7	WNITFCLLFQY-LFQYLCIFQNK	0.01039
10 – 140	130	WNITFCLLFQY-ENTIECSHPEA	0.01111
10 – 174	164	WNITFCLLFQY-PITSKCGKNSS	0.0107
17 – 140	123	LFQYLCIFQNK-ENTIECSHPEA	0.01038
17 – 174	157	LFQYLCIFQNK-PITSKCGKNSS	0.0104
140 – 174	34	ENTIECSHPEA-PITSKCGKNSS	0.02312
Weighted matching			
Peptide bonds			
10 – 17	WNITFCLLFQY – LFQYLCIFQNK		
140 – 174	ENTIECSHPEA – PITSKCGKNSS		
Predicted connectivity			
1-2, 3-4			

Prediction, refinement and validation of the 3D structure

Five 3D models (Figure 5) were predicted by LOMETS (Local meta-threading server) threading program, from the top 10 high-scoring templates on the I-TASSER tool (Wu and Zhang, 2007; Zheng *et al.*, 2019). Model 2, with the highest C-score was selected as a higher C-score was congruent with higher model quality. Subsequently, the GalaxyRefine program was used to refine the predicted 3D structures to improve the model quality, after which, the overall quality (Z-score) of the refined models were compared using ProSA-web server. The Z-score obtained prior to, and after refinement was -3.91 and -4.01, respectively. Ramachandran plot analysis through the PROCHECK server showed that unrefined model 2, had 90.6% residues in the most favored region, 7.5% in the allowed regions, and 0.6% residues in the allowed regions, with an overall quality factor of 94.46% as predicted by the ERRAT server. However, following refinement, the protein had 96.7% residues in the most favored region, 2.1% in the allowed regions, and 0.2% residues in the allowed regions, with an overall quality factor of 98.47% (Figure 5).

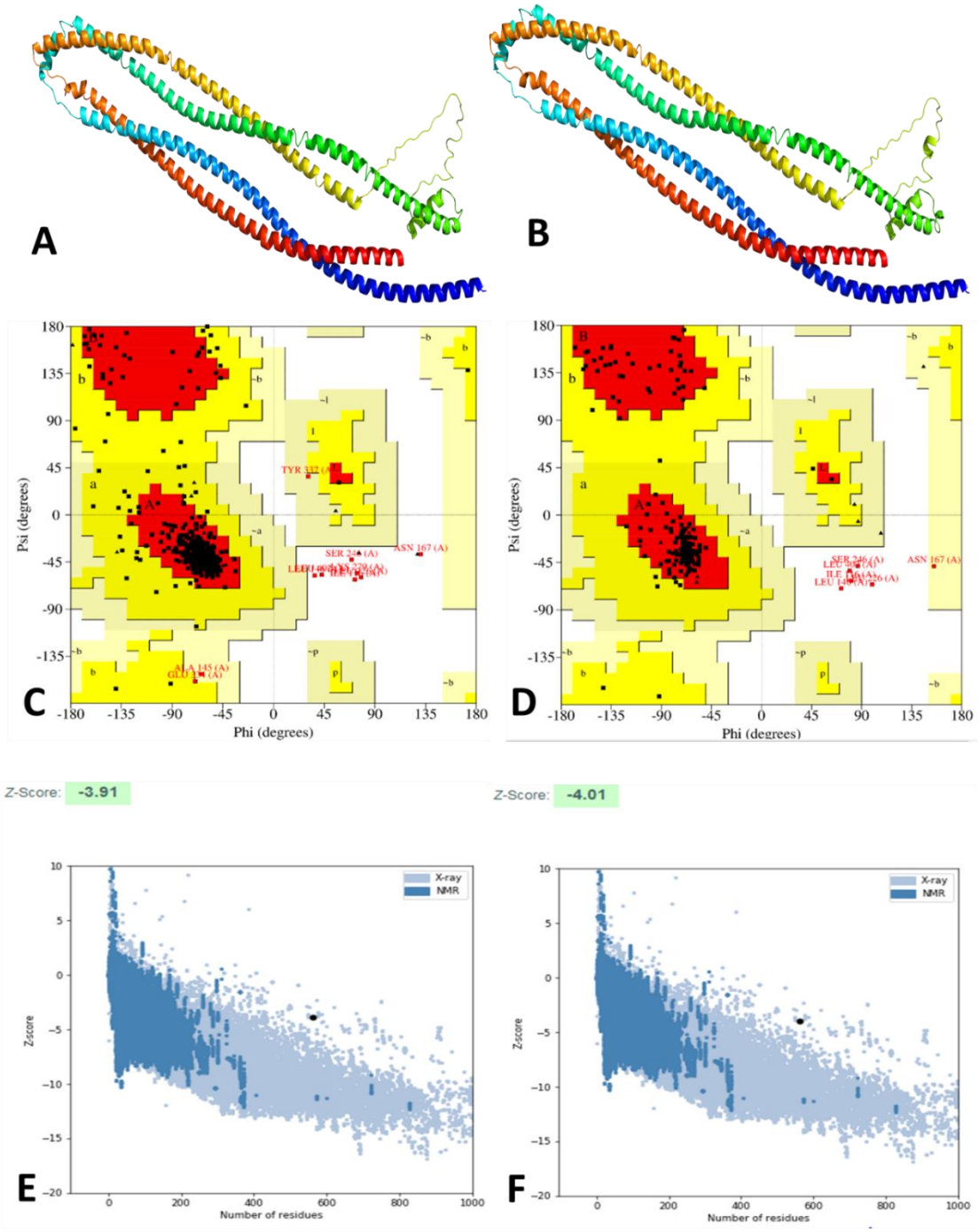


Figure 5: Tertiary structure and validation of the 3D structure before and after refinement. A) Cartoon representation of Q8IJ58 tertiary structure for unrefined model. B) Cartoon representation of Q8IJ58 tertiary structure after refinement. C) Ramachandran plot before refinement. D) Ramachandran plot after refinement. E) Z-score plot before refinement. F) Z-score plot after refinement

Continuous B-cell epitopes

Bcepred, ABCpred, and IEDB free online servers were employed to forecast the linear epitopes, and details are shown in Tables 4 and 5, and Figure 6, respectively. Bcepred was employed to evaluate the B-cells epitopes in terms of accessibility, mobility, exposed layer, turns, antigenic propensity, hydrophilicity, and polarity. The threshold, minimum, and maximum scores for Bepipred linear epitope prediction, beta-turn, surface accessibility, flexibility, antigenicity, and hydrophilicity using the IEDB online tool were: (0.554, 0.225, 0.709), (1.027, 0.663, 1.517), (1.000, 0.040, 4.706), (1.025, 0.912, 1.118), (0.988, 0.815, 1.204), and (2.903, -5.457, 8.314), respectively.

Table 4: B-Cell epitopes predicted on Q8IJ58 using Bcepred based on different parameters

Prediction parameters	Epitope sequence
Hydrophilicity	QNKVSTNDITKEDKRN, ENNKIEGKDGKKGSEASEKETKEKNSSNYHESSIDHNDNDNDNL, QSEEQTDTNFGVRDKTNNKK, ESTITNEKTQGKNSYHDDENTIE, VGGEKNRQNS, TSKCGKNSSDDSNHDEI, SEHERPSE, QVNKDNR, GNDGVHGATEDRKKARPQK, KGPNQERVER, EAQKENTVSGDGDKRGKVIYDENDDSL, PETSYNQDYS, PDKKSNMKN, EKFKDTS, HTKEEGP, HSVKEEQKDYHK, HGTKKKSEKV, TETNVSEQNDSLE, ISENEEEDDINN, QGEITEDEVEEEKEEI, EEEEYVDEQK, EKINKNEEKNV, KDIETTK
Flexibility	IFQNKVS, NDITKEDKRN, FNIKAENNKIEGKDGKKGSEASEKETKEKNSS, IDHNDNDN, NLLQSEEQ, FGVRDKTNN, STITNEKTQGK, ALVGGEKNRQ, KPITSKCGKNSSDDS, SIQVNKD, HGATEDRKKARPQKIKGPNQ, NTVSGDGDKRG, VIYDEND, NGLPDKKSNMKNIDLEKFKD, SIHTKEE, HSVKEEQ, NILHGTKKKSE, TNVSEQN, SLELKQR, NIISENEEE, DEVEEEK, EKINKNEE
Accessibility	QNKVSTNDITKEDKRNLRHRL,NIKAENNKIEGKDGKKGSEASEKETKEKNSSNYHESSIDHNDNDNDNLLQSEEQTDTNFGVRDKTNNKKLIESTITNEKTQGKNSYHDDENTIE, VGGEKNRQNSFA, TNKPITSKCGKNSSDDSNHDEIIDHHENLQ, PIFDQKDVS, VSEHERPSELPSD, IQVNKDNRIHL,HGATEDRKKARPQKIKGPNQERVERIHELTMDKFEAQKENTVSGDGDKRGKVIYDENDDSL, VPETSYNQDYSLYT, NGLPDKKSNMKNIDLEKFKDTSIHTKEEGPI, KIHSVKEEQKDYHKYEKNILHGTKKKSEKVYDLYD, PMKITETNVSEQNDSLELKQRTSWDRTNIISENEEEDDINNVEQLNEIQGEITEDEVEEEKEEIIEEEEYVDEQKMVNKKKIHDLEKINKNEEKNVITKDIETTKYTVTHLEKYNNIKKDVPNF, NLFSDYK
Turns	GVDNFNI, EKNSSNYHESSIDHNDNDNDNLL, KNSYHDDENT, GKNSSDDSNHDEIIDHHENLQ, GLHHGNDGV, YDENDDSL
Exposed surface	DITKEDKRNLRHR, IKAENNKIEGKDGKKGS, ASEKETKEKNSSNY, NDNDNDN, QSEEQTD, VRDKTNNKKLIE, NEKTQGKNSYHDD, GEKNRQNSF, NKPITSK, EHERPSE, QVNKDNRIH, ATEDRKKARPQKIKGPNQERVER, DKFEAQKENT, DGDKRGK, NGLPDKKSNMKNIDLEKFKDTS, SVKEEQKDYHKYEKNILHGTKKKSEKVYD, ELKQRTSW, SENEEEDD, EDEVEEEKEEI, EYVDEQKMVNKKKIHDLEKINKNEEKNV, KDIETTK, EKYNNIKKDVPN, SDYK
Polarity	NDITKEDKRNLRHRLGVD, IKAENNKIEGKDGKKGSEASEKETKEKNSSN, HESSIDHNDNDND, RDKTNNKKLIEST, NEKTQGKNSYHDDENTIECSHPEA, VGGEKNRQNS, SDDSNHDEIIDHHENLQI, DVVSEHERPSELP, QVNKDNRIHLDGH, VYEGLHHGN, HGATEDRKKARPQKIKGPNQERVERIHELTMDKFEAQKEN, SGDGDKRGKVIYDENDDSL, DKKSNMKNIDLEKFKDTSIHTKEEGPI, SKIHSVKEEQKDYHKYEKNILHGTKKKSEKVYD, SLELKQRTSWDR, IISENEEEDDINN, GEITEDEVEEEKEEIIEEEEYVDEQKMVNKKKIHDLEKINKNEEKNVITKDIETTK, TVTHLEKYNNIKK
Antigenic propensity	NITFCLLFQYLCIFQNKV, LRHRLGV, NLQIISLN, DVSIDVVSEH, IGSFQLVYEGLHH, LISNSLVPETSY, LYTINIL, PIVSKIHSV, KVYDLYDYILNWVF, KYTVTHL

Table 5: B-Cell epitopes predicted on Q8IJ58 using abcpred (artificial neural network)

Linear B Cell			
Rank	**Sequence**	**Start position**	**Score**
1	RVERIHELTMDKFEAQ	287	0.92
2	HESSIDHNDNDNDNLL	79	0.91
3	SEEQTDTNFGVRDKTN	96	0.89
3	EEEYVDEQKMVNKKKI	492	0.89
3	TNIISENEEEDDINNV	452	0.89
4	TVTHLEKYNNIKKDVP	532	0.88
5	GSEASEKETKEKNSSN	62	0.87
5	NGLPDKKSNMKNIDLE	350	0.87
5	GKVIYDENDDSLISNS	315	0.87
5	PQKIKGPNQERVERIH	277	0.87
5	YEGLHHGNDGVHGATE	255	0.87
6	KKSEKVYDLYDYILNW	407	0.86
6	KDTSIHTKEEGPIVSK	368	0.86
6	HGATEDRKKARPQKIK	266	0.86
6	SIVTNKPITSKCGKNS	163	0.86
7	KEEIIEEEEYVDEQKM	486	0.85
7	DSLISNSLVPETSYNQ	324	0.85
7	HDEIIDHHENLQIISL	184	0.85
8	FGPMKITETNVSEQND	424	0.84
8	HRLGVDNFNIKAENNK	38	0.84
8	KKLIESTITNEKTQGK	113	0.84
9	KNVITKDIETTKYTVT	519	0.83
9	QGEITEDEVEEEKEEI	474	0.83
10	IETTKYTVTHLEKYNN	526	0.82
10	EGPIVSKIHSVKEEQK	377	0.82
10	EKTQGKNSYHDDENTI	123	0.82
11	TSYNQDYSLYTINILN	335	0.81
11	GGEKNRQNSFALRSNS	148	0.81
12	HSVKEEQKDYHKYEKN	385	0.80
12	EHERPSELPSDSIQVN	219	0.80

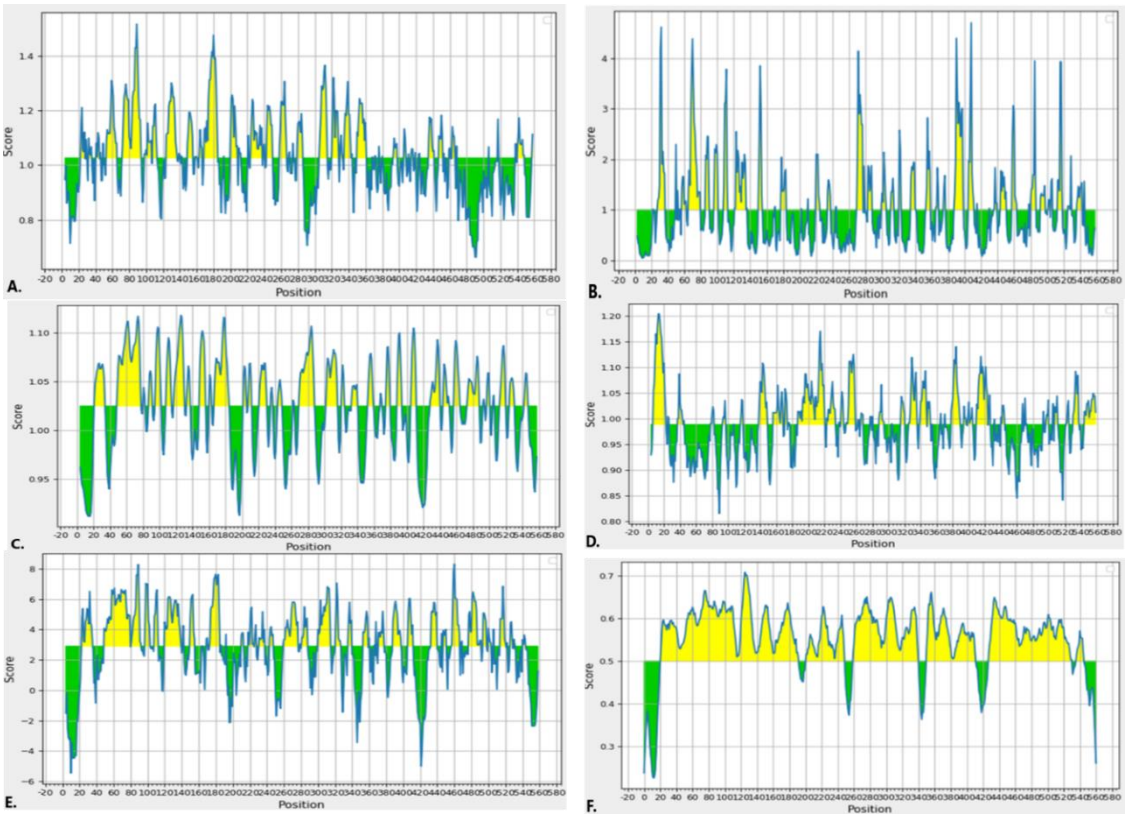


Figure 6: The output of IEDB server based on parameters including. A: Bepipred linear epitope prediction, B: ß-turn, C: surface accessibility, D: flexibility, E: antigenicity, and F: hydrophilicity. The higher scoring residues represent a higher probability to be a part of the epitope (those residues are colored in yellow on the graphs)

MHC-I and MHC-II epitope prediction

Epitopes were selected based on their minimum percentile ranks, which indicates a stronger affinity for the receptor molecule. Details are listed in Tables 6 and 7 respectively.

Table 6: Prediction of MHC-I binding epitopes on Q8IJ58 using IEDB

MHC I						
Allele	**Start**	**End**	**Length**	**Peptide**	**Score**	**Percentile rank**
HLA-A*01:01	408	416	9	KSEKVYDLY	0.927542	0.02
HLA-A*30:01	160	168	9	RSNSIVTNK	0.901201	0.01
HLA-A*30:01	165	173	9	VTNKPITSK	0.871095	0.01
HLA-A*01:01	531	539	9	YTVTHLEKY	0.840509	0.05
HLA-B*15:03	542	550	9	IKKDVPNFF	0.835142	0.04
HLA-B*42:01	546	554	9	VPNFFKTII	0.819245	0.1
HLA-B*15:03	405	413	9	TKKKSEKVY	0.803304	0.05
HLA-A*30:01	105	113	9	GVRDKTNNK	0.802502	0.02
HLA-A*30:01	522	530	9	ITKDIETTK	0.70414	0.04
HLA-A*30:01	290	298	9	RIHELTMDK	0.655643	0.06
HLA-A*30:02	531	539	9	YTVTHLEKY	0.646482	0.08
HLA-B*15:03	149	157	9	GEKNRQNSF	0.62278	0.16
HLA-A*30:02	336	344	9	SYNQDYSLY	0.622426	0.09
HLA-B*15:03	359	367	9	MKNIDLEKF	0.613363	0.17
HLA-B*15:03	288	296	9	VERIHELTM	0.580754	0.19
HLA-B*15:03	70	78	9	TKEKNSSNY	0.562377	0.21
HLA-B*15:03	218	226	9	SEHERPSEL	0.561551	0.21
HLA-A*30:02	408	416	9	KSEKVYDLY	0.550189	0.14
HLA-A*30:01	380	388	9	IVSKIHSVK	0.536983	0.11
HLA-B*15:03	37	45	9	RHRLGVDNF	0.528091	0.24
HLA-B*15:03	523	531	9	TKDIETTKY	0.525192	0.25
HLA-B*58:01	408	416	9	KSEKVYDLY	0.522888	0.3
HLA-B*58:01	531	539	9	YTVTHLEKY	0.520466	0.31
HLA-A*30:01	279	287	9	KIKGPNQER	0.517416	0.12
HLA-A*30:01	120	128	9	ITNEKTQGK	0.497222	0.14
HLA-A*02:01	240	248	9	HLDGHQSNI	0.493344	0.26
HLA-A*02:01	3	11	9	KIWNITFCL	0.492754	0.26
HLA-A*30:02	386	394	9	SVKEEQKDY	0.491572	0.17
HLA-B*15:03	199	207	9	LNAPPSPIF	0.467834	0.3
HLA-B*15:03	395	403	9	HKYEKNILH	0.447278	0.32

Table 7: Prediction of MHC-II binding epitopes on Q8IJ58 using NetMHCIIpan

MHC II			
Allele	**Peptide**	**% Rank**	**Binding level**
DRB1_0101			
	MDKFEAQKENTVSGD	0.51	SB
	ENLQIISLNAPPSPI	0.67	SB
	NLQIISLNAPPSPIF	0.70	SB
DRB1_0102			
	ENLQIISLNAPPSPI	0.20	SB
	NLQIISLNAPPSPIF	0.21	SB
	HENLQIISLNAPPSP	0.32	SB
DRB1_0401			
	PNFFKTIINLFSDYK	0.49	SB
	FGPMKITETNVSEQN	0.54	SB
	SNSLVPETSYNQDYS	0.61	SB
DRB3_0101			
	SIQVNKDNRIHLDGH	0.10	SB
	DSIQVNKDNRIHLDG	0.13	SB
	SDSIQVNKDNRIHLD	0.22	SB
DRB5_0101			
	VSKIHSVKEEQKDYH	0.74	SB
	ESTITNEKTQGKNSY	0.93	SB
	DVSIDVVSEHERPSE	1.02	WB

DRB1_0701			
	MDKFEAQKENTVSGD	1.91	WB
	RPSELPSDSIQVNKD	2.52	WB
	TMDKFEAQKENTVSG	2.60	WB
DRB1_1302			
	SDSIQVNKDNRIHLD	0.00	SB
	PSDSIQVNKDNRIHL	0.02	SB
	LPSDSIQVNKDNRIH	0.03	SB
HLA-DPA10201-DPB10101			
	IHELTMDKFEAQKEN	0.37	SB
	ETTKYTVTHLEKYNN	0.41	SB
	TKDIETTKYTVTHLE	0.42	SB
HLA-DQA10501-DQB10301			
	IQGEITEDEVEEEKE	2.76	WB
	QGEITEDEVEEEKEE	2.22	WB
	EIQGEITEDEVEEEK	4.05	WB
HLA-DQA10301-DQB10302			
	QGEITEDEVEEEKEE	0.10	SB
	DQKDVSIDVVSEHER	0.15	SB
	IQGEITEDEVEEEKE	0.22	SB

CTL epitopes prediction

The CTL epitope prediction was performed using the CTLpred server and the top ten high ranked epitopes were selected using the CTLpred server. More details in Table 8.

Table 8: Predicted CTL epitopes of Q8IJ58 using CTLpred (combined method)

Peptide rank	Start position	sequence	Score(ANN/SVM)
1	15	YLCIFQNKV	0.98/0.89035171
2	191	HENLQIISL	0.77/1.060624
3	376	EEGPIVSKI	0.86/0.88544261
4	22	KVSTNDITK	0.94/0.67764596
5	408	KSEKVYDLY	1.00/0.52661774
6	160	RSNSIVTNK	0.84/0.68106308
7	465	NNVEQLNEI	0.85/0.66813004
8	354	DKKSNMKNI	0.94/0.5284223
9	356	KSNMKNIDL	0.60/0.77278893
10	341	YSLYTINIL	0.79/0.48627807

Antigenicity and Allergenicity Prediction

The antigenicity was determined by using an alignment-free Vaxijen server and revealed that it has the antigenicity probability of 0.6609. Allergenicity prediction using Allertop server predicted Q8IJ58 to be a probable non-allergen.

Discussion

The World Health Organization (WHO) set the goal to eradicate malaria in 1948, and efforts have been intensified to this end. Correspondingly, malaria-related fatalities progressively decreased, but this progress was stalled due to the COVID-19 pandemic and attributed to the interference with malaria services and diagnostic practices (Albrecht-Schgoer *et al.*, 2022). Out of the 172 Plasmodium species, there are only five anthropophilic species, namely, *P. falciparum, P. vivax, P. malariae, P. ovale, and P. knowlesi,* with *P. falciparum* posing the most threat to life (Barber *et al.*, 2018; Pannu, 2019; Talapko *et al.*, 2019). To develop a novel vaccine, it is essential to identify prospective immunoprotective antigens of the parasite (Ghaffari *et al.*, 2020). Antigen characterization using bioinformatics tools is central to protein-based vaccine design and development. This study applies several bioinformatics tools to analyze the uncharacterized *Plasmodium falciparum* protein - Q8IJ58, to obtain information that would be useful in vaccine design research. To the best of our knowledge, this is the first *in silico* investigative study on Q8IJ58.

According to the results from ProtParam server, Q8IJ58 has a total of 561 amino acid residues, with a molecular weight (MW) of 64.9 kDa. The calculation for the stability index showed that the protein is unstable, with a value of 48.21. The aliphatic index (relative volume occupied by the aliphatic side chain; alanine, valine, isoleucine, and leucine) of the protein is predicted to be 71.00, which shows that the protein may be more stable within a wide range of temperatures (thermostable). The grand average of hydropathicity (GRAVY) score is -1.138, and the negative value indicates the protein is hydrophilic. Post-translational modification (PTM) has enabled structural and functional diversification of proteins beyond the genome (Chuh *et al.*, 2016). The identification of phosphorylation sites on proteins is important in analyzing the signaling networks and functional relationships between signaling proteins (Dephoure *et al.*, 2013). Accordingly, we employed the NetPhos 3.1 and CSS-Palm 4.0 servers to predict potential phosphorylation and acylation sites on Q8IJ58. The results showed that there are

60 potential PTM sites, with 56 phosphorylation sites and 4 acylation sites present in Q8IJ58, which may play a key role in the protein's function and affect its activity. TMHMM server predicted that Q8IJ58 has no transmembrane domain, therefore, it could interact with antigen-presenting cells (APCs) to trigger an immediate immune response.

Knowledge about the secondary structure of a protein is important for subsequent prediction of the protein interaction sites, and 3D tertiary structure. GOR4 and PRISPRED servers were used for the secondary structure prediction of our sequence. Results from GOR4 shows that Q8IJ58 comprises 28.3% alpha-helix, 17.83% extended strand, and 53.83 random coils. The alpha-helix and beta-turn have internal localization within the protein and are less likely to act as epitopes, rather, they maintain the structure of the protein due to the high chemical bond energy they contain. Random coils are less rigid regions of the protein and comprise over half the overall structure of Q8IJ58. This is good because they are more likely to act as epitopes. In addition, two disulfide bonds were predicted using the DiNNNA server. Disulfide bond formation is an important PTM, which affects the overall protein structures and functions (Winther and Thorpe, 2014).

The overall 3D structure of a protein is important for its functional determination. The 3D structure of Q8IJ58 was predicted by using the Iterative Threading ASSEmbly Refinement (I-TASSER). The I-TASSER server first retrieves template proteins using LOMETS. The top-scoring templates are ranked and selected based on the combination of their alignment Z score, program-specific confidence scores, and the sequence identity to the query protein. All of the independently identified threading templates have a normalized Z score >1, which indicates a good alignment (Roy *et al.*, 2010). Five 3D models are then constructed by MODELLER from the top five templates for easy/trivial targets, whereas, hard and very hard targets are constructed by L-BFGS (Wu and Zhang, 2007; Zheng *et al.*, 2019).

ProSAweb server was employed to evaluate the overall protein quality of the top five 3D models predicted for Q8IJ58, after refinement by the GalaxyRefine program. The Z score represents the overall model quality (Wiederstein and Sippl, 2007). Models 2 had the highest C-score of -1.29, which was akin to a higher model quality. It was subjected to refinement to improve the overall model quality. Subsequently, Ramachandran plot was generated to validate the model structures, using the PROCHECK server.

B cells, T-cells and the host immune system molecules recognize the epitope(s) present on an antigen; therefore, it is important to analyze the protein to determine its antigenicity and antigenic specificity, which are important criteria in the design of epitope-based vaccines. Humoral and cell-mediated immune response are important for protection against malaria; thus, multiple epitope prediction indices were used to obtain sufficient information about Q8IJ58, and these indices include surface accessibility, flexibility, antigenicity, etc. Linear B-cell epitopes were predicted using valid bioinformatics tools- ABCpred, and Bcepred, and IEDB. Data obtained shows that Q8IJ58 has probable epitopes that can make it a potential candidate in vaccine design. There were 30 top-scoring epitopes predicted on Q8IJ58 using ABCpred (listed in Table 5). The ABCpred server has 65.93% precision for epitope prediction using a recurrent neural network. Also, we utilized Bcepred to predict linear B-cell epitopes, using physicochemical properties such as hydrophilicity, flexibility/mobility, accessibility, polarity, exposed surface, and turns. The accuracy of Bcepred varies from 52.92% to 57.53%, and it has been proven that the combination of four amino acid properties (hydrophilicity, flexibility, polarity, and exposed surface) gives the highest accuracy of 58.70% (Saha and Raghava, 2004). The IEDB server also associated some protein parameters such as hydrophilicity, flexibility, accessibility, turns, exposed surface, polarity, and antigenic propensity to the location of continuous epitopes. Thirty high scoring MHC I and MHC II epitopes were selected for each, using the IEDB and NetMHCIIpan 4.0 servers, respectively. The fact that not all MHC binders act as T cell epitopes, required the need for better antigen prediction. The CTLpred server uses consensus and combined prediction of artificial neural network (ANN) and support vector machine (SVM) to improve sensitivity, accuracy, and epitope specificity. The antigenicity and allergenicity prediction are critical to vaccine development. Q8IJ58 was evaluated for its allergenicity and antigenicity, using the Vaxijen and AllerTOP servers, respectively. Results showed that the protein was immunogenic and non-allergen. From these results, it is now important to confirm these results experimentally in appropriate mouse models to ascertain the efficacy of Q8IJ58 as a putative vaccine candidate against falciparum malaria.

CONCLUSION

This study presents the results for the bioinformatics analysis of Q8IJ58. The data obtained showed that Q8IJ58 has strong antigenicity, flexibility, surface accessibility, and hydrophilicity indexes. The analyses of the B- and T-cell epitopes of Q8IJ58 indicates its potential as a promising vaccine candidate against *Plasmodium falciparum*. The basic and theoretical information about Q8IJ58 will be important for further *in vivo* studies to develop an effective Q8IJ58-based vaccine against malaria.

ACKNOWLEDGEMENT

All authors wish to acknowledge members of the Genomics and Molecular Biotechnology Research and Training Laboratory (GMBRTL) for reviewing this manuscript and also, the leadership of Prince Abubakar Audu University, Anyigba, Nigeria for providing the enabling environment for this research.

REFERENCES

Albrecht-Schgoer, K., Lackner, P., Schmutzhard, E. and Baier, G. (2022). Cerebral Malaria: Current Clinical and Immunological Aspects. *Front Immunol*, *13*, 863568. https://doi.org/10.3389/fimmu.2022.863568

Amlabu, E., Mensah-Brown, H., Nyarko, P. B., Akuh, O.-a., Opoku, G., Ilani, P., Oyagbenro, R., Asiedu, K., Aniweh, Y. and Awandare, G. A. (2018). Functional characterization of Plasmodium falciparum surface-related antigen as a potential blood-stage vaccine target. *The Journal of Infectious Diseases*, *218*(5), 778-790. doi.org/10.1093/infdis/jiy222

Arora, N., C Anbalagan, L. and Pannu, A. K. (2021). Towards eradication of malaria: is the WHO's RTS, S/AS01 vaccination effective enough? *Risk Management and Healthcare Policy*, 1033-1039. doi: 10.2147/RMHP.S219294.

Aurrecoechea, C., Brestelli, J., Brunk, B. P., Dommer, J., Fischer, S., Gajria, B., Gao, X., Gingle, A., Grant, G., Harb, O. S., Heiges, M., Innamorato, F., Iodice, J., Kissinger, J. C., Kraemer, E., Li, W., Miller, J. A., Nayak, V., Pennington, C., . . . Wang, H. (2009). PlasmoDB: a functional genomic database for malaria parasites. *Nucleic Acids Res*,

37(Database issue), D539-543. https://doi.org/10.1093/nar/gkn814

Barber, B. E., Grigg, M. J., Piera, K. A., William, T., Cooper, D. J., Plewes, K., Dondorp, A. M., Yeo, T. W. and Anstey, N. M. (2018). Intravascular haemolysis in severe Plasmodium knowlesi malaria: association with endothelial activation, microvascular dysfunction, and acute kidney injury. *Emerg Microbes Infect*, *7*(1), 106. https://doi.org/10.1038/s41426-018-0105-2

Bhasin, M., and Raghava, G. P. (2004). Prediction of CTL epitopes using QM, SVM and ANN techniques. *Vaccine*, *22*(23-24), 3195-3204. https://doi.org/10.1016/j.vaccine.2004.02.005

Birkett, A. J., Moorthy, V. S., Loucq, C., Chitnis, C. E. and Kaslow, D. C. (2013). Malaria vaccine R&D in the Decade of Vaccines: breakthroughs, challenges and opportunities. *Vaccine*, *31 Suppl 2*, B233-243. https://doi.org/10.1016/j.vaccine.2013.02.040

Blom, N., Sicheritz-Ponten, T., Gupta, R., Gammeltoft, S. and Brunak, S. (2004). Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence. *Proteomics*, *4*(6), 1633-1649. https://doi.org/10.1002/pmic.200300771

Burns, A. L., Dans, M. G., Balbin, J. M., de Koning-Ward, T. F., Gilson, P. R., Beeson, J. G., Boyle, M. J. and Wilson, D. W. (2019). Targeting malaria parasite invasion of red blood cells as an antimalarial strategy. *FEMS microbiology reviews*, *43*(3), 223-238. doi: 10.1093/femsre/fuz005.

Chou, P. Y., and Fasman, G. D. (1979). Prediction of beta-turns. *Biophys J*, *26*(3), 367-383. https://doi.org/10.1016/S0006-3495(79)85259-5

Chuh, K. N., Batt, A. R. and Pratt, M. R. (2016). Chemical Methods for Encoding and Decoding of Posttranslational Modifications. *Cell Chem Biol*, *23*(1), 86-107. https://doi.org/10.1016/j.chembiol.2015.11.006

Colovos, C., and Yeates, T. O. (1993). Verification of protein structures: patterns of nonbonded atomic interactions. *Protein Sci*, *2*(9), 1511-1519. https://doi.org/10.1002/pro.5560020916

Cowman, A. F., Tonkin, C. J., Tham, W. H. and Duraisingh, M. T. (2017). The Molecular Basis of Erythrocyte Invasion by Malaria Parasites. *Cell Host Microbe*, *22*(2), 232-245. https://doi.org/10.1016/j.chom.2017.07.003

Deléage, G. (2017). ALIGNSEC: viewing protein secondary structure predictions within large multiple sequence alignments. *Bioinformatics*, *33*(24), 3991-3992. doi: 10.1093/bioinformatics/btx521. PMID: 28961944.

Dephoure, N., Gould, K. L., Gygi, S. P. and Kellogg, D. R. (2013). Mapping and analysis of phosphorylation sites: a quick guide for cell biologists. *Mol Biol Cell*, *24*(5), 535-542. https://doi.org/10.1091/mbc.E12-09-0677

Dimitrov, I., Flower, D. R. and Doytchinova, I. (2013). AllerTOP-a server for in silico prediction of allergens. *BMC Bioinformatics* **14** (Suppl 6), S4 (2013). https://doi.org/10.1186/1471-2105-14-S6-S4

Doytchinova, I. A., and Flower, D. R. (2007). VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC bioinformatics*, *8*, 4. https://doi.org/10.1186/1471-2105-8-4

Egbewande, O. M. (2022). The RTS,S malaria vaccine: Journey from conception to recommendation. *Public Health Pract (Oxf)*, *4*, 100283. https://doi.org/10.1016/j.puhip.2022.100283

Emini, E. A., Hughes, J. V., Perlow, D. and Boger, J. (1985). Induction of hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide. *Journal of virology*, *55*(3), 836-839. doi: 10.1128/JVI.55.3.836-839.

Ferre, F., and Clote, P. (2005). DiANNA: a web server for disulfide connectivity prediction. *Nucleic Acids Res*, *33*(Web Server issue), W230-232. https://doi.org/10.1093/nar/gki412

Ghaffari, A. D., Dalimi, A., Ghaffarifar, F. and Pirestani, M. (2020). Structural predication and antigenic analysis of ROP16 protein utilizing immunoinformatics methods in order to identification of a vaccine against Toxoplasma gondii: An in silico approach. *Microb Pathog*, *142*, 104079. https://doi.org/10.1016/j.micpath.2020.104079

Gardner MJ, Hall N, Fung E, et al. Genome sequence of the human malaria parasite *Plasmodium falciparum*. Nature 2002; 419:498–511. doi: 10.1038/nature01097.

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S. e., Wilkins, M. R., Appel, R. D. and Bairoch, A. (2005). *Protein identification and analysis tools on the ExPASy server*. Springer. doi.org/10.1385/1-59259-890-0:571

Heo, L., Park, H. and Seok, C. (2013). GalaxyRefine: Protein structure refinement driven by side-chain repacking. *Nucleic Acids Res*, *41*(Web Server issue), W384-388. https://doi.org/10.1093/nar/gkt458

Karplus, P., and Schulz, G. (1985). Prediction of chain flexibility in proteins: a tool for the selection of peptide antigens. *Naturwissenschaften*, *72*(4), 212-213.

Kolaskar, A. S., and Tongaonkar, P. C. (1990). A semi-empirical method for prediction of antigenic determinants on protein antigens. *FEBS Lett*, *276*(1-2), 172-174. https://doi.org/10.1016/0014-5793(90)80535-q

Krogh, A., Larsson, B., von Heijne, G. and Sonnhammer, E. L. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol*, *305*(3), 567-580. https://doi.org/10.1006/jmbi.2000.4315

Larsen, J. E., Lund, O. and Nielsen, M. (2006). Improved method for predicting linear B-cell epitopes. *Immunome Res*, *2*, 2. https://doi.org/10.1186/1745-7580-2-2

McGuffin, L. J., Bryson, K. and Jones, D. T. (2000). The PSIPRED protein structure prediction server. *Bioinformatics*, *16*(4), 404-405. https://doi.org/10.1093/bioinformatics/16.4.404

Nakai, K., and Horton, P. (1999). PSORT: a program for detecting sorting signals in proteins and predicting their

subcellular localization. *Trends in Biochemical Sciences*, *24*(1), 34-35. doi: 10.1016/s0968-0004(98)01336-x.

Nezafat, N., Ghasemi, Y., Javadi, G., Khoshnoud, M. J. and Omidinia, E. (2014). A novel multi-epitope peptide vaccine against cancer: an *in silico* approach. *J Theor Biol*, *349*, 121-134. https://doi.org/10.1016/j.jtbi.2014.01.018

Pannu, A. K. (2019). Malaria today: advances in management and control. *Trop Doct*, *49*(3), 160-164. https://doi.org/10.1177/0049475519846382

Phillips, M., Burrows, J., Manyando, C., van Huijsduijnen, R. H., Van Voorhis, W. and Wells, T. (2017). Malaria Nat Rev Dis Primers, 3 (2017). 17050. doi.org/10.1038/nrdp.2017.50

Ren, J., Wen, L., Gao, X., Jin, C., Xue, Y. and Yao, X. (2008). CSS-Palm 2.0: an updated software for palmitoylation sites prediction. *Protein Eng Des Sel*, *21*(11), 639-644. https://doi.org/10.1093/protein/gzn039

Reynisson, B., Alvarez, B., Paul, S., Peters, B. and Nielsen, M. (2020). NetMHCpan-4.1 and NetMHCIIpan-4.0: improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data. *Nucleic Acids Res*, *48*(W1), W449-W454. https://doi.org/10.1093/nar/gkaa379

Romano, P., Giugno, R. and Pulvirenti, A. (2011). Tools and collaborative environments for bioinformatics research. *Brief Bioinform*, *12*(6), 549-561. https://doi.org/10.1093/bib/bbr055

Roy, A., Kucukural, A. and Zhang, Y. (2010). I-TASSER: a unified platform for automated protein structure and function prediction. *Nat Protoc*, *5*(4), 725-738. https://doi.org/10.1038/nprot.2010.5

Saha, S., and Raghava, G. P. (2006). Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. *Proteins*, *65*(1), 40-48. https://doi.org/10.1002/prot.21078

Saha, S., and Raghava, G. P. S. (2004). BcePred: prediction of continuous B-cell epitopes in antigenic sequences using physico-chemical properties. International conference on artificial immune systems, ICARIS 2004. Lecture Notes in Computer Science, vol 3239. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-30220-9_16

Talapko, J., Skrlec, I., Alebic, T., Jukic, M. and Vcev, A. (2019). Malaria: The Past and the Present. *Microorganisms*, *7*(6), 179. https://doi.org/10.3390/microorganisms7060179

Varo, R., Chaccour, C. and Bassat, Q. (2020). Update on malaria. *Med Clin (Barc)*, *155*(9), 395-402. https://doi.org/10.1016/j.medcli.2020.05.010

Vita, R., Mahajan, S., Overton, J. A., Dhanda, S. K., Martini, S., Cantrell, J. R., Wheeler, D. K., Sette, A. and Peters, B. (2019). The Immune Epitope Database (IEDB): 2018 update. *Nucleic Acids Res*, *47*(D1), D339-D343. https://doi.org/10.1093/nar/gky1006

WHO. (2021). World malaria report 2021. In: World Health Organization.

Wiederstein, M., and Sippl, M. J. (2007). ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res*, *35*(Web Server issue), W407-410. https://doi.org/10.1093/nar/gkm290

Winther, J. R., and Thorpe, C. (2014). Quantification of thiols and disulfides. *Biochim Biophys Acta*, *1840*(2), 838-846. https://doi.org/10.1016/j.bbagen.2013.03.031

Wu, S., and Zhang, Y. (2007). LOMETS: a local meta-threading-server for protein structure prediction. *Nucleic Acids Res*, *35*(10), 3375-3382. https://doi.org/10.1093/nar/gkm251

Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J. and Zhang, Y. (2015). The I-TASSER Suite: protein structure and function prediction. *Nat Methods*, *12*(1), 7-8. https://doi.org/10.1038/nmeth.3213

Zheng, W., Zhang, C., Wuyun, Q., Pearce, R., Li, Y. and Zhang, Y. (2019). LOMETS2: improved meta-threading server for fold-recognition and structure-based function annotation for distant-homology proteins. *Nucleic Acids Res*, *47*(W1), W429-W436. https://doi.org/10.1093/nar/gkz384