Vaccine Research
ISSN: 2423-4923
eISSN: 2383-2819

In-silico Analysis of Pneumococcal Heat-Shock Protein (DnaJ) to Predict Novel Multi-Epitope Vaccine Candidates

Elnaz Afshari[1], Reza Ahangari Cohan[2], Fattah Sotoodehnejadnematalahi[3], Seyed Fazlollah Mousavi[4,1]*

[1]Department of Biology, Science and Research Branch, Islamic Azad University, Tehran, Iran. [2]Department of Nanobiotechnology, New Technologies Research Group, Pasteur Institute of Iran, Tehran, Iran. [3]Department of Biology, Science and Research Branch, Islamic Azad University, Tehran, Iran. [4]Department of Microbiology, Pasteur Institute of Iran, Tehran, Iran.

ARTICLE INFO

Research Article

VacRes, 2022
Vol. 9, No.1, 11- 33
Received: June 09, 2022
Accepted: July 16, 2022
Pasteur Institute of Iran

***Corresponding Author:**
Seyed Fazlollah Mousavi;
Microbiology Department, Pasteur Institute of Iran, 69 Pasteur Ave. Tehran-13164, Iran.
Email: mousavi@pasteur.ac.ir
Tel: (+98) 2166402233

KEYWORDS: *Streptococcus pneumoniae*, DnaJ, Vaccine Candidate, *In-silico*

ABSTRACT

Introduction: To prevent pneumococcal infections, especially meningitis and bacteremia, and to overcome the serotype-dependent limitation of polysaccharide-based vaccines, the development of conserved protein-based vaccines is essential. This study aimed at investigate the *in-silico* analysis and epitope mapping of pneumococcal DnaJ for the first time, and to design the multi-epitope based vaccines with different categories by focusing on induction of both humoral and cellular immunities. **Methods:** We predicted B- and T-cell epitopes, IL-4, IL-17, IL-10, and IFN-γ inducer epitopes of DnaJ using Immunoinformatics tools. The immunogenicity and conservation score of the predicted epitopes among pneumococcal prevalent clinical serotypes, the immune simulation of DnaJ administration in mammals and potential regions involved in DnaJ-TLRs interactions were analyzed. Finally, we proposed three classes of multi-epitope DnaJ-based vaccine candidates. **Results:** This protein had 24 and 15 predicted linear B-cell and helper T-cell epitopes, respectively, with a conservation score of 86-100% among prevalent clinical pneumococcal serotypes. DnaJ also had many IL-4 and IFN-γ inducing epitopes and was considered an IL-10 and IL-17 inducer protein. The immune simulation showed induction of both humoral and cellular immunity against DnaJ. The residues at positions 274, 280, 292, 297, 300, 316-319, 333, 336-340, 358, 363-366, and 372 were predicted to be involved in DnaJ-TLR2 and DnaJ-TLR4 interactions. Three classes of proposed DnaJ-based constructs were based on only B-cell epitopes, only helper T-cell epitopes, and multi-epitopes of B- and T-cell and IL-17 epitopes. **Conclusion:** The results showed that although DnaJ has been reported to play an important role in cellular immunity, our results indicated the high potential of DnaJ to stimulate mucosal, humoral, and cellular immunity.

INTRODUCTION

Streptococcus pneumoniae (pneumococcus) is a Gram-positive pathogen of the respiratory system that causes a variety of non-invasive and invasive diseases with a high annual mortality rate in children under 5 years of age and the elderly or with compromised immune systems [1, 2]. The prevalence of pneumococcal infections has been reported differently around the world, and all deaths caused by pneumococci in children have been estimated at approximately 11% (8-12%) annually in developing countries [3]. In Iran, the prevalence of pneumococcus in carriers and patients are varied and has been reported as 20% and 1.5%, respectively [4]. In addition, since bacterial co- or secondary respiratory infections, especially pneumonia, have been regularly observed in influenza pandemics [5], pneumococcal respiratory co- or super-infections are gaining importance in the global pandemic of COVID-19 from 2020 [6]. Many studies have reported that *S. aureus* and *S. pneumoniae* were the most common pathogens with higher early co-infection in the first 48 hours from admission to the unit [6, 7].

Licensed vaccines have many limitations for the prevention of pneumococcal diseases, such as poor immunogenicity and low coverage of 98 pneumococcal serotypes, replacement of non-vaccine serotypes and the prevalence of non-encapsulated pneumococci in patients [8-11]. Many surface proteins have been studied as protein-based pneumococcal vaccine candidates. However, few of them have

been analyzed in clinical trials and their protective effects are still limited to a subset of pneumococcal strains [12, 13]. The candidate protein should mostly be well-conserved among pneumococcal serotypes and be able to stimulate T-helper cells, resulting in a strong memory response [13].

Prokaryotic heat-shock proteins (HSPs) are highly conserved molecular chaperones that are a virulence factor necessary for maintaining protein homeostasis during various stress processes. They contribute to bacterial attachment and invasion during the infections [14-16]. Pneumococcal DnaJ is a virulence factor belonging to a member of the HSP40 family [15]. The main feature of this protein is the presence of a highly conserved J-domain, containing 70 amino acids in its N-terminus with four helical structures [17]. Several studies have shown that immunization with DnaJ can stimulate cellular and mucosal immune responses in mice [18-21] and elevate protection against nasopharyngeal colonization in nasal or lung and invasive infections, caused by various pneumococcal serotypes [13]. Pneumococci are an extracellular pathogen and humoral immunity plays a major role in the clearing of pneumococci from invasive infections, especially bacteremia or meningitis. Therefore, Immunoinformatics-based analysis of DnaJ can help to understand its immunogenic profile, with a focus on regions that can stimulate the B- and T-helper2 cell-based immunity. Here, we used Immunoinformatics tools to characterize B- and T-cell epitopes, antigenicity, allergenicity, toxicity, as well as conservation of epitopes among clinical predominant serotypes of pneumococci (serotypes 1, 3, 4, 6A, 6B, 7F, 9V,10A, 11A, 12F, 14, 15B, 18C, 19F, 19A, 22A/F, 23F and 33F) DnaJs. We also analyzed the homology of DnaJ within the human proteome, cytokine inducer epitopes (IL-4, IL-10, IL-17, and IFN-γ), the mammalian immune system simulation using DnaJ administration, and potential regions of DnaJ involved in interacting with MHC-II, TLR-2, and TLR-4. Finally, we introduced the proposed DnaJ-based constructs to induce mucosal and humoral immunities against pneumococcal invasive infections.

MATERIALS AND METHODS

Sequence Retrieval and Initial Analysis

Reference sequence of *S. pneumoniae* DnaJ (WP_001066302.1) was retrieved from the National Center for Biotechnology Information (NCBI) https://www.ncbi.nlm.nih.gov. We used online server CELLO v.2.5 (http://cello.life.nctu.edu.tw/) [22], signalP-5.0 (https://services.healthtech.dtu.dk/service.php?SignalP-5.0) [23], and TMHMM (https://services.healthtech.dtu.dk/service.php?TMHMM-2.0) [24], to predict subcellular localization, signal peptide, and transmembrane helices of DnaJ, respectively. For achieving the conserved regions of the DnaJ sequence among the clinical prevalent serotypes of *S. pneumoniae* DnaJ (serotypes 1, 3, 4, 6A, 6B, 7F, 9V,10A, 11A, 12F, 14, 15B, 18C, 19F, 19A, 22A/F, 23F and 33F), the alignment of the DnaJ amino acid sequences was performed using Clustal Omega server (https://www.ebi.ac.uk/Tools/msa/clustalo/) [25].

Physicochemical Properties and Solubility Prediction

Using the Protparam online server (https://web.expasy.org/protparam/) and SOLpro server (http://scratch.proteomics.ics.uci.edu/), the physicochemical properties and the solubility upon overexpression of proteins in *E. coli* were evaluated, respectively [26, 27].

Secondary Structure Prediction, Tertiary Structure Homology Modeling, Refinement, and Validation

The secondary structure of DnaJ was predicted by GORV (https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_gor4.html) and PSIPRED 4.0 (http://bioinf.cs.ucl.ac.uk/psipred/)[28]. Since the 3D structure of DnaJ was partially modeled in the RCSB server (www.rcsb.org) with the PDB ID 6Jzb, we used the I-TASSER server (https://zhanggroup.org/I-TASSER/)[29] to model the entire DnaJ 3D structure. The best predicted model was selected and refined using the Galaxy Refine server (http://galaxy.seoklab.org/cgi-bin/submit.cgi?type=REFINE/)[30]. The final model was validated with ProSA (https://prosa.services.came.sbg.ac.at/prosa.php/) [31], PROCHECK, and ERRAT tools from the Saves server (https://saves.mbi.ucla.edu/)[32], to recognize the errors in the generated 3D model. PyMol software v.2.5 was used to exhibit a high-quality image of the predicted model.

Antigenicity, Toxicity, and Allergenisity Predictions

The antigenicity, toxicity, and allergenisity of DnaJ were predicted using the Vaxijen-v2 (http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html) [25] with cut-off ≥0.6, ToxinPred (https://webs.iiitd.edu.in/raghava/toxinpred/index.html) [33], and AlgPred 2.0 web tool (https://webs.iiitd.edu.in/raghava/algpred2/batch.html) [34], respectively.

Homology Analysis of Proteins with the Human Proteome

We used the PIR peptide matching program (https://research.bioinformatics.udel.edu/peptidematch/index.jsp) [35] to evaluate similarity of the DnaJ amino acid sequence with the human proteome.

Immunoinformatics Analyses

Determination of Potential B-cell epitopes

The BCpred (http://ailab.ist.psu.edu/bcpred/predict.html) [36] and IEDB servers (https://www.iedb.org/) [37] were applied to predict linear B-cell epitopes, surface accessibility, and antigenicity of DnaJ. The 3D structural of a validated 3D model of DnaJ was used for conformational B-cell epitope prediction using Ellipro sever (http://tools.iedb.org/ellipro/)[38].

Determination of Helper T-cell (HTL) Epitopes

IEDB server (http://tools.iedb.org/mhcii/)[39] was used for predicting MHC-II binding epitopes. We predicted 15 mer peptide binding affinities to eight Iranian HLA-II alleles [HLA-DRB* (01:01- 03:01- 04:01- 07:01- 08:01- 11:01-13:01-15:01)], as well as mouse MHC-II H2 alleles (I-Ad, I-Ab, and I-Ed) [40, 41, 28, 42].

Evaluation of Conservation of Predicted B- and T-Cell Epitope among Prevalent Clinical Serotypes of *Streptococcus pneumonia*

The IEDB Epitope Conservancy database (http://tools.iedb.org/conservancy/) [43] was used to predict

conservancy of predicted B- cell and HTL epitopes of DnaJ among clinical prevalent serotypes of *S. pneumoniae* (serotypes 1, 3, 4, 6A, 6B, 7F, 9V,10A, 11A, 12F, 14, 15B, 18C, 19F, 19A, 22A/F, 23F, and 33F) DnaJs.

Prediction of Cytokine Inducer Epitopes

For evaluating the effective mucosal and humoral immune responses against pneumococcal DnaJ, prediction of IL-4, IL-10, IL-17, and IFN-γ inducer epitopes were performed using IL4pred server (https://webs.iiitd.edu.in/raghava/il4pred/scan.php), IL-10Pred server (https://webs.iiitd.edu.in/raghava/il10pred/predict3.php), IL17eScan (http://metagenomics.iiserb.ac.in/IL17eScan/index.html), and IFNepitope server (https://webs.iiitd.edu.in/raghava/ifnepitope/predict.php), respectively [44, 45] .

***In-silico* Immune Response Simulation**

C-ImmSim server (http://150.146.2.1/C-IMMSIM/index.php) was used to *In-silico* immune simulation against DnaJ. Three injections of the DnaJ at intervals of 4 weeks, on days 1, 30, and day 60 was administered with entire simulation ran 1400 time steps (about 15 months) and the simulation volume of 10. HLA alleles of parameters were also set based on predominant human HLA alleles (HLA-A*1101, HLA-B*3501, and HLA-DRB1*0101) [46, 47, 42, 28].

Determination of the Potential Regions involved in DnaJ-TLRs and DnaJ-MHC Interactions using Molecular Docking

It was reported that DnaJ contributes to the activation of bone marrow-derived dendritic cells (BMDCs) and phagocytosis of macrophage, leading to the development of innate, humoral and cellular immunities through Toll-like receptors 4 and 2 (TLR4 and TLR2) [20, 21, 48]. For a more detailed understanding of the amino acid residues of DnaJ involved in the interacting interface with TLRs and MHC molecules, molecular docking was performed. For this purpose, the PDB structure of HLA-DR1 (DRB1*0101), TLR2, and TLR4 were retrieved from RCSB server (www.rcsb.org) with PDB ID 1AQD, 6NIG, and 3FXI, respectively [49]. Then PDB structures were refined by eliminating the ligand and H_2O from the structure using UCSF Chimera v.1.14 software. Docking simulations of DnaJ with above PDB structures were performed using ClusPro 2.0 server (http://nrc.bu.edu/cluster/). The model of ClusPro with the largest cluster size and the lowest binding free energy were checked for the interaction of amino acids using the DimPlot tool in LigPlot+ v.2.2.4 software and PyMolv. 2.5 software.

Designing of the DnaJ-based Constructs

We designed several multi-epitope constructs based on predicted B- and T-cell epitopes of DnaJ. The percentage of immunogenicity according to the VaxiJen score and its percentage of epitope conservation among DnaJ of clinical prevalent serotypes of pneumococci were considered as selection criteria. We also used IL-17 inducer epitope to design multi-epitope construct to induce and promote both mucosal and humoral immune responses against pneumococci.

Evaluation of Physicochemical Properties, Antigenicity, and other characteristics of proposed DnaJ-based constructs

We analyzed the five proposed DnaJ-based constructs for physicochemical properties, antigenicity, toxicity, solubility, transmembrain helix using ProtParam server (https://web.expasy.org/protparam/), Vaxijen-v2 (http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html) with cut-off ≥0.6, ToxinPred (https://webs.iiitd.edu.in/raghava/toxinpred/index.html), SOLpro server (http://scratch.proteomics.ics.uci.edu/), and TMHMM server (https://services.healthtech.dtu.dk/service.php?TMHMM-2.0), respectively.

Secondary and Tertiary Structure Prediction of the Constructs

The secondary structure of five proposed DnaJ-based constructs were predicted by GORV (https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_gor4.html) and PSIPRED 4.0 (http://bioinf.cs.ucl.ac.uk/psipred/). Then 3D structures of constructs were modeled using the I-TASSER server (https://zhanggroup.org/I-TASSER/).

Refinement and Validation of Tertiary Structure of the Constructs

The best predicted I-TASSER model for five proposed DnaJ-based constructs was selected and refined using the Galaxy Refine server (http://galaxy.seoklab.org/cgi-bin/submit.cgi?type=REFINE/)[30]. The final models were validated with ProSA (https://prosa.services.came.sbg.ac.at/prosa.php/) [31], PROCHECK, and ERRAT tools from the Saves server (https://saves.mbi.ucla.edu/)[32]. PyMol software v.2.5 was used to exhibit a high-quality image of the predicted model.

RESULTS

Sequence Retrieval and Initial Analysis

The DnaJ sequence was retrieved from the NCBI server. The results of the DnaJ subcellular localization, transmembrane topology, and signal peptide are shown in Table 1.

The result of the Clustal Omega alignment showed that the DnaJ amino acid sequence was highly conserved among clinical prevalent serotypes of *S. pneumoniae* (serotypes 1, 3, 4, 6A, 6B, 7F, 9V,10A, 11A, 12F, 14, 15B, 18C, 19F, 19A, 22A/F, 23F, and 33F). Among the 23 DnaJ sequences examined, only 11 of the 387 amino acids had a variable sequence. The position of variable amino acids in DnaJ includes amino acids 70, 73, 197, 201, 224, 297, 315, 355, 358 and 368.

Table 1. Structural and Physicochemical Characters of DnaJ

Structural and Physicochemical Characters							
Amino Acid Sequence	MNNTEFYDRLGVSKNASADEIKKAYRKLSKKYHPDINKEPGAEDKYKEVQEAYETLSDDQKRAAYDQYGAAGANGG FGGFNGAGGFGGFEDIFSSFFGGGGSSRNPNAPRQGDDLQYRVNLTFEEAIFGTEKEVKYHREAGCRTCNGSGAKPGTS PVTCGRCHGAGVINVDTQTPLGMMRRQVTCDVCHGRGKEIKYPCTTCHGTGHEKQAHSVHVKIPAGVETGQQIRLAG QGEAGFNGGPYGDLYVVVSVEASDKFEREGTTIFYNLNLNFVQAALGDTVDIPTVHGDVELVIPEGTQTGKKFRLRSKG APSLRGGAVGDQYVTVNVVTPTGLNDRQKVALKEFAAAGDLKVNPKKKGFFDHIKDAFDGE						
Trans-Membrane Helices	0	**VaxiJen Score**	0.7953 (Probable : ANTIGEN)	**Theoretical pI**	6.80	**(Asp + Glu)**	47
Sub-cellular Localization	Outer Membrane	**Toxicity**	Non Toxic	**Solubility Index**	0.92 (soluble)	**(Arg + Lys)**	46
Signal Peptide	0	**Allergenicity**	Non Allergic	**Instability index**	23.53 (stable)	**GRAVY**	-0.603
Molecular Weight (KDa)	40	**Similarity to Human Proteome**	No	**Aliphatic Index**	60.3	**Estimated Half-life**	30 hours (mammalian reticulocytes, *in vitro*). >20 hours (yeast, *in vivo*). >10 hours (*E. coli*, *in vivo*).

Physicochemical Properties

The results of physicochemical properties are shown in Table 1. Using the ProtParam server, the molecular weight (Mw) of DnaJ was predicted to be 40 kDa. According to the theoretical isoelectric point (pI) value, DnaJ was expected to be neutral in nature. Using the Solpro server, DnaJ was predicted to be soluble when overexpressed in *E. coli*. According to the instability index, DnaJ was classified as a stable protein (II of <40 indicates stability). The estimated negative GRAVY value of DnaJ means that the protein has hydrophilic nature and is able to interact with water molecules.

Secondary Structure Prediction

Using GOR V prediction server, we found that the DnaJ secondary structure contains 26.34% alpha-helix, 20.43% extended strand, and 53.23% random coil. The secondary structure of DnaJ using PSIPRED prediction is shown in Fig. 1.

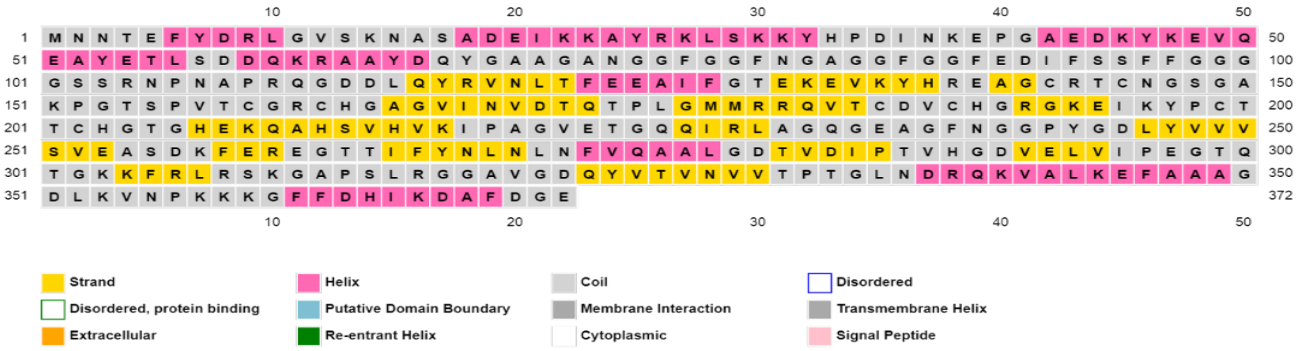

Fig. 1. Graphical representation of features of secondary structure of DnaJ sequence using PSIPRED server. The protein is predicted to comprise alpha helices (26.34%), beta strands (20.43%) and coils (53.23%). The alpha helix residues are pink, the beta strand residues are yellow and the coil residues are gray.

Tertiary Structure Homology Modeling, Refinement and Validation

I-TASSER server predicted five models for tertiary structure of DnaJ based on 10 threading templates. Model 1, with the highest C-score of -1.16 was selected for further refinement (Fig. 2a). This model had an estimated TM-score and RMSD of 0.52 ± 0.15 and 10.4 ± 4.6 Å, respectively. After refinement of predicted model using Galaxy Refine server, model one showed that the GDT-HA score was 0.928, which is usually between [0, 1] values, and a higher score indicates conservative refinement. RMSD score (0.475) and MolProbity score (2.308) indicate aggressive refinement and a more physically realistic model, respectively. The clash score and poor rotamers were 21 and 1, respectively. Ramachandran plot analysis of DnaJ model in procheck server revealed that among the 372 residues, 263 (86.8%) and 37 (12.3%) in the protein were in the most favored and in allowed regions, respectively. Only three residues (1%) were in the disallowed region, indicating that the predicted model is acceptable (Fig. 2b). The quality and potential errors in the 3D model were verified by ProSA-web and ERRAT. The overall quality factor of the chosen model after the refinement was 83.85% and ProSA-web

Z-score was −6.88 (Fig. 2b). The DnaJ fell close in the score range commonly found in native proteins of comparable size. The details of quality of DnaJ structure before and after refinement are shown in Table 2.

Table 2. The results of validation of DnaJ structure before and after refinement using Ramachandran plot statistics from PROCHECK and Z-score from ProSA web servers.

ITASSER	Before Refinement					After Refinement				
			Ramachandran plot					Ramachandran plot		
C-score	ERRAT quality factor	Z-score	Most favored zones (%)	Allowed zones (%)	Disallowed zones (%)	ERRAT quality factor	Z-score	Most favored zones (%)	Allowed zones (%)	Disallowed zones (%)
-1.16	87.87	-6.54	68.3	29.4	2.3	83.85	-6.88	86.8	12.3	1

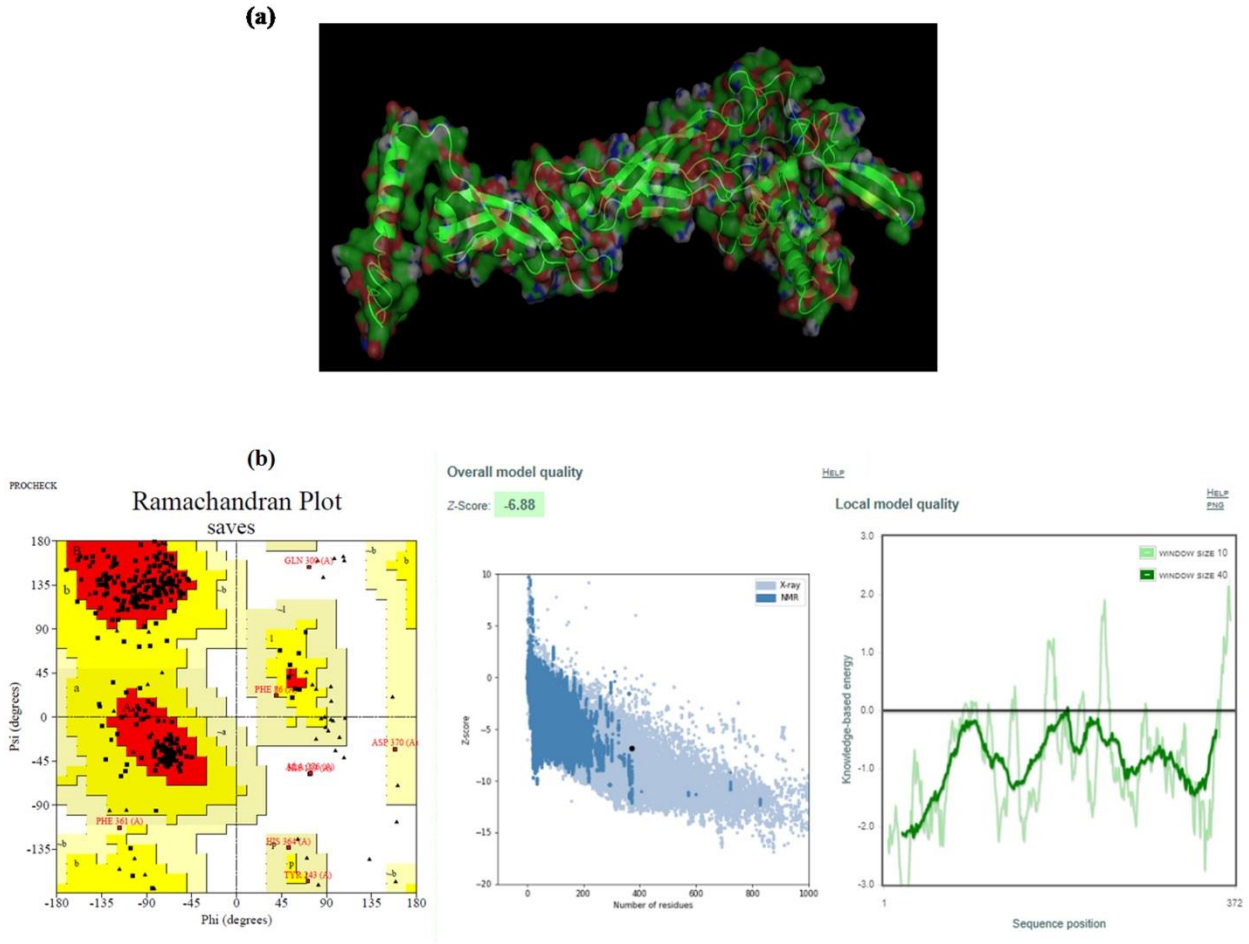

Fig. 2. 3D homology modeling and validation of DnaJ. **(a)** The I-Tasser 3D homology modeling of the DnaJ was displayed by PyMol v.2.5 software. **(b)** Validation of the model after refinement with Ramachandran plot and ProSA web plot. ProSA-web analysis shows the Z-score of -6.88, and the plot of residue scores that shows local model quality by plotting energies as a function of amino acid sequence position are also shown.

Antigenicity, Toxicity, Allergenisity Predictions of DnaJ

The results in Table 1 shows that DnaJ is immunogen, nontoxic and non-allergic. The results of the PIR peptide-matching program showed that the DnaJ had no similarity with the human proteome.

B-cell Epitope Prediction

The linear B-cell epitopes with high antigenicity, surface accessibility, and flexibility of DnaJ were predicted using BCPred and IEDB servers and provided in Table 3.

Table 3. Prediction of linear and discontinuous B-cell epitopes, antigenicity, surface accessibility, and epitope conservancy of DnaJ using BCpred and IEDB servers (Bepipred, Emini, kolaskar, Ellipro, and Epitope Conservancy tools of IEDB).

Bepipred Linear B-cell Epitope Prediction for DnaJ

Peptide Sequence	Position	VaxiJen Score	Conservation%
*EFYDRLGVSKNASADEIKKA	5-24	0.7149	100
RKLSKKYHPDINKEPGAEDKYKEVQEAYETLSDDQKRA AYDQYGAAGANGGFGGFNG	26-82	0.7540	8.7
*RNPNAPRQGDDLQYR	104-118	1.3442	100
FGTEKEVKYHREAGCRTCNGSGAKPGTSPVTCGRCHGA GVINVDTQTPLGMMRRQVTCDVCHGRGKE	128-194	0.8029	82.61
*GTGHEK	204-209	3.8809	100
*QGEAGFNGGP	233-242	0.6673	100
GTQTGKKFRLRSKGAPSLRGGAV	298-320	1.4225	86.96
*GLNDRQKV	334-341	0.4303	100
DLKVNPKKKGFF	351-362	0.5124	82.61

Ellipro Linear B-cell Epitope Prediction for DnaJ

Peptide Sequence	Position	Score	Conservation%
VVTPTGLNDRQKVALKEFAAAGDLKVNPKKKGFFDHIKDAFDGE	329-372	0.873	65.22
YDRLGVSKNASADEIKKAYRKLSKKYHPDINKEPGAEDKY	7-46	0.768	91.3
*SGAKPGTSPVTCGRCHGAGVINVDTQTPLGMMRRQVTCDVCHG	148-190	0.724	100
GKEIKYPCTTCHGT	192-205	0.684	91.3
*NFVQAALGDTVD	272-283	0.625	100
HGDVELVIPEGTQTGK	288-303	0.569	91.3
DQYGAAGAN	66-74	0.56	86.96
*GCRTCN	141-146	0.522	100

BCpred B-cell Epitope Prediction for DnaJ

Peptide Sequence	Position	Score	Conservation%
GAAGANGGFGGFNGAGGFGG	69-89	1	8.7
*AGQGEAGFNGGPYGDLYVVV	231-251	1	100
*FFGGGGSSRNPNAPRQGDDL	96-116	1	100
*GCRTCNGSGAKPGTSPVTCG	141-161	1	100
DINKEPGAEDKYKEVQEAYE	35-55	0.996	91.3
*GGAVGDQYVTVNVVTPTGLN	317-337	0.965	100
IPEGTQTGKKFRLRSKGAPS	295-315	0.959	86.96
GKEIKYPCTTCHGTGHEKQA	192-212	0.85	91.3
*VINVDTQTPLGMMRRQVTCD	167-187	0.793	100

Kolaskar & Tongaonkar Antigenicity for DnaJ (Threshold: 1.014)

Peptide Sequence	Position	Peptide Sequence	Position
KEVQEAYET	47-55	GDLYVVVSVEAS	244-255
QYRVNLT	116-122	LNFVQAA	271-277
TSPVTCGRC	154-162	GDTVDIPTVHGDVELVIPE	279-297
GAGVINV	164-170	GDQYVTVNVVTPT	321-333
QVTCDVCHG	182-190	KVALKEF	340-346
KYPCTTC	196-202	AGDLKVN	349-355
AHSVHVKIPAG	211-221		

Emini Surface Accessibility Prediction for DnaJ (Threshold: 1.00)

Peptide Sequence	Position	Peptide Sequence	Position
DEIKKAYRKLSKKYHPDINKEPGAEDKYKEVQEAY	19-53	ASDKFERE	254-261
TLSDDQKRAAYDQ	55-67	GTQTGKKFRLR	298-308
SSRNPNAPRQGDDLQYR	102-118	KVNPKKK	353-359
TEKEVKYHR	130-138		

*Epitopes with 100% conservation score.

According to B-cell epitope prediction servers, from 24 predicted linear B-cell epitopes, thirteen epitopes (which highlight in Table 3) had 100% conservation score among clinical prevalent serotypes of *S.pneumoniae* (serotypes 1, 3, 4, 6A, 6B, 7F, 9V, 10A, 11A, 12F, 14, 15B, 18C, 19F, 19A, 22A/F, 23F, and 33F) DnaJs. The antigenicity of predicted B-cell epitopes were analyzed using the Vaxijen-v2 server with cut-off ≥0.4. Some of the epitopes had VaxiJen antigenicity scores of 1 to 3.8. The predicted continuous B-cell epitopes (with score of >0.6) have been also shown in Fig. 3 and Table 4.

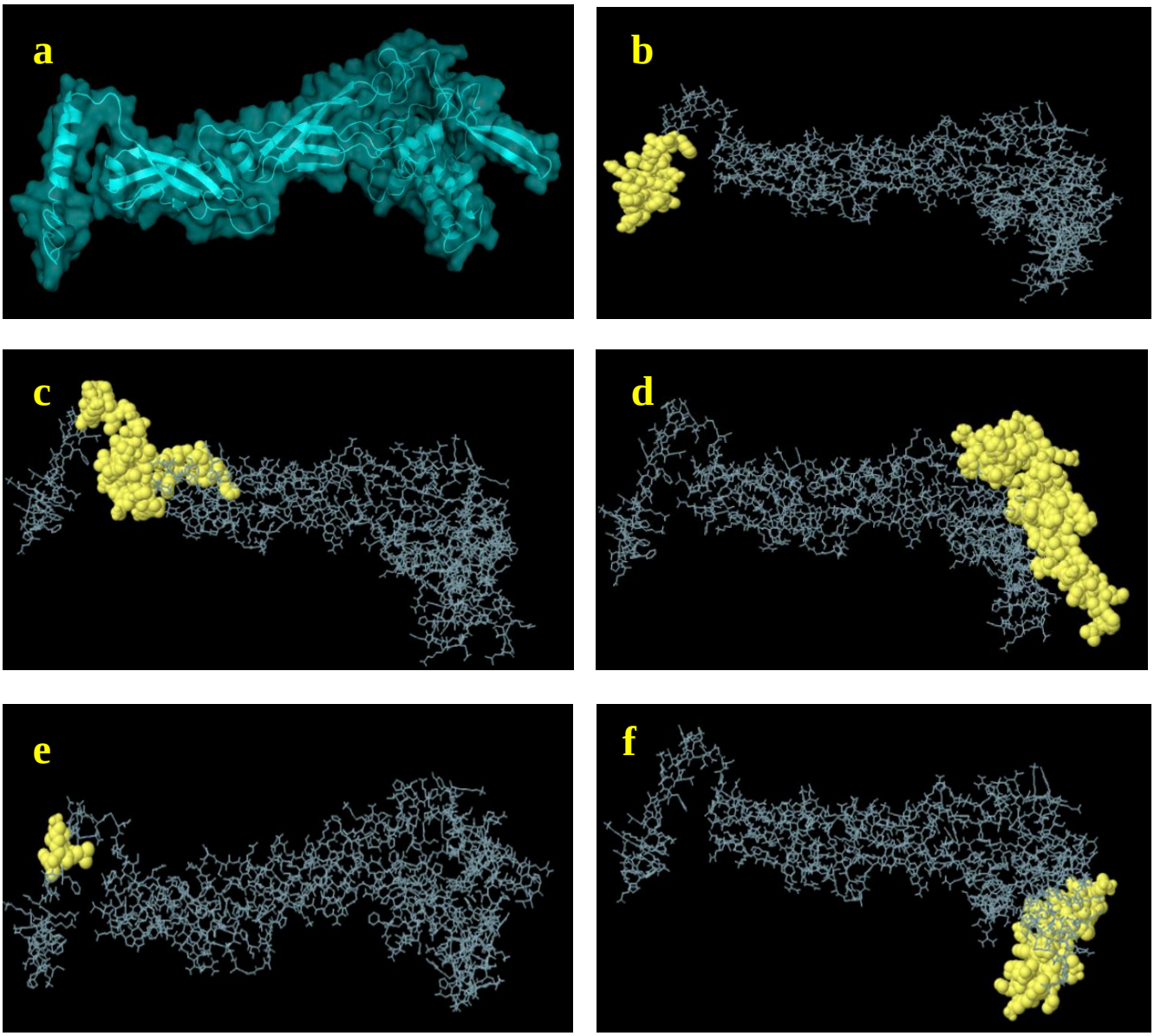

Fig. 3. The schematic results of Table 4 for predicted continuous B-cell epitopes of DnaJ using the IEDB server. a) The 3D structure of DnaJ for epitope mapping analysis. b to f) Predicted continuous B-cell epitopes from high to low antigenicity score. The residues with a score higher than the threshold (default value is 0.5) are predicted to be continuous B-cell epitopes and are colored yellow on the graph.

Table 4. Elipro Predicted Discontinuous B-cell Epitopes for DnaJ

No.	Residues	Number of residues	Score
1	A:F346, A:A347, A:A348, A:A349, A:G350, A:D351, A:L352, A:K353, A:V354, A:N355, A:P356, A:K357, A:K358, A:K359, A:G360, A:F361, A:F362, A:D363, A:H364, A:I365, A:K366, A:D367, A:A368, A:F369, A:D370, A:G371, A:E372	27	0.903
2	A:V341, A:A342, A:L343, A:K344, A:E345	5	0.859
3	A:Y7, A:L10, A:G11, A:V12, A:S13, A:K14, A:N15, A:A16, A:S17, A:A18, A:D19, A:E20, A:I21, A:K22, A:K23, A:A24, A:Y25, A:R26, A:K27, A:L28, A:S29, A:K30, A:K31, A:Y32, A:H33, A:P34, A:D35, A:I36, A:N37, A:K38, A:E39, A:P40, A:G41, A:A42, A:E43, A:D44, A:K45, A:Y46, A:V49, A:Y53	40	0.766
4	A:M1, A:N2, A:R62, A:A63, A:D66, A:Q67, A:Y68, A:G69, A:A70, A:A71, A:G72, A:A73, A:N74, A:G75, A:G141, A:C142, A:R143, A:T144, A:C145, A:N146, A:G147, A:S148, A:G149, A:A150, A:K151, A:P152, A:G153, A:T154, A:S155, A:P156, A:V157, A:T158, A:C159, A:G160, A:R161, A:C162, A:H163, A:G164, A:A165, A:G166, A:V167, A:I168, A:N169, A:V170, A:D171, A:T172, A:Q173, A:T174, A:P175, A:L176, A:G177, A:M178, A:M179, A:R180, A:R181, A:Q182, A:V183, A:T184, A:C185, A:D186, A:V187, A:C188, A:H189, A:G190, A:G192, A:K193, A:E194, A:I195, A:K196, A:Y197, A:P198, A:C199, A:T200, A:T201, A:C202, A:H203, A:G204, A:T205, A:G206, A:E208	80	0.657
5	A:N272, A:F273, A:V274, A:Q275, A:A276, A:A277, A:L278, A:G279, A:D280, A:T281, A:V282, A:D283, A:P285, A:T286, A:H288, A:G289, A:D290, A:V291, A:E292, A:L293, A:V294, A:I295, A:P296, A:E297, A:G298, A:T299, A:Q300, A:T301, A:G302, A:K303, A:V329, A:V330, A:T331, A:P332, A:T333, A:G334, A:L335, A:N336, A:D337, A:R338, A:Q339, A:K340	42	0.645

Helper T-cell (HTL) Epitope Prediction

The result of predicted helper T-cell epitopes of Dnaj using IEDB is shown in Table 5. According to rough guideline of server for MHCII binding prediction, using IEDB recommended 2.2 method, the lower number of percentile rank indicates higher affinity, so the predicted epitopes with a percentile rank < 5 were selected. From 76 predicted helper T-cell epitopes, 55 epitopes had 100% conservancy among of clinical prevalent serotypes of *S. pneumoniae* (serotypes 1, 3, 4, 6A, 6B, 7F, 9V, 10A, 11A, 12F, 14, 15B, 18C, 19F, 19A, 22A/F, 23F, and 33F) DnaJs (Table 5). Also, some predicted B-cell epitopes were also predicted as HTL epitopes.

Table 5. Predicted Helper T-cell epitopes for DnaJ using IEDB server.

Predicted MHCII Binding epitopes (Percentile Rank<5)									
Peptide Sequence	Allele Rank Score	Perc.	Vaxi.	Conser.	Peptide Sequence	Allele Rank Score	Perc.	Vaxi.	Conser.
IFYNLNLNFVQAALG	HLA-DRB1*01:01	2.10	0.99	100	YRKLSKKYHPDINKE	HLA-DRB1*11:01	0.65	0.77	91.3
AGVINVDTQTPLGMM	HLA-DRB1*03:01	2.50	-0.02	100	AYRKLSKKYHPDINK	HLA-DRB1*08:01	0.85	0.62	100
FYNLNLNFVQAALGD	HLA-DRB1*01:01	2.60	0.96	100	DEIKKAYRKLSKKYH	HLA-DRB1*08:01	0.85	0.41	100
GAGVINVDTQTPLGM	HLA-DRB1*03:01	2.70	0.14	100	EIKKAYRKLSKKYHP	HLA-DRB1*08:01	0.85	0.53	100
TIFYNLNLNFVQAAL	HLA-DRB1*01:01	2.80	0.91	100	IKKAYRKLSKKYHPD	HLA-DRB1*08:01	0.85	0.42	100
GTTIFYNLNLNFVQA	HLA-DRB1*04:01	2.90	0.82	100	KAYRKLSKKYHPDIN	HLA-DRB1*08:01	0.85	0.87	100
IFYNLNLNFVQAALG	HLA-DRB1*04:01	2.90	0.99	100	KKAYRKLSKKYHPDI	HLA-DRB1*08:01	0.85	0.91	100
TIFYNLNLNFVQAAL	HLA-DRB1*04:01	2.90	0.91	100	YRKLSKKYHPDINKE	HLA-DRB1*08:01	0.85	0.77	91.3
TTIFYNLNLNFVQAA	HLA-DRB1*04:01	2.90	0.8	100	GMMRRQVTCDVCHGR	HLA-DRB1*13:01	1.5	0.06	82.61
GVINVDTQTPLGMMR	HLA-DRB1*03:01	3	-0.32	100	LGMMRRQVTCDVCHG	HLA-DRB1*13:01	1.5	-0.3	100
NLNLNFVQAALGDTV	HLA-DRB1*01:01	3.60	1	100	MMRRQVTCDVCHGRG	HLA-DRB1*13:01	1.5	0.36	82.61
YNLNLNFVQAALGDT	HLA-DRB1*01:01	4.10	1.22	100	MRRQVTCDVCHGRGK	HLA-DRB1*13:01	1.5	0.58	82.61
LNLNFVQAALGDTVD	HLA-DRB1*01:01	4.30	1.04	100	PLGMMRRQVTCDVCH	HLA-DRB1*13:01	1.5	-0.4	100
EGTTIFYNLNLNFVQ	HLA-DRB1*04:01	4.40	0.76	100	QTPLGMMRRQVTCDV	HLA-DRB1*13:01	1.5	-0.2	100
NLNFVQAALGDTVDI	HLA-DRB1*01:01	4.60	0.78	100	TPLGMMRRQVTCDVC	HLA-DRB1*13:01	1.5	-0.3	100
ASDKFEREGTTIFYN	HLA-DRB1*04:01	4.70	0.26	100	DDLQYRVNLTFEEAI	HLA-DRB1*08:01	2.5	0.92	100
DKFEREGTTIFYNLN	HLA-DRB1*04:01	4.70	0.3	100	DLQYRVNLTFEEAIF	HLA-DRB1*08:01	2.5	0.91	100
GDDLQYRVNLTFEEA	HLA-DRB1*08:01	2.50	1.1	100	TGKKFRLRSKGAPSL	H2-IEd	2.75	1.10	86.96
LQYRVNLTFEEAIFG	HLA-DRB1*08:01	2.50	0.82	100	AYDQYGAAGANGGFG	H2-IAb	3.1	1.12	86.96
QGDDLQYRVNLTFEE	HLA-DRB1*08:01	2.50	1.25	100	ADEIKKAYRKLSKKY	H2-IEd	3.69	0.35	100
QYRVNLTFEEAIFGT	HLA-DRB1*08:01	2.50	0.78	100	KAYRKLSKKYHPDIN	H2-IEd	3.8	0.87	100
YRVNLTFEEAIFGTE	HLA-DRB1*08:01	2.50	0.76	100	AYRKLSKKYHPDINK	H2-IEd	4.25	0.62	100
ASADEIKKAYRKLSK	HLA-DRB1*11:01	3.20	0.33	100	YDQYGAAGANGGFGG	H2-IAb	4.55	1.18	86.96
EGTTIFYNLNLNFVQ	HLA-DRB1*15:01	4.50	0.76	100	FRLRSKGAPSLRGGA	HLA-DRB1*08:01	0.25	1.29	86.96
REGTTIFYNLNLNFV	HLA-DRB1*15:01	4.60	0.91	100	GKKFRLRSKGAPSLR	HLA-DRB1*08:01	0.25	1.32	86.96
QTPLGMMRRQVTCDV	HLA-DRB1*11:01	4.70	-0.24	100	KFRLRSKGAPSLRGG	HLA-DRB1*03:01	0.25	1.32	86.96
SDKFEREGTTIFYNL	HLA-DRB1*04:01	4.70	0.53	100	KKFRLRSKGAPSLRG	HLA-DRB1*08:01	0.25	1.01	86.96
HGAGVINVDTQTPLG	HLA-DRB1*03:01	3.70	0.11	100	QTGKKFRLRSKGAPS	HLA-DRB1*08:01	0.25	1.19	86.96
EIKKAYRKLSKKYHP	HLA-DRB1*11:01	0.02	0.53	100	TGKKFRLRSKGAPSL	HLA-DRB1*08:01	0.25	1.10	86.96
IKKAYRKLSKKYHPD	HLA-DRB1*11:01	0.04	0.42	100	TQTGKKFRLRSKGAP	HLA-DRB1*08:01	0.25	1.18	86.96

DEIKKAYRKLSKKYH	HLA-DRB1*11:01	0.05	0.41	100	ADEIKKAYRKLSKKY	HLA-DRB1*11:01	0.34	0.35	100
KKAYRKLSKKYHPDI	HLA-DRB1*11:01	0.15	0.91	100	SADEIKKAYRKLSKK	HLA-DRB1*11:01	0.60	0.51	100
KAYRKLSKKYHPDIN	HLA-DRB1*11:01	0.23	0.87	100	AYRKLSKKYHPDINK	HLA-DRB1*11:01	0.65	0.62	100
FRLRSKGAPSLRGGA	HLA-DRB1*08:01	0.25	1.29	86.96	EIKKAYRKLSKKYHP	H2-IEd	0.98	0.53	100
DEIKKAYRKLSKKYH	H2-IEd	1.23	0.41	100	TQTGKKFRLRSKGAP	H2-IEd	2.38	1.18	86.96
IKKAYRKLSKKYHPD	H2-IEd	1.25	0.42	100	RAAYDQYGAAGANGG	H2-IAb	2.55	1.35	86.96
KKAYRKLSKKYHPDI	H2-IEd	1.65	0.91	100	AAYDQYGAAGANGGF	H2-IAb	2.65	0.83	86.96
QTGKKFRLRSKGAPS	H2-IEd	2.08	1.19	86.96	GKKFRLRSKGAPSLR	H2-IEd	2.75	1.32	86.96

(Perc.: Percentile Rank, Vaxi.: VaxiJen Score, Conser.: Conservancy).

Prediction of IL-4, IL-10, IL-17, and INF-γ Inducer Epitopes

Approximately 100 antigenic regions (15mer) of DnaJ with the potential to induce IL-4 were predicted using the IL4pred server with a SVM score ranging from 0.3 to 0.7 and are shown in Table 6. Using the IFNepitope server, DnaJ was scanned and predicted to have many IFN-γ inducing MHC class II binder peptides throughout its sequence (Table 7). The result of the IL-10 and IL-17inducer sequences of DnaJ is shown in Table 8. DnaJ was predicted to be an IL-10 and IL-17 inducer protein with scores of 0.6.

Table 6. The best IL-4 inducing peptide of DnaJ (SVM Score > 0.3)

DnaJ IL-4 Inducer Sequence	Score	DnaJ IL-4 Inducer Sequence	Score	DnaJ IL-4 Inducer Sequence	Score	DnaJ IL-4 Inducer Sequence	Score
KEFAAAGDLKVNPKK	0.69	TGLNDRQKVALKEFA	0.47	YVVVSVEASDKFERE	0.41	LQYRVNLTFEEAIFG	0.37
EFAAAGDLKVNPKKK	0.68	NNTEFYDRLGVSKNA	0.46	VVVSVEASDKFEREG	0.41	ASDKFEREGTTIFYN	0.37
EFYDRLGVSKNASAD	0.59	LTFEEAIFGTEKEVK	0.46	VIPEGTQTGKKFRLR	0.41	KFEREGTTIFYNLNL	0.37
TEFYDRLGVSKNASA	0.58	GAVGDQYVTVNVVTP	0.46	NDRQKVALKEFAAAG	0.41	VQEAYETLSDDQKRA	0.36
DRLGVSKNASADEIK	0.58	GLNDRQKVALKEFAA	0.46	RQKVALKEFAAAGDL	0.41	APRQGDDLQYRVNLT	0.36
RLGVSKNASADEIKK	0.58	LNDRQKVALKEFAAA	0.46	SADEIKKAYRKLSKK	0.40	DLQYRVNLTFEEAIF	0.36
LGVSKNASADEIKKA	0.58	NLTFEEAIFGTEKEV	0.45	DKYKEVQEAYETLSD	0.40	QYRVNLTFEEAIFGT	0.36
GVSKNASADEIKKAY	0.58	TQTGKKFRLRSKGAP	0.45	VVSVEASDKFEREGT	0.40	TQTPLGMMRRQVTCD	0.36
NTEFYDRLGVSKNAS	0.57	GDQYVTVNVVTPTGL	0.45	VSVEASDKFEREGTT	0.40	SDKFEREGTTIFYNL	0.36
GVETGQQIRLAGQGE	0.57	DLKVNPKKKGFFDHI	0.45	VTVNVVTPTGLNDRQ	0.40	DKFEREGTTIFYNLN	0.36
VETGQQIRLAGQGEA	0.57	YRVNLTFEEAIFGTE	0.44	LKEFAAAGDLKVNPK	0.40	ALKEFAAAGDLKVNP	0.36
VNLTFEEAIFGTEKE	0.56	EGTQTGKKFRLRSKG	0.44	QTPLGMMRRQVTCDV	0.39	VNPKKKGFFDHIKDA	0.36
QYVTVNVVTPTGLND	0.56	DRQKVALKEFAAAGD	0.44	YVTVNVVTPTGLNDR	0.39	QKRAAYDQYGAAGAN	0.35
PAGVETGQQIRLAGQ	0.54	GDLKVNPKKKGFFDH	0.44	VNVVTPTGLNDRQKV	0.39	RGGAVGDQYVTVNVV	0.35
IPAGVETGQQIRLAG	0.50	VSKNASADEIKKAYR	0.43	KNASADEIKKAYRKL	0.38	VALKEFAAAGDLKVN	0.35
ETGQQIRLAGQGEAG	0.49	SKNASADEIKKAYRK	0.43	EAYETLSDDQKRAAY	0.38	LKVNPKKKGFFDHIK	0.35
RVNLTFEEAIFGTEK	0.48	GDLYVVVSVEASDKF	0.43	NPNAPRQGDDLQYRV	0.38	NPKKKGFFDHIKDAF	0.35
TGKKFRLRSKGAPSL	0.48	DLYVVVSVEASDKFE	0.43	PNAPRQGDDLQYRVN	0.38	YKEVQEAYETLSDDQ	0.34
VGDQYVTVNVVTPTG	0.48	LYVVVSVEASDKFER	0.43	SVEASDKFEREGTTI	0.38	PRQGDDLQYRVNLTF	0.34
FAAAGDLKVNPKKKG	0.48	GTQTGKKFRLRSKGA	0.43	VEASDKFEREGTTIF	0.38	EGTTIFYNLNLNFVQ	0.34
AGVETGQQIRLAGQG	0.47	DQYVTVNVVTPTGLN	0.43	EASDKFEREGTTIFY	0.38	GTTIFYNLNLNFVQA	0.34
IPEGTQTGKKFRLRS	0.47	TFEEAIFGTEKEVKY	0.42	KVNPKKKGFFDHIKD	0.38	GGAVGDQYVTVNVVT	0.34
PEGTQTGKKFRLRSK	0.47	PTGLNDRQKVALKEF	0.42	NASADEIKKAYRKLS	0.37	PKKKGFFDHIKDAFD	0.34
QTGKKFRLRSKGAPS	0.47	MNNTEFYDRLGVSKN	0.41	QEAYETLSDDQKRAA	0.37	KEVQEAYETLSDDQK	0.33
AVGDQYVTVNVVTPT	0.47	ASADEIKKAYRKLSK	0.41	NAPRQGDDLQYRVNL	0.37	EVQEAYETLSDDQKR	0.33

Table 7. The best IFN-γ inducing peptide of DnaJ.

DnaJ IFN-γ Inducer Sequence	Score	DnaJ IFN-γ Inducer Sequence	Score	DnaJ IFN-γ Inducer Sequence	Score	DnaJ IFN-γ Inducer Sequence	Score
SADEIKKAYRKLSKK	0.105	GGFNGAGGFGGFEDI	0.399	GRCHGAGVINVDTQT	0.293	FRLRSKGAPSLRGGA	0.370
DEIKKAYRKLSKKYH	0.184	GAGGFGGFEDIFSSF	0.046	RCHGAGVINVDTQTP	0.308	RLRSKGAPSLRGGAV	0.163
EIKKAYRKLSKKYHP	0.544	FGGFEDIFSSFFGGG	0.246	CHGAGVINVDTQTPL	0.418	LRSKGAPSLRGGAVG	0.343
IKKAYRKLSKKYHPD	0.421	GGFEDIFSSFFGGGG	0.739	HGAGVINVDTQTPLG	0.444	SKGAPSLRGGAVGDQ	0.017
KKAYRKLSKKYHPDI	0.364	GFEDIFSSFFGGGGS	0.361	GAGVINVDTQTPLGM	0.280	KGAPSLRGGAVGDQY	0.144
KAYRKLSKKYHPDIN	0.049	FEDIFSSFFGGGGSS	0.603	AGVINVDTQTPLGMM	0.370	GAPSLRGGAVGDQYV	0.247
TLSDDQKRAAYDQYG	0.013	EDIFSSFFGGGGSSR	0.543	GVINVDTQTPLGMMR	0.291	PKKKGFFDHIKDAFD	0.120
LSDDQKRAAYDQYGA	0.081	DIFSSFFGGGGSSRN	0.654	VINVDTQTPLGMMRR	0.109	KKKGFFDHIKDAFDG	0.078
SDDQKRAAYDQYGAA	0.343	IFSSFFGGGGSSRNP	0.400	QVTCDVCHGRGKEIK	0.086	KKGFFDHIKDAFDGE	0.057
DDQKRAAYDQYGAAG	0.244	FSSFFGGGGSSRNPN	0.449	VTCDVCHGRGKEIKY	0.069	KGFFDHIKDAFDGE	0.060
DQKRAAYDQYGAAGA	0.577	SSFFGGGGSSRNPNA	0.406	TCDVCHGRGKEIKYP	0.154	GFFDHIKDAFDGE	0.133
QKRAAYDQYGAAGAN	0.526	SFFGGGGSSRNPNAP	0.206	CDVCHGRGKEIKYPC	0.029	FFDHIKDAFDGE	0.290
KRAAYDQYGAAGANG	0.395	FFGGGGSSRNPNAPR	0.165	DVCHGRGKEIKYPCT	0.027	FDHIKDAFDGE	0.363
RAAYDQYGAAGANGG	0.400	FGGGGSSRNPNAPRQ	0.085	PAGVETGQQIRLAGQ	0.100	DHIKDAFDGE	0.100
YDQYGAAGANGGFGG	0.064	GGGGSSRNPNAPRQG	0.122	ETGQQIRLAGQGEAG	0.003	HIKDAFDGE	0.290
DQYGAAGANGGFGGF	0.081	EEAIFGTEKEVKYHR	0.051	TGQQIRLAGQGEAGF	0.243		
GAAGANGGFGGFNGA	0.266	EAIFGTEKEVKYHRE	0.200	GQQIRLAGQGEAGFN	0.176		
AAGANGGFGGFNGAG	0.004	AIFGTEKEVKYHREA	0.200	FVQAALGDTVDIPTV	0.166		
AGANGGFGGFNGAGG	0.530	IFGTEKEVKYHREAG	0.196	VQAALGDTVDIPTVH	0.011		
GANGGFGGFNGAGGF	0.586	FGTEKEVKYHREAGC	0.322	QAALGDTVDIPTVHG	0.206		
ANGGFGGFNGAGGFG	0.440	GTEKEVKYHREAGCR	0.186	AALGDTVDIPTVHGD	0.232		
NGGFGGFNGAGGFGG	0.831	TEKEVKYHREAGCRT	0.395	ALGDTVDIPTVHGDV	0.113		

GGFGGFNGAGGFGGF	0.990	EKEVKYHREAGCRTC	0.260	LGDTVDIPTVHGDVE	0.081		
GFGGFNGAGGFGGFE	0.744	TCGRCHGAGVINVDT	0.061	GDTVDIPTVHGDVEL	0.006		
FGGFNGAGGFGGFED	0.461	CGRCHGAGVINVDTQ	0.028	TVDIPTVHGDVELVI	0.195		

Table 8. The best IL-10 and IL-17 inducing peptides of DnaJ.

Protein	Peptide Sequence of IL-10 Inducer	VaxiJen Score
DnaJ	MNNTEFYDRLGVSKNASADEIKKAYRKLSKKYHPDINKEPGAEDKYKEVQEAYETLSDDQKRAAYDQYGA	0.69
Protein	**Peptide Sequence of IL-17 Inducer**	
DnaJ	QKRAA (epitope ID 138227) (as part of HLA-DRB1*04:01 chain)	0.62

In-Silico Immune Response Simulation

We used the C-ImmSim server to show the DnaJ administrations and the humoral and cellular response of the mammalian immune system against DnaJ. The result showed an increase in the primary response of the IgM titer. Secondary and tertiary responses observed after booster doses of DnaJ, an increase in proliferation of T- and B-cell populations with isotype switching (IgM to IgG and the formation of the IgG1 and IgG2 subclasses), and formation of memory cells were observed. In addition, an increase in natural killer cells, dendritic cell responses and macrophage activity were also found. In parallel, the immune simulation also showed that IFN-γ, IL-2, IL-10 and IL-4 production was stimulated after immunization, resulting in humoral and cellular immune protection responses (Fig. 4). The results of the IL4pred, IL-10pred, and IFNepitope predictions were consistent with the results of C ImmSim simulation.

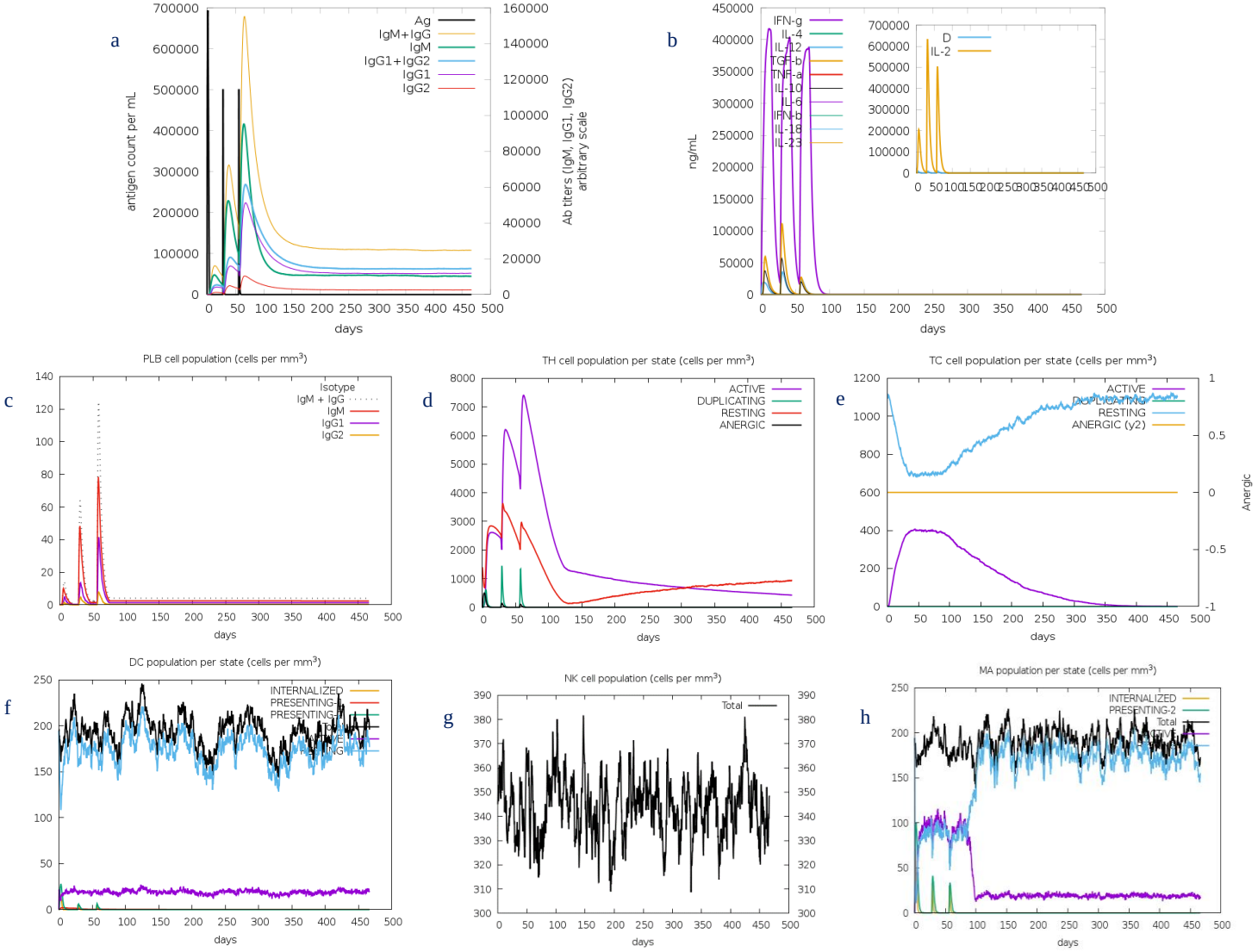

Fig. 4. *In-silico* immune response simulation using DnaJ administration. (a) Variety of immunoglobulin production in response to DnaJ injections (DnaJ antigen, and IgM, IgG1, and IgG2 subclasses are shown as black and colored peaks, respectively). (b) Cytokines and interleukin levels. The insert plot indicates the IL-2 level with the Simpson index, D shown by the dotted line, as a measure of diversity and danger signal together with leukocyte growth factor IL-2. (c) The B-cell cell population evolution after the three administrations. (d) and (e) The evolution of T-helper and T-cytotoxic cell populations after the DnaJ injections. (f) Dendritic cell population evolution. (g) Macrophage population evolution, (h) Natural killer cell population evolution.

Molecular Docking

The ClusPro online server performed molecular protein-protein docking between DnaJ and HLADRB1*01: 01 (the most common binding allele in the Iran population). Cluster No. 0.00 of DnaJ-HLADRB1 docked complex with 63 members having the lowest energy of -809.9 Kcal/mol was selected for further analysis. The interaction surface residues of the docked complex were checked with Dimpolt tools in LigPlot+ software and visualized using PyMol software (Fig. 5). A total of 7 DnaJ residues coupled with 4 and 4 residues of A and B chains from HLADRB1*01:01 molecule, [(Arg^{104}-Glu^{55}, Thr^{174}-Asn^{62}, Arg^{181}-Gln^{9}, Thr^{184}-Ser^{53}) and (Arg^{161}-Asp^{76}, Gln^{173}-Gln^{70}, Gln^{173}-Arg^{71}, Asp^{186}-Ser^{88})], respectively. Altogether, a number of 11 hydrogen bonds were formed between the DnaJ residues and HLADRB1*01:01 molecule (Fig. 5a, 5b, 5c).

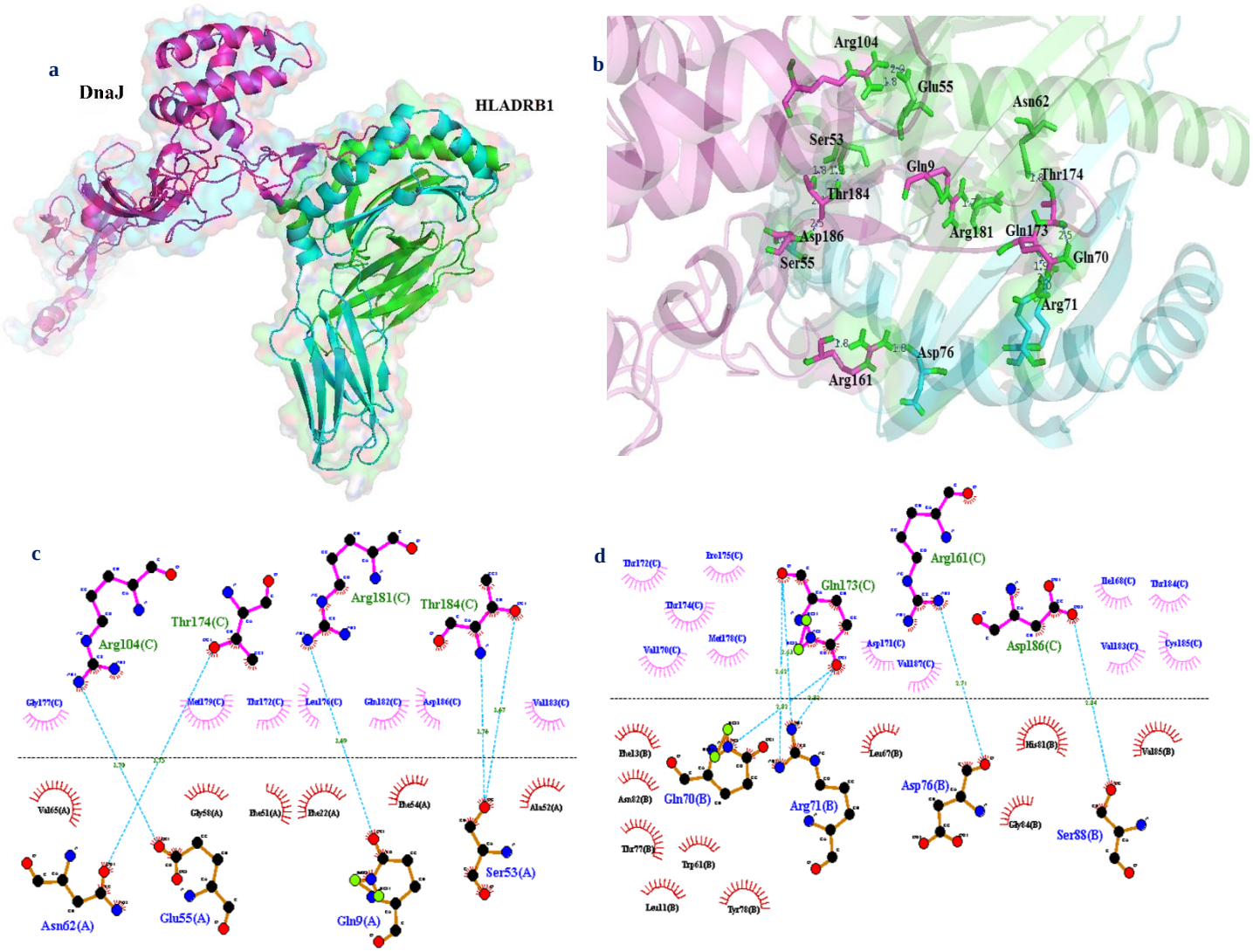

Fig. 5. Molecular docking of the DnaJ with HLADRB1*01:01 (Chain A and B). a) The 3D structure of DnaJ-HLADRB1 docked complex. The cartoon depictions of the DnaJ - HLADRB1*01:01 complex illustrated using the PyMol software. The DnaJ, chain A and B of HLADRB1*01:01 have been shown in magenta, green, and cyan, respectively. The lowest energy score of this complex model was -809.9 Kcal/mol, indicating good binding affinity. b) and c) Dimplot interaction plots between DnaJ residues and HLADRB1*01:01 A (b) and B (c) chains in docked complex. DnaJ residues, HLADRB1*01:01 residues, hydrogen bonds, and non-bonded residues are exhibited in green, blue, blue dashed lines, and red/pink eyelashes, respectively.

Cluster No. 0.00 of DnaJ-TLR-2 docked complex with 42 members having the lowest energy of -865.2 Kcal/mol was selected for further analysis. The interaction surface residues of the docked complex were checked with Dimpolt tools in LigPlot+ software and visualized using PyMol software (Fig. 6). A total of 8 and 6 residues of DnaJ coupled with 9 and 8 residues of B and D chains from the TLR-2 molecule, respectively. In total, a number of 20 hydrogen bonds, 2 salt bonds and many hydrophobic interactions were formed between the DnaJ residues and TLR-2 molecule (Fig. 6a, 6b, 6c). The details of the interactions are shown in Tables 9a and 9b.

Cluster No. 0.00 of the ClusPro prediction result of DnaJ-TLR-4 docked complex with 37 members having the lowest energy of -864.7 Kcal/mol was selected for further analysis. The interaction of the docked complex were checked with Dimpolt tools in LigPlot+ software and visualized by PyMol software (Fig. 7). A total of 12 and 3 residues of DnaJ coupled with 13 and 5 residues of A and B chains from the TLR-4 molecule,

respectively. In total, a number of 21 hydrogen bonds and 5 salt bonds, and many hydrophobic interactions were formed between the DnaJ residues and TLR-4 molecule (Fig. 7a, 7b, 7c). The details of the interactions are shown in Tables 10a and 10b.

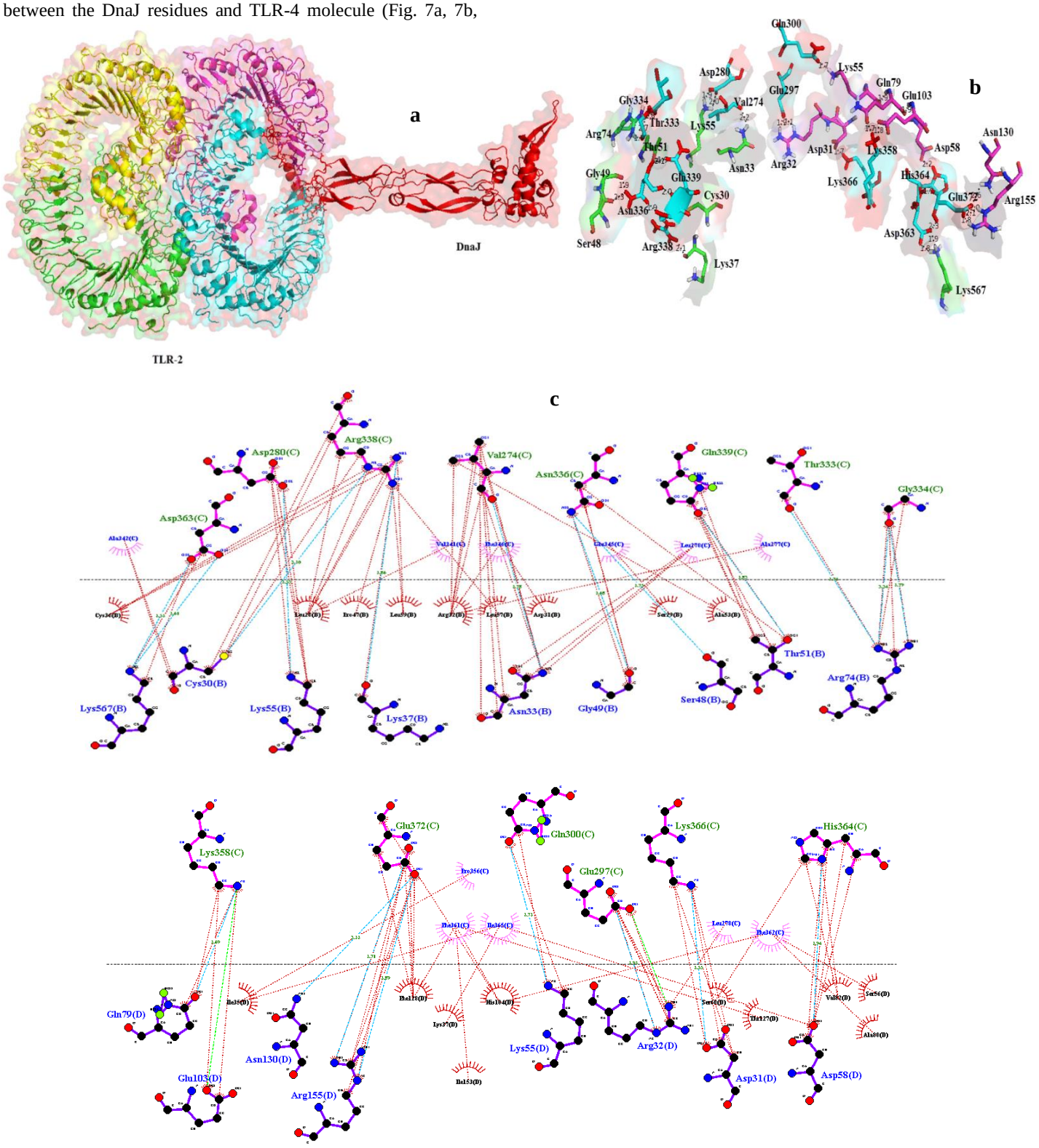

Fig. 6. Molecular docking of the DnaJ with TLR-2 (Chain B and D). a) The 3D structure of DnaJ-TLR-2 docked complex. The cartoon depictions of the DnaJ - TLR-2 complex illustrated using the PyMol software. The DnaJ, chain B and D of TLR-2 have been shown in red, magenta, and cyan, respectively. The lowest energy score of this complex model was -865.2 Kcal/mol, indicating good binding affinity. b) and c) Dimplot interaction plots between DnaJ residues and TLR-2 B (b) and D (c) chains in docked complex. DnaJ residues, TLR-2 residues, hydrogen bonds, residues with hydrophobic bonds are exhibited in green, blue, blue dashed lines, and red/pink eyelashes, respectively.

Table 9. Analysis of Dimplot 2D-interaction plot between DnaJ residues and TLR-2 molecule in docked complex.

a

DnaJ – TLR-2 docked complex							
DnaJ				**TLR-2 (B chain)**			
Atom Name	**Res. No.**	**Res. Name**	**Type of Bond**	**Atom Name**	**Res. No.**	**Res. Name**	**H-Bond Distance (A°)**
O	334	Gly	**Hydrogen Bonds**	NH1	74	Arg	2.79
O	334	Gly		NH2	74	Arg	3.34
O	333	Thr		NH2	74	Arg	2.74
OE1	339	Gln		OG1	51	Thr	2.83
ND2	336	Asn		O	49	Gly	2.68
ND2	336	Asn		O	48	Ser	2.75
O	274	Val		ND2	33	Asn	2.78
NH2	338	Arg		O	37	Lys	2.86
NE	338	Arg		SG	30	Cys	3.20
OD2	280	Asp		NZ	55	Lys	2.53
OD1	363	Asp		NZ	567	Lys	2.61
OD2	363	Asp		NZ	567	Lys	2.52
DnaJ				**TLR-2 (D chain)**			
Atom Name	**Res. No.**	**Res. Name**	**Type of Bond**	**Atom Name**	**Res. No.**	**Res. Name**	**H-Bond Distance (A°)**
ND1	364	His	**Hydrogen Bonds**	OD2	58	Asp	2.94
NZ	366	Lys		OD1	31	Asp	2.55
OE2	297	Glu		NE	32	Arg	2.92
OE1	300	Gln		NZ	55	Lys	2.72
OE2	372	Glu		NH2	155	Arg	2.71
OE1	372	Glu		NE	155	Arg	2.89
OE1	372	Glu		ND2	130	Asn	3.22
NZ	358	Lys		OE1	79	Gln	2.69
OE1	297	Glu	**Salt Bridges**	NH2	32	Arg	-
NZ	358	Lys		OE2	103	Glu103	

b

Interacting residues of DnaJ	**Interacting residues of TLR-2 (B chain)**	**Type of Bond**
Gly334, Thr333, Gln339, Asn336, Ala277, Leu278, Val274, Glu345, Phe346, Val341, Arg338, Asp280, Asp363, Ala342	Arg74, Thr51, Gly49, Ala53, Ser29, Asn33, Asp31, Leu57, Arg32, Leu59, Pro47, Leu28, Lys37, Cys36, Lys55, Cys30, Lys567	Hydrophobic Bond
Interacting residues of DnaJ	**Interacting residues of TLR-2 (D chain)**	**Type of Bond**
His364, Phe362, Leu278, Lys366, Glu297, Gln300, Ile365, Phe361, Pro356, Glu372, Lys 358	Ser56, Ala80, Val82, Asp58,Thr127, Asp31, Ser60, Arg32, Lys55, His104, Ile153, Lys37, Phe128, Arg155, Asn130, Ile35, Glu103, Gln79	Hydrophobic Bond

Cluster No. 0.00 of the ClusPro prediction result of DnaJ-TLR-4 docked complex with 37 members having the lowest energy of -864.7 Kcal/mol was selected for further analysis. The interaction of the docked complex were checked with Dimpolt tools in LigPlot+ software and visualized by PyMol software (Fig. 7). A total of 12 and 3 residues of DnaJ coupled with 13 and 5 residues of A and B chains from the TLR-4 molecule, respectively. In total, a number of 21 hydrogen bonds and 5 salt bonds, and many hydrophobic interactions were formed between the DnaJ residues and TLR-4 molecule (Fig. 7a, 7b, 7c). The details of the interactions are shown in Tables 10a and 10b.

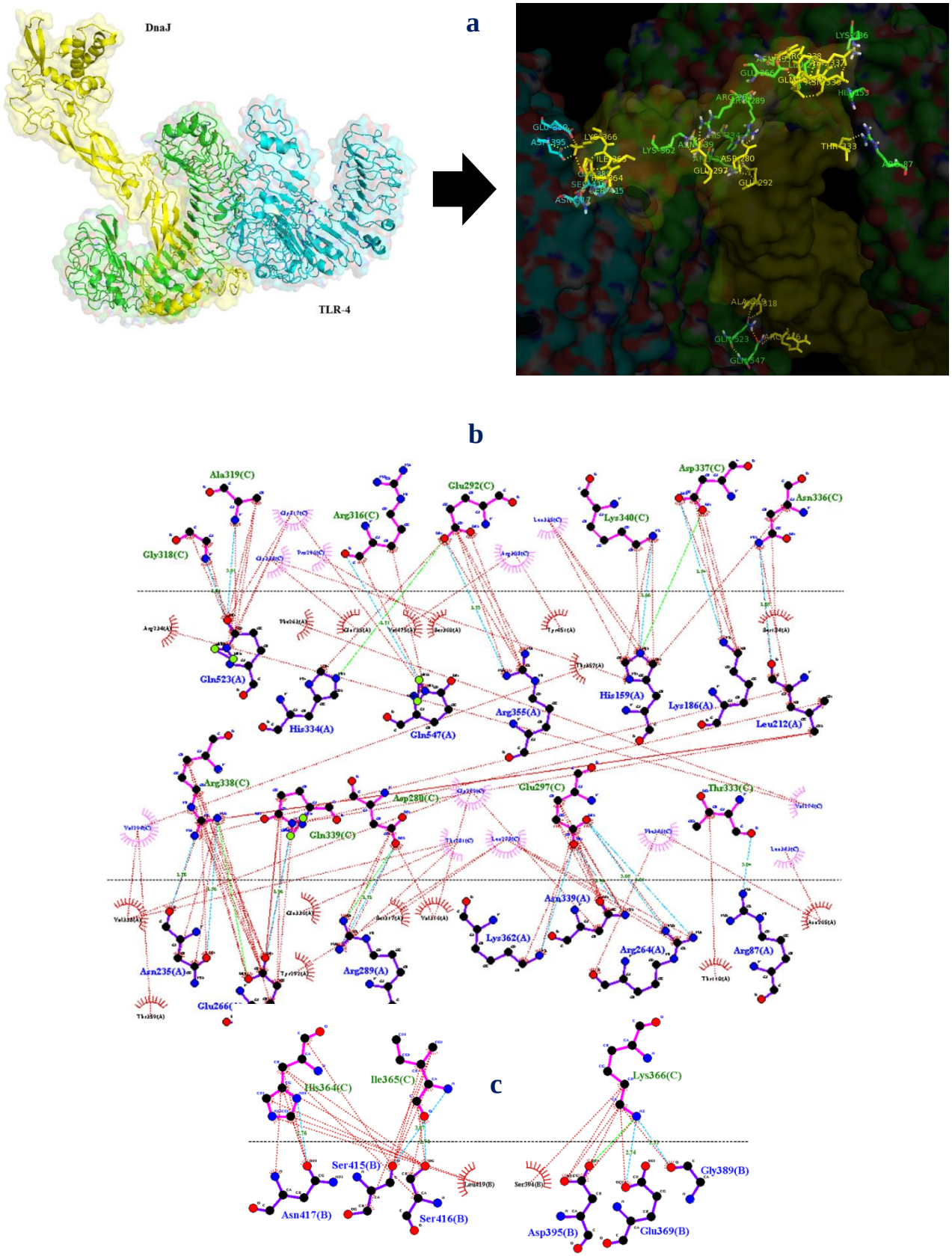

Fig. 7. Molecular docking of the DnaJ with TLR-4 (Chain A and B). **a)** The 3D structure of DnaJ-TLR-4 docked complex. The cartoon depictions of the DnaJ - TLR-4 complex illustrated using the PyMol software. The DnaJ, chain A and B of TLR-4 have been shown in yellow, green, and cyan, respectively. The lowest energy score of this complex model was -864.7 Kcal/mol, indicating good binding affinity. **b)** and **c)** Dimplot interaction plots between DnaJ residues and TLR-4 A **(**b) and B (c) chains in docked complex. DnaJ residues, TLR-4 residues, hydrogen bonds, residues with hydrophobic bonds are exhibited in green, blue, blue dashed lines, and red/pink eyelashes, respectively.

Table 10. Analysis of Dimplot 2D-interaction plot between DnaJ residues and TLR-4 molecule in docked complex.

a

DnaJ – TLR-4 docked complex							
DnaJ				**TLR-4 (A chain)**			
Atom Name	**Res. No.**	**Res. Name**	**Type of Bond**	**Atom Name**	**Res. No.**	**Res. Name**	**H-Bond Distance (A°)**
ND2	336	Asn	**Hydrogen Bonds**	O	212	Leu	2.85
OD2	337	Asp		NZ	186	Lys	2.54
NZ	340	Lys		NE2	159	His	2.66
OE1	292	Glu		NH2	355	Arg	2.73
O	316	Arg		HE22	547	Gln	2.71
N	319	Ala		OE1	523	Gln	3.01
N	318	Gly		OE1	523	Gln	2.81
O	333	Thr		NH1	87	Arg	3.04
OE2	297	Glu		NH1	264	Arg	3.01
OE2	297	Glu		NH2	264	Arg	3.05
OE1	297	Glu		ND2	339	Asn	2.96
OE1	297	Glu		NZ	362	Lys	2.46
OD2	280	Asp		NH2	289	Arg	2.72
NE2	339	Gln		OE1	266	Glu	2.96
NH2	338	Arg		OD1	235	Asn	2.56
NH1	338	Arg		O	235	Asn	2.78
OD1	337	Asp	**Salt Bridges**	NE2	159	His	-
OE1	292	Glu		NE2	334	His	
OD2	280	Asp		NH1	289	Arg	
NH2	338	Arg		OE2	266	Glu	
DnaJ				**TLR-4 (B chain)**			
Atom Name	Res. No.	Res. Name	Type of Bond	Atom Name	Res. No.	Res. Name	H-Bond Distance (A°)
NZ	366	Lys	**Hydrogen Bonds**	O	389	Gly	3.33
NZ	366	Lys		OE2	369	Glu	2.74
N	365	ILe		O	415	Ser	3.07
O	365	ILe		OG	416	Ser	2.74
ND1	364	His		OD1	417	Asn	2.76
NZ	366	Lys	**Salt Bridges**	OD2	395	Asp	-

b

Interacting residues of DnaJ	**Interacting residues of TLR-4 (A chain)**	**Type of Bond**
Asn336, Asp337, Lys340, Glu292, Arg316, Leu335, Arg308, Pro296, Gly317, Gly334, Ala319, Gly318, Val274, Leu343, Thr333, Phe346, Glu297, Leu278, Gly279, Thr281, Asp280, Gln339, Arg338, Val294	Leu212, Lys186, Ser184, His159, Thr357, Tyr451, Arg355, Ser360, Val475, Glu135, Gln547, His334, Phe263, Gln523, Arg234, Asn265, Thr110, Arg264, Asn339, Lys362, Val316, Ser317, Arg289, Tyr292, Glu336, Glu266, Val338, Asn235, Thr359	**Hydrophobic Bond**
Interacting residues of DnaJ	**Interacting residues of TLR-4 (B chain)**	**Type of Bond**
Lys366, Ile365, His364	Gly389, Glu369, Asp395, Ser394, Leu419, Ser416, Ser415, Asn417	**Hydrophobic Bond**

Designing of the DnaJ-based Constructs

We designed five multi-epitope constructs (construct 1 to 5) based on predicted B- and T-cell epitopes of DnaJ using GPGPG flexible linker. The B-cell epitope-based vaccine candidate have 3 categories (Fig. 8). As shown in Fig. 8, we selected the epitopes with Vaxijen score > 0.4 and 100% conservancy (construct1) (Fig. 8a), epitopes with Vaxijen score >1 and >85% conservancy (construct2) (Fig. 8b), and epitopes with Vaxijen score >1 and 100% conservancy (construct3) (Fig. 8c). In helper T-cell epitope-based vaccine candidate, we selected the MHCII binding epitopes with Vaxijen score >1 and 100% conservancy (construct4) (Fig. 8d). We also designed multi-epitopes of B- and T- cell and IL-17 inducer epitopes with Vaxijen score >1 and 100% conservancy to induce and promote both mucosal and humoral immune responses against pneumococci (construct5; Fig. 8e).

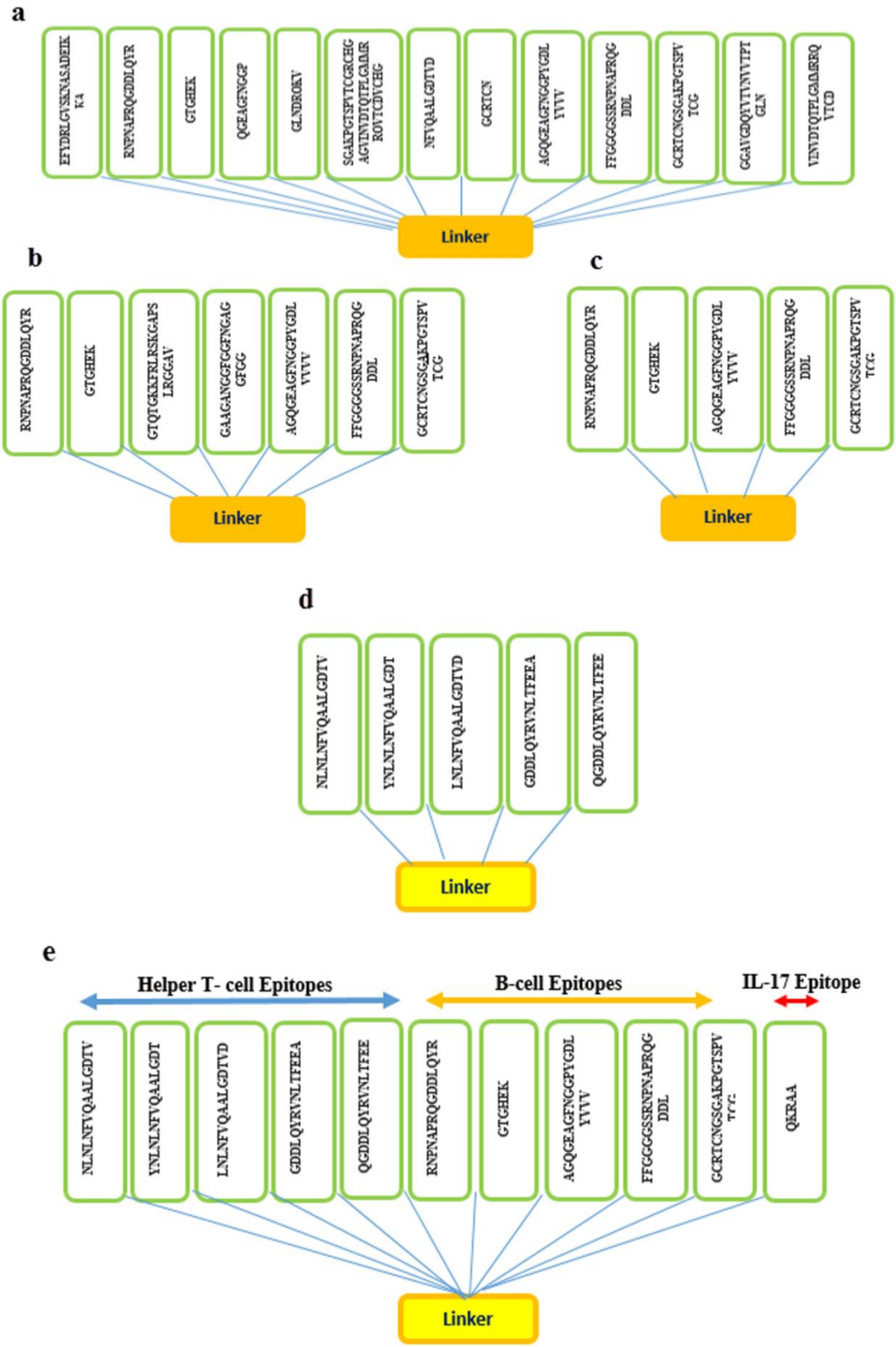

Fig. 8. Five multi-epitope Dnaj-based vaccine candidates. **a)** B-cell epitope-based vaccine candidate with Vaxijen score >0.4 and 100% conservancy, **b)** B-cell epitope-based vaccine candidate with Vaxijen score >1 and >85% conservancy, **c)** B-cell epitope-based vaccine candidate with with Vaxijen score >1 and 100% conservancy, **d)** helper T-cell epitope-based vaccine candidate, **e)** multi B- and T-cell epitopes DnaJ-based vaccine candidate.

Evaluation of Characteristics of Proposed DnaJ-based Constructs

We analyzed the five proposed DnaJ-based constructs for physicochemical properties, antigenicity, toxicity, solubility, transmembrain helix. The results are shown in Table 11. Using the ProtParam server, the molecular weight (MW) of constructs was predicted to be ranging from 11 to 29 kDa. According to the theoretical isoelectric point (pI) value, constructs (1 and 3), constructs (4 and 5), and construct 2 were expected to be neutral, acidic and basic in nature, respectively. Using the Solpro server, all constructs were predicted to be soluble when overexpressed in *E. coli*. According to the instability index (II), all constructs were classified as a stable protein (II of <40 indicates stability). The estimated negative GRAVY value of all constructs means that the constructs have hydrophilic nature and are able to interact with water molecules. All constructs were antigen and non-toxin, and had half-life of 30 hours in mammalian reticulocytes, *in vitro*, > 20 hours in yeast, *in vivo*, and >10 hours in *E. coli*, *in vivo*.

Table 11. Structural and Physicochemical characters of proposed DnaJ-based constructs.

Construct	Molecular Weight	VaxiJen Score	Toxicity	pI*	II**	GRAVY	Aliphatic Index	(Asp + Glu)	(Arg + Lys)	Trans-Membrane Helices	Solubility Index
1	28.88	1	Non Toxin	7.6	20.12	-0.69	41.54	20	21	0	0.94 (Soluble)
2	16.13	1.27	Non Toxin	9.85	26.75	-0.89	26.95	7	14	0	0.53 (Soluble)
3	11.46	1.21	Non Toxin	7.91	26.76	-1.116	24.82	7	8	0	0.7 (Soluble)
4	11.59	1.3	Non Toxin	5.44	15.58	-0.6	72.22	12	2	0	0.97 (Soluble)
5	22.55	1.32	Non Toxin	5.95	20.67	-0.77	48.35	19	12	0	0.51 (Soluble)

*pI: Theoretical Isoelectric Point
** II: Instability Index.

Secondary and Tertiary Structures

The predicted secondary structure of five proposed DnaJ-based constructs using the PSIPRED server were shown in Figure 9.

Using GOR V prediction server, we found that the construct 1 secondary structure contains 2.05% alpha helix, 24.23% extended strand, and 73.72% random coil. Construct 2 secondary structure contains 2.99% alpha helix, 8.3% extended strand, and 88.62% random coil. Construct 3 secondary structure contains 12.28% alpha helix and 87.72% random coil. Construct 4 secondary structure contains 34.26% alpha helix, 4.63% extended strand, and 61.11% random coil. Construct 5 secondary structure contains 12.5% alpha helix, 12.95% extended strand, and 74.55% random coil. 3D structures of all constructs were also modeled using the I-TASSER server and shown in Fig. 10. In all predicted models, Model 1 of I-TASSER, with the highest C-score, was selected for further refinement.

Fig. 9. Graphical representation of features of secondary structure of five DnaJ-based constructs using PSIPRED server. The alpha helix residues are pink, the beta strand residues are orange and the coil residues are gray.

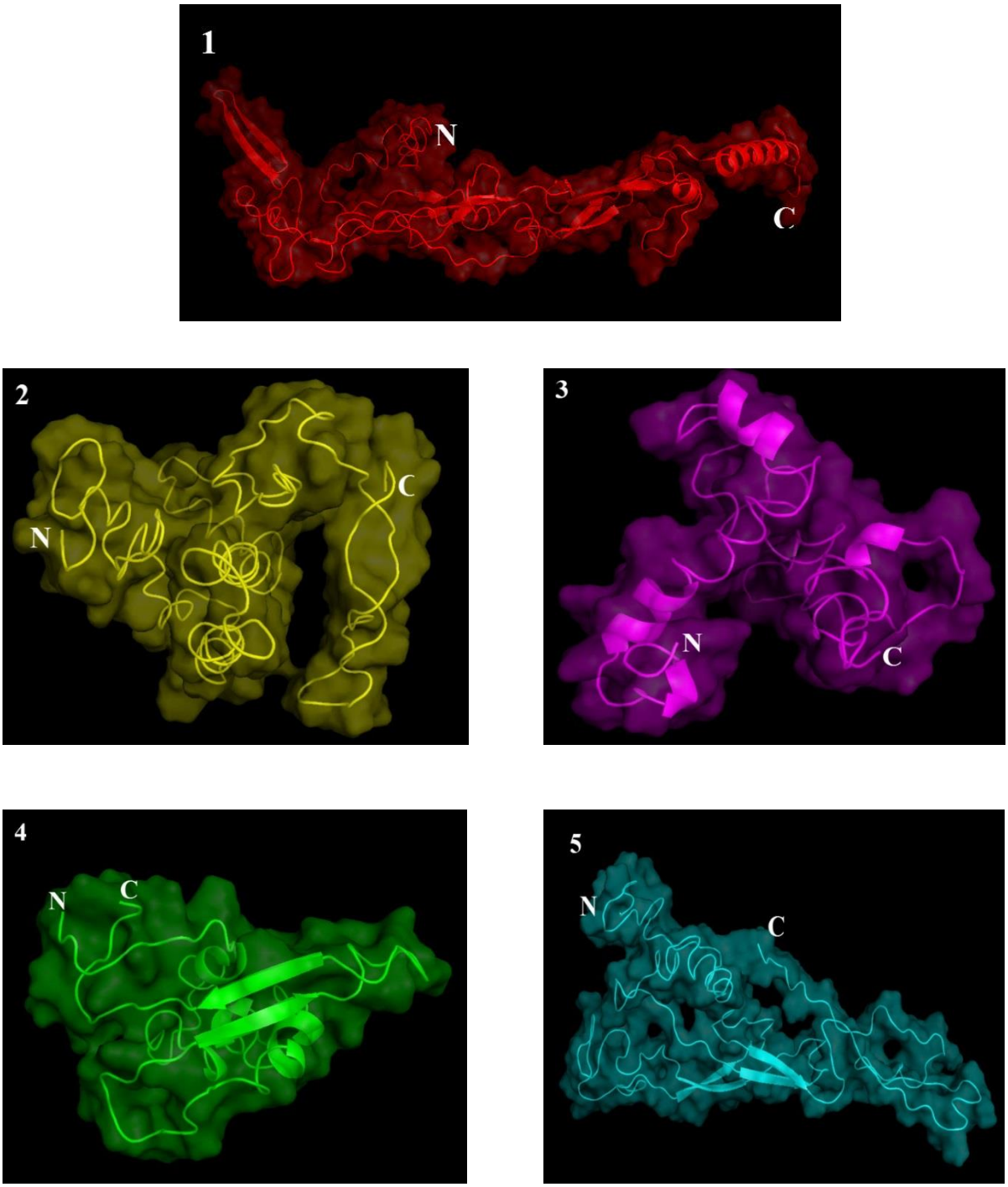

Fig. 10. I-TASSER 3D homology modeling of the DnaJ-based constructs 1 to 5. The structures were displayed by PyMol v.2.5 software. The estimated TM-score and RMSD for construct 1 to 5 were (0.46 ± 0.15 and 11 ± 4.6Å), (0.31 ± 0.10 and 13.7 ± 4.0Å), (0.35 ± 0.12 and 11.5 ± 4.5Å), (0.29 ± 0.09 and 13.1 ± 4.1Å), (0.34 ± 0.11 and 13.7 ± 4.0Å), respectively.

Refinement and Validation of Tertiary Structures

The best predicted I-TASSER model for five proposed DnaJ-based constructs was selected and refined using the Galaxy Refine server. GDT-HA score for all constructs were about 0.8 to 0.9 which is usually between [0, 1] values, and a higher score indicates conservative refinement. RMSD score and MolProbity score indicate aggressive refinement and a more physically realistic model, respectively. The details of Galaxy Refine server for all constructs before and after refinement were shown in Table 12. All refined constructs were analyzed for verifying the quality and potential errors in the 3D model structure using ProSA web and Ramachandran Plot (Figure 11). All constructs fell close in the score range commonly found in native proteins of comparable size. The details of all constructs validation before and after refinement using Ramachandran plot statistics from PROCHECK and Z-score from ProSA web servers were shown in Table 13.

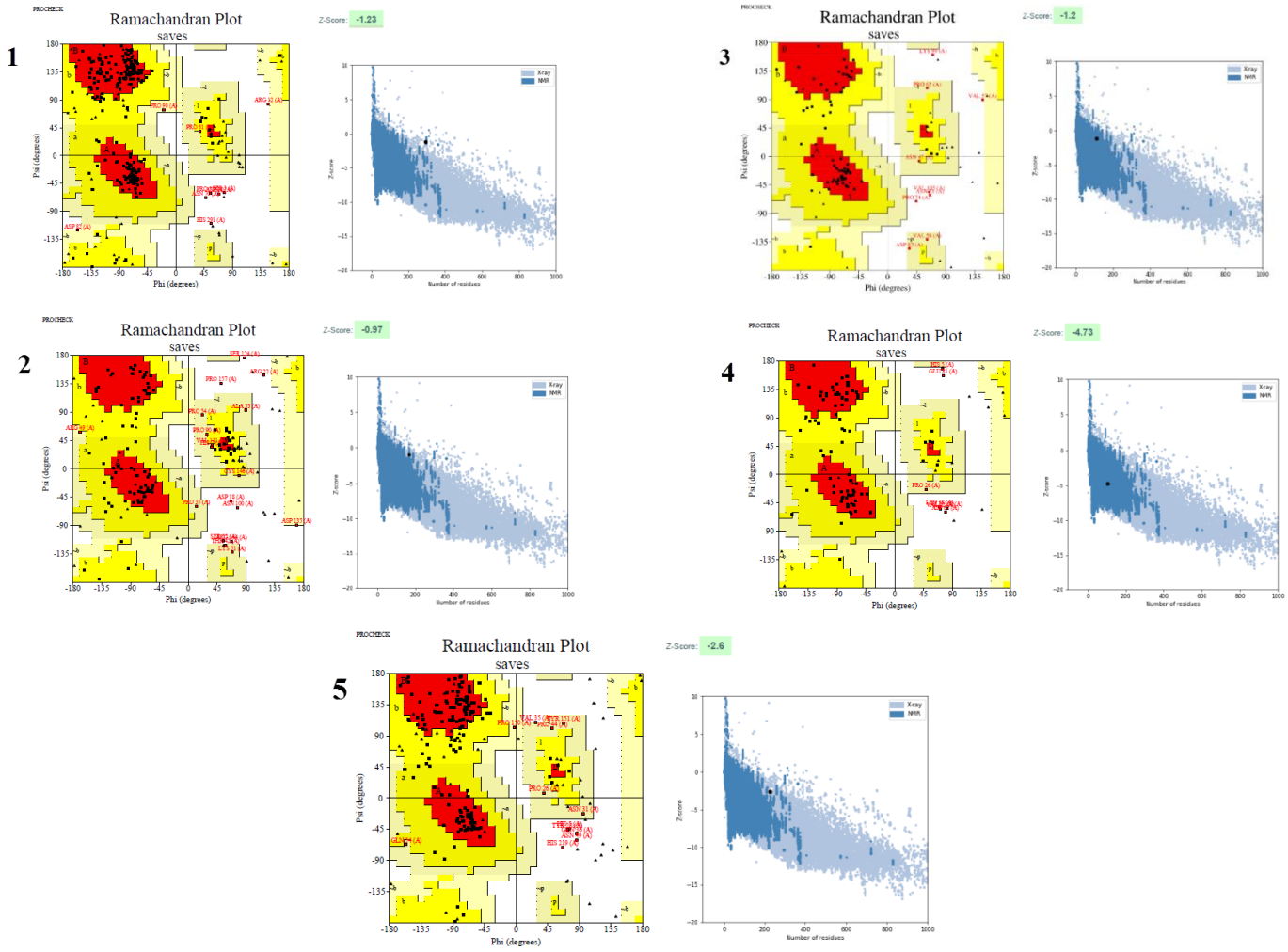

Fig. 11. Validation of the construct 1 to 5 after refinement with Ramachandran plot and ProSA web plot.

Table 12. The details of Galaxy Refine server for all constructs before and after refinement.

Construct	Model	GDT-HA	RMSD	MolProbity	Clash score	Poor rotamers	Rama favored
Construct 1	Initial	1.0000	0.000	3.264	11.2	12.7	67.7
	After refinement (model 1)	0.9215	0.508	2.430	19.0	0.5	85.2
Construct 2	Initial	1.0000	0.000	3.777	18.5	21.4	43.6
	After refinement (model 5)	0.8713	0.630	3.064	23.8	2.9	69.7
Construct 3	Initial	1.0000	0.000	3.546	14.1	16.7	51.8
	After refinement (model 2)	0.8772	0.614	2.356	10.3	1.3	79.5
Construct 4	Initial	1.0000	0.000	3.942	26.9	25.0	51.9
	After refinement (model 5)	0.8866	0.600	2.885	39.6	0.0	73.6
Construct 5	Initial	1.0000	0.000	3.327	15.2	7.7	50.9
	After refinement (model 2)	0.8873	0.582	2.781	26.3	1.3	75.2

Table 13. The details of all constructs validation before and after refinement using Ramachandran plot statistics from PROCHECK and Z-score from ProSA web servers.

Constructs	ITASSER C-score	Before Refinement					After Refinement				
		ERRAT quality factor	ProsA Z-score	Ramachandran plot			ERRAT quality factor	ProsA Z-score	Ramachandran plot		
				Most favored zones (%)	Allowed zones (%)	Disallowed zones (%)			Most favored zones (%)	Allowed zones (%)	Disallowed zones (%)
1	-2.11	78.27	-1.14	59.2 (106aa)	37.5 (67aa)	3.4 (6aa)	74.47	-1.23	79.3 (142aa)	17.9 (32aa)	2.8 (5aa)
2	-3.7	43.39	-0.37	35.9 (33aa)	56.5 (52aa)	7.6 (7aa)	43.58	-0.97	51.1 (47aa)	42.4 (39aa)	6.5 (6aa)
3	-3.25	7452	-0.72	36.4 (24aa)	49.1 (39aa)	4.5 (3aa)	52.88	-1.2	60.6 (40aa)	33.3 (22aa)	6.1 (4aa)
4	-3.92	70.21	-4.73	49.4 (40aa)	44.4 (36aa)	6.2 (5aa)	53.84	-4.73	66.7 (54aa)	25.9 (21aa)	7.4 (6aa)
5	-3.4	81.95	-1.99	47.5 (67aa)	48.2 (68aa)	4.3 (6aa)	71.25	-2.6	70.9 (100aa)	25.5 (36aa)	3.5 (5aa)

DISCUSSION

Due to the high prevalence of pneumococci with multi-drug resistance patterns as well as the serotype-dependent limitation and low coverage of polysaccharide-based licensed pneumococcal vaccines on the market, priority must be given to design and the development of surface-exposed and highly conserved protein-based vaccines against all pneumococcal strains. These vaccine candidates must be able to stimulate both the humoral and cellular immune systems along with immunological memory [50]. Pneumococcal heat-shock protein 40 (DnaJ) is highly conserved virulence factor which contributes to pneumococcal colonization, stimulates the innate immune response, and subsequent mucosal immunity [51]. In pneumococcal meningitis or other invasive infections, B-cell based immune responses are more important. Because DnaJ showed protection against pneumococcal infection and is a highly conserved protein, it is a very suitable candidate for development of pneumococcal protein-based vaccines. However, since DnaJ mainly stimulates cellular immunity, it is very valuable to identify and study the immune regions of the protein, focusing on the regions that stimulate humoral immunity.

In-silico techniques or Immunoinformatics studies could be very applicable to reduce the cost and time of experimental research [52, 53] and helpful for epitope mapping of vaccine candidate proteins [54]. To our knowledge, this is the first study of *in-silico* analysis of DnaJ. We first evaluated the conserved regions of the DnaJ amino acid sequence among clinical prevalent pneumococci. The results showed that not only all serotypes had DnaJ protein, indicating the conservation of this protein in pneumococci, but also their amino acid sequences were more conserved. Only 11 of the 387 amino acids had a variable sequence at positions of 70, 73, 197, 201, 224, 297, 315, 355, 358 and 368. We analyzed DnaJ for its physicochemical characteristic. The molecular weight of DnaJ was 40 kDa. It has reported that the proteins with a molecular weight of less than 100 kDa are suitable for vaccine design due to its easy expression and purification steps [42]. The SOLpro prediction result demonstrated the soluble feature of DnaJ when overexpressed in *E. coli* host. The DnaJ was expected to be almost neutral in nature depending on the theoretical isoelectric point. The aliphatic index (indicating thermostability) and grand average of hydropathicity (GRAVY) were estimated at 60.3 and -0.603, respectively. The negative GRAVY value means that the protein is hydrophilic in nature and may interact with water molecules. The *in-vivo* half-life, which estimates the time for half the amount of protein to be destroyed after synthesis in the cell, has been estimated 30, 20, and 10 hours in mammals, yeast, and *E. coli*, respectively. The instability index was also computed 23.53 and this categorize the protein as stable (II of <40 indicates stability) [27]. Understanding the secondary and tertiary structures of the target protein is critical to vaccine design. The secondary structure of DnaJ contained 30.34% alpha-helix, 20.43% extended strand, and 53.23% random coil using GOR V prediction server. It has been reported that the important shapes of “structural antigens” are natively unfolded protein regions and alpha-helical coiled-coils peptides. Both structural forms have the capacity to retreat into their native structure and therefore be identified by antibodies naturally induced in response to infection [46]. The DnaJ 3D structure was modeled using I-TASSER server. This server is one of the best and most widely used servers for designing three-dimensional protein structures. I-TASSER server uses the multiple threading alignments from PDB to identify structural templates, and designs the 3D structures using repetitive fragment assembly simulations [55]. Using structural refinement servers, we could improve the overall quality factor of the initial DnaJ 3D model predicted by I-TASSER and in Ramachandran plot, disallowed region residues were reduced from 2.3% to 1% after the refinement process. Ramachandran plot also revealed that most of the residues are lacated in the

favoured and allowed regions (99%); demonstrating that the overall model quality is satisfactory. The structural refinement servers optimized the hydrogen-bonding network, minimized the atomic energy of the model, and improved the 3D structure by the molecular dynamics simulation.

Then we also predicted linear and conformational B-cell, and T-cell epitopes using different databases. According to B-cell epitope prediction servers (BCPred, IEDB, and Ellipro), 24 predicted linear B-cell epitopes had 86-100% conservation among prevalent clinical pneumococcal serotypes. Among them, only 13 epitopes were 100% conserved, mostly at positions 5-24, 96-116, 104-118, 141-146, 141-161, 167-187, 148-190, 204-209, 233-242, 231-251, 272-283, 317-337, and 334-341, and sequences of EFYDRLGVSKNASADEIKKA, R N P N A P R Q G D D L Q Y R, G T G H E K, Q G E AG F N G G P, G L N D R Q K V, S G A K P G T S P V T C G R C H G A G V I N V D T Q T P L G M M R R Q V T C D V C H G, N F V Q A A L G D T V D, G C R T C N, A G Q G E A G F N G G P Y G D L Y V V V, F F G G G G S S R N P N A P R Q G D D L , G C R T C N G S G A K P G T S P V T C G , G G A V G D Q Y V T V N V V T P T G L N, V I N V D T Q T P L G M M R R Q V T C D. According to the VaxiJen score, the predicted linear B-cell epitopes had a score ranging from 0.4 to 3.88. The highest percentage of the VaxiJen score (score>1) can be considered more immunogenic. The highest percentage of VaxiJen score was found in GTGHEK, G T Q T G K K F R L R S K G A P S L R G G A V, RNPNAPRQGDDLQYR,GAAGANGGFGGFNGAGGFGG,A GQGEAGFNGGPYGDLYVVV,FFGGGGSSRNPNAPRQGD DL, GCRTCNGSGAKPGTSPVTCG epitopes with a VaxiJen score of 3.88, 1.42, 1.34, 1, 1, 1, and 1, respectively. Among them, only certain epitopes G T G H E K, RNPNAPRQGDDLQYR, AGQGEAGFNGGPYGDLYVVV, F F G G G G S S R N P N A P R Q G D D L, and GCRTCNGSGAKPGTSPVTCG were 100% conserved among clinical prevalent pneumococcal serotypes. Although DnaJ has been reported to play a major role in cellular immunity, our immunocomputing results may indicated the high potential of DnaJ to stimulate humoral immunity.

We also investigated MHC-II binding epitopes for eight HLA-II alleles [HLA-DRB* (01:01- 03:01- 04:01- 07:01- 08:01- 11:01-13:01- 15:01)], and mouse MHC-II H2 alleles (I-Ad, I-Ab, and I-Ed). The conservancy percentage and VaxiJen score were calculated for predicted helper T-cell epitopes. Of 76 predicted epitopes according to the IEDB server with a score <5, 55 epitopes were 100% conserved and 21 epitopes had a conservancy score of 82.61 to 91.3%. Among the fully conserved epitopes, some were non-immunogenic with a VaxiJen score < 0.5. In total, 45 epitopes had a VaxiJen score ranging from 0.6 to 1.35 and only 15 epitopes had a VaxiJen score ≥ 1 and conservation score from 86.96 to 100% with a sequence of N L N L N F V Q A A L G D T V, Y N L N L N F V Q A A L G D T, L N L N F V Q A A L G D T V D, G D D L Q Y R V N L T F E E A, Q G D D L Q Y R V N L T F E E, F R L R S K G A P S L R G G A, Q T G K K F R L R S K G A P S, T G K K F R L R S K G A P S L, A Y D Q Y G A A G A N G G F G, G K K F R L R S K G A P S L R, Y D Q Y G A A G A N G G F G G , K F R L R S K G A P S L R G G, K K F R L R S K G A P S L R G, T Q T G K K F R L R S K G A P, RAAYDQYGAAGANGG. From above epitopes, only five epitopes were 100% conserved and had VaxiJen score ≥ 1, with sequence of N L N L N F V Q A A L G D T V, Y N L N L N F V Q A A L G D T, L N L N F V Q A A L G D T V D, G D D L Q Y R V N L T F E E A, Q G D D L Q Y R V N L T F E E. Furthermore, the prediction results of cytokine inducer epitopes indicated that DnaJ might be able to induce the production of IL-4 and IFN-γ, suggesting that this protein may play a role in activating both humoral and cellular immunity.

To develop the effective mucosal immunity against extracellular bacteria, the Th1/Th17-polarized immune response is critical [44]. In this study, we analyzed the IL-17 and IFN-γ inducer epitopes as symbolic cytokines of Th17 and Th1, respectively. IL-17 epitope of DnaJ was identified as QKRAA sequence with VaxiJen score 0.6. In agreement with Cui et al.[13], our immuoinformatics investigation showed that DnaJ might induce mucosal and systemic immunity with stimulating the production of high levels of IL-10, IFN-γ and IL-17A cytokines. The results of immune simulation servers are accordance with other our epitope mapping of DnaJ and experimental analysis of DnaJ which have been reported so far.

ClusPro and DimPlot results of DnaJ and HLA-DRB1*01:01 (the most common binding allele in the Iran population) docking complex showed the lowest energy binding of -809.63 Kcal/mol and 63 cluster members indicating good binding affinity and coupling of this protein with human MHCII via 11 hydrogen bonds. It was reported that DnaJ contributes to the activation of bone marrow-derived dendritic cells (BMDCs) and phagocytosis of macrophage, leading to the development of innate, humoral and cellular immunities through Toll-like receptors 4 and 2 (TLR-4 and TLR-2) [20, 21, 48]. For a more detailed understanding of the amino acid residues of DnaJ involved in the interacting interface with TLRs molecules, molecular docking was performed. Docking complex of DnaJ with TLR-2 and TLR-4 showed the lowest energy binding of -865.2 Kcal/mol with 42 cluster members and -864.7 Kcal/mol with 37 cluster members via 20 and 21 hydrogen bonds, 2 and 5 salt bonds and many hydrophobic interactions indicating good binding affinity and coupling of DnaJ with TLR-2 and TLR-4, respectively. The DnaJ residues of (Gly334, Thr333, Gln339, Asn336, Val274, Arg338, Asp280, Asp363, His364, Lys366, Glu297, Gln300, Glu372, Lys358, Ala277, Leu278, Glu345, Phe346, Val341, Ala342, Phe361, Phe362, Leu278, Ile365, Pro356) and (Asn336, Asp337, Lys340, Glu292, Arg316, Ala319, Gly318, Thr333, Glu297, Asp280, Glu339, Arg338, Lys366, Ile365, His364, Asp337, Leu335, Arg308, Pro296, Gly317, Gly334, Val274, Val294, Leu343, Leu278, Phe346, Gly279, Thr281) were predicted to perform roles in the interactions with TLR-2 and TLR-4, respectively.

Finally, we could predict first the immunogenic profile (epitope mapping) of DnaJ for the first time using Immunoinformatics to design the multi-epitopes vaccine

candidates for stimulating the B- and T-helper2 cell-based immunity against DnaJ. Then, we considered the percentage of immunogenicity according to the VaxiJen score and percentage of epitope conservation among DnaJ of clinical prevalent serotypes of pneumococci as the selection criteria. Finally, we introduced five constructs as novel interesting multi-epitope DanJ-based vaccine candidate. We also used IL-17 inducer epitope to design multi-epitope construct to induce and promote both mucosal and humoral immune responses against pneumococci. All constructs were immunogenic, nontoxic, soluble, and stable. The secondary and 3D structures of all construct were predicted and modeled. The structures were validated. In the future, the efficacy of the predicted multi-epitope DnaJ-based vaccine candidates should be confirmed in laboratory and animal models.

ACKNOWLEDGEMENT

The authors thank the Pasteur Institute of Iran, especially the personnel of the Department of Microbiology and Nanobiotechnology at the Pasteur Institute of Iran for their assistance and cooperation with this project.

CONFLICT OF INTEREST

The authors declare they have no conflict of interests.

REFERENCES

1. Weiser JN, Ferreira DM, Paton JC. Streptococcus pneumoniae: transmission, colonization and invasion. Nature Reviews Microbiology. 2018;16(6):355-67. doi:https://doi.org/10.1038/s41579-018-0001-8.
2. Engholm DH, Kilian M, Goodsell DS, Andersen ES, Kjærgaard RS. A visual review of the human pathogen Streptococcus pneumoniae. FEMS microbiology reviews. 2017;41(6):854-79. doi:https://doi.org/10.1093/femsre/fux037.
3. Avarvand AY, Halaji M, Zare D, Hasannejad-Bibalan M, Ebrahim-Saraie HS. Prevalence of Invasive Streptococcus pneumoniae Infections among Iranian Children: A Systematic Review and Meta-Analysis. Iranian Journal of Public Health. 2021;50(6):1135. doi:https://doi.org/10.18502%2Fijph.v50i6.6412.
4. Yousefi M, Mohammadi M, Afshar D, Nazari-Alam A. Evaluation of frequency, drug resistance and serotyping of streptococcus pneumoniae in iran: A systematic review. Journal of Babol University of Medical Sciences. 2021;23(1):1-10.
5. Brueggemann AB, van Rensburg MJJ, Shaw D, McCarthy ND, Jolley KA, Maiden MC et al. Changes in the incidence of invasive disease due to Streptococcus pneumoniae, Haemophilus influenzae, and Neisseria meningitidis during the COVID-19 pandemic in 26 countries and territories in the Invasive Respiratory Infection Surveillance Initiative: a prospective analysis of surveillance data. The Lancet Digital Health. 2021;3(6):1-11. doi:https://doi.org/10.1016/S2589-7500(21)00077-7.
6. Cohen R, Finn T, Babushkin F, Geller K, Alexander H, Shapiro M et al. High rate of bacterial respiratory tract co-infections upon admission amongst moderate to severe COVID-19 patients. Infectious Diseases. 2021:1-11. doi:https://doi.org/10.1080/23744235.2021.1985732.
7. Baskaran V, Lawrence H, Lansbury LE, Webb K, Safavi S, Zainuddin NI et al. Co-infection in critically ill patients with COVID-19: an observational cohort study from England. Journal of medical microbiology. 2021;70(4):1-9. doi: https://doi.org/10.1099/jmm.0.001350.
8. Akbari E, Negahdari B, Faraji F, Behdani M, Kazemi-Lomedasht F, Habibi-Anbouhi M. Protective responses of an engineered PspA recombinant antigen against Streptococcus pneumoniae. Biotechnology Reports. 2019;24:1-7. doi:https://doi.org/10.1016/j.btre.2019.e00385.
9. Feldman C, Anderson R. Current and new generation pneumococcal vaccines. Journal of Infection. 2014;69(4):309-25. doi:https://doi.org/10.1016/j.jinf.2014.06.006.
10. Lagousi T, Basdeki P, Routsias J, Spoulou V. Novel protein-based pneumococcal vaccines: assessing the use of distinct protein fragments instead of full-length proteins as vaccine antigens. Vaccines. 2019;7(1):9-27. doi:https://doi.org/10.3390/vaccines7010009.
11. Yuki Y, Uchida Y, Sawada S-i, Nakahashi-Ouchida R, Sugiura K, Mori H et al. Characterization and Specification of a Trivalent Protein-Based Pneumococcal Vaccine Formulation Using an Adjuvant-Free Nanogel Nasal Delivery System. Molecular Pharmaceutics. 2021;18(4):1582-92. doi:https://doi.org/10.1021/acs.molpharmaceut.0c01003.
12. Moens L, Hermand P, Wellens T, Wuyts G, Derua R, Waelkens E et al. Identification of SP1683 as a pneumococcal protein that is protective against nasopharyngeal colonization. Human Vaccines & Immunotherapeutics. 2018;14(5):1234-42. doi:https://doi.org/10.1080/21645515.2018.1430541.
13. Cui Y, Zhang X, Gong Y, Niu S, Yin N, Yao R et al. Immunization with DnaJ (hsp40) could elicit protection against nasopharyngeal colonization and invasive infection caused by different strains of Streptococcus pneumoniae. Vaccine. 2011;29(9):1736-44. doi:https://doi.org/10.1016/j.vaccine.2010.12.126.
14. Barriot R, Latour J, Castanié-Cornet M-P, Fichant G, Genevaux P. J-Domain proteins in bacteria and their viruses. Journal of molecular biology. 2020;432(13):3771-89. doi:https://doi.org/10.1016/j.jmb.2020.04.014.
15. Cai Y, Yan W, Xu W, Yin Y, He Y, Wang H et al. Screening and identification of DnaJ interaction proteins in Streptococcus pneumoniae. Current microbiology. 2013;67(6):732-41. doi:https://doi.org/10.1007/s00284-013-0424-4.
16. Cao J, Zhang X, Gong Y, Zhang Y, Cui Y, Lai X et al. Protection against pneumococcal infection elicited by immunization with multiple pneumococcal heat shock proteins. Vaccine. 2013;31(35):3564-71. doi:https://doi.org/10.1016/j.vaccine.2013.05.061.
17. Zarouchlioti C, Parfitt DA, Li W, Gittings LM, Cheetham ME. DNAJ Proteins in neurodegeneration: essential and protective factors. Philosophical Transactions of the Royal Society B: Biological Sciences. 2018;373(1738):20160534. doi:https://doi.org/10.1098/rstb.2016.0534.
18. Khan MN, Bansal A, Shukla D, Paliwal P, Sarada S, Mustoori SR et al. Immunogenicity and protective efficacy of DnaJ (hsp40) of Streptococcus pneumoniae against lethal infection in mice. Vaccine. 2006;24(37-39):6225-31. doi:https://doi.org/10.1016/j.vaccine.2006.05.074.
19. Qiu Y, Zhang X, Wang H, Zhang X, Mo Y, Sun X et al. Heterologous prime-boost immunization with live SPY1 and DnaJ protein of Streptococcus pneumoniae induces strong Th1 and Th17 cellular immune responses in mice. Journal of Microbiology. 2017;55(10):823-9. doi:https://doi.org/10.1007/s12275-017-7262-1.
20. Wu Y, Cui J, Zhang X, Gao S, Ma F, Yao H et al. Pneumococcal DnaJ modulates dendritic cell-mediated Th1 and Th17 immune responses through Toll-like receptor 4 signaling pathway. Immunobiology. 2017;222(2):384-93. doi:https://doi.org/10.1016/j.imbio.2016.08.013.
21. Wang X, Yuan T, Yuan J, Su Y, Sun X, Wu J et al. Expression of Toll-like receptor 2 by dendritic cells is essential for the DnaJ-ΔA146Ply-mediated Th1 immune response against Streptococcus pneumoniae. Infection and Immunity. 2018;86(3):e00651-17. doi:https://doi.org/10.1128/IAI.00651-17.
22. Yu CS, Chen YC, Lu CH, Hwang JK. Prediction of protein subcellular localization. Proteins: Structure, Function, and Bioinformatics. 2006;64(3):643-51. doi:https://doi.org/10.1002/prot.21018.
23. Nielsen H, Tsirigos KD, Brunak S, von Heijne G. A brief history of protein sorting prediction. The protein journal. 2019;38(3):200-16. doi:https://doi.org/10.1007/s10930-019-09838-3.

24. Möller S, Croning MD, Apweiler R. Evaluation of methods for the prediction of membrane spanning regions. Bioinformatics. 2001;17(7):646-53. doi:https://doi.org/10.1093/bioinformatics/17.7.646.
25. Madeira F, Pearce M, Tivey AR, Basutkar P, Lee J, Edbali O et al. Search and sequence analysis tools services from EMBL-EBI in 2022. Nucleic acids research. 2022;50(W1):W276-W9. doi:https://doi.org/10.1093/nar/gkac240.
26. Yu M, Zhu Y, Li Y, Chen Z, Sha T, Li Z et al. Design of a Novel Multi-Epitope Vaccine Against Echinococcus granulosus in Immunoinformatics. Frontiers in immunology. 2021;12:1-17. doi:https://dx.doi.org/10.3389%2Ffimmu.2021.668492.
27. Gasteiger E, Hoogland C, Gattiker A, Wilkins MR, Appel RD, Bairoch A. Protein identification and analysis tools on the ExPASy server. The proteomics protocols handbook. 2005:571-607. doi:https://doi.org/10.1385/1-59259-890-0:571.
28. Ahmadi K, Pouladfar G, Kalani M, Faezi S, Pourmand MR, Hasanzadeh S et al. Epitope-based Immunoinformatics study of a novel Hla-MntC-SACOL0723 fusion protein from Staphylococcus aureus: Induction of multi-pattern immune responses. Molecular immunology. 2019;114:88-99. doi:https://doi.org/10.1016/j.molimm.2019.05.016.
29. Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y. The I-TASSER Suite: protein structure and function prediction. Nature methods. 2015;12(1):7-8. doi:https://doi.org/10.1038/nmeth.3213.
30. Lee GR, Heo L, Seok C. Effective protein model structure refinement by loop modeling and overall relaxation. Proteins: Structure, Function, and Bioinformatics. 2016;84:293-301. doi:https://doi.org/10.1002/prot.24858.
31. Wiederstein M, Sippl MJ. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic acids research. 2007;35(suppl_2):1-4. doi:https://doi.org/10.1093/nar/gkm290.
32. Laskowski RA, Rullmann JAC, MacArthur MW, Kaptein R, Thornton JM. AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. Journal of biomolecular NMR. 1996;8(4):477-86. doi:https://doi.org/10.1007/BF00228148.
33. Gupta S, Kapoor P, Chaudhary K, Gautam A, Kumar R, Consortium OSDD et al. In silico approach for predicting toxicity of peptides and proteins. PloS one. 2013;8(9):e73957. doi:https://doi.org/10.1371/journal.pone.0073957.
34. Saha S, Raghava GPS. AlgPred: prediction of allergenic proteins and mapping of IgE epitopes. Nucleic acids research. 2006;34(suppl_2):W202-W9. doi:https://doi.org/10.1093/nar/gkl343.
35. Chen C, Li Z, Huang H, Suzek BE, Wu CH, Consortium U. A fast peptide match service for UniProt knowledgebase. Bioinformatics. 2013;29(21):2808-9. doi:https://doi.org/10.1093/bioinformatics/btt484.
36. Nosrati MC, Ghasemi E, Shams M, Shamsinia S, Yousefi A, Nourmohammadi H et al. Toxoplasma gondii ROP38 protein: bioinformatics analysis for vaccine design improvement against toxoplasmosis. Microbial pathogenesis. 2020;149:1-11. doi:https://doi.org/10.1016/j.micpath.2020.104488.
37. Chen J, Liu H, Yang J, Chou K-C. Prediction of linear B-cell epitopes using amino acid pair antigenicity scale. Amino acids. 2007;33(3):423-8. doi:https://doi.org/10.1007/s00726-006-0485-9.
38. Ponomarenko J, Bui H-H, Li W, Fusseder N, Bourne PE, Sette A et al. ElliPro: a new structure-based tool for the prediction of antibody epitopes. BMC bioinformatics. 2008;9(1):1-8. doi:https://doi.org/10.1186/1471-2105-9-514.
39. Wang P, Sidney J, Kim Y, Sette A, Lund O, Nielsen M et al. Peptide binding predictions for HLA DR, DP and DQ molecules. BMC bioinformatics. 2010;11(1):1-12. doi:https://doi.org/10.1186/1471-2105-11-568.
40. van de Garde MD, van Westen E, Poelen MC, Rots NY, van Els CA. Prediction and validation of immunogenic domains of pneumococcal proteins recognized by human CD4+ T cells. Infection and immunity. 2019;87(6):1-18. doi:https://doi.org/10.1128/IAI.00098-19.
41. Dorosti H, Eslami M, Negahdaripour M, Ghoshoon MB, Gholami A, Heidari R et al. Vaccinomics approach for developing multi-epitope peptide pneumococcal vaccine. Journal of Biomolecular Structure and Dynamics. 2019;37(13):3524-35. doi:https://doi.org/10.1080/07391102.2018.1519460.
42. Sanami S, Zandi M, Pourhossein B, Mobini G-R, Safaei M, Abed A et al. Design of a multi-epitope vaccine against SARS-CoV-2 using Immunoinformatics approach. International journal of biological macromolecules. 2020;164:871-83. doi:https://doi.org/10.1016/j.ijbiomac.2020.07.117.
43. Bui H-H, Sidney J, Li W, Fusseder N, Sette A. Development of an epitope conservancy analysis tool to facilitate the design of epitope-based diagnostics and vaccines. BMC bioinformatics. 2007;8(1):1-6. doi:https://doi.org/10.1186/1471-2105-8-361.
44. Poolman JT. Expanding the role of bacterial vaccines into life-course vaccination strategies and prevention of antimicrobial-resistant infections. npj Vaccines. 2020;5(1):1-12. doi:https://doi.org/10.1038/s41541-020-00232-0.
45. Goodarzi NN, Bolourchi N, Fereshteh S, Badmasti F. Introduction of novel putative immunogenic targets against Proteus mirabilis using a reverse vaccinology approach. Infection, Genetics and Evolution. 2021;95:105045. doi:https://doi.org/10.1016/j.meegid.2021.105045.
46. Shey RA, Ghogomu SM, Esoh KK, Nebangwa ND, Shintouo CM, Nongley NF et al. In-silico design of a multi-epitope vaccine candidate against onchocerciasis and related filarial diseases. Scientific reports. 2019;9(1):1-18. doi:https://doi.org/10.1038/s41598-019-40833-x.
47. Umar A, Haque A, Alghamdi YS, Mashraqi MM, Rehman A, Shahid F et al. Development of a Candidate Multi-Epitope Subunit Vaccine against Klebsiella aerogenes: Subtractive Proteomics and Immuno-Informatics Approach. Vaccines. 2021;9(11):1373-92. doi:https://doi.org/10.3390/vaccines9111373.
48. Su Y, Li D, Xing Y, Wang H, Wang J, Yuan J et al. Subcutaneous immunization with fusion protein DnaJ-ΔA146Ply without additional adjuvants induces both humoral and cellular immunity against Pneumococcal infection partially depending on TLR4. Frontiers in Immunology. 2017;8:686. doi:https://doi.org/10.3389/fimmu.2017.00686.
49. Murthy VL, Stern LJ. The class II MHC protein HLA-DR1 in complex with an endogenous peptide: implications for the structural basis of the specificity of peptide binding. Structure. 1997;5(10):1385-96. doi:https://doi.org/10.1016/S0969-2126(97)00288-8.
50. Converso T, Assoni L, André G, Darrieux M, Leite LCdC. The long search for a serotype independent pneumococcal vaccine. Expert review of vaccines. 2020;19(1):57-70. doi:https://doi.org/10.1080/14760584.2020.1711055.
51. Cui J, Ma C, Ye G, Shi Y, Xu W, Zhong L et al. DnaJ (hsp40) of Streptococcus pneumoniae is involved in bacterial virulence and elicits a strong natural immune reaction via PI3K/JNK. Molecular immunology. 2017;83:137-46. doi:https://doi.org/10.1016/j.molimm.2017.01.021.
52. Shafaghi M, Shabani AA, Minuchehr Z. Rational design of hyper-glycosylated human luteinizing hormone analogs (a bioinformatics approach). Computational biology and chemistry. 2019;79:16-23.
53. Dorosti H, Eslami M, Negahdaripour M, Ghoshoon MB, Gholami A, Heidari R et al. Vaccinomics approach for developing multi-epitope peptide pneumococcal vaccine. Journal of biomolecular Structure & Dynamics. 2019;37(13):3524-35.
54. Tomar N, De RK. Immunoinformatics: an integrated scenario. Immunology - Wiley Online Library. 2010;131(2):153-68.
55. Yang J, Zhang Y. I-TASSER server: new development for protein structure and function predictions. Nucleic acids research. 2015;43(W1):W174-W81. doi:https://doi.org/10.1093/nar%2Fgkv342.