SCIENTIFIC REPORTS
natureresearch

OPEN

Received: 5 March 2019
Accepted: 23 August 2019
Published online: 16 September 2019

Immunoinformatics approaches to explore *Helicobacter Pylori* proteome (Virulence Factors) to design B and T cell multi-epitope subunit vaccine

Mazhar Khan[1], Shahzeb Khan[2], Asim Ali[1], Hameed Akbar[4], Abrar Mohammad Sayaf[2], Abbas Khan[2,3] & Dong-Qing Wei[3]

Helicobacter Pylori **is a known causal agent of gastric malignancies and peptic ulcers. The extremophile nature of this bacterium is protecting it from designing a potent drug against it. Therefore, the use of computational approaches to design antigenic, stable and safe vaccine against this pathogen could help to control the infections associated with it. Therefore, in this study, we used multiple immunoinformatics approaches along with other computational approaches to design a multi-epitopes subunit vaccine against *H. Pylori*. A total of 7 CTL and 12 HTL antigenic epitopes based on c-terminal cleavage and MHC binding scores were predicted from the four selected proteins (CagA, OipA, GroEL and cagA). The predicted epitopes were joined by AYY and GPGPG linkers. B-defensins adjuvant was added to the N-terminus of the vaccine. For validation, immunogenicity, allergenicity and physiochemical analysis were conducted. The designed vaccine is likely antigenic in nature and produced robust and substantial interactions with Toll-like receptors (TLR-2, 4, 5, and 9). The vaccine developed was also subjected to an *in silico* cloning and immune response prediction model, which verified its efficiency of expression and the immune system provoking response. These analyses indicate that the suggested vaccine may produce particular immune responses against *H. pylori*, but laboratory validation is needed to verify the safety and immunogenicity status of the suggested vaccine design.**

Helicobacter pylori causes multiple infections, such as mucosal lymph tissue, peptic ulcers and gastric carcinoma[1–3]. *H. pylori* is mainly associated with gastric infections, along with other diseases. The risk factors for gastric cancer are probably the interaction between host and pathogen virulence factors[4]. Pathogenic factors of *H. pylori* cause chronic inflammation, resulting in tissues narcosis, lesions and subsequently stomach cancer[5–8]. *H. Pylori* related infections are more frequently correlated with multiple known disease causing factors. Not all but a paramount number of these factors are believed to increase the threat of infection and thus upturns the production of pro-inflammatory cytokines. Predominant stomach disorders and virulent proteins are linked with *H. pylori* infection, among them the widely known virulence factors of *H. pylori* are, Cytotoxin associated genes (Cag), the Pathogenicity island (PaI), and are thought to play crucial role in eliciting inflammation[9]. According to available scientific studies, comparing the cagA strains having very few phosphorylated motifs to those having phosphorylation in abundance, the later are more involved in gastric carcinoma[10–12], while OipA is mainly

[1]The CAS Key Laboratory of Innate Immunity and Chronic Diseases, Hefei National Laboratory for Physical Sciences at Microscale, School of Life Sciences, CAS Center for Excellence in Molecular Cell Science, University of Science and Technology of China (USTC), Collaborative Innovation Center of Genetics and Development, Hefei, 230027, Anhui, China. [2]Centre for Biotechnology and Microbiology, University of Swat, Swat, Khyber Pakhtunkhwa, Pakistan. [3]Department of Bioinformatics and Biological Statistics, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai, 200240, P.R. China. [4]Laboratory of Cellular Dynamics, School of Life Sciences, University of Science and Technology of China (USTC), Anhui Sheng, P.R. China. Correspondence and requests for materials should be addressed to D.-Q.W. (email: dqwei@sjtu.edu.cn)

involved in duodenal ulcers[13,14]. Further, human gastric biopsy samples revealed the involvement of OipA[15,16] and cagA in gastric cancer and duodenal ulceration[17,18].

Numerous and drastic effects of the VacA on epithelial cells pave a way to gastric cancer. The increased risk of cancer is correlated with the active form of VacA[19]. The innate immune response is principally dependent on the pattern recognition receptors recognizing the pathogen-associated molecular patterns of the pathogen[20]. The pattern recognition receptor Nod1 is an intracellular receptor; it recognizes the PAMP and is stimulated by CagA strains. Toll-like receptors are the most studied pathogen recognition receptors (PRRs), TLRs family comprises of eleven proteins, and each of them uniquely binds to different PAMPs, they are expressed on surface of the cells[21]. According to the transfected cell line studies, *H. Pylori* elicits the expression of pro-inflammatory genes via different receptors (TLR2, TLR4, TLR5 and TLR9)[22–25]. High level of Alpha 1, 2 and 3 defensins along with β-defensins are quantified in infections caused by *H. pylori* such as intestinal cancer. Higher fold expression can be obtained via using these defensins in *in-silico* cloning.

The pathogenic role of these proteins in *H. pylori* infections makes them more ideal factors for vaccine designing, which will provide long term immunity against *H. pylori* infections. A potent vaccine against the *H. pylori* infection has not yet designed. Literatures on the mechanism of interactions between *H. pylori* and human immune responses are available that can be exploited for designing of a potent vaccine[26]. Epitopes are the antigenic portions of the pathogens that are recognized by host immune system and humoral or cell mediated immunity is triggered against it[27]. Immediately after infection antigen-presenting cells (APCs) activates the cytotoxic T lymphocytes, that have dominant role in killing the infected cells[28]. Peptides bound to MHC acts on surface of *H. pylori* infected cells[29]. Hence, a wide ranges of high binding affinity epitopes are displayed by MHC molecules.

Recent advance in the field of immunoinformatics and the availability of diverse array of tools to design epitope vaccines has profoundly increased the research on the development of novel potential vaccine candidates. The use immunoinformatics approaches and tools are reasonably effective for the development of precise and stable multi-epitope subunit vaccine[30]. Immunoinformatics is accurate, reliable and rapid approach for developing vaccine against virulent pathogens. Taking into account the feasibility and advantages of the vaccine designed through immunoinformatics approaches, this study also aims to design a multi-epitopes subunit vaccine against *H. pylori*. Herein, we used multiple antigenic proteins from *H. pylori* proteome to design B and T-cell epitopes from them. Prediction of HTL epitope was performed and the final vaccine composed of multiple epitopes was constructed. We further used molecular docking, thermodynamics stability profiling, *in-silico* expression and an agent-based modeling tool to verify the stability, expression and immune response reaction provoked by the final vaccine.

Methodology

Collection of proteins. Four proteins were selected for vaccine designing against *H. pylori*, including CagA, OipA, GroEL and VacA based on their role in *H. pylori* infection. Proteins that are important for virulence of a pathogen are suitable for designing a vaccine. The workflow of this scientific study is shown in Fig. 1.

MHC-I binding epitopes (CTL) prediction. CTL epitopes were predicted for four virulent proteins (VacA, GroEL OipA, and CagA,) of *H. pylori* using NetCTL1.2 (http://www.cbs.dtu.dk/services/NetCTL/)[31], an online tool. The server predicts 9-mer epitopes for the query protein based on transport-associated protein, C-terminal cleavage scores and MHC-I binding score. Artificial neural network is employed by the server for C-terminal cleavage and binding of peptides to MHC-I whereas Tap transport efficiency is obtained by the weight matrix. The epitopes were predicted at 0.75(default) threshold.

MHC-II binding epitopes (HTL) prediction. MHC-II binding, 15-mer epitopes, were predicted by a web tool IEDB MHC-II for each of the selected virulent proteins (CagA, GroEL, OipA, and VacA) against a set of seven human HLAs given as;, HLA-DRB1 * 15:01, HLA-DRB4 * 01:01, HLA-DRB3 * 01:01, HLA-DRB5*01:01, HLA-DRB1 * 03:01 HLA-DRB3 * 02:02, HLA-DRB1 * 07:01. The server defines the affinity of peptides towards the MHC-II on the basis of IC_{50} score. IC_{50} values less than 50 nM shows highest binding affinity towards MHC-II, IC_{50} less than 500 nM indicates midrange affinity whereas IC_{50} less than 5000 nM corresponds to lowest binding affinity. The percentile rank is inversely related to the IC_{50} score.

B-cell epitopes prediction. B lymphocytes are the most important factor of the immune system. It is responsible for secreting antibodies that in turn, provides long term immunity. For the prediction of B-cell linear epitopes, we employed an online server BCPred (http://ailab.ist.psu.edu/bcpred/)[32], the server exploits kernel technique to predict 20-mer linear B-cell epitopes. BCPred uses a support vector machine (SVM) algorithm for B-cell (linear) epitopes prediction.

For the prediction of discontinuous B cell epitopes, we used a web server ElliPro (http://tools.iedb.org/ellipro/). Residues clustering algorithms along with Thornton's technique is employed by ElliPro suite for prediction of B-cell (conformational) epitopes. The server exploits Modeler v9.20 in order to create 3D coordinates of the epitopes predicted by the server. The server assigns a PI (protrusion index) value to each predicted epitope[33].

Construction of vaccine sequence. A set of CTL and HTL epitopes were selected on the basis of its high binding score and non-allergenic nature. In order to make the final multi-epitopes vaccine construct, the selected CTL epitopes were linked together by AAY linker whereas GPGPG linkers were used for HTL epitopes. Epitopes representation and proper separation are enhanced by linkers. The linkers are significant for two reasons; (1) linkers effectively prevent the formation of neo-epitopes (junctional epitopes), (2) and improve epitope presentation[34–37]. Furthermore, in order to improve immunogenicity, human β-defensins was added at N-terminus of the vaccine using EAAAK linker as an adjuvant.

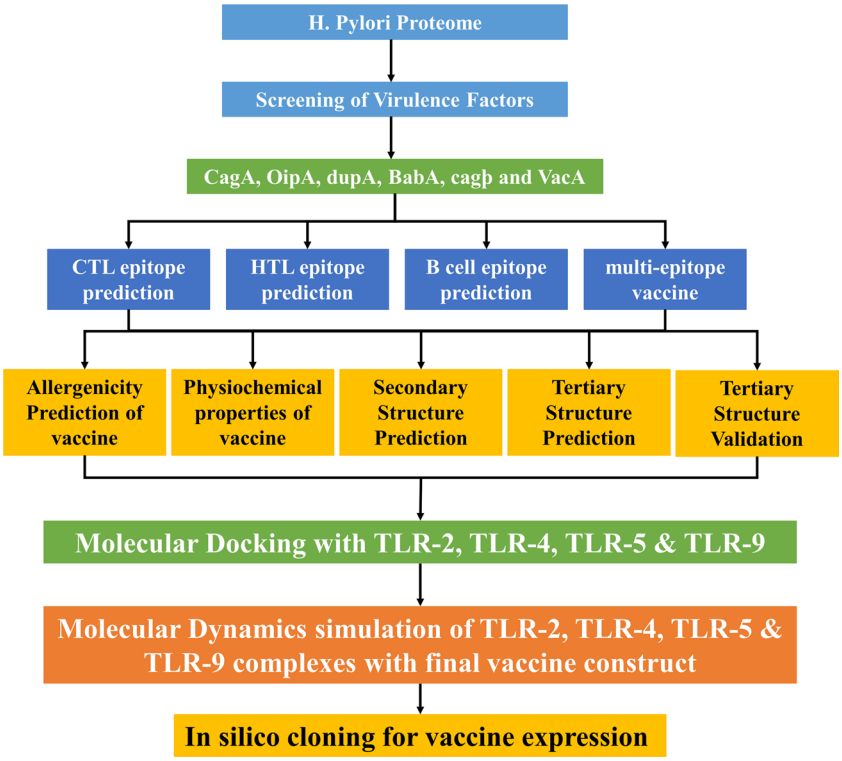

Figure 1. Workflow of this scientific study. The complete methodology is shown is four steps: the selection of proteins suitable for vaccine designing, the second step comprised of T and B cell epitopes prediction, construction of vaccine by joining together the epitopes, at last, molecular docking with TLRs and MD simulations for stability of complexes and finally *in silico* expression was performed.

Allergenicity profiling. To make sure the vaccine don't elicit any allergen response we used an online antigenicity prediction server AlgPred (http://www.imtech.res.in/raghava/algpred/), the server uses several algorithms (SVMc + MAST + IgEepitope + ARPs BLAST) to check allergenicity of query sequence[38]. The server results are reliable and have 85% precision at −0.4 threshold.

Vaccines antigenicity profiling. Antigenicity evaluation is an important step in vaccine designing. In this study, we utilized two servers for antigenicity prediction, ANTIGENpro (http://scratch.proteomics.ics.uci.edu/)[39] and Vaxijen 2.0 (http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html). For antigenicity evaluation of the query sequence the former uses microarray data, it employs pathogen independent and sequence-based approach to predict the antigenicity value. However, the later predicts antigenicity on the basis of physicochemical properties for analysis of the query sequence. Vaxijen have 70 to 89% precision depending on the organism.

Physiochemical properties evaluation. For assessment of several physicochemical properties of the vaccine sequence an online freely accessible web server ProtParam (http://web.expasy.org/protparam/)[40] was utilized. The server calculates amino acid composition, molecular weight, aliphatic score, *in-vitro* half-life, *in vivo* half-life, instability index, theoretical pI, and GRAVY.

Prediction of secondary structure. For the computation of secondary structure, we used a web server PSIPRED (http://bioinf.cs.ucl.ac.uk/psipred/)[41] for the vaccine sequence, the server provides results with high accuracy. Proteins showing homology to our vaccine construct were identified via PSI-Blast, these sequences were then used to create PSSM (position-specific scoring matrix). Furthermore, the server uses neural network (feed forward) to process the PSSM and predict the secondary structure elements.

Vaccine 3D structure prediction. For the prediction of 3D structure for our vaccine sequence, we employed a widely used, freely available online web server RaptorX (http://raptorx.uchicago.edu/)[42]. The server along with 3D structure prediction, can predict disordered regions, solvent accessibility, secondary structures, binding sites and contacts. RaptorX predicts the absolute global quality and comparable global quality for each of the residues of the query sequence. In order to check the 3D structure, we used Pymol software.

Vaccine 3D structure refinement. The 3D structure was subjected to an online server Galaxy Refine (http://galaxy.seoklab.org/)[43] for further improvement. The server employs the CASP10 technique in order to refine the query 3D structure. The side chains of proteins are reconstructed by CASP10 technique followed by repacking and

use of simulations of the 3D structure for relaxation. Galaxyrefine improved the structural and global quality of the 3D structure. YASARA[44] software was used for energy minimization and correction of the structure.

Vaccine 3D structure validation. Three freely available web tools ProSA-web (https://prosa.services.came.sbg.ac.at/prosa.php)[45], ERRAT (http://services.mbi.ucla.edu/ERRAT/)[46] and RAMPAGE (http://mordred.bioc.cam.ac.uk/rapper/~rampage.php)[47] were utilized for the refined 3D structure validation. Prosa-Web highlight and plots overall excellence scores of the errors calculated in the query 3D structure, ERRAT server focus on the nonbonded interactions within the given structure whereas RAMPAGE server investigates Ramachandran plot, it utilizes PROCHECK principles for validation via Ramachandran plot. A separate plot is drawn for Proline-glycine residues.

Interaction analysis vaccine with TLR receptors. For the evaluation of the interaction between the vaccine and human toll-like receptors 2, 4, 5 and 9 we employed PatchDock (http://bioinfo3d.cs.tau.ac.il/PatchDock/)[48]. It computes separating scores, surface fix coordinating scores, and portrayal of atomic shape for molecular docking. The server divides both the vaccine and TLRs into small patches in agreement with the surface shape. These small patches resemble unique shapes, which can distinguish between puzzle pieces visually. Another algorithm does superimposition of these small patches after identification of the patches. Furthermore, for refining and re-scoring the molecular docking complexes another server was employed, FireDock[49] to get the best structure. The refined complexes provided by FireDock are based on several factors such as atomic contact energy, partial electrostatics and vdW.

Molecular dynamics simulation. Amber 14 setup was utilized for all the complexes to perform MD simulations[50]. Addition of Sodium ions neutralized the system by using "tleap". A TIP3P water box was used to solvate the system. To remove the clashes and constraints in the system energy minimization was carried out on AMBER version 14. The minimized system was used for MD analysis using PMEMD.cuda. A cutoff radius of 10 Å for non-bond interactions was considered. SHAKE and Particle-Mesh Ewald (PME) algorithms were utilized[51]. Post simulation analysis of the ten nanoseconds trajectories was performed using CPPTRAJ and PTRAJ[52]. Finally, RMSDs and RMSFs of all the systems were calculated.

Validation

Codon optimization and *in-silico* vaccine expression. Jcat was used for reverse translation and optimization of codons in order to achieve maximum expression in *E. coli* cellular machinery, Jcat calculated the GC content and CAI scores for the query sequence in order to ensure maximum expression. Prokaryotic ribosome and restriction binding sites and rho-independent termination options were selected. Restriction sites of Xhol and Ndel were added to the reverse translated sequence. The final vaccine was then cloned into pET-28a(+) plasmid using snap-gene software.

Immune Simulation. C-ImmSim[53] is an online server, which uses an agent-based modelling approach to estimate the effect of a foreign particle, antigen, on the immune system. The server reflects the immune response against the antigen using PSSM method. Antibodies, cytokines and interferon production upon the injection of the vaccine is calculated. In addition, Th1 and Th2 responses are also forecasted by the webserver. Default parameters were used to plot the Simpson Index or D (a measure of diversity).

Results

***Helicobacter pylori* protein sequences retrieval.** The amino acid sequences of the selected proteins GroEL, OipA, CagA and VacA of *Helicobacter pylori* were retrieved from Gene Bank using GI: 446963037, GI: 446632395, GI: 2498230 and GI 15645505 to design multi-epitope subunit vaccine for immune response against *H. pylori* infection. The selection of these proteins are based on their target efficacy and reliability. Immunogenic behavior by vaccine was enhanced by fusing the human β-defensins 4A protein (Uniprot ID: O15263) as adjuvant, which regulates the immune response by coupling with vaccine protein.

Antigenicity of selected *H. pylori* proteins. Vaxijen2.0[54], a web server, determined the antigenicity of each protein. The results from the server showed that each protein possesses antigenic properties. Scores revealed by this server includes 0.46 for GroEL, 0.56 for OipA, 0.45 for CagA and 0.58 for VacA respectively.

Cytotoxic and HTL epitopes prediction. A total of 72 CTL epitopes were predicted for all of the proteins by the NetCTL 1.2. Among the total 72 epitopes, only seven epitopes based on the defined criteria, MHC binding score and non-allergenic nature, were selected. Similarly, the CTL epitopes given in Supplementary Table S1 by using the IEDB MHC-II server were predicted for the selected proteins. The HTL epitopes were predicted for a set of seven HLAs (HLA-DRB1 * 03:01, HLA-DRB4 * 01:01, HLA-DRB1 * 07:01, HLA-DRB3 * 02:02, HLA-DRB1 * 15:01, HLA-DRB3 * 01:01 and HLA-DRB5 * 01:01). HTL epitopes are given in Table 1. A variable number of HTL epitopes such three from GroEL, CagA and OipA each at different positions and four from VacA were predicted by the server.

Vaccine construction. Twelve HTL and seven CTL epitopes based on their binding scores, antigenic nature and the non-allergenic property is given in Table 2 were joined together by using GPGPG and AAY linkers to construct the final multi-epitopes vaccine. Adjuvant was attached to the N-terminus of the vaccine for the protection from degradation. EAAAK linker was used to join the adjuvant to the CTL epitopes. On the other hand, AAY and GPGPG linkers joined the CTL and HTL epitopes. Figure 2 is showing the final structure of the vaccine construct

Protein	S.NO	Allele	Start	End	Peptide sequence	method	Percentile rank
GroeL	1	HLA-DRB3*01:01	336	350	GHSHDVKDRVAQIKT	Consensus (comb.lib./smm/nn)	0.11
	2	HLA-DRB3*01:01	337	351	HSHDVKDRVAQIKTQ	Consensus (comb.lib./smm/nn)	0.11
	3	HLA-DRB1*03:01	186	200	LDVVEGMQFDRGYLS	Consensus (smm/nn/sturniolo)	0.88
oipA	1	HLA-DRB1*07:01	219	233	NLTPFNQVKSRTIFQ	Consensus (comb.lib./smm/nn)	0.01
	2	HLA-DRB1*07:01	220	234	LTPFNQVKSRTIFQL	Consensus (comb.lib./smm/nn)	0.01
	3	HLA-DRB1*07:01	191	205	KASRHVFRKSSGLVI	Consensus (comb.lib./smm/nn)	0.05
cagA	1	HLA-DRB1*03:01	806	820	SRVEQVLADLKNFSK	Consensus (smm/nn/sturniolo)	0.09
	2	HLA-DRB1*15:01	116	130	KFGDQRYQIFTSWVS	Consensus (smm/nn/sturniolo)	0.42
	3	HLA-DRB1*15:01	119	133	DQRYQIFTSWVSHQK	Consensus (smm/nn/sturniolo)	0.42
VacA	1	HLA-DRB3*02:02	1148	1162	FAFFRNALVLKPSVG	NetMHCIIpan	0.17
	2	HLA-DRB3*02:02	1149	1163	AFFRNALVLKPSVGV	NetMHCIIpan	0.5
	3	HLA-DRB5*01:01	153	167	GGDLDVNMQKATLRL	Consensus (smm/nn/sturniolo)	0.63

Table 1. Helper T-Lymphocytes epitopes are given in the table along with their scores predicted by IEDB MHC class II server.

Sequence	Combine score	Type of Epitope
TTTATVLAY	3.34	CTL epitope
KVNYYSDDY	1.4	CTL epitope
YSDDYGDKL	2.02	CTL epitope
GISQLREEY	1.07	CTL epitope
NRDARAIAY	1.17	CTL epitope
GIDTGNGGF	0.84	CTL epitope
FASNLGMRY	2.49	CTL epitope
GHSHDVKDRVAQIKT	0.11	HTL epitope
HSHDVKDRVAQIKTQ	0.11	HTL epitope
LDVVEGMQFDRGYLS	0.88	HTL epitope
NLTPFNQVKSRTIFQ	0.01	HTL epitope
LTPFNQVKSRTIFQL	0.01	HTL epitope
KASRHVFRKSSGLVI	0.05	HTL epitope
SRVEQVLADLKNFSK	0.09	HTL epitope
KFGDQRYQIFTSWVS	0.42	HTL epitope
DQRYQIFTSWVSHQK	0.42	HTL epitope
FAFFRNALVLKPSVG	0.17	HTL epitope
AFFRNALVLKPSVGV	0.5	HTL epitope
GGDLDVNMQKATLRL	0.63	HTL epitope

Table 2. A list of both HTL and CTL epitopes used to model the final vaccine structure based on their respective scores. These HTL and CTL epitopes sequences were used in the final structure.

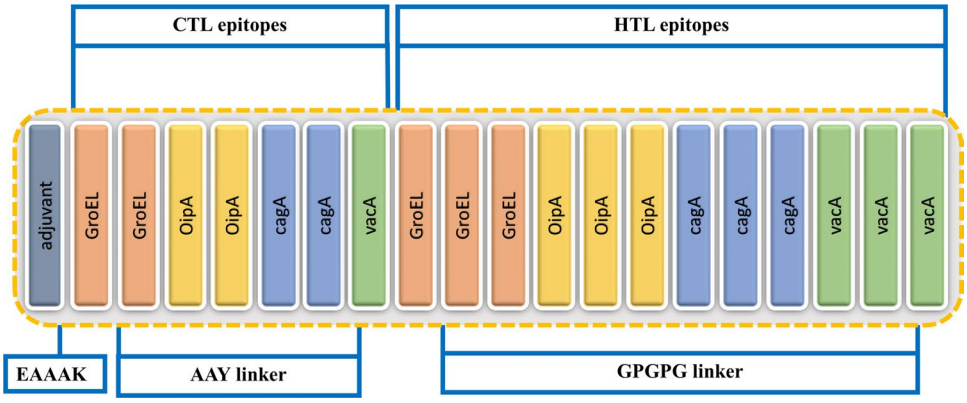

Figure 2. Structural arrangement of the final vaccine candidate constructed from CTL and HTL epitopes using different linkers.

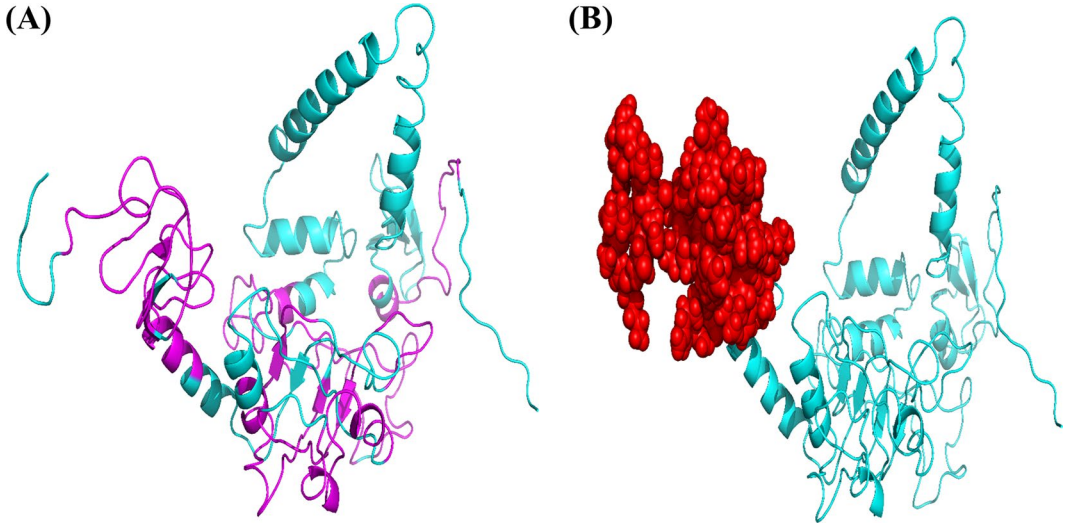

Figure 3. Prediction of linear B-cell epitopes by BCPred while discontinuous epitopes are predicted by ElliPro server. Panel A is showing B-cell (linear) epitopes (magenta) highlighted in vaccine 3D structure while B-cell (Conformational) epitopes are depicted in panel B (red color).

including. The predicted epitopes were evaluated by BLASTp to avoid epitopes homologous with human proteins. Only 16% similarity was reported which is due to the human β-defensins sequence attached at the N-terminus.

B cell epitopes prediction. BCPred predicted 12 (linear) B-cell epitopes of 20 amino acids in length each with scores ≥0.90 were selected. Likewise, ElliPro suite predicted the discontinuous B-cell epitopes. In total, 51 amino acids were classified as B-cell (conformational) epitopes with a score of 0.819. The server default parameters were applied to predict these epitopes. The predicted conformational and linear epitopes are given in Fig. 3A,B.

Physio-chemical properties calculations. Several essential properties such as allergenicity, molecular weight, half-life and instability were calculated. The server predicted that the final vaccine construct is non-allergenic by predicting −0.42 score. The default threshold −0.4 was used. The basic nature of the vaccine was confirmed by determining the Isoelectric point. PI score 9.64 with 41.25 kDa molecular weight (MW) was calculated by the server. On the other hand, half-life >30, >20 and >10 hours, *in vitro* (in mammal's reticulocyte) yeast *E. coli* (*in vivo*) was calculated. Instability index of 17.28 confirmed that the vaccine would be stable in the experimental setup. GRAVY was calculated as −0.261, while the aliphatic coefficient was 68.28. In addition, online servers ANTIGENpro and VaxiJen have calculated 0.68 and 0.93 antigenic scores for the vaccine design, suggesting that the vaccine is immunogenic and can trigger an adequate immune effect.

Secondary structure elements prediction. A web server PSIPRED was used to predict the secondary structure, shown in Fig. 4, of the vaccine protein. The forecasted secondary structure elements include 35.8% α-helix, 12.9% β-sheet and 51.3% coils.

3D structure modelling and refinement. 3-dimensional structure of the multi-epitope final vaccine was modelled by using RaptorX web tool. The final structure obtained from this server has been shown in Fig. 5. Multi-template modelling predicted the 3D structure in 3 domains. For each domain, different templates were utilized. Templates with accession numbers 1FD3, 2LXO for domain-1, 1IOK, 5CDI, 3RTK, 1GRL, 1WE3 for domain-2 while 5MZ5, 2GOK, 4E57, 3B3J and 4WA0 were used for domain-3 modeling. It was found that 3RTK with *p*-value 7.38e-05, attained score 83 and 18% of identity. Galaxy Refine tool was utilized then to refine the selected model-1 which was then evaluated using RMSD (0.442), MolProbity (2.001), Ramachandran plot (93.3), GDT-HA (0.9365), Poor rotamers (0.3) and clash score (11.2). YASARA energy minimization tool was used to further improve the quality of the vaccine structure.

3D structure validation. Validity of the refined 3-dimensional modelled structure of the multi-epitope vaccine protein was carried out by using RAMPAGE, which verified that 82.5% residues of the vaccine are in the favoured region, 13.9% in allowed while only 3.6% in outlier area. ProSA-web further assessed the quality of the model, which reported the Z-score of −3.05 while 86.3% confidence was reported by the ERRAT server. The plots obtained from RAMPAGE server and PROSA-web are given in Fig. 6.

Interaction analysis of the vaccine with TLR receptors. Multiple TLR receptors (TLR-2, 4, 5 and 9) were used as receptors and are given in Fig. 7. Patchdock server reported the top ten interaction models ranked by geometry of the protein's surface and electrostatic complementarity. Refinement and rescoring of the top complexes was performed on FireDock web tool which reported the best model with electrostatic interactions (6.12), Van Der Waals associations (−15.09), atomic contact energy (1.05), and binding free energy (−11.97).

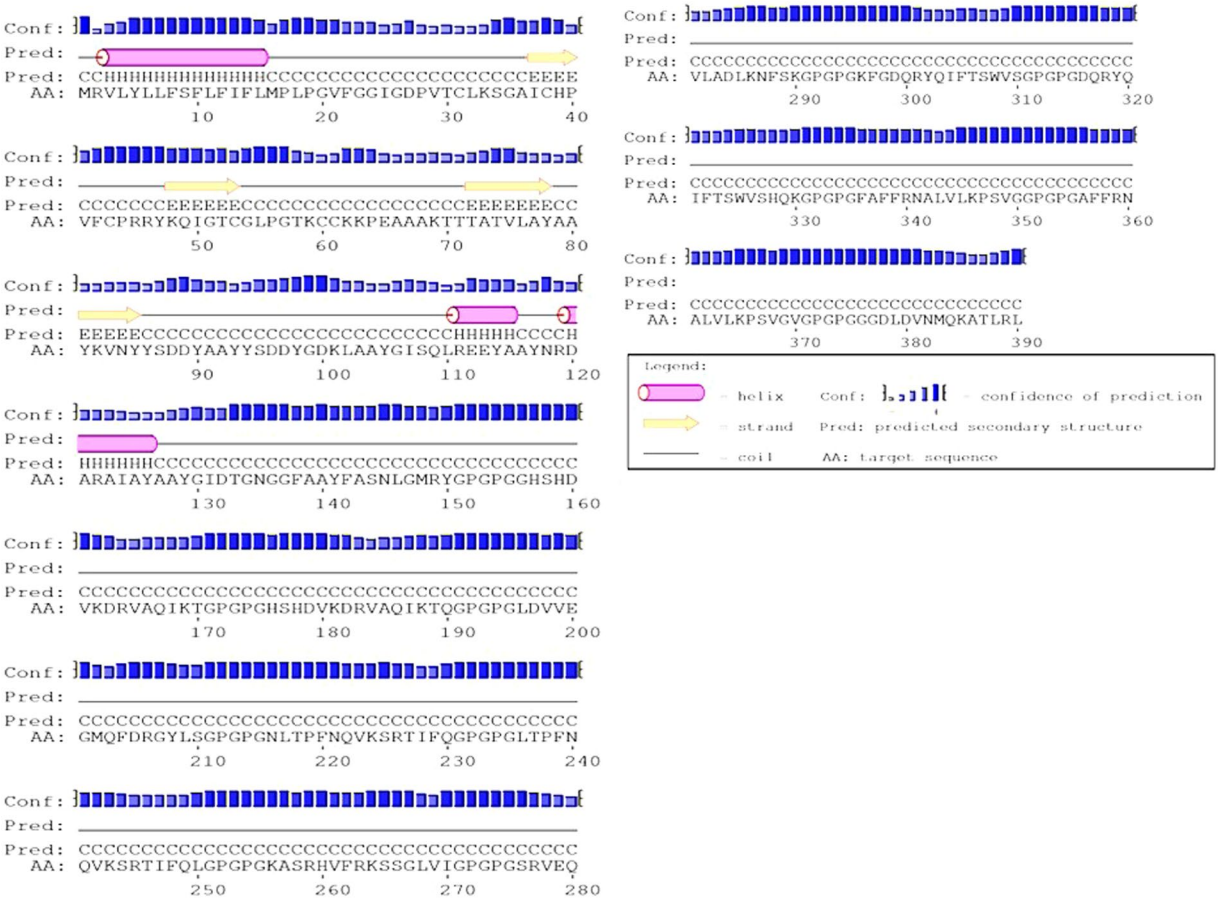

Figure 4. PSIPRED web tool predicted secondary structure elements. The structure was divided as 12.9% β-strand, 35.8% α-helix, and 51.3% coils.

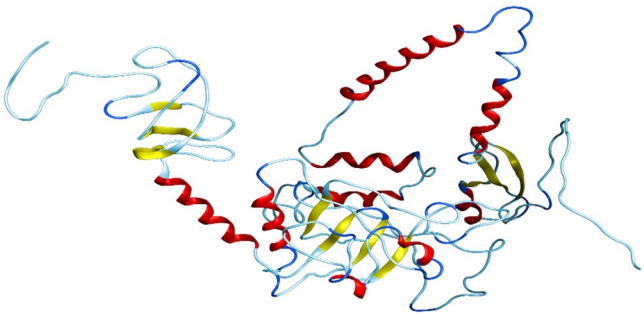

Figure 5. 3D structure of the final vaccine modelled by using multi-templates. The respective elements such as loops are given in sky blue colour, helices are in red while beta-sheets are coloured yellow.

MD simulation of TLRs-vaccine complexes. Complexes of final vaccine construct with the selected receptors (TLR-2, 4, 5 and 9) were subjected to MD simulations to calculate the residual stability and fluctuations. RMSD was evaluated to quantify each system's stability while RMSF was evaluated as shown in Fig. 8 for residual fluctuation. Ten nanoseconds simulations for all the systems were established, which reported variable RMSDs such as 0.7 nm (TLR-2), 0.7 nm (TLR-5), 0.5–0.6 nm (TLR-4) and 0.7 nm for TLR-9. Residual fluctuation (RMSF) was observed within the allowed range with the exception of a few greater fluctuating residues.

Codon optimization and *in-silico* vaccine expression. Herein, Jcat, was utilized to quantify the expression level of the multi-epitopes vaccine *Escherichia coli* (K12 strain). A total of 1170 nucleotides were used as input. Codon optimization was performed and Codon Adaptation Index (CAI) was calculated which was found to be 0.95 with GC contents of 55%. These results indicate the better expression of the final multip-epitope vaccine in *E. coli* (K-12 strain). It has been reported that 35% to 70% GC contents for better expression is required.

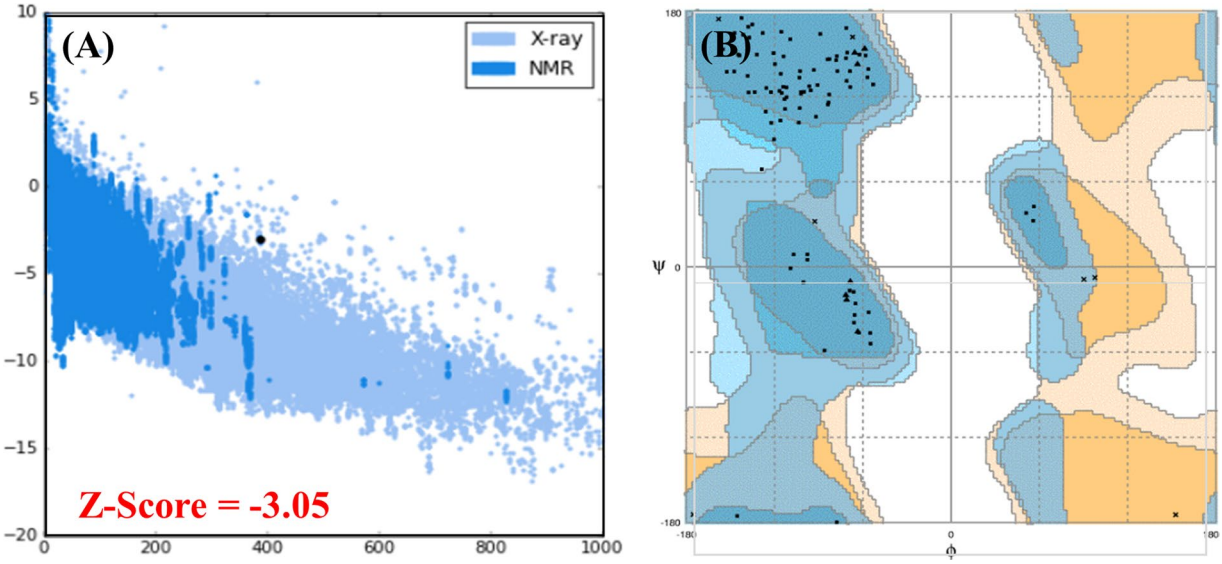

Figure 6. Ramachandran plot and PROSA-web server analysis are given in the above figure. The Z-score (−3.05) predicted by the PROSA-web is given in panel (A) while the distribution of each amino acid in the favoured, allowed and disallowed region are shown by using Ramachandran plot in the panel (B). Overall 82.5% residues were reported to be in favoured region, 13.9% in allowed while 3.6% were reported to be in disallowed region respectively.

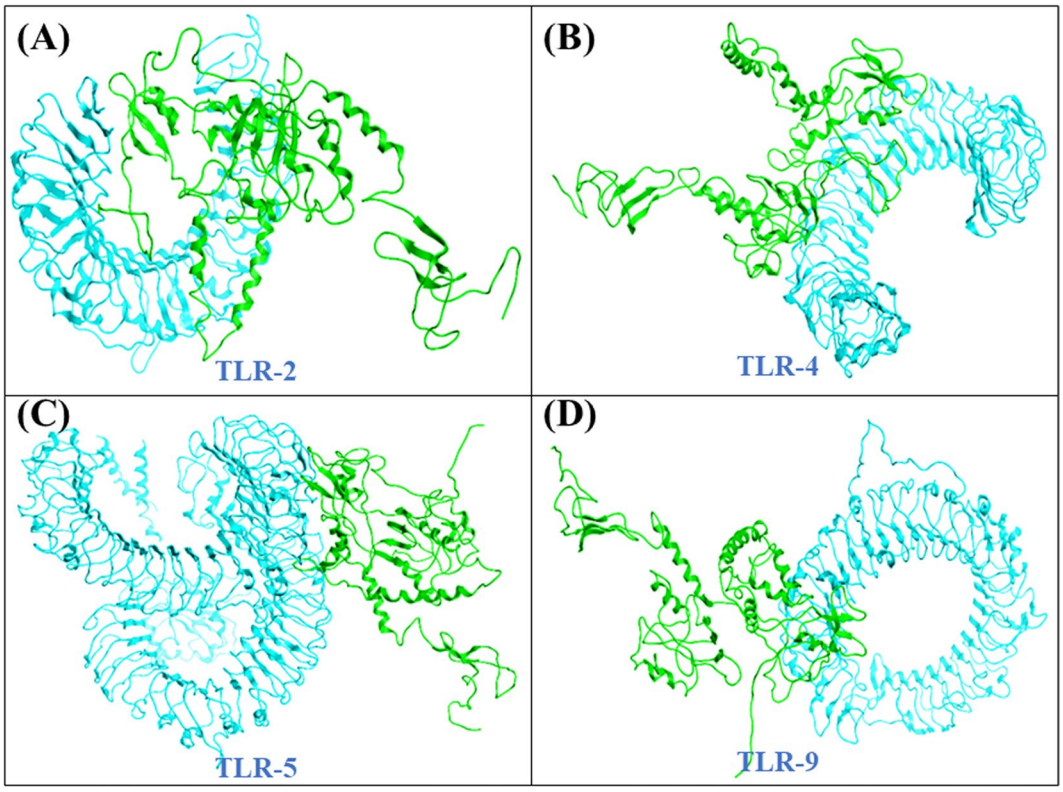

Figure 7. Refined docked complexes of all the TLRs with the vaccine construct. Docking of the final vaccine against each receptor is given in their respective panels. (**A**) TLR-2, (**B**) TLR-4, (**C**) TLR-5 and (**D**) TLR-9. The magenta colour is showing each receptor while green colour is the final vaccine structure.

Restriction sites (NdeI and XhoI) were added 5′ and 3′ ends and cloning of the optimized nucleotide sequence in the pET28a (+) vector was performed. The overall construct of this vaccine along with the vector and restriction sites are given in Fig. 9.

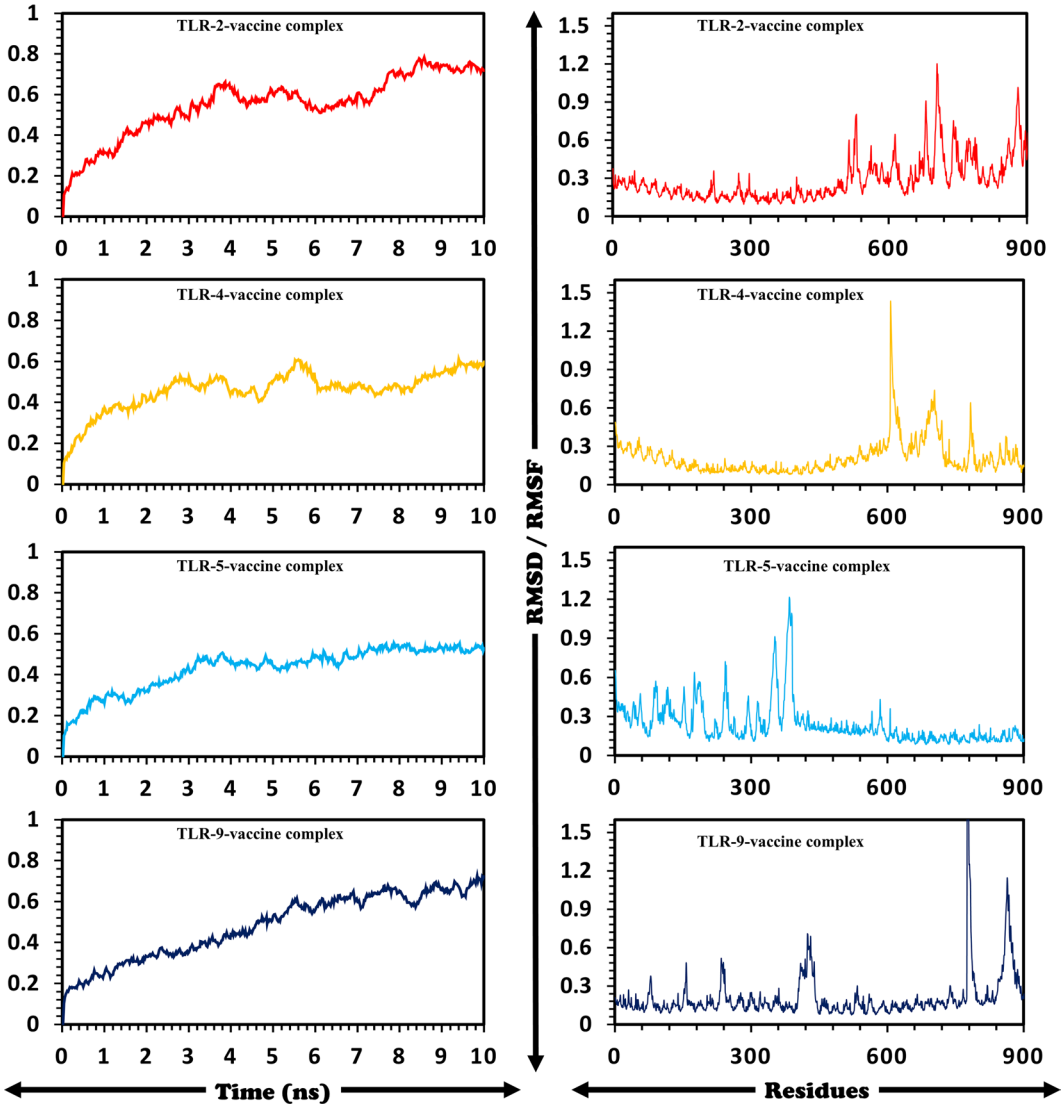

Figure 8. Residual stability and fluctuations i-e (RMSDs) and RMSFs for all the systems are given. Ten nanoseconds simulations were carried out. The left panel (RMSDs) while RMSFs are given in the right panel.

Immune simulation. It was found that results from the C-ImmSim server are compatible with prior experimental studies' reported the immune responses in *H. pylori* infections. Figure 10 shows the figures obtained from C-ImmSim server. Figure 10(A,C) show that the main reaction triggers the production of IgG and IgM antibodies, while the secondary reaction indicates enhanced levels of IgG1 + IgG2, IgM, and IgG + IgM (B-cell populations) antibodies. Earlier screening of *H. pylori* patients was reported to have IgM and IgG antibodies[55]. Furthermore, IFN- γ, Th1 and Th2 responses were also tested. Previously studies revealed that the *H. pylori* infections are characterized by Th1, Th2 and IFN- γ and both Th1 and Th2 responses are required for the protection[56]. Herein, Fig. 10(B,C), the IFN- γ concentration and TH cell population are reported to be high. Thus, these findings show that our vaccine design could effectively trigger the immune response and provide the basis for immunity against *H. pylori*-associated infections.

Discussion

To control infectious diseases, the use of antibiotics is not only expensive but is also precariously resulting in the continuous generation of resistant microbes. Alternatively, vaccination can be effectively applied to a vast population to prevent infections. Conversely, to the classical approaches for vaccinations, today's cutting-edge research has made possible the generation of "subunit vaccines" which constitutes of the particular pathogenic protein sequences of the microbes and are fully capable of effective stimulation of immune response. Due to the availability of comprehensive information about the genomes and proteome of the microbes designing novel and efficient subunit vaccines has become more practically possible. As conventional vaccine development approaches are now almost outdated due to their low effectiveness and high costs in terms of budget and time. Designing vaccines via immunoinformatics approach is comparatively stable, safe, inexpensive, specific and more effective.

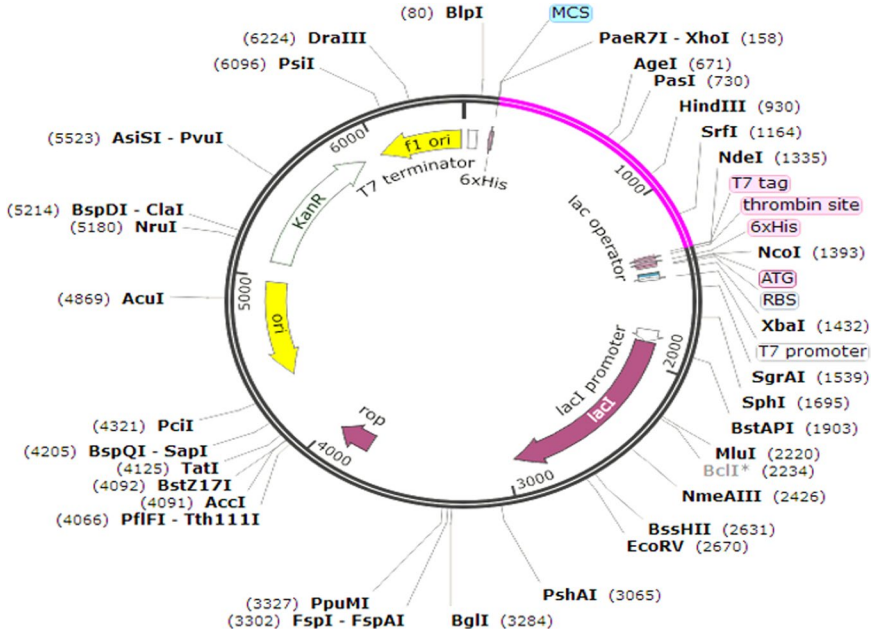

Figure 9. Structure of *In-silico* vector pET28a (+) construct including vaccine along with the vector and restriction sites. The vaccine sequence is highlighted in pink colour.

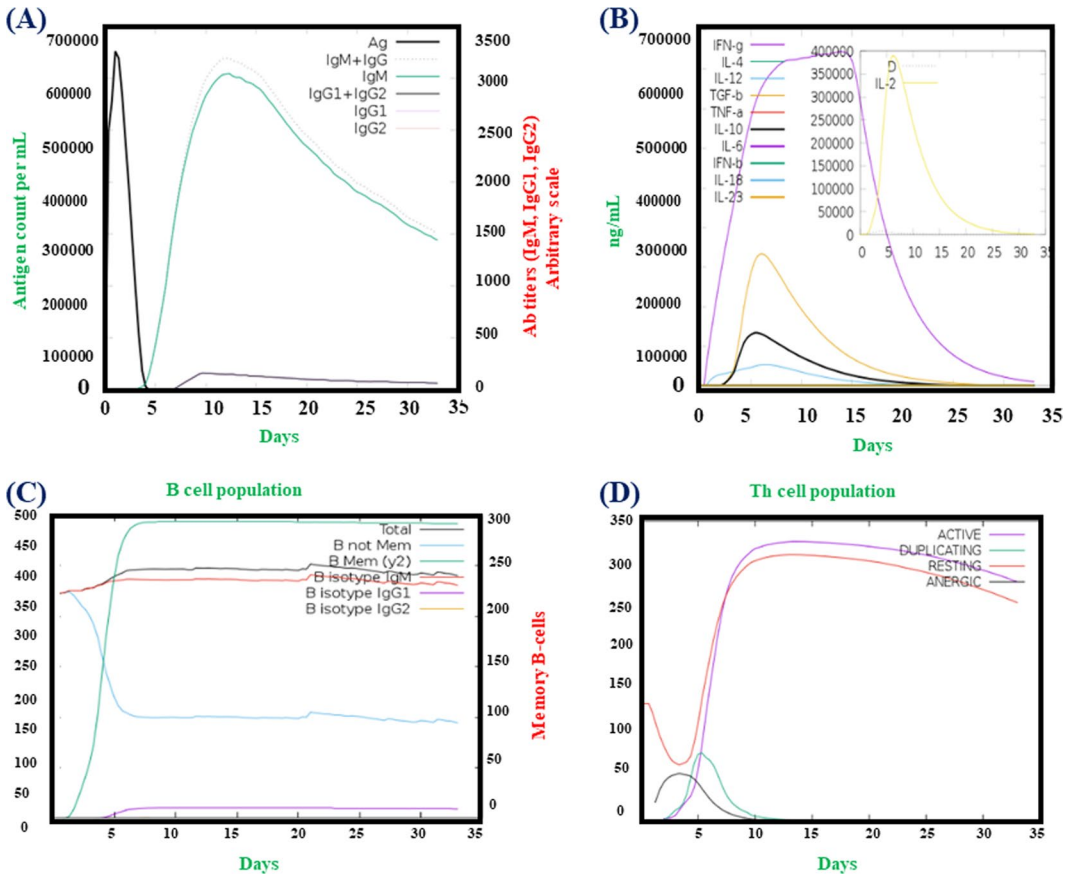

Figure 10. C-ImmSim presentation of an *in silico* immune simulation with the chimeric peptide. (**A**) Antibodies production (the black vertical lines is antigen. (**B**) Cytokines level, (**C**,**D**) shows the B and T-cell population respectively.

Utilizing the cutting-edge immune-informatics approach proteome of the *H. pylori* was exploited to identify the target sequences for creating the subunit vaccine against the *H. pylori*. As previously documented, the most recognizable toxic factors in *H. pylori* infections are cag, PaI, VacA, cagþ, dupA, OipA, dupA and BabA. To assess the immunogenic potency of these proteins comprehensive analysis was carried out using online servers, and online tools were applied to generate a potent subunit vaccine design. Additionally, the B-lymphocyte and T-lymphocyte epitopes were determined from the chosen protein sequences. Generally, the T cell receptors (TCR) on T-lymphocytes are responsible, to generate the immune response, after activated by the antigen-presenting cells (APCs) or MHC bound antigens, and these two molecules MHC-I and MHC-II. The immune-informatics scrutiny of the *H. pylori* proteome determines, that the proposed vaccine protein covers a comprehensive number of high affinities MHC Class I, II and B-cell linear epitope based on physio-chemical property and structural features. And the generation of plasma and memory B cells provides future protection from specific antigens or pathogen-associated antigens. Here our final proposed vaccine design is composed of significant MHC I & II binding epitopes assisted by the adjuvant improves the host immunity. To counter the allergic response to the designed vaccine, the involved key parameters were carefully assessed. PSIPRED v3.3 and RaptorX servers were utilized for the SS and 3D structures accordingly. To further refine the 3D structure GalaxyRefine server was utilized and the refined 3D model was additionally validated utilizing multiple servers for its quality. To assess the interactions among TLR-2, TLR-4, TLR-5, and TLR-9 and the designed vaccine, docking was performed, and for the stability of docking complexes, MD simulations were applied.

For maximized expression and optimization of codons, Jcat software was applied and K12 strain of *E. coli* was used. The GC contents and CAI calculated by Jcat indicated high level expression, the software validation confirmed high solubility of the proposed vaccine with required expression in *E. coli*. Disulfide bonds were used in order to provide stability to vaccine.

Immunoinformatics methods used in this scientific study validated the stability, effectiveness and high-level expression of vaccine protein in *E. coli* host. Furthermore, the use of agent-based model predicted the accurate response upon the injection of the vaccine sequence. The vaccine is specific, non-allergenic, antigenic and can effectively control *H. pylori* infections, further clinical trials are required to check the efficacy of vaccine.

Conclusion

One of the major causes of gastric disorder is *Helicobacter Pylori*. This scientific study used immunoinformatics to design a novel and potent multi-epitope vaccine against *H. pylori* to design novel multi-epitope vaccine, which could provide both types of immunity. Molecular docking, thermodynamics stability profiling, *in-silico* expression and an agent-based modelling tool to verify the stability, expression and immune response reaction provoked by the final vaccine. The result we got increases the experimental validity and can be helpful in designing of an effective vaccine against *H. pylori* infection.

References

1. Floch, P., Mégraud, F. & Lehours, P. Helicobacter pylori strains and gastric MALT lymphoma. *Toxins* **9**, 132 (2017).
2. Peek, R. M. Jr. & Blaser, M. J. Helicobacter pylori and gastrointestinal tract adenocarcinomas. *Nature Reviews Cancer* **2**, 28 (2002).
3. Suerbaum, S. & Michetti, P. Helicobacter pylori infection. *New England Journal of Medicine* **347**, 1175–1186 (2002).
4. Atherton, J. C. & Blaser, M. J. Coadaptation of Helicobacter pylori and humans: ancient history, modern implications. *The Journal of clinical investigation* **119**, 2475–2487 (2009).
5. Backert, S. & Tegtmeyer, N. Type IV secretion and signal transduction of Helicobacter pylori CagA through interactions with host cell receptors. *Toxins* **9**, 115 (2017).
6. Khatoon, J., Prasad, K., Prakash Rai, R., Ghoshal, U. & Krishnani, N. Association of heterogenicity of Helicobacter pylori cag pathogenicity island with peptic ulcer diseases and gastric cancer. *British journal of biomedical science* **74**, 121–126 (2017).
7. Merino, E., Flores-Encarnación, M. & Aguilar-Gutiérrez, G. R. Functional interaction and structural characteristics of unique components of Helicobacter pylori T4 SS. *The FEBS journal* **284**, 3540–3549 (2017).
8. Yuan, X.-y *et al*. Helicobacter pylori with East Asian-type cagPAI genes is more virulent than strains with Western-type in some cagPAI genes. *Brazilian journal of microbiology* **48**, 218–224 (2017).
9. Censini, S. *et al*. cag, a pathogenicity island of Helicobacter pylori, encodes type I-specific and disease-associated virulence factors. *Proceedings of the National Academy of Sciences* **93**, 14648–14653 (1996).
10. Yamakawa, A. *et al*. Correlation between Variation of the 3′ Region of the cagA Gene in Helicobacter pylori and Disease Outcome in Japan. *The Journal of Infectious Diseases* **186**, 1621–1630, https://doi.org/10.1086/345374 (2002).
11. Argent, R. H. *et al*. Determinants and consequences of different levels of CagA phosphorylation for clinical isolates of Helicobacter pylori. *Gastroenterology* **127**, 514–523, https://doi.org/10.1053/j.gastro.2004.06.006 (2004).
12. Yamaoka, Y. *et al*. Relationship between the cagA 3′ repeat region of Helicobacter pylori, gastric histology, and susceptibility to low pH. *Gastroenterology* **117**, 342–349, https://doi.org/10.1053/gast.1999.0029900342 (1999).
13. Yamaoka, Y. *et al*. Importance of Helicobacter pylori oipA in clinical presentation, gastric inflammation, and mucosal interleukin 8 production. *Gastroenterology* **123**, 414–424, https://doi.org/10.1053/gast.2002.34781 (2002).
14. Yamaoka, Y. *et al*. Helicobacter pylori outer membrane proteins and gastroduodenal disease. *Gut* **55**, 775–781, https://doi.org/10.1136/gut.2005.083014 (2006).
15. Yamaoka, Y. *et al*. Helicobacter pylori infection in mice: Role of outer membrane proteins in colonization and inflammation. *Gastroenterology* **123**, 1992–2004, https://doi.org/10.1053/gast.2002.37074 (2002).
16. Yoshida, H. *et al*. The Evaluation of Putative Virulence Factors of Helicobacter pylori for Gastroduodenal Disease b Use of a Short-Term Mongolian Gerbil Infection Model. *The Journal of Infectious Diseases* **185**, 341–347, https://doi.org/10.1086/338772 (2002).
17. Gerhard, M. *et al*. Clinical relevance of the Helicobacter pylori gene for blood-group antigen-binding adhesin. *Proceedings of the National Academy of Sciences* **96**, 12778–12783, https://doi.org/10.1073/pnas.96.22.12778 (1999).
18. Rad, R. *et al*. The Helicobacter pylori Blood Group Antigen-Binding Adhesin Facilitates Bacterial Colonization and Augments a Nonspecific Immune Response. *The Journal of Immunology* **168**, 3033–3041, https://doi.org/10.4049/jimmunol.168.6.3033 (2002).
19. Ki, M.-R. *et al*. Role of vacuolating cytotoxin VacA and cytotoxin-associated antigen CagA of Helicobacter pylori in the progression of gastric cancer. *Molecular and Cellular Biochemistry* **396**, 23–32, https://doi.org/10.1007/s11010-014-2138-8 (2014).
20. Nürnberger, T., Brunner, F., Kemmerling, B. & Piater, L. Innate immunity in plants and animals: striking similarities and obvious differences. *Immunological Reviews* **198**, 249–266, https://doi.org/10.1111/j.0105-2896.2004.0119.x (2004).

21. Netea, M. G., Van der Graaf, C., Van der Meer, J. W. M. & Kullberg, B. J. Toll-like receptors and the host defense against microbial pathogens: bringing specificity to the innate-immune system. *Journal of Leukocyte Biology* **75**, 749–755, https://doi.org/10.1189/jlb.1103543 (2004).
22. Torok, A. M., Bouton, A. H. & Goldberg, J. B. Helicobacter pylori Induces Interleukin-8 Secretion by Toll-Like Receptor 2- and Toll-Like Receptor 5-Dependent and -Independent Pathways. *Infection and Immunity* **73**, 1523–1531, https://doi.org/10.1128/iai.73.3.1523-1531.2005 (2005).
23. Ding, S.-Z., Torok, A. M., Smith, M. F. Jr. & Goldberg, J. B. Toll-like Receptor 2-Mediated Gene Expression in Epithelial Cells During Helicobacter pylori Infection. *Helicobacter* **10**, 193–204, https://doi.org/10.1111/j.1523-5378.2005.00311.x (2005).
24. Peek, R. M. *et al*. Helicobacter pylori Flagellin Evades Toll-Like Receptor 5-Mediated Innate Immunity. *The Journal of Infectious Diseases* **189**, 1914–1920, https://doi.org/10.1086/386289 (2004).
25. Ishihara, S. *et al*. Essential Role of MD-2 in TLR4-Dependent Signaling during Helicobacter pylori-Associated Gastritis. *The Journal of Immunology* **173**, 1406–1416, https://doi.org/10.4049/jimmunol.173.2.1406 (2004).
26. Sutton, P. & Boag, J. M. Status of vaccine research and development for Helicobacter pylori. *Vaccine*, https://doi.org/10.1016/j.vaccine.2018.01.001 (2018).
27. Larsen, J. E. P., Lund, O. & Nielsen, M. Improved method for predicting linear B-cell epitopes. *Immunome research* **2**, 2–2, https://doi.org/10.1186/1745-7580-2-2 (2006).
28. Khan, A. *et al*. Computational identification, characterization and validation of potential antigenic peptide vaccines from hrHPVs E6 proteins using immunoinformatics and computational systems biology approaches. *PloS one* **13**, e0196484–e0196484, https://doi.org/10.1371/journal.pone.0196484 (2018).
29. Ali, A. *et al*. Immunoinformatic and systems biology approaches to predict and validate peptide vaccines against Epstein–Barr virus (EBV). *Scientific Reports* **9**, 720, https://doi.org/10.1038/s41598-018-37070-z (2019).
30. Sakharkar, K. R., Sakharkar, M. K. & Chandra, R. *Post-Genomic Approaches in Drug and Vaccine Development*. Vol. 5 (River Publishers, 2015).
31. Larsen, M. V. *et al*. Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction. *BMC bioinformatics* **8**, 424 (2007).
32. EL-Manzalawy, Y., Dobbs, D. & Honavar, V. Predicting linear B-cell epitopes using string kernels. *Journal of Molecular Recognition: An Interdisciplinary Journal* **21**, 243–255 (2008).
33. Ponomarenko, J. *et al*. ElliPro: a new structure-based tool for the prediction of antibody epitopes. *BMC bioinformatics* **9**, 514 (2008).
34. Livingston, B. *et al*. A rational strategy to design multiepitope immunogens based on multiple Th lymphocyte epitopes. *The Journal of Immunology* **168**, 5499–5506 (2002).
35. Saadi, M., Karkhah, A. & Nouri, H. R. Development of a multi-epitope peptide vaccine inducing robust T cell responses against brucellosis using immunoinformatics based approaches. *Infection, Genetics and Evolution* **51**, 227–234 (2017).
36. Eslami, M., Nezafat, N., Negahdaripour, M. & Ghasemi, Y. Computational approach to suggest a new multi-target-directed ligand as a potential medication for Alzheimer's disease. *Journal of Biomolecular Structure and Dynamics*, 1–15 (2019).
37. Dorosti, H. *et al*. Vaccinomics approach for developing multi-epitope peptide pneumococcal vaccine. *Journal of Biomolecular Structure and Dynamics* **37**, 3524–3535 (2019).
38. Saha, S. & Raghava, G. AlgPred: prediction of allergenic proteins and mapping of IgE epitopes. *Nucleic acids research* **34**, W202–W209 (2006).
39. Cheng, J., Randall, A. Z., Sweredoski, M. J. & Baldi, P. SCRATCH: a protein structure and structural feature prediction server. *Nucleic acids research* **33**, W72–W76 (2005).
40. Gasteiger, E. *et al*. In *The proteomics protocols handbook* 571–607 (Springer, 2005).
41. McGuffin, L. J., Bryson, K. & Jones, D. T. The PSIPRED protein structure prediction server. *Bioinformatics* **16**, 404–405 (2000).
42. Källberg, M. *et al*. Template-based protein structure modeling using the RaptorX web server. *Nature protocols* **7**, 1511 (2012).
43. Ko, J., Park, H., Heo, L. & Seok, C. GalaxyWEB server for protein structure prediction and refinement. *Nucleic acids research* **40**, W294–W297 (2012).
44. Krieger, E. & Vriend, G. YASARA View—molecular graphics for all devices—from smartphones to workstations. *Bioinformatics* **30**, 2981–2982 (2014).
45. Wiederstein, M. & Sippl, M. J. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic acids research* **35**, W407–W410 (2007).
46. Colovos, C. & Yeates, T. O. Verification of protein structures: patterns of nonbonded atomic interactions. *Protein science* **2**, 1511–1519 (1993).
47. Lovell, S. C. *et al*. Structure validation by Cα geometry: ϕ, ψ and Cβ deviation. *Proteins: Structure, Function, and Bioinformatics* **50**, 437–450 (2003).
48. Schneidman-Duhovny, D., Inbar, Y., Nussinov, R. & Wolfson, H. J. PatchDock and SymmDock: servers for rigid and symmetric docking. *Nucleic acids research* **33**, W363–W367 (2005).
49. Andrusier, N., Nussinov, R. & Wolfson, H. J. FireDock: fast interaction refinement in molecular docking. *Proteins: Structure, Function, and Bioinformatics* **69**, 139–159 (2007).
50. Pearlman, D. A. *et al*. AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Computer Physics Communications* **91**, 1–41 (1995).
51. SalomonFerrer, R., Case, D. A. & Walker, R. C. An overview of the Amber biomolecular simulation package. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **3**, 198–210 (2013).
52. Roe, D. R. & Cheatham, T. E. III. PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data. *Journal of chemical theory and computation* **9**, 3084–3095 (2013).
53. Rapin, N., Lund, O. & Castiglione, F. Immune system simulation online. *Bioinformatics* **27**, 2013–2014 (2011).
54. Doytchinova, I. A. & Flower, D. R. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC bioinformatics* **8**, 4 (2007).
55. Wilson, K. T. & Crabtree, J. E. Immunology of Helicobacter pylori: insights into the failure of the immune response and perspectives on vaccine studies. *Gastroenterology* **133**, 288–308, https://doi.org/10.1053/j.gastro.2007.05.008 (2007).
56. Taylor, J. M., Ziman, M. E., Canfield, D. R., Vajdy, M. & Solnick, J. V. Effects of a Th1- versus a Th2-biased immune response in protection against Helicobacter pylori challenge in mice. *Microbial pathogenesis* **44**, 20–27, https://doi.org/10.1016/j.micpath.2007.06.006 (2008).

Acknowledgements

Mazhar Khan is financially supported by CAS-TWAS presidential scholarship. We acknowledge CAS-TWAS for financial support. Dong-Qing Wei is supported by the grants from the Key Research Area Grant 2016YFA0501703 of the Ministry of Science and Technology of China, the National Natural Science Foundation of China (Contract no. 61832019, 61503244), the Natural Science Foundation of Henan Province (162300410060) and Joint Research Funds for Medical and Engineering and Scientific Research at Shanghai Jiao Tong University (YG2017ZD14). The computations were partially performed at the Center for High Performance Computing, Shanghai Jiao Tong University.

Author Contributions

M.K., S.K., A.K. and A.A. conceptualized the methodology. M.K., H.A., A.M.S. and S.K. did the analysis. M.K., A.K. and D.Q.W. wrote the manuscript. D.Q.W. supervised the study. All the authors approved the manuscript.

Additional Information

Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-019-49354-z.

Competing Interests: The authors declare no competing interests.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2019

Supplementary Materials

Immunoinformatics approaches to explore *Helicobacter Pylori* proteome (Virulence Factors) to design B and T cell multi-epitope subunit vaccine

Mazhar Khan[1], Shahzeb Khan[2], Asim Ali[1], Hameed Akbar[4], Abrar Muhammad Sayaf[2], Abbas Khan[2, 3], Dong-Qing Wei[3*]

[1]The CAS Key Laboratory of Innate Immunity and Chronic Diseases, Hefei National Laboratory for Physical Sciences at Microscale, School of Life Sciences, CAS Center for Excellence in Molecular Cell Science, University of Science and Technology of China (USTC), Collaborative Innovation Center of Genetics and Development, Hefei, 230027, Anhui, China.

[2]Centre for Biotechnology and Microbiology, University of Swat, Swat, Khyber Pakhtunkhwa, Pakistan.

[3]Department of Bioinformatics and Biological Statistics, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai 200240, P.R China.

[4]Laboratory of Cellular Dynamics, School of Life Sciences, University of Science and Technology of China (USTC).

Mazhar Khan
mazharsw@mail.ustc.edu.cn
Shahzeb Khan
khanshahxeb173@gmail.com
Asim Ali
Asim21pk@gmail.com
Hameed Akbar,
Hameedakbar4181@mail.ustc.edu.cn
Abrar Muhammad Sayaf
Chemdr07@gmail.com
Abbas Khan
abbaskhan@sjtu.edu.cn
Dong-Qing Wei (corresponding author)
dqwei@sjtu.edu.cn

Table S1: Cytotoxic T-Lymphocytes epitopes identified.

Protein	Residue NO	Peptide sequence	MHC Binding affinity	Rescale binding affinity	C-terminal cleavage affinity	Transport affinity	Prediction score	MHC-I Binding
Gi:446963037, chaperonin GroEL	7	FSDSARNLL	0.3712	1.5762	0.6218	0.8250	1.7107	yes
	88	TTTATVLAY	0.7203	3.0584	0.9620	2.8160	3.3435	yes
	152	NSDHNIGKL	0.2194	0.9317	0.8385	1.0100	1.1080	yes
	210	MTAQLDNAY	0.7529	3.1966	0.6508	3.0680	3.4476	yes
	213	QLDNAYILL	0.2023	0.8589	0.9539	0.9650	1.0502	yes
	221	LTDKKISSM	0.2744	1.1652	0.7842	0.1590	1.2908	yes
	300	ISEELGLSL	0.1333	0.5658	0.9464	0.9210	0.7538	yes
	351	QIASTTSDY	0.3531	1.4990	0.9512	3.0930	1.7963	yes
	513	TTEATVHEI	0.2253	0.9566	0.8508	0.3960	1.1041	yes
Gi:446632395, outer inflammatory protein OipA	15	LHAERNGFY	0.1278	0.5427	0.4242	2.8970	0.7512	yes
	80	DSNKIASRF	0.1159	0.4922	0.9662	2.2660	0.7504	yes
	147	NTDLLINWT	0.2062	0.8756	0.0314	-0.9710	0.8318	yes
	168	RVKGLSIFY	0.1380	0.5858	0.8687	3.4370	0.8880	yes
	183	TLDANTLKK	0.1654	0.7023	0.9658	0.4190	0.8681	yes
	273	KVNYYSDDY	0.2689	1.1415	0.8007	2.9710	1.4102	yes
	277	YSDDYGDKL	0.4390	1.8640	0.8498	0.7470	2.0288	yes
Gi:2498230, Cytotoxicity associated immune-dominant antigen cagA	36	KVDNVVASF	0.1380	0.5860	0.9768	2.6130	0.8632	yes
	65	GISQLREEY	0.1879	0.7979	0.9289	2.7310	1.0738	yes
	91	FIDKSNDLI	0.2135	0.9063	0.1925	0.2530	0.9479	yes
	125	FTSWVSHQK	0.1566	0.6647	0.9669	0.3530	0.8274	yes
	182	RTDQKFMGV	0.2884	1.2244	0.8152	0.3700	1.3652	yes
	212	WLDIFLSFI	0.2194	0.9314	0.6385	0.2120	1.0378	yes
	249	TTDIQGLPP	0.2120	0.9002	0.0692	-0.1040	0.9054	yes
	284	GVADIDPNY	0.1639	0.6960	0.9685	2.7340	0.9780	yes
	316	EPEKVSLLY	0.1555	0.6604	0.9779	2.2810	0.9211	yes
	418	LSEKEKEKF	0.1442	0.6125	0.2986	2.5760	0.7860	yes
	463	ITEFNNGDL	0.1547	0.6567	0.9209	0.8600	0.8378	yes
	505	FVDYSNFKY	0.6991	2.9683	0.9638	2.8280	3.2542	yes
	669	NRDARAIAY	0.2057	0.8733	0.9332	3.1460	1.1706	yes
	769	KSDLENSVK	0.1625	0.6898	0.2787	0.5900	0.7611	Yes
	785	VTDKVDNLN	0.2400	1.0190	0.0529	-1.3500	0.9594	Yes
	834	NTGKNSELY	0.6100	2.5898	0.9445	2.6720	2.8651	Yes
	894	STEPIYAKV	0.1867	0.7926	0.9713	0.3430	0.9554	Yes
	910	VASPEEPIY	0.1391	0.5905	0.5934	2.9860	0.8288	Yes
	964	LSASPEPIY	0.2576	1.0939	0.9139	2.9560	1.3788	Yes
	974	TIDDLGGPF	0.1615	0.6859	0.5831	2.6200	0.9043	Yes
	1015	VSEAKAGFF	0.3085	1.3100	0.0717	2.6470	1.4531	Yes

	1037	TKKNVMNLY	0.1299	0.5517	0.9505	3.0050	0.8445	Yes
	1055	SLSAKLDNY	0.2276	0.9665	0.7782	3.1470	1.2406	Yes
	1129	YSDSFKFST	0.2677	1.1366	0.0424	-1.0000	1.0930	Yes
	1150	FTHFLANAF	0.1713	0.7272	0.2691	2.4180	0.8885	yes
	1154	LANAFSTGY	0.3152	1.3381	0.9606	2.7820	1.6213	Yes
Gi:15645505, vacuolating cytotoxin autotransporter vacA	135	DMKDAVGTY	0.1680	0.7134	0.9777	2.7750	0.9988	Yes
	231	SKNAEISLY	0.1459	0.6196	0.8845	2.9870	0.9016	Yes
	273	YSTINTSKV	0.1589	0.6745	0.7829	0.3530	0.8096	Yes
	277	NTSKVTGEV	0.2254	0.9571	0.9523	0.3350	1.1167	Yes
	372	KTEIQPTQV	0.1484	0.6303	0.8454	0.2070	0.7674	Yes
	415	LTTNAAHLH	0.1956	0.8307	0.0678	-0.5350	0.8141	Yes
	416	TTNAAHLHI	0.2052	0.8712	0.6959	0.4460	0.9979	Yes
	456	RVNNQVGGY	0.2121	0.9005	0.9688	3.2610	1.2088	Yes
	478	GTDTKNGTA	0.3002	1.2744	0.5297	-0.7580	1.3160	Yes
	484	GTATFNNDI	0.1992	0.8456	0.5317	0.4920	0.9499	Yes
	501	KVDAHTANF	0.1437	0.6103	0.8337	2.6520	0.8680	Yes
	511	GIDTGNGGF	0.1393	0.5913	0.9123	2.4090	0.8486	Yes
	659	YSQFSNLTI	0.1945	0.8258	0.8400	0.3840	0.9710	Yes
	703	NVDSATGFY	0.7135	3.0293	0.8036	3.0650	3.3031	Yes
	731	LLKAKIIGY	0.1093	0.4640	0.9608	2.9930	0.7578	Yes
	793	SMVNNPDNY	0.1581	0.6711	0.4466	3.1280	0.8944	Yes
	893	RSKDIDTLY	0.3762	1.5971	0.9760	3.3480	1.9109	Yes
	916	LIDSHDAGY	0.4998	2.1219	0.8986	2.9290	2.4031	Yes
	958	TSSLQTLSL	0.1453	0.6169	0.9545	1.0710	0.8136	Yes
	986	NIDSFAQRL	0.1341	0.5693	0.9640	1.0850	0.7682	Yes
	1045	GTSAGVDAY	0.5717	2.4275	0.8444	2.6680	2.6876	Yes
	1111	GSDQSSLNF	0.4972	2.1112	0.9169	2.2680	2.3621	Yes
	1123	LLRDLNQSY	0.1458	0.6189	0.9582	3.1510	0.9202	Yes
	1128	NQSYNYLAY	0.3870	1.6433	0.8923	3.0810	1.9312	Yes
	1136	YSAATRASY	0.5348	2.2705	0.5821	3.1160	2.5136	Yes
	1138	AATRASYGY	0.1250	0.5308	0.9481	3.1390	0.8299	Yes
	1194	ASANVEARY	0.4364	1.8530	0.9373	3.2670	2.1570	Yes
	1195	SANVEARYY	0.2018	0.8567	0.5286	3.0900	1.0905	Yes
	1203	YYGDTSYFY	0.1700	0.7216	0.9619	2.9540	1.0136	Yes
	1280	FASNLGMRY	0.5205	2.2099	0.9725	2.8550	2.4986	Yes