Determination of B- and T- cell epitopes for Helicobacter pylori *cagPAI: An in silico approach*

Sevgi Kalkanlı Taş[1], Duygu Kırkık[2], Kübra Öztürk[2], Alpaslan Tanoğlu[3]

[1]Department of Immunology, Health Sciences University School of Medicine, İstanbul, Turkey
[2]Department of Medical Biology, Health Sciences University School of Medicine, İstanbul, Turkey
[3]Department of Gastroenterology, Sultan Abdulhamit Training and Research Hospital, İstanbul, Turkey

Cite this article as: *Kalkanlı Taş S, Kırkık D, Öztürk K, Tanoğlu A. Determination of B- and T- cell epitopes for* Helicobacter pylori *cagPAI: An in silico approach. Turk J Gastroenterol 2020; 31(10): 713-20.*

ABSTRACT

Background/Aims: Helicobacter pylori *is classified as a gram-negative bacteria and can cause significant diseases, including gastric cancer, mucosa-associated lymphoid tumor, peptic ulcer, and chronic gastritis. Recent studies have shown that some autoimmune diseases are also associated with* H. pylori*. In the past decades, polymorphisms of certain genes of* H. pylori*, mechanisms and strains of* H. pylori*, and new therapeutic approaches have continued to be defined. Bioinformatic tools continue to be used in drug design and vaccine design. This study aimed to investigate the cag pathogenicity island (cagPAI) of* H. pylori *using an in silico approach, which could contribute to vaccine studies.*

Materials and Methods: *The pathogenicity island of H. pylori was obtained from GenBank and analyzed with ClustalW software. Structures of cag Virb11 (Hp0525) and an inhibitory protein (Hp1451) were obtained, and codon optimization and secondary and tertiary structure prediction for the cagPAI of* H. pylori *were analyzed using Garnier-Osguthorpe-Rabson IV secondary structure prediction method and self-optimized prediction method with alignment software. The BcePred prediction server was used to distinguish linear B-cell epitopes, and prediction of T-cell was obtained with NetCTL and MHCPred.*

Results: *According to the physicochemical parameters, the cagPAI of* H. pylori *was analyzed and found to be stable, and 2 B-cell epitopes of cagPAI of* H. pylori *and 2 T-cell epitopes of cagPAI were found in this study.*

Conclusion: *B- and T-cell epitopes that we have identified can induce both humoral and cellular immune responses. Thus, these epitopes have a potential for vaccine studies. Consequently, this in silico analysis should be combined with other pieces of evidence, including experimental data, to assign function.*

Keywords: *CagA protein,* Helicobacter pylori*, computational biology*

INTRODUCTION

Helicobacter pylori is a commonly occurring bacteria that infects numerous people worldwide (1). The prevalence of *H. pylori* infection varies from country to country; however, it has been reported to be 82.5% in Turkey (2). Recent studies showed a relationship between *H. pylori* and many significant diseases, including chronic gastritis, peptic ulcer, mucosa-associated lymphoid tumor/gastric lymphoma, idiopathic thrombocytopenic purpura, functional dyspepsia, unexplained iron deficiency anemia, as well as stomach cancer (3). *H. pylori* is classified as a gram-negative bacteria that has important enzymes for its metabolism and survival, including urease, catalase, and oxidase. In addition, *H. pylori* includes more than 35 strains with different genome sequences (4). To be able to accommodate host colonization, a single strain of *H. pylori* creates many different variants and causes widespread infections. The polymorphic variability of *H. pylori* leads to a genetic predisposition factor for infection (5).

Toll-like receptors (TLRs) are known to be transmembrane proteins that serve as sensors for microbial factors. The role of TLRs in the immune pathogenesis of *H. pylori* infection has been shown in many studies (6), and particularly, TLR4-positive cells were found in the gastric epithelia and monocytes/macrophages of superficial gastritis, atrophy, dysplasia, or carcinoma. Furthermore, there are studies that show a high correlation between TLR4 expression and the severity of inflammation. It is well known that many bacterial components have an important role in disease pathogenesis. For *H. pylori*, there is a pathogenicity island in type 1 strains. This island has various virulence components, such as cytotoxin-associated gene A (cagA) and cytotoxin-associated gene E (cagE) (7). These virulence factors are associated with more se-

Corresponding Author: ***Duygu Kırkık; dygkirkik@gmail.com***
Received: ***March 19, 2019*** *Accepted:* ***November 9, 2019***
© Copyright 2020 by The Turkish Society of Gastroenterology · Available online at turkjgastroenterol.org
DOI: ***10.5152/tjg.2020.19154***

vere gastroduodenal diseases (8). In the previous decade, researchers demonstrated that purified lipopolysaccharide (LPS) extracted from a type 1 *H. pylori* strain from the gastric parietal cells increases the transcription and translation of TLR4 in guinea pigs (9). *H. pylori* infection is known to be an inducer of TLR4 expression in several types of gastrointestinal epithelial cells.

Moreover, vaccine studies have a very important role in both prophylaxis and treatment. These studies showed that more powerful innate and acquired immune responses might be generated using different epitopes (10). In many previous studies, epitope prediction for different pathogens was performed using an *in silico* approach (11). Therefore, the characteristic structure of the cag pathogenicity island (cagPAI) of *H. pylori* was analyzed using bioinformatics (Figure 1) (12). In the literature, variants of the TLR4 gene, with polymorphic changes, have been associated with *H. pylori* infection. In this study, we aimed to investigate the cagPAI of *H. pylori* using an *in silico* approach and bioinformatic methods, which may contribute to a better understanding of interactions between *H. pylori* and its host cells. Our study may also contribute to vaccine studies upon experimental confirmation of potential B-and T-cell epitopes.

MATERIALS AND METHODS

Ethics Committee approval and informed consent

The term *in silico* indicates the use of individualized computer simulation in the development or regulatory evaluation of a medicinal product, medical device, or medical intervention. This study aimed to investigate the cagPAI of *H. pylori* using an *in silico* approach; therefore, no ethical approval or informed consent was required for our study.

Sequence analysis

The database GenBank was used to obtain sequences of antigens, and the software ClustalW was used to perform multiple sequence alignments (https://www.ebi.ac.uk/Tools/msa/clustalw2/). Gene designer v.1.1 program and Prot-Param (https://web.expasy.org/protparam/) are 2 different standalone software programs that were used to analyze the cagPAI of *H. pylori*. With the help of these, multiparameter characterizations were performed, and basic physicochemical parameters of the cagPAI of *H. pylori* were analyzed.

MAIN POINTS

- *The cagPAI of* H. pylori *contains immune dominant epitopes from different antigenic proteins including ureB, vacA, cagI and cagE.*
- *We have found 2 strong candidates for both B-cell epitopes and T-cell epitopes which can induce both humoral and cellular immune responses.*
- *Consequently,* in vivo *studies would support our findings and discover all epitopes of immune system. Therefore, they have the potential to develop recombinant vaccines.*

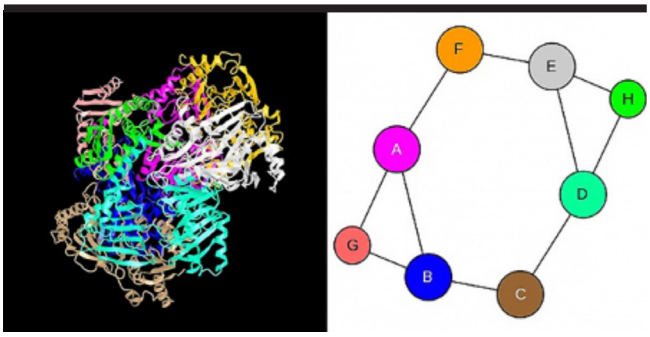

Figure 1. Structure of Cag Virb11 (Hp0525) and an Inhibitory Protein (Hp1451).

Codon optimization

The Kazusa codon database was used for optimization. The information required for the *in silico* analysis of genes and optimization of the cagPAI was provided by the reverse translation online tool (https://www.kazusa.or.jp/codon/), and optimization of the cagPAI and *in silico* gene analysis was performed using the reverse translation tool (www. bioinformatics. org/ sms2/rev_trans.html). JavaScript programs were then used to analyze and format the protein and DNA sequences (www.ualberta.ca/~stothard/software.html) (13). The cagPAI was designed for cloning and expression in *Escherichia coli* BL21, and immunogenicity of the full antigen and its subunit vaccine were determined with the VaxiJen server. Essential restriction enzyme sites (*MseI*, *Eco*RI) were present at the ends of the sequence, and other restriction sites were removed from the synthetic construct for the subsequent purpose of cloning. TAA, an effective stop codon, was used to increase the efficiency of translational termination in the prokaryotic host.

Secondary and tertiary structure prediction for cagPAI of H. pylori

The software Garnier-Osguthorpe-Rabson IV secondary structure prediction method (GOR-IV) and self-optimized prediction method with alignment (SOPMA) are the online tools (https://npsa-prabi.ibcp.fr) that were used to predict the secondary structure of the cagPAI of *H. pylori*. The server Predict Protein was then used to an-

alyze the sequence and predict the protein structure and function. These software programs provided additional information, such as the percentage of random coils, alpha-helices, beta-sheets, and solvent accessibility. The stability of the 3- dimensional (3D) structure of the cagPAI of *H. pylori* was determined using PubMed.

Prediction of B-cell epitopes and mRNA secondary structure

Antibodies of the immune system use B-cell epitopes of molecules for recognition. Therefore, it is important to gain knowledge about B-cell epitopes to design an effective subunit vaccine. Bcepred software was used to obtain continuous (linear) B-cell epitopes (http://crdd.osdd.net/raghava/bcepred/). ABCpred server was then chosen to analyze the linear B-cell epitope regions (http://crdd.osdd.net/raghava/abcpred/). The structure of folded mRNA of the cagPAI was determined and was analyzed for optimization, and the energetic stability of the predicted structures was determined (http://rna.tbi.univie.ac.at//cgi-bin/RNAWebSuite/RNAfold.cgi).

Prediction of T-cell epitopes

NetCTL server was used to detect common epitopes, which bind to both major-histocompatibility-complex (MHC) class molecules as well as to count the total number of interacting MHC alleles (www.cbs.dtu.dk/services/NetCTL/). This server is a predictor of T-cell epitopes along a protein sequence. This tool at the immune epitope database predicts the half-maximal (50%) inhibitory concentration (IC50) values for the peptides binding to specific MHC molecules. MHCPred (http://www.ddg-pharmfac.net/mhcpred/MHCPred/) was used to predict epitope candidates on the basis of the processing of peptides in the cell.

RESULTS

GOR-IV is the most common and accurate method of secondary structure prediction. If the sequence resemblance to a protein of recognized 3D structure is undetectable, the secondary structure can be used for prediction. GOR-IV examines the sequence to identify the alpha-helix; beta-sheet, turn, or random coil; ambiguous states; pi helix; beta bridge; and secondary structure at each position depending on a 17-amino-acid sequence window. Alpha-helices are one of the important secondary structures in a protein. The alpha-helix has a right-handed helix conformation that is stabilized by hydrogen bonds between the carbonyl (CO) groups and amino (NH) groups of the 4th amino acid from the C-terminus (14). The result of our experiment for the cagPAI random coil was 43.96%. This means that many regions

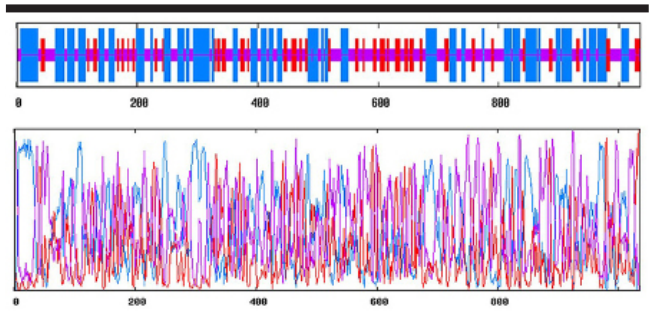

Figure 2. The GOR IV secondary structure prediction method was used to find alpha helix, beta sheet, turn or random coil, ambiguous states, pi helix, beta bridge secondary structure at each position. Alpha helix (Hh) : 408 is 39.42%, 310 helix (Gg) : 0 is 0.00%, Pi helix (Ii) : 0 is 0.00%, Beta bridge (Bb) : 0 is 0.00%, Extended strand (Ee) : 172 is 16.62%, Beta turn (Tt) : 0 is 0.00%, Bend region (Ss) : 0 is 0.00%, Random coil (Cc) : 455 is 43.96%, Ambiguous states (?) : 0 is 0.00%. Other states: 0 is 0.00%.

of the protein chain are not defined by secondary structure. They are not described by any systematic hydrogen bonding pattern and so are identified as random coils. Random coils are discovered in 2 locations in proteins, at the termini or in loops (14, 15).

Our results showed that the 1,035 residues were made up of 172 (16.62%) strands, 408 (39.42%) alpha-helices, and 455 (43.96%) random coils, as shown in Figure 2a. SOPMA is a program that is used to find secondary structures of proteins. According to SOPMA, the 1,035 length of sequences was made up of 160 (15.46%) strands, 471 (45.51%) alpha-helices, 343 (33.14%) random coils, and 61 (5.89%) beta-turns as shown in Figure 2b. Our parameters were analyzed with a window of 17 amino acids, a similarity threshold of 8, and 4 states. A signal peptide was not identified in the initial region of protein sequencing, and these results were double checked with the SignalP 4.1 server as shown in Figure 3. There was no significant difference between the 2 programs.

ABCpred server was used to predict B-cell epitopes (Table 1), and the database gave a tabular list. Thus, the predicted B-cell epitopes were ranked according to the score obtained by a trained recurrent neural network. The higher the score, the higher is the probability of the peptide being an epitope. Table 2 shows the physicochemical properties of proteins determined by BcePred prediction server. There are some significant factors that determine the potentiality of B-cell epitopes. These can be listed as hydrophilicity, secondary structure, flexibility, surface accessibility, and antigenicity. In addition, a potential

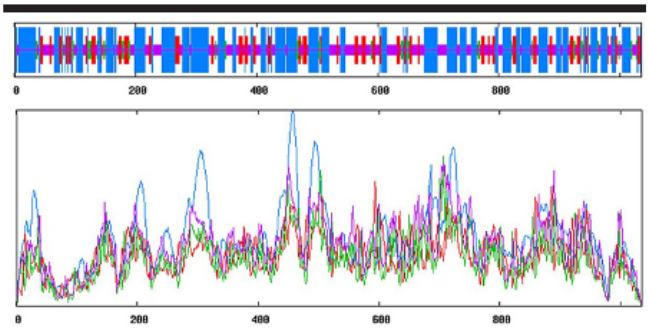

Figure 3. The results were found by SOPMA. Alpha helix(Hh) :471 is 45.51%,310 helix (Gg):0 is 0.00%, Pi helix(Ii) :0 is 0.00%, Beta bridge (Bb):0 is 0.00 %, Extended strand (Ee):160 is 15.46%, turn (Tt) : 61 is 5.89%, Bend region (Ss) : 0 is 0.00%,Random coil (Cc):343 is 33.14%,Ambiguous states (?): 0 is 0.00%,Other states: 0 is 0.00%.

B-cell epitope must have a beta-turn motif in the amino acid residues. The BCEpred server was used to predict the physicochemical properties of our protein (Table 2). VaxiJen predicted protective antigens and calculated the whole antigenicity of the chimeric protein with a score of 1.1899, which was 0.4 times higher than threshold.

Parameters, such as theoretical isoelectric point, extinction coefficients, atomic composition, instability index, aliphatic index, grand average of hydropathicity (GRAVY), and estimated half-life, were computed using the tool Prot-Param. According to the results, the total number of negatively charged residues (Asp+Glu) was 123, and the total number of positively charged residues (Arg+Lys) was 121. Extinction coefficients were in units of M^{-1} cm^{-1}, at 280 nm measured in water. The instability index was computed to be 34.97, which classified the protein as stable. Aliphatic index was analyzed as 86.88, and GRAVY was −0.186. The N-terminus of the sequence considered was M (Met), and the estimated half-life was 30 hours (mammalian reticulocytes, *in vitro*), >20 hours (yeast, *in vivo*), and >10 hours (*Escherichia coli, in vivo*).

The structure of folded mRNA of the cagPAI was determined as shown in Figure 4a and analyzed for optimization and energetic stability of the predicted structures. The result of the thermodynamic ensemble prediction was −448.40 kcal/mol. The frequency of the Minimum Free Energy (MFE) structure in the ensemble was 0.00%. The ensemble diversity was 452.06. Figure 4b shows a dot plot containing base pair probabilities. The centroid secondary structure in dot-bracket notation with a minimum free energy of −181.80 kcal/mol is shown in Figure 4c. A mountain plot representation of the MFE structure, the thermody-

Table 1. Predicted linear B-cell epitopes using ABCPred software. Higher is the score of the peptide, higher is the probability of it to be a useful epitope.

Rank	Sequence	Start position	Score
1	SATKTTKTHDDKTNDS	462	0.92
1	GGRVMKAKTNKARGKS	241	0.91
2	TERPYLRIHPHASETA	554	0.91
2	HSSVMAKVADMAGSVS	288	0.91
3	NATTNHDKSSDNHDNT	609	0.90
3	EQENCEENDMNNDKTS	585	0.90
4	GKSTDWSTKYVKMASD	254	0.90
4	TSEQENCEENDMKAKA	91	0.89
4	SESEQENCESTARTSE	78	0.89
4	THDGDKSTTDSGDNTS	627	0.89
4	SESEQENCESTARTSE	570	0.89
4	HELICACTERPYLRIH	547	0.89
4	VRKGDSSSAKRHGVNK	393	0.89
4	YHDVTRAYMYWAGVKK	162	0.89
4	AGDKMGNDRRKKYRNK	145	0.89
5	GTSADGDWAGDKMGND	137	0.89
5	NATANHDKSSDNHDNT	439	0.88
5	KRHGVNKDVMRWNSDT	402	0.88
6	AAGWSTGRNYGKTRNY	194	0.88
6	ANNSDSTYRVRKGDSS	384	0.87
7	SKKGSTGAGSSARANN	371	0.87
7	KRTKSDTNYDNNYTMD	659	0.86
7	TECAGPATHGENICIT	518	0.86
7	SGLYCSYLASEESCHE	48	0.86
7	ASRGVSTKDWNKRGSK	357	0.86
7	VRSGDTSSASRGVSTK	349	0.86
7	VASTTAANMMRNKMDG	216	0.86
7	TRNYDARRDVVASTTA	206	0.86
8	GAAKTSARWMDDGTSA	125	0.86
8	TDSGDNTSSSNGSADK	635	0.85
8	DKTNDSNDKGSKKNTV	472	0.85
9	GRNYGKTRNYDARRDV	200	0.85
9	THGENICITYISLAND	525	0.84
9	GSTGASGYVMVKKHAD	312	0.84
9	NMMVVSADHATSGANA	179	0.84
10	RMLINEARPARENTAK	17	0.84
10	MDSKTHSDSSNATANH	429	0.84
11	TANGDKTVKNNNMDSK	417	0.83
12	NHDNTTKTHDGDKSTT	620	0.83
12	KKARKVDDKRTKSDTN	651	0.82
12	SSDNHDNTSSDNHDNS	447	0.81
12	DVMRWNSDTANGDKTV	409	0.81
12	AASVVGCSTGNVHASS	106	0.81
13	SSSNGSADKKKARKVD	642	0.80
13	TSKKADSDSSNATTNH	599	0.80
13	ASEESCHERICHIACL	56	0.80
13	APLINEARFRAGMENT	488	0.80

ABCPred: Artificial neural network-based B-cell epitope prediction server

Table 2. Physicochemical properties of linear B-cell epitopes identified by BcePred prediction server. Thresholds: 1.9 (hydrophilicity), 2 (flexibility), 1.9 (accessibility), 2.4 (turns), 2.3 (exposed surface) 1.8 (polarity), 1.9 (antigenic propensity).

Parameters of prediction	Predicted B-cell epitopes
Hydrophilicity	RPARENTAKN, INTRANSG, ASEESCHER, ASETASESEQENCESTARTSEQENCEENDMKAK, GGAAKTS, MDDGTSADGDWAGDKMGNDRRKKYRNKSYHD, SADHATSGA, STGRNYGKTRNYDARRD, RNKMDGD, KAKTNKARGKST, KMASDKNSKRYGVRTTDSRAAR, GSVSKKTNAG, KGSTGASG, KKHADRSASGA, STVADNTNSRV, TVRSGDTSSASRGVSTKD, NKRGSKKGSTGAGSSARANNSDSTYRVRKGDSSSAKRHG, NSDTANGDKTVKNNNMDSKTHSDSSNATANHDKSSDNHDNTSSDNHDNSATKTTKTHDDKTNDSNDKGSKKNT, RPARENTECAGPATHGEN, ANDPRTE, HASETASESEQENCESTARTSEQENCEENDMNNDKTSKKADSDSSNATTNH DKSSDNHDNTTKTHDGDKSTTDSGDNTSSSNGSADKKKARKVDDKRTKSDTNYDNNYTMDDDTSK
Flexibility	RPARENT, REINTRAN, LIASETASESEQ, NCESTARTSEQ, AGDKMGNDRRKKYRN, YWAGVKK, NMMRNKM, AAYNSGGRVMKAKTNKARGKS, VKMASDKNSKRYGVRTTD, DMAGSVSKKTNAGVKGSTGA, KHADRSA, TVADNTNS, YTVRSGDTSSASRGVS,KDWNKRGSKKGSTGAGSSARANNSDSTYRVRKGDSSSAK, NGDKTVKNNNMDSKTHSDSSNATANHDKSSDNHDNTSSDNHDNSATKTTKTHDDKTNDSNDKGSKK, SLANDPR, PHASETASESEQ, NCESTARTSEQ, NDMNNDKTSKKADSDSSNATTNHDKSS DNHDNTTKTHDGDKSTTDSGDNTSSSNGSADKKKARKVDDKRTKSDTMDDDTS
Accessibility	INEARFRA, NEARPARENTAKNLYT, MREINTRANSG, EESCHER, SETASESEQENCESTARTSEQENCEENDMKAKA, GDKMGNDRRKKYRNKSYHDVTR, AGVKKRNMMV, WSTGRNYGKTRNYDARRDV, ANMMRNKMDGDW, RVMKAKTNKARGKSTDWSTKYVKMASDKNSKRYGVRTTDSRAAR, GSVSKKTNAGVK, VKKHADRSA, ADNTNSRVYT, RSGDTSSASRGVSTKDWNKRGSKKGSTG, SSARANNSDSTYRVRKGDSSSAKRHGVNKDVMRWNSDTANGDKTVKNNNMDSKTHSDSSNA TANHDKSSDNHDNTSSDNHDNSATKTTKTHDDKTNDSNDKGSKKNTVAP, INEARFRA, NEARPARENTEC, SLANDPRTEINHE, TERPYLRIH, SETASESEQENCESTARTSEQENCEENDMNNDKTSKKADSDSSNATTNHDKSSDN HDNTTKTHDGDKSTTDSGDNTSSSNGSADKKKARKVDDKRTKSDTNYDNNYTMDDDTSK
Turns	VADNTNSRV, ARANNSDSTY, TVKNNNMDSKTHSDSSNATANHDKSSDNHDNTSSDNHDNSAT, DKTNDSNDKG, ENDMNNDKT, ADSDSSNATTNHDKSSDNHDNTTK, DSGDNTSSSNGSA, SDTNYDNNYTM
Exposed surface	RENTAKN, RTSEQEN, ENDMKAK, KMGNDRRKKYRNKSYHD, VKKRNMMV, RNYGKTRNYDARRD, MKAKTNKARGKS, KMASDKNSKRYG, VKKHADRS, STKDWNKRGSKKG, STYRVRK, KRHGVNK, DKTVKNNN, KTTKTHDDKTNDSNDKGSKKNT, RTSEQEN, NDMNNDKTSKKADSDS, KTHDGDK, GSADKKKARKVDDKRTKSDTNYD, DDDTSK
Polarity	INEARFRAGME, NEARPARENTAKN, MREINTRA, ASEESCHERICHIA, ETASESEQENCES, ARTSEQENCEENDMKAK, DKMGNDRRKKYRNKSYHD, WAGVKKRNMMV, RNYGKTRNYDARRDVV, RNKMDGD, GRVMKAKTNKARGKST, SDKNSKRYGVRT, SRAARVHS, YVMVKKHADRSA, TKDWNKRGSKKGS, STYRVRKGDSSSAKRHGVNKD, DSKTHSD, NHDKSSDNHD, TKTTKTHDDKTNDSNDKGSKKNT, INEARFRAGME, NEARPARENTECA, PRTEINHELI, TERPYLRIHPHASETASESEQENCES, ARTSEQENCEEND, NDKTSKKADSDS, NHDKSSDNHD, TKTHDGDKST, GSADKKKARKVDDKRTKSDT, MDDDTSK
Antigenic propensity	SGLYCSYL, SCHERICHI, SVVGCSTG, YVMVKKH, NICITYISL, INHELIC, PYLRIHPH

BcePred: Prediction of linear B-cell epitopes, using physico-chemical properties

namic ensemble of RNA structures, and the centroid structure is presented. Figure 5 shows the positional entropy for each position. Results were computed using RNAfold 2.4.11. An equivalent command line call would have been RNAfold -p -d2 --noLP<reversetrans.fa>reversetrans.out.

DISCUSSION

H. pylori is a commonly seen pathogen and is known as a type of class I human carcinogen (16). An unusual link has been established between *H. pylori* and some commonly seen diseases, including gastritis, peptic ulcers, and other inflammatory diseases. There are many research studies on *H. pylori* aimed at finding the most suitable and successful eradication treatment, and many patients with this infection receive antibiotic therapies. The eradication therapies are not satisfactorily successful in all patients. The virulence factors of this bacteria play a pivotal role in the induction of mucosal inflammation and develop-

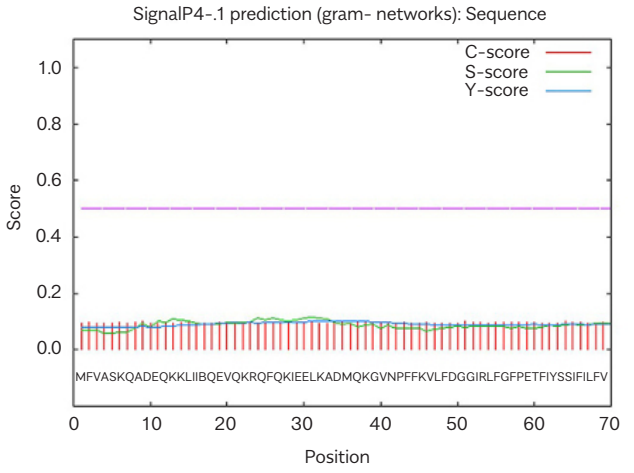

Figure 4. Signal peptide was not found by Signal1P 4.1 Server.

ment of gastric cancer. The best characterized bacterial virulence factors include cagPAI, cagA, and vacA. In the past years, studies have shown that many cagA-positive strains, most of which are able to produce cagA protein, retain all the other genes of the cagPAI, whereas most cagA-negative strains completely lack the cagPAI. This indicates that the existence of the cagA gene can be used as a marker of increased virulence linked to cagPAI. Therefore, Censini et al. (17) examined partial deletion of the cagPAI, suggesting that the positions of the cagA gene and other cagPAI genes should be separately explored.

B-cell epitopes are significant in the improvement of peptide vaccines. For the recognition of diseases, their prediction in the antigen sequence is a vital and complicated problem. Although most antigenic determinants of proteins are discontinuous, there is a potential to mimic the epitopes using synthetic peptides. This is because B-cell epitopes present on a microbial antigen can cross-react, causing molecular mimicry of an epitope on a human autoantigen. As the microbial antigen and the self-antigen are only partly analogous, there will be no T-cell tolerance to sequences on other parts of the microbial antigen; therefore, T cells to these sequences will be present and can provide help for B cells that recognize the cross-reactive epitope. Studies have shown the importance of predicting linear B-cell epitopes (18, 21). *In silico* programs are also valuable to characterize the B-cell epitopes and class-II T-cell epitopes.

In this study, our design was a pathogenic island of *H. pylori* obtained from the GenBank and analyzed with ClustalW software. The structures of cag virb11 (Hp0525) and an in-

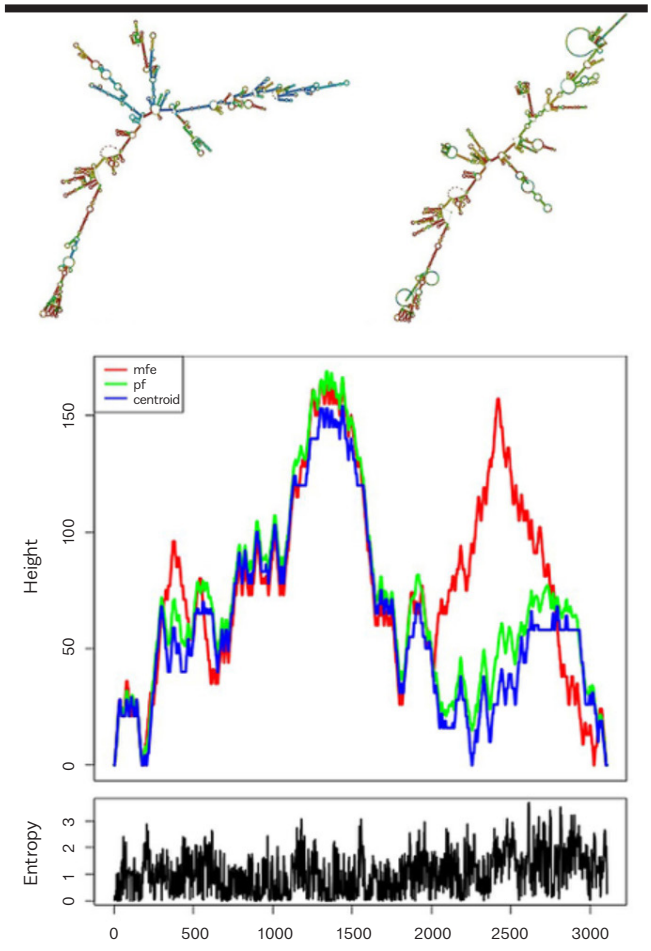

Figure 5. A mountain plot representation of the Maximum Free Energy structure, the thermodynamic ensemble of RNA structures, and the centroid structure. Additionally, we present the positional entropy for each position.

hibitory protein (Hp1451) were shown. Codon optimization and secondary- and tertiary-structure prediction for the cagPAI of *H. pylori* were analyzed. According to the physicochemical parameters, the cagPAI of *H. pylori* was analyzed and found to be stable. The physicochemical properties are important, and the BcePred prediction server was used to distinguish linear B-cell epitopes, with the precision of this process being 58.70%, a little better than any single residue property. The BcePred server gave the highest accuracy of 58.70% at a threshold of 2.38, which has been improved to calculate B-cell epitopes in an antigen sequence.

In addition, *H. pylori* is a type of carcinogen found to be associated with diseases, such as gastritis and peptic ulcers, and has a high incidence of infection in developing countries (22). Generally, antibiotic therapies are

successful, but sometimes, they are not effective (23). Therefore, vaccine development with recombinant DNA technology can be useful (24, 25). In our study, we used a bioinformatics approach to predict the potential B-cell and T-cell epitopes. We found that the cagPAI of *H. pylori* contains immune-dominant epitopes from distinct antigenic proteins, which include ureB, vacA, cagI, and cagE.

H. pylori and LPS are able to stimulate the expression of TLR4 in patients with gastric cancer and induce inflammatory responses. Moreover, it has been found that application of *H. pylori* and its vaccine increases the expression of TLR4 in the gastric mucosa and reduces *H. pylori* colonization. The B-cell and T-cell epitopes that we found may be able to induce both humoral and cellular immune responses. This *in silico* analysis should be combined with other pieces of evidence, including experimental data, to assign function. This analysis could help to determine whether the epitopes that we have predicted have the potential for vaccine studies.

B-cell epitopes for the protein could be predicted on the basis of solvent accessibility, antigenicity, flexibility, hydrophilicity, and secondary structure analyses. Only the epitopes with high prediction scores as continuous and discontinuous B-cell epitopes were selected by ABCpred. Our study has shown that sequences of ASDKNSKRYGVRTTDSRAA (start position: 166, stop position: 184) and DVVAST (start position: 113, stop position 118) were 2 strong candidates for B-cell epitopes. The sequence of ASDKNSKRYGVRTTDSRAA has the best value of accessibility and hydrophilicity. The high hydrophilicity and accessibility of this sequence made it a strong B cell candidate.

T-cell epitopes with a high potential to interact with human leukocyte antigen alleles were selected. The presence of amino acids with a high accessible value, such as LYS(R), SER(S), ASP(D), ASN(N), GLY(G), and ALA(A), would possibly result in the protein expressed in soluble form, which is important. Prediction of T-cell epitopes was confirmed by the NETCTL server, and 14 MHC ligands were identified at 0.75 threshold. The MHCPred software was also used to identify the common T-cell epitopes. After 2 sequences of RAYMYWAGV, VMAKVADMA was selected as a strong T-cell epitope. In conclusion, all epitopes would be discovered in the immune system *in vivo* in the future so that they have a potential to develop recombinant vaccine.

Ethics Committee Approval: N/A

Informed Consent: N/A.

Peer-review: Externally peer-reviewed.

Author Contributions: Concept - S.K.T., D.K., A.T.; Design - S.K.T., A.T.; Supervision- S.K.T.; Resource - S.K.T., D.K., A.T.; Data Collection and/or Processing - K.Ö., D.K.; Analysis and/or Interpretation - S.K.T., A.T.; Literature Search - S.K.T., D.K., A.T.; Writing - S.K.T., D.K., A.T.; Critical Reviews - S.K.T., D.K., A.T.

Conflict of Interest: The authors have no conflict of interest to declare.

Financial Disclosure: The authors declared that this study has received no financial support.

REFERENCES

1. Kekilli M, Onal IK, Ocal S, et al. Inefficacy of triple therapy and comparison of two different bismuth-containing quadruple regimens as a firstline treatment option for helicobacter pylori. Saudi J Gastroenterol 2016; 22: 366-9. [Crossref]
2. Ozaydin N,Turkyilmaz SA, Cali S. Prevalence and risk factors of Helicobacter pylori in Turkey: a nationally-representative, cross-sectional, screening with the (1)(3)C-Urea breath test. BMC Public Health 2013; 13: 1215. [Crossref]
3. Goodwin CS, Armstrong JA, Marshall BJ. Campylobacter pyloridis, gastritis, and peptic ulceration. J Clin Pathol 1986; 39: 353-65. [Crossref]
4. Kusters JG, Vliet AH, Kuipers EJ. Pathogenesis of Helicobacter pylori infection. Clin Microbiol Rev 2006; 19: 449-90. [Crossref]
5. Perez-Perez GI, Rothenbacher D, Brenner H. Epidemiology of Helicobacter pylori infection. Helicobacter 2004; 9: 1-6. [Crossref]
6. Smith SM. Role of Toll-like receptors in Helicobacter pylori infection and immunity. World J Gastrointest Pathophysiol 2014; 5: 133-46. [Crossref]
7. Day AS, Jones JN, Lynett JT, et al. cagE is a virulence factor associated with Helicobacter pylori-induced duodenal ulceration in children. J Infect Dis 2000; 181: 1370-5. [Crossref]
8. Blaser MJ, Berg DE. Helicobacter pylori genetic diversity and risk of human disease. J Clin Invest 2001; 107: 767-73. [Crossref]
9. Kawahara T, Teshima S, Oka A, et al. Type I Helicobacter pylori lipopolysaccharide stimulates toll-like receptor 4 and activates mitogen oxidase 1 in gastric pit cells. Infect Immun 2001; 69: 4382-9. [Crossref]
10. Mora-Solano C, Collier JH. Engaging adaptive immunity with biomaterials. J Mater Chem B 2014; 2: 2409-21. [Crossref]
11. Potocnakova L, Bhide M, Pulzova LB. An Introduction to B-Cell Epitope Mapping and In Silico Epitope Prediction. J Immunol Res 2016; 2016: 6760830. [Crossref]
12. Huleatt JW, Jacobs AR, Tang J et al. Vaccination with recombinant fusion proteins incorporating Toll-like receptor ligands induces rapid cellular and humoral immunity. Vaccine 2007; 25: 763-75. [Crossref]
13. Stothard P. The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. Bio Techniques 2000; 28: 1102-4. [Crossref]
14. Leiro V, Moreno PM, Sarmento B, et al. Bioinspired Materials for Medical Application. 2017; 1-44. [Crossref]
15. Fitzkee NC, Rose GD. Reassessing random-coil statistics in unfolded proteins. Proc Natl Acad Sci USA 2004; 101: 12497-502. [Crossref]

16. Panchenko AR, Madej T. Structural similarity of loops in protein families: toward the understanding of protein evolution. BMC Evol Biol 2005; 5: 10. [Crossref]
17. Kalali B, Mejías-Luque R, Javaheri A, et al. H. pylori virulence factors: influence on immune system and pathology. Mediators Inflamm 2014; 2014: 426309.[Crossref]
18. Censini S, Lange C, Xiang Z et al. cag, a pathogenicity island of Helicobacter pylori, encodes type I-specific and disease-associated virulence factors. Proc Natl Acad Sci USA 1996; 93: 14648-53.
19. Chauhan V, Singh MP, Ratho RK. Identification of T cell and B cell epitopes against Indian HCV-genotype-3a for vaccine development- An in silico analysis. Biologicals 2018; 53: 63-71. [Crossref]
20. Saha S, Raghava GP. Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. Proteins 2006; 65: 40-8. [Crossref]
21. Larsen JE, Lund O, Nielsen M. Improved method for predicting linear B-cell epitopes. Immunome Res 2006; 2: 2. [Crossref]
22. Singh H, Ansari HR, Raghava GP. Improved method for linear B-cell epitope prediction using antigen's primary sequence. PLoS One 2013; 8: 62216. [Crossref]
23. Chmiela M, Karwowska Z, Gonciarz W, et al. Host pathogen interactions in Helicobacter pylori related gastric cancer. World J Gastroenterol 2017; 23: 1521-40. [Crossref]
24. Hu Y, Zhu Y, Lu NH. Novel and Effective Therapeutic Regimens for Helicobacter pylori in an Era of Increasing Antibiotic Resistance. Front Cell Infect Microbiol 2017; 7: 168. [Crossref]
25. Tiptiri-Kourpeti A, Spyridopoulou K, Pappa A, et al. DNA vaccines to attack cancer: Strategies for improving immunogenicity and efficacy. Pharmacol Ther 2016; 165: 32-49. [Crossref]