Original Article

Applying Immunoinformatics Methods to Identify Potential T and B Cell Epitopes in the CagA Protein of *Helicobacter pylori*

Mostafa Akbarian Rokni[1], Somayyeh Talebishelimaki[2] Sakineh Poorhosein Fookolaee[2*], Mohammad Taha Saadati Rad[3], Mina Owrang[2]

1. Department of Nursing, School of Nursing & Midwifery, Babol University of Medical Sciences, Babol, Iran.
2. School of Medicine, Islamic Azad University- Sari Branch, Sari, Iran.
3. Young Researchers and Elite Club, Islamic Azad University- Sari Branch, Sari, Iran.

Received: April 1, 2019; Accepted: April 30, 2019

Abstract

Background and Aim: *Helicobacter pylori* is not only identified as a leading cause of chronic active gastritis and peptic ulcer disease in humans, but also it is considered as a risk factor for the development of gastric adenocarcinoma and MALT lymphoma. This study aims to predict specific epitopes for the utility of designing peptide vaccine against *H. pylori* infection by targeting invasive, virulent and membrane associated proteins CagA. **Materials and Methods:** In the present study, various immunoinformatics approaches have been applied to design a potential epitope-based vaccine against *H. pylori* infection. For prediction of linear epitopes, the sequence of CagA was submitted to ABCpred, BCPREDS, Bcepred, Bepipred and Ellipro servers. DiscoTope 2.0 and B-pred servers were also used for the prediction of conformational epitopes. In addition, prediction of T-cell epitopes was carried out by CTLPred. **Results:** The obtained results demonstrated 277 conformational B-Cell epitopes in addition to predicted high score linear B and T cell epitopes in CagA protein. **Conclusion:** These predicted epitopes might be used to design a vaccine against *H. pylori* and thus, could be validated in model hosts to verify their efficacy as vaccine.

Keywords: CagA; *Helicobacter pylori;* B cell Epitope; T cell Epitope; In silico.

***Corresponding Author:** Sakineh Poorhosein Fookolaee; Email: a52_amin@yahoo.com

Please cite this article as: Akbarian Rokni M, Talebishelimaki S, Poorhosein Fookolaee S, Saadati Rad MT, Owrang M. Applying Immunoinformatics Methods to Identify Potential T and B cell Epitopes in the CagA Protein of *Helicobacter pylori*. Arch Med Lab Sci. 2019;5(3):1-6. https://doi.org/10.22037/amls.v5i3.30561

Introduction

Helicobacter pylori is a microaerophilic spiral-shaped lophotrichous and gram-negative bacterium that colonizes in the gastric lumen of primates, including humans [1]. *H. pylori* was identified as the cause of chronic active gastritis and peptic ulcer disease in humans and is considered to be a risk factor for the development of gastric adenocarcinoma and mucosa-associated lymphoid tissue (MALT) lymphoma [2]. Strains of *H. pylori* are grouped into two broad families tentatively named type I and type II, which are based on whether they express or not the vacuolating cytotoxin (VacA) and the CagA antigen (cytotoxin-associated gene A) [3]. An increasing body of evidence has shown that patients with duodenitis, duodenal ulcers, and gastric tumors are most often infected by type I strains, which suggests that CagA and the co-expressed cytotoxin play a role in its pathogenicity [4]. Identification of the common immunogenic proteins of *H. pylori* may be a major step toward the development of purer, better-defined, and probably more-efficient vaccines. Such vaccines would confer cross-protection against a wide range of pathogenic *H. pylori* strains [5]. *H. pylori* vaccines development has been initiated from 2011 to predict novel T-cell epitopes. However, to the best of

our knowledge, there are no large pharmaceutical or biotechnology companies that have a current *H. pylori* vaccine development program. Therefore, the development of most *H. pylori* vaccines is currently at very early stage phases (Phase I or preclinical). For example, EpiVax, Helicovaxor®, and Lp220 are current vaccine candidates that are in preclinical phase [6]. These vaccines are composed of purified or recombinant components of *H. pylori* antigens with an adjuvant.

Current evidence indicates that CagA can be a good candidate for developing a new generation of vaccines that induce broad protection against *H. pylori* [7]. Active immunotherapy with peptide vaccines, which are designed to be chimeric with multi-epitopes of B cells and T helper cells, can induce generation of an adaptive immune response. Several experimental techniques are currently available for selection of suitable B and T cell epitopes [8]. The experimental approaches applied for detecting immunogenic regions are often laborious and resource-intensive. Computational techniques are fast, scalable, and cost-effective for B and T cell epitopes prediction, for focusing experimental investigations and for better understanding of antigen-antibody interactions [9, 10]. B cell epitopes can be classified into two types: linear (continuous) and conformational (discontinuous). While linear epitopes comprise residues that are continuous in the sequence, conformational epitopes are composed of amino acids that are not neighboring in primary sequence and are brought into close proximity in the folded protein structure [11]. Moreover, sequence- and structure-based methods investigating the binding affinity of peptides to MHC I and MHC II molecules and other parameters may aid in the search for new antigens in order to support vaccine development [12]. The present research was conducted to predict new B and T cell epitopes for the CagA of *H. pylori*.

Methods

Retrieval of protein sequences. All sequences of CagA were retrieved from UniProt (https://www.uniprot.org) in FASTA format.

Linear B cell epitopes prediction for CagA. For prediction of linear epitopes, CagA sequence was submitted to ABCpred, BCPREDS, Bcepred, Bepipred and Ellipro servers (13, 14). The hidden Markov model, Thornton's method, Support Vector Machine classifiers, Recurrent Neural Network and physico-chemical properties of amino acids were applied to predict the linear B cell epitopes. Only, the linear peptides which were predicted frequently by 3 or more servers were selected.

Conformational B cell epitope prediction for CagA. DiscoTope 2.0 and B-pred servers were used for the prediction of conformational epitopes from the entire PDB structure of CagA that obtained from the homology modeling method (15). More recently, the sensitivity and specificity of DiscoTope 2.0 server for the prediction of vaccine against *H. pylori* were 0.47 and 0.75, respectively (16). The DiscoTope method incorporates a new spatial neighborhood description and a half-sphere exposure as a surface measure based on the protein structure and epitope propensity scores and predicts residues that can be involved in B-cell epitopes. B-pred is a web-based platform for scoring and predicting B-cell epitopes based on the structures of the potential immunological proteins. The method scores the peptides set of a protein based on the average solvent exposure, by a filter on filtering the average local model quality for each peptide.

T cell epitope prediction for CagA. Prediction of T-cell epitopes was carried out by CTLPred at http://www.imtech.res.in/raghava/ctlpred/. This server provides a direct method for the prediction of cytotoxic T lymphocyte (CTL) epitopes crucially based on Quantitative Matrix (QM), Support Vector Machine (SVM), and Artificial Neural Network (ANN) in subunit vaccine design [17].

MHC class I and II epitope prediction. Prediction of the promiscuous MHC class-I binding peptides was carried out by using of ProPred-I at http://www.imtech.res.in/raghava/propred1/. Also, this online server predicts MHC-I binding peptides for 47 various alleles [18]. Prediction of binding peptides to class II MHC molecules was performed by means of RANKPEP online server [19].

Results

The predicted linear B cell epitopes of the Cag A are shown in Table 1. The predicted B cell epitopes were ranked according to their score obtained by trained recurrent neural network. Higher score of the peptide means the higher probability to be as epitope. As listed in the table, peptides were predicted by different servers used in the current study. The peptide "NETIDQTRTPDQTQSQ" had highest score compared to the other epitopes. These peptides reside in 3-968 regions of the Cag A.

Table 1. Predicted B cell epitopes according to their score obtained by trained recurrent neural network.

Rank	Sequence	Start position	Score
1	NETIDQTRTPDQTQSQ	3	0.96
2	QKPIVDKNDRDNRQAF	48	0.94
2	SQTAFDPQQFINNLQV	17	0.94
3	GGVGQAAGFPLKRHDK	940	0.93
3	DALGNDRIAFVSKKDT	442	0.93
4	QPDIATTTTDIQGLPP	242	0.92
4	ENIIQPPIPDDKEKAE	148	0.92
5	EAEKNGGPTGGDWLDI	200	0.91
6	GKKADKALDREKNVTL	479	0.90
7	GDLSYTLKDYGKKADK	469	0.89
7	AGNGGFGDKHDWNATV	325	0.89
8	DGISQLREEYSNKAIK	64	0.88
8	SGLVIAGGEKGINNPS	362	0.88
8	DVEGVADIDPNYKFNQ	281	0.88
9	PEPIYATIDDLGGPFP	968	0.87
9	QEEIRNKVDFMEFLAQ	395	0.87

Discotope and B-pred predicted 277 B-Cell epitope residues out of 656 total residues corresponding to conformational epitopes, approximately located in the region of linear B cell epitopes (Table 2). Further analysis for solvent accessible areas and relative solvent accessibility of all the residues on the PDB structures using Naccess program and NetSurfP server defined that predicted conformational B cell epitopes had higher solvent accessible and their residues were exposed on the surface (Fig 1).

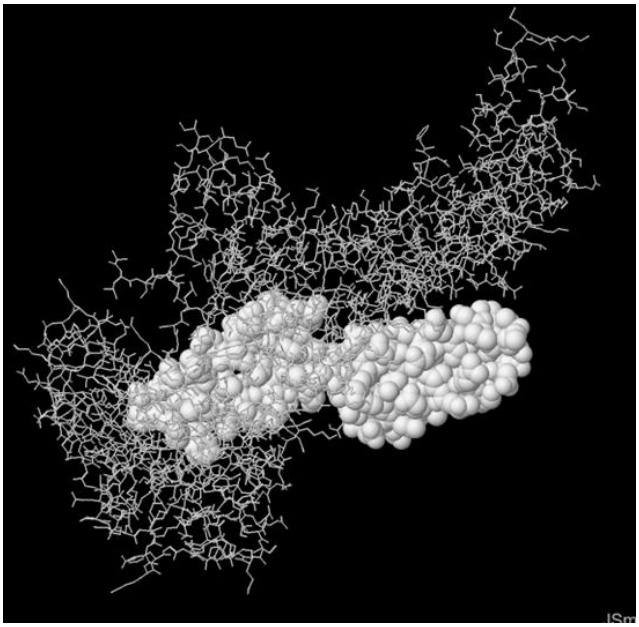

Fig 1. 3D structure of high score predicted discontinuous B cell epitopes. Predicted conformational B cell epitopes has higher solvent accessible and their residues were exposed on the surface.

Analysis using a combined algorithm integrating MHC epitope binding and T cell epitopes generated 3 T cell epitopes from Cag A of *H. pylori* that interacted with different possible MHC I alleles with the IC50 value <100 nM. These processed peptides were then analyzed by SMM based IEDB MHC I prediction tool. A good epitope should also interact with as many as MHC alleles. Thus, among the total peptides, only those peptides which interacted with minimum four MHC class I alleles as well as were found to be the core sequences of 15mer MHC class II alleles, were selected from each protein (Table 3).

Table 2. Predicted discontinuous B cell epitopes according to their score

No.	Residues	Number of residues	Score
1	P:I736, P:N737, P:P738, P:E739, P:W740, P:I741, P:S742, P:K743, P:V744, P:E745, P:N746, P:L747, P:N748, P:A749, P:A750, P:L751, P:N752, P:E753, P:F754, P:K755, P:N756, P:F762, P:S763, P:K764, P:V765, P:T766, P:Q767, P:A768, P:K769, P:S770, P:D771, P:L772, P:E773, P:N774, P:S775, P:V776, P:K777, P:D778, P:V779, P:I780, P:I781, P:N782, P:Q783, P:K784, P:V785, P:T786, P:D787, P:K788, P:V789, P:D790, P:N791, P:L792, P:N793, P:Q794, P:A795, P:V796, P:S797, P:V798, P:A799, P:K800, P:A801, P:M802, P:G803, P:D804, P:F805, P:S806, P:R807, P:V808, P:E809, P:Q810, P:V811, P:L812, P:A813, P:D814, P:L815, P:K816, P:N817, P:F818, P:S819, P:K820, P:E821, P:Q822	82	0.803
2	P:P546, P:D547, P:L548, P:N549, P:N550, P:L551, P:A552, P:I553, P:T554, P:S555, P:F556, P:V557, P:R558, P:R559, P:N560, P:L561, P:E562, P:N563, P:K564, P:L565, P:T566, P:A567, P:K568, P:G569, P:L570, P:S571, P:L572, P:Q573, P:E574, P:A575, P:N576, P:K577, P:L578, P:I579, P:K580, P:D581, P:L583, P:S584, P:R626, P:H628, P:L629, P:E630, P:K631, P:E632, P:V633, P:E634, P:K635, P:K636, P:L637, P:E638, P:S639, P:K640, P:S641, P:N645, P:K646, P:M647, P:E648, P:A649, P:K650, P:A651, P:Q652, P:A653, P:N654, P:S655, P:Q656, P:K657, P:D658, P:E659, P:I660, P:F661, P:A662, P:L663, P:I664	73	0.784
3	P:L69, P:E71, P:E72, P:Y73, P:S74, P:N75, P:K76, P:A77, P:I78, P:N80, P:P81, P:T82, P:K83, P:K84, P:N85, P:Q86, P:Y87, P:F88, P:S89, P:D90, P:F91, P:I92, P:D93, P:K94, P:S95, P:N96, P:D97, P:L98, P:I99, P:N100, P:K101, P:D102, P:N103, P:L104, P:I105, P:D106, P:V107, P:E108, P:S109, P:S110, P:T111, P:K112, P:S113, P:F114, P:Q115, P:K116, P:F117, P:G118, P:D119, P:Q120, P:R121, P:Y122, P:Q123, P:I124, P:K186, P:F187, P:M188, P:G189, P:V190, P:F191, P:D192, P:E193, P:S194, P:L195, P:K196, P:E197, P:R198, P:Q199, P:E200, P:A201, P:E202, P:K203, P:N204, P:G205, P:G206, P:P207, P:T208, P:G209, P:G210, P:D211, P:W212, P:L213, P:D214, P:I215, P:L217, P:S218	86	0.728
4	P:L699, P:K700, P:F702, P:S703, P:K704, P:S705, P:F706, P:D707, P:E708, P:F709, P:K710	11	0.718
5	P:E148, P:N149, P:I150, P:I151, P:Q152, P:P153, P:P154, P:I155, P:P156, P:D157, P:D158, P:K159, P:E160, P:K161, P:E163	15	0.699
6	P:E318, P:K319, P:G341, P:Y342, P:K343, P:N349, P:V350	7	0.675
7	P:A303, P:L304, P:S305, P:S306, P:V307, P:L308, P:G310, P:S311, P:H312, P:N313, P:G314, P:I315, P:V320, P:S321, P:L322, P:Y324, P:G326, P:N327, P:G328, P:G329, P:F330, P:H334	22	0.621
8	P:E588, P:K617, P:D618, P:E620, P:K621, P:S622, P:R624, P:K625	8	0.602
9	P:Y508, P:S509, P:F537	3	0.584
10	P:K455, P:D456, P:T457, P:K458, P:H459	5	0.568
11	P:D45, P:P46, P:D47, P:Q48	4	0.557
12	P:N608, P:D610, P:K613	3	0.504

Table 3. Predicted T cell epitopes according to their score using different server

Peptide Rank	Start Position	Sequence	Score (ANN/SVM)
1	22	DPQQFINNL	0.97/1.2919966
2	967	SPEPIYATI	0.94/1.2222893
3	436	DSKAYLDAL	0.96/1.0907681

Discussion

To find out remedial alternatives, the identification of new bio-targets is, therefore, highly anticipated. Understanding of molecules involved in pathogenesis has the potential to yield new and exciting strategies for therapeutic intervention [20]. The CagA protein antigen of *H. pylori* is arguably the most well-studied virulence factor of *H. pylori* that facilitates its colonization into the intestinal tract [21]. In this study, we made an attempt to predict novel T and B cell epitopes which could be tested for their efficacy in eliciting immunity through humoral and cell mediated immune responses. However, the activation of cytotoxic T-cells requires recognition of specific peptides bound to MHC class I molecules and while sequences from pathogens provide a huge amount of potential vaccine candidates, it is estimated that only one in 100 to 200 peptides actually binds to a particular MHC. Therefore, a good computational prediction method can significantly reduce the number of peptides

that have to be synthesized and tested. With the advent of computers and informatics, new approaches have been devised that facilitate immunoinformatics which targets the use of mathematical and computational approaches to predict T-cell and B-cell immune epitopes [22]. Identification of B-cell epitopes is rather important to immunodetection and immunotherapeutic applications since an epitope as the minimal immune unit is strong enough to elicit a potent humoral immune response with no harmful side effects to human body [11]. In this study, linear B-cell epitopes were chosen with two different algorithms, BepiPred and ABCpred. Only overlapping peptides which were chosen by both algorithms, as well as satisfied the scores of VaxiJen and passed transmembrane topology were selected as potential B-cell epitopes and subjected to further analysis using the parameters of surface accessibility, hydrophilicity, flexibility, and beta-turn. By cross-referencing of all the data, we predicted that the peptide sequences were capable of inducing the desired immune response as B cell epitope. For T cell epitope prediction, plenty of algorithms are freely available, and in this study, we employed IEDB analysis tool which is possibly the most wide-ranging database offering several B cell and T cell epitope-related analysis and prediction tools, as well as provides both intrinsic biochemical and extrinsic context dependent information about them. Initial analysis of the data showed that most of the predicted epitopes interacted with the highest numbers of MHC class I alleles, were the core peptides of a maximum number of MHC class II binding predictions and also demonstrated large world population coverage [16]. Further analysis found out the highest population coverage, interacted with 6-9 MHC class I alleles and also were found to be overlapping with a good number of 15mer peptides of MHC class II alleles. All epitopes showed 100 % conservancy among the majority of strains. In summary, in this study, we made an attempt to predict novel B and T cell epitope against Cag A of *H. pylori* which could be tested for their efficacy in eliciting immunity through humoral and cell mediated immune responses. The results of our study provide computational data for the identification and screening of epitopes, and may be used for the development of epitope vaccines that have an enhanced safety and efficacy. Our findings are based on sequence analysis and computational predictions. However, to prove the effectiveness of mounting an immune response, both in vitro and in vivo studies are required along with this in silico study.

Conclusion

In this research, we used immunoinformatics methods to predict appropriate vaccine against *H. pylori*. Our data revealed 277 conformational B-Cell epitopes in addition to predicted high score linear B and T cell epitopes in Cag A protein. All sequences were joined to each other by proper linkers. Epitopes were evaluated as nonallergenic, antigenic, soluble, with safety and efficacy. These predicted epitopes might be used to design a vaccine against *H. pylori* and thus, could be validated in model hosts to verify their efficacy as vaccine.

Conflicts of Interest

The authors declare that they have no conflict of interest.

References

1. Kuipers E, Pena A, Festen H, Meuwissen S, Uyterlinde A, Roosendaal R, et al. Long-term sequelae of Helicobacter pylori gastritis. The Lancet. 1995;345(8964):1525-8.

2. Parsonnet J, Friedman GD, Vandersteen DP, Chang Y, Vogelman JH, Orentreich N, et al. Helicobacter pylori infection and the risk of gastric carcinoma. New England Journal of Medicine. 1991;325(16):1127-31.

3. Suerbaum S, Michetti P. Helicobacter pylori infection. New England Journal of Medicine. 2002;347(15):1175-86.

4. Censini S, Lange C, Xiang Z, Crabtree JE, Ghiara P, Borodovsky M, 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. 1996;93(25):14648-53.

5. Del Giudice G, Covacci A, Telford JL, Montecucco C, Rappuoli R. The design of vaccines against Helicobacter pylori and their development. Annual review of immunology. 2001;19(1):523-63.

6. Suttona P, Boag JM. Status of vaccine research and development for Helicobacter pylori. Vaccine. 2019;37:7295–9.

7. Zhang S, Moise L, Moss SF. H. pylori vaccines: why we still don't have any. Human vaccines. 2011;7(11):1153-7.

8. Hayman WA, Brandt ER, Relf WA, Cooper J, Saul A, Good MF. Mapping the minimal murine T cell and B cell epitopes within a peptide vaccine candidate from the conserved region of the M protein of group A streptococcus. International immunology. 1997;9(11):1723-33.

9. Rothbard JB, Taylor WR. A sequence pattern common to T cell epitopes. The EMBO journal. 1988;7(1):93-100.

10. Yasser E-M, Honavar V. Recent advances in B-cell epitope prediction methods. Immunome research. 2010;6(2):S2.

11. Blythe MJ, Flower DR. Benchmarking B cell epitope prediction: underperformance of existing methods. Protein Science. 2005;14(1):246-8.

12. Nielsen M, Lund O, Buus S, Lundegaard C. MHC class II epitope predictive algorithms. Immunology. 2010;130(3):319-28.

13. Saha S, Raghava G, editors. BcePred: prediction of continuous B-cell epitopes in antigenic sequences using physico-chemical properties. International Conference on Artificial Immune Systems; 2004: Springer.

14. Saha S, Raghava G. Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. Proteins: Structure, Function, and Bioinformatics. 2006;65(1):40-8.

15. Haste Andersen P, Nielsen M, Lund O. Prediction of residues in discontinuous B-cell epitopes using protein 3D structures. Protein Science. 2006;15(11):2558-67.

16. Urrutia-Baca VH, Gomez-flores R, De La Garza-Ramos MA, Tamez-guerra P, Lucio-sauceda DG, Rodríguez-padilla MC. Immunoinformatics Approach to Design a Novel Epitope-Based Oral Vaccine Against Helicobacter pylori. J Comput Biol. 2019;26:1177–90.

17. Bhasin M, Raghava GP. Prediction of CTL epitopes using QM, SVM and ANN techniques. Vaccine. 2004;22(23-24):3195-204.

18. Singh H, Raghava G. ProPred: prediction of HLA-DR binding sites. Bioinformatics. 2001;17(12):1236-7.

19. Reche PA, Glutting J-P, Zhang H, Reinherz EL. Enhancement to the RANKPEP resource for the prediction of peptide binding to MHC molecules using profiles. Immunogenetics. 2004;56(6):405-19.

20. Franke R, van Gerven J. Theoretical drug design methods. 1984.

21. Akopyants NS, Clifton SW, Kersulyte D, Crabtree JE, Youree BE, Reece CA, et al. Analyses of the cag pathogenicity island of Helicobacter pylori. Molecular microbiology. 1998;28(1):37-53.

22. Bergmann-Leitner ES, Chaudhury S, Steers NJ, Sabato M, Delvecchio V, Wallqvist AS, et al. Computational and experimental validation of B and T-cell epitopes of the in vivo immune response to a novel malarial antigen. PloS one. 2013;8(8):e71610.