Original Article   Open Access

Prediction of T-cell epitopes for designing a reverse vaccine against streptococcal bacteria

Samira Ebrahimi[1], Hassan Mohabatkar[1],*

Department of Biotechnology, Faculty of Advanced Sciences and Technologies, University of Isfahan, Isfahan, Iran

ABSTRACT

Streptococcal bacteria are among dangerous human pathogens with major prevalence worldwide. A good vaccine against streptococcal bacteria should have epitopes that confer protection from infection by different streptococcal bacteria types. we aimed was to recognize the most immunogenic and conserved epitopes of streptococcal bacteria, which could be a potential candidate for vaccine development. Nineteen different M proteins of different streptococcal bacteria were chosen and analyzed. Nine-mer epitopes able to simulate both cells mediate and humoral immunity were predicted. Molecular docking was applied in order to measure free binding energy of selected epitopes. Final epitopes were analyzed if they were conserved among different streptococcal bacteria. The identified epitopes require experimental validation for their potential use in peptide vaccines.

Keywords: Epitope prediction; Molecular docking; M protein; Streptococcal bacteria; Vaccine design.

INTRODUCTION

Streptococcal bacteria belong to the phylum Firmicutes and the order lactobacillus [1]. Streptococcal bacteria are among the most important human pathogens responsible for a large variety of diseases among humans ranging from neonates to the elderly [2]. Group A streptococci cause at least 517000 deaths because of severing invasive diseases, 616 million cases of pharyngitis, and 111 million cases of pyoderma [3]. Group B streptococci can cause a wild range of diseases including meningitis, pneumonia, and sepsis. Direct mother-to-baby transmission of the pathogen is the main cause of infections in neonates [4]. Other groups of streptococcal bacteria cause different dangerous human diseases such as brain abscess, pulmonary abscess, and severe infections.

In establishing infection in the host, an initial and crucial step is the attachment of bacteria to the mucosa and dermal membranes [5]. The main role of bacterial surface proteins is to interact with host cells. Hence, these proteins are potential targets for designing a vaccine against bacterial infections and diseases [6]. M protein is one of the major virulence factors of invasive streptococcal bacteria isolates [7]. This protein is a surface protein that contains two

*Corresponding Author: Department of Biotechnology, Faculty of Advanced Sciences and Technologies, University of Isfahan, Isfahan, Iran
Tel: +98-313-7934391
Fax: +98-313-7932342
E. mail: h.mohabatkar@ast.ui.ac.ir

polypeptide chains. M protein has a sorting part in carboxy-terminal that is composed of LPXTG motif required for the exact attachment of M protein onto the cell wall [8]. M proteins have an affinity for albumin, plasminogen, fibrinogen, IgG, IgA, and C4b-binding protein [9] and they have a crucial role in the variety of diseases that can be evoked by streptococcal bacteria [10]. Following the work of Rebecca Lancefield that proved antibodies against the N-terminal of M protein could bring protection in a specific manner, vaccine development against streptococcal bacteria focused on M protein [11].

To reduce costs and experimental efforts in vaccine design, prediction of potentially immunogenic epitopes in a given protein would be a useful way [12]. In the development of vaccines against pathogens, the epitope-driven vaccine is an attractive concept [13]. In these vaccines, selected epitopes should be conserved across pathogen variants and their stages. Also, selected epitopes should evoke the immune response in the desired way. Epitopes ideally must cover a major population and have a binding affinity with more than one major histocompatibility complex (MHC) allele [14].

In this study, a reverse vaccinology approach was applied for epitope prediction of different M proteins of streptococcal bacteria. Epitope prediction is based on predicting epitopes that stimulate both humoral and cell-mediated immunity. Molecular docking was applied to model the interaction between the final epitopes with HLA-A1 allele. Conservation of peptides among studied M proteins was checked. Finally, selected epitopes were checked if they were promiscuous epitopes, which means they could be epitopes for several MHC class I and II alleles.

MATERIALS AND METHODS

Selection of sequences: The sequence of 19 different M proteins from different streptococcal bacteria, from NCBI and UniProt were selected. M proteins with their accession numbers are shown in Table 1.

Table 1: Selected M proteins of streptococcal bacteria with their accession number

Name of Bacteria	M protein	Accession Number
agalactiae Streptococcus	M protein repeat protein	KLL29245.1
S. pyogene	Antiphagocytic M protein, type 3	WP_011055052.1
S. pyogenes	Antiphagocytic M protein Emm	WP_014407886.1
S. pyogenes GAS	M protein type 1	NP_269973.1
S. pyogenes	M protein, serotype 5	WP_023079553.1
S. pyogenes	M12_STRPY M protein	P19401
S. pyogenes	M protein, serotype 24	WP_032465389.1
S. equi subsp. zooepidemicus	M-protein	AGV40770.1
S. equi	M-protein	AAB71984.1
S. equi subsp. ruminatorum	M-protein	KED03515.1
S. sp. 'group G'	M protein	AAA26928.1
S. canis	M protein	WP_003045324.1
s. dysgalactiae	Antiphagocytic M protein	WP_015016601.1
S. pyogenes	M protein, serotype 6	WP_023610762.1
S. pyogenes	M protein, serotype 49	WP_023612186.1
S. parauberis	M protein	WP_004347161.1
S. pneumoniae	M protein repeat family protein	WP_001854666.1
S. pyogenes	M protein	WP_011529059.1
S. dysgalactiae	M protein, serotype 24	WP_048327566.1

T-cell epitope prediction: In the present investigation, MAPPP [15], ComPred [16], HLArestrictor [17], MHCPred [18], NetMHC [19], ProPred I [20], RANKPEP [21], SVMHC

[22], Vaxitope [23], and IEDB [24] were used as epitope prediction tools. The HLA-A1 restricted 9-mer epitopes of each M protein were predicted by the above-mentioned tools. For each tool and each M protein, top 10 predicted epitopes were selected, and those recognized by at least five tools were selected for further analysis.

Applying filters B-Cell epitope prediction: The main purpose of this part of the investigation was to select 9-mer T-cell epitopes, which were part of 20-mer linear B-cell epitopes. Selected epitopes can stimulate both humoral and cell-mediate immunity. ABCpred software (www.imtech.res.in/raghava/abcpred) was used in order to predict antigenicity linear non-overlapping 20-mer B-cell epitopes [25]. B-cell epitopes with score> 0.6 were analyzed.

Antigenicity of the epitopes: VaxiJen (www.ddgpharmfac.net/vaxijen) was used to identify if epitopes are antigenic. vaxiJen classified antigens based on the physicochemical properties of proteins [26]. Epitopes with VaxiJen score >0.7 were selected.

Binding affinity analysis: Binding affinity was predicted by means of MHCPred server. The results of binding affinity prediction between HLA-A1 allele and selected epitopes were given in terms of inhibitory concentration [IC50] [18]. Epitopes with the IC50 value less than 1000 nm for HLA-A1 were selected.

Molecular docking of the epitopes binding to HLA-A1: The 3D structure of selected epitopes binding to HLA-A1 was needed in order to do molecular docking. The 3D structure of selected epitopes was designed using PEPstr server, which predicts the tertiary structure of peptide from the sequence given as input [27]. In addition, the 3D structure of HLA-A1 was retrieved from protein data bank (PDB ID: 4NQV). Epitopes were docked into HLA-A1 allele using ClusPro. The ClusPro program is one of the widely used tools for protein-protein docking [28].

Selected T-cell epitopes conservation: Final epitopes were checked if they were conserved among different studied M proteins.

Selection of promiscuous epitopes: In the last step, conserved epitopes were checked if they were promiscuous. Promiscuous T-cell epitopes are those, which could bind to different MHC class I and II alleles. ProPred I was used as an MHC class I epitope prediction server [20], and ProPred was used to predict MHC class II epitopes [29].

RESULTS

Top 10 T-cell epitopes of each of 10 servers recognized by at least 10 servers were identified. The numbers of selected epitopes for each protein are shown in Table 2. In order to select final practical epitopes, different filters were applied. B-cell epitopes with score >0.6 were predicted in order to find epitopes, which stimulate both humoral and cell-mediated immune responses. Antigenicity of the epitopes was measured by VaxiJen server, and epitopes with score <0.7 were omitted. Binding affinity between epitopes and HLA-A1 allele was measured by MHCPred server, and epitopes with IC50 >1000 nm were selected for each protein. Molecular docking between epitopes and HLA-A1 alleles was applied and they showed negative free binding energy in all cases. Different features of final epitopes, which fulfill all mentioned criteria, are shown in Table 3.

Conservation of the selected T-cell epitopes of each M protein among other M proteins was checked. 9 epitopes were conserved among studied M proteins, as listed in Table 3. The final

step of this investigation is to find out if final conserved epitopes are promiscuous. The numbers of MHC class I and II alleles bound to final epitopes are shown in Table 4.

Table 2: Number of predicted epitopes by at least 5 servers for each M protein.

Name of bacteria	**M protein**	**Number of Predicted T-cell Epitopes**
S. agalactiae	M protein repeat protein	28
S. pyogenes	Antiphagocytic M protein, type 3	39
S. pyogenes	Antiphagocytic M protein Emm	30
S. pyogenes M1 GAS	M protein type 1	27
S. pyogenes	M protein, serotype 5	22
S. pyogenes	M12_STRPY M protein	31
S. pyogenes	M protein, serotype 24	29
S. equi subsp. zooepidemicus	M-protein	25
S. equi	M-protein	34
S. equi subsp. ruminatorum CECT 5772	M-protein	38
Streptococcus sp. 'group G'	M protein	34
S. canis	M protein	27
S. dysgalactiae	Antiphagocytic M protein	38
S. pyogenes	M protein, serotype 6	36
S. pyogenes	M protein, serotype 49	34
S. parauberis	M protein	31
S. pneumoniae	M protein repeat family protein	25
S. pyogenes	M protein	35
S. dysgalactiae	M protein, serotype 24	30

Table 3: Different feathers of final conserved epitopes

Conserved T-cell Epitope	**Epitope VaxiJen score**	**No. of methods**	**HLA-A1 (IC50)**	**Binding free energy (Kcal/mol)**	**No. of alleles bounded MHC I+MHC II**
LTEKQNRVS	1.3667	5	371.54	-777.6	1+1
EAELNNLKA	0.8786	7	119.12	-572.3	3+0
IADLETKLK	1.2518	8	115.61	-696.7	1+2
LSDKRHQQE	1.8616	7	56.36	-850.5	1+0
ALEEANSKL	1.0475	9	75.34	-603.2	11+2
ELEAKHQKL	1.4469	7	45.71	-540.9	8+1
ALTAEHQKL	0.8842	5	99.77	-671.1	10+0
LTAEHQKLK	1.8546	6	240.44	-691.4	4+1
GHQHAHNEY	0.9194	5	37.84	-1013.4	3+3

DISCUSSION

The introduction of genomic technologies such as recombinant DNA technology, in silico analysis, and proteomics revoluted the approach of studying bacterial pathogens and vaccine design. Genomic technologies say that every antigenic part of a pathogen can induce a protective immune response and should be tested to be used in vaccine design [30]. One such approach is reverse vaccinology, in which the mining of genomic and proteomic information of the organism using bioinformatics-based software is done [31].

Initial and crucial step in colonization and establishing infection in the host is the adhesion of bacteria to the mucous and dermal membrane [32]. Surface proteins of pathogens are mostly pathogenic and cause bacterial pathogenicity. Surface proteins are potential targets for vaccine

development mostly at preventing bacterial disease and infections [33]. One of the major virulence factors of streptococcal bacteria is M protein, the main role of which is the presentation of opsonophagocytosis [34]. This protein is the favored target of vaccine design [35].

Table 4: Conserved epitopes of different studied M proteins

Conserved Epitope	Proteins with conserved epitopes
LTEKQNRVS	M protein, type 3 of S.pyogenes
	M protein, serotype 24 of *S. pyogenes*
EAELNNLKA	M protein, type 3 of *S. pyogenes*
	M protein, serotype 24 of *S. pyogenes*
	M protein, serotype 12 of *S. pyogenes*
IADLETKLK	M protein, type 3 of *S. pyogenes*
	M protein, serotype 24 of *S. pyogenes*
LSDKRHQQE	M protein, type 3 of *S. pyogenes*
	M protein, serotype 24 of *S. pyogenes*
ALEEANSKL	M protein of *S. dysgalactiae*
	M protein, type 1 of *S. pyogenes*
	M protein, type 5 of *S. pyogenes*
	M protein, serotype 12 of *S. pyogenes*
	M protein, serotype 24 of *S. dysgalactiae*
	M protein of group G
ELEAKHQKL	M protein Emm of *S. pyogenes*
	M protein, type 49 of *S. pyogenes*
ALTAEHQKL	M protein Emm of *S. pyogenes*
	M protein of *S. pyogenes*
LTAEHQKLK	M protein Emm of *S. pyogenes*
	M protein of *S. pyogenes*
GHQHAHNEY	M protein, serotype 12 of *S. pyogenes*
	M protein, serotype 24 of *S. dysgalactiae*

Early studies in vaccine design against streptococcal bacteria were focused on using whole protein [36]. Later studies showed contribution of epitopes to the pathogenicity of bacteria. Consequently, the use of full-length M protein as streptococcal bacterial vaccine candidates was prohibited. Later studies targeted immunogenic amino-terminal region of M proteins [37] and conserved C-terminal region of M protein [38]. Until now, efforts to develop a useful universal vaccine against streptococcal bacteria have not been unsuccessful. Therefore, in this investigation, the main aim was the detection of antigenic conserved epitopes in different M proteins. In order to recognize the most appropriate candidates for vaccine design, different filters were sequentially applied. Finally, selected T-cell epitopes, which are antigenic and capable of inducing antibody production, are detected. The identified epitopes require experimental validation for their potential use in peptide vaccines.

In conclusion, in this work, antigenicity testing, B-cell and T-cell epitope prediction, and molecular docking has been done for different streptococcal M proteins. Our results show that nine different are appropriate candidates for peptide vaccine. Molecular docking results show that these nine peptides have low free binding energy in binding to MHC alleles, which means these peptides are good binders and can simulate the immune response. *In vitro* and *in vivo* immunological tests should be performed in order to validate the suitability of the epitopes for vaccine development.

Conflict of Interest: There is nothing to declare.

REFERENCES

1. Ryan KJ, Ray CG. Medical microbiology. McGraw Hill. 2004;4:370.

2. Kreikemeyer B, Gámez G, Margarit I, Giard J-C, Hammerschmidt S, Hartke A, Podbielski A. Genomic organization, structure, regulation and pathogenic role of pilus constituents in major pathogenic Streptococci and Enterococci. Int J Med Microbiol 2011;301:240-251.

3. Carapetis JR, Steer AC, Mulholland EK, Weber M. The global burden of group A streptococcal diseases. Lancet Infect Dis 2005;5:685-694.

4. Maisey HC, Quach D, Hensler ME, Liu GY, Gallo RL, Nizet V, Doran KS. A group B streptococcal pilus protein promotes phagocyte resistance and systemic virulence. FASEB J 2008;22:1715-1724.

5. Oehmcke S, Shannon O, Mörgelin M, Herwald H. Streptococcal M proteins and their role as virulence determinants. Clin Chim Acta 2010;411:1172-1180.

6. Wei Z, Fu Q, Liu X, Xiao P, Lu Z, Chen Y. Identification of Streptococcus equi ssp. zooepidemicus surface associated proteins by enzymatic shaving. Vet Microbiol 2012;159:519-525.

7. Hamula CL, Peng H, Wang Z, Tyrrell GJ, Li XF, Le XC. An improved SELEX technique for selection of DNA aptamers binding to M-type 11 of Streptococcus pyogenes. Methods 2016;97:51-57.

8. Fischetti VA, Pancholi V, Schneewind O. Conservation of a hexapeptide sequence in the anchor region of surface proteins from gram-positive cocci. Mol Microbiol 1990;4:1603-1605.

9. Su YF, Chuang WJ, Wang SM, Chen WY, Chiang-Ni C, Lin YS, Wu JJ, Liu CC. The deficient cleavage of M protein-bound IgG by IdeS: insight into the escape of Streptococcus pyogenes from antibody-mediated immunity. Mol Immunol 2011;49:134-142.

10. Berge A, Kihlberg B-M, Sjöholm AG, Björck L. Streptococcal protein H forms soluble complement-activating complexes with IgG, but inhibits complement activation by IgG-coated targets. J Biol Chem 1997;272:20774-20781.

11. Moreland NJ, Waddington CS, Williamson DA, Sriskandan S, Smeesters PR, Proft T, Steer AC, Walker MJ, Baker EN, Baker MG, Lennon D, Dunbar R, Carapetis J, Fraser JD. Working towards a group A streptococcal vaccine: report of a collaborative Trans-Tasman workshop. Vaccine 2014;32:3713-3720.

12. Burt FJ, Samudzi RR, Randall C, Pieters D, Vermeulen J, Knox CM. Human defined antigenic region on the nucleoprotein of Crimean-Congo hemorrhagic fever virus identified using truncated proteins and a bioinformatics approach. J Virol Methods 2013;193:706-712.

13. Gershoni JM, Roitburd-Berman A, Siman-Tov DD, Freund NT, Weiss Y. Epitope mapping. BioDrugs 2007;21:145-156.

14. Soria-Guerra RE, Nieto-Gomez R, Govea-Alonso DO, Rosales-Mendoza S. An overview of bioinformatics tools for epitope prediction: implications on vaccine development. J Biomed Inform 2015;53:405-414.

15. Hakenberg J, Nussbaum AK, Schild H, Rammensee H-G, Kuttler C, Holzhütter HG, Kloetzel PM, Kaufmann SH, Mollenkopf HJ. MAPPP: MHC class I antigenic peptide processing prediction. Appl Bioinformatics. 2003;2:155-158.

16. Lata S, Bhasin M, Raghava GPS. Application of machine learning techniques in predicting MHC binders. Methods Mol Biol 2007;409:201-215.

17. Larsen ME, Kloverpris H, Stryhn A, Koofhethile CK, Sims S, Ndung'u T, Goulder P, Buus S, Nielsen M. HLArestrictor--a tool for patient-specific predictions of HLA restriction elements and optimal epitopes within peptides. Immunogenetics 2011;63:43-55.

18. Guan P, Doytchinova IA, Zygouri C, Flower DR. MHCPred: a server for quantitative prediction of peptide–MHC binding. Nucleic Acids Res. Oxford Univ Press; 2003;31:3621-3624.

19. Lundegaard C, Lamberth K, Harndahl M, Buus S, Lund O, Nielsen M. NetMHC-3.0: accurate web accessible predictions of human, mouse and monkey MHC class I affinities for peptides of length 8-11. Nucleic Acids Res 2008;36:W509-512.

20. Singh H, Raghava GPS. ProPred1: prediction of promiscuous MHC Class-I binding sites. Bioinformatics 2003;19:1009-1014.

21. 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;56:405-419.

22. Dönnes P, Kohlbacher O. SVMHC: a server for prediction of MHC-binding peptides. Nucleic Acids Res 2006;34:W194-197.

23. Xiang Z, He Y. Vaxign: a web-based vaccine target design program for reverse vaccinology. Procedia Vaccinol 2009;1:23-29.

24. Vita R, Overton JA, Greenbaum JA, Ponomarenko J, Clark JD, Cantrell JR, Wheeler DK, Gabbard JL, Hix D, Sette A, Peters B. The immune epitope database (IEDB) 3.0. Nucleic Acids Res 2014;43:D405-412.

25. Saha S, Raghava GPS. Prediction of continuous B‐cell epitopes in an antigen using recurrent neural network. Proteins Struct Funct Bioinforma 2006;65:40-48.

26. Doytchinova IA, Flower DR. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. BMC Bioinformatics. Bio Med Central 2007;8:4.

27. Kaur H, Garg A, Raghava GPS. PEPstr: a de novo method for tertiary structure prediction of small bioactive peptides. Protein Pept Lett 2007;14:626-631.

28. Comeau SR, Gatchell DW, Vajda S, Camacho CJ. ClusPro: an automated docking and discrimination method for the prediction of protein complexes. Bioinformatics 2004;20:45-50.

29. Singh H, Raghava GPS. ProPred: prediction of HLA-DR binding sites. Bioinformatics 2001;17:1236-1237.

30. Capecchi B, Serruto D, Adu-Bobie J, Rappuoli R, Pizza M. The genome revolution in vaccine research. Curr Issues Mol Biol 2004;6:17-28.

31. Yu NY, Wagner JR, Laird MR, Melli G, Rey S, Lo R, Dao P, Sahinalp SC, Ester M, Foster LJ, Brinkman FS. PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. Bioinformatics 2010;26:1608-1615.

32. Talay SR. Gram-Positive Adhesins. In: Concepts in bacterial virulence. Karger Publishers; 2005.p.90-113.

33. Saraav I, Pandey K, Sharma M, Singh S, Dutta P, Bhardwaj A, et al. Predicting promiscuous antigenic T cell epitopes of Mycobacterium tuberculosis mymA operon proteins binding to MHC Class I and Class II molecules. Infect Genet Evol 2016;44:182-189.

34. Perez-Casal J, Caparon MG, Scott JR. Introduction of the emm6 gene into an emm-deleted strain of Streptococcus pyogenes restores its ability to resist phagocytosis. Res Microbiol 1992;143:549-558.

35. Batzloff MR, Hayman WA, Davies MR, Zeng M, Pruksakorn S, Brandt ER, Good MF. Protection against group A streptococcus by immunization with J8-diphtheria toxoid: contribution of J8-and diphtheria toxoid-specific antibodies to protection. J Infect Dis 2003; 187:1598-1608.

36. Massell BF, Honikman LH, Amezcua J. Rheumatic fever following streptococcal vaccination: report of three cases. JAMA 1969;207:1115-1119.

37. Hu MC, Walls MA, Stroop SD, Reddish MA, Beall B, Dale JB. Immunogenicity of a 26-valent group A streptococcal vaccine. Infect Immun 2002;70:2171-2177.

38. Guilherme L, Faé KC, Higa F, Chaves L, Oshiro SE, Freschi de Barros S, Puschel C, Juliano MA, Tanaka AC, Spina G, Kalil J. Towards a vaccine against rheumatic fever. Clin Dev Immunol 2006;13:125-132.