

# *In Silico* Investigation for Evaluation of the Potential of the SclA Protein in *Streptococcus pyogenes*

Maryam Pourhajibagher<sup>1</sup>; Abbas Bahador<sup>1,\*</sup>

<sup>1</sup>Department of Microbiology, School of Medicine, Tehran University of Medical Sciences, Tehran, IR Iran

\*Corresponding author: Abbas Bahador, Department of Microbiology, School of Medicine, Tehran University of Medical Sciences, Tehran, IR Iran. Tel: +98-2188955810, E-mail: abahador@tums.ac.ir

Received: April 5, 2014; Revised: February 10, 2015; Accepted: May 18, 2015

**Background:** *Streptococcus pyogenes* is an important pathogen that is associated with a range of infections in humans, and causes common and severe invasive diseases. Currently, antimicrobial therapy is the first choice for the treatment of *S. pyogenes*; however, the emergence of antimicrobial resistance and side effects of antibacterial drugs is increasing. Consequently, there is an increased demand for novel drug targets and vaccine design.

**Objectives:** To develop an effective vaccine against *Streptococcus pyogenes* (group A streptococcus), we described a novel collagen-like surface protein of *S. pyogenes* which is important virulence factors

**Materials and Methods:** In this study, we focused on the SclA protein of *S. pyogenes* and characterized it using bioinformatic tools to introduce it as a candidate novel drug as a candidate for use in vaccine design. The secondary structure was determined and the 3D structure was modeled using SWISS-MODEL workspace. The immune epitope database analysis (IEDB) resource was used to predict regions of SclA that are likely to be recognized as epitopes.

**Results:** The SclA protein is present on the cell surface of the cell and has interact with a common ligand by its hypervariable NH<sub>2</sub>-terminal regions. The IEDB showed that the maximum peptide length that is likely to be predicted as an epitope is of 6 amino acids, from amino acid 26 to 31, with a score of 4.791. This epitope can be considered for use in Antibody and drug design.

**Conclusions:** Data from this study about SclA were not sufficient and further studies are needed; however, the information here suggests that SclA could be a candidate for further research on the design of drugs and vaccines against *S. pyogenes* infections.

**Keywords:** Computer Simulation; Streptococcal Collagen-Like Protein A; Vaccine Candidate; *Streptococcus pyogenes*

## 1. Background

*Streptococcus pyogenes* (group A streptococcus, GAS) is an important pathogen associated with a range of infections in humans (1). This organism causes common diseases such as pharyngitis, erysipelas, and impetigo, but sometimes can also cause severe invasive diseases such as toxic shock-like syndrome and necrotizing fasciitis (2, 3). Currently, antimicrobial therapy is the first choice for the treatment of GAS; however, antimicrobial resistance is a common problem worldwide and the consumption of antimicrobials has been identified as a driving force for the emergence and transmission of antimicrobial resistance in GAS (4). *Streptococcus pyogenes* has very complex virulence factor (5). The cell wall of *S. pyogenes* is associated with a range of proteins, a large number of which are encoded by the *emm* gene family: M proteins are the members that contain the most important structures for virulence (5-7).

Other cell wall proteins in *S. pyogenes* include streptococcal collagen-like surface protein (SclA) (8). SclA is transcriptionally regulated by Mga and has a highly conserved signal sequence; in a previous study, SclA was found in 12 *S. pyogenes* strains tested by PCR (9). Since the M protein is a major virulence factor of GAS, it could be a

candidate for use in vaccine design (10). However, previous studies have shown that some epitopes of M proteins can induce autoimmunity and cross-react with human tissues (9, 10), because of which they may not be suitable for use in vaccine design.

## 2. Objectives

In this study, we characterized the SclA protein of *S. pyogenes*, using online and off-line computational tools for the development of new vaccines to prevent and treat pyogenesis.

## 3. Materials and Methods

### 3.1. Retrieval of the Target Sequence

The candidate organism for the present study was *S. pyogenes*. The protein sequence of SclA (GenBank: CAC33784.1) was obtained from NCBI (<http://www.ncbi.nlm.nih.gov>), and a BLAST alignment was performed against Protein Data Bank (PDB) entries to find similar sequences.

### 3.2. Functional Characterization of SclA

Functional characterization of the SclA protein sequence was performed using Uniprot (<http://expasy.org>). Primary structure prediction was computed using the ExPASy ProtParam server (<http://expasy.org/cgi-bin/protpraram>).

### 3.3. Protein Topology Prediction and 3D Structure of SclA

The secondary structure of the SclA protein was obtained using the PSIPRED Protein Sequence Analysis Workbench (<http://bioinf.cs.ucl.ac.uk/psipred/>). Using the TMHMM data bank, we predicted transmembrane helices based on a hidden Markov model and predicted the locations of transmembrane intracellular and extracellular regions. Garnier-Osguthorpe-Robson (GOR) IV was used to predict alpha helices, beta sheets, and random coil secondary structures. The three-dimensional structure of SclA was obtained using <http://www.ncbi.nlm.nih.gov/Structure/VAST>.

### 3.4. B Cell Epitope Prediction

The immune epitope database analysis resource (IEDB-AR, <http://tools.immuneepitope.org>) provides a collection of web-based tools for the prediction and analysis of immune epitopes. In this study, IEDB was used to predict the regions of SclA that are likely to be recognized as epitopes in the context of a B response. Conservancy analysis for epitopes and the percentages of protein sequence matches at identity  $\geq 100\%$  were also calculated.

## 4. Results

*In silico* analysis can be used for identifying drug targets, designing vaccines, and choosing a rapid strategy for disease control. The sequence of the SclA protein was obtained from the NCBI protein sequence database and a BLAST search was performed against Protein Data Bank (PDB) entries for sequence similarity. SclA was found to be associated with 3HQV\_B, a protein from *Rattus norvegicus*. The score for protein alignment between SclA and 3HQV\_B was 30.0, with a 46% identity. PROSITE analysis showed that SclA was a surface-localized protein. Primary structure prediction showed that SclA had 123 amino acid residues and an estimated molecular weight of 13137.6. The total number of positively charged residues (Arg + Lys) was 22 and the total number of negatively charged residues (Asp + Glu) was 21. The very high aliphatic index (48.54) and instability index (36.76) provided an estimate of the stability of protein in a test tube. The Grand Average Hydropathicity (GRAVY) value was -0.956.

The secondary structure prediction showed that the extended strand (10.57%) and alpha helix (36.59%) were the least frequent, whereas the random coil was more frequent at 52.85%. As the 3D structure of SclA had not been

reported in Protein Data Bank (PDB), a BLASTp similarity search was performed against the PDB database and 3HQV\_B, where a complete query coverage was obtained (Figure 1). By using PSIPRED, we showed the location and spatial arrangement of each amino acid separately (Figure 2). TMHMM predicted the locations of transmembrane intracellular and extracellular regions and showed that the SclA protein was present on the outside of the cell (Figure 3). The results of the IEDB showed that the maximum length of a peptide that is likely to be predicted as an epitope was of 6 amino acids between positions 26 and 31, with an epitope sequence of QERTRK and a score of 4.791 (Figure 4).

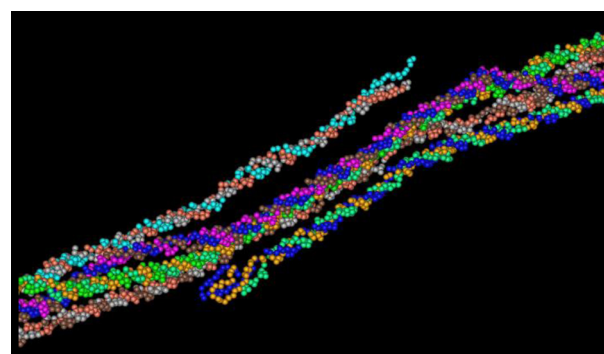


Figure 1. Modeled Spatial Configuration of the SclA Protein

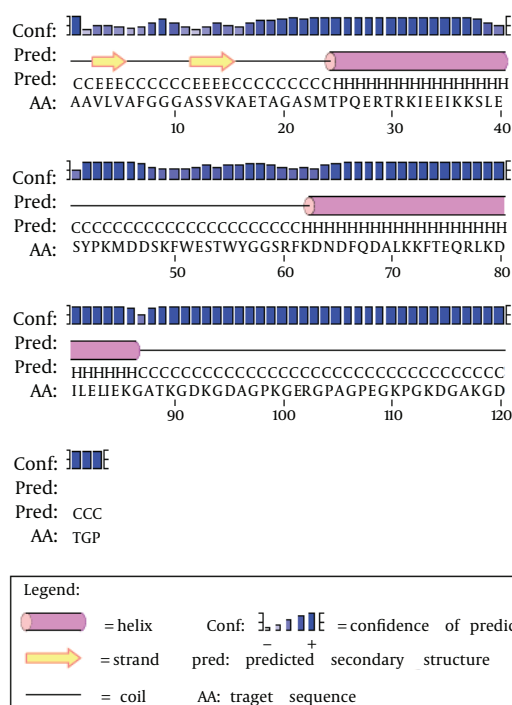


Figure 2. Secondary Structure Prediction of the SclA Protein of *Streptococcus pyogenes* using PSIPRED

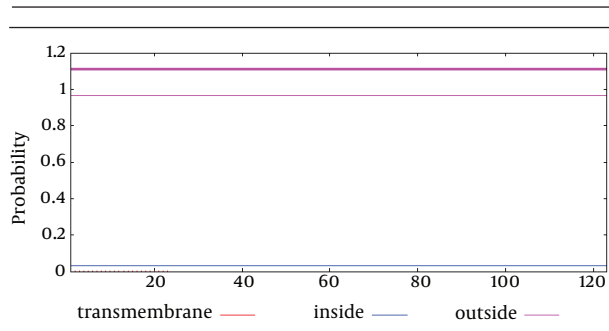


Figure 3. Graphical Representation of the Location of the SclA Protein

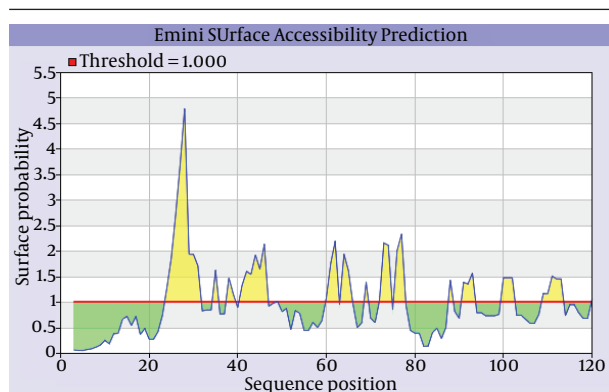


Figure 4. Emini Surface Accessibility Prediction

## 5. Discussion

*Streptococcus pyogenes* is one of the most pathogenic bacteria infecting children and adolescents. GAS is involved with a wide range of infections and diseases that are important causes of morbidity and mortality, primarily in developing countries, causing > 500,000 deaths annually (2, 11). It is therefore important to treat and prevent these diseases with drugs and vaccines. Currently, antimicrobial therapy is the first choice for the treatment of GAS; however, the emergence of antimicrobial resistance and side effects of antibacterial drugs is increasing (12), resulting in an increased demand for novel drug targets and vaccine design. On the surface of *S. pyogenes*, 79 surface-localized proteins were reconnoitered and each of them was identified in one growth phase. For example, the expression of SclA was consistent with a previously published report showing that SclA was transcribed in the mid-log phase (13).

The SclA protein contains a signal sequence, a cell wall anchor containing the LPXTGX motif, and proline-rich repeats in the COOH-terminal part; the hypervariable NH<sub>2</sub>-terminal regions of SclA show that conserved secondary structures interact with a common ligand and have common functions (14). For drug target and vaccine design against many bacterial pathogens, conserved protein sequences should be used. However, vaccine protection was demonstrated to be variable between different

strains of the same serotype and different genotypes (15). Unfortunately, an effective vaccine to prevent GAS infection is yet to be developed. Secreted or surface proteins or lipoproteins are favored candidates for vaccines because they are the most easily accessible microbial antigens to the immune system (14, 16). As SclA is a surface protein, it may be a suitable candidate for vaccine design (13).

The focus of the current study was therefore to characterize the SclA protein of *S. pyogenes* using computational tools. SclA had 123 amino acid residues with a molecular weight 13137.6. The total number of positively charged residues was more than that of negatively charged residues. The N-terminal sequence is considered to contain Alanin (Ala). Garnier-Osguthorpe-Robson (GOR) IV comparative analysis showed that a random coil is predominantly present, followed by an extended strand and an alpha helix. TMHMM determined that the SclA protein is present on the outside of the cell. The results of IEDB showed that the maximum length of the peptide that is likely to be predicted as an epitope is of 6 amino acids, from amino acid 26 to 31, with a score of 4.791 and a protein sequence identity match of 100.00% (1/1). These epitopes can be considered for antibody and drug design.

The information on the characterization of the SclA protein of *S. pyogenes* obtained from these studies alone is not sufficient. Nonetheless, this information suggests that SclA could be a candidate for more rigorous studies on drug and vaccine design against *S. pyogenes*.

## Acknowledgements

The authors would like to express their deep thanks to all staff of medical science and Dr. Masoud Tohidfar for help in the study.

## Authors' Contributions

Maryam Pourhajibagher developed the original idea and the protocol, abstracted and analyzed data, and drafted the manuscript. Abbas Bahador contributed to the development of the critical revision of the manuscript for important intellectual content, and is the guarantor.

## References

1. Barinov A, Loux V, Hammani A, Nicolas P, Langella P, Ehrlich D, et al. Prediction of surface exposed proteins in *Streptococcus pyogenes*, with a potential application to other Gram-positive bacteria. *Proteomics*. 2009;9(1):61-73.
2. Cole JN, Ramirez RD, Currie BJ, Cordwell SJ, Djordjevic SP, Walker MJ. Surface analyses and immune reactivities of major cell wall-associated proteins of group A streptococcus. *Infect Immun*. 2005;73(5):3137-46.
3. Carapetis JR, Steer AC, Mulholland EK, Weber M. The global burden of group A streptococcal diseases. *Lancet Infect Dis*. 2005;5(11):685-94.
4. Bergman M, Huikko S, Pihlajamaki M, Laippala P, Palva E, Huovinen P, et al. Effect of macrolide consumption on erythromycin resistance in *Streptococcus pyogenes* in Finland in 1997-2001. *Clin Infect Dis*. 2004;38(9):1251-6.
5. Chatellier S, Ihendyane N, Kansal RG, Khambaty F, Basma H, Norrby-Teglund A, et al. Genetic relatedness and superantigen

- expression in group A streptococcus serotype M1 isolates from patients with severe and nonsevere invasive diseases. *Infect Immun.* 2000;**68**(6):3523-34.
6. Facklam RF, Martin DR, Lovgren M, Johnson DR, Efstratiou A, Thompson TA, et al. Extension of the Lancefield classification for group A streptococci by addition of 22 new M protein gene sequence types from clinical isolates: emm103 to emm124. *Clin Infect Dis.* 2002;**34**(1):28-38.
  7. Caparon MG, Scott JR. Identification of a gene that regulates expression of M protein, the major virulence determinant of group A streptococci. *Proc Natl Acad Sci U S A.* 1987;**84**(23):8677-81.
  8. Rasmussen M, Eden A, Bjorck L. SclA, a novel collagen-like surface protein of *Streptococcus pyogenes*. *Infect Immun.* 2000;**68**(11):6370-7.
  9. Whatmore AM. *Streptococcus pyogenes* sclB encodes a putative hypervariable surface protein with a collagen-like repetitive structure. *Microbiology.* 2001;**147**(Pt 2):419-29.
  10. Factor SH, Levine OS, Harrison LH, Farley MM, McGeer A, Skoff T, et al. Risk factors for pediatric invasive group A streptococcal disease. *Emerg Infect Dis.* 2005;**11**(7):1062-6.
  11. Herwald H, Cramer H, Morgelin M, Russell W, Sollenberg U, Norrby-Teglund A, et al. M protein, a classical bacterial virulence determinant, forms complexes with fibrinogen that induce vascular leakage. *Cell.* 2004;**116**(3):367-79.
  12. Krucso B, Gacs M, Libisch B, Hunyadi ZV, Molnar K, Fuzi M, et al. Molecular characterisation of invasive *Streptococcus pyogenes* isolates from Hungary obtained in 2004 and 2005. *Eur J Clin Microbiol Infect Dis.* 2007;**26**(11):807-11.
  13. Lamagni TL, Darenberg J, Luca-Harari B, Siljander T, Efstratiou A, Henriques-Normark B, et al. Epidemiology of severe *Streptococcus pyogenes* disease in Europe. *J Clin Microbiol.* 2008;**46**(7):2359-67.
  14. Boekhorst J, de Been MW, Kleerebezem M, Siezen RJ. Genome-wide detection and analysis of cell wall-bound proteins with LPxTG-like sorting motifs. *J Bacteriol.* 2005;**187**(14):4928-34.
  15. Luca-Harari B, Darenberg J, Neal S, Siljander T, Strakova L, Tanna A, et al. Clinical and microbiological characteristics of severe *Streptococcus pyogenes* disease in Europe. *J Clin Microbiol.* 2009;**47**(4):1155-65.
  16. McMillan DJ, Geffers R, Buer J, Vlamincx BJ, Sriprakash KS, Chhatwal GS. Variations in the distribution of genes encoding virulence and extracellular proteins in group A streptococcus are largely restricted to 11 genomic loci. *Microbes Infect.* 2007;**9**(3):259-70.