Biologicals xxx (2017) 1–13

Contents lists available at ScienceDirect

Biologicals

journal homepage: www.elsevier.com/locate/biologicals

# Towards a peptide-based vaccine against *Shigella sonnei*: A subtractive reverse vaccinology based approach

Shehneela Baseer [a, 1], Sajjad Ahmad [a, 1], Kara E. Ranaghan [b], Syed Sikander Azam [a, *]

[a] Computational Biology Lab, National Center for Bioinformatics (NCB), Faculty of Biological Sciences, Quaid-i-Azam University, Islamabad, Pakistan
[b] Centre for Computational Chemistry, University of Bristol, Bristol, United Kingdom

ARTICLE INFO

*Article history:*
Received 22 April 2017
Received in revised form
11 July 2017
Accepted 8 August 2017
Available online xxx

*Keywords:*
*S. sonnei*
Vaccine
Epitope
TolC
PhoE
Outer membrane porin protein

ABSTRACT

*Shigella sonnei* is one of the major causes of shigellosis in technically advanced countries and reports of its unprecedented increase are published from the Middle East, Latin America, and Asia. The pathogen exhibits resistance against first and second line antibiotics which highlights the need for the development of an effective broad-spectrum vaccine. A computational based approach comprising subtractive reverse vaccinology was used for the identification of potential peptide-based vaccine candidates in the proteome of *S. sonnei* reference strain (53G). The protocol revealed three essential, host non-homologous, highly virulent, antigenic, conserved and adhesive vaccine proteins: TolC, PhoE, and outer membrane porin protein. The cellular interactome of these proteins supports their direct and indirect involvement in biologically significant pathways, essential for pathogen survival. Epitope mapping of these candidates reveals the presence of surface exposed 9-mer B-cell-derived T-cell epitopes of an antigenic, virulent, non-allergen nature and have broad-spectrum potency. In addition, molecular docking studies demonstrated the deep binding of the epitopes in the binding groove and the stability of the complex with the most common binding allele in the human population, DRB1*0101. Future characterization of the screened epitopes in order to further investigate the immune protection efficacy in animal models is highly desirable.


## 1. Introduction

Shigellosis, a severe and life-threatening diarrheal infection, is a major health concern in both industrialized and non-industrialized countries [1]. Annually, 91 million cases of shigellosis are reported worldwide [2] and it is associated with significant mortality and morbidity [3]. Shigellosis is placed 6th on the list of high mortality rate diseases in China [4] and is ranked 3rd in the United States among gastrointestinal diseases [5]. Annually, Shigellosis alone is responsible for 125 million infections and 14,000 deaths in Asia [3]. Almost 90.5% of shigellosis cases are caused by *Shigella sonnei* together with *Shigella flexneri*. In developed countries, *S. sonnei* is the most frequent pathogen responsible for shigellosis [6]; however, it has also been identified in cases in Asia, the Middle East, and Latin America. Outbreaks of shigellosis have been reported from several countries: Australia [6], Korea [7], Bangladesh [8] and Taiwan [9]. Shigella enterotoxin 1 (ShET-1), Shigella enterotoxin 2 (ShET-2) the invasion plasmid antigen H gene (IpaH) [10,11], and type 3 secretion system all contribute to the successful survival and pathogenesis of *S. sonnei* [12,13]. In the United States, *S. sonnei* is resistant to oral antibiotics such as trimethoprim, sulfamethoxazole, and ampicillin. Resistance to fluoroquinolone is also increasing progressively [14] with 2% of Shigella isolates in the US showing resistance [15]. *S. sonnei* is also resistant to ciprofloxacin [16], which is the first-line treatment remedy against shigellosis for adults [17].

Unfortunately, no licensed vaccine is available against *S. sonnei* infections [18]. The vaccine candidate protein against Shigella proposed by the National Institute of Child Health and Human Development (NICHHD) and Laboratory of Developmental and Molecular Immunology (LDMI), −O-SP-, induces poor immunogenic responses. In order to establish long lasting and strong T-cell immune response, NICHHD and LDMI conjugated the O-SP-protein covalently with a lipopolysaccharide (LPS) protein carrier. However, the O-SP antigen immunity is stereotype specific [19]. The Pasteur

* Corresponding author. National Center for Bioinformatics (NCB), Faculty of Biological Sciences, Quaid-i-Azam University, Islamabad 45320, Pakistan.
*E-mail addresses:* ssazam@qau.edu.pk, syedazam2008@gmail.com (S.S. Azam).
[1] Both the authors contributed equally to the work.

http://dx.doi.org/10.1016/j.biologicals.2017.08.004
1045-1056/

Institute designed a glycoconjugate vaccine comprised of a synthetic "mimic" of oligosaccharide from *S. flexineria2a* and is currently in clinical trials [19]. The Navarra University in Spain used outer membrane vesicles (OMVs) to develop an acellular vaccine which comprises 40% LPS and IpaB, IpaC, IpaD, OmpC/OmpF and OmpA major antigens. Despite these efforts, no vaccine based treatment is currently available against *S. sonnei* and human health will benefit from the identification of broad-spectrum vaccine candidates with improved immunogenic efficacy against *S. sonnei.*

As peptide-based vaccines are more specific and are easy to produce [20], we have focused our study on the identification of peptide vaccine candidates in the proteome of *S. sonnei.* With the recent developments in immunology, biochemistry, molecular biology, proteomics, and genomics, the field of conventional vaccinology has transformed into Reverse Vaccinology (RV) [21]. RV circumvents the hurdles of cost, time duration and accuracy associated with traditional vaccinology and has been applied successfully in designing a vaccine against serogroup B. *meningococcal* infections [22]. The RV protocol comprises *in silico* filters that prioritize proteins in the pathogen proteome with high probability as vaccine candidates. This methodology was applied to screen the proteome of *S. sonnei* for the identification of novel vaccine candidates. We strongly belive that the outcomes of this study will provide better guidance for future vaccine design and development against *S. sonnei*.

## 2. Material and methods

### 2.1. Proteome subtraction

The complete proteome of *S. sonnei* (reference strain 53G) [23] was retrieved from the Genome database available at the National Center for Biotechnology Information (NCBI) and characterized for redundancy through the CD-Hit web server (http://weizhongli-lab.org/cdhit_suite/cgi-bin/index.cgi?cmd=cd-hit) [24]. Here, a non-redundant set of proteins was retrieved by eliminating paralogous sequences from the proteome sharing 80% sequence identity. The non-redundant protein sequences were then used in a BLASTP search against the DEG database (http://tubic.tju.edu.cn/deg/) [25] using a Perl script provided by Computational Biology Lab at the National Center for Bioinformatics. Proteins with an E-value cut-off ($10^{-4}$), sequence identity $\geq$ 30% and bit score $\geq$ 100 were picked out as essential proteins. Screening of essential proteins is important as such proteins are necessary for pathogen survival and their deletion can lead to cell growth arrest, therefore, can be attractive vaccine targets [26]. To filter the host non-homologous proteins, the pathogen essential proteins were aligned against the human proteome (*Homo sapien*, taxonomic ID: 9606) (http://blast.ncbi.nlm.nih.gov/Blast.cgi/). Proteins with sequences identity $\leq$35% and an E-value cut-off of $10^{-4}$ were chosen as human non-homologs. Removal of host homologous sequences is crucial as such proteins generate cross-reactivity with the host proteins giving rise to adverse autoimmune responses [27]. The final filter of the subtractive proteomics process was to deduce proteins which are surface exposed and interact with biotic and abiotic factors of the extracellular environment [28]. The remainder of the proteome was examined for pathogen exoproteome and secretome through PSORTb (http://db.psort.org/) [29] and CELLO (http://cello.life.nctu.edu.tw/). Additionally, those recognized as extracellular and outer membranous were cross-checked by CELLO2GO (http://cello.life.nctu.edu.tw/cello2go/) [30] to achieve consistency in the results.

### 2.2. Virulent proteome evaluation

Virulent proteins mediate severe infection pathways in the host, more efficiently leading to disease compared to non-virulent proteins, and are thus suitable candidates for vaccine development. In this context, the exoproteome and secretome of the pathogen were used in a BLASTP search against the Virulence Factor Database (VFDB) (http://www.mgc.ac.cn/VFs/main.htm) [31] to screen proteins with an identity of $\geq$50% and bit score $\geq$ 100.

### 2.3. Physicochemical characterization

The virulent proteins were physicochemically characterized to identify experimentally suitable proteins. Three key parameters were considered: molecular weight, number of transmembrane helices and adhesion-like properties. First, the molecular weight of each protein was determined using the Expasy server [32]. Proteins weighing <110 kDa were considered due to their easy purification in subsequent wet lab analysis [33]. Computation of transmembrane helices was performed by HMMTOP (http://www.enzim.hu/hmmtop) and TMHMM (http://www.cbs.dtu.dk/services/TMHMM-2.0). Proteins with no more than 1 transmembrane helix were collected as such proteins can be cloned and expressed efficiently [34,35]. The adhesion probability of the shortlisted proteins was then investigated using SPAAN (ftp://203.195.151.45) [36]. Adhesive proteins aid in bacterial adherence, colonization to host tissues and subsequent infection, and are therefore, valuable candidates in vaccine development [37]. Proteins with an adhesion probability greater than 0.5 were selected from this analysis.

### 2.4. Epitopes mapping

Predicting epitopes with the potential to stimulate both B and T-cell immunity is imperative for epitope-based vaccine development [35]. The VaxiJen [38] server was used to predict the antigenic nature of the selected proteins. This method uses an approach based on auto cross covariance transformation of protein sequences into uniform vectors of principal amino acid properties. Proteins with antigenicity value greater than 0.4 were categorized as antigenic and analyzed for B-cell epitopes. Prediction of B-cell epitopes was done by accessing BCPred (http://ailab.ist.psu.edu/bcpred/predict.html) [39] with epitope length set to a 20-mer. The predicted epitopes were then examined for exposed topology via TMHMM. The exposed B-cell epitopes were finally tested for T-cell immunity. Binding alleles for both classes of MHC i.e. MHC-I and MHC-II were determined by Prophred I (http://www.imtech.res.in/raghava/propred1/) [40] and Prophred (http://www.imtech.res.in/raghava/propred/) respectively [41]. T-cell epitopes which bind to more than 15 alleles of MHC- I and II were selected and analyzed further for common binding alleles [34]. Using MHCpred, the $IC_{50}$ value for common T-cell epitopes was determined with a threshold of <100 nM. To ensure the virulent and antigenic nature of the T-cell epitopes, VirulentPred and VaxiJen analyses were performed, respectively. The 20-mer B-cell epitopes from each prioritized protein, for which a T-cell epitope was successfully predicted, were characterized further using the IBED server for flexibility, hydrophilicity, surface accessibility, antigenicity, and Beta turns [42]. The results from this analysis was cross-referred with that of T-cell epitopes and a consensus peptide was determined.

Please cite this article in press as: Baseer S, et al., Towards a peptide-based vaccine against *Shigella sonnei*: A subtractive reverse vaccinology based approach, Biologicals (2017), http://dx.doi.org/10.1016/j.biologicals.2017.08.004