Baseer, S., Ahmad, S., Ranaghan, K. E., & Azam, S. S. (2017). Towards a peptide-based vaccine against Shigella sonnei: A subtractive reverse vaccinology based approach. *Biologicals*. https://doi.org/10.1016/j.biologicals.2017.08.004

Title Page

Article title

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

Author list

Shehneela Baseer[1†], Sajjad Ahmad[1†], Kara E. Ranaghan[2], Syed Sikander Azam[1]*

[†] Both the authors contributed equally to the work.

Addresses

[1] Computational Biology Lab, National Center for Bioinformatics (NCB), Faculty of Biological Sciences, Quaid-i-Azam University, Islamabad, Pakistan

[2] Centre for Computational Chemistry, University of Bristol, Bristol, United Kingdom.

***Corresponding author:**

Syed Sikander Azam,
National Center for Bioinformatics (NCB),
Faculty of Biological Sciences,
Quaid-i-Azam University, Islamabad-45320, Pakistan.
Tel: 0092-51-90644130

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.

Keywords

S. sonnei; Vaccine; Epitope; TolC; PhoE; Outer membrane porin protein

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 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 methodolodgy 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 ≥ 30% and bit score ≥ 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 ≤ 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 ≥ 50% and bit score ≥ 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.

2.5. Epitope conservation

The development of a peptide-based, broad-spectrum vaccine truly requires the conservation of the selected antigenic peptide across all completely annotated strains of the pathogen. To achieve this, prioritized proteins from *S. sonnei* strains were retrieved and aligned through the CLC sequence viewer to examine epitope conservancy [43].

2.6. Epitope allergenicity

An increased number of vaccines are now recognized to cause allergic reactions [44]. To avoid this, the conserved epitopes were scanned for allergenicity through SORTALLER (http://sortaller.gzhmu.edu.cn/) which predicts sequence allergenicity with specificity and sensitivity of 98.4% and 98.6%, respectively [45]. The epitopes were cross-checked by AllerTop (http://www.ddg-pharmfac.net/AllerTOP/) to achieve consistency in results [46].

2.7. Interacting network

The cellular interactome of the prioritized proteins was analyzed using Search Tool for the Retrieval of Interacting Genes (STRING) (http://string-db.org/) [47] for both direct and indirect interactions. Knowledge of the interaction network of the selected proteins is important to understand the inhibitory impact of these proteins on pathogen survival [48]. The goal of the STRING database is to assemble, evaluate and disseminate protein–protein association information, in a user-friendly and complete manner. The database contains information from several sources, including computational predictions, experimental data and community text groups.

2.8. Target proteins validation

Two vaccine protein prediction tools were used to validate the target proteins identified by the protocol: VacSol [49] and VaxiGen (http://www.violinet.org/vaxign/) [50]. Both the servers evaluated the shortlisted proteins for homology with human proteome, essentiality for pathogen viability, surface exposure, virulence, antigenicity and adhesive properties.

2.9. Structure prediction

The 3D structure of the epitope proteins was needed for the next stages of analysis [51]. Proteins structure prediction was done to view topology of predicted epitopes. Further it aids in visualizing and understanding the structure and its relationship with other vital cellular proteins [35]. A BLASTP search was carried out against the Protein Databank (PDB) to retrieve proteins with resolved 3D structures or template structures for homology modelling for those proteins with undetermined structures. Five structure prediction tools were used for comparative structure prediction: Modeller [52], Phyre2 [53], Swiss-Model [54], RaptorX [55] and M4T [56]. The quality of the predicted structures was evaluated based on the scores given by the Verify-3D [57], ERRAT [58] and Z-score of ProSA servers [59]. Quality evaluation was important as proteins with complete residues mapped are ideal for epitope topology analysis and molecular docking studies [34-35]. The best structure for each protein was selected and minimized using UCSF Chimera for a total of 750 steps under Tripos force field (TFF). Energy minimization of the predicted structures was carried out to improve the stability of the structures and remove steric hindrances.

2.10. Pepitope analysis

The exomembrane topology of the screened epitopes was viewed using the pepitope server [60-61]. Epitopes with an exposed surface are favourable as they are easily recognised by the host immune system producing a specific and accurate response.

2.11. Molecular docking of the epitopes

Molecular docking of the selected epitopes with the most common binding allele in the human population (DRB1*0101) was carried out to provide structural insight into the proposed protein-

peptide complexes. The crystal structure of the allele was retrieved from the PDB (PDB Id, 1AQD). GalaxyPepDock [62], an on-line server for protein-peptide docking, was used to perform the docking. GalaxyPepDock implements a similarity-based docking scheme. It identifies templates from experimentally resolved structure databases to predict the protein structure, followed by an energy-based optimization to provide structural flexibility. The selection of complex was based on the protein structure similarity score, interaction similarity score, and estimated accuracy. Furthermore, numbers of hydrogen bonding was another important consideration. Hydrogen bonding play a significant role in specifying interaction between peptide and receptor and aid in stability of peptide in protein active site. The best-docked complex was visualized through UCSF Chimera [63] and LigPlot [64].

3. Results and Discussion

3.1. Differential proteome mining

Differential proteome mining is an approach designed to examine the complete pathogen proteome in a stepwise manner. The approach was applied in the proposed framework in combination with RV for progressive subtraction of the pathogen proteome and antigenic epitope mapping of potential peptide-based vaccine candidates against *S. sonnei* (**Fig. 1**). The number of proteins brought forward in each step of the subtractive proteomic screening process is outlined in **Fig. 2**. The proteome of the *S. sonnei* reference strain 53G encompasses a sum of 4573 proteins and was retrieved from Genome database at NCBI. The proteome was filtered through the CD-Hit server to remove paralogous sequences with an identity of 80%. Redundant proteins are paralogous sequences and emerge because of duplication events. As such sequences do not have any influence on the organism's survival, it is appropriate to remove such proteins from the core proteome. It was revealed by CD-Hit that the non-redundant proteome of the pathogen contains 4386 (96 %) proteins and these proteins were passed on to the next stage of the protocol. A BLASTP search of the non-redundant proteins using the DEG server revealed which of the 4386 proteins are essential for *S. sonnei* survival. The search selected 1295 proteins as essential for *S. sonnei* survival, removing 3091 non-essential proteins from further analysis. The specificity filter against the human proteome using a BLASTP search revealed no significant hits or identity < 35% for 617 proteins, subsequently termed as human-non-homologs. Proteins homologous to the host cause an autoimmune response because of the cross-reactivity of the target and host proteins [65] and it is important to minimize this risk. Subcellular localization was done for extracellular and outer membrane proteins, which constitute the exoproteome and secretome of the pathogen respectively. Targeting these proteins is important because of their association with pathogenicity, aiding in pathogen adherence, invasion, host tissue proliferation and ultimately in successful survival. The dataset of essential proteins was examined for their subcellular localization using three tools (see Methods). Proteins that were localized in the outer membrane or extracellular matrix using all 3 tools were selected. PsortB showed 26 outer membranes, 5 extracellular, 185 unknown, 41 periplasmic, 364 inner membrane and 674 cytoplasmic proteins (**Fig.3**). Cello demonstrated 62 proteins as outer membrane, 14 extracellular, 1 unknown, 163 periplasmic, 743 cytoplasmic and 313 inner membranes (**Fig. 3**). Similarly, Cello2Go revealed 26 extracellular, 47 outer membrane, 96 periplasmic, 514 inner membrane and 873 cytoplasmic proteins (**Fig. 3**). By comparative analysis, out of 617 essential and non-homologous proteins, 55 were outer membrane and extracellular, thus selected for further study**.** Targeting the exoproteome and secretome of *S. sonnei* is vital in the search for vaccine candidates as surface proteins are in frequent contact with the host environment and contribute to a number of pathogenic activities of the bacterial pathogen.

3.2. Screening *S. sonnei* virulent proteins

Virulent proteins allow bacterial pathogens to overcome host immune responses, helping them to survive the demanding and competitive environment. Proteins identified by the differential

proteome mining were used in a BLASTP search against the VFDB to select virulent proteins. The blast screen identified fourteen proteins: Cation transporter, E3 ubiquitin--protein ligase, outer membrane protein (TolC), major curlin subunit protein, Outer membrane protein A (OmpA), fimbrial protein (LpfC), Putative outer membrane porin protein (nmpC), Porin protein, Phophoporin protein E (PhoE), Outer membrane protein E, Outer membrane protein C (OmpC), Outer membrane protein F (OmpF), membrane protein (nmpC) and rhsA protein (**Table 1)**. Cation transporter is a transporter protein responsible for the transportation of small molecules across the cell and between the cells. It also enhances bacterial resistance by acting as an efflux pump, thus lowering the drug concentration [66]. E3 ubiquitin--protein ligase is involved in altering cell physiology and prompts bacterial survival in host tissues. The protein also interferes with the host ubiquitin pathway, modulating host inflammatory responses and facilitating bacterial colonization within host T-cells [67]. TolC is also an outer membrane transporter and directs movement of small molecules in and out of the cell [68], particularly small toxic molecules. In the past few years, TolC efflux substrates have attracted substantial interest as a platform for designing drugs against TolC [69], highlighting the role of TolC in the increasing problem of multidrug resistance. In the outer membrane of gram-negative bacteria, nmpC, PhoE, OmpC, OmpF, and Porin proteins are arranged in a manner to form ion-selective channels for the passage of small hydrophilic molecules (up to ~600 D) [70-71]. The combination of these proteins then directs the movement of charged atoms or small charged molecules into and out of the cell or between the cells [72]. The two component assembly and transport system in gram-negative bacteria gives rise to the formation of fimbriae and pili, both of which contribute significantly to the pathogenicity of the host by aiding in its adherence to biotic and abiotic surfaces [73-75]. The majority of molecule and ion movements occurs through this transporter system into, out of, or within a cell, or between cells. OmpA is supposed to perform a key role, along with other bacterial constituents, in the structural reliability of the outer membrane and is considered an important factor in the virulence of numerous human pathogens [76]. Major curlin subunit proteins are thin, aggregative curli fibers necessary for adhesion. This protein binds to laminin, plasminogen, fibronectin, major histocompatibility complex (MHC) class I molecules and human contact phase proteins. Curli fibers are encoded by the Csg gene cluster and are assembled in a unique fashion in the extracellular matrix [77]. Rhs proteins of gram-negative bacteria are weakly linked with wall-associated protein A (WapA) from gram-positive bacteria. Rhs toxin C-terminal domains play a critical role in the inhibition of neighbouring cells and also encodes diverse immunity proteins through which cognate toxins are neutralized [78].

3.3. Vaccine candidate prioritization

As the later stages of vaccine development will involve many wet lab experiments, the physiochemical properties of the proteins are also an important factor to consider during the screening. Including a filter based on these properties at the *in silico* screening stage will save time and resources further in the development process. Molecular weight is one of the most important parameters for proteins in the context of vaccine development. Proteins less than 110 kDa in weight

were selected as they can be easily purified during the subsequent validation process. It was found that the molecular weight of 13 proteins turned out to be < 110 kDa, while the remaining protein, RhsA, was excluded due to its higher molecular weight. The molecular weight of the proteins was in the following order; nmpC (13.85 kDa), major curlin subunit protein (15 kDa), OmpA (37.21 kDa), PhoE (38.93 kDa), OmpF (39.35 kDa), nmpC (39.63 kDa), OmpC (40.52 kDa), porin protein (41.03 kDa), Outer membrane protein E (41.27 kDa), cation transporter (50.7 kDa), TolC (53.76 kDa), E3 ubiquitin--protein ligase (64.92 kDa) and LpfC (92.74 kDa). The low molecular weight proteins were examined further to identify their transmembrane topology. Proteins with only one or two transmembrane helices are desirable because of their easy cloning and expression [79]. Out of 13 proteins, 9 proteins (Cation transporter, TolC, major curlin subunit, nmpC, PhoE, Outer membrane protein E, OmpC, nmpC and Porin protein) contain no more than 1 transmembrane helix and were selected for the next stage of analysis. Adhesive proteins are involved in the initial stages of bacterial infection: adherence to the host-cell followed by colonization and finally by infection. In view of the significance of adhesive proteins in bacterial pathogenesis, blocking such proteins could provide a vital means to prevent bacterial infection. Adhesion probability analysis of the 9 proteins revealed 8 proteins as adhesive with values greater > 0.5: TolC (0.677), major curlin subunit (0.71), nmpC (0.614), PhoE (0.583), Outer membrane protein E (0.582), OmpC (0.544), nmpC (0.502) and Porin protein (0.522). The cation transporter protein was predicted to have an adhesion probability of 0.354 and was not included further analysis. The final prioritized set of proteins suitable for epitope mapping is summarized in **Fig. 4**.

3.4. Epitope Mapping

Screening of proteins with the potential of provoking and binding to adaptive humoral and cell-mediated immunity products is a crucial step in the vaccine development process. The prioritized set of proteins were first analyzed for their antigenic potential in epitope mapping phase through VaxiJen. VaxiJen analysis confirmed the antigenic nature of six proteins with an antigenic score of > 0.4. The antigenic score of the proteins were in the following order: TolC (0.5434), PhoE (0.7480), Outer membrane protein E (0.7418), OmpC (0.7081), nmpC (0.7081) and porin protein (0.6526). The remaining 2 proteins, major curlin subunit (0.3451) and nmpC (0.2345) were excluded due to their non-antigenic nature. The antigenic analysis was followed by epitope mapping where antigenic determinants with the ability of binding to the B-cell of the human immune system were determined. B-cell epitopes with the length of a 20mer and cut-off score > 0.8 were identified for each antigenic protein. B-cell epitopes of each antigenic protein were further refined based on the exposed topology.

In order to design B-cell-derived T-cell epitopes, the selected B-cell epitope of each protein was evaluated for their binding with MHC class I and MHC class II molecules [80]. T-cell epitopes which bind to maximum numbers of MHC-I and MHC-II molecules were considered. Peptides “YQGGMVNSQ” and “MVNSQVKQA” from TolC were selected as they share a total of 13 and 33 of MHC-I and MHC-II alleles, respectively. Similarly, the “WGLSTTYDL” peptide in case PhoE protein was considered as it shares 34 alleles in both classes. In case of porin protein, the

"FGISSTYVY" peptide which binds to 25 alleles was preferred. Lastly, for Outer membrane protein E, the "YVLSKGKDI" and "VLSKGKDIE" peptides which bind to 31 and 8 MHC alleles, respectively, were selected. The affinity of the T-cell epitopes for the DRB1*0101 allele was determined using MHCPred. DRB1*0101 is the most common allele in the human population and can produce accurate, specific and profound antigenic responses. Those epitopes with IC_{50} values lower than 100 nm were considered. Protein peptides "MVNSQVKQA", "WGLSTTYDL", "FGISSTYVY" and "YVLSKGKDI" were considered for further studies as they have values less than the cut-off i.e. 49.09, 37.33, 17.50 and 17.86 respectively. Peptides with higher IC_{50} values like "YQGGMVNSQ" and "VLSKGKDIE" were excluded from the pipeline because their binding affinity was very low. Antigenic B-cell-derived T-cell epitopes with $IC_{50} < 100$ nm were further analyzed for virulence and antigenicity through VirlentPred and VaxiJen. The epitope of TolC protein 'MVNSQVKQA' was chosen because of its binding with 33 MHC molecules (3 MHC-I and 30 MHC-II), low IC_{50} for DRB1*0101 and high virulent and antigenic score (1.05 and 1.16). Similarly, the 'WGLSTTYDL' epitope from PhoE was considered; (12 MHC-I and 22 MHC-II, total 34 MHC molecule binding, IC_{50} value, 37.33, VirulentPred value, 1.0595 and antigenic value of 0.8486). In the case of putative outer membrane porin protein (nmpC), the peptide 'FGISSTYVY" was selected. The epitope binds to 11-MHC-I and 14-MHC-II alleles (Total-25 MHC molecules) and has an IC_{50} value of 17.50 nm, VirulentPred score of 0.9892 and antigenicity Score of 0.8995 (**Table 2**). The remaining epitopes were not considered as they do not fulfill the criteria. The screened epitopes were favored further by the IBED server. The epitopes were found to have higher accessibility, hydrophobicity, flexibility and antigenic score (**S-Table 1**).

3.5. Epitope allergenicity and conservation

Vaccine antigens, like drugs, have the potential to cause allergic reactions. Due to increased vaccination practices, even mild reactions can lead to severe complications [81]. It is important to identify any possible allergenicity problems at an early stage in the vaccine design process. All three epitopes were identified as non-allergens by SORTALLER with the following allergenic scores: 'MVNSQVKQA' (0.263), "WGLSTTYDL" (0.263) and "FGISSTYVY" (0.263). The epitopes were cross-checked by AllerTop and were found non-allergic. The conservation of these epitopes was then assessed to aid the design of a broad-spectrum vaccine. As the pathogen has different strains, designing an effective vaccine against all the strains is desirable as through one antigen we can lock the survival potency of all the pathogen strains. For this, the sequence of each epitope protein was retrieved from the pathogen proteome and aligned through the CLC sequence viewer. It was observed that both the proteins and epitopes were highly conserved among all the four complete sequenced strains of the pathogen thus can act as attractive targets for peptide-based vaccine designing (**Fig. 5**). Furthermore, conservation of the shortlisted epitopes was investigated among other serogroups of Shigella. It was revealed that the shortlisted are also conserved in different serogroups. The porin protein epitope is conserved *S. flexneri* (50 % conservation), PhoE

epitope is conserved in *S. boydii* (50 % conservation) while TolC epitope is conserved *S. flexneri* and *S. boydii* (75% conservation). The peptides can be used alone in combination with next generation adjuvants to increase its immunogenicity. Furthermore, the peptides can be used as fusion peptides for enhancing antigenicity.

3.6. Target proteins validation

To ensure the vaccine like nature of these proteins, we validate these proteins using two tools based on the RV principle. Through both tools, we concluded that our selected three proteins completely exhibit all the ingredients of ideal vaccine proteins for subsequent wet lab studies. It was found through VaxiGen analysis that Porin protein resides in the outer membrane matrix have an adhesion probability value of 0.52 and transmembrane helices. Similarly, PhoE and TolC proteins were also found to be outer membranous with adhesion probability values of 0.58 and 0.67 respectively. The number of transmembrane helices was 0 in both proteins. According to VacSol, the proteins were regarded as human non-homologous, essential for the pathogen survival, highly virulent and physicochemical friendly.

3.7. Interaction network of TolC, PhoE and Porin protein

Investigating interaction maps for the selected proteins at the cellular level can identify many significant interactions with proteins crucial for pathogen survival, as well as among therapeutic targets. Therefore, it was important to consider the inhibitory impact of these proteins on pathogen survival. All the three proteins were observed to interact with essential proteins of the organism through direct and indirect pathways, all of which are necessary for *S. sonnei* survival and pathogenesis. Specifically, TolC protein was found to directly connect with PhoE protein and indirectly with porin protein. In addition to interactions with each other, these 3 proteins also revealed interactions with several important *S. sonnei* proteins involved in vital metabolic pathways of the pathogen. TolC showed interactions with Mdtc, emrK, acrB, acrD, acrA, macB, macA, OmpA, YgiB and SKp. The function of proteins Mdtc, acrA, acrB, emrK and acrD is not fully understood; however, it has been suggested that they are involved in lipid transport and multidrug efflux systems. MacA and MacB are part of the tripartite MacAB-TolC efflux system. MacA motivates the ATPase activity of MacB by helping the closed MacB ATP-bound state. MacB is a non-canonical ABC transporter that covers transmembrane domains, which form an ATP-binding domain responsible for energy generation and a pore in the inner membrane [82]. Similarly, PhoE and porin protein (S1887) showed interaction with OmpA and YgiB protein (**Fig. 6**).

Efflux pumps play a major role in drug extrusion and antibiotic resistance in gram-negative bacteria [83-84]. The TolC-dependent efflux system is not only responsible for the exclusion of toxic compounds, but also important for the transport of intracellular metabolites like porphyrin, excess cysteine and enterobactin [85-86]. AcrAD-tolC and mdtABC-tolC efflux pumps are documented for their involvement in bile salt resistance [87]. AcrAB-tolC efflux plays a major

role in the multidrug resistance of *S. sonnei* [88]. EmrKY operons are important for multidrug resistance in a drug-hyper susceptible strain which lacks the constitutive multidrug efflux pump genes acrA and acrB [89]. OmpA is one of the abundant outer membrane proteins of bacterial cells and has a role in the opening and closing of the diffusion channel [90]. The host immune system primarily targets OmpA during infection due to their involvement in the invasion of new born meningitis of epithelial cells [91-92]. OmpA also functions as a colicin and phage receptor [93].

3.7. Structure prediction and pepitope analysis

The structures of the final screened epitopes that fulfill all the parameters of being a potential candidate for peptide-based vaccine development were predicted to provide more insight into their exomembrane topology. The 3D structures of all the three proteins were not present in PDB, therefore, a comparative structure prediction approach was applied. For TolC, the model generated by Phyre (**Fig. 7**) was considered to be superior to the structures generated by other tools. Based on the structure evaluation analysis, the structure contains the highest number of residues (428) mapped to the most favorable region, 6 were in the generously allowed region, 13 in additional allowed regions and 3 in disallowed regions. The structure was also found to have modest ERRAT (78.17), Verify 3D (53.75 %) and ProsA (-6.8) values (**Fig. 8**). For PhoE protein, a structure predicted by Modeller (**Fig. 7**) was selected as the majority of the residues (280) were positioned in most favored regions, 39 in additional allowed regions, 8 in the generously allowed region and 1 in disallowed region. The ERRAT, Verify 3D and ProsA scores were in the following order: 26.301, 78.51 and -1.73 (**Fig. 8**). The Swiss model structure for Porin protein (**Fig. 7**) was the highest scoring with 277 residues in the most favored regions, 36 in additional allowed regions, 6 in the generously allowed region and 2 in disallowed region. The ERRAT, Verify 3D and ProsA values for the structure were 86.7, 74.20 and -3.6, respectively (**Fig. 8** and **S-Table. 2**). All the epitopes were found to have a surface exposed topology and not a globular protein structure (**Fig. 9**). Surface expression of the epitopes is important for the recognition of epitopes by the host immune system which results in a strong immune response. The exposed topology of the 9-mer epitopes of all the three proteins suggests their greater potency should be explored by experimental vaccinology in animal models.

3.8. Epitope docking

Molecular docking was carried out to interpret the binding mode and interactions of the epitopes in the binding pocket of the DRB*0101 allele. The best complex provided for each protein was visualized using Chimera and Ligplot. The epitopes tend to bind deeply in the binding groove of the allele and formed stable complexes. The epitope of TolC protein “MVNSQVKQA” binds to the binding pocket formed by chain A and chain B of the DRB1*0101 allele with an accuracy of 0.95 Similarly, the PhoE epitope “WGLSTTYDL” favored binding into the binding pocket of chain A and B of the DRB1*0101 allele with an accuracy of 0.955. The Porin protein epitope “FGISSTYVY” also tends to bind in the binding pocket of chain A and chain B of the DRB*0101 allele with an accuracy score of 0.85. Receptor protein residues involved in hydrogen bonding

with the ‘WGLSTTYDL’ epitope of the TolC protein include Gln7, Asn60, Leu243, Trp237, Asn67, Arg247, Asn258 and Ser51. Hydrophilic interactions were observed with the following residues: Ile436, Lys439, Ala432, Val429, Leu607, Glu375, Gln373, Tyr618, Phe418, Phe553, Trp601, Leu551, Asp597 and Tyr600 (**Fig. 10**). Similarly, with the PhoE epitope ‘MVNSQVKQA’, the following residues were involved in hydrogen bonding: Trp237, Asn67, Arg247, Asn258, Ser51, Gln7, Asn60, and Leu243. The hydrophobic residues found in contact with the epitopes were Ile70, Met71, Val63, Asp233, Asp64, Phe189, Glu9, Phe20, Leu187, Phe52, Gly262, Ile5, Phe22, Val261, Ile29, Phe265, Ala50, Tyr254, His257, Gly56, Gln246 and Tyr223 (**Fig. 11**). In the case of the ‘FGISSTYVY’ epitope of putative Porin protein, Ser51, Asn60, Glu9, Asn67, Ser213, Gln246, Arg247, Gln7 and Asn258 of DRB1*0101 are involved in hydrogen bonding, while hydrophobic interactions were observed for Ile70, Asp233, Val63, Trp237, Tyr223, Leu243, Tyr254, Phe52, His257, Phe30, Phe22, Phe265, Glu53, Gly56, Phe189, Leu187, Val214, Cys206, Met71 and Trp185 (**Fig. 12**).

Conclusion

The current work successfully identified novel peptides of virulent, essential and antigenic proteins, which could evoke substantial immune responses, thus can act highly attractive targets for peptide vaccine development against *S. sonnei*. All three proteins (TolC, PhoE, and outer membrane porin protein) are completely conserved in all four completely annotated strains of *S. sonnei* and interact directly and indirectly with several vital proteins of the pathogen crucial for survival. The methodology employed in the current work could provide an attractive alternative approach to tackle the dissemination of *S. sonnei* resistance strains and in providing vaccine-based treatment. We also suggest future work for characterizing these targets in animal models for their role in protecting the host against *S. sonnei* virulence.

Supplementary Files

Supplementary Table.1. IBED server characterization of 20mer B-cell epitopes from each prioritized vaccine protein.

Supplementary Table.2. Stereochemical properties evaluation of predicted structures of prioritized 3 vaccine proteins.

Conflict of Interest

The authors declare that they have no conflict of interest.

Funding Information

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Acknowledgment

Authors are highly grateful to International Foundation for Science (IFS) under grant number 5546-1 and Higher Education Commission (HEC), Pakistan for granting the financial assistance.

Figure Captions

Fig. 1. The designed framework used in the current study for peptide-based vaccine proteins mining in the proteome of *S. sonnei***.**

Fig. 2. Number of proteins brought forward in each step of the subtractive proteomics process.

Fig. 3. Quantitative representation of subcellular localization of essential proteins from PsortB, CELLO and CELLO2GO.

Fig. 4. AA) represent virulent proteins analyzed from VFDB, AB) indicates proteins having weight < 110 kDa, AC) represent proteins that have transmembrane helices < 1, AD) demonstrated adhesion probability of proteins while AE) represents 3 putative vaccine candidates.

Fig. 5. Conservation of the screened epitopes in all completely sequenced strains of *S. sonnei.*

Fig. 6. Protein- protein interactions of the three prioritized putative vaccine proteins.

Fig. 7. Predicted optimal structures for the prioritized three putative vaccine proteins.

Fig. 8. Structural quality evaluation of predicted structures of the three prioritized putative vaccine proteins.

Fig: 9. A topological view of selected epitopes on the surface of the protein.

Fig. 10. The binding mode and interactions of the TolC epitope in the binding cavity of the DRB*0101 allele.

Fig. 11. The binding mode and interactions of the PhoE epitope in the binding cavity of the DRB*0101 allele.

Fig. 12. The binding mode and interactions of the Porin protein epitope in the binding cavity of the DRB*0101 allele.

References

1. Jha P, Chaloupka FJ, Moore J, Gajalakshmi V, Gupta PC, Peck R, Jamison DT, Breman JG, Measham AR, Alleyne G, Claeson M. Disease control priorities in developing countries. 2nd Ed. New York: Oxford University Press; 2006.
2. Kuo CY, Su LH, Perera J, Carlos C, Tan BH, Kumarasinghe G, So T, Van PH, Chongthaleong A, Song JH, Chiu CH. Antimicrobial susceptibility of Shigella isolates in eight Asian countries, 2001-2004. J Microbiol Immunol Infect 2008 Apr; 41(2):107-11.
3. Zaidi MB, Estrada-García T. Shigella: a highly virulent and elusive pathogen. Curr Trop Med Rep 2014; 1:81-87.
4. Mehata S, Duan GC. Molecular mechanism of multi-drug resistance in Shigella isolates from rural China. Nepal Med Coll J 2011; 13(1):27-9.
5. Cimmons M. Rapid food-borne pathogen ID system is making a difference. ASM News 2000; 66: 617.
6. McCall B, Stafford R, Cherian S, Heel K, Smith H, Corones N, Gilmore S. An outbreak of multi-resistant *Shigella sonnei* in a long-stay geriatric nursing center. Commun Dis Intell 2000; 24(9):272-5.
7. Seol SY, Kim YT, Jeong YS, Oh JY, Kang HY, Moon DC, Kim J, Lee YC, Cho DT, Lee JC. Molecular characterization of antimicrobial resistance in *Shigella sonnei* isolates in Korea. J Med Microbiol 2006; 55(7):871-7.
8. Talukder KA, Islam Z, Dutta DK, Islam MA, Khajanchi BK, Azmi IJ, Iqbal MS, Hossain MA, Faruque AS, Nair GB, Sack DA. Antibiotic resistance and genetic diversity of *Shigella sonnei* isolated from patients with diarrhea between 1999 and 2003 in Bangladesh. J Med Microbiol 2006; 55(9):1257-63.
9. Wei HL, Wang YW, Li CC, Tung SK, Chiou CS. Epidemiology and evolution of genotype and antimicrobial resistance of an imported *Shigella sonnei* clone circulating in central Taiwan. Diagnostic microbiology and infectious disease 2007; 58(4):469-75.
10. Sousa MÂ, Mendes EN, Collares GB, Péret-Filho LA, Penna FJ, Magalhães PP. Shigella in Brazilian children with acute diarrhea: prevalence, antimicrobial resistance and virulence genes. Mem Inst Oswaldo Cruz 2013; 108(1):30-5.
11. Yoshida S, Handa Y, Suzuki T, Ogawa M, Suzuki M, Tamai A, Abe A, Katayama E, Sasakawa C. Microtubule-severing activity of Shigella is pivotal for intercellular spreading. Science 2006; 314(5801):985-9.
12. Coster TS, Hoge CW, VanDeVerg LL, Hartman AB, Oaks EV, Venkatesan MM, Cohen D, Robin G, Fontaine-Thompson A, Sansonetti PJ, Hale TL. Vaccination against shigellosis with attenuated Shigella flexneri 2a strain SC602. Infect and Immun 1999 Jul 1; 67(7):3437-43.
13. Sansonetti PJ. Rupture, invasion and inflammatory destruction of the intestinal barrier by Shigella, making sense of prokaryote–eukaryote cross-talks. FEMS Microbiol Rev 2001; 25(1):3-14.
14. Gu B, Cao Y, Pan S, Zhuang L, Yu R, Peng Z, Qian H, Wei Y, Zhao L, Liu G, Tong M. Comparison of the prevalence and changing resistance to nalidixic acid and ciprofloxacin of

Shigella between Europe–America and Asia–Africa from 1998 to 2009. Int J Antimicro Ag 2012; 40(1):9-17.
15. Centers for Disease Control and Prevention. National antimicrobial resistance monitoring system: enteric bacteria. 2001 annual report. National Antimicrobial Resistance Monitoring System (NARMS), Atlanta, Ga. 2003.
16. Bowen A, Hurd J, Hoover C, Khachadourian Y, Traphagen E, Harvey E, Libby T, Ehlers S, Ongpin M, Norton JC, Bicknese A. Importation and domestic transmission of *Shigella sonnei* resistant to ciprofloxacin—United States, May 2014–February 2015. Morb Mortal Wkly Rep 2015; 64(12):318-20.
17. Centers for Disease Control and Prevention. CDC Health Information for International Travel 2014: The Yellow Book. Oxford University Press; 2013 Apr 22.
18. Mani S, Wierzba T, Walker RI. Status of vaccine research and development for Shigella. Vaccine. ELSEVIER 2016; 34(26):2887-94.
19. Kotloff KL, Nataro JP, Blackwelder WC, Nasrin D, Farag TH, Panchalingam S, Wu Y, Sow SO, Sur D, Breiman RF, Faruque AS. Burden and etiology of diarrheal disease in infants and young children in developing countries (the Global Enteric Multicenter Study, GEMS): a prospective, case-control study. The Lancet 2013; 382(9888):209-22.
20. Li W, Joshi MD, Singhania S, Ramsey KH, Murthy AK. Peptide vaccine: progress and challenges. Vaccines. 2014; 2(3):515-36.
21. Nascimento IP, Leite LC. Recombinant vaccines and the development of new vaccine strategies. Braz J Med Biol Res 2012; 45(12):1102-11.
22. Giuliani MM, Adu-Bobie J, Comanducci M, Aricò B, Savino S, Santini L, Brunelli B, Bambini S, Biolchi A, Capecchi B, Cartocci E. A universal vaccine for serogroup B meningococcus. PNAS 2006; 103(29):10834-9.
23. Formal SB, Kent TH, May HC, Palmer A, Falkow S, LaBrec EH. Protection of monkeys against experimental shigellosis with a living attenuated oral polyvalent dysentery vaccine. J Bacteriol 1966; 92(1):17-22.
24. Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 2006; 22(13):1658-9.
25. Zhang R, Ou HY, Zhang CT. DEG: a database of essential genes. Nucleic Acids Res 2004; 32(1):271-2.
26. Green ER, Mecsas J. Bacterial Secretion Systems–An overview. Microbiology spectrum. 2016; 4(1).
27. Cusick MF, Libbey JE, Fujinami RS. Molecular mimicry as a mechanism of autoimmune disease. Clinical reviews in allergy & immunology 2012; 42(1):102-11.
28. Berne C, Ducret A, Hardy GG, Brun YV. Adhesins involved in attachment to abiotic surfaces by Gram-negative bacteria. Microbiology spectrum 2015; 3(4).
29. Nancy YY, 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(13):1608-15.
30. Yu CS, Cheng CW, Su WC, Chang KC, Huang SW, Hwang JK, Lu CH. CELLO2GO: a web server for protein subCELlular LOcalization prediction with functional gene ontology annotation. PLoS One 2014; 9(6):e99368.
31. Chen L, Xiong Z, Sun L, Yang J, Jin Q. VFDB 2012 update: toward the genetic diversity and molecular evolution of bacterial virulence factors. Nucleic Acids Res 2012; 40(D1):D641-5.
32. Gasteiger E, Hoogland C, Gattiker A, Duvaud SE, Wilkins MR, Appel RD, Bairoch A. ExPASy—the proteomics server for in-depth protein knowledge and analysis. Nucleic Acids Res. 31, 3784–3788.
33. Parvege MM, Rahman M, Hossain MS. Genome-wide Analysis of Mycoplasma hominis for the Identification of Putative Therapeutic Targets. Drug target insights 2014; 8:51.
34. Naz A, Awan FM, Obaid A, Muhammad SA, Paracha RZ, Ahmad J, Ali A. Identification of putative vaccine candidates against Helicobacter pylori exploiting exoproteome and secretome: a reverse vaccinology based approach. Infection, Genetics and Evolution. 2015; 32:280-91.
35. Barh D, Barve N, Gupta K, Chandra S, Jain N, Tiwari S, Leon-Sicairos N, Canizalez-Roman A, dos Santos AR, Hassan SS, Almeida S. Exoproteome and secretome derived broad spectrum novel drug and vaccine candidates in Vibrio cholerae targeted by Piper betel derived compounds. PloS one. 2013; 8(1):e52773.
36. Sachdeva G, Kumar K, Jain P, Ramachandran S. SPAAN: a software program for prediction of adhesions and adhesion-like proteins using neural networks. Bioinformatics 2005; 21(4):483-91.
37. An YH, Friedman RJ. Concise review of mechanisms of bacterial adhesion to biomaterial surfaces. Journal of Biomedical Materials Research Part A 1998; 43(3):338-48.
38. He Y, Xiang Z, Mobley HL. Vaxign: the first web-based vaccine design program for reverse vaccinology and applications for vaccine development. BioMed Research International 2010.
39. EL-Manzalawy Y, Dobbs D, Honavar V. Predicting linear B-cell epitopes using string kernels. J Mol Recognit 2008; 21(4):243-55.
40. Singh H, Raghava GP. ProPred1: prediction of promiscuous MHC Class-I binding sites. Bioinformatics 2003; 19(8):1009-14.
41. Singh H, Raghava GP. ProPred: prediction of HLA-DR binding sites. Bioinformatics 2001; 17(12):1236-7.
42. Fieser TM, Tainer JA, Geysen HM, Houghten RA, Lerner RA. Influence of protein flexibility and peptide conformation on reactivity of monoclonal anti-peptide antibodies with a protein alpha-helix. PNAS 1987; 84(23):8568-72.
43. CLC bio A/S. The CLC Main Workbench 6.8 is developed by Science Park Aarhus. Finlandsgade, 8200 Aarhus N, and Denmark 2013; 10-12.

44. Bousquet J, Lockey R, Malling HJ. Allergen immunotherapy: therapeutic vaccines for allergic diseases A WHO position paper. Journal of Allergy and Clinical Immunology 1998; 102(4):558-62.
45. Zhang L, Huang Y, Zou Z, He Y, Chen X, Tao A. SORTALLER: predicting allergens using substantially optimized algorithm on allergen family featured peptides. Bioinformatics 2012; 28(16):2178-9.
46. Dimitrov I, Flower DR, Doytchinova I. AllerTOP-a server for in silico prediction of allergens. BMC bioinformatics. 2013; 14(6):S4.
47. Szklarczyk D, Franceschini A, Kuhn M, Simonovic M, Roth A, Minguez P, Doerks T, Stark M, Muller J, Bork P, Jensen LJ. The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. Nucleic Acids Res 2011; 39(1):561-8.
48. Kitano H. Systems biology: a brief overview. Science 2002; 295(5560):1662-4.
49. Rizwan M, Naz A, Ahmad J, Naz K, Obaid A, Parveen T, Ahsan M, Ali A. VacSol: a high throughput in silico pipeline to predict potential therapeutic targets in prokaryotic pathogens using subtractive reverse vaccinology. BMC bioinformatics 2017; 18(1):106.
50. Doytchinova IA, Flower DR. VaxiJen: a server for prediction of protective antigens, tumor antigens and subunit vaccines. BMC bioinformatics 2007; 8(1):4.
51. Xie H, Guo XM, Chen H. Making the most of fusion tags technology in structural characterization of membrane proteins. Molecular biotechnology 2009; 42(2):135-45.
52. Sali A, Blundell TL. Comparative protein modeling by satisfaction of spatial restraints. J Mol Biol 1993; 234(3):779-815.
53. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ. The Phyre2 web portal for protein modeling, prediction and analysis. Nature protocols 2015; 10(6):845-58.
54. Schwede T, Kopp J, Guex N, Peitsch MC. SWISS-MODEL: an automated protein homology-modeling server. Nucleic Acids Res 2003; 31(13):3381-5.
55. Källberg M, Wang H, Wang S, Peng J, Wang Z, Lu H, Xu J. Template-based protein structure modeling using the RaptorX web server. Nature protocols 2012; 7(8):1511-22.
56. Fernandez-Fuentes N, Madrid-Aliste CJ, Rai BK, Fajardo JE, Fiser A. M4T: a comparative protein structure modeling server. Nucleic Acids Res 2007; 35(2):363-8.
57. Eisenberg D, Lüthy R, Bowie JU. [20] VERIFY3D: Assessment of protein models with three-dimensional profiles. Methods Enzymol 1997; 277:396-404.
58. Colovos C, Yeates TO. Verification of protein structures: patterns of non-bonded atomic interactions. Prot Sci 1993; 2(9):1511-9.
59. Wiederstein M, Sippl MJ. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Res 2007; 35(2):407-10.
60. Mayrose I, Penn O, Erez E, Rubinstein ND, Shlomi T, Freund NT, Bublil EM, Ruppin E, Sharan R, Gershoni JM, Martz E. Pepitope: epitope mapping from affinity-selected peptides. Bioinformatics 2007; 23(23):3244-6.
61. Rubinstein ND, Mayrose I, Halperin D, Yekutieli D, Gershoni JM, Pupko T. Computational characterization of B-cell epitopes. Mol Immunol 2008; 31; 45(12):3477-89.

62. Lee H, Heo L, Lee MS, Seok C. GalaxyPepDock: a protein–peptide docking tool based on interaction similarity and energy optimization. Nucleic Acids Res 2015; 43: 431–435.
63. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. UCSF Chimera—a visualization system for exploratory research and analysis. J. Comput. Chem 2004; 25(13):1605-12.
64. Laskowski RA, Swindells MB. LigPlot+: multiple ligand–protein interaction diagrams for drug discovery. J. Chem. Inf. Model 2011; 51 (10):2778–2786.
65. Ueda H, Howson JM, Esposito L, Heward J, Chamberlain G, Rainbow DB, Hunter KM, Smith AN, Di Genova G, Herr MH, Dahlman I. Association of the T-cell regulatory gene CTLA4 with susceptibility to autoimmune disease. Nature 2003; 423(6939):506-11.
66. Outten FW, Huffman DL, Hale JA, O'Halloran TV. The independent cue and cus Systems confer copper tolerance during aerobic and anaerobic growth in *Escherichia coli*. J. Biol. Chem 2001; 276(33):30670-7.
67. Bhavsar AP, Brown NF, Stoepel J, Wiermer M, Martin DD, Hsu KJ, Imami K, Ross CJ, Hayden MR, Foster LJ, Li X. The Salmonella type III effector SspH2 specifically exploits the NLR co-chaperone activity of SGT1 to subvert immunity. PLoS Pathog 2013; 9(7):e1003518.
68. Fralick JA. Evidence that TolC is required for functioning of the Mar/AcrAB efflux pump of Escherichia coli. J Bacteriol 1996; 178(19):5803-5.
69. Sulavik MC, Houseweart C, Cramer C, Jiwani N, Murgolo N, Greene J, DiDomenico B, Shaw KJ, Miller GH, Hare R, Shimer G. Antibiotic Susceptibility Profiles of *Escherichia coli* Strains Lacking Multidrug Efflux Pump Genes. Antimicrob Agents Chemother 2001; 45(4):1126-36.
70. Jap BK, Walian PJ. Biophysics of the structure and function of porins. Q Rev Biophys 1990; 23(04):367-403.
71. Nikaido H. Porins and specific channels of bacterial outer membranes. Mol Microbiol 1992; 6(4):435-42.
72. Jacob-Dubuisson F, Striker R, Hultgren SJ. Chaperone-assisted self-assembly of pili independent of cellular energy. J. Biol. Chem 1994; 269(17):12447-55.
73. Schifferli DM, Alrutz MA. Permissive linker insertion sites in the outer membrane protein of 987P fimbriae of Escherichia coli. J. Bacteriol 1994; 176(4):1099-110.
74. MacIntyre S, Henning U. The role of the mature part of secretory proteins in translocation across the plasma membrane and in regulation of their synthesis in *Escherichia coli*. Biochimie 1990; 72(2-3):157-67.
75. Van Rosmalen M, Saier MH. Structural and evolutionary relationships between two families of bacterial extra cytoplasmic chaperone proteins which function cooperatively in fimbrial assembly. Res Microbiol 1993; 144(7):507-27.
76. Krishnan S, Prasadarao NV. Outer membrane protein A and OprF: versatile roles in Gram-negative bacterial infections. FEBS Journal 2012; 279(6):919-31.
77. Gophna U, Barlev M, Seijffers R, Oelschlager TA, Hacker J, Ron EZ. Curli fibers mediate internalization of *Escherichia coli* by eukaryotic cells. Infect Immun 2001; 69(4):2659-65.

78. Koskiniemi S, Lamoureux JG, Nikolakakis KC, de Roodenbeke CT, Kaplan MD, Low DA, Hayes CS. Rhs proteins from diverse bacteria mediate intercellular competition. PNAS 2013; 110(17):7032-7.
79. Korepanova A, Gao FP, Hua Y, Qin H, Nakamoto RK, Cross TA. Cloning and expression of multiple integral membrane proteins from Mycobacterium tuberculosis in Escherichia coli. Protein Science. 2005; 14(1):148-58.
80. Barth RJ, Fisher DA, Wallace PK, Channon JY, Noelle RJ, Gui J, Ernstoff MS. A randomized trial of ex vivo CD40L activation of a dendritic cell vaccine in colorectal cancer patients: tumor-specific immune responses are associated with improved survival. Clin Cancer Res 2010; 16(22):5548-56.
81. Chung EH. Vaccine allergies. Clin Exp Vaccine Res 2014; 3(1):50-7.
82. Kobayashi N, Nishino K, Yamaguchi A. Novel Macrolide-Specific ABC-Type Efflux Transporter in *Escherichia coli*. J Bacteriol 2001; 183(19):5639-44.
83. Levy SB. Active efflux mechanisms for antimicrobial resistance. Antimicrob. Agents Chemother 1992; 36(4):695.
84. Yang H, Duan G, Zhu J, Lv R, Xi Y, Zhang W, Fan Q, Zhang M. The AcrAB-TolC pump is involved in multidrug resistance in clinical *Shigella flexneri* isolates. Microb Drug Resist 2008; 14(4):245-9.
85. Tatsumi R, Wachi M. TolC-dependent exclusion of porphyrins in Escherichia coli. J Bacteriol 2008; 190(18):6228-33.
86. Bleuel C, Große C, Taudte N, Scherer J, Wesenberg D, Krauß GJ, Nies DH, Grass G. TolC is involved in enterobactin efflux across the outer membrane of *Escherichia coli*. . J Bacteriol 2005; 187(19):6701-7.
87. Nishino K, Latifi T, Groisman EA. Virulence and drug resistance roles of multidrug efflux systems of *Salmonella enterica serovar Typhimurium*. Mol Microbiol 2006; 59(1):126-41.
88. Yang H, Duan G, Zhu J, Lv R, Xi Y, Zhang W, Fan Q, Zhang M. The AcrAB-TolC pump is involved in multidrug resistance in clinical *Shigella flexneri* isolates. Microb Drug Resist 2008; 14(4):245-9.
89. Kato A, Ohnishi H, Yamamoto K, Furuta E, Tanabe H, Utsumi R. Transcription of emrKY is regulated by the EvgA-EvgS two-component system in Escherichia coli K-12Biosci. Biotechnol. Biochem 2000; 64(6):1203-9.
90. Koebnik R. Structural and Functional Roles of the Surface-Exposed Loops of the β-Barrel Membrane Protein OmpA from *Escherichia coli*. J Bacteriol 1999; 181(12):3688-94.
91. Prasadarao NV, Wass CA, Weiser JN, Stins MF, Huang SH, Kim KS. Outer membrane protein A of *Escherichia coli* contributes to invasion of brain microvascular endothelial cells. Infect Immun 1996; 64(1):146-53.
92. Prasadarao NV, Wass CA, Stins MF, Shimada H, Kim KS. Outer membrane protein A-promoted actin condensation of brain microvascular endothelial cells is required for *Escherichia coli* invasion. Infect Immun 1999; 67(11):5775-83.

93. Power ML, Ferrari BC, Littlefield-Wyer J, Gordon DM, Slade MB, Veal DA. A naturally occurring novel allele of *Escherichia coli* outer membrane protein A reduces sensitivity to bacteriophage. Appl Environ Microbiol 2006; 72(12):7930-2.

Table 1. The virulent exoproteome and secretome of *S. sonnei* screened from the essential and human non-homologous proteome

Protein name	Non-paralogous	Host Non-Homologous	Essentiality	CELLO2GO	CELLO	PSORTB	Virulent		
							Status	Identity	Bit Score
Cation transporter	Yes	Yes	Yes	OM	OM	OM	Yes	90	796
E3 ubiquitin--protein ligase	Yes	Yes	Yes	EC		EC	Yes	97	1108
TolC	Yes	Yes	Yes	OM	OM	OM	Yes	70	600
nmpC	Yes	Yes	Yes	OM	OM	OM	Yes	83	586
Phosphoporin PhoE	Yes	Yes	Yes	OM	OM	OM	Yes	61	446
OmpA	Yes	Yes	Yes	OM	OM	OM	Yes	95	638
Porin protein	Yes	Yes	Yes	OM	OM	OM	Yes	71	555
OmpE	Yes	Yes	Yes	OM	OM	OM	Yes	65	457
OmpC	Yes	Yes	Yes	OM	OM	OM	Yes	63	438
OmpF	Yes	Yes	Yes	OM	OM	OM	Yes	61	431
LpfC	Yes	Yes	Yes	OM	OM	OM	Yes	99	1689
NmpC	Yes	Yes	Yes	OM	OM	OM	yes	62	145
RhsA	Yes	Yes	Yes	OM	OM	OM	Yes	67	1759
Major curlin subunit	Yes	Yes	Yes	EC	EC	EC	Yes	86	251

OM, Outer Membraneous, EC, Extracellular

Table 2. B-cell derived T-cell epitopes of the three prioritized putative vaccine proteins against *S. sonnei*

B-cell epitopes	T-cell epitopes	MHC-I	MHC-II	Location	Total No	Antigenicity	VirulentPred	IC_{50}	Allerginicity
PIYQGGMVNSQVKQAQYNFV	MVNSQVKQA	3	30	319-327	33	1.1633	1.0595	49.09	0.263
RHENGDGWGLSTTYDLGMGF	WGLSTTYDL	22	12	200-208	34	0.8486	1.0595	37.33	0.263
ANGDGFGISSTYVYDGFGIG	FGISSTYVY	14	11	198-206	25	0.8995	0.9892	17.5	0.263

MHC-I, Major Histocompatibility complex I, MHC-II, Major Histocompatibility complex II