Finding Potential Therapeutic Targets against *Shigella flexneri* through Proteome Exploration

Mohammad Uzzal Hossain[1], Md. Arif Khan[2], Abu Hashem[3], Md. Monirul Islam[1], Mohammad Neaz Morshed[2], Chaman Ara Keya[4] and Md. Salimullah[5]*

[1] Department of Biotechnology and Genetic Engineering, Life Science Faculty, Mawlana Bhashani Science and Technology University, Tangail, Bangladesh, [2] Department of Science and Humanities, Military Institute of Science and Technology, Mirpur Cantonment, Dhaka, Bangladesh, [3] Microbial Biotechnology Division, National Institute of Biotechnology, Savar, Bangladesh, [4] Department of Biochemistry and Microbiology, North South University, Bashundhara, Dhaka, Bangladesh, [5] Molecular Biotechnology Division, National Institute of Biotechnology, Savar, Bangladesh

Background: *Shigella flexneri* is a gram negative bacteria that causes the infectious disease "shigellosis." *S. flexneri* is responsible for developing diarrhea, fever, and stomach cramps in human. Antibiotics are mostly given to patients infected with shigella. Resistance to antibiotics can hinder its treatment significantly. Upon identification of essential therapeutic targets, vaccine and drug could be effective therapy for the treatment of shigellosis.

Methods: The study was designed for the identification and qualitative characterization for potential drug targets from *S. flexneri* by using the subtractive proteome analysis. A set of computational tools were used to identify essential proteins those are required for the survival of *S. flexneri*. Total proteome (13,503 proteins) of *S. flexneri* was retrieved from NCBI and further analyzed by subtractive channel analysis. After identification of the metabolic proteins we have also performed its qualitative characterization to pave the way for the identification of promising drug targets.

Results: Subtractive analysis revealed that a list of 53 targets of *S. flexneri* were human non-homologous essential metabolic proteins that might be used for potential drug targets. We have also found that 11 drug targets are involved in unique pathway. Most of these proteins are cytoplasmic, can be used as broad spectrum drug targets, can interact with other proteins and show the druggable properties. The functionality and drug binding site analysis suggest a promising effective way to design the new drugs against *S. flexneri*.

Conclusion: Among the 53 therapeutic targets identified through this study, 13 were found highly potential as drug targets based on their physicochemical properties whilst only one was found as vaccine target against *S. flexneri*. The outcome might also be used as module as well as circuit design in systems biology.

Keywords: *S. flexneri*, drug target, therapeutics, metabolic proteins, proteome

INTRODUCTION

In developing countries, *S. flexnari* is the foremost reason of bacillary dysentery among the four species of Shigella. Annually, 1.1 million deaths occur out of 164.7 million cases of Shigellosis worldwide and children <5 years of age are the worst victim of it (Bardhan et al., 2010). Depending on the combinations of antigenic determinants existing on the O antigen of the cell envelope lipopolysaccharide (LPS)(2–14), *S. flexneri* is divided into 19 serotypes viz. serotypes 1a, 1b, 1c, 1d, 2a, 2b, 3a, 3b, 4a, 4av, 4b, 5a, 5b, X, Xv, Y, Yv, F6, and 7b (Simmons and Romanowska, 1987; Kotloff et al., 1999; Stagg et al., 2009; Ye et al., 2010; Foster et al., 2011; Sun et al., 2011, 2012; Luo et al., 2012; Perepelov et al., 2012).

Shigella, as Gram-negative facultative human pathogens, cause intestinal infections with sign and symptoms such as fever, abdominal cramps and watery, or bloody diarrhea. Recent evidence suggests that the third prominent reason of global infant mortality is diarrhea (Black et al., 2010). Children below the age of five are mostly affected by Shigella (Kotloff et al., 1999; Peng et al., 2002). Majority of endemic dysentery are caused by *S. flexneri* in regions of the world having less facility of sanitation. Also, improper use of antibiotic rendered resistance to shigella. Therefore, antimicrobials development could be the better alternatives for the prevention of antibiotic-resistant Shigella as it is non-toxic, cheap, easy to apply and produce life-long immunity (Kärnell et al., 1995; Coster et al., 1999; Mukhopadhaya et al., 2003, 2006; Katz et al., 2004; Ranallo et al., 2005; Paterson, 2006). Recently, it has been reported that Shigella serotypes including *S. flexneri* 2a, 3a, *S. dysenteriae* 1, and *S. sonnei* are targeted for vaccine development. In developing countries, the first three are more widespread while the last serotypes found in regions where sanitation standard is high (Jennison and Verma, 2004; Mukhopadhaya et al., 2006).

Antibiotic is mostly given for the treatment with shigella infection. But the high failure rate is increasing day by day due to acquired resistance to commonly used antibiotics (Nessar et al., 2012). To combat these infections new drugs discovery are necessary due to emerging multi-drug resistance and absence of optimal treatment. Therefore, to identify novel drug target(s) are one of the key ways of drug discovery. The present study is aimed to identify potential drug targets in *S. flexneri* by subtractive proteome analysis. The traditional way of drug discovery needs more times, expensive experiments and also laborious efforts whereas the computational way could be effective alternative way which could accelerate of drug discovery process within a very short time. The identification of drug targets has been growing more by coupling of "omics" data viz., genomics, proteomics, and metabolomics and the utilization of computational approaches. The sequencing of genome and revealing of proteome of disease causing organisms are advancing the search of drug targets in the field of drug research based on essential genes of specific pathogen, interacting factors of host-pathogen, proteins persistence, resistance genes/resistance-associated proteins, metabolic pathways, prediction of gene expression levels (Galperin and Koonin, 1999; Yeh et al., 2004; Briken, 2008; Raman et al.,

2008; Barh et al., 2011; Vetrivel et al., 2011). To identify novel drug targets these approaches have already been utilized in several life threatening pathogens, including *Mycobacterium tuberculosis* (Anishetty et al., 2005; Asif et al., 2009), *M. leprae* (Shanmugam and Natarajan, 2010), *M. ulcerans* (Butt et al., 2012), *Helicobacter pylori* (Sarkar et al., 2012), *Streptococcus pneumonia* (Singh et al., 2007), *Yersinia pestis* (Sharma and Pan, 2012), and *Pseudomonas aeruginosa* (Sakharkar et al., 2004). The foremost criteria for the identification of promising therapeutic candidates are essentiality and selectivity/specificity. To avoid the undesired interactions of host-pathogen which could occur to death of the pathogen by mediating inhibition of important proteins, the targets must be specific as drugs interact to host proteins. The present study integrates the various computational methods for the identification and characterization of the drug targets of *S. flexneri* which enables us to identify 53 potential therapeutic targets based on essentiality and specificity. The qualitative characterization of 53 therapeutic candidates predicts the uniqueness in metabolic pathway, capability to act as a broad spectrum target, the cellular location, cellular function, functional association with metabolic proteins, and druggability properties.

MATERIALS AND METHODS

An *in silico* systematic method consists of three stages is applied to classify and illustrate possible drug targets against *S. flexneri*. At the stage I, the protein dataset was collected from the NCBI–FTP site for the analysis. Clarification of protein datasets through subtractive channel of analysis was done in stage II and possible drug targets found from stage I and II were qualitatively characterized in stage III.

Stage I: Mining of Protein Datasets

The whole protein sequences of *S. flexneri* 2a strain 2457T were retrieved from NCBI (http://www.ncbi.nlm.nih.gov/) protein database as FASTA format.

Stage II: Subtractive Channel of Analysis

Protein datasets were further selected and subjected to be qualified by passing through a sequence of subtractive proteome analysis. In this process, highly discerning and effective drug targets could be identified.

Identification of Paralog Proteins

Paralog proteins were identified by exposing *S. flexneri* proteins to CD-hit suite (Huang et al., 2010). The CDHIT server can be utilized for the "sequence identity" among the proteins (Fasta file). In this server users can exploit 10–90% sequence identity in the "sequence identity cut-off" box depending on their requirement (Huang et al., 2010). It has been widely accepted to set 60% sequence identity as cut-off to maintain a rigid criteria to remove duplicate proteins (Dutta et al., 2006; Barh and Kumar, 2009; Rahman et al., 2014; Mondal et al., 2015; Hasan et al., 2016). Therefore, we have used CDHIT server (http://weizhong-lab.ucsd.edu/cdhit_suite/cgi-bin/index. cgi?cmd=cd-hit) in which 0.6 (60%) was manually set in

"sequence identity cut-off" box for the stringent selection of duplicate proteins. The duplicates were omitted and for further selection those were designated from the remaining set of non-paralog proteins that have in excess of 100 amino acids.

Identification of Orthologs in Gut Flora

The non-duplicate proteins that were described in the preceding step, analyzed to explore their similarity with the proteome of human gut microbiota (Fujimura et al., 2010). BLASTing (Altschul et al., 1990) of those proteins against the gut flora proteome available from literature with 0.0001 as *e*-value threshold helped to further escape from the orthologs.

Identification of Orthologs in Human Proteome

BLASTp (Altschul et al., 1990) was done for the qualified proteins resulted from stage I against non-redundant database of *H. sapiens* with an estimated threshold value of 0.0001. Those proteins are selected for the next step that shows no hits for the above mentioned *e*-value.

Essentiality Analysis

To identify essential proteins by BLASTp searching against Database of Essential Genes (DEG; Zhang and Lin, 2009), non-homologous proteins were screened. Protein alignments associated with expect value of <0.0001(Barh et al., 2011; Sharma and Pan, 2012) were considered as more significant hits.

Metabolic Pathway Analysis

Metabolic pathway analysis was done by KAAS server at KEGG to classify the possible targets of the human non-homologous essential proteins of *S. flexneri* which has been assimilated from DEG (Moriya et al., 2007). Functional annotation of genes is obtained through KAAS by *comparing* BLAST beside manually created KEGG GENES database. The result comprises of KO (KEGG Orthology) task that identify the metabolic proteins.

Stage III: Qualitative Characterization of the Short-Listed Targets

Detection of Proteins Involved In Unique Pathways

KEGG (Kyoto Encyclopedia of Genes and Genomes) Genome Database (Kanehisa et al., 2014) was applied to find the distinctive metabolic pathways of *S. flexneri* in contrast to *H. sapiens*. In order to compare the metabolic pathway three letter codes that are particular for host and pathogen was placed in the Genome Comparison and Combination box. Eventually KEGG Genome generated pathway maps were recognized as distinctive pathways. Human non-homologous metabolic proteins that were selected formerly were subjected to screening to disclose their association in those distinctive pathways.

Cellular Localization Analysis

In this study, PSORTb 3.0.2(Yu et al., 2010), CELLO 2.0(Yu et al., 2006), Signal IP 4.1(Petersen et al., 2011), and Phobius (Käll et al., 2007) were used to identify the location of the short-listed proteins in various modules like SVM, S-TMHMM, and SCL-BLAST with proteins of known localization from bacteria (Gram-positive and negative) and archaea as training set. These server were used for the accurate localization of identified targets.

Broad Spectrum Analysis

BLASTp (Altschul et al., 1990) search was done beside a wide-range of pathogenic bacteria to explore proteins with an estimated threshold value of 0.005 to find broad spectrum targets. In the broad spectrum analysis, overall 240 disease-causing bacteria from different genus including with other serotypes of *S. flexneri* were used. Cluster of Orthologous Groups of proteins (COG) search was used to identify homologs of the short listed targets in other pathogenic bacteria by means of COGnitor from NCBI that matches the query sequence with the COG database.

Interactome Analysis

STRING 9.0 (Szklarczyk et al., 2011) was used to build a protein-protein interaction network for each of the authorized targets. All interactors with low along with medium confidence score (<0.700) were removed from the network to escape false positives and false negatives.

Functionality Analysis

INTERPROSCAN, a tool that incorporates various protein signature recognition methods and databases, was used to predict the role of the hypothetical proteins from the list of possible targets (Mulder and Apweiler, 2007).

Binding Site Analysis

Local meta-threading-server (LOMETS; Wu and Zhang, 2007) was used for the selection of best template from 9 locally-installed threading programs (FFAS-3D, HHsearch, MUSTER, pGenTHREADER, PPAS, PRC, PROSPECT2, SP3, and SPARKS-X). Then, Modeller 9.17 (Webb and Sali, 2014) was used to build the 3D model of metabolic hypothetical proteins. A set of tools Procheck (Laskowski et al., 1993), Verify3D (Eisenberg et al., 1997), and proSA (Wiederstein and Sippl, 2007) were utilized for the model quality assessment. Thereafter, COFACTOR server (Roy et al., 2012) was also employed to reveal out the binding site of 6 metabolic hypothetical proteins.

Druggability Analysis

In the current study, a similarity search was done in contrast to Drug Bank 3.0 target collection for all of the target (Knox et al., 2011).

Docking Analysis

We have performed Autodock vina (Trott and Olson, 2010) for the analysis of binding affinity to the druggable targets. We retrieved all the interacting drugs (.pdb files) from 6 druggable targets which were found by virtual screening in Drugbank database. Then we generated the .pdbqt files of these 6 druggable targets for docking experiments. Blind docking was performed for the identification of most effective binding site of these drugs. A grid box parameter for covering the whole protein was set for all docking runs.

Furtheremore, we have performed ProtParam (http://web.expasy.org/protparam/) and VaxiJen (http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html) server for the identification of most suitable drug/vaccine target.

RESULTS AND DISCUSSION

Data Acquisition

A number of potential therapeutic targets in *S. flexneri* is detected and categorized through an *in silico* method. The strategy applied a hierarchy of subtractive analysis where functionally essential proteins of *S. flexneri* derived from whole proteome of 13,503 proteins dataset (Drug Target, Data Sheet S1). The number of subtractive channel was used to select the potential candidates that could serve either drug or vaccine for therapeutics treatment. Overall, the strategy composed of stages of data mining, subtractive channel of analysis and qualitative characterization.

Subtractive Channel of Analysis

Subtraction of Duplicate and Mini Proteins

CD-HIT suite identified the duplicate or paralog proteins from the proteome. This tool sorted out 4559 proteins as non-duplicate at sequence identity upto 60% (Drug Target, Data Sheet S2). Proteins that showed above 60% matching were considered as duplicates or paralogs in this analysis. We have chosen 60% similarity as a cut-off to maintain a very stringent selection criteria for the identification of the most effective therapeutic targets. Also, we considered to keep one protein from two identical sequences (>60% similarity) as they might be similar for protein domain, motifs, binding site etc. The proteins found from non-duplicate analysis have various length distribution. Proteins were excluded from the analysis having length of less than 100 amino acids known as mini proteins (Wang et al., 2008; Kumar et al., 2010; Barh et al., 2011). Prokaryotic genomes contain a high level of mini proteins which have key role in numerous biological phenomenon as well as regulatory purposes (Kumar et al., 2010). These mini proteins were deleted from the non-duplicate proteins as they are less likely to represent the essential therapeutics candidate (Drug Target, Data Sheet S3). In addition to this, the larger amino acid sequence has the probability to be involved in essential metabolic pathways (Haag et al., 2012).

Subtraction of Orthologs in Gut Flora

A critical step in this study is to identify proteins that are non-homologous to gut flora proteins for circumventing extreme lethal effects in host. It is reported that around 10^{14} (Kärnell et al., 1995) microorganisms exist in the gastrointestinal tract of a normal healthy human (Fujimura et al., 2010). As gut microbiota maintains a symbiotic relationship, it helps in metabolism by fermenting indigestible food particles along with defense from colonization of pathogenic bacteria in gut (Rabizadeh and Sears, 2008). If gut flora proteins are spoiled accidentally it may decline the microbiota which may cause harm to the host. Therefore, Human gut flora proteins were subjected to analyze for the assurance of non-similarity with our selected proteins. BLASTp was employed to recognize the gut homologs with our selected proteins and gut flora proteins. Eventually, we have identified 2708 gut flora similar proteins (Drug Target, Data Sheet S4 and **Figure 1**).

Subtraction of Orthologs in Human proteome

Identification of pathogen specific protein is the main goal of this analysis. The importance of this step is to reduce unwanted cross reactivity of the drug and thus to inhibit its binding to the active sites of the homologous proteins in host (Sarkar et al., 2012). In *in silico* drug target identification method, the first step is considered as the filtration of homologous proteins to human proteome (Anishetty et al., 2005; Sarkar et al., 2012). This non-similarity analysis was carried out for the gut flora non-similarity proteins datasets. BLASTp was used for similarity search in contrast to whole proteome of *H. sapiens* (host) with *e*-value threshold of 0.0001 (Altschul et al., 1990). Those Proteins were considered as close homologs that exhibit hits with proteins of human proteome. From the analysis of 2708 input proteins 1987 homologous proteins were omitted and 721 proteins were nominated that are non-homologous to human (Drug Target, Data Sheet S5 and **Figure 1**).

Identification of Essential Genes in *S. flexneri*

DEG server was used for the screening the essential genes of *S. flexneri* from human proteome non-homologous protein list with an estimated value of 0.0001. DEG 6.1 is a storehouse of genes necessary for the survival of an organism. It comprises 10,618 essential genes from prokaryotic and eukaryotic organisms. These types of proteins were considered as essential and it was clearly demonstrated that similar proteins which are crucial in one organism are likely to be essential in another. A potential drug target possesses a crucial feature for the existence of the pathogen and must be an indispensable protein (Sarkar et al., 2012). The 67 proteins (Drug Target, Data Sheet S6 and **Figure 1**) out of 721 non-homologous input proteins were nominated for the consecutive analysis and considered as vital for the existence of the pathogen as they have homologs with not more than the given threshold value (Supplementary Table, S1). Proteins showing no hit against DEG were omitted from the analysis and regarded as non-essential.

Metabolic Pathway Analysis

The output of this server assists to identify the potential drug targets by revealing the KEGG pathways as well as KO (KEGG orthology) assignments. About 53 proteins are involved in metabolic pathways obtained from the essential proteins (Drug Target, Data Sheet S7 and **Figure 1**). These 53 proteins (Supplementary Table, S2) have the key role in metabolism for the bacterial survival.

Qualitative Characterization of Metabolic Pathway Proteins

Unique Pathway Analysis

Besides the identification of metabolic pathway proteins we have also analyzed the unique metabolic pathway proteins which answers the disputable question whether the metabolic pathway proteins are also present or not in host. Here, we have found 11 unique metabolic pathway proteins that are only present in the bacterial metabolic pathway (**Table 1**). These unique proteins were found in the pathways: Purine metabolism, Pyrimidine metabolism, Fructose and mannose metabolism, Amino sugar and nucleotide sugar metabolism, Lipopolysaccharide biosynthesis, Pyruvate metabolism, Propanoate metabolism Butanoate metabolism, Lysine biosynthesis, Terpenoid

FIGURE 1 | Schematic representation of whole work.
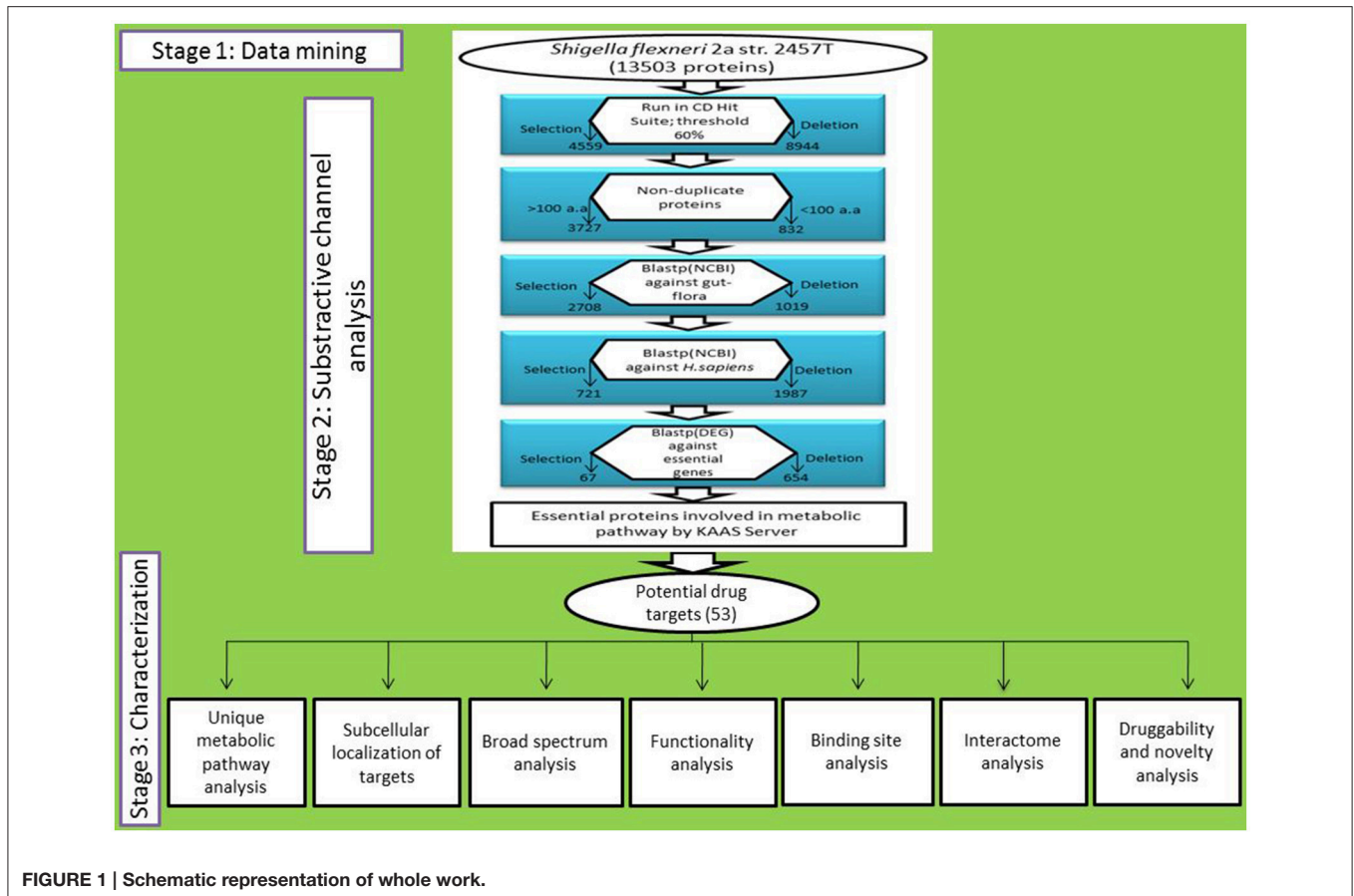
TABLE 1 | Proteins involved in unique pathways.

SN	Accession no.	Name of protein	KO number (gene name)	Pathways
1	NP_839575.1	DNA-directed RNA polymerase subunit alpha	K03040	Purine metabolism, Pyrimidine metabolism
2	NP_838872.1	Mannose-specific IIA component	K02794(manX)	Fructose and mannose metabolism, Amino sugar, and nucleotide sugar metabolism
3	NP_838706.1	Arabinose-5-phosphate isomerase	K06041	Lipopolysaccharide biosynthesis
4	NP_838628.1	Formate C-acetyltransferase	K00656 (pflD)	Pyruvate metabolism, Propanoate metabolism Butanoate metabolism
5	NP_835770.1	4-hydroxy-tetrahydrodipicolinate reductase	K00215(dapB)	Lysine biosynthesis
6	NP_835768.1	4-hydroxy-3-methylbut-2-enyl diphosphate reductase	K03527 (ispH, lytB)	Terpenoid backbone biosynthesis
7	AAP19547.1	PTS system, ascorbate-specific IIC component	K03475	Phosphotransferase system (PTS)
8	EFS15406.1	penicillin-binding protein 1C	K05367 (pbpC)	Peptidoglycan biosynthesis
9	EFS14577.1	Flagellar FliJ protein	K02413 (fliJ)	Flagellar assembly
10	EFS13661.1	Arginine N-succinyltransferase	K00673 (astA)	Arginine and proline metabolism
11	EFS12253.1	RNA polymerase nonessential primary-like sigma factor	K03087	Bacterial pathogenic cycle

backbone biosynthesis, Phosphotransferase system (PTS), Peptidoglycan biosynthesis, Flagellar assembly, Arginine and proline metabolism and bacterial pathogenic cycle.

Subcellular Localization

Proteins can be found in five possible subcellular locations, specifically, cytoplasm, periplasm, plasma membrane, outer membrane, and extracellular. The importance of the localization study is to depict the protein as drug or vaccine target. Surface membrane proteins and cytoplasmic proteins can be used as vaccine and drug targets respectively (Barh et al., 2011). Protein databases like UniProt contain information about subcellular location of some proteins. PSORTb 3.0.2 (Yu et al., 2010), CELLO 2.0 (Yu et al., 2006), SignalIP (Petersen et al., 2011), and Phobius (Käll et al., 2007) tools were utilized for the prediction of subcellular localization.

Therefore, we have analyzed the essential human non-homologous metabolic pathway proteins (53) for the prediction of their subcellular localization. From these analysis we have found 29 targets as cytoplasmic, 14 targets as inner membrane, 6 targets as periplasmic, 3 targets as outer membrane and only 1 target as extracellular (**Table 2**).

Broadspectrum Analysis

The progress of drug resistance can be reduced greatly by such types of broadspectrum and pathogen specific targets analysis (Raman et al., 2008). In this study, list of pathogenic bacteria described in literature was well-thought out (Griffith et al., 2007; Raman et al., 2008; Shenai et al., 2009). It is theorized from the similarity analysis in contrast to each of the pathogen that close homologs existing in more number of pathogens are more expected to be a "promising broad spectrum target" (Raman et al., 2008). Therefore, we have also analyzed broadspectrum targets by comparing the sequences analysis between the 240 pathogens shigella species as well as other clinically important bacterial pathogens with the selected 53 essential metabolic proteins (Drug Target, Data Sheet S7). This was done by the BLASTp analysis. About 25 targets were found to have close identity with less than 10 bacterial pathogens, 10 targets in less than 20 bacteria, 8 targets in less than 30, 7 less than 40 bacteria, and 3 in less than 50

bacterial pathogens (**Table 3**). From these results, it is concluded that these 3 (NP_836465.1, NP_836278.1 and NP_835876.1) or 7 (NP_837443.1, NP_837438.1, NP_836948.1, NP_836937.1, NP_836681.1, NP_836675.1 and NP_836672.1) targets are the exclusive proteins of Shigella.

Interactome Analysis

After data mining, finally identified protein data analyzed and depicted a PPI network which is shown in (**Figure 2**). The networking was established by STRING analysis. In STRING, one protein is interacted with a number of proteins and showed the strength of interaction as score. The interacting score depends on Neighborhood in the genome, Gene fusions, Co-occurrence across genomes, Co-Expression, Association in curated databases and text mining (Jensen et al., 2009). Protein network contains high confidence interactors with score more than or equal to 0.700. Based on the variation in the critical network parameter values the potentiality of the targets was determined. In the bacterial metabolic system, the significance of the query protein was figured out from the number of interacting proteins (nodes) and interactions (edges) interrupted on its deletion (Kushwaha and Shakya, 2010). In low confidence score (0.150), all of the proteins interacted with each other. In highest confidence score (0.900), 30S ribosomal protein (rpsD), DNA-directed RNA

TABLE 2 | Subcellular localization of drug targets.

Cytoplasmic	Inner membrane	Periplasmic	Outer membrane	Extracellular
NP_839574.2	NP_836827.2 (TM;6)	NP_839600.1 (SP)	NP_837438.1	EFS11306.1
NP_839575.1	NP_839768.1 (SP),	NP_837676.1	EFS15865.1	
NP_839064.1	(TM;1)	NP_836681.1 (SP),	EFS15439.1	
NP_838943.1	NP_839521.1 (TM;12)	(TM;1)		
NP_838894.1	NP_839165.1	EFS14577.1		
NP_838872.1	NP_836948.1 (TM;5)	EFS13874.1 (TM;1)		
NP_838722.1	NP_836937.1 (TM;6)	EFS12950.1 (SP)		
NP_838706.1	NP_836465.1 (TM;5)			
NP_838628.1	AAP19547.1 (TM;12)			
NP_837679.1	EFS15406.1			
NP_837604.1	EFS14933.1TM;2)			
NP_837597.1	EFS13325.1			
NP_837443.1	EFS11623.1 (TM;10)			
NP_836675.1	EFS10693.1 (TM;7)			
NP_836672.1	NP_837444.1			
NP_836278.1				
NP_835876.1				
NP_835770.1				
NP_835873.1				
NP_835768.1				
AAP19293.1				
AAP18497.1				
AAP16677.1				
EFS15144.1				
EFS15122.1				
EFS13661.1				
EFS12253.1				
EFS11712.1				
EFS10930.1				

Here, SP, Signal Peptide; TM, Transmembrane Helices.

TABLE 3 | Broad spectrum analysis of essential proteins.

Hit pathogenic bacterial protein (240 Pathogenic bacteria)				
=>10	=>20	=>30	=>40	=>50
EFS11712.1	NP_837679.1	NP_835770.1	NP_837443.1	NP_836465.1
EFS11623.1	NP_837676.1	NP_835768.1	NP_837438.1	NP_836278.1
EFS11306.1	NP_837604.1	AAP19547.1	NP_836948.1	NP_835876.1
EFS10930.1	NP_837597.1	AAP19293.1	NP_836937.1	
EFS10693.1	NP_837444.1	AAP18497.1	NP_836681.1	
NP_836827.2	EFS15406.1	AAP16677.1	NP_836675.1	
NP_839574.2	EFS15144.1	EFS15865.1	NP_836672.1	
NP_839768.1	EFS15122.1	EFS15439.1		
NP_839600.1	EFS14933.1			
NP_839575.1	EFS14577.1			
EFS13874.1				
EFS13661.1				
EFS13325.1				
EFS12950.1				
NP_839521.1				
NP_839165.1				
NP_839064.1				
NP_838943.1				
NP_838894.1				
EFS12253.1				
NP_838872.1				
NP_838722.1				
NP_838706.1				
NP_838628.1				
NP_837679.1				

polymerase subunit alpha (rpoA), RNA polymerase sigma factor RpoS (RpoS), RNA polymerase-binding transcription factor (dksA), anti-RNA polymerase sigma 70 factor (rsd), Holliday junction resolvase (ruvC), Holliday junction DNA helicase (RuvA), 7,8-dihydropteroate synthase (folP), and 2-amino-4-hydroxy-6-hydroxymethyldihydropteridine pyrophosphokinase (folK) were shown to interact with each other (**Figure 2**). The highest confidence score of interaction found in between folP and folk which was 0.999. The homologous genes of those proteins are neighbors of other species and in addition, putative homologs are reported to interact in other species.

Functionality Analysis

From the subtractive analysis we have found 6 hypothetical proteins (Drug Target, Data Sheet S8) which are uncharacterized proteins from the list of 53 potential targets. Attempts were made to characterize their function. To predict their function we have used INTERPROSCAN for the characterization of their functional domain, Molecular function, and Biological process (**Table 4**). But Molecular function and Biological process also was none predicted in some of the cases.

Binding Site Analysis

We have selected the best template from the Local meta-threading-server (LOMETS; Wu and Zhang, 2007) where more than one threading program showed same template with higher score (Align Length, Coverage, Z score and

Confidence score; Supplementary Figure S1). After selecting the template we have prepared the required files for sequence of interest and template (.ali, .pir, .bin, and python file). We have run python programming based Modeller software (Modeller 9.17) for generating the number of models (Webb and Sali, 2014). Then we have followed the both DOPE (Discrete Optimized Protein Energy) and GA341 method for selecting the best model from the number of generated models. As GA341 is not as good as DOPE at differentiating between high and low quality model, we have calculated the DOPE score as it is designed for selecting the best structure from a collection of models built by MODELLER. We have selected the best model which shown the lowest DOPE score as it is reported that lower the DOPE score the better is the model. Thereafter, we have checked the quality assessment of every model built by Modeller. We have utilized Ramachandran Plot (Laskowski et al., 1993), Verify3D (Eisenberg et al., 1997), and ProSA (Wiederstein and Sippl, 2007) for the model assessment. All the three tools showed satisfactory results which are summarized in Supplementary Table, S3, Figures, S2, S3. Through these results we have confirmed the predicted model of each hypothetical protein as high-quality. All the predicted model showed above 90% residues in favorable region (Supplementary Table, S3). Verify3D and ProSA also confirmed that the built model are good (Supplementary Figures, S2, S3). We have also analyzed the binding site and binding site residues of these hypothetical proteins (**Table 5** and Supplementary Figures, S4–S9).

Druggability Analysis

High binding affinity to the drug like molecule is a major feature of a "druggable target." Drug Bank contains 6816 experimental and FDA-approved drugs, 4326 drug targets and 169 drug enzymes/carriers. The module estimates the degree of similarity by using the similarity search option BLASTp program (Mondal et al., 2015). Presence of targets acts as assign for their druggable property. In contrast, its absence designates the novelty of the target and therefore, categorized as "novel target" (Knox et al., 2011). From the selected 53 targets we have found 6 targets (Drug Target, Data Sheet S9) which showed the similarity with approved, investigational, and experimental drugs when the BLASTp search was done against DrugBank database with cut-off parameters (**Table 6**). We have observed that target EFS15406.1 showed the highest similarity with 15 drugs which could act as inhibitors (both approved and investigational). The docking energy also revealed that the interacting drugs have better binding affinity to their respective targets (**Table 6**). Therefore, out of 53 therapeutic targets, 47 could be considered as novel drug targets as the remaining 6 targets showed druggable properties.

Implications for Drug and Vaccine Development

The 53 targets identified through the current study for vaccine and drug design against *S. flexneri* will make inroad for the development of effective drug(s) and vaccine(s) (**Figure 3**). Among the 53 targets 43 were found as novel drug targets (**Figure 3** and **Table 6**). The development of such therapeutics

FIGURE 2 | Interactome analysis of the final metabolic proteins.

TABLE 4 | Functionality analysis of Hypothetical proteins.

Name of hypothetical protein	Gene ontology	Functional domain	Molecular function	Biological process
NP_839521.1	GO:0055085	Major facilitator superfamily (4–385)	None predicted	Transmembrane transport
NP_837604.1	GO:0008237	Metallopeptidase, catalytic domain (44–257)	Metallopeptidase activity	None predicted
NP_837438.1	None predicted	Duplicated Hybrid motif (169–408)	None predicted	None predicted
NP_836675.1	GO:0048037	NAD(P)-binding domain, coA-binding protein	Cofactor binding	None predicted
AAP19547.1	GO:0009401	Panther (21–480)	None predicted	phosphoenolpyruvate-dependent sugar phosphotransferase system
AAP16677.1	None predicted	Putative zinc- or iron-chelating domain containing protein	None predicted	None predicted

might be targeted based on their qualitative characteristics. *S. flexneri* specific unique metabolic pathway might be a target for drug development (**Table 1**). Drug and vaccine development might also consider cellular location of the targets: Bacterial surface appears to be important for the immunogenicity and surface proteins are more accessible for vaccination.

Therefore, four (4) surface proteins NP_837438.1, EFS15865.1, EFS15439.1, and EFS11306.1 might be targeted for the vaccine design.

However, after identification of this vaccine targets we have analyzed their probable antigen and allergenecity assessment. And we have found NP_837438.1 could be an

TABLE 5 | Predicted Binding site for hypothetical proteins and their interacting residues.

Hypothetical Protein	Cscore[LB]	PDB Hit	TM-score	RMSD[a]	IDEN[a]	Cov.	BS-score	Lig. Name	Predicted binding site residues
NP_839521.1	0.01	3mk7D	0.335	7.03	0.035	0.628	0.84	CA	64, 66, 71, 115
	0.01	2z1qA	0.386	5.90	0.086	0.575	0.62	FAD	34, 37, 38, 45, 94
	0.01	1n38A	0.355	6.67	0.037	0.575	0.73	CH1	277, 278, 309, 312, 313, 314, 375
	0.01	2fonB	0.386	5.96	0.084	0.567	0.62	FAD	16, 77, 78, 118, 120, 123
NP_837604.1	0.74	3khiA	0.801	0.18	0.807	0.802	1.58	ZN	124, 161, 165, 224
	0.01	3kllA	0.428	6.35	0.054	0.741	0.88	MAL	217, 226, 227, 228, 229
NP_837438.1	0.24	2gu1A	0.721	0.56	0.417	0.725	1.48	ZN	314, 318, 395
	0.01	1fiqA	0.143	5.82	0.032	0.207	0.87	FES	345, 346, 363, 384, 385, 386, 387, 392
	0.01	2ckjC	0.329	6.76	0.048	0.523	0.83	FES	345, 346, 363, 384, 386, 387, 389, 392
	0.01	2ckjA	0.330	6.76	0.048	0.523	0.88	FES	330, 331, 332, 334, 350, 351, 358, 359
	0.01	3eubS	0.160	5.78	0.060	0.232	0.91	FES	331, 332, 333, 352, 353, 356, 357
NP_836675.1	0.45	3q9uA	0.692	2.46	0.409	0.826	1.38	COA	37, 38, 40, 42, 45, 46, 64, 65, 66, 67, 68, 96, 97, 100, 104, 119, 120, 143, 145, 146
	0.03	1pgqA	0.603	3.22	0.115	0.774	0.93	2AM	63, 64, 65, 94, 95, 104, 108
AAP19547.1	0.01	3eubS	0.163	4.76	0.055	0.211	0.84	FES	352, 353, 377, 378, 379, 382, 384
	0.01	2ckjC	0.330	7.93	0.031	0.579	0.86	FES	384, 386, 387, 390, 401, 402, 403, 404
	0.01	2ckjB	0.314	7.20	0.039	0.504	0.83	FES	225, 387, 389, 392, 394, 399
	0.01	2ckjA	0.324	7.81	0.028	0.558	0.86	FES	41, 42, 43, 47, 65, 66, 67, 68
AAP16677.1	0.57	3o7bA	0.646	3.48	0.145	0.903	0.90	SAH	64, 65, 66, 83, 84, 86, 98, 111, 112, 113, 114, 115, 116, 118, 121

Here, Cscore[LB] (Confidence score of predicted binding site) predicts a reliable ligand-binding site ranging in between [0 and 1]. BS-score measures between template binding site and predicted binding site. BS-score >1 reflects a significant binding site between the predicted and template. TM-score measures the global structural similarity between query and template protein. RMSD[a] ensures the structurally aligned by TM-align. IDEN[a] measures the percentage of sequence identity in the structurally aligned region. Cov. embodies the coverage of global structural alignment.

ideal target for the epitope based peptide vaccine as it showed probable antigenecity and non-allergenecity (Supplementary Table, S4).

Membrane proteins are the gateways to the cell: many nutrients, ions, waste products, and even DNA and proteins enter and leave cells via proteins which are tightly controlled, maintaining the integrity of the cell. Drugs often target membrane proteins; therefore, understanding their molecular structure helps us design better drugs to cure diseases. We have found 14 targets identified through this analysis can be suitable for drug design (**Table 2**).

Besides, Membrane localized proteins are sometimes difficult to purify and assay (Duffield et al., 2010) and therefore, cytoplasmic proteins are more favorable as drug targets. We have found 35 targets which might be suitable for drug design (**Table 2**).

A promising drug target should have better physicochemical properties such as increased hydrophobicities, *in vivo* half-lives, propensity for being membrane bound and the stability of protein (Bull and Doig, 2015). Our analysis also revealed 14 targets (NP_836827.2, NP_839521.1, NP_837679.1, NP_837676.1, NP_837444.1, NP_837443.1, NP_836948.1, NP_836675.1, NP_835770.1, AAP19547.1,

EFS14933.1, EFS13325.1, EFS11623.1, EFS10693.1) among the 49 suitable drug targets which showed protein stability, increased half-life propensity, greater hydrophobicity etc. (Supplementary Table, S5). We observed 13 targets as novel candidate where only NP_835770.1 target showed druggable (**Figure 3**).

Also, the functions of each identified targets including hypothetical targets may be exploited for the identification or design drug/vaccine for effective therapy against *S. flexneri*. In the perspectives of broadspectrum analysis we have identified some targets which might be used as "promising broadspectrum targets" as they cover the targets more than 40–50 pathogenic bacteria including *S. flexneri*. From the 13 novel drug targets we investigated NP_837443.1 as the most suitable drug candidate which covers the target similarity of 40 pathogenic bacteria including Shigella (**Figure 3**). The identification of the suitable target(s) is the key device for drug discovery. Our identified promising target (NP_837443.1) plays an important role in Holliday junction which exchanges the segments of genetic information during recombination as well as Helicase activity for the survival of *S. flexneri*. For the drug discovery against this target we need to identify the conserved domain responsible for Holliday junction. The effective binding site and its active

TABLE 6 | Druggability properties of metabolic human non-homologous essential proteins.

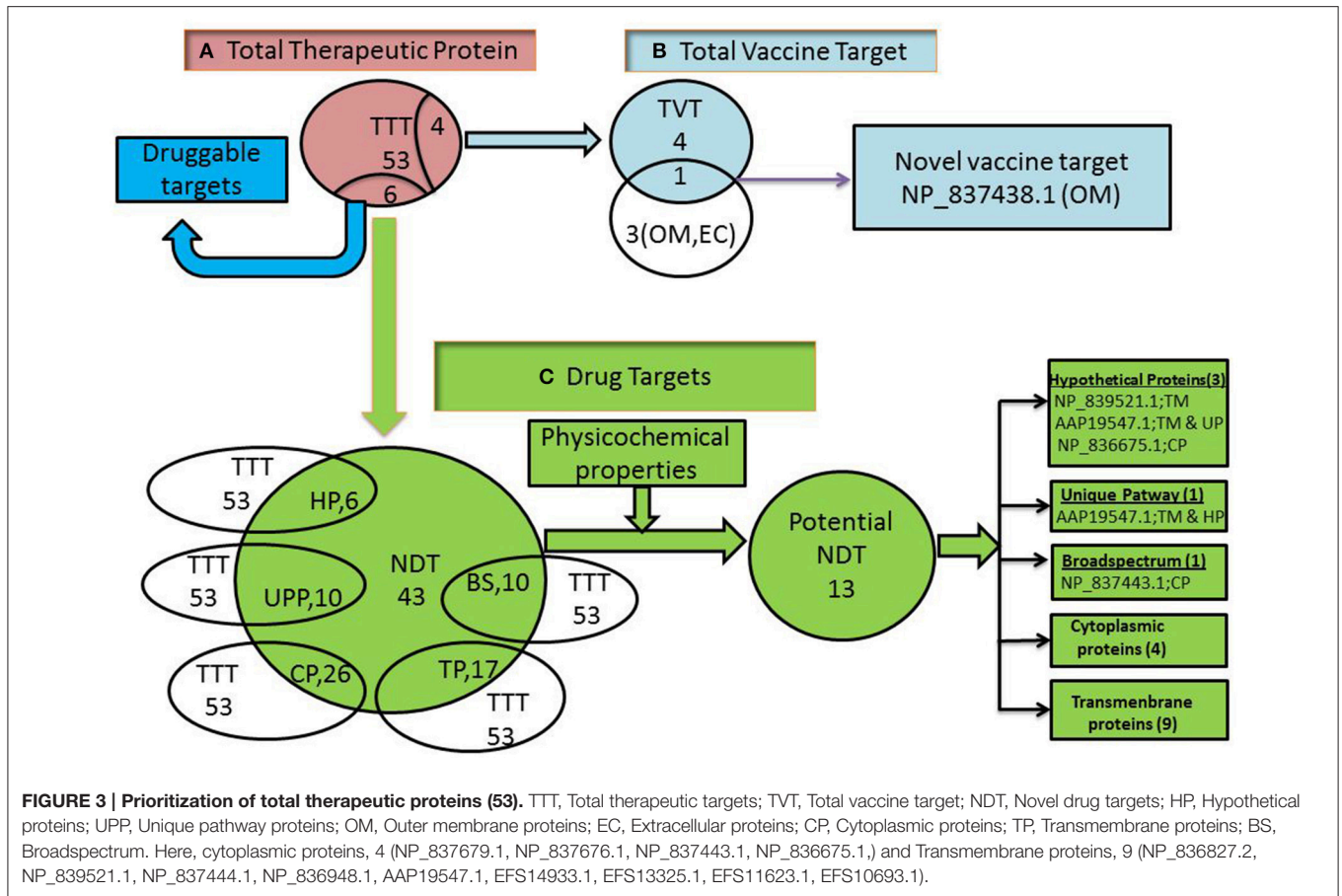| Protein ID	DrugBank ID	Drug name	Drug group	Pharmacological action	Docking energy (Kcal/mol)	Chemical formula
NP_835873.1	DB00233	Aminosalicylic Acid	Approved	Unknown	−7.8	$C_7H_7NO_3$
NP_835770.1	DB04267	Dipicolinic Acid	Experimental	Unknown	−8.1	$C_7H_5NO_4$
AAP18497.1	DB00259	Sulfanilamide	Approved	Inhibitor	−7.1	$C_6H_8N_2O_2S$
	DB00264	Sulfisoxazole	Approved, Vet_approved	Inhibitor	−7.5	$C_{11}H_{13}N_3O_3S$
	DB00576	Sulfamethizole	Approved, Vet_approved	Inhibitor	−8.7	$C_9H_{10}N_4O_2S_2$
	DB00634	Sulfacetamide	Approved	Inhibitor	−6.9	$C_8H_{10}N_2O_3S$
	DB01015	Sulfamethoxazole	Approved	Inhibitor	−8.0	$C_{10}H_{11}N_3O_3S$
	DB01298	Sulfacytine	Approved	Inhibitor	−7.8	$C_{12}H_{14}N_4O_3S$
	DB01581	Sulfamerazine	Approved, Vet_approved	Inhibitor	−8.8	$C_{11}H_{12}N_4O_2S$
	DB01582	Sulfamethazine	Approved, Vet_approved	Inhibitor	−6.5	$C_{12}H_{14}N_4O_2S$
	DB06729	Sulfaphenazole	Approved	Inhibitor	−7.0	$C_{15}H_{14}N_4O_2S$
EFS15406.1	DB00274	Cefmetazole	Approved	Inhibitor	−8.3	$C_{15}H_{17}N_7O_5S_3$
	DB00303	Ertapenem	Approved, Investigational	Inhibitor	−8.4	$C_{22}H_{25}N_3O_7S$
	DB00430	Cefpiramide	Approved	Inhibitor	−6.9	$C_{25}H_{24}N_8O_7S_2$
	DB00438	Ceftazidime	Approved	Inhibitor	−8.0	$C_{22}H_{22}N_6O_7S_2$
	DB01327	Cefazolin	Approved	Inhibitor	−7.7	$C_{14}H_{14}N_8O_4S_3$
	DB01328	Cefonicid	Approved	Inhibitor	−6.8	$C_{18}H_{18}N_6O_8S_3$
	DB01329	Cefoperazone	Approved	Inhibitor	−7.8	$C_{25}H_{27}N_9O_8S$
	DB01331	Cefoxitin	Approved	Inhibitor	−7.9	$C_{16}H_{17}N_3O_7S_2$
	DB01332	Ceftizoxime	Approved	Inhibitor	−6.7	$C_{13}H_{13}N_5O_5S_2$
	DB01333	Cefradine	Approved	Inhibitor	−7.2	$C_{16}H_{19}N_3O_4S$
	DB01414	Cefacetrile	Approved	Inhibitor	−7.0	$C_{13}H_{13}N_3O_6S$
	DB01415	Ceftibuten	Approved	Inhibitor	−6.1	$C_{15}H_{14}N_4O_6S$
	DB01598	Imipenem	Approved	Inhibitor	−6.6	$C_{12}H_{17}N_3O_4S$
	DB04570	Latamoxef	Approved	Inhibitor	−7.9	$C_{20}H_{20}N_6O_9S$
	DB06211	Doripenem	Approved, Investigational	Antagonist, Inhibitor	−6.8	$C_{15}H_{24}N_4O_6S$
EFS13874.1	DB02767	3-Hydroxy-Myristic Acid	Experimental	Unknown	−7.2	$C_{14}H_{28}O_3$
	DB04147	Lauryl Dimethylamine-N-Oxide	Experimental	Unknown	−6.9	$C_{14}H_{31}NO$
	DB08231	MYRISTIC ACID	Experimental	Unknown	−7.5	$C_{14}H_{28}O$
EFS11712.1	DB01862	Isopropyl beta-D-thiogalactopyranoside	Experimental	Unknown	−6.0	$C_9H_{18}O_5S$
	DB08297	ORTHONITROPHENYL-BETA-D-FUCOPYRANOSIDE	Experimental	Unknown	−7.2	$C_{12}H_{15}NO_7$

residues of this domain should also be identified for the drug design. Finally, drug-target binding affinity of this binding site, ADMET (Absorption, Distribution, Metabolism, and Excretion), QSAR (Quantitative Structural Activity Relationship), and toxicity measurement will lead the way to inhibit the function of this target protein. Further, animal model experiment must be needed to confirm the drug efficacy against this target.

As the selection of effective therapeutics targets are the crucial things for the vaccine and drug design (Hasan et al., 2015; Khan et al., 2015; Oany et al., 2015; Hossain et al., 2016a,b; Hossain and Oany, 2016d) as well as for the building of networking (protein-protein interaction; Hossain et al., 2016c),

our subtractive strategy will have a great impact in therapeutics discovery against *S. flexneri*.

CONCLUSION

Subtractive channel analysis in proteome of *S. flexneri* 2a str. 2445T suggest 53 drug targets and further with their qualitative characterization stage explicates unique metabolic pathways, localization of targets, broadspectrum, functionality, interactome, and druggable properties. This study might facilitate the development of drug and vaccines against *S. flexneri* as well as come to be handy in clinical interest with identification of drug candidates against other pathogens.

FIGURE 3 | Prioritization of total therapeutic proteins (53). TTT, Total therapeutic targets; TVT, Total vaccine target; NDT, Novel drug targets; HP, Hypothetical proteins; UPP, Unique pathway proteins; OM, Outer membrane proteins; EC, Extracellular proteins; CP, Cytoplasmic proteins; TP, Transmembrane proteins; BS, Broadspectrum. Here, cytoplasmic proteins, 4 (NP_837679.1, NP_837676.1, NP_837443.1, NP_836675.1,) and Transmembrane proteins, 9 (NP_836827.2, NP_839521.1, NP_837444.1, NP_836948.1, AAP19547.1, EFS14933.1, EFS13325.1, EFS11623.1, EFS10693.1).

AUTHOR CONTRIBUTIONS

MS: Conceived, designed, and guided the study, analyzed the data, helped in drafting and performed critical revision. CK: Guided the study, acquisition and analyzed the data, helped in drafting the manuscript. MM and AH: Guided the study, analyzed the data, and helped in drafting the manuscript. MK and MI: Helped in Bioinformatics analysis and drafted the manuscript. MH: Helped to design the study, performed bioinformatics analysis, drafted, and developed the manuscript and performed critical revision. All authors have approved the manuscript.

SUPPLEMENTARY MATERIAL

REFERENCES

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. doi: 10.1016/S0022-2836(05)80360-2

Anishetty, S., Pulimi, M., and Pennathur, G. (2005). Potential drug targets in *Mycobacterium tuberculosis* through metabolic pathway analysis. *Comput. Biol. Chem.* 29, 368–378. doi: 10.1016/j.compbiolchem.2005.07.001

Asif, S. M., Asad, A., Faizan, A., Anjali, M. S., Arvind, A., Neelesh, K., et al. (2009). Dataset of potential targets for *Mycobacterium tuberculosis* H37Rv through comparative genome analysis. *Bioinformation* 4, 245–248. doi: 10.6026/97320630004245

Bardhan, P., Faruque, A. S., Naheed, A., and Sack, D. A. (2010). Decrease in shigellosis related deaths without Shigella spp.-specific interventions, Asia. *Emerging Infect. Dis.* 16, 1718–1723. doi: 10.3201/eid1611.090934

Barh, D., and Kumar, A. (2009). *In silico* identification of candidate drug and vaccine targets from various pathways in *Neisseria gonorrhoeae*. *In Silico Biol.* 9, 225–231. doi: 10.3233/ISB-2009-0399

Barh, D., Tiwari, S., Jain, N., Ali, A., Santos, A. R., Misra, A. N., et al. (2011). *In silico* subtractive genomics for target identification in human bacterial pathogens. *Drug Dev. Res.* 72, 162–177. doi: 10.1002/ddr.20413

Black, R. E., Cousens, S., Johnson, H. L., Lawn, J. E., Rudan, I., Bassani, D. G., et al. (2010). Global, regional, and national causes of child mortality in 2008: a systematic analysis. *Lancet* 375, 1969–1987. doi: 10.1016/S0140-6736(10)60549-1

Briken, V. (2008). Molecular mechanisms of host-pathogen interactions and their potential for the discovery of new drug targets. *Curr. Drug Targets* 9, 150–157. doi: 10.2174/138945008783502449

Bull, S. C., and Doig, A. J. (2015). Properties of protein drug target classes. *PLoS ONE* 10:e0117955. doi:10.1371/journal.pone.0117955

Butt, A. M., Nasrullah, I., Tahir, S., and Tong, Y. (2012). Comparative genomics analysis of mycobacterium ulcerans for the identification of putative essential genes and therapeutic candidates. *PLoS ONE* 7:e43080. doi: 10.1371/journal.pone.0043080

Coster, T. S., Hoge, C. W., VanDeVerg, L. L., Hartman, A. B., Oaks, E. V., Venkatesan, M. M., et al. (1999). Vaccination against shigellosis with attenuated *Shigella flexneri* 2a strain SC602. *Infect. Immun.* 67, 3437–3443.

Duffield, M., Cooper, I., McAlister, E., Bayliss, M., Ford, D., and Oyston, P. (2010). Predicting conserved essential genes in bacteria: *in silico* identification of putative drug targets. *Mol. Biosyst.* 6, 2482–2489. doi: 10.1039/c0mb00001a

Dutta, A., Singh, S. K., Ghosh, P., Mukherjee, R., Mitter, S., and Bandyopadhyay, D. (2006). *In silico* identification of potential therapeutic targets in the human pathogen *Helicobacter pylori*. *In Silico Biol.* 6, 43–47.

Eisenberg, D., Lüthy, R., and Bowie, J. U. (1997). [20] VERIFY3D: assessment of protein models with three-dimensional profiles. *Methods Enzymol.* 277, 396–404. doi: 10.1016/S0076-6879(97)77022-8

Foster, R. A., Carlin, N. I. A., Majcher, M., Tabor, H., Ng, L. K., and Widmalm, G. (2011). Structural elucidation of the O-antigen of the *Shigella flexneri* provisional serotype 88–893: structural and serological similarities with *S. flexneri* provisional serotype Y394 (1c). *Carbohydr. Res.* 346, 872–876. doi: 10.1016/j.carres.2011.02.013

Fujimura, K. E., Slusher, N. A., Cabana, M. D., and Lynch, S. V. (2010). Role of the gut microbiota in defining human health. *Expert Rev. Anti Infect. Ther.* 8, 435–454. doi: 10.1586/eri.10.14

Galperin, M. Y., and Koonin, E. V. (1999). Searching for drug targets in microbial genomes. *Curr. Opin. Biotechnol.* 10, 571–578. doi: 10.1016/S0958-1669(99)00035-X

Griffith, D. E., Aksamit, T., Brown-Elliott, B. A., Catanzaro, A., Daley, C., Gordin, F., et al. (2007). An official ATS/IDSA statement: diagnosis, treatment, and prevention of nontuberculous mycobacterial diseases. *Am. J. Respir. Crit. Care Med.* 175, 367–416. doi: 10.1164/rccm.200604-571ST

Haag, N. L., Velk, K. K., and Wu, C. (2012). *In silico* identification of drug targets in methicillin/multidrug-resistant Staphylococcusaureus. *Int. J. Adv. Life Sci.* 4, 21–32.

Hasan, M. A., Khan, M. A., Datta, A., Mazumder, M. H. H., and Hossain, M. U. (2015). A comprehensive immunoinformatics and target site study revealed the corner-stone towards Chikungunya virus treatment. *Mol. Immunol.* 65, 189–204. doi: 10.1016/j.molimm.2014.12.013

Hasan, M. A., Khan, M. A., Sharmin, T., Mazumder, M. H. H., and Chowdhury, A. S. (2016). Identification of putative drug targets in Vancomycin-resistant *Staphylococcus aureus* (VRSA) using computer aided protein data analysis. *Gene* 575, 132–143. doi: 10.1016/j.gene.2015.08.044

Hossain, M. U., Hashem, A., Keya, C. A., and Salimullah, M. (2016a). Therapeutics insight with inclusive immunopharmacology explication of human rotavirus a for the treatment of Diarrhea. *Front. Pharmacol.* 7:153. doi: 10.3389/fphar.2016.00153

Hossain, M. U., Khan, M. A., Rakib-Uz-Zaman, S. M., Ali, M. T., Islam, M. S., Keya, C. A., et al. (2016b). Treating diabetes mellitus: pharmacophore based designing of potential drugs from *Gymnema sylvestre* against insulin receptor protein. *Biomed Res. Int.* 14:3187647. doi: 10.1155/2016/3187647

Hossain, M. U., Oany, A. R., Ahmad, S. A. I., Hasan, M. A., Khan, M. A., and Siddikey, M. A. A. (2016d). Identification of potential inhibitor and enzyme-inhibitor complex on trypanothione reductase to control Chagas disease. *Comput. Biol. Chem.* 65, 29–36. doi: 10.1016/j.compbiolchem.2016.10.002

Hossain, M. U., Shibly, A. Z., Omar, T. M., Zohora, F. T., Santona, U. S., Hossain, M. J., et al. (2016c). Towards finding the linkage between metabolic and age-related disorders using semantic gene data network. *Bioinformation* 12, 22–26. doi: 10.6026/97320630012022

Huang, Y., Niu, B., Gao, Y., Fu, L., and Li, W. (2010). CD-HIT suite: a web server for clustering and comparing biological sequences. *Bioinformatics* 26, 680–682. doi: 10.1093/bioinformatics/btq003

Jennison, A. V., and Verma, N. K. (2004). *Shigella flexneri* infection: pathogenesis and vaccine development. *FEMS Microbiol. Rev.* 28, 43–58. doi: 10.1016/j.femsre.2003.07.002

Jensen, L. J., Kuhn, M., Stark, M., Chaffron, S., Creevey, C., Muller, J., et al. (2009). STRING 8–a global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Res.* 37, D412–D416. doi: 10.1093/nar/gkn760

Käll, L., Krogh, A., and Sonnhammer, E. L. (2007). Advantages of combined transmembrane topology and signal peptide prediction—the Phobius web server. *Nucleic Acids Res.* 35, W429–W432. doi: 10.1093/nar/gkm256

Kanehisa, M., Goto, S., Sato, Y., Kawashima, M., Furumichi, M., and Tanabe, M. (2014). Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Res.* 42, 199–205. doi: 10.1093/nar/gkt1076

Katz, D. E., Coster, T. S., Wolf, M. K., Trespalacios, F. C., Cohen, D., Robins, G., et al. (2004). Two studies evaluating the safety and immunogenicity of a live, attenuated *Shigella flexneri* 2a vaccine (SC602) and excretion of vaccine organisms in North American volunteers. *Infect. Immun.* 72, 923–930. doi: 10.1128/IAI.72.2.923-930.2004

Kärnell A., Li, A., Zhao, C. R., Karlsson, K., Nguyen, B. M., and Lindberg, A. A. (1995). Safety and immunogenicity of the auxotrophic *Shigella flexneri* 2a vaccine SFL1070 with a deleted aroD gene in adult Swedish volunteers. *Vaccine* 13, 88–89. doi: 10.1016/0264-410X(95)80017-8

Khan, M. A., Hossain, M. U., Rakib-Uz-Zaman, S. M., and Morshed, M. N. (2015). Epitope -based peptide vaccine design and target site depiction against Ebola viruses: an immunoinformatics study. *Scand. J. Immunol.* 82, 25–34. doi: 10.1111/sji.12302

Knox, C., Law, V., Jewison, T., Liu, P., Ly, S., Frolkis, A., et al. (2011). DrugBank 3.0: a comprehensive resource for 'omics' research on drugs. *Nucleic Acids Res.* 39, D1035–D1041. doi: 10.1093/nar/gkq1126

Kotloff, K. L., Winickoff, J. P., Ivanoff, B., Clemens, J. D., Swerdlow, D. L., Sansonetti, P. J., et al. (1999). Global burden of Shigella infections: implications for vaccine development and implementation of control strategies. *Bull. World Health Organ.* 77, 651–666.

Kumar, G. S., Sarita, S., Kumar, G. M., Pant, K. K., and Seth, P. K. (2010). Definition of potential targets in Mycoplasma Pneumoniae through subtractive genome analysis. *J. Antivir. Antiretrovir.* 2, 038–041. doi: 10.4172/jaa.1000020

Kushwaha, S. K., and Shakya, M. (2010). Protein interaction network analysis–approach for potential drug target identification in *Mycobacterium tuberculosis*. *J. Theor. Biol.* 262, 284–294. doi: 10.1016/j.jtbi.2009.09.029

Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* 26, 283–291. doi: 10.1107/S0021889892009944

Luo, X., Sun, Q., Lan, R., Wang, J., Li, Z., Xia, S., et al. (2012). Emergence of a novel *Shigella flexneri* serotype 1d in China. *Diagn. Microbiol. Infect. Dis.* 74, 316–319. doi: 10.1016/j.diagmicrobio.2012.06.022

Mondal, S. I., Ferdous, S., Jewel, N. A., Akter, A., Mahmud, Z., Islam, M. M., et al. (2015). Identification of potential drug targets by subtractive genome analysis of *Escherichia coli* O157: H7: an *in silico* approach. *Adv. Appl. Bioinform. Chem.* 8, 49. doi: 10.2147/AABC.S88522

Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A. C., and Kanehisa, M. (2007). KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res.* 35, W182–W185. doi: 10.1093/nar/gkm321

Mukhopadhaya, A., Mahalanabis, D., and Chakrabarti, M. K. (2006). Role of *Shigella flexneri* 2a 34 kDa outer membrane protein in induction of protective immune response. *Vaccine* 24, 6028–6036. doi: 10.1016/j.vaccine.2006.03.026

Mukhopadhaya, A., Mahalanabis, D., Khanam, J., and Chakrabarti, M. K. (2003). Protective efficacy of oral immunization with heat-killed *Shigella flexneri* 2a in animal model: study of cross protection, immune response and antigenic recognition. *Vaccine* 21, 3043–3050. doi: 10.1016/S0264-410X(03)00111-7

Mulder, N., and Apweiler, R. (2007). InterPro and InterProScan: tools for protein sequence classification and comparison. *Methods Mol. Biol.* 396, 59–70. doi: 10.1007/978-1-59745-515-2_5

Nessar, R., Cambau, E., Reyrat, J. M., Murray, A., and Gicquel, B. (2012). *Mycobacterium abscessus*: a new antibiotic nightmare. *J. Antimicrob. Chemother.* 67, 810–818. doi: 10.1093/jac/dkr578

Oany, A. R., Ahmad, S. A. I., Hossain, M. U., and Jyoti, T. P. (2015). Highly conserved antigenic epitope regions in RNA dependent RNA polymerase-L of Crimean-Congo haemorrhagic fever virus: insights about novel vaccine. *Adv. Appl. Bioinform. Chem.* 8, 1–10. doi: 10.2147/AABC.S75250

Paterson, D. L. (2006). Resistance in gram-negative bacteria: Enterobacteriaceae. *Am. J. Med.* 119, S20–S28. doi: 10.1016/j.amjmed.2006.03.013

Peng, X., Luo, W., Zhang, J., Wang, S., and Lin, S. (2002). Rapid detection of Shigella species in environmental sewage by an immunocapture PCR with universal primers. *Appl. Environ. Microbiol.* 68, 2580–2583. doi: 10.1128/AEM.68.5.2580-2583.2002

Perepelov, A. V., Shekht, M. E., Liu, B., Shevelev, S. D., Ledov, V. A., Sof'ya, N. S., et al. (2012). *Shigella flexneri* O-antigens revisited: final elucidation of the O-acetylation profiles and a survey of the O-antigen structure diversity. *FEMS Immunol. Med. Microbiol.* 66, 201–210. doi: 10.1111/j.1574-695X.2012.01000.x

Petersen, T. N., Brunak, S., von Heijne, G., and Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods* 8, 785–786. doi: 10.1038/nmeth.1701

Rabizadeh, S., and Sears, C. (2008). New horizons for the infectious diseases specialist:how gut microflora promote health and disease. *Curr. Infect. Dis. Rep.* 10, 92–98. doi: 10.1007/s11908-008-0017-8

Rahman, M. A., Noore, M. S., Hasan, M. A., Ullah, M. R., Rahman, M. H., Hossain, M. A., et al. (2014). Identification of potential drug targets by subtractive genome analysis of *Bacillus anthracis* A0248: an *in silico* approach. *Comput. Biol. Chem.* 52, 66–72. doi: 10.1016/j.compbiolchem.2014.09.005

Raman, K., Yeturu, K., and Chandra, N. (2008). targetTB: a target identification pipeline for *Mycobacterium tuberculosis* through an interactome, reactome and genome-scale structural analysis. *BMC Syst. Biol.* 2:109. doi: 10.1186/1752-0509-2-109

Ranallo, R. T., Fonseka, C. P., Cassels, F., Srinivasan, J., and Venkatesan, M. M. (2005). Construction and characterization of bivalent *Shigella flexneri* 2a vaccine strains SC608(pCFAI) and SC608(pCFAI/LTB) that express antigens from enterotoxigenic *Escherichia coli*. *Infect. Immun.* 73, 258–267. doi: 10.1128/IAI.73.1.258-267.2005

Roy, A., Yang, J., and Zhang, Y. (2012). COFACTOR: an accurate comparative algorithm for structure-based protein function annotation. *Nucleic Acids Res.* 40, W471–W477. doi: 10.1093/nar/gks372

Sakharkar, K. R., Sakharkar, M. K., and Chow, V. T. (2004). A novel genomics approach for the identification of drug targets in pathogens, with special reference to *Pseudomonas aeruginosa*. *In Silico Biol.* 4, 355–360.

Sarkar, M., Maganti, L., Ghoshal, N., and Dutta, C. (2012). *In silico* quest for putative drug targets in *Helicobacter pylori* HPAG1: molecular modeling of candidate enzymes from lipopolysaccharide biosynthesis pathway. *J. Mol. Model.* 18, 1855–1866. doi: 10.1007/s00894-011-1204-3

Shanmugam, A., and Natarajan, J. (2010). Computational genome analyses of metabolic enzymes in *Mycobacterium leprae* for drug target identification. *Bioinformation* 4, 392–395. doi: 10.6026/97320630004392

Sharma, A., and Pan, A. (2012). Identification of potential drug targets in *Yersinia pestis* using metabolic pathway analysis: MurE ligase as a case study. *Eur. J. Med. Chem.* 57C, 185–195. doi: 10.1016/j.ejmech.2012.09.018

Shenai, S., Rodrigues, C., and Mehta, A. (2009). Rapid speciation of 15 clinically relevant mycobacteria with simultaneous detection of resistance to rifampin, isoniazid, and streptomycin in *Mycobacterium tuberculosis* complex. *Int. J. Infect. Dis.* 13, 46–58. doi: 10.1016/j.ijid.2008.03.025

Simmons, D. A., and Romanowska, E. (1987). Structure and biology of *Shigella flexneri* O antigens. *J. Med. Microbiol.* 23, 289–302. doi: 10.1099/00222615-23-4-289

Singh, S., Malik, B. K., and Sharma, D. K. (2007). Metabolic pathway analysis of *S. pneumoniae*: an *in silico* approach towards *drug-design*. *J. Bioinform. Comput. Biol.* 5, 135–153. doi: 10.1142/S0219720007002564

Stagg, R. M., Tang, S.-S., Carlin, N. I. A., Talukder, K. A., Cam, P. D., and Verma, N. K. (2009). A novel glucosyltransferase involved in O-antigen modification of *Shigella flexneri* serotype 1c. *J. Bacteriol.* 191, 6612–6617. doi: 10.1128/JB.00628-09

Sun, Q., Knirel, Y. A., Lan, R., Wang, J., Senchenkova, S. N., Jin, D., et al. (2012). A novel plasmid-encoded serotype conversion mechanism through addition of phosphoethanolamine to the O-antigen of *Shigella flexneri*. *PLoS ONE* 7:e46095. doi: 10.1371/journal.pone.004609

Sun, Q., Lan, R., Wang, Y., Wang, J., Luo, X., Zhang, S., et al. (2011). Genesis of a novel *Shigella flexneri* serotype by sequential infection of serotype-converting bacteriophages SfX and SfI. *BMC Microbiol.* 11:269. doi: 10.1186/1471-2180-11-269

Szklarczyk, D., Franceschini, A., Kuhn, M., Simonovic, M., Roth, A., Minguez, P., et al. (2011). The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Res.* 39, D561–D568. doi: 10.1093/nar/gkq973

Trott, O., and Olson, A. J. (2010). AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* 31, 455–461. doi: 10.1002/jcc.21334

Vetrivel, U., Subramanian, G., and Dorairaj, S. (2011). A novel *in silico* approach to identify potential therapeutic targets in human bacterial pathogens. *HUGO J.* 5, 25–34. doi: 10.1007/s11568-011-9152-7

Wang, F., Xiao, J., Pan, L., Yang, M., Zhang, G., Jin, S., et al. (2008). A systematic survey of mini-proteins in bacteria and archaea. *PLoS ONE* 3:e4027. doi: 10.1371/journal.pone.0004027

Webb, B., and Sali, A. (2014). Comparative protein structure modeling using modeller. *Curr. Protoc. Bioinformatics* 47, 5.6.1–5.6.32. doi: 10.1002/0471250953.bi0506s47

Wiederstein, M., and Sippl, M. J. (2007). ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res.* 35(Suppl. 2), W407–W410. doi: 10.1093/nar/gkm290

Wu, S., and Zhang, Y. (2007). LOMETS: A local meta-threading-server for protein structure prediction. *Nucleic Acids Res.* 35, 3375–3338. doi: 10.1093/nar/gkm251

Ye, C., Lan, R., Xia, S., Zhang, J., Sun, Q., Zhang, S., et al. (2010). Emergence of a new multidrug-resistant serotype X variant in an epidemic clone of *Shigella flexneri*. *J. Clin. Microbiol.* 48, 419–426. doi: 10.1128/JCM.00614-09

Yeh, I., Hanekamp, T., Tsoka, S., Karp, P. D., and Altman, R. B. (2004). Computational analysis of *Plasmodium falciparum* metabolism: organizing genomic information to facilitate drug discovery. *Genome Res.* 14, 917–924. doi: 10.1101/gr.2050304

Yu, C. S., Chen, Y. C., Lu, C. H., and Hwang, J. K. (2006). Prediction of protein subcellular localization. *Proteins* 64, 643–651. doi: 10.1002/prot.21018

Yu, N. Y., Wagner, J. R., Laird, M. R., Melli, G., Rey, S., Lo, R., et al. (2010). PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics* 26, 1608–1615. doi: 10.1093/bioinformatics/btq249

Zhang, R., and Lin, Y. (2009). DEG 5.0, a database of essential genes in both prokaryotes and eukaryotes. *Nucleic Acids Res.* 37, D455–D458. doi: 10.1093/nar/gkn858