EUROPEAN JOURNAL OF BIOMEDICAL AND PHARMACEUTICAL SCIENCES
http://www.ejbps.com

ISSN 2349-8870
Volume: 5
Issue: 5
736-747
Year: 2018

IN SILICO IDENTIFICATION AND CHARACTERIZATION OF POTENTIAL DRUG TARGETS IN HUMAN PATHOGEN SHIGELLA FLEXNERI

***Desai Sharav and [1] Meshram Dhananjay**

*Department of Pharmaceutical Microbiology and Biotechnology, Pioneer Pharmacy Degree College, Vadodara, Gujarat, India.
[1]Department of Pharmaceutical Quality Assurance, Pioneer Pharmacy Degree College, Vadodara, Gujarat, India.

***Corresponding Author: Dr. Desai Sharav**
Department of Pharmaceutical Microbiology and Biotechnology, Pioneer Pharmacy Degree College, Vadodara, Gujarat, India.

Article Received on 05/03/2018 **Article Revised on 26/03/2018** **Article Accepted on 15/04/2018**

ABSTRACT
Shigellosis is a significant cause of morbidity and mortality in children amongst developing countries. Newly emerging Antibiotic-resistant strains from all over the world are also one of the alarming conditions in the treatment and management of infection clinically. Such conditions calls for immediate investigation of all available proteins in the species that can act as a target for controlling the disease. The traditional discovery process can result in an enormous amount of time with effort. In the present work novel *in-silico* ordered approach is used to identify and characterize potential drug target in *Shigella flexneri* species. Three phase of analysis used to screen proteins in the species and at every phase, the proteins were filtered according to define parameters. In phase, I proteins from five sets (Chokepoint, Pathway, Plasmid, Virulence, and resistant) collected from described sources. Total of 1,276 proteins resulted from the phase I analysis. Phase II studies resulted in 220 proteins with several of target characteristics. Phase III studies led to a final list of three proteins namely CytidylateKinase, Ribunuclease and 2-amino-4-hydroxy-6-hydroxymethyl dihydro pteridine pyrophosphokinase. Protein homology modeling and structure refinement studies produced five models for each of the target. ERRAT profile, QMEAN6, Z-Score and distribution of residues using Ramachandran plot were considered as parameters of validation. Structures were analyzed for normal mode analysis using the Bio3D package in R to promote molecular dynamic studies. Proteins resulted can be developed as a potential drug target for infections occurring by *Shigella flexneri*.

KEYWORDS: Computer Simulation, Drug Resistance, Gastrointestinal Microbiota In Silico Models, Protein Interaction Maps, Protein Sequence Analysis, Virulence Factors.

INTRODUCTION
Shigella are gram-negative, Facultatively anarobic, rod shaped enterobacteria causing bacillary dysentery or "Shigellosis" in man and other primates. They share a typical characteristic with Escherichia, and their genetic analogy indicates that they are subtypes of Escherichia.[1] All Shigella invades the colonic mucosa and induce abdominal aches and disturbance. The presence of erythrocytes, PMNs and mucous in the stool forms the support for the diagnosis of shigellosis. Four subgroups (*S. flexneri, S. sonnei, S, dysenteriae, and S. boydii*) represents the Shigella group with many serotypes. Serotypes vary by the composition of lipopolysaccharide gram o- negative membrane.[2] Shigella classified as a leading ranking pathogen and is yet the leading source of childhood morbidity and fatality. Several types of research confirmed that *Shigella flexneri* and *Shigella sonnei* continues the contributing source of shigellosis in the developing countries.[3] In the present article primary focus is done on the *Shigella flexneri* as is endemic to the growing countries, and one of the contributing cause of infection. All subtypes of *S.flexneri* like 2a, 3a, 4a, 6 cause shigellosis, but in the developed countries, 2a found to be a most frequent pathogen. The development of Multi-Drug Resistance strains for Shigella flexneri 2a str. 301 has been published from Iran, Bangladesh, USA, China, Indonesia, Vietnam, and India. Overuse and easy availability of the useful antibiotics also created Resistance in Shigella species. Virulence genes and proteins emerge from both, 220 kb plasmid known as virulence plasmid and Chromosome also encodes virulence proteins like lipopolysaccharide membrane and temperature-dependent expression of virulence genes.[4] There is a list of the proteins involved in the invasion of pathogens in the host, and many proteins are so important that if their role inhibited, the infection could be avoided. It has been proven that traditional Drug target discovery demands a massive amount of time with cost and efforts. All the proteins with amino acid sequences are accessible for examination and every

protein should be considered to overcome as a probable rug target. Use of the computer in the pharmaceutical discovery process ended in time reduction for the development process, extended treatment adequacy, strategy preferences are advancing, the percentage of rug failure reduced. Several aspects of targets like target essentiality, Resistance proteins, host and protein interactions, chokepoint enzymes present in the pathway, pathway analysis can be examined. Because of computational approach several pathogenic microorganisms like Mycobacterium tuberculosis, Mycobacterium Leprae, Helicobacter pylori has been broadly examined and many targets about precision, selectivity, and essentiality has been identified. The present manuscript includes a computational way to identify Drug targets in *Shigella flexneri 2a str. 30*. The investigation ended with three targets and are missing in the host, a prerequisite for the persistence of cell, can be employed as a wide-spectrum target. The investigation further revealed the site of the target in the cell as it is essential to produce targets for vaccine or Drug. As 3D structures were not available for targets also were produced using homology modeling and developed models were improved and verified for further consideration.

MATERIALS AND METHODS

In the staged study, to find and characterize the possible drug target in Shigella flexneri 2a.301 novel In- silico idea is employed. The presented design is constituted of three-stage of evaluation Phase I, incorporated retrieval of proteins especially for Chokepoint, Plasmid, Pathway proteins, Proteins reasoned for virulence and proteins that resist the antibiotics. Phase II studies commenced with the subtractive analysis which performed to eliminate the proteins present in the humans. In Phase III proteins derived from Phase I and Phase II investigations were quantitatively characterized to overcome as a probable drug target. The complete workflow is describing (Figure 1& Figure 2) various steps utilized in the study.

Phase I: Protein Dataset Mining

Choke Point Analysis: Special chokepoint proteins present in the metabolic pathway of Shigella flexneri 2a str. 301 were retrieved utilizing chokepoint analysis. Enzymes engaged in the chokepoint reactions are called as chokepoint proteins. Chokepoint proteins can be in from the producing side or can be of from the consuming side. Stopping either of the reactions or the enzymes will lead to the deprivation of that product or may appear in accumulation of the product or substrate that may generate a lethal result in the organisms. Pathway tools accessible from SRI international utilized to collect the chokepoint proteins from Shigella flexneri 2a str. 301 and human. Comparative study was done to analyze and exclude the shared chokepoint proteins present in both human and Shigella flexneri 2a str. 301. Common proteins were excluded to evade the side returns of drug that may emerge during the treatment.[5-8]

Plasmid Protein Analysis: *Shigella flexneri 2a str. 301* is constituted of 221 kb separate plasmid. The plasmid is consisting of 262 proteins with several of the clinically significant proteins including quinoline resistance, infection invasion gene, and some virulence proteins. It indicates that plasmid proteins may serve as possible drug targets. The complete protein sequence of the analogous plasmid was fetched from the NCBI database with an Accession number NC_004851.1. The sequence received was passed through the subtractive channel of analysis.[1,9]

Pathway analysis: Comparative pathway study conducted for Human and Shigella flexneri 2a str. 301. Details of pathways were achieved utilizing pathway tools from SRI international. The pathway classified as a distinct pathway and common pathway. The pathway present alone in the Shigella flexneri 2a str. 301 was viewed as a distinct pathway, and the protein sequences collected. Similarly pathways present in both human and Shigella flexneri 2a str. 301were recognized as common pathway. Unique proteins were analyzed by comparative study and were fetched from NCBI.[10-12]

Virulence Factor Analysis: Severity of disease and establishment of organism in to host depends on the virulence proteins. Virulence factor for Shigella flexneri 2a str. 301 were gathered from the Virulence factor database (VFDB). VFDB forms of 74 genera of bacteria with 926 bacterial strains and 1796 Virulence factors.[13-16] From the correlation table in the database, four species of Shigella with 13 strains yields 23 virulence factors. Collected proteins were again moved through the subtractive channel of study.

Resistance genes and proteins analysis: Drug resistance is prevalent issue for anti-microbial therapy. Shigella species are likewise seen to develop protection for the antibiotics like quinolones and tetracycline.[17-20]. Proteins which are recognized to give resistance in Shigella flexneri 2a str.

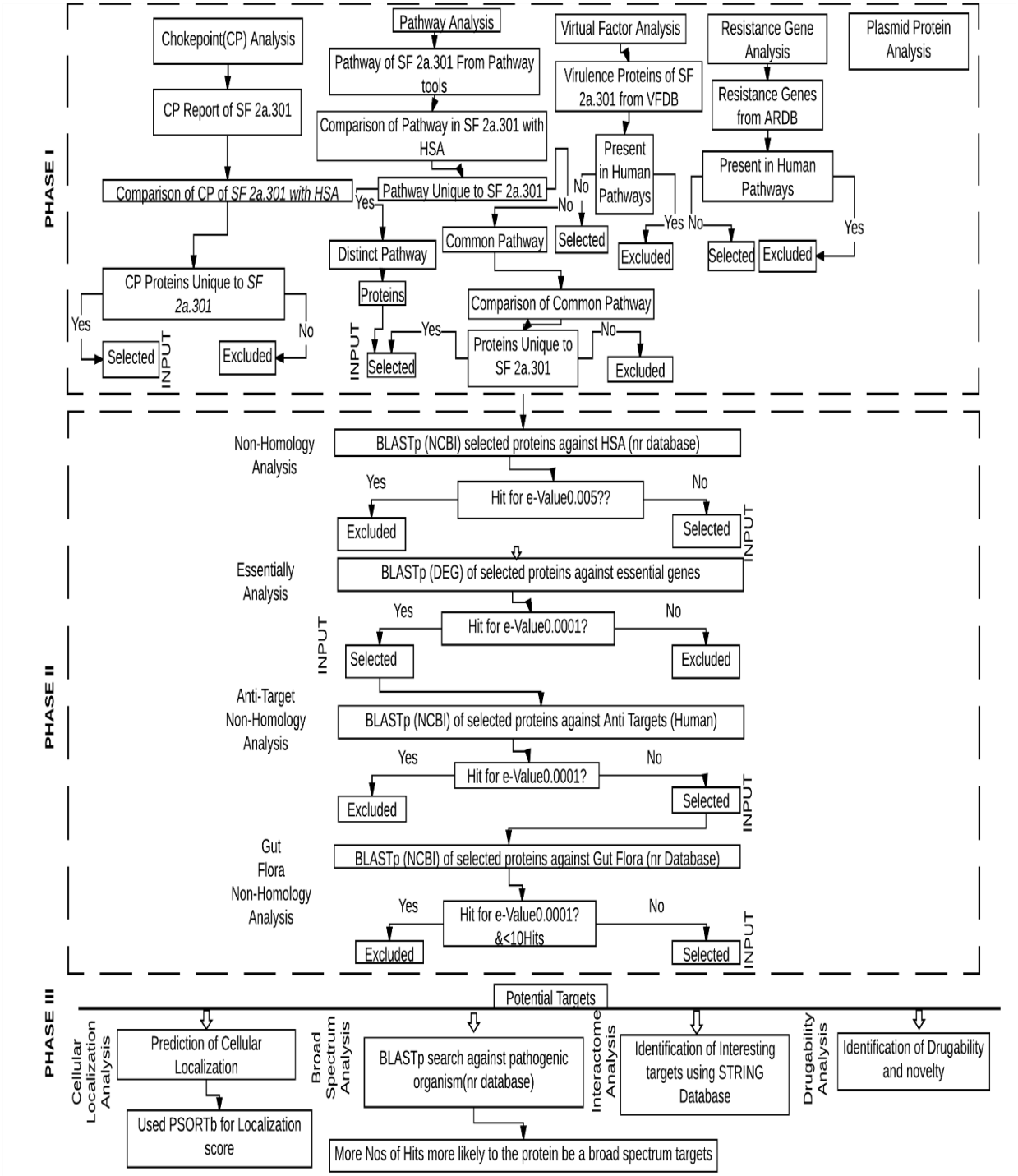

Figure 1. The flowchart is representing various steps in the Analysis.

301 compiled from the Antibiotic Resistance Gene Database. This database cover 1737 species with the resulting number of genes 17237 for 249 antibiotics. By manually examining the database 38 resistance genes and their analogous proteins were saved. Proteins raised in the human were omitted from the study, and unique proteins analyzed[21].

Phase II: Subtractive Channel of Analysis

List of the proteins came from the Phase I survey were significantly interpreted to obtain a selective and extremely competent drug target in Shigella flexneri 2a str. 301.

Non-Homology Analysis: study was planned to classify the pathogen-limited proteins i.e. they must be non-homologs to humans. Non-homology study is crucial to reduce the cross-reactivity and unwanted side effect of the drug for the target. All the proteins came from the Phase I were later evaluated against nr human database using BLASTp program maintained by NCBI with the expected value 0.005. Proteins demonstrating no substantial relationship with a human were chosen for a further phase of evaluation.[22]

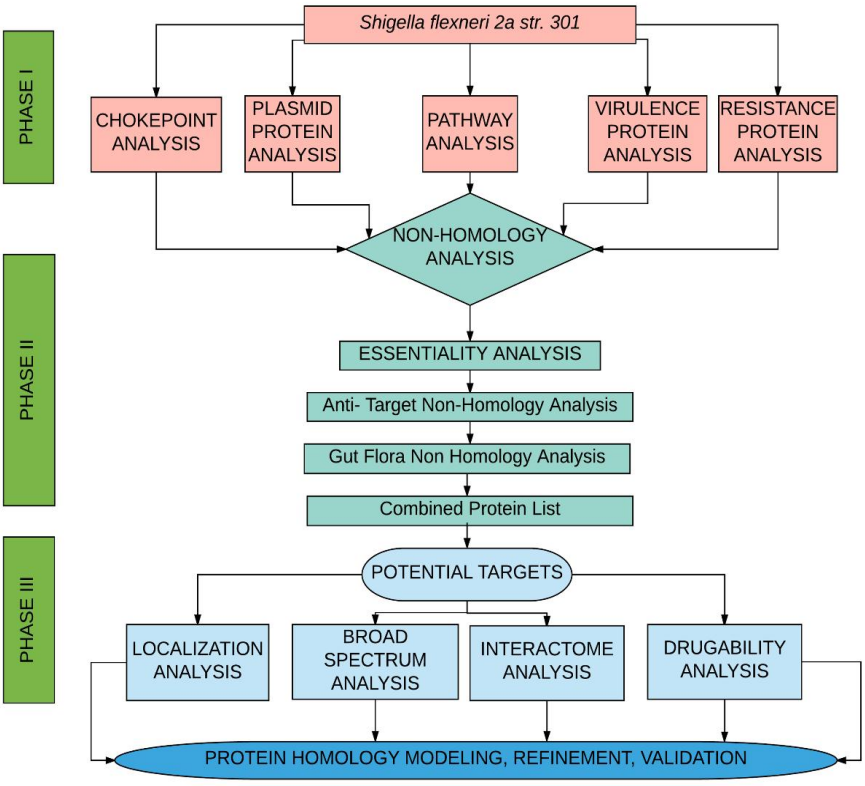

Figure 2. Whole wok Flowchart.

Essentiality analysis: Study conducted on the presumption that proteins sensible for one series of organisms will likewise demand for another set of organisms. Essentiality study was driven out using the Database of Essential genes and proteins. The database is constructed of 53,038 relevant genes and 786 fundamental non-coding sequences. What list of the proteins emerged in the preceding list exposed to similarity search against DEG V.14.0 using BLASTp program with an Expected value of 0.0001 [22]. Proteins reaching similarity with 0.0001 e-value or fewer were meaningful hits and advanced as an input for the later list of the proteins.[23,24]

Anti-target non-homology analysis: The call Anti-targets is adopted for the proteins present in the host and usually time may be docked with the drug targeting pathogen proteins. This accidental targeting host proteins will induce intense pharmacological actions.[25] Anti-target database was established which made up 203 proteins. Example of the proteins included in the database is a human ether-a-go-go-related gene (hERG), the pregnant X receptor (PXR), constitutive androstane receptor (CAR), and P-glycoprotein (P-gp). Local BLASTp search done against the database adopting an e-value of 0.005. Proteins showing no similarity were chosen as input for the further study.[22]

Gut flora non-homology analysis: Around 1014 microflora seen to display in the Gut of human. They have a reciprocal relation by fermenting the undigested food particles and further provide protection against the colonization of pathogenic microorganisms.[26,27] Unplanned loss of gut microflora may precipitate extreme side effects. Protein sequences produced from the preceding list were examined against the human gut microflora database provided by NCBI with an e-value of 0.0001.Proteins reaching similarity with over ten proteins were omitted from the list. The derived proteins were inputs for later phase of investigation.[22]

Proteins made up the first and second phase were merged to create the shared list. This simple list was significantly investigated.

Phase III: Qualitative Characterization of shortlisted targets

Cellular localization Analysis: The study planned to reveal proteins in the cell. Shigella flexneri 2a str. 301 is gram-negative bacteria. The proteins can be displayed in the outer membrane, cytoplasm, cell membrane, periplasm membrane. Localization investigation was conducted to determine whether the protein sufficient to grow as a drug target or vaccine target. PSORTb server applied to determine location of proteins in the cell. At PSORTb main homepage option chosen as gram-negative with a result outcome as tab limited short format. PSORTb performs interpretation by studying the position of accessible proteins in the database. FASTA formatted sequence was referred to the server.[28,29]

Broad Spectrum Analysis: the early list of proteins next studied to search their presence in popular pathogenic bacteria. List used for the examination included 177

pathogenic bacteria with many of the Non-Shigella species. Proteins were investigated using BLASTp tool with an e-value of 0.005. Proteins present in the number of species show the essentiality of the proteins and can be expanded as a target present in multiple pathogenic species. The derived proteins were likewise investigated for their cluster of orthologous groups using web server for meta genomic study which measures the query sequence with the COG database.[30]

Interactome analysis: Derived protein list was adopted to set up the protein network analysis using STRING 10.0 STRING applies a theory of Experimental data, gene-based analysis, varied proteins study, and curated pathway database. Protein network investigation was considered applying a high interactors score of higher than or matched to 0.700. Some interacting proteins (nodes) and interactors (edges) were utilized to scrutinize the list.[31,32]

Druggability Analysis: Proteins could emerge as possible drugs if they bind with the targets. Druggability study was run utilizing the drug target database of Drug bank.it subsists of thousands of the drug targets with FDA recommended drugs, enzyme careers, membrane proteins and more. E-value was kept 0.00001 to examine targets in the database. Proteins showing target correlation with available drug and functionality made up the final list of proteins.[33,34]

Homology modeling, refinement, and evaluation of the 3D structure: Protein sequence of the resulted list submitted to PHYRE2 server ((Protein Homology/analogy Recognition Engine version 2) from Imperial College of London.[35] Structure refinement done using the Galaxy WEB server at the Computational Biology Lab in the Department of Biochemistry, Seoul National University.[36] The precise models were verified by several verification tools to choose the finest model and determine the quality of that model. ERRAT is a protein structure verification method to check the improvement of crystallographic model construction and refining, controlled by the National Health Institute, University of California, USA. The Z- the score worked out by the PROSA web tool from the Center for Applied Molecular Engineering, Division of Bioinformatics, and University of Salzburg, Austria. It measures the variation of the overall energy of the structure belonging to an energy distribution obtained from random conformations located in native proteins.[37] The SWISS-MODEL workspace applied further to investigate the structure. The workspace consists of several assessment tools integrated within like ANOLEA, QMEAN6 server (Qualitative Model Energy Analysis), which measures the universal and resident quality of the models (Biozentrum, University of Basel, Switzerland, and Procheck from the European Bioinformatics Institute, Welcome Trust Genome Campus, Hinxton, UK)[38-44]

Normal Mode Analysis: Bio3D package in R V. 3.4.1 was employed to examine the Normal Mode Analysis for three protein structures. Normal Mode Analysis is a computational approach to study the flexibility of protein structures. Three Parameters mode index eigenvalues, mode index frequency, and residue index fluctuations were obtained.[45,46]

RESULTS AND DISCUSSION

Phase I Mining of Protein Datasets

Chokepoint Analysis

Chokepoint summary of both host and pathogen was made utilizing Pathway Tools 14.5.[6,47] Pathway/Genome Database (PGDB) describes 4180 genes from the individual chromosome and 262 genes from the plasmid of Shigella flexneri 2a str. 301. Of 4180 genes, 4052 are protein-coding genes, and 128 are RNA genes. Pathway Tools consider both metabolic pathway and no metabolic-pathway reactions for generation of chokepoint report. Chokepoint report generated a list of enzymes for both producing and consuming sides. Of 823 chokepoint enzymes of Shigella flexneri 2a.301, 408 were from producing, 415 from consuming. Chokepoint enzymes unique to *Shigella flexneri 2a.301* with a relation to Human host were chosen (356 proteins), and enzymes common to both host and pathogen were eliminated (131).

Pathway Analysis

In-silico comparative investigation of H. sapiens and Shigella flexneri 2a.301pathways referred to distinct metabolic processes was done. Shigella flexneri 2a.301possesses 420 pathways, which are distributed as Biosynthesis, Degradation, Utilization, Assimilation, Generation of Precursor Metabolites and Energy and Signal transduction pathway. Unique pathways present in Shigella flexneri 2a.301 and devoid in H. sapiens (316 pathways) were described as distinct pathways, whereas pathways found in both host and pathogen were referred as common pathways. Sequences of 556 proteins from distinct pathways and 33 unique proteins from common pathways were collected and crossed through the subtractive channel of study (Phase II).

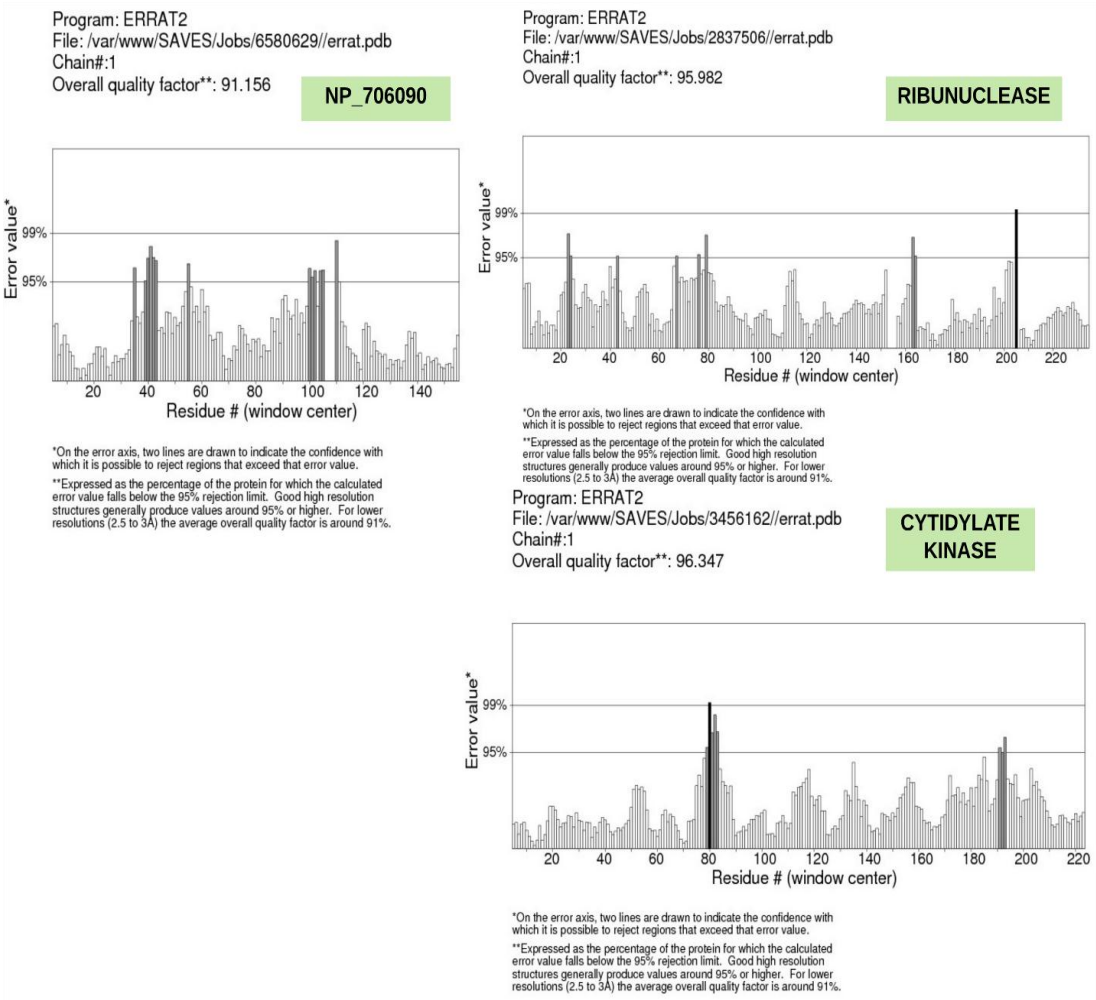

Figure 3: ERRAT plot shows error values for residues. The Y-axis represents the error value, and the X-axis represents the amino acid residues of the protein model. An error value exceeding 99% confidence level indicates a poorly-modeled region.

Plasmid Protein Analysis

All of 262 proteins sequences were recovered from the NCBI database. All the proteins present in the plasmid were taken as input for the non-homology analysis (**Error! Reference source not found.**).

Virulence factor Analysis: pathogenomics identification was done in the Virtual factor Database, and 31 proteins were detected as proteins providing Virulence to the Shigella flexneri 2a str. 301 and were unique to the pathogen.

Table 1: Results of Three Phase Analysis

Phase	Analysis	Method	Sequences	Input	Passed	failed
I.	**Mining of Protein Data**					
I.	Chokepoint analysis	SF reactions against HSA choke point proteins: P*408,C*415	823	823	356 P*193 C*163 (A1)	467
II.	Plasmid protein analysis	Protein sequences for plasmid retrieved from NCBI web server(Accession: NC_ NC_004851.1)	262	262	262 (B1)	--------
III.	Pathway analysis	Distinct pathway enzymes:556 Unique proteins from common pathway:33	589	589	589 (C1)	--------
IV.	Virulence factor database	Presence of virulence protein in KEGG/NCBI	31	31	31 (D1)	--------
V.	AntiBiotic Resistance protein	ARDB database used to retrieve resistance proteins.	38	38	38 (E1)	--------
II.	**Subtractive analysis of Phase I. protein.**					
I.	Non Homology analysis	1276 SF proteins against nr HSA database (BLASTp)	1276	**1276** A1=356 B1=262	**868** A1=257 B1=244	**408** A1=99 B1=18

				C1=589 D1=31 E1=38	C1=321 D1=22 E1=24	C1=268 D1=09 E1=14
II.	Essentiality analysis	868 SF proteins against DEG database	868	**868** A1=257 B1=244 C1=321 D1=22 E1=24	**579** A1=217 B1=64 C1=264 D1=12 E1=22	**289** A1=40 B1=180 C1=57 D1=10 E1=02
III.	Anti-target non-homology analysis	579 SF proteins were analyzed against Anti-targets in humans	579	**579** A1=217 B1=64 C1=264 D1=12 E1=22	**576** A1=216 B1=63 C1=263 D1=12 E1=22	**03** A1=01 B1=01 C1=01 D1=00 E1=00
IV.	Gut flora nonhomology analysis		576	**576** A1=216 B1=63 C1=263 D1=12 E1=22	220 A1=68 B1=40 C1=91 D1=07 E1=14	356 A1=148 B1=23 C1=172 D1=05 E1=08
III.	**Qualitative characterization of short listed sequences.**					
I.	Cellular localization analysis	CELLO & PSORTb protein location	220	220	138	82
II.	Broad spectrum analysis	Homology search against 177 pathogenic species	138	138	24	114
III.	Interactome analysis	Protein interaction analysis using string V.10.0	24	24	11	13
IV.	Druggability analysis	Performed against DrugBank targets, TTD targets	11	11	04	07

Resistance genes database: Total of 38 resistance genes found using antibiotic resistance database. Genes covered in the investigations were (baca) seen to produce resistance for bacitracin, (mdtn) causing resistance for acriflavine, puromycin, t_chloride, (macb) for macrolide and several others for proven antibiotics. (**Error! Reference source not found.**).

Table 2: Refined models with validation scores.

Models	ProSA Z-Score	ERRAT Quality	QMEAN6 Score	Ramachandran Plot			
				Co[1] (%)	AA[2] (%)	GA[3] (%)	DA[4] (%)
NP_706090							
Initial	-7.29	68.874	0.09	91.90	7.40%	0.00%	0.70%
MODEL 1	7.49	85.714	0.27	92.60	6.60%	0.00%	0.70%
MODEL 2	-7.44	84.932	0.66	94.10	5.10%	0.00%	0.70%
MODEL 3	-7.53	91.156	0.00	94.90	4.40%	0.00%	0.70%
MODEL 4	-7.43	79.592	0.04	94.10	5.10%	0.70%	0.00%
MODEL 5	-7.40	86.577	0.09	94.10	5.10%	0.00%	0.70%
CytidylateKinase							
Initial	-7.56	81.279	0.24	90.70	7.40%	0.50%	1.50%
MODEL 1	-7.38	94.977	0.65	93.10	5.90%	0.50%	0.50%
MODEL 2	-7.48	91.244	0.75	95.10	3.90%	0.50%	0.50%
MODEL 3	-7.56	93.088	0.61	95.10	3.90%	0.50%	0.50%
MODEL 4	-7.47	92.694	0.60	93.60	5.40%	0.50%	0.50%
MODEL 5	-7.55	96.247	0.69	94.10	4.90%	0.50%	0.50%
Ribunuclease							
Initial	-7.42	82.609	-0.48	94.20	4.40%	1.00%	0.50%
MODEL 1	-7.68	95.946	0.30	94.20	4.40%	1.00%	0.50%
MODEL 2	-7.68	93.213	0.33	93.70	5.30%	0.50%	0.50%
MODEL 3	-7.56	95.982	0.35	95.10	3.90%	0.50%	0.50%
MODEL 4	-7.67	94.545	0.53	95.10	3.90%	0.50%	0.50%
MODEL 5	-7.79	95.964	0.38	94.20	4.90%	0.50%	0.50%

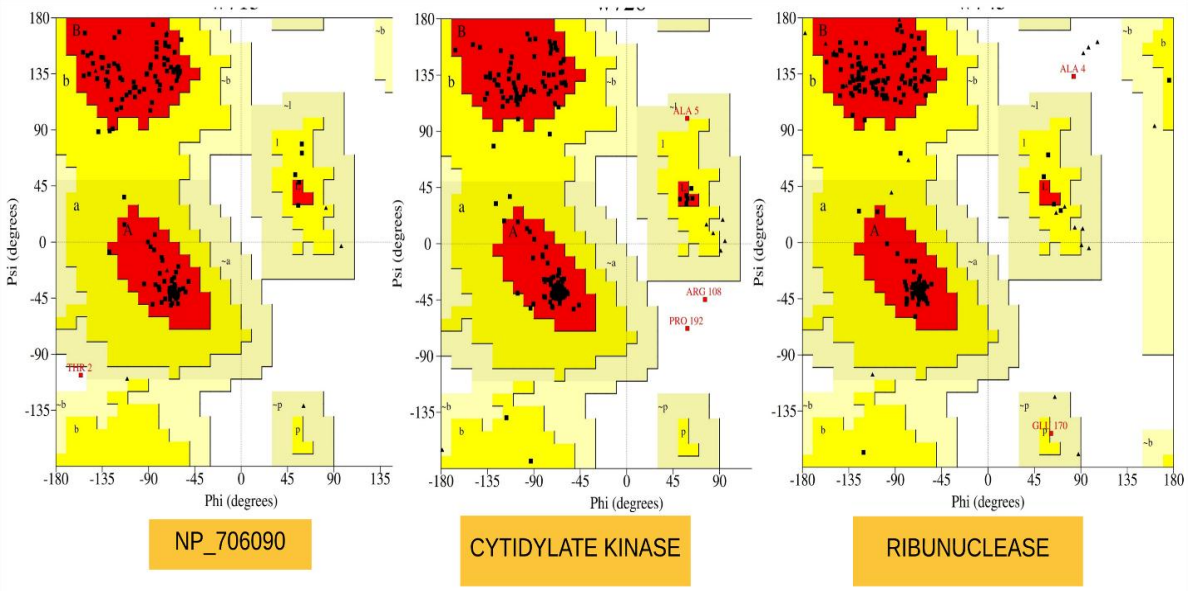

Figure 4: Ramachandran plot of the Best refined models determined by Procheck. The most favored regions are marked as A, B, and L. The additional allowed regions are indicated as a, b, l, and p. All non-glycine and proline residues are shown as filled black squares, whereas glycines (non-end) are shown as filled black triangles. Disallowed residues are colored red.

Phase II subtractive channel of Analysis

Selectivity, specificity, and essentiality are the expected characteristic for any drug. Phase II investigations were planned for stated demands.

Non Homology Analysis: In any drug discovery and design the prior and foremost step is to eliminate the drug target that is present in both pathogen and host. As the presence of the host can result in acute pharmacological consequences. A total of 1276 Shigella flexneri 2a.301 proteins listed in phase I were subjected to a homology search against the whole proteome of H. sapiens (host) using BLASTp with an E-value threshold of 0.005.[14] Among 1276 proteins 257 from chokepoint analysis, 244 proteins from plasmid protein analysis, 321 proteins from pathway analysis, and 22 were from virulence factor analysis and 24 proteins from antibiotic resistance genes resulted in the second list of proteins.

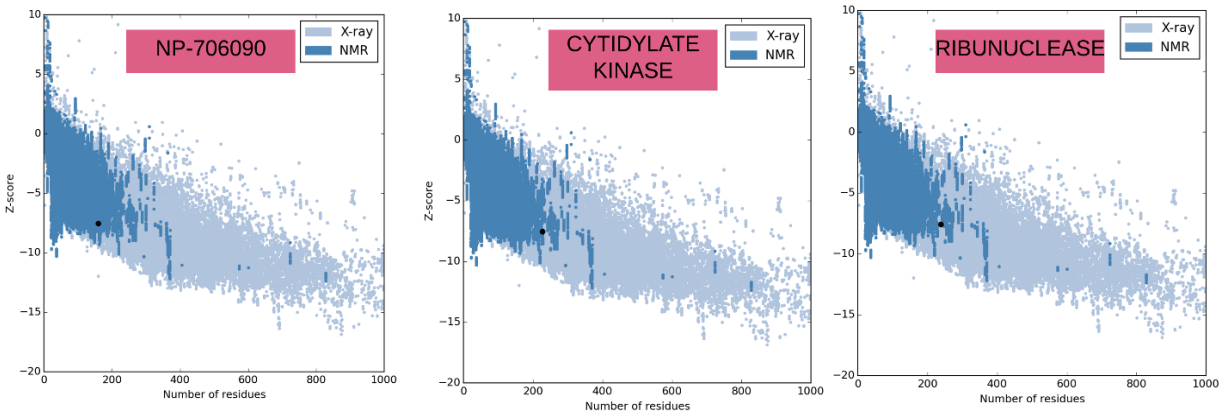

Figure 5: The Z-score plot of Three Proteins determined by ProSA. The Z-score is, within the range of experimental native structures of similar sizes.

Essentiality analysis: The non-homologous protein list was significantly filtered based on essentiality using DEG server with an expected value of 0.0001. Of 868 input proteins found as non-homologous to human, 579 proteins that have homologs with fewer than the threshold value were fundamental for the durability of the pathogen and made for the subsequent investigations. Proteins not offering any hits (289) against DEG were non-fundamental and removed from the study.

Anti-Target Non-Homology Analysis: identification of proteins nonhomologous to the Human Anti target protein (essential human proteins) is an essential step for any drug design studies. it is done to avoid the extreme Pharmacological reactions. Each of the early list pathogen proteins exposed to an identity check against a list of 203 anti-targets using BLASTp. The E-value threshold was arrayed to 0.005 and proteins exhibiting hits below this threshold were homologs. Of 579 proteins

in the last list, 576 were non-homologous, and 03 were close homologs to anti-targets. Proteins in homologs list excluded as drugs inhibiting such targets may interfere with host metabolism causing severe adverse effects.

Gut Flora Analysis: The proteins produced from the past list examined against the Human gut flora database maintained by NCBI metagenomic project using an e-value of 0.0001. Proteins producing fewer than ten homologs with an E-value threshold were selected and preferred for the further qualitative studies. From 576 proteins provided as input, 220 proteins have been determined, and 356 proteins were getting higher hits (>10) were removed. Proteins reaching all the selection principles given in phase II were regarded as potential drug targets in Shigella flexneri 2a.301.

Emerged proteins from the Phase I and Phase II were incorporated to serve as the decisive list of target proteins and the same list was later considered for selectivity, specificity, and functionality.

PHASE III: Qualitative Characterization of the Short-listed Targets

Cellular Localization Analysis: To indicate the short-specified targets as drug and vaccine targets, their placement within the bacterial cell was detected applying prediction server PSORTb. Target proteins situated in the cytoplasm can be employed as drug targets, whereas extracellular and membrane-bound proteins can function as vaccine targets. Based on the localization score. Total of 138 out of 220 proteins was separated as cytoplasmic and cytoplasmic membrane targets (Table 1).

Broad Spectrum Analysis: BLASTp search against 177 pathogenic bacteria reported that 24 proteins found a relationship in over 100 species (Table 1). This examination showed that evaluation of such protein as drug target might be appropriate to limit other pathogenic infections as well. From the ends, it is again obvious that related proteins in the list likewise found in the non Shigella species. Results of the broad spectrum analysis again showed that the proteins identified in the virulence factor database were not related to other pathogenic bacteria as the virulence proteins identified to be precise for the Shigella flexneri 2a str. 301.

Interactome analysis: Metabolic protein network investigation was conducted to discover the relevance of short-listed protein. Network map was set up using STRING protein database. Proteins having a part in higher pathway regarded as significant for the cell. Total of 11 proteins identified to develop a strong network amongst other proteins and they were picked up as an active target for Druggability analysis. Selected protein sequences were fetched from the NCBI protein database (Table 1).

Druggability analysis

In the prevailing idea, the Druggability of the short-record possible targets was assessed by sequence comparison search against targets from DrugBank. BLASTp search against DrugBank targets with FDA recognized drugs, reported that NP_706090.1 is homologs to (2-amino-4-hydroxy-6-hydroxymethyl dihydro pteridine pyrophosphokinase, Folic acid synthesis protein FOL1 and 2-amino-4-hydroxy-6-hydroxymethyl dihydro pteridine pyrophosphokinase) with an e-value of 0. This obvious target has one established inhibitors Sulfacetamide. Also, NP_706829.1 (cytidylate kinase) revealed homology with Cytidine kinase and NP_709422.2 (Ribonuclease PH) revealed identity with t-RNA nucleotidyltransferase.

Protein Homology Modelling, Refinement and Validation: The decisive list of the three proteins was in the FASTA form referred to Phyre2 server for the Homology modeling. Model built using Intensive modeling choice. PDB formatted structures recovered from protein databank.

Structures referred to Galaxy Refine server for the precision. Five models were developed for each of the three structures. Last, validation of the processed structures was driven by finding scores and structural properties.

As it is obvious from the () Model no.3 for NP_706090 and Ribunuclease and Model no.5 for CytidylateKinase identified to be relevant for the ERRART quality (Figure 3), Z-score (Figure 5), QMEAN6 score. From the table (Table 2), the validation scores are higher related to the crude model. Ramachandran plot (Figure 4) generated by urging the relevant model to Procheck server. Plot produced again indicate the preferred structures are of immense quality as the better % of residues fall under the most favored regions.

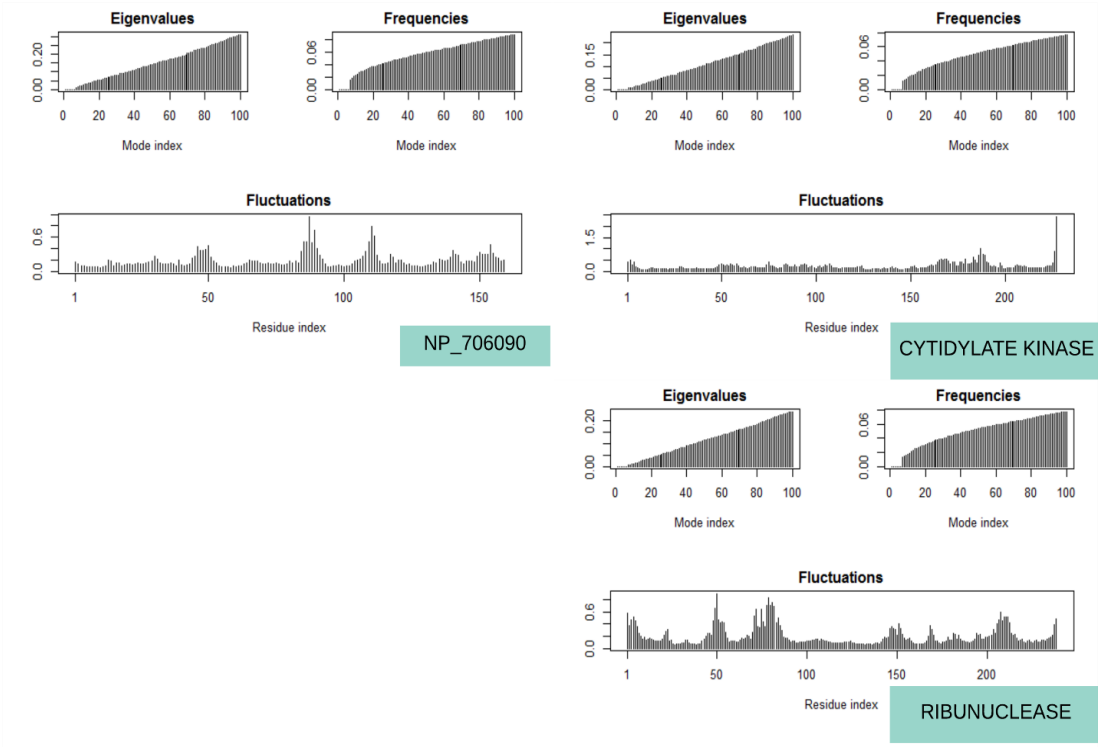

Figure 6: Normal Mode Analysis for Refined and Validated Structures.

Normal Mode Analysis: NMA evaluation gives knowledge for the robustness of the structure at the conditions. In the study, mode frequency provided the stiffness of the structure. Eigenvalues were taken to look at the fluctuation along the principal axes. Last, fluctuations values were detected on the belief that at any conditions variations can be determined to utilize the single harmonic design for potential (Figure 6). NMA analysis for the indicated three structures can direct the Dynamic studies and look at the variations from the global mode.[48]

CONCLUSION

Three possible drug targets are determined to adopt a harmonized In-silico approach. Targets described were moved from three phases of evaluation which drained them affirming to their essentiality, homology with the host, their behavior in the metabolic pathway and moreover the extra-chromosomal in the plasmids were likewise studied. With all the raise, the proteins after the Phase, I study was later determined for pathogenic specific virulence factors and resistance producing proteins. The third phase was considered for selectivity, specificity by matching with gut flora database and broad spectrum pathogenic analysis. Three stage of filtration produced a list of particular three proteins complied with all the criteria for acceptable drug target., the homology modeling, refinement and validation exercises produced structures which can produce target ligand docking studies. Normal Mode Analysis managed can again serve to create the molecular dynamic studies for the protein in various solvent and conditions. Further protein and ligand binding complex can still be simulated applying the primary NMA studies. The technique adopted for drug target study can further be applied to another group of pathogenic organisms. Structures referred to Galaxy Refine server for the precision. Five models were developed for each of the three structures. Last, validation of the processed structures was driven by finding scores and structural properties.

Conflicts of Interest

Authors declare no conflict of interest.

REFERENCES

1. Wei J, Goldberg MB, Burland V, Venkatesan MM, Deng W, Fournier G, et al. Complete genome sequence and comparative genomics of shigella flexneri serotype 2a strain 2457t. *Infection and immunity*, 2003; 71(5): 2775-86.
2. Anderson M, Sansonetti PJ, Marteyn BS. Shigella diversity and changing landscape: Insights for the twenty-first century. *Frontiers in Cellular and Infection Microbiology*, 2016; 6: 45. doi: 10.3389/fcimb.2016.00045
3. Guerin PJ, Brasher C, Baron E, Mic D, Grimont F, Ryan M, et al. Case management of a multidrug-resistant shigella dysenteriae serotype 1 outbreak in a crisis context in sierra leone, 1999-2000. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 2004; 98(11): 635-43. doi: 10.1016/j.trstmh.2004.01.005
4. Moss JE, Cardozo TJ, Zychlinsky A, Groisman EA. The selc-associated shi-2 pathogenicity island of shigella flexneri. *Molecular Microbiology*, 1999; 33(1): 74-83. doi: 10.1046/j.1365-2958.1999.01449.x

5. Karp PD, Paley SM, Krummenacker M, Latendresse M, Dale JM, Lee TJ, et al. Pathway tools version 13.0: Integrated software for pathway/genome informatics and systems biology. *Briefings in bioinformatics*, 2010; 11(1): 40-79. doi: 10.1093/bib/bbp043
6. Karp PD, Paley SM, Krummenacker M, Latendresse M, Dale JM, Lee TJ, et al. Pathway tools version 13.0: Integrated software for pathway/genome informatics and systems biology. *Briefings in bioinformatics*, 2010; 11(1): 40-79. doi: 10.1093/bib/bbp043
7. Kushwaha SK, Shakya M. Protein interaction network analysis--approach for potential drug target identification in mycobacterium tuberculosis. *Journal of theoretical biology*, 2010; 262(2): 284-94. doi: 10.1016/j.jtbi.2009.09.029
8. Shanmugham B, Pan A. Identification and characterization of potential therapeutic candidates in emerging human pathogen mycobacterium abscessus: A novel hierarchical in silico approach. *PLoS ONE*, 2013; 8(3): e59126. doi: 10.1371/journal.pone.0059126
9. Venkatesan M, Fernandez-Prada C, Buysse JM, Formal SB, Hale TL. Virulence phenotype and genetic characteristics of the t32-istrati shigella flexneri 2a vaccine strain. *Vaccine*, 1991; 9(5): 358-63.
10. Caspi R, Altman T, Billington R, Dreher K, Foerster H, Fulcher CA, et al. The metacyc database of metabolic pathways and enzymes and the biocyc collection of pathway/genome databases. *Nucleic Acids Research*, 2014; 42(D1): D459-D71. doi: 10.1093/nar/gkt1103
11. Caspi R, Billington R, Ferrer L, Foerster H, Fulcher CA, Keseler IM, et al. The metacyc database of metabolic pathways and enzymes and the biocyc collection of pathway/genome databases. *Nucleic Acids Research*, 2016; 44(D1): D471-D80. doi: 10.1093/nar/gkv1164
12. Romero P, Wagg J, Green ML, Kaiser D, Krummenacker M, Karp PD. Computational prediction of human metabolic pathways from the complete human genome. *Genome Biology*, 2004; 6(1): R2. doi: 10.1186/gb-2004-6-1-r2
13. Buchrieser C, Glaser P, Rusniok C, Nedjari H, D'Hauteville H, Kunst F, et al. The virulence plasmid pwr100 and the repertoire of proteins secreted by the type iii secretion apparatus of shigella flexneri. *Mol Microbiol*, 2000; 38(4): 760-71.
14. Lan R, Stevenson G, Reeves PR. Comparison of two major forms of the shigella virulence plasmid pinv: Positive selection is a major force driving the divergence. *Infection and immunity*, 2003; 71(11): 6298-306.
15. Venkatesan MM, Goldberg MB, Rose DJ, Grotbeck EJ, Burland V, Blattner FR. Complete DNA sequence and analysis of the large virulence plasmid of shigella flexneri. *Infection and immunity*, 2001; 69(5): 3271-85. doi: 10.1128/iai.69.5.3271-3285.2001
16. Xiong Z, Tang X, Yang F, Zhang X, Yang J, Chen L, et al. Comparison of the virulence plasmid genomes of two strains of shigella which lost the ability to bind congo red. *Science in China Series C, Life sciences*, 2006; 49(2): 141-8.
17. Bhattacharya D, Purushottaman SA, Bhattacharjee H, Thamizhmani R, Sudharama SD, Manimunda SP, et al. Rapid emergence of third-generation cephalosporin resistance in shigella sp. Isolated in andaman and nicobar islands, india. *Microbial drug resistance (Larchmont, NY)*, 2011; 17(2): 329-32. doi: 10.1089/mdr.2010.0084
18. Ye C, Lan R, Xia S, Zhang J, Sun Q, Zhang S, et al. Emergence of a new multidrug-resistant serotype x variant in an epidemic clone of shigella flexneri. *Journal of clinical microbiology*, 2010; 48(2): 419-26. doi: 10.1128/jcm.00614-09
19. Zhang N, Lan R, Sun Q, Wang J, Wang Y, Zhang J, et al. Genomic portrait of the evolution and epidemic spread of a recently emerged multidrug-resistant shigella flexneri clone in china. *Journal of clinical microbiology*, 2014; 52(4): 1119-26. doi: 10.1128/jcm.02669-13.
20. Zhu JY, Duan GC, Yang HY, Fan QT, Xi YL. Atypical class 1 integron coexists with class 1 and class 2 integrons in multi-drug resistant shigella flexneri isolates from china. *Current microbiology*, 2011; 62(3): 802-6. doi: 10.1007/s00284-010-9790-3.
21. Liu B, Pop M. Ardb—antibiotic resistance genes database. *Nucleic Acids Research*, 2009; 37(1): D443-D7. doi: 10.1093/nar/gkn656
22. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *Journal of molecular biology*, 1990; 215(3): 403-10. doi: 10.1016/s0022-2836(05)80360-2
23. Luo H, Lin Y, Gao F, Zhang C-T, Zhang R. Deg 10, an update of the database of essential genes that includes both protein-coding genes and noncoding genomic elements. *Nucleic Acids Research*, 2014; 42(D1): D574-D80. doi: 10.1093/nar/gkt1131
24. Zhang R, Lin Y. Deg 5.0, a database of essential genes in both prokaryotes and eukaryotes. *Nucleic Acids Res.*, 2009; 37: D455-8. doi: 10.1093/nar/gkn858
25. Recanatini M, Bottegoni G, Cavalli A. In silico antitarget screening. *Drug discovery today Technologies*, 2004; 1(3): 209-15. doi: 10.1016/j.ddtec.2004.10.004
26. Fujimura KE, Slusher NA, Cabana MD, Lynch SV. Role of the gut microbiota in defining human health. *Expert review of anti-infective therapy*, 2010; 8(4): 435-54. doi: 10.1586/eri.10.14
27. Rabizadeh S, Sears C. New horizons for the infectious diseases specialist: How gut microflora promote health and disease. *Current infectious disease reports*, 2008; 10(2): 92-8.

28. Yu CS, Chen YC, Lu CH, Hwang JK. Prediction of protein subcellular localization. *Proteins*, 2006; 64(3): 643-51. doi: 10.1002/prot.21018
29. Yu NY, Wagner JR, Laird MR, Melli G, Rey S, Lo R, et al. Psortb 3.0: Improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics*, 2010; 26(13): 1608-15. doi: 10.1093/bioinformatics/btq249
30. Wu S, Zhu Z, Fu L, Niu B, Li W. Webmga: A customizable web server for fast metagenomic sequence analysis. *BMC Genomics*, 2011; 12: 444. doi: 10.1186/1471-2164-12-444
31. Jensen LJ, Kuhn M, Stark M, Chaffron S, Creevey C, Muller J, et al. String 8--a global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Res*., 2009; 37: D412-6. doi: 10.1093/nar/gkn760
32. Szklarczyk D, Franceschini A, Kuhn M, Simonovic M, Roth A, Minguez P, et al. The string database in 2011: Functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Res*., 2011; 39: D561-8. doi: 10.1093/nar/gkq973
33. Knox C, Law V, Jewison T, Liu P, Ly S, Frolkis A, et al. Drugbank 3.0: A comprehensive resource for 'omics' research on drugs. *Nucleic Acids Res*., 2011; 39: D1035-41. doi: 10.1093/nar/gkq1126
34. Robertson JG. Mechanistic basis of enzyme-targeted drugs. *Biochemistry*, 2005; 44(15): 5561-71. doi: 10.1021/bi050247e
35. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE. The phyre2 web portal for protein modeling, prediction and analysis. *Nat Protocols*, 2015; 10(6): 845-58. doi: 10.1038/nprot.2015.053
36. Ko J, Park H, Heo L, Seok C. Galaxyweb server for protein structure prediction and refinement. *Nucleic Acids Research*, 2012; 40(W1): W294-W7. doi: 10.1093/nar/gks493
37. Wiederstein M, Sippl MJ. Prosa-web: Interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Research*, 2007; 35(2): W407-W10. doi: 10.1093/nar/gkm290
38. Arnold K, Bordoli L, Kopp J, Schwede T. The swiss-model workspace: A web-based environment for protein structure homology modelling. *Bioinformatics*, 2005; 22(2): 195-201. doi: 10.1093/bioinformatics/bti770
39. Biasini M, Bienert S, Waterhouse A, Arnold K, Studer G, Schmidt T, et al. Swiss-model: Modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic acids research*, 2014; 42(W1): W252-W8.
40. Guex N, Peitsch MC, Schwede T. Automated comparative protein structure modeling with swiss-model and swiss-pdbviewer: A historical perspective. *ELECTROPHORESIS*, 2009; 30(S1): S162-S73. doi: 10.1002/elps.200900140
41. Kiefer F, Arnold K, Künzli M, Bordoli L, Schwede T. The swiss-model repository and associated resources. *Nucleic Acids Research*, 2008; 37(1): D387-D92. doi: 10.1093/nar/gkn750
42. Melo F, Feytmans E. Assessing protein structures with a non-local atomic interaction energy. *Journal of molecular biology*, 1998; 277(5): 1141-52. doi: 10.1006/jmbi.1998.1665
43. Benkert P, Künzli M, Schwede T. Qmean server for protein model quality estimation. *Nucleic Acids Research*, 2009; 37. (Web Server issue):W510-W4. doi: 10.1093/nar/gkp322
44. Benkert P, Tosatto SC, Schomburg D. Qmean: A comprehensive scoring function for model quality assessment. *Proteins*, 2008; 71(1): 261-77. doi: 10.1002/prot.21715
45. Skjærven L, Yao X-Q, Scarabelli G, Grant BJ. Integrating protein structural dynamics and evolutionary analysis with bio3d. *BMC Bioinformatics*, 2014; 15(1): 399. doi: 10.1186/s12859-014-0399-6
46. Yao X-Q, Skjærven L, Grant BJ. Rapid characterization of allosteric networks with ensemble normal mode analysis. *The Journal of Physical Chemistry B*., 2016; 120(33): 8276-88. doi: 10.1021/acs.jpcb.6b01991
47. Karp PD. Pathway databases: A case study in computational symbolic theories. *Science*, 2001; 293. doi: 10.1126/science.1064621
48. Bahar I, Lezon TR, Bakan A, Shrivastava IH. Normal mode analysis of biomolecular structures: Functional mechanisms of membrane proteins. *Chemical reviews*, 2010; 110(3): 1463-97. doi: 10.1021/cr900095e.