1 **A comprehensive computational investigation into the conserved virulent proteins of *Shigella* sp**

2 **unveils potential siRNA candidates as a new therapeutic strategy against shigellosis**

3 Parag Palit[1], Farhana Tasnim Chowdhury[2], Namrata Baruah[3], Bonoshree Sarkar[2], Sadia Noor Mou[2],

4 Mehnaz Kamal[1], Towfida Jahan Siddiqua[1], Zannatun Noor[1, *], Tahmeed Ahmed[1]

5 Author Affiliations:

6 [1]International Centre for Diarrhoeal Disease Research, Bangladesh (icddr,b), Dhaka, Bangladesh

7 [2]Department of Biochemistry and Molecular Biology, University of Dhaka, Dhaka, Bangladesh

8 [3]Department of Biological Sciences and Bioengineering, Indian Institute of Technology Kanpur (IIT-K),

9 Uttar Pradesh, India

10 [*]Corresponding author: Zannatun Noor, PhD. Email: zannatun@icddrb.org

11

12

13

14

15

16

17

18

19

20

21 **Abstract**

22 *Shigella* sp account for the second-leading cause of deaths due to diarrheal diseases among children of

23 less than 5 years of age. Emergence of multi-drug resistant *Shigella* isolates and the lack of availability of

24 *Shigella* vaccines have made the efforts in the development of new therapeutic strategies against

25 shigellosis very pertinent. In our study we have analyzed a total of 241 conserved sequences from a 15

26 different conserved virulence genes of *Shigella* sp and through extensive rational validation using a

27 plethora of computational algorithms; we primarily obtained fifty eight small-interfering RNA (siRNA)

28 candidates. Further extensive computational validation showed only three siRNA candidates that were

29 found to exhibit substantial functional efficacy, be non-immunogenic and have a thermodynamically

30 stable and sterically feasible and thereby acceptable tertiary structure. These siRNA candidates are

31 intended to suppress the expression of the virulence genes, namely: IpgD (siRNA 9) and OspB (siRNA 15

32 and siRNA 17) and thus act as a prospective tool in the RNA interference (RNAi) pathway. However, the

33 findings of our study require further wet lab validation and optimization for regular therapeutic use in the

34 treatment of shigellosis.

35 **Keywords: *Shigella*, siRNA, computational algorithms.**

36

37

38

39

40

41

42

Introduction

Shigellosis can be attributed for approximately 12.5% cases of diarrheal mortality on a global scale; accounting for about 163400 cases of annual deaths, 54900 of these cases being children of less than 5 years of age (1). Global burden of shigellosis is mainly attributable to *S. flexneri* and *S. sonnei* whereas, *S. boydii* is uncommon outside South-east Asia (2). In regions with rapid industrialization, development in water sanitation and in economically developing regions, *S. sonnei* has gained prevalence in the epidemiological shift in places like Vietnam, Thailand and Bangladesh (3-6). *S. sonnei* is also predominant in traveler's diarrhea (7).

Although antibiotic resistant strains are being reported against ciprofloxacin, azithromycin, ceftriaxone are continued to be utilized as mainstay treatment strategies (8). The growing rise of antimicrobial resistance to *Shigella* sp. has thus led to the need for development of newer therapeutic strategies against Shigellosis more pertinent (9-15). A multi-centric study reported that 85% of cases occurring in LMIC could be attributable to *S. flexneri* 2a, 3a and 6 together with *S. sonnei* and therefore, a quadrivalent vaccine targeting these strains is expected to provide significant protection in endemic regions (16). Serotype based vaccines include conjugate vaccine, carbohydrate vaccine, live-attenuated or killed whole cell vaccines (17-20). However, due to the lack of ideal animal models, no *Shigella* vaccine has reached the stage of commercialization till date (21).

The concept of designing a single therapeutic candidate with substantial efficacy and potency against all *Shigella* sp would thereby involve targeting the multiple conserved proteins involved in the process of invasion and pathogenesis. The virulence-related proteins or Vir proteins such as VirA helps in entry and intracellular motility, golgi fragmentation in host (22, 23), VirB activates the invasion proteins (24), VirF, involved in cell invasion and activation of VirG and VirB (25). VirG, also called IcsA, helps in actin polymerization and hence movement of the bacteria from one cell to another (26). The invasion plasmid antigens, Ipa, are the key regulators of invasion and pathogenicity and also activate other related proteins (27-30). IpgB is an invasion effector protein of the Type III Secretion system (T3SS), involved in

68   membrane ruffling for cellular entry of *Shigella* (31). IpgD is another effector involved in entry and

69   survival of the bacteria (32-34). There are many membrane excretion proteins and surface presentation

70   antigens together called as Mxi-Spa proteins that make up the T3SS and help in pathogenesis and

71   invasion. In particular, Spa 33 belonging to this class of gatekeeper proteins controlling T3SS protein

72   secretion into host cells, helps in the secretion of Ipa proteins (35), while MxiC which interacts with the

73   Ipa proteins to mediate the entry and survival of the bacteria (36). Other T3SS effectors such as OspB,

74   OspF and OspG modulate host cell signaling pathways, transcription and manipulate the host

75   inflammatory response (34, 37, 38).

76   Recent developments in the quest for newer therapeutic strategies against a number of infectious

77   pathogens involves the utilization of the concept of post-transcriptional gene silencing by designing

78   siRNA molecules specific to the pathogen (39). Small-interfering RNA (siRNA) represent a class of

79   exogenous double-stranded non-coding RNA that binds with specific sequences of a messenger RNA

80   (mRNA) and promotes the process of degradation of the mRNA, eventually halting the process of

81   transcription (40). Concurrent works involving the use of a plethora of computational tools and

82   algorithms have been design potential therapeutic siRNA candidates against a number of infectious

83   organisms, including *P.vivax* (41) and *L.donovani* (42), the COVID-19 pandemic causing SARS-CoV-2

84   (43) and against a number of bacterial pathogens, such as: *M. tuberculosis* (44), *Salmonella* (45) and

85   *Listeria* (46).

86   In this study, we have aimed to design novel therapeutic options against all members of the *Shigella*

87   genus that are pathogenic to humans (*S.flexneri, S.dysenteriae, S.sonnei* and *S.boydii*) through the design

88   of siRNA candidates against a number of conserved *Shigella* proteins, involved in the process of invasion

89   and pathogenesis. We have targeted a total of 15 conserved virulent proteins of *Shigella* sp, including:

90   IcsA/Vir G, IpaA, IpaB, IpaC, IpaJ, IpgB, IpgD, MxiC, OspB, OspF, OspG, Spa33, VirA, VirB and Vir F

91   and through the use of rigorous computational tools and algorithms to select the most feasible and

92   effective siRNA candidates against all the *Shigella* sp.

93   **Results**

94   **Retrieval of nucleotide sequences of conserved virulent proteins of *Shigella* sp**

95   A total of 241 conserved sequences (shown in **Supplementary File 1**) from the 15 different conserved

96   virulent *Shigella* proteins were obtained from subsequent multiple sequence alignment analysis, done

97   using Clustal Omega.

98   **Designing siRNA candidates by using a combination of first-generation and second-generation**

99   **algorithms**

100   siDirect 2.0, a highly efficient web based computational tool was used to predict potential target specific

101   siRNA candidates on the basis of three first generation algorithms governing the sequence preference of

102   siRNA, namely: Ui-Tei, Amarzguioui and Reynolds rules (U, R, A rules) (47) . This tool predicted a total

103   of 58 potential siRNA candidates from the 241 conserved regions of the 15 conserved virulent *Shigella*

104   proteins, which were found to comply with the three first generation algorithms for siRNA designing, i.e.-

105   U, R, A rules.

106   Out of these 58 siRNA candidates that fulfilled the U, R, A rules, only 38 siRNA candidates were found

107   to comply with the 90% threshold set by the i-SCORE Designer software for the second generation

108   algorithms for siRNA designing (48). Table 1 illustrates the results of the analysis obtained from the first

109   and second generation algorithms of siRNA designing. NCBI BLAST program was used to confirm off-

110   target resemblance of siRNAs and none of the 38 siRNA candidates that had fulfilled both the first and

111   second-generation algorithms for siRNA designing had displayed any off-target similarity with the human

112   genome.

113 **Table 01: Effective siRNA candidates for each of the conserved virulence associated *Shigella* genes as analyzed by both the first**

114 **generation and second generation algorithms for designing siRNAs along with their respective free energy of folding and GC content.**

siRNA	Gene	Target Sequence	Predicted siRNA duplex siRNA candidate at 37 °C (5' to 3')	Combined second generation algorithm Scoring*	Free energy of self-folding of guide strand (kcal/mol)	GC Content (%)
siRNA 1	IcsA	TGGAATGGATGCGTGGTATATAA	AUAUACCACGCAUCCAUUCCA GAAUGGAUGCGUGGUAUAUAA	<90% threshold	-----	-----
siRNA 2	IcsA	TACGGATTAACTCTGATATTATG	UUAUAUAUCCGCUAUUAGCAA GCUAAUAGCGGAUAUAUAAUU	>90% threshold	1.0	28.57
siRNA 3	IcsA	TTGCTAATAGCGGATATATAATT	UUAUAUAUCCGCUAUUAGCAA GCUAAUAGCGGAUAUAUAAUU	>90% threshold	-1.30	-----
siRNA 4	IcsA	TGGTGATGTACAGGTTAACAATT	UUGUUAACCUGUACAUCACCA GUGAUGUACAGGUUAACAAUU	>90% threshold	1.50	38.10
siRNA 5	IpaJ	ATGGAATTAGGTGCAAGAAGAAG	UCUUCUUGCACCUAAUUCCAU GGAAUUAGGUGCAAGAAGAAG	>90% threshold	1.80	38.10
siRNA6	IpaJ	TGGAATTAGGTGCAAGAAGAAGT	UUCUUCUUGCACCUAAUUCCA GAAUUAGGUGCAAGAAGAAGU	>90% threshold	1.80	38.10
siRNA7	IpgB	AAGGATCTTACAAATCTAGTATC	UACUAGAUUUGUAAGAUCCUU GGAUCUUACAAAUCUAGUAUC	>90% threshold	1.70	28.57
siRNA8	IpgD	AGCCATGATGGAACGATTAGATA	UCUAAUCGUUCCAUCAUGGCU CCAUGAUGGAACGAUUAGAUA	<90% threshold	-----	-----

siRNA 9	IpgD	GAGGAATGTTGACAAGCTTAATG	UUAAGCUUGUCAACAUUCCUC	>90% threshold	1.80	38.10
			GGAAUGUUGACAAGCUUAAUG			
siRNA 10	IpgD	TGGAAATCCAAGAGATGAATACT	UAUUCAUCUCUUGGAUUUCCA	<90% threshold	-----	-----
			GAAAUCCAAGAGAUGAAUACU			
siRNA 11	MxiC	TTCAGCTATACAGGCTAAATTAT	AAUUUAGCCUGUAUAGCUGAA	<90% threshold	-----	-----
			CAGCUAUACAGGCUAAAUUAU			
siRNA 12	MxiC	GTGAAAGTGAGCAAATTCTTACT	UAAGAAUUUGCUCACUUUCAC	<90% threshold	-----	-----
			GAAAGUGAGCAAAUUCUUACU			
siRNA 13	MxiC	GAGGATTCTGTAGTGTATCAAAC	UUGAUACACUACAGAAUCCUC	>90% threshold	1.70	38.10
			GGAUUCUGUAGUGUAUCAAAC			
siRNA 14	OspB	ATGTACAAACAATCATTTCAAGA	UUGAAAUGAUUGUUUGUACAU	>90% threshold	1.70	23.80
			GUACAAACAAUCAUUUCAAGA			
siRNA 15	OspB	CTGCTGAAAGTCTTTCTTGTATC	UACAAGAAAGACUUUCAGCAG	>90% threshold	1.70	38.10
			GCUGAAAGUCUUUCUUGUAUC			
siRNA 16	OspB	ATCTTTGCTAGAGCAGATAAAAA	UUUAUCUGCUCUAGCAAAGAU	>90% threshold	-1.40	-----
			CUUUGCUAGAGCAGAUAAAAA			
siRNA 17	OspB	ATGAAAGACTGTGGTATTCTAAA	UAGAAUACCACAGUCUUUCAU	>90% threshold	1.60	33.33
			GAAAGACUGUGGUAUUCUAAA			
siRNA 18	OspF	ATGCTTTCTGCGAATGAAAGATT	UCUUUCAUUCGCAGAAAGCAU	<90% threshold	-----	-----
			GCUUUCUGCGAAUGAAAGA			
siRNA 19	OspF	TGGAAGATAACTGATATGAATCG	AUUCAUAUCAGUUAUCUUCCA	>90% threshold	1.80	28.57
			GAAGAUAACUGAUAUGAAUCG			

siRNA 20	OspF	ATGGAAGATAACTGATATGAATC	UUCAUAUCAGUUAUCUUCCAU	>90% threshold	1.80	28.57
			GGAAGAUAACUGAUAUGAAUC			
siRNA 21	OspF	TCGCAATATAGTGCTTTATTACT	UAAUAAAGCACUAUAUUGCGA	>90% threshold	-2.0	-----
			GCAAUAUAGUGCUUUAUUACU			
siRNA 22	OspG	GCCCATTCTCGGTAAGTTAATAG	AUUAACUUACCGAGAAUGGGC	<90% threshold	-----	-----
			CCAUUCUCGGUAAGUUAAUAG			
siRNA 23	OspG	CAGCTGATATCCCTGATAATATA	UAUUAUCAGGGAUAUCAGCUG	>90% threshold	1.50	38.10
			GCUGAUAUCCCUGAUAAUAUA			
siRNA 24	OspG	ATCTACAGTTGATATGTAAATTG	AUUUACAUAUCAACUGUAGAU	<90% threshold	-----	-----
			CUACAGUUGAUAUGUAAAUUG			
siRNA 25	OspG	ATCCATTACGATCTTAATACAGG	UGUAUUAAGAUCGUAAUGGAU	>90% threshold	1.60	28.57
			CCAUUACGAUCUUAAUACAGG			
siRNA 26	OspG	CGCAATATTTATGCTGAATATTA	AUAUUCAGCAUAAAUAUUGCG	>90% threshold	-2.40	-----
			CAAUAUUUAUGCUGAAUAUUA			
siRNA 27	Spa33	GACAATCAATGAACTAAAAATGT	AUUUUUAGUUCAUUGAUUGUC	<90% threshold	-----	-----
			CAAUCAAUGAACUAAAAAUGU			
siRNA 28	Spa33	AACTAAAAATGTATGTAGAAAAC	UUUCUACAUACAUUUUUAGUU	>90% threshold	1.60	19.05
			CUAAAAAUGUAUGUAGAAAAC			
siRNA 29	Spa33	ATGTATGTAGAAAACGAATTATT	UAAUUCGUUUUCUACAUACAU	>90% threshold	1.60	28.10
			GUAUGUAGAAAACGAAUUAUU			
siRNA 30	Spa33	TTCAAGTTTCCCGATGACATAGT	UAUGUCAUCGGGAAACUUGAA	>90% threshold	-1.30	-----
			CAAGUUUCCCGAUGACAUAGU			
siRNA 31	VirF	TTCAACAAATCCTTCTTGATATT	UAUCAAGAAGGAUUUGUUGAA	<90% threshold	-----	-----
			CAACAAAUCCUUCUUGAUAUU			

siRNA 32	VirF	TGGCGTCTTTCTGATATTTCAAA	UGAAAUAUCAGAAAGACGCCA	>90% threshold	1.80	38.10
			GCGUCUUUCUGAUAUUUCAAA			
siRNA 33	VirF	GTCTTTCTGATATTTCAAATAAC	UAUUUGAAAUAUCAGAAAGAC	>90% threshold	-1.40	-----
			CUUUCUGAUAUUUCAAAUAAC			
siRNA 34	VirF	AACTTGAATTTATCAGAAATAGC	UAUUUCUGAUAAAUUCAAGUU	<90% threshold	-----	-----
			CUUGAAUUUAUCAGAAAUAGC			
siRNA 35	VirA	GCCTGAACAACGAGTTATTAACA	UUAAUAACUCGUUGUUCAGGC	<90% threshold	-----	-----
			CUGAACAACGAGUUAUUAACA			
siRNA 36	VirA	CTGAACAACGAGTTATTAACAAT	UGUUAAUAACUCGUUGUUCAG	<90% threshold	-----	-----
			GAACAACGAGUUAUUAACAAU			
siRNA 37	VirA	TACGAAGTTAGCTCATCAATATT	UAUUGAUGAGCUAACUUCGUA	>90% threshold	-0.90	-----
			CGAAGUUAGCUCAUCAAUAUU			
siRNA 38	VirA	ACGAAGTTAGCTCATCAATATTA	AUAUUGAUGAGCUAACUUCGU	>90% threshold	1.60	33.33
			GAAGUUAGCUCAUCAAUAUUA			
siRNA 39	VirA	CTCCAGAAAGTCGTCAAGTATCA	AUACUUGACGACUUUCUGGAG	>90% threshold	1.60	42.86
			CCAGAAAGUCGUCAAGUAUCA			
siRNA 40	VirB	CTCCATTCTGGTAATAAAGTTTC	AACUUUAUUACCAGAAUGGAG	<90% threshold	-----	-----
			CCAUUCUGGUAAUAAAGUUUC			
siRNA 41	VirB	AACGAATGTACGCGATCAAGAAT	UCUUGAUCGCGUACAUUCGUU	<90% threshold	-----	-----
			CGAAUGUACGCGAUCAAGAAU			
siRNA 42	VirB	GAGATTGATGGTAGAATTGAAAT	UUCAAUUCUACCAUCAAUCUC	>90% threshold	1.80	33.33
			GAUUGAUGGUAGAAUUGAAAU			
siRNA 43	VirB	AACTAGCAAACGATATACAAACA	UUUGUAUAUCGUUUGCUAGUU	>90% threshold	1.80	28.57
			CUAGCAAACGAUAUACAAACA			

siRNA 44	VirB	TAGTTCTACACTACCAATATTAA	AAUAUUGGUAGUGUAGAACUA	>90% threshold	-1.10	-----
			GUUCUACACUACCAAUAUUAA			
siRNA 45	IpaA	GGGAAAGAAGATGTGTTAAGAAG	UCUUAACACAUCUUCUUUCCC	<90% threshold	-----	-----
			GAAAGAAGAUGUGUUAAGAAG			
siRNA 46	IpaA	CACAGTATTCGGAACTAATTATA	UAAUUAGUUCCGAAUACUGUG	>90% threshold	1.90	30.96
			CAGUAUUCGGAACUAAUUAUA			
siRNA 47	IpaA	TTGCACTATAGCACAACAACACA	UGUUGUUGUGCUAUAGUGCAA	<90% threshold	-----	-----
			GCACUAUAGCACAACAACACA			
siRNA 48	IpaA	CTCCTCAATACTGAAGTATCATC	UGAUACUUCAGUAUUGAGGAG	>90% threshold	-3.20	-----
			CCUCAAUACUGAAGUAUCAUC			
siRNA 49	IpaA	TCCGTTCTACCACACTCTATATC	UAUAGAGUGUGGUAGAACGGA	>90% threshold	1.70	42.86
			CGUUCUACCACACUCUAUAUC			
siRNA 50	IpaA	TTCAACCATACTCCAGATAATTC	AUUAUCUGGAGUAUGGUUGAA	>90% threshold	1.50	35.71
			CAACCAUACUCCAGAUAAUUC			
siRNA 51	IpaB	CACCAAAGTCATTAAATGCAAGT	UUGCAUUUAAUGACUUUGGUG	>90% threshold	1.50	33.33
			CCAAAGUCAUUAAAUGCAAGU			
siRNA 52	IpaB	AAGAAATACAACTCACTATCAAA	UGAUAGUGAGUUGUAUUUCUU	>90% threshold	1.50	28.57
			GAAAUACAACUCACUAUCAAA			
siRNA 53	IpaB	CAGTTAAAGACAGGACATTGATT	UCAAUGUCCUGUCUUUAACUG	>90% threshold	1.80	35.71
			GUUAAAGACAGGACAUUGAUU			
siRNA 54	IpaB	CTCAATTGATGGCAACCTTTATT	UAAAGGUUGCCAUCAAUUGAG	>90% threshold	1.40	35.71
			CAAUUGAUGGCAACCUUUAUU			
siRNA 55	IpaB	CTCCTTTCAGATGCATTTACAAA	UGUAAAUGCAUCUGAAAGGAG	<90% threshold	-----	-----
			CCUUUCAGAUGCAUUUACAAA			

siRNA 56	IpaB	GGCCAATTGCAGGAAGTAATTGC	AAUUACUUCCUGCAAUUGGCC	<90% threshold	-----	-----
			CCAAUUGCAGGAAGUAAUUGC			
siRNA 57	IpaC	TTGAAGAAGAAGAACAACTAATC	UUAGUUGUUCUUCUUCUUCAA	>90% threshold	1.80	30.96
			GAAGAAGAAGAACAACUAAUC			
siRNA 58	IpaC	AAGAAGAAGAACAACTAATCAGT	UGAUUAGUUGUUCUUCUUCUU	<90% threshold	1.70	30.963
			GAAGAAGAACAACUAAUCAGU			

115   Footnotes:

116   *Combined Score obtained from the results of the second-generation tools for siRNA designing, namely: i-SCORE, s-Biopredsi, Katoh and DSIR.

117 **Free energy of self-folding and evaluation of GC content**

118 Functional of siRNA largely depends on molecular structure. For this, extensive efforts have been put on

119 predicting the secondary structure of the of the RNA molecules (49). Benchmark of molecular structural

120 accuracy of the siRNA was set as the "Minimum Free Energy (MFE)" (50). The minimization of free

121 energy is an established phenomenon in computational structural biology on the basis that at an state of

122 equilibrium, the molecule folds into the state of least energy (51).The minimum free energy of folding

123 was calculated by using the RNAstructure web server to check the stability of the predicted siRNA guide

124 strands. From the above mentioned analysis, 30 siRNA candidates were found to yield a positive free

125 energy of folding and thus were considered for further analysis (Table 1).

126 The GC content of siRNA is a major determinant of the stability of the secondary structure of the siRNA,

127 whereby a GC content ranging between 30–60% is considered sufficient for the execution of its action

128 (52). In our study, out of the 30 siRNA candidates that exhibited a positive free energy of folding in

129 addition to complying with all the primary and secondary algorithms, only 19 candidates were found to

130 have a GC content in between 30-60% and thus were selected for subsequent analysis (Table 1).

131 **Evaluation of binding energy of siRNA and target and visualization of secondary structures of**

132 **siRNA-target duplex**

133 Precise prediction of binding energy between siRNA and target is integral for proper understanding of the

134 binding of siRNA to the target and for the subsequent assessment of functional efficiency of the siRNA

135 (53)-(54)-(55). RNA structure, an online web server was used for the estimation of the hybridization

136 energy between the siRNA-target duplex. The thermodynamics of the siRNA-target interaction and the

137 details of the calculation of this binding energy between siRNA and target have been published elsewhere

138 (56)-(57). Table 2 shows the binding energies of the siRNA with the target sequence.

139   **Table 02: Designed siRNA molecules with their respective free energy of binding of siRNA with target, melting temperature of siRNA-**

140   **target duplex, target accessibility, functional efficiency, and immunogenicity designed siRNA.**

siRNA	Target sequence	Free Energy of Binding with Target (kcal/mol)	Melting temperature of siRNA-target duplex (^{0}C)	End-diff* (kcal/mol)	Break-targ. ΔG** (kcal/mol)	Probability of being efficient siRNA	siRNA validity score (Binary)	Immunogenicity
siRNA 4	TGGTGATGTACAGGTTAACAATT	-32.6	79.1	1.76	-1.8	0.951	0.901	Non immunomodulatory (IMscore:4.2)
siRNA 5	ATGGAATTAGGTGCAAGAAGAAG	-32.1	79.8	0.17	-0.1	0.951	0.95	Non immunomodulatory (IMscore:3.5)
siRNA 6	TGGAATTAGGTGCAAGAAGAAGT	-32.5	78.2	0.17	-0.1	0.951	0.95	Non immunomodulatory (IMscore:4.2)
siRNA 9	GAGGAATGTTGACAAGCTTAATG	-31.8	78.2	2.33	-1.4	0.961	0.967	Non immunomodulatory (IMscore:3.4)
siRNA 13	GAGGATTCTGTAGTGTATCAAAC	-32.7	78.7	2.33	-0.7	0.974	0.975	Non immunomodulatory (IMscore:2.6)
siRNA 15	CTGCTGAAAGTCTTTCTTGTATC	-31	78.3	2.09	-1.8	0.967	1.026	Non immunomodulatory (IMscore:2.7)
siRNA 17	ATGAAAGACTGTGGTATTCTAAA	-30.1	78.9	0.03	-0.9	0.947	1.026	Non immunomodulatory (IMscore:3.1)
siRNA 23	CAGCTGATATCCCTGATAATATA	-32.4	79.1	2.09	-0.9	0.96	0.989	Immunomodulatory (IMscore:4.7)

siRNA 32	TGGCGTCTTTCTGATATTTCAAA	-31.9	76.6	1.76	-1.8	0.957	1.009	Immunomodulatory (IMscore:4.8)
siRNA 38	ACGAAGTTAGCTCATCAATATTA	-29.4	74.9	1.7	-0.3	0.931	1.012	Non immunomodulatory (IMscore:4.0)
siRNA 39	CTCCAGAAAGTCGTCAAGTATCA	-33	80.4	2.16	-0.3	0.952	0.949	Non immunomodulatory (IMscore:2.8)
siRNA 42	GAGATTGATGGTAGAATTGAAAT	-30.2	74	1.87	-1	0.968	0.965	Immunomodulatory (IMscore:5.6)
siRNA 46	CACAGTATTCGGAACTAATTATA	-28.6	74.6	1.23	-0.3	0.936	0.971	Non immunomodulatory (IMscore:2.2)
siRNA 49	TCCGTTCTACCACACTCTATATC	-34.3	82	1.03	-0.1	0.935	1.006	Non immunomodulatory (IMscore:1.1)
siRNA 50	TTCAACCATACTCCAGATAATTC	-30.3	78.5	1.46	0	0.886	0.886	-----
siRNA 51	CACCAAAGTCATTAAATGCAAGT	-29	76.1	0.13	-0.3	0.818	0.818	-----
siRNA 53	CAGTTAAAGACAGGACATTGATT	-31.3	78.1	0.79	-1.1	0.926	0.909	Immunomodulatory (IMscore:5.6)
siRNA 54	CTCAATTGATGGCAACCTTTATT	-31	78	1.23	-1	0.891	-----	-----
siRNA 57	TTGAAGAAGAAGAACAACTAATC	-27.8	76.2	0.33	-0.2	0.906	0.999	Immunomodulatory (IMscore:7.3)

141

142 Footnotes:

143 * The free energy difference between the 5' and 3' end of the antisense strand of siRNA.

144 ** The free energy cost for opening base pairs in the region of complementarity to the target.

145   Secondary structures of siRNA with their respective target provide a very efficient computational

146   estimation for both the structure and for the thermodynamics of RNA-RNA interaction (53).

147   RNAstructure program predicts the most stable secondary structures of target-siRNA duplex and

148   minimizes the folding energy. 37 °C is the temperature that was chosen to predict the folded structure.

149   The secondary duplex of candidate siRNA molecules and their corresponding targets are elucidated in

150   **Figure 2**.

151

Determination of heat stability of siRNA-target duplex

Heat stability analysis of the siRNA-target duplex, a key factor in the assessment of the stability of secondary structure and functional efficiency of siRNA (58), was conducted by using the OligoWalk server. The results of the heat stability analysis of siRNA-target duplex showed that all the potential 19 siRNA candidates that had fulfilled all the criteria for effective siRNA assessed so far, (as shown in Table 1) had a melting temperature of greater than -70$^{\mathrm{o}}$C (Table 2). Henceforth, the melting temperatures of each of the siRNA-target duplex structure were found to be substantially greater than the physiological temperature of 37.4$^{\mathrm{o}}$C indicating towards the maintenance of integrity of secondary structure of siRNA in the host physiological system.

Prediction of functional efficiency and target accessibility of the siRNA candidates

Table 2 delineates the summary of the results of the analysis of the functional efficiency and target accessibility as determined by the OligoWalk web server, of each of the siRNA candidates that had fulfilled the threshold for all the criteria set in Table 1. In Table 2, the "End-diff" score for each of the 19 potential siRNA candidates indicates free energy difference between the 5′ and 3′ end of the antisense strand of siRNA (59); whereby the siRNA candidates that have showing positive "End-diff" scores are ranked to have a high functional efficiency. In our study, all the siRNA candidates demonstrated a positive "End-diff" score, thus indicating towards a high functional efficiency of the candidate siRNAs. Moreover, siRNA 9, siRNA 13, siRNA 39, siRNA 15 and siRNA 23 were found to exhibit more positive "End-diff" scores compared to the rest of the siRNA candidates and thus can be classified as having the highest functional efficiency.

Consequently, target accessibility was determined by the "Break targ. ΔG" score, which is free energy account for the opening of base pairs in the region of complementarity to the target (58). Candidate siRNAs demonstrating a less negative "Break targ. ΔG" score are classified as having greater target

175 accessibility (59)-(58). In our study, siRNA 5, siRNA 6, siRNA 49 and siRNA 50 showed the least

176 negative "Break targ. ΔG" score and thus can be classified as having the greatest target accessibility.

177 Another score delineated in Table 2 as "Probability of being efficient siRNA", manifests both the

178 functional efficiency and target accessibility of the designed siRNA candidates (58)-(59). In our analysis,

179 siRNA 50, siRNA 51 and siRNA 54 demonstrated a "Probability of being efficient siRNA" score of less

180 than 0.9 and were not considered for further analysis. The rest of the siRNA candidates yield a

181 "Probability of being efficient siRNA" score of greater than 0.9 and thus can be predicted to have high

182 functional efficiency and substantial target accessibility.

183 **Validation of the functional efficiency of the siRNA candidates**

184 Among the 16 predicted siRNA candidates that had fulfilled the threshold for all the criteria set so far, 11

185 siRNA candidates (siRNA 4, siRNA 5, siRNA 6, siRNA 9, siRNA 13, siRNA 23, siRNA 39, siRNA 42,

186 siRNA 46, siRNA 53 and siRNA 57) showed a siRNA validity score between 0.8-1.0, following the

187 binary pattern prediction approach indicating high functional efficiency (60). On the other hand, five

188 siRNA candidates (siRNA 15, siRNA 17, siRNA 32, siRNA 38 and siRNA 49) showed a binary

189 validation score of greater than 1.0, which manifests towards the highest functional efficiency (60).

190 **Prediction of immunotoxicity of the siRNA candidates**

191 We evaluated the immunotoxicity of the 16 siRNA candidates for which functional efficiency was

192 validated in the previous step. 12 out of the 16 siRNA candidates were found to be non-immunogenic on

193 the basis of the default threshold IMscore set for assessing potential immunogenicity of query siRNA

194 sequences and were considered for subsequent analysis. siRNA 23, siRNA 32, siRNA 42 and siRNA 53

195 were found to be immunogenic (IMscore greater than the threshold of 4.5) and were discarded.

Designing of tertiary structure of the siRNA candidates and validation

Respective tertiary/3D structures for each of the 12 non-immunogenic siRNA candidates was generated using the web-based RNA modeling software, RNAComposer (61) and the individual 3D structures were saved in pdb format for downstream application. Subsequently, MOLProbity web server was used to validate the individual tertiary structures of the non-immunogenic siRNA candidates and the 3D structures are then filtered for clash score/the number of serious steric overlaps (>0.4 Å) per 1000 atoms (62). Validation scores for the predicted 3D structures of the siRNA candidates have been shown in **Table 3**.

In our study, a total of three siRNA candidates, namely: siRNA 9, siRNA 15 and siRNA 17 showed an acceptable tertiary structure with all the scores of the analyzed criteria for nucleic acid geometry such as: probability of wrong sugar puckers, bad backbone confirmations, bad bonds and bad angles being below the 5% threshold of acceptance for tertiary structures. **Figure 3** illustrates the tertiary structures of siRNA 9, siRNA 15 and siRNA 17 as visualized using the PyMol Molecular Graphics System (v1.8.4).

209 **Table 03: Validation scores for the predicted 3D structures of the 12 non-immunogenic siRNA**

210 **candidates.**

siRNA	Nucleic acid Geometry				All Atom Contacts [Clashsccore, Percentile] **
	Probability of wrong sugar puckers* (%)	Bad backbone confirmations* (%)	Bad bonds* (%)	Bad angles* (%)	
siRNA 4	0.00	0.00	0.00	0.00	36.81, 9th
siRNA 5	0.00	0.00	0.00	0.00	30.67, 15th
siRNA 6	0.00	0.00	0.00	0.00	36.81, 9th
siRNA 9	0.00	0.00	0.00	0.00	19.76, 33rd
siRNA 13	0.00	0.00	0.00	0.00	24.1, 23rd
siRNA 15	0.00	4.76	0.00	0.00	19.23, 35th
siRNA 17	0.00	0.00	0.00	0.00	19.61, 34th
siRNA 38	4.76	57.14	0.00	0.00	22.52, 26th
siRNA 39	0.00	0.00	0.00	0.00	22.46, 26th
siRNA 46	0.00	0.00	0.00	0.00	29.76, 16th
siRNA 49	0.00	28.57	0.00	0.00	19.03, 35th

211 Footnotes:

212 * 5% threshold is allowed for acceptance of the criteria for a valid 3D structure

213 ** 100th percentile is the best among the structures for a comparable resolution, 0th percentile ranks to be

214 the worst. clashscore is the parameter that represents a comparable resolution.

215

216

217

218

219

220

221

222

223

Discussion

224

225 Our computational study is concurrent with the current efforts in the development of new therapeutic

226 strategies to combat the persistent conundrum surrounding the emergence of multi-drug resistance among

227 *Shigella* sp. To the best of our knowledge this is the first study that has envisaged proposing a novel

228 treatment option for shigellosis by targeting the virulent proteins which have been reported to be

229 conserved across all known isolates of *Shigella* sp.

230 In our study, we have used the designing of siRNA candidates that are expected to target the expression

231 of the specific conserved virulence genes of *Shigella* sp through the mediation of a process known as gene

232 silencing (63). Design of siRNA candidates represents a new therapeutic strategy aimed to induce RNA

233 specific inhibition (64), whereby an effective siRNA must fulfill all the threshold criteria set by the first-

234 generation (Ui-Tei, Amarzguioui and Reynolds rules) and second-generation (i-SCORE, s-Biopredsi,

235 DSIR and Katoh rules) algorithms for siRNA designing (65, 66). In this current study, for all the 15

236 aforementioned conserved virulence genes of *Shigella* sp, we had obtained a total of 58 siRNA candidates

237 that fulfilled the first-generation algorithms, among which only 38 siRNA (Table 1) candidates satisfied

238 all the second-generational algorithms. Consequently, off-target silencing of siRNAs can potentially lead

239 to undesired toxicities (67) and none of the siRNA candidates that satisfied both the first and second

240 generation algorithms were found to exhibit such off-target activity within the human genome.

241 The folded secondary structure of siRNA is integral for assessment of functional efficiency of siRNA (68)

242 and for the evaluation of structural stability and accuracy of RNA, minimum free energy (MFE) is

243 considered as a standard parameter (50). In our study, out of all the siRNA candidates satisfying both

244 first- and second-generation algorithms, 30 siRNA candidates (Table 1) showed a positive MFE value,

245 indicating thermodynamic infeasibility of self-folding. Subsequently, only 19 of these 30 siRNA

246 candidates with a positive MFE value were found to have GC content within the range of 30-60%. GC

247 content between 30% and 60% is recommended for siRNA sequence, since there is a considerable

248   reciprocal correlation between GC-content and RNAi activity (52), whereby a GC content of siRNA

249   within the aforementioned range exhibits stronger inhibitory effect (69).

250   Prediction of secondary structure between siRNA and target mRNA acts as an integral cue to the selection

251   of specific siRNA target site (69). Random folding of siRNA may lead to the inhibition of its RNAi

252   activity with an inappropriate secondary structure of siRNA-duplex can result in impediment to the RISC

253   (RNA-induced silencing complex) formation (70). Henceforth, evaluation of the thermodynamic outcome

254   for the interaction between the siRNA candidate and the target mRNA involves the assessment of the sum

255   of the energy required to unravel the binding site and the energy gained from the resultant hybridization

256   process (56). In our study, all the 19 siRNAs that were analyzed for feasibility in binding to target mRNA

257   exhibited highly negative ΔG values for binding to target sequences, thus indicating thermodynamically

258   feasible target binding.

259   The designing of therapeutic siRNA candidates involves the analysis of heat stability of the siRNA for the

260   evaluation of its in vivo stability and functional efficiency (71). In our study, all the 19 siRNA candidates

261   analyzed for heat stability were found to exhibit melting temperatures considerably greater than the

262   physiological temperature, thus indicating structural integrity in the host system. Consequently,

263   assessment of target accessibility of the siRNA candidates indicates the efficiency of efficiency of RISC

264   mediated endonucleolytic cleavage, which is the final step in the biological mechanism of gene silencing

265   by siRNA (72). siRNA 5, siRNA 49 and siRNA 50 exhibited the best scores for assessment of target

266   accessibility; i.e- least negative "Break targ. ΔG" values.

267   The immune system is armed with the required machinery to recognize foreign RNA sequences, resulting

268   in the mediation of activation of pattern recognition regions (PRR) for the clearance of the exogenous

269   components (73). Thereby, the immunogenicity of an RNA sequence in the case of siRNA-based therapy

270   may lead to immunotoxicity (74). Our assessment of immunogenicity of the candidate siRNAs showed

271   that a total of 12 siRNAs were non-immunogenic (Table 2) and thus were considered for tertiary structure

272   validation. Owing to the small number of siRNA molecules being evaluated by X-ray crystallography,

273   RMN spectroscopy, and cryoelectronic microscopy (cryo-EM), tertiary structure prediction of siRNA is

274   integral in understanding the respective structure-function relationship (75). Among the 12 siRNA

275   candidates that were analyzed for tertiary structure validation, only three siRNA candidates (siRNA 9,

276   siRNA 15 and siRNA 17) satisfied all threshold scores for nucleic acid chemistry parameters such as:

277   RNA sugar puckers, RNA backbone conformations and bond angles (Table 3).

278   Henceforth, the siRNA candidates that were found to have an acceptable tertiary structure are intended to

279   halt the translation of distinct conserved virulence genes of *Shigella* sp. siRNA 9 is intended to bind to the

280   mRNA of Ipg D gene, in the process halting the expression of the IpD protein that is involved with the

281   entry of the bacteria through ruffling of the host membrane (34). Thus, siRNA 9 is expected to protect the

282   host system through the maintenance of host membrane integrity. siRNA 15 and siRNA 17 are intended

283   to target the expression of the OspB gene, thereby halting the process of host inflammatory response and

284   resulting tissue damage (34).

285   **Limitations of the study**

286   Despite the long-term prospects of the findings intended for the development of a novel therapeutic

287   strategy against shigellosis, our study has several limitations. First of all, our analysis is solely based on

288   the sequences of the isolates reported in the NCBI database and may not necessarily be applicable for

289   isolates with novel mutations in these conserved virulence genes that may be reported in the future.

290   Moreover, the implementation of siRNA in regular therapeutic applications is a distant prospect and

291   requires extensive wet lab validation regarding the mode of delivery of the siRNA into the host system as

292   well as modifications in the siRNA sequences may be required for enhanced efficacy in the host system.

293   **Conclusion**

294   In this study, we have proposed three distinct siRNA candidates that were found to have substantial

295   functional efficacy, non-immunogenic and have an acceptable tertiary structure that is integral for the

296   effectiveness of siRNA in the host system. Our study provides insights into the development of a new

297   form of therapeutics against shigellosis in the form of siRNA. Such future therapeutic strategies may have

298   promising implications in combating the rapid development of anti-microbial resistance among *Shigella*

299   sp and the emergence of multi-drug resistant isolates.

300

301

302

303

304

305

306

307   **Methods and Materials**

308   An overview of the methodology followed in our study is illustrated by **Figure 1**.

309   **Sequence retrieval and multiple sequence alignment for determination of conserved regions**

310   Complete gene sequences for 15 different conserved virulent proteins of multiple isolates of *Shigella* sp.

311   were obtained from NCBI Nucleotide (https://www.ncbi.nlm.nih.gov/nucleotide). These conserved

312   virulent proteins of *Shigella* sp. included: IcsA (from 40 isolates), IpaA (from 64 isolates), IpaB (from 55

313   isolates), IpaC (from 41 isolates), IpaJ (from 57 isolates), IpgB (from 46 isolates), IpgD (from 42

314   isolates), MxiC (from 62 isolates), OspB (from 68 isolates), OspF (from 28 isolates), OspG (57 isolates),

315   Spa33 (from 21 isolates), VirA (from 50 isolates), VirB (from 61 isolates) and VirF (68 isolates).

316   Identification of conserved regions from these 15 different virulent proteins of *Shigella* sp was done by

317   multiple sequence alignment using Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/).

318   **Figure 1. Overview of the workflow of the methodology followed in the study.**

319 **Recognition of target sequence and des igning of potential siRNA candidates**

320 For the purpose of identification of target sequence and siRNA designing, siDirect 2.0

321 (http://sidirect2.rnai.jp/) an efficient and target-specific siRNA designing tool was used (47). This tool

322 employs a combination of first-generation algorithms for siRNA designing, namely: Ui-Tei,

323 Amarzguioui, Reynolds rules (URA rules) along with a melting temperature of less than 21.5 °C as the

324 absolute parameters for prediction of potential siRNA duplex formation (76). Moreover, a multitude of

325 other components that were also taken into consideration on the concept of the URA rules as shown in

326 Supplementary Table 1 have been used in a number of previously published literatures (42).

327 The i-SCORE Designer, an online software (48) was used for the validation of the potential siRNA

328 candidates obtained from siDirect 2.0. This web-based software utilizes multiple second generation

329 algorithms in addition to the first-generation algorithms for siRNA designing (U, R, A rules), among

330 which i-SCORE, s-Biopredsi, Katoh and DSIR rules are the prominent second-generation algorithms

331 (77)-(78)-(79). Only those siRNA candidates that were found to fulfill the threshold set by the criteria for

332 the second-generation algorithms were subsequently selected for further downstream validation.

Determination of off-target similarity

Off-target sequence similarity for the guide strands of the siRNA candidates was checked using the BLAST tool (http://www.ncbi.nlm.nih.gov/blast) against the entire GenBank database with default threshold value of 10 and BLOSUM 62 as parameter.

Prediction of free energy of folding and calculation of GC content

The free energy of folding of the guide strand of the siRNA candidates was assessed using RNAstructure (https://rna.urmc.rochester.edu/RNAstructureWeb/), a web-based tool for prediction of secondary structure of RNA (80). Only those siRNA candidates that had exhibited a positive free energy of folding (positive ΔG) were used for the subsequent process of GC content calculation by using DNA/RNA GC Content Calculator (http://www.endmemo.com/bio/gc.php) and siRNA candidates with a GC content of 30-60% were subsequently selected for downstream validation.

Evaluation of the thermodynamics secondary structure formed between siRNA and target and visualization of siRNA-target binding

Investigation of the thermodynamics for the secondary structure formed between the guide strand and subsequent validation of the siRNA-target duplex was done using the web-based tool for RNA secondary structure prediction, RNAStructure (https://rna.urmc.rochester.edu/RNAstructureWeb/). This tool determines the hybridization energy and base pairing from two RNA sequences by following the functional algorithm of McCaskill's partition to compute probabilities of base pairing, realistic communication energies and equilibrium concentrations of duplex structures (80)-(81).

Determination of heat stability and prediction of functional efficiency and target accessibility of the potential siRNA candidates

For the determination of heat stability of the guide strand as well as for the assessment of functional efficiency and target accessibility of the predicted siRNA candidates, we used the features of OligoWalk

356   (http://rna.urmc.rochester.edu/cgi-bin/server_exe/oligowalk/oligowalk_form.cgi), a web-based server for

357   calculating thermodynamic features of sense-antisense hybridization (82). This tool operates through a

358   designated query involving the sequence for the guide strand of the siRNA candidate and the subsequent

359   results obtained are expressed as: "End-diff (free energy difference between the 5' and 3' end of the

360   antisense strand of siRNA)", which indicates denotes the functional efficiency of the siRNA candidate;

361   "Break-targ. ΔG (free energy cost for opening base pairs in the region of complementarity to the target)",

362   which signifies the target accessibility of the designed siRNA and "Probability score of being efficient

363   siRNA", which is calculated on the basis of both target accessibility and functional efficiency of the

364   designed siRNA (82)-(59).

365   **Validation of the functional efficiency of the siRNA candidates**

366   siRNAPred server (http://crdd.osdd.net/raghava/sirnapred/index.html) was used for the validation of the

367   functional efficiency of the predicted siRNA candidates against the Main21 dataset using support vector

368   machine algorithm and the binary pattern prediction approach (60). Validation scores from the server

369   greater than 1 predicts very high efficiency, scores ranging from 0.8-0.9 predict high efficiency and scores

370   ranging from 0.7-0.8 predicts moderate efficiency (60).

371   **Prediction of immunotoxicity of the predicted siRNA candidates**

372   For the purpose of prediction of immunotoxicity of the designed siRNA candidates, imRNA

373   (https://webs.iiitd.edu.in/raghava/imrna/sirna.php), a web-based server consisting of various components

374   integral for the designing of RNA-based therapeutics was employed (83). IMscore of 4.5 set as default

375   was used for the prediction of immunogenicity of potential siRNA candidates.

376   **Designing of tertiary (3D) structure of the siRNA candidates and validation**

377   Designed siRNA candidates that had surpassed the immunogenicity filter set in the previous step were

378   subject to tertiary structure prediction using RNAComposer (http://rnacomposer.cs.put.poznan.pl/), an

379   automated tool that predicts the tertiary structure from a linear siRNA sequence (84)-(85). Subsequently,

380   the tertiary structures of these siRNA candidates were validated using the MOLprobity web server

381   (http://molprobity.biochem.duke.edu/) (62), which uses the pdb file of the tertiary structure of the siRNA

382   as input to predict the validity of the tertiary structure on the basis of all-atom contacts and geometry,

383   RNA sugar puckers, RNA backbone conformations, hydrogen bonds and Van der Waals forces (62)-(86).

384   The tertiary structures of the candidate siRNAs with an acceptable tertiary structure were visualized using

385   the PyMOL Molecular Graphics System (v1.8.4).

386   **Supplementary Information**

387   **Supplementary File 1. List of conserved sequences from the 15 virulent genes of *Shigella* sp.**

388   **Supplementary Table 1. Additional rules applied along with the URA rules for siRNA designing.**

389   **Funding**

390   This research received no external funding.

391   **Acknowledgments**

392   The authors thank Mohammad Umer Sharif Shohan, Lecturer, Department of Biochemistry and

393   Molecular Biology, University of Dhaka for his insightful suggestions during the work. ICDDR,B is

394   grateful to the Government of Bangladesh, Canada, Sweden, and the UK for providing unrestricted core

395   support. The authors gratefully acknowledge these donors for their support and commitment to the

396   endeavors of ICDDR,B.

397   **Author Contributions**

398   P.P, N.B and Z.N conceived the study; P.P and F.T.C designed the study; B.S, S.N.M, M.U.S.S and

399   U.F.C were involved in the acquisition of the data; P.P, M.K, T.J.S and S.N were involved in the data

400   acquisition. P.P, T.J.S, Z.N and TA were involved in drafting of the manuscript. All authors read and

401   approved of the manuscript.

402   **Institutional Review Board Statement**

403   Not applicable.

404   **Informed Consent Statement**

405   Not applicable.

406   **Data Availability Statement**

407   All relevant data are contained within this article.

408   **Conflict of Interest**

409   The authors declare no conflict of interest.

References

410

411 1.   Troeger C, Forouzanfar M, Rao PC, Khalil I, Brown A, Reiner Jr RC, et al. Estimates of global,
412 regional, and national morbidity, mortality, and aetiologies of diarrhoeal diseases: a systematic analysis
413 for the Global Burden of Disease Study 2015. The Lancet Infectious Diseases. 2017;17(9):909-48.

414 2.   Baker S, The HC. Recent insights into Shigella: a major contributor to the global diarrhoeal
415 disease burden. Current opinion in infectious diseases. 2018;31(5):449.

416 3.   Thompson CN, Duy PT, Baker S. The rising dominance of Shigella sonnei: an intercontinental
417 shift in the etiology of bacillary dysentery. PLoS Negl Trop Dis. 2015;9(6):e0003708.

418 4.   Vinh H, Nhu NTK, Nga TVT, Duy PT, Campbell JI, Hoang NVM, et al. A changing picture of
419 shigellosis in southern Vietnam: shifting species dominance, antimicrobial susceptibility and clinical
420 presentation. BMC infectious diseases. 2009;9(1):204.

421 5.   Bangtrakulnonth A, Vieira AR, Lo Fo Wong DM, Pornreongwong S, Pulsrikarn C, Sawanpanyalert
422 P, et al. Shigella from humans in Thailand during 1993 to 2006: spatial-time trends in species and
423 serotype distribution. Foodborne pathogens and disease. 2008;5(6):773-84.

424 6.   Ud-Din AI, Wahid SU, Latif HA, Shahnaij M, Akter M, Azmi IJ, et al. Changing trends in the
425 prevalence of Shigella species: emergence of multi-drug resistant Shigella sonnei biotype g in
426 Bangladesh. PloS one. 2013;8(12):e82601.

427 7.   Riddle MS. Is a Shigella vaccine needed for travellers and the military? Journal of travel
428 medicine. 2018;25(1):tay049.

429 8.   Klontz KC, Singh N. Treatment of drug-resistant Shigella infections. Expert review of anti-
430 infective therapy. 2015;13(1):69-80.

431 9.   Murphy GS, Bodhidatta L, Echeverria P, Tansuphaswadikul S, Hoge CW, Imlarp S, et al.
432 Ciprofloxacin and loperamide in the treatment of bacillary dysentery. Annals of internal medicine.
433 1993;118(8):582-6.

434 10.   Darton TC, Tuyen HT, Newton PN, Dance DA, Phetsouvanh R, Davong V, et al. Azithromycin
435 resistance in Shigella spp. in Southeast Asia. Antimicrobial agents and chemotherapy. 2018;62(4).

436 11.   Williams PC, Berkley JA. Guidelines for the treatment of dysentery (shigellosis): a systematic
437 review of the evidence. Paediatrics and international child health. 2018;38(sup1):S50-S65.

438 12.   Kotloff KL, Riddle MS, Platts-Mills JA, Pavlinac P, Zaidi AK. Shigellosis. The Lancet.
439 2018;391(10122):801-12.

440 13.   Brown JD, Willcox SJ, Franklin N, Hazelton B, Howard P, Reinten T, et al. Shigella species
441 epidemiology and antimicrobial susceptibility: the implications of emerging azithromycin resistance for
442 guiding treatment, guidelines and breakpoints. Journal of Antimicrobial Chemotherapy.
443 2017;72(11):3181-6.

444 14.   Camacho AI, Irache JM, Gamazo C. Recent progress towards development of a Shigella vaccine.
445 Expert review of vaccines. 2013;12(1):43-55.

446   15.   Giersing BK, Porter CK, Kotloff K, Neels P, Cravioto A, MacLennan CA. How can controlled human
447   infection models accelerate clinical development and policy pathways for vaccines against Shigella?
448   Vaccine. 2019;37(34):4778-83.

449   16.   Livio S, Strockbine NA, Panchalingam S, Tennant SM, Barry EM, Marohn ME, et al. Shigella
450   isolates from the global enteric multicenter study inform vaccine development. Clinical Infectious
451   Diseases. 2014;59(7):933-41.

452   17.   Taylor DN, Trofa A, Sadoff J, Chu C, Bryla D, Shiloach J, et al. Synthesis, characterization, and
453   clinical evaluation of conjugate vaccines composed of the O-specific polysaccharides of Shigella
454   dysenteriae type 1, Shigella flexneri type 2a, and Shigella sonnei (Plesiomonas shigelloides) bound to
455   bacterial toxoids. Infection and immunity. 1993;61(9):3678-87.

456   18.   Ravenscroft N, Braun M, Schneider J, Dreyer AM, Wetter M, Haeuptle MA, et al.
457   Characterization and immunogenicity of a Shigella flexneri 2a O-antigen bioconjugate vaccine candidate.
458   Glycobiology. 2019;29(9):669-80.

459   19.   Cohen D, Atsmon J, Artaud C, Meron-Sudai S, Gougeon M-L, Bialik A, et al. Safety and
460   immunogenicity of a synthetic carbohydrate conjugate vaccine against Shigella flexneri 2a in healthy
461   adult volunteers: a phase 1, dose-escalating, single-blind, randomised, placebo-controlled study. The
462   Lancet Infectious Diseases. 2020.

463   20.   Kotloff KL, Taylor DN, Sztein MB, Wasserman SS, Losonsky GA, Nataro JP, et al. Phase I
464   evaluation of ΔvirG Shigella sonnei live, attenuated, oral vaccine strain WRSS1 in healthy adults.
465   Infection and immunity. 2002;70(4):2016-21.

466   21.   Barry EM, Pasetti MF, Sztein MB, Fasano A, Kotloff KL, Levine MM. Progress and pitfalls in
467   Shigella vaccine research. Nature reviews Gastroenterology & hepatology. 2013;10(4):245.

468   22.   Germane KL, Ohi R, Goldberg MB, Spiller BW. Structural and functional studies indicate that
469   Shigella VirA is not a protease and does not directly destabilize microtubules. Biochemistry.
470   2008;47(39):10241-3.

471   23.   Campbell-Valois F-X, Sachse M, Sansonetti PJ, Parsot C. Escape of actively secreting Shigella
472   flexneri from ATG8/LC3-positive vacuoles formed during cell-to-cell spread is facilitated by IcsB and VirA.
473   MBio. 2015;6(3).

474   24.   Tobe T, Nagai S, Okada N, Adter B, Yoshikawa M, Sasakawa C. Temperature-regulated
475   expression of invasion genes in Shigella flexneri is controlled through the transcriptional activation of
476   the virB gene on the large plasmid. Molecular microbiology. 1991;5(4):887-93.

477   25.   Emanuele AA, Adams NE, Chen Y-C, Maurelli AT, Garcia GA. Potential novel antibiotics from HTS
478   targeting the virulence-regulating transcription factor, VirF, from Shigella flexneri. The Journal of
479   antibiotics. 2014;67(5):379-86.

480   26.   Phalipon A, Sansonetti PJ. Shigella's ways of manipulating the host intestinal innate and adaptive
481   immune system: a tool box for survival? Immunology and cell biology. 2007;85(2):119-29.

482   27.   Martinez-Becerra FJ, Kissmann JM, Diaz-McNair J, Choudhari SP, Quick AM, Mellado-Sanchez G,
483   et al. Broadly protective Shigella vaccine based on type III secretion apparatus proteins. Infection and
484   immunity. 2012;80(3):1222-31.

485 28.   Martinez-Becerra FJ, Chen X, Dickenson NE, Choudhari SP, Harrison K, Clements JD, et al.
486 Characterization of a novel fusion protein from IpaB and IpaD of Shigella spp. and its potential as a pan-
487 Shigella vaccine. Infection and immunity. 2013;81(12):4470-7.

488 29.   Chitradevi STS, Kaur G, Uppalapati S, Yadav A, Singh D, Bansal A. Co-administration of rIpaB
489 domain of Shigella with rGroEL of S. Typhi enhances the immune responses and protective efficacy
490 against Shigella infection. Cellular & molecular immunology. 2015;12(6):757-67.

491 30.   Heine SJ, Diaz-McNair J, Martinez-Becerra FJ, Choudhari SP, Clements JD, Picking WL, et al.
492 Evaluation of immunogenicity and protective efficacy of orally delivered Shigella type III secretion
493 system proteins IpaB and IpaD. Vaccine. 2013;31(28):2919-29.

494 31.   Handa Y, Suzuki M, Ohya K, Iwai H, Ishijima N, Koleske AJ, et al. Shigella IpgB1 promotes
495 bacterial entry through the ELMO–Dock180 machinery. Nature cell biology. 2007;9(1):121-8.

496 32.   Sun CH, Wacquier B, Aguilar DI, Carayol N, Denis K, Boucherie S, et al. The Shigella type III
497 effector IpgD recodes Ca2+ signals during invasion of epithelial cells. The EMBO journal.
498 2017;36(17):2567-80.

499 33.   Niebuhr K, Jouihri N, Allaoui A, Gounon P, Sansonetti PJ, Parsot C. IpgD, a protein secreted by
500 the type III secretion machinery of Shigella flexneri, is chaperoned by IpgE and implicated in entry focus
501 formation. Molecular microbiology. 2000;38(1):8-19.

502 34.   Mattock E, Blocker AJ. How do the virulence factors of Shigella work together to cause disease?
503 Frontiers in cellular and infection microbiology. 2017;7:64.

504 35.   Schuch R, Maurelli AT. Spa33, a cell surface-associated subunit of the Mxi-Spa type III secretory
505 pathway of Shigella flexneri, regulates Ipa protein traffic. Infection and immunity. 2001;69(4):2180-9.

506 36.   Botteaux A, Sory MP, Biskri L, Parsot C, Allaoui A. MxiC is secreted by and controls the substrate
507 specificity of the Shigella flexneri type III secretion apparatus. Molecular microbiology. 2009;71(2):449-
508 60.

509 37.   Zurawski DV, Mumy KL, Faherty CS, McCormick BA, Maurelli AT. Shigella flexneri T3SS effectors
510 OspB and OspF target the nucleus to down-regulate the host inflammatory response via interactions
511 with retinoblastoma protein. Molecular microbiology. 2009;71(2):350.

512 38.   Kim DW, Lenzen G, Page A-L, Legrain P, Sansonetti PJ, Parsot C. The Shigella flexneri effector
513 OspG interferes with innate immune responses by targeting ubiquitin-conjugating enzymes. Proceedings
514 of the National Academy of Sciences. 2005;102(39):14046-51.

515 39.   Whitehead KA, Langer R, Anderson DG. Knocking down barriers: advances in siRNA delivery.
516 Nature reviews Drug discovery. 2009;8(2):129-38.

517 40.   Pei Y, Tuschl T. On the art of identifying effective and specific siRNAs. Nature methods.
518 2006;3(9):670-6.

519 41.   Islam MO, Palit P, Shawon J, Hasan MK, Mahmud A, Mahfuz M, et al. Exploring novel therapeutic
520 strategies against vivax malaria through an integrated computational investigation to inhibit the
521 merozoite surface protein– 1 of Plasmodium vivax. Informatics in Medicine Unlocked. 2020;21:100471.

522 42.   Chowdhury FT, Shohan MU, Islam T, Mimu TT, Palit P. A Therapeutic Approach Against
523 Leishmania donovani by Predicting RNAi Molecules Against the Surface Protein, gp63. Current
524 Bioinformatics. 2019;14(6):541-50.

525 43.   Chowdhury UF, Shohan MUS, Hoque KI, Beg MA, Siam MKS, Moni MA. A computational
526 approach to design potential siRNA molecules as a prospective tool for silencing nucleocapsid
527 phosphoprotein and surface glycoprotein gene of SARS-CoV-2. Genomics. 2020.

528 44.   Mohanty PS, Sharma S, Naaz F, Kumar D, Raikwar A, Patil SA. Inhibition of Mycobacterium
529 tuberculosis tRNA-Ligases Using siRNA-Based Gene Silencing Method: A Computational Approach.
530 Journal of Computational Biology. 2020;27(1):91-9.

531 45.   Thornbrough JM, Hundley T, Valdivia R, Worley MJ. Human genome-wide RNAi screen for host
532 factors that modulate intracellular Salmonella growth. PloS one. 2012;7(6):e38097.

533 46.   Kühbacher A, Emmenlauer M, Rämo P, Kafai N, Dehio C, Cossart P, et al. Genome-wide siRNA
534 screen identifies complementary signaling pathways involved in Listeria infection and reveals different
535 actin nucleation mechanisms during Listeria cell invasion and actin comet tail formation. MBio.
536 2015;6(3).

537 47.   Naito Y, Yoshimura J, Morishita S, Ui-Tei K. siDirect 2.0: updated software for designing
538 functional siRNA with reduced seed-dependent off-target effect. BMC bioinformatics. 2009;10(1):1-8.

539 48.   Ichihara M, Murakumo Y, Masuda A, Matsuura T, Asai N, Jijiwa M, et al. Thermodynamic
540 instability of siRNA duplex is a prerequisite for dependable prediction of siRNA activities. Nucleic acids
541 research. 2007;35(18):e123.

542 49.   Hajiaghayi M, Condon A, Hoos HH. Analysis of energy-based algorithms for RNA secondary
543 structure prediction. BMC bioinformatics. 2012;13(1):22.

544 50.   Mathews DH. Predicting a set of minimal free energy RNA secondary structures common to two
545 sequences. Bioinformatics. 2005;21(10):2246-53.

546 51.   Ding Y, Chan CY, Lawrence CE. RNA secondary structure prediction by centroids in a Boltzmann
547 weighted ensemble. Rna. 2005;11(8):1157-66.

548 52.   Liu Y, Chang Y, Zhang C, Wei Q, Chen J, Chen H, et al. Influence of mRNA features on siRNA
549 interference efficacy. Journal of bioinformatics and computational biology. 2013;11(03):1341004.

550 53.   Shao Y, Chan CY, Maliyekkel A, Lawrence CE, Roninson IB, Ding Y. Effect of target secondary
551 structure on RNAi efficiency. Rna. 2007;13(10):1631-40.

552 54.   Umu SU, Gardner PP. A comprehensive benchmark of RNA–RNA interaction prediction tools for
553 all domains of life. Bioinformatics. 2017;33(7):988-96.

554 55.   Nazipova NN, Shabalina SA. Understanding off-target effects through hybridization kinetics and
555 thermodynamics. Springer; 2020.

556 56.   Mückstein U, Tafer H, Hackermüller J, Bernhart SH, Stadler PF, Hofacker IL. Thermodynamics of
557 RNA–RNA binding. Bioinformatics. 2006;22(10):1177-82.

558 57.   Bernhart SH, Tafer H, Mückstein U, Flamm C, Stadler PF, Hofacker IL. Partition function and base
559 pairing probabilities of RNA heterodimers. Algorithms for Molecular Biology. 2006;1(1):3.

560   58.   Reynolds A, Leake D, Boese Q, Scaringe S, Marshall WS, Khvorova A. Rational siRNA design for
561   RNA interference. Nature biotechnology. 2004;22(3):326-30.

562   59.   Mathews DH. Using OligoWalk to identify efficient siRNA sequences. RNA Therapeutics:
563   Springer; 2010. p. 107-19.

564   60.   Kumar M, Lata S, Raghava G, editors. siRNApred: SVM based method for predicting efficacy
565   value of siRNA. Proceedings of the first international conference on Open Source for Computer Aided
566   Drug Discovery (OSCADD); 2009: CSIR-IMTECH.

567   61.   Biesiada M, Purzycka KJ, Szachniuk M, Blazewicz J, Adamiak RW. Automated RNA 3D structure
568   prediction with RNAComposer. RNA Structure Determination: Springer; 2016. p. 199-215.

569   62.   Davis IW, Leaver-Fay A, Chen VB, Block JN, Kapral GJ, Wang X, et al. MolProbity: all-atom
570   contacts and structure validation for proteins and nucleic acids. Nucleic acids research.
571   2007;35(suppl_2):W375-W83.

572   63.   Dana H, Chalbatani GM, Mahmoodzadeh H, Karimloo R, Rezaiean O, Moradzadeh A, et al.
573   Molecular mechanisms and biological functions of siRNA. International journal of biomedical science:
574   IJBS. 2017;13(2):48.

575   64.   Ryther R, Flynt A, Phillips Jr, Patton J. siRNA therapeutics: big potential from small RNAs. Gene
576   therapy. 2005;12(1):5-11.

577   65.   Taxman DJ, Livingstone LR, Zhang J, Conti BJ, Iocca HA, Williams KL, et al. Criteria for effective
578   design, construction, and gene knockdown by shRNA vectors. BMC biotechnology. 2006;6(1):1-16.

579   66.   ElHefnawi M, Kim T, Kamar MA, Min S, Hassan NM, El-Ahwany E, et al. In silico design and
580   experimental validation of siRNAs targeting conserved regions of multiple hepatitis C virus genotypes.
581   PloS one. 2016;11(7):e0159211.

582   67.   Jackson AL, Linsley PS. Recognizing and avoiding siRNA off-target effects for target identification
583   and therapeutic application. Nature reviews Drug discovery. 2010;9(1):57-67.

584   68.   Hajiaghayi M, Condon A, Hoos HH. Analysis of energy-based algorithms for RNA secondary
585   structure prediction. BMC bioinformatics. 2012;13(1):1-11.

586   69.   Heale BS, Soifer HS, Bowers C, Rossi JJ. siRNA target site secondary structure predictions using
587   local stable substructures. Nucleic acids research. 2005;33(3):e30-e.

588   70.   Vickers TA, Wyatt JR, Freier SM. Effects of RNA secondary structure on cellular antisense activity.
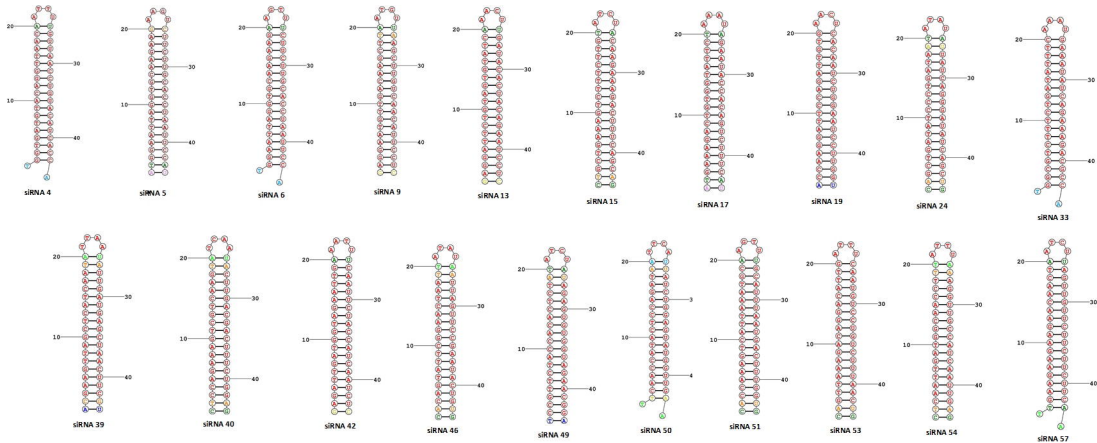589   Nucleic acids research. 2000;28(6):1340-7.

590   71.   Nawrot B, Sipa K. Chemical and structural diversity of siRNA molecules. Current topics in
591   medicinal chemistry. 2006;6(9):913-25.

592   72.   Tafer H, Ameres SL, Obernosterer G, Gebeshuber CA, Schroeder R, Martinez J, et al. The impact
593   of target site accessibility on the design of effective siRNAs. Nature biotechnology. 2008;26(5):578-83.

594   73.   Broz P, Monack DM. Newly described pattern recognition receptors team up against
595   intracellular pathogens. Nature Reviews Immunology. 2013;13(8):551-65.

596   74.   Fedorov Y, Anderson EM, Birmingham A, Reynolds A, Karpilow J, Robinson K, et al. Off-target
597   effects by siRNA can induce toxic phenotype. Rna. 2006;12(7):1188-96.

598   75.   Hajdin CE, Ding F, Dokholyan NV, Weeks KM. On the significance of an RNA tertiary structure
599   prediction. Rna. 2010;16(7):1340-9.

600   76.   Naito Y, Ui-Tei K. Designing functional siRNA with reduced off-target effects. siRNA Design:
601   Springer; 2013. p. 57-68.

602   77.   Takasaki S, Kotani S, Konagaya A. An effective method for selecting siRNA target sequences in
603   mammalian cells. Cell cycle. 2004;3(6):788-93.

604   78.   Huesken D, Lange J, Mickanin C, Weiler J, Asselbergs F, Warner J, et al. Design of a genome-wide
605   siRNA library using an artificial neural network. Nature biotechnology. 2005;23(8):995-1001.

606   79.   Katoh T, Suzuki T. Specific residues at every third position of siRNA shape its efficient RNAi
607   activity. Nucleic acids research. 2007;35(4):e27.

608   80.   Reuter JS, Mathews DH. RNAstructure: software for RNA secondary structure prediction and
609   analysis. BMC bioinformatics. 2010;11(1):1-9.

610   81.   Seetin MG, Mathews DH. RNA structure prediction: an overview of methods. Bacterial
611   Regulatory RNA: Springer; 2012. p. 99-122.

612   82.   Lu ZJ, Mathews DH. OligoWalk: an online siRNA design tool utilizing hybridization
613   thermodynamics. Nucleic acids research. 2008;36(suppl_2):W104-W8.

614   83.   Chaudhary K, Nagpal G, Dhanda SK, Raghava GP. Prediction of Immunomodulatory potential of
615   an RNA sequence for designing non-toxic siRNAs and RNA-based vaccine adjuvants. Scientific reports.
616   2016;6(1):1-11.

617   84.   Antczak M, Popenda M, Zok T, Sarzynska J, Ratajczak T, Tomczyk K, et al. New functionality of
618   RNAComposer: application to shape the axis of miR160 precursor structure. Acta Biochimica Polonica.
619   2016;63(4):737-44.

620   85.   Popenda M, Szachniuk M, Antczak M, Purzycka KJ, Lukasiak P, Bartol N, et al. Automated 3D
621   structure composition for large RNAs. Nucleic acids research. 2012;40(14):e112-e.

622   86.   Williams CJ, Headd JJ, Moriarty NW, Prisant MG, Videau LL, Deis LN, et al. MolProbity: More and
623   better reference data for improved all-atom structure validation. Protein Science. 2018;27(1):293-315.

624

siRNA 4 · siRNA 5 · siRNA 6 · siRNA 9 · siRNA 13 · siRNA 15 · siRNA 17 · siRNA 19 · siRNA 24 · siRNA 33

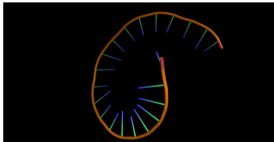siRNA 39 · siRNA 40 · siRNA 42 · siRNA 46 · siRNA 49 · siRNA 50 · siRNA 51 · siRNA 53 · siRNA 54 · siRNA 57

Probability >= 99%
99% > Probability >= 95%
95% > Probability >= 90%
90% > Probability >= 80%
80% > Probability >= 70%
70% > Probability >= 60%
60% > Probability >= 50%
50% > Probability
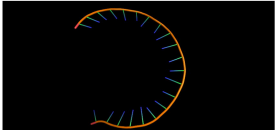
siRNA 9
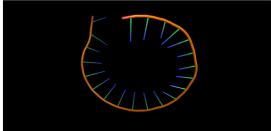
siRNA 15

siRNA 17

Designing of siRNA candidates by first-generation and second-generation algorithms for siRNA designing

Determination of off-target silencing by NCBI BLAST

Evaluation of minimum free energy (MFE) of folding by RNAstructure software and assessment of GC content by using DNA/RNA GC Content Calculator

Assessment of free energy of target binding, evaluation of target accessibility and analysis of heat stability of siRNA-target duplex by OligoWalk software

Validation of the functional efficiency of the predicted siRNA candidates by siRNAPred server

Assessment of immunogenicity of the siRNA candidates by imRNA server

Generation of 3D structure of the siRNA candidates by RNAComposer tool and validation of the generated tertiary structures by MOLprobity web server