



An *in-silico* approach to design potential siRNAs against the *ORF57* of Kaposi's sarcoma-associated herpesvirus

Anisur Rahman, Shipan Das Gupta, Md. Anisur Rahman*, Saheda Tamanna

Department of Biotechnology and Genetic Engineering, Faculty of Science, Noakhali Science and Technology University, Noakhali 3814, Bangladesh

Kaposi's sarcoma-associated herpesvirus (KSHV) is one of the few human oncogenic viruses, which causes a variety of malignancies, including Kaposi's sarcoma, multicentric Castleman disease, and primary effusion lymphoma, particularly in human immunodeficiency virus patients. The currently available treatment options cannot always prevent the invasion and dissemination of this virus. In recent times, siRNA-based therapeutics are gaining prominence over conventional medications as siRNA can be designed to target almost any gene of interest. The ORF57 is a crucial regulatory protein for lytic gene expression of KSHV. Disruption of this gene translation will inevitably inhibit the replication of the virus in the host cell. Therefore, the ORF57 of KSHV could be a potential target for designing siRNA-based therapeutics. Considering both sequence preferences and target site accessibility, several online tools (i-SCORE Designer, Sfold web server) had been utilized to predict the siRNA guide strand against the ORF57. Subsequently, off-target filtration (BLAST), conservancy test (fuzznuc), and thermodynamics analysis (RNAcofold, RNAalifold, and RNA Structure web server) were also performed to select the most suitable siRNA sequences. Finally, two siRNAs were identified that passed all of the filtration phases and fulfilled the thermodynamic criteria. We hope that the siRNAs predicted in this study would be helpful for the development of new effective therapeutics against KSHV.

Keywords: guide strand, human herpesvirus-8, Kaposi's sarcoma, RNA interference, siRNA

Introduction

Kaposi's sarcoma (KS) is a multifocal hyperproliferative neoplasm of the vascular or lymphatic endothelium. KS lesions may also occur on the skin surface of the upper and lower extremities or on the visceral organs like the lung and spleen [1-4]. Four epidemiological subtypes (Classic, Endemic, Epidemic, and Iatrogenic) of KS have been described so far and Kaposi's sarcoma-associated herpesvirus (KSHV), also known as human herpesvirus-8 was identified as the primary etiologic agent of these subtypes [5-7]. Besides, KSHV causes two lymphoproliferative diseases and an inflammatory syndrome: primary effusion lymphoma, multicentric Castleman disease, and KSHV inflammatory cytokine syndrome [8-10].

KSHV has the highest prevalence in sub-Saharan Africa, the intermediary prevalence in the Mediterranean, and low prevalence in northern Europe, North America, and Asia [4]. Previously, KS was 20,000 times more persistent in acquired immune deficiency syndrome (AIDS) patients than in the general population [11]. With the advent of combina-

tion antiretroviral therapy, the incidence of AIDS-related (epidemic) KS substantially decreased in developed countries [12]. But the rapid progression of AIDS and inaccessibility to antiretroviral drugs worsened the condition in several African countries [13]. Countries like Malawi, Uganda, Zimbabwe, and Swaziland have experienced a multitude of KS incidences since the onset of the AIDS epidemic in the early 1980s, making KS the most common cancer in males and the second most common cancer in females [14,15].

The KSHV genome has approximately 140.5 kb long unique region (LUR) flanked by 25-30 kb of direct terminal repeats [16,17]. The LUR comprises 90 open reading frames (ORFs) that separately express during the latent or lytic phase of viral infection [18]. Though *ORF50/Rta* (replication and transcription activator) initiates the entire KSHV lytic cycle, its completion depends on *ORF57* that ensures the orderly expression of other early and late lytic genes [19]. The *ORF57/Mta* (mRNA transcript accumulation) is a 455 amino acid long nuclear protein with no substantial sequence homology to any known cellular proteins [20]. This post-transcriptional regulator plays several vital roles in KSHV pathogenesis, for example, binding and stabilization of intron-less transcripts, facilitates intron splicing, nuclear accumulation of coding/non-coding RNAs, cytoplasmic accumulation of mRNAs, promoting protein translation, etc. [21]. Disruption of *ORF57* showed to interrupt the KSHV lytic gene expression and eventually halt virion production [22]. Therefore, it could be a potent target for therapeutic intervention to suppress the replication of KSHV in the host cells.

RNA interference is a unique molecular therapeutic technique in which specific non-coding RNAs, such as siRNA or miRNA, silence or downregulate the target mRNA in a sequence-specific manner. Although siRNA and miRNA share a common pathway, they have distinct functions, and in many cases, siRNA outcompeted miRNA in silencing activity. Since siRNA designing allows more sophistication in reducing off-target effects, it can inhibit one particular target mRNA. Besides, modern transfection techniques such as lipofection and nanoparticles have greatly simplified its delivery into cells [23]. siRNA-based treatment has been successfully demonstrated in influenza, hepatitis C, severe acute respiratory syndrome coronavirus 2, cancer, different autoimmune, and neurodegenerative diseases [24-28]. Hence, considering the clinical significance of KSHV, the role of *ORF57* in viral pathogenesis, and the therapeutic potentiality of RNA interference, we aimed in the present study to design effective siRNA(s) against *ORF57* of KSHV.

Methods

The overall pictorial representation of this study is depicted in

Fig. 1 in a stepwise manner.

Data mining

The coding sequence (CDS) of *ORF57* was obtained from the KSHV reference sequence (NC_009333.1) in the NCBI GenBank (<https://www.ncbi.nlm.nih.gov/genbank/>) database [29]. This reference CDS was used as a query in the NCBI nucleotide BLAST tool (<http://www.ncbi.nlm.nih.gov/blast>) to identify all the available CDS of *ORF57* of KSHV [30]. During BLAST search, non-redundant nucleotide collection (nr/nt) was selected as a database and the search was restricted to only KSHV sequences by selecting txid: 37296 in the organism section. The maximum number of target sequences was set to 1,000 and other parameters were kept as default. BLAST analysis resulted in a total of 76 sequences with 100% query coverage and a percent identity \geq 98.64%. The accession numbers of these 76 sequences were retrieved, and the CDS of *ORF57* from each accession were downloaded from NCBI using command-line code.

Generation of consensus sequence

In the UGENE v34.0 software [31], the CDS of all collected *ORF57* were subjected to multiple sequence alignment (MSA) by the MUSCLE [32] tool with default parameters. UGENE employs JalView [33] as a default algorithm to generate a consensus sequence from MSA. The consensus sequence constructed from this MSA analysis was considered to be the representative sequence of the *ORF57* of all KSHV strains.

Prediction of siRNA

The consensus sequence obtained from MSA was used to predict potential siRNAs against the *ORF57* gene. In this regard, two parameters (e.g., sequence features and target site accessibility) were taken into consideration [34]. The i-SCORE Designer web tool (https://www.med.nagoya-u.ac.jp/neurogenetics/i_score/i_score.html) has been deployed for sequence-based siRNA design [35]. This tool computes nine distinct algorithm scores (Ui-Tei, Amarzguioui, Hsieh, Takasaki, s-Biopredsi, i-Score, Reynolds, Kato, and DSIR) for siRNA prediction by analyzing different nucleotide preferences from target mRNA.

Sequence-focused algorithms can further be subdivided into two groups (rule-based and machine-learning aided) depending on their calculation nature [34]. Ui-Tei, Amarzguioui, and Reynolds scoring schemes had been taken into account for this study as rule-based approaches. The conditions of those rules are listed in [Supplementary Table 1](#). The i-Score (inhibitory score) algorithm, which implements a linear regression model to predict siRNAs, was

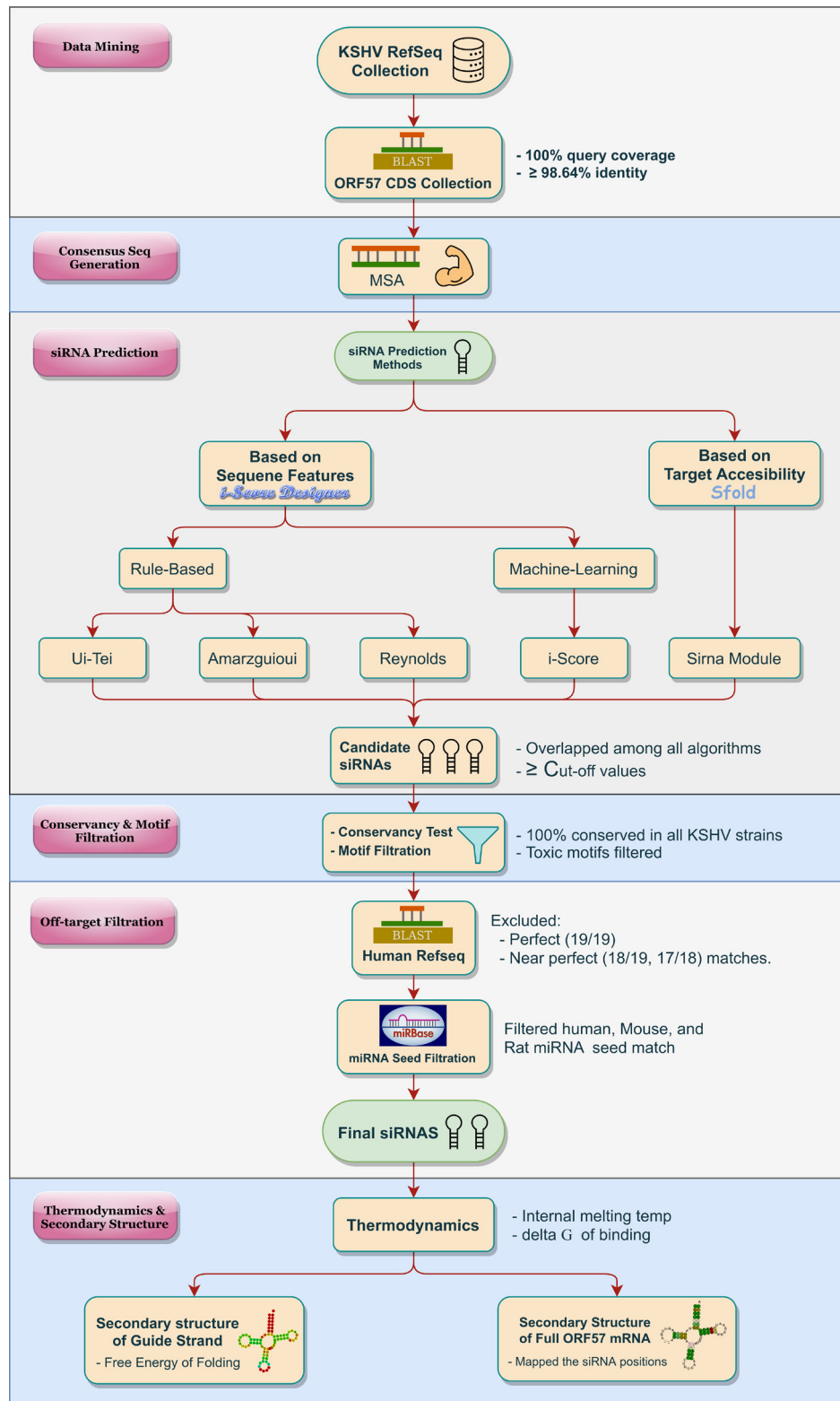


Fig. 1. Schematic diagram of the overall methodology. Designing of effective siRNA/(s) against the ORF57 of Kaposi's sarcoma-associated herpesvirus (KSHV) was divided into several phases. After initial sequence data collection and representative consensus sequence generation for ORF57, five distinct algorithms were adopted for siRNA prediction. Final siRNAs were selected through a rigorous filtration process (conservancy, toxic motif, off-target, miRNA seed). The secondary structure and thermodynamic properties of each siRNA were also evaluated. CDS, coding sequence; MSA, multiple sequence alignment.

employed for machine-learning aided siRNA designing. This method solely examines nucleotide preferences at each position to estimate the score. Finally, the Sirna module of the Sfold web server (<http://sfold.wadsworth.org>) was used to identify putative siRNAs based on target accessibility [36]. But the module also incorporates the sequence rules as well as accessibility criteria [37].

Only the results that scored equal to or greater than the recommended cutoff values of each of the five algorithms (Table 1) were taken. The siRNAs that overlapped across all five algorithms were then selected for subsequent analysis.

Conservancy test and motif filtration

Every selected siRNA target was tested for 100% conservancy among all KSHV strains. In the next step, the occurrence of some sequence motifs (“UGUGU,” “GUCCUCAA,” “AUCGAU(N)nGGGG,” “UGGC”) was screened for every siRNA candidate. Because such motifs are suggested to be immune-stimulatory by some experimental evidence and should be avoided during the design of siRNA [41]. These steps were conducted by employing the *fuzznuc* command-line tool from EMBOSS software package [42].

Filtration of off-target sites

A two-step filtration process was performed to assess the likelihood that candidate siRNAs would cause off-target effects. In the first step, the NCBI nucleotide BLAST was employed to screen a perfect (19/19) or near-perfect (18/19, 17/19) match of the human RefSeq mRNA database against both sense and antisense strands of candidate siRNAs [43,44]. BLAST’s default parameters are inefficient for evaluating very short sequences of siRNAs, thereby, some parameters have been customized according to the guidelines reported in Birmingham et al. [45]. The BLAST options used for this study to identify off-targets are listed in Supplementary Table 2. siRNAs showing complete or nearly complete complementarity with off-target mRNA are assumed to have off-target effects, and

Table 1. siRNA design algorithms and their cutoff values

Algorithms	Cutoff-value	Reference
Based on sequence features		
Rule-based		
Reynolds	≥ 6	[38]
Amarzguioui	≥ 3	[39]
Ui-Tei	la & lb	[40]
Machine learning approach		
i-score	≥ 66	[35]
Based on target accessibility		
Sfold Sirna module	≥ 12	[37]

therefore have been rejected.

In the second step of the off-target screening, the seed region (2nd to 7th nucleotide from 5’ end) of the selected siRNAs was compared to the seed region (2nd to 8th nucleotide from 5’ end) of miRNAs found in humans, mice, and rats. All the human, mouse, and rat miRNAs have been downloaded from miRBase (<http://www.mirbase.org/index.shtml>), a publicly accessible microRNA database of annotated miRNA sequences [46]. The siRNAs, whose seed region matched with the seed region of human, rat, and mouse miRNAs were eliminated in this phase.

Thermodynamic analysis

The internal melting temperature (T_m) of the sense strand of each candidate siRNA duplex was calculated by the OligoEvaluator analysis tool (<http://www.oligoevaluator.com>). To estimate the free energy of heterodimer binding (ΔG) between target mRNA and siRNA guide strand RNAfold web server (<http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi>) was used with defaults parameters [47]. This program calculates the base-pairing pattern and hybridization energy of interacting RNA strands by utilizing thermodynamic and kinetic properties. To estimate base-pairing probabilities it implements an extension of McCaskill’s partition function algorithm [48]. The frequency of the minimum free energy (MFE) structure in the ensemble and ΔG for heterodimer binding was calculated for every candidate siRNA duplex (sense and antisense strand).

Determination of secondary structure of guide strand and free energy of folding

The MaxExpect algorithm of the RNA Structure web server (<https://rna.urmc.rochester.edu/RNAstructureWeb/>) has been employed to appraise the secondary structure of the guide strand [49]. In addition to the structure prediction, the MaxExpect program produces CT files for each structure. The CT file obtained for each of the siRNA guide strands was then used as input for another RNA Structure web server tool called efn2. This efn2 algorithm calculates the folding free energy change of the secondary structure of the guide strand from the CT file [50].

Prediction of the secondary structure of full mRNA and mapping target sites

RNAalifold server from Vienna RNA Web Services has been employed to predict the secondary structure of *ORF57* mRNA [51]. This server generates a consensus secondary structure from the alignment of multiple related DNA or RNA sequences. The MSA file, previously generated by the MUSCLE algorithm, was supplied

as input in this phase. The new RNAalifold with the RIBOSUME scoring option was chosen as the RNAalifold version. Besides, other settings were kept in the default option. This tool provides an MFE structure that was downloaded in Vienna format. Vienna RNA Web Service also has a visualization tool called *forna* server that can be used to design and import RNA secondary structures interactively [52]. The secondary structure of full *ORF57* was drawn on this server with the Vienna file retrieved from the previous tool. In addition, the target sites of the chosen siRNAs have been mapped within the structure of this full mRNA.

Results

Prediction and selection of siRNA

After performing BLAST using CDS of *ORF57* obtained from KSHV RefSeq, we got *ORF57* sequences of 76 strains and isolates of KSHV for subsequent analysis. Accession numbers of all selected strains and isolates are listed in [Supplementary Table 3](#). A 1.3 kb long consensus sequence was generated from the 76 *ORF57* sequences using MSA of the MUSCLE tool (see [Supplementary Table 4](#)). Following a systematic screening, twelve candidate siRNAs were identified that scored at or above the recommended threshold value for each of the five algorithms specified above ([Supplementary Table 5](#)).

Out of these twelve, eight siRNAs displayed 100% conserved target sequences across all KSHV strains. The next step was to eliminate siRNAs with immune-stimulatory motifs (GUCCUCAA, UGUGU, AUCGAU, and UGGC) in their guide strands. Two of our siRNA guide strands were found to have the “UGGC” motif and thereby excluded from the candidate list. Though there is no clear evidence, some studies suggest avoiding siRNA candidates with low complexity motifs such as “AAAA”, “CCCC”, “GGGG”, or “UUUU” [41]. The Sfold algorithm filters out such guide strands in the case of siRNA duplexes with a total score of 12 or higher. The siRNAs that have passed the filtration steps so far have been named from siRNA_1 to siRNA_6 for ease of exposition ([Supplementary Table 6](#)).

For off-target filtration, BLAST analysis was performed with both strands of candidate siRNAs against the human genome to filter out the undesired siRNAs. Two siRNA duplexes (siRNA_3 and siRNA_6) were filtered out and excluded from the final list as those

possessed nearly identical (17/19 and 17/18) sequence segments in respect to the human genome ([Supplementary Table 7](#)). In the next phase of off-target filtration, siRNA_2 and siRNA_5 were excluded from the candidate list as their seed regions were shown to be similar to those of human miRNA seeds ([Supplementary Table 8](#)). There are differing opinions on the position and length of the seed region of miRNA (6-8 mers), but in higher mammals, the seed region is regarded to be in the 2nd to 8th position (7-mers) from the 5' end [45,53]. Therefore, in order to be on the safe side, the 2nd to 8th nucleotide (7-mers) of miRNA was considered as the seed region. Our recommended siRNA_1 and siRNA_4 satisfied all of the sequence properties and filtration conditions of this study. Eventually, siRNA_1 and siRNA_4 were found to meet all of the sequence properties and filtration conditions and hence selected as final siRNAs ([Table 2](#)). It is noteworthy to mention that, seed region of siRNA_4 matched with a mouse miRNA seed.

The percentage of GC content in each siRNA was also noted because low GC content can result in poor and nonspecific binding, whereas high GC content prevents the helicase and RNA-Induced Silencing Complex (RISC) complex from unwinding the siRNA duplex [54]. Many studies have proposed various acceptable GC content limits [38,39]. Considering all nucleotide preferences, Fakhr et al. [55] suggested that the GC content of siRNA should be between 36%–52%. All of the siRNAs selected in this study have a GC content within this range.

Thermodynamic attributes

The silencing machinery of siRNA is largely modulated by the thermodynamic stability of nucleotide base pairing [56]. The internal melting temperature (T_m) and free energy change (ΔG) between siRNA seed and mRNA target are reliable markers of the thermodynamic stability of such heteroduplexes [57]. Different thermodynamics properties of the selected siRNAs are listed in [Table 3](#). Both siRNAs were found to have internal melting temperatures (T_m) below 60°C. The free energy of heterodimer binding between siRNA and target mRNA is the consequence of two energy contributions. The first one is the energy used to open the binding site, and the other is the energy obtained from hybridization. The RNAfold web server calculates the free energy of heterodimer binding (ΔG) as per the following equation:

Table 2. Final siRNAs

Rank	Name	Start positions	Sense	Antisense	GC (%)
1	siRNA_1	294	CAGUAAACAGGUACGGUAA	UUACCGUACCUGUUUACUGgu	42.1
2	siRNA_4	700	CGACGAACUCAUAAACAAA	UUUGUUUUAUGAGUUCGUCguc	36.8

Table 3. Thermodynamics results

Rank	Name	Internal melting Tm (°C)	Frequency of the MFE structure in the ensemble (%)	ΔG for heterodimer binding (kcal/mol)	Free energy of folding (kcal/mol)
1	siRNA_1	54.2	50.54	-29.16	0.3
2	siRNA_4	48.9	42.73	-27.42	0.5

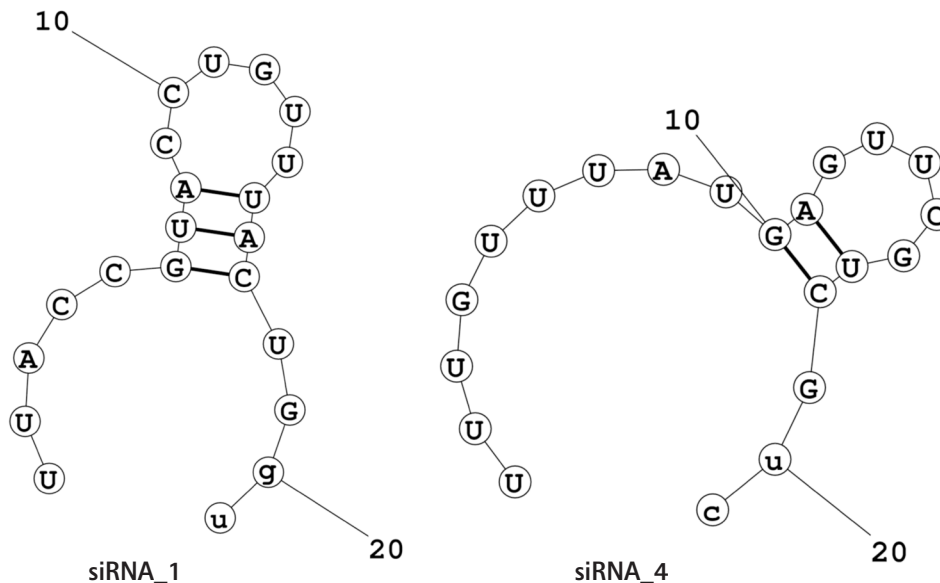


Fig. 2. Folding structure of guide RNAs. Potential intra-oligomer binding within the secondary structure of siRNA guide strands. The free energies for the folding structures of siRNA guide strands were also generated from these secondary structures. siRNA guide strand with the positive free energy of folding has a higher probability of binding with the target mRNA. Here, the positive free energy of the siRNA_1 and siRNA_4 for the folding structure of their guide strands are 0.3 kcal/mol and 0.5 kcal/mol, respectively.

$$\Delta G_{\text{Binding}} = \Delta G_{\text{AB}} - \Delta G_{\text{A}} - \Delta G_{\text{B}} [47].$$

The net ΔG value of the mRNA-siRNA duplex should be negative for better interaction. The higher the negative ΔG value, the more stable the duplex will form between siRNA and target mRNA [58]. But for proper siRNA silencing effect, this ΔG value should not be too high or too low, and a ΔG value between -35 and -27 kcal/mol yields better performance [59]. The values of ΔG for siRNA_1 and siRNA_4 were -29.16 kcal/mol and -27.42 kcal/mol, respectively. In the ensemble of secondary structures, the MFE structure of both siRNAs seemed to have a relatively higher frequency (50.54% for siRNA_1 and 42.73% for siRNA_4). The free energy for the folding of the selected siRNA guide strands was also calculated, along with their probable folding structures (Fig. 2). Both siRNA_1 and siRNA_4 tended to fold with a positive free energy (0.3 kcal/mol and 0.5 kcal/mol, respectively). Then, the selected siRNAs were ranked based on their off-target and thermodynamic properties, with siRNA_1 being the first and siRNA_4 in

the second position. Finally, target sites of the two selected siRNAs were mapped onto the secondary structure of this mRNA (Fig. 3).

Discussion

There are several algorithms, standalone tools, and web servers available for designing siRNA against target sequences, each with its own set of benefits and drawbacks [55]. These tools consider a diverse array of mechanisms and features to screen out the best possible siRNAs, and yet none of them can adopt the all-in-one strategy. That's why a somewhat manual hybrid approach was employed in this study. We utilized a combination of different algorithms (Reynolds, Amarzguioui, Ui-Tei, i-Score, and Sfold) to acquire siRNAs that met most of the prerequisites to be an effective gene silencer. For example, Reynolds, Amarzguioui, Ui-Tei algorithms are some of the most widely used rule-based algorithms, and many tools implement these algorithms for initial siRNA prediction [34,41,60,61]. Second-generation algorithms like i-Score, s-Bio-

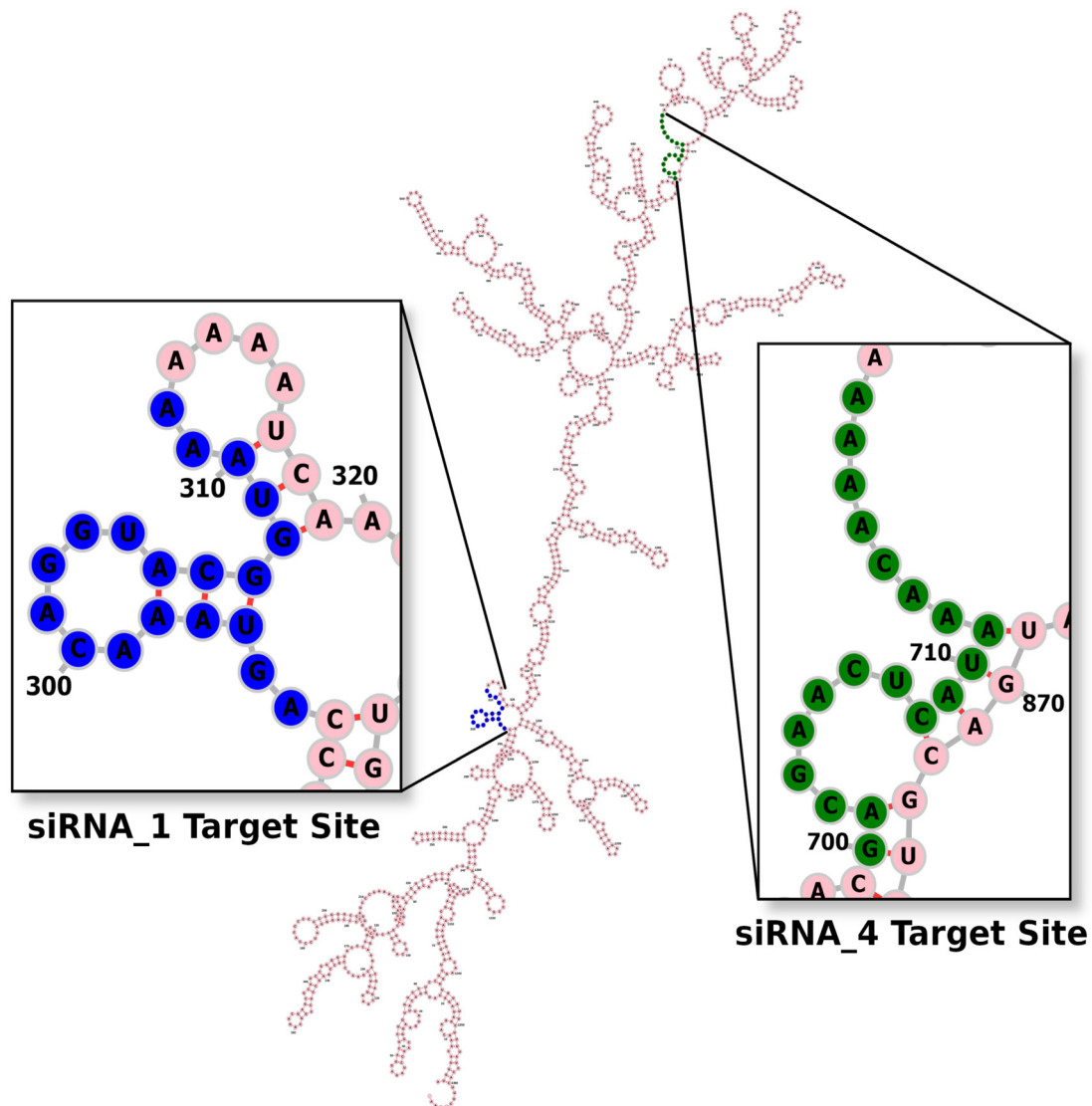


Fig. 3. Secondary structure of the target mRNA. From the multiple sequence alignment file generated by the MUSCLE tool, the RNAalifold server produced a consensus secondary structure of the entire ORF57 mRNA of Kaposi's sarcoma-associated herpesvirus. Target sites of both siRNA_1 and siRNA_4 (highlighted in blue and green color, respectively) seem to have an unpaired 5' or 3' end, which will favor their

predsi, and DSIR implement different machine-learning methods to predict the most promising siRNAs with almost equivalent accuracy [35,60]. On the other hand, Sfold uses a statistical algorithm-based probability profile approach to predict siRNA accessible regions in the secondary structure of target mRNA [36]. While other algorithms perform calculations based on 19-mer siRNA sequences, s-Biopredsi and DSIR consider 21-mer for their analysis [35]. That's why from the second-generation category, only the i-Score algorithm was taken into account to ensure uniformity in siRNA length and computation. Only those siRNAs were picked from all of the predicted siRNAs that met every key criterion of

these five methods. Furthermore, the full conservation of these siRNA target sites was also verified to ensure that our proposed siRNAs are effective against all KSHV strains.

Previously it was reported that siRNA sequences with "GUCCU-UCAA" and "UGUGU" motifs can be immune-stimulatory. For example, the "UGUGU" motif induces IFN type 1 and causes downregulation of nonspecific genes [62,63]. Another experiment demonstrated that monocytes are stimulated to generate large amounts of IL-12 when the CpG motif "AUCGAU" is present in RNA oligonucleotides along with a poly-G tail [64]. The existence of the "UGGC" motif in the siRNA guide strand has been found to

decrease cell viability [65]. The siRNAs which contained these sorts of motifs were omitted in this phase of screening.

After the conservancy and motif filtration analysis, both the sense and antisense of the selected siRNAs were investigated for the presence of off-target human gene and miRNA seed regions. The siRNA-targeted sequence of viral mRNA should not be identical or nearly identical to any human gene, otherwise, it will create an undesired silencing effect [66]. Off-target effects can be induced by either strand of siRNA (sense or antisense) [45]. Furthermore, siRNA guide strands with seed regions identical to the human miRNA were sieved out in the second phase of off-target filtration. Because miRNA-like off targeting may induce mRNA translation inhibition, which can also result in gene-nonspecific downregulation [67].

For effective siRNA activity, target sites should not be located in SNP sites, and/or in between the first 75 bases of mRNA's start codon or in intron sites [55]. Since all of the target sites for the candidate siRNAs of this study are fully conserved across all mRNA sequences, it can be assured that these target sequences are not in SNP sites. None of the selected siRNAs' target sites are within the first 75 bases of mRNA. Finally, the use of only coding sequences in this analysis eradicated any possibilities of the intron sequences being found in target sites (Table 2). GC content of siRNA has a significant impact on its silencing efficacy, and our suggested siRNAs have GC content in the recommended range [55].

The performance of siRNA is influenced not only by sequence features but also by structural and thermodynamic properties [68]. The incorporation of RNA strands in the RISC complex can be influenced by the thermodynamics of the siRNA duplex. More specifically, RNA strands having lower binding stability at the 5' end of the guide strand are preferred to be integrated into the RISC. Furthermore, the thermodynamic features of nucleotide base-pairing between the siRNA guide strand seed region and off-target mRNAs are primarily responsible for its off-target silencing activity [69]. Therefore, after selecting the desired siRNAs through a stringent filtration procedure, their thermodynamic attributes were also evaluated. The existence of hairpin structures in target mRNA sites can be anticipated by calculating the internal melting temperature (T_m) of the sense strands. Especially, if the T_m value of the siRNA sense strand is greater than 60°C, there is a high probability of hairpin formation, which will eventually reduce the knockdown ability [70]. None of our suggested siRNAs exceed this threshold value of T_m (Table 3).

One of the most essential factors for implementing siRNA target prediction algorithms is the computational estimation of RNA-RNA binding interaction. This RNA-RNA binding interaction is indicated by the free energy of binding (ΔG) between a specific siRNA and its

target mRNA. In case of weak duplex (mRNA-siRNA) stability, the RISC would not have enough opportunity to cleave the target mRNA. Contrarily, the passenger strand or target mRNA will separate slowly after cleavage by the RISC complex if the siRNA forms a very stable duplex with its antisense strand. All of the finalized siRNAs in this study have free energy values of heterodimer binding (ΔG) in between the recommended range (Table 3).

Another important indicator of the strength of base-pairing in a seed duplex complex is the frequency of MFE structure in the ensemble of secondary structures. A high frequency of MFE structure in the ensemble is correlated to a stiff seed duplex binding [71]. The accessibility of siRNA for binding, which can be projected by its folding free energy change, is found to be highly correlated with its efficacy for hybridization with target mRNA [72]. According to prior findings, an RNA molecule with the positive free energy of folding has a greater likelihood of binding to a target site because it will be more accessible [73]. Furthermore, intra-oligomer binding within the antisense strand of siRNA can significantly impede its accessibility for target mRNA [72]. That's why potential folding structures of selected siRNA guide strands were also predicted in this analysis (Fig. 2). Both siRNAs reported in this investigation (siRNA_1 and siRNA_4) appeared to have a high frequency of MFE structure and positive free energy of folding (Table 3).

Aside from choosing a siRNA guide strand, predicting the secondary structure of the target mRNA is also very imperative for RNAi activity. The previous evidence indicated that the secondary structure of complete mRNA should be postulated, as nucleotides distal from the target site can also modify its (target site's) structure [74]. But the number of potential secondary structures increases exponentially as the length of the sequence increases, making reliable mRNA secondary structure prediction challenging [72]. For that reason, the RNAalifold server, which predicts the consensus secondary structure of mRNA from MSA among several related sequences, was used in this experiment. Target mRNA with unpaired regions at either 5'-end or 3'-end can be silenced more effectively than a fully paired target [75]. As shown in Fig. 3, the target sites of our selected siRNAs contain an unpaired 5' or 3' end.

ORF57 of KSHV plays a vital role in viral lytic replication. Dimerization of ORF57 stabilizes the protein structure and is crucial for its functional activity. Each monomer of ORF57 homodimer is encoded from a single gene and consists of two distinct domains: intrinsically disordered N-terminal domain (NTD) with no defined structural motifs (amino acid residues 1–166), and helix-rich C-terminal domain (CTD) (amino acid residues 167–455) [76]. The NTD harbors three nuclear localization signals (NLSs) and interacts with several cellular factors [20,77]. The CTD has an N-ter-

minal 'arm' stretching from residues 167 to 222 and a C-terminal globular domain (aa residues 223–455) having a conserved zinc-binding motif. These three structural elements (arm, globular interface, and zinc-binding motif) are equally significant for ORF57 dimerization as the 'arm' region from one monomer docks on the globular surface of a neighboring monomer in an antiparallel fashion, whereas the C-terminus end (aa residues 445–454) is locked into the globular domain of the same monomer [76].

Interestingly, both the siRNA_1 and siRNA_4 target regions are located in the CDS of N and C-terminal domains of ORF57, respectively. Three NLSs are found in the NTD and the siRNA_1 target sequence (aa residues 98-104) overlaps with the NLS1 (aa residues 101-107). A previous study demonstrated that disruption of NLSs inhibits the nuclear translocation process of ORF57 [77]. Our designed siRNA_1 targets the NLS1 coding region and could potentially disturb the translation process of the NTD CDS of ORF57. Besides, ORF57-CTD facilitates the dimerization, stability, and function of the protein. Functional studies also revealed the dissociation of ORF57 dimer upon deletion/point mutation of either one of three structural elements as mentioned earlier. siRNA_4 designed in this study falls within the CDS of the globular interface (aa residues 234-239) of ORF57-CTD. This globular structure maintains the electrostatic interaction with the interface residues of CTD of adjacent monomer to stabilize the dimer [76]. Thus, our proposed siRNA_4 could also inhibit this function by degrading the target mRNA.

It is well established that siRNA-induced post-transcriptional gene silencing starts with the assembly of the RISC [78,79]. The mRNA molecule is cut exactly by cleaving the phosphodiester bond between the target nucleotides which are paired to siRNA residues [80]. The functional ORF57 protein is a dimer of two identical subunits. This indicates the subunits of ORF57 protein are translated from the transcripts of a single copy gene. Hence, the disruption of the ORF57 transcripts by siRNAs could stop the translation process completely. Therefore, we believe that either siRNA_1 or siRNA_4 could be sufficient enough to suppress the *ORF57* gene expression completely or at least partially.

There is currently no specific treatment for KSHV-related diseases. The treatment of choice for KSHV patients predominantly depends on various parameters, like the tumor location, a variant of KS, rate of progression, distribution of the lesions, the severity of the symptoms, and immune competence [81]. Although medicines such as rituximab, acyclovir, and others are currently used to treat KSHV-related complications, these are not specific therapies [82,83]. Besides that, severe side effects like kidney toxicity, neutropenia, and neurotoxicity have made the treatments more challeng-

ing [84]. That's why additional studies are required to explore new drugs for KSHV associated diseases. In this circumstance siRNA-based therapy might be a viable alternative as the inhibitory effect of siRNA on different herpes virus replication has already been reported in several studies [85-87]. This research is such an effort to accelerate the discovery of new treatments for KSHV-related diseases. Two potential siRNAs have been screened in this study through a series of comprehensive filtration steps that will hopefully inhibit the translation of the *ORF57* gene in KSHV. Since our suggested siRNAs meet all of the requirements for an effective siRNA, it can be expected that they'll be able to inhibit the infection against all KSHV strains. But, as this selection method was entirely based on computational prediction, proper in vitro and in vivo validation is albeit necessary.

ORCID

Anisur Rahman: <https://orcid.org/0000-0001-8279-9810>

Shipan Das Gupta: <https://orcid.org/0000-0002-3236-519X>

Md. Anisur Rahman: <https://orcid.org/0000-0001-5344-2767>

Saheda Tamanna: <https://orcid.org/0000-0001-8720-2390>

Authors' Contribution

Conceptualization: MAR. Data curation: MAR, SDG. Formal analysis: AR. Methodology: AR, SDG, MAR, ST. Writing - original draft: AR, ST. Writing - review & editing: MAR, SDG.

Conflicts of Interest

No potential conflict of interest relevant to this article was reported.

Supplementary Materials

Supplementary data can be found with this article online at <http://www.genominfo.org>.

References

1. Kagu MB, Nggada HA, Garandawa HI, Askira BH, Durosinmi MA. AIDS-associated Kaposi's sarcoma in Northeastern Nigeria. *Singapore Med J* 2006;47:1069-1074.
2. Tamburro KM, Yang D, Poisson J, Fedoriw Y, Roy D, Lucas A, et al. Vironome of Kaposi sarcoma associated herpesvirus-inflammatory cytokine syndrome in an AIDS patient reveals co-infec-

- tion of human herpesvirus 8 and human herpesvirus 6A. *Virology* 2012;433:220-225.
3. Giffin L, Damania B. KSHV: pathways to tumorigenesis and persistent infection. *Adv Virus Res* 2014;88:111-159.
 4. Goncalves PH, Uldrick TS, Yarchoan R. HIV-associated Kaposi sarcoma and related diseases. *AIDS* 2017;31:1903-1916.
 5. Moore PS, Chang Y. Detection of herpesvirus-like DNA sequences in Kaposi's sarcoma in patients with and those without HIV infection. *N Engl J Med* 1995;332:1181-1185.
 6. Gao SJ, Kingsley L, Li M, Zheng W, Parravicini C, Ziegler J, et al. KSHV antibodies among Americans, Italians and Ugandans with and without Kaposi's sarcoma. *Nat Med* 1996;2:925-928.
 7. Cesarman E, Damania B, Krown SE, Martin J, Bower M, Whitby D. Kaposi sarcoma. *Nat Rev Dis Primers* 2019;5:9.
 8. Cesarman E, Chang Y, Moore PS, Said JW, Knowles DM. Kaposi's sarcoma-associated herpesvirus-like DNA sequences in AIDS-related body-cavity-based lymphomas. *N Engl J Med* 1995;332:1186-1191.
 9. Soulier J, Grollet L, Oksenhendler E, Cacoub P, Cazals-Hatem D, Babinet P, et al. Kaposi's sarcoma-associated herpesvirus-like DNA sequences in multicentric Castleman's disease. *Blood* 1995;86:1276-1280.
 10. Uldrick TS, Wang V, O'Mahony D, Aleman K, Wyvill KM, Marshall V, et al. An interleukin-6-related systemic inflammatory syndrome in patients co-infected with Kaposi sarcoma-associated herpesvirus and HIV but without Multicentric Castleman disease. *Clin Infect Dis* 2010;51:350-358.
 11. Beral V, Peterman TA, Berkelman RL, Jaffe HW. Kaposi's sarcoma among persons with AIDS: a sexually transmitted infection? *Lancet* 1990;335:123-128.
 12. Roshan R, Labo N, Trivett M, Miley W, Marshall V, Coren L, et al. T-cell responses to KSHV infection: a systematic approach. *Oncotarget* 2017;8:109402-109416.
 13. Agaba PA, Sule HM, Ojoh RO, Hassan Z, Apena L, Mu'azu MA, et al. Presentation and survival of patients with AIDS-related Kaposi's sarcoma in Jos, Nigeria. *Int J STD AIDS* 2009;20:410-413.
 14. Wabinga HR, Parkin DM, Wabwire-Mangen F, Mugerwa JW. Cancer in Kampala, Uganda, in 1989-91: changes in incidence in the era of AIDS. *Int J Cancer* 1993;54:26-36.
 15. Parkin DM, Sitas F, Chirenje M, Stein L, Abratt R, Wabinga H. Part I: Cancer in Indigenous Africans: burden, distribution, and trends. *Lancet Oncol* 2008;9:683-692.
 16. Renne R, Lagunoff M, Zhong W, Ganem D. The size and conformation of Kaposi's sarcoma-associated herpesvirus (human herpesvirus 8) DNA in infected cells and virions. *J Virol* 1996;70:8151-8154.
 17. Russo JJ, Bohenzky RA, Chien MC, Chen J, Yan M, Maddalena D, et al. Nucleotide sequence of the Kaposi sarcoma-associated herpesvirus (HHV8). *Proc Natl Acad Sci U S A* 1996;93:14862-14867.
 18. Neipel F, Albrecht JC, Fleckenstein B. Cell-homologous genes in the Kaposi's sarcoma-associated rhadinovirus human herpesvirus 8: determinants of its pathogenicity? *J Virol* 1997;71:4187-4192.
 19. Gradoville L, Gerlach J, Grogan E, Shedd D, Nikiforow S, Metroka C, et al. Kaposi's sarcoma-associated herpesvirus open reading frame 50/Rta protein activates the entire viral lytic cycle in the HH-B2 primary effusion lymphoma cell line. *J Virol* 2000;74:6207-6212.
 20. Majerciak V, Zheng ZM. KSHV ORF57, a protein of many faces. *Viruses* 2015;7:604-633.
 21. Nekorchuk M, Han Z, Hsieh TT, Swaminathan S. Kaposi's sarcoma-associated herpesvirus ORF57 protein enhances mRNA accumulation independently of effects on nuclear RNA export. *J Virol* 2007;81:9990-9998.
 22. Majerciak V, Pripuzova N, McCoy JP, Gao SJ, Zheng ZM. Targeted disruption of Kaposi's sarcoma-associated herpesvirus ORF57 in the viral genome is detrimental for the expression of ORF59, K8alpha, and K8.1 and the production of infectious virus. *J Virol* 2007;81:1062-1071.
 23. Lam JK, Chow MY, Zhang Y, Leung SW. siRNA versus miRNA as therapeutics for gene silencing. *Mol Ther Nucleic Acids* 2015;4:e252.
 24. Leconet W, Petit P, Peraldi-Roux S, Bresson D. Nonviral delivery of small interfering RNA into pancreas-associated immune cells prevents autoimmune diabetes. *Mol Ther* 2012;20:2315-2325.
 25. Karim ME, Tha KK, Othman I, Borhan Uddin M, Chowdhury EH. Therapeutic potency of nanoformulations of siRNAs and shRNAs in animal models of cancers. *Pharmaceutics* 2018;10:65.
 26. Qureshi A, Tantray VG, Kirmani AR, Ahangar AG. A review on current status of antiviral siRNA. *Rev Med Virol* 2018;28:e1976.
 27. Alterman JF, Godinho B, Hassler MR, Ferguson CM, Echeverria D, Sapp E, et al. A divalent siRNA chemical scaffold for potent and sustained modulation of gene expression throughout the central nervous system. *Nat Biotechnol* 2019;37:884-894.
 28. Mehta A, Michler T, Merkel OM. siRNA Therapeutics against respiratory viral infections-what have we learned for potential COVID-19 therapies? *Adv Healthc Mater* 2021;10:e2001650.
 29. Sayers EW,avanaugh M, Clark K, Ostell J, Pruitt KD, Karsch-Mizrachi I. GenBank. *Nucleic Acids Res* 2020;48:D84-D86.
 30. Zhang Z, Schwartz S, Wagner L, Miller W. A greedy algorithm for aligning DNA sequences. *J Comput Biol* 2000;7:203-214.

31. Okonechnikov K, Golosova O, Fursov M, team U. Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics* 2012; 28:1166-1167.
32. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 2004;32:1792-1797.
33. Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ. Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 2009;25:1189-1191.
34. Tafer H. Bioinformatics of siRNA design. (Gorodkin J, Ruzzo WL, eds.). In: *RNA Sequence, Structure, and Function: Computational and Bioinformatic Methods* Totowa: Humana Press, 2014. pp. 477-490.
35. Ichihara M, Murakumo Y, Masuda A, Matsuura T, Asai N, Jijiwa M, et al. Thermodynamic instability of siRNA duplex is a prerequisite for dependable prediction of siRNA activities. *Nucleic Acids Res* 2007;35:e123.
36. Ding Y, Lawrence CE. Statistical prediction of single-stranded regions in RNA secondary structure and application to predicting effective antisense target sites and beyond. *Nucleic Acids Res* 2001;29:1034-1046.
37. Ding Y, Chan CY, Lawrence CE. Sfold web server for statistical folding and rational design of nucleic acids. *Nucleic Acids Res* 2004;32:W135-W141.
38. Reynolds A, Leake D, Boese Q, Scaringe S, Marshall WS, Khvorovva A. Rational siRNA design for RNA interference. *Nat Biotechnol* 2004;22:326-330.
39. Amarzguioui M, Prydz H. An algorithm for selection of functional siRNA sequences. *Biochem Biophys Res Commun* 2004;316:1050-1058.
40. Ui-Tei K, Naito Y, Takahashi F, Haraguchi T, Ohki-Hamazaki H, Juni A, et al. Guidelines for the selection of highly effective siRNA sequences for mammalian and chick RNA interference. *Nucleic Acids Res* 2004;32:936-948.
41. Matveeva O. What parameters to consider and which software tools to use for target selection and molecular design of small interfering RNAs. (Taxman DJ, ed.). In: *siRNA Design: Methods and Protocols* Totowa: Humana Press, 2013. pp. 1-16.
42. Rice P, Longden I, Bleasby A. EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet* 2000;16:276-277.
43. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol* 1990;215:403-410.
44. Pruitt KD, Tatusova T, Maglott DR. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res* 2007;35: D61-D65.
45. Birmingham A, Anderson E, Sullivan K, Reynolds A, Boese Q, Leake D, et al. A protocol for designing siRNAs with high functionality and specificity. *Nat Protoc* 2007;2:2068-2078.
46. Kozomara A, Griffiths-Jones S. miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res* 2014;42:D68-D73.
47. Gruber AR, Lorenz R, Bernhart SH, Neubock R, Hofacker IL. The Vienna RNA websuite. *Nucleic Acids Res* 2008;36: W70-W74.
48. Bernhart SH, Tafer H, Muckstein U, Flamm C, Stadler PF, Hofacker IL. Partition function and base pairing probabilities of RNA heterodimers. *Algorithms Mol Biol* 2006;1:3.
49. Reuter JS, Mathews DH. RNAstructure: software for RNA secondary structure prediction and analysis. *BMC Bioinformatics* 2010;11:129.
50. Zuber J, Mathews DH. Estimating uncertainty in predicted folding free energy changes of RNA secondary structures. *RNA* 2019;25:747-754.
51. Bernhart SH, Hofacker IL, Will S, Gruber AR, Stadler PF. RNAalifold: improved consensus structure prediction for RNA alignments. *BMC Bioinformatics* 2008;9:474.
52. Kerpedjiev P, Hammer S, Hofacker IL. Forna (force-directed RNA): Simple and effective online RNA secondary structure diagrams. *Bioinformatics* 2015;31:3377-3379.
53. Ashraf MI, Leema AA. Identification of micro-RNA seed sequences and other possible conserved motifs: an information theoretic approach. *J RNAi Gene Silencing* 2017;13:570-572.
54. Ishizuka A, Siomi MC, Siomi H. A *Drosophila* fragile X protein interacts with components of RNAi and ribosomal proteins. *Genes Dev* 2002;16:2497-2508.
55. Fakhr E, Zare F, Teimoori-Toolabi L. Precise and efficient siRNA design: a key point in competent gene silencing. *Cancer Gene Ther* 2016;23:73-82.
56. Ui-Tei K, Nishi K, Takahashi T, Nagasawa T. Thermodynamic control of small RNA-mediated gene silencing. *Front Genet* 2012;3:101.
57. Ui-Tei K, Naito Y, Nishi K, Juni A, Saigo K. Thermodynamic stability and Watson-Crick base pairing in the seed duplex are major determinants of the efficiency of the siRNA-based off-target effect. *Nucleic Acids Res* 2008;36:7100-7109.
58. Shatizadeh Malekshahi S, Arefian E, Salimi V, Mokhtari Azad T, Yavarian J. Potential siRNA molecules for nucleoprotein and M2/L overlapping region of respiratory syncytial virus: in silico design. *Jundishapur J Microbiol* 2016;9:e34304.
59. Matveeva OV, Kang Y, Spiridonov AN, Saetrom P, Nemtsov VA, Ogurtsov AY, et al. Optimization of duplex stability and terminal

- asymmetry for shRNA design. *PLoS One* 2010;5:e10180.
60. Naito Y, Yoshimura J, Morishita S, Ui-Tei K. siDirect 2.0: updated software for designing functional siRNA with reduced seed-dependent off-target effect. *BMC Bioinformatics* 2009;10:392.
 61. ElHefnawi M, Kim T, Kamar MA, Min S, Hassan NM, El-Ahwany E, et al. *In silico* design and experimental validation of siRNAs targeting conserved regions of multiple hepatitis C virus genotypes. *PLoS One* 2016;11:e0159211.
 62. Hornung V, Guenther-Biller M, Bourquin C, Ablasser A, Schlee M, Uematsu S, et al. Sequence-specific potent induction of IFN- α by short interfering RNA in plasmacytoid dendritic cells through TLR7. *Nat Med* 2005;11:263-270.
 63. Judge AD, Sood V, Shaw JR, Fang D, McClintock K, MacLachlan I. Sequence-dependent stimulation of the mammalian innate immune response by synthetic siRNA. *Nat Biotechnol* 2005;23:457-462.
 64. Sugiyama T, Gursel M, Takeshita F, Coban C, Conover J, Kaisho T, et al. CpG RNA: identification of novel single-stranded RNA that stimulates human CD14+CD11c+ monocytes. *J Immunol* 2005;174:2273-2279.
 65. Fedorov Y, Anderson EM, Birmingham A, Reynolds A, Karpilow J, Robinson K, et al. Off-target effects by siRNA can induce toxic phenotype. *RNA* 2006;12:1188-1196.
 66. Bartoszewski R, Sikorski AF. Editorial focus: understanding off-target effects as the key to successful RNAi therapy. *Cell Mol Biol Lett* 2019;24:69.
 67. Aleman LM, Doench J, Sharp PA. Comparison of siRNA-induced off-target RNA and protein effects. *RNA* 2007;13:385-395.
 68. Heale BS, Soifer HS, Bowers C, Rossi JJ. siRNA target site secondary structure predictions using local stable substructures. *Nucleic Acids Res* 2005;33:e30.
 69. Ui-Tei K, Naito Y, Zenno S, Nishi K, Yamato K, Takahashi F, et al. Functional dissection of siRNA sequence by systematic DNA substitution: modified siRNA with a DNA seed arm is a powerful tool for mammalian gene silencing with significantly reduced off-target effect. *Nucleic Acids Res* 2008;36:2136-2151.
 70. Vidarsdottir L, Goroshchuk O, Kolosenko I, Palm-Apergi C. Designing siRNA and evaluating its effect on RNA targets using qPCR and western blot. In: *Oligonucleotide-Based Therapies: Methods and Protocols* (Gissberg O, Zain R, Lundin KE, eds.). New York: Springer, 2019. pp. 59-72.
 71. Zuker M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* 2003;31:3406-3415.
 72. Lu ZJ, Mathews DH. Efficient siRNA selection using hybridization thermodynamics. *Nucleic Acids Res* 2008;36:640-647.
 73. Singh S, Gupta SK, Nischal A, Khattri S, Nath R, Pant KK, et al. Design of potential siRNA molecules for hepatitis delta virus gene silencing. *Bioinformatics* 2012;8:749-757.
 74. Patzel V. *In silico* selection of active siRNA. *Drug Discov Today* 2007;12:139-148.
 75. Gredell JA, Berger AK, Walton SP. Impact of target mRNA structure on siRNA silencing efficiency: a large-scale study. *Biotechnol Bioeng* 2008;100:744-755.
 76. Yuan F, Gao ZQ, Majerciak V, Bai L, Hu ML, Lin XX, et al. The crystal structure of KSHV ORF57 reveals dimeric active sites important for protein stability and function. *PLoS Pathog* 2018;14:e1007232.
 77. Majerciak V, Yamanegi K, Nie SH, Zheng ZM. Structural and functional analyses of Kaposi sarcoma-associated herpesvirus ORF57 nuclear localization signals in living cells. *J Biol Chem* 2006;281:28365-28378.
 78. Lee YS, Nakahara K, Pham JW, Kim K, He Z, Sontheimer EJ, et al. Distinct roles for *Drosophila* Dicer-1 and Dicer-2 in the siRNA/miRNA silencing pathways. *Cell* 2004;117:69-81.
 79. Tomari Y, Zamore PD. Perspective: machines for RNAi. *Genes Dev* 2005;19:517-529.
 80. Orban TI, Izaurralde E. Decay of mRNAs targeted by RISC requires XRN1, the Ski complex, and the exosome. *RNA* 2005;11:459-469.
 81. Fatahzadeh M. Kaposi sarcoma: review and medical management update. *Oral Surg Oral Med Oral Pathol Oral Radiol* 2012;113:2-16.
 82. Veraitch O, Bower M, Shackleton D, Stebbing J. Rituximab therapy for HIV-associated multicentric Castleman disease. *HIV Ther* 2010;4:281-284.
 83. Poole CL, James SH. Antiviral therapies for herpesviruses: current agents and new directions. *Clin Ther* 2018;40:1282-1298.
 84. Steinberg I, Kimberlin DW. Acyclovir dosing and acute kidney injury: deviations and direction. *J Pediatr* 2015;166:1341-1344.
 85. Wang K, Yang C, Ye J, Zeng F, Duan Y, Zheng Y, et al. Inhibition activity of herpes virus (HSV) replication by α -TIF siRNA-loaded PLGA-TPGS nanoparticles *in vitro* and *in vivo*. *J Biomed Nanotechnol* 2017;13:717-726.
 86. Gatault P, Jones IKA, Meyer C, Kreklywich C, Alexander T, Smith PP, et al. Rat and human cytomegalovirus ORF116 encodes a virion envelope glycoprotein required for infectivity. *Virology* 2021;557:23-33.
 87. Lu JF, Jin TC, Zhou T, Lu XJ, Chen JP, Chen J. Identification and characterization of a tumor necrosis factor receptor like protein encoded by Cyprinid Herpesvirus 2. *Dev Comp Immunol* 2021;116:103930.