Chapter 11

Computational Approaches for Designing Efficient and Specific siRNAs

Suman Ghosal, Shaoli Das and Jayprokas Chakrabarti

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/50125

1. Introduction

Small RNA mediated RNA interference (RNAi) is a widely adopted mechanism towards immunity in plants and invertebrates. Two types of small RNAs- small interfering RNA (siRNA) and microRNA (miRNA), play key role in RNA interference either through cleaving or through translational repression of the target mRNA by guiding RNA induced silencing complex (RISC) to its target site. siRNA is a small RNA (19-23 nucleotides) which is complementary to part of their target mRNA [1]. siRNAs are very efficient in target gene knockdown that makes synthetic siRNA's perfect choice for use in experiments for silencing genes to examine their function. In addition, siRNA have good potential in drug development for therapeutic purpose [2]. Exogenous synthetic siRNAs are designed to target a part of the coding region in the target mRNA [3]. But, as evident from experiments, all siRNAs are not equally efficient in target gene silencing. The potency of siRNAs is largely dependent upon the selection of the region it targets. A displacement of 5-6 nucleotide position hugely alters the efficiency of siRNA. The reason behind this alteration of efficiency is the alteration in local sequence and structural features of the target region. These variations in sequence and structural features correlate with target accessibility, RISC loading or stimulation of immune response. A lot of study has been done in the field of rational siRNA designing to find appropriate parameters that facilitate designing of effective siRNAs [4]. Several commercial suppliers and non-profit educational institutes contribute to the research for searching appropriate siRNA selection parameters for improving potency of designed siRNAs. Progresses have been made in targeting success rate compared to the early days- from as low as 0-10% targeting success rate, today siRNAs have reached average 50% targeting success rate. Still there are many scopes to improve siRNA designing. For efficient designing, the siRNA selection parameters must be arranged and weighted in such a way that ensures optimal result while selecting the siRNA target site. There are some previously suggested guidelines about parameter weight assignment like rational or

weighted methods. These guidelines were made in the early days of siRNA selection research with small number of data, when experimentally knockdown validated siRNA dataset was scarce. But with increasing amount of knockdown validated siRNA datasets, new parameter optimization methods must come out to ensure selection of potent siRNAs in a bigger scenario. Today, high throughput siRNA screening experiments have become a common technique to investigate thousands of gene functions at a time to study some specific pathway. These experiments use siRNA libraries targeting transcripts in a genome wide range. But the success of these experiments relies on the knockdown success rate of the siRNAs in the library. After two decades of research, still many of these large scale screening experiments fail because of a large number of non-functional siRNAs present in the libraries used in these experiments. So, still there are needs for improvement in the siRNA selection algorithm and efficient parameter optimization.

Another main challenge in computational siRNA designing lies in the specificity issue of siRNAs. Exogenous siRNAs often induce off-target effects that arise from near perfect or imperfect sequence complementarity with other mRNAs resulting in false positive phenotype during RNAi-based study of gene functions. The type of gene regulation resulting from imperfect sequence complementarity resembles target regulation by endogenous miRNAs- hence this type of off-targeting is called miRNA-like off-targeting. Minimization of miRNA-like off-targets involves choosing a siRNA with a seed region sequence that has fewer targets in the 3′ UTR of other mRNAs. The siRNA seed region is a 6 or 7 nucleotide sequence (the 2-8th nucleotide position) from 5′ end of siRNA guide strand and finding complementary sequence with this 6 or 7 nucleotide seed in the 3′ UTRs of unintended transcripts results in a huge number of potential targets. A majority of these predicted targets are not practically relevant, as in practice, a large number of these predicted off-targets may not be silenced at all. Like miRNA targets, siRNA targets are also dependent on target accessibility and other sequence features around the target site. So, a more rational approach is needed when predicting the siRNA off-targets resulting from partial sequence complementarity. This chapter focuses on current siRNA designing parameters and the approaches towards minimizing the off-target effects as provided by the in-silico siRNA designing solutions and then discuss a bit about a customized off-target reducing algorithm that will be useful to avoid particular genes from being off-targeted.

2. Guidelines for selecting potent siRNAs

As discussed earlier, potency of siRNAs varies greatly with selection of target region. The parameters responsible for effectiveness of siRNAs being target accessibility, uniqueness of target region, absence of SNP sites etc. Other parameters worth considering while selecting the target site is the consideration of alternatively spliced isoforms of the target gene [5].

A parameter which greatly influences siRNA potency is RISC loading of the intended antisense strand [6]. Only one strand of the siRNA duplex enters into RISC and guides the complex to the target mRNA for its silencing. So, the choice of strand that enters into RISC

complex plays the vital role in the whole silencing process. Generally RISC complex chooses the siRNA strand which has weaker 5′ end binding energy. So this thermodynamic property should be ensured while selecting siRNAs. Also there are a number of sequence compositions that are said to be important for recognition of the intended strand by RISC and some parameters for enhancing the stability of siRNA duplex are also worth considering while selecting siRNAs. These rules can be divided into three categories 1) Selection of target region, 2) Structural and Thermodynamic Consideration and 3) Sequence Characteristics.

2.1. Selection of target region

Selection of the target region for siRNA should satisfy some constraints to ensure effective silencing. Target site should be analyzed for its position within the mRNA, presence of polymorphic site or other considerations like homology with other mRNAs or its alternatively spliced isoforms.

2.1.1. Selection of target starting site

Generally siRNAs are targeted to a part of the coding region of an mRNA. The target site should be deep inside the open reading frame (ORF) of the target mRNA for efficient silencing. Generally it is advised to start from 50-100 nucleotides downstream of the AUG start codon. Besides targeting the coding region, some siRNAs may also be designed to target the 3′ un-translated region (UTR) of an mRNA. This type of targeting is especially done in case of experiments conducted for restoration of original phenotype. In such cases it is advised to start targeting regions at least 15-20 nucleotides downstream of the stop codon.

2.1.2. Consideration for alternative splicing

In eukaryotic organisms frequent alternative splicing results in diversification of mRNAs. To account for the alternative splicing, it is necessary to evaluate a common target region where siRNAs can be designed to knockdown a specific mRNA isoform or multiple mRNA isoforms from a target gene.

2.1.3. Absence of homology with other mRNAs

Cellular mRNAs with 15/16 or more consecutive base match with the siRNA, are likely to be silenced and degraded by the siRNA. So target sites with 15 or more consecutive base homology with any other mRNA should be avoided.

2.1.4. Avoiding target sites having polymorphic locus

Target regions having single nucleotide polymorphic (SNP) sites should be avoided.

2.2. Structural and thermodynamic consideration

siRNA potency largely depends upon structural constraints of the target region. Heavily structured sites are less likely to be bound by siRNAs as these sites are not accessible by siRNAs. The relative binding energy of the 5′ and 3′ ends of the siRNA with the target site play a vital role in the choice of strand to be incorporated into RISC complex and thus is one of the most important parameter to be considered during siRNA design.

2.2.1. Presence of Secondary structure

It has been suggested that presence of local secondary structures (stem loops) in the target site restricts its accessibility to RISC and hence reduces the efficiency of the siRNA. So it is necessary to filter out those potential inaccessible target sites with strong secondary structures. The prediction of local secondary structure can be made by numerous RNA secondary structure prediction tools or packages like Mfold [7] or Vienna RNA package [8] - that mainly predict minimum free energy secondary structure of a RNA sequence.

2.2.2. Thermodynamic property for efficient RISC loading

In a siRNA duplex, antisense strand with relatively low energy in 5′ end is favourable for its loading into RISC complex. So, there should be difference in binding energy between the 5′ end of the sense and antisense strand.

2.3. Sequence characteristics

Years of research for finding appropriate designing parameters identified some sequence parameters enriched within efficient siRNAs. These sequence characteristics often contribute to efficient RISC loading or siRNA sequence specificity or stability issues.

2.3.1. Position specific nucleotide composition

Sequence analysis of effective siRNAs revealed many position specific nucleotide compositions for enhancing potency of the siRNA. Some of these preferences are listed below in table 1-

2.3.2. Sequence feature for efficient RISC entry

siRNA guide strands with low energy at 5′ end are favored for entering the RISC complex. So, presence of at least three (A/U)s in the seven nucleotides at the 3' end of the sense strand is preferable.

2.3.3. siRNA duplex stability

Target sites with low GC content (generally less than 55%) has a greater potential for being functional siRNA site, as too high GC content can impede the loading of siRNAs into RISC

complex. Also, too low GC content is not favourable because too low GC content can destabilize the siRNA duplex and reduce their affinity to target mRNA binding. Analysis of effective siRNAs showed G/C content between 35% and 60% is most favorable.

Position specific nucleotides
Presence of A base at position 19 of the sense strand.
Presence of U base at position 10 of the sense strand.
Presence of A base at position 3 of the sense strand.
A base other than G or C at position 19 of the sense strand.
A base other than G at position 13 of the sense strand.
Presence of A base at the 2nd nucleotide position of the sense strand.
Presence of C base at the 4th nucleotide position of the sense strand.
Absence of C base at the 6th nucleotide position of the sense strand.
Absence of U base at the 7th nucleotide position of the sense strand.
Presence of C base at the 9th nucleotide position of the sense strand.
Presence of A base at the 17th nucleotide position of the sense strand.
Absence of C base at the 18th nucleotide position of the sense strand.
No occurrences of four or more identical nucleotides in a row.
No occurrences of G/C stretch of length 7 or longer.

Table 1. Position specific nucleotide composition prefered in functional siRNAs

3. Choice of appropriate parameters

All the parameters discussed above are not equally important for selection of efficient siRNAs. By far, many research groups have conducted studies for evaluation of effective parameter sets for siRNA selection. Gong et al. studied 276 known siRNA selection parameters on a sufficiently large set of 3277 experimentally validated siRNAs targeting 1518 genes to identify common parameters that effectively distinguishes functional siRNAs from non functional ones [9]. They were able to identify 34 features associated with improved siRNA efficacy among which 27 features were associated with greater than 70% efficacy. They examined combination of siRNA features to find their cooperative effects on potent siRNA selection and used a disjunctive rule merging (DRM) algorithm to generate a bunch of non-redundant rules set to efficiently predict functional siRNAs and lower the false positive predictions. Table 2 list 17 features set associated with greater than 90% efficacy and used for optimal features combination.

siRNA selection parameter
F1 2nd nucleotide = A
F2 4th nucleotide = C
F3 6th nucleotide ≠ C
F4 7th nucleotide ≠ U
F5 9th nucleotide = C
F6 17th nucleotide = A
F7 18th nucleotide ≠ C
F8 19th nucleotide = (A/U)
F9 At least three (A/U)s in the seven nucleotides at the 3' end
F10 No occurrences of four or more identical nucleotides in a row
F11 No occurrences of G/C stretches of length 7 or longer
F12 G/C content is between 35 and 60%
F13 Tm is between 20 and 60°C
F14 Binding energy of N16–N19 > -9 KCal/Mol
F15 Binding energy of N16–N19 – binding energy of N1–N4 is between 0 and 1 KCal/Mol
F16 Local folding potential (mean) ≥ -22.72 KCal/Mol
F17 Target site is on CDS

Table 2. Feature sets predicted to be associated with greater siRNA efficacy as described by Gong et al.

4. Prediction of siRNA potency

Computational prediction of siRNA potency relies on assessment of appropriate designing parameters combined with their optimal weight distribution. Since the early era of siRNA designing researches, many studies are made for finding optimal weights for siRNA selection parameters. Raynolds et al. proposed a method for rational siRNA designing with appropriate parameter weights, by empirical study of 180 experimentally validated efficient siRNAs [10]. The designed siRNAs were given a score based on weighted summation of these parameters and a score threshold was used to identify efficient siRNAs. Table 3 lists the parameters used in the study with their weight distributions.

siRNA selection parameter	Parameter weight
GC content 30% to 52%	Satisfying this criteria earns 1 point
Occurence of 3 or more A/U base pair at position 15-19 of sense strand	Each A/U base pair in this region earns 1 point
Low internal stability at target site (melting temperature T_m>-20^{o}c)	Satisfying this criteria earns 1 point
Presence of A at position 19 of the sense strand	Satisfying this criteria earns 1 point
Presence of A at position 3 of the sense strand	Satisfying this criteria earns 1 point
Presence of U at position 10 of the sense strand	Satisfying this criteria earns 1 point
Absense of G or C at position 19 of the sense strand	Failure to satisfy this criteria decreases 1 point
Absense of G at position 13 of the sense strand	Failure to satisfy this criteria decreases 1 point
Threshold for efficient siRNAs	score>=6

Table 3. Parameters used in Raynold's algorithm with their weights

Since then many siRNA designing algorithm worked on different weight distribution schemes for improved prediction of siRNA potency and some even used machine learning algorithms.

4.1. Use of machine learning algorithms for classification of functional siRNAs

After many years of research about the guidelines for selection of effective siRNAs, we are a few steps ahead in the process of improving the targeting success rate. But for better targeting success, the siRNA selection parameters provided in various guidelines needs to be optimized. Still there is no reliable guideline for optimization of weights of siRNA selection parameter. Machine learning algorithms like Support vector machine or artificial neural network can serve excellent purpose, when trained with sufficient volume of biologically validated siRNA data sets [11]. Some online siRNA designing tools (like BioPredsi and Genescript siRNA target finder) use machine learning algorithms for classification of effective siRNAs from non-effective ones.

4.1.1. Use of artificial neural network for siRNA classification

Artificial neural networks (ANNs), as they aim to mimic the working of biological networks through a connectionist approach to computation, provide a powerful method of identifying highly complex traits in data sets. ANNs are generally very efficient classifiers in case of complex patterns in the given data set as they can adaptively change their weighting parameters during the learning process. ANNs have been broadly applied in the biological sciences. The prediction quality and generalization capabilities of an ANN of fixed size depend on a sufficiently large training set of directly comparable data points.

Biopredsi siRNA designing algorithm from Novartis lab used Stuttgart Neural Net Simulator to train algorithms on a data set of 2182 randomly selected siRNAs targeted to 34 mRNA species [11]. It reliably predicted activity of 249 siRNAs of an independent test set (Pearson coefficient $r = 0.66$) and siRNAs targeting endogenous genes at mRNA and protein levels.

4.1.2. Support Vector machine based classification

Support Vector Machine (SVM) is a non-probabilistic binary linear classifier. Given a set of training examples, each marked as belonging to one of two categories, an SVM training algorithm builds a model that assigns new examples one of the two categories. An SVM is called the maximum margin classifier that optimizes the margin between the example points belonging to two classes so that their gap is maximized.

A newly developed siRNA designing tool enables improved selection of potent siRNAs by application of a Support Vector machine based optimization of a set of eight siRNA selection parameters. The support vector machine is trained with the feature set of 200 highly efficient and 200 poorly efficient siRNA candidates, collected from siRecords, a database of validated siRNAs [12]. The support vector machine is trained using a Gaussian kernel and Sequential Minimal Optimization (SMO) algorithm [13]. It has been tested with huge number of experimentally validated data samples from four different sources and gave sufficiently good result.

5. Experimentally validated siRNA datasets

The effectiveness of the siRNA designing rules should be tested on biologically validated siRNA datasets. On the early days of RNAi research, these biologically validated datasets were scarce. But now, with emerging high throughput technologies, large amount of validated siRNA data is being generated. Some databases are created by manual curation of literature describing validation of siRNA mediated silencing. siRecords [12] is one such database where siRNAs are marked with their respective silencing efficacy (low, medium, high and very high). MIT siRNA database [14] consists of siRNAs designed by Qiagen with validated knockdown efficiency and marked with mRNA knockdown level.

6. Improving specificity of siRNAs

The specificity of siRNAs is a big issue in siRNA mediated gene silencing experiments. Exogenous siRNAs are reported to have off-target effects arising from either silencing unintended targets or toxic effects arising from their recognition by innate immune system [15].

The recognition of siRNAs by innate immune system can result from interferon response triggered by double stranded siRNA duplex or sequence dependent stimulation of toll like receptors. Avoiding some sequence motifs and a constraint related to the siRNA duplex length can effectively reduce immune response stimulation [16].

siRNAs silence unintended transcripts in mainly two ways: transcripts with near perfect complementarity are cleaved while transcripts with imperfect complementarity are translationally repressed. mRNAs other than intended targets which exhibit near perfect sequence complementarity with the siRNA are likely to be degraded by the siRNA. This kind of off-targets can be avoided by choosing targets sites that do not have a large number of consecutive base homology with any other mRNA.

siRNAs down regulate a set of transcripts with 3′ UTR complementarity to the 5′ portion of the corresponding siRNA guide strand. These 5′ ends of the guide strand resemble the seed region of endogenous microRNA and are responsible for target recognition. Such off-targets are regulated by translational repression like miRNA target regulation. This kind of off-target cannot be fully avoided but can be reduced by computational design.

6.1. Stimulation of innate immune response

siRNAs can induce potential unwanted effects by activating innate immune system. Exogenous siRNAs are prone to be recognized by Toll-like receptors (TLRs), mainly TLR7, TLR8 and TLR9. TLR7 and TLR8 recognize synthetic siRNAs in a sequence dependent manner [16]. There seems to be preferential recognition of GU-rich sequences. AU rich sequences can also be immune stimulatory. Selecting siRNA sequences lacking GU rich regions can provide siRNAs with low immune stimulatory activity. Also presence of the motif "GUCCUUCAA" the 4-base motif "UGGC" in the siRNA is known to be immune stimulatory [17]. So, this motif should be avoided in the time of designing of siRNAs. The length of the siRNA is also an important factor for stimulation of immune response- the minimum length of siRNA to be recognized by innate immune system is in the range of 19 nucleotides.

6.2. Near perfect complementarity with other mRNAs

mRNAs other than intended targets which exhibit near perfect sequence complementarity with the siRNA are likely to be degraded by the siRNA. This kind of off-targets can be avoided by choosing targets sites that do not have many consecutive base homologies with any other mRNA. Actually siRNAs can potentially silence transcripts with more than 11 base complementarity including base matches corresponding its 9^{th}-11^{th} nucleotides. But as finding unique 11 base target site is impossible, the siRNA designing algorithms try to find unique target sites that do not have 15 or more consecutive base homology with other transcripts.

6.3. miRNA-like off-target effect

siRNAs down regulate a set of transcripts with 3′ UTR complementarity to the 5′ portion of the corresponding siRNA guide strand. These 5′ ends of the guide strand resemble the seed region of endogenous miRNA which is responsible for target recognition. Such off-targets are regulated by translational repression like miRNA target regulation. That is why this

kind of off-target effect is called miRNA-like off-target effect [18, 19]. This kind of off-target cannot be fully avoided but can be reduced by computational design. Consideration for minimization of such off-target effect involves imposing a threshold for number of off-target genes. All of the present day online siRNA designing techniques consider only the quantitative approach for minimizing miRNA-like off-target effect by restricting the number of off-targets [20]. To go beyond mere quantitative approach and look for the functional correlation between the on-target and the genes off-targeted by the siRNA will certainly prove to be beneficial in minimizing miRNA-like off-target effect. A newly developed siRNA designing tool is aimed for such off-target minimization considering functional correlation of the off-target and the direct target (explained in section 9).

7. Prediction of siRNA off-targets

Prediction of miRNA-like off-targets involves finding the seed complementarity of the siRNA with the 3'UTR of a non-target mRNA. But considering only seed region complementarity identifies a large number of off-targets that could not be actually targeted. So, a more rational approach is needed for prediction of siRNA off-targets that needs understanding the miRNA target recognition procedure. Parameters like local AU content near the seed region or accessibility of the target site within the 3'UTR of the predicted off-target play roles in siRNA off-target detection also [21]. To have a more reliable prediction of off-targets, some siRNA designing solutions consider the stability factor of the duplex formed by siRNA guide strand seed region and mRNA 3' UTR target as the off-targets forming duplex with lower stability are less likely to be actually silenced by the siRNA [22]. These tools examine thermodynamic property of the siRNA seed-mRNA duplex and carefully choose siRNAs with seed sequences that are predicted to form less thermodynamically stable duplex with the target mRNA. Some siRNA designing solutions consider conservation of such target regions among closely related species to determine the candidate mRNAs most likely to be silenced by miRNA-like mechanism [23]. Table 4 gives the mechanism adopted by different online siRNA designing solutions for minimizing off-targets.

In a detailed investigation of all possible seed sequence and their frequency of complimenting the 3' UTRs of human mRNAs, it is shown that the seeds can be classified into low, medium and high frequency classes according to the number 3' UTR sites targeted by them [24]. The low frequency seeds have targets around 350, while the medium and high frequency seeds have targets around 2500 and more than 4800 respectively. So, it is obvious that siRNAs having a seed region that falls into the low frequency group will have fewer off-targets. But then in many cases, presence of such seed sequences can decrease the potency of the siRNA because they often contain stretches of identical nucleotides or other features unfavorable for a potent siRNA. So, there is a tradeoff between potency and specificity of a siRNA which have to be dealt with in their computational designing. Seed complement frequency (SCF) is the frequency of the complement of the hexamer/heptamer seed region within 3' UTR of an mRNA [24]. It is a major parameter which greatly enhances specificity of siRNA off-target prediction, but at the cost of decreasing sensitivity. Some microRNA target site features as uncovered by combining computational and experimental approaches, also apply to the

siRNA off-target prediction problem. A study reported potential silencing of transcripts having consecutive 11 or more bases complementarity with miRNA or siRNAs including siRNA bases 9-12. These transcripts are more likely to be cleaved by the siRNA. Some other off-target prediction parameters include secondary structure analysis for target accessibility prediction and A/U base richness near target site [25]. For reliable off-target prediction, an optimized combination of all the above mentioned parameters is needed.

siRNA designing tool	**Approaches towards minimization of miRNA-like off-target effects**
BIOPREDsi [http://www.biopredsi.org/].	siRNA seed region match in combination with 10 or more bases of additional homology to unintended target genes is used for pediction of off-targets. siRNAs predicted to have a large number of off-targets are rejected.
siRNA Target Finder [https://www.genscript.com/ssl-bin/app/rnai].	Depending on their proprietary off-target prediction database, a threshold of seed match for a specific species is applied which is computed by a probability model which has general biological significance. siRNAs predicted to have a large number of off-targets are rejected.
siDESIGN Center [http://www.dharmacon.com/sidesign].	Chooses siRNAs that has a seed sequence with lower seed compliment frequency.
siDirect [http://sidirect2.rnai.jp].	siDirect selects siRNAs with lower melting temperature (Tm value) at the seed region, which contains 7 nucleotides at positions 2-8 from 5′ end of the guide strand as the capability of siRNA to induce this seed-dependent off-target effect is highly correlated with the thermodynamic stability of the duplex formed between the seed region of the siRNA guide strand and its target mRNA.
siDRM [http://sidrm.biolead.org/]	For each candidate siRNA, *siDRM* checks and reports if its seed region (position 2–8) has full homology to the 3′ UTR region of another transcript, and this homology region is followed by four consecutive mismatches. siRNAs predicted to have a large number of off-targets are rejected.
Whitehead WI siRNA Selection Program [http://jura.wi.mit.edu/bioc/siRNAext/].	Predicts off-targets based on the seed region complementarity as well as conservation of the target site among related species (human, mouse, rat, dog and chicken). siRNAs predicted to have a large number of off-targets are rejected.

Table 4. Off-target minimization techniques of different siRNA designing tools.

8. Discussion about selected siRNA designing tools

Several siRNA sequence selection algorithms have been developed in the past decade that relies on intrinsic sequence, stability and target accessibility features of functional siRNAs. Different siRNA selection algorithms follow different set of rules derived from some well-known siRNA design parameters as discussed above. In general, these algorithms rely on features like- low GC content, absence of siRNA self-alignment, absence of internal repeat, thermodynamic conditions favouring efficient RISC entry, absence of homology to other mRNAs and some position specific nucleotide compositions. Few of them also consider silencing of alternatively spliced isoforms of the given gene. Different algorithms use different techniques for combination of parameters and their weight distribution- ranging from empirical observation to sophisticated machine learning. In spite of a large number of online siRNA design solutions, few of them consider miRNA-like off-targeting potential of synthetic siRNAs. Consideration for minimization of such off-target effect involves imposing a threshold for number of off-target genes. Table 5 lists some of the online siRNA designing solutions with their designing parameters.

siRNA selection tool	**Thermodynamic deference between sense and antisense strand 5′ end**	**Target accessibility**	**Alternative splicing**	**Off-target(near complementary)**	**miRNA-like off-target**
Ambion siRNA Target Finder [http://www.ambion.com/techlib/misc/siRNA_finder.html].	Considered	Not Considered	Not Considered	Considered	Not Considered
AsiDesigner [http://sysbio.kribb.re.kr:8080/AsiDesigner/menuDesigner.jsf]	Considered	Considered	Considered	Considered	Not considered
BIOPREDsi [http://www.biopredsi.org/].	Considered	Not Considered	Not Considered	Considered	Considered
siRNA Target Finder [https://www.genscript.com/ssl-bin/app/rnai].	Considered	Considered	Not Considered	Considered	Considered
BLOCK-iT RNAi Designer [https://rnaidesigner.invitrogen.com/rnaiexpress].	Considered	Considered	Not Considered	Considered	Not Considered
IDT RNAi Design [http://www.idtdna.com/Scitools/Applications/RNAi/RNAi.aspx].	Considered	Considered	Not Considered	Considered	Not Considered
MicroSynth siRNA design [http://www.microsynth.ch/499.0.html].	Considered	Not Considered	Not Considered	Considered	Not Considered

siRNA selection tool	Thermodynamic deference between sense and antisense strand 5′ end	Target accessibility	Alternative splicing	Off-target(near complementary)	miRNA-like off-target
Promega siRNA Target Designer [http://www.promega.com/siRNADesigner/program/].	Considered	Considered	Not Considered	Considered	Not Considered
siDESIGN Center [http://www.dharmacon.com/sidesign].	Considered	Not Considered	Considered	Considered	Considered
siDRM [http://sidrm.biolead.org/]	Considered	Considered	Not Considered	Considered	Considered
siDirect [http://sidirect2.rnai.jp]	Considered	Considered	Not Considered	Considered	Considered
Imgenex siRNA tool [http://www.imgenex.com/sirna_tool.php].	Considered	Considered	Not Considered	Considered	Not Considered
siSearch [http://sonnhammer.cgb.ki.se/siSearch/siSearch_1.7.html].	Considered	Considered	Not Considered	Considered	Not Considered
Whitehead WI siRNA Selection Program [http://jura.wi.mit.edu/bioc/siRNAext/].	Considered	Considered	Not Considered	Considered	Considered

Table 5. A comparison of siRNA designing parameters in different siRNA designing tools

9. Importance of functional off-target filtering

Considering only quantity of the off-targets and not the functions of individual off-targets can lead to inefficient handling of the miRNA-like off-target issue. Often in siRNA screening experiments, it has been reported that the desired output is affected because of silencing of unintended off-targets those sometimes are themselves member of the upstream pathway components of the direct target gene [26]. In such cases it can be useful to avoid specifically some off-targets that can cause more harmful or undesirable effects. It should be considered that during silencing process siRNA should not silence any mRNA from the same pathway the target mRNA is part of. In case of investigation of a gene function, if any gene from the same pathway is silenced rather than target gene then it will be difficult to investigate the actual phenotype of silencing the gene under investigation. For e.g. in a siRNA screening experiment designed for novel members of the transforming growth factor (TGF)-b pathway in a human keratinocyte cell line, dominant off-target effect was observed due to unintended silencing of two known upstream pathway components, the TGF-b receptors 1 and 2 (TGFBR1 and TGFBR2). Such off-target silencing activity poses threats of confusing and misleading results. Also the siRNAs suggested by the online siRNA selection tools often are predicted to have off-targets that belong to the same pathway or somehow related to the direct target.

Das et al reported designing of a siRNA designing tool using a simple approach towards minimizing miRNA-like off-target effect through user feedback. Here, the user can actually choose from the list of potential off-target genes, the off-targets he/she wants to filter out, by considering the effect of silencing of those off-target genes. This tool statistically evaluates present day siRNA design rules such as- low GC content, absence of long stretches of identical nucleotides, thermodynamic conditions favouring efficient RISC entry, absence of homology to other mRNAs, absence of immune stimulatory motifs in the RISC entering strand of the siRNA duplex and some position specific nucleotide compositions, in a database of validated siRNAs [12] used in experiments to examine the threshold parameters. A support vector machine, trained with the optimal features set, is used for classifying potential and effective siRNAs. Moreover, with other parameters, it predicts the potential miRNA-like off-target genes for each candidate siRNA, sets a threshold for the number of off-targets to minimize miRNA-like off-target effect and presents the list of predicted off-target genes. A feedback mechanism allows the user to choose specific genes that needs to be filtered out from the list of predicted off-target genes recursively until his/her needs are met. This technique gives a more rational approach towards handling the miRNA-like off- target issue.

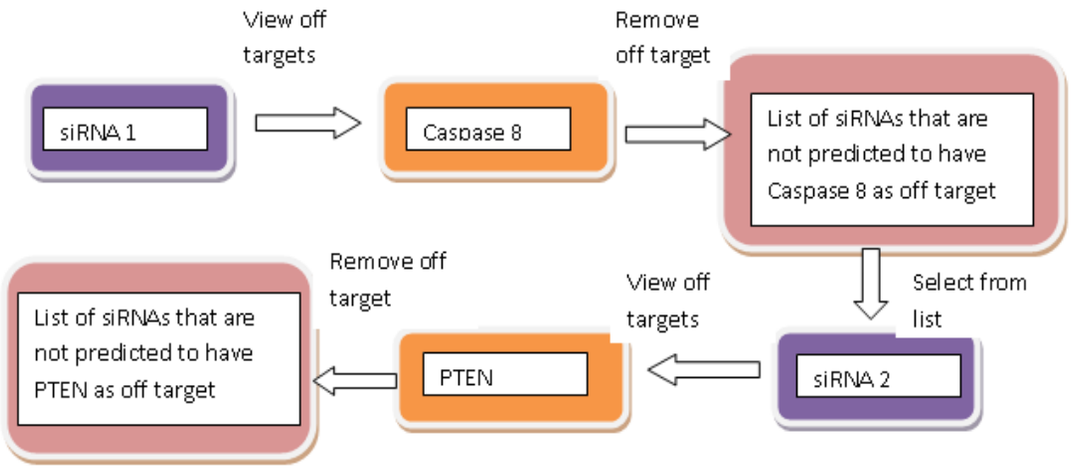

Figure 1. An example of off-target filtering by user feedback

10. Conclusion

As many guidelines are available for selection of efficient siRNAs and they promise to improve efficacy in silencing, much progress has been made in designing siRNAs with improved targeting success rate. Now the main challenge in in-silico siRNA design is reducing the unintended effects arising from inadvertent targets. While the off-targets with near perfect sequence complementarity with the given siRNA is taken care of much efficiently by targeting against a unique region of the target mRNA, which can be found by homology searching with other mRNAs of the organism, miRNA-like off-targets remain to be addressed with efficiency. The few online siRNA designing solutions those consider for reducing such off-targets rely on reducing the number of off-targets. Even then a compromise has to be made between potency and specificity as siRNA sequences that could give higher specificity often comes with lower potency. Now approaching this specificity

issue arising from miRNA-like off-targets of a siRNA from a different point of view, like the functional off-target filtering discussed in previous section, may prove to be beneficial and may emerge as a new paradigm for designing efficient siRNAs with customized specificity.

Author details

Suman Ghosal, Shaoli Das and Jayprokas Chakrabarti
Indian Association for the Cultivation of Science, Kolkata, India

Acknowledgement

We thank Sanga Mitra and Smarajit Das of Indian Association for the Cultivation of Science for their valuable suggestions.

11. References

[1] Grosshans H. & Filipowicz W. (2008). The expanding world of small RNAs. *Nature,* Vol. 451, (April 2008), pp. (414)
[2] De Fougerolles A., Vornlocher H.P., Maraganore J., Lieberman J. (2007). Interfering with disease: a progress report on siRNA-based therapeutics. *Nature Reviews Drug Discovery,* Vol. 6, (June 2007), pp. (443)
[3] Elbashir S.M., Harborth J., Lendeckel W., Yalcin A., Weber K., Tuschl T. (2001). Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature,* Vol. 411, (May 2001), pp. (494)
[4] Pei Y. & Tuschl T. (2006). On the art of identifying effective and specific siRNAs. *Nature Methods,* Vol. 3, (September 2006), pp. (670)
[5] Park Y.K., Park S.M., Choi Y.C., Lee D., Won M., Kim Y.J. (2008). AsiDesigner: exon-based siRNA design server considering alternative splicing. *Nucleic Acids Research,* Vol. 36, (September 2008), pp. (97)
[6] Khvorova A., Reynolds A., Jayasena S.D. (2003). Functional siRNAs and miRNAs Exhibit Strand Bias. *Cell*, Vol. 115, (November 2003), pp. (505)
[7] Zuker M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* Vol. 31, (July 2003), pp. (3406)
[8] Hofacker I.L. (2003). Vienna RNA secondary structure server. *Nucleic Acids Research*, Vol. 31, (July 2003), pp. (3429)
[9] Gong W., Ren Y., Xu Q., Wang Y., Lin D., Zhou H., Li T. (2006). Integrated siRNA design based on surveying of features associated with high RNAi effectiveness. BMC Bioinformatics. Vol. 7, (November, 2006), pp. (516)
[10] Reynolds A., Leake D., Boese Q., Scaringe S., Marshall W.S., Khorova A. (2004). Rational siRNA design for RNA interference. *Nature Biotechnology,* Vol. 22, (September 2004), pp. (326)
[11] Huesken D., Lange J., Mickanin C., Weiler J., Asselbergs F., Warner J., Meloon B., Engel S., Rosenberg A., Cohen D., Labow M., Reinhardt M., Natt F., Hall J. (2005). Design of a genome-wide siRNA library using an artificial neural network. *Nature Biotechnology,* Vol. 23, (July 2005), pp. (995)

[12] Ren Y., Gong W., Xu Q., Zheng X., Lin D., Wang Y., Li T. (2006). siRecords: an extensive database of mammalian siRNAs with efficacy ratings. *Bioinformatics,* Vol. 22, (January 2006), pp. (1027)
[13] Platt J. C. (1999) Fast training of support vector machines using sequential minimal optimization, Advances in kernel methods. *MIT Press Cambridge*, pp. (185)
[14] MIT siRNA database [http://web.mit.edu/sirna/sirnas-human.html].
[15] Jackson A.L. & Linsley P.S. (2010). Recognizing and avoiding siRNA off-target effects for target identification and therapeutic application. *Nature Review Drug Discovery,* Vol. 9, (January 2010), pp. (57)
[16] Judge D., Sood V., Shaw J.R., Fang D., McClintock K., MacLachlan I. (2005). Sequence-dependent stimulation of the mammalian innate immune response by synthetic siRNA. *Nature Biotechnology,* Vol. 23, (March 2005), pp. (457)
[17] Fedorov Y., Anderson E.M., Birmingham A., Reynolds A., Karpilow J., Robinson K., Leake D., Marshall W.S., Khvorova A. (2006). Off-target effects by siRNA can induce toxic phenotype. *RNA,* Vol. 12, (March 2006), pp. (1188)
[18] Birmingham A., Anderson E.M., Reynolds A., Ilsley-Tyree D., Leake D., Fedorov Y., Baskerville S., Maksimova E., Robinson K., Karpilow J., Marshall W.S., Khvorova A. (2006). 3' UTR seed matches, but not overall identity, are associated with RNAi off-targets. *Nature Methods,* Vol. 3, (March 2006), pp. (199)
[19] Burchard J., Jackson A.L., Malkov V., Needham R.H.V., Tan Y., Bartz S.R., Dai H., Sachs A.B., Linsley P.S. (2009). MicroRNA-like off-target transcript regulation by siRNAs is species specific. *RNA,* Vol. 15, (February 2009), pp. (308)
[20] Wang L. & Forest Y.M. (2004). A Web-based design center for vector-based siRNA and siRNA cassette. *Bioinformatics,* Vol. 20, (September 2004), pp. (1818)
[21] Nielsen C.B., Shomron N., Sandberg R., Hornstein E., Kitzman J., Burge C.B. (2007). Determinants of targeting by endogenous and exogenous microRNAs and siRNAs. *RNA,* Vol. 13, (November 2007), pp. (1894)
[22] Ui-Tei K., Naito Y., Nishi K., Juni A., Saigo K. (2008). 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 Research,* Vol. 36, (November 2008), pp. (7100)
[23] Yuan B., Latek R., Hossbach M., Tuschl T., Lewitter F. (2004). siRNA Selection Server: an automated siRNA oligonucleotide prediction server. *Nucleic Acids Research*, Vol. 32, (July 2004), pp. (130)
[24] Anderson E. M., Birmingham A., Baskerville S., Reynolds A., Maksimova E., Leake D., Fedorov Y., Karpilow J. & Khvorova A. (2008). Experimental validation of the importance of seed complement frequency to siRNA specificity. *RNA,* Vol. 14, (May 2008), pp. (853)
[25] Kiryu H., Terai G., Imamura O., Yoneyama H., Suzuki K., Asai K. (2011). A detailed investigation of accessibilities around target sites of siRNAs and miRNAs. *Bioinformatics,* Vol. 27, (July 2011), pp. (1788)
[26] Schultz N., Marenstein D.R., De Angelis D.A., Wang W.Q., Nelander S., Jacobsen A., Marks D.S., Massagué J., Sander C. (2011). Off-target effects dominate a large-scale RNAi screen for modulators of the TGF-β pathway and reveal microRNA regulation of TGFBR2. *Silence,* Vol. 2, (March 2011)