Chapter 1

What Parameters to Consider and Which Software Tools to Use for Target Selection and Molecular Design of Small Interfering RNAs

Olga Matveeva

Abstract

The design of small gene silencing RNAs with a high probability of being efficient still has some elements of an art, especially when the lowest concentration of small molecules needs to be utilized. The design of highly target-specific small interfering RNAs or short hairpin RNAs is even a greater challenging task. Some logical schemes and software tools that can be used for simplifying both tasks are presented here. In addition, sequence motifs and sequence composition biases of small interfering RNAs that have to be avoided because of specificity concerns are also detailed.

Key words: siRNA efficiency, siRNA specificity, Duplex stability, Nucleotide preferences, RNA structure, Software-tools

1. Introduction

1.1. siRNA- and shRNA-Mediated Gene Silencing

Designing any molecule to mimic nature's design is not an easy task. Knowing the pathways that nature has taken and roles that the molecule plays in these pathways simplifies the mission. Small RNA interfering (RNAi) molecules have been extensively studied, yet there are substantial gaps in our knowledge of their natural function and processing. For this reason, bioinformatics of RNAi design represents a combination of solid statistical facts and educated guesses.

What is the function of small interfering RNA (siRNA) in mammals? It is well established that in plants and insects short siRNA duplexes can be generated from long double-stranded RNA by the enzyme Dicer in response to viral infections that promote viral RNA destruction (1). Whether RNAi plays a similar role as a viral infection-fighting agent in mammals is as yet unknown. siRNA

Debra J. Taxman (ed.), *siRNA Design: Methods and Protocols*, Methods in Molecular Biology, vol. 942,
DOI 10.1007/978-1-62703-119-6_1, © Springer Science+Business Media, LLC 2013

in a form that resembles a Dicer cleavage product (two paired 21 nt synthetic oligonucleotides with a 19 nt duplex region and 3′ end 2 nt overhangs) has become a popular research tool. After their introduction into a cell, siRNAs can cause specific target mRNA degradation. Thus, the biological consequences of the down-regulation of any gene can be studied.

siRNA-mediated silencing of mammalian genes uses synthetic oligonucleotides transfected into cells. An alternative approach employs the expression of short hairpin RNAs (shRNAs) in cells following delivery of expression plasmids or viral vectors (2, 3). shRNAs are artificial analogs of endogenous microRNAs (miRNAs), the vast class of small noncoding RNA molecules that regulate the stability and translation of their target mRNAs. Precursors of miRNAs (pre-miRNAs) are stable hairpins, which are encoded in plant and animal genomes. miRNAs play important regulatory roles in animals and plants. Most of these molecules appear to regulate the expression of a diverse set of genes. This regulation takes place during embryonic development, apoptosis, tissue regeneration, and so forth (for review, see ref. 4).

The approach to gene silencing based on synthesized siRNAs is fast and simple. The shRNA-based approach is more laborious and time consuming, but it is becoming increasingly popular. Compared to chemically synthesized siRNAs, the shRNA methodology offers advantages in silencing longevity and lower costs for genome-wide studies. Also, gene therapy is a particularly promising application for shRNAs. It is believed that transcription of shRNA delivers lower intracellular concentrations of siRNA-like products, compared to synthetic siRNA oligonucleotides transfected into cells. Lower intracellular concentrations achieved through the natural process of transcription for extended periods of time can yield more specific silencing effects. It is assumed that after shRNA molecules are processed in cells they enter the same enzymatic pathways as siRNAs.

One of the two strands of an siRNA duplex, the "guide strand," enters the RNA-induced silencing complex (RISC). RISC loaded with a guide strand acquires the ability to cleave RNA sequences that become base paired with the guide strand. The other duplex strand, the "passenger" strand, is the first "victim" of such newly acquired enzymatic ability. It gets cleaved and its fragments dissociate from the duplex. After passenger strand cleavage and dissociation, the guide strand becomes free from any base pairing; consequently, it becomes available for interaction with a new partner (5, 6). This new partner must feature some complementarity to the guide strand RNA sequence. Thus, an mRNA can be cleaved and destroyed if it includes a region complementary to the guide strand. In relation to the mRNA, the siRNA "guide strand" is synonymous to the "antisense strand," while the "passenger strand" is synonymous to the "sense strand."

The RISC core protein complex contains a member of the Argonaute (Ago) family of proteins (7). So far, variable number of different Ago protein-coding genes have been characterized in different organisms (8). In mammals there are eight Ago genes (9, 10). All of these eight genes have been identified in the human genome (11). Among them, only the product of Ago2 mediates RNA cleavage directed by siRNA (12).

RISC can cleave mRNA between residues base paired to nucleotides 10 and 11 of the siRNA, and the cleavage itself does not require ATP. Multiple rounds of mRNA cleavage can be guided by the same siRNA as long as it remains associated with the Ago complex. It is likely that the release of the cleaved mRNA products involves an RNA helicase, which is dependent on ATP, because several proteins associated with the RNAi pathway in *Drosophila* and other organisms contain RNA helicase/ATPase domains (13). The cleavage of the human siRNA-mediated passenger strand and of mRNA is associated solely with an Ago2-containing RISC (Ago2-RISC) (12, 14).

Because processing of artificial siRNAs and shRNAs in cells utilizes the main components of the cellular RNAi machinery, the design of new versions of these molecules should allow provision for successful interaction with RISC and mRNA targets.

1.2. History of Development of Algorithms for Predicting siRNA Silencing Efficiency

Independently selected siRNA duplexes for different mRNA target regions can have vastly different silencing efficiency. In 2003, two independent studies made very important contributions to understanding how this happens (15, 16).

The experiments performed in the first study (15) demonstrated that the rate of RISC entry might be very different for the two strands of an siRNA duplex. Both the absolute and relative 5′ terminal duplex stabilities of the two siRNA strands determine the speed of the process. In other words, RISC prefers the strand whose 5′ end more loosely pairs with its complement. Such a strand enters RISC fast with consequent efficient and fast cleavage of the target. Conversely, a strand whose 5′ end pairs tightly with its complement enters RISC slowly with an elevated chance of target cleavage failure. The statistical analysis performed in the other study (16) revealed that efficient and inefficient siRNAs differ in their 5′ antisense strand terminal duplex stability. The 5′ ends of efficient guide strands pair less stably with their complements. The findings of both studies are consistent with one another; fast RISC entry is a prerequisite for efficient siRNA guide strand cleavage. Thus, asymmetry in siRNA terminal duplex stability defines which strand enters RISC efficiently and which strand guides efficient target cleavage. The asymmetry can be evaluated by the calculation of difference in G/C or A/T nucleotide content at siRNA duplex ends; however, it is more correctly captured by the difference in 5′ terminal free energy (ddG) of the guide and passenger strand.

While asymmetry in siRNA duplex stability is an important feature defining molecular silencing efficiency, it is not the only

important feature. Studies from 2004 onwards, some bioinformatic, focused on analysis of other features defining siRNA silencing efficiency as well as on ways to predict it. The logic of these studies followed a common series of steps. A database of siRNA with variable efficiency was assembled, sequence features (parameters) associated with efficient siRNAs were identified, and recommendations were formulated for how to use all of these features for experimental molecular design. The first published study resulted from analysis of more than a hundred siRNA molecules with variable efficiency (17). The list of features that were found to be associated with siRNA efficiency includes optimal G/C content, low terminal duplex stability of sense strand, lack of inverted repeats, and certain nucleotide preferences in the sense strand. A second published study that resulted from analysis of a few dozen siRNAs also suggested a list of features that included base preference of A or U at the 5′ end of the antisense strand; G or C at the 5′ end of the sense strand; at least five A/U residues in the 5′ terminal one-third of the antisense strand; and the absence of any GC stretch of more than 9 nt in length (18). The authors of both works suggested that the probability of an siRNA candidate being efficient could be evaluated according to the number of these sequence features associated with the candidate. The studies described above involved only "human intelligence" without sophisticated machine learning techniques—"artificial intelligence."

The sequence features associated with siRNA efficiency even in these early and not-so-sophisticated bioinformatics studies can be separated into at least four different categories. The first category relates to asymmetry in siRNA terminal duplex stability, the second to total siRNA duplex stability, the third to base preferences at different siRNA strand positions, and the fourth to the necessity of avoiding certain nucleotide motifs in siRNA sequences. Some of these categories are overlapping. For example, as mentioned above, asymmetry in siRNA terminal duplex stability can be roughly evaluated through certain base preferences at the 5′ and 3′ ends of siRNA strands. Total siRNA duplex stability can be roughly approximated through calculation of the total GC content of any particular duplex strand.

Computer models that can deliver siRNA design candidates with high probability of being efficient require reliable input parameters (sequence features associated with siRNA efficiency), large experimental datasets, and good data processing schemes. Almost all studies that were published in 2004 and in 2005 used comparatively small databases (no more than a few hundred data points). However, exploration of different machine learning techniques for the purpose of siRNA design begins even with small databases. The initial explorations involved Regression Trees (RT) (19, 20), Genetic Programming (GP) (21), Generalized String Kernel (GSK) combined with Support Vector Machine (SVM) (22, 23), and so forth.

A large experimental database (more than 2,000 siRNA data points) was published by Novartis in 2005 (24). This database, as well as others (25) that were compiled from experiments described in the literature, became fuel for the next round of studies that explored advanced statistical analysis and Artificial Intelligence for developing algorithms for efficient siRNA design. These studies utilized Neuronal Networking (24, 26) and supervised learning of a Radial Basis Function (RBF) network combined with Regression Trees (27), Linear Regression (28–30), SVM (31–34), Random Forest Regression (35), and RBF (36). The siRNA design algorithms utilizing large experimental databases were called second-generation algorithms.

There are a number of statistical tools which can be helpful for evaluating the performance of the algorithms listed above. When a predictive method outputs a continuous measure of siRNA efficiency, then correlation between experimental and theoretical values can be calculated. When a predictive method outputs a binary measure of efficiency, for example, siRNA is predicted to be either efficient or inefficient, then Receiver Operating Characteristic or Precision Recall Curves can be created. Both types of curves characterize an algorithm's performance from slightly different angles (37). Despite the existence of all of these statistical techniques and the expansion of approaches for siRNA design, very few studies addressed algorithm comparison issues (30, 38). A more recent work was published in 2007 (30), and it revealed that four algorithms (24, 29, 30, 39) out of 11 used for comparison fared better than the others.

Although considerable success has been reported in predicting siRNA activity, very few studies deal with the analysis of shRNA features related to their silencing efficiency (3, 25, 40). The difference between the fate of chemically synthesized siRNA and intracellular transcribed shRNA within the cell is in the transcription and enzymatic processing. shRNA has to be first transcribed and then cleaved by RNases to enter the RISC pathway, while siRNA enters this pathway without transcription or cleavage. Two popular types of shRNA molecules are currently experimentally used. The first type is miRNA-like shRNAs. Their design is based upon one particular miRNA (miR30) and they are frequently employed for loss-of-function assays (41, 42). These molecules have long (more than 21 nt) sometimes partially mismatched stems, and undergo processing by RNaseIII-like endoribonucleases (Drosha and Dicer). The second type is shRNAs with short perfectly paired stems (19–21 nt). They are more frequently used in experiments that require silencing of individual genes (2).

It is likely that the processing of miR30-based shRNAs depends on the rates of Drosha and Dicer cleavages. Both enzymes demonstrate certain nucleotide preferences for their cleavage sites. This is probably the reason why approaches that try to transfer siRNA design to miRNA-like shRNAs have a tendency to fail (40). It has been suggested, however, that the processing of shRNAs of

the second type from shorter (19 nt) stems is not Dicer dependent (43). Perhaps single-strand RNases (for example, representatives of the RNase A gene superfamily) are involved in the processing of shRNAs with short stems. Thus, the ability of such shRNAs to silence genes might depend on the susceptibility of their loop sequences to RNase A cleavage. The 9 nt "UUCAAGAGA" loop was described as the optimal configuration for a potent silencing trigger (2, 44). Both types of shRNA enter the RISC after enzymatic cleavage, and at this stage the pathways of chemically synthesized and intracellularly transcribed molecules merge.

This scenario suggests that design rules for shRNA duplexes, which emerge after the enzymatic cleavage step, should be highly similar to those for siRNA duplexes. However, the precise location of enzymatic cleavage sites as well as cleavage site sequence requirements are not always well characterized. Moreover, efficient transcription considerations should also affect shRNA design rules. Thus, bioinformatics of shRNA design belongs primarily to the future. Future progress needs to be fueled by new experimental databases that permit careful analysis of transcription and cleavage preferences of all the enzymatic machinery involved in the shRNA pathway. One such database was recently published and hopefully is going to be a great asset for the design of efficient miR30-based shRNAs (40).

1.3. List of siRNA Sequence Features Associated with Specific Silencing Efficiency

The sequence features found to be associated with siRNA-specific silencing efficiency can be subdivided into at least seven categories. A list of these categories is provided below, along with the references to the studies in which the relationship between the experimental siRNA silencing efficiency and feature presence were studied. Additionally, the rationales for incorporating these features are presented in greater detail in Subheadings 1.4–1.7 below.

Category 1. Asymmetry in terminal duplex stability in siRNA (15–18, 25, 31, 34, 39, 45–47).
Category 2. Total siRNA duplex stability (a rough approximation can be made through the evaluation of GC content) (17, 21, 25, 27, 31, 34, 39, 40, 45, 47).
Category 3. Internal local siRNA duplex stability (30, 39, 40).
Category 4. Base preferences at different siRNA strand positions (17, 20, 23, 24, 31, 36, 39, 40, 45, 47–50).
Category 5. Frequency of occurrence or avoidance of certain motifs in the complete sequence and at certain positions along the siRNA (22, 29, 31, 51).
Category 6. Guide siRNA strand (31, 33, 34, 52, 53) and/or target mRNA secondary structure (33, 34, 47, 54–61).
Category 7. Negative influence of cross-hybridization potential (39, 62).

The siRNA sequence features belonging to the first two categories most strongly affect silencing efficiency. In one way or another they are included in all siRNA design algorithms.

1.4. Optimal Asymmetry in Terminal Duplex Stability

High asymmetry in terminal duplex stability is responsible for a high rate of RISC loading for the siRNA cleavage guidance strand and a low rate of RISC loading for the passenger strand. This rate difference eliminates strand competition for RISC access. Theoretically, this competition is not a problem when the RISC molecular concentration is higher than the concentration of siRNAs. However, even in such cases, RISC loading with the passenger strand is not a desirable event because RISC being loaded with the passenger strand can make an extra contribution to nonspecific cleavages. In other words, optimization of the terminal asymmetry in siRNA duplexes increases both the efficiency and specificity of silencing.

Two scenarios for interaction between the siRNA strands and RISC are possible (Fig. 1). The first scenario happens when the concentration of RISC is below that of the siRNA and the second when the concentration of RISC is equal to, or above, that of the siRNA. According to the first scenario, the competition between the strands for RISC entry happens if the 5′ ends of both strands are loosely paired and does not occur when one of the 5′ ends is tightly paired with its complement. According to the second scenario, there is no competition between the two strands regardless of how tightly their 5′ ends are paired with the complements. The absence of competition between the strands can be a consequence of RISC concentration excess or terminal asymmetry in siRNA duplexes.

If experiments are performed in RISC excess, symmetrical or asymmetrical duplexes will appear efficient (no strand competition). If experiments are done in siRNA excess only asymmetrical duplexes will appear efficient (strand competition exists but only for symmetrical duplexes). It is important to understand that sometimes it appears that asymmetry is not needed because in a RISC excess scenario symmetrical duplexes with loosely paired ends will look efficient.

Even though considerable time has passed since the 2003 discovery that asymmetry in siRNA terminal duplex stability defines which strand is loaded into RISC (15), the optimal method for calculating this asymmetry is still debatable. Mismatches in the first four nucleotides from the 5′ end of an siRNA duplex affect the rate of strand entry into RISC. This led to the conclusion that 4 nt should be considered for asymmetry calculation. However, careful statistical analysis has demonstrated that the contribution of the first nucleotide to siRNA efficiency is the strongest, while that of the fourth is the weakest (63).

Analysis of four independent experimental databases revealed that a ddG value equal to, or above, 2 kcal/mol calculated for the two terminal nucleotides of an siRNA duplex corresponds to the highest silencing activity (25). Almost all siRNA duplexes with two U/A at the 5′ end of the cleavage guidance strand and two G/C

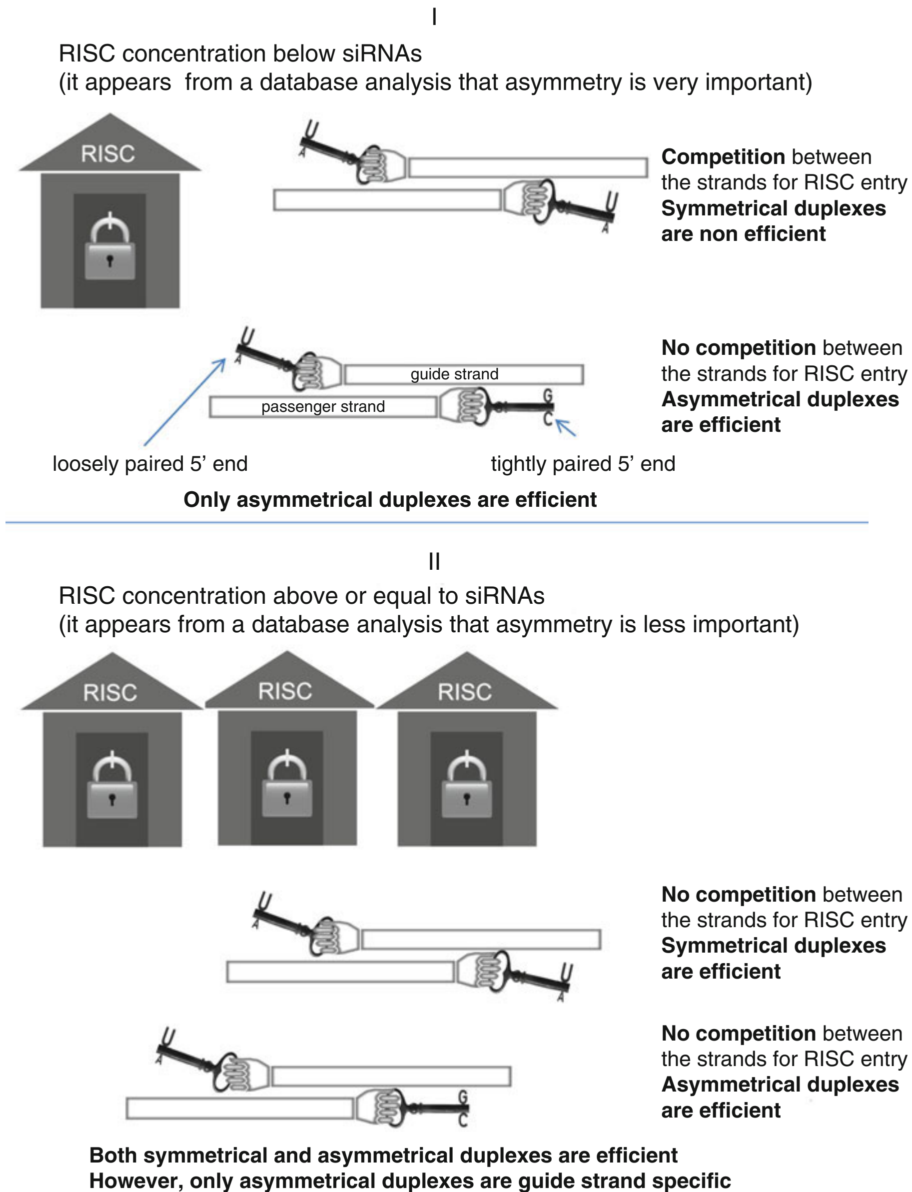

Fig. 1. Two scenarios of siRNA–RISC interaction. Two scenarios for interaction between siRNA strands and RISC are possible. The first scenario (I) happens when the concentration of RISC is below that of siRNA, and the second (II) when the concentration of RISC is equal to, or above, that of siRNA. According to the first scenario, the competition between the strands for RISC entry happens if the 5′ ends of both strands are loosely paired and does not occur when one of the 5′ ends is tightly paired with its complement. According to the second scenario, there is no competition between both strands regardless of how tightly their 5′ ends are paired with the complements. The absence of competition between the strands can be a consequence of RISC concentration excess or terminal asymmetry in siRNA duplexes.

at its 3′ end belong to the category of those with optimal ddG. It was noticed that the guide strand of efficient siRNA and shRNA molecules frequently starts with U rather than with A (40). This bias is most likely explained by AGO2 binding preferences (64). Consequently, selection of siRNAs with U at the first position from

the 5′ end of the cleavage guidance strand, U/A at the second position, as well as two G/C at their 3′ end should deliver the best silencing efficiency and specificity in RISC loading. That is why in Fig. 1 "U" is presented with larger symbol in comparison with "A" in the RISC entry key.

For shRNAs, transcription start nucleotide preferences have to be considered to achieve optimal duplex asymmetry. The transcription of shRNAs is usually performed from Polymerase III promoters H1 or U6 (2, 3). The H1 promoter favors adenine at the first position of a potential encoded shRNA, while a U6 promoter favors guanine.

1.5. Optimal siRNA Duplex Stability

siRNA duplex stability should be neither too high nor too low for efficient siRNA functioning and have dG values between −35 and −27 kcal/mol (25). Low duplex stability results in slow formation and short lifetime of cleavage guidance strand–target duplexes, with consequent inadequate opportunity for RNA cleavage to occur. On the other hand, the siRNA passenger strand of duplexes that are too stable may dissociate too slowly even after cleavage by RISC. The same problem can occur with target RNA cleavage products. In addition, stable siRNA duplexes are usually GC rich and GC-rich guide strands, or their target regions, are usually constituents of stable RNA secondary structures. The optimum value of duplex stability of most efficient siRNAs or shRNAs might be concentration dependent, however.

1.6. Optimal Profile of Internal Local siRNA Duplex Stability and Nucleotide Preferences

The nucleotide preferences at terminal siRNA strand positions of efficient molecules are related to terminal duplex asymmetry, so they are easily explainable. However, it is still unclear why some nucleotides are preferred at certain nonterminal positions of siRNA strands and some are avoided. Speculatively, the preference for A at the 10th position of a guidance strand has been revealed in a number of studies (17, 20, 30, 34), and may be related to RISC cleavage preference, because this cleavage occurs between residues base paired to nucleotides 10 and 11 of the siRNA. Perhaps preferential low stability (A/U enrichment) of some base pairs (6th, 7th, 12th, 13th, and 14th from the 5′ end of the guide strand) in the siRNA duplex (30, 34) is needed for easier dissociation of the cleaved RNA product and enhances RISC turnover. Some evidence was obtained that this dissociation proceeds energetically uphill because it involves an ATP-dependent RNA helicase P (10).

1.7. Specificity Problems

Early reports related to siRNA target gene silencing specificity were very optimistic; they suggested that siRNA is a highly specific tool for targeted gene knockdown (65, 66). However, later microarray studies showed that in addition to the intended target, siRNAs down regulate many unintended target transcripts (67–69). Intended and unintended transcript targets may be silenced

with indistinguishable kinetics, though unintended transcripts are generally down regulated to a smaller degree (67). Off-target silencing is now widely recognized as a complication of any studies involving siRNAs (70, 71). It seems that to a large extent, target silencing effects are mediated by the participation of artificial siRNAs in natural miRNA pathways. The miRNA translational suppression pathway is directed by imperfect base pairing between the target and the miRNA guide strand (72). The specificity of this base-pairing depends on the 6–7 base "seed region" at the 5′ end of the guide strand of an miRNA. This miRNA-type seed-matched, off-target silencing by siRNAs is very common. Most of the siRNA unintended targets share sequence complementarity in their 3′ UTR regions with residues 1–8 (or contain seed motif matching nucleotides 1–6, 2–7, or 3–8) of the siRNA guide strand (67–69, 73).

The investigation of "seed region" complementarity and siRNA off-target silencing or miRNA down regulation revealed several interesting facts. It was found that the stability of a duplex between the siRNA "seed region" and an off-target region, measured as free energy dG, and the degree of gene silencing correlate with each other (74). A correlation was also discovered between the frequency of a seed hexamer being present in the 3′ UTR portion of a transcriptome and siRNA off-targeting effects. In other words, siRNAs with the most unique hexamers in their seed regions are less prone to off targeting (74). It was also found that extra complementarity within the 3′ end of a guide strand increases the miRNA targeting effect. The effect was also increased if miRNA target seed regions were located within AU-rich regions of 3′ UTRs (75). The latter observation might indicate that AU-rich mRNA regions are usually not involved in stable base-pairing within RNA secondary structure and consequently are more likely to be available for intermolecular interaction with miRNA.

siRNAs can be potent inducers of interferons (IFNs) and inflammatory cytokines both in vivo and in vitro (76–78). These findings are promoting a growing concern among researchers whether activation of an immune response can be systematically avoided during gene silencing experiments. The problem of immune-stimulatory motifs can be addressed through certain chemical modifications; however, this approach has its own limitations. For example, it cannot be used for in vivo transcribed shRNA, which along with chemically synthesized siRNA, is a popular tool for gene silencing experiments.

A basic series of steps for incorporating the design criteria presented in Subheading 1.3 and further detailed in Subheadings 1.4–1.7 is provided below. Suggested Web resources that can best help to achieve these criteria are also provided.

2. Materials

Computer with Internet access.

3. Methods

Steps for designing siRNAs according to the seven criteria categories provided in Subheading 1.3, and further discussed in Subheadings 1.4–1.7, are provided below. Step 1 can be used as a broad method for selecting siRNAs that satisfy criteria categories 1–5. The remaining steps will help to refine the selection of siRNAs with consideration of target secondary structure as well as siRNA specificity requirements to satisfy criteria categories 6–7.

1. Use one of the following Web tools for filtering out inefficient siRNAs according to the features from the categories 1–5 in Subheading 1.3: http://www5.appliedbiosystems.com/tools/siDesign/ or http://www.med.nagoya-u.ac.jp/neurogenetics/i_Score/i_score.html (see Notes 1 and 2).
2. (Optional) Use Web tools for filtering out inefficient siRNAs according to the optimal accessibility of the guide strand and target secondary structures (see Note 3): http://rna.tbi.univie.ac.at/cgi-bin/RNAxs?hakim=1, http://sfold.wadsworth.org/cgi-bin/sirna.pl, or http://rna.urmc.rochester.edu/servers/oligowalk.
3. (Optional) Use siRNA candidates that are in overlapping pools in outputs from steps 1 and 2.
4. Filter out siRNA candidates that are vulnerable to nonspecific silencing efficiency due to the presence of immune-stimulating motifs if you are using any software other than http://www5.appliedbiosystems.com/tools/siDesign/. Three sequence motifs, "UGUGU," "GUCCUUCAA," and "AUCGAU(N)nGGGG," should be included in an immune-motif avoidance list. In addition, U-rich sequences or sequences with biased nucleotide content, such as (G+U)>>(C+A) or with A+U or G+U rich motifs, should also be avoided (see Note 4).
5. Filter out siRNA candidates with the motif "UGGC," which is associated with reduced cell viability (see Note 5).
6. Filter out siRNA candidates with motifs of low sequence complexity such as "GGGG," "UUUU," "CCCC," and "AAAA" (see Note 6).
7. Use Washington University BLAST (WU-BLAST) to filter out the least specific siRNAs: http://informatics-eskitis.griffith.edu.au/SpecificityServer/ (see Notes 7–9).
8. Use for your experiments siRNA candidates with the top scores from step 1 that pass all the filters listed above.

4. Notes

1. A large number of Web services for siRNA design are currently available. However, very few of them use the most relevant updated information, so I have recommended two Web sites that are most committed to predicting efficient siRNAs:

   http://www5.appliedbiosystems.com/tools/siDesign/ (34). This Web service filters out siRNA candidates with miRNA "seed" matches, and toxic and immunostimulatory motifs.

   http://www.med.nagoya-u.ac.jp/neurogenetics/i_Score/i_score.html (27) This Web site allows the calculation of siRNA efficiency scores based upon nine independent siRNA design tools. In addition it provides thermodynamic information related to the dG value of an siRNA duplex, dG of guide strand secondary structure, and dinucleotide dG values at the 5′ and 3′ ends. However, the Web service does not filter out siRNAs with continuous runs of identical nucleotides or other non-desirable siRNA sequence motifs. It does not consider the BLASTN score.

2. Two independent studies suggested that the ability of the siRNA guide strand to cross hybridize with multiple off-target regions (39, 62) diminishes this strand's ability to silence its own target. Recently, a downloadable software was described that can perform siRNA design steps that can substitute for steps 1, 2, and 7 (79).

3. There are reported correlations between the openness of the secondary structures of the siRNA guide strand and target mRNA. I have recommended three Web tools that offer design of siRNAs based upon consideration of these sequence features. It has to be noted that for http://sfold.wadsworth.org/cgi-bin/sirna.pl the restriction in mRNA target length is 250 nt. Two other Web sites can deal with longer mRNAs. However, all the listed Web services do not filter out siRNAs with continuous runs of identical nucleotides or some other non-desirable siRNA sequence motifs.

4. Varieties of different specialized cellular receptors have evolved for recognition of pathogen-associated molecular patterns (PAMPs) that trigger immune responses when their ligands (viral or bacterial components) become present (80, 81). Some of these receptors are restricted to immune cells only, while others are present in all cells. Toll-like receptors (TLRs) are type I transmembrane proteins involved in the functioning of innate immunity by recognizing pathogen-specific molecules including DNA and RNA. The human TLR7 and TLR8 immune response appears to be stimulated strongest with U-rich sequences (82–84). AU-rich motifs mostly were

shown to activate TLR8 and GU-rich motifs both TLR7 and TLR8 receptors (85). A- and U-rich sequences stimulate TLR8, which triggers the production of both IFNα and TNFα from monocytes (84, 85), while G- and U-rich sequences trigger TLR7, thereby causing IFNα production from pDCs (86). Certain siRNA sequences are particularly immuno-stimulatory, such as "GUCCUUCAA" (77) and "UGUGU" (78, 87). Experimental evidence demonstrates that the latter motif triggers induction of IFN type 1 and nonspecific gene down regulation through TLR7. However, some siRNA sequences, independent of their GU or AU enrichment, are potent inducers of IFN-alpha production. One study showed that the presence of the CpG motif "AUCGAU" in RNA oligonucleotides together with a poly-G tail stimulates monocytes to produce large amounts of IL-12 (88).

5. A strong correlation has been found between the presence of motif "UGGC" in an siRNA cleavage guidance strand and reduced cell viability (89).
6. Clear answers are elusive about why certain motifs are preferred or avoided in efficient siRNA molecules. These parameters are poorly studied. However, it is widely believed that it is desirable to avoid targeting mRNA regions with low sequence complexity, particularly those represented by continuous runs of four identical nucleotides. It is assumed that the "UUUU" motif should be avoided during design of shRNAs because it represents an RNA III polymerase termination site.
7. NCBI BLASTN allows the fast discrimination of siRNA design candidates according to their homology with potential off-target hits. It can be also used for simple homology searches. For BLAST searching, it is recommended to discard any candidate with a score above 30 (34). However, BLASTN default parameters have been demonstrated to be inappropriate for siRNA design and other parameters have been suggested (90).
8. In general, siRNA off-target silencing is a function of many characteristics of duplexes between the guide strand and multiple regions in a transcriptome that are partially complementary to this strand (off-target hits) (91). So far, there is no ideal software for careful screening and characterization of these imperfect mismatched duplexes.
9. It is preferable to design siRNAs without "seed regions" present in known miRNAs. miRNA-like off targeting that involves mRNA translation inhibition can lead to gene-nonspecific down regulation (92). The list of these regions can be extracted from the sequences of mature miRNAs. A database of these sequences is available from http://www.mirbase.org/ftp.shtml. However, the siRNA design tool available from http://www5.appliedbiosystems.com/tools/siDesign/ performs this function automatically.

Acknowledgement

The author is very grateful to John Atkins for careful reading of this manuscript and very constructive comments. The work was supported in part by Fred Hutchinson Cancer Center grant DK056465 to animal core facility and by Russian Ministry of Science and Education grant 11.G34.31.0034 to Novosibirsk State University.

References

1. Van Rij RP, Andino R (2006) The silent treatment: RNAi as a defense against virus infection in mammals. Trends Biotechnol 24:186–193
2. Brummelkamp TR, Bernards R, Agami R (2002) A system for stable expression of short interfering RNAs in mammalian cells. Science 296:550–553
3. Paddison PJ et al (2002) Short hairpin RNAs (shRNAs) induce sequence-specific silencing in mammalian cells. Genes Dev 16:948–958
4. Bartel DP (2009) MicroRNAs: target recognition and regulatory functions. Cell 136:215–233
5. Matranga C et al (2005) Passenger-strand cleavage facilitates assembly of siRNA into Ago2-containing RNAi enzyme complexes. Cell 123:607–620
6. Rand TA et al (2005) Argonaute2 cleaves the anti-guide strand of siRNA during RISC activation. Cell 123:621–629
7. Hutvagner G, Simard MJ (2008) Argonaute proteins: key players in RNA silencing. Nat Rev Mol Cell Biol 9:22–32
8. Höck J, Meister G (2008) The Argonaute protein family. Genome Biol 9:210–210
9. Meister G, Tuschl T (2004) Mechanisms of gene silencing by double-stranded RNA. Nature 431:343–349
10. Carmell MA et al (2002) The Argonaute family: tentacles that reach into RNAi, developmental control, stem cell maintenance, and tumorigenesis. Genes Dev 16:2733–2742
11. Sasaki T et al (2003) Identification of eight members of the Argonaute family in the human genome small star, filled. Genomics 82:323–330
12. Meister G et al (2004) Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs. Mol Cell 15:185–197
13. Witold F (2005) RNAi: the nuts and bolts of the RISC machine. Cell 122:17–20
14. Liu J, Carmell MA et al (2004) Argonaute2 is the catalytic engine of mammalian RNAi. Science 305:1437–1441
15. Schwarz DS et al (2003) Asymmetry in the assembly of the RNAi enzyme complex. Cell 115:199–208
16. Khvorova A et al (2003) Functional siRNAs and miRNAs exhibit strand bias. Cell 115:209–216
17. Reynolds A et al (2004) Rational siRNA design for RNA interference. Nat Biotechnol 22: 326–330
18. Ui-Tei K et al (2004) Guidelines for the selection of highly effective siRNA sequences for mammalian and chick RNA interference. Nucleic Acids Res 32:936–948
19. Chalk AM et al (2004) Improved and automated prediction of effective siRNA. Biochem Biophys Res Commun 319:264–274
20. Jagla B et al (2005) Sequence characteristics of functional siRNAs. RNA 11:864–872
21. Saetrom P (2004) Predicting the efficacy of short oligonucleotides in antisense and RNAi experiments with boosted genetic programming. Bioinformatics 20:3055–3063
22. Teramoto R et al (2005) Prediction of siRNA functionality using generalized string kernel and support vector machine. FEBS Lett 579: 2878–2882
23. Jia P et al (2006) Demonstration of two novel methods for predicting functional siRNA efficiency. BMC Bioinform 7:271
24. Huesken D et al (2005) Design of a genome-wide siRNA library using an artificial neural network. Nat Biotechnol 23:995–1001
25. Matveeva OV et al (2010) Optimization of duplex stability and terminal asymmetry for shRNA design. PLoS One 5:e10180
26. Ge G et al (2005) Prediction of siRNA knockdown efficiency using artificial neural network models. Biochem Biophys Res Commun 336: 723–728
27. Takasaki S, Kawamura Y, Konagaya A (2006) Selecting effective siRNA sequences by using radial basis function network and decision tree learning. BMC Bioinform 7(Suppl 5):S22

28. Ichihara M et al (2007) Thermodynamic instability of siRNA duplex is a prerequisite for dependable prediction of siRNA activities. Nucleic Acids Res 35:e123
29. Vert J-P et al (2006) An accurate and interpretable model for siRNA efficacy prediction. BMC Bioinform 7:520
30. Matveeva O et al (2007) Comparison of approaches for rational siRNA design leading to a new efficient and transparent method. Nucleic Acids Res 35:e63
31. Ladunga I (2007) More complete gene silencing by fewer siRNAs: transparent optimized design and biophysical signature. Nucleic Acids Res 35:433–440
32. Peek AS, Behlke MA (2007) Design of active small interfering RNAs. Curr Opin Mol Ther 9:110–118
33. Lu ZJ, Mathews DH (2008) Efficient siRNA selection using hybridization thermodynamics. Nucleic Acids Res 36:640–647
34. Wang X et al (2009) Selection of hyperfunctional siRNAs with improved potency and specificity. Nucleic Acids Res 37:e152
35. Jiang P et al (2007) RFRCDB-siRNA: improved design of siRNAs by random forest regression model coupled with database searching. Comput Methods Programs Biomed 87:230–238
36. Takasaki S (2009) Methods for selecting effective siRNA sequences by using statistical and clustering techniques. Methods Mol Biol 487:1–39
37. Davis J, Goadrich M (2006) The relationship between Precision-Recall and ROC curves. ACM Press, New York, NY, pp 233–240
38. Saetrom P, Snøve O Jr (2004) A comparison of siRNA efficacy predictors. Biochem Biophys Res Commun 321:247–253
39. Shabalina SA, Spiridonov AN, Ogurtsov AY (2006) Computational models with thermodynamic and composition features improve siRNA design. BMC Bioinform 7:65
40. Fellmann C et al (2011) Functional identification of optimized RNAi triggers using a massively parallel sensor assay. Mol Cell 41:733–746
41. Silva JM et al (2008) Profiling essential genes in human mammary cells by multiplex RNAi screening. Science 319:617–620
42. Schlabach M et al (2008) Cancer proliferation gene discovery through functional genomics. Science 319:620–624
43. Siolas D et al (2005) Synthetic shRNAs as potent RNAi triggers. Nat Biotechnol 23:227–231
44. Boudreau RL, Monteys AM, Davidson BL (2008) Minimizing variables among hairpin-based RNAi vectors reveals the potency of shRNAs. RNA 14:1834–1844
45. Amarzguioui M, Prydz H (2004) An algorithm for selection of functional siRNA sequences. Biochem Biophys Res Commun 316: 1050–1058
46. Walton SP, Wu M, Gredell JA, Chan C (2010) Designing highly active siRNAs for therapeutic applications. FEBS J 277:4806–4813
47. Gong W et al (2006) Integrated siRNA design based on surveying of features associated with high RNAi effectiveness. BMC Bioinform 7:516
48. Takasaki S, Kotani S, Konagaya A (2004) An effective method for selecting siRNA target sequences in mammalian cells. Cell Cycle 3: 790–795
49. Holen T (2006) Efficient prediction of siRNAs with siRNA rules 1.0: an open-source JAVA approach to siRNA algorithms. RNA 12: 1620–1625
50. Takasaki S (2009) Selecting effective siRNA target sequences by using Bayes' theorem. Comput Biol Chem 33:368–372
51. Peek AS (2007) Improving model predictions for RNA interference activities that use support vector machine regression by combining and filtering features. BMC Bioinform 8:182
52. Patzel V et al (2005) Design of siRNAs producing unstructured guide-RNAs results in improved RNA interference efficiency. Nat Biotechnol 23:1440–1444
53. Köberle C, Kaufmann SHE, Patzel V (2006) Selecting effective siRNAs based on guide RNA structure. Nat Protoc 1:1832–1839
54. Bohula EA et al (2003) The efficacy of small interfering RNAs targeted to the type 1 insulin-like growth factor receptor (IGF1R) is influenced by secondary structure in the IGF1R transcript. J Biol Chem 278:15991–15997
55. Yoshinari K, Miyagishi M, Taira K (2004) Effects on RNAi of the tight structure, sequence and position of the targeted region. Nucleic Acids Res 32:691–699
56. Heale BS et al (2005) siRNA target site secondary structure predictions using local stable substructures. Nucleic Acids Res 33:e30
57. Brown KM, Chu C-Y, Rana TM (2005) Target accessibility dictates the potency of human RISC. Nat Struct Mol Biol 12:469–470
58. Overhoff M et al (2005) Local RNA target structure influences siRNA efficacy: a systematic global analysis. J Mol Biol 348:871–881
59. Schubert S et al (2005) Local RNA target structure influences siRNA efficacy: systematic analysis of intentionally designed binding regions. J Mol Biol 348:883–893
60. Shao Y et al (2007) Effect of target secondary structure on RNAi efficiency. RNA 13: 1631–1640

61. Tafer H et al (2008) The impact of target site accessibility on the design of effective siRNAs. Nat Biotechnol 26:578–583
62. Alsheddi T et al (2008) siRNAs with high specificity to the target: a systematic design by CRM algorithm. Mol Biol (Mosk) 42:163–171
63. Tilesi F et al (2009) Design and validation of siRNAs and shRNAs. Curr Opin Mol Ther 11: 156–164
64. Frank F, Sonenberg N, Nagar B (2010) Structural basis for 5(prime)-nucleotide base-specific recognition of guide RNA by human AGO2. Nature 465:818–822
65. Elbashir SM et al (2001) Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. Nature 411:494–498
66. Amarzguioui M et al (2003) Tolerance for mutations and chemical modifications in a siRNA. Nucleic Acids Res 31:589–595
67. Jackson AL et al (2003) Expression profiling reveals off-target gene regulation by RNAi. Nat Biotechnol 21:635–637
68. Jackson AL et al (2006) Widespread siRNA "off-target" transcript silencing mediated by seed region sequence complementarity. RNA 12:1179–1187
69. Birmingham A et al (2006) 3′ UTR seed matches, but not overall identity, are associated with RNAi off-targets. Nat Meth 3: 199–204
70. Echeverri CJ et al (2006) Minimizing the risk of reporting false positives in large-scale RNAi screens. Nat Meth 3:777–779
71. Vankoningsloo S et al (2008) Gene expression silencing with "specific" small interfering RNA goes beyond specificity—a study of key parameters to take into account in the onset of small interfering RNA off-target effects. FEBS J 275: 2738–2753
72. Doench JG, Sharp PA (2004) Specificity of microRNA target selection in translational repression. Genes Dev 18:504–511
73. Lin X et al (2005) siRNA-mediated off-target gene silencing triggered by a 7 nt complementation. Nucleic Acids Res 33:4527–4535
74. Anderson EM et al (2008) Experimental validation of the importance of seed complement frequency to siRNA specificity. RNA 14: 853–861
75. Grimson A et al (2007) MicroRNA targeting specificity in mammals: determinants beyond seed pairing. Mol Cell 27:91–105
76. Sioud M (2005) Induction of inflammatory cytokines and interferon responses by double-stranded and single-stranded siRNAs is sequence-dependent and requires endosomal localization. J Mol Biol 348:1079–1090
77. Hornung V et al (2005) Sequence-specific potent induction of IFN-alpha by short interfering RNA in plasmacytoid dendritic cells through TLR7. Nat Med 11:263–270
78. Judge AD et al (2005) Sequence-dependent stimulation of the mammalian innate immune response by synthetic siRNA. Nat Biotechnol 23:457–462
79. Mysara M et al (2011) MysiRNA-designer: a workflow for efficient siRNA design. PLoS One 6:e25642
80. Armant MA, Fenton MJ (2002) Toll-like receptors: a family of pattern-recognition receptors in mammals. Genome Biol 3:reviews3011.1–reviews3011.6
81. Akira S, Hemmi H (2003) Recognition of pathogen-associated molecular patterns by TLR family. Immunol Lett 85:85–95
82. Heil F et al (2004) Species-specific recognition of single-stranded RNA via toll-like receptor 7 and 8. Science 303:1526–1529
83. Sioud M (2006) Single-stranded small interfering RNA are more immunostimulatory than their double-stranded counterparts: a central role for 2′-hydroxyl uridines in immune responses. Eur J Immunol 36:1222–1230
84. Goodchild A et al (2009) Sequence determinants of innate immune activation by short interfering RNAs. BMC Immunol 10:40
85. Forsbach A et al (2008) Identification of RNA sequence motifs stimulating sequence-specific TLR8-dependent immune responses. J Immunol 180:3729–3738
86. Gantier MP et al (2008) TLR7 is involved in sequence-specific sensing of single-stranded RNAs in human macrophages. J Immunol 180: 2117–2124
87. Stewart CR et al (2011) Immunostimulatory motifs enhance antiviral siRNAs targeting highly pathogenic avian influenza H5N1. PLoS One 6:e21552
88. Sugiyama T et al (2005) CpG RNA: identification of novel single-stranded RNA that stimulates human CD14+CD11c+ monocytes. J Immunol 174:2273–2279
89. Fedorov Y et al (2006) Off-target effects by siRNA can induce toxic phenotype. RNA 12: 1188–1196
90. Birmingham A et al (2007) A protocol for designing siRNAs with high functionality and specificity. Nat Protoc 2:2068–2078
91. Dahlgren C et al (2008) Analysis of siRNA specificity on targets with double-nucleotide mismatches. Nucleic Acids Res 36:e53
92. Alemán LM, Doench J, Sharp PA (2007) Comparison of siRNA-induced off-target RNA and protein effects. RNA 13:385–395