ELSEVIER

In silico selection of active siRNA

Volker Patzel

Max-Planck-Institute for Infection Biology, Department of Immunology, Charitéplatz 1, D-10117 Berlin, Germany

RNA interference (RNAi) mediated by short interfering RNA (siRNA) represents a powerful reverse genetics tool, and siRNAs are attracting increasing interest as potential therapeutics. Progress in the design of functional siRNAs has significantly contributed to our understanding of cellular RNA silencing pathways and vice versa. Parameters related to RNA sequence and structure have a strong impact on various steps along the silencing pathway and build the backbone of many siRNA design tools. Recent work has demonstrated that there is more to siRNA design than enhancement of gene silencing activity. Current efforts aim at avoidance of off-target effects, the understanding of siRNA-triggered immunostimulation, and evasion of interference with cellular regulatory RNA. Molecular features determining the biological functions of siRNA and their meaning for computational (*in silico*) selection are the focus of this review.

RNAi is a powerful evolutionarily conserved defence mechanism against viral invasion, transposon expansion and post-transcriptional regulation of gene expression. As such, RNAi has been quickly adapted for use as a unique reverse genetics tool. Today, siRNAs are widely used for functional gene knockdown and target validation and have moved into the spotlight as a novel group of therapeutics [1,2].

In mammalian systems, RNAi is triggered by siRNAs and micro RNAs (miRNAs) [3,4]. siRNA and miRNA duplexes comprise complementary RNA strands of 17–23 nt. Whereas miRNAs frequently form imperfect duplexes, siRNAs are generally designed as perfectly base-paired double strands. One of the two strands, termed the 'guide strand', is included in the RNA-induced silencing complex (RISC). The other strand, termed the 'passenger strand', is excluded and destroyed. Only when the antisense strand is chosen for RISC formation is the RISC guided to the mRNA target to induce gene silencing.

In mammalian cells, siRNA-triggered RNAi begins with formation of the RISC-loading complex, which incorporates the steps of siRNA duplex recognition and definition of the guide and passenger strands. Subsequent steps include formation and activation of the RISC, cleavage and release of the passenger strand, targeting and cleavage of mRNA, and release of the cleaved target sequence before targeting of further mRNA molecules [5–7]. The RISC is a multiprotein complex containing one member of the Argonaute (Ago) family in its core. Whereas an miRNA-induced RISC (miRISC) can contain Ago1, Ago2, Ago3 or Ago4, siRNA-induced mRNA cleavage and passenger strand cleavage are associated exclusively with an Ago2-containing RISC (Ago2–RISC) in humans [5,8,9].

Inefficient delivery and low bioactivity represent key obstacles to *in vivo* applications of siRNA. In principle, $L-21$ siRNAs containing guide strands of 21 nt can be directed against a mRNA of length L. Statistically, however, only a few gene-specific siRNAs show satisfactory silencing activity leading to detectable phenotypes [10]. Whereas improvement of delivery still represents a challenging task for drug development based on nucleic acids, the definition of criteria for designing active siRNA sequences has made significant progress. siRNAs, however, can be prone to non-specific effects and cellular degradation, and often many probes need to be designed to identify one that works.

siRNA design takes into consideration general aspects including RNA synthesis, delivery, biodistribution, bioavailability and metabolism, which are experimentally determined and can be modulated by chemical modifications. These general aspects do not depend on individual sequences and have no role in computational (*in silico*) selection of siRNA, which focuses on the identification of

Corresponding author: Patzel, V. (patzel@mpiib-berlin.mpg.de)

1359-6446/06/$ - see front matter doi:10.1016/j.drudis.2006.11.015

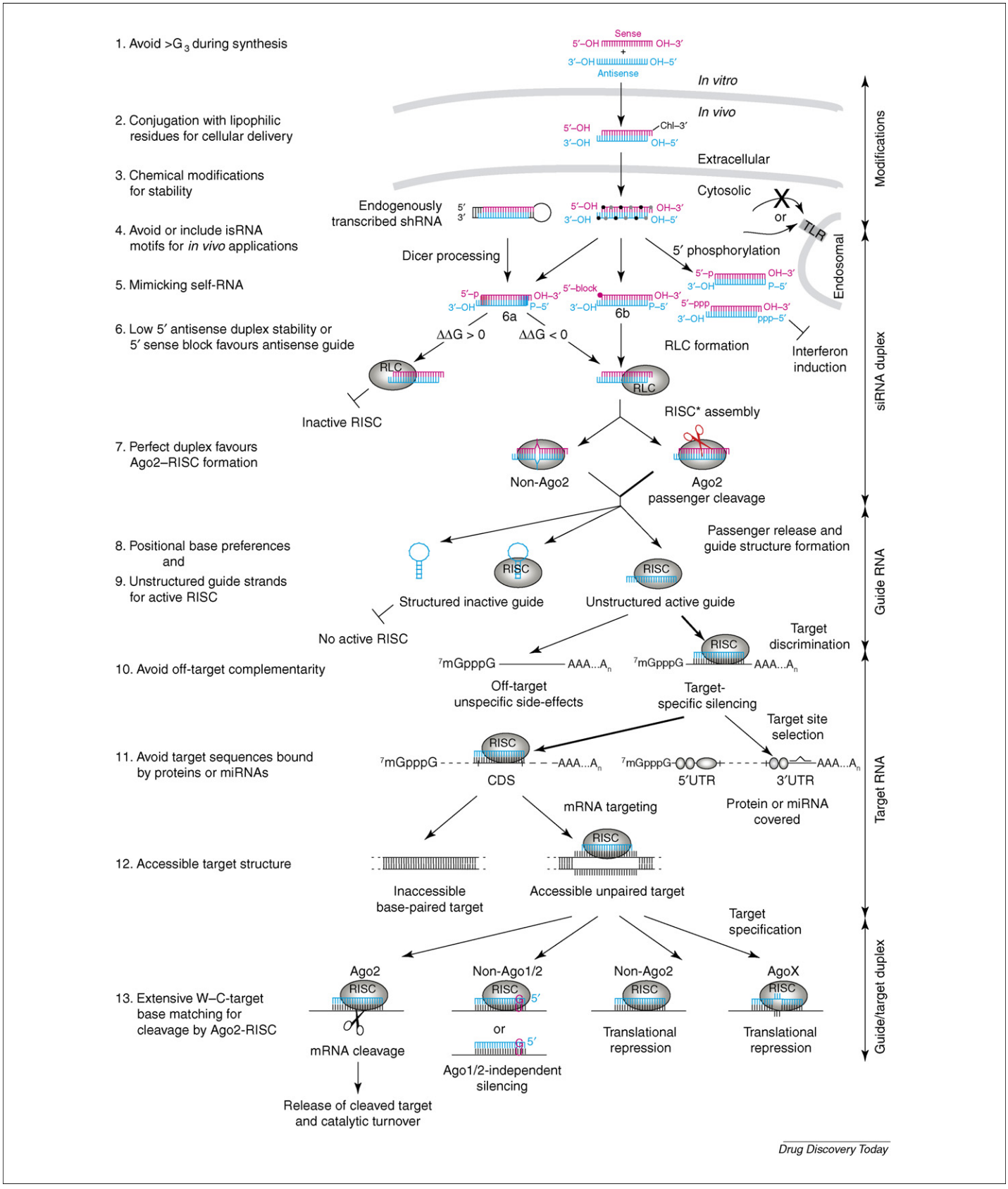

FIGURE 1

Impact of molecular characteristics of siRNA on RNA preparation, delivery and stability, and at different steps along the cellular silencing pathway. Contiguous stretches of more than three G bases ($>G_3$) cause difficulties during RNA preparation and should be avoided (1). Conjugation with lipophilic residues can enhance cellular delivery (2), and chemical modifications of the backbone, ribose and bases can improve siRNA stability in serum and the cytosol (3). Depending on the application and target, immunostimulating RNA (isRNA) motifs should be avoided or included in the duplex (4). Mimicking self-RNA by nucleoside modification,

individual sequences and on structure-related issues. Gene silencing activity and immunostimulation – that is, non-self recognition and activation of the interferon system – are other aspects of siRNA design and are closely associated with specific RNA sequences or structures.

In the development of computational tools for siRNA prediction, empirical and rational approaches have been used to identify selectable characteristics of active siRNA species. The molecular features most relevant to siRNA activity and their meaning for the RNAi mechanism are the focus of this review. Such computable parameters of activity can be weighted and implemented into selection programs. The development of reliable computational design tools is central to the successful use of siRNA in all fields of application.

Hunting for selectable characteristics of active siRNA

The key to the (*in silico*) design of active siRNAs are (computationally) selectable characteristics. Initial experiments focused on the attributes of siRNA duplexes of 19–25 nt with 3′ dinucleotide overhangs characteristic of products generated by RNase-III-like cleavage. The experiments, primarily performed in cell lysates, led to the development of simple conventional guidelines thought to govern the selection of functional siRNAs [11–13] (Box 1). Although this approach did not reliably predict functional siRNAs, and investigators needed to screen at least three or four sequences per target mRNA for functional validation, its parameters have been implemented in many design programs.

Strategies for identifying siRNA selection rules have also adopted lessons learned from antisense technologies with a focus on the accessibilities of local target structures and target binding affinities. Conflicting results have been reported about a putative correlation between siRNA activity and target structure accessibility [14–18]. Furthermore, knowledge about nuclease resistance and cellular delivery has been transferred from antisense oligodeoxyribonucleotide development to siRNA development, including stabilizing modifications and conjugation with lipophilic residues for improved biodistribution and membrane trafficking.

Statistical sequence analyses of randomly selected, active siRNAs have identified distinct base preferences at key positions of siRNAs attributed to functional gene silencing [19–22]. With respect to the silencing process, it is often unclear which of the two strands (sense, antisense, or both) must incorporate these identified base preferences. Furthermore, on the level of individual sequences, there is only limited agreement on these positions, possibly because the data sets used are too small to ensure statistical significance or because other parameters are also crucial for activity [23,24].

More rational approaches have led to the identification of the most reliable parameters related to siRNA activity. Some of those approaches have addressed the issue of why the two strands of siRNA duplexes are not equally eligible for assembly into the RISC [25,26]. The frequently observed strand bias has been found to be caused by thermodynamic asymmetries of the siRNA duplex. The design of duplexes with lower stability at the 5′ antisense terminus, as compared with the 5′ sense terminus, favours the formation of an antisense RISC and gene silencing activity. It is likely, however, that successful RNAi depends on parameters more complex than base positions.

A powerful method of identifying complex traits in data sets is provided by artificial neural networks (ANNs). For example, an ANN has been successfully trained to predict functional siRNA by using a data set of >2000 randomly selected siRNA sequences directed against 34 targets [27]. However, this approach is based exclusively on sequence data sets and RNA structures are not considered. By contrast, another approach has addressed whether intramolecular structures of the siRNA antisense (guide) strand have an impact on gene silencing [23]. Computational analyses of RNA guide strand secondary structures revealed a linear correlation between the length of free (unpaired) 3′ ends and RNAi. Extrapolation of this quantitative structure–activity relationship led directly to completely unstructured guide strands, which can be regarded as 21-nt free 3′ ends, although no unstructured guide strand was included in the original set of data [23].

Empirical and rational approaches, in addition to knowledge transferred from antisense technologies, have disclosed various siRNA-linked parameters that can significantly influence RNA preparation, cellular delivery, intracellular stability, specificity and silencing activity.

BOX 1

Guidelines for conventional siRNA design

In conventional siRNA design, 21-nt siRNAs forming 19-bp duplexes with 2-nt 3′ overhangs are produced. These are directed against the AA(N_{19})TT motif (where *N* is any nucleotide) within the coding sequence of the mRNA target. Alternatively, *NA*(N_{21}) motifs are chosen. The N_{19} defines the 19-bp core duplex and AA, TT and *NA* represent the overhangs [13].

Non-coding regions and the sequence following the start codon (~75–100 bases) are avoided to prevent the targeting of regions of mRNA occupied by translational or regulatory proteins or of regions that are potentially polymorphic. A G+C content of between 30 and 70% is tolerated and a G+C content of 50% is regarded as most favourable. Lastly, siRNAs with homology to sequences other than the target message are excluded to ensure target specificity.

Molecular features of active siRNA

Here, I discuss the most relevant molecular features of active siRNA and their relevance for different steps along the silencing pathway (Figure 1).

usage of 2-nt 3′ overhangs or Dicer substrate siRNA, or avoidance of 5′ triphosphates facilitates entry into sequence-specific silencing pathways (5). Low stability of the 5′ end of the antisense duplex ($\Delta\Delta G < 0$) and blocking the 5′ end of the sense strand specify antisense siRNAs as guide strands during RISC-loading complex (RLC) formation (6), and perfect duplex base matching favours Ago2–RISC formation (7). Positional base preferences (8) and unstructured guide strands (9) correlate with siRNA activity. Sequences with off-target complementarity can cause off-target silencing and should be avoided (10). Accessible target sequences (11) (namely, the coding region) and accessible local target structures (12), such as loops, bulges or joints, might favour RNAi. Extensive W–C target-base matching is required for mRNA cleavage by Ago2–RISC (13). Imperfect target-base matching and/or non-Ago2–RISCs result in translational repression. Wobble pairing with the target at 5′ terminal regions of the guide strand might lead to Ago1- or Ago2-independent silencing. Numbering matches that in the text and in Table 1. Abbreviation: TLR, Toll-like receptor.

(1) Avoidance of sequence motifs interfering with RNA synthesis and purification

Guanine-rich RNA sequences can form Hoogsten-paired quartets of G residues, or so-called 'tetrads', which give rise to unusual structural characteristics and biological activities [28]. Oligomeric sequences containing consecutive stretches of more than three G bases can form G-tetraplexes consisting of four tetrads. These G-tetraplexes can lead to the formation of stable tetrameric aggregates that cause difficulties during synthesis and purification [29]. To avoid these difficulties and unexpected biological activities, siRNA duplexes containing contiguous stretches of more than three G bases should be avoided.

(2) Modifications for enhanced cellular uptake

Owing to the large molecular weight and negative charge of siRNA duplexes, effective cellular uptake and intracellular delivery represent key challenges to using RNAi *in vivo*. To enhance cellular uptake, siRNAs have been conjugated with lipophilic derivatives of cholesterol, lauric acid or lithocholic acid [30]. Although modifications at the 5′ end of the antisense strand impair silencing, 5′ sense, 3′ sense and 3′ antisense modifications do not or only marginally disturb silencing activity [31]. Conjugation with lipophilic residues at the 3′ end of the sense strand are recommended for systemic administration of siRNA *in vivo*.

(3) RNA stabilizing modifications

Unmodified siRNA is prone to nuclease-mediated degradation in serum and the cytosol, which has a negative impact on its activity. Chemical modifications of the phosphate backbone (e.g. phosphorothioate linkages), the ribose (e.g. locked nucleic acids, 2′-deoxy-2′-fluorouridine, 2′-*O*-methyl), and/or the base (e.g. 2′-fluoropyrimidines) increase the resistance of siRNA to nuclease [32,33]. Such modifications are recommended for siRNA that is used *in vivo*. Hydrolysis mediated by ribonuclease A and spontaneous hydrolysis are both accelerated at pyrimidine-A dinucleotides, and are influenced by hydrogen bonds and neighbouring nucleotides [34]. Avoiding such motifs can slightly stabilize unmodified siRNA, although not to the extent of chemical modifications.

(4) Inclusion or avoidance of immunostimulatory sequence motifs

Experimental evidence suggests that gene silencing activity and immunostimulation have to be treated as independent functions of RNA oligonucleotides [35,36]. Immunorecognition of RNA depends on length, distinct sequence motifs, single- versus double-strand configuration, and nucleoside modifications. Duplexes of less than 30 nt that comply with the typical design criteria of siRNA are short enough to evade immunorecognition by cytosolic double-stranded RNA (dsRNA) receptors, but are too long to bypass sequence-dependent recognition through Toll-like receptor 7. Recognition of the motif 5′-GUCCUUCAA-3′ by Toll-like receptor 7 triggers the induction of type I interferons and non-specific downregulation of gene expression [35]. Other cytokine-inducing motifs are 5′-UGUGU-3′ [37] and tetrad-forming poly(G) stretches [38,39]. Systematic analyses of sequence motifs are expected to identify additional motifs with even higher activity, and comprehensive information on sequence-dependent immunological activity will allow us to predict the immunostimulatory properties of any given RNA oligonucleotide in the near future (G. Hartmann, pers. commun.).

Depending on the application, either separating or combining gene silencing and immunostimulation in a single RNA molecule might be desirable. For functional target validation *ex vivo*, strong silencing is required and immunostimulating RNA motifs might be tolerated. Conversely, for many *in vivo* studies aimed at inhibiting specific gene expression, immunostimulating RNA motifs have to be excluded. Lastly, combining silencing and immunostimulation might represent a powerful tool for treatment of viral infection and cancer [36].

(5) Mimicking self-RNA

The immune system distinguishes between self and non-self RNA molecules. Whereas immunostimulation by short RNA depends on sequence motifs [35,37], immunological activity is less sequence-specific for long RNA molecules such as mRNA [40]. Experimental evidence suggests that the immunological activity of long self-RNA in mammals is either masked by naturally occurring nucleoside modifications or prevented by compartmentalization [35,41]. This observation has led to the idea that mimicking self-RNA by introducing nucleoside modifications might avoid any unwanted immunostimulatory activity of siRNA [37]. Nucleoside modifications of the siRNA antisense strand abrogate both the immunological and the silencing activity of the duplex, whereas modification of two nucleosides in the siRNA sense strand can suffice to inhibit the immunological activity coming from either strand without impairing the silencing activity. Thus, masking immunostimulating motifs by nucleoside modification represents a promising strategy to suppress immunostimulation by siRNA.

Recently, an additional structural basis for discriminating between self and non-self dsRNA in mammalian cells has been reported [42]. This study shows that for self recognition siRNAs require the 2-nt 3′ overhangs characteristic of Dicer products; siRNAs lacking these overhangs lead to non-self recognition mediated by the RNA helicase RIG-1, resulting in activation of the interferon system. This finding has only a minor impact on current siRNA design because most siRNAs used are duplexes with 2-nt 3′ overhangs (conventional siRNAs) that conform to this rule for self recognition. An earlier study, however, reported that siRNAs of 25–30 nt that are substrates of Dicer but are not immunostimulatory can be up to 100-fold more potent than conventional 21-nt siRNAs, indicating that the direct linkage between the endogenous production of siRNA with 2-nt 3′-overhangs and RISC formation represents a promising strategy for siRNA design [43]. Most recently, it has been demonstrated that 5′-triphosphate RNA is the ligand for RIG-I-mediated non-self recognition, indicating that there is more to cytoplasmic RNA recognition than long dsRNA [44].

(6) Specification of guide and passenger strand

(a) Thermodynamics of siRNA duplexes determines choice of guide and passenger strand

Thermodynamic asymmetries of the siRNA duplex can favour formation of the sense or antisense RISC, and thereby influence silencing activity. The absolute and relative stabilities of the base

pairs at the 5′ ends of the two siRNA strands can be used as a measure of terminal thermodynamic duplex stability [26]. In brief, duplexes with A–U or G•U base pairs at the 5′ end of the antisense strand and G–C base pairs at the 5′ end of the sense strand are preferable.

Alternatively, complete thermodynamic duplex profiles can be monitored [25]. To evaluate such profiles, the Gibbs free-energy values (ΔG) representing the internal average stabilities of subsequences within an siRNA duplex are calculated. Lower thermodynamic duplex stabilities (higher ΔG value) at the 5′ end of the antisense strand (ΔG_{as}) as compared with the 5′ end of the sense strand (ΔG_s) favour the selection of antisense siRNAs as guide strands, and thus the formation of a silencing-competent antisense RISC [25,26]. Given that $\Delta\Delta G = \Delta G_s - \Delta G_{as}$, duplexes with $\Delta\Delta G < 0$ should be selected.

(b) Blocking 5′ phosphorylation specifies sense siRNA as the passenger strand

Incorporation of the sense strand into the RISC competes with the silencing-competent antisense RISC or with miRNA-mediated gene regulation for cellular resources of the silencing machinery, which lessens the efficacy of the silencing. Off-target effects caused by a sense RISC cause further problems. RISC assembly requires a 5′-phosphorylated guide strand and 5′-hydroxyl residues of siRNA are phosphorylated either before or on formation of the RISC in the cytosol [31]. Thus, blocking 5′ phosphorylation of the sense strand by replacing the 5′-hydroxyl group with, for example, an aminopropyl phosphate residue, represents an elegant way to avoid sense RISC formation and all of its associated side-effects [31].

(7) Duplex base-matching influences choice of Ago proteins for RISC

In addition to Ago2, siRNA can recruit other members of the Ago family for RISC formation [45]. These RISCs lead to translational repression – a silencing mechanism that is less efficient than mRNA cleavage mediated by Ago2–RISC. To achieve high silencing activities, it is important to know which parameters determine the choice of Ago protein for RISC formation and thus the silencing pathway.

For some time it was assumed that Ago proteins are chosen randomly according to their cellular abundance. Results obtained with Ago1 or Ago2 knockdown cells suggest, however, that the structure of the effector duplex influences the choice of the Ago protein for the RISC [23]. Silencing induced by distinct mismatched duplexes has been found to be Ago1-dependent, whereas that induced by perfect Watson–Crick (W–C) base-paired duplexes is dependent on Ago2. Thus, on the basis of effector duplex base matching, researchers can choose between the Ago1-dependent translational block (by using a mismatched duplex) and the formation of Ago2–RISC (by using a perfect duplex), which is required but not sufficient for more powerful Ago2-dependent mRNA cleavage.

(8) Positional base preferences correlate with siRNA activity

Statistical analyses have revealed that positional base preferences at key positions and composition features correlate with siRNA activity [19–22]. These parameters might contribute to efficient processing at different steps and have been included in various design algorithms (Box 2).

BOX 2

Base preferences at key positions correlate with siRNA activity

Several base preferences at key positions and composition features have been reported to correlate with siRNA activity [19–22], including an A or U at the antisense 5′ end, a G or C at the sense 5′ end, and at least five or three A or U bases in the 3′ terminal third or at positions 15–19, respectively, of the sense strand.

Furthermore, an A at position 3, a U at position 10, a base other than G at position 13, and an A or bases other than G or C at position 19, each with respect to the 5′ end of the sense strand, have been described to correlate with RNAi. In addition, the absence of GC stretches of more than 9 nt and a G+C content of 30–52% have been regarded as favourable.

(9) Guide structure has an impact on silencing strength

In antisense RNA-mediated gene suppression, structures of antisense RNAs have a crucial role in both the efficient interaction with and translational block of mRNA targets [46]. In RNAi, antisense siRNAs guide the RISC to the target message, thereby mediating gene silencing. Although favourable sequences for such guide RNAs were investigated early on, their corresponding structures have not been considered for some time. Recently, my colleagues and I [23] reported a strong inverse correlation between the degree of secondary structure formation in guide RNA and gene silencing by siRNA. We found that unstructured guide strands that completely lack complementary bases or in which internal base-pairing is thermodynamically unlikely confer the strongest silencing, whereas structures with base-paired ends are inactive. To derive maximum benefit from the guide structure concept, completely unstructured guide strands should be selected (Box 3).

(10) Avoidance of off-target complementarities

Partial complementarity of as few as 11 contiguous complementary nucleotides between RISC-associated guide strands (sense or

BOX 3

RNA secondary structure: prediction and value

RNA structures incorporate integrated information about the underlying sequences. As a result, structures often represent a more comprehensive basis for investigating and understanding the molecular functions associated with biopolymers than do sequences. Statistically, ~6200 different 21-nt sequences can form the same secondary structure [58], indicating that a single structure–function relationship can cover a plenitude of sequence–function relationships.

Secondary structures of RNA can be predicted by using *in silico* tools for thermodynamic structure prediction. These tools are based on the assumption that the lowest free-energy structure (the minimum free-energy structure) is the most probable fold; however, higher energy (suboptimal) folds can also exist and can be functionally relevant. Minimum free-energy structures and suboptimal folds can be predicted by mfold, RNAfold and other related programs [60,61]. These tools do not include structure search and alignment functions. For directed selection of active guide RNA structures, an automated search program has been developed [58]. To obtain useful secondary structure predictions of mRNA targets, it is recommended that the whole target sequence is analyzed and that suboptimal folds are considered [62].

BOX 4

RNA sequence alignments

The S–W alignment and the BLASTn alignment of the National Center for Biotechnology Information are frequently used for comparisons of RNA sequence. The S–W dynamic programming sequence alignment algorithm is most suitable for addressing the aspects of imperfect and suboptimal alignments of short siRNAs with mRNA targets [63]. BLASTn uses a more heuristic algorithm and sacrifices sensitivity to gain speed [64].

When using word size seven, as recommended for BLASTn searches for short, near-exact matches, ~6% of all alignments for 21-nt RNAs with three mismatches are missed [65]. Thus, a significant fraction of sequences filtered by BLASTn are not sufficiently unique to avoid off-target activity. As a consequence, BLASTn might give acceptable results for some *ex vivo* applications of siRNA, but is unsuitable for designing siRNAs for functional target validation and applications *in vivo*. Notably, WU-BLASTn, which provides gapped alignments with statistical significance estimates, represents a fast and sensitive alternative to classical ungapped BLASTn [66].

antisense) and mRNAs other than the target mRNAs – so-called 'off-targets' – can lead to non-target silencing and undesirable side-effects [47]. siRNA sequences with off-target complementarity or homology to the transcriptome of interest can be identified and excluded on the basis of sequence alignments (Box 4). In particular, seed matches in 3′ untranslated regions (3′ UTRs) – in other words, sense or antisense siRNAs with matches in the hexamer or pentamer seed region (positions 2–7 or 2–8) to the 3′ UTRs of off-targets – must be avoided [48].

Minimization of off-target complementarity is recommended for functional target validation and *in vivo* applications of siRNA. Notably, for statistical reasons, seed matches can be strongly reduced but cannot be completely avoided on the basis of sequence alignments. Chemical modifications, such as 2′-*O*-methyl ribosyl substitutions at position 2 of the guide strand, represent a promising option to reduce off-target activity associated with unavoidable seed matches [49]. If siRNA duplexes with sense strands blocked at the 5′ end are used, sequence alignments and chemical modifications can be confined to the antisense strand.

(11) Avoidance of target sequences bound by regulatory proteins or miRNAs

In conventional siRNA design, untranslated target regions are avoided. More recent studies have shown, however, that siRNA can repress the expression of an mRNA target with partially complementary binding sites in its 3′ UTR in a way that is clearly distinct from siRNA-mediated mRNA cleavage and is much like the function of miRNAs [50]. Nevertheless, the functionality of target sites within the 3′ UTR can be influenced by direct or nearby binding of miRNAs and proteins [51]. Additional, alternative poly(A) sites could make target sites within 3′ UTRs unviable. Thus, coding sequences and 3′ UTRs can be targeted, although 3′ UTRs should be targeted with some reservations.

TABLE 1

Design recommendations for *in vitro*, *ex vivo* and *in vivo* applications of siRNA[a]

Criteria[b]	*In vitro*	*Ex vivo*	*In vivo*
(1) Avoidance of more than three contiguous G bases	Yes	Yes	Yes
(2) Modifications for enhanced cellular uptake	No	No	Yes
(3) RNA stabilizing modifications	No	Yes[c]	Yes
(4) Inclusion or avoidance of immunostimulatory motifs	No	Yes[d]	Yes
(5) Mimicking self-RNA	Yes[e]	Yes[e]	Yes
(6) Specification of guide and passenger strand	Yes	Yes	Yes
(7) Duplex base matching for Ago2–RISC formation	Yes	Yes	Yes
(8) Positional base preferences	Yes	Yes	Yes
(9) Unstructured guide strand	Yes	Yes	Yes
(10) Avoidance of off-target complementarity	No	Yes[f]	Yes[g]
(11) Avoid protein and miRNA bound target sequence	Yes[h]	Yes[h]	Yes[h]
(12) Accessible target structure	Yes	Yes[i]	Yes[i]
(13) Extensive W–C target-base matching[j]	Yes	Yes	Yes
SSE[k]	Yes	Yes	Yes
esiRNA	No	Yes	No

[a] *In vitro* applications include siRNA validation, *ex vivo* applications include knockdown and target validation, and *in vivo* applications include target knockdown.
[b] Numbering matches that in the text and in Figure 1.
[c] Recommended for prolonged activity, use in primary cells, and downregulation of stable gene products.
[d] Recommended for some immunological assays.
[e] Only 2-nt 3′ overhangs are recommended; nucleoside modifications are not required.
[f] I recommend S–W alignments; however, BLASTn can be used.
[g] S–W alignments are strictly recommended. Additional chemical modifications can also be considered.
[h] If guide-strand-related features are very promising, 3′ UTRs can also be targeted.
[i] Only if guide-strand-related criteria remain fulfilled.
[j] To identify novel attributes associated with RNAi, imperfect target-base-matching guide strands and non-W–C base-pairs can be considered.
[k] Sequence space expansion; recommended if no promising candidates can be otherwise predicted.

BOX 5

Comparison of siRNA activity predictions

To compare the functionality of siRNAs predicted with different tools, the so-called '*F* value' has been defined [19]. *F* values reflect the reduction of mRNA levels achieved within 24–48 h of transfecting cells with 100 nM siRNA, a concentration at which most siRNAs reach their silencing potential. To differentiate further among groups of highly active siRNAs, lower RNA concentrations can be used. *F* values strongly depend on the experimental conditions; in other words, on the concentration of siRNA used.

Another value for comparing siRNA prediction tools is the correlation or Pearson coefficient, *r*. This describes the linear correlation between siRNA activity, represented by the relative knockdown or IC_{50} value (inhibitory concentration at which 50% knockdown is observed), and a predictable physical parameter or score calculated by the algorithm.

(12) Thermodynamic accessibility of target structure favours RNAi

In principle, unstructured local sites within the mRNA target are more accessible for hybridization with complementary sequences than are base-paired regions. Whereas the activity of an antisense oligodeoxyribonucleotide is known to depend strongly on the accessibility of the target structure, there is only partial agreement on the meaning of target structures in RNAi [14–18] and target accessibilities are not suitable to predict and to understand siRNA activity with single-nucleotide resolution [23]. Thus, comprehensive target accessibility predictions can support siRNA design but are rated lower than other selection criteria (Box 3).

(13) Extensive W–C target-base matching is required but not sufficient for mRNA cleavage and catalytic turnover

In vitro and *in vivo* studies have shown that an miRISC can function as a siRNA-associated RISC (siRISC) to cleave the target RNA; similarly, an siRISC can function as an miRISC to repress translation [50,52,53]. In humans, most potent gene silencing is mediated by mRNA cleavage and requires both an Ago2–RISC and a guide RNA with extensive complementarity to the target sequence [54]. Non-Ago2–RISCs containing perfectly target-base matching guide sequences and Ago2–RISCs containing imperfectly target-base matching guide sequences will not be competent for mRNA cleavage.

Furthermore, wobble pairing with the target can influence the specificity and activity of the siRNA. However, conflicting results have been reported on the positional effects of wobble pairing [23,55,56], which can be explained by the fact that the same base exchanges that result in wobble paring can also influence the guide structure. Consequently, guide structures should be monitored to be able to distinguish between guide-structure-related effects and positional effects of target-base wobbling. For siRNAs with active guide structures, wobble pairing with the target impairs silencing and possibly leads to Ago1- or Ago2-independent silencing in 5′ terminal regions of guide strands, but is tolerated in central positions [23].

Most of the parameters discussed in this section are computable (1, 4, 6a, 8–13), others (2, 3, 5, 6b, 7) are exclusively subject to chemical synthesis. An alternative strategy, which does not require any kind of design, uses endoribonuclease-prepared siRNA (esiRNA) [57]. These RNAs are mixtures of target-specific siRNAs in which individual sequences including those with off-target complementarity do not reach the concentrations required for off-target silencing or immunostimulation. Consequently, esiRNAs have low potential to cause non-specific side-effects [57]. The disadvantage of esiRNAs is that they usually have lower silencing activities as compared with rationally designed single siRNAs. Nevertheless, the use of esiRNAs represents a simple, cost-effective way for functional target validation *in vitro*.

In silico selection of siRNA

Many of the criteria and guidelines described above have been incorporated into public and commercial design tools that facilitate the selection of potent siRNAs. Consideration of all of the parameters associated with siRNA activity might mean that the stringency of selection is too high to identify any candidate sequence. In particular, analyses of short target sequences can easily suffer from this restriction. As a result, criteria have to be weighted and those of minor importance might have to be omitted, depending on target sequence and length. The process of weighting is usually based on empirically selected factors. Expanding knowledge on parameters attributed to functional siRNAs, together with increasing fields of applications (e.g. *ex vivo*, *in vivo*, including or excluding immunostimulation), requires flexible tools for custom-made siRNA design. Some general recommendations for the criteria that might be

TABLE 2

Subset of tools available for *in silico* selection of active siRNA[a]

Name	Design features	Link	Source	Refs
siRNA Selection Server	GSe, Align[b]	http://jura.wi.mit.edu/bioc/siRNA	OS Res	[67]
si*DESIGN* Center	GSe, Align[b]	http://www.dharmacon.com/sidesign/default.aspx	OS	[19]
*S*fold (Sirna module)	GSe, TSt	http://sfold.wadsworth.org/sirna.pl	OS	[68]
BIOPRED*si*	GSe, Align[b]	http://www.biopredsi.org/start.html	OS Res	[27]
siRNA*scout*	GSe, Align[b], GSt, SSE, TSt[c]	http://www.stz-nad.com	CS	[23,58]
siRNArules 1.0	GSe	http://sourceforge.net/projects/sirnarules	OS	[21]

[a] Abbreviations: CS, commercial source; GSe, guide sequence-based design; GSt, guide structure-based design; OS, open source; Res, for academic research only; SSE, sequence space expansion; TSt, target-structure-based design.
[b] Align indicates BLASTn or S–W alignment.
[c] Available on request from the author.

considered for selection depending on the application are given in Table 1.

Expansion of the space of target-specific guide RNAs represents an elegant option to design more target-specific (including more active) siRNAs for selection [23]. The space of guide RNAs can be expanded by performing A to G and C to U base exchanges within guide sequences (the corresponding U to C and G to A exchanges in sense strands must be considered at RNA synthesis). This increases the number of complementary guide siRNAs by more than a factor of 1000 for target sequences with statistical base contents. These base exchanges induce wobble pairing with the target, but preserve target complementarity and thus silencing activity. The computational process of sequence space expansion can be biased towards the generation of active guide sequences and/or structures by means of constraints (V.P., C. Köberle and S.H.E. Kaufmann, unpublished data).

For many design tools, the selection criteria and their weighting functions are not visible. By comparison, some powerful programs identify siRNAs leading to more than 90% or 80% mRNA knockdown (≥F90/≥F80) with frequencies of about 40% or 60%, making them useful tools [19]. The highest described Pearson coefficients thereby range from $r = 0.52$ ($n = 526$) [21] to $r = 0.66$ ($n = 249$) [27] (Box 5). A novel tool that includes guide structure analyses and the sequence space expansion option needs further validation ($r = 0.89$, $n = 17$) [23,58]. A collection of siRNA design tools is listed in Table 2.

Although considerable success has been reported concerning the prediction of siRNA activity, progress in the design of functional short hairpin RNAs (shRNAs) has been rather slow. Like miRNAs, shRNAs are transcribed endogenously, depend on nuclear export, and undergo processing by RNase-III-like endoribonucleases. Because shRNAs enter the cellular silencing pathways at an earlier stage than siRNAs, other steps might become rate limiting, which is probably the reason why approaches trying to transfer siRNA design to shRNAs have failed. However, the largely shared pathway of shRNA and miRNA raises another problem. A recent study has reported fatality in mice owing to competition between shRNAs and miRNAs for limiting cellular factors, resulting in oversaturation of vital cellular pathways [59]. Although the current model argues that siRNAs and shRNAs share later stages of the silencing pathway, emerging evidence indicates that siRISC and miRISC might be distinct types of complex and that structures of RNA effector molecules can bias silencing pathways [23,54]. It is a key challenge of future siRNA and shRNA design to actively avoid interference with the processing and action of cellular regulatory RNA. Another way to suppress this kind of 'RNA interference' in RNAi is to enhance the siRNA or shRNA activity, thereby minimizing the active dose.

A flow chart of rational siRNA *in silico* selection is shown in Figure 2. Initial steps encompass target sequence input, generation of reverse complement and simulation of space of guide sequences. Depending on the application (*in vitro*, *ex vivo*, *in vivo*) and target (cellular, viral, cancer), various parameters of siRNA activity are calculated. For best results, guide-structure-related and guide-sequence-related features need to be considered. Be aware that some published base-preference rules refer to positions of the sense strand and applying them to the guide strand will produce incorrect predictions. The inclusion of target-structure-related issues for selection is rather optional. On the basis of BLASTn or

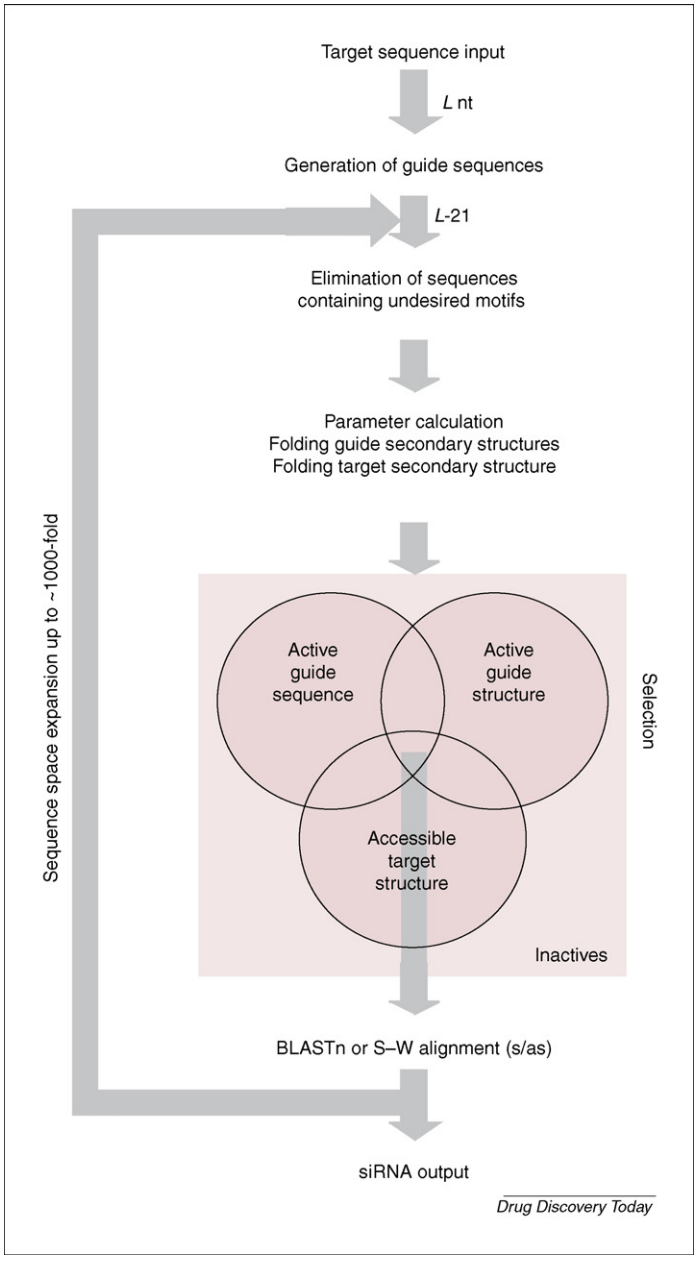

FIGURE 2

Rational siRNA *in silico* selection. The procedure begins with input of the target sequence, which is then reversed and complemented. The antisense siRNA sequence space is generated by subdivision of the reverse complement into 1-nt-shifted overlapping 21-nt siRNAs, which represent the putative guide sequences. Successively, 21-nt siRNAs containing undesirable motifs or their complements are excluded from the procedure. These motifs include those interfering with RNA preparation or immunostimulatory motifs. For the remaining sequences, various sets of parameters are calculated and screened for signatures of active siRNA molecules. These parameters are related to the sequences and secondary structures of the guide strands and to the secondary structure of the mRNA target. Some published base-preference rules relate to the sense strand, and applying them to the guide strand will produce incorrect predictions. High priority should be given to guide RNA and duplex-related issues. Target-structure-related issues are controversial and should be given lower priority. By using BLASTn or S–W analyses, candidate sequences with potential identity to off-targets are identified and eliminated. If no or too few candidates are predicted, more antisense siRNAs can be generated and subjected to selection by using the sequence space expansion option. Lastly, the candidate sequences selected are ranked on the basis of empirical scoring functions. Abbreviations: as, antisense siRNA; s, sense siRNA.

Smith–Waterman (S–W) analyses, sequences with off-target homology and/or complementarity are excluded and selection might be extended to guide strands that form wobble base pairs with the target. Lastly, selected candidate sequences are ranked on the basis of empirical scoring functions.

Conclusion

The vast collection of data emerging from genome-wide siRNA gene knockdown studies, along with rapidly increasing computer power, high-throughput screening assays and novel computational approaches, are paving the way for the development of highly reliable tools for the custom design of functional siRNA for all fields of application. Increasing value needs to be attached to the role of RNA structures that influence RNA silencing in various ways: (i) structures of endogenously transcribed precursors are assumed to be relevant for efficient processing; (ii) structures of the effector duplex have an impact on the silencing pathway; (iii) guide structures and (iv) local target structures influence the strength of silencing; and (v) structures of duplexes between guide and target RNA monitor for target specificity. At present, computational siRNA design markedly reduces the costs of reagents and labour in early stage pharmaceutical research. Improvements in our ability to predict cross-silencing and immunostimulatory activities will facilitate the progression of siRNA-based drugs to clinical trials.

Acknowledgements

I thank Ulrike Jung, Christian Köberle and Stefan H.E. Kaufmann for critical and helpful comments on this review, and Stefan H.E. Kaufmann for support.

Competing financial interest statement

The author declares competing financial interests. The author is avocationally affiliated to the Steinbeis Transfer Center for Nucleic Acids Design (STZ-NAD), which offers services relating to nucleic acids design. As a full-time employee of the Max-Planck-Society, which is an institution of academic research, the author assures that declared interests did not influence the objectivity, integrity or perceived value of this review.

References

1 Zamore, P.D. and Haley, B. (2005) Ribo-gnome: the big world of small RNAs. *Science* 309, 1519–1527
2 Downward, J. (2005) RNA interference libraries prove their worth in hunt for tumor suppressor genes. *Cell* 121, 813–815
3 Valencia-Sanchez, M.A. *et al.* (2006) Control of translation and mRNA degradation by miRNAs and siRNAs. *Genes Dev.* 20, 515–524
4 Elbashir, S.M. *et al.* (2001) Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature* 411, 494–498
5 Matranga, C. *et al.* (2005) Passenger-strand cleavage facilitates assembly of siRNA into Ago2-containing RNAi enzyme complexes. *Cell* 123, 407–420
6 Rand, T.A. *et al.* (2005) Argonaute2 cleaves the anti-guide strand of siRNA during RISC activation. *Cell* 123, 621–629
7 Gregory, R.I. *et al.* (2005) Human RISC couples microRNA biogenesis and posttranscriptional gene silencing. *Cell* 123, 631–640
8 Liu, J. *et al.* (2004) Argonaute2 is the catalytic engine of mammalian RNAi. *Science* 305, 1437–1441
9 Meister, G. *et al.* (2004) Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs. *Mol. Cell* 15, 185–197
10 Harborth, J. *et al.* (2001) Identification of essential genes in cultured mammalian cells using small interfering RNAs. *J. Cell Sci.* 114, 4557–4565
11 Elbashir, S.M. *et al.* (2001) RNA interference is mediated by 21- and 22-nucleotide RNAs. *Genes Dev.* 15, 188–200
12 Elbashir, S.M. *et al.* (2001) Functional anatomy of siRNAs for mediating efficient RNAi in *Drosophila melanogaster* embryo lysate. *EMBO J.* 20, 6877–6888
13 Elbashir, S.M. *et al.* (2002) Analysis of gene function in somatic mammalian cells using small interfering RNAs. *Methods* 26, 199–213
14 Holen, T. *et al.* (2002) Positional effects of short interfering RNAs targeting the human coagulation trigger tissue factor. *Nucleic Acids Res.* 30, 1757–1766
15 Xu, Y. *et al.* (2003) Effective small interfering RNAs and phosphorothioate antisense DNAs have different preferences for target sites in the luciferase mRNAs. *Biochem. Biophys. Res. Commun.* 306, 712–717
16 Vickers, T.A. (2003) Efficient reduction of target RNAs by small interfering RNA and RNase H-dependent antisense agents. A comparative analysis. *J. Biol. Chem.* 278, 7108–7118
17 Heale, B.S.E. *et al.* (2005) SiRNA target site secondary structure predictions using local stable substructures. *Nucleic Acids Res.* 33, e30
18 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
19 Reynolds, A. *et al.* (2004) Rational siRNA design for RNA interference. *Nat. Biotechnol.* 22, 326–330
20 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
21 Holen, T. (2006) Efficient prediction of siRNAs with siRNArules 1.0: an open-source JAVA approach to siRNA algorithms. *RNA* 12, 1620–1625
22 Shabalina, S.A. *et al.* (2006) Computational models with thermodynamic and composition features improve siRNA design. *BMC Bioinformatics* 7, 65–80
23 Patzel, V. *et al.* (2005) Design of siRNAs producing unstructured guide-RNAs results in improved RNA interference efficiency. *Nat. Biotechnol.* 23, 1440–1444
24 Saetrom, P. and Snove, O., Jr (2004) A comparison of siRNA efficacy predictors. *Biochem. Biophys. Res. Commun.* 321, 247–253
25 Khvorova, A. *et al.* (2003) Functional siRNAs and miRNAs exhibit strand bias. *Cell* 115, 209–216
26 Schwarz, D.S. *et al.* (2003) Asymmetry in the assembly of the RNAi enzyme complex. *Cell* 115, 199–208
27 Huesken, D. *et al.* (2005) Design of a genome-wide siRNA library using an artificial neural network. *Nat. Biotechnol.* 23, 995–1001
28 Shafer, R.H. and Smirnov, I. (2001) Biological aspects of DNA/RNA quadruplexes. *Biopolymers* 56, 209–227
29 Pan, B. *et al.* (2004) Synthesis, purification and crystallization of guanine-rich RNA oligonucleotides. *Biol. Proced. Online* 6, 257–262
30 Lorenz, C. *et al.* (2004) Steroid and lipid conjugates of siRNAs to enhance cellular uptake and gene silencing in liver cells. *Bioorg. Med. Chem. Lett.* 14, 4975–4977
31 Chiu, Y.L. and Rana, T.M. (2002) RNAi in human cells: basic structural and functional features of small interfering RNA. *Mol. Cell* 10, 549–561
32 Zhang, H.Y. *et al.* (2006) RNA interference with chemically modified siRNA. *Curr. Top. Med. Chem.* 6, 893–900
33 Morrissey, D.V. *et al.* (2005) Potent and persistent *in vivo* anti-HBV activity of chemically modified siRNAs. *Nat. Biotechnol.* 23, 1002–1007
34 Kierzek, R. (1992) Hydrolysis of oligoribonucleotides: influence of sequence and length. *Nucleic Acids Res.* 20, 5073–5077
35 Hornung, V. *et al.* (2005) Sequence-specific potent induction of IFN-α by short interfering RNA in plasmacytoid dendritic cells through TLR7. *Nat. Med.* 11, 263–270
36 Schlee, M. *et al.* (2006) siRNA and isRNA: two edges of one sword. *Mol. Ther.* 14, 463–470
37 Judge, A.D. *et al.* (2005) Sequence-dependent stimulation of the mammalian innate immune response by synthetic siRNA. *Nat. Biotechnol.* 23, 457–462
38 Krug, A. *et al.* (2001) Identification of CpG oligonucleotide sequences with high induction of IFN-alpha/beta in plasmacytoid dendritic cells. *Eur. J. Immunol.* 31, 2154–2163
39 Sugiyama, T. *et al.* (2005) CpG RNA: identification of novel single-stranded RNA that stimulates human $CD14^{+}CD11c^{+}$ monocytes. *J. Immunol.* 174, 2273–2279
40 Scheel, B. *et al.* (2005) Toll-like receptor-dependent activation of several human blood cell types by protamine-condensed mRNA. *Eur. J. Immunol.* 35, 1557–1566
41 Kariko, K. *et al.* (2005) Suppression of RNA recognition by Toll-like receptors: the impact of nucleoside modification and the evolutionary origin of RNA. *Immunity* 23, 165–175

42 Marques, J.T. *et al.* (2005) A structural basis for discriminating between self and nonself double-stranded RNAs in mammalian cells. *Nat. Biotechnol.* 24, 559–565
43 Kim, D.H. *et al.* (2005) Synthetic dsRNA Dicer substrates enhance RNAi potency and efficacy. *Nat. Biotechnol.* 23, 222–226
44 Hornung, V. *et al.* (2006) 5′-Triphosphate RNA is the ligand for RIG-1. *Science* 314, 994–997
45 Martinez, J. *et al.* (2002) Single-stranded antisense siRNAs guide target RNA cleavage in RNAi. *Cell* 110, 563–574
46 Patzel, V. and Sczakiel, G. (1998) Theoretical design of antisense RNA structures substantially improves annealing kinetics and efficacy in human cells. *Nat. Biotechnol.* 16, 64–68
47 Jackson, A.L. *et al.* (2003) Expression profiling reveals off-target gene regulation by RNAi. *Nat. Biotechnol.* 21, 635–637
48 Birmingham, A. *et al.* (2006) 3′ UTR seed matches, but not overall identity, are associated with RNAi off-targets. *Nat. Methods* 3, 199–204
49 Jackson, A.L. *et al.* (2006) Position-specific chemical modification of siRNAs reduces off-target transcript silencing. *RNA* 12, 1197–1205
50 Doench, J.G. *et al.* (2003) siRNAs can function as miRNAs. *Genes Dev.* 17, 438–442
51 Rajewsky, N. (2006) L(ou)sy miRNA targets? *Nat. Struct. Mol. Biol.* 13, 754–755
52 Hutvagner, G. and Zamore, P.D. (2002) A microRNA in a multiple turnover RNAi enzyme complex. *Science* 297, 2056–2060
53 Zeng, Y. *et al.* (2003) MicroRNAs and small interfering RNAs can inhibit mRNA expression by similar mechanisms. *Proc. Natl. Acad. Sci. U. S. A.* 100, 9779–9784
54 Tang, G. (2005) siRNA and miRNA: an insight into RISCs. *Trends Biochem. Sci.* 30, 106–114
55 Doench, J.G. and Sharp, P.A. (2004) Specificity of microRNA target selection in translational repression. *Genes Dev.* 18, 504–511
56 Holen, T. *et al.* (2005) Tolerated wobble mutations in siRNAs decrease specificity, but can enhance activity *in vivo*. *Nucleic Acids Res.* 33, 4704–4710
57 Yang, D. *et al.* (2002) Short RNA duplexes produced by hydrolysis with Escherichia coli RNase III mediate effective RNA interference in mammalian cells. *Proc. Natl. Acad. Sci. U. S. A.* 99, 9942–9947
58 Köberle, C. *et al.* (2006) Selecting effective siRNAs based on guide RNA structure. *Nat. Protocols* 1, 1832–1839
59 Grimm, D. *et al.* (2006) Fatality in mice due to oversaturation of cellular microRNA/short hairpin RNA pathways. *Nature* 441, 537–541
60 Zuker, M. and Stiegler, P. (1981) Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucleic Acids Res.* 9, 133–148
61 Hofacker, I.L. (2003) Vienna RNA secondary structure server. *Nucleic Acids Res.* 31, 3429–3431
62 Scherr, M. *et al.* (2000) RNA accessibility prediction: a theoretical approach is consistent with experimental studies in cell extracts. *Nucleic Acids Res.* 28, 2455–2461
63 Smith, T.F. and Waterman, M.S. (1981) Identification of common molecular subsequences. *J. Mol. Biol.* 47, 195–197
64 Altschul, S.F. *et al.* (1990) Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410
65 Snøve, O., Jr and Holen, T. (2004) Many commonly used siRNAs risk off-target activity. *Biochem. Biophys. Res. Commun.* 319, 256–263
66 Lopez, R. *et al.* (2003) WU-Blast2 server at the European Bioinformatics Institute. *Nucleic Acids Res.* 31, 3795–3798
67 Yuan, B. (2004) siRNA Selection Server: an automated siRNA oligonucleotide prediction server. *Nucleic Acids Res.* 32, W130–W134
68 Ding, Y. *et al.* (2004) Sfold web server for statistical folding and rational design of nucleic acids. *Nucleic Acids Res.* 32, W135–W141