Chapter 2
RNA Secondary Structure Prediction and Gene Regulation by Small RNAs

Ye Ding

2.1 Introduction

RNA molecules are involved in some of the cell's most fundamental processes that include catalysis, pre-mRNA splicing and RNA editing, and regulation of transcription and translation. To a large degree, the function of a regulatory RNA molecule is determined by its structure. Computational methods for modeling RNA secondary structure provide useful initial models for solving the tertiary structure by crystallography or nuclear magnetic resonance (NMR). The problem of computational prediction of secondary structure for a single RNA sequence dates back to the early 1970s [99]. Free energy minimization has been an important method for such prediction. The partition function approach by McCaskill enables rigorous computation of base-pair probabilities and heat capacity [70]. In recent years, there has been increasing interest in ensemble-base approaches that extend the pioneering work of McCaskill. In this chapter, we briefly review these developments. Gene silencing by RNA interference and posttranscriptional gene regulation by microRNAs are fundamental discoveries in molecular biology. Rational design of short interfering RNAs for improving potency of gene silencing and regulatory target prediction for microRNAs are two important computational problems. We here review work from our group and others to show that target mRNA secondary structure is important for both efficient gene silencing and microRNA target recognition.

Y. Ding (✉)
Wadsworth Center, New York State Department of Health, New York, USA
e-mail: yding@wadsworth.org

Y. Ding
Center for Medical Science, 150 New Scotland Avenue, Albany, NY 12208, USA

J. Feng et al. (eds.), *Frontiers in Computational and Systems Biology*,
Computational Biology 15,
DOI 10.1007/978-1-84996-196-7_2, © Springer-Verlag London Limited 2010

2.2 RNA Secondary Structure Prediction

RNA plays a variety of important functional roles that include catalysis, RNA splicing, and regulation of transcription and translation. These roles are typically carried out at specific RNA structural sites, often through molecular interactions or conformational change. Hence, the function of an RNA molecule is primarily determined by its secondary and tertiary structures. RNA tertiary interactions involve secondary structure elements and are substantially weaker than secondary interactions. Thus, to a large extent, the free energies in secondary structure represent the thermodynamics of RNA folding. The tendency for RNA folding to be primarily driven by secondary structure features is a tremendous advantage for structural and functional studies on RNAs. Furthermore, computational RNA tertiary structure prediction without experimental information is an intractable problem, and the thermodynamics of tertiary interactions have not been well characterized. In addition, RNA secondary structure is well conserved in evolution. For these reasons, computational algorithms have focused on RNA secondary structure prediction in the last several decades. Given an RNA sequence, a secondary structure is simply defined by a list of base pairs, typically Watson–Crick (G•C or A–U) and Wobble G–U. As shown by Fig. 2.1 for a predicted minimum free-energy structure for *Xlo* 5S rRNA, helices and loops of various types represent basic structural elements of RNA secondary structure.

2.2.1 Free Energy Minimization

In structural computational biology, free-energy minimization for prediction of macromolecular folding is a long-established paradigm. It assumes that, at equilibrium, the solution to the underlying molecular folding problem is unique and that the molecule folds into the lowest-energy state. Also, it is implicitly assumed that the free energies of individual structural motifs are additive. This paradigm had been the foundation for prediction of RNA secondary structure for several decades [67, 68, 74, 99, 116]. For RNA secondary structure prediction, free-energy parameters for basic structural motifs are estimated or extrapolated from chemical melting experiments [67, 68, 110]. The discrete optimization problem is ill-conditioned, in that the prediction is sensitive to small changes in the energy parameters [53, 115]. Furthermore, there is substantial uncertainty in the energy parameters, particularly for loops. For these reasons, efficient algorithms have been developed for not only computing the minimum free energy (MFE) structure, but also for generating a heuristic set of suboptimal structures [67, 68, 116]. An alternative approach computes all suboptimal foldings within an energy increment above the MFE [109]. The exponential growth in the number of these foldings motivated the development of the RNAshapes method for the efficient representation of the near-optimal foldings [36]. The complete suboptimal approach addresses the low-energy end of the

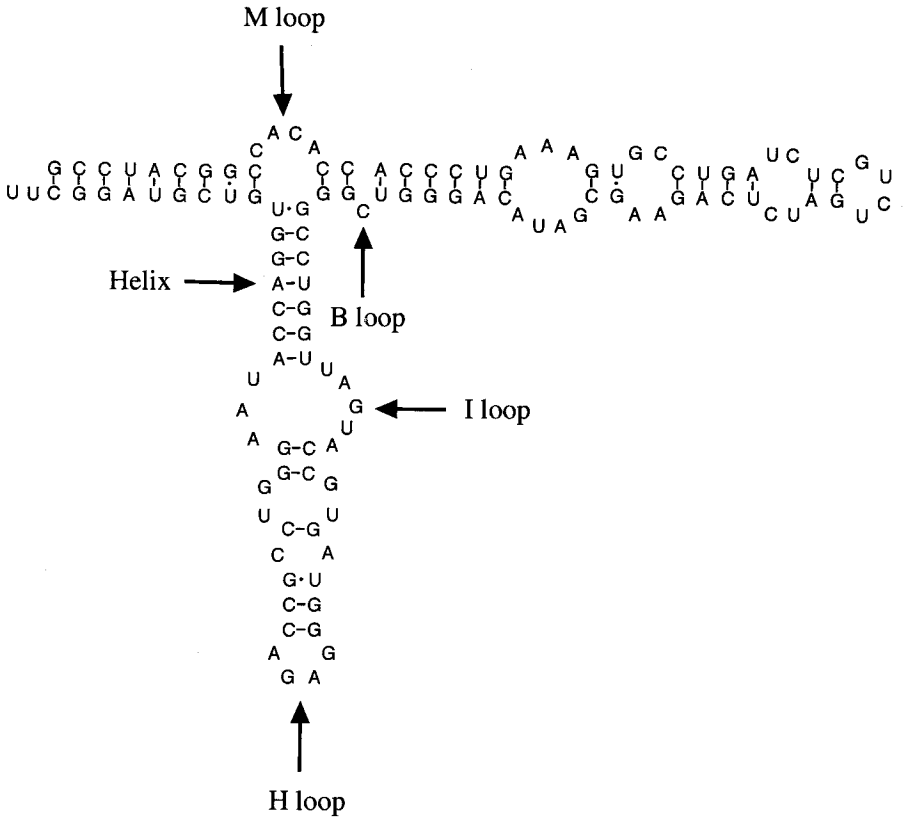

Fig. 2.1 The minimum free-energy structure for *Xlo* 5S rRNA and all types of secondary structural elements: helix (formed by stacked base pairs), bulge loop (B loop), interior loop (I loop), hairpin loop (H loop), and multibranched loop (M loop)

unweighted energy landscape. Neither approach guarantees an unbiased representation of the Boltzmann-weighted ensemble. The free-energy minimization algorithm [116] and the algorithm for computing suboptimal structures [109] have been extended for two or more interacting RNAs [3].

2.2.2 *Partition Function Approach*

In a drastic departure from free-energy minimization, the partition function approach pioneered by McCaskill (1990) [70] laid the foundation for statistical characterizations of the equilibrium ensemble of RNA secondary structures. In particular, base-pair probabilities can be calculated. Similar to its MFE counterpart, the algorithm for computing partition function and base-pair probabilities is cubic and requires quadratic storage. The significance of base-pair probabilities has been

further demonstrated in two studies. For base pairs in the MFE structure, those with higher probabilities have higher predictive accuracy measured by positive predictive value [64]. The positive predictive value is the percentage of base pairs in the predicted structure that are in the structure determined by comparative sequence analysis. Thus, base-pair probabilities provide measures of confidence for MFE predictions. That study was based on an extended partition function algorithm that accommodated coaxial stacking and more recent energy parameters. Furthermore, base-pair probabilities are found to be less affected by uncertainties in energy parameters than is the MFE structure [53]. The McCaskill algorithm has also been extended to include a class of pseudoknots [29, 30]. Like the partition function, the mean and variance (and any moments in general) of the Boltzmann-weighted free-energy distribution can be calculated, and these ensemble characteristics are reported to be useful for distinguishing biological sequences from random sequences [71]. A partition function algorithm for k-point mutants of an RNA sequence has recently been described [17]. For modeling the hybridization of two nucleic acid molecules, the Zuker group was the first to compute partition function and base-pair probabilities [21]. These developments are indicative of a paradigm shift towards ensemble-based approaches.

2.2.3 *Statistical Sampling Approach*

In the traceback step of an RNA folding algorithm, base pairs are generated one at a time according a chosen principle (e.g., energy minimization or probabilistic sampling as discussed below) to form a secondary structure. The long-standing problem of a statistical representation of probable foldings can be addressed by a sampling extension of the partition function approach [24]. In the traceback step, the conditional probabilities computed with partition functions are used to sample a new base-pair or unpaired base(s), given partially formed structure. Thus, the essence of the sampling algorithm is stochastic traceback. The Boltzmann distribution in statistical mechanics gives the probability of a secondary structure I at equilibrium as $\exp[-E(I)/RT]/U$, where $E(I)/$ is the free energy of the structure, R is the gas constant, T is the absolute temperature, and U is the partition function for all admissible secondary structures of the RNA sequence, i.e., $U = \sum_I \exp[-E(I)/RT]$. The sampling algorithm generates a sample of secondary structures in proportion to their Boltzmann probabilities, guaranteeing a statistical representation of the Boltzmann-weighted ensemble.

A statistical sample of the ensemble allows sampling estimates of the probabilities of any structural motifs, from the simplest elements of base pair and unpaired base, to loops of various types, to more complex structures consisting of stems and loops that may be of special interest in a given application. In particular, probability profiling of single-stranded regions in RNA secondary structure is directly applicable to the rational design of mRNA-targeting nucleic acids [22–26]. The Boltzmann-weighted density of states (BWDOS) [24] characterizes the weighted

energy landscape, whereas a density-of-states algorithm [19], applicable only to short sequences, describes the *unweighted* landscape. A structure sample can also be used for computation of other characteristics of the Boltzmann ensemble. For example, the mean and the variance of the free-energy distribution can be estimated by a sample, whereas exact calculations require laborious algorithm development [71]. In principle, a sampling extension can also be developed for a partition function algorithm including pseudoknots. In this case, base-pair probabilities can be estimated by a sample, and the estimates should closely approximate those computed by a high-order algorithm [30].

A sample of moderate size drawn from the ensemble of an enormous number of possible structures is sufficient to guarantee statistical reproducibility in the estimates of typical sampling statistics. The reproducibility is best demonstrated when two independent samples do not have a single structure in common [24, 28]. These seemingly surprising observations are fully expected for an exact sampling algorithm. The sampling algorithm is the basis of the Sfold RNA software package [26] and has been implemented into other RNA folding software including UNAfold [63], Vienna RNA package [42], and RNAstructure [65]. Sampling was adapted for probabilistic representation of structure shapes for RNA sequences of moderate length or longer [107]. A method has been presented to speed up the sampling step [80].

2.2.4 Cluster and Centroid Representation of Boltzmann Ensemble

In the sampled ensemble, distinct structural clusters were observed [24]. This observation suggested that the Boltzmann ensemble could be efficiently represented by clusters. Automated clustering procedure and tools have been developed for this purpose [13, 27, 28]. The procedure returns three to four clusters on average. Another advantage of clustering is that the centroid structure, as the single best representative of the cluster, can be easily identified with little computational cost. The centroid of any set of structures is defined as the structure in the whole ensemble that has the shortest total distance to structures in the set. For the base-pair distance between two structures, the centroid is simply the structure formed by all base pairs having a frequency > 0.5 in the structure set [27]. The clusters, together with their probabilities (estimated by frequencies in the sample) and their centroids, present a complete and efficient statistical characterization of the Boltzmann ensemble. Similar to the reproducibility of ensemble-level sampling statistics [24], the clusters and centroids are also statistically reproducible from one sample to another, even when the two independent samples do not share a single structure [28]. The centroid of the sampled ensemble and the best cluster centroid provide alternative structural predictions. It was a surprising finding that these predictions are substantially improved over the minimum free energy predictions [27], a result that further validates ensemble-based approaches. The idea of centroid has generated considerable interest. Generalized centroid estimators for bioinformatics problems in particular RNA secondary structure prediction have been proposed [12, 38].

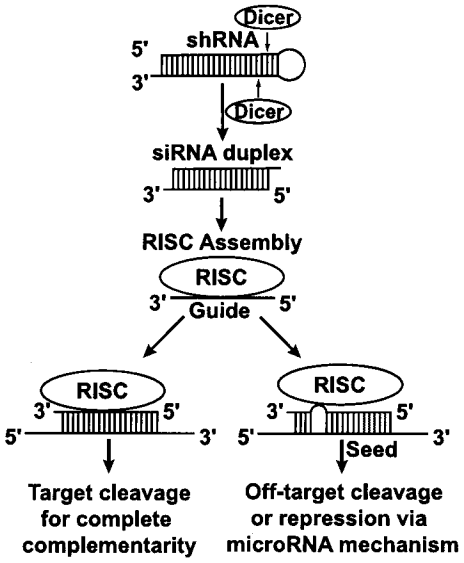

Fig. 2.2 Post-transcriptional regulation by shRNAs or siRNAs. An shRNA (with a typically 19–29 bp stem) can be processed by Dicer into an siRNA. The guide strand in the assembled RISC guides target recognition by complementary base-pairing. Target cleavage by RNAi machinery is triggered by perfect complementarity. Partial complementarity can induce off-target mRNA cleavage or repression of gene expression via microRNA pathway, for which the seed base-pair match (in red) involving nt 2 to nt 7 or 8 of the 5′ end of the guide strand is reported to be important

2.3 Gene Silencing by Small Interfering RNAs

RNA interference (RNAi) is a sequence-specific gene silencing mechanism that is induced by double-stranded RNA (dsRNA) homologous to the target gene [35]. RNAi can be mediated either by small interfering RNAs (siRNAs) of about 21 nt with two-nucleotide 3′ overhang [33], or by stably expressed short hairpin RNAs (shRNAs) that are processed by Dicer into siRNAs [9, 76]. During activation of the RNA-induced silencing complex (RISC), the guide (antisense) strand of the siRNA duplex is preferentially assembled into the RISC when the stem formed by the 5′ end and its complement is less stable than the one formed by the 3′ end and its complement [49, 90]; the "passenger" (sense) strand is cleaved by Argonaute2 (Ago2), the catalytic component of RISC [69, 82]. The antisense strand guides Ago2 to cleave mRNA by perfect base-pairing with the complementary site in the target (Fig. 2.2). In comparison with antisense oligos or trans-cleaving ribozymes for gene knockdown, RNAi generally offers greater potency and target specificity. As the method of choice for loss-of-function studies in mammalian systems and drug target validation, RNAi has revolutionized basic biology study and drug discovery research. In addition, novel RNAi-based therapeutic agents for treating a variety of human diseases have been under development, most notably by Alnylam Pharmaceuticals.

2.3.1 Design Rules for Improving Potency

Large variation in the efficiency of siRNAs for different sites on the same target is commonly observed [43]. Usually, only a small proportion of randomly selected

siRNAs are potent. Thus, there has been a great interest in determining rules for the improvement of RNAi design. A number of empirical rules on siRNA duplex features have been reported. These include the asymmetry rule for siRNA duplex ends, which requires that the 5′ end of the antisense strand forms a stem with its complement that is less stable than the stem formed by the 5′ end of the sense strand [49, 90]. The asymmetry rule is strongly related to the requirements of high A/U content at the 5′ end of the antisense strand and high G/C at the 5′ end of the sense strand [84, 101]. A number of position-specific nucleotide preferences and other siRNA sequence features have been proposed [78, 84]. In addition, the importance of target secondary structure and accessibility has been suggested by several studies based on computational modeling of target structure and accessibility [41, 50, 51, 61, 62, 89, 94, 97] and was supported by compelling evidence based on experimentally assessed accessibility [2, 7, 54, 75, 106, 108].

2.3.2 *Structure Based Assessment of Target Accessibility*

A number of approaches have been published for quantifying target site accessibility for rational design of RNA-targeting nucleic acids. Based on target structures predicted by RNA folding algorithms, these methods are either probabilistic or energetic. Probabilistic methods assess the probability that a base or a block of bases is single stranded [23, 70, 73], whereas energetic methods model the energy exchanges of the hybridization process [59–61, 66, 93–95], arguably providing more refined measures of accessibility. For example, consider two target sites with (nearly) equal probability of being single stranded. If one site has high AU and the other has high GC, then the energetic costs for disrupting the target structure and the stabilities of the hybrid could be quite different for the two sites. In data analysis for some of our studies, energy measures were observed to give improved correlations than probabilistic measures. Thus, our efforts have focused on energetic models. Below, we briefly discuss several major methods.

The Sfold structure sample [24, 26] allows computation of both probabilistic measures [23] and energetic measures of target accessibility [59, 60, 93–95]. It is well established that a single-stranded block of 4–5 nts can facilitate the nucleation step of the hybridization [40, 112]. Thus, a moderate structure sample is sufficient for revealing potential effective sites by using block size of 4 nts for accessibility profiling [23]. The major advantage of using the structure sampling algorithm is that the time-consuming partition-function calculation for the whole target sequence only needs to be computed once. Folding constraints such as maximum nucleotide distance L for two bases to form a pair can be imposed for "local" folding. Such local folding was found to be significant for prokaryotic applications [93]. For prokaryotes, transcription and translation are tightly coupled events so that the target mRNA is unlikely to be able to fold globally. In contrast, eukaryotic mRNAs are first transcribed in nucleus and then transported to cytoplasm where they can conceivably fold globally before they engage in interactions with other molecules

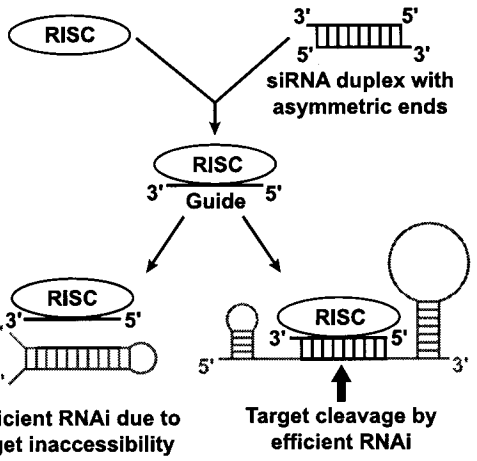

Fig. 2.3 A proposed simple model for efficient RNAi. RISC assembly is facilitated by asymmetric ends of siRNA duplex; target recognition via intermolecular base-pairing is aided by structural accessibility at the target site. The combination of the upstream effect of duplex asymmetry and the downstream effect of target accessibility is generally essential for potent gene silencing

in the cytoplasm for regulation of gene translation. Global folding using Sfold sampling algorithm can reveal highly unstructured sites that are well "conserved" in the likely mRNA structure population. These well-predicted sites can be valuable for the selection of effective target sites.

Target site disruption energy, $\Delta G_{disruption}$, is the energy cost of local disruption of the mRNA structure so that the binding site becomes completely single stranded [94]. A largely single-stranded (i.e., structurally accessible) site does not require substantial structure alteration for the guide siRNA strand to bind to the target. $\Delta G_{disruption}$ is a quantitative measure of the structural accessibility at the target site and is calculated based on target secondary structures predicted by Sfold [26] to address the likely population of mRNA structures. We found in data analysis, as illustrated by Fig. 2.3, that target accessibility is an important determinant of RNAi activity and the asymmetry of siRNA duplex asymmetry is important for facilitating RISC assembly [94]. We also found that the commonly observed negative effect of high siRNA GC-content on RNAi potency is due to generally poor target accessibility for a high GC target site which is likely to have stable secondary structure [14], rather than the likelihood that the high GC siRNA guide strand may form stable intramolecular secondary structure as previously suggested [78].

An alternative to the local disruption assumption is the global disruption model. For this model, as a result of siRNA:mRNA hybridization, the base pairs outside the target site can be rearranged so that the mRNA adopts a new globally altered structure. In this case, the free energy of the target secondary structure after siRNA binding must be recalculated by refolding the mRNA with the binding site constrained to be unpaired. This constraint option has been implemented in Sfold and available through the Sfold web server [26]. However, refolding will cost a hefty computational price. This global model is essentially equivalent to an approach based on exact calculation of ensemble free energies from initial folding and refolding [61]. This approach makes the assumption that the target will reestablish structure equilibrium after siRNA binding. The analysis of siRNA datasets in our study suggests that target cleavage by RNAi machinery appear to be rather rapid so that the target

may not have time to refold before cleavage [94]. While this issue warrants further investigation, data analysis using ensemble energies also confirmed the importance of target secondary structure in RNAi activity [61].

An extension of the McCaskill algorithm [70] can compute the probability that a block of nucleotides is single stranded [73]. However, for each block, this extension requires recomputation of the partition functions for the entire RNA and is too time consuming to be efficient for scanning through all possible blocks of a long RNA in the search of best target sites. To handle this problem for RNAi application, a short local RNA folding window of size W was used, along with L and block length u [97]. These treatments introduce substantial uncertainty in computational analysis. Indeed, for u, the empirically selected optimal values are quite different for two training datasets [97], raising the concern of the general applicability of optimal parameter values learned from one source of data. For a specific mRNA, because it is not possible to have accurate information on its independent folding domains which may be better predicted individually, the overall prediction accuracy would be compromised by a prespecified local folding window length that does not suit this specific mRNA. The major findings from this study are the same as we previously reported [94], i.e., target accessibility as a down stream factor in the RNAi pathway and duplex asymmetry for facilitating RISC assembly [49, 90] are two most important factors for RNAi efficiency.

2.3.3 *Specificity and Off-targeting*

Gene silencing by RNAi can be highly gene-specific [16, 92]. Single base-pair mismatches could drastically alter RNAi efficacy [33, 43], and siRNAs can be designed to discriminate the wild-type and mutant alleles of many genes that differ by just s single nucleotide [91]. However, off-target gene regulation by RNAi has been observed [45, 88]. Each strand of an siRNA duplex can possibly be assembled into the RISC to guide recognition of both fully and partially complementary mRNAs [79]. Off-target activity results from partial complementarity for nontargeted genes. Off-targeting can induce measurable phenotypes [58] and thus represents a major impediment to large-scale phonotypic screening applications of RNAi. While chemical modification of siRNA duplexes may reduce off-target effects [15, 46], it is essential to take into account the issue of target specificity in the design of siRNAs or shRNAs. Microarray studies suggest that off-targeting is mainly associated with perfect 3′ UTR matches for nucleotide positions 2–7 or 8 (hexamer or heptamer "seed" [57]; see Fig. 2.2) of the 5′ end of the siRNA guide strand [5, 47]. The seed region is an important determinant for target recognition by microRNAs [57]. However, in these microarray studies, the number of mRNAs with seed matches is far greater than the number of actual off-targets. In addition, two recent studies reported either a lack of enrichment for either 3′ UTR or seed matches [100] or a substantial number of off-targets that do not have a seed match [103]. These observations strongly indicate that additional factors responsible for off-target effects remain to

be identified. In general, it is advisable that an siRNA or shRNA contains at least three mismatches to any other genes in the genome of the species under study, that known single-nucleotide polymorphism should be avoided, and that common exons of alternatively spliced mRNAs should be avoided as well [77]. For a complete suite of RNAi design tools, it is essential to address both the issue of gene silencing potency and the issue of targeting specificity. This is particularly important for large-scale loss-of-function screens by using siRNA libraries [32] or shRNA libraries [10, 87, 96].

2.4 Posttranscriptional Gene Regulation by MicroRNAs

MicroRNAs are endogenous noncoding RNAs (ncRNAs) of ~22 nt and are among the most abundant regulatory molecules in multicellular organisms. microRNAs typically negatively regulate specific mRNA targets through essentially two mechanisms: (1) when a microRNA is perfectly or nearly perfectly complementary to mRNA target sites, as is the case for most plant microRNAs, it causes mRNA target cleavage [85]; and (2) a microRNA with incomplete complementarity to sequences in the 3′ untranslated region (3′ UTR) of its target (as is the case for most animal microRNAs) can cause translational repression or mRNA destabilization [34]. microRNAs regulate diverse developmental and physiological processes in animals and plants [1, 6, 11, 31, 102]. Besides animals and plants, microRNAs have also been discovered in viruses [18].

2.4.1 Target Identification Using Sequence Features

Identification and experimental validation of microRNA targets are essential for understanding the regulatory functions of this important class of ncRNAs. The targets and functions of plant microRNAs are relatively easy to identify due to the near-perfect complementarity [85]. By contrast, the incomplete target complementarity typical of animal microRNAs implies a huge regulatory potential but also presents a challenge for target identification. A number of algorithms have been developed for predicting animal microRNA targets. A common approach relies on a "seed" model based on a critical observation by Lai [52], wherein the target site is assumed to form strictly Watson–Crick (WC) pairs with bases at positions 2 through 7 or 8 of the 5′ end of the microRNA (see Fig. 2.2). In the stricter, "conserved seed" formulation of the model, perfect conservation of the 5′ seed match in the target is required across multiple species [56, 57]. One well-known exception to the seed model is the interaction between *let-7* on *lin-41* in *C. elegans*, as shown by Fig. 2.4, for which G–U pair and unpaired base(s) are present in the seed regions of two binding sites with experimental support [104]. While the seed model is supported as a basis for identifying many well-conserved microRNA targets [81], two studies suggest that G–U or mismatches in the seed region can be well tolerated and that conserved seed

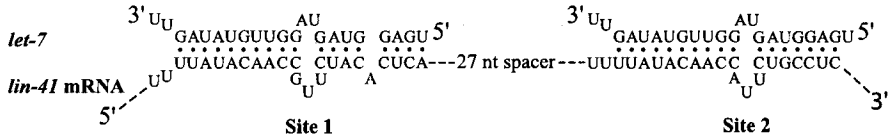

Fig. 2.4 *let-7* regulates *lin-41* by complementary base-pairing at two sites in the 3′ UTR of the *lin-41* mRNA [104, 105]. Neither the bulged A in the seed region for site 1 (in *red*, at position 5 from the 5′ end of the 27 nt spacer) nor the wobble G–U pair in the seed region for site 2 (in *red*, with U at position 6 of the 5′ end of *let-7*) meets the requirements of the seed model [56, 57] that bases 2 to 7 or 8 of the miRNA 5′ end must form Watson–Crick pairs with its target (for the color version, see Color Plates on p. 389)

match does not guarantee repression [20, 72]. These suggest that the seed model may represent only a subset of functional target sites and that additional factors are involved in further defining target specificity at least for some cases with conserved seed matches. A comprehensive study led to the proposal of three classes of target sites: "canonical", "seed", and "3′ compensatory" [8]. A canonical site pairs well with a microRNA on both the 5′ end and the 3′ end; a seed site has strong pairing to the 5′ end of the microRNA, with little or no pairing required on the 3″ end to stabilize the hybrid; a 3′ compensatory site requires strong pairing to the 3′ end of the microRNA to compensate for weak pairing on the 5′ end. Most genetically validated target sites appear to be of the canonical configuration, including the sites for *let-7: lin-41* (see Fig. 2.4). In addition to seed match, a number of features of site context have been proposed for enhancing targeting specificity [37]. More recently, functional target sites within the protein coding region of mouse mRNAs have been reported, and four of five validated mouse targets do not contain sites with seed match [98]. Also interestingly, a new class of human microRNA targets was reported to contain interaction sites in both the 5′ UTR and the 3′ UTR, and the 3′ end of the microRNAs are primarily involved in target binding for 5′ UTR sites [55].

2.4.2 *A Target Structure-Based Model for MicroRNA: Target Hybridization*

To attempt to understand the exceptions to the seed model and to develop target prediction methodology that does not rely on but can incorporate sequences features such as seed match, we considered the secondary structure of target mRNA that has been found to be important for other types of mRNA-targeting nucleic acids including siRNAs. We developed a model for modeling the interaction between a microRNA and a target as a two-step hybridization reaction (see Fig. 2.5): nucleation at an accessible target site, followed by hybrid elongation to disrupt local target secondary structure and form the complete microRNA-target duplex [59]. Nucleation potential and hybridization energy are two key energetic characteristics of the model. In this model, the role of target secondary structure on the efficacy

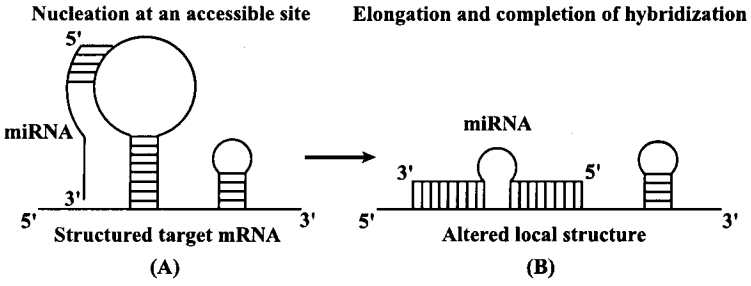

Fig. 2.5 Two-step model of hybridization between a small (partially) complementary nucleic acid molecule and a structured mRNA: (1) nucleation at an accessible site of at least 4 or 5 unpaired bases (**A**); (2) elongation through "unzipping" of the nearby helix, resulting in altered local target structure (**B**)

of repression by microRNAs is taken into account, by employing the Sfold program to address the likelihood of a population of structures that coexist in dynamic equilibrium for a specific mRNA molecule. This model can accurately account for the sensitivity to repression by *let-7* of both published and rationally designed mutant forms of the *Caenorhabditis elegans lin-41* 3′ UTR, and for the behavior of many other experimentally tested microRNA-target interactions in *C. elegans* and *Drosophila melanogaster*. The model is particularly effective in accounting for certain false positive predictions obtained by other methods. The model also performed well in a study of mammalian and viral microRNA targets [60].

In a more recent study [39], we analyzed a set of 3404 transcripts in *C.* elegans that were suggested by immunoprecipitation (IP) to be the targets for worm microRNAs [111]. Enrichment analyses by comparing targets and nontargets (i.e., transcripts absent in the IP dataset) revealed several important parameters. These include 5′ seed and modifications, structural accessibility of both the target site and the 25 nt-region upstream of the target site as assessed by target structures predicted by Sfold, the nucleation potential and the total energy change of the hybridization described in our previous work [59]. We developed a method to incorporate these significant parameters into worm microRNA target predictions. This method was found to make much better predictions than several well-known algorithms. Surprisingly, for this large target dataset, there was a lack of correlation for the contextual features based on analysis of microarray data for a small number of microRNAs [37]. In an independent study, three prediction parameters were analyzed for 6,387 candidate microRNA-target interactions between 114 human microRNAs and 890 mRNA transcripts, with patterns of expression across 88 human tissue samples [44]. These three parameters are the total energy change of the hybridization [59], the context score based on contextual features [37], and a core for measuring site conservation. It was found that only the total energy change of the hybridization is predictive of paired microRNA and mRNA expression data. Thus, the results from analyses of these large datasets not only further support our structure-based hybridization model but also cast doubts on the general applicability of the contextual features proposed from analysis of a relatively small microarray dataset. Results from microarray data

may not be highly reliable, due to inherent limitations and difficulty in the interpretation of the microarray data. For example, it has been shown that the secondary structure of the target is important for microarray probe design and data interpretation [83]. However, this issue has been largely overlooked in the analysis of microarray data. The importance of target structural accessibility is also supported by several other studies [48, 86, 113, 114].

2.5 Concluding Remarks

The paradigm-shifting work by McCaskill has inspired the developments of extended partition function algorithms for modeling single molecular folding and hybridization of two nucleic acid molecules, sampling extension and clustering representation of sampled ensemble. These methods enable characterizations of the equilibrium structure ensemble that are not possible with the use of free energy minimization.

For improving the potency of RNAi, target structure is clearly an important factor in the design of siRNAs. Several existing methods use different assumptions and treatments in parameter calculations for RNAi design. It is not clear whether one approach is superior to the other. Clearly, analyses of large datasets would be needed to compare these methods and to further investigate relevant issues such as the validity of global or local target folding. Off-targeting by RNAi is a major impediment for large-scale RNAi screening. 3′ UTR seed match can explain some but not all of observed off-target effects. It remains a challenge to identify additional factors responsible for off-target effects for improving the specificity of gene silencing.

It has been established that microRNAs can also target protein coding regions and 5′ UTR, in addition to 3′ UTR, and target binding can primarily involve the 5′ end or the 3′ end, or both ends of the microRNA. Seed match may represent a major class of target sites; however, it remains a challenge to estimate how large this class and other classes of targets are, which will require large-scale carefully designed experiments and analysis. Because seed pairing and contextual features are learned from small number of highly expressed microRNAs [4], the ratios of different classes of targets may well depend on the abundance of the microRNAs. It is conceivable that strong pairing for both 5′ end and the 3′ end can be essential for a microRNA of low abundance. The strength of microRNA-target hybridization would depend on the expression levels of the microRNA and its target. Incorporation of concentrations of microRNAs and target mRNAs will be a logical step for extension of hybridization modeling, as some data for expression levels have become available. Since target binding by microRNA can lead to two regulatory outcomes, translational repression or target mRNA degradation, it is an open question whether it is possible to predict the two outcomes.

Acknowledgements The Computational Molecular Biology and Statistics Core at the Wadsworth Center is acknowledged for providing computing resources. This work was supported in part by National Science Foundation grant DBI-0650991 and National Institutes of Health grant GM068726 to Y.D.

References

1. V. Ambros. The functions of animal microRNAs. *Nature*, **431**(7006):350–355, 2004.
2. S.L. Ameres, J. Martinez, and R. Schroeder. Molecular basis for target RNA recognition and cleavage by human RISC. *Cell*, **130**(1):101–112, 2007.
3. M. Andronescu, Z.C. Zhang, and A. Condon. Secondary structure prediction of interacting RNA molecules. *J Mol Biol*, **345**(5):987–1001, 2005.
4. D.P. Bartel. MicroRNAs: target recognition and regulatory functions. *Cell*, **136**(2):215–233, 2009.
5. A. Birmingham, E.M. Anderson, A. Reynolds, D. Ilsley-Tyree, D. Leake, Y. Fedorov, S. Baskerville, E. Maksimova, K. Robinson, J. Karpilow, W.S. Marshall, and A. Khvorova. 3′ UTR seed matches, but not overall identity, are associated with RNAi off-targets. *Nat Methods*, **3**(3):199–204, 2006.
6. M. Boehm and F. Slack. A developmental timing microRNA and its target regulate life span in C. elegans. *Science*, **310**(5756):1954–1957, 2005.
7. E.A. Bohula, A.J. Salisbury, M. Sohail, M.P. Playford, J. Riedemann, E.M. Southern, and V.M. Macaulay. 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**(18):15991–15997, 2003.
8. J. Brennecke, A. Stark, R.B. Russell, and S.M. Cohen. Principles of MicroRNA-target recognition. *PLoS Biol*, **3**(3):e85, 2005.
9. T.R. Brummelkamp, R. Bernards, and R. Agami. A system for stable expression of short interfering RNAs in mammalian cells. *Science*, **296**(5567):550–553, 2002.
10. F. Buchholz, R. Kittler, M. Slabicki, and M. Theis. Enzymatically prepared RNAi libraries. *Nat Methods*, **3**(9):696–700, 2006.
11. G.A. Calin, M. Ferracin, A. Cimmino, G. Di Leva, M. Shimizu, S.E. Wojcik, M.V. Iorio, R. Visone, N.I. Sever, M. Fabbri, R. Iuliano, T. Palumbo, F. Pichiorri, C. Roldo, R. Garzon, C. Sevignani, L. Rassenti, H. Alder, S. Volinia, C.G. Liu, T.J. Kipps, M. Negrini, and C.M. Croce. A MicroRNA signature associated with prognosis and progression in chronic lymphocytic leukemia. *N Engl J Med*, **353**(17):1793–1801, 2005.
12. L.E. Carvalho and C.E. Lawrence. Centroid estimation in discrete high-dimensional spaces with applications in biology. *Proc Natl Acad Sci USA*, **105**(9):3209–3214, 2008.
13. C.Y. Chan, C.E. Lawrence, and Y. Ding. Structure clustering features on the Sfold Web server. *Bioinformatics*, **21**(20):3926–3928, 2005.
14. C.Y. Chan, C.S. Carmack, D.D. Long, A. Maliyekkel, Y. Shao, I.B. Roninson, and Y. Ding. A structural interpretation of the effect of GC-content on efficiency of RNA interference. *BMC Bioinform*, **10**(1):S33, 2009.
15. P.Y. Chen, L. Weinmann, D. Gaidatzis, Y. Pei, M. Zavolan, T. Tuschl, and G. Meister. Strand-specific 5′-O-methylation of siRNA duplexes controls guide strand selection and targeting specificity. *RNA*, **14**(2):263–274, 2008.
16. J.T. Chi, H.Y. Chang, N.N. Wang, D.S. Chang, N. Dunphy, and P.O. Brown. Genomewide view of gene silencing by small interfering RNAs. *Proc Natl Acad Sci USA*, **100**(11):6343–6346, 2003.
17. P. Clote, J. Waldispuhl, B. Behzadi, and J.M. Steyaert. Energy landscape of k-point mutants of an RNA molecule. *Bioinformatics*, **21**(22):4140–4147, 2005.
18. B.R. Cullen. Viruses and microRNAs. *Nat Genet*, **38**:S25–30, 2006.
19. J. Cupal, C. Flamm, A. Renner, and P.F. Stadler. Density of states, metastable states, and saddle points exploring the energy landscape of an RNA molecule. *Proc Int Conf Intell Syst Mol Biol*, **5**:88–91, 1997.
20. D. Didiano and O. Hobert. Perfect seed pairing is not a generally reliable predictor for miRNA-target interactions. *Nat Struct Mol Biol*, **13**(9):849–851, 2006.
21. R.A. Dimitrov and M. Zuker. Prediction of hybridization and melting for double-stranded nucleic acids. *Biophys J*, **87**(1):215–226, 2004.

22. Y. Ding. Rational statistical design of antisense oligonucleotides for high throughput functional genomics and drug target validation. *Stat Sin*, **12**:273–296, 2002.
23. Y. Ding and C.E. Lawrence. Statistical prediction of single-stranded regions in RNA secondary structure and application to predicting effective antisense target sites and beyond. *Nucleic Acids Res*, **29**(5):1034–1046, 2001.
24. Y. Ding and C.E. Lawrence. A statistical sampling algorithm for RNA secondary structure prediction. *Nucleic Acids Res*, **31**(24):7280–7301, 2003.
25. Y. Ding and C.E. Lawrence. Rational design of siRNAs with the Sfold software. In K. Appasani, editor, *RNA Interference: from Basic Science to Drug Development*, pages 129–138. Cambridge University Press, Cambridge, 2005.
26. Y. Ding, C.Y. Chan, and C.E. Lawrence. Sfold web server for statistical folding and rational design of nucleic acids. *Nucleic Acids Res*, **32**:W135–141, 2004. (Web Server issue)
27. Y. Ding, C.Y. Chan, and C.E. Lawrence. RNA secondary structure prediction by centroids in a Boltzmann weighted ensemble. *RNA*, **11**(8):1157–1166, 2005.
28. Y. Ding, C.Y. Chan, and C.E. Lawrence. Clustering of RNA secondary structures with application to messenger RNAs. *J Mol Biol*, **359**(3):554–571, 2006.
29. R.M. Dirks and N.A. Pierce. A partition function algorithm for nucleic acid secondary structure including pseudoknots. *J Comput Chem*, **24**(13):1664–1677, 2003.
30. R.M. Dirks and N.A. Pierce. An algorithm for computing nucleic acid base-pairing probabilities including pseudoknots. *J Comput Chem*, **25**(10):1295–1304, 2004.
31. D.V. Dugas and B. Bartel. MicroRNA regulation of gene expression in plants. *Curr Opin Plant Biol*, **7**(5):512–520, 2004.
32. C.J. Echeverri and N. Perrimon. High-throughput RNAi screening in cultured cells: a user's guide. *Nat Rev Genet*, **7**(5):373–384, 2006.
33. S.M. Elbashir, J. Harborth, W. Lendeckel, A. Yalcin, K. Weber, and T. Tuschl. Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature*, **411**(6836):494–498, 2001.
34. W. Filipowicz, S.N. Bhattacharyya, and N. Sonenberg. Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? *Nat Rev Genet*, **9**(2):102–114, 2008.
35. A. Fire, S. Xu, M.K. Montgomery, S.A. Kostas, S.E. Driver, and C.C. Mello. Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature*, **391**(6669):806–811, 1998.
36. R. Giegerich, B. Voss, and M. Rehmsmeier. Abstract shapes of RNA. *Nucleic Acids Res*, **32**(16):4843–4851, 2004.
37. A. Grimson, K.K. Farh, W.K. Johnston, P. Garrett-Engele, L.P. Lim, and D.P. Bartel. MicroRNA targeting specificity in mammals: Determinants beyond seed pairing. *Mol Cell*, **27**(1):91–105, 2007.
38. M. Hamada, H. Kiryu, K. Sato, T. Mituyama, and K. Asai. Prediction of RNA secondary structure using generalized centroid estimators. *Bioinformatics*, **25**(4):465–473, 2009.
39. M. Hammell, D. Long, L. Zhang, A. Lee, C.S. Carmack, M. Han, Y. Ding, and V. Ambros. mirWIP: microRNA target prediction based on microRNA-containing ribonucleoprotein-enriched transcripts. *Nat Methods*, **5**:813–819, 2008.
40. M.R. Hargittai, R.J. Gorelick, I. Rouzina, and K. Musier-Forsyth. Mechanistic insights into the kinetics of HIV-1 nucleocapsid protein-facilitated tRNA annealing to the primer binding site. *J Mol Biol*, **337**(4):951–968, 2004.
41. B.S. Heale, H.S. Soifer, C. Bowers, and J.J. Rossi. siRNA target site secondary structure predictions using local stable substructures. *Nucleic Acids Res*, **33**(3):e30, 2005.
42. I.L. Hofacker. Vienna RNA secondary structure server. *Nucleic Acids Res*, **31**(13):3429–3431, 2003.
43. T. Holen, M. Amarzguioui, M.T. Wiiger, E. Babaie, and H. Prydz. Positional effects of short interfering RNAs targeting the human coagulation trigger Tissue Factor. *Nucleic Acids Res*, **30**(8):1757–1766, 2002.
44. J.C. Huang, B.J. Frey, and Q.D. Morris. Comparing sequence and expression for predicting microRNA targets using GenMiR3. In *Pacific Symposium on Biocomputing, volume 13*, pages 52–63, 2008.

45. A.L. Jackson, S.R. Bartz, J. Schelter, S.V. Kobayashi, J. Burchard, M. Mao, B. Li, G. Cavet, and P.S. Linsley. Expression profiling reveals off-target gene regulation by RNAi. *Nat Biotechnol*, **21**(6):635–637, 2003.
46. A.L. Jackson, J. Burchard, D. Leake, A. Reynolds, J. Schelter, J. Guo, J.M. Johnson, L. Lim, J. Karpilow, K. Nichols, W. Marshall, A. Khvorova, and P.S. Linsley. Position-specific chemical modification of siRNAs reduces "off-target" transcript silencing. *RNA*, **12**(7):1197–1205, 2006.
47. A.L. Jackson, J. Burchard, J. Schelter, B.N. Chau, M. Cleary, L. Lim, and P.S. Linsley. Widespread siRNA "off-target" transcript silencing mediated by seed region sequence complementarity. *RNA*, **12**(7):1179–1187, 2006.
48. M. Kertesz, N. Iovino, U. Unnerstall, U. Gaul, and E. Segal. The role of site accessibility in microRNA target recognition. *Nat Genet*, **39**(10):1278–1284, 2007.
49. A. Khvorova, A. Reynolds, and S.D. Jayasena. Functional siRNAs and miRNAs exhibit strand bias. *Cell*, **115**(2):209–216, 2003.
50. R. Kretschmer-Kazemi Far and G. Sczakiel. The activity of siRNA in mammalian cells is related to structural target accessibility: a comparison with antisense oligonucleotides. *Nucleic Acids Res*, **31**(15):4417–4424, 2003.
51. I. Ladunga. More complete gene silencing by fewer siRNAs: transparent optimized design and biophysical signature. *Nucleic Acids Res*, **35**(2):433–440, 2007.
52. E.C. Lai. Micro RNAs are complementary to 3′ UTR sequence motifs that mediate negative post-transcriptional regulation. *Nat Genet*, **30**(4):363–364, 2002.
53. D.M. Layton and R. Bundschuh. A statistical analysis of RNA folding algorithms through thermodynamic parameter perturbation. *Nucleic Acids Res*, **33**(2):519–524, 2005.
54. N.S. Lee, T. Dohjima, G. Bauer, H. Li, M.J. Li, A. Ehsani, P. Salvaterra, and J. Rossi. Expression of small interfering RNAs targeted against HIV-1 rev transcripts in human cells. *Nat Biotechnol*, **20**(5):500–505, 2002.
55. I. Lee, S.S. Ajay, J.I. Yook, H.S. Kim, S.H. Hong, N.H. Kim, S.M. Dhanasekaran, A. Chinnaiyan, and B.D. Athey. New class of microRNA targets containing simultaneous 5′-UTR and 3′-UTR interaction sites. *Genome Res*, 2009.
56. B.P. Lewis, I.H. Shih, M.W. Jones-Rhoades, D.P. Bartel, and C.B. Burge. Prediction of mammalian microRNA targets. *Cell*, **115**(7):787–798, 2003.
57. B.P. Lewis, C.B. Burge, and D.P. Bartel. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell*, **120**(1):15–20, 2005.
58. X. Lin, X. Ruan, M.G. Anderson, J.A. McDowell, P.E. Kroeger, S.W. Fesik, and Y. Shen. siRNA-mediated off-target gene silencing triggered by a 7 nt complementation. *Nucleic Acids Res*, **33**(14):4527–4535, 2005.
59. D. Long, R. Lee, P. Williams, C.Y. Chan, V. Ambros, and Y. Ding. Potent effect of target structure on microRNA function. *Nat Struct Mol Biol*, **14**:287–294, 2007.
60. D. Long, C.Y. Chan, and Y. Ding. Analysis of microRNA-target interactions by a target structure based hybridization model. In *Pacific Symposium on Biocomputing*, volume *13*, pages 64–74, 2008.
61. Z.J. Lu and D.H. Mathews. Efficient siRNA selection using hybridization thermodynamics. *Nucleic Acids Res*, **36**(2):640–647, 2008.
62. K.Q. Luo and D.C. Chang. The gene-silencing efficiency of siRNA is strongly dependent on the local structure of mRNA at the targeted region. *Biochem Biophys Res Commun*, **318**(1):303–310, 2004.
63. N.R. Markham and M. Zuker. UNAFold: software for nucleic acid folding and hybridization. *Methods Mol Biol*, **453**:3–31, 2008.
64. D.H. Mathews. Using an RNA secondary structure partition function to determine confidence in base pairs predicted by free energy minimization. *RNA*, **10**(8):1178–1190, 2004.
65. D.H. Mathews. RNA secondary structure analysis using RNAstructure. In *Curr Protoc Bioinformatics*, Chapter 12: Unit 12.16, 2006.
66. D.H. Mathews, M.E. Burkard, S.M. Freier, J.R. Wyatt, and D.H. Turner. Predicting oligonucleotide affinity to nucleic acid targets. *RNA*, **5**(11):1458–1469, 1999.

67. D.H. Mathews, J. Sabina, M. Zuker, and D.H. Turner. Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J Mol Biol*, **288**(5):911–940, 1999b.
68. D.H. Mathews, M.D. Disney, J.L. Childs, S.J. Schroeder, M. Zuker, and D.H. Turner. Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proc Natl Acad Sci USA*, **101**(19):7287–7292, 2004.
69. C. Matranga, Y. Tomari, C. Shin, D.P. Bartel, and P.D. Zamore. Passenger-strand cleavage facilitates assembly of siRNA into Ago2-containing RNAi enzyme complexes. *Cell*, **123**(4):607–620, 2005.
70. J.S. McCaskill. The equilibrium partition function and base pair binding probabilities for RNA secondary structure. *Biopolymers*, **29**(6–7):1105–1119, 1990.
71. I. Miklos, I.M. Meyer, and B. Nagy. Moments of the Boltzmann distribution for RNA secondary structures. *Bull Math Biol*, **67**(5):1031–1047, 2005.
72. K.C. Miranda, T. Huynh, Y. Tay, Y.S. Ang, W.L. Tam, A.M. Thomson, B. Lim, and I. Rigoutsos. A pattern-based method for the identification of microRNA binding sites and their corresponding heteroduplexes. *Cell*, **126**(6):1203–1217, 2006.
73. U. Muckstein, H. Tafer, J. Hackermuller, S.H. Bernhart, P.F. Stadler, and I.L. Hofacker. Thermodynamics of RNA–RNA binding. *Bioinformatics*, **22**(10):1177–1182, 2006.
74. R. Nussinov and A.B. Jacobson. Fast algorithm for predicting the secondary structure of single-stranded RNA. *Proc Natl Acad Sci USA*, **77**(11):6309–6313, 1980.
75. M. Overhoff, M. Alken, R.K. Far, M. Lemaitre, B. Lebleu, G. Sczakiel, and I. Robbins. Local RNA target structure influences siRNA efficacy: a systematic global analysis. *J Mol Biol*, **348**(4):871–881, 2005.
76. P.J. Paddison, A.A. Caudy, E. Bernstein, G.J. Hannon, and D.S. Conklin. Short hairpin RNAs (shRNAs) induce sequence-specific silencing in mammalian cells. *Genes Dev*, **16**(8):948–958, 2002.
77. P.J. Paddison, M. Cleary, J.M. Silva, K. Chang, N. Sheth, R. Sachidanandam, and G.J. Hannon. Cloning of short hairpin RNAs for gene knockdown in mammalian cells. *Nat Methods*, **1**(2):163–167, 2004.
78. V. Patzel, S. Rutz, I. Dietrich, C. Koberle, A. Scheffold, and S.H. Kaufmann. Design of siRNAs producing unstructured guide-RNAs results in improved RNA interference efficiency. *Nat Biotechnol*, **23**(11):1440–1444, 2005.
79. Y. Pei and T. Tuschl. On the art of identifying effective and specific siRNAs. *Nat Methods*, **3**(9):670–676, 2006.
80. Y. Ponty. Efficient sampling of RNA secondary structures from the Boltzmann ensemble of low-energy: the boustrophedon method. *J Math Biol*, **56**(1–2):107–127, 2008.
81. N. Rajewsky. microRNA target predictions in animals. *Nat Genet*, **38**:S8–13, 2006.
82. T.A. Rand, S. Petersen, F. Du, and X. Wang. Argonaute2 cleaves the anti-guide strand of siRNA during RISC activation. *Cell*, **123**(4):621–629, 2005.
83. V.G. Ratushna, J.W. Weller, and C.J. Gibas. Secondary structure in the target as a confounding factor in synthetic oligomer microarray design. *BMC Genomics*, **6**(1):31, 2005.
84. A. Reynolds, D. Leake, Q. Boese, S. Scaringe, W.S. Marshall, and A. Khvorova. Rational siRNA design for RNA interference. *Nat Biotechnol*, **22**(3):326–330, 2004.
85. M.W. Rhoades, B.J. Reinhart, L.P. Lim, C.B. Burge, B. Bartel, and D.P. Bartel. Prediction of plant microRNA targets. *Cell*, **110**(4):513–520, 2002.
86. H. Robins, Y. Li, and R.W. Padgett. Incorporating structure to predict microRNA targets. *Proc Natl Acad Sci USA*, **102**(11):4006–4009, 2005.
87. D.E. Root, N. Hacohen, W.C. Hahn, E.S. Lander, and D.M. Sabatini. Genome-scale loss-of-function screening with a lentiviral RNAi library. *Nat Methods*, **3**(9):715–719, 2006.
88. P.C. Scacheri, O. Rozenblatt-Rosen, N.J. Caplen, T.G. Wolfsberg, L. Umayam, J.C. Lee, C.M. Hughes, K.S. Shanmugam, A. Bhattacharjee, M. Meyerson, and F.S. Collins. Short interfering RNAs can induce unexpected and divergent changes in the levels of untargeted proteins in mammalian cells. *Proc Natl Acad Sci USA*, **101**(7):1892–1897, 2004.

89. S. Schubert, A. Grunweller, V.A. Erdmann, and J. Kurreck. Local RNA target structure influences siRNA efficacy: systematic analysis of intentionally designed binding regions. *J Mol Biol*, **348**(4):883–893, 2005.
90. D.S. Schwarz, G. Hutvagner, T. Du, Z. Xu, N. Aronin, and P.D. Zamore. Asymmetry in the assembly of the RNAi enzyme complex. *Cell*, **115**(2):199–208, 2003.
91. D.S. Schwarz, H. Ding, L. Kennington, J.T. Moore, J. Schelter, J. Burchard, P.S. Linsley, N. Aronin, Z. Xu, and P.D. Zamore. Designing siRNA that distinguish between genes that differ by a single nucleotide. *PLoS Genet*, **2**(9):e140, 2006.
92. D. Semizarov, L. Frost, A. Sarthy, P. Kroeger, D.N. Halbert, and S.W. Fesik. Specificity of short interfering RNA determined through gene expression signatures. *Proc Natl Acad Sci USA*, **100**(11):6347–6352, 2003.
93. Y. Shao, Y. Wu, C.Y. Chan, K. McDonough, and Y. Ding. Rational design and rapid screening of antisense oligonucleotides for prokaryotic gene modulation. *Nucleic Acids Res*, **34**(19):5660–5669, 2006.
94. Y. Shao, C.Y. Chan, A. Maliyekkel, C.E. Lawrence, C.E. Roninsonx, and Y. Ding. Effect of target secondary structure on RNAi efficiency. *RNA*, **13**(10):1631–1640, 2007.
95. Y. Shao, S. Wu, C.Y. Chan, J.R. Klapper, E. Schneider, and Y. Ding. A structural analysis of in vitro catalytic activities of hammerhead ribozymes. *BMC Bioinf*, **8**(1):469, 2007.
96. D. Shirane, K. Sugao, S. Namiki, M. Tanabe, M. Iino, and K. Hirose. Enzymatic production of RNAi libraries from cDNAs. *Nat Genet*, **36**(2):190–196, 2004.
97. H. Tafer, S.L. Ameres, G. Obernosterer, C.A. Gebeshuber, R. Schroeder, J. Martinez, and I.L. Hofacker. The impact of target site accessibility on the design of effective siRNAs. *Nat Biotechnol*, **26**(5):578–583, 2008.
98. Y. Tay, J. Zhang, A.M. Thomson, B. Lim, and I. Rigoutsos. MicroRNAs to Nanog, Oct4 and Sox2 coding regions modulate embryonic stem cell differentiation. *Nature*, **455**(7216):1124–1128, 2008.
99. I. Tinoco, Jr., O.C. Uhlenbeck, and M.D. Levine. Estimation of secondary structure in ribonucleic acids. *Nature*, **230**(5293):362–367, 1971.
100. C. Tschuch, A. Schulz, A. Pscherer, W. Werft, A. Benner, A. Hotz-Wagenblatt, L.S. Barrionuevo, P. Lichter, and D. Mertens. Off-target effects of siRNA specific for GFP. *BMC Mol Biol*, **9**:60, 2008.
101. K. Ui-Tei, Y. Naito, F. Takahashi, T. Haraguchi, H. Ohki-Hamazaki, A. Juni, R. Ueda, and K. Saigo. Guidelines for the selection of highly effective siRNA sequences for mammalian and chick RNA interference. *Nucleic Acids Res*, **32**(3):936–948, 2004.
102. E. van Rooij, L.B. Sutherland, X. Qi, J.A. Richardson, J. Hill, and E.N. Olson. Control of stress-dependent cardiac growth and gene expression by a microRNA. *Science*, **316**(5824):575–579, 2007.
103. S. Vankoningsloo, F. de Longueville, S. Evrard, P. Rahier, A. Houbion, A. Fattaccioli, M. Gastellier, J. Remacle, M. Raes, P. Renard, and T. Arnould. 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**(11):2738–2753, 2008.
104. M.C. Vella, E.Y. Choi, S.Y. Lin, K. Reinert, and F.J. Slack. The C. elegans microRNA let-7 binds to imperfect let-7 complementary sites from the lin-41 3′UTR. *Genes Dev*, **18**(2):132–137, 2004.
105. M.C. Vella, K. Reinert, and F.J. Slack. Architecture of a validated microRNA:target interaction. *Chem Biol*, **11**(12):1619–1623, 2004.
106. T.A. Vickers, S. Koo, C.F. Bennett, S.T. Crooke, N.M. Dean, and B.F. Baker. Efficient reduction of target RNAs by small interfering RNA and RNase H-dependent antisense agents. A comparative analysis. *J Biol Chem*, **278**(9):7108–7118, 2003.
107. B. Voss, R. Giegerich, and M. Rehmsmeier. Complete probabilistic analysis of RNA shapes. *BMC Biol*, **4**:5, 2006.
108. E.M. Westerhout, M. Ooms, M. Vink, A.T. Das, and B. Berkhout. HIV-1 can escape from RNA interference by evolving an alternative structure in its RNA genome. *Nucleic Acids Res*, **33**(2):796–804, 2005.

109. S. Wuchty, W. Fontana, I.L. Hofacker, and P. Schuster. Complete suboptimal folding of RNA and the stability of secondary structures. *Biopolymers*, **49**(2):145–165, 1999.
110. T. Xia, J. SantaLucia, Jr., M.E. Burkard, R. Kierzek, S.J. Schroeder, X. Jiao, C. Cox, and D.H. Turner. Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson–Crick base pairs. *Biochemistry*, **37**(42):14719–14735, 1998.
111. L. Zhang, L. Ding, T.H. Cheung, M.Q. Dong, J. Chen, A.K. Sewell, X. Liu, J.R. Yates 3rd, and M. Han. Systematic identification of C. elegans miRISC proteins, miRNAs, and mRNA targets by their interactions with GW182 proteins AIN-1 and AIN-2. *Mol Cell*, **28**(4):598–613, 2007.
112. J.J. Zhao and G. Lemke. Rules for ribozymes. *Mol Cell Neurosci*, **11**(1–2):92–97, 1998.
113. Y. Zhao, E. Samal, and D. Srivastava. Serum response factor regulates a muscle-specific microRNA that targets Hand2 during cardiogenesis. *Nature*, **436**(7048):214–220, 2005.
114. Y. Zhao, J.F. Ransom, A. Li, V. Vedantham, M. von Drehle, A.N. Muth, T. Tsuchihashi, M.T. McManus, R.J. Schwartz, and D. Srivastava. Dysregulation of cardiogenesis, cardiac conduction, and cell cycle in mice lacking miRNA-1-2. *Cell*, **129**(2):303–317, 2007.
115. M. Zuker. Calculating nucleic acid secondary structure. *Curr Opin Struct Biol*, **10**(3):303–310, 2000.
116. M. Zuker and P. Stiegler. Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucleic Acids Res*, **9**(1):133–148, 1981.