Predicting oligonucleotide affinity to nucleic acid targets

D. H. Mathews, M. E. Burkard, S. M. Freier, J. R. Wyatt and D. H. Turner

RNA 1999 5: 1458-1469

RNA (1999), *5*:1458–1469. Cambridge University Press. Printed in the USA.

Predicting oligonucleotide affinity to nucleic acid targets

DAVID H. MATHEWS,[1] MARK E. BURKARD,[1] SUSAN M. FREIER,[2] JACQUELINE R. WYATT,[2] and DOUGLAS H. TURNER[1]

[1]Department of Chemistry, University of Rochester, Rochester, New York 14627-0216, USA
[2]Division of Molecular and Structural Biology, Isis Pharmaceuticals, Carlsbad, California 92008, USA

ABSTRACT

A computer program, OligoWalk, is reported that predicts the equilibrium affinity of complementary DNA or RNA oligonucleotides to an RNA target. This program considers the predicted stability of the oligonucleotide-target helix and the competition with predicted secondary structure of both the target and the oligonucleotide. Both unimolecular and bimolecular oligonucleotide self structure are considered with a user-defined concentration. The application of OligoWalk is illustrated with three comparisons to experimental results drawn from the literature.

Keywords: antisense oligonucleotides; bimolecular RNA secondary structure; hybridization; primer extension; reverse PCR; reverse transcription; RNA secondary-structure prediction

INTRODUCTION

For a target RNA or DNA sequence of N nucleotides, there are $N - L + 1$ complementary oligonucleotides of length L. Which of these oligomers will bind most tightly? The answer to this question is important for the design of antisense oligomers, polymerase chain reaction (PCR) and reverse transcription primers, and hybridization probes.

To predict the affinity of an oligomer to a target, a first approximation is obtained by examining the duplex stability of the oligomer-target helix, but self structure of both oligomer and target reduce the stability of this interaction. The self structure of the target is usually unimolecular at biological concentrations. For an oligomer probe of higher concentration than the target, the self structure can be either unimolecular or bimolecular. Tertiary structure is generally less stable than secondary structure and can therefore be neglected as an approximation (Crothers et al., 1974; Hilbers et al., 1976; Banerjee et al., 1993; Jaeger et al., 1993; Mathews et al., 1997).

This article presents a computer program, OligoWalk, that predicts the equilibrium affinity of complementary DNA or RNA oligonucleotides to a structured RNA target. The target can be a known structure or, if unknown, a set of predicted secondary structures. Duplex stability, local secondary structure in target, and both intermolecular and intramolecular secondary structure in the oligonucleotide are considered. Three examples drawn from the literature are used to make comparisons with predicted binding and illustrate the use of the program.

Reprint requests to: Douglas H. Turner, Department of Chemistry, University of Rochester, Rochester, New York 14627-0216, USA; e-mail: turner@chem.chem.rochester.edu.

RESULTS

Oligomer-target binding

The OligoWalk program requires five inputs from the user. First, a target RNA structure is selected. This can be a secondary structure determined by comparative sequence analysis (James et al., 1989; Pace et al., 1999) or predicted by an RNA folding algorithm (Mathews et al., 1999) with or without refinement with chemical and/or enzymatic cleavage data (Ehresmann et al., 1987; Knapp, 1989; Mathews et al., 1997, 1999). A set of predicted suboptimal structures can also be included in the calculation of affinity. The user selects oligomer length, concentration, and chemical structure, that is, RNA or DNA. Finally, the user chooses the method for considering target structure, that is, either local or global reorganization upon oligomer binding and either a single structure or an ensemble of structures are used for considering target structure.

OligoWalk calculates the equilibrium affinity of each complementary oligomer using the scheme illustrated

in Figure 1. Each oligomer can be (1) in a folded, unimolecular structure, $O_{F–U}$, (2) in a folded, bimolecular structure, $O_{F–B}$, or (3) unfolded, O_U. The target RNA can be folded in the region of complementarity, T_F, or unfolded in the region of complementarity, T_U. Finally, the target can be bound to oligomer, O-T. The affinity of oligomer to target is expressed as an overall Gibb's free-energy change of self structured oligomer and target associating into an oligomer-target complex. As detailed in the methods section, OligoWalk calculates the overall free-energy change of binding at 37 °C for each possible oligomer according to:

$$\Delta G^\circ_{overall} = \Delta G^\circ_3 + R(310.15) \times [\ln(K_{1eff} + 1) + \ln(K_2 + 1)]. \quad (1)$$

As usual, free-energy change at 37 °C is related to the equilibrium constant, K, according to

$$\Delta G^\circ = -R(310.15)\ln K, \quad (2)$$

R is the gas constant, and 310.15 K is the absolute temperature equivalent to 37 °C. ΔG°_3 is the free-energy change attributed to oligomer-target helix formation from unstructured oligomer and target, K_2 is the equilibrium constant for the folding of local target structure, and K_{1eff} is derived from an effective equilibrium between unfolded and both unimolecularly and bimolecularly folded oligomer

$$K_{1eff} = \frac{[O_{F–U}] + [O_{F–B}]}{[O_U]} = \left(\frac{4K_{1B}[O_{total}]}{-1 - K_{1U} + \sqrt{(1 + K_{1U})^2 + 8K_{1B}[O_{total}]}} \right) - 1. \quad (3)$$

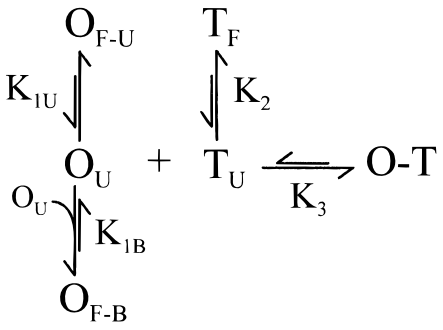

FIGURE 1. The equilibrium scheme used by OligoWalk. $O_{F–U}$ and $O_{F–B}$ are oligomer in a folded unimolecular and bimolecular structure, respectively. O_U is the unfolded oligomer. The target RNA is either folded in the complementary region, T_F, or single stranded in the complementary region, T_U. O-T is the target with oligomer bound. Each equilibrium is governed by an equilibrium constant, K, defined to be forward in the more folded direction.

The derivation of equation 3 is shown in Materials and Methods. A consequence of this equation is that the relative contributions of unimolecular and bimolecular oligonucleotide structure depends on the relative stability of these two structures and the total concentration of oligonucleotide. To evaluate equation 3, ΔG°_{1U} and ΔG°_{1B} are determined by secondary-structure prediction using either RNA or DNA parameters (SantaLucia & Allawi, 1997; SantaLucia, 1998; Xia et al., 1998; Mathews et al., 1999). To predict ΔG°_{1B}, the secondary-structure prediction algorithm of Zuker (1989a,b) was modified to allow bimolecular structure formation as described below. The value of ΔG°_3, the stability of base pairing of oligomer to target, is determined using nearest-neighbor parameters for either RNA–RNA (Xia et al., 1998) or RNA–DNA (Sugimoto et al., 1995) helices.

For the free-energy penalty associated with breaking of the target's local secondary structure, ΔG°_2, four different forms of the calculation are available (Figure 2). The binding of the oligomer will disrupt preexisting structure in the region of complementarity. The disruption can be treated as a local disruption only, or the target structure can rearrange to optimize secondary structure. Local disruption assumes that the target secondary structure is fixed outside the region of oligomer binding, whereas rearrangement assumes a true equilibrium of the target secondary structure. Either may be important for a cellular target, because kinetics or associated proteins may or may not allow rearrangement. For each option, suboptimal predicted structures of the target can be included in the calculation. This is important for targets with secondary structure predicted by free-energy minimization without experimental constraints.

Form I of Figure 2 is the simplest form of the calculation. It considers local disruption of structure on a single target secondary structure:

$$\Delta G^\circ_2 = \Delta G^\circ_{37}(T_F) - \Delta G^\circ_{37}(T_U) \quad (4)$$

where T_U is the target structure without any base pairs in the region of oligomer complementarity. Free-energy changes are approximated with nearest-neighbor free-energy parameters using the program Efn2 (Mathews et al., 1999).

Form II considers local structure for a set of suboptimal structures. Each structure contributes to the free-energy penalty for disruption of structure in proportion to a Boltzmann weight:

$$\Delta G^\circ_2 = \frac{\sum_i [(e^{-\Delta G^\circ_i/(310.15R)})(\Delta G^\circ_i - \Delta G^{\circ\prime}_i)]}{\sum_i e^{-\Delta G^\circ_i/(310.15R)}}. \quad (5)$$

The summations are over all suboptimal structures provided. ΔG°_i is the free energy of suboptimal structure *i*

	Include suboptimal predicted structures for Target:	
Disruption of Secondary Structure:	No -	Yes -
Local only -	**Form I**	**Form II**
Allow rearrangement of Target when oligomer is bound -	**Form III**	**Form IV**

FIGURE 2. The four forms for calculation of disrupted structure in target. The user can choose to include suboptimal structures in the calculation. Suboptimal structures are important when the target secondary structure is predicted by free-energy minimization without experimental constraints. The disruption in secondary structure can be considered either locally only or with target rearrangement to optimize overall free energy. Thus, four forms of the calculation are available for these options.

and $\Delta G^{\circ\prime}_i$ is the free energy of suboptimal structure *i* without base pairs in the region of complementarity. Each is calculated with the Efn2 program (Mathews et al., 1999).

Form III allows rearrangement of secondary structure upon oligomer binding with a single target secondary structure:

$$\Delta G^\circ_2 = \Delta G^\circ(T_F) - \Delta G^\circ(T'_U) \quad (6)$$

where T'_U is the target refolded with each nucleotide in the region of complementarity forced to be unpaired. The free energy of the entered structure is calculated with Efn2 and the refolded structure is predicted using free-energy minimization (Mathews et al., 1999).

Form IV allows rearrangement of suboptimal secondary structures for each of a set of suboptimal structures upon oligomer binding. Equation 5 is used for this calculation with $\Delta G^{\circ\prime}_i$ as the lowest free-energy structure predicted when nucleotides bound to the oligomer are forced to be unpaired (Mathews et al., 1999).

Bimolecular secondary structure prediction

To determine the free-energy change for formation of the bimolecular oligomer–oligomer structure, the algorithm of Zuker (Zuker & Stiegler, 1981; Zuker, 1989a) was modified to predict bimolecular secondary structures. This is done by predicting secondary structure for two sequences joined by a virtual linker composed of nucleotides that cannot pair. For the loop region occupied by the linker, the algorithm applies the free energy of intermolecular initiation rather than a loop penalty. This method allows many conformations of intermolecular structure, but it cannot predict those that mimic a pseudoknot, most notably the kissing hairpin motif (Chang & Tinoco, 1994, 1997; Gregorian & Crothers, 1995; Comolli et al., 1998). In the future, however, this method for bimolecular structure prediction could be combined with recent work in dynamic programming (Rivas & Eddy, 1999) to predict kissing hairpins.

Display of affinity calculations

The OligoWalk program is part of the RNAstructure package for Microsoft Windows. The program has a user friendly, graphical interface that displays the predicted free energies in a window as illustrated in Figure 3. The target sequence is shown 5′ to 3′ with one complementary oligomer at a time. The top portion displays the current oligomer number, the type of oligomer (DNA or RNA), and the oligomer concentration. The five free-energy changes including $\Delta G^\circ_{overall}$ (Fig. 1) are displayed for the current oligomer. Also displayed is the oligomer-target melting temperature, T_m, without considering oligomer or target self structure. Below, the bar graph displays the overall free energy, $\Delta G^\circ_{overall}$, and the duplex free energy, ΔG°_3. The other free-energy parameters can be graphed alternatively. Clicking on the oligomer will display either the oligomer intramolecular or intermolecular self structure. The arrows and “go” buttons are for navigation across the target sequence. The data for all oligomers can be output to a tab-delimited text file that is suitable for reading into other programs, such as Microsoft Excel.

Comparison of OligoWalk predictions to data from the literature

OligoWalk predictions of oligonucleotide-target affinity were compared to three types of experiments drawn from the literature. The first comparison is to a cell-free RNase-H cleavage assay (Ho et al., 1998). The second is to the cellular antisense efficacy of 22 oligomers (Ho et al., 1996). The third is to cell-free and cellular accessibilities of sickle β-globin mRNA to splicing by an exogenous ribozyme (Lan et al., 1998).

RNase-H cleavage

The RNase-H cleavage assay is a method for determining regions of an RNA target that are accessible to DNA oligomers (Ho et al., 1996, 1998; Birikh et al., 1997; Lima et al., 1997; Milner et al., 1997). It is directly

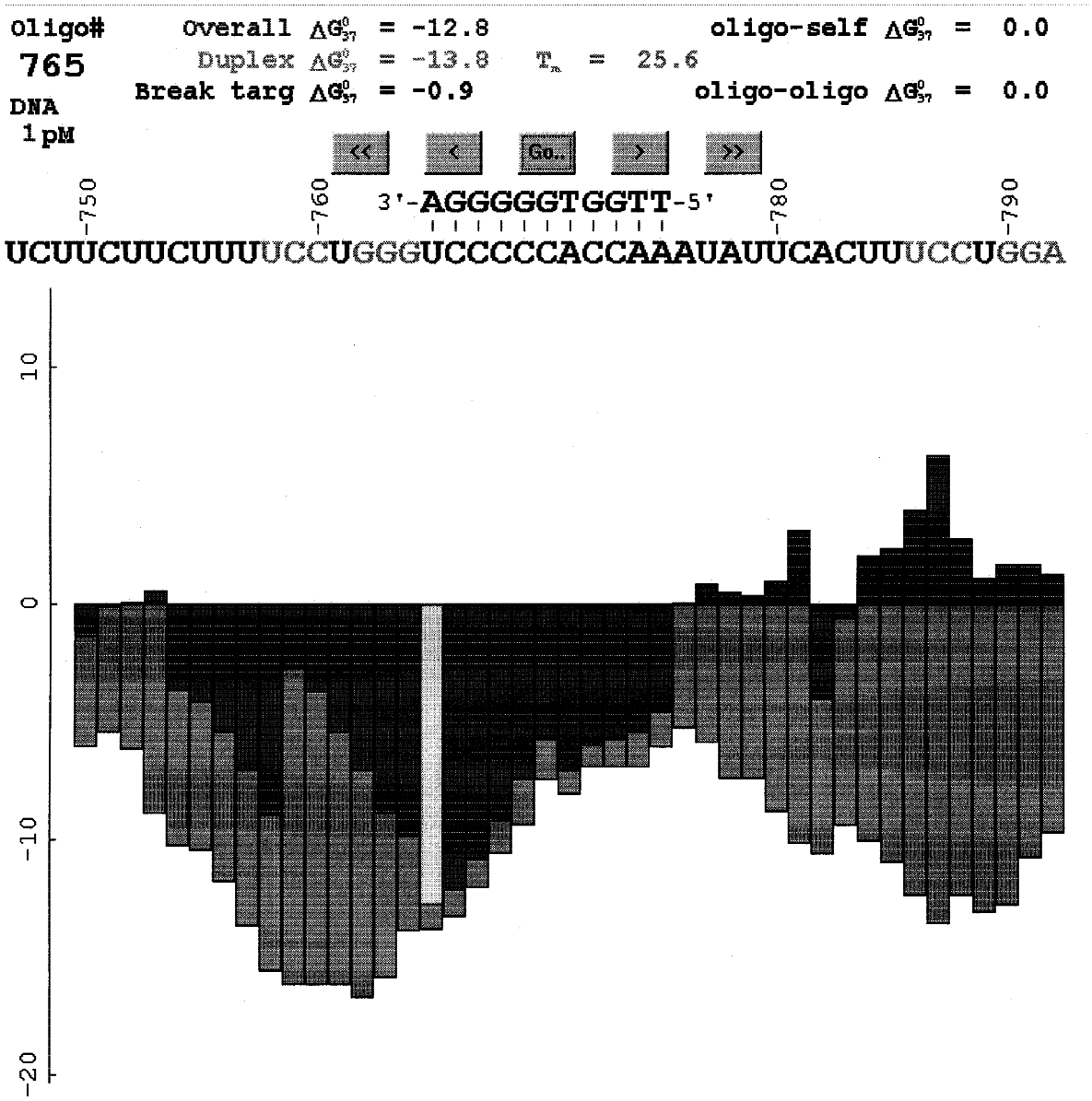

FIGURE 3. A view of the OligoWalk display window. OligoWalk displays predicted free-energy changes in the Microsoft Windows environment. Shown here is the oligomer predicted to have the highest affinity to the angiotensin type-I mRNA (GenBank accession number X62295). The dark bars graph the $\Delta G^\circ_{overall}$ for each possible 11-mer with its 3′ end paired to the nucleotide shown above the bar. The grey bars are the corresponding ΔG°_2 (duplex). The lightest bar indicates the $\Delta G^\circ_{overall}$ for the currently displayed oligomer (#765). The numbering of the sequence is for the coding region of the gene. Above the sequence are the free-energy changes at 37 °C for each equilibrium from Figure 1 and the overall free-energy change. Duplex ΔG°_{37} is ΔG°_3; Break targ ΔG°_{37} is ΔG°_2; oligo-self ΔG°_{37} is ΔG°_{1U}; and oligo-oligo ΔG_{37} is ΔG°_{1B}. Lighter nucleotides in the target sequence are base paired in the lowest free-energy structure.

relevant to the problem of designing effective antisense oligomers because a major mechanism of antisense function is the stimulation of target RNA cleavage by endogenous RNase-H (Chiang et al., 1991). Ho et al. (1998) developed a method of RNase-H cleavage involving semirandom DNA oligomers containing phosphodiester and phosphorothioate linkages, for which the exact footprint of bound oligomers can be inferred. This method was used to map the RNase-H activating DNA 11-mers to the angiotensin type-I receptor mRNA, AT_1 (GenBank accession number X62295).

OligoWalk was used to predict the affinity of 11-mers to the same sequence used for the RNase mapping, a 1,253-nt transcript encompassing the entire coding region of the AT_1 receptor mRNA. First, a set of optimal and suboptimal target structures was predicted using RNAstructure (Mathews et al., 1999). Then, OligoWalk calculations were run for an oligomer concentration of 1 pmol (the concentration of each oligomer in the experimental assay), considering local target structure only, and considering the set of suboptimal target structures (Form II of Fig. 2).

The predicted $\Delta G^\circ_{overall}$ values range from −12.8 to +13.4 kcal/mol. The two oligomers with the lowest $\Delta G^\circ_{overall}$, therefore predicted to have the highest target affinities, correspond to the longest accessible region determined by RNase-H mapping (Ho et al., 1998). Figure 4A shows the partitioning of all possible complementary 11-nt oligomers into seven categories of overall free energy of binding, $\Delta G^\circ_{overall}$. It also plots the percentage of those oligomers that stimulated RNase-H cleavage, demonstrating a correlation between $\Delta G^\circ_{overall}$ and RNase-H-demonstrated binding. The correlation suggests that the predictions could be useful for reducing the number of oligomers surveyed by tenfold or more while still identifying a large number of oligomers that are capable of binding to target. Figure 4B shows the partitioning of the oligomers into six

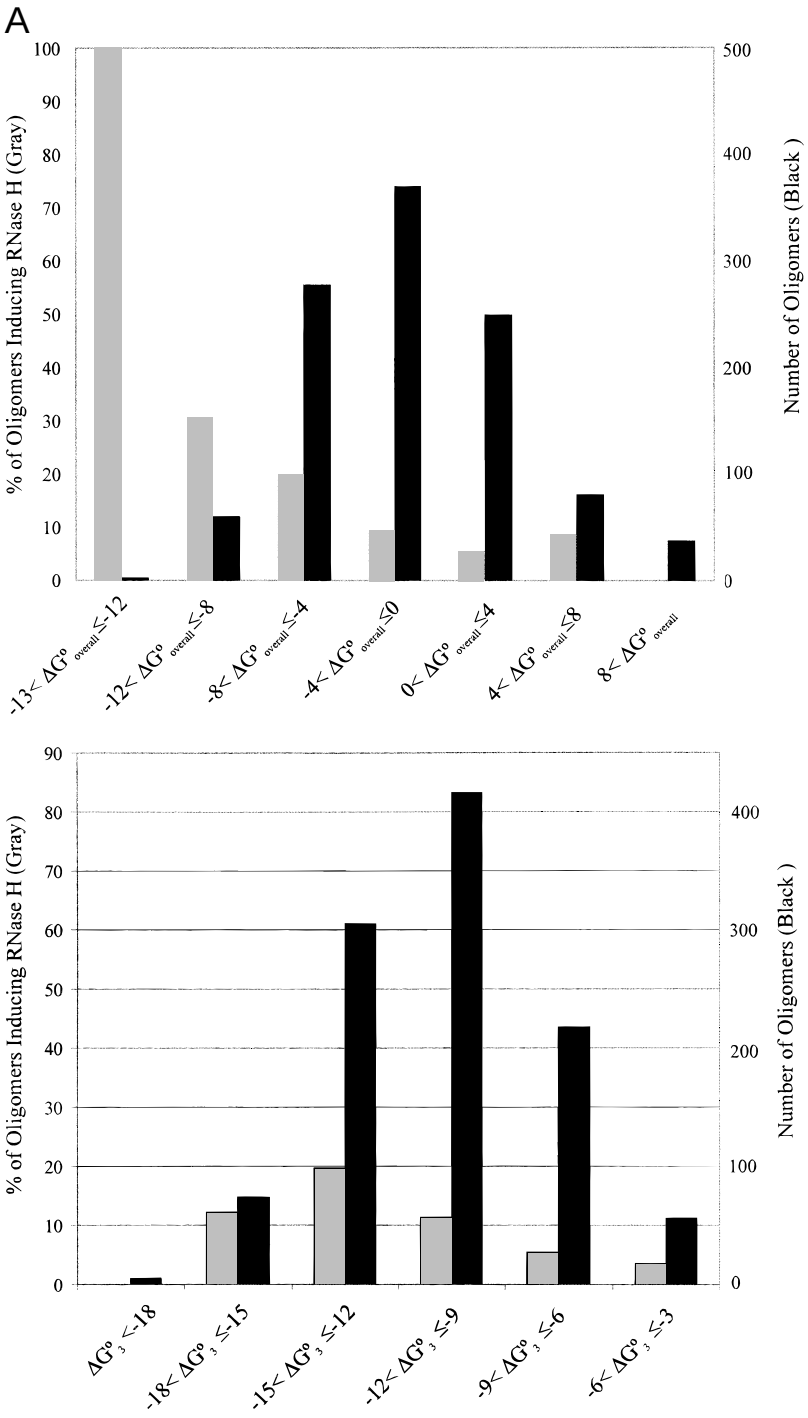

FIGURE 4. Comparison of free-energy predictions with sequences inducing RNase-H cleavage. **A**: The 11-mer complementary oligomers to the coding region of AT_1 are divided into seven bins by overall free energy of binding (in kcal/mol). **B:** The complementary oligomers are divided into bins of duplex free energy, ΔG°_3. The black bars show the total number of oligomers within a range of free energy. The gray bars plot the percentage of those oligomers that are known to bind accessible sites as based on RNase-H mapping (Ho et al., 1998). These plots illustrate that $\Delta G^\circ_{overall}$ correlates better to accessibility than ΔG°_3.

categories of duplex free energy, ΔG°_3. This plot shows little correlation between duplex free energy and RNase-H-demonstrated binding and therefore shows the utility of calculating $\Delta G^\circ_{overall}$, which considers both oligomer and target self structure.

Antisense efficacy

Ho et al. (1996) examined the antisense efficacy of 22 DNA phosphorothioate 20-mers against the human multidrug resistance-1 mRNA (GenBank accession number M14758), which encodes for P-glycoprotein. Twenty of the 22 antisense oligomers were designed to target regions that were determined to be accessible by an RNase-H assay. The effect of antisense oligonucleotides on production of P-glycoprotein was evaluated with a rhodamine assay. Rhodamine, detectable by fluorescence, is actively pumped from cells by P-glycoprotein, and loss of fluorescence is correlated with an antisense-mediated mRNA reduction.

The secondary structure of the first 800 nt of the target mRNA was predicted (Mathews et al., 1999). Ho et al. (1996) report that RNase-H experiments with 800-, 1,259-, and 3,000-nt transcripts demonstrate the same accessibility. OligoWalk was run with DNA oligomers at a concentration of 1 μM (an upper bound on the intracellular concentration), considering local structure only, and considering the set of suboptimal target structures (Form II from Fig. 2).

The antisense efficacy did not correlate to the $\Delta G^\circ_{overall}$. In this case, a correlation is present between ΔG°_3 (duplex stability) and antisense efficacy, with a correlation coefficient of 0.91. Figure 5 shows the percent inhibition plotted against predicted duplex stability.

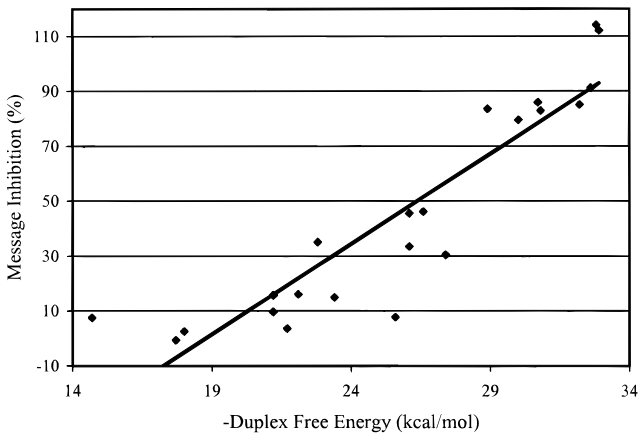

FIGURE 5. Percent inhibition of message as a function of duplex free energy. Antisense efficacy (as percent inhibition) of 22 20-mer DNA oligomers plotted against predicted duplex stability, ΔG°_3. The efficacy data are from Ho et al. (1996) and the line is a best fit to the data. The percent inhibition by antisense oligomer is relative to verapamil, a competitive inhibitor of P-glycoprotein, and so it is possible to have greater than 100% efficacy (Ho et al., 1996). The antisense efficacy did not correlate with $\Delta G^\circ_{overall}$.

This suggests that, when designing antisense oligonucleotides to a target with experimentally determined accessibilities, duplex stability is an important parameter for predicting efficacy.

Accessibility to ribozyme

Lan et al. (1998) used a ribozyme to repair sickle β-globin mRNA mutations. This requires a ribozyme with a correct β-globin exon sequence on the 3′ end and a 5′ internal guide sequence (IGS) of 6 nt complementary to the target mRNA upstream of the sickle mutation. Because of constraints on the ribozyme reaction, the 5′ end of the IGS must be a G that pairs to a U in the target.

Lan et al. (1998) used a library of ribozymes with random internal guide sequences to screen by cloning for suitable IGS complementary regions. U61 in the mRNA sequence was identified as a successful target in five of nine clones for both cell-free and cellular screening and it was the only position to be identified in both the cell-free and cellular experiments.

OligoWalk was used to calculate overall binding affinity for RNA hexamers. The 626-nt secondary structure for the whole target was predicted by free-energy minimization (Mathews et al., 1999). OligoWalk calculations were made with an oligomer concentration of 1 μM (an upper bound on the intracellular ribozyme concentration), considering local structure only, and considering the set of suboptimal target structures (Form II from Fig. 2). There are 16 Us in the 70-nt region 5′ to the sickle mutation, and, therefore, the 16 oligomers with 5′ As were considered as potential IGS mimics. This is reasonable because the free-energy increments for terminal G-U and A-U pairs are similar (Freier et al., 1986). The oligomer with the second lowest $\Delta G^\circ_{overall}$ is the oligomer that binds to U61, the target found most often by screening (see Table 1; Lan et al., 1998). The oligomer with the lowest $\Delta G^\circ_{overall}$ also binds to a region that was accessible to splicing. These two hexamers predicted to be most favorable account for 11 of the 15 clones that showed splicing to the region 5′ to the sickle mutation. The range of overall free energies of binding for oligomers with a 5′ A was -3.9 to $+5.0$ kcal/mol. The five target regions of mRNA, 5′ to position 70, identified by either cell-free or cellular screening by at least one clone each ranged between -3.9 and $+0.6$ kcal/mol in $\Delta G^\circ_{overall}$.

DISCUSSION

OligoWalk extends prior methods for predicting oligomer-target affinity (Rychlik & Rhoads, 1989; Stull et al., 1992) by quantifying more structural considerations. The program by Rychlik and Rhoads (1989) considers duplex stability and simple oligomer self structures. The method of Stull et al. (1992) produces histograms for the free energy of duplex stability and the

TABLE 1. The affinity of 5′A oligonucleotides to sickle β-globin mRNAs.

Target U	Oligomer	$\Delta G^\circ_{overall}$ (kcal/mol)	ΔG°_3 (kcal/mol)	ΔG°_2 (kcal/mol)	ΔG°_{1B} (kcal/mol)	Cell-free clones	Cellular clones
68	AGGAGU	−3.9	−8.0	−4.1	0	1	
61	AGGUGC	−3.7	−9.0	−5.3	−0.8	5	5
65	AGUCAG	−2.9	−6.8	−3.9	0		
52	AUGGUG	−2.3	−6.7	−4.4	0		
6	AAAUGU	−1.6	−3.1	−1.4	0		
25	ACACAG	−1.3	−6.6	−5.3	0		
70	ACAGGA	−1.3	−8.0	−6.7	0		
30	AGUGAA	−1.3	−5.6	−4.3	0		1
23	ACAGUU	−1.2	−5.4	−4.2	0		
21	AGUUGU	−1.2	−5.4	−4.2	0		
38	AGGUUG	−0.6	−6.5	−5.9	0	2	
26	AACACA	0.6	−5.4	−6.0	0		1
10	AAGCAA	1.6	−5.3	−6.9	−2.7		
9	AGCAAA	1.6	−5.3	−6.9	−2.7		
12	AGAAGC	4.6	−6.8	−11.4	0		
55	ACCAUG	5.0	−6.7	−11.7	−1.8		

The oligomers with 5′ adenines are sorted by $\Delta G^\circ_{overall}$. The cell-free and cellular clones columns report the number of clones found that spliced at the U in the region 5′ to the sickle mutation (Lan et al., 1998). The ΔG°_{1U} was 0 kcal/mol for each oligomer shown.

free-energy loss of breaking target structure. These histograms are then used to identify oligomers that may bind tightly. OligoWalk predicts the overall free energy of binding by accounting for duplex stability, oligomer self structure, and target structure using the most current sets of thermodynamic parameters for nucleic acid secondary structure (Sugimoto et al., 1995; SantaLucia et al., 1996; SantaLucia, 1998; Xia et al., 1998; Mathews et al., 1999; Peyret et al., 1999).

OligoWalk calculations are illustrated using three types of experimental data: the accessibility to DNA oligomers as defined by RNase-H cleavage (Ho et al., 1998) and antisense efficacy (Ho et al., 1996) and the accessibility to splicing by a ribozyme (Lan et al., 1998). Several additional sets of experimental data are available that could have been chosen to demonstrate the application of this program (Chiang et al., 1991; Bennett et al., 1994; Dean et al., 1994; Stepkowski et al., 1994; Zarrinkar & Williamson, 1994; Duff et al., 1995; Lee et al., 1995; d'Hellencourt et al., 1996; Miraglia et al., 1996; Stewart et al., 1996; Matveeva et al., 1997; Tu et al., 1998). The three that were chosen cover a variety of experimental techniques, each of which required oligonucleotide binding to an RNA target. These three examples contained data for many oligonucleotides and also used RNA targets of less than 1,300 nt in length to allow for relatively accurate predictions of target structure. For the three examples used, the overall free energy of binding, $\Delta G^\circ_{overall}$, correlated with experimental data in the RNase-H and ribozyme comparisons. For the comparison to antisense efficacy, duplex stability, ΔG°_3, correlated with efficacy, but the experimentally tested oligomers were preselected to target accessible regions based on an Rnase-H assay.

Although correlations are found between predicted thermodynamics and experimental data, the OligoWalk calculation does not predict all oligomer effects. For example, even though the oligomers predicted to have the most favorable overall free-energy-induced RNase-H cleavage and were useful as a ribozyme IGS, other low free-energy oligomers were not effective. Moreover, for the IGS mimic to U26, the overall free energy of binding was predicted to be unstable, but it functioned in the cellular screen (Table 1).

These discrepancies presumably arise from some of the approximations inherent in OligoWalk. First, for these three tests, the target secondary structures were predicted by free-energy minimization only. For RNAs with known secondary structures, the lowest free-energy structure contains 73% of known base pairs and the best suboptimal structure contains 86% of known base pairs on average (Mathews et al., 1999). Structural inaccuracies in the target secondary structure result in an error in ΔG°_2. Presumably, OligoWalk's predictive power will be improved when mapping experiments and/or sequence comparison are used to constrain the predicted target secondary structure.

A second limitation in OligoWalk is the simplified model. The equilibrium model shown in Figure 1 assumes that the oligomers will be self structured in only two conformations at most, one intramolecular and one intermolecular. For longer oligomers, on the order of 20 or more nucleotides, several favorable oligomer conformations may compete with binding to the target. In the case of RNase-H cleavage assays, a complex library of DNA oligomers is employed where there could be intermolecular interactions between oligomers of different sequences.

OligoWalk considers only secondary structure in thermodynamic predictions. Tertiary structure may also play a role in stabilizing or destabilizing the binding of some oligomers. For example, the binding strength of the IGS with 5′ exons of various ribozymes is enhanced by tertiary interactions that OligoWalk does not consider (Bevilacqua & Turner, 1991; Pyle & Cech, 1991; Strobel & Cech, 1995; Testa et al., 1997, 1998).

For both the RNase-H assay (Ho et al., 1998) and antisense efficacy experiments (Ho et al., 1996), phosphorothioate linkages replaced some or all of the natural phosphodiester linkages, respectively. This substitution is known to decrease the stability of binding (Freier & Altmann, 1997), but the oligomers were approximated with natural phosphodiester-linked DNA bases for the OligoWalk calculations because nearest-neighbor parameters are not available for RNA-phosphorothioate helices.

Antisense efficacy may be inherently difficult to predict because of other complications. For example, the region targeted by an antisense oligonucleotide may need to have some biological relevance, such as an AUG start codon or regulatory region. Proteins may cover certain regions of the target. Oligomer sequence and self structure may influence delivery or cellular localization. Furthermore, it is not clear whether or not kinetics play a role in cellular efficacy. In vitro binding experiments are not highly correlated with cell-based data. In the case of C-raf mRNA, the correlation coefficient between in vitro mRNA accessibility, measured with an RNase-H assay, and cellular antisense activity for 20 oligomers was only 0.61 (Matveeva et al., 1998).

OligoWalk provides a method for making comparisons between thermodynamic predictions and experimental results. It is an improvement over prior methods because it calculates the overall binding free energy, taking into account the target structure, oligonucleotide structure, and duplex stability. The relative thermodynamic contributions of unimolecular and bimolecular oligonucleotide structure are considered by including oligonucleotide concentration in the overall free-energy calculation. Although many factors are neglected by OligoWalk, correlations between predictions and experimental results suggest that the program may serve as a useful tool in conjunction with experimental techniques for the selection of oligonucleotides designed to bind RNA. The limited data presented here suggest it may have predictive value, but its true utility for prediction awaits extensive comparisons between high quality experimental data and these predictions. Such comparisons should help identify other factors that are important for predicting oligonucleotide binding affinity and antisense efficacy.

OligoWalk is part of the RNAstructure software package, available for free download at the Turner Lab Homepage (http://rna.chem.rochester.edu). RNAstructure is a program for Microsoft Windows 95, 98, or NT that was programmed in C++. An intermolecular secondary structure prediction feature is also added to the package.

MATERIALS AND METHODS

Effective equilibrium between structured and unstructured oligomer

The equilibrium in Figure 1 shows the steps involved in the binding of oligomer to target. The quantity of interest is the "overall equilibrium constant" for the association of native oligomer and native target in an oligomer-target complex:

$$K_{overall} = \frac{[O - T]}{[O_{unbound}][T_{unbound}]}, \quad (7)$$

where

$$[O_{unbound}] = [O_{total}] - [O\text{-}T] = [O_{F-U}] + 2[O_{F-B}] + [O_U] = [O_F] + [O_U], \quad (8)$$

$$[T_{unbound}] = [T_{total}] - [O\text{-}T] = [T_F] + [T_U], \quad (9)$$

$[O_F]$ is the total concentration of folded oligomer strands, counting both unimolecular and bimolecular structures. $K_{overall}$ and its associated $\Delta G^\circ_{37\ overall}$ quantify the preference of oligomer to bind target despite self structures.

Using equations 8 and 9, equation 7 can be expressed as:

$$K_{overall} = \frac{[O - T]}{([O_F] + [O_U])([T_F] + [T_U])}. \quad (10)$$

Given the free energies for each equilibrium constant, the overall equilibrium constant can be calculated. From the definition of the equilibrium constant:

$$K_{1B} = \frac{[O_{F-B}]}{[O_U]^2}, \quad (11)$$

$$K_{1U} = \frac{[O_{F-U}]}{[O_U]}, \quad (12)$$

$$K_2 = \frac{[T_F]}{[T_U]}, \quad (13)$$

$$K_3 = \frac{[O - T]}{[O_U][T_U]}. \quad (14)$$

If $[O_{total}] \gg [T_{total}]$, then $[O_{total}] \gg [O\text{-}T]$ and $[O_U]$ can be expressed as a function of K_{1B}, K_{1U}, and $[O]_{total}$, the total oligomer concentration:

$$[O_{total}] = [O_U] + 2[O_{F-B}] + [O_{F-U}]; \quad (15)$$

substituting 11 and 12 into 15 yields

$$[O_{total}] = [O_U] + 2[O_U]^2 K_{1B} + [O_U] K_{1U}. \quad (16)$$

Equation 16 is solved for $[O_U]$ with the quadratic equation

$$[O_U] = \frac{-1 - K_{1U} \pm \sqrt{(1 + K_{1U})^2 + 8K_{1B}[O_{total}]}}{4K_{1B}}. \quad (17)$$

The negative solution to equation 17 is rejected because the concentration of single-stranded oligomer must be greater than or equal to zero. Define an effective equilibrium constant, K_{1eff}, as

$$K_{1eff} = \frac{[O_F]}{[O_U]} = \frac{[O_{total}] - [O_U]}{[O_U]} = \frac{[O_{total}]}{[O_U]} - 1; \quad (18)$$

substituting equation 17 into equation 18 yields

$$K_{1eff} = \frac{[O_F]}{[O_U]} = \left(\frac{4K_{1B}[O_{total}]}{-1 - K_{1U} + \sqrt{(1 + K_{1U})^2 + 8K_{1B}[O_{total}]}} \right) - 1. \quad (19)$$

When $K_{1U} \gg [O_{total}]K_{1B}$, equation 19 is numerically unstable. In this case, unimolecular folding is much more important than bimolecular folding of the oligomer and $K_{1eff} \cong K_{1U}$. Now, using equations 13, 14, and 18, equation 10 can be written as

$$K_{overall} = \frac{K_3}{(K_{1eff} + 1)(K_2 + 1)}, \quad (20)$$

which means[3]

$$\Delta G^\circ_{overall} = \Delta G^\circ_3 + RT[\ln(K_{1eff} + 1) + \ln(K_2 + 1)]. \quad (21)$$

T_m calculation for oligomer-target duplex

OligoWalk calculates a melting temperature, T_m, for the duplex formation between oligomer and target, K_3 in Figure 1. This calculation neglects the self structure in the oligomer and target and assumes that $[O_U] \gg [T_U]$. The T_m is then the temperature at which half of the target is bound:

$$[O\text{-}T] = [T_U]. \quad (22)$$

Therefore, from equations 22 and 14, at T_m,

$$K_3 = \frac{1}{[O_{total}]}. \quad (23)$$

Using equation 23 and

$$\Delta G^\circ_{37} = \Delta H^\circ_3 - T\Delta S^\circ_3 = -RT \ln K_3, \quad (24)$$

where ΔH°_3 and ΔS°_3 are calculated from nearest neighbor parameters for oligo-target binding, then

$$T_m\,(\text{in } ^\circ\text{C}) = \frac{\Delta H^\circ}{\Delta S^\circ + R \ln[O]_{total}} - 273.15. \quad (25)$$

Here the -273.15 converts the T_m from Kelvin to centigrade. Note that the T_m is dependent on the oligomer concentration.

Bimolecular secondary structure prediction

To predict a secondary structure formed by two oligomers of sequence S1 and S2 of equal concentration, a third sequence, S, is created. S starts with S1, continues with a 3-nt linker region, and finishes with S2. S is submitted to a modified version of the dynamic programming algorithm of Zuker and co-workers (Zuker, 1989a; Zuker & Stiegler, 1981). The intermolecular linker is not allowed to base pair and can appear in four types of loops: hairpin loops, bulge/internal loops, multibranch loops (junctions), or exterior loops. A loop that contains the intermolecular linker is not actually a loop because it contains the ends of both sequences, so its free energy is that of intermolecular initiation plus the free energy of dangling ends.

The constraint array, *Fce*(*i*, *j*), used to apply folding constraints for unimolecular folding (Mathews et al., 1999), is used to speed bimolecular folding. *Fce*(*i*, *j*) is a triangular character array ($i > j$) that is set to a unique value if and only if the intermolecular linker occurs between nucleotides *i* and *j*. The single-stranded regions of hairpin loops and bulge/internal loops are checked for the intermolecular linker by consulting *Fce*(*i*, *j*) for single-stranded regions from $i + 1$ to $j - 1$. The unique value indicates that the loop really contains the ends of the sequences.

When the intermolecular linker appears in an exterior loop region, the two sequences, S1 and S2, each form intramolecular secondary structures independent of each other. For this case, the intermolecular initiation free energy is not added to the free energies of the structures.

To correctly model the free energy of multibranch loops that contain the intermolecular linker, a new free-energy array, W2, is added to the original V and W arrays (Zuker, 1989b; Zuker & Stiegler, 1981). W2 is the minimum of six quantities:

$$W2(i,j) = \text{minimum}(W2_1, W2_2, W2_3, W2_4, W2_5, W2_6), \quad (26)$$

[3] In cases when no stable secondary structures can be formed by the oligomer, the oligomer structure predicted by the dynamic programming algorithm is the single strand and the program assigns a $\Delta G^\circ = 0$ kcal/mol. This is the unpaired structure, available to bind with the target and therefore K_{leff} is replaced with 0. Similarly, when the target nucleotides are unpaired before binding by the oligomer, K_2 is replaced with 0.

where

$$W2_1(i,j) = V(i,j) + \text{penalty}(i,j) + K,$$

$$W2_2(i,j) = V(i+1,j) + Ed(j,i+1,i) + \text{penalty}(i+1,j) + K,$$

$$W2_3(i,j) = V(i,j-1) + Ed(j-1,i,j) + \text{penalty}(i,j-1) + K,$$

$$W2_4(i,j) = V(i+1,j-1) + Ed(j,i+1,i) + Ed(j-1,i,j) + \text{penalty}(i,j-1) + K,$$

$$W2_5(i,j) = W2(i,j-1),$$

if j is not the middle of the intermolecular linker,

$$= W2(i,j-1) + \text{intermolecular initiation} - K,$$

if j is the middle of the linker,

$$W2_6(i,j) = W2(i+1,j),$$

if i is not the middle of the intermolecular linker,

$$= W2(i+1,j) + \text{intermolecular initiation} - K,$$

if i is the middle of the linker.

$V(i,j)$ is the lowest free energy for the nucleotide segment from i to j with i paired to j. $Ed(i,j,k)$ is the free energy of nucleotide k dangling on the terminal pair of i and j. K is a large integer, 16,000, chosen so that twice K does not overflow a short integer variable. Penalty(i,j) is the free-energy penalty for a terminal A-U or G-U base pair or zero for G-C closing pairs (Xia et al., 1998).

W is also modified to include conditions about the intermolecular linker:

$$W(i,j) = \text{minimum}(W_1, W_2, W_3, W_4, W_5, W_6), \quad (27)$$

where

$$W_1(i,j) = V(i,j) + \text{penalty}(i,j) + c,$$

$$W_2(i,j) = V(i+1,j) + Ed(j,i+1,i) + \text{penalty}(i+1,j) + b + c,$$

$$W_3(i,j) = V(i,j-1) + Ed(j-1,i,j) + \text{penalty}(i,j-1) + b + c,$$

$$W_4(i,j) = V(i+1,j-1) + Ed(j,i+1,i) + Ed(j-1,i,j) + \text{penalty}(i,j-1) + 2b + c,$$

$$W_5(i,j) = W(i,j-1) + b,$$

if j is not the middle of the intermolecular linker,

$$= W(i,j-1) + K, \text{ if } j \text{ is the middle of the linker,}$$

$$W_6(i,j) = W(i+1,j) + b,$$

if i is not the middle of the intermolecular linker,

$$= W(i+1,j) + K, \text{ if } i \text{ is the middle of the linker.}$$

The constants a, b, and c are from the free-energy approximation for a multibranch loop (Mathews et al., 1999):

$$\Delta G^\circ \text{ (multibranch loop)} = a + bN + cH, \quad (28)$$

with N, the number of unpaired nucleotides in the loop, and H, the number of helices exiting the loop. These modifications will result in an unfavorable W if it includes the intermolecular linker in the loop, while W2 is unfavorable if it does not encompass the intermolecular linker.

The array WM is defined as usual (Mathews et al., 1999):

$$WM(i,j) = \min[W(i,k) + W(k+1,j)], \text{ with } i \le k < j. \quad (29)$$

A second array, WM2 is defined similarly for W2:

$$WM2(i,j) = \min[W2(i,k) + W2(k+1,j)], \text{ with } i \le k < j. \quad (30)$$

The lowest free energy for a multibranch loop closed by a pair between i and j is then VM, defined as the minimum of eight terms:

$$VM = \min(VM_1, VM_2, VM_3, VM_4, VM_5, VM_6, VM_7, VM_8), \quad (31)$$

where

$$VM_1(i,j) = WM(i+1,j-1) + \text{penalty}(i,j),$$

$$VM_2(i,j) = a + b + c + Ed(i,j,i+1) + WM(i+2,j-1) + \text{penalty}(i,j),$$

$$VM_3(i,j) = a + b + c + Ed(i,j,j-1) + WM(i+1,j-2) + \text{penalty}(i,j),$$

$$VM_4(i,j) = a + 2b + c + Ed(i,j,i+1) + Ed(i,j,j-1) + W(i+2,j-2) + \text{penalty}(i,j),$$

$$VM_5(i,j) = WM2(i+1,j-1) + \text{penalty}(i,j),$$

$$VM_6(i,j) = Ed(i,j,i+1) + WM2(i+2,j-1) + \text{penalty}(i,j),$$

$$VM_7(i,j) = Ed(i,j,j-1) + WM2(i+1,j-2) + \text{penalty}(i,j),$$

$$VM_8(i,j) = Ed(i,j,i+1) + Ed(i,j,j-1) + W2(i+2,j-2) + \text{penalty}(i,j).$$

In this way, if the minimum is from terms 1–4, it is a true multibranch loop with the correct penalty terms. If the minimum is from terms 5–8, it contains the bimolecular linker, and intermolecular initiation is substituted for the multibranch loop penalties.

Time requirements

The run time for OligoWalk scales with $O(N^2)$, where N is the length of the sequence, if only local structure is considered

upon oligomer binding. With refolding, the run time scales with $O(N^4)$. For the sickle hemoglobin mRNA, 626 nt, the calculation considering only local structure and oligomers of 6 nt takes less than 4 min on a Pentium-II 233 MHZ computer with 64 MB of RAM. When refolding the whole sequence upon oligomer binding, the calculation takes about 3.5 h.

Prediction of RNA secondary structures

The RNA target secondary structures were predicted by free-energy minimization (Mathews et al., 1999). Suboptimal structures were included with a 20% sort, window size of zero, and a maximum of 750 structures.

ACKNOWLEDGMENTS

The authors thank Dr. S.M. Testa for helpful comments about the OligoWalk user interface and T.W. Barnes for helpful discussions. This work was supported by National Institutes of Health Grant GM22939 to D.H.T. D.H.M. is a Messersmith Fellow and M.E.B. is a Hooker Fellow. D.H.M. and M.E.B. are trainees in the Medical Scientist Training Program funded by National Institutes of Health Grant T32 GM07356.

Received June 2, 1999; returned for revision July 2, 1999; revised manuscript received August 16, 1999

REFERENCES

Banerjee AR, Jaeger JA, Turner DH. 1993. Thermal unfolding of a group I ribozyme: The low temperature transition is primarily a disruption of tertiary structure. *Biochemistry 32*:153–163.

Bennett CF, Condon TP, Grimm S, Chan H, Chiang M. 1994. Inhibition of endothelial cell adhesion molecule expression with antisense oligonucleotides. *J Immunol 152*:3530–3540.

Bevilacqua PC, Turner DH. 1991. Comparison of binding of mixed ribose-deoxyribose analogues of CUCU to a ribozyme and to GGAGAA by equilibrium dialysis: Evidence for ribozyme specific interactions with 2′ OH groups. *Biochemistry 30*:10632–10640.

Birikh KR, Berlin YA, Soreq H, Eckstein F. 1997. Probing accessible sites for ribozymes on human acetylcholinesterase RNA. *RNA 3*:429–437.

Chang KY, Tinoco I Jr. 1994. Characterization of a "kissing" hairpin complex derived from the human immunodeficiency virus genome. *Proc Natl Acad Sci USA 91*:8705–8709.

Chang KY, Tinoco I Jr. 1997. The structure of an RNA "kissing" hairpin complex of the HIV TAR hairpin loop and its complement. *J Mol Biol 269*:52–66.

Chiang M, Chan H, Zounes MA, Freier SM, Lima WF, Bennett CF. 1991. Antisense oligonucleotides inhibit intercellular adhesion molecule 1 expression by two distinct mechanisms. *J Biol Chem 266*:18162–18171.

Comolli LR, Pelton JG, Tinoco I Jr. 1998. Mapping of a protein-RNA kissing hairpin interface: Rom and Tar-Tar*. *Nucleic Acids Res 26*:4688–4695.

Crothers DM, Cole PE, Hilbers CW, Schulman RG. 1974. The molecular mechanism of thermal unfolding of *Escherichia coli* formylmethionine transfer RNA. *J Mol Biol 87*:63–88.

d'Hellencourt CL, Diaw L, Cornillet P, Guenounou M. 1996. Inhibition of TNFα and LT in cell-free extracts and in cell culture by antisense oligonucleotides. *Biochim Biophys Acta 1317*:168–174.

Dean NM, McKay R, Condon TP, Bennett CF. 1994. Inhibition of protein kinase C-α expression in human A549 cells by antisense oligonucleotides inhibits induction of intercellular adhesion molecule 1 (ICAM-1) mRNA by phorbol esters. *J Biol Chem 269*:16416–16424.

Duff JL, Monia BP, Berk BC. 1995. Mitogen-activated protein (MAP) kinase is regulated by the MAP kinase phosphatase (MKP-1) in vascular smooth muscle cells. *J Biol Chem 270*:7161–7166.

Ehresmann C, Baudin F, Mougel M, Romby P, Ebel J, Ehresmann B. 1987. Probing the structure of RNAs in solution. *Nucleic Acids Res 15*:9109–9128.

Freier SM, Altmann K. 1997. The ups and downs of nucleic acid duplex stability: Structure-stability studies on chemically-modified DNA:RNA duplexes. *Nucleic Acids Res 25*:4429–4443.

Freier SM, Kierzek R, Caruthers MH, Neilson T, Turner DH. 1986. Free energy contributions of G·U and other terminal mismatches to helix stability. *Biochemistry 25*:3209–3223.

Gregorian RSJ, Crothers DM. 1995. Determinants of RNA hairpin loop-loop complex stability. *J Mol Biol 248*:968–984.

Hilbers CW, Robillard GT, Shulman RG, Blake RD, Webb PK, Fresco R, Riesner D. 1976. Thermal unfolding of yeast glycine transfer RNA. *Biochemistry 15*:1874–1882.

Ho SP, Bao Y, Lesher T, Malhotra R, Ma LY, Fluharty SJ, Sakai RR. 1998. Mapping of RNA accessible sites for antisense experiments with oligonucleotide libraries. *Nat Biotechnol 16*:59–63.

Ho SP, Britton DHO, Stone BA, Behrens DL, Leffet LM, Hobbs FW, Miller JA, Trainor GL. 1996. Potent antisense oligonucleotides to the human multidrug resistance-1 mRNA are rationally selected by mapping RNA-accessible sites with oligonucleotide libraries. *Nucleic Acids Res 24*:1901–1907.

Jaeger L, Westhof E, Michel F. 1993. Monitoring of cooperative unfolding of the sunY group I intron of bacteriophage T4. *J Mol Biol 234*:331–346.

James BD, Olsen GJ, Pace NR. 1989. Phylogenetic comparative analysis of RNA secondary structure. *Methods Enzymol 180*:227–239.

Knapp G. 1989. Enzymatic approaches to probing RNA secondary and tertiary structure. *Methods Enzymol 180*:192–212.

Lan N, Howrey RP, Lee S, Smith CA, Sullenger BA. 1998. Ribozyme-mediated repair of sickle β-globin mRNAs in erythrocyte precursors. *Science 280*:1593–1596.

Lee C, Chen HH, Hoke G, Jong JS, White L, Kang Y. 1995. Antisense gene suppression against human ICAM-1, ELAM-1, and VCAM-1 in cultured human umbilical vein endothelial cells. *Shock 4*:1–10.

Lima WF, Brown-Driver V, Fox M, Hanecak R, Bruice TW. 1997. Combinatorial screening and rational optimization for hybridization to folded hepatitis C virus RNA of oligonucleotides with biological antisense activity. *J Biol Chem 272*:626–638.

Mathews DH, Banerjee AR, Luan DD, Eickbush TH, Turner DH. 1997. Secondary structure model of the RNA recognized by the reverse transcriptase from the R2 retrotransposable element. *RNA 3*:1–16.

Mathews DH, Sabina J, Zuker M, Turner DH. 1999. Expanded sequence dependence of thermodynamic parameters provides improved prediction of RNA secondary structure. *J Mol Biol 288*:911–940.

Matveeva O, Felden B, Audlin S, Gesteland RF, Atkins JF. 1997. A rapid in vitro method for obtaining RNA accessibility patterns for complementary DNA probes: Correlation with intracellular pattern and known RNA structures. *Nucleic Acids Res 25*:5010–5016.

Matveeva O, Felden B, Tsodikov A, Johnston J, Monia BP, Atkins JF, Gesteland RF, Freier SM. 1998. Prediction of antisense oligonucleotide efficacy by in vitro methods. *Nat Biotechnol 16*:1374–1375.

Milner N, Mir KU, Southern EM. 1997. Selecting effective antisense reagents on combinatorial oligonucleotide arrays. *Nat Biotechnol 15*:537–541.

Miraglia L, Geiger T, Bennett CF, Dean NM. 1996. Inhibition of interleukin-1 type I receptor expression in human cell-lines by an antisense phosphorothioate oligodeoxynucleotide. *Int J Immunopharm 18*:227–240.

Pace NR, Thomas BC, Woese CR. 1999. Probing RNA structure, function, and history by comparative analysis. In: Gesteland RF, Cech TR, Atkins JF, eds, *The RNA world*, 2nd ed. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press. pp 113–141.

Peyret N, Seneviratne PA, Allawi HT, SantaLucia J Jr. 1999. Nearest-neighbor thermodynamics and NMR of DNA sequences with internal A·A, C·C, G·G, and T·T mismatches. *Biochemistry 38*:3468–3477.

Pyle AM, Cech TR. 1991. Ribozyme recognition of RNA by tertiary interactions with specific ribose 2′-OH groups. *Nature 350*:628–631.

Rivas E, Eddy SR. 1999. A dynamic programming algorithm for RNA structure prediction including pseudoknots. *J Mol Biol 285*:2053–2068.

Rychlik W, Rhoads RE. 1989. A computer program for choosing oligonucleotides for filter hybridization, sequencing and in vitro amplification of DNA. *Nucleic Acids Res 17*:8543–8551.

SantaLucia J Jr. 1998. A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proc Natl Acad Sci USA 95*:1460–1465.

SantaLucia J Jr, Allawi HT. 1997. Thermodynamics and NMR of internal G·T mismatches in DNA. *Biochemistry 36*:10581–10594.

SantaLucia J Jr, Allawi HT, Seneviratne PA. 1996. Improved nearest-neighbor parameters for predicting DNA duplex stability. *Biochemistry 35*:3555–3562.

Stepkowski SM, Tu Y, Condon TP, Bennett CF. 1994. Blocking of heart allograft rejection by intercellular adhesion molecule-1 antisense oligonucleotides alone or in combination with other immunosuppressive modalities. *J Immunol 153*:5336–5346.

Stewart AJ, Canitrot Y, Baracchini E, Dean NM, Deeley RG, Cole SPC. 1996. Reduction of expression of the multidrug resistance protein (MRP) in human tumor cells by antisense phosphorothioate oligonucleotides. *Biochem Pharmacol 51*:461–469.

Strobel SA, Cech TR. 1995. Minor-groove recognition of the conserved G·U pair at the *Tetrahymena* reaction site. *Science 267*:675–679.

Stull RA, Taylor LA, Szoka FC Jr. 1992. Predicting antisense oligonucleotide inhibitory efficacy: A computational approach using histograms and thermodynamic indices. *Nucleic Acids Res 20*:3501–3508.

Sugimoto N, Nakano S, Katoh M, Matsumura A, Nakamuta H, Ohmichi T, Yoneyama M, Sasaki M. 1995. Thermodynamic parameters to predict stability of RNA/DNA hybrid duplexes. *Biochemistry 34*:11211–11216.

Testa SM, Gryaznov SM, Turner DH. 1998. Antisense binding enhanced by tertiary interactions: Binding of phosphorothioate and N3′ → P5′ phosphoramidate hexanucleotides to the catalytic core of a group I ribozyme from mammalian pathogen *Pneumocystis carinii*. *Biochemistry 37*:9379–9385.

Testa SM, Haidaris CG, Gigliotti F, Turner DH. 1997. A *Pneumocystis carinii* group I intron ribozyme that does not require 2′ OH groups on its 5′ exon mimic for binding to the catalytic core. *Biochemistry 36*:15303–15314.

Tu G, Cao Q, Zhou F, Israel Y. 1998. Tetranucleotide GGGA motif in primary RNA transcripts: Novel target site for antisense design. *J Biol Chem 273*:25125–25131.

Xia T, SantaLucia J Jr, Burkard ME, Kierzek R, Schroeder SJ, Jiao X, Cox C, Turner DH. 1998. Parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson–Crick pairs. *Biochemistry 37*:14719–14735.

Zarrinkar PP, Williamson JR. 1994. Kinetic intermediates in RNA folding. *Science 265*:918–924.

Zuker M. 1989a. On finding all suboptimal foldings of an RNA molecule. *Science 244*:48–52.

Zuker M. 1989b. The use of dynamic programming algorithms in RNA secondary structure prediction. In: Waterman MS, ed. *Mathematical methods for DNA sequences*. Boca Raton, Florida: CRC Press. pp 159–184.

Zuker M, Stiegler P. 1981. Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucleic Acids Res 9*:133–148.