Predicting antisense oligonucleotide inhibitory efficacy: a computational approach using histograms and thermodynamic indices

Robert A. Stull, Leslie A. Taylor¹ and Francis C. Szoka, Jr.*

Departments of Pharmacy and Pharmaceutical Chemistry and ¹Computer Graphics Laboratory, University of California, San Francisco, CA 94143-0446, USA

Received December 23, 1991; Revised and Accepted June 12, 1992

ABSTRACT

Antisense oligonucleotides (ASOs) are designed to bind to a specific mRNA and selectively suppress its translation. To facilitate selection of optimal ASO targets, we have developed three thermodynamic indices to evaluate putative structural complexes important in ASO action. These indices are: a secondary structure score (Sscore), which estimates the strength of local mRNA secondary structures at the ASO target site; a duplex score (Dscore), which estimates the $\Delta G_{\text{formation}}$ for the ASO:mRNA target sequence duplex; and a competition score (Cscore), which is the difference between the Dscore and the Sscore. We also present two histograms to graphically display these indices from different regions of the mRNA. The indices are compared to the inhibition reported in five studies of ASO-mediated suppression of gene expression. The Dscore is the most consistent predictor of ASO efficacy in four of the five studies (r^2 from 0.44 to 0.99), while the results of the fifth study could not be predicted by any thermodynamic or physical index. Thus the Dscores and their histogram may prove useful in selection of ASO targets.

INTRODUCTION

Antisense oligodeoxyribonucleotides (ASOs) have become widely used to elucidate gene product function in cell culture (1, 2, 3; for reviews 4, 5, 6). Gene-specific antisense oligomers were first shown to inhibit viral replication in RSV infected chick embryo fibroblasts (7, 8), and this approach has proved applicable to other viral systems (9, 10, 11). Consequently oligonucleotides have been proposed as a novel class of therapeutic agents, especially for the treatment of viral or neoplastic disease (4).

One obstacle to applying ASOs for therapeutic purposes is selection of the most effective target sequences from the many available on a given mRNA. For example, a 1500 nucleotide mRNA sequence affords 1486 potentially different 15-mer ASO target sites. Many experiments have shown that certain ASOs are more effective than others in suppressing gene expression.

A rational method for determining those ASO sequences most likely to produce strong inhibition of gene expression would streamline the empirical approaches currently used to select ASOs.

The efficacy of sequence specific inhibition by an ASO will depend, in part, upon its mechanism of action. These mechanisms include blocking ribosomal binding to and translocation along the mRNA, blocking the binding of regulatory proteins to the mRNA, and providing a double stranded heteroduplex substrate for the enzyme RNase H. In each mechanism, the oligomer must hybridize with the mRNA in order to effect inhibition. We therefore hypothesized that ASO efficacy is primarily related to the ability of the oligomer to hybridize with its mRNA target sequence, and that analysis of factors influencing ASO hybridization would allow prediction of ASO efficacy.

Binding of the ASO to its mRNA target is a competitive process; at its binding site, the ASO competes with proteins or ribosomal binding (external competition) and with mRNA secondary and tertiary structure (internal competition). A reasonable first approach to predicting the hybridization capacity and hence the efficacy of an ASO involves computations of the thermodynamic stability of each of the complexes formed by the competing processes. The binding energies for each of the many proteins and ribosomal structures that interact with the RNA are not currently known. Therefore, not much information is available to evaluate the external competitive factors influencing ASO:mRNA interactions. However, the internal competition can be analyzed using computer algorithms that estimate the thermodynamics of mRNA secondary structure (for review, see 12). By extension, these algorithms also afford analysis of ASO:mRNA duplex formation.

In this paper we describe two histograms which conveniently summarize the output from RNA folding algorithms. We also describe three indices to predict the efficacy of an ASO in suppressing gene expression. The first index (S[tructure]score) is based upon the predicted secondary structure of the mRNA target region. The second index (D[uplex]score) is based upon the duplex formation energy of the ASO and the mRNA target region. The third index (C[ompetition]score) combines secondary

* To whom correspondence should be addressed

structure predictions and the duplex formation energy. The usefulness of these indices was tested by comparing our predictions of ASO efficacy to the results obtained in five published studies of ASO-mediated inhibition of gene expression (13, 14, 15, 16, 17). Of these three indices, we find the Dscores most consistently possess a statistically significant correlation with inhibition reported in these published studies.

METHODS AND ALGORITHMS

Folding Algorithm

The MONTE CARLO mRNA folding algorithm has been described previously (18, 19). We have updated the energy values used during folding to those determined at 37°C, with 1M concentration of reactants and products (20, 21). We have also implemented the equation employed by Jaeger and coworkers for extrapolating the length dependence of loop destabilization energies (21), but do not include their calculations for terminal unpaired nucleotides and mismatches, nor their bonus energies for stable 'tetraloop' sequences, nor their linear approximation for multibranched loops. G-U pairing is the only non-Watson-Crick base pairing accommodated by MONTE CARLO. Due to hardware limitations, sequences folded by MONTE CARLO were generally less than 3300 nucleotides. All MONTE CARLO folding sets were composed of ten foldings of the particular RNA sequence being examined.

Histograms

MONTE CARLO, and many other RNA folding algorithms, generate a population of structures through multiple foldings of a single sequence. GENHISTO was created to summarize these folding sets in various histograms (Fig. 1). In general, sequence information is plotted along the abscissa, while numerical scores are plotted along the ordinate. Two types of histograms were generated: an energy scaled histogram, and a window-score histogram.

Energy Scaled Histograms

We devised a histogram in which the relative stabilities of local secondary structures are plotted. This approach is similar to the P-num and S-num plots described by Jaeger and coworkers (21), but allows a weighting of each nucleotide score according to the energies of the substructures in which it is found.

According to the principle of additivity, the free energy of local structure (stem) formation is defined by

$$\Delta G_{\text{stem}} = \sum \Delta G_{\text{helical regions}} + \sum \Delta G_{\text{bulge}} + \Delta G_{\text{hairpin}} \quad (\text{eq 1})$$

This energy can be redistributed equally among the base-paired nucleotides (helical regions) by defining an energy redistribution function

$$E_3 = \Delta G_{\text{stem}}/B \quad (\text{eq 2})$$

Figure 1. (A) Energy scaled histogram. Base-paired nucleotides are assigned energy scores based upon the distribution functions described in the Materials and Algorithms section. The RNA sequence is plotted along the abscissa, while the energy score of each nucleotide is plotted along the ordinate. This histogram was produced by GENHISTO using upon data supplied by the MONTE CARLO algorithm. (B) Windowing histogram. This histogram displays the Dscore, Cscore (and by extension the Sscore) for each window of m nucleotides across the mRNA. In this figure, $m = 15$. The scores for each window are attached to the first nucleotide appearing in the window. The sequence of the window can then be deduced by reading the next $m-1$ nucleotides downstream in the histogram. For example, the bars beneath the asterisk represent the score for the sequence CUCCCGGAUGCACCA. Dscores, Cscores and Sscores were computed by EWINDOW using source files produced by the folding algorithms. The histogram was generated by GENHISTO using output files produced by EWINDOW.

where the free energy of stem formation is divided by the number of *base-paired* nucleotides, *B*, appearing in the helical regions. The resulting E_3 value is then assigned, as an $E_{3,N}$ score, to each nucleotide in the stem that is base-paired. All base-paired nucleotides in the same stem receive equal $E_{3,N}$ values, while those nucleotides appearing in hairpin, bulge, bifurcation or interior loops receive zero $E_{3,N}$ scores.

The MONTE CARLO algorithm generates n foldings for a given mRNA sequence; hence each nucleotide will receive n $E_{3,N}$ scores. These $E_{3,N}$ scores are weighted to produce WE_3 scores by

$$\overline{WE_{3,N}} = \frac{\left[\sum_{i=1}^n (E_{3,N_i}) * e^{-\frac{(\Delta G_{\text{global},i} - \Delta G_{\text{global},\min})}{RT}} \right]}{n} \quad (\text{eq. 3})$$

The exponential term is a Boltzmann factor designed to weight the local $E_{3,N}$ scores from a given global mRNA folding, i , relative to the other global foldings in the folding set. $\Delta G_{\text{global},i}$ is the free energy of formation for the global mRNA structure produced by folding i , while $\Delta G_{\text{global},\min}$ is the free energy of formation for the most stable global mRNA structure in the folding set. The weighted scores are summed and then normalized by dividing by n , to produce a single, energy-weighted, composite global mRNA structure. The result of these computations is to attach structural energies to individual nucleotides (eq. 2), and to weight these energies according to the strength of the local structure and the stability of the global structure of which the local structure is a part (eq. 3).

An energy scaled histogram (Fig. 1A) is generated when the normalized $WE_{3,N}$ scores are plotted versus the sequence of the mRNA.

EWINDOW program, window-score histograms

A windowing program, EWINDOW, is used to calculate a duplex formation score (Dscore), a competition score (Cscore), and a secondary structure score (Sscore) for each window of length m nucleotides across the entire mRNA sequence.

The Dscore is based upon the $\Delta G_{\text{formation}}$ of a perfectly matched duplex between an RNA oligomer of length m and the mRNA sequence appearing within the window. The $\Delta G_{\text{formation}}$ is computed using the same nearest neighbor model thermodynamic parameters used to calculate the mRNA folding energies, and includes a helix initiation penalty of 3.4 kcal/mole (20). $\Delta G_{\text{formation}}$ is related to the Dscore by

$$\text{Dscore} = \Delta G_{\text{duplex formation}}/2 \quad (\text{eq. 4})$$

This equation is analogous to equation 2 and represents the special case of distributing the formation energy of a perfectly matched duplex (no bulges or loops) over all the participating nucleotides and then summing the values for one strand of the duplex (i.e. the ASO nucleotides).

The secondary structures found for the mRNA sequence present within a given window can be quantified by a Sscore, which is a summing of the normalized $WE_{3,N}$ scores for each nucleotide within the window. We can define a Cscore by

$$\text{Cscore} = \text{Dscore} - \sum_{j=1}^m \overline{(WE_{3,N})_j} = \text{Dscore} - \text{Sscore} \quad (\text{eq. 5})$$

The Cscore is an approximation comparing the equilibrium thermodynamics between the ASO:mRNA duplex and predicted mRNA secondary structure at the ASO target site.

A window-score histogram is obtained by plotting the Dscore and Cscore for each window across the mRNA sequence (Fig. 1B). Dscores appear as grey bars, while Cscores appear in black. Both scores for each window are attached to the *first* nucleotide of the window sequence, so that the complete window sequence can be deduced by reading the next $m-1$ nucleotides downstream. Tabular Cscore data files are also generated for user reference.

Mean Stacking Temperature Calculations

A mean stacking temperature (MST) score was also calculated (22, 23) and tested as a predictive index. Like the Dscore, the MST score is an sequence based estimator of duplex stability. However, it is computed for DNA:DNA duplex rather than a RNA:RNA duplex.

Statistics

Correlations between reported experimental inhibition and predicted inhibition using a specific prediction index (i.e. Dscore, Cscore, Sscore) were determined by fitting the data using a least-squares method provided by a statistical analysis software package. All results are reported as coefficients of determination (r^2). P values were computed using $t = r \sqrt{(n-2)} / \sqrt{(1-r^2)}$. A correlation was rejected as being statistically insignificant if the P value was greater than 0.05.

Inhibition Studies Analyzed and Sequences Examined

We examined five studies of ASO-mediated suppression of gene expression. The genes targeted in these five studies were human dihydrofolate reductase (DHFR; 13), rabbit β -globin (14, 15), human *c-myc* (16) and human intracellular adhesion molecule-1 (ICAM-1; 17, 34). The GenBank™ data base (24) supplied the human genomic dihydrofolate reductase sequence (accession number X00855; 25, 26), the rabbit genomic β -globin sequence (J00659; 27, 28, 29), the human genomic *c-myc* sequence (J00120; 30, 31), and the human ICAM-1 cDNA sequence (M24283; 32, 33). Each sequence was trimmed to account for mRNA transcript cap sites when known; genomic DNA sequences were also converted into cDNA sequences by removal of intron sequences.

Computational Environment

GENHISTO and EWINDOW are available upon request from L. Taylor, email address: ltaylor@cgl.ucsf.EDU, or from R. Stull, email address: bstull@cgl.ucsf.EDU. DSCORES, a version of the windowing program which calculates only the Dscores for various target regions and is Macintosh compatible, is also available from R. Stull.

RESULTS

General approach of experiments

To test if the thermodynamic indices developed above are useful predictors of ASO efficacy, we examined published studies of ASO-mediated inhibition of gene expression and compared the computed indices to the reported inhibition achieved with different ASOs. We analyzed studies carried out both in cell-free translation systems and in cultured cells. Cell free systems afford the simplest analysis of ASO mechanism and effect, since issues

of ASO transport and stability are not necessarily involved in ASO efficacy. Analysis of ASO efficacy in cultured cells may be confounded by transport and stability variables, but obviously provide a more realistic application for the indices.

Analysis of ASO action in Cell-free Translation Systems

Rabbit reticulocyte lysates. The first study we examined involved the human dihydrofolate reductase (DHFR) gene. In this study by Maher and Dolnick (13), mRNA isolated from a methotrexate-resistant, DHFR gene-amplified cell line was translated in rabbit

reticulocyte lysate in the presence of nine different phosphodiester oligonucleotides (Table 1). The correlations between the observed inhibition of DHFR expression and our indices of prediction are shown in Table 2.

The correlations for the indices based upon duplex formation (Dscore; $r^2 = 0.740$ ($P < 0.005$)) and upon competition between duplex formation and mRNA secondary structure present at the target site (Cscore; $r^2 = 0.808$ ($P < 0.001$)) are both significant. We found no significant correlation between predicted secondary structure (Sscore) provided by Martinez's MONTE

Table 1. Antisense phosphodiester oligomers used in cell-free translation systems

Oligo Name	Oligo Length	Oligo Sequence (5' to 3')	Target Site on mRNA	Dscore	Cscore ^a	Inhibition (%)
Human DHFR						
A20	20	GCTCAAGTTTGGCGCGAAAT	5' UTR ^b	-16.3	-14.2	74
B14	14	CTCCGAGCCCGCTC	5' UTR	-14.1	-11.7	40
B16	16	CCTCCGAGCCCGCTCG	5' UTR	-16.5	-13.9	48
B18	18	ACCTCCGAGCCCGCTCGT	5' UTR	-18.6	-15.7	71
B20	20	GACCTCCGAGCCCGCTCGTT	5' UTR	-20.2	-17.2	72
C14	14	AGCAGCGGGAGGAC	5' UTR	-13.9	-12.4	28
C20	20	CATGACAGCAGCGGGAGGAC	start	-19.2	-17.2	80
D11	11	GCGAACCAACC	coding	-9.0	-8.4	75
D20	20	TGCAGTTTAGCGAACCAACC	coding	-16.7	-15.6	23
Rabbit β-globin						
Goodchild <i>et al.</i> (1988) (ref 14)						
β 1	20	GCAAGTAAACACAGTTGTGT	5' UTR	-14.9	-11.4	23
β 2	20	CAGATGCACCATCTGTCTG	start	-16.3	-12.8	61
β 3	20	GGGTAGACAACCAGCAGCCT	coding	-18.9	-14.5	18
β 4	20	TTGTTTATAACAGCATTTGC	coding	-14.0	-9.9	43
β 5	22	CAGTTGTGTCAAAAGCAAGTGT	cap	-16.6	-15.6	67
β 6	23	TTTTGGGGGATTGCAAGTAAACA	5' UTR	-17.7	-13.9	47
β 7	25	CAGTTGTGTCAAAAGCAAGTGTCCC ^c	cap	-16.6	-15.6	75
β 8	48	(β 6 + β 7)	cap	-39.8	-34.2	89
Cazenave <i>et al.</i> (1987) (ref 15)						
17 Glo _[3-19]	17	TTGTGTCAAAAGCAAGT	cap	-11.9	-11.1	85
17 Glo _[51-67]	17	GACAGATGCACCATCT	start	-13.6	-11.0	95
11 Glo	11	TTGTGTCTGTT	start	-7.1	-4.8	75
17 Glo _[113-129]	17	CACCAACTTCTTCCACA	coding	-13.3	-10.6	95

^aFrom foldings using MONTE CARLO.

^bUntranslated region.

^cCap nucleotides complementary to the underlined nucleotides.

Table 2. Coefficients of determination (r^2) between predictive indices and reported inhibition in cell-free systems

Index		Human DHFR ⁽¹³⁾		Rabbit β -Globin	
		All Targets @ 200 mM (n = 9) ^b	Ribosome Competitors ^a @ 200 mM (n = 4) ^b	Goodchild study ⁽¹⁴⁾ @ 5.2 mM (n = 6) ^{b,d}	Cazenave study ⁽¹⁵⁾ @ 6 mM (n = 4) ^c
duplex	Dscore	0.740 ($P < 0.005$)	0.962 ($P < 0.02$)	0.015 (N. S.)	0.933 ($P < 0.05$)
mRNA 2° structure	Sscore	0.181 (N. S.)	0.573 (N. S.)	0.499 (N. S.)	0.129 (N. S.)
competition	Cscore	0.808 ($P < 0.001$)	0.948 ($P < 0.05$)	0.098 (N. S.)	0.710 (N. S.)
	Length	0.943 ($P < 0.001$)	0.990 ($P < 0.01$)	0.215 (N. S.)	0.758 (N. S.)

^aThose ASO targets where protein or ribosomal binding occurs (e. g. A20, C20, D11, D20 oligomers listed in Table 1).

^bPerformed in rabbit reticulocyte cell-free translation system.

^cCell-free translation performed in wheat germ extracts.

^dOur structural analyses do not include cap nucleotides. Therefore the binding sites for β 5 and β 7 (Table 3) are identical, and the reported inhibition for this target was considered 71% (average of β 5 and β 7).

CARLO program and reported inhibition. This was to be expected since the DHFR mRNA in these experiments was first heated to 65°C to eliminate secondary structure, and therefore suggests the correlations observed for the Cscore merely reflect the contribution of the Dscore to the Cscore. Using the length as a predictive index of ASO efficacy produces the best correlation with experimental results ($r^2 = 0.943$ ($P < 0.001$)). Weaker but significant correlations were observed between length and reported inhibition for experiments at 100 μM ($r^2 = 0.403$ ($P < 0.05$)) and 50 μM ($r^2 = 0.702$ ($P < 0.005$)) oligomer concentration. Dscores, Cscores, and Sscores did not correlate with inhibition at these lower concentrations (data not shown).

Rabbit reticulocyte lysates have variable, but generally very low levels of RNase H activity (35, 36). This suggests that ASOs used in this system act primarily through a RNase H independent (i.e. physical occlusion) mechanism. One might expect that ASOs targeted to regions of the mRNA at which the ribosomal subunits bind (e.g. within approximately the first fifteen nucleotides downstream of the m⁷G cap (37) or adjacent to or across the AUG translation initiation codon (38)) might therefore be more effective inhibitors of gene expression than ASOs targeted elsewhere along the mRNA.

A subgroup of ASOs in the Maher and Dolnick study can be classified as 'ribosome competitors' (e.g. A20, C14, C20, D11, D20). For the ribosome competing ASOs, length correlates best with reported inhibition ($r^2 = 0.990$ ($P < 0.01$)). The correlations for the Dscores ($r^2 = 0.962$ ($P < 0.02$)) and Cscores ($r^2 = 0.948$ ($P < 0.05$)) are also excellent.

We also examined the results of ASO-mediated inhibition of translation of rabbit β -globin mRNA in rabbit reticulocyte lysates (14). These studies by Goodchild and coworkers used eight different phosphodiester ASOs targeted to the globin mRNA. When examining ASOs $\beta 1$ – $\beta 8$ at 5.2 μM concentration, we found no significant correlations between observed inhibition and any predictive indices (Table 2).

The results of the study by Goodchild and colleagues clearly suggest that in rabbit reticulocyte lysates, ASOs targeted to the cap site or the start codon work better than those targeted to other regions of the mRNA (Table 1). However, when analyzing ASOs $\beta 2$, $\beta 5$, $\beta 7$ and $\beta 8$, no significant correlations were obtained between inhibition and Dscore, Cscore, Score or length (data not shown).

Wheat Germ Extracts. Cazenave and colleagues (15) used an 11-mer and three 17-mer ASOs targeted to rabbit β -globin mRNA in a wheat germ cell-free translation system (Table 1). At 6 μM oligomer concentration, Dscores correlate very well with reported inhibition ($r^2 = 0.933$ ($P < 0.05$); Table 2). Wheat germ extracts contain significant amounts of RNase H activity, and RNase H mediated processes are likely to be the major mode by which ASOs produce inhibition in this cell-free translation system (35). Any region of the mRNA accessible to the ASO should allow duplex formation, and thereby permit RNase H to degrade the mRNA. More energetically favored duplexes should render the mRNA more sensitive to RNase H, and therefore result in greater inhibitions. The strong correlation between inhibition and Dscores for all targets in the wheat germ system is consistent with RNase H mediated ASO inhibition. For the same ASO experiments, the Cscores, Sscores, and ASO length do not correlate with reported inhibition (Table 2).

For experiments with oligomer concentrations below 6 μM , no significant correlations were obtained between reported inhibitions of β -globin translation and any of the predictive indices (data not shown).

Analysis of ASO action in Intact Cells

Predictions for human *c-myc* inhibition in HL-60 cells. Bacon and Wickstrom (16) selected multiple ASO targets located between the 5' cap site and the tenth codon in the coding sequence of human *c-myc* mRNA. They tested phosphodiester ASOs

Table 3. Antisense phosphodiester against human *c-myc* in HL-60 cells⁽¹⁶⁾

Name	Length	Sequence	Target Site ^a	Dscore	Cscore ^a	Inhibition (%) ^b
E ^c	15	CTCGCATTATAAAGG	orf(?) ^{d,e,f}	N D	N D	–25 ^g
F	15	GCACAGCTCGGGGGT	cap ^e	–15.3	–11.2	38
G	15	CCGGCGGTGGCGGCC	5' UTR ^{e,h}	–17.1	–11.7	35
H	15	GCCCCGAAAACCGGC	5' UTR ^e	–14.5	–12.5	16
I	15	CGCCCGGCTCTTCCA	5' UTR	–14.9	–8.9	8
J	15	TGGGCCAGAGGCGAA	5' UTR	–14.8	–13.6	21
K	15	GTGTTGTAAGTTCCA	5' UTR	–10.3	–9.3	11
L	15	GAGGCTGCTGGTTT	splice	–12.5	–7.2	23
M	15	GGGGCATCGTCGCGG	start	–15.2	–13.6	15
N	12	GTTGAGGGGCAT	start	–	10.1	–10.0 6
O	15	AACGTTGAGGGGCAT	start	–	12.6	–12.4 17
Q	15	CTCAAGTCGCATGAG	control	N D	N D	(52) ^k
R	18	GCTAACGTTGAGGGGCAT	start	–	–15.7	–8.6 28
S	15	CCTGTTGGTGAAGCT	coding	–14.3	–12.3	0

^aSpliced P1 transcript.

^bSpecific inhibition calculated by subtracting out inhibition due to oligomer Q.

^cNot included in correlation data.

^dTarget site does not appear in the P1 transcripts.

^eTarget site does not appear in the P2 transcripts.

^fUntranslated region.

^gOligomer was stimulatory.

^hWith cap-sensitive region (38).

^jControl oligomer with sequence/composition derived from scrambling oligomer O sequence.

^kFactor referred to in note b.

against these targets when attempting to suppress the *c-myc* gene in cultured HL-60 cells. Since a genomic *c-myc* sequence was available and because ASOs presumably may bind to either hnRNA or mRNA within the cell, we folded both unspliced and spliced *c-myc* mRNA sequences. All reported inhibition values were adjusted prior to calculating correlation coefficients to account for non-sequence specific inhibition (e. g. oligomer Q; Table 3). Since oligomer E produced a stimulation of *c-myc* p65 expression (Table 3), it was not included in any of the best fit correlation calculations.

When considering all the possible ASO targets on the spliced (sP1) or unspliced (usP1) transcripts, only the Dscores have statistically significant correlations to reported inhibition (Table 5). The coefficient of determination for the targets on the usP1 transcript ($r^2 = 0.497$ ($P < 0.02$)) is slightly better than that for the ASO targets on the sP1 transcript ($r^2 = 0.437$ ($P < 0.05$)). However, this improvement simply reflects elimination of oligomer L from the usP1 correlation data set. We excluded oligomer L because it is capable of partial binding to two distinct sites on the usP1 transcript, and we were uncertain how partial binding would affect computation of Dscores. Indices incorporating secondary structure considerations (Cscores, Sscores) do not correlate to reported inhibition when examining the ASO targets on either transcript.

When examining the ribosome competitor set of ASOs (e. g. oligomers F, G, M, N, O, and R; Table 5), Dscore correlations to the reported specific inhibition improve ($r^2 = 0.665$ ($P < 0.05$), Table 5). The correlations between Sscore and reported inhibition are also significant for both transcripts (sP1, $r^2 = 0.735$ ($P < 0.05$); usP1, $r^2 = 0.878$ ($P < 0.01$)). However, the Sscore correlations are negative (increasingly strong secondary structures correlated to increasingly efficacious ASOs). The anomalous behavior of the Sscores compensates for the correlations obtained with the Dscores, and explains why the Cscores are such poor indicators of ASO efficacy for this subset of ASOs (sP1, $r^2 = 0.022$; usP1, $r^2 = 0.292$).

The *c-myc* gene also contains two other promoters, termed P0 (39) and P2 (40). The results obtained for both the spliced and unspliced transcripts originating from the P0 promoter were

similar to those obtained for the corresponding transcripts from the P1 promoter (data not shown). In the case of the spliced and unspliced transcripts originating from the P2 promoter, there was no statistically significant correlation obtained between any of the three indices and reported inhibition. It should be noted that the targets for ASOs F, G, and H do not appear on the P2 transcripts (Table 3).

We also used Martinez's implementation of the Zuker optimal folding algorithm (41) to fold a 610 nucleotide fragment of the sP1 transcript containing all 12 *c-myc* ASO binding sites. However, no significant correlation between Sscores or Cscores and reported inhibition was observed (data not shown). Thus use of a different folding algorithm did not improve the predictive values of the indices incorporating secondary structure considerations.

Predictions for human ICAM-1 inhibition in HUVECs and A549 cells. Chiang and colleagues (17) examined the effects of multiple phosphorothioate ASOs upon cytokine induced expression of ICAM-1 in human umbilical vein endothelial cells (HUVEC) and in human lung carcinomal cells (A549). They selected targets located across the entire mRNA sequence (Table 4) and tested each ASO over a range of concentrations in both cell lines. For computing the correlations, we used the percent of inhibition occurring at the midpoint of each range of concentrations (i. e. $0.06 \mu\text{M}$ in HUVECs and $0.8 \mu\text{M}$ in A549 cells). These inhibition values are shown in Table 4. Each value was adjusted to eliminate a non-sequence specific inhibition of ICAM-1 reported for a control oligonucleotide (e. g. ISIS 1821; Table 4) prior to calculating correlation coefficients.

Chiang and colleagues selected their ASO targets using a published ICAM-1 cDNA sequence whose cap site had not been determined (32). Recently, the cap-site was mapped for human ICAM-1 transcripts isolated from cytokine-stimulated epithelial cells (33). This mapping study suggests that the target for ISIS 1930 lies upstream of the ICAM-1 transcriptional initiation site and therefore would not be found on the ICAM-1 mRNA sequence. The study also suggests that the target site for ISIS 1571 spans the cap site, and therefore would consist of 9

Table 4. Antisense phosphorothioates against human ICAM-1 mRNA in HUVEC and A549 cells⁽¹⁷⁾

Name	Length	Sequence	Target Site	Dscore	Cscore	Inhibition (%) ^a	
						HUVEC	A549
ISIS 1930	20	GGGCGCGTGATCCTTATAGC	5' UTR ^{b,c}	N D ^d	N D ^d	0	0
ISIS 1571	20 ^b	GAGGAGCTCAGCGTCGACTG ^e	5' UTR	N D ^d	N D ^d	0	0
ISIS 1931	19	CATAGCGAGGCTGAGGTTG	start AUG	-16.5	-14.7	45	50
ISIS 1570	18	TGGGAGCCATAGCGAGGC	start AUG	-17.9	-15.6	55	45
ISIS 1934	20	TGCCCCATCAGGGCAGTTTGA	coding	-18.4	-13.8	10	30
ISIS 1574	18	CGATGGGCAGTGGGAAAG	coding	-15.9	-10.7	30	0
ISIS 1938	20	CCTGTCCCGGATAGGTTCA	3' UTR	-18.5	-12.5	45	50
ISIS 1939	20	CCCCCACTTCCCCTCTC	3' UTR	-20.3	-10.9	75	70
ISIS 2302	20	GCCCAAGCTGGCATCCGTC	3' UTR	-19.8	-17.3	75	70
ISIS 1572	20	GACACTCAATAAATAGCTGGT	3' UTR	-14.4	-8.2	5	30
ISIS 1573	18	GAGGCTGAGGTGGGAGGA	3' UTR	-18.1	-15.0	42	5
ISIS 1940	20	TTGAGAAAGCTTTATTAAC	poly(A)	-12.4	-12.4	0	0
ISIS 1821	18	CATGGCGCGGGCCGCGGG	control ^f	N D	N D	(45) ^g	(20) ^g

^aFor HUVECs, specific inhibition at $0.06 \mu\text{M}$ oligomer; for A549 cells, inhibition at $0.8 \mu\text{M}$ oligomer.

^bAccording to sequence of Tomassini (32) (See Results).

^cUntranslated region.

^dNot determined (see Results).

^eLength of hybrid may only be the underlined nucleotides (see Results).

^fOligo targeted to 5-lipoxygenase mRNA.

^gConsidered non-specific inhibition, and subtracted from remaining oligomers.

nucleotides rather than 20 (Table 4). We were uncertain how this partial binding would affect the activity of ISIS 1571, and presumed ISIS 1930 would not be able to produce any effects through an antisense mechanism. Consequently, we did not include inhibition data for either ASO in our correlations.

For all 10 ASO targets in both cell lines, the Dscores correlated significantly with specific inhibition ($r^2 = 0.658$ ($P < 0.005$) for HUVECs; $r^2 = 0.489$ ($P < 0.02$) for A549s). No statistically significant correlations could be drawn between the Sscores or Cscores and observed inhibition in either HUVECs or A549 cells (Table 5).

To control for possible length dependent effects upon ASO uptake and therefore ASO inhibition, we also examined only the 20-mer ASOs used in the ICAM studies. The correlations between Dscore and observed inhibition were significant in both cell lines ($r^2 = 0.699$ ($P < 0.05$) for HUVECs; $r^2 = 0.820$ ($P < 0.02$) for A549s), while Cscores and Sscores for this subgroup did not correlate with reported inhibition in either cell line. This result suggests that the Dscores are best used to compare ASOs of identical length in the absence of information concerning intracellular concentrations of ASO.

We also used Martinez's implementation of the Zuker optimal folding algorithm (41) to fold nine, staggered and overlapping 600 nucleotide fragments of the ICAM-1 transcript. Again, use of a different folding algorithm did not produce improvements in the correlation between Sscores or Cscores and reported inhibition (data not shown).

Comparison of Dscore with other physical parameters of oligomers

Often, the melting temperature of a short DNA duplex is estimated by a simple 'two plus four' rule based upon the GC content of the oligomer. Simple GC compositional analysis (expressed as percent of ASO nucleotides that are G or C) did not correlate with reported inhibition for any of the cell free-systems, nor for the entire set or any subset of the *c-myc* ASOs (data not shown). For the ICAM-1 experiments, a significant correlation was observed between ASO GC content and specific inhibition in HUVECs ($r^2 = 0.694$ ($P < 0.005$)), while no

significant correlation was observed for the data from the A549 cells ($r^2 = 0.230$). Therefore a simple 'two plus four' rule for computing the T_m of the duplex is not a useful predictor of inhibition. This result illustrates the importance of ASO sequence rather than composition when predicting the efficacy of the ASO.

The ICAM-1 inhibition studies included information concerning the measured T_m values for duplexes between the phosphorothioate ASOs and complementary phosphodiester ASOs (data not shown). While the experimentally determined T_m 's correlate with observed inhibition as well as our Dscores for the HUVEC system ($r^2 = 0.692$ ($P < 0.02$)), there was no correlation between measured T_m 's and the specific inhibition in A549 cells ($r^2 = 0.171$).

McGraw and colleagues have published a program that computes a Mean Stacking Temperature (MST) for a DNA-DNA duplex (23). The MST is designed to account for sequence effects when estimating the thermal stability of a DNA duplex. No significant correlation was observed between MST and specific inhibition for the *c-myc* targets ($r^2 = 0.268$). In the ICAM-1 studies, a significant correlation between the MST and observed inhibition was obtained for the HUVEC data ($r^2 = 0.640$, ($P < 0.01$)), while the correlation between MST and inhibition for the A549 cells was not significant ($r^2 = 0.243$).

Why doesn't predicted mRNA structure correlate with ASO efficacy?

Certainly the binding of an ASO is important to inhibition, even when approximated by a Dscore. However, the lack of consistent significant correlations between inhibition and indices which incorporate considerations of mRNA secondary structure (Sscore or Cscore) suggests several interpretations. The structure prediction algorithms we used may not be reliable enough for this type of analysis, and therefore it is unreasonable to expect a high correlation with experimental results. If mRNA secondary structure is an important determinant of ASO efficacy, better prediction algorithms may improve the Sscore and Cscore values and hence their correlation with experimental results. It is also possible that mRNA secondary structure plays a minor role in determining ASO efficacy. This could reflect a general lack of

Table 5. Coefficients of determination (r^2) for predictive indices and reported inhibition in cultured cells

		sP1	<i>c-myc</i>	usP1	ICAM-1	A549 Cells
		(n = 12)		(n = 11) ^a	HUVECs	
duplex	Dscore	0.437 ($P < 0.05$)		0.497 ($P < 0.02$)	All Targets (n = 10) 0.658 ($P < 0.005$)	0.489 ($P < 0.02$)
	mRNA 2° structure	0.061 (N. S.) ^b		0.032 (N. S.)	0.075 (N. S.)	0.133 (N. S.)
	competition	0.155 (N. S.)		0.327 (N. S.)	0.237 (N. S.)	0.079 (N. S.)
		Ribosome Binding Competitors ^c (n = 6)			All 20-MER Targets ^d (n = 6)	
duplex	Dscore	0.665 ($P < 0.05$)		0.665 ($P < 0.05$)	0.699 ($P < 0.05$)	0.820 ($P < 0.02$)
	mRNA 2° structure	0.735 ($P < 0.05$)		0.878 ($P < 0.01$)	0.169 (N. S.)	0.346 (N. S.)
	competition	0.022 (N. S.)		0.292 (N. S.)	0.188 (N. S.)	0.101 (N. S.)

^aOligomer L not included correlation

^bNot significant

^cOligomers F, G, M, N, O, R from table 3

^dOligomers 1572, 1934, 1938, 1939, 1940, 2303 from table 4

stable mRNA secondary structure outside of defined functional elements. Alternatively, it may reflect a capacity of ASOs, when present at sufficient concentrations, to 'melt out' all but the strongest local secondary structures of the mRNA. Finally, the role of tertiary mRNA structure and mRNA binding proteins in ASO binding are not addressed by the folding algorithms or by our thermodynamic analysis.

It should also be noted that our thermodynamic indices assume ASO, mRNA and duplex concentrations are 1M. The indices make no attempt to account for the concentrations of the ASO or mRNA involved in the ASO:mRNA reaction or the mRNA folding reaction. A more rigorous thermodynamic approach would include terms for the concentration of these reactants and products. When making predictions concerning ASO efficacy within intact cells, such a rigorous approach is limited by the lack of knowledge of intracellular concentrations of mRNA and ASO. Our approach is an attempt to predict which of the many possible ASOs from a given sequence has the greatest probability of producing inhibition. Therefore, the ASO concentration and effective concentration of target mRNA are assumed to be constant when comparing different ASOs. The failure of the Cscores to consistently and accurately predict ASO efficacy may be in part due to this omission of concentration terms.

Conclusions

Our results suggest Dscores are useful indicators of ASO efficacy. The Dscores displayed as histograms along the target mRNA sequence permit rapid determination of the ASOs most likely to produce strong inhibition. This strategy of using Dscores to select optimal ASO sequences may work best when comparing ASO targets that overlap a specific mRNA region (i. e. determining the best 15-mer spanning the start codon) An added advantage of the Dscore is that it can be calculated by a quick and simple sequence analysis algorithm, and therefore, is not computationally expensive.

ACKNOWLEDGEMENTS

This work has been supported by NIH GM 30163 (RAS, F.Szoka, principal investigator) and in part, utilized the facilities of the UCSF Computer Graphics Lab (RR-1081; R.Langridge, principal investigator). We would like to thank Dr. C.F.Bennett of ISIS Pharmaceuticals for kindly providing the results of the ICAM inhibition studies prior to publication, Dr. H.Martinez for helpful discussions, Dr. P.Szoka, T.Wolfsberg, and D.Knoell for comments concerning the manuscript, and Dr. R.Siegel for advice concerning the statistical analysis. R.A.S. would also like to thank Dr. L.Ceci for useful comments concerning the EWINDOW program.

REFERENCES

- Cope, F., and Wille, J. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 5590–5594.
- Heikkila, R., Schwab, G., Wickstrom, E., Loke, S.L., Pluznik, D.V., Watt, R., and Neckers, L.M. (1987) *Nature* **328**, 445–449.
- Teichman-Weinberg, A., Littauer, U.Z. and Ginzberg, I. (1988) *Gene* **72**, 297–307.
- Zon, G. (1988) *Pharmaceut. Res.* **5**, 539–549.
- Hélène, C. and Toulmé, J.J. (1990) *Biochim. Biophys. Acta* **1049**, 99–125.
- Cohen, J.S. (ed.) (1989). *Oligonucleotides: Antisense Inhibitors of Gene Expression* Macmillan/CRC Press, London/Miami.
- Stephenson, M.L. and Zamecnik, P.C. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 285–288.
- Zamecnik, P.C. and Stephenson, M.L. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 280–284.
- Matsukura, M., Zon, G., Shinozuka, K., Robert-Guroff, M., Shimada, T., Stein, C.A., Mitsuya, H., Wong-Staal, F., Cohen, J.S., and Broder, S. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 4244–4248.
- Smith, C.C., Aurelian, L., Reddy, M.P., Miller, P.S. and Ts'o, P.O.P. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 2787–2791.
- Zaia, J.A., Rossi, J.J., Murakawa, G.J., Spallone, P.A., Stephens, D.A., Kaplan, B.E., Eritja, R., Wallace, R.B., and Cantin, E. (1988) *J. Virol.* **62**, 3914–3917 (1988).
- Zuker, M. (1989) *Methods Enzymol* **180**, 262–288.
- Maher, L.J. and Dolnick, B.J. (1987) *Arch. Biochem. Biophys.* **253**, 214–220.
- Goodchild, J., Carroll, E. and Greenberg, J.R. (1988) *Arch. Biochem. Biophys.* **263**, 401–409.
- Cazenave, C., Loreau, N., Thuong, N.T., Toulmé, J.J. and Hélène, C. (1987) *Nucleic Acids Res.* **15**, 4717–4736.
- Bacon, T.A. and Wickstrom, E. (1991) *Oncogene Res.* **6**, 13–19.
- Chiang, M.-Y., Chan, H., Zounes, M.A., Freier, S.M., Lima, W.F. and Bennett, C.F. (1991) *J. Biol. Chem.* **266**, 18162–18171.
- Martinez, H. (1988) *Nucleic Acids Res.* **16**, 1789–1798.
- Martinez, H. (1984). *Nucleic Acids Res.* **12**, 323–334.
- Freier, S.M., Kierzek, R., Jaeger, J.A., Sugimoto, N., Caruthers, M.H., Neilson, T., and Turner, D.H. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 9373–9388.
- Jaeger, J.A., Turner, D.H. and Zuker, M. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 7706–7710.
- Gotoh, O. and Tagashira, Y. (1981) *Biopolymers* **20**, 1033–1042.
- McGraw, R.A., Steffe, E.K. and Baxter, S.M. (1990) *BioTechniques* **8**, 674–678.
- Bilofsky, H.S. and Burks, C. (1988) *Nucleic Acids Res.* **16**, 1861–1864.
- Yang, J.K., Masters, J.N. and Attardi, G. (1984) *J. Mol. Biol.* **176**, 169–187.
- Chen, M.-J., Shimada, T., Moulton, A.D., Cline, A., Humphries, R.K., Maizel, J., and Nienhuis, A.W. (1984) *J. Biol. Chem.* **259**, 3933–3943.
- Baralle, F.E. (1977) *Cell* **10**, 549–558.
- Efstratiadis, A., Kafatos, F.C. and Maniatis, T. (1977) *Cell* **10**, 571–585.
- Proudfoot, N.J. (1977) *Cell* **10**, 559–570.
- Colby, W.W., Chen, E.Y., Smith, D.H. and Levinson, A. (1983) *Nature* **301**, 722–725.
- Gazin, C., Dupont de Dinechin, S., Hampe, A., Masson J.-M., Marint, P., Stehelin, D., and Galibert, F. (1984) *EMBO J* **3**, 383–387. H., Lenoir, G., and Leder, P. (1983) *Cell* **34**, 779–787.
- Tomassini, J.E., et al. *Proc. Natl. Acad. Sci. USA* **86**, 4907–4911 (1989).
- Degitz, K., Lian-Jie, L. and Caughman, S.W. (1991) *J. Biol. Chem.* **266**, 14024–14030.
- C. F. Bennett, personal communication
- Minshull, J. and Hunt, T. (1986) *Nucleic Acids Res.* **14**, 6433–6451.
- Walder, R.Y. and Walder, J.A. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 5011–5015.
- Lawson, T.G., Ray, B.K., Dodds, J.T., Grifo, J.A., Abramson, R.D., Merrick, W.C., Betch, D.F., Weith, H.L., and Thach, R.E. (1986). *J. Biol. Chem.* **261**, 13979–13989.
- Shakin, S.H. and Liebhafner, S.A. (1986) *J. Biol. Chem.* **261**, 16018–16025.
- Bentley, D.L. and Groudine, M. (1986) *Mol. Cell. Biol.* **6**, 3481–3489.
- Batley, J., Moulding, C., Taub, R., Murphy, W., Stewart, T., Potter, H., Lenoir, G. and Leder, P. (1983) *Cell* **34**, 779–787.
- Zuker, M. and Steigler, P. (1981) *Nucleic Acids Res.* **9**, 133–148.