BIOINFORMATICS

Vol. 18 no. 12 2002
Pages 1567–1575

Computational antisense oligo prediction with a neural network model

Alistair M. Chalk and Erik L.L. Sonnhammer*

Center for Genomics and Bioinformatics, Karolinska Institutet, S-17177 Stockholm, Sweden

Received on September 4, 2001; revised on April 5, 2002; May 27, 2002; accepted on June 6, 2002

ABSTRACT

Motivation: The expression of a gene can be selectively inhibited by antisense oligonucleotides (AOs) targeting the mRNA. However, if the target site in the mRNA is picked randomly, typically 20% or less of the AOs are effective inhibitors *in vivo*. The sequence properties that make an AO effective are not well understood, thus many AOs need to be tested to find good inhibitors, which is time consuming and costly. So far computational models have been based exclusively on RNA structure prediction or motif searches while ignoring information from other aspects of AO design into the model.

Results: We present a computational model for AO prediction based on a neural network approach using a broad range of input parameters. Collecting sequence and efficacy data from AO scanning experiments in the literature generated a database of 490 AO molecules. Using a set of derived parameters based on AO sequence properties we trained a neural network model. The best model, an ensemble of 10 networks, gave an overall correlation coefficient of 0.30 ($p = 10^{-8}$). This model can predict effective AOs (>50% inhibition of gene expression) with a success rate of 92%. Using these thresholds the model predicts on average 12 effective AOs per 1000 base pairs, making it a stringent yet practical method for AO prediction.

Availability: A prediction server is available at http://www.cgb.ki.se/AOpredict

Contact: alistair.chalk@cgb.ki.se

INTRODUCTION

Antisense oligonucleotides (AOs) contain 10–30 nucleotides complementary to a specific sub-sequence of an mRNA target, which are designed to bind to targets by standard Watson–Crick base pairing rules. The bound duplex is a substrate for RNase-H which cleaves the RNA while leaving the AO intact. The AO inhibits gene expression in a specific and reversible manner, a process termed 'Gene knock-down'. For a comprehensive review of the topic see (Crooke, 2000).

There are many laboratory-based strategies for selecting AOs. One of the most successful is the 'gene-walk' approach, in which 15 or more AOs are evaluated for a gene in order to find a sufficiently effective AO. A faster method selects mRNA regions that are accessible to RNase-H cleavage and therefore more likely to be an effective site for AOs (Ho *et al.*, 1998). Milner *et al.* (1997) use an array of 1–17mer oligodeoxynucleotides to measure heteroduplex formation potential. In general the experimental approaches are time consuming and expensive.

AO design is not as simple as synthesizing an AO that is complementary to the mRNA. Factors such as likelihood to form hairpin or dimer structures and the accessibility of the region of mRNA targeted are important factors in AO efficacy. AOs are generally selected to avoid these problems, however even for careful designs the success rate is low, typically 20% or less of AOs are effective (Myers and Dean, 2000).

There are many examples in the literature of experimental groups attempting to correlate AO sequence properties with efficacy. A correlation between binding energy (AO-RNA) and efficacy has been observed (Ho *et al.*, 1996, 1998). Particular target secondary structures have been shown to correlate with efficacy (Vickers *et al.*, 2000; Bacon and Wickstrom, 1991; Laptev *et al.*, 1994). However the correlations are not consistently detected across studies. This variation can be due to many factors including biases in the selection of the AOs, varying experimental conditions, or, in cases where computational RNA folding prediction was used, limitations in the structure prediction methods.

The properties thought most likely to affect AO efficacy can be divided into classes: (a) binding energy parameters consisting of RNA–AO binding energy, measures of accessibility of the RNA to the AO and AO–AO binding energies (dimerization and hairpin energy); (b) sequence motifs (these could affect the AO or RNA structure, RNase-H recognition, half-life); and (c) 'Other'—cellular uptake, protein interactions, etc.

*To whom correspondence should be addressed.

AO selection can be based on either experimental or theoretical approaches (for a review, see Sczakiel, 2000). Computational approaches to AO design have so far focused on prediction of the structure of the target mRNA and from this deriving the accessibility of target regions (e.g. Sczakiel *et al.*, 1993; Mathews *et al.*, 1999; Patzel *et al.*, 1999; Walton *et al.*, 1999; Amarzguioui *et al.*, 2000; Song *et al.*, 2000; Scherr *et al.*, 2000; Sczakiel, 2000; Toschi, 2000; Ding and Lawrence, 2001). Perhaps the most successful method is that of Ding and Lawrence (2001), using a statistical sampling of secondary structures to predict accessible regions to find effective AOs for rabbit β-globin. In general, methods have not been evaluated on a broad range of gene targets. Another method is to look for motifs that occur more often in effective AOs. In a survey of 42 AOs with high efficacy from selected studies, 20 contain the TCCC motif. 10 sequence motifs have been identified with a correlation to AO efficacy by Matveeva *et al.* (2000). While no proof of mechanism has been reported, possible explanations include cellular uptake, preferential binding to RNase-H, stability of the DNA:RNA duplex and effect of motifs on AO/target structure.

The challenge is hence to discover general principles that hold across all AO studies. One approach to discover such principles is to explore a diverse range of sequence properties and incorporate the factors that affect AO efficacy into a computational model for AO design. This requires both a database of tested AOs, such as that produced by Giddings *et al.* (2000), and machine learning methods of model building. The database should be based on large AO screening experiments to ensure comparability. The model must be able to handle the inherent noise in the system, whilst being able to find the complex relationship between the various AO properties and the AO's efficacy.

A popular machine learning approach that effectively handles noise and complex relationships in a robust way is the artificial neural network (NN). For an introduction to neural networks see Baldi and Brunak (1998) and Rumelhart *et al.* (1986). In this paper we present a data compilation of AOs with experimentally determined efficacy, and apply a neural network model to predict AO efficacy.

METHOD

Database

A database was assembled from a selection of AO publications. Published data was incorporated for which: (a) at least 6 AOs were tested under the same experimental conditions, although more than one gene target were allowed; (b) efficacy of the AOs were presented as a percentage of the control level of the target gene expression, either as RNA or protein. No papers were reported matching these criteria before 1990, as is consistent with Giddings *et al.* (2000). Accompanying this data is the full RNA sequence and accession number (where available) together with positional coordinates of the AOs and the position of the coding sequence. Publication details, cell line used and the chemistry of the AOs (unmodified, phosphorothioate, phosphoramidite) were also recorded in the database. The database consists of 490 oligonucleotides from 20 studies testing AO efficacy on 16 genes. The essential information in the database is AO sequence and efficacy expressed as (100%–[% of control expression])/100; efficacies range between 0 and 1, representing no inhibition and complete inhibition of gene expression respectively. A summary of the contents is shown in Table 1. For the cases where the same AO is tested in two different laboratories, or twice by the same laboratory the average efficacy is used.

Derived parameters as network input

A set of parameters was derived from the information contained in the AO sequence collection, including values for: (1) base composition (Number of A/C/G/T, %GC content): (2) RNA–AO binding properties (binding energy, enthalpy, entropy): (3) RNA–AO terminal properties (3' binding energy, 5' binding energy); (4) AO–AO binding properties (Hairpin energy and quality, Dimer energy); and (5) 9 of the 10 verified sequence motifs correlated with efficacy from Matveeva *et al.* (2000). Binding energy calculations were completed using thermodynamic parameters from Sugimoto *et al.* (1995). The calculation of dimer energy was made using an ungapped alignment with stacking energies taken from Sugimoto *et al.* (1996) and a uniform penalty 0.5 for mismatches. Hairpin energy was calculated using mfold (Walter *et al.*, 1994). Hairpin quality (the number of hydrogen bonds in the stem of the stem–loop structure) is calculated using the Stemloop module of the GCG Package (Devereux *et al.*, 1984). Parameters describing cellular uptake and protein interactions were not included, as we have no explicit way of modeling them.

Efficacy values were first rounded such that all efficacy values are between [0,1], negative efficacy levels occur when the target gene is up-regulated with respect to the level of control. Any biological data set contains noise, however careful the collection method. We strive to reduce this noise by removing 93 data points with efficacy less than 0.05. The reasoning is that there are many factors that can cause an experiment or part of the experiment to fail (giving efficacy <0.05), while there are few situations that can cause high efficacy values to be so noisy. This 'noise' is illustrated by observing the correlation between dH/length and efficacy, which is -0.11 using the full data set and improves to -0.19 when removing AOs in the range 0–0.05. It is also interesting to note that for binding

Table 1. The antisense target genes in AOdb 1.0. Each antisense oligo is annotated with the antisense efficacy and its position in the target gene. The EMBL accession number is given where known

Description / Swiss-prot/ TREMBL id	EMBL accession	No AOs	Reference
AG2R_RAT	X62295	33	Ho *et al.* (1998)
β-globin/adenovirus introns	–	11	Hodges and Crooke (1995)
CA11_HUMAN	Z74615	21	Laptev *et al.* (1994)
DUS1_MOUSE	X61940	11	Duff *et al.* (1995)
Hepatitis C virus sequences	–	20	Hanecak *et al.* (1996)
I11R_HUMAN	M27492	36	Miraglia *et al.* (1996)
ICA1_HUMAN	J03132	4	Bennett *et al.* (1994)
ICA1_HUMAN	J03132	10	Chiang *et al.* (1991)
ICA1_HUMAN	J03132	40	Lee *et al.* (1995)
ICA1_HUMAN	M31585	10	Stepkowski *et al.* (1994)
KPCA_HUMAN	X52479	20	Dean *et al.* (1994)
KRAF_HUMAN	X03484	34	Monia *et al.* (1996)
LEM2_HUMAN	M30640	18	Bennett *et al.* (1994)
LEM2_HUMAN	M30640	17	Lee *et al.* (1995)
MDR1_HUMAN	M14758	22	Ho *et al.* (1996)
MRP1_HUMAN	L05628	15	Stewart *et al.* (1996)
Q13896	M20789	6	Colige *et al.* (1993)
MYC_HUMAN	V00568	13	Bacon and Wickstrom (1991)
TNFA_HUMAN	M10988	29	Mayne *et al.* (1999)
TNFA_RAT	D00475	42	Tu *et al.* (1998)
TNFB_HUMAN	M16441	14	Lefebvre d'Hellencourt (1996)
TPA_MOUSE	J03520	21	Stutz *et al.* (1997)
VCA1_HUMAN	M60335	15	Bennett *et al.* (1994)
VCA1_HUMAN	M60335	28	Lee *et al.* (1995)

energy (delta G) the correlation coefficient is zero for the set of AOs with efficacy 0–0.5, while a correlation coefficient of -0.17 is observed for the AOs with efficacy in 0.5–1.0. This observation held for a large number of the 23 AO properties explored, with correlations increasing with removal of low efficacy AOs.

Several strategies were employed to determine the desirable input parameters for the network. First all potential parameters were derived. Next two strategies were used to create input parameter sets: (A) removal of highly redundant parameters using the following method. If the correlation between two parameters was above a defined cutoff the parameter with the lowest correlation coefficient with efficacy was removed. This procedure is repeated until no two parameters are correlated above the defined cutoff. Two sets (denoted b and c) were generated using 0.8 and 0.65 as cutoff values; and (B) correlated data set. Only parameters with a high correlation with efficacy were used. Two sets (denoted d and e) were generated using 0.1 and 0.15 as cutoffs for the correlation coefficients (data not shown.). The parameter combinations are presented in Table 2.

Neural Network

A schematic view of the neural network (NN) model for predicting AOs is shown in Figure 1. The neural network source code is based on C code written by Dr Mikael Boden, with additions made by the authors for the simulation requirements. For the current application, inline training was employed using a standard sigmoid activation function with backpropagation. For all simulations the learning rate was set to 0.2, with momentum $= 0.5$. The presentation order of the input vectors remained constant throughout training. The continuous input nodes of the network were normalized so that 95% of values for each node were in the range $[0, 1]$. Motif presence/absence was encoded in a binary fashion, 0 denotes no motif, while 1 denotes one or more occurrences of the motif.

Early stopping and cross validation

For our NN simulations we define three sets of AOs: a training set, an early stopping set and a test set. The early stopping set is used in order to decide when to stop the training, and should not contain AOs similar to those in the test set. After this point is reached the test set (as yet unseen by the network) is used to test the accuracy of the network. An example is shown in Figure 2. We tested several different early stopping set sizes (20, 40, 60, 80—data not shown), and selected 40 as the early stopping set size.

When considering a multitude of cross-validation methods we asked ourselves the question: 'How related is the

Table 2. The reduced parameter input sets were based on selections from set a, which includes all input parameters

	Input set					
		Non-redundant		Efficacy-correlation		
Parameter	a	b	c	d	e	Correlation
Contains AGAG	x	x	x	x	x	0.18
Contains CAGT	x	x	x	x	x	−0.18
Contains CCCC	x	x	x	x		−0.13
Contains GAGT	x	x	x			0.09
Contains GGGA	x	x	x	x	x	0.26
Contains GTGG	x	x	x	x	x	0.17
Contains TGGC	x	x	x	x		0.05
Contains TTA	x	x	x			−0.02
Contains TTT	x	x	x			−0.10
Delta G/Oligo length	x	x		x	x	−0.18
Delta H/Oligo length	x	x	x	x	x	−0.19
Delta S/Oligo length	x			x		−0.14
Dimer energy	x	x	x			0.00
GC content	x			x		0.11
Hairpin energy	x	x				0.09
Hairpin length	x	x				−0.09
Hairpin quality	x	x	x	x		−0.11
Number of A's in AO	x	x				−0.04
Number of C's in AO	x	x	x	x	x	0.23
Number of G's in AO	x	x	x	x	x	−0.15
Number of T's in AO	x	x	x			−0.06
3′ dG	x	x	x	x		0.14
5′ dG	x	x	x			0.03

Two non-redundant sets were created as follows. Set b: potentially redundant parameters correlated to each other at > 0.8 were removed (the parameter with the highest correlation to efficacy was kept.); and set c: as (b) using 0.65 as a cut-off. Two sets based on correlation between efficacy and individual parameters were created. Set d: all parameters having individual correlation coefficients to efficacy $\geqslant 0.1$; and set e: as (d) having a correlation $\geqslant 0.15$.

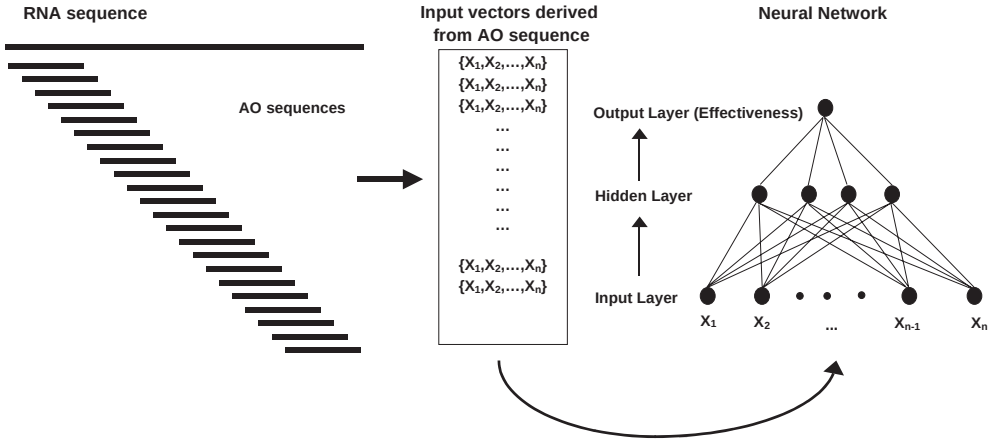

Fig. 1. Schematic diagram of the method for training and prediction. All possible AOs are generated from the target gene. A vector of derived parameters is then calculated for each AO. Each vector is used as input to the neural network, resulting in a predicted efficacy. *n* is defined as the maximum number of input parameters for the network.

efficacy of AOs on a single gene?' We collated all possible pairs of AOs on the same gene, and computed the correlation between all pairs: pairs that were overlapping, between 1 and 50 bp apart, and those greater than 50 bp apart. All pairs in a distance category were combined from different genes by normalizing the ranks of the efficacies

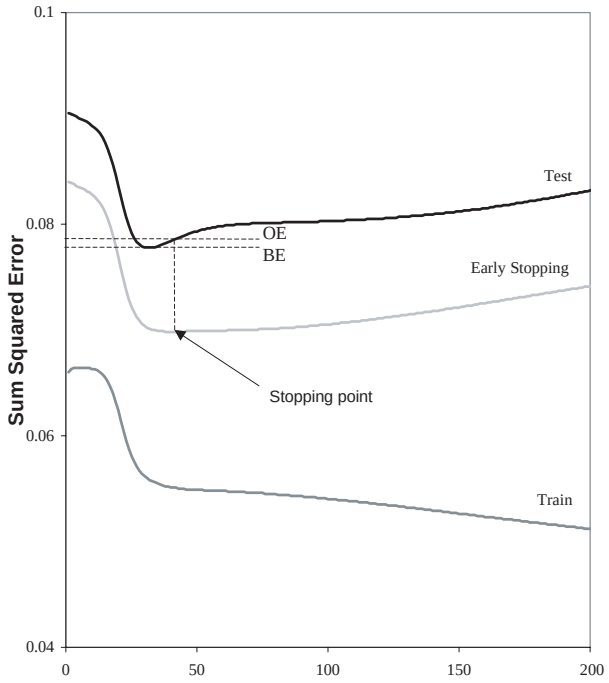

Fig. 2. Error profile of a network training run. An example of NN training shows that optimal performance for training, early stopping, and test sets may not occur at the same epoch. The learning rate was 0.2 with a momentum of 0.5. The architecture of the network was 14-10-1. BE = Best observed error on the test set. OE = Observed error on the test set when the early stopping set reaches its best performance.

so that the scale was independent of the number of AOs. The results for the normalized rank correlation in each category are shown in Table 3. No correlation was found between AOs greater than 50 bp apart, hence we selected this cut-off for further testing. This approach allows us to use more of the data points in the training of the model than the standard method of cross-validation, where all AOs from the same gene are disallowed from the training and early stopping sets. Our method is illustrated in Figure 3.

We trained each network until the error on the early stopping set reached minimum. We then use this network to predict the test example's efficacy. Network performance is measured as the Spearman rank correlation for the set of all pairs of test examples (observed, predicted). In all cases we averaged the results from 10 networks to improve the robustness of the results.

RESULTS

We compiled a database of 490 AO molecules from 20 AO screening studies on 16 different genes, see Table 1. The database, called AOdb, contains all the relevant data on the AO sequences and their efficacy, as well as the

Table 3. Rank correlation between AOs in AOdb. (a) Where several studies study the same gene the gene is counted as a single entity (default behavior). (b) In cases where several groups study the same gene, each study is calculated separately

All pairs of AOs	Examples	Correlation	*P*-value
(a)			
Overlapping	501	0.132	0.001
Within 50 bp	723	0.079	0.017
Greater than 50 bp apart	4685	0.025	0.018
(b)			
Overlapping	405	0.164	0.001
Within 50 bp	592	0.114	0.003
Greater than 50 bp apart	3614	0.039	0.003

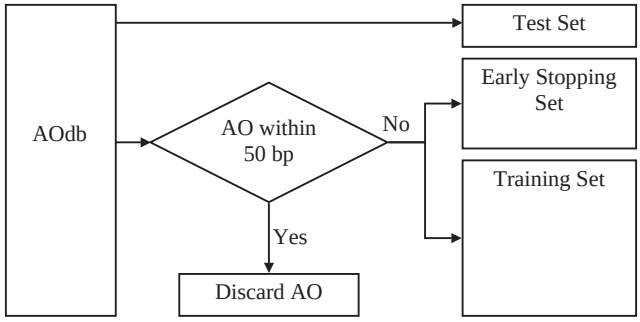

Fig. 3. Division of database into training, early stopping and test sets. Creation of the sets is done in 3 steps. (a) One sequence is taken out from the database as the test set. (b) The remaining AOs are checked to see if they are within 50 bp of the test AO. (c) A portion of the remaining AOs is randomly selected for the early stopping set while all remaining AOs are used as the training set.

corresponding cDNA sequences. AOdb can be used as a resource for testing models for AO prediction, and is available on request.

We used those AOs in AOdb with position information to train NN models for predicting AO efficacy. We explored a multitude of architectures, using previously described input parameter sets, and varied the number of hidden units. The best architecture was observed using input set (d), with 11 hidden units. The resulting Spearman rank correlation (observed vs predicted) was 0.30 ($p = 10^{-8}$).

How much confidence should one have in a given prediction? The distribution of observed efficacies compared to predicted efficacies (Figure 4) shows that AOs predicted to have high efficacy are more reliable than low efficacy predictions. We can quantitate the reliability by looking at predictions above a defined threshold X and observe how many are correctly predicted to have an observed efficacy above a threshold Y. Using the 10-ensemble net-

Table 4. Prediction results. Accuracy of predictions using several levels of stringency and accuracy. Results for using a distance of 50 as a cut-off, input set d, 11 hidden units

Prediction cutoff	Observed cutoff 0.5	0.8
0.5	40/46 (87%)	19/46 (41%)
0.6	24/26 (92%)	13/26 (50%)
0.7	11/12 (92%)	6/12 (50%)

works gives the following result: Of all AOs predicted above 0.6, 92% were observed to have an efficacy above 0.5. Of these, 50% were highly effective (having an observed efficacy above 0.8; see Table 4). This is considerably higher than randomly picked AOs from AOdb, which gives 52% above 0.5 efficacy and 17% above 0.8.

For comparison we tested a more standard method of cross-validation, involving removing all AOs targeting the same gene as the test AO from the training and early stopping sets. The resulting performance was better when looking at the rank correlation statistic (0.34) while AOs predicted above 0.8 efficacy showed similar results, while results for predicting AOs with efficacy above 0.5 were diminished (data not shown).

How many effective AOs can one expect to predict in a stretch of DNA? We applied the model to a set of 17 randomly chosen mRNA sequences, with a total length of 55 750 bp. No sequences showed homology to any genes in AOdb using BLAST (Altschul *et al.*, 1997) with an e-value cutoff of 10.0. The 327 ensembles generated during training processed the sequences. AOs of high efficacy were predicted at the rate of 12 and 51 AOs per 1000 bp, using a cutoff of 0.6 and 0.5 respectively. This suggests that the method is useful for AO selection in practice. Obviously, because the model is tuned to be stringent, it does not find all effective AOs. In fact, the false negative rate is high—on AOdb it is over 80%. As AOdb only contains a sparse sampling of AOs from each gene, the true false negative rate is unknown, but it is clearly very high. This is however acceptable if the goal is to find a practical number of effective AOs in a given gene while avoiding false positives.

DISCUSSION

While the proposed mechanisms by which AOs inhibit gene expression are relatively simple, the efficacy of a given AO is determined by many complex factors of which we have limited understanding. Antisense studies have produced small correlations between one parameter (e.g. RNA structure, motif content, binding energy) and efficacy. However these correlations are sometimes conflicting and vary between studies. Our method of model generation relies on a broad data set from many studies and target genes to generate a robust system for predicting effective AOs for any given gene.

We removed AOs within 50 bp when training, based on observation that such pairs have some correlation between their efficacies. Yet there are many examples of overlapping AOs that have strikingly different efficacies. An answer to this question and greater understanding of the AO prediction problem requires both a data set containing many series of overlapping AOs and methods for extracting explicitly structural information from the target region. RNA structural information is implicit in our model, using motif content and energy parameters. Subsequent studies of RNA folding combined with more detailed data sets of overlapping AOs have the potential to greatly increase our understanding of AO design.

The input parameters used in the model contains many of the motifs found by Matveeva and colleagues (Matveeva *et al.*, 2000), which is expected due to the similarity of the data sets used. Perhaps surprising is the lack of correlation between dimer energy and efficacy. An explanation for this could be a bias in AOdb—one would expect the majority of AOs in AOdb to be designed to avoid high dimer energies. Hairpin energy had the lowest correlation of any parameter used in the model, again this lower correlation could relate to a pre-screening process used by experimentalists when designing AOs. The highest correlations found were generally energy (directly as binding energy, indirectly as %GC content) or motif based. Energy terms can be expected to be good indicators as binding is a critical part of this antisense process, and indeed a number of studies in the area focus on the thermodynamics of the process. The reasons for motif content being a good predictor are unknown, but several hypotheses exist (see **Introduction**).

Because of the complexity and multitude of factors affecting AO efficacy, we were unable to reach a higher correlation coefficient between predicted and observed efficacy than 0.30 for all data points. However, most of this variability is found for AOs with predicted efficacies in the range 0–0.5, while AOs in the upper range show less variation (see Figure 3). Therefore the suitable application of this model is the prediction of highly effective AOs. The model can predict effective AOs (>50% inhibition of gene expression) with a success rate of 87% at 51 predicted high efficacy AOs per 1000 bp. A higher success rate can be obtained at the cost of a lower prediction rate.

When comparing predicted efficacy values to experimental studies there are several issues that should be noted. Firstly the model does not predict all AOs of high efficacy, but a subset of them. This is an issue of sensitivity versus accuracy. We consider that as long as a sufficient number of high efficacy AOs are predicted

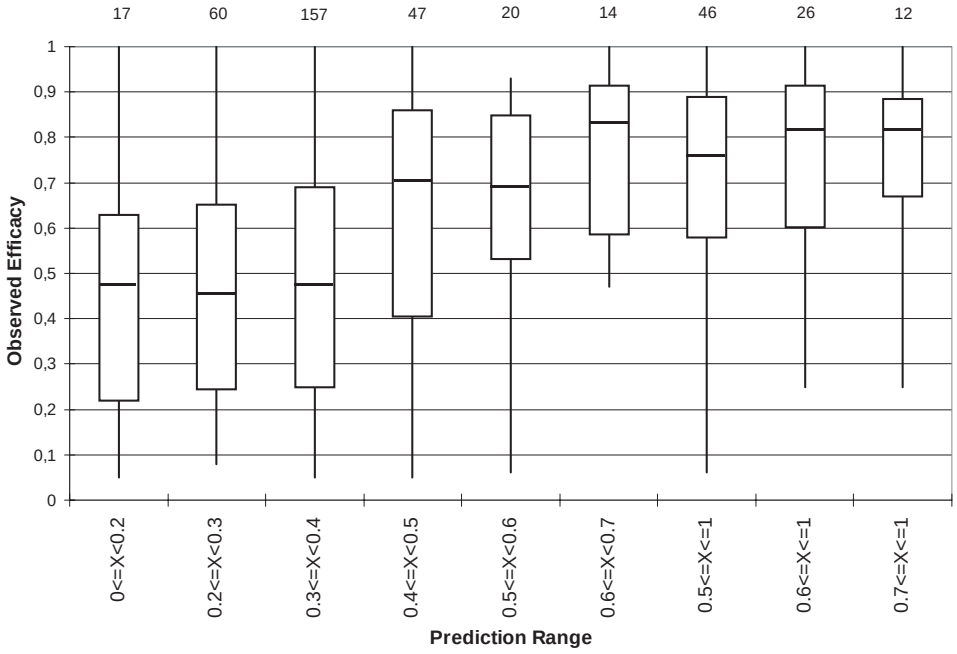

Fig. 4. The distribution of the observed efficacy for each sub-section of predictions is shown. The box represents the upper and lower quartiles with the median enclosed; the lines represent the minimum and maximum observed efficacy. The numbers on top indicate the number of AOs predicted in each range. This result is for an ensemble model incorporating 10 networks with the [14-11-1] architecture.

per gene then we should be as selective as possible. It is however difficult to validate this using published studies. We will often identify AOs that have not been tested, while known AOs of high efficacy will not necessarily be picked up by our model. The ideal scenario for testing the method would be to validate the predicted AOs *in vivo* and/or *in vitro*.

The model often produces a curve with adjacent AOs having correlated predictions, i.e. it tends to predict regions of similar efficacy. This correlation is mainly caused by regions of similar composition and by motifs being shared by overlapping AOs.

Our model can be used in combination with other methods to strengthen our predictions. A specificity screen to check for perfect or near-perfect matches to other target genes helps us avoid non-specific AOs. A more explicit incorporation of (predicted) accessible sites in the target RNA should improve accuracy if they are sufficiently reliable. A web server interface to the model is available at http://www.cgb.ki.se/AOpredict.

ACKNOWLEDGEMENTS

We would like to acknowledge Dr G.C. Tu for the list of AO references, Dr Timothy Bailey for helpful discussions on machine learning, and Dr Boden for his neural network software source code.

REFERENCES

Altschul,S.F., Madden,T.L., Schaffer,A.A., Zhang,J., Zhang,Z., Miller,W. and Lipman,D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.

Amarzguioui,M., Brede,G., Babaie,E., Grotli,M., Sproat,B. and Prydz,H. (2000) Secondary structure prediction and in vitro accessibility of mRNA as tools in the selection of target sites for ribozymes. *Nucleic Acids Res.*, **28**, 4113–4124.

Bacon,T.A. and Wickstrom,E. (1991) Walking along human c-myc mRNA with antisense oligodeoxynucleotides: maximum efficacy at the 5′ cap region. *Oncogene Res.*, **6**, 13–19.

Baldi,P.F. and Brunak,S. (1998) *Bioinformatics: The machine learning approach*. MIT Press, Cambridge, MA, pp. 91–104.

Bennett,C.F., Condon,T.P., Grimm,S., Chan,H. and Chiang,M.Y. (1994) Inhibition of endothelial cell adhesion molecule expression with antisense oligonucleotides. *J. Immunol.*, **152**, 3530–3540.

Chiang,M.Y., Chan,H., Zounes,M.A., Freier,S.M., Lima,W.F. and Bennett,C.F. (1991) Antisense oligonucleotides inhibit intercellular adhesion molecule 1 expression by two distinct mechanisms. *J. Biol. Chem.*, **266**, 18162–18171.

Colige,A., Sokolov,B.P., Nugent,P., Baserga,R. and Prockop,D.J. (1993) Use of an antisense oligonucleotide to inhibit expression of a mutated human procollagen gene (COL1A1) in transfected mouse 3T3 cells. *Biochemistry*, **32**, 7–11.

Crooke,S.T. (2000) Progress in antisense technology: the end of the beginning. *Methods Enzymol.*, **313**, 3–45.

Dean,N.M., McKay,R., Condon,T.P. and Bennett,C.F. (1994) Inhibition of protein kinase C-alpha 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.

Devereux,J., Haeberli,P. and Smithies,O. (1984) A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res.*, **12**, 387–395.

Ding,Y. and Lawrence,C.E. (2001) Statistical prediction of single-stranded regions in RNA secondary structure and application to predicting effective antisense target sites and beyond. *Nucleic Acids Res.*, **29**, 1034–1046.

Duff,J.L., Monia,B.P. and Berk,B.C. (1995) Mitogen-activated protein (MAP) kinase is regulated by the MAP kinase phosphatase (MKP-1) in vascular smooth muscle cells. Effect of actinomycin D and antisense oligonucleotides. *J. Biol. Chem.*, **270**, 7161–7166.

Giddings,M.C., Matveeva,O.V., Atkins,J.F. and Gesteland,R.F. (2000) ODNBase-a web database for antisense oligonucleotide effectiveness studies. *Bioinformatics*, **16**, 843–844.

Hanecak,R., Brown-Driver,V., Fox,M.C., Azad,R.F., Furusako,S., Nozaki,C., Ford,C., Sasmor,H. and Anderson,K.P. (1996) Antisense oligonucleotide inhibition of hepatitis C virus gene expression in transformed hepatocytes. *J. Virol.*, **70**, 5203–5212.

Ho,S.P., Britton,D.H., Stone,B.A., Behrens,D.L., Leffet,L.M., Hobbs,F.W., Miller,J.A. and Trainor,G.L. (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.

Ho,S.P., Bao,Y., Lesher,T., Malhotra,R., Ma,L.Y., Fluharty,S.J. and Sakai,R.R. (1998) Mapping of RNA accessible sites for antisense experiments with oligonucleotide libraries. *Nat. Biotechnol.*, **16**, 59–63.

Hodges,D. and Crooke,S.T. (1995) Inhibition of splicing of wild-type and mutated luciferase-adenovirus pre-mRNAs by antisense oligonucleotides. *Mol. Pharmacol.*, **48**, 905–918.

Laptev,A.V., Lu,Z., Colige,A. and Prockop,D.J. (1994) Specific inhibition of expression of a human collagen gene (COL1A1) with modified antisense oligonucleotides. The most effective target sites are clustered in double-stranded regions of the predicted secondary structure for the mRNA. *Biochemistry*, **33**, 11033–11039.

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

Lefebvre d'Hellencourt,C., Diaw,L., Cornillet,P. and Guenounou,M. (1996) Inhibition of human TNF alpha and LT in cell-free extracts and in cell culture by antisense oligonucleotides. *Biochim. Biophys. Acta*, **1317**, 168–174.

Mathews,D.H., Sabina,J., Zuker,M. and Turner,D.H. (1999) Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J. Mol. Biol.*, **288**, 911–940.

Matveeva,O.V., Tsodikov,A.D., Giddings,M., Freier,S.M., Wyatt,J.R., Spiridonov,A.N., Shabalina,S.A., Gesteland,R.F. and Atkins,J.F. (2000) Identification of sequence motifs in oligonucleotides whose presence is correlated with antisense activity. *Nucleic Acids Res.*, **28**, 2862–2865.

Mayne,M., Ni,W., McKenna,R. and Power,C. (1999) Antisense oligodeoxynucleotides targeting internal exon sequences efficiently regulate TNF-alpha expression. *Antisense Nucleic Acid Drug Dev.*, **9**, 135–144.

Milner,N., Mir,K.U. and Southern,E.M. (1997) Selecting effective antisense reagents on combinatorial oligonucleotide arrays. *Nat. Biotechnol.*, **15**, 537–541.

Miraglia,L., Geiger,T., Bennett,C.F. and Dean,N.M. (1996) Inhibition of interleukin-1 type I receptor expression in human cell-lines by an antisense phosphorothioate oligodeoxynucleotide. *Int. J. Immunopharmacol.*, **18**, 227–240.

Monia,B.P., Johnston,J.F., Geiger,T., Muller,M. and Fabbro,D. (1996) Antitumor activity of a phosphorothioate antisense oligodeoxynucleotide targeted against C-raf kinase. *Nat. Med.*, **2**, 668–675.

Myers,K.J. and Dean,N.M. (2000) Sensible use of antisense: how to use oligonucleotides as research tools. *Trends Pharmacol. Sci.*, **21**, 19–23.

Patzel,V., Steidl,U., Kronenwett,R., Haas,R. and Sczakiel,G. (1999) A theoretical approach to select effective antisense oligodeoxyribonucleotides at high statistical probability. *Nucleic Acids Res.*, **27**, 4328–4334.

Rumelhart,D.E., McClelland,J.L. and the PDP research group, (1986) *Parallel Distributed Processing. Explorations in the Microstructure of Cognition, Volume I: Foundations*. MIT Press, Cambridge, MA.

Scherr,M., Rossi,J.J., Sczakiel,G. and Patzel,V. (2000) RNA accessibility prediction: a theoretical approach is consistent with experimental studies in cell extracts. *Nucleic Acids Res.*, **28**, 2455–2461.

Sczakiel,G. (2000) Theoretical and experimental approaches to design effective antisense oligonucleotides. *Front. Biosci.*, **5**, D194–201.

Sczakiel,G., Homann,M. and Rittner,K. (1993) Computer-aided search for effective antisense RNA target sequences of the human immunodeficiency virus type 1. *Antisense Res. Dev.*, **3**, 45–52.

Song,H.F., Tang,Z.M., Yuan,S.J. and Zhu,B.Z. (2000) Application of secondary structure prediction in antisense drug design targeting protein kinase C-alpha mRNA and QSAR analysis. *Acta Pharmacol. Sin.*, **21**, 80–86.

Stepkowski,S.M., Tu,Y., Condon,T.P. and Bennett,C.F. (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,A.J., Canitrot,Y., Baracchini,E., Dean,N.M., Deeley,R.G. and Cole,S.P. (1996) Reduction of expression of the multidrug resistance protein (MRP) in human tumor cells by antisense phosphorothioate oligonucleotides. *Biochem. Pharmacol.*, **51**, 461–469.

Stutz,A., Huarte,J., Gubler,P., Conne,B., Belin,D. and Vassalli,J.D. (1997) In vivo antisense oligodeoxynucleotide mapping reveals masked regulatory elements in an mRNA dormant in mouse oocytes. *Mol. Cell Biol.*, **17**, 1759–1767.

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

Sugimoto,N., Nakano,S., Yoneyama,M. and Honda,K. (1996) Improved thermodynamic parameters and helix initiation factor to predict stability of DNA duplexes. *Nucleic Acids Res.*, **24**, 4501–4505.

Toschi,N. (2000) Influence of mRNA self-structure on hybridization: computational tools for antisense sequence selection. *Methods*, **22**, 261–269.

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

Vickers,T.A., Wyatt,J.R. and Freier,S.M. (2000) Effects of RNA secondary structure on cellular antisense activity. *Nucleic Acids Res.*, **28**, 1340–1347.

Walter,A.E., Turner,D.H., Kim,J., Lyttle,M.H., Muller,P., Mathews,D.H. and Zuker,M. (1994) Coaxial stacking of helixes enhances binding of oligoribonucleotides and improves predictions of RNA folding. *Proc. Natl Acad. Sci. USA*, **91**, 9218–9222.

Walton,S.P., Stephanopoulos,G.N., Yarmush,M.L. and Roth,C.M. (1999) Prediction of antisense oligonucleotide binding affinity to a structured RNA target. *Biotechnol. Bioeng.*, **65**, 1–9.