Analysis of Antisense Oligodeoxynucleotides Based on Sequence Motif Content to Predict Effectiveness

Michael Giddings[1,3] giddings@genetics.utah.edu
Olga Matveeva[1,3] olgam@howard.genetics.utah.edu
Alex Tsodikov[2]
Raymond F. Gesteland[1] **John F. Atkins**[1]

[1] Department of Human Genetics, 15 N. 2030E., University of Utah, SLC, UT, USA 84112
[2] Huntsman Cancer Institute, University of Utah, SLC, UT, USA 84112
[3] Contributed equally to this work

Keywords: antisense oligodeoxynucleotides , neural network prediciton, sequence motif

1 Introduction

As interest in antisense oligodeoxynucleotide technology continues to grow, the search for a computational means of predicting the efficacy particular oligos has increased as well. Antisense technology allows the targeted down-regulation of an mRNA by application of a complimentary (antisense) phosphorothioate DNA oligo, typically about 20 nt in length. Much remains unknown about which particular target on the mRNA will be most effective, and it is costly and time consuming to perform *in vivo* screenings of multiple oligos to determine the site [1]. This has led to the interest in a computational solution to determining effective oligos/sites. At present, most studies have focused on two primary means of predicting RNA site (oligo) efficacy. One approach examines the predicted secondary structure of the target RNA for sites that might be most accessible to oligo binding [2]. A related but distinct approach calculates binding energies between oligo and target, with more recent studies taking disruption energies of secondary structures into account [3, 4]. These studies have demonstrated some successes, but have not as yet achieved the level of accuracy that would be expected if structural and energetic considerations were the only factors acting *in vivo* to determine antisense efficacy. The work of Tu and colleagues indicated that the short tetranucleotide TCCC, when present in the oligo, greatly improves its chances of having a significant effect on RNA expression [5]. We have further studied correlations between short (3-mer, 4-mer) motifs and oligo efficacy, and have determined there are several motifs which have a strong correlation with antisense action on an RNA target. We have utilized two distinct machine-learning methods to demonstrate that motif-based analysis can be used to predict antisense efficacy with similar or better accuracy as that of the other types of studies.

2 Methods and Results

Statistical analyses of 2, 3, and 4-mer motifs within antisense oligodeoxynucleotides was performed to determine the correlation between certain motifs and oligo effectiveness. The analysis, performed on a database of 349 oligos collected from the literature, demonstrated a strong correlation between certain motifs and the effectiveness of an oligo when tested *in vivo*. Some 4-mer motifs have a strong positive correlation with antisense activity, whereas a few 4-mer and 3-mer motifs demonstrate a negative correlation with antisense activity. These results were further verified by examining the correlations for all data sets where oligos for one particular RNA had been removed. Thus, the resulting correlations were demonstrated to be present independent of the biases possibly present in any particular RNA experiment.

To utilize this for effectiveness prediction based on sequence data, two independent methods were explored, both of which demonstrate an ability to predict effectiveness based on motif content alone. One method applied was Artificial Neural Networks, and the other was Logistic Regression analysis. For both methods cross-validation was performed by training the algorithm on a subset of the database, and using the then-trained system to predict unseen test cases from the database.

The predictions provide a significant improvement over trial-and-error oligo selection. For example, of the 10% of oligos predicted most effective by the neural network, cross validation indicates 43% will be active by causing a decline in mRNA expression by $\geq 75\%$. This is far better than trial-and-error oligo screening, in which less than 1 in 10 oligos screened are generally found effective at the 75% expression reduction level.

The antisense oligo Database collected for this work, and a neural network prediction system based on oligo or RNA input is available at http://antisense.genetics.utah.edu/.

References

[1] Branch, A.D., A good antisense molecule is hard to find, *Trends Biochem. Sci.*, 23(2):45–50, 1998.

[2] Patzel, V., *et al.*, A theoretical approach to select effective antisense oligodeoxyribonucleotides at high statistical probability, *Nucleic Acids Res.*, 27(22):4328–34, 1999.

[3] Stull, R.A., L.A. Taylor, and F.C. Szoka, Jr., Predicting antisense oligonucleotide inhibitory efficacy: a computational approach using histograms and thermodynamic indices, *Nucleic Acids Res.*, 20(13):3501–8, 1992.

[4] Walton, S.P., *et al.*, Prediction of antisense oligonucleotide binding affinity to a structured RNA target, *Biotechnol Bioeng*, 65(1):1–9, 1999.

[5] Tu, G.C., *et al.*, Tetranucleotide GGGA motif in primary RNA transcripts. Novel target site for antisense design, *J. Biol. Chem.*, 273(39):25125–31, 1998.