# Maximum Sequence Alignment Fails to Predict Off-targeted Gene Regulation by RNAi

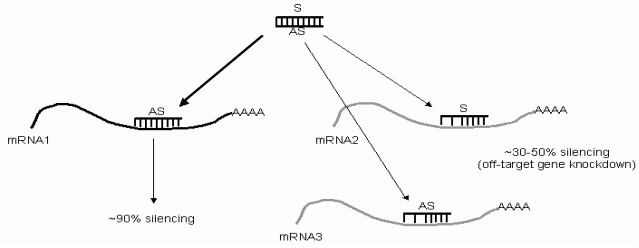
Amanda Birmingham, Emily M.Anderson, William S. Marshall, Anastasia Khvorova
Dharmacon, Inc.
birmingham.a@dharmacon.com

## Abstract

*We have employed various sequence alignment algorithms and scoring techniques to determine whether current computational tools accurately predict genes that will be off-targeted by the RNA interference (RNAi) pathway. Our studies show that distributions of maximum alignment scores for off-targeted and untargeted genes are statistically indistinguishable, indicating that maximum complementarity by itself is an unsatisfactory predictor of off-targeting. Interestingly, a highly significant association was observed between off-targeting and exact complementarity between the seed region (bases 2-7) of siRNA and their off-targeted genes. This pattern has been previously recognized in microRNA-mediated gene knockdown and suggests a distinctive role for the 5' terminus of these strands in RNAi-triggered gene suppression.*

## 1. Introduction: RNAi and Off-target Effects

RNA interference (RNAi) is a revolutionary technique that uses small interfering RNAs (siRNA) to induce gene silencing. RNAi has been employed by researchers in a wide range of functional genomics applications. While early studies predicted RNAi would provide precise gene knockdown, recent investigations [1] have demonstrated that siRNA can induce down-regulation of unintended targets that contain as few as 11 bases of complementarity between the sense or antisense strand of the siRNA and secondary ("off") target.




**Off-Target Gene Regulation. Off-target-siRNA pairings typically have less than 100% complementarity and modulate gene expression 30-50%.**

## 3. siRNA Design

To increase siRNA specificity, siRNA design programs typically incorporate local alignment algorithms (such as BLAST or Smith-Waterman) to anticipate and minimize off-targeting [2,3]. This approach is based on the assumption that since perfect complementarity predicts targeting, imperfect complementarity (or identity, with respect to the opposite strand) will predict off-targeting. Though many siRNA design centers have incorporated this approach to eliminate siRNAs having matches of 15/19 (or even fewer) nt with potential off-target genes, little or no research has been done to confirm that it succeeds in accurately predicting and eliminating off-targeted genes for levels of complementarity lower than 18-19 bases.

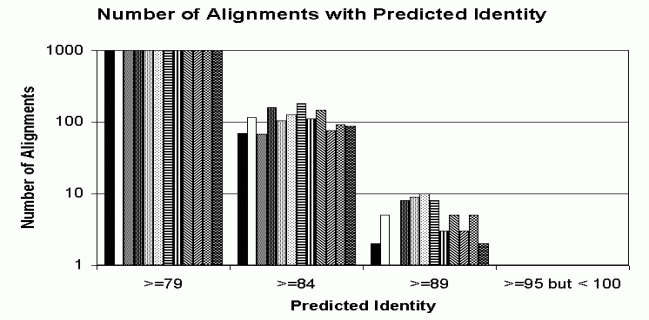
## 4. Prediction Experiments

In order to test the efficacy of using overall sequence complementarity in predicting off-targets, we designed twelve siRNAs (19bp) against three different genes. These duplexes were introduced into cells and the off-targets were identified by microarray analysis.

Simultaneously, Smith-Waterman [4], the optimal local alignment algorithm, was implemented in C# to identify the 1000 best alignments (15/19 or more bases of complementarity) between each test siRNA and the $20{,}000^{+}$ genes represented on the Agilent microarray chip used in these studies. This procedure employed commonly used reward/penalty parameters including a match reward = 2, mismatch penalty = -2, and linear gap penalty = -3.

The results of these studies indicate that the cutoff of 15/19 bp of complementarity is incapable of accurately predicting off-target effects. While wet lab (microarray) experiments demonstrated that each siRNA generated between 5 and 75 different off-targets, the in silico alignment algorithm identified a minimum of 1000 targets (having 15/19 or more alignments) for each duplex. Thus, if the 15/19 cutoff accurately predicted off-targeting, 1-2 orders of magnitude more off-targets should have been found in the microarray experiments. Off-target predictions

at higher cutoffs (with complementarity levels of fewer than 18 bases) were also over-enriched. Overall, (1) only 26 of the predicted 12000 alignments were experimentally validated off-targets, and (2) only 26 of the 366 experimentally validated off-targets were identified by in silico methods, suggesting this methodology grossly overestimates the number of off-targets (false negatives) and serves as a poor predictor of true off-targeted genes.



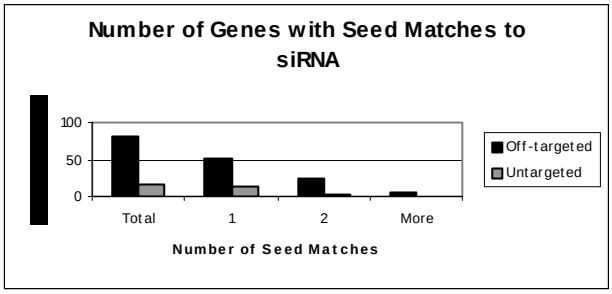

**Predicted Off-target Counts. Bars represent the number of alignments at or above the indicated level of identity for the twelve tested siRNAs.**

Alignment results are notably sensitive to the scoring parameters used. To test whether the poor predictions obtained with commonly used scoring parameters were due to unsuitable constraints, 196 siRNA/mRNA pairs representing known off-target interactions and a second set of (196) siRNA/mRNA pairs with no discernable off-target interactions were scored using twenty-four unique scoring schemes. Examples of new parameters employed in these new schemes include 1) treating both Watson-Crick and wobble pairs as matches, 2) differential scoring of matches and mismatches depending on the two bases involved, and 3) altered gap penalties that allowed only single vs. multi-base gaps. Analysis of the two data sets using each of the new scoring parameters provided distributions that were statistically indistinguishable, suggesting that none of them successfully differentiated off-targeted from untargeted genes.

Overall, these experiments support the conclusion that maximum complementarity with an siRNA, as determined by sequence alignment, does not predict off-targeted genes in any but the most extreme cases. This is contradictory to the assumptions underlying most siRNA design programs.

## 5. Seed Region Complementarity

It has been suggested that siRNAs share the RNAi pathway with micro RNAs (miRNAs), endogenous non-coding ~21-23 nt sequences that have imperfect complementarity to their targets and induce post-transcriptional silencing. Recent studies by several labs [5,6] have shown that the seed region of miRNAs (bases 2-7) frequently have 100% complementarity with one or more regions of their cognate, targeted gene(s). To investigate whether siRNAs share seed complementarity with known off-targets, 98 siRNA/mRNA off-target pairs were analyzed. For each pair, the number of instances of exact complementarity between the siRNA's seed region and the off-targeted mRNA were identified. The same process was performed on a control group that showed no off-target interactions. The results of this study identified a highly significant association between off-targeting and one or more instances of exact seed complementarity. These findings suggest that siRNA off-targeting may operate by a mechanism similar to that of miRNA targeting, and implies a distinctive role for the 5' terminus of active siRNA strands in RNAi-triggered gene suppression.




**Seed Matches and Off-targeting. The total number of genes with at least one seed match to their siRNA for off-targeted and untargeted sets.**

## References


[1] Jackson AL, Bartz SR, Schelter J, Kobayashi SV, Burchard J, Mao M, Li B, Cavet G, Linsley PS. "Expression profiling reveals off-target gene regulation by RNAi", *Nat Biotechnol.*, 2003 Jun;21(6):635-7.

[2] Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs", *Nucleic Acids Res.*, 1997 Sep 1;25(17):3389-402.

[3] Naito Y, Yamada T, Ui-Tei K, Morishita S, Saigo K. "siDirect: highly effective, target-specific siRNA design software for mammalian RNA interference", *Nucleic Acids Res.*, 2004 Jul 1;32(Web Server issue):W124-9.

[4] Smith TF, Waterman MS. "Identification of common molecular subsequences", *J Mol Biol.*, 1981 Mar 25;147(1):195-7.

[5] Lewis BP, Burge CB, Bartel DP. "Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets", *Cell*, 2005 Jan 14;120(1):15-20.

[6] Brennecke J, Stark A, Russell RB, Cohen SM. "Principles of microRNA-target recognition", *PLoS Biol*, 2005 Mar;3(3):e85.