An Algorithm for Generating Small RNAs Capable of Epigenetically Modulating Transcriptional Gene Silencing and Activation in Human Cells

Amanda Ackley1, Alexandra Lenox1, Kenneth Stapleton1, Stuart Knowling1, Tim Lu1, Kenny S Sabir2, Peter K Vogt1 and Kevin V Morris1,3

Small noncoding antisense RNAs (sasRNAs) guide epigenetic silencing complexes to target loci in human cells and modulate gene transcription. When these targeted loci are situated within a promoter, long-term, stable epigenetic silencing of transcription can occur. Recent studies suggest that there exists an endogenous form of such epigenetic regulation in human cells involving long noncoding RNAs. In this article, we present and validate an algorithm for the generation of highly effective sasRNAs that can mimic the endogenous noncoding RNAs involved in the epigenetic regulation of gene expression. We validate this algorithm by targeting several oncogenes including AKT-1, c-MYC, K-RAS, and H-RAS. We also target a long antisense RNA that mediates the epigenetic repression of the tumor suppressor gene DUSP6, silenced in pancreatic cancer. An algorithm that can efficiently design small noncoding RNAs for the epigenetic transcriptional silencing or activation of specific genes has potential therapeutic and experimental applications.

Molecular Therapy—Nucleic Acids (2013) 2, e104; doi:10.1038/mtna.2013.33; published online 9 July 2013

Subject Category: siRNAs, shRNAs, and miRNAs

Introduction

Recent studies have revealed a mechanism of transcriptional regulation involving long, noncoding RNAs. These RNAs are antisense to protein-coding genes and affect transcription by epigenetic modification of the chromatin.1 Parts of this endogenous regulatory pathway overlap with steps known from transcriptional gene silencing (TGS) with small noncoding antisense RNAs (sasRNAs) or small interfering RNAs (siRNAs).2 SasRNAs, designed to target promoter sequences, can induce silent-state epigenetic changes at the targeted loci, mediating long-term, stable TGS.3 These small RNAs are exogenous to the cell, but there is also emerging evidence that their mechanisms of action involve endogenous RNA-mediated transcriptional controls.4–6 Intervention in such endogenous mechanisms could have important medical applications.

Transcriptional controls are important in cancer. The gain of function in oncogenes can result from transcriptional upregulation, and loss of function in tumor suppressor genes is often due to transcriptional silencing.7–9 Silenced genes typically show characteristic epigenetic marks at their promoters.10,11 Some tumor suppressor genes, such as p15, p21, and PTEN, are regulated by endogenous antisense transcripts.4–6 The epigenetic silencing of p21 and PTEN4–6 is mechanistically similar to TGS induced by promoter-targeting sasRNAs.2 These observations suggest the existence of endogenous RNA-mediated transcriptional regulation that works through epigenetic modification and involves natural antisense transcripts.1,12,13

To date, however, while there are extensive siRNA design algorithms for RNA interference-mediated gene silencing,14,15 the design of promoter-targeting sasRNAs, which would mimic the endogenous natural antisense transcript pathway, has not been developed. We surmised, based on previous observations,16,17 that TGS works through an RNA–RNA interaction at the targeted promoter. It may then be possible to rationally design sasRNAs capable of inducing TGS in human cells.16,18

In this article, we present an algorithm for the identification of small RNAs that can mediate TGS. We include examples of TGS of several oncogenes and TGS that affects the promoter of an endogenous natural antisense transcript. Effective TGS operating on the promoter of a natural antisense transcript can elevate the transcription from the sense promoter, offering a method to activate gene expression.

Results

Small RNAs effective in the induction of TGS

SasRNAs, siRNAs, or short hairpin RNAs can induce TGS of various gene promoters in human cells.2 However, not all small RNAs show this effect on transcription. We compared effective versus noneffective small RNAs from published literature in order to identify shared features that are correlated with transcriptional regulator activity (Supplementary Tables S1 and S2, respectively). Small RNAs that function in initiating TGS typically show a string of purines from ~4 to 12bp in the region targeting the promoter sequence (Supplementary Table S1). Long noncoding RNAs, which are the endogenous drivers of TGS in human cells, also

1Department of Molecular and Experimental Medicine, The Scripps Research Institute, La Jolla, California, USA; 2Garvan Institute of Medical Research, Sydney, Australia; 3School of Biotechnology and Biomolecular Sciences, The University of New South Wales, Sydney, Australia. Correspondence: Kevin V Morris, Department of Molecular and Experimental Medicine, The Scripps Research Institute, 10550 N. Torrey Pines Road, La Jolla, California 92037, USA. E-mail: kmorris@scripps.edu or Peter K Vogt, Department of Molecular and Experimental Medicine, The Scripps Research Institute, 10550 N. Torrey Pines Road, La Jolla, California 92037, USA. E-mail: pkvogt@scripps.edu

Keywords: epigenetic; non-coding RNA; oncogene; transcription; tumor suppressor gene

Received 15 May 2013; accepted 15 May 2013; advance online publication 9 July 2013. doi:10.1038/mtna.2013.33
contain strings of purines, suggesting it is of biological relevance in RNA-directed control of gene expression.\textsuperscript{19,20} We speculated that this purine string was mechanistically relevant for TGS. In order to examine this possibility, we produced a computer algorithm taking into account the observed strings of purines as well as GC pairing (Figure 1a). In using this algorithm (http://www.morrislab.unsw.edu.au/TGSalgorithm/Algorithm/purine_finder.html), sequences from a desired target locus, generally a gene promoter, are copied and pasted into the “paste sequence” box. The algorithm then generates and displays the top five candidate targeting sequences in the "output sequence" box (Figure 1a). The candidate targeting sequences are of variable length from ~18 to 30 bp and are displayed in the sense orientation.

We first validated this algorithm for TGS of several oncogenes. The promoters of AKT-1, MYC, K-RAS, and H-RAS were targeted, and the top candidate sasRNAs were generated (Figure 1b). sasRNAs that effectively induce TGS were identified for all of the targeted oncogene promoters with a success rate greater than 50%. Most effective was an sasRNA targeting AKT-1 (as-A3, Figure 1c) and one directed to MYC (as-M1, Figure 1d). Transcription of K-RAS and H-RAS was also susceptible to algorithm-designed sasRNAs (Supplementary Figure S1a,b). A mutational analysis of the algorithm

In order to examine the role of the algorithm-generated targeting sequence and of structural parameters in the induction of TGS, we performed a mutational analysis. We chose the functionally active small RNAs as-A3 and as-M1 (Figure 1c,d) to introduce two kinds of mutations. In the first kind, we changed the sequence of the conserved purine string (Figure 1b). These mutations are subsequently referred to as sequence changes (Figure 2a and Supplementary Figure S2a). In the second kind of mutations, we introduced structural changes that affect the predicted folding of the sasRNA. These mutations are referred to as structural changes (Figure 2a and Supplementary Figure S2a). One fundamental difference between these two classes of mutations can be found in their predicted overall thermodynamic stability (Figure 2a and Supplementary Figure S2a). The thermodynamic stability affects the ability of RNA to interact with target molecules. In these experiments, structural changes in the sasRNAs affected their TGS-inducing activity (Figure 2c and Supplementary Figure S2c, respectively). Changes in the purine string had a small effect on the induction of TGS. However, in the AKT-targeted series, the effect of multiple sequence changes appeared additive (Supplementary Figure S2b). Unexpectedly, one of the mutations in Myc1 proved functional (Figure 2b).
Figure 2  The effects of mutations on small noncoding antisense RNA (sasRNA) function. (a) Changes in the sequence and structure of the algorithm-defined sasRNAs for MYC. The effects of (b) sequence versus (c) structural mutations on sasRNA-mediated transcriptional gene silencing of MYC. (d) M-fold was used to generate the predicted secondary structures for the functional versus nonfunctional sasRNAs. For b and c, the averages of triplicate-treated cultures are shown with the SEM and P values from a paired t-test.

sasRNA, termed as-M10, as well as the other active sasRNAs, retained the purine string in the predicted loop of the construct when compared by the M-fold program with inactive constructs (Figure 2d and Supplementary Figure S2d). This observation may suggest that the sequence GAAC in promoter-associated transcripts, or GUUC in the loop of the small promoter-targeting RNA, is of functional significance in the induction of TGS in human cells (Figure 1b and Supplementary Figure S2e).

Small RNA-induced TGS alters the chromatin architecture of the target locus

Previous studies of TGS induced by either small antisense or long antisense RNA have identified a role of DNA methyltransferase 3A (DNMT3a), histone deacetylase 1 (HDAC-1), or long antisense RNA have identified a role of DNA methyltransferase 3A (DNMT3a), histone deacetylase 1 (HDAC-1), or Enhancer of zeste 2 (EZH2).3,6,21–23 In order to determine whether these chromatin-modifying proteins are involved in TGS induced by as-M1, as-M10, or as-A3, we suppressed these chromatin modifiers with RNA interference (Supplementary Figure S3). The action of the sasRNAs was abrogated by the suppression of DNMT3a, HDAC-1, or Ezh2, suggesting that these epigenetic remodeling components are involved in the TGS induced by the algorithm-designed sasRNA constructs (Figure 3a).

To further confirm that the constructs as-M1 and as-M10 actively induce TGS of the MYC gene, we performed quantitative reverse transcription-PCR for the unspliced MYC transcripts following treatment of the cells with as-M1 or as-M10. The assessment of unspliced transcripts, an indication of ongoing transcription, suggested that both as-M1 and as-M10 suppress transcription of MYC (Figure 3b). In these and previous experiments, as-M10 showed higher activity and was, therefore, used in chromatin immunoprecipitation to survey the epigenetic landscape of the MYC gene after induction of TGS. These tests revealed a modest gain of H3K27me3 specifically at the as-M10 target locus as well as a reduction in histone 3 at the target locus and various regions around the as-M10-targeted sequence (Figure 3c). These observations suggest that the epigenetic architecture of the MYC gene was modified by the action of as-M10, similar to previous observations with sasRNAs directing TGS in human cells.3,21,23–25

Collectively, these data show that the experimental algorithm is capable of identifying regions in promoters that are susceptible to the induction of sasRNA-mediated TGS.
observations also suggest that the retention of the purine string in the loop predicted by the M-fold program may be important for activity.

Targeting an endogenous antisense transcript affects gene expression

Endogenous antisense transcripts (natural antisense transcripts) can modulate the epigenetic states of some genes.¹ In order to explore possible mechanisms for this activity, we used the new algorithm for selecting TGS-active sasRNAs to target the putative promoter of an endogenous antisense transcript, EST AK124265. EST AK124265 was found on the UCSC genome browser to align antisense to the promoter of dual-specificity phosphatase (DUSP6) (Figure 4a). DUSP6 is a regulator of extracellular signal-regulated kinase. It is epigenetically silenced in some pancreatic cancers with significant DNA methylation found at the DUSP6 promoter.²⁶⁻²⁷ We surmised that AK124265 is endogenously active in controlling the transcription and epigenetic states of DUSP6.

To explore this hypothesis, we first targeted an siRNA construct to a transcribed noncoding RNA sequence of AK124265 and determined the effects on the expression of AK124265 and of DUSP6 (Supplementary Figure S4a). We observed the expected reduction in the level of AK124265 and measured a concomitant increase in the level of DUSP6 (Supplementary Figure S4b). This observation suggested a link between AK124265 and DUSP6. We, therefore, attempted to control the levels of AK124265 by targeting the promoter of this natural antisense transcript. We used algorithm-generated sasRNA constructs directed to four regions upstream of the predicted transcriptional start site of AK124265 (Figure 4d). These regions showed promoter activity in a luciferase reporter assay (Figure 4b). Only one of the algorithm-generated sasRNA constructs, as-AK4, induced elevated expression of DUSP6 (Figure 4d) while
suppressing AK124265 (Figure 4c). The suppression of DNMT3a or DNMT3a, HDAC-1, and EZH2 in combination also induced marked increases in DUSP6 (Figure 4e), suggesting that DUSP6 is endogenously regulated by AK124265 by RNA-directed TGS.

Discussion

In this study, we have used algorithm-designed sasRNAs to upregulate and to downregulate transcription from specific genes. Our observations suggest that in human cells, such sasRNAs can affect promoter activity of protein-coding and of noncoding transcripts. Targeting the promoter of a protein-coding transcript can reduce transcription of the gene, whereas targeting of the promoter of the corresponding native antisense transcript results in the elevation of gene expression. The induction of TGS by siRNAs that target promoter sequences has been documented for several genes (reviewed in ref. 2). Available evidence suggests that this activity involves chromatin-modifying mechanisms. In cell culture, TGS is stable over prolonged time periods due to an epigenetic-based mechanism involving both histone and DNA methylation being directed to the small RNA-targeted promoter loci. The computer algorithm presented in this paper permits the design of RNA constructs that efficiently induce TGS. Important features of this TGS include targeting of promoter sequences at one to three nucleosomes upstream of the dominant transcription start site and the involvement of chromatin-modifying proteins. Constructs designed by the algorithm have proved effective in reducing transcriptional activities from the promoters of MYC and of H-RAS. The transcriptional activity of MYC has been reduced with siRNAs before, but one of these constructs did not target the promoter upstream and probably worked through blockage of RNA polymerase II rather than through epigenetic change. The algorithm described here is empirically derived and incorporates both sequence and structural features. It enhances the success rate of designing sasRNA constructs that are effective in transcriptional regulation but should be considered a “β” version and is open to modification and improvement.

The successful targeting of natural antisense transcripts with small RNAs is less common. In mammalian cells, antisense transcripts were first discovered by Williams and Fried 1986; another antisense transcript identified early...
was that of MYC. Today endogenous antisense transcription is recognized as pervasive and as an integral cellular activity. The regulatory potential of antisense transcripts was immediately recognized, and their main activity was initially linked to interference with protein translation. However, more recent observations suggest an important role of natural antisense transcripts in the regulation of gene transcription by epigenetic modification of chromatin.

We have used the newly developed computer algorithm to target sasRNAs for the first time to a putative promoter of a natural antisense transcript, EST AK124265. This natural antisense transcript spans the promoter of DUSP6, a tumor suppressor which is transcribed from the complementary DNA strand. The underlying assumption in targeting the promoter of EST AK124265 is that this antisense transcript functions as a negative regulator of DUSP6 whose promoter is heavily methylated and epigenetically silenced in pancreatic cancers. The transcriptional derepression of DUSP6 by sasRNA directed at the EST AK124265 promoter provides initial support for this assumption. More detailed studies on the molecular mechanism of this transcriptional activation are currently in progress.

sasRNAs directed to promoter sequences can be used both to inhibit and to activate transcription, depending on the sense of the target DNA strand. In general, the inhibition of active gene function can be readily accomplished. Options besides sasRNA-induced TGS include microRNAs that interfere with translation or direct inhibition of the target protein by small molecules or biomolecules. In contrast, restoration of lost function is a much greater challenge. Natural antisense transcripts may be widely involved in epigenetic gene silencing (discussed in refs. 13, 14), and targeting the promoters of such transcripts could provide a valuable technique for the reactivation of silenced genes. The algorithm presented here will facilitate this novel approach to the restoration of gene expression.

**Materials and methods**

**Cell culture and sasRNA validation studies.** Human 293HEK and CF PAC cell lines were transfected with the sasRNA pU6M2-based expressing plasmids at a ratio of ~1 µg/10⁶ cells. The “Control” used in this study was the U6M2 plasmid without any sasRNA expressed. Transfection of 293HEK and CF PAC cells was carried out using Lipofectamine 2000. Transfected cultures were assessed 72 hours post-transfection and gene expression was determined by quantitative reverse transcription-PCR.

**Small RNA-expressing vectors.** SasRNAs were cloned into the pU6M2 expression vector (Supplementary Figure S5). For each expressed small RNA, two oligonucleotides were designed containing a 5’ Bgl-II and 3’ Kpn-I restriction sites (Supplementary Table S3). The oligonucleotides were then treated with T4 polynucleotide kinase and heat-inactivated at 65 °C for 20 minutes. The oligonucleotides were mixed, incubated in a boiling water bath for 1 minute, and then allowed to slowly cool to room temperature. The oligonucleotides were then diluted 1:25 in water, digested, and ligated into the pU6M2 plasmid that had been treated the same way. Positive clones were collected (Qiagen QIAprep Miniprep on the Qiacube; Qiagen, Valencia, CA) and screened by PCR.

*An Algorithm for Small RNA TGS*  
Ackley et al.
strings from 4 to 12bp of the particular promoter-targeted locus, based on observations with those small noncoding RNAs that have been observed previously to direct TGS in human cells (contrast Supplementary Table S1 with Supplementary Table S2). Those promoter-targeted loci which are utilized in the algorithm are generally one to three nucleosomes upstream of the dominant transcription start site as these regions tend to favor the involvement of chromatin-modifying proteins. The resulting sasRNA target sites should be assessed using BLAT in the UCSC genome browser to confirm whether targeting is specific to the intended target loci and the other loci are not secondarily targeted.

Supplementary material

Figure S1. Algorithm-generated small antisense RNA target validation.

Figure S2. Point mutation analysis of sasRNAs targeted to the AKT-1 promoter reveals a structural aspect to the algorithm-designed functional asRNAs.

Figure S3. Suppression of epigenetic regulatory factors.

Figure S4. siRNA targeting DUSP6 and AK124265 in CF-PAC cells.

Figure S5. The U6M2 insertion for generating asRNAs.

Table S1. Functional TGS targets.

Table S2. Nonfunctional TGS targets.

Table S3. Oligonucleotides used in this study.

Acknowledgments.

Research reported in this publication was supported by the National Institute of Allergy and Infectious Diseases of the National Institutes of Health under award number R01AI084406 and P01AI09783-01 (K.V.M.) and by the National Institute of Cancer of the National Institutes of Health under award numbers CA151574 and RO1 CA153124 (P.K.V.). A provisional patent filing for the algorithm has been initiated as US Patent 12-11422-12.

Table S1. Functional TGS targets.

Table S2. Nonfunctional TGS targets.

Table S3. Oligonucleotides used in this study.

Acknowledgments.

Research reported in this publication was supported by the National Institute of Allergy and Infectious Diseases of the National Institutes of Health under award number R01AI084406 and P01AI09783-01 (K.V.M.) and by the National Institute of Cancer of the National Institutes of Health under award numbers CA151574 and RO1 CA153124 (P.K.V.). A provisional patent filing for the algorithm has been initiated as US Patent 12-11422-12.

Table S1. Functional TGS targets.

Table S2. Nonfunctional TGS targets.

Table S3. Oligonucleotides used in this study.

Acknowledgments.

Research reported in this publication was supported by the National Institute of Allergy and Infectious Diseases of the National Institutes of Health under award number R01AI084406 and P01AI09783-01 (K.V.M.) and by the National Institute of Cancer of the National Institutes of Health under award numbers CA151574 and RO1 CA153124 (P.K.V.). A provisional patent filing for the algorithm has been initiated as US Patent 12-11422-12.

Table S1. Functional TGS targets.

Table S2. Nonfunctional TGS targets.

Table S3. Oligonucleotides used in this study.

Acknowledgments.

Research reported in this publication was supported by the National Institute of Allergy and Infectious Diseases of the National Institutes of Health under award number R01AI084406 and P01AI09783-01 (K.V.M.) and by the National Institute of Cancer of the National Institutes of Health under award numbers CA151574 and RO1 CA153124 (P.K.V.). A provisional patent filing for the algorithm has been initiated as US Patent 12-11422-12.

Table S1. Functional TGS targets.

Table S2. Nonfunctional TGS targets.

Table S3. Oligonucleotides used in this study.

Acknowledgments.

Research reported in this publication was supported by the National Institute of Allergy and Infectious Diseases of the National Institutes of Health under award number R01AI084406 and P01AI09783-01 (K.V.M.) and by the National Institute of Cancer of the National Institutes of Health under award numbers CA151574 and RO1 CA153124 (P.K.V.). A provisional patent filing for the algorithm has been initiated as US Patent 12-11422-12.

Table S1. Functional TGS targets.

Table S2. Nonfunctional TGS targets.

Table S3. Oligonucleotides used in this study.

Acknowledgments.

Research reported in this publication was supported by the National Institute of Allergy and Infectious Diseases of the National Institutes of Health under award number R01AI084406 and P01AI09783-01 (K.V.M.) and by the National Institute of Cancer of the National Institutes of Health under award numbers CA151574 and RO1 CA153124 (P.K.V.). A provisional patent filing for the algorithm has been initiated as US Patent 12-11422-12.

Table S1. Functional TGS targets.

Table S2. Nonfunctional TGS targets.

Table S3. Oligonucleotides used in this study.

Acknowledgments.

Research reported in this publication was supported by the National Institute of Allergy and Infectious Diseases of the National Institutes of Health under award number R01AI084406 and P01AI09783-01 (K.V.M.) and by the National Institute of Cancer of the National Institutes of Health under award numbers CA151574 and RO1 CA153124 (P.K.V.). A provisional patent filing for the algorithm has been initiated as US Patent 12-11422-12.

Table S1. Functional TGS targets.

Table S2. Nonfunctional TGS targets.

Table S3. Oligonucleotides used in this study.

Acknowledgments.

Research reported in this publication was supported by the National Institute of Allergy and Infectious Diseases of the National Institutes of Health under award number R01AI084406 and P01AI09783-01 (K.V.M.) and by the National Institute of Cancer of the National Institutes of Health under award numbers CA151574 and RO1 CA153124 (P.K.V.). A provisional patent filing for the algorithm has been initiated as US Patent 12-11422-12.