Effective expression of small interfering RNA in human cells

Cynthia P. Paul1, Paul D. Good1, Ira Winer2, and David R. Engelke1,2*

In many eukaryotes, expression of nuclear-encoded mRNA can be strongly inhibited by the presence of a double-stranded RNA (dsRNA) corresponding to exon sequences in the mRNA (refs 1,2). The use of this “RNA interference” (RNAi) in mammalian studies had lagged well behind its utility in lower animals because uninterrupted RNA duplexes longer than 30 base pairs trigger generalized cellular responses through activation of dsRNA-dependent protein kinases. Recently it was demonstrated that RNAi can be made to work in cultured human cells by introducing shorter, synthetic duplex RNAs (~20 base pairs) through liposome transfection. We have explored several strategies for expressing similar short interfering RNA (siRNA) duplexes within cells from recombinant DNA constructs, because this might allow long-term target-gene suppression in cells, and potentially in whole organisms. Effective suppression of target gene product levels is achieved by using a human U6 small nuclear RNA (snRNA) promoter to drive nuclear expression of a single RNA transcript. The siRNA-like parts of the transcript consist of a 19–base pair siRNA stem with the two strands joined by a tightly structured loop and a U-U-3’ overhang at the end of the antisense strand. The simplicity of the U6 expression cassette and its widespread transcription in human cell types suggest that this mode of siRNA delivery could be useful for suppressing expression of a wide range of genes.


*Corresponding author (engelke@umich.edu).
transfected cells. Production of siRNA constructs nonspecifically obstruct protein synthesis. Figure 2B RNA blot analysis has shown that high levels (104–105 RNA molecule) of the lamin A/C protein in transfected cell nuclei are transfected. For testing recombinant DNAs, cells were cotransfected (red) substantially disappears from most cells, presumably those that staining (blue). As expected from previous work, the lamin A/C signal was marked by 4,6-diamidino-2-phenylindole (DAPI) Figure 2A were visualized with 4,6-diamidino-2-phenylindole (DAPI) expression of the siRNA constructs. However, as seen elsewhere in this issue (Lee et al., p. 500), the synthesis of the siRNA as independent strands from U6 promoter can also be effective.

To make the siRNA duplex as one short transcript, an RNA insert was used that contains the 19-nucleotide sense strand of the target, followed by a UUCG tetraloop sequence, the antisense strand, and a stem overhangs. Results with synthetic siRNA4 suggest that such 3′ overhangs can increase efficacy. RNA blot analysis has shown that high levels (~105–106 RNA molecules/cell) of nearly full-length RNAs can be expressed from these cassettes. The “hairpin siRNAs” give comparable expression levels, although there is a complex pattern of breakdown products as well as the full-length product (not shown).

Figure 2 shows elimination of lamin A/C (red), β-Gal (green), or the overlay of the two signals. Without the siRNA inserts, cells transfected with any of the expression cassette plasmids do not have detectably reduced lamin A/C signal (shown only for U6+1 in Fig. 2B). When either U6+1 or U6+27 cassettes were used with anti-lamin hairpin siRNA inserts, dramatic reductions of lamin A/C signal were observed relative to the untransfected cells in the same fields. Transfected cells receiving the U6+27-siRNA expression cassettes gave the most consistent and greatest lamin A/C reductions (>90%, Table 1), similar to synthetic siRNA (~95%). This might reflect a threshold effect caused by lower levels of the U6+1-expressed siRNA1.

Figure 2C shows lamin A/C–β-Gal overlay panels for control RNAs expressed from the U6+27 cassette. Expression of only the sense or only the antisense strands of the siRNA in U6+27 did not affect lamin levels, reinforcing the notion that the observed reduction in Figure 2B requires the duplex, a hallmark of siRNA action.

We next tested a U6+27 hairpin siRNA construct with the order of the strands reversed to determine the specific need for an accessible 3′ overhang on the antisense strand of the duplex. Some models for siRNA function predict that siRNA degradation of the target message is amplified by annealing of the antisense strand to the mRNA and extension to a longer duplex with an RNA-dependent RNA polymerase. This condition would indicate the need to have an accessible antisense 3′ terminus so that it can be extended. Surprisingly, there was a significant reduction of the lamin signal with the reversed-strand construct, although it was not as consistent or effective as the original orientation. It is not clear why the reversed-strand construct causes partial reduction of the lamin signal. It is possible that small amounts of breakdown products with 3′-UU overhang are created on the antisense strand of the reversed construct by 3′ exonuclease digestion or a discrete endonuclease cleavage between the strands. Alternatively, these hairpin siRNAs, when expressed within the cells, might not need to act exclusively through primer extension amplification. Although the active form of the nuclear-expressed RNAs will require long-term investigation, we recommend that siRNA transcripts have the sense strand first, followed by a tetraloop and antisense strand ending with a 3′ overhang created by the poly(U) terminator.

Previous studies of siRNA-mediated target cleavage by extracts in vitro suggested that the 5′ termini of one or both strands might need to be phosphorylated, and that this might be needed for efficient assembly into obligatory ribonucleoprotein complexes6,7. Results pre-

![Image](http://biotech.nature.com)

**Table 1. Effect of siRNA and expression cassettes on the levels of the lamin A/C protein in transfected cell nuclei**

<table>
<thead>
<tr>
<th>Construct</th>
<th>Percentage lamin A/C in transfected vs. nontransfected cells</th>
</tr>
</thead>
<tbody>
<tr>
<td>pAVU6+27 No insert</td>
<td>130 ± 5</td>
</tr>
<tr>
<td>Synthetic anti-lamin siRNA</td>
<td>5 ± 2</td>
</tr>
<tr>
<td>pAVU6+27 Anti-lamin siRNA hairpin</td>
<td>9 ± 5</td>
</tr>
<tr>
<td>pAVU6+27 Sense strand only</td>
<td>130 ± 40</td>
</tr>
<tr>
<td>pAVU6+27 Antisense strand only</td>
<td>130 ± 30</td>
</tr>
<tr>
<td>pAVU6+27 Reverse-strands hairpin</td>
<td>25 ± 14</td>
</tr>
</tbody>
</table>
Figure 2. Effects of siRNA constructs on lamin A/C levels. HeLa cells were transfected with either synthetic siRNA or recombinant DNA cassettes expressing different small RNAs from different RNA polymerase III promoters. Cells were stained with DAPI (blue) or with antibodies to lamin A/C (red) or β-Gal (green). (A) Synthetic siRNA or no RNA transfections, showing that lamin A/C staining of the nuclear periphery is largely abolished in most cells, with only low levels of residual red staining in nuclear interiors. (B) Transfection with U6 promoter cassettes either without an siRNA insert (U6+1, no insert) or containing the anti-lamin siRNA shown in Figure 1B (U6+1 siRNA and U6+27 siRNA). Transfected cell cytoplasts are green, whereas nuclei from untransfected cells show no green cytoplasm. Empty expression cassettes have no apparent effect on lamin A/C levels (only empty U6+1 is shown), while transfected cells (green) using siRNA-expressing constructs have little remaining lamin A/C (red). (C) Overlay panels of β-Gal and lamin A/C signal after transfection with different control insertions shown in Figure 1B. Quantitative assessment of remaining lamin A/C signals in transfected cells compared to untransfected cells on the same slide is given in Table 1.

Experimental protocol

Materials. Lipofectin, Plus reagent, and Oligofectamine were purchased from Invitrogen (Carlsbad, CA), as were synthetic DNA oligonucleotides for cloning and probes. Cy3-2'-O-methyl RNA oligonucleotide hybridization probes were from Trilink (San Diego, CA). Synthetic siRNA oligonucleotides were from Dharmacon (Lafayette, CO). Anti-lamin A/C monoclonal antibodies were purchased from Santa Cruz Biotechnology (Santa Cruz, CA) (sc-7292, used at 1 μg/ml); rabbit anti-β-Gal antibodies were from Molecular Probes (Eugene, OR) (A-11312, 1 μg/ml); Oregon green 488–labeled goat anti-rabbit secondary antibodies were from Molecular Probes (O-10383, 5 μg/ml); and cyanin-3 (Cy3)-labeled goat anti-mouse secondary antibodies were from Amersham-Pharmacia Biotech (Piscataway, NJ) (PA 43002, 1 μg/ml); and cyanin-3 (Cy3)-labeled goat anti-mouse secondary antibodies were from Amersham-Pharmacia Biotech (Piscataway, NJ) (PA 43002, 1 μg/ml).

Figure 3. Localization of U6+27 siRNA transcripts. Two days after transfection with the U6+27 anti-lamin A/C cassette plasmid, cells were fixed and stained for nuclear DNA (DAPI, blue) and probed with a Cy3-labeled 2'-O-methyl oligoribonucleotide (red) complementary to the antisense strand of the siRNA. As expected from work with previous U6 expression constructs9,10, the U6+27 siRNA pattern was primarily in a nuclear speckled pattern. Nuclear and cytoplasmic background staining by the Cy3-oligonucleotide in the absence of U6+27-siRNA (“Mock”) was minimal.
Cassettes were cloned in pAV vectors, derived from pCWBSVN (ref. 11) by placing the promoter modules between BamHI and HindIII sites, after modifying the vector. Modifications included destruction of the BamHI site downstream of the Neo cassette, and removal of all sites between the original Salt and Xhol sites, inclusive, by cleavage and religation. After inserting the cassettes, a new polynicker was created between the HindIII and Salt sites.

Sequence data suggests that >75% of 21 bp tags would be expected to occur only once in the genome size of the human genome (Table 1A). Likewise, similar analyses based on actual sequence information from ~16,000 known genes suggest that >75% of 21 bp tags would be expected to occur only once in the human genome, with the remaining tags matching duplicated genes or repeated sequences (as discussed below). In contrast, conventional SAGE tags of 14 bp do not allow unique assignment of tags to genomic sequences, though they do allow such assignment to the much less complex compendium of expressed sequence tags (ESTs) and previously characterized mRNAs of other species, and have made another ~10,000–20,000 gene predictions of lower confidence, supported by various types of in silico evidence, including homology studies, domain searches, and ab initio gene predictions.

Acknowledgments

This work was supported by NIH grant AI40936 to D.R.E. and the Medical Scientist Training Program at the University of Michigan. We thank Gary R. Kunkel for the cloned, human U6 snRNA gene, and Michael Imperiale and Eric Fearon for other materials.

Competing interests statement

The authors declare competing financial interests: see the Nature Biotechnology website (http://biotech.nature.com) for details.

Received 28 December 2001; accepted 19 March 2002

A remaining challenge for the human genome project involves the identification and annotation of expressed genes. The public and private sequencing efforts have identified ~15,000 sequences that meet stringent criteria for genes, such as correspondence with known genes from humans or other species, and have made another ~10,000–20,000 gene predictions of lower confidence, supported by various types of in silico evidence, including homology studies, domain searches, and ab initio gene predictions.

The computational methods have limitations, both because they are unable to identify a significant fraction of genes and exons and because they are unable to provide definitive evidence about whether a hypothetical gene is actually expressed. As the in silico approaches identified a smaller number of genes than anticipated, we wondered whether high-throughput experimental analyses could be used to provide evidence for the expression of hypothetical genes and to reveal previously undiscovered genes. We describe here the development of such a method—called long serial analysis of gene expression (LongSAGE), an adaption of the original SAGE approach—that can be used to rapidly identify novel genes and exons.

The LongSAGE method (Fig. 1) generates 21 bp tags derived from the 3′ ends of transcripts that can rapidly be analyzed and matched to genomic sequence data. The method is similar to the original SAGE approach, but uses a different type IIS restriction endonuclease (Mmel) and incorporates other modifications to produce longer transcript tags. The resulting 21 bp tag consists of a constant 4 bp sequence representing the restriction site at which the transcript was cleaved, followed by a unique 17 bp sequence derived from an adjacent sequence in each transcript. Theoretical calculations show that >99.8% of 21 bp tags are expected to occur only once in the genome size of the human genome (Table 1A). Likewise, similar analyses based on actual sequence information from ~16,000 known genes suggest that >75% of 21 bp tags would be expected to occur only once in the human genome, with the remaining tags matching duplicated genes or repeated sequences (as discussed below). In contrast, conventional SAGE tags of 14 bp do not allow unique assignment of tags to genomic sequences, though they do allow such assignment to the much less complex compendium of expressed sequence tags (ESTs) and previously characterized mRNAs of other species, and have made another ~10,000–20,000 gene predictions of lower confidence, supported by various types of in silico evidence, including homology studies, domain searches, and ab initio gene predictions.