shRNA design

shRNAs or short hairpin RNAs are artificial constructs that can be inserted into a genome and expressed endogenously**[5]**. The expressed hairpins can then fold to form dsRNA, and Drosha and Dicer can then act on these hairpins to create mature sequence, used by the RISC complex to target the genes. The design used in a library created by the Hannon lab at CSHL [24] is shown in figure 2. If a 29 mer is used for the hairpin as shown in the figure, then the first 22 mer from the end is believed to be used for creating the siRNA. siRNA design considerations are used to design 19 mers that are then extended two base pairs at the 5' end and one at the 3' end. We give procedures below for constructing the designs. We use a concrete example to illustrate the concepts.

Let the loop sequence be TTGG, restriction site I be EcoRI (GAATTC) and restriction site II be SalI (GTCGAC). The U6 promoter sequence is TTGTGGAAAGGACGAAACACC and the U6 terminator is TTTTT. Let the accession of interest be NM_001033, whose current version number is 2. Then the following steps will lead to good shRNA designs,

1) Using the procedures given for siRNA design, we find that the 19 mer CAAAGTATGGTATAAGAAA is one of the good designs for NM_001033.2.
2) Extending it two base pairs in the 5' direction and one in the 3' direction gives us the 22 mer TGCAAAGTATGGTATAAGAAAC
3) Extending it seven base pairs along the 3' direction gives us the 29 mer TGCAAAGTATGGTATAAGAAACAGTTTAC.
4) Generate the anti-sense sequence GTAAACTGTTTCTTATACCATACTTTGCA.
5) Create hairpin by concatenating anti-sense loop and sense, in that order, GTAAACTGTTTCTTATACCATACTTTGCA<u>TTGG</u>TGCAAAGTATGGTATAAGAAACAGTTTAC.
6) Use a folding program like mfold [25] to confirm that the hairpin folds correctly. The result is shown in Figure 3.
7) Add the U6 promoter sequence to the 5' end of the hairpin <u>TTGTGGAAAGGACGAAACACC</u>GTAAACTGTTTCTTATACCATACTTTGCATTGGTGCAAAGTATGGTATAAGAAACAGTTTAC.
8) Add a SalI site to the 5' end of the hairpin <u>GTCGAC</u>TTGTGGAAAGGACGAAACACCGTAAACTGTTTCTTATACCATACTTTGCATTGGTGCAAAGTATGGTATAAGAAACAGTTTAC.
9) Add the U6 terminator sequence to 3' end of hairpin GTCGACTTGTGGAAAGGACGAAACACCGTAAACTGTTTCTTATACCATACTTTGCATTGGTGCAAAGTATGGTATAAGAAACAGTTTAC<u>TTTTT</u>.
10) Add an EcoRI site to the 3' end of the hairpin GTCGACTTGTGGAAAGGACGAAACACCGTAAACTGTTTCTTATACCATACTTTGCATTGGTGCAAAGTATGGTATAAGAAACAGTTTACTTTTT<u>GAATTC</u>
11) Make sure that only one copy of each restriction site and the U6 terminator occur in the sequence

12) Reverse Complement the whole construct, and send for DNA synthesis
GAATTCAAAAAGTAAACTGTTTCTTATACCATACTTTGCACCAATGCA
AAGTATGGTATAAGAAACAGTTTACGGTGTTTCGTCCTTTCCACAAGT
CGAC

The exact restriction site and terminator and promoter sequences depend upon the lab procedures involved in inserting these sequences into plasmids. The U6 promoter ensures that the gene is expressed and the terminator ensures that only the hairpin gets expressed, that is, there is no transcriptional run through. The barcode at the end is a random 60 mer that is unique to each hairpin allowing identification of the hairpin, either via microarrays or via the use of PCR.

Another method of constructing these is to use the context of a known miRNA. An miRNA with a target strand of length 22 is picked, and the target sequence is replaced with the anti-sense strand from the design above. The complementary strand is also replaced, taking care to preserve the bulges, loops and types of mismatches. This will probably ensure that the precursor sequence, generated by Drosha, is properly and efficiently processed by the Dicer and gets associated properly with the RISC complex.

miR-30 design protocols.

1. For given Accession, check whether there are 19-mers for that accession in the Pre-designed database. To check this, use Blast to compare given sequence with pre-designed blast database.If there are designs , then take them and store in csv format which will be used for shRNA design.
2. If there is no 19-mer for given sequence, then extract the 19-mer from the sequence based on the matrix scoring method described above. matrices on experimental derived 19-mers.
3. Design 6 such 19-mer for each given sequence and put them in csv format. (Format : Accession Start Rank comment 19-mer)
4. Above designed 19-mer is only good if it is unique for given gene. To check uniqueness, blast this 19-mer with non-redundant blast database described above. If we find more than 1 hit, then eliminate the 19-mer.
5. Put all 19-mers (pre-designed and newly designed) into one .csv file.
6. Extend each 19-mer to 22-mer by first finding 19-mer in sequence and taking 2 upstream bases and 1 downstream base of that 19-mer from the sequence.
7. This 22-mer is used to design shRNA, as described in the next few steps.
8. Sense = last 21 bases of 22-mer.
9. anti-sense = complement of Sense.
10. We need to check first base of 22-mer and add appropriate base in Sense and anti-sense seq.
    a) if first_base = A then sense = C + sense , anti-sense = T + anti-sense
    b) if first_base = G then sense = A + sense , anti-sense = C + anti-sense
    c) if first_base = T then sense = C + sense , anti-sense = A + anti-sense

d) if first_base = C then sense = A + sense , anti-sense = G + anti-sense

11. Now , top = GCG + sense-seq + T bottom = CGT + anti-sense + AT. Hairpin = prebind+ top + LOOP + reverse(bottom) + postbind where LOOP = "GTGAAGCCACAGATG", prebind ="TGCTGTTGACAGTGA postbind = "CTACTGCCTCGGA

The oligo is synthesized as a DNA, and then cloned into a plasmid using PCR, grown and then inserted into a retroviral vector. For transient effects, the plasmid can be transfected into the cell with the effects seen about 48 hours after transfection. For permanent insertion, the vector is used to infect the cell.

Building a Library of shRNAs.

Library construction requires making a collection of shRNAs that can be ordered off-the-shelf. The collection needs to be fairly complete, covering every major functional group and almost all known genes (at least the major variants of each gene type). In addition, the collection needs to contain constructs that can be trusted to work. This is required since it is impossible to have functional assays for each individual construct. Instead, the library is most likely to be used for silencing collections of genes in functional groups, such as kinases. Building the library involves the following steps,

1) Design at least 3 shRNAs per gene. All of the design principles for siRNAs apply. In addition, you can pick oligos that can work for both mouse and human genes (this has indeed been done for the library at CSHL), allowing use of common oligos in both species.
2) Barcodes (random 60 mer sequences synthesized in a lab) are inserted into the constructs.
3) Set up a lab tracking system (LIMS) and make the data accessible through a web-based access system. All the data can be put into a mysql database, and a web interface can be set up, using an Apache server, along with cgi scripts. The details can be found in many books, and we will not go into them here. This step is crucial as it allows web-based access to the data for wet-bench scientists.
4) The shRNAs must be verified by sequencing, before use. Use a mysql database to track the fate of the sequencing activities.
5) Build a BLAST database of all the target sequences in the library, so users can input names and get back oligos against their genes. This eliminates problems with annotations, errors in naming etc.
6) Keep track of the barcodes for each construct, as these are crucial for analysis of the assays.

A library has been constructed along these lines and will soon be available from the Hannon lab at CSHL[24].

Use of the Library.

The library allows large-scale studies that are prohibitively expensive with siRNAs. We can now perform genome-wide analysis in a way that was not feasible earlier. Steps involved in using an shRNA library.