The Stedman Guide to Designing and/or Selecting RNAi Targets, v1.0, 2008

RNAi targets can be chosen several ways, as detailed below. In general, we've found that, very roughly, about two-thirds of the RNAi targets chosen using options 1, 3, or 4 will lead to 70% or better silencing. As a result, it's generally advisable to choose several targets per gene of interest.

RNAi experiments should always include at least a scrambled RNAi control. You may use the commercially available control offered by the relevant vendor, but our experience is that some of these can affect phenotype. As such, a scrambled control (GAGACCCUAUCCGUGAUUA) was developed by the Stedman Center and has no effect on β-cell viability or insulin secretion. In vector form, the plasmid and adenoviral database designation for this scrambled control is FB102 and Ad-FB102, respectively. Regardless of which scrambled control is used, it'd be advisable to include an untransfected control at least once to monitor non-specific effects of the transfected duplexes.

1. Order "pre-selected" siRNA duplexes from Qiagen, ABI or another source. Several companies have developed their own proprietary target search algorithm and offer a number of "pre-selected" duplexes against fully annotated Genbank mRNAs. Note that "pre-selected" doesn't mean "pre-validated", and there's no guarantee any particular pre-selected duplexes will be effective. If you want that, order the more-expensive "pre-validated" duplexes.

If you order commercially available duplexes, you will only get the target sequences when you receive the duplexes. You may be able to transfect these duplexes and test their efficacy before designing and ordering the corresponding DNA oligos that are subcloned into the adenoviral shuttle vector.

2. Search the literature for published targets. It's possible someone already silenced the gene you're interested in. If so, the target sequence ought to be published.

3. Use on-line algorithms, such as that on the Applied Biosystems and IDT websites, to design siRNA targets.

http://www.ambion.com/techlib/misc/siRNA_finder.html

https://www.idtdna.com/Scitools/Applications/RNAi/RNAi.aspx

Note: Though rare, some otherwise effective RNAi targets may contain internal elements that make them unsuitable for a vector-based approach. As such, if one is to use a vector-based approach, all sequences should be checked for the presence of these elements before selecting them (see *Tom's Rules Summary* and *Eliminate targets that contain undesirable internal elements* below).

4. Design your own using Tom's Rules as a guide (Note from Tom: There are likely many mechanistic reasons why any particular siRNA or shRNA is or is not

effective, and few of these reasons have been elucidated. As such, “Tom’s Rules” are not intended to be comprehensive and immutable. There are likely many potentially effective targets within any mRNA sequence that may break one or more of these rules. I have found, however, that targets identified using these rules work well. Happy silencing! tb)

If taking this approach for a vector-based approach, once your RNAi targets are identified, you will go directly to ordering 2 complementary DNA oligos to clone into the viral shuttle vector, i.e. you do not need to order siRNA duplexes to pre-test, though one may do this if they prefer. Before making virus, however, you can test the RNAi shuttle clone for its ability to silence by transiently transfection.

Tom’s Rules Summary

1. **Generate a G+C plot for your target mRNA, and identify G+C Transition Areas.**

2. **Within the G+C Transition Areas, identify 19 to 23mers whose 5’ and 3’ ends are strong and weak, respectively. Mutate 3’ end to further de-stabilize if applicable.**

3. **Eliminate targets that contain undesirable internal elements, such as internal run of Gs or Cs. If planning a vector-based approach, also eliminate incompatible targets, such as those that contain, or end up encoding a pre-mature T5 termination signal.**

Explanations:

1: Choose a target that will be accessible by the RNAi Silencing Complex (RISC), so focus in on G+C Transition Areas: Obviously, the RISC has to have access to the target within the mRNA to be effective, so the mRNA 2° structure around the target should be relatively open.

Locating regions of low 2° structure: It’s generally accepted that, to a first degree, G+C content correlates with 2° structure, so a G+C plot of a target mRNA, created using DSGene or MacVector, offers a window into the general 2° structure of a transcript (see below and Figure 1).

During nucleic acid hybridization, “G”s hydrogen bond to “Cs relatively strongly, while “A”s hydrogen bond to “T”s (“U”s in RNA). Areas on the G+C content plot that display relatively high or low G+C content, are therefore likely to be domains of the mRNA target that are internally rich in 2° structure, so targeting these regions for RNAi isn’t generally preferred.

In contrast to high or low G+C regions, *transition* regions may be excellent places to identify effective RNAi targets. Transition regions are those regions on the G+C plot that connect high and low G+C domains. They are presumed

to contain less mRNA 2° structure since they merely connect a pair of domains that we already assumed to contain independent 2° structural domains.

A transition region can be identified as having one or more of the following characteristics:

1. It is a stretch of at least 40bp with a clear positive or negative slope.
2. It generally crosses the 50% G+C mark roughly halfway through its run.
3. It often connects a strong G+C peak to a strong G+C valley.

A number of, but not all, transition areas are shown in Figure 1 (arrows).

To create a G+C plot: With your sequence open in DSGene under Analyze choose nucleic Acid Analysis toolbox. Choose Base composition and add it to selected profiles. Click on Base composition within the selected profiles window and some choices will appear. Choose C+G and click OK. You will then get a results window with a plot of the G/C content.

2: The RNAi Silencing complex (RISC) must take up the correct strand, so choose targets with strong and weak 5' and 3' ends, respectively.
Although, free, bioactive siRNAs (purchased as siRNA duplexes, or expressed intracellularly off a vector) are short double-stranded RNAs (dsRNAs), *only 1 of the RNA strands is actually incorporated into the silencing complex*. Obviously, only 1 of these strands is complementary to the mRNA target, and should be the one incorporated, but how does the silencing complex (RISC) know which is the appropriate strand? The answer is that it doesn't – it simply takes any strand that's easy to take (see Figure 2). Taking this into consideration when selecting and designing RNAi targets will help the RISC incorporate only the correct strand. Otherwise, the silencing complex may sometimes or always take the wrong strand, in which case the net efficacy of your siRNA will be diluted or even blocked entirely.

To incorporate a single RNA strand from a dsRNA duplex, the RISC first recognizes the 3' overhang at either end of the duplex. It then "grabs" the 5' end of the opposite strand, and using its helicase activity, tries to separate and unwind the strands. If it's able to get the bases (particularly the first base) apart, at this end, it continues to separate, unwind and incorporate that particular strand by its 5' end. The key to having the correct strand incorporated is therefore to select targets that are G/C- and A/T-rich at their extreme 5' and 3' ends, respectively. In other words, look for N19-N23 stretches that begin with G or C and end in A or T (see Figure 2A).

To destabilize the 3' end of the duplex further and insure the correct strand gets incorporated, however, you may consider targets that end in C, in which case, the C is mutated to a U. Since Us and Gs do hybridize, but do so weakly, the 3' end of the duplex is de-stabilized and the correct non-mutated strand is easily incorporated into the RISC (see Figure 2B).

3: Eliminate targets that contain undesirable internal elements. When designing siRNA duplexes, undesirable elements include internal run of 4 or more Gs or Cs in the target sequence. When placed sequentially, 4 Gs can form a particularly stable cruciform structure, which will wreak havoc with the synthesis of the duplex, the ability of the 2 strands of the dsRNA to completely anneal to each other, and the ability of the bioactive RISC to completely anneal to the target. A consecutive run of 4 Cs is to be avoided, since the complementary strand of the duplex will then contain a run of 4 consecutive Gs.

If planning a vector-based approach, also eliminate targets that might be troublesome with this approach, such as those that contain, or end up encoding a pre-mature T5 termination signal. See special notes for vector-based approach.

4: On targeting the ORF or UTR: You might wonder whether it's more effective to target the Open Reading Frame (ORF) or one (5' or 3') of the Untranslated Regions (UTRs). Some prefer to target the ORF, if, for no other reason, because that's where most are already located. Others may target the 3' UTR, feeling that, as a regulatory element, it ought to be particularly sensitive when targeted. As of this writing, however, I'm not aware of any clear difference in inherent efficacy between targeting the ORF and 3'UTR, and we've seen effective and ineffective siRNAs that target either region.

On the other hand, there may be some very practical reasons to specifically target the ORF or the 3'UTR: First, within an mRNA transcript, ORFs tends to be relatively conserved between orthologues, since they encode the corresponding protein. As such, if you're interested in silencing a given gene from 2 different species, say rat and mouse, you're more likely to find conserved RNAi targets in the ORF. Conversely, sequence homology tends to degenerate in the 3' UTR sequences of orthologous mRNAs, so if you want a species-specific knockdown, consider RNAi targets in the 3'UTR.

Finally, targeting the 3'UTR allows researchers to perform rescue experiments much easier, i.e. restoring wild-type phenotype in silenced cells by co-expressing a heterologous copy of that same gene. Obviously, this experiment, considered the Gold Standard RNAi Experiment, is difficult to do when the heterologous copy of the gene contains the RNAi target. By targeting the 3'UTR of the host transcript, one can co-express a cloned copy that lacks this region.

5. Special considerations for vector-based shRNAs:

a) **Targets should start with G:** For efficient RNAi, you need sufficient quantities of the shRNA expressed in the cell. For this, cells must not only be able to be transfected or transduced efficiently, but the shRNA

must also be able to be transcribed efficiently. Vector based systems typically use the H1 or U6 RNA polymerase III promoter to drive expression of the shRNA, and these promoters tend to prefer a purine (G or A) at the transcription start site. The system used in our lab uses the H1 promoter and orients the first base of the target to be the first transcribed base, so the target should start with a G or an A (though we have seen effective silencing using targets starting with C). Given that the 5' end of the target should be G/C-rich (see Rule 2), I suggest selecting targets that begin with G.

b) **Targets should end with an A or T.** Targets may also end in a C, which may or may not have to be mutated to a T in the sense strand (See Rule 2). These requirements, however, may not be as strict with vector-based expression of shRNAs as they are for siRNA duplexes for the following reason: The RNA sequence of the hairpin loop begins with "UU" and ends with "GA". As such the actual stem of the expressed shRNA stem/loop itself may elongated by a U/A bond, and possibly even a very weak U/G bond. As a result, the processing enzyme Dicer may process the shRNA after the weak U/A or U/G bond. If so, mature siRNAs, by virtue of their precursor's loop, may always have a weak 3' end, insuring the correct strand is taken up.

c) **Targets should not end with "TTT"…**otherwise, the 3' T3 of the target, when cloned next to the T2 of the loop, will form a T5 sequence, which is the termination signal for polIII polymerase. In other words, you won't get any silencing because you won't get any shRNA expressed.

d) **Avoid internal runs of 4 or more As or Ts.** Again, T5 is an RNA polIII termination. A run of 4 or more "A"s in the sense strand will mean a run of 4 or more "T"s in the complement. Although a run of 5 "T"s is thought to be needed for termination, this requirement may not be absolute, so play it safe and avoid runs of 4 or more.

e) **Targets may end in "GAA", but**…be aware that in this case, when constructing the 64mer sense DNA oligo, the target's terminal "GAA" will be placed next to the loop sequence's initial "TTC", forming an E*co*RI site. Since miniprepped shRNA vector clones are screened by looking for a gel-shifted small E*co*RI/H*ind*III fragment, the presence of this internal E*co*RI site will result in a false negative, since it will prevent the band from shifting. Be aware of this possibility. If your target ends in "GAA", you may screen miniprep clones using E*co*RI only.

Figure 1. G+C content of rat Pdx-1 (NM_022852). Domains of high and low G+C content may, to a first approximation, contain their own 2° structure. In contrast, Transition Areas can be thought of as domain connectors, relatively free of strong 2° structure, and therefore make excellent places to begin identifying RNAi targets (see text). Some, but not all transition areas are indicated with a black arrow. Red Arrow shows position of Pdx-1 "E" target, which, when targeted by it's respective siRNA reduced Pdx-1 mRNA levels >90%, despite a relatively low overall G+C content.

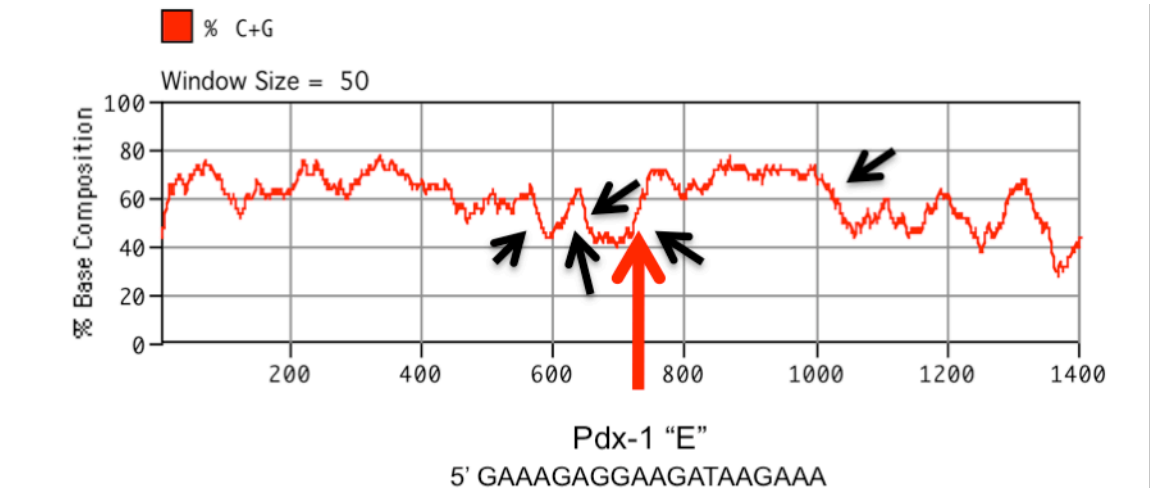

Figure 2. **Schematic representation of the 5' and 3' ends of an siRNA duplex.** (**A**) The RNAi silencing complex (RISC) recognizes the 3' di-nucleotide overhand on each end of a duplex, and attempts incorporate the complementary strand by complementary 5' end (red arrows). Since the underlined strand is identical to the target sequence, the bioactive strand is the complementary strand, so strong and weak 5' and 3' ends of the duplex, respectively, enable the RISC to incorporate only the correct strand. (**B**) A target that ends in C may be used effectively if the 3' end is de-stabilized after mutation of the terminal sense C to a U. With a strong 5' end, and a very weak 3' end, the RISC is able to incorporate only the correct strand.