SABiosciences

TECHNICAL ARTICLE

Did Your RNAi Experiment Work? Reliably Validating RNA Interference with qRT-PCR

Bill Wang, Song Tian, Qiong Zhou, and Xiao Zeng

SABiosciences 6951 Executive Way, Frederick, MD 21704 USA
Phone: +1 (301) 682-9200 Fax: +1 (301) 682-7300 Web: www.SABiosciences.com
Email: support@SABiosciences.com

Abstract: RNA interference (RNAi) technology for gene knockdown has become a powerful tool for gene function studies. As the popularity of RNAi technology grows, so does the frustration it causes for many researchers. How do you know with certainty if your RNAi experiment is working? Many prefer Western blots measuring the target protein directly, because its knockdown is the most relevant to the observable phenotype under study. However, in practice, the antibody to a given target protein may not always be readily available. Besides, the long half-lives of many proteins may mask the RNAi effect at the protein level, particularly in transient transfections. Therefore, a direct measurement at the mRNA level is always needed to more directly verify that RNA interference is decreasing the amount of the transcript. Real-time RT-PCR is the "gold" standard for measuring steady state mRNA levels. However, when used for RNAi validation, a few important factors need to be considered carefully, especially those contributing to data variation. This white paper addresses the sources of variation that must be dealt with when validating RNAi experiments.

Introduction

Although quantitative, real-time PCR (qPCR) is hailed for its broad detection dynamic range (up to 9-logs), only a very narrow range is needed for RNAi validation (usually less than one (1)-log for each target gene). As a matter of fact, the exponential nature of its readout makes the detection of small changes in mRNA levels technically challenging. For example, a two-fold change in mRNA level (i.e., a 50% knock-down by RNAi) translates into a $\Delta\Delta C_t$ value of only one (1). Real-time PCR experts often cite that the systematic error of a PCR instrument could be as high as 0.5 cycles between wells. Therefore, if extra care is not taken during reaction setup, an observed $\Delta\Delta C_t$ value of one cycle will not truly indicate a knock-down of 50 percent. The following discussion provides a detailed analysis of all sources of variance that may lead to the erroneous interpretation of RNAi validation.

We also provide a real example demonstrating how to keep track of those variances in order to estimate confidence intervals (C. I.) at the 95% level. Using this theoretical error model and our standardized experimental protocol, we have tested 329 SureSilencing™ pre-designed shRNA Plasmids. Among those 329 designs, 221 (67.2%) sequences have achieved successful knockdown. This success rate is twice that reported by The RNAi Consortium, a resource more commonly used by researchers and other commercial manufacturers of shRNA alike. Based on this empirically determined success rate, we will also show how we can guarantee that our SureSilencing shRNA Plasmid design algorithm delivers at least two successfully effective sequences per gene.

Sources of Variance in RNAi Validation

1. **Transfection Efficiency:** This factor is not directly related to real-time PCR detection, but we will demonstrate that without consistently achieving an 80% or higher transfection efficiency (TE) of the specific cell line under study, the RNAi effect measured by real-time RT-PCR is not reliable.
2. **PCR Reproducibility:** Most commercial real-time PCR instruments are stable enough to give reasonable reproducibility. However, some technical skills are required to lower the differences between C_t values across technical real-time PCR replicates. We will recommend that the standard deviation in threshold cycle determinations (SD_{PCR} or SD C_t) should be no higher than 0.20.
3. **Biological Sample Consistency:** Biological replicates are absolutely required for the validation process. Standard error decreases with the number of knockdown determinations performed. We will illustrate that the minimum number of required biological replicates (N) is 3.
4. **PCR Amplification Efficiency:** Because we use the $\Delta\Delta C_t$ method for relative quantification of mRNAs, 90 % or higher PCR amplification efficiencies for both the gene of interest and the housekeeping (or reference) gene amplicons are required. Although it is beyond the scope of this discussion, the use in this study of RT² qPCR Primer Assays designed by SABiosciences achieves this goal.

Discussion of Knockdown Efficiency and the Theoretical Error Model

The first set of equations (presented at right) tells us that in order to calculate knockdown efficiency (KD), we need to collect all the C_t values (to use the $\Delta\Delta C_t$ method), as well as the transfection efficiency (TE). Through a series of calculations and estimations of the standard deviations (SD), the second set of equations leads us to the standard error (SE) of the mean $\Delta\Delta C_t$ value. Using the SE, we can attach the 95% confidence interval (95% C.I., with z score = 1.96) to any measured $\Delta\Delta C_t$ value. A narrower 95% C.I. indicates a more reliable data set generated in the validation process. By converting the extremes to KD values using Equation 1, the 95% C.I. establishes a range centered on an "observed" or measured KD value, which we will call the "implicated" KD. Using the relationship between the observed KD and the implicated KD, we can build a model and test the contribution that variation makes to the implicated KD values. As demonstrated in the "thought experiments" of Figure 1 and summarized in Table 1, this theoretical error model allows us to establish minimal criteria for parameters of the validation process to assure that the observed KD is distinguishable from the implicated KD of a less efficient knockdown.

In Figure 1, Panel A starts with an experiment having the following "measured" variables:
1. Transfection Efficiency (TE) = 80%
2. Standard Deviation of PCR C_t Values (SD_{PCR}) = 0.2
3. Number of Biological Replicates (N) = 3

Using our two sets of equations, we can assign a range (95% C.I.) as the implicated KD for any given observed KD. For example, the observed 70% KD in Panel A has a 95% chance of representing a 50% to 84% implicated KD value. Under these conditions, we can confidently claim that a given shRNA with a 70% observed KD is statistically different from one with a 24% observed KD, because their implicated KD ranges are not likely to overlap ($p < 0.05$, student t-test). We can then determine how changes to all three of our "measured" variables (TE, SD_{PCR} and N) affect the implicated KD of an observed KD. As shown in Panel B, if the TE is reduced to 60%, the confidence of our 70% observed KD drops tremendously. The implicated KD range of an observed 70% KD has expanded to between 44% and 96%. The validity of this 70% observed KD is now much in doubt, because it can barely distinguish itself from a shRNA with a 1% observed KD. Similarly, if we were less certain about PCR consistency (by setting SD_{PCR} = 0.3 in Panel C), or we were to reduce the biological replicates (by setting N = 1 in Panel D), the observed 70% KD would become

Knockdown Efficiency Calculation and Theoretical Error Model

The following two sets of equations were essential in building our theoretical error model to estimate a 95% confidence interval (C.I.) for the final knockdown efficiency (KD).

First Set of Equations

$$KD = \frac{1-2^{\Delta\Delta C_t}}{TE}$$

$$\Delta\Delta C_t = \Delta C_t^{T} - \Delta C_t^{NC}$$

$$\Delta C_t^{T} = GOI\ C_t^{T} - HKG\ C_t^{T}$$

$$\Delta C_t^{NC} = GOI\ C_t^{NC} - HKG\ C_t^{NC}$$

KD, Knockdown Efficiency

TE, Transfection Efficiency

$\Delta\mathbf{C_t^T}$, the difference in C_t values between the gene of interest (**GOI C_t**) and the housekeeping (or reference) gene (**HKG C_t**) in the Target (T) gene shRNA transfection

$\Delta\mathbf{C_t^{NC}}$, the difference in C_t values between the gene of interest (**GOI C_t**) and the housekeeping (or reference) gene (**HKG C_t**) in the Negative Control (NC) shRNA transfection

Second Set of Equations

$$SD_{PCR} = GOI\ SD_{PCR} = HKG\ SD_{PCR}$$

$$SD\ \Delta C_t = \sqrt{(GOI\ SD_{PCR})^2 + (HKG\ SD_{PCR})^2} = \sqrt{2 \times (SD_{PCR})^2} = \sqrt{2} \times SD_{PCR}$$

$$SD\ \Delta C_t^{BIO} = 1.1 \times SD_{PCR}$$

$$Mean\ SD\ \Delta C_t = \sqrt{(SD\ \Delta C_t^{BIO})^2 + \frac{(SD\ \Delta C_t)^2}{N}}$$

$$SD\ \Delta\Delta C_t = \sqrt{2 \times (Mean\ SD\ \Delta C_t)^2} = \sqrt{2} \times Mean\ SD\ \Delta C_t$$

$$SE = \frac{SD\ \Delta\Delta C_t}{\sqrt{N}}$$

$$C.I. = Mean\ \Delta\Delta C_t \pm (z \times SE)$$

SD_{PCR}, Standard Deviation (SD) in C_t values
(Indicates technical PCR replicate consistency.)
We assume that the SD in the gene of interest C_t values (GOI SD_{PCR}) is equal to that of the housekeeping (or reference) gene C_t values (HKG SD_{PCR}).

SD $\Delta\mathbf{C_t}$, SD in ΔC_t values between technical PCR replicates

SD $\Delta\mathbf{C_t^{BIO}}$, SD in ΔC_t values between biological replicates
We assume that SD ΔC_t^{BIO} is slightly greater than the SD among technical PCR replicates (SD_{PCR}).

SD $\Delta\Delta\mathbf{C_t}$, SD in $\Delta\Delta C_t$ values (composite indicator of all variances in the experiment)

N, Number of biological replicates in the experiment

SE, Standard Error in $\Delta\Delta C_t$

C.I., Confidence Interval for $\Delta\Delta C_t$; **z**, z score

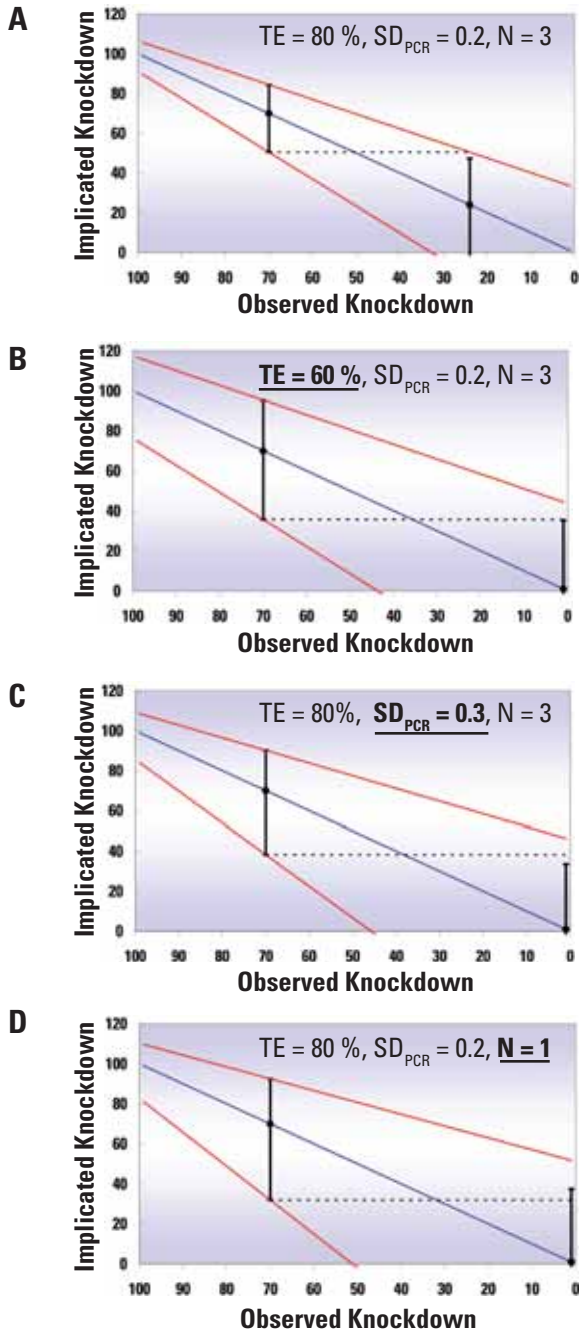

Figure 1: Low Transfection Efficiencies, High qPCR Variability, and Too Few Biological Replicates Significantly Weaken the Reliability of the Knockdown Measurement.

The variance in RNAi knockdown measurements are estimated based on Equations 1 and 2. The middle diagonal lines represent "observed" knockdown (KD) values. The lower and upper outer diagonal lines represent the lower and upper limits of the 95% C.I. for the observed KD defining the "implicated" KD. The error bars and dotted lines denote how two KD measurements can be seen as different with reasonable and statistically significant confidence ($p < 0.05$). TE, transfection efficiency; SD_{PCR}, standard deviation of C_t values from technical PCR replicates; N, number of RNAi transfection biological replicates.

Based on this analysis (summarized in Table 1), we conclude that the confidence in determining shRNA knockdown efficiency is very sensitive to three key sources of variances: 1) the transfection efficiency (TE) of the cell line of interest, 2) the PCR consistency (SD_{PCR}), and 3) the number of biological replicates (N).

Table 1: RNAi Knockdown Measurements are Sensitive to Variance in Several Parameters

The implicated KD for a 70 % observed KD obtained under different theoretical experimental conditions was calculated using the error model described in the text. The "Maximum Distinguishable Implicated KD" refers to highest observed KD with an implicated KD that can be distinguished from the "implicated KD" of an observed 70 % KD with a high degree of confidence (95 % interval).

TE	SD_{PCR}	N	Implicated KD	Maximum Distinguishable Implicated KD
80 %	0.20	3	70 (50, 84)	24 (-12, 50)
60 %	0.20	3	70 (36, 95)	1 (-56, 44)
80 %	**0.3**	3	70 (38, 90)	1 (-71, 46)
80 %	0.20	**1**	70 (32, 92)	1 (-85, 52)

The Total Variance in RNAi Validation, a Real RNAi Case Study

The following real case study demonstrates how the error model can be applied by validating a STAT3 RNAi design from one of SABiosciences' pre-designed SureSilencing™ shRNA Plasmids.

Materials and Methods

Cell Culture and shRNA Delivery: 293H cells (Invitrogen) were cultured in D-MEM supplemented with 10% FBS and 1X non-essential amino acids (Invitrogen) for no more than 15 passages. Genome-wide, gene-specific shRNA targeting sequences were designed using a proprietary algorithm and cloned into the pGeneClip™ hMGFP vector (Promega) to generate SureSilencing™ shRNA plasmids. Transfection grade SureSilencing™ plasmid (0.8 mg) mixed with Lipofectamine 2000 reagent (Invitrogen, 3 mL) was delivered to 80,000 cells in a 24-well plate format. Culture media were changed 24 hours after transfection. Transfection efficiency was estimated by following the expression of GFP using fluorescence microscopy. After 48 hours, total RNA was extracted using the ArrayGrade™ Total RNA Isolation Kit with gDNA cleanup by TURBO DNase™ (Ambion).

Real-Time RT-PCR: cDNA was synthesized from total RNA using the ReactionReady™ First Strand cDNA Synthesis Kit. Real-time PCR was performed using RT^2 SYBR Green qPCR Master Mixes on the Bio-Rad iCycler real-time PCR system or the Stratagene Mx3000 real-time PCR system. β-Actin was chosen as the housekeeping gene for normalization. Threshold cycle values (C_t) were collected and used for "$\Delta\Delta C_t$" analysis. Gene knockdown efficiency was calculated in the multi-step process described in the text.

The Total Variance in RNAi Validation, Case Study Results

Table 2 lists all of the collected C_t values. In this experiment, STAT3 is the target gene of interest (GOI), while β-actin (ACTB) is the housekeeping gene (HKG). Biological triplicate (N = 3, in the rows labeled 1, 2 and 3) test (T) RNA samples were collected, each from an independent transfection of cells with the STAT3 shRNA. Biological triplicate negative control (NC) RNA samples were also collected from independent transfections of cells with a shRNA having a random sequence and no homology to the mammalian transcriptome. For each of the six biological samples, three RT-PCR analyses were performed for both STAT3 and ACTB side-by-side (technical replicates, in the columns labeled 1, 2, and 3) for a total of 36 reactions and C_t values (3 biological replicates x 3 technical replicates x 2 genes x 2 shRNA = 36). The resulting C_t values are recorded in the appropriate cells of the table according to transfected shRNA, gene tested, biological replicate, and technical replicate.

Table 2: Raw Data for STAT3 Knockdown

	shRNA for STAT3				Negative Control shRNA			
	N	C_t^T1	C_t^T2	C_t^T3	N	C_t^T1	C_t^T2	C_t^T3
STAT3	1	27.00	26.70	27.20	1	24.40	24.60	24.70
	2	27.00	27.00	27.00	2	24.30	24.30	24.30
	3	26.70	26.70	26.90	3	25.10	25.20	25.30
ACTB	1	17.80	17.80	17.90	1	17.90	18.20	18.10
	2	17.90	18.20	17.70	2	17.60	18.10	17.80
	3	17.90	18.00	17.60	3	18.10	18.10	18.20

Table 3: ΔC_t Calculations

shRNA for STAT3

	N	Mean GOI C_t^T	GOI SD_{PCR}^T	Mean ΔC_t^T	SD ΔC_t^T
STAT3	1	26.97	0.25	9.13	0.26
	2	27.00	0.00	9.07	0.25
	3	26.77	0.12	8.93	0.24

	N	Mean HKG C_t^T	HKG SD_{PCR}^T
ACTB	1	17.83	0.06
	2	17.93	0.25
	3	17.83	0.21

Negative Control shRNA

	N	Mean GOI C_t^{NC}	GOI SD_{PCR}^{NC}	Mean ΔC_t^{NC}	SD ΔC_t^{NC}
STAT3	1	24.57	0.15	6.50	0.22
	2	24.30	0.00	6.47	0.25
	3	25.2	0.10	7.07	0.12

	N	Mean HKG C_t^{NC}	HKG SD_{PCR}^{NC}
ACTB	1	18.07	0.15
	2	17.83	0.25
	3	18.13	0.06

Table 4: Mean ΔC_t Calculations

shRNA for STAT3			Negative Control shRNA		
Mean ΔC_t^T	SD $\Delta C_t^{BIO}T$	Mean SD ΔC_t^T	Mean ΔC_t^{NC}	SD $\Delta C_t^{BIO}NC$	Mean SD ΔC_t^{NC}
9.04	0.10	0.18	6.68	0.34	0.36

Finally, using the data in Table 4, we calculate the Mean $\Delta\Delta C_t$, the difference in the test and control samples' Mean ΔC_t values. The final composite SD $\Delta\Delta C_t$ is used to define the Standard Error (SE) and the 95% Confidence Interval (95% C.I., z = 1.96) for the $\Delta\Delta C_t$ value (Table 5). Because we determined the transfection efficiency (TE) to be 90%, we calculate our observed knockdown efficiency (KD) to be 89.5%. By converting the extreme $\Delta\Delta C_t$ 95% C.I. values into KD and normalizing for the TE in the same way, we observe the 95% Confidence Interval for the KD value to be between 81.7% and 95.7%, defining the implicated KD for this shRNA design. Therefore, this real life example successfully and reliably identifies the tested STAT3 shRNA as a design that has a 95% chance of knocking down this gene's expression by at least 81.7% and as much as 95.7%. This demonstration also illustrates more directly how the validation of any shRNA (or siRNA or RNAi) sequence can be performed with confidence using real-time RT-PCR.

Table 5: $\Delta\Delta C_t$ and KD Calculations

Mean $\Delta\Delta C_t$	SD $\Delta\Delta C_t$	SE	$\Delta\Delta C_t$ 95% CI
2.36	0.39	0.23	(1.92, 2.80)

TE	KD	KD 95% CI
90 %	89.5 %	(81.7 %, 95.7 %)

Batch Validation of SABiosciences' SureSilencing™ Pre-Designed shRNA Plasmids

Using the protocol and calculations demonstrated above for STAT3 shRNA, we have tested a total of 329 shRNA designs targeting 86 different genes (an average of roughly four designs per gene) generated by the algorithm[1] used for SABiosciences' SureSilencing™ shRNA Plasmids. To help interpret the knockdown and confidence interval results, we adopted the following empirical standards to categorize those 329 shRNA designs.

- Successful Design: KD ≥ 70.0 %, and lower 95 % C.I. extreme ≥ 55.5%
- Failed Design: KD < 33.3 %, higher 95 % C.I. extreme < 55.5%

Our results, summarized in Table 6 and Figure 2, indicate that 221 designs belong to the Successful Group, yielding a 67.2% success rate. In comparison, we notice that the shRNA library in The RNAi Consortium (TRC) collection, used by other commercial manufacturers of plasmid-based shRNA, has only a 31 ~ 38% success rate reported using a similar definition for a "successful" design[2,3]. Two out of four of our designs were Successful for 74 of the 86 genes tested, yielding an 86.0 percent success rate on a per gene basis. A binomial distribution method estimates that the algorithm should yield an 89.3 percent per gene success rate when applied to the design of shRNA sequences for every gene annotated by the NCBI in the human, mouse, and rat genome. Therefore, the delivery of two successful SureSilencing™ shRNA Plasmids for any human, mouse, or rat gene is an enforceable guarantee.

Table 6: SABiosciences' shRNA Design Algorithm Relative Sequence and Gene Success Rates

	Designs Tested	Successful Designs	Success Rate
SABiosciences Training Set	329	221	67.2
The RNAi Consortium	256[2], 54[3]	98[2],17[3] (calculated from reported numbers)	38[2], 31[3]

	Genes Tested	Successful Genes	Success Rate
SABiosciences Training Set	86	74	86.0
Genome-Wide Prediction	N/A	N/A	89.3

Figure 2: Validation of Pre-Designed SureSilencing™ shRNA Plasmids. The figure depicts the results from testing, using the protocol described in the text, a subset of 174 shRNA designs targeting 47 different genes from the complete training set used in this study. The observed knockdown rates (KD) for Successful (top trace) and Failed (bottom trace) shRNA designs were calculated and plotted by rank in ascending order. The Y-axis error bar demarcates the upper and lower extremes of the defined by 95% C.I. for each KD. The darker lines indicate the KD cutoff for our standards of successful (70.0%) or failed (33.3 %) designs. The dashed line indicates the minimum for the lower 95% C.I. extreme cutoff and the maximum for the higher 95% C.I. extreme cutoff for our standards of a Successful or Failed design, respectively (55.5%).

Summary

We have examined all major contributing factors essential to RNAi validation using real-time RT-PCR. We have concluded that transfection efficiency (TE), technical consistency (SD_{PCR}) and the number of biological replicated (N) are the three major key indicators to monitor in order to establish a reliable RNAi validation protocol. Transfection efficiencies should be consistently 80 percent or greater. The standard deviation in the technical replicate raw C_t values from the real-time PCR analyses should no greater than 0.2. In order to achieve the required biological consistency, no fewer than three biological replicates of the target gene-specific and the negative control shRNA transfections are required. We have also used this error model to assess our own SureSilencing™ shRNA design algorithm and proved that it out-performs the version used for the current TRC collections and other commercial shRNA plasmids.

References

1. Zhou H, Zeng X, Wang Y and Seyfarth BR. A Three-Phase Algorithm for Computer Aided siRNA Design. Informatica. 2006 30:357-364.
2. Root DE, Hacohen N, Hahn WC, Lander ES, Sabatini DM. "Genome-scale loss-of-function screening with a lentiviral RNAi library." Nat Methods. 2006 Sep;3(9):715-9.
3. Moffat J, Grueneberg DA, Yang X, et. al. "A lentiviral RNAi library for human and mouse genes applied to an arrayed viral high-content screen. Cell. 2006 Mar 24;124(6):1283-98.

SureSilencing™ shRNA Plasmids

Guaranteed Knock Down for Every Human, Mouse, and Rat Gene

SureSilencing™ shRNA Plasmids specifically knock down the expression of every human, mouse, or rat gene by RNA interference. For each gene, we provide four separate short hairpin RNA (shRNA) designs packaged in a plasmid backbone. Our experimentally verified shRNA design algorithm ensures the maximum gene-specificity and RNAi efficacy. We guarantee that you will see 70% or greater knock down for your targeted gene with at least two of the four shRNA plasmids.* The availability of two effective sequences allows you to properly control for non-specific and off-target effects.

Unlike chemically synthesized siRNA, plasmid-based shRNA provides a renewable source of RNA interference. You can amplify enough plasmid to complete any number of projects in your lab. Purchase of SureSilencing™ shRNA plasmids includes complete shRNA sequence information.

- **Enrich or Select:**
  Use Neomycin or Puromycin Resistance for stable selection and study the long term effects of gene suppression.
  Use GFP Marker for FACS enrichment and the study of short term effects of gene suppression.
- **Guaranteed!*:**
  Knock down expression of the target gene by at least 70 percent. Control for non-specific and off target effects.
- **Convenient and Cost-Effective:**
  Use standard plasmid-based and lipid-mediated transfection methods. Plasmids provide a renewable source of RNA Interference.

* **The SureSilencing™ shRNA Guarantee:** At least two of the provided pre-designed SureSilencing™ shRNA Plasmids are guaranteed to knock down expression of the targeted gene at the RNA level by at least 70 percent as measured by real-time qRT-PCR in transfected cells upon FACS-based enrichment for GFP expression or selection for neomycin or puromycin resistance as described in the User Manual.

shRNA Feature Summary

Tested	Experimentally verified design algorithm ensures efficacy.
Specificity	Minimizes off-target effect with the Smith-Waterman specificity search.
Plasmid Backbone	U1 promoter transcribes moderate shRNA amount: minimal off-target effect & toxicity Bacterial origin of replication and ampicillin-resistance marker: amplify unlimited supply Mammalian Expression Markers: **Neomycin** or **Puromycin** Resistance: select stably transfected cells, study long term effects **GFP Marker**: enrich or track transiently transfected cells, study short term effects

shRNA Plasmid Ordering Guide

Find the shRNA for your genes at: www.SABiosciences.com/RNAisearch.php

Each SureSilencing™ shRNA Plasmid Set comes with four shRNA designs targeting your gene of interest along with a negative control

Cat. No.	Description	Price
KX#####N	SureSilencing™ shRNA Plasmid Set with Neomycin Resistance	$ 799
KX#####P	SureSilencing™ shRNA Plasmid Set with Puromycin Resistance	$ 799
KX#####G	SureSilencing™ shRNA Plasmid Set with GFP Marker	$ 799

X = H for Human, M for Mouse, or R for Rat

SABiosciences™