



Strand Analysis, a free online program for the computational identification of the best RNA interference (RNAi) targets based on Gibbs free energy

Tiago Campos Pereira¹, Vinícius D'Ávila Pascoal Bittencourt¹, Rodrigo Secolin¹,
Cristiane de Souza Rocha¹, Ivan de Godoy Maia² and Iscia Lopes-Cendes¹

¹*Departamento de Genética Médica, Faculdade de Ciências Médicas,
Universidade Estadual de Campinas, Campinas, SP, Brazil.*

²*Departamento de Genética, Instituto de Biologia, Universidade Estadual Paulista, Botucatu, SP, Brazil.*

Abstract

The RNA interference (RNAi) technique is a recent technology that uses double-stranded RNA molecules to promote potent and specific gene silencing. The application of this technique to molecular biology has increased considerably, from gene function identification to disease treatment. However, not all small interfering RNAs (siRNAs) are equally efficient, making target selection an essential procedure. Here we present Strand Analysis (SA), a free online software tool able to identify and classify the best RNAi targets based on Gibbs free energy (ΔG). Furthermore, particular features of the software, such as the free energy landscape and ΔG gradient, may be used to shed light on RNA-induced silencing complex (RISC) activity and RNAi mechanisms, which makes the SA software a distinct and innovative tool.

Key words: RNAi, siRNA, siRNA design, software.

Received: October 23, 2006; Accepted: January 31, 2007.

The RNA interference (RNAi) gene silencing technique is a recently developed technology that allows potent and specific gene silencing through the use of double-stranded RNA molecules (dsRNAs; Fire *et al.*, 1998). The RNAi technique is widely used for identification of gene function (reverse genetics), functional genomics (Fraser *et al.*, 2000), to combat pathogens (Gitlin *et al.*, 2002; Mohammed *et al.*, 2003), as a therapeutic tool in cancer (Brummelkamp *et al.*, 2002) and some specific genetic disorders (Xia *et al.*, 2004), in the generation of biotechnological products (Ogita *et al.*, 2003) and for the construction of model animals (Fedoriw *et al.*, 2004).

In mammals, however, dsRNAs trigger antiviral responses and cell death ensues, so in these models small interfering RNAs (siRNAs) are the molecules of choice for RNAi studies because they are too small to trigger such responses. Molecules of siRNA possess a well defined structure *i.e.* a 21-mer duplex, two-nucleotide 3' overhang and a 5' phosphate. It is interesting to note that siRNAs directed to different regions of a specific transcript display widely different silencing efficiencies (Holen *et al.*, 2002), possibly in part due to the fact that, intracellularly, siRNAs are incorporated into an RNA-induced silencing complex (RISC) containing

slicer endonuclease activity. Slicer cleaves one strand of siRNA while keeping the other strand (the guide strand) to direct target RNA cleavage (supplementary data, Figure S1: Rand *et al.*, 2005). If the antisense strand remains in the RISC, efficient silencing occurs but if the sense strand remains in the RISC silencing is reduced or even compromised (Khvorova *et al.*, 2003; Schwarz *et al.*, 2003). Two independent research groups (Khvorova *et al.*, 2003; Schwarz *et al.*, 2003) have shown that the thermodynamic features of the siRNA termini, defined in terms of Gibbs free energy (ΔG , kcal mol⁻¹), determines the guide strand choice. Thus, Tuschl's rules, the well known protocol for siRNAs design (Elbashir *et al.*, 2001) now including ΔG via computational and systematic calculations, would reduce the time and costs involved in RNAi experiments.

In this paper we present Strand Analysis (SA), a free online program (see internet resources section) for the identification of the best RNAi targets based on thermodynamic features (Khvorova *et al.*, 2003, Table 1). The SA program computes ΔG in kcal mol⁻¹, the higher the ΔG value then the more preferentially will the antisense strand be kept within the RISC slicer domain thus resulting in better efficiency. As shown in Figure 1, the SA program has two different entry modes "Oligo Analysis" (OA mode) and "Sequence analysis" (SA mode). The OA mode can be used to compute single pre-selected 23-mer targets derived from messenger RNA (DNA or RNA

Table 1 - Standard Gibbs energy of activation ΔG values (kcal mol^{-1}) used for calculating the internal stability of RNA duplexes. The equation used was $\Delta G = S_{\text{Gas}} - S_{\text{Gs}}$, where $S_{\text{Gas}} = \text{Sum of the } \Delta G \text{ values for the first four nucleotides in the anti-sense 5' region}$, $S_{\text{Gs}} = \text{Sum of the } \Delta G \text{ values for the first four nucleotides in the sense 5' region}$. Note that $\Delta G \leq 0$ for the non-functional strand and $\Delta G > 0$ for the functional strand. Table modified from Khvorova *et al.*, 2003).

First nucleotide base pair	Second nucleotide ΔG values (kcal mol^{-1})			
	A	C	G	U
A-U	1.1	2.4	1.9	1.1
C-G	2.2	3.3	2.2	1.9
G-C	2.7	3.8	3.3	2.4
G-U	1.5	2.7	2.2	1.4
U-A	1.4	2.6	2.2	1.1

format) and presents the results as ΔG values, with positive ΔG values for the “active guide strand” and null or negative ΔG values for the “non-active guide strand”. The SA mode scans all the query sequence and calculates the ΔG values for all the 23-mer targets along the sequence to produce a list of ΔG values which may be set as a function of target position along the transcript or as the best values in decreasing order. Alternative outputs are the identification of only active or non-active strands. The SA ΔG gradient varies from $+9.3 \text{ kcal mol}^{-1}$ to $-9.3 \text{ kcal mol}^{-1}$ and may be used for special purposes in molecular analysis, as for example when a haploinsufficiency (50% silencing) would be more interesting than a knockdown (99.9% silencing). More exact molecular analysis may now be possible using this gradient principle, uncovering new phenotypes resulting from partial silencing. The correlation between ΔG values and silencing efficiency has been well-characterized (Khvorova *et al.*, 2003; Schwarz *et al.*, 2003) and was reproduced in our laboratory during the experimental validation of the SA program (supplementary data, Figure S2).

When working with H1/U6-based vectors for the production of short-hairpins it is important to avoid four thymines (Ts) or adenines (As) in a row and, likewise, four guanines (Gs) or cytosines (Cs) in a row should also be avoided when chemical synthesis is the choice. The SA program takes these factors into consideration and presents a warning message when such motifs are found during analyses.

The input file for the SA program is a .txt file, with the first line format as “> name of the gene” and its coding sequence (CDS, with or without numbers or spaces, DNA or RNA sequence) in the lines below. The .txt format output file is automatically generated in the same folder as the input file and presents i) position of the first siRNA nucleotide along the input sequence, ii) the ΔG value, iii) the siRNA structure (anti-parallel misaligned duplex for didactic visualization) and iv) the resulting siRNA oligos both on 5'-3' orientation (for ordering).

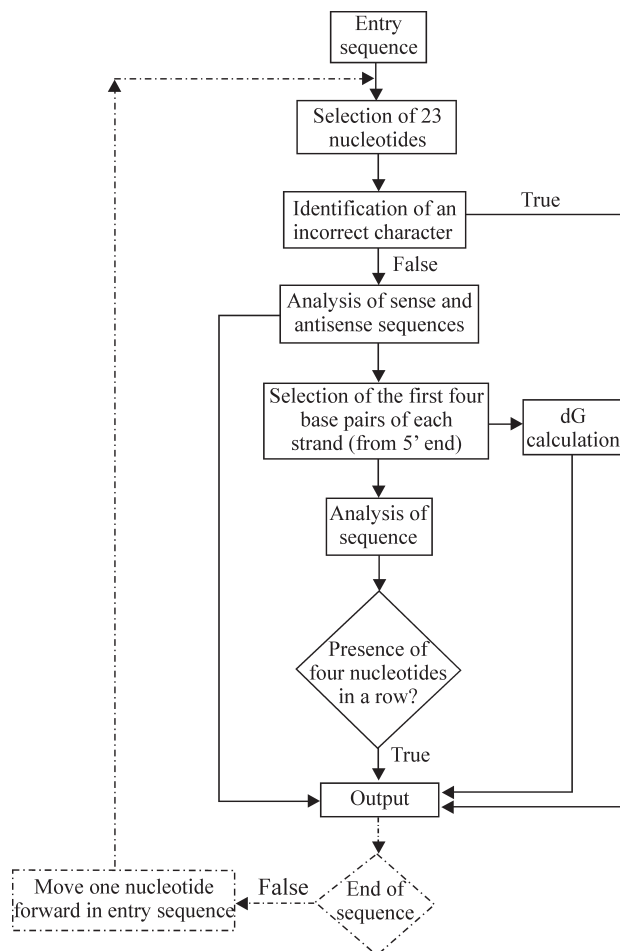


Figure 1 - Strand Analysis (SA) flowchart. The SA software has two different modes: ‘oligo analysis’ for 23-mer pre-selected targets (continuous lines only) and ‘sequence analysis’ for complete transcript scanning (all lines).

The SA program calculates the ΔG values of specific 23-mer targets or performs a complete scanning of the query sequence, listing the best targets by position or by the best ΔG values. Given that optimal siRNAs are selected, we believe that the SA program will improve knockdown efficiencies in RNAi experiments. When using RNAi to combat viral replication for example, targeted genomes may extend to tens of kilobases. The SA program can scan such large sequences presenting the few excellent targets (ΔG value greater than 6.0), which would not be identified by random choice. For example, a SA scan of the HIV genome (Genbank AF033819) indicated that the best target is located in position 1940, within the “pol” gene ($\Delta G = 8.5$). Furthermore, the ΔG gradient may also be used to shed some light on RISC activity and the mechanism of RNAi.

The SA program also displays a ΔG -based landscape along the gene sequence in a graphic format, which facilitates visualization of gene (or genomic) “ ΔG hotspots” where many siRNAs may be used (supplementary data, Figure S3). Furthermore, since RNAi acts as an antiviral

system, such landscapes may provide insight into changes in viral genomes and adaptations which occur over time under such pressure, aspects which are currently under investigation in our laboratory.

The SA program was implemented on the Linux platform, is web based and written in the Perl programming language, which is widely used in bioinformatics. With a small source code of only 7.9 kb the SA program shows good performance, taking only 2.3 s to run a sequence of 20,000 bases, and can be used along with other bioinformatics tools developed in our laboratory. The SA program is freely available, but is not open source.

It is important to note that the SA program must be used in combination with other computational tools for the design of siRNAs (Tuschl's rules) and not alone. For example, it is important to exclude 23-mer targets with strong secondary structures, a task that may be performed using Gene Runner (see internet resources section). Strand Analysis has already been registered at the Brazilian Patent Office (Instituto Nacional de Propriedade Industrial, INPI) under number 00068371.

Although there are other web-based programs used for siRNA design (Pei *et al.*, 2006), some of them are very slow, not user friendly or do not even consider thermodynamic features in their calculations. Those which do include thermodynamic parameters compute them along with many other factors, generating a ranking that is not a function of ΔG alone, thus making selection based on free energy difficult. Our Strand Analysis (SA) program distinguishes itself from its counterparts by providing the following advantages: i) the results are displayed in RNA format for both strands in the 5' to 3' orientation; ii) the ability to view positive or negative values alone or altogether; iii) the fact that the list of standard Gibbs energy values (ΔG) result may be set as a function of target position along the transcript or as ΔG values; and iv) the ΔG landscape may be analyzed along the gene sequence, thus providing a distinct and innovative tool.

Acknowledgments

This work was supported by the Brazilian agency Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), process number 02/01828-8-TCP and 03/10900-7). I.G.M. and I.L.C. are recipients of research fellowships from the Brazilian agency Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Brazil.

The authors declare no competing interests.

References

Brummelkamp TR, Bernards R and Agami R (2002) Stable suppression of tumorigenicity by virus-mediated RNA interference. *Cancer Cell* 2:243-247.

- Elbashir SM, Harborth J, Lendeckel W, Yalcin A, Weber K and Tuschl T (2001) Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature* 411:494-498.
- Fedoriw AM, Stein P, Svoboda P, Schultz RM and Bartolomei MS (2004) Transgenic RNAi reveals essential function for CTCF in H19 gene imprinting. *Science* 303:238-240.
- Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE and Mello CC (1998) Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* 391:806-811.
- Fraser AG, Kamath RS, Zipperlen P, Martinez-Campos M, Sohrmann M and Ahringer J (2000) Functional genomic analysis of *C. elegans* chromosome I by systematic RNA interference. *Nature* 408:325-330.
- Gitlin L, Karelsky S and Andino R (2002) Short interfering RNA confers intracellular antiviral immunity in human cells. *Nature* 418:430-434.
- Holen T, Amarzguoui M, Wiiger MT, Babaie E and Prydz H (2002) Positional effects of short interfering RNAs targeting the human coagulation trigger Tissue Factor. *Nucleic Acids Res* 30:1757-1766.
- Khvorova A, Reynolds A and Jayasena SD (2003) Functional siRNAs and miRNAs exhibit strand bias. *Cell* 115:209-216.
- Mohammed A, Dasaradhi PV, Bhatnagar RK, Chauhan VS and Malhotra P (2003) *In vivo* gene silencing in *Plasmodium berghei* - A mouse malaria model. *Biochem Biophys Res Commun*. 309:506-511.
- Ogita S, Uefuji H, Yamaguchi Y, Koizumi N and Sano H (2003) Producing decaffeinated coffee plants. *Nature* 423:823.
- Pei Y and Tuschl T. (2006) On the art of identifying effective and specific siRNAs. *Nat Methods* 3:670-6.
- Rand TA, Petersen S, Du F and Wang X (2005) Argonaute2 cleaves the anti-guide strand of siRNA during RISC activation. *Cell* 123:621-619.
- Schwarz DS, Hutvagner G, Du T, Xu Z, Aronin N and Zamore PD (2003) Asymmetry in the assembly of the RNAi enzyme complex. *Cell* 115:199-208.
- Xia H, Mao Q, Eliason SL, Harper SQ, Martins IH, Orr HT, Paulson HL, Yang L, Kotin RM and Davidson BL (2004) RNAi suppresses polyglutamine-induced neurodegeneration in a model of spinocerebellar ataxia. *Nat Med* 10:816-820.

Internet Resources

Generunner program at www.generunner.com.

The Strand Analysis (SA) program is available for academic use at <http://lgm.fcm.unicamp.br:9001/cgi-bin/SA/SA.cgi>.

Supplementary Material

The following online material is available for this article:

Figure S1

Figure S2

Figure S3

This material is available as part of the online article from <http://www.scielo.br/gmb>.

Associate Editor: Luciano da Fontoura Costa

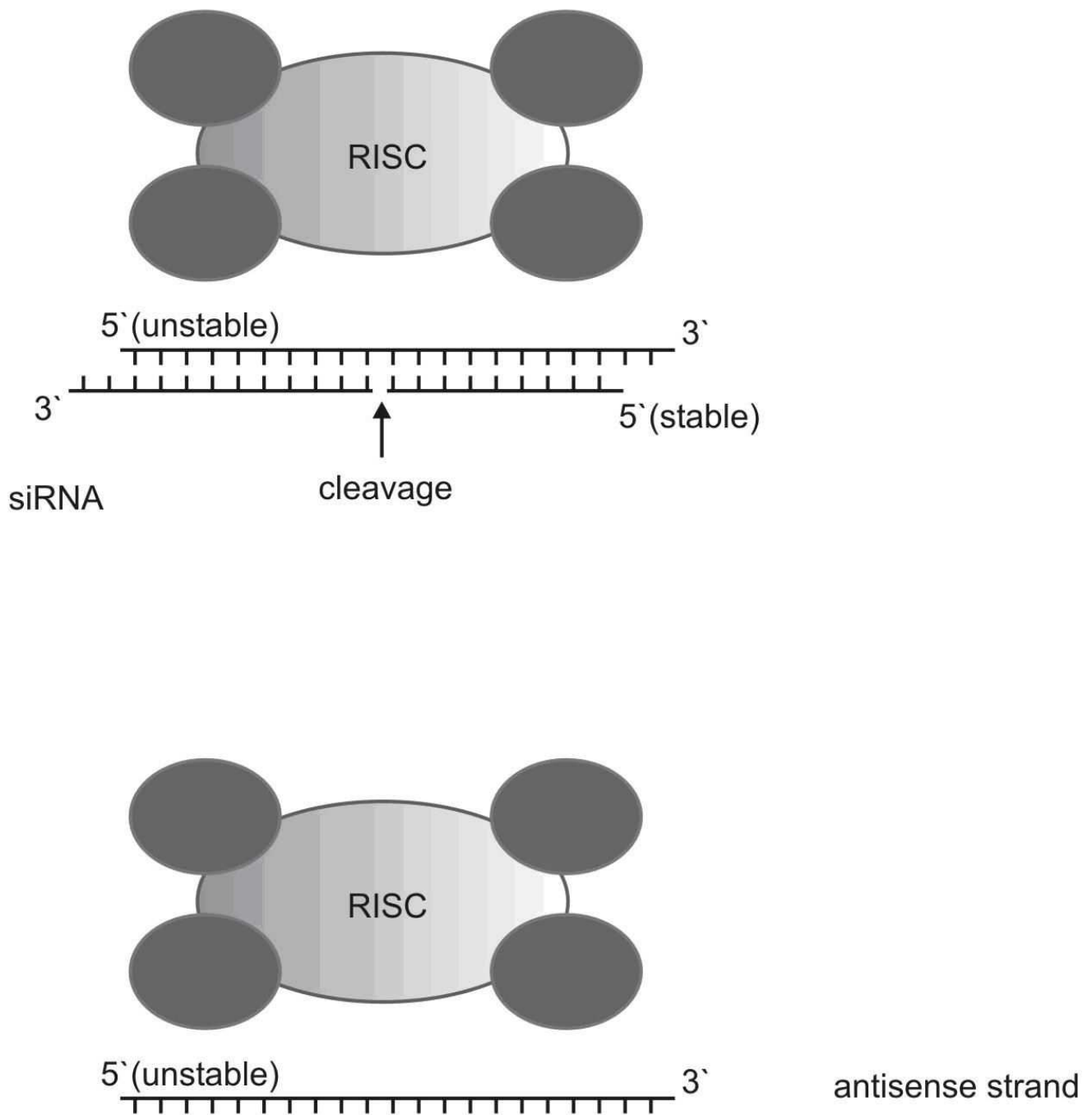


Figure S1. Role of thermodynamic stability features of small interfering RNAs (siRNAs) termini in RNA-induced silencing complex (RISC) activity. The more stable 5' strand is cleaved while the more unstable 5' strand is kept within the RISC for target RNA cleavage. Strand Analysis displays RNAi targets for which the antisense strands remains in RISC, i.e. which have a positive Gibbs free energy (ΔG) value.

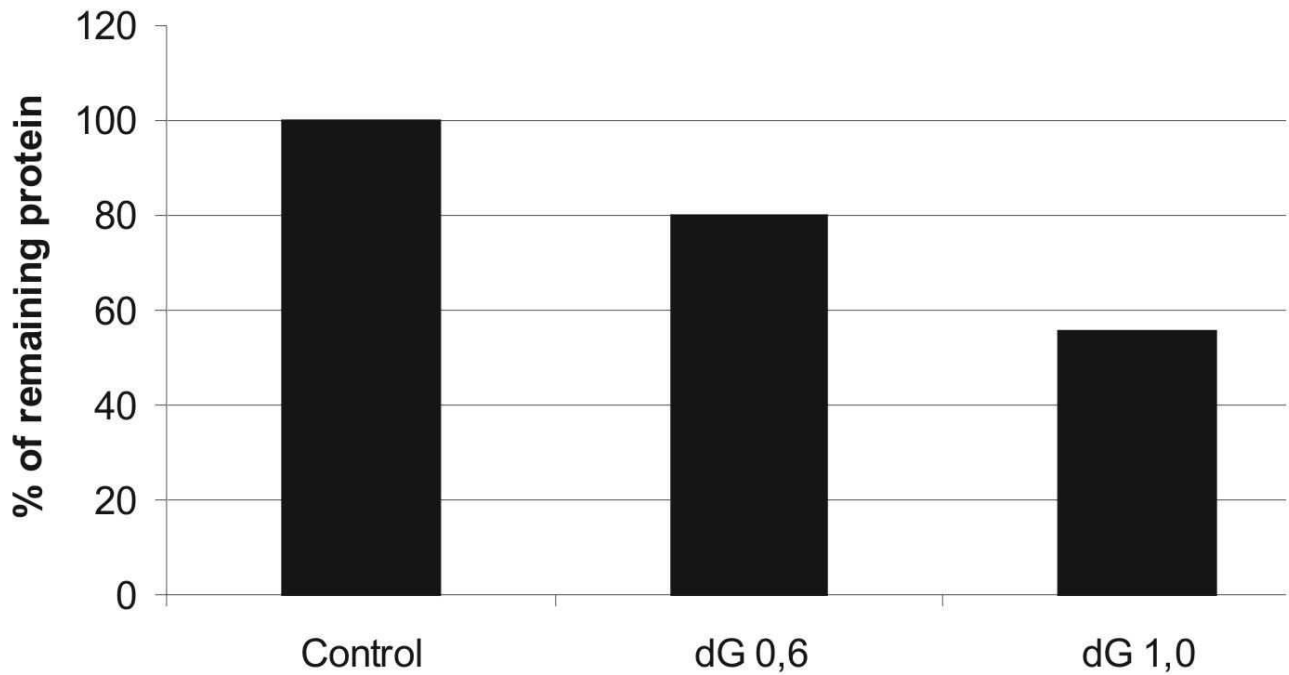


Figure S2. Gibbs free energy (ΔG) values clearly correlate with silencing efficiencies. As a practical example of such correlation, two small interfering RNAs (siRNAs) directed against the MeCP2 gene were identified in our laboratory using the Strand Analysis (SA) program and evaluated *in vivo* in mice using a hydrodynamic transfection protocol*. Western blot followed by densitometric analysis of the gel clearly confirmed that the greater the ΔG value, the more efficient was the siRNA (below). As a negative control, mice were injected with phosphate buffered saline (PBS) only. *McCaffrey AP *et al.* (2002) RNA interference in adult mice. *Nature*. 418:38-9.

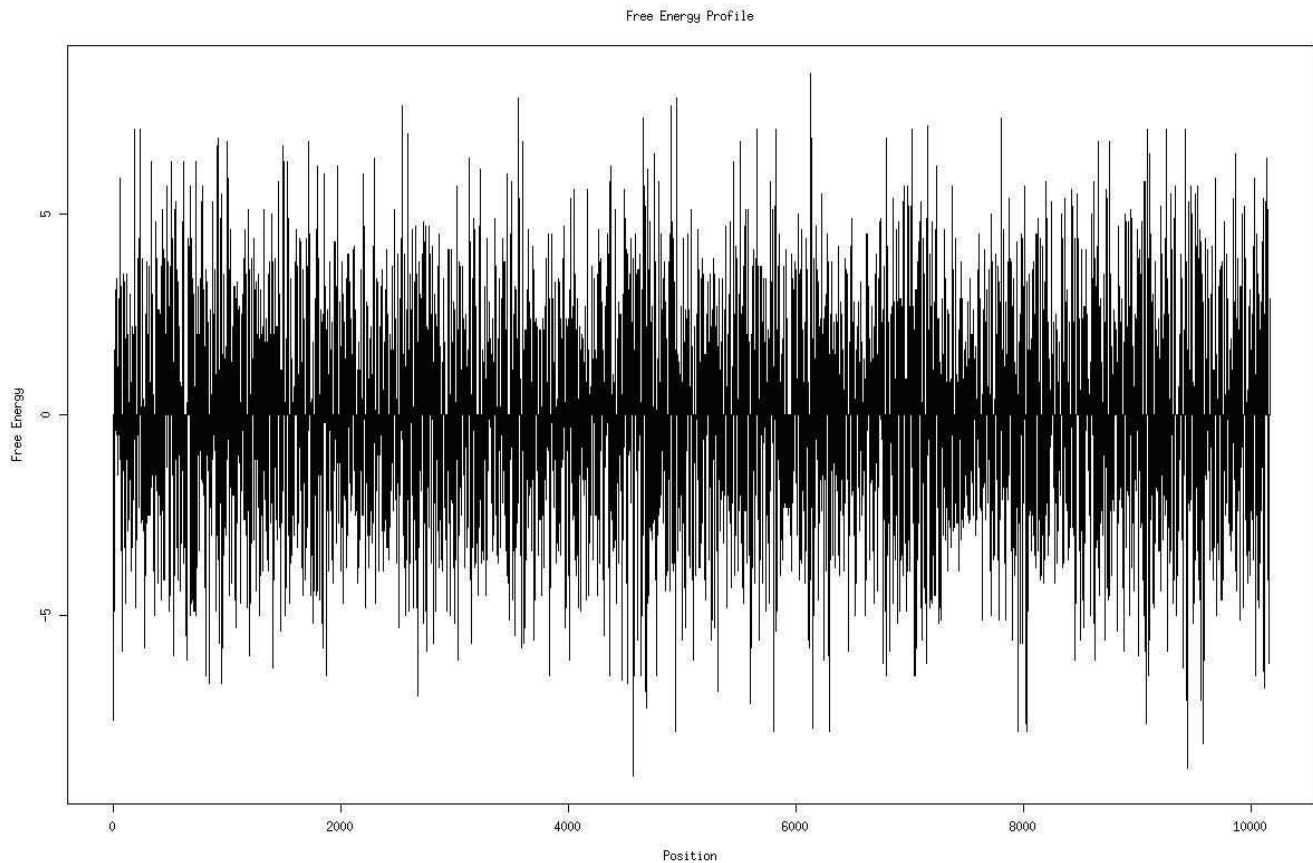


Figure S3. The Gibbs free energy (ΔG) landscape of a gene sequence. The Strand Analysis (SA) program calculates the ΔG value of each small interfering RNA (siRNA) along a specific DNA sequence and displays them in a comprehensive fashion, see graphic below using the human MeCP2 sequence (AF158180), with ΔG in the Y axis and the gene position in the X axis. This analysis indicates possible ΔG hotspots along the sequence and may also provide information regarding the ΔG profile of viral genomes, thus shedding light on viral adaptation to host RNAi responses.