Short Communication

Analysis of energetically biased transcripts of viruses and transposable elements

Rodrigo Secolin[1], Vinícius D'Ávila Bitencourt Pascoal[1,*], Iscia Lopes-Cendes[1] and Tiago Campos Pereira[2]

[1]*Departamento Genética Médica Faculdade de Ciências Médicas, Universidade de Campinas, Campinas, SP, Brazil.*
[2]*Departamento de Biologia, Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, SP, Brazil.*

Abstract

RNA interference (RNAi) is a natural endogenous process by which double-stranded RNA molecules trigger potent and specific gene silencing in eukaryotic cells and is characterized by target RNA cleavage. In mammals, small interfering RNAs (siRNAs) are the trigger molecules of choice and constitute a new class of RNA-based antiviral agents. In an efficient RNAi response, the antisense strand of siRNAs must enter the RNA-induced silencing complex (RISC) in a process mediated by thermodynamic features. In this report, we hypothesize that silent mutations capable of inverting thermodynamic properties can promote resistance to siRNAs. Extensive computational analyses were used to assess whether continuous selective pressure that promotes such mutations could lead to the emergence of viral strains completely resistant to RNAi (*i.e.,* prone to transfer only the sense strands to RISC). Based on our findings, we propose that, although synonymous mutations may produce functional resistance, this strategy cannot be systematically adopted by viruses since the longest RNAi-refractory sequence is only 10 nt long. This finding also suggests that all mRNAs display fluctuating thermodynamic landscapes and that, in terms of thermodynamic features, RNAi is a very efficient antiviral system since there will always be sites susceptible to siRNAs.

Keywords: RNAi, siRNA, synonymous mutation, viral evolution.

RNA interference (RNAi) is a process by which double-stranded RNA molecules (dsRNAs) trigger potent, specific post-transcriptional gene silencing in eukaryotic cells and is characterized by target RNA cleavage (Fire *et al*., 1998; Martinez *et al*., 2002). RNAi has been successfully applied in reverse genetics (for gene function identification), biotechnology (development of genetically modified plants and animals) and experimental therapeutics (by silencing key replication genes in pathogens) (Castanotto and Rossi, 2009). Some of the natural roles of RNAi include controlling viral replication (Ding, 2010) and transposon mobilization within the cellular environment (Levin and Moran, 2011).

Small interfering RNAs (siRNAs), the triggers of choice for gene knockdown in mammalian cells, consist of two short (21 nucleotides, nt; Figure S1) RNA strands

that promote effective cleavage of target RNAs (Elbashir *et al*., 2001). Once a viral sequence (a gene or the complete genome) is available, specific siRNAs can be designed against it to control pathogen replication *in vitro* or *in vivo*.

The basic mechanism of siRNA action has been established. Inside the infected cell, siRNAs are transferred to an endogenous protein complex known as RISC (RNA-induced silencing complex) where one strand is cleaved (referred to as the *passenger strand*) while the other is retained intact (referred to as the *guide strand*). RISC then patrols the cytoplasm to search for transcripts with perfect complementary sequences to the guide strand, after which the identified RNA targets are then cleaved, *i.e.,* silenced (Castanotto and Rossi, 2009).

Efficient gene silencing requires that the antisense RNA strand (complementary to the targeted transcript) be the guide strand. This strand choice is not random in the siRNA pathway since the strand with the thermodynamically more unstable 5' terminus is preferentially retained in the RISC (Khvorova *et al*., 2003; Schwarz *et al*., 2003). This characteristic is determined by the four base pairs (bp) at the two termini of the siRNA (hereafter simply referred

Send correspondence to Tiago Campos Pereira. Departamento de Biologia, Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, Universidade de São Paulo, Avenida Bandeirantes 3900, 14040-901 Ribeirão Preto, SP, Brazil. E-mail: tiagocampospereira@ffclrp.usp.br.
*Present address: Departamento de Ciências Básicas, Universidade Federal Fluminense, 24020-420 Nova Friburgo, RJ, Brazil.

to as 'termini'). The energy difference between these termini determines which strand will be the guide strand.

To date, three resistance mechanisms are known for artificial siRNA-based agents: 1) virally-encoded suppressor proteins (Hao *et al.*, 2011), 2) negative regulators such as eri-1 protein, non-toxic concentrations of cadmium and low temperatures (Kameda *et al.*, 2004), and 3) point mutations in the targeted transcript (Holz *et al.*, 2012). Suppressor proteins and negative regulators can sequester/cleave siRNAs or hinder their biosynthesis, whereas point mutations hinder target cleavage by RISC because of imperfect base pair complementarity. In the latter situation, re-sequencing the targeted site allows the design of a new effective siRNA (restoration of perfect base pair complementarity).

Viruses are under selective pressure since RNAi is a natural antiviral system. We therefore reasoned that under such conditions viruses could escape through a different mechanism, *i.e.,* synonymous mutations that invert the thermodynamic energy within the target site (Figure S1). This special case differs entirely from a simple point mutation since re-sequencing of the targeted site would not allow the design of an effective siRNA because the guide strand would be the sense strand.

We decided to investigate whether, under hypothetical continuous selective pressure, viral genomes could evolve to develop a biased energy gradient along their entire sequences, *i.e.,* be able to transfer only the sense strands to RISC (Figure S2). If this were possible then such viruses would be refractory to RNAi defense systems. To test this hypothesis, we undertook computational analyses with a newly developed computer program designed to generate energetically biased transcripts (EBTs), *i.e.*, an RNA from which all derived siRNAs tend to transfer the sense strand to RISC.

One means of generating an EBT is to align siRNA termini in a biased fashion (Figure S2). Initially, to evaluate 4-mer terminus length sequences (the natural size of siRNA termini), we calculated the thermodynamic values based on Khvorova *et al.* (2003) for all possible 1,024 5-mer RNA sequences. For this, let N be the vector of nucleotides {A, U, G, C}, Z the vector of 2-mer RNA sequences {AA, AC, AG, AU, CA, CC, CG, CU, GA, GC, GG, GU, UA, UC, UG, UU} and $Q(Z_x Z_y)$ the thermodynamic value for each 2-mer RNA within the 5-mer sequence. The algorithm initially evaluated whether this sequence showed increasing or decreasing order of Q values and then advanced one nucleotide N_j forward in the 5-mer sequence (Z_{i+1}). From this new nucleotide position, new Q values were calculated for all 2-mers backwards along the sequence ($Z_{i-3}...Z_{i+1}$). If, after these calculations, the sequence retained its increasing or decreasing ordering then the algorithm included a new nucleotide ahead of the sequence, otherwise it stopped. This analysis was done to include all of the nucleotides over time in N_j (Figure 1). The algorithm was implemented in a
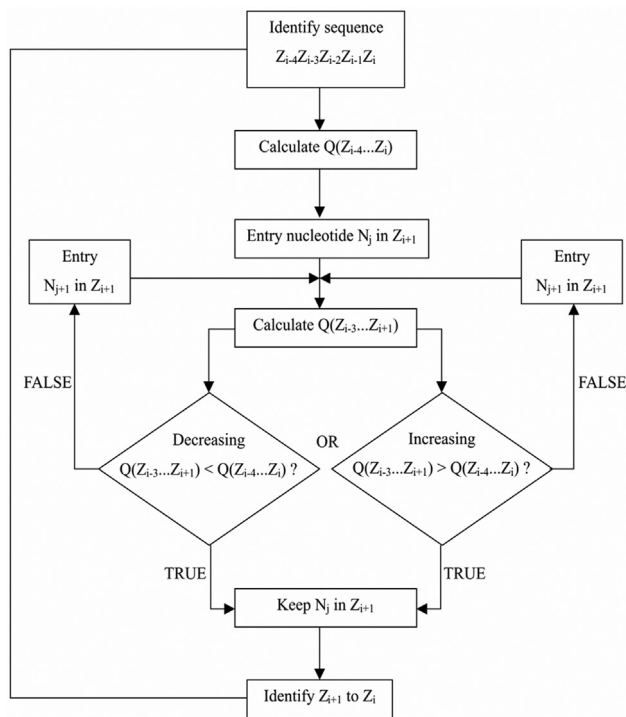

Figure 1 - Algorithm workflow. The algorithm calculates the free energy of each siRNA terminus (5 bp) in a stepwise fashion. For example, the free energy of the hypothetical terminus ACCGAU corresponds to the sum of the free energies of the following 2-mer RNA sequences (Z): AC+CC+CG+GA+AU. These Z values were originally described by Khvorova *et al.* (2003). Once all possible termini have their free energies determined, the algorithm aligns them based on nucleotide sequence and decreasing and increasing free energies (performed separately). N – vector of nucleotides {A, U, G, C}, Z – vector of 2-mer RNA sequences {AA, AC, AG, AU, CA, CC, CG, CU, GA, GC, GG, GU, UA, UC, UG, UU} and Q – thermodynamic value from each 2-mer RNA within the 5-mer sequence.

computer program (T-ASSEMBLER) using Perl scripts and is available from: http://www.laboratoriomultiusuario.com.br/index.php?option=com_content&view=article&id=75&Itemid=74.

We also determined the biased alignment by considering termini of different sizes, ranging from 4 to 10 base pairs. This upper limit was based on the fact that conventional siRNA consists of a central region of 19 bp such that the longest possible non-overlapping terminus is 9 bp in size. Alignments were done using T-ASSEMBLER after changing the corresponding parameters. The EBT sequences generated were translated in all possible six frames using the freeware program *Gene Runner* ver. 3.05 (Hastings Software, Inc.).

Our hypothesis was that EBTs would emerge from the selective constraint that RNAi exerted upon viruses and transposable elements. A BLAST search (Altschul *et al.*, 1990) for short, nearly exact matches was used to determine whether our computer-generated EBTs were detectable in these two classes of genetic elements. In view of the time required for processing, only the longest generated EBT

(10 nt in length) was run in searches against viral sequences. The BLAST search was optimized by altering some parameters in which the virus name was indicated in "Organism Optional" and the "Low Complexity Regions" filter was removed. The viruses searched were human immunodeficiency virus (GenBank no. 11676), hepatitis C virus (GenBank no. 11103), rabies virus (GenBank no. 11292), influenza A virus (GenBank no. 11320) and human herpesvirus 1 (GenBank no. 10298).

We tested seven sequences (all of which were the longest EBTs for conventional termini) for transposon-related hits in the human genome. The search was done by indicating 'RNA reference sequences (refseq_rna)' as the "Database", '*Homo sapiens*' in "Organism Optional", 'transposon'/'retrotransposon' in "Entrez Query Optional" and removing the "Low Complexity Regions" filter. All other parameters were set as default. We considered only hits that displayed complete identity to the query.

Computational analyses using T-ASSEMBLER revealed that for 4 bp-termini the number of termini with unique nucleotide sequences was 1,024; the number of termini with unique free energy (G) values was 72 and the number of unique longest termini was 7. However, these EBTs can be no longer than ten nucleotides (Table 1). Each one of these decanucleotides consisted of the biased alignment of seven 4-bp termini.

This group of EBTs (5' to 3': UUUUAGGGUG, UUUUACGGUC, UUUUACCGUC, UUUUACCCUC, UUUUACGGGA, UUUUACCGGA, UUUUACCCGA) encoded tripeptide chains in four reading frames (+1, +2, -1 and -2). The sequence UUUUACGGUC, for example, might encode for FTV (frame +2), FYG (+1), KRD (-1) and KVT (-2).

As expected, an increase in terminus size was accompanied by increases in almost all of the other parameters, including the G values (Table 1). For 5-bp termini, for example, the number of termini with unique sequences quadruplicated, the number of domains with unique G values increased by 26 units, the longest aligned sequence was extended by 3 bp and the range of energy difference increased by 2.2 units (from 9.3 kcal/mol in 4-bp termini to 11.5 kcal/mol in 5-bp termini). In a hypothetical 9-bp termini scenario, although the number of termini with unique sequences expanded to over one million, the number with unique G values increased only to 285. More importantly, the longest aligned sequence was only 19 nt.

Only one of the seven longest EBTs (UUUUAGGGUG) identified in 4-bp termini was used as a query in the search for identical hits in GenBank. This sequence showed total identity to two sites ("envelope glycoprotein" and "truncated gag protein") in human immunodeficiency virus and to three sites in hepatitis C virus: (polyprotein gene, genomic sequence and E1-E2 region). Three other sites were identified in influenza virus A: hemagglutinin, neuraminidase and segment 4. Only partial identities were obtained with two other viruses: rabies virus (genomic sequence; 9 out of 10 nt) and *human herpesvirus 1* isolate (UL4 gene; 9 out of 10 nt).

There were no perfect matches (hits) for transposable elements. Most hits were only partial (7-9 nt), as was the case for the "*Homo sapiens* piggyBac transposable element derived 3 (PGBD3)" for which the sequence UUUUACCCUC showed partial identity with TTTTACCCT. This set of seven EBTs frequently displayed hits with the same group of transposable elements, namely, *Homo sapiens* piggyBac transposable element derived (PGBD), SET domain and mariner transposase fusion gene, *Homo sapiens* tigger transposable element derived 2 and *Homo sapiens* harbinger transposase derived 1 (HARBI1). A search for retrotransposon sequences produced similar results, as shown in Table 2.

RNAi is a natural immune system (Ding, 2010) that acts by dicing the invading viral dsRNAs. The result is a pool of virus-derived siRNAs that in turn is used to load RISCs against this same replicating pathogen until it is eliminated. Alternatively, synthetic siRNAs may be designed and administered as an antiviral therapy.

A careful analysis of literature reports yielded experimental data in accordance with our basic hypothesis. One of these examples involved hepatitis B virus (HBV) that

Table 1 - Features of the termini analyzed and the sequence alignments that retained the thermodynamic gradient. The lowest and highest free energy (G) values were obtained for all of the sequences that were aligned.

Terminus length (bp)	Number of termini with unique sequences	Number of termini with unique free energy (G) values	Size (nt) of the longest aligned sequence	Number of unique longest sequences	Lowest terminus free energy (G) value (kcal/mol)	Highest terminus free energy (G) value (kcal/mol)
4	1024	72	10	7	-13.7	-4.4
5	4096	98	13	4	-17.0	-5.5
6	16 384	120	14	5	-20.3	-6.6
7	65 536	144	16	151	-23.6	-7.7
8	262 144	166	17	190	-26.9	-8.8
9	1 048 576	210	18	223	-29.8	-9.9
10	4 194 304	285	19	290	-32.8	-10.1

Table 2 - Identification of EBTs within viral and transposable element sequences. The accession numbers listed refer to the viral/transposable element sequences with identity to the ETB.

EBT sequence	Virus/transposable element	Sequence	Accession number	Alignment region	E-value
UUUUAGGGUG	HIV	envelope glycoprotein	JF506161.1	TTTTAGGGTG (positions 1056-1047)	609
	Hepatitis C virus	polyprotein gene	FJ515103.1	TTTTAGGGTG (positions 39-48)	109
	Rabies virus	genomic sequence	JQ944709.1	TTTTAGGGT (positions 6456-6464)	73
	Influenza virus A	hemagglutinin	CY108710.1	TTTTAGGGTG (positions 945-936)	441
	Human herpesvirus 1	UL4 gene	AY240369.1	TTTTAGGGT (positions 314-306)	48
UUUUAGGGUG	PEG10*	-	NM_001172438.1	TTTTAGGGTG (positions 6057-6066)	0.17
	PGBD2**	-	NM_170725.2	TTTTAGGG (positions 1302-1295)	3.4
UUUUACGGUC	RTL1*	-	NM_001134888.2	TTTTACG (positions 1878-1884)	10
	SETMAR**	-	NM_001243723.1	TACGGTC (positions 973-979)	13
UUUUACCGUC	RTL1*	-	NM_001134888.2	TTACCGTC (positions 3785-3792)	2.6
	PGBD3**	-	NM_170753.2	TTTTACC (positions 1237-1231)	13
UUUUACCCUC	RGAG4*	-	NM_001024455.3	TACCCTC (positions 1108-1114)	10
	PGBD3**	-	NM_170753.2	TTTTACCCT (positions 1237-1229)	0.85
UUUUACGGGA	RTL1*	-	NM_001134888.2	TTTTACG (positions 1878-1884)	10
	SETMAR**	-	NM_001243723.1	TTTTACG (positions 1628-1634)	13
UUUUACCGGA	PEG10*	-	NM_001172438.1	TACCGGA (positions 2496-2490)	10
	PGBD4**	-	NM_152595.4	TTTACCGG (positions 721-728)	3.4
UUUUACCCGA	GIN1*	-	NM_017676.2	TACCCGA (positions 3072-3066)	10
	SETMAR**	-	NM_001243723.1	TTTACCCG (positions 68-61)	3.4

*Retrotransposon. **Transposon. GIN1: gypsy retrotransposon integrase 1, PEG10: paternally expressed 10, PGBD (2-4): piggyBac transposable element derived (2-4), RGAG4: retrotransposon gag domain containing 4, RTL1: retrotransposon-like 1, SETMAR: SET domain and mariner transposase fusion gene.

was isolated from patients, multiplied in cell cultures and challenged with an RNAi-based treatment (Wu *et al.*, 2005). A plasmid expressing a short hairpin RNA (shRNA) was designed against a conserved region encoding two overlapping genes ("Pol" and "S") in the same strand (frame +1 and +3). This process led to the selection of a resistant variant after a few cycles of replication *in vitro*. Sequencing of these clones revealed that they displayed a single point mutation in the 3' terminus of the target site (corresponding to the 5' terminus of the antisense strand of the siRNA). More intriguingly, this mutation was silent for both open reading frames and thus affected the fifth nucleotide in the terminus.

Calculation of the free energy associated with a 4-bp terminus requires the use of five base pairs (Khvorova *et al.*, 2003). When the specific mutation within the HBV sequence was analyzed the original siRNA used to control viral replication was found to have a ΔG of -0.2 kcal/mol (ΔG = the G value of the antisense strand minus the G value of the sense strand). This value was close to zero, indicating that both sense and antisense strands had nearly the same chance of being retained in the RISC. However, the point mutation created a ΔG of -1.0 kcal/mol that led to retention of the sense strand in RISC. This report by Khvorova *et al.*

(2003) provided very clear experimental evidence of the basic principle investigated here, namely, inversion of the thermodynamic features by a silent mutation in viruses escaping siRNA-based therapies.

Since RNAi exerts selective pressure on viruses and transposable elements it would be expected that, in the long term, these synonymous mutations would convert their genomes into inverted energetically-biased sequences (EBTs). Such viruses would be refractory to RNAi. To test this hypothesis, we used computational analyses to generate EBTs and verify whether they occurred within viral and transposable element-related sequences.

Although the number of termini with unique sequences for conventional 4-bp termini was large (1,024), the longest EBTs were restricted to only ten nucleotides – shorter than a single siRNA. Consequently, it is impossible to have viral transcripts (or genomes) that are refractory throughout their entire sequences. If the siRNA structure were different, *i.e.*, termini 5-9 bp in length, the number of unique sequences would be 4,096 and 1,048,576 respectively. However, although this parameter increases geometrically, all the others expand arithmetically (Table 1). Thus, even in an extreme situation in which the siRNA terminus was extended to its maximum length (half of the entire structure - 9 bp), EBTs would be not longer than 18 bp.

Consequently, although synonymous mutations capable of inverting terminal energetics represent a potentially new mechanism of resistance, they can be used for only a limited number of sites within viral transcripts, without systematically covering the entire viral genome. Based on these considerations, we predict that a considerable number of sites in a pathogen's genome will always remain susceptible to siRNAs as it is not possible to create a long EBT. Hence, in terms of thermodynamics, RNAi is indeed a very efficient antiviral system.

One of the EBTs identified (UUUUAGGGUG) was found in all of the viruses analyzed (some of the most relevant to human medicine), with all of the viruses having 2-3 sites with full identity to this sequence. Additionally, all seven decanucleotide EBTs displayed hits to exactly the same restricted group of transposable elements. Although these sites were identified, they did not correspond to RNAi-refractory domains since they were shorter than a single siRNA (21 nt). These regions probably emerged as a result of other evolutionary forces rather than selection for refractoriness to gene silencing.

Previous work has shown that several RNAs display "thermodynamic landscapes" that continuously fluctuate between positive and negative values (Pereira *et al*., 2007). These landscapes do not exhibit long regions of uninterrupted ascending (or descending) values, *i.e.,* a thermodynamic gradient, and the positive/negative values correlate directly with the accessibility/inaccessibility to siRNAs. Consequently, all of the RNAs analyzed displayed amenable/refractory regions to RNAi.

The findings of the present study agree with the observation in the preceding paragraph. Since our computational analyses revealed that it was not possible to align thermodynamic domains over long extensions, *i.e.*, it was impossible to generate long EBTs, then RNAs will naturally fluctuate between positive and negative thermodynamic values. More importantly, since we undertook extensive, complete alignment procedure, our current findings generalize the previous observations by providing evidence that no long EBTs exist and that all RNAs therefore have fluctuating thermodynamic landscapes. In thermodynamic terms, all RNAs display sites amenable to RNA interference.

It is noteworthy that greater differences between the thermodynamic stabilities of termini result in a trend of only one strand preferentially entering RISC. However, cases of miRNA precursors in which both strands (5' and 3') occur in RISC have been reported. This may occur because: (a) the difference between termini stabilities is close to zero (such that both strands tend to enter RISC) or (b) because accessory proteins help to load the unfavored strand. This situation does not invalidate our hypothesis since condition (a) is also takes into consideration termini thermodynamics and situation (b) may apply only to endogenous miRNA sequences and not to viral-derived siRNAs.

The first evolutionary resistance strategies may have been based on synonymous mutations that inverted the termini thermodynamics. However, as shown here, this explanation cannot be systematically applied to the entire genome so that other more elaborate strategies were required. Suppressor proteins, which represent a much more challenging evolutionary acquisition, must have emerged in the face of such pressure.

The new mechanism of resistance to siRNA-mediated silencing proposed here is based on certain silent mutations that cause the inversion of thermodynamic features within the targeted site, leading to inefficient loading of RISC with the sense strand. We believe that this strategy could be adopted by viruses of any nature (single-stranded/double stranded viruses, DNA or RNA viruses and viruses that infect insects, plants or animals) since RNAi acts on messenger RNAs that are present in any viral life cycle.

An analysis of experimental data reported in the literature clearly supports the concept of synonymous mutations as a strategy to escape siRNAs. Although RNAi exerts selective pressure on viruses, our computational analyses revealed that this strategy cannot be systematically adopted by viruses since the longest possible EBT is only 10 nt long. Our data therefore do not support our initial hypothesis that viral genomes developed a biased energy gradient along their entire sequences. Nevertheless, our findings confirm previous observations and allow the generalization that all RNAs display fluctuating thermodynamic landscapes. They also reveal that, thermodynamically, RNAi is a very sophisticated antiviral system from which pathogens cannot systematically escape by synonymous mutation.

Acknowledgments

References

Altschul SF, Gish W, Miller W, Myers EW and Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403-410.

Castanotto D and Rossi JJ (2009) The promises and pitfalls of RNA-interference-based therapeutics. Nature 457:426-433.

Ding SW (2010) RNA-based antiviral immunity. Nat Rev Immunol 10:632-644.

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.

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.

Hao X, Lu A, Sokal N, Bhagwat B, Leung E, Mao R, Reade R, Wu Y, Rochon D and Xiang Y (2011) Cucumber necrosis virus p20 is a viral suppressor of RNA silencing. Virus Res 155:423-432.

Holz CL, Albina E, Minet C, Lancelot R, Kwiatek O, Libeau G and Servan de Almeida R (2012) RNA interference against animal viruses: How morbilliviruses generate extended diversity to escape small interfering RNA control. J Virol 86:786-95.

Kameda T, Ikegami K, Liu Y, Terada K and Sugiyama T (2004) A hypothermic-temperature-sensitive gene silencing by the mammalian RNAi. Biochem Biophys Res Commun 315:599-602.

Khvorova A, Reynolds A and Jayasena SD (2003) Functional siRNAs and miRNAs exhibit strand bias. Cell 115:209-216.

Levin HL and Moran JV (2011) Dynamic interactions between transposable elements and their hosts. Nat Rev Genet 12:615-627.

Martinez J, Patkaniowska A, Urlaub H, Lührmann R and Tuschl T (2002) Single-stranded antisense siRNAs guide target RNA cleavage in RNAi. Cell 110:563-574.

Pereira TC, Pascoal VDB, Secolin R, Rocha CS, Maia IG and Lopes-Cendes I (2007) Strand Analysis, a free online program for the computational identification of the best RNA interference (RNAi) targets based on Gibbs free energy. Genet Mol Biol 30:1206-1208.

Schwarz DS, Hutvágner G, Du T, Xu Z, Aronin N and Zamore PD (2003) Asymmetry in the assembly of the RNAi enzyme complex. Cell 115:199-208.

Wu HL, Huang LR, Huang CC, Lai HL, Liu CJ, Huang YT, Hsu YW, Lu CY, Chen DS and Chen PJ (2005) RNA interference-mediated control of hepatitis B virus and emergence of resistant mutant. Gastroenterology 128:708-716.

Supplementary Material

The following online material is available for this article:

Figure S1 - Escape from the RNAi response through inverted thermodynamics.

Figure S2 - A target refractory to gene silencing.

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

Associate Editor: Carlos F.M. Menck

Supplementary Fig. 1. Escape from the RNAi response through inverted thermodynamics. A virus may evade a specific siRNA by mutating its target sequence (through synonymous mutations, for example), inverting the terminus thermodynamics (G) from AUUA (-4.7 kcal^{-mol}, the guide strand is the antisense strand) to ACUG (-7.6 kcal^{-mol} ; the guide strand is the sense strand). Free energy values were calculated according to Khvorova *et al*., 2003. Gray arrows indicate the terminus from which the unwinding process initiates. Q: Glutamine K: Lysine, L: Leucine.
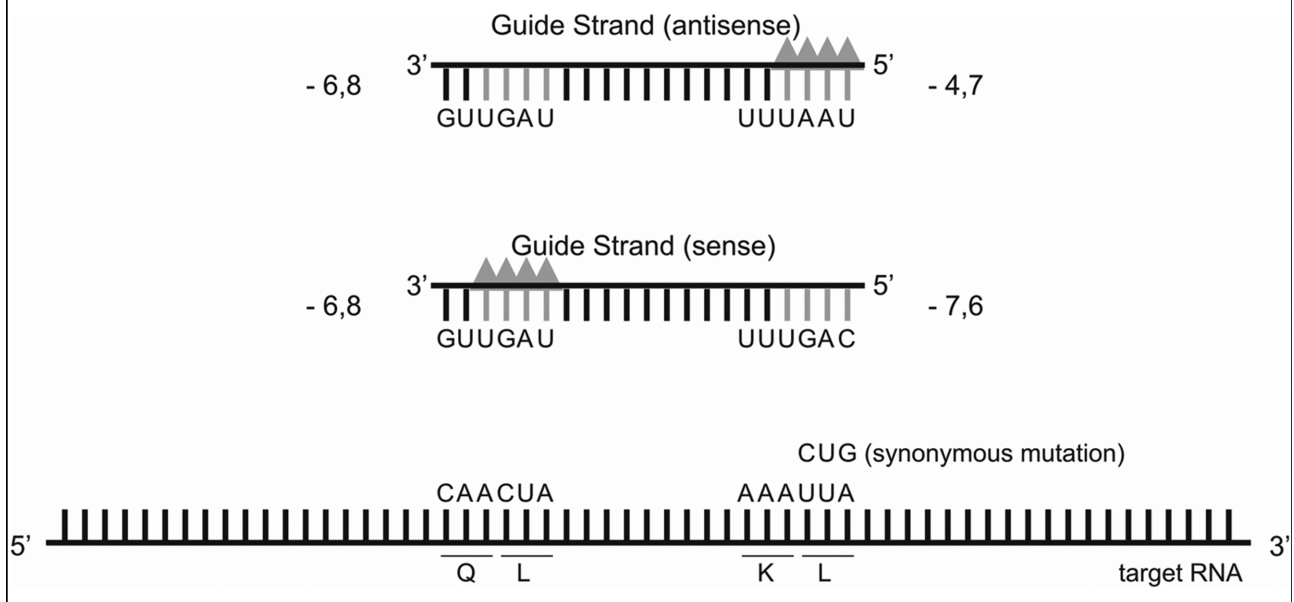
Supplementary Fig. 2. A refractory target to gene silencing. Such molecule is an RNA sequence with a biased thermodynamic gradient throughout its sequence (represented by the box flowing from black to white). Its 5' end begins with the highest terminus thermodynamics value (-4,4 kcal^{-mol}; gray arrowheads) to the lowest (-13,4 kcal^{-mol}; gray lines). siRNAs against this transcript cannot promote cleavage since they tend to keep the sense strand into RISC.