Structure and sequence motifs of siRNA linked with in vitro down-regulation of morbillivirus gene expression

SERVAN de ALMEIDA R., KEITA D., LIBEAU G., ALBINA E.
CIRAD, EMVT, UR-15, Campus International de Baillarguet, 34398, Montpellier, France

The most challenging task in RNA interference is the design of active small interfering RNA (siRNA) sequences. Numerous strategies have been published to select siRNA. They have proved effective in some applications but have failed in many others. Nonetheless, all existing guidelines have been devised to select effective siRNAs targeting human or murine genes. They may not be appropriate to select functional sequences that target genes from other organisms like viruses. In this study, we have analyzed 62 siRNA duplexes of 19 bases targeting three genes of three morbilliviruses. In those duplexes, we have checked which features are associated with siRNA functionality. Our results suggest that the intramolecular secondary structure of the targeted mRNA contributes to siRNA efficiency. We also confirm that the presence of at least the sequence motifs U13, A or U19, as well as the absence of G13, collaborate to increase siRNA knockdown rates. Additionally, we observe that G11 is linked with siRNA efficacy. We believe that an algorithm based on these findings may help in the selection of functional siRNA sequences directed against viral genes.

INTRODUCTION

RNA interference (RNAi) pathway is triggered by longer double-stranded precursor RNA, which is cleaved in fragments of 21-25 nt, the small interfering RNAs (siRNAs), that recognize, combine, and degrade the target mRNA (2). Recent studies have shown that the most critical point in applying RNAi is the identification of potent siRNAs and that their functionality is affected by the duplex nucleotide composition and the accessibility of target RNA (1, 3, 4, 5). To solve this problem, various guidelines or algorithms to optimize the design of effective siRNAs have been published but although be effective under certain circumstances, they have failed in many others. Here, we carried out an analysis on 62 siRNA duplexes directed against three genes of morbilliviruses. We evaluated the participation of various features previously reported necessary for siRNA functionality.

MATERIALS AND METHODS

siRNA design

Sixty-two 19-bases long siRNAs were designed against three morbilliviruses, the human measles virus (MV), the bovine rinderpest virus (RPV), and the small ruminant "peste des petits ruminants" virus (PPRV). The design was made Ambion (Cenix algorithm) for 26 molecules or in-house for the other 36, without particular attention to their sequences. We therefore considered features like GC content of the duplex ranged between 32% and 58%, low internal stability of the 3′ terminal of the sense strand, and some desired bases in strategic positions, as S1 and W19.

Test of siRNA activity

Vero cells were transfected with 100, 50, 25 and 12.5nM of siRNAs and Lipofectamine 2000 (Invitrogen) using usual methodology. Twenty-four hours after transfection, cells were infected by using a MOI of 0.1 of PPRV, RPV or MV and 4 days later, the siRNA silencing effect was evaluated by flow cytometry. Inhibition level of 70% was established as cut-off value for siRNA efficacy.

Validation of positively correlated features with siRNA functionality

We analyzed the secondary structure of siRNA targets and the intrinsic sequence of siRNAs to define a procedure that would optimize the identification of effective siRNA. Secondary structures of mRNA targets were defined using the Mfold software. Last, the frequency of the 12 criteria (Tab.1) compiled from the literature and are related to siRNA activity, was determined on functional (*f*) versus non-functional (*nf*) siRNAs.

Tab 1: Criteria evaluated in the siRNA sequences

Criteria	Authors
1. At least four out of seven A/U residues in the 5′ terminal of the antisense strand* and four out of seven G/C residues in the 3′ terminal of the antisense strand	Ui-Tei et al. (2004)
2. Absence G, C stretches >4 nt	Ding et al. (2004), Henschel et al. (2004)
3. S1	Schwarz et al. (2003), Amarzguioui and Prydz (2004), Jagla et al. (2005)
4. A3	Reynolds et al. (2004), Schubert et al. (2005), Pekarik (2005)
5. A6	Amarzguioui and Prydz (2004)
6. U10	Amarzguioui and Prydz (2004), Reynolds et al. (2004), Yoshinari et al. (2004), Pekarik (2005), Shabalina et al. (2006)
7. W19	Amarzguioui and Prydz (2004), Reynolds et al. (2004), Huesken et al. (2005), Pekarik (2005), Shabalina et al. (2006)
8. U13	Pekarik (2005), Huesken et al. (2005)
9. Absence of G13	Arziman et al. (2005) Pekarik (2005)
10. S11	Jagla et al. (2005)
11. G16	Hsieh et al. (2004)
12. U18	Huesken et al. (2005)
13. Evaluation of secondary structure of mRNA target	Bohula et al. (2003), Luo and Chang (2004), Yoshinari et al. (2004), Heale et al. (2005), Overhoff et al. (2005), Schramm and Ramey (2005), Schubert et al. (2005), Westerhout et al. (2005), Yiu et al. (2005)

RESULTS AND DISCUSSION

Determination of siRNA functionality

Twelve out of 62 (19%) siRNA duplexes were functional according to the criterion of a minimum reduction of 70% of protein expression. We identified four siRNAs able to knockdown the N gene of PPRV, 5 active on the N gene of RPV and 3 active on the N gene of MMV.

Sequence stability and siRNA functionality

For the incorporation of the antisense strand into RISC complex, the siRNA duplex should have a weak binding affinity at the 5′ end of the antisense strand (2). As a measure of stability asymmetry, we considered a minimum of 4 out of 7 A/U residues in the 3′ end of the sense strand (5) and a minimum of 4 out of 7 G/C residues in the 5′ end of the sense strand. Nine out of our 12 functional siRNAs (75%) had a high incidence of A/U at the 3′ end of the sense whereas 38 out of 50 (76%) non-functional siRNAs also showed this criterion. Moreover, only 50% of functional siRNAs showed 4 out of 7 G/C residues in the sense 5′ end whereas 54% of non-functional siRNAs followed this rule. Interestingly, we observed that 8 out of 12 (>66%) of functional siRNAs had simultaneously a higher A/U content at the 5′ end of the antisense strand and a lower A/U content at the 5′ end of the sense strand while 24 out of 50 (48%) of non-functional siRNAs did not. Although the difference was not significant ($p \geq 0.05$), we believe that this parameter could probably be more relevant for siRNA activity than the strict requirement of a minimum number of A/U or G/C nucleotide-pairs at the siRNA extremities (5).

Secondary structure and siRNA functionality

We found that hairpin structures and target sites located between two very close branches possibly contributed to the inefficacy of siRNAs (Fig 1).

One hundred percent of our (*f*) siRNAs targeted mRNA positions, which were accessible in more than 60% of the predicted secondary RNA structures and 68% of (*nf*) siRNAs targeted positions with less than 60% of accessible secondary mRNA structures. The frequency of the accessible secondary structures in active siRNAs was significantly higher ($p \leq 0.05$) than in non-active siRNAs.

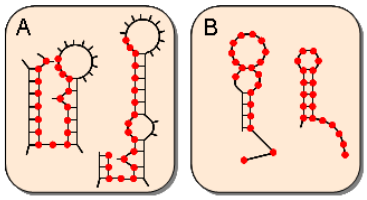

Fig. 1: Example of non-functional target sites (red). A. Target sites located between two very close branches. B.Hairpins.

Although, the structure stability could determine the inefficacy of our siRNAs sequences, a threshold of paired bases that confer this stability has to be determined. Ideally, the correlation between the stability level of the target structure and RNAi efficiency has to be done by a mutational analysis of one target instead of comparing different siRNAs with intrinsically different RNAi efficacy. This kind of evaluation was not possible with our set of data since we have analyzed siRNAs with different targets and nucleotide compositions.

Specific sequence motifs and siRNA functionality

The incidence of the nucleotides was determined for (*f*) and (*nf*) siRNA groups. Additionally, the incidence of the four nucleotides was determined at each position, and the (*f*) and (*nf*) sequences were compared (Fig. 2). This analysis confirmed the strong dominance of motif U13 (differential incidence of 49%), the absence of G13 (-42%), U10 (36%) and A/U19 (40%) in our (*f*) sequences. In contrast to other reports, motifs S1 (C = -15%; G = 18%), A3 (22%), A6 (-5%), and U18 (4%) were not dominant in the (*f*) siRNAs in this study. Moreover, the new motif G11 (differential incidence of 59% and a strong statistical difference ($p \leq 0.0001$) was over-represented in (*f*) siRNAs and under-represented in (*nf*) sequences. Additionally, other motifs were moderately over-represented like G2 (30%), U4 (34%), G7 (26%), C9 (30%), G16 (30%) and G15 (26%) or under-represented like A5 (-26%), U8 (-26%), G9 (-30%) and A15 (-30%).

The sequence that we can finally recommend as a model for the selection of functional siRNAs is a RNA duplex of 19 nucleotides with a 2 nucleotides 3′ overhang and the sense strand consisting of 5′- N1, G2, N3-9, U10, G11, N12, U(-G)13, N14-18 and W19- dTdT-3′ (where U(-G) at position 13 means the presence of U13 and absence of G13).

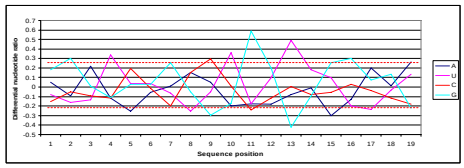

Fig. 2: Differential nucleotide incidence in functional versus non-functional siRNAs. For each 19 positions, the relative percentages of functional siRNAs with the four nucleotides were determined. Differences higher than 25% were considered over-representations of the corresponding nucleotides in functional siRNAs. Conversely, results lower than -25% were considered under-representations.

To optimize the identification of effective siRNA, The sequences were selected using as eliminatory criterion an mRNA accessibility of at least 60%. After, the requirement for G2, U10, G11, U13, absence of G13 and W19 and a stability asymmetry between the 5′ and 3′ termini of the sense strand were confronted to the selected ones using an inclusive cut-off score of any four out of these seven features. This strategy allowed the selection of >83% functional siRNAs while only 8% of non-functional siRNAs could not be rejected (Fig. 3).

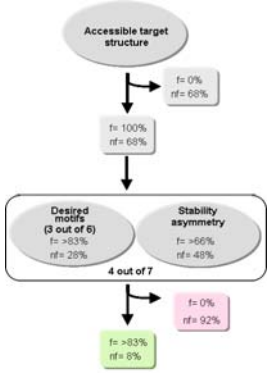

Fig. 2: Strategy for identification of effective siRNA. The procedure considers as potentially functional firstly the 19-nt siRNAs that target accessible sequences in at least 60% of the predicted mRNA secondary structures. Subsequently, the screened siRNAs are evaluated based on the stability asymmetry between the 5′ and 3′ termini of the sense strand (higher A/U content in the 3′ termini and a lower A/U content in the 5′ termini) and the desirable motifs (G2, U10, G11, U13, absence of G13 and W19). The siRNAs that fulfill 4 out of 7 of these latter criteria are finally selected. The percentages of our sequences that showed these attributes are in the green area. Abbreviations: f, functional siRNA; nf, non-functional siRNA.

ACKNOWLEDGEMENTS

This research was supported by Marie Curie International Fellowship within the 6th European Community Framework Program. This study was also partially granted by the European Union Network of Excellence EPIZONE, n°FOOD-CT-2006-016236.

REFERENCES

1. Amarzguioui & Prydz, 2004. *Biochem. Biophys. Res. Commun.* 316, 1050-1058.
2. Elbashir *et al.*, 2001. *EMBO Journal* 411, 494-498.
3. Fire *et al*, 1998. *Nature* 391, 806-811.
4. Heale *et al.*, 2005. *Nucleic Acids Res.* 33(3), e30.
5. Reynolds *et al.*, 2004. Nat. Biotechnol. 22, 326-330.

Centre de coopération Internationale pour la Recherche en Agronomie pour le Développement

CIRAD, UMR 15
Control of exotic and emerging animal diseases
TA30/G
Campus Baillarguet
34398 Montpellier cedex 5
France

Contact: Renata SERVAN de ALMEIDA
renata.almeida@cirad.fr