TROD: T7 RNAi Oligo Designer

Peter Dudek and Didier Picard*

Département de Biologie Cellulaire, Université de Genève, Sciences III, Quai Ernest-Ansermet 30, CH-1211 Genève 4, Switzerland

ABSTRACT

We have developed T7 RNAi Oligo Designer (TROD), a web application for RNA interference studies. TROD greatly facilitates the design of oligodeoxynucleotide sequences for the *in vitro* production of siRNA duplexes with T7 RNA polymerase. Given a query cDNA sequence, the program scans for appropriate target sequences based on the constraints of the T7 RNA polymerase method and published criteria for RNA interference with siRNAs. The output is an ordered and prioritized list of ready-to-order DNA oligonucleotide sequences, with links to perform a BLAST search to ascertain target sequence specificity. The TROD web service is available at http://www.cellbio.unige.ch/RNAi.html.

INTRODUCTION

The recent discovery of the process of RNA interference (RNAi) is rapidly leading to the development of novel tools for specific inhibition of gene expression. In mammalian cells, short (21–24 nt) interfering dsRNA species (siRNAs) have been shown to be the most effective (1). Recently, we developed a simple and inexpensive method for the production of siRNAs *in vitro* (2). In brief, it capitalizes on the ability of the T7 RNA polymerase to use partially double-stranded DNA oligonucleotides as templates to synthesize short RNA strands that can be annealed and applied to down-regulate expression of a specific gene of interest. For a comprehensive review including protocols, see Donzé *et al*. (3). A critical step in RNAi studies is choosing target sequences (4,5). Although two recent publications have provided insights into the rules that govern the selection of the proper RNA strand by the nucleolytic RISC complex (6,7), only a few criteria have been established to aid in this effort. They include low GC content (30–50%), an AT-rich complementary dinucleotide at the 3′ end of the antisense siRNA strand, a destabilizing base pair and internal AT-rich stretch at the 5′ end of the antisense strand, and a lack of long stretches of a single nucleotide, particularly G. Moreover, because T7 RNA polymerase has a strong preference for a G as initiation nucleotide (8),

the T7-based method is associated with an additional constraint: for a siRNA of 22 nt, this means a requirement for G and C at the +1 and +20 positions of the target sequence, respectively. In particular, for large genes or large numbers of genes, the process of target sequence selection becomes wearisome if done manually. Here, we describe a web-based program that greatly facilitates this step.

APPLICATION

The $\underline{T}$7 $\underline{R}$NAi $\underline{O}$ligo $\underline{D}$esigner (TROD) web application (Figure 1) was developed using Perl and the Bioperl module (9), and can, therefore, be deployed on almost any platform, and be accessible by any HTML-browsing application. Whereas most other online siRNA design tools primarily focus on selecting siRNA sequences for expensive commercial chemical synthesis, TROD is tailored to researchers using the T7 RNA polymerase method [reviewed in (3)]. Our goal was to offer a simple tool that facilitates the design, selection and ordering of the three synthetic DNA oligonucleotide sequences that are required. TROD can easily be coupled with recently developed high-throughput methods for siRNA screening in which the T7 RNA polymerase method is used (10). The program requires a single cDNA sequence as input, in the form of either a GenBank accession code, a file, or entered directly. It first scans for all occurrences of the following sequence: $N_2GN_{17-20}C$. The default is $N_2GN_{18}C$, but the user has the option of defining a different length. As mentioned above, the G and C nucleotides are required for efficient synthesis of the siRNA strands by T7 RNA polymerase. This sequence constraint is not a serious handicap, since the probability of the target (for one set length) appearing in an average DNA sequence is 1/16. In contrast, another recently reported method and corresponding web application (11) for designing T7 DNA oligonucleotides is far more limiting because of additional constraints, with a maximum hit probability of 1/64 (or 1/256 for an ideal sequence). Note that the frequency of hits that conform to the $GN_{17-20}C$ rule increases with GC content, whereas GC content is inversely correlated with target sequences that also fulfil the AT-richness criteria.

From each one of the captured target sequences, TROD generates two DNA oligonucleotides with an appended T7

---

*To whom correspondence should be addressed. Tel: +41 22 379 6813; Fax: +41 22 379 6928; Email: picard@cellbio.unige.ch
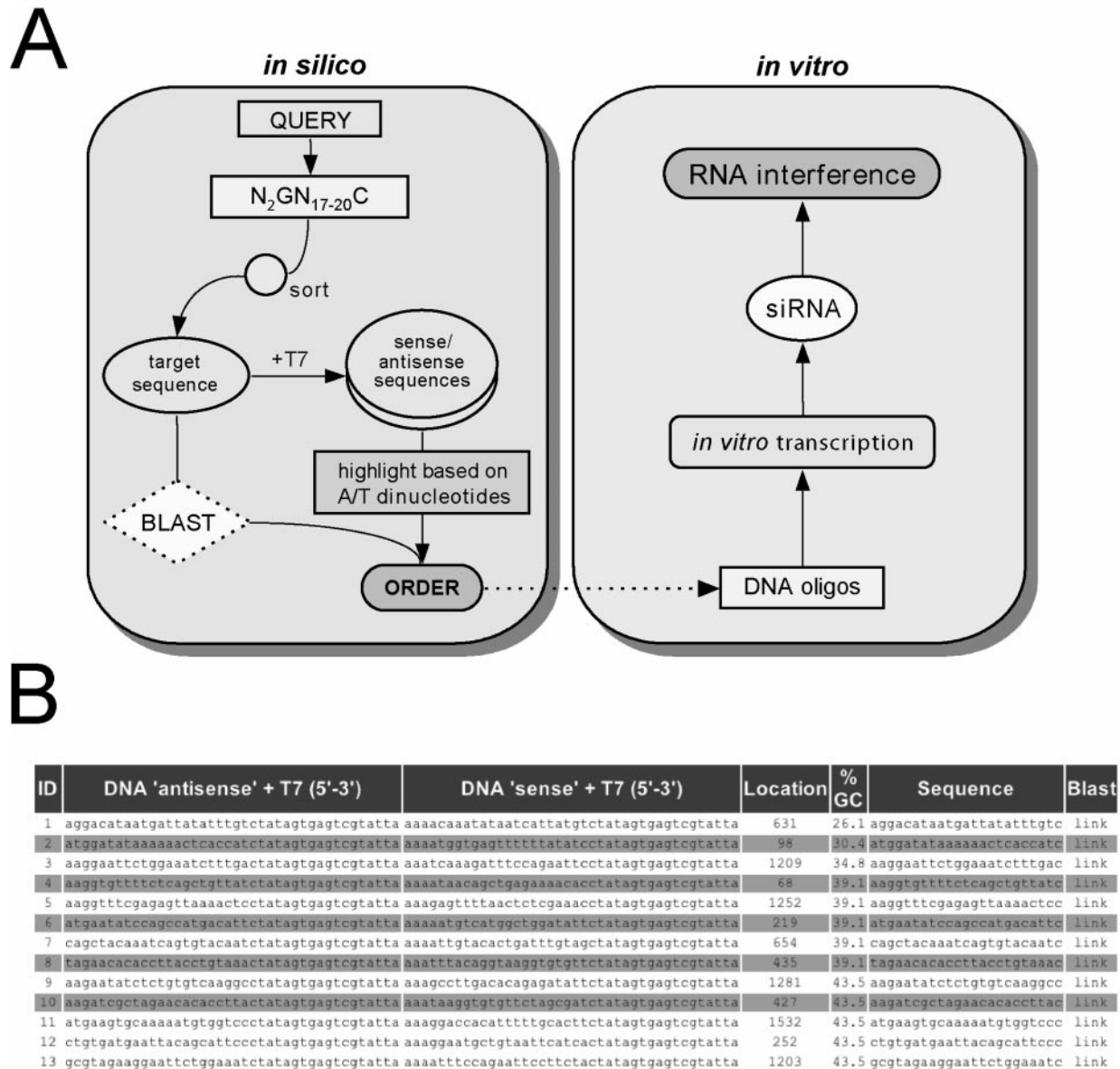
Figure 1. TROD overview. (**A**) Workflow diagram showing steps from program algorithm to experiment. (**B**) Sample program output for the human ERβ cDNA (GenBank accession no.: AB006590).

promoter sequence, which will allow the production of the RNA sense and antisense strands of the siRNA duplex. By definition, we refer to the DNA oligonucleotides that are needed to generate the sense and antisense RNA strands as 'sense' and 'antisense', respectively. Since it is the antisense RNA strand that directs target recognition, the entire transcribed portion of the DNA oligonucleotide is complementary to the target mRNA. On the other hand, since AT-richness is also preferred for the overlapping 3′ dinucleotide of the RNA sense strand, the program itself adds AA to its DNA template by default. Moreover, the program pairs the 5′ G of the antisense RNA strand with a U on the sense strand as suggested by Schwarz *et al*. (6). As a result, the base pair at the 5′ end of the antisense strand is destabilized, favouring the selection of the antisense RNA strand by the RNAi machinery (6,7). Note that the 5′ G of the antisense strand remains complementary to the C present at the 3′ end of the target sequence selected

according to the $GN_{17-20}C$ rule. An interesting reciprocal 'side-effect' of the $GN_{17-20}C$ constraint of our method is that the GC base pair at the other end of the siRNA duplex has a stabilizing effect, further biasing strand selection towards the antisense strand. The output page presents a list of ready-to-order DNA oligonucleotide sequences (5′ to 3′). The list is sorted either by GC content or by the location of the targeted sequence within the query sequence. Sequences that contain the preferred AT-rich dinucleotides at the 3′ end and adjacent to the 5′ G of the antisense siRNA strand (5′-$GW_2N_{15-18}CW_2$-3′, where W = A or T) are highlighted. Furthermore, since the target sequences are small (21–24 nt), a BLAST link is provided for each to ascertain target specificity.

One of the major drawbacks in manually designing siRNA oligos is the sheer number of possible target sequences, a matter that is complicated further by the additional $N_2GN_{17-20}C$ constraint. TROD's automated approach is ideal

for generating and sorting through these data. For example, the AT-rich protein phosphatase 2Cα2 cDNA sequence (GenBank accession no.: AF070670) contains 394 possible siRNAs with $N_2GN_{17-20}C$ sequences. Using TROD's report and the additional $W_2GN_{15-18}W_2C$ criterion, one can narrow the focus to 35 candidate siRNAs. Conversely, a similarly sized GC-rich cDNA for a protein such as MARCKS (GenBank accession no.: M68956) contains many more possible siRNAs totalling 657, but with only 11 'ideal' sequences.

FUTURE DEVELOPMENT

Since the mechanisms that confer target specificity are still not fully understood, few criteria have been identified to aid in designing siRNA sequences, which is hampering efforts to develop programs that automate this crucial step. Our program is a step towards this goal, and can be continually updated to include newly discovered selection criteria.

ACKNOWLEDGEMENTS

REFERENCES

1. Zamore,P.D., Tuschl,T., Sharp,P.A. and Bartel,D.P. (2000) RNAi: double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals. *Cell*, **101**, 25–31.
2. Donzé,O. and Picard,D. (2002) RNA interference in mammalian cells using siRNAs synthesized with T7 RNA polymerase. *Nucleic Acids Res.*, **30**, e46.
3. Donzé,O., Dudek,P. and Picard,D. (2004) siRNA production by *in vitro* transcription. In Sohail,M. (ed.), *Gene Silencing by RNA Interference: Technology and Application*. CRC Press LLC, in press.
4. Dykxhoorn,D.M., Novina,C.D. and Sharp,P.A. (2003) Killing the messenger: short RNAs that silence gene expression. *Nat. Rev. Mol. Cell. Biol.*, **4**, 457–467.
5. Elbashir,S.M., Harborth,J., Weber,K. and Tuschl,T. (2002) Analysis of gene function in somatic mammalian cells using small interfering RNAs. *Methods*, **26**, 199–213.
6. Schwarz,D.S., Hutvágner,G., Du,T., Xu,Z., Aronin,N. and Zamore,P.D. (2003) Asymmetry in the assembly of the RNAi enzyme complex. *Cell*, **115**, 199–208.
7. Khvorova,A., Reynolds,A. and Jarasena,S.D. (2003) Functional siRNAs and miRNAs exhibit strand bias. *Cell*, **115**, 209–216.
8. Milligan,J.F., Groebe,D.R., Witherell,G.W. and Uhlenbeck,O.C. (1987) Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates, *Nucleic Acids Res.*, **15**, 8783–8798.
9. Stajich,J.E., Block,D., Boulez,K., Brenner,S.E., Chervitz,S.A., Dagdigian,C., Fuellen,G., Gilbert,J.G.R., Korf,I., Lapp,H. *et al.* (2002) The Bioperl toolkit: Perl modules for the life sciences. *Genome Res.*, **12**, 1611–1618.
10. Kumar,R., Conklin,D.S. and Mittal,V. (2003) High-throughput selection of effective RNAi probes for gene silencing, *Genome Res.*, **13**, 2333–2340.
11. Paddison P.J., Caudy A.A., Bernstein,E., Hannon, G.J. and Conklin,D.S. (2002) Short hairpin RNAs (shRNAs) induce sequence specific silencing in mammalian cells. *Genes Dev.*, **16**, 948–958.