W448–W450 *Nucleic Acids Research, 2006, Vol. 34, Web Server issue*
doi:10.1093/nar/gkl214

siVirus: web-based antiviral siRNA design software for highly divergent viral sequences

Yuki Naito[1,*], Kumiko Ui-Tei[1,2], Toru Nishikawa[3], Yutaka Takebe[4] and Kaoru Saigo[1]

[1]Department of Biophysics and Biochemistry, Graduate School of Science, [2]Undergraduate Program for Bioinformatics and Systems Biology, School of Science and, [3]Department of Computer Science, Graduate School of Information Science and Technology, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan and [4]Laboratory of Molecular Virology and Epidemiology, AIDS Research Center, National Institute of Infectious Diseases, 1-23-1 Toyama, Shinjuku-ku, Tokyo 162-8640, Japan

Received February 14, 2006; Revised and Accepted March 24, 2006

ABSTRACT

siVirus (http://siVirus.RNAi.jp/) is a web-based online software system that provides efficient short interfering RNA (siRNA) design for antiviral RNA interference (RNAi). siVirus searches for functional, off-target minimized siRNAs targeting highly conserved regions of divergent viral sequences. These siRNAs are expected to resist viral mutational escape, since their highly conserved targets likely contain structurally/functionally constrained elements. siVirus will be a useful tool for designing optimal siRNAs targeting highly divergent pathogens, including human immunodeficiency virus (HIV), hepatitis C virus (HCV), influenza virus and SARS coronavirus, all of which pose enormous threats to global human health.

INTRODUCTION

RNA interference (RNAi) is now widely used to knockdown gene expression in a sequence-specific manner, making it a powerful tool not only for studying gene function, but also for therapeutic purposes, including antiviral treatments (1–4). Currently, the replication of a wide range of viruses can be inhibited successfully using RNAi, with both short interfering RNAs (siRNAs) and siRNA expression vectors (5).

In mammalian RNAi, the efficacy of each siRNA varies widely depending on its sequence; only a limited fraction of randomly designed siRNAs is highly effective. Many experiments have been conducted to clarify possible sequence requirements of functional siRNAs. Of these, our work incorporates guidelines from three major studies (6–8) of selecting functional siRNAs. However, designing functional siRNAs that target viral sequences is problematic because of their extraordinarily high genetic diversity. For example, about 500 entries of near full-length sequences of HIV-1 group M, which is largely responsible for global pandemic, are stored in the sequence databases, but it proved impossible to select a common 21mer from among all of them. Moreover, RNAi-resistant viral mutants achieved through point mutation or deletion emerge rapidly when targeting viruses in cell culture. These problems suggest a strong need to select highly conserved target sites for designing antiviral siRNAs. Furthermore, the off-target silencing effects of siRNA are also a serious problem that could affect host gene expression (9). Off-target silencing effects arise when an siRNA has sequence similarities with unrelated genes. In antiviral RNAi, it is desirable to minimize off-target effects against human genes.

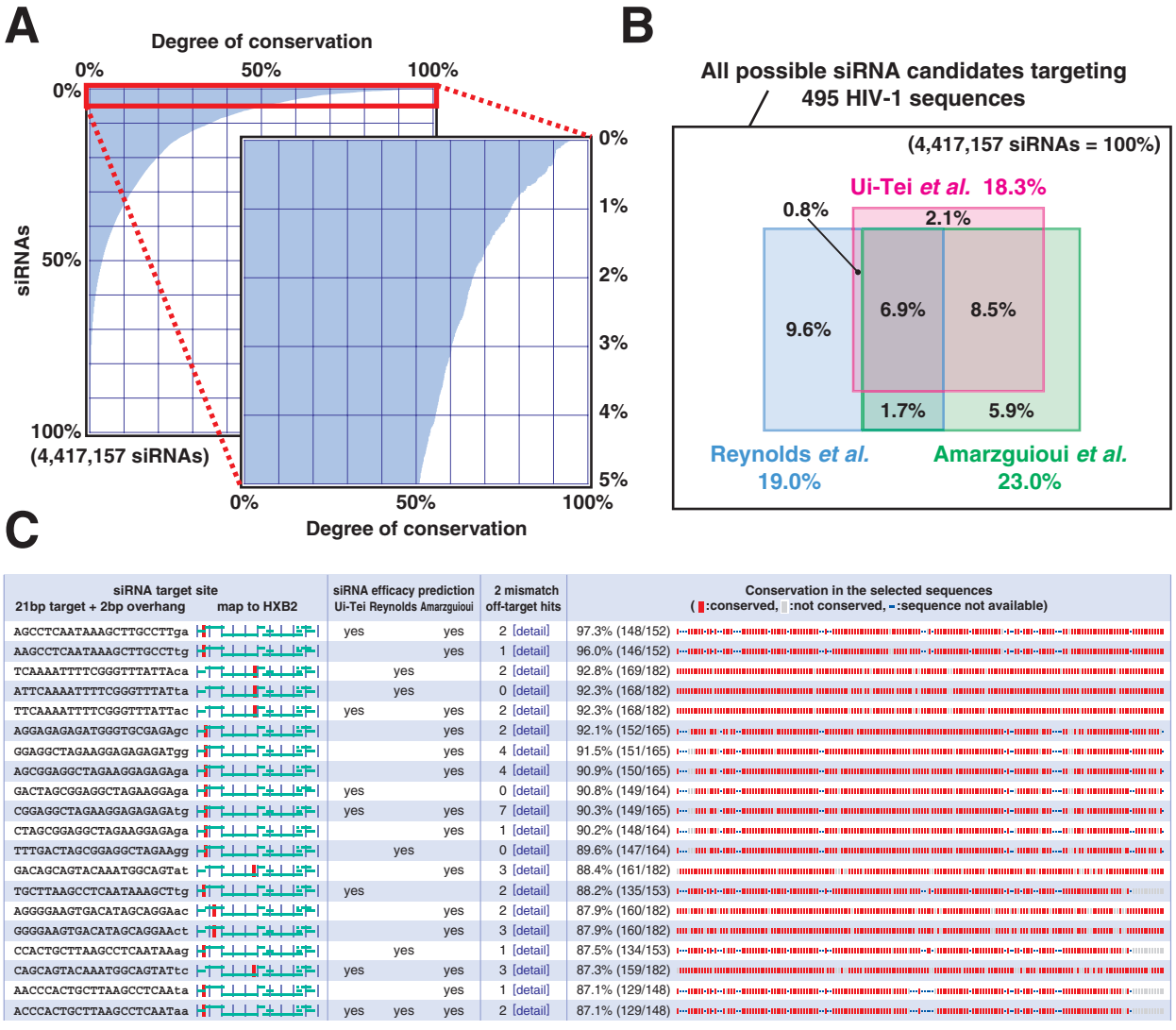
Consequently, only a limited fraction of 21mers is suitable for use as antiviral siRNAs. In this study, we developed a novel web-based online software system, siVirus, which provides functional, off-target minimized siRNAs targeting highly conserved regions of divergent viral sequences.

METHODS

Selection of highly conserved siRNA target sites

Highly conserved siRNA sequences are selected based on their *degree of conservation*, defined as the proportion of viral sequences that are targeted by the corresponding siRNA, with complete matches (i.e. 21/21 matches). All possible siRNA candidates targeting every other position of user-selected viral sequences are generated and their degrees of conservation are computed. Users can arbitrarily specify a set of viral sequences for the computation; e.g. sequences can be selected from a specific geographic region(s) or a specific genotype(s) to design the best siRNAs tailored to specific user needs. siVirus also accepts user's own sequences in a multi-FASTA format and shows whether each siRNA can target the posted sequences.

*To whom correspondence should be addressed. Tel: +81 3 5841 4404; Fax: +81 3 5841 4400; Email: y-naito@RNAi.jp



siRNA target site 21bp target + 2bp overhang	map to HXB2	siRNA efficacy prediction Ui-Tei	Reynolds	Amarzguioui	2 mismatch off-target hits	Conservation in the selected sequences (▮:conserved, ▯:not conserved, -:sequence not available)
AGCCTCAATAAAGCTTGCCTTga		yes		yes	2 [detail]	97.3% (148/152)
AAGCCTCAATAAAGCTTGCCTtg				yes	1 [detail]	96.0% (146/152)
TCAAAATTTTCGGGTTTATTAca			yes		2 [detail]	92.8% (169/182)
ATTCAAAATTTTCGGGTTTATta			yes		0 [detail]	92.3% (168/182)
TTCAAAATTTTCGGGTTTATTac		yes		yes	2 [detail]	92.3% (168/182)
AGGAGAGAGATGGGTGCGAGAgc				yes	2 [detail]	92.1% (152/165)
GGAGGCTAGAAGGAGAGAGATgg				yes	4 [detail]	91.5% (151/165)
AGCGGAGGCTAGAAGGAGAGAga				yes	4 [detail]	90.9% (150/165)
GACTAGCGGAGGCTAGAAGGAga		yes			0 [detail]	90.8% (149/164)
CGGAGGCTAGAAGGAGAGAGAtg		yes		yes	7 [detail]	90.3% (149/165)
CTAGCGGAGGCTAGAAGGAGAga				yes	1 [detail]	90.2% (148/164)
TTTGACTAGCGGAGGCTAGAAgg			yes		0 [detail]	89.6% (147/164)
GACAGCAGTACAAATGGCAGTat				yes	3 [detail]	88.4% (161/182)
TGCTTAAGCCTCAATAAAGCTtg		yes			2 [detail]	88.2% (135/153)
AGGGGAAGTGACATAGCAGGAac				yes	2 [detail]	87.9% (160/182)
GGGGAAGTGACATAGCAGGAAct				yes	3 [detail]	87.9% (160/182)
CCACTGCTTAAGCCTCAATAAag			yes		1 [detail]	87.5% (134/153)
CAGCAGTACAAATGGCAGTATtc		yes		yes	3 [detail]	87.3% (159/182)
AACCCACTGCTTAAGCCTCAAta				yes	1 [detail]	87.1% (129/148)
ACCCACTGCTTAAGCCTCAATaa		yes	yes	yes	2 [detail]	87.1% (129/148)

Figure 1. (**A**) The degree of conservation is calculated for all possible siRNA candidates (total 4 417 157) targeting every other position of 495 HIV-1 sequences. (**B**) The efficacy predictions of these 4 417 157 siRNA candidates based on three different guidelines: Ui-Tei *et al*. (6), Reynolds *et al*. (7) and Amarzguioui *et al*. (8). (**C**) Typical output of siVirus for designing anti-HIV siRNAs. Sequence information, efficacy predictions, off-target search results and the degrees of conservation are shown.

siRNA efficacy prediction

In mammalian RNAi, the efficacy of each siRNA varies markedly depending on its sequence; hence, several groups have reported guidelines for selecting functional siRNAs. siVirus incorporates the guidelines of Ui-Tei *et al*. (6), Reynolds *et al*. (7) and Amarzguioui *et al*. (8) and shows whether each siRNA satisfies these guidelines.

Off-target searches

Off-target searches were performed for each siRNA using siDirect (10,11). siVirus shows the number of off-target hits within two mismatches against the non-redundant database of human transcripts (10).

Database maintenance

Currently, siVirus incorporates viral genome sequences of HIV-1, HCV, influenza A virus and SARS coronavirus. These sequences were downloaded from the Los Alamos HIV Sequence Database (http://hiv-web.lanl.gov/), the Los Alamos HCV Sequence Database (12), the NCBI Influenza Virus Sequence Database (http://www.ncbi.nlm.nih.gov/genomes/FLU/FLU.html), and NCBI GenBank (13), respectively. siVirus will be updated continuously as these databases are revised. We also plan to incorporate other viruses if sufficient numbers of their sequences are available.

RESULTS AND DISCUSSION

To design anti-HIV siRNA, we analyzed the 495 near full-length HIV-1 sequences listed in Supplementary Table 1. A total of 4 417 157 possible siRNA candidates (i.e. substrings of length 21) targeting every other position of the HIV-1 sequences were produced from the 495 viral sequences. The analysis of these siRNA candidates revealed that highly

conserved siRNAs constituted only 0.3% of the possible siRNAs if >90% conservation is expected (Figure 1A). The fraction is still as small as 0.8% even if the threshold of the conservation is relaxed to 80%. On the other hand, siRNAs predicted to be functional by one or more guidelines (6–8) constituted 35.5% of the 4 417 157 siRNAs (Figure 1B). Taken together, siRNAs that are >80% conserved, and satisfy at least one guideline constitute only 0.2% of the siRNAs. In this condition, 20–30 siRNAs can be designed for each full-length sequence of HIV-1. These indicate that most of the randomly designed siRNAs are not suited for targeting HIV-1 efficiently.

Figure 1C shows typical output from siVirus for designing anti-HIV siRNAs. A total of 182 sequences from HIV-1 subtypes B, C and CRF01_AE, which are the most prevalent HIV-1 genotypes circulating in Asia, were selected. The results were sorted by their degree of conservation, and filtered to display siRNAs that satisfy at least one efficacy guideline. The off-target search results against human genes are also shown. It is desirable to select an siRNA that has less off-target hits.

To test the validity of siVirus, 35 siRNAs satisfying the guideline by Ui-Tei *et al.* (6) were designed against the conserved regions of HIV-1 genomes using siVirus and were assayed for inhibition of viral replication. Among them, 31 siRNAs effectively inhibited HIV-1 replication by >80% when each siRNA duplex was transfected at 5 nM (Y. Naito, K. Ui-Tei, K. Saigo and Y. Takebe, unpublished data).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

This work was supported in part by grants from the Ministry of Education, Culture, Sports, Science and Technology of Japan to K.S., K.U.-T. and Y.T., and by grants from the Ministry of Health, Labour and Welfare of Japan to Y.T. Funding to pay the Open Access publication charges for this article was provided by the Ministry of Education, Culture, Sports, Science and Technology of Japan. Y.N. is a Research Fellow of the Japan Society for the Promotion of Science.

Conflict of interest statement. None declared.

REFERENCES

1. Fire,A., Xu,S., Montgomery,M.K., Kostas,S.A., Driver,S.E. and Mello,C.C. (1998) Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature*, **391**, 806–811.
2. Mello,C.C. and Conte,D. Jr (2004) Revealing the world of RNA interference. *Nature*, **431**, 338–342.
3. Hannon,G.J. and Rossi,J.J. (2004) Unlocking the potential of the human genome with RNA interference. *Nature*, **431**, 371–378.
4. Voinnet,O. (2005) Induction and suppression of RNA silencing: insights from viral infections. *Nature Rev. Genet*., **6**, 206–220.
5. Leonard,J.N. and Schaffer,D.V. (2006) Antiviral RNAi therapy: emerging approaches for hitting a moving target. *Gene Ther*., **13**, 532–540.
6. Ui-Tei,K., Naito,Y., Takahashi,F., Haraguchi,T., Ohki-Hamazaki,H., Juni,A., Ueda,R. and Saigo,K. (2004) Guidelines for the selection of highly effective siRNA sequences for mammalian and chick RNA interference. *Nucleic Acids Res*., **32**, 936–948.
7. Reynolds,A., Leake,D., Boese,Q., Scaringe,S., Marshall,W.S. and Khvorova,A. (2004) Rational siRNA design for RNA interference. *Nat. Biotechnol*., **22**, 326–330.
8. Amarzguioui,M. and Prydz,H. (2004) An algorithm for selection of functional siRNA sequences. *Biochem. Biophys. Res. Commun*., **316**, 1050–1058.
9. Jackson,A.L. and Linsley,P.S. (2004) Noise amidst the silence: off-target effects of siRNAs? *Trends Genet*., **20**, 521–524.
10. Naito,Y., Yamada,T., Ui-Tei,K., Morishita,S. and Saigo,K. (2004) siDirect: highly effective, target-specific siRNA design software for mammalian RNA interference. *Nucleic Acids Res*., **32**, W124–W129.
11. Yamada,T. and Morishita,S. (2005) Accelerated off-target search algorithm for siRNA. *Bioinformatics*, **21**, 1316–1324.
12. Kuiken,C., Yusim,K., Boykin,L. and Richardson,R. (2005) The Los Alamos hepatitis C sequence database. *Bioinformatics*, **21**, 379–384.
13. Benson,D.A., Karsch-Mizrachi,I., Lipman,D.J., Ostell,J. and Wheeler,D.L. (2006) GenBank. *Nucleic Acids Res*., **34**, D16–D20.