In: RNA Interference Research Progress
Editors: R.T. Lyland and I.B. Browning, pp. 3-11
ISBN: 978-1-60456-446-4
© 2008 Nova Science Publishers, Inc.

Evaluation of Published Rational siRNA Design Algorithms Using Firefly *luciferase* Gene as a Reporter

Yuki Naito, Kaoru Saigo and Kumiko Ui-Tei*
Department of Biophysics and Biochemistry,
Graduate School of Science, University of Tokyo,
7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan

Abstract

In mammalian RNA interference (RNAi), a limited fraction of the short interfering RNA (siRNA) is capable of successful interference, since the efficiency of RNAi depends highly on the siRNA sequence used. Several different algorithms for selecting functional siRNA sequences have been published. To find a robust siRNA selection procedure by integrating these algorithms, we performed validation studies on some published rational siRNA design algorithms. In this study, we used a single reporter assay with the *firefly luciferase* gene, since it is likely that the efficiencies of siRNA sequences vary between target genes, partially depending on the detection sensitivity of the expression of each gene. The results suggest that it is best to select siRNAs that generally satisfy a combination of some reliable integrated algorithms. Our results should be considered when conducting RNAi experiments.

Introduction

RNA interference (RNAi) is used in loss-of-function genetics to assess the functions of genes involved in various biological processes, and the therapeutic applications of RNAi are currently being investigated [1]. RNAi is an evolutionarily conserved process by which specific genes are silenced using short interfering RNA (siRNA) molecules [2]. In mammalian cells, the extent of RNAi depends on the

* E-mail address: ktei@biochem.s.u-tokyo.ac.jp. Address correspondence to Dr. Kumiko Ui-Tei, Department of Biophysics and Biochemistry, Graduate School of Science, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan

sequence being used; only a fraction of the possible siRNAs is capable of successful interference [3, 4]. Although several databases and other sources offer experimentally validated siRNAs, the number of available target genes is limited; thus, it is difficult to use these siRNAs for genome-wide experiments. The best way to select well-authenticated functional siRNAs is to apply reliable published integrated algorithms based on experimental evidence. For this purpose, we performed validation studies using the firefly luciferase gene as a reporter.

Results and Discussion

Distribution of siRNAs Predicted to be Functional by the Rational siRNA Design Algorithms in Human Genome

Among the published algorithms for selecting siRNA sequences, at first three algorithms I [4], II [5], and III [6] were compared because they are based on data from biological analyses. In algorithm I, the parameters for selecting effective siRNA sequences are as follows: A/U at the 5'-end of the guide strand, G/C at the 5'-end of the passenger strand, at least four A/U residues in the 5'-terminal third of the guide strand, and the absence of any long GC stretches. In Algorithm I, siRNAs simultaneously satisfying all of these conditions are defined as functional. In algorithm II, the following 8 characteristics have been shown to be associated with siRNA functionality: low G/C content, low internal stability at the 3'-terminus of the passenger strand, and the absence of inverted repeats. Furthermore, the passenger strand should contain A, U, and A at positions 3, 10, and 19, respectively, from the 5'-end of the passenger strand. Additionally, the 5'-terminus of the guide strand should not be G/C, and G should not be present at position 13 from the 5'-end of the passenger strand. In this algorithm, each parameter was scored (see Table 1) and siRNAs having a score ≥ 6 are determined as functional. Algorithm III recommended that the A/U content of the three terminal nucleotides is different between the ends of the duplex, and the 5'-terminus of the guide strand and its passenger strand partner should be A/U, while the 5'-terminus of the passenger strand and its guide strand partner should be G/C. However, U and G each at position 1 and 19 of the passenger strand should be avoided. Moreover, functional siRNAs contain A at position 6. In this algorithm, functional siRNAs are also determined by algorithm-derived score (see Table 1).

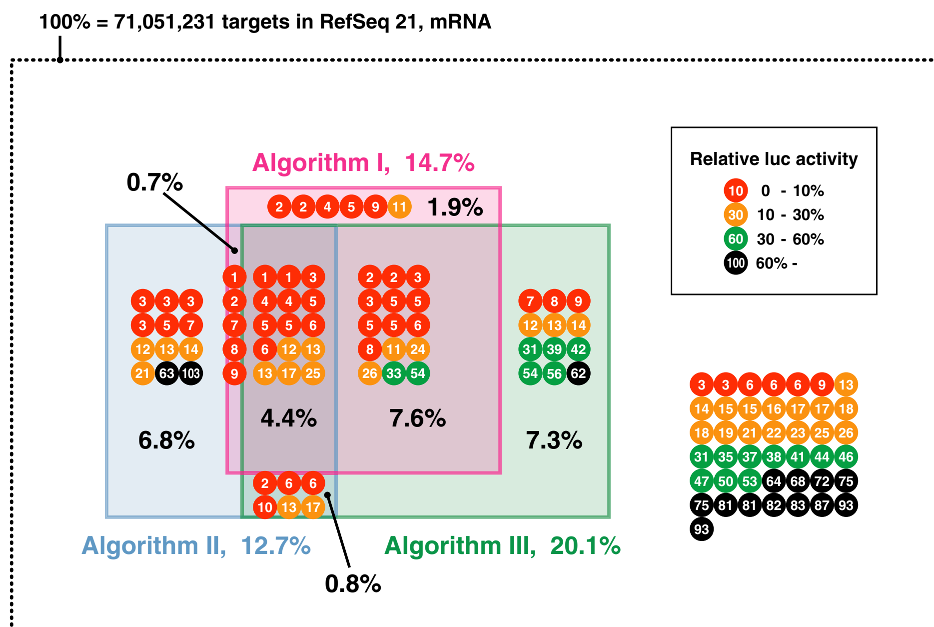

Figure 1. The distribution of human siRNA sequences and their RNAi activities against the firefly *luc* gene. In total, 71,051,231 human sequences, each 21 bp in length, extracted from the RefSeq mRNA database (version 21) were chosen as siRNA candidates and classified using algorithms I, II, and III. The percentages indicate the proportions of the sequences that were classified into each compartment. Red, orange, green, and black circles indicate the relative levels of luc activity (0–10, 10–30, 30-60, and 60–100%, respectively). The numbers inside each circle indicate the actual values.

All human siRNA sequences (21 bp each, 7.1×10^7 total) were extracted from the RefSeq database (version 21) by dividing into pieces of 21 bp and classified using the three algorithms. Their correlation is presented in Figure 1. The percentages of functional siRNA sequences satisfying algorithms I, II, and III were 14.7, 12.7, and 20.1 %, respectively. Nearly 90 % of the siRNAs satisfying algorithm I were predicted to be functional by algorithm II or III, or both. About 85 % of the siRNAs simultaneously predicted to be functional by algorithms II and III were also predicted to be functional by algorithm I. More than 50 % of the siRNAs labeled as functional by algorithm II were not predicted to be functional by algorithm III, and 74 % of the functional siRNAs from algorithm III were not predicted to be functional by algorithm II. For genome-wide experiments, it is desirable that functional siRNA sequences can be provided for as many genes as possible. Our computational analyses indicate that at least an siRNA sequence predicted to be functional by algorithm I, II, and III, respectively, can be designed for all of human genes (Table 2). Although all of the siRNAs determined to be functional by both of algorithms I and II, II and III, or all of three algorithms reduced luc activity equal to or less than 33 %, functional siRNA sequences were not able to be selected for 6, 6 or 8 genes in all of 24,261 transcripts registered in RefSeq database. The results indicated that about 8 millions of siRNA sequences (12 % of total 21 bp sequences) is necessary for selecting at least one siRNA sequence against all of human genes. Therefore, it may be advantageous to use combined algorithms covering as many sequences as possible in some circumstances, especially when more than one siRNA are recommended for each gene.

Table 1. Classification of siRNA sequences for firefly luc gene.

Serial No.	siRNA No.[1)]	Target sequence[2)]	Relative luc activity (%)	Predicted functionality[3)]							Algorithm I					Algorithm II								Algorithm III							Algorithm IV					Algorithm V	Algorithm VI	Algorithm VII
				I	II	III	IV	V	VI	VII	i	ii	iii	iv	Total	I	II	IV	V	VI	VII	VIII	Total	1	2	3	4	5	6	Total	1	2	3	4	Total	Score	Score	Result
1	5**	aagacgccaaaaacataaagaaa	17	F	F	F	F	F	NF	F	1	1	1	1	4	1	4	1	0	0	0	0	6	1	1	0	0	0	1	3	1	1	1	1	4	79.1	5	F
2	6*	agacgccaaaaacataaagaaag	7	NF	NF	F	NF	F	NF	NF	1	0	1	1	3	0	4	1	0	0	0	0	5	1	0	0	1	0	1	3	1	1	0	1	3	81.7	4	NF
3	12*	caaaaacataaagaaaggcccgg	93	NF	NF	NF	NF	NF	NF	NF	0	0	0	1	1	1	0	0	1	0	-1	0	1	-3	0	0	1	0	0	-2	1	0	0	0	1	55.5	3	NF
4	14**	aaaacataaagaaaggcccggcg	81	NF	NF	NF	NF	NF	NF	F	0	0	0	1	1	1	0	0	0	0	-1	-1	-1	-2	0	0	1	-1	0	-2	1	0	0	0	1	48.0	3	F
5	18*	cataaagaaaggcccggcgccat	93	NF	NF	NF	NF	NF	NF	NF	0	0	0	0	0	0	0	0	1	0	-1	0	0	-3	0	-1	1	0	0	-2	0	0	0	0	0	40.4	0	NF
6	24*,***	gaaaggcccggcgccattctatc	83	NF	NF	NF	NF	NF	NF	NF	1	0	1	0	2	0	4	1	0	0	0	0	5	0	0	0	0	0	1	1	0	1	0	1	2	65.7	3	NF
7	29***	gcccggcgccattctatccgctg	47	NF	NF	NF	NF	NF	NF	NF	0	1	0	1	2	0	1	0	0	1	-1	0	1	0	1	0	0	0	0	1	1	0	1	1	3	52.9	3	NF
8	30*,**,***	cccggcgccattctatccgctgg	54	NF	NF	F	F	NF	NF	NF	1	1	0	1	3	0	1	0	0	1	0	0	2	1	1	0	0	0	1	3	1	1	1	1	4	69.8	3	NF
9	31***	ccggcgccattctatccgctgga	22	NF	NF	NF	NF	NF	NF	NF	0	1	0	1	2	0	1	0	0	0	-1	0	0	1	1	0	0	-1	0	1	0	0	1	0	1	67.3	2	NF
10	33**	ggcgccattctatccgctggaag	14	NF	NF	F	NF	NF	NF	NF	1	1	0	1	3	0	2	1	0	0	0	0	3	1	1	0	0	0	1	3	1	1	1	0	3	70.5	2	NF
11	34**	gcgccattctatccgctggaaga	31	NF	NF	F	F	F	NF	NF	1	1	0	1	3	1	3	1	0	1	0	-1	5	2	1	0	0	0	1	4	1	1	1	1	4	76.8	3	NF
12	35***	cgccattctatccgctggaagat	18	NF	NF	NF	NF	NF	NF	NF	0	1	0	1	2	1	2	0	1	0	-1	0	3	1	1	0	0	-1	0	1	0	0	1	1	2	65.2	3	NF
13	36*,**,***	gccattctatccgctggaagatg	11	F	NF	NF	NF	F	NF	NF	1	1	1	1	4	1	3	1	0	0	0	0	5	0	1	0	0	0	1	2	0	1	1	1	3	80.3	3	NF
14	37***	ccattctatccgctggaagatgg	44	NF	NF	NF	NF	NF	NF	NF	1	0	1	1	3	1	4	0	0	0	0	-1	4	-1	0	0	1	0	1	1	0	1	0	1	2	71.1	2	NF
15	40**	ttctatccgctggaagatggaac	9	F	F	NF	NF	F	NF	NF	1	1	1	1	4	1	3	1	1	0	0	0	6	-1	1	0	0	0	1	1	0	1	1	1	3	77.9	2	NF
16	42*,***	ctatccgctggaagatggaaccg	87	NF	NF	NF	NF	NF	NF	NF	0	0	1	1	2	1	2	0	0	0	-1	0	2	0	0	0	0	0	0	0	1	0	0	1	2	55.7	3	NF
17	47***	cgctggaagatggaaccgctgga	17	NF	NF	NF	NF	NF	NF	NF	0	1	0	1	2	0	1	0	0	0	-1	0	0	0	1	0	1	-1	0	1	0	0	1	0	1	66.6	3	NF
18	48*,***	gctggaagatggaaccgctggag	81	NF	NF	NF	NF	NF	NF	NF	0	0	0	1	1	0	1	0	0	0	-1	0	0	0	0	-1	0	-1	0	-1	0	0	0	0	0	51.8	2	NF
19	49***	ctggaagatggaaccgctggaga	8	NF	NF	F	NF	F	NF	NF	1	1	0	1	3	0	2	1	1	0	0	0	4	0	1	0	1	0	1	3	1	1	1	0	3	88.1	3	NF
20	53***	aagatggaaccgctggagagcaa	17	NF	NF	NF	NF	NF	NF	F	0	1	0	1	2	0	2	0	0	0	-1	-1	0	-1	1	0	1	0	0	1	0	0	1	0	1	60.3	1	F
21	54*,***	agatggaaccgctggagagcaac	64	NF	NF	NF	NF	NF	NF	NF	1	0	0	1	2	0	2	1	0	0	0	-1	2	-1	0	0	1	0	1	1	0	1	0	1	2	63.3	1	NF
22	55***	gatggaaccgctggagagcaact	9	NF	NF	NF	NF	NF	NF	NF	1	0	1	1	3	0	3	1	0	1	0	0	5	1	0	-1	0	0	1	2	1	1	0	1	3	74.2	3	NF
23	179**	ccgttcggttggcagaagctatg	5	F	NF	NF	NF	F	NF	NF	1	1	1	1	4	1	3	1	0	0	0	-1	4	0	1	0	0	0	1	2	0	1	1	1	3	92.1	3	NF
24	184**	cggttggcagaagctatgaaacg	6	F	F	F	F	F	NF	NF	1	1	1	1	4	1	4	1	0	0	0	0	6	1	1	0	0	0	1	3	1	1	1	1	4	89.2	5	NF
25	257**	ttcaattctttatgccggtgttg	53	NF	NF	NF	NF	NF	NF	NF	1	1	0	1	3	1	2	0	1	0	0	0	4	0	1	0	0	0	1	2	1	1	1	0	3	69.8	3	NF
26	265**	tttatgccggtgttgggcgcgtt	75	NF	NF	NF	NF	NF	NF	NF	0	0	0	1	1	0	0	0	0	0	-1	-1	-2	-3	0	-1	0	-1	0	-4	0	0	0	0	0	36.9	2	NF
27	272**	cggtgttgggcgcgttatttatc	13	F	F	F	NF	F	NF	NF	1	1	1	1	4	1	5	1	0	0	0	0	7	2	1	0	0	0	1	4	0	1	1	1	3	84.4	4	NF
28	295**	ggagttgcagttgcgcccgcgaa	68	NF	NF	NF	NF	NF	NF	NF	0	0	0	1	1	0	0	0	0	1	-1	-1	-1	-2	0	0	0	-1	0	-3	1	0	0	0	1	48.7	2	NF
29	296**	gagttgcagttgcgcccgcgaac	23	NF	NF	NF	NF	F	NF	NF	1	1	0	1	3	0	1	1	0	0	0	0	2	-1	1	0	1	0	1	2	0	1	1	0	2	75.1	2	NF
30	309**	gcccgcgaacgacatttataatg	6	F	F	F	F	F	NF	NF	1	1	1	1	4	1	5	1	0	0	0	0	7	3	1	0	1	0	1	6	1	1	1	1	4	87.7	5	NF
31	329	atgaacgtgaattgctcaacagt	7	F	F	NF	F	F	NF	NF	1	1	1	1	4	1	3	1	1	1	0	0	7	0	1	0	0	0	1	2	1	1	1	1	4	86.3	5	NF
32	370**	accgtggtgttcgtttccaaaaa	3	F	NF	F	NF	F	NF	NF	1	1	1	1	4	1	3	1	0	0	0	0	5	2	1	0	0	0	1	4	0	1	1	1	3	96.5	5	NF
33	387	caaaaaggggttgcaaaaaattt	5	NF	F	NF	NF	F	NF	NF	1	0	1	1	3	0	5	0	1	1	0	0	7	0	0	0	0	0	1	1	1	1	0	1	3	80.5	4	NF
34	411	gaacgtgcaaaaaaagctcccaa	72	NF	NF	NF	NF	NF	NF	NF	0	0	0	1	1	1	1	0	0	0	-1	0	1	-1	0	0	0	0	0	-1	1	0	0	0	1	51.6	3	NF
35	412	aacgtgcaaaaaaagctcccaat	9	NF	NF	F	NF	F	NF	F	1	1	0	1	3	1	2	1	0	0	0	-1	3	0	1	0	1	0	1	3	1	1	1	0	3	81.5	3	F
36	421	aaaaagctcccaatcatccaaaa	21	NF	F	NF	NF	F	NF	F	1	0	1	1	3	1	3	1	1	0	0	0	6	-1	0	0	0	0	1	0	1	1	0	1	3	83.7	4	F
37	423	aaagctcccaatcatccaaaaaa	17	NF	F	F	NF	F	F	F	1	0	1	1	3	1	4	1	0	1	0	0	7	2	0	0	0	0	1	3	1	1	0	1	3	88.0	6	F
38	428**,***	tcccaatcatccaaaaaattatt	4	F	F	F	NF	F	NF	NF	1	1	1	1	4	0	5	1	1	0	0	0	7	2	1	0	0	0	1	4	0	1	1	1	3	88.4	5	NF
39	431	caatcatccaaaaaattattatc	63	NF	F	NF	NF	F	NF	NF	1	0	1	1	3	0	5	1	0	0	0	0	6	1	0	0	0	0	1	2	1	1	0	1	3	75.5	4	NF
40	507**	atctcatctacctcccggtttta	13	NF	NF	F	NF	NF	F	NF	1	1	0	1	3	1	3	0	0	0	0	0	4	2	1	0	0	0	1	4	0	1	1	1	3	70.8	6	NF
41	515**	tacctcccggttttaatgaatac	25	F	F	F	NF	F	NF	NF	1	1	1	1	4	1	4	0	0	1	0	0	6	2	1	0	0	0	1	4	1	1	1	1	4	80.6	4	NF
42	575	ttgcactgatcatgaactcctct	9	F	NF	NF	F	F	NF	NF	1	1	1	1	4	1	2	0	1	0	0	0	4	0	1	0	0	0	1	2	1	1	1	1	4	78.2	4	NF
43	583	atcatgaactcctctggatctac	26	F	NF	F	NF	NF	F	NF	1	1	1	1	4	1	3	0	0	0	0	0	4	0	1	0	1	0	1	3	0	1	1	1	3	73.2	6	NF
44	630**	gcctcatagaactgcctgcgtga	39	NF	NF	F	NF	NF	NF	NF	1	1	0	1	3	1	2	0	0	0	0	0	3	0	1	0	1	0	1	3	0	1	1	0	2	62.8	4	NF
45	655	ttctcgcatgccagagatcctat	5	F	NF	F	NF	NF	NF	NF	1	1	1	1	4	0	3	0	0	0	0	0	3	0	1	0	1	0	1	3	0	1	1	1	3	72.6	3	NF
46	658**	tcgcatgccagagatcctatttt	5	F	F	F	NF	F	F	NF	1	1	1	1	4	1	4	0	1	0	0	0	6	2	1	0	0	0	1	4	1	1	1	1	4	85.4	6	NF
47	694	attccggatactgcgattttaag	2	NF	F	F	NF	F	NF	NF	1	0	1	1	3	1	5	1	0	1	0	-1	7	2	0	-1	1	0	1	4	1	1	0	1	3	85.3	5	NF
48	695**	ttccggatactgcgattttaagt	5	F	F	F	NF	F	NF	NF	1	1	1	1	4	1	5	1	0	0	0	0	7	3	1	0	0	0	1	5	0	1	1	1	3	93.4	4	NF
49	721	gttccattccatcacggttttgg	12	NF	F	NF	NF	NF	NF	NF	1	0	1	1	3	1	4	0	0	1	0	0	6	2	0	-1	0	0	1	3	1	1	0	1	3	65.7	5	NF
50	734**	acggttttggaatgtttactaca	6	F	NF	F	F	F	NF	NF	1	1	1	1	4	0	4	1	0	0	0	0	5	1	1	0	0	0	1	3	1	1	1	1	4	97.6	5	NF
51	749	ttactacactcggatatttgata	3	NF	F	NF	NF	F	F	NF	1	0	1	1	3	1	4	1	0	0	0	0	6	0	0	0	1	0	1	2	0	1	0	1	2	84.6	6	NF
52	750	tactacactcggatatttgatat	8	F	F	NF	NF	F	NF	NF	1	1	1	1	4	1	4	0	1	0	0	0	6	0	1	0	0	0	1	2	0	1	1	1	3	78.3	5	NF
53	758**	tcggatatttgatatgtggattt	2	F	NF	F	F	F	F	NF	1	1	1	1	4	0	3	0	1	0	0	0	4	1	1	0	0	0	1	3	1	1	1	1	4	91.2	6	NF
54	765***	tttgatatgtggatttcgagtcg	19	NF	NF	NF	NF	NF	NF	NF	1	0	1	1	3	1	2	0	1	0	0	0	4	0	0	-1	0	0	1	1	0	1	0	1	2	72.4	4	NF
55	766***	ttgatatgtggatttcgagtcgt	50	NF	NF	NF	NF	NF	NF	NF	0	1	0	1	2	1	2	0	0	0	-1	0	2	-1	1	0	0	0	0	0	1	0	1	1	3	71.3	4	NF
56	767***	tgatatgtggatttcgagtcgtc	38	NF	NF	NF	NF	NF	NF	NF	0	0	0	1	1	1	2	0	1	1	-1	0	4	-2	0	0	0	-1	0	-3	1	0	0	0	1	61.9	3	NF
57	768***	gatatgtggatttcgagtcgtct	25	NF	NF	NF	NF	NF	NF	NF	1	0	0	1	2	1	2	0	0	1	0	-1	3	-2	0	-1	0	0	1	-1	1	1	0	1	3	61.9	4	NF
58	769***	atatgtggatttcgagtcgtctt	14	NF	NF	NF	NF	NF	NF	NF	0	0	0	1	1	1	2	0	0	1	-1	0	3	-1	0	0	0	0	0	-1	1	0	0	1	2	65.4	4	NF
59	770***	tatgtggatttcgagtcgtctta	13	NF	NF	NF	NF	NF	NF	NF	1	0	0	1	2	1	2	0	0	0	0	-1	2	0	0	-1	1	0	1	2	0	1	0	1	2	70.7	5	NF
60	771***	atgtggatttcgagtcgtcttaa	2	F	NF	F	NF	F	NF	NF	1	1	1	1	4	1	3	0	0	0	0	0	4	1	1	0	0	0	1	3	0	1	1	1	3	82.6	5	NF
61	772***	tgtggatttcgagtcgtcttaat	3	NF	F	NF	NF	F	F	NF	1	0	1	1	3	1	4	1	0	0	0	0	6	2	0	-1	0	0	1	3	1	1	0	1	3	80.1	6	NF
62	773***	gtggatttcgagtcgtcttaatg	3	F	F	F	NF	F	NF	NF	1	1	1	1	4	1	4	1	1	0	0	-1	6	2	1	0	0	0	1	4	0	1	1	1	3	84.5	5	NF
63	774**,***	tggatttcgagtcgtcttaatgt	1	F	F	F	F	F	NF	NF	1	1	1	1	4	1	5	0	0	1	0	0	7	1	1	0	0	0	1	3	1	1	1	1	4	83.7	5	NF
64	775***	ggatttcgagtcgtcttaatgta	37	NF	NF	NF	NF	NF	NF	NF	0	0	1	1	2	1	4	0	0	0	-1	0	4	-1	0	0	0	-1	0	-2	0	0	0	1	1	55.2	4	NF

Table 1. Continued

65	776***	gatttcgagtcgtcttaatgtat	31	NF	NF	NF	NF	NF	F	NF	1	0	1	1	3	1	4	0	0	0	0	0	5	-1	0	-1	1	0	1	1	0	1	0	1	2	58.3	7	NF
66	777***	atttcgagtcgtcttaatgtata	7	NF	F	NF	NF	F	F	NF	1	0	1	1	3	1	4	1	0	1	0	0	7	0	0	-1	0	0	1	1	1	1	0	1	3	80.9	6	NF
67	778***	tttcgagtcgtcttaatgtatag	15	NF	NF	NF	NF	F	F	NF	1	0	1	1	3	1	4	0	0	0	0	0	5	2	0	-1	0	0	1	3	0	1	0	1	2	77.2	6	NF
68	779***	ttcgagtcgtcttaatgtataga	1	F	F	F	NF	F	F	NF	1	1	1	1	4	1	4	1	1	1	0	0	8	2	1	0	0	0	1	4	1	1	1	1	4	93.5	6	NF
69	780***	tcgagtcgtcttaatgtatagat	3	NF	NF	NF	NF	F	F	NF	0	1	1	1	3	1	4	0	0	1	-1	0	5	1	1	0	0	-1	0	1	1	0	1	1	3	83.2	6	NF
70	781***	cgagtcgtcttaatgtatagatt	35	NF	NF	NF	NF	F	NF	NF	1	0	1	1	3	0	4	1	0	0	0	-1	4	0	0	0	0	0	1	1	1	1	0	1	3	81.5	4	NF
71	782***	gagtcgtcttaatgtatagattt	54	F	NF	F	F	NF	NF	NF	1	1	1	1	4	0	4	0	0	0	0	0	4	1	1	0	0	0	1	3	1	1	1	1	4	74.7	4	NF
72	783***	agtcgtcttaatgtatagatttg	6	NF	NF	NF	NF	F	NF	NF	1	0	1	1	3	0	4	0	0	1	0	0	5	2	0	-1	0	0	1	3	1	1	0	1	3	78.4	4	NF
73	857	tgctggtgccaaccctattctcc	24	F	NF	F	F	NF	NF	NF	1	1	1	1	4	1	4	0	0	0	0	0	5	1	1	0	0	0	1	3	1	1	1	1	4	65.5	5	NF
74	891**	aagcactctgattgacaaatacg	5	F	F	F	F	F	NF	F	1	1	1	1	4	1	5	1	1	1	0	0	9	2	1	0	0	0	1	4	1	1	1	1	4	95.9	5	F
75	903	tgacaaatacgatttatctaatt	14	NF	F	NF	NF	F	NF	NF	1	0	1	1	3	0	4	1	1	0	0	0	6	1	0	0	0	0	1	2	1	1	0	1	3	88.8	3	NF
76	904**	gacaaatacgatttatctaattt	13	F	F	F	F	F	NF	NF	1	1	1	1	4	0	4	0	1	1	0	0	6	1	1	0	1	0	1	4	1	1	1	1	4	83.5	5	NF
77	998	tgccaggtatcaggcaaggatat	5	F	NF	F	F	F	NF	NF	1	1	1	1	4	1	3	0	1	0	0	0	5	1	1	0	0	0	1	3	1	1	1	1	4	80.4	4	NF
78	1044	tattctgattacacccgaggggg	75	NF	NF	NF	NF	NF	NF	NF	0	0	0	1	1	1	1	0	0	0	-1	0	1	-2	0	-1	1	-1	0	-2	0	0	0	0	0	50.3	3	NF
79	1048**	ctgattacacccgagggggatga	6	NF	NF	NF	NF	NF	NF	NF	1	1	0	1	3	0	2	0	0	0	0	-1	1	0	1	0	0	0	1	2	0	1	1	0	2	68.2	0	NF
80	1054	acacccgagggggatgataaacc	10	NF	F	F	NF	F	NF	NF	1	0	1	1	3	1	5	1	0	0	0	0	7	2	0	0	1	0	1	4	0	1	0	1	2	80.6	3	NF
81	1063***	ggggatgataaaccgggcgcggt	6	NF	NF	NF	NF	NF	NF	NF	0	1	0	1	2	0	0	0	1	0	-1	-1	-1	-1	1	0	1	-1	0	0	1	0	1	0	2	51.2	2	NF
82	1064*,***	gggatgataaaccgggcgcggtc	46	NF	NF	NF	NF	NF	NF	NF	0	1	0	0	1	0	0	0	0	0	-1	-1	-2	-2	1	0	0	-1	0	-2	0	0	1	0	1	47.9	2	NF
83	1065***	ggatgataaaccgggcgcggtcg	15	NF	NF	NF	NF	NF	NF	NF	1	0	0	0	1	0	1	0	0	0	0	-1	0	-1	0	0	1	0	1	1	0	1	0	0	1	44.4	0	NF
84	1074***	accgggcgcggtcggtaaagttg	5	F	NF	F	F	NF	NF	NF	1	1	1	1	4	0	4	0	0	1	0	-1	4	2	1	0	0	0	1	4	1	1	1	1	4	71.2	3	NF
85	1075*,**,***	ccgggcgcggtcggtaaagttgt	8	F	NF	F	NF	F	NF	NF	1	1	1	1	4	0	4	0	0	0	0	0	4	2	1	0	0	0	1	4	0	1	1	1	3	83.3	3	NF
86	1076***	cgggcgcggtcggtaaagttgtt	3	NF	NF	NF	NF	NF	F	NF	0	1	1	1	3	0	3	0	0	0	-1	0	2	2	1	0	0	-1	0	2	0	0	1	1	2	76.3	3	NF
87	1081	gcggtcggtaaagttgttccatt	2	F	NF	NF	F	F	NF	NF	1	1	1	1	4	1	3	1	0	0	0	0	5	0	1	0	0	0	1	2	1	1	1	1	4	91.3	4	NF
88	1085*,**,***	tcggtaaagttgttccatttttt	5	F	NF	F	NF	F	NF	NF	1	1	1	1	4	0	5	0	0	0	0	0	5	2	1	0	1	0	1	5	0	1	1	1	3	91.5	5	NF
89	1110	agcgaaggttgtggatctggata	1	F	F	NF	F	F	NF	NF	1	1	1	1	4	1	2	1	1	1	0	0	6	0	1	0	0	0	1	2	1	1	1	1	4	92.4	4	NF
90	1111	gcgaaggttgtggatctggatac	4	F	NF	NF	NF	F	F	NF	1	1	1	1	4	1	3	0	1	0	0	0	5	0	1	0	0	0	1	2	0	1	1	1	3	83.9	6	NF
91	1120	gtggatctggataccgggaaaac	6	NF	F	F	F	F	NF	NF	1	1	0	1	3	1	3	1	1	1	0	0	7	2	1	0	0	0	1	4	1	1	1	1	4	88.5	5	NF
92	1137	gaaaacgctgggcgttaatcaaa	13	NF	F	NF	NF	NF	NF	NF	1	0	1	1	3	1	4	1	1	0	0	0	7	-1	0	0	0	0	1	0	0	1	0	1	2	63.7	2	NF
93	1186**	atgattatgtccggttatgtaaa	2	F	F	NF	NF	F	NF	NF	1	1	1	1	4	1	4	1	0	0	0	0	6	0	1	0	0	0	1	2	0	1	1	1	3	96.0	3	NF
94	1190	ttatgtccggttatgtaaacaat	103	NF	F	NF	NF	NF	F	NF	1	0	1	1	3	1	4	1	0	1	0	-1	6	0	0	0	0	0	1	1	1	1	0	1	3	72.5	6	NF
95	1228**	gccttgattgacaaggatggatg	26	NF	NF	NF	NF	NF	F	NF	1	1	0	1	3	1	3	1	0	0	0	-1	4	-1	1	0	0	0	1	1	0	1	1	1	3	76.0	3	NF
96	1234	attgacaaggatggatggctaca	3	NF	F	NF	NF	F	NF	NF	1	0	1	1	3	1	2	1	1	1	0	0	6	0	0	-1	1	0	1	2	1	1	0	1	3	86.5	2	NF
97	1308**	ccgcctgaagtctctgattaagt	4	F	F	F	NF	F	NF	NF	1	1	1	1	4	1	5	1	0	0	0	0	7	3	1	0	1	0	1	6	0	1	1	1	3	97.8	5	NF
98	1346**	tggctcccgctgaattggaatcc	3	F	NF	F	NF	NF	NF	NF	1	1	1	1	4	1	3	0	0	0	0	0	4	2	1	0	0	0	1	4	0	1	1	1	3	74.3	4	NF
99	1352**	ccgctgaattggaatccatcttg	11	F	NF	F	NF	F	NF	NF	1	1	1	1	4	1	3	0	0	0	0	0	4	1	1	0	1	0	1	4	0	1	1	1	3	81.1	4	NF
100	1429***	gccggtgaacttcccgccgccgt	21	NF	NF	NF	NF	NF	NF	NF	0	1	0	1	2	0	0	0	0	1	-1	0	0	0	1	0	1	0	0	2	1	0	1	0	2	51.3	2	NF
101	1430*,***	ccggtgaacttcccgccgccgtt	82	NF	NF	NF	NF	NF	NF	NF	0	1	0	0	1	0	0	0	0	0	-1	-1	-2	-1	1	0	1	-1	0	0	0	0	1	0	1	55.5	0	NF
102	1431**,***	cggtgaacttcccgccgccgttg	16	NF	NF	NF	NF	NF	NF	NF	1	1	0	0	2	0	1	0	0	0	0	0	1	0	1	0	0	0	1	2	0	1	1	0	2	55.5	3	NF
103	1431	cggtgaacttcccgccgccgttg	16	NF	NF	NF	NF	NF	NF	NF	1	1	0	0	2	0	1	0	0	0	0	0	1	0	1	0	0	0	1	2	0	1	1	0	2	55.5	3	NF
104	1439*,**,***	ttcccgccgccgttgttgttttg	62	NF	NF	F	NF	NF	NF	NF	1	1	1	0	3	0	4	0	0	0	0	-1	3	3	1	0	0	0	1	5	0	1	1	1	3	65.8	4	NF
105	1440*,**,***	tcccgccgccgttgttgttttgg	33	F	NF	F	F	F	NF	NF	1	1	1	1	4	0	4	0	0	1	0	0	5	3	1	0	0	0	1	5	1	1	1	1	4	81.4	4	NF
106	1441*,***	cccgccgccgttgttgttttgga	41	NF	NF	NF	NF	F	NF	NF	0	1	1	1	3	0	4	0	0	1	-1	0	4	2	1	0	0	-1	0	2	1	0	1	1	3	75.8	3	NF
107	1482	ggaaaaagagatcgtggattacg	3	NF	F	NF	NF	F	NF	NF	1	0	1	1	3	0	4	1	1	1	0	0	7	0	0	0	0	0	1	1	1	1	0	1	3	79.9	3	NF
108	1509**	cagtcaagtaacaaccgcgaaaa	42	NF	NF	F	NF	F	NF	NF	1	1	0	1	3	1	2	1	0	0	0	0	4	1	1	0	0	0	1	3	0	1	1	0	2	92.1	4	NF
109	1512	tcaagtaacaaccgcgaaaaagt	12	NF	F	F	NF	F	NF	NF	1	0	1	1	3	1	5	1	0	0	0	0	7	1	0	0	1	0	1	3	0	1	0	1	2	86.9	5	NF
110	1522**	accgcgaaaaagttgcgcggagg	56	NF	NF	F	NF	F	NF	NF	1	1	0	1	3	0	1	1	0	0	0	-1	1	1	1	0	1	0	1	4	0	1	1	0	2	81.1	1	NF
111	1539	cggaggagttgtgtttgtggacg	2	F	NF	NF	F	F	NF	NF	1	1	1	1	4	1	2	1	0	1	0	0	5	0	1	0	0	0	1	2	1	1	1	1	4	93.5	5	NF
112	1565**	taccgaaaggtcttaccggaaaa	12	NF	NF	F	NF	F	NF	NF	1	1	0	1	3	1	2	1	0	0	0	0	4	2	1	0	1	0	1	5	0	1	1	1	3	88.7	3	NF
113	1586**	aactcgacgcaagaaaaatcaga	12	F	F	F	F	F	NF	F	1	1	1	1	4	1	4	1	0	0	0	0	6	1	1	0	0	0	1	3	1	1	1	1	4	82.1	5	F
114	1601	aaatcagagagatcctcataaag	6	NF	F	F	NF	F	NF	F	1	0	1	1	3	1	4	1	0	0	0	0	6	1	0	0	1	0	1	3	1	1	0	1	3	83.0	5	F

1) siRNA number indicates the nucleotide position within the coding region of the target firefly *luciferase* mRNA, corresponding to the 3' siRNA-passenger-strand end.

2) Target sequence corresponding from 3' siRNA-passenger-strand end to 3' siRNA-guide-strand end (23nt).

3) siRNA predicted to be functional is shown as "F", not functional, "NF".

*siRNAs used in Ui-Tei, K., et al., Nucleic Acids Res., 32, 936–948, 2004.

**siRNAs used in Ui-Tei, K., et al., Methods in Mol. Biol., 361, 201–216, 2007.

***siRNAs used in Ui-Tei, K., et al., New Research on Antisense Element (Genetics), in press.

[Scoring criteria of each condition]

Algorithm I (Functional = 4, Not functional = 0~3)

i 5' terminus of the passenger strand; G/C → 1, A/U → 0
ii 5' terminus of the guide strand; A/U → 1, G/C → 0
iii 5' terminus 7 bp double stranded region of the guide strand; A/U≥4 → 1, A/U<4 → 0
iv 19bp double stranded region; G/C≤9 → 1, G/C>9 → 0

Algorithm II (Functional ≥ 6, Not functional < 6)
I 19bp double stranded region; 7≤G/C≤10 → 1, 0≤G/C<7 → 0, 10<G/C≤19 → 0
II A/U contents at position 15–19 of the passenger strand; A/U=1 → 1, A/U=2 → 2, A/U=3 → 3, A/U=4 → 4, A/U=5 → 5
III Tm of potential internal hairpin, This parameter cannot be evaluated.
IV At position 19 of the passenger strand; A → 1, others → 0
V At position 3 of the passenger strand; A → 1, others → 0
VI At position 10 of the passenger strand; U → 1, others → 0
VII At position 19 of the passenger strand; G/C → -1, A/U → 0
VIII At position 13 of the passenger strand; G → -1, others → 0

Algorithm III (Functional ≥ 3, Not functional < 3)
1 AU differential ; (A/U in the three 5' terminal of the guide strand) – (A/U in the three 5' terminal of the passenger strand) → -3~3
2 At position 1 of the passenger strand; G/C → 1, A/U → 0
3 At position 1 of the passenger strand; U → -1, others → 0
4 At position 6 of the passenger strand; A → 1, others → 0
5 At position 19 of the passenger strand; G → -1, others → 0
6 At position 19 of the passenger strand; A/U → 1, G/C → 0

Algorithm IV (Functional = 4, Not functional = 0~3)
1 A/U at position 10 of the passenger strand → 1
2 A/U at position 19 of the passenger strand → 1
3 G/C at position 1 of the passenger strand → 1
4 More than three A/U between position 13–19 of the passenger strand → 1

Algorithm V (Functional ≥ 75.0%, Not functional < 75.0%, defined in this study)

Algorithm VI (Functional ≥ 6, Not functional < 5)

<References>
Algorithm I: Ui-Tei, K. et al. Nucleic Acids Res., 32, 936, 2004.
Algorithm II: Reynolds, A. et al. Nat. Biotechnol., 22, 326, 2004.
Algorithm III: Amarzguioui, M. and Prydz, H. Biochem. Biophys. Res. Commun., 316, 1050, 2004.
Algorithm IV; Jagla, B. et al. RNA, 11, 864, 2005.
Algorithm V: Vert, J-P. et al. BMC Bioinformatics, 7, 520, 2006.
Algorithm VI: Chalk, A.M. et al. Biochem. Biophys. Res. Commun., 319, 264, 2004.
Algorithm VII: Elbashir, S.M. et al. Genes Dev., 15, 188, 2001.

Table 2. Functionality of siRNAs and the number of unselectable transcripts (Total=24,261)

Algorithm*			Percentage of siRNA		No. of transcript for which functional siRNA can not be designed
			Relative luc activity		
I	II	III	≤10%	≤33%	No. of gene
○			73	98	0
	○		66	95	0
		○	56	88	0
○	○		75	100	6
	○	○	67	100	6
○		○	67	97	0
○	○	○	67	100	8
			14	51	-

*Circle; the algorithm is satisfied, blank column; the algorithm is not satisfied.

Functionalities of the Selected siRNAs Evaluated by a Reporter Assay

To evaluate their functionalities, 114 siRNA sequences were designed against the firefly *luc* gene according to the scoring criteria of each algorithm, score 4, ≥ 6 and ≥ 3 are functional in algorithm I, II and III, respectively (Table 1). Their effects were investigated using a Dual-Luciferase Reporter Assay System [4] in duplicated experiments (Figure 1). The concentration of siRNA used was 50nM. The relative level of luc activity was reduced to below 10 % by 73, 66, and 56 % of the siRNAs predicted to be functional by algorithms I, II, and III, respectively (Figure 1 and Table 2). Only 14 % of the siRNAs predicted to be not functional exhibited luc activity less than 10 %. Fourty of 41 (98 %) of siRNAs predicted to be functional by algorithm I reduced luc activity equal to or less than 33 % (Table 2). Furthermore, we compared the scores of siRNAs that only be predicted functional in one algorithm but not any other two algorithms. All the siRNAs predicted to be functional in algorithm II or III but not algorithms I and III, or I and II showed marginal score 3 in algorithm I. The average scores of siRNAs predicted to be functional only in algorithm I or II also showed close scores 2.0 or 1.5, respectively, in algorithm III. Although siRNAs predicted to be functional in algorithm I alone showed an average score of 4.7 in algorithm II, those predicted functional in algorithm III alone was as low as 1.2. The result clearly suggested that it would be much safer to use at least two different algorithms I and II in designing effective siRNAs successfully. So the assessments of the 114 *luc* siRNAs evaluated in the other algorithms IV-VII published were also performed (Table 1) [7-10]. The relative luc activity was reduced to below 10 % by 64, 67, 56 and 33 % of siRNAs, and below 30 % by 91, 92, 93 and 89 % of siRNAs predicted to be functional by algorithms IV, V, VI and VII, respectively. And 100 % of siRNAs simultaneously selected by at least 4 among 7 algorithms reduced relative luc activity below 33 %, and 71 % of them reduced the activity below 10 %. In contrast, 9 % of siRNAs which were not selected to be functional by any algorithms reduced the luc activity below 10 %.

Other Insight

Mismatched siRNA may occasionally inactivate genes other than the target, an undesired side effect designated as the “off-target effect”. For selecting siRNA sequences, it is also important to avoid off-target effects especially for therapeutic applications. The molecular basis for this remains to be clarified, though mRNA cleavage requires a nearly strict nucleotide sequence identity between the mRNA target portion and guide strand of siRNA. Thus, at least some fraction of undesirable siRNAs, giving rise to the off-target effect through destabilization of mRNAs other than the target, may be eliminated by computer-based homology search [11]. siRNA is also shown to be capable of causing translational inhibition depends on the "seed" sequence, 2-8 nucleotides from the 5’end of the guide strand [12]. The homology search of possible seed sequences (4^7 sequences) for human RefSeq mRNA database (version 20) indicated that most of all 7 nucleotide seed sequences are proved to be completely matched with more than two genes. Furthermore, about 65 % of 7 nucleotide sequences matched with multi-sites in over 200 human genes (our unpublished data). If translational inhibition is taken up by any seed sequences, the off-target effect could not be avoided by using any siRNA. Although further study for revealing the molecular mechanism of off-target effect is expected, one promissing procedure for reducing off-target effects without any mechanistic implication is to use multiple siRNAs at a time. The siRNA sequences completely matched with a target gene but each seed sequence is distinct should lead specific interference without actual off-target effects.

Conclusion

In this report, we compared the rational siRNA design algorithms for selecting reliable siRNA sequences by the uniform assay system using the firefly *luc* gene as a target. The result suggests that it is efficient to select siRNAs commonly satisfying the reliable algorithms combined as mentioned above. However, it is noted that the experiments targeting endogenous genes might provide more information. Our findings should be considered when conducting RNAi experiments.

Acknowledgments

This work was supported in part by Special Coordination Fund for promoting Science and Technology to K.S., and grants from the Ministry of Education, Culture, Sports, Science and Technology of Japan to K.U-T. and K.S.

References

[1] Dorsett, Y. & Tuschl, T. (2004). siRNAs: applications in functional genomics and potential as therapeutics. *Nat. Rev. Drug Discov.,* **3**, 318-329.

[2] Elbashir, S.M., Harborth, J., Lendeckel, W.,Yalcin, A.,Weber, K. & Tuschl, T. (2001). Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature,* **24***,* 494–498.

[3] Holen, T., Amarzguioui, M., Wiiger, M.T., Babaie, E. & Prydz, H. (2002). Positional effect of short interfering RNAs targeting the human coagulation trigger Tissue Factor. *Nucleic Acids Res.,* **30***,* 1757-1766.

[4] Ui-Tei, K., Naito, Y., Takahashi, F., Haraguchi, T., Ohki-Hamazaki, H., Juni, A., Ueda, R. & Saigo, K. (2004). Guidelines for the selection of highly effective siRNA sequences for mammalian and chick RNA interference, *Nucleic Acids Res.,* **32**, 936–948.

[5] Reynolds, A., Leake, D., Boese, Q., Scaringe, S., Marshall, W.S. & Khvorova, A. (2004). Rational siRNA design for RNA interference. *Nat. Biotechnol.,* **22***,* 326-330.

[6] Amarzguioui, M. & Prydz, H. (2004). An algorithm for selection of siRNA sequences. *Biochem, Biophys. Res. Commun.,* **316**, 1050-1058.

[7] Jagla, B., Aulner, N., Kelly, P. D., Song, D., Volchuk, A., Zatorski, A., Shum, D., Mayer, T., De Angelis, D. A., Ouerfelli, O., Rutishauser, U. & Rothman, J. E. (2007). Sequence characteristics of functional siRNAs. *RNA,* **11***,* 864-872.

[8] Vert, J.-P., Foveau, N., Lajaunie, C. & Vandenbrouck, Y. (2006). An accurate and interpretable model for siRNA efficacy prediction. *BMC Bioinformatics,* **7***,* 520.

[9] Chalk, A. M., Wahlestedt, C. & Sonnhammer, E. L. L. (2004). Improved and automated prediction of effective siRNA. *Biochem. Biophys. Res. Commun.,* **319***,* 264-274.

[10] Elbashir, S. M., Martinez, J., Patkaniowska, A., Lendeckel, W. & Tuschl, T. (2001). Functional anatomy of siRNAs for mediating efficient RNAi in *Drosophila melanogaster* embryo lysate. *EMBO J.,* **20***,* 6877-6888.

[11] Naito, Y., Yamada, T., Ui-Tei, K., Morishita, S. & Saigo, S. (2004). siDirect: highly effective, target-specific siRNA design software for mammalian RNA interference, *Nucleic Acids Res.,* **32***,* W124-129.

[12] Lewis, B.P., Burge, C.B. & Bartel, D.P. (2005). Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell,* **120***,* 15-20.