

Research Article

Triplet Analysis That Identifies Unpaired Regions of Functional RNAs

Junji Kawakami,^{1,2} Yoshie Yamaguchi,^{1,2} and Naoki Sugimoto^{1,2}

¹Department of Nanobiochemistry, FIRST, Konan University, 7-1-20 Minatojima-minamimachi, Chuo-ku, Kobe 650-0047, Japan

²Frontier Institute for Biomolecular Engineering Research (FIBER), Konan University, 7-1-20 Minatojima-minamimachi, Chuo-ku, Kobe 650-0047, Japan

Correspondence should be addressed to Junji Kawakami, kawakami@konan-u.ac.jp

Received 15 May 2011; Accepted 28 June 2011

Academic Editor: Daisuke Miyoshi

Copyright © 2011 Junji Kawakami et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We developed a novel method for analyzing RNA sequences, deemed triplet analysis, and applied the method in an *in vitro* RNA selection experiment in which HIV-1 Tat was the target. Aptamers are nucleic acids that bind a desired target (bait), and to date, many aptamers have been identified by *in vitro* selection from enough concentrated libraries in which many RNAs had an obvious consensus primary sequence after sufficient cycles of the selection. Therefore, the higher-order structural features of the aptamers that are indispensable for interaction with the bait must be determined by additional investigation of the aptamers. In contrast, our triplet analysis enabled us to extract important information on functional primary and secondary structure from minimally concentrated RNA libraries. As a result, by using our method, an important unpaired region that is similar to the bulge of TAR was readily predicted from a partially concentrated library in which no consensus sequence was revealed by a conventional sequence analysis. Moreover, our analysis method may be used to assess a variety of structural motifs with desired function.

1. Introduction

In vitro selection of nucleic acids is a powerful method used to isolate novel functional molecules [1–6]. Generally, *in vitro* selection involves many rounds of selection followed by sequencing and analysis of several recovered clones and determination of a consensus sequence. These consensus sequences are often highly functional molecules that have promise as pharmaceutical and chemical agents [7–9]. However, there are several inherent biases that often affect *in vitro* selection procedures, for example, the abundance of particular motifs in the library, differential efficiency of reversetranscription or amplification by PCR due to stable secondary structures. Therefore, confirmation that a particular consensus sequence is optimal is difficult and usually requires identification of a naturally occurring, functional counterpart. For the selection of aptamers, which interact with desired targets, a consensus motif may be a sequence in a local minimum rather than an absolute minimum in an energy diagram that represents interactions between all

possible aptamers and the bait or target. Furthermore, it should be hard to identify general rules of intermolecular interactions from only such “highly evolved” motifs. To overcome this problem, the sequence changes in the library during the early stage of the selection should be analyzed. Evaluation of the sequences that are gradually concentrated from the starting library should reveal a general rule for interactions, that is, information that is indispensable for designing new artificial functional molecules.

Here, we developed a method, designated triplet analysis, to identify important sequence characteristics in the primary sequences that have been recovered from an RNA library after only several rounds of selection with a target molecule. The first step of triplet analysis is extraction of all nucleotide triplets from the primary sequence (Figure 1). For example, a sequence AGCGUCCA would be separated to six triplets: AGC, GCG, CGU, GUC, UCC, and CCA. The next step is reconstruction of the parent sequence of these triplets. If these six triplets could be extracted, one would easily reconstruct the parent sequence, AGCGUCCA. Thus, a

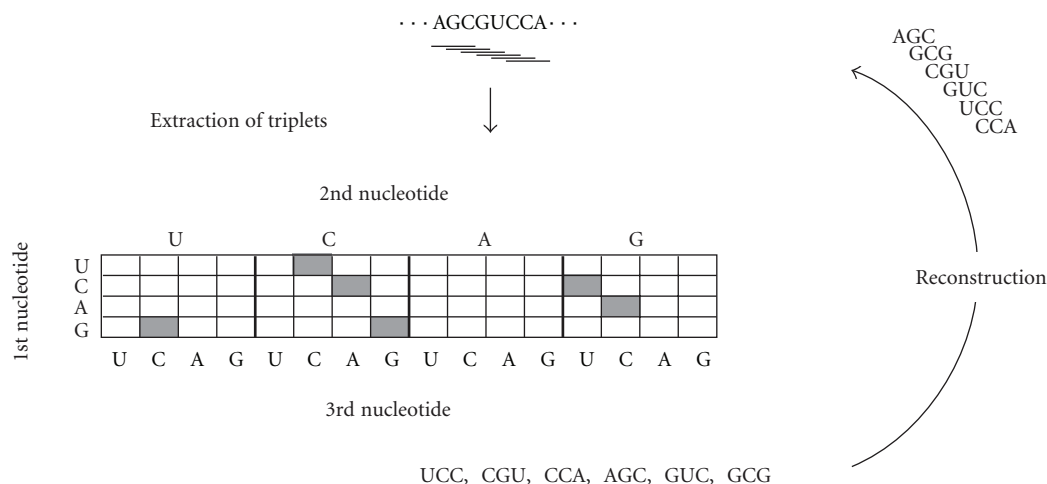


FIGURE 1: The scheme for the basic triplet analysis of a library. Frequent successive sequences could be determined by this method.

frequent successive sequence should be revealed from an analysis of frequent triplets. We chose the HIV-1 Tat protein as our experimental bait for our *in vitro* RNA selection because a naturally occurring binding RNA motif, TAR, and artificially selected binding motifs have been identified [10, 11]. Therefore, we could compare our finding with these previously identified binding motifs to judge the validity of our analysis method.

2. Materials and Methods

2.1. In Vitro Selection of RNA Aptamers. A library comprised of 78-nucleotide RNAs that each contained a region of 30 randomized nucleotides was prepared using chemically synthesized 97-bp DNAs that each had a T7 promoter sequence as templates and *in vitro* transcription. Approximately 1 nmol of RNA was generated from every 100 pmol of template DNA using the AmpliScribe T7 transcription kit (EPICENTRE). The transcripts were purified from 8% denaturing polyacrylamide gels that contained 7 M urea by means of crush and soak extraction and subsequent ethanol precipitation. The bait, Tat peptide (RKKRRGRRR), was synthesized using an Fmoc strategy on a solid support of Fmoc-NH-SAL resin (N- α -9-fluorenylmethoxy-carbonyl-super acid labile polystyrene resin, Watanabe Chem.), and this synthesis method produces an amide at the C-terminus of the bait peptide, as described previously.

The purified RNA was incubated in a binding buffer containing 2.5 mM Tris-HCl (pH 7.6), 100 mM NaCl, 2.0 mM MgCl₂ and then passed through a nitrocellulose filter (HAWP, MILLIPORE) three times to remove RNAs with filter binding properties. A sample of the RNA library (100 pmol) was incubated with 5 pmol of Tat peptide in 50 μ L of the binding buffer for 60 min at 25°C, and then the mixture was passed through another nitrocellulose filter. After washing with 1.5 mL of the binding buffer, RNAs bound to Tat on the filter were collected and amplified by subsequent reverse transcription, PCR, and transcription (RNA PCR kit, PE Applied Biosystems).

2.2. Dot Blot Analysis.

2.3. Cloning and DNA Sequencing. To analyze the binding ability of the RNA library, RNAs were labeled internally with [α -³²P] UTP (Amersham) during transcription *in vitro*. Labeled RNA (10 pmol) was mixed with 2.5 pmol of Tat peptide in 50 μ L of the binding buffer. After a 60 min incubation, the mixture was passed through a nitrocellulose filter. The filter was washed with the binding buffer, and radioactive RNAs on the filter were visualized and quantified with the BIO-Imaging Analyzer, BAS 2000 (Fuji Film).

After six rounds of selection and amplification, the ends of the double-stranded PCR products were blunted by T4 DNA polymerase (TAKARA). The blunted DNAs were purified and concentrated by ethanol precipitation and phosphorylated at their 5' end with T4 polynucleotide kinase (TOYOBO). The resulting DNAs were ligated into the *Hinc* II site of separate pUC118 plasmids using T4 DNA ligase (TAKARA). The ligated plasmid DNAs were transformed into *E. coli* MV1184, which were then cultured on LB plates containing ampicillin, IPTG (isopropyl- β -D-thiogalactoside), and X-gal (5-bromo-4-chloro-3-indolyl- β -D-galactoside). After blue-white selection, 20 white colonies were randomly selected and cultured in LB medium containing ampicillin. The recombinant plasmid DNA was recovered from *E. coli* clones using the alkaline lysis procedure. Nucleotide sequences were determined using the BigDye-termination method and an ABI PRISM 310 Genetic Analyzer (PE). Of the 20 clones analyzed, 17 contained a library sequence as the insert.

3. Results and Discussion

3.1. Simple Triplet Analysis to Identify Important Primary Structure. From the library shown in Figure 2(a), RNAs that bound Tat were concentrated as described in the previous section. The binding ability of the library RNAs on each round was analyzed by dot blot analysis, and after six rounds of selection, accumulation of Tat-binding RNAs on the

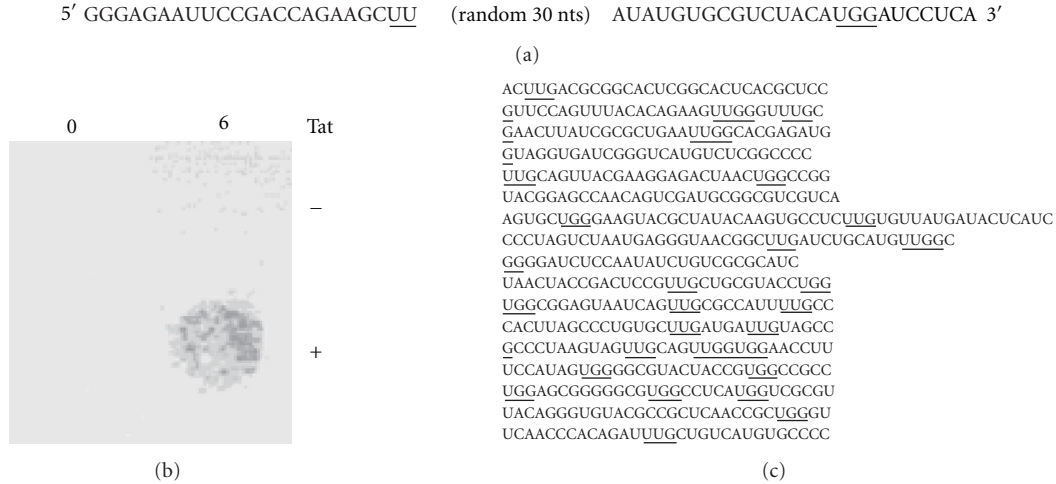


FIGURE 2: The RNA library used in this study. (a) The primary sequence of the constant region and the location of the 30 random nucleotides within the 78-nucleotide RNAs that compose the library are indicated. (b) Tat-binding ability of the libraries before selection (0) and after six rounds of the selection (6). Dot blot analysis was carried out with 10 pmol 32 P-labeled RNA and 2.5 pmol Tat peptide in 50 μ L of the binding buffer. (c) Primary sequence of the 30-nucleotide random region from 17 clones of the library after six rounds of the selection. The frequent triplets after the triplet subtraction, UUG and UGG, are underlined.

TABLE 1: Number and frequency of each triplet obtained from the primary sequences of the clones after the selection^(a).

	Second		U				C			A			G					
	First	Third	U	C	A	G	U	C	A	G	U	C	A	G	U	C	A	G
Number of triplet	U		5	1	5	16	6	5	9	8	4	12	6	6	9	14	8	15
	C		7	10	7	10	7	8	6	7	8	7	5	8	8	13	4	10
	A		4	8	4	9	11	6	5	8	4	8	0	5	14	4	4	4
	G		11	9	9	11	10	14	7	12	9	3	6	6	8	13	7	13
Frequency of triplet	U		0.29	0.06	0.29	0.94	0.35	0.29	0.53	0.47	0.24	0.71	0.35	0.35	0.53	0.82	0.47	0.88
	C		0.41	0.59	0.41	0.59	0.41	0.47	0.35	0.41	0.47	0.41	0.29	0.47	0.47	0.76	0.24	0.59
	A		0.24	0.47	0.24	0.53	0.65	0.35	0.29	0.47	0.24	0.47	0.00	0.29	0.82	0.24	0.24	0.24
	G		0.65	0.53	0.53	0.65	0.59	0.82	0.41	0.71	0.53	0.18	0.35	0.35	0.47	0.76	0.41	0.76

^(a) Abundant triplets with frequency above 0.75 are indicated by boldface.

filter increased as shown in Figure 2(b). The RNAs which recovered after six rounds of selection were cloned using the *E. coli* MV1184/pUC118 system and sequenced. The primary sequences of 17 clones are shown in Figure 2(c).

Apparently, there were no regions of obvious primary sequence similarity shared among these sequences, and a conventional sequence alignment method [12] did not reveal any consensus sequence (data not shown). All triplets in the primary sequences of the 30-nucleotide randomized regions were extracted. Of all 64 possible triplets, 23 triplets had frequencies (the number of triplets divided by the number of clones) above 0.5 (Table 1). In other words, these 23 triplets occurred in a half of the clones stochastically. Notably, eight triplets, UUG, UGC, UGG, CGC, AGU, GCC, GGC, and GGG had, frequencies over 0.75 (bold values in Table 1). Successive sequences of over five nucleotides available from these triplets were UUGCC, UUGGC, and UUGGG. However, these pentanucleotide sequences were found in only one or two of the 17 clones though the eight triplets should potentially be an important component of

a Tat-binding RNA. This finding indicated that the longer consensus sequence had not been concentrated within six rounds of selection and that the process was in a relatively early stage of selection.

3.2. Triplet Subtraction Analysis to Identify the Important Secondary Structure. Further analysis, called triplet subtraction, was carried out as shown in Figure 3 to obtain detailed information, including RNA secondary structural features. In this analysis, surpluses of triplets are defined as the number of a triplet after taking away the number of all complementary triplets. For example, the numbers of AGC, GCG, and CGU subtracted by the numbers of GCU, CGC, and ACG, respectively, were calculated and listed. If a consensus sequence is in a fully matched stem region, there is no remaining triplet after the subtraction. On the other hand, if a consensus region is unpaired (i.e., in a bulge or loop), only the triplets concerned with the unpaired region are revealed after the subtraction. For example, six triplets—GCG, CGU, GUC, UCC, GGG, and

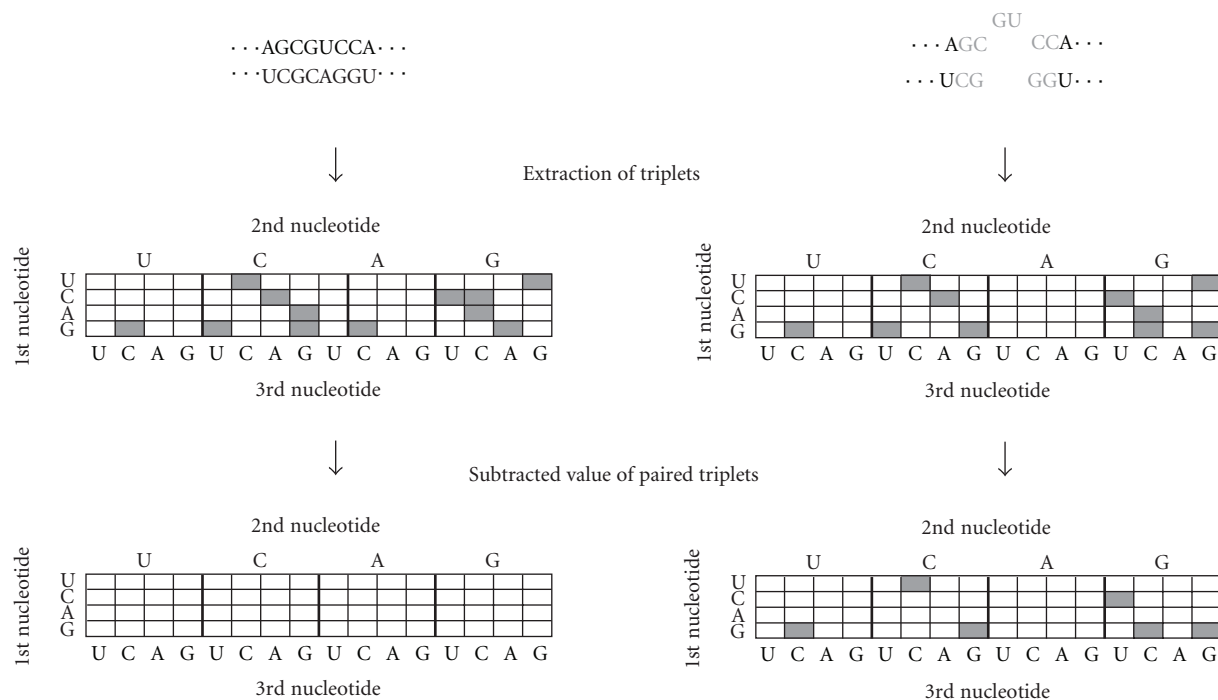


FIGURE 3: The scheme for the triplet subtraction from a library. Only the triplets from unpaired regions could be determined.

GGC—should be revealed from a paired region with a GU bulge, AGCGUCCA/UGGGCU, as shown in Figure 3. Thus, one can easily predict a frequent sequence around an unpaired region that might contribute to the function of an RNA. The results of the triplet subtraction analysis are summarized in Table 2 and Figure 4. As listed in Table 2, only two triplets UUG and UGG had a frequency over 0.5 after the subtraction. This value of 0.5 was surprisingly high because all the triplet frequency should be zero for a truly random library. Though UUU and GGG were the most frequent type of all NUU and GGN (where N indicates U, C, A, or G) triplets, respectively, the frequencies of these triplets were lower than those of UUG or UGG. Therefore, the most probable consensus sequence predicted from these triplets is UUGG. As listed in Table 2, many NNU and GNN might exist in unpaired region and the pattern of these triplets was widely spread, whereas UUG and UGG stood out among the NUN and NGN sequences, respectively. These findings indicated that the second U and/or the third G of the UUGG would be unpaired bases. The frequency of GGN triplets was moderate compared with that of UUG triplets, and, therefore, it was not clear whether the third G was unpaired. Nevertheless, this G might be an unpaired base because CCN triplets seemed not to be concerned with an unpaired region. Additionally, neither AAN nor ANN triplets seemed to occur in unpaired regions of secondary structure. Similar features were not found for NNC. These findings strongly indicated that an unpaired region closed by a UA base pair at their 5' end would not occur in motifs that bound to Tat, that is, the first UU of the UUGG should be unpaired bases. Ultimately, we could determine

that a sequence, NUUGG/CN with unpaired UUG, was an important structural motif that should have significant role in an RNA-Tat interaction (Figure 5(a), left). Possibly, a small portion of Tat-binding RNAs may contain an NUUGG/CCN sequence in which the UU doublet is unpaired as shown in Figure 5(a), right. This conclusion did not conflict with the results from a prediction of the secondary structure of RNA based on our thermodynamic parameters (data not shown).

3.3. Validation of the Triplet Analysis. TAR, the naturally occurring Tat-binding RNA motif, has a pyrimidine bulge, UCU, UUU, or UU, closed by a GC base pair at its 3' end as shown in Figure 5(b) [14–17]. The arginine side residues in Tat stack with the first U of the bulge and make hydrogen bonds to the GC base pair next to the 3' end of the bulge. These interactions along with electrostatic interactions at phosphate backbone are indispensable for the specific Tat-TAR interaction. Therefore, the essential structural motif of TAR that mediates binding with Tat is NNU(Y)UG(A)/(U)CNN with U(Y)U bulge, where Y indicates a pyrimidine nucleotide [11, 18]. Similar structural features were seen in RNA aptamers selected with arginine as a target. A general motif of these arginine aptamers was GUNGA/UCC with a UN bulge as shown in Figure 5(c) [13, 19]. Our UUGG motif closely resembled these indispensable structural features of TAR and arginine aptamers; therefore, we were able to confirm the validity of our triplet analysis method. The UUGG motif was revealed from the early stage of Tat-binding RNA selection, and therefore, the UUGG motif might be the ancestor of other Tat-binding RNAs and have fundamental property as a binding motif.

TABLE 2: Frequency of each triplet after subtraction^(a).

Second		U				C				A				G			
First	Third	U	C	A	G	U	C	A	G	U	C	A	G	U	C	A	G
U		0.29	—	—	0.65	0.12	—	0.06	0.24	—	0.18	0.06	—	0.24	0.41	—	0.53
C		0.12	0.24	0.06	0.12	0.18	—	—	—	—	—	—	—	—	0.06	—	0.18
A		—	—	—	0.06	—	—	—	—	—	—	—	—	0.18	—	—	—
G		0.18	0.35	—	0.24	0.35	0.06	—	—	0.06	—	0.29	—	0.12	—	0.12	0.29

^(a) Dash (—) indicates the value less than or equal to zero.

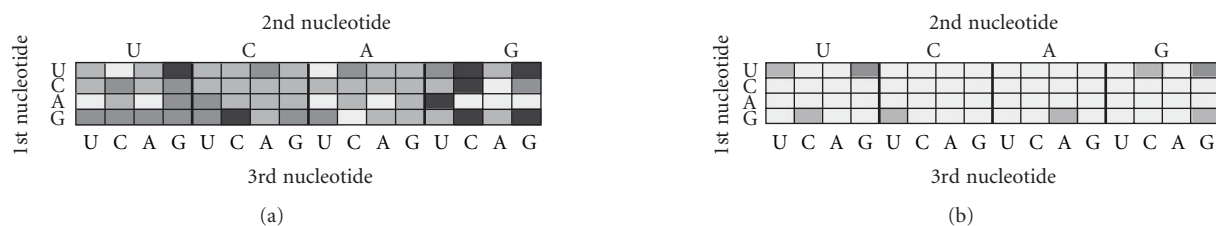


FIGURE 4: The frequency of the triplets before (a) and after (b) subtraction. The changes in gray scale (from light to dark) indicate changes in frequency (from 0–25%, 26–50%, 51–75%, to 76–100%).

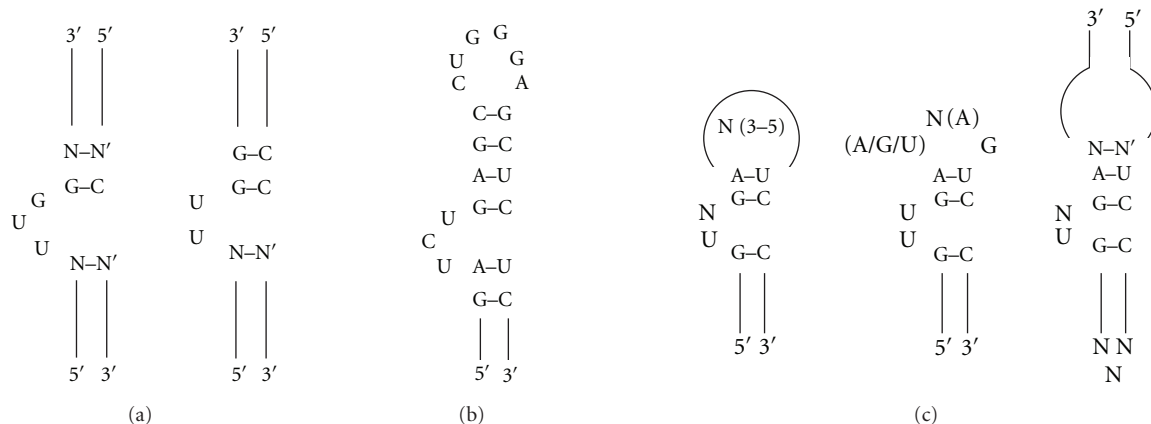


FIGURE 5: Secondary structures of Tat and/or arginine binding RNAs. (a) UUGG motif obtained in the present study. (b) Tat-binding motif of TAR [12]. (c) Several motifs of arginine aptamers from *in vitro* selection experiments [13].

In summary, we developed triplet analysis method to analyze a library. From our analysis of a small number of clones which recovered from a library after only a few rounds of selection, we could determine an important sequence motif though no consensus sequence was revealed by conventional analysis. Furthermore, unpaired regions in the selected RNAs, that may play important roles in many functional nucleic acids, could be readily predicted using this method. The resulting UUGG motif, NUUGG/CN with an unpaired UUG, resembled the essential features of Tat-binding RNAs. This motif may be a fundamental sequence of Tat- and arginine-binding motifs, and researchers may be able to construct a Tat-binding RNA library based on this information for obtaining “higher evolved” motifs.

Acknowledgments

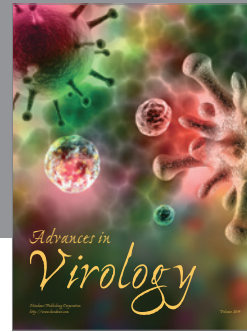
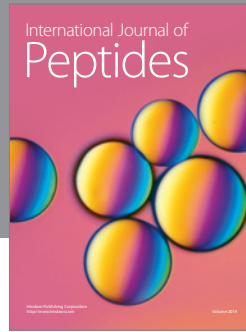
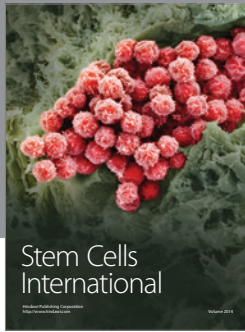
This work was supported in part by Grant-in-Aid for Scientific Research, the “Core Research” Project (2009–2014), and the “Academic Frontier” Project (2004–2009)

from the Ministry of Education, Culture, Sports, Science and Technology, Japan.

References

- [1] A. D. Ellington and J. W. Szostak, “In vitro selection of RNA molecules that bind specific ligands,” *Nature*, vol. 346, no. 6287, pp. 818–822, 1990.
- [2] C. Tuerk and L. Gold, “Systemic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase,” *Science*, vol. 249, no. 4968, pp. 505–510, 1990.
- [3] D. L. Robertson and G. F. Joyce, “Selection in vitro of an RNA enzyme that specifically cleaves single-stranded DNA,” *Nature*, vol. 344, no. 6265, pp. 467–468, 1990.
- [4] J. Tang and R. R. Breaker, “Rational design of allosteric ribozymes,” *Chemistry and Biology*, vol. 4, no. 6, pp. 453–459, 1997.
- [5] B. J. Tucker and R. R. Breaker, “Riboswitches as versatile gene control elements,” *Current Opinion in Structural Biology*, vol. 15, no. 3, pp. 342–348, 2005.

- [6] D. S. Wilson and J. W. Szostak, "In vitro selection of functional nucleic acids," *Annual Review of Biochemistry*, vol. 68, pp. 611–647, 1999.
- [7] J. Liu, Z. Cao, and Y. Lu, "Functional nucleic acid sensors," *Chemical Reviews*, vol. 109, no. 5, pp. 1948–1998, 2009.
- [8] S. M. Nimjee, C. P. Rusconi, and B. A. Sullenger, "Aptamers: an emerging class of therapeutics," *Annual Review of Medicine*, vol. 56, pp. 555–583, 2005.
- [9] D. López-Colón, E. Jiménez, M. You, B. Gulbakan, and W. Tan, "Aptamers: turning the spotlight on cells," *Wiley Interdisciplinary Reviews: Nanomedicine and Nanobiotechnology*, vol. 3, no. 3, pp. 328–340, 2011.
- [10] R. Yamamoto, M. Katahira, S. Nishikawa, T. Baba, K. Taira, and P. K. R. Kumar, "A novel RNA motif that binds efficiently and specifically to the Tat protein of HIV and inhibits the trans-activation by Tat of transcription in vitro and in vivo," *Genes to Cells*, vol. 5, no. 5, pp. 371–388, 2000.
- [11] A. Marozzi, R. Meneri, M. Giacca, M. I. Gutierrez, A. G. Siccardi, and E. Ginelli, "In vitro selection of HIV-1 TAR variants by the Tat protein," *Journal of Biotechnology*, vol. 61, no. 2, pp. 117–128, 1998.
- [12] J. P. Davis, N. Janjić, B. E. Javornik, and D. A. Zichi, "Identifying consensus patterns and secondary structure in SELEX sequence sets," *Methods in Enzymology*, vol. 267, pp. 302–312, 1996.
- [13] J. Tao and A. D. Frankel, "Arginine-binding RNAs resembling TAR identified by in vitro selection," *Biochemistry*, vol. 35, no. 7, pp. 2229–2238, 1996.
- [14] M. Guyader, M. Emerman, and P. Sonigo, "Genome organization and transactivation of the human immunodeficiency virus type 2," *Nature*, vol. 326, no. 6114, pp. 662–669, 1987.
- [15] M. A. Muesing, D. H. Smith, and D. J. Capon, "Regulation of mRNA accumulation by a human immunodeficiency virus trans-activator protein," *Cell*, vol. 48, no. 4, pp. 691–701, 1987.
- [16] B. Berkhout, R. H. Silverman, and K. T. Jeang, "Tat transactivates the human immunodeficiency virus through a nascent RNA target," *Cell*, vol. 59, no. 2, pp. 273–282, 1989.
- [17] C. Dingwall, I. Ernberg, M. J. Gait et al., "HIV-1 tat protein stimulates transcription by binding to a U-rich bulge in the stem of the TAR RNA structure," *EMBO Journal*, vol. 9, no. 12, pp. 4145–4153, 1990.
- [18] J. Tao and A. D. Frankel, "Specific binding of arginine to TAR RNA," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 89, no. 7, pp. 2723–2726, 1992.
- [19] Y. Yang, M. Kochoyan, P. Burgstaller, E. Westhof, and M. Famulok, "Structural basis of ligand discrimination by two related RNA aptamers resolved by NMR spectroscopy," *Science*, vol. 272, no. 5266, pp. 1343–1347, 1996.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

