J. Biochem. 122, 878-882 (1997)

Identification of the Nucleotides in the A-Rich Bulge of the *Tetrahymena* Ribozyme Responsible for an Efficient Self-Splicing Reaction[1]

Yoshiya Ikawa,* Ayako Okada,* Hiroshi Imahori,† Hideaki Shiraishi,* and Tan Inoue*,[2]

*Department of Chemistry, Faculty of Science, Kyoto University, Sakyo-ku, Kyoto 606-01; and †The Institute of Scientific and Industrial Research, Osaka University, Mihoga-oka, Ibaraki, Osaka 657

Received for publication, June 25, 1997

P5abc is a large extension of the P5 element characteristic of subclasses IC1 and IC2 of group I introns. It has a conserved region termed the A-rich bulge, that is responsible for activation of the *Tetrahymena* self-splicing intron. By employing a modified color-colony assay system, we identified four adenosines in the bulge that are responsible for an efficient splicing reaction. On comparison with the X-ray crystal structure of the P4-5-6 domains of the *Tetrahymena* intron, three adenosines at positions 183, 184, and 186 were found to be identical to those significantly contributing to the formation of its tertiary structure. However, our results show that an adenosine at 187 is involved in the formation of a Watson-Crick base pair with U135, although it forms a Hoogsteen base pair in the crystal structure.

Key words: A-rich bulge, *in vivo* selection, P5abc, ribozyme, *Tetrahymena*.

The *Tetrahymena* LSU intron, which belongs to subclass IC1 of the group I introns, is capable of performing efficient splicing reactions in the complete absence of protein molecules (*1, 2*). The intron consists of essential core elements and several non-essential peripheral elements (Fig. 1A) (*3-6*). P5abc, a large extension of the P5 element that is commonly found in subclasses IC1 and IC2 (*2*), is an extensively studied peripheral element. A distinctive property of this domain is its function as an independent activator RNA (*7*). A mutant *Tetrahymena* intron lacking P5abc is hardly capable of splicing under standard conditions, but can be activated through the formation of a RNA-RNA complex with separately prepared P5abc RNA. It has also been revealed that the folding of the P5abc domain requires a lower magnesium concentration than the folding of the rest of the intron (*8*).

A structural element termed the A-rich bulge exists within the P5abc domain (Fig. 1B). Its size and distance from the P4 element in the predicted secondary structure model are highly conserved (*2, 9*). A mutant *Tetrahymena* intron lacking the bulge exhibits very weak activity similar to that seen for the mutant lacking P5abc (*10, 11*). Several studies involving mutants with base-substitutions in the bulge have indicated that the bulge is important for the folding and activation of the intron (*10, 12*). Recently, the crystal structure of P4-5-6 RNA, that is a part of the *Tetrahymena* intron including P5abc and is capable of forming a stable structure independent from the rest of the intron (*11*), has shown that the bulge is a key region for the folding of the RNA (*13*).

The studies, however, were incomplete for determining the contribution of individual nucleotides in the bulge required for the self splicing reaction. It remains to be determined whether the structure of the P4-5-6 RNA is identical to its catalytically active form. To elucidate the roles of the nucleotides in the activity, those required for a highly efficient splicing reaction in *Escherichia coli* were determined by using a modified color-colony assay system (*14-18*).

MATERIALS AND METHODS

General Methods and Materials—[α-^{32}P]ATP was obtained from ICN. T7 RNA polymerase was prepared as described by Davanloo *et al.* (*19*). Other enzymes were purchased from Takara, Toyobo and New England Biolabs. *E. coli* JM109 competent cells were purchased from Toyobo. Oligonucleotides were synthesized by Sawady Technology (Tokyo). Sequencing was performed with a *Bca* BEST™ dideoxy sequencing kit (Takara).

Construction of Variant Pools—A plasmid, pTZ⊿Arb, encoding a mutant *Tetrahymena* intron lacking the A-rich bulge (positions 183-188) was constructed from pTZIVSU (*18*) by use of PCR (*26*). *Eco*NI-digested pTZ⊿Arb (0.5 ng) was used as the PCR template to construct a plasmid pool containing variant *Tetrahymena* introns with randomized bases in the A-rich bulge (*26*). To introduce mixed bases in place of the wild type sequence, the following sets of primers were employed. For variant pool 1: Y48 (5′-ATA CCC TTT GCA AGG CCA TCT CAA AGT TTT-3′) and Y50 (5′-GGT NNN NNN CTG ACG GAC ATG GTC CTA A-3′). For variant pool 2: Y48 and Y51 (5′-GGT NXN AXN CTG ACG GAC ATG GTC CTA A-3′). For variant pool 3: Y52 (5′-AAG TTT CCC CTG AGA CTT GGT NCT GAA CGG

[1] This work was supported by Grants-in-Aid for Scientific Research on Priority Areas and the Encouragement of Young Scientists from the Ministry of Education, Science, Sports and Culture, Japan, and the Tokuyama Foundation.

[2] To whom correspondence should be addressed. Tel: +81-75-753-3995, Fax: +81-75-753-3996, E-mail: tan@kuchem.kyoto-u.ac.jp

CTG TTG ACC-3′) and Y53 (5′-TGA GAT GGC CTT GCA AAG GGT ATG GTA GTA NGC TGA CGG ACA T-3′) (N and X represent mixed bases consisting of equal amounts of A, G, T, and C or mixtures of G, T, and C, respectively.)

Preliminary In Vitro Selection of Variant Pool 1—Preliminary selection of variant pool 1 was carried out using the *in vitro* selection system described previously (*20*). To generate the 5′-half RNA for the bimolecular self-ligation reactions, PCR with primers D8 and D6 (*20*) was performed using plasmids from variant pool 1 as templates.

Color Colony Assay Involving E. coli (18)—E. coli JM109 strain was employed for transformation, followed by plating on LB agar plates containing ampicillin (50 μg/ml), IPTG (100 μg/ml), and X-gal (40 μg/ml). The plates were incubated at 37°C for 16 h, and then stocked at 4°C for further color development. Color scoring was performed 24 h and one week after plating. Clones were divided into three classes according to the degree of the blue color. Colonies exhibiting a blue color within 24 h after plating were designated as blue. Colonies not initially blue (within 24 h) but which turned blue within one week were designated as pale blue. Colonies remaining white after one week were designated as white.

RESULTS AND DISCUSSION

Color-Selection System for Analysis of the A-Rich Bulge—We employed a selection system for analysis of the A-rich bulge involving the randomization-selection technique (*21*) combined with a color colony assay method employing *E. coli* (*18*). Pools of variant *Tetrahymena* introns containing randomized nucleotides in the A-rich bulge were subjected to selection. With the system, clones containing active variant introns that produce functional β-galactosidase α-fragments as the result of the splicing reaction can be distinguished by their ability to form blue colonies on α-complementation on agar plates containing X-gal. Since the self-splicing of the *Tetrahymena* intron in *E. coli* proceeds as fast as that in the *Tetrahymena* nucleus (*22*), the splicing conditions in *E. coli* might be related to the physiological conditions in *Tetrahymena*.

Selection of Active Variants from Variant Pool 1—We first attempted to select active introns from a pool of variant introns in which the six bases of the A-rich bulge (positions 183-188) were replaced with six completely randomized bases; the variant pool (termed pool 1) contains $4^6 = 4{,}096$ different sequences (Fig. 1B). Prior to the selection in *E. coli*, three rounds of *in vitro* selection were carried out using our previously reported selection system (*20*) to discard inactive variants. The resulting pool was cloned into a pTZIVSU vector (*20*), followed by introduction into *E. coli*. After plating, colonies harboring active variants were selected according to their color and grouped into three classes, blue (active), pale blue (weakly active), and white (inactive) colonies (see "MATERIAL AND METHODS"). Among over 500 colonies, 8-10% were blue, and approximately 50% of the remaining colonies were pale blue and the rest were white.

Randomly selected fourteen clones from each class were subjected to sequencing (Fig. 3). The identified sequences were divided into six groups on the basis of the presence of adenosine residues at particular sites, as shown in Fig. 2.

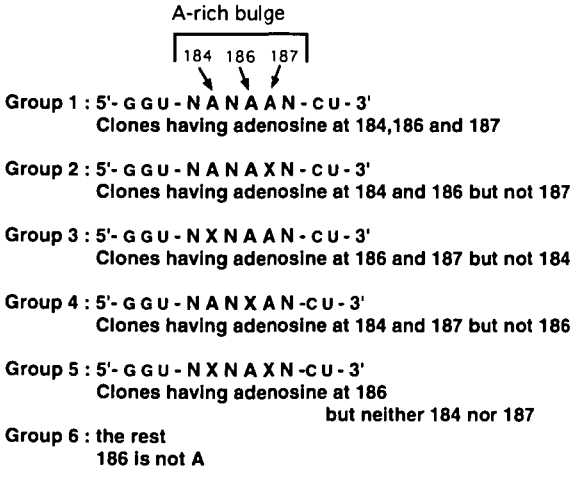

Fig. 2. **Groups of variant *Tetrahymena* introns classified according to the sequences of their A-rich bulges.** N represents nucleotide A, G, C, or U. X indicates nucleotide G, C, or U.

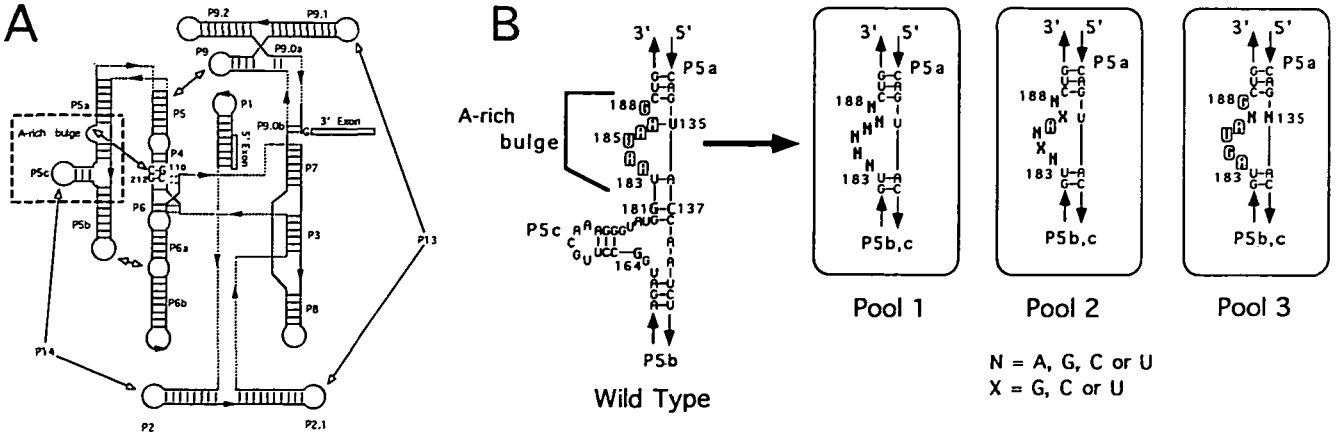

Fig. 1. **A: Schematic representation of the secondary structure of the *Tetrahymena* group I intron ribozyme (*24, 25*).** Structural elements composed of Watson-Crick and G-U wobble base-pairs are denoted by P*i*. Arrows with white arrowheads indicate long-range interactions. **B: Nucleotide sequences of the A-rich bulge region of the wild type *Tetrahymena* ribozyme, variant pools 1, 2, and 3.**

The *in vitro* splicing activities of the clones from all three classes were examined and found to correlate well with their *in vivo* (*E. coli*) activities (data not shown, *14-18*). The relationship between the sequences and colors of the colonies is summarized in Table I.

The clones having three adenosines at positions 184, 186, and 187 (group 1) were found among the blue (active) colonies without exception. In contrast, these three adenosines were not present in the weakly active (pale blue) or the inactive (white) clones (Fig. 3 and Table I). This suggests that the sequence represented by 5′-NANAAN-3′ (where N is A, G, C, or U) is the consensus sequence sufficient for an efficient splicing in *E. coli*. To test this hypothesis, we conducted the color-colony assay by employing a variant pool of introns that contained the three adenosines and randomized nucleotides (5′-NANAAN-3′, where N is A, G, C, or U) in the A-rich bulge. As anticipated, more than 80% of the colonies exhibited the blue color.

Importance of A186—All clones from the blue colonies had an adenosine at position 186 (Fig. 3), suggesting that A186 is the most critical nucleotide for the activation. This is also supported by the fact that A186 is completely conserved among naturally existing A-rich bulges. To verify the role of A186, mutant introns having G, C, or U at position 186 were prepared and then subjected to the color-colony assay. As expected, their activities were distinctively weaker than that of the wild type (*10*), and thus they were placed in the pale blue (weakly active) category (data not shown).

According to the structure of the P4-P5-P6 domains determined by X-ray crystallography (*13*), the A-rich bulge interacts with the P4 helix as well as with the three-helix junction of the P5abc domain (consisting of P5a, P5b, and P5c). In the crystal structure, A186 interacts with three nucleotides (C137, G181, and G164, see Fig. 1B) in the three-helix junction *via* as many as five hydrogen bonds, suggesting that the hydrogen bonds play important roles in an efficient splicing reaction.

Selection of Active Variants from Variant Pool 2—In addition to A186, 13 of the 14 blue colony clones had adenine(s) at position 184 and/or 187 (Fig. 3 and Table I), suggesting that two adenosines are also responsible for the activation. The contribution of these adenosines is also suggested by the fact that 5 of 6 clones in group 4 that possess the adenosines are weakly active (pale blue), albeit in the absence of A186 (Fig. 3 and Table I).

We investigated whether A184 and A187 are essential for efficient splicing. We performed selection using a newly designed pool of variant introns (termed pool 2, Fig. 1B) having A186, that is essential for efficient splicing (Fig. 3). A184 and A187 were replaced with mixed nucleotides consisting of G, C, and U, and the remaining three nucleotides (positions 183, 185, and 188) were replaced with completely randomized nucleotides. Without preliminary selection *in vitro*, active variants in pool 2 (consisting of 576 variants) were directly selected in *E. coli*. The sequences of randomly picked clones from blue colonies are shown in Fig. 4.

An adenosine at a position corresponding to position 183 of the *Tetrahymena* intron can be seen highly frequently among naturally occurring A-rich bulges (*2, 23*). No such conservation was observed on the selection from pool 1 (Fig. 3). However, from pool 2, 11, of 14 clones had an adenosine at position 183 and the rest had a guanosine. This

TABLE I. **Distribution of the blue (active), pale-blue (weakly active), and white (inactive) clones from variant pool 1 in sequence groups.**

Colony color	Sequence groups					
	1	2	3	4	5	6
Blue (active)	5	4	4	0	1	0
Pale blue (weakly active)	0	1	3	5	3	2
White (inactive)	0	0	0	1	3	10

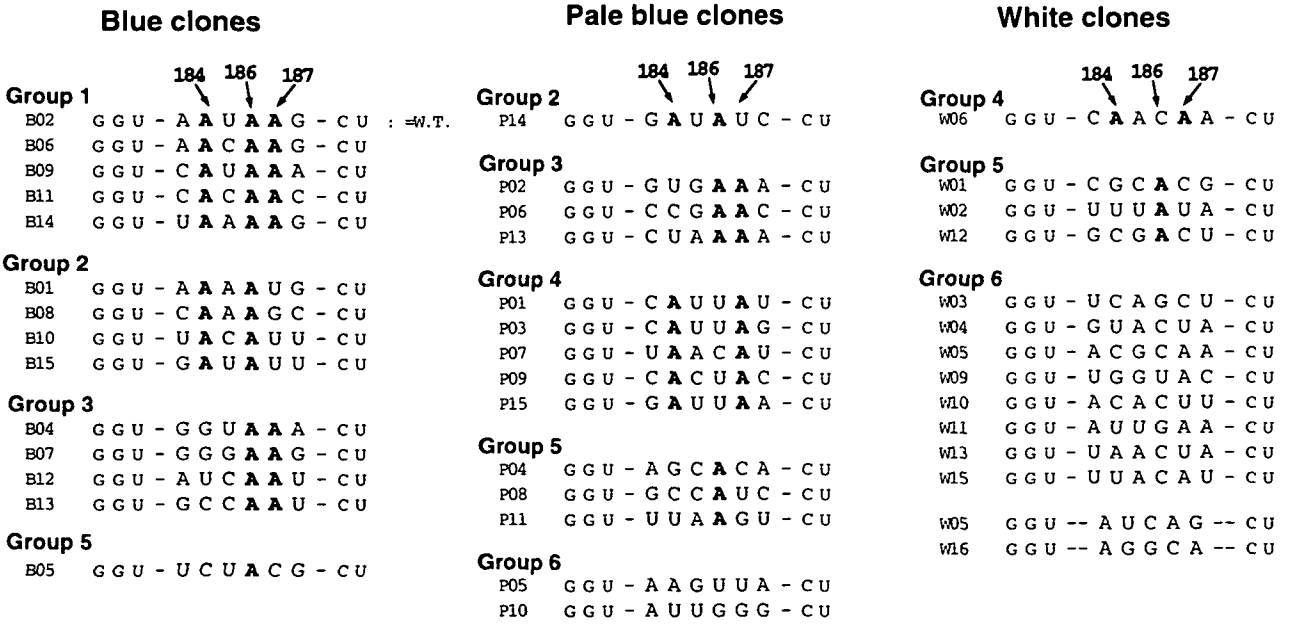

Fig. 3. **Sequence alignment of the A-rich bulge regions (positions 183-188) of individual clones isolated from variant pool 1.** The highly conserved adenines (A184, A186, and A187) are indicated in boldface.

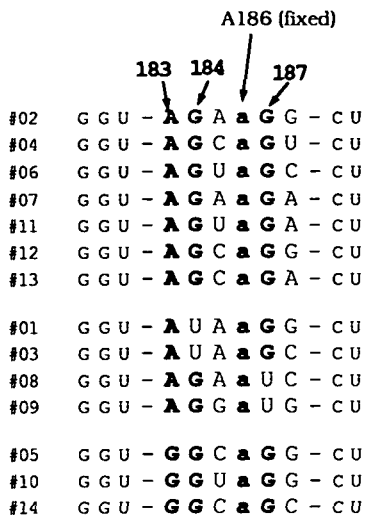

Fig. 4. **Sequence alignment of the A-rich bulge regions (positions 183-188) of individual clones isolated from blue colonies from variant pool 2.** The highly conserved nucleotides (183, 184, and 187) are indicated in bold capitals. The fixed adenine (A186) in the A-rich bulge is indicated in a bold lowercase letter.

TABLE II. **Phylogeny of the naturally existing A-rich bulges of 47 introns of subclasses IC1 and IC2 (*23*).** Relationship of the nucleotides corresponding to U135 and A187 of the *Tetrahymena* intron.

Position 135	Position 187			
	A	G	U	C
A	0	0	4	0
G	0	0	0	2
U	35	0	0	0
C	0	6	0	0

suggests that A183 is dispensable for an efficient splicing reaction if either A184 or A187 is present. In other words, in the absence of A184 and A187, it seems that A183 becomes indispensable for an efficient splicing reaction.

In the crystal structure, A183 forms two hydrogen bonds through N3 of its adenine residue and O2′ of its ribose. Because the hydrogen bonds of A183 in the crystal structure are not significantly affected by its replacement with a guanosine, it is difficult to explain why position 183 showed a preference for an adenosine to a guanosine in our selection (Fig. 4), like in the case of naturally existing bulges (*2, 23*).

Of the 14 active clones from pool 2, 10 had guanosines at both positions 184 and 187, the remaining 4 having a guanosine at either position, indicating that the function(s) of A184 and A187 can be partially substituted by guanosines. A184, that is absolutely conserved among naturally existing A-rich bulges, forms three hydrogen bonds in the crystal structure of the P4-P5-P6 domains. Two of the hydrogen bonds can be retained if A184 is replaced with a guanosine (*13*). However, it is difficult to explain the fact that the function of A187 can be partially substituted by a guanosine. Because A187 forms a Hoogsteen base pair with U135 in the crystal structure, a guanosine does not seem to form the pair. Phylogenetic analysis of naturally existing A-rich bulges revealed that the nucleotides equivalent to A187 and U135 form a Watson-Crick base pair with no exception (Table II). The fact that the function of A187 can be partially substituted with a guanosine suggests that the A187-U135 base pair can be partially replaced by a G-U wobble pair.

TABLE III. **Phylogeny of the A-rich bulges of 21 active clones from pool 3.** Relationship of the nucleotides at positions 135 and 187 of the *Tetrahymena* intron.

Position 135	Position 187			
	A	G	U	C
A	1	1	7	0
G	0	0	0	2
U	4	1	0	0
C	0	5	0	0

Relationship between U135 and A187—To analyze more extensively the relationship of the nucleotides at positions 135 and 187, we constructed variant pool 3, that consisted of variant introns having the sequence, 5′-AGUANG-3′ (where N is A, G, U, or C at position 187), in place of the wild type bulge and a randomized nucleotide at position 135 (Fig. 1B), and attempted selection. Active variants from pool 3 (consisting of 16 variants) were directly selected in *E. coli*. Approximately 20% of the total colonies showed the blue color. Twenty-one active clones were randomly picked and the nucleotides at their positions 135 and 187 are summarized in Table III. Of the 21 clones, 19 can form a Watson-Crick or G-U Wobble base pair between positions 135 and 187, suggesting that the naturally observed phylogeny (Table II) is functionally important in the mechanism of the activation underlying the group I intron ribozymes.

We wish to thank Dr. Ruth Yu for critical reading of the manuscript.

REFERENCES

1. Kruger, K., Grabowski, P.J., Zaug, A.J., Sands, J., Gottschling, D.E., and Cech, T.R. (1982) Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA interventing sequence of *Tetrahymena*. *Cell* **31**, 147-157
2. Michel, F. and Westhof, E. (1990) Modelling of the three-dimensional architecture of group I catalytic introns based on comparative sequence analysis. *J. Mol. Biol.* **216**, 585-610
3. Jaeger, L., Michel, F., and Westhof, E. (1996) The structure of group I ribozymes in *Catalytic RNA* (Eckstein, F., and Lilley, D.M.J., eds.) pp. 33-52, Springer, Berlin
4. Joyce, G.F., van der Horst, G., and Inoue, T. (1989) Catalytic activity is retained in the *Tetrahymena* group I intron despite removal of the large extension of element P5. *Nucleic Acids Res.* **17**, 7879-7889
5. Beaudry, A.A. and Joyce, G.F. (1990) Minimum secondary structure requirements for catalytic activity of a self-splicing group I intron. *Biochemistry* **29**, 6534-6539
6. Doudna, J.A. and Szostak, J.W. (1989) Miniribozymes, small derivatives of the *sunY* intron, are catalytically active. *Mol. Cell. Biol.* **9**, 5480-5483
7. van der Horst, G., Christian, A., and Inoue, T. (1991) Reconstitution of a group I intron self-splicing reaction with an activator RNA. *Proc. Natl. Acad. Sci. USA* **88**, 184-188
8. Celander, D.W. and Cech, T.R. (1991) Visualizing the higher order folding of a catalytic RNA molecule. *Science* **251**, 401-407
9. Collins, R.A. (1988) Evidence of natural selection to maintain a functional domain outside of the 'core' in a large subclass of group I introns. *Nucleic Acids Res.* **16**, 2705-2715
10. Pace, U. and Szostak, J. W. (1991) Mutations in a semiconserved region of the *Tetrahymena* intron. *FEBS Lett.* **280**, 171-174
11. Murphy, F.L. and Cech, T.R. (1993) An independently folding domain of RNA tertiary structure within the *Tetrahymena*

ribozyme. *Biochemistry* **32**, 5291-5300
12. Murphy, F.L. and Cech, T.R. (1994) GAAA tetraloop and conserved bulge stabilize tertiary structure of a group I intron domain. *J. Mol. Biol.* **236**, 49-63
13. Cate, J.H., Gooding, A.R., Podell, E., Zhou, K., Golden, B.L., Kundrot, C.E., Cech, T.R., and Doudna, J.A. (1996) Crystal structure of a group I ribozyme domain: principles of RNA packing. *Science* **273**, 1678-1685
14. Price, J.V. and Cech, T.R. (1985) Coupling of *Tetrahymena* ribosomal RNA splicing to beta-galactosidase expression in *Escherichia coli*. *Science* **228**, 719-722
15. Waring, R.B., Ray, J.A., Edwards, S.W., Scazzocchio, C., and Davies, R.W. (1985) The *Tetrahymena* rRNA intron self-splices in *E. coli*. *Cell* **40**, 371-380
16. Been, M.D. and Cech, T.R. (1986) One binding site determines sequence specificity of *Tetrahymena* pre-rRNA self-splicing, trans-splicing, and RNA enzyme activity. *Cell* **47**, 207-216
17. Flor, P.J., Flanegan, J.B., and Cech, T.R. (1989) A conserved base pair within helix P4 of the *Tetrahymena* ribozyme helps to form the tertiary structure required for self-splicing. *EMBO J.* **8**, 3391-3399
18. Williamson, C.L., Desai, N.M., and Burke, J.M. (1989) Compensatory mutations demonstrate that P8 and P6 are RNA secondary structure elements important for processing of a group I intron. *Nucleic Acids Res.* **17**, 675-689
19. Davanloo, P., Rosenberg, A.H., Dunn, J.J., and Studier, S.W. (1984) Cloning and expression of the gene for bacteriophage T7 RNA polymerase. *Proc. Natl. Acad. Sci. USA* **81**, 2035-2039
20. Williams, K.P., Imahori, H., Fujimoto, D.N., and Inoue, T. (1994) Selection of novel forms of a functional domain within the *Tetrahymena* ribozyme. *Nucleic Acids Res.* **22**, 2003-2009
21. Williams, K.P. and Bartel, D.P. (1996) *In vitro* selection of catalytic RNA in *Catalytic RNA* (Eckstein, F. and Lilley, D.M.J., eds.) pp. 367-382, Springer, Berlin
22. Zhang, F., Ramsay, E.S., and Woodson, S.A. (1995) *In vivo* facilitation of *Tetrahymena* group I intron splicing in *Escherichia coli* pre-ribosomal RNA. *RNA* **1**, 284-292
23. Damberger, S.H. and Gutell, R.R. (1994) A comparative database of group I intron structures. *Nucleic Acids Res.* **22**, 3508-3510
24. Cech, T.R., Damberger, S.H., and Gutell, R.H. (1994) Representation of the secondary and tertiary structure of group I introns. *Nature Struct. Biol.* **1**, 273-280
25. Lehnert, V., Jaeger, L., Michel, F., and Westhof, E. (1996) New loop-loop interactions in self-splicing introns of subgroup IC and ID. *Chem. Biol.* **3**, 993-1009
26. Imai, Y., Matsushima, Y., Sugimura, T., and Terada, M. (1991) A simple and rapid method for generating a deletion by PCR. *Nucleic Acids Res.* **19**, 2785