

A Deoxyribozyme That Synthesizes 2',5'-Branched RNA with Any Branch-Site Nucleotide

Elizabeth D. Pratico¹, Yangming Wang¹ and Scott K. Silverman*

Department of Chemistry and ¹Department of Biochemistry,
University of Illinois at Urbana-Champaign, 600 South Mathews Avenue, Urbana, Illinois 61801

Table of Contents

| | |
|--|-------------|
| Data for changes to L substrate sequences | page S2 |
| Changes to L substrate 3'-overhang sequence..... | page S2 |
| Sequences of the proposed Ty1 branched RNA | pages S3–S4 |
| Verification of branch-site locations by partial alkaline hydrolysis..... | page S4 |
| Hairpin approach for synthesis of the proposed Ty1 branched RNA..... | page S4 |
| Optimizing the disruptor approach for synthesis of the proposed Ty1 branched RNA | pages S4–S6 |

Data for changes to L substrate sequences

The data sets from which the Figure 5 entries were derived are shown in Figure X1.

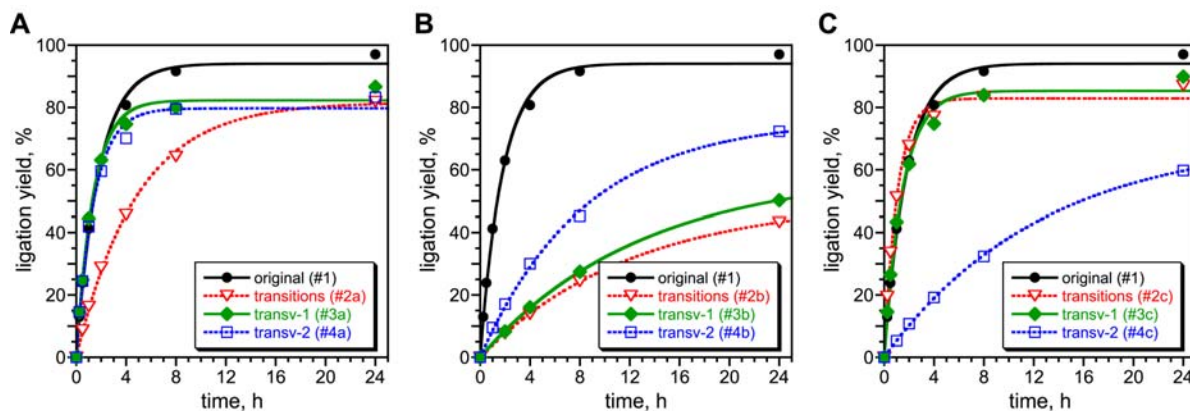


Figure X1. Effects on 6CE8 ligation activity of changes to the L substrate sequence for RNA nucleotides that interact with the DNA binding arms. Note that the data is organized differently from that in Figure 5. Here, panel A shows the effects of nucleotide changes up to and including position -2 ; panel B shows the effects of nucleotide changes up to and including position -1 ; and panel C shows the effects of nucleotide changes up to and including position -1 in the RNA but -2 in the DNA while retaining G at -1 in the DNA. The full sequences are shown in Figure 5. Incubation conditions: 50 mM HEPES, pH 7.5, 150 mM NaCl, 2 mM KCl, and 20 mM MnCl_2 at 37 °C.

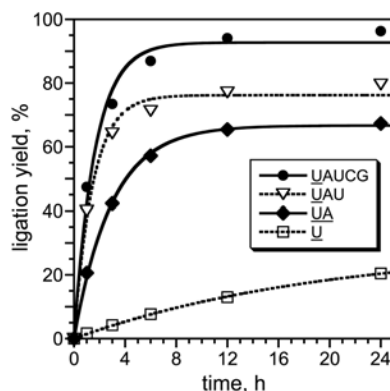
Changes to L substrate 3'-overhang sequence

Figure X2. Effects on 6CE8 ligation activity of changes to the L substrate 3'-overhang sequence. The branch-site \underline{U} and all overhanging RNA nucleotides to its 3'-side are shown. Incubation conditions were as listed in the Figure X1 caption. k_{obs} values for this particular experiment (top to bottom, h^{-1}): 0.61, 0.68, 0.34, and 0.051.

Sequences of the proposed Ty1 branched RNA

In Figures 8 and 9 are shown partial sequences for the substrates used to prepare the proposed Ty1 branched RNA. Complete sequence information for the substrates is provided here. The full left-hand (L) substrate sequence is as follows, written 5' to 3':

```

gguauguugg aauagaaauc aacuaucauc uacuaacuag uauuuacaau acuaguauau
uaucauauac gguguuagaa gaugacgcaa augaugagaa auagucaucu aaauuagugg
aagcugaaac gcaaggauug auaauguaau aggaucaaug aaauaaaaca uauaaaauga
ugauaaaaau auuuauagaa uuguguagaa uugcagauuc ccuuuuauugg auuccuaaau
ccuUgaggag aacuucuagu auauucugua uaccuaauau uauagccuuu aucaacaaug

```

This 300-nt sequence is derived from the 3' Ty1 RNA sequence (J. Boeke et al., *Cell* **1985**, *40*, 491-500; GenBank accession number M18706). Nucleotides 5581–5880 of the Ty1 RNA are shown using the numbering from the cited *Cell* manuscript. The color scheme follows that used in Figure 8. The branch-site uridine (nt 5824) proposed in ref. 4 is in boldface. Nucleotides 5585–5824 (red, 240 nt) correspond to the Ty1 U3 region, where the proposed branch-site U is the last nt of the U3 region. Embedded within the U3 region is the CYC3 sequence (pink italics), identified empirically as pairing with the CYC5 sequence in U5 (see R sequence below; G. Cristofari et al., *EMBO J.* **2002**, *21*, 4368-4379). The first four nt GGUA of the listed L sequence (nt 5581–5584) naturally are found just to the 5'-side of the U3 region and were included here to permit transcription of L by T7 RNA polymerase. These are the last four nt of the polypyrimidine tract (PPT) identified in the cited *EMBO J.* manuscript. Nucleotides 5825–5880 (green, 56 nt) correspond to the Ty1 R (“repetitive”) region, which should not be confused with our abbreviation “R substrate” for “right-hand substrate”. Underlined within the red Ty1 U3 region are the 16 nucleotides retained for the “short L arm” experiments of Figure 8; these correspond to the nucleotides of the core of the proposed branch, including the proposed branch-site uridine. Coincidentally, this particular stretch of nucleotides begins with 5'-GG..., thereby permitting T7 RNA polymerase transcription without the need to append any unnatural nucleotides. Underlined within the green Ty1 R region are the four nucleotides GAGG retained for the “short L tail” experiments of Figure 8. The length of the full L substrate is 4 + 240 + 56 = 300 nt. For the L substrate with long arm but short tail, the length is 4 + 240 + 4 = 248 nt. For the L substrate with short arm and short tail, the length is 16 + 4 = 20 nt.

The full right-hand (R) substrate sequence is as follows, written 5' to 3':

```

gaggagaacu ucuaguauau ucuguauacc uauuuuuaua gccuuuauca acaauggaau
cccaacaauu aucucaacau ucacccaauu cucaUGGUAG CGCCugugcu ucgguuacuu
cuaaggaagu ccacacaaau caagauccgu uagacguuuc agcuuccaaa acagaagaau
gugagaaggc uuccacuaag gc

```

This 202-nt sequence is derived from the 5' Ty1 RNA; nucleotides 241–442 are shown. The color scheme follows that used in Figure 8. Nucleotides 241–296 (blue, 56 nt) correspond to the Ty1 R region; this sequence is identical to the green portion of the L sequence shown above. Underlined within the blue Ty1 R region are the 17 nucleotides retained for the “short R arm” experiments of Figure 8. Nucleotides 297–334 (purple, 38 nt) correspond to the Ty1 U5 region. Immediately 3' of the U5 region is the primer binding site (PBS), comprising nt 335–344 (underlined purple, 10 nt). Continuing 3' of the U5/PBS region are nt 345–442 (98 nt), with an arbitrary ending point. Within this 3'-most region are box 0, box 1, and box 2.1 (each underlined), as well as the CYC5 sequence (pink italics, encompassing

box 2.1) identified empirically as pairing with the CYC3 sequence in U3. The length of the full R substrate is $56 + 38 + 10 + 98 = 202$ nt. For the R substrate with short arm, the length is 17 nt.

Verification of branch-site locations by partial alkaline hydrolysis

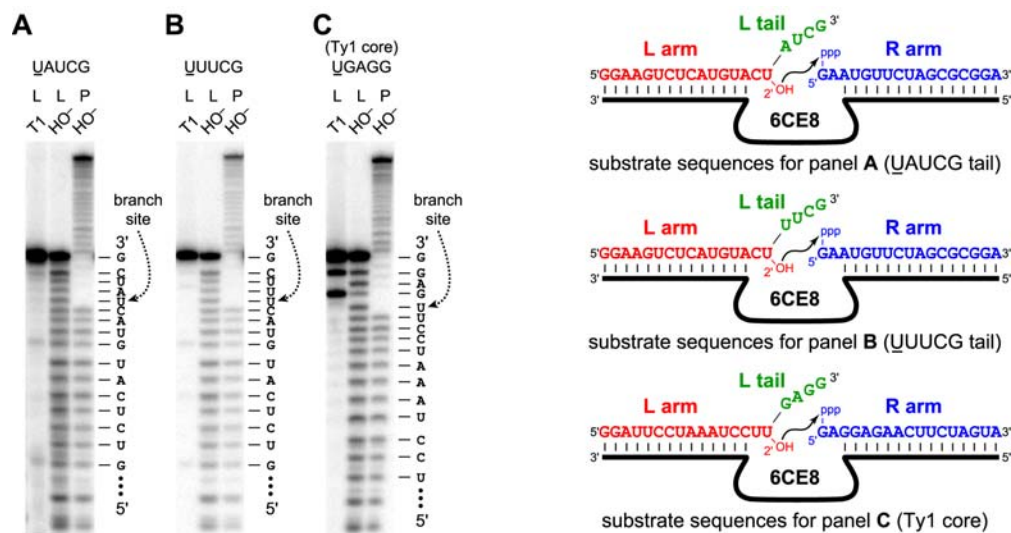


Figure X3. Verifying branch-site locations for the 6CE8 ligation products. (A) Product from the selection substrates, where the left-hand substrate has tail sequence UAUCG (U = branch site). This is a repeat of the experiment shown in Figure 4B (branch-site X = U); note that the spurious cleavage band noted in the Figure 4B caption is not observed here. (B) Product from the selection substrates, where the left-hand substrate has tail sequence UUUCG (U = branch site). The branch-site U is maintained despite the potential ambiguity of three adjacent uridines. (C) Product from the Ty1 core sequences. The ligation product is from Figure 8B, first set of lanes. The branch-site U is as expected.

Hairpin approach for synthesis of the proposed Ty1 branched RNA

The hairpin approach shown in Figure 9A was implemented by extending the L substrate with the 35-nt RNA sequence ...CUCCAGCCAUCCAGAACAGAAUAUCUAGAAGUUC-3'. The underlined portion forms the stem with the 5'-portion of the L tail. The new L substrate is 335 nt long; the R substrate remains 202 nt long; and the initially formed branched product is therefore 537 nt long. The 10–23 deoxyribozyme used to cleave the branched ligation product formed via the hairpin approach was 5'-CTGGATGGCTGGAGGGCTAGCTACAACGAATTGTTGATAAAGGCT-3', with the binding arm sequences underlined. Cleavage by 10–23 led to the final proposed Ty1 branched RNA, which is 502 nt long.

Optimizing the disruptor approach for synthesis of the proposed Ty1 branched RNA

The disruptor approach shown in Figure 9B can be implemented in several variations, depending on the choice of disruptor sequence. We optimized the disruptor oligonucleotide by varying the number of mismatches and G•T or U•G wobble pairs between the L tail and the disruptor oligo (Figure X4A). Five different disruptors 1–5 were designed with progressively less complementarity between the RNA and DNA, seeking a balance between the disruptor binding to the L tail and the disruptor *not* binding to the R arm. Disruptor 1 is fully complementary to both the L tail sequence and the extension sequence appended at the 3'-end of the L tail. Disruptor 2 is derived from disruptor 1 by introducing 8 mismatches in the middle of the RNA:DNA complementarity region (the extension sequence remains fully complemented, and the portion of the L tail closest to the ligation site also remains fully

complemented). Disruptor 3 is similarly derived from disruptor 1 by introducing 16 mismatches. Disruptor 4 is a variant in which 6 mismatches and 14 wobble pairs are introduced. Finally, disruptor 5 has numerous mismatches and wobble pairs extending throughout the L tail region, with only the extension region fully complemented.

All five disruptors were tested individually to optimize the ligation yield, using either 1, 3, or 6 equivalents of disruptor oligo relative to the L substrate (1×, 3×, or 6×). All five disruptors measurably improved the yield relative to the absence of disruptor (0×; Figure X4B). However, clear differences were observed for efficacy among the disruptor variants. The fully complementary disruptor 1 improved the ligation yield to nearly half that found with only a short L tail as the positive control; however, excess disruptor 1 was detrimental (compare 6× versus 3×). Disruptor 2 was somewhat less effective than disruptor 1 although not harmed by a modest excess of disruptor, and in turn disruptor 3 was less effective than disruptor 2. Disruptor 4 was approximately as effective as disruptor 1 yet not as sensitive to using excess disruptor (note that 6× was better than 3×, unlike the case for disruptor 1). Finally, disruptor 5 was the least effective of those tested. Overall, disruptor 4 was determined to be the most effective of those tested, and it was used at 6× concentration relative to the L substrate in the experiments of Figure 9.

For the disruptor approach (regardless of the particular disruptor sequence), the L substrate was extended with the 38-nt sequence ...CGGCCAAGGUGGCCGACAGGAGCCUGUCUACACUCUG-3', as shown in Figure X4. The new L substrate is 338 nt long; the R substrate remains 202 nt long; and the initially formed branched product is therefore 540 nt long. The 10–23 deoxyribozyme used to cleave the ligation product formed via the disruptor approach was 5'-CACCTTGGCCGGGCTAGCTACAACGAATTGTTGATAAAGGC-3'. Cleavage by 10–23 led to the final proposed Ty1 branched RNA, which is 502 nt long.

A



B

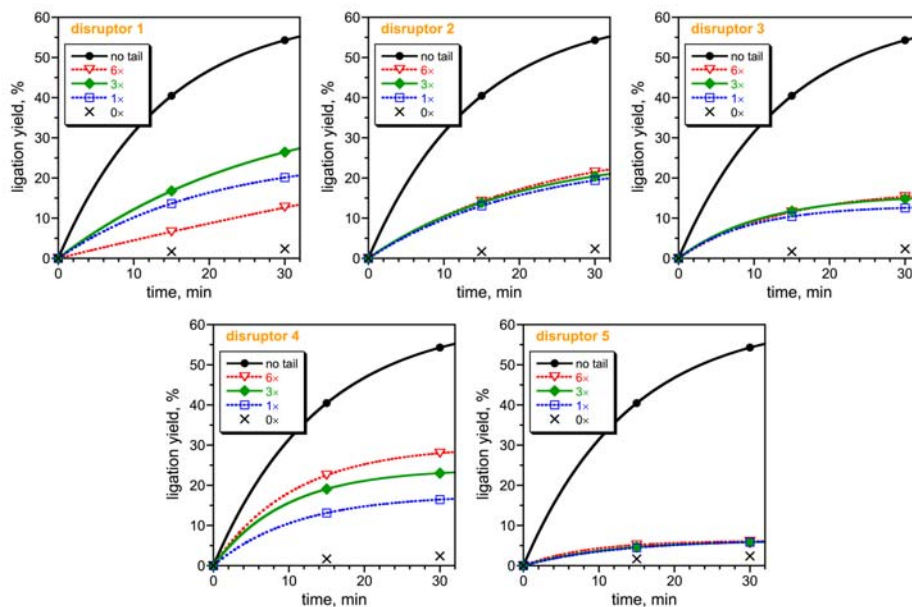


Figure X4. Sequences and testing of the DNA disruptor oligonucleotides. (A) Sequences of the five tested disruptors (orange), showing mismatches (red) and wobble pairs (purple). (B) Testing each disruptor alongside a positive control assay (L substrate with only a short 4-nt tail; no disruptor present or needed, denoted “no tail” in the plots) and a negative control assay (long L tail with no disruptor included, denoted 0x in the plots). Each disruptor D was tested with L:E:R:D = 1:3:6:x, where $x = 1, 3, \text{ or } 6$ (denoted 1x, 3x, and 6x in the plots).