Angewandte Chemie
Eine Zeitschrift der Gesellschaft Deutscher Chemiker

Supporting Information

69451 Weinheim, Germany

Efficient One-Step Synthesis of Biologically Related Lariat RNAs by a Deoxyribozyme

Yangming Wang and Scott K. Silverman*

Department of Chemistry, University of Illinois at Urbana-Champaign, 600 South Mathews Avenue, Urbana, IL 61801 (USA)

RNA substrates and lariat products

The YBL059W (yeast), β-globin IVS1 (human), and ACT1 (yeast) lariat RNAs were prepared from linear RNA substrates transcribed using T7 RNA polymerase. The YBL059W transcription used a DNA template prepared by annealing two synthetic DNA oligonucleotides (obtained from IDT; Coralville, IA).[1] The β-globin transcriptions used a DNA template prepared by PCR from synthetic DNA oligonucleotides. The ACT1 transcription used a DNA template prepared by PCR using the pT7ACT(–6) plasmid, which was a gift from J. Staley (Univ. of Chicago). Analogues of the ACT1 transcript with modifications at the branch-site nucleotide were prepared by transcription from PCR products that were obtained using pT7ACT(–6) as the template. DNA and RNA samples were purified by denaturing PAGE and quantified by UV absorbance (A_{260}) as described previously.[2]

The two yeast RNAs share the consensus sequences UACUAAC/GU (the branch-site adenosine is underlined), as shown in Figure 2. The β-globin RNA has the sequence CACUGAC/GU, which differs from the yeast consensus at the two italicized positions. The YBL059W sequence is 5'-**GU**AUGCAUAGGCAAUAACUUCGGCCUCAUACUCAAAGAACACGUU**UACUAAC**AUAACUUAUUUACAUAG-3' (the branch-site adenosine is underlined; the yeast consensus elements for the 5'-splice site and branch-site regions are in boldface). The β-globin[3]and ACT1[4] sequences were derived from the published sequences. The β-globin sequence is shown in full in Figure S3. For the YBL059W, β-globin, and ACT1 RNAs, the linear substrates are 69, 130, and 309 nt in length, respectively. The lariat loops are 51, 94, and 266 nt, respectively; the remaining 18, 36, and 43 nt nucleotides form the single-stranded 3'-tail. The loops of these lariats correspond to rings of 307, 565, and 1597 atoms, respectively (loop size is 6*n*+1 atoms, where *n* = number of nucleotides in the loop).

Deoxyribozyme activity assays

For the assays of Figure 2, 1 pmol of 5'-^{32}P-radiolabeled L substrate (plus 4 pmol of unradiolabeled L), 15 pmol of deoxyribozyme, and 30 pmol of R substrate (L:deoxyribozyme:R = 1:3:6) were incubated in 10 μL volume containing 50 mM HEPES, pH 7.5, 150 mM NaCl, 2 mM KCl, and 20 mM $MnCl_2$ at 37 °C. The Mn^{2+} was added from a 100 mM aqueous stock solution. The assays of Figure 3 used 5 pmol of linear substrate (which was quantified by UV absorbance) and 10 pmol of deoxyribozyme. See our previous report for details of the annealing and incubation procedures.[5] Aliquots from the reaction solutions were quenched onto stop solution (80% formamide, 1× TBE [89 mM each Tris and boric acid, pH 8.3], and 0.025% each xylene cyanol and bromophenol blue), electrophoresed on 20% PAGE, and imaged using a PhosphorImager. Values of k_{obs} and final yield were obtained by fitting the yield versus time data directly to first-order kinetics; i.e., yield = $Y \bullet (1 - e^{-kt})$, where $k = k_{obs}$ and Y = final yield. The $K_{d,app}$ for Mn^{2+} was ~30 mM (data not shown), and Mn^{2+} was typically used at 20 mM to suppress nonspecific RNA degradation that is observed at high Mn^{2+} concentrations.

Comprehensive demonstration of the generality of the 6BX22 deoxyribozyme

A summary of the sequence requirements of the 6BX22 deoxyribozyme is shown in Figure 2a. These requirements were determined by systematic experiments in which the RNA substrates were varied and deoxyribozyme activity was assayed. Figure 2b shows the experiment in which all nucleotides of the binding arms remote from the branch site were varied systematically by transversions, with maintenance of high ligation activity (see Figure 2b caption for details). In addition, Figure 2c documents that the branch-site nucleotide itself may be varied with good yield for A, U, and C and detectable ligation activity for branch-site G.

The RNA nucleotide of the L substrate immediately 5' of the branch site, UACUA (see Figure 2a), was not known to be base-paired with DNA at the outset of these experiments.[6] However, covariation experiments demonstrated that an RNA:DNA base pair indeed exists at this position (Figure S1a), thereby permitting any RNA nucleotide to be used at this site. In all cases, partial alkaline hydrolysis of the branched product verified that the site of branching remained unchanged (data not shown). For the unpaired 5'-GU position of the R substrate, testing of all four RNA nucleotides showed that any may be used, although G has lower rate and yield (Figure S1b). At the UACUAAC position of the L substrate, a U nucleotide led to <1% ligation activity (data not shown), suggesting that the C nucleotide is required. The sequence of the 3'-tail of the L substrate (after the UACUAAC sequence) was not varied systematically in these initial experiments. However, the synthesis of the three biologically related lariat RNAs shown in Figure 3 demonstrates that a wide variety of 3'-tail sequences and lengths are tolerated.

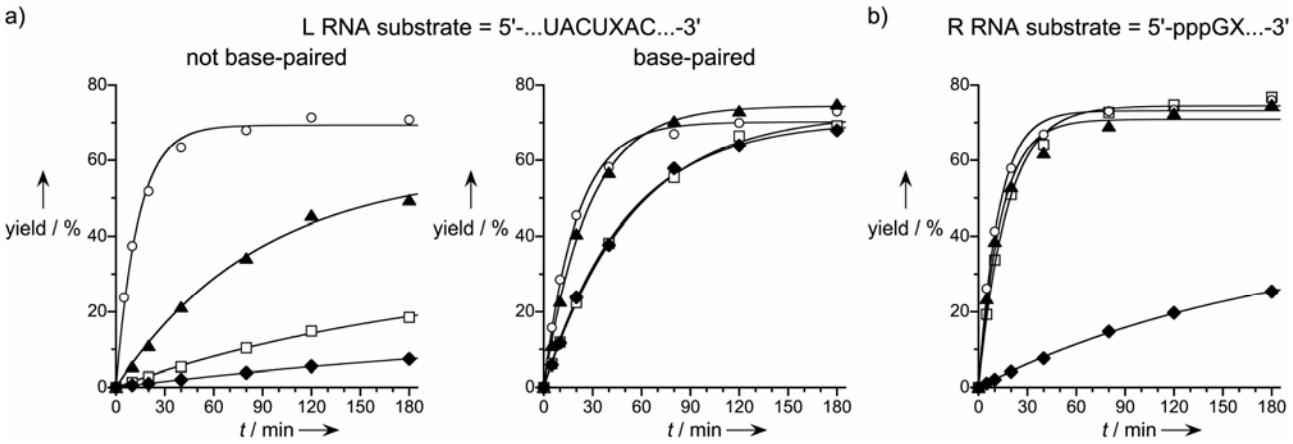

Figure S1. Additional experiments to establish the generality of the 6BX22 deoxyribozyme. a) Demonstrating an RNA:DNA base pair at the UACUA position of the L substrate. Because of this complementarity, any RNA nucleotide may be used at this position with proper choice of the corresponding DNA nucleotide. The L RNA sequence was either ...UACUA, UACUG, UACUC, or UACUU (circles, triangles, squares, and diamonds). The corresponding 6BX22 DNA nucleotide was T in all cases for the "not base-paired" experiment. The DNA nucleotide was the Watson-Crick match (i.e., T, C, G, or A), for the "base-paired" experiment, in which the k_{obs} values were 0.050 ± 0.003, 0.037 ± 0.001, 0.019 ± 0.001, and 0.020 ± 0.001 min^{-1} (<3-fold variation). b) Demonstrating generality at the GU position of the R substrate. The R RNA sequence was 5'-GU..., GA..., GC..., or GG... (circles, triangles, squares, and diamonds). The k_{obs} values were 0.081 ± 0.005, 0.072 ± 0.007, 0.057 ± 0.003, and 0.0055 ± 0.0003 min^{-1}. All errors are standard deviations from exponential curve fits.

The predicted secondary structure of 6BX22 when complexed with its RNA substrates, as generated by the mfold program,[7] is shown in Figure S2. The three-dimensional structure of 6BX22 is unknown, and the comprehensive structural biology experiments necessary to determine its three-dimensional structure have not yet been performed.

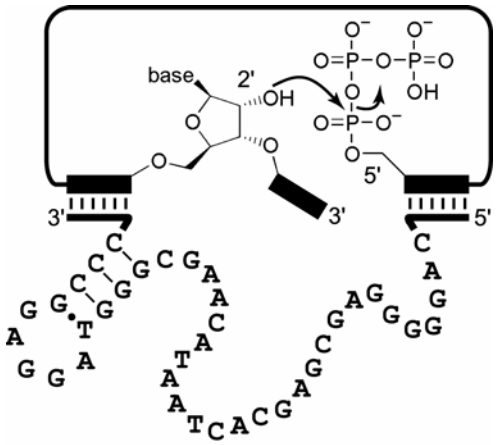

Figure S2. Mfold-predicted secondary structure of the 6BX22 deoxyribozyme.

RNA secondary structures and use of disruptor DNA oligonucleotides to enhance lariat synthesis

To enable deoxyribozyme-catalyzed synthesis of lariat RNA, the DNA binding arms of the deoxyribozyme must bind with the appropriate nucleotides of the linear RNA substrate (Figure 1). Intramolecular secondary structure within the linear RNA substrate can potentially compete with binding of the deoxyribozyme, which would inhibit lariat formation. The mfold-predicted secondary structures of the YBL059W and β-globin linear RNA substrates are shown in Figure S3. For YBL059W (69 nt), predicted secondary structures are shown in Figure S3a for both the linear substrate and the lariat product (the RNA nucleotides colored blue and red interact with the DNA binding arms of the deoxyribozyme). Of these two structures, the first is more relevant when considering the binding of the deoxyribozyme during the lariat-formation reaction; the lariat itself is shown here only for comparison. Because the deoxyribozyme-catalyzed yield of lariat was high, intramolecular secondary structure formation by the YBL059W linear RNA substrate was functionally inconsequential during deoxyribozyme-catalyzed lariat synthesis. Consistent with this, disruptor DNA oligonucleotides did not increase the rate or yield of lariat synthesis when tested (data not shown).

For β-globin (130 nt), the predicted secondary structure for the linear substrate is shown in Figure S3b. The RNA nucleotides colored blue and red interact with the DNA binding arms of the deoxyribozyme. (A number of variant secondary structures are of comparable energy, but in all cases the blue and red nucleotides and the branch-site adenosine are located within regions of considerable base pairing.) Because the blue and red RNA nucleotides are involved in extensive secondary structure interactions with other (black) nucleotides of the substrate, lariat formation using this RNA substrate proceeded well only when disruptor DNA oligonucleotides were present. The disruptors (orange and green) were designed to base-pair with the black RNA nucleotides that form secondary structure elements with blue and red RNA nucleotides, thereby freeing the latter residues to interact with the deoxyribozyme binding arms. Because of the nature of Watson-Crick complementarity, the disruptors are inherently complementary to the deoxyribozyme binding arms themselves; this is true for all RNA positions within the binding arms at which an intramolecular Watson-Crick interaction exists within the RNA substrate. To disfavor unproductive binding of the disruptors to the deoxyribozyme, key

mismatches were strategically made within the disruptor sequences, as denoted by "x" in Figure S3b; disruptor positions not mismatched in this fashion are marked with "*", and disruptor positions that cannot bind to the corresponding deoxyribozyme nucleotide are denoted by "o". For the β-globin substrate, only when both of the disruptors shown in Figure S3b were included was the yield of lariat RNA optimal. For example, the yield at the 6-h timepoint (see Figure 3) was 33% with both disruptors; 3% with just the orange or green disruptor; and 0.3% with neither disruptor.

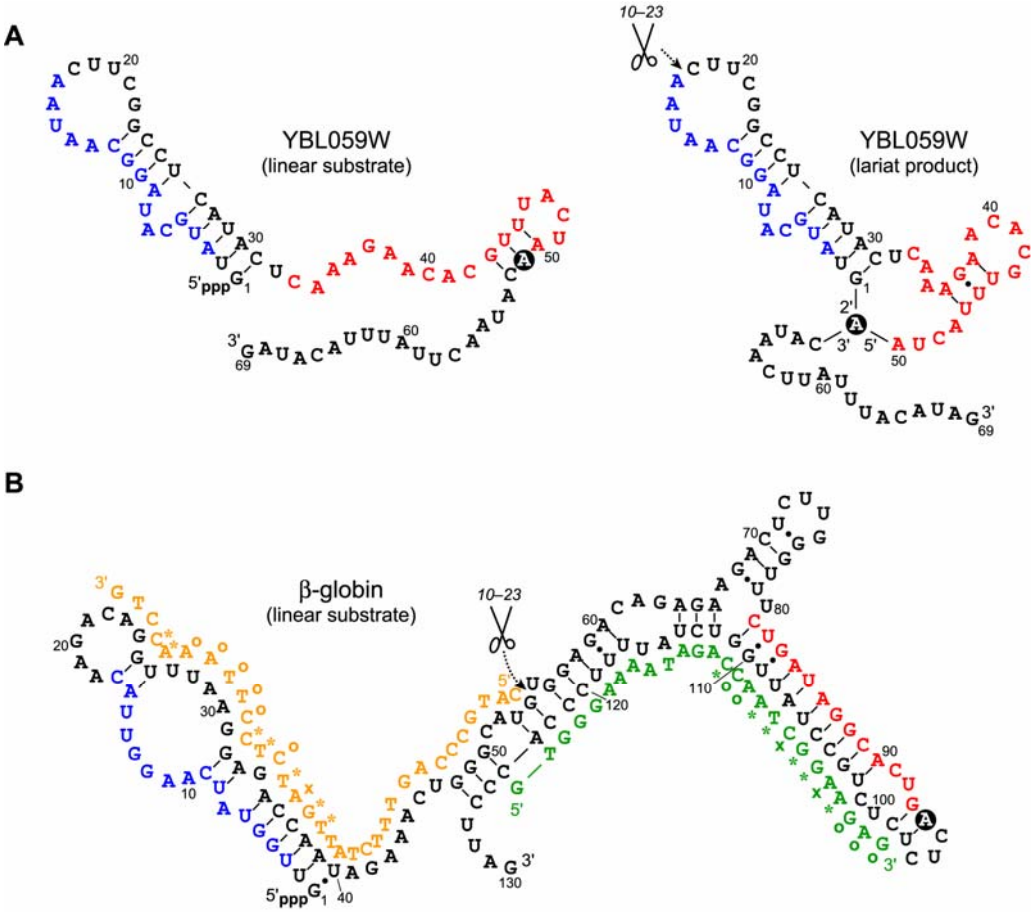

Figure S3. Mfold-predicted secondary structures of the YBL059W and β-globin linear RNA substrates. The branch-site adenosine is highlighted. The red RNA nucleotides comprise the L binding arm, and the blue nucleotide comprise the R binding arm (the 6BX22 sequence was synthesized with DNA binding arms that are the Watson-Crick complement of these RNA sequences, with the 39-nt enzyme region embedded within the binding arms). The lowest-energy mfold structure is shown, although other structures can be formed from the given sequence. a) YBL059W. Structures are shown for both the linear substrate and lariat product. In the latter instance, mfold was used to predict the structure from the 5'-pppG to the branch-site A, and the 3'-tail (which was not predicted to adopt any secondary structure on its own) was appended. Marked with a small arrow is the site of 10–23 cleavage for lariat characterization (see Figure S4 below). b) β-globin. The orange and green DNA sequences are disruptor oligonucleotides that together were required for the best lariat yields (see text for explanation, including the *, x, and o symbols).

Our guideline for designing a disruptor sequence is that mismatches ("x") are intentionally introduced when four or more contiguous Watson-Crick base pairs exist within the intramolecular RNA structure in a region where the deoxyribozyme needs to bind. Although these disruptor mismatches destabilize the desired disruptor-RNA interaction, this effect is presumably small because of the extensive disruptor-RNA complementarity along the remaining length. However, the disruptor-deoxyribozyme interactions are more severely destabilized by these mismatches, because the disruptor-deoxyribozyme interactions were not very strong to begin with, considering that the intramolecular RNA secondary structure is imperfect (i.e., there are G-U wobble pairs and unpaired nucleotides). The use of disruptor oligonucleotides has a side benefit in that single-stranded RNA regions are converted to RNA:DNA duplex structure, which helps to prevent nonspecific RNA degradation. This is useful regardless of whether or not sequestration of RNA secondary structure by disruptors is necessary to enable lariat synthesis. Overall, for any particular RNA substrate, we recommend testing of several disruptor oligonucleotides with varying design details, to optimize the yield of lariat synthesis.

For the ACT1 RNA, the disruptor 5'-CTAAACATATAATATAaCAACAAAAAGAATGAAGC-3' was used (the lowercase nucleotide denotes an intentional mismatch). As indicated in the text, a disruptor oligonucleotide was helpful but not absolutely required for ACT1 lariat synthesis, primarily by modestly enhancing the ligation rate (by less than two-fold) and by suppressing nonspecific RNA degradation. The protective effect of the disruptor was sufficiently large that it was used in the preparative experiment shown in Figure 4.

Experimental evidence for the YBL059W, β-globin, and ACT1 lariat RNA structures

As shown in Figure S4, evidence for the lariat RNA structures was obtained using assays similar to those described in detail in our previous report.[8] Lariats were synthesized using RNA substrates transcribed with α-^{32}P-CTP, such that the substrates were internally radiolabeled. In each panel of Figure S4, the lariat synthesized in the first of five sets of lanes was subjected either to cleavage by debranching enzyme (Dbr; second set of lanes)[9] or to cleavage by a 10–23 deoxyribozyme that targets a specific site within the lariat loop (third set of lanes; see Figure S3 above for site of 10–23 cleavage within the YBL059W lariat).[10] The PAGE-purified 10–23 cleavage product was also subjected to Dbr cleavage (fourth set of lanes), resulting in two fragments that match the 10–23 cleavage products obtained from the linear substrate standard RNA (fifth set of lanes). In all cases, the expected pattern of bands was observed, thereby confirming the lariat RNA structure.

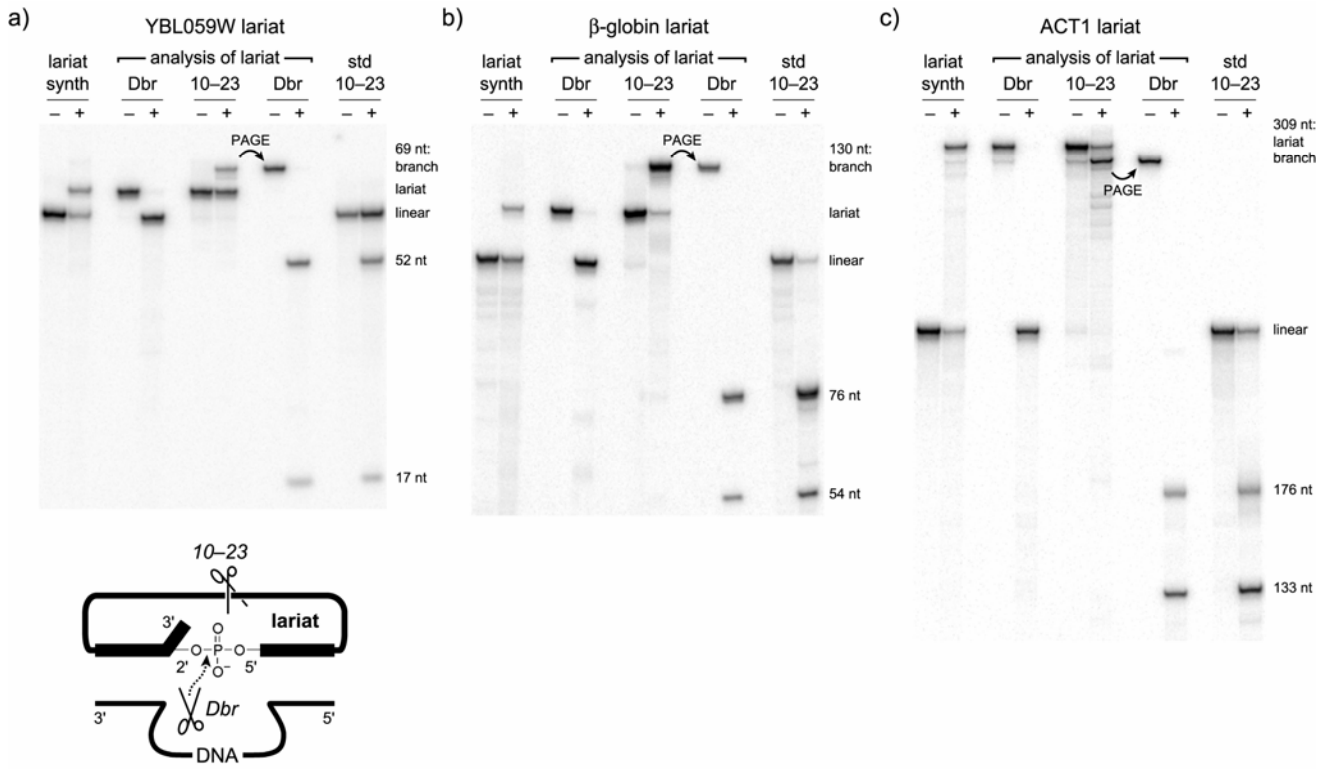

Figure S4. Demonstrating lariat RNA structures for the three lariat RNAs synthesized by the 6BX22 deoxyribozyme. a) YBL059W lariat (12% PAGE). b) β-globin lariat (8% PAGE). c) ACT1 lariat (8% PAGE). For lariat synthesis, + denotes 1.5 h incubation (YBL059W, ACT1) or 6 h incubation (β-globin) as described in the Experimental Section. For Dbr cleavage, + denotes 15 min incubation with ~75 ng yeast debranching enzyme Dbr (previously purified by glycerol gradient gel electrophoresis) in 20 mM HEPES, pH 7.5, 125 mM KCl, 0.5 mM $MgCl_2$, and 1 mM DTT at 30 °C. For 10–23 cleavage, + denotes 1.5 h incubation with at least five-fold excess 10–23 deoxyribozyme in 50 mM HEPES, pH 7.5, 150 mM NaCl, 2 mM KCl, and 5 mM $MnCl_2$ at 37 °C. The schematic diagram below panel a) schematically indicates the cleavage sites for Dbr and the 10–23 deoxyribozyme on the lariat RNA.

References for Supporting Information

[1] J. F. Milligan, D. R. Groebe, G. W. Witherell, O. C. Uhlenbeck, *Nucleic Acids Res.* **1987**, *15*, 8783; J. F. Milligan, O. C. Uhlenbeck, *Methods Enzymol.* **1989**, *180*, 51.
[2] Y. Wang, S. K. Silverman, *Biochemistry* **2003**, *42*, 15252.
[3] R. A. Spritz, P. Jagadeeswaran, P. V. Choudary, P. A. Biro, J. T. Elder, J. K. deRiel, J. L. Manley, M. L. Gefter, B. G. Forget, S. M. Weissman, *Proc. Natl. Acad. Sci. USA* **1981**, *78*, 2455; R. Reed, T. Maniatis, *Cell* **1985**, *41*, 95; K. B. Hall, M. R. Green, A. G. Redfield, *Proc. Natl. Acad. Sci. USA* **1988**, *85*, 704.
[4] R. Ng, J. Abelson, *Proc. Natl. Acad. Sci. USA* **1980**, *77*, 3912.
[5] A. Flynn-Charlebois, Y. Wang, T. K. Prior, I. Rashid, K. A. Hoadley, R. L. Coppins, A. C. Wolf, S. K. Silverman, *J. Am. Chem. Soc.* **2003**, *125*, 2444.
[6] Y. Wang, S. K. Silverman, *Biochemistry* **2005**, *44*, 3017.
[7] J. SantaLucia, Jr., *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 1460; M. Zuker, *Nucleic Acids Res.* **2003**, *31*, 3406.
[8] Y. Wang, S. K. Silverman, *J. Am. Chem. Soc.* **2003**, *125*, 6880.
[9] S. L. Ooi, C. Dann, 3rd, K. Nam, D. J. Leahy, M. J. Damha, J. D. Boeke, *Methods Enzymol.* **2001**, *342*, 233; K. Nam, R. H. Hudson, K. B. Chapman, K. Ganeshan, M. J. Damha, J. D. Boeke, *J. Biol. Chem.* **1994**, *269*, 20613.
[10] S. W. Santoro, G. F. Joyce, *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 4262.