# Supplementary Information

# Small, highly-active DNAs that hydrolyze DNA

Hongzhou Gu, [†,‡] Kazuhiro Furukawa, [†] Zasha Weinberg, [†,‡] Daniel F. Berenson, [§] and Ronald R. Breaker[*,†,‡,§]

[†]Department of Molecular, Cellular and Developmental Biology, Yale University, New Haven, Connecticut, 06520 United States

[‡]Howard Hughes Medical Institute, New Haven, Connecticut, 06520 United States

[§]Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, Connecticut, 06520 United States

[*]To whom correspondence should be addressed. E-mail: ronald.breaker@yale.edu

## Supplementary Methods

**Oligonucleotides.** The initial double-stranded 145mer DNA pool was generated by extension of the 80mer DNA library 1 (5′-pGGTGCTACAGCCATA$(N)_{50}$GACTGCATCACGAAG) and the 80mer template DNA library 2 (5′-GCGTGTACCAGTATG$(N)_{50}$CTTCGTGATGCAGTrC). The underlined nucleotides are complementary. DNA extension was carried out using SuperScript II reverse transcriptase (Invitrogen) according to the manufacturer's directions. A ribonucleotide (rC) was included at the 3′ terminus of library 2 to permit degradation of the template strand by treatment with 0.25 M NaOH at 90 °C for 5 min. 145-nucleotide DNA products were purified by denaturing 8% PAGE.

Single-stranded 145mer DNAs in subsequent rounds of selection, PCR amplification was performed with primer 1 (5′-pGGTGCTACAGCCATA) and primer 2 (5′-GCGTGTACCAGTATrG). Again, the rG was included at the 3′ terminus of primer 2 to permit degradation of the template strand by treatment with NaOH, and 145-nucleotide DNA products were purified by denaturing 8% PAGE. Full-length (~145mer) precursors of each deoxyribozyme chosen for analysis were purchased from the W. M. Keck Biotechnology Resource Laboratory at Yale University. Truncated deoxyribozyme representatives and synthetic DNA markers for mapping cleavage site were purchased from Sigma-Aldrich.

**Ligation of linear single-stranded DNAs with CircLigase to form circles.** CircLigase is a thermostable ATP-dependent single-stranded DNA ligase that catalyzes intramolecular ligation of single-stranded DNA templates. The ligation substrates must carry 5′-phosphate a 3′-hydroxyl groups on the donor and acceptor DNAs, respectively. CircLigase joins ends of single-stranded DNA in the absence of a complementary sequence. With the reaction buffer provided by the vendor, a final concentration of 5 mM $MgCl_2$ is present in the ligation reaction. Since we sought DNAs that might undergo cleavage in $Mg^{2+}$, we prepared a new ligation buffer wherein $Mg^{2+}$ is replaced with $Mn^{2+}$. This buffer contains 50 mM MOPS (pH 7.5 at 23°C), 100 mM KCl, 2.5 mM $MnCl_2$, 0.05 mM ATP, and 1 mM DTT. We observed only minor differences in the yield of ligation products between the vendor-supplied buffer and our ligation buffer.

**Sample preparation for exact mass spectroscopy.** A 100 μL sample containing 150 pmol of substrate and 200 pmol of enzyme DNAs (unlabeled) were incubated in 50 mM HEPES (pH 7.05 at 23°C), 100 mM NaCl, 20 mM $MgCl_2$, and 2 mM $ZnCl_2$ at 37°C for 40 min. The products were precipitated by centrifuging with the addition of 400 μL ethanol, dried at 23°C for 20 min, and the DNAs were resuspended in 10 μL $dH_2O$. The sample was analyzed (Novatia) to establish the monoisotopic mass of the cleaved products.

**Bioinformatic searches for natural sequences that conform to engineered deoxyribozymes.** We searched the bacterial and archaeal sequences in RefSeq[1] version 44, eukaryotic and viral sequences in RefSeq version 41. We also searched a set of metagenome sequences used in a previous study,[2] which were extracted from IMG/M,[3] assembled Human Microbiome Project sequences,[4] CAMERA,[5] MG-RAST,[6] and GenBank.[7]

Searches were performed by using our alignments of the deoxyribozyme classes as queries for covariance model[8] searches as implemented by the Infernal software package, version 1.0.[9] Predicted matches were evaluated manually based on their reported E-values and based on their consistency with biochemical data arising from our study. We also conducted searches using RNAmotif,[10] defining patterns based on the secondary structure reported in this work and the results of our mutational studies. Nucleotides that could not be mutated without significantly affecting deoxyribozyme speed were treated as fixed in RNAmotif patterns. The natural DNA sequences reported in Table S1 are a combination of sequences found using one or both of these search methods.

**Table S1.** Natural DNA sequences with similarities to class I deoxyribozymes. The locations of potential deoxyribozyme sequences are expressed in the form ACCESSION/START-END, where ACCESSION identifies the sequence containing the potential deoxyribozyme (e.g. a chromosome), START is the nucleotide number of the 5′-most nucleotide of the potential deoxyribozyme and END is the 3′-most nucleotide. Thus, if START>END, the potential deoxyribozyme is located on the reverse-complement strand of the containing sequence. The source of the sequence is indicated

as follows: "RefSeq" = RefSeq database,[1] "GOS" = Global Ocean Survey,[11] "PPool" = water treatment plant plasmid pool,[12] "HMP" = Human Microbiome Project assembled contig,[13] "BBAY" = Botany Bay,[14] "Strom" = Stromalite microbial fraction,[5] "MetaHIT" = MetaHIT Consortium sequencing of the human gut.[15] Some of these sequences were derived from larger databases cited in the Methods section. Red nucleotides refer to the highly conserved nucleotides in the consensus sequence of the class I deoxyribozymes.

Original Natural DNA Sequences

NS1 (NW_002763791.1/276-326)
5′-TGCAGCTGTTGAAGAGGGATAGATACAGTATCCGTAGTTGAGCTAGCTGTA
RefSeq

NS2 (NW_002032911.1/557153-557188)
TTTTGTAGTTGAGCTGGTCAAACAGTTGAAGCAAAG
RefSeq

NS3 (NW_001811629.1/1391829-1391858)
AATATTAGTTGAGGAATTGTTGAAGATATT
RefSeq

NS4 (JCVI_READ_1101733062292/441-411)
TCCTGGAGTTGAGCCGCAAGTTGAAGCAGGA
GOS

NS5 (NW_001794727.1/1449431-1449522(approx))
AGGCTCTGGTTGAGAGAGAAAGTTGGAGGAGCCT
RefSeq

NS6 (NT_167259.1/542797-542886(approx))
TAGATGTGGTTGAGGCATTGTTGAAGCATCTG
RefSeq

NS7 (NZ_ABUK01000001.1/4928840-4928804)
GATAGTCCCAGTTGAGAATATCGTTGAAGGGGCCATC
RefSeq

NS8 (NC_010336.1/1022349-1022388)
GATGTCTATAGTTGAGAAATTTGTTGAAGCATTAGATATC

RefSeq

NS9 (85012654/292-257)
GACTCTAGTAGTTGAGGAATTAGTTGAAGAAGAGTC
GOS

NS10 (2036027165/340-386)
TCGACCGTCGTTGAGTTCGGCCACGAAGTCCGGGTAGAAGCGGTCGG
PPool

NS11 (NW_002198982.1/878019-878049)
ATTCAGTGTTGAGGCCATAGTTGAAGTGAAT
RefSeq

NS12 (NC_008469.1/1178261-1178236)
CTTCCTTGAGTTTCATAGTGTTGAAG
RefSeq

NS13 (NW_001888394.1/48725-48754)
TTATTTAGTTGAGGGGGCGTTGGAGAATAA
RefSeq

NS14 (NC_009601.1/80764-80734)
GAATTTTGTTGAGAGCTCCGTTGCAGAATTC
RefSeq

NS14 (SRS050925_LANL_scaffold_3227/9387-9357)
GGCCCGTGGTTGAGGTGCTCGTTGCAGCGGGCC
HMP

NS15 (JCVI_SCAF_1101668626060/709-738)
CATGCAAGTTGAGCAACCGTTGAAGGCATG
GOS

NS16 (NCBI_BBAY_READ_1106073142478/401-366)
CGAAGCTGGTTGAGAGCCTCTGTTGGAGGCTTCG
BBAY

NS17 (SRR001063.296208/35-64)
GTACCCCGTTGAGAGGACGTTGAAGGGTAC
Strom

NS18 (NW_001283854.1/277967-278056(approx))
AACTACTTTGTTGAGGAATCGTTGTAGAGTAGTT
RefSeq

NS19 (NW_001867407.1/3494541-3494630(approx))
GGGAGATGTTGAGAATCAGTTGCAGCTCCC
RefSeq

NS20 (NW_001493109.1/512952-512861(approx))
GCCCAGGGGTTGAGACCAGACGTTGGAGCTGGGC
RefSeq

NS21 (NW_001030737.1/3464208-3464298(approx))
TCAGCTACTGTTGAGTTCATTGTTGGAGTAGCTGA
RefSeq

NS22 (NT_039353.7/87857789-87857699(approx))
GTTTTATACCGTTGAGTGACAAGTTGCAGTATAAAAC
RefSeq

NS24 (NT_039578.7/27203945-27203854(approx))
GACCCACAGGTTGAGAACTAGTGTTGTAGGTGGGTC
RefSeq

NS25 (NW_002198506.1/13163585-13163676(approx))
TTCAGCCTGTTGAGCTAGAAGGTTGAAGGCTGAA
RefSeq

NS26 (NC_012070.1/871748-871837(approx))
CGTTGTTGTTGAGCATGTGTTGAAGCAACG
RefSeq

NS27 (NW_001092856.1/891102-891012(approx))
GTGCCGCCGTTGAGGCAGGGGTTGGAGCGGCAC
RefSeq

NS28 (scaffold1186_1_MH0002/265-355(approx))
TCGACGCTGATGGGCCCGTGGTTGAGGTGCTCGTTGCAGCGGGCCATCGCTCGG
MetaHIT

Sequences of the NS4 variants

NS4-gtoT
TCCTGTAGTTGAGCCGCAAGTTGAAGCAGGA

NS4-Ext.
GGCAAGAAGGCCAGCAATTCGATCCTCAGCTTCCTGGAGTTGAGCCGCAAGTTGA
AGCAGGACGGCCCCGATGCTGATCTCGTGGAGGTGG

NS4-SH
GCATGTCCTGGAGTTGAGCCGCAAGTTGAAGCAGGACATGC

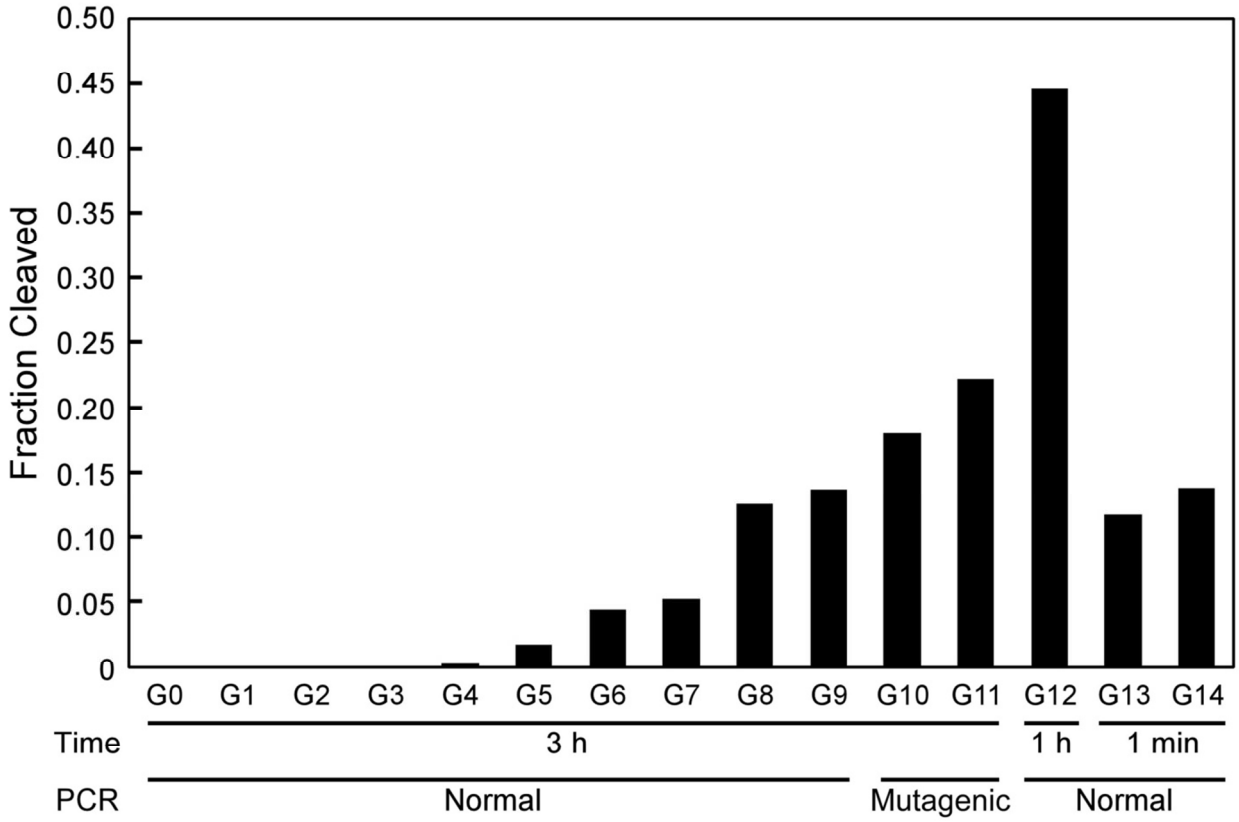
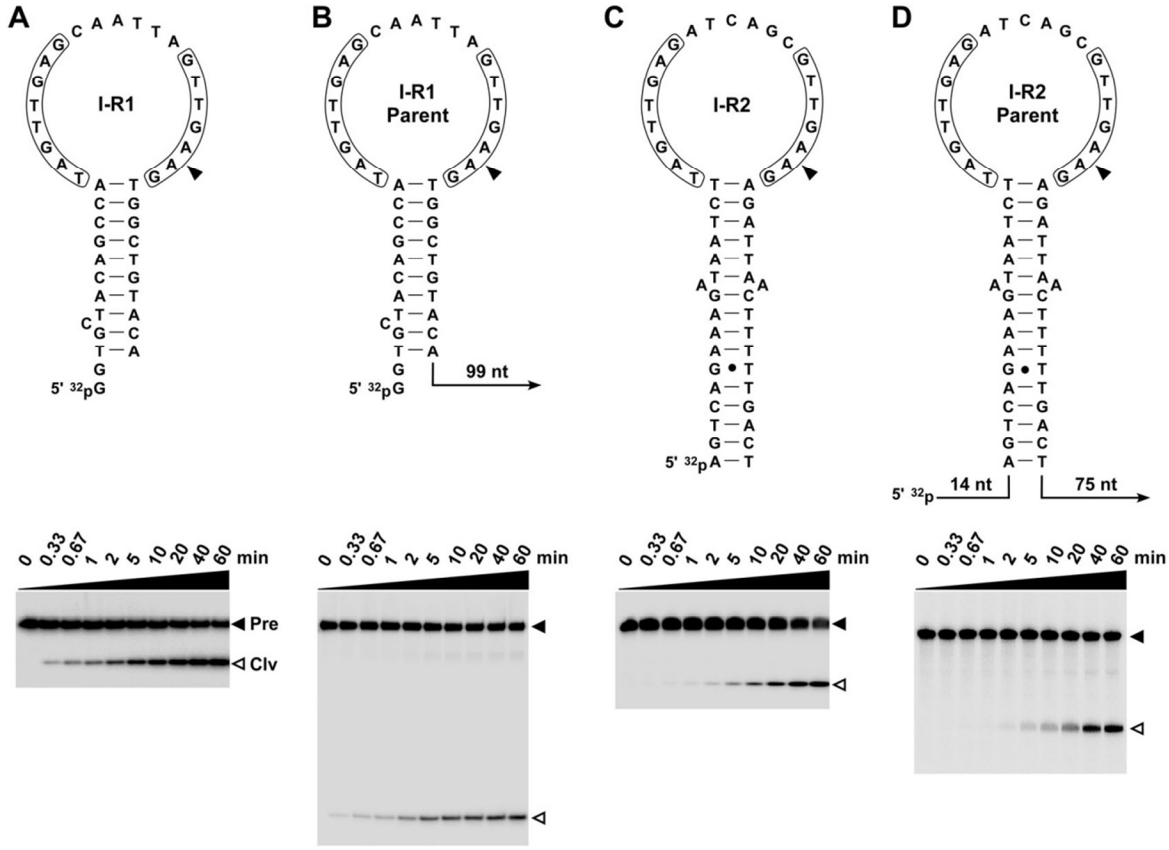
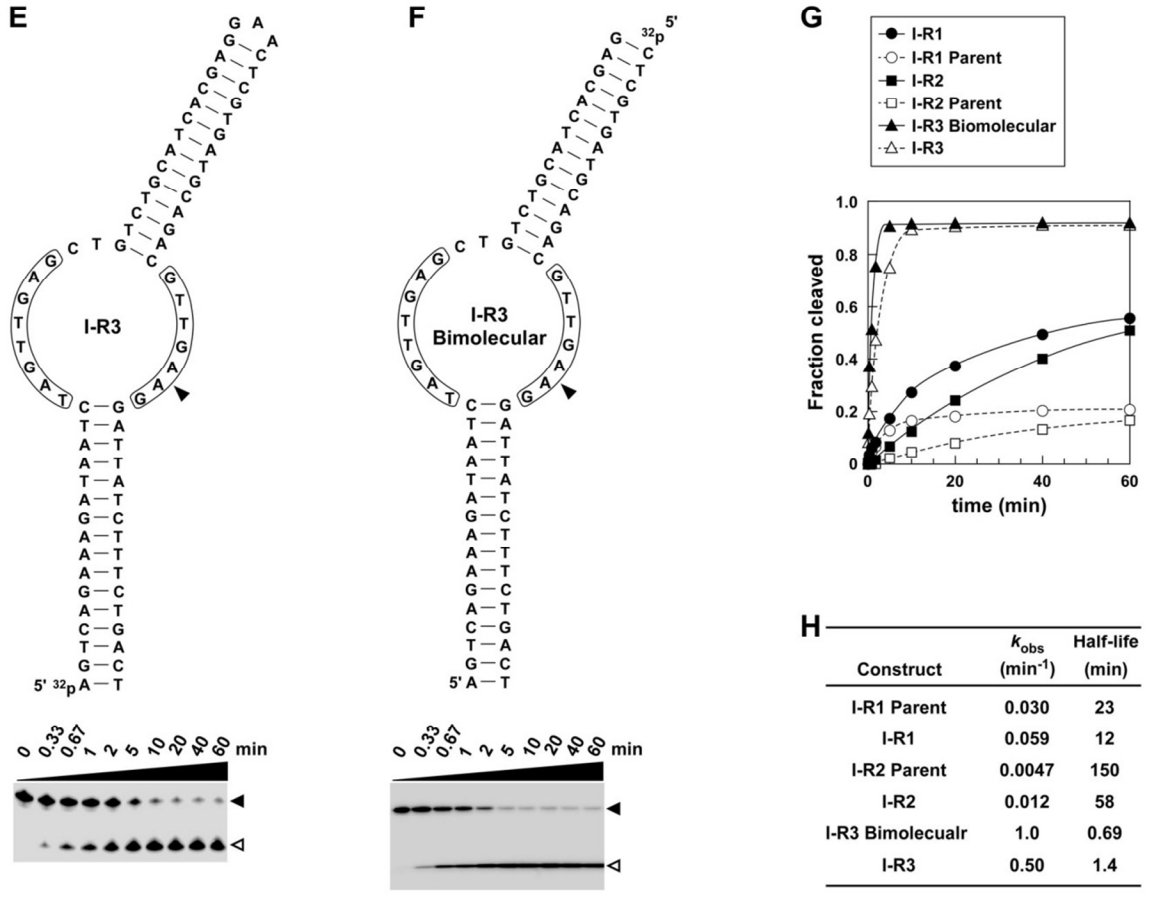



**Figure S1.** *In vitro* selection of self-cleaving deoxyribozymes. Plot of the fraction of DNA cleaved versus the population (G0 through G14), where G0 designates the original DNA pool before selection and G14 designates the DNA pool resulting from 14 rounds of selective amplification.







H

| Construct | $k_{obs}$ ($min^{-1}$) | Half-life (min) |
|---|---|---|
| I-R1 Parent | 0.030 | 23 |
| I-R1 | 0.059 | 12 |
| I-R2 Parent | 0.0047 | 150 |
| I-R2 | 0.012 | 58 |
| I-R3 Bimolecualr | 1.0 | 0.69 |
| I-R3 | 0.50 | 1.4 |

**Figure S2.** Sequences, secondary structure models, and kinetic parameters for representative class I deoxyribozymes. (A-F) Trace amounts of each representative deoxyribozyme (5′ $^{32}$P-labeled) were incubated at 37°C in reaction buffer. Only the substrate DNA was radiolabeled for the bimolecular I-R3 construct. Samples exposed for different incubation times were subjected to denaturing 10% PAGE to separate deoxyribozyme 5′ cleavage products from uncleaved DNAs. Filled and open arrowheads designate uncleaved and 5′ cleavage products, respectively. Encircled nucleotides are highly conserved among class I deoxyribozymes. (G) Plot of the fraction cleaved versus the incubation time for certain class I representatives and their precursors. Plots were derived from the gels in a-f. (H) Rate constant and half-life values for class I representatives and their precursors (parent). Values were calculated based on the curves depicted in G. Note that I-R1 and I-R2 cleaved faster compared to their parent constructs, which might be caused by misfolding of the longer DNAs.

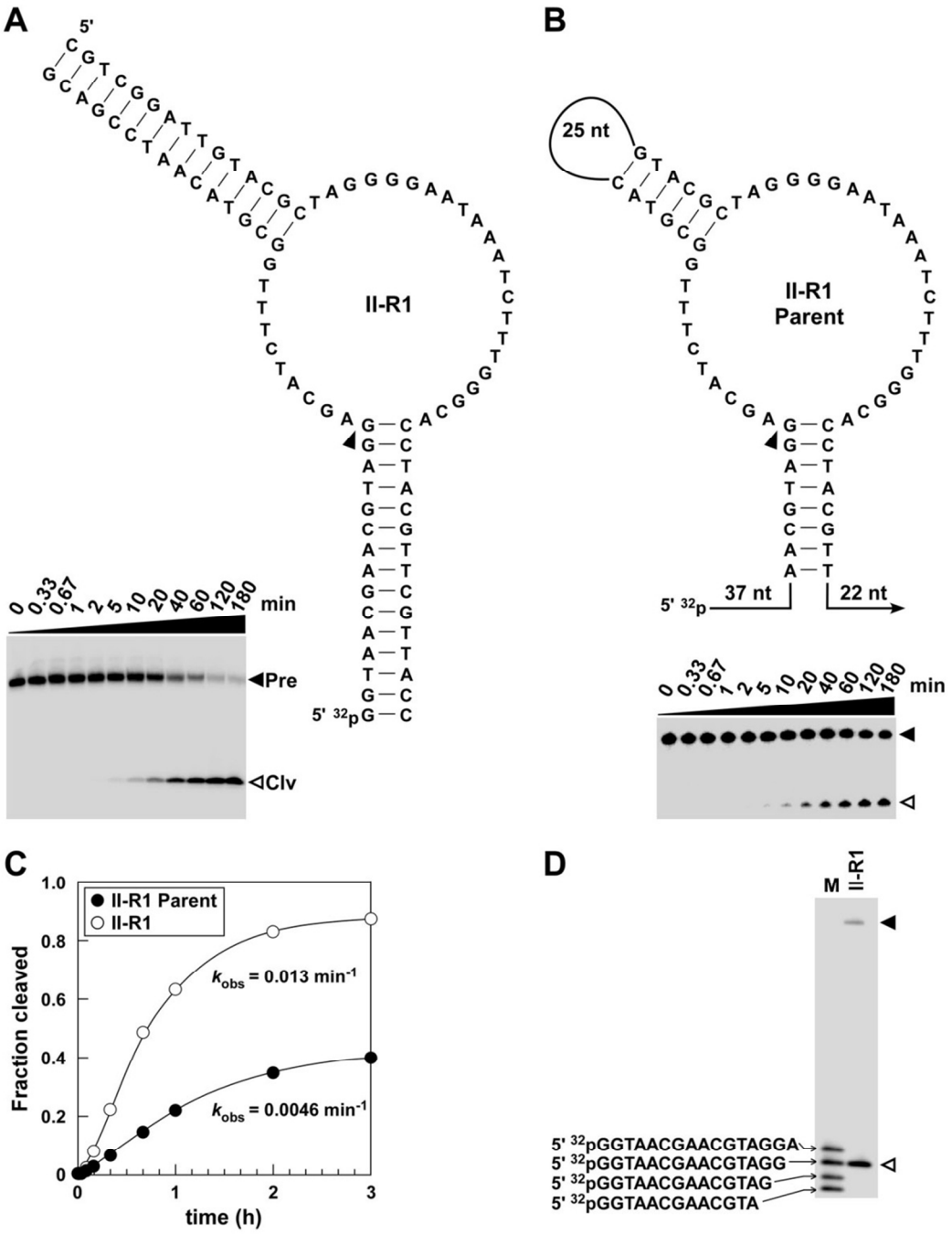



**Figure S3.** Sequences, secondary structure models, and kinetic parameters for representative class II deoxyribozymes. (A) and (B) Trace amounts of 5′ $^{32}$P-labeled substrate or parent DNAs as depicted were incubated at 37°C in reaction buffer. Other annotations are as described in the legend to Figure S2 A-F. (C) Plot of the fraction cleaved versus the incubation time for II-R1 and its parent. See the legend to Figure S2G for additional details. (D) Mapping the cleavage site (arrowhead) of II-R1 by

comparison of the 5′ cleavage product with 5′ $^{32}$P-labeled synthetic DNA markers (M). A trace amount of 5′ $^{32}$P-labeled II-R1 was incubated at 37°C in reaction buffer for 1 h, and then separated by using denaturing 10% PAGE.

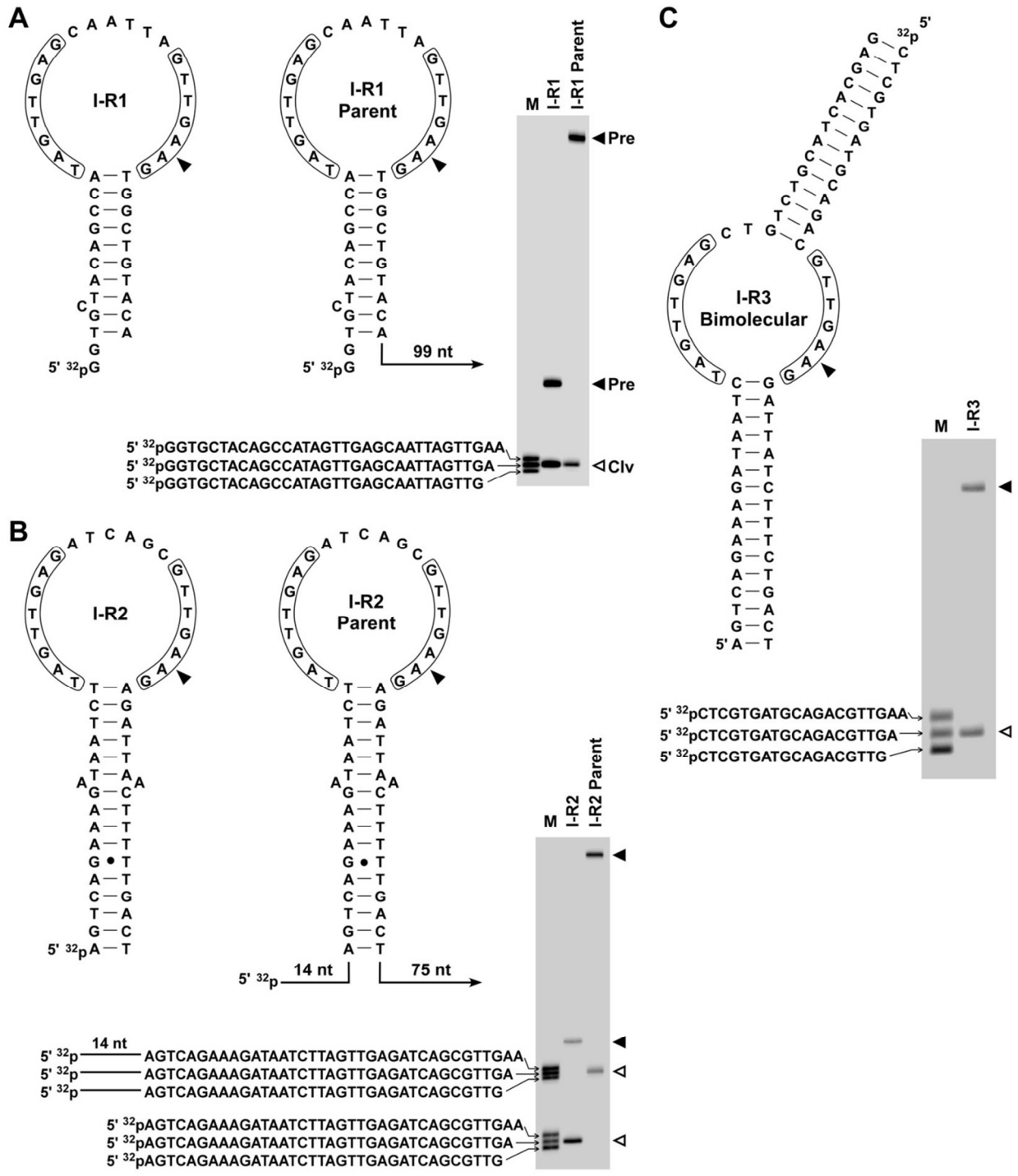



**Figure S4.** Mapping class I deoxyribozyme cleavage sites by comparison with synthetic DNA markers. (A-C) Mapping the cleavage sites of I-R1, I-R2, I-R3 and their derivatives. See the legend of Figure S3D for details.

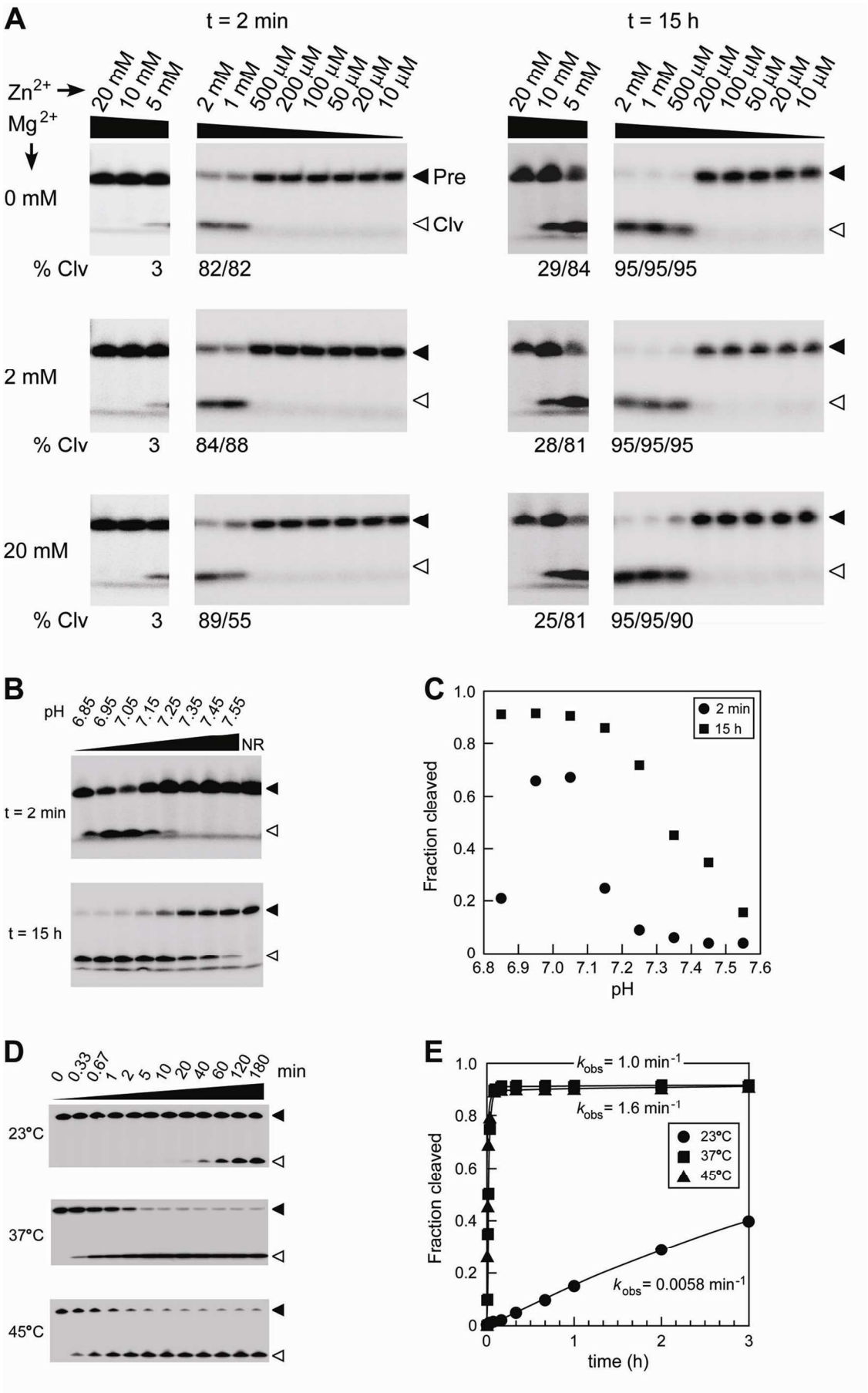

A
t = 2 min
t = 15 h
Zn2+
Mg2+
20 mM
10 mM
5 mM
2 mM
1 mM
500 µM
200 µM
100 µM
50 µM
20 µM
10 µM
Pre
Clv
0 mM
% Clv
3
82/82
29/84
95/95/95
2 mM
84/88
28/81
20 mM
89/55
25/81
95/95/90
B
pH
6.85
6.95
7.05
7.15
7.25
7.35
7.45
7.55
NR
t = 2 min
t = 15 h
C
2 min
15 h
Fraction cleaved
pH
D
0
0.33
0.67
1
2
5
10
20
40
60
120
180
min
23°C
37°C
45°C
E
kobs = 1.0 min-1
kobs = 1.6 min-1
kobs = 0.0058 min-1
23°C
37°C
45°C
Fraction cleaved
time (h)

**Figure S5.** Activity of I-R3 with various concentrations of $Mg^{2+}$/$Zn^{2+}$ and the pH and temperature dependence of I-R3. (A) Depicted are denaturing 10% PAGE separations of 5′ $^{32}$P-labeled I-R3 reaction products generated upon incubation in the presence of various divalent metal ion concentrations. In addition to the concentrations of $Mg^{2+}$ and $Zn^{2+}$ noted, the reaction buffer contains 50 mM HEPES (pH 7.05 at 23°C), and 100 mM NaCl. Samples were incubated for 2 min (or 15 h) and were subjected to denaturing 10% PAGE. The optimal $Zn^{2+}$ concentration is between 1 and 2 mM. Below 200 μM or above 20 mM $Zn^{2+}$, I-R3 did not exhibit any detectable activity over a 15 h incubation. (B) Denaturing 10% PAGE separations of 5′ $^{32}$P-labeled I-R3 reaction products generated upon incubation at various pH values. NR designates no reaction. (C) Plot of the fraction of DNA cleaved versus the pH for the reactions depicted in B. The optimal pH range for I-R3 is 6.95 to 7.05. (D) Denaturing 10% PAGE separations of 5′ $^{32}$P-labeled I-R3 reaction products generated upon incubation at various temperatures under otherwise standard reaction conditions. (E) Plot of the fraction of DNA cleaved versus the incubation time for the reactions depicted in D. The $k_{obs}$ values 0.0058 $min^{-1}$, 1.0 $min^{-1}$, and 1.6 $min^{-1}$, respectively, were obtained at 23°C, 37°C, and 45°C.

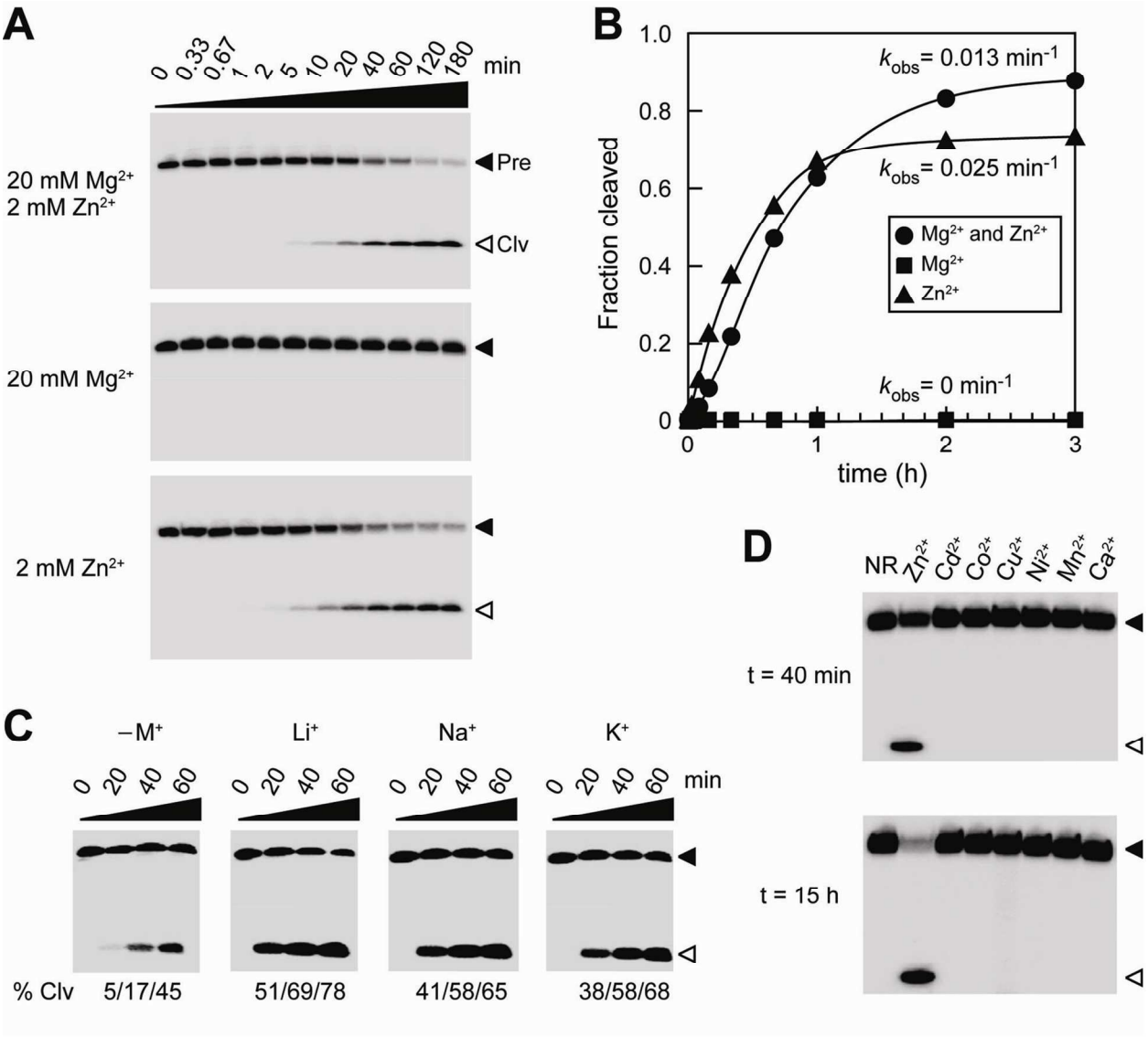



**Figure S6.** Analysis of deoxyribozyme II-R1 with various mono- and divalent cations. (A) Denaturing 10% PAGE separations of 5′ $^{32}$P-labeled II-R1 reaction products generated upon incubation in the presence of $Zn^{2+}$ and/or $Mg^{2+}$. Other annotations are as described in the legend to Figure S5. (B) Plot of the fraction of DNA cleaved versus time for the gels depicted in A. (C) Denaturing 10% PAGE separations of 5′ $^{32}$P-labeled II-R1 reaction products generated upon incubation in the absense (-$M^{2+}$) or the presence of various monovalent cations at 100 mM as indicated. Although the type of monovalent cation is not important, the absense of a monovalent cation causes a reduction in deoxyribozyme activity. (D) Denaturing 10% PAGE separations of 5′ $^{32}$P-labeled II-R1 reaction products generated upon incubation in the presence of various divalent metal

ions. Mixtures were incubated with 2 mM of the divalent cations as indicated for 40 min and 15 h. NR designates no reaction.

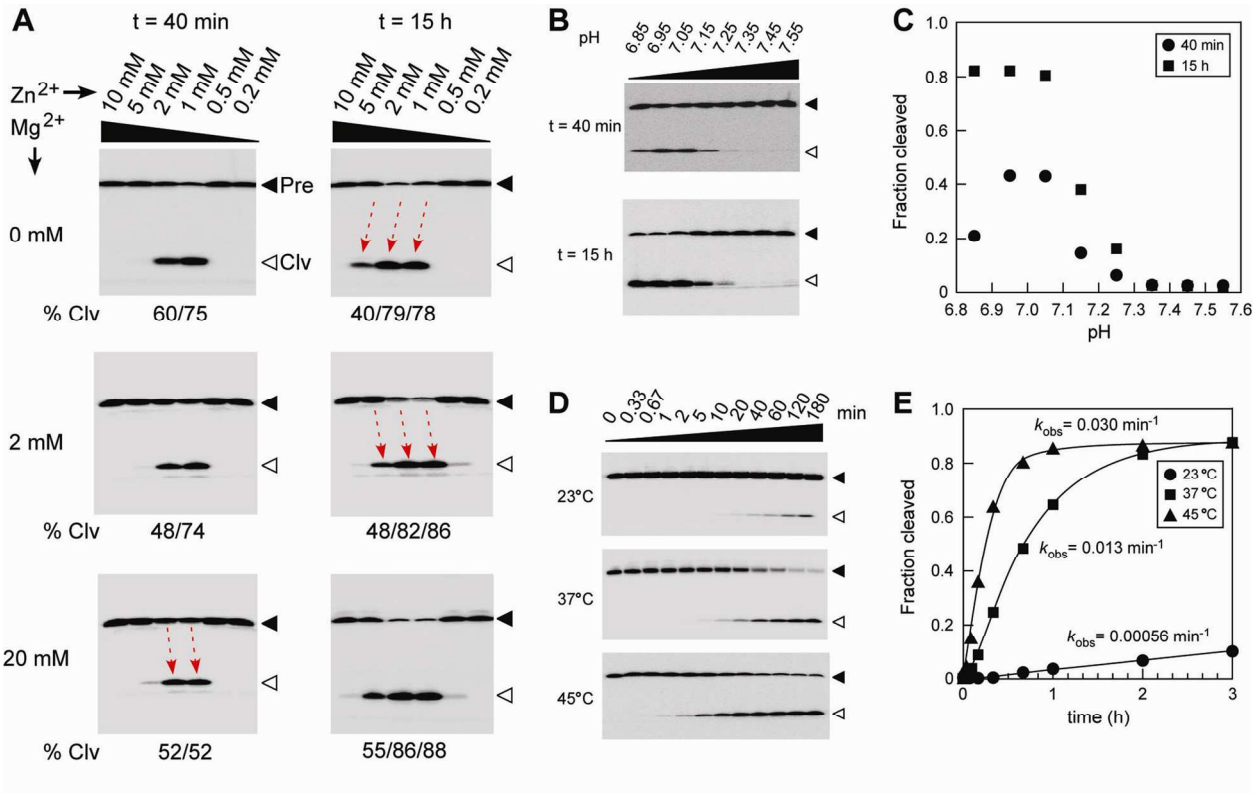



**Figure S7.** Activity of II-R1 with various concentrations of $Mg^{2+}$/$Zn^{2+}$ and the pH and temperature dependence of II-R1. (A) Depicted are denaturing 10% PAGE separations of 5′ $^{32}$P-labeled II-R1 reaction products generated upon incubation in the presence of various divalent metal ion concentrations. In addition to the concentrations of $Mg^{2+}$ and $Zn^{2+}$ noted, the reaction buffer contains 50 mM HEPES (pH 7.05 at 23°C), and 100 mM NaCl. Samples were incubated for 40 min or 15 h and were subjected to denaturing 10% PAGE. Higher $Mg^{2+}$ concentration causes inhibition of II-R1 (see the 40 min timepoint). However higher $Mg^{2+}$ concentration increases the overall fraction of II-R1 that cleaves after 15 h. The optimal $Zn^{2+}$ concentration is between 1 mM and 2 mM. Below 0.5 mM or above 10 mM $Zn^{2+}$, II-R1 exhibited almost no activity in the 15 h incubation. The gel lanes slightly shifted upon work-up, and arrows designate bands that are in the same lanes. (B) Denaturing 10% PAGE separations of 5′ $^{32}$P-labeled II-R1 reaction products generated upon incubation at various pH values. (C) Plot of the fraction of DNA cleaved versus the pH for the reactions depicted in *B*. The optimal pH range for II-R1 is 6.95 to 7.05. (D) Denaturing 10% PAGE separations of 5′ $^{32}$P-labeled II-R1 reaction products generated upon incubation at various temperatures under otherwise standard reaction conditions. (E) Plot of the fraction of DNA cleaved versus the incubation for the

reactions depicted in D. The $k_{obs}$ values 0.00056 $min^{-1}$, 0.013 $min^{-1}$, and 0.030 $min^{-1}$, respectively, were obtained at 23°C, 37°C, and 45°C.

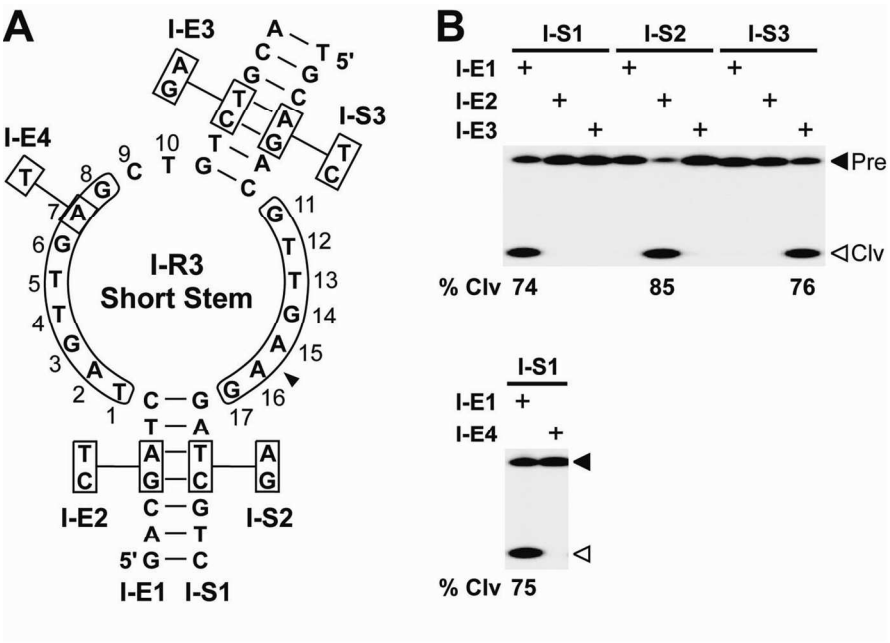



**Figure S8.** Analysis of I-R3 for the base-pairing in the stems and a conserved nucleotide in the catalytic core. (A) A truncated version of I-R3 with shortened stems and its variant sequences examined for DNA hydrolysis activity. E and S designate enzyme and substrate strands, respectively. Boxed nucleotides identify mutations and their locations. (B) Denaturing 10% PAGE separations of the products of various deoxrybozyme reactions. Trace amounts of 5′ $^{32}$P-labeled substrate strands were incubated with an excess of enzyme strands as indicated at 37°C under standard reaction conditions for 20 min.

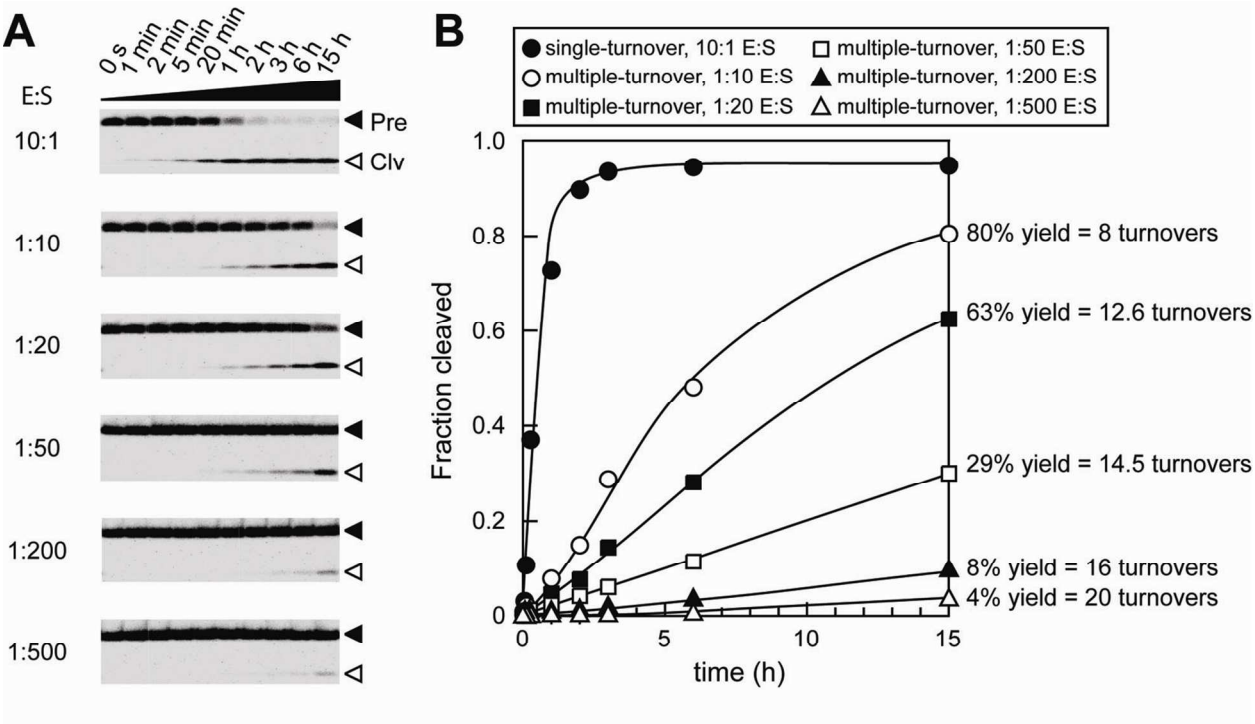



**Figure S9.** Analysis of I-R3 for multiple turnover. (A) Denaturing 10% PAGE separation of products of single and multiple turnover assays with 5′ $^{32}$P-labeled I-R3 substrate (S) DNAs and various concentrations of enzyme (E) DNAs based on the short-stem construct depicted in Figure S8. (B) Plot of the fraction cleaved versus the incubation time for the gels depicted in A. Turnover values were calculated from the fraction cleaved at the final (15 h) timepoint.

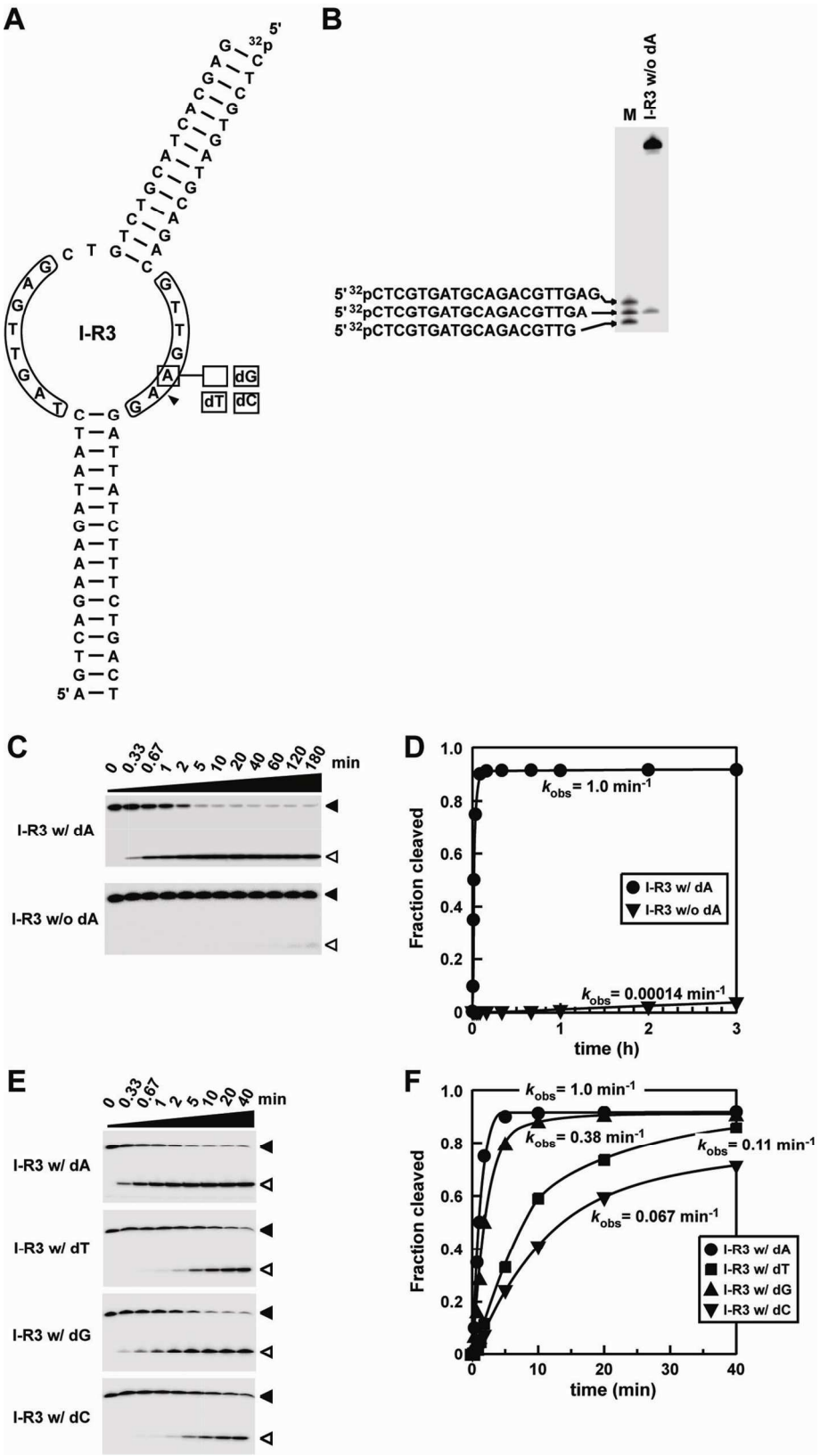

A
B
M
I-R3 w/o dA
5' 32pCTCGTGATGCAGACGTTGAG
5' 32pCTCGTGATGCAGACGTTGA
5' 32pCTCGTGATGCAGACGTTG
I-R3
dG
dT
dC
C
min
0 0.33 0.67 1 2 5 10 20 40 60 120 180
I-R3 w/ dA
I-R3 w/o dA
D
Fraction cleaved
time (h)
kobs = 1.0 min-1
kobs = 0.00014 min-1
I-R3 w/ dA
I-R3 w/o dA
E
0 0.33 0.67 1 2 5 10 20 40 min
I-R3 w/ dA
I-R3 w/ dT
I-R3 w/ dG
I-R3 w/ dC
F
Fraction cleaved
time (min)
kobs = 1.0 min-1
kobs = 0.38 min-1
kobs = 0.11 min-1
kobs = 0.067 min-1
I-R3 w/ dA
I-R3 w/ dT
I-R3 w/ dG
I-R3 w/ dC

**Figure S10.** Analysis of cleavage site sequence requirements. (A) Sequence, secondary structure, and nucleotide variations made to a bimolecular I-R3 construct. Boxed nucleotides indicate the nucleotide changes tested. Empty box designates a deletion. (B) Denaturing 10% PAGE separation of products of the I-R3 constructs using 5′ $^{32}$P-labeled substrate lacking one of the two deoxyadenosine moieties flanking the cleavage site. M designates a marker lane containing the synthetic DNAs as annotated. (C) Denaturing 10% PAGE of time-course reactions for I-R3 constructs using 5′ $^{32}$P-labeled substrates containing the typical dA residues flanking the cleavage site, or substrates that lack one of the two dA residues. (D) Plot of the fraction cleaved versus the incubation time for I-R3 with normal and modified substrates. Data is derived from C. The $k_{obs}$ for the normal substrate is approximately 10,000-fold greater than that observed for the substrate carrying the deletion. (E) Denaturing 10% PAGE of time-course reactions for I-R3 constructs using 5′ $^{32}$P-labeled substrates containing the typical dA residues flanking the cleavage site, or substrates wherein the dA 5′ of the cleavage site is changed to another type. The $k_{obs}$ for the normal substrate is approximately 9-, 3- and 15-fold greater than that observed for the substrates carrying dT, dG and dC residues, respectively.

## References


1. Pruitt, K. D.; Tatusova, T.; Maglott, D. R. *Nucleic Acids Res.* **2007**, *35*, D61-65.
2. Baker, J. L. et al. *Science* **2012**, *335*, 233-235.
3. Markowitz, V. M. et al. *Nucleic Acids Res.* **2008**, *36*, D534-538.
4. Gevers, D. et al. *PLoS Biol.* **2012**, *10*, e1001377.
5. Sun, S. et al. *Nucleic Acids Res.* **2011**, *39*, D546-551.
6. Meyer, F. et al. *BMC Bioinformatics* **2008**, *9*, 386.
7. Benson, D. A.; Karsch-Mizrachi, I.; Lipman, D. J.; Ostell, J.; Wheeler, D. L. *Nucleic Acids Res.* **2008**, *36*, D25-30.
8. Eddy, S. R.; Durbin, R. *Nucleic Acids Res.* **1994**, *22*, 2079-2088.
9. Nawrocki, E. P.; Kolbe, D. L.; Eddy, S. R. *Bioinformatics* **2009**, *25*, 1335-1337.

10. Macke, T. J. et al. *Nucleic Acids Res.* **2001**, *29*, 4724-4735.
11. Seshadri, R.; Kravitz, S. A.; Smarr, L.; Gilna, P.; Frazier, M. *PLoS Biol.* **2007**, *5*, 394-397.
12. Markowitz, V. M. et al. *PLoS One* **2012**, *7*, e40151.
13. Human Microbiome Project Consortium *Nature* **2012**, *486*, 215-221.
14. Thomas, T. et al. *ISME J.* **2010**, *4*, 1557-1567.
15. Qin, J. et al. *Nature* **2010**, *464*, 59-65.