

Electronic Supplementary Information

Controlling the direction of site-selectivity and regioselectivity in RNA ligation by Zn²⁺-dependent deoxyribozymes that use 2',3'-cyclic phosphate RNA substrates

Diana M. Kost, Joseph P. Gerdt, P. I. Pradeepkumar and Scott K. Silverman*

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

Deoxyribozyme sequences

Unique sequences of all deoxyribozyme enzyme regions are listed below. Some of the enzyme regions are 39 or 41 nt rather than 40 nt due to a nucleotide deletion or addition during the selection process. For two pairs of the 8CW deoxyribozymes, the sequences could not be determined confidently due to overlap within one sequencing run between the two members of the pair, presumably due to overlapping *E. coli* colonies harboring plasmids with the two different deoxyribozyme sequences. Nonetheless, in each such case, the RNA linkage assignment is clear because partial alkaline hydrolysis assay of the mixture of products clearly revealed that no unnatural linkage was present, and the 8–17 and complement/100 mM Mg²⁺/pH 9.0 assays were similarly unequivocal. For each 7CX and 8CW deoxyribozyme, the binding arm to the 5'-side of the enzyme region is 5'-CGAAGTCGCCATCTCTTC..., and the binding arm to the 3'-side of the enzyme region is ...ATAGTGAGTCGTATTA-3'. For each 12BB deoxyribozyme, the binding arm to the 5'-side of the enzyme region is 5'-CGAAGTCGCCATCTC..., and the binding arm to the 3'-side of the enzyme region is ...GTGAGTCGTATTA-3'. Ligation yields are shown for deoxyribozymes prepared by solid-phase synthesis using either the parent (Par) or transversion (Tv) RNA substrate sequences outside of the unchanged UA↓GG nucleotides near the ligation site. Yields were determined only for deoxyribozymes that form 3'–5' linkages. n.d., yield not determined using a deoxyribozyme prepared by solid-phase synthesis.

7CX deoxyribozymes

<u>Name</u>	<u>Enzyme region sequence, 5' to 3'</u>	<u>Linkage</u>	<u>Yield (Par)</u>	<u>Yield (Tv)</u>
7CX2	GTGAGAGCCATTGTGCATTGTGGTGCAACTGTACGCGCCT	3'–5'	55%	46%
7CX3	GTGAGAGCAGGTTTGGCACGCGTGCCACTAACCCGCGCCT	3'–5'	n.d.	50%
7CX4	GTGGGAGCCTGATGGAAGAATGGGAGAACAGCACGCGCCT	3'–5'	56%	11%
7CX5	GGGAGAGCTAGGAGGCGCTGAGTCTCGGCTCCTCGCGCCT	3'–5'	n.d.	35%
7CX6	GTGAGAGCAAGCGAGGGATACATGAACCGAAACCGCGCCT	3'–5'	44%	43%
7CX7	GTGAGAGCAGTAAACGGAAGGCCCTCTTCCTGTCCGCGCCT	3'–5'	n.d.	50%
7CX8	GTGAGAGCATGTGGCACCCCTGGCAGGAACCTAACGCGCCT	3'–5'	56%	19%
7CX9	ACTGAGAGCCTCGATGAATGCCTAATTATATCGTCGCGCC	3'–5'	57%	32%
7CX10	GTGAGAGCAATACGTATGAAGATATTGAAGTTTCGCGCCT	3'–5'	55%	57%
7CX12	GTGAGAGCATCACTGTAACCGCCGTGCGGTGGACGCGCCT	3'–5'	53%	10%
7CX13	GGGAGAGCAGTTATGATGCATGAGCCATTGACCGCGCCT	3'–5'	n.d.	34%
7CX14	GGGAGAGCAGCGATGCAAGACATTGGGGCGGAGCCGCGCT	3'–5'	n.d.	15%
7CX21	TGAGAGCATTTCCGGCGGAGCTCTACGGGGCCGCGCCT	3'–5'	n.d.	<1%
7CX11	TTTAGTCAGGAGTAGACATCGATGATTGCTAATCGAACCT	2'–5'	n.d.	n.d.

8CW deoxyribozymes

<u>Name</u>	<u>Enzyme region sequence, 5' to 3'</u>	<u>Linkage</u>	<u>Yield (Par)</u>	<u>Yield (Tv)</u>
8CW1	GTGAGAGAACC GCGGTGTAGACACAGATCGCGGCGCGCCT	3'-5'	n.d.	45%
8CW8	TACAAGGTGGGAGGAGGGAGCACCCGATGCGGCATATCGTG	3'-5'	n.d.	50%
8CW11	GGGAGAGCAGTTATGACGCATAAGCTATTGACCGCGCCT	3'-5'	n.d.	15%
8CW15	GTGAGAGCAATCACTTTGGGTAGGTACGGGTGAACGCGCT	3'-5'	n.d.	52%
8CW16	GTTAGAGCCAAACACGTTTGTGTGTCAGCGGGTTTCGCGCCT	3'-5'	n.d.	40%
8CW20	GGAAGAGCAATGTACTCCGACGTCAGAGGATATCGCGCCT	3'-5'	n.d.	39%
8CW17	-	3'-5'	n.d.	n.d.
8CW22	-	3'-5'	n.d.	n.d.
8CW3	ACGCCACCATTAAAGAGCATCGCCGGAATAGCGGGGACT	2'-5'		
8CW4	GCACGGCCGATTGGGGGGCCCAATGACTTGATCATTGCT	2'-5'		
8CW9	ACACGCGTTACGGACTGAGCAGATTCAGGCCAAACTGCCT	2'-5'		
8CW12	ACATGTGCGGTAATGAGGCGTAGACAAATAGAGATACCT	2'-5'		
8CW13	GGGGAAGACAAGATAGTCGAGGGGGACGCGCTCTCAGCCT	2'-5'		
8CW14	GCGGGGTTAAGCCACGAATGCGGGGCAAAGCGTACCCCT	2'-5'		
8CW19	-	2'-5'		
8CW21	-	2'-5'		

12BB deoxyribozymes (sequences of 12BB1, 2, 5, 6, 8, and 12 are from ref. 9)

<u>Name</u>	<u>Enzyme region sequence, 5' to 3'</u>	<u>Linkage</u>	<u>Yield (Par)</u>	<u>Yield (Tv)</u>
12BB5	CAGCGCGATTGAGTGCGTGATTGAAGCTCGGGGTTGGTTA	3'-5'	n.d.	n.d.
12BB12	CAGAGCCCCCTTACGTACAGCCTTTTTAGGTAACCGGGGAG	3'-5'	n.d.	n.d.
12BB24	CAGCGCGATTGGGGGCGTGATTGAAGCTCGGGGTTGGTTA	3'-5'	n.d.	n.d.
12BB26	ACTATACCACCGAGATTCGAATTGGAGCAGTAGTGGCTTG	3'-5'	n.d.	n.d.
12BB33	TGTAGCCTTCTGAAGGTTGGCTGGTTCGGCGAGGTGGGAA	3'-5'	n.d.	n.d.
12BB36	ACCGGTCCCTGTCGGTCAAGGAGTGGCACTGAGGAGAAT	3'-5'	n.d.	n.d.
12BB41	ACGCGAGGGGAGTTCAATCGCTTGTTCGGCAAGGTCGGGA	3'-5'	n.d.	n.d.
12BB1	CCGCTCCGATTGGTGGAGTCTATTGGGGCTGTAGGCGAG	2'-5'		
12BB2	ACCGCGGGAGCTACGTTAGTGGTAACTGCTTGTAGGCGAG	2'-5'		
12BB21	AAGGCAGCAGGCGGCATTTTTTGTCCGTACGTTCTCCTATA	2'-5'		
12BB22	ACCGCGGGAGTTTCGTTAGTGGTACTGCTTGTAGGCGAG	2'-5'		
12BB25	GTAGAGTGGTCCGGTTATCGCCTTAGATAACCATTGCTGT	2'-5'		
12BB27	CCGCTCCGATTGGTGAAGTCTATTGGGGCTGTAGGCGAG	2'-5'		
12BB38	CCGCTGCGGAGGTTGTACGCGTCTGTGGCTTGTAGGCGAG	2'-5'		
12BB6	TACACCTATTATGGTTTTCGTGAGGGGTGTGGCTGGTGCTG	unnat		
12BB8	CACGCTGACTAGCTTCGTGAGGGGGTGTGATAGATGCGG	unnat		
12BB23	TACACCTATAATGGTTTTCGTGAGGGGTGTGGCTGATGCTG	unnat		
12BB30	CCTGTACTGCGTGTCTCAAATCAGCCGGGTGTGTGAACTC	unnat		
12BB32	TGTATGTGGGGGTGTGTATCAGTCTACTGTGGCTTAAGC	unnat		