DNA-Catalyzed Covalent Modification of Amino Acid Side Chains in Tethered and Free Peptide Substrates

On Yi Wong, P. I. Pradeepkumar, and Scott K. Silverman*

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

Table of Contents

Highly preorganized catalyst/substrate architecture used in our previous studies ..... page S2

Characterization of the tripeptide substrates CYA, CSA, and CAA ..... page S3

Assays for additional 9NG deoxyribozymes ..... page S6

Assays for additional 11MN deoxyribozymes ..... page S6

Assays for additional 15MZ deoxyribozymes ..... page S7

Determination of $K_{d,app}$ for 15MZ36 with free CYA tripeptide ..... page S7

Assays for 6QG deoxyribozymes ..... page S8

Assays for 15NZ deoxyribozymes ..... page S8

Dependence of 15MZ36 free peptide substrate reactivity on 3′-terminal composition ..... page S9

MALDI-MS analyses of deoxyribozyme products and their DTT and RNase T1 digestions ..... page S10

Mfold-predicted secondary structures of new deoxyribozymes ..... page S11

References for Supporting Information ..... page S11

Highly preorganized catalyst/substrate architecture used in our previous studies

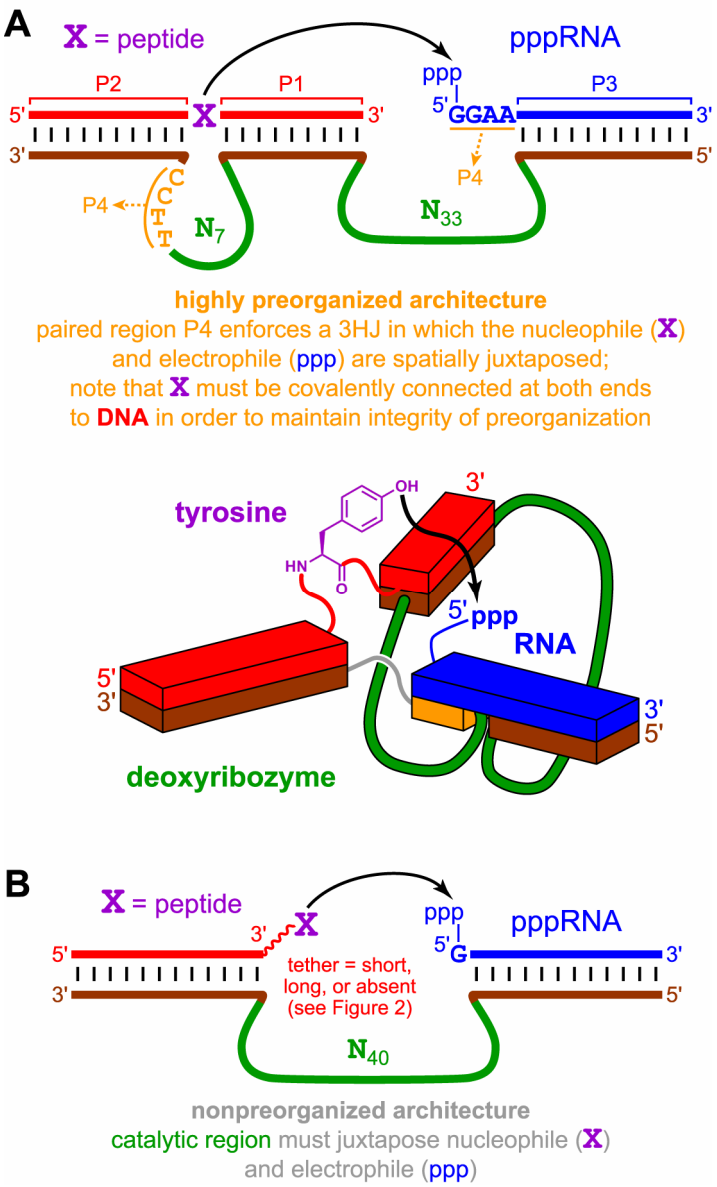
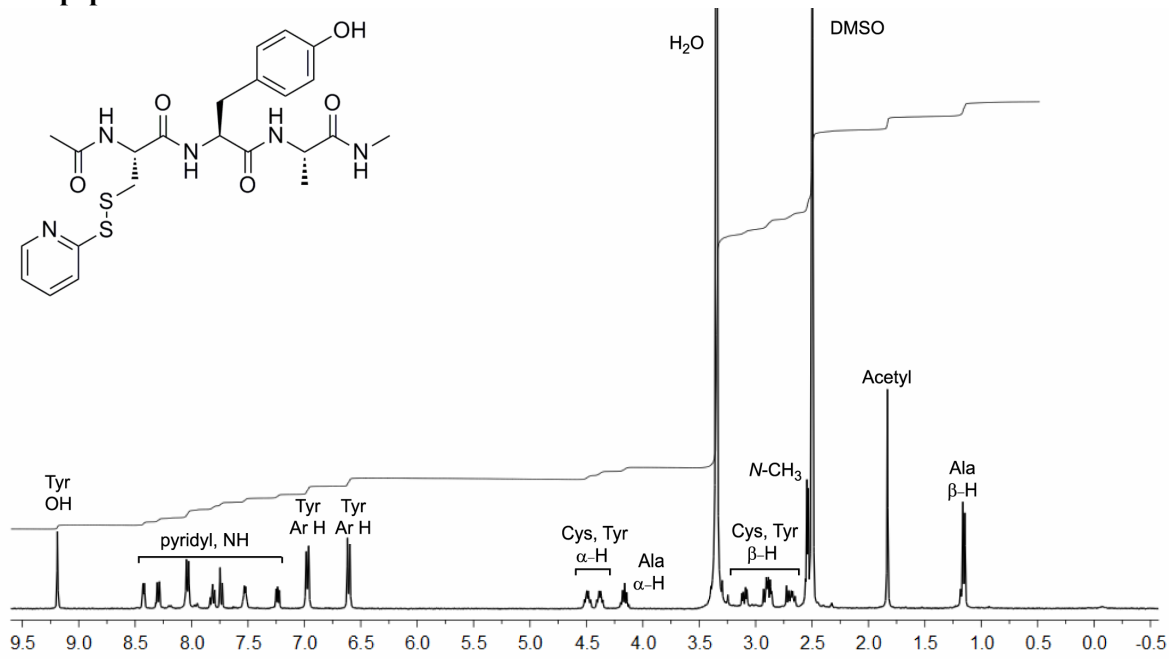


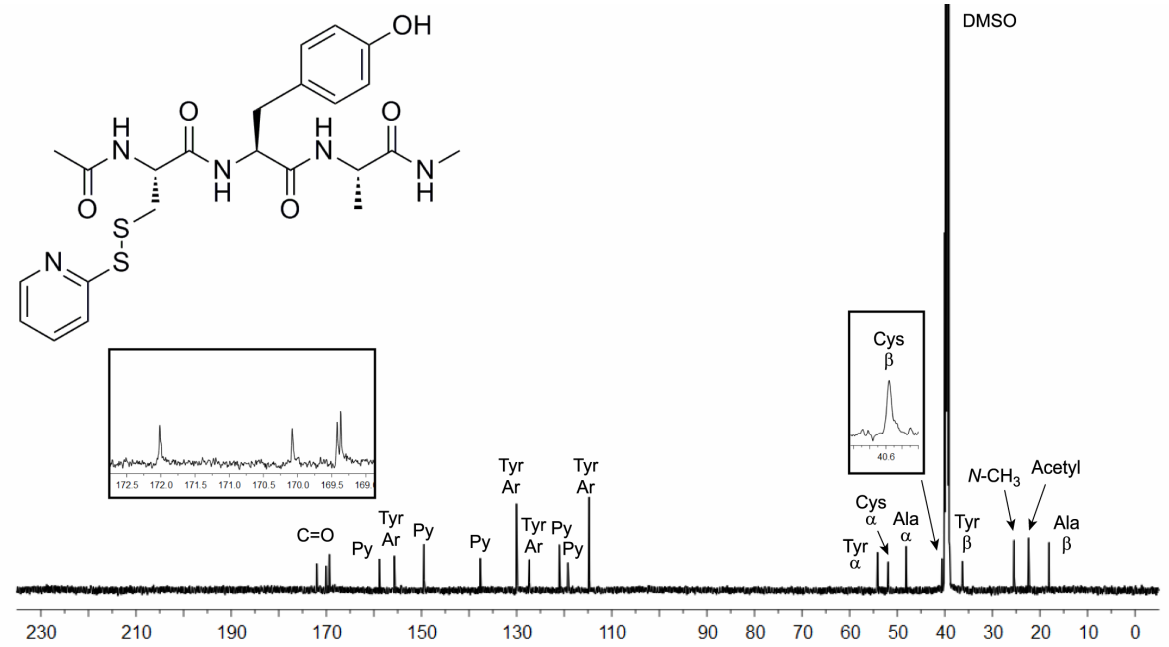
Figure S1. Highly preorganized catalyst/substrate architecture for DNA-catalyzed peptide side chain reactivity, as used in our previous studies (*1, 2*). (A) Formation of paired region P4 enforces a "three-helix junction" (3HJ) (*3, 4*) in which the nucleophile (X) and electrophile (ppp) are spatially juxtaposed. Note that X must be covalently connected twice to DNA (i.e., DNA-peptide-DNA connectivity) in order to maintain the integrity of the preorganization. Therefore, replacement of X with a free peptide that is entirely untethered to DNA will not allow formation of the 3HJ, and the overall approach inherently cannot be used to achieve covalent modification of free peptides. (B) In sharp contrast, the open architecture used for the first time in the current study (see also Figure 2) is amenable to catalytic function with a free peptide, because the tether that is attached to the peptide may be dispensable.

Characterization of the tripeptide substrates CYA, CSA, and CAA

CYA tripeptide



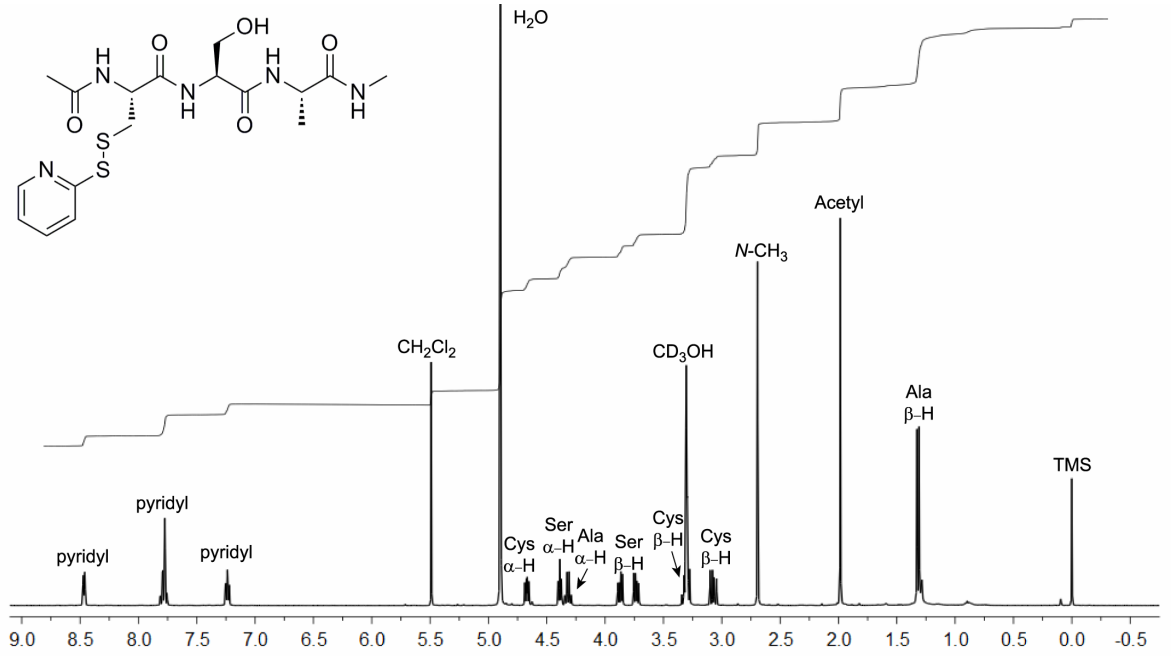
^{1}H NMR (400 MHz, DMSO-d_6): δ 9.19 (s, 1H), 8.43 (d, J = 4.8 Hz, 1H), 8.30 (d, J = 8.0 Hz, 1H), 8.04 (d, J = 7.7 Hz, 2H), 7.81 (td, J = 7.7 Hz, <2 Hz, 1H), 7.74 (d, J = 8.1 Hz, 1H), 7.53 (m, 1H), 7.24 (dd, J = 6.5, 5.0 Hz, 1H), 6.97 (d, J = 8.4 Hz, 2H), 6.61 (d, J = 8.4 Hz, 2H), 4.49 (td, J = 8.5, 5.1 Hz, 1H), 4.38 (td, J = 8.1, 5.0, 1H), 4.16 (quintet, J = 7.1 Hz, 1H), 3.10 (dd, J = 13.3, 4.9 Hz, 1H), 2.93-2.87 (m, 2H), 2.68 (dd, J = 13.9, 8.8 Hz, 1H), 2.54 (d, J = 4.6 Hz, 3H), 1.83 (s, 3H), 1.15 (d, J = 7.2 Hz, 3H) ppm.



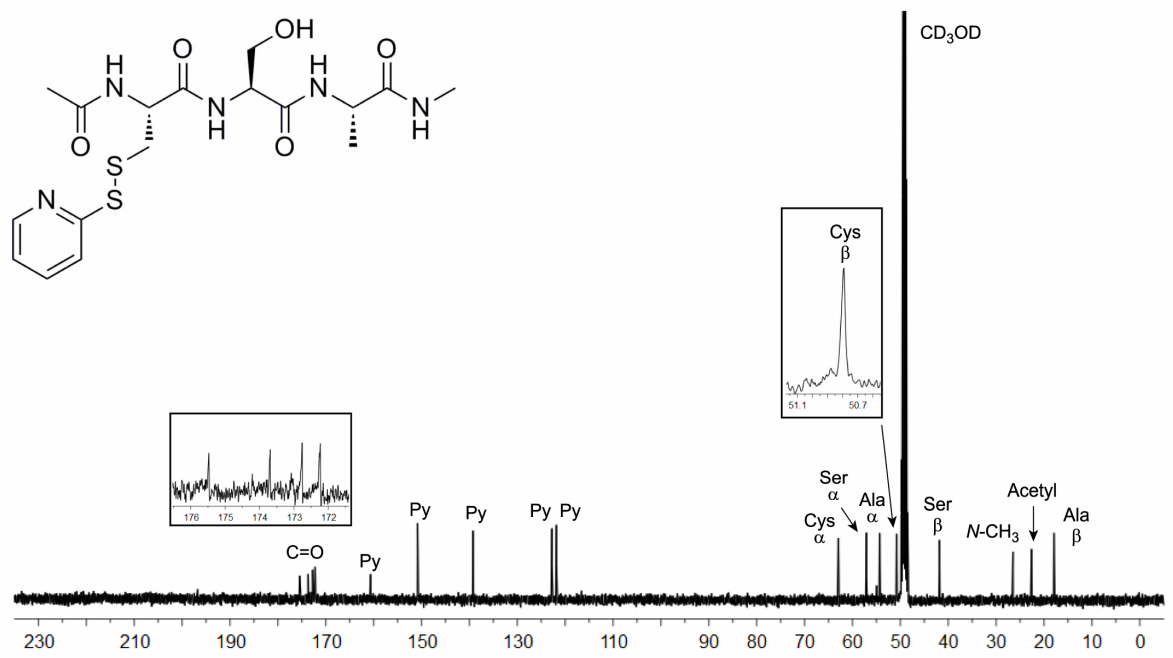
^{13}C NMR (126 MHz, DMSO-d_6): δ 172.01, 170.08, 169.42, 169.37, 158.85, 155.70, 149.45, 137.63, 130.04, 127.30, 121.05, 119.20, 114.73, 54.12, 51.87, 48.12, 40.58, 36.30, 25.41, 22.35, 18.06 ppm.

ESI-HRMS: *m/z* calcd. for $C_{23}H_{29}N_5O_5S_2$ $[M+H]^+$ 520.1688, found 520.1705 (Δm +0.0017, error +3.3 ppm).

CSA tripeptide



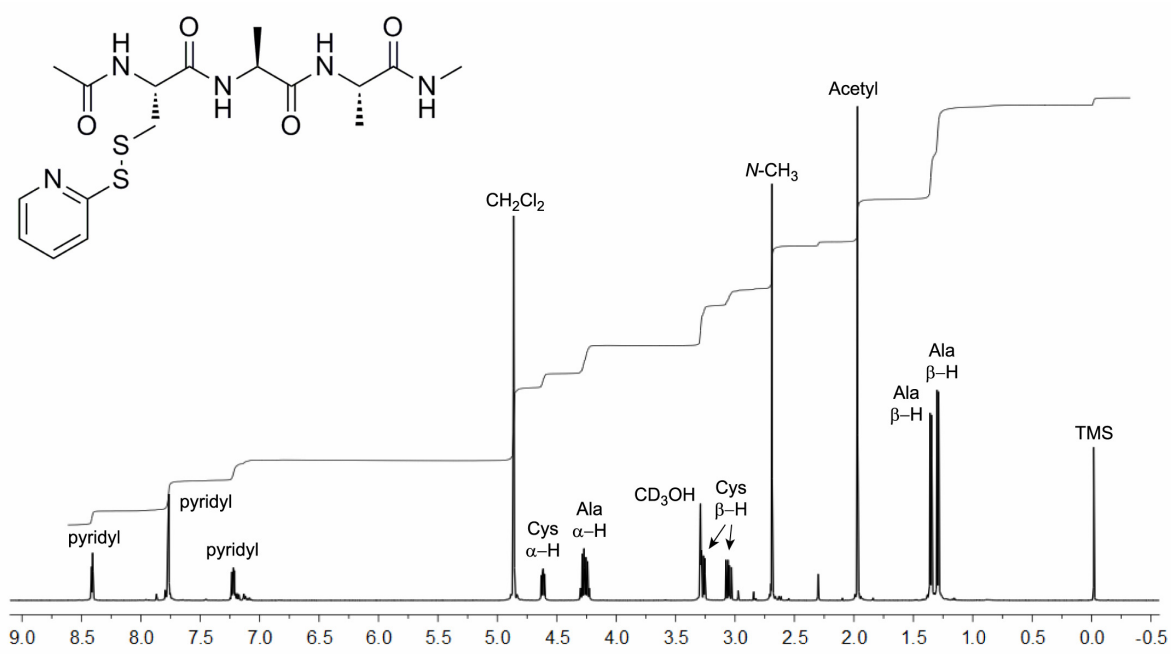
^{1}H NMR (400 MHz, CD_3OD): δ 8.47 (ddd, *J* = 5.0, 1.6, 1.2 Hz, 1H), 7.82-7.75 (m, 2H), 7.24 (ddd, *J* = 6.5, 4.9, 1.9 Hz, 1H), 4.67 (dd, *J* = 8.5, 5.5 Hz, 1H), 4.39 (t, *J* = 5.7 Hz, 1H), 4.30 (q, *J* = 7.3 Hz, 1H), 3.87 (dd, *J* = 10.9, 5.2 Hz, 1H), 3.73 (dd, *J* = 10.8, 6.0 Hz, 1H), 3.29 (dd, *J* = 13.8, 5.6 Hz , 1H), 3.06 (dd, *J* = 13.9, 8.5 Hz, 1H), 2.69 (s, 3H), 1.99 (s, 3H), 1.32 (d, *J* = 7.3 Hz, 3H) ppm.



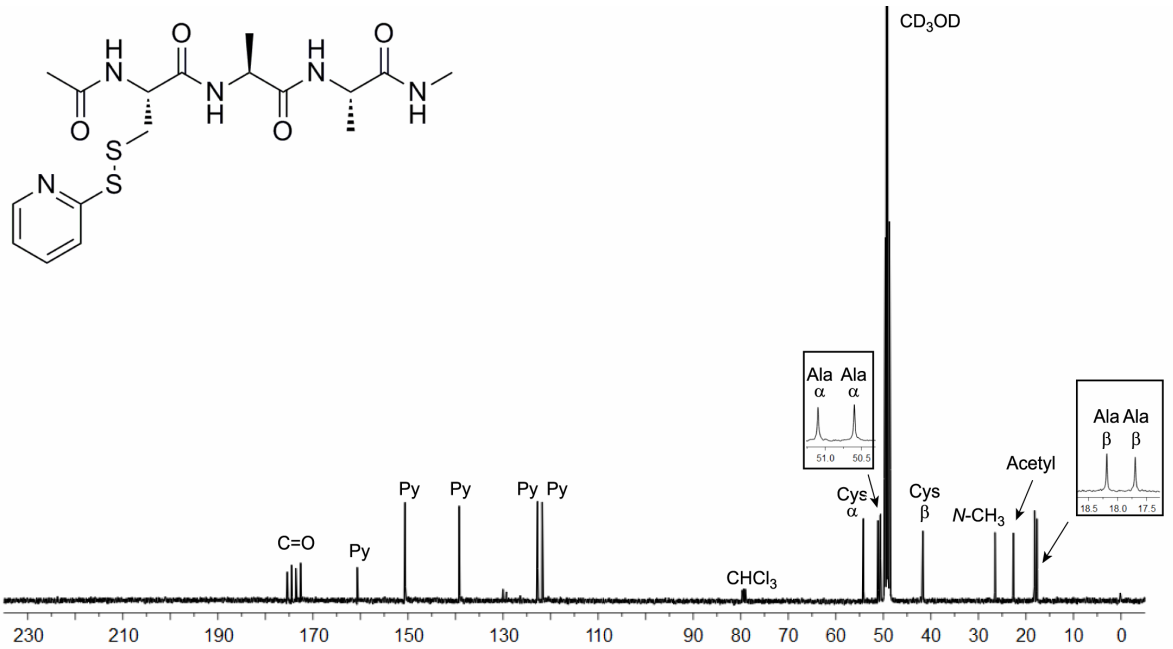
^{13}C NMR (101 MHz, CD_3OD): δ 175.46, 173.68, 172.75, 172.22, 160.66, 150.82, 139.27, 122.79, 121.84, 62.95, 57.09, 54.35, 50.79, 41.83, 26.49, 22.62, 17.93 ppm.

ESI-HRMS: *m/z* calcd. for $C_{17}H_{25}N_5O_5S_2$ $[M+Na]^+$ 466.1185, found 466.1187 (Δm +0.0002, error +0.4 ppm).

CAA tripeptide



^{1}H NMR (500 MHz, CD_3OD): δ 8.43 (ddd, *J* = 4.9, 1.7, 1.0 Hz, 1H), 7.83-7.77 (m, 2H), 7.24 (ddd, *J* = 6.0, 4.9, 2.5 Hz, 1H), 4.64 (dd, *J* = 8.6, 5.4 Hz, 1H), 4.30 (q, *J* = 7.2 Hz, 1H), 4.27 (q, *J* = 7.2 Hz, 1H), 3.30 (dd, *J* = 13.9, 5.4Hz, 1H), 3.07 (dd, *J* = 13.9, 8.6 Hz, 1H), 2.71 (s, 3H), 1.99 (s, 3H), 1.36 (d, 7.2 Hz, 3H), 1.29 (d, 7.2 Hz, 3H) ppm.



^{13}C NMR (101 MHz, CD_3OD): δ 175.40, 174.52, 173.62, 172.59, 160.69, 150.62, 139.24, 122.74, 121.77, 54.20, 51.10, 50.60, 41.69, 26.51, 22.66, 18.19, 17.69 ppm.

ESI-HRMS: *m/z* calcd. for $C_{17}H_{25}N_5O_4S_2$ $[M+H]^+$ 428.1426, found 428.1422 (Δm –0.0004, error –0.9 ppm).

Assays for additional 9NG deoxyribozymes

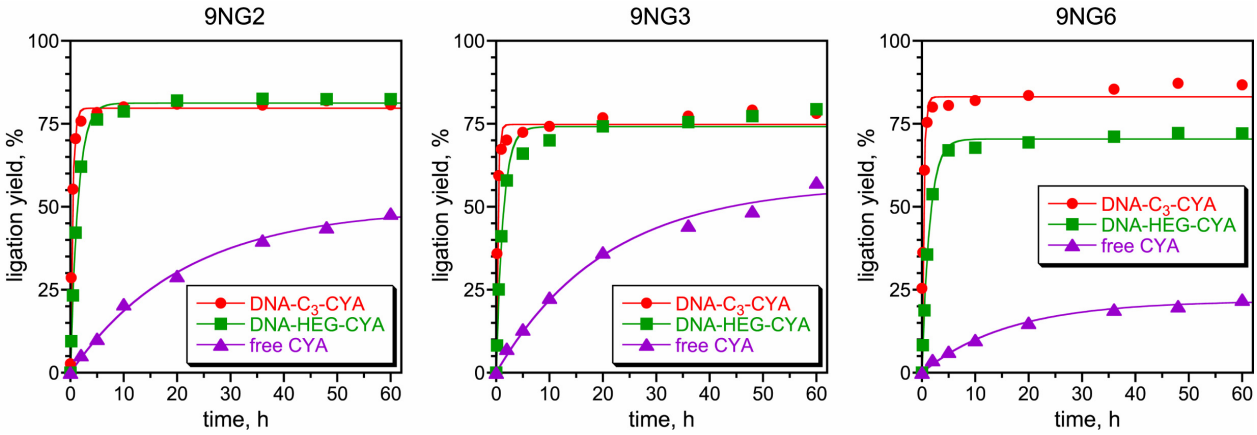


Figure S2. Kinetic data for three 9NG deoxyribozymes other than 9NG14 (compare with Figure 5A). Data for the remaining two 9NG deoxyribozymes, 9NG5 and 9NG15, was comparable (not shown).

Assays for additional 11MN deoxyribozymes

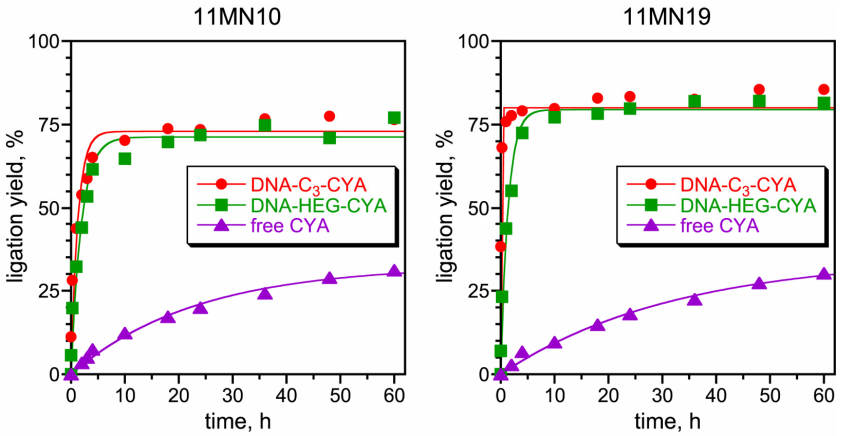


Figure S3. Kinetic data for 11MN deoxyribozymes other than 11MN5 (compare with Figure 5B).

Assays for additional 15MZ deoxyribozymes

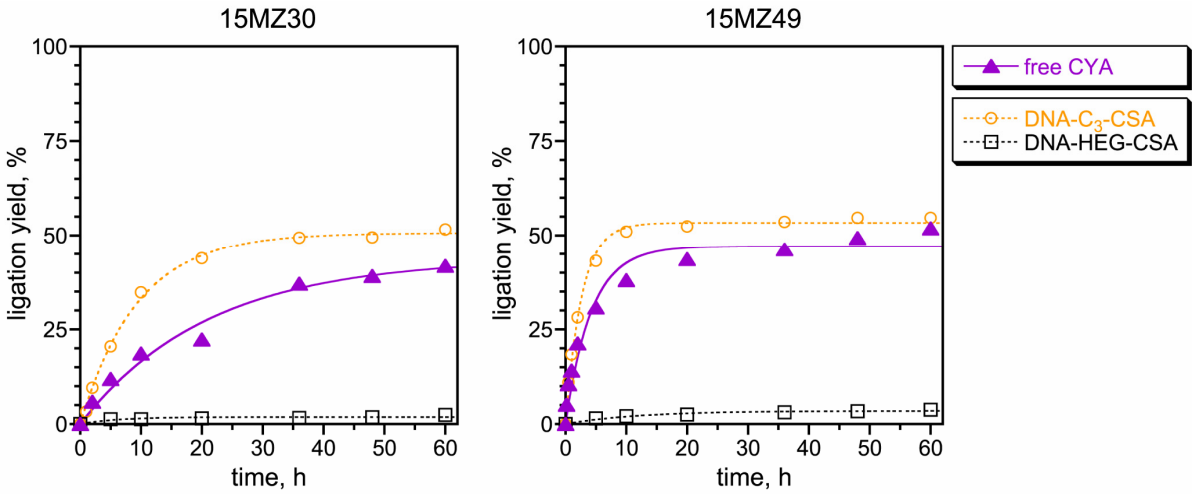


Figure S4. Kinetic data for 15MZ deoxyribozymes other than 15MZ36 (compare with Figure 6). Each additional deoxyribozyme was analyzed with the three illustrated substrates to assess potential improvement in activity relative to 15MZ36.

Determination of $K_{d,app}$ for 15MZ36 with free CYA tripeptide

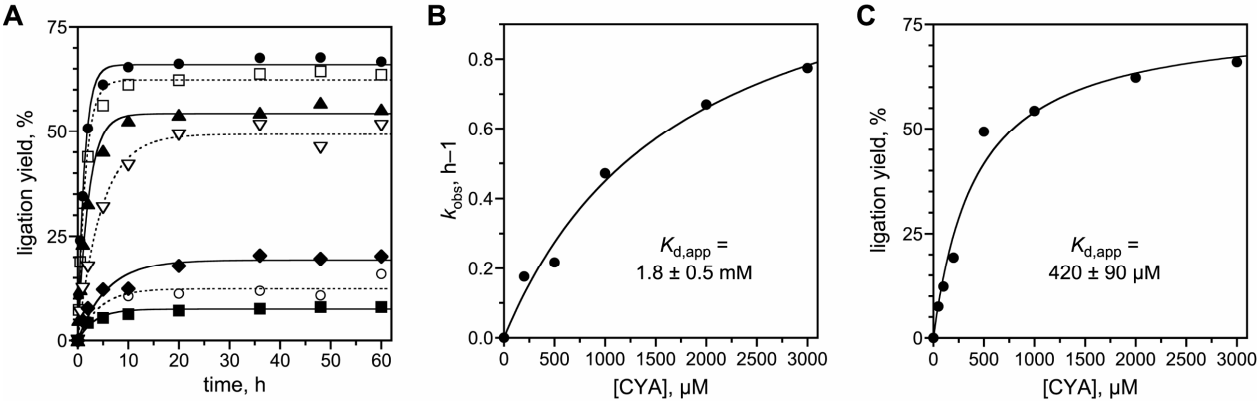


Figure S5. Determination of $K_{d,app}$ for 15MZ36 with the free CYA tripeptide substrate. (A) 15MZ36 ligation yield was evaluated with CYA concentrations of 50, 100, 200, and 500 μM as well as 1, 2, and 3 mM. Data were fit to the standard equation $Y = Y_{max} \cdot C/(K_d + C)$, where C is the tripeptide concentration. (B) From k_{obs} data, the $K_{d,app}$ value for the free CYA tripeptide substrate was 1.8 ± 0.5 mM. (C) From yield data, the $K_{d,app}$ value was 420 ± 90 μM.

Assays for 6QG deoxyribozymes

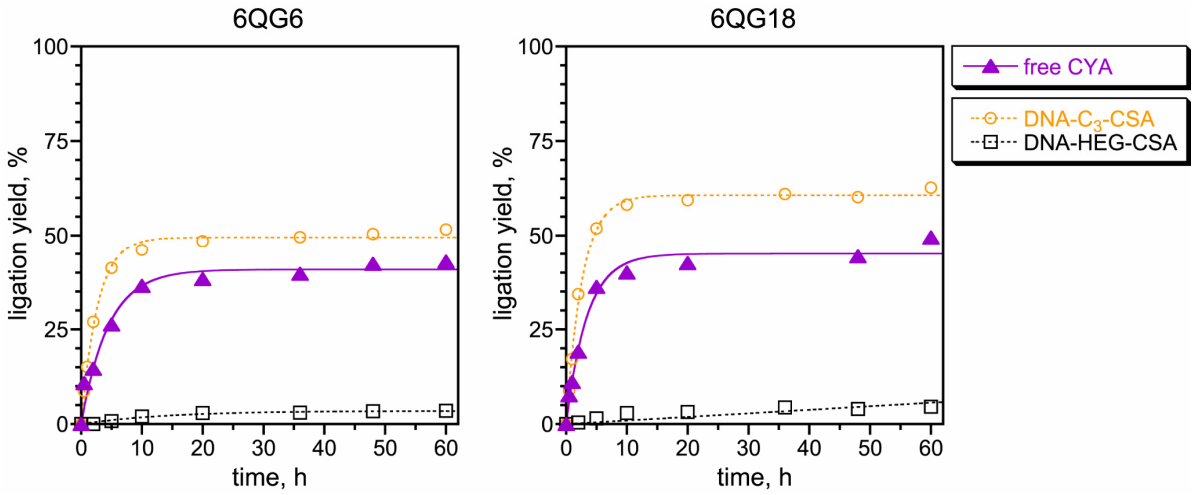


Figure S6. Kinetic data for 6QG deoxyribozymes (compare with Figure 6). Each deoxyribozyme was analyzed with the three illustrated substrates to assess potential improvement in activity relative to 15MZ36. In the reselection experiment that led to these deoxyribozymes, the partially (25%) randomized pool derived from the 15MZ36 sequence was selected with the DNA-C_3-CSA substrate with 15 h incubation for three rounds and 1 h incubation for three more rounds, at which point the pool ligation yield was 13%, and cloning was performed.

Assays for 15NZ deoxyribozymes

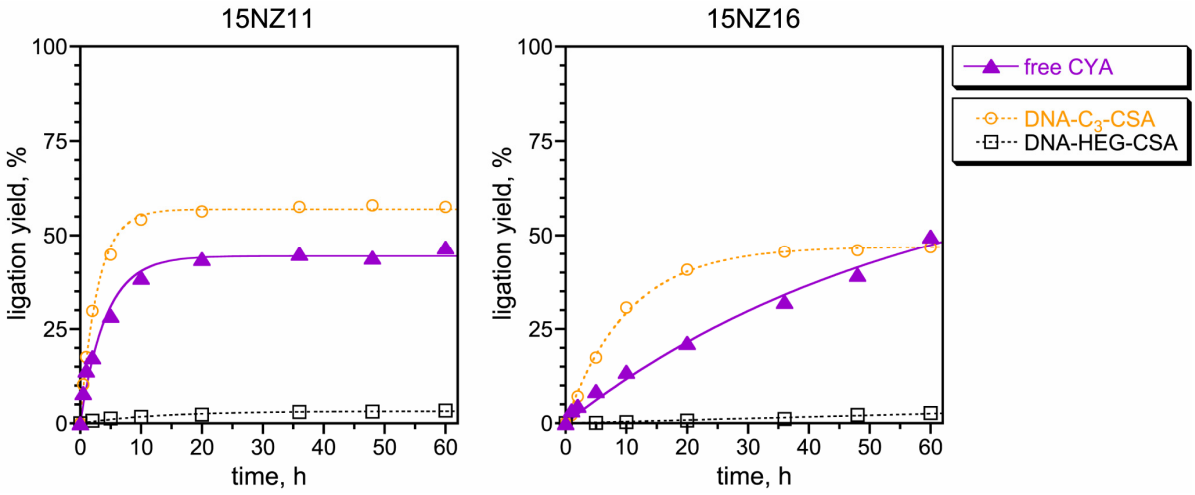


Figure S7. Kinetic data for 15NZ deoxyribozymes (compare with Figure 6). Each deoxyribozyme was analyzed with the three illustrated substrates to assess potential improvement in activity relative to 15MZ36. In the selection experiment that led to these deoxyribozymes, the round 11 pool from the selection with the DNA-TEG-OH substrate (2 h incubation) was then continued for two rounds with the DNA-C_3-CYA substrate (2 h incubation) and two more rounds with DNA-C_3-CYA (10 min incubation), at which point the pool ligation yield was 30%, and cloning was performed.

Dependence of 15MZ36 free peptide substrate reactivity on 3′-terminal composition of deoxyribozyme

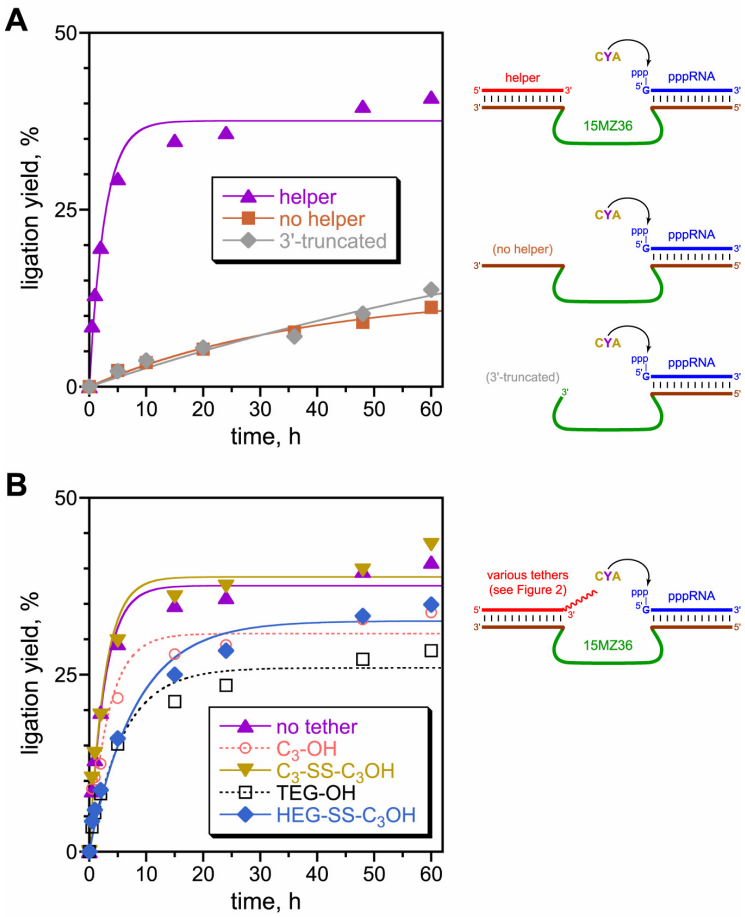


Figure S8. Dependence of 15MZ36 free peptide substrate reactivity on composition of the 3′-terminal region of the deoxyribozyme. (A) Effect of omitting helper oligonucleotide or truncating 15MZ36 3′-binding arm. (B) Effect of including various combinations of tether atoms at 3′-end of helper oligonucleotide. For both panels, similar observations were made for 9NG14 and 11MN5 (data not shown). See Figure 2 for C_3 and HEG structures. TEG denotes tri(ethylene glycol). SS-C_3-OH denotes the protected disulfide linker (IDT) not yet reduced to form a thiol.

MALDI-MS analyses of deoxyribozyme products and their DTT and RNase T1 digestions

deoxyribozyme and substrate	mass calcd. [a]	mass found	error, % (found – calcd.)
10KC3 with **DNA-C_3-CYA**			
ligation product	12026.7	12025.6	–0.01
DTT digestion (L product) [b]	5948.7	5950.2	+0.03
DTT digestion (R product) [b]	6079.0	6080.2	+0.02
RNase T1 digestion	6782.4	6781.2	–0.02
11MN5 with **DNA-HEG-CYA**			
ligation product	12371.0	12369.2	–0.02
DTT digestion (L product) [b]	6293.0	6295.2	+0.04
DTT digestion (R product) [b]	6079.0	6079.3	+0.01
RNase T1 digestion	7126.7	7126.8	0.00
15MZ36 with **DNA-C_3-CSA**			
ligation product	11950.6	11948.1	–0.02
DTT digestion (L product) [b]	5948.7	5948.9	+0.003
DTT digestion (R product) [b]	6002.9	6002.7	–0.003
RNase T1 digestion	6706.3	6714.8	+0.13
15MZ36 with **untethered CYA**			
ligation product	6167.1	6165.9	–0.02
DTT digestion [c]	6079.0	6078.7	–0.005
RNase T1 digestion	924.8 [d]	924.5 [d]	–0.03

Table S1. MALDI mass spectrometry analyses of key deoxyribozyme products and their DTT and RNase T1 digestions. See Figure 7 for a depiction of the digestion reactions. All MALDI mass spectra were obtained in the mass spectrometry laboratory of the UIUC School of Chemical Sciences. See the Experimental Procedures for reaction details.

[a] Calculated masses are for $[M–H]^-$, because the mass spectra were obtained in negative ion mode except as indicated.

[b] "L product" refers to the left-hand product (DNA anchor oligonucleotide + tethered thiol) and "R product" refers to the right-hand product (tripeptide + RNA), as shown in Figure 7.

[c] When the substrate is a free tripeptide, DTT digestion leads only to the tripeptide-RNA product as detectable by mass spectrometry.

[d] Mass spectrum obtained in positive ion mode, $[M+H]^+$.

Mfold-predicted secondary structures of new deoxyribozymes

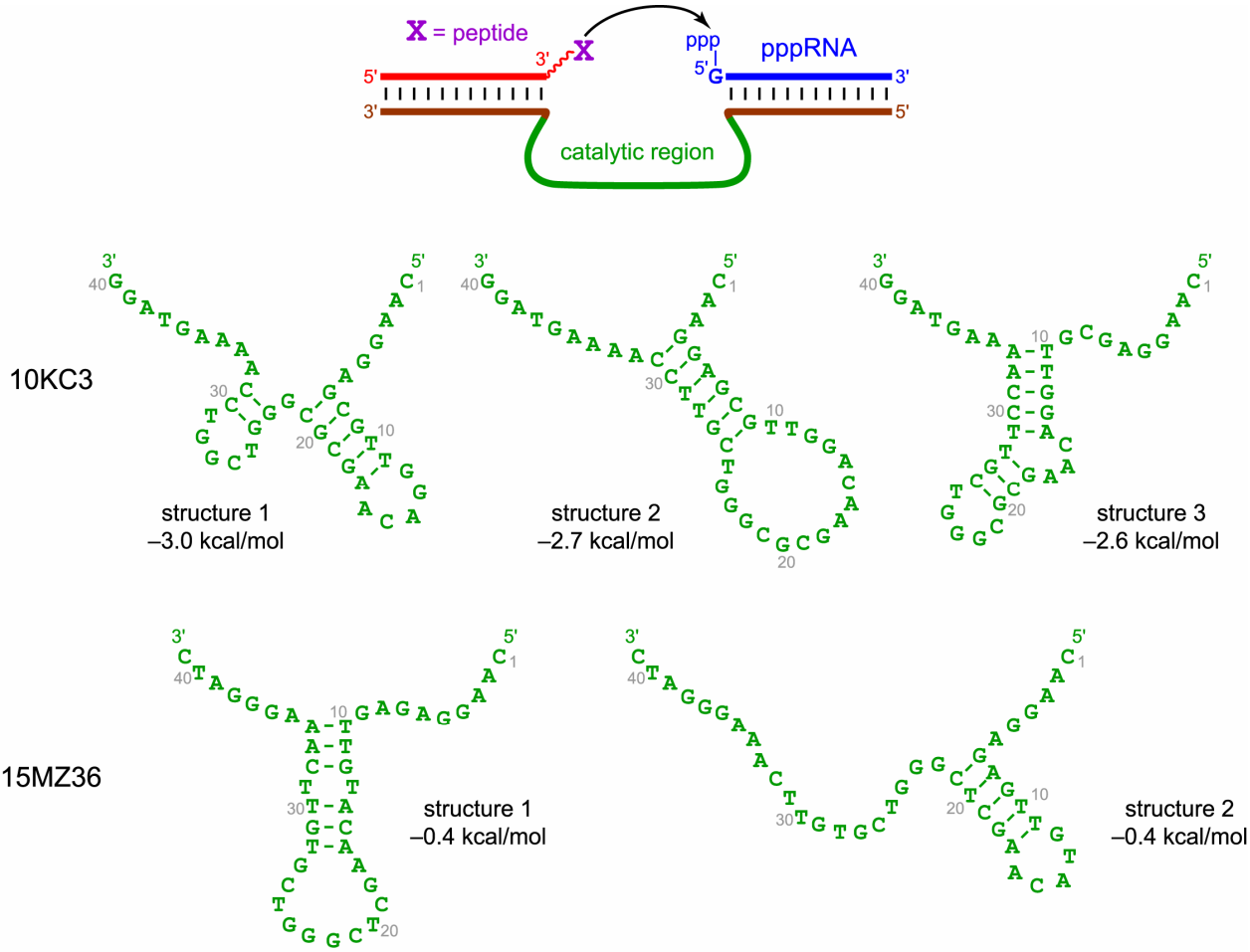


Figure S9. Predicted secondary structures of the 10KC3 and 15MZ36 deoxyribozymes, as computed using mfold (*5*). Only the catalytic region sequences are shown. For 10KC3, the three predicted structures are all of modest folding energy (–3.0 to –2.6 kcal/mol), and the predicted stem-loop element(s) are located in different places in each structure. For 15MZ36, the two predicted structures are of essentially negligible folding energy (each –0.4 kcal/mol), and again the stem-loop elements are in different places. Experimentally distinguishing and validating such secondary structures would require considerable work, and such findings would not immediately clarify either tertiary structure or catalytic mechanism.

References for Supporting Information

1. Pradeepkumar, P. I., Höbartner, C., Baum, D. A., and Silverman, S. K. (2008) DNA-Catalyzed Formation of Nucleopeptide Linkages, *Angew. Chem. Int. Ed. 47*, 1753-1757.
2. Sachdeva, A., and Silverman, S. K. (2010) DNA-Catalyzed Serine Side Chain Reactivity and Selectivity, *Chem. Commun. 46*, 2215-2217.
3. Coppins, R. L., and Silverman, S. K. (2004) A DNA Enzyme that Mimics the First Step of RNA Splicing, *Nat. Struct. Mol. Biol. 11*, 270-274.
4. Coppins, R. L., and Silverman, S. K. (2005) A Deoxyribozyme That Forms a Three-Helix-Junction Complex with Its RNA Substrates and Has General RNA Branch-Forming Activity, *J. Am. Chem. Soc. 127*, 2900-2907.
5. Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction, *Nucleic Acids Res. 31*, 3406-3415.