Structure-Function Correlations Derived from Faster Variants of an RNA Ligase Deoxyribozyme

Tracey K. Prior, Daniel R. Semlow, Amber Flynn-Charlebois, Imran Rashid, and Scott K. Silverman*

Department of Chemistry, University of Illinois at Urbana-Champaign, 600 S. Mathews Ave., Urbana, Illinois 61801

Figures and Tables in this Supplementary Material are prefixed by the letter X (e.g., Figure X1) to distinguish them from those in the manuscript. All references cited by number are from the manuscript. See the manuscript's Materials and Methods Section and ref. 1 for further experimental details.

Calculations of extent of pool randomization

Each deoxyribozyme strand was prepared with its 40-nucleotide DNA enzyme region subjected to 25% randomization at each position relative to the parent sequence (this is a typical level of randomization; ref. 10). It is straightforward to calculate the distribution of nucleotide changes per molecule relative to the parent sequence as a function of the fraction parent nucleotide at each position. Let x = fraction "correct" nucleotide at each individual position (x has the same value for each nucleotide position in a particular selection pool). Define P(n) = the probability of having a total of n changes relative to the parent sequence. It is readily shown that $P(n) = x^{m-n} \bullet (1-x)^n \bullet {}_mC_n$, where m = the length of the sequence and ${}_mC_n$ denotes the combinatorial function of m objects taken n at a time. For a 40-nucleotide enzyme region, $m = 40$, P(n) is completely determined by x. Calculated values of P(n) for various values of x are shown in Figure X1.

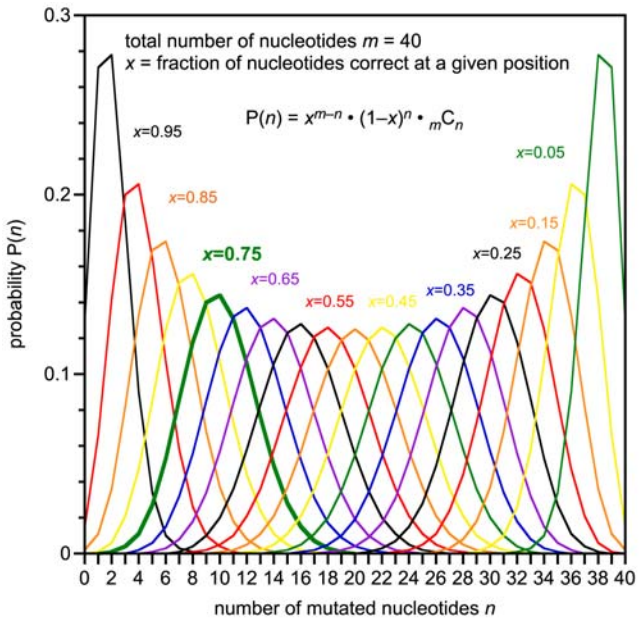

Figure X1. Calculation of probability distribution of number of nucleotide changes n as a function of fraction correct nucleotide x for the re-selections, with $m = 40$ for the 40-nucleotide random region. See text for explanation.

Although the most likely number of nucleotide changes is $n = 10$ for $x = 0.75$ (i.e., 25% randomization; left-most green curve), there is a significant tail on either side of the distribution, including many sequences with more than $n = 20$ mutations. For $n = 30$ mutations with $x = 0.75$,

$P(30) = 4.1\times10^{-11}$, and the sum of P(30) through P(40) is 4.6×10^{-11}. Therefore, in 200 pmol (1.2×10^{14} molecules) of the randomized pool synthesized with $x = 0.75$, about 4900 molecules are expected to have exactly 30 mutations relative to the parent sequence, and about 5500 molecules are expected to have 30 or more mutations. For $n = 25$ mutations with $x = 0.75$, $P(25) = 4.8\times10^{-7}$, and the sum of P(25) through P(40) is 5.9×10^{-7}. Therefore, in 200 pmol of the randomized pool synthesized with $x = 0.75$, about 58 million molecules are expected to have exactly 25 mutations relative to the parent sequence, and about 71 million molecules are expected to have 25 or more mutations.

Deoxyribozyme sequence alignments and preliminary kinetic characterizations

In Tables X1–X3 are sequence alignments for the deoxyribozymes related to 7Z81, 7Z48, and 7Z101. The sequences are written in the 5'-to-3' direction. Only one of the two binding arms is shown. The 5'-side DNA binding arm (termed the right-hand DNA binding arm, because it binds to the right-hand RNA substrate; see Figure X1) is not shown because it was 5'-CGAAGTCGCCATCTC-3' in all sequenced clones, as used during selection. In contrast, the 3'-side (left-hand) DNA binding arm often has mutations that arose due to the use of *Taq* polymerase during the selection. Only the left-hand binding arm is susceptible to such mutation because it is amplified by *Taq* polymerase rather than originating in an oligonucleotide primer during each selection round. The sequences in the tables are color-coded. The enzyme region consensus is black; nucleotide differences from the consensus are blue. The left-hand DNA binding arm is violet; mutations within this binding arm are grey.

clone	7Z81 and related sequences, 5' to 3'
N_{40} pool:	**NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN**GTGAGTCGTATTA
7Z10	ACGGCGAGTTTCATGGAGCGATTGGGAGGTTAGCTCTAGTG**T**GTCGTATT**G**
7Z67	ACGGCGAGTTTCATGGAGCGATTGGGAGGTTAGCTCTAGTG**T**GTCGTATT**G**
7Z81	ACGGCGAGTTTCATGGAG**T**GATTGGGAGGTTAGCTCTAGTG**T**GTCGTATTA
7Z88	ACGGCGAGTTTCATGGAGCGATTGGGAGGTTAGCTCTAGTG**T**GTCGTATT**G**
7Z97	ACGGCGAGTTTCATGGAGCGATTGGGAGGTTAGCTCTAGTG**T**GTT**T**GTATT**G**
7Z98	ACGGCGAGTTTCATGGAGCGATTGGGAGGTTAGCTCTAGTG**T**GTCGTATT**G**
7Z108	ACGGCGAGTTTCATGGAGCGATTGGGAGGTTAGCTCTAGTG**G**GTCGT**GG**TA

Table X1. Sequences for 7Z81 and related deoxyribozymes. See text for details. When the single **T** in 7Z81 was changed to C, the ligation activity remained approximately unchanged (data not shown).

clone	7Z48 and related sequences, 5' to 3'
N_{40} pool:	**NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN**GTGAGTCGTATTA
7Z14	TGGGGGCCGGTCTGCGTGCCTGATTGGGAGGTTAGCTCTAGTGAGTCGT**G**TTA
7Z48	**AC**GGGGCCGGT**T**TGCGTGCCTGATTGGGAGGTTAGCTCTAGTGAGTCGT**G**TTA
7Z70	TGGGGGCCGGTCTGCGTGCCTGATTGGGAGGTTAGCTCTAGTGAGTCGT**G**TTA
7Z93	TGGGGGCCGGTCTGCGTGCCTGATTGGGAGGTTAGCTCTAGTGAGTCGTATT**G**
7Z99	**A**GGGGGCCGGTCTGCGTGCCTGATTGGGAGGTTAGCTCTAGTGAGTCGTATT**G**
7Z103	**A**GGGGGCCGGTCTGCGTGCCTGATTGGGAGGTTAGCTCTAG**C**GAGTCGTATTA
7Z105	TGGGGGCCGGTCT**A**CGTGCCTGATTGGGAGGTTAGCTCTAGTGAGTCGT**G**TTA

Table X2. Sequences for 7Z48 and related deoxyribozymes.

clone	7Z101 and related sequences, 5' to 3'
N_{40} pool:	**NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN**GTGAGTCGTATTA
7Z7	CCCGAGGAGGGGCGG**A**GGGATTTGGTGTGGAGTTTCATTCGTG**G**GTT**G**TATT**G**
7Z50	CCCGAGGAGGGGCGGGGGGATTTGGTGTGGAGTTTCATTCGTG**G**GTCGTATT**G**
7Z61	CCCGAGGAGGGGCGG**A**GGGATTTGGTGTGGAGTTTCATTCGTG**G**GTCGT**G**TTA
7Z65	CCCGAGGAGGGGCGGGGGGACTTGGTGTGGAGTTTCATTCGTG**G**GTCGTATTA
7Z101	CCCGAGGAGGGGCGGGGGGACTTGGTGTGGAGTTTCATTCGTG**T**GTCGTATT**G**
7Z102	CCCGAGGAGGGGCGGGGGGACTTGGTGTGGAGTTTCATTCGTG**T**GT**G**GTATT**G**

Table X3. Sequences for 7Z101 and related deoxyribozymes.

Predicted secondary structures for the 7Z101 deoxyribozyme

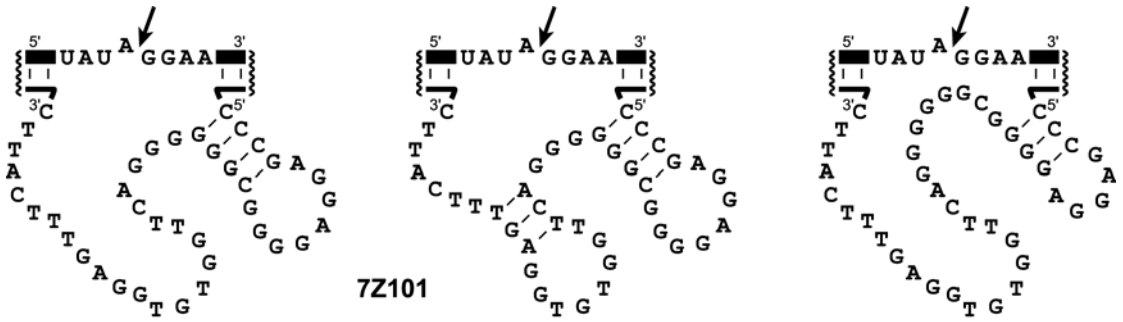

Figure X3. Secondary structures predicted by mfold for the 7Z101 deoxyribozyme. No predicted structure is strongly preferred; the computed ΔG for each is within 1 kcal/mol of zero at 0–10 mM $MgCl_2$.