Electronic Supplementary Material (ESI) for RSC Chemical Biology.

Supporting Information for

Mutant polymerases capable of 2' fluoro-modified nucleic acid synthesis and amplification with improved accuracy

Trevor A. Christensen, Kristi Y. Lee, Simone Z. P. Gottlieb, Mikayla B. Carrier and Aaron M. Leconte*

Supporting Information – Contents

I. Supplementary Tables and Figures

Table S1. Genotype of mutant polymerases studied. Amino acid changes are shown relative to the Stoffel fragment of Taq DNA polymerase

Mutant Name	Amino Acid Changes
SFP1	I614E, E615G, E742R, A743R
SFP4	I614E, E615G, E742K, A743R
SFP7	I614E, E615G, K314R, E520G, F598L, A608V, E742G
SFM4-3	I614E, E615G, V518A, N583S, D655N, E681K, E742Q, M747R
SFM4-6	I614E, E615G, D655N, L657M, E681K, E742N, M747R
SFM4-9	I614E, E615G, N415Y, V518A, D655N, L657M, E681V, E742N, M747R
SFM4-6:LVL*	I614E, E615G, D655N, L657M, E681K, E742N, M747R, Q782L, H784L
SFM4-6:ETL*	I614E, E615G, D655N, L657M, E681K, E742N, M747R, A661E, I665T, F667L
SFM4-3:LVL*	I614E, E615G, V518A, N583S, D655N, E681K, E742Q, M747R, Q782L, H784L
SFM4-3:ETL*	I614E, E615G, V518A, N583S, D655N, E681K, E742Q, M747R, A661E, I665T, F667L
SFP1:LVL*	I614E, E615G, E742R, A743R, Q782L, H784L
SFP1:ETL*	I614E, E615G, E742R, A743R, A661E, I665T, F667L

*created for this study

Table S2. Error rate and spectrum of XNA synthesis by SFP1, SFP4, SFP7, SFM4-3, SFM4-6, and SFM4-9

Error Type	SFP1	SFP4	SFP7	SFM4-3	SFM4-6	SFM4-9	Template
Reads analyzed	27333	62451	123984	23993	3327	164470	84851
Total errors	5.63	7.38	12.17	6.90	18.04	16.67	3.23
Insertions	0.26	0.28	0.64	0.35	1.14	1.02	0.16
Deletions	2.80	3.17	4.42	2.87	8.42	5.48	2.66
Substitutions	2.57	3.93	7.12	3.68	8.48	10.16	0.41
C→T	6.09	9.62	12.12	7.25	14.86	23.68	0.55
T→C	0.59	2.24	4.49	1.12	3.09	5.06	0.30
G→A	0.55	1.34	1.50	1.89	4.27	1.71	0.30
A→G	1.09	2.01	5.40	1.05	3.20	5.17	0.16
T→A	0.33	0.21	1.32	0.63	1.69	1.50	0.05
T→G	0.19	0.02	0.21	0.28	0.51	0.26	0.06
A→C	0.14	0.74	1.14	0.17	0.95	1.50	0.18
A→T	0.44	1.03	2.36	0.55	1.48	4.07	0.05
G→C	0.15	0.05	0.16	0.48	0.90	0.18	0.03
G→T	0.04	0.25	0.47	0.07	0.33	0.74	0.00
C→A	0.54	0.18	0.48	1.01	2.12	0.82	0.06
C→G	0.12	0.10	0.97	0.20	0.51	1.05	0.09

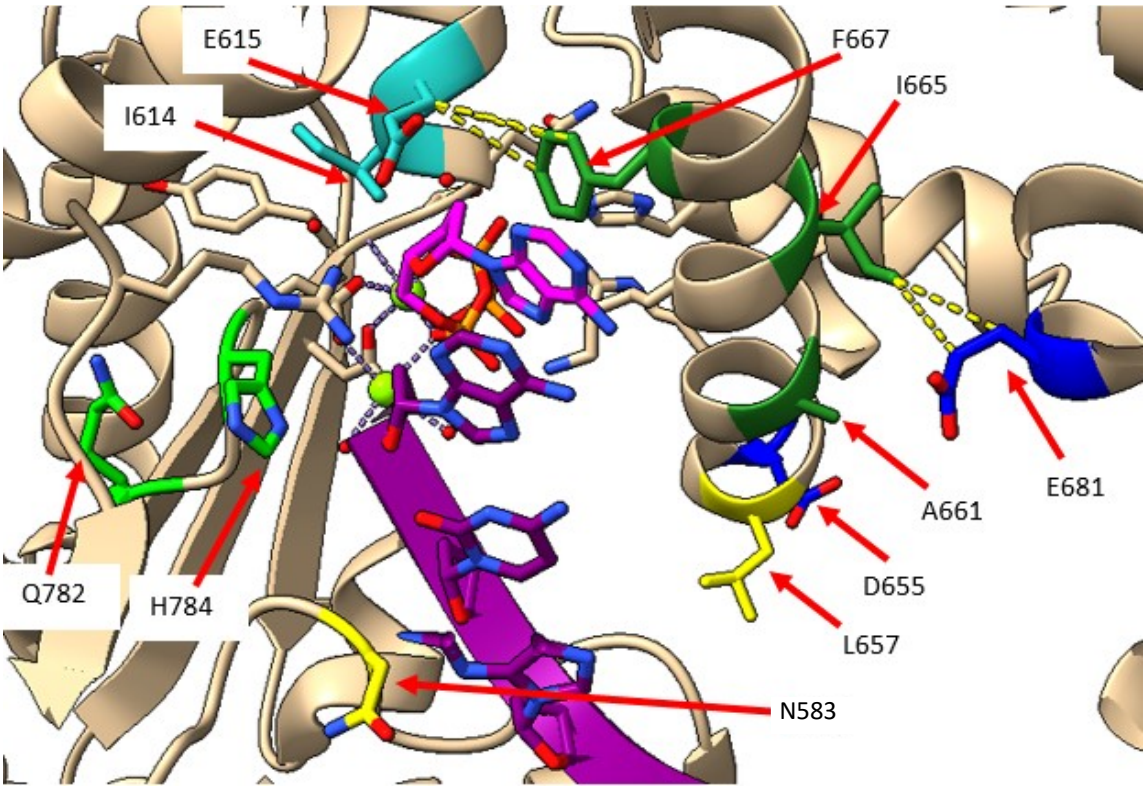


Figure S1. ChimeraX rendering of Taq polymerase (PDB:1QSY) active site with various XNA polymerase mutations and their interactions (yellow dashes) within 5 Angstroms. Magenta = incoming nucleotide, purple = primer strand, dark green = ETL, light green = LVL, light blue = 4-3, 4-6 and P1 mutations, dark blue = 4-3 and 4-6 mutations, yellow = unique to 4-3 (N583) or 4-6 (L657).

Table S3. Error rate and spectrum of XNA synthesis by SFM4-6, SFM4-6:LVL, and SFM4-6:ETL on 100bp template.

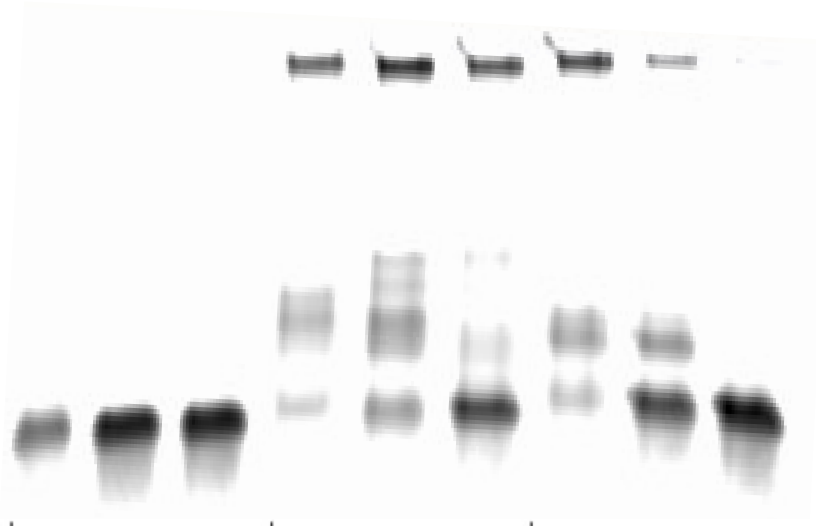
Error Type	SFM4-6	SFM4-6:LVL	SFM4-6:ETL	Template
Reads analyzed	3327	2924	2480	1104
Total errors	18.04	11.44	11.33	3.82
Insertions	1.14	1.05	2.77	0.10
Deletions	8.42	5.78	5.77	3.01
Substitutions	8.48	4.61	2.78	0.71
C→T	14.86	6.98	4.12	1.08
T→C	3.09	1.92	0.96	0.46
G→A	4.27	2.27	2.16	0.46
A→G	3.20	2.18	1.72	0.23
T→A	1.69	1.05	0.27	0.08
T→G	0.51	0.20	0.24	0.08
A→C	0.95	0.84	0.07	0.08
A→T	1.48	0.93	0.10	0.08
G→C	0.90	0.38	0.10	0.08
G→T	0.33	0.26	0.14	0.00
C→A	2.12	1.05	0.82	0.08
C→G	0.51	0.32	0.21	0.15

Table S4. Error rate and spectrum of XNA synthesis by SFM4-6, SFM4-6:LVL, and SFM4-6:ETL on 150bp template

Error Type	SFM4-6	SFM4-6:LVL	SFM4-6:ETL	Template
Reads analyzed	80086	7829	18230	28007
Total errors	19.16	13.78	12.48	4.13
Insertions	6.65	3.34	3.75	1.12
Deletions	9.64	7.13	5.97	2.54
Substitutions	2.87	3.31	2.76	0.46

Table S5. Error rate and spectrum of XNA synthesis by SFM4-3, SFM4-3:LVL, SFM4-3:ETL, SFP1, SFP1:LVL, and SFP1:ETL on 100bp template

Error Type	SFP1	SFP1: LVL	SFP1: ETL	Template	SFM4-3	SFM4-3: LVL	SFM4-3: ETL	Template
Reads analyzed	27333	32513	47610	20829	23993	39069	36841	22536
Total errors	5.63	3.90	4.00	2.98	6.90	5.62	5.50	3.15
Insertions	0.26	0.16	0.15	0.13	0.35	0.30	0.27	0.18
Deletions	2.80	2.19	2.14	2.56	2.87	2.68	2.79	2.60
Substitutions	2.57	1.55	1.72	0.29	3.68	2.64	2.45	0.37
C→T	7.25	4.32	3.99	0.34	6.09	3.10	4.41	0.23
T→C	1.12	0.90	0.66	0.22	0.59	0.38	0.41	0.20
G→A	1.89	1.34	1.17	0.26	0.55	0.91	0.53	0.19
A→G	1.05	1.07	1.30	0.14	1.09	0.80	0.77	0.10
T→A	0.63	0.69	0.70	0.06	0.33	0.15	0.08	0.03
T→G	0.28	0.21	0.19	0.05	0.19	0.10	0.09	0.06
A→C	0.17	0.20	0.17	0.26	0.14	0.02	0.06	0.20
A→T	0.55	0.35	0.33	0.03	0.44	0.06	0.02	0.03
G→C	0.48	0.36	0.28	0.01	0.15	0.20	0.01	0.00
G→T	0.07	0.09	0.08	0.01	0.04	0.04	0.02	0.00
C→A	1.01	0.86	0.71	0.06	0.54	0.34	0.42	0.04
C→G	0.20	0.14	0.19	0.03	0.12	0.07	0.04	0.08



	No enzyme			SFM4-6			SFM4-6:LVL		
Termini (primer/temp.)	C/G	T/G	G/G	C/G	T/G	G/G	C/G	T/G	G/G
Fully extended primer				90%	73%	52%	85%	14%	3%

Figure S2. Matched and mismatched primer-termini extension by SFM4-6 and SFM4-6:LVL. Fully extended primer is measured by quantifying the band fluorescence intensity of the full-length band relative to the total fluorescence intensity in the lane. These data are identical to data shown in Figure 3, but presented with less contrast.

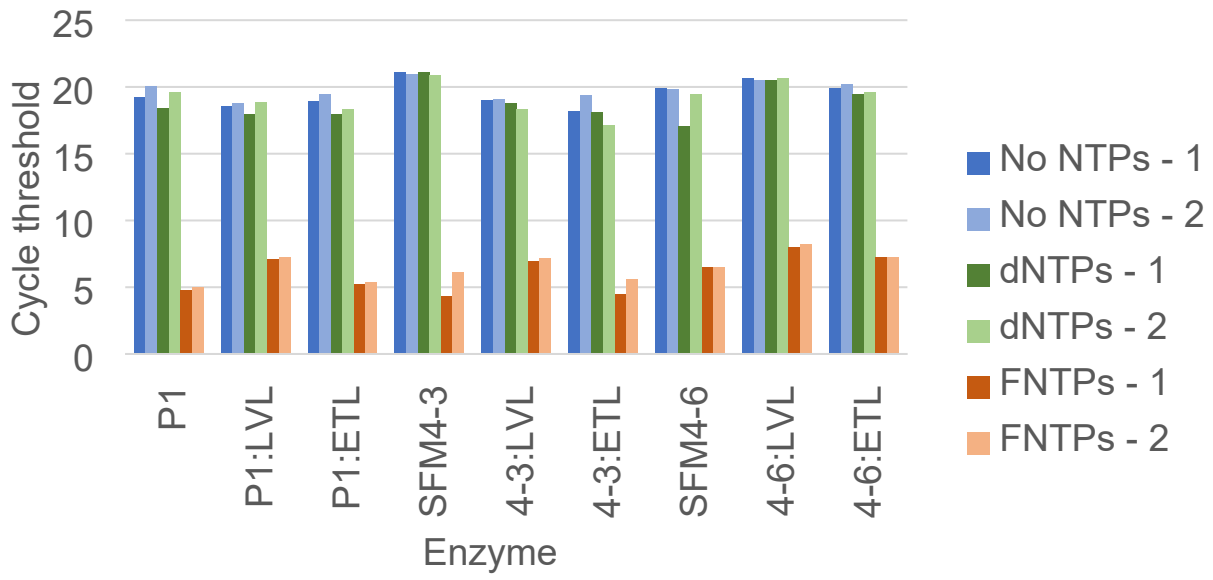


Figure S3. Efficiency of XNA synthesis as measured by cycle threshold values from quantitative reverse transcription and amplification reactions. The reactions vary based on the enzyme used and the NTPs added in the XNA synthesis reaction. The purified products of the XNA synthesis were subjected to a one-pot reverse transcription and amplification reaction with addition of SYBR green dye using Q5 DNA polymerase. Two independent synthesis reactions were performed for each enzyme.

Table S6. Total error rate and spectrum during PCR amplification of partially substituted DNA by SFM4-3, SFM4-3:LVL, and SFM4-3:ETL

Error Type	SFM4-3	SFM4-3:LVL	SFM4-3:ETL	Template
Reads analyzed	38014	82624	25553	9243
Total	61.46	23.84	24.23	1.60
Insertions	6.10	4.57	4.99	0.37
Deletions	21.60	12.90	13.61	1.01
Substitutions	33.76	6.37	5.64	0.22
C→T	98.49	9.73	15.42	0.14
T→C	21.95	2.11	3.52	0.15
G→A	11.46	3.52	2.41	0.08
A→G	16.24	3.27	4.96	0.14
T→A	12.93	0.91	0.98	0.02
T→G	1.16	0.40	0.28	0.02
A→C	0.85	0.25	0.31	0.03
A→T	4.62	0.75	1.26	0.03
G→C	0.80	0.25	0.13	0.06
G→T	2.45	0.92	0.50	0.04
C→A	15.96	6.37	7.03	0.12
C→G	1.82	0.34	0.10	0.05

IIa. Molecular Cloning

In this study, we created six new polymerases. For each, either A661E, I665T, F667L (abbreviated as “ETL”) or Q879L, H881L (abbreviated as “LVL”) were encoded in a commercially synthesized dsDNA (Integrated DNA Technologies). For ETL mutations, a unique dsDNA had to be created for each polymerase due to mutations in the parent enzymes in close proximity in primary sequence space. For LVL mutations, a single dsDNA could be used for the creation of all three mutant enzymes. The synthetic dsDNA was ligated to a vector containing the remainder of the polymerase gene (generated via PCR) in a Gibson assembly.

Sequence-verified plasmids were amplified via PCR using Phusion high-fidelity DNA polymerase with its accompanying 5x Phusion HF or GC buffer (New England Biosciences), 500μM dNTPs (Thermo Scientific), 5-100ng plasmid, and 125-500nM of the primers (Table S7). An Arktik thermal cycler (Thermo Scientific) was used to run the reaction under the following conditions: 98ºC for 30s, [98 ºC for 10s, 60-70ºC for 30s, 72ºC for 210s] x 25 cycles, 72 ºC 10 min, 4ºC hold. The resulting PCR products were analyzed via a 1% agarose (Research Products International) gel with GelRed (Biotium) for visualization using a UV gel box. Successful PCRs were purified using Qiagen PCR Purification Kit (Qiagen) according to the Qiagen protocol, and the concentration of DNA was measured on a NanoDrop 2000 Spectrophotometer (ThermoScientific).

Using HiFi Assembly Master Mix kit (New England Biolabs) a synthetic insert was ligated to the PCR-amplified vector according to the manufacturer’s protocol using 2:1 – 5:1 ratios of insert:vector and 1 hr of incubation at 50 ºC. The resulting plasmids were then transformed into DH5α cells (New England Biolabs) using the manufacturer’s protocol, and plated on LB/agar/carbenicillin (50μg/mL). Following overnight incubation at 37 ºC, individual colonies were picked and innoculated 20mL Luria Broth, Lennox (Research Products International) supplemented with carbenicillin (50μg/mL), and incubated at 37 ºC with shaking in a C24 shaking incubator (New Brunswick Scientific Class series).

The plasmids from the overnight cultures above were purified using the QIAprep Spin Plasmid Purification Kit (Qiagen) according to the Qiagen protocol. After successful isolation, concentrations were measured using a NanoDrop 2000 (Thermo Scientific). Plasmids were sequenced using Sanger Sequencing (Genewiz).

Table S7. Primers and templates used in the construction and sequence verification of novel XNA polymerases

Type	Primer code	Primer Sequence (5' to 3')
Vector Amplification	AL414 (for installing LVL mutations)	CACCACTGAGATCCGGCTGCTAACAAAGC
	AL415 (for installing LVL mutations)	CTCCAGCCTGGGGAAGAGCTTCAC
	AL431 (for installing ETL mutations)	GAGGCCCAGGCCTTCATTGAG
	AL432 (for installing ETL mutations)	CTCCGTGTGGATGTCCCGC
Insert	AL413 (LVL for all three mutants)	CTCATGAAGCTGGCTATGGTGAAGCTCTTCCCCAGGCT GGAGGAAATGGGGGCCAGGATGCTCCTTCTGGTCCTGGACGAGCT GGTCCTCGAGGCCCCAAAAGAGAGGGCGGAGGCCGTGGCACGGCT GGCCAAGGAGGTCATGGAGGGGGTGTATCCCCTGGCCGTGCCCCT GGAGGTGGAGGTGGGGATAGGGGAGGACTGGCTCTCCGCCAAGGA GGCGGCCGCACTCGAGCACCATCACCATCACCACTGAGATCCGGC TGCTAACAAAGC
	AL435 (ETL for SFM4-6)	AACCTGATCCGGGTCTTCCAGGAGGGGCGGGACATCCA CACGGAGACCGCCAGCTGGATGTTCGGCGTCCCCCGGGAGGCCGT GAACCCCCTGATGCGCCGGGAGGCCAAGACCACCAACCTAGGGGT CCTCTACGGCATGTCGGCCCACCGCCTCTCCCAGAAGCTAGCCAT CCCTTACGAGGAGGCCCAGGCCTTCATTGAG
	AL436 (ETL for SFM4-3)	AACCTGATCCGGGTCTTCCAGGAGGGGCGGGACATCCA CACGGAGACCGCCAGCTGGATGTTCGGCGTCCCCCGGGAGGCCGT GAACCCCATGATGCGCCGGGAGGCCAAGACCACCAACCTAGGGGT CCTCTACGGCATGTCGGCCCACCGCCTCTCCCAGAAGCTAGCCAT CCCTTACGAGGAGGCCCAGGCCTTCATTGAG
	AL438 (ETL for SFP1)	AACCTGATCCGGGTCTTCCAGGAGGGGCGGGACATCCA CACGGAGACCGCCAGCTGGATGTTCGGCGTCCCCCGGGAGGCCGT GGACCCCCTGATGCGCCGGGAGGCCAAGACCACCAACCTAGGGGT CCTCTACGGCATGTCGGCCCACCGCCTCTCCCAGGAGCTAGCCAT CCCTTACGAGGAGGCCCAGGCCTTCATTGAG

IIb. XNA polymerase expression and purification.

Expression, purification, and quantification of XNA polymerases were performed according to previous protocols.[1]

IIc. XNA polymerase enzyme assays.

For 100bp XNA accuracy, mismatch extension, and XNA PCR assays, see manuscript 'Experimental Section'.

150bp fidelity assay. The 150bp fidelity assay broadly followed the same steps as the 100bp fidelity assay, with a few notable exceptions. Most importantly, a different template (AL-K040) was used instead of the original AL-K021 (Table 4). Otherwise, the solutions were annealed and incubated in the same way as above. Once the resulting inserts were purified, slightly modified

conditions were used to PCR amplify them for sequencing, including 2µL template instead of just 1µL. The PCR program was significantly modified: 98ºC for 30s, 98 ºC for 5s, 50ºC for 15s, 72ºC for 60s, [98 ºC for 10s, 67ºC for 20s, 72ºC for 25s] x 17, 72 ºC 5 min, and 4ºC hold. Due to widely varying effectiveness as shown by running the resulting PCR products on 2% agarose (Research Products International) gels, a qPCR was performed using SYBR green dye (Bio-Rad Laboratories, Inc.), a C1000 Thermal Cycler (Bio-Rad), and the above PCR conditions, with the key difference that it ran for 30 cycles instead of 17 with quantification after each step.

Table S8. Sequences of oligonucleotides used in enzyme assays

Primer	Primer Sequence
AL-K017	/IR700/CTTTTAAGAACCGGACGAACGACACTCGTTTGCAGTAGCC
AL-K021	ATGCTGTCTACACGCAAGCTTACATTAAGACTCGCCATGTTACGATCTG CCTCAGGTAAGTACAGCCTTGAATCGTCACTGGCTACTGCATACGAGTG TC
AL-K038	/IR700/CTTTTAAGAACCGGACGAACGACACTCGTTTGCAGTAGCT
AL-K039	/IR700/CTTTTAAGAACCGGACGAACGACACTCGTTTGCAGTAGCG
AL-K040	TACGACAGATGTGCGTTGTTTTCTAGTAGTAGTACCTATCGTTCTGGCT GATGGTCCCGAAGGTTTCGTACATGTGGAAGCACTGAAGGGTAAACGGT GGGCCGAAGTTGCTCTCGGAACTTAGCAGTGACCGATGACGTTTGCTCA CAG
AL-K057-T75	CTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATAC GAGCCGGAAGCATAAAGTGTAAAGCC
K058-T75P-for	CTGTTTCCTGTGTGAAATTGTTAT
AL-K059-T57P-rev	GGCTTTACACTTTATGCTTCCG

Table S9. Oligonucleotides used in sample preparation for next-generation DNA

Primer	Primer Sequence
AL359_HTSamplicon_F	TTCTCGCCAAAGACCTGAGCGTTCTGGCCCTGAGGATGCTGTCT ACACGCA
AL360_HTSamplicon_R	TTTCCCGCAAGGAGCCCATGTGGGCCGATCTTCTGAGTGACGAT TCAAGGCT
AL443_HTS_barcode01	TTTCCCGCAAGGAGCCCATGTGGGCCGATCTTCTGGTCGTGATA GTGACGATTCAAGGCT
AL443_HTS_barcode02	TTTCCCGCAAGGAGCCCATGTGGGCCGATCTTCTGACCACTGTA GTGACGATTCAAGGCT
AL443_HTS_barcode03	TTTCCCGCAAGGAGCCCATGTGGGCCGATCTTCTGTGGATCTGA GTGACGATTCAAGGCT

AL443_HTS_barcode04	TTTCCCGCAAGGAGCCCATGTGGGCCGATCTTCTGCCGTTTGTA GTGACGATTCAAGGCT
AL443_HTS_barcode05	TTTCCCGCAAGGAGCCCATGTGGGCCGATCTTCTGTGCTGGGTA GTGACGATTCAAGGCT
AL443_HTS_barcode06	TTTCCCGCAAGGAGCCCATGTGGGCCGATCTTCTGGAGGGGTTA GTGACGATTCAAGGCT

References

[1] S. L. Rosenblum, A. G. Weiden, E. L. Lewis, A. L. Ogonowsky, H. E. Chia, S. E. Barrett, M. D. Liu, A. M. Leconte, *Chembiochem* **2017**, *18*, 816-823.