Electronic Supplementary Material (ESI) for Organic & Biomolecular Chemistry.

Supplemental Information

Systematic study of constraints imposed by modified nucleoside triphosphates with protein-like side chains for use in in vitro selection

Yajun Wang, Nicole Ng, Erkai Liu, Curtis H. Lam, David M. Perrin*

Department of Chemistry, University of British Columbia, 2036 Main Mall, Vancouver,

British Columbia, Canada, V6T 1Z1

*To whom correspondence should be addressed. Email: dperrin@chem.ubc.ca

Table of Contents

General materials.

dA^{im}TP and dU^{ga}TP were synthesized according to literatures[1,2]. dC^{aa}TP was obtained from TriLink Bio Technologies. Non-modified oligonucleotides were purchased from Integrated DNA Technologies (Coralville, IA). Modified oligonucleotides (with protection groups) were synthesized by Dr. Sabatino, at the department of Biochemistry, Seton Hall University. All oligonucleotides were purified by 10-20% denaturing PAGE (7 M urea). Ultrapure dNTPs were obtained from Fermentas. Vent exo$^-$ DNA polymerase and T4 polynucleotide kinase were obtained from New England Biolabs. Sequenase V2.0 and pyrophosphatase were purchased from Affymetrix. γ-^{32}P-ATP and α-^{32}P-dGTP were purchased from Perkin Elmer. Sephadex G25 resin was obtained from GE. UV spectrometry was performed on a Beckman Coulter DU800 spectrophotometer.

Oligonucleotides (ON, shown 5' to 3').

C^{m}C^{m}C^{m}TGCGGAGGGGCTGCCAGTA^{im}GTCTTGTGGCGTTCGTTTGTTCGCGGC GCGCT^{Inv} **ON 1**,

C^{m}C^{m}C^{m}TGCGGAGGGGCTGCCAGTA^{im}GU^{ga}C^{aa}TTGTGGCGTTCGTTTGTTCGCG GCGCGC^{Inv} **ON 2,**

C^{m}C^{m}C^{m}TGCGGAGGGGCTGCCAGTAGTCTTGTGGCGTTCGTTTGTTCGCGGCG CGCT^{Inv} **ON 3**,

CCACAAGACTACTGGCAG **ON 4**,

AGCGCGCCGCGAACAAACGAACGCCACAAGAC **ON 5**,
AGCGCGCCGCGAACAAACGAACGCCACAAG **ON 6**,

AGCGCGCCGCGAACAAACGAACGCCACAA **ON 7**,

GCGCTCGCGCGGCGTGCTGTACCATCCTGACGCTCCACACGTCCCAATCGAT ACCACTGTAACTGTTGGCGCAGGCCGACGC **ON 8**,

Biotin-T$_{40}$CCCGGGTTTTT^r(GCGUGCCCGUCUGUUGG)TTTTGCGTCGGCCTGC

GCCAACAG **ON 9**,

GCGCTCGCGCGGCGTGC **ON 10**,

GCGTCGGCCTGCGCCAACAG **ON 11**.

In oligonucleotide **1, 2** and **3**, "C^{m}" designates 2'-OMe ribocytidine, and "T^{Inv}" represents an inverted deoxythymidine at the 3'-end of the oligonucleotide. The three consecutive 2'-OMe ribocytidine at the 5'-end and the inverted deoxythymidine at the 3'-end were constructed for enhanced stability for cell based studies in the future. "r" designates a stretch of RNA bases in **ON 9**.

Buffers.

1 pH variant phosphate buffer for thermal melting study: 10 mM sodium phosphate (pH 6, and 8.5), 150 mM KCl.

2 1X thermopol buffer: 20 mM Tris-HCl (pH 8.8), 10 mM $(NH_4)_2SO_4$, 10 mM KCl, 2 mM $MgSO_4$, 0.1% Triton® X-100.

3 1X reaction buffer for Sequenase V2.0: 40 mM Tris-HCl (pH 8.0), 50 mM NaCl, 20 mM $MgCl_2$.

4 Wash buffer: 50 mM Tris-HCl (pH 7.5), 200 mM NaCl, 1 mM EDTA.

5 Neutralization buffer: 25 mM sodium cacodylate (pH 6.0), 1 mM EDTA.

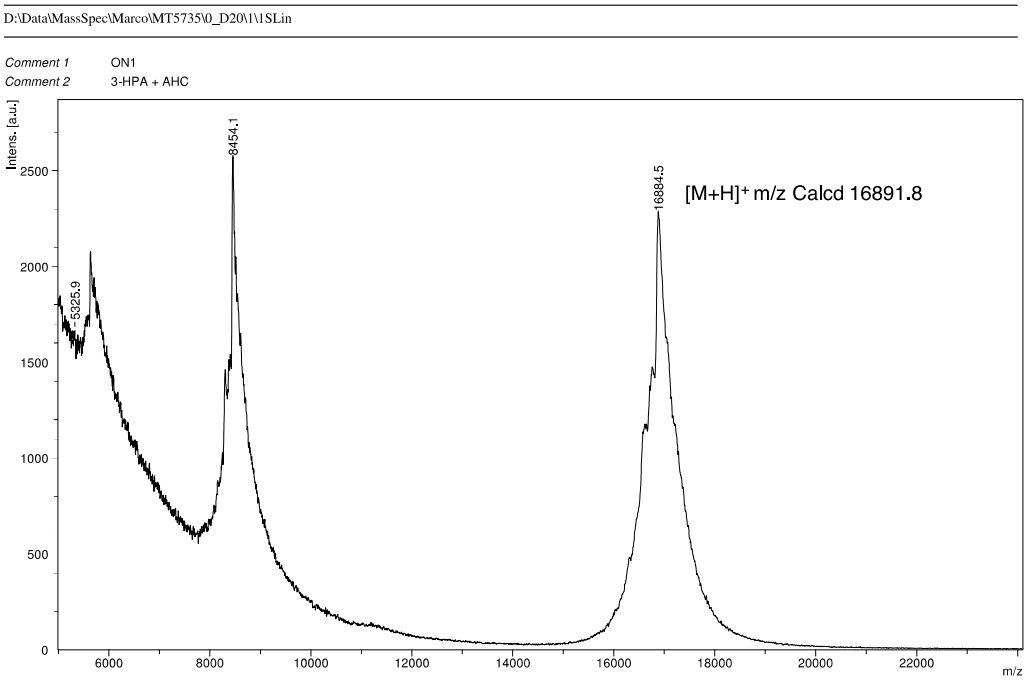
6 Cleavage buffer: 50 mM sodium cacodylate (pH 7.4), 200 mM NaCl, 1 mM EDTA.

7 Gel elution buffer: 10 mM Tris-HCl, (pH 8.0), 1% $LiClO_4$.

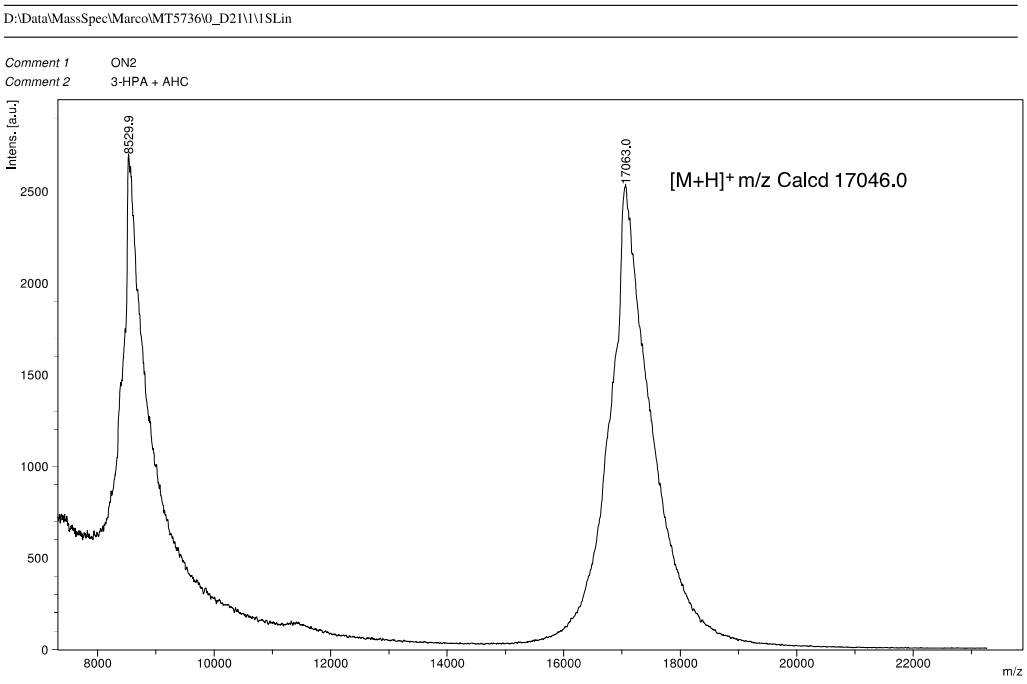
De-protection and MALDI-TOF characterization of oligonucleotides.

ON 1 and **ON 3** were cleaved off the resin and fully de-protected by direct incubation in concentrated ammonium hydroxide (28-30%, wt) at 55°C overnight after receiving.[3] **ON 2** was firstly treated with piperidine/H_2O (1:1, v/v) at room temperature overnight to remove the β-cyanoethoxycarbonyl groups, meanwhile cleave the oligos off the resin, and then the supernatant was recovered and dried by speed-vac prior to the same concentrated ammonium hydroxide (28-30%, wt) treatment as **ON 1** and **ON 3**.[4,5] Following lyophilization, all the crude oligonucleotides were dissolved in 10 mM Tris-HCl (pH 8.0) buffer, and then subjected to 10% denaturing PAGE (7 M urea) purification. ON 1, 2 and 3 were analyzed and confirmed their identity by MALDI-TOF mass spectrometry. Dada were acquired on a Bruker autoflex MALDI-TOF spectrometer with matrix 3-hydroxypicolinic acid (3-HPA) in positive ion mode in the Mass Spectrometry Laboratory at the department of Chemistry, University of British Columbia. Their respective mass spectrums are shown in **Figure S1**.

A


D:\Data\MassSpec\Marco\MT5735\0_D20\1\1SLin
Comment 1 ON1
Comment 2 3-HPA + AHC
Intens. [a.u.]
8454.1
16884.5
5325.9
[M+H]+ m/z Calcd 16891.8
m/z

B


D:\Data\MassSpec\Marco\MT5736\0_D21\1\1SLin
Comment 1 ON2
Comment 2 3-HPA + AHC
Intens. [a.u.]
8529.9
17063.0
[M+H]+ m/z Calcd 17046.0
m/z

C

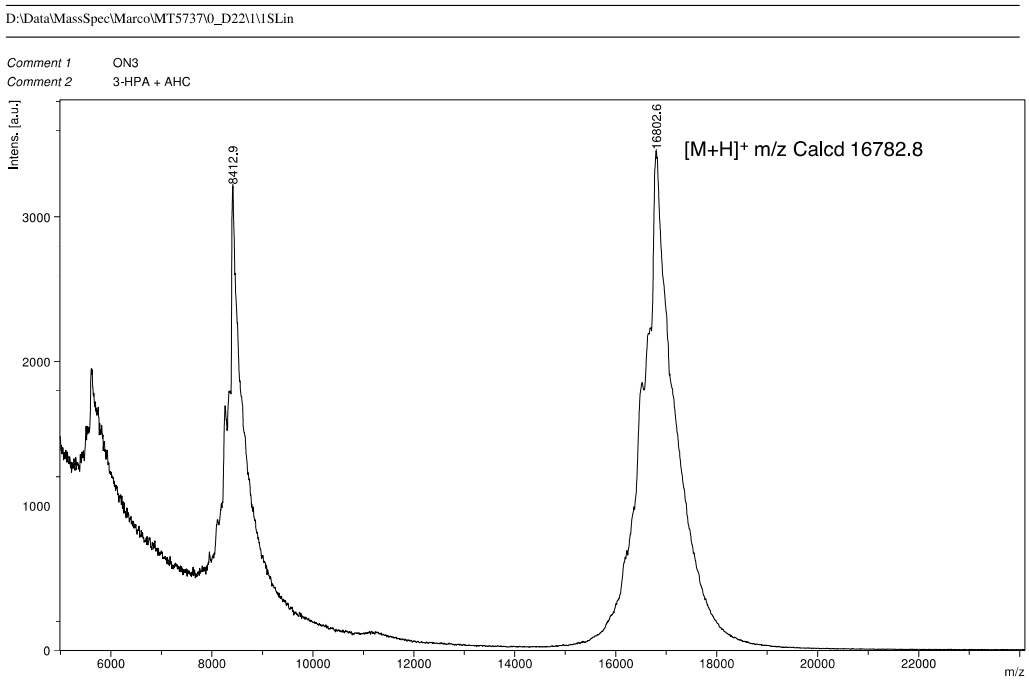


Figure S1 MALDI mass spectrometry analysis of ON1, 2 and 3. The mass spectrums of ON 1, 2 and 3 are shown in (A), (B) and (C), respectively.

Thermal melting temperature.

Thermal melting temperatures were measured on Beckman Coulter DU800 spectrophotometer using its "Experimental T_m Analysis" program. Oligonucleotides **1, 2 or 3** at a final concentration of 1 μM were heated up to 95°C with **4** at one-to-one ratio in pH variant buffers, respectively, and then allowed to cool down to room temperature slowly to form a duplex, prior to the thermal melting temperature measurement. Samples were monitored for absorbance at 260 nm from 20–85 °C at a rate of 1°C/min with data points at every 1°C. To minimize the buffer concentration change caused by evaporation during the process of thermal melting, mineral oil was applied to the top of the solutions inside the Teflon-stopped 1 cm path length quartz microcells. Three separate experiments of forward and reverse melting temperatures were obtained for each buffer condition (pH 6 and 8.5, **Figure S2**).

A

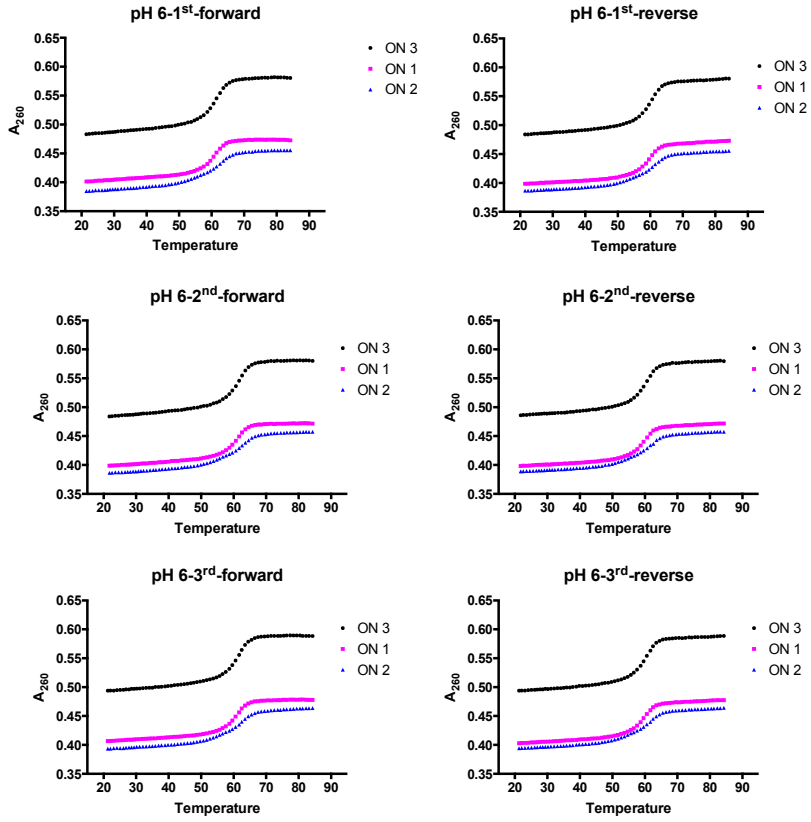
5'-C^{m}C^{m}C^{m}TGCGGAGGGGCTGCCAGTA GTCTTGTGGCGTTCGTTTGTTCGCGGCGCGCT^{Inv}-3' **ON 3**

5'-C^{m}C^{m}C^{m}TGCGGAGGGGCTGCCAGTA^{im}GU^{ga}C^{aa}TTGTGGCGTTCGTTTGTTCGCGGCGCGCT^{Inv}-3' **ON 2**

5'-C^{m}C^{m}C^{m}TGCGGAGGGGCTGCCAGTA^{im}GTCTTGTGGCGTTCGTTTGTTCGCGGCGCGCT^{Inv}-3' **ON 1**

3'-GACGGTCAT CAGAACACC-5' **ON 4**

B



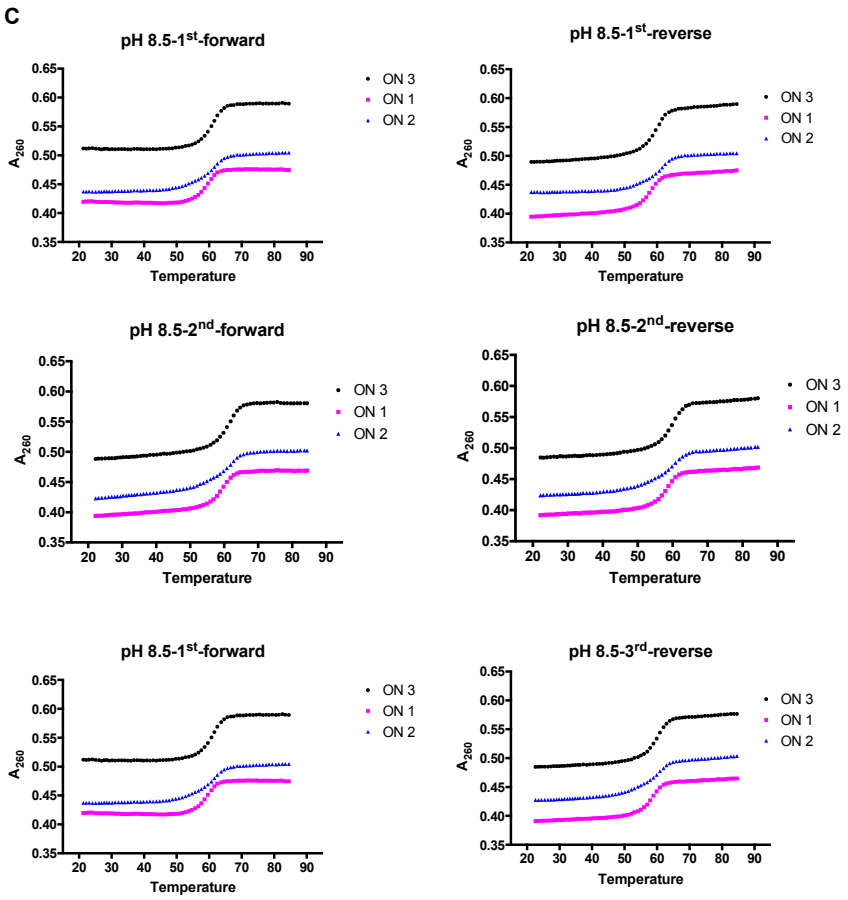


Figure S2 Thermal melting study. (A) Oligonucleotide sequences used for thermal denaturation study. ON 1, ON 2, and ON 3 in duplex with a complementary strand ON 4 for thermal melting studies. (B) Thermal melting curves of three separate experiments of forward and reverse melting processes at pH 6. (C) Thermal melting curves of three separate experiments of forward and reverse melting processes at pH 8.5.

Full-strand synthesis continuing beyond multiple modified base pairs.

The primer elongation continuing beyond multiple modified base pairs assay was carried out under almost the same nonforcing conditions as single-nucleotide insertion survey. 2 µl of 20 µM 5'-^{32}P-labeled 29 nt primer (**ON 7**) was annealed to an equal amount of modified template (**ON 2**) and unmodified control template (**ON 3**), respectively, followed by incubation with 0.5 µl (1 unit) of Vent exo$^-$ DNA polymerase at 72°C for 5 min to form enzyme•duplex complex (**Figure S3A**). Then, 10 µl of nucleotide mixture containing 10

μM of each dATP, dCTP, dGTP, and TTP was added into 10 μl of pre-formed enzyme•duplex complex. The primer extension assay was immediately initiated by incubation at 72°C. 2 μl of reaction was taken into 10 μl of gel loading buffer at different time points, including 5 min, 10 min, 30 min, 120 min, and 180 min. Samples were separated by denaturing 20% PAGE (7 M urea) and visualized by PhosphorImager (**Figure S3B**).

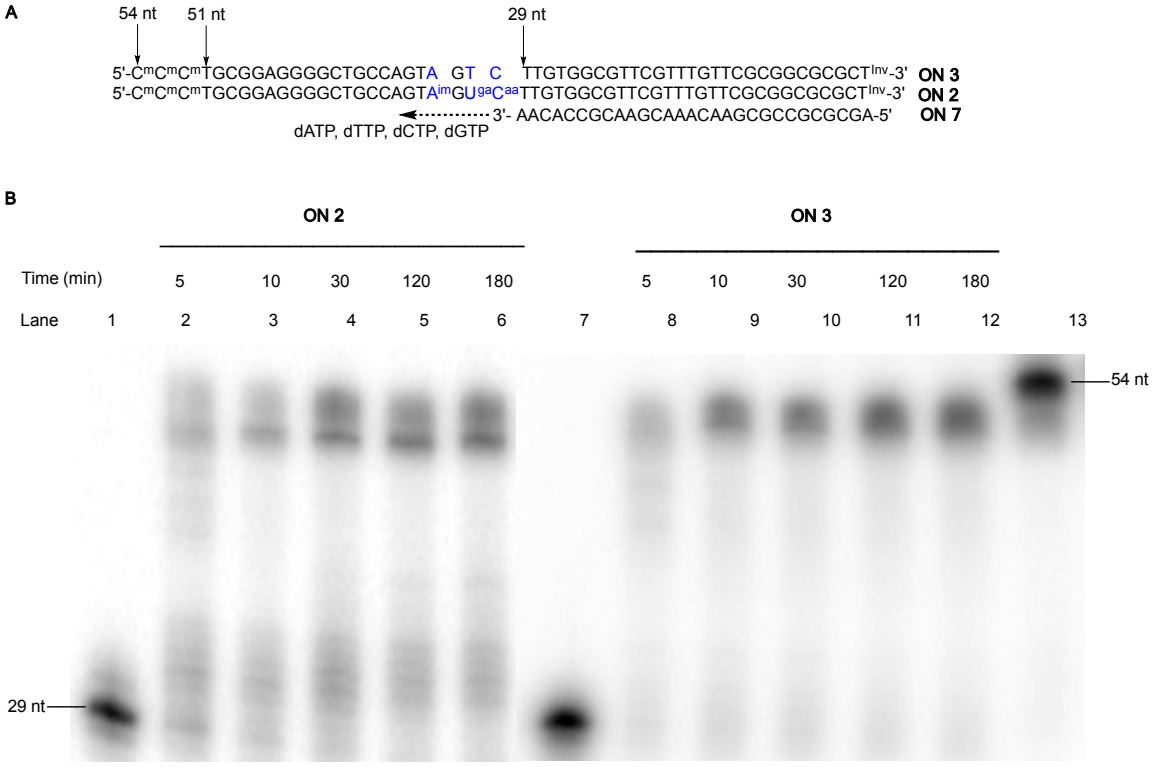


Figure S3 Strand synthesis continuing beyond multiple modified base pairs by Vent exo$^-$ DNA polymerase. (A) Template-primer duplexes used to evaluate strand synthesis progress. Template **ON 2** contains all the three modifications immediately downstream of the primer, and modification free **ON 3** was used as reference. (B) Strand synthesis results in the presence of all four natural dNTPs through 180 min reaction time. Lane 1 and lane 7, 5'-^{32}P labeled primer, 29 nt (**ON 7**); lane 2 to lane 6, strand synthesis progress on template **ON 2**; lane 13: 5'-^{32}P labeled **ON 2** as size-marker for full-strand synthesis. The observed multiple bands on the top part of gel pictures representing full-length products were attributed to the three consecutive 2'-OMe-rC at the very 5'-end of the templates (**ON 2** and **ON 3**). All the data were taken using 1 μM template-primer duplex and dNTP mixtures containing 10 μM of each dATP, TTP, dCTP, and dGTP at 72°C.

Standing-start single-nucleotide insertion survey.

The primers (**ON 5, ON 6,** and **ON 7**) for standing-start single-nucleotide insertion were 5'-labeled with γ-^{32}P-ATP by T4 polynucleotide kinase and purified with 15% denaturing PAGE gel (7 M urea) before using. The insertion reaction was carried out in 20 μl of 1X thermopol buffer (Buffer **2**) under nonforcing condition: the duplex concentration was 1 μM, the tested nucleotide concentration was 10 μM, and the Vent exo$^-$ DNA polymerase used for each 20 μl reaction was 1 unit, which is sufficient amount for incorporating 10 nmole of dNTP in 30 min at 75°C. Firstly, 9.5 μl of annealing mixture containing 1 μl of each of 20 μM labeled primer and template, 2 μl of 10X thermopol buffer, 0.2 μl of 100 mM MgSO4 (supplemented Mg^{2+} for further stabilization of enzyme•duplex complex), as well as 5.3 μl of H_2O was heated to above 95°C for 5 min, then cooled down to room temperature slowly to form template-primer duplex. At this point, 0.5 μl (1 unit) of Vent exo$^-$ DNA polymerase was added to the annealed duplex, and incubate at 72°C for 5 min to facilitate the formation of enzyme•duplex complex. Then, 10 μl of 20 μM of tested dNTP was added into the enzyme•duplex complex, and the insertion assay was readily initiated by 72°C incubation. The reaction was stopped by adding into equal volume of gel loading buffer (20 μl of formamide stop solution containing 25 mM EDTA, 0.01% bromophenol blue and 0.01% xylene cyanole) after designed reaction time. Reaction products were separated by denaturing 20% PAGE, then visualized by PhosphorImager.

To further test if dA^{im} forms mispairs with dATP, dCTP, and dGTP after long-time incubation, 20 μl of insertion reaction with each dNTP was set up as described above, respectively. 2 μl of reaction was sampled into 10 μl of gel loading buffer at different time points including 2 min, 5 min, 30 min and 180 min, through 180 min reaction time at 72°C. Autoradiograph (**Figure S4**) showed that no dA^{im}-dC or dA^{im}-dG mispairs formed at all even after 180 min reaction time. Less than 5% of the 32 nt primer was elongated for one base opposite the template dA^{im} in the presence of dATP, which is highly likely attributed to the "A-rule". [6]

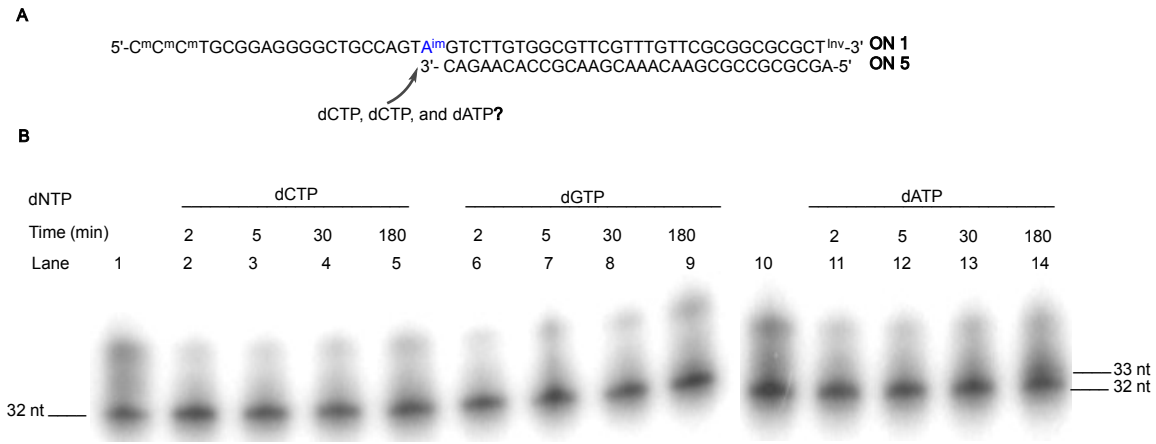


Figure S4 Standing-start single-nucleotide insertion opposite dA^{im} by Vent exo^{-} DNA polymerase. (A) Template-primer duplex used for single-nucleotide insertion assay: the dA^{im} (in blue) in template (**ON 1**) is just downstream of the 3'-end of a 32 nt primer (**ON 5**). (B) Single-nucleotide insertion result with dCTP, dGTP, and dATP at different time points through 180 min reaction time, respectively. Lane 1 and lane 10, 5'-^{32}P labeled primer, 32 nt (**ON 5**); lane 2 to lane 5, single-nucleotide insertion progress with dCTP; lane 6 to lane 9, single-nucleotide insertion progress with dGTP; lane 11 to lane 14, single-nucleotide insertion progress with dATP, less than 5% of 33 nt product formed as shown in lane 14, which represents dA^{im}-dA mispair.

Modified Dz7-38-32 transcription, self-cleavage, and gel purification of self-cleavage product.

15 pmole of 5'-biotinylated chimeric DNA/RNA primer (**ON 9**) serving as substrate for all-RNA cleavage in *cis* was annealed to equal amount of **ON 8**, the synthetic template for transcribing Dz7-38-32, before the initiation of template directed transcription of modified strand (as shown in Figure S5A). Then the transcription reaction was performed using 13 unit of Sequenase V2.0 at 32°C for 4h in 40 µl of 1X Sequenase buffer condition (40 mM Tris-HCl (pH 8.0), 50 mM NaCl, 20 mM $MgCl_2$) containing 5 mM DTT, 0.1 unit of pyrophosphatase, 50 µM dA^{im}TP, 25 µM of dC^{aa}TP, 10□µM of each dU^{ga}TP and dGTP, and ~15 µCi of α-^{32}P-dGTP. Mineral oil was used to cover the top to reduce the effect caused by evaporation and condensation. At the end of reaction time, EDTA was added to a final concentration of 25 mM to quench the polymerization reaction. To prepare the

"dummy" control, exactly the same procedures were followed as the transcription of Dz7-38-32 above. Instead of including dNTP mixtures (dA^{im}TP, dC^{aa}TP, dU^{ga}TP, and dGTP), equal volume of dH_2O was used to supplement the 40 μl of transcription reaction.

The resulting reaction mixture from transcription was incubated with 50 μl of buffer **4** (wash buffer) pre-washed streptavidin beads for 15 min at room temperature to immobilize the DNA duplex on beads. Following two times of wash with 100 μl of wash buffer **4** after beads incubation, the template strand was stripped away by 5 times of quick wash (no longer than 30 seconds per wash) with 100 μl of 0.1 M NaOH containing 1 mM EDTA. The resulting modified strand sticking to streptavidin beads was immediately neutralized by using 200 μl of buffer **5** followed by a final 100 μl of water wash. The modified DNA on beads was then allowed to fold and cleave in 100 μl of cleavage buffer **6** for 30 min. Then the cleavage reaction was magnetized, and reaction supernatant containing self-cleaved Dz7-38-32 was recovered from beads, which was then precipitated using 1% $LiClO_4$ in acetone and washed with ethanol. The modified DNA was re-suspended in formamide loading buffer containing 25 mM EDTA, 0.01% bromophenol blue and 0.01% xylene cyanole, and resolved by 7% denaturing PAGE (7 M urea). Once again, the same treatments were applied to both the DNAzyme and the "dummy" control. The complete scheme of DNAzyme transcription, self-cleavage, and subsequent amplification of gel purified self-cleavage product is shown in **Figure S5B**. Gel pieces corresponding to the desired self-cleavage products (red rectangulars in lane 2 and lane 5, **Figure S5C**) were sliced off, eluted by using gel elution buffer **7**.

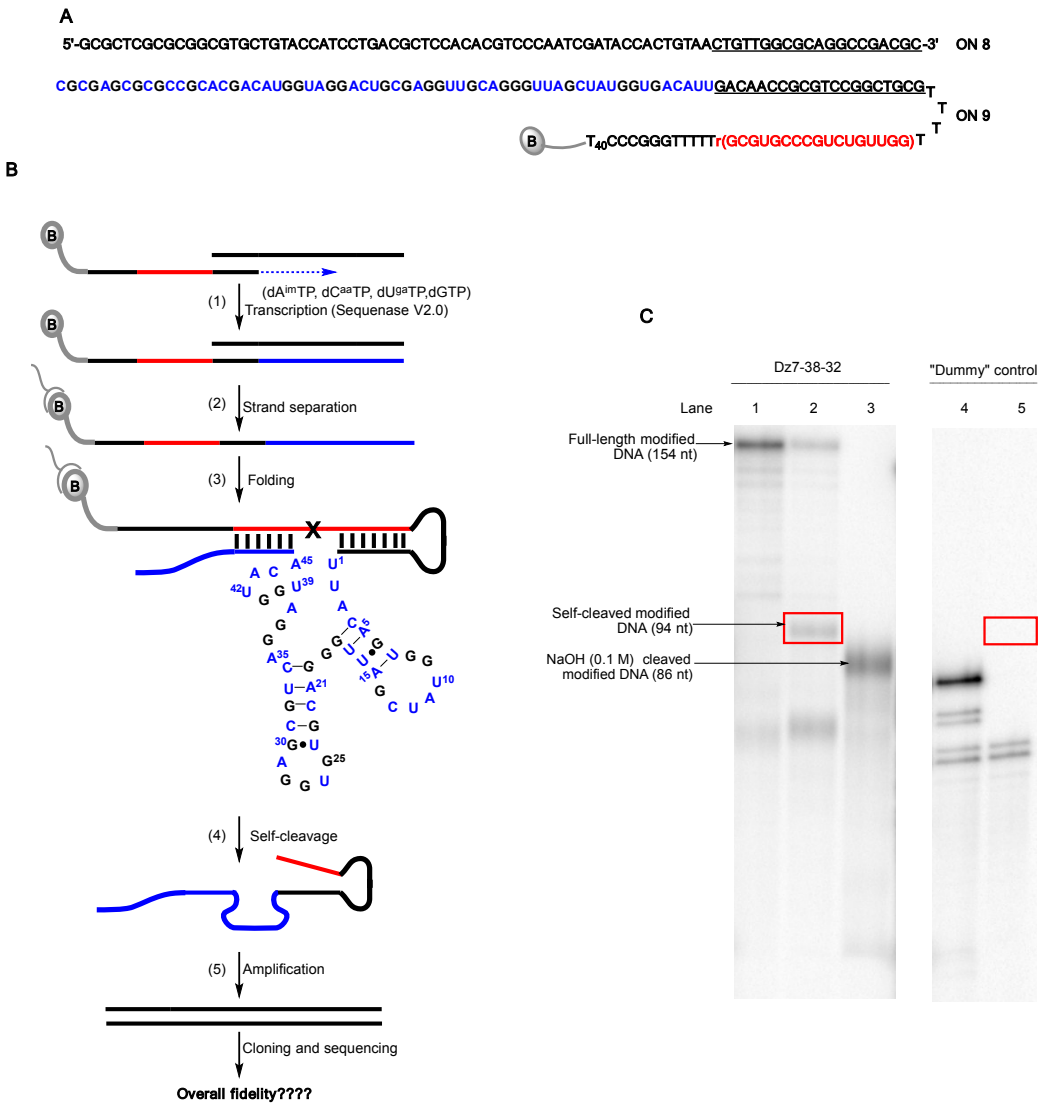


Figure S5 Transcription, self-cleavage, and purification of Dz7-38-32 and "dummy" control. (A) Transcription of self-cleaving Dz7-38-32 by template (ON 8) directed primer (ON 9) extension reaction. Template-primer hybridization sequence regions are underlined; all the nucleotides in blue are modified dA^{im} (A), dC^{aa} (C) and dU^{ga} (U). (B) Schematic representations of a replication cycle of Dz7-38-32. (1) Synthetic template directed transcription of modified DNAzyme, (2) catalytic strand separation from unmodified template, (3) and (4) catalytic folding and self-cleavage, and (5) amplification of self-cleavage product by *Taq* or Vent exo$^-$ DNA polymerase. (C) PAGE gel purification of self-cleavage product. Lane 1: full-length Dz7-38-32; lane 2: precipitated self-cleavage product; lane 3: 0.1 M NaOH treated full-length product as size marker for self-cleavage product; lane 4: 0.1 M NaOH treated "dummy" sample; lane 5: "self-cleaved" "dummy" sample. Self-cleavage products in red rectangulars in lane 2 (Dz7-38-32) and lane 5 ("dummy" control) were sliced off for gel elution and subsequent PCR reaction.

Amplification of self-cleaved modified DNA.

The gel eluted DNAs (self-cleaved DNAzyme and "dummy" control), which were re-suspended into 20 µl of dH_2O, were used as templates for PCR amplification using Vent exo$^-$ followed by *Taq* DNA polymerase or directly by *Taq* DNA polymerase, respectively. The amplification reactions were performed in 200 µl of 1X Thermopol buffer (20 mM Tris-HCl (pH 8.8), 10 mM $(NH_4)_2SO_4$, 10 mM KCl, 2 mM $MgSO_4$, 0.1% Triton® X-100), supplemented with 1 mM of $MgSO_4$ 7 µM of primers (**ON 10** and **ON 11**), 0.3 mM of each of the four natural dNTPs, and 0.1 unit/µl of Vent exo$^-$ or 0.25 unit/µl of *Taq* DNA polymerase. The reactions were cycled as follows for 25 cycles: 15 seconds at 95°C, 15 seconds at 58°C and 40 seconds at 75°C. To generate PCR products with 3'-A overhangs required by subsequent TOPO cloning, a final extension step of 30 min at 75°C was added to the *Taq* amplification reaction. To monitor DNA amplification, which was used to assay the self-cleaved modified DNA for possible unmodified template DNA contamination, aliquots (1 µl) were removed at cycle 0, 5, 10, 15 and 25 and stored in fridge. Once all the cycles were taken, the samples were analyzed on a 2% agarose gel stained with ethidium bromide (**Figure S6**).

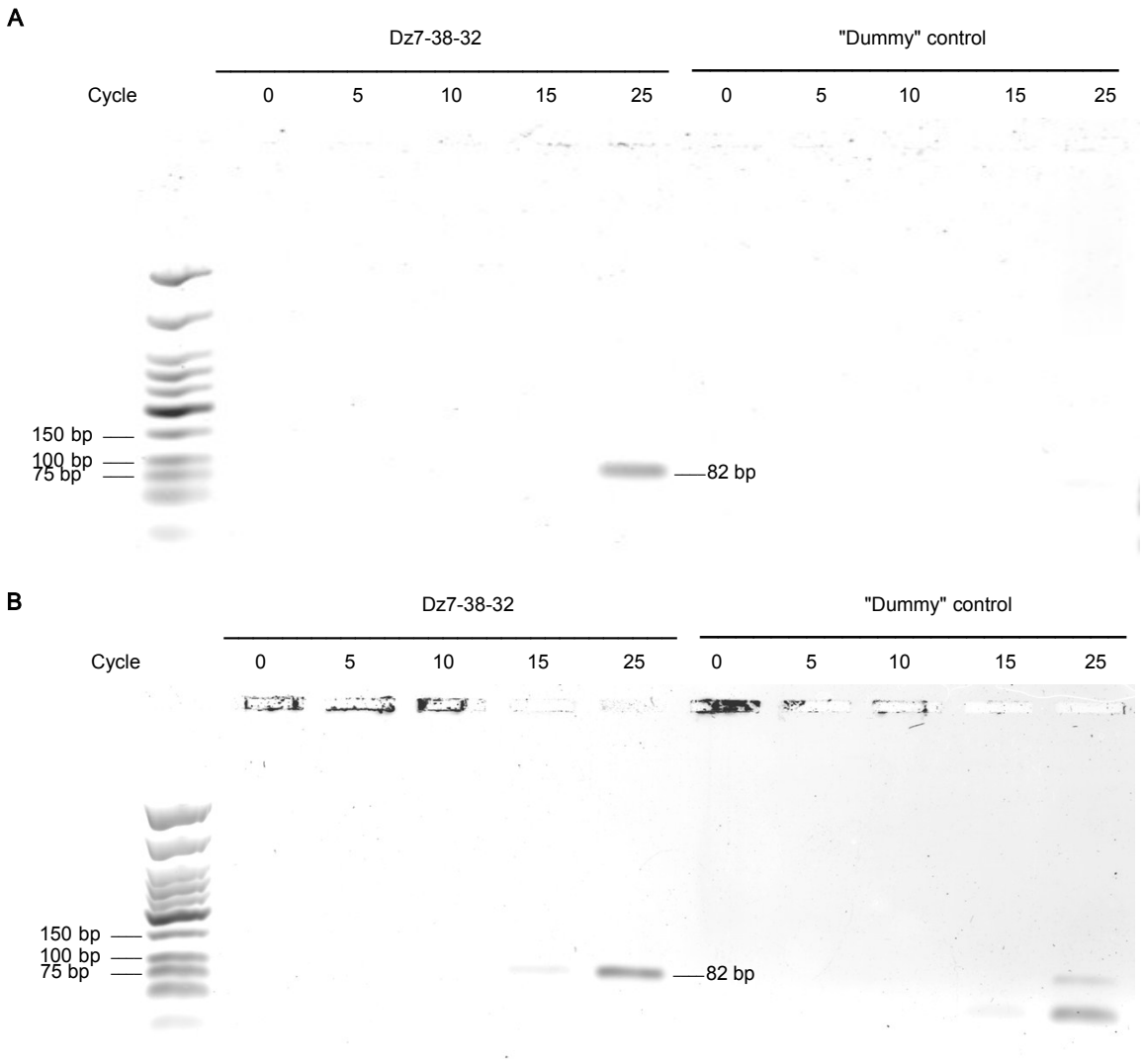


Figure S6 PCR assay for unmodified template DNA contaminants. PCR analysis confirmed that correct DNA amplicons (82 bp) were only observed for amplification of self-cleaved Dz7-38-32, but not for the "dummy" control, which ruled out the small but nonzero possibility that the amplicons were resulting from amplification of co-migrating unmodified DNAs. (A) Amplification of self-cleaved Dz7-38-32 using Vent exo$^-$ DNA polymerase. (B) Amplification of self-cleaved Dz7-38-32 using *Taq* DNA polymerase.

Fidelity analysis by cDNA sequencing.

PCR amplified DNA was purified using 2% agarose gel. Gel slice containing the amplicons were purified using GeneJet Gel Extraction Kit first according to the manufacture's instruction before they were TOPO cloned into the pCR2.1-TOPO vector. The vector was then used to transform *E.coli* DH5α following a normal chemical transformation protocol. White transformant colonies were picked by Blue-White screening on LB Agar plate containing 100 mg/L ampicillin for inoculation. Plasmids were prepared by using PureLink Quick Plasmid Miniprep Kit (Invitrogen), and sequenced by the Nucleic Acid Protein Service Unit of UBC (NAPS-UBC) using the M13R sequencing primer. The alignment of the individual sequences was done on the MEGA 6.06 platform (**Figure S7**).

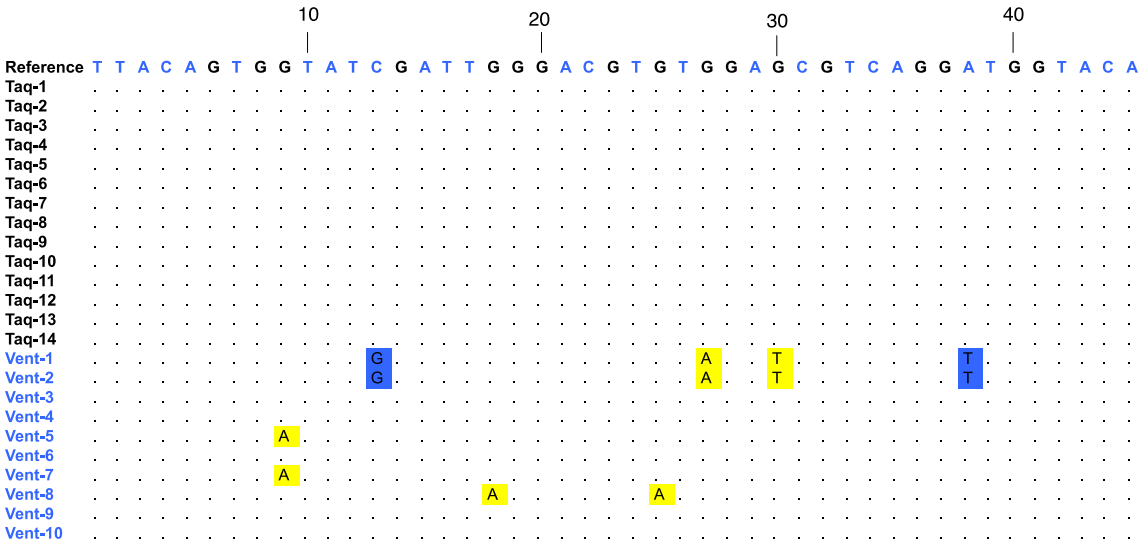


Figure S7 cDNA sequence alignment used to evaluate modified DNA replication fidelity of a complete replication cycle. All individual sequences were aligned in reference to the original catalytic sequence. 10 individual sequences derived from amplicons amplified by Vent exo$^-$, denoted Vent-1 to Vent-10, revealed 12 mutations (highlighted in blue and yellow) through a total of 450 positions in the modified sequence region, only 4 of these mutations were caused by modified nucleosides (highlighted in blue). All 14 individual sequences derived from amplicons amplified by *Taq* DNA polymerase, denoted Taq-1 to Taq-14, revealed an overall fidelity of 100% with no mutations observed through a total of 630 modified positions.

References

(1) Lam, C.; Hipolito, C.; Perrin, D. M. *Eur. J. Org. Chem.* **2008**, *2008*, 4915–4923.

(2) Hollenstein, M.; Hipolito, C. J.; Lam, C. H.; Perrin, D. M. *Nucleic Acids Research* **2009**, *37*, 1638–1649.

(3) Lermer, L.; Hobbs, J.; Perrin, D. M. *Nucleosides and Nucleotides* **2002**, *21*, 651–664.

(4) Prakash, T. P.; Püschl, A.; Manoharan, M. *Nucleosides and Nucleotides* **2007**, *26*, 149–159.

(5) Lam, C. H.; Perrin, D. M. *Bioorg. Med. Chem. Lett.* **2010**, *20*, 5119–5122.

(6) Goodman, M. F.; Creighton, S.; Bloom, L. B.; Petruska, J. *Crit. Rev. Biochem. Mol. Biol.* **1993**, *28*, 83–126.