Electronic Supplementary Information

Formation of Instantaneous Nick for Highly Efficient Adenylation of Oligonucleotides by Ligase without Subsequent Jointing

Kunling Hu,[a] Wenhua Sun,[a] Hui Chen,[a] Jian Luo,[a] Ziting Song,[a] Ran An,[*,a,b] Makoto Komiyama[c] and Xingguo Liang[*,a,b]

[a] State Key Laboratory of Marine Food Processing & Safety Control, College of Food Science and Engineering, Ocean University of China, Qingdao, No. 1299 Sansha Road, P. R. China.

[b] Laboratory for Marine Drugs and Bioproducts, Pilot National Laboratory for Marine Science and Technology (Qingdao), Qingdao, No. 1 Wenhai Road, P. R. China.

[c] Research Center for Advanced Science and Technology (RCAST), The University of Tokyo, 4-6-1 Komaba, Meguro, Tokyo 153-8904, Japan.

[*]Corresponding Authors: liangxg@ouc.edu.cn, ar@ouc.edu.cn.

Table of Contents:

Materials and Methods

Materials. All the DNA/RNA oligonucleotides (Table S1) were purchased from Shanghai Sangon Biotech (Shanghai, China). T4 polynucleotide kinase was from Thermo Scientific (Pittsburgh, PA, USA). The fluorescent dye of GelRed was from Invitrogen (Pittsburgh, PA, USA). Other chemicals are from Sinopharm Chemical Reagent Co., Ltd. (Shanghai, China). The T4 DNA ligase utilized in this study was sourced from Thermo Scientific (5 Weiss U/μl). One Weiss unit is equivalent to approximately 200 cohesive end ligation units (CEU, 1 Weiss Unit=200 CEU). One CEU is defined as the amount of enzyme required to give 50% ligation of HindIII fragments of lambda DNA in 30 min at 16°C.

Phosphorylation reactions by T4 polynucleotide kinase. Before the reaction of adenylation, the 5′-end structure of oligonucleotides should be phosphate, which could be modified in advance during synthesis or use T4 polynucleotide kinase (Thermo Scientific; Pittsburgh, PA, USA) to convert hydroxyl to phosphate. For a typical phosphorylation reaction, 40 μM oligonucleotides, 0.5 U/μL T4 polynucleotide kinase, 1× T4 polynucleotide kinase buffer (50 mM Tris-HCl (pH7.5@25°C), 10 mM MgCl$_2$, 5.0 mM DTT and 0.10 mM spermidine), and 0.24 mM ATP. For the high concentration preparation reactions, 100 μM oligonucleotides, 1.0 U/μL T4 polynucleotide kinase, 1× T4 polynucleotide kinase buffer, and 1.2 mM ATP. The reaction was carried out at 37°C for 12 h. T4 polynucleotide kinase was deactivated at 75°C for 10 min.

Previous adenylation reaction replied on acceptor and template strand with a C-T mismatch. The reaction conditions followed the design of Patel et al. in 2008. The typical reaction was performed with a final volume of 20 μL which include 4.0 μM ssDNA, 5.0 μM template, 6.0 μM acceptor, 1× T4 Dnl buffer (40 mM Tris-HCl (pH7.8), 10 mM MgCl$_2$, 10 mM DTT). The process of annealing should be carried out by heating at 95°C for 3.0 min and gradient cooling to 25°C for 5.0 min(0.1^{o}C/s). The acceptor and template sequence forms a mismatch (C-T). Then, 6.0 mM ATP and 0.25 U/μL T4 Dnl were added and incubated at 25°C. T4 Dnl was deactivated at 75°C for 10 min, followed by electrophoresis analysis (12% denaturing PAGE with 8.0 M urea and 25% formamide) and imaged with Image Lab.

Previous adenylation reaction replied on an aid strand with a mini-hairpin and a C-C mismatch. The reaction conditions followed the design of Lou et al. in 2022. The typical reaction was performed with a final volume of 20 μL which include 4.0 μM ssDNA, 6.0 μM aid-DNA, 1× T4 Dnl buffer (40 mM Tris-HCl (pH7.8), 10 mM MgCl$_2$, 10 mM DTT, 0.5 mM ATP) and 0.25 U/μL T4 Dnl. The process of annealing should be carried out before adding ligase. Then the samples were incubated at 25°C. Products were separated by 12% denaturing PAGE (8.0 M urea and 25% formamide) and imaged with Image Lab. T4 Dnl was deactivated at 75°C for 10 min, followed by electrophoresis analysis (12% denaturing PAGE with 8.0 M urea and 25% formamide) and imaged with Image Lab.

Dynamic nick adenylation reaction. The optimized reaction was performed with a final volume of 20 μL including 4.0 μM ssDNA, 6.0 μM aid-DNA, 1× T4 Dnl buffer (40 mM Tris-HCl (pH7.8), 10 mM

MgCl$_2$, 10 mM DTT), 2.0 mM ATP and 0.25 U/μL T4 Dnl. The process of annealing should be carried out before adding ligase. Then the samples were incubated at 37°C. To verify the effect of different dynamic structures on the efficiency of adenylation, samples with different hairpin structure were incubated at 37°C for 0.5–24 h. To find the optimum conditions for adenylation, different reaction temperature (16°C, 25°C and 37°C) and concentration of ATP (0.01–2.0 mM) were carried out. For high-concentration preparation of AppDNA, the reaction included 100 μM ssDNA, 120 μM aid-DNA, 1.0 U/μL T4 Dnl, 1× T4 Dnl buffer (2.0 mM ATP), 37°C for a certain time. All samples were analyzed by 12% dPAGE (8.0 M urea and 25% formamide) and imaged with Image Lab.

Ligation of adaptors (pSMRT-3T or AppSMRT-3T) and pT56-3A. When pSMRT-3T (phosphate at 5') played as the adaptor, the sample included 0.5 μM pT56-3A, 2.5 μM pSMRT-3T, 0.25 U/μL T4 Dnl, 1× T4 Dnl buffer (0.5 mM ATP), and incubated at 25°C for 0.5, 2.0, 12 h. When AppSMRT-3T (adenyl pyrophosphoryl moiety at 5') played as the adaptor, the sample included 0.5 μM pT56-3A, 2.5 μM AppSMRT-3T, 0.25 U/μL T4 Dnl, 1× T4 Dnl buffer (no ATP), and incubated at 25°C for 0.5, 2.0, 12 h. Notes, T4 Dnl may be adenylated ligase, therefore deadenylation for enzyme should be carried out by dialysis.

Electrophoresis analysis of products and evaluation of the adenylated yield. After the adenylation reaction, the products were diluted with the loading buffer (10 mM Tris-HCl (pH7.6) 0.03% bromophenol blue, 0.03% xylene cyanol FF, 60% glycerol 60 mM EDTA) and subjected to 12% denatured polyacrylamide gel containing 8.0 M urea and 25% formamide, stained by GelRed.

The substrates and products were quantified by Gel Doc XR+ Imaging System (Bio-Rad) and Image lab software v 3.0 (Bio-Rad). The yield of AppDNA/RNA or ligation were calculated as the following equations:

$$Yield\ of\ AppDNA(\%) = \frac{P}{S + P + L} \times 100\%$$

$$Yield\ of\ Ligation(\%) = \frac{L}{S + P + L} \times 100\%$$

Here, S is the band intensity of the substrate, P and L are and the product of adenylation and ligation, respectively.

T_m values of hairpins calculating by mfold. The melting temperature (T_m) of the hairpin structures in aid-DNA were calculated by "Homodimer simulations" of "the mfold WebServer" (http://www.unafold.org). Salt conditions: 10 mM Na$^+$ and 10 mM Mg^{2+}, other conditions are defaulted ones of this online program.

Mass Spectroscopy Analysis. The adenylated products, generated from high-concentration reactions, were then purified using a gel extraction kit. A direct injection approach was used for introducing the sample to the mass spectrometer (Thermo LCQ Deca XP Max) using a syringe pump with an electrospray ionization source in positive mode with the following parameters: gas flow, 10 L/min; nebulizer pressure, 30 psi; drying gas temperature, 325°C; and capillary voltage, 3100 V. The results were analyzed by Xcalibur software version 2.0.7 (Thermo Fisher Scientific, Inc).

Table S1. Sequences used in this study.

Name	Sequence (5′→3′)	Length (nt)
P1	ATCACCGACTGCCCATAGAGAGGAAAGCGGAGGCGTAGTGGTT	43
Acceptor	GAATTCTAATACGACTCACTATC	23
P1-template	ATGGGCAGTCGGTGATTATAGTGAGTCGTATTAGAATTC	39
P1-aidccmis	ATGGGCAGTCGGTGACCACTGAAAGTC	27
Hp1(Aid-P1)	TTTTTTTGGGCAGTCGGTGATCCCTCTTTGAG	32
Hp2	GCTATCCTAGCGTATGAGATGGCTAGGATAGCCCCTCTTTGAG	43
Hp3	GCTATCCTAGCGTATGAGATGGCTAGGATAGCCCCGCCTTTGGCG	45
Hp4	GCTATCCTAGCGTATGAGATGGCTAGGATAGCCCCTCGAAGAG	43
Hp5	GCTATCCTAGCGTATGAGATGGCTAGGATAGCCCATCTCTCGAAGAG	47
Hp6	GCTATCCTAGCGTATGAGATGGCTAGGATAGCCTCTCGCCGAAGGCG	47
Hp7	GCTATCCTAGCGTATGAGATGGCTAGGATAGCCTATCTCGCCGAAGGCG	49
Hp8	ACTATCCTAGCGTATGAGATGGCTAGGATAGTCCCTCTTTGAG	43
Hp9	CCTATCCTAGCGTATGAGATGGCTAGGATAGGCCCTCTTTGAG	43
Hp10	TCTATCCTAGCGTATGAGATGGCTAGGATAGACCCTCTTTGAG	43
Hp11	ACTATCCTAGCGTATGAGATGGCTAGGATAGTCCCGCCTTTGGCG	45
Hp12	CCTATCCTAGCGTATGAGATGGCTAGGATAGGCCCGCCTTTGGCG	45
Hp13	TCTATCCTAGCGTATGAGATGGCTAGGATAGACCCGCCTTTGGCG	45
Y-adaptor	GATCGGAAGAGCACACGTCTGAACTCCAGTCACATCTCGTATGCCGTCTTCTGCTTG	57
Y-template	AGACGTGTGCTCTTCCGATCTATAGTGAGTCGTATTAGAATTC	43
Y-aidccmis	GTGTGCTCTTCCGATCCACTGAAAGTC	27
Hp14(Aid-Y)	TTTTTTTGTGCTCTTCCGATCCCCTCTTTGAG	32
Hp15	GCTATCCTAGCGTATGAGATGGCTAGGATAGCCCCGCTTTGAG	43
Hp16	GCTATCCTAGCGTATGAGATGGCTAGGATAGCCCCGATTTGAG	43
Hp17	GCTATCCTAGCGTATGAGATGGCTAGGATAGCCCATCTCCACTATTTTAGTGG	53
BGI-Ada	AGTCGGAGGCCAAGCGGTCTTAGGAAGACAACAACTCCTTGGCTCACA	48
Aid-BGI	TTTTTTGCTTGGCCTCCGACTCCCTCTTTGAG	32
SMRT-3T	ATCTCTCTCTTTTCCTCCTCCTCCGTTGTTGTTGTTGAGAGAGATT	46
Aid-SMRT	TTTTTTGGAAAAGAGAGAGATCCCTCTTTGAG	32
NEBnext-Ada	GATCGGAAGAGCACACGTCTGAACTCCAGTCACACTCTTTCCCTACACGACGCTCTTCCGATCT	64
Aid-NEBnext	TTTTTTAGTTCAGACGTGTGCTCTTCCGATCCCCTCTTTGAG	42
Ill3S-Ada	CTCTCTTTCGAACGACACGTCTGAACTCCAGTCAC*C*A*ACACTCTTTCCCTACACGACCGTTCTGTTGAGAG	71
Aid-Ill3S	TTTTTTGTGTCGTTCGAAAGAGAGCCCTCTTTGAG	35
AGATAG	AGATAG	6
AGATAA	AGATAA	6
CGATAG	CGATAG	6
CGCTAG	CGCTAG	6
AGACAA	AGACAA	6
AAATCG	AAATCG	6
CACTAG	CACTAG	6
CACCAG	CACCAG	6
CACCCG	CACCCG	6
Hp2-CA	GCTATCCTAGCGTATGAGATGGCTAGGATAGCACCTCTTTGAG	43
Hp2-CG	GCTATCCTAGCGTATGAGATGGCTAGGATAGCGCCTCTTTGAG	43
Hp2-CT	GCTATCCTAGCGTATGAGATGGCTAGGATAGCTCCTCTTTGAG	43
Hp2-AC	GCTATCCTAGCGTATGAGATGGCTAGGATAGCCACTCTTTGAG	43

Hp2-GC	GCTATCCTAGCGTATGAGATGGCTAGGATAGCCGCTCTTTGAG	43
Hp2-TC	GCTATCCTAGCGTATGAGATGGCTAGGATAGCCTCTCTTTGAG	43
R8D28	rGrCrCrArArUrGrCCAGTCTGATAAGACTCTTCCTCATTCGC	36
Aid-RNA	TTTTTTCAGACTGGCATTGGCCCCTCTTTGAG	32
T56-3A-F	ACGTTGTAAAACGACGGCCAGTGCCAAGCTTGCATGCCTGCAGGTCGACTCTAGAGA	57
T56-3A-R	CTCTAGAGTCGACCTGCAGGCATGCAAGCTTGGCACTGGCCGTCGTTTTACAACGTA	57

The "*" indicates the position of phosphorothioated modification. The 'r' in "R8D28" indicates a nucleotide of RNA.

Table S2. Calculated Mass and Measured Values of Various Structures Analyzed by Mass Spectrometry.

Name	5′-OH Molecular Weight (Da)	5′-PO$_4^{2-}$ Molecular Weight (Da)	5′-App Molecular Weight (Da)	The difference in measurement values between 5′-PO$_4^{2-}$ and 5′-App	The difference in calculated mass between 5′-OH and 5′-App
SMRT-3T	13996.88	14077.1 (cal. 14075.88)	14405.7 (cal. 14404.88)	+328.6 (cal.329.21)	+408.82 (cal.408.18)
NEBNext -Ada	19472.45	19553.6 (cal. 19551.45)	19883.4 (cal. 19880.45)	+329.8 (cal.329.21)	+410.95 (cal.408.18)
ILL3S -Ada	21628.78	21709.1 (cal.22038.78)	22039.8 (cal.22038.78)	+330.7 (cal.329.21)	+411.02 (cal.408.18)

The symbol "+" indicates an increase in mass compared 5′-App to the other structure, and "cal." in parentheses indicates the calculated mass.

The calculated mass (Structure is 5′-OH) was calculated as the following equations:

Calculated mass= $(An \times 313.21) + (Cn \times 289.18) + (Gn \times 329.21) + (Tn \times 304.19) - 61$

Here, An, Cn, Gn and Tn represent the number of A, C, G, and T bases in the sequence, respectively.
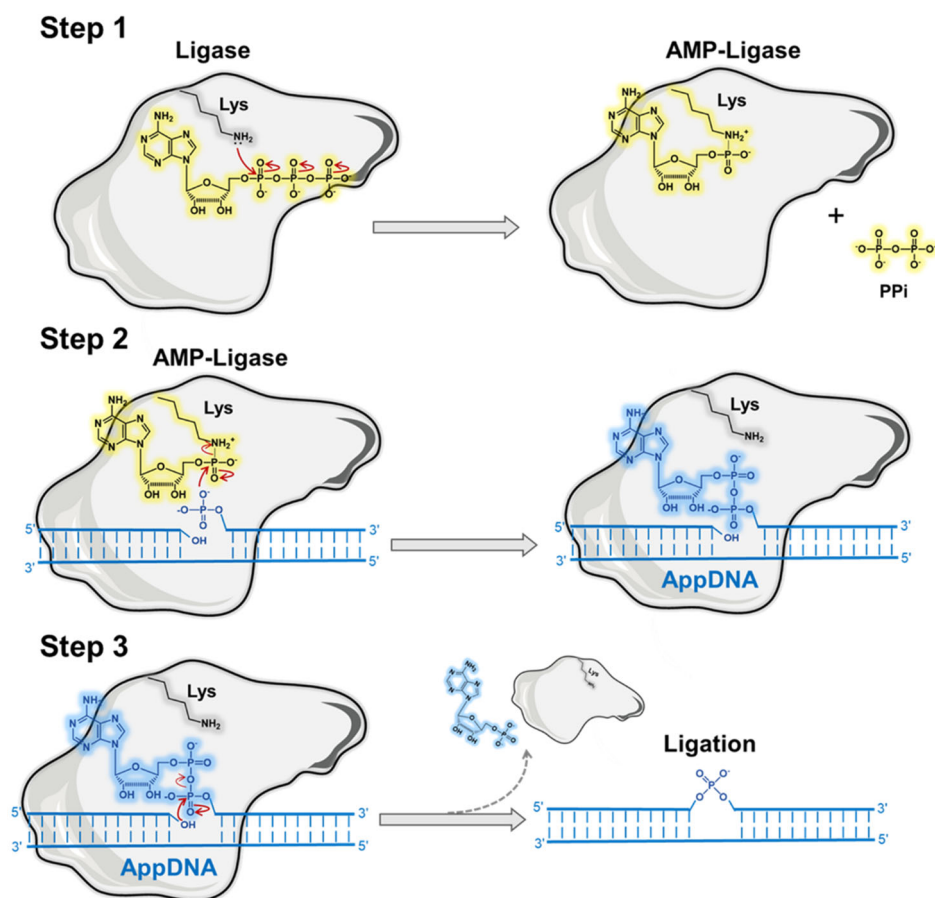
Figure S1. Process of the nick sealing reaction catalyzed by DNA ligase. Step 1 of the main reaction scheme illustrates the generation of E-AMP complex using ATP. Then, AMP is transferred to the 5′-phosphate group to form the AppDNA. The last step including a nucleophilic attack by 3′-OH to form phosphodiester bond. The ligation mechanism of T4 Dnl was proposed by Rossi et al. (ref No.13 in the main text, also shown as follows).

Reference

Rossi, R.; Montecucco, A.; Ciarrocchi, G.; Biamonti, G., Functional characterization of the T4 DNA ligase: a new insight into the mechanism of action. *Nucleic Acids Res*. **1997**, *25*, 2106–2113.
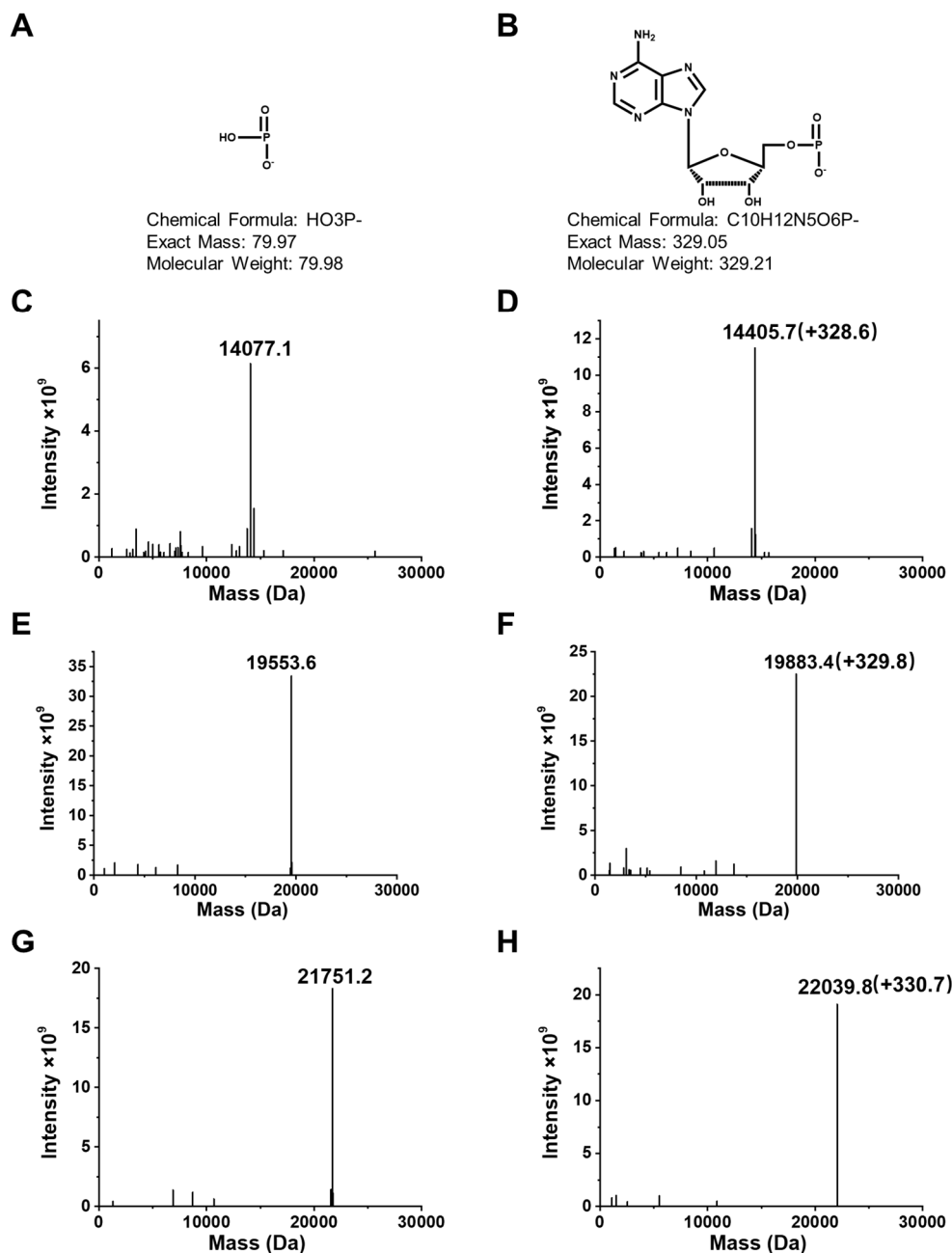
Figure S2. Mass spectroscopy analysis of phosphorylated and adenylated product. **A)** and **B)** shows the structures and masses of nucleic acids at 5′end after phosphorylated and adenylated reactions, respectively. **C, E, G)** represent the mass spectroscopy analysis of phosphorylated SMRT-3T, NEBNext-Ada and ILL3S-Ada, respectively. **D, F, H)** represent the mass spectroscopy analysis of products obtained from adenylated strategy (this study), samples were SMRT-3T, NEBNext-Ada and ILL3S-Ada, respectively.

As shown in **Figure S2B**, post-adenylation treatment should result in a calculated mass increase of 329.21. In comparison with the measured values of phosphorylated products (**Figure S2C, E, G**), the products of SMRT-3T, NEBNext-Ada and ILL3S-Ada exhibit mass increments of 328.6, 329.6, and 330.7, respectively. This observation confirms that these products are indeed adenylated products. The detailed calculated masses are listed in **Table S2**.
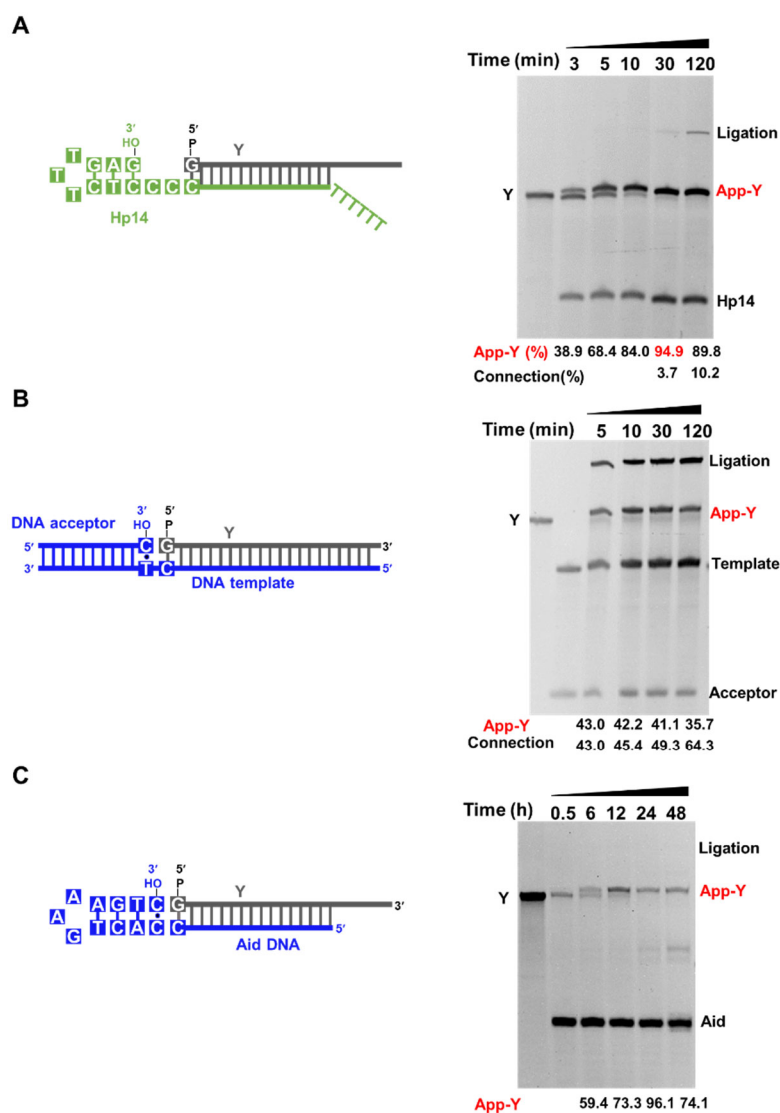
Figure S3. Preparation of App-Y based on T4 Dnl. **A)** Sequences design replied on optimized dynamic nick strategy (This study) and electrophoresis analysis (12% dPAGE, 8.0 M urea, 25% formamide) of adenylation time course. **B, C)** Sequence designs of approaches and by Patel et al. and Luo et al., respectively. Conditions for **A)** and **C)**: 6.0 μM aid-DNA/Hp14, 0.5 mM ATP, 37°C; Conditions for **B):** 5.0 μM template, 6.0 μM acceptor, 6.0 mM ATP, 25°C. Other conditions: 4.0 μM Y, 0.25 U/μL T4 Dnl, 1× T4 Dnl buffer.

Applying the novel strategy proposed in this study, we were able to achieve 95% adenylation of Y adaptor (Illumina sequencing platform) in 30 mins (**Figure S3A**). However, when other strategies were employed, they either resulted in an inability to suppress the product of ligation (**Figure S3B**, 43%) or required an excessively long reaction time (**Figure S3C**, 24 h).

Reference

Patel, M. P.; Baum, D. A.; Silverman, S. K., Improvement of DNA adenylation using T4 DNA ligase with a template strand and a strategically mismatched acceptor strand. *Bioorg. Chem.* **2008**, *36*, 46–56.

Luo, J.; Chen, H.; An, R.; Liang, X., Efficient preparation of AppDNA/AppRNA by T4 DNA ligase aided by a DNA involving mismatched mini-hairpin structure at its 3′ side. *Bull. Chem. Soc. Jpn.* **2022**, *95*, 1380–1388.
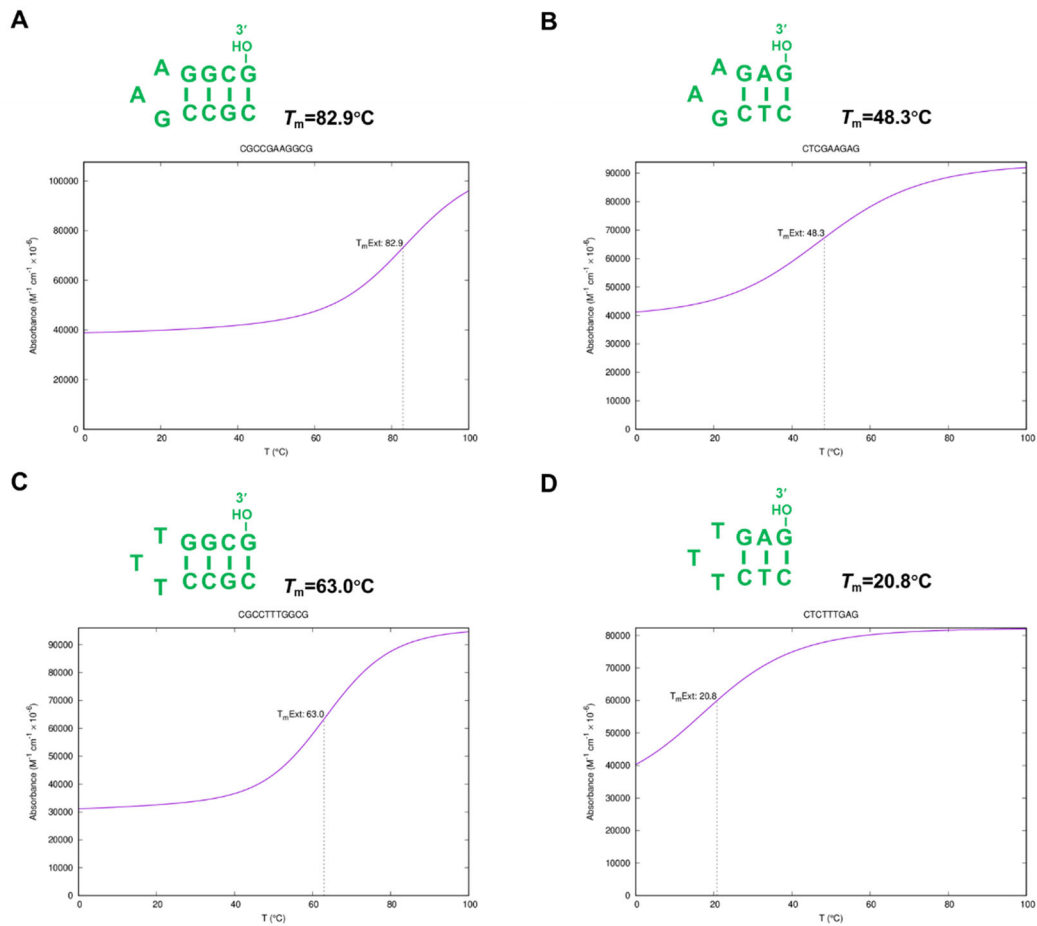
Figure S4. T_m of hairpin structure in aid-DNA. **A–D)** Sequence of the hairpins in aid-DNA and T_m calculated by the "Homodimer Simulations" mode of mfold (http://www.unafold.org/),which can be used to describe the stability of the hairpin structure. Salt conditions: 10 mM Na+ and 10 mM Mg2+, other conditions are defaulted ones of this online program.
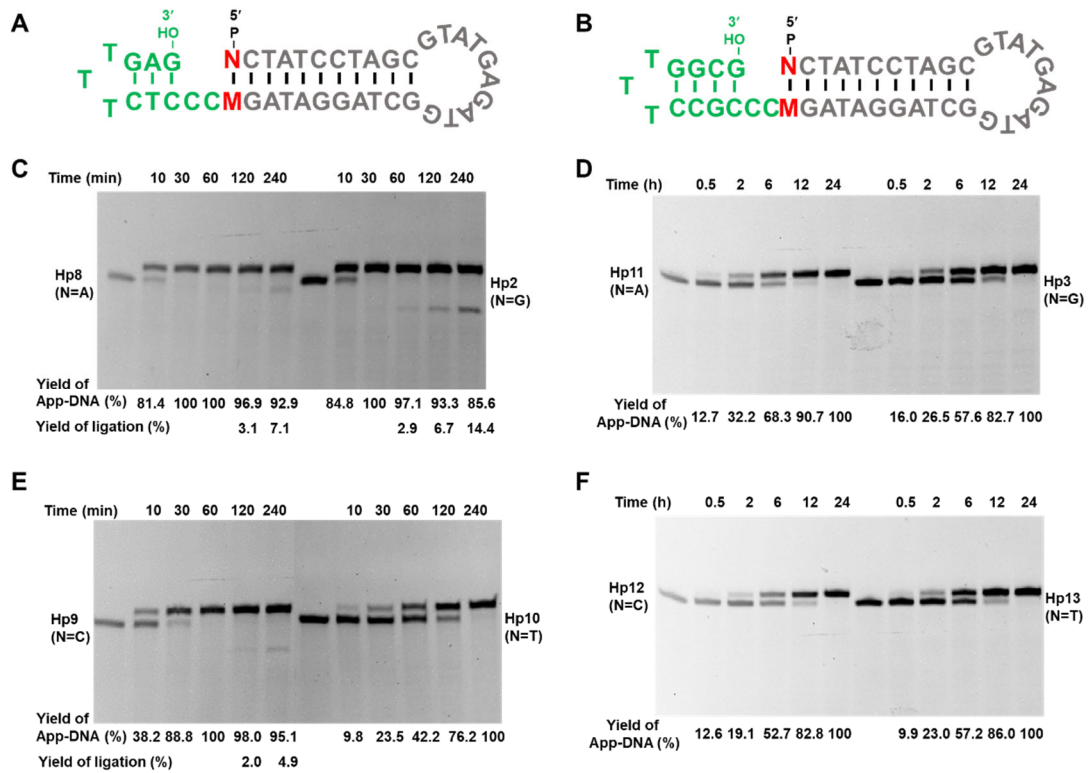
Figure S5. Effect of competition on adenylation and ligation. **A)** and **C)** Sequences design for experiment. In the ssDNA, sequences complementary to **Hp7** and **Hp17** are denoted in blue, while sequences that mismatch are highlighted in yellow and marked in red font. **B)** and **D)** Yield of product of adenylation and ligation. Reaction conditions: 4.0 μM **Hp7** or **Hp17**, 6.0 μM ssDNA (No.1–6), 0.25 U/μL T4 Dnl, 1× T4 Dnl buffer (0.5 mM ATP), 37°C for 2.0 h, analyzed by 12% dPAGE.

As shown in **Figure 3G**, **Hp7** is capable of achieving an 82% yield of adenylation with just a reaction time of 2.0 h. However, when an ssDNA is introduced as a competitor into the adenylation reaction, it has been observed that even ssDNA with a high degree of mismatch, such as No. 5–9, possesses the ability to slow down the progression of the adenylation reaction **(Figure S5A and B)**. These findings highlight the crucial role of the dynamic changing between the gap state and nick state in the adenylation reaction. It's noteworthy that even a minor disturbance can lead to a reduction in the reaction rate.

Furthermore, **Figures S5C and D** provide additional evidence for the importance of the hairpin structure. Compared to **Hp7** with stable hairpin, the open conformation with a 3′ overhang **(Hp17)** shows a reduction in both adenylation and ligation efficiency, except in cases where there is more complementarity at the 3′ end (ssDNA No.1,3,4). Hence, we believe that a moderately stable hairpin facilitates the rapid transition to the instantaneous nick state, which in turn enhances its propensity for ligase's binding.

Figure S6. Effect of base pair flanking the nick at 5′ side on adenylation. **A)** and **B)** Sequences design. **C–F)** Electrophoresis analysis of the adenylated products. Reaction conditions: 4.0 μM ssDNA, 0.25/μL U T4 Dnl, 1× T4 Dnl buffer (0.5 mM ATP), 37ºC.
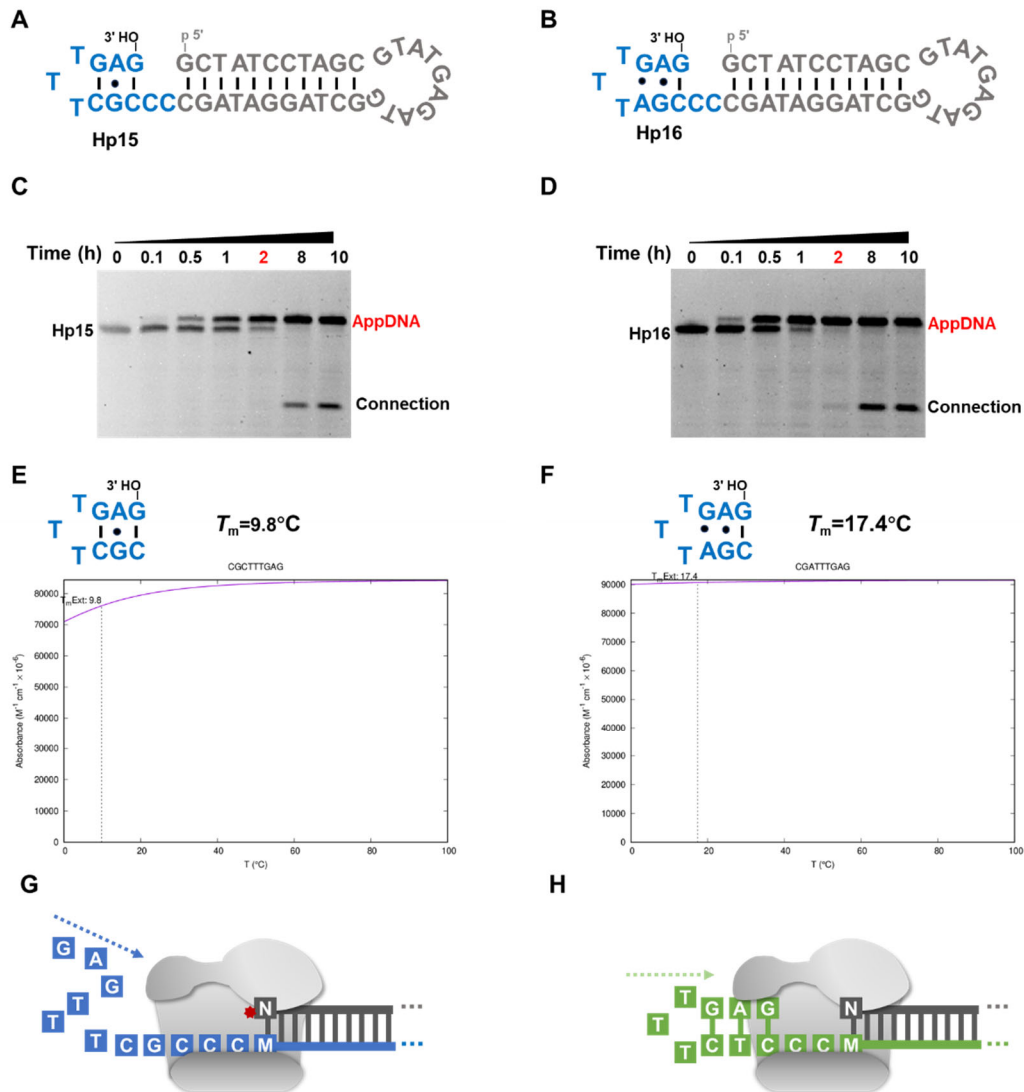
Figure S7. Effect of unstable hairpin structure on efficiency of adenylation. **A)** and **B)** Sequences design of aid-DNA with unstable hairpin. **C)** and **D)** Electrophoresis analysis of the adenylation reaction of **Hp15,16**, respectively. Reaction conditions: 4.0 μM ssDNA, 0.25 U/μL T4 Dnl, 1× T4 Dnl buffer (0.5 mM ATP), 37°C. **E)** and **F)** the T_m of unstable hairpin sequence (**Hp15,16**) calculated by mfold. **G)** Proposed schematic view of the binding of T4 Dnl to aid-DNA with unstable hairpin. **H)** Proposed schematic view of the binding of T4 Dnl to aid-DNA with stable hairpin.

When mismatches are introduced into the stem, the efficiency of adenylation decreases, which could be attributed to the inability of the ligase to bind to the stem in advance. An unstable hairpin structure will make it more difficult to form the instantaneous nick state, consequently slowing down both the adenylation and ligation reaction. For those stable hairpin like **Hp2**, the stem is already in the binding domain of T4 Dnl, the dynamic changes occur more rapidly.
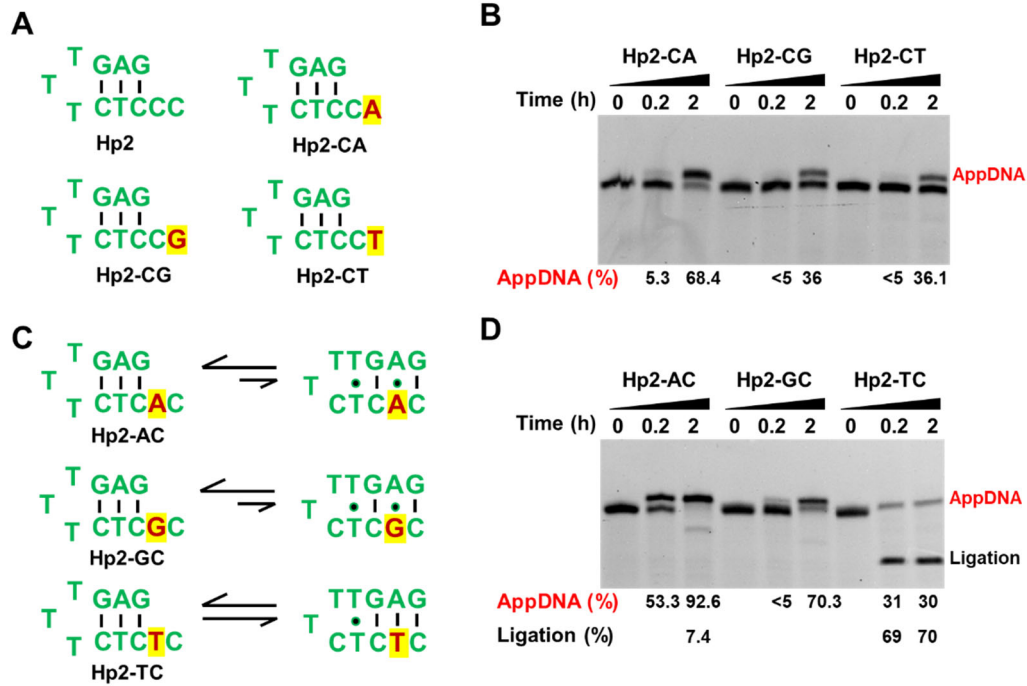
Figure S8. Effect of the instantaneous nick state's forming. **A)** and **C)** Sequences design of hairpin in aid-DNA. **B)** and **D)** Electrophoresis analysis of the adenylated products. Reaction conditions: 4.0 μM ssDNA, 0.25/μL U T4 Dnl, 1× T4 Dnl buffer (0.5 mM ATP), 37°C.

Altering the sequence within the gap region of **Hp2** from CC to CA, CG, or CT (altering the first base on the 5'side) disrupts the formation of the instantaneous nick state due to mismatches (**Figure S8A**). The results indicate a significant decrease in adenylation rate compared to the original **Hp2** (**Figure 3B**, nearly 100% adenylation is achieved within 0.5 h). On the other hand, maintaining the first base on the 5'side of the gap region enables the formation of the nick state, while altering the second base (**Figure S8C**). When the nick state occurs with three consecutive complementary bases (**Hp2-TC**), it fails to inhibit subsequent ligation reactions, resulting in a 70% ligation yield. This underscores the necessity for the nick state to be instantaneous. Only unstable structures can prompt enzyme dissociation, halting the reaction at the adenylation step. Changing the second C to A or G also allows adenylation with slightly lower efficiency than **Hp2**. Therefore, the 3'part of **Hp2** is chosen as the structure of aid-DNA for subsequent adenylation.
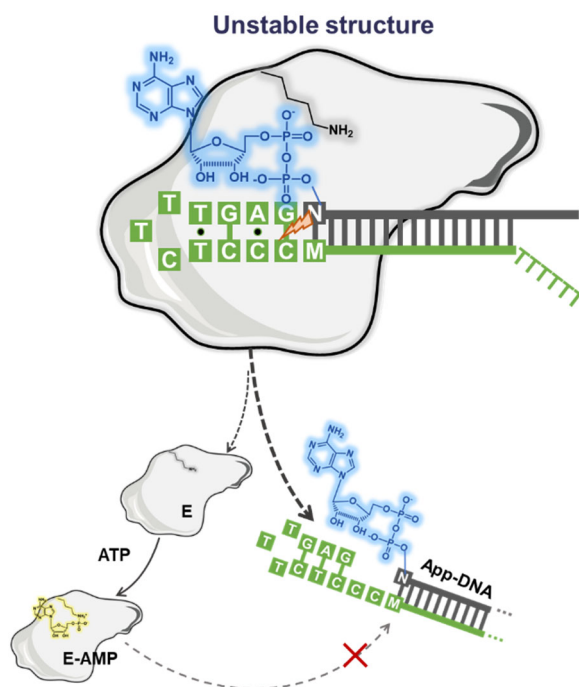
Figure S9. Proposed Mechanisms for Efficient Adenylation Using the Instantaneous Nick Strategy.

The entire reaction process involves conformational changes between step 2 adenylation and step 3 final ligation (the E-AMP-DNA complex). The unstable nick state cannot transition to a form suitable for step 3 ligation, causing the reaction to halt at the adenylation step. Additionally, as the E-AMP-DNA complex undergoes conformational changes, it dissociates. Then ligase becomes E-AMP at high ATP concentrations, which then cannot interact with AppDNA.
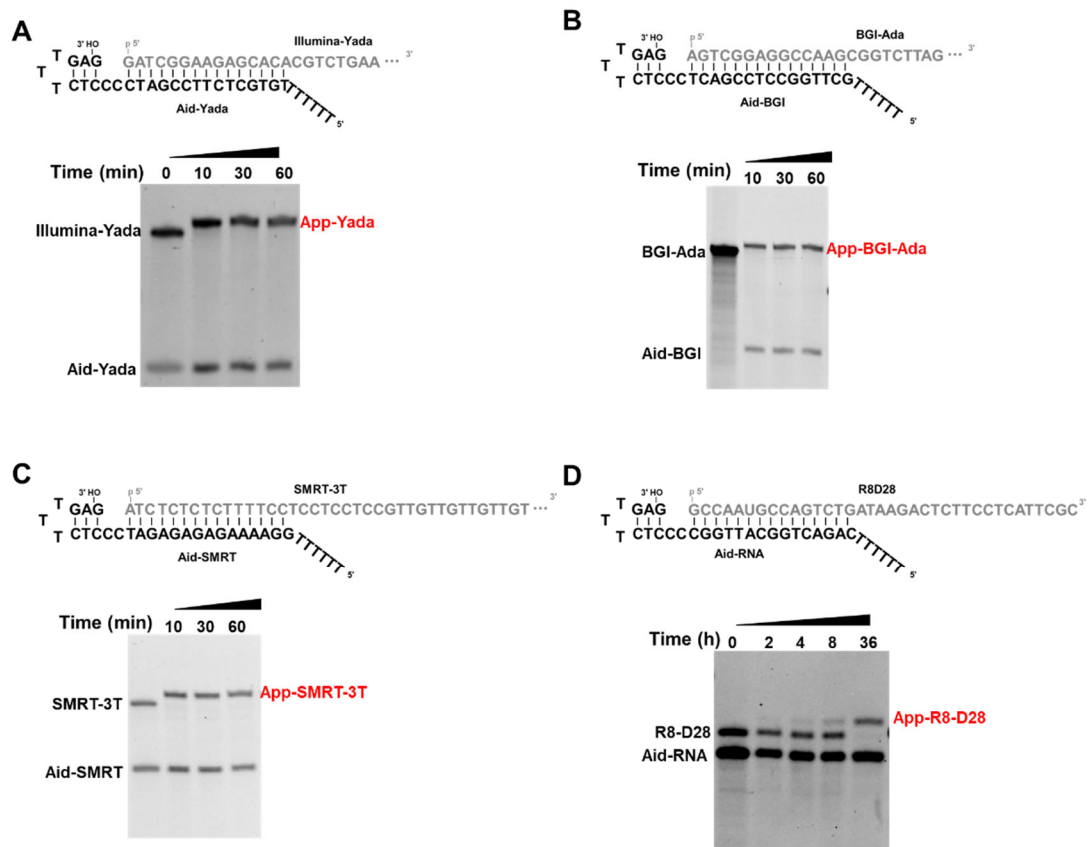
Figure S10. Optimized adenylation conditions applied in other commercialized sequencing adaptors and RNA. **A–D)** Electrophoresis analysis of the adenylated products. Reaction conditions: 4.0 μM ssDNA, 6.0 μM aid-DNA, 0.25 U/μL T4 Dnl, 1× T4 Dnl buffer (2.0 mM ATP), 37°C.
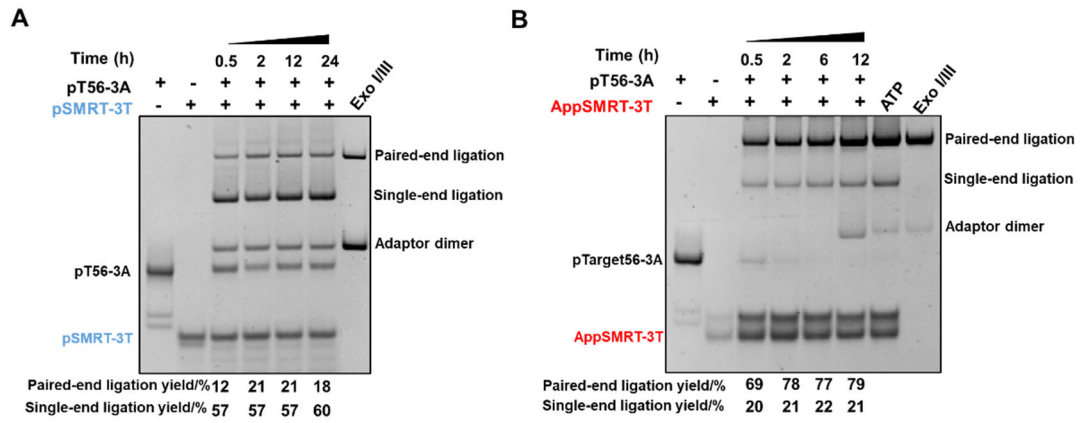
Figure S11. Ligation between adaptors (pSMRT-3T or AppSMRT-3T) and pT56-3A. **A)** Electrophoresis analysis of the ligation products of pSMRT-3T and pT56-3A. Ligation conditions: 0.5 μM pT56-3A, 2.5 μM pSMRT-3T, 0.25 U/μL T4 Dnl, 1× T4 Dnl buffer (0.5 mM ATP), 25ºC. **B)** Electrophoresis analysis of the ligation products of AppSMRT-3T and pT56-3A. Ligation conditions: 0.5 μM pT56-3A, 2.5 μM AppSMRT-3T, 0.25 U/μL T4 Dnl, 1× T4 Dnl buffer (no ATP), 25ºC.