An ATP-dependent ligase with substrate flexibility involved in assembly of the peptidyl nucleoside antibiotic polyoxin

Rong Gong,[1] Jianzhao Qi,[1] Pan Wu,[1] You-sheng Cai,[1] Hongmin Ma,[1] Yang Liu,[1] He Duan,[1] Meng Wang,[1] Zixin Deng,[1,2] Neil P. J. Price,[3] Wenqing Chen[1,2*]

[1]Key Laboratory of Combinatorial Biosynthesis and Drug Discovery, Ministry of Education, and School of Pharmaceutical Sciences, Wuhan University, Wuhan 430071, China.

[2]State Key Laboratory of Microbial Metabolism, and School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai 200030, China.

[3]Agricultural Research Service, US Department of Agriculture, National Center for Agricultural Utilization Research, Peoria, IL, USA.

For correspondence: **Wenqing Chen,** Key Laboratory of Combinatorial Biosynthesis and Drug Discovery, Ministry of Education, and School of Pharmaceutical Sciences, Wuhan University, Wuhan 430071, China. ***E-mail:** wqchen@whu.edu.cn, **Tel:** +86-27-68756713

Table of Contents

1. Supplementary Materials and Methods

1.1 Enzymes, chemicals, and reagents

The restriction enzymes and T4 DNA ligase used in this study were the products of New England Biolabs, and DNA polemerase was purchased from TOYOBO Life Science (high fidelity KOD plus) or Takara (rTaq).

1.2 In-frame deletion of the target *polG* gene by PCR-targeting strategy

For targeted inactivation of *polG* on pCHW201, a kanamycin resistance cassette (*neo*) from SuperCos1 was amplified with primers polGtgtF and polGtgtR (Table S2), and then recombined into pCHW201 by PCR-targeting strategy (1) to yield pCHW201/*polG-neo*. After that, the *neo* cassette was deleted *in vitro* by XbaI-SpeI double digestion (unique sites), and religated to generate the in-frame deletion scar of the target gene (pCHW201/*ΔpolG*) (Fig. S1). The in-frame deletion was subsequently confirmed by PCR using primers polGidF and polGidR (Table S2).

1.3 Cultivation and metabolite analysis of the CXR14::pCHW201 and its derivative

CXR14::pCHW201 and its derivative CXR14::pCHW201*/ΔpolG* were both cultivated in liquid fermentation medium according to the protocol described previously (2). Then the culture broth was acidified to pH3.0-4.0 with oxalic acid, and purified by Dowex 50W × 8(H^+), then concentrated and filtered for HPLC or LC-HRMS analysis.

For the purification of the target metabolites, the samples were separated on a HPLC (Shimadzu LC-20AT) equipped with ZORBA SB-C18 Column (Agilent, 5 μm, 9.4 × 250 mm) with a linear gradient of 5%-35% phase B (phase A, aqueous 0.15% TFA; phase B, MeOH) over 30 min at flow rate of 3 ml/min, the elutions were monitored by a DAD detector at 263 nm. The preparation of CPOAA was performed on HPLC (Shimadzu HPLC-ELSD) with a reverse phase C18 column (Inertsil ODS-3, 4.6 × 250 mm, 5 μm) following the same program as above except for the flow rate (1 ml/min).

1.4 Enzymatic preparation of NIK-Cx

The enzymatic reaction consisting of 50 mM Tris-HCl (pH 8.0), 5 mM NIK-X and 1 mg/ml Pronase E was incubated at 25°C for 16 hr (3). NIK-Cx was purified by HPLC using the gradient described above for POL preparation, but monitored at 287 nm by a DAD detector.

1.5 Production and purification of PolG and its variants in *E. coli.*

The optimized *polG* structural gene (Table S1) was cloned to the NdeI-EcoRI sites of pET28a, and the expression construct was subsequently transformed into *E. coli* Rosseta(DE3)/pLysS.

Expression and purification for His_6-tagged proteins were performed as follows: cells were grown at 37°C in LB medium (50 μg/ml kanamycin and 25 μg/ml chloramphenicol) until an OD_{600} of 0.6. Cooling for 10 min on ice, the cells were then induced with 0.15 mM (final conc.) β-D-1-thiogalactopyranoside (IPTG) for 24 hr at 15°C. After that, the cells were harvested by centrifugation (5,000 r/min, 10 min, 4°C), re-suspended in 30 ml lysis buffer (25 mM Tris, pH 8.0, 150 mM NaCl, 5 mM imidazole) and lysed by ultrasonication on ice. Cellular debris was removed by centrifugation (10,000 r/m, 20 min, 4°C). After incubation with 1 ml of Ni-NTA agarose resin (Qiagen) on ice for 1 hr, the supernatant (soluble proteins)-resin mixture was loaded onto a gravity flow column. Washed with washing buffer (25 mM Tris, pH 8.0, 150 mM NaCl, and 20 mM imidazole) and the proteins were subsequently eluted with elution solution (25 mM Tris-Cl, pH 8.0, 150 mM NaCl, and 200 mM imidazole), the purified protein was evaluated by 12% SDS-PAGE. Protein concentration was determined by the Bradford assay using serum albumin as the standard. PolG yield in *E. coli* Rosseta(DE3)/pLysS was 847.59 μg/l. Finally, the proteins were flash-frozen and stored at -40°C for further use.

2. Supplementary Figures

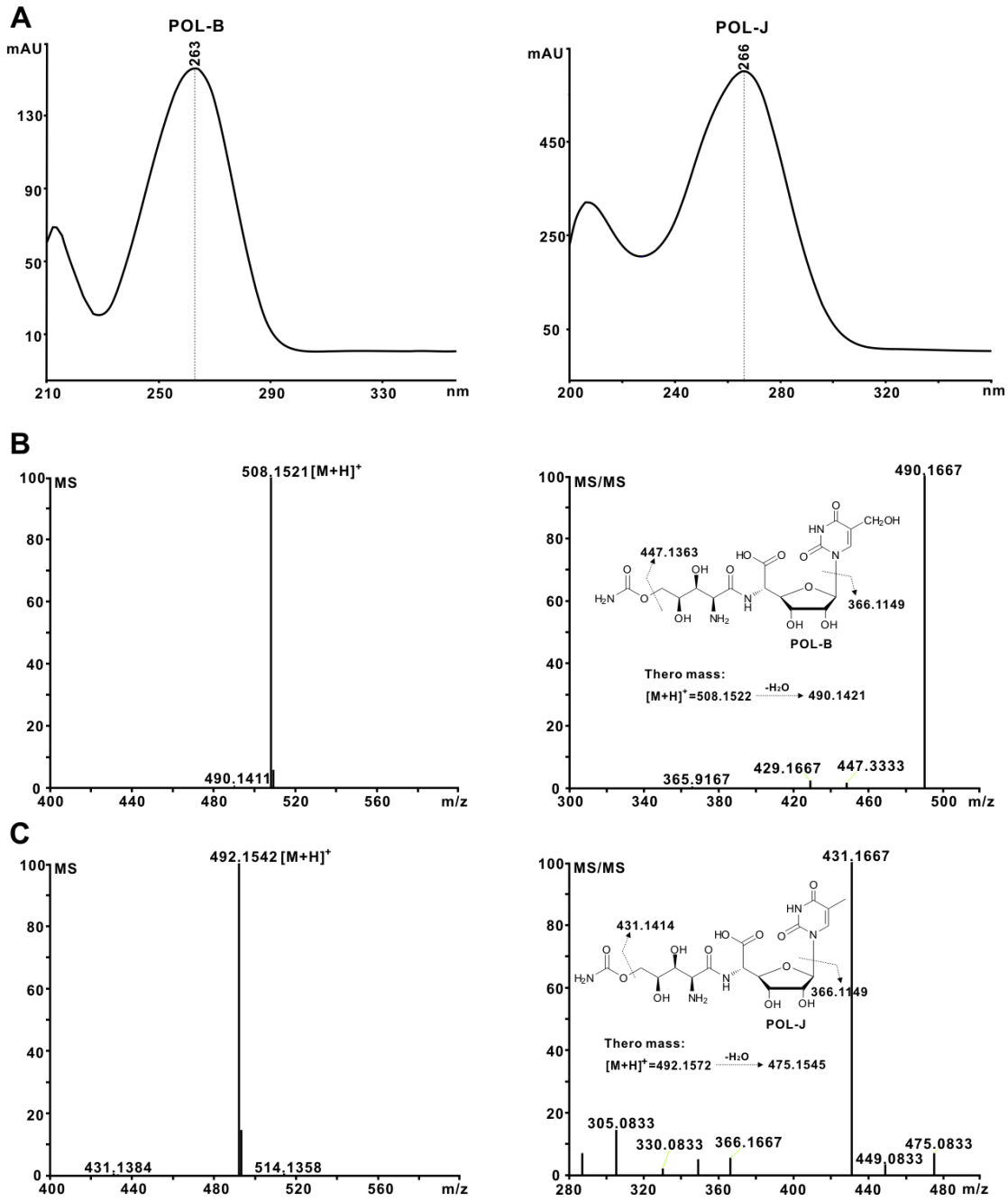


Figure S1. HPLC and LC-HRMS analysis of the metabolites produced from the CXR14::pCHW201.

(A) The UV spectrum of POL-B and POL-J. (B) LC-HRMS (left) and LC-HRMS/MS (right) analysis of POL-B produced by CXR14::pCHW201. (C) LC-HRMS (left) and LC-HRMS/MS (right) analysis of POL-J produced by CXR14::pCHW201.

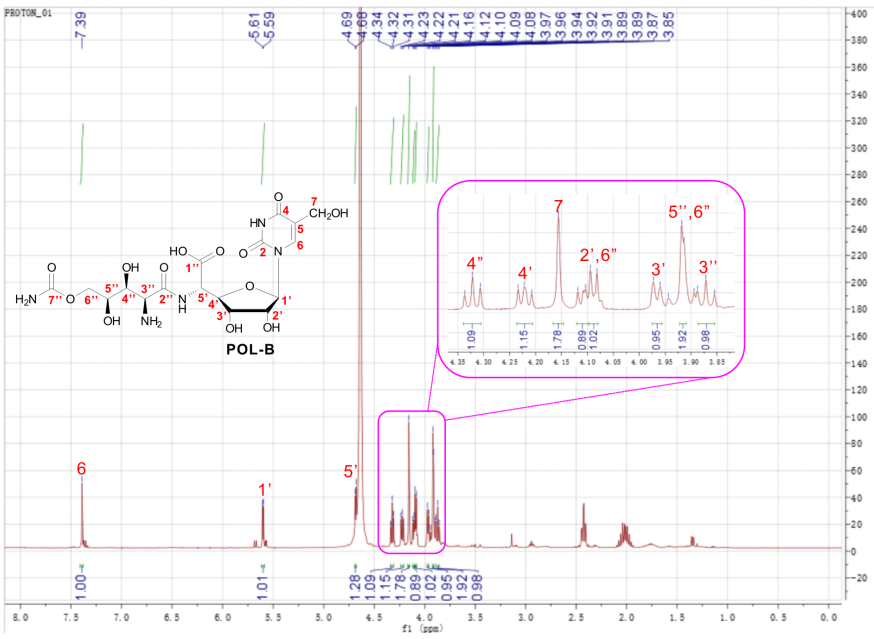


Figure S2A.^{1}H NMR data of POL-B.

^{1}H NMR (400 MHz, D_2O) δ 7.39 (1H, s), 5.61 (1H, d, *J* = 8.0 Hz), 4.69 (1H, d, *J* = 4.0 Hz), 4.32 (1H, t, *J* = 8.0 Hz), 4.22 (1H, t, *J* = 4.0 Hz), 4.16 (2H, s), 4.12 (1H, d, *J* = 8.0 Hz), 4.09(1H, d, *J* = 4.0 Hz), 3.96 (1H, t, *J* = 4.0 Hz), 3.92 (2H, d, *J* = 4.0 Hz), 3.89-3.85 (1H, m).

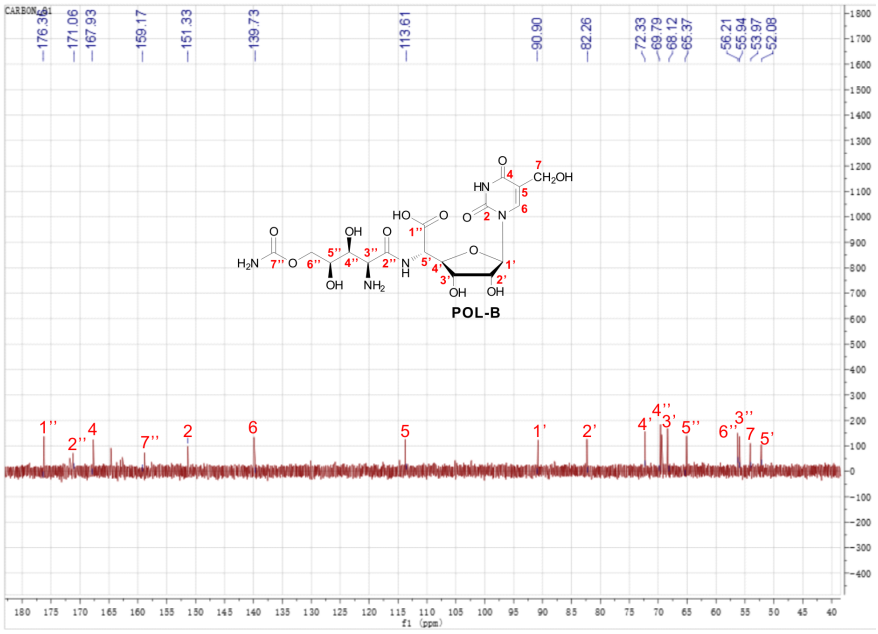


Figure S2B. ^{13}C NMR data of POL-B.

^{13}C NMR (400 MHz, D_2O), δ 176.36, 171.06, 167.93, 159.17, 151.33, 139.73, 113.61, 90.90, 82.26, 72.33, 69.79, 68.12,65.37, 56.21, 55.94, 53.97, 52.08.

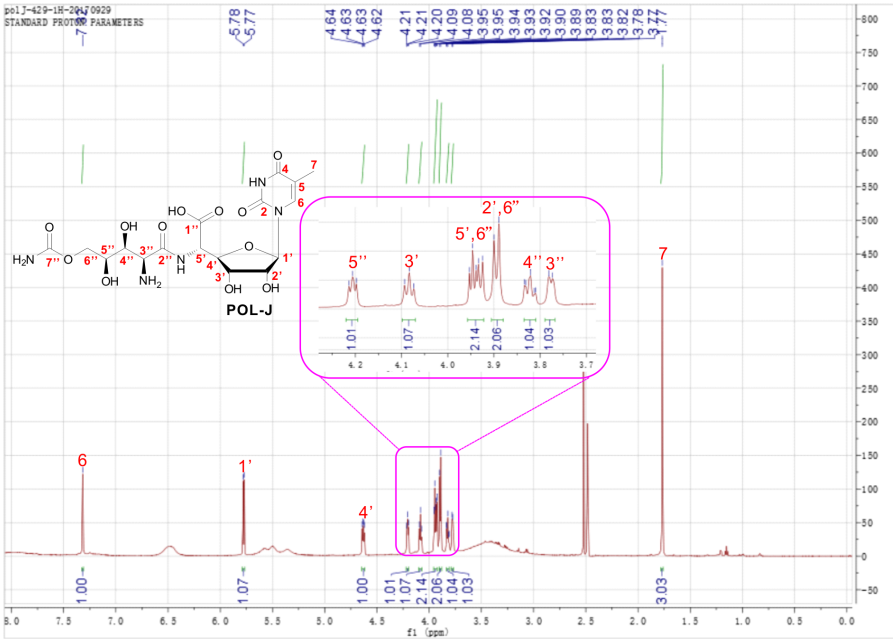


Figure S3A. ^{1}H NMR data of POL-J.

^{1}H NMR (600 MHz, DMSO-d_6) δ 7.32 (1H, s), 5.78 (1H, d, *J* = 6.0 Hz), 4.63 (1H, t, *J* = 6.0 Hz), 4.21 (1H, d, *J* = 6.0 Hz), 4.08 (1H, t, *J* = 6.0 Hz), 3.92-3.95 (2H, m), 3.90 (2H, d, *J* = 6.0 Hz), 3.82 (1H, t, *J* = 6.0 Hz), 3.78 (1H, d, *J* = 6.0 Hz), 1.77 (3H, s).

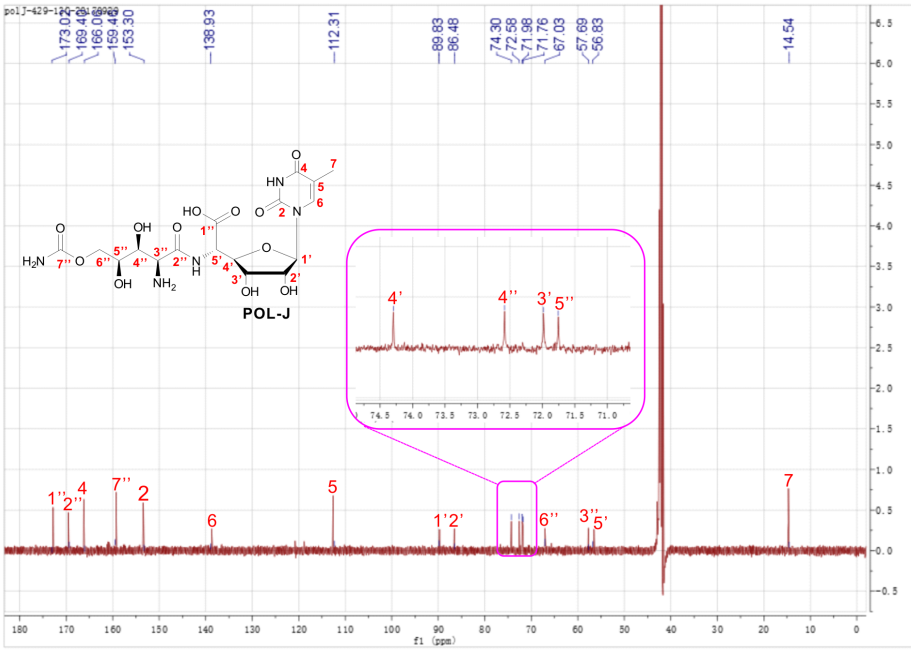


Figure S3B. ^{13}C NMR data of POL-J.

^{13}C NMR (600 MHz, DMSO-d_6), δ 173.02, 169.40, 166.06, 159.46, 153.30, 138.93, 112.31, 89.83, 86.48, 74.30, 72.58, 71.98, 71.76, 67.03, 57.69, 56.83, 14.54.

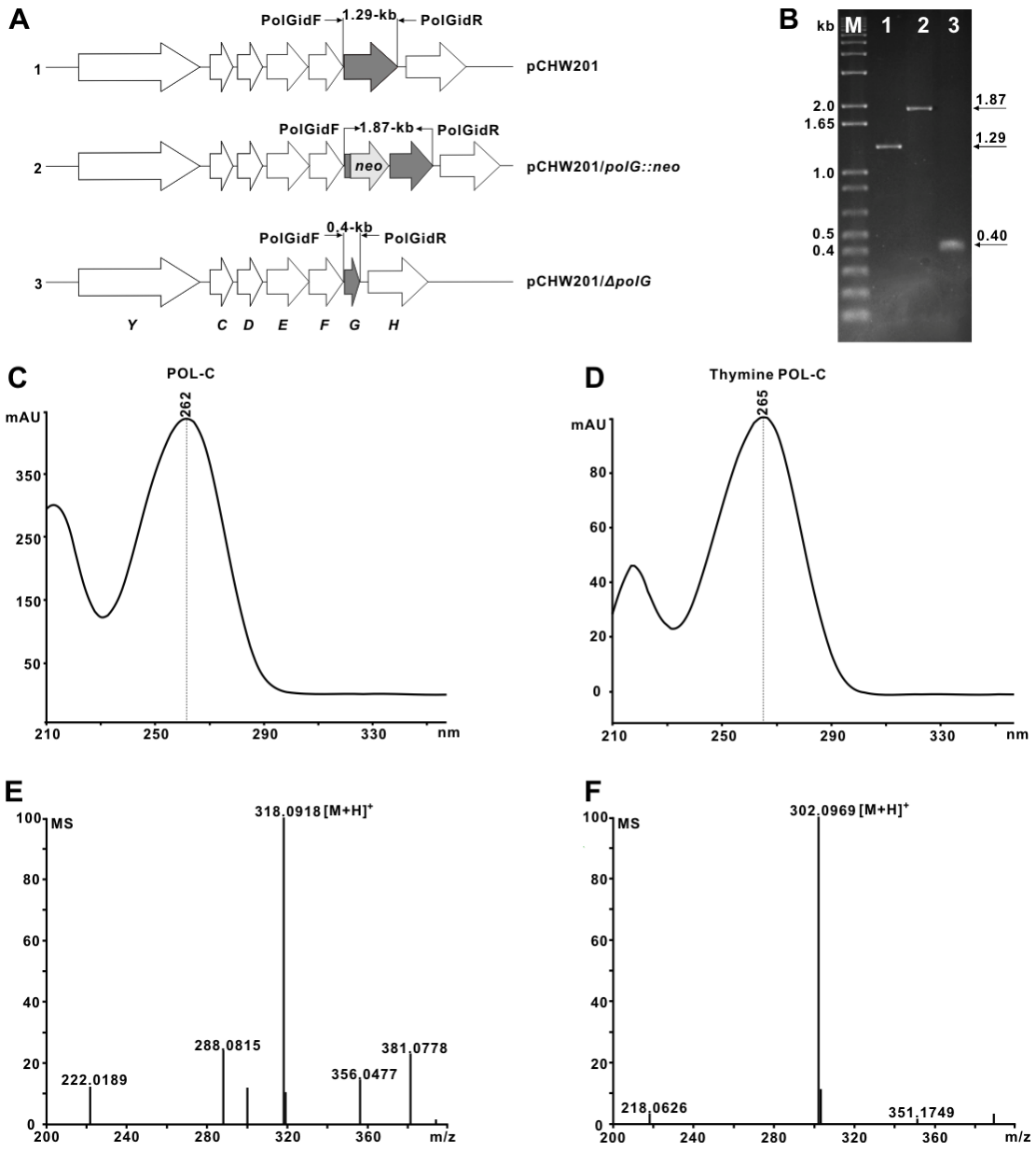


Figure S4. In-frame deletion of *polG* on pCHW201 and HPLC/LC-HRMS analysis of the target metabolites produced by CXR14::pCHW201/*ΔpolG* recombinant.

(A) Schematic illustration for the construction of *polG via* PCR-targeting technology. (B) PCR identification of the in-frame deletion mutant. (C) UV spectrum of POL-C. (D) UV spectrum of thymine POL-C. (E) LC-HRMS analysis of POL-C produced by CXR14::pCHW201/*ΔpolG*. (F) LC-HRMS analysis of thymine POL-C produced by CXR14::pCHW201/*ΔpolG*.

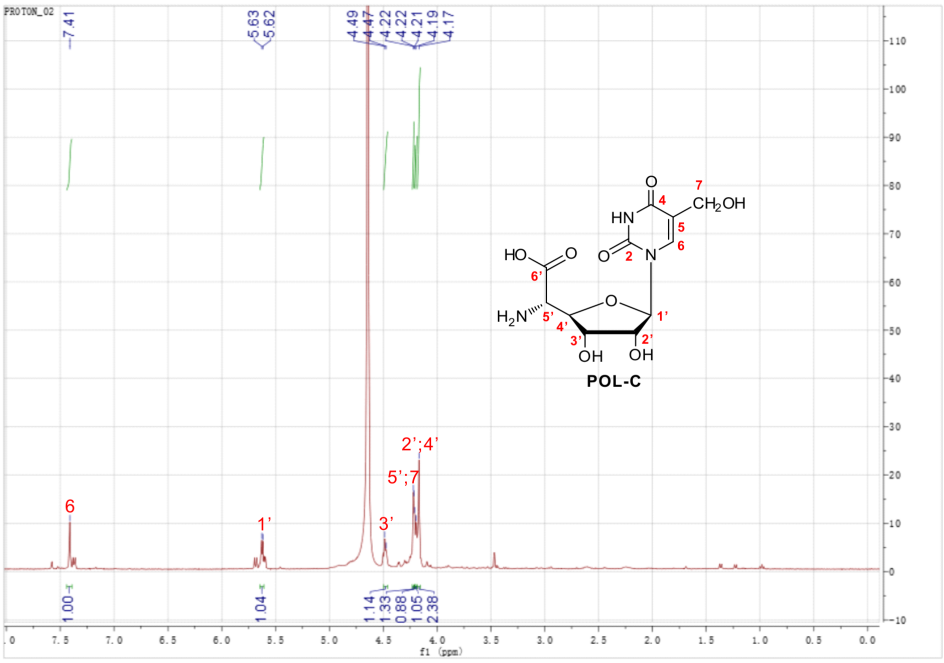


Figure S5A. ^{1}H NMR data of POL-C.

^{1}H NMR (400 MHz, D_2O) δ 7.41 (1H, s), 5.63 (1H, d, *J* = 4.0 Hz), 4.49 (1H, d, *J* = 8.0 Hz), 4.22 (1H, d, *J* = 4.0 Hz), 4.21 (2H, s), 4.19 (1H, s), 4.17 (1H, s).

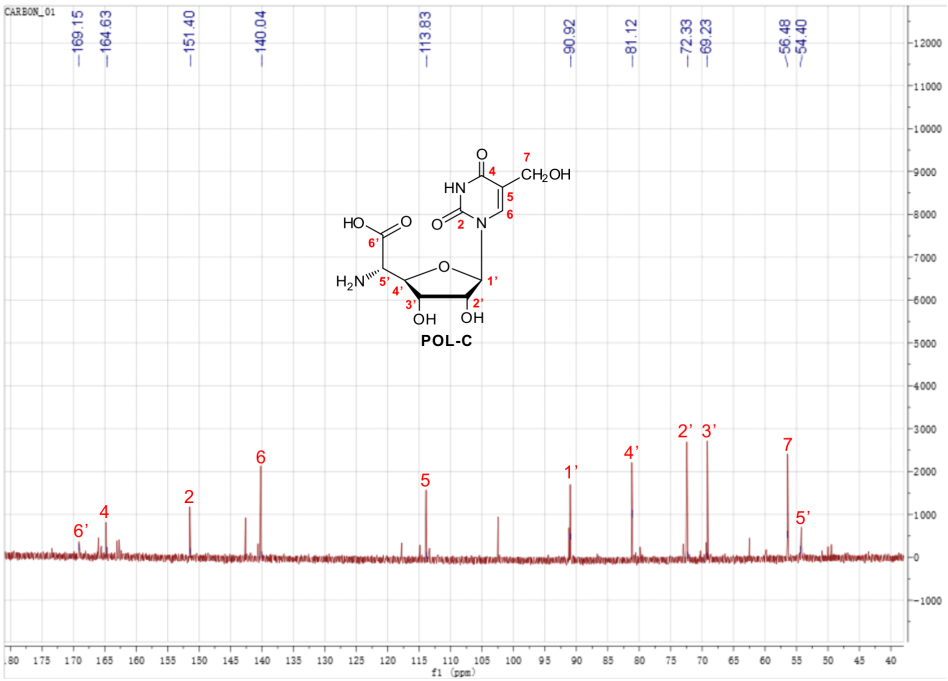


Figure S5B. ^{13}C NMR data of POL-C.

^{13}C NMR (400 MHz, D_2O), δ 169.15, 164.63, 151.40, 140.04, 113.83, 90.92, 81.12, 72.33, 69.23, 56.48, 54.40.

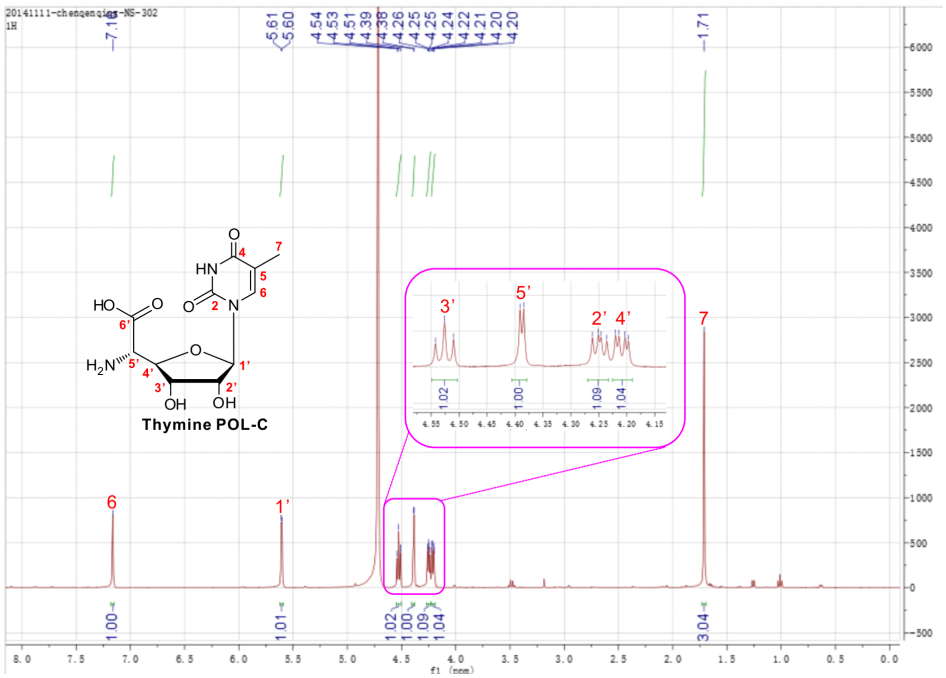


Figure S6A. ^{1}H NMR data of thymine POL-C.

^{1}H NMR (400 MHz, D_2O) δ 7.16 (1H, s), 5.61 (1H, d, *J* = 4.0 Hz), 4.53 (1H, t, *J* = 8.0 Hz), 4.39 (1H, d, *J* = 4.0 Hz), 4.25 (1H, m), 4.21 (1H, m), 1.71 (3H, s).

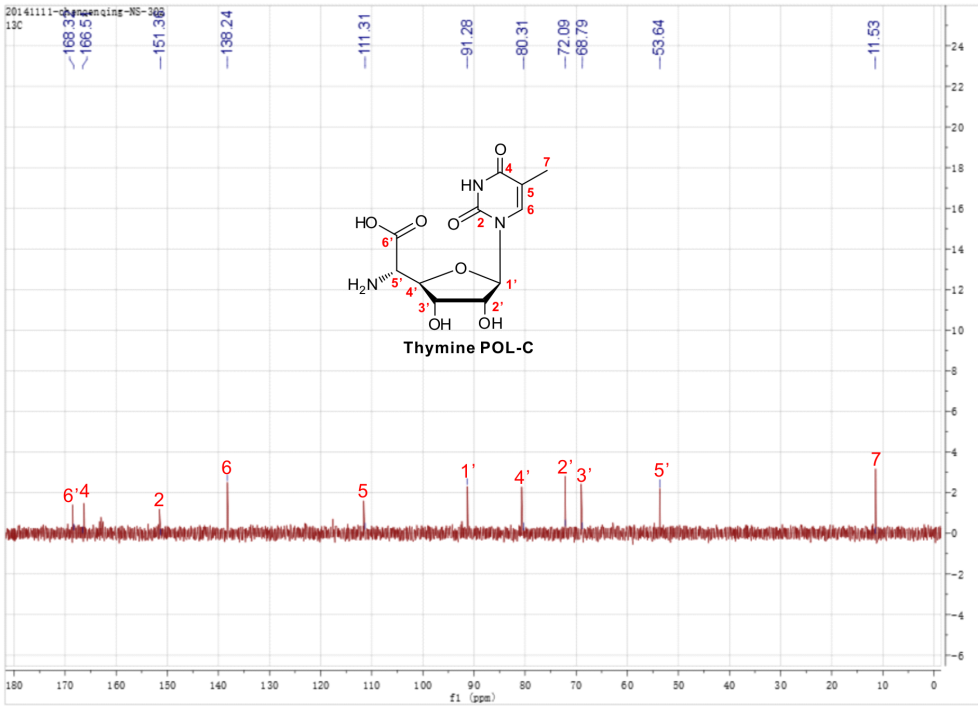
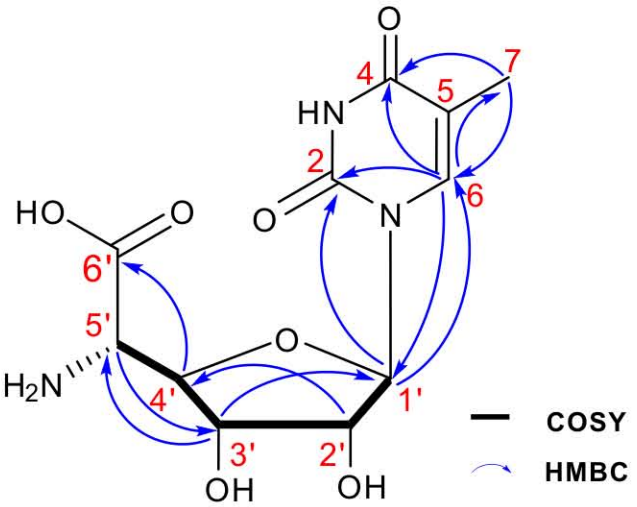


Figure S6B. ^{13}C NMR data of thymine POL-C.

^{13}C NMR (400 MHz, D_2O), δ 168.32, 166.51, 151.36, 138.24, 111.31, 91.28, 80.31, 72.09, 68.79, 53.64, 11.53.



Thymine POL-C

Figure S6C. 2D NMR analysis of thymine POL-C.

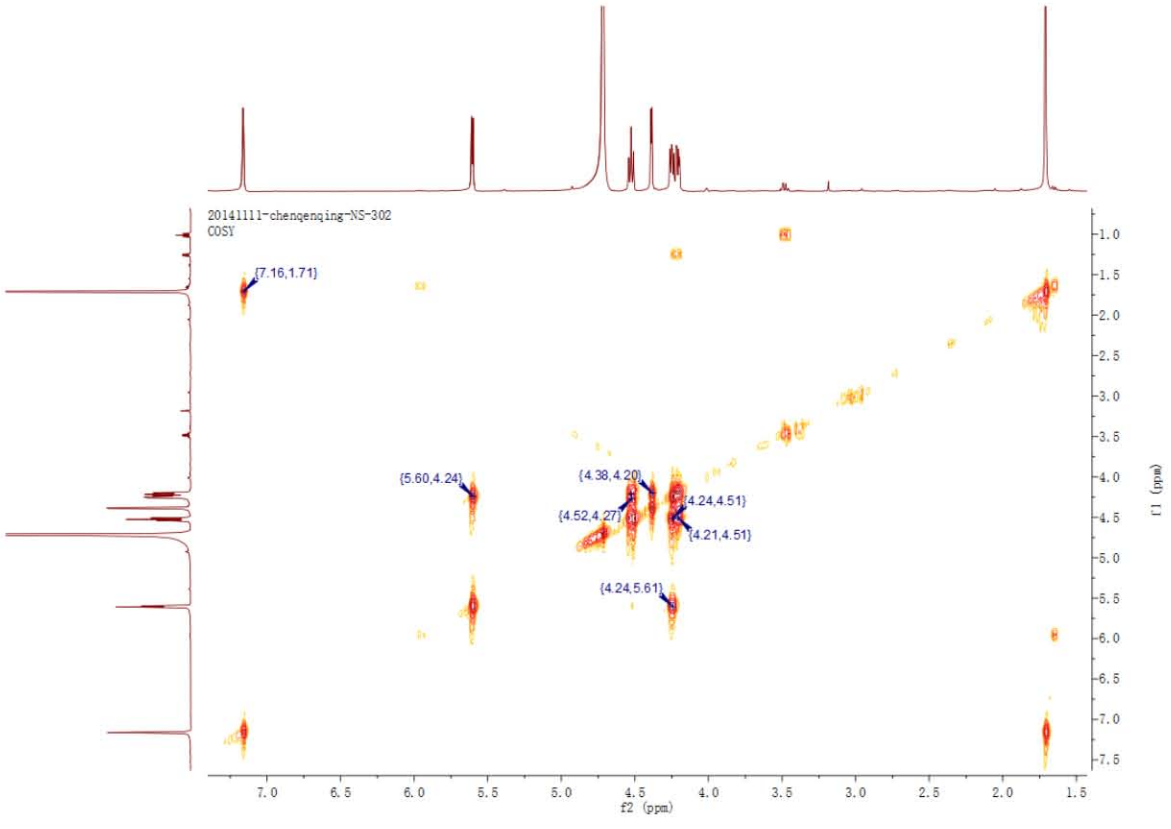


Figure S6D. COSY NMR data of thymine POL-C

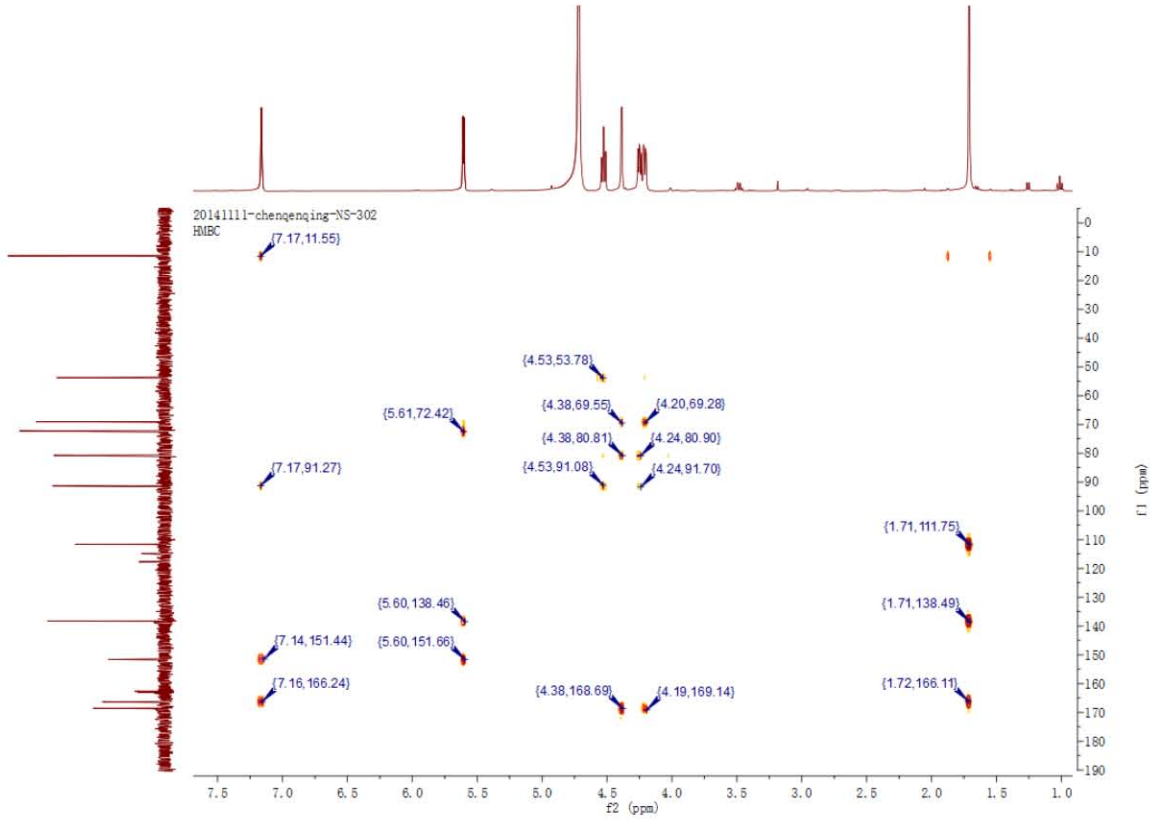


Figure S6E. HMBC NMR data of thymine POL-C.

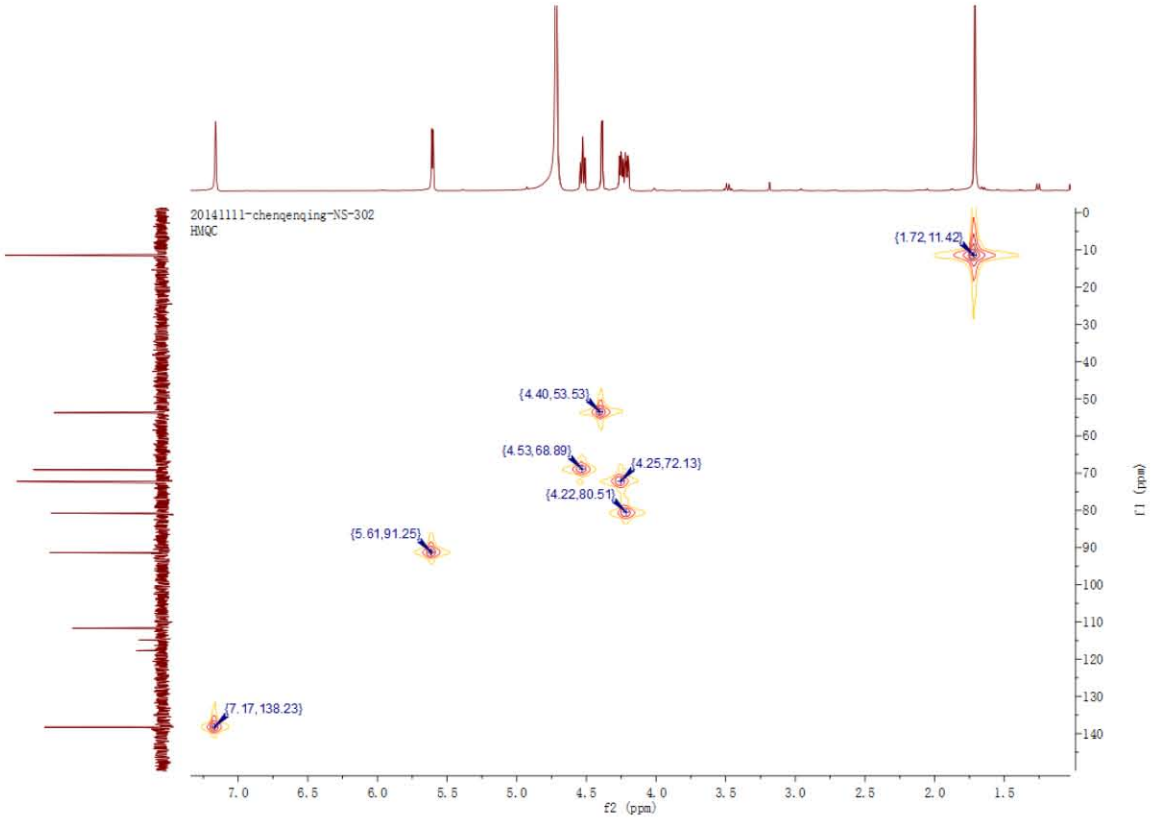


Figure S6F. HMQC NMR data of thymine POL-C.

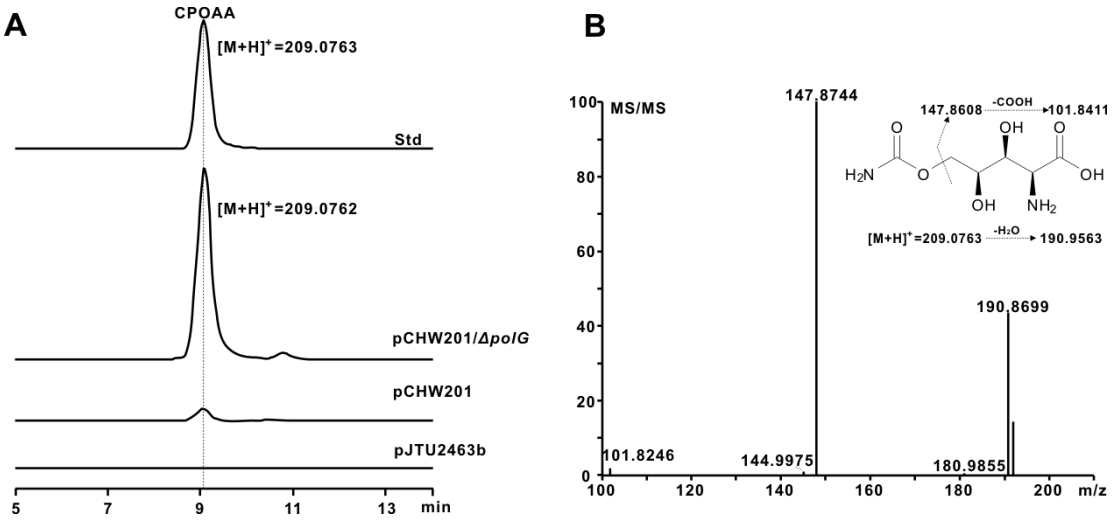


Figure S7. LC-HRMS analysis of CPOAA produced from CXR14::pCHW201/*ΔpolG* recombinant.

(A) Extract ion chromatography (EIC) analysis of the target metabolite (CPOAA) produced from CXR14::pCHW201*/ΔpolG* recombinant. (B) LC-HRMS/MS analysis of CPOAA produced from CXR14::pCHW201*/ΔpolG* recombinant and the fragmentation pattern of the authentic CPOAA standard.

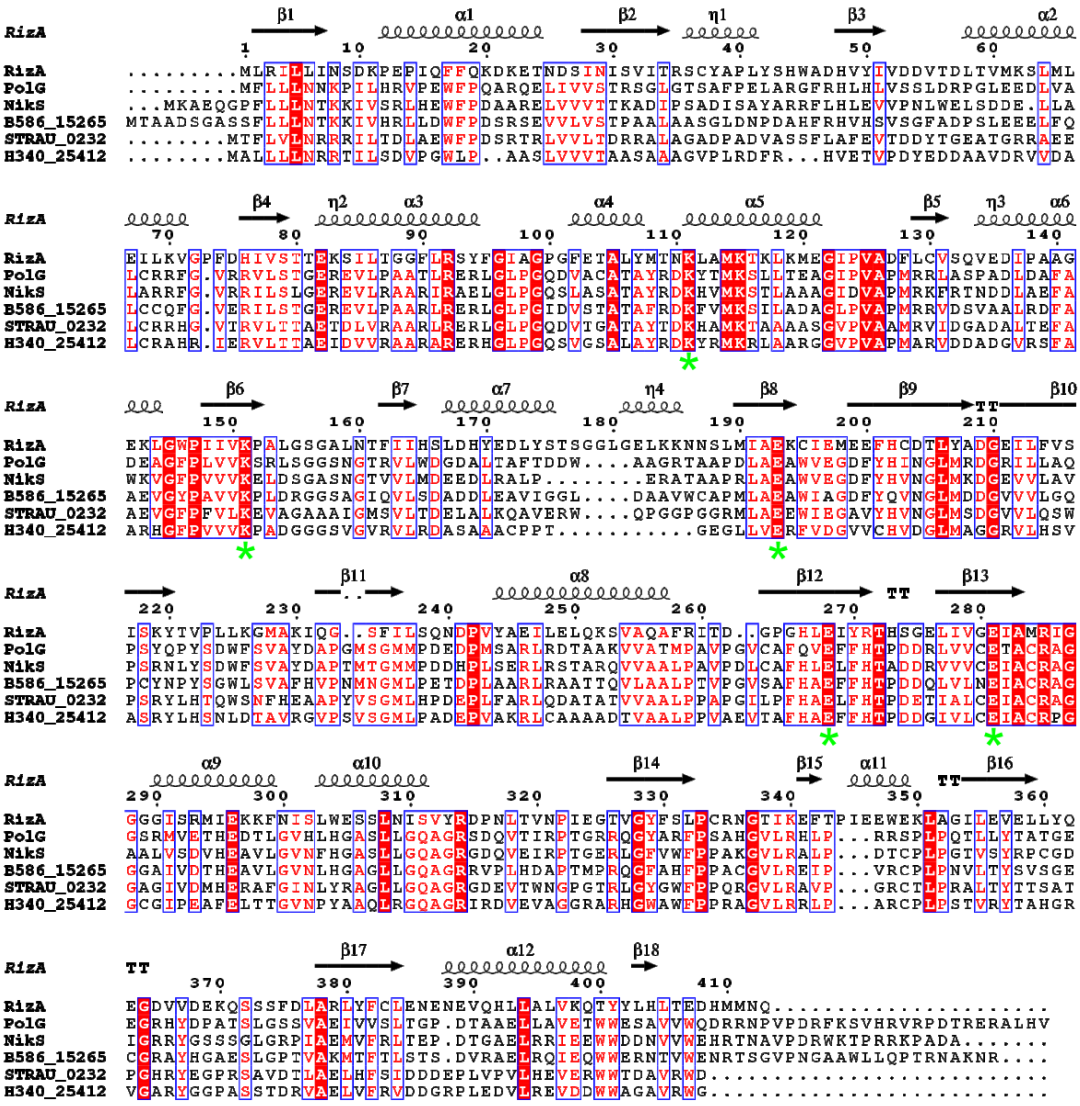


Figure S8. Bioinformatics analysis of PolG with its homologs.

Sequence alignment of PolG with its homologs using ESPript (4). Secondary structure of RizA (PDB: 4WD3) is shown on the top. RizA (GenBank: BAG72134.1) from *Bacillus subtilis subsp. subtilis*, NikS (GenBank:CAC11141.1) from *Streptomyces tendae*, STRAU_0232 (GenBank:EPH46679.1) from *Streptomyces aurantiacus* JA 4570, H340_25412 (GenBank: EME97668.1) from *Streptomyces mobaraensis* NBRC 13819. The conserved ATP-binding sites are highlighted with green asterisk.

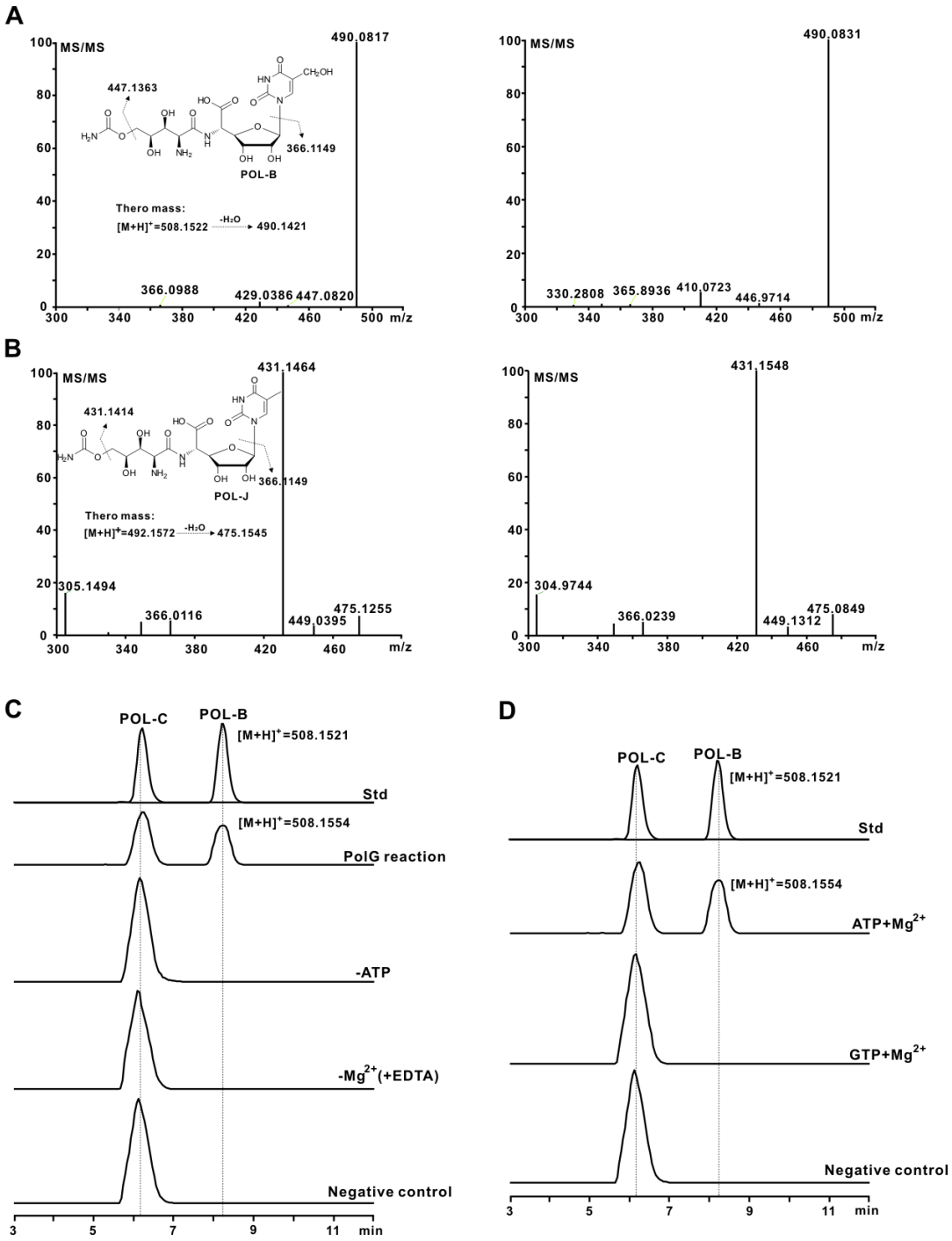


Figure S9. LC-HRMS analysis of the PolG-catalyzed reactions.

(A) LC-HRMS/MS analysis of the authentic POL-B standard and the target product (POL-B) from the PolG-catalyzed reaction using POL-C as substrate. (B) LC-HRMS/MS analysis of the authentic POL-J standard and the target product (POL-J) from the PolG-catalyzed reaction using thymine POL-C as substrate. (C) LC-MS analysis of the PolG reactions; the results indicated that both ATP and Mg^{2+} are indispensible for PolG activity. (D) LC-MS analysis of the PolG-catalyzed reactions (with ATP replaced by GTP).

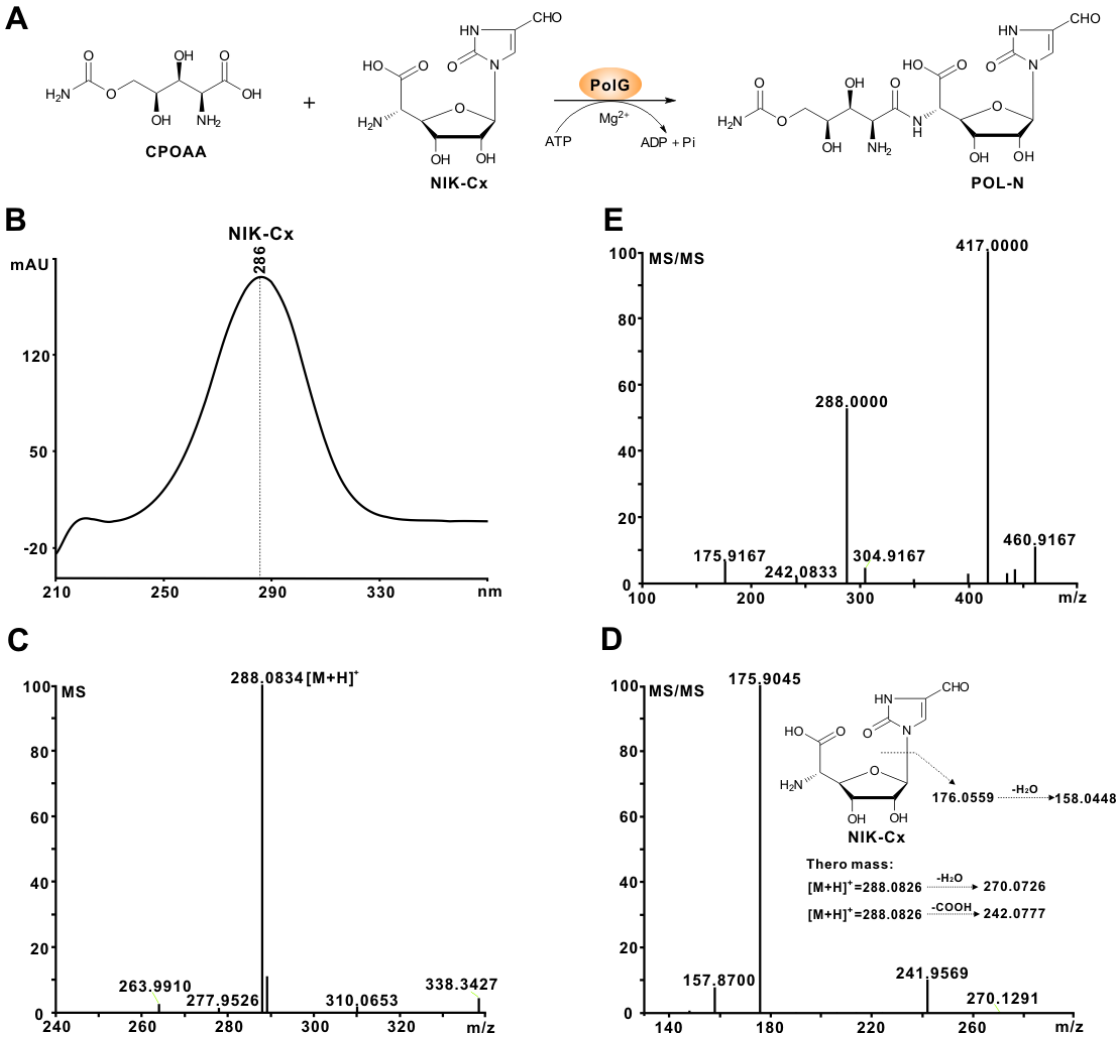


Figure S10. HPLC and LC-HRMS analysis of PolG-catalyzed reaction using NIK-Cx as substrate.

(A) Scheme of the PolG-catalyzed reaction. (B) UV spectrum of NIK-Cx. (C) LC-HRMS analysis of NIK-Cx. (D) LC-HRMS/MS analysis of NIK-Cx and the fragmentation pattern of NIK-Cx. (E) LC-HRMS/MS analysis of the authentic POL-N standard.

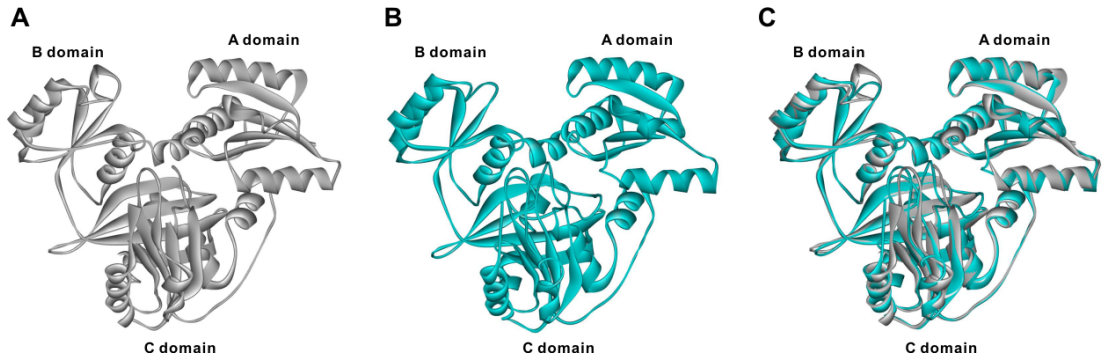


Figure S11. Structural comparison of RizA and PolG proteins.

(A) Structure of RizA (PDB: 4WD3). (B) Structure model of PolG. (C) Superimposition of PolG structure (Cyan) and RizA (Gray).

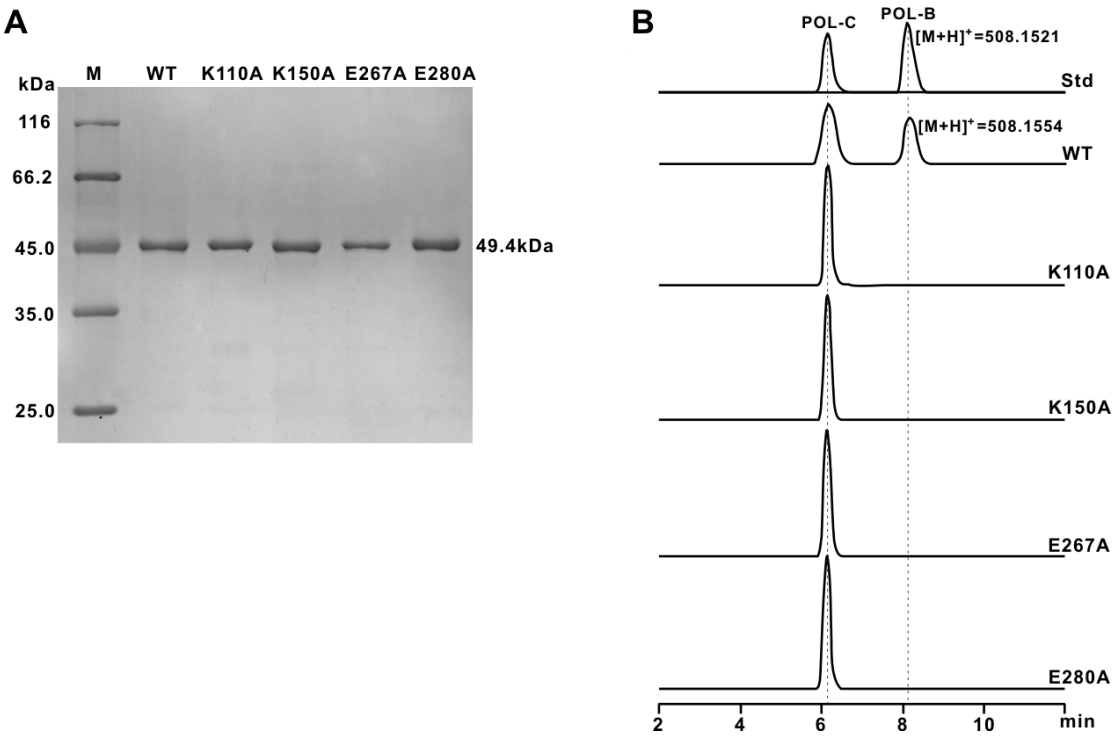


Figure S12. SDS-PAGE analysis and *in vitro* characterization of PolG variants.

(A) SDS-PAGE analysis of PolG and its variants (5 μg per lane). (B) Extract ion chromatograpy (EIC) analysis of the reactions catalyzed by PolG and its variants using POL-C as substrate.

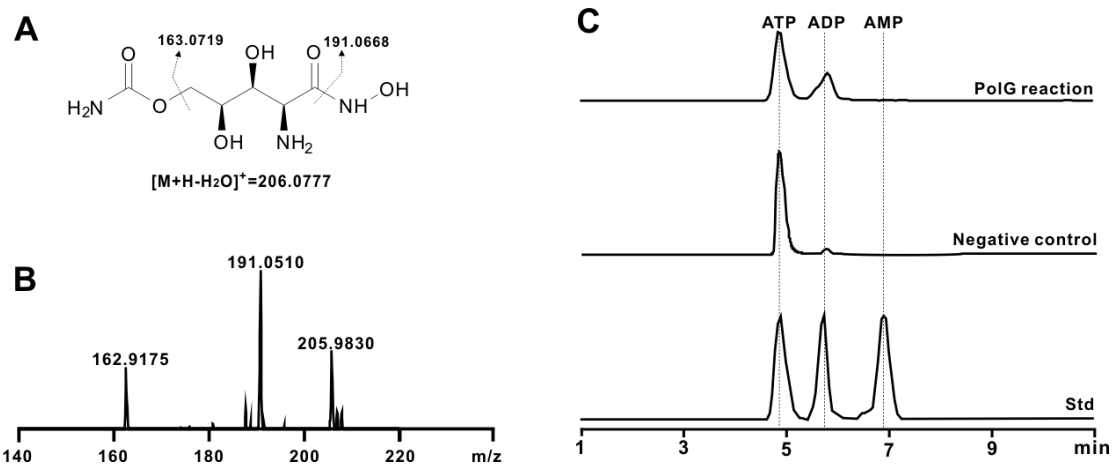


Figure S13. LC-HRMS analysis of the proposed mechanism of PolG-catalyzed reaction.

(A) Theoretical fragmentation pattern of compound **2**. (B) LC-HRMS/MS analysis of the compound **2** from the PolG-catalyzed reaction (with NH_2OH added). (C) Extract ion chromatograpy (EIC) analysis of the deduced ADP and AMP products from the PolG-catalyzed reaction.

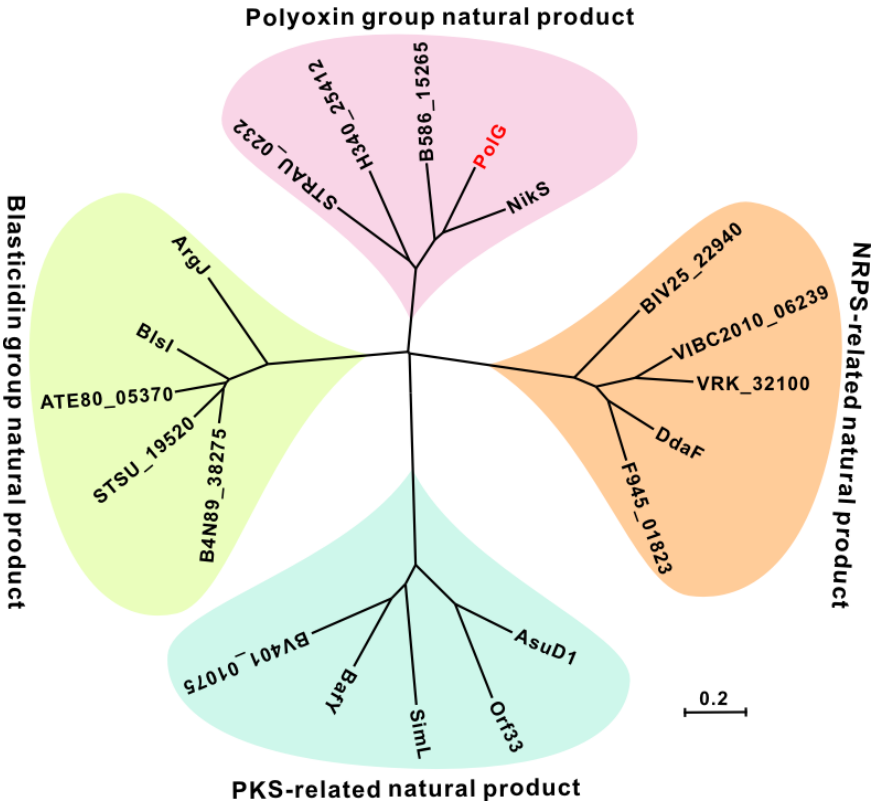


Figure S14. Phylogenetic analysis of PolG against other ATP-dependent ligases.

NikS (GenBank: CAC11141.1) from *Streptomyces tendae*, STRAU_0232 (GenBank: EPH46679.1) from *Streptomyces aurantiacus* JA 4570, H340_25412 (GenBank: EME97668.1) from *Streptomyces mobaraensis* NBRC 13819, B586_15265 (GenBank: ALL56250.1) from *Mycobacterium haemophilum* DSM 44634. DdaF (GenBank: ADN39488.1) from *Pantoea agglomerans*, F945_01823 (GenBank: EPF73944.1) from *Acinetobacter rudis* CIP 110305, VIBC2010_06239 (GenBank: EFP97961.1) from *Vibrio caribbeanicus* ATCC BAA-2122, VRK_32100 (GenBank: KUI97647.1) from *Vibrio sp*. MEBiC08052, BIV25_22940 (GenBank: OIJ94455.1) from *Streptomyces sp*. MUSC 14. BafY (GenBank: ADC79614.1) from *Streptomyces lohii*, SimL (GenBank: AAG34183.1) from *Streptomyces antibioticus*, AsuD1 (GenBank: ADI58645.1) from *Streptomyces nodosus subsp*. *asukaensis*, Orf33 (GenBank: AAX98208.1) from *Streptomyces aizunensis*, BV401_01075 (GenBank: AQA09292.1) from *Streptomyces autolyticus*. BlsI (GenBank: AAP03123.1) from *Streptomyces griseochromogenes*, ArgJ (GenBank: AGG35700.1) from *Streptomyces arginensis*, B4N89_38275 (GenBank: OPC78500.1) from *Streptomyces scabrisporus*, ATE80_05370 (GenBank: KUH39827.1) from *Streptomyces kanasensis*, STSU_19520 (GenBank: EIF90686.1) from *Streptomyces tsukubensis* NRRL18488.

3. Supplementary Tables

Table S1. Optimized sequence for PolG

ATGTTTTTACTGCTGAATAACAAGCCGATTCTGCACCGCGTTCCGGAATGGTTTCCGCAGGCCCGTCAGGAACTGATTGTG
GTGAGTACCCGTAGTGGTCTGGGTACCAGCGCCTTCCCGGAACTGGCACGTGGCTTTCGTCATCTGCACCTGGTTAGCAG
TCTGGATCGCCCTGGCCTGGAAGAGGATCTGGTGGCACTGTGCCGCCGTTTTGGCGTGCGCCGTGTTCTGAGCACCGGCG
AACGCGAGGTTCTGCCTGCAGCCACACTGCGCGAACGTCTGGGTCTGCCGGGTCAGGATGTGGCCTGTGCCACCGCCTAC
CGCGATAAATACACCATGAAAAGCCTGCTGACAGAAGCCGGCATTCCGGTTGCCCCTATGCGTCGTCTGGCCAGCCCTGC
AGACTTAGATGCATTCGCCGACGAAGCCGGCTTTCCGCTGGTGGTGAAAAGTCGCTTAAGCGGCGGCAGTAATGGTACA
CGCGTGCTGTGGGATGGCGATGCCCTGACCGCATTTACCGATGATTGGGCCGCCGGTCGTACAGCAGCACCTGATCTGGC
AGAAGCCTGGGTGGAGGGTGACTTCTACCACATCAATGGCCTGATGCGCGATGGTCGCATTCTGCTGGCACAGCCGAGT
TACCAGCCGTACAGCGATTGGTTCAGTGTGGCTTATGATGCCCCGGGTATGAGTGGCATGATGCCGGATGAAGATCCGAT
GAGCGCACGTCTGCGTGATACCGCCGCCAAAGTGGTTGCCACCATGCCGGCCGTTCCGGGTGTTTGCGCCTTTCAGGTGG
AGTTCTTTCACACCCCGGATGACCGCCTGGTTGTGTGTGAAACCGCATGTCGCGCCGGTGGTAGCCGTATGGTGGAAACC
CACGAAGATACACTGGGCGTGCATCTGCATGGTGCCAGTCTGCTGGGTCAAGCAGGCCGCAGTGATCAGGTTACCATTC
GTCCGACCGGCCGTCGCCAGGGTTATGCACGCTTTCCGAGCGCACACGGCGTTCTGCGTCATTTACCTCGCCGCAGCCCG
TTACCGCAGACACTGCTGTATACCGCAACCGGTGAAGAGGGCCGCCATTATGATCCGGCAACCAGCCTGGGCAGTAGCG
TGGCAGAAATTGTGGTGAGTCTGACAGGCCCGGATACAGCAGCCGAATTACTGGCAGTGGAAACCTGGTGGGAAAGCG
CCGTGGTTTGGCAGGATCGCCGTAATCCGGTTCCTGATCGCTTTAAAAGCGTGCATCGCGTTCGCCCGGATACACGCGAA
CGTGCATTACACGTGTAA

Table S2. PCR primers used in this study

Primers	Sequence (5'--3')
polGtgtF	GGACTGGGTACCAGCGCCTTCCCCGAGCTCGCGCGCGGCTCTAGAGCTATTCCAGAAGT
polGtgtR	CAAGAGGGTCTGCGGCAGGGGGCTTCGCCGCGGCAGGTGACTAGTCTGGATGCCGACG
polGidF	GGGCGTGACGTGCTCAGACAT
polGidR	CCCGCAGGCGAAAGGAACT
polGexF	GTCCATATGTTTTTACTGCTGAA
polGexR	GGAATTCTTACACGTGTAATGCAC
K110A-F	CGCGATGCGTACACCATGAAAAG
K110A-R	GGTGTACGCATCGCGGTAGGCGG
K150A-F	GTGGTGGCGAGTCGCTTAAGC
K150A-R	GCGACTCGCCACCACCAGCG
K188A-F	CTGGCAGCGGCCTGGGTGGAGG
K188A-R	CCAGGCCGCTGCCAGATCAGGTGC
E267A-F	CAGGTGGCGTTCTTTCACACCCCG
E267A-R	GAAAGAACGCCACCTGAAAGGCGC
E280A-F	GTGTGTGCGACCGCATGTCGCGC
E280A-R	GCGGTCGCACACACAACCAGGCGG

4. Supplementary References

1. Gust B, Challis GL, Fowler K, Kieser T, Chater KF. 2003. PCR-targeted *Streptomyces* gene replacement identifies a protein domain needed for biosynthesis of the sesquiterpene soil odor geosmin. Proc Natl Acad Sci U S A 100:1541-6.
2. Chen W, Huang T, He X, Meng Q, You D, Bai L, Li J, Wu M, Li R, Xie Z, Zhou H, Zhou X, Tan H, Deng Z. 2009. Characterization of the polyoxin biosynthetic gene cluster from *Streptomyces cacaoi* and engineered production of polyoxin H. J Biol Chem 284:10627-38.
3. Moon M, Van Lanen SG. 2010. Characterization of a dual specificity aryl acid adenylation enzyme with dual function in nikkomycin biosynthesis. Biopolymers 93:791-801.
4. Robert X, Gouet P. 2014. Deciphering key features in protein structures with the new ENDscript server. Nucleic Acids Res 42:W320-4.