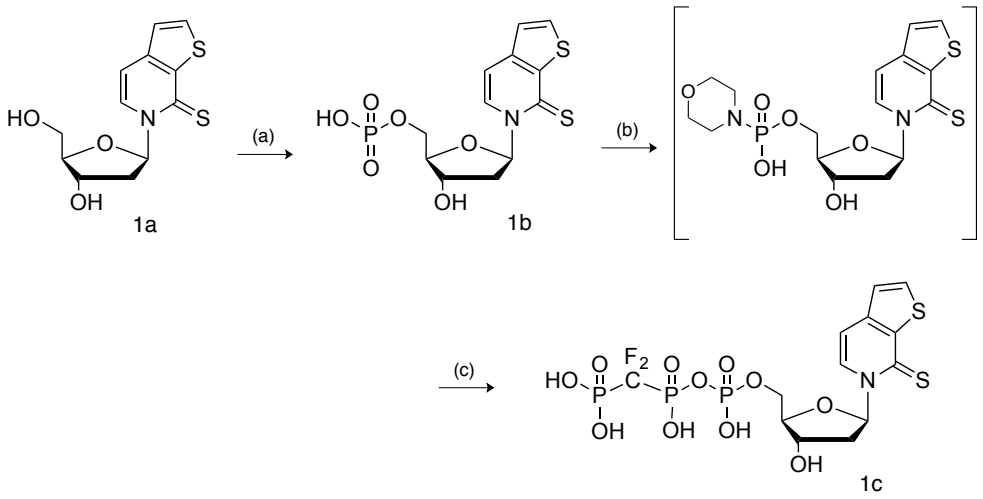
Supporting Information

Stabilized Unnatural Nucleotide Triphosphates for the *in vivo* Expansion of the Genetic Alphabet

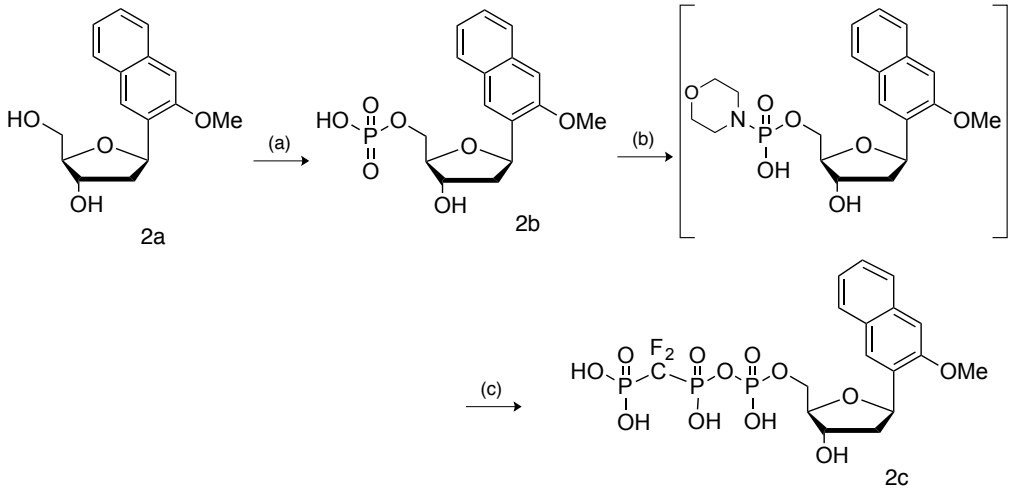
Aaron W. Feldman, Vivian T. Dien, Floyd E. Romesberg*

Department of Chemistry, The Scripps Research Institute,
10550 N. Torrey Pines Road, La Jolla, California, 92037
floyd@scripps.edu

Contents **Page**



***Scheme S1*. Synthesis of dTPT3TP^{CF2}**. (a) Proton sponge, $POCl_3$, $(MeO)_3P$, -15 °C, 3 h; (b) morpholine, DCC, *t*BuOH/H_2O (50:50), reflux, 5.5 h; (c) Bu_3N•O_3P-CF_2-PO_3H, DMSO 3 d.



***Scheme S2*. Synthesis of dNaMTP^{CF2}.** (a) Proton sponge, $POCl_3$, $(MeO)_3P$, -15 °C, 3 h; (b) morpholine, DCC, *t*BuOH/H_2O (50:50), reflux, 5.5 h; (c) Bu_3N•O_3P-CF_2-PO_3H, DMSO 3 d.

Synthetic procedures and compound characterization

General synthesis of monophosphates:

Nucleoside (0.1 mmol) and proton sponge (27.9 mg, 0.13 mmol) were placed in a reaction flask along with a magnetic stirbar and dried overnight under vacuum with P_2O_5. The solids were then taken up in freshly distilled trimethyl phosphate (0.47 mL, 4.0 mmol, dried over activated 4 Å molecular sieves) and cooled to -18 °C in a bath of salt and ice. Freshly distilled $POCl_3$ (12.1 μL, 0.13 mmol) was added dropwise, and the reaction was stirred

between -10 °C and -18 °C for 3 h. The reaction was quenched by the addition of 0.5 mL 0.5M TEAB pH 7.5. The crude reaction mixture was purified directly by ion exchange chromatography (DEAE Sephadex A-25; GE Healthcare) with a linear gradient of 0 to 1.2 M TEAB pH 7.5.

General synthesis of β,γ-CF_2 nucleotide triphosphates:

A solution of monophosphate (27.8 μmol) and morpholine (14.5 μL, 166.8 μmol) in 50% *t*BuOH/H_2O (1.0 mL) was stirred at room temperature for 15 min before heating to a reflux. A solution of *N,N'*-dicyclohexylcarbodiimide (34.4 mg, 166.6 μmol) in *t*BuOH (0.2 mL) was added dropwise to the refluxing reaction, and the resulting mixture was allowed to reflux an additional 5.5 h, monitoring reaction progress by ^{31}P NMR. Upon cooling to room temperature, solvent was removed *in vacuo* and residue was redissolved in H_2O. Solid side products were filtered over celite, rinsing with H_2O. The combined aqueous solution was extracted with diethyl ether three times. The resulting aqueous layer was concentrated (Savant SpeedVac concentrator) and dried overnight on high vacuum.

The tributylammonium salt of difluoromethylene bisphosphonate (63.8 mg, 83.1 μmol) was dissolved in anhydrous DMSO (0.5 mL). Dried morpholidate (16.7 μmol) was dissolved in anhydrous DMSO (0.5 mL), and the resulting solution was added to the reaction vial dropwise. Upon sealing under argon, the reaction was allowed to stir at room temperature for 3 days, monitoring reaction progress by HPLC. The crude reaction mixture was purified by ion exchange chromatography (DEAE Sephadex A-25) with a linear gradient of 0 to 1.2 M TEAB pH 7.5. Upon concentration (SpeedVac), any remaining inorganic phosphonate was removed via further purification by reverse-phase (C18) HPLC (linear gradient of 5% to 35% acetonitrile in 0.1 M TEAB, pH 7.5), and then concentrated to dryness (SpeedVac).

Compound 1b.

^{1}H NMR (600 MHz, Deuterium Oxide) δ 8.37 (d, *J* = 7.2 Hz, 1H), 7.99 (d, *J* = 5.3 Hz, 1H), 7.44 (d, *J* = 7.2 Hz, 1H), 7.37 (d, *J* = 5.4 Hz, 1H), 7.33 (t, *J* = 6.4 Hz, 1H), 4.62 (dt, *J* = 6.3, 4.1 Hz, 1H), 4.32 (m, 1H), 4.17 (dddd, *J* = 40.6, 11.6, 4.9, 3.5 Hz, 2H), 2.76 (ddd, *J* = 14.1, 6.4, 4.2 Hz, 1H), 2.30 (dt, *J* = 14.1, 6.4 Hz, 1H). ^{13}C NMR (151 MHz, D_2O) δ 168.19, 143.24, 140.74, 138.50, 130.52, 124.53, 111.16, 90.45, 86.04, 85.98, 69.87, 63.49, 63.46, 58.38, 39.88. ^{31}P NMR (162 MHz, D_2O) δ 1.43 (s). (MALDI-TOF$^-$, matrix: 9-aminoacridine) (m/z): [M-H]- calcd for $C_{12}H_{13}NO_6PS_2^-$, 361.99; found, 361.99. (33.3 μmol, 33.3% yield)

Compound 2b.

^{1}H NMR (600 MHz, Deuterium Oxide) δ 7.97 (s, 1H), 7.88 (d, *J* = 8.2 Hz, 1H), 7.80 (d, *J* = 8.2 Hz, 1H), 7.49 (t, *J* = 7.5 Hz, 1H), 7.41 (t, *J* = 7.5 Hz, 1H), 7.30 (s, 1H), 5.52 (dd, *J* = 10.2, 5.9 Hz, 1H), 4.53 – 4.46 (m, 1H), 4.22 – 4.16 (m, 1H), 4.03 – 3.89 (m, 5H), 2.39 (ddd, *J* = 13.6, 5.9, 2.1 Hz, 1H), 2.13 (ddd, *J* = 13.4, 10.1, 5.8 Hz, 1H). ^{13}C NMR (151 MHz, D_2O) δ 154.55, 133.30, 130.27, 127.33, 126.22, 125.89, 124.80, 123.71, 105.25, 84.96, 75.33,

72.49, 64.54, 54.97, 48.12, 41.71. ^{31}P NMR (162 MHz, D_2O) δ 2.50. (MALDI-TOF$^-$, matrix: 9-aminoacridine) (m/z): [M-H]- calcd for $C_{16}H_{18}O_7P^-$, 353.08; found, 353.02. (17.8 μmol, 17.8% yield, two steps)

Compound 1c.

^{19}F NMR (376 MHz, DMSO-d_6) δ -120.32 (t, J = 84.4 Hz). ^{31}P NMR (162 MHz, Deuterium Oxide) δ 4.22 – 1.26 (m), -4.70 – -6.55 (m), -10.94 (d, J = 29.2 Hz). HRMS calcd for $C_{13}H_{16}F_2NO_{11}P_3S_2$ $[MH]^+$ 557.9418; found, 557.9418. (5.7 μmol, 25.9% yield, two steps)

Compound 2c.

^{19}F NMR (376 MHz, DMSO-d_6) δ -120.19 (t, J = 84.4 Hz). ^{31}P NMR (162 MHz, Deuterium Oxide) δ 4.22 – 2.61 (m), -4.08 – -5.38 (m), -10.67 (d, J = 30.2 Hz). HRMS calcd for $C_{17}H_{21}F_2O_{12}P_3$ $[MH]^+$ 549.0287; found, 549.0287. (3.6 μmol, 15.2% yield, two steps)

Degradation of nucleotide triphosphates

For analysis of nucleotide degradation, a 10-μL injection was eluted by a linear gradient of 0–10% B (for d**TPT3** samples) or 0–40% B (for d**NaM** samples) over 40 min at a flow rate of 1 mL/min. Buffer A: 95% 0.1 M TEAB, pH 7.5, 5% acetonitrile. Buffer B: 20% 0.1 M TEAB, pH 7.5, 80% acetonitrile. Elution was monitored by UV/Vis absorption at 260, 280, 326, 365, 550 nm.

Gel-based kinetic assay

Primer oligonucleotides (4 μM of **P1** for insertion of d**TPT3** triphosphates, **P2** for insertion of d**NaM** triphosphates; see Table S1 for sequences) were 5'-radiolabeled with T4 polynucleotide kinase (New England Biolabs) and [γ-^{32}P]-ATP (Perkin-Elmer) and annealed to template oligonucleotides (**O1** for insertion of d**TPT3** triphosphates, **O2** for insertion of d**NaM** triphosphates) by heating to 95 ºC for 5 minutes, followed by slow cooling to room temperature. An aliquot of a 0.04–5.6 μM stock solution of unnatural triphosphate and a 5 nM Klenow Fragment polymerase solution were mixed in equal volumes (2.5 μL). The dXTP-polymerase (2×) and primer:template (2×, 80 nM, in 2× Klenow reaction buffer) solutions were incubated at 37 ºC for at least 2 minutes. Reactions were initiated by combining the two solutions in equal volumes (5 μL), incubating at 37 ºC for 20 s–5 min, and quenching with a 20 μL solution of loading dye (95% formamide, 20 mM EDTA, and sufficient amounts of bromophenol blue and xylene cyanol).

In vivo UBP replication

Assembly of UBP-containing plasmids. Inserts for Golden Gate assembly bearing the UBP were generated by PCR of sequence **O1** (0.025 ng per 50 μL reaction) using primers **P3** and **P4** and conditions described previously,[2] under

the following thermocycling protocol (times denoted as mm:ss): [96 °C 0:30 | 25 × [96 °C 0:30 | 47 °C 0:30 | 68 °C 4:00]].

Golden Gate assembly reactions were prepared by combining destination plasmid pUCX2 (1 μg),[1] PCR insert (3:1 insert:plasmid molar ratio), T4 DNA ligase (533 U), BsaI-HF (53.3 U), and ATP (1 mM) in 1× NEB CutSmart buffer (final volume 80 μL). The reactions were thermocycled under the following conditions: [37 °C 20:00 | 40× [37 °C 5:00 | 16 °C 10:00 | 22 °C 5:00] | 37 °C 20:00 | 50 °C 15:00 | 70 °C 30:00]. Following thermocycling, T5 exonuclease (13.3 U) and additional BsaI-HF (26.6 U) were added, and the reactions were incubated at 37 °C for an additional hour. Assembled plasmids were purified on a Zymo-Spin I column, and quantified fluorometrically using a Qubit fluorometer (Thermo-Fisher). Refer to Table S1 for list of sequences.

Transformation and replication of UBP-containing plasmids. The *E. coli* SSO was made electrocompetent as described previously,[1] and an aliquot (50 μL) was transformed with 2 ng pINF plasmid. Upon electroporation, cells were immediately diluted with media supplemented with chloramphenicol (950 μL). An aliquot of diluted cells (10 μL) was immediately further diluted 10-fold with media supplemented with the same media, but further supplemented with d**NaM**TP (250 μM, 1000 μM, or 2000 μM) and d**TPT3**^{X} (3 μM or 0.75 μM, X denoting either the unmodified triphosphate or the β,γ-CF_2 analog) for plasmid recovery at concentrations corresponding to outgrowth conditions. Transformed cells were incubated at 37 °C for 1 h before diluting each sample (10 μL) 10-fold into fresh media supplemented with ampicillin for pINF selection, as well as fresh unnatural triphosphates (at conditions identical to recovery for each sample). Samples were then further incubated at 37 °C, monitoring for growth. For experiments testing "permissive" conditions, cells were harvested by pelleting (9,000 r.c.f., 10 min) at an OD_{600} of ~0.7. To test more stringent conditions, cells were allowed to grow overnight (18 h) into deep saturation (OD_{600} ~ 9.0) before pelleting.

Analysis of pINF plasmids by biotin shift PCR. Plasmids were isolated for analysis using commercial miniprep kits (ZR Plasmid Miniprep Classic, Zymo Research), and analyzed for UBP retention by biotin shift PCR, as described previously.[1,2] PCR conditions were as follows: plasmid miniprep (0.5 μL, diluted 5-fold for samples grown to saturation), dNTPs (400 μM), 1× SYBR Green, $MgSO_4$ (2.2 mM), primers (**P5**/**P6,** 1 μM each), d**5SICS**TP and d**MMO2**^{Bio}TP (65 μM each), OneTaq DNA polymerase (0.018 U/μL), and DeepVent DNA polymerase (0.007 U/μL) in 1× OneTaq standard reaction buffer (final volume 15 μL). Thermocycling conditions were as follows: [20 × [95 °C 0:15 | 52 °C 0:15 | 68 °C 4:00]]. An aliquot (1 μL) of each reaction was mixed with a solution of streptavidin (2.5 μL, 2 μg/μL, Promega), and incubated at room temperature for 10 min. Streptavidin-bound amplicon was then resolved from unbound amplicon on a 6% polyacrylamide (29:1 acrylamide:bis-acrylamide) TBE gel, at 120 V for 25 min. Gels were then stained with 1× SYBR Gold (Thermo Fisher) and imaged using a Molecular Imager Gel Doc XR+ equipped with a 520DF30 filter (Bio-Rad).

UBP retention was measured by densitometric analysis of the gels (Quantity One 1-D Analysis Software, Bio-Rad) from the biotin shift assay and calculation of a percent raw shift, which equals the intensity of the streptavidin-shifted band divided by the sum of the intensities of the shifted and unshifted bands. Reported UBP retentions are percent raw shift for a given sample normalized to the percent raw shift of the input pINF plasmid used.

References

(1) Zhang, Y.; Lamb, B.; Feldman, A. W.; Zhou, A. X.; Lavergne, T.; Li, L.; Romesberg, F. E. *Proc. Natl. Acad. Sci. USA* **2017**, *in press*.

(2) Malyshev, D. A.; Dhami, K.; Lavergne, T.; Chen, T.; Dai, N.; Foster, J. M.; Correa, I. R., Jr.; Romesberg, F. E. *Nature* **2014**, *509*, 385–388.

Table S1. Oligonucleotides and plasmids used in this study.

Name	*Sequence (5' — 3')*
O1	CTGTTTCCTGTGTGAAATTGTTATCCGCTCACA-d**NaM**-TTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCC
O2	GGCTTTACACTTTATGCTTCCGGCTCGTATGTTGTGTGGAA-d**TPT3**-TGTGAGCGGATAACAATTTCACACAGGAAACAG
P1	TCCGGCTCGTATGTTGTGTGGAA
P2	TGTGAAATTGTTATCCGCTCACA
P3	ATGGGTCTCCAGTGGCTGTTTCCTGTGTGAAATTGTTATCCGC
P4	ATGGGTCTCTTCGTTGGCTTTACACTTTATGCTTCCGGC
P5	CTGTTTCCTGTGTGAAATTGTTATCC
P6	GGCTTTACACTTTATGCTTCCG

pUCX2 (3174 bp)

AACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCC
CCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGC
GAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCCAATACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATTCATTAATGC
AGCTGGCACGACAGGTTTCCCGACTGGAAAGCGGGCAGTGAGCGCAACGCAATTAATGTAAGTTAGCTCACTCATTAG
GCACCCCACTATGACCATGATTACGCCAAGCTTGCATGCCTGCAGGTCGACTCTAGAGGATCCCCGGGTACCGAGCTC
GAATTCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATCGCCTTGCAGCACA
TCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCGCACCGATCGCCCTTCCCAACAGTTGCGCAGCCTGAATGG
CGAATGGCGCCTGATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTCACACCGCATATGGTGCACTCTCAGTAC
AATCTGCTCTGATGCCGCATAGTTAAGCCAGCCCCGACACCCGCCAACACCCGCTGACGCGCCCTGACGGGCTTGTCT
GCTCCCGGCATCCGCTTACAGACAAGCTGTGACCGTCTCCGGGAGCTGCATGTGTCAGAGGTTTTCACCGTCATCACC
GAAACGCGCGAGACGAAAGGGCCTCGTGATACGCCTATTTTTATAGGTTAATGTCATGATAATAATGGTTTCTTAGAC
GTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCC
GCTCATGAGACAATAACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGT
CGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGCTGGTGAAAGTAAAAGATGC
TGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCC
CGAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTATTGACGCCGGGCA
AGAGCAACTCGGTCGCCGCATACACTATTCTCAGAATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTAC
GGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCATAACCATGAGTGATAACACTGCGGCCAACTTACTTCTGAC
AACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGGGATCATGTAACTCGCCTTGATCGTTGGGA
ACCGGAGCTGAATGAAGCCATACCAAACGACGAGCGTGACACCACGATGCCTGTAGCAATGGCAACAACGTTGCGCAA
ACTATTAACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAATAGACTGGATGGAGGCGGATAAAGTTGCAGG
ACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGCTCTCGCGG
TATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTAT
GGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTAACTGTCAGACCAAGTTTACTC
ATATATACTTTAGATTGATTTAAAACTTCATTTTTAATTTAAAAGGATCTAGGTGAAGATCCTTTTTGATAATCTCAT
GACCAAAATCCCTTAACGTGAGTTTTCGTTGAGACCATTCTCACCAATAAAAAACGCCCGGCGGCAACCGAGCGTTCT
GAACAAATCCAGATGGAGTTCTGAGGTCATTACTGGATCTATCAACAGGAGTCCAAGCGAGCTCTCAGTCCTGCTCCT
CGGCCACGAAGTGCACGCAGTTGCCGGCCGGGTCGCGCAGGGCGAACTCCCGCCCCCACGGCTGCTCGCCGATCTCGG
TCATGGCCGGCCCGGAGGCGTCCCGGAAGTTCGTGGACACGACCTCCGACCACTCGGCGTACAGCTCGTCCAGGCCGC
GCACCCACACCCAGGCCAGGGTGTTGTCCGGCACCACCTGGTCCTGGACCGCGCTGATGAACAGGGTCACGTCGTCCC
GGACCACACCGGCGAAGTCGTCCTCCACGAAGTCCCGGGAGAACCCGAGCCGGTCGGTCCAGAACTCGACCGCTCCGG
CGACGTCGCGCGCGGTGAGCACCGGAACGGCACTGGTCAACTTGGCCATGGTTTAGTTCCTCACCTTGTCGTATTATA
CTATGCCGATATACTATGCCGATGATTAATTGTCAACGACGGGTCTCCCACTGAGCGTCAGACCCCGTAGAAAAGATC
AAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTG
GTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACT
GTTCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATC
CTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAG
GCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATAC
CTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTC
GGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTC
TGACTTGAGCGTCGATTTTTGTGTGCTCGTCAGGGGGGCGGAGCCTATGGAAA