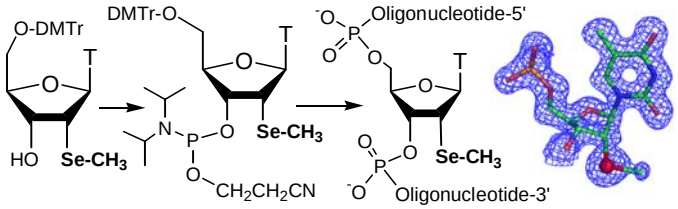
Table of Content



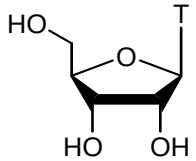
Synthesis of a 2'-Se-Thymidine Phosphoramidite and Its Incorporation into Oligonucleotides for Crystal Structure Study

Jia Sheng, Jiansheng Jiang, Jozef Salon, Zhen Huang*

SUPPORTING MATERIAL

General Section

Most solvents and reagents were purchased from Sigma, Fluka, or Aldrich (p.a.) and used without purification unless mentioned otherwise. Triethylamine (TEA) was dried over KOH (s) and distilled under argon. When necessary, solid reagents were dried under high vacuum. Reactions with compounds sensitive to air or moisture were performed under argon. Solvent mixtures are indicated as volume/volume ratios. Thin layer chromatography (TLC) was run on Merck 60 F_{254} plates (0.25 mm thick; R_f values in the text are for the title products), and visualized under UV-light or by a Ce-Mo staining solution (phosphomolybdate, 25 g; $Ce(SO_4)_2 \cdot 4H_2O$, 10 g; H_2SO_4, 60 mL, conc.; H_2O, 940 mL) with heating. Preparative TLC was performed using Merck 60 F_{254} pre-coated plates (2 mm thick). Flash chromatography was performed using Fluka silica gel 60 (mesh size 0.040-0.063 mm) using a silica gel:crude compound weight ratio of ca. 30:1. ^{1}H, ^{13}C and ^{31}P-NMR spectra were recorded using Bruker-300 or 400 (300 or 400 MHz). All chemical shifts (δ) are in ppm relative to tetramethylsilane and all coupling constants (J) are in Hz. High resolution mass spectrum (HRMS) analysis was performed at Scripps Center for Mass Spectrometry, California.

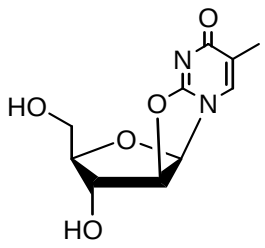


5-Methyluridine or ribothymidine (2)

Thymine (3.80 g, 30.23 mmol) and 1-O-acetyl-2,3,5-tri-O-benzoylribose (12.7 g, 25.4 mmol) were suspended in dry acetonitrile (350 mL), followed by adding trimethylsilyl

chloride (3.2 mL, 25.9 mmol), hexamethyldisilazane (5.3 mL, 25.4 mmol), and tin chloride (2.97 mL, 25.4 mmol), then the suspension was heated to reflux for 1.5 h. The solution was concentrated in vacuum (approximately 20 mL) and dissolved in methylene chloride (100 mL). The organic solution was washed with water (20 mL), saturated sodium bicarbonate (2 X 20 mL), and brine (20 mL) , dried over $MgSO_4$ and concentrated to give 2',3',5'-tri-O-benzoyl-5-methyluridine as a white solid (14.36 g). ^{1}H-NMR (400 MHz, $CDCl_3$) δ:1.61 (s, 3H, 5-CH_3), 4.72-4.78 (m, 1H, H-2'), 4.68 and 4.92 (2x dd, J_1 =2.4 Hz, J_2 = 12 Hz, 2H, H-5'), 5.78 (t, J =6.0 Hz, 1H, H-3'), 4.92-4.97 (m, 1H, H-4'), 6.43 (d, J =6.2 Hz, 1H, H-1'), 7.18-8.14 (m, 16H, aromatic H and H-6). The white solid (14.36 g, 25.2 mmol), without further purification, was dissolved in methanol (350 mL). Sodium methoxide (8.16 g, 151.2 mmol) was then added, and the reaction mixture was stirred at room temperature overnight. The solution was neutralized with Dowex 50 X 8-4200 ion-exchange resin (approximately 5 g, monitored by wet pH paper), concentrated, and dissolved in water (100 mL), which then be lyophilized after extracting with ethyl ether (2 X 150 mL). The lyophilized sample was recrystallized from absolute ethanol to give compound **2** as a white solid (5.76 g, 88% in two steps).

^{1}H-NMR (400 MHz, D_2O) δ:1.78 (s, 3H, 5-CH_3), 3.72 and 3.83 (2 X dd, J_1 =4.0 Hz, J_2 = 12.8 Hz, 2H, H-5'), 4.02 (m, 1H, H-4'), 4.15 (t, J =2.6 Hz, 1H, H-3'), 4.25 (t, J =2.5 Hz, 1H, H-2'), 5.82 (d, J =6.0 Hz, 1H, H-1'), 7.60 (s, 1H, H-6). ^{1}H-NMR spectrum is identical to the literature (18).

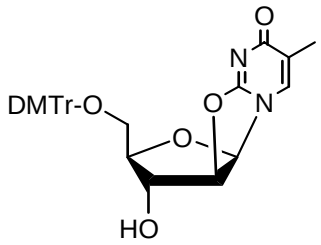


2,2'-Anhydro[1-(β-D-arabinofuranosyl)-5-methyluridine]

Diphenylcarbonate (7.0 g, 0.033 mol) and sodium bicarbonate (0.16 g, 1.8 mmol) were added to the DMF (25 mL) solution of 5-methyluridine (5.6 g, 0.022 mol). The mixture was heated to reflux, and carbon dioxide was generated. After 1.5 hour, the darkened solution was concentrated to approximately 4 mL under reduced pressure. The black residue was poured

into diethylether (200 mL) under vigorous stirring, and a brown gum product was formed. The ether was decanted and the residue was dissolved in a minimum amount of methanol. The crude product was precipitated again with fresh ether (150 mL) to yield a stiff light brown gum. The ether was decanted and the gum was dried under high vacuum over night to give a solid that was easily crushed to give a light tan powder (4.4 g, 85% crude yield). The material could be used directly in the following step without further purification.

^{1}H-NMR (400 MHz, d-DMSO) δ: 1.79 (s, 3H, 5-CH_3), 3.12-3.25 (m, 2H, H-5'), 4.05-4.11 (m, 1H, H-4'), 4.35-4.41 (m, 1H, H-3'), 5.19 (d, *J* = 5.6 Hz, 1H, H-2'), 6.29 (d, *J* =6.2 Hz, 1H, H-1'), 7.76 (s, 1H, H-6). ^{1}H-NMR spectrum is identical to the literature (19).



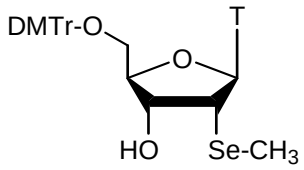
2,2'-Anhydro-1-[2'-deoxy-5'-O-(4,4-dimethoxytrityl)-β-D-arabinofuranosyl-5-methyluridine] (3)

The starting material, 2,2'-anhydro[1-(β-D-arabinofuranosyl)-5-methyluridine] (4.0 g non-purified after the cyclization step), was co-evaporated with dry pyridine (3 X 5 mL) and then dissolved in dry pyridine (25 mL). A first half of dimethoxytrityl chloride (2.36 g, 6.95 mmol) was added and the solution was stirred at room temperature for one hour. Then a second half of dimethoxytrityl chloride was added and the reaction was completed after stirred for additional one hour, followed by quenching with methanol (4 mL). The solution was evaporated and co-evaporated with CH_3CN (2 X 5 mL). The residue was then dissolved in $CHCl_3$ (40 mL) and the solution was washed with saturated sodium bicarbonateand (2 X 15 mL) and saturated sodium chloride (2 X 15 mL). The organic phase was dried over Na_2SO_4 (s) and evaporated after the filtration of the drying agent. The obtained crude product (6.9 g) was purified on a silica gel column equilibrated with methylene chloride, and the pure product was eluted with a methanol/CH_2Cl_2 gradient (methanol in CH_2Cl_2, 0-3%). The fractions containing the product were combined, evaporated and dried on high vacuum overnight to yield the pure product (**3**, 7.80 g) in 87% yield (calculated from ribothymidine).

^{1}H-NMR (400 MHz, $CDCl_3$) δ: 1.92 (s, 3H, 5-CH_3), 2.94-3.09 (m, 2H, H-5'), 3.79 (s, 6H, 2x CH_3O), 4.36-4.42 (m, 1H, H-4'), 4.49-4.51 (m, 1H, H-3'), 5.26 (dd, J_1 =1.4 Hz, J_2 = 5.8 Hz, 1H, H-2'), 6.15 (d, J =6.0 Hz, 1H, H-1'), 6.76-6.80 (m, 4H, aromatic), 7.17 - 7.35 (m, 10H, H-6 and 9 aromatic protons).

^{13}C-NMR ($CDCl_3$) δ: 55.23 (OCH_3), 63.01 (C-5'), 75.63 (C-3'), 86.23 (Ar-C) 87.41 (C-4'), 89.09 (C-2'), 90.40 (C-1') , 118.75 (C-5), 113.21, 126.95, 127.90, 129.84, 135.45, 144.37, 158.50, 159.27 (Ar-C), 131.22 (C-6), 159.27 (C-2), 172.86 (C-4).

HRMS (ESI-TOF): molecular formula, $C_{31}H_{30}N_2O_7$; $[M+H]^+$:543.2051 (calc.543.2053).



2'-methylseleno-5'-O-(4,4'-dimethoxytrityl)-5-methyluridine (4)

$NaBH_4$ (0.56 g, 14.72 mmol) was placed in a 50 ml round ask and suspended in dry THF (15 mL) under vigorous stirring. Dimethyl diselenide ($CH_3SeSeCH_3$, 1.47 ml, 14.72 mmol) was slowly injected in an ice-water bath under dry argon. Anhydrous ethanol (2 ml) was added dropwisely until gas bubbles started in the yellow mixture. The reaction mixture turned white or light yellow after 30 min, then a solution of **3** (4.0 g, 7.36 mmol) in THF (8 ml) was injected into the flask. The ice-water bath was removed after the addition, and the reaction was heated at 50 ºC and monitored by TLC (5% CH_3OH in CH_2Cl_2, product Rf = 0.47). The reaction was completed in 3 h, and the solvent was removed under reduced pressure. The residue was dissolved in methylene dichloride (20 mL) and the solution was washed with NaCl solution (sat., 3 X 10 mL). The organic layer was dried over anhydrous $MgSO_4$, followed by the filtration and solvent evaporation. The residue was then purified on a silica gel column (equilibrated with CH_2Cl_2) and the column was eluted with a methanol/methylene chloride gradient (CH_3OH in CH_2Cl_2, 0-3%) to afford the pure foamy product (**4**, 4.24 g) in 82 % yield.

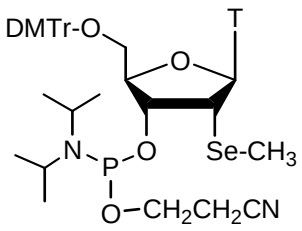
^{1}H-NMR (400 MHz, $CDCl_3$) δ: 1.45 (s, 3H, 5-CH_3), 2.12 (s, 3H, CH_3Se), 2.82-2.87 (m, 1H, OH), 3.39-3.60 (dd, J_1 =2.0 Hz, J_2 = 10.6Hz, 2H, H-5'), 3.62-3.66 (dd, J_1 =5.2 Hz, J_2

= 8.8 Hz , 1H, H-2'), 3.82 (s, 6H, CH_3O), 4.23-4.25 (m, 1H, H-4'), 4.38-4.40 (dd, J_1 =2.0 Hz, J_2 = 5.2 Hz, 1H, H-3'), 6.26 (d, J = 8.8 Hz, 1H, H-5), 6.81-6.88 (m, 4H, aromatic), 7.25-7.43 (m, 9H, aromatic), 7.62 (d, J = 1.2 Hz, 1H, H-6), 8.62 (br, 1H, NH).

^{13}C-NMR ($CDCl_3$) δ: 4.69 ($SeCH_3$), 11.69 (5-CH_3), 50.17 (C-2'), 55.30 (OCH_3), 63.78 (C-5'), 72.28 (C-3'), 84.52 (C-4'), 87.08 (C-1'), 87.25 (Ar-C), 111.79 (C-5), 113.33, 127.31, 128.14, 130.11, 135.12 (Ar-C), 135.19 (C-6), 144.18 (Ar-C), 150.46 (C-2), 158.82 (Ar-C), 163.49 (C-4).

HRMS (ESI-TOF): molecular formula, $C_{32}H_{34}N_2O_7Se$; $[M-H]^-$: 637.1446

(calc.637.1453).



3'-O-(2-cyanoethyl-N,N-diisopropylphosphoramidite)-2'-methylseleno-5'-O(4,4'-dimethoxytrityl)-5-methyluridine (5)

The completely dried starting material **4** (0.9 g , 1.41 mmol) was dissolved in dry CH_2Cl_2 (5 ml) and the 2-cyanoethyl *N*,*N*-diisopropylchlorophosphoramidite (2.0 g) was slowly added into the solution, followed by a few drops of N,N-diisopropylethylamine. The reaction mixture was stirred at room temperature under dry nitrogen for 3 h at room temperature until all the starting materials were converted. Then, the reaction mixture was quenched with $NaHCO_3$ (10 ml, sat.), and extracted with CH_2Cl_2 (3 X 10 ml). The combined organic layer was washed with NaCl (3 X 10 ml, sat.) and dried over anhydrous $MgSO_4$, followed by filtration and solvent evaporation. The crude product was re-dissolved in CH_2Cl_2 (2 ml), and precipitated in petroleum ether (200 ml) under vigorous stirring, generating white powder. The petroleum ether solution was then decanted carefully. The crude product was further purified on a flash column packed with aluminum oxide (activated, neutral, 150 mesh, 58 Å) by eluding with a solution of 30%

EtOAc in CH_2Cl_2 containing 1% triethylamine. The desired pure product was obtained as an white foam (1.06 g, 90 %).

Alternative synthesis:

The starting material **4** (2.4 g , 3.76 mmol) and the catalyst 5-Benzylthio-1H-Tetrazole (0.06 g, 0.3 mmol) were placed in a 100 ml round flask and dried on high vacuum over night. Dry CH_2Cl_2 (45 ml) was added followed by adding 2-cyanoethyl N,N,N,N-tetraisopropyl phosphane (1.36 g, 4.5 mmol) and a few drops of N,N-diisopropylethylamine. The reaction mixture was stirred at 0°C under dry argon for 30 min, followed by stirring for 2 h at room temperature until most of the starting material was converted. The reaction mixture was then quenched with $NaHCO_3$ (10 ml, sat.) by stirring for 15 min, and extracted with CH_2Cl_2 (3 X 10 ml). The combined organic layer was washed with NaCl (3 X 10 ml, sat.) and dried over anhydrous $MgSO_4$ for 15 min, followed by the filtration and solvent evaporation. The crude product was re-dissolved in CH_2Cl_2 (3 ml), and precipitated in petroleum ether (200 ml) under vigorous stirring, generating white powder. The petroleum ether solution was then decanted carefully. The crude product was dissolved in CH_2Cl_2 and loaded on a column packed with aluminum oxide (activated, neutral, 150 mesh, 58 Å) and equilibrated with CH_2Cl_2. The column was then eluded with a solution of 30% EtOAc in CH_2Cl_2 containing 1% triethylamine. The product fractions were collected and evaporated under the reduced pressure. The purified compound was re-dissolved in 3 ml of CH_2Cl_2, followed by the precipitation again in petroleum ether (200 mL, as indicated above) to yield a white foamy product 5 as the diastereomer mixture (2.6 g) in 83 % yield.

^{1}H-NMR (400 MHz, $CDCl_3$) δ: 0.89-1.42 (m, 24H, 8xCH_3-ipr), 1.40 (2x s, 6H, 2x 5-CH_3), 2.10 (2x s, 6H, 2x CH_3Se), 2.42 and 2.69 (2x t, *J* =7.8 Hz, 4H, 2x CH_2-CN), 3.48-3.72 (m, 12H, 4x CH-ipr, C$\underline{H}_2$-CH_2-CN, 2x H-2', 2x 2H-5'), 3.82 (s, 12H, 4x CH_3O), 3.90-4.05 (m, 2H, C$\underline{H}_2$-CH_2-CN), 4.22 and 4.31 (2x m, 2H, 2x H-4'), 4.58-4.72 (m, 2H, 2x H-3'), 6.39 and 6.42 (d and d, *J* =8.6 Hz, 2H, 2x H-1'), 6.79-6.92 (m, 8H, aromatic protrons), 7.20-7.45 (m, 18H, aromatic protons), 7.59 and 7.62 (2x s, 2H, H-6), 8.18-8.31 (br, 2H, 2x NH).

^{13}C-NMR ($CDCl_3$) δ: 4.57 and 4.71 ($SeCH_3$), 11.53 and 11.59 (5-CH_3), 20.11, 20.18, 20.46, 20.53, 22.50, 24.38, 24.45, 24.58, 24.66, 24.77,24.85, 43.23, 43.30, 43.35, 43.42, 46.36, 46.43, 47.38, 47.43, 50.36 (C-2', CH_2-CH_2-CN and C-ipr), 55.31 (OCH_3), 63.28, 63.33 (C-5'), 74.91 and 75.09 (C-3'), 84.97 (C-4'), 87.29 (Ar-C), 89.38 and 89.25 (C-1'), 111.69 (C-5), 113.46, 117.29, 126.99, 127.81, 128.54, 130.21, 130.42, 135.22, 135.40 (Ar-C), 135.44 (C-6), 144.10 (Ar-C), 150.37 (C-2), 158.85 (Ar-C), 163.34 (C-4).
HRMS (ESI-TOF): molecular formula, $C_{41}H_{51}N_4O_8PSe$; $[M+H]^+$:839.2675

(calc.839.2682).

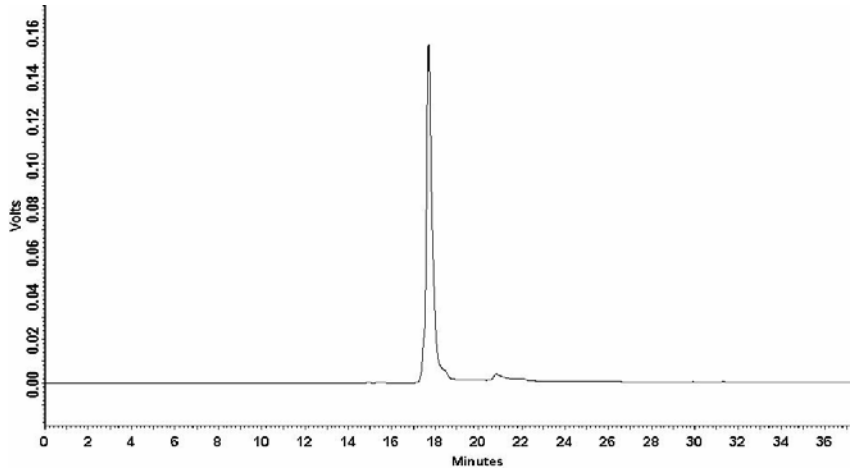


Figure 1. Reversed-phase HPLC analysis of Purified DMTr-off-9-mer : (5'-ATGGT$_{Se}$GCTC-3'). The sample was eluted (6 mL/min) with a linear gradient from buffer A (20 mM triethylammonium acetate, pH 6.9-7.1) to 70% buffer B (50% acetonitrile, 20 mM triethyl ammonium acetate) in 20 min. Its retention time is 14.8 min.

Synthesis and purification of 2'-Se-functionalized DNA oligonucleotides (6)

All DNA oligonucleotides were synthesized by solid-phase synthesis using an ABI392 DNA/RNA synthesizer on a 1 μmol scale. The concentration of the Se-modified thymidine phosphoramidite (**5**) was identical to that of the standard phosphoramidite (0.1 M in acetonitrile). The coupling reaction was carried out using a 5-BMT solution (0.3 M) in acetonitrile with a coupling time of 25 seconds. All the oligonucleotides were prepared with DMTr-On, followed by the cleavage from the CPG solid support (Beads) and the deprotection with concentrated ammonia solution under 55 ºC over night. After the

DMTr-on purification by HPLC, the detritylation of the oligonucleotides was performed by the treatment of 3% trichloroacetice acid (aqueous solution) for five minutes, followed by neutralization to pH 7.0 with triethylamine and extraction with petroleum ether to remove the by-product DMTr-OH. The DMTr-off oligonucleotides were purified again by HPLC. A typical HPLC profile of the Se-oligonucleotides is shown in Figure 1, and a typical spectrum of MS analysis of the Se-oligonucleotides is shown in Figure 2.

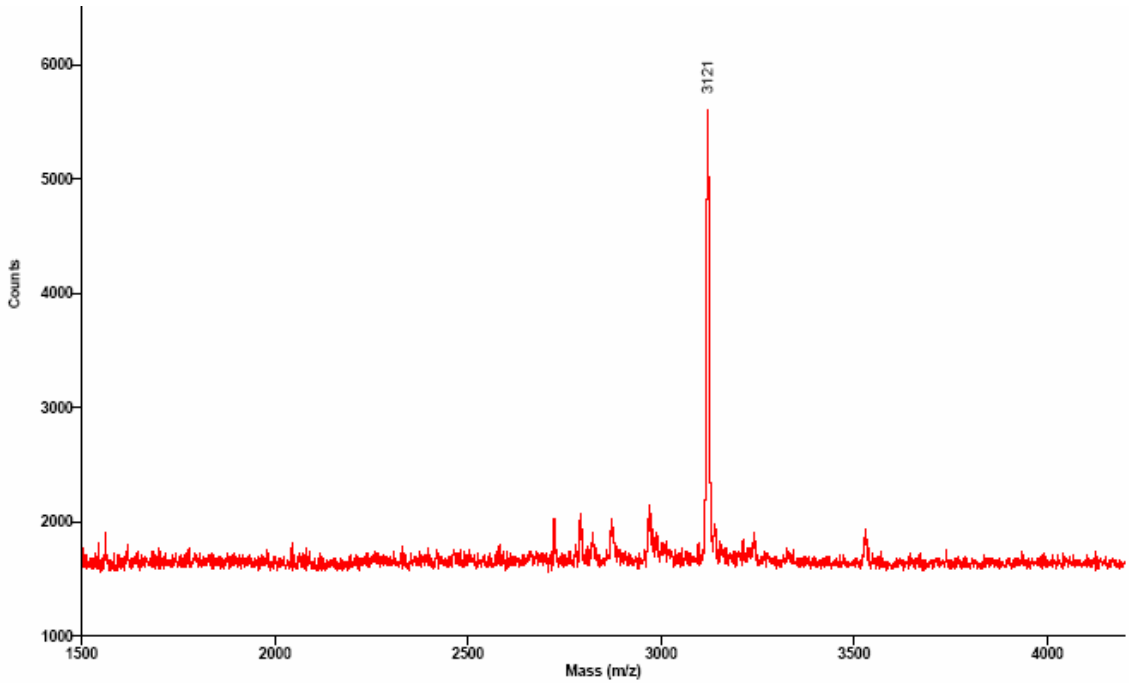
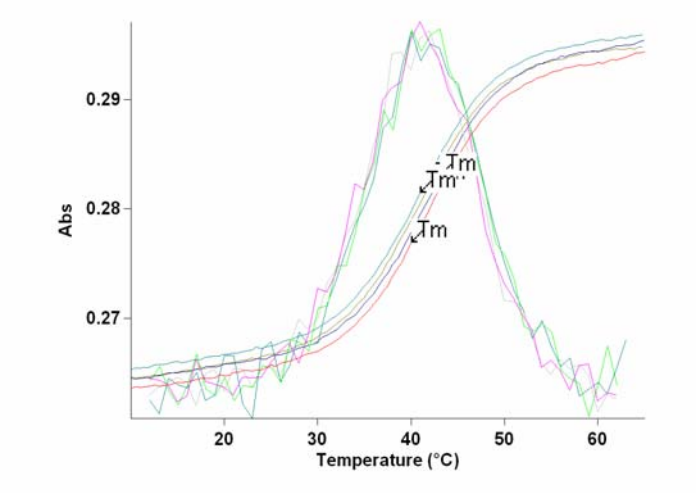


Figure 2. MALDI-TOF MS analysis of the DMTr-off Se-DNA ($GCGT_{Se}ATACGC$). Molecular formula: $C_{98}H_{125}N_{38}O_{58}P_9Se$; FW: 3120.4; found $[M+H]^+$: 3121; calculated $[M+H]^+$: 3121.4

UV-melting temperature experiements

The experiements were performed using the samples (2 μM DNA duplexes) disolved in the buffer of 50 mM NaCl, 10 mM NaH_2PO_4-Na_2HPO_4 (pH 6.5), 0.1 mM EDTA and 10 mM $MgCl_2$. The samples were heated to 85 ºC and allowed to cool down to room temperature slowly. These experiements were carried out by Cary 300 UV-Visible Spectrophotometer with a temperature controller at a heating rate of 0.5 ºC/min. Typical denaturing curves are shown in Figure 3 and 4.

A.



B.

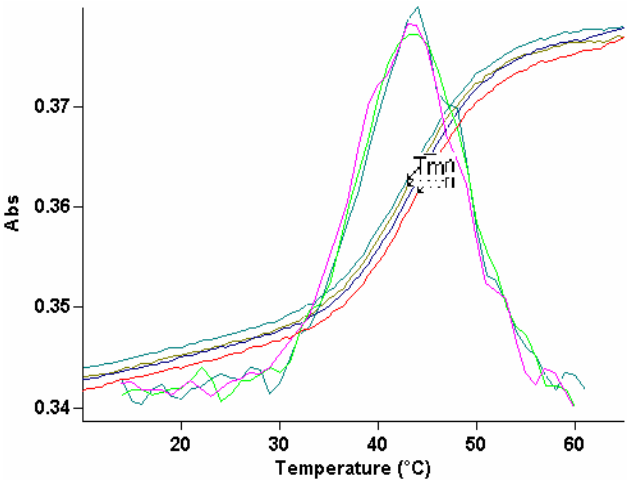


Figure 3. UV melting curves of DNAs. (A) The native DNA duplex: 5'-ATGGTGCTC-3' & 5'-GAGCACCAT-3' (Tm= 42.6 ± 1.4 ºC). (B) The Se-DNA duplex: 5'-ATGGT$_{Se}$GCTC-3' & 5'-GAGCACCAT-3' (Tm= 42.9 ± 1.1 ºC).

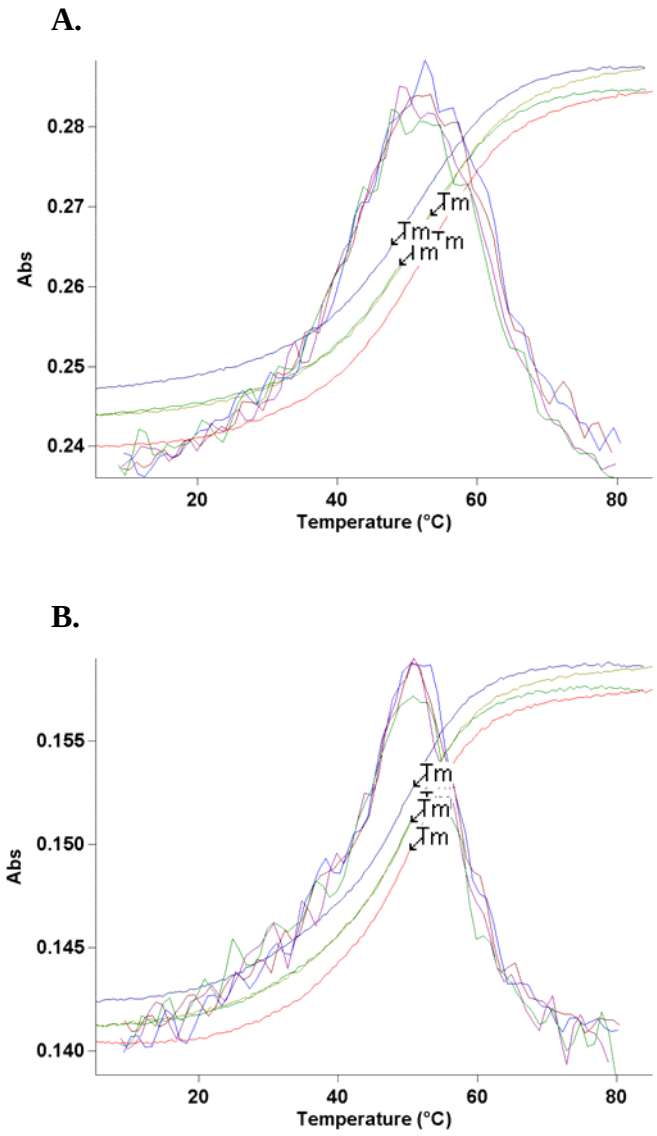


Figure 4. UV melting curves of RNAs. (A) The native tRNA TΨC loop sequence: 5'-CUGUG(rT)UCGAUCCACAG-3' (Tm= 52.4 ± 1.1 ºC). (B) The Se-tRNA TΨC loop sequence: 5'-CUGUGT_{Se}UCGAUCCACAG-3' (Tm= 51.7 ± 0.4 ºC).

Table 1. X-ray data collection, phasing and refinement statistics for 2HC7. Two data sets were collected: the data collected at selenium K-edge (0.9793 Å) was used for SAD phasing, and the data collected at 1.100 Å was used for the refinement.

Data Collection	λ = 0.9793 Å	λ=1.100 Å
Resolution range, Å (last shell)	40.0-1.50 (1.55-1.50)	40.0 - 1.40 (1.45 - 1.40)
Unique reflections	3786 (349)	4577 (492)
Completeness, %	99.7 (99.7)	98.8 (92.3)
R_{merge} ,%	6.4 (29.7)	3.6 (34.3)
$\langle I/\sigma(I)\rangle$	13.5 (4.3)	13.1 (2.1)
Redundancy	13.3 (13.6)	12.7 (10.0)
R-Cullis	0.352	
Phasing power	4.053	
Figure of merit	0.421	
Refinement		
Resolution range, Å (last shell)	18.72 - 1.40 (1.49 - 1.40)	
Number of reflections	4252 (492)	
R_{work} , %	17.0 (22.1)	
R_{free} , %	18.7 (24.4)	
Number of atoms		
Nucleic Acid (single)	162	
Heavy atom	1 (Se)	
Water	43	
R.m.s. deviations		
Bond length, Å	0.015	
Bond angle,	1.9	
Average B-factors, $Å^2$		
All atoms	18.5	
Wilson plot	16.1	
Overall anisotropic B-values		
B11/B22/B33	-0.86/-0.86/1.71	
Bulk solvent correction		
Solvent density, e/$Å^3$	0.34	
B-factors, $Å^2$	45.1	
Coordinates error (c.-v.), 5Å		
Esd. from Luzzatt plot, Å	0.15	
Esd. from SIGMAA, Å	0.08	

Crystallization and X-ray crystal structure determination of the Se-DNA

We found that the Se-DNA (GT_{Se}GTACAC, self-complementary) crystallized in two to three weeks in the native buffer, where the native DNA (GTGTACAC) crystallized over two months.[21,22] When the Nucleic Acid Mini Screen kit (Hampton Research, with 24 diversified crystallization buffers) was screened, to our pleasant surprise, we found that the Se-DNA crystallized overnight in 20 out of 24 buffers (buffer #5-24), and these

crystals also diffracted well. An investigation is going on to find out why and how the 2'-Se-T substitution facilitate the crystallization process. In contrast, the native DNA did not crystallize at all over many weeks in these Hampton buffers. The size of these Se-containing crystals is approximately 0.1x 0.1 x 0.1 mm. The structure of the Se-DNA crystals, grown in the buffer #7 of the kit (10% v/v MPD, 40 mM sodium cacodylate pH 6.0, 12 mM spermine tetra-HCI, 80 mM potassium chloride and 20 mM magnessium chloride), was obtained at 1.40 Å resolution via the Se SAD phasing and the refinement using the data collected at 1.100 Å. The structure resolution of the Se-DNA (1.40 Å) is also higher than that of the native with the same space group (2.0 Å). Data collection, phasing, and refinement statistics of the determined Se-DNA structure (2HC7) are listed in Table 1.