Published online 5 November 2008 *Nucleic Acids Research, 2008, Vol. 36, No. 22* ***7009–7018***
doi:10.1093/nar/gkn843

Derivatization of DNAs with selenium at 6-position of guanine for function and crystal structure studies

Jozef Salon, Jiansheng Jiang, Jia Sheng, Oksana O. Gerlits and Zhen Huang*

Department of Chemistry, Georgia State University, Atlanta, GA 30303, USA

Received September 16, 2008; Revised and Accepted October 14, 2008

ABSTRACT

To investigate nucleic acid base pairing and stacking via atom-specific mutagenesis and crystallography, we have synthesized for the first time the 6-Se-deoxyguanosine phosphoramidite and incorporated it into DNAs via solid-phase synthesis with a coupling yield over 97%. We found that the UV absorption of the Se-DNAs red-shifts over 100 nm to 360 nm ($\epsilon = 2.3 \times 10^4\ M^{-1}\ cm^{-1}$), the Se-DNAs are yellow colored, and this Se modification is relatively stable in water and at elevated temperature. Moreover, we successfully crystallized a ternary complex of the Se-G-DNA, RNA and RNase H. The crystal structure determination and analysis reveal that the overall structures of the native and Se-modified nucleic acid duplexes are very similar, the selenium atom participates in a Se-mediated hydrogen bond (Se…H–N), and the ^{Se}G and C form a base pair similar to the natural G–C pair though the Se-modification causes the base-pair to shift (approximately 0.3 Å). Our biophysical and structural studies provide new insights into the nucleic acid flexibility, duplex recognition and stability. Furthermore, this novel selenium modification of nucleic acids can be used to investigate chemogenetics and structure of nucleic acids and their protein complexes.

INTRODUCTION

Selenium derivatization of proteins via selenomethionine has revolutionized protein X-ray crystallography via multiwavelength anomalous dispersion (MAD), and two thirds of new crystal structures of proteins have been determined via this strategy recently (1,2). Indirect derivatization of nucleic acids with the Se-labeled proteins for structure determination of five ribozymes, such as the hepatitis δ virus ribozyme and the flexizyme (3,4), was also reported. Inspired by these advances, our research group pioneered and developed covalent incorporation of selenium into DNAs and RNAs (5–8) for structure determination via MAD or single-wavelength anomalous dispersion (SAD) phasing. This novel research area has attracted many attentions and research activities in chemical synthesis, biochemistry and structural biology (9–15).

Besides structural study, we are exploring chemogenetic investigation of nucleic acid function by the atom-specific substitution of oxygen (atomic radius, 0.73 Å) with selenium (1.16 Å, from the same elemental Family VIA in the periodic table) as an atomic probe (12,15). As the genetic information storage, replication, and transcription are achieved via base-pairing, stacking interaction and size-and-shape impact of the nucleobase pairs, extensive research has been focused on studying the recognition and stability of nucleobase pairs and double-stranded structures (16–20). Interestingly, our recent study via replacement of thymidine 4-oxygen with selenium in DNA (^{Se}T) has revealed that DNA is flexible and able to accommodate a large atom. In addition, we have discovered that the thymidine 4-selenium atom forms a hydrogen bond (Se···H–N) with the adenosine 6-amino group in DNA duplex (12), and the Se-nucleobase-derivatized DNA has an X-ray crystal structure virtually identical to the corresponding native DNA structure. This selenium substitution, which leads to the Se-mediated hydrogen bond, the Se-nucleobase stacking interaction and the Se-modified duplex, provides a unique opportunity to obtain new insights into the base-paring and stacking interactions, and the duplex recognition and stability.

Despite the synthesis of the 2′-deoxy-6-selenoguanosine and its derivatives several decades ago (21,22), synthesis of nucleic acids containing the Se-guanine remained a challenge, even though 6-S-purines have been introduced into nucleic acids (23,24). To explore the Se derivatization of guanine for structure and function studies, we report here the synthesis of a 6-Se-2′-deoxyguanosine phosphoramidite, its incorporation into oligonucleotides, and Se-DNA duplex stability investigation via UV-melting study. We also report here the observation of the colored DNAs containing the 6-Se-guanine and the Se-G DNA stability in water and at an elevated

*To whom correspondence should be addressed. Tel: +1 404 413 5535; Fax: +1 404 413 5505; Email: huang@gsu.edu

temperature. Moreover, we crystallized a ternary complex of the Se-G-DNA, RNA and RNase H, and determined its X-ray crystal structure. This is the first structure determination of a protein-nucleic acid complex on the basis of Se-derivatized nucleic acids and MAD phasing. We also report here that the overall nucleic acid structures of the native and modified duplexes are very similar, that the selenium atom forms a Se-hydrogen bond (Se···H–N), and the ^{Se}G and C form a base pair similarly to the native G–C pair, though the 6-selenium modification results in the base-pair shift for approximately 0.3 Å. This Se-base-pair shift leads to a reduction in the base-stacking interaction, which explains the decrease in UV-melting temperatures of the modified duplexes, comparing to the corresponding native ones. Our exciting and novel discoveries will open a new research avenue in structure and function studies of nucleic acids as well as their protein complexes.

MATERIALS AND METHODS

Synthesis of 6-Se-deoxyguanosine phosphoramidite (3)

N^2-[2-(4-*tert*-butylphenoxy)acetyl]-6-(2-cyanoethyl)seleno-5′-*O*-(4,4′-dimethoxytriphenylmethyl)-2′-deoxyguanosine (**2**). 2,4,6-(Triisopropylbenzene)sulfonyl chloride (300 mg, 1 mmol, 1.5 eq., TIBS–Cl) dissolved in CH_2Cl_2 (1 ml) was added to the solution of **1** (500 mg, 0.66 mmol), 4-dimethylaminopyridine (15 mg, 0.12 mmol, DMAP), and triethylamine (0.18 ml, 1.32 mmol, 2 eq., TEA) in CH_2Cl_2 (2 ml) under argon. This reaction was stirred at room temperature for 15 min (monitored by silica gel TLC to confirm the completion; 5% MeOH in CH_2Cl_2) and then injected to a solution of the sodium selenide ($NCCH_2CH_2SeNa$). The selenide solution (25) was prepared by injecting the $NaBH_4$ suspension (150 mg in 4 ml of EtOH) into a flask containing di-(2-cyanoethyl) diselenide (700 mg, 2.64 mmol, 8 eq.) dissolved in ethanol (15 ml) on an ice-bath and under argon. After the selenium incorporation reaction was completed in an hour (monitored by TLC, 5% MeOH in CH_2Cl_2, product $R_f = 0.43$), water (10 ml) was added to the reaction flask. The solution was extracted with CH_2Cl_2 (3×20 ml). The combined organic layer was dried over $MgSO_4$ (s), filtered and evaporated under reduced pressure. The crude product was then dissolved in methylene chloride (5 ml) and purified on a silica gel column equilibrated with methylene chloride. The column was eluded with a step-wise gradient of methanol-methylene chloride mixtures (CH_2Cl_2, 0.5%, 1.0%, 2.0% MeOH in CH_2Cl_2, 300 ml each) to afford product **2** as a white foam (460 mg, 80% yield over two reactions). Intermediate: 5′-*O*-(4,4′-Dimethoxytriphenylmethyl)-N^2-[2-(4-*tert*-butylphenoxy)acetyl]-6-*O*-[2,4,6-(triisopropyl-benzene)sulfonyl]-2′-deoxyguanosine. ^{1}H-NMR (CD_2Cl_2) δ: 1.28–1.32 (m, 18H, 6× CH_3-ipr), 1.36 (s, 9H, 3× CH_3-*t*Bu), 2.59–2.65 and 2.76–2.83 (2× m, $J_{2'-1'} = 6.4$ Hz, 2H, H-2′), 2.99 (h, $J = 6.8$ Hz, 1H, CH-ipr), 3.16 (br, 1H, OH), 3.35 and 3.47 (2× dd, $J_{5'-4'} = 4.0$ and $J_{5'-5'} = 10.4$ Hz, 2H, H-5′), 3.78 (s, 6H, 2× OCH_3), 4.21–4.31 (m, 3H, 2× CH-ipr and H-4′), 4.64 (s, 2H, CH_2–O), 4.85–4.88 (m, 1H, H-3′), 6.60 (*t*, $J_{1'-2'} = 6.4$ Hz, 1H, H-1′), 6.76–7.50 (m, 19H, CH-arom), 8.16 (s, 1H, H-8), 8.66 (br, 1H, NH); ^{13}C-NMR (CD_2Cl_2) δ: 23.24 and 24.30 (CH_3-ipr), 29.93 and 34.36 (CH-ipr), 31.20 (CH_3-*t*Bu), 34.07 (C-*t*Bu), 40.36 (C-2′), 55.16 (OMe), 64.23 (C-5′), 68.11 (CH_2-O), 72.31 (C-3′), 84.64 (C-1′), 86.44 (C-arom), 86.83 (C-4′), 113.05, 114.41, 124.06, 126.54, 126.81, 127.75, 128.04, 129.92, 130.08 (CH-arom), 120.57 (C-5), 131.18, 135.55, 135.88, 144.76, 145.18, 150.31, 150.76, 154.52, 154.68, 154.92, 158.63, 158.67, 166.40 (C=O). HRMS (ESI–TOF): molecular formula, $C_{58}H_{67}N_5O_{10}S$; $[M + H]^+$: 1026.4678 (calc. 1026.4681). Compound **2**: ^{1}H-NMR (CD_2Cl_2) δ: 1.35 (s, 9H, 3× CH_3-*t*Bu), 2.47 (br, 1H, OH), 2.56–2.62 and 2.80–2.88 (2× m, $J_{2'-1'} = 6.4$ Hz, 2H, H-2′), 3.16 (t, $J = 7.2$ Hz, 2H, SeC$\underline{H}_2$CH$_2$CN), 3.34 and 3.47 (2× dd, $J_{5'-4'} = 4.0$ and $J_{5'-5'} = 10.0$ Hz, 2H, H-5′), 3.58 (*t*, $J = 7.2$ Hz, 2H, Se–CH$_2$–C$\underline{H}_2$–CN), 3.79 (s, 6H, 2× OCH_3), 4.18–4.21 (m, 1H, H-4′), 4.68 (s, 2H, CH_2–O), 4.80–4.84 (m, 1H, H-3′), 6.49 (*t*, $J_{1'-2'} = 6.4$ Hz, 1H, H-1′), 6.78–7.43 (m, 17H, CH-arom), 8.10 (s, 1H, H-8), 8.96 (br, 1H, NH); ^{13}C-NMR (CD_2Cl_2) δ: 18.92 and 19.01 (SeCH$_2$CH$_2$CN), 30.94 (CH_3–*t*Bu), 33.83 (C–*t*Bu), 40.04 (C-2′), 54.95 (OMe), 63.80 (C-5′), 67.76 (CH_2–O), 72.00 (C-3′), 84.11 (C-1′), 86.17 (C-arom), 86.27 (C-4′), 112.79, 114.07, 126.39, 126.57, 127.56, 127.76, 129.73 (CH-arom), 118.93 (CN), 131.49 (C-5), 135.38, 135.42, 144.54, 144.99, 154.63, 158.39 (C-arom), 148.37 (C-4), 150.80 (C-2), 158.01 (C-6), 165.55 (C=O). HRMS (ESI–TOF): molecular formula, $C_{46}H_{48}N_6O_7Se$; $[M + Na]^+$: 899.2651 (calc. 899.2642).

N^2-[2-(4-*tert*-butylphenoxy)-acetyl]-6-(2-cyanoethyl)-seleno-5′-*O*-(4,4′-dimethoxytriphenylmethyl)-2′-deoxy-guanosine 3′-*O*-(2-cyanoethyl)-*N*,*N*-diisopropylamino phosphoramidite (**3**). Compound **2** (250 mg, 0.29 mmol) and 5-(benzylthio)-1*H*-tetrazole (27 mg, 0.15 mmol) were dried on a high vacuum overnight. Under argon, dry methylene chloride (1 ml) was added into the flask to dissolve them, followed by injection of 2-cyanoethyl *N,N,N,N*-tetraisopropylphosphorodiamidite (103 mg, 0.34 mmol, 1.2 eq.). The solution was stirred under argon at room temperature for 30 min. Reaction completion was indicated by TLC [CH_2Cl_2/EtOAc (7:3), product ($R_f = 0.58$ and 0.65): a mixture of two diastereomers]. The reaction was quenched with $NaHCO_3$ (3 ml, sat.), stirred for 5 min, and extracted with CH_2Cl_2 (3×5 ml). The combined organic layer was dried over $MgSO_4$ (s), filtered, and evaporated under reduced pressure. The crude products were purified on a silica gel column equilibrated with CH_2Cl_2, and eluted with CH_2Cl_2/EtOAc (7:3). After dissolving in CH_2Cl_2 (1 ml), these products were precipitated from pentane (200 ml) to yield the titled compound (**3**) as a white powder (230 mg, 75% yield). ^{1}H-NMR (CD_2Cl_2, two sets of signals from a mixture of two diastereoisomers) δ: 1.15–1.24 (m, 24H, 8× CH_3-ipr), 1.36 (s, 18H, 6× CH_3–*t*Bu), 2.51 and 2.65 ($2 \times t$, $J = 6.4$ Hz, 4H, 2× OCH_2CH_2CN), 2.68–2.78 and 2.88–2.95 (2× m, $J_{2'-1'} = 6.4$ Hz, 4H, 2× H-2′), 3.21 and 3.22 ($2 \times t$, $J = 7.2$ Hz, 4H, 2× SeC$\underline{H}_2$CH$_2$CN), 3.35–3.50 (m, 4H, 2× H-5′), 3.58 (*t*, $J = 7.2$ Hz, 4H, 2× SeC$\underline{H}_2$CH$_2$CN), 3.63–3.91 (m, 8H, 4× CH-ipr, 2× O–C$\underline{H}_2$–CH$_2$–CN), 3.80 and 3.82 (2× s, 12H, 4× OCH_3), 4.29–4.37 (m, 2H, 2× H-4′), 4.71 and 4.72 (2× s, 4H, 2× CH_2–O), 4.78–4.88

(m, 2H, 2× H-3′), 6.44 (*t*, $J_{1'\text{-}2'}$ = 6.4 Hz, 2H, 2× H-1′), 6.78–7.45 (m, 34H, CH-arom), 8.13 and 8.14 (2× s, 2H, 2× H-8), 8.91 and 8.92 (2× br, 2H, 2× NH); ^{13}C-NMR (CD_2Cl_2) δ: 19.14 ($SeCH_2CH_2CN$), 19.43 and 19.45 ($SeCH_2CH_2CN$), 20.20, 20.27 and 20.36, 20.43 (OCH_2CH_2CN), 24.30, 24.34, 24.37, 24.41 (CH_3-ipr), 31.21 (CH_3-*t*Bu), 34.08 (C-*t*Bu), 39.44, 39.49 and 39.54, 39.57 (C-2′), 43.22, 43.34 (CH-ipr), 55.18 and 55.21 (OMe), 58.15, 58.34 and 58.29, 58.48 (OCH_2CH_2CN), 63.35 and 63.69 (C-5′), 68.08 (CH_2–O), 73.38, 73.55 and 73.85, 74.03 (C-3′), 84.48 and 84.52 (C-1′), 85.85, 85.91 and 86.09, 86.13 (C-4′), 86.55 (C-arom), 112.59, 114.33, 126.61, 126.80, 126.83, 127.79, 128.01, 128.08, 129.99, 130.02, 130.06 (CH-arom), 117.66 and 117.82 (OCH_2CH_2CN), 119.29 ($SeCH_2CH_2CN$), 131.76 and 131.77 (C-5), 135.60, 135.65, 144.77, 145.12, 145.16, 155.00, 155.02, 158.64, 158.66 (C-arom), 148.72 and 148.76 (C-4), 151.13 and 151.16 (C-2), 158.29 and 158.35 (C-6), 165.76 (C = O); ^{31}P-NMR (CD_2Cl_2, using H_3PO_4 as the standard) δ: 149.2. HRMS (ESI–TOF): molecular formula, $C_{55}H_{65}N_8O_8PSe$; $[M + H]^+$: 1077.3902 (calc. 1077.3901).

Synthesis, purification and analysis of the Se-G DNAs

All syntheses were carried out in a 1-μmol scale and with the DMTr-on. The 6-Se-G derivatized oligonucleotides were prepared using the ultramild CE phosphoramidites (dA, dC and dG). Concentration of the Se-modified phosphoramidite (**3**) in acetonitrile was the same as those of the conventional phosphoramidites (0.1 M). The phosphoramidite coupling reaction was carried out using BTT activator (0.325 M). After the synthesis, the Se-oligonucleotides were cleaved from the solid support and fully deprotected overnight at room temperature with potassium carbonate (1 ml, 0.05 M in anhydrous methanol). The supernatant was evenly divided into two 2-ml Eppendorf tubes, followed by the addition of water (1 ml per tube) and triethylammonium acetate buffer [0.5 ml per tube, 2 M triethylammonium acetate (TEAAc), pH 7.1]. After filtration with 0.45 μ filter, the Se-DNA oligonucleotides were purified by reversed-phase high performance liquid chromatography (RP-HPLC) twice, with DMTr-on and DMTr-off.

The purification was carried out on Welchrom XB-C18 column (21.1 × 250 mm, 10 μ) with buffer A (10 mM TEAAc, pH 7.1) and buffer B (10 mM TEAAc, pH 7.1, in 50% acetonitrile). A flow rate of 6 ml/min and a gradient (starting from buffer A) were used with buffer B increased by 4.5% (for DMTr-on) or 2% (for DMTr-off) every minute over 20 min. By monitoring under 260 nm and 360 nm, the Se-DNAs were collected and lyophilized. For the 5′-DMTr deprotection, the lyophilized Se-DNAs with DMTr-on were treated with trichloroacetic acid solution (final concentration: 0.3% w/w) for 3 min, followed by the addition of TEAAc buffer (2 M) to adjust the pH to ≈7. Similarly, the Se-DNAs with DMTr-off were purified by HPLC again, followed by lyophilization. The purified Se-DNAs with DMTr-off were re-dissolved in water and analyzed by HPLC, UV and MS to confirm the high quality and integrity. Similarly, Se-DNA HPLC analysis was performed on a Welchrom XB-C18 column (4.6 × 250 mm, 5 μ) using the same buffer system. A flow rate of 1.0 ml/min and a gradient (starting from buffer A) of reaching 30% buffer B in 20 min were used.

UV absorption, thermo-stability, and duplex melting studies of the Se-G DNAs

To determine the extinction coefficient of DNA ^{Se}G by comparing with the native nucleotide, we synthesized and purified GG, ^{Se}GG and ^{Se}G^{Se}G dimers, and their UV spectra were studied. By taking the advantage of HPLC separation and UV analysis, we have developed a useful HPLC–UV approach to allow separation and accurately measure and calculate the extinction coefficients of the base-modified nucleotides. The HPLC conditions are the same as those in the oligonucleotides analysis. To study the Se-DNA thermo-stability, 6-Se-G-DNAs were heated at 60°C for 1 h in the buffer of 100 mM NaH_2PO_4-Na_2HPO_4 (pH 7.6), followed by UV and HPLC analysis.

We measured the melting temperature of the Se-DNA-derivatized duplexes along with those of the native duplexes. Prior to acquisition of the melting curves, duplexes were annealed by heating to 70°C for 2 min, followed by slowly cooling to 5°C and keeping at the temperature for 3 h. Denaturation curves were acquired at 260 nm and 1 cm path length at heating or cooling rates of 0.5°C/min, using a UV-Vis spectrophotometer equipped with a six-sample thermo-stated cell block and a temperature controller. The experiments were performed using the samples (DNA duplexes, 1.0 μM) dissolved 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 $MgCl_2$.

RESULTS AND DISCUSSION

Synthesis of 6-Se-2′-deoxyguanosine phosphoramidite (3)

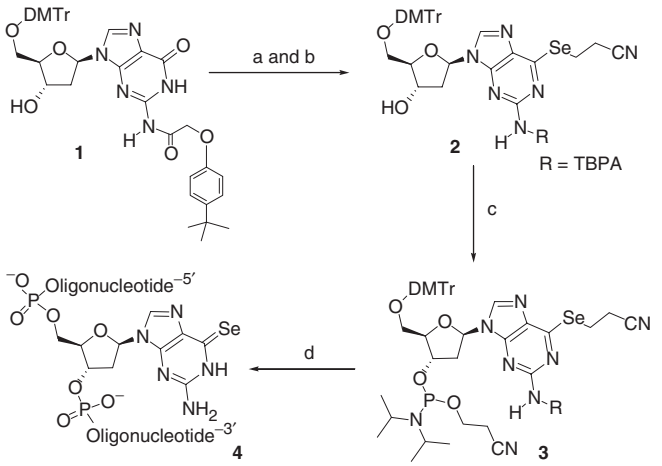
Our development of the 2-cyanoethyl-seleno protection and deprotection for the 4-Se-thymidine DNA synthesis (12) encouraged us to protect the 6-Se-functionality on deoxyguanosine with the same protecting group. In addition, the 2-cyanoethyl protecting group can be removed under ultramild conditions (0.05 M K_2CO_3 in methanol). Since strong basic conditions (such as NH_3 treatment) can cause deselenization, we decided to use (4-*tert*-butylphenoxy)acetyl (TBPA) as the protecting group (26) for the 2-NH_2 of this 6-Se-deoxyguanosine phosphoramidite (**3**), which can also be removed under the ultramild condition. Our synthesis (Scheme 1) started from the partially protected deoxyguanosine derivative (**1**). To avoid protection of the 3′-hydroxyl group, we have developed a condition that allows selective sulfonylation at the 6-position of deoxyguanosine in the presence of the free 3′-OH group, using 2,4,6-triisopropylbenzenesulfonyl chloride (27). Without purification of the activated intermediate, the protected 6-Se-deoxyguanosine derivative (**2**) was obtained (80% yield over two steps) by the substitution of the activating group at the 6-position with sodium 2-cyanoethylselenide, which was generated by the reduction of di-(2-cyanoethyl) diselenide with $NaBH_4$ (25).

The 6-Se-deoxyguanosine derivative (**2**) was converted to the corresponding phosphoramidite (**3**) in a satisfactory yield.

Synthesis of the 6-Se-G DNAs

Strong basic conditions (such as NH_3 treatment) for nucleobase deprotection cause the deselenization of 6-Se-G (^{Se}G), thus the ultramild protecting group [(4-*tert*-butylphenoxy)acetyl] (26) is used for the 2-NH_2 of **3**. The 6-Se-G derivatized oligonucleotides were prepared using the ultramild CE phosphoramidites (dA, dC and dG; 28), **3**, and BTT activator (29). These ultramild protecting groups can be removed under the ultramild deprotection condition (the K_2CO_3 treatment; 12). When the oligonucleotides contain many dG residues, phenoxyacetic anhydride (Pac_2O), instead of acetic anhydride, is used in the capping step to avoid the dG acetylation, which is difficult to remove under the K_2CO_3 treatment.

To measure the coupling efficiency of the 6-Se-G phosphoramidite (3), we synthesized 5′-DMTr-^{Se}GG dinucleotide, analyzed it by RP-HPLC (Figure 1), and compared it with the native 5′-DMTr-GG synthesis and analysis (Supplementary Data), which indicated a high coupling yield (over 97%). Typical HPLC, UV and MS analyses of the ^{Se}G-DNAs are shown in Figure 2. More MS data of the synthesized ^{Se}G-DNAs are shown in Table 1. The purified Se-DNAs are yellow colored, which is the first observation of the colored DNAs containing the 6-Se-deoxyguanosine. In addition, our synthesis and analysis indicated that the 6-Se-G functionality of the Se-DNAs is relatively stable under aqueous conditions and air.



Scheme 1. Synthesis of the 6-(2-cyanoethyl)seleno guanosine phosphoramidite (**3**) and oligonucleotides containing the 6-Se-G (**4**). Reagents and conditions: (**a**) TIBS, DMAP, TEA, CH_2Cl_2, room temperature; (**b**) diselenide, $NaBH_4$/EtOH, −5°C, 80% yield in two steps; (**c**) phosphoramidite, BTT, CH_2Cl_2, 75% yield; (**d**) solid-phase synthesis. TIBS = 2,4,6-triisopropylbenzene-1-sulfonyl chloride, DMAP = 4-dimethylaminopyridine, CH_2Cl_2 = dichloromethane, TEA = triethylamine, EtOH = ethanol, $NaBH_4$ = sodium borohydride, diselenide = $(NCCH_2CH_2Se)_2$, phosphoramidite = 2-cyanoethyl tetraisopropyl-phosphorodiamidite, BTT = 5-(benzylthio)-1*H*-tetrazole.

Determination of extinction coefficient of DNA ^{Se}G (ε_{360}^{SeG}) and UV spectroscopic studies

To determine the extinction coefficient of DNA ^{Se}G by comparing with the native nucleotide, we synthesized and purified GG, ^{Se}GG and ^{Se}G^{Se}G dimers, and their

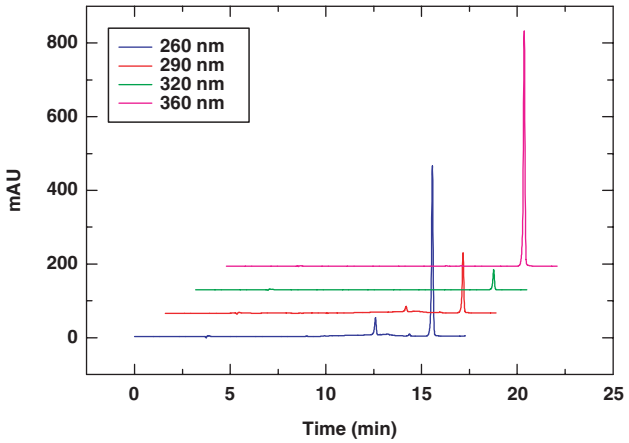


Figure 1. RP-HPLC analysis of the deprotected and crude 5′-DMTr-^{Se}GG dimer. The crude dimer was monitored at four different wavelengths: blue 260 nm, red 290 nm, green 320 nm, and pink 360 nm. HPLC conditions: Welchrom C18-XB column (4.6 × 250 mm, 5 μ), 25°C, 1 ml/min, gradient from buffer A to 30% buffer B in 20 min.

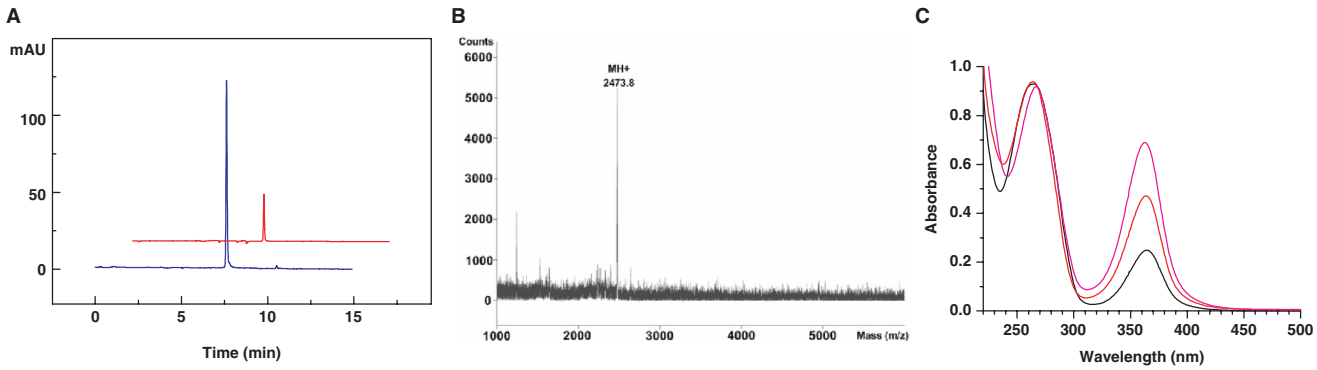


Figure 2. HPLC, MS and UV analyses of the ^{Se}G-DNAs. (**A**) RP-HPLC analysis of 5′-d(GAATCA-^{Se}G-GTGTC)-3′ [monitored at 260 nm (blue) and 360 nm (red)]. The sample was analyzed on a Welchrom XB-C18 column (4.6 × 250 mm, 5 μ) at a flow of 1.0 ml/min and with a linear gradient of 5 to 50% B in 10 min, with a retention time of 7.6 min. Buffer A: 10 mM TEAAc (pH 7.1); B: 60% acetonitrile in 10 mM TEAAc (pH 7.1). (**B**) MS analysis of 5′-d(GT-^{Se}G-TACAC)-3′. Molecular formula: $C_{78}H_{99}N_{30}O_{45}P_7Se$; $[M+H]^+$: 2473.8 (calcd: 2473.6). (**C**) UV spectra of the ^{Se}G-DNAs containing one ^{Se}G (ATG-^{Se}G-TGCAC, black), two ^{Se}Gs (ATG-^{Se}G-T-^{Se}G-CAC, red), and three ^{Se}Gs (AT-^{Se}G-^{Se}G-T-^{Se}G-CAC, pink).

UV spectra are presented in Figure 3. By taking the advantage of HPLC separation and UV analysis, we have developed this useful HPLC–UV approach to accurately measure and calculate the extinction coefficients of the base-modified nucleotides. Our experimental results indicate that ^{Se}G in DNA absorbs at both 254 nm and 360 nm (λ_{max} = 267 nm and 360 nm) while native G (λ_{max} = 254 nm) does not absorb at 360 nm (Figure 3A). The ^{Se}G λ_{max} values of ^{Se}GG (λ_{max} = 359 nm) and ^{Se}G^{Se}G (λ_{max} = 361 nm) are virtually identical; the average λ_{max} of ^{Se}G is 360 nm. Since the extinction coefficient of G at 260 nm ($\varepsilon^{G}_{260} = 1.22 \times 10^4\,M^{-1}cm^{-1}$) is known (30), we performed HPLC analysis of ^{Se}GG and ^{Se}G^{Se}G under both 260 nm and 360 nm (Figure 3B and C), and their peak areas were quantified, respectively. First, the absorption ratio at 260 nm and 360 nm of ^{Se}G^{Se}G ($A^{Se_G}_{260}/A^{Se_G}_{360}$) was calculated and determined as α value. In Figure 3B, the α value is used to calculate the 260-nm absorption contribution from ^{Se}G of the ^{Se}GG. The net 260-nm absorption from G of the ^{Se}GG (A^{G}_{260}) is obtained by subtraction of the ^{Se}G 260-nm contribution from the total ^{Se}GG absorption at 260 nm (Figure 3B). Thus, we deduced Equation (3) from Equation (1) and (2) presenting the ^{Se}GG. Since $A^{Se_G}_{360}$can be directly measured and A^{G}_{260}can be accurately calculated from this ^{Se}GG in Figure 3B, we determined $\varepsilon^{Se_G}_{360}$ as $2.3 \times 10^4\,M^{-1}cm^{-1}$. Similarly, from Equation (2) and (4) presenting the ^{Se}G^{Se}G, we deduced Equation (5) and calculated the ratio of $A^{Se_G}_{260}/A^{Se_G}_{360}$in Figure 3C, thereby accurately determining $\varepsilon^{Se_G}_{260}$ as $5.3 \times 10^3\,M^{-1}\,cm^{-1}$.

Table 1. MALDI-TOF MS data of the ^{Se}G-modified oligonucleotides

Entry	Se-oligonucleotides	Measured (calcd.) m/z
a	5′-T-^{Se}G-T-3′ $C_{30}H_{40}N_9O_{19}P_2Se$: FW 971.6	$[M + H]^+$: 973 (973)
b	5′-TT-^{Se}G-T-3′ $C_{40}H_{52}N_{11}O_{24}P_3Se$: FW 1242.8	$[M + H]^+$: 1244 (1244)
c	5′-ATG-^{Se}G-TGCTC-3′ $C_{88}H_{112}N_{32}O_{53}P_8Se$: FW 2792.8	$[M + H]^+$: 2793 (2794)
d	5′-ATG-^{Se}G-T-^{Se}G–CTC-3′ $C_{88}H_{112}N_{32}O_{52}P_8Se_2$: FW 2855.7	$[M + H]^+$: 2858 (2857)
e	5′-AT-^{Se}G-^{Se}G-T-^{Se}G–CTC-3′ $C_{88}H_{112}N_{32}O_{51}P_8Se_3$: FW 2918.7	$[M + H]^+$: 2920 (2920)
f	5′-GT-^{Se}G-TACAC-3′ $C_{78}H_{99}N_{30}O_{45}P_7Se$: FW: 2472.6	$[M + H]^+$: 2474 (2474)
g	5′-G-^{Se}G-GTACAC-3′ $C_{78}H_{98}N_{33}O_{44}P_7Se$: FW 2497.6	$[M + H]^+$: 2498 (2499)
h	5′-GC-^{Se}G-TATACGC-3′ $C_{97}H_{123}N_{38}O_{57}P_9Se$: FW: 3091.0	$[M + H]^+$: 3092 (3092)
i	5′-GCG-^{Se}G-ATACGC-3′ $C_{97}H_{122}O_{56}N_{41}P_9Se$: FW: 3115.4	$[M + H]^+$: 3116 (3116)

$$A^{o}_{260}\mathrm{G} = \varepsilon^{G}_{260} \times \mathrm{C} \times \mathrm{d} \quad \mathbf{1}$$

$$A^{Se_G}_{360} = \varepsilon^{Se_G}_{360} \times \mathrm{C} \times \mathrm{d} \quad \mathbf{2}$$

$$\varepsilon^{Se_G}_{360} = \varepsilon^{G}_{260} \times A^{Se_G}_{360}/A^{G}_{260} \quad \mathbf{3}$$

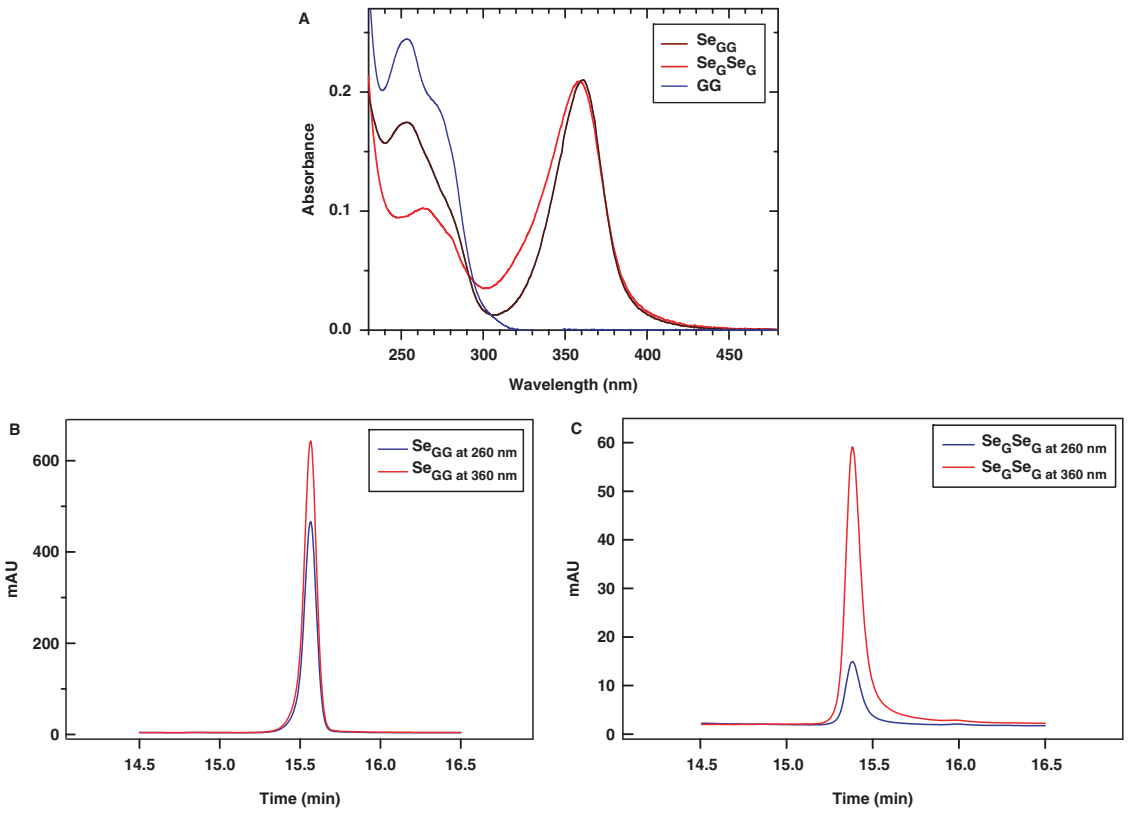


Figure 3. Calculation of $\varepsilon^{Se_G}_{360}$ via UV and HPLC analyses. (**A**) UV absorption spectra of GG dimer (blue line), ^{Se}GG dimer (black line), and ^{Se}G^{Se}G dimer (red line); (**B**) RP-HPLC analysis of ^{Se}GG dimer at 260 nm (blue line) and 360 nm (red line); (**C**) RP-HPLC analysis of ^{Se}G^{Se}G dimer at 260 nm (blue line) and 360 nm (red line).

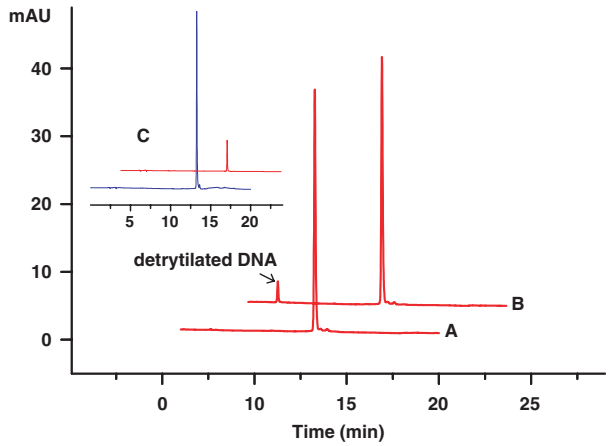


Figure 4. Thermostability studies of the 6-Se-G-DNA. The sample [5′-DMT-d(GAATCA-^{Se}G-GTGTC)-3′] was dissolved in a 100 mM phosphate buffer (pH 7.6) and analyzed by HPLC at 360 nm. (**A**) before heating; (**B**) after heating at 60°C for 1 h. (**C**) HPLC analysis of the Se-G-DNA, monitored at both 267 nm and 360 nm, before heating.

$$A_{260}^{Se_G} = \varepsilon_{260}^{Se_G} \times C \times d \quad \textbf{4}$$

$$\varepsilon_{260}^{Se_G} = \varepsilon_{360}^{Se_G} \times A_{260}^{Se_G} / A_{360}^{Se_G} \quad \textbf{5}$$

Excitingly, we observed that invisible DNA turns into colored DNA via the single atom replacement with selenium, while natural DNAs are colorless. Comparing with the native deoxyguanosine nucleotide (UV $\lambda_{max} = 254$ nm, $\epsilon = 1.22 \times 10^4\,M^{-1}\,cm^{-1}$; 30), the UV spectrum of the 6-Se-deoxyguanosine nucleotide ($\lambda_{max} = 360$ nm, $\epsilon = 2.3 \times 10^4\,M^{-1}\,cm^{-1}$) reveals a higher absorption and a large red-shift over 100 nm, thereby leading to the appearance of yellow color. This Se-nucleotide visualization is probably due to the ease of the delocalization of the selenium electrons on the nucleobase, requiring less energy for the electron excitation, thereby resulting in the large UV red-shift. In contrast, the 6-*S*-deoxyguanosine nucleotide (6-*S*-dG, $\lambda_{max} = 339$ nm; 24) shows a smaller red-shift from the deoxyguanosine nucleotide and remains colorless.

Thermostability studies of the 6-Se-G SeNAs

The DNA containing 6-Se-G [5′-DMTr-d(GAATCA-^{Se}G-GTGTC)-3′] was heated at 60°C for 1 h in the buffer of 100 mM NaH_2PO_4-Na_2HPO_4 (pH 7.6) (Figure 4). Besides a small amount of the DNA detritylation (≈5%), deselenization of the Se-G-DNA was insignificant (≤1%, determined by the Se-DNA total decrease), which shows that the 6-Se-G functionality is relatively stable in aqueous solution at the elevated temperature. Since stability in the air is required in most crystallization experiments, we have monitored the Se-derivatized DNAs for weeks by HPLC, and found insignificant deselenization.

UV-melting study of the Se-G DNA duplexes

The UV-melting temperatures were measured (T_m, Table 2) to examine the impact of the ^{Se}G residue incorporation on the thermodynamic stability of DNA duplexes. Two typical melting tempeature curves of the native and modified duplexes are shown in Figure 5. It has been observed in the literature (31) that the 6-*S*-G modification can slightly destablize nucleic aicd duplex and cause drop in the UV-melting temperature (*T*m) up to ≈3°C per *S*-modification. Since selenium atomic size is larger than sulfur atomic size, we expected that the 6-Se substitution can cause more decrease in the duplex stability and melting temperature as well. As predicted, these Se-DNA duplexes with single, double or triple Se-G-modifications indeed show significant decrease in melting temperatures (up to 11°C per modification on DNA with one Se atom, Table 2).

Interestingly, we also observed that *T*ms of the Se-DNA duplexes in Entry a and b of Table 2, where the Se-modifications are close to the 5′ or 3′ terminus, are just 3–4°C lower than that of the corresponding native complex. In addition, a small change in T_m was

Table 2. UV Melting temperatures of the ^{Se}G-modified oligonucleotides

Entry	DNA pairs	Se-DNA T_m (native) °C
a	5′-CGTACC TACAGTT-^{Se}G-T-3′ 3′-GCATGGATGTCAA—C-A-5′	51.1 ± 0.2 (55.0 ± 0.1)
b	5′-Py-A-^{Se}G–A-ACTGTAGGTACG 3′-T—C-^{Br}U-TGACATCCATGC-5′	55.2 ± 0.1 (58.3 ± 0.1)
c	5′-TACTAAC-^{Se}G-TAGTA-3′	47.9 ± 0.1 (54.0 ± 0.3)
d	5′-GAATCC-^{Se}G–CTGTC-3′ 3′-CTT AGG–C-GACAG-5′	42.0 ± 0.1 (53.0 ± 0.1)
e	5′-GAATCT-^{Se}G–CTGTC-3′ 3′-CTT AGA–C-GACAG-5′	40.0 ± 0.2 (48.0 ± 0.2)
f	5′-GAATCA-^{Se}G-GTGTC-3′ 3′-CTT AGT—C-CACAG-5′	38.4 ± 0.2 (47.0 ± 0.1)
g	5′-GC-^{Se}G-TATACGC-3′	28.5 ± 0.3 (38.0 ± 0.2)
h	5′-ATG-^{Se}G-TGCTC-3′ 3′-TAC—C-ACGAG-5′	32.3 ± 0.3 (42.5 ± 0.3)
i	5′-ATG-^{Se}G-T-^{Se}G–CTC-3′ 3′-TAC—C-A—C-GAG-5′	17.0 ± 0.5 (42.5 ± 0.3)
j	5′-AT-^{Se}G-^{Se}G-T-^{Se}G–CTC-3′ 3′-TA—C—C-A—C-GAG-5′	9.6 ± 1.0 (42.5 ± 0.3)

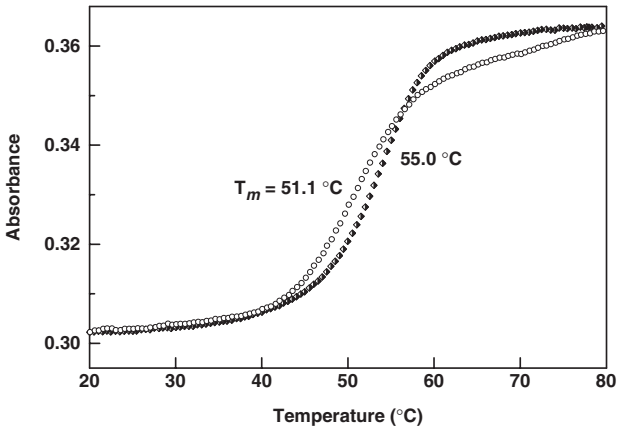


Figure 5. Normalized melting temperature curves of the non- and Se-modified DNA duplexes. The Se-DNA duplex: 5′-CGTACCTAC AGTT-^{Se}G-T-3′ and 5′-ACAACTGTAGGTACG-3′ (Open circle, T_m = 51.1°C); the corresponding native DNA duplex: (filled diamond, 55.0°C).

also observed when the Se-modification was introduced to a region containing a bulge or a flipping nucleotide. For instance, the self-complementary sequence in Entry c forms two A-bulges in the duplex, where these two A-nucleotides flip out in the crystal structure (32). T_m of this duplex with the Se-modification close to these two As drops only 2-3°C per Se-modification. These results suggest that the degree of the Se-G-duplex destabilization is dependent on the modification location, sequence and secondary structure of nucleic acids. Since the large size of selenium atom, which requires more space surrounding the modification site, is the major factor of the duplex instability, less thermo-destabilization is observed when the Se-modification is introduced to a position close to secondary structures that are more dynamic and flexible. This unique effect of the selenium modification can be taken advantage of in identifying and studying secondary structures of nucleic acids, such as DNAzymes and ribozymes, and nucleic acid-protein complexes (33–36). For structural studies of nucleic acids and their protein complexes, it is better to place this modification close to the termini of DNA or RNA, pyrimidines, or the internal bulges or loops. It is also a good idea to put the modification in RNA or DNA loops.

Structure of Se-DNA/RNA in a protein complex

In order to further study the Se-G-modification and Se-derivatized duplex of nucleic acids, we attempted to crystallize a Se-DNA/RNA/RNase H ternary complex and determine its X-ray crystal structure. Furthermore, we attempted to demonstrate the proof of principle on the structural determination of protein-nucleic acid complexes via Se-derivatized nucleic acids and phasing. By using the RNase H complex as a model system, which was determined previously (37,38), we successfully demonstrated the structure determination of a protein–nucleic acid complex on the basis of the nucleic acid Se-derivatization and MAD phasing. Besides facilitation of the structure determination, our selenium-modification study has revealed new insights into RNase H catalysis. The crystallization, structure refinement and determination results, and the enzyme mechanism study will be published elsewhere.

In this report, we focus here on the study of the Se-G-modification and its structural impact on base pairing and stacking interaction. It is easier to synthesize, derivatize, and purify DNAs than RNAs, thus the DNA portion of the DNA/RNA hybrid (5′-AT**G**TC**G**-p-3′/5′-UCGACA-3′; one-base overhang at both 5′-ends) was derivatized with the selenium functionality on two Gs (G3 and G6). The plasmid expressing *Bacillus halodurans* RNase H (D132N mutant; 37,38) was a kind gift from Yang's laboratory at the National Institute of Health. The 3D crystal structure of the Se-DNA/RNA/RNase H complex (PDB ID: 2R7Y) was successfully determined on the basis of the DNA Se-derivatization and phase information obtained from selenium scattering.

Both the native and Se-DNA-derivatized complexes were crystallized in C2 space group with similar unit cell dimensions. Our study reveals that the protein structures of both the native (2.70 Å resolution, PDB ID: 2G8U; 37,38) and modified (1.80 Å resolution, PDB ID: 2R7Y) complexes are virtually identical, and that the nucleic acid global structures of the native and Se-modified duplexes are very similar (Figure 6) though the nucleobases shift locally (Figure 6A–D). Probably due to flexibility of the over-hung ends, more structural differences are observed at the DNA and RNA termini. The distance between 1-NH of G3 in the DNA sequence and the N3 of C5 in the RNA sequence, and the distance between exo-2-NH_2 of G3 and exo-2-O of C5 are 3.16 Å and 2.59 Å, respectively (the corresponding H-bond lengths of the native G–C pair: 2.99 Å and 2.95 Å). These distances indicate the retention of the two native hydrogen bonds of the G3–C5 base pair. Since the Se atomic radius is 0.43 Å larger than that of O and hydrogen bond length is normally 2.7–3.2 Å, the distance (3.48 Å) between the G3 exo-6-Se and C5 exo-4-NH_2 (the native H-bond length: 2.99) indicates a selenium-mediated H-bond (Figure 6E). Thus, the ^{Se}G3–C5 base pair consists of three hydrogen bonds (exo-6-Se/exo-4-NH_2, 1-NH/N3, and exo-2-NH_2/exo-2-O). Interestingly, the hydrogen bond length (2.59 Å) between G3 exo-2-NH_2 and C5 exo-2-O is 0.36 Å shorter than the corresponding native bond (2.95 Å).

The bond length comparisons between the native and Se-mediated H-bonds are summarized in Table 3. Furthermore, the bond length (3.48 Å) of the Se···H–N hydrogen bond (exo-6-Se/exo-4-NH_2) in the ^{Se}G3-C5 are very close to the bond length (3.35 Å) of the Se···H–N hydrogen bond (exo-4-Se/exo-6-NH_2) previously discovered within the ^{Se}T-A base pair (12). Similarly, the ^{Se}G6 and C2 also form three H-bonds and behave in the same way (Table 3). Therefore, we have demonstrated that the ^{Se}G and C form a base pair that is similar to the natural G–C pair. Consistently, our UV-melting results agree with our structure study. The crystal structure indicates that in order to accommodate the large Se atom and to form the Se···H–N bond within the base pair and duplex, each ^{Se}G–C base pair unwinds slightly via minor adjustment of the sugar pucker (Figure 6D and E) and shifts ~0.3 Å distance within the major groove (Figure 6D), when comparing to the corresponding native G–C pair. Our crystal structure results explain the decrease in the melting temperature after the sulfur or selenium modification, which causes the base-pair shift and reduces stacking interaction, thereby destabilizing the duplex structure. Our results also explain why larger selenium atom destabilizes the duplexes more than smaller sulfur atom (31). In addition, the results are consistent with our observation of the duplex slight destabilization when the Se-modifications are introduced to the positions close to the 5′ or 3′ terminus, bulge loops or flipping nucleotides, which are more dynamic, flexible and capable of accommodating a larger atom.

CONCLUSIONS

In summary, we have synthesized the first 6-Se-deoxyguanosine phosphoramidite and incorporated it into oligonucleotides, via solid-phase synthesis under ultramild conditions, in nearly quantitative yield. We also

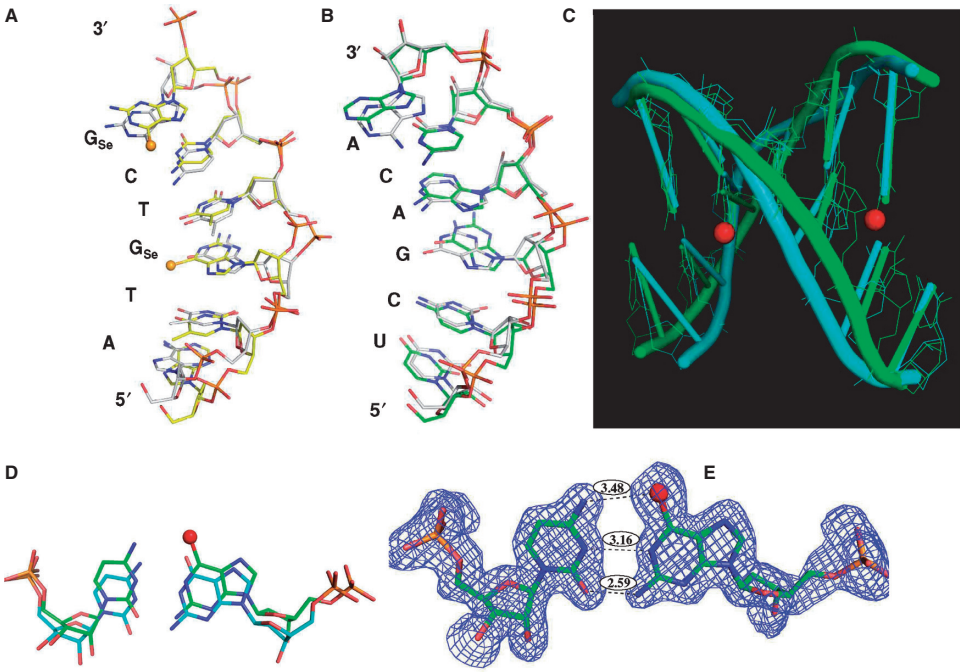


Figure 6. The superimposed global and local structures of the 6-Se-G-modified (2R7Y) and native (2G8U) DNA/RNA duplexes (5′-ATGTCG-p-3′/5′-UCGACA-3′) of the nucleic acid–protein complex; the balls represent selenium atoms in the Se-derivatized DNA (5′-AT-^{Se}G-TC-^{Se}G-p-3′). (**A**) The structure of the Se-DNA sequence (2R7Y, in yellow) is superimposed over the corresponding native (2G8U, in grey); (**B**) The structure of the RNA sequence (2R7Y, in green) is superimposed over the corresponding native (2G8U, in grey); (**C**) The duplex structure of the Se-DNA/RNA hybrid (2R7Y, in green) is superimposed over the corresponding native (2G8U, in cyan); (**D**) The comparison of the Se-modified (in green) and native (in cyan) G3/C5 base-pair structures; (**E**) The Se-G3/C5 base pair (2R7Y) with the experimental electron density shows three hydrogen bonds (exo-6-Se/exo-4-NH_2, 1-NH/N3, and exo-2-NH_2/exo-2-O) with bond lengths in 3.48 Å, 3.16 Å and 2.59 Å, respectively.

Table 3. Hydrogen bond lengths of Se-mediated H-bonds comparing to the native ones

Base pair	H-bond	Se-modified base pair bond length (Å)	Native base pair bond length (Å)
6-Se-G3/C5 in 2R7Y	exo-6-Se/exo-4-NH_2	3.48	2.99
	1-NH/N3	3.16	2.99
	exo-2-NH_2/exo-2-O	2.59	2.95
6-Se-G6/C2 in 2R7Y	exo-6-Se/exo-4-NH_2	3.43	3.07
	1-NH/N3	3.19	3.15
	exo-2-NH_2/exo-2-O	2.75	3.15
4-Se-T/A in 2NSK	exo-4-Se/exo-6-NH_2	3.35	2.87
	3-NH/N1	3.02	2.78

discovered that the Se-G-containing DNAs are yellow and with strong UV absorption at 360 nm ($\epsilon = 2.3 \times 10^4\,M^{-1}\,cm^{-1}$). In addition, this Se-functionality is stable in aqueous solution and at the elevated temperature. Moreover, we crystallized a ternary complex of the Se-G-DNA, RNA and RNase H, and revealed that the global nucleic acid structures of the native and Se-modified duplexes are very similar, the ^{Se}G and C form a base pair similar to the native G–C pair, and the selenium atom participates in a Se-mediated hydrogen bond (Se···H–N). The G–C base pair accommodates the large selenium atom by shifting approximately 0.3 Å. As this base-pair shift reduces the stacking interaction, our structure work provides a clear picture on why the Se-modification or *S*-modification causes decrease in the UV-melting temperature. Our study also points out where to better place Se atom by incorporating it to the positions close to the termini or flexible secondary structures.

The Se-modified nucleic acids have been used for the first time in derivatization and structural determination of a protein–nucleic acid complex via Se-nucleic acid and MAD phasing. Our studies shed new light on the nucleic acid stability, flexibility, and duplex recognition governed by stacking interaction, base-pairing, and size-and-shape impact of the base pairs. In addition, this novel Se-modification of nucleic acids can be used to carry out chemogenetic and spectroscopic investigation of nucleic acids and their protein complexes, and to perform the macromolecular structure study via crystallography by MAD or SAD phasing. Besides the applications in structure and function studies, the Se-G DNAs (yellow color) may also have great potentials in construction of colored DNA nanoscale devices and structures (39) as well as in nucleic acid diagnosis (40).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Dr Wei Yang at National Institute of Health for the kind gift of RNase H-expressing plasmid, and Drs Anand Saxena and Michael Becker at NSLS beamline X12C and X25 for their help in the data collection.

FUNDING

The GCC Distinguished Cancer Clinicians and Scientists Award, and the National Science Foundation (MCB-0517092 and CHE-0750235). Funding for open access charge: National Science Foundation (MCB-0517092).

Conflict of interest statement. None declared.

REFERENCES:

1. Hendrickson,W.A., Pahler,A., Smith,J.L., Satow,Y., Merritt,E.A. and Phizackerley,R.P. (1989) Crystal structure of core streptavidin determined from multiwavelength anomalous diffraction of synchrotron radiation. *Proc. Natl Acad. Sci. USA*, **86**, 2190–2194.
2. Hendrickson,W.A. (2000) Synchrotron crystallography. *Trends Biochem. Sci.*, **25**, 637–643.
3. Ferre-D,A.R., Zhou,K.H. and Doudna,J.A. (1998) Crystal structure of a hepatitis δ virus ribozyme. *Nature*, **395**, 567–574.
4. Xiao,H., Murakami,H., Suga,H. and Ferre-D,A.R. (2008) Structural basis of specific tRNA aminoacylation by a small in vitro selected ribozyme. *Nature*, **454**, 358–361.
5. Carrasco,N., Ginsburg,D., Du,Q. and Huang,Z. (2001) Synthesis of selenium-derivatized nucleosides and oligonucleotides for X-ray crystallography. *Nucleos. Nucleot. Nucl.*, **20**, 1723–1734.
6. Du,Q., Carrasco,N., Teplova,M., Wilds,C.J., Egli,M. and Huang,Z. (2002) Internal derivatization of oligonucleotides with selenium for X-ray crystallography using MAD. *J. Am. Chem. Soc.*, **124**, 24–25.
7. Caton-Williams,J. and Huang,Z. (2008) Biochemistry of selenium-derivatized naturally occurring and unnatural nucleic acids. *Chem. Biodivers.*, **5**, 396–407.
8. Sheng,J. and Huang,Z. (2008) Selenium Derivatization of Nucleic Acids for Phase Determination in Nucleic Acid X-ray Crystallography. *Int. J. Mol. Sci.*, **9**, 258–271.
9. Teplova,M., Wilds,C.J., Wawrzak,Z., Tereshko,V., Du,Q., Carrasco,N., Huang,Z. and Egli,M. (2002) Covalent incorporation of selenium into oligonucleotides for X-ray crystal structure determination via MAD: proof of principle. *Biochimie*, **84**, 849–858.
10. Moroder,H., Kreutz,C., Lang,K., Serganov,A. and Micura,R. (2006) Synthesis, oxidation behavior, crystallization and structure of 2′-methylseleno guanosine containing RNAs. *J. Am. Chem. Soc.*, **128**, 9909–9918.
11. Egli,M., Pallan,P.S., Pattanayek,R., Wilds,C.J., Lubini,P., Minasov,G., Dobler,M., Leumann,C.J. and Eschenmoser,A. (2006) Crystal structure of homo-DNA and nature's choice of pentose over hexose in the genetic system. *J. Am. Chem. Soc.*, **128**, 10847–10856.
12. Salon,J., Sheng,J., Jiang,J., Chen,G., Caton-Williams,J. and Huang,Z. (2007) Oxygen replacement with selenium at the midine 4-position for the Se base pairing and crystal structure studies. *J. Am. Chem. Soc.*, **129**, 4862–4863.
13. Jiang,J., Sheng,J., Carrasco,N. and Huang,Z. (2007) Selenium derivatization of nucleic acids for crystallography. *Nucleic Acids Res.*, **35**, 477–485.
14. Watts,J.K., Johnston,B.D., Jayakanthan,K., Wahba,A.S., Pinto,B.M. and Damha,M.J. (2008) Synthesis and biophysical characterization of oligonucleotides containing a 4′-selenonucleotide. *J. Am. Chem. Soc.*, **130**, 8578–8579.
15. Caton-Williams,J. and Huang,Z. (2008) Synthesis and DNA-polymerase incorporation of colored 4-selenothymidine triphosphate for polymerase recognition and DNA visualization. *Angew. Chem. Int. Ed. Engl.*, **47**, 1723–1725.
16. Piccirilli,J.A., Krauch,T., Moroney,S.E. and Benner,S.A. (1990) Enzymatic incorporation of a new base pair into DNA and RNA extends the genetic alphabet. *Nature*, **343**, 33–37.
17. Moran,S., Ren,R.X. and Kool,E.T. (1997) A thymidine triphosphate shape analog lacking Watson-Crick pairing ability is replicated with high sequence selectivity. *Proc. Natl Acad. Sci. USA*, **94**, 10506–10511.
18. Chaput,J.C., Sinha,S. and Switzer,C. (2002) 5-propynyluracil. diaminopurine: an efficient base-pair for non-enzymatic transcription of DNA. *Chem. Commun. (Camb)*, 1568–1569.
19. Zimmermann,N., Meggers,E. and Schultz,P.G. (2002) A novel silver(i)-mediated DNA base pair J. *Am. Chem. Soc.*, **124**, 13684–13685.
20. Leconte,A.M., Hwang,G.T., Matsuda,S., Capek,P., Hari,Y. and Romesberg,F.E. (2008) Discovery, characterization, and optimization of an unnatural base pair for expansion of the genetic alphabet. *J. Am. Chem. Soc.*, **130**, 2336–2343.
21. Mautner,H.G., Chu,S.H., Jaffe,J.J. and Sartorelli,A.C. (1963) The Synthesis and Antineoplastic Properties of Selenoguanine, Selenocytosine and Related Compounds. *J. Med. Chem.*, **6**, 36–39.
22. Milne,G.H. and Townsend,L.B. (1974) Synthesis and antitumor activity of alpha-and beta-2′-deoxy-6-selenoguanosine and certain related derivatives. *J.Med. Chem.*, **17**, 263–268.
23. Coleman,R.S., Arthur,J.C. and McCary,J.L. (1997) Covalent cross-linking of duplex DNA using 4-thio-2′-deoxyuridine as a readily modifiable platform for introduction of reactive functionality into oligonucleotides. *Tetrahedron*, **53**, 11191–11202.
24. Kadokura,M., Wada,T., Seio,K. and Sekine,M. (2000) Synthesis of 4-thiouridine, 6-thioinosine, and 6-thioguanosine 3′,5′-O-bisphosphates as donor molecules for RNA ligation and their application to the synthesis of photoactivatable TMG-capped U1 snRNA fragments. *J. Org. Chem.*, **65**, 5104–5113.
25. Logan,G., Igunbor,C., Chen,G.-X., Davis,H., Simon,A., Salon,J. and Huang,Z. (2006) A Novel and Simple Strategy for Incorporation, Protection, and Deprotection of Selenium Functionality. *Synlett*, **10**, 1554–1558.
26. Boal,J.H., Wilk,A., Harindranath,N., Max,E.E., Kempe,T. and Beaucage,S.L. (1996) Cleavage of oligodeoxyribonucleotides from controlled-pore glass supports and their rapid deprotection by gaseous amines. *Nucleic Acids Res.*, **24**, 3115–3117.
27. Palin,R., Grovea,S.J. A., Prossera,A.B. and Zhang,M.-Q. (2001) Mono-6-(O-2,4,6-triisopropylbenzenesulfonyl)-γ-cyclodextrin, a novel intermediate for the synthesis of mono-functionalised γ-cyclodextrins. *Tetrahedron Lett.*, **42**, 8897–8899.
28. Vu,H., McCollum,C., Jacobson,K., Theisen,P., Vinayak,R., Spiess,E. and Andrus,A. (1990) Fast oligonucleotide deprotection phosphoramidite chemistry for DNA synthesis. *Tetrahedron Lett.*, **31**, 7269–7272.
29. Welz,R. and Müller,S. (2002) 5-(Benzylmercapto)-1H-tetrazole as activator for 2′-O-TBDMS phosphoramidite building blocks in RNA synthesis. *Tetrahedron Lett.*, **43**, 795–797.
30. Cavaluzzi,M.J. and Borer,P.N. (2004) Revised UV extinction coefficients for nucleoside-5′-monophosphates and unpaired DNA and RNA. *Nucleic Acids Res.*, **32**, e13.
31. Bohon,J. and de los Santos,C.R. (2005) Effect of 6-thioguanine on the stability of duplex DNA. *Nucleic Acids Res.*, **33**, 2880–2886.
32. Berglund,J.A., Rosbash,M. and Schultz,S.C. (2001) Crystal structure of a model branchpoint-U2 snRNA duplex containing bulged adenosines. *RNA*, **7**, 682–691.
33. Salehi-Ashtiani,K., Luptak,A., Litovchick,A. and Szostak,J.W. (2006) A genomewide search for ribozymes reveals an HDV-like sequence in the human CPEB3 gene. *Science*, **303**, 1788–1792.
34. Silverman,S.K. (2008) Catalytic DNA (deoxyribozymes) for synthetic applications-current abilities and future prospects. *Chem. Commun. (Camb.)*, **30**, 3467–85.
35. Morikawa,K. and Shirakawa,M. (2000) Three-dimensional structural views of damaged-DNA recognition: T4 endonuclease V, E. coli Vsr protein, and human nucleotide excision repair factor XPA. *Mutat. Res.*, **460**, 257–275.
36. Benoff,B., Yang,H., Lawson,C.L., Parkinson,G., Liu,J., Blatter,E., Ebright,Y.W., Berman,H.M. and Ebright,R.H. (2002) Structural basis of transcription activation: the CAP-alpha CTD-DNA complex. *Science*, **297**, 1562–1566.

37. Nowotny,M., Gaidamakov,S.A., Crouch,R.J. and Yang,W. (2005) Crystal structures of RNase H bound to an RNA/DNA hybrid: substrate specificity and metal-dependent catalysis. *Cell*, **121**, 1005–1016.
38. Nowotny,M. and Yang,W. (2006) Stepwise analyses of metal ions in RNase H catalysis from substrate destabilization to product release. *Embo J.*, **25**, 1924–1933.
39. Ding,B. and Seeman,N.C. (2006) Operation of a DNA robot arm inserted into a 2D DNA crystalline substrate. *Science*, **314**, 1583–1585.
40. Lee,J.G., Cheong,K.H., Huh,N., Kim,S., Choi,J.W. and Ko,C. (2006) Microchip-based one step DNA extraction and real-time PCR in one chamber for rapid pathogen identification. *Lab Chip.*, **6**, 886–895.