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 (ε = 2.3 × 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 SeG 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 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 (5\textsuperscript{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-pairing 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

© 2008 The Author(s)

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/2.0/uk/) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.
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 SeG 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 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 phosphoramide (3)

\[ N^2\text{-}[2-(4-tert-butylphenoxy)acetyl]-6-(2-cyanoethyl)seleno-5'-O-(4,4'\text{-dimethoxytriphenylmethyl})-2'-deoxyguanosine (2) \]

Selenolysis of 1 (300 mg, 1 mmol, 1.5 eq., TiBS-Cl) dissolved in CHCl\(_3\) (1 ml) was added to the solution of NaBH\(_4\) (150 mg in 4 ml of EtOH) into a flask containing di-(2-cyanoethyl) diselenide (700 mg, 2.56-2.62) and triethylamine (0.18 ml, 1.32 mmol, 2 eq., TEA) in CHCl\(_3\) (2 ml) under argon. This reaction was stirred at room temperature for 15 min (monitored by TLC, completed in an hour (monitored by TLC, 5% MeOH in CH2Cl2). The reaction was completed in an hour (monitored by TLC, 5% MeOH in CH2Cl2) and then injected to a solution of the sodium selenide (NCCH2CH2SeNa). The selenide solution (25 ml) was prepared by injecting the sodium selenide (NCCH2CH2SeNa) to a solution of the sodium selenide (NCCH2CH2SeNa). This reaction was stirred at room temperature for 5 min (monitored by TLC, completed in an hour (monitored by TLC, 5% MeOH in CH2Cl2) and then injected to a solution of the sodium selenide (NCCH2CH2SeNa). The selenide solution (25 ml) was prepared by injecting the sodium selenide (NCCH2CH2SeNa) to a solution of the sodium selenide (NCCH2CH2SeNa). This reaction was stirred at room temperature for 5 min (monitored by TLC, completed in an hour (monitored by TLC, 5% MeOH in CH2Cl2) and then injected to a solution of the sodium selenide (NCCH2CH2SeNa). The selenide solution (25 ml) was prepared by injecting the sodium selenide (NCCH2CH2SeNa) to a solution of the sodium selenide (NCCH2CH2SeNa). This reaction was stirred at room temperature for 5 min (monitored by TLC, completed in an hour (monitored by TLC, 5% MeOH in CH2Cl2) and then injected to a solution of the sodium selenide (NCCH2CH2SeNa). The selenide solution (25 ml) was prepared by injecting the sodium selenide (NCCH2CH2SeNa) to a solution of the sodium selenide (NCCH2CH2SeNa). This reaction was stirred at room temperature for 5 min (monitored by TLC, completed in an hour (monitored by TLC, 5% MeOH in CH2Cl2) and then injected to a solution of the sodium selenide (NCCH2CH2SeNa). The selenide solution (25 ml) was prepared by injecting the sodium selenide (NCCH2CH2SeNa) to a solution of the sodium selenide (NCCH2CH2SeNa). This reaction was stirred at room temperature for 5 min (monitored by TLC, completed in an hour (monitored by TLC, 5% MeOH in CH2Cl2) and then injected to a solution of the sodium selenide (NCCH2CH2SeNa). The selenide solution (25 ml) was prepared by injecting the sodium selenide (NCCH2CH2SeNa) to a solution of the sodium selenide (NCCH2CH2SeNa). This reaction was stirred at room temperature for 5 min (monitored by TLC, completed in an hour (monitored by TLC, 5% MeOH in CH2Cl2) and then injected to a solution of the sodium selenide (NCCH2CH2SeNa). The selenide solution (25 ml) was prepared by injecting the sodium selenide (NCCH2CH2SeNa) to a solution of the sodium selenide (NCCH2CH2SeNa). This reaction was stirred at room temperature for 5 min (monitored by TLC, completed in an hour (monitored by TLC, 5% MeOH in CH2Cl2) and then injected to a solution of the sodium selenide (NCCH2CH2SeNa).
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 μm filter, the 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 SeG by comparing with the native nucleotide, we synthesized and purified GG, SeGG and SeSeG 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 NaH2PO4-Na2HPO4 (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 NaH2PO4-Na2HPO4 (pH 6.5), 0.1 mM EDTA and 10 mM MgCl2.

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 K2CO3 in methanol). Since strong basic conditions (such as NH3 treatment) can cause deselenization, we decided to use (4-tert-butylphenoxymethyl)acetyl (TBPA) as the protecting group (26) for the 2-NH2 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 2,6-di-tert-butylbenzenesulfonyl 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 NaBH4 (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 NH₃ treatment) for nucleobase deprotection cause the deselenization of 6-Se-G (SeG), thus the ultramild protecting group [(4-tert-butylyphenoxo)acetyl] (26) is used for the 2'-NH₂ 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 K₂CO₃ treatment; 12). When the oligonucleotides contain many dG residues, phenoxyacetic anhydride (Pac₂O), instead of acetic anhydride, is used in the capping step to avoid the dG acetylation, which is difficult to remove under the K₂CO₃ treatment.

To measure the coupling efficiency of the 6-Se-G phosphoramidite (3), we synthesized 5'-DMTr-SeGG dimerucleotide, analyzed it by RP-HPLC (Figure 1), and compared it with the native 5'-DMTr-seG synthesis and analysis (Supplementary Data), which indicated a high coupling yield (over 97%). Typical HPLC, UV and MS analyses of the SeG-DNAs are shown in Figure 2. More MS data of the synthesized SeG-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.

**Determination of extinction coefficient of DNA**

To determine the extinction coefficient of DNA ScG by comparing with the native nucleotide, we synthesized and purified GG, ScGG and ScG-SeG dimers, and their

![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, CH₂Cl₂, room temperature; (b) diselenide, NaBH₄/EtOH, −5°C; 80% yield in two steps; (c) phosphoramidite, BTT, CH₂Cl₂, 75% yield; (d) solid-phase synthesis. TIBS = 2,4,6-trisisopropylbenzene-1-sulfonyl chloride, DMAP = 4-dimethylaminopyridine, TEA = triethylamine, EtOH = ethanol, NaBH₄ = sodium borohydride, diselenide = (NCCH₂CH₂Se)₂, phosphoramidite = 2-cyanoethyl tetraisopropyl-phosphoramidite, BTT = 5-(benzylthio)-1H-tetrazole.](image1)

![Figure 1. RP-HPLC analysis of the deprotected and crude 5'-DMTr-SeGG 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 x 250 mm, 5 µ), 25°C, 1 ml/min, gradient from buffer A to 30% buffer B in 20 min.](image2)

![Figure 2. HPLC, MS and UV analyses of the ScGG-DNAs. (A) RP-HPLC analysis of 5'-d(GAATCA-ScGG-GTGTC)-3' [monitored at 260 nm (blue) and 360 nm (red)]. The sample was analyzed on a Welchrom XB-C18 column (4.6 x 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(TGTG-ScGG-TACAC)-3'. Molecular formula: C₇₈H₈₉N₃₀O₄₅P₇Se; [M + H]⁺: 2473.8 (calcld: 2473.6). (C) UV spectra of the ScGG-DNAs containing one ScG (ATG-ScG-TGCAC, black), two ScGs (ATG-ScG-T-G-CAC, red), and three ScGs (ATG-ScG-T-G-TGCAC, pink).](image3)
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 $^{79}$G in DNA absorbs at both 254 nm and 360 nm ($\lambda_{\text{max}} = 267$ nm and 360 nm) while native G ($\lambda_{\text{max}} = 254$ nm) does not absorb at 360 nm (Figure 3A). The $^{79}$G $\lambda_{\text{max}}$ values of $^{79}$GG ($\lambda_{\text{max}} = 359$ nm) and $^{79}$G$^{79}$G ($\lambda_{\text{max}} = 361$ nm) are virtually identical; the average $\lambda_{\text{max}}$ of $^{79}$G is 360 nm. Since the extinction coefficient of G at 260 nm ($e_{260} = 1.22 \times 10^4 \text{M}^{-1}\text{cm}^{-1}$) is known (30), we performed HPLC analysis of $^{79}$GG and $^{79}$G$^{79}$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 $^{79}$G$^{79}$G ($A_{260}^{G} / A_{360}^{G}$) was calculated and determined as $\alpha$ value. In Figure 3B, the $\alpha$ value is used to calculate the 260-nm absorption contribution from $^{79}$G of the $^{79}$GG. The net 260-nm absorption from G of the $^{79}$GG ($A_{260}^{G}$) is obtained by subtraction of the $^{79}$G 260-nm contribution from the total $^{79}$GG absorption at 260 nm (Figure 3B). Thus, we deduced Equation (3) from Equation (1) and (2) presenting the $^{79}$GG. Since $A_{360}^{G}$ can be directly measured and $A_{360}^{G}$ can be accurately calculated from this $^{79}$GG in Figure 3B, we determined $e_{260}^{G}$ as $3.3 \times 10^3 \text{M}^{-1}\text{cm}^{-1}$. Similarly, from Equation (2) and (4) presenting the $^{79}$G$^{79}$G, we deduced Equation (5) and calculated the ratio of $A_{260}^{G} / A_{360}^{G}$ in Figure 3C, thereby accurately determining $e_{360}^{G}$ as $3.3 \times 10^3 \text{M}^{-1}\text{cm}^{-1}$.

$A_{260}^{G} = e_{260}^{G} \times C \times d$  \hspace{1cm} \text{1}

$A_{360}^{G} = e_{360}^{G} \times C \times d$  \hspace{1cm} \text{2}

$e_{360}^{G} = e_{260}^{G} \times A_{260}^{G} / A_{360}^{G}$  \hspace{1cm} \text{3}

### Table 1. MALDI-TOF MS data of the $^{79}$G-modified oligonucleotides

<table>
<thead>
<tr>
<th>Entry</th>
<th>Se-oligonucleotides</th>
<th>Measured (calcd.) m/z</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>5’-T-$^{79}$G-T-3’</td>
<td>[M + H]^+ : 973 (973)</td>
</tr>
<tr>
<td>b</td>
<td>5’-TT-$^{79}$G-T-3’</td>
<td>[M + H]^+ : 1244 (1244)</td>
</tr>
<tr>
<td>c</td>
<td>5’-ATG-$^{79}$G-TGCTC-3’</td>
<td>[M + H]^+ : 2793 (2794)</td>
</tr>
<tr>
<td>d</td>
<td>5’-ATG-$^{79}$G-$^{79}$G-CTC-3’</td>
<td>[M + H]^+ : 2858 (2857)</td>
</tr>
<tr>
<td>e</td>
<td>5’-AT-$^{79}$G-$^{79}$G-$^{79}$GTGCTC-3’</td>
<td>[M + H]^+ : 2920 (2920)</td>
</tr>
<tr>
<td>f</td>
<td>5’-GT-$^{79}$G-TAGTAC-3’</td>
<td>[M + H]^+ : 2474 (2474)</td>
</tr>
<tr>
<td>g</td>
<td>5’-G-$^{79}$G-GTACAC-3’</td>
<td>[M + H]^+ : 2498 (2499)</td>
</tr>
<tr>
<td>h</td>
<td>5’-GC-$^{79}$G-GTATACGC-3’</td>
<td>[M + H]^+ : 3092 (3092)</td>
</tr>
<tr>
<td>i</td>
<td>5’-GCG-$^{79}$G-ATACGC-3’</td>
<td>[M + H]^+ : 3116 (3116)</td>
</tr>
</tbody>
</table>

![Figure 3](https://example.com/figure3.png)

**Figure 3.** Calculation of $e_{360}^{G}$ via UV and HPLC analyses. (A) UV absorption spectra of GG dimer (blue line), $^{79}$GG dimer (black line), and $^{79}$G$^{79}$G dimer (red line); (B) RP-HPLC analysis of $^{79}$GG dimer at 260 nm (blue line) and 360 nm (red line); (C) RP-HPLC analysis of $^{79}$G$^{79}$G dimer at 260 nm (blue line) and 360 nm (red line).
The UV-melting study of the 6-Se-G DNA duplexes

The UV-melting temperatures were measured \( (T_m, \text{Table 2}) \) to examine the impact of the Se-G residue incorporation on the thermodynamic stability of DNA duplexes. Two typical melting temperature 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 destabilize nucleic acid 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 \( T_m \)s 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
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′-ATGTGp-3′/5′-UCGAC-A-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 Sc-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 SeG3–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 SeG3–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 ScT–A base pair (12). Similarly, the SeG6 and C2 also form three H-bonds and behave in the same way (Table 3). Therefore, we have demonstrated that the ScG 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 SeG–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
discovered that the Se-G-containing DNAs are yellow and with strong UV absorption at 360 nm (\(\varepsilon = 2.3 \times 10^4 \text{M}^{-1} \text{cm}^{-1}\)). 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 SeG 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.

---

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

<table>
<thead>
<tr>
<th>Base pair</th>
<th>H-bond</th>
<th>Se-modified base pair bond length (Å)</th>
<th>Native base pair bond length (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td>6-Se-G3/C5 in 2R7Y</td>
<td>exo-6-Se/exo-4-NH₂</td>
<td>3.48</td>
<td>2.99</td>
</tr>
<tr>
<td></td>
<td>1-NH/N3</td>
<td>3.16</td>
<td>2.99</td>
</tr>
<tr>
<td></td>
<td>exo-2-NH₂/exo-2-O</td>
<td>2.59</td>
<td>2.95</td>
</tr>
<tr>
<td>6-Se-G6/C2 in 2R7Y</td>
<td>exo-6-Se/exo-4-NH₂</td>
<td>3.43</td>
<td>3.07</td>
</tr>
<tr>
<td></td>
<td>1-NH/N3</td>
<td>3.19</td>
<td>3.15</td>
</tr>
<tr>
<td></td>
<td>exo-2-NH₂/exo-2-O</td>
<td>2.75</td>
<td>3.15</td>
</tr>
<tr>
<td>4-Se-T/A in 2NSK</td>
<td>exo-4-Se/exo-6-NH₂</td>
<td>3.35</td>
<td>2.87</td>
</tr>
<tr>
<td></td>
<td>3-NH/N1</td>
<td>3.02</td>
<td>2.78</td>
</tr>
</tbody>
</table>
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). Funding for open access charge: National Science Foundation (MCB-0517092).

Conflict of interest statement. None declared.

REFERENCES:


