J|A|C|S
COMMUNICATIONS

Published on Web 03/28/2007

Oxygen Replacement with Selenium at the Thymidine 4-Position for the Se Base Pairing and Crystal Structure Studies

Jozef Salon, Jia Sheng, Jiansheng Jiang, Guexiong Chen, Julianne Caton-Williams, and Zhen Huang*

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

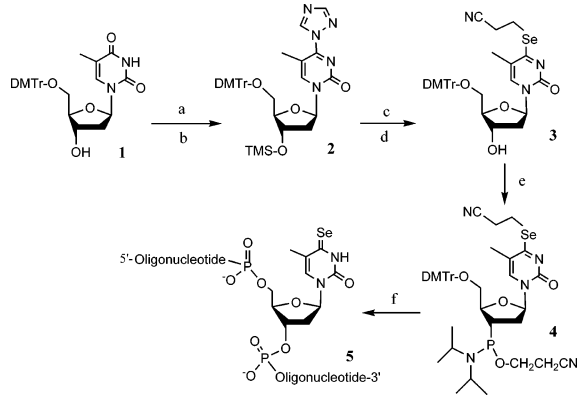
Received November 15, 2006; E-mail: huang@gsu.edu

Nucleic acids participate in many important biological functions in living systems, including genetic information storage, gene expression, and catalysis.[1−4] The T−A and C−G base pairing and stacking allow the formation of the stable DNA duplex for genetic information storage, transcription, and replication. The replacement of oxygen on the nucleobases with sulfur[5,6] has provided more insights on the DNA duplex stability, recognition, and replication at the atomic level.[7,8] The 2-thio thymidine causes the slight stabilization of the T−A paring and DNA duplex, while the 4-thio thymidine causes the slight destabilization of them.[7,8] Recent studies on these sulfur modifications have revealed the enhanced base-paring selectivity[8a] and the replication efficiency and fidelity,[8b] especially with the 2-thio thymidine, probably due to discouraging the undesired T−G wobble base pairing by the larger sulfur atom at the 2-position of the thymine.

As selenium (atomic radius, 1.16 Å), another element member of family VIA in the periodic table, is much larger than oxygen (0.73 Å), the replacement of O with Se on the nucleobases of nucleic acids will most likely provide more advantages over the natives and reveal more insights on the base-pairing selectivity, the duplex stability and recognition, and the replication efficiency and fidelity. Furthermore, the Se derivatization by replacing O can facilitate X-ray crystal structure study of nucleic acids via multiwavelength anomalous dispersion (MAD) phasing,[9] and the nucleobase Se derivatization can provide unique opportunities over the Se-derivatized nucleic acids (SeNA) with Se on other positions for the structure study.[10,11] In addition, specific pyrimidines in natural tRNAs have been derivatized by Se on the nucleobases, and the Se function has been barely understood, though it was suggested that the Se substitution might be involved in the tRNA anticodon.[12] Despite the synthesis of the Se-containing nucleobases and nucleosides several decades ago,[13] synthesis of the nucleic acids containing the Se nucleobases remained a challenge. To explore the Se derivatization on the nucleobases for the function and structure studies, we report here the first synthesis of the 4-Se-thymidine phosphoramidite (Scheme 1), its incorporation into oligonucleotides, the ^{Se}T−A base-pair formation, and the thermostability and crystal structure studies of the DNAs containing 4-Se-thymidine modification.

To replace the 4-oxygen with selenium, we activated the partially protected thymidine derivative (**1**) at the 4-position via the formation of triazolide **2** in two reactions (87% yield).[14] We also developed di(2-cyanoethyl) diselenide reagent for the Se introduction.[15] After reducing the diselenide reagent to the corresponding selenol with $NaBH_4$, the protected Se functionality was introduced to the 4-position of **2** by displacing the 4-triazolyl activating group, achieving the simultaneous Se incorporation and protection for the Se−T intermediate (**3**). Due to the undesired removal of the 2-cyanoethyl group on Se while removing the 3′-TMS group by the fluoride treatment, we developed a mild condition to remove the 3′-TMS group by using 10% triethylamine in methano (81% yield). Alternatively, we removed the 3′-TMS group first by the fluoride treatment, followed by the Se incorporation (85% yield in two reactions). Finally, Se-thymidine derivative **3** was converted to the corresponding phosphoramidite **4** in 92% yield.[10c]

Scheme 1. Synthesis of 4-Se-Thymidine Phosphoramidite (**4**) and Oligonucleotides Containing 4-Se−T (**5**)[a]



[a] Reagents and conditions: (a) TMS-Im and CH_3CN; (b) triazole−$POCl_3$−Et_3N; (c) $(NCCH_2CH_2Se)_2/NaBH_4$, EtOH; (d) 10% Et_3N in MeOH; (e) 2-cyanoethyl *N*,*N*-diisopropylchlorophosphoramidite and *N*,*N*-diisopropylethylamine in CH_2Cl_2; (f) the solid-phase synthesis.

The 4-Se−T phosphoramidite (**4**) was examined on solid-phase nucleotide synthesis. As expected, it is compatible with the coupling reaction, acetylation, I_2 oxidation, and trichloroacetic acid treatment to remove the 5′-DMTr group, without causing deselenization. Unfortunately, the concentrated ammonia treatment, which is commonly used for deprotecting the nucleobases and the 2-cyanoethyl groups from O and S functionalities,[6] caused the deselenization. To prevent the deselenization, we found the deprotecting condition by using the potassium carbonate in methanol, which removes the 2-cyanoethyl group on Se in 2 h. The oligonucleotides synthesized using the phosphoramidites containing the ultramild nucleobase deprotection groups can also be fully deprotected overnight with the same potassium carbonate solution.[16] The synthesized Se DNAs were purified and analyzed by HPLC and MS (Table 1 and Supporting Information). A typical crude Se DNA HPLC profile, which is almost identical to that of the corresponding native, is shown in Figure 1. Less than 2% short oligonucleotides were formed in this Se decamer synthesis, indicating the high yield of the Se-thymidine coupling (99%).

The stability study of 5′-ATGG^{Se}TGCTC-3′ indicated that 4-Se−T was stable at 60 °C over 3 h without significant decomposition. In addition, the UV-melting temperature of the duplexes of 5′-ATGG^{Se}TGCTC-3′ and 5′-GAGCACCAT-3′ (38.6 °C) is close to that of the corresponding native duplex (39.2 °C), suggesting no significant perturbation caused by the Se modification on thymine. We have also successfully crystallized the 4-Se−T-containing DNA (5′-G-dU_{Se}-G-^{Se}T-A-C-A-C-3′, self-complemen-

10.1021/ja0680919 CCC: $37.00

Table 1. MS Analysis of the DNAs Containing 4-Se−T

entry	Se oligonucleotides	measured (calcd) *m/z*
1	5′-^{Se}TT-3′ $C_{20}H_{17}N_4O_{12}PSe$: FW 610.1	$[M - H^+]^-$: 609.0 (609.1)
2	5′-DMTr-^{Se}T^{Se}TT-3′ $C_{51}H_{58}N_6O_{19}P_2Se_2$: FW 1280.1	$[M - H^+]^-$: 1279.0 (1279.1)
3	5′-T^{Se}TTT-3′ $C_{40}H_{53}N_8O_{25}P_3Se$: FW 1218.1	$[M + H^+]^+$: 1219.0 (1219.1)
4	5′-TT^{Se}TTT-3′ $C_{50}H_{66}N_{10}O_{32}P_4Se$: FW 1522.2	$[M + H^+]^+$: 1523.0 (1523.2)
5	5′-G^{Se}TGTACAC-3′ $C_{78}H_{99}N_{30}O_{45}P_7Se$: FW 2473	$[M + H^+]^+$: 2476 (2474)
6	5′-ATGG^{Se}TGCTC-3′ $C_{88}H_{112}N_{32}O_{53}P_8Se$: FW 2792.4	$[M + H^+]^+$: 2792.5 (2793.4)
7	5′-GCG^{Se}TATACGC-3′ $C_{97}H_{123}N_{38}O_{57}P_9Se$: FW 3091.0	$[M + H^+]^+$: 3091.6 (3092.0)

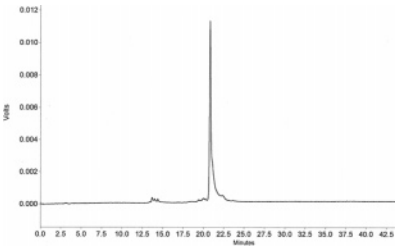


Figure 1. HPLC analysis of crude 5′-DMTr-GCG(^{Se}T)ATACGC-3′ after cleavage from the solid support and the deprotection of the bases and backbone (retention time: 21.0 min).

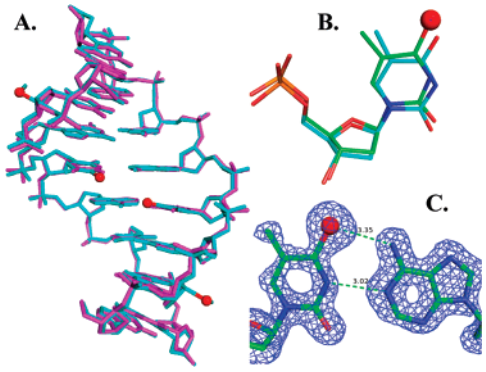


Figure 2. The global and local structures of the 4-Se−T DNA [(5′-G-dU_{Se}-G-^{Se}T-ACAC-3′)$_2$]. (A) The duplex structure of the modified DNA (2NSK, in cyan) is superimposed over the native (1DNS, in pink). (B) The comparison of the modified (in green) and native (in cyan) local T4 structures. (C) The Se base pair of T4−A5 with the experimental electron density.

tary), where the dU_{Se} (2′-Se-dU) was used to facilitate the crystal growth.[17] Though the Se functionality is relatively stable in air, DTT was still used in the crystallization buffers to prevent the oxidative deselenization of 4-Se-thymidine. The determined crystal structure (1.50 Å resolution) of the Se DNA (Figure 2) is superimposable over the native structure in the same tetragonal space group,[18] indicating that the O replacement by Se does not cause the significant structure perturbation (Figure 2A). The large Se atom is accommodated by a slight shift (Figure 2B), revealing the flexibility of the DNA duplex structure. Furthermore, the atomic distance (3.02 Å) between the thymine N3 and the adenine N1 indicates a hydrogen bond formation (Figure 2C). In addition, considering that the Se atomic radius is 0.43 Å larger than that of O, and that the distance between T4 exo-O4 and A5 exo-N6 in the native structure is 2.87 Å,[18] the distance (3.35 Å) between the exo-Se4 and exo-N6 indicates a Se-mediated hydrogen bond formation. The observation of the longer Se−hydrogen bond (usual hydrogen bond length: 2.8−3.2 Å) in the crystal structure is consistent with the 4-Se DNA UV-melting study.

In summary, we have synthesized the 4-Se thymidine phosphoramidite (**4**) and incorporated it into oligonucleotides with quantitative yield. The Se substitution on the nucleobase is relatively stable under the elevated temperature. By the UV-melting and crystal structure studies, we have further demonstrated that the O substitution with Se on the nucleobase does not cause the significant duplex structure perturbation. The crystal structure study also reveals the accommodation of the large Se atom on the thymine by the DNA duplex and the formation of the Se-mediated hydrogen bond within the T−A base pair. This work will stimulate the studies on the Se substitution on other nucleobases and positions, opening a new avenue for further exploring the base-paring selectivity governed by the hydrogen bond and base shape, the duplex structure and flexibility, the DNA replication efficiency and fidelity, and the polymerase recognition of the modified nucleobases. Moreover, the Se derivatization on the nucleobases will largely facilitate X-ray crystal structure studies of nucleic acids as well as their protein complexes.

Acknowledgment. We thank Dr. Anand Saxena at NSLS beamline X12C for his help in the data collection. This work was supported by GSU and the NSF (MCB-0517092).

Supporting Information Available: Synthetic procedures, ^{1}H and ^{13}C NMR, HPLC, and MS analytical data. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Storz, G. *Science* **2002**, *296*, 1260.
(2) Eddy, S. R. *Nat. Rev. Genet.* **2001**, *2*, 919.
(3) Latham, M. P.; Brown, D. J.; McCallum, S. A.; Pardi, A. *ChemBioChem* **2005**, *6*, 1492.
(4) Blount, K. F.; Uhlenbeck, O. C. *Annu. Rev. Biophys. Biomol. Struct.* **2005**, *34*, 415.
(5) (a) Lezius, A. G.; Scheit, K. H. *Eur. J. Biochem.* **1967**, *3*, 85. (b) Sprinzl, M.; Scheit, K. H.; Cramer, F. *Eur. J. Biochem.* **1973**, *34*, 306.
(6) Coleman, R. S.; Kesicki, E. *J. Am. Chem. Soc.* **1994**, *116*, 11636.
(7) Kutyavin, I. V.; Rhinehart, R. L.; Lukhtanov, E. A.; Gorn, V. V.; Meyer, R. B.; Gamper, H. B. *Biochemistry* **1996**, *35*, 11170.
(8) (a) Sismour, A. M.; Benner, S. A. *Nucleic Acids Res.* **2005**, *33*, 5640. (b) Sintim, H. O.; Kool, E. T. *J. Am. Chem. Soc.* **2006**, *128*, 396.
(9) (a) Yang, W.; Hendrickson, W. A.; Crouch, R. J.; Satow, Y. *Science* **1990**, *249*, 1398. (b) Ferre-D'Amare, A. R.; Zhou, K.; Doudna, J. A. *Nature* **1998**, *395*, 567. (c) Du, Q.; Carrasco, N.; Teplova, M.; Wilds, C. J.; Egli, M.; Huang, Z. *J. Am. Chem. Soc.* **2002**, *124*, 24. (d) Teplova, M.; Wilds, C. J.; Wawrzak, Z.; Tereshko, V.; Du, Q.; Carrasco, N.; Huang, Z.; Egli, M. *Biochimie* **2002**, *84*, 849. (e) Serganov, A.; Keiper, S.; Malinina, L.; Tereshko, V.; Skripkin, E.; Hobartner, C.; Polonskaia, A.; Phan, A. T.; Wombacher, R.; Micura, R.; Dauter, Z.; Jaschke, A.; Patel, D. J. *Nat. Struct. Mol. Biol.* **2005**, *12*, 218.
(10) (a) Carrasco, N.; Ginsburg, D.; Du, W.; Huang, Z. *Nucleosides Nucleotides Nucleic Acids* **2001**, *20*, 1723. (b) Wilds, C. J.; Pattanayek, R.; Pan, C.; Wawrzak, Z.; Egli, M. *J. Am. Chem. Soc.* **2002**, *124*, 14910. (c) Carrasco, N.; Buzin, Y.; Tyson, E.; Halpert, E.; Huang, Z. *Nucleic Acids Res.* **2004**, *32*, 1638.
(11) (a) Carrasco, N.; Caton-Williams, J.; Brandt, G.; Wang, S.; Huang, Z. *Angew. Chem., Int. Ed.* **2006**, *45*, 94. (b) Brandt, G.; Carrasco, N.; Huang, Z. *Biochemistry* **2006**, *45*, 8972. (c) Egli, M.; Pallan, P. S.; Pattanayek, R.; Wilds, C. J.; Lubini, P.; Minasov, G.; Dobler, M.; Leumann, C. J.; Eschenmoser, A. *J. Am. Chem. Soc.* **2006**, *128*, 10847. (d) Moroder, H.; Kreutz, C.; Lang, K.; Serganov, A.; Micura, R. *J. Am. Chem. Soc.* **2006**, *128*, 9909.
(12) (a) Chen, C. S.; Stadtman, T. C. *Proc. Natl. Acad. Sci. U.S.A.* **1980**, *77*, 1403. (b) Ching, W. M.; Alzner-DeWeerd, B.; Stadtman, T. C. *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *82*, 347.
(13) (a) Mautner, H. C. *J. Am. Chem. Soc.* **1956**, *78*, 5292. (b) Milne, G. H.; Townsend, L. B. *J. Med. Chem.* **1974**, *17*, 263. (c) Shiue, C.-Y.; Chu, S.-H. *J. Org. Chem.* **1975**, *40*, 2971.
(14) Buzin, Y.; Carrasco, N.; Huang, Z. *Org. Lett.* **2004**, *6*, 1099.
(15) Logan, G.; Igunbor, C.; Chen, G.-X.; Davis, H.; Simon, A.; Salon, J.; Huang, Z. *Synlett* **2006**, *10*, 1554.
(16) Vu, H.; McCollum, C.; Jacobson, K.; Theisen, P.; Vinayak, R.; Spiess, E.; Andrus, A. *Tetrahedron Lett.* **1990**, *31*, 7269.
(17) Jiang, J.; Sheng, J.; Carrasco, N.; Huang, Z. *Nucleic Acids Res.* **2007**, *35*, 477.
(18) Jain, S.; Zon, G.; Sundaralingam, M. *Biochemistry* **1989**, *28*, 2360.

JA0680919