J|A|C|S COMMUNICATIONS

Published on Web 01/28/2010

High Fidelity of Base Pairing by 2-Selenothymidine in DNA

Abdalla E. A. Hassan, Jia Sheng, Wen Zhang, and Zhen Huang*

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

Received November 3, 2009; E-mail: huang@gsu.edu

The base pairs are the contributors to the sequence-dependent recognition of nucleic acids, genetic information storage, and high fidelity of DNA polymerase replication and RNA polymerase transcription. The base hydrogen bonding and stacking also contribute to the folding and structure of DNA and RNA.[1] To increase RNA structure and function diversities, wobble base pairs (such as U/G) are involved in RNAs, including ribozymes and viral and rRNAs.[2] However, the wobble base pairing, where T (or U) pairs with G instead of A, reduces specific base-pairing recognition[3] and compromises the high fidelity of the enzymatic polymerization, especially when G or T on the template is modified.[4,5] The wobble base pairing can increase misrecognition and misincorporation during nucleic acid polymerization.[5,6]

Since the T/G wobble pairing (or U/G) is caused by the shift of hydrogen bonding via the 2-exo-oxygen participation within the base pairs (Figure 1) and the 2-exo-oxygen atom is not involved in T/A (or U/A) base pairing, the 2-position is critical for the discrimination between T/A and T/G pairs (or U/A vs U/G). Thus, we hypothesized the discrimination of the wobble base pairs by manipulating the steric and electronic effects at the 2-exo position. In addition to a poor hydrogen-bonding ability, selenium (atomic radius, 1.16 Å) has a much larger radius than oxygen (0.73 Å) and sulfur (1.02 Å) in the same elemental family. Thus, introduction of a selenium atom at position 2 can largely increase the electronic and steric effects, resulting in strong base-pairing discrimination, presumably better than the incorporation of a sulfur atom.[7] This atom-specific selenium substitution of the 2-oxygen of thymidine provides a unique chemical strategy to enhance the base pairing specificity. Furthermore, because of the noninvolvement of the 2-exo-oxygen in T/A pairing, its atomic replacement with selenium would most likely allow the normal T/A pairing while discouraging T/G wobble pairing. Herein we report the first synthesis of the novel 2-Se-thymidine (^{Se}T) derivative, its phosphoramidite, and the Se-DNAs. Moreover, our biophysical and structural studies of the 2-Se-T DNAs reveal that the bulky selenium atom at the 2-position can largely increase the mismatch discrimination (including the wobble pairing discrimination) while maintaining the ^{Se}T/A virtually the same as the native T/A base pair, without significant perturbation. This atom-specific substitution and probing provide a novel tool to investigate the specific recognition of base pairs, which is the basis for high fidelity during replication, transcription, and translation.

Though the selenium functionality was introduced into the 2-position of uridine over three decades ago,[8] selenium has not been incorporated into the 2-position of thymidine and DNA due to synthetic challenges. Following our successful selenium incorporation at the pyrimidine 4-positions via the C-4 activation and displacement,[9] our successful synthesis started from the 5′-trityl-protection of 2-thiothymidine derivative **1** (Scheme 1). To activate the 2-thio moiety mildly, we alkylated the 2-thio-functionality of **2** with CH_3I to give **3**.[10] Treatment of **3** with freshly prepared NaSeH gave a clean selenization reaction, and **4** was isolated in 82% yield. The protection of the 2-seleno moiety of **4** was finally achieved with ICH_2CH_2CN, giving **5** in 91% yield. Phosphitylation[9] of **5** gave Se-phosphoramidite derivative **6**. 2-Se-thymidine (^{Se}T) phosphoramidite **6** was found to be compatible with the conditions of the solid-phase synthesis, and the stability of the protected 2-Se-T moiety allows us to successfully synthesize the Se-oligonucleotides using the ultramild protecting groups.[9] The coupling of **6** into DNA is similar to the native DNA synthesis. The synthesized Se-DNAs were purified and analyzed by HPLC and MS (Table S1 and Supporting Information).

Our UV-denaturation studies showed that the melting temperatures of the native and Se-modified DNA duplexes (Tables 1 and S2) are almost the same (~0.4 °C difference), suggesting that the

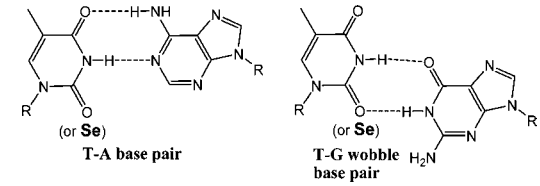
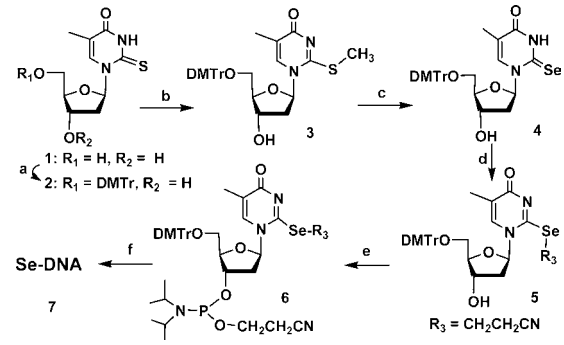


Figure 1. Native and Se-modified T/A base pair and T/G wobble pair.

Scheme 1. Synthesis of 2-Se-T phosphoramidite (**6**) and Se-DNA (**7**)[a]



[a] Reagents and conditions: (a) DMTr-Cl, Pyridine, DMAP, rt; (b) DBU, DMF, CH_3I; (c) Se, $NaBH_4$, EtOH; (d) I-CH_2CH_2CN, *i*-Pr_2NEt, CH_2Cl_2; (e) (*i*-$Pr_2N)_2P(Cl)OCH_2CH_2CN$, (*i*-$Pr)_2NEt$, CH_2Cl_2; (f) Solid-phase synthesis.

Table 1. Melting Temperatures of Native and ^{Se}T-DNA Duplexes

Sequences and Mismatches	Pairs	T_m (°C)/$-\Delta G°$ (kcal/mol)[11]	ΔT_m (°C)
I: 5′-CTTCT **T**GTCCG-3′ 3′-GAAGA**A**CAGGC-5′	T/A	42.6/10.06	−
I + 3′-GAAGA**T**CAGGC-5′	T/T	30.9/7.05	−11.7
I + 3′-GAAGA**G**CAGGC-5′	T/G	35.6/8.23	−7.0
I + 3′-GAAGA**C**CAGGC-5′	T/C	28.0/6.11	−14.6
II: 5′-CTTCT^{Se}**T** GTCCG-3′ 3′-GAAGA**A**CAGGC-5′	^{Se}T/A	42.2/9.84	−0.4
II + 3′-GAAGA**T**CAGGC-5′	^{Se}T/T	31.9/7.55	−10.7
II + 3′-GAAGA**G**CAGGC-5′	^{Se}T/G	28.9/7.18	−13.7
II + 3′-GAAGA**C**CAGGC-5′	^{Se}T/C	23.9/5.58	−18.7

10.1021/ja909330m

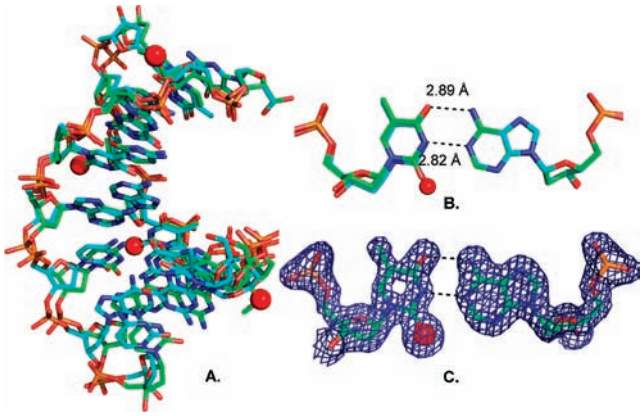


Figure 2. Global and local structures of the 2-Se-T-DNA [(5′-GdU$_{2'\text{-Se}}$-G-^{Se}T-ACAC-3′)$_2$], with a resolution of 1.58 Å. (A) The superimposed comparison of the Se-DNA duplex (3HGD, in green) with its native counterpart (1D78, in cyan). The red balls represent the selenium atoms. (B) The superimposed comparison of the local ^{Se}T4/A5 (in green) and native T4/A5 (in cyan) base pairs. (C) The experimental electron density map of the ^{Se}T4/A5 base pair with $\sigma = 1.0$.

oxygen substitution with selenium at the 2-position of thymidine does not cause significant perturbation. The melting temperature decrease of the ^{Se}T/T mismatched duplex (ΔT_m = 10.7 °C) by comparison with the matched native duplex is very similar to that of the T/T mismatched duplex (ΔT_m = 11.7 °C), suggesting the same level of the T/T mismatch discrimination. Moreover, the T_m decrease of the ^{Se}T/G wobble-paired duplex (ΔT_m = 13.7 °C) is much more than that of the native T/G wobble-paired duplex (ΔT_m = 7.0 °C), suggesting a stronger discrimination against the T/G wobble pairing after the 2-Se incorporation (Tables 1 and S2). Interestingly, the T_m decrease of the ^{Se}T/C-mismatched duplex (ΔT_m = 18.7 °C) is also much higher than that of the native T/C-mismatched duplex (ΔT_m = 14.6 °C), suggesting a stronger selectivity against the T/C mismatch after the 2-selenium substitution. It appears that the 2-Se-modification of thymidine significantly increases the selectivity against T/G wobble and T/C mismatches while retaining the same level of strong discrimination against the T/T mismatched pair and the same high fidelity of the native T/A matched base pair. Results of thermal dynamic calculation[11] (Tables 1 and S3) are consistent with the UV-melting studies.

Furthermore, this increase of the base pair specificity is consistent with our crystal structure study. We have solved the X-ray crystal structure of the Se-DNA (5′-G-dU$_{2'\text{-Se}}$-G-^{Se}T-ACAC-3′)$_2$ at the atomic resolution (Figure 2), via the crystallization facilitation using the 2′-Se-dU moiety.[12] Superimposition of the determined Se-DNA crystal structure (1.58 Å) over the corresponding native in the same tetragonal space group[13] reveals that these two structures are very similar (Figure 2A). Moreover, we found that the ^{Se}T/A base pair is virtually identical to the native T/A pair (Figure 2B and 2C). As expected, the large selenium atom fits well in the structure, since the 2-exo-position of T is not involved in the hydrogen bond formation with A. Furthermore, our biophysical and structural studies suggest that the bulky 2-Se atom (a weak hydrogen-bond acceptor) may discourage the wobble pairing by both the steric hindrance and the electronic effect: the weaker hydrogen-bonding ability of the 2-Se atom compared to the 2-O atom in the native wobble pair (Figure 1). The bulky 2-Se atom does not interact with the pairing A and thus causes no disruption on the T/A pairing, which is consistent with our UV-melting study. Therefore, the 2-Se atom bulkiness and electronic effect are probably the main factors responsible for the discrimination against the formation of the wobble ^{Se}T/G base pair, which is in a good agreement with our UV-melting and structural results on the native and Se-modified T/G wobble pairs.

In summary, we have developed the novel chemistry to first synthesize the 2-Se-derivatized thymidine, 2-Se-T phosphoramidite, and 2-Se-T DNAs. Our biophysical studies on the 2-Se-DNAs indicate that the perfectly matched Se-DNA duplexes have almost the same stability as the native ones. Interestingly, the 2-Se-substitution largely increases the specificity of the base pair recognition by further discouraging the T/G wobble and T/C base pairs, providing a unique chemical strategy to further enhance base-pairing fidelity. Consistently, our crystal structure study further reveals that the selenium-atomic substitution does not significantly alter the native T/A base pairing and overall duplex structure. Our experimental results indicate that this 2-Se-substitution facilitates the higher specificity of the thymidine pairing with the natural nucleobases. Moreover, this 2-Se-derivatized thymidine provides a useful tool in derivatization and phasing for X-ray crystal structure studies of nucleic acids and their protein complexes. The Se-atom-specific probing will open new research opportunities for further investigating base-pair recognition and the high fidelity of DNA polymerase replication, RNA polymerase transcription, and mRNA translation.

Acknowledgment. We thank Dr. Alexei Soares at NSLS X12B for his help in diffraction data collection and Dr. David Wilson and Dr. Douglas Turner for their discussions on thermal dynamic calculations. This work was supported by a GCC Distinguished Cancer Clinicians and Scientists award and the NSF (CHE-0750235 and MCB-0824837).

Supporting Information Available: Detailed experimental procedures, ^{1}H and ^{13}C NMR, and MALDI-MS spectra, UV-melting, and crystal diffraction data. These materials are available free of charge via the Internet at http://pubs.acs.org.

References

(1) (a) Watson, J. D.; Crick, F. H. C. *Nature* **1953**, *171*, 737–738. (b) Sussman, J. L.; Kim, S. *Science* **1976**, *192*, 853–858. (c) Drew, H. R.; Wing, R. M.; Takano, T.; Broka, C.; Tanaka, S.; Itakura, K.; Dickerson, R. E. *Proc. Natl. Acad. Sci. U.S.A.* **1981**, *78*, 2179–2183. (d) Cate, J. H.; Gooding, A. R.; Podell, E.; Zhou, K.; Golden, B. L.; Kundrot, C. E.; Cech, T. R.; Doudna, J. A. *Science* **1996**, *273*, 1678. (e) Montange, R. K.; Batey, R. T. *Nature* **2006**, *441*, 1172. (f) Egli, M.; Pallan, P. S. *Annu. Rev. Biophys. Biomol. Struct.* **2007**, *36*, 281. (g) Neidle, S. *Principles of Nucleic Acid Structure*; Academic Press: London, 2007.
(2) (a) Zhang, L.; Doudna, J. A. *Science* **2002**, *295*, 2084. (b) Schrader, O.; Baumstark, T.; Riesner, D. *Nucleic Acids Res.* **2003**, *31*, 988.
(3) (a) Kennard, O. *J. Biomol. Struct. Dyn.* **1985**, *3*, 205. (b) Herschlag, D. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 6921.
(4) (a) Swann, P. F. *Mutat. Res.* **1990**, *233*, 81. (b) Basu, A. K.; Loechler, E. L.; Leadon, S. A.; Essigmann, J. M. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 7677. (c) Spratt, T. E.; Levy, D. E. *Nucleic Acids Res.* **1997**, *25*, 3354. (d) Eoff, R. L.; Irimia, A.; Egli, M.; Guengerich, F. P. *J. Biol. Chem.* **2007**, *282*, 1456.
(5) Yoshida, K.; Tosaka, A.; Kamiya, H.; Murate, T.; Kasai, H.; Nimura, Y.; Ogawa, M.; Yoshida, S.; Suzuki, M. *Nucleic Acids Res.* **2001**, *29*, 4206.
(6) (a) Huang, M. M.; Arnheim, N.; Goodman, M. F. *Nucleic Acids Res.* **1992**, *20*, 4567. (b) Hubner, A.; Kruhoffer, M.; Grosse, F.; Krauss, G. *J. Mol. Biol.* **1992**, *223*, 595.
(7) (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.
(8) (a) Wise, D. S.; Townsend, L. B. *J. Heterocycl. Chem.* **1972**, *9*, 1461. (b) Shiue, C. Y.; Chu, S. H. *J. Org. Chem.* **1975**, *40*, 2971.
(9) (a) Salon, J.; Sheng, J.; Jiang, J.; Chen, G.; Caton-Williams, J.; Huang, Z. *J. Am. Chem. Soc.* **2007**, *129*, 4862. (b) Caton-Williams, J.; Huang, Z. *Angew. Chem., Int. Ed.* **2008**, *47*, 1723.
(10) Connolly, B. A.; Newman, P. C. *Nucleic Acids Res.* **1989**, *17*, 4957.
(11) Turner, D. H; Petersheim, M. *Biochemistry* **1983**, *22*, 256–263.
(12) (a) Jiang, J.; Sheng, J.; Carrasco, N.; Huang, Z. *Nucleic Acids Res.* **2007**, *35*, 477–485. (b) Sheng, J.; Salon, J.; Gan, J.-H.; Huang, Z. *Sci. China, Ser. B: Chem.* **2010**, *53*, 78–85.
(13) Jain, S.; Zon, G.; Sundaralingam, M. *Biochemistry* **1989**, *28*, 2360–2364.

JA909330M