Selenium derivatization of nucleic acids for crystallography

Jiansheng Jiang[1,2], Jia Sheng[1], Nicolas Carrasco[1] and Zhen Huang[1,*]

[1]Department of Chemistry, Georgia State University, Atlanta, GA 30302, USA and [2]Department of Biology, Brookhaven National Laboratory, Upton, NY 11973, USA

PDB ID codes[+]

ABSTRACT

The high-resolution structure of the DNA (5′-GTGTACA-C-3′) with the selenium derivatization at the 2′-position of T2 was determined via MAD and SAD phasing. The selenium-derivatized structure (1.28 Å resolution) with the 2′-Se modification in the minor groove is isomorphorous to the native structure (2.0 Å). To directly compare with the conventional bromine derivatization, we incorporated bromine into the 5-postion of T4, determined the bromine-derivatized DNA structure at 1.5 Å resolution, and found that the local backbone torsion angles and solvent hydration patterns were altered in the structure with the Br incorporation in the major groove. Furthermore, while the native and Br-derivatized DNAs needed over a week to form reasonable-size crystals, we observed that the Se-derivatized DNAs grew crystals overnight with high-diffraction quality, suggesting that the Se derivatization facilitated the crystal formation. In addition, the Se-derivatized DNA sequences crystallized under a broader range of buffer conditions, and generally had a faster crystal growth rate. Our experimental results indicate that the selenium derivatization of DNAs may facilitate the determination of nucleic acid X-ray crystal structures in phasing and high-quality crystal growth. In addition, our results suggest that the Se derivatization can be an alternative to the conventional Br derivatization.

INTRODUCTION

X-ray crystallography is a powerful tool for 3D structure determinations of DNA and RNA molecules and nucleic acid–protein complexes, which provides insight into the structure–function relationships of DNA–drug complexes, functional RNAs, (such as catalytic RNAs and snoRNAs), and RNA–protein and DNA–protein interactions (1–7). Besides the difficulties related to crystallization, however, heavy-atom derivatization for phase determination, a limiting factor in nucleic acid X-ray crystallography, has largely slowed down structural determination of new structures, including DNA–protein complexes. The conventional approaches (8,9) for DNA and RNA derivatization, such as heavy-atom soaking and co-crystallization, have proved to be much more difficult for nucleic acids than for proteins, probably because nucleic acids often lack specific binding sites for metal ions. In the conventional halogen derivatization, bromine (K-edge, 0.920 Å) is used to derivatize DNAs via 5-bromo-deoxyuridine (thymidine analog) and RNAs via 5-bromo-uridine (8,9) in multi-wavelength anomalous dispersion (MAD) phasing. The derivatives containing the bromine derivatization are relatively stable because of formation of vinyl halides. However, it is difficult to introduce bromine to other positions because of the chemical nature of the halogen. For instance, as bromide is a good leaving group, bromine placed on the ribose 2′-position can be displaced by the pyrimidine exo-2-oxygen, the purine 3-nitrogen, or the nearby nucleotide phosphate groups. Though the 8-bromo-purine nucleosides can been easily prepared (10), the 8-Br-substituted purine bases tend to adopt a syn-conformation (11), which perturbs the local structure. Therefore, in the conventional halogen derivatization, the halogens are primarily limited to the 5-position of uracil (sometimes cytosine), which places the halogen in the major groove. As the major groove of A-form dsDNA or dsRNA is narrow and deep, the halogen derivatization may perturb these structures. The halogen derivatization can also cause base-stacking disruption, hydration pattern change and other structural perturbations. In addition, halogen derivatives, such as the bromine derivatives, are light sensitive, and long-time exposure to X-ray or ultraviolet (UV) sources may cause decomposition (12,13). These factors may have caused failures of bromine derivative crystallization, diffraction, and MAD phasing in many cases.

The selenomethionine derivatization method (14–17) has revolutionized the protein X-ray crystallography, and over

---

two-thirds of novel protein structures have been determined recently by the selenomethionine MAD or single-wavelength anomalous dispersion (SAD) phasing (18,19). Indirect derivatization method of RNA using selenomethionine-labeled U1A for phase and structure determination was also successfully demonstrated though it is labor-intensive (20). Recently, we were the first (Huang, Egli and the co-workers) to demonstrate the direct derivatization of DNAs and RNAs with selenium for nucleic acid X-ray crystallography via MAD phasing (21–27). This atom-specific derivatization has great potential in structural and functional studies of nucleic acids. So far, this novel derivatization methodology has been utilized in X-ray crystal structure studies of RNA and DNA molecules by several laboratories (23,24,28–30), including structure analysis and determination of ribozymes and riboswitches. To further facilitate the selenium derivatization of DNAs and RNAs, the enzymatic strategies have also been developed to prepare selenium-derivatized nucleic acids (SeNA) via DNA polymerization and RNA transcription (31,32). As the synthesis and purification of SeNAs are much easier than those of proteins, nucleic acid–protein complexes may also be derivatized by SeNAs instead of their protein counterparts. It appears that the selenium derivatization strategy has great potential in X-ray crystal structure determination of nucleic acids and their protein complexes.

Our UV thermal denaturation and NMR studies of the Se-modified duplexes have indicated the same stability as the non-modified DNAs and RNAs (25–27). In order to further investigate the selenium derivatization strategy, we determined the structure of the DNA (5′-GTGTACAC-3′) at high-resolution (1.28 Å) using the 2′-Se derivatization and also report here the structural comparison between the Se and Br derivatization. It is worth to mention that the Se derivatization is placed on the sugar ring and located in the minor groove of the duplex, while the Br derivatization occurs to the base and is located in the major groove. The structure study reveals that the DNA duplex structure with the Se derivation is identical to the native structure (33) while the Br derivatization causes local structure perturbations in all three bromine-derivatized DNA structures. This report demonstrates that the Se derivatization can be an alternative to the conventional Br derivatization. In addition, we also demonstrated that the Se derivatization may facilitate determination of nucleic acid crystal structures in phasing and high-quality crystal growth.

MATERIALS AND METHODS

Synthesis and incorporation of selenium into oligonucleotides

For the simplicity of the synthesis, 2′-Me-Se-dU (22) is used to replace T2 in the self-complementary DNA (GTGTACAC, Figure 1A). In order to directly compare the Se derivatization and the conventional Br derivatization strategy, we also determined the DNA structure with the Br derivatization at the T4 position (Figure 1B), in the presence and absence of the Se derivatization at the T2 position. The native DNA (5′-GTGTACAC-3′), the Se-DNA (5′-GdU$_{Se}$GTACAC-3′), the Br-DNA (5′-GTGdU$_{Br}$ACAC-3′), and the Se-Br-DNA
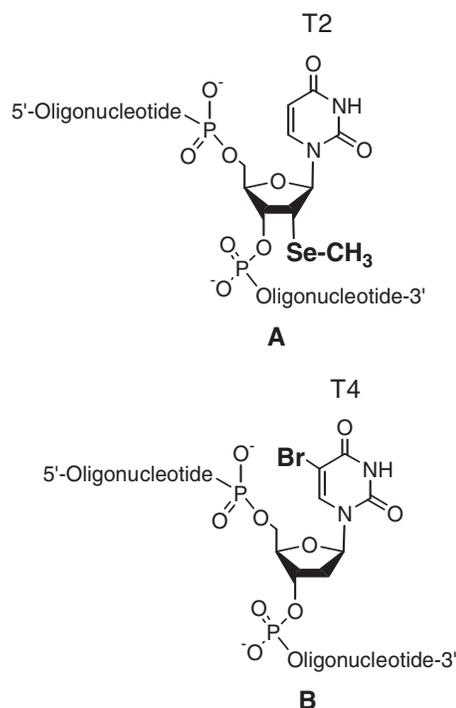
Figure 1. (**A** and **B**) Structures of the 2′-Se modification and 5-Br modification of DNA oligonucleotide.

(5′-GdU$_{Se}$GdU$_{Br}$AC-A-C-3′) were synthesized on solid phase by following the developed strategy (22,26) and were purified by high-performance liquid chromatography (HPLC) using a 21.2 × 250 mm Zorbax, RX-C8 column at a flow rate of 10 ml/min [Buffer A: 20 mM triethylammonium acetate (TEAAc, pH 7.1); Buffer B: 20 mM TEAAc (pH 7.1) in 50% aqueous acetonitrile]. The purified oligonucleotides were analyzed by MS and HPLC to confirm their sequences and purity.

Crystallization

These purified oligonucleotides (1 mM) were first heated to 90°C for 1 min, and the samples were then allowed to cool slowly to 25°C. Though the Se-DNA crystallized in the native buffer (33,34), crystallization conditions were still screened using the Nucleic Acid Mini Screen kit (Hampton Research, www.hamptonresearch.com/). A total of 1 μl of the oligonucleotide solutions (1 mM) was typically used in each of these screens. Crystallization was carried out using the hanging drop method by vapor diffusion at 25°C. In the cases of the Se-DNA and the Se-Br-DNA, their crystals with diffraction quality were identified in many identical buffer conditions. Crystals of the Se-DNA (5′-GdU$_{Se}$GTA-CAC-3′), grown in buffer #7 of the kit [10% v/v MPD, 40 mM sodium cacodylate (pH 6.0), 12 mM spermine tetra-HCl, 80 mM potassium chloride and 20 mM magnessium chloride], were identified to give the highest diffraction resolution, while crystals of the Se-Br-DNA (5′-GdU$_{Se}$G-dU$_{Br}$ACAC-3′), grown in buffer #9 [10% v/v MPD, 40 mM sodium cacodylate (pH 6.0), 12 mM spermine tetra-HCl, 80 mM sodium chloride, 12 mM potassium chloride and 20 mM magnessium chloride], gave highest resolution. The Br-DNA crystal was

grown under the native conditions (33, 34, 40% v/v MPD, 4 mM spermine and 8 mM magnesium chloride). The best Br-DNA crystal was obtained after a week under the lower temperature (15°C). The size of these crystals generating good diffraction is approximately 0.1 × 0.1 × 0.1 mm.

Data collection

30% glycerol or PEG 6000 was used as a cryoprotectant while data collection was taken under the liquid nitrogen stream at 99°K. The Se-DNA and Br-DNA data were collected at beam line X12C, and the Se-Br-DNA datasets were collected at X25, X29 and X12 in NSLS of Brookhaven National Laboratory. A number of crystals were scanned to find the one with strong anomalous scattering at the K-edge absorption of selenium and/or bromine. The distance of the detector to the crystals was set to 90 mm. The chosen wavelengths for selenium and bromine MAD are listed in Tables 1 and 2. The crystals were exposed for 10 or 15 s per image after one degree rotation, and a total of 180 images were taken for each dataset. The additional reference datasets were collected at 1.10 Å wavelength and were used for the final structure refinement. All data were processed using HKL2000 and DENZO/SCALEPACK (35).

Structure determination and refinement

The positions of the selenium and bromine atoms were identified by the heavy-atom search scripts provided by CNS (36). Following the refinement of the selenium and bromine positions, the experimental phases were calculated and extended using the MAD or SAD procedures in CNS. The initial phased maps were then improved by the solvent flattening and density modification procedure. The model of single-strand of DNA was manually built into the electron density map using O (37), and then the refinement was carried out. The refinement protocol includes simulated annealing, positional refinement, restrained B-factor refinement and bulk solvent correction. The stereo-chemical topology and geometrical restrain parameters of DNA/RNA (38) have been applied. The topologies and parameters for modified dU with selenium (UMS) and dU with bromine (BRU) were constructed and applied. After several cycles of refinement and the model rebuilding, a number of highly ordered waters were added. Final, the occupancies of selenium and bromine

were refined. Cross-validation (39) with a 5–10% test set was monitored during the refinement. The σ_A-weighted maps (40) of the ($2m|F_o| - D|F_c|$) and the difference ($m|F_o| - D|F_c|$) density maps were computed and used throughout the model building.

RESULTS AND DISCUSSION

Crystallization facilitation by the Se derivatization

We performed the crystallization screening to study the crystallizability of these modified oligonucleotides, and found that the Se-DNA formed quality crystals in many buffers of the Hampton kit overnight (approximate size: 0.1 × 0.1 × 0.1 mm). The Se-Br-DNA also formed quality crystals overnight in the same buffers, though the crystal sizes were smaller (approximately 0.05 × 0.05 × 0.05 mm). However, both the Br-DNA and the native DNA did not grow crystals in any buffers of the kit (24 buffers) over several months. While native DNA and the Br-DNA crystallize in the native conditions (33,34), it is worth to mention that these two Se-derivatized DNAs can also crystallize in the native conditions.

To further demonstrate this interesting observation, we have synthesized and purified four other native DNAs [(41,42), PDB ID: 395D & 1QPH] and their selenium derivatives (Se-DNAs): 5′-GdU$_{Se}$ACGTAC-3′, 5′-GTAC-G-CGdU$_{Se}$AC-3′, 5′-CCGdU$_{Se}$ACGTACGG-3′, and 5′-GA-CC-ACGTGGdU$_{Se}$C-3′. The Se derivatization was incorporated into different regions of these DNA molecules, including the 5′, middle and 3′ regions. These self-complementary DNA molecules were screened using the Hampton kit. High-quality single crystals (average size: 0.1 × 0.1 × 0.1 mm) of these Se-DNA sequences were formed in three days under many different buffer conditions: 5′-GdU$_{Se}$ACGTAC-3′ in buffer #9, #12, #13, #22 and #23, 5′-GTACGCGdU$_{Se}$AC-3′ in buffer #8, #10, #12, #17, #19 and #24, 5′-CCGdU$_{Se}$ACGTA-CGG-3′ in buffer #11, #12 and #21, and 5′-GACCACG-TGGdU$_{Se}$C-3′ in buffer #18, #22 and #23. The corresponding native DNA sequences, however, did not crystallize over two weeks in any of these 24 buffers. It seems that it is quite challenging to crystallize the native molecules. In summary, these observations suggest that these Se-derivatized DNAs can crystallize in broader buffer conditions and have better crystallizability than these native and Br-derivatized DNAs.

Table 1. Summary of data collection and phasing statistics for Se-DNA (1Z7I)

	Selenium K-edge Peak	Inflection	Remote
Wavelength Å	0.9790	0.9794	0.9400
Resolution range Å (last shell)	50.0–1.4 (1.49–1.4)	50.0–1.4 (1.49–1.4)	50.0–1.4 (1.49–1.4)
Unique reflections	4478 (441)	4425 (420)	4396 (402)
Completeness %	97.6 (99.1)	96.3 (94.0)	95.2 (90.0)
R_{merge} %	6.9 (40.2)	7.2 (44.0)	7.3 (45.0)
$<I/\sigma(I)>$	14.0 (3.8)	13.9 (3.9)	13.2 (3.5)
Redundancy	13.0 (10.5)	11.5 (9.7)	10.5 (9.7)
R-Cullis (Friedel)	0.513 (0.330)	0.371 (0.279)	−(0.428)
Phasing power λ_3/λ_t (Friedel)	2.026 (4.345)	3.155 (5.021)	−(3.405)
Figure of merit (Friedel)	0.426 (0.508)	0.502 (0.539)	−(0.380)
Overall figure of merit	0.817		

$R_{merge} = \Sigma|I - <I>|/\Sigma I$.

Our observations indicate that the 2′-selenium derivatization may facilitate the crystal growth, which is probably assisted by the better packing of the Se-derivatized DNA. As the bulky 2′-Se modification may be able to lock the sugar pucker into the 2′-exo conformation, the created rigidity of the Se-derivatized DNAs may facilitate better molecular packing and faster crystal growth. Furthermore, the structure determination of the modified DNAs (the Se-DNA, the Se-Br-DNA and the Br-DNA), grown under different buffer conditions, indicated that the different buffers may not change the crystal packing and the space group of the crystal form.

MAD and SAD phasing

As one Se atom should enable the phase determination for DNAs and RNAs with up to 30 nt (23), one selenium or one bromine atom in the octamer DNA is sufficient for MAD or SAD phasing. As shown in Table 1, we have used the Se-DNA data only up to 1.4 Å for MAD and SAD phasing. The overall figures of merit (FOM) of the initial phases for both Se-DNA MAD and SAD datasets were 0.816 and 0.442, respectively. This produced an excellent electron density map showing a single-strand of DNA after solvent flattening and density modification. The reference data were used as the amplitudes of the structure factors and combined with the phase probabilities from MAD/SAD, and the phases were extended from 1.4 to 1.28 Å for the refinement of Se-DNA. The MAD datasets for the Se-Br-DNA were available to up to 1.8 Å resolutions. The data at the selenium K-edge and the bromine K-edge were processed separately. As shown in Table 2, the overall FOMs of the initial phases were 0.731 and 0.761, respectively from the selenium and bromine K-edge datasets. Both produced excellent electron density maps clearly showing the DNA strands. The experienced phases were extended to 1.5 Å for the refinement of the Se-Br-DNA structures. Likewise, the Br-DNA structure was determined. The data collection summary, phasing statistics and refinement statistics of the Se-DNA, Se-Br-DNA, and Br-DNA are shown in Tables 1–3.

These Se and/or Br modified DNA structures have the same tetragonal space group $P4_32_12$, as the native structure (33). These duplex models are generated by applying the symmetry operator (y, x, −z) of $P4_32_12$. Figure 2 shows

the electron density maps with the duplex models of the Se and Br-derivatized DNA structures. The resulting structures of the 2′-Se-DNA (1Z7I), the Se-Br-DNA (2DLJ) and the Br-DNA (2H05) were refined at 1.28, 1.5 and 1.8 Å resolution, respectively. The second Se-Br-DNA structure (2GPX), containing a barium ion, was refined at 1.6 Å resolution; its electron density map (identical to 2DLJ) is not shown. Figure 3 illustrates the details of the electron density maps of the modified dUs in the structures. Figure 3A, B and C represent the structures of the 2′-Se-dU in the Se-DNA, 5-Br-dU in the Se-Br-DNA and 5-Br-dU in the Br-DNA, respectively.

Structures of the derivatized DNAs

The Se-DNA structure (1Z7I, in blue, Figure 4A) is superimposable over the native A-form DNA structure (33, 1DNS, in pink), which has the same tetragonal space group ($P4_32_12$). The overall r.m.s.d. of the Se-DNA over the native is low (0.328 Å) and mainly contributed by the first and the last bases (as shown in Figure 4A), and the main structure is considered as the same as the native, although the spermine molecule bound to the native was not observed in the derivatized structure. The superimposed local dU_{Se} structure of the Se-DNA (1Z7I, in blue) over T2 of the native structure (1DNS, in pink) is shown in Figure 4B. The derivatized local structure of the Se-DNA is also superimposable over that of the second native (34, 1D78, in gray, Figure 4B), which has a different space group ($P6_122$). Therefore, the selenium modification causes no local perturbations when it is compared with both native forms. In addition, superimposing the dU_{Se} in this Se-DNA structure (space group $P4_32_12$, Figure 4C) with the dU_{Se} in the reported DNA structure (23, GCGTAdU_{Se}A-CGC, 1MA8 in yellow, space group $P2_12_12_1$) indicates no difference, suggesting that the Se-derivatized local structure is stable and independent from its sequence environment.

Figure 4D and E show the superimposed Br-DNA (2H05, in green) and Se-Br-DNA (2DLJ in cyan) over the native (1DNS, in pink), respectively. We observed the local perturbations in both Br-derivatized DNA structures, thought the global structures were very close to the native structure with the same crystal form (1DNS, 33). The overall r.m.s.d. values of the Se-Br-DNA and the Br-DNA over the native are 0.713 and 0.642 Å, respectively. In addition, the

Table 2. Data collection and phasing statistics for Se-Br-DNA (2DLJ)

| | Selenium K-edge | | | Bromine K-edge | | |
	Peak	Inflection	Remote	Peak	Inflection	Remote
Wavelength Å	0.9790	0.9794	0.9400	0.9196	0.9199	0.9400
Resolution range Å (last shell)	50.0–1.80 (1.86–1.80)	50.0–1.80 (1.86–1.80)	50.0–1.80 (1.86–1.80)	50.0–1.80 (1.86–1.80)	50.0–1.80 (1.86–1.80)	50.0–1.80 (1.86–1.80)
Unique reflections	2283 (222)	2284 (219)	2288 (220)	2290 (220)	2292 (220)	2288 (220)
Completeness %	99.8 (100.0)	99.8 (100.0)	99.9 (100.0)	99.8 (100.0)	99.9 (100.0)	99.9 (100.0)
R_{merge} %	7.8 (36.7)	7.2 (31.8)	6.7 (32.9)	7.9 (39.5)	7.9 (41.0)	7.8 (40.4)
$I/\sigma(I)$	11.4 (3.1)	12.2 (3.7)	12.1 (3.7)	11.4 (3.1)	11.1 (2.6)	11.1 (2.7)
Redundancy	12.8 (12.8)	12.9 (13.0)	12.9 (13.0)	12.9 (13.1)	12.9 (13.0)	12.9 (13.0)
R-Cullis (Friedel)	0.709 (0.455)	0.469 (0.359)	−(0.510)	0.686 (0.417)	0.587 (0.407)	−(0.482)
Phasing power (Friedel)	1.221 (3.223)	2.365 (4.043)	−(2.826)	1.205 (4.002)	1.646 (4.053)	−(3.395)
Figure of merit (Friedel)	0.321 (0.429)	0.425 (0.477)	−(0.311)	0.324 (0.465)	0.385 (0.478)	−(0.356)
Overall figure of merit	0.731			0.761		

$R_{merge} = \Sigma|I - <I>|/\Sigma I$.

Table 3. Data collection and refinement statistics in Se-DNA, Se-Br-DNA and Br-DNA structures

Structure (PDB ID)	Se-DNA (1Z7I)	Se-Br-DNA (2D LJ)	Se-Br-DNA (2GPX)	Br-DNA (2H05)
Data collection				
Space group	$P4_32_12$	$P4_32_12$	$P4_32_12$	$P4_32_12$
Cell dimensions: $a = b, c$ (Å),	42.076, 23.935	42.369, 23.744	42.252, 24.133	42.048, 24.270
Resolution range Å (last shell)	50.00–1.28 (1.36–1.28)	50.00–1.50 (1.55–1.50)	50.00–1.60 (1.66–1.60)	50.00–1.80 (1.86–1.80)
Unique reflections	6325 (552)	3777 (354)	3161 (297)	2248 (203)
Completeness %	95.3 (70.0)	99.7 (99.2)	99.7 (99.3)	99.4 (95.8)
R_{merge} %	8.0 (32.8)	5.1 (27.8)	8.4 (49.9)	4.5 (32.3)
$I/\sigma(I)$	14.8 (4.3)	14.4 (4.9)	13.1 (4.2)	12.1 (2.7)
Redundancy	12.4 (10.7)	12.3 (9.1)	12.4 (9.0)	12.1 (7.1)
Refinement				
Resolution range Å (last shell)	20.81–1.28 (1.36–1.28)	18.16–1.50 (1.59–1.50)	18.90–1.60 (1.70–1.60)	18.81–1.80 (1.91–1.80)
R_{work} %	17.7 (23.7)	18.3 (18.9)	18.7 (21.3)	21.3 (27.3)
R_{free} %	20.0 (23.1)	22.8 (22.7)	23.8 (24.4)	24.8 (41.3)
Number of reflections	5275 (552)	3597 (473)	2909 (289)	2036 (247)
Number of atoms				
Nucleic Acid (single)	161	160	160	160
Heavy-atoms and ion	1 Se	1 Se, 1 Br	1 Se, 1 Br, 1 Ba^{2+}	1 Br
Water	44	43	31	33
R.m.s.d				
Bond length Å	0.012	0.011	0.010	0.009
Bond angle	1.8	1.6	1.8	1.9
Average B-factors Å^2				
All atoms	15.0	18.8	25.9	32.5
Wilson plot	14.9	18.2	23.6	29.8
Overall anisotropic B-values				
B11/B22/B33	0.70/0.70/−1.41	−0.29/−0.29/0.58	0.31/0.31/−0.62	2.20/2.20/−4.41
Bulk solvent correction				
Solvent density e/Å^3	0.30	0.32	0.34	0.32
B-factors Å^2	22.2	34.5	40.05	41.04
Coordinates error (c.-v.) 5Å				
Esd. from Luzzatt plot Å	0.15	0.19	0.22	0.23
Esd. from SIGMAA Å	0.10	0.05	0.12	0.20

$R_{merge} = \Sigma|I - <I>|/\Sigma I$.
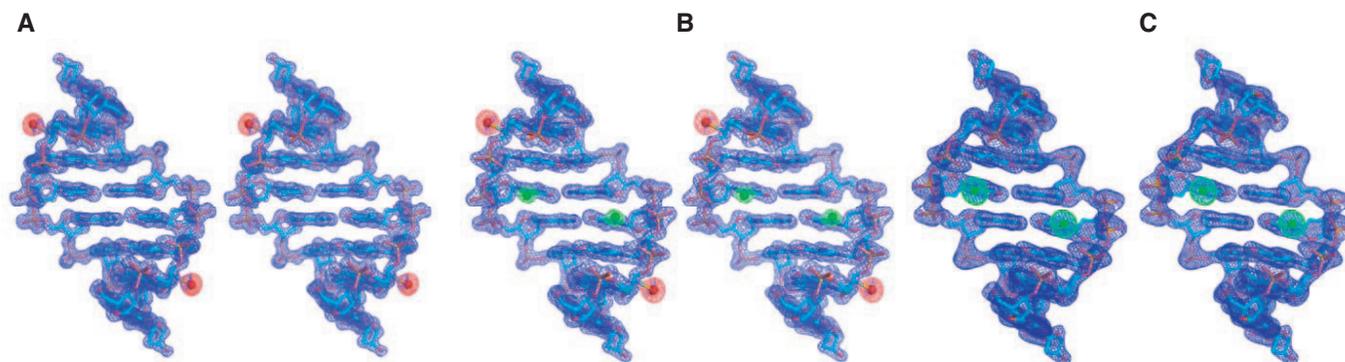
Figure 2. Electron density maps and models of the derivatized DNA duplexes (Stereo view). Red and green balls represent Se and Br, respectively. Contours are at the 1.2σ level. (**A**) The structure of the Se-DNA ($GdU_{Se}GTACAC$) at 1.28 Å resolution (1Z7I). (**B**) the structure of the Se-Br-DNA ($GdU_{Se}GdU_{Br}ACAC$) at 1.5 Å resolution (2DLJ). (**C**) The structure of the Br-DNA ($GTGdU_{Br}ACAC$) at 1.8 Å resolution (2H05).

comparison of the Br-DNA dU_{Br} with the T4 of the native DNA (1DNS) and the Se-DNA indicated that the Br incorporation could alter the hydration pattern (Figure 5A and B) and the water networking (Figure 6) at the derivatized site, though the sugar pucker of the dU_{Br} was almost not changed from that of the native (Figure 5A). The alterations at the dG3 site of the Br-DNA (Figure 5C and Table 4), including the dG sugar pucker change and ~107 degree rotation about its $C_{4'}$–$C_{5'}$ bond, were also observed. This rotation caused large local perturbations on the dG3, with its $C_{5'}$ and P deviations by

1.30 and 0.63 Å, respectively. The $C_{2'}$ of the Br-DNA T2 is also deviated by 0.60 Å (Figure 5D), which also created the flattened sugar pucker as the dG3 (Figure 5C). Unsurprisingly, the same perturbations were also observed with the Se-Br-derivatized DNA (2DLJ) while the perturbations were not observed with the Se-derivatized DNA (1Z7I). These perturbations caused by the Br derivatization may explain why some Br-derivatized DNAs have been difficult to crystallize even after successful crystallization of their natives. Our observation may serve as a reminder in the bromine derivatization.
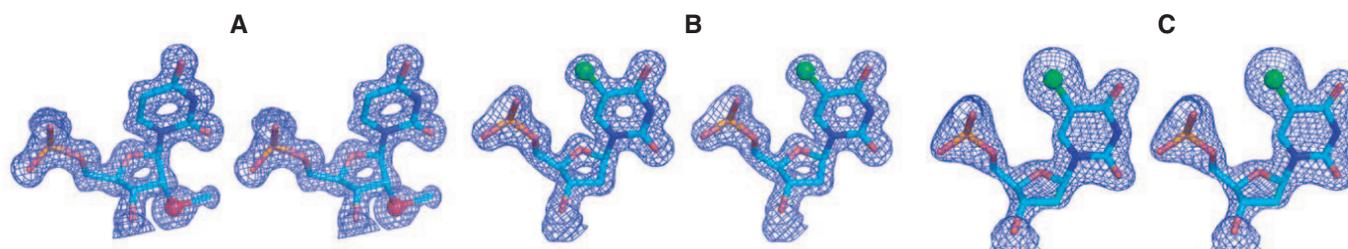
Figure 3. Electron density maps and models of the Se and Br modified nucleotides in the derivatized DNAs (Stereo view). Red and green balls represent Se and Br, respectively. Contours are at the 1.2σ level. (**A**) The structure of the 2′-Se-dU in d(GdU$_{Se}$GTACAC) at 1.28 Å resolution (1Z7I). (**B**) The structure of the 5-Br-dU in d(GdU$_{Se}$GdU$_{Br}$ACAC) at 1.50 Å resolution (2DLJ). (**C**) The structure of the 5-Br-dU in d(GTGdU$_{Br}$ACAC) at 1.80 Å resolution (2H05).
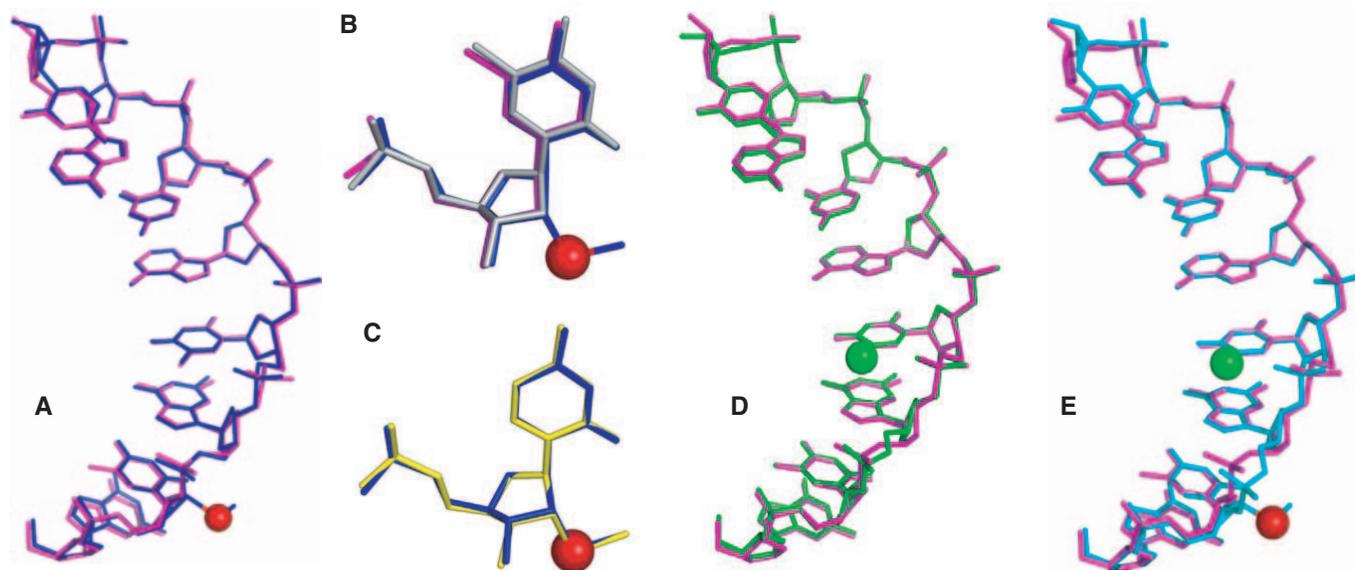
Figure 4. Superimposed comparison of the native and the derivatized DNA structures (red and green balls represent Se and Br, respectively). (**A**) The comparison of the Se-DNA (GdU$_{Se}$GTACAC, 1Z7I in blue) and the native (GTGTACAC, 1DNS in pink). (**B**) The comparison of the Se-DNA dU$_{se}$ (1Z7I in blue) and the two native T2 (1DNS in pink and 1D78 in gray). (**C**) The comparison of the dU$_{se}$ of the Se-DNA (1Z7I in blue) and the reported DNA (GCGTAdU$_{Se}$ACGC, 1MA8 in yellow). (**D**) The comparison of the Br-DNA (GTGdU$_{Br}$ACAC, 2H05 in green) and the native (1DNS in pink). (**E**) The comparison of the Se-Br-DNA (GdU$_{Se}$GdU$_{Br}$ACAC, 2DLJ in cyan) and the native (1DNS in pink).
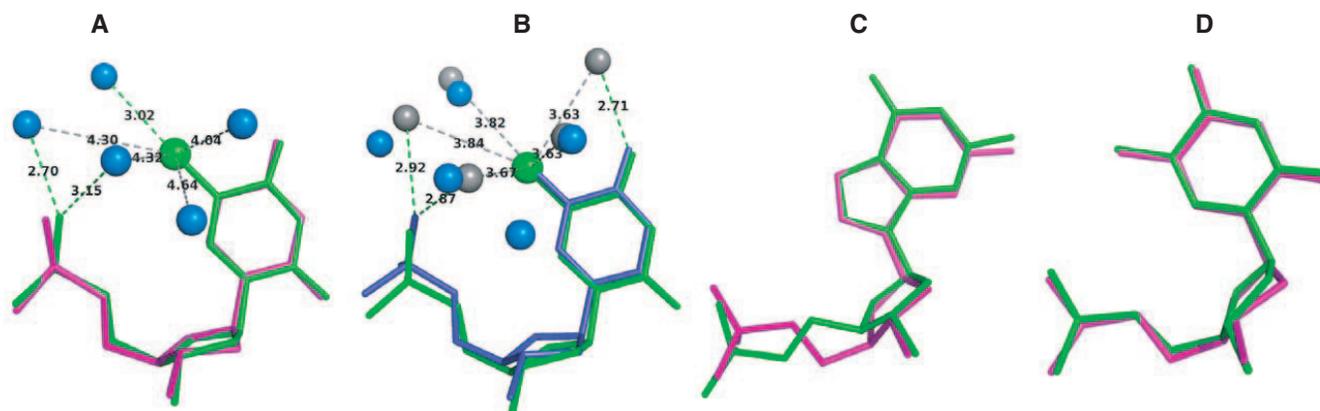
Figure 5. The superimposed comparison of the native and Br-DNA local structures (green ball represents Br). The green and gray dash lines represent hydrogen and non-hydrogen bonds, respectively. (**A**) The comparison of the dU$_{Br}$ of the Br-DNA (GTGdU$_{Br}$ACAC, 2H05 in green, and water molecules in blue balls) and the T4 of the native DNA (GTGTACAC, 1DNS in pink). (**B**) The comparison of the dU$_{Br}$ of the Br-DNA (2H05 in green, and its water molecules in blue balls) and the T4 of the Se-DNA (GdU$_{Se}$GTACAC, 1Z7I in blue, its water molecules in gray). (**C**) The comparison of the dG3 of the Br-DNA (in green) and the native (1DNS in pink). (**D**) The comparison of the T2 of the Br-DNA (in green) and the native (1DNS in pink).
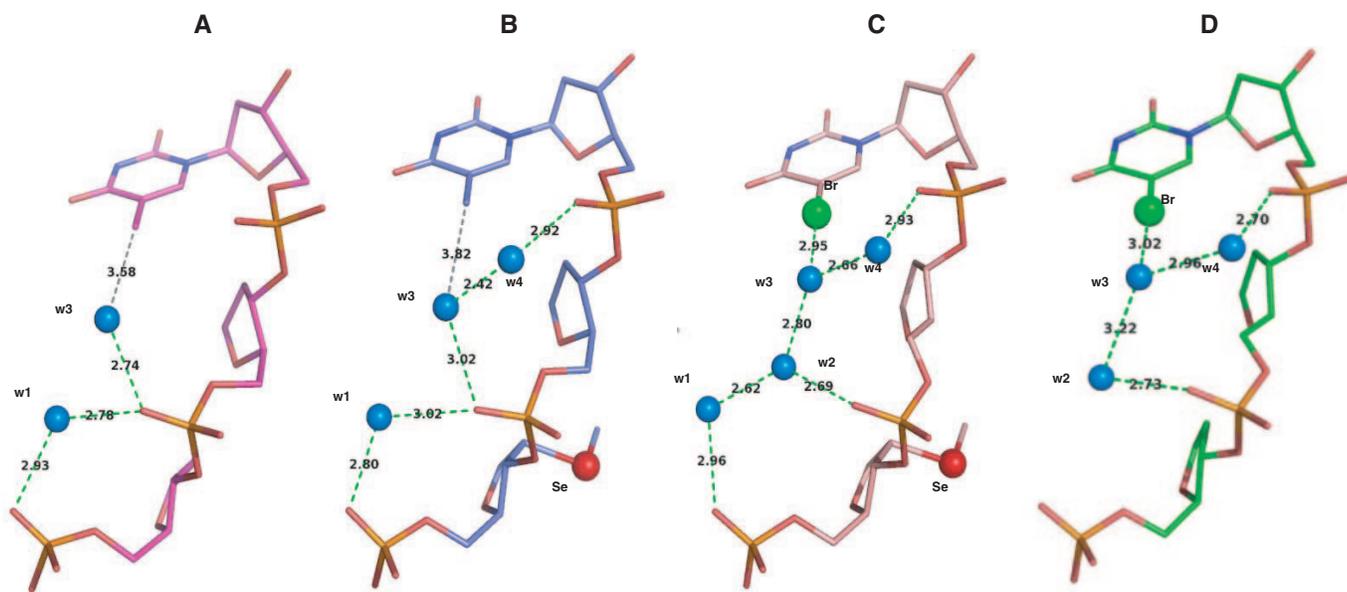
Figure 6. The bromine hydration pattern and the local structures. Water molecules, Se, and Br are in blue, red and green balls, respectively. The green and gray dash lines represent hydrogen and non-hydrogen bonds, respectively. (**A**) The TGT backbone structure of the native DNA (GTGTACAC, 1D78). (**B**) The dU$_{Se}$GT backbone structure of the Se-DNA (GdU$_{Se}$GTACAC, 1Z7I). (**C**) The dU$_{Se}$GdU$_{Br}$ backbone structure of the Se-Br-DNA structure (GdU$_{Se}$GdU$_{Br}$ACAC, 2GPX). (**D**) The TGdU$_{Br}$ backbone structure of the Br-DNA (GTGdU$_{Br}$ACAC, 2H05).

Table 4. Summary of the atom deviations [the Br-DNA (2H05) versus the native structure (1DNS)]

Atom (Å)	$C_{2'}$	$C_{4'}$	$C_{5'}$	$O_{5'}$	P
dU$_{Br}$ versus T4	0.25	0.11	0.24	0.08	0.08
dG3 versus dG3	0.27	0.53	1.30	1.01	0.63
T2 versus T2	0.60	0.08	0.30	0.09	0.22

Local perturbations and hydration patterns

The native d(GTGTACAC) structure in tetragonal space group [1DNS, (33)] did not include any water molecules in the refined structure, while the other native structure in hexagonal space group [1D78, (34)], determined at 1.4 Å resolution, has included 37 waters. The TGT local structures of the tetragonal native DNA (GTGTACAC), the hexagonal native, and the Se-DNA are identical (Figures 4A, 6A and B). Instead of comparing with those of the tetragonal native structure due to the absence of the water molecules, we compared the water networking and local backbone structure of the Se-DNA structure with those of the hexagonal native structure. The water molecules were found forming hydrogen bonds to the oxygen atoms of the 5'-phosphate groups of TGT in the native and Se-derivatized structures (Figure 6A and B). In the hexagonal native structure (1D78, Figure 6A), two water molecules (W1 and W3) formed hydrogen bonds (the green dash lines) to the oxygen atoms of the T2 and dG3 5'-phosphate groups. In the Se-DNA structure (1Z7I, Figure 6B), three water molecules (W1, W3 and W4) formed the hydrogen bonds to each other and to the oxygen atoms of the dU$_{Se}$, dG3 and T4 5'-phosphate groups, forming the same water networking as the native. In the Se-Br-DNA structure (2GPX in Figure 6C), four water molecules (W1, W2, W3 and W4) were found in the region, connecting with each

other and the oxygen atoms of the dU$_{Se}$, dG3 and T4 5'-phosphate groups. These four water molecules formed a different water networking from those of the native and Se-DNA structures. Similar to the Se-Br-DNA structure (Figure 6C), the same water networking (W2, W3 and W4) was found in the Br-DNA (GTGdU$_{Br}$ACAC, 2H05 and Figure 6D). Interestingly, in both the Br-DNA and Se-Br-DNA structures, the W2 water insertion, which can probably facilitate the 5'-phosphate perturbation and the $C_{4'}$-$C_{5'}$ bond rotation of the dG3, was observed. This inserted water (W2) formed the hydrogen bonds with W1, the dG3 5'-phosphate oxygen, and W3, which was pushed closer to the dU$_{Br}$ bromine atom in both structures. In both the Br-DNA structure (Figure 6D) and the Se-Br-DNA structure (Figure 6C), we observed that ordered W3 molecules are very close to the bromine atom (3.02 Å and 2.95 Å, respectively), which suggests a hydrogen bond formation. This hydrogen bond may weaken the Br-C5 bond, which probably is one reason of the debromination reported previously (12).

Our experimental results indicate that though it does not cause the change in the sugar pucker of the dU$_{Br}$, the Br derivatization placed in major groove caused significant conformation changes of the sugar puckers in the dG3 and T2 and the large changes of the hydration pattern and water molecule networking, which may lead to the local perturbation. Unlike the bromine derivatization, the 2'-selenium derivatization is placed in the minor groove, and the hydration alteration in the minor groove was not observed, probably due to disorder of the water molecules near the 2'-position of T2. No ordered water molecules near the 2'-positions in the Se-derivatized and native structures (33,34) have been observed. Probably, the hydration alteration caused by the 2'-selenium modification is insignificant.

CONCLUSIONS

To facilitate X-ray crystal structure studies of nucleic acids, we are in a process of developing selenium derivatization strategies by synthesizing SeNA via atom-specific replacement of oxygen with selenium and studying SeNA structures by comparing them with the structure-known natives and the conventional bromine derivatization. By using both MAD and SAD phasing, we have demonstrated the X-ray crystal structure determination of the 2′-Se-derivtized DNA at high-resolution. The Se-derivatized DNA structure (1.28 Å resolution) is identical to the native structure with the same crystal form (33, 2.0 Å resolution). Besides the demonstration of the formation of these Se-derivatized crystals with high-diffraction quality and in broader buffer conditions, we also demonstrated that these Se-derivatizated DNAs crystallized faster than these native and Br-derivatized DNAs. Furthermore, by comparing with the conventional bromine derivative, which retained the same global structure as the native, we found that the bromine derivatization caused the local perturbations, such as the backbone rotation, the water networking disruption, and the hydrate pattern alteration. Our experimental results indicate that the selenium derivatization may assist the studies of the nucleic acid X-ray crystal structures in two major ways: phasing and crystal growth with high-quality, which leads to structure determination with high-resolution. These experimental results suggest that the selenium derivatization can be an alternative to the conventional bromine derivatization. This Se derivatization strategy via the atom-specific substitution will significantly facilitate X-ray crystal structure studies of nucleic acids and their protein complexes.

REFERENCES

1. MacRae,I.J., Zhou,K., Li,F., Repic,A., Brooks,A.N., Cande,W.Z., Adams,P.D. and Doudna,J.A. (2006) Structural basis for double-stranded RNA processing by Dicer. *Science*, **311**, 195–198.
2. Jacobs,S.A., Podell,E.R. and Cech,T.R. (2006) Crystal structure of the essential N-terminal domain of telomerase reverse transcriptase. *Nature Struct. Mol. Biol.*, **13**, 218–225.
3. Adams,P.L., Stahley,M.R., Kosek,A.B., Wang,J. and Strobel,S.A. (2004) Crystal structure of a self-splicing group I intron with both exons. *Nature*, **430**, 45–50.
4. Parkinson,G.N., Lee,M.P.H. and Neidle,S. (2002) Crystal structure of parallel quadruplexes from human telomeric DNA. *Nature*, **417**, 876–880.
5. Peersen,O.B., Ruggles,J.A. and Schultz,S.C. (2002) Dimeric structure of the *Oxytricha nova* telomere end-binding protein alpha-subunit bound to ssDNA. *Nature Struct. Biol.*, **9**, 182–187.
6. Lu,M. and Steitz,T.A. (2000) Structure of *Escherichia coli* ribosomal protein L25 complexed with a 5S rRNA fragment at 1.8 A resolution. *Proc. Natl Acad. Sci. USA*, **97**, 2023–2028.
7. Cate,J.H., Gooding,A.R., Podell,E., Zhou,K., Golden,B.L., Szewczak,A.A., Kundrot,C.E., Cech,T.R. and Doudna,J.A. (1996) RNA tertiary structure mediation by adenosine platforms. *Science*, **273**, 1696–1699.
8. Ogata,C.M., Hendrickson,W.A., Gao,X. and Patel,D. (1989) Crystal structure of chromomycin–DNA complex. *J. Abstr. Am. Cryst. Assoc. Mtg. Ser.*, **217**, 53.
9. Xiong,Y. and Sundaralingam,M. (2000) Crystal structure of a DNA.RNA hybrid duplex with a polypurine RNA r(gaagaagag) and a complementary polypyrimidine DNA d(CTCTTCTTC). *Nucleic Acids Res.*, **28**, 2171–2176.
10. Ikehara,M., Uesugi,S. and Kaneko,M. (1978) Direct bromination of adenine nucleosides and nucleotides: 8-Bromoadenosine and its 5′-phosphate. In Townsend,L.B. and Tipson,R.S. (eds), *Nucleic Acid Chemistry*. Wiley, NY, Vol. **2**, pp. 837–841.
11. Tavale,S.S. and Sobell,H.M. (1970) Crystal and molecular structure of 8-bromoguanosine and 8-bromoadenosine, two purine nucleosides in the syn conformation. *J. Mol. Biol.*, **48**, 109–123.
12. Ennifar,E., Carpentier,P., Ferrer,J.-L., Walter,P. and Dumas,P. (2002) X-ray-induced debromination of nucleic acids at the Br k absorption edge and implication for MAD phasing. *Acta Crystallogrphica*, **D58**, 1262–1268.
13. Willis,M.C., Hicke,B.J., Uhlenbeck,O.C., Cech,T.R. and Koch,T.H. (1993) Photocrosslinking of 5-iodouracil-substituted RNA and DNA to proteins. *Science*, **262**, 1255–1257.
14. 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.
15. Yang,W., Hendrickson,W.A., Crouch,R.J. and Satow,Y. (1990) Structure of ribonuclease H phased at 2-A resolution by MAD analysis of the selenomethionyl protein. *Science*, **249**, 1398–1405.
16. Hendrickson,W.A., Horton,J.R. and LeMaster,D.M. (1990) Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure. *EMBO J.*, **9**, 1665–1672.
17. Hendrickson,W.A. (1991) Determination of macromolecular structures from anomalous diffraction of synchrotron radiation. *Science*, **254**, 51–58.
18. Hendrickson,W.A. (2001) Synchrotron crystallography. *Trends Biochem. Sci.*, **25**, 637–643.
19. Jiang,J.-S. and Sweet,R.M. (2004) Protein data bank depositions from synchrotron sources. *J. Synchrotron Rad.*, **11**, 319–327.
20. Ferre-D'Amare,A.R., Zhou,K. and Doudna,J.A. (1998) Crystal structure of a hepatitis delta virus ribozyme. *Nature*, **395**, 567–574.
21. Carrasco,N., Ginsburg,D., Du,Q. and Huang,Z. (2001) Synthesis of selenium-derivatized nucleosides and oligonucleotides for X-ray crystallography. *Nucleosides, Nucleotides and Nucleic Acids*, **20**, 1723–1734.
22. 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.
23. Teplova,M., Wilds,C.J., 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.
24. Wilds,C.J., Pattanayek,R., Pan,C., Wawrzak,Z. and Egli,M. (2002) Selenium-assisted nucleic acid crystallography: use of DNA phosphoroselenoates for MAD phasing. *J. Am. Chem. Soc.*, **124**, 14910–14916.
25. Buzin,Y., Carrasco,N. and Huang,Z. (2004) Synthesis of selenium-derivatized cytidine and oligonucleotides for X-ray crystallography using MAD. *Organic Lett.*, **6**, 1099–1102.
26. Carrasco,N., Buzin,Y., Tyson,E., Halpert,E. and Huang,Z. (2004) Selenium derivatization and crystallization of DNA and RNA oligonucleotides for X-ray crystallography using multiple anomalous dispersion. *Nucleic Acids Res.*, **32**, 1638–1646.

27. Salon,J., Chen,G., Portilla,Y., Germann,M.W. and Huang,Z. (2005) Synthesis of a novel 2′-Se-uridine phosphoramidite and its incorporation into oligonucleotides for structure study. *Organic Lett.*, **7**, 5645–5648.

28. 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.

29. Serganov,A., Yuan,Y., Pikovskaya,O., Polonskaia,A., Malinina,L., Phan,A., Hobartner,C., Micura,R., Breaker,R. and Patel,D. (2004) Structural basis for discriminative regulation of gene expression by adenine- and guanine-sensing mRNAs. *Chem. Biol.*, **11**, 1729–1741.

30. Serganov,A., Keiper,S., Malinina,L., Tereshko,V., Skripkin,E., Hobartner,C., Polonskaia,A., Phan,A.T., Wombacher,R., Micura,R. *et al.* (2005) Structural basis for Diels-Alder ribozyme-catalyzed carbon-carbon bond formation. *Nature Struct. Mol. Biol.*, **12**, 218–224.

31. Carrasco,N. and Huang,Z. (2004) Enzymatic synthesis of phosphoroselenoate DNA using thymidine 5′-(α-P-Seleno)-triphosphate and DNA polymerase for X-ray crystallography using MAD. *J. Am. Chem. Soc.*, **126**, 448–449.

32. Carrasco,N., Caton-Williams,J., Brandt,G., Wang,S. and Huang,Z. (2006) Efficient enzymatic synthesis of phosphoroselenoate RNA using adenosine 5′-(α-P-seleno)triphosphate. *Angew. Chem.*, **118**, 100–103.

33. Jain,S., Zon,G. and Sundaralingam,M. (1989) Base only binding of spermine in the deep groove of A-DNA octomer d(GTGTACAC). *Biochemistry*, **28**, 2360–2364.

34. Thota,N., Li,X.H., Bingman,C. and Sudaralingam,M. (1993) High-resolution refinement of the hexagonal A-DNA octomer d(GTGTACAC) at 1.4 Å. *Acta Cryst.* **D49**, 282–291.

35. Otwinowski,Z. and Minor,W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Meth. Enzymol.*, **276**, 307–326.

36. Brűnger,A.T., Adams,P.D., Clore,G.M., DeLano,W.L., Gros,P., Grosse-Kunstleve,R.W., Jiang,J.-S., Kuszewski,J., Nilges,N., Pannu,N.S. *et al.* (1998) Crystallography and NMR system (CNS): a new software system for macromolecular structure determination. *Acta Cryst.*, **D54**, 905–921.

37. Jones,T.A., Zou,J.Y., Cowan,S.W. and Kjeldgaard,M. (1991) Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta. Cryst.* **A47**, 110–118.

38. Parkinson,G., Vojtechovsky,J., Clowney,L., Brunger,A.T. and Berman,H.M. (1996) New parameters for refinement of nucleic acid containing structures. *Acta Cryst.*, **D52**, 57–64.

39. Brűnger,A.T. (1992) The free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature*, **355**, 472–475.

40. Read,R.J. (1986) Improved fourier coefficients for maps using phases from partial structures with errors. *Acta Cryst.*, **A42**, 140–149.

41. Bingman,C.A., Zon,G. and Sundaralingam,M. (1992) Crystal and molecular structure of the A-DNA dodecamer d(CCGTACGTACGG). Choice of fragment helical axis. *J. Mol. Biol.*, **227**, 738–756.

42. Langlois,E.B., Dautant,A., Courseille,C. and Precigoux,G. (1993) Orthorhombic crystal structure of the A-DNA octamer d(GTACGTAC). Comparison with the tetragonal structure. *Eur. J. Biochem.*, **213**, 673–682.