Selenium derivatization and crystallization of DNA and RNA oligonucleotides for X-ray crystallography using multiple anomalous dispersion

Nicolas Carrasco, Yuri Buzin, Elizabeth Tyson, Elizer Halpert and Zhen Huang*

Department of Chemistry, Brooklyn College, and Program of Biochemistry and Chemistry, The Graduate School, The City University of New York, 2900 Bedford Avenue, Brooklyn, NY 11210, USA

Received January 16, 2004; Revised and Accepted February 13, 2004

ABSTRACT

We report here the solid phase synthesis of RNA and DNA oligonucleotides containing the 2'-selenium functionality for X-ray crystallography using multiwavelength anomalous dispersion. We have synthesized the novel 2'-methylseleno cytidine phosphoramidite and improved the accessibility of the 2'-methylseleno uridine phosphoramidite for the synthesis of many selenium-derivatized DNAs and RNAs in large scales. The yields of coupling these Se-nucleoside phosphoramidites into DNA or RNA oligonucleotides were over 99% when 5-(benzylmercapto)-1H-tetrazole was used as the coupling reagent. The UV melting study of A-form dsDNAs indicated that the 2'-selenium derivatization had no effect on the stability of the duplexes with the 3'-endo sugar pucker. Thus, the stems of functional RNA molecules with the same 3'-endo sugar pucker appear to be the ideal sites for the selenium derivatization with 2'-Se-C and 2'-Se-U. Crystallization of the selenium-derivatized oligonucleotides is also reported here. The results demonstrate that this 2'-selenium functionality is suitable for RNA and A-form DNA derivatization in X-ray crystallography.

INTRODUCTION

Nucleic acids play a variety of important roles in biological systems, including the transfer and regulation of genetic information, and the RNA catalytic functions found in vivo in protein synthesis and in the life cycle of some viruses (1,2). RNAs and DNAs with catalytic and binding functions have also been identified via in vitro selection (3). Furthermore, the recent discovery of noncoding small RNAs in diverse organisms has enormously expanded the repertoire of functions of nucleic acids (4,5). This vast array of biologically active RNAs and DNAs has promoted a new front of research into the field of structural analysis to elucidate their three-dimensional (3D) structure and function relationships. To date, in combination with synchrotron radiation, multiwavelength anomalous dispersion (MAD) and new methodologies in introducing anomalously scattering atoms into molecules, X-ray crystallography has become the most powerful and widely used technique for the 3D structure determination of proteins, DNAs and RNAs (6–9). The area of protein X-ray crystallography has been revolutionized by replacing methionine with selenomethionine (10,11). The use of selenomethionyl proteins has proved to be excellent for MAD phasing by the exploitation of the unique absorption K edge of selenium (12.6578 keV, 0.9795 Å), which is readily accessible by synchrotron radiation (7,11).

On the other hand, DNA and RNA X-ray structure determination has been more challenging due to inherent difficulties in creating suitable heavy atom derivatives for MAD phasing. Currently, there are three conventional methods for creating heavy atom derivatives for nucleic acids: (i) soaking and co-crystallization with heavy metal ions; (ii) RNA derivatization via selenomethionyl proteins; (iii) derivatization with halogens. Soaking and co-crystallization have not been very successful in nucleic acid X-ray crystallography, as opposed to the scenario in proteins, probably because of nonspecific binding of heavy metal ions and backbone cleavage by the ions. The derivatization of RNAs using the selenomethionyl protein U1A has been recently explored by Doudna’s research group. Although it has led to the successful elucidation of two ribozyme structures by X-ray crystallography (12,13), this method is labor-intensive. In addition, the utilization of selenomethionyl proteins that bind intrinsically to RNAs has facilitated the elucidation of a few other RNA crystal structures (14,15). Similarly, halogens (bromine or iodine) have been pursued as potential X-ray scattering centers in nucleic acids. The phosphoramidites of 5-halogen-uridine, 5-halogen-2'-deoxyuridine (a mimic of thymidine) and 5-halogen-2'-deoxycytidine have been synthesized for the halogen derivatization of DNA and RNA oligonucleotides via solid phase synthesis (16). However, there are two major problems associated with halogen derivatives. First, since the chemical incorporation of halogens is primarily limited to the 5-position of pyrimidine nucleosides, this lack of multiple choices of positions may limit derivatization freedom to avoid base-stacking disruptions, other structural perturbations, and possible crystalizability problems. Secondly, recent reports have indicated...
that the light sensitivity of the halogen derivatives can lead to their decomposition after long-time exposure to X-ray (17,18).

In an effort to increase the repertoire of anomalously scattering atoms for MAD phasing in nucleic acid X-ray crystallography, Huang, Egli and co-workers have undertaken chemical and enzymatic approaches to covalently introduce selenium into DNAs and RNAs (19–23). We report here the derivationization of biological function-related DNAs and RNAs using our novel 2'-methylseleno cytidine phosphoramidite and the previously reported 2'-methylseleno uridine phosphoramidite (20). We also present major improvements in the synthesis of 2'-methylseleno uridine phosphoramidite as well as new advances in the highly efficient synthesis of Se-DNA and Se-RNA oligonucleotides using 5-(benzylmercapto)-1H-tetrazole (5-BMT) as the coupling reagent. Crystallization of some selenium-derivatized oligonucleotides is also described.

**MATERIALS AND METHODS**

**Synthesis of the 2'-selenium-derivatized uridine and cytidine phosphoramidites**

2,2'-O-anhydro-1-(β-d-arabinofuranosyl)-uracil (2). Uridine (50 g, 201 mmol) and diphenyl carbonate (48 g, 224 mmol) were placed in a round flask, and N,N-dimethylformamide (50 ml) was added. The slurry was heated in an oil bath at 100°C. Dry sodium bicarbonate (400 mg) was then added, and a watch glass was used to cover the flask. The reaction mixture was stirred at 0°C and monitored by TLC (methanol/methylene chloride, 2:8). After completion, the reaction was cooled to room temperature, filtered, and washed by methanol three times (each time 15 ml). The white powdered product was dissolved on high vacuum overnight (37.9 g, 82%; wt 226). 1H-NMR (CDCl₃) δ (p.p.m.): 2.99–3.18 (m, 2H, H-5'), 3.69 (s, 6H, 2×CH₂O), 4.29–4.36 (m, 1H, H-4'), 4.42–4.46 (m, 1H, H-3'), 5.20–5.25 (m, 1H, H-2'), 5.92 (d, J = 7.5 Hz, 1H, H-5), 6.07 (d, J = 5.7 Hz, 1H, H-1'), 6.98–6.79 (m, 4H, Ar-H), 7.11–7.30 (m, 10H, H-6, 9 Ar-H).

13C-NMR (CDCl₃) δ (p.p.m.): 55.20 (OCH₃), 62.96 (C-5'), 75.60 (C-3'), 86.16 (Ar-C), 87.64 (C-4'), 89.42 (C-2'), 90.31 (C-1'), 109.64 (C-5), 113.18, 126.91, 127.91, 129.83, 135.40, 144.39, 158.45 (Ar-C), 135.93 (C-6), 159.75 (C-2'), 172.69 (C-4'). FAB-HRMS: C₃₉H₅₉N₅O₇ (M⁺), 529.1976 (calc. 529.1974).

2,2'-O-(4,4-dimethoxytrityl)-2'-methylseleno-2'-deoxyuridine (4). NaN₃ (1.1 g, 6 mmol) was placed in a 250 ml round flask and suspended in dry THF (45 ml) under vigorous stirring. Dimethyl diselenide (CH₃SeSeCH₃, 1.99 ml, 19.9 mmol) was slowly injected, and the suspension was placed in an ice-water bath under dry argon. Anhydrous ethanol (5 ml) was added dropwise. Gas bubbles started to occur in the yellow mixture. The reaction mixture turned colorless after 60 min, and a solution of 3 (10.5 g, 19.9 mmol) in THF (20 ml) was injected to the flask. The reaction mixture turned colorless after 60 min, and a solution of 3 (10.5 g, 19.9 mmol) in THF (20 ml) was injected to the flask. The reaction mixture turned colorless after 60 min, and a solution of 3 (10.5 g, 19.9 mmol) in THF (20 ml) was injected to the flask. The reaction mixture turned colorless after 60 min, and a solution of 3 (10.5 g, 19.9 mmol) in THF (20 ml) was injected to the flask. The reaction mixture turned colorless after 60 min, and a solution of 3 (10.5 g, 19.9 mmol) in THF (20 ml) was injected to the flask. The reaction mixture turned colorless after 60 min, and a solution of 3 (10.5 g, 19.9 mmol) in THF (20 ml) was injected to the flask. The reaction mixture turned colorless after 60 min, and a solution of 3 (10.5 g, 19.9 mmol) in THF (20 ml) was injected to the flask. The reaction mixture turned colorless after 60 min, and a solution of 3 (10.5 g, 19.9 mmol) in THF (20 ml) was injected to the flask. The reaction mixture turned colorless after 60 min, and a solution of 3 (10.5 g, 19.9 mmol) in THF (20 ml) was injected to the flask. The reaction mixture turned colorless after 60 min, and a solution of 3 (10.5 g, 19.9 mmol) in THF (20 ml) was injected to the flask. The reaction mixture turned colorless after 60 min, and a solution of 3 (10.5 g, 19.9 mmol) in THF (20 ml) was injected to the flask. The reaction mixture turned colorless after 60 min, and a solution of 3 (10.5 g, 19.9 mmol) in THF (20 ml) was injected to the flask. The reaction mixture turned colorless after 60 min, and a solution of 3 (10.5 g, 19.9 mmol) in THF (20 ml) was injected to the flask. The reaction mixture turned colorless after 60 min, and a solution of 3 (10.5 g, 19.9 mmol) in THF (20 ml) was injected to the flask. The reaction mixture turned colorless after 60 min, and a solution of 3 (10.5 g, 19.9 mmol) in THF (20 ml) was injected to the flask. The reaction mixture turned colorless after 60 min, and a solution of 3 (10.5 g, 19.9 mmol) in THF (20 ml) was injected to the flask. The reaction mixture turned colorless after 60 min, and a solution of 3 (10.5 g, 19.9 mmol) in THF (20 ml) was injected to the flask. The reaction mixture turned colorless after 60 min, and a solution of 3 (10.5 g, 19.9 mmol) in THF (20 ml) was injected to the flask. The reaction mixture turned colorless after 60 min, and a solution of 3 (10.5 g, 19.9 mmol) in THF (20 ml) was injected to the flask. The reaction mixture turned colorless after 60 min, and a solution of 3 (10.5 g, 19.9 mmol) in THF (20 ml) was injected to the flask. The reaction mixture turned colorless after 60 min, and a solution of 3 (10.5 g, 19.9 mmol) in THF (20 ml) was injected to the flask.

**UV (in acetonitrile), λmax: 233.8, 281.2 nm. IR (KBr): 3450 (br), 3080, 3030, 2940, 1705, 1610, 1580, 1470, 1410, 1320, 1270, 1260, 1190, 1095, 1055, 820, 770, 705, 580 cm⁻¹.**

1H-NMR (CDCl₃) δ (p.p.m.): 2.99–3.18 (m, 2H, H-5'), 3.69 (s, 6H, 2×CH₂O), 4.29–4.36 (m, 1H, H-4'), 4.42–4.46 (m, 1H, H-3'), 5.20–5.25 (m, 1H, H-2'), 5.92 (d, J = 7.5 Hz, 1H, H-5), 6.07 (d, J = 5.7 Hz, 1H, H-1'), 6.98–6.79 (m, 4H, Ar-H), 7.11–7.30 (m, 10H, H-6, 9 Ar-H).

13C-NMR (CDCl₃) δ (p.p.m.): 55.20 (OCH₃), 62.96 (C-5'), 75.60 (C-3'), 86.16 (Ar-C), 87.64 (C-4'), 89.42 (C-2'), 90.31 (C-1'), 109.64 (C-5), 113.18, 126.91, 127.91, 129.83, 135.40, 144.39, 158.45 (Ar-C), 135.93 (C-6), 159.75 (C-2'), 172.69 (C-4'). FAB-HRMS: C₃₉H₅₉N₅O₇ (M⁺), 529.1976 (calc. 529.1974).
under dry nitrogen for 30 min and the ice bath was then removed. The mixture was further stirred for 2 h at room temperature. Reaction completion was indicated by TLC [5% CH$_2$OH/CH$_2$Cl$_2$, product $R_f = 0.37$]. The reaction mixture was then quenched with NaHCO$_3$ (20 ml, sat.), stirred for 15 min, and extracted with CH$_2$Cl$_2$ (3 × 100 ml). The combined organic layer was washed with NaCl (100 ml, sat.) and dried over anhydrous MgSO$_4$ for 15 min, followed by filtration and solvent evaporation. The crude product was re-dissolved in CH$_2$Cl$_2$ (20 ml), and this solution was added dropwise to petroleum ether (1000 ml) under vigorous stirring; a white precipitate was formed. The petroleum ether solution was decanted carefully (sometimes filtration was necessary). The crude product was then loaded into a silica gel column that was equilibrated with 20% EtOAc/hexane containing 0.5% triethylamine. The column was eluted with increasing stepwise gradient of EtOAc/hexane in the presence of 0.5% triethylamine (20, 30, 40, 50, and 60%, 400 ml each). The pooled fractions containing the pure compound were re-dissolved in CH$_2$Cl$_2$, transferred into a small round flask, and evaporated under reduced pressure, and re-dissolved in 20 ml of CH$_2$Cl$_2$. This solution was precipitated and extracted with CH$_2$Cl$_2$ (3 × 100 ml). The combined organic layer was washed with saturated NaCl (100 ml). The combined organic layer was washed with saturated NaCl (100 ml) and dried over MgSO$_4$ (s) before evaporation. The crude product was purified on silica gel column chromatography with hexane/CH$_2$Cl$_2$ 1:1. The column was eluted with hexane/CH$_2$Cl$_2$, 1:1, 1:3, and pure CH$_2$Cl$_2$, and then with a methanol/CH$_2$Cl$_2$ gradient (0.5, 1, 2, 3 and 4% methanol in CH$_2$Cl$_2$) to afford product 9 (2.13 g, Fw 622.6) as a white foam (86% yield). $^1$H-NMR (CDCl$_3$) δ: 1.2 (br, 1H, OH), 1.99 (s, 3H, CH$_3$Se), 3.38 and 3.46 (2x dd, J = 1.8, 9.6 Hz, 2H, H-5'), 3.51 (dd, J = 4.5, 4.8 Hz, 1H, H-2'), 3.68 (s, 6H, CH$_3$O), 4.13–4.17 (m, 1H, H-4'), 4.44–4.47 (m, 1H, H-3'), 5.40 (d, J = 7.5 Hz, 1H, H-5), 6.32 (d, 1H, J = 5.9 Hz, H-1'), 6.76–6.82 (d, 2J = 5.9 Hz, 4H, aromatic), 7.09–7.31 (m, 9H, aromatic), 7.87 (d, J = 7.5 Hz, H-1, H-6), 12.1 (br, 2H, NH$_2$). $^{13}$C-NMR (CDCl$_3$) δ: 5.25 (SeCH$_3$), 51.56 (C-2'), 56.23 (OCH$_3$), 64.6 (C-5'), 72.04 (C-3'), 85.52 (C-4'), 86.97 (Ar-C), 88.76 (C-1'), 114.23, 128.84, 129.08, 129.60, 130.46, 133.73, 136.95, 142.23 (C-6), 144.05, 158.40 (Ar-C), 160.40 (C-2'), 167.26 (C-4'). HRMS (MALDI-FTMS): C$_9$H$_{13}$N$_3$O$_7$Se [M+Na]$^+$: 646.1429 (calc. 646.1427).

5'-O-(4,4-dimethoxytrityl)-2'-methylseleno-2'-deoxyoxycytosine (9a). THF (10 ml), dry triethylamine (1.11 ml, 8 mmol) and 1-(trimethylsilyl)imidazole (0.44 ml, 3 mmol) were added to a 25 ml round flask containing compound 9 (dry 1.25 g, 2 mmol) under argon at room temperature. After 15 min of stirring, a catalytic amount of N,N-dimethylamino-pyridine (DMAP, 20 mg) and acetic anhydride (371 mg, 3.6 mmol) was added. The reaction mixture was stirred for another 20 min. The solvents were evaporated under reduced pressure and the resultant residue was dissolved in EtOAc (50 ml). The precipitated salt was removed by filtration and the filtrate was evaporated again under reduced pressure. This residue was dissolved in THF (10 ml), and tetrabutylammonium fluoride (4 ml, 1 M, 4 mmol) was added. The mixture was stirred for 1 h to remove the 3'-TMS group (monitored on silica gel TLC in 5% MeOH/CH$_2$Cl$_2$). After evaporation of THF, the crude product was dissolved in CH$_2$Cl$_2$ and purified on a silica gel column, equilibrated with hexane/CH$_2$Cl$_2$, 1:1. The column was first eluted with hexane/CH$_2$Cl$_2$, 1:1, 1:3, and pure CH$_2$Cl$_2$, and then with a methanol/CH$_2$Cl$_2$ gradient (0.5, 1, 2 and 3% methanol in CH$_2$Cl$_2$) to afford the desired product 9a (1.26 g, Fw 664.6) as a white foam (95% yield). $^1$H-NMR (CDCl$_3$) δ: 2.15 (s, 3H, CH$_3$Se), 2.23 (s, 3H, CH$_3$CO), 3.29 (br, 1H, OH), 3.49 and 3.57 (2x dd, J = 1.9, 9.8 Hz, 2H, H-5'), 3.68 (dd, J = 4.4, 4.7 Hz, 1H, H-2'), 3.79 (s, 6H, CH$_3$O), 4.16–4.27 (m, 1H, H-4'), 4.42–4.53 (m, 1H, H-3'), 6.34 (d, 1H, J = 6.1 Hz, H-1'), 6.81–6.95 (d, J = 6.1 Hz, 4H, aromatic), 7.19 (d, J = 7.5 Hz, 1H, H-5'), 7.26–7.52 (m, 9H, aromatic), 8.32 (d, J = 7.5 Hz, 1H, H-6), 9.89 (br, 1H, NH). $^{13}$C-NMR (CDCl$_3$) δ: 4.58 (SeCH$_3$), 24.79 (CH$_3$CO), 46.23 (C-2'), 55.23 (OCH$_3$), 62.23 (C-5'), 69.92 (C-3'), 84.47 (C-4'), 87.21 (Ar-C), 90.26 (C-1'),
N\textsuperscript{4}-acetyl-3'-O-(2-cyanoethyl-N,N-diisopropylphosphoramidite)-5'-O-dimethoxytrityl-2'-methylseleno-2'-deoxy-cytosine (10). See the synthesis of compound 5. A white foamy product resulted (Fw: 864.8: 1.2 g, 92%). \textsuperscript{1}H-NMR (CDCl\textsubscript{3}) \( \delta \): 1.04–1.35 (m, 24H, 8x CH\textsubscript{3}ipr), 2.13 (s, 6H, 2x CH\textsubscript{2}CO), 2.16 and 2.19 (2x s, 6H, 2x CH\textsubscript{2}Se), 2.41 and 2.67 (2x t, \textit{J} = 7.5 Hz, 4H, 2x CH\textsubscript{2}CN), 3.45–3.78 (m, 12H, 4x CH\textsubscript{3}O), 3.92–4.03 (m, C\textsubscript{6}H\textsubscript{5}O\textsubscript{C}), 4.36–4.41 (m, 2H, H-2\textsuperscript{n}), 4.65–4.71 (m, 2H, H-3\textsuperscript{n}), 6.44 (d, 2H, \textit{J} = 5.1 Hz, H-1\textsuperscript{a}), 6.82–6.94 (d, \textit{J} = 8.7 Hz, 8H, aromatic), 7.09 and 7.14 (2x \textit{d}, \textit{J} = 7.5 Hz, 2H, H-5\textsuperscript{a}), 7.27–7.46 (m, 18H, aromatic), 8.31 and 8.35 (2x \textit{d}, \textit{J} = 7.5 Hz, 2H, H-6\textsuperscript{a}), 10.56 (br, 2H, NH). \textsuperscript{13}C-NMR (CDCl\textsubscript{3}) \( \delta \): 46.55 (SeCH\textsubscript{3}), 19.12, 19.19, 20.56, 22.69 and 22.96 (CH\textsubscript{3}CO), 24.77, 24.83, 43.34, 43.47, 45.32 and 45.38 (C-2\textsuperscript{f}), 46.77, 47.64, 47.55 (OCH\textsubscript{3}), 57.95, 58.34, 61.94 (C-5\textsuperscript{f}), 73.32 and 73.46 (C-3\textsuperscript{f}), 84.36 (C-4\textsuperscript{f}), 87.15 (Ar-C), 90.92 (C-1\textsuperscript{a}), 97.22 (C-5), 113.32, 117.25, 127.23, 127.99, 128.30, 130.09, 130.23, 135.14, 135.32, 144.16 (Ar-C), 144.46 (C-6\textsuperscript{f}), 155.39 (C-2\textsuperscript{f}), 158.81 (Ar-C), 163.09 (C-4\textsuperscript{f}), 171.20 (CO\textsubscript{2}Se). \textsuperscript{3}P-NMR (CDCl\textsubscript{3}) \( \delta \): 148.67, 149.05. HRMS (MALDI-FTMS): C\textsubscript{42}H\textsubscript{52}N\textsubscript{5}O\textsubscript{8}PSe \([\text{M+H}]^+\): 866.2795 (calc. 866.2791).

Synthesis of 2'-Se-functionalized RNA and DNA oligonucleotides

All DNA and RNA oligonucleotides were synthesized chemically on a 1.0 or 10 \( \mu \)mol scale using an ABI392 DNA/RNA Synthesizer (24–26). The concentration of the Se-nucleoside phosphoramidites was identical to that of the conventional phosphoramidites (0.1 M in acetonitrile). Coupling was performed on control pore glass (CPG-500) immobilized with trichloroacetic acid in methylene chloride. Syntheses were performed on a Zorbax SB-C18 column (4.6 \( \times \) 250 mm) at a flow rate of 1.0 ml/min using the same buffer system. The DMTr-on oligonucleotides were eluted with up to 90\% buffer B in 25 min in a linear gradient, while the DMTr-off oligonucleotides were eluted with up to 40\% buffer B in a linear gradient in the same period of time. The collected fractions were lyophilized; the purified compounds were redisolved in RNase-free water. The pH was adjusted to 7.0 after the final purification of the Se-oligonucleotides without the DMTr group.

Electrospray mass spectrometry analysis

Crude and purified oligonucleotides containing the selenium derivatization were analyzed by LC-MS using electrospray negative ion mode. The general analytical procedures for the liquid chromatography were elution (1 ml/min, 4.5 \( \times \) 150 mm 300SB-C8 column) with buffer A (5 \( \mu \)M ammonium acetate, pH 6.5) for 2 min, and then elution with a linear gradient from buffer A to 100\% buffer B (60\% acetonitrile and 40\% of buffer A) in 13 min.

Thermodenaturization of duplex DNAs

Solutions of the duplex DNAs (2 \( \mu \)M) were prepared by dissolving the DNAs in a buffer containing NaCl (90 \( \mu \)M), sodium phosphate (10 mM, pH 7.2), and EDTA (1 mM). The solutions were then heated to 95\°C for 2 min, cooled slowly to room temperature, and stored at 5\°C overnight before measurement. Prior to thermal denaturation, helium was bubbled through the samples. Denaturation curves were acquired at 254 nm at a heating rate of 0.5\°C/min using an 8453 UV-Visible Spectrometer from Agilent Technologies. This system is equipped with a Peltier Temperature Controller. The data were analyzed in accordance with the convention of Puglisi and Tinoco (28).

Crystallization of the Se-derivatized oligonucleotides

The purified oligonucleotides (2 \( \mu \)M) containing selenium labels were first heated to 90\°C for 1 min, and the samples were then allowed to slowly cool to 25\°C. Crystallization conditions were first screened using the nucleic acid screening kits from Hampton Research. To minimize the amount of material used, 1 or 2 \( \mu \)l of the appropriate oligonucleotide
solutions were typically used in each of these screens. Crystallization was carried out using the hanging drop method by vapor diffusion at 25 and 4°C.

RESULTS AND DISCUSSION

Synthesis of 2'-'methylseleno uridine and cytidine phosphoramidites

On the basis of the synthesis of 2'-methylseleno uridine phosphoramidite 5 (20), the corresponding 2'-methylseleno cytidine derivative was first attempted from cytidine derivative 7 (Scheme 1). This partially protected derivative was first mesylated at the 2'-position, followed by the displacement of the 2'-mesyl group with the cytosine exo-2-oxygen under basic conditions. To obtain compound 8 for the incorporation of the selenium functionality, the 3'-TBDMS group was removed using tetrabutylammonium fluoride. Unfortunately, a low yield in selenium incorporation was obtained when sodium methyl selenide was used to generate cytidine derivative 9. The poor substitution by sodium methyl selenide is probably because the exo-2-oxygen in cytosine is a much poorer leaving group in comparison to that in uridine. An alternative approach, via conversion of uridine derivative 4 to cytidine 9, was later explored successfully (29).

We reported previously the synthesis of 2'-methylseleno uridine phosphoramidite 5 (Scheme 1) using a relatively expensive uridine derivative (5'-O-DMTr-3'-O-TBDMS uridine). This synthesis required four steps to generate the key intermediate 4 (20). To improve the synthesis of this intermediate, we started the reaction using uridine, a relatively inexpensive starting material (29). Thus, intermediate 4 was made in three simple steps and the synthesis has been run at a 50 g scale. This uridine intermediate 4 was converted to cytidine derivative 9 via triazolide 4a that was generated in situ. After acetylation of the base, 2'-methylseleno cytidine phosphoramidite 10 was made as in the case of 5. This new advancement in the synthesis of intermediate 4 has facilitated the preparation of the 2'-methylseleno pyrimidine phosphoramidites in large scales.

Design of 2'-Se-derivatized RNAs and DNAs

Collaborative investigation with Egli and co-workers has previously demonstrated that the 2'-deoxyriboses containing the 2'-methylseleno functionality in the furanose ring displayed 3'-endo conformations in the crystal structure of the selenium-derivatized oligonucleotides, which is consistent with the geometry that is adopted by both RNA and A-form DNA (21). These studies indicated that this type of selenium

Scheme 1. (a) (Ph)2CO3, Na2CO3, DMF. (b) DMTr-Cl, Py. (c) NaSeCH3, EtOH-THF. (d) 2-Cyanomethyl N,N-diisopropyl-chlorophosphoramidite and N,N-diisopropylmethyamine in CH2Cl2. (e) Synthesis of oligonucleotides on solid phase. (f) MsCl, TEA, THF. (g) Toluene/tetrahexylammonium hydrogen sulfate, Na2CO3 (sat.). (h) (Bu)4N+ F-, THF. (i) TMS-Im, then Ac2O, TEA and DMAP in THF. (j) TMS-Im, then POCl3-triazole-TEA in CH2CN. (k) NH4OH.
derivatization may not significantly perturb the structure and stability of RNA and A-form DNA duplexes.

Consequently, we have derivatized a number of biologically and structurally important RNAs using this derivatization strategy for structural analysis. Though the 3D structure of an RNA molecule is difficult to foretell, its secondary structure is relatively easy to predict. In order to avoid structural perturbation, we decided to incorporate this 2'-Se derivatization into a stem region, which is a major element in the secondary structure of RNA.

To illustrate this principle, several RNA oligonucleotides were chosen and derivatized at specific positions through the solid phase synthesis. A 12mer RNA fragment of the HIV-1 Rev binding element (30) was selected and derivatized with both 2'-Se-U and 2'-Se-C at three different positions (Fig. 1A). To better understand G-U wobble base-pairing of RNA (31), (GGCGUGCC)₂ was selected for selenium derivatization at both U and C sites (Fig. 1B). Derivatization of this duplex RNA 8mer has been attempted previously using conventional approaches without success in the structure determination. Likewise, other RNAs were selected for the selenium derivatization on their stems (Fig. 1C and D).

There are many DNA short repeats in human genomic sequences that are implicated in some genetic diseases (32,33). Studies of the structures and functions of these DNA repeats can provide invaluable insights into the disease mechanisms. As an exploration, two DNA purine repeats (TGGAGGGAGAT and TAGGAGGGAGAT) with unknown structures have been derivatized using this novel Se-derivatization strategy. The selenium modification was introduced at their 5'-termini to minimize possible perturbation. Similarly, a purine-rich DNA (GGAAGTTTGGGAT) was derivatized internally (bold T) for the structural analysis. In order to further study DNA structure and to understand possible perturbation caused by the 2'-Se derivatization, other functional DNAs (see Table 1), including Z-DNA (34), A-form DNA (35), and transcription promoter DNA 32mer (36), were also derivatized analogously using the 2'-Se-C and 2'-Se-U. As a demonstration, seven selenium labels were successfully incorporated into the promoter DNA, which was confirmed by MS analysis (see Table 1).

Synthesis of the 2'-Se-derivatized RNAs and DNAs

We first attempted the synthesis of selenium-functionalized RNA oligonucleotides using the 2'-O-TBDMS protecting group in combination with 1H-tetrazole as the coupling reagent. Unfortunately, this procedure failed to produce a satisfactory coupling yield under the experimental conditions. Although no further experiments were performed to investigate the specific causes of the low yield, this result prompted us to explore 5-BMT as the coupling reagent in combination with the nucleoside phosphoramidites containing the 2'-O-TOM protection (27). We found that the 2'-methylselenolymidine phosphoramidites were as reactive as the ordinary deoxynucleoside phosphoramidites. Coupling of the Se-phosphoramidites was performed in 25 s, and the coupling yields were over 99% when 5-BMT was used. The yields were confirmed by RP-HPLC analysis. As examples, the HPLC profiles of the crude 2'-Se-U RNA 8mer with the DMTr group and the purified 2'-Se-U RNA 8mer without the DMTr group are shown in Figure 2.

Partial detritylation of the purified DMTr-on oligonucleotides was observed during lyophilization. Analysis with HPLC or TLC indicated that ~50% of the oligonucleotide lost the DMTr group (data not shown). We traced the cause of the detritylation to an acidic environment (pH 4.5) generated during the evaporation of the TEA-acetone buffer. Degradation of both Se-DNA and Se-RNA oligonucleotides was also observed when these oligonucleotides were treated with the conventional approach to remove the 5'-DMTr groups. We found that treatment using 2% aqueous trichloroacetic acid for only 1.5 min was sufficient to remove the DMTr groups without causing degradation of these selenium-derivatized oligonucleotides.

In order to investigate the stability of the selenium functionality in the iodine oxidation during the solid phase synthesis, we conducted the oxidation using 20 mM I₂ for 20 s. Interestingly, no measurable oxidation of the selenide functionality was observed in most of the cases. However, when the selenide moiety was close to the 3'-terminus, 2-5% of the selenoxide oxidized from the selenide product was occasionally observed by LC-MS. This selenoxide displayed an extra 16 Da of mass over the corresponding selenide on MS spectra (data not shown). We find that the 2'-selenium derivatized oligonucleotides are quite stable, and there is no detectable oxidation or degradation under air for months.

After HPLC purification, the derivatized oligonucleotides were confirmed by electrospray mass spectrometry (negative ion mode). Two typical MS spectra of the derivatized RNA oligonucleotides are shown in Figure 3. The molecular peaks with several different charges are observed. The MS analytical data of all Se-DNA and Se-RNA oligonucleotides synthesized and presented in this report are shown in Table 1. These data confirm the introduction of selenium labels using both 2'-Se-C and 2'-Se-U phosphoramidites, including incorporation of seven selenium labels per nucleic acid molecule in a DNA promoter.

Thermodenaturation and stability of oligonucleotide duplexes containing the 2'-methylseleno derivatization

To further study the stability of the oligonucleotide duplexes containing the 2'-selenium derivatization, one set of A-form DNA oligonucleotides (GTGTACAC)₂ containing different modifications was synthesized (Table 2). Though the melting temperature of the 2'-MeSe octamer was less than that of the 2'-MeO octamer, it was slightly higher than that of the native octamer. This result suggests that the two derivatized U₈SeMe
residues have no significant effect on duplex stability of the A-form DNA. This result is consistent with the previous UV melting study of the A-form DNA duplex [(GCGTAdUSeMeACGC)₂] (21).

Crystallization of selenium-derivatized oligonucleotides
An A-form DNA with a self-complementary sequence (5'-GTGTACAC-3') (35) was chosen for selenium derivatization and crystallization studies. It was found that these oligonucleotides (Table 2) were able to crystallize in many identical conditions. As examples, photos of crystals of the octamer (native, 5'-GU₃MeGTACAC-3'), the selenium-derivatized octamer (Se-Oct, 5'-GU₃MeGTACAC-3'), and the selenium-bromide derivatized octamer (SeBr-Oct, 5'-GU₃MeGdU₃BrACAC-3') are shown in Figure 4. These native and derivatized crystals were crystallized in the same conditions, and they appear to have the same morphology.

CONCLUSIONS
Derivatization of DNA and RNA with selenium represents a new strategy to facilitate structural determination by X-ray
crystallography via MAD phasing. We have recently achieved the covalent incorporation of selenium into nucleic acids for MAD phasing through collaboration with Egli and coworkers (19–23). This strategy involves the replacement of specific oxygen atoms in the nucleotide building blocks with selenium, followed by chemical or enzymatic incorporation of the modified building blocks into DNAs or RNAs. As an important part of our ongoing research on derivatizing DNA and RNA molecules for crystallographic phasing, we have reported here the chemical synthesis of selenium-labeled oligonucleotides with important structural or biological properties. To achieve the Se-derivatization in more desired sites, the novel 2'-methylseleno cytidine phosphoramidite has been synthesized, and the accessibility of the 2'-methylseleno uridine phosphoramidite has been advanced. DNA and RNA oligonucleotides derivatized with selenium at the specific cytidine and uridine sites have been prepared in large scales using the solid phase approach. In this novel strategy, specific incorporation of the 2'-selenium functionality to the stems of A-form octamer DNAs Melting temperature (°C) 21.2 24.8 21.5

**Table 2.** UV melting temperatures of the A-form DNAs

<table>
<thead>
<tr>
<th>A-form octamer DNAs</th>
<th>Melting temperature (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Native octamer (5'-GTGTACAC-3')</td>
<td>21.2</td>
</tr>
<tr>
<td>2'-MeO-octamer (5'-GUOMaGTACAC-3')</td>
<td>24.8</td>
</tr>
<tr>
<td>2'-MeSe-octamer (5'-GUSeMaGTACAC-3')</td>
<td>21.5</td>
</tr>
</tbody>
</table>

**Figure 4.** Photos of crystals of the native and Se-derivatized octamers. (a) Native-Oct. (b) Se-Oct. (c) Se/Br-Oct. Sizes of the crystals range from 0.1 × 0.1 to 0.4 × 0.4 mm.
ACKNOWLEDGEMENTS

We thank Drs Helen Berman, Martin Egli, Stephen R. Holbrook, Catherine L. Lawson, Benoit Masquida, Dinshaw Patel and Eric Westhof for their discussions, enthusiasm and screenings on these RNA and DNA oligonucleotides derivatized with selenium. We also thank Dr Clifford E. Soll at Hunter College for assisting in MS data collection. This work was supported by PSC-CUNY Awards (64203-00-33 and 65356-00-34), CUNY Project Programs (80210-0302 and 80209-0408) and National Institute of Health grant (GM069703).

REFERENCES