

Covalent incorporation of selenium into oligonucleotides for X-ray crystal structure determination via MAD: proof of principle

Marianna Teplova ^a, Christopher J. Wilds ^a, Zdzislaw Wawrzak ^b, Valentina Tereshko ^a,
Quan Du ^c, Nicolas Carrasco ^c, Zhen Huang ^c, Martin Egli ^{a,*}

^a Department of Biological Sciences, Vanderbilt University, Nashville, TN 37235, USA

^b DND-CAT Synchrotron Research Center, Sector 5, Advanced Photon Source, Argonne National Laboratory, Argonne, IL 60439, USA

^c Department of Chemistry, Brooklyn College, and Program of Biochemistry and Chemistry, The Graduate School, The City University of New York, NY 11210, USA

Received 22 March 2002; accepted 17 May 2002

Abstract

Selenium was incorporated into an oligodeoxynucleotide in the form of 2'-methylseleno-uridine (U_{Se}). The X-ray crystal structure of the duplex $[d(GCGTA)U_{Se}d(ACGC)]_2$ was determined by the multiwavelength anomalous dispersion (MAD) technique and refined to a resolution of 1.3 Å, demonstrating that selenium can selectively substitute oxygen in DNA and that the resulting compounds are chemically stable. Since derivatization at the 2'- α -position with selenium does not affect the preference of the sugar for the C3'-endo conformation, this strategy is suitable for incorporating selenium into RNA. The availability of selenium-containing nucleic acids for crystallographic phasing offers an attractive alternative to the commonly used halogenated pyrimidines. © 2002 Société française de biochimie et biologie moléculaire / Éditions scientifiques et médicales Elsevier SAS. All rights reserved.

Keywords: Anomalous dispersion; DNA; MAD phasing; Nucleic acids; RNA; Selenium labeling; X-ray crystallography

1. Introduction

Determination of protein crystal structures with the multiwavelength anomalous dispersion (MAD) technique [1–3] has gained widespread popularity over the last few years. Selenomethionyl proteins now account for about two thirds of all new protein crystal structures phased by MAD [4]. MAD phasing can be accomplished without the need of a potentially time consuming search for heavy atom derivatives. With more X-ray synchrotron beamlines coming on line worldwide, collection of diffraction data at multiple wavelengths has become ever more routine. Selenium (K edge, 0.979 Å [5]) has proven to be a very effective anomalous scattering center for MAD experiments and can be readily introduced into recombinant proteins in the form of selenomethionine [6]. Selenium in place of sulfur leads to only minimal changes in geometry and hydrophobicity and

crystals of Se-labeled proteins exhibit a high degree of isomorphism with their wild type counterparts.

By comparison, covalent derivatization of nucleic acids for subsequent crystallographic phasing is thus far limited to incorporation of 5-halogen-2'-deoxyuridine or 5-halogen-2'-deoxycytidine. Bromine exhibits a 0.920 Å theoretical K edge and MAD phasing has been performed on bromine-containing DNA crystal structures (e.g. [7,8]). In addition, numerous DNA and RNA crystal structures were determined with multiple isomorphous replacement (MIR), using brominated fragments either alone or in combination with additional derivatives ([9–12] and numerous other examples). However, applications of either MAD or MIR with bromoderivatives are often not successful in practice. Disruptions or alterations of base-stacking and other structural perturbations caused by bromo-derivatization at the 5-position of pyrimidines may hamper crystallization efforts. In the majority of cases, halogen derivatives cannot be crystallized under native conditions. Moreover, derivative crystals sometimes do not diffract as well as the native ones. A further potential problem associated with halogen derivatives is that the modified

* Corresponding author. Tel.: +1-615-343-8070; fax: +1-615-343-6707.
E-mail address: martin.egli@vanderbilt.edu (M. Egli).

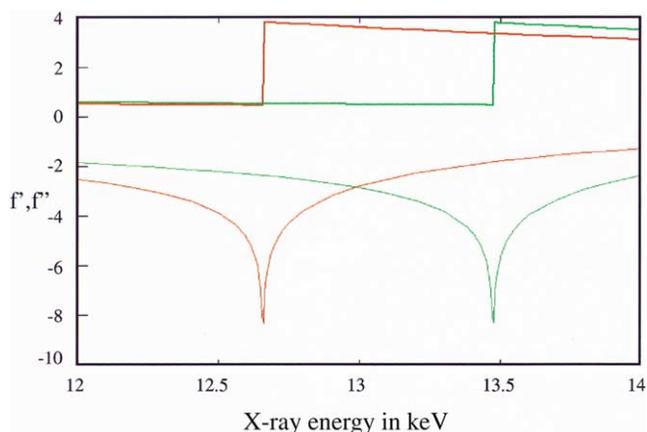


Fig. 1. Location of the K-edges for selenium and bromine (<http://www.bmsc.washington.edu/index.html>); the curves represent the theoretical values of the real and imaginary dispersion coefficients f' and f'' (thin and thick lines, respectively) as a function of X-ray energy ($1 \text{ \AA} = 12\,398.5 \text{ eV}$).

nucleotides are light sensitive. Thus, long-time exposure of halogenated oligonucleotides to X-ray or UV sources may cause decomposition [13].

Selenium and oxygen are in the same Family VIA in the periodic table, which allows for selective replacement of oxygen with selenium. The anomalous signals of bromine and selenium are comparable (Fig. 1), indicating that similar numbers of bromine and selenium atoms need to be incorporated into larger RNA or DNA molecules in order to successfully use MAD phasing. Incorporation of many bromine atoms with limited choice of positions may cause significant changes in the native structures and potential problems in crystallization. Thereby, the use of bromine derivatives is probably limited to oligonucleotides. Compared with bromine, selenium leaves more choices in terms of incorporation site. In principle, selenium can selectively replace nucleotide oxygen atoms in different chemical and geometrical environments, including the 2'-, 3'- and 5'-ribose oxygens, the furanose oxygen, the non-bridging phosphate oxygens, or oxygens in the nucleobases. Many of these oxygen atoms are not involved in intramolecular or intermolecular contacts, and the space surrounding them may be sufficient to accommodate the larger selenium atom.

Selenium has previously been incorporated into DNA in place of non-bridging [14–16] and bridging phosphate oxygen atoms [17]. The phosphoroselenoates in which one of the non-bridging phosphate oxygens is substituted by selenium were found to decompose to phosphate with a half-life of around 30 days [16]. As antisense compounds, phosphoroselenoates were less active than the corresponding phosphorothioates and they were also more toxic to cells. Although they may therefore not be useful as therapeutic agents, the slow loss of non-bridging selenium in phosphodiester does not a priori rule out the use of the phosphoroselenoate modification for crystallographic phasing.

To examine the feasibility of using selenium covalently incorporated into an oligonucleotide for phasing crystallo-

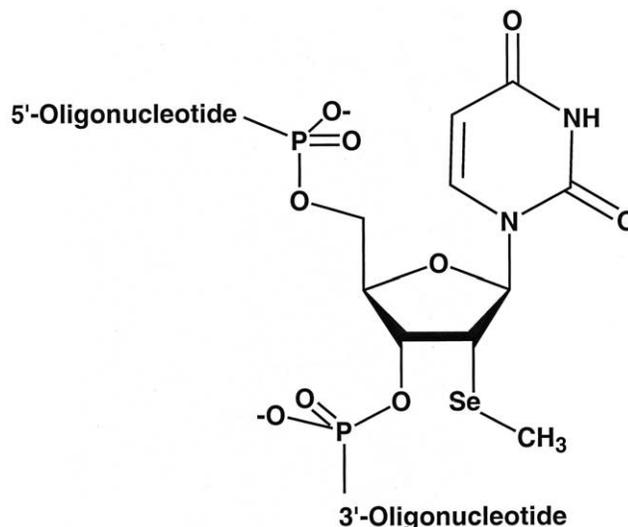


Fig. 2. Structure of 2'-methylseleno-uridine.

graphic diffraction data, we synthesized 2'-methylseleno-uridine (U_{Se} , Fig. 2) and produced a self-complementary decamer DNA oligonucleotide containing the modified building block via solid phase synthesis. A preliminary report of the synthesis of U_{Se} as well as the crystallization of the modified decamer duplex along with evidence for the presence of selenium in the crystal based on X-ray fluorescence spectra has been published earlier (Fig. 3) [18]. Here, we describe details of the synthesis, purification and crystallization of the selenium-modified DNA decamer GCGTAU $_{Se}$ ACGC and provide details of the MAD structure determination and the structural consequences of selenium incorporation.

2. Materials and methods

2.1. Oligonucleotide synthesis and purification

The synthesis of the 2'-SeCH₃ ribo U phosphoramidite was described previously (Fig. 4) [18]. All other oligonucleotide reagents and 3'-CE deoxyphosphoramidites were purchased from Glen Research (Sterling, Virginia). The DNA decamer GCGTAU $_{Se}$ ACGC was synthesized on a 2- μ mol scale on an Applied Biosystems Inc. 381A DNA synthesizer. Coupling times of 90 s and 10 min were used for 3'-CE deoxyphosphoramidites and the 2'-SeCH₃ phosphoramidite, respectively, and the 5'-trityl group was retained following synthesis. All couplings were greater than 75% as judged by the trityl assay. Cleavage of the oligonucleotide from the solid support and deprotection was achieved using 28% NH₄OH, 55 °C, for 8 h. Reverse phase (RP) HPLC analysis and purification were carried out on an Applied Biosystems Inc. chromatograph with a Hewlett-Packard Hypersil ODS-5 column (4.6 \times 200). A 1% gradient of acetonitrile in 0.03 M triethylammonium acetate buffer (pH 7.0) was used with a flow rate of 1.0 ml/min. The crude trityl-on oligo-

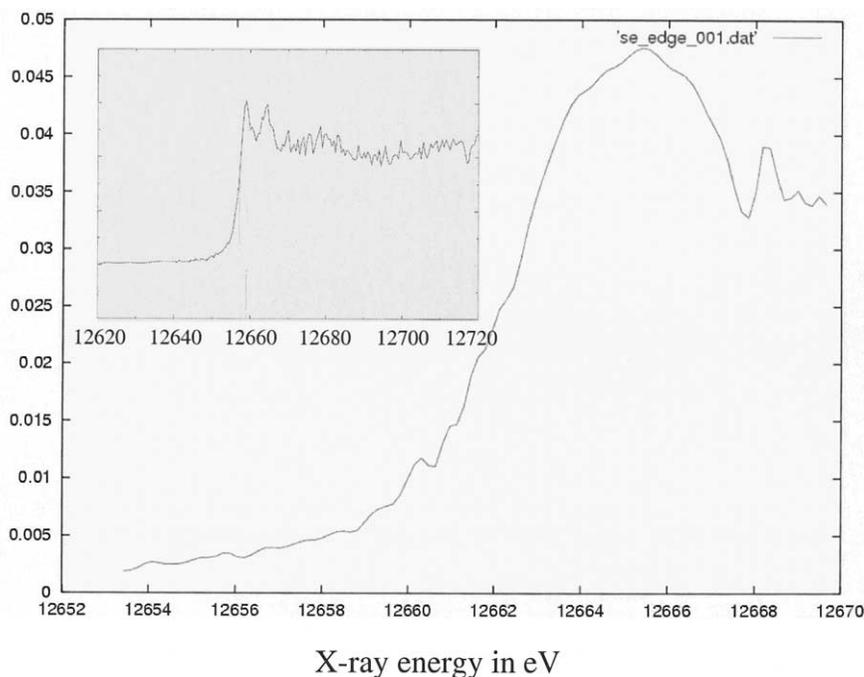


Fig. 3. X-ray fluorescence spectrum of the $d(\text{GCGTA})\text{U}_{\text{Se}}d(\text{ACGC})$ crystal around the Se K-edge ($1 \text{ \AA} = 12\,398.5 \text{ eV}$). The theoretical value for the Se K-edge is $12\,657.8 \text{ eV}$ (0.9795 \AA). The inset depicts a corresponding spectrum obtained for a single crystal of the SeMet-enriched protein Maf from *Bacillus subtilis* [36]. Note the different scales for the x -axis (energy) in the two drawings.

nucleotides were purified by RP HPLC, detritylated in 80% acetic acid for 30 min, and then purified again by RP HPLC. The GCGTAUACGC and GCGTAU_{MeO}ACGC ($\text{U}_{\text{MeO}} = 2'$ -methoxy-U) oligonucleotides were synthesized and purified following standard protocols.

2.2. UV thermal denaturation studies

Oligonucleotide solutions were prepared by dissolving each oligonucleotide in a buffer consisting of 90 mM

sodium chloride, 10 mM sodium phosphate, and 1 mM EDTA (pH 7.0) to give a final concentration of $1 \mu\text{M}$ in duplex. Extinction coefficients were calculated using the nearest neighbor approximation [19]. It was assumed that the 2'-methylseleno and 2'-methoxy substitutions did not change the coefficients. The solutions were then heated to $85 \text{ }^\circ\text{C}$ for 15 min, cooled slowly to room temperature, and stored at $4 \text{ }^\circ\text{C}$ overnight before measurements were taken. Prior to the thermal run, samples were degassed by placing them in a speed-vac concentrator for 2 min. Denaturation

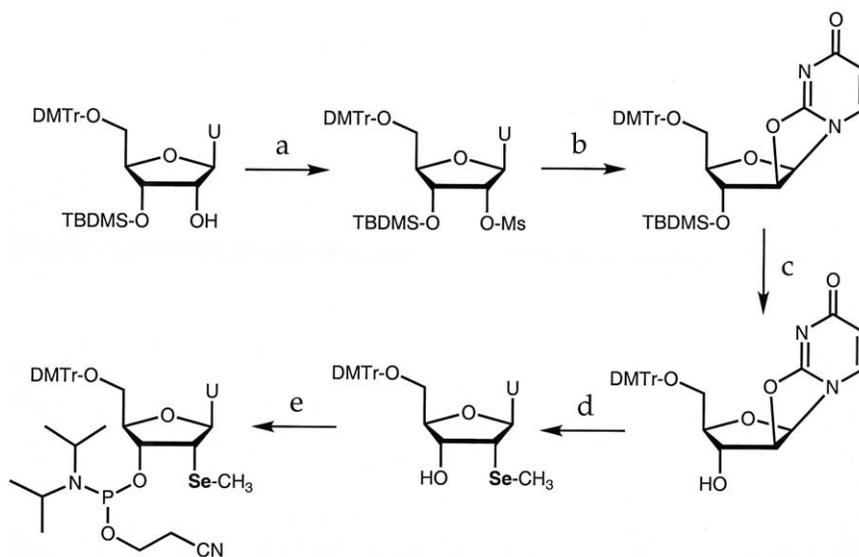


Fig. 4. Synthesis of nucleoside phosphoramidites containing selenium at the 2'-position [18]. (A) $\text{MsCl}/\text{THF}/\text{TEA}$, 95% yield; (B) toluene/tetrahexylammonium hydrogen sulfate/ Na_2CO_3 (sat.), 96% yield; (C) $(\text{Bu})_4\text{N}^+\text{F}^-$, 95% yield; (D) NaSeCH_3 , 96%; (E) $\text{PCI}(\text{OCH}_2\text{CH}_2\text{CN})\text{N}(\text{iPr})_2$, 92% yield.

Table 1
UV melting temperatures (T_m 's) of native and 2'-methoxy- (U_{MeO}) and 2'-methylseleno-modified (U_{Se}) oligodeoxynucleotides

Oligonucleotide	T_m (°C)
GCGTAUACGC	58.2
GCGTAU _{MeO} ACGC	63.3
GCGTAU _{Se} ACGC	58.8

curves were acquired at 260 nm at a heating rate of 0.5 °C/min, using a Varian CARY model 3E spectrophotometer fitted with a six-sample thermostated cell block and a temperature controller. The data were analyzed in accordance with the convention of Puglisi and Tinoco [20] (Table 1).

2.3. Crystallizations

Crystals of the selenium-modified decamer duplex were grown by the hanging drop vapor diffusion method, using a commercially available sparse matrix screen (Hampton Research Inc., Laguna Niguel, CA) [21]. Crystals for data collection were obtained by mixing equal volumes of an aqueous DNA solution (2 mM) and a solution containing 40 mM sodium cacodylate (pH 7.0), 12 mM sodium chloride, 80 mM potassium chloride, 12 mM spermine tetrahydrochloride and 10% (v/v) 2-methyl-2,4-pentanediol (solution 19, Nucleic Acid Mini Screen) and equilibrating the droplets over 1 ml of a 35% v/v 2-methyl-2,4-pentanediol reservoir solution. Several other conditions also resulted in crystal growth. Crystals typically appeared within one week and belong to the orthorhombic space group $P2_12_12_1$ with unit cell constants $a = 24.56$ Å; $b = 43.97$ Å; $c = 45.34$ Å and one duplex per asymmetric unit. The crystals were shock-frozen directly in the mother liquor for data collection.

Table 2
Data collection and phasing statistics

	Wavelength (Å)		
	0.97915 Inflection point	0.97892 Peak	0.94196 Remote
Resolution range, Å (last shell)	20–1.3 (1.35–1.3)		
Unique reflections	23 332	22 943	21 232
Redundancy	6.9	6.1	2.8
Completeness, % (last shell)	99.8 (100.0)	98.2 (100.0)	90.5 (91.3)
R_{merge}^a , % (last shell)	4.9 (26.3)	4.2 (20.5)	3.2 (24.1)
R_{Cullis}^b			
Centrics/accentrics	0.79/0.74	0.43/0.69	–
All	0.74	0.64	
Phasing power ^c			
Centrics/accentrics	0.74/0.70	1.82/1.48	–
All	0.71	1.52	
Overall figure of merit (before density modification)			
Centrics/accentrics	0.89/0.75		
All	0.76		

^a $R_{merge} = \sum_{hkl} \sum_i |I(hkl)_i - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i \langle I(hkl) \rangle$ for i measurements of the intensity I of a reflection hkl .

^b $R_{Cullis} = \sum |F_{((\lambda))} \pm F_{(\lambda)}| - |F_{h(\lambda),c}| / \sum |F_{((\lambda))} \pm F_{(\lambda)}|$, where $F_{h(\lambda),c}$ is the calculated heavy structure factor.

^cPhasing power = $\langle F_{h((\lambda))} \rangle / E$, where $\langle F_{h((\lambda))} \rangle$ is the RMS heavy atom structure factor and E is the residual lack-of-closure error.

2.4. MAD data collection and crystal structure determination

X-ray diffraction data to a maximum resolution of 1.3 Å were collected at 100 K on the 5-ID beamline of the Dupont–Northwestern–Dow Collaborative Access Team (DND-CAT) at the advanced photon source (APS), Argonne, Illinois. MAD data from the K edge of selenium were collected on a single 2'-methylseleno-modified crystal. High- and low-resolution data sets were collected separately. Before data collection, a fluorescence scan centered around the selenium K-edge was recorded in order to accurately tune the X-ray beam to the three wavelengths (Fig. 3), i.e. those at the inflection point (edge), the absorption maximum (peak), and ca. 460 eV above the peak wavelength (remote) (Table 2). The three wavelengths were alternated after each sweep of 45° in phi with a 1° oscillation. All data were integrated and merged using DENZO and SCALEPACK, respectively [22]. Selected statistics of the anomalous data collection are listed in Table 2. A summary of the magnitude of the anomalous effect based on the two selenium atoms per asymmetric unit as a function of data resolution is given in Table 3.

The positions of the two selenium atoms per duplex were determined with CNS-solve [23]. Following refinement of the heavy atom parameters, experimental MAD phases were calculated with the same program. Solvent flattening and density modification significantly improved the initial map, revealing excellent density for all twenty nucleotides of the duplex (Fig. 5). The structural model of an A-form DNA duplex [24] was used to fit atom positions into the modified experimental electron density. Initial positional and isotropic B-factor refinements of the duplex were carried out with CNS [23], using inflection point data (Table 2) and setting aside 5% of the reflections as a test set to monitor the R_{free} .

Table 3
Magnitude of anomalous effect due to selenium as evaluated by the so-called χ^2 test [22]

Friedel pairs	Shell limit (Å)		Average I	Average		Norm. χ^2	Linear R-factor	Square R-factor
	Lower	Upper		Error	Stat.			
Unmerged	20.00	2.80	114 168	7742	942	0.991	0.035	0.036
	2.80	2.22	28 340	728	200	0.998	0.032	0.034
	2.22	1.94	13 429	462	87	1.016	0.051	0.053
	1.94	1.76	7663	300	68	1.016	0.051	0.055
	1.76	1.64	4902	193	59	1.005	0.054	0.054
	1.64	1.54	2325	106	50	1.017	0.070	0.066
	1.54	1.46	1225	69	48	1.028	0.096	0.090
	1.46	1.40	874	60	50	1.010	0.120	0.109
	1.40	1.35	569	56	51	1.019	0.169	0.150
	1.35	1.30	449	55	53	0.998	0.205	0.181
	All reflections		15 788	865	148	1.010	0.042	0.038
Merged	20.00	2.80	116 954	6535	6535	17.513	0.049	0.059
	2.80	2.22	28 218	563	563	58.877	0.061	0.070
	2.22	1.94	13 503	358	358	26.921	0.059	0.066
	1.94	1.76	7737	229	229	15.771	0.061	0.067
	1.76	1.64	4947	148	148	12.671	0.060	0.061
	1.64	1.54	2354	80	80	7.676	0.060	0.059
	1.54	1.46	1217	51	51	6.069	0.073	0.072
	1.46	1.40	868	44	44	4.099	0.075	0.072
	1.40	1.35	568	41	41	2.664	0.086	0.079
	1.35	1.30	449	40	40	2.095	0.095	0.086
	All reflections		17 013	767	767	15.082	0.054	0.060

[25]. Anisotropic B-factor refinement for all DNA atoms and selected solvent molecules was performed with SHELX-97 [26]. After an R_{free} of 20.0% was reached, all reflections to 1.3 Å resolution were used in the final rounds of refinement. Fixed positions for hydrogen atoms based on the positions of heavier atoms were included in the refinement, resulting in an R_{work} of 14.9%. Selected refinement parameters and root mean square (r.m.s.) deviations from standard geometry are listed in Table 4.

3. Results and discussion

3.1. Covalent incorporation of selenium into oligonucleotides

Nucleic acids for X-ray crystallographic studies are usually prepared by solid phase synthesis (DNA and RNA oligonucleotides) or in vitro transcription (RNAs > 35 nucleotides long), using nucleoside phosphoramidites and triphosphates, respectively. In our initial exploration of the suitability of selenium for determining crystal structures of nucleic acids via MAD, we focused on the synthesis of the selenium-labeled T and U phosphoramidite building blocks. As one selenium atom should enable phase determination for RNAs or DNAs of up to 30 nucleotides length and the frequency of a building block in a nucleotide sequence is high (25% average rate), selenium-labeled T and U building blocks should satisfy most of the needs of oligonucleotide derivatization.

In principle, any oxygen of a given nucleotide building block (A, C, G, U or T), including the 2', 3', 4' and 5' ribose, nucleobase, and α -phosphate non-bridging oxygens, may be

subject to selenium replacement. In some cases, selective substitution of a single oxygen atom by selenium in a nucleotide may not result in significant functional or structural perturbations. However, certain substitutions, particularly those affecting oxygens involved in inter-strand (i.e. Watson–Crick) or intra-strand hydrogen bonds (i.e. C2'–H2' (n) \cdots O4' ($n+1$) [27]), may not be tolerated and modification of such centers is not advisable. We chose to introduce selenium at the 2'-position of the ribose sugar (Fig. 2). Modification of the 2'- α -position with a methylseleno moiety appeared compatible with solid phase synthesis and the 2'-substituent was expected to conform to the geometric boundaries of A-form DNA and RNA duplexes. An obvious limitation of this modification strategy is that the sugar pucker is being restricted to the C3'-endo region. Thus, structures of B-form DNAs are not amenable to derivatization; however, the 2'-methylseleno modification is suitable for derivatizing small- and medium-sized RNA molecules.

The uridine phosphoramidite with selenium at the 2'-position can be synthesized in five steps, starting from partially protected uridine [18] (Fig. 4). Displacement of the mesyl protection group at the 2'-position by the uracil exo-2-oxygen under basic conditions in a two-phase reaction system involving a phase transfer catalyst results in formation of anhydro-uridine in high yield. The 3'-TBDMS group is then removed by fluoride treatment, followed by nucleophilic substitution at the 2'-carbon with sodium methylselenide in THF to furnish the 2'-methylseleno nucleoside. The methyl group prevents oxidation of the selenium functionality. Finally, the selenium-labeled nucleoside is converted to the phosphoramidite building block by treating the former with 2-cyanoethyl-N,N-diisopropylchlorophosphoramidite (Fig. 4). Apart from extending the

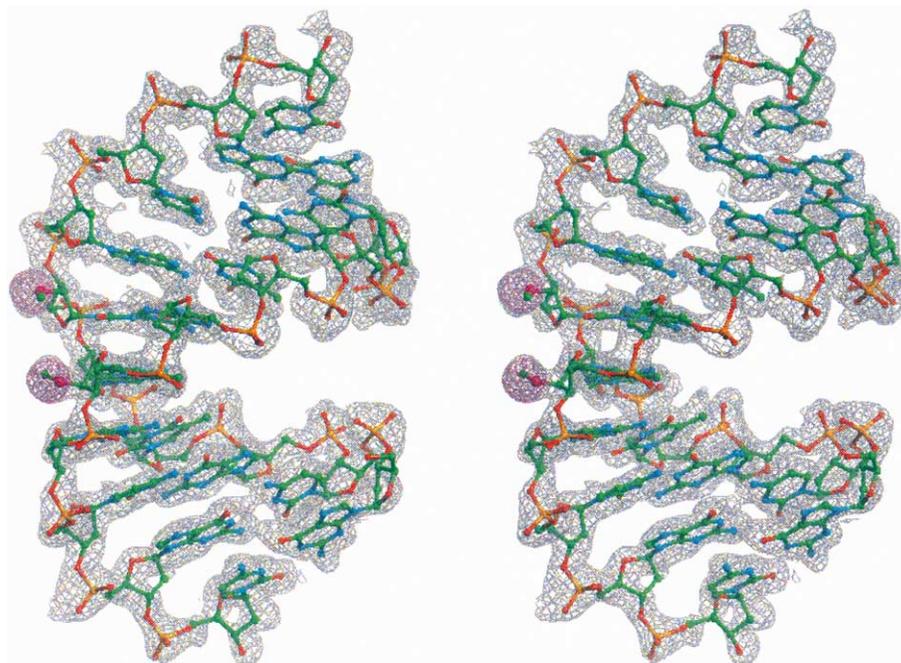


Fig. 5. Stereo diagram depicting experimental electron density (gray) calculated at 1.3 Å resolution with MAD phases and superimposed on the final structure of the decamer DNA duplex $[d(\text{GCGTA})\text{U}_{\text{Se}}d(\text{ACGC})]_2$. The anomalous difference Fourier electron density based on all data to 1.3 Å resolution is shown in pink. Both maps are contoured at the 1σ level. DNA atoms are colored green, cyan, red and orange for carbon, nitrogen, oxygen and phosphorus, respectively, and the two selenium atoms per asymmetric unit are highlighted in pink.

coupling times for U_{Se} phosphoramidite, standard solid phase synthesis protocols were followed to synthesize the DNA decamer and other DNA and RNA oligonucleotides [18]. Compared with the yields of regular DNA and RNA couplings (typically >98%), those observed for U_{Se} are somewhat lower. However, they are still in a range that permits good overall yields for a 15mer oligonucleotide. Moreover, incorporation of a single U_{Se} is sufficient for phasing diffraction data of oligonucleotide crystals. After deprotection of the 2'-selenium-modified oligonucleotide, ion exchange or reverse phase (RP) high-performance liquid chromatography (HPLC) can be used for purification. The decamer $\text{GCGTAU}_{\text{Se}}\text{ACGC}$ was purified twice by

Table 4
Refinement statistics

Parameter	Value
Resolution (Å)	10.0–1.3
Reflections: $F_o \geq 4\sigma[F_o]/\text{all}$	21 189/23 279 ^a
R_{work}^b : $F_o \geq 4\sigma[F_o]/\text{all}$ (%)	14.9/15.4
$R_{\text{free}}^{b,c}$: $F_o \geq 4\sigma[F_o]/\text{all}$ (%)	17.3/17.8
No. of nucleic acid atoms	628
No. of water molecules	146
<i>RMS deviations</i>	
Bond length (Å)	0.015
Bond angles (1–3 distances, Å)	0.023
<i>Average temperature factors</i>	
DNA atoms (Å ²)	18.4 (s.d. 5.4)
Solvent molecules (Å ²)	38.5 (s.d. 11.9)

^aInflection point data (0.97915 Å).

^b $R = \sum |F_o| - |F_c| / \sum |F_o|$.

^cFor 5% of the reflections; test set reflections were included in the very final rounds of refinement.

RP-HPLC (trityl-on and -off), using a standard triethylammonium acetate buffer/acetonitrile eluent protocol.

3.2. Stability of oligodeoxynucleotides containing 2'-methylseleno U

UV melting experiments were conducted to examine the effect of incorporation of a single U_{Se} residue on the thermodynamic stability of the DNA decamer GCGTAUACGC . We measured the melting temperatures (T_m 's) of the duplex $[d(\text{GCGTA})\text{U}_{\text{Se}}d(\text{ACGC})]_2$ along with those of the native duplex $[d(\text{GCGTAUACGC})]_2$ and a duplex $[d(\text{GCGTA})\text{U}_{\text{MeO}}d(\text{ACGC})]_2$, containing a single 2'-methoxy-uridine (U_{MeO}). The results are listed in Table 1. Incorporation of two U_{Se} residues has virtually no effect on the thermal stability of the decamer duplex relative to the all-DNA control. The T_m of $[d(\text{GCGTA})\text{U}_{\text{Se}}d(\text{ACGC})]_2$ is approximately half a degree higher than that of the all-DNA, but it is clearly less stable than the duplex featuring a single 2'-methoxy-uridine per strand.

It is well known that 2'-methoxy substitution of DNA leads to increased stability of duplexes between modified DNAs and RNA relative to the parent DNA:RNA hybrids [28,29]. The increase in T_m amounts to about 1.5 °C per methoxy substitution, but the origins of this gain in stability are not completely understood. In addition, RNA duplexes are typically more stable thermodynamically than DNA duplexes of identical sequence [30]. Both substituents, 2'-OH and 2'-OMe conformationally preorganize the sugar for the C3'-endo type pucker. In addition, the 2'-hydroxyl

groups act as hydrogen bond donors and acceptors, thus engaging in stable hydration motifs that laterally span the groove [31]. Therefore, the RNA minor groove is well hydrated. By comparison, the minor groove of 2'-*O*-methyl RNA (2'-methoxy DNA) is dry [32]. The methyl groups are directed into the groove (antiperiplanar orientation of the C3'-C2'-O_{2'}-CH₃ torsion) and are in van der Waals contact with the edges of base pairs [33]. Clearly, opposing effects appear to be at work in terms of hydration of (2'-OH) RNA and 2'-*O*-methyl RNA, both resulting in increased thermodynamic stability of the respective duplexes relative to DNA.

3.3. MAD structure determination

In order to select optimal wavelengths for MAD data collection an X-ray fluorescence spectrum was recorded from a GCGTAU_{Se}ACGC crystal. In addition to the precise locations of peak and inflection point, the shape of the K edge X-ray spectrum has been used as an indicator of the electronic integrity of selenium sites in protein crystals [34]. Compared with the K edge X-ray spectrum of a protein crystal (see Fig. 3, inset) that of the U_{Se}-modified DNA crystal exhibits a somewhat broader peak and the maximum is shifted to higher energy (by about 5 eV, Fig. 3). The cause of these differences is unclear, but it is unlikely that partial oxidation of selenium or anisotropy of the Se K edge are involved. The particular shape and position of the K edge may be related to the electronic environment of the selenium atom (i.e. the gauche orientation to O_{4'}). Clearly, as more selenium-modified nucleic acid crystals are being investigated and different incorporation sites are being explored, we may gain a better understanding of the relation between the chemical environment in DNA or RNA and peak shape.

Datasets at three wavelengths to a maximum resolution of 1.3 Å were collected from a single crystal (Table 2). Comparison of the intensities of Friedel pairs reveals a striking effect of the two selenium atoms per crystallographic asymmetric unit (Table 3). The magnitude of the anomalous differences is undoubtedly due in part to the presence of one selenium atom per single-stranded DNA decamer. This is approximately comparable to a frequency of one Se-Met residue per 20 amino acids in a protein. However, when comparing the relative magnitudes of the anomalous difference, we need to bear in mind that the occupancies of selenium sites (we are not referring to crystallographic occupancy here) in the U_{Se}-modified DNA can be assumed to be 100% due to the covalent nature of the C2'-Se bond. On the contrary, replacement of methionine by seleno-methionine during protein overexpression will typically not proceed with 100% efficiency. A further difference may be constituted by the flexibility of the Se-Met side chain and the 2'-methylseleno-modified sugar in proteins and double helical DNA, respectively. In most cases the sugar moiety of nucleotides in duplexes is well ordered.

Terminal residues and single stranded regions (i.e. bulges) may constitute exceptions. As expected, both 2'-methylseleno substituents in the crystal structure of the modified decamer were ordered and all atoms were refined with (crystallographic) occupancies of 1.0.

Initially, the locations of both selenium atoms were determined from anomalous difference density maps (Fig. 5), relying on all three data sets (peak, inflection point and reference). After refinement of the two Se sites, Fourier electron density maps were computed, revealing excellent density for the entire duplex (Fig. 5). A model based on an earlier atomic-resolution structure of the decamer [24] was fit into the density and served as the starting point of the refinement. Initial rounds of positional and temperature factor refinements as well as simulated annealing were carried out with the program CNS [23]. At a later stage, the CGLS-type constraint least squares refinement in the program SHELXL-97 [26] in combination with anisotropic temperature factor refinement for all DNA atoms as well as selected water molecules was carried out. No bond length and angle constraints were used for the C2'-Se and Se-CH₃ bonds and the C2'-Se-CH₃ angle, respectively. An example of the final (2F_o-F_c) Fourier sum electron density surrounding U_{Se} residues is depicted in Fig. 6.

We also examined whether data collected at the peak wavelength alone would allow phasing of the DNA decamer structure (single-wavelength anomalous dispersion, SAD). Indeed, Fourier synthesis based on just these data revealed unambiguous density for the whole duplex. Thus, incorporation of selenium into nucleic acids and careful determination of the peak wavelength is sufficient for SAD phasing.

3.4. Structure of [d(GCGTA)U_{Se}d(ACGC)]₂

The structure of the U_{Se}-modified DNA duplex is isomorphous to that of the GCGTAT_MACGC decamer (T_M = 2'-methoxy-3'-methylphosphonate-thymidine) [24]. In [d(GCGTA)U_{Se}d(ACGC)]₂ the two selenium atoms are separated by 9.1 Å across the minor groove (Fig. 7). The A5pT_M(U_{Se})6:A15pT_M(U_{Se})16 base pair steps in the two duplexes exhibit virtually identical conformations, illustrating that Se-modification at the sugar 2'-position does not distort the A-form duplex geometry (Fig. 7).

All sugars in the decamer duplex adopt C3'-endo pucker and the pseudorotation phase angles (*P*) for the 2'-modified sugars of residues U_{Se}6 and U_{Se}16 are 13° and 20°, respectively. Both, the 2'-methylseleno and the 2'-methoxy substituents adopt antiperiplanar orientation around the C2'-Se and C2'-O bonds, respectively. The C3'-C2'-Se-CH₃ torsion angle for residue U_{Se}6 is -175° (C3'-C2'-O-CH₃ = -171° for T_M6 in the reference structure). The C3'-C2'-Se-CH₃ torsion angle for residue U_{Se}16 is -172° (C3'-C2'-O-CH₃ = -175° for T_M16 in the reference structure). In residue U_{Se}6, the C2'-Se and Se-CH₃ bond lengths measure 2.02 and 1.98 Å, respectively, and in residue U_{Se}16, these bond lengths are 2.00 and 2.03 Å,

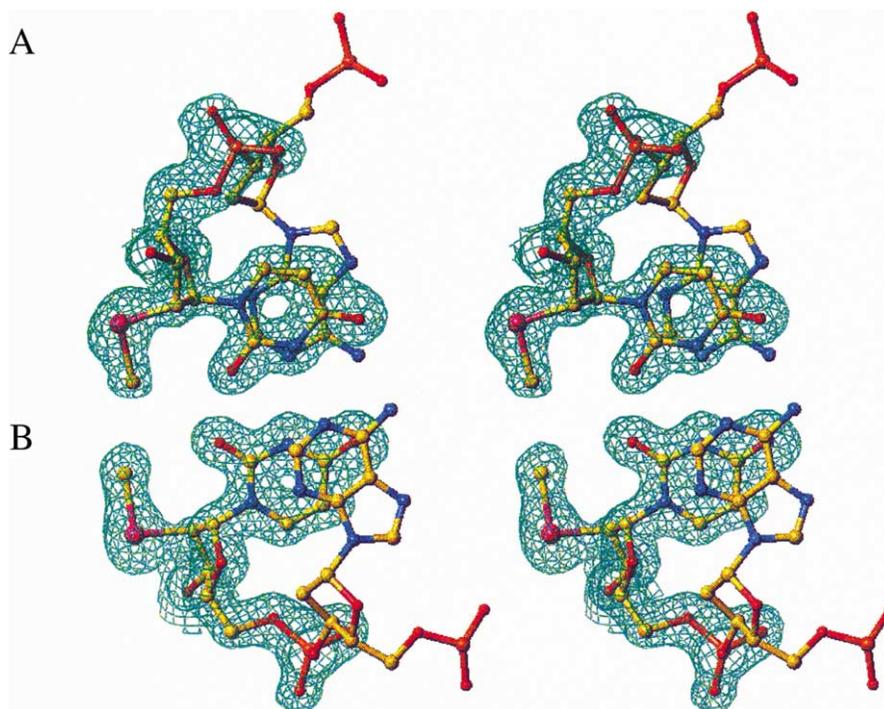


Fig. 6. Stereo diagram depicting the base pair steps (A) A5pU_{Se}6 and (B) A15pU_{Se}16, viewed approximately along the normal to the base planes. Final sum ($2F_o - F_c$) Fourier electron density (1σ level) at 1.3 Å resolution around 2'-methylseleno-uridines is shown in green. DNA atoms are colored yellow, blue, red and orange for carbon, nitrogen, oxygen and phosphorus, respectively, and Se atoms are highlighted in pink.

respectively. By comparison, the C2'-O₂' and O₂'-CH₃ bond lengths in the reference structure are shorter by approximately 0.6 Å (T_M6 : 1.37 and 1.43 Å, resp.; T_M16 : 1.43 and 1.43 Å resp.; Fig. 7).

Selenium is more hydrophobic than oxygen or sulfur, but the higher hydrophobicity is of little consequence in terms of the number of water molecules surrounding the 2'-methoxy and 2'-methylseleno substituents. In the GCGTAT_M-ACGC structure, both 2'-oxygens are engaged in a hydrogen bond to single waters. The hydrogen bond distances are 2.83 and 3.03 Å for O₂' of T_M6 and T_M16 , respectively. The selenium atom of residue U_{Se}6 is surrounded by three water molecules at distances of 2.63, 3.24 and 3.34 Å. However, the selenium of residue U_{Se}16 is not involved in interactions with solvent molecules. Thus, introduction of selenium at the sugar 2'-position does not appear to affect minor groove hydration to a significant extent.

3.5. Coordinates

The final coordinates and structure factors have been deposited in the Brookhaven Protein Databank (1MA8).

4. Conclusions and outlook

Covalent incorporation of selenium into DNA and RNA oligonucleotides at the sugar 2'-position in combination with MAD or SAD opens up a new route to rapid determination of nucleic acid crystal structures and potentially of

structures of protein-nucleic acid complexes. The approach presented here excludes applications to nucleic acid fragments with a B-form duplex conformation. Moreover, longer nucleic acids that are not accessible to solid phase synthesis cannot be labeled with selenium using our current protocol. However, the possibility to derivatize oligonucleotides with selenium does add to the armamentarium of crystallographic phasing tools for solving new nucleic acid structures. Since preparation of heavy atom derivatives of nucleic acid crystals is often not successful, these tools are currently mostly limited to incorporation of 5-halogenated pyrimidines.

A major advantage of a derivatization strategy using selenium compared to 5-bromo U (C) or 5-iodo U (C) is that selenium can be incorporated in place of oxygen at a variety of sites in nucleic acids, including base, sugar and phosphate group. We have recently incorporated selenium into an oligodeoxynucleotide in place of one of the non-bridging phosphate oxygens. The two resulting diastereoisomers (Se-*R_p* and Se-*S_p*) were separated by ion-exchange HPLC and crystals diffracting to atomic-resolution were grown for both. Phosphoroselenoates were reported to decompose with a half-life of 30 days [16]. However, the self-complementary oligonucleotides containing a single P-Se moiety per strand were stable in the crystallization droplets for months and subsequent structure determination using MAD or SAD revealed fully occupied Se sites and excellent experimental maps. The opportunity to incorporate selenium at different sites within the framework of nucleic acids greatly enhances the versatility of the Se-derivatization

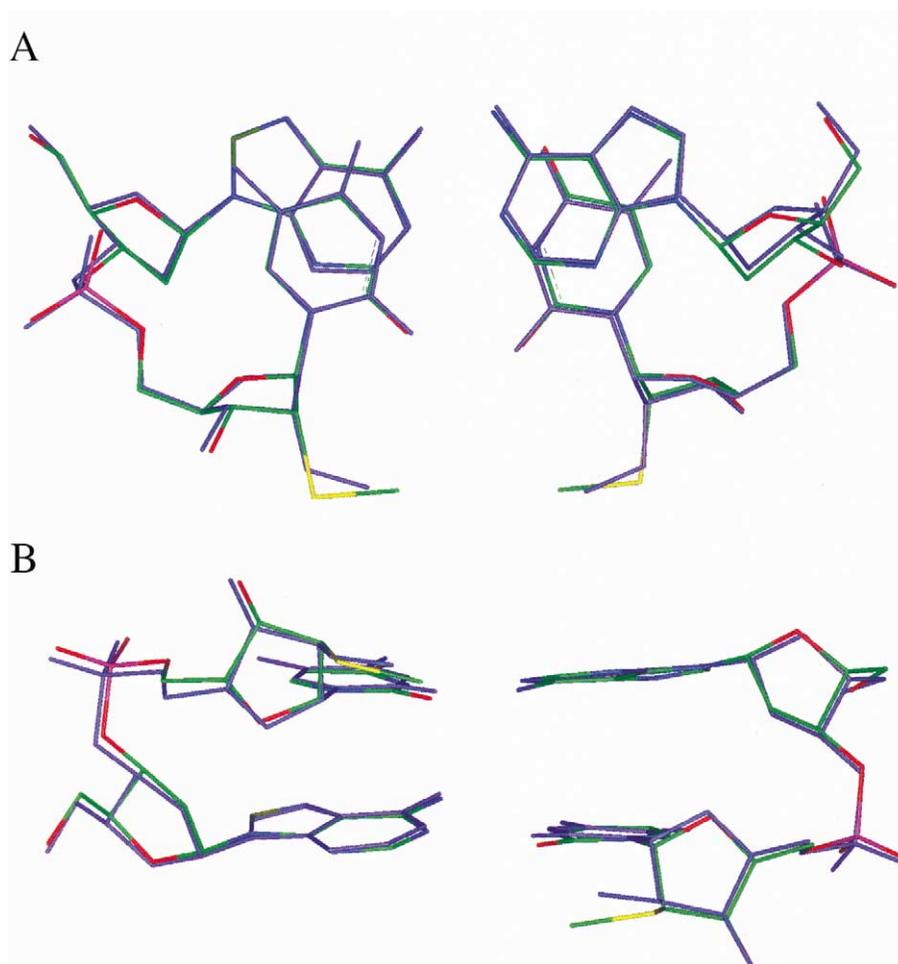


Fig. 7. Superposition of the A5pU_{Se}6:A15pU_{Se}16 and A5p(2'MeO-3'MP)T6:A15p(2'MeO-3'MP)T16 (purple bonds) base pair steps in the GCGTAU_{Se}ACGC and GCGTAT_MACGC [24] duplexes, respectively. The views are (A) roughly along the normal to the top base pair (top) and (B) rotated around the horizontal by 90°, into the minor groove. The 2'MeO-3'MP T (T_M) residue is 2'-methoxy-3'-methylenephosphonate-thymidine. Atoms of the base pair step featuring the 2'-methylseleno-uridines are colored green, red, blue, magenta and yellow for carbon, oxygen, nitrogen, phosphorus and selenium, respectively.

approach. Not only can DNA and RNA oligonucleotides be readily modified, but phosphoroselenates can also be synthesized for any nucleic acid analog as long as it features phosphates in its backbone. In a decamer, for example, nine phosphate groups are available for P–Se modification, potentially resulting in a total of 18 derivative strands after separation of all diastereoisomeric pairs. This modification strategy is sequence-independent and, while involving a considerable synthetic effort, will undoubtedly lead to at least one selenoated oligonucleotide for which crystals can be grown. Possible partial decomposition of the phosphoroselenates may not hamper the crystallographic phase determination as even a selenium site of 50% occupancy can be expected to provide a sufficient anomalous contribution.

Finally, it may be possible to extend the selenium derivatization approach to longer RNAs or DNAs that are not accessible via solid phase synthesis. For RNA, incorporation of selenium can be achieved by enzymatic *in vitro* transcription using UTPs carrying selenium at the 5'-position [35] or in place of the exocyclic 2-oxygen. In

addition, longer RNAs can also be prepared by ligation of a long fragment prepared by *in vitro* transcription and a shorter one containing selenium and prepared by solid phase synthesis.

Acknowledgements

Financial support from the National Institutes of Health (grant GM-55237 to ME) is gratefully acknowledged. ZH is supported by PSC-CUNY research awards (69674-00-29 and 62392-00-31) and the New Research Dimension Fund and CJW is the recipient of a post-graduate scholarship from the National Science and Engineering Research Council of Canada (NSERC). We are grateful to Professor Paul Miller, Johns Hopkins University, and Dr. Anne Noronha for their assistance with thermal melting experiments. Portions of this work were performed at the DuPont–Northwestern–Dow collaborative access team (DND-CAT) Synchrotron Research Center located at Sector 5 of the Advanced Photon Source. DND-CAT is supported by the

E.I. DuPont de Nemours & Co., The Dow Chemical Company, the US National Science Foundation through Grant DMR-9304725 and the State of Illinois through the Department of Commerce and the Board of Higher Education Grant IBHE HECA NWU 96. Use of the Advanced Photon Source was supported by the US Department of Energy, Basic Energy Sciences, Office of Energy Research under Contract No. W-31-102-Eng-38.

References

- [1] W.A. Hendrickson, A. Pähler, J.L. Smith, Y. Satow, E.A. Merritt, R.P. Phizackerley, Crystal structure of core streptavidin determined from multiwavelength anomalous diffraction of synchrotron radiation, *Proc. Natl. Acad. Sci. USA* 86 (1989) 2190–2194.
- [2] W. Yang, W.A. Hendrickson, R.J. Crouch, Y. Satow, Structure of ribonuclease H phased at 2-Å resolution by MAD analysis of the selenomethionyl protein, *Science* 249 (1990) 1398–1405.
- [3] W.A. Hendrickson, Determination of macromolecular structures from anomalous diffraction of synchrotron radiation, *Science* 254 (1991) 51–58.
- [4] W.A. Hendrickson, Synchrotron crystallography, *Trends Biochem. Sci.* 25 (2001) 637–643.
- [5] J.R. Helliwell, *Macromolecular crystallography with synchrotron radiation*, Cambridge University Press, Cambridge, UK, 1992.
- [6] S. Doublé, Preparation of selenomethionyl proteins for phase determination, *Meth. Enzymol.* 276 (1997) 523–530.
- [7] C.M. Ogata, W.A. Hendrickson, X. Gao, D. Patel, Crystal structure of chromomycin-DNA complex, *J. Abstr. Am. Cryst. Assoc. Mtg. Ser.* 2 (1989) 17–53.
- [8] X. Shui, M.E. Peek, L.A. Lipscomb, Q. Gao, C. Ogata, B.P. Roques, C. Garbay-Jaureguiberry, A.P. Wilkinson, L.D. Williams, Effects of cationic charge on three-dimensional structures of intercalative complexes: structure of a bis-intercalated DNA complex solved by MAD phasing, *Curr. Med. Chem.* 7 (2000) 59–71.
- [9] R. Wing, H. Drew, T. Takano, C. Broka, S. Tanaka, K. Itakura, R.E. Dickerson, Crystal structure analysis of a complete turn of B-DNA, *Nature* 287 (1980) 755–758.
- [10] A.D. DiGabriele, T.A. Steitz, A DNA dodecamer containing an adenine tract crystallizes in a unique lattice and exhibits a new bend, *J. Mol. Biol.* 231 (1993) 1024–1039.
- [11] L. Su, L. Chen, M. Egli, J.M. Berger, A. Rich, A minor groove RNA triplex in the crystal structure of a viral pseudoknot involved in ribosomal frameshifting, *Nature Struct. Biol.* 6 (1999) 285–292.
- [12] L. Zhang, J.A. Doudna, Structural insights into group II intron catalysis and branch-site selection, *Science* 295 (2002) 2084–2088.
- [13] M.C. Willis, B.J. Hicke, O.C. Uhlenbeck, T.R. Cech, T.H. Koch, Photocrosslinking of 5-iodouracil-substituted RNA and DNA to proteins, *Science* 262 (1993) 1255–1257.
- [14] M.J. Nemer, K.K. Ogilvie, Ribonucleotide analogs having novel internucleotide linkages, *Tetrahedron Lett.* 21 (1980) 4149–4152.
- [15] W.J. Stec, G. Zon, W. Egan, Automated solid-phase synthesis, separation, and stereochemistry of phosphorothioate analogs of oligodeoxyribonucleotides, *J. Am. Chem. Soc.* 106 (1984) 6077–6079.
- [16] K. Mori, C. Boiziau, C. Cazenave, M. Matsukura, C. Subasinghe, J.S. Cohen, S. Broder, J.J. Toulmé, C.A. Stein, Phosphoroselenoate oligodeoxynucleotides: synthesis, physico-chemical characterization, anti-sense inhibitory properties and anti-HIV activity, *Nucleic Acids Res.* 17 (1989) 8207–8219.
- [17] Y. Xu, E.T. Kool, Rapid and selective selenium-mediated autoligation of DNA strands, *J. Am. Chem. Soc.* 122 (2000) 9040–9041.
- [18] Q. Du, N. Carrasco, M. Teplova, C.J. Wilds, M. Egli, Z. Huang, Internal derivatization of oligonucleotides with selenium for X-ray crystallography using MAD, *J. Am. Chem. Soc.* 104 (2002) 24–25.
- [19] P.N. Borer, Optical properties of nucleic acids, absorption, and circular dichroism spectra., in: G.D. Fasman (Ed.), *Handbook of Biochemistry and Molecular Biology, Nucleic Acids*, third ed. vol. I, CRC Press, Cleveland, OH, 1975, pp. 589.
- [20] J.D. Puglisi, I.J. Tinoco, Absorbance melting curves of RNA, *Meth. Enzymol.* 180 (1989) 304–325.
- [21] I. Berger, C. Kang, N. Sinha, M. Wolters, A. Rich, A highly efficient 24-condition matrix for the crystallization of nucleic acid fragments, *Acta Cryst. D* 52 (1996) 465–468.
- [22] Z. Otwinowski, W. Minor, Processing of X-ray diffraction data collected in oscillation mode, *Meth. Enzymol.* 276 (1997) 307–326.
- [23] A.T. Brünger, P.D. Adams, G.M. Clore, W.L. DeLano, P. Gros, R.W. Grosse-Kunstleve, J.S. Jiang, J. Kuszewski, M. Nilges, N.S. Pannu, R.J. Read, L.M. Rice, T. Simonson, G.L. Warren, Crystallography and NMR system: a new software suite for macromolecular structure determination, *Acta Cryst. D* 54 (1998) 905–921.
- [24] M. Egli, V. Tereshko, M. Teplova, G. Minasov, A. Joachimiak, R. Sanishvili, C.M. Weeks, R. Miller, M.A. Maier, H. An, P.D. Cook, M. Manoharan, X-ray crystallographic analysis of the hydration of A- and B-form DNA at atomic resolution, *Biopolymers, Nucleic Acid Sciences* 48 (2000) 234–252.
- [25] A.T. Brünger, Free R value: a novel statistical quantity for assessing the accuracy of crystal structures, *Nature* 355 (1992) 472–475.
- [26] G.M. Sheldrick, T.R. Schneider, SHELXL: High-resolution refinement, *Meth. Enzymol.* 277 (1997) 319–343.
- [27] P. Auffinger, E. Westhof, Rules governing the orientation of the 2'-hydroxyl group in RNA, *J. Mol. Biol.* 274 (1997) 54–63.
- [28] A. De Mesmaeker, R. Häner, P. Martin, H.E. Moser, Antisense oligonucleotides, *Acc. Chem. Res.* 28 (1995) 366–374.
- [29] S.M. Freier, K.H. Altmann, The ups and downs of nucleic acid stability: structure-stability studies on chemically modified DNA-RNA duplexes, *Nucleic Acids Res.* 25 (1997) 4429–4443.
- [30] K.B. Hall, L.W. Laughlin, Thermodynamic and structural properties of pentamer DNA.DNA, RNA.RNA, and DNA.RNA duplexes of identical sequence, *Biochemistry* 30 (1991) 10606–10613.
- [31] M. Egli, S. Portmann, N. Usman, RNA hydration: a detailed look, *Biochemistry* 35 (1996) 8489–8494.
- [32] D.A. Adamiak, W.R. Rypniewski, J. Milecki, R.W. Adamiak, The 1.19 Å X-ray structure of 2'-O-Me(CGCGCG)₂ duplex shows dehydrated RNA with 2-methyl-2,4-pentanediol in the minor groove, *Nucleic Acids Res.* 29 (2001) 4144–4153.
- [33] P. Lubini, W. Zürcher, M. Egli, Crystal structure of an oligodeoxynucleotide duplex containing 2'-O-methylated adenosines, *Chem. Biol.* 1 (1994) 39–45.
- [34] J.L. Smith, A. Thompson, Reactivity of selenomethionine – dents in the magic bullet? *Structure* 6 (1998) 815–819.
- [35] N. Carrasco, D. Ginsburg, Q. Du, Z. Huang, Synthesis of selenium-derivatized nucleosides and oligonucleotides for X-ray crystallography, *Nucleosides, Nucleotides and, Nucleic Acids* 20 (2001) 1723–1734.
- [36] G. Minasov, M. Teplova, G.C. Stewart, E.V. Koonin, W.F. Anderson, M. Egli, Functional implications from crystal structures of the conserved *Bacillus subtilis* Maf protein with and without dUTP, *Proc. Natl. Acad. Sci. USA* 97 (2000) 6328–6333.