

Cite this: *Chem. Sci.*, 2019, 10, 10948

All publication charges for this article have been paid for by the Royal Society of Chemistry

Solid-phase synthesis and structural characterisation of phosphoroselenolate-modified DNA: a backbone analogue which does not impose conformational bias and facilitates SAD X-ray crystallography†

Patrick F. Conlon,^a Olga Eguagie,^a Jordan J. Wilson,^a Jamie S. T. Sweet,^a Julian Steinhögl,^b Klaudia Englert,^c Oliver G. A. Hancox,^b Christopher J. Law,^d Sarah A. Allman,^b James H. R. Tucker,^c James P. Hall^{b,e} and Joseph S. Vyle^{*a}

Oligodeoxynucleotides incorporating internucleotide phosphoroselenolate linkages have been prepared under solid-phase synthesis conditions using dimer phosphoramidites. These dimers were constructed following the high yielding Michaelis–Arbuzov (M–A) reaction of nucleoside *H*-phosphonate derivatives with 5′-deoxythymidine-5′-selenocyanate and subsequent phosphorylation. Efficient coupling of the dimer phosphoramidites to solid-supported substrates was observed under both manual and automated conditions and required only minor modifications to the standard DNA synthesis cycle. In a further demonstration of the utility of M–A chemistry, the support-bound selenonucleoside was reacted with an *H*-phosphonate and then chain extended using phosphoramidite chemistry. Following initial unmasking of methyl-protected phosphoroselenolate diesters, pure oligodeoxynucleotides were isolated using standard deprotection and purification procedures and subsequently characterised by mass spectrometry and circular dichroism. The CD spectra of both modified and native duplexes derived from self-complementary sequences with A-form, B-form or mixed conformational preferences were essentially superimposable. These sequences were also used to study the effect of the modification upon duplex stability which showed context-dependent destabilisation (−0.4 to −3.1 °C per phosphoroselenolate) when introduced at the 5′-termini of A-form or mixed duplexes or at juxtaposed central loci within a B-form duplex (−1.0 °C per modification). As found with other nucleic acids incorporating selenium, expeditious crystallisation of a modified decanucleotide A-form duplex was observed and the structure solved to a resolution of 1.45 Å. The DNA structure adjacent to the modification was not significantly perturbed. The phosphoroselenolate linkage was found to impart resistance to nuclease activity.

Received 16th August 2019
Accepted 11th October 2019

DOI: 10.1039/c9sc04098f

rsc.li/chemical-science

Introduction

The automated solid-phase synthesis of oligonucleotides using phosphoramidite chemistry¹ (Fig. 1A) is a platform technology which has been transformative in a wide range of applications²

including the measurement^{3,4} and manipulation^{5,6} of gene expression, DNA nanotechnology^{7,8} and data storage.⁹

Whilst simple Watson–Crick base-pairing rules can be used to programme both intramolecular and intermolecular assembly of three-dimensional structures from simple duplexes,^{10,11} the tertiary interactions of folded nucleic acids derived from more complex assemblies and their recognition, especially by nucleic acid binding proteins or small molecule effectors, typically requires structural elucidation at the atomic level.^{12,13} Characterisation of novel structural motifs including such complexes by X-ray crystallography is augmented by the introduction of an anomalous heavy atom scattering centre.^{14,15}

Anomalous scattering of X-rays at the Se–K edge has been of considerable value in facilitating structural biology studies of proteins following the pioneering work of Hendrickson and co-

^aSchool of Chemistry and Chemical Engineering, Queen's University Belfast, David Keir Building, Stranmillis Road, Belfast, BT9 5AG, UK. E-mail: j.vyle@qub.ac.uk

^bReading School of Pharmacy, University of Reading, Whiteknights, Reading RG6 6AP, UK. E-mail: james.hall@reading.ac.uk

^cSchool of Chemistry, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK

^dSchool of Biological Sciences, Queen's University Belfast, 15 Chlorine Gardens, Belfast BT9 5AH, UK

^eDiamond Light Source, Chilton, Didcot, Oxfordshire, OX11 0DE, UK

† Electronic supplementary information (ESI) available: PDB ID: 6S7D. See DOI: 10.1039/c9sc04098f

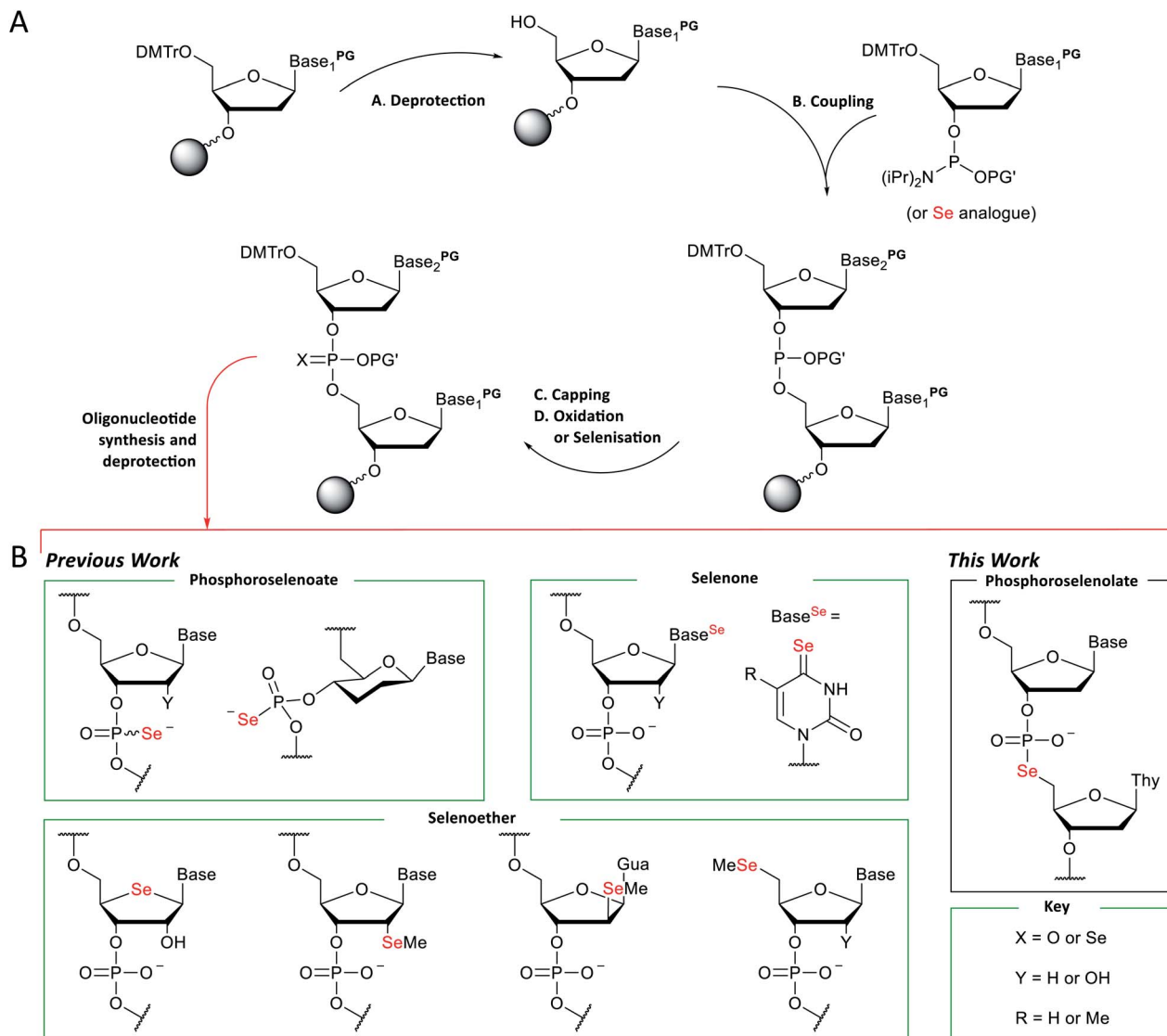


Fig. 1 (A) General scheme for the solid-phase synthesis of oligonucleotides using phosphoramidite chemistry. (B) Examples of selenium-modified nucleic acid analogues prepared using this chemistry.

workers using either selenomethionine¹⁶ or a high-affinity ligand.¹⁷ The molecular biology techniques required to engineer polypeptide sequences containing a mutated selenomethionine residue have now become routine and as this residue does not disrupt protein folding, it is widely used in MAD (multiple wavelength anomalous dispersion) or SAD (single wavelength anomalous diffraction) applications.¹⁸ Concurrent increased access to third generation synchrotron sources which generate more focussed, tunable X-rays of high brightness and stability coupled with automation of the screening of crystallisation conditions has engendered a considerable upsurge in protein crystal structure solution such that in the year 1990, 132 protein structures were deposited in the protein databank whereas in 2018 this number had risen to 10 306. Of the 137 000 protein crystal structures in the PDB repository by the end of 2018, over 10% had been solved using SAD or MAD methods. These methodologies are particularly important for “first in

class” structures of proteins and nucleic acids, where new structural folds or ligand-induced distortions exclude using other approaches to structure solution such as molecular replacement (which relies upon prior knowledge of a structure based upon its sequence). However, the true impact of using Se-SAD and Se-MAD may never fully be known as once the structure of a new fold or motif has been determined, this can then be used as the basis for solving future structures without requiring anomalous data. SAD and MAD phasing is therefore a great facilitator of structural science and is used in drug development programmes worldwide, due to the expeditious ability to determine pharmacophore–target interactions. However, for a structure to be biologically “valid” the incorporation of a heavy atom (such as Se or Br) into the crystal needs to have little or no effect on the conformation or topology of the molecule, whilst also not modifying the molecular arrangement of any potential binding sites. This is a particular problem when working with



nucleic acids as the most commonly used nucleotide analogues, which contain heavy atom groups, either alter the topology of the nucleic acid (shifting its conformation) or change the groups present in either the major or minor groove, both of which are binding sites for most DNA-targeting drugs.

Methodology for the synthesis of selenium-modified nucleic acid (Fig. 1B) sequences under solid-phase conditions was first described three decades ago.¹⁹ In this report, selenisation of support-bound *H*-phosphonate precursors required treatment every 3–4 hours over three or four days and gave mixtures of the diastereomeric phosphoroselenoates. Subsequently, more efficient selenisation protocols^{20,21} and stereoselective synthesis²² have been developed and this modification has enabled the solution of refractory structures such as homo-DNA.²³ However, the sensitivity of intermediates during oligomerisation and short half-life of the final product (*ca.* 30 days under ambient conditions in aqueous solution) has limited their application, especially *in vivo*. Likewise, nucleobase analogues which incorporate a selenone function can be accessed using concise synthetic pathways^{24–28} but are sensitive towards oxidation, hydrolysis and photochemical degradation.

A third class of selenium-modified oligonucleotides in which selenoether functions have been introduced at the 2',^{29–32} 4',^{33,34} or terminal 5'-positions³⁵ or appended to a nucleobase³⁶ exhibit greater resilience towards the conditions typically employed during standard solid-phase oligonucleotide synthesis using phosphoramidite chemistry although mitigation of side-reactions arising from selenoether oxidation is required^{37,38} and replacement of an oxygen at a stereogenic centre provides additional synthetic demands, *e.g.*, up to 14-steps are required to prepare 4'-seleno-4'-deoxynucleosides from *D*-ribose.^{39,40} The 2'-selenomethyl ether analogues (especially pyrimidine derivatives) are considerably more accessible (requiring only six steps to prepare on large scale)⁴¹ and can also promote crystallisation of modified oligomers.^{29,42–45} However, although the furanoside ring of individual nucleosides modified in this fashion adopt a conformation typically found in B-form DNA,⁴⁶ when incorporated into oligodeoxynucleotide duplexes, the location of the 2'-selenomethyl function in the minor groove provides conformational steering towards the formation of RNA-like A-form duplexes. Although the epimeric arabino-2'-selenomethyl analogue reverses this trend,³¹ full details of its preparation have not yet been reported and the steric bulk of both isomers can interfere with groove-binding interactions.⁴⁶

Replacement of a bridging oxygen (3' or 5') within an internucleotide phosphate diester by selenium gives rise to phosphoroselenolate analogues. These are achiral and potentially more stable than the corresponding phosphoroselenoates but have been documented in only a very limited number of communications.^{47–49}

Alkylation of the lead salt of *O*,*O*-diethylphosphoroselenoate was first reported in 1911⁵⁰ and this strategy has been applied to the preparation of internucleotide phosphoroselenolate linkages using 5'-halothymidine derivatives either in DMF⁴⁷ or under aqueous conditions.⁴⁸ In the former case, characterisation of the dimer product was equivocal indicating that perhaps both *O*- and Se-alkylation had occurred.^{51–53}

Subsequently, Xu and Kool exploited related chemistry for the synthesis of oligomers incorporating this analogue following template-directed chemical ligation. However, consistent with the extreme sensitivity of phosphoroselenoate monoesters towards ariel oxidation in aqueous solution,⁵⁴ incomplete reaction was observed. In contrast, Michaelis-Arbuzov (M-A) reaction of a nucleoside 3'-*H*-phosphonate with a 5'-selenocyanate enabled synthesis of a fully-characterised phosphoroselenolate-bridged dimer (TpSedT) with high efficiency.⁴⁹

Herein, we describe the further developments to this methodology enabling the isolation of pure dinucleoside phosphoroselenolate triesters in good to excellent yields, their subsequent phosphitylation and efficient introduction into oligodeoxynucleotides under solid-phase conditions. Furthermore, we report a model DNA crystal structure incorporating this modification which evinces an A-form morphology with minimal perturbation of the backbone.

Results and discussion

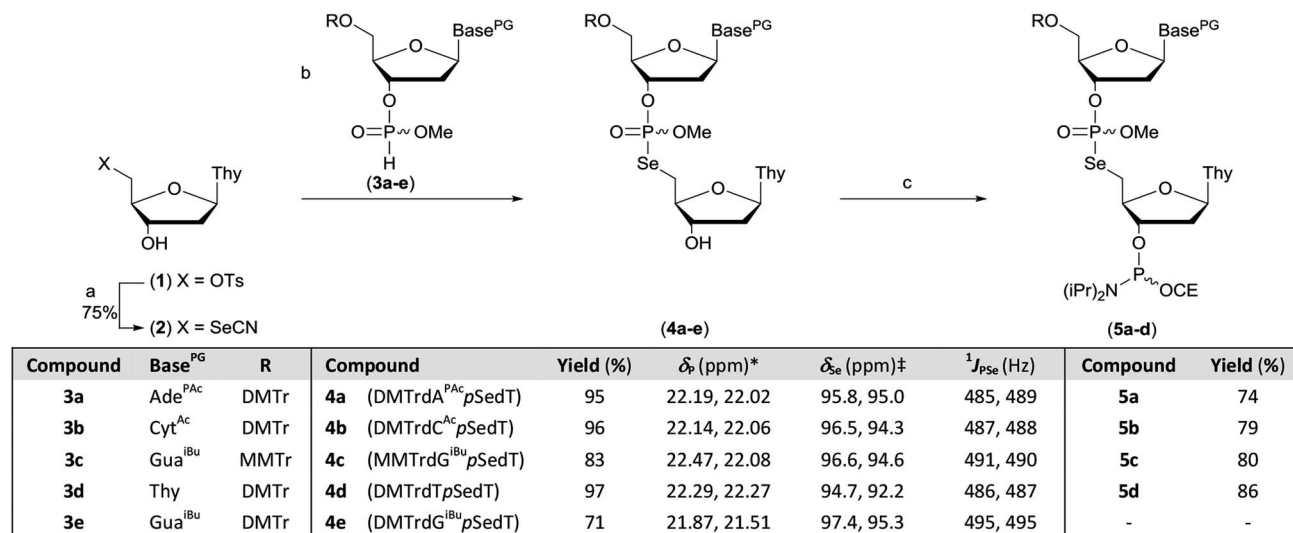
Preparation of dinucleotide phosphoramidites

The protected M-A product from the synthesis of TpSedT⁴⁹ appeared to be an attractive target as efficient dimer coupling has been reported during the solid-phase construction of oligonucleotides containing sulfur.^{55,56} Furthermore, this strategy would avoid the preparation and use of intermediates containing a labile chalcogen-P(III) bond.^{57,58} We therefore sought to optimise the synthesis and isolation of the protected dinucleoside phosphoroselenolate triester from the corresponding selenocyanate.

5'-Deoxythymidine-5'-selenocyanate has been prepared on small scales in solution over 24 hours⁵⁹ or using liquid assisted grinding over 9–11 hours.^{60,61} In order to prepare this material more efficiently, we therefore examined the application of microwaves which has been reported to enhance the rate of *S*-selective alkylation of the thiocyanate anion.⁶² In the presence of excess potassium selenocyanate, clean transformation of 5'-tosylthymidine (**1**) into the corresponding 5'-selenocyanate (**2**) was observed within 90 minutes following irradiation (20 W, 100 °C) of the stirred reaction mixture in acetonitrile (Scheme 1). As previously observed,⁶⁰ quenching of the excess selenocyanate with benzyl bromide greatly facilitated subsequent purification by silica gel chromatography and pure 5'-deoxythymidine 5'-selenocyanate (**2**) was isolated in 75% yield.

During optimisation of both the reaction conditions and subsequent purification of the M-A product derived from **2** and cyanoethyl-protected nucleoside 3'-*H*-phosphonate diesters, several issues became apparent. Specifically, the neutral silylating agent, *N*,*O*-bis(trimethylsilyl)acetamide (BSA), which served both to render the selenonucleoside soluble in chloroform and enhance the rate of the M-A reaction also promoted decyanoethylation of both the starting material and the product. Furthermore, subsequent quenching of the reaction following addition of water resulted in detritylation in the absence of a strong organic base but if such a base was used





Scheme 1 Preparation of dinucleotide phosphoramidites **5a–d**. Reagents and conditions: (a) (i) KSeCN (1.5 eq.), MeCN, MW power 20 W, 100 °C, 90 min; (ii) BnBr (0.6 eq.), MeOH, rt, 60 min. (b) **3a–e** (1.5 eq.), 2,6-lutidine (5 eq.), MeCN, rt, 30–60 min. (c) (i) CIP(OCE)NiPr₂ (2 eq.), iPr₂NEt (3 eq.), DCM, rt, 45 min; (ii) MeOH (1.5 eq.), rt, 15 min. *in MeCN (**4d**) or DCM, †in DMSO-*d*₆. Key. Base^{PG}: Ade^{PAC}: 9-(*N*⁶-phenoxyacetyl)adeninyl. Cyt^{Ac}: 1-(*N*⁴-acetylcytosinyl). Gua^{IBu}: 9-(*N*²-isobutyryl)guaninyl. Thy: 1-thyminyl. DMTr: 4,4'-dimethoxytrityl. MMTr: 4-methoxytrityl. CE: 2-cyanoethyl. *p*: P(O)OMe. Ts: *p*-toluenesulfonyl.

during work-up, further decyanoethylation of the phosphor-selenolate was evident.⁶³

In order to ameliorate issues associated with β -elimination of the cyanoethyl function, methyl-protection of the phosphate was adopted. This provides greater resilience towards phosphate deprotection in the presence of stronger organic bases such as triethylamine and has recently been utilised for the preparation of di- and trinucleotide phosphoramidites.^{63–65} The methyl-protected 3'-*H*-phosphonates (**3a–e**) were prepared from the corresponding phosphoramidites under modified literature conditions.⁶⁶ Following extractive work-up, the materials were pure by ³¹P NMR.

However, in our hands, these *H*-phosphonates were less stable towards dephosphorylation than those derived from the cyanoethyl-protected precursors and low levels of the nucleoside congeners were identified by TLC as contaminants. Thus, once isolated, the solids were stored at –20 °C and used within 24 hours. To accommodate the presence of these impurities, excess *H*-phosphonate (1.5 eq.) was used in the M–A reaction. In a further change to the previous published procedure, acetonitrile was used as solvent and in the presence of 2,6-lutidine, nucleoside selenocyanate **2** could be dissolved in this mixture under gentle heating without silylation. Addition of this solution to thymidine *H*-phosphonate (**3d**) in acetonitrile at room temperature gave complete and clean M–A reaction within 30 minutes. We note that these conditions contrast with those typically employed to promote such reactions with chalcogen electrophiles following conversion of the *H*-phosphonate into the corresponding phosphite tautomer in the presence of a silylating agent or a strong base. The protected dimers (**4a–e**) were isolated as mixtures of diastereoisomers following silica gel column chromatography. Phosphitylation of the dimers under modified literature conditions⁶⁷ enabled isolation of the corresponding phosphoramidites derived from 5'-dimethoxytrityl-protected dA

(**5a**), dC (**5b**) or T (**5d**) and 5'-monomethoxytrityl-protected dG (**5c**) in greater than 70% yields. During attempted processing of **4e** (DMTrdG^{IBu}*p*SedT) under the same conditions, detritylation was evident and therefore further reaction to the corresponding phosphoramidite was not attempted.

Analysis of the dimer phosphoramidites by ³¹P NMR showed expected resonances associated with two separate diastereomeric P(III) and P(V) phosphorus centres (e.g., Fig. 2). Low levels (<4%) of putative *H*-phosphonate derivatives were also apparent and as this can be associated with autocatalytic degradation,^{68,69} the stability of a dimer phosphoramidite in solution was examined by ³¹P NMR. At 4 °C, **5a** (0.1 M in acetonitrile)

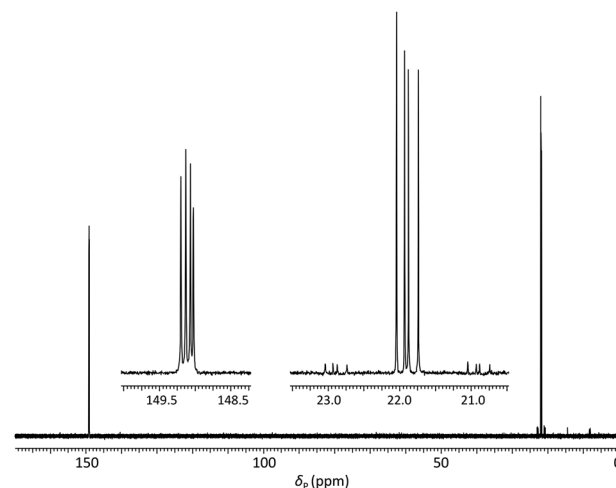


Fig. 2 ³¹P{¹H} NMR of phosphoramidite **5a** (prepared from DMTrdA-^{PAC}*p*SedT – **4a**) in MeCN with (inset) expansions of resonances associated with P(III) and P(V) nuclei.



remained untransformed over 24 days. A separately prepared solution of the same composition was stored on the DNA synthesiser at ambient temperature. After 24 days, solid deposition was observed and ^{31}P NMR analysis of the combined materials (solid and liquid) showed over 80% degradation of the phosphoramidite although no transformation of the phosphoroselenolate function was evident.

Preparation of phosphoroselenolate-modified oligodeoxynucleotides under solid-phase conditions *via* coupling of dimer phosphoramidites

Optimisation of solid-phase synthesis conditions was performed with the dimer phosphoramidite **5d** (derived from DMTrdTpSedT) using manual syringe addition of all reagents and acetonitrile washes to prepare a model homothymidine-derived pentamer incorporating two phosphoroselenolate linkages (**ODN 1** – Table 1). Under optimised conditions, quantitative coupling (by trityl release) of **5d** in the presence of the activator 5-*S*-benzylthiotetrazole was observed over five minutes with subsequent oxidation using 0.02 M iodine. Duplicate, one μmol -scale syntheses of the trityl-on, model pentamer were performed. Removal of phosphate diester protecting groups and cleavage from the controlled pore glass (CPG) support required a two-step procedure. Initial demethylation of the phosphoroselenolate triesters was effected following treatment with a solution of 0.2 M sodium diethyldithiocarbamate (NaDEC) in acetonitrile for 30 minutes at room temperature. This method was found to be equally efficient in a side-by-side comparison with standard literature conditions using 1 : 2 : 2 (v/v/v) thiophenol/triethylamine/dioxane.⁵⁵ Excess reagents were removed following washing with acetonitrile, the CPG dried and suspended in 1 : 1 (v/v) 35% (w/v) NH_3 (aq)/40% (w/v) MeNH_2 (aq) (AMA) at room temperature for 60 minutes. Following removal of volatile amines, the crude tritylated oligomer **ODN 1** was purified using reversed phase HPLC. Finally, treatment with aqueous acetic acid under standard conditions effected cleavage of the 5'-DMTr function and following desalting, pure **ODN 1** was isolated in 33% yield. Analysis of **ODN 1** by ^{31}P NMR (Fig. 3) showed four resonances associated with the two internucleotide phosphate diesters ($\delta_{\text{P}} = -1.19, -1.25$) and two internucleotide phosphoroselenolate diesters ($\delta_{\text{P}} = 10.56, 10.48$).

Satellite peaks associated with ^{31}P – ^{77}Se coupling were consistent with a phosphorus–selenium single bond ($J_{\text{PSe}} =$

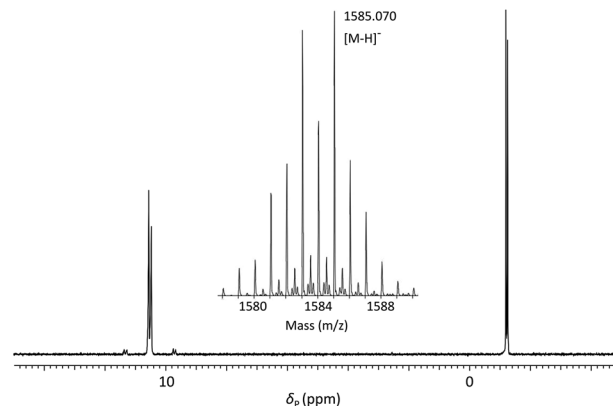


Fig. 3 $^{31}\text{P}\{^1\text{H}\}$ NMR of **ODN 1** in 60 : 40 (v/v) H_2O : D_2O and (inset) ESI-mass spectrum showing isotope pattern associated with molecular ion.

393 Hz and 390 Hz)⁷⁰ although somewhat weaker than that found in the corresponding triester.

ODN 1 was further characterised using ESI-MS (Fig. 3 – inset) which gave a molecular ion peak ($[\text{M} - \text{H}]^-$) displaying the expected isotope distribution pattern.

In order to further examine the scope of the manual dimer phosphoramidite coupling and deprotection protocol developed for **ODN 1**, a series of self-complementary 10-mer sequences designed to form A-form duplexes (**ODN 2–5**) were prepared. Chain extension of support-bound octanucleotides using phosphoramidite **5d** was performed using the same cycle as previously described. However, less efficient couplings were observed using **5a–c** and therefore an extended double coupling⁷¹ was employed with these phosphoramidites.

Following oxidation, the columns were washed with acetonitrile and dried under argon. Deprotection using NaDEC followed by AMA liberated the crude, tritylated oligodeoxynucleotides which were purified by RP-HPLC and reduced *in vacuo*. Detritylation of the purified oligomers was accomplished under literature conditions using an extended treatment to fully remove the monomethoxytrityl function from **ODN 4**.

The modified Dickerson–Drew dodecamer⁷² sequence **ODN 6** was prepared under automated conditions using the phosphoroselenolate-linked dimer phosphoramidite **5a** which was incorporated in place of the central AT motif. In addition to the extended coupling of **5a**, the standard cycle was modified to

Table 1 Phosphoroselenolate-modified oligodeoxynucleotide sequences prepared on solid support using dinucleotides phosphoramidites **5a–d**

Phosphoramidite (coupling time)	Oligodeoxynucleotide sequence 5' → 3'	Yield ^a (%)	Mol. ion ^b (calcd)
5d (5 min)	ODN 1 d(TSeTTSeTT)	657 nmol (33%)	1585.070 (1585.112)
5a (2 × 7.5 min)	ODN 2 d(AsTCCCGGGAT)	541 nmol (27%)	3090.370 (3090.472)
5b (2 × 7.5 min)	ODN 3 d(CSeTCCCGGGAG)	251 nmol (13%)	3091.365 (3091.468)
5c (2 × 7.5 min)	ODN 4 d(GSeTCCCGGGAC)	484 nmol (24%)	3091.351 (3091.468)
5d (5 min)	ODN 5 d(TSeTCCCGGGAA)	337 nmol (17%)	3090.360 (3090.472)
5a (2 × 7.5 min)	ODN 6 d(CGCGAAsSeTTCGCG)	198 nmol (20%)*	3708.450 (3708.450)

^a Determined using $\epsilon_{260 \text{ nm}}$ of the corresponding native sequence on a 2 μmol scale (* from 1 μmol). ^b m/z $[\text{M} - \text{H}]^-$ (^{80}Se).



include a wash using 10% (v/v) pyridine/MeCN following the oxidation step. In the absence of this wash, the CPG retained iodine residues which were found to engender lower yields of the modified oligomers. The subsequent five monomer coupling cycles were performed under standard conditions and gave efficient nucleotide incorporation according to trityl yields.

Once complete, the support-bound dodecamer was treated with a solution of 150 mM dithiothreitol (DTT) in 1 : 1 (v/v) H₂O : absolute ethanol to remove potential oxidised selenium species^{37,45} and further demethylated using NaDEC as previously described. Further deprotection with AMA enabled the pure trityl-on oligomer to be isolated following RP-HPLC purification. This material was then detritylated and desalted under standard conditions to afford pure **ODN 6** in 20% yield. Under the same conditions the corresponding native sequence was isolated in 45% yield. All selenium-containing oligodeoxynucleotides were characterised by mass spectrometry in which the *m/z* ratios exhibited the expected isotope distribution patterns and did not show any peaks associated with selenoxide formation.

Preparation of a model phosphoroselenolate-modified oligodeoxynucleotide under solid-phase conditions via Michaelis–Arbuzov chemistry

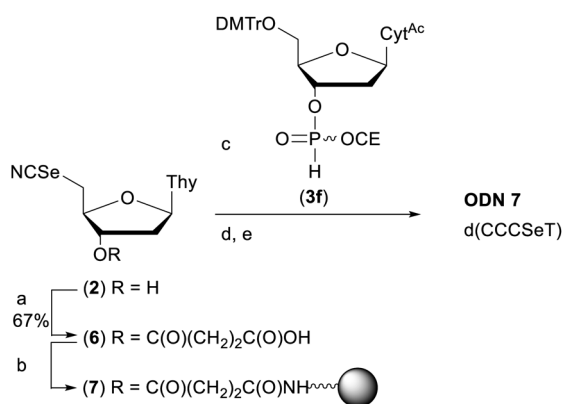
In order to demonstrate the general utility of the M–A reaction for constructing internucleotide phosphoroselenolates under solid-phase conditions, this chemistry was used to prepare a tetranucleotide sequence which, in its native form, crystallises in the four-stranded, i-motif conformation.^{73,74} Thus, **ODN 7** was prepared from the corresponding support-bound nucleoside selenocyanate (Scheme 2). Succinylation of selenonucleoside **2** under standard conditions⁷⁵ gave complete consumption of starting material. Following extraction with cold citric acid, the succinyl ester (**6**) was isolated contaminated with the corresponding symmetrical diselenide (30 mol%). This crude material was treated with excess condensing agent (DCC) in the presence of 4-nitrophenol⁷⁶ and following removal of

precipitated DCU, the mixture of activated esters reacted with amino-functionalised controlled pore glass. The CPG-bound selenonucleoside (**7**) thus obtained was reacted with the nucleoside *H*-phosphonate **3f**. After 60 minutes, trityl assay indicated that the support loading (39 μmol g^{−1}) was consistent with that typically used in solid-phase DNA synthesis indicating that efficient M–A reaction had occurred.

Using this material (1.75 μmol), chain extension was performed under manual conditions using standard monomer phosphoramidites using the cycle developed for **ODN 1**. Subsequent deprotection with AMA liberated the crude tritylated tetramer which was purified by RP-HPLC, detritylated and desalted to afford pure **ODN 7**.

Enzymatic digestion

Further confirmation of the integrity of the internucleotide phosphoroselenolate linkage during both coupling and downstream processing was provided by enzymatic digestion of both **ODN 5** and the corresponding native sequence. Using commercially-sourced snake venom with both phosphodiesterase and phosphatase activities, complete digestion of the native 10-mer to the corresponding nucleosides was found within eight hours at 37 °C (Fig. S2†). Under the same conditions, **ODN 5** liberated the nucleosides dA, dC and dG in the expected ratios but less than 1 mol% of T (*t_R* = 18.1 min) was observed (Fig. 4). However, a single major peak with significantly longer retention time (*t_R* = 38.7 min) was identified as the dimer TpSedT following coinjection with an authentic sample.⁴⁹ Previously, Stec and coworkers examined nuclease digestion of putative TpSedT and described using ten times the quantity of enzyme required to effect cleavage of the corresponding phosphorothioate substrate. During this treatment, formation of the corresponding diselenide; (SedT)₂, was observed and this compound is assumed to be the origin of a novel hydrophobic material (*t_R* = 45.1 min) which was also apparent during extended treatment of pure TpSedT with snake venom. Early studies on the digestion of dimers incorporating a 5'-thiothymidylate (e.g., TpSdT) also reported slow cleavage activities^{77,78} although Xu and Kool subsequently described snake venom digestion of a 20-mer containing a single internal modification which was shown to proceed at essentially the



Scheme 2 Preparation of **ODN 7**. Reagents and conditions: (a) succinic anhydride (1 eq.), DMAP (0.8 eq.), pyridine, rt, overnight; (b) (i) *p*-nitrophenol (1 eq.), DCC (2.5 eq.), DCM, rt, 2 h; (ii) amino SynBase 500/110 CPG (500 mg), DMF, Et₃N (4 eq.), rt, 5 h. (c) **3f**, MeCN, rt, 1 h; (d) solid-phase DNA synthesis using manual syringe addition. (e) (i) AMA, rt, 2 h. (ii) 80% AcOH, rt, 1 h. See ESI† for full details.

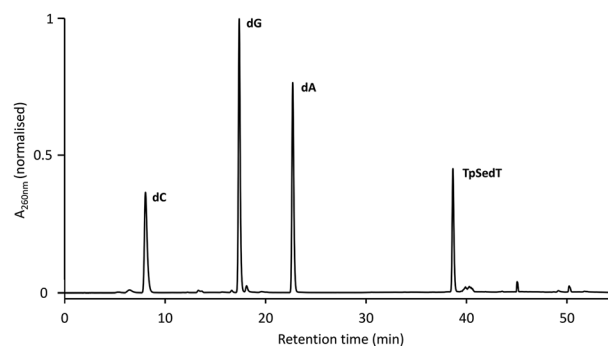


Fig. 4 RP-HPLC chromatogram of the snake venom digest of **ODN 5** (after 8 hours at 37 °C). For full conditions see ESI.†

same rate as the unmodified sequence and cleaved the internal phosphorothiolate linkage.⁷⁹

Although the scissile bond (3'O-P) is not modified in the phosphoroselenolate analogues examined in this study, it is assumed that distortion in the phosphoryl moiety arising out of the extended bond lengths and compressed bond angle associated with the P-Se-C5' bridge (*vide infra*) inhibits productive binding of the dimer.

UV thermal denaturation studies

The effect of the phosphoroselenolate linkage upon duplex stability was investigated by comparing thermal denaturation of the modified oligomers with the corresponding native sequences (Table 2). Sequence-dependent destabilisation of duplexes designed to exhibit A-form conformations was observed following phosphoroselenolate substitution of the 5'-terminal phosphodiester linkage (entries 1–4). The largest reductions in melting temperatures were observed for sequences in which the native oligomers had the greatest stabilities by virtue of the presence of a terminal C-G (entry 2) or G-C (entry 3) base pair and in these circumstances, ΔT_m values of -6.2°C and -4.9°C (respectively) were observed. In contrast, introduction of a phosphoroselenolate-linkage downstream from a terminal A-T base pair (entry 1) had a minor effect upon duplex stability ($\Delta T_m = -0.7^\circ\text{C}$) although the effect upon the T-A (entry 4) pairing was more significant ($\Delta T_m = -4.0^\circ\text{C}$).

The effect of introducing juxtaposed phosphoroselenolate bridges into the more extended backbone typical of B-form helices was tested using the Dickerson–Drew dodecamer sequence **ODN 6** (entry 5). A decrease in stability of 1.9°C was observed in comparison with the native oligomer. Xu and Kool reported the effect upon duplex stability of substituting a single internal phosphate diester-linkage within a 17-mer (B-form) sequence with either a phosphorothiolate or phosphoroselenolate analogue.⁴⁸ Duplexes formed between these chalcogen-modified oligomers and the native complement displayed ΔT_m values of -3.0°C (Se) and -2.0°C (S). These values are consistent with those obtained in the current study in which two modifications are introduced per duplex.

Solution phase conformation studies

The effects of introducing the phosphoroselenolate modification upon the conformations of self-complementary duplex DNA sequences were investigated using circular dichroism. The spectral shape and peak positions of native and modified

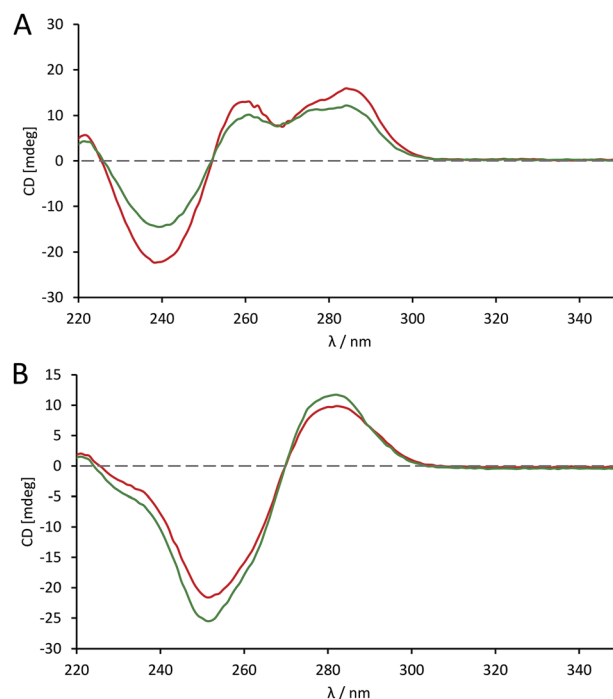


Fig. 5 CD spectra comparing the conformation of phosphoroselenolate-modified sequences (RED) and the corresponding native (all phosphodiester) sequences (GREEN). (A) ODN 4; (B) ODN 6.

duplexes displayed a high level of congruence, with only small changes in band intensity observed (Fig. S5–S9†). For example, conservation of the mixed A/B conformation adopted in solution by **ODN 4** (Fig. 5A) highlights that the modification is not perturbing or directing the conformation of the duplex in a borderline case.

The advantage of the phosphoroselenolate modification over ribose modifications, such as replacement of O with Se in the ring, which typically leads to formation of an A-form duplex,³³ is clear. It should also be noted that both the native and modified dodecamer (Fig. 5B) exhibited the classical B-form again illustrating that the modification has had little effect on the solution conformation of the oligonucleotide. In all cases the incorporation of the Se into the DNA backbone means that the environment in the major and minor grooves is unaffected, unlike with modifications such as 5-bromo-dU or most selenium modified DNA. Whilst some of these modifications (such as a 2'-SeMe analogues)⁴² may not significantly affect the native conformation if this is A-form DNA,

Table 2 Melting temperatures (T_m) of self-complementary phosphoroselenolate-modified and native oligodeoxynucleotide duplexes

Entry	Sequence 5' → 3'	X =	$T_m/^\circ\text{C}$	X =	$T_m/^\circ\text{C}$	$\Delta T_m^a/^\circ\text{C}$
1	d(AXCCCGGGAT)	SedT (ODN 2)	46.9	T	47.6	-0.7
2	d(CXCCCGGGAG)	SedT (ODN 3)	48.1	T	54.3	-6.2
3	d(GXCCCGGGAC)	SedT (ODN 4)	50.7	T	55.6	-4.9
4	d(TXCCCGGGAA)	SedT (ODN 5)	43.0	T	47.0	-4.0
5	d(CGCGAAXTCGCG)	SedT (ODN 6)	55.9	T	57.8	-1.9

^a $\Delta T_m = T_m(\text{X} = \text{SedT}) - T_m(\text{X} = \text{T})$. Conditions: ssDNA (10 μM), NaP_i (10 mM, pH 7.0), NaCl (100 mM).



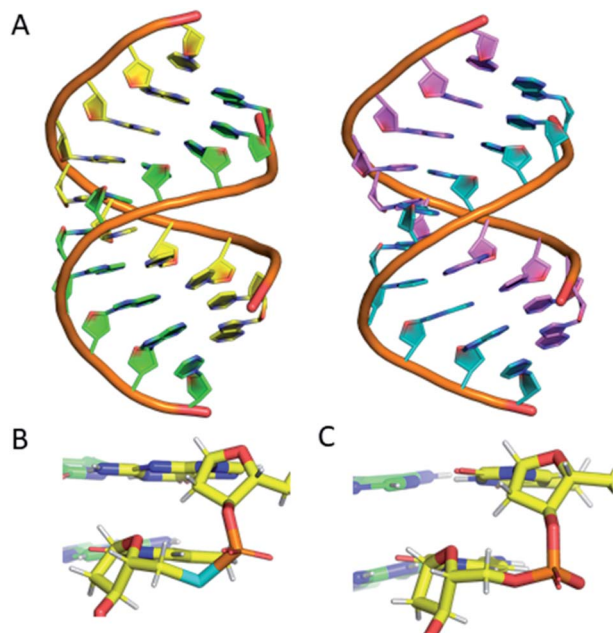


Fig. 6 (A) (left) The DNA structure reported here (ODN 4) and (right) a standard A-form DNA 10-mer.⁸¹ (B) The phosphoroselenolate modification (cyan) compared with (C) the naturally occurring backbone at the terminal AC step. The chains are coloured according to atom type, with carbon in red, green, cyan or magenta (as a function of DNA chain), nitrogen in blue, oxygen in red, selenium in cyan, phosphorus in orange and hydrogen in white. PDB ID: 6S7D.†

the addition of groups into possible binding sites could affect the validity of drug–DNA binding studies, which is not an issue with a phosphoroselenolate modification.

Crystal structure

To confirm that the structure of the nucleic acid is not significantly perturbed by the introduction of a phosphoroselenolate modification, crystallisation trials of oligodeoxynucleotide sequences containing the group were performed. This resulted in the crystal structure of d(GSeTCCCGGGAC) – ODN 4 – being obtained, which allows for visualisation of the modification at 1.45 Å resolution.

The structure shows that the overall conformation of the duplex is that of A-DNA (Fig. 6A). The average twist of the helix is 31.8° which compares favourably to a standard value of 32.7°. However, at the modification site the T*–A base pair has a twist of ca. 41°, which can be attributed to the increased bond lengths from P–Se (2.25 Å) and Se–C5' (1.94 Å) compared to the oxygen-containing equivalent (P–O5' 1.60 Å, O5'–C5' 1.42 Å).

Apart from this slight increase in bond lengths, and subsequent increase in helical twist angle, the base pair containing the phosphoroselenolate modification is otherwise consistent with a standard A-DNA base pair (Fig. 6B and C), with parameters such as roll, tilt, rise and slide all within the normal range for this conformation (analysis performed using W3DNA).⁸⁰ Whilst it could be expected that the increased twist or bond lengths could affect the sugar pucker of the nucleotide, this is

not the case, with the nucleotide possessing a C3'-endo sugar pucker, consistent with that of A-DNA.

The nucleotide 3' to the modified base also possesses a C3'-endo pucker. The base 5' to the modification site possesses a non-standard pucker, C2'-exo, which is attributed to the proximity of a neighbouring strand, forming a crystal contact within the lattice. This structure therefore confirms that the introduction of the modification does not significantly perturb the solid-state conformation of the duplex.

Conclusions

For the first time, a phosphoroselenolate-modified oligomer has been prepared *de novo* using phosphoramidite chemistry with only minor modification to the standard solid-phase DNA synthesis cycle. All modified oligonucleotides were characterised by electrospray and/or MALDI mass spectrometry which gave molecular ion masses (with appropriate isotope distribution patterns) consistent with the single-stranded sequence and if self-complementary, the corresponding duplex. No M+16 peaks were apparent in these spectra suggesting that either the oxidation conditions did not lead to oxygen transfer to selenium or that downstream processing of such modified oligomers leads to their removal. This was further confirmed by ³¹P NMR analysis of a double-labelled model pentamer.

Unlike previous generations of DNA analogues in which selenium is introduced within the sugar moiety, the phosphoroselenolate modification described here is readily accessible requiring only four high yielding steps and does not require handling highly air-sensitive materials. Furthermore, the furanose pseudorotation angles of the most common modification, 2'-SeMe, typically results in conformational steering towards an A-form duplex. Using the related 2'-SMe analogue, such conformational restrictions were not observed and both A- and B-DNA duplex crystal structures were solved^{82,83} although longer collection times (and concomitant DNA damage) are associated with the weaker anomalous scattering from sulfur at the CuK_α wavelength. Furthermore, both analogues interfere with minor groove binding interactions. Finally, the expeditious crystallisation processes observed using other selenium-modified nucleic acids was maintained with the phosphoroselenolate.

Selenium-SAD has recently been applied to the *de novo* phasing of data generated from an X-ray-free electron laser including serial femtosecond crystallography^{84,85} and therefore offers the potential to observe conformational changes within nucleic acids during folding, recognition and processing.

The resistance of the internucleoside phosphoroselenolate linkage towards the exonuclease activity of snake venom offers the potential to both facilitate *in vivo* activity of therapeutic nucleic acid sequences⁸⁶ and to probe biochemical processes in the same fashion as have phosphorothiolate^{56,87} and we envisage that the ability to label an oligonucleotide backbone using 100% ⁷⁷Se may provide a valuable tool for investigating the biological processing of such materials.⁸⁸ Towards these future goals we are currently investigating the application of



this chemistry for the preparation of phosphoroselenolate-linkages derived from other selenonucleoside analogues.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

This work was funded by: DEL (P. F. C.); EPSRC (EP/R511602/1, EP/G007578/1 – O. E.); University of Reading, School of Pharmacy (O. G. A. H.); ERASMUS+ (J. S.); BBSRC equipment grant BB/M025624/1 “Next Generation DNA Synthesis” (J. H. R. T.) and by the authors (J. S. V., O. E.). J. P. H. gratefully acknowledges the provision of beamtime at Diamond Light Source (MX20077-1 and SM20214-1). We thank Rick Cosstick/James Gaynor (Liverpool) for their kind support in automated solid-phase synthesis. Cameron Thorpe (Oxford) and Sabine Muller/Bettina Appel (Greifswald) provided constructive discussions on the phosphorylation procedure. Andrew Cummings (Q. U. B.) provided technical support for microwave reactions. Catherine McKeen (LGC Link) provided useful discussions relating to support bound octamers. We acknowledge Darren Baskerville and Conor McGrann (Q. U. B.) and The Centre for Chemical and Materials Analysis in the School of Chemistry at the University of Birmingham for running mass spectra. Richard Murphy provided NMR technical support. J. S., O. G. A. H., S. A. A. & J. P. H. would like to thank the Chemical Analysis Facility at the University of Reading for access to instruments.

Notes and references

- M. H. Caruthers, *Science*, 1985, **230**, 281–285.
- S. Kosuri and G. M. Church, *Nat. Methods*, 2014, **11**, 499–507.
- R. K. Saiki, S. Scharf, F. Faloona, K. B. Mullis, G. T. Horn, H. A. Erlich and N. Arnheim, *Science*, 1985, **230**, 1350–1354.
- C. A. Heid, J. Stevens, K. J. Livak and P. M. Williams, *Genome Res.*, 1996, **6**, 986–994.
- M. Smith, *Annu. Rev. Genet.*, 1985, **19**, 423–462.
- M. Jinek, K. Chylinski, I. Fonfara, M. Hauer, J. A. Doudna and E. Charpentier, *Science*, 2012, **337**, 816–821.
- H. Shen, Y. Wang, J. Wang, Z. Li and Q. Yuan, *ACS Appl. Mater. Interfaces*, 2019, **11**, 13859–13873.
- N. C. Seeman and H. F. Sleiman, *Nat. Rev. Mater.*, 2017, **3**, 17068.
- L. Ceze, J. Nivala and K. Strauss, *Nat. Rev. Genet.*, 2019, **20**, 456–466.
- P. W. K. Rothmund, *Nature*, 2006, **440**, 297–302.
- E. S. Andersen, M. Dong, M. M. Nielsen, K. Jahn, R. Subramani, W. Mamdouh, M. M. Golas, B. Sander, H. Stark, C. L. P. Oliveira, J. S. Pedersen, V. Birkedal, F. Besenbacher, K. V. Gothelf and J. Kjems, *Nature*, 2009, **459**, 73–76.
- J. Sheng, J. Gan and Z. Huang, *Med. Res. Rev.*, 2013, **33**, 1119–1173.
- R. Satange, C. K. Chang and M. H. Hou, *Nucleic Acids Res.*, 2018, **46**, 6416–6434.
- W. A. Hendrickson, *Q. Rev. Biophys.*, 2014, **47**, 49–93.
- B. T. Wimberly, D. E. Brodersen, W. M. Clemons, R. J. Morgan-Warren, A. P. Carter, C. Vornheln, T. Hartsch and V. Ramakrishnan, *Nature*, 2000, **407**, 327–339.
- W. A. Hendrickson, J. R. Horton and D. M. LeMaster, *EMBO J.*, 1990, **9**, 1665–1672.
- W. A. Hendrickson, A. Pähler, J. L. Smith, Y. Satow, E. A. Merritt and R. P. Phizackerley, *Proc. Natl. Acad. Sci. U. S. A.*, 1989, **86**, 2190–2194.
- X. D. Su, H. Zhang, T. C. Terwilliger, A. Liljas, J. Xiao and Y. Dong, *Crystallogr. Rev.*, 2015, **21**, 122–153.
- K. Mori, C. Boiziau, C. Cazenave, M. Matsukura, C. Subasinghe, J. S. Cohen, S. Broder, J. J. Toulmé and C. A. Stein, *Nucleic Acids Res.*, 1989, **17**, 8207–8219.
- C. J. Wilds, R. Pattanayek, C. Pan, Z. Wawrzak and M. Egli, *J. Am. Chem. Soc.*, 2002, **124**, 14910–14916.
- K. Tram, X. Wang and H. Yan, *Org. Lett.*, 2007, **9**, 5103–5106.
- P. Guga, A. Maciaszek and W. J. Stec, *Org. Lett.*, 2005, **7**, 3901–3904.
- M. Egli, P. Lubini and P. S. Pallan, *Chem. Soc. Rev.*, 2007, **36**, 31–45.
- J. Salon, J. Sheng, J. Jiang, G. Chen, J. Caton-Williams and Z. Huang, *J. Am. Chem. Soc.*, 2007, **129**, 4862–4863.
- J. Salon, J. Jiang, J. Sheng, O. O. Gerlits and Z. Huang, *Nucleic Acids Res.*, 2008, **36**, 7009–7018.
- A. E. A. Hassan, J. Sheng, W. Zhang and Z. Huang, *J. Am. Chem. Soc.*, 2010, **132**, 2120–2121.
- T. Habuchi, T. Yamaguchi, H. Aoyama, M. Horiba, K. R. Ito and S. Obika, *J. Org. Chem.*, 2019, **84**, 1430–1439.
- C. Liczner, V. Grenier and C. J. Wilds, *Tetrahedron Lett.*, 2018, **59**, 38–41.
- Q. Du, N. Carrasco, M. Teplova, C. J. Wilds, M. Egli and Z. Huang, *J. Am. Chem. Soc.*, 2002, **124**, 24–25.
- C. Höbartner and R. Micura, *J. Am. Chem. Soc.*, 2004, **126**, 1141–1149.
- W. Zhang, J. W. Szostak and Z. Huang, *Front. Chem. Sci. Eng.*, 2016, **10**, 196–202.
- K. Morihiro, T. Kodama, Kentefu, Y. Moai, R. N. Veedu and S. Obika, *Angew. Chem., Int. Ed.*, 2013, **52**, 5074–5078.
- J. K. Watts, B. D. Johnston, K. Jayakanthan, A. S. Wahba, B. M. Pinto and M. J. Damha, *J. Am. Chem. Soc.*, 2008, **130**, 8578–8579.
- Y. Inagaki, N. Minakawa and A. Matsuda, *Nucleic Acids Symp. Ser.*, 2008, **52**, 329–330.
- N. Carrasco, D. Ginsburg, Q. Du and Z. Huang, *Nucleosides, Nucleotides Nucleic Acids*, 2001, **20**, 1723–1734.
- A. Nuthanakanti, M. A. Boerneke, T. Hermann and S. G. Srivatsan, *Angew. Chem., Int. Ed.*, 2017, **56**, 2640–2644.
- C. Höbartner, R. Rieder, C. Kreutz, B. Puffer, K. Lang, A. Polonskaia, A. Serganov and R. Micura, *J. Am. Chem. Soc.*, 2005, **127**, 12035–12045.
- N. Tarashima, K. Hayashi, M. Terasaki, H. Taniike, Y. Inagaki, K. Hirose, K. Furukawa, A. Matsuda and N. Minakawa, *Org. Lett.*, 2014, **16**, 4710–4713.
- H. Taniike, Y. Inagaki, A. Matsuda and N. Minakawa, *Tetrahedron*, 2011, **67**, 7977–7982.



- 40 K. Ishii, N. Saito-Tarashima, M. Ota, S. Yamamoto, Y. Okamoto, Y. Tanaka and N. Minakawa, *Tetrahedron*, 2016, **72**, 6589–6594.
- 41 R. Micura, C. Höbartner, R. Rieder, C. Kreutz, B. Puffer, K. Lang and H. Moroder, *Curr. Protoc. Nucleic Acid Chem.*, 2006, **27**, 1.15.1–1.15.34.
- 42 J. Salon, J. Sheng, J. Gan and Z. Huang, *J. Org. Chem.*, 2010, **75**, 637–641.
- 43 J. Sheng, J. Jiang, J. Salon and Z. Huang, *Org. Lett.*, 2007, **9**, 749–752.
- 44 J. Sheng, J. Salon, J. Gan and Z. Huang, *Sci. China: Chem.*, 2010, **53**, 78–85.
- 45 H. Moroder, C. Kreutz, K. Lang, A. Serganov and R. Micura, *J. Am. Chem. Soc.*, 2006, **128**, 9909–9918.
- 46 R. A. Thompson, A. M. Spring, J. Sheng, Z. Huang and M. W. Germann, *J. Biomol. Struct. Dyn.*, 2015, **33**, 289–297.
- 47 L. A. Wozniak, M. Sochacki, H. Mitsuya, S. Kageyama and W. J. Stec, *Bioorg. Med. Chem. Lett.*, 1994, **4**, 1033–1036.
- 48 Y. Xu and E. T. Kool, *J. Am. Chem. Soc.*, 2000, **122**, 9040–9041.
- 49 O. Eguagie, P. F. Conlon and J. S. Vyle, *Tetrahedron Lett.*, 2016, **57**, 5000–5002.
- 50 P. Pistichimuka, *J. Prakt. Chem.*, 1911, **84**, 746–760.
- 51 G. Mielniczak and A. Lopusiński, *Heteroat. Chem.*, 2003, **14**, 121–127.
- 52 M. del C. Michelini, N. Russo, S. Alcaro and L. A. Wozniak, *Tetrahedron*, 2012, **68**, 5554–5563.
- 53 L. A. Wozniak, M. Gra, Z. Kaminski and W. J. Stec, *Phosphorus, Sulfur Silicon Relat. Elem.*, 2008, **183**, 1076–1081.
- 54 M. Kullberg and J. Stawinski, *Nucleosides, Nucleotides Nucleic Acids*, 2005, **24**, 659–661.
- 55 R. Cosstick and F. Eckstein, *Biochemistry*, 1985, **24**, 3630–3638.
- 56 A. Grandas, W. S. Marshall, J. Nielsen and M. H. Caruthers, *Tetrahedron Lett.*, 1989, **30**, 543–546.
- 57 R. Cosstick and J. S. Vyle, *Nucleic Acids Res.*, 1990, **18**, 829–835.
- 58 K. Misiura and W. J. Stec, *Bioorg. Med. Chem. Lett.*, 1994, **4**, 1037–1040.
- 59 A. M. Belostotskii, J. Lexner and A. Hassner, *Tetrahedron Lett.*, 1999, **40**, 1181–1184.
- 60 O. Eguagie, P. F. Conlon, F. Ravalico, J. S. T. Sweet, T. B. Elder, L. P. Conway, M. E. Lennon, D. R. W. Hodgson and J. S. Vyle, *Beilstein J. Org. Chem.*, 2017, **13**, 87–92.
- 61 O. Eguagie, J. S. Vyle, P. F. Conlon, M. A. Gilea and Y. Liang, *Beilstein J. Org. Chem.*, 2018, **14**, 955–970.
- 62 Y. Ju, D. Kumar and R. S. Varma, *J. Org. Chem.*, 2006, **71**, 6697–6700.
- 63 R. Suchsland, B. Appel, M. Janczyk and S. Müller, *Appl. Sci.*, 2019, **9**, 2199.
- 64 M. Hassler, Y. Q. Wu, N. Mallikarjuna Reddy, T. H. Chan and M. J. Damha, *Tetrahedron Lett.*, 2011, **52**, 2575–2578.
- 65 R. Suchsland, B. Appel and S. Müller, *Beilstein J. Org. Chem.*, 2018, **14**, 397–406.
- 66 J. S. Vyle, N. H. Williams and J. A. Grasby, *Tetrahedron Lett.*, 1998, **39**, 7975–7978.
- 67 C. Thorpe, S. Epple, B. Woods, A. H. El-Sagheer and T. Brown, *Org. Biomol. Chem.*, 2019, **17**, 5341–5348.
- 68 J. S. Hargreaves, R. Kaiser and P. K. Wolber, *Nucleosides, Nucleotides Nucleic Acids*, 2015, **34**, 691–707.
- 69 A. H. Krotz, C. Rentel, D. Gorman, P. Olsen, H. J. Gaus, J. V. McArdle and A. N. Scozzari, *Nucleosides, Nucleotides Nucleic Acids*, 2004, **23**, 767–775.
- 70 C. Glidewell and E. J. Leslie, *J. Chem. Soc., Dalton Trans.*, 1977, 527–531.
- 71 L. J. Yu, C. J. Wiederholt, J. N. Patro, K. Haraguchi and M. M. Greenberg, *J. Org. Chem.*, 2005, **70**, 141–149.
- 72 H. R. Drew, R. M. Wing, T. Takano, C. Broka, S. Tanaka, K. Itakura and R. E. Dickerson, *Proc. Natl. Acad. Sci. U. S. A.*, 1981, **78**, 2179–2183.
- 73 C. H. Kang, I. Berger, C. Lockshin, R. Ratliff, R. Moyzis and A. Rich, *Proc. Natl. Acad. Sci. U. S. A.*, 1994, **91**, 11636–11640.
- 74 H. A. Assi, M. Garavís, C. González and M. J. Damha, *Nucleic Acids Res.*, 2018, **46**, 8038–8056.
- 75 S. P. Adams, K. S. Kavka, E. J. Wykes, S. B. Holder and G. R. Galluppi, *J. Am. Chem. Soc.*, 1983, **105**, 661–663.
- 76 R. Zuckermann, D. Corey and P. Schultz, *Nucleic Acids Res.*, 1987, **15**, 5305–5321.
- 77 A. F. Cook, *J. Am. Chem. Soc.*, 1970, **92**, 190–195.
- 78 S. Chladek and J. Nagyvary, *J. Am. Chem. Soc.*, 1972, **94**, 2079–2085.
- 79 Y. Xu and E. T. Kool, *Nucleic Acids Res.*, 1998, **26**, 3159–3164.
- 80 S. Li, W. K. Olson and X.-J. Lu, *Nucleic Acids Res.*, 2019, **47**, W26–W34.
- 81 C. A. Frederick, G. J. Quigley, M. K. Teng, M. Coll, G. A. van der Marel, J. H. van Boom, A. Rich and A. H.-J. Wang, *Eur. J. Biochem.*, 1989, **181**, 295–307.
- 82 P. S. Pallan, T. P. Prakash, F. Li, R. L. Eoff, M. Manoharan and M. Egli, *Chem. Commun.*, 2009, 2017–2019.
- 83 J. Harp, P. Pallan and M. Egli, *Crystals*, 2016, **6**, 125.
- 84 J. M. Martin-Garcia, C. E. Conrad, J. Coe, S. Roy-Chowdhury and P. Fromme, *Arch. Biochem. Biophys.*, 2016, **602**, 32–47.
- 85 C. H. Yoon, H. DeMirci, R. G. Sierra, E. H. Dao, R. Ahmadi, F. Aksit, A. L. Aquila, A. Batyuk, H. Ciftci, S. Guillet, M. J. Hayes, B. Hayes, T. J. Lane, M. Liang, U. Lundström, J. E. Koglin, P. Mgbam, Y. Rao, T. Rendahl, E. Rodriguez, L. Zhang, S. Wakatsuki, S. Boutet, J. M. Holton and M. S. Hunter, *Sci. Data*, 2017, **4**, 170055.
- 86 R. L. Setten, J. J. Rossi and S. ping Han, *Nat. Rev. Drug Discovery*, 2019, **18**, 421–446.
- 87 R. Cosstick and J. Gaynor, *Curr. Org. Chem.*, 2008, **12**, 291–308.
- 88 Z. Veres, L. Tsai, T. D. Scholz, M. Politino, R. S. Balaban and T. C. Stadtman, *Proc. Natl. Acad. Sci. U. S. A.*, 1992, **89**, 2975–2979.

