Chemical Science

EDGE ARTICLE

Enzymatic synthesis of 2′-methylseleno-modified RNA†

Vanessa Siegmund,[a] **Tobias Santner,**[b] **Ronald Micura**[*b] **and Andreas Marx**[*a]

Selenium-derivatization of RNA is a powerful and advantageous alternative to conventional heavy atom derivatization techniques that are required for the phasing of X-ray crystallographic diffraction data. Among several possibilities, the 2′-methylseleno (2′-SeCH$_3$) modification has been most widely explored and was responsible for a series of important RNA structure determinations, such as the Diels–Alder ribozyme or complexes of antibiotics to HIV dimerization initiation site (DIS) RNA. So far, 2′-SeCH$_3$-RNA has only been accessible by chemical solid-phase synthesis for sizes of up to 50 nucleotides and up to about 100 nucleotides in combination with enzymatic ligation procedures. To overcome this limitation, here we present the enzymatic synthesis of 2′-SeCH$_3$-RNA to open up access for the preparation of long selenium-modified RNA sequences, which cannot be accomplished by conventional chemical synthesis. Therefore, we first elaborated a synthetic route towards the 2′-methylseleno-2′-deoxyribonucleoside triphosphates of cytosine and uridine (2′-SeCH$_3$–CTP and 2′-SeCH$_3$–UTP). With these crucial derivatives in hand, we found that mutants of T7 RNA polymerase are able to incorporate 2′-SeCH$_3$–CMP and 2′-SeCH$_3$–UMP into RNA, while the wild-type polymerase fails to do so. This study demonstrates the efficient enzymatic synthesis of 2′-SeCH$_3$-modified RNA and, thus, provides a thorough foundation for an alternative derivatization strategy in X-ray crystallographic structure analysis of larger RNAs. Such efforts are currently highly requested because of the steadily increasing number of novel non-coding RNAs whose structural features remain to be elucidated.

Introduction

Modified nucleic acids are substantial for the development of new oligonucleotide-based methods and technologies, since they possess altered properties in comparison to their natural counterparts. In addition to the conventional chemical synthesis of modified oligonucleotides on a solid support, the enzymatic synthesis of modified DNA and RNA by nucleic acid polymerases represents an alternative route to overcome the inherent limitations of automated chemical synthesis, such as oligonucleotide size restriction and degree of modification.

By using modified 2′-deoxynucleoside triphosphates (dNTPs), DNA polymerases can incorporate modifications into DNA in primer extension reactions and even amplify modified DNA using PCR. Modifications attached to the nucleobase at C5 of pyrimidines or C7 of 7-deazapurines are tolerated by many DNA polymerases.[1–9] Therefore, DNA polymerases are widely employed to introduce functionalities into DNA for a variety of applications. Enzymatic synthesis of modified RNA is often accomplished by *in vitro* transcription using T7 RNA polymerase in the presence of either modified ribonucleoside triphosphates (NTPs) or of so-called initiator nucleotides to site-specifically modify the 5′-end of the RNA.[10–14] When modified NTPs are used for the internal labeling of RNA, the modifications are attached to the nucleobase or at the phosphates.[15–18] Nevertheless, modifications at the 2′-position of the ribose are of great interest for further usage of the modified RNAs in siRNA technologies or as aptamers due to their increased stability and resistance towards ribonucleases, their enhanced pharmacokinetic properties and their ability to reduce immunostimulatory effects.[19–22] Since wild-type T7 RNA polymerase is not able to efficiently incorporate nucleotides with modifications at the 2′-position into RNA, mutant enzymes have been generated. These T7 RNA polymerase mutants can be used to efficiently incorporate nucleotides with modified 2′-groups, like 2′-H, 2′-F, 2′-NH$_2$, 2′-OMe and 2′-N$_3$, into RNA for further application.[23–26]

We have tested two of those mutants for their ability to incorporate 2′-methylseleno (2′-SeCH$_3$)-modified nucleotides into RNA. The 2′-SeCH$_3$ group offers the possibility for

[a] Department of Chemistry and Konstanz Research School Chemical Biology, University of Konstanz, Universitätsstraße 10, 78457 Konstanz, Germany. E-mail: andreas.marx@uni-konstanz.de; Fax: +49-7531-885140; Tel: +49-7531-885139

[b] Institute of Organic Chemistry and Center for Molecular Biosciences (CMBI), Leopold-Franzens University, Innrain 52a, 6020 Innsbruck, Austria. E-mail: ronald.micura@uibk.ac.at; Fax: +43-512-5072892; Tel: +43-512-5075210

† Electronic supplementary information (ESI) available: ^{1}H and ^{13}C-NMR spectra of compounds **1**–**7**, ^{31}P-NMR spectra of compounds **4** and **7**. SDS-PAGE gel of purified T7 RNA polymerases. See DOI: 10.1039/c1sc00404b

application in X-ray crystallography since selenium derivatization of RNA represents a powerful method for structure determination using the multiwavelength anomalous dispersion technique (MAD).[27–30] 2′-Methylseleno-modified RNA has been successfully used in several X-ray crystallographic studies and played a crucial role in the structure determination of the Diels–Alder ribozyme.[31–33] The availability of all four 2′-SeCH$_3$ nucleoside phosphoramidite building blocks permits great flexibility for adequate positioning within the RNA target during solid phase synthesis.[34–38] In addition, structural studies using the 2′-methylseleno-modified RNA target sequences showed that, for the majority of tested sequences, this type of modification did not affect the crystallization behavior or the native fold and was reported to even support the crystallization process.[37,39] Another advantage of 2′-SeCH$_3$-modified RNA is its increased resistance towards X-ray-induced radiation damage and it is thus an attractive alternative to commonly used 5-halogen-RNA.[39]

Currently, 2′-SeCH$_3$-modified RNA is accessible in sizes of up to 50 nucleotides by chemical synthesis and up to about 100 nucleotides in combination with enzymatic ligation techniques.[33,40] Since both of these preparation techniques are laborious and are only applicable to oligonucleotides of restricted sizes, they pose severe limitations to making this method viable in practice. For the large number of novel and highly relevant non-coding RNA sequences of 50–200 nucleotides that await structural characterization, an enzymatic approach would be of great interest to overcome these drawbacks. However, until now, no RNA polymerase has been reported that is able to incorporate 2′-methylseleno-modified nucleotides. Motivated to overcome the limitations of chemical RNA synthesis on solid supports, our efforts focused on the challenge to develop an enzymatic approach for the preparation of 2′-SeCH$_3$-modified RNA (Fig. 1a).

In short, we synthesized the 2′-methylseleno ribonucleoside triphosphates of cytidine and uridine (2′- SeCH$_3$–CTP and 2′-SeCH$_3$–UTP (Fig. 1b)). Subsequently, we tested wild-type T7 RNA polymerase, as well as two mutants, for their acceptance of the 2′-methylseleno-modified nucleotides in *in vitro* transcription experiments. Those mutants, M1 (Y639F, H784A) and M2 (Y639V, H784G, E593G, V685A), have been described for their enhanced ability to incorporate 2′-OCH$_3$- and 2′-N$_3$-modified NTPs.[25,26] Here, we found that the wild-type T7 RNA polymerase is not able to incorporate the modified nucleotides into RNA during *in vitro* transcription, while both of the mutants accept 2′-SeCH$_3$–CTP and 2′-SeCH$_3$–UTP as substrates and efficiently incorporate them into RNA. Therefore, this study

creates the foundation for 2′-SeCH$_3$ derivatization of large-size RNA, which is of high potential for crystallographic X-ray structure analysis.

Materials and methods

Synthesis of 2′-SeCH$_3$ pyrimidine nucleoside triphosphates 4 and 7

General. ^{1}H and ^{13}C NMR spectra were recorded on a Bruker DRX 300 MHz or Bruker UltraShield™ Plus 600 MHz instrument. The chemical shifts are reported relative to TMS and referenced to the residual proton signal of the deuterated solvent. ^{1}H and ^{13}C assignments were based on COSY and HSQC experiments. Mass spectrometry (MS) experiments were performed on a Finnigan LCQ Advantage MAX ion trap instrumentation (Thermo Fisher Scientific); samples were analyzed in the positive- or negative-ion mode. Reaction control was performed *via* analytical thin-layer chromatography (TLC, Macherey-Nagel) on silica plates with a fluorescent indicator. Flash column chromatography was carried out on silica gel 60 (70–230 mesh). Chemical reagents and solvents were purchased from commercial suppliers (Sigma-Aldrich, Acros) and used without further purification. Organic solvents for reactions were dried overnight over freshly activated molecular sieves (4 Å). All reactions were carried out under an argon atmosphere. The triphosphates were purified using ion-exchange chromatography (DEAE Sephadex A-25; 0.05 M - 1 M triethylammonium bicarbonate (TEAB) buffer and reversed-phase chromatography (Merck Lobar LiChroprep RP-18; 0–40% 50 mM triethylammonium acetate (TEAA) buffer – acetonitrile) using a GE Healthcare Äkta Prime system. The yields of the triphosphate products were determined by UV spectroscopy (IMPLEN Nanophotometer) using the following extinction coefficients: ε (CTP, 260 nm) 7400; ε (UTP, 260 nm) 9900.

3′-Acetyl-5′-O-(4,4′-dimethoxytrityl)-2′-methylseleno-2′-deoxyuridine (2). Compound 1[41–43] (600 mg; 0.96 mmol) was coevaporated twice with anhydrous pyridine and then dissolved in pyridine (3 mL) and acetic anhydride (3 mL; 31.7 mmol). The reaction mixture was stirred at room temperature for 6 h and evaporated to dryness under reduced pressure. Compound 2 was purified by column chromatography on SiO$_2$ with ethylacetate/hexane, 4/6. Yield: 440 mg of 2 as a slightly yellow foam (69%). TLC (ethylacetate/hexane, 1/1): R_f 0.42. ^{1}H-NMR (300 MHz, CDCl$_3$): δ 2.08 (s, 3H, SeCH$_3$); 2.12 (s, 3H, COCH$_3$); 3.52 (m, 2H, H1-C(5′), H2-C(5′)); 3.60 (m, 1H, H–C(2′)); 3.79 (s, 6H, 2 × OCH$_3$); 4.15 (m, 1H, H–C(4′)); 5.33 (d, J = 8.2 Hz, H–C(5)); 5.47 (dd, J = 2.7 Hz, J = 5.7 Hz, 1H, H–C(3′)); 6.30 (d, J = 7.9 Hz, 1H, H–C(1′)); 6,84 (m, 4H, C (ar)); 7.29 (m, 9H, H–C(ar)); 7,72 (d, J = 8.2 Hz, 1H, H–C(6)) ppm. ^{13}C-NMR (150 MHz, CDCl$_3$): δ 4.82 (SeCH$_3$); 20.81 (COCH$_3$); 45.10 (C(2′)); 55.28 (2 × OCH$_3$ (DMT)); 63.04 (C(5′)); 74.89 (C (3′)); 83.21; 88.63 (C(1′)); 102.96 (C(5); 113.39 (C(ar)); 127.34 (C (ar)); 128.09 (C(ar)); 130.10 (C(ar)); 139.63; 144.12 (C(6)); 150.46; 158.63; 158.86; 162.8; 170.04 ppm. ESI-MS (m/z): [M + Na]$^+$ calcd for C$_{33}$H$_{34}$N$_2$O$_8$SeNa, 689.14; found 689.35.

3′-Acetyl-2′-methylseleno-2′-deoxyuridine (3). Compound 2 (411 mg; 0.56 mmol) was dissolved in CH$_2$Cl$_2$ (24 mL) and
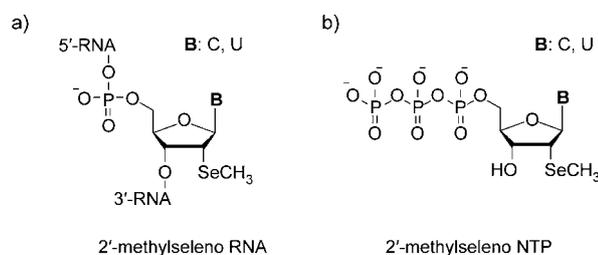

Fig. 1 a) The chemical structure of 2′-methylseleno RNA, b) the chemical structure of the synthesized 2′-methylseleno–CTP and 2′-methylseleno–UTP.

methanol (475 µl) was added. Then, a 30% solution of CF_3COOH in CH_2Cl_2 (950 µl) was injected in and the reaction progress was monitored by TLC. After 90 min, the reaction was finished and the reaction mixture was washed with an aqueous saturated $NaHCO_3$ solution, dried over Na_2SO_4 and evaporated under reduced pressure. Compound **3** was purified by column chromatography on SiO_2 with ethylacetate/hexane, 7/3. Yield: 146 mg of **3** as a white solid (72%). TLC (ethylacetate/hexane = 1/1): R_f 0.21. ^{1}H-NMR (300 MHz, DMSO-d_6): δ 1.95 (s, 3H, SeCH_3); 2.10 (s, 3H, COCH_3); 3.64 (m, 3H, H1-C(5′), H2-C(5′) and H–C(2′)); 4.03 (m, 1H, HO–C(5′)); 5.27 (m, 1H, H–C(3′)); 5.35 (dd, J = 1.4 Hz, J = 5.6 Hz, 1H, H–C(4′)); 5.76 (dd, J = 7.9 Hz, J = 1.6 Hz, 1H, H–C(5)); 6.10 (d, J = 9.3 Hz, 1H, H–C(1′)); 7.87 (d, J = 8.1, 1H C(6)); 11.44 (s, 1H, H–N(3)) ppm. ^{13}C-NMR (150 MHz, DMSO-d_6): δ 4.28 (SeCH_3); 21.10 (COCH_3); 40.55 (C(2′)); 61.67 (C(5′)); 75.57 (C(3′)); 87.91 (C(1′)); 103.24 (C(5); 140.41 (C(6)); 151.17; 163.27; 169.97 ppm. ESI-MS (m/z): [M + Na]$^+$ calcd for $C_{12}H_{16}N_2O_6SeNa$, 387.01; found 387.17.

2′-Methylseleno-2′-deoxyuridine 5′-triphosphate (4). Compound **3** (42 mg; 115 µmol) was coevaporated three times with anhydrous pyridine (2 mL) and dried over P_2O_5 in a desiccator for four hours. The residue was dissolved in dioxane (345 µL) and pyridine (115 µL). Then, a freshly prepared 1 M solution of 2-chloro-4H-1,2,3-dioxaphosphorin-4-one in dioxane (150 µL; 150 µmol) was added. After 10 min, a 0.5 M solution of tris(tetrabutylammonium) hydrogenpyrophosphate in DMF (370 µL; 185 µmol) was injected and the suspension was stirred for another 10 min. Then, a 1% solution of I_2 in pyridine/H_2O (98/2) (2.25 mL; 180 µmol) was injected in and quenched after 15 min by the addition of a few drops of 5% aqueous $NaHSO_3$. The reaction mixture was evaporated under reduced pressure and the residue dried under high vacuum. The residue was dissolved in H_2O (10 mL) and stirred for 30 min. Subsequently, saturated NH_4OH solution (20 mL) was added and, after one hour, the solution was evaporated and the residue dried. Compound **4** was purified using ion-exchange chromatography and reversed-phase chromatography. Yield after lyophilization: 48.3 µmol of **4** as a slightly yellow foam (42%). ^{1}H-NMR (300 MHz, D_2O): δ 1.19 (t, J = 6.1 Hz, NCH$_2$CH_3); 2.07 (s, 3H, SeCH_3); 3.11 (q, J = 6.1 Hz, NCH_2CH$_3$); 3.73 (dd, J = 8.6 Hz, J = 14.2 Hz, 1H, H–C(2′)); 4.31 (m, 3H, H–C(4′), H1-C(5′), H2-C(5′)); 4.66 (m, 1H, H–C(3′)); 6.05 (d, J = 8.1 Hz, 1H, H–C(5)); 6.34 (d, J = 8.8 Hz, 1H, H–C(1′)); 8.01 (d, J = 8.1 Hz, 1H, H–C(6)) ppm. ^{13}C-NMR (150 MHz, CDCl$_3$): δ 3.22 (SeCH_3); 8.28 (NCH$_2$CH$_3$); 46.33 (C(2′)); 46.72 (NCH_2CH$_3$); 65.58 (C(5′)); 72.35 (C(3′)); 89.10 (C(1′)); 103.41 (C(5); 141.52 (C(6)); 151.94; 165.95; 180.21 ppm. ^{31}P-NMR (121 MHz, D_2O): δ −10.98 (d, J = 19.6 Hz, 1P, Pγ); −10.48 (d, J = 20.0 Hz, 1P, Pα); −22.84 (triplettoid, J = 19.8 Hz, J = 20.0 Hz, 1P, Pβ) ppm. ESI-MS (m/z): [M − H]$^-$ calcd for $C_{10}H_{16}N_2O_{14}P_3Se$, 560.90; found 561.17.

N^4-Acetyl-3′-*tert*.-butyldimethylsilyl-2′-methylseleno-2′-deoxycytidine (6). Compound **5**[41,42] (400 mg; 0.52 mmol) was dissolved in a 3% solution of Cl_2HCCOOH in ClCH$_2$CH$_2$Cl (10 mL; 3.7 µmol). The reaction progress was monitored by TLC and after 15 min the reaction was finished. Then, the mixture was washed with aqueous saturated $NaHCO_3$ and evaporated under reduced pressure. Compound **6** was purified by column chromatography on SiO_2 with CH_2Cl_2/CH_3OH, 100/0 to 96/4. Yield: 198 mg of **6** as a white solid (81%). TLC (CH_2Cl_2/CH_3OH = 9/1): R_f 0.42. ^{1}H-NMR (300 MHz, CDCl$_3$): δ 0.10 (s, 3H, (CH$_3$)$_3$CSi(CH_3)$_2$); 0.15 (s, 3H, (CH$_3$)$_3$CSi(CH_3)$_2$); 0.92 (s, 9H, (CH_3)$_3$CSi(CH$_3$)$_2$); 1.83 (s, 3H, SeCH_3); 2.27 (s, 3H, CH$_3$CO); 3.71 (m, 1H, H1-C(5′)); 3.93 (m, 1H, H2-C(5′)); 4.13 (m, 3H, HO–C(5′), H–C(3′), H–C(2′)); 4.53 (dd, J = 1.9 Hz, J = 5.3 Hz, 1H, H–C(4′)); 5.78 (d, J = 7.9 Hz, 1H, H–C(1′)); 7.48 (d, J = 7.4 Hz, 1H, H–C(5)); 7.91 (d, J = 7.5 Hz, 1H, H–C(6)); 9.62 (s, 1H, NH) ppm. ^{13}C-NMR (150 MHz, DMSO-d_6): δ −4.81 ((CH$_3$)$_3$CSi(CH$_3$)$_2$); −4.61 ((CH$_3$)$_3$CSi(CH$_3$)$_2$); 1.01; 3.35 (SeCH_3); 18.15; 24.95 (CH$_3$CO); 25.75 ((CH$_3$)$_3$CSi(CH$_3$)$_2$); 46.23 (C(2′)); 62.33 (C(5′)); 74.45 (C(4′)); 88.65 (C(3′)); 97.30 (C(5)); 100.00 (C(1′)); 148.32 (C(6)); 155.27; 163.21; 170.99 ppm. ESI-MS (m/z): [M + Na]$^+$ calcd for $C_{18}H_{31}N_3O_5SeSiNa$, 500.11; found 500.27.

2′-Methylseleno-2′-deoxycytidine 5′-triphosphate (7). Compound **6** (45 mg; 94 µmol) was coevaporated three times with anhydrous pyridine (2 mL) and dried over P_2O_5 in an desiccator for four hours. The residue was dissolved in dioxane (750 µL) and pyridine (195 µL). Then, a freshly prepared 1 M solution of 2-chloro-4H-1,2,3-dioxaphosphorin-4-on in dioxane (122 µL; 122 µmol) was added. After 10 min, a 0.5 M solution of tris(tetrabutylammonium) hydrogenpyrophosphate in DMF (300 µL; 150 µmol) was injected in and the suspension was stirred for another ten minutes. Subsequently, a 1% solution of I_2 in pyridine/H_2O (98/2) (1.86 mL; 149 µmol) was injected and quenched after 15 min by the addition of a few drops of 5% aqueous $NaHSO_3$. The reaction mixture was evaporated under reduced pressure and the residue dried under high vacuum. The residue was dissolved in H_2O (10 mL) and stirred for 30 min. Then, saturated NH_4OH solution (20 mL) was added and after one hour, the solution was evaporated and the residue dried. Subsequently, it was dissolved in 1 M TBAF in THF (1 mL; 1 mmol) and stirred for one hour. The solution was evaporated under reduced pressure and dried under high vacuum. Compound **7** was purified using ion-exchange chromatography and reversed-phase chromatography. Yield after lyophilization: 10.9 mg (=19.4 µmol) of **7** as a white foam (21%). ^{1}H-NMR: (300 MHz, D_2O): δ 1.24 (t, J = 6.1 Hz, NCH$_2$CH_3); 1.89; 1.93 (s, 3H, SeCH_3); 3.15 (q, J = 6.1 Hz, NCH_2CH$_3$); 3.60 (dd, J = 8.4 Hz, J = 5.6 Hz, 1H, H-C(2′)); 4.19 (m, 2H, H1-C(5′), H2-C(5′)); 4.24 (m, 1H, H–C(4′)); 4.62 (m, 1H, H–C(3′)); 6.16 (d, J = 7.5 Hz, 1H, H–C(5)); 6.35 (d, J = 8.6 Hz, 1H, H–C(1′)); 7.93 (d, J = 7.6 Hz, 1H, H–C(6)) ppm. ^{13}C-NMR (150 MHz, CDCl$_3$): δ 3.12 (SeCH_3); 8.35 (NCH$_2$CH$_3$); 46.74 (NCH_2CH$_3$); 46.82 (C(2′)); 65.69 (C(5′)); 72.47 (C(3′)); 84.98 (C(4′)); 89.4 (C(1′)); 97.30 (C(5); 141.79 (C(6)); 153.69; 157.19; 165.71; 180.21 ppm. ^{31}P-NMR: (121 MHz, D_2O) δ −10.98 (d, J = 19.6 Hz, 1P, Pγ); −10.48 (d, J = 20.0 Hz, 1P, Pα); −22.84 (triplettoid, J = 19.8 Hz, J = 20.0 Hz, 1P, Pβ) ppm. ESI-MS (m/z): [M − H]$^-$ calcd for $C_{10}H_{16}N_2O_{14}P_3Se$, 559.91; found 560.07.

Enzyme preparation and *in vitro* transcription assays

Generation of T7 RNA polymerase mutants M1 and M2. The *E. coli* codon optimized gene for T7 RNA polymerase wild-type was ordered from Geneart. The gene was cloned into the expression vector pGDR11,[44] which is a derivative of the pQE31

vector (Qiagen) carrying the *lacI$_q$* gene, using *Hind*III and *Sph*I restriction endonucleases (Fermentas). Introduction of the mutations Y639F and H784A for Mutant 1 (M1) was performed according to the QuickChange Site-Directed Mutagenesis protocol (Stratagene) using Phusion High-Fidelity DNA polymerase (Finnzymes). Mutations Y639V, H784G, E593G and V685A for Mutant 2 (M2) were introduced using the Quick-Change Multi Site-Directed Mutagenesis Kit (Stratagene). Sequences were confirmed after each step of mutagenesis by sequencing at GATC Biotech. All of the oligonucleotide primers were purchased from Metabion.

Expression and purification of T7 RNA polymerase wild-type and mutants M1 and M2. Plasmids carrying the corresponding T7 RNA polymerase gene were transformed into *E. coli* BL21 (DE3) cells (Stratagene). Expression was performed in 1 l LB-medium (Lennox, LB broth) supplemented with 100 μg mL^{-1} carbenicillin. Cells were grown at 37 °C to an OD$_{600}$ of 0.6–0.7 and expression was induced by adding IPTG to a final concentration of 1 mM. After 5 h of expression, cells were harvested by centrifugation (4.500 × g for 20 min, 4 °C) and pellets stored at −80 °C until further use. Purification of the polymerases bearing an N-terminal polyhistidine (6 × His) tag was performed by nickel affinity chromatography using Ni-IDA chelating sepharose fast flow on an ÄKTA Purifier FPLC system (GE Healthcare). Therefore, the cell pellets were resuspended in 50 mL lysis buffer (50 mM Tris-HCl pH 8.0, 300 mM NaCl, 0.1% (v/v) Triton X-100, 0.1 mg ml^{-1} lysozyme, 1 mM Benzamidine, 1 mM PMSF) and incubated on ice for 30 min. Cells were disrupted by sonication and lysates cleared by centrifugation (22.000 × g for 30 min, 4 °C). The supernatant was loaded on a Ni-IDA column equilibrated with buffer A (50 mM Tris-HCl pH 8.0, 300 mM NaCl) and proteins eluted by gradient-elution with buffer B (50 mM Tris-HCl pH 8.0, 300 mM NaCl, 250 mM imidazol). Fractions containing the purified protein were pooled and the buffer exchanged with 2 × storage buffer (100 mM Tris-HCl pH 8.0, 200 mM NaCl, 10 mM 1,4-dithiothreitol (DTT), 2 mM EDTA, 0.2% (v/v) Triton X-100) using a Vivaspin 50.000 MWCO PES concentrator (Sartorius Stedim Biotech). Glycerol was added to a final concentration of 50% (v/v) and the proteins were stored at −20 °C. Purified enzyme concentrations were determined *via* a Bradford protein assay (Roth). The proteins were adjusted to the same concentration and their qualities verified by SDS-PAGE (12%) stained with colloidal Coomassie blue (Roth).

In vitro transcription assays. DNA templates (Metabion or Biomers) containing the T7 promoter sequence were prepared by PCR using Phusion High-Fidelity DNA polymerase (Finnzymes) and purified by phenol–chloroform extraction and ethanol precipitation. Transcription reactions using NTPs (Fermentas) and modified 2′-methylseleno NTPs were performed for 2 h at 37 °C in 40 mM Tris-HCl pH 8.0, 2 mM spermidine, 10 mM 1,4-dithiothreitol (DTT) and 6 mM MgCl$_2$, with templates at 0.5 μM and RNA polymerases at 0.5 μM final concentration. Nucleotides were used at 1 mM final concentration in a 20 μL reaction volume and transcripts were labeled by the inclusion of 1% (v/v) 3000 Ci mmol^{-1} [α-^{32}P]GTP (Hartmann Analytic). Reactions were stopped by adding an equal volume of 90% (v/v) formamide, 50 mM EDTA, 0.01% (w/v) bromophenol blue and 0.01% (w/v) xylene cyanol and the transcripts were resolved by electrophoresis on 12% denaturing polyacrylamide gels. Visualization was performed by phosphorimaging using a Molecular Imager FX (Bio Rad).

Time course experiments. Transcription reactions were set up as described above. Aliquots (2 μL) were withdrawn from the reaction mixture at the corresponding time points and added to an equal volume of 90% (v/v) formamide, 50 mM EDTA, 0.01% (w/v) bromophenol blue and 0.01% (w/v) xylene cyanol. The transcripts were resolved by electrophoresis on 12% denaturing polyacrylamide gels, visualized by phosphorimaging and quantified using QuantityOne (Bio Rad).

RNA quantification. For RNA quantification, the transcription reactions, as described above, were scaled up to 400 μL. After 2 h at 37 °C, the reactions were stopped by the addition of 50 units DNase I (Fermentas, RNase-free) and the DNA templates were digested for 15 min at 37 °C. The RNA transcripts were precipitated with EtOH at −20 °C overnight. The precipitate was dissolved in equal volumes of ddH$_2$O and 90% (v/v) formamide, 50 mM EDTA, 0.01% (w/v) bromophenol blue and 0.01% (w/v) xylene cyanol and purified by electrophoresis on an 8% denaturing polyacrylamide gel. The excised bands were crushed and the RNA eluted into ddH$_2$O at 4 °C overnight. After EtOH precipitation, the RNA was dissolved in ddH$_2$O and the concentrations were measured using a BioPhotometer at 260 nm (Eppendorf).

Results and discussion

Synthesis

To investigate the enzymatic synthesis of 2′-methylseleno RNA, we first synthesized and characterized the nucleoside triphosphate of 2′-SeCH$_3$ uridine. The functionalized derivative **1** (Scheme 1) represented the starting material, which was readily obtained along the lines of a previously introduced synthetic route.[41] Subsequent treatment with acetic anhydride resulted in quantitative acetylation of the 3′-hydroxyl group to yield compound **2**, followed by selective cleavage of the dimethoxytrityl group under acidic conditions to give derivative **3**. Introduction of the triphosphate moiety at the free 5′-hydroxyl group of **3** was achieved according to the Ludwig and Eckstein approach,[45] using 2-chloro-4*H*-1,2,3-benzodioxaphosphorin-4-one, pyrophosphate and subsequent iodine oxidation. Finally, cleavage of the 3′-*O*-acetyl furnished the desired 2′-SeCH$_3$ uridine derivative **4**.

Likewise, the synthesis of the 2′-SeCH$_3$ cytidine triphosphate **7** started from the readily available 3′-*O*-silylated compound **5** (Scheme 1).[41] The selective cleavage of the dimethoxytrityl group under acidic conditions yielded derivative **6** and this was followed by the introduction of the triphosphate moiety at the free 5′-hydroxyl group of **6**, according to the Ludwig and Eckstein approach.[45] Finally, cleavage of the 3′-*O*-TBDMS furnished the desired 2′-SeCH$_3$ cytidine derivative **7**.

Enzymes

Next, we wanted to test whether T7 RNA polymerase, an enzyme frequently used in *in vitro* transcription for the synthesis of RNA,
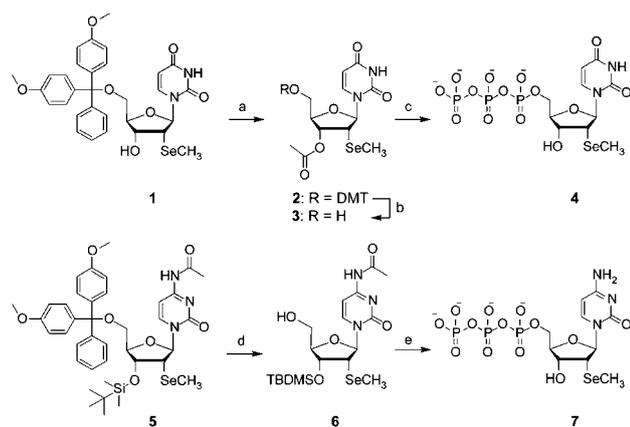
Scheme 1 The synthesis of 2′-methylseleno uridine triphosphate **4** and 2′-methylseleno cytidine triphosphate **7**. Reaction conditions: (a) 33 equiv. acetic anhydride in pyridine at room temperature for 10 h, 69%; (b) 1.2% CF_3COOH in CH_2Cl_2 at room temperature for 15 min, 72%; (c) (i): 1.3 equiv. 2-chloro-4H-1,2,3-benzodioxaphosphorin-4-one, in dioxane/pyridine, room temperature, 10 min, (ii): 1.6 equiv. $(HNBu_3)_2H_2P_2O_7$ in DMF and NBu_3 at room temperature for 10 min, (iii): 1.2 equiv. I_2 in pyridine/H_2O at room temperature for 15 min, (iv): 5% (v/v) aqueous $NaHSO_3$ solution, room temperature, 5 min, (v): 28% NH_4OH solution at room temperature for 1 h, 42%; (d) 3% Cl_2HCOOH in $ClCH_2CH_2Cl$ at room temperature for 15 min, 81%; (e) (i): 1.3 equiv 2-chloro-4H-1,2,3-benzodioxaphosphorin-4-one in dioxane/pyridine at room temperature for 10 min, (ii): 1.6 equiv $(HNBu_3)_2H_2P_2O_7$ in DMF and NBu_3 at room temperature for 10 min, (iii): 1.2 equiv I_2 in pyridine/H_2O at room temperature for 15 min, (iv): 5% (v/v) aqueous $NaHSO_3$ solution at room temperature for 5 min, (v): 28% NH_4OH solution at room temperature for 1 h, 21%.

is able to accept the modified substrates. We generated T7 RNA polymerase wild-type, as well as two mutants, M1 (Y639F, H784A) and M2 (Y639V, H784G, E593G, V685A), that are known for their increased propensity to tolerate 2′-modifications.[25,26] The enzymes were heterologously expressed in *E. coli* containing an N-terminal 6 × His-tag and purified by affinity chromatography. Interestingly, both mutants have mutations at amino acid positions Y639 and H784. The amino acid position Y639 has been implicated before in substrate recognition and an interaction with the 2′-position of incoming ribonucleotides.[46,47] Both amino acid residues are located close to each other in the active site of a transcribing T7 RNA polymerase (Fig. 2).

In vitro transcription

For *in vitro* transcription using T7 RNA polymerase, a double stranded DNA template (112 nucleotides) containing the T7 promoter sequence was designed to allow the incorporation of 12 C or U residues into the RNA transcript (89 nucleotides). The RNA transcripts were labeled by the addition of [α-^{32}P]GTP to the enzymatic reactions and the transcripts were visualized using a phosphorimager after separation by denaturating polyacrylamide gel electrophoresis (PAGE).

We first tested whether the wild-type T7 RNA polymerase was able to incorporate 2′-SeCH$_3$–CMP into an 89mer RNA transcript containing 12 C residues (Fig. 3a). PAGE analysis demonstrated that the full-length transcript could not be
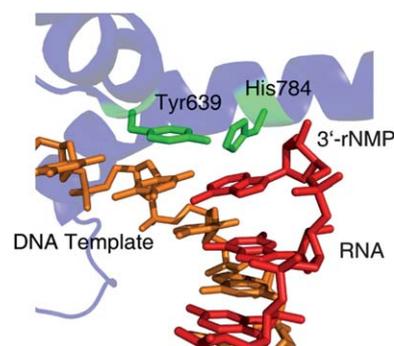

Fig. 2 The location of Tyr639 and His784 as seen in the crystal structure of a transcribing T7 RNA polymerase (PDB 1MSW). The DNA template is shown in orange, RNA transcript in red, interesting amino acid residues in green and the RNA polymerase in blue.

detected either when CTP was replaced by its 2′-methylseleno-modified analog (Fig. 3b, lane 2, WT) nor in the control experiment, where CTP was absent (Fig. 3b, lane 3, WT). We then tested the ability of the mutants M1 and M2 to incorporate 2′-SeCH$_3$–CMP into RNA. The results indicate that both mutants
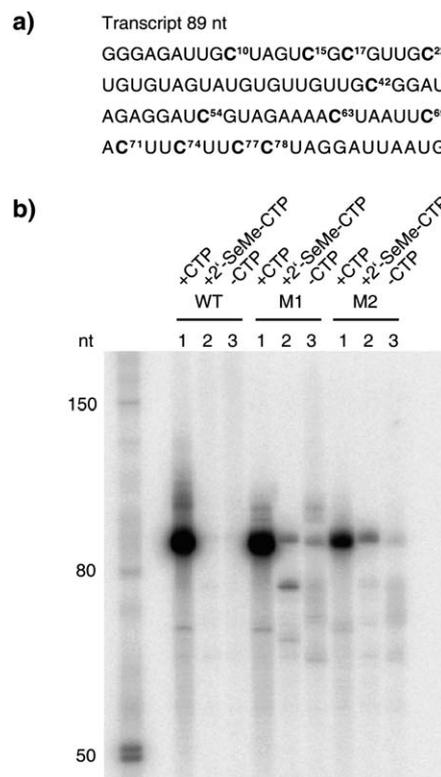

Fig. 3 The enzymatic incorporation of 2′-SeCH$_3$–CMP into RNA. a) The sequence of the RNA transcript (89 nt). Positions where CMP or 2′-SeCH$_3$–CMP were incorporated into RNA are highlighted. b) The PAGE analysis of the transcription reactions comparing T7 RNA polymerase wild-type (WT) with mutant M1 and mutant M2 after 2 h of incubation at 37 °C. In each set, lane 1 shows the reaction with all four NTPs, lane 2 shows the reaction where CTP was replaced by 2′-SeCH$_3$–CTP and lane 3 shows the control reaction without CTP but with ATP, GTP and UTP.

are using 2′-SeCH$_3$–CTP as a substrate and the full-length transcript is formed (Fig. 3b, lane 2, M1 and M2). In case of mutant M1, a small amount of the full-length transcript was also formed in the control reaction without any CTP, leading to the assumption that this mutant is less accurate on the provided DNA template and misincorporation is taking place (Fig. 3b, lane 3, M1). This was also observed for mutant M2, but in insignificantly low amounts (Fig. 3b, lane 3, M2). With respect to the applications for X-ray crystallography, the resulting heterogeneous mixture of transcripts could lead to difficulties during the crystallization process.

In case of mutant M1, the polymerase discontinues transcription when incorporating two 2′-SeCH$_3$–CMP moieties consecutively at positions 77 and 78 to form a shorter, truncated transcript (Fig. 3b, lane 2, M1). The positive control with all four NTPs demonstrated that the three tested enzymes are highly active when their natural substrates are provided (Fig. 3b, lane 1).

We performed a time-course experiment to gain insights into the incorporation efficiencies of 2′-SeCH$_3$–CMP into RNA, comparing T7 RNA polymerase wild-type with the mutants M1 and M2. The PAGE analysis clearly indicates that 2′-SeCH$_3$-modified RNA is formed in case of the mutants M1 and M2, but not when using the wild-type polymerase (Fig. 4a). The quantification of the full-length transcript shows that mutant M2 is approximately twice as efficient as mutant M1 when incorporating 2′-SeCH$_3$–CMP (Fig. 4b).

Next, we determined whether 2′-SeCH$_3$–UTP is a substrate for T7 RNA polymerase. Therefore, a DNA template calling for an 89mer transcript, where all of the C residues from our previous experiments were exchanged by U and *vice versa*, was used (Fig. 5a).

We performed transcription reactions where UTP was replaced by its 2′-SeCH$_3$ counterpart. In the case of the wild-type T7 RNA polymerase, no full-length transcript could be detected after PAGE analysis (Fig. 5b, lane 2, WT). For the mutants, M1 and M2, the PAGE analysis indicates that 2′-SeCH$_3$–UMP was efficiently incorporated (Fig. 5b, lane 2, M1 and M2). No full-length transcript was detected in the control reactions without
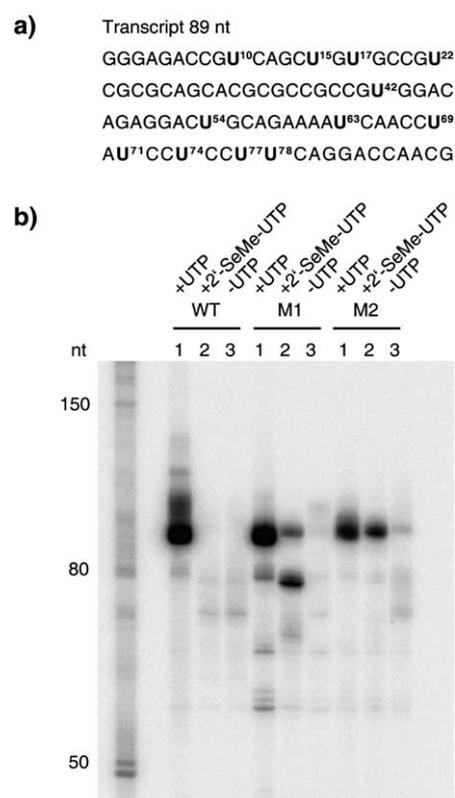

Fig. 5 The enzymatic incorporation of 2′-SeCH$_3$–UMP into RNA. a) The sequence of the RNA transcript (89 nt). Positions where UMP or 2′-SeCH$_3$–UMP were incorporated into RNA are highlighted. b) PAGE analysis of the transcription reactions, comparing T7 RNA polymerase wild-type (WT) with mutant M1 and mutant M2 after 2 h of incubation at 37 °C. In each set, lane 1 shows the reaction with all four NTPs, lane 2 shows the reaction where UTP was replaced by 2′-SeCH$_3$–UTP and lane 3 shows the control reaction without UTP but with ATP, GTP and CTP.

any UTP (Fig. 5b, lane 3) leading to the assumption that the formation of the full-length transcript in the reactions where 2′-SeCH$_3$–UTP was added was due to the incorporation of
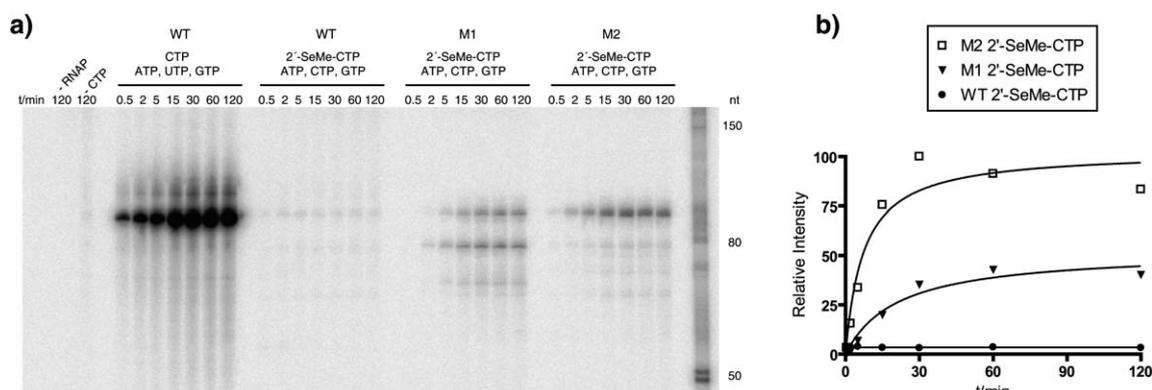

Fig. 4 A time-course experiment for the incorporation of 2′-SeCH$_3$–CMP into RNA. a) PAGE analysis of the transcription reactions using T7 RNA polymerase wild-type (WT), mutant M1 and mutant M2. Aliquots of the reaction mixtures were withdrawn at the corresponding time points and quenched with an equal volume of stop solution containing 50 mM EDTA. Control reactions were done with no polymerase (−RNAP) and with no CTP but with ATP, GTP and UTP (−CTP). b) Quantification of the time-course experiment for the formation of RNA in the presence of 2′-SeCH$_3$–CTP.
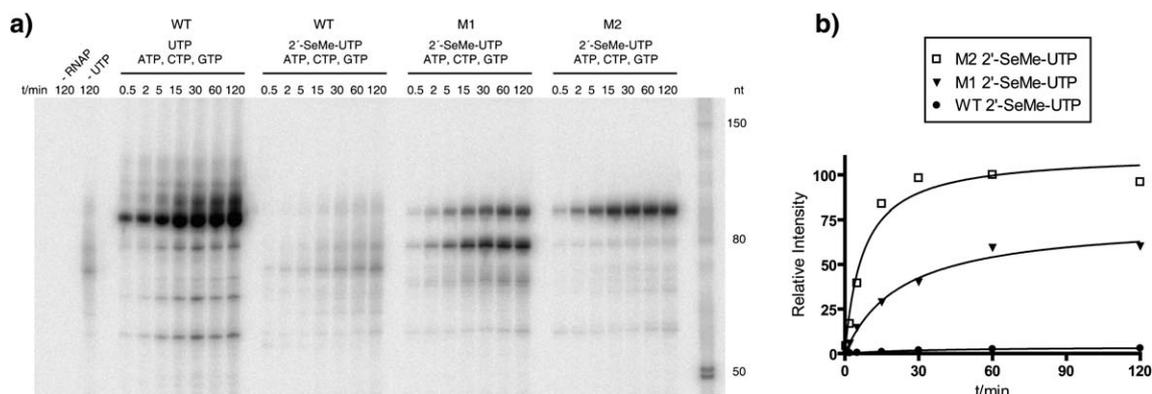
Fig. 6 A time-course experiment for the incorporation of 2′-SeCH$_3$–UMP into RNA. a) PAGE analysis of the transcription reactions using T7 RNA polymerase wild-type (WT), mutant M1 and mutant M2. Aliquots of the reaction mixtures were withdrawn at the corresponding time points and quenched with an equal volume of stop solution containing 50 mM EDTA. Control reactions were done with no polymerase (−RNAP) and with no UTP but with ATP, GTP and CTP (−UTP). b) Quantification of the time-course experiment for the formation of RNA in the presence of 2′-SeCH$_3$–UTP.

2′-SeCH$_3$–UMP into RNA. As already seen in our previous experiments, after incorporating two 2′-SeCH$_3$–UMP moieties consecutively at positions 77 and 78, mutant M1 was stalling, leading to a shorter transcript (Fig. 5b, lane 2, M1). The time-course experiment demonstrated that 2′-SeCH$_3$–UMP was efficiently incorporated into RNA by mutants M1 and M2, but not when using the wild-type T7 RNA polymerase (Fig. 6a). The quantification of the full-length transcript showed that mutant M2 is most efficiently incorporating 2′-SeCH$_3$–UMP (Fig. 6b).

To determine the quantity of total RNA that can be synthesized by the T7 RNA polymerase mutant M2 in the presence of either 2′-SeCH$_3$–UTP or UTP, we performed large scale transcription assays under the same conditions as used before and PAGE-purified the RNA transcripts. This step was necessary to remove truncated RNA molecules from the reaction mixtures that would have influenced the quantification of the desired full-length transcripts. With this procedure, it was possible to recover from a 400 µL transcription assay 90 µg (3.12 nmol) of unmodified RNA synthesized by wild-type T7 RNA polymerase, 35.3 µg (1.22 mol) of unmodified RNA synthesized by T7 RNA polymerase mutant M2 and 19.5 µg (0.65 nmol) of 2′-SeCH$_3$-modified RNA synthesized by M2, respectively. These results are consistent with the ratios observed in the time-course experiment, where the amount of unmodified RNA generated with the wild-type enzyme was determined to be approximately six-fold higher than the amount of 2′-SeCH$_3$-modified RNA synthesized by the mutant M2.

Conclusions

In summary, we have demonstrated the enzymatic synthesis of 2′-methylseleno-modified RNA by using the T7 RNA polymerase mutants M1 (Y639F, H784A) and M2 (Y639V, H784G, E593G, V685A) known from the literature.[25,26] The experiments indicate that 2′-SeCH$_3$–CMP and 2′-SeCH$_3$–UMP were successfully incorporated into RNA. The best results were obtained for the incorporation of 2′-SeCH$_3$-modified U using mutant M2. The resulting RNA is highly selenium-modified, since all 12 C or U moieties within the transcript were replaced by

their corresponding 2′-SeCH$_3$-modified counterpart. Our approach can be used, in particular, for the preparation of long 2′-SeCH$_3$-modified RNA targets that are inaccessible by chemical solid-phase synthesis, to determine their three-dimensional structure or to study their interaction with proteins by X-ray crystallography using the multiwavelength anomalous dispersion technique (MAD). We could show that both T7 RNA polymerase mutants that have been found either by site-directed mutagenesis studies or by directed evolution for their increased acceptance of 2′-modified NTPs could also be used for the synthesis of 2′-SeCH$_3$-modified RNA. These results are promising and a step towards an engineered RNA polymerase that processes all four 2′-SeCH$_3$–NTPs efficiently by means of directed protein evolution.

Acknowledgements

Notes and references

1 O. Thum, S. Jager and M. Famulok, *Angew. Chem., Int. Ed.*, 2001, **40**, 3990–3993.
2 S. Jager, G. Rasched, H. Kornreich-Leshem, M. Engeser, O. Thum and M. Famulok, *J. Am. Chem. Soc.*, 2005, **127**, 15071–15082.
3 J. P. Anderson, B. Angerer and L. A. Loeb, *BioTechniques*, 2005, **38**, 257–264.
4 P. Capek, H. Cahova, R. Pohl, M. Hocek, C. Gloeckner and A. Marx, *Chem.–Eur. J.*, 2007, **13**, 6196–6203.
5 T. Ohbayashi, M. Kuwahara, M. Hasegawa, T. Kasamatsu, T. Tamura and H. Sawai, *Org. Biomol. Chem.*, 2005, **3**, 2463–2468.
6 G. Giller, T. Tasara, B. Angerer, K. Muhlegger, M. Amacker and H. Winter, *Nucleic Acids Res.*, 2003, **31**, 2630–2635.
7 T. Tasara, B. Angerer, M. Damond, H. Winter, S. Dorhofer, U. Hubscher and M. Amacker, *Nucleic Acids Res.*, 2003, **31**, 2636–2646.
8 P. M. Gramlich, C. T. Wirges, J. Gierlich and T. Carell, *Org. Lett.*, 2008, **10**, 249–251.
9 M. Kuwahara, J. Nagashima, M. Hasegawa, T. Tamura, R. Kitagata, K. Hanawa, S. Hososhima, T. Kasamatsu, H. Ozaki and H. Sawai, *Nucleic Acids Res.*, 2006, **34**, 5383–5394.

10 B. Seelig and A. Jaschke, *Bioconjugate Chem.*, 1999, **10**, 371–378.
11 J. Wolf, V. Dombos, B. Appel and S. Muller, *Org. Biomol. Chem.*, 2008, **6**, 899–907.
12 D. Williamson, M. J. Cann and D. R. Hodgson, *Chem. Commun.*, 2007, 5096–5098.
13 R. Fiammengo, K. Musilek and A. Jaschke, *J. Am. Chem. Soc.*, 2005, **127**, 9271–9276.
14 F. Huang, J. He, Y. Zhang and Y. Guo, *Nat. Protoc.*, 2008, **3**, 1848–1861.
15 J. D. Vaught, T. Dewey and B. E. Eaton, *J. Am. Chem. Soc.*, 2004, **126**, 11231–11237.
16 J. F. Milligan and O. C. Uhlenbeck, *Methods Enzymol.*, 1989, **180**, 51–62.
17 N. Carrasco, J. Caton-Williams, G. Brandt, S. Wang and Z. Huang, *Angew. Chem., Int. Ed.*, 2005, **45**, 94–97.
18 S. G. Srivatsan and Y. Tor, *J. Am. Chem. Soc.*, 2007, **129**, 2044–2053.
19 E. L. Chernolovskaya and M. A. Zenkova, *Curr. Opin. Mol. Ther.*, 2010, **12**, 158–167.
20 M. A. Behlke, *Oligonucleotides*, 2008, **18**, 305–319.
21 J. K. Watts, G. F. Deleavey and M. J. Damha, *Drug Discovery Today*, 2008, **13**, 842–855.
22 G. Mayer, *Angew. Chem., Int. Ed.*, 2009, **48**, 2672–2689.
23 R. Sousa and R. Padilla, *EMBO J*, 1995, **14**, 4609–4621.
24 R. Padilla and R. Sousa, *Nucleic Acids Res.*, 1999, **27**, 1561–1563.
25 R. Padilla and R. Sousa, *Nucleic Acids Res.*, 2002, **30**, 138e.
26 J. Chelliserrykattil and A. D. Ellington, *Nat. Biotechnol.*, 2004, **22**, 1155–1160.
27 M. Teplova, C. J. Wilds, Z. Wawrzak, V. Tereshko, Q. Du, N. Carrasco, Z. Huang and M. Egli, *Biochimie*, 2002, **84**, 849–858.
28 N. Carrasco, D. Ginsburg, Q. Du and Z. Huang, *Nucleosides, Nucleotides Nucleic Acids*, 2001, **20**, 1723–1734.
29 Q. Du, N. Carrasco, M. Teplova, C. J. Wilds, M. Egli and Z. Huang, *J. Am. Chem. Soc.*, 2002, **124**, 24–25.
30 N. Carrasco, Y. Buzin, E. Tyson, E. Halpert and Z. Huang, *Nucleic Acids Res.*, 2004, **32**, 1638–1646.
31 S. Freisz, K. Lang, R. Micura, P. Dumas and E. Ennifar, *Angew. Chem., Int. Ed.*, 2008, **47**, 4110–4113.
32 A. Serganov, S. Keiper, L. Malinina, V. Tereshko, E. Skripkin, C. Hobartner, A. Polonskaia, A. T. Phan, R. Wombacher, R. Micura, Z. Dauter, A. Jaschke and D. J. Patel, *Nat. Struct. Mol. Biol.*, 2005, **12**, 218–224.
33 C. Hobartner, R. Rieder, C. Kreutz, B. Puffer, K. Lang, A. Polonskaia, A. Serganov and R. Micura, *J. Am. Chem. Soc.*, 2005, **127**, 12035–12045.
34 B. Puffer, H. Moroder, M. Aigner and R. Micura, *Nucleic Acids Res.*, 2008, **36**, 970–983.
35 Y. Buzin, N. Carrasco and Z. Huang, *Org. Lett.*, 2004, **6**, 1099–1102.
36 J. Sheng, J. Jiang, J. Salon and Z. Huang, *Org. Lett.*, 2007, **9**, 749–752.
37 H. Moroder, C. Kreutz, K. Lang, A. Serganov and R. Micura, *J. Am. Chem. Soc.*, 2006, **128**, 9909–9918.
38 J. Sheng and Z. Huang, *Chem. Biodiversity*, 2010, **7**, 753–785.
39 V. Olieric, U. Rieder, K. Lang, A. Serganov, C. Schulze-Briese, R. Micura, P. Dumas and E. Ennifar, *RNA*, 2009, **15**, 707–715.
40 K. Lang and R. Micura, *Nat. Protoc.*, 2008, **3**, 1457–1466.
41 C. Hobartner and R. Micura, *J. Am. Chem. Soc.*, 2004, **126**, 1141–1149.
42 R. Micura, C. Hobartner, R. Rieder, C. Kreutz, B. Puffer, K. Lang and H. Moroder, *Curr. Protoc. Nucleic Acid Chem.*, 2007, (Chapter 1), Unit 1 15.
43 B. Puffer, H. Moroder, M. Aigner and R. Micura, *Nucleic Acids Res.*, 2008, **36**, 970–983.
44 R. Peist, A. Koch, P. Bolek, S. Sewitz, T. Kolbus and W. Boos, *J. Bacteriol.*, 1997, **179**, 7679–7686.
45 J. Ludwig and F. Eckstein, *J. Org. Chem.*, 1989, **54**, 631–635.
46 Y. Huang, F. Eckstein, R. Padilla and R. Sousa, *Biochemistry*, 1997, **36**, 8231–8242.
47 L. G. Brieba and R. Sousa, *Biochemistry*, 2000, **39**, 919–923.