DOI: 10.1002/hlca.202000016

FULL PAPER

HELVETICA chimica acta

Synthesis and Incorporation of k^2U into RNA

Milda Nainyte[a] and Thomas Carell*[a]

[a] Department of Chemistry, Ludwig-Maximilians-Universität, Butenandtstr. 5–13, DE-81377 Munich, e-mail: Thomas.carell@lmu.de

Dedicated to Prof. *François Diederich* on the occasion of his retirement

Lysidine (k^2C) is one of the most modified pyrimidine RNA bases. It is a cytidine nucleoside, in which the 2-oxo functionality of the heterocycle is replaced by the ε-amino group of the amino acid lysine. As such, lysidine is an amino acid-containing RNA nucleoside that combines directly genotype (C-base) with phenotype (lysine amino acid). This makes the compound particularly important in the context of theories about the origin of life and here especially for theories that target the origin of translation. Here, we report the total synthesis of the U-derivative of lysidine (k^2U), which should have the same base pairing characteristics as k^2C if it exists in the isoC-like tautomeric form. To investigate this question, we developed a phosphoramidite building block for k^2U, which allows its incorporation into RNA strands. Within RNA, k^2U can base pair with the counter base U and isoG, confirming that k^2U prefers an isoC-like tautomeric structure that is also known to dominate for k^2C. The successful synthesis of a k^2U phosphoramidite and its use for RNA synthesis now paves the way for the preparation of a k^2C phosphoramidite and RNA strands containing k^2C.

Keywords: lysidine, modified RNA bases, RNA, nucleosides, origin of life, prebiotic chemistry.

Introduction

RNA contains a vast variety of modified nucleosides.[1] Many of the modified bases are just methylated versions of the canonical nucleosides A, C, G and U, but other are highly modified. This is achieved with the help of dedicated biosynthesis machineries.[2] From a prebiotic point of view in which genotype-phenotype discussions address the question of whether life started with nucleic acids or peptides, the most interesting modified nucleosides are those which are modified with amino acids.[3] These molecules are chemical structures between genotype and phenotype. They directly merge the properties of nucleic acids with those of amino acids. As such, RNA containing these amino acid-modified bases could have been a central element for the origin of life and for the origin of translation as already discussed by *Grosjean* and others.[4–7] RNA containing these amino acids could establish catalytic properties next to information encoding functions. Catalytic RNAs in turn are the central elements in all discussions about how life emerged from simple starting materials.[8,9]

One of the most interesting amino acid modified nucleosides is lysidine (k^2C),[10,11] which is a cytidine base to which a lysine amino acid is attached with the ε-amino group to the C(2) position (*Figure 1*). Agmatidine (agm^2C)[12,13] is a close relative, which features the guanidinium group found in the amino acid arginine (*Figure 1*). We also want to mention the amino acid modified RNA base puromycin in which the amino acid is attached to the C(3′)-OH group. This base is in use for ribosomal studies.[14,15] In our effort to investigate the properties of RNA containing amino acid-modified nucleosides as units that equip such RNA potentially with peptide-like properties, we were interested to study the lysine modified uridine base. This compound is a close derivative of lysidine, which we name k^2U. k^2U is in principle the deamination product of k^2C which has likely occurred under prebiotic conditions, where the bases were exposed to potentially hot aqueous environments.[16,17] Under

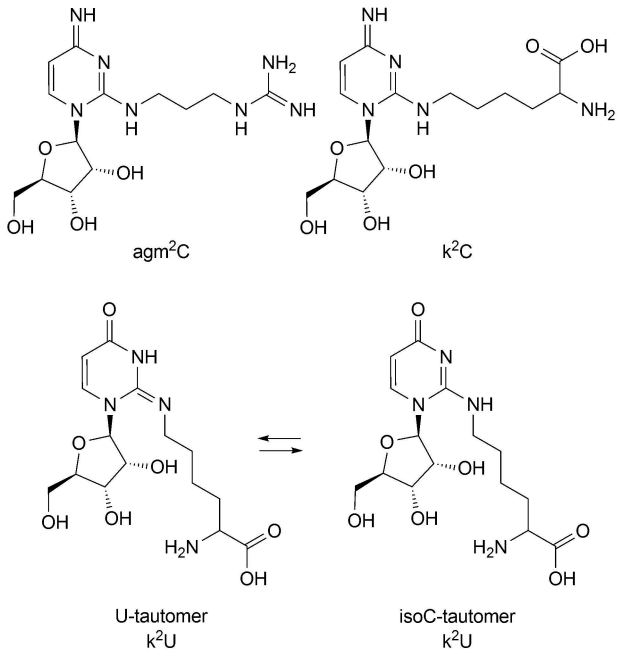


Figure 1. Depiction of agmatidine (agm^{2}C), lysidine (k^{2}C) and of the potential prebiotically relevant deamination product (k^{2}U) of k^{2}C.

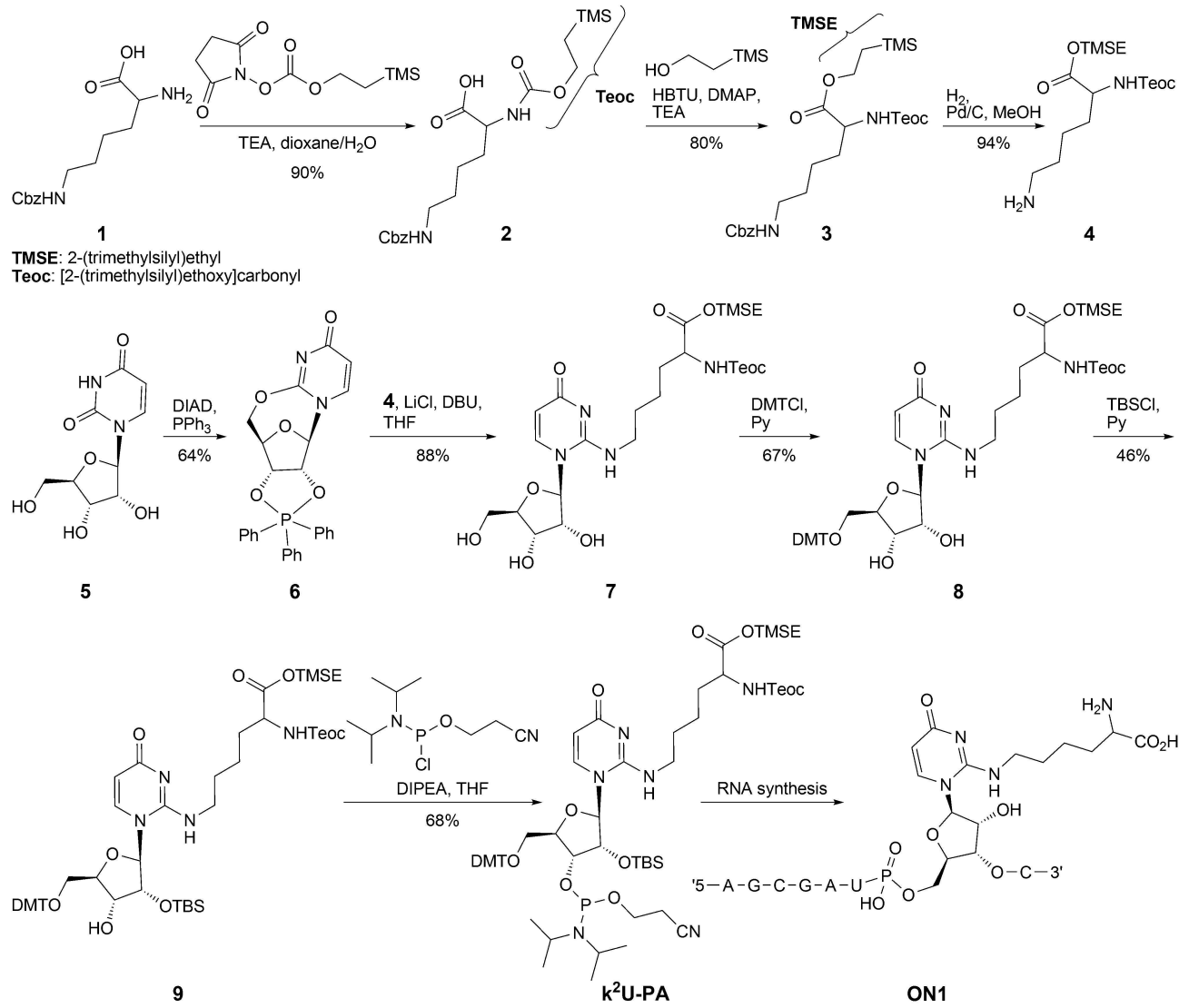
these conditions, deamination of k^{2}C to k^{2}U is an expected process. k^{2}U can in principle exist is two tautomeric states. In one it has U-type base pairing properties, while in the second it should behave like an isoC similar to k^{2}C, which also exists in a quinoid-like tautomeric structure (*Figure 1*)[10,18–23]. In this latter scenario, k^{2}U and k^{2}C are supposed to have similar base pairing properties. A major difference could be that while k^{2}C is protonated, k^{2}U is likely not. k^{2}C protonation is supposed to occur at N(6) and hence it does not change the base pairing properties.[10,24]

Results and Discussion

To investigate how k^{2}U is affecting the structure and properties of RNA, we started the synthesis of k^{2}U as its phosphoramidite building block (**k^{2}U-PA**) and investigated procedures that allow its incorporation into RNA. The synthesis of the **k^{2}U-PA** is depicted in *Scheme 1*. The synthesis was started with carboxybenzoyl- (Cbz) protected lysine **1**, which we converted with 1-({[2-(trimethylsilyl)ethoxy]carbonyl}oxy)pyrrolidine-2,5-dione into the Cbz- and Teoc-protected lysine compound **2**. Subsequent protection of the carboxy group with 2-(trimethylsilyl)ethanol furnished the fully protected lysine amino acid **3**. After Cbz-deprotection, we obtained the amino acid coupling partner **4**. At the nucleoside side, we treated uridine **5** under *Mitsunobu* conditions with DIAD and PPh_3 to obtain the literature-known 2′,3′-protected cyclouridine compound **6**.[25] Coupling of this intermediate with the lysine building block **4** in the presence of LiCl and DBU furnished the protected lysine coupled uridine derivative **7**. We next protected the primary 5′-OH group with DMTCl and the secondary 2′-OH group with TBSCl to give the intermediate **9**.[26–28] Compound **9** was finally converted into the phosphoramidite building block using a standard procedure.[29] Purification of the **k^{2}U-PA** was difficult due to its high polarity. We needed to use a rather polar mixture of dichloromethane/acetone (8:3) as the mobile phase for the chromatographic purification. This provided, however, the target compound **k^{2}U-PA** in a total yield of 12% in just eight steps in a good but not excellent purity.

To study the properties of k^{2}U in RNA, we next inserted the compound into an RNA strand using solid phase RNA synthesis.[30–33] The oligonucleotide synthesis was performed using standard coupling conditions. This allowed us to achieve a coupling yield of 30%. Importantly, the incorporation of the next canonical bases, *e.g.*, coupling of uridine to k^{2}U, was not affected. We achieved elongation yields that reached typically 95–98%. All together the method provided enough material for all further studies. Optimization of the k^{2}U-incorporation yield was consequently not performed. In addition, we noted that the purity of the obtained strands was high allowing rapid separation of the target k^{2}U containing oligonucleotide. After the RNA synthesis, we removed the Teoc- and TMSE-protecting groups from the lysine moiety with saturated $ZnBr_2$ solution in isopropanol/nitromethane (1:1) at r.t. over night. It is interesting that the RNA strand is stable under these quite *Lewis* acidic conditions. Our observation, however, agrees with an earlier report.[34] Own attempts to cleave the Teoc- and TMSE-protecting groups with HF·Et_3N complex provided only partial deprotection which gave a mixture of products. In our hands, the reported $ZnBr_2$ method gave superior results.[33] RNA degradation was not observed.

We subsequently cleaved the oligonucleotide from the solid support and removed the protecting groups from the canonical bases with a mixture of methylamine and ammonia (1:1) at 65 °C for 5 min. *Figure 2* shows the sequence of the prepared oligonucleotide together with the raw HPL-chromatogram and the MALDI-TOF spectrum. These data prove the integrity



Scheme 1. Synthesis of k^{2}U and of its phosphoramidite building block **k^{2}U-PA**.

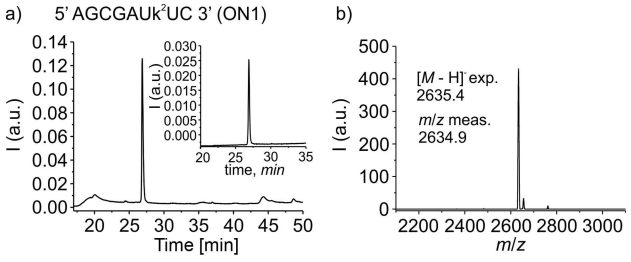
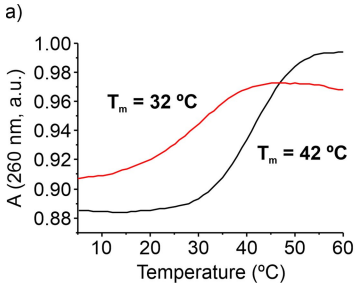


Figure 2. a) Sequence of the k^{2}U-containing RNA strand together with the raw-HPL chromatogram of **ON1**. The inset shows the HPL-chromatogram of the purified **ON1**. b) MALDI-TOF mass spectrum of **ON1** after purification.

and the high purity of the synthesized k^{2}U-containing RNA strand.

We finally studied how the k^{2}U base affects the stability of the RNA duplex. *Figure 3* shows the melting curves of the k^{2}U-containing oligonucleotide (k^{2}U:A base pair, red) in comparison to the unmodified RNA duplex containing a U:A base pair (black). The table in *Figure 3* depicts all melting points measured for k^{2}U facing any of the four canonical bases, together with all possible combinations of canonical bases as reference strands. While the U:A reference strand melts at 42 °C, replacement of U by k^{2}U reduces the melting temperature (k^{2}U:A) to just 32 °C, which is a dramatic destabilization. This is unexpected and not explainable with a U-type tautomeric structure, because the lysine residue can in principle point out of the large shallow groove of the RNA-duplex in A-



	T_m, °C		T_m, °C
U:A	42	k^2U:A	32
U:G	38	k^2U:G	35
U:C	29	k^2U:C	33
U:U	34	k^2U:U	38

Figure 3. a) Depiction of melting curve with the duplex containing U:A base pair (black), and k^2U:A (red); b) table of all melting points measured for k^2U.

conformation. Because the C(2)-O atom, which is replaced by the lysine residue, does not take part in the H-bonding to the A-counterbase as depicted in *Figure 1*, the lysine residue should not affect base pairing so strongly if k^2U would exist in the U-tautomeric form. Further melting point studies in which we exchanged systematically the counter base showed the k^2U is unable to undergo any productive base pairing. The only slight stabilization that we detected is in the k^2U:U situation (38 °C), which is compared to the U:U (34 °C) situation stabilized by 4 °C.

This rather large global destabilization is best explained, if we assume that the k^2U base exists indeed not in the typical U-tautomer but in the hemiquinoid tautomeric state known from isoC (*Figure 4*). As such, k^2U behaves like an isoC tautomer, which indeed prohibits k^2U to form productive base pairs with any of the other canonical bases particularly with the purine bases A and G. The slight interaction with U is in this scenario explained because the U counterbase can get engaged with isoC-tautomeric k^2U with two H-bonds. To investigate this isoC type

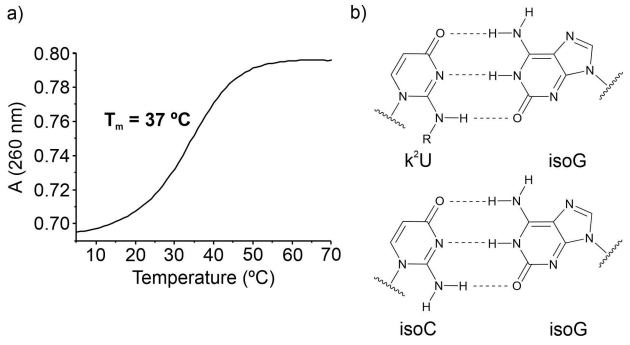


Figure 4. a) Depiction of melting curve with the duplex containing k^2U:isoG base pair; b) depiction of base pairing of isoC:isoG and k^2U:isoG.

structure of k^2U in more detail, we created a duplex in which k^2U is facing isoG as the counterbase. Indeed, in this situation we measure a higher melting point of 37 °C (*Figure 4*) close to the k^2U:U situation (38 °C), in fully agreement with the idea that k^2U is a lysine modified isoC derivative that has an k^2C like tautomeric structure. Further support for the idea that the k^2U base exists predominantly in the quinoid tautomeric structure comes from NMR data. In our compounds with the k^2U base the typical NH ^{1}H-NMR signal around $\delta = 9.5$ ppm was not observed. Instead we observed the NH signal at around $\delta = 7.0$ ppm, in line with the quinoid tautomeric structure of k^2U (*Figure 4*). This shift to around $\delta = 7.0$ ppm agrees with literature data about such compounds.[35]

Conclusions

Here, we report the development of a phosphoramidite building block of k^2U which is a close deamination-based relative of the non-canonical base k^2C. We show that the modified base can be incorporated into RNA strands using standard phosphoramidite chemistry in combination with a three-stage deprotection protocol, in which we first cleave the protecting groups at the lysine residue, followed by deprotection of the nucleobases and cleavage from the resin. Melting point studies and NMR data show that k^2U exists in a tautomeric state that resembles the situation in isoC. As such, k^2U destabilizes RNA duplexes dramatically allowing only limited interactions with U and most importantly isoG as counterbases.

Experimental Section

General Methods

Chemicals were purchased from *Sigma–Aldrich*, *TCI*, *Fluka*, *ABCR*, *Carbosynth* or *Acros organics* and used without further purification. Strands containing canonical bases and isoG were purchased from *Metabion*. The solvents were of reagent grade or purified by distillation, unless otherwise specified. Reactions and chromatography fractions were monitored by qualitative thin-layer chromatography (TLC) on silica gel *F254* TLC plates from *Merck KGaA*. Flash chromatography was performed on *Geduran® Si60* (40–63 μm) silica gel from *Merck KGaA*. NMR spectra were recorded on *Bruker AVIIIHD 400* spectrometers (400 MHz). ^{1}H-NMR shifts were calibrated to the residual solvent resonan-

ces: (D_6)DMSO (2.50 ppm), $CDCl_3$ (7.26 ppm), (D_6)acetone (2.05 ppm). ^{13}C-NMR shifts were calibrated to the residual solvent: (D_6)DMSO (39.52 ppm), $CDCl_3$ (77.16 ppm), (D_6)acetone (29.84 ppm). All NMR spectra were analyzed using MestRENOVA 10.01.1 from Mestrelab Research S. L. High resolution mass spectra were measured on the spectrometer *MAT 90* (ESI) from *Thermo Finnigan GmbH*. Analytical RP-HPLC was performed on an analytical HPLC *Waters Alliance* (*2695 Separation Module*, *2996 Photodiode Array Detector*) equipped with the column *Nucleosil 120–2C18* from *Macherey Nagel* using a flow of 0.5 ml/min, a gradient of 0–30% of buffer *B* in 45 min was applied. Preparative RP-HPLC was performed on a HPLC *Waters Breeze* (*2487 Dual λ Array Detector*, *1525 Binary* HPLC Pump) equipped with the column *VP 250/32C18* from *Macherey Nagel* using a flow of 5 ml/min, a gradient of 0–25% of buffer *B* in 45 min was applied for the purifications. Oligonucleotides were purified using the following buffer system: buffer *A*: 100 mM NEt_3/AcOH (pH 7.0) in H_2O and buffer *B*: 100 mM NEt_3/AcOH in 80% (*v*/*v*) acetonitrile. The pH values of buffers were adjusted using a *MP 220* pH-meter (*Metter Toledo*). Oligonucleotides were detected at wavelength: 260 nm. Melting profiles were measured on a *JASCO V-650* spectrometer. Calculation of concentrations was assisted using the software OligoAnalyzer 3.0. For strands containing artificial bases, the extinction coefficient of their corresponding canonical-only strand was employed without corrections. Matrix-assisted laser desorption/ionization-time-of-flight (MALDI-TOF) mass spectra were recorded on a *Bruker Autoflex II*. For MALDI-TOF measurements, the samples were desalted on a 0.025 μm *VSWP* filter (*Millipore*) against ddH_2O and co-crystallized in a 3-hydroxypicolinic acid matrix (HPA).

N^6-[(Benzyloxy)carbonyl]-N^2-{[2-(trimethylsilyl)-ethoxy]carbonyl}lysine (**2**). The reaction was performed according to the published procedure.[36]

To a stirred suspension of cbz-protected lysine (**1**) (3 g, 10.7 mmol, 1 equiv.) in dioxane/water (1:1) mixture, TEA (2.23 ml, 16 mmol, 1.5 equiv.) was added. To the resultant solution, Teoc–OSu (3.2 g, 12.3 mmol, 1.1 equiv.) was added and the mixture was left to stir overnight at room temperature. Afterwards, the mixture was diluted with water, acidified with saturated $KHSO_4$ solution and extracted with diethyl ether. The combined organic layers were washed with water, dried over Na_2SO_4 and evaporated *in vacuo*. Yield: 85%. ^{1}H-NMR (400 MHz, $CDCl_3$): 7.35–7.33 (*m*, 5 H); 5.50–5.45 (*m*, 1 H); 5.09 (*s*, 2 H); 4.41–4.32 (*m*, 1 H); 4.15 (*t*, *J* = 8.4, 2 H); 3.22–3.13 (*m*, 2 H); 1.86–1.72 (*m*, 2 H); 1.51–1.40 (*m*, 4 H); 0.97 (*t*, *J* = 8.4, 2 H); 0.02 (*s*, 9 H). ^{13}C-NMR (101 MHz, $CDCl_3$): 176.4; 168.9; 156.8; 136.6; 128.6; 128.4; 128.3; 66.9; 63.7; 53.5; 40.6; 31.9; 29.5; 22.3; 17.7; −1.4. HR-ESI-MS: 425.2103 ($C_{20}H_{33}N_2O_6Si^+$, $[M+H]^+$; calc. 425.2108).

2-(Trimethylsilyl)ethyl N^6-[(benzyloxy)carbonyl]-N^2-{[2-(trimethylsilyl)ethoxy]carbonyl}lysinate (**3**). To the stirred solution of compound **2** (1.2 g, 2.82 mmol, 1 equiv.) in CH_2Cl_2/DMF (1:1) mixture, HBTU (1.61 g, 4.2 mmol, 1.5 equiv.), TEA (0.788 ml, 5.65 mmol, 2 equiv.) and DMAP (0.034 g, 0.27 mmol, 0.1 equiv.) were added. Then, 2-(trimethylsilyl)ethanol (0.486 ml, 4.2 mmol, 1.2 equiv.) was added and the mixture was left to stir overnight at room temperature. Then, CH_2Cl_2 was evaporated and product was extracted with diethyl ether and washed with water. The organic phase was dried and evaporated. The residue was purified by flash chromatography eluting with hexane/AcOEt (4:1). Yield: 80%. ^{1}H-NMR (400 MHz, $CDCl_3$): 7.35–7.30 (*m*, 5 H); 5.19 (*d*, *J* = 8, 1 H); 5.08 (*s*, 2 H); 4.81 (br. *s*, 1 H); 4.32–4.26 (*m*, 1 H); 4.24–4.18 (*m*, 2 H); 4.14 (*t*, *J* = 9, 2 H); 3.18 (*q*, *J* = 6.6, 2 H); 1.82–1.62 (*m*, 2 H); 1.55–1.50 (*m*, 2 H); 1.42–1.34 (*m*, 2 H); 1.00–0.96 (*m*, 4 H); 0.04 (*s*, 9 H); 0.02 (*s*, 9 H). ^{13}C-NMR (101 MHz, $CDCl_3$): 172.8; 156.6; 156.4; 136.7; 128.6; 128.4; 128.2; 66.7; 63.9; 63.5; 53.7; 40.7; 32.4; 29.5; 22.4; 17.8; 17.5; −1.36; −1.39. HR-ESI-MS: 525.2817 ($C_{25}H_{45}N_2O_6Si_2^+$, $[M+H]^+$; calc. 525.2816).

2-(Trimethylsilyl)ethyl N^2-{[2-(trimethylsilyl)ethoxy]carbonyl}lysinate (**4**). Cbz-protected lysine derivative **3** (0.7 g, 1.33 mmol, 1 equiv.) was dissolved in methanol. Pd/C (0.096 g) was added, and the mixture was stirred at room temperature under H_2 at atmospheric pressure. After 2 h, the mixture was filtered through *Celite* and evaporated. Yield: 94%. ^{1}H-NMR (400 MHz, $CDCl_3$): 5.18 (*d*, *J* = 8, 1 H); 4.33–4.26 (*m*, 1 H); 4.23–4.18 (*m*, 2 H); 4.14 (*t*, *J* = 8.5, 2 H); 2.69 (br. *s*, 2 H); 1.80–1.34 (*m*, 6 H); 0.98 (*q*, *J* = 8.5, 4 H); 0.04 (*s*, 9 H); 0.03 (*s*, 9 H). ^{13}C-NMR (101 MHz, $CDCl_3$): 172.9; 156.4; 64.0; 63.4; 53.8; 42.2; 32.6; 28.1; 22.6; 17.8; 17.4; −1.37; −1.40. HR-ESI-MS: 391.2444 ($C_{17}H_{39}N_2O_4Si_2^+$, $[M+H]^+$; calc. 391.2448).

2′,3′-*O*-(Triphenylphosphoranediyl)-O^2,5′-cyclouridine (**=(3a*R*,4*R*,12*R*,12a*R*)-2,2,2-Triphenyl-4,5,12,12a-tetrahydro-2*H*,3a*H*,8*H*-4,12-epoxy-2λ^5-[1,3,2]dioxaphospholo[4,5-*e*]pyrimido[2,1-*b*][1,3]-oxaz-

ocin-8-one; 6). The reaction was performed according to the published procedure.[25]

To a stirred solution of uridine **5** (5 g, 20.5 mmol, 1 equiv.) in THF, PPh_3 (16.1 g, 61.4 mmol, 3 equiv.) was added. Then DIAD (12.1 ml, 61.4 mmol, 3 equiv.) was added dropwise, and the reaction was left to stir at room temperature overnight. Afterwards the precipitate was filtered, washed with THF and water, and then dried. Yield: 64%. ^{1}H-NMR (400 MHz, (D_6)DMSO): 8.05 (*d*, *J* = 7.5, 1 H); 7.67–7.51 (*m*, 2 H); 7.47–7.17 (15 H); 5.96–5.85 (*m*, 2 H); 4.84 (*dd*, *J* = 6.6, 13.2, 1 H); 4.73 (*s*, 1 H); 4.60–4.46 (*m*, 2 H); 4.16 (*d*, *J* = 12.7, 1 H). ^{13}C-NMR (101 MHz, (D_6)DMSO): 170.7; 157.0; 144.8; 143.6; 130.8; 128.5; 127.6; 109.0; 97.8; 85.4; 76.8; 74.4; 72.7. HR-ESI-MS: 487.1412 ($C_{27}H_{24}N_2O_5P^+$, $[M+H]^+$; calc. 487.1423).

2-(Trimethylsilyl)ethyl N^6-(4-oxo-1-pentofuranosyl-1,4-dihydropyrimidin-2-yl)-N^2-{[2-(trimethylsilyl)ethoxy]carbonyl}lysinate (**7**). To the dry LiCl (0.043 g, 1.02 mmol, 5 equiv.) THF was poured followed by the addition of DBU (0.156 g, 1.02 mmol, 5 equiv.) and amine **4** (0.16 g, 0.41 mmol, 2 equiv.). Then, cyclouridine derivative **6** (0.1 g, 0.2 mmol, 1 equiv.) was added, and the mixture was left to stir for 2 h at room temperature. In time, the solution became pink and clear. Afterwards, the mixture was evaporated, and the residue was purified by flash chromatography eluting with MeCN/H_2O (95:5, *v/v*). Yield: 88%. ^{1}H-NMR (400 MHz, (D_6)DMSO): 7.82 (*s*, 1 H); 7.58 (*d*, *J* = 7.7, 1 H); 7.44 (*dd*, *J* = 7.7, 4.6, 1 H); 7.12 (*t*, *J* = 5.2, 1 H); 5.53 (*d*, *J* = 7.7, 1 H); 5.50 (*d*, *J* = 7.0, 1 H); 5.44 (*t*, *J* = 4.6, 1 H); 5.39 (*d*, *J* = 7.0, 1 H); 5.29 (*d*, *J* = 4.3, 1 H); 4.16–4.08 (*m*, 4 H); 4.04–3.97 (*m*, 4 H); 3.20 (*q*, *J* = 6.7, 2 H); 2.74 (*t*, *J* = 7.6, 1 H); 1.66–1.43 (*m*, 4 H); 1.37–1.26 (*m*, 2 H); 0.95–0.89 (*m*, 4 H); 0.02 (*s*, 18 H). ^{13}C-NMR (101 MHz, (D_6)DMSO): 172.7; 163.5; 156.5; 150.9; 140.9; 101.9; 87.7; 73.7; 70.0; 62.7; 62.0; 60.9; 53.9; 51.9; 30.8; 30.1; 26.7; 22.6; 17.5; 16.9; −1.32; −1.35. HR-ESI-MS: 617.3028 ($C_{26}H_{49}N_4O_9Si_2^+$, $[M+H]^+$; calc. 617.3038).

2-(Trimethylsilyl)ethyl N^6-(1-{5-*O*-[bis(4-methoxyphenyl)(phenyl)methyl]pentofuranosyl}-4-oxo-1,4-dihydropyrimidin-2-yl)-N^2-{[2-(trimethylsilyl)-ethoxy]carbonyl}lysinate (**8**). To the solution of nucleoside **7** (0.45 g, 0.73 mmol, 1 equiv.) in dry pyridine, DMTCl (0.37 g, 1.1 mmol, 1.5 equiv.) was added. The mixture was left to stir overnight at room temperature. The product was extracted with AcOEt and washed with sat. $NaHCO_3$ solution. The combined organic layers were evaporated, and the residue was purified by flash chromatography eluting with CH_2Cl_2/MeOH (10:1, *v/v*) containing 0.02% of pyridine. Yield: 67%. ^{1}H-NMR (400 MHz, (D_6)acetone): 7.86 (*d*, *J* = 7.8, 1 H); 7.49–7.44 (*m*, 2 H); 7.38–7.20 (*m*, 8 H); 7.02 (*t*, *J* = 5.3, 1 H); 6.92–6.87 (*m*, 4 H); 6.48 (*d*, *J* = 8.2, 1 H); 5.71 (*d*, *J* = 4.3, 1 H); 5.57 (*d*, *J* = 7.7, 1 H); 4.69 (br. *s*, 1 H); 4.49–4.47 (*m*, 1 H); 4.43–4.41 (*m*, 1 H); 4.23 (*q*, *J* = 3.5, 1 H); 4.18–4.04 (*m*, 5 H); 3.77 (*s*, 6 H); 3.47 (*d*, *J* = 3.5, 2 H); 3.41–3.30 (*m*, 2 H); 1.83–1.52 (*m*, 4 H); 1.48–1.38 (*m*, 2 H); 1.00–0.92 (*m*, 4 H); 0.03 (*s*, 9 H); 0.02 (*s*, 9 H). ^{13}C-NMR (101 MHz, (D_6)acetone): 173.4; 172.2; 159.7; 157.3; 154.3; 145.7; 139.1; 136.4; 136.2; 131.0; 128.9; 128.7; 127.7; 114.0; 106.4; 92.1; 87.5; 85.4; 75.5; 71.2; 63.6; 63.1; 55.6; 55.1; 41.9; 32.2; 23.9; 18.3; 17.9; −1.3; −1.4. HR-ESI-MS: 919.4326 ($C_{47}H_{67}N_4O_{11}Si_2^+$, $[M+H]^+$; calc. 919.4345).

2-(Trimethylsilyl)ethyl N^6-(1-{5-*O*-[bis(4-methoxyphenyl)(phenyl)methyl]-2-*O*-[*tert*-butyl(dimethyl)silyl]pentofuranosyl}-4-oxo-1,4-dihydropyrimidin-2-yl)-N^2-{[2-(trimethylsilyl)ethoxy]carbonyl}lysinate (**9**). DMT-protected nucleoside **8** (0.1 g, 0.1 mmol, 1 equiv.) was dissolved in pyridine. Then, imidazole (0.022 g, 0.3 mmol, 3 equiv.) and TBSCl (0.049 g, 0.3 mmol, 3 equiv.) were added. The mixture was left to stir overnight at room temperature. Then, the mixture was poured into saturated $NaHCO_3$ solution and extracted with CH_2Cl_2. Organic layers were evaporated, and the residue was purified by flash chromatography eluting with CH_2Cl_2/acetone (8:3, *v/v*) containing 0.02% of pyridine. Yield: 46%. ^{1}H-NMR (400 MHz, (D_6)acetone): 7.49–7.44 (*m*, 2 H); 7.38–7.29 (*m*, 9 H); 6.91 (*d*, *J* = 8.9, 5 H); 5.67 (*d*, *J* = 5.3, 1 H); 5.53 (*d*, *J* = 7.7, 1 H); 4.45 (*t*, *J* = 5.3, 1 H); 4.41–4.35 (*m*, 1 H); 4.21–4.14 (*m*, 4 H); 4.13–4.06 (*m*, 4 H); 3.79 (*s*, 6 H); 3.55 (*dd*, *J* = 11.0, 3.0, 1 H); 3.38 (*dd*, *J* = 11.0, 3.0, 1 H); 3.31 (*t*, *J* = 6.7, 2 H); 1.81–1.67 (*m*, 2 H); 1.65–1.50 (*m*, 2 H); 1.47–1.34 (*m*, 2 H); 1.02–0.95 (*m*, 2 H); 0.86 (*s*, 9 H); 0.11 (*s*, 3 H); 0.05 (*s*, 9 H); 0.03 (*s*, 9 H). ^{13}C-NMR (101 MHz, (D_6)acetone): 173.4; 170.9; 159.7; 157.2; 154.3; 145.6; 138.9; 136.2; 136.1; 131.0; 128.9; 128.7; 127.8; 114.0; 107.8; 91.4; 87.7; 85.9; 75.9; 71.7; 64.2; 63.5; 63.0; 55.5; 55.1; 41.9; 32.2; 26.1; 24.1; 18.7; 18.3; 17.9; −1.37; −1.41; −4.5; −4.7. HR-ESI-MS: 1033.5205 ($C_{53}H_{81}N_4O_{11}Si_3^+$, $[M+H]^+$; calc. 1033.5210).

2-(Trimethylsilyl)ethyl N^6-[1-(5-*O*-[bis(4-methoxyphenyl)(phenyl)methyl]-2-*O*-[*tert*-butyl(dimethyl)silyl]-3-*O*-{(2-cyanoethoxy)[di(propan-2-yl)amino]phosphanyl}pentofuranosyl)-4-oxo-1,4-dihydropyrimidin-2-yl]-N^2-{[2-(trimethylsilyl)ethoxy]car-

bonyl}lysinate (**k²U-PA**). Compound **9** (0.1 g, 0.09 mmol, 1 equiv.) was co-evaporated three times in dry pyridine. Then, the compound was dissolved in dry THF and cooled to 0 °C. DIPEA (67 µl, 0.39 mmol, 4 equiv.) and 2-cyanoethyl-*N*,*N*-diisopropyl chlorophosphoramidite (54 µl, 0.24 mmol, 2.5 equiv.) were added under nitrogen atmosphere. The mixture was brought to room temperature and stirred for 2.5 h. Then, the reaction was quenched by the addition of saturated $NaHCO_3$ solution and extracted with CH_2Cl_2. The combined organic layers were dried and concentrated. The residue was purified by flash chromatography eluting with CH_2Cl_2/acetone (8:3, *v*/*v*) containing 0.3% of pyridine. Yield: 68%. ^{31}P-NMR (162 MHz, (D_6)acetone): 149.9; 148.6. HR-ESI-MS: 1233.6278 ($C_{62}H_{98}N_6O_{12}PSi_3^+$, $[M+H]^+$; calc. 1233.6288).

Synthesis and Purification of Oligonucleotide

The oligonucleotide was synthesized on a 1 µmol scale using a DNA automated synthesizer (*Applied Biosystems 394 DNA/RNA Synthesizer*) with standard phosphoramidite chemistry. The phosphoramidites of canonical ribonucleotides were purchased from *Glen Research* and *Sigma-Aldrich*. Oligonucleotide containing k²U nucleoside was synthesized in DMT-OFF mode using phosphoramidites (Bz–A, Dmf-G, Ac–C, U) with BTT in MeCN as an activator, DCA in CH_2Cl_2 as a deblocking solution and Ac_2O in pyridine/THF as a capping reagent. After the synthesis the solid support was treated with saturated solution of $ZnBr_2$ in ^{i}PrOH/$MeNO_2$ (1:1, *v*/*v*, 1 ml) and left overnight at room temperature. Then the beads were washed with water and 0.1 M EDTA solution. The cleavage and deprotection of CPG-bound oligonucleotide were performed with aq. NH_4OH/$MeNH_2$ solution (1:1, *v*/*v*, 1 ml) for 5 min at 65 °C. The resin was removed by filtration, washed with H_2O and the solution was evaporated under reduced pressure. The residue was subsequently heated with a solution of triethylamine trihydrofluoride (125 µl) in DMSO (50 µl) at 65 °C for 1.5 h. Upon cooling in an ice bath, AcONa (3 M, 25 µl) and BuOH (1 ml) were added. The resulting suspension was vortexed and cooled in a freezer (−80 °C) for 1 h. After the centrifugation, supernatant was removed, and the remaining oligonucleotide pellet was dried under vacuum. The oligonucleotide was analyzed and purified using RP-HPLC. The structural integrity of RNA was analyzed by MALDI-TOF mass measurement.

Acknowledgements

We would like to thank the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) for funding through the programs SFB1309 (325871075), SPP-1784. We also thank the European Research Council (ERC-AG) under the European Union's Horizon 2020 research and innovation program (Project ID: 741912, EpiR) for additional funding. We thank Dr. *Tynchtyk Amatov* for the help with the initial synthesis.

Author Contribution Statement

All syntheses and characterization of the compounds and measurements were carried out by *M. N. T. C.* designed the study and supervised the experimental work. *M. N.* and *T. C.* wrote the manuscript.

References

[1] M. A. Machnicka, K. Milanowska, O. O. Oglou, E. Purta, M. Kurkowska, A. Olchowik, W. Januszewski, S. Kalinowski, S. Dunin-Howkawicz, K. M. Rother, M. Helm, J. M. Bujnicki, H. Grosjean, 'MODOMICS: a database of RNA modification pathways–2013 update', *Nucleic Acids Res.* **2013**, *41*, D262–D267.

[2] M. T. Watts, I. Tinoco, 'Role of hypermodified bases in transfer RNA. Solution properties of dinucleoside monophosphates', *Biochemistry* **1978**, *17*, 2455–2463.

[3] H. Grosjean, E. Westhof, 'An integrated, structure- and energy-based view of the genetic code', *Nucleic Acids Res.* **2016**, *44*, 8020–8040.

[4] C. Schneider, S. Becker, H. Okamura, A. Crisp, T. Amatov, M. Stadlmeier, T. Carell, 'Noncanonical RNA Nucleosides as Molecular Fossils of an Early Earth–Generation by Prebiotic Methylations and Carbamoylations', *Angew. Chem. Int. Ed.* **2018**, *57*, 5943–5946; *Angew. Chem.* **2018**, *130*, 6050–6054.

[5] M. Di Giulio, 'Reflections on the Origin of the Genetic Code: A Hypothesis', *J. Theor. Biol.* **1998**, *191*, 191–196.

[6] H. Grosjean, V. de Crécy-Lagard, G. R. Björk, 'Aminoacylation of the anticodon stem by a tRNA-synthetase paralog: relic of an ancient code?', *Trends Biochem. Sci.* **2004**, *29*, 519–522.

[7] E. Szathmáry, 'The origin of the genetic code: amino acids as cofactors in an RNA world', *Trends Genet.* **1999**, *15*, 223–229.

[8] T. R. Cech, A. J. Zaug, P. J. Grabowski, '*In Vitro* splicing of the ribosomal RNA precursor of tetrahymena: involvement of a guanosine nucleotide in the excision of the intervening sequence', *Cell* **1981**, *27*, 487–496.

[9] C. Guerrier-Takada, K. Gardiner, T. Marsh, N. Pace, S. Altman, 'The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme', *Cell* **1983**, *35*, 849–857.

[10] S. Yokoyama, T. Muramatsu, N. Horie, K. Nishikawa, A. Matsuda, T. Ueda, Z. Yamaizumi, Y. Kuchino, S. Nishimura, T. Miyazawa, 'A novel lysine-substituted nucleoside in the first position of the anticodon of minor isoleucine tRNA from *Escherichia coli*', *Pure Appl. Chem.* **1989**, *61*, 573–576.

[11] B. J. Kopina, C. T. Lauhon, 'Efficient Preparation of 2,4-Diaminopyrimidine Nucleosides: Total Synthesis of Lysidine and Agmatidine', *Org. Lett.* **2012**, *14*, 4118–4121.

[12] D. Mandal, C. Köhrer, D. Su, S. P. Russell, K. Krivos, C. M. Castleberry, P. Blum, P. A. Limbach, D. Söll, U. L. RajBhandary, 'Agmatidine, a modified cytidine in the anticodon of archaeal $tRNA^{Ile}$, base pairs with adenosine but not with guanosine', *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 2872–2877.

[13] Y. Ikeuchi, S. Kimura, T. Numata, D. Nakamura, T. Yokogawa, T. Ogata, T. Wada, T. Suzuki, T. Suzuki, 'Agmatine-conjugated cytidine in a tRNA anticodon is essential for AUA decoding in archaea', *Nat. Chem. Biol.* **2010**, *6*, 277–282.

[14] S. V. Melnikov, N. F. Khabibullina, E. Mairhofer, O. Vargas-Rodrigues, N. M. Reynolds, R. Micura, D. Söll, Y. S. Polikanov, 'Mechanistic insights into the slow peptide bond formation with D-amino acids in the ribosomal active site', *Nucleic Acids Res.* **2019**, *47*, 2089–2100.

[15] S. Sothiselvam, S. Neuner, L. Rigger, D. Klepacki, R. Micura, N. Vázquez-Laslop, A. S. Mankin, 'Binding of Macrolide Antibiotics Leads to Ribosomal Selection against Specific Substrates Based on Their Charge and Size', *Cell Rep.* **2016**, *16*, 1789–1799.

[16] R. Shapiro, 'Prebiotic cytosine synthesis: A critical analysis and implications for the origin of life', *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 4396–4401.

[17] H. Lönnberg, R. Käppi, 'Competition between the hydrolysis and deamination of cytidine and its 5-substituted derivatives in aqueous acid', *Nucleic Acids Res.* **1985**, *13*, 2451–2456.

[18] R. M. Voorhees, D. Mandal, C. Neubauer, C. Köhrer, U. L. RajBhandary, V. Ramakrishnan, 'The structural basis for specific decoding of AUA by isoleucine tRNA on the ribosome', *Nat. Struct. Mol. Biol.* **2013**, *20*, 641–643.

[19] A. Soma, Y. Ikeuchi, S. Kanemasa, K. Kobayashi, N. Ogasawara, T. Ote, J. Kato, K. Watanabe, Y. Sekine, T. Suzuki, 'An RNA-Modifying Enzyme that Governs Both the Codon and Amino Acid Specificities of Isoleucine tRNA', *Mol. Cell* **2003**, *12*, 689–698.

[20] Y. Ikeuchi, A. Soma, T. Ote, J. Kato, Y. Sekine, T. Suzuki, 'Molecular Mechanism of Lysidine Synthesis that Determines tRNA Identity and Codon Recognition', *Mol. Cell* **2005**, *19*, 235–246.

[21] C. Switzer, S. E. Moroney, S. A. Benner, 'Enzymatic incorporation of a new base pair into DNA and RNA', *J. Am. Chem. Soc.* **1989**, *111*, 8322–8323.

[22] J. A. Piccirilli, S. A. Benner, T. Krauch, S. E. Moroney, 'Enzymatic incorporation of a new base into DNA and RNA extends the genetic alphabet', *Nature* **1990**, *343*, 33–37.

[23] T. Suzuki, K. Miyauchi, 'Discovery and characterization of $tRNA^{Ile}$ lysidine synthetase (TilS)', *FEBS Lett.* **2010**, *584*, 272–277.

[24] T. Muramatsu, K. Nishikawa, F. Nemoto, Y. Kuchino, S. Nishimura, T. Miyazawa, S. Yokoyama, 'Codon and amino-acid specificities of a transfer RNA are both converted by a single post-transcriptional modification', *Nature* **1988**, *336*, 179–181.

[25] J. Kimura, K. Yagi, H. Suzuki, O. Mitsunobu, 'Studies on Nucleosides and Nucleotides. VIII. Preparation and Reactions of Triphenylphosphoranediylnucleosides', *Bull. Chem. Soc. Jpn.* **1980**, *53*, 3670–3677.

[26] P. T. Gilham, H. G. Khorana, 'Studies on Polynucleotides. A New and General Method for the Chemical Synthesis of the C_5''-C_3'' Internucleotidic Linkage. Synthesis of Deoxyribo-dinucleotides', *J. Am. Chem. Soc.* **1958**, *80*, 6212–6222.

[27] M. Smith, D. H. Rammler, I. H. Goldberg, H. G. Khorana, 'Studies on Polynucleotides. XIV. Specific Synthesis of the C_3''-C_5'' Interribonucleotide Linkage. Syntheses of Uridylyl-($3''\rightarrow5''$)-Uridine and Uridylyl-($3''\rightarrow5''$)-Adenosine', *J. Am. Chem. Soc.* **1962**, *84*, 430–440.

[28] K. K. Ogilvie, A. L. Schifman, C. L. Penney, 'The synthesis of oligoribonucleotides. III. The use of silyl protecting groups in nucleoside and nucleotide chemistry. VIII', *Can. J. Chem.* **1979**, *57*, 2230–2238.

[29] N. Usman, K. K. Ogilvie, M.-Y. Jiang, R. J. Cedergren, 'Automated Chemical Synthesis of Long Oligoribonucleotides Using 2′-O-Silylated Ribonucleoside 3′-*O*-Phosphoramidites on a Controlled-Pore Glass Support: Synthesis of a 43-Nucleotide Sequence Similar to the 3′-Half Molecule of an *Escherichia coli* Formylmethionine tRNA', *J. Am. Chem. Soc.* **1987**, *109*, 7845–7854.

[30] R. Kierzek, M. H. Caruthers, C. E. Longfellow, D. Swinton, D. H. Turner, S. M. Freier, 'Polymer-Supported RNA Synthesis and Its Application to Test the Nearest-Neighbor Model for Duplex Stability', *Biochemistry* **1986**, *25*, 7840–7846.

[31] I. Hirao, M. Ishikawa, K. Miura, 'Solid-phase synthesis of oligoribonucleotides', *Nucleic Acids Symp. Ser.* **1985**, *16*, 173–176.

[32] H. Tanimura, T. Fukazawa, M. Sekine, T. Hata, J. W. Efcavitch, G. Zon, 'The practical synthesis of RNA fragments in the solid phase approach', *Tetrahedron Lett.* **1988**, *29*, 577–578.

[33] H. Tanimura, M. Sekine, T. Hata, 'Chemical Synthesis of RNA Fragments Related to the C4 N Hypothesis', *Nucleosides Nucleotides* **1986**, *5*, 363–383.

[34] F. Ferreira, F. Morvan, 'Silyl protecting groups for oligonucleotide synthesis removed by a $ZnBr_2$ treatment', *Nucleosides Nucleotides Nucleic Acids* **2005**, *24*, 1009–1013.

[35] A. A.-H. Abdel-Rahman, E. B. Pedersen, C. Nielsen, 'Synthesis of 5-methylisocytidine derivatives', *Monatsh. Chem.* **1996**, *127*, 455–459.

[36] R. E. Shute, D. H. Rich, 'Synthesis and Evaluation of Novel Activated Mixed Carbonate Reagents for the Introduction of the 2-(Trimethylsilyl)ethoxycarbonyl(Teoc)-Protecting Group', *Synthesis* **1987**, *4*, 346–349.

Received January 21, 2020
Accepted February 18, 2020

