Chemical Science

ROYAL SOCIETY OF CHEMISTRY

EDGE ARTICLE

View Article Online
View Journal | View Issue

Cite this: *Chem. Sci.*, 2021, **12**, 8149

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

Novel dual methylation of cytidines in the RNA of mammals†

Ming-Yu Cheng,[a] Xue-Jiao You,[a] Jiang-Hui Ding,[a] Yi Dai,[a] Meng-Yuan Chen,[a] Bi-Feng Yuan *[ab] and Yu-Qi Feng [ab]

RNA modifications play critical roles in regulating a variety of physiological processes. Methylation is the most prevalent modification occurring in RNA. Three isomeric cytidine methylation modifications have been reported in RNA, including 3-methylcytidine (m^3C), *N*4-methylcytidine (m^4C), and 5-methylcytidine (m^5C), in mammals. Aside from the single methylation on the nucleobase of cytidines, dual methylation modifications occurring in both the 2′ hydroxyl of ribose and the nucleobase of cytidines also have been reported, including *N*4,2′-*O*-dimethylcytidine (m^4Cm) and 5,2′-*O*-dimethylcytidine (m^5Cm). m^4Cm has been found in the 16S rRNA of *E. coli*, while m^5Cm has been found in the tRNA of terminal thermophilic archaea and mammals. However, unlike m^4Cm and m^5Cm, the presumed dual methylation of 3,2′-*O*-dimethylcytidine (m^3Cm) has never been discovered in living organisms. Thus, the presence of m^3Cm in RNA remains an open question. In the current study, we synthesized m^3Cm and established a liquid chromatography-electrospray ionization-tandem mass spectrometry (LC-ESI-MS/MS) method to determine the dimethylation of cytidines, m^3Cm, m^4Cm and m^5Cm. Under optimized analytical conditions, m^3Cm, m^4Cm and m^5Cm can be clearly distinguished. Using the method, we discovered the existence of m^3Cm in the RNA of mammals. The identified m^3Cm is a novel modification that hasn't been reported in the three-domain system, including archaea, bacteria, and eukaryotes. We confirmed that m^3Cm mainly existed in the small RNA (<200 nt) of mammals. In addition, we identified, for the first time, the presence of m^4Cm in the 18S rRNA of mammalian cells. The stable isotope tracing monitored by mass spectrometry demonstrated that *S*-adenosyl-L-methionine was a methyl donor for all three dimethylations of cytidines in RNA. The discovery of m^3Cm broadens the diversity of RNA modifications in living organisms. In addition, the discovery of m^3Cm and m^4Cm in mammals opens new directions in understanding RNA modification-mediated RNA processing and gene expression regulation.

Received 8th April 2021
Accepted 7th May 2021

DOI: 10.1039/d1sc01972d

rsc.li/chemical-science

Introduction

DNA cytosine methylation (5-methylcytosine, 5 mC) is the most important epigenetic modification that can regulate gene expression in genomes of higher eukaryotes.[1–3] Similar to the 5 mC on DNA, RNA modifications recently were also discovered to have critical roles in regulating a variety of physiological processes.[4,5] Up to date, over 150 structurally distinct modified nucleosides have been identified in various RNA species in all three kingdoms of life.[6]

Methylation is the most prevalent modification occurring in RNA.[6] Three isomeric cytidine methylation modifications have been reported in RNA, including 3-methylcytidine (m^3C), *N*4-methylcytidine (m^4C), and 5-methylcytidine (m^5C).[6] m^3C was initially discovered in yeast RNA and later on was identified to be present in both the rRNA and tRNA of eukaryotes.[7] Recently, m^3C has also been found in the mRNA of mammals. m^3C carries positive charges and could affect the secondary structure of RNA, change protein–RNA interactions and affect translation.[8,9] m^4C is mainly located in the decoding center of mammalian mitochondrial 12S rRNA and is critical for mitochondrial protein synthesis and respiratory function.[10–12] m^5C has long been known to be present in rRNA, mRNA, tRNA and snRNA in eukaryotes.[13,14] It was reported that the distribution and level of m^5C in mRNA were critical for mRNA stability and protein translation.[15,16] m^5C in rRNA has been recognized to be involved in ribosome biogenesis[17] and translational regulation.[18]

Aside from the single methylation on the nucleobase of cytidines (m^3C, m^4C and m^5C), some dual methylation modifications occurring in both the 2′ hydroxyl of ribose and the nucleobase of cytidines also have been reported. For example, 5,2′-*O*-dimethylcytidine (m^5Cm) has been found in the tRNA of terminal thermophilic archaea and mammalian cells.[19,20] *N*4,2′-

[a]*Sauvage Center for Molecular Sciences, Department of Chemistry, Wuhan University, Wuhan 430072, China. E-mail: bfyuan@whu.edu.cn*

[b]*School of Health Sciences, Wuhan University, Wuhan 430071, China*

† Electronic supplementary information (ESI) available: Evaluation of the purity of isolated 18S rRNA by real-time quantitative PCR; enzymatic digestion of RNA; Tables S1–S6; Fig. S1–S6. See DOI: 10.1039/d1sc01972d

O-Dimethylcytidine (m^4Cm) has been found at position 1402 of the 16S rRNA of *E. coli* and participates in the fine-tuning of the local structure of the P site and the correct identification of the start codon, thus playing an important role in modulating the fidelity of decoding.[21,22] However, unlike m^4Cm and m^5Cm, their counterpart of the dual methylation of 3,2′-*O*-dimethylcytidine (m^3Cm) has never been discovered in living organisms in the three-domain system, including archaea, bacteria, and eukaryotes.[6] Thus, the presence of m^3Cm in RNA remains an open question.

RNA modifications generally have low *in vivo* abundance. Uncovering the presence and potential roles of dimethylated cytidines needs appropriate qualitative and quantitative methods. It was reported that the level of m^5Cm in the total RNA of mammalian cells was approximately several modifications per 10^6 nucleosides.[23] Although m^4Cm was identified to be present in the 16S rRNA of *E. coli*, it hasn't been reported in the RNA of mammals. Moreover, m^3Cm, m^4Cm and m^5Cm are structurally isomers and have exactly the same molecular weights, which causes the distinct determination of these dimethylated cytidines a challenging task.

Some methods have been developed for determining RNA modifications, such as two-dimensional cellulose thin layer chromatography (2D-TLC),[24] liquid chromatography (LC) coupled with ultraviolet[25] or mass spectrometry (MS) detection.[26–29] Capillary electrophoresis with ultraviolet[30] or MS detection[31,32] also has been frequently employed in detecting RNA modifications. Because of the high detection sensitivity and good capability for deciphering structures, LC-MS recently has become a more widely used platform for analyzing low-abundant nucleic acid modifications.[26,33,34]

In the current study, we established a liquid chromatography-electrospray ionization-tandem mass spectrometry (LC-ESI-MS/MS) method to determine the dimethylation of cytidines, m^3Cm, m^4Cm and m^5Cm. Under optimized analytical conditions, m^3Cm, m^4Cm and m^5Cm can be clearly distinguished. Using the method, we identified, for the first time, the existence of m^3Cm in small RNA (<200 nt) and m^4Cm in the 18S rRNA of mammalian cells. Furthermore, stable isotope tracing experiments indicated that *S*-adenosyl-L-methionine (SAM) was a methyl donor for all three dimethylations of cytidines in RNA.

Experimental section

Chemicals and reagents

The nucleosides and modified nucleosides, including cytidine (C), adenosine (A), uridine (U), guanosine (G), N^4,2′-*O*-dimethylcytidine (m^4Cm), 5,2′-*O*-dimethylcytidine (m^5Cm), 2′-*O*-methylcytidine (Cm), and cytidine-$^{13}C_5$ (rC-$^{13}C_5$), were purchased from various commercial sources. The detailed information (CAS numbers, molecular formulas, and molecular weights) of these nucleoside standards can be found in Table S1 in the ESI.† 3,2′-*O*-dimethylcytidine (m^3Cm) was synthesized in the current study. The chemical structures of m^3Cm, m^4Cm, and m^5Cm are shown in Fig. 1.

CIAP (calf intestinal alkaline phosphatase) and S1 nuclease were obtained from Takara Biotechnology (Dalian, China).

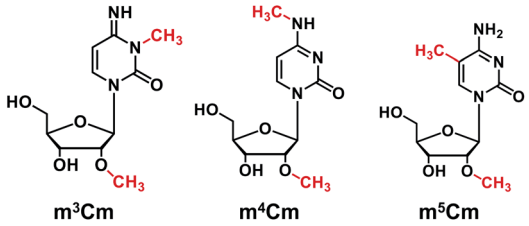

Fig. 1 Chemical structures of m^3Cm, m^4Cm and m^5Cm.

RPMI-1640 medium, Dulbecco's Modified Eagle medium (DMEM) and fetal bovine serum were obtained from Thermo Fisher Scientific (Beijing, China). L-Methionine-(methyl-D_3) (D_3-Met) was purchased from Aladdin Industrial Inc. (Shanghai, China). Dimethyl sulfate (DMS) was purchased from Macklin Biochemical Technology Co., Ltd (Shanghai, China). LCMS grade methanol (MeOH) was purchased from FTSCI Co., Ltd (Wuhan, China). Analytical grade formic acid (FA), ammonium bicarbonate and ammonia hydroxide were purchased from Sinopharm Chemical Reagent Co., Ltd (Shanghai, China).

Synthesis and characterization of 3,2′-*O*-dimethylcytidine

m^3Cm was synthesized from 2′-*O*-methylcytidine (Cm) according to a previous report.[35] The synthesis route is shown in Fig. 2A. Briefly, 2 mg of Cm in 800 μL of NaAc–HAc buffer (pH 4.3) reacted with 200 μL of DMS at 37 °C for 1 h. Then 1 mL of trichloromethane was added to the mixture to extract DMS. The generated m^3Cm was separated and purified by HPLC. A self-made C18-T reversed-phase column (10.0 × 250 mm) was used for the separation. The mobile phases consisted of H_2O (solvent A) and MeOH (solvent B). An isocratic gradient of 95% A and 5% B was utilized at a flow rate of 2.0 mL min^{-1}. The synthesized m^3Cm was characterized by high-resolution mass spectrometry and nuclear magnetic resonance (NMR) analysis.

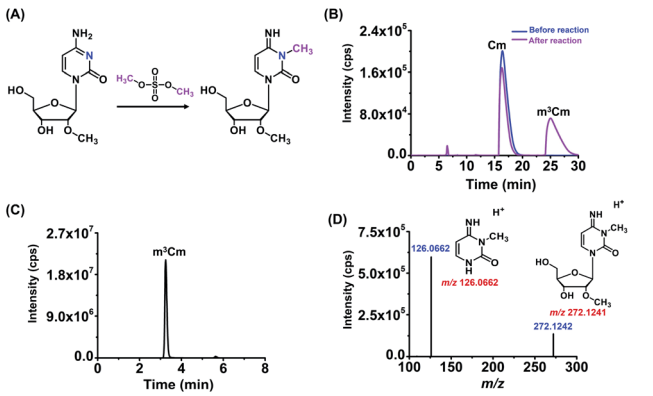

Fig. 2 Synthesis and characterization of m^3Cm by high-resolution mass spectrometry. (A) The chemical reaction for the synthesis of m^3Cm. (B) Chromatogram for the separation and purification of m^3Cm. (C) The extracted-ion chromatogram of the synthesized m^3Cm. (D) MS/MS spectrum of the synthesized m^3Cm. Shown in red are theoretical *m/z*; shown in blue are measured *m/z*.

Biological samples

The human cervical carcinoma cell line (HeLa), human hepatic cell line (HL7702), human liver carcinoma cell line (HepG2) and human breast cancer cell line (MCF-7) were obtained from the China Center for Type Culture Collection. HeLa, HepG2 and MCF-7 cells were grown in DMEM medium, and HL-7702 cells were cultured in RPMI-1640 medium at 37 °C in a 5% CO_2 atmosphere. The DMEM and RPMI-1640 media were supplemented with 10% fetal bovine serum, 100 U mL^{-1} penicillin, and 100 μg mL^{-1} streptomycin.

Purification of different RNA species

Total RNA was extracted using TRIzol reagent (Sigma Aldrich, Beijing, China) according to the manufacturer's recommended procedure. Small RNA (<200 nt) was purified using an E.Z.N.A. MiRNA kit (Omega Bio-Tek Inc., Norcross, GA, USA). For the separation of 18S rRNA and 28S rRNA, the obtained total RNA was further processed by agarose gel electrophoresis-based purification to remove the potential trace level of contamination of the 16S rRNA of *E. coli*. 18S rRNA and 28S rRNA were separated and purified with 1.0% agarose gel (low melting point). The agarose gel was stained with GelRed (Invitrogen) and visualized by using a gel documentation system (Tanon, Shanghai, China). 18S rRNA and 28S rRNA were excised from the agarose gel and recovered using a Zymoclean Gel RNA Recovery kit (Zymo Research). The total RNA of *E. coli* was extracted using a Bacteria Total RNA Isolation kit (Sangon Biotech, Shanghai, China). The 16S rRNA of *E. coli* was separated and purified with 1.0% agarose gel (low melting point) in a similar way to that for the purification of the 18S rRNA of mammalian cells.

Evaluation of the purity of 18S rRNA

The purity of the isolated 18S rRNA from mammalian cells was evaluated by real-time quantitative PCR. The reverse transcription of 18S rRNA from mammalian cells and 16S rRNA of *E. coli* was carried out using a PrimeScript™ RT reagent kit with gDNA Eraser (Takara Biotechnology) according to the manufacturer's recommended procedure.

The real-time quantitative PCR mixture includes 12.5 μL of TB Green Premixe Ex Taq (Takara Biotechnology), 10 μM forward primer (1 μL), 10 μM reverse primer (1 μL), 8.5 μL of H_2O, and 2 μL of cDNA product. The PCR amplification was performed on a CFX Connect real-time system (Bio-Rad Laboratories, Hercules, USA). The program includes denaturation at 95 °C for 10 s, annealing at 57 °C for 30 s, and elongation at 72 °C for 60 s for 40 cycles. The primer sequences include 16S rRNA forward primer (5′-TCAAATGAATTGACGGGGGC-3′), 16S rRNA reverse primer (5′-AGGCCCGGGAACGTATTCAC-3′), 18S rRNA forward primer (5′-GTAACCCGTTGAACCCCATT-3′), and 18S rRNA reverse primer (5′-CCATCCAATCGGTAGTAGCG-3′).

Enzymatic digestion of RNA

RNA was digested by the neutral enzymatic digestion method according to a previous study.[36] The detailed procedure can be found in the ESI.†

LC-ESI-MS/MS analysis

The LC-ESI-MS/MS analysis was performed on a Shimadzu 8045 mass spectrometer (Kyoto, Japan) equipped with an electrospray ionization (ESI) source (Turbo Ionspray) and a Shimadzu LC-30AD UPLC system (Tokyo, Japan). The chromatographic conditions were optimized to achieve the baseline separation of m^3Cm, m^4Cm and m^5Cm (detailed optimization procedure can be found in the ESI†). The native nucleosides and dimethylated cytidines were detected by multiple reaction monitoring (MRM) under positive mode. The MRM mass spectrometric parameters were optimized, and the optimized parameters are listed in Table S2 in the ESI.†

High-resolution mass spectrometry analysis

The dimethylated cytidines were enriched using a LC20AT HPLC system. A Hisep C18-T reversed-phase column (5 μm, 4.6 × 250 mm, Weltech Co., Ltd, Wuhan, China) was employed for the separation. 0.05% FA/H_2O (pH 2.8, solvent A) and MeOH (solvent B) were used as the mobile phases. Gradients of 0–5 min, 5% B; 5–40 min, 5–45% B; 40–45 min, 45% B; 45–47 min, 45–5% B; 47–55 min, 5% B were used. The flow rate was set at 0.8 mL min^{-1}.

The HPLC-enriched dimethylated cytidines were characterized by using a high-resolution LTQ-Orbitrap Elite mass spectrometer (Thermo Fisher Scientific, Waltham, MA, USA), which was equipped with an ESI source and a Dionex Ultimate 3000 UPLC system (Thermo Fisher Scientific, Waltham, MA, USA). The MS analysis was performed in positive-ion mode with full scan detection (m/z 70–350) at a resolution of 60 000. The collision energy of collision induced dissociation (CID) was 15 eV. The source and ion transfer parameters applied were as follows: capillary temperature, 350 °C; heater temperature, 300 °C; auxiliary gas flow, 15 arbitrary units; sheath gas flow, 35 arbitrary units; capillary voltage, 35 V; spray voltage, 3.5 kV; the S-lens RF level, 60%. Data analysis was performed using Xcalibur v3.0.63 (Thermo Fisher Scientific, Waltham, MA, USA).

Stable isotope labelling tracing

As for stable isotope tracing experiments by mass spectrometry, HeLa cells were cultured in DMEM medium which was supplemented with 0.3 mM of D_3-Met to label RNA with the CD_3 group. HeLa cells were harvested after culturing for 72 h. Small RNA (<200 nt) and 18S rRNA were extracted followed by enzymatical digestion and LC-ESI-MS/MS analysis.

Results and discussion

Characterization of the synthesized m^3Cm

In this study, we aimed to investigate the existence status of three dimethylations of cytidines, including m^3Cm, m^4Cm and m^5Cm, in mammals. m^3Cm has never been found in living organisms in the three-domain system, including archaea, bacteria, and eukaryotes. m^4Cm was previously reported to exist in the 16S rRNA of *E. coli*, but hasn't been reported to be present in mammalian cells.[21] m^5Cm exists in the tRNA of eukaryotes.[19] Here, we systematically investigate the presence of three

dimethylated cytidines of m^3Cm, m^4Cm and m^5Cm in mammalian cells.

Because there haven't been any previous reports regarding m^3Cm and m^3Cm standard is not commercially available, it is essential to synthesize an m^3Cm standard for the confident determination of its potential presence in living organisms. The chemical reaction for the synthesis of m^3Cm is shown in Fig. 2A. Cytidine can react with DMS to produce different products under different pH conditions. We found that the methylation mainly occurred at the *N*3 position of the nucleobase at pH 4.3; but at pH 8.0–10.0, the methylation occurred at both the *N*3 and *N*4 positions of the nucleobase. Thus, the synthesis of m^3Cm was carried out in NaAc–HAc buffer under pH 4.3.

HPLC was employed to separate and purify m^3Cm. In addition to the substrate of Cm (16.4 min), we observed a new peak that occurred at 25.0 min after the chemical reaction (Fig. 2B). The new peak was then collected and analyzed by high-resolution mass spectrometry. The results showed that the precursor ion (*m*/*z* 272.1242) and fragment ions (*m*/*z* 126.0662) of the synthesized product were identical to the theoretical *m*/*z* of m^3Cm (*m*/*z* 272.1241 and 126.0662, Fig. 2C and D), indicating that the compound should be the desired m^3Cm. The chromatographic retention time of this synthesized compound (3.2 min) is different from those of the m^4Cm standard (2.7 min) and m^5Cm standard (3.5 min) (see the next section). In addition, NMR analysis further confirmed the synthesized m^3Cm (Fig. S1 and S2 in the ESI†). Collectively, these results demonstrated that the desired compound of m^3Cm was successfully synthesized.

Separation of the isomers of m^3Cm, m^4Cm and m^5Cm

m^3Cm, m^4Cm and m^5Cm are structurally isomers and have exactly the same molecular weights. Their MRM transitions are also the same during LC-ESI-MS/MS analysis (Table S2 in the ESI†). Thus, complete chromatographic separation of m^3Cm, m^4Cm and m^5Cm is essential and important to distinguish these isomers. In this respect, we optimized the separation conditions for these isomers, including the mobile phases and separation gradients.

We first used 0.05% FA in H_2O with different pH (2.8, 3.0, 4.0 and 5.0) as solvent A and MeOH as solvent B for the chromatographic separation. The gradient 1 was first employed for the separation at a flow rate of 0.3 mL min^{-1}. The results showed that 0.05% FA in H_2O (pH 2.8) offered the best separation for m^3Cm, m^4Cm and m^5Cm (Fig. S3A–D in the ESI†). In addition, we also examined 2 mM NH_4HCO_3 as solvent A and MeOH as solvent B for the separation. However, m^3Cm and m^4Cm cannot be separated under these conditions (Fig. S3E in the ESI†). Thus, 0.05% FA in H_2O (pH 2.8) (solvent A) and MeOH (solvent B) were employed as the mobile phases. We next evaluated the gradients in the separation of m^3Cm, m^4Cm and m^5Cm. The results showed that gradient 2 offered better separation resolution towards m^3Cm, m^4Cm and m^5Cm than gradient 1 (comparing Fig. S3A and S3F in the ESI†). In addition, we also examined the effect of flow rate on the separation of these isomers. The results demonstrated that the peaks of these isomers were narrower at a flow rate of 0.3 mL min^{-1} than those at 0.2 mL min^{-1} (comparing Fig. S3F and S3G in the ESI†).

Collectively, the optimal chromatographic separation conditions for m^3Cm, m^4Cm and m^5Cm were 0.05% FA in H_2O (pH 2.8) (solvent A) and MeOH (solvent B) as the mobile phases with using the gradient 2 at a flow rate of 0.3 mL min^{-1}. Under the optimized separation conditions, m^3Cm, m^4Cm and m^5Cm can be well separated (Fig. 3A), which provides a good fundamental to readily determine these isomers.

Determination of m^3Cm, m^4Cm and m^5Cm in the RNA of mammalian cells

With the m^3Cm, m^4Cm and m^5Cm standards, we next detected these dimethylated cytidines in mammalian RNA. We initially analyzed these dimethylated cytidines in the total RNA of HeLa cells and *E. coli.* The preliminary results showed that all three dimethylated cytidines can be detected in the total RNA of HeLa cells (Fig. 3B); however, only m^4Cm was detected in *E. coli* (Fig. 3C). It should be noted that in the sample with only adding enzymes and omitting the RNA sample, we observed tiny peaks whose retention times were similar to those of m^3Cm, m^4Cm and m^5Cm standards (Fig. 3D), indicating that the enzymes might contain trace levels of these modifications. However, the intensities of these peaks were much lower than those detected from total RNA, which should not affect the determination of these dimethylated cytidines.

We next detected m^3Cm, m^4Cm and m^5Cm in different RNA species from HeLa cells. The overall results showed that m^3Cm and m^5Cm were mainly present in small RNA (<200 nt), while m^4Cm was mainly present in 18S rRNA. m^4Cm has been reported to be present in the 16S rRNA of *E. coli*.[21] In this study, we aimed to investigate the presence of m^3Cm, m^4Cm and m^5Cm

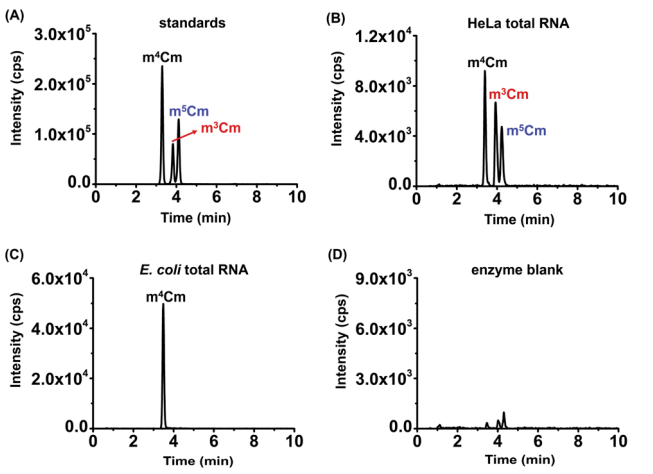

Fig. 3 Examination of m^3Cm, m^4Cm and m^5Cm in the total RNA of HeLa cells and *E. coli*. (A) The extracted-ion chromatograms of m^3Cm, m^4Cm and m^5Cm standards under the optimized separation conditions. (B) The extracted-ion chromatograms of m^3Cm, m^4Cm and m^5Cm from the total RNA of HeLa cells. (C) The extracted-ion chromatograms of m^4Cm from the total RNA of *E. coli*. (D) The extracted-ion chromatograms of m^3Cm, m^4Cm and m^5Cm from an enzyme blank.

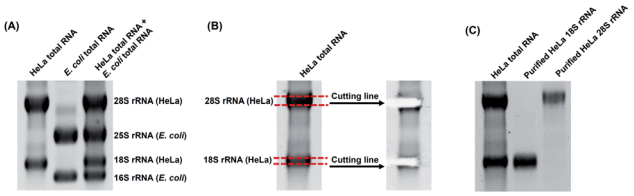

Fig. 4 Purification of the 18S rRNA and 28S rRNA of HeLa cells by agarose gel electrophoresis. (A) Separation of total RNAs from HeLa cells and *E. coli*. (B) Purification of the 18S rRNA and 28S rRNA of HeLa cells. (C) Analysis of the gel-purified 18S rRNA and 28S rRNA of HeLa cells.

in mammalian cells. It is possible that the cultured mammalian cells contain trace levels of bacterial contamination. In this respect, it is critical to obtain highly pure rRNA for the determination of m^4Cm in mammalian cells. To exclude the potential contamination of the rRNA of *E. coli* in the isolated rRNA of mammalian cells, we employed agarose gel electrophoresis to separate and purify the 18S and 28S rRNA. The results showed that both the 18S rRNA and 28S rRNA of HeLa cells can be well differentiated from the 16S rRNA and 25S rRNA of *E. coli* (Fig. 4A). The bands of the 18S rRNA and 28S rRNA of HeLa cells were then cut and purified (Fig. 4B and C). The real-time quantitative PCR results showed that the purity of the isolated 18S rRNA of HeLa cells was higher that 99.999% (Fig. S4 in the ESI†).

The retention times of two compounds detected in the small RNA (<200 nt) of HeLa cells were similar to that of m^3Cm and m^5Cm standards (Fig. 5A, B, D and E); the retention time of one compound detected in the 18S rRNA of HeLa cells was similar to that of the m^4Cm standard (Fig. 5C and F), indicating the existence of these dimethylated cytidines in different RNA species. Furthermore, we separately added the standards of m^3Cm, m^4Cm and m^5Cm to the digested nucleosides from the small RNA (<200 nt) or 18S rRNA of HeLa cells. It can be seen that the spiked standards had the same retention times as those detected in HeLa cells and increased the peak intensities (Fig. 5G–I), suggesting that the detected compounds should be m^3Cm, m^4Cm and m^5Cm.

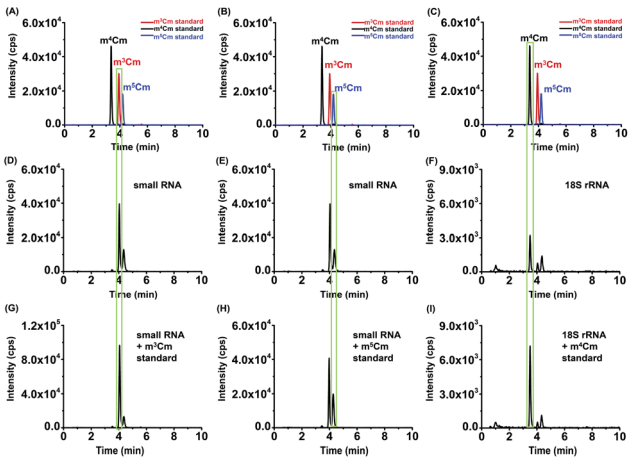

Fig. 5 Determination of m^3Cm, m^4Cm and m^5Cm in different RNA species of HeLa cells. (A)–(C) The extracted-ion chromatograms of m^3Cm, m^4Cm and m^5Cm standards. (D)–(E) The extracted-ion chromatograms of m^3Cm and m^5Cm from the small RNA (<200 nt) of HeLa cells. (F) The extracted-ion chromatograms of m^4Cm from the 18S rRNA of HeLa cells. (G) The extracted-ion chromatograms of m^3Cm from the small RNA of HeLa cells with a spiked m^3Cm standard. (H) The extracted-ion chromatograms of m^5Cm from the small RNA (<200 nt) of HeLa cells with a spiked m^5Cm standard. (I) The extracted-ion chromatograms of m^4Cm from the 18S rRNA of HeLa cells with a spiked m^4Cm standard.

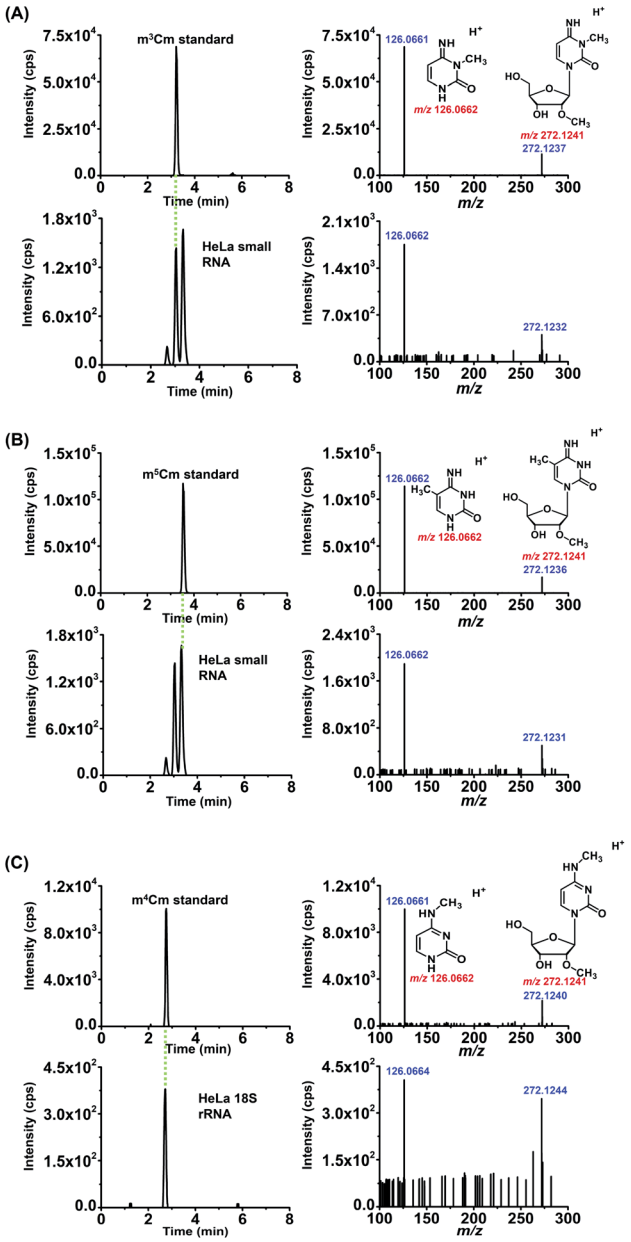

Fig. 6 Identification of m^3Cm, m^4Cm and m^5Cm in the RNA of HeLa cells by high-resolution mass spectrometry. (A) The extracted-ion chromatograms and product ion spectra from the m^3Cm standard (above) and small RNA (<200 nt) of HeLa cells (below). (B) The extracted-ion chromatograms and product ion spectra from the m^5Cm standard (above) and small RNA (<200 nt) of HeLa cells (below). (C) The extracted-ion chromatograms and product ion spectra from the m^4Cm standard (above) and 18S rRNA of HeLa cells (below).

We also confirmed these dimethylated cytidines from HeLa cells by high-resolution MS analysis. The enzymatically digested cytidine modifications from the small RNA (<200 nt) or 18S rRNA of HeLa cells were first enriched by offline HPLC before high-resolution MS analysis (Fig. S5 in the ESI†). The MS analysis showed that the precursor ions and product ions (*m*/*z* shown in blue) of the detected compounds in the RNA of HeLa cells were identical to their corresponding theoretical values (*m*/*z* shown in red) as well as to those of the standards (Fig. 6), further confirming the detected compounds of m^3Cm, m^4Cm and m^5Cm. Taken together, the results suggested that m^3Cm and m^5Cm were present in the small RNA (<200 nt) of mammals and m^4Cm was present in the 18S rRNA of mammals.

Metabolic labeling of cells

To further confirm the presence of m^3Cm, m^4Cm and m^5Cm in the RNA of mammals, we conducted stable isotope tracing experiments by mass spectrometry. ATP and L-methionine (Met) could be converted into *S*-adenosyl-L-methionine (SAM) by methionine adenosyltransferase and SAM is a methyl group donor for the methylation of nucleic acids.[37] If these dimethylated cytidines exist in mammalian cells, then the culturing of cells in DMEM medium containing isotopically labeled L-methionine (D_3-Met) would lead to the transfer of the CD_3 group from D_3-Met to m^3Cm, m^4Cm and m^5Cm. Theoretically, single and dual CD_3 might be added to these dimethylated cytidines.

As for m^3Cm, three compounds, D_3-m^3Cm (CD_3 labeled nucleobase), m^3Cm-D_3 (CD_3 labeled ribose), D_3-m^3Cm-D_3 (CD_3 labeled both nucleobase and ribose), can be theoretically formed in the RNA of cells that were cultured in a medium containing D_3-Met (Fig. S6 in the ESI†). We cultured human HeLa cells in DMEM medium supplied with 0.3 mM D_3-Met. Then 18S rRNA and small RNA (<200 nt) were isolated and analyzed by our developed method. The results showed that, in addition to m^3Cm, all three compounds, D_3-m^3Cm, m^3Cm-D_3 and D_3-m^3Cm-D_3, were distinctly detected (Fig. 7). The concentration of unlabeled methionine in the DMEM medium is 0.2 mM. Thus, the theoretical percentage of D_3-Met in the medium is approximately 60% of the total methionine. It can be seen that approximately 60% of the measured m^3Cm carried the CD_3 nucleobase group and CD_3 ribose group (Table S3 in the ESI†), which is equivalent to the theoretical percentage of D_3-Met in total methionine in the medium. Similar to m^3Cm, CD_3 labeled m^4Cm and m^5Cm were both clearly detected (Fig. 7). Collectively, the results further confirmed the presence of these three dimethylated cytidines in mammalian cells and indicated that SAM was the methyl donor for the dual methylation of cytidines in RNA.

Quantification of m^3Cm, m^4Cm and m^5Cm in RNA

Calibration curves were constructed to quantitatively measure the dimethylated cytidines in RNA. A mixture of cytidine standards with a series of amounts ranging from 0.05 nmol to 5 nmol (for cytidine) or from 0.5 fmol to 500 fmol (for m^3Cm, m^4Cm, and m^5Cm) and fixed amounts of isotope internal standards (rC-13C[5]) were prepared. The calibration curves were

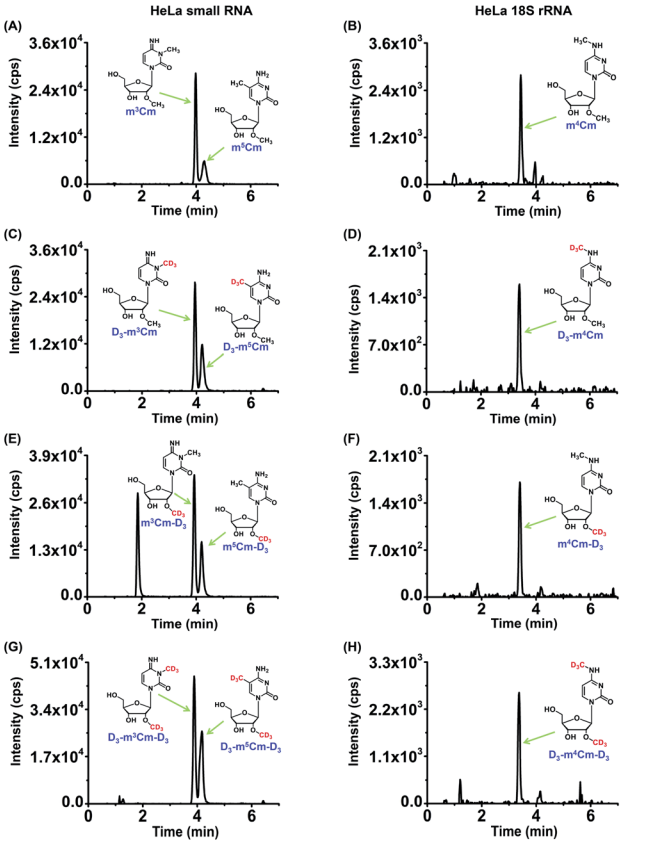

Fig. 7 Stable isotope tracing experiments. (A) and (B) The extracted-ion chromatograms of dimethylated cytidines from the small RNA (<200 nt) and 18S rRNA of HeLa cells (MRM: 272.1 → 126.1). (C) and (D) The extracted-ion chromatograms of dimethylated cytidines that carried the CD_3 nucleobase group from the small RNA (<200 nt) and 18S rRNA of HeLa cells (MRM: 275.1 → 129.1). (E) and (F) The extracted-ion chromatograms of dimethylated cytidines that carried the CD_3 ribose group from the small RNA (<200 nt) and 18S rRNA of HeLa cells (MRM: 275.1 → 126.1). (G) and (H) The extracted-ion chromatograms of dimethylated cytidines that carried the CD_3 nucleobase group and the CD_3 ribose group from the small RNA (<200 nt) and 18S rRNA of HeLa cells (MRM: 278.1 → 129.1).

constructed by plotting the peak area ratios (analytes/IS) against the amounts of analytes with triplicate measurements. The results showed that good linearities were obtained with the coefficients of determination (R^2) being greater than 0.99 (Table S4 in the ESI†). The accuracy and reproducibility of the method were evaluated with the relative errors (REs) and intra- and inter-day relative standard deviations (RSDs) being less than 13.7% and 14.9%, respectively (Tables S4 and S5 in the ESI†), demonstrating that good accuracy and reproducibility were achieved.

In addition to HeLa cells, these modifications of m^3Cm, m^4Cm and m^5Cm were also detected and quantified in other human cells, including human HL-7702 hepatic cells, human HepG2 hepatocellular carcinoma and human MCF-7 breast adenocarcinoma cells. The quantification results showed that the measured m^3Cm ranged from 0.0003% to 0.0008%, m^4Cm ranged from 0.0004% to 0.001%, and m^5Cm ranged from

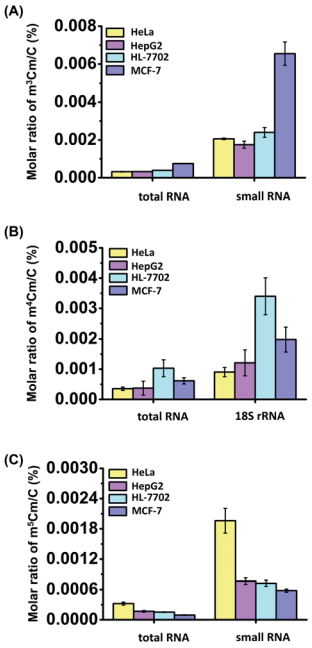

Fig. 8 Quantification of m^3Cm, m^4Cm and m^5Cm in RNA from different cell lines (HeLa, HepG2, HL-7702 and MCF-7 cells). (A) Quantification of m^3Cm in total RNA and small RNA (<200 nt). (B) Quantification of m^4Cm in total RNA and 18S rRNA. (C) Quantification of m^5Cm in total RNA and small RNA (<200 nt). The error bars represent the standard deviations of the mean from data of three independent experiments.

0.0001% to 0.0003% in total RNA (Fig. 8 and Table S6 in the ESI†). In small RNA (<200 nt), the m^3Cm ranged from 0.002% to 0.007%, and m^5Cm ranged from 0.0006% to 0.002% (Fig. 8 and Table S6 in the ESI†). The measured levels of m^5Cm in total RNA and small RNA (<200 nt) are comparable with those in previous studies.[23] In 18S rRNA, the m^4Cm ranged from 0.0009% to 0.003% (Fig. 8 and Table S6 in the ESI†).

The results clearly indicated that m^3Cm and m^5Cm are mainly present in small RNA (<200 nt) and m^4Cm is mainly present in 18S rRNA. Previous studies demonstrated that tRNA accounts for 90% of small RNA (<200 nt).[38] Therefore, the m^3Cm detected in small RNA (<200 nt) most likely comes from tRNA. The identified m^3Cm is a totally new modification that has never been found in living organisms. As for m^4Cm, it is firstly identified to be present in the RNA of mammals in the current study. Although the potential roles of m^3Cm and m^4Cm in RNA are still elusive, we speculate that m^3Cm and m^4Cm could enhance the structural stability of tRNA and 18S rRNA. It was reported that the methyltransferase-like protein 6 (METTL6) could catalyze the formation of m^3C in serine tRNA.[39] FtsJ RNA 2′-*O*-methyltransferase 1 (FTSJ1) is involved in the 2′-*O*-methylation of cytidine in the anti-codon region of tRNA.[40] Thus, we speculate that METTL6 may methylate cytidine in tRNA to form m^3C, which is then further methylated by FTSJ1 to form m^3Cm. Future study on identifying enzymes responsible for the formation and removal of the methyl groups of m^3Cm and m^4Cm will promote uncovering the functions of these dual methylations of cytidines in RNA.

Conclusions

In summary, using the developed LC-ESI-MS/MS method, we identified the existence of a novel modification of m^3Cm in the RNA of mammals. We confirmed that m^3Cm mainly exists in the small RNA (<200 nt) of mammals. The identified m^3Cm is a totally new modification that has never been found in living organisms. In addition, we identified, for the first time, the presence of m^4Cm in 18S rRNA of mammalian cells. The stable isotope tracing monitored by mass spectrometry demonstrated that SAM was the methyl donor for these dimethylations of cytidines in RNA. The discovery of m^3Cm broadens the diversity of RNA modifications in living organisms. Moreover, the discovery of m^3Cm and m^4Cm in mammals indicates the new layer of RNA epigenetic modifications. Future functional investigating of these dual methylations of cytidines in RNA will unveil the complexity in the regulation of biological processes through RNA modifications.

Author contributions

M. Y. C. and B. F. Y. designed the experiments and interpreted the data. M. Y. C., X. J. Y., and J. H. D. performed the synthesis and purification of m^3Cm and carried out the metabolic labelling and mass spectrometry analysis. Y. D. and M. Y. C. performed the cell culturing and RNA isolation. Y. Q. F. interpreted the mass spectrometry data. M. Y. C. and B. F. Y. wrote the manuscript.

Conflicts of interest

The authors declare no competing financial interests.

Acknowledgements

The work is supported by the National Natural Science Foundation of China (22074110, 21635006, and 21721005), and the Fundamental Research Funds for the Central Universities (2042021kf0212).

References

1 C. Luo, P. Hajkova and J. R. Ecker, *Science*, 2018, **361**, 1336–1340.
2 T. Liu, C. J. Ma, B. F. Yuan and Y. Q. Feng, *Sci. China: Chem.*, 2018, **61**, 381–392.
3 Y. Feng, N. B. Xie, W. B. Tao, J. H. Ding, X. J. You, C. J. Ma, X. Zhang, C. Yi, X. Zhou, B. F. Yuan and Y. Q. Feng, *CCS Chem.*, 2020, **2**, 994–1008.
4 K. Chen, B. S. Zhao and C. He, *Cell Chem. Biol.*, 2016, **23**, 74–85.
5 X. Li, X. Xiong and C. Yi, *Nat. Methods*, 2016, **14**, 23–31.
6 P. Boccaletto, M. A. Machnicka, E. Purta, P. Piatkowski, B. Baginski, T. K. Wirecki, V. de Crecy-Lagard, R. Ross, P. A. Limbach, A. Kotter, M. Helm and J. M. Bujnicki, *Nucleic Acids Res.*, 2018, **46**, D303–D307.

7 A. Noma, S. Yi, T. Katoh, Y. Takai, T. Suzuki and T. Suzuki, *RNA*, 2011, **17**, 1111–1119.
8 F. Liu and C. He, *J. Biol. Chem.*, 2017, **292**, 14704–14705.
9 C. J. Ma, J. H. Ding, T. T. Ye, B. F. Yuan and Y. Q. Feng, *ACS Chem. Biol.*, 2019, **14**, 1418–1425.
10 I. Laptev, E. Shvetsova, S. Levitskii, M. Serebryakova, M. Rubtsova, V. Zgoda, A. Bogdanov, P. Kamenski, P. Sergiev and O. Dontsova, *Nucleic Acids Res.*, 2020, **48**, 8022–8034.
11 H. Chen, Z. Shi, J. Guo, K. J. Chang, Q. Chen, C. H. Yao, M. C. Haigis and Y. Shi, *J. Biol. Chem.*, 2020, **295**, 8505–8513.
12 L. Van Haute, A. G. Hendrick, A. R. D'Souza, C. A. Powell, P. Rebelo-Guiomar, M. E. Harbour, S. Ding, I. M. Fearnley, B. Andrews and M. Minczuk, *Nucleic Acids Res.*, 2019, **47**, 10267–10281.
13 L. Trixl and A. Lusser, *Wiley Interdiscip. Rev.: RNA*, 2019, **10**, e1510.
14 L. Fu, C. R. Guerrero, N. Zhong, N. J. Amato, Y. Liu, S. Liu, Q. Cai, D. Ji, S. G. Jin, L. J. Niedernhofer, G. P. Pfeifer, G. L. Xu and Y. Wang, *J. Am. Chem. Soc.*, 2014, **136**, 11582–11585.
15 J. E. Squires, H. R. Patel, M. Nousch, T. Sibbritt, D. T. Humphreys, B. J. Parker, C. M. Suter and T. Preiss, *Nucleic Acids Res.*, 2012, **40**, 5023–5033.
16 Y. Yang, L. Wang, X. Han, W. L. Yang, M. Zhang, H. L. Ma, B. F. Sun, A. Li, J. Xia, J. Chen, J. Heng, B. Wu, Y. S. Chen, J. W. Xu, X. Yang, H. Yao, J. Sun, C. Lyu, H. L. Wang, Y. Huang, Y. P. Sun, Y. L. Zhao, A. Meng, J. Ma, F. Liu and Y. G. Yang, *Mol. Cell*, 2019, **75**, 1188–1202 e1111.
17 M. Schosserer, N. Minois, T. B. Angerer, M. Amring, H. Dellago, E. Harreither, A. Calle-Perez, A. Pircher, M. P. Gerstl, S. Pfeifenberger, C. Brandl, M. Sonntagbauer, A. Kriegner, A. Linder, A. Weinhausel, T. Mohr, M. Steiger, D. Mattanovich, M. Rinnerthaler, T. Karl, S. Sharma, K. D. Entian, M. Kos, M. Breitenbach, I. B. Wilson, N. Polacek, R. Grillari-Voglauer, L. Breitenbach-Koller and J. Grillari, *Nat. Commun.*, 2015, **6**, 6158.
18 S. Blanco and M. Frye, *Curr. Opin. Cell Biol.*, 2014, **31**, 1–7.
19 S. M. Huber, P. van Delft, A. Tanpure, E. A. Miska and S. Balasubramanian, *J. Am. Chem. Soc.*, 2017, **139**, 1766–1769.
20 C. G. Edmonds, P. F. Crain, T. Hashizume, R. Gupta, K. O. Stetter and J. A. McCloskey, *J. Chem. Soc., Chem. Commun.*, 1987, 909–910.
21 S. Kimura and T. Suzuki, *Nucleic Acids Res.*, 2010, **38**, 1341–1352.
22 J. L. Nichols and B. G. Lane, *Biochim. Biophys. Acta*, 1966, **119**, 649–651.
23 Y. Feng, C. J. Ma, J. H. Ding, C. B. Qi, X. J. Xu, B. F. Yuan and Y. Q. Feng, *Anal. Chim. Acta*, 2020, **1098**, 56–65.
24 H. Grosjean, G. Keith and L. Droogmans, *Methods Mol. Biol.*, 2004, **265**, 357–391.
25 H. M. Liebich, S. Muller-Hagedorn, M. Bacher, H. G. Scheel-Walter, X. Lu, A. Frickenschmidt, B. Kammerer, K. R. Kim and H. Gerard, *J. Chromatogr. B: Anal. Technol. Biomed. Life Sci.*, 2005, **814**, 275–283.
26 B. Chen, B. F. Yuan and Y. Q. Feng, *Anal. Chem.*, 2019, **91**, 743–756.
27 C. B. Qi, H. P. Jiang, J. Xiong, B. F. Yuan and Y. Q. Feng, *Chin. Chem. Lett.*, 2019, **30**, 553–557.
28 C. B. Qi, J. H. Ding, B. F. Yuan and Y. Q. Feng, *Chin. Chem. Lett.*, 2019, **30**, 1618–1626.
29 X. J. You, T. Liu, C. J. Ma, C. B. Qi, Y. Tong, X. Zhao, B. F. Yuan and Y. Q. Feng, *Anal. Chem.*, 2019, **91**, 10477–10483.
30 K. Petru, J. Siroka, L. Bydzovska, L. Krcmova and M. Polasek, *Electrophoresis*, 2014, **35**, 2546–2549.
31 F. Yuan, X. H. Zhang, J. Nie, H. X. Chen, Y. L. Zhou and X. X. Zhang, *Chem. Commun.*, 2016, **52**, 2698–2700.
32 Y. Yu, S. H. Zhu, F. Yuan, X. H. Zhang, Y. Y. Lu, Y. L. Zhou and X. X. Zhang, *Chem. Commun.*, 2019, **55**, 7595–7598.
33 W. Y. Lai, J. Z. Mo, J. F. Yin, C. Lyu and H. L. Wang, *TrAC, Trends Anal. Chem.*, 2019, **110**, 173–182.
34 M. D. Lan, B. F. Yuan and Y. Q. Feng, *Chin. Chem. Lett.*, 2019, **30**, 1–6.
35 H. Bredereck, H. Haas and A. Martini, *Chem. Ber.*, 1948, **81**, 307–313.
36 Q. Y. Cheng, J. Xiong, C. J. Ma, Y. Dai, J. H. Ding, F. L. Liu, B. F. Yuan and Y. Q. Feng, *Chem. Sci.*, 2020, **11**, 1878–1891.
37 B. J. Landgraf, E. L. McCarthy and S. J. Booker, *Annu. Rev. Biochem.*, 2016, **85**, 485–514.
38 D. Su, C. T. Chan, C. Gu, K. S. Lim, Y. H. Chionh, M. E. McBee, B. S. Russell, I. R. Babu, T. J. Begley and P. C. Dedon, *Nat. Protoc.*, 2014, **9**, 828–841.
39 V. V. Ignatova, S. Kaiser, J. S. Y. Ho, X. Bing, P. Stolz, Y. X. Tan, C. L. Lee, F. P. H. Gay, P. R. Lastres, R. Gerlini, B. Rathkolb, A. Aguilar-Pimentel, A. Sanz-Moreno, T. Klein-Rodewald, J. Calzada-Wack, E. Ibragimov, M. Valenta, S. Lukauskas, A. Pavesi, S. Marschall, S. Leuchtenberger, H. Fuchs, V. Gailus-Durner, M. H. de Angelis, S. Bultmann, O. J. Rando, E. Guccione, S. M. Kellner and R. Schneider, *Sci. Adv.*, 2020, **6**, eaaz4551.
40 M. P. Guy, M. Shaw, C. L. Weiner, L. Hobson, Z. Stark, K. Rose, V. M. Kalscheuer, J. Gecz and E. M. Phizicky, *Hum. Mutat.*, 2015, **36**, 1176–1187.