Oxime formation coordination-directed detection of genome-wide thymine oxides with nanogram-scale sample input†

Feng Xiao,‡a Qi Wang,‡a Kaiyuan Zhang,a Chaoxing Liu,†b Guangrong Zou*a and Xiang Zhou‡a

Natural chemical modifications of nucleic acids play a vital role in life processes. Compared to other epigenetic modifications, there are multiple ways to quantify the methylated derivatives of cytosine. However, simple and convenient methods for detecting and quantifying thymine derivatives are scarce because they are found in tiny quantities in biological systems. Additionally, exploring easy ways to detect these derivatives can also throw light on their biological significance. This manuscript reports a novel strategy to quantify 5-formyluracil (5fU) and 5-hydroxymethyluracil (5hmU). Differences between modified and unmodified bases are accumulated and amplified by arranging phi29 DNA polymerase to repeat through a circular template labeled thymidine. In combination with real-time quantitative rolling circle amplification (RCA), low-abundance thymine oxides can be quantified precisely. The global levels of 5fU and 5hmU were analyzed in different biological samples, using only 40 ng of sample input on a laboratory real-time PCR instrument. The reported strategy was executed hassle-free and, in principle, can be extended to design methods for detecting other epigenetic modifications in nucleotides that are rare in biological systems.

Introduction

In addition to oxidized cytosine derivatives, cells also harbor natural thymidine (T)-modified nucleosides 5-hydroxymethyluracil (5hmU) and 5-formyluracil (5fU).1–3 Generated by enzyme-mediated or non-enzymatic pathways, they are present in the genomic DNA of various organisms ranging from bacteriophages to mammals.4,5 As the modified nucleobase counterpart of 5-formylcytosine (5fC),6–10 5fU has garnered significant interest from researchers. The fact that some tumors have been reported to have higher levels of 5fU than adjacent normal tissue has increased the importance of these derivatives.11 5fU is mainly obtained from thymine oxidation by UV light, reactive oxygen species, hydrogen peroxide, and other oxidants.12–15 It can cause gene mismatch, miscoding, alteration of DNA structures, and interference with DNA–protein interactions.16,17 Different forms of 5fU and 5hmU are synthesized by two different pathways in biological organisms, i.e., by deamination of 5hmC by AID/APOBEC enzymes,18,19 and by oxidation of thymine by ten-eleven translocation enzymes.2 All of this suggests that 5hmU and 5fU have specific functions other than triggering DNA repair. Because of their extremely low abundance in genomes and the subtle differences in their structures, there are very few methods for detecting and quantifying them effectively and accurately.20,21 Additionally, most assays/methods cannot ensure high sensitivity as well as cost efficiency with nanogram-scale sample input.

PCR based quantitative detection methods mainly rely on the inability of modified bases and their derivatives to the polymerase-mediated amplification,22–24 but are usually performed under cyclic temperature-controlled conditions and require a long amplification time (1–2 h). Quantitative detection methods based on Liquid Chromatography with Mass Spectrometry (LC-MS) are widely accepted and highly reliable. Generally, after oligonucleotides are degraded into nucleotides, they are separated by chromatography and are then quantified by MS. Various chemical derivatization and enrichment methods have been developed to improve the sensitivity of detection.11,12,25,26 However, expensive instrumentation, high sample input (1–10 μg), and lengthy analysis time (>24 h) limit its widespread application. Compared with LC-MS-based quantitative methods, fluorescence-based quantitative methods have the advantage of being more economical and convenient.27–30 The chemical structure of 5fU contains an aldehyde group on the pyrimidine ring that can be modified with amine, hydrazine, aminoxyl, and indantrione derivatives.31–35 Besides, 5hmU can

† These authors contributed equally.
also be easily oxidized to 5fU by K RuO₄. However, 5fC and an abasic site (AP site) contain an aldehyde group in their structure, similar to 5fU. Thus, it is challenging to find an appropriate chemical reagent to selectively label and detect 5fU and 5hmU throughout the genome without any interference from 5fC and AP. In addition, traditional probes are usually in excess and have been found to bind very well to the template of DNA modiﬁed with similar slopes with increasing reaction time. This result conﬁrmed no signiﬁcant polymerization efficiencies for DNA–T, DNA–5fU, DNA–5hmU, and DNA–5fU from DNA–5hmU oxidation (Fig. 2A). DNA templates without adaptor ligation showed almost no ﬂuorescent signal (Fig. S7†). The slopes and R² values from all possible ﬁts were arranged. In the ﬂuorescence curve of a typical RCA reaction, a high correlation coefﬁcient can be obtained by collecting data in 10 min (Fig. 2B–D, and Table S2, † R² > 0.9995).

Compared with DNA–5fU templates without chemical labeling, the slope of the DNA–5fU template modiﬁed with AQA was reduced by about 67% (Fig. 3A). However, the slope of DNA–T and DNA–5hmU showed no signiﬁcant differences postreaction with AQA under the same conditions (Fig. 3B). We then attempted to get a linear correlation between 5fU content and the efﬁciency difference of polymerization reaction (ΔS) for quantitative estimation of 5fU. The correlation indicated that with the increase in DNA–5fU–AQA content (C₅fU, mixed samples with different proportions of DNA–5fU–AQA and DNA–

**Fig. 1** The OFCRCA strategy for detecting genome-wide thymine oxides. (A) Oxidation of 5hmU and AQA labeling of 5fU. (B) Rationale for signal generation by the OFCRCA strategy.

**Results and discussion**

We started by examining the selectivity and efﬁciency of reactions of AQA to a series of DNAs (15 nt–5fU, 15 nt–5fC, 15 nt–5hmU, and 15 nt–T in Table S1†). After the ethanol precipitation step, the final DNA products were analyzed by 20% denatured PAGE. Under the natural reaction conditions, only the single band in the lane of 5fU migrated slowly compared to other lanes due to an increase in molecular weight due to AQA labeling (Fig. S1†). The PAGE results show that 5fU could be labeled by AQA speciﬁcally, unlike other bases. Further, Matrix-Assisted Laser Desorption Ionization Time of Flight Mass Spectrometry (MALDI-TOF-MS) data also indicated the formation of the desired 5fU–AQA nucleotide product (Fig. S2†). Two DNA strands (15 nt–EA and 28 nt–EB) containing 5fU sites were designed, and phi29 DNA polymerase was used for chain extension experiments to verify if selective labeling of 5fU by AQA has a blocking effect on the polymerase reaction (Fig. S3†). As expected, primers could hardly be extended to full length with the template of DNA modiﬁed with AQA (lane 2). However, significant extension efﬁciency was observed in the group of DNA templates that were not labeled with AQA.

We prepared a model to verify the OFCRCA strategy further. Two 80 bp DNA strands, one containing two 5fU sites in the duplex and the other with thymines at the same positions, were generated by PCR ampliﬁcation. Post AQA labeling and purifying by Spin Column Wash Tubes, the 80 bp DNA products were ligated with a predesigned hairpin adaptor (SL). The ratio between DNA and adaptors was ﬁrst optimized to guarantee high ligation efﬁciency. It was observed that a ratio of 1 : 4 (DNA : adaptor) results in a relatively good connection efﬁciency (Fig. S4†). Next, the conditions for ampliﬁcation reactions were also optimized, including the ratio of template DNA and primer and the concentration of phi29 DNA polymerase (Fig. S5†). To ensure the correct RCA reaction, we cut the RCA product with a restriction enzyme Eam1104, which has a restriction site in the SL adaptor, after adding the second primer (SPSL). The digested products were then subjected to PAGE analysis. The main products obtained were 130 bp and 260 bp in length, corresponding to the chain with two ligated adaptors (Fig. S6†). The appearance of the 260 bp band was attributed to insufficient degradation. Under optimal conditions, the ﬂuorescence intensity from RCA increased linearly with similar slopes with increasing reaction time. This result conﬁrmed no signiﬁcant polymerization efficiencies for DNA–T, DNA–5fU, DNA–5hmU, and DNA–5fU from DNA–5hmU oxidation (Fig. 2A). DNA templates without adaptor ligation showed almost no ﬂuorescent signal (Fig. S7†). The slopes and R² values from all possible ﬁts were arranged. In the ﬂuorescence curve of a typical RCA reaction, a high correlation coefﬁcient can be obtained by collecting data in 10 min (Fig. 2B–D, and Table S2, † R² > 0.9995).

Compared with DNA–5fU templates without chemical labeling, the slope of the DNA–5fU template modiﬁed with AQA was reduced by about 67% (Fig. 3A). However, the slope of DNA–T and DNA–5hmU showed no signiﬁcant differences postreaction with AQA under the same conditions (Fig. 3B). We then attempted to get a linear correlation between 5fU content and the efﬁciency difference of polymerization reaction (ΔS) for quantitative estimation of 5fU. The correlation indicated that with the increase in DNA–5fU–AQA content (C₅fU, mixed samples with different proportions of DNA–5fU–AQA and DNA–
The success of OFCRCA in detecting 5fU prompted us further to extend this method to the quantification of 5hmU. To evaluate the exact content of 5hmU in different samples, we plotted a linear curve of the decrease in the polymerization efficiency of phi29 DNA polymerase caused by the varying concentrations of DNA–5fU–AQA (oxidized from DNA–5hmU) (Fig. 4A). The equation represented the perfect fit of the regression line, $\Delta S = 0.3914C_{5fU} - 4.101$ ($R^2 = 0.9938$, Fig. 4B), indicating that the 5hmU content could also be calculated from the $\Delta S$ value.

The reliability of OFCRCA was verified by quantifying three artificially prepared DNA–5fU samples in parallel ($n = 3$) before applying the strategy for detecting 5fU in the genome. The results showed 94.5–112.4% recovery with small relative standard deviations (Table S3†), which confirmed the accuracy of this method. The high accuracy of the method encouraged us to extend the method to quantify 5fU in biological samples. First, genomic dsDNA extracted from MCF-7 cells was used to construct a quantitative calibration curve to detect genomic 5fU. The dsDNA (extracted from MCF-7) was then sheared into small dsDNA fragments (about 200 bps) using ultrasound following the manufacturer’s instructions (Fig. S8†). The formed dsDNA fragments were subjected to NEBNext ultra end repair and dA-tailing, followed by ligation to SL. Amplification reactions with the genomic sample before and after their reaction with AQA showed a clear polymerization efficiency difference, with about 55.9% reduction in the slope of the time-dependent fluorescence spectrum (Fig. 5A). The decrease in the slope represents the global 5fU levels in MCF-7 dsDNA (0.00103% in DNA), which was quantified by LC-MS (Fig. S9†). A quantitative calibration plot for 5fU in genomic DNA was obtained by analyzing the gradual dilution of AQA-tagged MCF-7 dsDNA fragments. The $\Delta S$ increased linearly with the 5fU content in the range of 0.000103–0.001030%, with a linear correlation of $\Delta S = 5.162C_{5fU} + 3.573$ ($R^2 = 0.9859$, Fig. 5B and C). The content of 5fU was quantified to be 6.5 per10^6 dNs in LO2 genomic

**T**, the polymerization efficiency of phi29 DNA polymerase was considerably reduced. This trend was consistent with the results of previous chain extension experiments (Fig. 3C). Correspondingly, $\Delta S$ increased linearly from 12.9 to 67.3%, with an increase in $G_{5fU}$ from 20 to 100%. The linear equation can represent the perfect fit of the regression line, $\Delta S = 0.6811C_{5fU} - 1.695$ ($R^2 = 0.9972$, Fig. 3D), indicating that the 5fU content can be calculated from the $\Delta S$ value. The detection limit was 5.15% for DNA–5fU–AQA (at 3SD/slope), representing 8.24 fmol of 5fU. These results indicated the promising potential of OFCRCA for the genome-wide detection of 5fU. The main challenge for quantitative detection of 5fU in genetic samples is the high variation in the types of modified bases, the limited abundance of 5fU, and the variation in different cell types.

**Fig. 2**  Effect of DNA modification on the polymerization efficiency of polymerase. (A) Time-dependent fluorescence spectra of the RCA reactions with DNA–T, DNA–5fU, DNA–5hmU, and DNA–5fU from 5hmU oxidation. (B) Data processing for calculating all the slopes from the plot of the RCA reaction with DNA–T; cycle$\text{i}$ represents the initial screening cycle site; cycle$\text{f}$ represents the selected computation interval (see the ESI for details†). (C) Data processing for screening $R^2$. (D) The screened confidence interval in the plot of the RCA reaction with DNA–T.

**Fig. 3**  Feasibility of OFCRCA for the detection of 5fU. (A) Time-dependent fluorescence spectra of the RCA reaction with DNA–5fU before and after labeling with AQA. (B) Histogram of $\Delta S$% for the RCA reaction with DNA–T, DNA–5hmU, and DNA–5fU before and after labeling with AQA. (C) Time-dependent fluorescence spectra of the diverse proportions of DNA–5fU–AQA. (D) Histogram of $\Delta S$ versus DNA–5fU–AQA content during the RCA reaction with diverse proportions of DNA–5fU–AQA. The values represent the means ± SD from three independent measurements.

**Fig. 4**  Feasibility of improved OFCRCA for the detection of 5hmU. (A) Time-dependent fluorescence spectra of the diverse proportions of DNA–oxi–5hmU–AQA. (B) Histogram of $\Delta S$ versus DNA–oxi–5hmU–AQA content during the RCA reaction with diverse proportions of DNA–oxi–5hmU–AQA. The values represent the means ± SD from three independent measurements.
and LO2 was then subjected to OFCRCA to quantify the total 5fU per 10^6 dNs in genomic DNA of MCF-7 (Fig. 5D).

The oxidation from the total 5fU content. Consequently, 5hmU was subtracting the amount of 5fU present in the sample before content a

...genomic DNA by OFCRCA was relative error of quantitative detection for 5fU levels in LO2...© 2022 The Author(s). Published by the Royal Society of Chemistry

In summary, we have developed a novel OFCRCA strategy capable of sensitively quantifying the global levels of thymine oxides with nanogram scale sample input, including 5hmU and 5fU, that are difficult to quantify by existing methods as they are found in minimal amounts in organisms. The OFCRCA method is reliable, cost-effective, and convenient because multiple samples can be analyzed simultaneously and on one plate with standard laboratory equipment. The strategy will also be a promising option for analyzing numerous samples' thymine oxides and other nucleoside modifications.

Conclusions

In summary, we have developed a novel OFCRCA strategy capable of sensitively quantifying the global levels of thymine oxides with nanogram scale sample input, including 5hmU and 5fU, that are difficult to quantify by existing methods as they are found in minimal amounts in organisms. The OFCRCA method is reliable, cost-effective, and convenient because multiple samples can be analyzed simultaneously and on one plate with standard laboratory equipment. The strategy will also be a promising option for analyzing numerous samples’ thymine oxides and other nucleoside modifications.
