Trends in Analytical Chemistry 54 (2014) 24–35

Contents lists available at ScienceDirect

Trends in Analytical Chemistry

journal homepage: www.elsevier.com/locate/trac

Review

Recent advances in the analysis of 5-methylcytosine and its oxidation products

Bi-Feng Yuan *, Yu-Qi Feng

Key Laboratory of Analytical Chemistry for Biology and Medicine, Ministry of Education, Department of Chemistry, Wuhan University, Wuhan 430072, PR China

ARTICLE INFO

Keywords:
5-Carboxylcytosine
5-Formylcytosine
5-Hydroxymethylcytosine
5-Methylcysine
Disease biology
DNA methylation
Epigenetic modification
Genome-wide distribution analysis
Overall analysis
Oxidation product

ABSTRACT

DNA methylation (5-methylcytosine, 5-mC) is an important epigenetic mark that has regulatory roles in a broad range of biological processes and diseases. Aberrant DNA methylation is associated with a wide variety of human diseases. Recently, novel cytosine modifications with potential regulatory roles, such as 5-hydroxymethylcytosine (5-hmC), 5-formylcytosine (5-foC), and 5-carboxylcytosine (5-caC), were discovered. Systematic investigation of the functions of 5-mC and its oxidation products promotes understanding of the mechanism underlying association of epigenetic modifications with disease biology. In this respect, remarkable advances have been made in developing methods for investigating the occurrence and the localization of these cytosine modifications. In this review, we mainly focus on the recent methodological advances in the analysis of the total levels of 5-mC and its oxidation products (5-hmC, 5-foC and 5-caC). In addition, we summarize and discuss new methods for mapping the genome-wide distribution of 5-hmC, 5-foC and 5-caC.

Contents

Abbreviations: 5-caC, 5-Carboxylcytosine; 5-cadC, 2′-Deoxy-5-caboxycytidine; 5-foC, 5-Formylcytosine; 5-fodC, 2′-Deoxy-5-formylcytidine; 5-gmC, β-Glucosyl-5-hydroxymethylcytosine; 5-gmdC, β-Glucosyl-5-hydroxymethyl-2′-deoxycytidine; 5-hmC, 5-Hydroxymethylcytosine; 5-hmdC, 2′-Deoxy-5-hydroxymethylcytidine; 5-mC, 5-Methylcytosine; 5-mdC, 2′-Deoxy-5-methylcytidine; 5-mdCMP, 2′-Deoxy-5-methylcytidine-3′-monophosphate; β-GT, T4 Bacteriophage β-glucosyltransferase; BSTFA, *N,O*-bis-(Trimethylsilyl)-trifluoroacetamide; CAD, Collisionally-activated dissociation; CE, Capillary electrophoresis; CMS, Cytosine 5-methylenesulfonate; DNMT, DNA cytosine-5-methyltransferase; ECL, Electrochemiluminescence; EDC, 1-Ethyl-3-[3-dimethylaminopropyl]-carbodiimide hydrochloride; ELISA, Enzyme-linked immunosorbent assay; GC-MS, Gas chromatography-mass spectrometry; HCC, Hepatocellular carcinoma; HILIC, Hydrophilic interaction liquid chromatography; HPLC, High-performance liquid chromatography; ICP-MS, Inductively-coupled plasma mass spectrometry; LC, Liquid chromatography; LC-MS, Liquid chromatography-mass spectrometry; LIF, Laser-induced fluorescence; LOD, Limit of detection; LOQ, Limit of quantification; MBD, Methyl-binding domain; MRM, Multiple-reaction monitoring; MS/MS, Tandem mass spectrometry; NGS, Next-generation sequencing; oxBS-Seq, Oxidative bisulfite sequencing; SAX/RP, Strong anion-exchange/reversed-phase; SMRT, Single-molecule real-time; TAB-Seq, TET-assisted bisulfite sequencing; TET, Ten–Eleven translocation proteins; TLC, Thin-layer chromatography; UPLC, Ultra-performance liquid chromatography.

* Corresponding author. Tel./Fax: +86 2768755595.
E-mail address: bfyuan@whu.edu.cn (B.-F. Yuan).

0165-9936/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.trac.2013.11.002

1. Introduction

DNA methylation, consisting of the addition of a methyl group at the fifth position of cytosine (5-methylcytosine, 5-mC) at the CpG dinucleotide site, is one of the most well-studied epigenetic marks in mammals [1]. DNA methylation is reversible and involved in diverse physiological functions, including maintenance of chromosomal integrity, X-chromosome inactivation, and transcriptional suppression of genes. DNA methylation is crucial for mammalian development and cellular differentiation [2]. Aberrant DNA methylation is a well-recognized hallmark of many diseases, such as heart disease, diabetes, neurological disorders and cancers [3,4]. A variety of investigations have therefore focused on characterizing DNA methylation and its roles associated with pathogenesis [5].

Recently, it was discovered that the Ten–Eleven translocation (TET) proteins were capable of catalyzing the sequential oxidation of 5-mC to 5-hydroxymethylcytosine (5-hmC), 5-formylcytosine (5-foC), and finally to 5-carboxylcytosine (5-caC) [6–9]. The resulting 5-foC and 5-caC can be further recognized and cleaved by thymine-DNA glycosylase, thereby restoring unmethylated cytosine via base-excision repair machinery [10]. These reports indicate that active DNA demethylation may be achieved through a multistep oxidation of 5-mC with the generation of various forms of intermediates of 5-hmC, 5-foC and 5-caC (Fig. 1). Besides, a recent study suggests mammalian DNA cytosine-5-methyltransferases (DNMTs) can directly convert 5-mC in DNA to cytosine *in vitro*, indicating that DNMT proteins may act as both methyltransferases and demethylases [11]. However, the reported *in-vitro* demethylase activity of DNMTs needs further confirmation and the significance of this novel activity needs to be demonstrated *in vivo*.

These identified novel DNA modifications were considered with potential regulatory roles [12]. 5-hmC is now viewed to be "the sixth base" of the genome in mammals besides A, C, T, G, and 5-mC [13], which raises the possibility that 5-hmC may act as an epigenetic mark associated with transcriptional regulation [14]. 5-hmC also plays important roles in maintenance and differentiation of embryonic stem cells [15–17]. In addition, 5-hmC is found to be enriched in brain tissues and accumulates with age, so it is suspected of playing regulatory roles in neurodevelopment and neurological diseases as well as aging [7,18,19]. Furthermore, some studies demonstrated that 5-hmC significantly decreased in various tumors, indicating that 5-hmC is associated with tumor formation and development [20–23]. Recently, we and others demonstrated that 5-foC and 5-caC slowed down RNA polymerase II transcription and changed the fidelity of DNA replication and transcription, which may suggest possible functional roles of these modifications on replication and transcription [24,25].

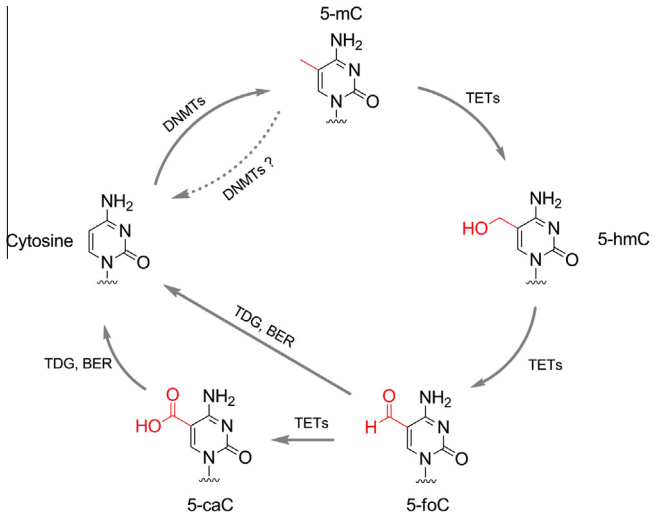

Fig. 1. The structures of the cytosine derivatives. 5-mC modification is catalyzed by DNA methyltransferases (DNMTs). 5-mC can be demethylated through the oxidation of 5-mC by TET proteins to produce 5-hmC, 5-foC and 5-caC. Alternatively, DNMTs may serve as bifunctional enzymes and can directly demethylate 5-mC to cytosine *in vitro*. TDG, Thymine-DNA glycosylase; BER, Base-excision repair.

Although multiple cytosine modifications beyond 5-mC have now been identified, they differ in their abundance within the genome, with 5-mC being present at a frequency about 10-fold to 100-fold higher than that of 5-hmC, and 5-hmC being about 40-fold to 1000-fold higher than that of 5-foC and 5-caC [8,12,26]. In the past decade, remarkable advances have been made in the development of new methods for the investigation of the occurrence and the localization of these cytosine modifications. These methods can mainly be divided into two categories: overall detection, and bisulfate conversion-based single-base resolution detection. As for the bisulfate conversion-based single-base resolution detection of 5-mC, some reviews summarized well the detection methods developed [27–29]. However, few reviews have surveyed the overall detection of 5-mC and its oxidation products, so, in this review, we mainly focus on the recent methodological advances in the overall analysis of 5-mC and its oxidation products, 5-hmC, 5-foC and 5-caC. Also, since, in the past three years, tremendous progress has been made in understanding the biological function of 5-hmC, 5-foC and 5-caC as a direct result of the rapid development of their genome-wide mapping methods, we summarize and discuss these newly established methods for their genome-wide distribution analysis. Further, we discuss their relative utility and drawbacks and provide specific examples that utilized these technologies to address important biological questions.

2. Overall analysis of 5-mC and its oxidation products

Reported methods for the determination of overall 5-mC and its oxidation products mainly include liquid chromatography (LC), capillary electrophoresis (CE), LC-mass spectrometry (LC-MS), gas chromatography-MS (GC-MS), thin-layer chromatography (TLC), chemical derivatization-based detection, single-molecule detection, and immuno-based detection (Fig. 2). These techniques rely on the release of DNA components, such as 2′-deoxynucleotides, 2′-deoxynucleosides, or nucleobases, by means of enzymatic or chemical treatment followed by determination of the corresponding components. The information of different methods is summarized in Table 1.

2.1. Liquid chromatography

Genome-wide DNA methylation can be evaluated by chromatographic techniques, which are, in general, highly quantitative, reproducible, but less sensitive. These assays are based on the chromatographic separation of different nucleobases, 2′-deoxynucleosides or 2′-deoxynucleotides obtained by enzymatic or chemical hydrolysis of DNA. Upon hydrolysis, these DNA components are typically separated by reversed-phase [30–32] or cation-exchange high-performance LC (HPLC) [33]. DNA can be hydrolyzed with formic acid to nucleobases; in this case, the complete elimination of RNA is required, since both DNA and RNA contain 5-mC [34], which can therefore affect the accurate quantification of DNA methylation. The baseline separation of the DNA components is essential, since the analysis relies heavily on the chromatographic separation to avoid co-elution of analytes. In addition, a relatively large amount of genomic DNA (~1–50 μg) is usually needed, due to the low sensitivity.