

Development of methods derived from iodine-induced specific cleavage for identification and quantitation of DNA phosphorothioate modifications

Sucheng Zhu

Shanghai Jiao Tong University

Tao Zheng

Shanghai Jiao Tong University

Lingxin Kong

Shanghai Jiao Tong University

Jinli Li

Shanghai Jiao Tong University

Bo Cao

Qufu Normal University

Yihua Sun

Shanghai Jiao Tong University

Ying Chen

Shanghai Jiao Tong University

Zixin Deng

Shanghai Jiao Tong University

Delin You (✉ dlyou@sjtu.edu.cn)

Shanghai Jiao Tong University <https://orcid.org/0000-0003-1644-9825>

Methodology

Keywords: DNA phosphorothioate (PT) modifications, Iodine-induced cleavage (ICA), ICDS, PT-IC-Seq

Posted Date: December 23rd, 2019

DOI: <https://doi.org/10.21203/rs.2.19448/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Biomolecules on October 28th, 2020. See the published version at <https://doi.org/10.3390/biom10111491>.

Abstract

Background DNA phosphorothioate (PT) modifications is a novel modification that occurs on the DNA backbone, which refers to a non-bridging phosphate oxygen replaced by sulfur. This exclusive DNA modification widely distributes in bacteria and archaea but has not been found in eukaryotes to date. PT modification renders DNA nuclease tolerance and serves as a constitute element of bacterial restriction-modification (R-M) defensive system and more biological functions are awaiting exploration. Identification and quantification of the bacterial PT modifications are thus critical to better understanding their biological functions.

Results In this report, we describe three detailed methods based on PT-specific iodine-induced cleavage and high-throughput next-generation sequencing technologies: an iodine-induced cleavage assay (ICA), a deep sequencing of iodine-induced cleavage at PT site (ICDS) and an iodine-induced cleavage PT sequencing (PT-IC-Seq), for the investigation of PT modifications. Using these approaches, we have identified the presence of PT modifications and quantized the frequency of PT modifications in bacteria. These characterizations contributed to the high-resolution genomic mapping of PT modifications, in which the distribution of PT modification sites on the genome was marked accurately and the frequency of the specific modified sites was reliably obtained.

Conclusion Here, we provide time-saving and less labor consuming methods for both of qualitative and quantitative analysis of genomic PT modifications. The application of these methodologies will provide new opportunities for better understanding the biology of the PT modifications and opens door for future further systematically study.

Introduction

Naturally occurring epigenetic modifications of DNA have been found prevalent in organism from all domains of life [1]. Usually, those modifications involve a variety of chemical groups (like methyl group, amino acids, polyamines, monosaccharides, and disaccharides) appended to the nucleobase portion of a nucleotide. These modifications don't alter the specificity of base pairing but play important role in protection and genetic regulation. DNA phosphorothioate (PT) modifications, in which the non-bridging phosphate oxygen is replaced by sulfur, occurs to the sugar-phosphate backbone rather than the nucleobase in a sequence- and R_p stereo-specific manner [2–4].

The discovery of natural PT modifications originated from DNA degradation (Dnd) phenotype during electrophoresis of genomic DNA (gDNA) isolated from streptomyces lividans 1326 [5]. Since then the Dnd phenotype has been considered as the typical method for detection of PT modification. This modification turned out to be the sulphur-for-oxygen substitution supported by the isotope incorporation during mass spectrometric analysis and all identified physiological PT internucleotide linkages was in the unique stereochemistry of R_p configuration [6]. PT modification made DNA susceptible to oxidative cleavage by a peracid derivative of Tris formed at the electrophoretic anode [7]. Meanwhile, the modified DNA

possessed enhanced resistance to nuclease (like nuclease P1), which set stage for the generation of PT-linked dinucleotides facilitating the quantification of genomic PT modification by liquid chromatography-mass spectrometry (LC-MS) [3]. Further studies have demonstrated that this modification is governed by the *dndABCDE* cluster in *Streptomyces lividans* 1326 [6], and deleting of *dnd* cluster could abolish the modification [8]. PT modification has been found in many other bacteria and the functional study of the *dnd* clusters revealed the necessity of the five proteins [9]. DndA is regarded as a pyridoxal 5'-phosphate-dependent cysteine desulphurase and supplying sulphur for PT modification, DndC exhibited homology to 3'-phosphoadenosine-5'-phosphosulphate (PAPS) reductase and showed ATP pyrophosphatase activity in vitro [10], DndD carried the conserved ATP-binding motif and was suggested to provide energy for stabilizing DNA or site-specific DNA nicking [11], and DndE has been showed to bind the nicked double-stranded DNA substrates preferentially in a sequence independent manner [12, 13]. DndB negatively regulated the PT modification as the deletion of *dndB* aggravated DNA degradation and increased PT frequencies [14]. The biochemical study of PT modification revealed that PT-modifying enzymes DndACDE function as a large protein complex [15]. However, it is still poorly understood that how the Dnd proteins incorporate the sulphur into DNA synergistically. Up to date, PT modifications have been found in over 1000 bacterial and archaeal strains but not in eukaryotes, and their biological functions are still poorly understood [9]. In some bacteria, the *dndABCDE* contribute to the constitution of the defensive R-M system with *dndFGH*, by which the PT-carrying host could distinguish and attack non-PT-modified foreign DNA [16]. Other biological function of PT modification remains to be explored, which require more characterization of PT modification in a variety of bacteria. To this end, quantitative analysis of PT modifications and whole genomic landscapes description of the location and frequency of PT modification are thus important in bacteria. The tolerance of PT-modified DNA enabled the sequence identification and frequency quantitation by LC-MS/MS in the two-nucleotide sequence context after hydrolysis of the DNA samples by nucleases [3], however this method is complicated and labor-intensive.

Long before the discovery of natural PT modifications, the modification was developed by artificial chemical approaches to protect oligodeoxynucleotides against nuclease degradation [17, 18]. Previous works showed that compared with the native form of DNA, the PT-modified DNA has special chemical properties and is susceptible to cleavage by some chemical reagents, such as 2-iodoethanol [19]. Subsequently it found that three products: 3'- or 5'- fracture phosphate DNA and desulfurization produce normal phosphate two ester bonds DNA, are produced through selective cleavage of PT linkages by 2-iodoethanol [20]. Moreover, the higher cutting efficiency (nearly 100%) for the DNA containing PT modifications could be achieved under the optimized reaction system [21–24]. Taking advantage of these special properties, an iodine-induced cleavage assay was developed to rapidly determine whether there is PT modification on the DNA. This method is easy to follow compared with LC-MS/MS and has become the most common method used for PT modifications discovery [25–28]. As PT modification is a partial and highly dynamic modification, more techniques should be developed to locate the PT modification sites and uncover the diverse sequence selectivity and quantity in different bacteria, even in an individual cell of a population because of the heterogeneity of PT modification [25, 29].

Here, we describe a series of methods based on PT-specific iodine-induced cleavage and high-throughput next-generation sequencing technologies: ICA, ICDS and PT-IC-Seq [25, 29](See Fig. 1 for a schematic overview). Among them, the ICA, a method of rapidly detecting PT modification status in the genome, has been developed by using electrophoresis pattern of the iodine-cleaved gDNA. The DNA samples were cleaved by iodine solution under an optimized condition, and the presence of PT modifications was firstly detected by electrophoresis in agarose gel, and then confirmed by LC-MS/MS. The ICDS technology combining high throughput Illumina sequencing with iodine specific cleavage maps the locating of PT modified sites in the genome. The PT-modified sites were cleaved specifically by iodine, and the resultant cutting sites were then marked with special labels. The tag location sites could be detected by the two-generation sequencing, which in turn fulfilled the determination of the distribution and localization of genomic PT modification sites. The PT-IC-Seq technology was an updated derivative of ICDS with higher throughput and resolutions. For the quantitative determination of PT modification percentage at each site, the DNA samples were firstly cleaved by iodine at selected phosphorothioate linkages, followed by tag ligations and high-throughput sequencing. No enrichment process was required for mapping sequenced reads to the reference genome. Taken together, these methods could be applied to investigate and quantitatively analyze PT modifications at whole genomic landscapes, and will expand our knowledge of PT modifications in the future.

Material And Methods

Materials

E. coli strain DH10B was purchased from Novagen. *E. coli* strain B7A was obtained from Dr Jaquelyn Fleckenstein (Departments of Medicine and Molecular Sciences, University of Tennessee Health Science Center).

Mono nucleosides standard samples (A, G, C, T) and PT-modified dinucleotide standard samples (G_{PS}A, G_{PS}T) were synthesized by Sangon Biotech Co. Ltd. (Shanghai).

MicroSpin G-25 columns ([GE Healthcare](#), #27-5325-01).

Pipettes, tips and microcentrifuge tubes ([Eppendorf](#))

Reagents

Ethanol (Sigma-Aldrich, #32205).

Acetic acid (Sigma-Aldrich, #33209).

I₂ (Sigma-Aldrich, #D7295-20X).

dNTP (Sigma-Aldrich, #D7295-20X).

Nuclease P1 (Sigma-Aldrich, #N8630).

DNA ladder (Thermo Fisher Scientific, #10787018).

DNA Gel Loading Dye (Thermo Fisher Scientific, #R0611).

FastAP Thermosensitive Alkaline Phosphatase (Thermo Fisher Scientific, #EF0654).

[Klenow Fragment](#) (3'→5' exo-) (New England BioLabs, #M0212L).

Buffer 2 (New England BioLabs, #B7002S).

dATP (New England BioLabs, #N0440S).

T4 DNA ligase (New England BioLabs, #M0202S).

Chemicals

Agarose ([Sigma-Aldrich](#), #A9539).

NaCl (Sigma-Aldrich, #746398).

Na₂HPO₄ ([Sigma-Aldrich](#), #795410).

EDTA (Sigma-Aldrich, #E9884).

ZnCl₂ (Sigma-Aldrich, #793523).

Thiourea (Sigma-Aldrich, #T8656).

Sodium acetate ([Sigma-Aldrich](#), #791741).

Tryptone (Sigma-Aldrich, #T7293).

Yeast Extract (Sigma-Aldrich, #Y1625).

Tris (Sigma-Aldrich, #RDD008).

Kits

Bacterial DNA Kit (OMEGA, #D3350-02).

Plasmid Mini Kit (OMEGA, #D6943-02).

Bacterial DNA Kit (OMEGA, #D3350-02).

Cycle Pure Kit (OMEGA, #D6492-02).

Quant-iT™ PicoGreen™ dsDNA Assay Kit (Thermo Fisher Scientific, #P11496).

Quick Blunting™ kit (New England BioLabs, #E1201L).

NEBNext DNA Library Prep Reagent Set for Illumina (New England BioLabs, #E6000L).

Equipment

Horizontal Electrophoresis Systems ([BIO-RAD](#), #09-0298).

PCR Systems (BIO-RAD, T100™ Thermal Cycler).

HPLC Systems ([Agilent](#), 1260 Infinity II Analytical-Scale LC Purification System).

LC-MS/MS Systems (Agilent, HPLC 1260 - QQQ 6470).

Bioanalyzer Instrument 2100 (Agilent, #G2939BA).

Heated Circulating Baths (Thermo Fisher Scientific, #TSCIR35).

NanoDrop 2000 (Thermo Fisher Scientific, #ND2000CLAPTOP).

Qubit™ 4 Fluorometer (Thermo Fisher Scientific, #Q33226).

UV Spectrophotometer (HACH, #DR6000).

Constant Temperature Heater (Eppendorf, #5383000078).

Sonicator (Bandelin, Sonoplus HD2070 with Sonotrode MS72).

SPRIworks Fragment Library System (Beckman).

Freeze Dryer (Labconco).

Primers

Duplex tag-sequence

F1: 5'-Phos/TTTAACCGCGAATTCAG/ddC/ -3'(DNA, HPLC-purified)

R1: 5'-GCTGGAATTCGCGGTAAAT-3' (DNA, HPLC-purified)

Enriched PCR primers

F2: 5'-CAAGCAGAAGACGGAATACGA-3'(DNA, PAGE-purified)

R2: 5'-AATGATACGGCGACCACCGAGATCTACACTCTTCCCTACACGACG

CTCTTCCGATCTGCTGGAATTCGCGGTAAAT-3' (DNA, PAGE-purified)

Sample preparation (using *E. coli* B7A as an example)

1. Isolate a single colony from a freshly streaked selective plate, and regrown in LB medium at 37 °C for overnight or to logarithmic phase.
2. Harvest cells by centrifugation at 10,000 rpm for 2 min.
3. Aspirate and discard the media.
4. Extract gDNA using the Bacterial DNA Kit according to the manufacturer's instructions.
5. Store eluted DNA at -80 °C until ready for use in following experiments.

[Note: For optimal DNA yields, the starting culture volume should be based on culture cell density. From experience, the cell density at 600 nm (OD₆₀₀) of 2.0~3.0 is recommended.]

ICA assay

1. Prepare the 10 × ICA Reaction Buffer (500 mM Na₂HPO₄, pH 9.0, 30 mM iodine).
2. Add ~1 µg gDNA with 1 × ICA Reaction Buffer to clean microcentrifuge tube. Mix well, and incubate at 65 °C for 10 min and then slow cooled (0.1 °C/s) to 4 °C using a thermal cycler.
3. Analyze reaction sample using ethidium bromide-stained 1.0% agarose gel electrophoresis in 1 × TAE buffer containing 50 µM thiourea.
4. At the same time, analyze the above samples using the LC-MS/MS method (see section of Quantification of PT modifications in DNA by LC-MS/MS).

[Note: Always prepare a fresh iodine for every use, and confirm slow cooling process. Reduce voltage and increase gel running times possible during electrophoresis. These tips may yield better results.]

The standard curve of mono nucleosides

1. Dilute the mono nucleosides standards (C and T) to six different concentrations (0 ng/µl, 2 ng/µl, 4 ng/µl, 10 ng/µl, 20 ng/µl, 40 ng/µl).

[Note: The total number of A is equal to the number of T, and similarly the number of C is equal to the number of G. Thus, we could choose two mono nucleosides to draw the standard curve.]

2. Analysis samples using the High Performance Liquid Chromatography (HPLC) method, and the analysis of results as follow:

With the concentration of mono nucleosides standards (ng/ μ l) as the abscissa (X) and the peak area values of HPLC as the ordinate (Y), the standard curve was drawn.

[Note: Always prepare a fresh set of standards and record the samples volume for every use.]

The standard curve of dinucleotides

1. Dilute the *Rp* configuration of the PT modified dinucleotide standards to seven different concentrations (0 fmol/ μ l, 2 fmol/ μ l, 10 fmol/ μ l, 20 fmol/ μ l, 40 fmol/ μ l, 100 fmol/ μ l, 200fmol/ μ l).
2. Add \sim 100 fmol/ μ l the *Sp* configuration of PT modified dinucleotides to each standard used as internal reference.
3. Analysis samples using the LC-MS/MS (see section Quantification of PT modifications in DNA by LC-MS/MS), and the analysis of results as follow:

With the concentration of *Rp* configuration of the PT modified dinucleotide (fmol/ μ l) as the abscissa (X) and the ratio of the standard peak area to internal reference peak area values measured by LC-MS/MS as the ordinate (Y), the standard curve was drawn.

[Note: Always prepare a fresh set of standards and record the samples volume for every use.]

Quantification of PT modifications in DNA by LC-MS/MS

1. Prepare the nuclease P1 Hydrolysis Buffer (0.5 mM ZnCl₂, 30 mM sodium acetate, pH 5.3).
2. Add \sim 20 μ g purified DNA, 2 U of nuclease P1 and Hydrolysis Buffer to a clean microcentrifuge tube. Mix well, and incubate at 50 °C for 2 h using a thermal cycler.
3. Then mix in 100 mM Tris-HCl, pH 8.0 and 15 U of FastAP Alkaline Phosphatase, and incubate at 37 °C for 2 h for the totally dephosphorylation.
4. Remove the enzymes by ultrafiltration at 10000 rpm for 10 min.

[Note: This step is critical for removal of enzymes that may interfere with follow-up experiments.]

5. Add \sim 200 μ l deionized water to ultrafiltration tube and repeat step 4 until all of the sample has been transferred to the [collection](#) tube.
6. Collect the filtrate in a clean microcentrifuge tube and dry the sample.

[Note: For subsequent LC-MS/MS analysis, the volume of sample is as small as possible. [Concentrated sample using the Freeze Dryer may yield better results.](#)]

7. Resuspend and elute the sample from step 6 in 40 μ L deionized water.

8. Quantitative analysis sample by LC-MS/MS [2], and the analysis of results according to standard curves as follow:

The hydrolyzed mono nucleotides were quantified by HPLC according to the standard curve of mono nucleosides. Meanwhile, the PT modified dinucleotides were quantified by LC-MS/MS according to the standard curve of dinucleotides. Thus, the number of PT modified dinucleotides in a unit length of DNA can be calculated.

ICDS assay

1. Cleave ~20 µg purified gDNA by iodine as described above in ICA methods.

2. Remove residual iodine and salts using [MicroSpin G-25 columns](#).

[Note: This step is critical for removal of residual iodine and salts that may interfere with follow-up experiments.]

3. Add 10 U of FastAP Alkaline Phosphatase to the sample at 37 °C for 1h for remove terminal phosphate groups.

4. Inactivate the enzyme by heating at 75 °C for 5 min and slow cooled (0.1 °C/s) to 4 °C to assure proper complementary re-annealing.

5. Clean up the sample using the Cycle Pure Kit and [elute it in 30 µl MilliQ water](#).

6. Blunt end of the break sites using the Quick Blunting™ Kit at room temperature for 30 min.

7. Inactivate the enzyme by heating at 75 °C for 10 min and slow cooled as before.

8. [Clean up and elute the sample as before](#).

9. Then mix in 1x NEB Buffer 2, 0.1 mM dATP and 15 U of [Klenow Fragment \(3'→5' exo-\)](#), and incubate at 37 °C for 30 min for the 3'-deoxyadenylation (namely A-tailing).

10. Inactivate the enzyme by heating at 75 °C for 20 min and slow cooled as before.

11. Clean up and elute the sample as before.

12. Combine 3 µM of custom duplex tag-sequence (see section of Primers) with 3'-deoxyadenylated ends by T4 DNA ligase at 16 °C for 16 h.

13. Inactivate the enzyme by heating at 75 °C for 10 min and slow cooled as before.

14. Clean up and elute the sample as before.

[Note: The quality of the DNA samples was measured using NanoDrop 2000 and their concentration was measured using the Quant-iT™ PicoGreen™ dsDNA Assay Kit.]

15. Shear the DNA samples to fragment lengths between 150-350 bp by sonication, and then ligate standard Illumina sequencing adaptor using the SPRIworks Fragment Library System.

[Note: To fragment the DNA samples to a size range of 150-350 bp, using a probe sonicator at an amplitude of 20% with 20 seconds “ON” and 10 seconds “OFF” (10 min total), while on ice to avoid excessive heat.]

16. Concentrate samples using the Freeze Dryer and elute it in 30 µl MilliQ water.

17. Enrich the iodine-cleaved DNA linked unique tag by PCR amplification for 15 cycles using F2 and R2 as the primers.

[Note: During the PCR amplification, one of the primers used matches the marked segments attached to the ends of cleavage by iodine (F2), while the other matches the standard Illumina sequencing adaptor (R2). This allows only fragments containing the iodine-cleaved ends to be amplified, which achieves the enrichment of the PT-modified molecules.]

18. Sequence the libraries constructed from step 17 on the Illumina HiSeq X Ten platform.

[Note: Agilent Technologies 2100 Bioanalyser is used to confirm successful library generation and Life Technologies Qubit 3.0 Fluorometer for quantification.]

19. Analysis the ICDS sequencing data, and the details as follow:

After completing Illumina sequencing, the reads containing tag were trimmed for adaptor and tag and done quality control as follows:

(1) Clipping the adapter sequences.

(2) Removing non-A, G, C, T bases of the 5' end.

(3) Trimming low-quality base (less than Q20).

(4) Removing reads with more than 10% of “N” calls.

(5) Filtering small fragments with less than 25 bp after clipping the adapter sequences and quality control, and then aligned to reference genomes by Burrows-Wheeler Aligner (BWA) and the position-wise coverage values were calculated by using the custom python script. The GAAC/GTTC sites will be defined as PT modified sites if their reads above 50 and ended at this site. Meanwhile, 10 non GAAC/GTTC sites were randomly selected as control.

PT-IC-Seq assay

1. Cleave ~20 µg purified gDNA by iodine as described above in ICA methods.

2. Remove residual iodine and salts using MicroSpin G-25 columns.

[Note: This step is critical for removal of residual iodine and salts that may interfere with follow-up experiments.]

3. Shear the DNA samples to fragment lengths of 150–350 bp by sonication.

[Note: To fragment the DNA samples to a size range of 150–350 bp, using a probe sonicator as before.]

4. According to instructions provided with the NEBNext DNA Library Prep Reagent Set for Illumina, the resulting fragments were end-repaired, adenylated at the 3' ends and ligated to Illumina paired-end adaptors.

5. Amplify DNA fragments by PCR for 15 cycles using standard Illumina adapter-specific primers.

6. Sequence the libraries constructed from step 5 on the Illumina HiSeq X Ten platform.

[Note: Agilent Technologies 2100 Bioanalyser is used to confirm successful library generation and Life Technologies Qubit 3.0 Fluorometer for quantification.]

7. Analysis the PT-IC-Seq data, and the details as follow:

All sequencing reads were trimmed for adaptor and low-quality bases and aligned to reference genomes by BWA for creating SAM files. Then, SAM files were converted to BAM files that were piled up using samtools and the results were visually performed using the Integrated Genomics Viewer 2.3 software (IGV; Broad Institute, Cambridge, MA, USA). Meanwhile, the position-wise reads number obtained from fragment terminals and across the same site were calculated respectively by using the custom python script. The PT modification frequency of each GAAC/GTTC sites were calculated by the number of reads ended at this site divided by all of the number of reads ended and crossed the same sites. In order to eliminate the false-ended reads arising from random shearing the DNA, 10 non GAAC/GTTC sites used as the control and their average PT modification frequency were calculated and used them as thresholds. The modification ratios at each PT sites in the whole genome will be calculated and analyzed if the modification frequency of those sites were above the set thresholds.

Results

ICA assay

For the feasibility assessment of ICA method, the gDNA isolated from *E. coli* B7A containing PT modified genes was tested as an example. The ICA assay showed that treatment of the PT-modified gDNA with iodine resulted in smaller fragments distribution compared with the control treated without iodine, whereas the gDNA from *E. coli* DH10B lacking PT gene cluster was not cleaved (Fig. 2A). The results

demonstrated that iodine could cleave PT modified DNA with high efficiency. This finding was consistent with the previous cleavage studies [18-21], so this assay is feasible to detect presence of PT modifications in the genome. To consolidate the above results, we re-analyzed PT modifications in gDNA of *E. coli* B7A by LC-MS/MS. As shown in Fig. 2B and Fig. 2C, the PT-containing dinucleotides (GpsA or GpsT) after hydrolysis of the gDNA by nucleases, is consistent with the retention time of the standard that possessed PT modifications with R_p configuration. These results confirmed that the ICA approach could rapidly detect PT modification status in the genome.

The standard curve of mono nucleosides

According to above the method, standard curves of different mono nucleosides standards were shown in Fig. 3.

The standard curve of dinucleotides

According to above the method, standard curves of different PT modified dinucleotide standards were shown in Fig. 4.

Quantification of PT modifications in DNA by LC-MS/MS

According to above the method, the dinucleotides of gDNA isolated from *E. coli* B7A was quantitatively analyzed as an example. As shown in Fig.5, we have detected PT modifications levels occurred in GpsA and GpsT contexts at 325 ± 8 and 361 ± 11 PTs per 10^6 nt, respectively in 20 μ g gDNA (Fig. 5), which is consistent with previous researches [3].

ICDS assay

Examples using ICDS method to characterize the PT-modified on the gDNA isolated from *E. coli* B7A have been reported by Cao et al. [25] and Li et al. [29]. They demonstrated that the method was feasible for discerning PT sites in the whole genome and can draw the high-resolution genomic maps of PT modifications. However, it has been shown that this approach was not applied in single-stranded modifications, such as *Vibrio cyclitrophicus* FF75 [25].

PT-IC-Seq assay

As the PT modifications have been enriched during PCR amplification step, the ICDS method can only be used for determining PT sites but not for quantification. Taking the limits of ICDS method into consideration, a novel approach named PT-IC-Seq was developed.

PT-IC-Seq method has been applied to quantify the PT modification on the gDNA isolated from *E. coli* B7A [29]. It demonstrated that the method was feasible for quantitatively characterize the genomic landscape of PT modifications and determine PT modification frequency at each modified site. Moreover,

the successful application of this approach in other PT modified strains, such as *S. enterica serovar Cerro* 87 [29], have supported its generality for future further discovery in bacteria.

Discussion

The nucleic acids contain diverse chemical modifications, which exerts important influence in a variety of life processes [30–32]. Among these modifications, DNA methylation is the most known and has essential roles in cellular processes, such as genome regulation, development and disease [33–35]. The recently discovered PT modifications occurring on DNA backbone was a novel DNA modification, in which the non-bridging phosphate oxygen is replaced by sulfur. This exclusive DNA modification is widespread in bacteria within a sequence selective and Rp stereo-specific manner but not found in mammal cell, and their physiological functions remain still poorly understood. Thus, determination and quantitative analysis of the PT modifications are essential for exploring and understanding their biological functions. We provide here a comprehensive description for the entire process of a series of approaches for analyzing PT modifications in the genome so that it can be easily adopted by researchers.

Since the initial 2014 report on PT modifications detection using the ICA assay, the technique has been widely used for PT modifications discovery. An important advantage of this method is its simplicity, because the detection of PT modifications can be completed in a relatively short time at low cost. The main goal of ICA method is to rapidly detect the presence of PT modifications in the genome. The treatment of iodine introduced a strand break at the PT-modified site, providing the basis for new technology development. During the development of the method, we found that the quality of iodine was key to successful experimentation according to previous results. From experience, a fresh iodine solution should be prepared when using every time. Meanwhile, the reduction of voltage and extending gel running time during electrophoresis process of ICA assay may help yielding better results. Importantly, we also used the LC-MS/MS method to further confirm the results after the ICA assay, which ensure the accuracy of the results.

In order to further understand the characteristics of PT modifications in the context of the genomic landscape, we developed two highly novel approaches: ICDS and RT-IC-Seq, which integrate the integration of PT-specific iodine-induced cleavage with high-throughput next-generation sequencing technology. As illustrated in Fig. 1, the ICDS technology involves selective cleavage of PT-modified site by iodine, subsequent marking the break sites with special labels and two-generation sequencing. This approach has been used to determine the distribution and localization of genome PT modification sites. Because of enriched amplicons with PT modifications at the PCR amplification step, this method can only be used for determining PT sites but not for quantification. As a complement to ICDS technology, the PT-IC-Seq method has higher throughput and resolutions to quantitatively determine the frequency of PT modification on the genome. In the PT-IC-Seq, the gDNA was selectively cleaved at of PT-modified site by iodine, followed by tag ligations and high-throughput sequencing. No enrichment process was involved during the whole process, so the modification was quantified authentically. Unfortunately, neither the

ICDS nor the RT-IC-Seq method can be used to analyze strain possessing single-stranded PT modifications, such as *Vibrio cyclitrophicus* FF75. In fact, when the PT modification occur on a single stranded, the treatment of iodine couldn't introduce double-strand break at modified sites. So, ICDS and RT-IC-Seq methods are bistranded PT modifications specific methods. Even so, this limitation may not impact the application for PT modifications research, because almost all reported PT modifications occur on double-strand of DNA to date, except for a few bacterial such as the *Vibrio cyclitrophicus* FF75.

Conclusions

In summary, we describe a series of methods coupling PT-specific iodine-induced cleavage with high-throughput next-generation sequencing technologies, to identify and quantify PT modifications in the genome. Using the ICA method, we have achieved rapid detection of PT modification status in the genome. Moreover, in order to characterize the PT modification on whole genomic landscapes, the ICDS technology was developed. We have successfully applied it to identify features of this epigenetic mark at any one of genomic positions. Based on the success, a high-resolution map of locations for PT modified sites has been achieved. Furthermore, the PT-IC-Seq technology was developed on the basis of the ICDS method, which was able to quantitatively analyze the frequency of specific PT modified sites on the genome. Overall, the protocols not only pave a path to a better understanding of the biology of the PT modification, but also serve as a useful technique suitable for the investigation of PT modification related studies.

Abbreviations

PT

Phosphorothioate

R- M

S- Restriction-modification

ICA

Iodine-induced cleavage assay

ICDS

Deep sequencing of iodine-induced cleavage at PT site

PT-IC-Seq

Iodine-induced cleavage PT sequencing

Dnd

DNA degradation

gDNA

Genomic DNA

LC-MS/MS

Liquid chromatography-coupled tandem quadrupole mass spectrometry

OD

References

1. Weigle P, Raleigh EA. Biosynthesis and Function of Modified Bases in Bacteria and Their Viruses. *Chemical reviews*. 2016;116(20):12655-87.
2. Wang L, Chen S, Xu T, Taghizadeh K, Wishnok JS, Zhou X, et al. Phosphorothioation of DNA in bacteria by *dnd* genes. *Nature chemical biology*. 2007;3(11):709-10.
3. Wang L, Chen S, Vergin KL, Giovannoni SJ, Chan SW, DeMott MS, et al. DNA phosphorothioation is widespread and quantized in bacterial genomes. *Proc Natl Acad Sci U S A*. 2011;108(7):2963-8.
4. Ou HY, He X, Shao Y, Tai C, Rajakumar K, Deng Z. *dndDB*: a database focused on phosphorothioation of the DNA backbone. *PloS one*. 2009;4(4):e5132.
5. Zhou X, Deng Z, Firmin JL, Hopwood DA, Kieser T. Site-specific degradation of *Streptomyces lividans* DNA during electrophoresis in buffers contaminated with ferrous iron. *Nucleic Acids Res*. 1988;16(10):4341-52.
6. Zhou X, He X, Liang J, Li A, Xu T, Kieser T, et al. A novel DNA modification by sulphur. *Mol Microbiol*. 2005;57(5):1428-38.
7. Ray T, Weaden J, Dyson P. Tris-dependent site-specific cleavage of *Streptomyces lividans* DNA. *FEMS Microbiol Lett*. 1992;75(2-3):247-52.
8. Xu T, Liang J, Chen S, Wang L, He X, You D, et al. DNA phosphorothioation in *Streptomyces lividans*: mutational analysis of the *dnd* locus. *BMC Microbiol*. 2009;9:41.
9. Tong T, Chen S, Wang L, Tang Y, Ryu JY, Jiang S, et al. Occurrence, evolution, and functions of DNA phosphorothioate epigenetics in bacteria. *Proc Natl Acad Sci U S A*. 2018;115(13):E2988-e96.
10. You D, Wang L, Yao F, Zhou X, Deng Z. A novel DNA modification by sulfur: *DndA* is a NifS-like cysteine desulfurase capable of assembling *DndC* as an iron-sulfur cluster protein in *Streptomyces lividans*. *Biochemistry*. 2007;46(20):6126-33.
11. Yao F, Xu T, Zhou X, Deng Z, You D. Functional analysis of *spfD* gene involved in DNA phosphorothioation in *Pseudomonas fluorescens* Pf0-1. *FEBS letters*. 2009;583(4):729-33.
12. Chen F, Lin K, Zhang Z, Chen L, Shi X, Cao C, et al. Purification, crystallization and preliminary X-ray analysis of the *DndE* protein from *Salmonella enterica* serovar Cerro 87, which is involved in DNA phosphorothioation. *Acta crystallographica Section F, Structural biology and crystallization communications*. 2011;67(Pt 11):1440-2.
13. Hu W, Wang C, Liang J, Zhang T, Hu Z, Wang Z, et al. Structural insights into *DndE* from *Escherichia coli* B7A involved in DNA phosphorothioation modification. *Cell research*. 2012;22(7):1203-6.
14. He W, Huang T, Tang Y, Liu Y, Wu X, Chen S, et al. Regulation of DNA phosphorothioate modification in *Salmonella enterica* by *DndB*. *Scientific reports*. 2015;5:12368.

15. Cao B, Zheng X, Cheng Q, Yao F, Zheng T, Ramesh Babu I, et al. In vitro analysis of phosphorothioate modification of DNA reveals substrate recognition by a multiprotein complex. *Scientific reports*. 2015;5:12513.
16. Cao B, Cheng Q, Gu C, Yao F, DeMott MS, Zheng X, et al. Pathological phenotypes and in vivo DNA cleavage by unrestrained activity of a phosphorothioate-based restriction system in *Salmonella*. *Mol Microbiol*. 2014;93(4):776-85.
17. Eckstein F, Gish G. Phosphorothioates in molecular biology. *Trends Biochem Sci*. 1989;14(3):97-100.
18. Braasch DA, Corey DR. Locked nucleic acid (LNA): fine-tuning the recognition of DNA and RNA. *Chem Biol*. 2001;8(1):1-7.
19. Gish G, Eckstein F. DNA and RNA sequence determination based on phosphorothioate chemistry. *Science*. 1988;240(4858):1520-2.
20. Nakamaye KL, Gish G, Eckstein F, Vosberg HP. Direct sequencing of polymerase chain reaction amplified DNA fragments through the incorporation of deoxynucleoside alpha-thiotriphosphates. *Nucleic Acids Res*. 1988;16(21):9947-59.
21. Mag M, Luking S, Engels JW. Synthesis and selective cleavage of an oligodeoxynucleotide containing a bridged internucleotide 5'-phosphorothioate linkage. *Nucleic Acids Res*. 1991;19(7):1437-41.
22. Alexeeva EV, Shpanchenko OV, Dontsova OA, Bogdanov AA, Nierhaus KH. Interaction of mRNA with the *Escherichia coli* ribosome: accessibility of phosphorothioate-containing mRNA bound to ribosomes for iodine cleavage. *Nucleic Acids Res*. 1996;24(12):2228-35.
23. Shpanchenko OV, Dontsova OA, Bogdanov AA, Nierhaus KH. Structure of 5S rRNA within the *Escherichia coli* ribosome: iodine-induced cleavage patterns of phosphorothioate derivatives. *RNA (New York, NY)*. 1998;4(9):1154-64.
24. Blanusa M, Schenk A, Sadeghi H, Marienhagen J, Schwaneberg U. Phosphorothioate-based ligase-independent gene cloning (PLICing): An enzyme-free and sequence-independent cloning method. *Analytical biochemistry*. 2010;406(2):141-6.
25. Cao B, Chen C, DeMott MS, Cheng Q, Clark TA, Xiong X, et al. Genomic mapping of phosphorothioates reveals partial modification of short consensus sequences. *Nat Commun*. 2014;5:3951.
26. Zheng T, Jiang P, Cao B, Cheng Q, Kong L, Zheng X, et al. DndEi Exhibits Helicase Activity Essential for DNA Phosphorothioate Modification and ATPase Activity Strongly Stimulated by DNA Substrate with a GAAC/GTTC Motif. *The Journal of biological chemistry*. 2016;291(3):1492-500.
27. Cheng Q, Cao B, Yao F, Li J, Deng Z, You D. Regulation of DNA phosphorothioate modifications by the transcriptional regulator DptB in *Salmonella*. *Mol Microbiol*. 2015;97(6):1186-94.
28. Ahlgren NA, Chen Y, Needham DM, Parada AE, Sachdeva R, Trinh V, et al. Genome and epigenome of a novel marine Thaumarchaeota strain suggest viral infection, phosphorothioation DNA modification and multiple restriction systems. *Environ Microbiol*. 2017;19(6):2434-52.

29. Li J, Chen Y, Zheng T, Kong L, Zhu S, Sun Y, et al. Quantitative mapping of DNA phosphorothioatome reveals phosphorothioate heterogeneity of low modification frequency. *PLoS genetics*. 2019;15(4):e1008026.
30. Carell T, Brandmayr C, Hienzsch A, Muller M, Pearson D, Reiter V, et al. Structure and function of noncanonical nucleobases. *Angewandte Chemie (International ed in English)*. 2012;51(29):7110-31.
31. Jones PA. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nature reviews Genetics*. 2012;13(7):484-92.
32. Chen K, Zhao BS, He C. Nucleic Acid Modifications in Regulation of Gene Expression. *Cell Chem Biol*. 2016;23(1):74-85.
33. Jaenisch R, Bird A. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat Genet*. 2003;33 Suppl:245-54.
34. Li E, Beard C, Jaenisch R. Role for DNA methylation in genomic imprinting. *Nature*. 1993;366(6453):362-5.
35. Robertson KD. DNA methylation and human disease. *Nature reviews Genetics*. 2005;6(8):597-610.

Declarations

Acknowledgements

We thank Wei Zhang at Shanghai Jiao Tong University for her help with the Mass Spectrometry analysis, and Dr Jaquelyn Fleckenstein at University of Tennessee Health Science Center for supplying *E. coli* strain B7A.

Funding

This work was supported by grants from the National Natural Science Foundation of China (31630002, 31770038, 31470183, 31700029, 21661140002 and 31170085); National Key R&D Program of China (2018YFA0900400) from the Ministry of Science and Technology; Shanghai Pujiang Program from the Shanghai Municipal Council of Science and Technology (12PJD021); China Postdoctoral Science Foundation (2017M620151).

Availability of Data and Materials

Datasets and materials are available from the corresponding author on reasonable request.

Author information

Affiliations

State Key Laboratory of Microbial Metabolism, Joint International Research Laboratory of Metabolic and Developmental Sciences, and School of Life Sciences & Biotechnology, Shanghai Jiao Tong University, Shanghai 200030, China

Contributions

SCZ and DLY designed the research. SCZ performed the experiments. TZ, L XK, JLL, YHS, YC, BC and ZXD analyzed and interpreted the data. SCZ, L XK and DLY wrote the manuscript. All authors read and approved the final manuscript.

Corresponding author

Correspondence to Delin You.

Ethics declarations

Not Applicable.

Ethics Approval

Not Applicable.

Consent for Publication

Not applicable.

Competing Interests

The authors declare that they have no competing interest.

Author details

¹State Key Laboratory of Microbial Metabolism, Joint International Research Laboratory of Metabolic and Developmental Sciences, and School of Life Sciences & Biotechnology, Shanghai Jiao Tong University, Shanghai, China

²College of Life Sciences, Qufu Normal University, Qufu, Shandong, China

Figures



Figure 1

Overview of the PT-specific iodine-induced cleavage related methods. Depiction of the key steps in the three experimental protocols from PT modifications identification to quantitative analysis.



Figure 2

The detection of PT modification. (A) ICA analysis of gDNA. Lanes 1-2, genomic DNA from E. coli B7A with (+) or without (-) iodine. Lanes 3-4, genomic DNA from E. coli DH10B with (+) or without (-) iodine. (B) and (C) LC-MS/MS analysis of the PT-linked dinucleotides from E. coli B7A and from E. coli DH10B. The PT-modified dinucleotide (GPSA or GPST) was used as standards.



Figure 3

The standard curve of (A) T and (B) C.



Figure 4

The standard curve of (A) GPSA and (B) GPST.



Figure 5

Quantitative analysis of genomic PT modification by LC-MS/MS. (A) and (B) LC-MS/MS analysis of the PT-linked dinucleotides from E. coli B7A. (C) HPLC analysis of gDNA from E. coli B7A. (D) The number of PT-modified dinucleotides per 106 nt of DNA in E. coli B7A. Error bars are calculated as the s.d. of three independent biological replicates.