A recent work by Jin *et al.* demonstrated the conversion of monomeric 5-methyldeoxycytidine (5mdC) to 5-formyldeoxycytidine (5fdC) through a photocatalytic pathway. The oxidation reaction was carried out in the presence of DMSO and also required an oxygen atmosphere. These reactions conditions are not compatible with applications on polynucleotides, such as DNA and RNA.

The present inventors have established an alternative method for the detection of 5mC and/or 5hmC in a polynucleotide.

Summary

In a general example the present disclosure provides a method for oxidising a polynucleotide containing a 5-methylcytosine (5mC) residue and/or a 5-hydroxymethylcytosine (5-hmC) residue. The oxidation product comprises a 5-formylcytosine (5fC) residue.

The oxidation method of the present disclosure is non-enzymatic and is carried out in the absence of an enzyme, such as a TET enzyme. Enzymatic methods of converting modified cytosine residues in a polynucleotide, such as through use of TET enzymes, can lead to sequence-specific biases, and in particular a bias to modified cytosine residues in a CpG context. Further, TET enzymes may oxidise 5mC or 5hmC residues in a polynucleotide to form 5caC as the major oxidation product, and therefore polynucleotides containing other oxidation products such as 5fC cannot be obtained using this method.

The present inventors have devised methods that allow the modified cytosine residues, 5mC and 5hmC, to be distinguished from canonical cytosine residues. The method can be performed on a nucleobase, or on a nucleoside, a nucleotide, or a polynucleotide comprising 5mC and/or 5hmC residues.

In a first aspect, the disclosure provides a method of identifying a modified cytosine residue in a sample nucleotide sequence, the method comprising

(i) providing a population of polynucleotides which comprise the sample nucleotide sequence;

(ii) oxidising the modified cytosine residue in the population to form a 5-formylcytosine (5fC) residue through a non-enzymatic, one-electron process;

(iii) labelling the 5-formylcytosine (5fC) residue; and

(iv) identifying the labelled residue within the population,

wherein the modified cytosine residue is selected from a 5-methylcytosine (5mC) residue and a 5-hydroxymethylcytosine (5hmC) residue.

During step (ii), the modified cytosine residue is oxidised at the carbon that is attached to the C5 position of the pyrimidine ring. The oxidation process forms a 5-formylcytosine (5fC) residue.

In some examples, the modified cytosine residue is a 5-methylcytosine (5mC) residue. In other examples, the modified cytosine residue is a 5-hydroxymethylcytosine (5hmC) residue.

The one-electron process includes a radical process and involves the generation of a radical. The one-electron process may involve hydrogen atom transfer (HAT) or single-electron transfer (SET).

The aldehyde group of 5fC provides a reactive handle for labelling during step (iii). Methods of functionalising 5fC through the aldehyde group are known in the art, including methods described in Raiber *et al.*, McInroy *et al.*, and US 2020/165661.

Advantageously, the conditions for oxidising the modified cytosines are suitable for use with polynucleotides. The oxidation reaction proceeds in a solvent system in which a polynucleotide is soluble. The reaction conditions, including the reaction temperature and pH, are compatible with polynucleotides and are selected to minimise polynucleotide degradation, such that a substantial amount of polynucleotides can be recovered for downstream analysis following oxidation. This is demonstrated on model oligodeoxyribonucleotides in the examples below.

Accordingly, the present inventors have devised methods that allow 5mC and 5hmC to be selectively targeted in the presence of canonical nucleobases within a polynucleotide. The oxidation product comprises 5fC, which is then labelled in step (iii). The labelled residue can subsequently be detected in step (iv) to identify the modified cytosine residue within the population of polynucleotides. The labelling may be by introduction of a detection tag or an isolation tag. The labelling may convert the 5fC to a residue having a different base-pairing pattern to cytosine, such as a uracil or thymine analogue, which can be subsequently detected by amplifying and/or sequencing the polynucleotide.

The oxidation in step (ii) may be performed in the absence of a TET enzyme, such as in the absence of an enzyme selected from TET1, TET2, and TET3.

Step (ii) may comprise oxidation of the modified cytosine residue in the presence of a radical initiator, to form a 5-formylcytosine (5fC residue). The radical initiator may be a metal-oxo species. In some examples, the oxidation in step (ii) may be performed in the presence of a radical initiator that is a photocatalyst, irradiating light, and water, and optionally a single-electron oxidant.

The photocatalyst may have an absorbance maximum in the range 300 nm to 600 nm. That is, the photocatalyst may absorb light in this range to form an excited state. In this way, the oxidation reaction may proceed in the presence of near-ultraviolet (UV) or visible light range

and does not require the use of short wavelength UV light (e.g. less than 300 nm), which may damage polynucleotides.

The photocatalyst may be an organic photocatalyst or a transition metal photocatalyst. Preferably, the photocatalyst is a transition metal photocatalyst, and more preferably the photocatalyst comprises a metal-oxo group.

Examples of a photocatalyst include polyoxometalates, such as tungsten polyoxometalates. Preferably, the photocatalyst is selected from decatungstic acid, phosphotungstic acid, and a salt thereof, and more preferably the photocatalyst is decatungstic acid or a salt thereof.

Step (ii) of the method may be performed in the presence of a single-electron oxidant. This helps to accelerate the oxidation reaction, particularly when the oxidation is performed in the presence of a photocatalyst, such that a good yield of 5fC is obtained before substantial degradation of the polynucleotide begins to occur. Preferably, the single-electron oxidant is an organic single-electron oxidant, such as *N*-fluorobenzenesulfonimide, 5-(trifluoromethyl)dibenzothiophenium tetrafluoroborate, and *N*-chlorosaccharin.

Step (iii) may comprise labelling the 5fC residue with a detection tag or an isolation tag. A detection tag may comprise a chromophore, a fluorescent label, a phosphorescent label or a radiolabel. An isolation tag may comprise a moiety that binds to a binding agent. The moiety that binds to a binding agent may be biotin. Labelling the 5fC residue in this way allows the polynucleotide comprising the modified cytosine to be identified within the population of polynucleotides, by methods that are well-known in the art.

Preferably, step (iii) comprises labelling the 5fC residue to alter the Watson-Crick base pairing pattern of the 5fC. For example, a nucleophilic probe may be introduced to the 5fC residue to form a derivatised residue having a different base-pairing pattern compared to cytosine. Preferably, the labelled residue is a uracil analogue. Examples of a suitable nucleophilic probe for this labelling include 1,3-indandione and malononitrile. When the 5fC residue is labelled in this way to alter its base-pairing pattern, the position of the 5fC residue may be identified by sequencing of the polynucleotide population.

The labelling of the 5fC residue in step (iii) may comprise deaminating the oxidised residue at the C4 position. Deamination of 5fC forms 5-formyluracil (5fU). The deaminated residue is thus a uracil analogue, and the base-pairing pattern is changed from that for cytosine. This change in base-pairing pattern allows the location of the modified cytosine residue to be identified within the population, such as by sequencing.

The deamination in step (iii) may also be accompanied by reduction of the residue, such as reduction of the pyrimidine ring. The deamination may be performed after the reduction. For

example, the 5fC residue may be reduced and then deaminated to form dihydrouracil (DHU). Methods for this transformation are described in WO 2019/136413.

Step (iv) may comprise the steps of:

(a) sequencing the polynucleotides in the population following step (iii) to produce a treated nucleotide sequence; and

(b) identifying the residue in the treated nucleotide sequence which corresponds to a modified cytosine residue in the sample nucleotide sequence.

This allows the location of the modified cytosine residue in the sample nucleotide sequence to be detected by sequencing, such as next-generation sequencing.

The polynucleotide may be DNA or RNA, or a mixture thereof.

The method of oxidising a 5mC or 5hmC residue provides 5fC in good yield. The method is thus advantageous over oxidation methods involving TET enzymes. Typically, TET enzymes oxidise 5mC residues in a polynucleotide to produce a mixture of 5hmC, 5fC and 5caC residues. When the oxidation is performed using a large excess of TET enzymes to the substrate, 5caC is formed as the major oxidation product. For example, Liu *et al.* report that the oxidation product of 5mC in a polynucleotide using *Naegleria* TET-like oxygenase (NgTET1) is almost entirely 5caC, with a 5fC yield of only 3%. Thus, the methods of the present disclosure can also be incorporated into a method of oxidising 5-methylcytosine (5mC) or 5-hydroxymethylcytosine (5hmC) residues in a polynucleotide to form 5fC residues in good yield.

In a second aspect, the disclosure provides a method of oxidising modified cytosine residues in a sample nucleotide sequence, the method comprising;

(i) providing a population of polynucleotides which comprise the sample nucleotide sequence,

(ii) oxidising the modified cytosine residues in the population to form 5-formylcytosine (5fC) residues, wherein the product mole ratio of 5-formylcytosine (5fC) residues to modified cytosine residues is 10:90 or more,

(iii) optionally labelling the 5-formylcytosine (5fC) residues, and

(iv) optionally identifying the labelled 5-formylcytosine (5fC) residues within the population,

wherein the modified cytosine residues are 5-methylcytosine (5mC) residues or 5-hydroxymethylcytosine (5hmC) residues.

The preferred features of the first aspect apply equally to the second aspect.

Preferably, the product mole ratio of 5fC to modified cytosine residues (i.e. either 5mC residues or 5hmC residues) in step (ii) is 20:80 or more, such as 30:70 or more.

The reaction product in step (ii) may be substantially free of oxidation products other than 5fC, such as 5hmC and 5caC. Where the modified cytosine residues provided in the population are 5mC residues, the mole ratio of 5fC product formed in step (ii) to 5hmC and/or 5caC may be 2:1 or higher, such as 5:1 or higher, such as 10:1 or higher, such as 50:1 or higher, such as 100:1 or higher. Where the modified cytosine residues provided in the population are 5hmC residues, the mole ratio of 5fC product formed in step (ii) to 5caC may be 2:1 or higher, such as 5:1 or higher, such as 10:1 or higher, such as 50:1 or higher, such as 100:1 or higher.

In an example, the disclosure provides a method of modifying a polynucleotide, the method comprising oxidising a 5-methylcytosine (5mC) residue and/or a 5-hydroxymethylcytosine (5hmC) residue in the polynucleotide through a non-enzymatic, one-electron process to form a 5-formylcytosine (5fC) residue.

The preferred features of the first aspect apply equally to the example.

In an example, the disclosure provides a method of oxidising 5-methylcytosine (5mC) or 5-hydroxymethylcytosine (5hmC), the method comprising oxidising the 5-methylcytosine (5mC) or 5-hydroxymethylcytosine (5hmC) through a non-enzymatic, one-electron process to form 5-formylcytosine (5fC).

The preferred features of the first aspect apply equally to the example.

In another aspect, the disclosure provides a method of oxidising a 5-methylcytosine (5mC) residue or a 5-hydroxymethylcytosine (5hmC) residue in a nucleoside, nucleotide or polynucleotide through a non-enzymatic, one-electron process to form a 5-formylcytosine (5fC) residue.

The preferred features of the first aspect apply equally to the aspect.

In yet another aspect, the present disclosure refers to use of a non-enzymatic radical initiator to oxidise a 5-methylcytosine (5mC) residue or a 5-hydroxymethylcytosine (5hmC) residue in a polynucleotide to form a 5-formylcytosine (5fC) residue. In an sexample, the disclosure provides use of a non-enzymatic radical initiator to oxidise a 5-methylcytosine (5mC) residue or a 5-hydroxymethylcytosine (5hmC) residue in a polynucleotide. The radical initiator may be a photocatalyst, which may be used in the presence of irradiating light, water, and optionally a single-electron oxidant.

The preferred features of the first aspect apply equally to the example.

In a further aspect, the disclosure provides a kit for use in a method described herein, comprising;

(a) a radical initiator, such as a photocatalyst, such as a polyoxometalate;

(b) a polymerase, and optionally.

(c) a single-electron oxidant, such as an organic single-electron oxidant, such as a compound selected from *N*-fluorobenzenesulfonimide, 5-(trifluoromethyl)dibenzothiophenium tetrafluoroborate, and *N*-chlorosaccharin.

These and other aspects and examples of the disclosure are described in further detail below.

Summary of the Figures

The present invention is described herein with reference to the figures listed below.

Figure 1 shows the results of a kinetic study of oxidising a sample comprising equimolar amounts of 5-methyldeoxycytidine, deoxyadenosine, deoxycytidine, deoxyguanosine and deoxythymidine. The solution was oxidised in the presence of 5 mol% $Na_4W_{10}O_{32}$ and 4 mM NFSI, in a 1:9 mixture of DMSO and water. The sample was irradiated at 365 nm, and the reaction was followed by LCMS over a reaction time of 3 hours.

Figure 2 shows the sequencing results for a 100mer single-stranded DNA model (5mC-100mer) using the method of the present invention. The signals obtained for 5mC and C positions in the 5mC-100mer are shown. At position 28, which corresponds to 5mC, 32% of reads were observed as thymine, i.e. a 5mC-to-T conversion of 32%. For the remaining positions, essentially all reads were observed as C.

Figure 3 shows the non-specific mutation rate observed for a 100mer ssDNA (5mC-100mer) using the method of the present invention.

Detailed Description of the Invention

The present invention provides a method for oxidising a polynucleotide containing a 5-methylcytosine (5mC) residue and/or a 5-hydroxymethylcytosine (5-hmC) residue. The oxidation product comprises a 5-formylcytosine residue.

Osberger *et al.* describe methods of using Fe complexes to selectively oxidise a C-H bond to a carbonyl group in amino acids and peptides by generating an Fe(IV)-oxo species *in situ*, which is also reviewed in White *et al.* It is not disclosed that this system can be applied to nucleosides, such as 5-methylcytosine. Also, the use of a strong oxidant such as H_2O_2 to generate the Fe(IV)-oxo species is described, which can degrade nucleosides and polynucleotides such as by depurination.

Jonasson *et al.* describe a method of oxidising a sample of 5mC to a mixture containing 5hmC, 5fC and 5caC. It is not disclosed that the method is suitable for use on 5mC when present as a residue in a polynucleotide. Further, the mixture of products formed by this

Claims:

1. A method of identifying a modified cytosine residue in a sample nucleotide sequence, the method comprising

(i) providing a population of polynucleotides which comprise the sample nucleotide sequence;

(ii) oxidising the modified cytosine residue in the population to form a 5-formylcytosine (5fC) residue through a non-enzymatic, one-electron process;

(iii) labelling the 5-formylcytosine (5fC) residue; and

(iv) identifying the labelled residue within the population,

wherein the modified cytosine residue is selected from a 5-methylcytosine (5mC) residue and a 5-hydroxymethylcytosine (5hmC) residue.

2. The method according to claim 1, wherein step (ii) is performed in the presence of a radical initiator.

3. The method according to claim 2, wherein the radical initiator is a metal-oxo species.

4. The method according to any one of claims 1 to 3, wherein in step (ii) the population of polynucleotides is irradiated with light in the presence of a photocatalyst and water.

5. The method according to claim 4, wherein the photocatalyst has an absorbance maximum in the range 300 nm to 600 nm.

6. The method according to claim 4 or claim 5, wherein the photocatalyst is a polyoxometalate.

7. The method according to claim 6, wherein the polyoxometalate comprises tungsten.

8. The method according to any one of claims 4 to 7, wherein the photocatalyst is selected from decatungstic acid, phosphotungstic acid, and a salt thereof.

9. The method according to any one of claims 1 to 8, wherein step (ii) is performed in the presence of a single-electron oxidant.

10. The method according to claim 9, wherein the single-electron oxidant is an organic single-electron oxidant.

11. The method according to claim 9 or claim 10, wherein the single-electron oxidant is selected from *N*-fluorobenzenesulfonimide, 5-(trifluoromethyl)dibenzothiophenium tetrafluoroborate, and *N*-chlorosaccharin.

12. The method according to any one of claims 1 to 11, wherein step (iii) comprises: (a) converting the 5-formylcytosine (5fC) residue to a uracil analogue, such as by reaction with a nucleophile;

(b) labelling the 5-formylcytosine (5fC) residue with a detection tag or an isolation tag;

(c) labelling the 5-formylcytosine (5fC) residue with an isolation tag, such as an isolation tag comprising biotin; and/or

(d) deaminating the 5-formylcytosine (5fC) residue at the C4 position, and optionally reducing the 5-formylcytosine (5fC) residue, such as reducing the pyrimidine ring.

13. The method according to any one of claims 1 to 12, wherein step (iv) comprises the steps of:

(a) sequencing the polynucleotides in the population following step (iii) to produce a treated nucleotide sequence; and

(b) identifying the residue in the treated nucleotide sequence which corresponds to a modified cytosine residue in the sample nucleotide sequence.

14. The method according to any one of claims 1 to 13, wherein the modified cytosine residue is a 5-methylcytosine (5mC) residue.

15. A method of oxidising modified cytosine residues in a sample nucleotide sequence, the method comprising;

(i) providing a population of polynucleotides which comprise the sample nucleotide sequence,

(ii) oxidising the modified cytosine residues in the population to form 5-formylcytosine (5fC) residues, wherein the product mole ratio of 5-formylcytosine (5fC) residues to modified cytosine residues is 10:90 or more,

(iii) optionally labelling the 5-formylcytosine (5fC) residues, and

(iv) optionally identifying the labelled 5-formylcytosine (5fC) residues within the population,

wherein the modified cytosine residues are 5-methylcytosine (5mC) residues or 5-hydroxymethylcytosine (5hmC) residues.

16. The method according to claim 15, wherein the product mole ratio of 5-formylcytosine (5fC) residues to 5-hydroxymethylcytosine (5hmC) residues and/or 5-carboxylcytosine (5caC) residues is 2:1 or higher.

17. A method of oxidising a 5-methylcytosine (5mC) residue or a 5-hydroxymethylcytosine (5hmC) residue in a nucleoside, nucleotide or polynucleotide through a non-enzymatic, one-electron process to form a 5-formylcytosine (5fC) residue.

18. Use of a non-enzymatic radical initiator to oxidise a 5-methylcytosine (5mC) residue or a 5-hydroxymethylcytosine (5hmC) residue in a polynucleotide to form a 5-formylcytosine (5fC) residue.

19. A kit for use in a method according to any of claims 1 to 14, comprising;

(a) a radical initiator, such as a photocatalyst, such as a polyoxometalate;

(b) a polymerase; and optionally

(c) a single-electron oxidant, such as an organic single-electron oxidant, such as a compound selected from *N*-fluorobenzenesulfonimide, 5-(trifluoromethyl)dibenzothiophenium tetrafluoroborate, and *N*-chlorosaccharin.