



Published in final edited form as:

*Chem Commun (Camb)*. ; 59(58): 8997–9000. doi:10.1039/d3cc02699j.

## Sequencing for oxidative DNA damage at single-nucleotide resolution with click-code-seq v2.0

Songjun Xiao<sup>a</sup>, Aaron M. Fleming<sup>a</sup>, Cynthia J. Burrows<sup>a,\*</sup>†

<sup>a</sup>Dept. of Chemistry, University of Utah, 315 S. 1400 East, Salt Lake City, UT 84112-0850, United States.

### Abstract

Oxidative damage to DNA nucleotides has many cellular outcomes that could be aided by the development of sequencing methods. Herein, the previously reported click-code-seq method for sequencing a single damage type is redeveloped to enable the sequencing of many damage types by making simple changes to the protocol (i.e., click-code-seq v2.0).

Oxidatively-derived damage to DNA occurs by reactive oxygen species (ROS) formed endogenously during metabolism and the inflammatory response or exogenously when cells are exposed to ionizing radiation.<sup>1</sup> The cellular milieu is a reducing context with bicarbonate buffer that can influence the chemistry of the ROS and products formed from DNA oxidation.<sup>2</sup> The guanosine (dG) nucleotide in DNA is the most electron-rich and is a dominant site of oxidation to yield the 8-oxo-7,8-dihydroguanosine (dOG) analog intracellularly.<sup>3,4</sup> Oxidation reactions can also occur on other bases, such as thymine to yield thymine glycol (dTg).<sup>5</sup> A long-standing question has been, where are these modifications formed in the genome? To answer this question for each oxidatively derived damage, methodologies must be developed that can sequence the DNA beyond reading the locations of the dA, dC, dG, and dT nucleotides.<sup>6</sup> The biological payoff from developing and employing such techniques is a better understanding of the role of site-specific oxidative DNA damage causing mutations found in a variety of diseases.<sup>7</sup> DNA oxidation at dG nucleotides may also function as an epigenetic-like regulator for the cellular response to stress, and knowing the sites of genomic dOG is essential for advancing this hypothesis.<sup>8</sup>

Many methods for sequencing genomic oxidation sites have been reported. These methods fall into two general categories: The first are those that enrich fragmented DNA for strands containing the oxidation site, after which sequencing reports the location of the oxidation site at the resolution of the average fragment length.<sup>5,9–12</sup> The second are those that introduce signatures at the site of modification, which enables single-nucleotide resolution of the modification sites.<sup>13–17</sup>

†Electronic Supplementary Information (ESI) available: complete method details, data for optimization of each step, phosphoramidite synthesis protocol and characterization, and modified oligonucleotide synthesis and characterization. See DOI: [10.1039/x0xx00000x](https://doi.org/10.1039/x0xx00000x)

\*To whom correspondence should be addressed: burrows@chem.utah.edu.

Conflicts of interest

There are no conflicts to declare.

A sub-category of the nucleotide-resolution methods is those that hijack the natural base excision repair (BER) pathway to utilize the specificity of a DNA glycosylase for removal of the damaged nucleotide followed by the installation of a reporter group to be found during sequencing (Fig. 1A).<sup>13,14,17</sup> Two reports that are background for the present report will be discussed. In 2015, the Burrows laboratory first demonstrated this approach by replacing the damaged nucleotide with a PCR-amplifiable unnatural base pair (UBP = dNaM:d5SICS,<sup>18</sup> Fig. 1B).<sup>13</sup> Sequencing for the dOG heterocycle was achieved by targeting the site with either the Fpg and OGG1 glycosylases, the latter being more selective than the former enzyme,<sup>19</sup> to replace the dOG with the UBP that was revealed by a sharp stop in a Sanger sequencing chromatogram. The universality of this approach was demonstrated by the selectivity of the NEIL1 glycosylase for the hyperoxidized dG base spiroiminodihydroantoin (dSp)<sup>20</sup> and by use of UDG<sup>21</sup> to label uracil (dU) resulting from dC amination in DNA. At the time, high-throughput sequencing technology was unable to locate an unnatural nucleotide in the DNA sequence; however, this could now change based on a recent report.<sup>22</sup> An improvement to this approach was reported in 2018 by the Sturla laboratory who showed that a BER-generated gap could be filled with a reporter sequence. Their method, “click-code seq,” used Fpg for the removal of dOG that was then replaced with a custom-synthesized 3'-O-propargyl-dGTP (Fig. 1C).<sup>14</sup> The installed alkyne group was then subjected to the click reaction (CuAAC) to append a 5'-azide-containing DNA strand of known sequence used to locate the dOG after sequencing. The Brown laboratory previously developed the 1,2,3-triazole linkage formed in this reaction and found it was compatible with polymerase bypass.<sup>23</sup> The click-code-seq protocol was employed on yeast genomic DNA;<sup>14</sup> however, the demand for the custom synthesis of a 3'-O-propargyl-dNTP imposes a barrier for application to damage at other DNA nucleotides.

In the present report, we were inspired by click-code-seq developed in the Sturla laboratory that allowed sequencing for dOG sites via high-throughput sequencing;<sup>14</sup> our goal was to adapt the methodology to function universally for targeting other types of DNA damage, similar to the approach using UBP to mark the damaged sites (Fig. 1).<sup>13</sup> To achieve the goal, the click partners were reversed in which the gap in the DNA formed by the glycosylase/endonuclease step was filled by a polymerase with a commercially-available 3'-azido-2',3'-dideoxynucleotide (3'-N<sub>3</sub>-ddNTP; Fig. 1D). The code DNA sequence was comprised of a 5'-alkynyl group formed by standard phosphoramidite synthesis of the monomer and solid-phase synthesis of the polymer of known sequence. We have termed the update in the methodology click-code-seq v2.0. The universal nature of the update results from all four 3'-N<sub>3</sub>-ddNTP being commercially available, and the code sequence needs to be synthesized only once and then applied to whichever damage-specific glycosylase and 3'-N<sub>3</sub>-ddNTP were selected. What follows is an outline of the methodology, a description of the synthetic protocol, optimizations conducted at each step, and application of this to sequence a dOG at a known site in a ~6,300-nt long plasmid DNA. Finally, we outline future DNA damage sites that could be sequenced by the method by simply changing the glycosylase and ordering a new 3'-N<sub>3</sub>-ddNTP from a commercial vendor.

The method reported was first developed and optimized on DNA oligomers made by standard solid-phase synthesis with dOG at a known location (Fig. S2, ESI†). The sequence selected for study flanks codon 12 of the *KRAS* gene in which there is an established G→T

transversion mutation found in many cancers that is also diagnostic of the mutation caused by dOG during replication.<sup>7</sup> The dOG-containing DNA strand was 5′-<sup>32</sup>P labelled and then mixed with the complementary strand to form duplex DNA (Figs. 2A and B lane 1). The chemical reactions were monitored by denaturing PAGE separation and phosphorimager autoradiography visualization (Fig. 2B).

The first step of the protocol is where selectivity in the DNA damage recognition occurs because BER glycosylases have evolved to target specific types of DNA lesions.<sup>19</sup> In the development of the method, the duplex DNA was treated with the bifunctional glycosylase Fpg that hydrolyses the damaged base from the sugar followed by catalyzing  $\beta,\delta$ -elimination of the phosphates to yield a 5′ fragment with a 3′ phosphate (Fig. 2 lane 2).<sup>19</sup> The 3′ phosphate must be removed prior to installation of the commercially available 3′-N<sub>3</sub>-ddGTP. In the original version of click-code-seq, this reaction was conducted with the endonuclease APE1;<sup>14</sup> however, we found the reaction is much more efficient using endonuclease IV (Endo IV; Figs. 2B lane 3 and S3, ESI†), as we previously reported.<sup>13</sup> The original click-code-seq protocol used Terminator DNA polymerase for installation of the 3′-O-propargyl-dGTP; however, in the present approach, we used the DNA polymerase Klenow fragment 3′→5′ exonuclease minus (Kf exo<sup>-</sup>) known to tolerate modified dNTPs,<sup>24</sup> which was found to install 3′-N<sub>3</sub>-ddGTP with nearly quantitative yield (Fig. 2B lane 4). The DNA product in this scenario has a 3′ azide on the DNA strand of interest, which is the reverse of the original click-code-seq having the alkyne at the same position and strand.

Production of the 5′-alkynyl-code DNA sequence first required synthesis of a 5′-alkynyl-dT phosphoramidite following a well-established protocol from the Brown laboratory (Fig. S4, ESI†).<sup>25</sup> The same laboratory also found the 1,2,3-triazole linkage formed after the click reaction is biocompatible and readily bypassed by a DNA polymerase.<sup>25</sup> Synthesis of phosphoramidites, especially on a dT nucleotide that does not require base protection, is a more tractable synthesis than reactions to furnish dNTPs particularly those with the G heterocycle.<sup>26</sup> The phosphoramidite formed was used in the solid-phase synthesis of the code DNA strand where it was installed as the final 5′ nucleotide (Fig. S5, ESI†). The code sequence was selected to minimize any off-target binding to the human genome (Fig. S2, ESI†). In the final step of the protocol, the click reaction was allowed to occur between the 3′-azide group installed where dOG previously resided and the 5′-alkynyl-code DNA sequence. The click reaction was optimized and found to reach yields >95% when the Cu:TBTA catalyst was used in a solution with 55% DMSO when allowed to react at 22 °C for 24 h (Figs. 2B lane 5 and S6, ESI†).

Demonstration of the protocol on a large DNA strand of biological origin was then conducted. Using a literature protocol,<sup>27</sup> a ~6300-nt long plasmid extracted from *E. coli* culture was subjected to a chemo-enzymatic synthesis method for installation of dOG at a site of our choosing (Figs. 3A and S7, ESI†). The *VEGF* promoter potential G-quadruplex forming sequence was selected for dOG introduction because this sequence, when oxidized in cellulo, induces mRNA synthesis via the epigenetic-like role of this

†Electronic Supplementary Information (ESI) available: complete method details, data for optimization of each step, phosphoramidite synthesis protocol and characterization, and modified oligonucleotide synthesis and characterization. See DOI: [10.1039/x0xx00000x](https://doi.org/10.1039/x0xx00000x)

modified nucleotide.<sup>27</sup> Employing the method as described on the dOG-containing plasmid, followed by direct PCR amplification of the click product using Phusion High-Fidelity DNA Polymerase, revealed the oxidation site upon Sanger sequencing (Fig. 3B). In the sequencing chromatogram, the plasmid sequence is observed with the *VEGF* sequence that is interrupted at the dOG site to continue with the code sequence introduced by the method. The PCR amplicon is a high-fidelity copy of the click conjugate, without any insertion and deletion of nucleotides. This validates the utility of the method on large DNA of biological origin where the site of oxidation is known. Studies in our laboratory routinely applied a gap-ligation method for verification of site selectively installed modification in plasmids;<sup>28</sup> a drawback to this approach were cases in which the modification resides in a dG run (e.g., a G-quadruplex forming sequence)—the data revealed the run was modified but not at which site. Click-code-seq has the advantage of sequencing modified plasmids to pinpoint the site of modification in poly-dG sequences (Fig. 3B). A detailed protocol for the method is provided in the Electronic Supplementary Information.

Herein, a modified version of click-code-seq (i.e., click-code-seq v2.0) is reported that retains the single-nt resolution of the original procedure and has the additional benefit of being applicable to more than one type of DNA damage. The requirements to target other modifications are met by the suite of BER glycosylases that target specific DNA lesions.<sup>19</sup> Using commercial 3'-N<sub>3</sub>-ddNTPs allows the gap formed by the glycosylase-endonuclease pair to be filled with the appropriate click reaction partner. Synthesis of one 5'-alkynyl-code DNA sequence allows targeting of any of the 3'-N<sub>3</sub>-ddNTPs enabling sequencing for the target modification of interest. Similar to click-code-seq, the update version of the method could map lesions that are ~20 nucleotides or more apart on the same strand to enable enough sequence of the target genome to be read for read alignment to the reference sequence. Our previous report that labelled three different types of DNA damage with an UBP demonstrates the feasibility of extending this approach to other lesions (Fig. 1B).<sup>13</sup> Furthermore, the original report by the Sturla laboratory demonstrated that the method can be deployed on the genomic scale.<sup>14</sup> Lastly, a recent report showed efficient polymerase bypass of a 1,2,3-triazole formed between a 3'-N<sub>3</sub>-ddG and a 5'-hexynyl terminated DNA strand;<sup>29</sup> this is relevant because both click reaction partners would then be commercially available ones, allowing future applications of this approach that avoid custom chemical synthesis altogether.

As a final note, the glycosylase introduces the selectivity in the DNA damage to be sequenced (Fig. 1A). While Fpg is commonly used for targeting dOG base paired with dC, this glycosylase also has one of the broadest substrate scopes and includes the dG modifications Fapy-dG paired with dC and the hyperoxidized dG lesions, dSp and dGh in any base pair context.<sup>30</sup> Greater selectivity for dOG paired with dC is achieved by using OGG1 as the glycosylase, but the Fapy-dG lesion paired with dC remains a substrate in either case.<sup>30,31</sup> Other DNA damage products that can be sequenced by the click-code-seq v2.0 method are provided in Table 1. These examples include dOG paired with dA using MutY,<sup>32</sup> dTg using NTHL1 or NEIL1,<sup>5,33,34</sup> thymine dimers (d(T<>T)) using T4 Endo V,<sup>35</sup> the hydantoins using NEIL1 or 3,<sup>20,36</sup> and dU by UDG.<sup>21</sup>

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

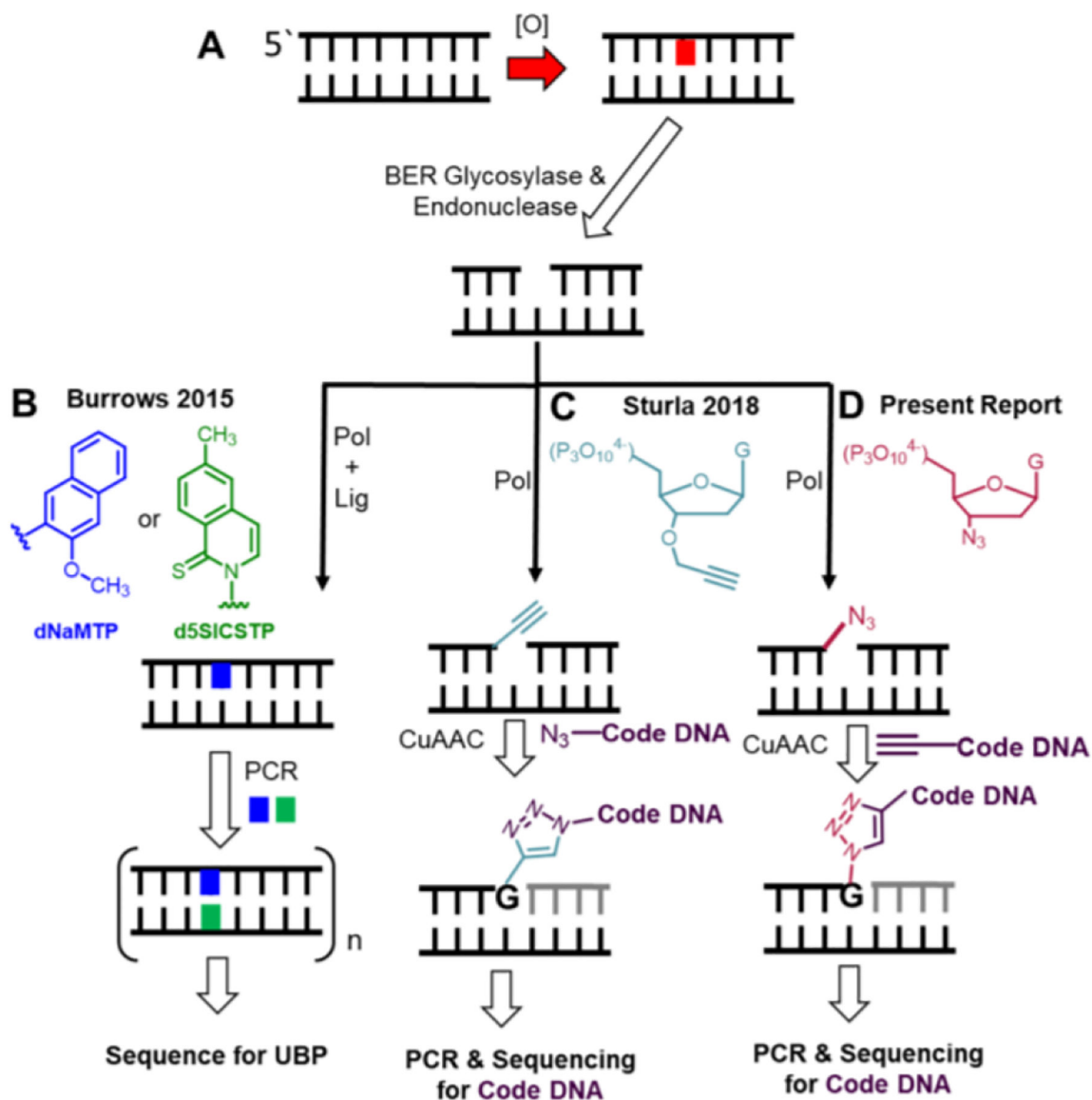
## Acknowledgments

We acknowledge financial support from the National Institute of General Medical Sciences grant no. R35 GM145237 and the University of Utah core facilities for synthesizing the DNA used in the studies and conducting the Sanger sequencing.

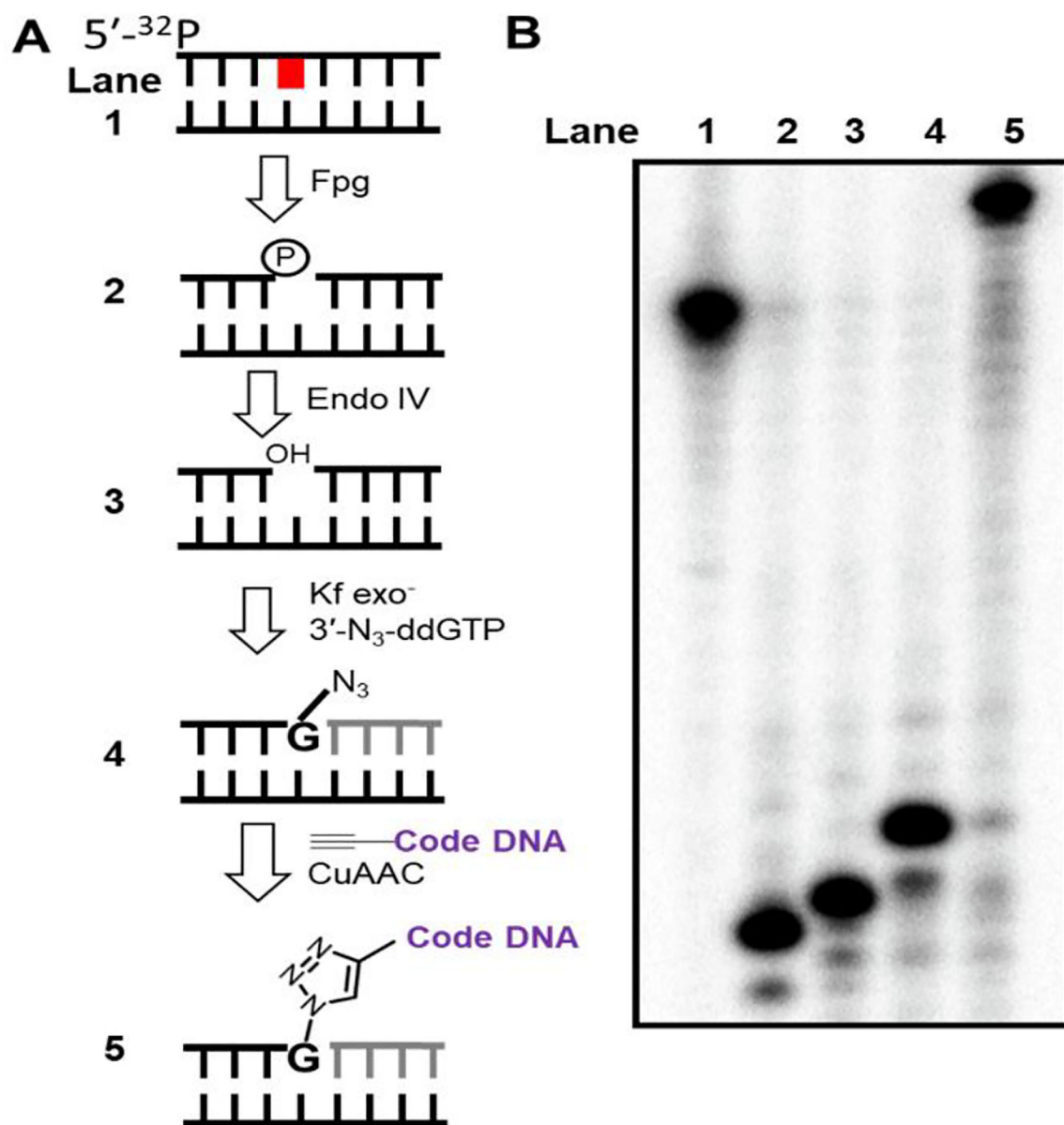
## Notes and references

1. Cadet J, Wagner RJ, Shafirovich V and Geacintov NE, *Int. J. Radiat. Biol.*, 2014, 90, 423. [PubMed: 24369822]
2. Fleming AM and Burrows CJ, *Chem. Soc. Rev.*, 2020, 49, 6524. [PubMed: 32785348]
3. Mangerich A, Knutson CG, Parry NM, Muthupalani S, et al., *Proc. Natl. Acad. Sci. U.S.A.*, 2012, 109, E1820. [PubMed: 22689960]
4. Gates K, *Chem. Res. Toxicol.*, 2009, 22, 1747. [PubMed: 19757819]
5. Tang F, Yuan J, Yuan B-F and Wang Y, *J. Am. Chem. Soc.*, 2022, 144, 454. [PubMed: 34978433]
6. Wu J, Sturla SJ, Burrows CJ and Fleming AM, *Chem. Res. Toxicol.*, 2019, 32, 345. [PubMed: 30807111]
7. Pfeifer G and Besaratinia A, *Hum. Genet.*, 2009, 125, 493. [PubMed: 19308457]
8. Fleming AM and Burrows CJ, *J. Am. Chem. Soc.*, 2020, 142, 1115. [PubMed: 31880930]
9. Ding Y, Fleming AM and Burrows CJ, *J. Am. Chem. Soc.*, 2017, 139, 2569. [PubMed: 28150947]
10. Poetsch AR, Boulton SJ and Luscombe NM, *Genome Biol.*, 2018, 19, 215. [PubMed: 30526646]
11. Amente S, Di Palo G, Scala G, Castrignanò T, et al., *Nucleic Acids Res.*, 2019, 47, 221. [PubMed: 30462294]
12. Fang Y and Zou P, *Biochemistry*, 2020, 59, 85. [PubMed: 31618020]
13. Riedl J, Ding Y, Fleming AM and Burrows CJ, *Nat. Commun.*, 2015, 6, 8807. [PubMed: 26542210]
14. Wu J, McKeague M and Sturla SJ, *J. Am. Chem. Soc.*, 2018, 140, 9783. [PubMed: 29944356]
15. An J, Yin M, Yin J, Wu S, et al., *Nucleic Acids Res.*, 2021, 49, 12252. [PubMed: 34788860]
16. Cao B, Wu X, Zhou J, Wu H, et al., *Nucleic Acids Res.*, 2020, 48, 6715. [PubMed: 32484547]
17. Gilat N, Fridman D, Sharim H, Margalit S, et al., *Biophys. Rep.*, 2021, 1, 100017.
18. Malyshev DA, Dhami K, Quach HT, Lavergne T, et al., *Proc. Natl. Acad. Sci. U. S. A.*, 2012, 109, 12005. [PubMed: 22773812]
19. David SS and Williams SD, *Chem. Rev.*, 1998, 98, 1221. [PubMed: 11848931]
20. Krishnamurthy N, Zhao X, Burrows CJ and David SS, *Biochemistry*, 2008, 47, 7137. [PubMed: 18543945]
21. Stivers JT and Drohat AC, *Arch. Biochem. Biophys.*, 2001, 396, 1. [PubMed: 11716455]
22. Ledbetter MP, Craig JM, Karadeema RJ, Noakes MT, et al., *J. Am. Chem. Soc.*, 2020, 142, 2110. [PubMed: 31985216]
23. El-Sagheer AH, Sanzone AP, Gao R, Tavassoli A, et al., *Proc. Natl. Acad. Sci. U.S.A.*, 2011, 108, 11338. [PubMed: 21709264]
24. Hottin A and Marx A, *Acc. Chem. Res.*, 2016, 49, 418. [PubMed: 26947566]
25. El-Sagheer AH and Brown T, *J. Am. Chem. Soc.*, 2009, 131, 3958. [PubMed: 19292490]
26. Burgess K and Cook D, *Chem. Rev.*, 2000, 100, 2047. [PubMed: 11749283]
27. Fleming AM, Ding Y and Burrows CJ, *Proc. Natl. Acad. Sci. U.S.A.*, 2017, 114, 2604. [PubMed: 28143930]
28. Riedl J, Fleming AM and Burrows CJ, *J. Am. Chem. Soc.*, 2015, 138, 491.
29. Schönegger ES, Crisp A, Müller M, Fertl J, et al., *Bioconjugate Chem.*, 2022, 33, 1789.

30. Krishnamurthy N, Haraguchi K, Greenberg MM and David SS, *Biochemistry*, 2008, 47, 1043. [PubMed: 18154319]
31. Leipold MD, Workman H, Muller JG, Burrows CJ, et al., *Biochemistry*, 2003, 42, 11373. [PubMed: 14503888]
32. Banda DM, Nunez NN, Burnside MA, Bradshaw KM, et al., *Free Radic. Biol. Med*, 2017, 107, 202. [PubMed: 28087410]
33. Yeo J, Goodman RA, Schirle NT, David SS, et al., *Proc. Natl. Acad. Sci. U. S. A*, 2010, 107, 20715. [PubMed: 21068368]
34. Carroll BL, Zahn KE, Hanley JP, Wallace SS, et al., *Nucleic Acids Res*, 2021, 49, 13165. [PubMed: 34871433]
35. Latham KA, Taylor J-S and Lloyd RS, *J. Biol. Chem*, 1995, 270, 3765. [PubMed: 7876117]
36. Liu M, Bandaru V, Bond JP, Jaruga P, et al., *Proc. Natl. Acad. Sci. U. S. A*, 2010, 107, 4925. [PubMed: 20185759]

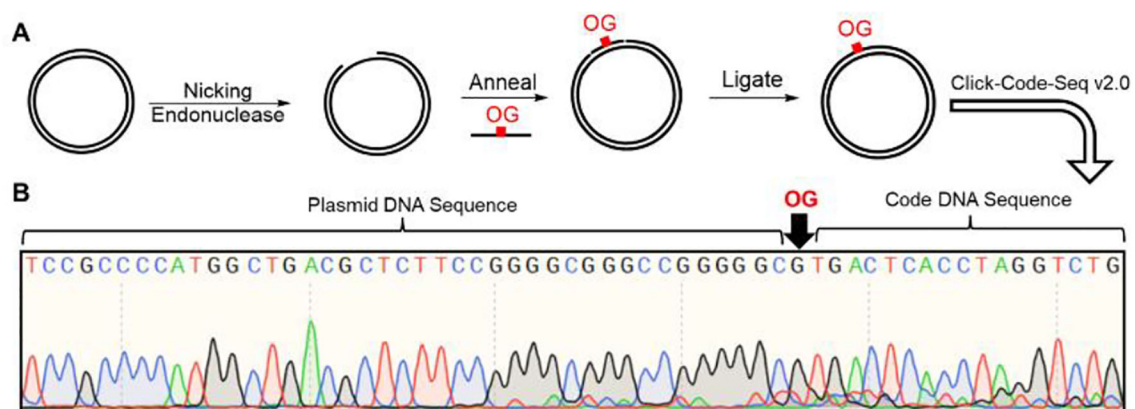


**Figure 1.** Prior (A, B, and C) BER methods for nucleotide-resolution mapping of oxidative damaged nucleotides that were inspirational for the (D) present method described. See Figure S1 (ESI<sup>†</sup>) for the complete structures of the 1,2,3-triazole linkages formed in panels C and D.



**Figure 2.**  
 (A) Overview of the method and (B) monitoring the reaction progress of click-code-seq v2.0 by PAGE and storage phosphor autoradiography.





**Figure. 3.** (A) DNA of a biological origin was chemo-enzymatically modified to contain a site-specific dOG, (B) which was used for validation of the methodology on a large substrate of known composition.

**Table 1.**

Glycosylases, the substrates they target, and the 3'-N<sub>3</sub>-ddNTP that can be used for sequencing the target substrates.

<b>BER Glycosylase</b>	<b>Substrate</b>	<b>3'-ddNTP</b>
Fpg	dOG:C, dFapy-G, dSp, and dGh	3'-N <sub>3</sub> -ddGTP
NEIL1-3	dSp and dGh or dTg	3'-N <sub>3</sub> -ddGTP or 3'-N <sub>3</sub> -ddTTP
UDG	dU	3'-N <sub>3</sub> -ddCTP
MutY	dOG:A	3'-N <sub>3</sub> -ddTTP
T4 Endo V	d(T<>T)	3'-N <sub>3</sub> -ddTTP
Endo IV	AP	3'-N <sub>3</sub> -ddNTP
Endo V	dI	3'-N <sub>3</sub> -ddATP

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript