This is an Accepted Manuscript, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this Accepted Manuscript with the edited and formatted Advance Article as soon as it is available.

You can find more information about Accepted Manuscripts in the Information for Authors.

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal’s standard Terms & Conditions and the Ethical guidelines still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this Accepted Manuscript or any consequences arising from the use of any information it contains.
While a methyl group is an inert moiety, pioneering work by bind its recognition sequence, ensuring site-specific labeling. Catalytic transfer will only occur after the enzyme successfully (N$_S$-adenosyl-L-methionine (SAM) cofactor onto adenine labeling is DNA methyltransferases (MTases) based labeling. In modifications methyltransferase based optical mapping is shown.

Site-specific DNA labeling holds the potential of unlocking promising advances in DNA-based applications. Through modifications of known positions on DNA, researchers are able to extract vital information about the genetic material, making site-specific DNA labeling an indispensable tool for species identification or medical diagnostics. In the last two decades, several approaches to such site-specific labeling have been developed, mainly relying on hairpin polyamides, triplex-helix-forming oligodeoxynucleotides or enzyme-directed DNA modifications.

Perhaps one of the most exciting examples of site-specific labeling is DNA methyltransferases (MTases) based labeling. In nature, these enzymes play a key role in the DNA methylation process by catalyzing the transfer of a methyl group from an S-adenosyl-L-methionine (SAM) cofactor onto adenine (N$_6$-position) or cytosine (N$_4$- or C$_5$-position). Generally, the catalytic transfer will only occur after the enzyme successfully binds its recognition sequence, ensuring site-specific labeling. While a methyl group is an inert moiety, pioneering work by Weinhold and coworkers shows that MTase enzymes tolerate non-natural cofactors to effectively transfer functionalized labels onto DNA, and such strategies have been developed to introduce (fluorescent) labels which allow for localization, capture, photolabile caging, photocrosslinking, selective scission, and is a fundamental technology in the field of optical mapping.

In variants of this technology, replacing the methionine side chain of SAM with an aziridine ring resulted in functional cofactors with a pending aziridine moiety. MTase enzymes are tricked into catalyzing the nucleophilic ring opening, in line with nucleophilic attack by the nucleobase, thus covalently coupling the whole cofactor onto DNA. By introducing a functional group on the nucleobase, these molecules are able to label DNA. This research was later extended by Rajski and co-workers, who used reactive nitrogen mustards for in situ aziridine generation.

However, during the enzymatic labeling, both research groups noticed a strong affinity of the labeled DNA for the enzymatic binding pocket, which inhibited further turnovers and thus forcing the use of stoichiometric amounts of enzyme to fully label DNA.

This downside prompted Weinhold et al. to develop more efficient cofactors, and a second class called the doubly activated cofactors was evaluated. Here, a double or triple bond is introduced at the β-position to the sulfonium center as part of the functionalized label which is transferred by the MTase enzymes. As only the label is transferred, this class is less affected by enzyme inhibition. The unsaturated bond stabilizes the transition state in the transfer reaction, and hence, no enzymatic transfer was observed for saturated labels.

Commonly, doubly activated cofactors are used to transfer a chemically reactive group, e.g. amine, azide or alkyne, followed by conjugation to a fluorescent dye. This class was rapidly adopted by the research community and extended with synthetic modifications introduced on the sulfonium center, the amino acid or the transferrable linker. While this method can be used to quickly introduce different labels, the conjugation of the dye often results in a lower yield. Hence, in...
pursuit of high degrees of fluorescent labeling, the research was continued to fluorescent SAM analogues which contain the fluorescent molecule and directly transfer this functionality to its substrate.\textsuperscript{11,33}

\textbf{Previous Work: Homocysteine based cofactors}

\textbf{This Work: Cysteine based cofactors}

\begin{center}
\begin{tabular}{c}
\textbf{Scheme 1} Formation of cofactors 2a and 2b. \\
\end{tabular}
\end{center}

\begin{center}
\begin{tabular}{c}
\textbf{Fig. 1} Summary of previous work and present work. \\
\end{tabular}
\end{center}

Unfortunately, the described pathways towards these non-natural cofactors are time consuming and often resulting in low yields. This is especially true for the final step, where $S$-adenosyl-L-homocysteine (SAH) is coupled with a highly electrophilic alkyl triflate or allyl bromide and a high excess of this linker (50-200 equivalents) is needed to drive the reaction towards completion (Fig. 1). As the reagent used in large excess actually contains the functionality introduced to the substrate later-on, this imposes significant restraints on the flexibility of the approach. Furthermore, even with this large excess, reported yields do not exceed 25% for a single epimer after HPLC purification. Efforts to increase the formation of the correct epimer by a chemoenzymatic synthesis pathway also failed to give reasonable yields for larger systems.\textsuperscript{31,34}

In our pursuit of advanced applications of enzymatic DNA labeling, we were interested in the development of doubly activated cofactors, unhindered by the synthetic issues discussed above. To this end, we developed and evaluated new cysteine based SAM analogues. Our findings show that they are easily accessible and possess comparable transfer behavior as their homocysteine counterparts. The late stage introduction of the amino acid moiety allows to reach functional MTase cofactors without sacrificing labor intensive molecules.

Our first efforts were aimed at the synthesis of cofactors which can be used in a two-step labeling protocol. In such schemes, a reactive functionality, e.g. azide or amine, is appended to the transferable group, for modification at either the cofactor stage or after transfer to the substrate. Thus, we started with the coupling of $S'$-thioadenosine\textsuperscript{33} and 6-azido-1-bromohex-2-yne. These starting materials are well described in literature and their synthesis can be carried out at gram scale.

The substitution reaction furnished thioether 1a in good yield, which is readily reduced to the amine substituted product 1b under Staudinger conditions. Much to our delight, these molecules are converted into the desired cofactors 2 in acidic media in the presence of one to four equivalents of the $\beta$-lactone (Scheme 1). In line with the soft nucleophilic character of the sulfur, the reaction with the strained $\beta$-lactone favors $\alpha$-carbon attack over nucleophilic acylation.\textsuperscript{36} For simple systems, conversion was quantitative without the formation of byproducts, and final removal of the protecting group proceeded smoothly after addition of a stronger acid. Though lactone salts (tosylate, tetrafluoroborate, trifluoroacetate) can provide the desired compounds directly upon reaction with the thioethers 1, reactions are sluggish, and the deprotection route is preferred. Due to the good conversion and low amount of byproducts formed, these cofactors can be used directly or with minimal purification (trituration or reverse phase silica gel filtration). While this is an effective and scalable method for cofactor synthesis, it should be noted that the formed cofactors are diastereomeric in nature, and the sulfonium centers are approximately a 50/50 mixture of R and S isomers. DiastereomERICALLY pure cofactors can be obtained from HPLC purification, though the inactive isomer does not hinder further applications. The cofactors can be stored for several months at low temperatures and low pH.

Shifting from homocysteine appended cofactors to cysteine containing systems will have an effect on the position of the cofactor within the enzymatic pocket. Fortunately, gel based DNA restriction analysis indicated that MTase enzymes tolerate cysteine based cofactors, though at increased concentration (Supporting Fig. S1-S3), proving that this pathway provides efficient access to azide and amine substituted functional cofactors.

Next, these cofactors were converted to fluorescent versions through coupling of their reactive handles with dyes. However, these reactions were always accompanied by severe sideproduct formation, attributed to high pH (amide coupling) or incompatible reagents (Cu-salts in click-chemistry). Thus, fluorescent dye attachment was effected before the actual introduction of the amino acid.
Here, it was postulated that enzyme binding might be influenced by the fluorescent molecules, which tend to be quite bulky and often carry charges. To counter these effects, we attempted to increase the distance between dye and cofactor. This distance can be acquired by adding aliphatic or oligoethylene glycol spacers. To evaluate the effect of the length and structure of the spacer, three different examples 3-5a were prepared: 1) no linker 2) an aliphatic extension of six carbons 3) extension with three ethylene glycol units, respectively. Chain extension proceeded smoothly when using N-Boc protected carboxylic acids and 1b. After removal of the protecting group, the free amine can be used in the next step without intermediate purification. Dye attachment was carried out through amide formation using similar reaction conditions as for the extension. In the presence of four equivalents of the β-lactone, the desired cofactors 3-5a (Fig. 2) were obtained in good conversions after reverse phase silica gel filtration.

To further extend this product portfolio of chain extended direct cofactors, the dye repertoire was expanded and a small library, spanning the visible spectrum, was prepared (Fig. 3, Full structures in Supp. information). To ensure solubility for cofactors conjugated to hydrophobic dyes, cofactor synthesis was started from the ethylene glycol extended molecule S2.

After successful synthesis of the fluorescent cofactors 3-5j, the compatibility in MTase based labeling was validated through gel based DNA restriction analysis (Fig. 4 and Supporting Fig. S4-S9). Interestingly, all extended cofactors 4-5 were accepted by the MTase enzyme M.TaqI. In a next step, the labeling performance was assessed through a counting assay (Supporting Fig. S10). Given their superior labeling performance, Rhodamine B containing cofactors 4 and 5a were selected to use in further optical mapping experiments. In short, DNA fragments of the bacteriophage lambda were labeled using the M.TaqI methyltransferase enzyme (recognition sequence 5’-TCGA-3’) and the fluorescent cofactor 4-5a. Labeled DNA is linearized on a zeonex coated coverslip selected to use in further optical mapping experiments. In short, DNA fragments of the bacteriophage lambda were labeled using the M.TaqI methyltransferase enzyme (recognition sequence 5’-TCGA-3’) and the fluorescent cofactor 4-5a. Labeled DNA is linearized on a zeonex coated coverslip.

**Fig. 3** General structure and normalized absorbance spectra of the synthesized fluorescent cofactors 5a-5j.

**Fig. 4** TaqI restriction assay on pUC19 DNA using oligonucleotide-treated M.TaqI (0.03 μg/μl) with 3, 4 and 5a. All samples were reacted with TaqI restriction enzyme unless stated otherwise. From left to right: GeneRuler 1 kb plus (ladder); 1. M.TaqI with 50 μM J. Name, 2013, 00, 1-3 | 3
This work was supported by the European Union’s Horizon 2020 research and innovation program under grant agreement No. 686271, “AD-Gut” and European Union’s Horizon 2020 research and innovation program under grant agreement No. 634890, “BeyondSeq”. J.H. gratefully acknowledges financial support of the Flemish government through long-term structural funding Methusalem (CASAS2, Meth/15/04). Mass spectrometry was made possible by the support of the Hercules Foundation of the Flemish Government (grant 20100225–7).

Conflicts of interest

J.H. and V.L. are co-founders of Chrometra, a company that commercializes the fluorescent direct cysteine based cofactors as part of a kit for methyltransferase-directed modification of DNA.

Notes and references

1. A. Gottfried and E. Weinhold, *Biochem. Soc. Trans.*, 2011, **39**, 623-628.
2. R. K. Neely, P. Dedeker, J. Hotta, G. Urbanaviciute, S. Klimašauskas and J. Hofkens, *Chem. Sci.*, 2010, **1**, 453-460.
3. R. K. Neely, J. Deen and J. Hofkens, *Biopolymers*, 2011, **95**, 298-311.
4. M. Levy-Sakin and Y. Ebenstein, *Current Opinion in Biotechnology*, 2013, **24**, 690-698.
5. M. Mrksich, M. E. Parks and B. Dervan, *J. Am. Chem. Soc.*, 1994, **116**, 7983-7988.
6. N. T. Thuong and C. Hélène, *Angew. Chem. Int. Ed. Engl.*, 1993, **32**, 666-690.
7. J. Deen, C. Vranken, V. Leen, R. K. Neely, K. P. F. Janssen and J. Hofkens, *Angew. Chem. Int. Ed.*, 2017, **56**, 5182-5200.
8. G. Pljevaljić, F. Schmidt and E. Weinhold, *ChemBioChem*, 2004, **5**, 265-269.
9. C. Dalhoff, G. Lukinavičius, S. Klimašauskas and E. Weinhold, *Nature Chemical Biology*, 2006, **2**, 31-32.
10. X. Cheng, *Annu. Rev. Biophys. Biomol. Struct.*, 1995, **24**, 293-318.
11. A. Grunwald, M. Dahan, A. Giesbertz, A. Nilsson, L. K. Nyberg, E. Weinhold, T. Ambjörnsson, F. Westerlund and Y. Ebenstein, *Nucleic Acids Research*, 2015, **43**, e117.
12. F. Kunkel, R. Lurz and E. Weinhold, *Molecules*, 2015, **20**, 20805–20822.
13. L. Anhäuser, F. Muttach and A. Rentmeister, *Chem. Commun.*, 2018, **54**, 449-451.
14. M. Heimes, L. Kolmar and C. Brieke, *Chem. Commun.*, 2018, **54**, 12718-12721.
15. F. Muttach, F. Mäsing, A. Studer and A. Rentmeister, *Chem. Eur. J.*, 2017, **23**, 5988-5993.
16. L. R. Comstock and S. R. Rajski, *J. Am. Chem. Soc.*, 2005, **127**, 14136-14137.
17. M. Pignot, C. Sietthoff, M. Linscheid and E. Weinhold, *Angew. Chem. Int. Ed.*, 1998, **37**, 2888-2891.
18. R. L. Wellner and S. R. Rajski, *ChemBioChem*, 2006, **7**, 243-245.
19. G. Lukinavičius, V. Lapienė, Z. Staševskij, C. Dalhoff, E. Weinhold and S. Klimašauskas, *J. Am. Chem. Soc.*, 2007, **129**, 2758-2759.
20. G. Lukinavičius, M. Tomkuvienė, V. Masevičius and S. Klimašauskas, *ACS Chem. Biol.*, 2013, **8**, 1134-1139.
21. W. Peters, S. Willnow, M. Duiskens, H. Kleine, T. Macherey, K. Duncan, D. Litchfield, B. Lüscher and E. Weinhold, *Angew. Chem. Int. Ed.*, 2010, **49**, 5170-5173.
22. C. Vranken, J. Deen, L. Dirix, T. Stakenborg, W. Dehaen, V. Leen, J. Hofkens and R. K. Neely, *Nucleic Acids Research*, 2014, **42**, e50.