

Dynamic S-adenosyl-L-methionine cofactor analogues for use in the Methyltransferase-directed Labelling of DNA

By

Andrew Allison Wilkinson

A thesis submitted to the University of Birmingham for the
degree of

Doctor of Philosophy



School of Chemistry
College of Engineering & Physical
Sciences
University of Birmingham
September 2019

UNIVERSITY OF
BIRMINGHAM

University of Birmingham Research Archive

e-theses repository

This unpublished thesis/dissertation is copyright of the author and/or third parties. The intellectual property rights of the author or third parties in respect of this work are as defined by The Copyright Designs and Patents Act 1988 or as modified by any successor legislation.

Any use made of information contained in this thesis/dissertation must be in accordance with that legislation and must be properly acknowledged. Further distribution or reproduction in any format is prohibited without the permission of the copyright holder.

Abstract

Current bioconjugation strategies are a versatile tool for the introduction of non-natural functionality to biomolecules including DNA, RNA, proteins. However, current strategies lack the versatility required to undergo sequential functionalisations. This is an area of increasing interest in chemical biology, as it facilitates multiple analyses of an individual biomolecule to help develop a wider understanding of its role within a system. Methyltransferase directed bioconjugation offers a site-selective approach for the introduction of a variety of functional groups to biomolecules, in complex mixtures.

Here we present the design and synthesis of a series of 'doubly-activated' methyltransferase cofactor analogues, based on the structure of the naturally occurring cofactor S-adenosyl-L-methionine. Specifically, we developed two cofactor analogues which included a dynamic covalent moiety and could undergo sequential modifications once introduced to a biomolecule of interest. Our approach uses a centrally located Schiff base in the form of either an oxime or hydrazone, with terminal azide functionality to allow for further modification once introduced to a biomolecule of interest.

Through a series of proof of concept experiments we have been able to rigorously test each cofactor analogue developed in the transalkylation of DNA with methyltransferases. We have then gone on to show that when present within the DNA backbone, a hydrazone moiety provides an excellent platform for the repeated functionalisation of DNA. This work is now being developed further in attempts to utilise this chemistry for unique bioconjugation applications in epigenetics and microscopy.

Acknowledgments

I would like to thank all who have helped, supported, guided and advised me throughout this research project. Without the support of others, it would not have been possible to achieve what I did during my time as a PhD student.

I would first like to thank both my PhD supervisors, Robert Neely and Francisco Fernandez-Trillo for allowing me the opportunity to work within their teams and on this project. It was a great project that at times was challenging but overall was incredibly rewarding. Their help and ideas throughout this helped to guide me to the success that I achieved.

Next, I would like to thank all those past and present from both research groups I worked within. I had a great time splitting my work between each team and felt like a key member in both. I loved being able to share all of my project highs and lows with them and appreciated all their kind words of wisdom supporting me both physically and scientifically. I would particularly like to thank those who have helped contribute to this project directly. Daniel Crisan and Nacho Insua for teaching me the ways of a chemistry lab. Elodie Jagu and Krystian Ubych for their never-ending synthetic chemistry knowledge and experience. Darren Smith, Nathaniel Wand and Qiang Su for teaching me the world of methyltransferases and helping me to develop labelling studies. And finally, Jack Kennefick and Ashleigh Rushton for allowing me to use their methyltransferase enzymes and helping to develop my biology beyond A-level.

Finally, I would like to thank my friends and family for their continuous encouragement and support throughout my project. Special thanks as well go to

the University Water Polo team where I was able to relieve stress and forget about the challenges of my day either in the pool or more often the pub. Without the team I would also never have met my incredibly supportive wife Rachel Wilkinson. Everything I have been through a small part is due to Rachel, whether it be pretending to understand my excitement of getting an experiment to work or picking me up after a tough day in the lab.

Table of Contents

| | | |
|----------|---|------------|
| 1 | <i>Introduction: Bioconjugation as a versatile tool for DNA modification</i> | 1 |
| 1.1 | Coupling Chemistries | 5 |
| 1.1.1 | Click Chemistry | 6 |
| 1.2 | DNA bioconjugation | 11 |
| 1.2.1 | Chemical Bioconjugation strategies for modifying DNA | 12 |
| 1.2.2 | Covalent, enzyme-directed DNA labelling | 17 |
| 1.2.3 | Nick Translation | 21 |
| 1.2.4 | Summary | 22 |
| 1.3 | Methyltransferase-directed bioconjugation to DNA | 23 |
| 1.3.1 | Expanding methyltransferase function | 24 |
| 1.3.2 | Examples of functionality built into DNA | 28 |
| 1.3.3 | Methyltransferase-directed bioconjugation applications | 30 |
| 1.4 | Conclusion | 33 |
| 1.4.1 | Overview | 33 |
| 1.4.2 | Aims & Objectives | 34 |
| 2 | <i>Design and synthesis of dynamic AdoMet analogues for use with Methyltransferase enzymes</i> | 37 |
| 2.1 | Background/Introduction | 37 |
| 2.2 | Precursor Synthesis | 43 |
| 2.2.1 | Hydroxyalkylation | 45 |
| 2.2.2 | Functional group introduction | 50 |
| 2.2.3 | Alcohol Activation | 52 |
| 2.3 | Cysteine cofactor synthesis | 59 |
| 2.3.1 | Thioadenosine synthesis | 60 |
| 2.3.2 | Linker coupling | 62 |
| 2.3.3 | Amino acid coupling and cofactor synthesis | 65 |
| 2.4 | Homocysteine Cofactor Synthesis | 68 |
| 2.4.1 | Cofactor Deprotection: | 70 |
| 2.4.2 | Cofactor purification | 70 |
| 2.4.3 | Cofactor analysis and stability | 73 |
| 2.4.4 | Aldehyde capping of Schiff base cofactors | 80 |
| 2.5 | Chapter Summary | 82 |
| 2.6 | Materials and Methods | 84 |
| 2.6.1 | General materials | 84 |
| 2.6.2 | HPLC analysis, concentration and yield determination | 84 |
| 2.6.3 | Chemical precursor synthesis | 85 |
| 2.6.4 | Cysteine cofactor synthesis | 92 |
| 2.6.5 | Homocysteine cofactor synthesis | 98 |
| 3 | <i>Probing cofactor analogue effectiveness</i> | 107 |
| 3.1 | Introduction | 107 |
| 3.2 | Gel electrophoresis | 109 |
| 3.2.1 | Cysteine Cofactors | 112 |
| 3.2.2 | M.TaqI transalkylation with homocysteine cofactors | 115 |
| 3.2.3 | M.MpeI DNA transalkylation | 126 |

| | | |
|------------|--|------------|
| 3.3 | Oligonucleotide HPLC | 132 |
| 3.3.1 | 14 bp oligonucleotide transalkylation | 133 |
| 3.3.2 | Hairpin transalkylation studies | 137 |
| 3.4 | Chapter summary | 146 |
| 3.5 | Materials and Methods | 147 |
| 3.5.1 | Materials | 147 |
| 3.5.2 | Gel Electrophoresis | 147 |
| 3.5.3 | Nucleic Acid Quantitative analysis | 150 |
| 4 | <i>Write, Erase, Reverse, Rewrite.....</i> | 158 |
| 4.1 | Reverse..... | 159 |
| 4.1.1 | HPLC | 159 |
| 4.1.2 | LCMS..... | 164 |
| 4.2 | DNA capture and release | 166 |
| 4.2.1 | Selectivity study..... | 168 |
| 4.2.2 | Optimisation | 170 |
| 4.3 | Beyond capture and release..... | 175 |
| 4.3.1 | Hydrazide Rewriteability..... | 175 |
| 4.3.2 | Unmethylated and Hydroxymethylated capture and release | 179 |
| 4.4 | Chapter summary | 186 |
| 4.5 | Methods..... | 187 |
| 4.5.1 | HPLC and LCMS reversibility studies | 187 |
| 4.5.2 | PCR product amplification, transalkylation, capture and release..... | 200 |
| 4.5.3 | PCR rewriting protocol and gel electrophoresis | 201 |
| 4.5.4 | Hydroxymethylated and unmethylated capture and release protocol..... | 202 |
| 5 | <i>Conclusions and Future Outlook.....</i> | 204 |
| 6 | <i>Appendix 1 - Compound Characterisation.....</i> | 208 |
| 7 | <i>Appendix 2 Draft Manuscript submitted for review</i> | 232 |
| 8 | <i>References.....</i> | 247 |

List of Figures

- Figure 1.1 Bioconjugation is achieved first by modification of the biomolecule of interest followed by coupling. Common application areas for bioconjugation include; 1) Surface attachment often achieved through singular modifications to a biomolecule tail, 2) Tag introduction to build in fluorescent labels or receptors for imaging or delivery studies, 3) Biomolecule Coupling typically sees attachment of proteins and nucleic acids. 2
- Figure 1.2 A) Huisgen cycloaddition traditionally used for bioconjugation is achieved using a copper catalyst in the presence of an azide and alkyne linking agent. B) Strain promoted azide alkyne cycloaddition (SPAAC) is achieved through the use of Di-benzocyclo-octyne's which remove the need for a copper catalyst due to the highly strained ring system which is released after coupling. 8
- Figure 1.3 Examples of cyclooctyne rings used in SPAAC to eliminate the use of copper as a catalyst. OCT - cyclooctyne, DIBO - dibenzylcyclooctyne, DBCO - dibenzotriazacyclooctyne, BARAC - biarylazacyclooctynone 10
- Figure 1.4 Alternative DNA bioconjugation strategies have used diels-alder cycloaddition. For example, Jaschke et.al. who incorporate norbornene derivatives during DNA synthesis which can be coupled to modified tetrazines where R_1 = aryl group and R_2 = Heteroaryl group. Cycloaddition is achieved in aqueous additions often achieving yields of over 90% depending on the R groups present. 11
- Figure 1.5 Schematic diagram showing how Label-IT® coupling agents can be used to covalently couple to nucleic acids for the introduction of fluorophores or haptens. Labelling is achieved through in-situ formation of an aziridine ring which is typically attacked by N⁷ of guanine forming a covalent link with the nucleic acid of interest. 14
- Figure 1.6 Covalent labelling of DNA using the random labelling kit FastTag®. The coupling reagent is first activated using heat or light to form the reactive nitrene which is attacked by primary amines in DNA in a ring expansion coupling. After coupling disulphide cleavage then releases the free thiol which can be used for tag introduction. 15
- Figure 1.7 ULYSIS nucleic acid labelling takes advantage of the universal linkage system for binding which sees substitution at the Platinum centre of the N⁷ of guanine with the leaving group forming a coordinate bond. 16
- Figure 1.8 PCR induced modifications are achieved through the introduction of synthetic nucleotides during PCR which are then used by the polymerase enzyme during synthesis and introduced randomly across the backbone. 18
- Figure 1.9 T4 RNA ligase can be used to introduce end modifications to DNA or RNA at the 3' end in the presence of ATP and a modified nucleotide bisphosphate like cytidine for the introduction of either fluorescent tags or biotin tags. 20

- Figure 1.10 Schematic diagram showing how nick translation can be used to introduce modifications to DNA. 1) DNase I binds to the DNA, 2) a nick is created in the phosphate backbone creating the 3' hydroxyl priming site. 3) DNA polymerase I binds to damaged DNA removing nucleotides. 4) DNA polymerase I simultaneously replaces the removed nucleotides with modified nucleotides present within the stock and rejoins the DNA. 22
- Figure 1.11 Pathways towards DNA modification where R groups highlight modification sites developed for the introduction of extended functionality. 1) Shows natural methylation with the class I methyl-donor cofactor S-adenosyl-L-methionine (AdoMet). 2) Demonstrates how entire nucleotides can be incorporated to DNA through Sequence-specific methyltransferase induced labelling (SMILing) with aziridine cofactor analogues when used with class I methyltransferases. 3) Identifies the pathway for methyltransferase-directed transfer of activated groups (mTAG) which is achieved with doubly-activated cofactor analogues. 24
- Figure 1.12 Aziridoadenosine AdoMet analogues alkylate DNA through the introduction of nucleotides into the backbone in a sequence specific manner depending on the methyltransferase used e.g. M.TaqI with the recognition site of 5'-TCGA-3'. 25
- Figure 1.13 Doubly-activated AdoMet analogues have shown success through introducing alkyl chains into the backbone of DNA as opposed to an entire nucleotide. With the only prerequisite to labelling success being an unsaturated bond beta to the sulfonium centre. 26
- Figure 1.14 Schematic representation of how doubly-activated analogues serve to stabilise the transition state during nucleophilic attack through p-orbital overlap. 27
- Figure 1.15 Shows the reported unsaturated functionality that has been used in doubly-activated cofactor analogues in the labelling of biomolecules. In the example of alkenes R_1 has included H, CH_3 , N_3 , $C=C$ and $C\equiv C$ ¹⁰⁵⁻¹⁰⁸. For alkynes R_2 has similarly included H, CH_3 , CH_2CH_3 , $C\equiv C$, N_3 , NH_2 ⁹⁴. For phenyl ring R_3 has been NO_2 , Anthracene, or norbornene^{101,102}. Alternatively heterogenous unsaturated systems including ketones have also been used¹⁰⁹ 29
- Figure 1.16 **Schematic representation of the reversible and rewritable modification of DNA.** Site-selective MTase-directed writing of DNA ① allows for modification via azide-alkyne cycloaddition ②. The introduced functionality can then be erased via dynamic exchange ③, to give an intermediate DNA that can be rewritten via Schiff-base formation ④ to give the original functionality introduced via MTase-labelling. Alternatively, this intermediate DNA can be further functionalized via standard conjugation techniques ⑤ to give functional DNAs for follow-on research. 35
- Figure 2.1 A) The natural synthetic pathway of S-adenosyl-L-methionine (AdoMet) with the enzyme Methionine Adenosyl Transferase (MAT). B) The general coupling mechanism for doubly-activated AdoMet analogues with AdoHcy and an activated alkyl linker

where X is typically a good leaving group like a halogen or sulfonate.

Figure 2.2 A) Highlights how Schiff base chemistry can be utilised repeatedly once incorporated into biomolecule's for multiplexed analysis. A biomolecule of choice will first have the Schiff base introduced where further functionalisation can occur followed by complete reversal to the free amine. After recovery of the amine modified biomolecule it will then be possible to repeat or end the cycle. B) Shows the two new cofactors designed containing either an oxime or hydrazone within the middle of the transferrable chain. C) The structure of a widely used previously reported cofactor with azide functionality, which has shown success with multiple methyltransferases.

Figure 2.3 Shows the structure of the final two cofactors with a change to the amino acid side chain. Synthesis of the cofactors was achieved from adenosine in simple and high yielding steps which is discussed in detail in section 2.3.

Figure 2.4 Shows the synthetic route adopted to create each activated linker. Each linker synthesis can be broken down into 3 key steps; hydroxyalkylation, functional group introduction and alcohol activation. For each linker the activating group was dependent on the analogue to be synthesised.

Figure 2.5 A) bromination of 6-heptyn-1-ol. B) an overlay of IR traces collected of 7-bromoheptyne (**1b**) compared to the starting alcohol (hept-6-yn-1-ol). Loss of OH at $3000-3500\text{ cm}^{-1}$ and introduction of sharp C-Br stretch at 540 cm^{-1} . B) Stacked ^1H NMR spectra of the alcohol (top) and bromide (bottom), complete OH removal is seen (4.4 ppm) as well as a shift and change in splitting of the CH_2 signal adjacent to the reactive site ($3.4 \rightarrow 3.6\text{ ppm}$).

Figure 2.6 Top: ^1H NMR of linker **1b**, Bottom: ^1H NMR of linker **2b** after hydroxyalkylation and purification. Two clear new peaks can be seen; a doublet of triplets at 4.0 ppm due to the CH_2 introduced and a triplet at 5.1 ppm from the OH.

Figure 2.7 A) Synthetic routes attempted for the synthesis of **2c**. B) ^1H NMR of the crude reaction mixture using nBuLi. Top = initial, Bottom = optimised. % conversion was measured by integration of the CH_2 being created at 4.0 ppm and normalised to the middle CH_2 's in the chain (2.2/1.5 ppm). If 100% conversion $4.0\text{ ppm int} = 2$. So % conversion $= (0.54/2) * 100 = 27\%$. After optimisation: $(1.34/2) * 100 = 67\%$. C) ^1H NMR of **2ci**, a doublet is seen for the CH_2 at 4.1 ppm and a triplet at 3.4 ppm (although residual Et_2O overlaps with this signal).

Figure 2.8 Characterisation of **2c** by ^1H NMR. Top: before D_2O shake, a clear peak for the OH can be seen at 5.0 ppm (same as seen for **2b**) and the carboxylic acid at 12.0 ppm. Bottom: ^1H NMR after D_2O shake. There is a complete removal of both OH peaks present in the NMR.

Figure 2.9 A) IR overlay of **2a** and **3a** following reaction with sodium azide. A clear new band can be seen corresponding to the $\text{N}=\text{N}$ stretch at 2100 cm^{-1} from the azide.

- Figure 2.10 A) Linker **3a** was chlorinated with a combination of thionyl chloride and pyridine with stirring at room temperature. B) IR overlay of linker **3a** before and after chlorination. A complete removal of the O-H stretch at 3450 cm^{-1} can be seen as well as the introduction of a new peak at 690 cm^{-1} corresponding to the C-Cl. B) ^1H NMR overlay of **3a** (Top) and **4ai** (bottom). Loss of OH can be seen at 5.0 ppm as well as a shift ($4.0 \rightarrow 4.4\text{ ppm}$) and change in splitting ($\text{dt} \rightarrow \text{t}$) of the CH_2 adjacent. 54
- Figure 2.11 A) ^1H NMR collected after tosylation and extraction of **3a**. The ratio of product to remaining tosylate was measured by comparing integrations of the aromatic peaks at 7.2 ppm and 7.8 ppm: $1.88/(1.88+0.34) = 0.85$. B) ^1H NMR of the failed tosylation of **3b**. 56
- Figure 2.12 MS data collected after synthesis and purification of linkers **4b/ciii**. Both compounds can be identified by their molecular weight in the $[\text{M}+\text{Na}]$ form with a clear bromide effect with a 1:1 ratio between 342/344 and 355/357 respectively. 58
- Figure 2.13 Synthesis of cysteine-based cofactors **AdoCys-6-N₃/NH₂** can be broken down in the 3 stages: A) synthesis of thioadenosine, B) linker coupling (and reduction if necessary), C) amino acid coupling and deprotection. 60
- Figure 2.14 Characterisation methods used for **8**. A) MS showing $[\text{M}^+]$ 284.1 and $[\text{M}+\text{Na}]$ 306.1, B) ^1H NMR C) Pendant ^{13}C NMR. 62
- Figure 2.15 A) ^1H NMR of **9a** after coupling and purification using **4aii**. B) ^1H NMR of **9a** after coupling and purification using **4ai**. C) IR of **9b** after reduction, where complete removal of azide at ca. 2100 cm^{-1} can be seen. D) ^1H NMR overlay of **9a** (top) vs. **9b** (bottom). A clear shift of the CH_2 can be seen from $3.3 \rightarrow 2.6\text{ ppm}$. 64
- Figure 2.16 Synthetic route employed for cysteine introduction to cofactors **AdoCys-6-N₃/NH₂** 65
- Figure 2.17 prep-HPLC traces of cysteine cofactor coupling after deprotection. A) **AdoCys-6-N₃**, Cofactor peaks identified by MS as T14.2/ T16.3 mins. Remaining starting material T29.3. Conversion = 74%. B) **AdoCys-6-NH₂**, Cofactor peaks at T11/ T11.8 mins. Attempted to separate as effectively as possible but some cross-over likely. Remaining starting material T28.0 conversion = 78%. 67
- Figure 2.18 A) General scheme for the optimised coupling of homocysteine cofactors. B) Optimisation experiment of cofactor coupling when using linker **4b**. 4 conditions were tested: room temperature (black), room temp. + catalytic TBAI (red), TBAI at $30\text{ }^\circ\text{C}$ (Blue) and AgClO_4 at $30\text{ }^\circ\text{C}$. Cofactor conversion was measured by integration of the three associated cofactor peaks seen during HPLC; AdoHcy, isomer 1 and isomer 2. Overall conversion was combined as the value for both isomers collected compared to remaining AdoHcy. 68
- Figure 2.19 Reaction scheme for deprotection of **10a/b** to form the oxime and hydrazone cofactors. The crude products from coupling were stirred in TFA for 2hrs dried and taken forward to be purified. 70

- Figure 2.20 prep-HPLC traces collected during purification after coupling and deprotection (if needed). Identification of peaks was confirmed using MS. A) purification of cofactor AdoHcy-6-N₃. 20 mM Ammonium formate buffer. AdoHcy = 17.1 mins, cofactor = 21.1, 23.0 mins. B) purification of **11a**. 20 mM Ammonium formate buffer. AdoHcy = 17.2 mins, cofactor = 23.6, 26.0 mins. C) purification of **11b**. 20 mM Ammonium acetate buffer. Cofactor = 17.4 mins, 19.3 mins, AdoHcy = 21.4 mins. 72
- Figure 2.21 A) Analytical HPLC analysis of each isomer of **AdoHcy-6-N₃** collected during prep-HPLC. Identification of each peak was completed by MS. Some degradation after final collection of the product can be seen. B) An example of the analytical HPLC traces collected over time during degradation studies. Loss of amino acid from the cofactor analogue can be seen as the key pathway for degradation through an increase in intensity of the peak at 34 mins. All fractions isolated were subsequently analysed by MS to confirm structures. C) A comparison of the percentage degradation seen when varying the pH at 50 °C. Degradation was measured by integration of all peaks associated with cofactor breakdown, including; adenine, AdoHcy, Ado-6-N₃ and Ado-6-N₃+H₂O. 74
- Figure 2.22 A) traditional degradation pathway proposed from previous studies for the loss of the amino acid from AdoMet analogues. B) A proposed route to explain rearrangement of the cofactor through attack on the amino acid. C) overlay of analytical HPLC collected after storage of each isomer collected for **11a**. Degradation can be seen through the formation of impurities Ade, AdoHcy, I and II. 77
- Figure 2.23 A) MS of **11a** after collection from freeze drying. Two clear peaks can be seen: 524 = **11a**, 564 = **11a**+Acetone. HRMS identified that the addition was due to the introduction of C₃H₄ which agrees with the introduction of acetone. B) MS sample of **11a** after removal of acetone, problems were found with a peak at 552 (+28) due to the addition of CO which was attributed to formate. C) Black = analytical HPLC trace of sample from A of **11a**. Red = **11a** + acetone. Complete removal of the cofactor at 15.6 mins is seen and an increase in intensity of the peak at 27.1 mins which was collected and confirmed to have a molecular weight of 564. 79
- Figure 2.24 A) Synthetic route adopted towards the capping of **11a/b**. B) LCMS traces after capping of each isomer of **11b**. The presence of **AdoHcy-8-Hyd** was confirmed for both isomers by the peak at 23.9/24.3 minutes. Excess aldehyde could also be seen to remain at 29 minutes and minor degradation through loss of the amino acid was also seen for both isomers at 32 minutes. For **AdoHcy-8-Ox** see sup. Fig. 2.11, 2.13. 81
- Figure 3.1 A) Schematic representation of DNA transalkylation followed by restriction using the plasmid pUC19 and the methyltransferase M.TaqI. Upon complete restriction the DNA is cut up in to 4 fragments of varying sizes. B) Computational modelling can be

used to predict the appearance of gels depending on the transalkylation efficiency of a methyltransferase with any cofactor. As the DNA is labelled less efficiently it is restricted more and more. Partial restriction sees the introduction of larger fragments of DNA. C) An experimental result showing the decrease in efficiency seen when transalkylation pUC19 with M.TaqI.

111

Figure 3.2 Gel electrophoresis comparing the transalkylation efficiency of each isomer collected of both **AdoCys-6-N₃/NH₂** [500µM]. A serial 2x dilution of enzyme M.TaqI, [15 ng/µl] was run for each faction to monitor its effect on transalkylation efficiency. Controls were set up using the natural cofactor AdoMet with further controls to monitor transalkylation without cofactor present and a further negative control to show complete restriction.

113

Figure 3.3 Crystal structures of AdoMet binding with M.TaqI highlighting the hydrogen bonding network (Blue dotted lines). 4 Hydrogen bonds are formed between AdoMet and the enzyme and 1 salt bridge: Adenine; N1 Phe90-NH, N6 As89-O, Ribose; O2' Glu71-COOH, Homocysteine; COOH Thr23-O, NH₃⁺ Glu45-O Ala47-O Cys48-O. A) Highlighting the enzyme binding pocket and how AdoMet fits within. B) Highlights the amino acid residues within the protein that AdoMet binds with¹⁵³.

114

Figure 3.4 Gel electrophoresis for both isomers collected of **AdoHcy-6-N₃** with M.TaqI and a 2x enzyme dilution series. Here the restriction control without M.TaqI was introduced to ensure observed protection was due to transalkylation. Each isomer was incubated at a concentration of 500 µM with M.TaqI (final concentration: [15 ng/µl]) for 1 hour at 50 °C followed by addition of the restriction enzyme for a further hour before electrophoresis. Some latent protection of the DNA in the presence of just M.TaqI can be seen in the control (11).

116

Figure 3.5 A) Synthetic pathway adopted for the synthesis of **10a**. Purification was achieved using 20 mM ammonium acetate buffer gradient with methanol. Retention times: isomer 1 = 45.0 minutes, isomer 2 = 47 minutes. B) prep-HPLC collected for the purification of **10a**. The two isomers were collected and identified by MS before being taken forward for transalkylation. C) Gel electrophoresis of each isomer of **10a** [500 µM] after purification. A serial 2x enzyme dilution was run using M.TaqI, highest concentration: [15 ng/µl].

119

Figure 3.6 A) Deprotection was achieved directly after coupling using neat TFA. After deprotection purification was achieved by prep-HPLC. Each isomer was again isolated and tested individually as a methyltransferase cofactor. B) Gel electrophoresis of the deprotected cofactor **11a** [500 µM], M.TaqI [15 ng/µl]. C) analytical HPLC chromatograms of each isomer sample after prep-HPLC showing separation of the two diastereoisomers, blue = isomer 1, black = isomer 2.

120

Figure 3.7 Gel electrophoresis of the capped hydroxylamine cofactor (**AdoHcy-8-Ox**). A serial 2x cofactor dilution was run, showing

- isomer 1 to be the less effective diastereoisomer (M.TaqI [7.5 ng/μl]). Restriction controls (lane 7, 14) were run at the highest cofactor concentration of 500 μM. 122
- Figure 3.8 Gel electrophoresis of both isomers of **AdoHcy-8-Hyd**. A serial 3x cofactor dilution series was run from 300-1.2 μM (M.TaqI [7.5 ng/μl]) to highlight and differences in transalkylation efficiencies. 123
- Figure 3.9 Plot of percentage alkylation vs. [Cofactor] to determine approximate K_X of both cofactors **AdoHcy-8-Ox/Hyd**. Percentage alkylation was calculated by comparison against restriction plot from Figure 3.1.B for M.TaqI. Approximate K_X **AdoHcy-8-Hyd** isomer 1 = 5.5 μM. 124
- Figure 3.10 Gel electrophoresis showing a cofactor dilution from 10-0.1 μM for **AdoHcy-8-Hyd**, **AdoHcy-8-Ox**, **AdoHcy-6-N₃**, and **AdoMet** with M.TaqI [7.5 ng/μl] and pUC19 (300 ng). Transalkylation was set-up in tandem and samples were spread across 2 gels run in parallel. 125
- Figure 3.11 A plot of alkylation efficiency against cofactor concentration for each cofactor analogue. Alkylation efficiency was calculated by comparison with the restriction map for M.TaqI. K_X can then be estimated for each cofactor analogue by looking at the point at which 50% alkylation occurs. **AdoMet** K_X = 0.15 μM, **AdoHcy-6-N₃** K_X = 0.7 μM, **AdoHcy-8-Hyd** K_X = 1.8 μM, **AdoHcy-8-Ox** K_X = 5.9 μM. 126
- Figure 3.12 A) Modelled HhaI restriction plot of pUC19 at varying transalkylation densities. B) pUC19 plasmid map highlighting the locations of each HhaI restriction site. C) A plot of fragment size after complete restriction of pUC19 with HhaI. Upon complete restriction two main bands are expected at approximately 100 and 350 bp. 128
- Figure 3.13 Gel electrophoresis of **AdoHcy-8-Hyd** and M.MpeI [8.5 μg/μl] in the transalkylation of pUC19. A serial 2x cofactor dilution series was run for each isomer. Transalkylation was run at 37 °C for 1 hour followed by individual purification of samples on silica columns before restriction with HhaI at 37 °C for a further hour. 129
- Figure 3.14 A) predicted restriction map of HaeII at varying transalkylation efficiencies for pUC19. B) Gel electrophoresis of **AdoHcy-8-Hyd** after transalkylation with M.MpeI [8.5 μg/μl] and restriction with HaeII with a serial 2x cofactor dilution series. 131
- Figure 3.15 A) Transalkylation of the 14 bp oligo was achieved using M.TaqI and either AdoMet or one of the successful AdoMet analogues. After transalkylation DNA was purified on silica columns before injection into the HPLC. B) HPLC traces of each analogue tested after transalkylation. All HPLC was run at 60 °C using 0.1 M triethylammonium acetate with an acetonitrile gradient. C) DNA sequence of both the forward and reverse strand of the 14 bp oligonucleotide highlighting the M.TaqI recognition site. Raw chromatograms in supplementary information figures S1-10. 134
- Figure 3.16 A) Watson and Crick hydrogen bonding between Adenine and Thymine. B) After methylation there is the introduction of a methyl group, however hydrogen bonding is not dramatically affected. C)

introduction of more complex functionality may affect the overall hydrogen bonding network of DNA at the modification site because of sterics caused by the new group. 136

Figure 3.17 A) Mass calculations for the expected increase in molecular weight with each cofactor analogue. B) A table plotting the observed masses for each strand of the 14 bp oligo after transalkylation and HPLC. For samples which appeared as the triply charged species the total mass was calculated by: $((M^3 \times 3) + 3)$. MS data supplementary figures S1-10 137

Figure 3.18 Design and sequences of hairpin oligonucleotides used for HPLC. Each consisted of a 32 bp (red) single stranded recognition sequence followed by either 1, 2 or 3 TCGA sites (green) with 4 bp spacing regions either side. 138

Figure 3.19 HPLC traces collected after transalkylation of hairpin DNA probes with **AdoHcy-6-N₃**. Transalkylation was carried out in the same way as for the 14 bp oligo with M.TaqI and cofactor for 1 hour and 50 °C followed by proteinase K treatment and purification using silica columns before injection in to the HPLC. Successful MS data collected for OP1 in supplementary figure S3.12, 3.13. Black traces show the unlabelled control for each hairpin and blue shows modified DNA after the initial transalkylation attempt. 139

Figure 3.20 A) LCMS analysis of OP2 transalkylation with **AdoHcy-6-N₃** and M.TaqI. Separation was achieved using a 0.1M TEAA buffer gradient with acetonitrile. Top = UV absorbance 200-800 nm, Bottom = TIC C) After LCMS the TIC spectra collected were then combined between 4.25-4.55 minutes and analysed using the software promass for masslynx by waters giving the molecular weights: 23,568.8, 23690.6, 23810.7 (Unlabelled = 23445), showing the 121 g mol⁻¹ increment expected. 142

Figure 3.21 A) HPLC traces collected of OP1. I: unlabelled hairpin. II: OP1 after transalkylation with **AdoHcy-6-N₃** and M.TaqI. III: OP1 after labelling with DBCO-TAMRA. Coupling was achieved at room temperature overnight with 100 equivalents of fluorophore in 20% DMSO. B) UV absorbance of OP1 after TAMRA labelling at 560 nm. Fluorescent tagging was achieved by coupling of 100 equivalents of fluorophore for each transalkylation site in 20% DMSO at room temperature overnight. 144

Figure 4.1 Schematic diagram of the dynamic cycle we hoped to create using the responsive cofactors **AdoHcy-8-ox/hyd**. After transalkylation you can **write** desired information into the DNA backbone using azide alkyne click chemistry. This information can then be **Erased** using the reversal conditions developed, at which point the DNA can either be **reversed** to its starting form or **rewritten** to introduce new functionality be it permanent or reversible. 158

Figure 4.2 The role of the competing reagent within Schiff base reversal. A) The Schiff base equilibrium can be artificially shifted through the introduction of a competing reagent in excess which reacts with either the original aldehyde or amine preventing reformation of

the original hydrazone or oxime. B) For HPLC studies, two competing reagents were tried which were; tri-hydroxybenzaldehyde and hydroxylamine. These were used due to the stability of the products they create in the formation of more stable Schiff bases than the original.

160

Figure 4.3 HPLC analysis of the 14 bp oligo after the first attempt at reversal A) **AdoHcy-8-ox** labelled DNA, Top = 8-ox modified oligo, middle = THBA reversal, bottom = hydroxylamine reversal. B) **AdoHcy-8-hyd** labelled DNA, Top = 8-hyd modified oligo, middle = THBA reversal, bottom = hydroxylamine reversal. All samples were incubated at 50 °C for 1 hour in 10 mM ammonium acetate pH 4.0 with 10 equivalents of the competing reagent. C) During HPLC each new peak was collected and then concentrated under vacuum before MS analysis. MS and HPLC traces section 4.5.1.1 Figure S4.14-21.

162

Figure 4.4 LCMS analysis of transalkylation, reversal and reattachment. A) Unlabelled 14 bp oligo, B) labelled with M.TaqI and **AdoHcy-8-Hyd**, C) hydrazone reversal, D) PEG-aldehyde reattachment. Reversal was achieved by heating labelled oligo at 50 °C for 1 hour in pH 4 ammonium acetate buffer (10 mM) and Hydroxylamine (100 equivalents, 2 mM). After reversal the DNA was not purified and an excess of PEG-aldehyde (200 equivalents, 4 mM) was added and incubated at 37 °C for 1 hour before final analysis. Control chromatograms section 4.4.1.2 S4.31-35.

165

Figure 4.5 The proposed method for capture and release of fragmented DNA. DNA is first transalkylated with an AdoMet analogue and M.MpeI using the labelling conditions used previously for gel electrophoresis followed by coupling of a cross-linking reagent (5x excess) before purification and capture onto streptavidin beads. DNA is then released using the reversible chemistry introduced with the optimised release conditions found to be 10 mM Ammonium Acetate, pH 6.8, 1 M NaCl, 20 mM H₂NOH.HCl.

168

Figure 4.6 A study to test the capture and release efficiency of a 203 bp fragment of DNA with M.MpeI and AdoMet analogues. A) DNA capture was achieved using streptavidin beads in a high salt (1 M NaCl) tris buffer (10 mM). After each capture attempt the supernatant DNA concentration was measured and transferred to a fresh batch of streptavidin beads. B) After capture all portions of beads were combined and resuspended in the high salt tris buffer followed by the addition of the release buffer (10% by volume, 100 mM Ammonium acetate pH 4.0) and CR (10 % by volume, 200 mM H₂NOH.HCl). Samples were then incubated at 50 °C for 1 hour and the supernatant DNA concentration was measured. The supernatant was then removed and the beads were resuspended in fresh release buffer for the process to be repeated.

169

Figure 4.7 Release optimisation assessing the effect of pH and CR concentration on release. Capture for all samples used identical

- conditions. A) Capture profiles for the three samples to assess pH affect. B) Release profiles showing the effect of adjusting pH on release. C) Capture for varying CR concentration. D) Release profiles showing the effect of the competing reagent on release. 171
- Figure 4.8 DNA capture and release with varying quantities of DNA. A) DNA capture. Capture was performed using 3 portions of 0.01 mg streptavidin dynabeads in 70 μ l 10 mM Tris 1 M NaCl. B) DNA release. Release was then achieved using 70 μ l 20 mM hydroxylamine in 10 mM ammonium acetate with 1 M NaCl and 0.01% SDS. 173
- Figure 4.9 Optimised capture and release results A) 400 ng of DNA was captured for each sample in 56 μ l of 10 mM Tris and 1 M NaCl. B) After capture each sample was released in 3 lots of 56 μ l 10 mM ammonium acetate pH 6.5, 0.01% SDS, 1 M NaCl, 20 mM hydroxylamine. The average value from the triplicate repeat was plotted with the error bars plotting the standard deviation of each result. 175
- Figure 4.10 Rewriteability was tested using the 203 bp fragment labelled with M.Mpel and **AdoHcy-8-Hyd** followed by capture and release using streptavidin beads. The released DNA was then coupled to an NHS fluorophore to prove the hydrazide reactivity. After each stage an aliquot of DNA was taken to be used for controls. For uncropped electrophoresis results see supplementary fig. S23. 177
- Figure 4.11 Gel electrophoresis testing the reversibility of 8-hyd labelled DNA completed by E. Jagu. A) Schematic representation of the labelling study carried out. DNA was first alkylated followed by direct attachment in solution of DBCO-TAMRA. Labelling was then reversed followed by subsequent attachment of NHS-Atto 647N. Further details of this experiment can be found in the attached manuscript in the appendix. 179
- Figure 4.12 A) A method for the purification of the three key epigenetic cytosine markers found within genomic DNA. Hydroxymethylated and methylated DNA can first be azide tagged followed by functionalisation onto streptavidin beads leaving methylated DNA in solution. Unmethylated DNA can be released using hydrazone reversal and Hydroxymethylated DNA by the denaturation of streptavidin. B) Capture performed to test the efficiency of transalkylation of the Hydroxymethylated and unmethylated DNA. C) Release profiles showing the selectivity of each release method. Unmethylated DNA was released over 4 attempts using hydroxylamine competition. Hydroxymethylated DNA was then released by boiling the beads in 0.1% SDS for 10 minutes. 181
- Figure 4.13 Capture and release of Hydroxymethylated and unmethylated DNA with desthiobiotin release. A) capture profiles of the five samples used to test the effectiveness of biotinylation vs. desthiobiotinylation. B) Hydrazone reversal was completed over 5 release attempts using hydroxylamine competition, followed by two attempts at desthiobiotin reversal using biotin competition

and a final release attempt was completed by boiling in 0.1% SDS.

184

Figure 5.1 Shows a 23 kb section of data collected after nanopore sequencing of M.TaqI labelled DNA with **AdoHcy-8-hyd** and fragmented E.coli DNA after transalkylation, capture and release. Genome coordinate – identifies the region of the genome being examined. Number of reads – a histogram plot of all fragments that passed through the nanopore highlighting density of fragments that passed through from each region of the genome. Individual reads – Shows an individual plot of every DNA fragment that passed through the nanopore placed where they align against the reference genome. M.TaqI sites – Identifies the location of all M.TaqI sites within this section of the E.coli genome analysed.

206

Figure 2 – MTase-directed writing of DNA. (A) Schematic representation of restriction assay and (B) gel electrophoresis of pUC19 following enzymatic treatment with M.TaqI and/or R.TaqI in the presence and absence of AdoMet (375 μ M) or AdoMet derivative **4b**. In the absence of M.TaqI-mediated alkylation (lanes 4, 8 and 12), pUC19 is cut into fragments, of which the largest three can be identified by gel electrophoresis. M.TaqI-mediated alkylation with AdoMet (lane 10) or derivative **4b** (lanes 1-3 and 5-6) results in partial to full protection from restriction by R.TaqI, with mainly open circular and supercoiled plasmid DNA being observed by gel electrophoresis. Controls in the absence of AdoMet derivatives (lanes 11 and 12), in the absence of M.TaqI (lanes 4, 8 and 12) and in the absence of R.TaqI (lane 9) are included. The gel clearly shows that catalysis of DNA transalkylation by M.TaqI with diastereomer I is less efficient than diastereomer II.

236

Figure 3 – Writing and erasing the introduced functionality. Analytical HPLC chromatograms of oligo DNA (top) and oligo DNA following incubation with M.TaqI and **4** (middle), and with M.TaqI and **4**, followed by incubation with 10 equiv. of H₂NOH·HCl in 10 mM ammonium acetate pH4 (bottom). HPLC conditions: 0.1 M triethylammonium acetate buffer, pH 7.0 (A)/MeCN (B) gradient at 60 °C. Under these conditions, oligo DNA melts and both strands of DNA can be observed independently.

237

Figure 4 – Re-writing DNA. (A) Schematic representation of *erasing* and *rewriting* of chemical functionality on oligo DNA using the dynamic AdoMet derivative **4a** and aldehyde **5**. Analytical HPLC chromatograms of oligo DNA following incubation with M.TaqI and **4a**, followed by incubation with 10 equiv. of H₂NOH·HCl in 10 mM ammonium acetate pH4 (B), and oligo DNA from B followed by incubation with aldehyde **5** (C). Chromatogram of oligo DNA following incubation with M.TaqI and **4a** (D) and of aldehyde **5** incubated with 10 equiv. of H₂NOH·HCl in 10 mM ammonium acetate pH4 (E) shown for comparison. HPLC conditions: 0.1 M triethylammonium acetate buffer, pH 7.0 (A)/MeCN (B) gradient at 60 °C. Under these conditions, oligo DNA melts and both strands of DNA can be observed independently.

238

Figure 5 – Dual functionalization. (A) Schematic representation of the dual functionalization of DNA: DNA was first *written* via site-selective MTase-directed labelling using dynamic cofactor **4a** ①. The obtained azide-functionalized DNA fragments were then *modified* with TAMRA-DBCO **6** via azide-alkyne cycloaddition ②. The introduced TAMRA was then *erased* via dynamic exchange ③, to give hydrazide-functionalized DNA fragments that were further functionalized with NHS activated Atto 647N **7** ⑤. Modification was monitored using gel electrophoresis. Conditions: DNA concentration; 7 ng/μL, release buffer; 10 mM Ammonium Acetate, pH 6.8, 1 M NaCl, 0.01% SDS. DNA stained with GelRed®. Gel was visualized using a Bio-Rad Pharos FX (GelRed®: excitation, trans-UV; emission filter, 590/110 nm; TAMRA: excitation, epi-green illumination; emission filter: 602/50 nm; Atto 647N **7**: excitation, epi-red illumination; emission filter: 700/50 nm). TAMRA channel was colored yellow and Atto 647N **7** was colored red for visualization. (B) GelRed® channel and (C) Composite image of TAMRA and Atto647n channels. Full chemical structure of the fluorescent dyes are available in the Supporting Information.

240

Figure 6 – Capture and release of DNA using MTase-directed labelling and dynamic exchange. (A) Schematic representation of functionalization of DNA fragments with a biotin tag using both MTase-directed labelling and subsequent DBCO conjugation. (B) Schematic representation of the capture and release of DNA fragments using the dynamic AdoMet derivative **4a**. (C) Percentage of DNA remaining following capture with magnetic beads and (D) percentage DNA released from the magnetic beads following treatment with H₂NOH·HCl (cycles 1-3) and denaturing conditions (reflux in denaturing buffer, cycles 4 and 5). DNA was either unlabeled (■), incubated with M.Mpel and **4a** (●), or incubated with M.Mpel and azide containing cofactor **8** (○). Conditions: DNA concentration; 7 ng/μL; capture buffer: 10 mM Tris, 1 M NaCl, pH 8.5; dynamic release buffer: 20 mM H₂NOH·HCl, 10 mM ammonium acetate, pH 6.8, 1 M NaCl, 0.01% SDS; denaturing buffer.

241

1 Introduction: Bioconjugation as a versatile tool for DNA modification

Bioconjugation is a general term for an array of different labelling reactions that see the formation of covalent linkages between any biomolecule (e.g. protein, DNA, RNA, carbohydrate, etc.) and compound of interest. Since its inception, bioconjugation has now developed into an expansive series of reactions that are routinely used within both research and industry. Bioconjugation reactions have allowed for the development of previously unachievable applications in which biomolecules can either be used to further our understanding of their structure and function; or to take advantage of their unique characteristics as tools in molecular biology. A multitude of applications are now possible using bioconjugation, however this is not our main focus. Instead we are interested in the methods by which bioconjugation is achieved, specifically with DNA. Comprehensive literature can be found around the applications in which bioconjugation reactions are most commonly used, three key areas of which are (Figure 1.1):

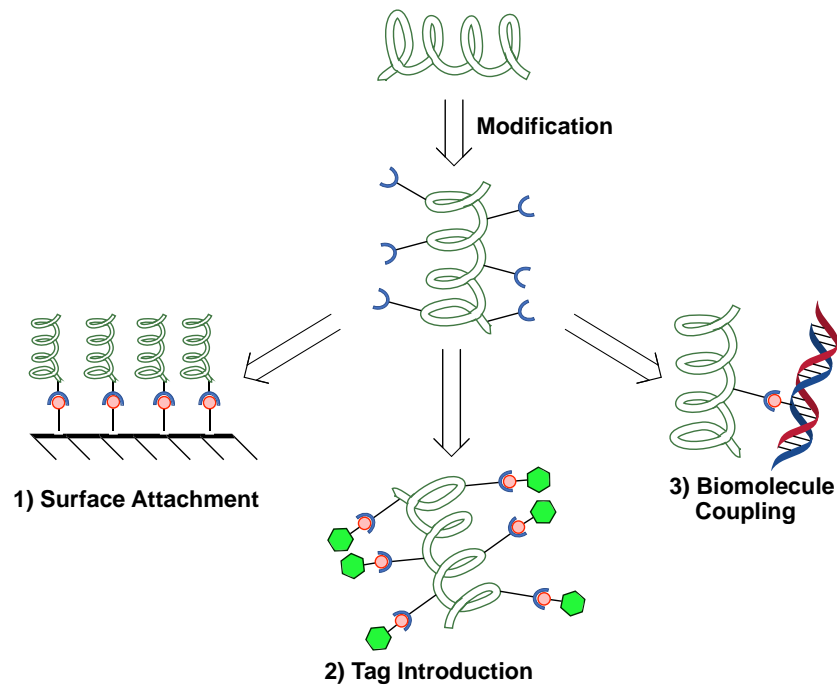


Figure 1.1 Bioconjugation is achieved first by modification of the biomolecule of interest followed by coupling. Common application areas for bioconjugation include; 1) Surface attachment often achieved through singular modifications to a biomolecule tail, 2) Tag introduction to build in fluorescent labels or receptors for imaging or delivery studies, 3) Biomolecule Coupling typically sees attachment of proteins and nucleic acids.

1) Surface attachment – This area of bioconjugation sees the attachment of biomolecules to nanoparticles¹, beads², and slides in an array of different applications like microarrays³, spherical nucleic acids⁴ and imaging⁵. It is an area of increasing interest as surface functionalisation allows the trapping of biomolecules in specific locations and configurations. Once attached to a surface it is then possible take advantage of the biomolecule of interest in new ways that may not be possible in cells or solution. This has aided the development of both biomolecule purification techniques with products like dynabeads®, and the creation of new theranostics like spherical nucleic acids (SNA's)⁶. In the example of SNA's Mirkin *et.al.* take advantage of thiol terminated synthetic nucleic acids which bind non-covalently to gold nanoparticles in an ordered, dense and linear fashion. Examples achieving

this have shown synergistic effects through amplification of both the nucleic acid and nanoparticle characteristics. Perhaps surprisingly, due to their dense negative charge, SNA's have worked very effectively as tools for *in vitro* gene delivery⁴.

- 2) Tag introduction – The next general application of bioconjugation is tag introduction. Here tags including fluorophores and haptens can be introduced to create fluorescent or targeting biomolecules. Doing this allows the targeted delivery or imaging of biomolecules with high specificity and selectivity. Through the attachment of tags to biomolecules it is possible to develop probes for *in vitro* imaging which can be used in applications like Fluorescence in situ hybridization (FISH) ⁷. Here fluorescently tagged DNA can be used to target chromosomal abnormalities across the genome dependant on the DNA sequence introduced. This is achieved through heat treatment of the tagged DNA with fixed cells and subsequent hybridisation before imaging. Other uses of biomolecular probes also include; monitoring enzymatic interactions, conformation changes and real-time imaging of biomolecules within cells⁸.
- 3) Biomolecule coupling – The final application area sees the covalent combination of two or more biomolecules. This is of interest for its potential in applications around diagnostic assays, molecular recognition studies, and the development of biomolecule nanostructures⁹. Typically, this sees the combination of proteins with nucleic acids of interest and benefits from the unique specificity provided by the nucleic acid and its Watson and crick base pairing which allows great flexibility as a building block. This can then be

combined with an array of proteins, each of which have their own unique characteristics and properties. This combination has proven crucial for the success of applications like immuno-PCR¹⁰ and DNA-Paint¹¹ both of which see the combination of short chain nucleic acids with modified antibodies. First developed by Sano *et.al.* in 1992, immuno-PCR sees the combination of PCR and its signal amplification power with ELISA's (enzyme linked immunosorbent assay) versatility and flexibility¹². Originally developed for the detection of bovine serum albumin (BSA) in which it increased detection limits by a factor of 10^5 , this technique has now progressed dramatically for use in both clinical and commercial studies¹³.

Here we are interested in exploring the methods used to achieve bioconjugation and the chemistry that can be introduced to improve conjugation favourability. This is an area of increasing interest within both research and industry as a way to work towards, for example, the improvement of traditional drug delivery systems like liposomes, nanoparticles and vesicular systems¹⁴. By developing specific bioconjugation strategies it is possible to improve upon traditional systems by introducing new features that can introduce flexibility, stability, or biodegradable structures. For example, in the case of oligonucleotides cellular internalisation can be improved through the attachment of conjugates like cell penetrating peptides (CPP), poly-ethylene glycol or GalNAc (N-acetylgalactose-amine)¹⁵. This bioconjugation can then be further developed by the introduction of cleavable sites like disulphide linkages for controlled oligonucleotide release after internalization and reduction in the case of CPP's¹⁶. We are particularly interested in these examples where dynamic covalent chemistry is incorporated

through bioconjugation and exploring applications around not only drug delivery, but also biomolecule purification, and the development of active conjugates¹⁷.

When looking at the way in which DNA bioconjugation is achieved there are two main factors that need to be considered. First is the chemistry by which coupling to the biomolecule of interest is achieved and second is how the reactivity is built into the DNA backbone for coupling. Both of which need careful consideration when developing any bioconjugation strategy because as yet a perfect answer without limitations has not been found. A further area that can also be considered but is less explored is the dynamic nature of systems developed. Traditionally the main aim of bioconjugation is the formation of permanent linkages, but with the inclusion of dynamic chemistries and linkage sites the creation of dynamic bioconjugates is possible.

1.1 Coupling Chemistries

Throughout the development of bioconjugation chemistries an array of biomolecules have been coupled to surfaces, tags and other biomolecules. However, comparatively few chemistries have been used to do this. This is because, in general, the chemical and physical properties desired for each bioconjugate are achieved by the biomolecule, tag or surface used, not the chemistry that couples them together. As such, coupling is typically done with simple chemical reactions that are high yielding, bio-orthogonal and do not undergo multiple side reactions to ensure the maximum chances of success. One such field of chemistry which can be used to achieve this being 'click chemistry'¹⁸.

1.1.1 Click Chemistry

Click chemistry is a general term coined by Sharpless *et.al.* in 2001 and describes a sub-set of coupling reactions used in bioconjugation due to their simplicity and high yields. It is defined not by a catalogue of individual starting materials or products but instead as a research ideology for reactions that produce minimal by-products, can be purified without chromatographic techniques, are modular in nature and stereospecific. Originally developed as a methodology for industry, click chemistry assists greatly in the development of large libraries of compounds which can be tested more rapidly in screening studies¹⁹.

Several coupling strategies have now been developed which are consistent with the principles of click chemistry, however only a select few are regularly employed in bioconjugation. This is because, for use in bioconjugation further criteria must be met for the reaction to be successful as coupling is often carried out in complex biological conditions. Some key factors to consider being:

- 1) Bio-orthogonality – In bioconjugation the chances of cross reactions during coupling increases greatly due to the complexity of biomolecules and the conditions in which they are stored. Typically, this means when carrying out bioconjugation, reactive groups orthogonal to those present within the biomolecule and nature are commonly chosen for coupling. Some examples for DNA being; aldehydes, azides, alkynes & thiols²⁰.
- 2) Coupling conditions – A key factor to consider as coupling may negatively impact the fundamental structure and function of the original biomolecule, reducing potential applications post conjugation. For example, if using reducing conditions with proteins, the quaternary structure may be altered

due to cleavage of disulphide bonds²¹. Or if using a copper catalyst in the coupling of DNA, fragmentation may be caused²². This can significantly change the original biomolecule structure meaning applications based around their function will not be reliable post coupling.

- 3) Solubility - Further to the reagents used it is also necessary, in DNA coupling to consider the solvents used. For example, when attaching fluorophores, it is often necessary to use small quantities of organic solvents like DMSO or alcohols to solubilise the fluorophore. However, if used in high quantities, these solvents may have adverse effects including DNA denaturation or precipitation during coupling. Additionally, it is also necessary to consider the effect introduction of new chemistries will have on biomolecules and how these may change their behaviour. Adding large quantities of bulky, hydrophobic groups may change biomolecule solubility and behaviour in solution again reducing the number of applications they can be used in.

One coupling chemistry that can meet many of the requirements above and is commonly used in bioconjugation especially with DNA is Huisgen cycloaddition. This reaction sees the coupling of azides and alkynes with a copper catalyst (Figure 1.2.A)²³. This reaction has garnered much support for its use in bioconjugation due to its high yields, selectivity, and biorthogonality^{24,25}. However, some drawbacks for this reaction do exist, including the oxidative damage the copper catalyst can cause²⁶. When used in DNA coupling, Cu(I) can cause significant fragmentation due to formation of reactive oxygen species (ROS) which can oxidise bases and lead to mutations or deletions²⁷. These

effects can be mitigated with the addition of polytriazole chelating ligands like tris(benzyltriazolylmethyl)amine to help stabilize Cu(I). Or, by adding radical scavengers like dimethylsulfoxide to prevent ROS formation, but ultimately cannot completely remove the risk of damage caused by copper²⁸. An additional drawback of this reaction is that it requires elevated temperatures when not using a catalyst and is highly exothermic in nature. Again, this can significantly impact DNA during bioconjugation through denaturation

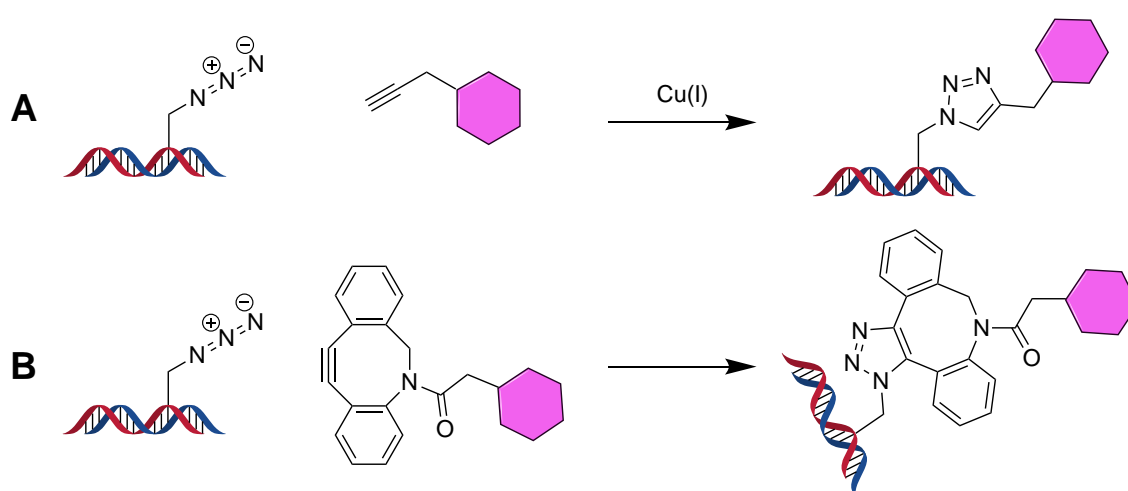


Figure 1.2 A) Huisgen cycloaddition traditionally used for bioconjugation is achieved using a copper catalyst in the presence of an azide and alkyne linking agent. B) Strain promoted azide alkyne cycloaddition (SPAAC) is achieved through the use of Di-benzocyclo-octyne's which remove the need for a copper catalyst due to the highly strained ring system which is released after coupling.

To eliminate the risk of DNA damage highly strained cyclooctyne rings can be used to remove Cu(I) as the thermodynamic benefit provided by strain release drives cycloaddition to completion without the need for a catalyst, whilst also maintaining good reaction kinetics (Figure 1.2.B)²⁹. This coupling mechanism is called strain promoted azide alkyne cycloaddition (SPAAC). Although originally developed as a tool for *in vivo* labelling studies to avoid Cu(I)'s cytotoxicity this method is also beneficial in DNA bioconjugation as the risk of DNA damage is removed³⁰. SPAAC has now developed into a valuable tool for DNA

bioconjugation and is used where changes to DNA integrity can significantly impact the quality of results. For example, in optical DNA mapping where sample fragmentation significantly decreases accuracy and reliability of maps produced^{22,31}.

In general, two classes of cyclooctynes exist that have been used in bioconjugation which are aliphatic cyclooctynes and dibenzoannulated cyclooctynes (Figure 1.3). Although originally developed around aliphatic cyclooctynes, dibenzoannulated rings are now preferred for SPAAC due to their increased reactivity³². Through the addition of benzyl groups to the cyclooctyne ring framework the quantity of sp^2 carbon atoms present within the ring increases which in turn increases ring strain³³.

Although increasing ring strain does bring with it increased reactivity, a compromise is often found with this modification due to their decreased stability. For instance, in the example of biarylazacyclooctynone (BARAC) ring strain is increased by the addition of a ketone within the octyne ring and has a very fast reaction rate, but the compound also rapidly decomposes³⁴. As such, for bioconjugation a middle ground is required where reaction rates are fast enough to provide efficient conjugation but where compounds also maintain relative stability. An example that fits these criteria is dibenzoazacyclooctyne (DBCO) which maintains stability and includes a carboxyl group that can be used for the coupling of tags³⁵.

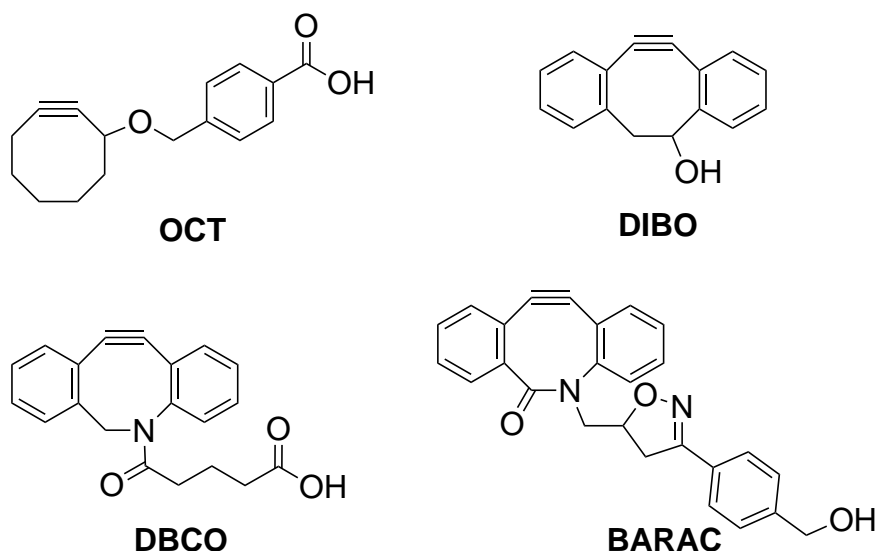


Figure 1.3 Examples of cyclooctyne rings used in SPAAC to eliminate the use of copper as a catalyst. OCT - cyclooctyne, DIBO - dibenzylcyclooctyne, DBCO - dibenzoazacyclooctyne, BARAC - biarylazacyclooctynone

Other click reactions also available for their use in DNA-bioconjugation include NHS-esters³⁶, carbonyls³⁷, thiols³⁸ and dienes³⁹. Often though, these chemistries also have drawbacks which reduce their versatility as tools in conjugation. For example, the group of Jaschke *et.al.* successfully incorporate norbornene derivatives into the backbone of oligonucleotides (Oligos) which can then react with modified tetrazines in a traditional [4+2] Diels Alder cycloaddition⁴⁰. Although only limited to synthetic DNA as the norbornene derivatives are incorporated during solid phase synthesis. The chemistry used is biorthogonal,

meaning coupling is highly selective and is not compromised by competing reactions.

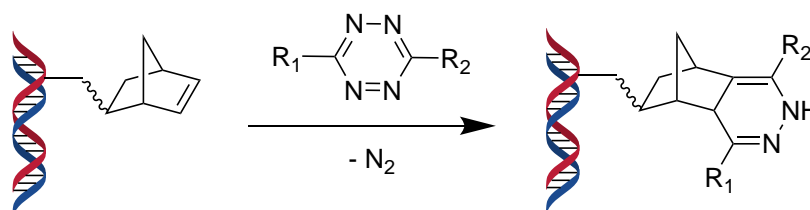


Figure 1.4 Alternative DNA bioconjugation strategies have used diels-alder cycloaddition. For example, Jaschke *et.al.* who incorporate norbornene derivatives during DNA synthesis which can be coupled to modified tetrazines where R₁ = aryl group and R₂ = Heteroaryl group. Cycloaddition is achieved in aqueous additions often achieving yields of over 90% depending on the R groups present.

Similarly, with thiols and NHS-esters good coupling conditions can be found that react with complementary functionalities built into DNA like terminal amines and thiols. With these functionalities though, selectivity can be reduced due to off-site targeting as the reactive groups are commonly present in biomolecules, like proteins in the case of amines and thiols. As such, coupling with these chemistries is a challenge *in vitro*. However, it is possible to overcome these drawbacks by exploring the way reactive functionalities are first introduced. For example, in the case of NHS-esters fluorophores can be pre-coupled to enzymatic cofactors that are subsequently introduced directly to DNA in solution⁴¹. This emphasises the importance when considering a bioconjugation strategy to not only consider the coupling chemistry used but also the pathway towards the final product and how this is achieved.

1.2 DNA bioconjugation

As discussed previously, when conjugating functional chemistries to DNA 'click chemistry' is adopted to carry out the couplings of any compound of interest.

However, what was not discussed was the way in which the reactive groups are first introduced to DNA to allow coupling to occur. Broadly the techniques used to introduce reactive groups to DNA can be classified into two key categories where modifications are either introduced chemically or enzymatically, both of which have unique advantages and disadvantages.

1.2.1 Chemical Bioconjugation strategies for modifying DNA

1.2.1.1 *End modifications*

The first conjugation strategy is end modifications on synthetic oligonucleotides. Through the use of modified phosphoramidites for 5' end modifications, or specialist resins for the 3' end, terminal chemistries are introduced that can be used for further couplings. Some examples of functionalities introduced are; amines, carboxylic acids, thiols, NHS esters, biotin derivatives and azides and alkynes. Once introduced, the reactive groups can be used as tethering sites for attachment of DNA using click chemistry to proteins⁴², surfaces⁴³, quantum dots⁴⁴, nanoparticles⁴⁵ and more. Alternatively, chemical tags like fluorophores and quenchers can also be introduced directly for imaging studies using pre-functionalised nucleotides. This technique was made possible by the automation of DNA synthesis in the early 1980's⁴⁶. After refinement and optimisation synthetically modified oligos can now be ordered and delivered within 24 hours of their design for a few pounds⁴⁷.

End modifications are a robust method for obtaining modified DNA due to their commercial availability, simplicity and low cost. Furthermore, they are beneficial as the quantity and location of modifications are known. However, this

modification strategy does also have weaknesses, like the fact that it is limited to synthetic DNA, so cannot be used for labelling genomic samples directly. Additionally, this modification strategy is limited in the length of oligonucleotide that can be made, as yields decrease with increasing chain length which dramatically increases cost^{25,48}.

1.2.1.2 *Non-specific covalent labelling*

Slattum *et.al.* developed covalent DNA labelling as a tool for the fluorescent labelling of plasmid DNA to monitor their movement through cells ⁴⁹. This has now been developed into a commercial labelling kit Label-IT®. This kit can be used with all nucleic acids, taking advantage of the *in-situ* generation of highly electrophilic N-mustard aziridine rings to promote coupling. Although a range of coupling reagents exist, each have the same three-part structure; the reactive halo-amine group, a positively charged linker, and the final tag region which contains the fluorophore or hapten of interest (Figure 1.5). It is proposed labelling occurs by the reagent first intercalating with DNA through electrostatic interactions arranging the coupling chemistry in the right proximity for attack. This is followed by formation of the aziridine ring and then final nucleophilic attack of DNA to form the covalent linkage. After coupling labelling efficiency is quantified by spectrofluorometric analysis, comparing the UV absorbance of the DNA to the fluorophore introduced.

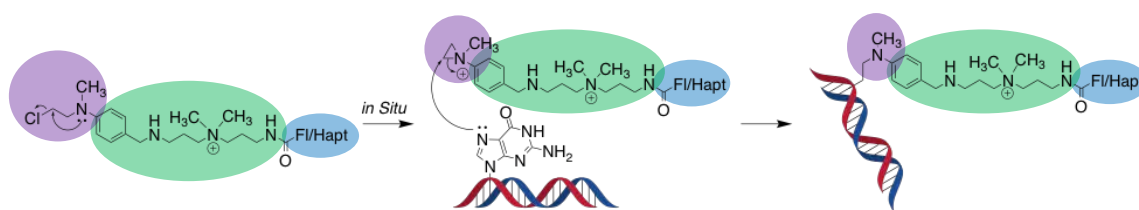


Figure 1.5 Schematic diagram showing how Label-IT® coupling agents can be used to covalently couple to nucleic acids for the introduction of fluorophores or haptens. Labelling is achieved through *in-situ* formation of an aziridine ring which is typically attacked by N^7 of guanine forming a covalent link with the nucleic acid of interest.

Primarily labelling is directed through attack of N_7 of guanine after formation of the aziridine intermediate. However, intercalation occurs randomly across the DNA sequence meaning specificity cannot be guaranteed as aziridine can also be attacked by other less nucleophilic amines when formed *in situ* in different positions across the backbone. With this method labelling density also needs to be considered as high fluorophore content decreases plasmid DNA (pDNA) transfection within cells, highlighting high modification densities do impact the behaviour of DNA⁵⁰.

A second commercially available non-specific labelling kit that has also been developed and alkylates DNA in a similar way to Label-IT® in the covalent modification of DNA is FastTag®, produced by Vector Laboratories⁵¹. This kit uses photoactivated aryl-azide chemistry to force the introduction of thiols into the DNA backbone through primary amines. Transalkylation is initiated by light or heat activation of the aryl azide coupling agent, stimulating nitrene formation, which then covalently reacts with primary amines in DNA (Figure 1.6). Once introduced the disulphide is then cleaved to release the free thiol where maleimide coupling can introduce a wide variety of permanent functionalities.

Typically, light activation is favoured with double stranded DNA as unnecessary heating can cause denaturation.

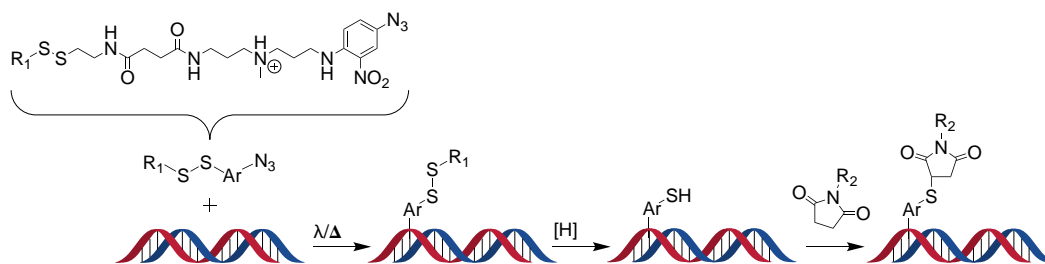


Figure 1.6 Covalent labelling of DNA using the random labelling kit FastTag®. The coupling reagent is first activated using heat or light to form the reactive nitrene which is attacked by primary amines in DNA in a ring expansion coupling. After coupling disulphide cleavage then releases the free thiol which can be used for tag introduction.

As with Label-IT®, FastTag® non-specifically reacts across the DNA backbone after the *in-situ* creation of an electrophilic intermediate and can introduce multiple modifications across the DNA. However, FastTag® has no targeting effects and can react readily with all primary amines within DNA, meaning labelling occurs with no specificity. FastTag® is beneficial as it can be used in *in vivo* applications due to the far-red dyes that can be coupled to DNA through maleimide chemistry after disulphide cleavage. For example, in transfection studies of labelled pDNA complexed with poly-ethyleneimine (PEI) in mice, where Qdot 655 could be introduced to DNA samples for monitoring transfection efficiency and DNA localisation⁵².

Additionally, FastTag® can also introduce alternative coupling agents as, after reduction, the free thiol moiety remaining can undergo reaction with any sulfhydryl reactive group. A further benefit with this chemistry not yet explored, is its potential as a responsive labelling system. By taking advantage of the dynamic disulphide bond and replacing R₁ with a chemical tag levels of

information could be written and erased into the initial modifications made to DNA (section 4.1) ⁵³.

The third distinct, non-specific labelling kit for DNA modifications is ULYSIS (Figure 1.7) ⁵⁴. In a similar manner to Label-IT®, ULYSIS targets N₇ guanine residues within DNA, but instead attaches non-covalently by ligand substitution at its square-planar Pt centre⁵⁵. The kits are available in a range of fluorescent dyes and have been used in photometric nucleic acid studies, including cDNA microarrays⁵⁶ and microRNA expression profiles⁵⁷. As with both other random labelling kits, ULYSIS can also introduce multiple modifications across DNA, which can be controlled with the reagent equivalents added. But, as with Label-IT® the target selectivity is not guaranteed and the dye may insert elsewhere in DNA.

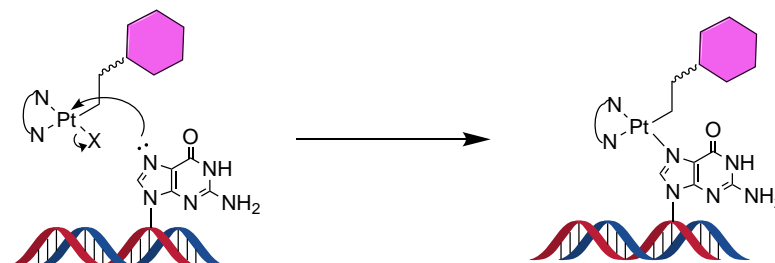


Figure 1.7 ULYSIS nucleic acid labelling takes advantage of the universal linkage system for binding which sees substitution at the Platinum centre of the N⁷ of guanine with the leaving group forming a coordinate bond.

In summary, the three distinct chemical labelling methods described highlight the versatility of non-specific labelling chemistries and their benefits as tools in the bulk modification of nucleic acids. These kits are beneficial as they do not alter the nucleic acid sequence like some enzymatic kits and often require one-step to label DNA. In addition, they are beneficial over intercalation of fluorophores as

tags are covalently bound to DNA. Something that must be considered with these labelling kits is the effect labelling may have on their interaction with any system, be it during or after labelling. For example, if completed in a biological mixture FastTag® labelling may non-specifically bind with proteins. Additionally, as the tag to DNA bp ratio increases the biological and physical properties of the labelled nucleic acid may change, especially in fluorescent labelling studies due to the hydrophobicity of dyes. Overall, as with any labelling chemistry there are trade-offs that need to be explored to identify the best labelling chemistry in each scenario.

1.2.2 Covalent, enzyme-directed DNA labelling

In addition to the chemical conjugation strategies outlined above, there are a second class which take advantage of natural enzymatic pathways. Enzyme mediated methods for bioconjugation offer similar opportunities to chemical alternatives but can alkylate native nucleic acids *in vitro*. As with chemical labelling strategies, depending on the enzyme or kit used, different specificity and targeting can be achieved. For instance, non-specific labelling with polymerases, end modifications with polynucleotide kinases, insertions with modified restriction endonucleases and sequence specific labelling with methyltransferases. All of which, with the exception of methyltransferases (which instead require modified AdoMet analogues), require modified nucleotides for labelling to be successful⁵⁸.

To facilitate enzymatic bioconjugation modified nucleotides have been developed which can introduce functionalities to nucleic acids, by the replacement or addition of bases to nucleic acids during labelling. As with chemical labelling strategies new functionalities introduced can then be used in the coupling of tags.

Some common modifications introduced include; fluorophores, biotin and click reactive groups. For pyrimidines modifications can be introduced either at C(4) or C(5) and with purines at C(8), depending on the synthetic pathway used^{59 60}. It is desirable for modifications to be introduced at carbon sites rather than amines present within bases to avoid interference in Watson and Crick base pairing.

1.2.2.1 Incorporation of modified nucleotides using PCR

A simple method for enzymatic modification is with polymerase enzymes during nucleic acid synthesis for both RNA and DNA by introducing either modified nucleotide triphosphates or deoxynucleotide triphosphates respectively during PCR. In the case of RNA polymerases, this can be done *in vitro* through the introduction of modified nucleotides to the reaction mixture in the presence of a DNA template yielding labelled RNA^{61,62}. Whilst for DNA a similar approach can also be taken but just with a change in the nucleotides and enzyme used⁶³.

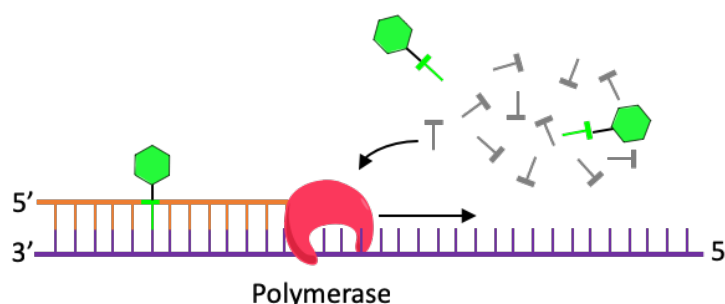


Figure 1.8 PCR induced modifications are achieved through the introduction of synthetic nucleotides during PCR which are then used by the polymerase enzyme during synthesis and introduced randomly across the backbone.

Using this approach, it is possible to incorporate many modifications to prepared DNA. Traditionally modifications were used for the detection of templated strands and was originally achieved with radio labelled nucleotides⁶⁴. However, these have now been replaced by fluorophores and biotin tags. Modified fragments can now also be used further in sequencing studies⁶⁵. As with chemical strategies

labelling density can be controlled by adjusting the ratio of modified nucleotides available, but again no specificity in the distribution of modifications is achieved. Although, in the case of PCR modifications depending on the nucleotide stock used, the exact base can be guaranteed, something not possible with random chemical strategies.

1.2.2.2 Enzymatic End Modifications

Enzymatic end modification of DNA is possible through manipulation of the enzyme terminal deoxynucleotide transferase (TdT). This enzyme is a polymerase but labelling conditions can be manipulated to promote the addition of 1-3 modifications to the 3' end of DNA. Primarily this enzyme is most efficient with single stranded DNA but can also be used in the labelling of 3' overhangs with lower efficiency⁶⁶. One benefit of this labelling is that it is template independent meaning no model strands are necessary to facilitate labelling.

Some common applications for TdT and its ability to introduce end modifications include RACE (rapid amplification of cDNA) and TUNEL (terminal deoxynucleotidyl transferase dUTP Nick end labelling) to facilitate cloning⁶⁷. It can also be used for a variety of detection, affinity and sequencing applications as well^{68,69}.

Alternatively 3' end modifications to RNA can also be achieved using the enzyme T4 RNA ligase, which introduces terminal 5' phosphates to a terminal 3' hydroxyl group (Figure 1.9)⁷⁰. Again, this labelling technique is template independent but does require the presence of single stranded RNA and ATP to proceed⁷¹.

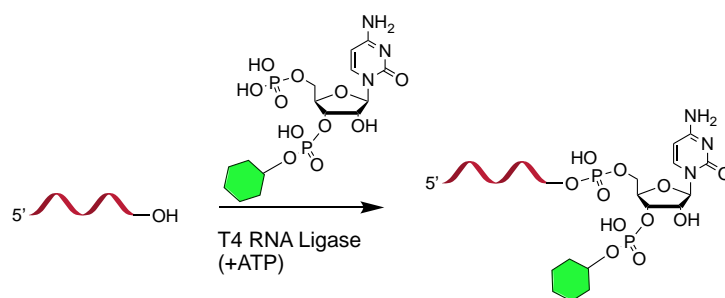


Figure 1.9 T4 RNA ligase can be used to introduce end modifications to DNA or RNA at the 3' end in the presence of ATP and a modified nucleotide bisphosphate like cytidine for the introduction of either fluorescent tags or biotin tags.

Further to T4 RNA ligase and TdT introducing nucleotides to the 3' end of nucleic acids, 5' end modifications are possible through the use of T4 Polynucleotide Kinase (PNK). Similarly to TdT, this enzyme is most effective in modifying single stranded nucleic acids and 5' overhangs, but is more versatile as it can modify shorter overhangs and blunt ends effectively and can also alkylate both DNA & RNA. One common application for PNK is in the phosphorylation of synthetic oligonucleotides to facilitate cloning⁷². However, it can also introduce common coupling groups like thiophosphate to facilitate the conjugation of alternative tags⁷³.

Overall, enzymatic end modification strategies provide an efficient means to introduce conjugation sites and probes to DNA and RNA. As with chemical end modifications a variety of functionalities can be introduced in a targeted way to both the 5' and 3' end of nucleic acids. Furthermore, there are also multiple commercial kits which, as with polymerase directed labelling, provide a simple means to achieve bioconjugated DNA. This, combined with the ability to add multiple probes make them an attractive proposition for labelling nucleic acids,

however they are less commonly used due to the cost of labelling compared to synthetic equivalents.

1.2.3 Nick Translation

A third method for the enzymatic modification of nucleic acids is nick translation (Figure 1.10). This method was originally developed Rigby and Berg in 1977 and takes advantage of natural cellular repair mechanisms for the insertion of modified nucleotides⁷⁴. To carry out labelling the nucleic acid of interest is first 'nicked' using a DNA endonuclease which creates terminal 3' hydroxyl groups at each nicking site that then act as a priming site for polymerase repair. Nicking enzymes work by introducing single strand breaks at specific recognition sites within double stranded DNA. After nicking the DNA is 'repaired' using DNA polymerase I which binds at each strand break and inserts a modified nucleotide from the reaction mixture stock. The polymerase then moves along the DNA with 5' to 3' exonuclease activity replacing each native base along the sequence, using the non-digested strand as a template. As with end modification enzymes, nick translation label in a template independent manner. Common applications for DNA labelled with this approach include fluorescence *in situ* hybridisation and radiolabelling^{75,76}.

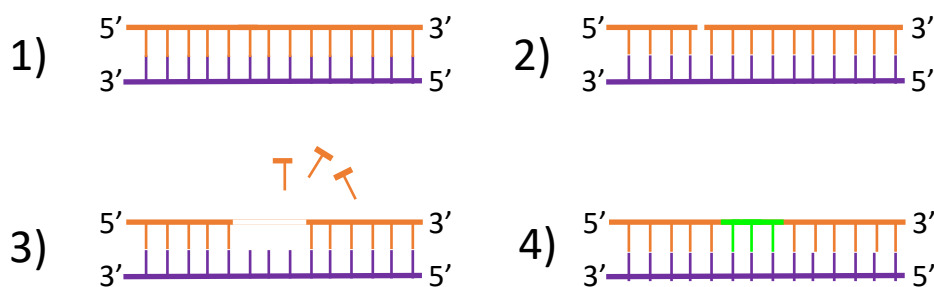


Figure 1.10 Schematic diagram showing how nick translation can be used to introduce modifications to DNA.

1) DNase I binds to the DNA, 2) a nick is created in the phosphate backbone creating the 3' hydroxyl priming site. 3) DNA polymerase I binds to damaged DNA removing nucleotides. 4) DNA polymerase I simultaneously replaces the removed nucleotides with modified nucleotides present within the stock and rejoins the DNA.

1.2.4 Summary

In summary, the bioconjugation strategies described give great flexibility in the introduction of modifications to nucleic acids for a variety of different applications. This can be achieved on both native and synthetic samples across the backbone depending on the method used either in a non-specific or targeted manner. These techniques are now widely understood and commercially available, often with one or two step labelling protocol, for any nucleic acid of interest. However, what has yet to be found is a technique that provides complete control on all nucleic acids over the location, density and selectivity of modifications. As such, this has been an area of interest in recent research which has begun to look at further enzymatic pathways where this might be possible. One area showing promise for this is through the use of methyltransferase enzymes and cofactor analogues which is the focus of this project.

1.3 Methyltransferase-directed bioconjugation to DNA

An emerging sequence specific field of bioconjugation sees the use of methyltransferase enzymes and cofactor analogues. Methyltransferases are a class of enzyme that exist within nature and play a key role in multiple pathways for the regulation of cells through the selective methylation of DNA, RNA, proteins and small molecules⁷⁷. In DNA, methylation directly impacts the function of a cell by affecting chromatin packing and gene expression⁷⁸. Through the simple introduction of a methyl group to nucleotides within DNA it methylation can silence or activate specific genes without causing any modification to the overall genetic code. Although originally not considered a significant factor in cell function, the under or over expression of methylation has now been directly linked to a host of diseases, especially hypermethylation of promoter regions in tumorigenesis^{79,80}.

Methyltransferases have drawn interest in bioconjugation reactions due their ability to label proteins, DNA, RNA and small molecules, whilst only depending on one methyl-donor cofactor, S-adenosyl-L-methionine (AdoMet). All methyltransferases which rely upon this methyl donor can be grouped within class I which are distinct as they each contain the Rossmann fold for binding AdoMet^{81,82}. The implication is that if the methyl-donor can successfully be redesigned to introduce more complex functionality, one labelling technique has the potential to be used with multiple biomolecules as a versatile tool for conjugation. A further benefit methyltransferases provide is sequence specificity, as each methyltransferase targets a single biomolecule and sequence. For example, the N⁶ adenine methyltransferase M.TaqI which targets DNA and only methylates adenine present within 5'-TCGAA-3'⁸³. This diversity and selectivity

offers unique opportunities for bioconjugation that cannot be achieved using traditional methods.

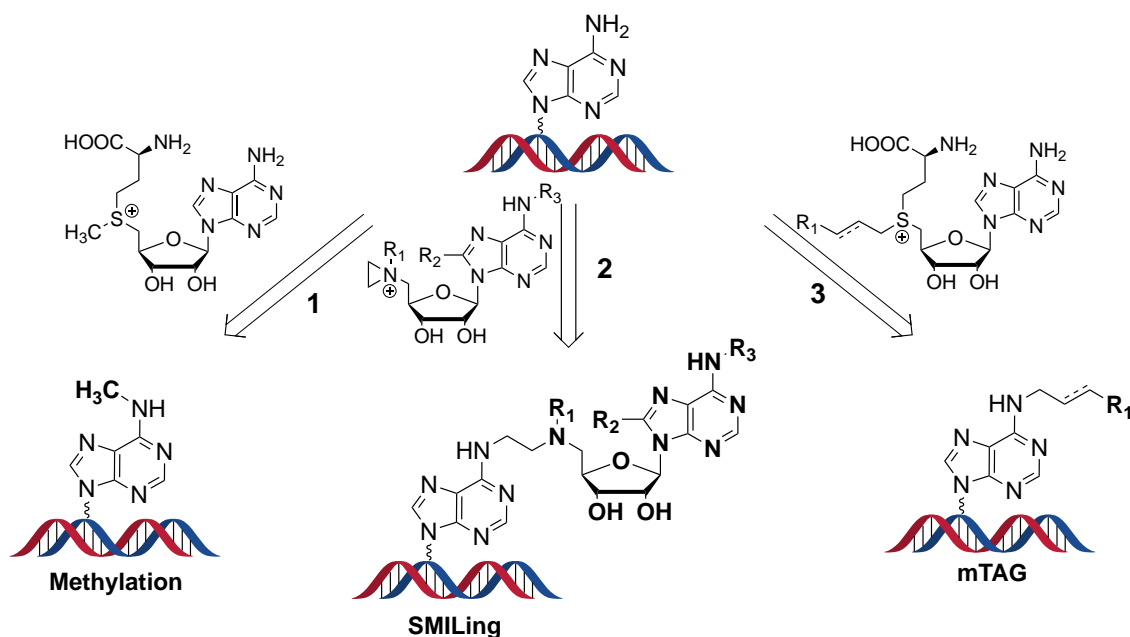


Figure 1.11 Pathways towards DNA modification where R groups highlight modification sites developed for the introduction of extended functionality. 1) Shows natural methylation with the class I methyl-donor cofactor S-adenosyl-L-methionine (AdoMet). 2) Demonstrates how entire nucleotides can be incorporated to DNA through Sequence-specific methyltransferase induced labelling (SMILing) with aziridine cofactor analogues when used with class I methyltransferases. 3) Identifies the pathway for methyltransferase-directed transfer of activated groups (mTAG) which is achieved with doubly-activated cofactor analogues.

1.3.1 Expanding methyltransferase function

Thus far, two major successes have taken advantage of the methyltransferase pathway for bioconjugation. These have seen the development of two new classes of AdoMet analogues which facilitate the introduction of more complex functionality to biomolecules. The earliest successful example of this arose through the development of aziridine based cofactors which introduce an entire nucleoside into the DNA backbone (Figure 1.12)⁸⁴. These were then further developed for the introduction of more usable functionality to DNA like azides,

alkynes and fluorophores^{85,86}. For example, Comstock *et.al.* successfully introduced azide functionalised adenosine to DNA using this pathway with the methyltransferases M.TaqI (TCGAA) and M.EcoRI (GAATTC)⁸⁷. Subsequently they were able to biotinylate the azide modified oligonucleotides by Staudinger ligation followed by immobilisation on agarose. Alternatively, the Weinhold group successfully developed a fluorescent aziridine cofactor analogue that fluorescently modified both pDNA and short oligonucleotides with the enzyme M.TaqI⁸⁸.

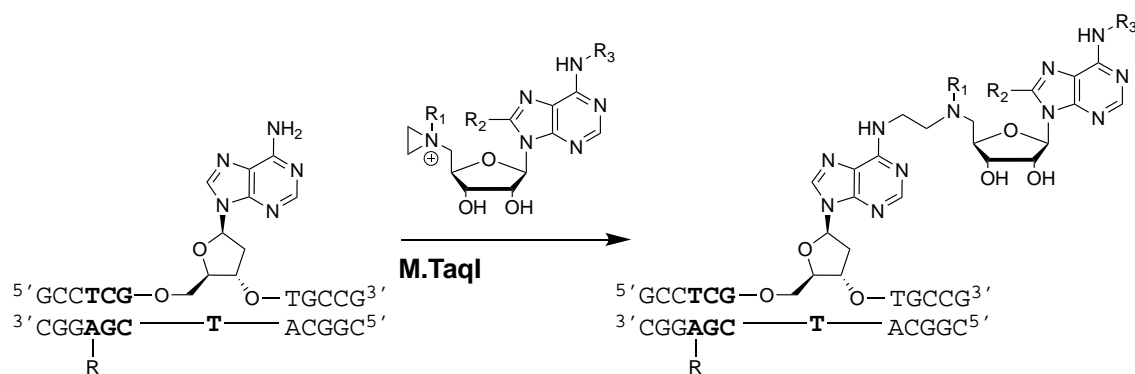


Figure 1.12 Aziridoadenosine AdoMet analogues alkylate DNA through the introduction of nucleotides into the backbone in a sequence specific manner depending on the methyltransferase used e.g. M.TaqI with the recognition site of 5'-TCGA-3'.

Although effective in the labelling of DNA, aziridine cofactor analogues do have drawbacks which severely limit their use as tools in bioconjugation reactions. The main drawback of these cofactors is, to get complete labelling stoichiometric amounts of methyltransferase are required, which is costly⁸⁹. This is because the product formed after labelling acts as a potent methyltransferase inhibitor preventing dissociation of the methyltransferase from the modification site⁹⁰. However, what these cofactors do show, is that it is possible to hijack the methyltransferase as a tool for sequence-specific delivery of chemical

modifications and if the cofactor analogue design can be refined to transalkylate catalytically this method has a lot of potential as a bioconjugation strategy.

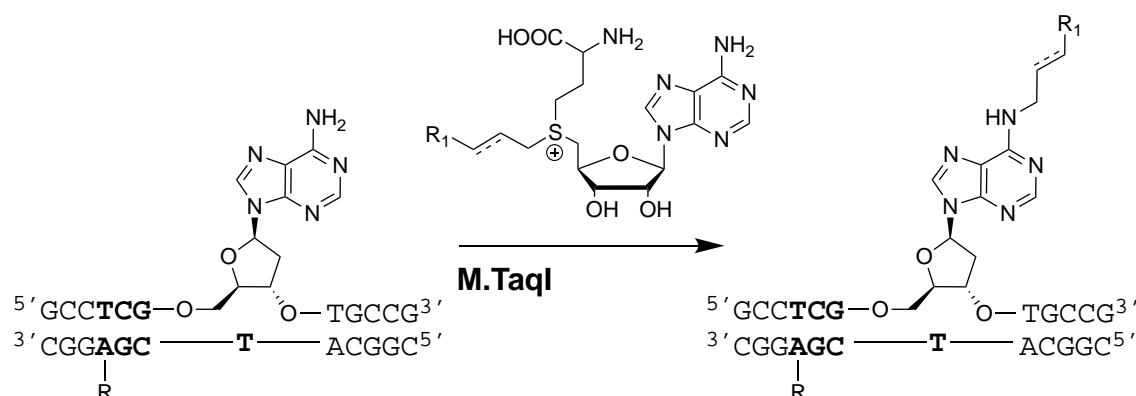


Figure 1.13 Doubly-activated AdoMet analogues have shown success through introducing alkyl chains into the backbone of DNA as opposed to an entire nucleotide. With the only prerequisite to labelling success being an unsaturated bond beta to the sulfonium centre.

The second class of AdoMet analogues designed have found more utility in bioconjugation reactions as they allow catalytic DNA modifications with methyltransferases. These so called 'doubly-activated' analogues are modified simply by replacing the methyl group attached to the sulfonium centre of AdoMet with varying alkyl side chains (Figure 1.13). Initial studies around these cofactors attempted the introduction of simple functionalities including ethyl and propyl substituents, but with limited success⁹¹. To facilitate efficient alkyl transfer an unsaturated bond is required beta to the sulfonium centre of the AdoMet analogue^{91,92}. It is thought this helps stabilise the transition state during alkylation through conjugative stabilisation, allowing transfer to occur (Figure 1.14). Since this discovery a catalogue of doubly-activated cofactors have now been formed. Their application with DNA methyltransferases is the focus of this review⁹³.

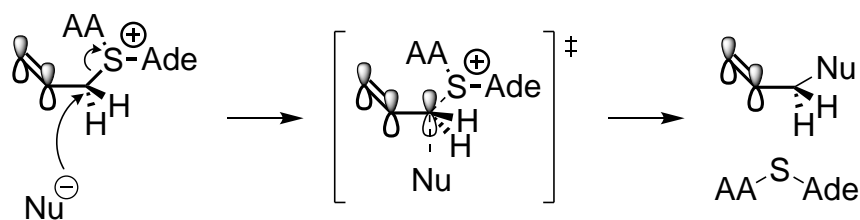


Figure 1.14 Schematic representation of how doubly-activated analogues serve to stabilise the transition state during nucleophilic attack through p-orbital overlap.

Further to their ability to allow catalytic modification of DNA by methyltransferases, doubly-activated analogues also have benefits over their aziridine counterparts due to their increased stability and ease of synthesis⁹⁴. The main synthetic approach adopted for these cofactors sees the coupling of an alkyl halide to the commercially available precursor S-Adenosyl-L-homocysteine (AdoHcy) followed by reversed phase HPLC separating the diastereoisomers formed. Alternatively, synthesis can also be achieved enzymatically *in situ* through coupling of chloroadenosine to L-methionine analogues in the presence of chlorinase SalL. This also removes the requirement for further purification, as only the naturally active diastereoisomer is formed^{95,96}.

Although doubly-activated analogues do have increased stability over aziridine cofactors they still encounter problems with degradation during labelling. As such, further to the introduction of more complex alkylating groups, work has also been carried out to improve their stability during labelling. Degradation of AdoMet analogues occurs following the same two key pathways seen with AdoMet through either loss of the amino acid or loss of adenine with the former by far the most prominent⁹⁴. In order to reduce these effects but maintain cofactor activity two key approaches have been tried. First sees carboxylic acid replacement within homocysteine with an isoteric tetrazole ring which drastically reduces

amino acid removal and maintains good labelling kinetics⁹⁷. Through replacement of the carboxylic acid a > 5-fold increase in cofactor analogue stability was seen compared to native AdoMet. This was achieved due to the tetrazole ring's increased steric bulk and decreased nucleophilicity preventing cyclisation. Whilst the second has explored the replacement of sulphur for heavier group 6 elements like selenium and tellurium, where tellurium had much improved stability due to a reduction in electronegativity⁹⁸. Through reducing the electronegativity of the chalcogenic centre a vast reduction in amino acid loss was seen, caused by a decreased electrophilicity of the surrounding carbon atoms again reducing cyclisation.

1.3.2 Examples of functionality built into DNA

Following the identification that an unsaturated bond beta to the sulfonium centre allows alkyl transfer with methyltransferases, a wide array of cofactors with varying functionalities have been developed⁹⁹. These have seen the use of a combination of unsaturated systems including ketones, alkynes, alkenes and benzylic functional groups (Figure 1.15). In the example of alkenes and alkynes similar R groups have been used in the literature including simple substituents like methyl and ethyl groups, click reactive groups including azides, amines and alkynes and extended functionality like fluorophores^{91,94,100}. Whilst for benzylic groups more complex functionality has been included. For example, Rentmeister *et.al.* have incorporated photoactive phenyl rings as both the transferrable group and unsaturated system through the inclusion of nitro groups or anthracene¹⁰¹. Additionally, they have also used phenyl rings in the unsaturated position to introduce norbornene derivatives which once introduced to DNA can be used in

traditional diels-alder cycloadditions with inverse electron demand¹⁰². Once introduced, further biomolecule functionalisation is then possible using reactions like azide alkyne cycloaddition. Through the use of a copper catalyst, cycloaddition has allowed the functionalisation of DNA with both a fluorophore and its attachment to beads^{22,103,104}. However, more recently dibenzocyclooctyne (DBCO) reagents have been preferred due to the reduced chance of oxidative damage they may cause to DNA (see section 1.1.1)^{29,31}.

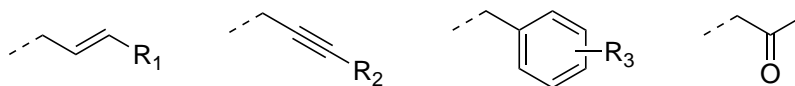


Figure 1.15 Shows the reported unsaturated functionality that has been used in doubly-activated cofactor analogues in the labelling of biomolecules. In the example of alkenes R_1 has included H, CH_3 , N₃, C=C and $\text{C}\equiv\text{C}$ ¹⁰⁵⁻¹⁰⁸. For alkynes R_2 has similarly included H, CH_3 , CH_2CH_3 , $\text{C}\equiv\text{C}$, N₃, NH_2 ⁹⁴. For phenyl ring R_3 has been NO_2 , Anthracene, or norbornene^{101,102}. Alternatively heterogenous unsaturated systems including ketones have also been used¹⁰⁹

Further to the introduction of simple side chains comprising an unsaturated bond and reactive tail group more complex modifications have also been introduced to DNA (Figure 1.15). For example, Ebenstein *et.al.* have successfully introduced fluorescently labelled cofactors directly to DNA using the methyltransferase M.TaqI¹¹⁰. This highlights the versatility of methyltransferase-directed labelling and the apparent lack of cofactor selectivity shown by these enzymes. This is of benefit, as it shows as long as the base AdoHcy structure is maintained there is a relative level of flexibility available in side chain design. However, it cannot be assumed all methyltransferases will show the same reduced cofactor specificity requirements between enzymes, as the cofactor binding pocket is not a conserved structure^{83,111}. More complex AdoMet analogues may reduce the

number of methyltransferases they are able to alkylate DNA with, as with each methyltransferase the binding pocket interactions and structures change depending on the cofactor analogue used. This could result in poorer binding of AdoMet analogues that contain bulky hydrophobic groups like fluorophores.

1.3.3 Methyltransferase-directed bioconjugation applications

One method in which methyltransferase-directed labelling has been used is the covalent attachment of fluorophores to DNA³¹. Although this can be achieved using traditional bioconjugation techniques, methyltransferase-directed labelling allows the quantity and location of the tags introduced to be known, meaning quantification of the fluorescence signal is possible. This is difficult to achieve using covalent attachment strategies which randomly introduce fluorophores to the DNA backbone with no sequence specificity (e.g. the Label-IT® kit)¹¹².

The specificity offered by methyltransferase-directed fluorescent labelling has also enabled the development of DNA mapping. Through the use of specific methyltransferases, fluorescent tags are introduced to genomic samples at known positions and can be used as markers within a genome. When linearized and imaged, alignment of the introduced markers can then create a fluorescent 'DNA barcode'¹¹³. This technique offers benefits over traditional approaches to optical mapping as it keeps the DNA intact during sample preparation and has much higher labelling densities¹⁰⁰. For example, a methyltransferase that has a 4 bp recognition site will occur one in every 256 bases, whilst optical restriction maps typically have modifications one in every 10-100 kb. This method suggests

a promising new way for extracting sequence information from DNA and has potential uses in the rapid identification of genomic DNA.

Methyltransferase-directed modification of DNA has also been applied in the study of the epigenome⁴¹. The epigenome is defined as structural modifications within a cell that impact on gene expression without affecting the DNA sequence. Thus far some of the key epigenetic modifications found to regulate gene expression include; 5-methyl-cytosine, 5-hydroxymethylcytosine, and protein/transcription factor methylation^{114,115}. Current methods for the analysis of the epigenome rely on the detection of 5-methyl-cytosine and include techniques like bisulfite conversion, affinity capture or restriction endonuclease methods¹¹⁶⁻¹¹⁹. Methyltransferase-directed labelling offers new opportunities as it instead targets unmethylated cytosine residues within the epigenome (unmethylome)⁴¹. This has the potential to improve targeting and simplify and reduce data collected for whole genome analysis which could assist in population-based studies. This has been achieved through the introduction of both amine and azide functionalities to genomic DNA using the CpG (Cytosine-phosphate-Guanine) specific methyltransferase M.SssI. CpG sites were specifically targeted for this process as it has been found within somatic cells this is where 5-methylcytosine almost exclusively exists¹²⁰. As with restriction enzymes, methyltransferases cannot modify methylated bases, meaning when labelling with AdoMet analogues unmethylated sites are selectively targeted.

After labelling at unmethylated sites, this DNA was selectively extracted through conjugation of either NHS/DBCO-biotin followed by capture onto streptavidin

beads. The DNA was then released by disulphide reduction yielding the purified unmethylated DNA fragments. No apparent differences were found between the azide and amine modified DNA with respect to capture and release efficiencies. However, some issues were found with azide modified DNA after release during PCR, which was attributed to the bulky DBCO modification that remained on the DNA. This should be considered when designing future methyltransferase-directed bioconjugation experiments as large hydrophobic modifications clearly impact the way in which DNA behaves with enzymes and in solution.

After release the DNA was sequenced and the results were compared to traditional approaches for studying DNA methylation like Methylated DNA Immuno Precipitation (MeDIP). Methyltransferase labelling showed clear benefits in data reduction as within the genome approximately 70% of CpG sites are permanently methylated and are removed during purification with this technique¹²¹. Furthermore, unmethylome capture was also able to avoid biasing DNA sampling toward high density CpG regions and gave consistently accurate results across the genome, something which can be problematic when using MeDIP^{122,123}.

A final, more recent application of methyltransferase-directed modification of DNA has seen the development of a new catalogue of reversible, photosensitive AdoMet analogues¹⁰¹. Here four analogues were designed which were similar in structure to doubly-activated cofactors but with no reactive tail group and the unsaturation required was provided by a phenyl ring as opposed to previously discussed alkenes and alkynes¹⁰². All analogues were successful in labelling of

DNA and after irradiation with UV light for 10 minutes complete reversal was seen for analogues containing either *o*-nitrobenzyl or 4-acetyl-2-nitrobenzyl groups. This technique offers an interesting new method for the reversible modification of DNA as after reversal the DNA is returned completely intact with no modifications present. Complete reversal is something which has not been previously reported for synthetic AdoMet analogues and more closely resembles the behaviour of methylation in nature which is a dynamic process. In the future further modifications will likely be introduced to these cofactors to allow further functionalisation after labelling, expanding their utility and potential applications.

1.4 Conclusion

1.4.1 Overview

Through the development of both enzymatic and chemical conjugation strategies it is now possible to manipulate DNA in a variety of ways each of which can help to further our understanding. Here a range of DNA bioconjugation strategies have been outlined with each of their strengths, weaknesses and applications identified. In general compromises are often found for each technique which limit their applications and transferability. For example, with non-specific labelling strategies like Label-IT® or the use of modified nucleotides during PCR, modifications can be introduced to both synthetic and natural DNA. However, both techniques lack control leading to indiscriminate labelling where the quantity and position of tags is not known. In the case of introducing end modifications to DNA, labelling is limited by the quantity of tags that can be introduced and the types of DNA that can be modified, like synthetic oligonucleotides with phosphoramidite chemistry.

Methyltransferase-directed labelling though helps combat many of these shortfalls in the existing toolbox of bioconjugation reactions. As a technique it allows for much greater control and flexibility over the number of modifications made to DNA. Crucially it is also sequence specific, meaning labelling only occurs in defined positions dictated by the enzyme used. These unique properties not only make it a useful tool in traditional bioconjugation applications but open up new avenues for exploration like in DNA mapping and next generation sequencing. Moreover, it does not suffer from off-site targeting effects or risk damaging DNA as it takes advantage of natural pathways to introduce labelling.

1.4.2 Aims & Objectives

The overall aim for this project is based upon the development of new applications for methyltransferase labelled DNA through the creation of responsive labelling systems. To achieve this goal, the first objective is to design and synthesise a new class of cofactor analogues which can introduce reversible functionality directly into DNA. For this, we plan to take advantage of previously explored dynamic covalent chemistry systems with a focus on Schiff base chemistry. Once successful, the cofactor analogues will then be tested in the labelling of DNA with methyltransferase enzymes to test their efficiency against previously reported structures. After labelling the reversible binding site now incorporated into the backbone will then be tested through reversibility studies to test the flexibility and efficiency of the Schiff bases used. It is then hoped to apply the responsively labelled DNA in a variety of applications with a specific interest

in selective capture and release for next generation sequencing studies. Overall though, this labelling system will be unique due to its ability to undergo multiple modification cycles through competitive binding, allowing for a variety of applications beyond capture and release (Figure 1.16).

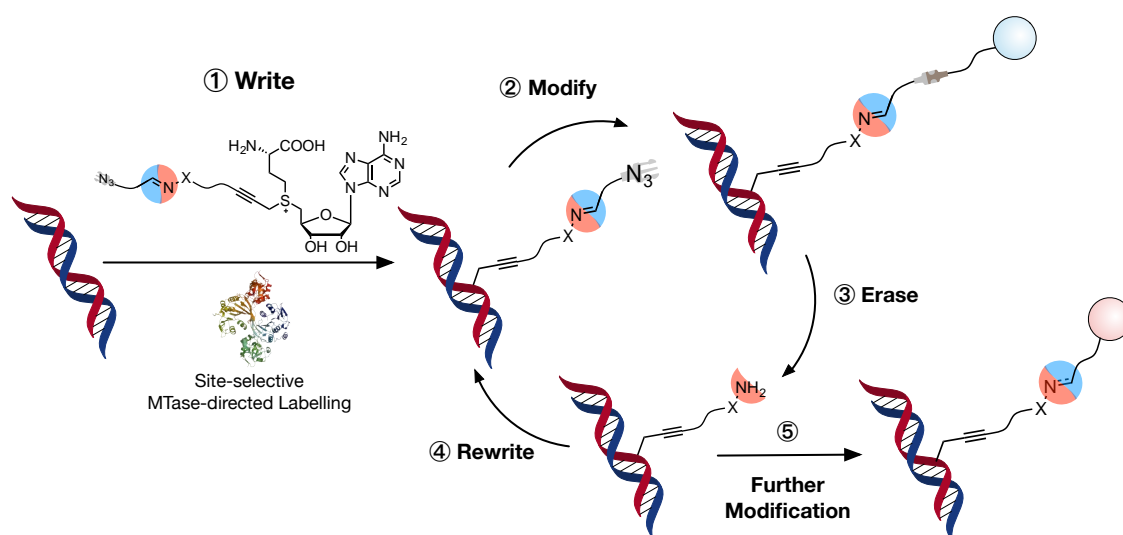


Figure 1.16 Schematic representation of the reversible and rewritable modification of DNA. Site-selective MTase-directed writing of DNA (1) allows for modification via azide-alkyne cycloaddition (2). The introduced functionality can then be erased via dynamic exchange (3), to give an intermediate DNA that can be rewritten via Schiff-base formation (4) to give the original functionality introduced via MTase-labelling. Alternatively, this intermediate DNA can be further functionalized via standard conjugation techniques (5) to give functional DNAs for follow-on research.

To achieve these objectives, three key stages will be explored, first of which is the design and synthesis of cofactor analogues (Chapter 2). In addition to the responsive cofactors desired a variety were also synthesised to either use as labelling efficiency comparators or as tests to try and improve the traditional synthetic pathway. Through the use of a new synthetic route it was hoped the scalability, yields and costs traditionally associated with doubly-activated analogues could be improved. Whilst a previously reported cofactor was also synthesised to be used as a benchmark throughout all labelling studies.

Following the successful synthesis of all cofactors the second stage in this project was to test their efficiency in the labelling of DNA (Chapter 3). Here the main aim was to gain a comprehensive understanding of cofactor interactions with both methyltransferases and DNA during labelling. Hopefully providing further confirmation around the structure and function of each cofactor. For these studies gel electrophoresis was initially used as a qualitative measure of labelling efficiency as seen in previous cofactor studies. This was then followed by the development of new quantitative labelling studies using LCMS to prove the structure of the modification introduced and give further confirmation of the original cofactor structure.

The third and final area explored was in testing the responsive cofactors developed. Here studies focused initially on reversibility testing but once successful, moved on to DNA capture and release followed by use in nanopore sequencing studies. In addition, a model system was also developed to test the efficiency of repeated modification cycles again through a combination of gel electrophoresis and LCMS.

2 Design and synthesis of dynamic AdoMet analogues for use with Methyltransferase enzymes

2.1 Background/Introduction

Through the synthesis of methyltransferase cofactor analogues, it has been possible to introduce a wide variety of different chemistries and tags into biomolecules of interest including DNA, RNA and proteins. Thus far, two main classes of cofactor analogue have been discovered which can achieve this, either those with an aziridine centre or so called doubly-activated analogues with unsaturated bonds beta to the sulfonium centre (see section 1.3.1)⁹³. Since their discovery both classes have successfully been used to introduce extended functionality into biomolecules, including azides, alkynes, fluorophores, and biotin into DNA, RNA and proteins⁹⁴. However, as discussed previously, applications of aziridine cofactors are limited due to their enzymatic inhibition caused post labelling by the complete introduction of nucleotides; meaning prohibitively large quantities of enzyme are needed for labelling on a useful scale. As such, for the purposes of this project, work focused on the development of new AdoMet analogues based on the structure of previously successful doubly-activated cofactors.

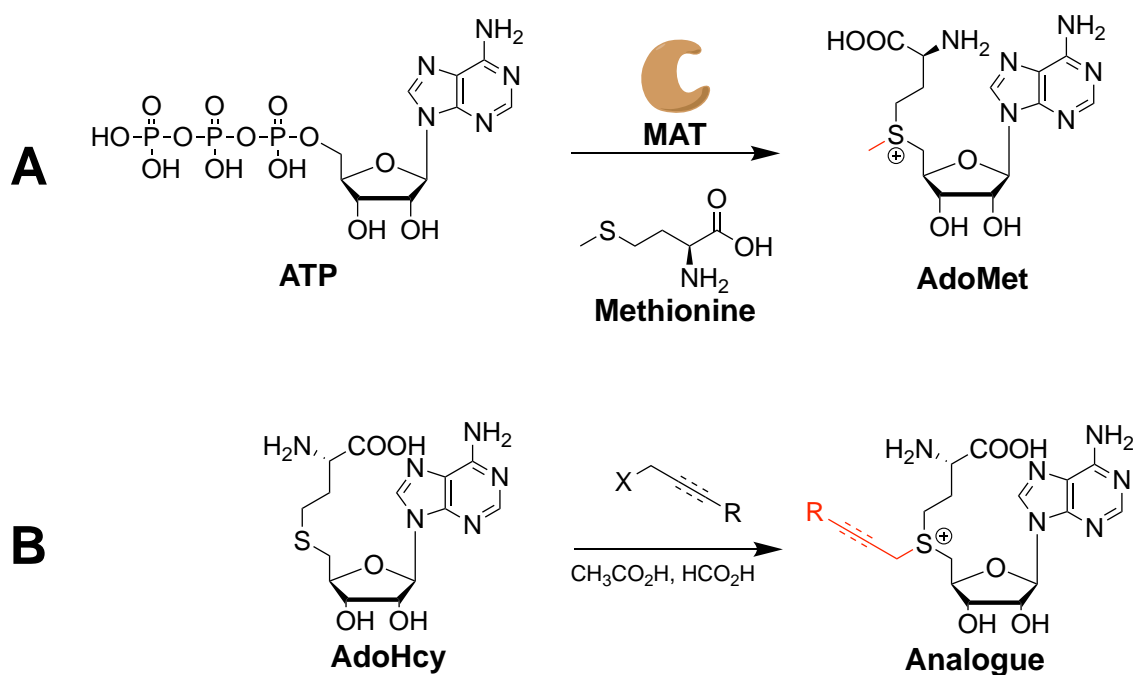


Figure 2.1 A) The natural synthetic pathway of S-adenosyl-L-methionine (AdoMet) with the enzyme Methionine Adenosyl Transferase (MAT). B) The general coupling mechanism for doubly-activated AdoMet analogues with AdoHcy and an activated alkyl linker where X is typically a good leaving group like a halogen or sulfonate.

Traditionally, the synthetic pathway employed for synthesis of doubly-activated analogues is simple and comprises one key step with creation of the sulfonium centre. In nature this is achieved through combination of methionine and adenosine triphosphate by the enzyme methionine adenosyl transferase (MAT) (Figure 2.1.A)²⁴. Some success has been had in the creation of AdoMet analogues enzymatically through the use of modified L-methionine substrates with chloro/fluoro-5'-deoxyadenosine by the chlorinase SalL from *salinispora tropica*^{95,96}. However, so far this has only shown limited effectiveness due to the reduced efficiencies of SalL with modified L-methionine substrates. As such, synthetic analogues are instead created by addition of an alkyl moiety to AdoHcy through nucleophilic attack at sulphur with activated alkyl precursors (Figure 2.1.B). Attack preferentially occurs at sulphur over the amine due to addition of

acid in the reaction mixture which decreases the nucleophilicity of the amines present. This reaction has now been widely explored and has led to the development of an array of cofactor analogues. Some drawbacks of this reaction are yields, scalability and a lack of diastereoselectivity, but overall the reaction has been widely utilised^{41,91,101,125}.

When designing new AdoMet analogues the base AdoHcy structure is rarely changed as this region forms many important contacts to the cofactor binding pockets of methyltransferases and so may significantly impact enzyme kinetics⁸³. As such, when designing new doubly-activated cofactor analogues the base AdoHcy structure is generally conserved and only the methyl group is modified, with alternative alkyl sidechains. In early analogues, this saw the introduction of simple replacements like ethyl and propyl groups, but poor transfer efficiencies limited their effectiveness as cofactor analogues⁹¹. To facilitate efficient alky transfer Dalhoff *et.al.* employed an unsaturated bond beta to the sulfonium centre. This provides conjugative stabilisation to the transition state through p-orbital overlap and increases reaction rates⁹². Once introduced, doubly-activated alkenyl and propargyl analogues were then used to effectively label DNA in a sequence specific and catalytic manner. Using this concept, an array of different activated cofactors have now been developed with multiple chain lengths and functional groups⁹³.

Here we develop a new range of responsive doubly-activated cofactor analogues. This area has been driven by the expanding number of applications in which methyltransferase-directed bioconjugation can be used. For example,

photocleavable cofactors can completely reverse biomolecule labelling when exposed to UV light and have potential applications in studying labelling in model natural systems³⁸. Whilst post modification labelling strategies like those developed by Klimasauskaus can be used for the selective targeting of biomolecules through the use of disulphide redox chemistry⁹⁸.

We are interested in furthering this area of research through the development of a new class of rewriteable cofactors, which can undergo multiple response cycles for serial biomolecule analysis (Figure 2.2.A). For this we identified Schiff base chemistry as it has been widely explored in alternative bioconjugation applications but hasn't yet been used as a responsive MTase labelling method¹²⁶⁻¹²⁸. Schiff base chemistry was highlighted as the dynamic chemistry of choice because the hydrolytic stability of bases formed is tuneable depending on the amine functional group used during coupling¹²⁹. Here we were specifically interested in the use of hydroxylamines and hydrazides as when coupled to carbonyls they form either an oxime or hydrazone respectively which can be cleaved in acidic solutions¹³⁰. We hoped to take advantage of this cleavability after conjugation of the oxime/hydrazone to DNA which will be discussed in detail in chapter 4.

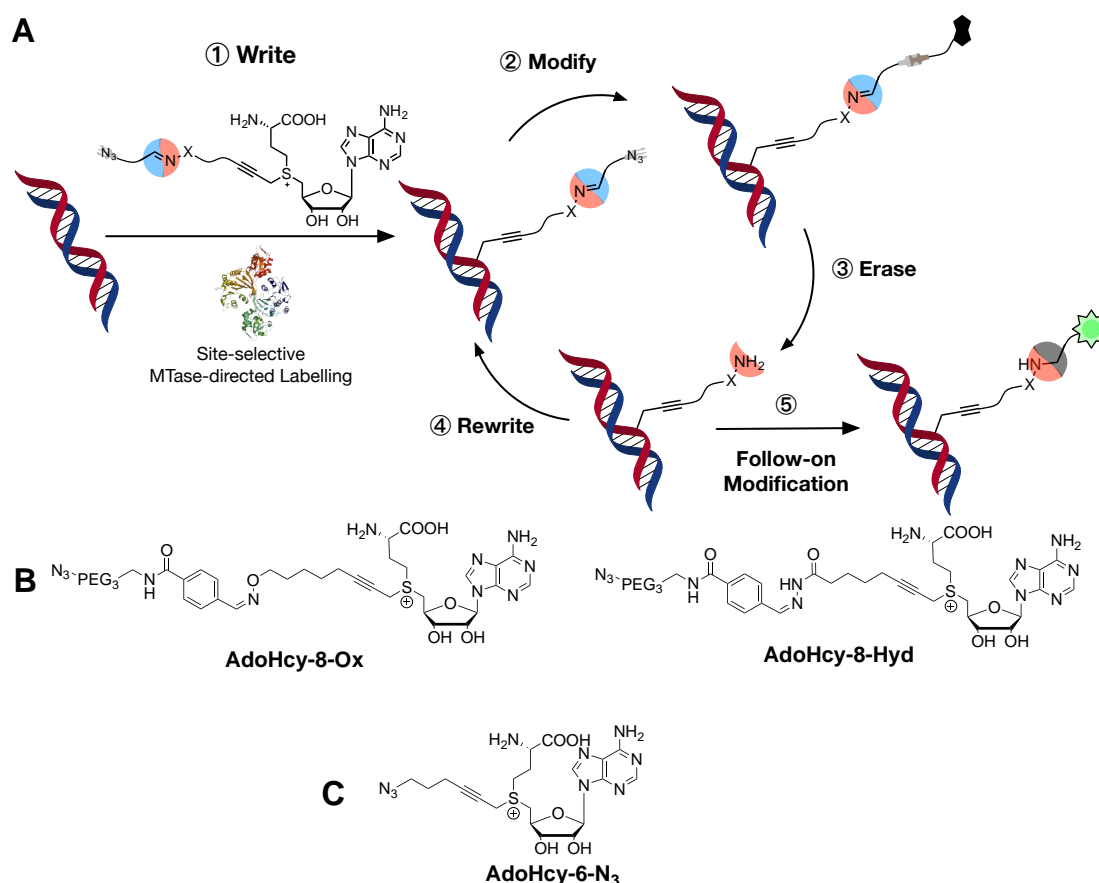


Figure 2.2 A) Highlights how Schiff base chemistry can be utilised repeatedly once incorporated into biomolecule's for multiplexed analysis. A biomolecule of choice will first have the Schiff base introduced where further functionalisation can occur followed by complete reversal to the free amine. After recovery of the amine modified biomolecule it will then be possible to repeat or end the cycle. B) Shows the two new cofactors designed containing either an oxime or hydrazone within the middle of the transferrable chain. C) The structure of a widely used previously reported cofactor with azide functionality, which has shown success with multiple methyltransferases.

Taking into consideration all areas discussed above we came up with the design and synthesis pathway for our new responsive cofactors **AdoHcy-8-Ox** and **AdoHcy-8-Hyd** (Figure 2.2.B). For this we adapted the synthesis pathway of a previously reported cofactor **AdoHcy-6-N₃**. First this sees the synthesis of an activated alkyl precursor containing both the reactive functionality and unsaturated bond required for doubly-activated cofactor analogues, followed by its coupling to AdoHcy¹³¹. Initially we synthesised the cofactor **AdoHcy-6-N₃** as

a standard to test methyltransferase activity with and to use as a comparator in future labelling studies against **AdoHcy-8-Ox/Hyd**. Each of the three cofactors were synthesised in the same way as previously, using AdoHcy in the presence of the relevant alkyl precursor and acid. However, the synthesis of the responsive cofactors varied slightly as the Schiff base was not introduced until after coupling. Instead the cofactors were coupled to AdoHcy with Boc protected versions of the amine precursor as it was originally intended to use the cofactors in either the free hydrazide or hydroxylamine form. After coupling the cofactors were deprotected, purified and finally had the Schiff base functionality introduced before labelling (see section 2.4.4).

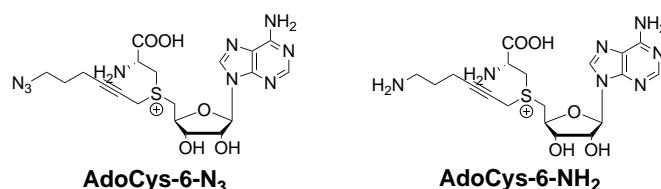


Figure 2.3 Shows the structure of the final two cofactors with a change to the amino acid side chain. Synthesis of the cofactors was achieved from adenosine in simple and high yielding steps which is discussed in detail in section 2.3.

In addition to the three cofactors outlined above we synthesised two more cofactors of interest where the amino acid side chain was replaced with cysteine (Figure 2.3). These were created as part of a separate project to investigate a more economical and scalable synthetic pathway for doubly-activated analogues. We also anticipated that amino acid substitution would yield more stable cofactors as the main degradation pathway of AdoMet, through loss of the amino acid, would become unfavourable due to the decreased stability of the 4 membered ring that would need to be formed⁹⁷. It was hoped this change would not lead to a significant decrease in enzymatic activity, as previously methyltransferases have shown large flexibility in donor cofactor analogue

structure. Furthermore, we hoped to improve reaction yields during creation of the sulfonium centre using lactone ring opening as opposed to traditional nucleophilic substitution. Another drawback of doubly-activated analogues is their low conversion in the coupling of linkers with AdoHcy⁹¹.

2.2 Precursor Synthesis

For all cofactors the first stages of synthesis required the creation of one of three precursors which could then be used in coupling to either thioadenosine or AdoHcy (Figure 2.4). For cofactors **AdoCys-6-N₃/NH₂** and **AdoHcy-6-N₃** the precursor **4a** was synthesised according to the literature⁹⁴. Briefly, this comprised; hydroxyalkylation of 5-chloropentyne, followed by azodination and subsequent sulfonate activation⁹⁴. After the synthesis of **4a** the two linkers required for **AdoHcy-8-Ox/Hyd** were designed. Here an extended C-8 alkyl chain was chosen to maximise separation between the nucleophilic amine and the sulfonium centre, hopefully improving cofactor stability⁹⁴. In an effort to simplify synthesis the same general three step protocol for **4a** was used.

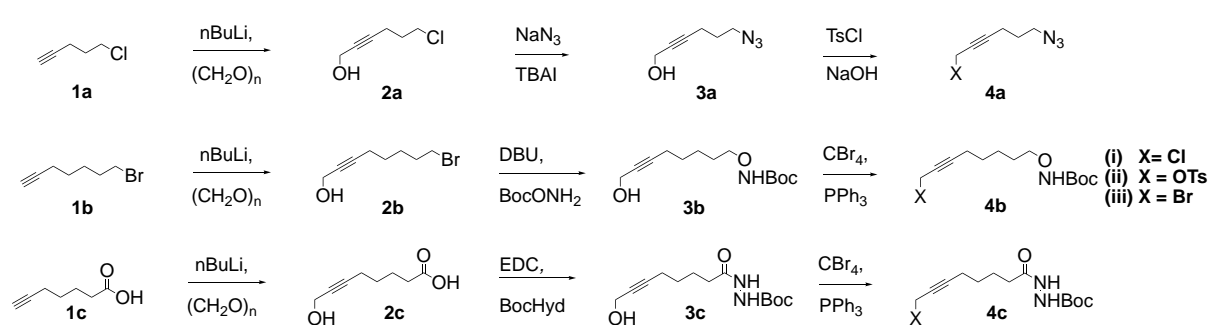


Figure 2.4 Shows the synthetic route adopted to create each activated linker. Each linker synthesis can be broken down into 3 key steps; hydroxyalkylation, functional group introduction and alcohol activation. For each linker the activating group was dependent on the analogue to be synthesised.

All starting materials were commercially available except for 7-bromoheptyne (**1b**), which was synthesised by bromination of 6-heptyn-1-ol (Figure 2.5.A).

Bromination was achieved using the Appel reaction with carbon tetrabromide (CBr_4) and triphenylphosphine (PPh_3)¹³². The procedure was taken from the literature and yields of over 90% were consistently achieved¹³³. The greatest difficulty for this reaction came with purification due to the difficulty in removing the organophosphorus reagents. To ensure complete removal, both cold filtration and flash column chromatography (FCC) were necessary. The presence of organophosphorus reagents was monitored by ^1H NMR where clear peaks in the aromatic region could be seen. Residual impurities were found to significantly impact yields during hydroxyalkylation and so necessitated thorough purification.

After purification the product was characterised using a combination of Infra-red spectroscopy (IR) and ^1H NMR (Figure 2.5). When overlaying the IR of the starting material against the product complete removal of the O-H stretch ($3000\text{--}3500\text{ cm}^{-1}$) and the introduction of a new sharp band at 540 cm^{-1} due to the C-Br stretch could be seen. Further confirmation came from the NMR where again complete removal of the OH peak was seen at 4.4 ppm. As well as a downfield shift ($3.4\rightarrow 3.6\text{ ppm}$) and change in splitting of the CH_2 nearest the reactive site ($\text{dt}\rightarrow\text{t}$). Following synthesis and purification, **1b** was then used directly in hydroxyalkylation.

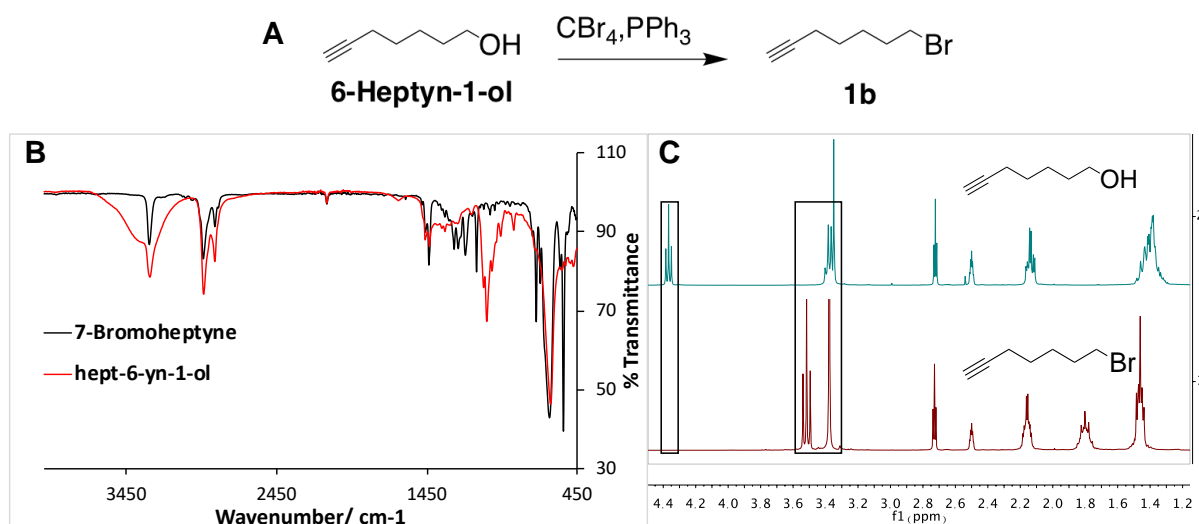


Figure 2.5 A) bromination of 6-heptyn-1-ol. B) an overlay of IR traces collected of 7-bromoheptyne (**1b**) compared to the starting alcohol (hept-6-yn-1-ol). Loss of OH at $3000\text{--}3500\text{ cm}^{-1}$ and introduction of sharp C-Br stretch at 540 cm^{-1} . C) Stacked ^1H NMR spectra of the alcohol (top) and bromide (bottom), complete OH removal is seen (4.4 ppm) as well as a shift and change in splitting of the CH_2 signal adjacent to the reactive site ($3.4 \rightarrow 3.6\text{ ppm}$).

2.2.1 Hydroxyalkylation

Hydroxyalkylation of **1a** was achieved using paraformaldehyde and *n*-butyl lithium ($n\text{BuLi}$)¹³⁴. No alterations were made to the original procedure and the product was subsequently purified by vacuum distillation. Distillation was simple to achieve as the product was found to be the first fraction collected and all remaining impurities were not volatile at the temperatures explored. Care had to be taken to ensure complete removal of all solvents prior to distillation as any that remained caused issues with vigorous boiling under reduced pressure. During vacuum distillation it was necessary to heat the crude product to $180\text{ }^\circ\text{C}$ and a distillation temperature of $120\text{ }^\circ\text{C}$ was recorded. This was significantly higher than the referenced distillation temperature of $55\text{--}57\text{ }^\circ\text{C}$. It is likely this change occurred due to variations in the vacuum pressure and was not due to a change in compound characteristics. This meant thermal degradation may have been an issue however, this was found not to be the case during subsequent

characterisation of the product. When running the distillation on a test scale (200-500 mg) the yield of purification was low, which was attributed to product adhering to the distillation apparatus. Increasing the scale of purification helped to improve yields greatly as the quantity of product sticking to the apparatus during distillation did not change.

After the successful synthesis of **1a** the same approach was used for **1b**. Residual impurities from the bromination of **1b** were found to cause significant reductions in yield. It is likely that any remaining organophosphorus reagents from the previous bromination could interact with *n*BuLi and either cause unwanted side reactions or reduce the expected equivalents of the base being added¹³⁵. For the purification of **1b** vacuum distillation was not attempted due to the increased chain length of the linker, which would significantly increase the distillation temperature. This, combined with the decreased stability of bromides meant thermal degradation during distillation would be a much larger risk. Instead it was possible to carry out purification using FCC where similar yields were achieved.

The clearest indicator for success of both hydroxyalkylations came from ¹H NMR. Here a clear introduction of the new CH₂ and OH peaks present could be seen (Figure 2.6, Sup. Fig. 2.17). For both linkers the CH₂ introduced appeared at approximately 4.0 ppm matching with the reference for **2a** with doublet or triplet splitting, whilst when resolved, the OH was seen at approximately 5.1 ppm. Once the structures were confirmed both linkers were then taken forward for functional group introduction.

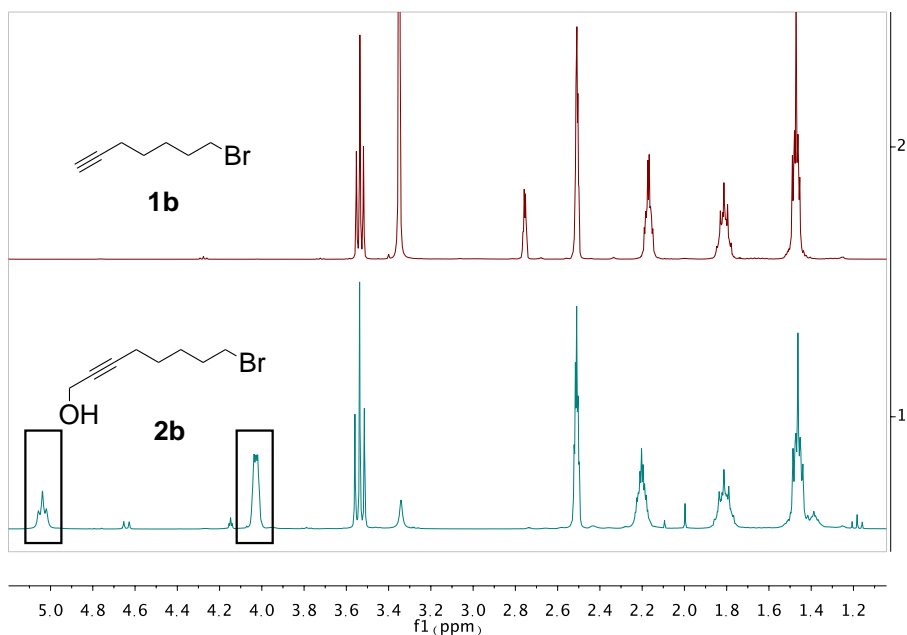


Figure 2.6 Top: ^1H NMR of linker **1b**, Bottom: ^1H NMR of linker **2b** after hydroxyalkylation and purification. Two clear new peaks can be seen; a doublet of triplets at 4.0 ppm due to the CH_2 introduced and a triplet at 5.1 ppm from the OH.

With the final hydroxyalkylation to form **2c**, the original procedure using paraformaldehyde and nBuLi had to be adapted as the carboxylic acid present is deprotonated preferentially to the alkyne, hence the number of equivalents of nBuLi were increased¹³⁶. Initially there were significant issues with low yields (ca. 10-15%) which were attributed to the low percentage conversion seen during reaction (Figure 2.7.B). This was due to precipitation of the starting material after conversion to its lithium salt, limiting its reactivity. In an attempt to improve yields, two approaches were taken; 1) using 5-bromovaleric acid and propargyl alcohol as an alternate route from the literature, 2) optimisation of the current procedure (Figure 2.7.A).

Coupling using 5-bromovaleric acid and propargyl alcohol was attempted in the synthesis of **2c** as it has been frequently used in the making of similar

compounds^{137 138 139}. In the coupling of **2c** however, synthesis was unsuccessful. Initially, complete removal of the starting material was observed. However, after purification, product characterisation did not show the result expected. Despite the mass corresponding to that of the expected product, the ¹H NMR showed the CH₂ peak introduced (4.1 ppm) with doublet splitting rather than the doublet of triplets observed for **2a/b** (Figure 2.7.C). Furthermore, the second peak introduced had an unexpected chemical shift (3.4 ppm) compared to **2a/b** (5.1ppm) and did not undergo proton exchange after a D₂O shake. As such, it is likely the reaction was unsuccessful and instead **2ci** was the predominant product (Figure 2.7.A). According to the literature preferential formation of **2ci** occurs when reagent addition is carried out at higher temperatures¹⁴⁰. This means, to allow for successful synthesis of **2c** using this approach, the temperature of reaction likely needs to be reduced well below the -78 °C used to gain complete thermodynamic control. However, further exploration into the effects of temperature control on this reaction were not explored as the yields of the original approach were dramatically improved, which was favourable as no competing reactions occurred.

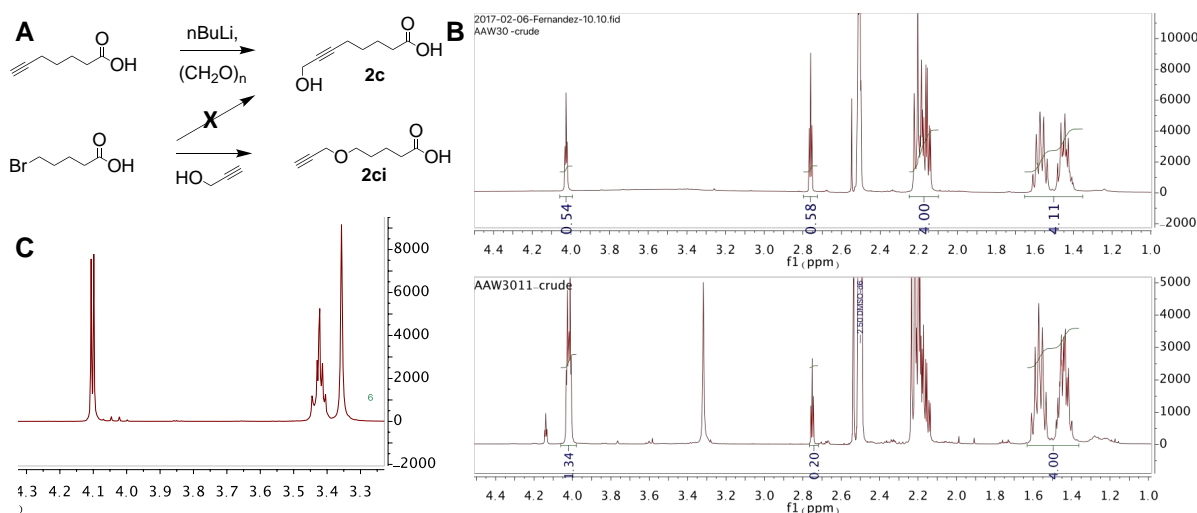


Figure 2.7 A) Synthetic routes attempted for the synthesis of **2c**. B) ^1H NMR of the crude reaction mixture using $n\text{BuLi}$. Top = initial, Bottom = optimised. % conversion was measured by integration of the CH_2 being created at 4.0 ppm and normalised to the middle CH_2 's in the chain (2.2/1.5 ppm). If 100% conversion 4.0 ppm int = 2. So % conversion = $(0.54/2) \times 100 = 27\%$. After optimisation: $(1.34/2) \times 100 = 67\%$. C) ^1H NMR of **2ci**, a doublet is seen for the CH_2 at 4.1 ppm and a triplet at 3.4 ppm (although residual Et_2O overlaps with this signal).

Optimisation of the original approach for the formation of **2c** was achieved through the introduction of hexamethylphosphoramide (HMPA) as a co-solvent and adjusting temperatures during addition (Figure 2.7.B)¹⁴¹. We believe inclusion of HMPA helped to improve the reaction twofold; first by improving the solubility of the lithium salts formed after deprotonation, and secondly by breaking down $n\text{BuLi}$ oligomers present into their reactive form. Additional improvements in conversion also came from adding paraformaldehyde at a lower temperature (-40°C) compared to **2a/b** (0°C). After reaction, purification was achieved using FCC and followed by ^1H NMR characterisation (Figure 2.8). As seen previously in the NMR of both **2a/b**, there was the introduction of a CH_2 and OH peak at 4.0 and 5.0 ppm respectively. Further confirmation of the alcohol could then be achieved by D_2O shake, where complete exchange was seen.

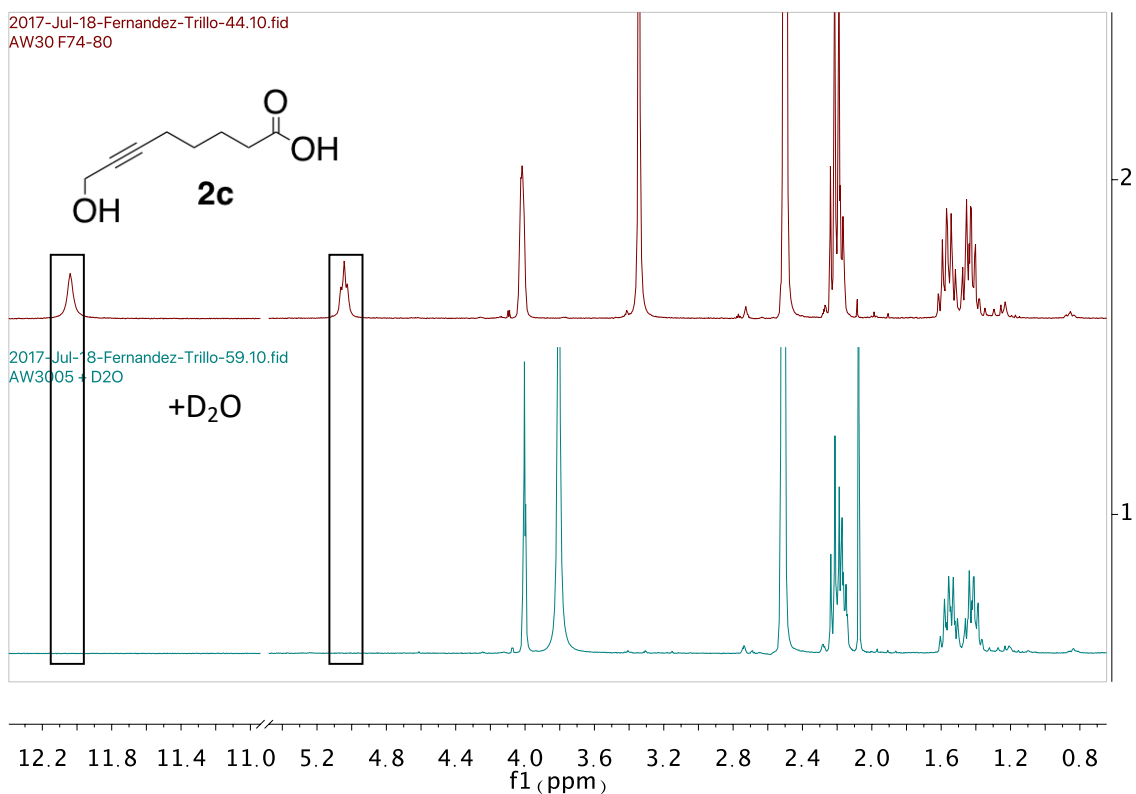


Figure 2.8 Characterisation of **2c** by ^1H NMR. Top: before D_2O shake, a clear peak for the OH can be seen at 5.0 ppm (same as seen for **2b**) and the carboxylic acid at 12.0 ppm. Bottom: ^1H NMR after D_2O shake. There is a complete removal of both OH peaks present in the NMR.

2.2.2 Functional group introduction

Following successful hydroxyalkylation the functional groups for each precursor needed to be introduced. For **3a** this was completed, by following the literature method with sodium azide and tetrabutylammonium iodide (TBAI). Whilst for **3b/c** this was carried out by the introduction of Boc (tert-butyloxycarbonyl) protected versions of the desired amines.

Azodination of **2a** was completed by heating in the presence of sodium azide and TBAI. Once complete the reaction was purified by extraction where success could then be measured by comparison with the reference and easily seen with both IR and UV (Figure 2.9). By overlaying the IR of **2a** and **3a** it was possible

to see the introduction of the azide stretching frequency at ca. 2100 cm^{-1} matching the reference.

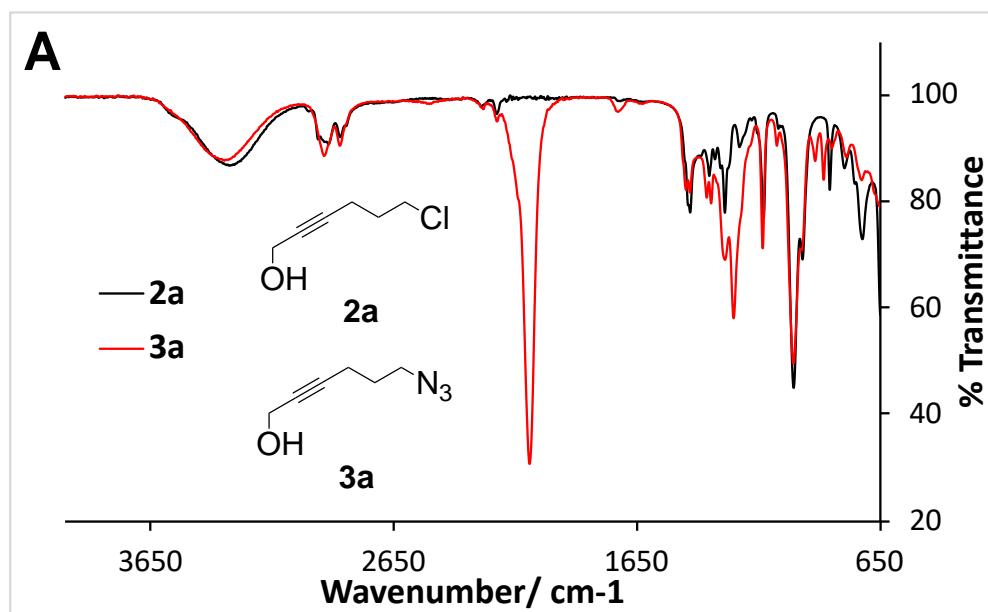


Figure 2.9 A) IR overlay of **2a** and **3a** following reaction with sodium azide. A clear new band can be seen corresponding to the $\text{N}=\text{N}$ stretch at 2100 cm^{-1} from the azide.

For the hydroxylamine and hydrazide functionalities of **3b/c** boc protecting groups were necessary due to the nucleophilicity of the free amine which would decrease the selectivity of future couplings. Boc was highlighted as a suitable protecting group for these linkers as it had been used in similar AdoMet analogues that have been synthesised and so is known to not prevent AdoHcy coupling¹³¹. Furthermore, another benefit of boc is that deprotection can be carried out in acidic conditions which is beneficial as cofactors are stable in acid and so no degradation should occur during later deprotection after creation of the sulfonium centre.

Hydroxylamine coupling of **2b** was achieved following a referenced procedure with N-boc-hydroxylamine in the presence of 1,8-diazabicyclo[5.4.0]undec-7-ene (DBU)¹⁴². Success of the reaction was monitored by TLC during the reaction and

a combination of NMR and MS after purification (Sup. Fig. S6.16-S1.34). The reaction typically went to completion allowing high yields post purification of up to 70%. For this reaction the product was purified by a combination of extraction and FCC before final analysis. Care had to be taken during extraction with thorough washing to ensure complete removal of DMF, which if present caused significant issues during FCC with poor impurity separation.

Hydrazide introduction of **2c** was achieved in a similar method as for **2b** but using 1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC) coupling with boc-hydrazide¹⁴³. Coupling was again confirmed by a combination of NMR and MS and the product was purified by extraction (Sup. Fig. S6.27-S1.46). Some difficulties were initially found with complete removal of the excess Boc-hydrazide used in coupling. However, complete removal was achieved by lowering the equivalents of tert-butyl carbazate added during reaction and also carrying out acidic washes during separation. These changes caused no significant impact on reaction efficiency and enabled yields of consistently above 60%.

2.2.3 Alcohol Activation

The final step in the synthesis of all precursors was alcohol activation. Typical leaving groups employed for alcohol activation in previous cofactors have included halogens and sulfonates¹⁴⁴. For this project a combination of chlorination, tosylation and bromination were all completed with varying degrees of success depending on the precursor.

2.2.3.1 Chlorination

The initial alcohol activation attempted was chlorination of **3a** using thionyl chloride in the presence of pyridine (Figure 2.10.A). Chlorination was initially chosen as the favoured method for alcohol activation due to its simplicity and high yields. Although it was expected there would be a compromise in its reactivity the ease of synthesis and long-term storage benefits outweighed this factor. The reaction was effective and could easily be monitored using a combination of IR and ^1H NMR (Figure 2.10). IR showed complete removal of the OH stretch at ca. 3300 cm^{-1} and the introduction of a new sharp band at 690 cm^{-1} for the C-Cl stretch. Whilst ^1H NMR showed removal of the OH peak at 5.0 ppm and a shift in the CH_2 adjacent to the reactive site from 4.0→4.4 ppm with a change in splitting. Combined these results highlighted what was expected from the functional group change and **4ai** was taken forward for the synthesis of cofactors **AdoHcy-6-N₃**, **AdoCys-6-NH₂/N₃**. However, we found the chloride showed limited reactivity with AdoHcy and unfortunately only proved reactive enough to couple to thioadenosine even with the use of catalysts like TBAI. As such, for the synthesis of **AdoHcy-6-N₃** it was necessary to explore alternative more reactive activating groups.

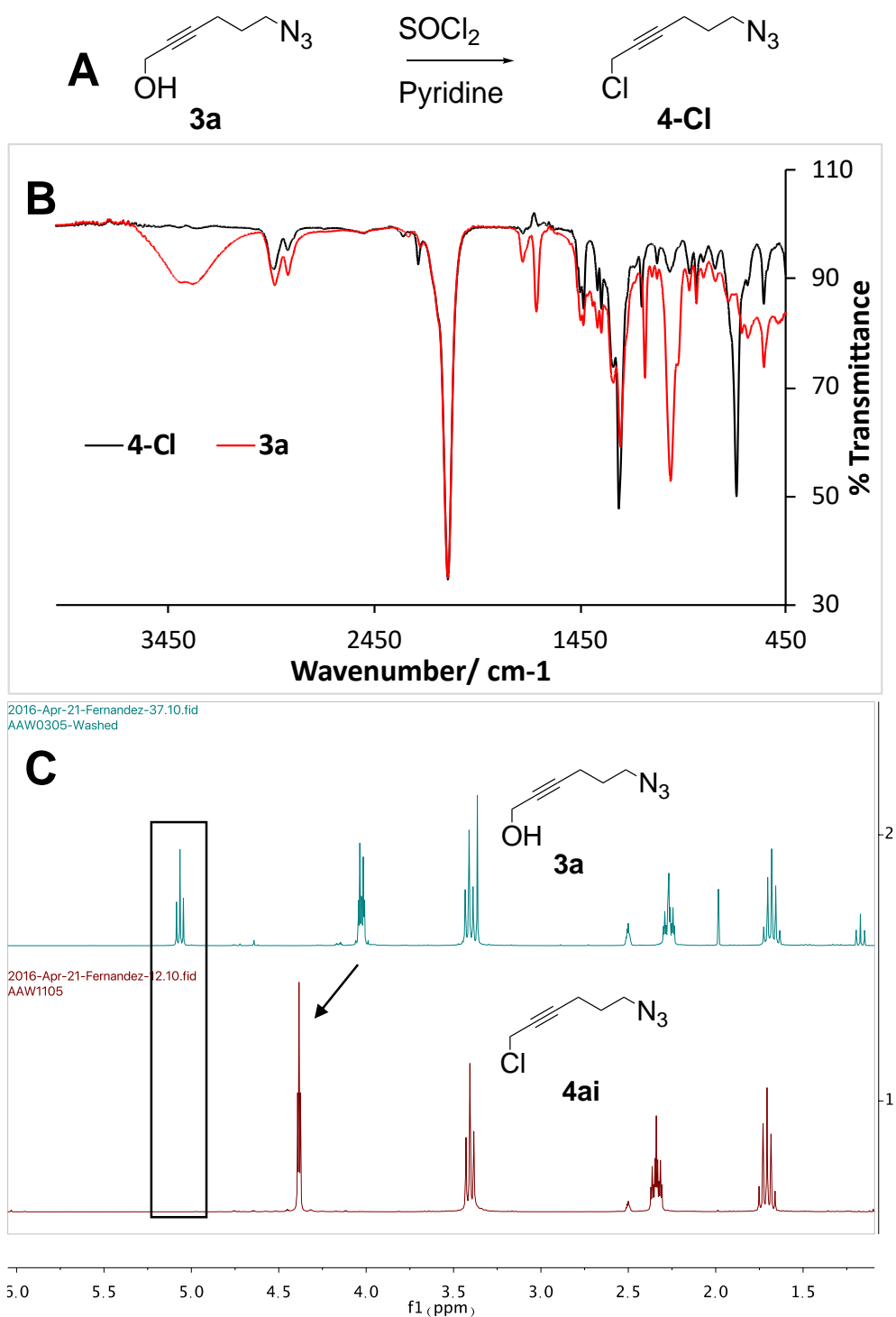


Figure 2.10 A) Linker **3a** was chlorinated with a combination of thionyl chloride and pyridine with stirring at room temperature. B) IR overlay of linker **3a** before and after chlorination. A complete removal of the O-H stretch at 3450 cm^{-1} can be seen as well as the introduction of a new peak at 690 cm^{-1} corresponding to the C-Cl. B) ^1H NMR overlay of **3a** (Top) and **4ai** (bottom). Loss of OH can be seen at 5.0 ppm as well as a shift ($4.0 \rightarrow 4.4\text{ ppm}$) and change in splitting ($\text{dt} \rightarrow \text{t}$) of the CH_2 adjacent.

2.2.3.2 Tosylation

Due to the unsuccessful coupling of **4ai** to AdoHcy, tosylation was explored for alcohol activation. Tosylate was the chosen sulfonate because it has greater reactivity than chloride but should remain relatively stable and easy to handle. The tosylate was formed by reaction of **3a** with *p*-toluenesulfonyl chloride (TsCl) in the presence of NaOH at 0 °C forming **4aii**. The reaction was then extracted and the crude product collected and analysed. Success of the reaction was observed using ¹H NMR through the introduction of two new doublets in the aromatic region (7.5/7.8 ppm) and a singlet (2.4 ppm) in a ratio of 2:2:3, matching the integrations of the alkyl chain (Figure 2.11.A). After extraction, column chromatography was carried out to remove excess TsCl and other impurities. However, this was unsuccessful as although good compound separation could be achieved by TLC complete removal of excess tosylate was not possible as it continuously bled through the column. As such the crude reaction product was tested directly in coupling with AdoHcy, which was successful and so further purification methods were not explored. Coupling was also tested with thioadenosine to form **AdoCys-6-NH₂/N₃** in an effort to minimise the compounds necessary for synthesis. But, this was unsuccessful as again purification after reaction was difficult, leading to thioadenosine coupling being continued with **4ai**.

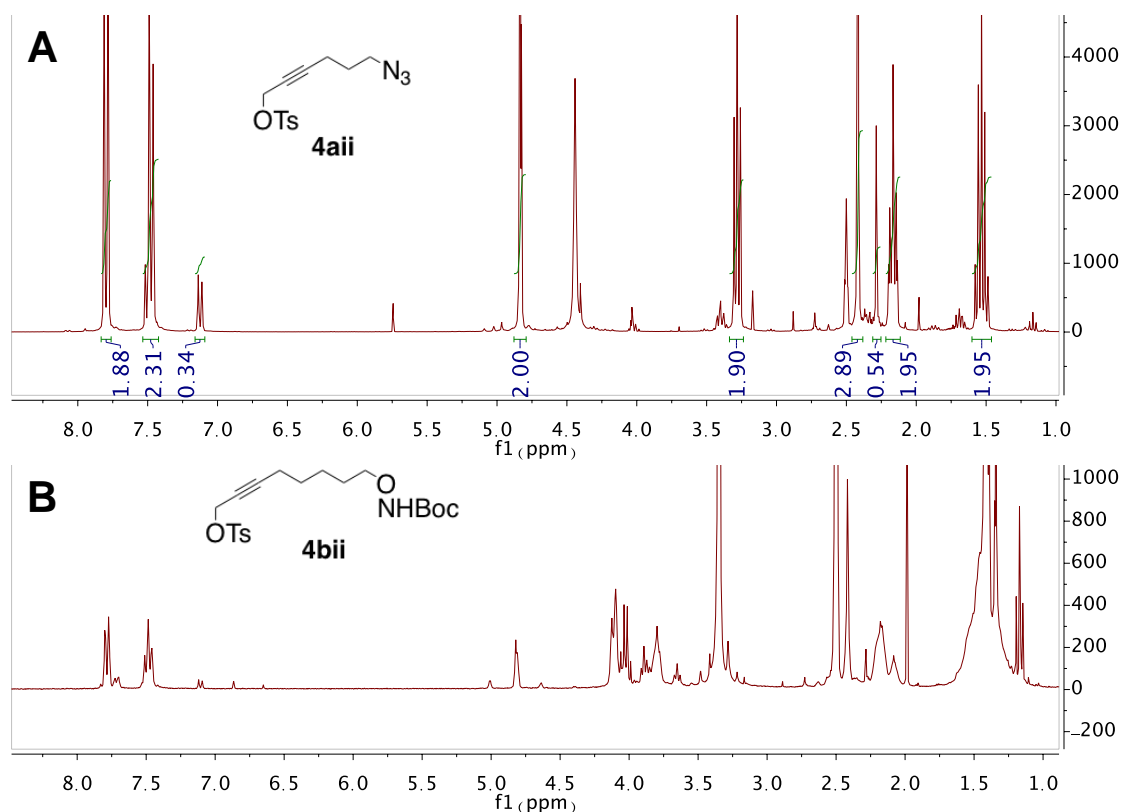


Figure 2.11 A) ¹H NMR collected after tosylation and extraction of **3a**. The ratio of product to remaining tosylate was measured by comparing integrations of the aromatic peaks at 7.2 ppm and 7.8 ppm: $1.88/(1.88+0.34) = 0.85$. B) ¹H NMR of the failed tosylation of **3b**.

Following the successful synthesis of **4aii** and its coupling to AdoHcy tosylation was also initially chosen as the activating group for precursor **3b**. However, this reaction proved to be unsuccessful due to issues with low yields (less than 10% after extraction) and purity (Figure 2.11.B). In an attempt to improve the reaction, the reaction solvent was changed as solubility was an issue due to the formation of precipitates which could not be removed during extraction. For this, a combination of solvents were tried with increasing polarity including ethyl acetate and diethyl ether but no improvements in yields were seen. As such, bromination was explored as an alternative activating group for this linker.

2.2.3.3 Bromination

Bromination was highlighted as a potential alternative to tosylation of **3b** as it had given very high yields previously in the formation of **1b** and was known to have similar reactivity to the tosylate for future couplings. For this the Appel reaction was again used but with a modified reaction work-up as filtration with cold hexane was not possible. Instead after reaction the crude product was filtered with a mixture of hexane:ethyl acetate (8:2) to remove the bulk of organophosphorous compounds followed by FCC for further purification.

After purification, confirmation of bromination was achieved using a combination of MS and ^1H NMR. MS highlighted the molecular weight expected along with the clear isotope effect for bromides with a 1:1 ratio between isotopes ^{78}Br and ^{80}Br (Figure 2.12). Whilst ^1H NMR showed a complete loss of the OH peak at 5.0 ppm and a downfield shift of the CH_2 adjacent to the reactive site by 0.2 ppm (4.0→4.2 ppm) in an identical way to the bromination of **1b** (Figure 2.5.B). Once successful, bromination was then also used as the activating group for precursor **3c** as it likely would not have been possible to tosylate (Sup. Fig. S6.31-S1.49).

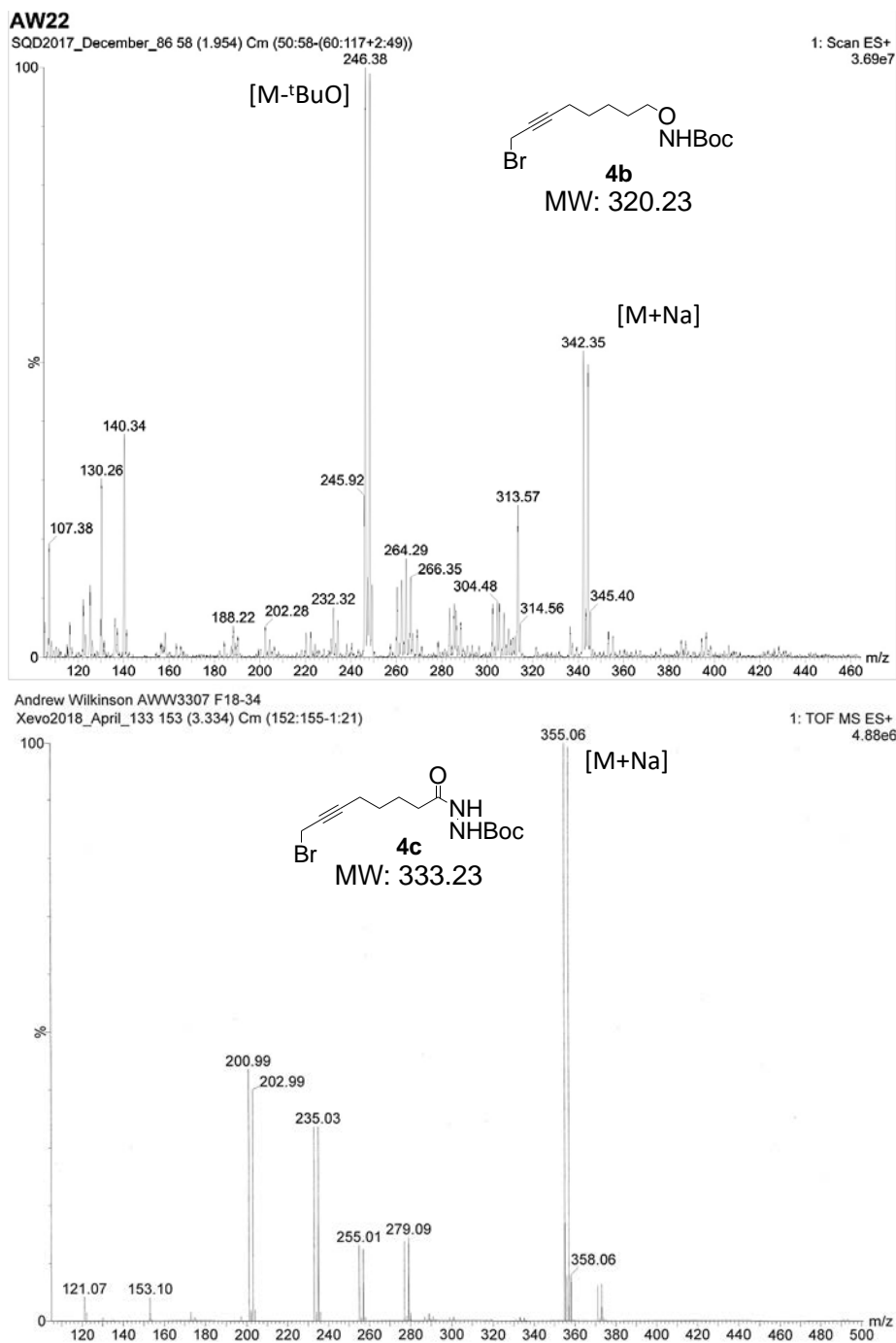


Figure 2.12 MS data collected after synthesis and purification of linkers **4b/ciii**. Both compounds can be identified by their molecular weight in the $[M+Na]$ form with a clear bromide effect with a 1:1 ratio between 342/344 and 355/357 respectively.

After alcohol activation linkers were then taken on to be used in couplings to either thioadenosine or AdoHcy to complete the cofactor synthesis. The first attempts at synthesis of cofactors were achieved using linker **4a** where it was

found that synthesis of all cofactors could be achieved. However, it was subsequently found that the cysteine cofactors **AdoCys-6-NH₂/N₃** showed limited labelling efficiency with methyltransferases, hence cysteine versions of **AdoHcy-8-Ox/Hyd** were not explored.

2.3 Cysteine cofactor synthesis

The first cofactors attempted using linker **4a** were **AdoCys-6-N₃/NH₂**. Here it was first necessary to create the reactive intermediate thioadenosine from isopropyl adenosine, followed by coupling of the linker and then subsequent amino acid introduction (Figure 2.13) ¹⁴⁵. Initially this was intended to be the synthetic route adopted for all cofactors including the hydrazide and hydroxylamine cofactors. However, as mentioned above, after the initial synthesis of **AdoCys-6-N₃/NH₂**, these cofactor analogues displayed poor effectiveness as cofactors compared to their homocysteine equivalent **AdoHcy-6-N₃**. This is discussed further in the following chapter.

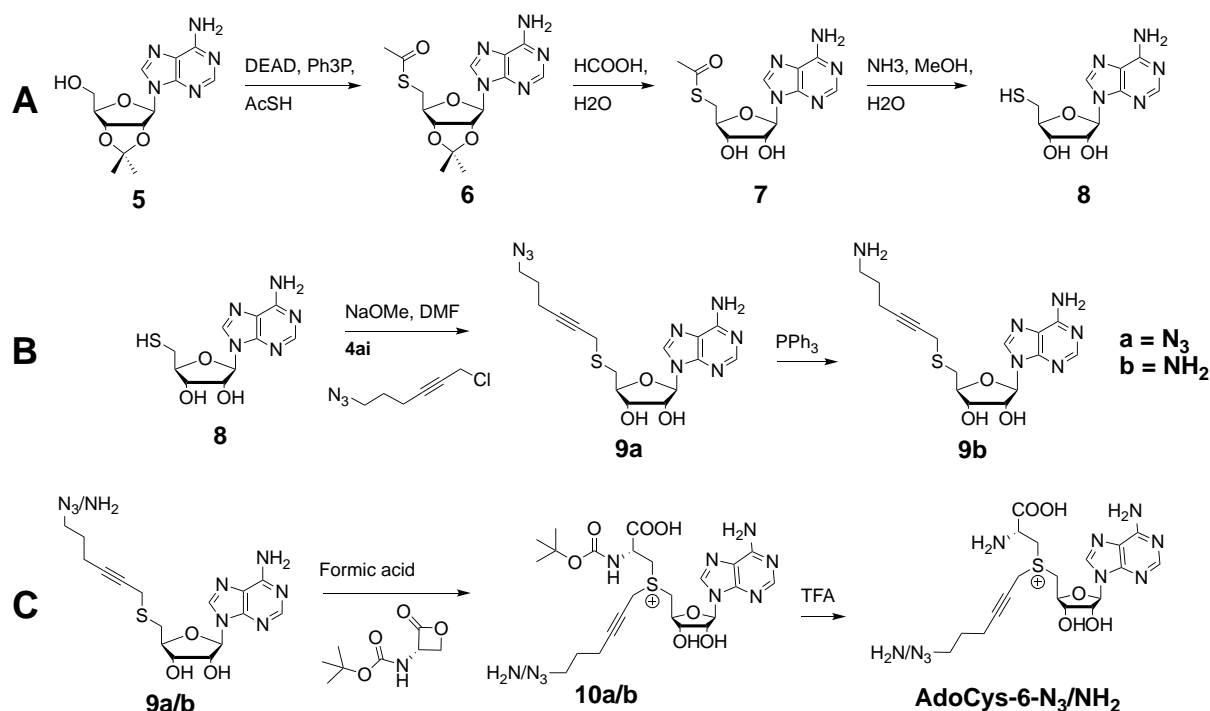


Figure 2.13 Synthesis of cysteine-based cofactors **AdoCys-6-N₃/NH₂** can be broken down in the 3 stages:

A) synthesis of thioadenosine, B) linker coupling (and reduction if necessary), C) amino acid coupling and deprotection.

2.3.1 Thioadenosine synthesis

Synthesis of thioadenosine was achieved by following the referenced procedure¹⁴⁵. This first required a Mitsunobu reaction to allow for the introduction of the thiol in its protected form¹⁴⁶. The reaction was successful and near quantitative yields of **6** were achieved. The success of the reaction could then be monitored by comparison with the referenced spectra (Sup. Fig. S6.37-S1.55)¹⁴⁵.

Once the desired thiol had been inserted synthesis of **8** was then completed by a two stage deprotection. First, removal of the acetonide group using formic acid, followed by thiol deprotection with a saturated ammonia solution. Care was taken with thiol deprotection ensuring a complete absence of oxygen. However, it was found the referenced procedure could be adapted to avoid the use of ammonia gas, by using commercially available saturated solutions. This was preferable as

using ammonia gas is dangerous and expensive. The procedure could further be simplified by replacing the solvent freeze/thaw cycles with bubbling argon through the solution to degas. The structure of **8** was then confirmed using a combination of NMR and MS. All of which showed very high purity and matched with the reference (Figure 2.14). Due to the expected instability of thioadenosine, the final deprotection was completed individually prior to each linker coupling to help maintain high yields and purity.

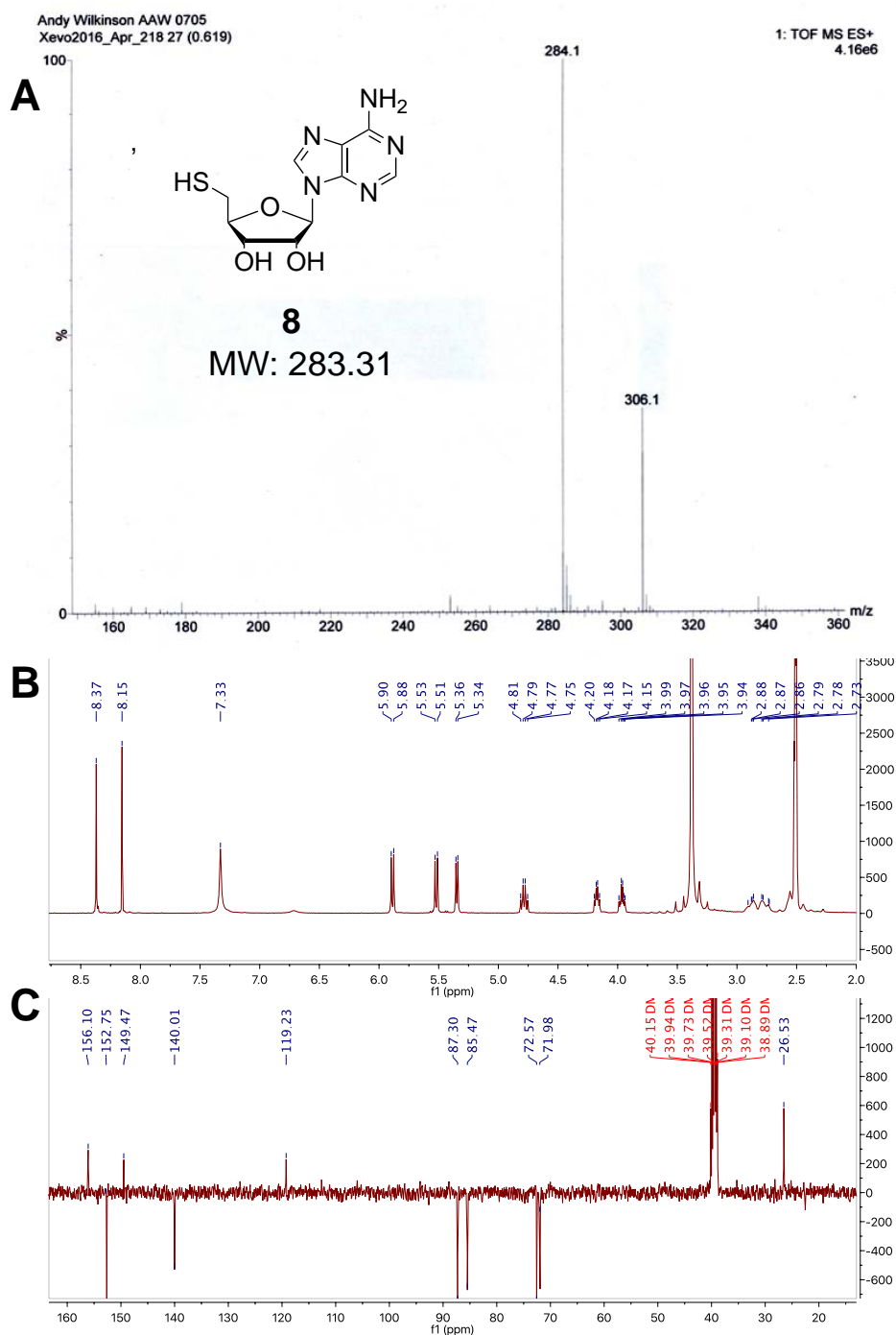


Figure 2.14 Characterisation methods used for **8**. A) MS showing $[M]^+$ 284.1 and $[M+Na]^+$ 306.1, B) 1H NMR C) Pendant ^{13}C NMR.

2.3.2 Linker coupling

After the synthesis of **8** coupling of the activated linker **4a** was achieved through heating in the presence of a strong base in DMF yielding **9a** (Figure 2.15.A). Nucleophilic substitution was successful with both the tosylate and chloride form

of **4a** but the chloride was used preferentially due to problems with purification when using **4aii**. When using **4aii** for coupling, after purification there still remained large quantities of excess tosylate which can clearly be seen in the ^1H NMR at 7.5, 7.0 and 2.2 ppm (Figure 2.15.B). This was the same issue that had occurred previously during the tosylation of **4aii** where remaining tosylate bled through the column during purification. As such, the chloride was used in all future couplings due to the purity of the product that could be collected. Confirmation of the final product then came from cross referencing with the patented procedure for this compound¹⁴⁷. Introduction of the linker CH_2 peaks could be seen (1.6, 2.3 ppm) as well as a shift in the diastereotopic protons from 2.8→2.9 ppm (Figure 2.15.C). The CH_2 closest to the sulphur centre from the linker also had a large upfield shift, however this overlapped with water and the final CH_2 peak at 3.3 ppm and was difficult to monitor directly.

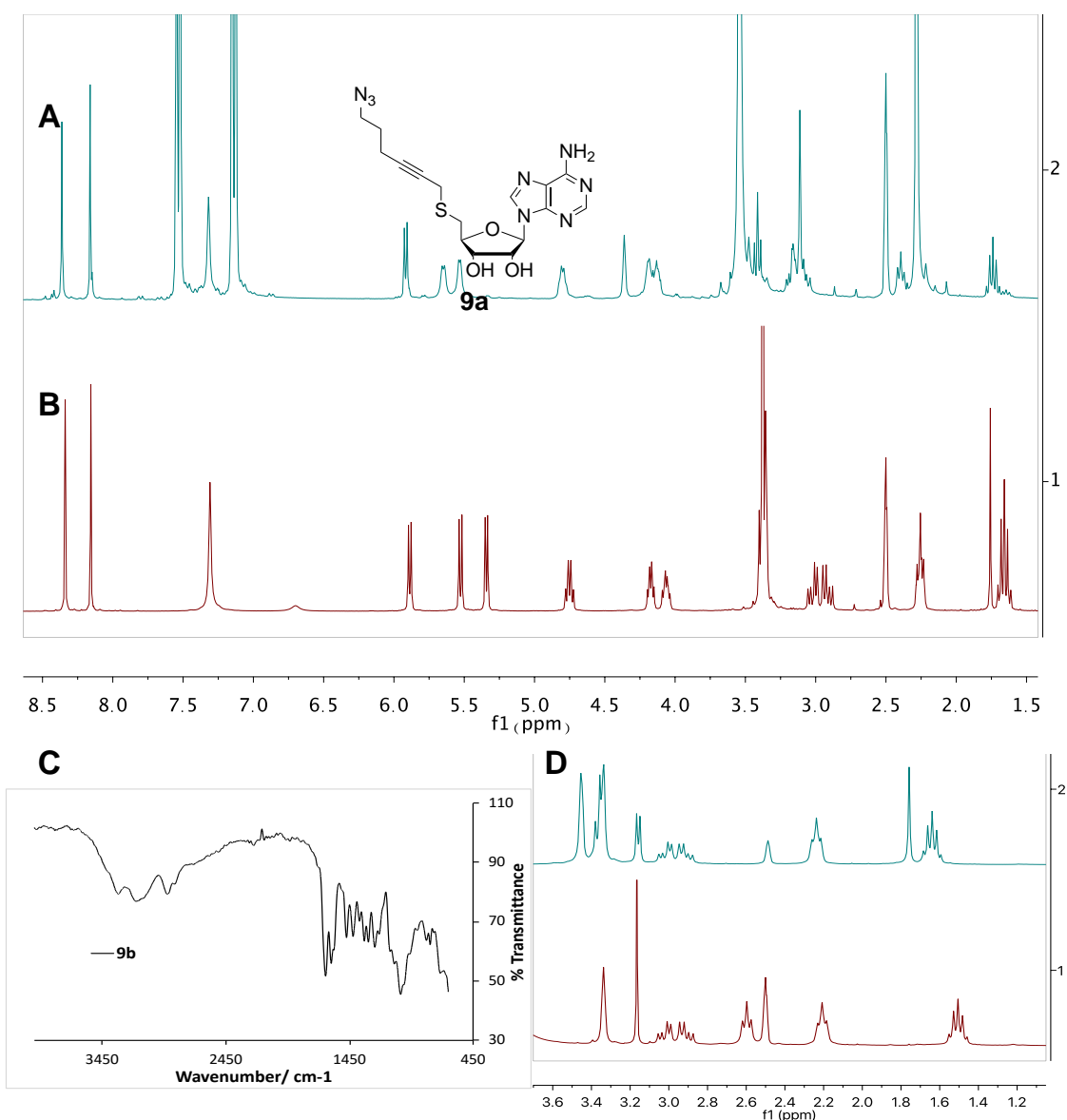


Figure 2.15 A) ^1H NMR of **9a** after coupling and purification using **4aii**. B) ^1H NMR of **9a** after coupling and purification using **4ai**. C) IR of **9b** after reduction, where complete removal of azide at ca. 2100 cm^{-1} can be seen. D) ^1H NMR overlay of **9a** (top) vs. **9b** (bottom). A clear shift of the CH_2 can be seen from 3.3 \rightarrow 2.6 ppm.

After introduction of the linker, **9a** could either be used directly in cysteine coupling to form **AdoCys-6-N₃** or undergo a Staudinger reduction to form the cofactor **AdoCys-6-NH₂**¹⁴⁸. As with all organophosphorous chemistry the largest difficulties in the reduction of **9b** were in purification. Here, the best method for purification was found to be silica pad chromatography. This saw the crude

product first dissolved in the minimum amount of DCM:MeOH (7:3) followed by loading on to a small quantity of silica. Due to the acidic nature of silica the product then bound to the column, allowing all impurities to be washed away. After by-product removal the product could then be eluted through basic substitution on the silica with ammonia. This was achieved by adding ammonia to the solvents at a 1% concentration using saturated methanol. Due to the basic nature of ammonia it then displaced the product from the column and allowed its elution. Once eluted the ammonia was then removed by bubbling argon through the solution, followed by rotary evaporation yielding **9b**.

Success of the reduction could then be monitored through a combination of IR and NMR (Figure 2.15.C,D). IR showed a complete removal of the azide stretch that had previously been seen at ca. 2100 cm^{-1} . Whilst ^1H NMR showed a clear upfield shift in the triplet CH_2 adjacent to the reactive site, from 3.3 \rightarrow 2.6 ppm. After confirmation of the structure, **9b** was then used directly in the synthesis of **AdoCys-6-NH₂**.

2.3.3 Amino acid coupling and cofactor synthesis

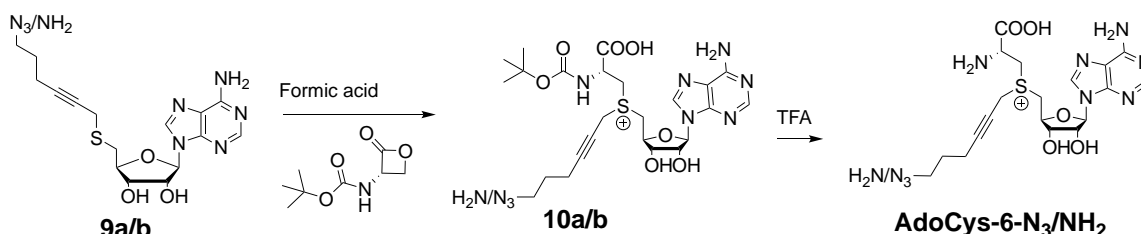


Figure 2.16 Synthetic route employed for cysteine introduction to cofactors **AdoCys-6-N₃/NH₂**

The final stage towards cofactors **AdoCys-6-N₃/NH₂** was achieved by coupling of **9a/b** to Boc-L-serine- β -lactone (BSL) followed by deprotection (Figure 2.16).

The reaction was carried out in a similar manner to the reported route for previous AdoMet analogues, i.e. through coupling in the presence of acid with a large excess of coupling reagent⁹¹. For both the azide and amine cofactors an identical procedure was applied which saw coupling overnight followed by removal of the boc protecting group and then final purification by prep-HPLC. Success of the reaction was then monitored by MS analysis of the peaks collected during prep-HPLC (Figure 2.17). For both cofactors percentage conversions of over 70% were achieved, assuming no impurities contributed to cofactor degradation. As expected there were a clear pair of peaks for each cofactor, which we suspected to be the two diastereoisomers that can be formed due to the formation of the new chiral sulfonium centre¹³¹. Separation of the individual fractions formed during HPLC was attempted so that in later studies it would be possible to monitor their individual effectiveness; but this proved challenging as the retention times were so similar, especially with **AdoCys-6-NH₂**⁹³. As such, the samples collected were likely a mixture of isomers albeit with a vast majority of one.

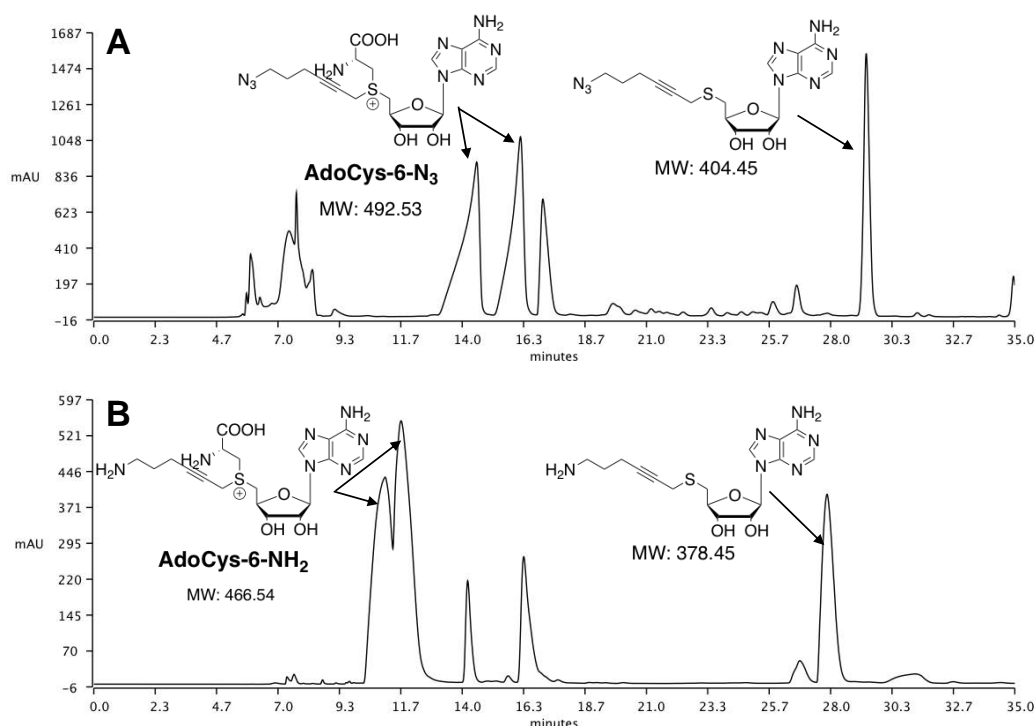


Figure 2.17 prep-HPLC traces of cysteine cofactor coupling after deprotection. A) **AdoCys-6-N₃**, Cofactor peaks identified by MS as T14.2/ T16.3 mins. Remaining starting material T29.3. Conversion = 74%. B) **AdoCys-6-NH₂**, Cofactor peaks at T11/ T11.8 mins. Attempted to separate as effectively as possible but some cross-over likely. Remaining starting material T28.0 conversion = 78%.

After identification, each fraction collected from the prep-HPLC was dried and resuspended in a solution of 0.1% formic acid for permanent storage and use with methyltransferases. 0.1% formic acid was the chosen storage buffer as the acidity helped to stabilise the cofactors during storage, but isn't too acidic for future use with enzymes¹⁴⁹. The concentration of each sample of cofactor was then calculated using UV absorbance at the λ_{max} of adenosine ($\epsilon_{260}=15,400 \text{ M}^{-1}\text{cm}^{-1}$) and stored at working cofactor concentrations of 15 mM for labelling studies. After synthesis each cofactor analogue was tested in combination with M.TaqI and their labelling efficiencies were compared to the homocysteine equivalent **AdoHcy-6-N₃**, which will be discussed in detail in chapter 3.

2.4 Homocysteine Cofactor Synthesis

The second category of cofactors to be synthesised were the doubly-activated analogues containing homocysteine, where each cofactor was synthesised following an adapted version of the original protocol (Figure 2.18.A) ⁹². This saw coupling of activated linkers in acid overnight, followed by extraction with diethyl ether and final purification by reversed phase HPLC. Typically though, this reaction has low yields due to low recovery and poor conversion of AdoHcy. In an attempt to improve coupling efficiencies and maximise yields we tested an array of different conditions using the linker **4biii** (Figure 2.18.B).

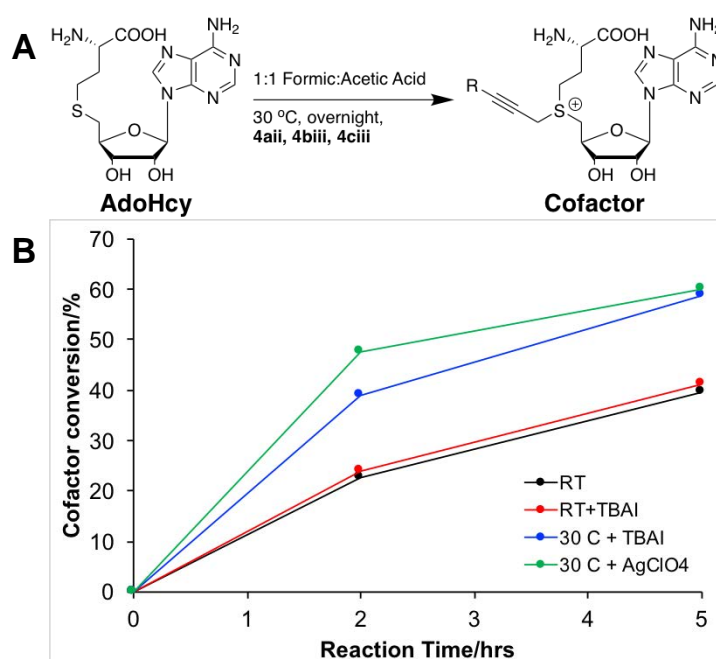


Figure 2.18 A) General scheme for the optimised coupling of homocysteine cofactors. B) Optimisation experiment of cofactor coupling when using linker **4b**. 4 conditions were tested: room temperature (black), room temp. + catalytic TBAI (red), TBAI at 30 °C (Blue) and AgClO₄ at 30 °C. Cofactor conversion was measured by integration of the three associated cofactor peaks seen during HPLC; AdoHcy, isomer 1 and isomer 2. Overall conversion was combined as the value for both isomers collected compared to remaining AdoHcy.

Tetrabutylammonium iodide (TBAI) was initially chosen as a catalyst for this reaction as it is commonly used with alkyl halides due to its ability to undergo halogen exchange. *In Situ* generation of the alky-iodide increases the electrophilicity of the alkyl side chain which should in turn increase the success of nucleophilic attack. Similarly, gentle heating was also tested in an attempt to improve the reaction. One worry prior to this test was that any benefit caused by heating could cause decomposition of the cofactor. However, this was not an issue as the reaction was run in acidic media and cofactor integrity was found to be maintained during HPLC of each test reaction. Through the inclusion of a catalyst and heating it was found the presence of TBAI did not significantly impact on the conversion of the reaction, but heating did. By increasing the temperature to just 30 °C a two-fold increase in coupling efficiency to nearly 60% was seen.

Further attempts to improve coupling efficiency were then made by testing the alternative catalyst AgClO₄. This catalyst was chosen as it has been used previously in the literature for AdoHcy coupling in the synthesis of doubly-activated analogues which we hoped would help to increase coupling efficiencies further^{106,150,151}. Silver salts can aid reaction efficiency by working as both a lewis-acid activator and a halide anion quencher. However, when testing the effectiveness of replacing TBAI with AgClO₄ no significant improvements were seen. Ultimately, neither TBAI or AgClO₄ significantly improved the yields of conversion and it was later found they could be removed entirely from coupling without reducing efficiency. Increased temperatures were also explored above 30 °C to see if this improved conversion further but a plateau in efficiency was found and problems with uncontrolled thermal Boc deprotection started to occur.

Once optimised the same protocol was then adopted and used for all linkers to form the three cofactors of interest **AdoHcy-6-N₃**, **AdoHcy-8-Ox/Hyd**. Here it was found similar coupling efficiencies could be achieved for all cofactors and no significant changes were found in the reactivity of tosylated or brominated linkers as expected. After coupling the cofactors were then extracted and deprotected where necessary before final purification by HPLC (Figure 2.20). Importantly, through improvement of the coupling efficiency during reaction it was possible to obtain reaction yields similar to those achieved with the cysteine-based cofactors.

2.4.1 Cofactor Deprotection:

After coupling, to obtain cofactors **AdoHcy-8-Ox/Hyd** it was necessary to first carry out a boc deprotection. For deprotection the extracted reaction mixtures were first freeze dried and then redissolved in neat trifluoroacetic acid (TFA) to cause boc deprotection. The crude reaction mixture was then stirred at room temperature for two hours before removal of TFA by blowing Argon over the mixture thoroughly. The crude product was then collected as a yellow oil and diluted in water before injection in to the prep-HPLC.

2.4.2 Cofactor purification

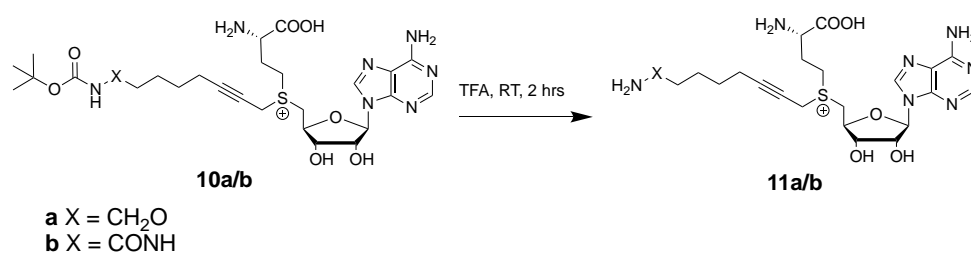


Figure 2.19 Reaction scheme for deprotection of **10a/b** to form the oxime and hydrazone cofactors. The crude products from coupling were stirred in TFA for 2hrs dried and taken forward to be purified.

Purification of all cofactors was carried out by reversed phase prep-HPLC using a C18 column and either an ammonium formate or acetate buffer. The main target of purification was to remove any unused AdoHcy from the reaction mixture as this serves as an inhibitor for methyltransferase enzymes and so would lower the potency of the cofactor formed. Purification further helps to remove any excess linker that may remain from coupling not removed during extraction. During purification it was possible to purify all cofactors formed and as seen previously for the cysteine cofactors, both diastereoisomers formed could be separated in a roughly 1:1 ratio (Figure 2.20)¹³¹. For cofactors **AdoHcy-8-Ox/Hyd** purification was originally carried out twice, once after coupling and then again after deprotection. However, we found that this significantly compromised overall yields and was unnecessary as during deprotection no competing reactions occurred. In addition, homocysteine coupling also provided fewer impurities during prep-HPLC compared to the cysteine cofactors which likely has a positive impact on overall yields.

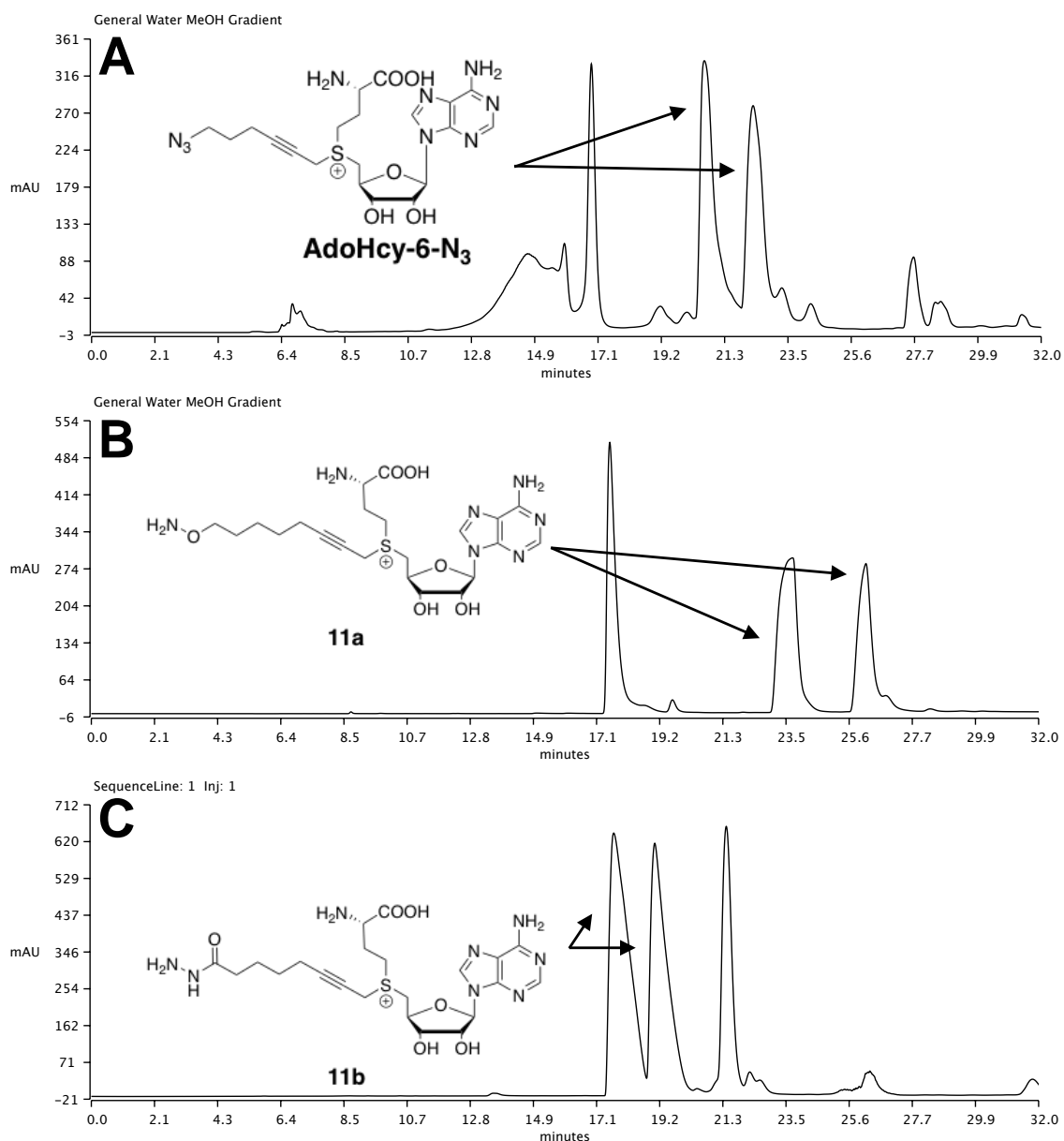


Figure 2.20 prep-HPLC traces collected during purification after coupling and deprotection (if needed). Identification of peaks was confirmed using MS. A) purification of cofactor AdoHcy-6-N₃. 20 mM Ammonium formate buffer. AdoHcy = 17.1 mins, cofactor = 21.1, 23.0 mins. B) purification of **11a**. 20 mM Ammonium formate buffer. AdoHcy = 17.2 mins, cofactor = 23.6, 26.0 mins. C) purification of **11b**. 20 mM Ammonium acetate buffer. Cofactor = 17.4 mins, 19.3 mins, AdoHcy = 21.4 mins.

As with the cysteine cofactors, following purification each fraction was collected and then analysed by MS for identification (Sup. Fig. S2.5-S1.10). Percentage conversion of the original coupling was then measured by integrating the area under each of the peaks. As stated previously, coupling efficiencies were

typically between 60-80% (Figure 2.20). Each fraction was then dried by low temperature rotary evaporation followed by resuspension in 0.1% acetic acid. The concentration of each cofactor was then measured by UV absorbance at 260 nm and the cofactor was again stored at working concentrations of 15 mM. Once stored the cofactor could then be further analysed by analytical HPLC to monitor for degradation and purity. Yields of cofactors could then be back calculated from the concentration and volume collected and were found to be between 7-10% per suspected diastereoisomer for all cofactors.

2.4.3 Cofactor analysis and stability

2.4.3.1 *AdoHcy-6-N₃*

Due to the known instability of AdoMet analogues, after purification it was necessary to further test each individual isomer collected for degradation that may have occurred during work-up and their stability in the enzymatic buffers that would be used for labelling. For **AdoHcy-6-N₃** it was found the product could be collected with good purity. However, as expected degradation from work-up was seen which can be quantified by comparing HPLC peak areas (isomer 1 = 61%, isomer 2 = 46%) (Figure 2.21). As can be seen from the analytical HPLC near complete isolation of isomer one could be achieved. However, isomer two could not be completely separated and some of isomer one remained. All peaks were identified through MS and the key degradation by-products identified include those seen previously for this compound; loss of the amino acid (34 mins, 15%), loss of adenine (5%, 10.2 mins) and loss of linker (2%, 12 mins). Further by-products between 18-19 mins and 31.2 mins can also be seen and were found to have an increase in molecular weight of 18 on either **AdoHcy-6-N₃** or **Ado-6-**

N₃. When comparing with the literature degradation of this compound it was concluded that this is likely due to hydration of the alkyne within the linker⁹⁴. Once analysed the cofactor stability was then tested before moving onto enzymatic studies.

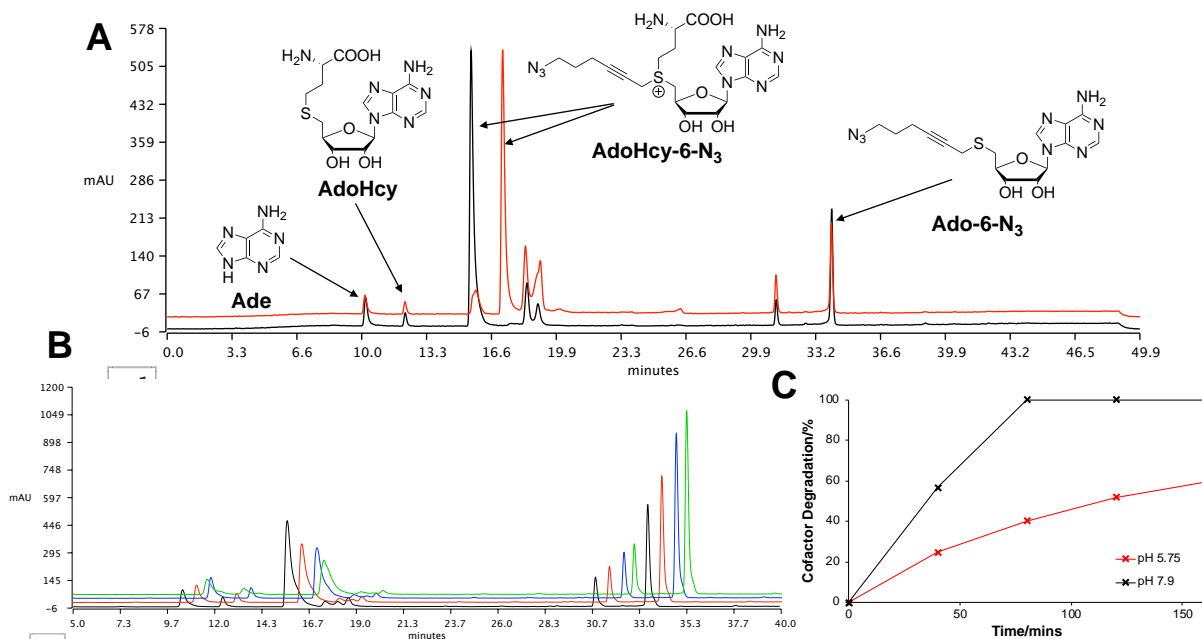


Figure 2.21 A) Analytical HPLC analysis of each isomer of **AdoHcy-6-N₃** collected during prep-HPLC. Identification of each peak was completed by MS. Some degradation after final collection of the product can be seen. B) An example of the analytical HPLC traces collected over time during degradation studies. Loss of amino acid from the cofactor analogue can be seen as the key pathway for degradation through an increase in intensity of the peak at 34 mins. All fractions isolated were subsequently analysed by MS to confirm structures. C) A comparison of the percentage degradation seen when varying the pH at 50 °C. Degradation was measured by integration of all peaks associated with cofactor breakdown, including; adenine, AdoHcy, Ado-6-N₃ and Ado-6-N₃+H₂O.

After analysis, an investigation into the stability of **AdoHcy-6-N₃** was carried out using analytical HPLC. The cofactor was tested by heating in the presence of the enzymatic buffers that would be used for future DNA labelling tests. Stability was initially tested at 50 °C as this was the working temperature for the first methyltransferase of interest M.TaqI. Two buffers were chosen, which were New

England Biolabs (NEB) cutsmart (pH 7.9, recommend for use with M.TaqI) and a more acidic MES-based cutsmart buffer (pH 5.75). Degradation was then monitored at 40-minute intervals where, as expected, the dominant pathway for degradation was through loss of the amino acid (Figure 2.21.B). Interestingly, no interconversion was seen from the suspected diastereoisomers collected during heating. This likely highlights that by modifying the R group attached to the sulfonium centre there has been an increase in the energy of interconversion compared to native AdoMet¹⁴⁹. This change is likely due to an increase the steric interactions around the sulfonium centre¹⁵². A lack of interconversion between isomers during degradation studies was positive though as it meant after isolation each isomer would remain in either the *S,S* (active) or *R,S* (inactive) form during enzymatic labelling which is another common degradation pathway for AdoMet.

When comparing the percentage degradation over time of the two buffers a clear improvement in stability was seen in the more acidic buffer with over a 3-fold decrease in cofactor degradation over time (Figure 2.21.C). This was expected, as the predominant pathway for degradation (amino acid removal) is known to be pH dependant due to the differing nucleophilicity of carboxyl groups vs. their carboxylate anions⁹⁸. This result highlighted that future labelling reactions may need to be optimised to gain maximum labelling efficiency as the cofactor and enzyme transalkylate most effectively in alternate buffers. It has been shown that cofactors are more stable in acidic solution, however the commercial buffer M.TaqI is slightly alkaline with a pH of 7.9. Further studies were also carried out on the second isomer where no significant differences in degradation patterns were seen. No further temperatures were studied for degradation as for all future

enzymes of interest 50 °C was the highest labelling temperature to be used. After analysis, **AdoHcy-6-N₃** was then taken forward to labelling studies with methyltransferases to test its labelling efficiency and to compare against the cysteine cofactors.

2.4.3.2 *AdoHcy-8-Ox/Hyd*

Following complete analysis of **AdoHcy-6-N₃** the same process was repeated for **11a**. Here it was found cofactor degradation became a much larger issue due to the increased reactivity of the free hydroxylamine. In addition to the predominant degradation pathways that were seen for **AdoHcy-6-N₃** (loss of amino acid ca. 15% 32.2 mins, adenine ca. 18% 12.5 mins), **11a** also underwent degradation following a new pathway through the introduction of a peak at 30.0 mins (Figure 2.22.C, 30%). Initially it was thought that this was likely due to hydration of the alkyne as seen previously but to a larger extent due to either a change in the functional group or alkyl chain length. However, MS analysis later showed that for both suspected isomers a m/z ratio equal to the cofactor was seen. As the molecular weight of the by-product and cofactor are identical it is likely that this degradation was caused rather by a rearrangement than decomposition. The most likely pathway for such, was transfer of the amino acid from the sulfonium centre to the hydroxylamine (Figure 2.22.B). This was proposed as the method for rearrangement as the hydroxylamine can serve as a better nucleophile than that of the carboxylic acid, which is accepted as the main method for amino acid removal (Figure 2.22.A)⁹⁷. This conclusion is also supported by the fact that this by-product shows a loss in chirality, as both isomers show the impurity with the same retention time indicating loss of the sulfonium centre. Also, during later HPLC studies with acetone no reaction was seen, indicating the likely

change/removal of the hydroxylamine functionality which is consistent with this conclusion (Figure 2.23.B).

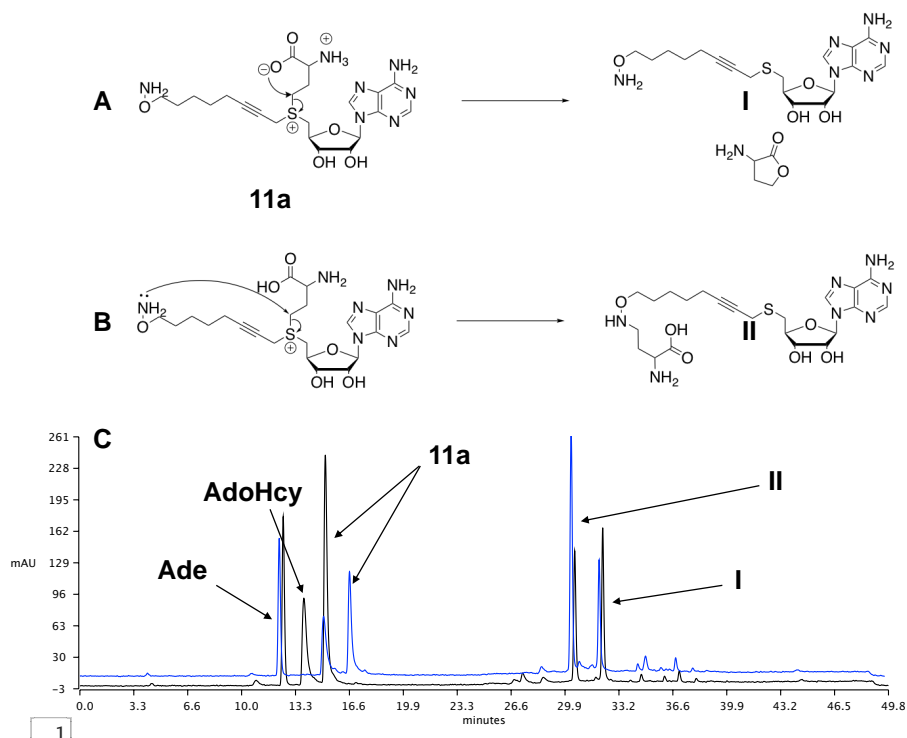


Figure 2.22 A) traditional degradation pathway proposed from previous studies for the loss of the amino acid from AdoMet analogues. B) A proposed route to explain rearrangement of the cofactor through attack on the amino acid. C) overlay of analytical HPLC collected after storage of each isomer collected for **11a**. Degradation can be seen through the formation of impurities Ade, AdoHcy, **I** and **II**.

Further to the issues found above with increased degradation of **11a**, problems were also found with this cofactor due to the reactivity of the hydroxylamine. During work-up great care had to be taken to remove any trace carbonyls as these could cause issues with cofactor purity. For example, during freeze drying a dry-ice/acetone bath was typically used, but acetone readily reacted with the free hydroxylamine to form a modified cofactor with a molecular weight of 564 g mol⁻¹. This was identified using a combination of MS and HPLC where the increase of 40 g mol⁻¹ from what was expected was confirmed as being due to the formation of the acetone-oxime Schiff base (Figure 2.23). Tandem MS

showed degradation of both the free hydroxylamine (524 g mol^{-1}) and acetone-oxime (564 g mol^{-1}) occurred following identical pathways through loss of the amino acid or loss of adenine, indicating any modification to be on the linker. Whilst HPLC showed addition of acetone to **11a** caused complete removal of free hydroxylamine (15.7 mins) along with an increased intensity of the peak corresponding to the impurity at 27.1 mins, which was confirmed by MS as 564 g mol^{-1} (Figure 2.23.C). Once identified, it was straightforward to remove this impurity by changing the freezing medium. However once one impurity was removed a new degradation pathway was identified. For example, after the removal of acetone, free hydroxylamine was also found to react with formate with a molecular weight of 552 g mol^{-1} (Figure 2.23.B). As such, to avoid the possibility of impurity formation we chose to cap the free hydroxylamine.

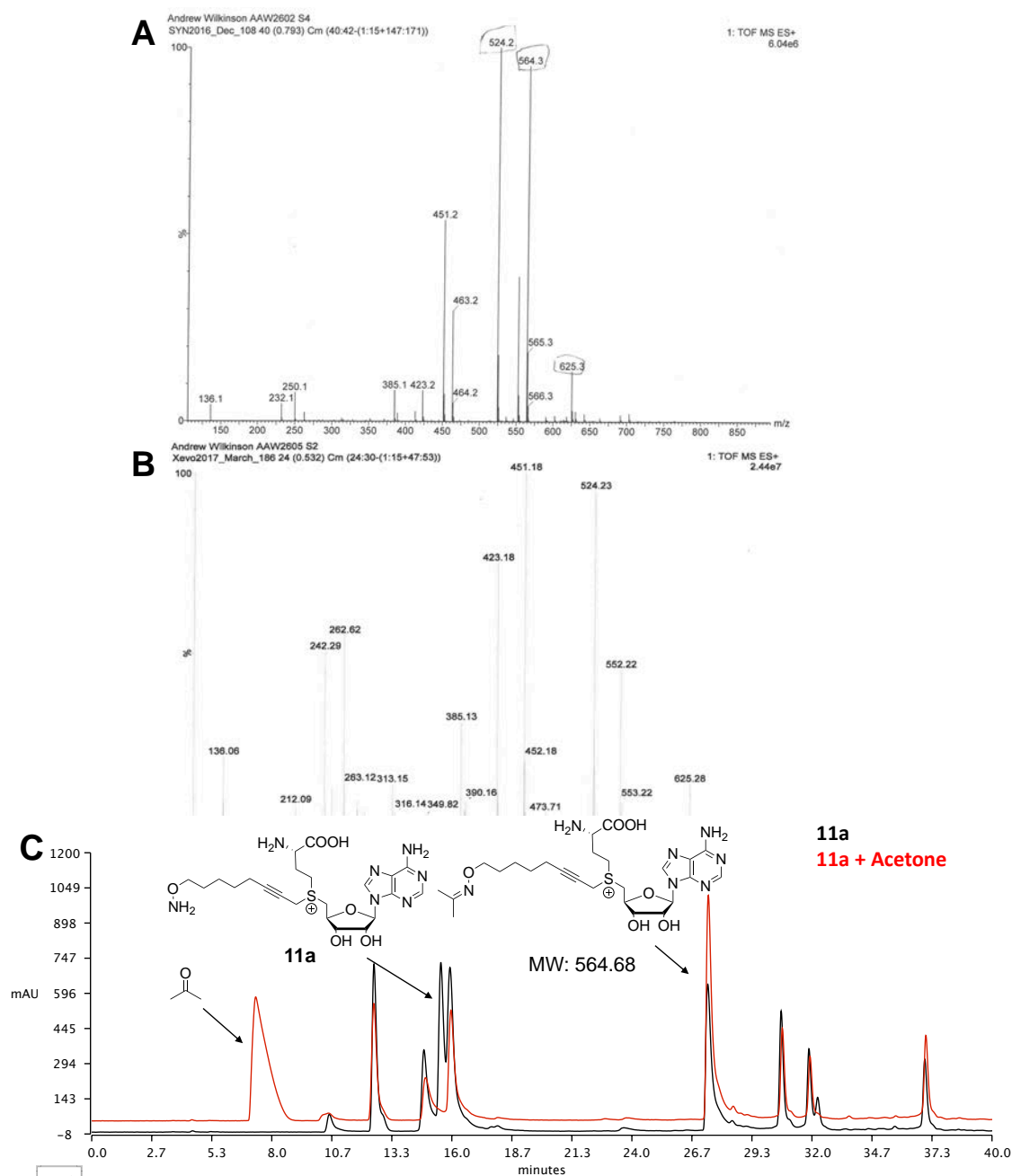


Figure 2.23 A) MS of **11a** after collection from freeze drying. Two clear peaks can be seen: 524 = **11a**, 564 = **11a**+Acetone. HRMS identified that the addition was due to the introduction of C_3H_4 which agrees with the introduction of acetone. B) MS sample of **11a** after removal of acetone, problems were found with a peak at 552 (+28) due to the addition of CO which was attributed to formate. C) Black = analytical HPLC trace of sample from A of **11a**. Red = **11a** + acetone. Complete removal of the cofactor at 15.6 mins is seen and an increase in intensity of the peak at 27.1 mins which was collected and confirmed to have a molecular weight of 564.

2.4.4 Aldehyde capping of Schiff base cofactors

In an effort to minimise degradation of the hydroxylamine/hydrazide cofactors and ensure the functionalisation state of the linker introduced during future labelling, it was decided to 'cap' the amine cofactors after deprotection. For capping, a bifunctional PEG aldehyde was chosen as it would maintain the reversible Schiff base desired, whilst also allowing for the introduction of a more stable click reactive group like an azide. It was hoped this would allow the continued development of two comparable reversible cofactor analogues whilst also providing a pathway that produced minimal impurities. In order to reduce the amount of new chemistry that had to be developed and optimised it was decided the best opportunity for capping was directly after purification. An alternative approach could have been through replacement of the PEG aldehyde for the Boc group on the original linker. However, this was not explored as during coupling there could have been unwanted degradation of either the hydrazone or oximes formed as they are known to be unstable in acidic solutions¹²⁹. The cofactors were capped after purification as if this was done prior to HPLC the cofactors would behave much more similarly and so be difficult to separate.

Although the issues with degradation had not been seen for the hydrazide **11b** it was assumed that they would also occur to a lesser extent due to its decreased reactivity. Therefore, both **11a/b** were capped in the same way. This process would also allow for direct comparison of the hydrazone and oxime cofactors in future labelling studies of DNA, which is desirable to compare their reversibility.

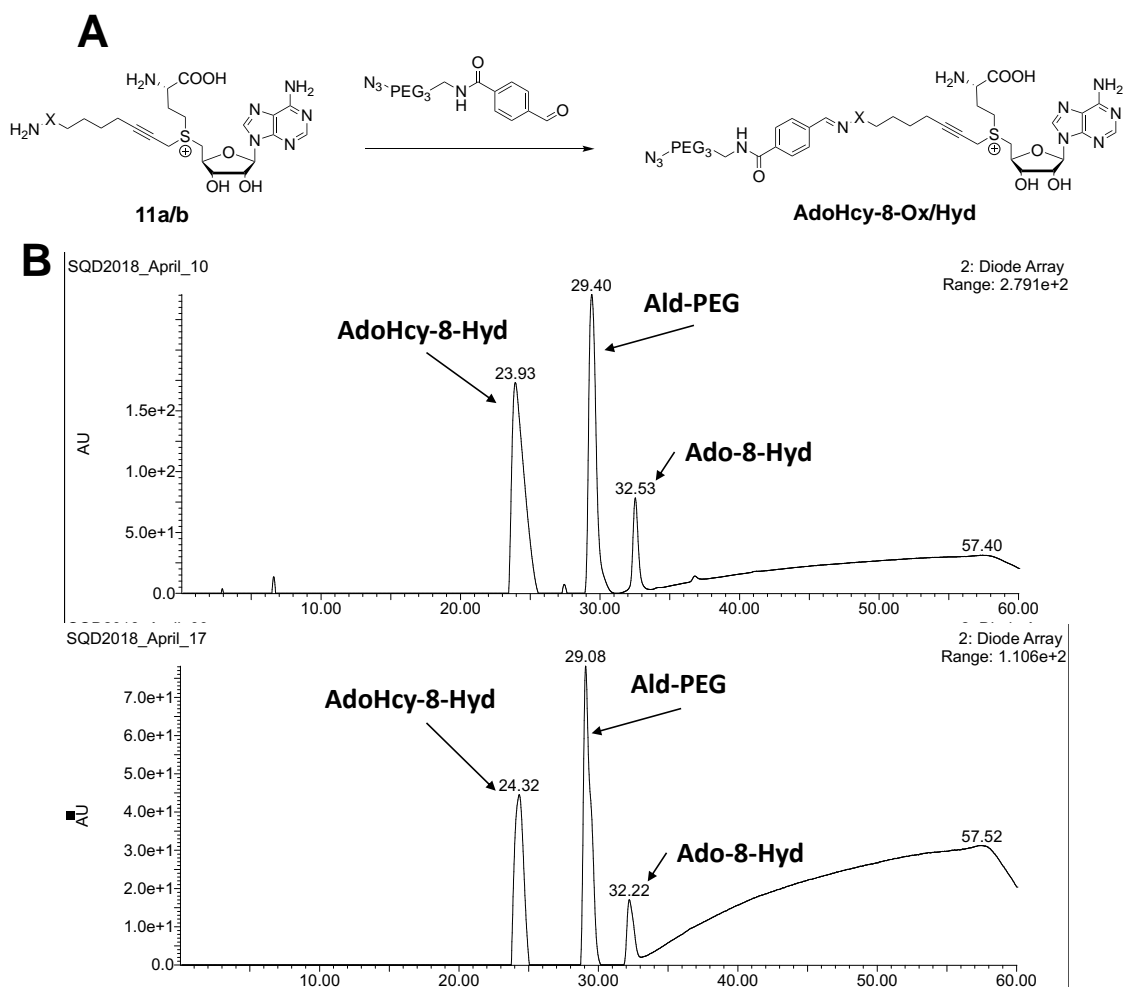


Figure 2.24 A) Synthetic route adopted towards the capping of **11a/b**. B) LCMS traces after capping of each isomer of **11b**. The presence of **AdoHcy-8-Hyd** was confirmed for both isomers by the peak at 23.9/24.3 minutes. Excess aldehyde could also be seen to remain at 29 minutes and minor degradation through loss of the amino acid was also seen for both isomers at 32 minutes. For **AdoHcy-8-Ox** see sup. Fig. 2.11, 2.13.

Capping was achieved by direct addition of the aldehyde to the fractions collected during HPLC which was then rolled at room temperature for half an hour. For the reaction 1.25 equivalents of aldehyde were added, which was calculated using the percentage conversion seen during purification. After rolling the product was collected directly by rotary evaporation and freeze drying in the same way as for **AdoHcy-6-N₃**. Coupling was found to go to completion quickly in the HPLC buffer due to the acidic nature of the ammonium acetate used. LCMS could then be used to check for degradation, where it was found the cofactor was successfully

formed and did not contain any of the impurities that had been seen previously with only minor degradation through loss of the amino acid (Figure 2.24.B). Final characterisation then came by using these cofactor analogues in transalkylation with methyltransferases on DNA. For the oxime cofactor it was more difficult to see separation of the excess aldehyde from the cofactor during LCMS, but it was possible to still confirm its presence (Sup. Fig. S2.11-S1.16).

This method proved very effective at both capping the amine functionality and also eliminating all impurities for both cofactors. As such, once capped it was possible to take both cofactors through to directly test in enzymatic studies which will be discussed in the next chapter. Additionally, no problems were found during storage with stability of the Schiff bases present which gave confidence for future labelling and was likely due, in part, to the excess aldehyde present.

2.5 Chapter Summary

After the successful synthesis of all cofactors and confirmation of their structure, each were taken forward for use with methyltransferases where their structure could further be confirmed through individual labelling studies. Here we highlighted not only the effective synthesis of each cofactor but also developed two new classes of cofactors that may open up completely new applications for methyltransferase-directed bioconjugation. In addition to this we have also greatly improved coupling efficiencies and yields compared to previously reported conditions by reducing the number of equivalents of alkyl halide required during coupling and by improving conversion. We hope the cofactors outlined in this Chapter have the potential to be used as new tools in bioconjugation with a

special emphasis on cysteine cofactors for use with mutant methyltransferases
in situ and Schiff base cofactors for the development of responsive biomolecules.

2.6 Materials and Methods

2.6.1 General materials

Where possible reagents were purchased from Sigma-Aldrich except 6-heptyn-1-ol which was purchased from Fluorochem and N-Boc-hydroxylamine purchased from Alfa Aesar. CutSmart® buffer, was purchased from New England BioLabs (NEB) and MES buffer made using salt purchased from Sigma-Aldrich. Flash chromatography was performed using Geduran Si 60 Å from Merck and TLC (F254) analysis was performed using 60 Å silica gel from VWR International. TLC plates were visualized by staining with potassium permanganate stain or UV-absorption. NMR data was acquired on a Bruker Avance III operating at 300 or 400 MHz. MS spectra were obtained on a Xevo® G2-XS-ToF (Waters) and Synapt-G2-S from electrospray ionization (ESI) and time-of-flight (TOF) measurement in negative or positive ion mode. IR spectra were acquired on Agilent Technologies Cary 600 Series FTIR Spectrometer. UV-Vis absorbance measurements were performed on Shimadzu BioSpec-nano.

2.6.2 HPLC analysis, concentration and yield determination

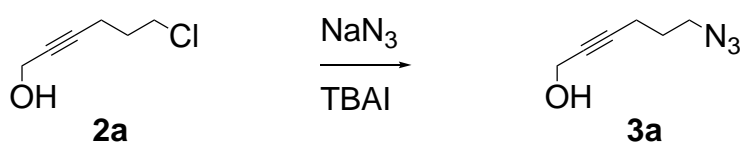
Analytical RP-HPLC was performed on Shimadzu LC-20 Prominence equipped with ACE 5 C18 (250 x 4.6 mm, flow rate 1 ml/min). Elution with 10 mM ammonium acetate pH 5.5/MeOH gradient: 3-30% MeOH over 20 minutes, 30-97% over 15 minutes hold 97% for 10 minutes. Preparative RP HPLC was performed on Agilent Technologies 1260 Infinity equipped with ACE 5 C18 (250 x 21.2 mm, 100 Å, flow rate 10 ml/min). Elution with 10 mM ammonium acetate pH 5.5/MeOH gradient: 3-30% MeOH over 30 minutes, 30-97% over 20 minutes, hold 97% 5 minutes.

The concentration and yields were determined by UV absorbance measurements at 260 nm performed in 0.1 % acetic acid. For the calculation molar extinction coefficient of adenine was used: $\epsilon_{260} = 15,400 \text{ dm}^3 \text{ mol}^{-1} \text{ cm}^{-1}$.

2.6.3 Chemical precursor synthesis

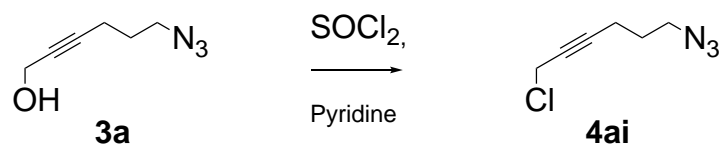
6-chloro hex-2-yn-1-ol (**2a**): This compound was synthesised as laid out by Belyk *et.al.* with no alterations made to the original procedure¹³⁴.

6-azido-hex-2-yn-1-ol (**3a**)



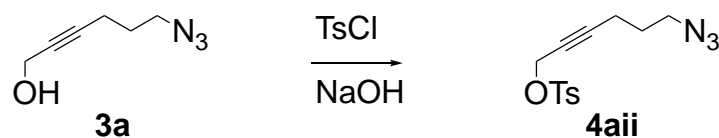
To a solution of **2a** (2.3 g, 17.36 mmol) in DMF (10 ml), Sodium Azide (2.28 g, 34.72 mmol) was added slowly, followed by the addition of tert-Butyl Ammonium Iodide (161.9 mg, 0.43 mmol). The reaction was heated to 70 °C and left to stir for 4 hrs. Once complete the reaction mixture was extracted with Ethyl Acetate (EtOAc) and H₂O. Following extraction, the organic layer was dried and the solvent was removed under reduced pressure yielding the product (1.67 g, 69%): $R_f = 0.72$ (9:1 DCM:EtOAc); $\nu_{\text{max}}(\text{neat})/\text{cm}^{-1}$ 2094 sharp (N₃), 2805-3000 w (C-H), 3103-3665 broad (O-H); δ_{H} (300 MHz, DMSO-d₆) [calibrated using (CH₃)₂SO resonance at 2.50 ppm] 1.71-1.80 (q, 2H, CH₂CH₂CH₂), 2.29-2.36 (tt, 2 H, CH₂CH₂C), 3.40 (t, 2 H, N₃CH₂), 4.23 (t, 2 H, CH₂OH); δ_{C} (400 MHz, DMSO-d₆) [calibrated using (CH₃)₂SO resonance at 39.52 ppm] 15.4 (CH₂C≡C), 27.5 (CH₂CH₂CH₂), 49.1 (OHCH₂), 49.6 (N₃CH₂), 81.1 (OHCH₂C≡C), 82.8 (C≡CCH₂OH).

6-azido 1-chloro-hex-2-yne (**4ai**)



A solution of **3a** (500 mg, 3.59 mmol) was made through dissolution in diethyl ether (1.02 ml) and cooled to -17 °C in an ice bath. Pyridine (0.54 mmol, 0.044 ml) was added to the solution dropwise followed by Thionyl Chloride (4.31 mmol, 0.324 ml). The reaction mixture was brought to room temperature and left to stir overnight. After overnight stirring the reaction mixture was extracted with NaHCO₃ and diethyl ether. The organic layer was dried and solvent removed under reduced pressure yielding **4ai** as a brown oil (417.5 mg, 2.83 mmol, 79%): $\nu_{\text{max}}(\text{neat})/\text{cm}^{-1}$ 2091 sh (N₃), 689 sh (C-Cl); δ_{H} (300 MHz, DMSO-d₆) [calibrated using (CH₃)₂SO resonance at 2.50 ppm] 1.78 (q, 2 H, CH₂CH₂CH₂), 2.35 (tt, 2 H, CH₂C≡C), 3.40 (t, 2 H, N₃CH₂CH₂), 4.13 (t, 2 H, CH₂Cl); δ_{C} (400 MHz, DMSO-d₆) [calibrated using (CH₃)₂SO resonance at 39.52 ppm] 15.5 (CH₂CH₂C≡C), 27.2 CH₂CH₂CH₂), 31.4 (CH₂Cl), 49.6 (N₃CH₂), 76.4 (ClCH₂C≡C), 85.9 (CH₂C≡CCH₂Cl).

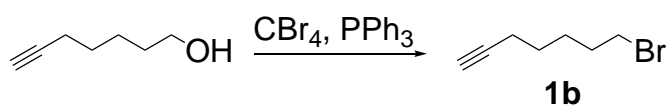
6-azido hex-2-yn-1-yl 6 methylbenzenesulfonate (**4aii**)



A 0.2 M solution of **3a** (300 mg, 2.16 mmol) was made with DCM (10.8 ml) and cooled to -10 °C. Over a period of 10 minutes Tosyl chloride (457.2 mg, 2.37 mmol) was added, followed by pellets of NaOH (436 mg, 10.79 mmol). The mixture was brought to room temperature and left to stir for three hours. Following completion of the reaction the product was extracted with ice water.

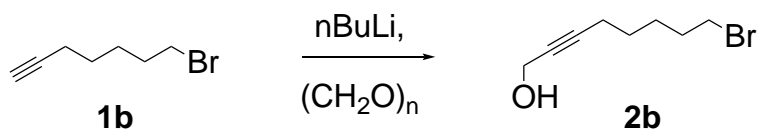
The organic layer was then dried and the solvent was removed under reduced pressure yielding the product as a yellow oil (0.61 g, 90%). δ_{H} (300 MHz, DMSO- d_6) [calibrated using $(\text{CH}_3)_2\text{SO}$ resonance at 2.50 ppm] 1.61 (p, 2 H, $\text{CH}_2\text{CH}_2\text{CH}_2$), 2.18 (tt, 2 H, $\text{CH}_2\text{C}\equiv\text{C}$), 3.27 (t, 2 H, $\text{N}_3\text{CH}_2\text{CH}_2$), 4.65 (t, 2 H, CH_2OTs), 7.32, 7.75 (d, 4 H, Ar-H).

7-Bromo-hept-1-yne (**1b**)



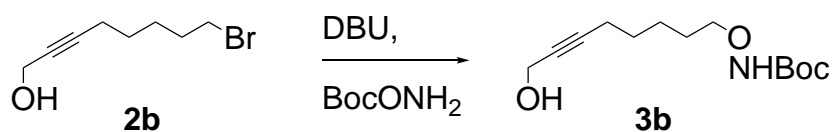
A solution of 6-heptyn-1-ol (5g, 44.6 mmol) was made in dry DCM (60 ml) and cooled on ice. To this triphenylphosphine (17.6 g, 67 mmol) was added, upon complete dissolution tetrabromomethane (22.2 g, 67 mmol) was added slowly. The reaction mixture was brought to room temperature and stirred for 1 hr. After completion, the solvent was removed under reduced pressure. Hexane was added to the crude product forming a white suspension. The hexane fraction was filtered, collected and then the solvent was removed. An oily residue remained which was purified by flash column chromatography with hexane: Yield = 91%, R_f = 0.45 (hexane); ^1H NMR (300 MHz, DMSO- d_6) δ 3.53 (t, J = 6.7 Hz, 2H), 2.75 (t, J = 2.7 Hz, 1H), 2.23 – 2.10 (m, 2H), 1.89 – 1.74 (m, 2H), 1.50 – 1.43 (m, 4H); ^{13}C NMR (101 MHz, DMSO) δ 84.3, 71.3, 35.0, 31.7, 27.0, 27.7, 17.6; IR (cm^{-1}) ν max: 3290, 2918, 1437, 1120, 635, 540.

8-bromooct-2-yn-1-ol (**2b**)



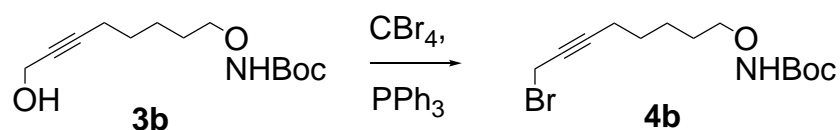
A solution of 7-bromohept-1-yne (**1b**) (20.56 mmol, 3600 mg) was made in Dry THF (12.3 ml) and cooled to -78 °C under Argon. To this a solution of nBuLi in hexanes (1.6 M, 13 ml) was added dropwise, whilst maintaining the temperature below -60 °C. The reaction mixture was then warmed to 0 °C in an ice bath at which point paraformaldehyde (1718 mg, 55.5 mmol) was added under a flow of argon and stirred for 30 minutes. The mixture was then warmed to room temperature and left to stir, the temperature was maintained below 30 °C until the exothermic reaction had stopped. The mixture was then heated to 45 °C for 2 hrs. Once complete the reaction was extracted with diethyl ether and sat. NH₄Cl. The organic layer was collected, dried over anhydrous sodium sulfate and the solvents were removed under reduced pressure to yield the crude product as an oil. Purification was completed by flash column chromatography (Hex: EtOAc 9:1). The product was then collected as a colourless oil: Yield = 55%; R_f = 0.15 (Hex: EtOAc 9:1); ¹H NMR (300 MHz, DMSO-*d*₆) δ 5.04 (t, *J* = 5.7 Hz, 1H), 4.03 (dt, *J* = 5.5, 2.1 Hz, 2H), 3.54 (t, *J* = 6.7 Hz, 2H), 2.20 (m, 2H), 1.88 – 1.75 (m, 2H), 1.52 – 1.40 (m, 4H); ¹³C NMR (101 MHz, CDCl₃) δ 86.1, 78.8, 77.2, 51.5, 33.7, 32.4, 27.8, 27.5, 18.7.

Tert-butyl((8-hydroxyoct-6-yn-1-yl)oxy)carbamate (**3b**)



To a solution of N-Boc Hydroxylamine (890 mg, 6.55 mmol) in DMF (4.3 ml), 8-bromooct-2-yn-1-ol (**2b**) (1200 mg, 5.85 mmol) and 1,8-Diazabicyclo[5.4.0]undec-7-ene (1000 mg, 6.55 mmol) were added. The solution was stirred at 50 °C for 20 hrs. Once complete, the reaction was extracted with DCM and 15% citric acid solution. The organic phases were dried over anhydrous sodium sulfate, collected and the solvent was removed under reduced pressure. A colourless oil was collected as the crude product. This was further purified by flash column chromatography (Hex: EtOAc 8:2). The product was collected as a colourless oil: Yield = 73 %; R_f = 0.27; ^1H NMR (300 MHz, DMSO- d_6) δ 9.91 (s, 1H), 5.03 (t, J = 5.9 Hz, 1H), 4.02 (dt, J = 5.9, 2.2 Hz, 2H), 3.66 (t, J = 6.2 Hz, 2H), 2.17 (tt, J = 6.7, 1.7 Hz, 2H), 1.40 (m, 15H); ^{13}C NMR (101 MHz, DMSO) δ 156.1, 84.0, 80.4, 79.4, 75.1, 49.1, 28.1, 28.0, 27.1, 24.8, 17.9; MS: m/z $[M+H]^+$ = 258.2.

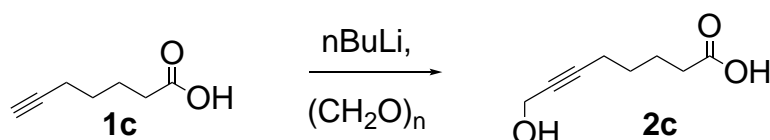
tert-butyl ((8-bromooct-6-yn-1-yl)oxy)carbamate (**4biii**)



A solution of *tert*-butyl((8-hydroxyoct-6-yn-1-yl)oxy)carbamate (**3b**) (1 g, 3.89 mmol) was made in dry DCM (5.2 ml) and cooled on ice. To this triphenylphosphine (1.53 g, 67 mmol) was added. Upon complete dissolution tetrabromomethane (1.94 g, 67 mmol) was added slowly. The reaction mixture was brought to room temperature and allowed to stir for 1 hr. After completion,

the solvent was removed under reduced pressure. Purification was completed using flash column chromatography (Hex: EtOAc 8:2): Yield = 67%; R_f 0.52 (Hex:EtOAc, 8:2); ¹H NMR (300 MHz, DMSO-*d*₆) δ 9.90 (s, 1H), 4.21 (t, *J* = 2.4 Hz, 2H), 3.66 (t, *J* = 6.2 Hz, 2H), 2.25 (tt, *J* = 6.9, 2.4 Hz, 2H), 1.40 (m, 15H); ¹³C NMR (101 MHz, DMSO) δ 156.0, 87.8, 79.3, 76.2, 75.0, 39.5, 28.0, 27.6, 27.0, 24.7, 18.0, 17.2; IR (cm⁻¹) ν max: 1712, 607; MS: *m/z* [M+Na] = 342.35/344.35, [M-^tBuOH] = 246.38/248.38.

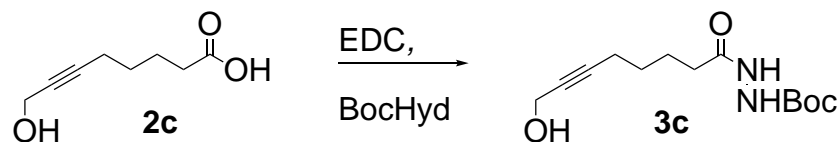
8-hydroxyoct-6-ynoic acid (**2c**)



A solution of 6-heptynoic acid (2 g, 15.87 mmol) was made in dry THF (42 ml) under argon, to this HMPA (34.9 mmol, 6.13 ml) was added and the solution was cooled to -78 °C. To this *n*-BuLi (1.6 M in hexanes, 34.9 mmol, 21.8 ml) was added dropwise whilst maintaining the temperature below -60 °C. The solution was then warmed to -40 °C and stirred for 1 hour. After 1-hour paraformaldehyde (1.47 g, 47.6 mmol) was added via powder funnel under an argon flow. The reaction mixture was then warmed to 45 °C for 4 hours. After reaction, the mixture was quenched with 1 M HCl to pH 4-5 and extracted with EtOAc. Organic layer was then dried over Na₂SO₄ and the EtOAc was removed by rotary evaporation giving the crude product. Purification was completed using flash column chromatography (silica gel, Hex:EtOAc, 6:4): Yield = 68%; R_f = 0.27 (Hex:EtOAc, 6:4); ¹H NMR (300 MHz, DMSO-*d*₆) δ 12.03 (s, 1H), 5.03 (s, 1H), 4.02 (d, *J* = 2.6 Hz, 2H), 2.29 – 2.14 (m, 4H), 1.63 – 1.50 (m, 2H), 1.50 – 1.39 (m, 2H); ¹³C NMR (101 MHz, DMSO) δ 174.3, 83.8, 80.5, 49.1, 33.1, 27.6, 23.7, 23.7,

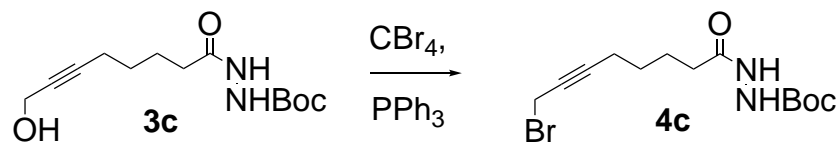
17.7; IR (cm⁻¹) ν max: 3003, 1700, 1411, 1218, 1133, 1002; MS: m/z [M-H] = 155.46.

tert-butyl 2-(8-hydroxyoct-6-ynoyl)hydrazine-1-carboxylate (**3c**)



8-hydroxyoct-6-ynoic acid (**2c**) (1.35 g, 8.65 mmol) and tert-butyl carbazate (1.4 g, 10.38 mmol) were dissolved in 2:1 THF:H₂O (13.5:6.75 ml). To this EDC.HCl (1.87 g, 9.52 mmol) was added slowly over 15 minutes. The mixture was left to stir for 3 hours and then extracted with EtOAc. The organic layer was washed with 0.1 M HCl, water and brine and the organic layer was collected, dried over anhydrous sodium sulfate and the solvent was removed under reduced pressure yielding the product as a white solid: Yield = 63%; ¹H NMR (400 MHz, DMSO-*d*₆) δ 9.47 (s, 1H), 8.66 (s, 1H), 5.04 (t, *J* = 5.9 Hz, 1H), 4.02 (dt, *J* = 5.9, 2.2 Hz, 2H), 2.19 (tt, *J* = 7.1, 2.2 Hz, 2H), 2.06 (t, *J* = 7.2 Hz, 2H), 1.58 (p, *J* = 7.3 Hz, 2H), 1.50 – 1.32 (m, 12H); ¹³C NMR (101 MHz, DMSO) δ 172.0, 155.6, 84.3, 80.9, 79.4, 49.6, 33.0, 28.5, 28.1, 24.7, 18.2; IR (cm⁻¹) ν max: 3300, 1668, 1370, 1245, 1157; MS: m/z [M+Na] = 294.15.

tert-butyl 2-(8-bromooct-6-ynoyl)hydrazine-1-carboxylate (**4ciii**)



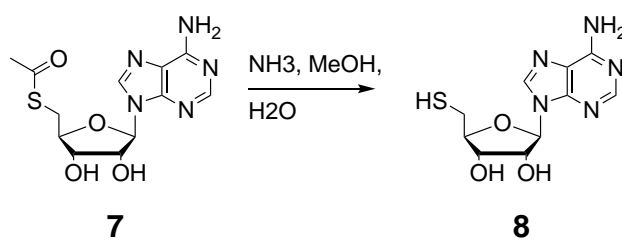
A solution of tert-butyl 2-(8-hydroxyoct-6-ynoyl)hydrazine-1-carboxylate (**3c**) (300 mg, 1.11 mmol) was made in dry DCM (3.33 ml) and cooled on ice. Triphenylphosphine (437 mg, 1.67 mmol) was added and left to dissolve, once dissolved tetrabromomethane (552 mg, 1.67 mmol) was added slowly. The

reaction was then brought to room temperature and left to stir for 1 hour. After reaction the solvent was removed under reduced pressure and the crude mixture was purified by flash column chromatography (Hex:EtOAc, 7:3): Yield = 55%; R_f = 0.15 (Hex:EtOAc 7:3); ^1H NMR (300 MHz, DMSO- d_6) δ 9.48 (s, 1H), 8.67 (s, 1H), 4.21 (t, J = 2.3 Hz, 2H), 2.27 (tt, J = 6.9, 3.4 Hz, 2H), 2.06 (t, J = 7.4 Hz, 2H), 1.65 – 1.31 (m, 13H); ^{13}C NMR (101 MHz, DMSO) δ 171.4, 155.2, 87.7, 78.9, 76.3, 54.9, 39.5, 32.5, 28.0, 27.3, 24.1, 17.9, 17.2; MS: m/z $[M+\text{Na}]$ = 355/357.08.

2.6.4 Cysteine cofactor synthesis

5'-Acetylthio-5'-deoxyadenosine (**7**): This compound was synthesised as laid out by Pignot *et.al* with no alterations made¹⁴⁵.

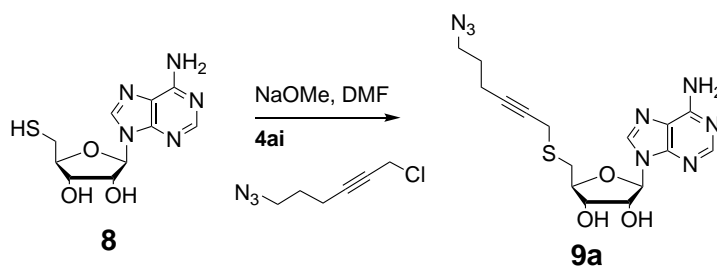
5'-Deoxy-5'-thioadenosine (**8**)



To eliminate traces of oxygen, nucleoside **7** (150 mg, 0.45 mmol) was sealed in a round bottom flask and degassed with Argon for 10 minutes. To this a solution of saturated NH_3 in MeOH (7 N, 15 ml) was added under Argon at 0 °C. Then a solution of saturated NH_3 in H_2O (35%, 6 ml) was added under Argon. The solution was kept at 0 °C and left to stir for 1.5 hrs, until the reaction was complete. The NH_3 was removed from solution by bubbling Argon in to the solution for 15 minutes. Once all traces of NH_3 had been removed, MeOH was removed under reduced pressure. The solution was then frozen and remaining solvent was removed by lyophilisation to yield **8** (119 mg, 94%) as a colourless

white powder. δ_{H} (300 MHz, DMSO- d_6) [calibrated using $(\text{CH}_3)_2\text{SO}$ resonance at 2.50 ppm] 2.51 (s, 1H, CH_2SH), 2.73-2.95(m, 2 H, 5'a-H, 5'b-H), 3.99 (dt, 1H, 4'-H), 4.20 (q, 1 H, 3' H), 4.80 (q, 1 H, 2'-H), 5.91 (d, 1 H, 1'-H), 7.35 (s, br., 2 H, D_2O exchangeable, NH_2), 8.18 (s, 1 H, 2-H), 8.38 (s, 1 H, 8-H); δ_{C} (400 MHz, DMSO- d_6) [calibrated using $(\text{CH}_3)_2\text{SO}$ resonance at 39.52 ppm] 26.5 (s, 5' CH_2), 72.0 (s, 3' C), 72.6 (s, 2' C), 85.5 (s, 4' C), 87.3 (s, 1' C), 119.2 (s, C-5), 140.0 (s, C-8), 149.5 (s, C-4), 152.8 (s, C-2), 156.1 (s, C-6); MS (TOF ES-) m/z 284.1 [M+H].

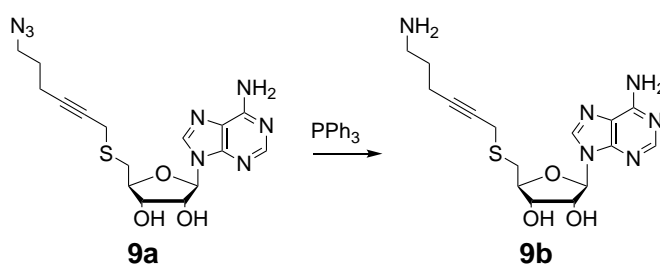
S-6-azido-hex-2-yne-5'-thioadenosine (**9a**)



To a solution of **8** (800 mg, 2.83 mmol), in DMF (22.6 ml), **4ai** (4.52 mmol) was added dropwise. After 5 minutes stirring NaOMe (247 mg, 4.52 mmol) was added. The resulting solution was heated to 70 °C and left to stir for 3 hours. After reaction the solvent was removed under reduced pressure and the product was purified by flash column chromatography. The fractions containing the product were combined and the solvent was removed yielding an off white solid (0.81 g, 2 mmol, 71%): R_f = 0.25 (9:1 DCM:MeOH); δ_{H} (300 MHz, DMSO- d_6) [calibrated using $(\text{CH}_3)_2\text{SO}$ resonance at 2.50 ppm] 1.66 (p, 2 H, $\text{CH}_2\text{CH}_2\text{CH}_2$), 2.25 (t, 2 H, $\text{CH}_2\text{C}\equiv\text{C}$), 2.99 (m, 2 H, 5'a-H, 5'b-H), 3.37, (t, 2 H, N_3CH_2), 4.07 (m, 1 H, 4'-H), 4.18 (m, 1 H, 3'-H), 4.76 (m, 1 H, 2'-H), 5.36 (d, 1 H, D_2O exchangeable, 3'-OH), 5.55 (d, 1 H, D_2O exchangeable, 2'-OH), 5.90 (d, 1 H, 1'

H), 7.32 (s, br., 2 H, D₂O exchangeable, NH₂), 8.16(s, 1H, 2-H), 8.34 (s, 1H, 8-H); δ_C (400 MHz, DMSO-d₆) [calibrated using (CH₃)₂SO resonance at 39.52 ppm] 14.4 (CH₂CH₂C \equiv C), 19.9 (5'CH₂), 28.0 (CH₂CH₂CH₂), 33.8 (C \equiv CCH₂S), 49.9 (N₃CH₂), 72.9 (2'/3' C, proven by HSQC), 77.4 (C \equiv CCH₂S), 82.1 (C \equiv CCH₂S), 83.9 (4' C), 87.8 (1' C), 119.5 (s, C-5), 140.2 (s, C-8), 149.7 (s, C-4), 153.0 (s, C-2), 156.3 (s, C-6); MS (TOF ES-) *m/z* 485.1 [M+H].

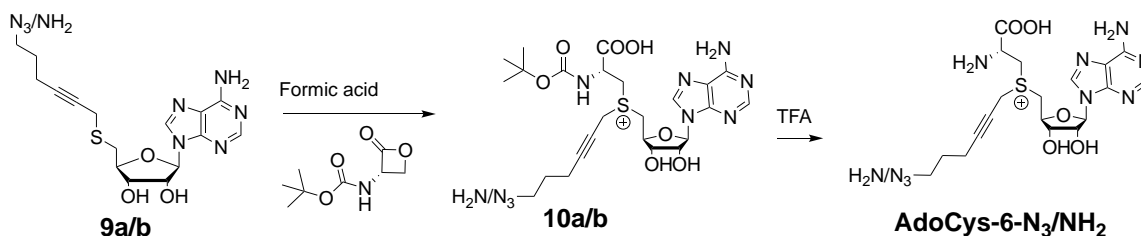
S-6-amino-hex-2-yne-5'-thioadenosine (**9b**)



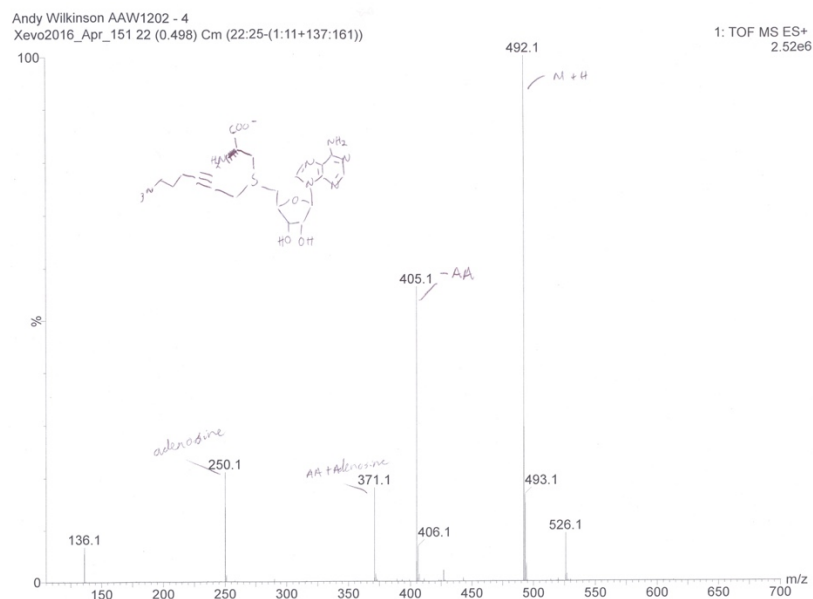
9a (210 mg, 0.52 mmol), was suspended in a 2:1 solution of THF:H₂O (7.8 ml). To this an excess of Triphenylphosphine (162 mg, 0.60 mmol) was added. The resulting solution was stirred under reflux for 3 hrs. Following completion the reaction was purified by column (silica pad, CH₂Cl₂:MeOH 7:3). The fractions containing the product were combined and the solvent was removed yielding an off white solid (91.3 mg, 0.24 mmol, 48%): *R_f* = 0 (9:1 DCM:MeOH); δ_H (300 MHz, DMSO-d₆) [calibrated using (CH₃)₂SO resonance at 2.50 ppm] 1.51 (p, 2 H, CH₂CH₂CH₂), 2.21 (tt, 2 H, CH₂CH₂C \equiv C), 2.60 (t, 2 H, NH₂CH₂), 2.96 (m, 2 H, 5'a-H, 5'b-H), 3.34 (t, 2 H, C \equiv CCH₂S), 3.83 (br s, 2 H, OH), 4.07 (m, 1 H, 4'-H), 4.18 (m, 1 H, 3'-H), 4.75 (t, 1 H, 2'-H), 5.89 (d, 1 H, 1'-H), 7.33 (s, br., 2 H, D₂O exchangeable, NH₂), 8.16, 8.34 (2 x s, 2 H, Ar-H); δ_C (400 MHz, DMSO-d₆) [calibrated using (CH₃)₂SO resonance at 39.52 ppm] 16.0 (CH₂CH₂C \equiv C), 20.1 (C \equiv CCH₂S), 30.0 (5'-CH₂), 31.5 (CH₂CH₂CH₂), 40.5 (NH₂CH₂), 73.1 (2'/3' C),

76.4 (C≡CCH₂S), 84.1(C≡CCH₂S), 84.1 (4' C), 87.9 (1' C), 119.5 (s, C-5), 140.2 (s, C-8), 149.7 (s, C-4), 153.0 (s, C-2), 156.3 (s, C-6).

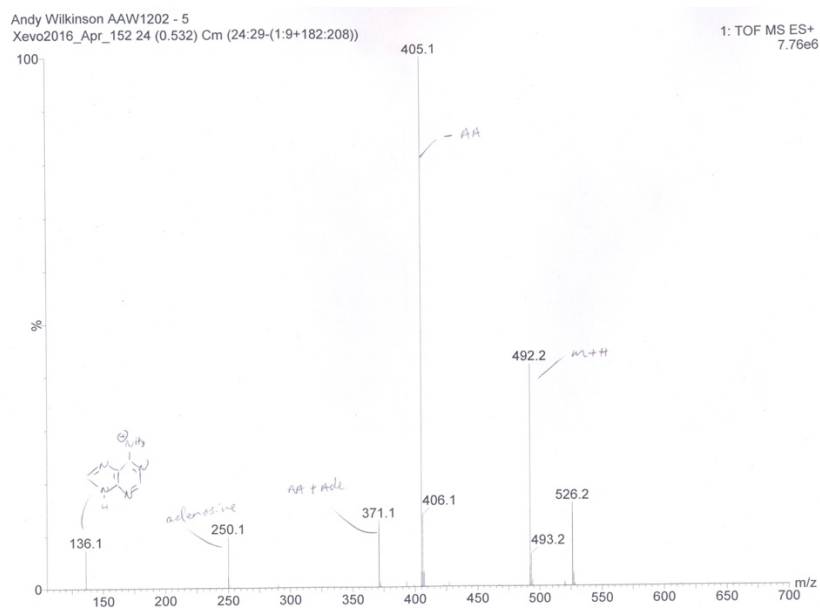
General Cysteine cofactor coupling:



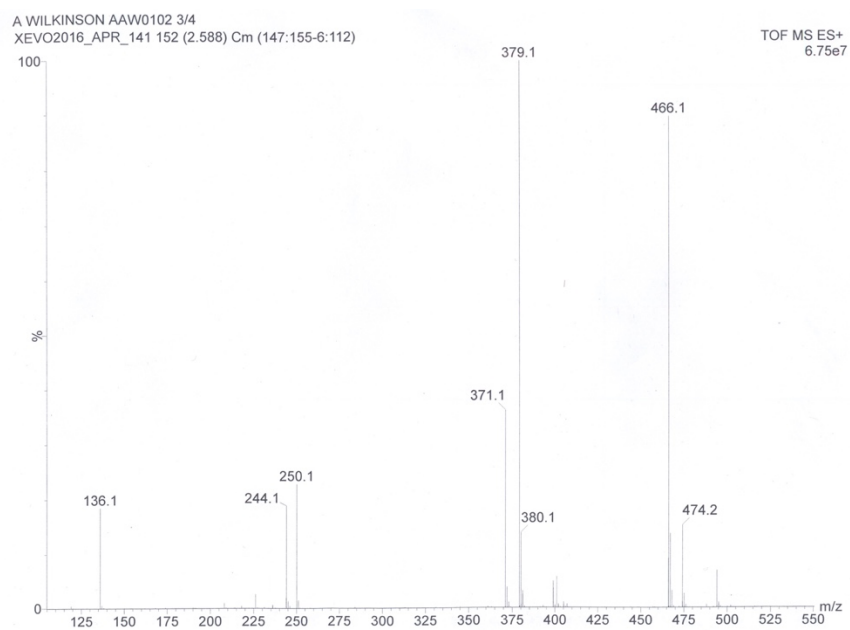
S-6-amino/azido-hex-2-yne-5' thioadenosine **9a/b** (0.16 mmol) is suspended in formic acid. To this N-Boc-L-serine-β-Lactone (22 mg, 0.24 mmol) was added. The reaction was flushed with Argon and stirred at room temperature overnight. After reaction the crude product was collected by extraction with diethyl ether and the aqueous layer was collected and dried. The dried product was then dissolved in trifluoroacetic acid (400 µl) and stirred at room temperature for 2 hours. The product was then dried and resuspended in H₂O (2 ml) before final purification by prep-HPLC. Retention times: **AdoCys-6-N₃** isomer 1 = 14.13 mins, isomer 2 = 16.05 mins; **AdoCys-6-NH₂** isomer 1 = 10.95 mins, isomer 2 = 11.47 mins; MS: m/z [M+H] = 492.53 (**AdoCys-6-N₃**), [M+H] = 466.54 (**AdoCys-6-NH₂**).



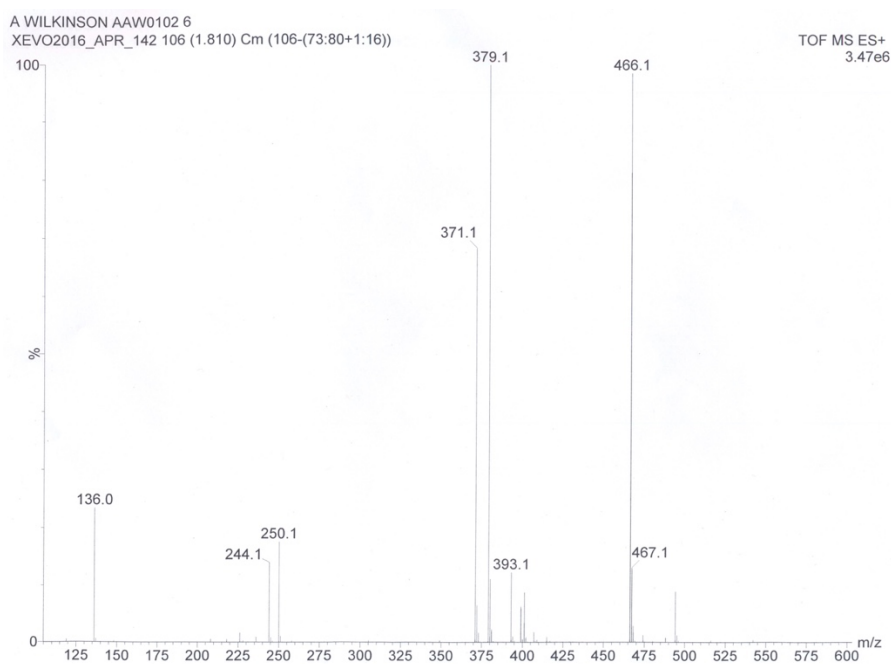
Sup. Fig. S2.1 Mass spectrum of isomer 1 collected during prep-HPLC of **AdoCys-6-N₃** $t = 14.13$ mins from Figure 2.17



Sup. Fig. S2.2 Mass spectrum of isomer 2 collected during prep-HPLC of **AdoCys-6-N₃** $t = 16.05$ mins from Figure 2.17



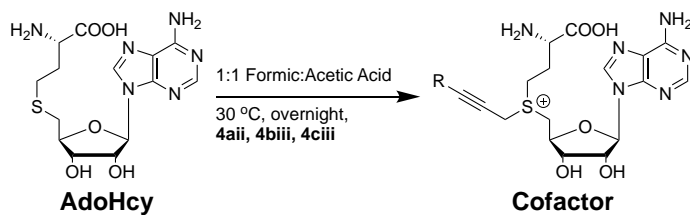
Sup. Fig. S2.3 Mass spectrum of isomer 1 collected during prep-HPLC of **AdoCys-6-NH₂** $t = 10.95$ mins from Figure 2.17



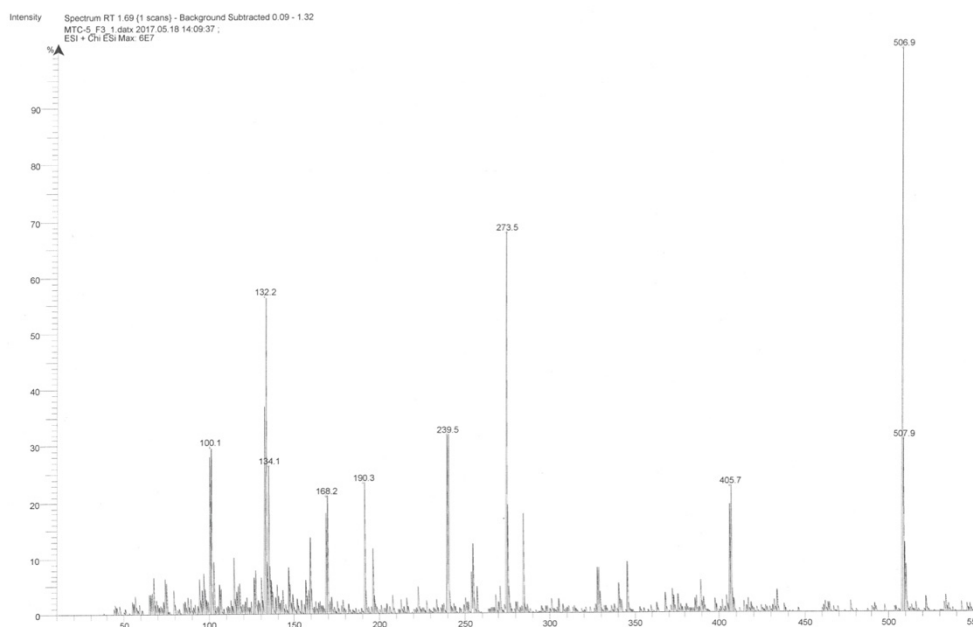
Sup. Fig. S2.4 Mass spectrum of isomer 2 collected during prep-HPLC of **AdoCys-6-NH₂** $t = 11.47$ mins from Figure 2.17

2.6.5 Homocysteine cofactor synthesis

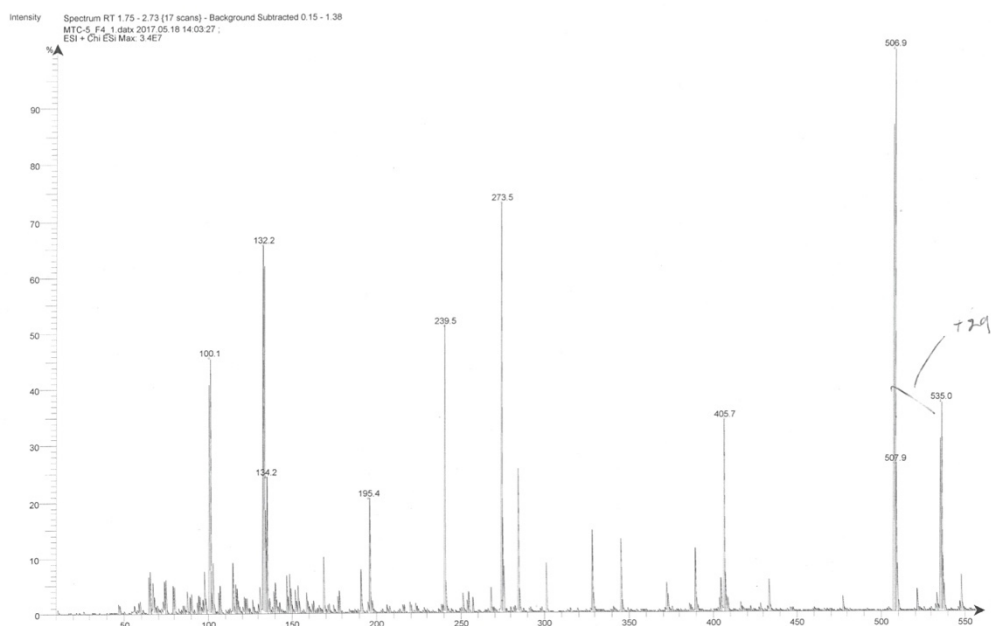
General Homocysteine cofactor coupling:



A solution of S-adenosyl-L-homocysteine (**AdoHcy**) (15 mg, 0.04 mmol) was made in a 1:1 mixture of Formic and acetic acid (300 μ l). Linker (1.2 mmol, 30 equivs) was then added dropwise, on ice. The reaction mixture was warmed to 30 °C and left to stir overnight. After overnight stirring the reaction mixture was extracted twice with diethyl ether and the aqueous layer was collected and then dried by lyophilisation. Each product was then either purified directly or taken forward for Boc deprotection. Retention times: **AdoHcy-6-N₃** isomer 1 = 20.57 mins, isomer 2 = 22.86 mins; MS: m/z [M+H] = 506 (**AdoHcy-6-N₃**).



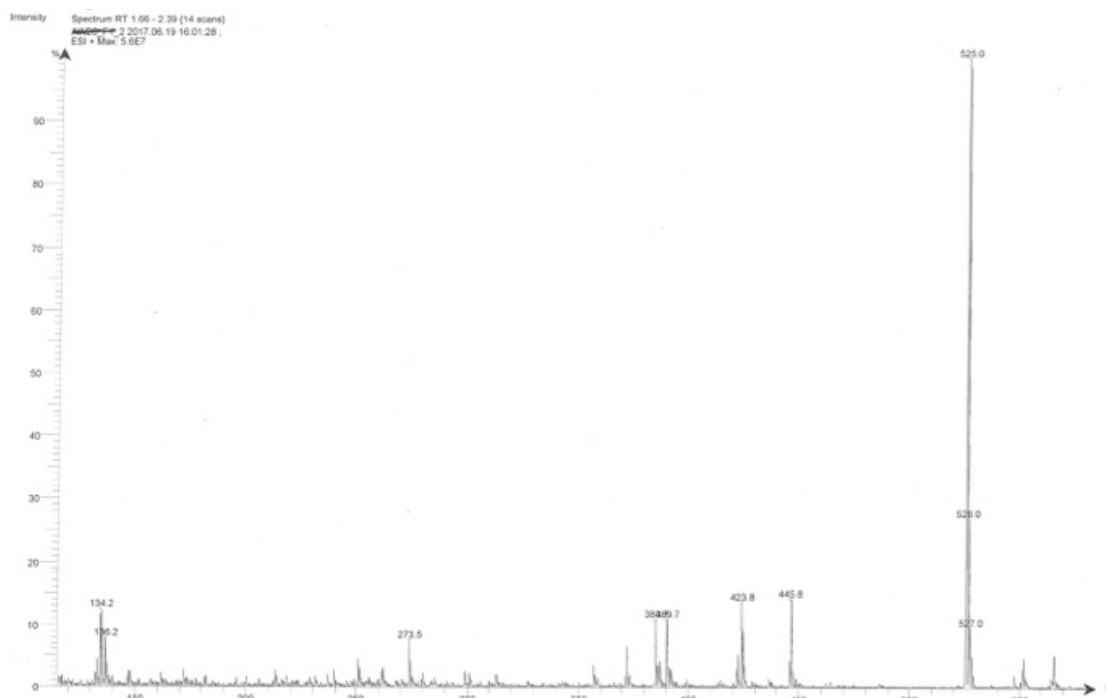
Sup. Fig. S2.5 Mass spectrum of isomer 1 collected during prep-HPLC of **AdoHcy-6-N₃** t = 20.57 mins from Figure 2.20



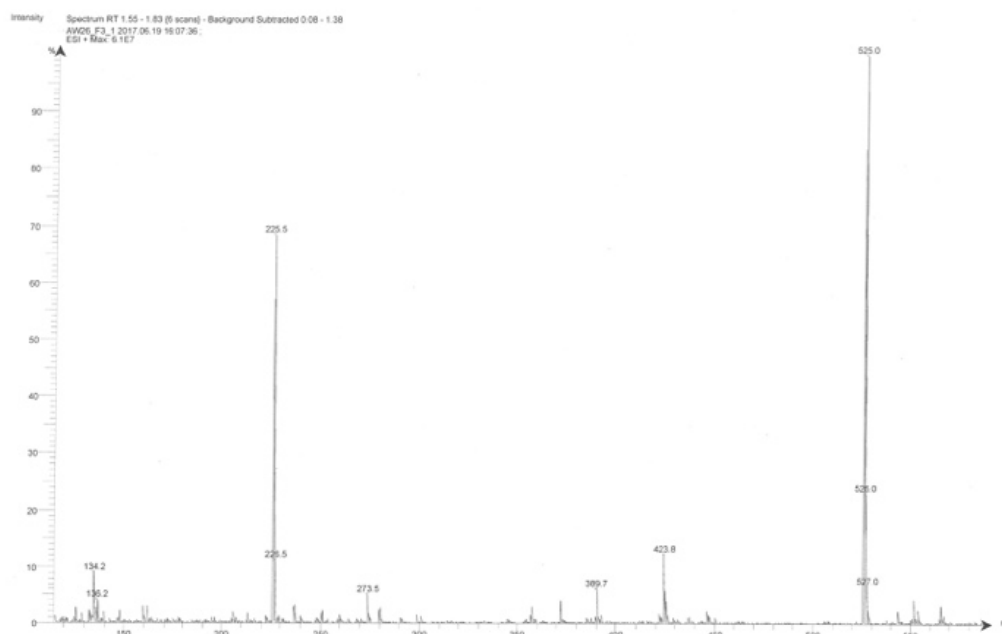
Sup. Fig. S2.6 Mass spectrum of isomer 2 collected during prep-HPLC of **AdoHcy-6-N₃** $t = 22.86$ mins from Figure 2.20

General method for cofactor boc deprotection:

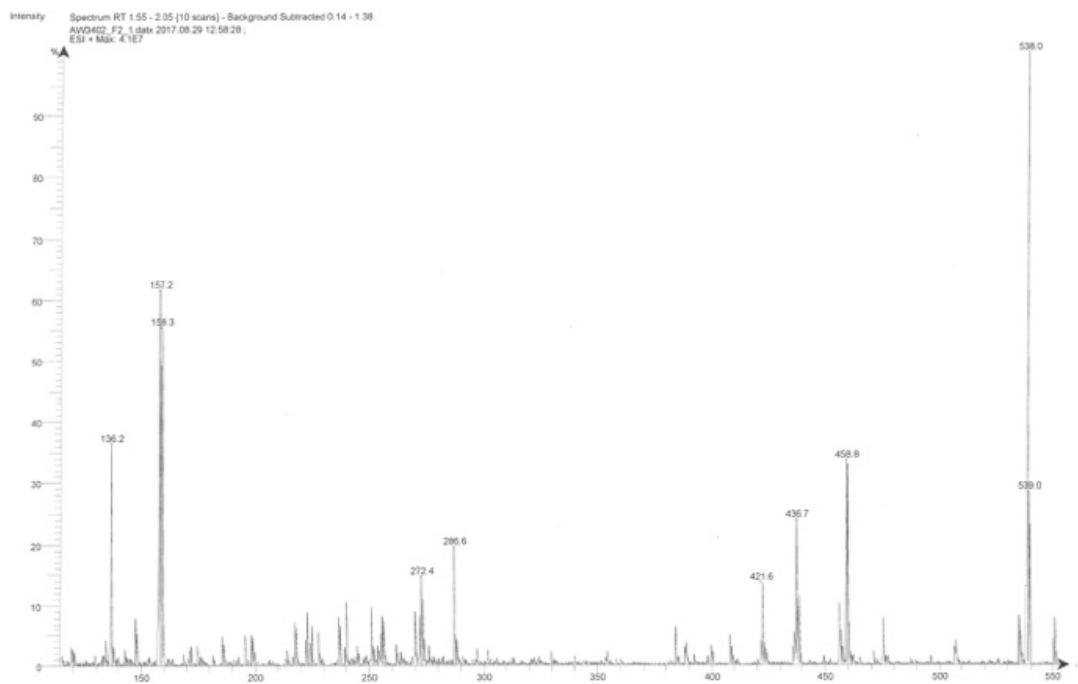
After coupling and lyophilisation the crude reaction product was dissolved in neat tri-fluoroacetic acid (300 μ l) and left to stir at room temperature for 2 hours. Once complete the TFA was removed with Argon yielding a brown oil. The crude product was then resuspended in 2 ml H₂O for prep-HPLC. Retention times: **11a** isomer 1 = 25.47 mins, isomer 2 = 28.24 mins; **11b** isomer 1 = 17.51 mins, isomer 2 = 18.73 mins; MS: m/z $[M+H] = 524$ (**11a**), $[M+H] = 538$ (**11b**).



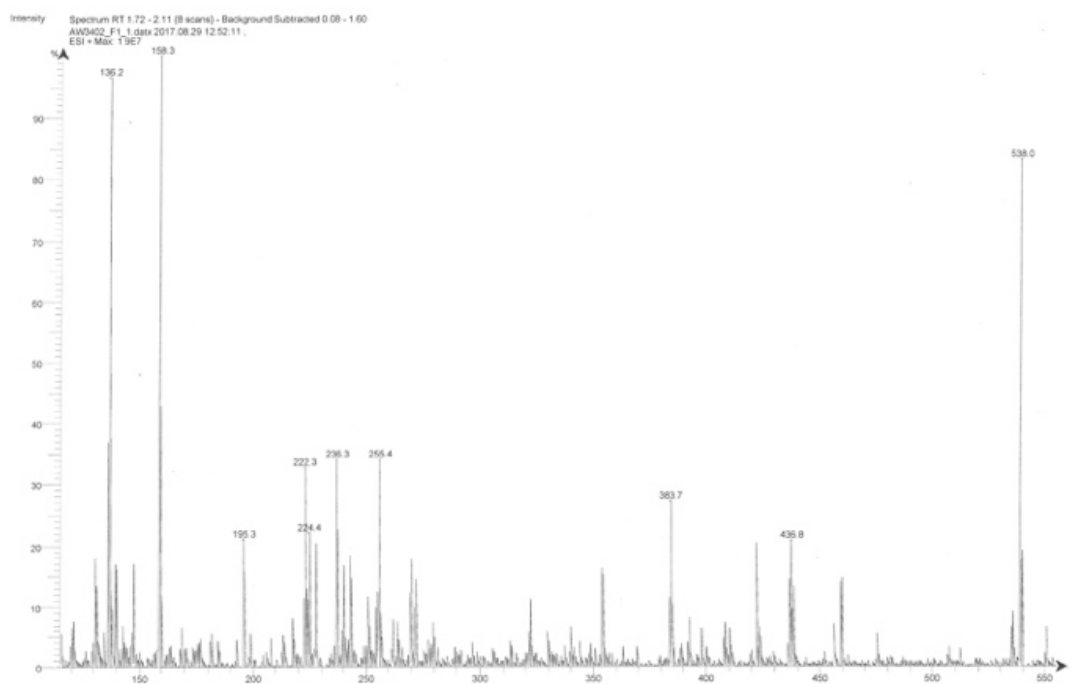
Sup. Fig. S2.7 Mass spectrum of isomer 1 collected during prep-HPLC of **11a** $t = 25.47$ mins from Figure 2.20



Sup. Fig. S2.8 Mass spectrum of isomer 2 collected during prep-HPLC of **11a** $t = 28.24$ mins from Figure 2.20

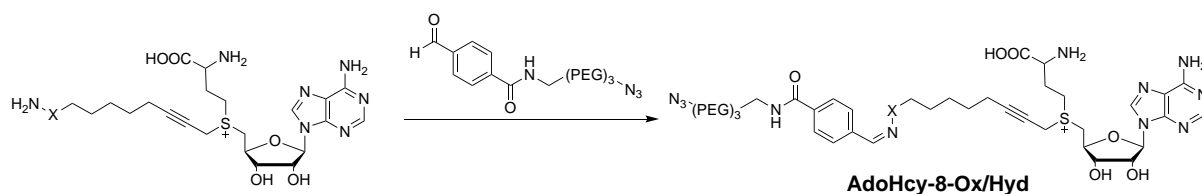


Sup. Fig. S2.9 Mass spectrum of isomer 1 collected during prep-HPLC of **11b** $t = 17.51$ mins Figure 2.20

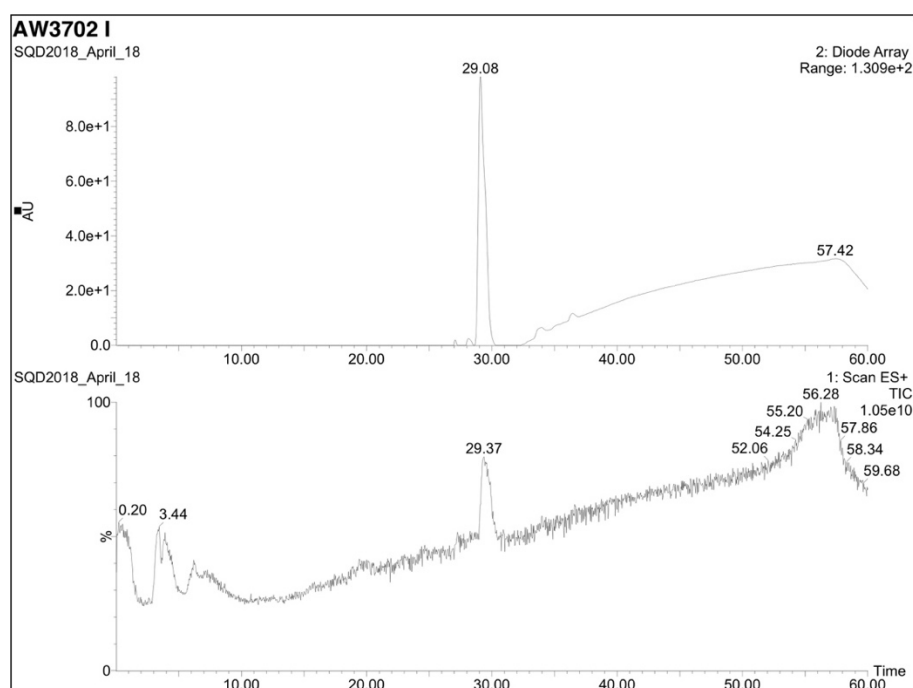


Sup. Fig. S2.10 Mass spectrum of isomer 2 collected during prep-HPLC of **11b** $t = 18.73$ mins Figure 2.20

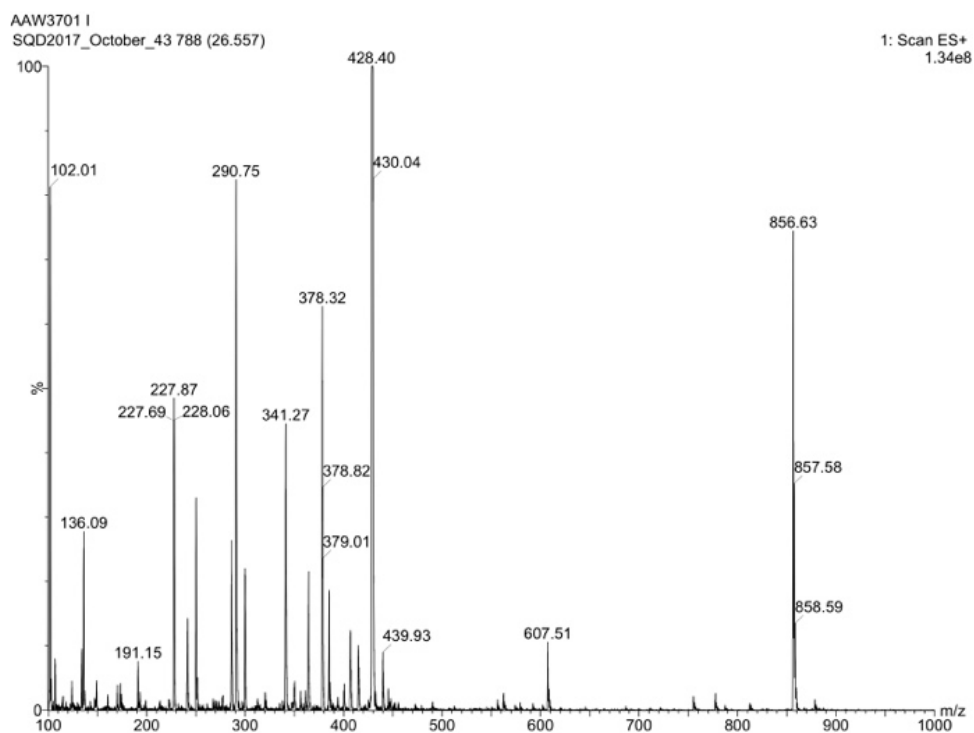
Synthesis of cofactors **AdoHcy-8-Ox/Hyd**



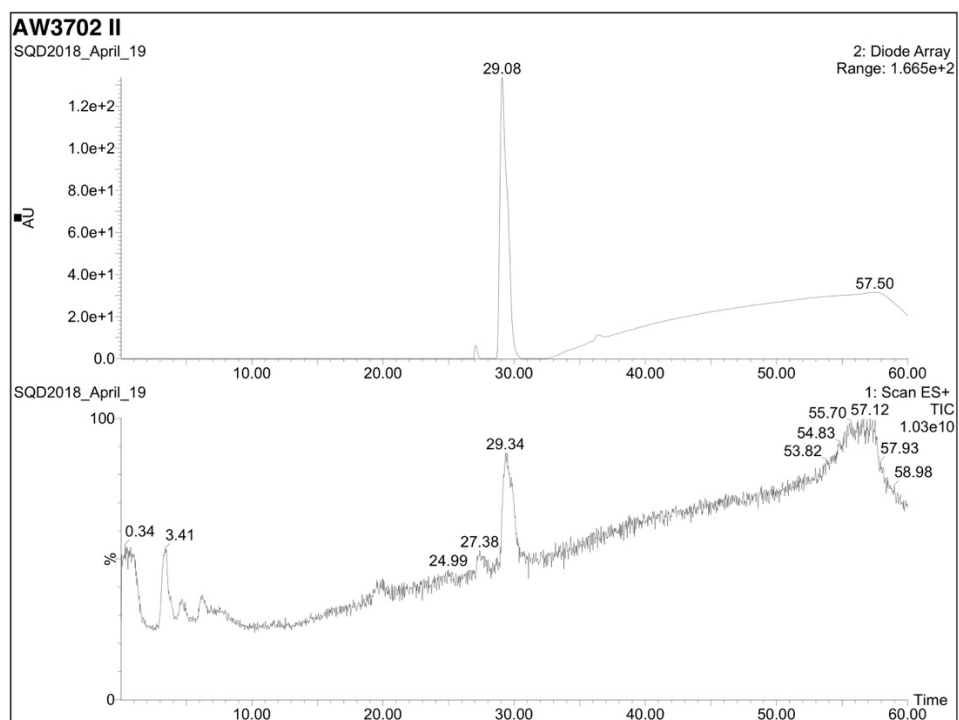
To the collected HPLC fractions aldehyde (1.2 equivs) was added and rolled for 30 mins at room temperature. The fractions were then dried by lyophilisation. Once dry, the solids were dissolved in 100 μ l 0.1% Acetic Acid and stored at -20 $^{\circ}$ C. HRMS-ESI(+): calculated for **AdoHcy-8-Hyd** $C_{38}H_{54}N_{11}O_{10}S$: 869.3728, found: 869.3729 $[M+H]^+$ HRMS-ESI(+): calculated for **AdoHcy-8-Ox** $C_{38}H_{53}N_{12}O_{10}S$: 856.3776, found: 856.3780 $[M+H]^+$.



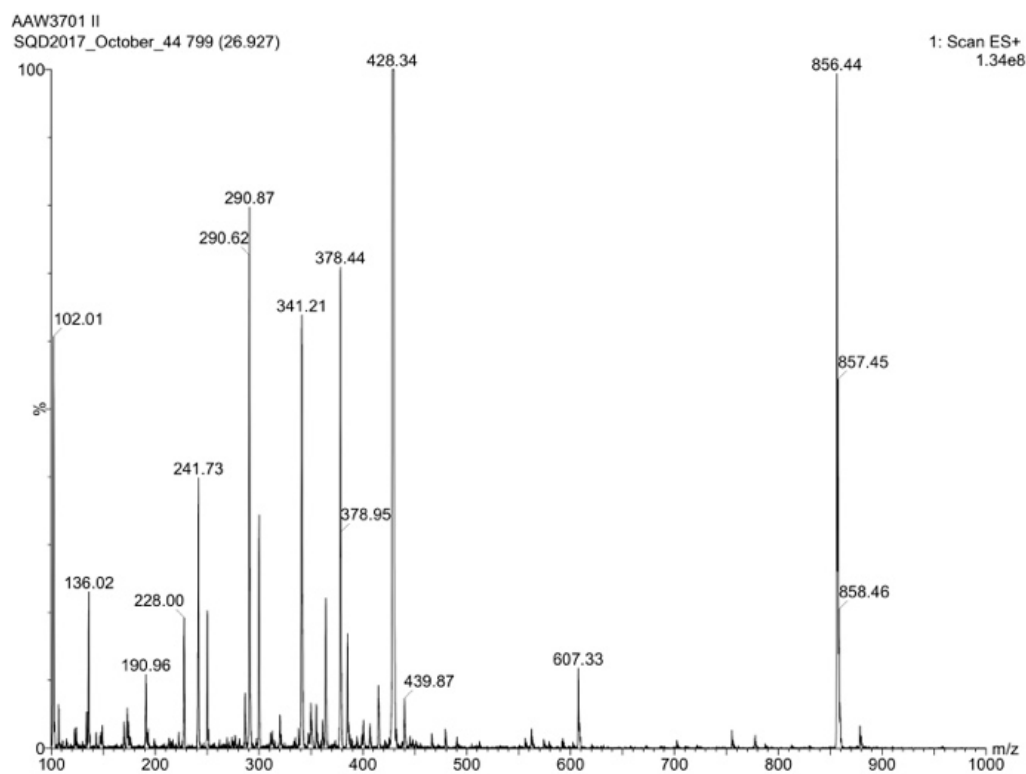
Sup. Fig. S2.11 LCMS trace of **AdoHcy-8-Ox** isomer 1 after aldehyde introduction. Top = Diode array 200-800 nm, bottom = TIC.



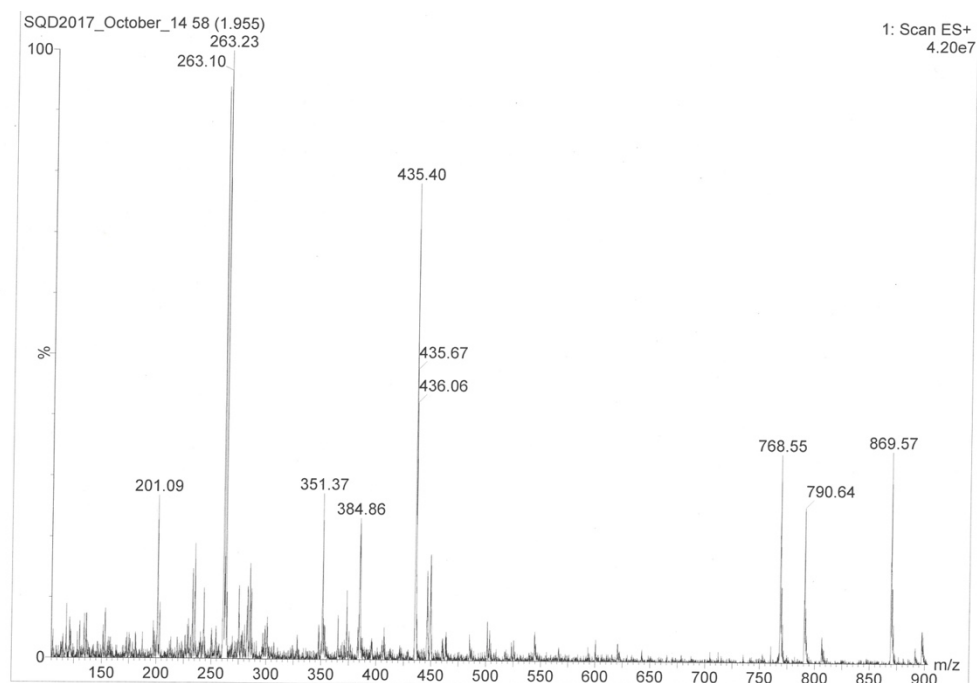
Sup. Fig. S2.12 Mass spectrum from LCMS of **AdoHcy-8-Ox** isomer 1 $t = 29.37$ minutes



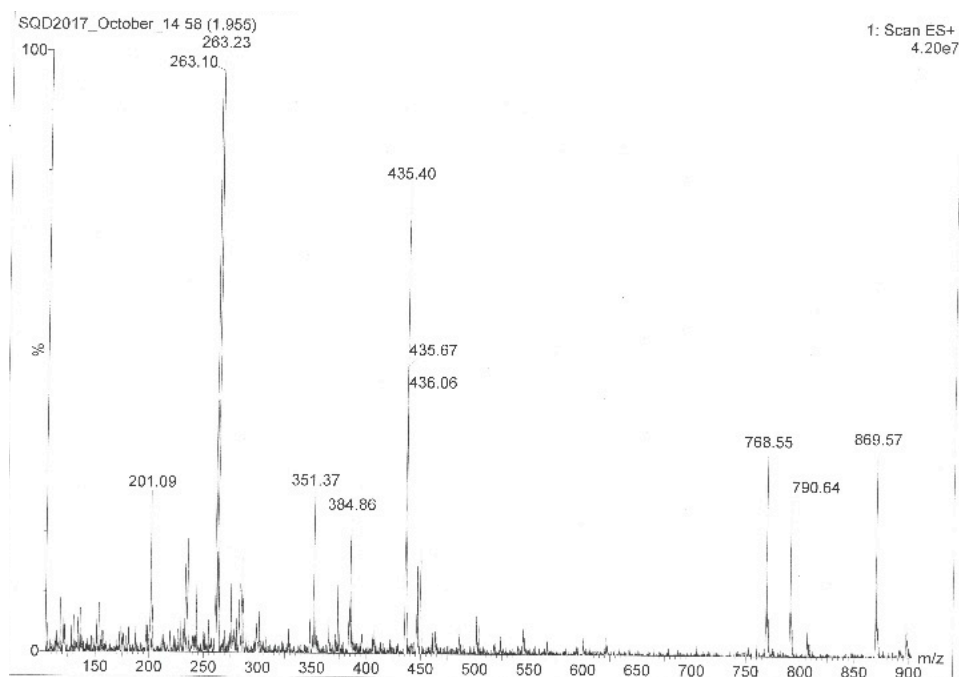
Sup. Fig. S2.13 LCMS trace of **AdoHcy-8-Ox** isomer 2 after aldehyde introduction. Top = Diode array 200-800 nm, bottom = TIC.



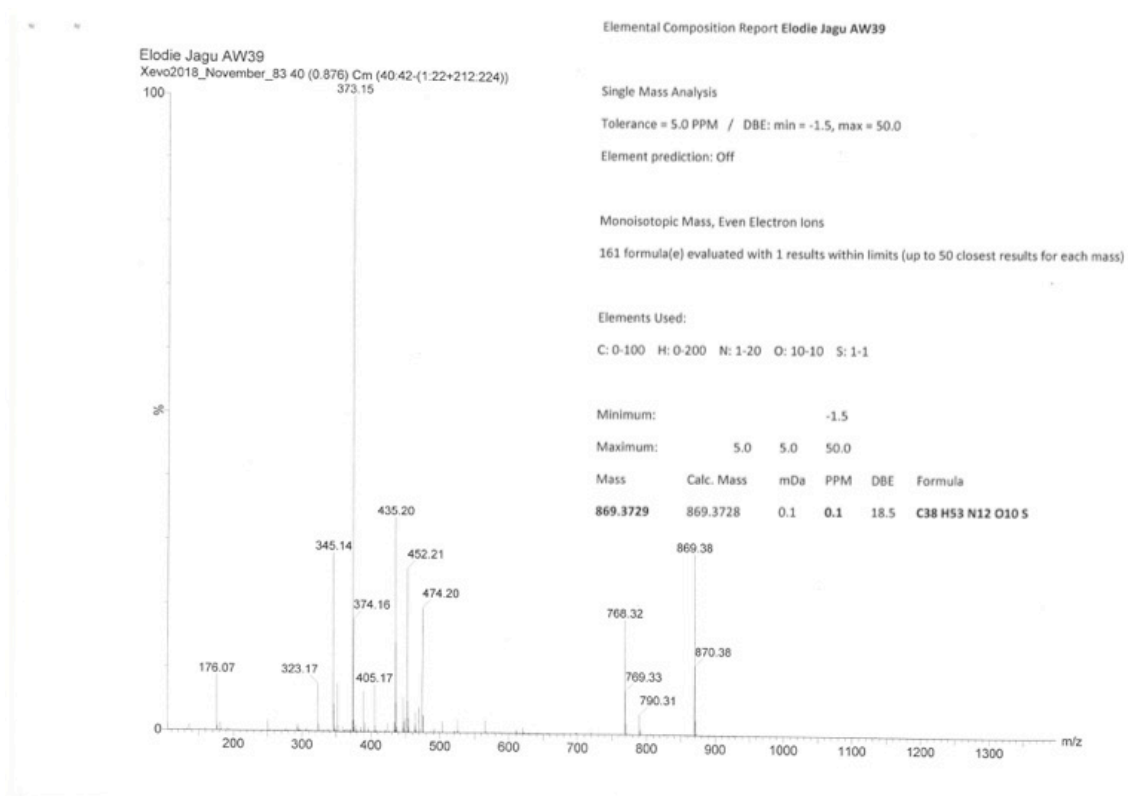
Sup. Fig. S2.14 Mass spectrum from LCMS of **AdoHcy-8-Ox** isomer 2 $t = 29.37$ minutes



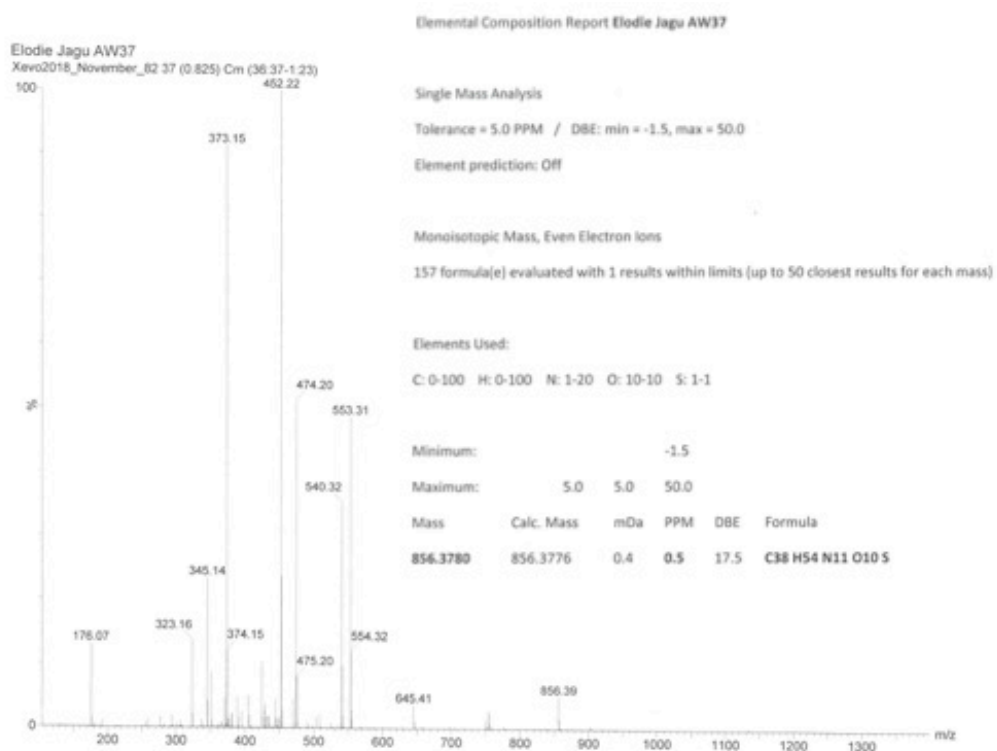
Sup. Fig. S2.15 Mass spectrum from LCMS of **AdoHcy-8-Hyd** isomer 1 $t = 23.93$ minutes (Figure 2.24.B)



Sup. Fig. S2.16 Mass spectrum from LCMS of **AdoHcy-8-Hyd** isomer 2 $t = 24$. minutes (Figure 2.24.B)



Sup. Fig. S2.17 HRMS of **AdoHcy-8-Hyd**



Sup. Fig. S2.18 HRMS of **AdoHcy-8-Ox**

3 Probing cofactor analogue effectiveness

3.1 Introduction

After the successful synthesis of each of the five doubly-activated analogues highlighted in chapter 1 it was then necessary to test their ability as cofactors in the transalkylation of DNA. To this end we identified two key DNA transalkylation assays that would allow us to carry out in depth analysis for each cofactor, which were gel electrophoresis and oligo-HPLC. Gel electrophoresis studies were based around those seen previously in the literature allowing initial assessment of the performance of each cofactor⁹¹. Oligo-HPLC was then used in tandem with MS to gain further evidence around transalkylation in an attempt to quantitatively identify the modifications introduced. In combination these assays could then give us a clear understanding of how each cofactor performed in the transalkylation of DNA highlighting the following key features:

Diastereomeric preference – As highlighted during prep-HPLC, two fractions were collected which we suspected to be the two diastereoisomers that can be formed upon creation of the sulfonium centre¹³¹. However, the structures of these cannot be confirmed by HPLC alone. As such, during electrophoresis we tested both potential isomers to identify any differences in their activity in an effort to identify which was the active *S,S* form. If each fraction collected during HPLC were diastereoisomers then a clear preference by each enzyme should be shown for the fraction containing the naturally active form.

Transalkylation efficiency – In addition to being able to identify the most effective isomer present, in combination with modelling during electrophoresis it is also

possible to gain an understanding around enzyme transalkylation efficiency in different conditions. This can be done by modelling the random introduction of modifications to DNA at different efficiencies and then predicting the fragments formed upon restriction. Doing this allows for comparative analysis with actual transalkylation results which helps to identify transalkylation efficiency and the best conditions to take forward for future studies. This analysis can then also be used to compare the efficiency of each cofactor analogue synthesised highlighting any compromises structural changes may have made to enzyme binding and transalkylation efficiency overall.

Palindromic alkylation – Another area examined whilst testing cofactor efficacy was if transalkylation with AdoMet analogues occurred on both target bases in a palindromic recognition sequence. This is something which has been suggested to occur through fluorescent labelling studies but as yet is something for which definitive evidence has not been provided²². By developing oligo-HPLC as an analytical tool for monitoring transalkylation we aimed to test this by individually analysing each strand of a short oligo.

Structure introduced – Furthermore, we aimed to collect modified DNA and confirm the identity of the modification by MS. Doing this would not only prove if the modification had been successful or not but would provide further evidence that alkylation had introduced the expected modification to DNA.

Modification functionality – The final area we wanted to explore through transalkylation studies was to test the functionality of the modifications

introduced, which we again aimed to achieve using HPLC. Here we hoped to prove that the reactive groups introduced during alkylation could be further utilised for introduction of more complex functionality like fluorophores, through click chemistry.

3.2 Gel electrophoresis

Initial studies into the enzymatic activity of each cofactor analogue with methyltransferases were completed using gel electrophoresis. This experiment was adapted from the literature for similar AdoMet analogues⁹¹. Here DNA (pUC19) was first labelled with the cofactor and methyltransferase of interest, followed by restriction with a complimentary endonuclease. For example, M.TaqI an DNA-adenine-N⁶ methyltransferase with the recognition sequence 5'-TCGA-3', combined with AdoMet analogues in the presence of DNA, can be used to alkylate all adenines within the 5'-TCGA-3' sequence. Transalkylation efficiency can then be tested with the complimentary restriction enzyme (TaqI), which has the same recognition sequence, but can only cut unmodified sites within the DNA backbone (Figure 3.1.A). The methyltransferase/cofactor transalkylation efficiency can then be monitored by running the DNA on a gel following transalkylation and restriction, where the DNA fragments are separated by size for analysis.

Knowing the sequence of the alkylated DNA and the location of each site it is possible to predict what a restriction gel will look like at different transalkylation efficiencies and which fragments are expected to appear and disappear. This can be achieved by modelling the introduction of random cuts and plotting the

intensity of each band that is formed and visualising this as a gel (Figure 3.1.B). For initial transalkylation studies the plasmid DNA pUC19 was used which is 2686 bp in length and contains four 5'-TCGA-3' alkylation sites. Due to the palindromic nature of double stranded DNA this equates to a total of eight potential points of alkylation per plasmid.

Furthermore, pUC19 exists as two forms, which are the supercoiled and open circular topological isomers and these appear as two distinct bands in a gel. When the plasmid is completely restricted with TaqI, pUC19 is cut in to 4 distinct fragments. On a standard 1% agarose gel though only 3 of these can be resolved due to the smallest fragment being too short to visualise well. At various transalkylation efficiencies between these two extremes different combinations of the 4 fragments are seen (Figure 3.1.B). As restriction becomes more complete though smaller fragments become more abundant. Comparisons between the experimental and modelled data can then be made to qualitatively estimate transalkylation efficiency.

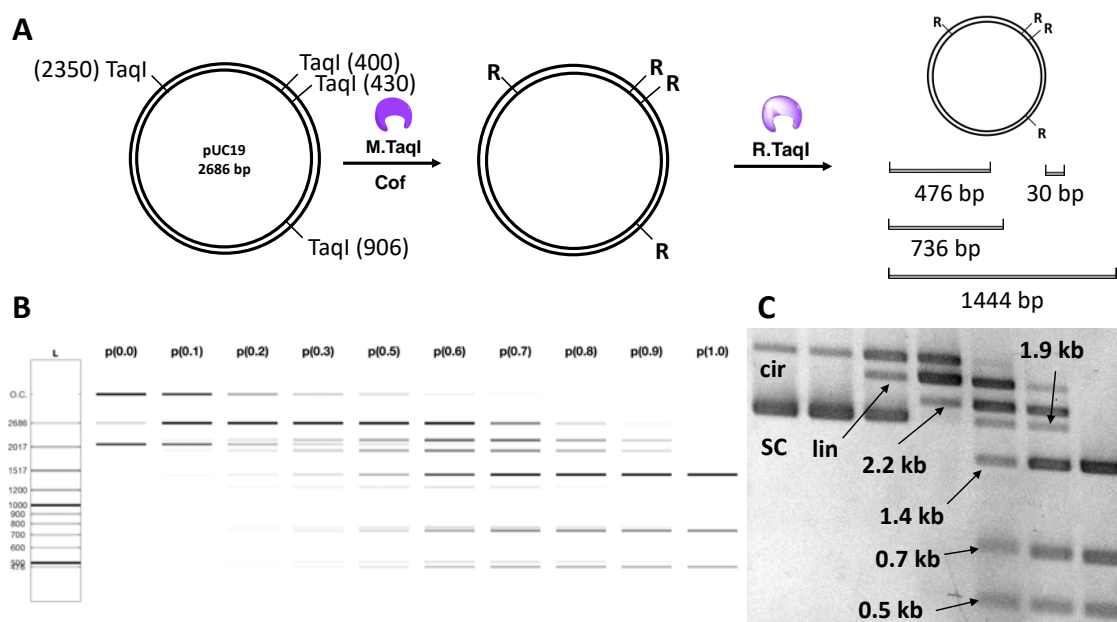


Figure 3.1 A) Schematic representation of DNA transalkylation followed by restriction using the plasmid pUC19 and the methyltransferase M.TaqI. Upon complete restriction the DNA is cut up in to 4 fragments of varying sizes. B) Computational modelling can be used to predict the appearance of gels depending on the transalkylation efficiency of a methyltransferase with any cofactor. As the DNA is labelled less efficiently it is restricted more and more. Partial restriction sees the introduction of larger fragments of DNA. C) An experimental result showing the decrease in efficiency seen when transalkylation pUC19 with M.TaqI.

To have confidence in the results achieved during electrophoresis multiple controls are run alongside the experiment. The controls run for these experiments were:

- AdoMet – Here the naturally occurring cofactor serves as a positive control to prove that the methyltransferase is active and is able to alkylate DNA.
- Methyltransferase control – This control is used to test the effect of the methyltransferase on DNA without the addition of any cofactor. It was necessary as the methyltransferases can contain small quantities of AdoMet due to the way they are made causing latent DNA protection and giving a false positive result. For example, in Figure 3.2, lane 16, a slight band can be seen at ca. 2 kb showing partial DNA protection.

- Negative controls – Run without the addition of methyltransferase or cofactor to test the restriction enzyme activity under the experimental conditions.
- Restriction enzyme control – Each cofactor analogue was tested without the addition of methyltransferase but with the addition of restriction enzyme to highlight if the cofactor impacts restriction activity. In initial protection assays this control was not run. However, later on it was introduced as the acidity of the cofactor storage buffer could sometimes stop the restriction enzyme from being able to effectively cut DNA, leading to a false positive result. Here DNA is combined with the restriction enzyme in the presence of the highest experimental concentration of cofactor used.

One benefit to protection assays is that they can provide useful information on the kinetics of alkylation. This is typically achieved through either an enzyme or cofactor dilution, to identify how much or little enzyme is needed to ensure complete DNA alkylation. It is also further possible from this to gain an understanding of the differences in enzyme turnover kinetics between each of the cofactors. Once this had been established a cofactor dilution could then be run to give information on the dissociation constant for each enzyme and cofactor analogue. This enables a direct comparison to be made between each AdoMet analogue and how it performs in comparison to others.

3.2.1 Cysteine Cofactors

The first cofactors tested by electrophoresis were **AdoCys-6-N₃/NH₂** which had very low transalkylation efficiencies (Figure 3.2). Here each isomer collected by HPLC was used in transalkylation whilst running a 2x enzyme dilution series. As

can be seen from the gel, transalkylation was ineffective. When comparing each isomer at the highest enzyme concentration to the previously predicted restriction maps a maximum efficiency of approximately 20% was seen. Interestingly it appears a greater level of DNA modification can be achieved with the azide compared to the amine cofactor. However, from this result it is not possible to elucidate if this is caused specifically by the change in functional group. Other factors that also may affect modification efficiency are cofactor purity and degradation.

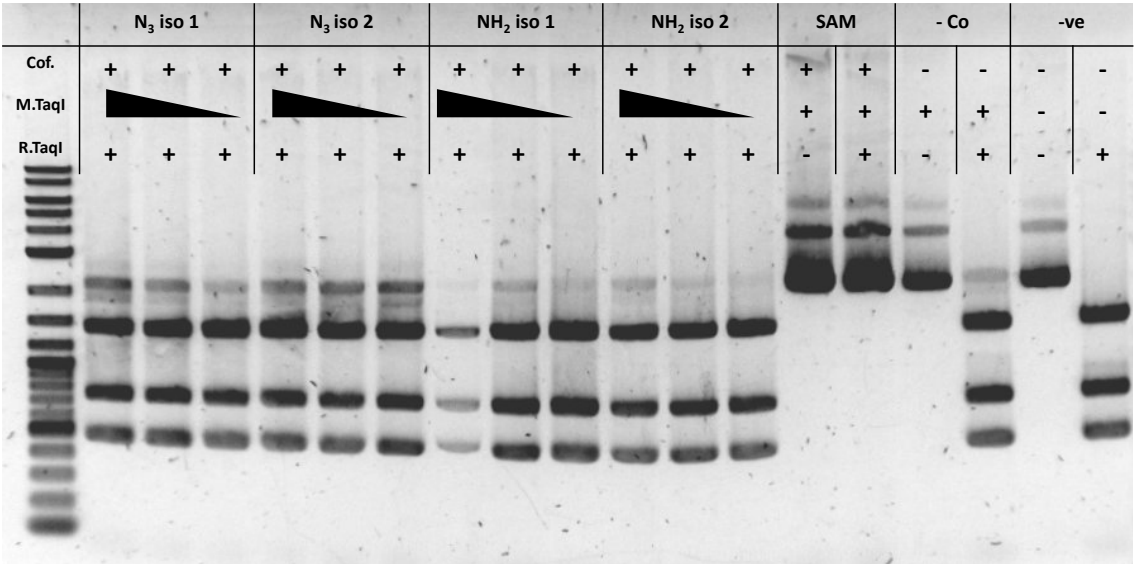


Figure 3.2 Gel electrophoresis comparing the transalkylation efficiency of each isomer collected of both **AdoCys-6-N₃/NH₂** [500μM]. A serial 2x dilution of enzyme M.TaqI, [15 ng/μl] was run for each faction to monitor its effect on transalkylation efficiency. Controls were set up using the natural cofactor AdoMet with further controls to monitor transalkylation without cofactor present and a further negative control to show complete restriction.

In an effort to improve the transalkylation efficiency of both cysteine cofactors, multiple conditions were attempted. For example, lowering the reaction pH, as poor transalkylation efficiency may be caused by *in situ* cofactor degradation. Other factors also explored were increasing/decreasing the concentrations of

both the enzyme and DNA. However, no significant improvements were seen and the transalkylation efficiency remained consistently low. When considering that complete transalkylation could be achieved using the analogous cofactor **AdoHcy-6-N₃** under the same conditions (Section 3.2.2.1) it was decided not to explore these cofactors further. Although they were synthetically cheaper, their poor performance as cofactors for methyltransferases negated these benefits.

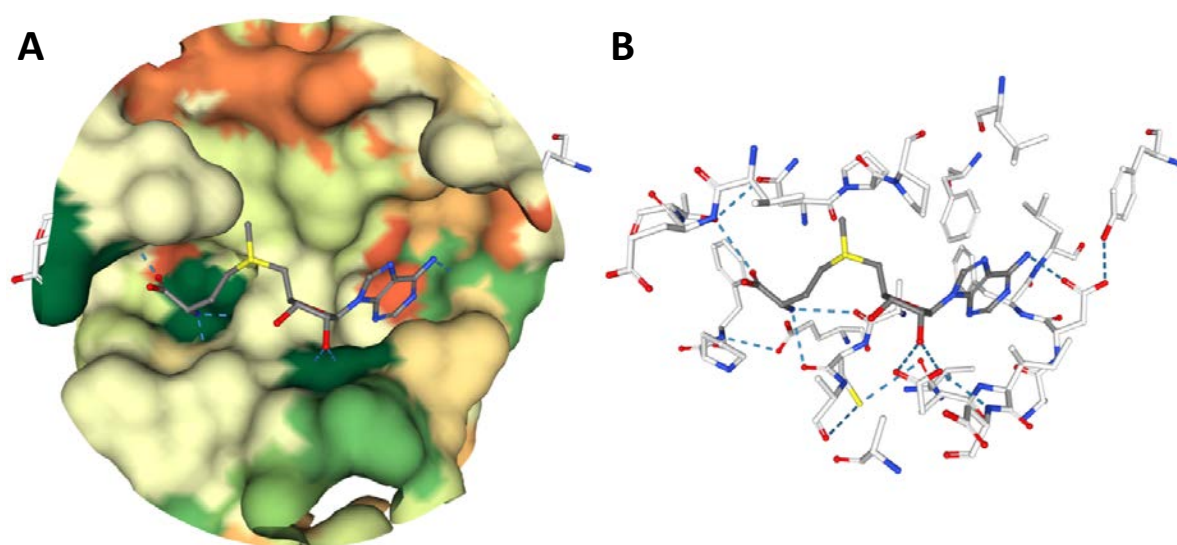


Figure 3.3 Crystal structures of AdoMet binding with *M. TaqI* highlighting the hydrogen bonding network (Blue dotted lines). 4 Hydrogen bonds are formed between AdoMet and the enzyme and 1 salt bridge: Adenine; N1 Phe90-NH, N6 As89-O, Ribose; O2' Glu71-COOH, Homocysteine; COOH Thr23-O, NH₃⁺ Glu45-O Ala47-O Cys48-O. A) Highlighting the enzyme binding pocket and how AdoMet fits within. B) Highlights the amino acid residues within the protein that AdoMet binds with¹⁵³.

In the initial alkylation study of the cofactor **AdoHcy-6-N₃** complete protection of pUC19 was achieved under the same conditions used to test the cysteine cofactors. This indicates changing the amino acid within the cofactor structure causes a significant reduction in the alkylation efficiency with methyltransferases. This is likely due to the elimination of the hydrogen bond and salt bridge formed by homocysteine upon cofactor binding (Figure 3.3)^{153,154}. Combined, these factors explain the significant reduction in activity seen when using cysteine

cofactors. In the future these cofactors will find use with mutant methyltransferases where the binding pocket has been specifically engineered to encourage cysteine binding. This would be of interest for *in vitro* studies as it may allow transalkylation to occur in the presence of AdoMet which is naturally present within cells.

3.2.2 M.TaqI transalkylation with homocysteine cofactors

3.2.2.1 *AdoHcy-6-N₃*

As explained above initial studies of the cofactor **AdoHcy-6-N₃** returned very positive results. After incubation of the cofactor in the presence of pUC19 and M.TaqI complete modification of each labelling site was seen (Figure 3.4). Here also saw the introduction of the new restriction control (lanes 4/8) to further prove labelling was due to alkylation and not incomplete DNA restriction by R.TaqI.

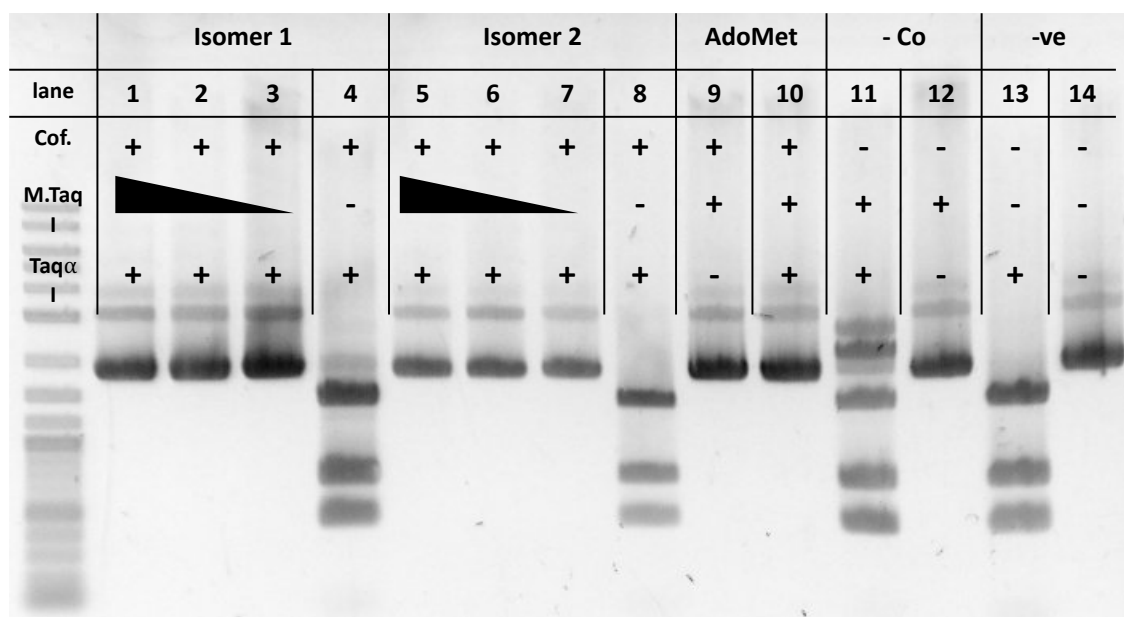


Figure 3.4 Gel electrophoresis for both isomers collected of **AdoHcy-6-N₃** with *M.TaqI* and a 2x enzyme dilution series. Here the restriction control without *M.TaqI* was introduced to ensure observed protection was due to transalkylation. Each isomer was incubated at a concentration of 500 μ M with *M.TaqI* (final concentration: [15 ng/ μ l]) for 1 hour at 50 $^{\circ}$ C followed by addition of the restriction enzyme for a further hour before electrophoresis. Some latent protection of the DNA in the presence of just *M.TaqI* can be seen in the control (11).

Within this gel, it appears complete protection of DNA can be achieved for both diastereoisomers collected, as indicated by the lack of restriction for all experimental lanes, whilst complete DNA fragmentation was seen in the restriction controls (lane 4, 8)¹³¹. However, the effectiveness of each cofactor may have been exaggerated due to the latent DNA protection seen in lane 11 which has the same enzyme concentration as lanes 1 and 5. As can be seen from the dilution series though, complete protection could still be achieved after a 4x enzyme dilution (final concentration: [3.8 ng/ μ l]). Indicating the starting enzyme concentration as unnecessarily high. Reducing enzyme concentration would decrease latent protection, as the quantity of residual AdoMet present would be reduced. Latent protection issues were not explored further in this project as enzyme purification improved greatly in later studies due to the removal

of residual AdoMet during protein production. This is qualitatively confirmed in later gels and quantitatively by HPLC (see section 3.3.1).

Despite the latent protection seen within this gel, another surprising feature was that both isomers showed effective transalkylation. This was unexpected as it has not previously been seen when testing DNA methyltransferases. It was felt the most likely reasons for this was due to either flipping at the pyramidal centre during heating, which can cause isomer interconversion¹⁵². Or due to incomplete separation of each isomer during prep-HPLC, meaning enough of the more favourable isomer was present to cause DNA alkylation. As shown in Chapter 2, cofactor interconversion does not occur with heating, likely due to steric hinderance caused by the bulky groups attached to the sulfonium centre. Therefore, it is unlikely thermal isomer interconversion was the cause of both isomers alkylating DNA. Additionally, it is also unlikely that incomplete purification was the cause of effective alkylation for both isomers as when complete separation is achieved, both isomers still remain effective. This was demonstrated during the testing of **11a**, which showed isomer one to be pure after prep-HPLC and not contain any of the more favourable isomer (Figure 3.6.C). As such, we instead expect the activity seen to be due to a lack of specificity of the enzyme M.TaqI due to its large cofactor binding pocket. The only other reported example found where methyltransferases also show this apparent lack of selectivity was in the alkylation of coumarins for the synthesis of antibiotics showing it to be quite a unique characteristic of M.TaqI¹⁵⁰. This is potentially important as if both isomers are effective cofactors for M.TaqI, purification by prep-HPLC is not necessary, which can simplify cofactor synthesis when used

with this methyltransferase. Unfortunately though, this effect was not seen in later studies with the enzyme M.Mpel.

3.2.2.2 *Schiff base cofactors AdoHcy-8-Ox/Hyd*

After the successful alkylation using **AdoHcy-6-N₃** work moved to testing the Schiff base cofactors. As they were new cofactors it was decided to test their efficiency after each stage of the coupling in the protected, deprotected and capped forms as they were synthesised.

The initial gel run for these cofactors was of the boc-protected hydroxylamine cofactor (**10a**). After the initial coupling, purification was achieved by prep-HPLC, followed by DNA transalkylation (Figure 3.5). An enzyme dilution was again used to assess any differences that may occur in the kinetics between **10a** and **AdoHcy-6-N₃**, which were not found. Again, at the highest enzyme concentration [15 ng/μl], which was the same as used for **AdoHcy-6-N₃**, some latent DNA protection occurred. Having found that there was not a significant difference in the behaviour of either cofactor, the enzyme concentration was reduced for all future protection assays to minimise the effects of residual AdoMet. Here both suspected isomers again showed complete protection at the highest enzyme concentration of [15 ng/μl].

To further probe the kinetics between each HPLC fraction, in this gel the enzyme was diluted to lower concentrations where isomer 1 was less effective in the alkylation of DNA. This highlights the alkylation kinetics of the two suspected isomers do vary. This indicates isomer one likely contains the less effective *R,S* form, whilst isomer two is the *S,S* form which has the same structural confirmation as natural AdoMet.

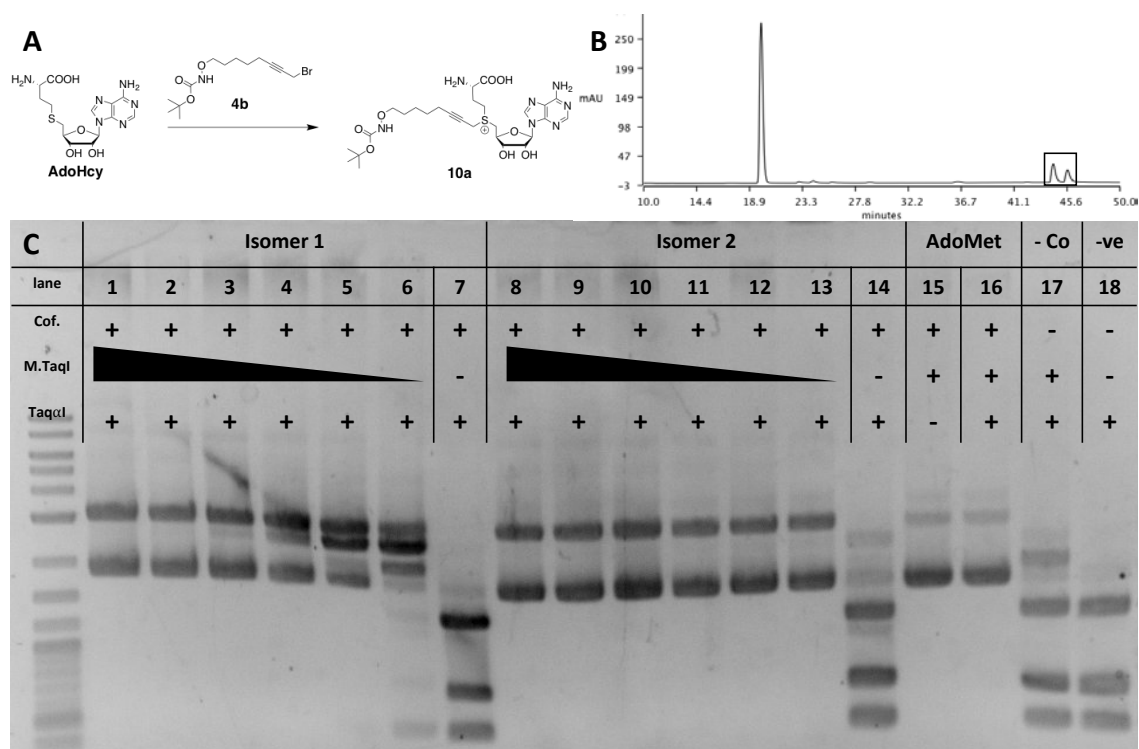


Figure 3.5 A) Synthetic pathway adopted for the synthesis of **10a**. Purification was achieved using 20 mM ammonium acetate buffer gradient with methanol. Retention times: isomer 1 = 45.0 minutes, isomer 2 = 47 minutes. B) prep-HPLC collected for the purification of **10a**. The two isomers were collected and identified by MS before being taken forward for transalkylation. C) Gel electrophoresis of each isomer of **10a** [500 μ M] after purification. A serial 2x enzyme dilution was run using M.TaqI, highest concentration: [15 ng/ μ l].

Following success of **10a** in the alkylation of DNA, deprotection was then tested prior to prep-HPLC purification using neat TFA. The deprotected cofactor was then worked up by extracting with diethyl ether and water followed by purification of the aqueous layer. After purification the deprotected cofactors were then tested in the transalkylation of DNA in the same way as previously with M.TaqI. Complete protection was again achieved at the highest enzyme concentration ([15 ng/ μ l]) and isomer 1 remained as the less effective cofactor analogue, as shown by its reduced transalkylation efficiency as the enzyme concentration was reduced to [3.8 ng/ μ l] (Figure 3.6).

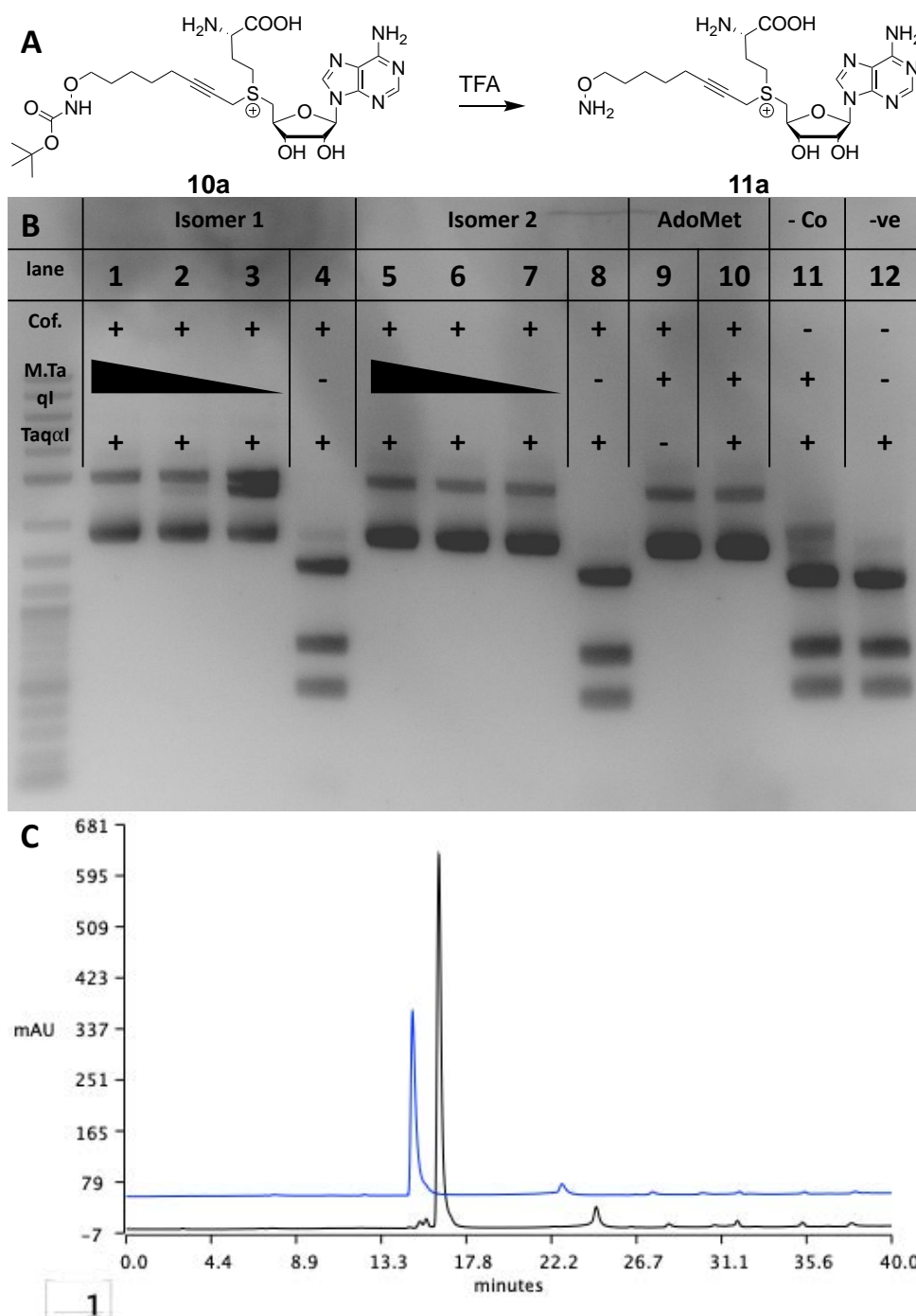


Figure 3.6 A) Deprotection was achieved directly after coupling using neat TFA. After deprotection purification was achieved by prep-HPLC. Each isomer was again isolated and tested individually as a methyltransferase cofactor. B) Gel electrophoresis of the deprotected cofactor **11a** [500 μ M], M.TaqI [15 ng/ μ l]. C) analytical HPLC chromatograms of each isomer sample after prep-HPLC showing separation of the two diastereoisomers, blue = isomer 1, black = isomer 2.

Interestingly, alkylation with the deprotected cofactor **11a** shows one of the main drawbacks of protection assays to measure transalkylation efficiency. Previously

in Chapter 2 we showed **11a** readily reacts with any trace impurities like acetone (Section 1.4.4). Something protection assays cannot identify as they give no information about the type of modification made to DNA, which, in this instance is likely a combination of the free and reacted hydroxylamine. This is a problem as to take cofactors forward to biomolecular applications we need confidence in the modifications they introduce to DNA which protection assays alone cannot provide. This led to the decision to 'cap' the free hydroxylamine and hydrazide prior to work-up to preserve the functionality introduced during alkylation. Furthermore, it also prompted the development of quantitative alkylation studies using LCMS (section 2.3.1).

After proof by LCMS that capping of the deprotected hydroxylamine had been successful and improved cofactor purity, the final cofactor was again tested by gel electrophoresis. Initially it was thought that by extending the alkyl side chain of the cofactor, a decrease in alkylation efficiency may be seen. However, this was not the case and efficiencies were the same as previously (Figure 3.7). We expected this may occur as the alkyl side chain may be too big for the cofactor to fit effectively in the enzyme binding pocket.

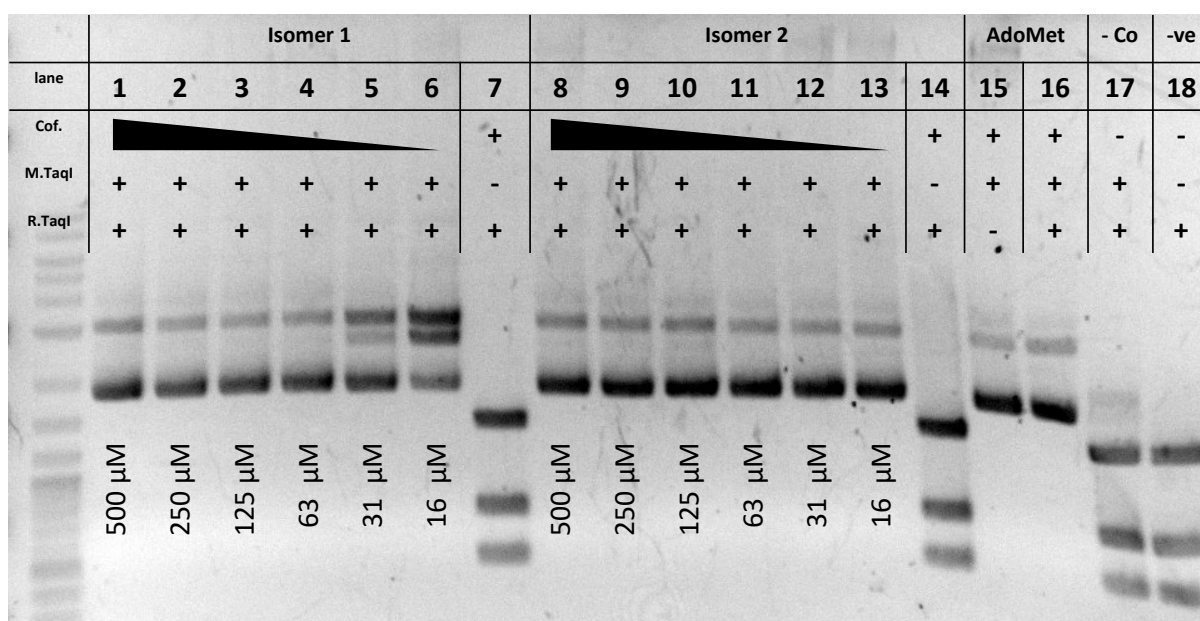


Figure 3.7 Gel electrophoresis of the capped hydroxylamine cofactor (**AdoHcy-8-Ox**). A serial 2x cofactor dilution was run, showing isomer 1 to be the less effective diastereoisomer (M.TaqI [7.5 ng/μl]). Restriction controls (lane 7, 14) were run at the highest cofactor concentration of 500 μM.

After probing the enzyme concentration and its effect on alkylation efficiency, and identifying the effective quantity of enzyme required to ensure complete transalkylation with both isomers. In the transalkylation of **AdoHcy-8-Ox** a cofactor dilution series was used instead. Figure 3.7 shows that for isomer 2 complete transalkylation was observed with a cofactor concentration of 16 μM. This concentration equates to only a 100-fold excess of cofactor per transalkylation site within pUC19. This is significant as a greater impact on cofactor analogue binding was expected as the binding pocket is a conserved structure and the change from a methyl to alkyl group is large. Previously high cofactor concentrations have been used in the alkylation of DNA, but this appears unnecessary allowing for reduced cofactor concentrations to be used in transalkylation. From this result it is also possible to see that transalkylation with isomer 2 is at least four times more effective in the alkylation of DNA when comparing the lowest concentration required for complete modification.

Following the successful synthesis of the oxime cofactor, the same process was reproduced for the hydrazone. Here a serial 3x cofactor dilution was run to further explore the concentrations at which transalkylation began to fail (Figure 3.8). As can be seen from the gel, isomer 2 still showed complete transalkylation at all concentrations, again highlighting it as the effective cofactor isomer to be taken forward for future studies.

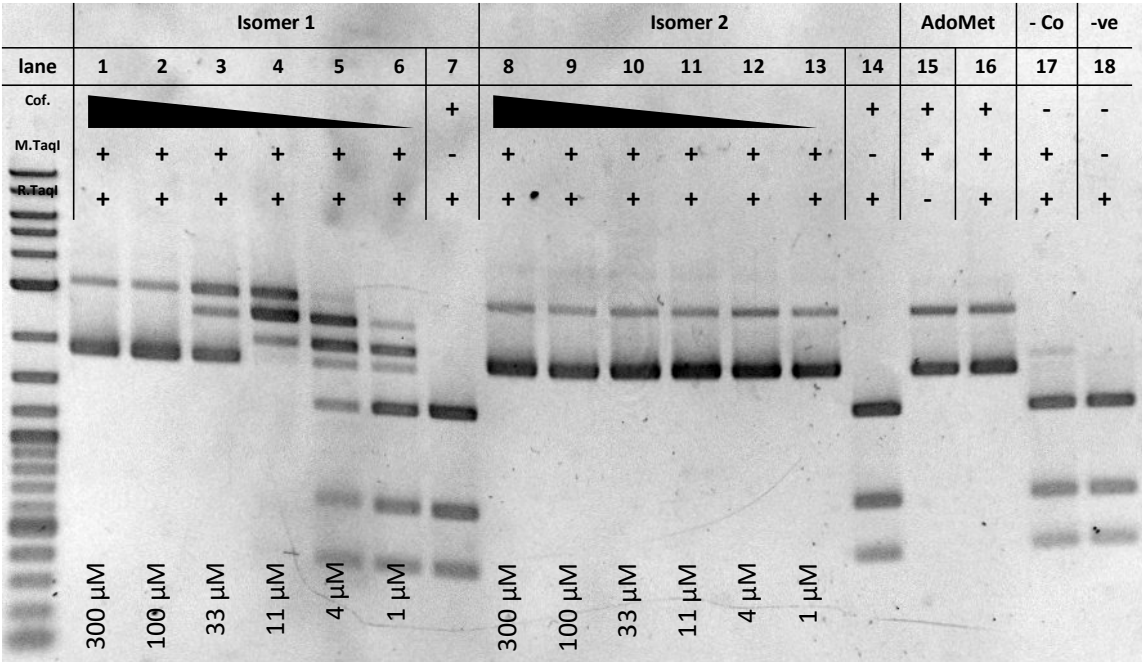


Figure 3.8 Gel electrophoresis of both isomers of **AdoHcy-8-Hyd**. A serial 3x cofactor dilution series was run from 300-1.2 μM (M.TaqI [7.5 ng/μl]) to highlight and differences in transalkylation efficiencies.

When analysing the gels of both **AdoHcy-8-Ox/Hyd** it is possible to see that both cofactors behave similarly with M.TaqI. For example, when comparing the transalkylation efficiencies of isomer 1, both cofactors begin to fail at approximately 30 μM. When comparing the data collected for these isomers it is possible to gauge an understanding of the differences in the way the cofactors bind and behave with methyltransferases. This can be calculated as an apparent binding constant (K_x) by plotting alkylation efficiency vs [cofactor] and comparing the concentration required to achieve 50% alkylation (Figure 3.9). Here at the

concentrations tested for **AdoHcy-8-Ox** no significant variation in alkylation efficiency, compared to **AdoHcy-8-Hyd**, was seen, and for **AdoHcy-8-Hyd** a K_x of approximately 5.5 μM can be calculated. Unfortunately for **AdoHcy-8-Ox** no K_x could be calculated as the data did not go to low enough concentrations, however an approximate K_x of between 6-8 μM can be predicted. Similarly, no K_x can be calculated for isomer 2 for each cofactor analogue as restriction did not take place at any of the concentrations explored.

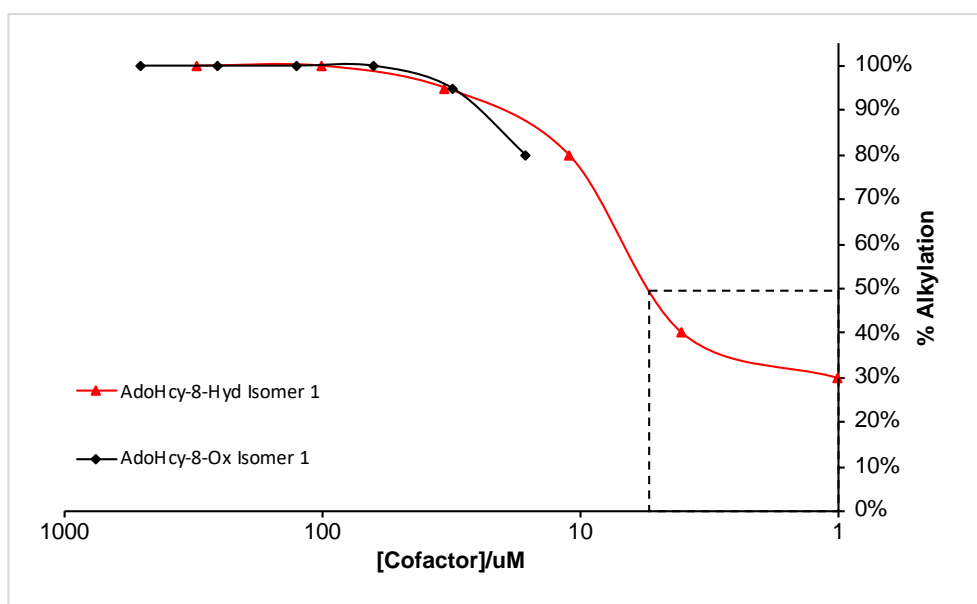


Figure 3.9 Plot of percentage alkylation vs. [Cofactor] to determine approximate K_x of both cofactors **AdoHcy-8-Ox/Hyd**. Percentage alkylation was calculated by comparison against restriction plot from Figure 3.1.B for *M.TaqI*. Approximate K_x **AdoHcy-8-Hyd** isomer 1 = 5.5 μM .

3.2.2.3 Cofactor efficiency comparison

In an effort to develop a more complete understanding around the enzyme kinetics with each cofactor analogue, and to compare each cofactor against **AdoMet**, one final gel was run with *M.TaqI*. This was run to compare the transalkylation efficiency of the *S,S* isomer of all cofactors to **AdoMet**. Here a 3x cofactor dilution series was run from 10-0.1 μM (Figure 3.10). A lower limit of 0.1

μM was chosen for transalkylation as all cofactors should begin to fail. This is because 1 equivalent of cofactor for each transalkylation site equated to a final concentration of $0.14\ \mu\text{M}$ for 300 ng pUC19, assuming 8 labelling sites.

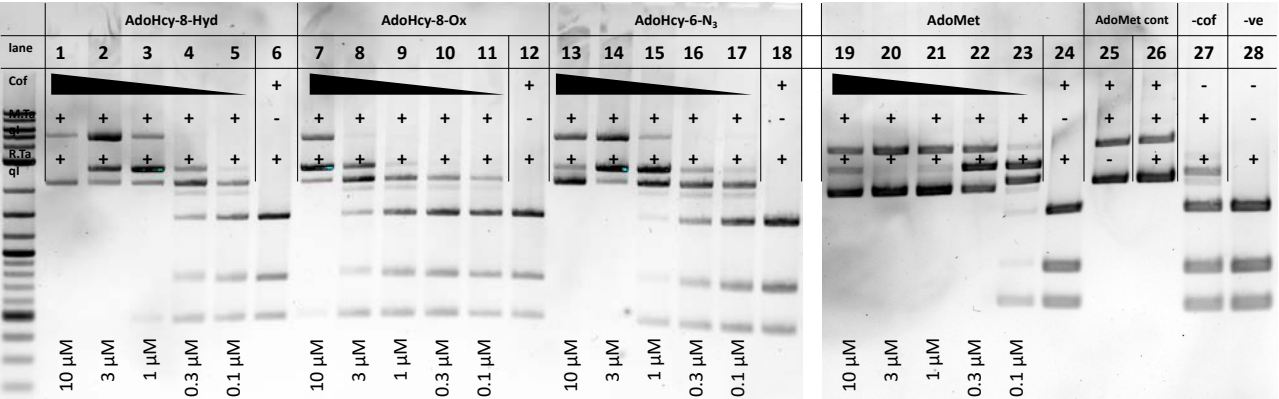


Figure 3.10 Gel electrophoresis showing a cofactor dilution from 10-0.1 μM for **AdoHcy-8-Hyd**, **AdoHcy-8-Ox**, **AdoHcy-6-N₃**, and **AdoMet** with *M.TaqI* [7.5 ng/ μL] and pUC19 (300 ng). Transalkylation was set-up in tandem and samples were spread across 2 gels run in parallel.

Through plotting and subsequent comparison of each cofactor analogue’s transalkylation efficiency it can be estimated that each analogue is at least 7x less efficient (Figure 3.11). This can be estimated using the predicted restriction map shown previously (Figure 3.1) and plotting the change in alkylation over time. Although only an estimation this data gives good insight into each cofactor analogues behaviour with *M.TaqI* compared to AdoMet. Similarly, this data can be used to compare the efficiency of the previously plotted, less favourable isomer of **AdoHcy-8-Hyd**, which shows transalkylation with the more favourable isomer to be approximately three times more efficient.

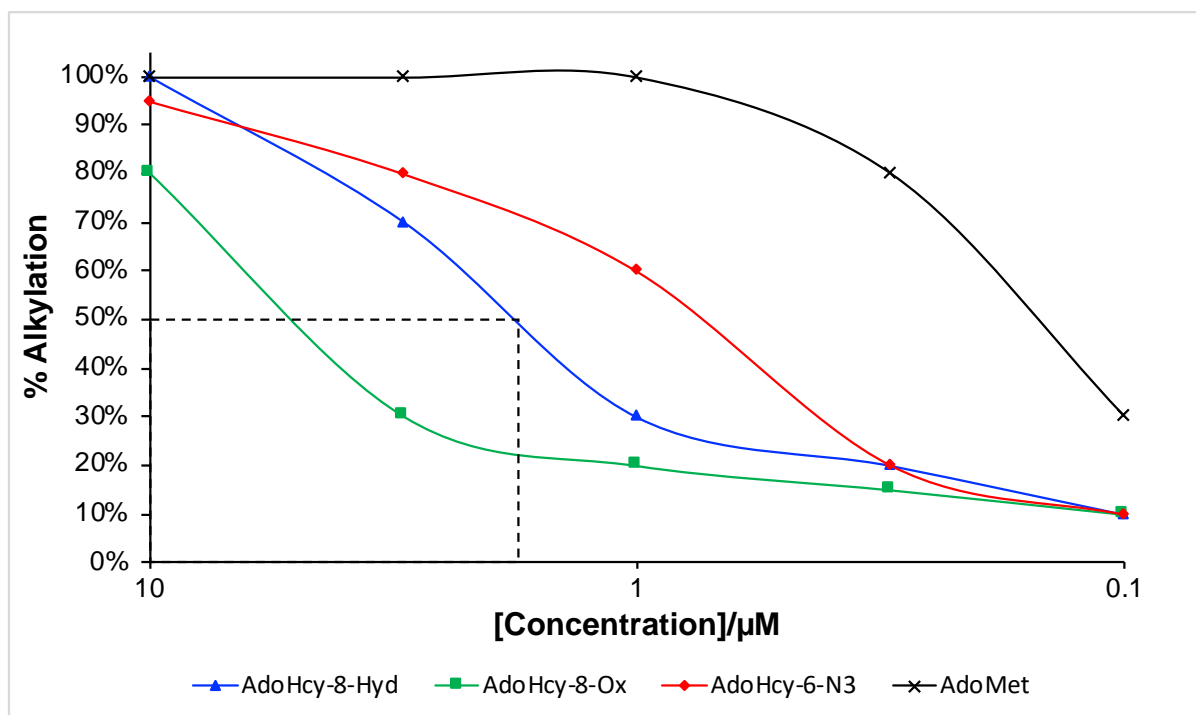


Figure 3.11 A plot of alkylation efficiency against cofactor concentration for each cofactor analogue. Alkylation efficiency was calculated by comparison with the restriction map for *M.TaqI*. K_X can then be estimated for each cofactor analogue by looking at the point at which 50% alkylation occurs. **AdoMet** $K_X = 0.15 \mu\text{M}$, **AdoHcy-6-N₃** $K_X = 0.7 \mu\text{M}$, **AdoHcy-8-Hyd** $K_X = 1.8 \mu\text{M}$, **AdoHcy-8-Ox** $K_X = 5.9 \mu\text{M}$.

3.2.3 M.MpeI DNA transalkylation

Following the successful testing of all cofactors with *M.TaqI* work moved to a new methyltransferase of interest, *M.MpeI*, which can be used as a tool to study epigenetics. *M.MpeI* is a cytosine methyltransferase with the recognition sequence of CG. The variant used in these experiments was a mutant containing modifications at (Q136A, N347A) to improve binding of the alkyl containing AdoMet analogues. Interestingly *M.MpeI* is an orphan methyltransferase meaning it does not have a complimentary restriction enzyme. Therefore, to monitor transalkylation with protection assays it is necessary to find a restriction enzyme that is directly affected by cytosine methylation but with a different recognition sequence. Two restriction enzymes were highlighted which meet

these criteria and were: HhaI (GCG'C) and HaeII (RGCGC'Y, where R = purine, Y = pyrimidine).

Within the plasmid pUC19, M.MpeI has 173 CpG transalkylation sites, meaning 346 potential modifications. The recognition site of HhaI is GCG'C which occurs 17 times within pUC19 and restriction is stopped by methylation of the central cytosine base. It was hoped, using this restriction enzyme a clear picture of M.MpeI transalkylation efficiency could be seen. Prior to transalkylation the restriction pattern was modelled in the same way as for TaqαI (Figure 3.12), which found restriction efficiency would be far more complex to analyse accurately due to the large quantity of fragments that can be formed. However, it would still be possible to get a qualitative understanding of M.MpeI activity.

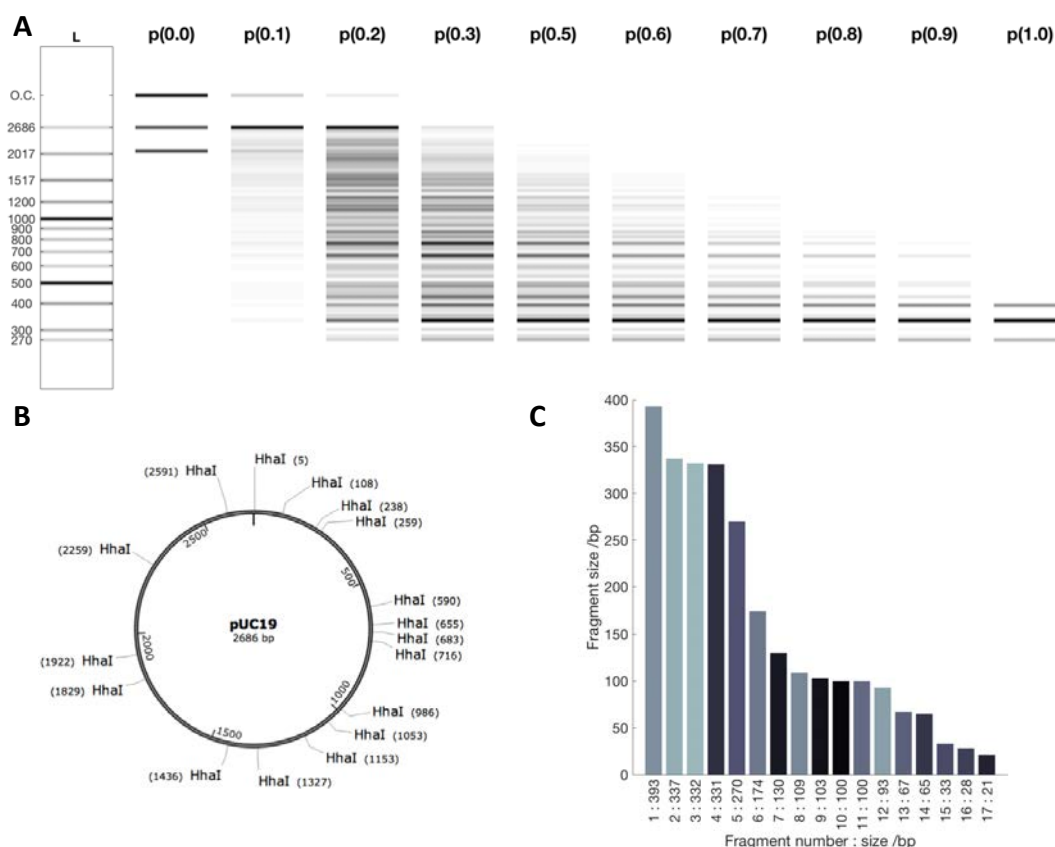


Figure 3.12 A) Modelled *HhaI* restriction plot of *pUC19* at varying transalkylation densities. B) *pUC19* plasmid map highlighting the locations of each *HhaI* restriction site. C) A plot of fragment size after complete restriction of *pUC19* with *HhaI*. Upon complete restriction two main bands are expected at approximately 100 and 350 bp.

The cofactor of most interest for use with M.MpeI was **AdoHcy-8-Hyd**, due to its ability to undergo rewriteable transalkylation (to be discussed in more detail later) which we hoped could be used in the capture and release of DNA. Through taking advantage of the Schiff base, it would be possible to first capture DNA by coupling of the terminal azide, followed by its release with cleavage of the Schiff base. As such, all gels for this enzyme were performed with this cofactor. It was then assumed that all cofactors would behave in a similar way, as seen for M.TaqI.

Due to the high density of transalkylation sites of M.Mpel within pUC19 it was necessary to add significantly more enzyme than was used for M.TaqI (ca. 500x increase) in order to facilitate high alkylation efficiencies. As such, after initial transalkylation with the cofactor and enzyme, purification of the DNA was carried out to remove the methyltransferase before restriction. This was to ensure complete removal of the methyltransferase from the DNA backbone and prevent interference of the methyltransferase with restriction. After restriction, the DNA was run on a gel to analyse the result (Figure 3.13)

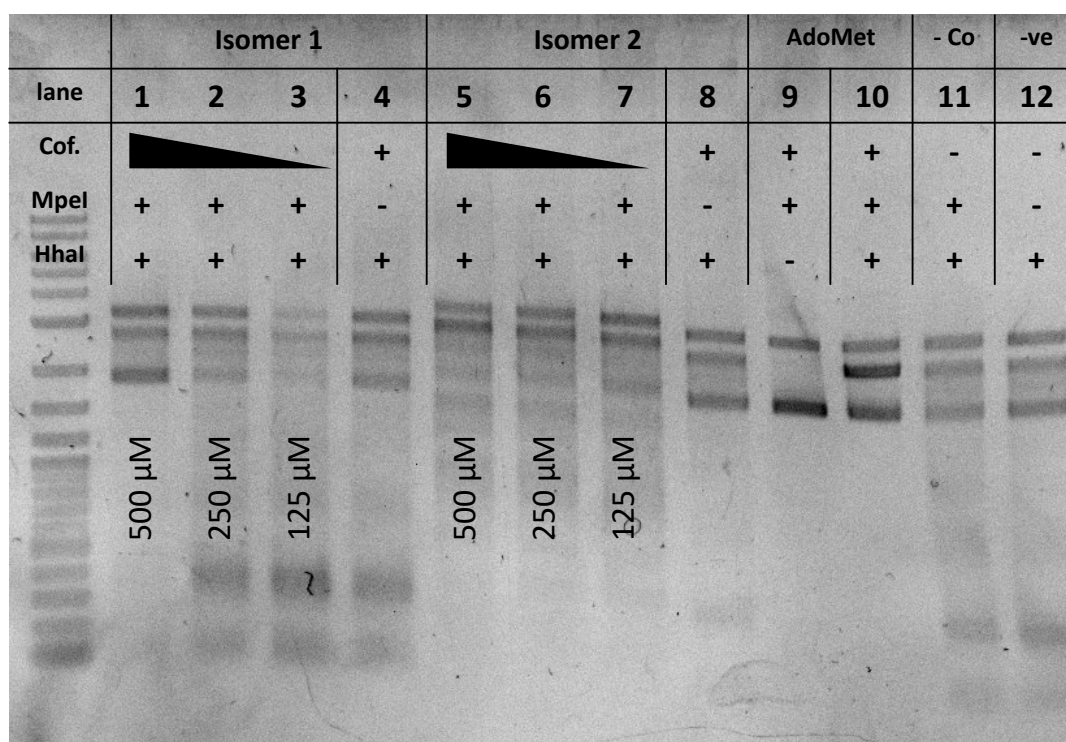


Figure 3.13 Gel electrophoresis of **AdoHcy-8-Hyd** and M.Mpel [8.5 μg/μl] in the transalkylation of pUC19. A serial 2x cofactor dilution series was run for each isomer. Transalkylation was run at 37 °C for 1 hour followed by individual purification of samples on silica columns before restriction with Hhal at 37 °C for a further hour.

In initial gels using Hhal restriction problems were found because of the inconsistent performance of this enzyme. As can be seen in lanes 4, 8 and 12,

the restriction controls which did not show the complete restriction. Due to the failure of these controls it is difficult to gather any information reliably from the lanes of interest. It was attempted to improve the performance of HhaI by switching the buffer used during restriction and increasing the quantity of restriction enzyme up to two times. But, neither change improved the results seen and HhaI continued to inconsistently restrict each sample. This meant a new restriction enzyme was required as no reliable information into alkylation efficiency could be obtained. For this HaeII was chosen, which has the recognition site of RGCGCY (R= purine (A/G), Y = pyrimidine (C/T)). One benefit of HaeII is that it only has 3 recognition sites within pUC19 meaning that monitoring the transalkylation reaction is simpler (Figure 3.14.A). However, analysis is slightly complicated as supercoiled pUC19 appears at a very similar retention to the main 1.9 kb fragment formed after complete restriction. This issue could be minimised though through running extended gels which led to greater separation between fragments.

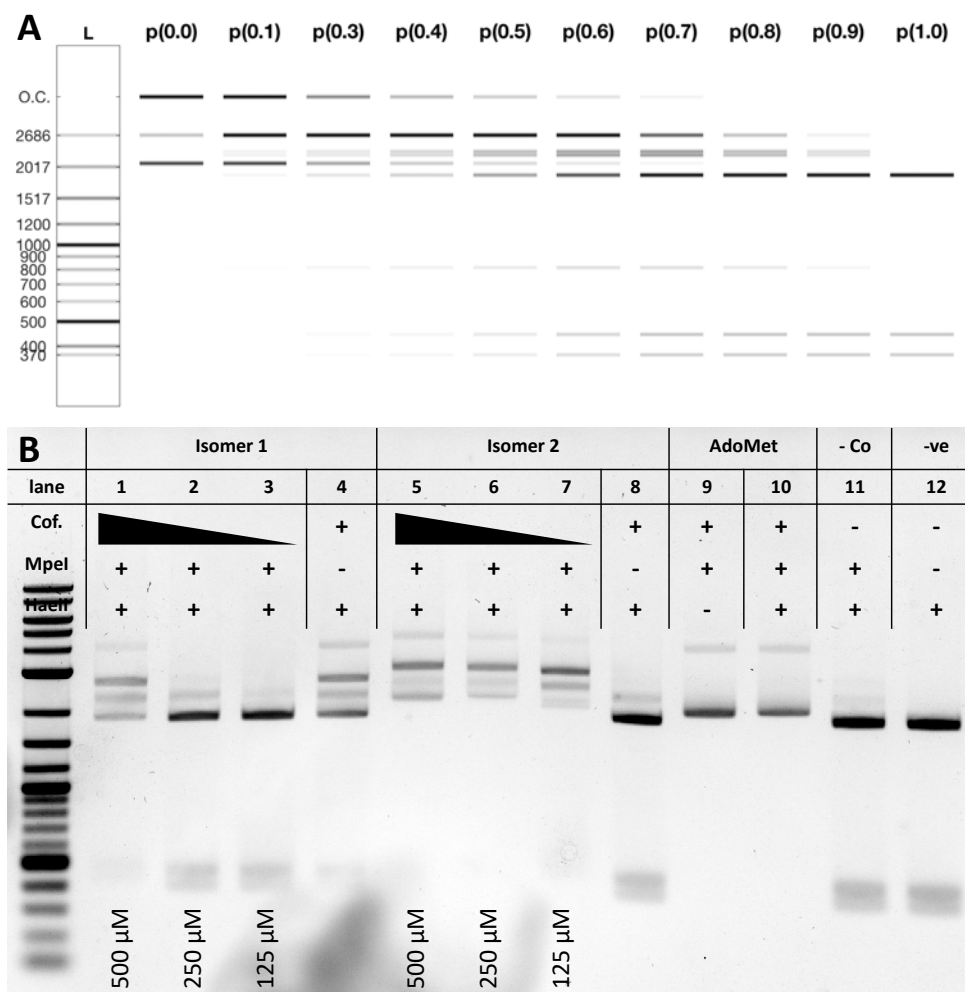


Figure 3.14 A) predicted restriction map of Haell at varying transalkylation efficiencies for pUC19. B) Gel electrophoresis of **AdoHcy-8-Hyd** after transalkylation with *M.Mpel* [8.5 μ g/ μ l] and restriction with Haell with a serial 2x cofactor dilution series.

Restriction with Haell appeared effective in the cutting of DNA as expected and enabled *M.Mpel* transalkylation to be monitored successfully (Figure 3.14.B). As can be seen from lane 12, complete fragmentation of the DNA was seen, proving Haell was digesting DNA. Furthermore, lane 11 also showed complete restriction, indicating limited latent protection was caused by the methyltransferase, despite the high concentrations of enzyme used. Unfortunately, the cofactor control in lane 4 failed, as did lane 1 which had the same overall cofactor concentration. This indicates high cofactor concentrations

caused inhibition of the restriction enzyme for this isomer. However, restriction could be achieved at the lower cofactor concentrations in lanes 2/3 which showed no transalkylation with the cofactor. From this we conclude that isomer one inhibits HaeII restriction and likely shows no activity as a cofactor for M.MpeI. This again fits with data collected previously for M.TaqI showing isomer one to be the less effective R,S diastereomer. Yet in this case, isomer one appears not to successfully transalkylate DNA with M.MpeI, better fitting with previously reported methyltransferases like M.HhaI¹³¹. This highlights M.TaqI's ability to alkylate with both isomers to be quite unique.

Isomer two showed clear transalkylation with efficiencies of approximately 90%. There was also a clear dilution effect seen with decreasing cofactor concentration. This shows these cofactor analogues can successfully be used to alkylate DNA with two distinct methyltransferases. Hopefully in the future the versatility of these cofactors will be further demonstrated through testing of each cofactor with more DNA methyltransferases and also expanding to protein and RNA methyltransferases.

3.3 Oligonucleotide HPLC

As discussed previously, although gel electrophoresis is a straightforward test for methyltransferase activity with a cofactor analogue, alone it does not provide any information about the type of modification that has been made to DNA. In order to further understand the structure of the cofactors developed and how they interact with methyltransferases we created a new technique to monitor transalkylation directly by HPLC. For this, two types of DNA oligonucleotide

(Oligo) were used, the first was a 14 bp double stranded oligo with a TCGA motif in the centre¹⁵⁴. The second type of oligo used were hairpins with either 1, 2, or 3 TCGA recognition sites.

3.3.1 14 bp oligonucleotide transalkylation

To first test the transalkylation of each of the successful AdoMet analogues by HPLC a 14 bp oligo was used. This oligo was taken from the literature and chosen because of its size, melting temperature and proven success as a substrate for methylation by M.TaqI¹⁵⁴. Size was an important consideration in initial HPLC experiments as the oligo needed to be as short as possible to maximise the chances of monitoring modifications to DNA by HPLC through changes in hydrophobicity and retention times. The calculated melting temperature for this oligo is 57.7 °C, meaning during transalkylation with M.TaqI, at 50 °C, the oligo remains double stranded; and then during HPLC, separation of the individual strands can be achieved by melting the DNA at 60 °C. During HPLC, fractions could then be collected and further characterised by MS. It was hoped that by HPLC it would be possible to monitor transalkylation efficiency and see shifts in DNA retention relating to the modifications made. Subsequent analysis of these peaks by MS can then be used to further confirm the identity of alkylations introduced to DNA.

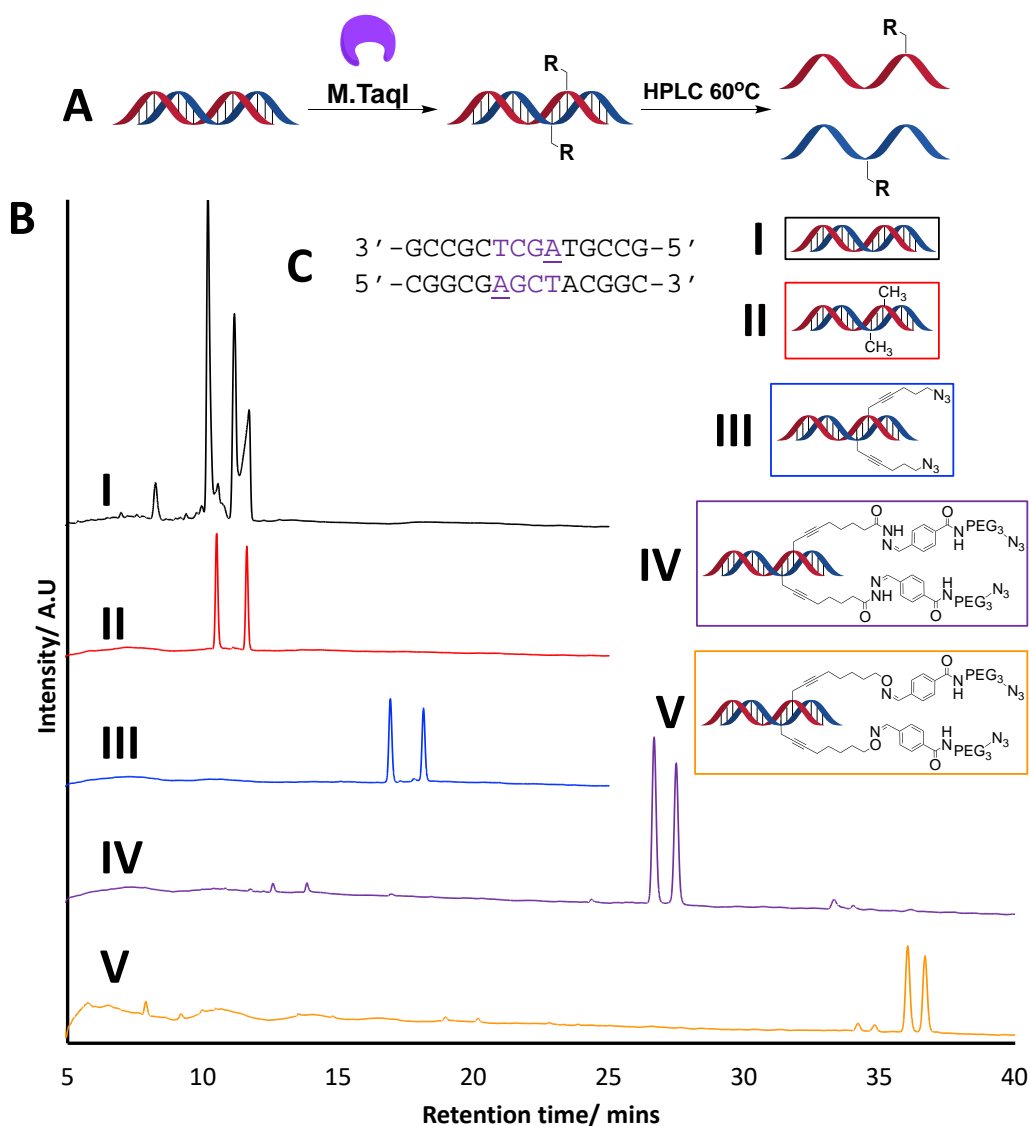


Figure 3.15 A) Transalkylation of the 14 bp oligo was achieved using M.TaqI and either AdoMet or one of the successful AdoMet analogues. After transalkylation DNA was purified on silica columns before injection into the HPLC. B) HPLC traces of each analogue tested after transalkylation. All HPLC was run at 60 °C using 0.1 M triethylammonium acetate with an acetonitrile gradient. C) DNA sequence of both the forward and reverse strand of the 14 bp oligonucleotide highlighting the M.TaqI recognition site. Raw chromatograms in supplementary information figures S1-10.

Before testing each analogue, an experiment was first run using the natural cofactor AdoMet to see if transalkylation could be monitored by both HPLC and MS (Figure 3.15.B.II). It was also necessary to test if complete transalkylation could be achieved on a comparatively large scale with a much higher density of

transalkylation sites compared to what was used for pUC19. Initial methylation showed transalkylation of the oligo to go to completion along with a small shift in retention of both strands, highlighting its success. Final confirmation of transalkylation then came from collection of each HPLC fraction followed by MS analysis showing an increase of 15 to the molecular weight of each strand as expected (sup. Fig. 3,4)

Following methylation, the oligo was then used to test modification with the synthetic cofactors **AdoHcy-6-N₃**, **AdoHcy-8-Hyd** and **AdoHcy-8-Ox** (Figure 3.15). For each cofactor the putative S,S isomer that showed the highest activity during electrophoresis was used. As can be seen in the HPLC traces there is a complete removal of unlabelled oligo and a shift for all samples in retention time, indicating alkylation went to completion. Figure 3.15 shows a clear link between the size of the expected modification and retention time. Interestingly, for transalkylation with the Schiff base cofactors some minor peaks can also be seen between 12-13 and 19-21 minutes for the hydrazone and oxime respectively. These likely occur from hydrolysis of the hydrazone or oxime bond during HPLC and is something that needed to be accounted for in future reversibility studies. The suspected free hydrazide has since been confirmed during reversibility studies by MS and gave confidence for the use of these cofactors in rewriting DNA (section 4.1).

Further to proving that each cofactor analogue alkylated DNA and introduced the anticipated modification, this experiment also conclusively showed transalkylation with AdoMet analogues occurred on both strands of the DNA, due

to the palindromic nature of the TCGA site. Although this is known to be the case in nature for methylation, it cannot be guaranteed the same would occur for alkylation with more complex functionalities⁸³. Instead, asymmetric alkylation may have predominated as the modified adenine may interfere with methyltransferase binding (Figure 3.16.C). However, this HPLC data conclusively shows this not to be the case.

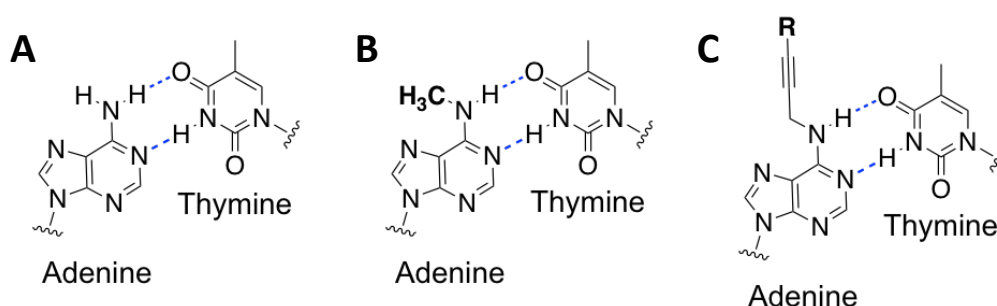



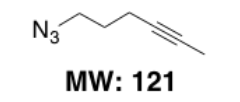
Figure 3.16 A) Watson and Crick hydrogen bonding between Adenine and Thymine. B) After methylation there is the introduction of a methyl group, however hydrogen bonding is not dramatically affected. C) introduction of more complex functionality may affect the overall hydrogen bonding network of DNA at the modification site because of sterics caused by the new group.

After HPLC of each reaction the fractions collected were concentrated and submitted for MS. Prior to analysis the expected molecular weight of each modification to be introduced was calculated which could then be used to calculate the expected molecular weight for each strand (Figure 3.17.A). Calculated masses were then compared to the observed molecular weights of each fraction which positively indicated alkylation had occurred in each case. Combined, the electrophoresis, HPLC and MS data prove each AdoMet analogue modifies DNA as designed and can be used in the alkylation of DNA.

A



MW: 14



MW: 121

B

| Sample | Observed weight | |
|-------------------------------|-----------------|-------------|
| | Strand 1 | Strand 2 |
| Unlabelled | 4289 | 4240 |
| Methylated | 4303 (+14) | 4254 (+14) |
| AdoHcy-6-N₃ | 4410 (+121) | 4361 (+121) |
| AdoHcy-8-Hyd | 4772 (+483) | 4724 (+484) |
| AdoHcy-8-Ox | 4760 (+471) | 4710 (+470) |

Figure 3.17 A) Mass calculations for the expected increase in molecular weight with each cofactor analogue.
 B) A table plotting the observed masses for each strand of the 14 bp oligo after transalkylation and HPLC.
 For samples which appeared as the triply charged species the total mass was calculated by: $((M^3 \times 3) + 3)$.
 MS data supplementary figures S1-10

3.3.2 Hairpin transalkylation studies

After the successful purification and analysis of modified oligos by HPLC, work moved towards monitoring more complex transalkylation experiments to help gain a better understanding of how alkylation occurs. For this, hairpin oligos with variable numbers of transalkylation sites were chosen (Figure 3.18). Hairpins were of specific interest for this project as it was hoped HPLC could be used as a method for purification, separating fully and partially labelled DNA. If successful, methyltransferase-directed transalkylation could then be used as a tool for the development of fluorescent DNA probes with multiple fluorophores. This will be of benefit in microscopy through the creation of brighter probes, which

will be easier to see and have decreased detection limits. Probe design and transalkylation was carried out by Ashleigh Rushton.



Figure 3.18 Design and sequences of hairpin oligonucleotides used for HPLC. Each consisted of a 32 bp (red) single stranded recognition sequence followed by either 1, 2 or 3 TCGA sites (green) with 4 bp spacing regions either side.

The design of all hairpins was similar, with a 32 bp single stranded detection region, followed by a double stranded section containing either 1, 2, or 3 TCGA sites with 4 thymine residues causing the hairpin loop (Figure 3.18). Prior to transalkylation each hairpin was formed by heating the DNA to 85 °C for 5 minutes and then slowly cooling to room temperature overnight. All oligos were initially labelled using the same quantity of enzyme and cofactor to monitor the effect of increasing the number of transalkylation sites on efficiency.

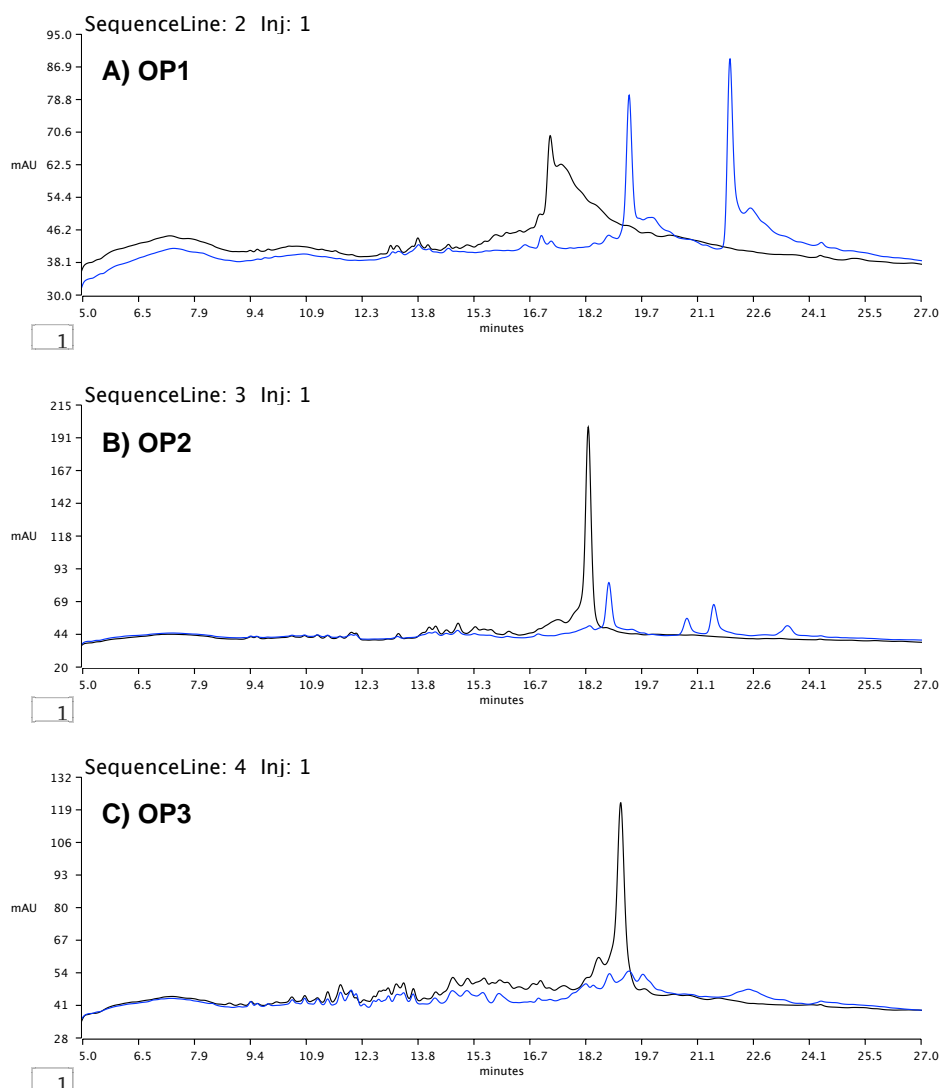


Figure 3.19 HPLC traces collected after transalkylation of hairpin DNA probes with **AdoHcy-6-N₃**. Transalkylation was carried out in the same way as for the 14 bp oligo with *M.TaqI* and cofactor for 1 hour and 50 °C followed by proteinase K treatment and purification using silica columns before injection in to the HPLC. Successful MS data collected for OP1 in supplementary figure S3.12, 3.13. Black traces show the unlabelled control for each hairpin and blue shows modified DNA after the initial transalkylation attempt.

Initial alkylation studies for each hairpin were carried out using the cofactor **Ado-Hcy-6-N₃** in the same way as for the 14 bp oligo. Alkylation was achieved by first incubating the annealed DNA with *M.TaqI* and cofactor. This was followed by proteinase K treatment and subsequent purification. DNA was purified prior to HPLC to remove any remaining cofactor and denatured protein that could complicate HPLC. As can be seen in the HPLC traces, alkylation of each hairpin

saw a complete removal of unlabelled DNA from the sample (Figure 3.19). For both hairpins OP1 and OP2 there was then the clear introduction of either two or four new peaks after transalkylation. This fit with what was expected as for each TCGA present within the double stranded region there were two possible transalkylation sites due to the palindromic nature of the sequence (Figure 3.18). However, for the hairpin OP3 no labelled DNA was seen as the quantity of DNA present was too low to resolve any peaks. This can be attributed to the increase to 6 possible peaks that should be observed for the different transalkylation densities of this probe.

After HPLC, identification of each peak was achieved using MS. For OP1 the two new peaks collected at 19 and 22 minutes were identified as having one and two labels respectively (Supplementary Figures S3.12, 3.13). As seen previously for the 14 bp oligo, there was an increase in 121 to the molecular weight of the hairpin for each alkyl group introduced which matches with the alkyl chain of **AdoHcy-6-N₃** which was the cofactor analogue used in transalkylation.

MS was also initially attempted with the fractions collected for the OP2 transalkylation however, the DNA concentration was too low. Instead, transalkylation was repeated and the DNA was purified to a higher concentration and used directly in LCMS (Figure 3.20.A). Again, LCMS showed a clear introduction of 4 new peaks, but with better peak distribution as a new batch of enzyme was used with apparent higher activity. Analysis of the total ion count (TIC) spectra collected during LCMS was used to calculate the molecular weight of each individual peak seen in the LC trace. Here the four peaks corresponding

to DNA carrying 1, 2, 3 and 4 labels. Further confirmation came when combining and analysing the TIC spectra for all peaks where peak intensities remained proportional to what was seen in the UV trace (Figure 3.20.B). LCMS was also attempted for OP3 however, was unsuccessful as the DNA concentrations for each peak were too low for detection. Due to the high quantity of enzyme required for each transalkylation experiment, analysis of OP3 was not explored further.

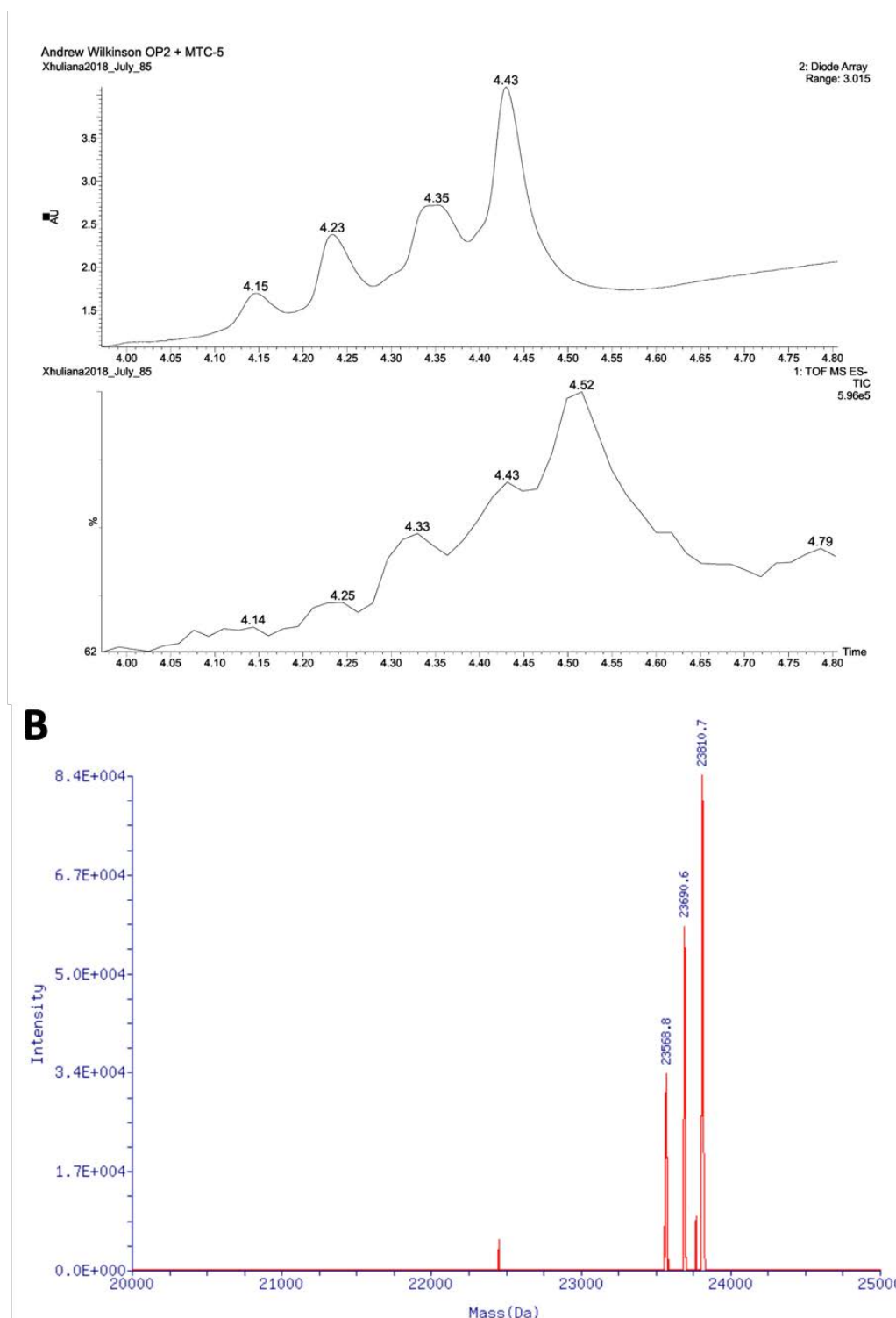


Figure 3.20 A) LCMS analysis of OP2 transalkylation with **AdoHcy-6-N₃** and *M.TaqI*. Separation was achieved using a 0.1M TEAA buffer gradient with acetonitrile. Top = UV absorbance 200-800 nm, Bottom = TIC C) After LCMS the TIC spectra collected were then combined between 4.25-4.55 minutes and analysed using the software promass for masslynx by waters giving the molecular weights: 23,568.8, 23690.6, 23810.7 (Unlabelled = 23445), showing the 121 g mol⁻¹ increment expected.

Another factor these results showed was that transalkylation with hairpin oligo's had reduced effectiveness compared to that of the 14 bp oligo used previously. This can be seen clearly in the transalkylation of OP1 as when carrying out alkylation the TCGA:M.TaqI ratio was the same as used previously for 14 bp studies, but transalkylation was approximately 50% less effective as it did not go to completion. One explanation for this could be a reduced favourability for alkylation at one of the transalkylation sites. If this is the case it is likely due to the short double stranded region at the tail end of the hairpin with only 4 bases before becoming single stranded (compared to 5 in the 14 bp oligo) which may reduce enzyme binding efficiency as there is not enough space for effective docking. Alternatively, it could also be due to the proximity of the hairpin to the transalkylation site which could also reduce enzyme binding efficiency. However, from this result alone this is difficult to prove. To prove this, it would be necessary to run a series of alkylation experiments in parallel with a range of OP1 probes which vary the position of the TCGA site in relation to the double stranded break region and seeing how transalkylation efficiency is impacted.

After the successful HPLC of alkylated hairpins, work moved towards fluorescent tagging of probes which was done using azide alkyne click chemistry. Traditionally coupling is achieved using a copper catalyst, however due to the oxidative damage that this can cause to DNA, this was avoided²⁶. Instead strain promoted azide alkyne click chemistry (SPAAC) was used with the commercially available dibenzocyclooctyne (DBCO) fluorophore, TAMRA²⁹. As the coupling efficiency of DBCO-TAMRA to DNA was unknown to us, coupling was done prior to HPLC to ensure any alkylated DNA (not fluorescently tagged) could be

removed during purification. Transalkylation was initially tested using the probe OP1 as analysis was the simplest and required the least DNA.

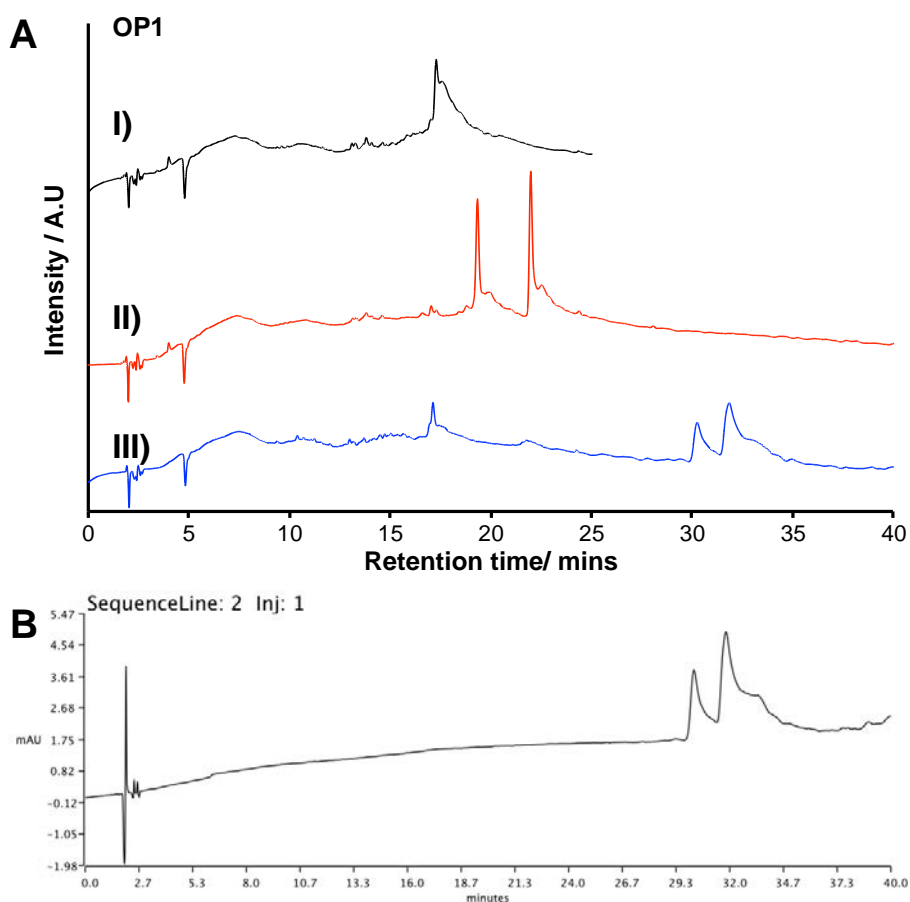


Figure 3.21 A) HPLC traces collected of OP1. I: unlabelled hairpin. II: OP1 after transalkylation with **AdoHcy-6-N₃** and *M.TaqI*. III: OP1 after labelling with DBCO-TAMRA. Coupling was achieved at room temperature overnight with 100 equivalents of fluorophore in 20% DMSO. B) UV absorbance of OP1 after TAMRA labelling at 560 nm. Fluorescent tagging was achieved by coupling of 100 equivalents of fluorophore for each transalkylation site in 20% DMSO at room temperature overnight.

HPLC of the fluorescently labelled oligo showed success through the introduction of two clear new peaks at 31 and 33 minutes (Figure 3.21). In the sample there were also residual peaks from unlabelled DNA at 17 minutes and a small proportion of doubly alkylated oligo at 22 minutes. However, this result does show that DBCO coupling was successful through almost complete removal transalkylated DNA. Also, this result gives further confidence in the structure of

the alkyl group introduced to DNA by showing that the azide introduced can be used for extended couplings. Additional confirmation of fluorescent tagging also came from monitoring absorbance at 560 nm, the λ_{max} of TAMRA, highlighting the new peaks contain TAMRA fluorophore.

Interestingly, the two new peaks formed had the same ratio at both 260 and 560 nm, which was unexpected (Figure 3.21). This is because the latter peak should have two fluorophores attached and so have twice the absorbance at 560 nm compared to one fluorophore. Prior to carrying out this experiment it was expected a linear increase in absorbance would be seen when increasing the number of fluorophores attached to DNA. As through space the fluorophores should be sufficiently separated to avoid interaction due to the length of alkyl chain separating each from the DNA. However, the relationship was clearly not linear and the addition of a second fluorophore did not have a significant impact on absorbance which indicates the fluorophores must interact when symmetrically attached to DNA. It's thought this observation is due to interaction of the fluorophores which causes homo-Forster resonance energy transfer (FRET), altering the fluorophores optical properties. Previously this effect has also been shown when coupling multiple fluorophores to proteins and antibodies, where non-linear increases in fluorescence are also seen^{155 156}. Similar behavior has been observed for fluorescent methyltransferase-directed transalkylation as discussed above³¹.

Further analysis on the fluorescence characteristics of these probes has not yet been explored. However, the evidence provided shows that methyltransferase

transalkylation followed by HPLC can successfully be used as a method to develop probes with multiple chemical tags. In the future this may provide use through attachment of not only fluorophores, but other chemical tags of interest like receptors for surfaces, antibodies or membranes.

3.4 Chapter summary

Through a combination of electrophoresis, HPLC and MS it has been possible to get a complete understanding of how each AdoMet analogue can be used in the transalkylation of DNA with methyltransferases. Using electrophoresis, it was first possible to clearly identify the success of alkylation with all analogues and further gain an understanding of the differences in transalkylation between each cofactor isomer. Following electrophoresis, a more quantitative understanding of transalkylation could be achieved through a combination of HPLC and MS. HPLC showed clear shifts in retention time of all oligos, dependant on the size and number of modifications made to DNA. MS could then be used to give further evidence about the exact structure of each cofactor. Having gained a more complete understanding of the interaction of each AdoMet analogue with methyltransferases it was then possible to move on to their use in applications with confidence. Further to this it is also possible to begin exploring the reversibility of the Schiff base introduced which could potentially create many previously unexplored new opportunities for AdoMet analogues through manipulation of its dynamic transalkylation.

3.5 Materials and Methods

3.5.1 Materials

TAMRA-dibenzylcyclooctyne conjugate was purchased from Click Chemistry Tools LLC. CutSmart® buffer, pUC19, proteinase K, HaeII, HhaI and TaqI were purchased from New England BioLabs (NEB). The DNA sequence for 14bp oligo was 3'-GCCGCTCGATGCCG-5' 5'-CGGCGAGCTACGGC-3', and was purchased from Integrated DNA technologies®. UV-Vis absorbance measurements were performed on Shimadzu BioSpec-nano. M.TaqI and M.MpeI were synthesised by other members of the Neely group and gifted for these studies.

3.5.2 Gel Electrophoresis

3.5.2.1 General Procedure M.TaqI

For all gels using M.TaqI a similar procedure was followed and included the creation of a master mix followed by dilution using either enzyme or cofactor. Volumes could be changed easily dependent on the number of lanes to be analysed. But in general each tube was set up to contain 300 ng pUC19, 1 µl buffer, 0.25 µl M.TaqI (variable if used in dilution), and 0.5 µl cofactor (variable concentrations). An example procedure for this which gave the highest transalkylation efficiencies is as follows:

On ice, a master mix was created by mixing 79.5 µl molecular grade water, 10 µl (10x NEB CutSmart buffer) and 3 µl pUC19 (1000ng/µl). 2x9.25 µl was taken for tubes 4, 8 (restriction controls) and 0.25 µl of each cofactor isomer (15 mM, concentration variable) and 0.5 µl water was added. To the remaining master mix 4 µl M.TaqI (0.3 mg/ml) was added and mixed. The master mix was then split

into 2x 19.5 µl (1,5) and 4x10 µl and labelled (2,3,6,7). A 2x serial dilution was made by adding 0.5 µl cofactor to the final concentration 300 µM to tubes 1/5 and mixed. 10 µl was then taken from tubes 1/5 and added to 2/6 and continued until tubes 6/7 discarding the final 10 µl. Additional controls for this experiment were also set up (Table 1). All samples were incubated at 50 °C for 1 hour before adding 0.5µl restriction enzyme (Taq αI) to all tubes except 9. Samples were again incubated for 1 hour at 50 °C. 0.5µl proteinase K was added to all tubes and incubated at 50 °C for 1 hour before being run on 1% agarose gel to analyse.

Table 1: Controls set up for M.TaqI gel electrophoresis

| | AdoMet control (9/10) | No cofactor (11) | No methyltransferase (12) |
|--------------------|--------------------------|---------------------|------------------------------|
| 10x NEB CuSmart | 2 µl | 1 µl | 1 µl |
| pUC19 (1000 ng/ul) | 1 µl | 0.5 µl | 0.5 µl |
| AdoMet (3.5 mM) | 0.5 µl | - | - |
| m.Taq.I | 0.125 µl | 0.125 | n/a |
| Water | 16.5 µl | 8.5 µl | 8.5 µl |

3.5.2.2 General Procedure M.MpeI

For all gels using M.MpeI a similar procedure was followed and included the creation of a master mix followed by cofactor dilution. Volumes could be changed easily dependent on the number of lanes to be analysed. But in general each tube was set up to contain 750 ng pUC19, 3.5 µl buffer, 2.5 µl M.MpeI, and 1.5 µl cofactor (variable concentrations). After transalkylation samples were then purified using Qiaquick columns and restricted with the chosen enzyme. An example procedure for this which gave the highest transalkylation efficiencies is as follows:

On ice, a master mix was created by mixing 254.25 μ l molecular grade water, 31.5 μ l (10x NEB CutSmart buffer), 6.75 μ l pUC19 (1000ng/ μ l) and 22.5 μ l M.Mpel (1.7 μ g/ μ l). The master mix was then split into 2x 67.7 μ l (tubes 1,5) and 4x35 μ l and labelled (2,3,6,7). A 2x serial dilution was made by adding 2.3 μ l cofactor (15 mM) to the final concentration 300 μ M to tubes 1/5 and mixed. 35 μ l was then taken from tubes 1/5 and added to 2/6 and continued until tubes 6/7 discarding the final 35 μ l. Restriction controls were then set-up for lanes 4 and 8 using 29.55 μ l molecular grade water, 0.75 μ l pUC19 (1000 ng/ μ l), 3.5 μ l cutsmart buffer and 1.2 μ l cofactor (15 mM). Additional controls for this experiment were also set up (Table 2). All samples were incubated at 37 °C for 1.5 hours followed by addition of 1.5 μ l proteinase K and incubation at 50 °C for 1 hour. Samples were then purified using Qiagen Qiaquick nucleic acid purification kit eluting into 25 μ l molecular grade water. The concentration of each sample was then measured on the Shimazu Biospec-nano followed by addition of 2 μ l of restriction enzyme to all except tube 9. Samples were again incubated at 37 °C for 1 hour with subsequent addition of 1.5 μ l proteinase K and final incubation for 1 hour at 50 °C. 300 ng of DNA was then taken from each sample and run on a 1% agarose gel at 120 V for 50 minutes followed by imaging

Table 2 Controls set up for M.Mpel Gel electrophoresis

| | AdoMet Control (9&10) | No Cofactor (11) | No methyltransferase (12) |
|--------------------------|-----------------------|------------------|---------------------------|
| 10x NEB CutSmart | 7 μ l | 3.5 μ l | 3.5 μ l |
| pUC19 (1000 ng/ μ l) | 1.5 μ l | 0.75 μ l | 0.75 μ l |
| AdoMet (32 mM) | 1 μ l | - | - |
| M.Mpel | 5 μ l | 2.5 μ l | - |
| Water | 55.5 μ l | 28.25 μ l | 30.75 μ l |

3.5.3 Nucleic Acid Quantitative analysis

3.5.3.1 *Transalkylation*

For each sample a solution of oligo (120 μ l, 10 μ M), buffer (40 μ l, 10x NEB cutsmart buffer), M.TaqI (4.5 μ l), water (189 μ l) and cofactor (6 μ l, 20 mM) was made. Samples were incubated at 50 °C for 1.5 hrs. After incubation proteinase K (2.5 μ l) was added and the samples were incubated at 50 °C for a further 1 hr. The samples were then purified using the Qiagen Qiaquick nucleotide clean up kit and eluted into 50 μ l water and their concentration was measured by Shimadzu BioSpec-nano. Samples were then taken and used directly in HPLC. The same transalkylation procedure was also used for hairpin oligonucleotides but volumes used sometimes varied dependant on the quantity of DNA required. In general though for each HPLC sample 2 μ g of DNA were required for successful analysis. Volumes could then be back calculated from this taking into account estimated purification yields.

3.5.3.2 *HPLC and Mass Spectrometry*

Analytical RP-HPLC for all oligonucleotides was performed on Agilent Technologies 1260 Infinity equipped with Phenomenex Gemini® C18 (150 x 4.6 mm, 5 μ m, 100 Å, flow rate 1 ml/min) at 60°C. Elution with a 0.1 M Triethyl amine acetate buffer, pH 7.0 /MeCN gradient: gradient system A: 5-18% MeCN over 25 mins, to 100% 5 mins, hold at 100% 10 mins, lower to 5% for 5 mins; system B: 5-31% MeCN over 50 mins, to 100% 10 mins, hold at 100% 5 mins, lower to 5% for 10 mins. For unlabelled and alkylated oligonucleotides gradient A was used and for all remaining samples system B was used. The UV-detection was carried at 260 nm and 560 nm for fluorescently labelled DNA.

After purification by HPLC prior to MS all samples were concentrated by Speedvac. After concentrating samples were then submitted for analysis to the University's analytical facility. MS spectra were obtained on a Xevo® G2-XS-ToF (Waters) and Synapt-G2-S from electrospray ionization (ESI) and time-of-flight (TOF) measurement in negative ion mode.

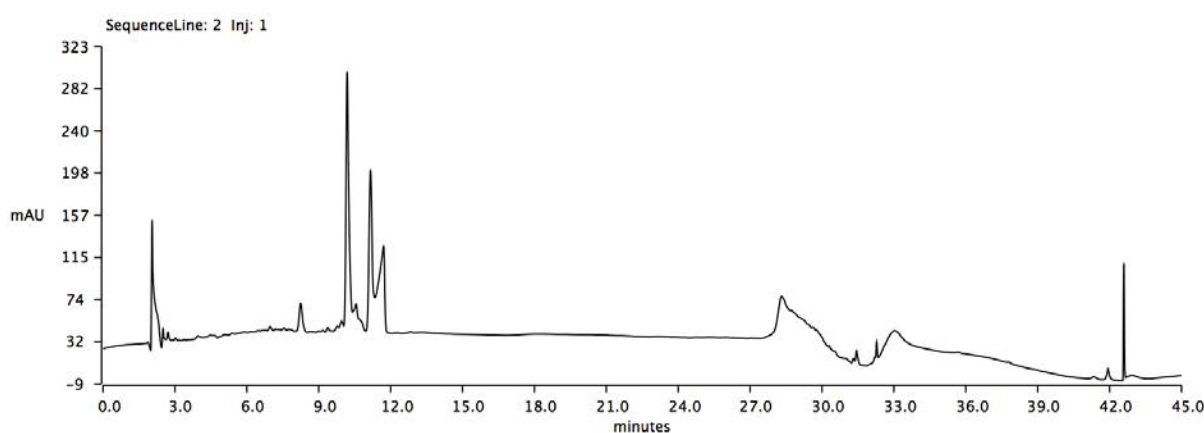


Figure S3.1 Analytical-HPLC chromatogram of 14 bp oligo DNA. Conditions for analytical HPLC: 0.1 M TEAA buffer against acetonitrile. 5-18% MeCN over 25 minutes, 18-100% over 5 minutes, hold 100% 10 minutes, drop to 5% and hold 5 minutes. Column Temp = 60 °C.

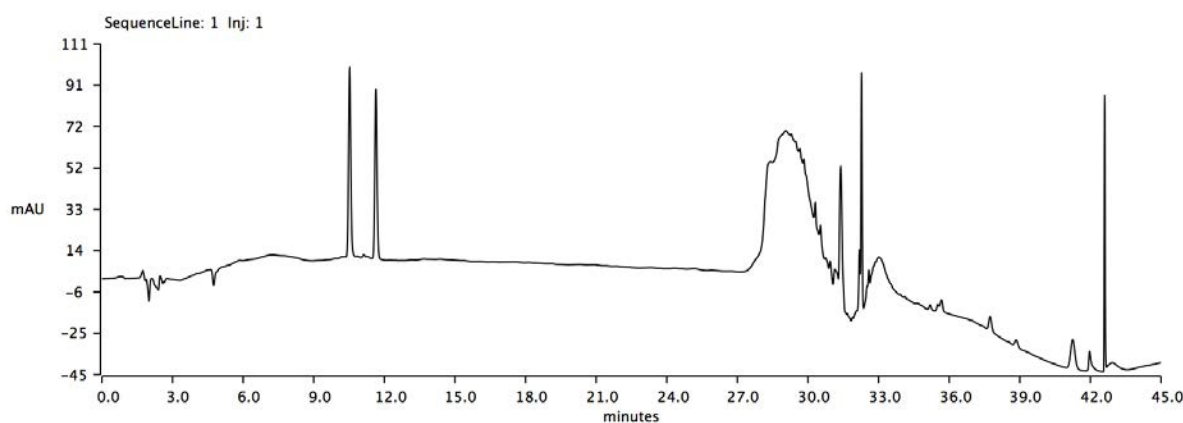


Figure S3.2 Analytical-HPLC chromatogram of 14 bp oligo DNA incubated in the presence of M.TaqI and AdoMet. Conditions for analytical HPLC: 0.1 M TEAA buffer against acetonitrile. 5-18% MeCN over 25 minutes, 18-100% over 5 minutes, hold 100% 10 minutes, drop to 5% and hold 5 minutes. Column Temp = 60 °C.

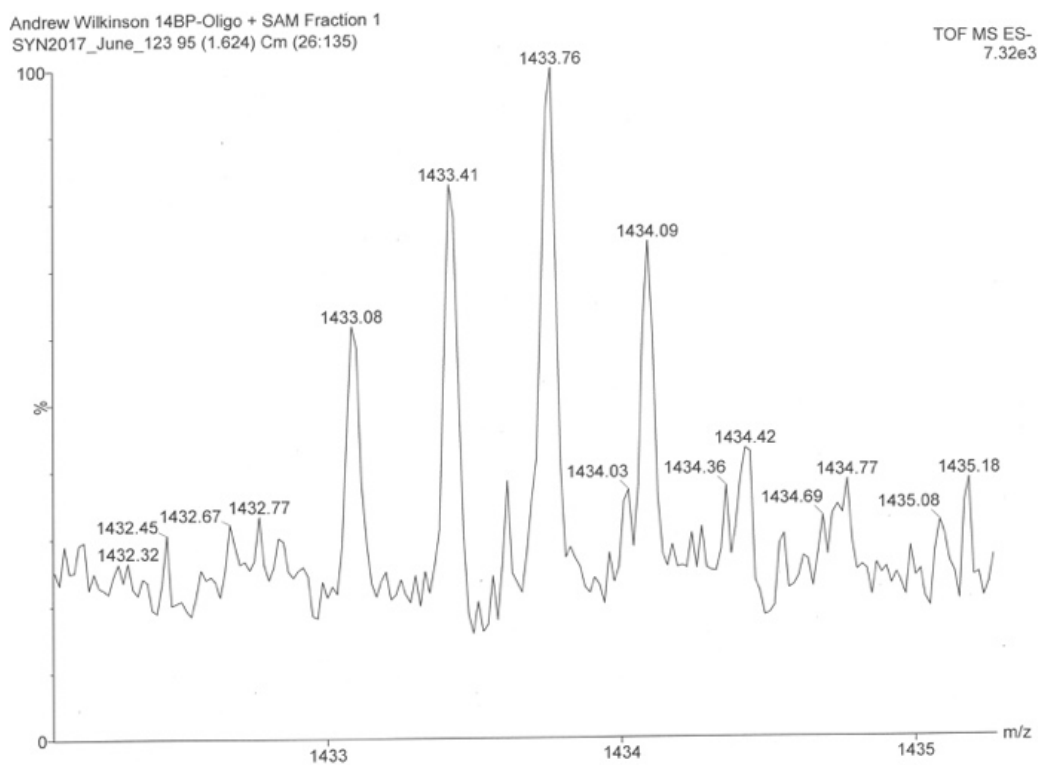


Figure S3.3. Mass spectrum of the peak at 10.53 min from Figure S3.2 (1433.41×3) + 3 = $4303.23 \text{ g mol}^{-1}$

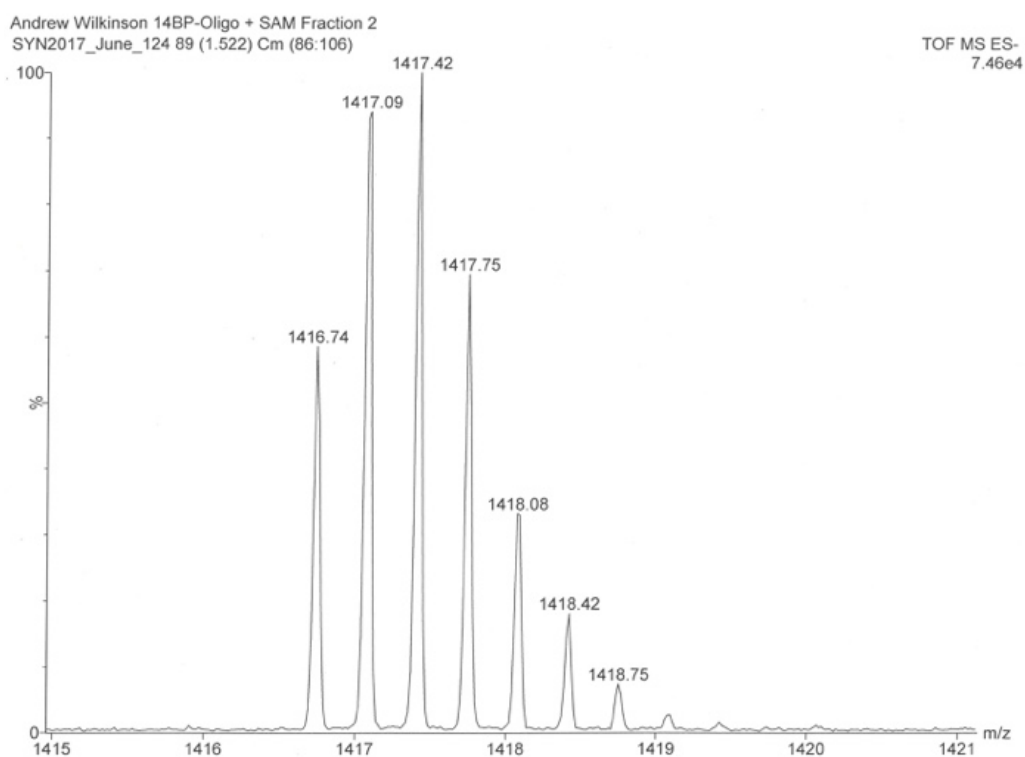


Figure S3.4. Mass spectrum of the peak at 11.64 min from Figure S3.2 (1417.09×3) + 3 = $4254.27 \text{ g mol}^{-1}$

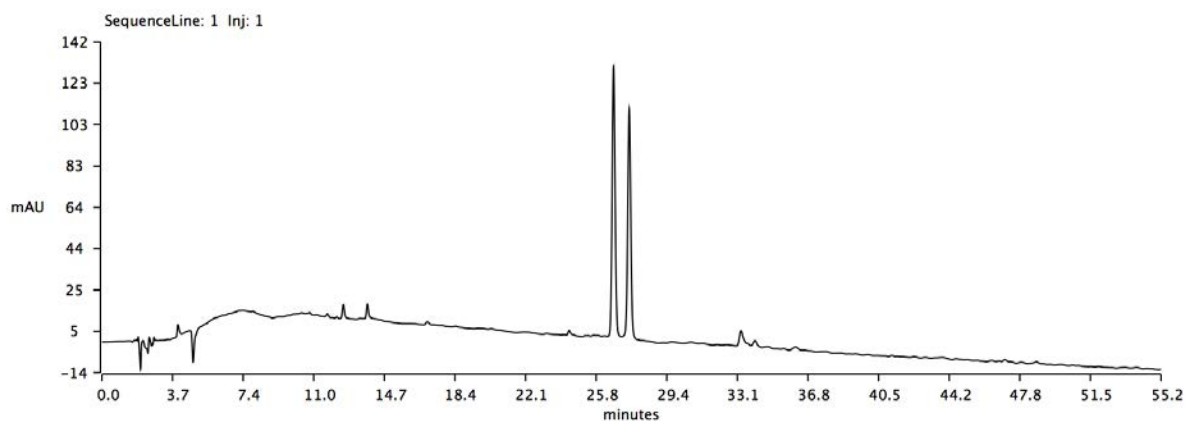


Figure S3.5 Analytical-HPLC chromatogram of 14 bp oligo DNA incubated in the presence of *M.TaqI* and **AdoHcy-8-Hyd**. Conditions for analytical HPLC: 0.1 M TEAA 5-36% over 50 minutes 60 °C

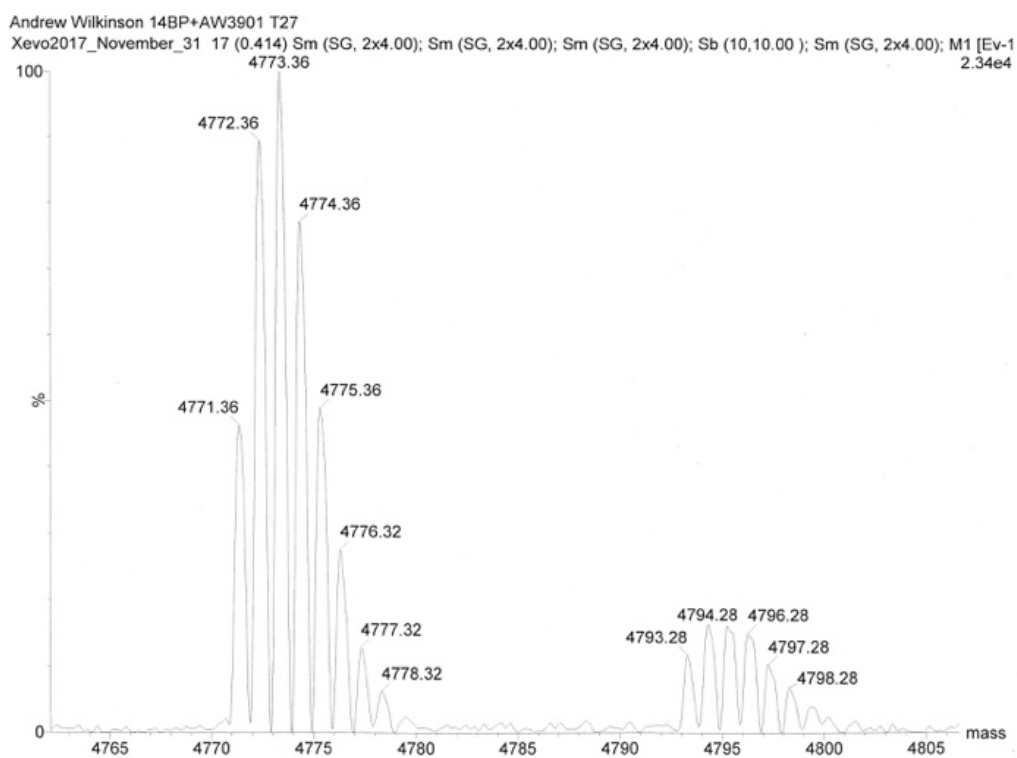


Figure S3.6. Mass spectrum of the peak at 27 min from Figure S3.5.

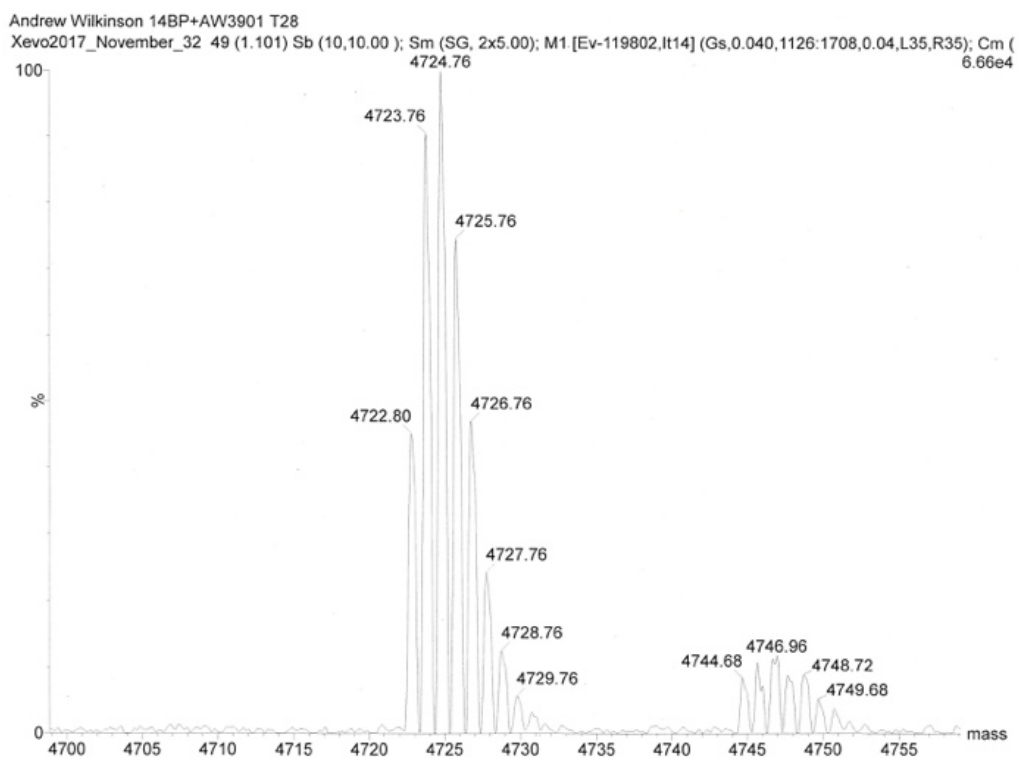


Figure S3.7. Mass spectrum of the peak at 28 min from Figure S3.5.

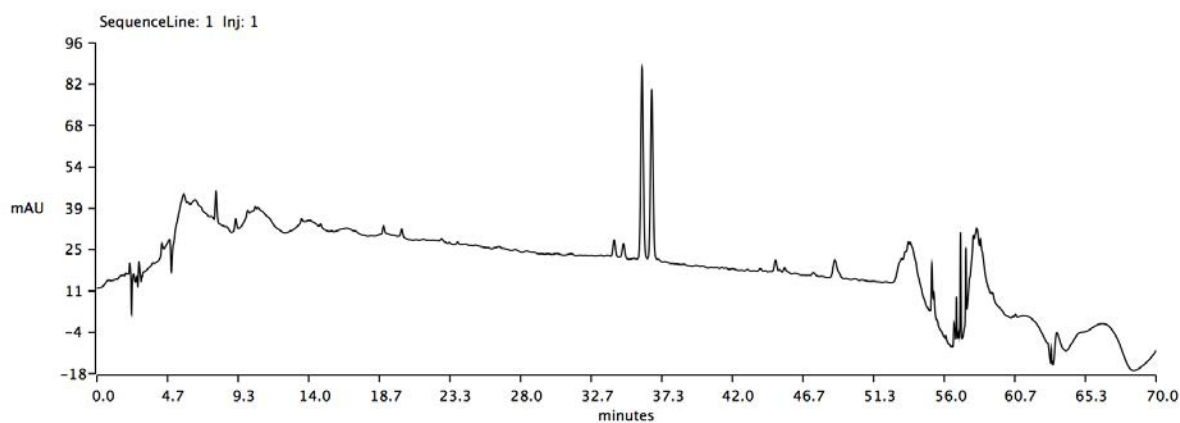


Figure S3.8 Analytical-HPLC chromatogram of 14 bp oligo DNA incubated in the presence of *M.TaqI* and **AdoHcy-8-Ox**. Conditions for analytical HPLC: 0.1 M TEAA buffer against acetonitrile. 5-18% MeCN over 25 minutes, 18-100% over 5 minutes, hold 100% 10 minutes, drop to 5% and hold 5 minutes. Column Temp = 60 °C.

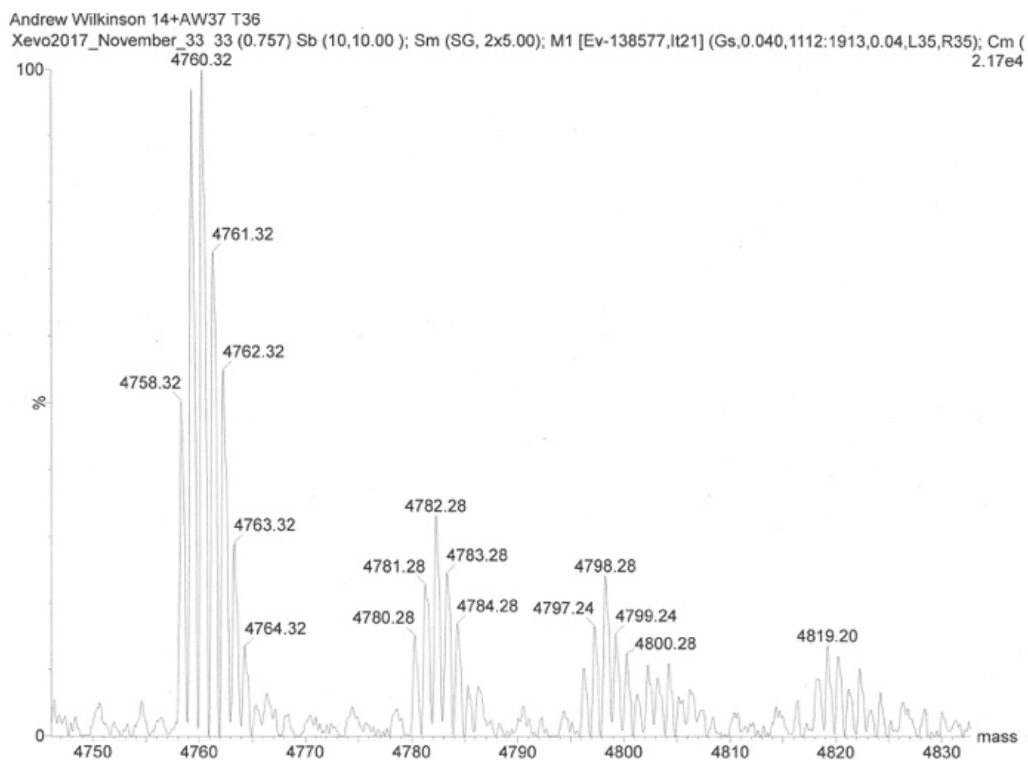


Figure S3.9. Mass spectrum of the peak at 36 min from Figure S3.8

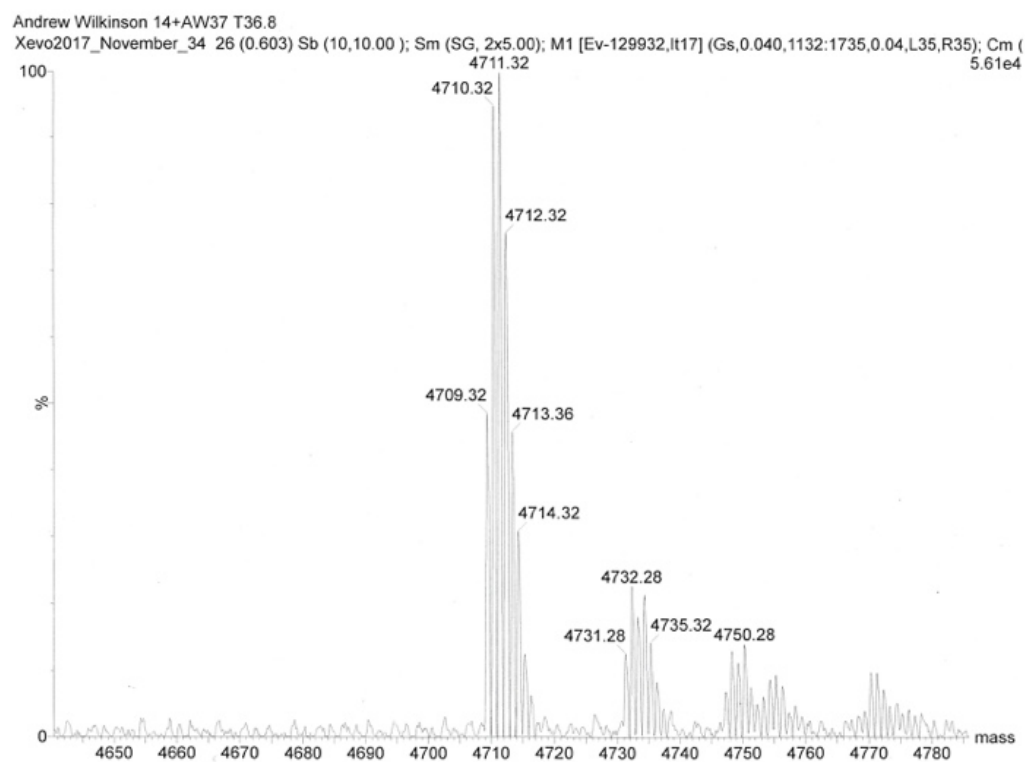


Figure S3.10. Mass spectrum of the peak at 36.8 min from Figure S3.8

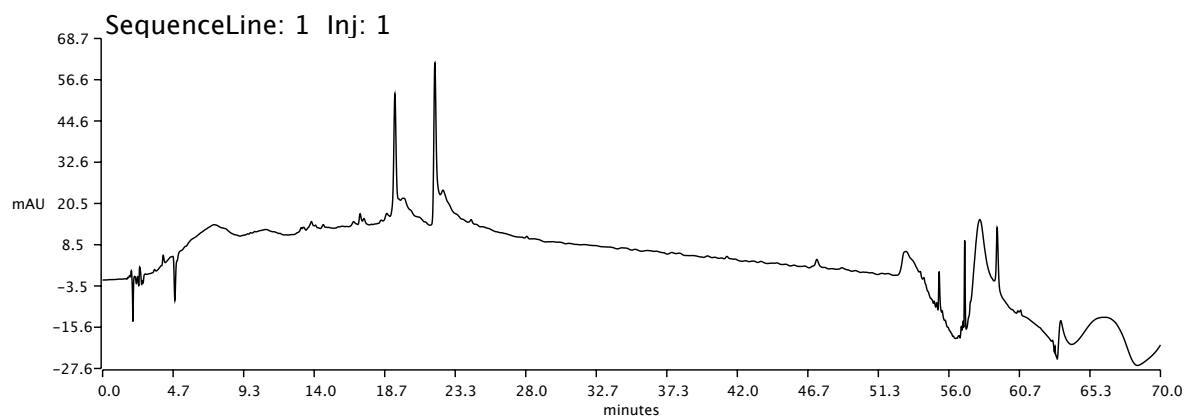


Figure S3.11 Analytical-HPLC chromatogram of Hairpin probe OP-1 DNA incubated in the presence of *M.TaqI* and **AdoHcy-6-N₃**. Conditions for analytical HPLC: 0.1 M TEAA buffer against acetonitrile. 5-36% MeCN over 50 minutes, 18-100% over 5 minutes, hold 100% 10 minutes, drop to 5% and hold 5 minutes. Column Temp = 60 °C.

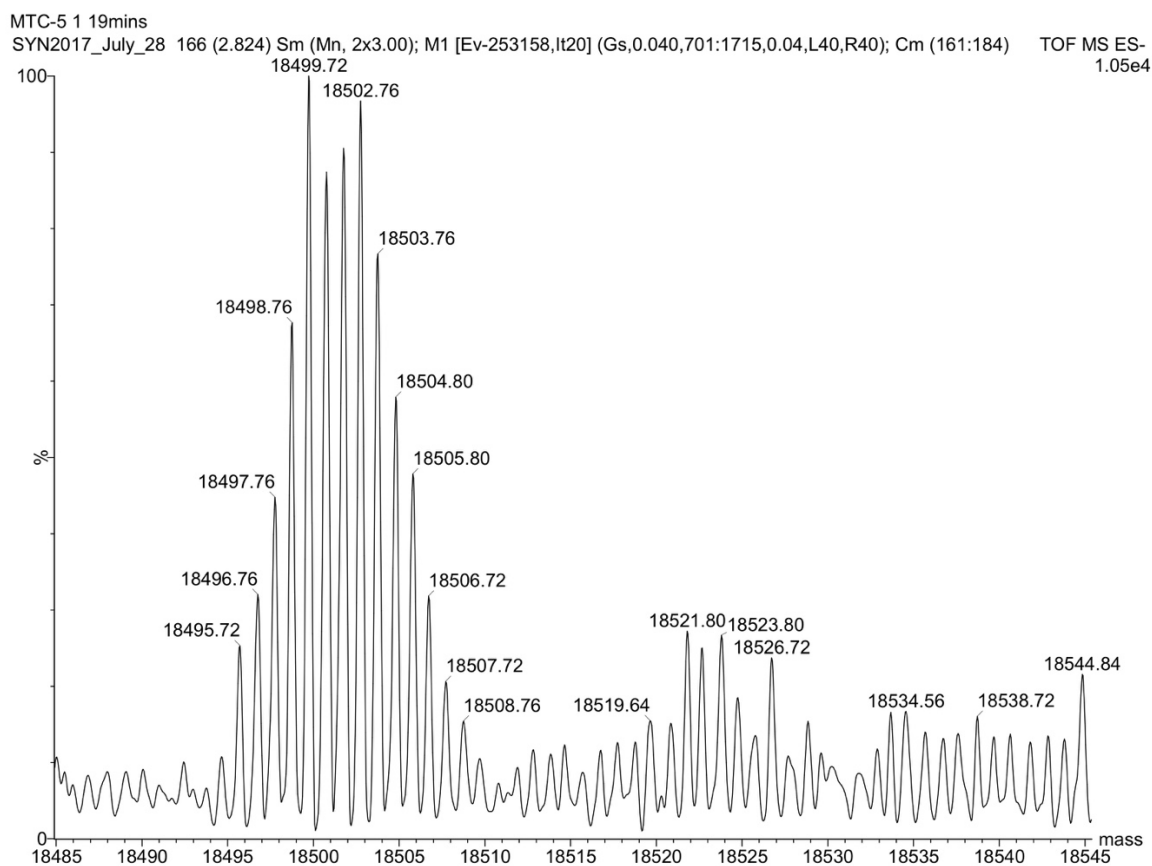


Figure S3.12. Mass spectrum of the peak at 19.0 min from Figure S3.11

MTC-5 2 22mins
 SYN2017_July_31 100 (1.708) Sm (Mn, 2x3.00); M1 [Ev-316064,lt21] (Gs,0.040,714:1995,0.04,L40,R40); Cm (96:113) TOF MS ES-
 1.23e4

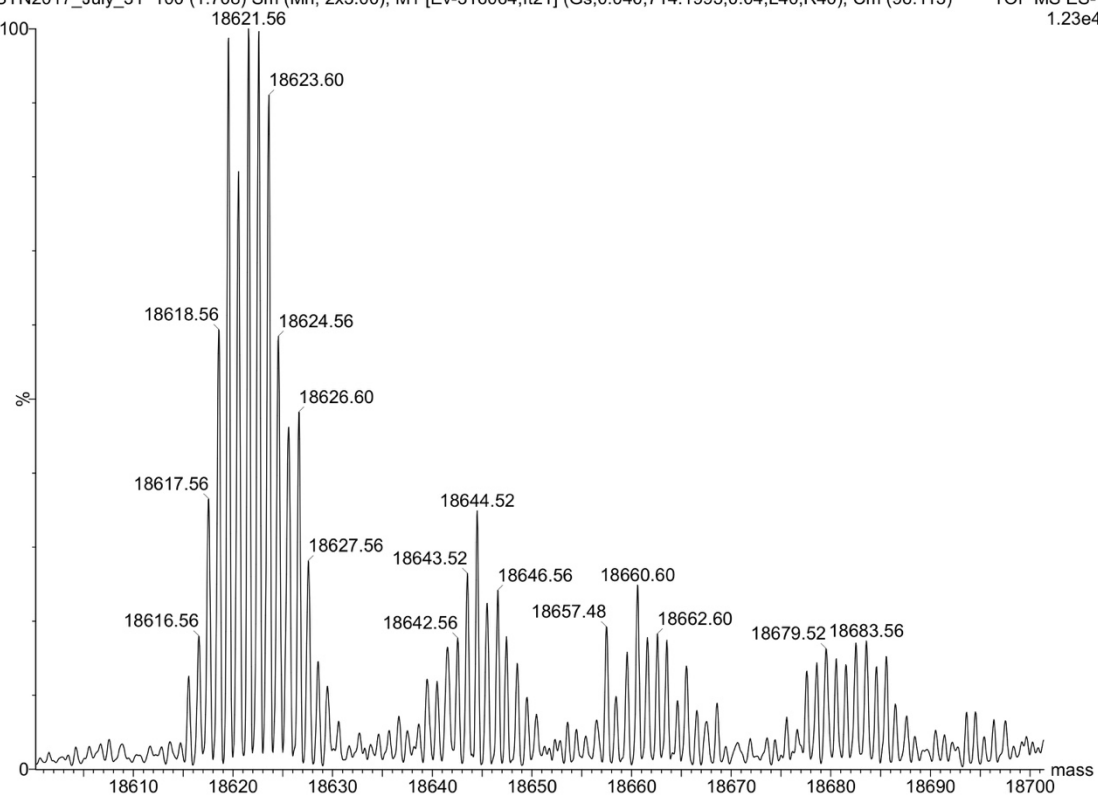


Figure S3.13. Mass spectrum of the peak at 19.0 min from Figure S3.11

4 Write, Erase, Reverse, Rewrite

After successfully proving cofactor effectiveness in the transalkylation of DNA with methyltransferases, work then focused on the development of the responsive cofactors **AdoHcy-8-ox/hyd**. Here we first tested the reversibility of each cofactor by HPLC using the 14 bp oligonucleotide described in Chapter 3. After developing successful reversal conditions model studies were designed and completed which tested each pathway in the dynamic cycle we hoped to create (Figure 4.1). We then used this as a platform to demonstrate the range of applications in which this dynamic chemistry could be used. For example, in epigenetic sequencing studies and the targeting of unmethylated sections of the genome as demonstrated previously by Kriukiene *et.al.*⁴¹

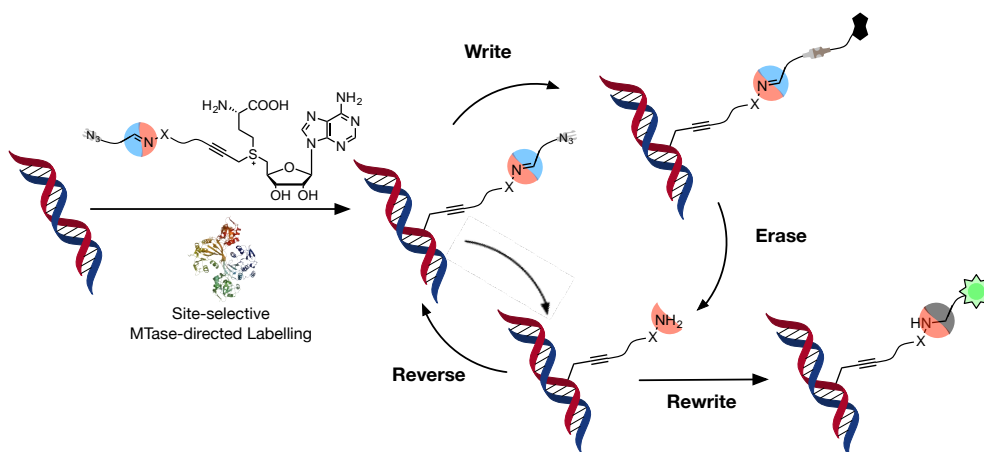


Figure 4.1 Schematic diagram of the dynamic cycle we hoped to create using the responsive cofactors **AdoHcy-8-ox/hyd**. After transalkylation you can **write** desired information into the DNA backbone using azide alkyne click chemistry. This information can then be **Erased** using the reversal conditions developed, at which point the DNA can either be **reversed** to its starting form or **rewritten** to introduce new functionality be it permanent or reversible.

At the heart of all steps within this modification cycle was the dynamic hydrazone or oxime functionality. Schiff bases were chosen for this purpose as their reversal is a well understood pathway that is commonly employed in dynamic covalent

chemistry and has been used with biomolecules successfully in the past¹⁷. For example, in the purification of glycans from protein complex protein mixtures where hydrazone chemistry is used to facilitate their capture and release from solution¹⁵⁷. They are especially useful in dynamic biomolecule transalkylation studies due to their scarcity in nature and the selectivity of the original reaction between imines and aldehydes, minimising the chances of cross reactions. Additionally, they were also chosen as the Schiff base present proved relatively stable during transalkylation. It was expected the oxime had greater stability than the hydrazone as it is less susceptible to hydrolysis, but what wasn't known was which would have adequate stability for the development of our responsive transalkylation pathway¹³⁰. As such, both were taken forward and tested equally under mild reversal conditions that would not affect the overall integrity of DNA to identify one which could be taken forward for future applications.

4.1 Reverse

4.1.1 HPLC

Initial reversibility studies were attempted using the 14 bp oligo used previously in Chapter 3 in combination with HPLC and MS. Schiff base reversal is widely understood for both oximes and hydrazones and is typically achieved through the use of an acid catalyst and the introduction of a competing reagent (CR)¹²⁹. The acid catalyst improves the rate of hydrolysis whilst the CR undergoes reaction with either the hydrolysed aldehyde or amine, preventing reformation of the original Schiff base (Figure 4.2.A).

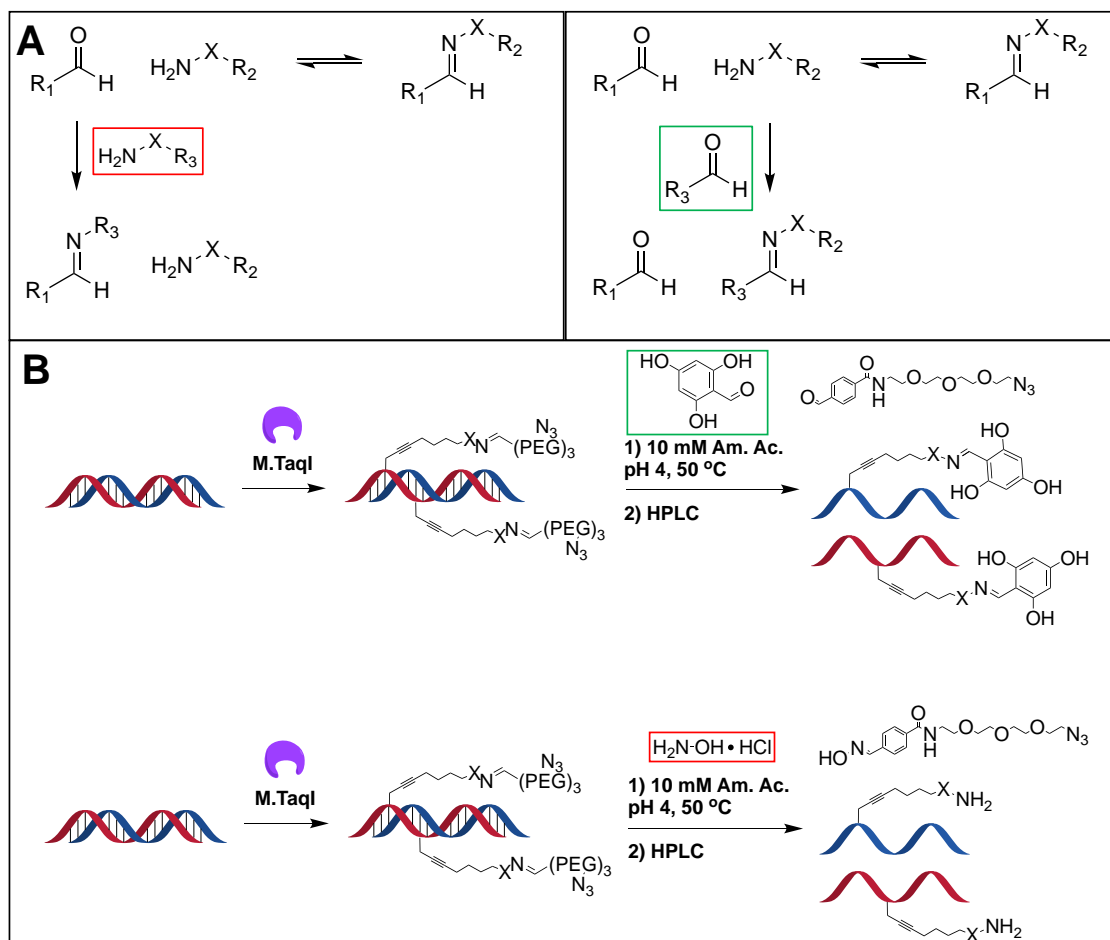


Figure 4.2 The role of the competing reagent within Schiff base reversal. A) The Schiff base equilibrium can be artificially shifted through the introduction of a competing reagent in excess which reacts with either the original aldehyde or amine preventing reformation of the original hydrazone or oxime. B) For HPLC studies, two competing reagents were tried which were; tri-hydroxybenzaldehyde and hydroxylamine. These were used due to the stability of the products they create in the formation of more stable Schiff bases than the original.

For reversal, two competing reagents were initially tested; trihydroxybenzaldehyde (THBA) and hydroxylamine (Figure 4.2.B). THBA was of interest as, it would show direct substitution of aldehydes was possible as a method for the introduction of new functionality to DNA after transalkylation. Whilst hydroxylamine would return the free hydroxylamine/hydrazide which could again be used to introduce new functionality, but in a two-step method. Careful consideration was taken for both CR's, as for reversal to be successful the new

Schiff base formed needed to be at least if not more stable than the original as detailed by Furukawa *et.al.*¹³⁰. Hence the choice of hydroxylamine which will lead to the formation of a stable oxime after competition and THBA which will form a phenyl Schiff base like the one present in the original base.

The initial attempt made at reversal was carried out using both hydrazone and oxime labelled oligos with each competing reagent to monitor any differences in behaviour. For reversal the DNA was first labelled in the same way as previously discussed using M.TaqI and either **AdoHcy-8-ox** or **AdoHcy-8-hyd** (Section 3.3.1). After transalkylation the DNA was then purified on silica columns, eluted into water and split in two. Once labelled the DNA was then acidified through the addition of an ammonium acetate buffer to pH 4.0 and a final concentration of 10 mM. This pH was chosen as according to the literature this should cause complete reversal but minimal hydrolytic damage to the DNA^{130,158}. Ten equivalents of either CR were then added and the samples were incubated at 50 °C for 1 hour. After incubation no further purification was carried out and the samples were analysed directly by HPLC (Figure 4.3).

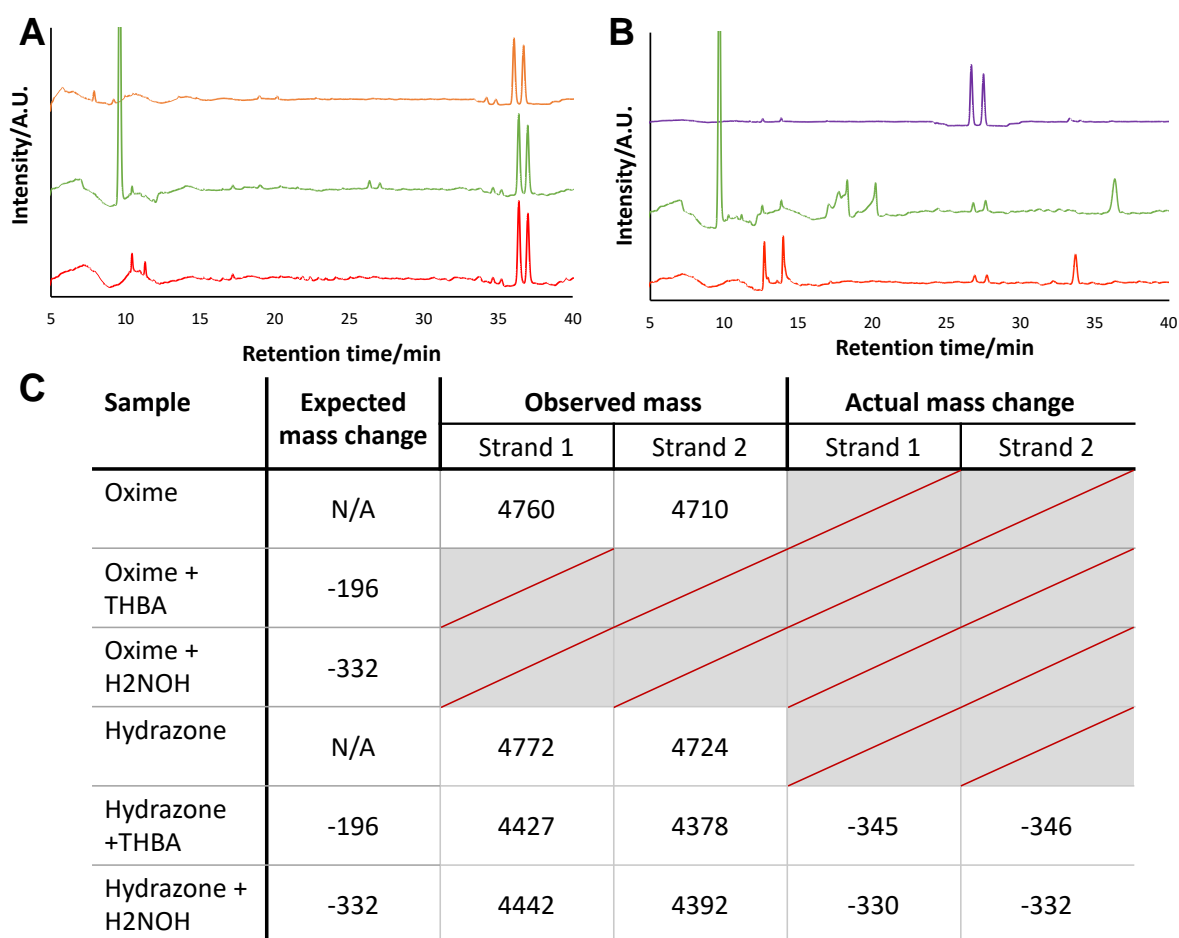


Figure 4.3 HPLC analysis of the 14 bp oligo after the first attempt at reversal A) **AdoHcy-8-ox** labelled DNA, Top = 8-ox modified oligo, middle = THBA reversal, bottom = hydroxylamine reversal. B) **AdoHcy-8-hyd** labelled DNA, Top = 8-hyd modified oligo, middle = THBA reversal, bottom = hydroxylamine reversal. All samples were incubated at 50 °C for 1 hour in 10 mM ammonium acetate pH 4.0 with 10 equivalents of the competing reagent. C) During HPLC each new peak was collected and then concentrated under vacuum before MS analysis. MS and HPLC traces section 4.5.1.1 Figure S4.14-21.

Oxime labelled DNA showed no reversal using either competing reagent (Figure 4.3.A). This stability of the oxime can be attributed to the decreased electrophilicity of the base, caused by resonance-induced electron delocalisation¹⁵⁹. However, with hydrazone labelled DNA there was the introduction of a pair of new peaks for both samples and reduction in intensity of the original hydrazone peaks, indicating reversal was likely successful (Figure 4.3.B). An approximate measure of efficiency for each process could then be

measured by integrating the area under each peak and was found to be approximately 90% and 85% for THBA and hydroxylamine respectively.

After HPLC each fraction collected from hydrazone reversal was analysed by MS to confirm the identity of the peaks and whether reversal had been successful (Figure 4.3.C). For the samples collected with hydroxylamine competition, the exact molecular weights predicted were seen showing the free hydrazide had been formed. This was further supported as the new peaks also overlaid with those previously attributed to free hydrazide present in the unreversed sample.

For the THBA reversal, we anticipated substitution of PEG-aldehyde with THBA and a corresponding reduction in molecular weight of 196 Da. However, a reduction of 345 Da was actually recorded for both strands, which is 14 Da lower than the free hydrazide, potentially corresponding to amide formation. We hypothesise that this occurs through successful substitution with THBA, followed by hydroxyl attack on the Schiff base leading to intramolecular decomposition. However, this has not been previously reported and similar Schiff bases have been successfully formed with THBA in the literature¹⁶⁰. After the failure of THBA substitution, further aldehydes were not tested for direct substitution due to the success of hydroxylamine reversal. Aldehyde substitution was also less favourable as commercial reagents to be used in future stages for DNA functionalisation like fluorophores are widely available in amine reactive cross-linking chemistries, but not as aldehydes.

After this initial study oxime reversal was not explored further as for initial applications hydrazone reversal was preferable as it could be achieved with high yields. Oxime reversal will likely require much harsher conditions with increased heating for longer periods in more acidic solutions. As such, hydrazone reversal using hydroxylamine as the CR was taken forward and explored further in proof of concept studies to see if it could be used as a tool for DNA analysis.

4.1.2 LCMS

Having proven hydrazone reversal could be selectively triggered through the use of a CR we were interested in exploring if reversal could be undone through PEG-aldehyde reattachment. To test this, we carried out a study using LCMS where aliquots were taken to be analysed after each modification step. For this LCMS was chosen over HPLC to minimise sample size and reduce handling between each step as purification would not be necessary. Alkylation was first shown to be successful through the introduction of two peaks at 6.21 and 6.33 minutes where the corresponding masses agreed with what had been seen previously (Figure 4.4.B). After transalkylation hydrazone reversal was achieved by incubating the DNA in an acidic buffer in the presence of hydroxylamine as the CR. For reversal 100 equivalents of the CR were used in an effort to completely shift the equilibrium to the formation of the hydrazone, which was 10 times more than used previously. Hydrazide formation was again successful and a higher reversal efficiency was achieved (>95%) (Figure 4.4.C). This reinforces addition of the competing reagent directly impacts the equilibrium between the hydrazone and hydrazide and is necessary for reversal. After reversal there was also the introduction of two further peaks at 4.24 and 4.40 minutes with lower intensity.

These peaks had the same molecular weights as the free hydrazide and were likely caused by a malfunction during sample injection.

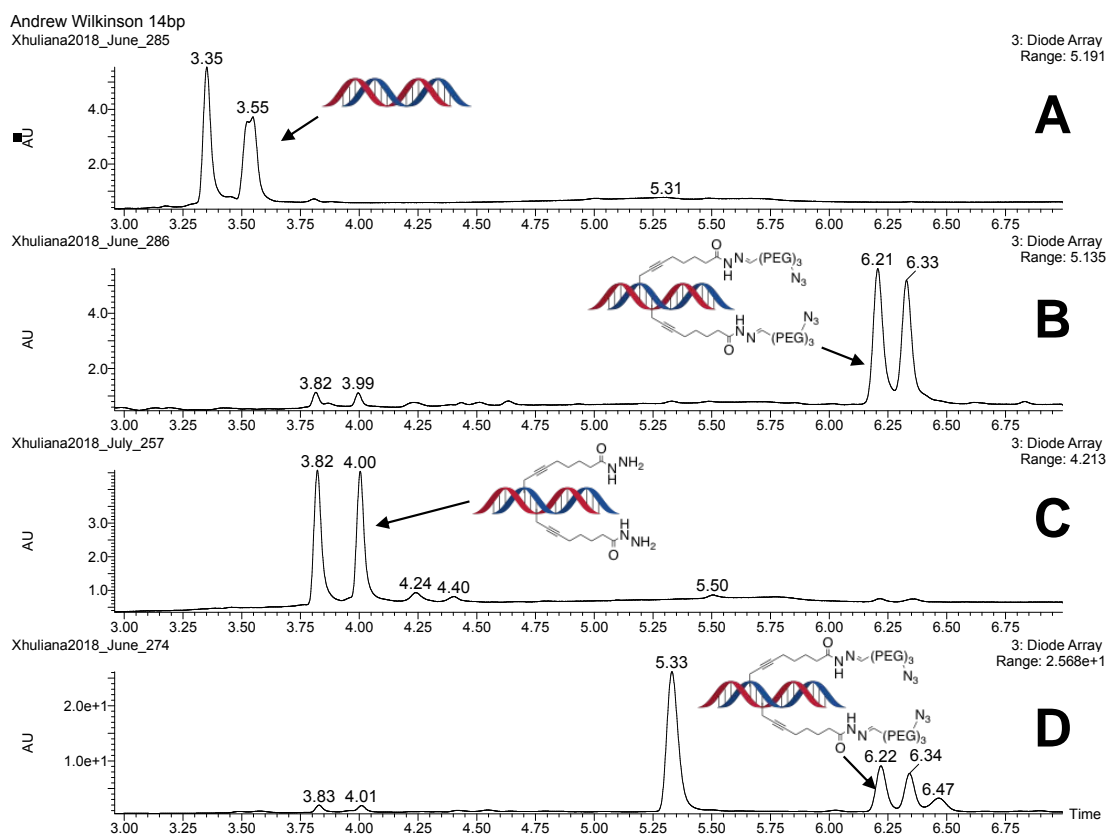


Figure 4.4 LCMS analysis of transalkylation, reversal and reattachment. A) Unlabelled 14 bp oligo, B) labelled with M. TaqI and AdoHcy-8-Hyd, C) hydrazone reversal, D) PEG-aldehyde reattachment. Reversal was achieved by heating labelled oligo at 50 °C for 1 hour in pH 4 ammonium acetate buffer (10 mM) and Hydroxylamine (100 equivalents, 2 mM). After reversal the DNA was not purified and an excess of PEG-aldehyde (200 equivalents, 4 mM) was added and incubated at 37 °C for 1 hour before final analysis. Control chromatograms section 4.4.1.2 S4.31-35.

Once reversal had been confirmed reattachment of PEG-aldehyde was achieved by addition of 200 equivalents of the original aldehyde. A large excess of aldehyde was necessary as the excess hydroxylamine that remained from reversal preferentially reacted with the aldehyde and needed to be quenched. During reattachment there was the introduction of 3 new peaks that initially could not be identified. However, after running a negative reversal control without DNA

these could be attributed to either; free PEG-aldehyde ($t = 8.1$ min), the aldehyde hydrate ($t = 5.3$ min), or PEG-CHNOH ($t = 7.6$) (see supp. Fig. S18-22). After the assignment of these peaks it was clear aldehyde reattachment had been successful and almost all oligo was returned to the pegylated form without any DNA damage caused (Figure 4.4.D).

This result highlights the hydrazone chemistry introduced during methyltransferase transalkylation is truly reversible and a degree of control over the position of equilibrium can be achieved. This is promising, as it is something that previously has not been shown with methyltransferase transalkylation and opens up the potential for many new applications. It also gave confidence that future studies around writing and erasing information into DNA using hydrazone chemistry were also possible.

4.2 DNA capture and release

Following the success of initial reversibility experiments, hydrazone transalkylation was taken forward to further explore its potential as a tool for DNA analysis and modification. We aimed to determine whether hydrazone reversal could maintain high efficiencies in a more experimentally applicable context through the selective capture and release of DNA. For this, a 203 bp fragment of DNA was used and modified using the CpG specific methyltransferase M.MpeI.

In conjunction with the CpG specific methyltransferase M.MpeI capture and release was of interest as when used on fragmented genomic DNA it can selectively capture and release unmethylated sections of the genome

(unmethylome)⁴¹. This is something that has previously been demonstrated with an alternative CpG methyltransferase (M.SssI) with disulphide cleavage for release but with low efficiencies (20-40% recovery). Through the use of our newly developed hydrazone chemistry we hoped to dramatically increase release efficiencies and also explore nanopore sequencing for the analysis of the released DNA.

To first test if capture and release of unmethylated DNA could be achieved using hydrazone reversal we designed a proof of concept experiment (Figure 4.5.A). For these initial studies a 203 bp fragment amplified by PCR from the plasmid pUC19 was used. This fragment was chosen as it had a high density of transalkylation sites (17 CpG sites) and was also of a similar length to the genomic DNA fragments that would be used in future studies. The protocol used for capture was adapted from a method that had been previously developed and optimised within our group. DNA was first labelled with **AdoHcy-8-Hyd** and M.MpeI followed by attachment of the cross-linking reagent DBCO-biotin. Attachment of the cross-linking reagent then allowed for the capture of labelled DNA onto streptavidin beads. After capture DNA release was then attempted using the reversal conditions employed previously.

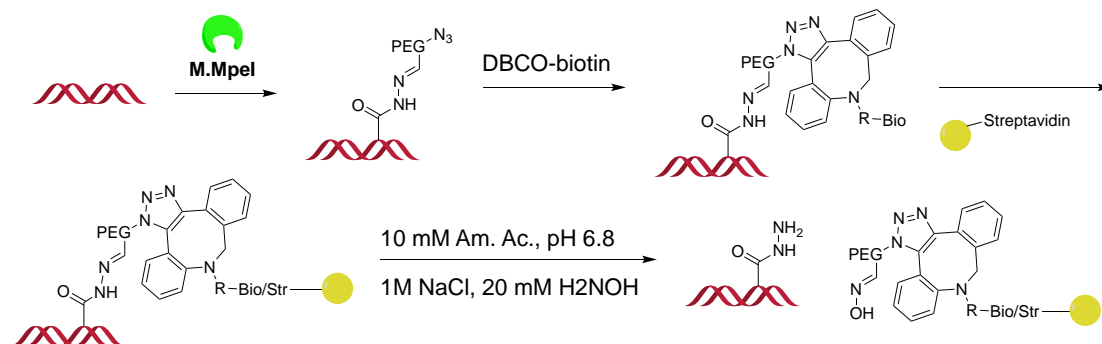


Figure 4.5 The proposed method for capture and release of fragmented DNA. DNA is first transalkylated with an AdoMet analogue and M.Mpel using the labelling conditions used previously for gel electrophoresis followed by coupling of a cross-linking reagent (5x excess) before purification and capture onto streptavidin beads. DNA is then released using the reversible chemistry introduced with the optimised release conditions found to be 10 mM Ammonium Acetate, pH 6.8, 1 M NaCl, 20 mM H₂NOH.HCl.

4.2.1 Selectivity study

Initial attempts at releasing DNA from the streptavidin beads using the same conditions as those previously used in HPLC experiments with 100 equivalents of CR show little to no release. To facilitate the release of any DNA, a large excess of CR was required (ca. 400x). Once release had been successful an initial study into testing the specificity of capture and release was run. For this there were three key samples: unlabelled DNA, **AdoHcy-6-N₃** labelled DNA and **AdoHcy-8-Hyd** labelled DNA (Figure 4.6). Unlabelled DNA was used to monitor non-specific interactions that may occur during capture, whilst **AdoHcy-6-N₃** labelled DNA was used to check release was selective. The selectivity of the release conditions needed to be monitored as there was a chance that the reversal conditions used may cause denaturation of the biotin/streptavidin linkage. **AdoHcy-6-N₃** labelled DNA can monitor this effect as it contains the same biotin/streptavidin linkage but no reversible site. Therefore, any DNA released within this sample would indicate denaturation did occur. Similarly, oxime labelled DNA was not chosen for this purpose as there was a chance that

it could release small quantities of DNA from reversal which would cause a false negative result.

Capture of the labelled DNA was successful with almost complete removal from solution for both samples. This showed, M.MpeI transalkylation followed by the subsequent coupling of DBCO-biotin had been effective in the modification of DNA. However, some non-specific interactions between the DNA and beads also occurred, as indicated by the unlabelled sample, likely reducing the amount of selectively captured DNA.

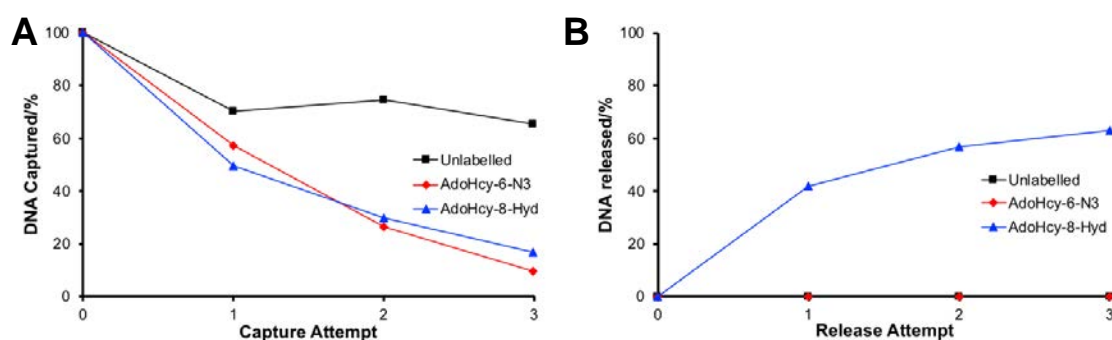


Figure 4.6 A study to test the capture and release efficiency of a 203 bp fragment of DNA with M.MpeI and AdoMet analogues. A) DNA capture was achieved using streptavidin beads in a high salt (1 M NaCl) tris buffer (10 mM). After each capture attempt the supernatant DNA concentration was measured and transferred to a fresh batch of streptavidin beads. B) After capture all portions of beads were combined and resuspended in the high salt tris buffer followed by the addition of the release buffer (10% by volume, 100 mM Ammonium acetate pH 4.0) and CR (10 % by volume, 200 mM $H_2NOH.HCl$). Samples were then incubated at 50 °C for 1 hour and the supernatant DNA concentration was measured. The supernatant was then removed and the beads were resuspended in fresh release buffer for the process to be repeated.

After three capture attempts all bead samples were resuspended and combined in the high salt tris buffer before attempting release. To achieve release, the pH was lowered through the addition of 10% by volume 100 mM ammonium acetate pH 4.0, followed by the addition of approximately 3700 equivalents of CR. Samples were then shaken at 50 °C for 1 hour. After incubation, the supernatant

DNA concentration was measured to monitor any release that may have occurred. No DNA was released in either the unlabelled or **AdoHcy-6-N₃** labelled samples, showing release was selective and the conditions used do not cause denaturation of the biotin/streptavidin linkage (Figure 4.6.B). For **AdoHcy-8-Hyd** labelled DNA after three release attempts a total of 60% of the captured DNA could be released. This equated to an overall DNA recovery efficiency of approximately 50%, 20% higher than the previously reported method. However, for this capture and release method to be successful in future applications we wanted to improve upon the overall recovery efficiency greatly. To do this we ran a series of optimisation studies analysing the effect of each component on release.

4.2.2 Optimisation

The first optimisation study looked into the effect of altering the pH and CR equivalents during release as we wanted to find the mildest conditions possible whilst maximising efficiency (Figure 4.7). For the development of a usable DNA release protocol mild conditions are desired to minimise the chance of damage through denaturation, cleavage, or chemical modification of DNA occurring during capture and release of experimental samples. For all 5 samples in this initial study capture was performed in an identical way and was not affected by the conditions that would be used for release. In an effort to reduce the amount of non-specific binding that may have occurred in the previous study, prior to capture each portion of beads was washed more thoroughly to ensure complete removal of any unwanted salts and surfactants.

After capture all samples were suspended in the high salt tris buffer followed by addition of ammonium acetate and the CR. When adjusting pH the quantity of CR added was not modified and had a final concentration of 20 mM for all samples (7400 equivalents). To adjust the pH a 100 mM ammonium acetate buffer was made and the pH was gradually adjusted with acetic acid for each sample. The final pH was then recorded after the combination of all parts of the release solution. Initially it was planned to check a pH range between 4-6, but ammonium acetate does not buffer correctly above 6 and so the final pH recorded for this sample was higher than expected after combination with Tris. As monitoring the pH on such a small scale was not possible the buffers were combined on a larger scale without the presence of beads or DNA for recording. It was then assumed the pH on a smaller scale would remain the same.

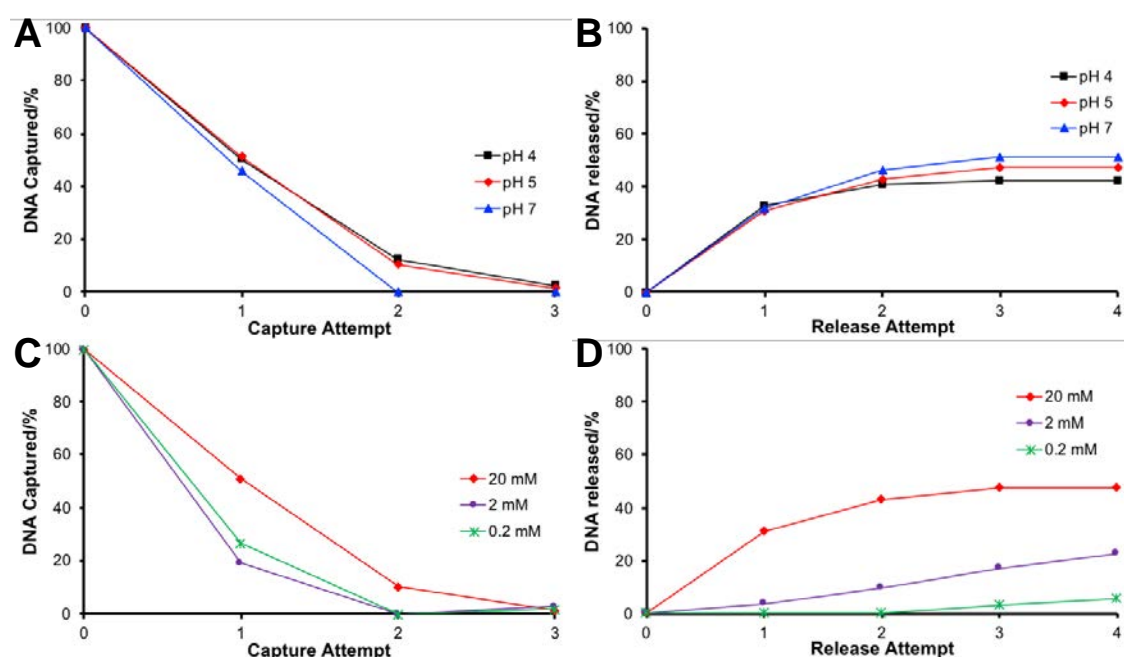


Figure 4.7 Release optimisation assessing the effect of pH and CR concentration on release. Capture for all samples used identical conditions. A) Capture profiles for the three samples to assess pH affect. B) Release profiles showing the effect of adjusting pH on release. C) Capture for varying CR concentration. D) Release profiles showing the effect of the competing reagent on release.

No effect on release was seen upon changing the pH, which was surprising as it was expected release would be less effective at more neutral pH's as hydrolysis is acid catalysed. Unfortunately for all samples there was also no improvement in the release efficiency seen compared to the previous selectivity study. The lack of effect of pH on the release efficiency was positive however, as it meant a more neutral pH could be used during release in future studies. This indicates that the driving force behind hydrazone reversal is more significantly impacted by the CR present rather than the acid catalyst.

Conversely a significant impact on release efficiency was found when monitoring the effect of changing the equivalents of CR present. For monitoring the CR effect all samples were run at a final pH of 5. The three concentrations explored were: 20 mM, 2 mM and 0.2 mM which equate to approximately 7400, 740, and 74 equivalents respectively. The equivalents were calculated from the average number of moles of DNA present within each sample and the theoretical maximum number of labels. As the concentration of CR present was increased there was a clear improvement in release efficiency showing the CR plays a key role in the hydrazide/hydrazone equilibrium position. Interestingly to get satisfactory release of the DNA from the beads far higher equivalents of CR were required compared to previous studies on DNA in solution (Section 4.1.1). We expect the reason for this to be caused by two key factors; firstly, the number of transalkylation sites has increased which makes release more difficult compared to nucleic acids with fewer alkylation sites. This is because, for release to occur every hydrazone attached to both the bead and DNA needs to be reversed. Secondly moving from solution to solid phase likely reduces the accessibility of

the Schiff bases present within the alkyl chain. In this study the equivalents of CR added at the highest concentration was double what was used in the selectivity study previously. However, no significant increase in efficiency was seen. This indicates that between a range of 3500-7500 equivalents the effect seen by hydroxylamine addition plateaus meaning this concentration can be reduced in future studies.

The final factors explored for the optimisation of DNA capture and release were through the removal of Tris from the release buffer and testing the effect of increasing the DNA concentration. Here four samples were tested for capture and release with varying quantities of DNA (Figure 4.8). Both capture and release for all samples were performed in identical conditions to see if there was an optimum amount of DNA that could be captured and released from one sample. Capture was again achieved using the high salt tris buffer with streptavidin coated dynabeads. Followed by release in a 10 mM ammonium acetate buffer pH 6.5 with 1 M NaCl and 20 mM Hydroxylamine. For release a small fraction of SDS (0.01%) was also added to prevent beads from sticking to sample vials.

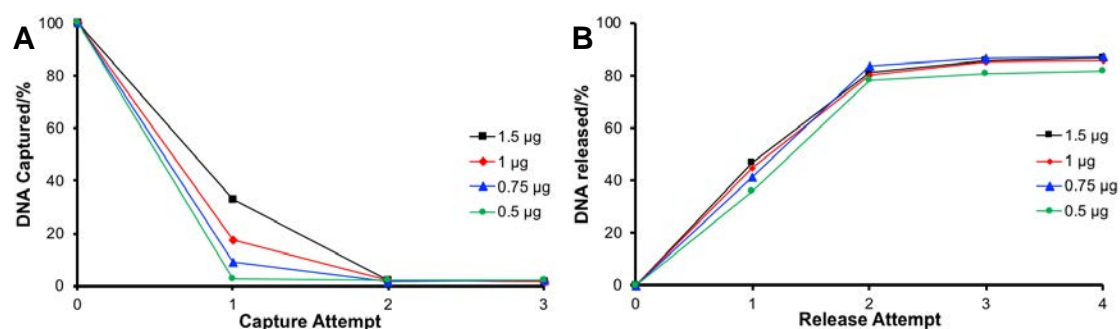


Figure 4.8 DNA capture and release with varying quantities of DNA. A) DNA capture. Capture was performed using 3 portions of 0.01 mg streptavidin dynabeads in 70 µl 10 mM Tris 1 M NaCl. B) DNA release. Release was then achieved using 70 µl 20 mM hydroxylamine in 10 mM ammonium acetate with 1 M NaCl and 0.01% SDS.

No effect was seen on the release efficiency when changing the amount of DNA present. However, capture was affected as the samples with less DNA were more efficient, but for all samples complete capture was still achieved after two attempts. Interestingly the overall release efficiency seen for all samples improved significantly compared to previous results with all samples releasing above 80% of the total DNA captured. This is either due to a reduction in the amount of DNA that is non-specifically bound to the beads during capture, which could be attributed to the use of a new batch of beads, or the removal of tris and addition of SDS to the release buffer. Regardless, the removal of the tris and addition of SDS did not negatively impact release and so was used in the future.

Following the optimisation of capture and release a final study on the PCR product DNA was run using the optimal conditions. Here the original three samples of unlabelled, **AdoHcy-6-N₃** labelled and **AdoHcy-8-Hyd** labelled were repeated in triplicate. Capture showed non-specific binding interactions between unlabelled DNA and the beads had been significantly reduced and also showed transalkylation was successful, as for both labelled samples almost complete capture was achieved. Additionally, for both capture and release, there was also very little variation recorded between repeats, showing the conditions to be reproducible and consistent. Unfortunately for hydrazone labelled DNA average release was > 100%, however this does fit within the error of instrument used to measure DNA concentrations. The most likely cause for this is due to more DNA being present than originally calculated due to an inaccurate initial reading of the starting DNA concentration.

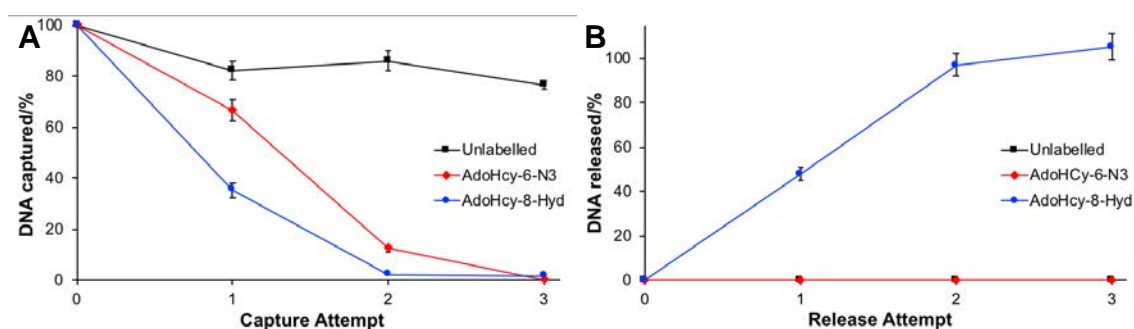


Figure 4.9 Optimised capture and release results A) 400 ng of DNA was captured for each sample in 56 μ l of 10 mM Tris and 1 M NaCl. B) After capture each sample was released in 3 lots of 56 μ l 10 mM ammonium acetate pH 6.5, 0.01% SDS, 1 M NaCl, 20 mM hydroxylamine. The average value from the triplicate repeat was plotted with the error bars plotting the standard deviation of each result.

What can be seen from this result is that after optimisation, non-specific binding has been reduced, DNA release remains specific, and reversal efficiencies have improved dramatically. Hence, after transalkylation information can be written into DNA which can subsequently be used and removed in the capture and release of DNA. After the successful development of this capture and release protocol we were then interested to test its potential in rewriting the remaining free hydrazide with permanent functionality to complete every step of the originally designed dynamic pathway. In addition, we were also interested in further probing the selectivity of this capture and release pathway. Here we intended to test two factors, first, the selectivity of capture. Secondly can release be selective when two types of DNA are captured on the same bead.

4.3 Beyond capture and release

4.3.1 Hydrazide Rewriteability

After the development of a capture and release method for unmethylated DNA we were then interested to explore the rewriteability of the free hydrazide that remains on the released DNA. This had already been partially proven during

LCMS of the 14 bp oligonucleotide previously through the reattachment of the PEG-aldehyde (Section 4.1.2). However, here we were interested in using the hydrazide as a primary amine in the formation of permanent linkages, not Schiff bases. If successful this chemistry could then be translated for use with any methyltransferase/biomolecule of interest, opening many new applications for this technique.

Rewriteability was tested by first labelling the PCR fragment used previously with M.Mpel and **AdoHcy-8-Hyd** followed by its capture and release. After release the DNA was then purified to remove any CR and buffers, followed by coupling of an NHS ester dye. Coupling was then monitored by electrophoresis where the fluorophore (Atto-488) was selectively excited and imaged using blue light. The presence of DNA was detected through the use of a nucleic acid stain (GelRed® Nucleic acid stain, Biotium) which is selectively excited by UV-transillumination. Both images could then be overlaid to see if the fluorophore signal spatially-overlapped with that of the DNA (Figure 4.10.D/E).

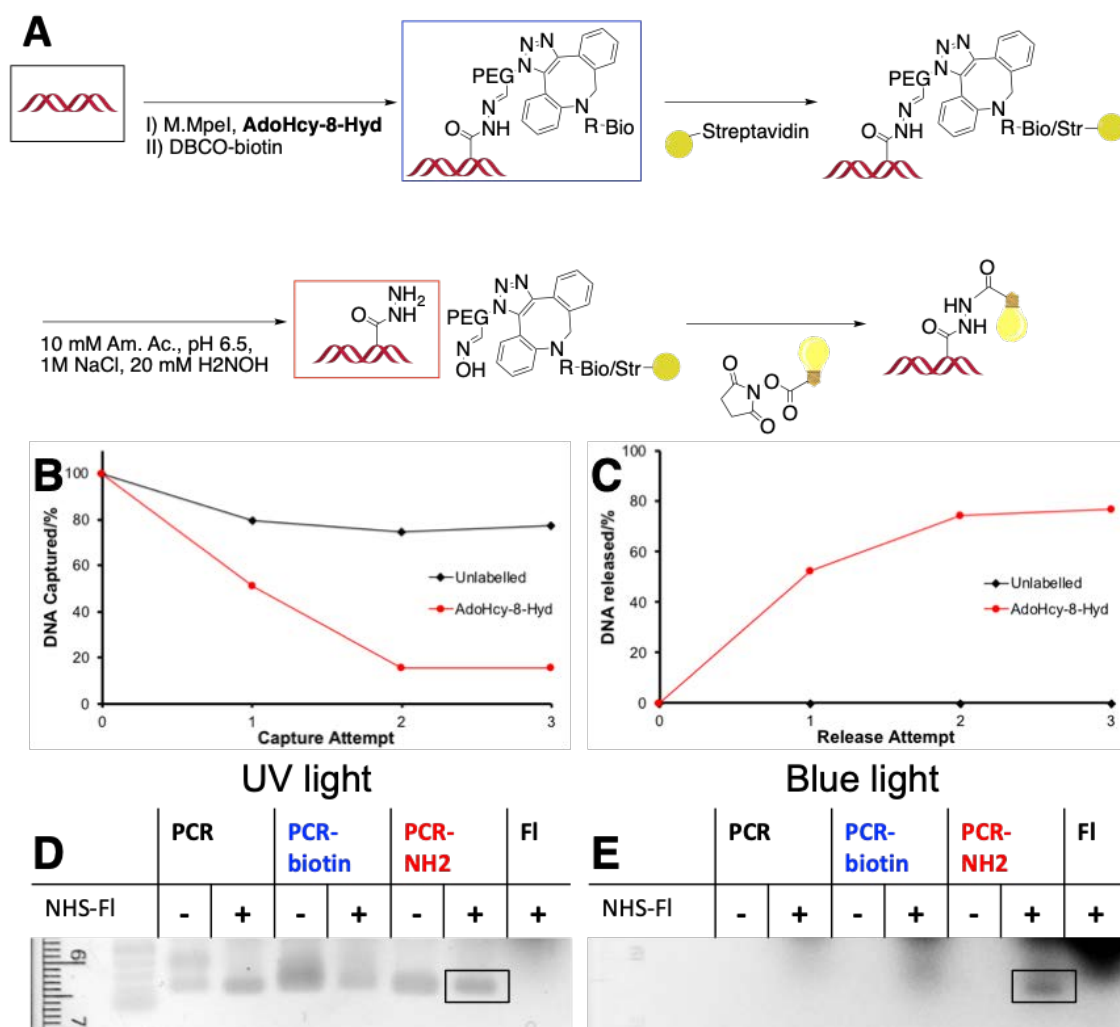


Figure 4.10 Rewriteability was tested using the 203 bp fragment labelled with M.Mpel and AdoHcy-8-Hyd followed by capture and release using streptavidin beads. The released DNA was then coupled to an NHS fluorophore to prove the hydrazide reactivity. After each stage an aliquot of DNA was taken to be used for controls. For uncropped electrophoresis results see supplementary fig. S23.

At each stage of the transalkylation process aliquots of DNA were taken to be run as controls to monitor interaction between the samples and fluorophore. Unlabelled DNA was tested to make sure no interactions occurred with the fluorophore. Biotin labelled DNA was tested to see if the NHS ester introduced could cause degradation of the Schiff base, like hydroxylamine when used as the CR. Due to the negative charge of the fluorophore, Atto-488, some signal was seen for all samples in which it was present as it ran through the gel. However,

through the use of a final fluorophore control these bands could be easily identified. For unlabelled DNA there were no new bands, showing the fluorophore did not interact with DNA non-specifically. But, for the biotin labelled control there was a slight shadow seen that partially overlaid with the DNA fragment. This was attributed to excess free fluorophore rather than direct attachment to the DNA as it was very diffuse, which the DNA band was not when imaged under UV light (Figure 4.10.E). The shadowing seen within this control is likely due to less effective DNA purification after transalkylation which led to excess fluorophore remaining in the sample.

When looking at the released DNA sample under blue light a faint new band could be seen which overlaid with that of the DNA fragment. This is consistent with the attachment of the fluorophore Atto-488 and suggests the free hydrazide is able to undergo reactions with amine reactive groups for the introduction of new functionalities to DNA. Since this experiment was completed it has now been further explored by Elodie Jagu a postdoc within the Neely group who further confirmed this result and successfully switched the fluorescent signal of DNA with two fluorophores (Figure 4.11). Interestingly from this experiment the impact bulky hydrophobic groups have on the overall behaviour of the DNA can also be seen through changes in retention of DNA samples in the gel. As the subsequent modifications are made to the DNA through steps 1-3 the retention gradually shifts as the modification becomes bulkier and more hydrophobic. When this is removed during reversal the DNA retention then returns to a position similar to unmodified DNA. This retention shift is then

switched again following final attachment of ATTO-488 but to a smaller extent as the dye is negatively charged.

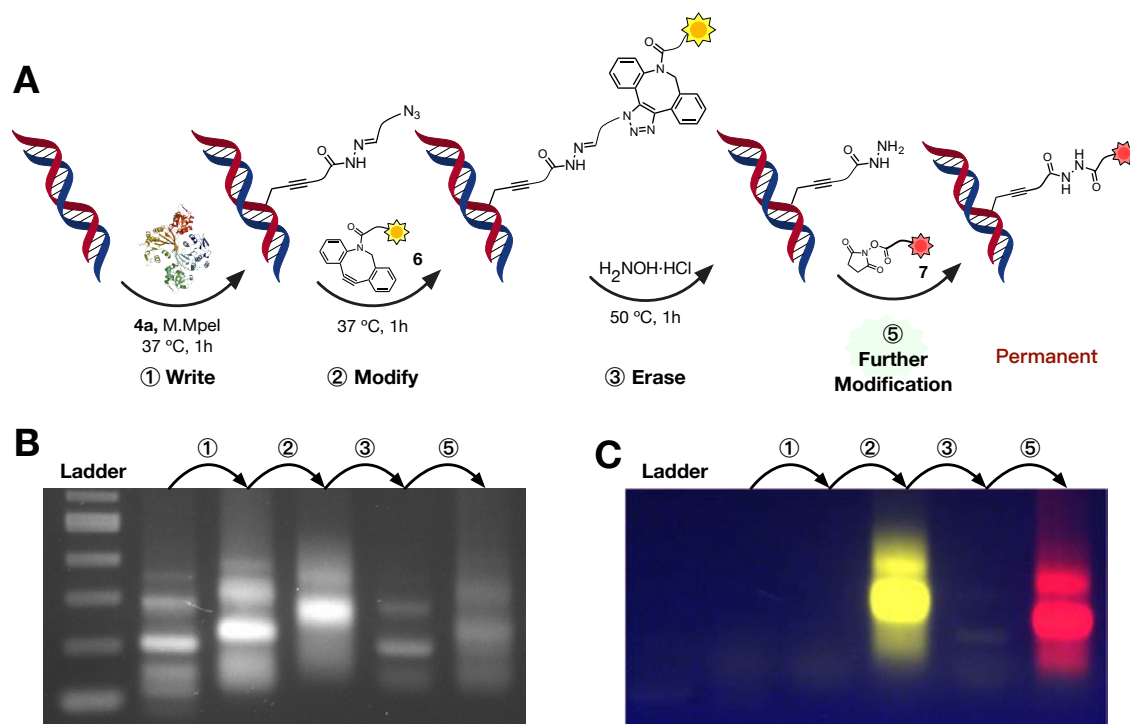


Figure 4.11 Gel electrophoresis testing the reversibility of 8-hyd labelled DNA completed by E. Jagu. A) Schematic representation of the labelling study carried out. DNA was first alkylated followed by direct attachment in solution of DBCO-TAMRA. Labelling was then reversed followed by subsequent attachment of NHS-Atto 647N. Further details of this experiment can be found in the attached manuscript in the appendix.

4.3.2 Unmethylated and Hydroxymethylated capture and release

The next proof concept experiment run with hydrazone labelled DNA focused on the capture and release of two types of DNA on one bead, followed by their selective release. This was of interest as a potential new tool to develop for the epigenetic study of genomic DNA. Within the genome cytosine has been shown to exist in five key epigenetic states which are; unmethylated, methylated and hydroxymethylated, formyl and carboxy each of which, when present in higher or lower than normal quantities, have been linked individually to disease¹⁶¹. Through the development of new techniques, it has been possible to get a clearer

understanding of the role of each of these markers and their location within a genome. However, these are often laborious and require extensive data analysis, and only give an insight into one type of modification. Here, we hoped to develop a one-pot method for the targeted transalkylation of both hydroxy and unmethylated genomic DNA followed by their selective capture and release; after which each of the three epigenetic markers will have been individually isolated (Figure 4.12.A). If successful, when twinned with sequencing, this would allow for the development of a single technique that gives a comprehensive understanding of the entire epigenome from one sample, which is not currently possible. This would have significant benefits in data reduction, simplification of sample preparation, and new analysis being possible looking at links between modification types and disease when twinned with nanopore sequencing.

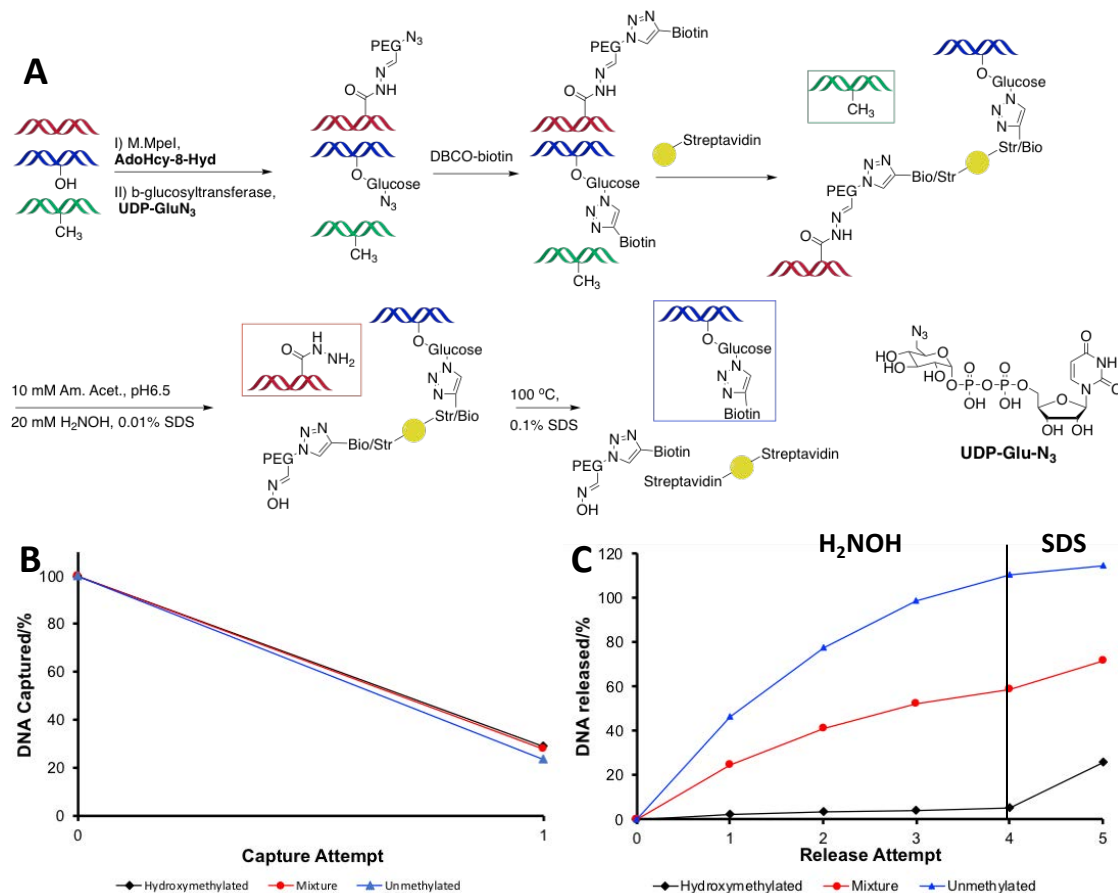


Figure 4.12 A) A method for the purification of the three key epigenetic cytosine markers found within genomic DNA. Hydroxymethylated and methylated DNA can first be azide tagged followed by functionalisation onto streptavidin beads leaving methylated DNA in solution. Unmethylated DNA can be released using hydrazine reversal and Hydroxymethylated DNA by the denaturation of streptavidin. B) Capture performed to test the efficiency of transalkylation of the Hydroxymethylated and unmethylated DNA. C) Release profiles showing the selectivity of each release method. Unmethylated DNA was released over 4 attempts using hydroxylamine competition. Hydroxymethylated DNA was then released by boiling the beads in 0.1% SDS for 10 minutes.

For this initial proof of concept experiment we first tested if capture and release of both Hydroxymethylated and unmethylated DNA was possible on one bead and if release could be selective. This was achieved through the use of the same 203 bp PCR fragment used previously and a new version of it where each cytosine had been replaced with hydroxymethylcytosine (Figure 4.12.A). Transalkylation of unmethylated DNA was achieved in the same way as

previously with M.Mpel. Whilst hydroxymethylated DNA was labelled separately using a T4 phage β -glucose transferase (T4-BGT) enzyme in conjunction with an analogue of its natural cofactor uridine diphosphate glucose azide (**UDP-Glu-N₃**), which selectively targets hydroxymethylcytosine¹⁶¹. Using this combination of enzymes allowed for the introduction of azide functionality into both types of DNA which could then be used in the same way as previously for capture on to streptavidin beads. Selective release would then be achieved by first carrying out hydrazone reversal to remove all unmethylated DNA, followed by hydroxymethylated DNA release by denaturation of the biotin streptavidin linkage, successfully allowing separation of each type of DNA.

The first study into the selectivity of release saw each DNA sample labelled individually and then combined into three samples of interest which were; 100% unmethylated, a 1:1 mixture of both hydroxy/unmethylated DNA and 100% hydroxymethylated. Capture of all samples was then completed once using streptavidin beads, where over 70% of DNA was captured for all samples (Figure 4.12.B). Capture efficiencies between all samples varied by only 3% indicating that the transalkylation efficiency achieved for both methods was equally effective. Also, by capturing over 70% of the DNA in one round it was clear transalkylation was successful for both enzymes with high efficiency.

After capture hydrazone reversal was then attempted using the optimal conditions identified previously. During release, good selectivity was seen with very limited DNA release for the hydroxymethylated sample during hydrazone reversal, showing the biotin/streptavidin linkage remains undamaged (Figure

4.12.C). Further evidence that transalkylation and capture for both hydroxy/unmethylated DNA was equally effective came from the 1:1 mixture which consistently had 50% less release than the unmethylated sample as expected. This indicates there is an actual 1:1 mixture of both DNA types on the beads as expected and further confirming transalkylation for both types of DNA was equally effective. After removing what we thought to be all of the hydrazone DNA in 4 release cycles Hydroxymethylated DNA was released by boiling the sample in 0.1% SDS. Here again release was equally effective for both relevant samples and approximately 20% of hydroxymethylated DNA was released. Unfortunately, further DNA clearly also remained in the unmethylated sample meaning release for the 1:1 mixture was likely a combination of both hydroxy and unmethylated DNA. This is an issue as we hoped to achieve complete purity of all samples which was not achieved. However, what can be seen is that if 100% hydrazone reversal can consistently be achieved, selective release of both DNA types should be possible after optimisation. A further issue also found in this initial experiment was that SDS release had low efficiencies which for actual genomic samples would cause significant issues due to the very low natural abundance of hydroxymethylcytosine within the genome (typically <1%)¹⁶².

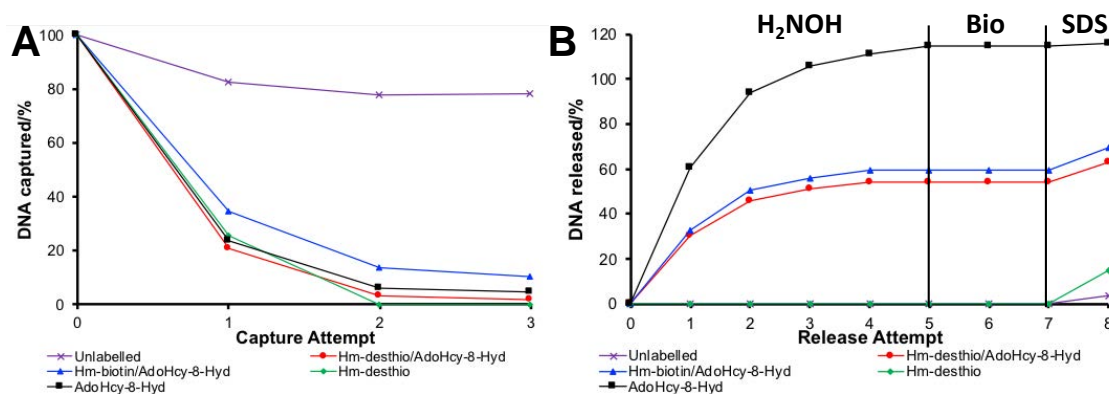


Figure 4.13 Capture and release of Hydroxymethylated and unmethylated DNA with desthiobiotin release.

A) capture profiles of the five samples used to test the effectiveness of biotinylation vs. desthiobiotinylation.

B) Hydrazine reversal was completed over 5 release attempts using hydroxylamine competition, followed by two attempts at desthiobiotin reversal using biotin competition and a final release attempt was completed by boiling in 0.1% SDS.

Due to the low release efficiencies seen with SDS release for hydroxymethylated DNA a new method using desthiobiotin was attempted. Desthiobiotin is an analogue of naturally occurring biotin that binds in the same way to streptavidin, but much less tightly. It is then possible to take advantage of this weaker interaction between desthiobiotin and streptavidin for reversal through biotin exchange in mild conditions¹⁶³. For this experiment five samples were set up in tandem which were: Unlabelled DNA, Hydroxymethylated desthiobiotin (Hm-Desthio) labelled DNA, AdoHcy-8-Hyd labelled DNA, a 1:1 mixture of Hm-Desthio/AdoHcy-8-Hyd labelled DNA and finally a 1:1 mixture of Hm-biotin/AdoHcy-8-Hyd labelled DNA. As can be seen for all labelled DNA samples, capture was effective with almost all DNA removed from solution (Figure 4.13.A). Whilst the unlabelled control showed that relatively little non-specific interactions occurred between the beads and DNA.

After capture, unmethylated DNA was then released with hydroxylamine competition. Complete reversal was achieved after five attempts as determined by the plateau seen for both samples containing a mixture of both hydroxymethylated and unmethylated DNA (Figure 4.13.B). After complete reversal was achieved biotin competition was then employed for the removal of hydroxymethylated DNA. Unfortunately, for both samples containing desthiobiotinylated DNA no release was seen after either attempt. For biotin competition two approaches were employed which were; 2 hours in 2 mM biotin shaking at room temperature followed by overnight in 4 mM biotin. Great difficulty was found during desthiobiotin reversal with the solubility of competing biotin which reduced the effective concentration of the added biotin. Alternative solvents were not used to increase biotin solubility as the effect they may have caused on the beads and DNA was unknown. As such, one final release attempt was made with SDS release to see if desthiobiotin was more easily released during protein denaturation, which it was not.

During SDS denaturation, a further issue was also found with the release of hydroxymethylated DNA in the unlabelled control. Here there was a clear release of small quantities of non-specifically bound DNA. This was a problem as it likely indicates the hydroxymethylated DNA collected during SDS release was impure and, if fragmented genomic DNA, would not be usable in sequencing. Although only a small issue in this proof of concept experiment, non-specific DNA release in genomic samples would be greatly exacerbated due to the much lower quantities of hydroxymethylcytosine present. As such, until the sensitivity and selectivity of hydroxymethylated DNA capture and release could be improved this

experiment was not explored further. However, this proof of concept study does show that it is possible to selectively capture and release different types of DNA modification. This may be of use when using multiple methyltransferases on a sample of DNA or when trying to capture both a protein and DNA sequence of interest from cells.

4.4 Chapter summary

Here we have successfully demonstrated after optimisation that complete reversal of the hydrazone introduced during transalkylation can be achieved whilst attached to DNA. Once this was achieved it was then possible through a series of proof of concept studies to successfully achieve each step in the rewriting cycle originally proposed. Having now demonstrated the versatility and pathways open when using this transalkylation chemistry it is now hoped to take this forward to a diverse range of applications be it sequencing studies, biomolecule purification or other DNA based assays. It is also hoped to further prove the utility of this chemistry with additional methyltransferases and biomolecules of interest to expand its application scope.

4.5 Methods

4.5.1 HPLC and LCMS reversibility studies

For all transalkylation of the 14 bp oligonucleotide the conditions used were the same as those seen previously in chapter 3. After transalkylation DNA was then purified using Qiagen Qiaquick nucleotide clean-up silica columns and concentrated in water before reversal was attempted.

Analytical RP-HPLC for oligonucleotides was performed on Agilent Technologies 1260 Infinity equipped with Phenomenex Gemini® C18 (150 x 4.6 mm, 5µm, 100 Å, flow rate 1 ml/min) at 60°C. Elution with a 0.1 M Triethyl amine acetate buffer, pH 7.0 /MeCN gradient: gradient system A: 5-18% MeCN over 25 mins, to 100% 5 mins, hold at 100% 10 mins, lower to 5% for 5 mins; system B: 5-31% MeCN over 50 mins, to 100% 10 mins, hold at 100% 5 mins, lower to 5% for 10 mins. For unlabeled and alkylated oligonucleotides gradient A was used and for all remaining samples system B was used. The UV-detection was carried at 260 nm

4.5.1.1 *General Reversal conditions*

For initial HPLC and LCMS studies the same general procedure towards reversal was used with only changing the competing reagent or its final concentration depending on the equivalents desired. The general procedure used for this was as follows:

To the labelled DNA, a solution of competing reagent in water (10 µl, X equivs) was added. The pH of the solution was then adjusted using 100 mM ammonium acetate buffer (pH 4.0, 7 µl) and the solution was made to a total volume of 70 µl. The samples were then incubated at 50 °C for 1.5 hrs and then stored in the fridge until analysis.

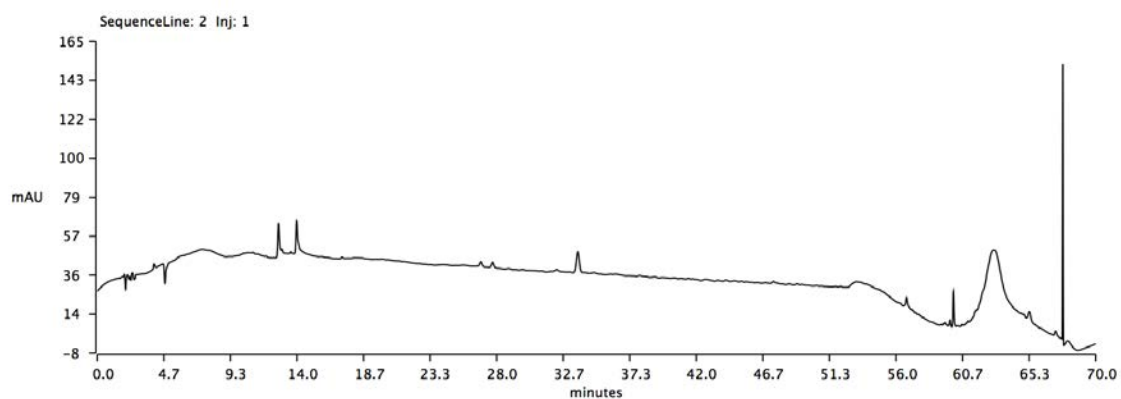


Figure S4.14 Analytical-HPLC chromatogram of 14 bp oligo DNA incubated in the presence of *M.TaqI* and **AdoHcy-8-Hyd**, followed by treatment with $H_2NOH-HCl$ (10 equivs) in a pH 4 10 mM ammonium acetate buffer. Conditions for analytical HPLC: System B

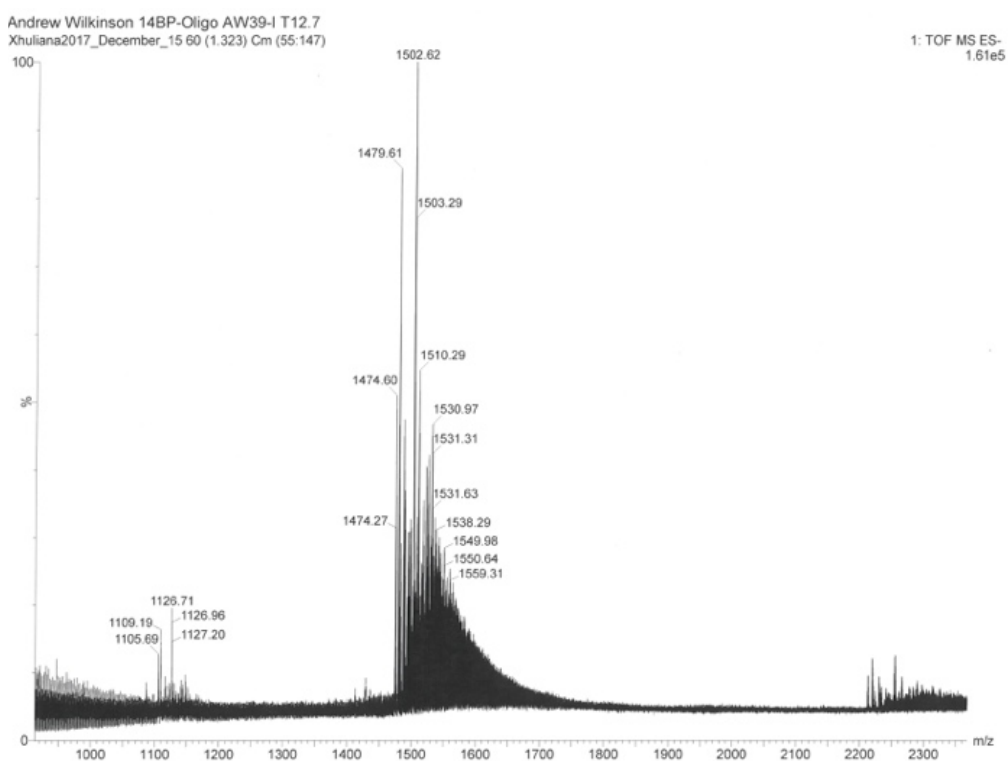


Figure S4.15. Mass spectrum of the peak at 12.7 min from Figure S4.14. Mass = $(1479.61 \times 3) + 3 = 4441.83 \text{ g mol}^{-1}$

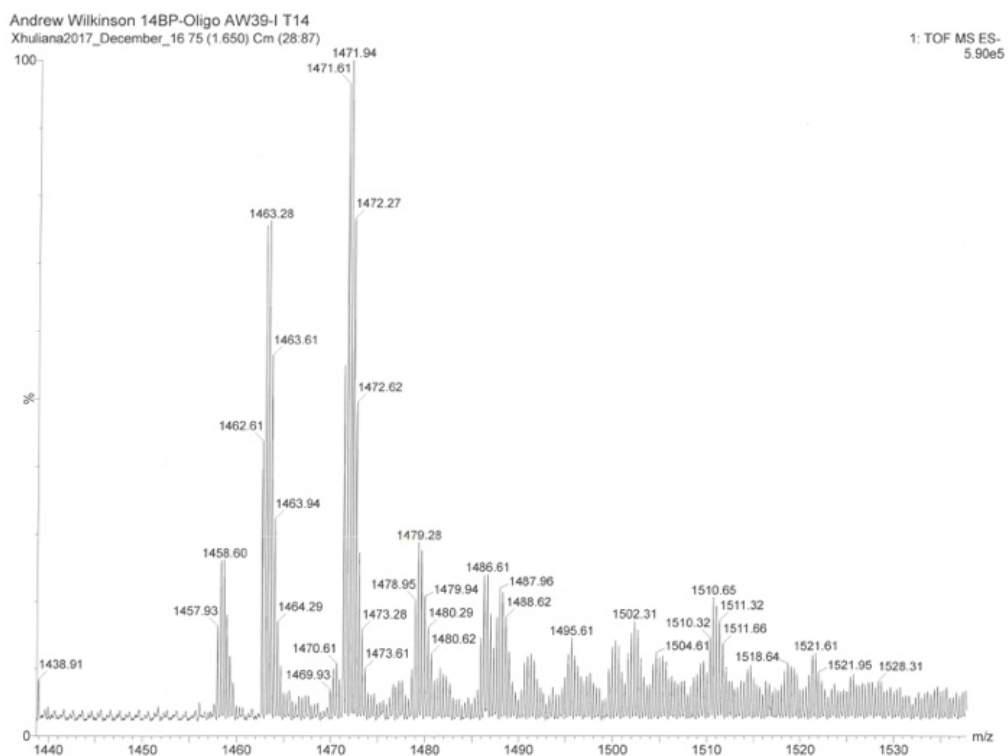


Figure S4.16. Mass spectrum of the peak at 14 min from Figure S4.14. Mass = $(1463.28 \times 3) + 3 = 4392.84 \text{ g mol}^{-1}$

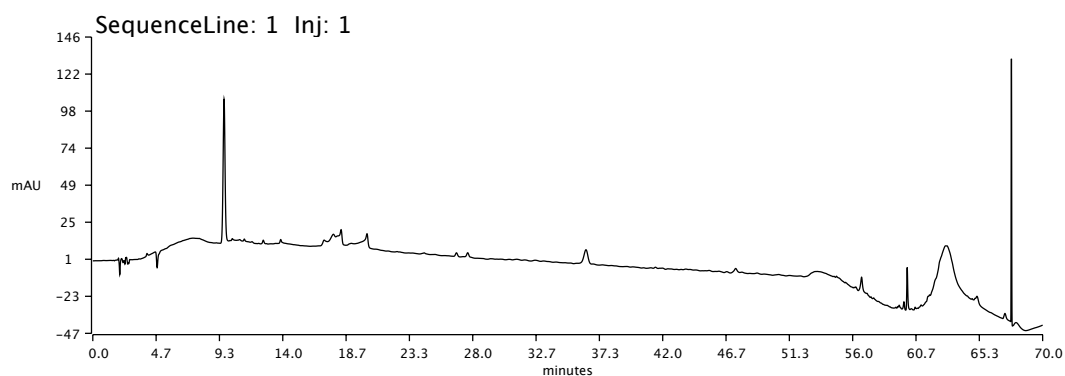


Figure S4.17. Analytical-HPLC chromatogram of 14 bp oligo DNA incubated in the presence of *M. TaqI* and **AdoHcy-8-Hyd**, followed by treatment with THBA (10 equivs) in a pH 4 10 mM ammonium acetate buffer. Conditions for analytical HPLC: System B

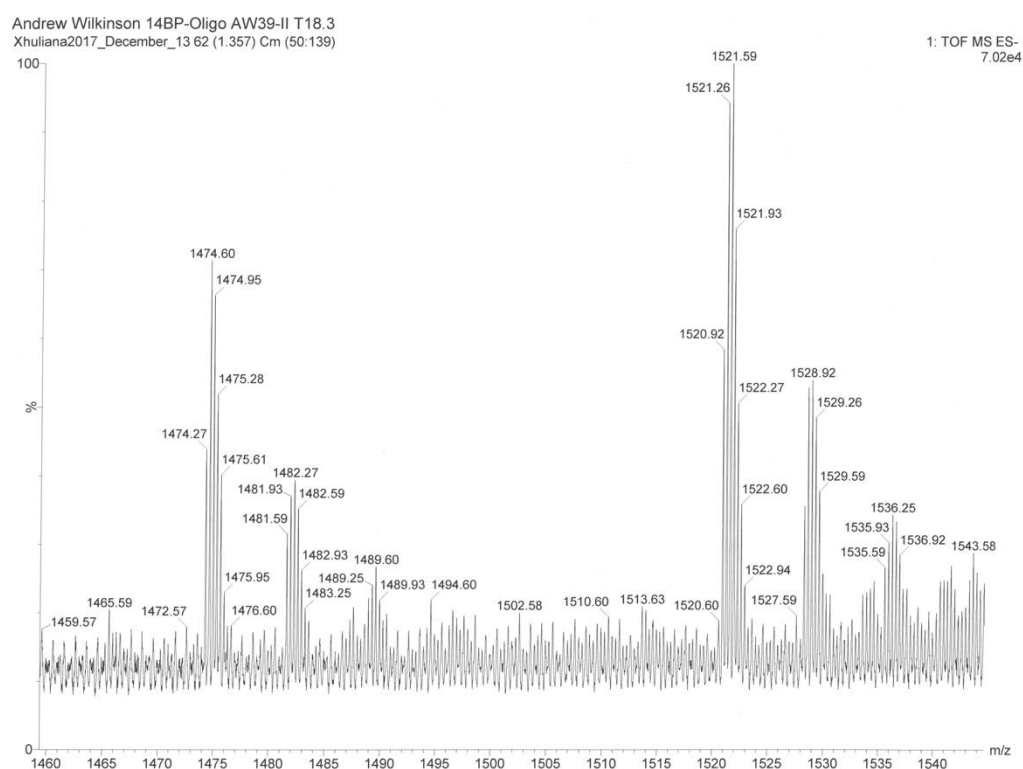


Figure S4.18. Mass spectrum of the peak at 18.3 min from Figure S4.17 Mass = $(1474.60 \times 3) + 3 = 4426.8 \text{ g mol}^{-1}$

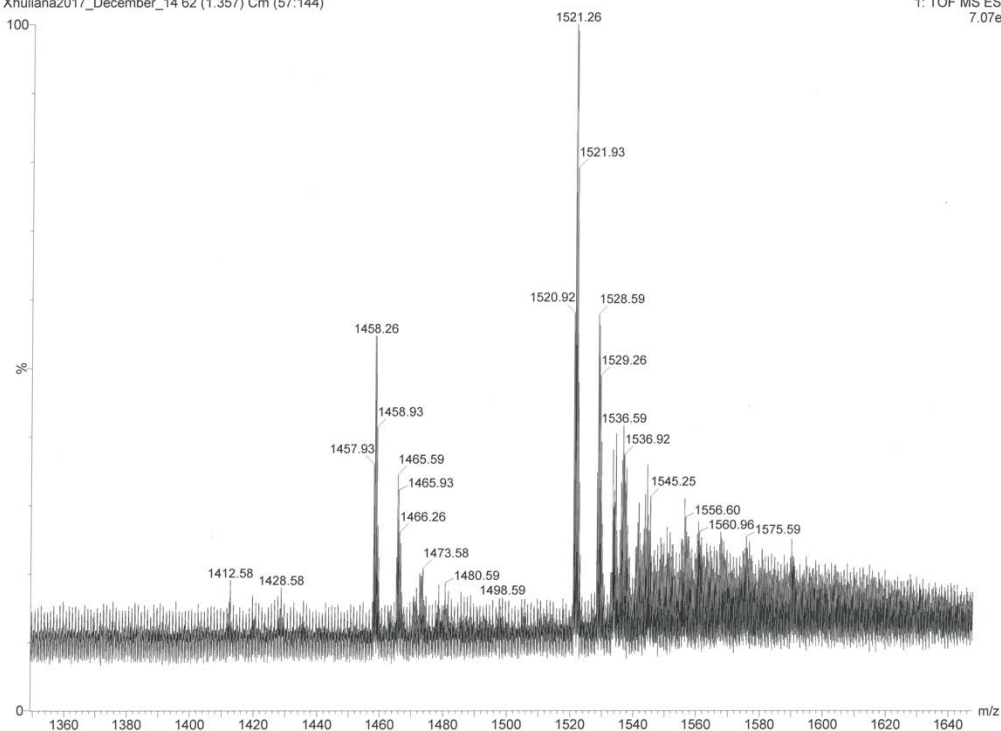


Figure S4.19. Mass spectrum of the peak at 20.3 min from Figure S4.17 Mass = $(1458.26 \times 3) + 3 = 4377.8 \text{ g mol}^{-1}$

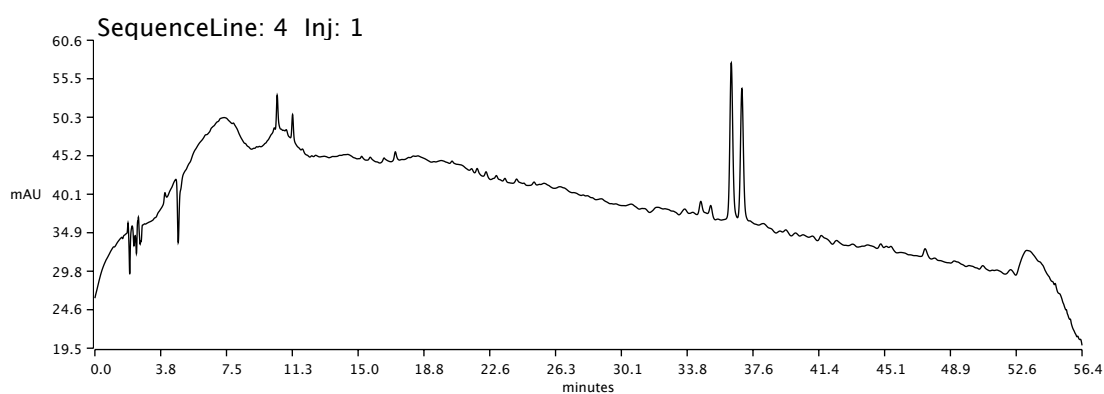


Figure S4.20. Analytical-HPLC chromatogram of 14 bp oligo DNA incubated in the presence of *M. TaqI* and **AdoHcy-8-Ox**, followed by treatment with $\text{H}_2\text{NOH.HCl}$ (10 equivs) in a pH 4 10 mM ammonium acetate buffer. Conditions for analytical HPLC: System B

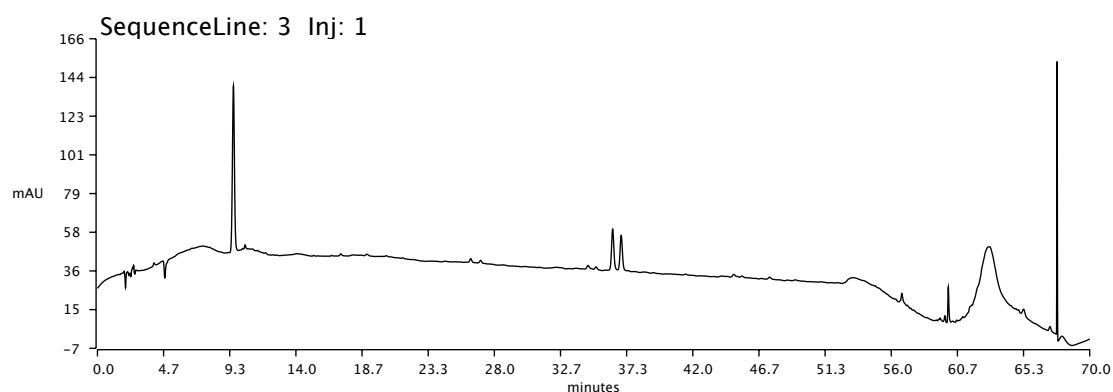


Figure S4.21. Analytical-HPLC chromatogram of 14 bp oligo DNA incubated in the presence of *M.TaqI* and **AdoHcy-8-Ox**, followed by treatment with THBA (10 equivs) in a pH 4 10 mM ammonium acetate buffer. Conditions for analytical HPLC: System B

4.5.1.2 LCMS reversal and relabeling

Andrew Wilkinson 14bp + AW39
Xhuliana2018_June_286

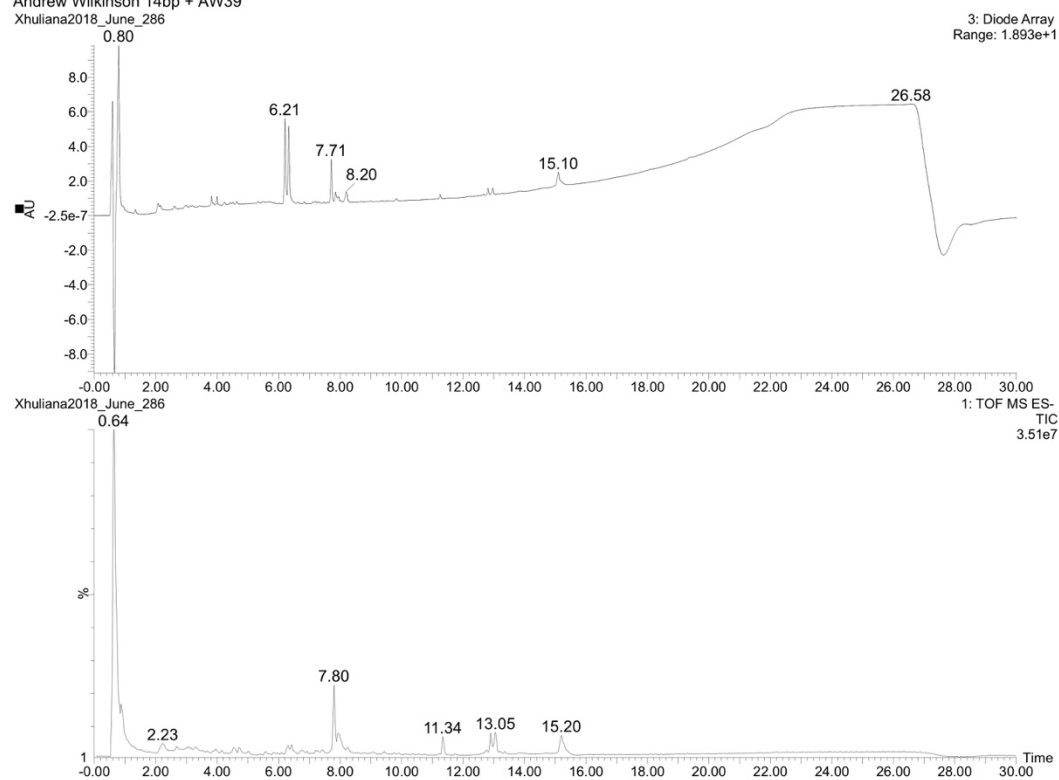


Figure S4.22. LCMS chromatogram of 14 bp oligo DNA incubated in the presence of *M.TaqI* and **AdoHcy-8-Hyd**. Top Diode array. Bottom TIC.

Andrew Wilkinson 14bp + AW39
Xhuliana2018_June_286 292 (6.304) Cm (289:292)

1: TOF MS ES-
7.11e4

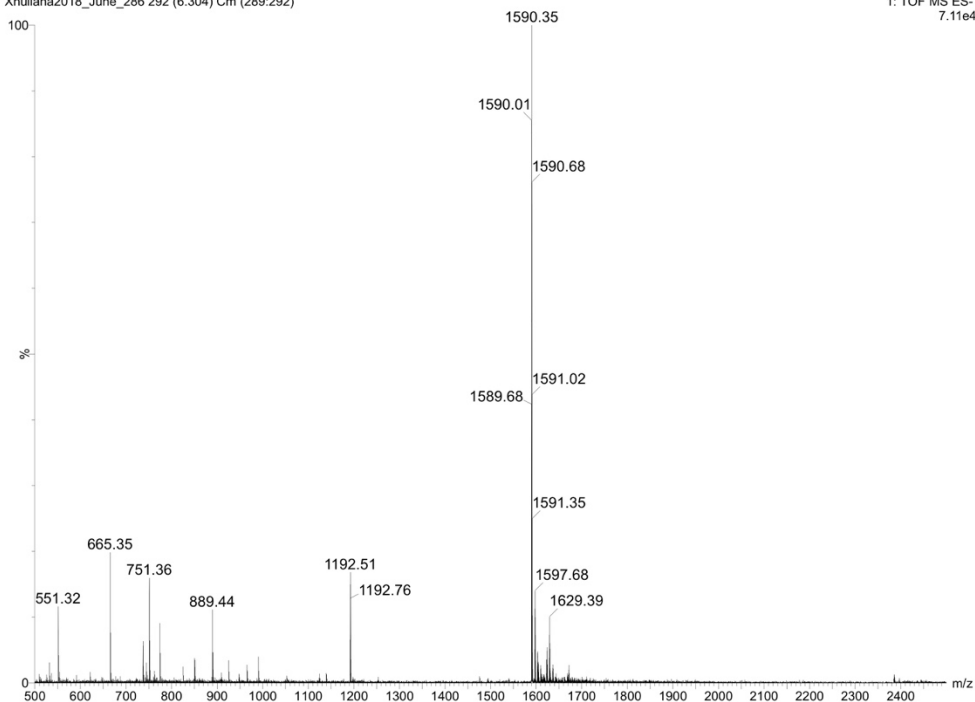


Figure S4.23. Mass spectrum of the peak at 6.3 min from Figure S4.22 Mass = $(1590.01 \times 3) + 3 = 4773.03 \text{ g mol}^{-1}$

Andrew Wilkinson 14bp + AW39
Xhuliana2018_June_286 297 (6.425) Cm (297:299)

1: TOF MS ES-
5.12e4

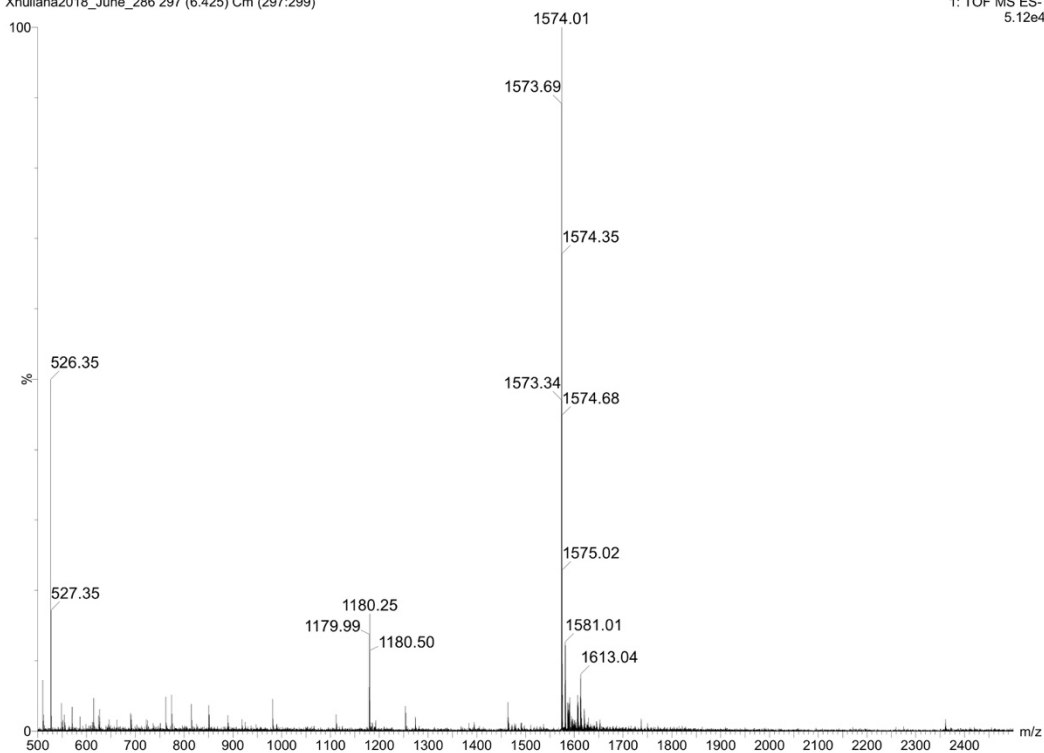


Figure S4.24. Mass spectrum of the peak at 6.4 min from Figure S4.22 Mass = $(1573.69 \times 3) + 3 = 4724.07 \text{ g mol}^{-1}$

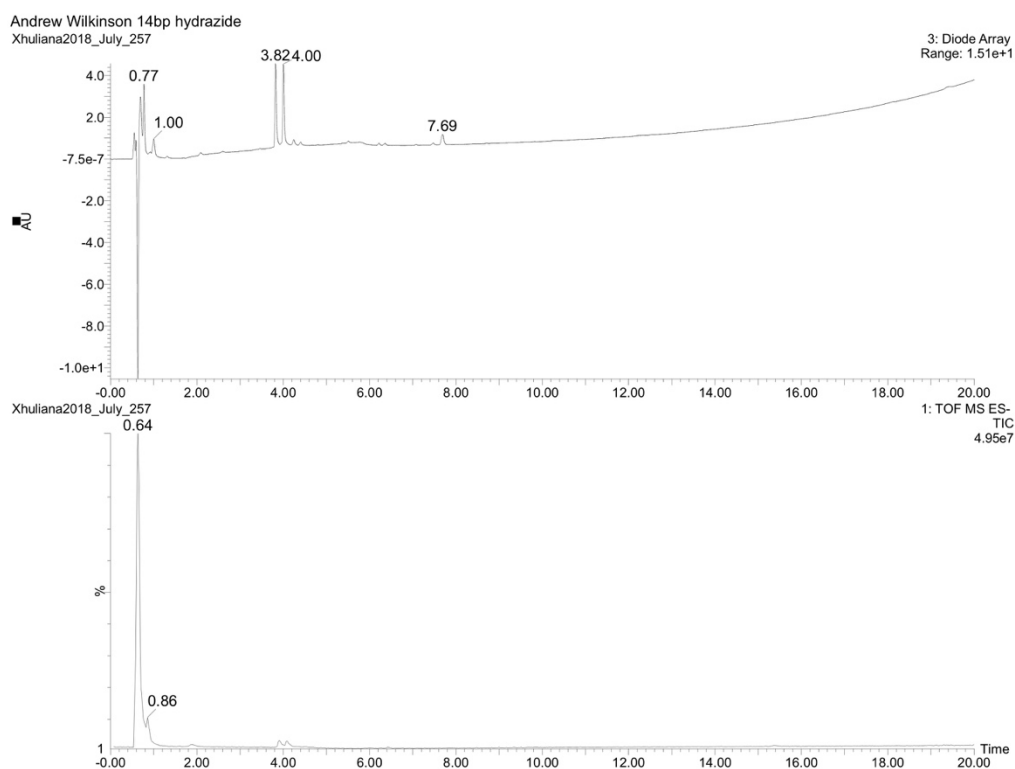


Figure S4.25. LCMS chromatogram of 14 bp oligo DNA incubated in the presence of *M.TaqI* and **AdoHcy-8-Hyd** followed by treatment with $H_2NOH.HCl$ (100 equivalents) in 10 mM Ammonium acetate pH 5.0 for 1 hr 50 °C. Top Diode array. Bottom TIC.

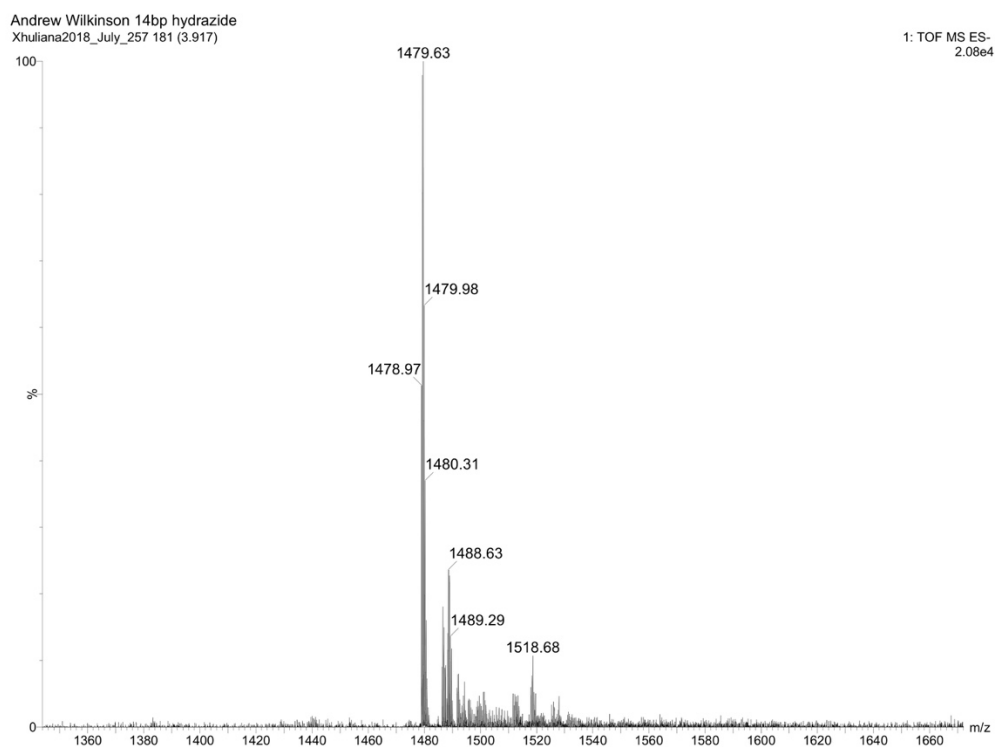


Figure S4.26. Mass spectrum of the peak at 3.8 min from Figure S4.25 Mass = $(1479.63 \times 3) + 3 = 4441.9 \text{ g mol}^{-1}$

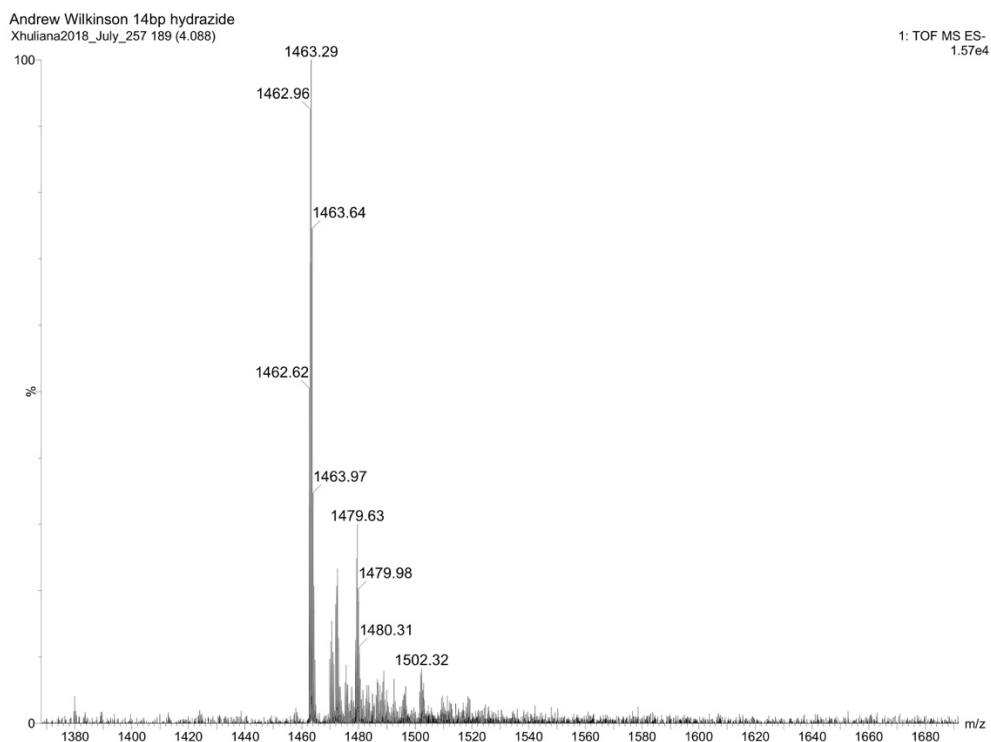


Figure S4.27. Mass spectrum of the peak at 4 min from Figure S4.25 Mass = $(1462.96 \times 3) + 3 = 4391.9 \text{ g mol}^{-1}$

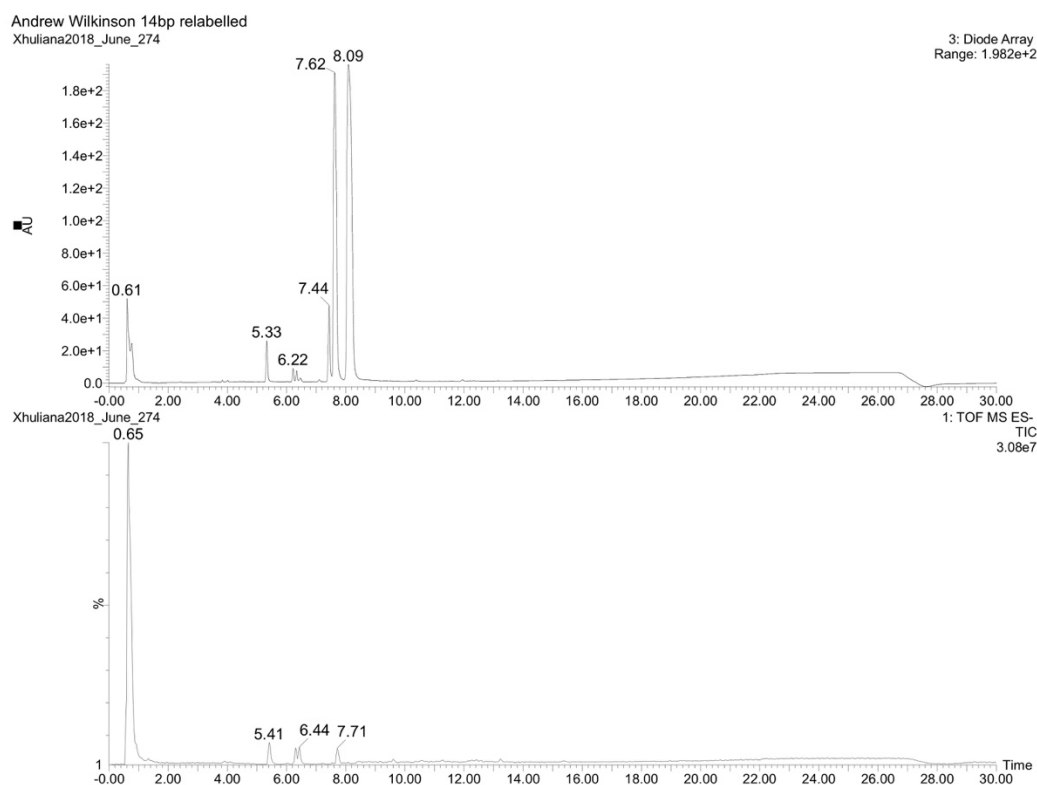


Figure S4.28. LCMS chromatogram of 14 bp oligo DNA incubated in the presence of *M.TaqI* and **AdoHcy-8-Hyd** followed by treatment with $\text{H}_2\text{NOH.HCl}$ (100 equivalents) in 10 mM Ammonium acetate pH 5.0 for 1 hr 50 °C. After reversal confirmation DNA was then relabelled by addition of Ald-Ph-PEG₃-N₃ (200 equivalents) Top Diode array. Bottom TIC.

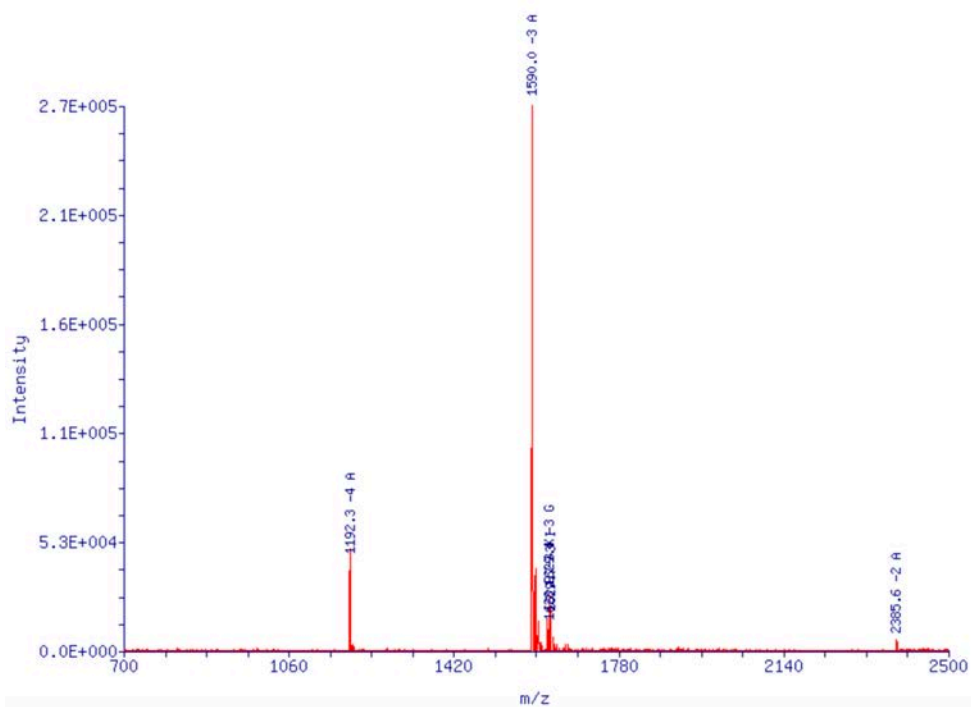


Figure S4.29. Mass spectrum of the peak at 6.22 min from Figure S4.28 Mass = $(1590.0 \times 3) + 3 = 4773 \text{ g mol}^{-1}$

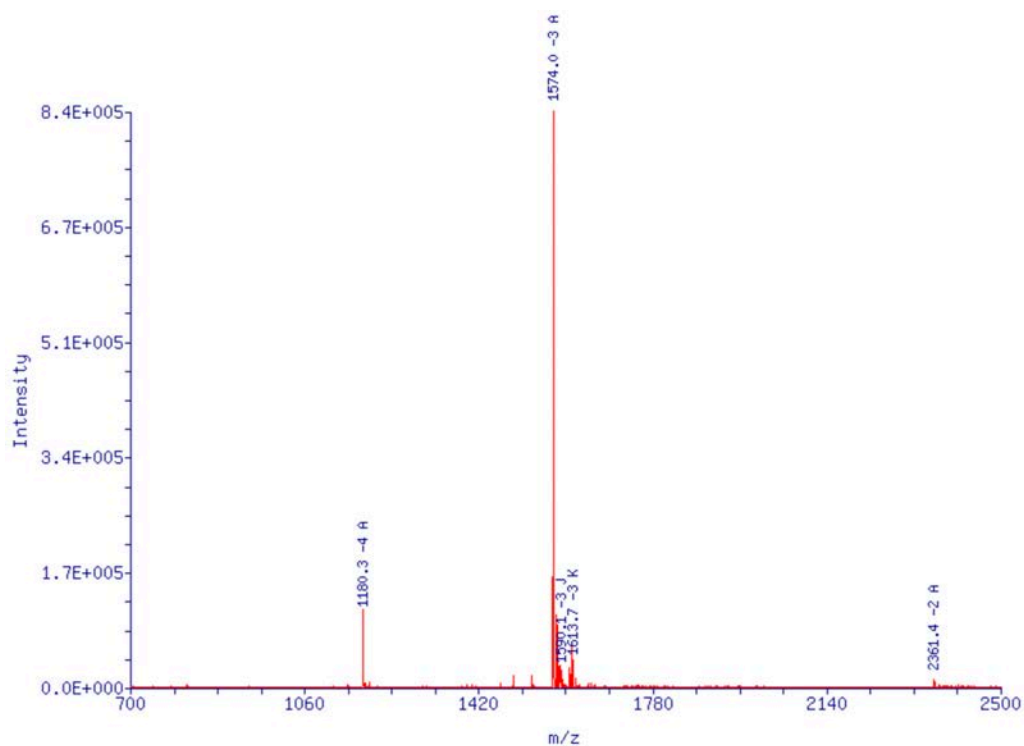


Figure S4.30. Mass spectrum of the peak at 6.40 min from Figure S4.28 Mass = $(1574.0 \times 3) + 3 = 4725 \text{ g mol}^{-1}$

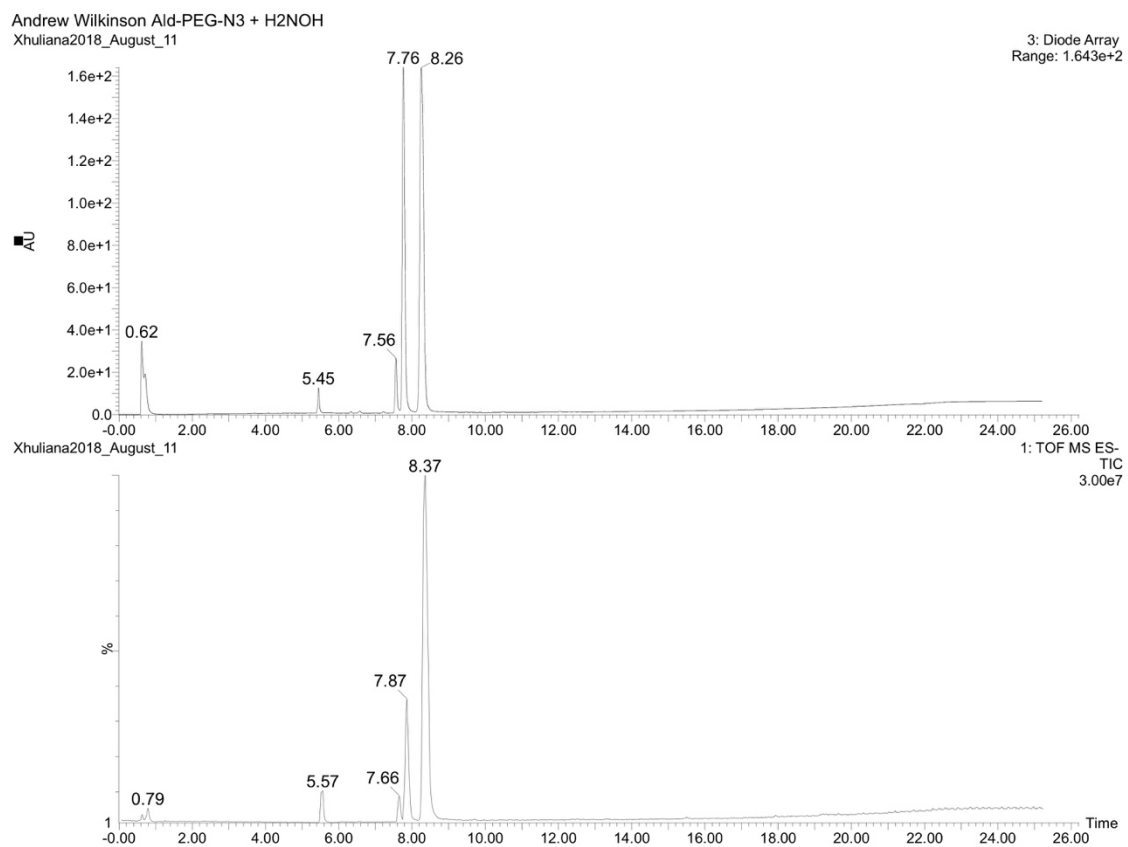


Figure S4.31. LCMS chromatogram of Ald-Ph-PEG₃-N₃ mixed with H₂NOH.HCl in ammonium acetate buffer Top Diode array. Bottom TIC.

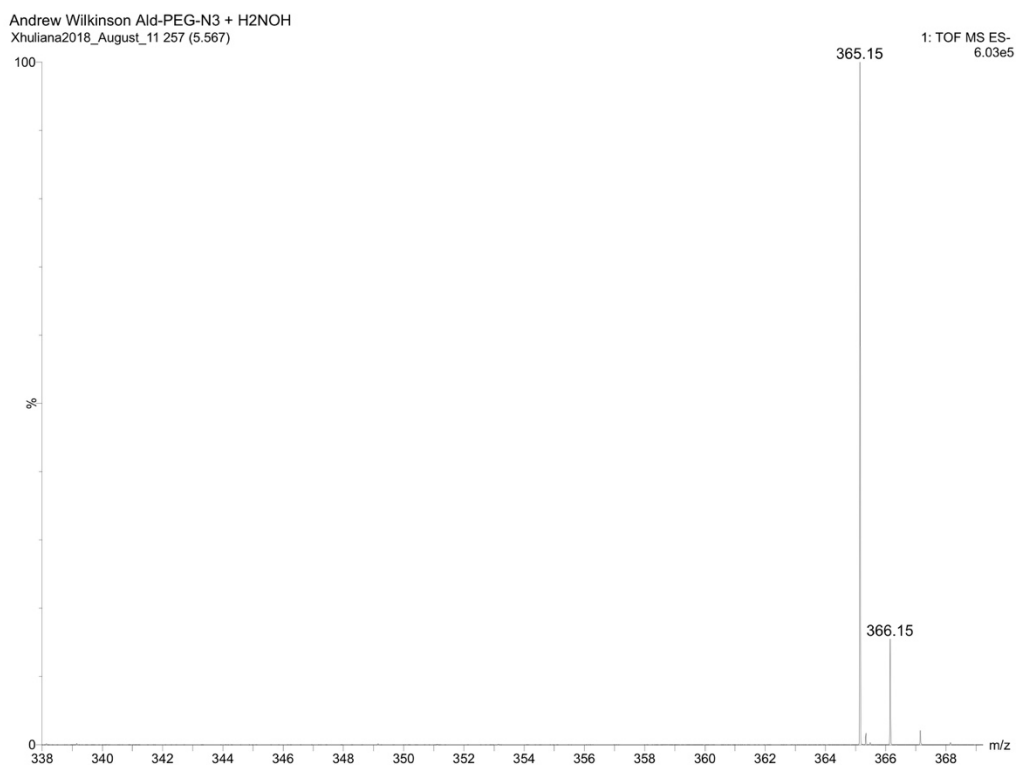


Figure S4.32. Mass spectrum of the peak at 5.57 min from Figure S4.31

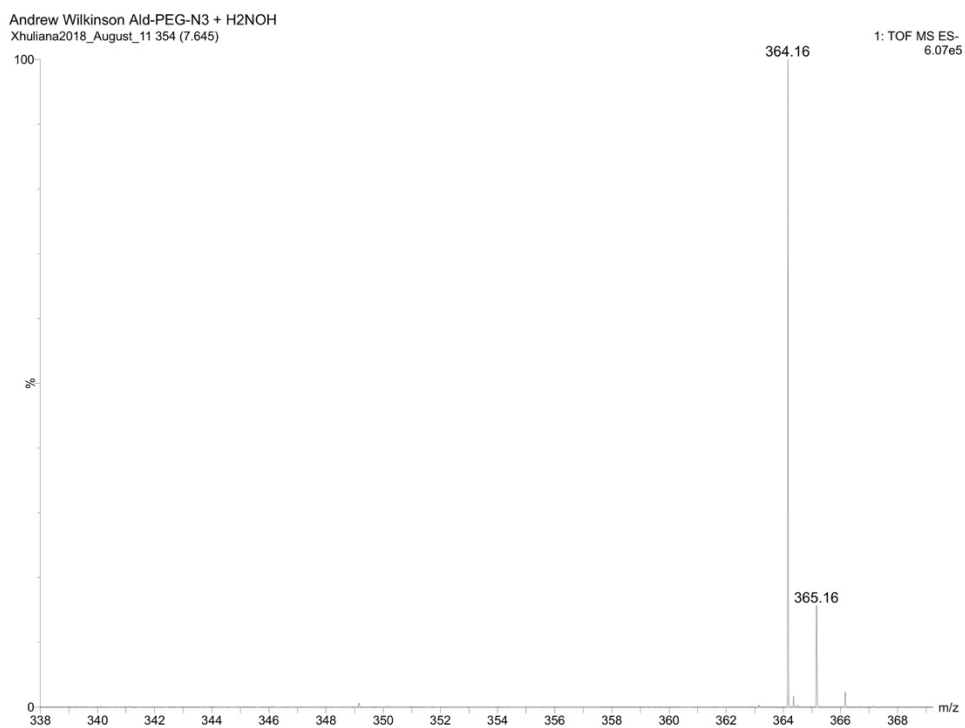


Figure S4.33. Mass spectrum of the peak at 7.65 min from Figure S4.31

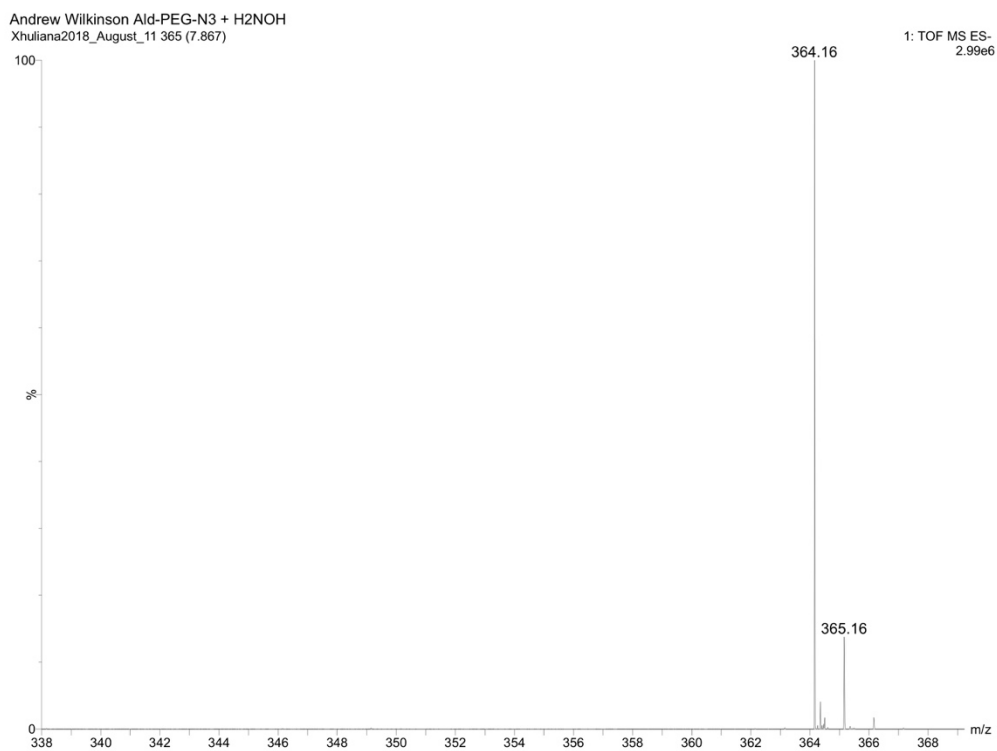


Figure S4.34. Mass spectrum of the peak at 7.87 min from Figure S31

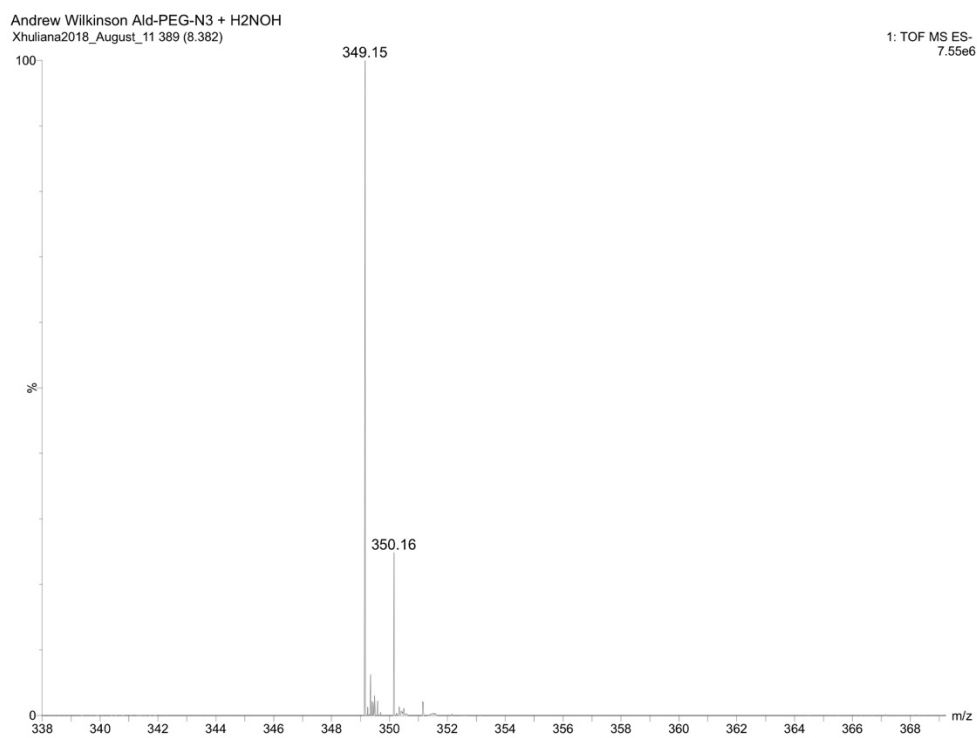


Figure S4.35. Mass spectrum of the peak at 8.38 min from Figure S4.31

4.5.2 PCR product amplification, transalkylation, capture and release

4.5.2.1 Amplification

The PCR product was prepared by amplification of a section of the plasmid pUC19 using the corresponding primers (Forward primer: 5'-GCC AGG AAC CGT AAA AAG-3' and Reverse primer: 5'-AGA AAG GCG GAC AGG TAT-3'). The DNA fragment amplified was 203 bp with 31 CpG sites. For amplification a mastermix of buffer (200 µl, NEB 2x highfidelity mastermix), Forward primer (4 µl, 100 µM), reverse primer (4 µl, 100 µM), pUC19 (2 µl, 10 ng/µl) and water (190 µl) was prepared. The mastermix was then aliquoted into 50 µl portions. Amplification was completed using the programme: 98 °C for 45 s, 40 cycles 98 °C for 15 s, 62 °C for 45 s, and 72 °C for 60 s. After amplification the DNA was purified on silica column (Sigma-Aldrich PCR clean up kit) and eluted into water.

4.5.2.2 Transalkylation

Once collected the DNA fragment was labelled using the CpG specific DNA methyltransferase M.MpeI. For transalkylation samples were made up of: DNA (66 µl, 40 ng/ µl), buffer (10 µl, 10x NEB cutsmart buffer), Cofactor (3 µl, 15 mM), M.MpeI (5 µl, 1.7 mg/ml), and water (16 µl). Samples were incubated at 37 °C for 1 hr after which proteinase K (2 µl) was added and incubated at 50 °C for 1 hr. Finally, a solution of diazo-DBCO-Biotin (2 µl, 50 mM, Jenabioscience) was added and incubated for a further 1 hr at 37 °C. After labelling the DNA was purified using a Qiagen purification kit and the DNA was eluted into a high salt Tris buffer (10 mM Tris, 1 M NaCl, pH 8.5, Tris A) ready for capture.

4.5.2.3 General Capture protocol

After transalkylation the DNA was captured onto streptavidin coated magnetic beads (dynabeads myone C1). For capture 10 µl of the bead stock was washed 2x with 70 µl Tris A and then all supernatant was removed. The bead pellet was then resuspended in the DNA solution previously eluted (500 ng, 70 µl, Tris A). This solution was shaken at room temperature for 15 minutes. After capture the bead mixture was placed on a magnetic rack and the DNA concentration of the eluent was measured. The eluent was then placed into a fresh bead pellet to capture the remaining DNA. This process was repeated 3x until the majority of DNA was captured. After capture the bead pellets were resuspended in water and washed 2x. The beads were then finally placed within the relevant buffer ready for release.

4.5.2.4 General release protocol

To release captured DNA (500 ng, 200 bp) the bead pellets were suspended in Ammonium Acetate buffer (63 µl, 11.1 mM, 1 M NaCl, 0.01% SDS, var. pH) and to this a solution of H₂NOH.HCl (7 µl, var. equivs.) was added. The bead solution was then shaken at 50 °C for 1 hr. After shaking the solution was placed on a magnet, the supernatant was removed and the DNA concentration was measured. This process was repeated until all DNA had been released.

4.5.3 PCR rewriting protocol and gel electrophoresis

6 µg PCR product was labelled with **AdoHcy-8-Hyd** in the same way as above, at each stage in the transalkylation process 600 ng was aliquoted, purified and stored for fluorophore coupling. After release each 600 ng aliquots was split into 2 one to be labelled and one to keep as an unlabelled control. Into 6 samples

were set up: unlabelled, unlabelled + Atto-488, labelled, labelled + Atto-488, Released, Released + Atto-488. To label each sample Atto-488 (1.5 equivalents) was added to each sample in DMSO followed by the addition of DMSO (5 μ l) to give a final volume of 25 μ l. Samples were then incubated in the fridge overnight before purification using Qiagen Qiaquick nucleotide clean up kits with elution in to water. After purification each sample was run on an 1% agarose gel at 120 V for 50 minutes before staining with Gel-red and final imaging under UV and blue light transillumination.

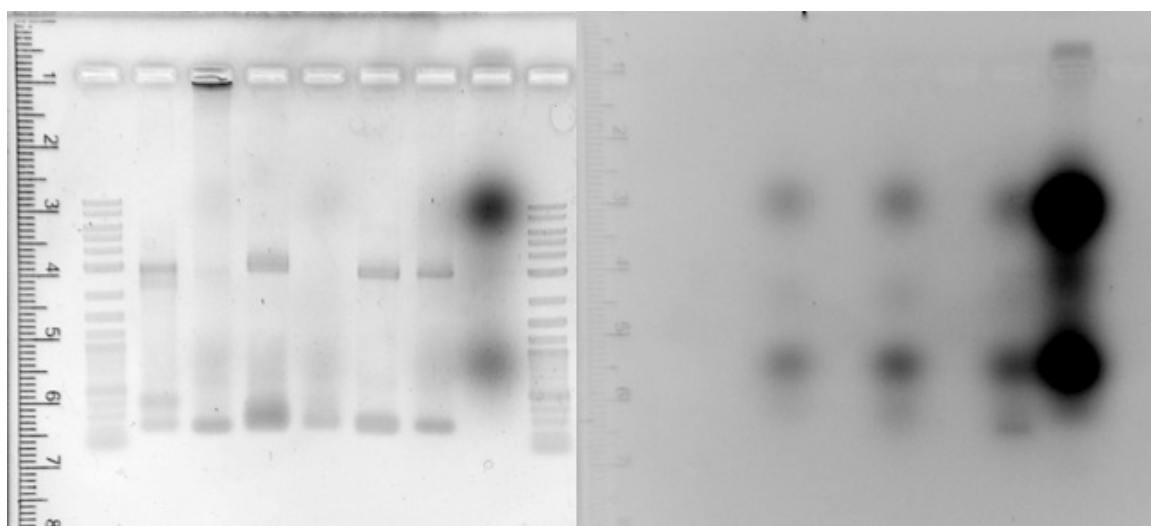


Figure S36. Uncropped Gel electrophoresis results after rewriteability test discussed in fig. 4.10. Left = UV-Transillumination. Right = Blue light illumination.

4.5.4 Hydroxymethylated and unmethylated capture and release protocol

4.5.4.1 Amplification

The hydroxymethylated PCR product was prepared in the same way as for the unmodified PCR product but by using an enriched stock containing modified 5'-hydroxymethylcytosine instead of cytosine.

4.5.4.2 Hydroxymethylated DNA transalkylation

Hydroxymethylated DNA (26 μ l, 64 ng/ μ l) was labelled by addition of T4 Phage β -glucosyltransferase (10,000 units/ml) in the presence of; cofactor (6-N₃-Uridine

diphosphoglucose, 3 mM, 1 μ l), new England biolabs buffer 4 (5 μ l), and water (17 μ l). DNA was incubated for 1 hour at 37 °C followed by addition of 1.5 μ l proteinase K and further incubation for 1 hour at 50 °C. Finally, a solution of diazo-DBCO-Biotin (2 μ l, 50 mM, Jenabioscience) was added and incubated for a further 1 hr at 37 °C. After transalkylation the DNA was purified using a Qiagen purification kit and the DNA was eluted into a high salt Tris buffer (10 mM Tris, 1 M NaCl, pH 8.5, Tris A) ready for capture.

Subsequent capture and release of all DNA mixtures was then completed in the same way as for unmethylated DNA samples using the protocols outlined in sections 4.5.3.3/4.5.3.4.

5 Conclusions and Future Outlook

I have demonstrated the successful design and synthesis of a small variety of novel AdoMet cofactor analogues for use in the methyltransferase-directed transalkylation of DNA. Each AdoMet analogue was then tested directly with two methyltransferases of interest, M.TaqI and M.MpeI, to check their efficacy as cofactor analogues in the transalkylation of DNA with methyltransferases. Oligo HPLC was used to further test each cofactor analogue and develop an understanding around the transalkylation process and what each methyltransferase introduced to DNA. This helped conclusively prove alkylation with doubly-activated analogues behaved as it does with methylation symmetrically on both the forward and reverse strand¹⁶⁴.

Although cysteine cofactor analogues had significantly compromised transalkylation efficiencies compared to their homo-cysteine equivalents, it is hoped these can be taken forward for their use with mutant methyltransferases during *in vitro* transalkylation studies. On the other hand, both Schiff base analogues showed robust and reproducible alkylation, with similar efficiencies compared to the more classical analogue **AdoHcy-6-N₃**. After developing confidence in their alkylation, we then tested each Schiff base cofactor analogue in a series of dynamic bioconjugation studies. Here we first identified if reversal was possible followed by testing the control and flexibility that could be achieved. After developing an understanding of reversal, proof of concept studies were run to highlight some of the potential applications for this alkylation technique.

Overall the Schiff base cofactors developed show great promise as tools for dynamic bioconjugation studies. We hope this is an area of growing interest as

not only do they effectively transalkylate DNA and allow for subsequent modification cycles, but they also combine the benefits traditionally afforded from dynamic covalent chemistry with those seen for methyltransferase-directed transalkylation. Particularly of interest for this project were the simplicity, yields and repeated reversal opportunities brought by the use of Schiff base chemistry and the specificity and diversity of methyltransferases brought by the natural range that exist and their sequence-specific targeting. It is hoped this technique can be taken forward for the development of commercially available kits to be used as a competitor to current transalkylation techniques for biomolecules, like biotin/streptavidin capture and release. Additionally, the scope of this technique also needs to be expanded to a wider variety of methyltransferases and alternative biomolecules like proteins and RNA. Through further exploration new applications will be identified which were not previously possible using methyltransferase-directed transalkylation alone.

One application that has already shown success for this transalkylation technique is in epigenomic sequencing studies. Following the development of the proof of concept study showing DNA can be successfully captured and released using methyltransferases and hydrazone reversal, this chemistry was then used within our group on genomic DNA.

Initially the selectivity of this method was tested on fragmented samples of *E.coli* genomic DNA using M.TaqI where near quantitative capture and release yields could be achieved. After capture and release the DNA was purified and then prepared for nanopore sequencing. Once sequenced the individual reads could

then be aligned against the reference genome to monitor the selectivity seen (Figure 5.1). When using M.TaqI a clear relation could be seen between the location of TCGA sites and the quantity of fragments that overlaid with these regions. This indicates transalkylation is site-selective and can be used to target genomic regions of interest in future studies.

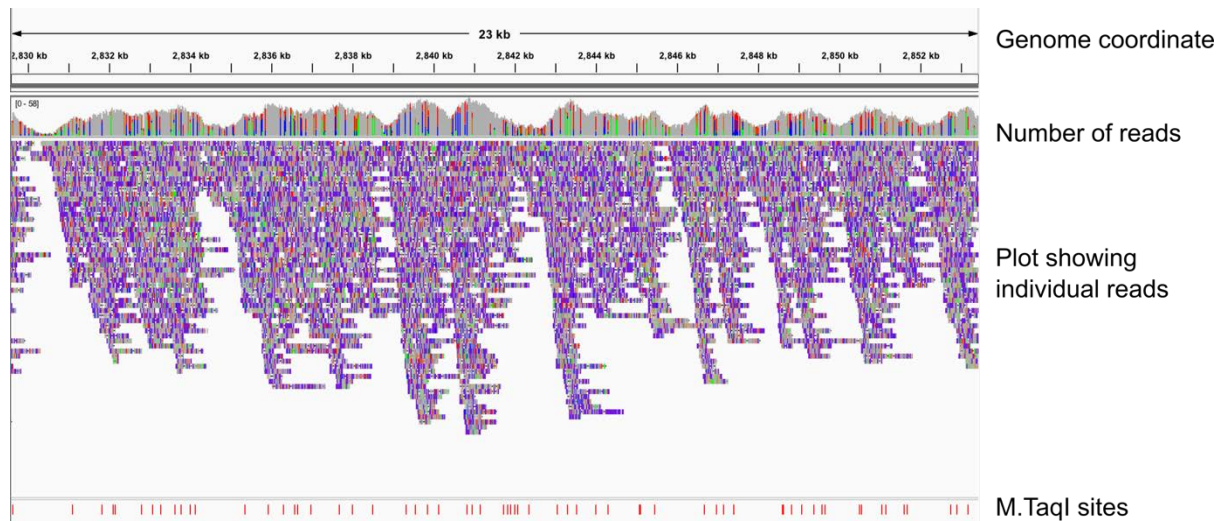


Figure 5.1 Shows a 23 kb section of data collected after nanopore sequencing of M.TaqI labelled DNA with **AdoHcy-8-hyd** and fragmented *E.coli* DNA after transalkylation, capture and release. Genome coordinate – identifies the region of the genome being examined. Number of reads – a histogram plot of all fragments that passed through the nanopore highlighting density of fragments that passed through from each region of the genome. Individual reads – Shows an individual plot of every DNA fragment that passed through the nanopore placed where they align against the reference genome. M.TaqI sites – Identifies the location of all M.TaqI sites within this section of the *E.coli* genome analysed.

One region of interest that has since been explored using this technique is CpG islands within the human genome as these can be identified using the CpG specific methyltransferase M.MpeI. These are of interest as they play a key role in gene expression and when hyper/hypo methylated have been linked directly to diseases like cancer³⁸. This was achieved using samples of human genomic head a neck cancer DNA. Unfortunately, due to the size of the human genome results cannot yet be relied upon as coverage was not extensive enough to

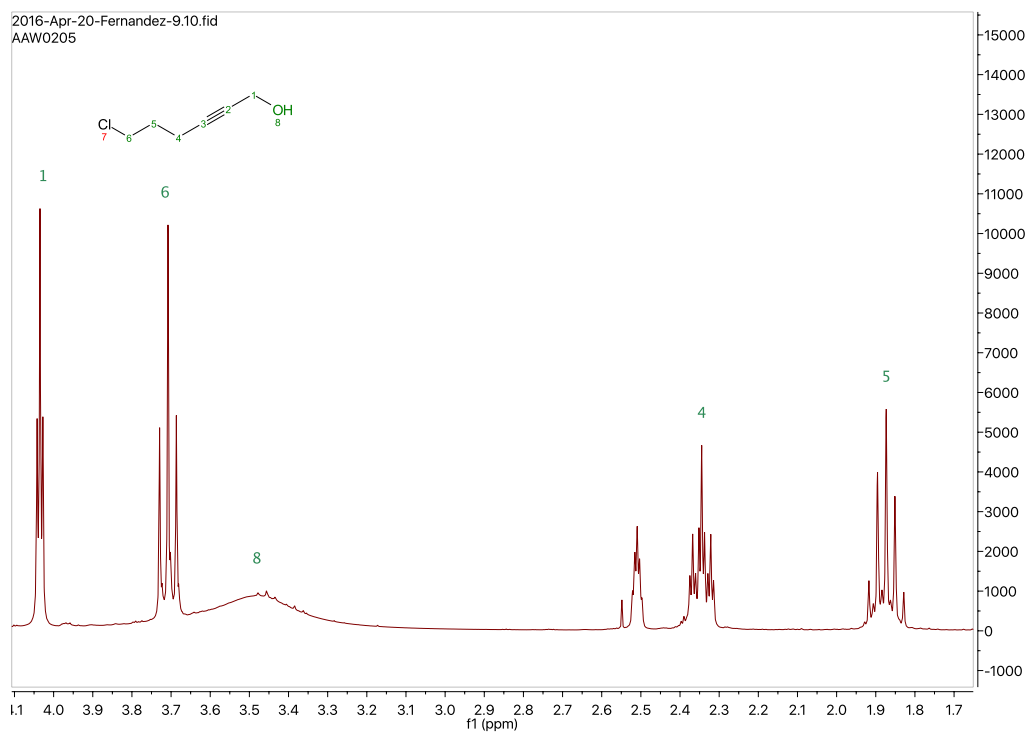
produce reliable results. In order to collect reliable data from nanopore sequencing it is necessary to get 30-40 times coverage of the genome, but with the human genome data only 1 times could be achieved due to its size. However, of the data that was collected there was a clear preference towards unmethylated regions when aligning against the reference for this genome.

In the future this study will be expanded to nanopore sequencing devices which are capable of effectively sequencing the entire genome and hopefully give a thorough understanding of the unmethylated sections of this genome. Furthermore, it is also hoped the data can be further analysed to identify the exact position of each modification by analysing the voltage recorded for each base. However, the most significant advantage this technique gives over currently available epigenetic kits (e.g. bisulphite conversion) is data reduction. In bisulphite sequencing to get a complete understanding it is necessary to sequence the entire genome twice, with methyltransferase-directed techniques, this theoretically only needs to be done once.

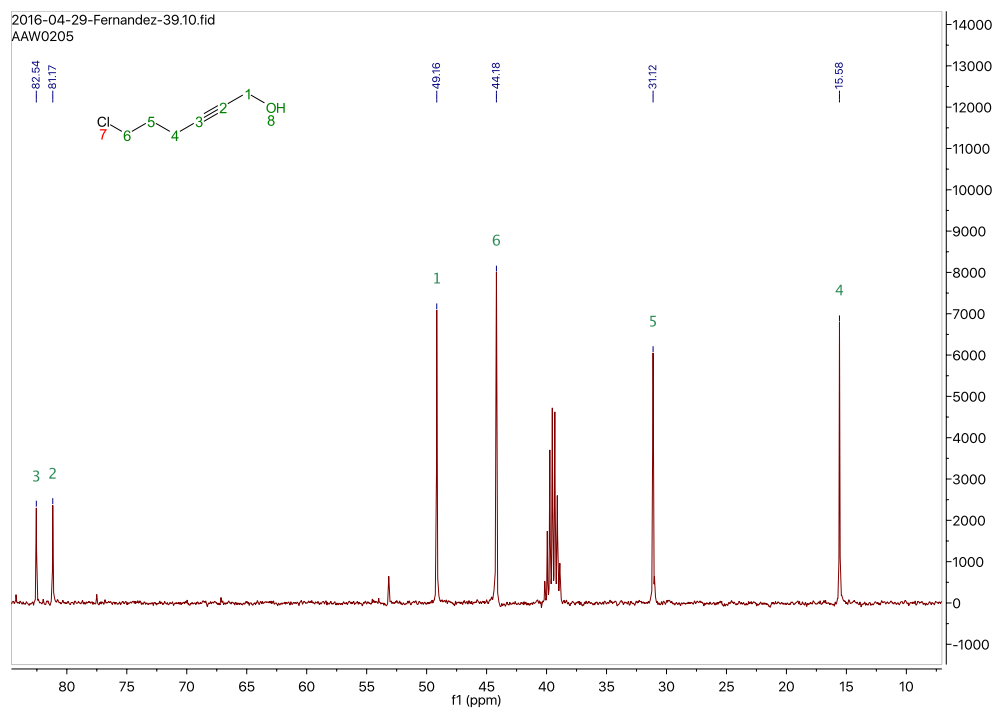
Other applications that have also been identified for this chemistry that will be explored further in the future are its use in super-resolution microscopy through controlled bleaching by *in situ* fluorophore removal; repeated modification studies combining optical DNA mapping with sequencing; and the selective capture and release of biomolecules from complex mixtures.

6 Appendix 1 - Compound Characterisation

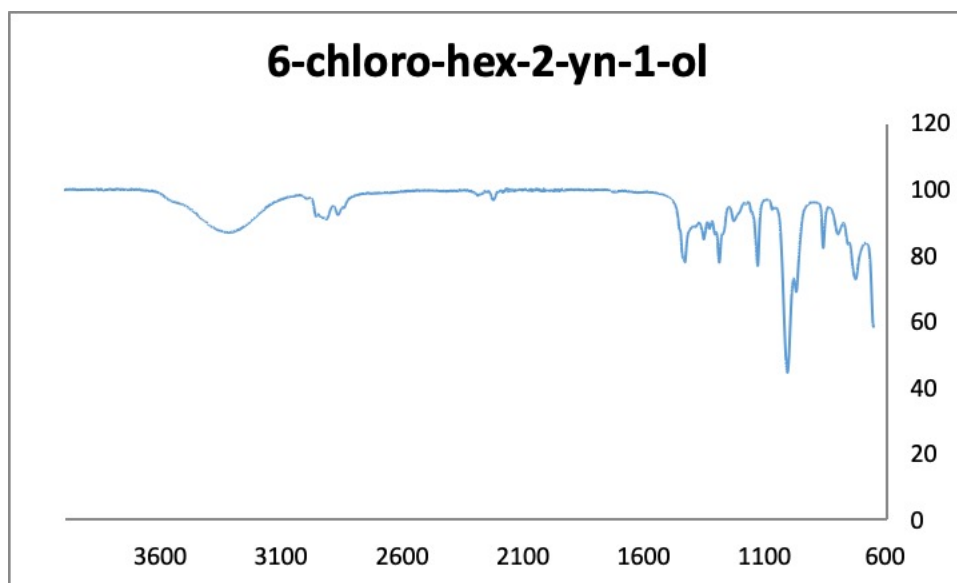
6-chloro hex-2-yn-1-ol (**2a**)



Sup. Fig. S6.1 ¹H NMR spectrum of 6-chloro hex-2-yn-1-ol (**2a**)

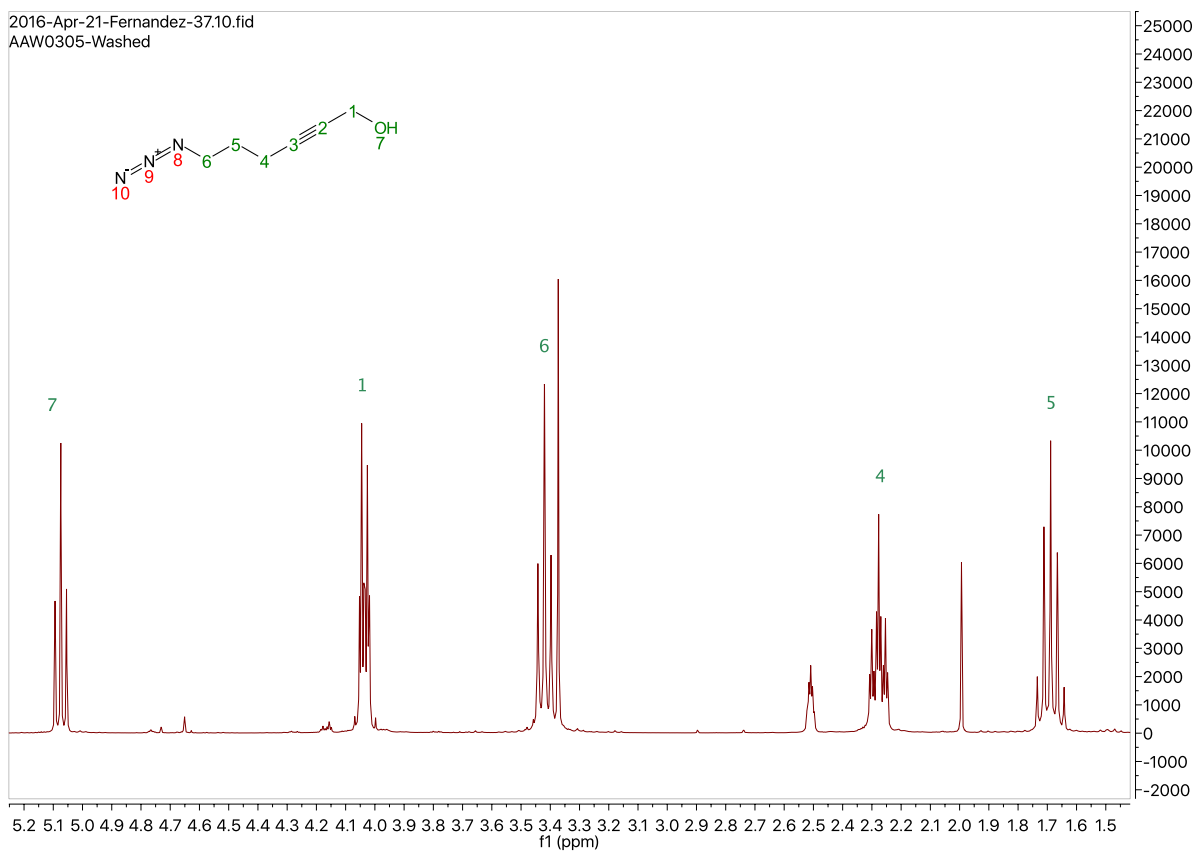


Sup. Fig. S6.2 ¹³C NMR spectrum of 6-chloro hex-2-yn-1-ol (**2a**)

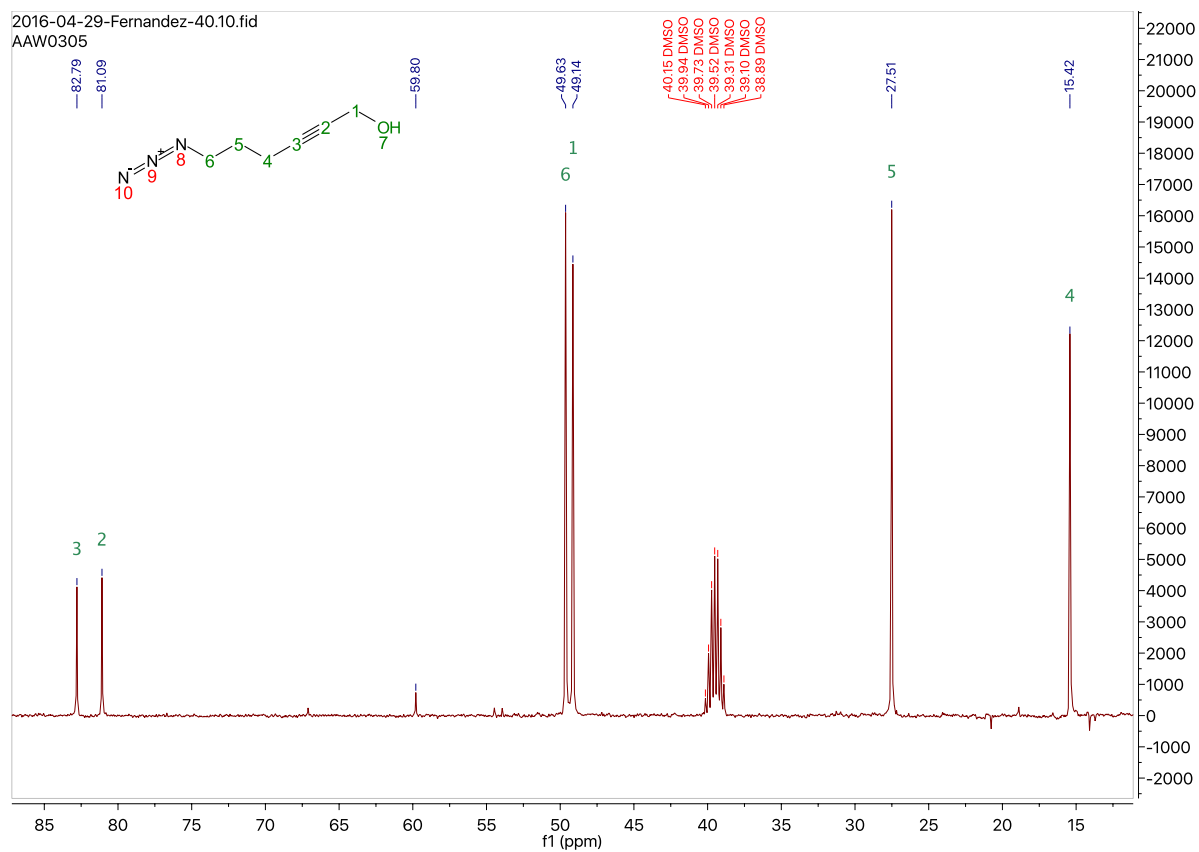


Sup. Fig. S6.3 Infrared spectrum of 6-chloro hex-2-yn-1-ol (**2a**)

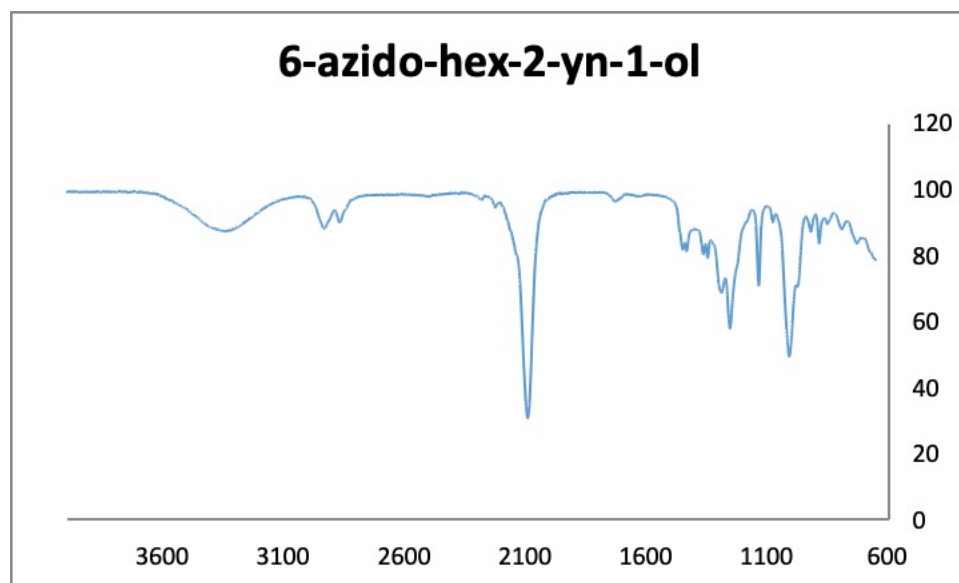
6-azido hex-2-yn-1-ol (**3a**)



Sup. Fig. S6.4 ^1H NMR spectrum of 6-azido hex-2-yn-1-ol (**3a**)

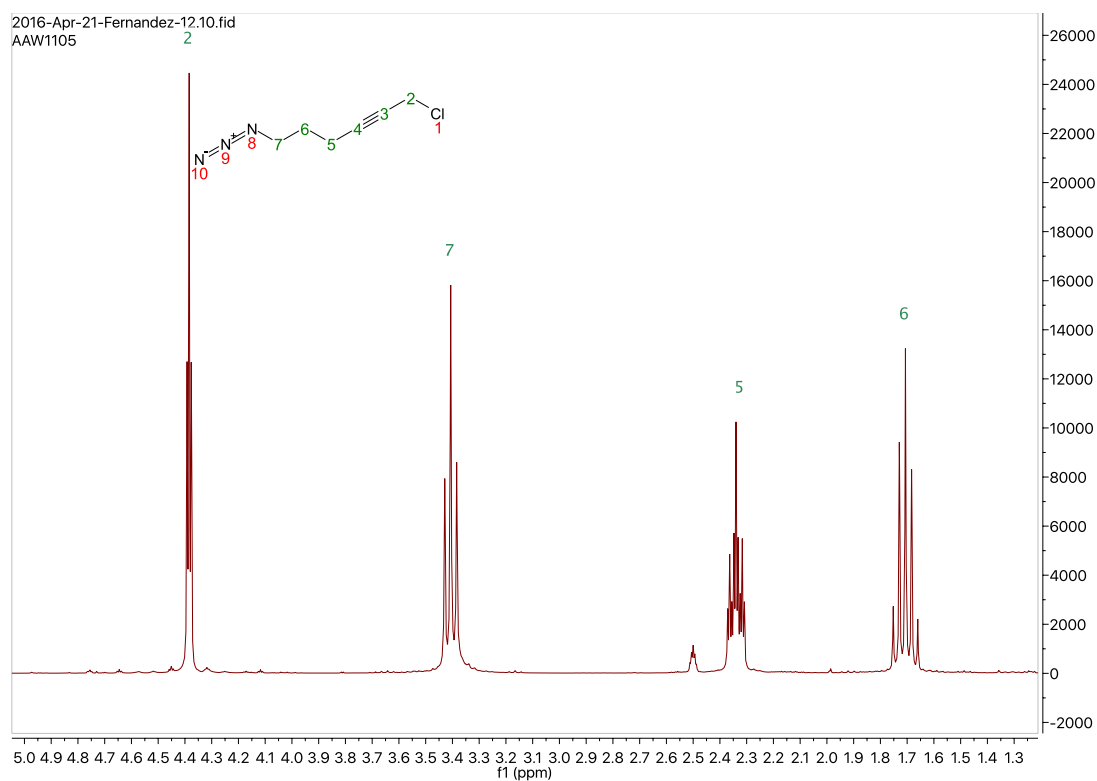


Sup. Fig. S6.5 ^{13}C NMR spectrum of 6-azido hex-2-yn-1-ol (3a)

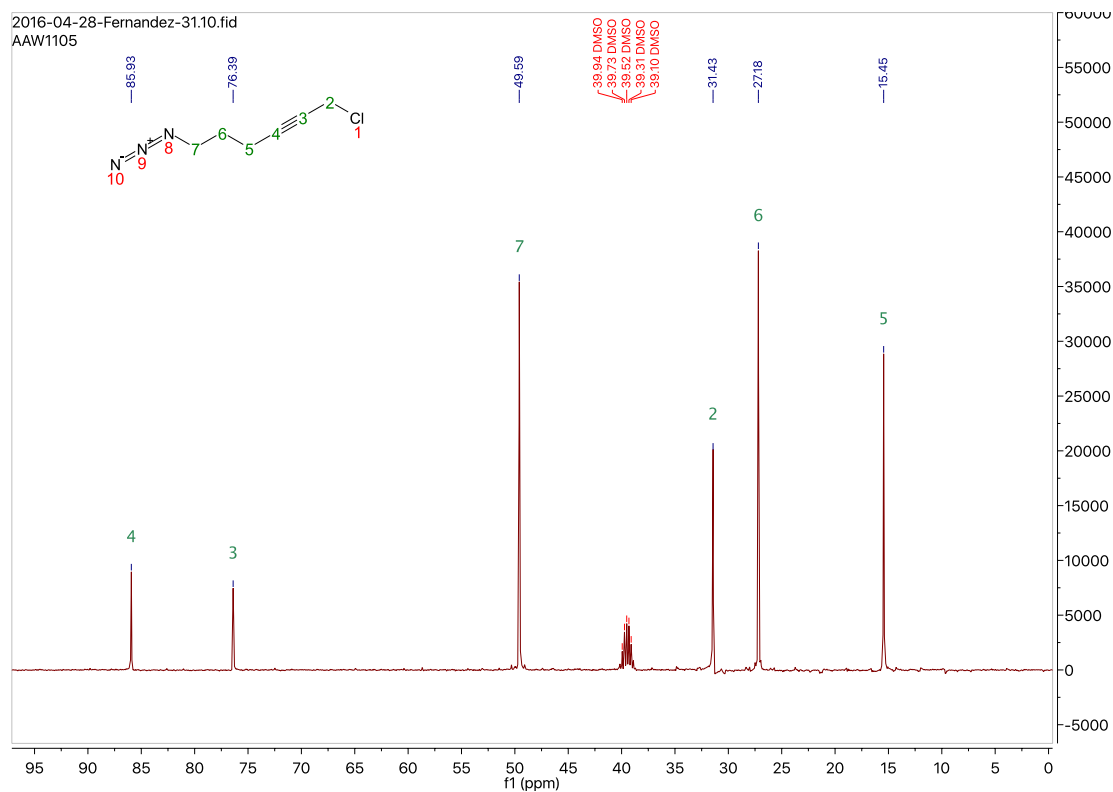


Sup. Fig. S6.6 Infrared spectrum of 6-azido hex-2-yn-1-ol (3a)

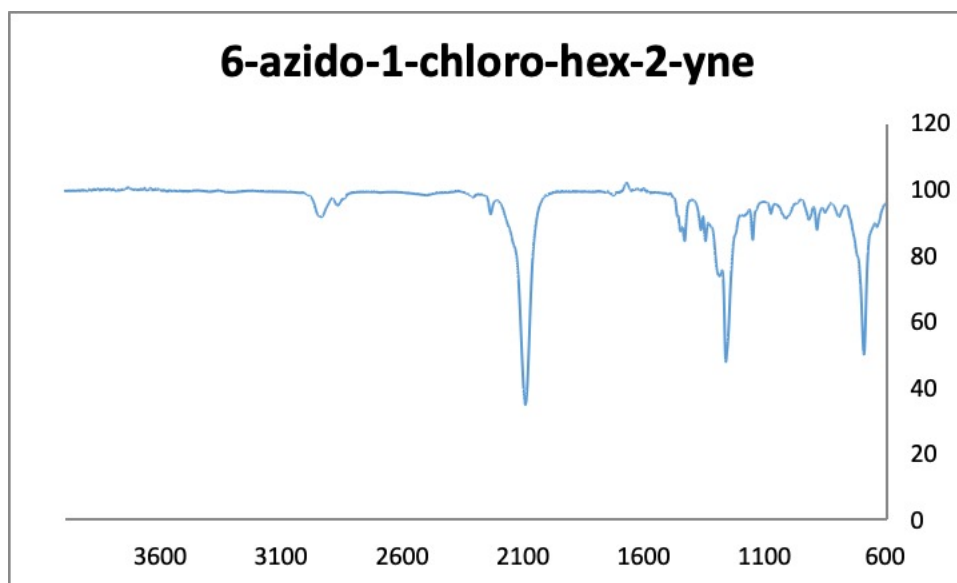
6-azido-1-chlorohex-2-yne (4ai)



Sup. Fig. S6.7 ¹H NMR spectrum of 6-azido-1-chloro hex-2-yne (4ai)

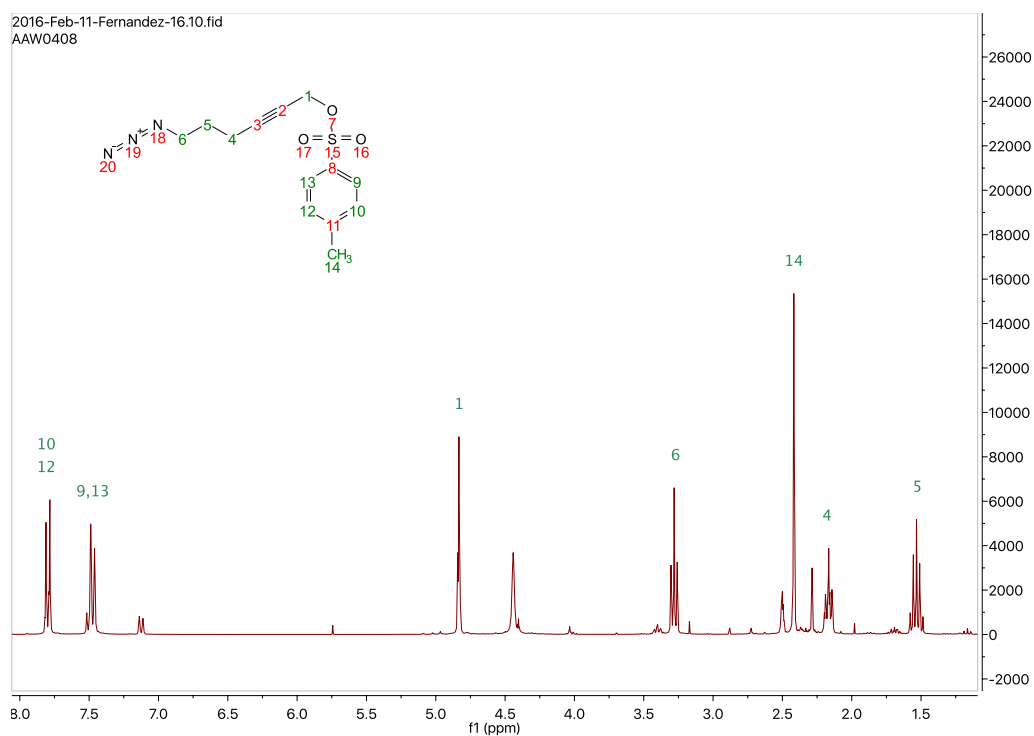


Sup. Fig. S6.8 ¹³C NMR spectrum of 6-azido-1-chloro hex-2-yne (4ai)



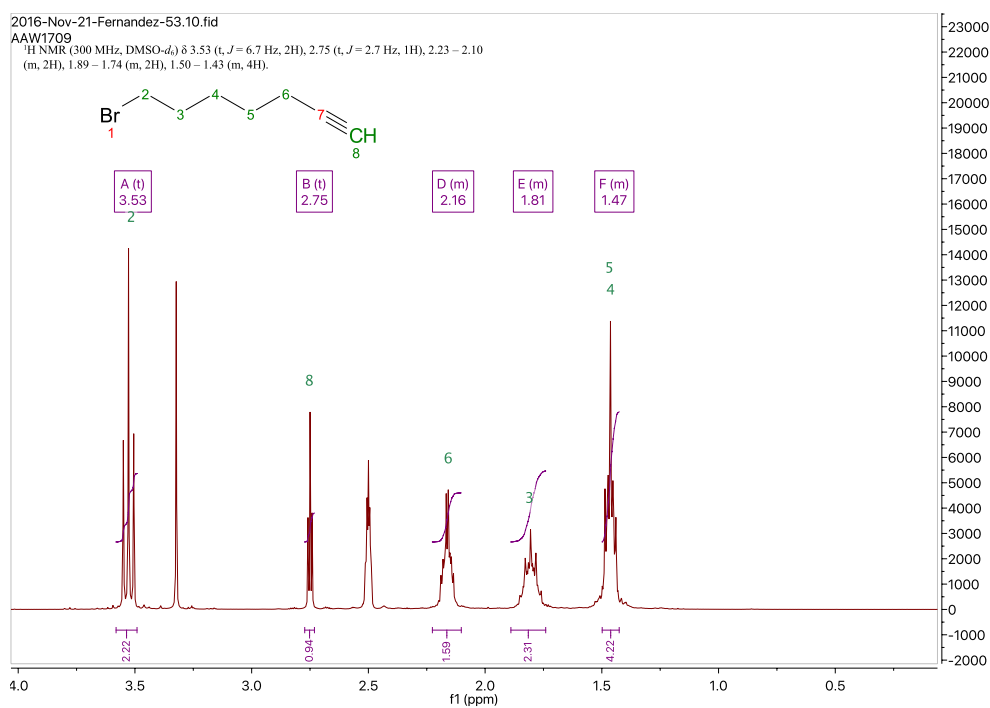
Sup. Fig. S6.9 Infrared spectrum of 6-azido-1-chloro hex-2-yne (4ai)

6-azidoex-2-yn-1-yl 6 methylbenzenesulfonate (**4a**_{ii})

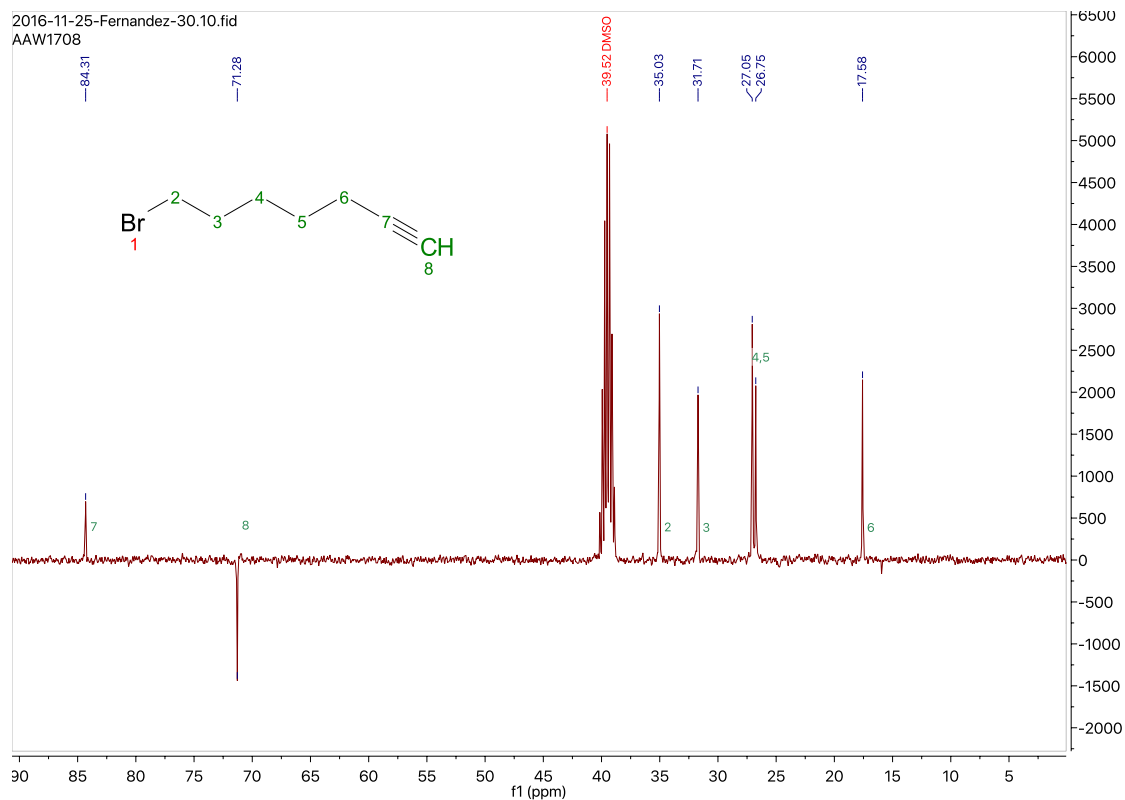


Sup. Fig. S6.10 ¹H NMR spectrum of 6-azidoex-2-yn-1-yl 6 methylbenzenesulfonate (**4a**_{ii})

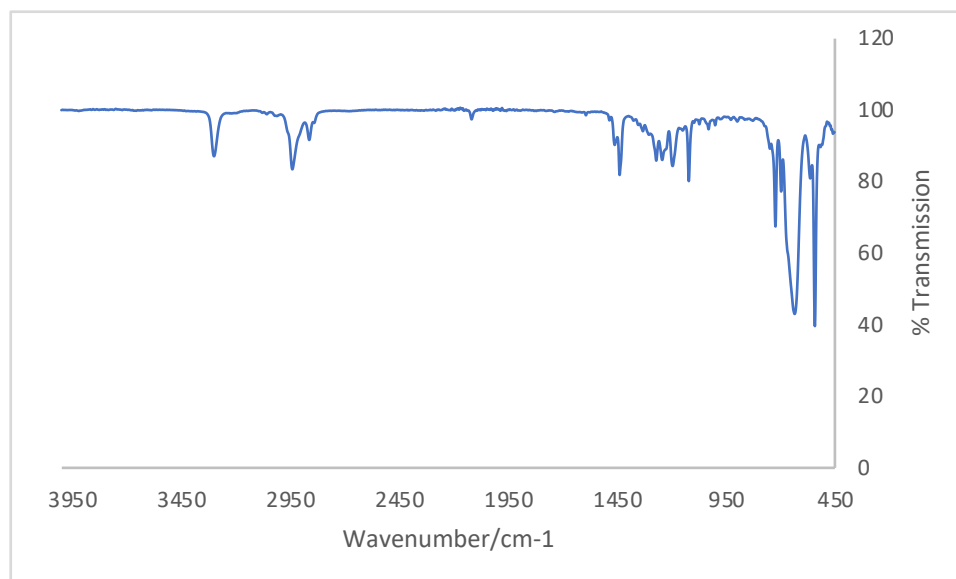
7-bromo-hept-1-yne (**1b**)



Sup. Fig. S6.11 ¹H-NMR spectrum of 7-bromo-hept-1-yne (**1b**)

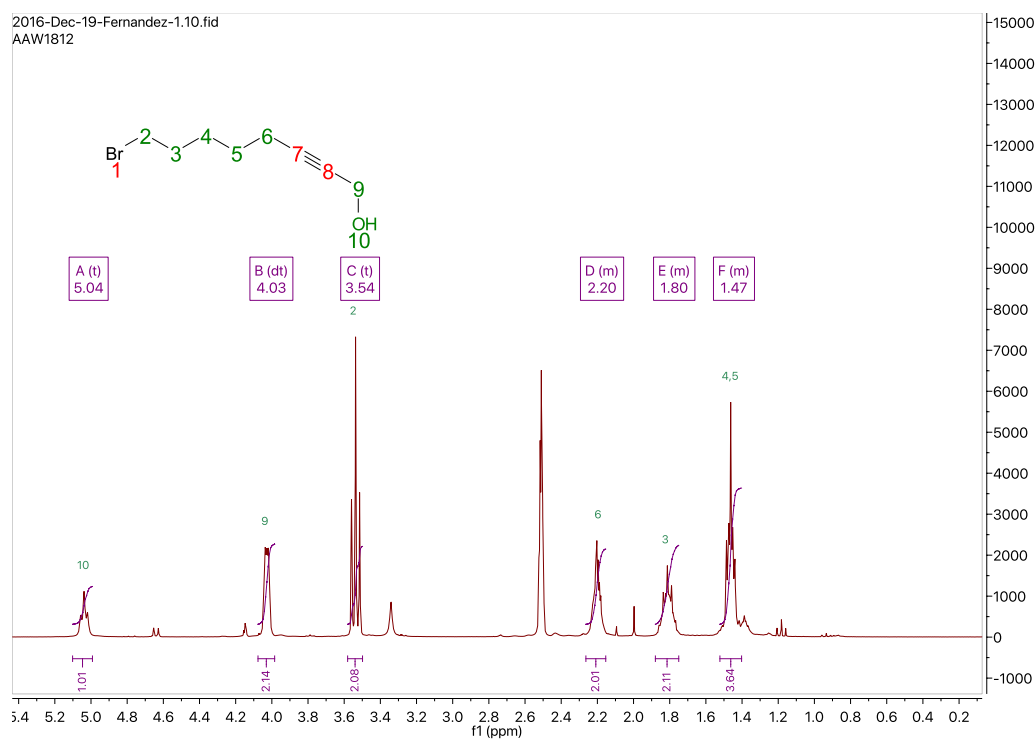


Sup. Fig. S6.12 ^{13}C -NMR spectrum of 7-bromo-hept-1-yne (**1b**)

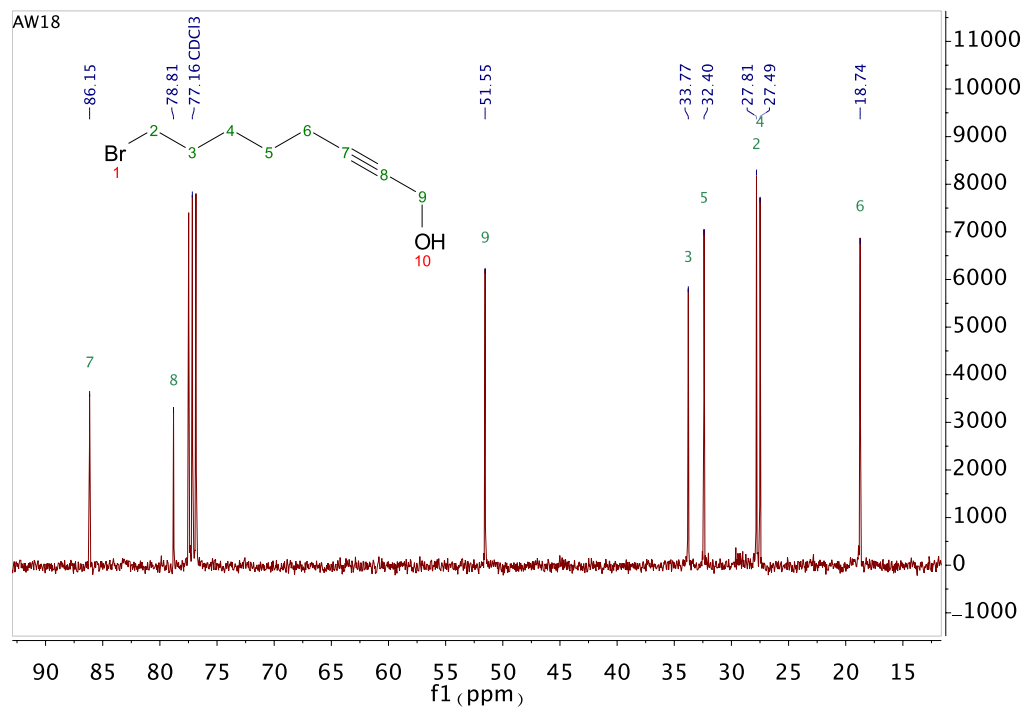


Sup. Fig. S6.13 Infrared spectrum of 7-bromo-hept-1-yne (**1b**)

8-bromooct-2-yn-1-ol (**2b**)

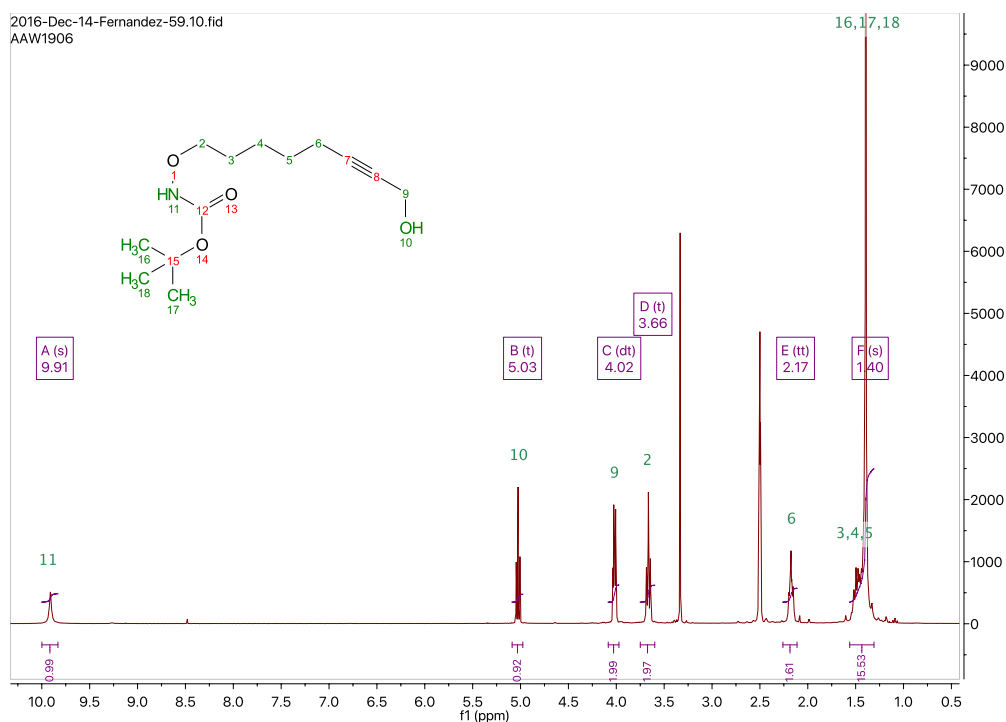


Sup. Fig. S6.14 ^1H -NMR spectrum of 8-bromooct-2-yn-1-ol (**2b**)

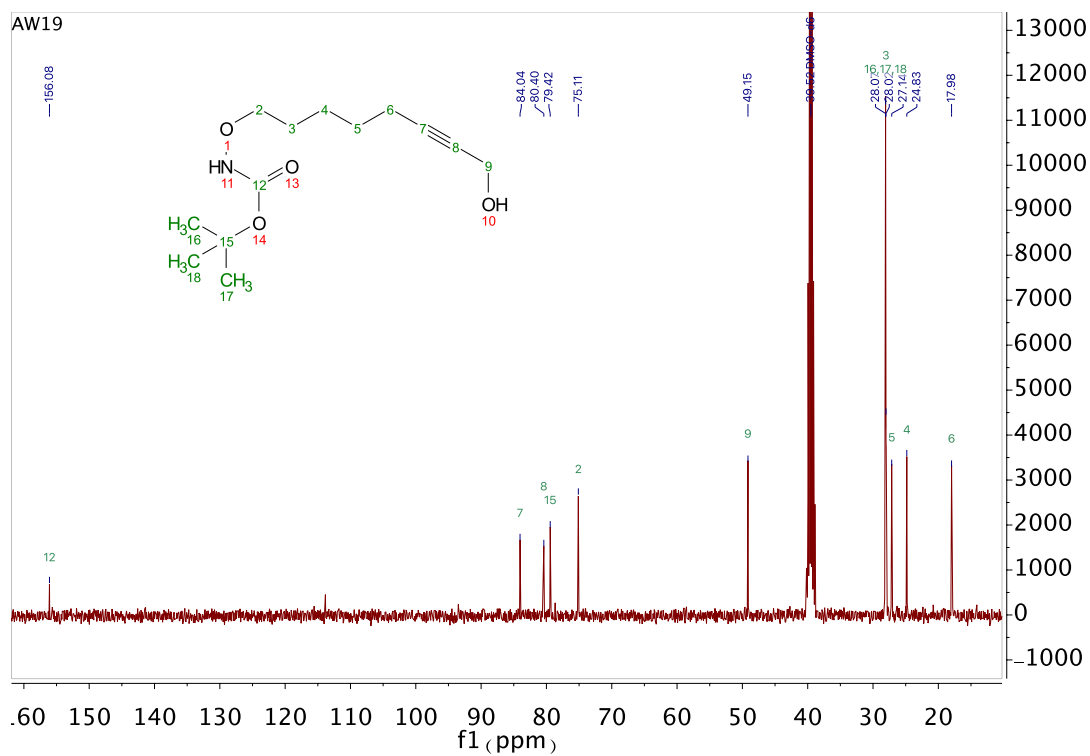


Sup. Fig. S6.15 ^{13}C NMR spectrum of 8-bromooct-2-yn-1-ol (**2b**)

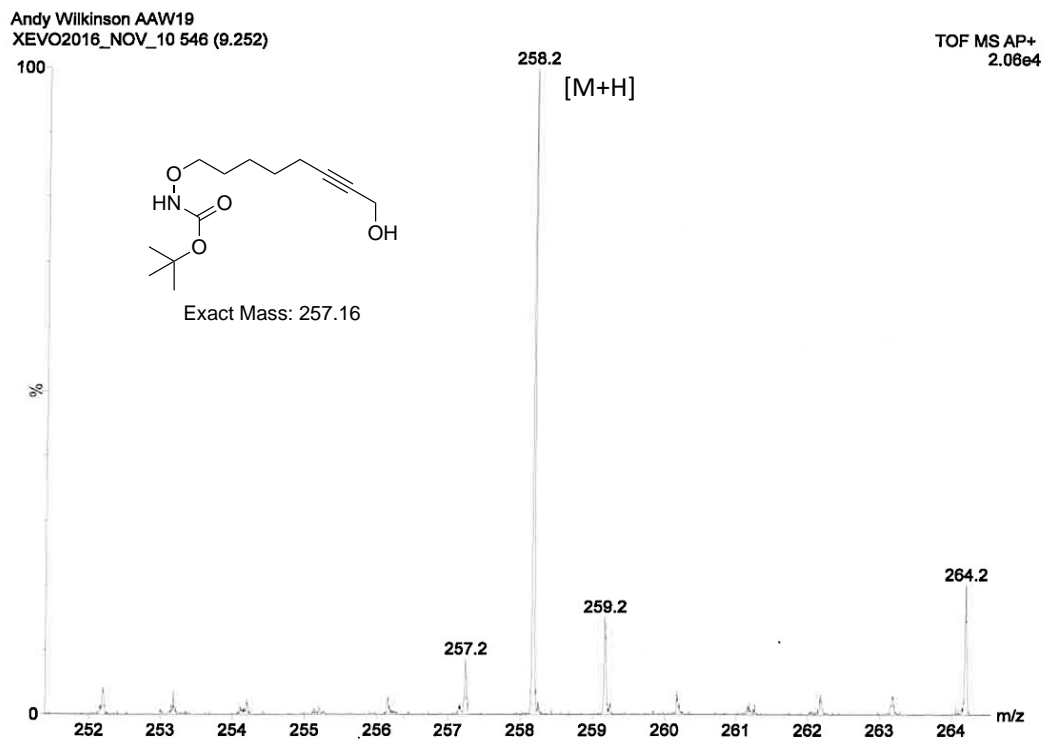
Tert-butyl ((8-hydroxyoct-6-yn-1-yl)oxy)carbamate (**3b**)



Sup. Fig. S6.16 ^1H -NMR spectrum of tert-butyl ((8-hydroxyoct-6-yn-1-yl)oxy)carbamate (**3b**)

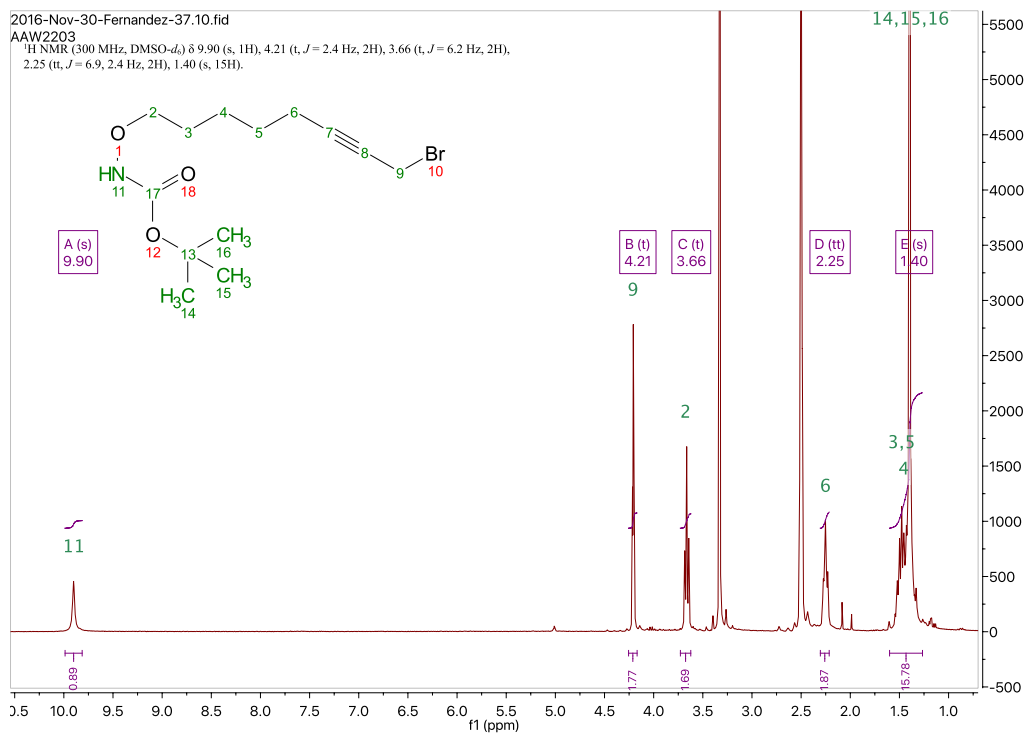


Sup. Fig. S6.17 ^{13}C NMR spectrum of tert-butyl ((8-hydroxyoct-6-yn-1-yl)oxy)carbamate (**3b**)

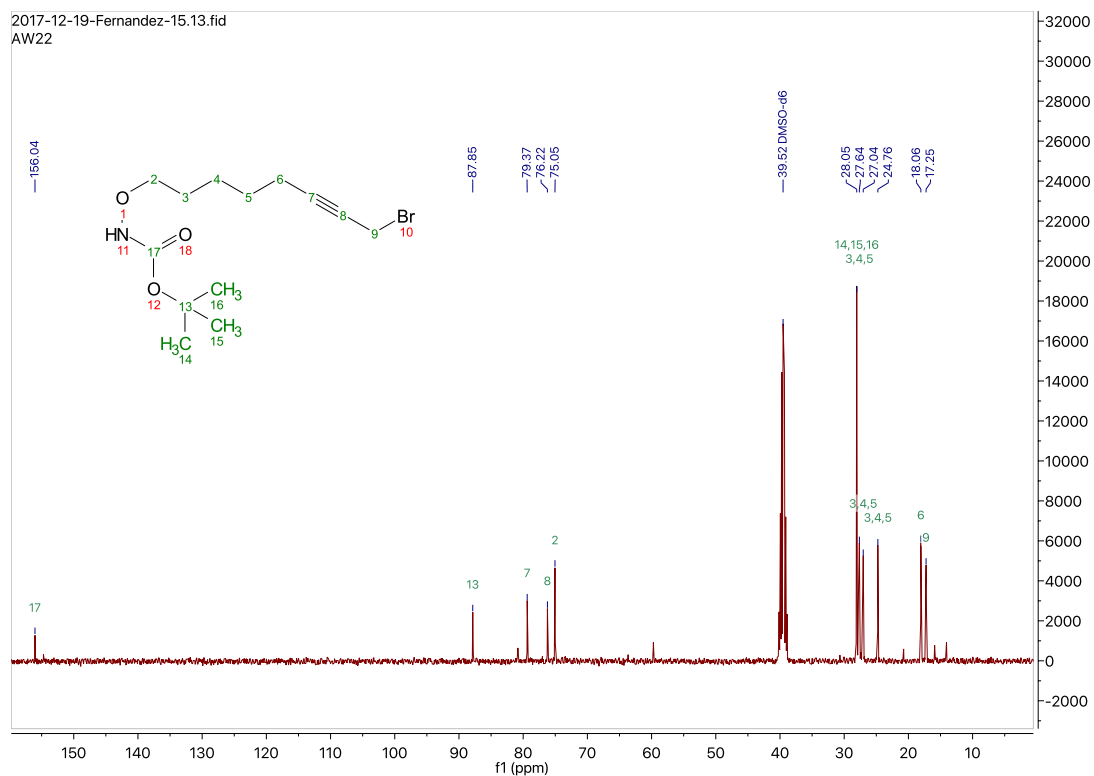


Sup. Fig. S6.18 Mass spectrum of *tert*-butyl ((8-hydroxyoct-6-yn-1-yl)oxy)carbamate (**3b**)

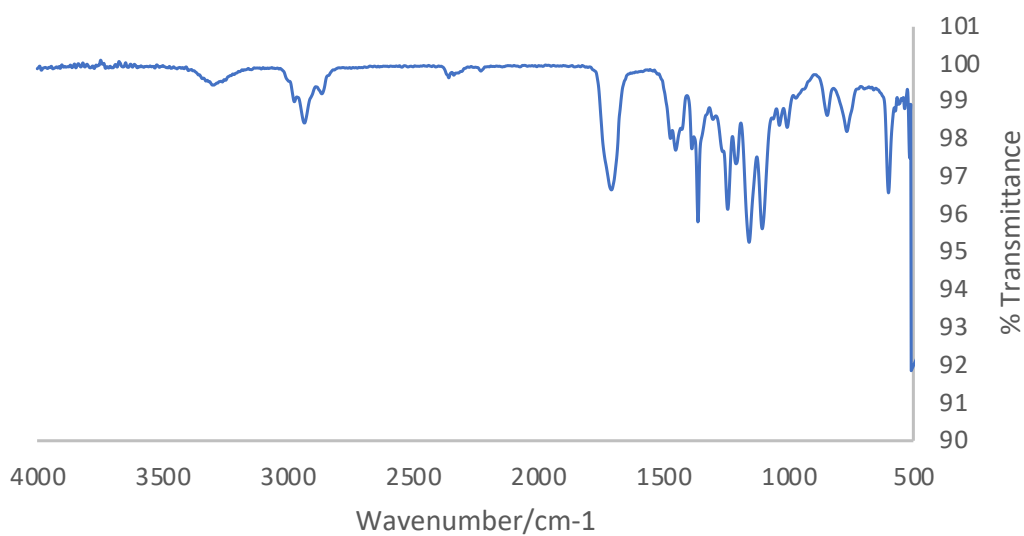
Tert-butyl ((8-bromooct-6-yn-1-yl)oxy)carbamate (**1b**)



Sup. Fig. S6.19 ¹H-NMR spectrum of *tert*-butyl ((8-bromooct-6-yn-1-yl)oxy)carbamate (**4biii**)



Sup. Fig. S6.20 ¹³C-NMR spectrum of *tert*-butyl ((8-bromooct-6-yn-1-yl)oxy)carbamate (**4biii**)

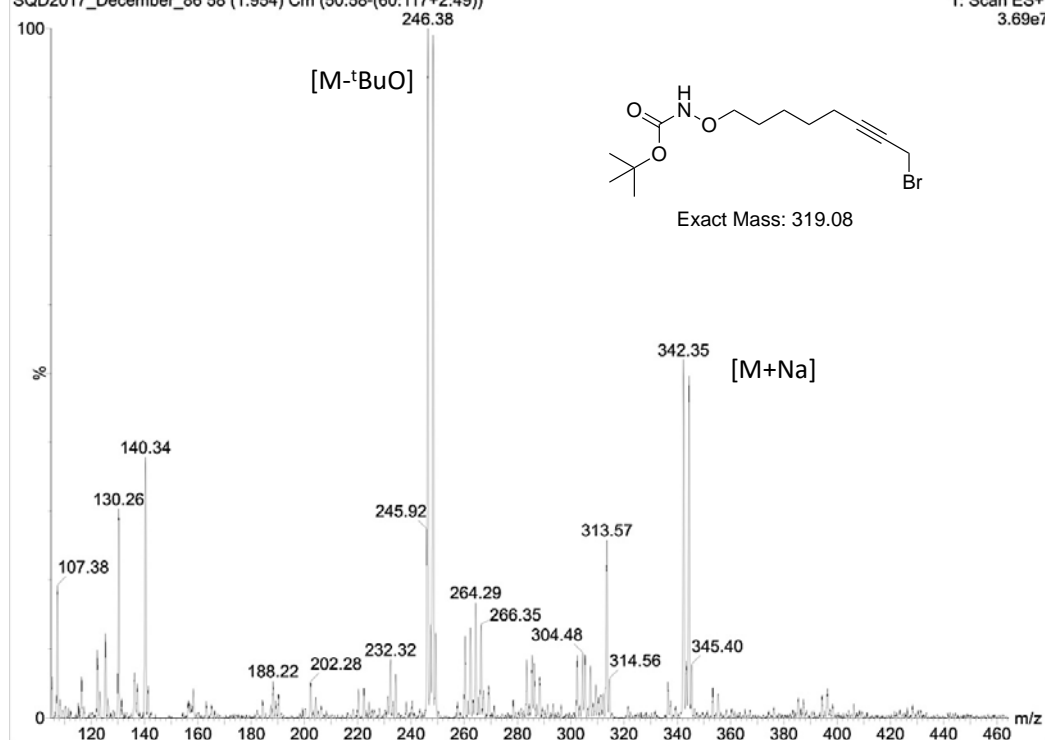


Sup. Fig. S6.21 Infrared spectrum of *tert*-butyl ((8-bromooct-6-yn-1-yl)oxy)carbamate (**4biii**)

AW22

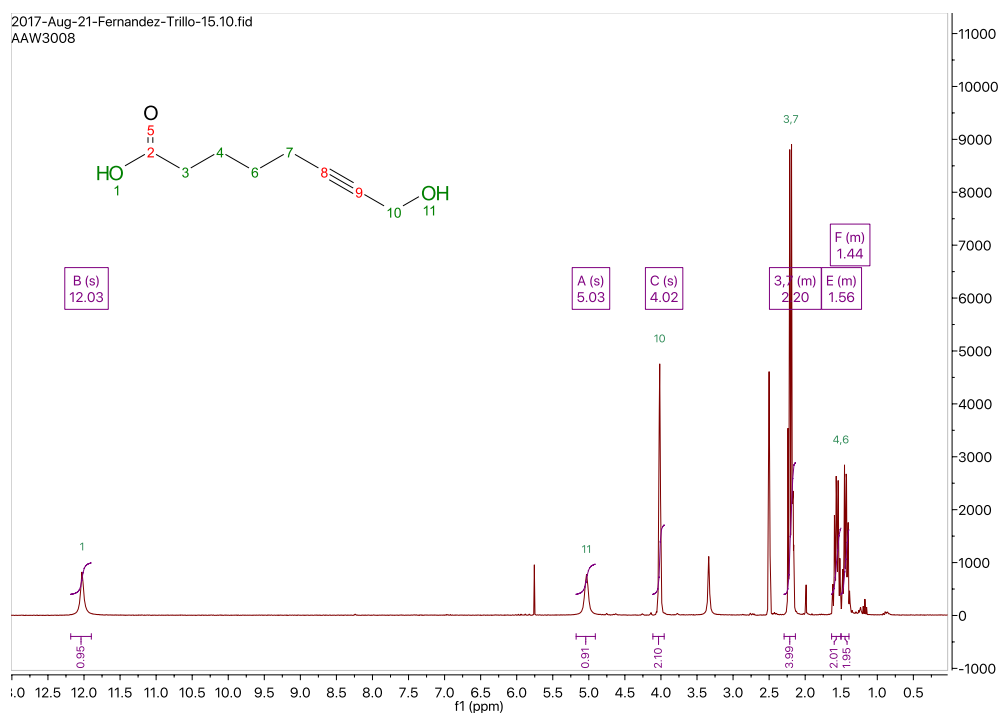
SQD2017_December_86 58 (1.954) Cm (50:58-(60:117+2:49))

1: Scan ES+
3.69e7

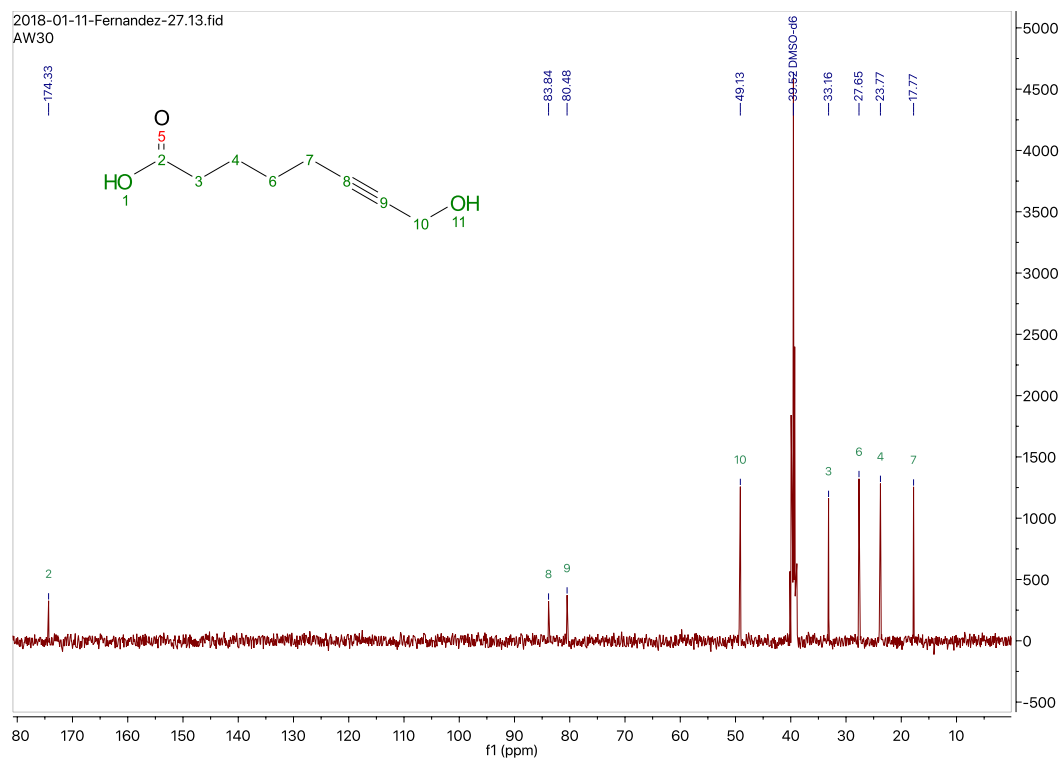


Sup. Fig. S6.22 Mass spectrum of tert-butyl ((8-bromooct-6-yn-1-yl)oxy)carbamate (4biii)

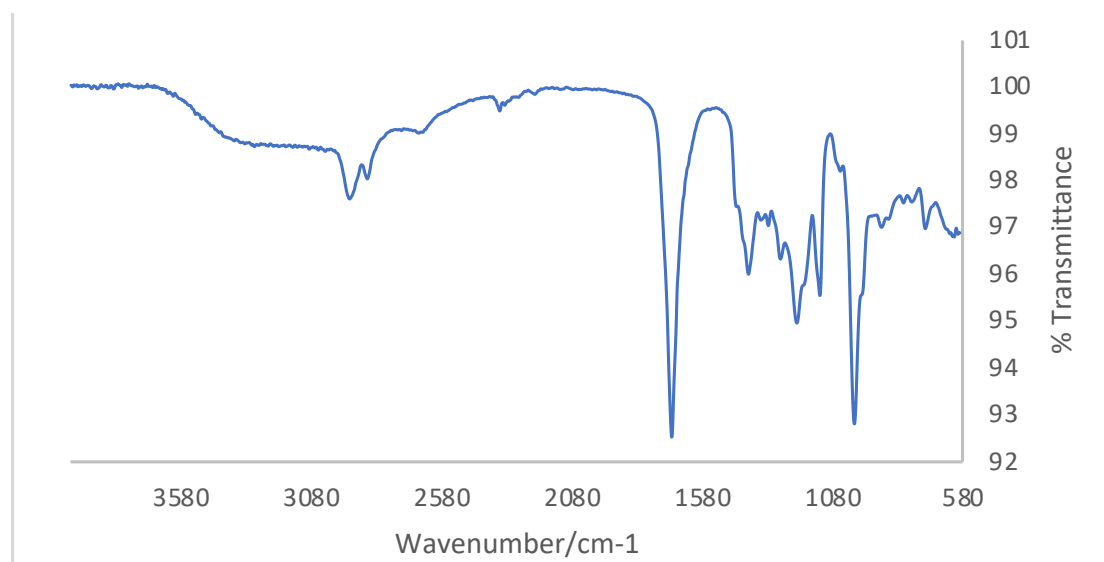
8-hydroxyoct-6-ynoic acid (**2c**)



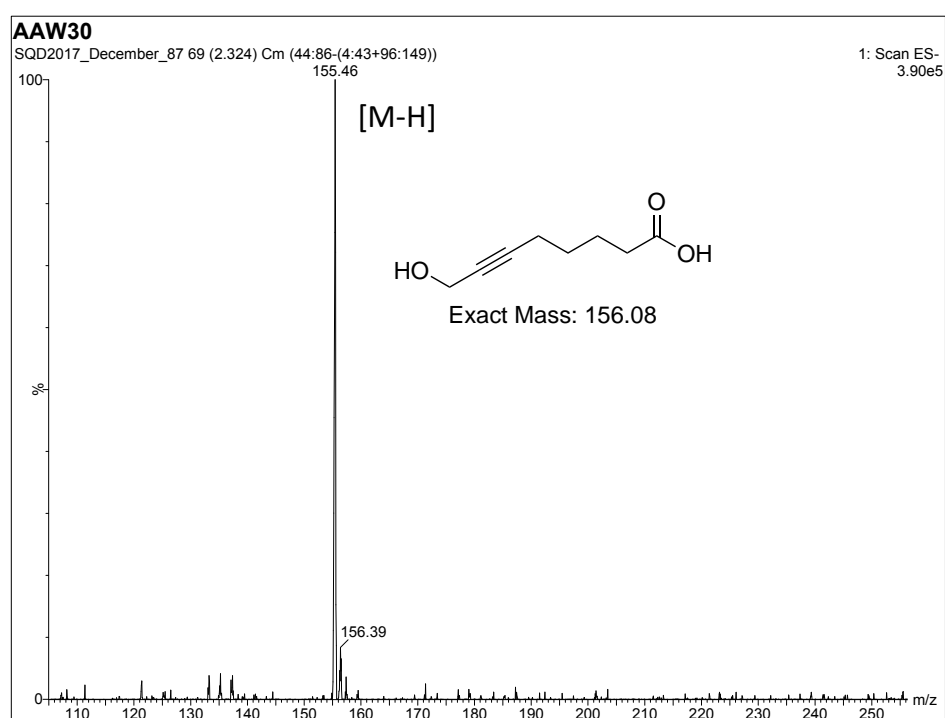
Sup. Fig. S6.23 ^1H -NMR spectrum of 8-hydroxyoct-6-ynoic acid (**2c**)



Sup. Fig. S6.24 ^{13}C -NMR spectrum of 8-hydroxyoct-6-ynoic acid (**2c**)

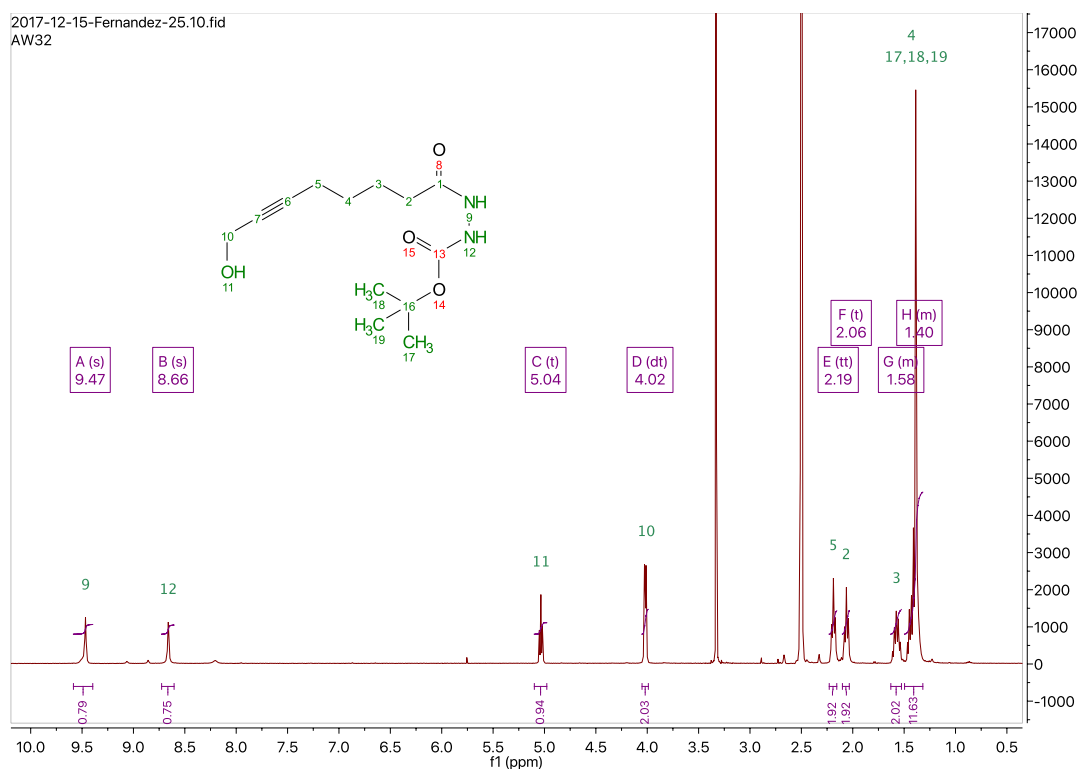


Sup. Fig. S6.25 Infrared spectrum of 8-hydroxyoct-6-ynoic acid (**2c**)

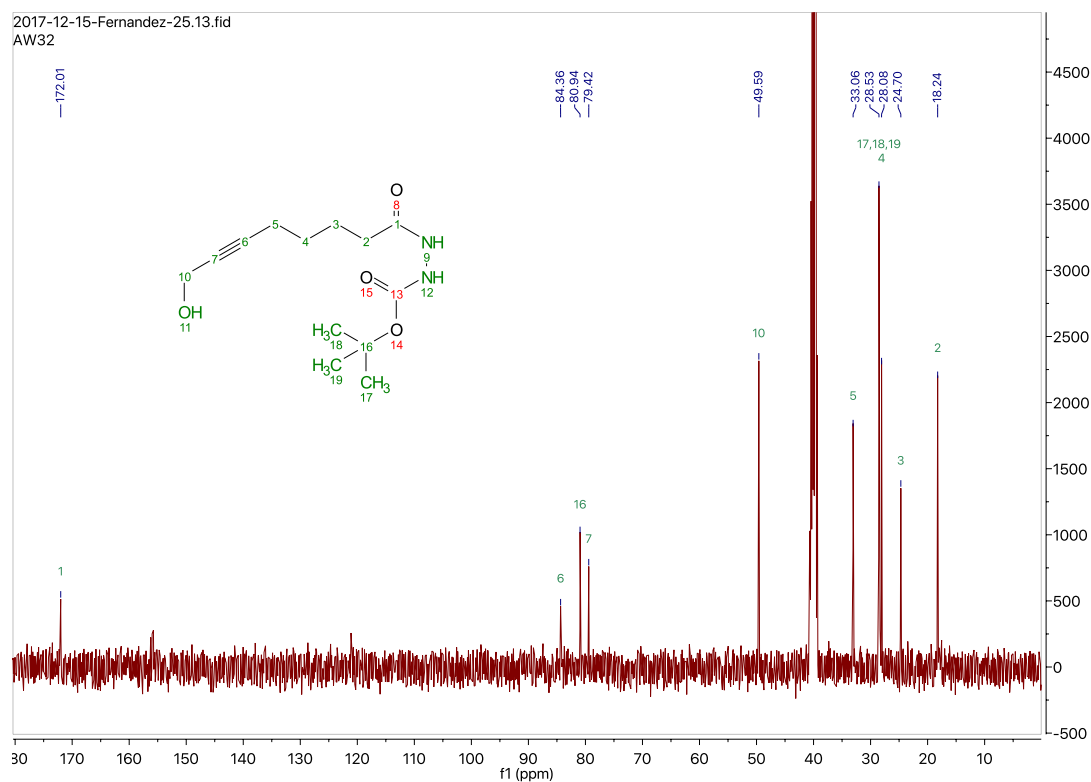


Sup. Fig. S6.26 Mass spectrum of 8-hydroxyoct-6-ynoic acid (**2c**) (negative ion)

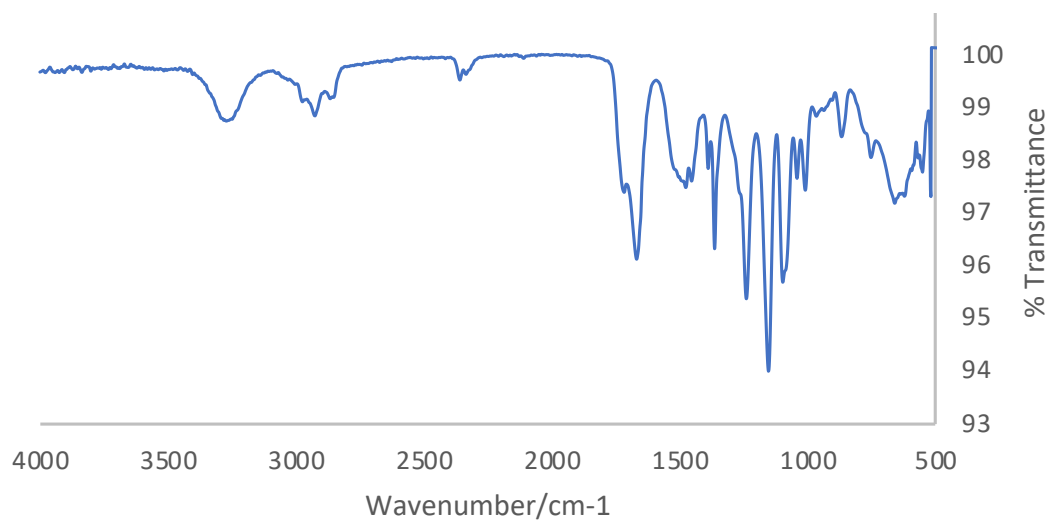
tert-butyl 2-(8-hydroxyoct-6-ynoyl)hydrazine-1-carboxylate (**3c**)



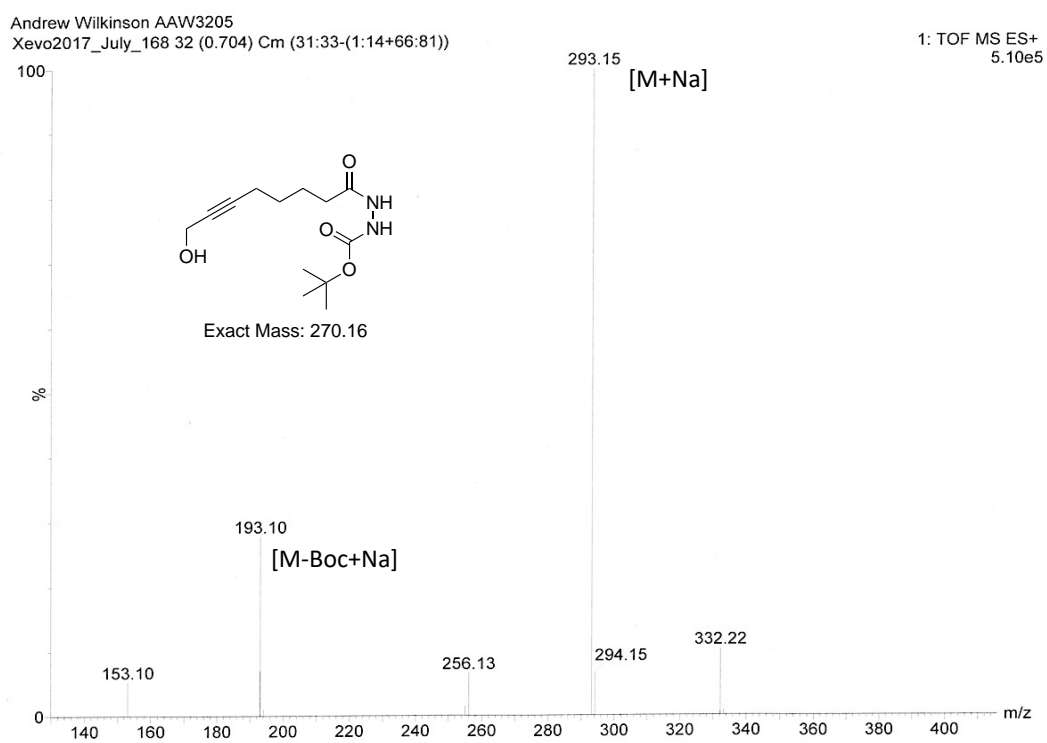
Sup. Fig. S6.27 ^1H -NMR spectrum of tert-butyl 2-(8-hydroxyoct-6-ynoyl)hydrazine-1-carboxylate (**3c**)



Sup. Fig. S6.28 ^{13}C -NMR spectrum of tert-butyl 2-(8-hydroxyoct-6-ynoyl)hydrazine-1-carboxylate (**3c**)

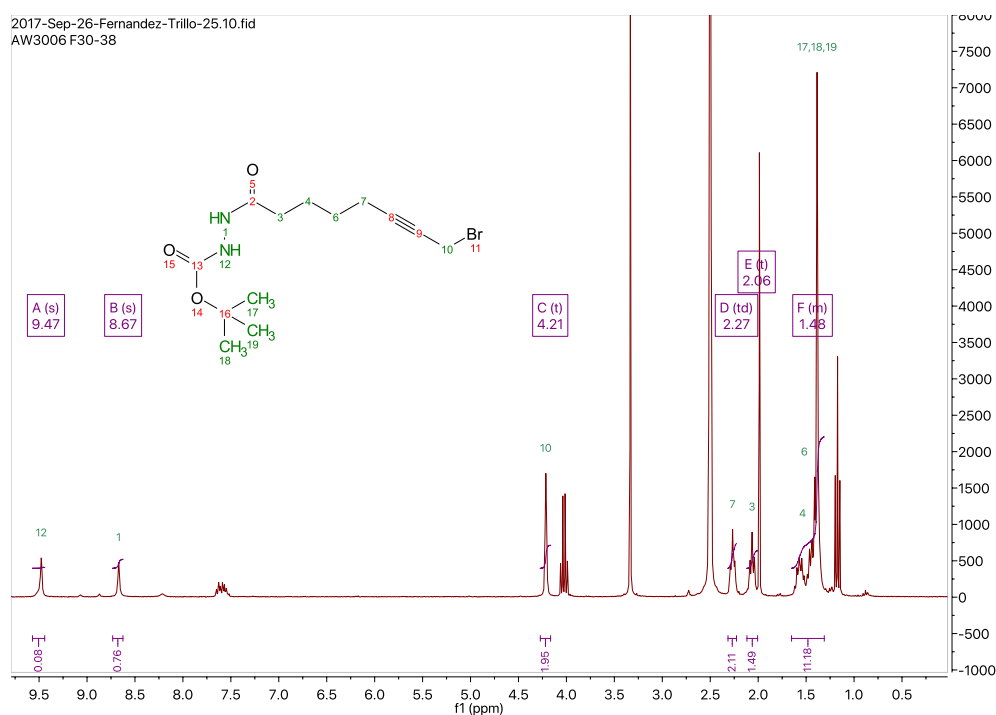


Sup. Fig. S6.29 Infrared spectrum of *tert*-butyl 2-(8-hydroxyoct-6-ynyl)hydrazine-1-carboxylate (**3c**)

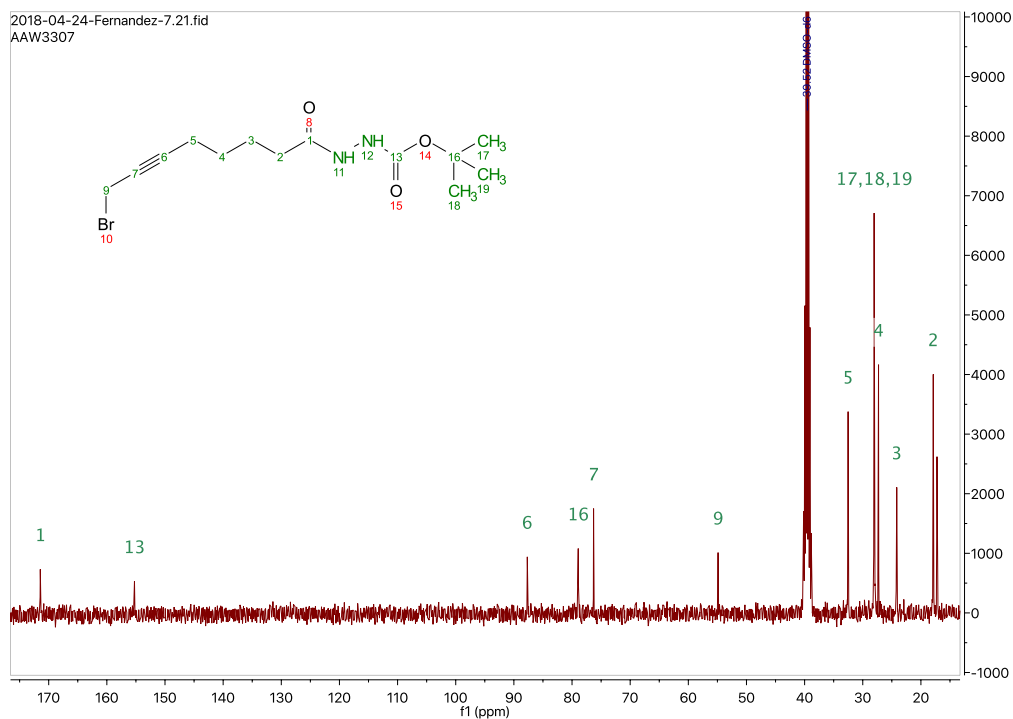


Sup. Fig. S6.30 Mass spectrum of *tert*-butyl 2-(8-hydroxyoct-6-ynyl)hydrazine-1-carboxylate (**3c**)

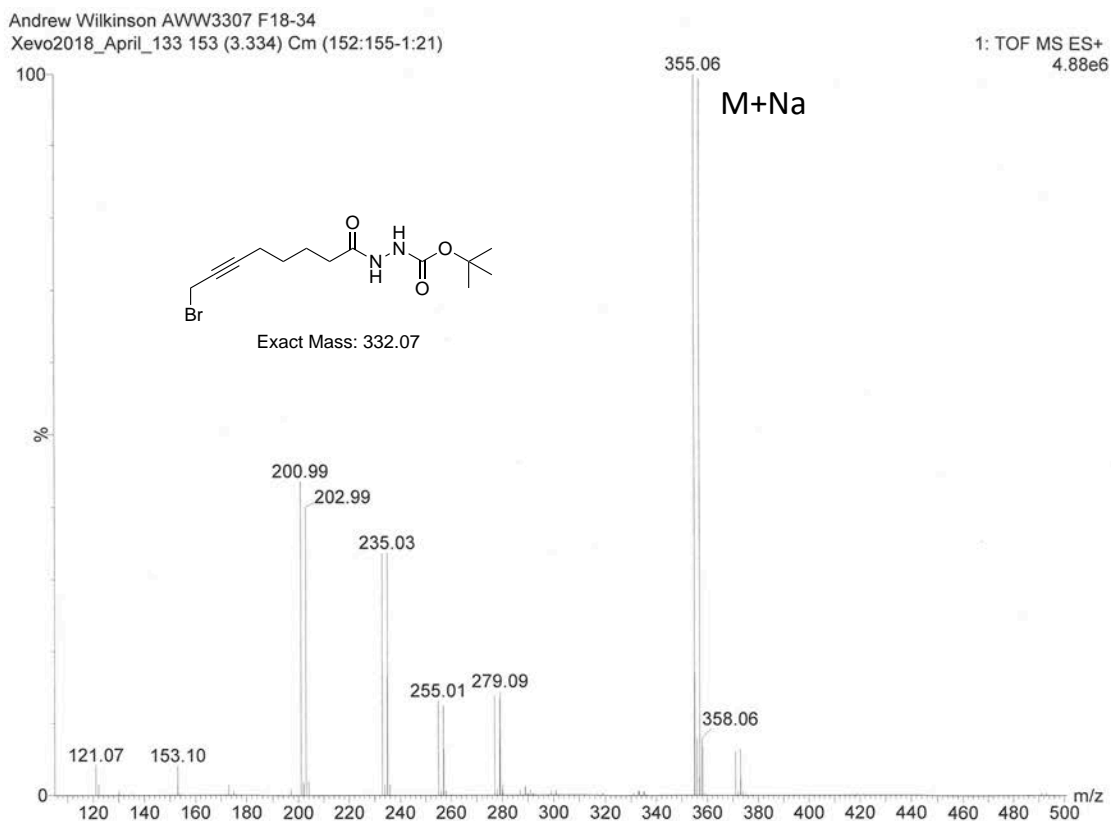
tert-butyl 2-(8-bromooct-6-ynoyl)hydrazine-1-carboxylate (**4ciii**)



Sup. Fig. S6.31 ^1H -NMR spectrum of tert-butyl 2-(8-bromooct-6-ynoyl)hydrazine-1-carboxylate (**4ciii**)

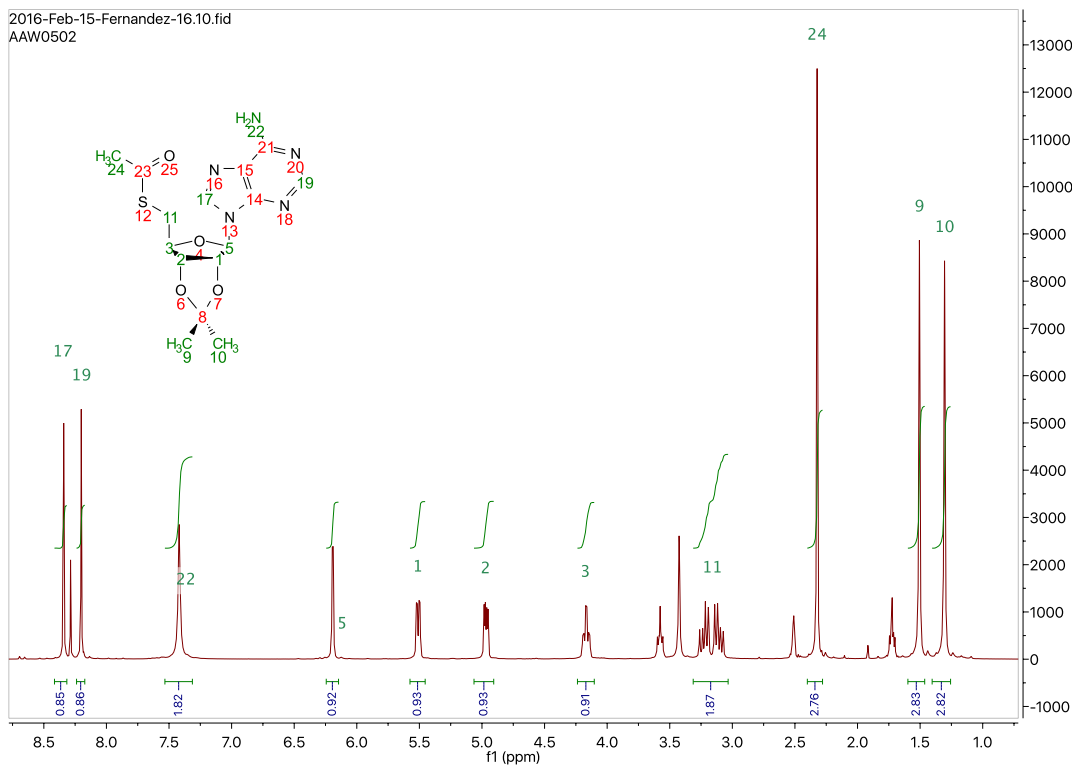


Sup. Fig. S6.32 ^{13}C NMR spectrum of tert-butyl 2-(8-bromooct-6-ynoyl)hydrazine-1-carboxylate (**4ciii**)

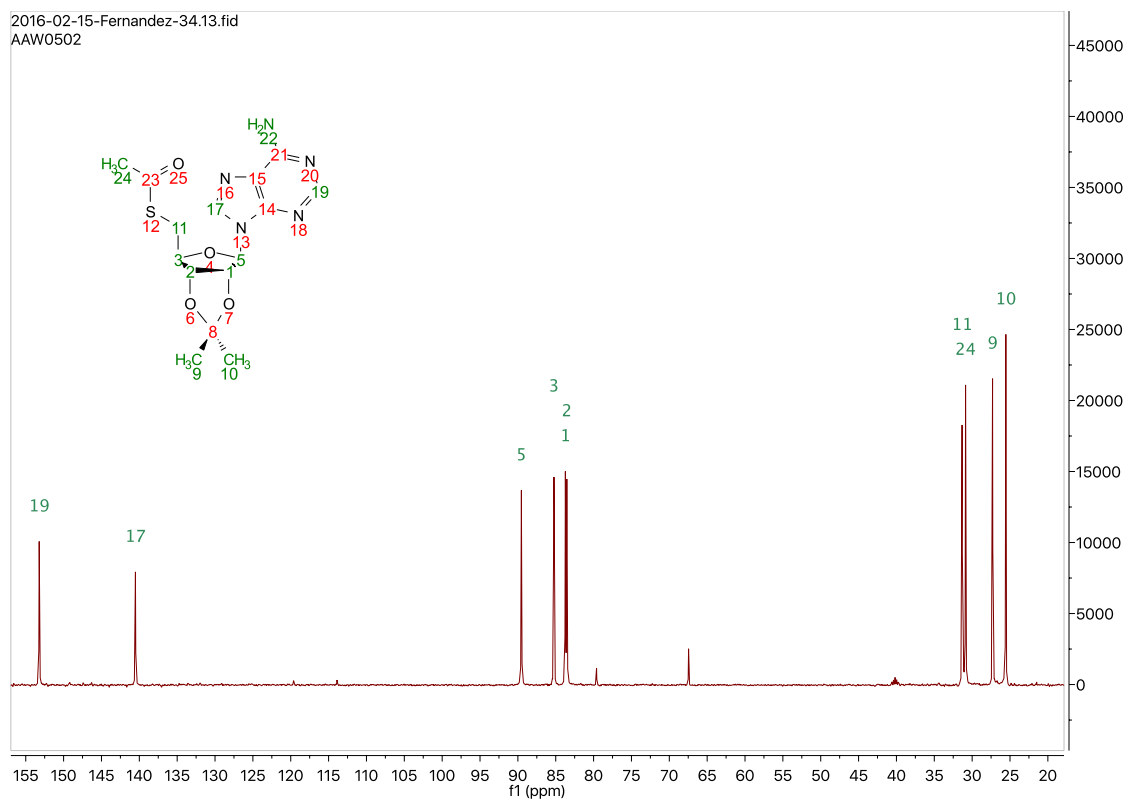


Sup. Fig. S6.33 Mass spectrum of *tert*-butyl 2-(8-bromooct-6-ynoyl)hydrazine-1-carboxylate (**4ciii**)

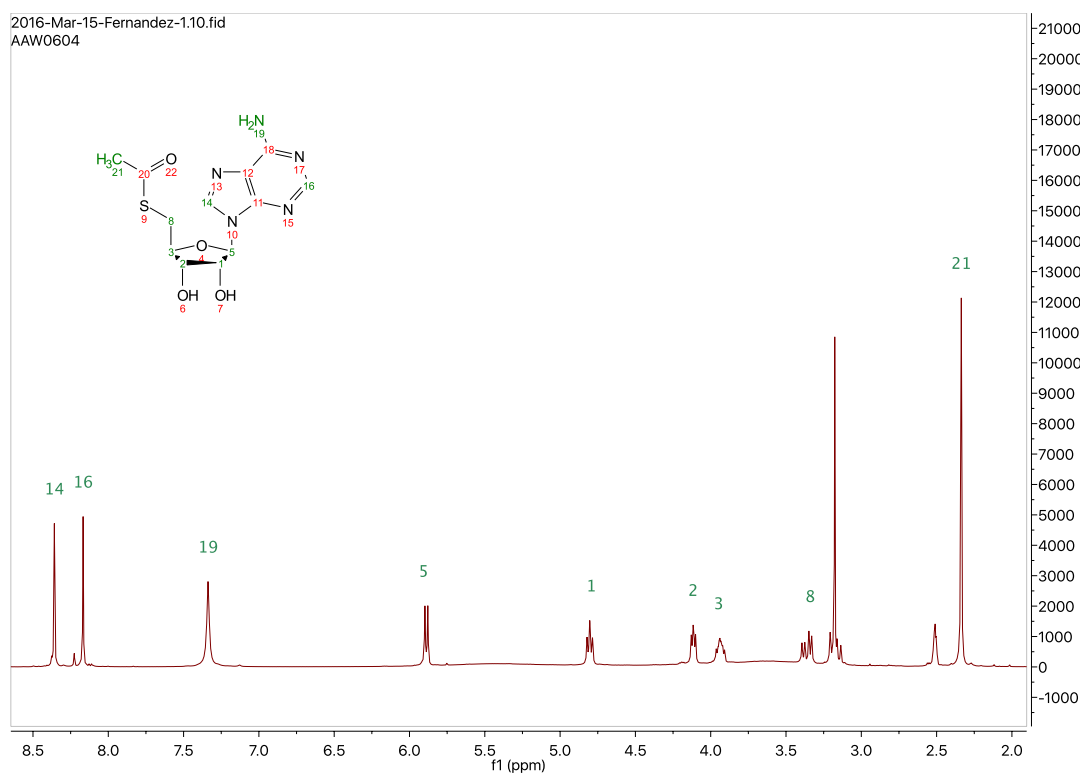
5'-Acetylthio-5'-deoxy-2',3'-O-isopropylideneadenosine (**6**)



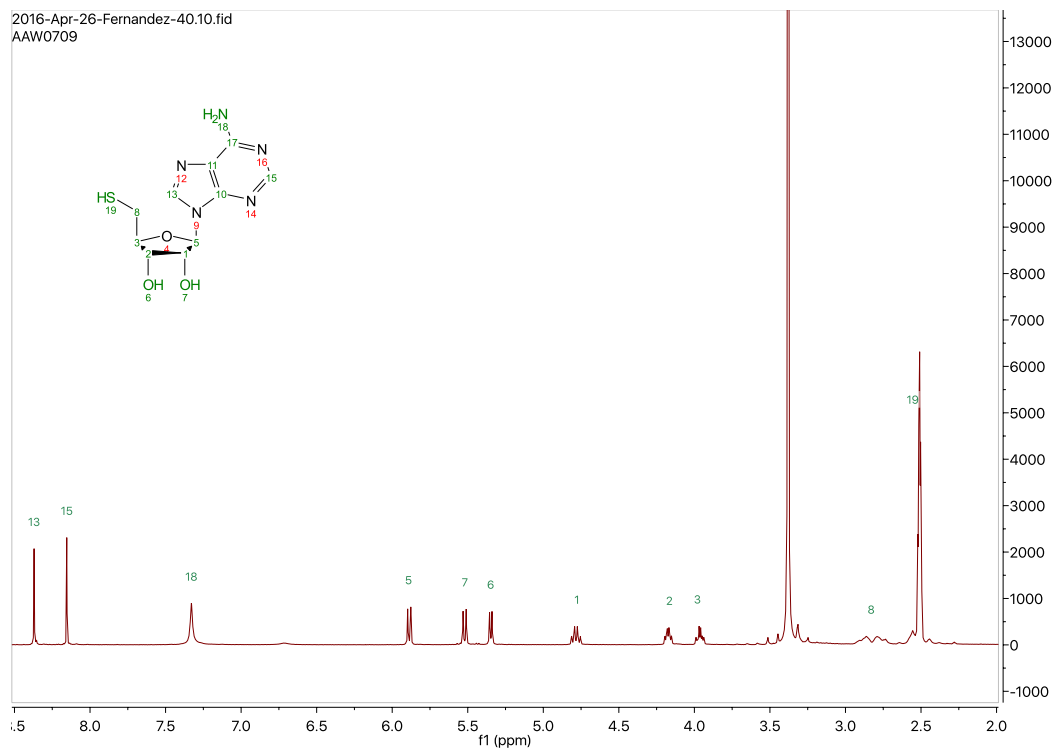
Sup. Fig. S6.34 ^1H NMR spectrum of 5'-Acetylthio-5'-deoxy-2',3'-O-isopropylideneadenosine (**6**)



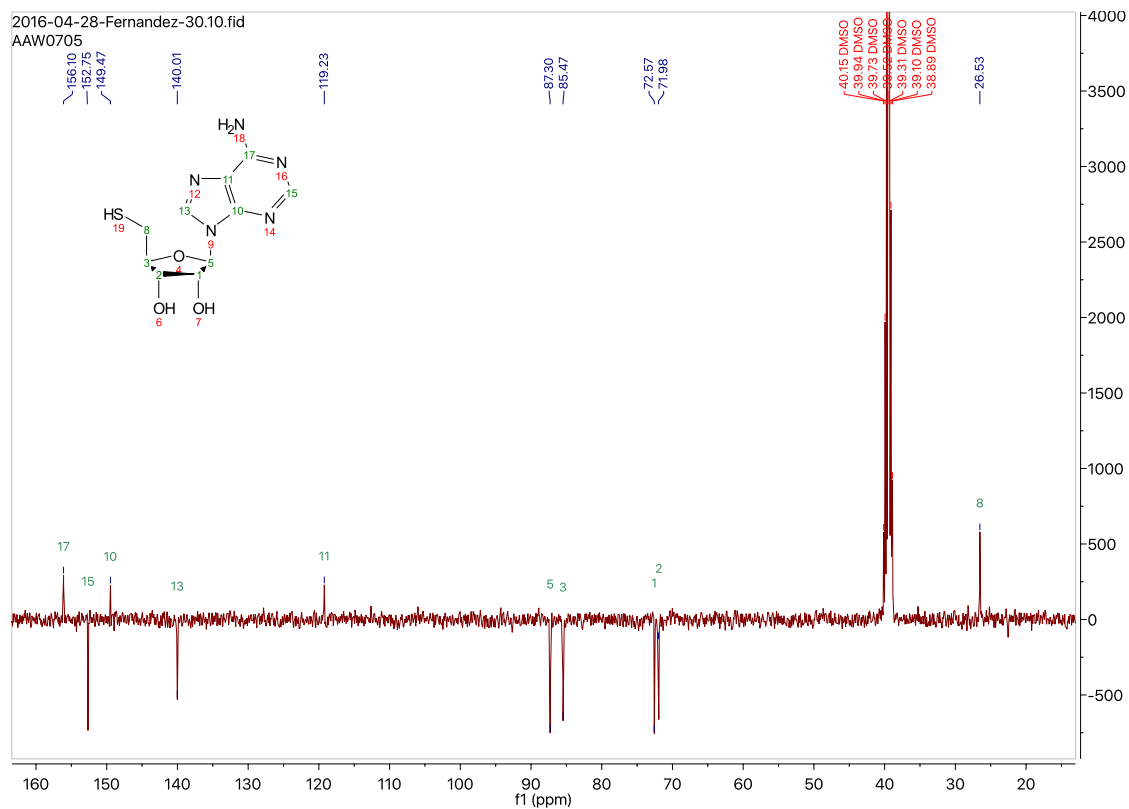
5'-Acetylthio-5'-deoxyadenosine (7)



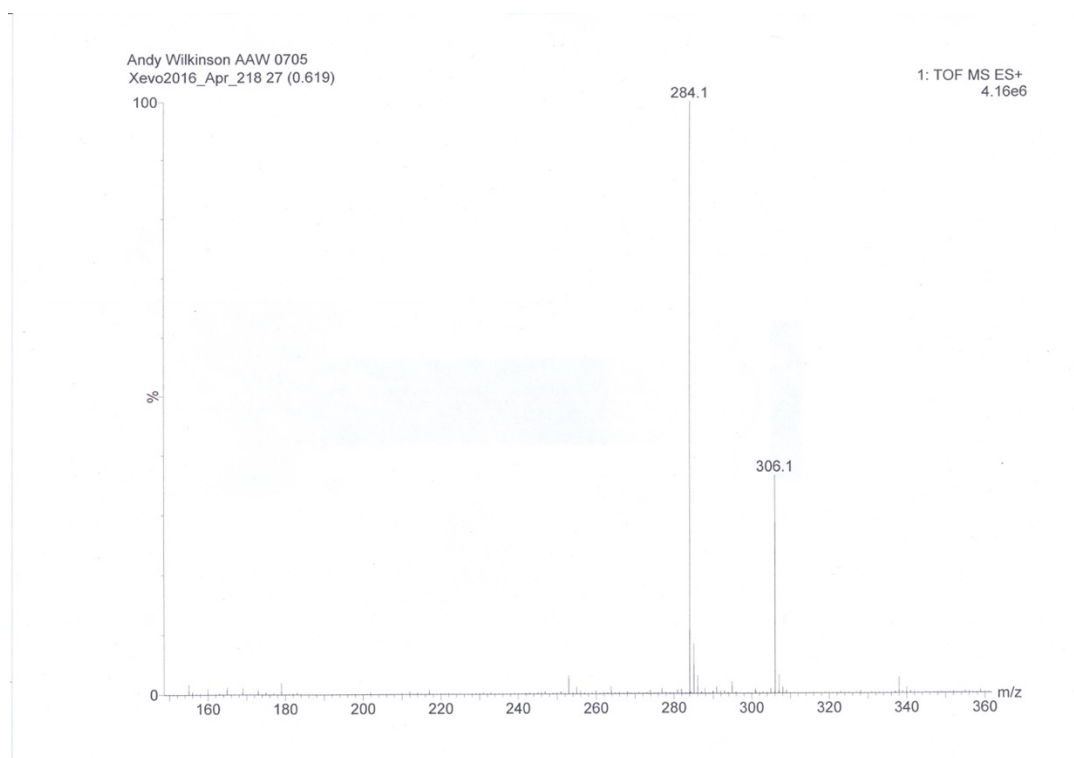
Thioadenosine (8)



Sup. Fig. S6.37 ^1H NMR spectrum of thioadenosine (8)

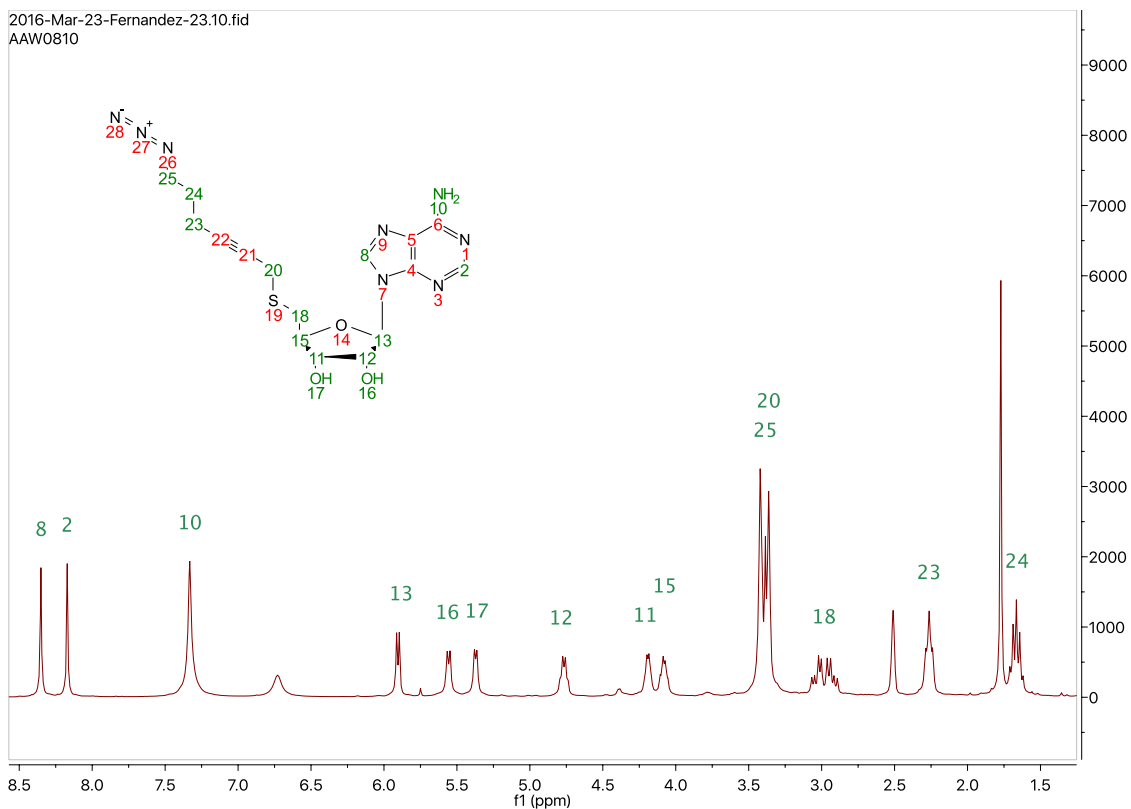


Sup. Fig. S6.38 Pendant ^{13}C NMR spectrum of thioadenosine (8)

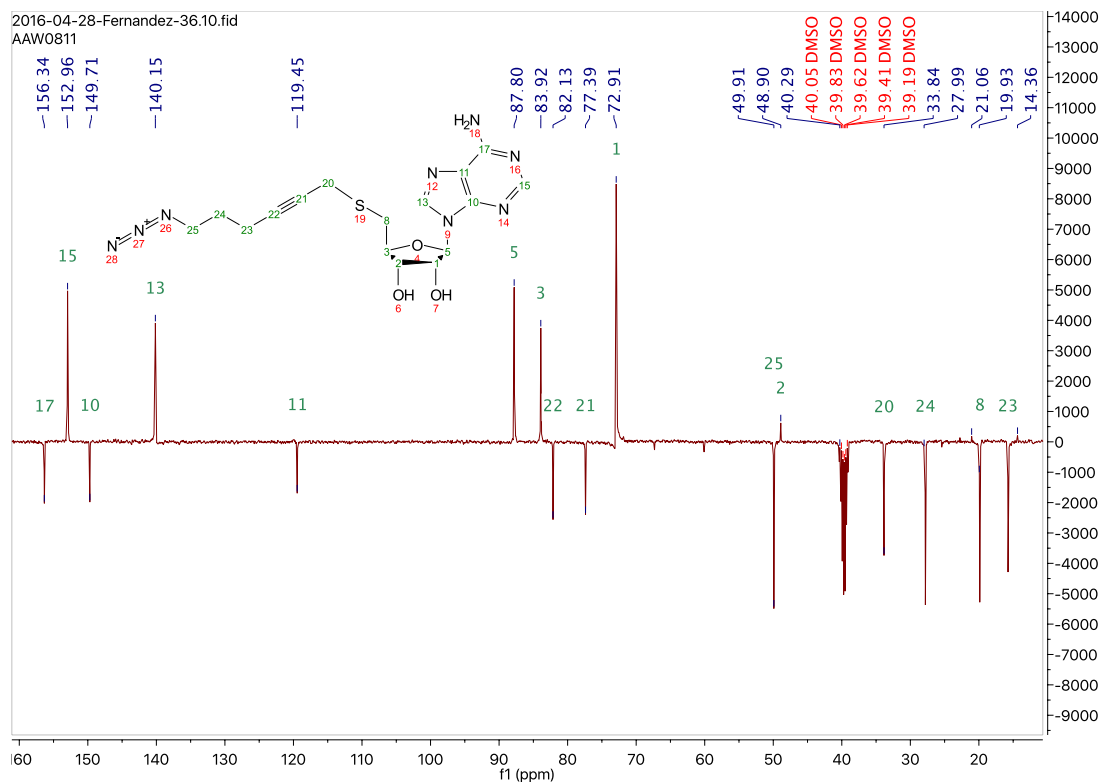


Sup. Fig. S6.39 Mass spectrum of thioadenosine (**8b**)

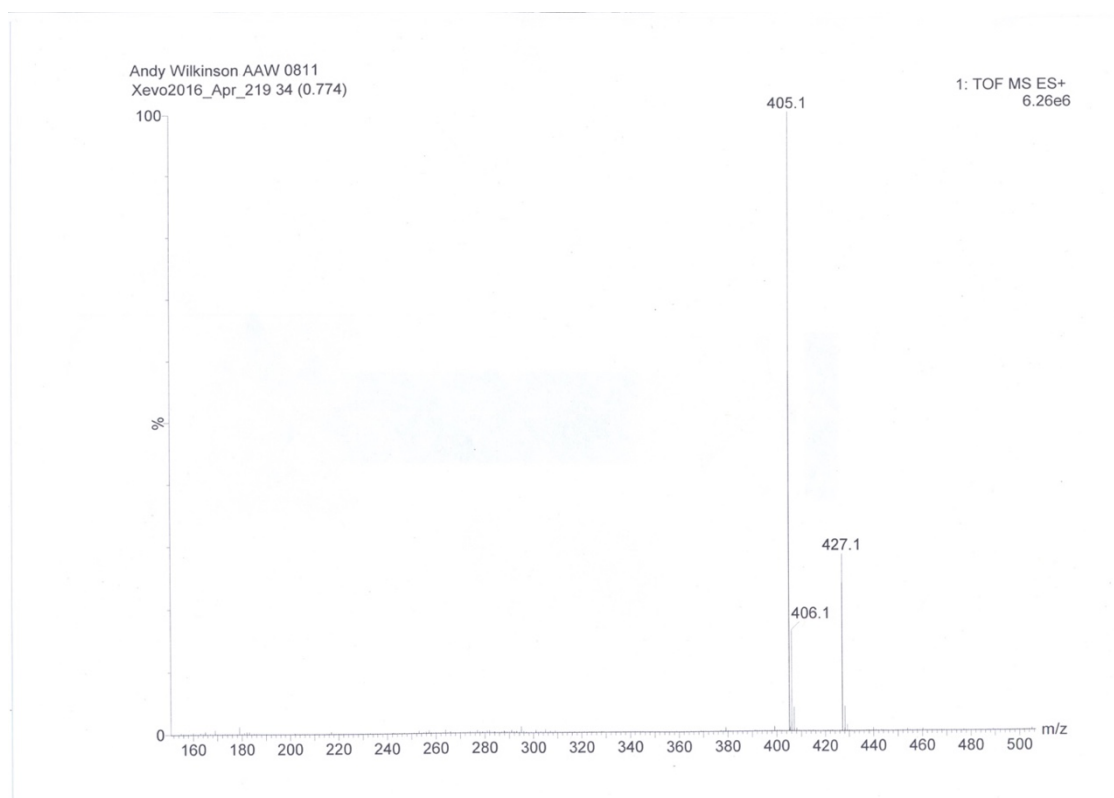
S-6-azido-hex-2-yne-5'-thioadenosine (**9a**)



Sup. Fig. S6.40 ¹H NMR spectrum of S-6-azido-hex-2-yne-5'-thioadenosine (**9a**)

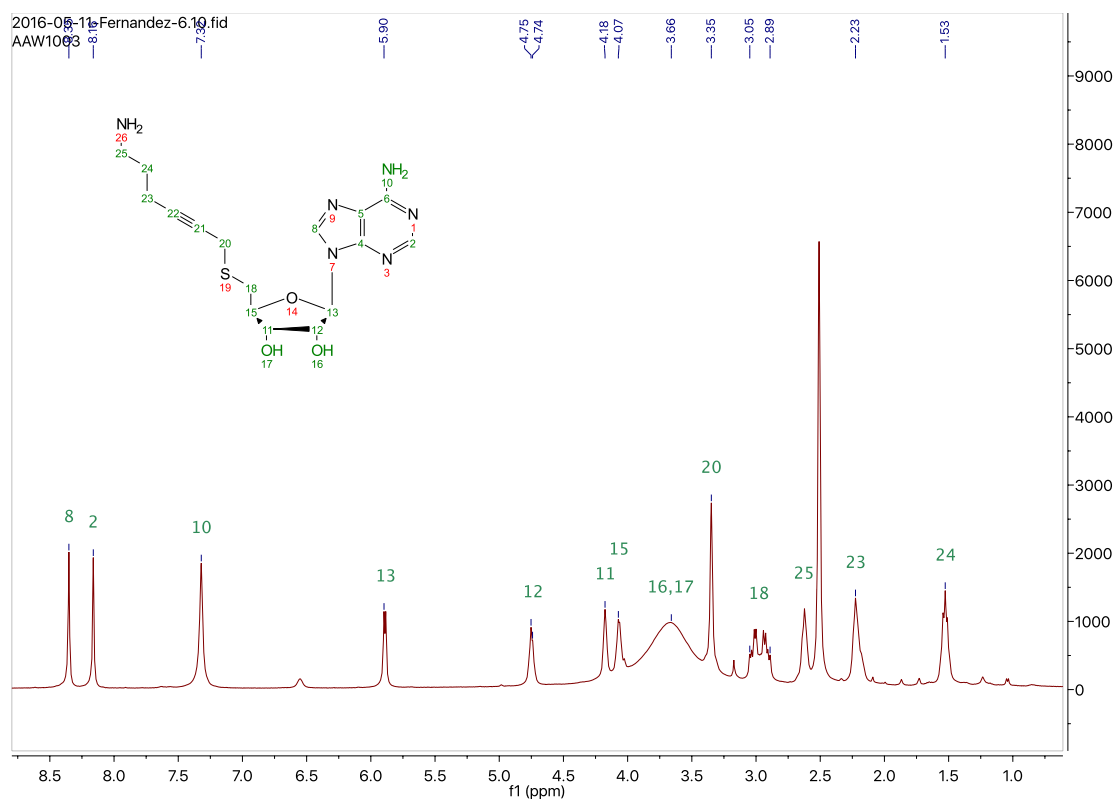


Sup. Fig. S6.41 Pendant ¹³C NMR spectrum of *S*-6-azido-hex-2-yne-5'-thioadenosine (**9a**)

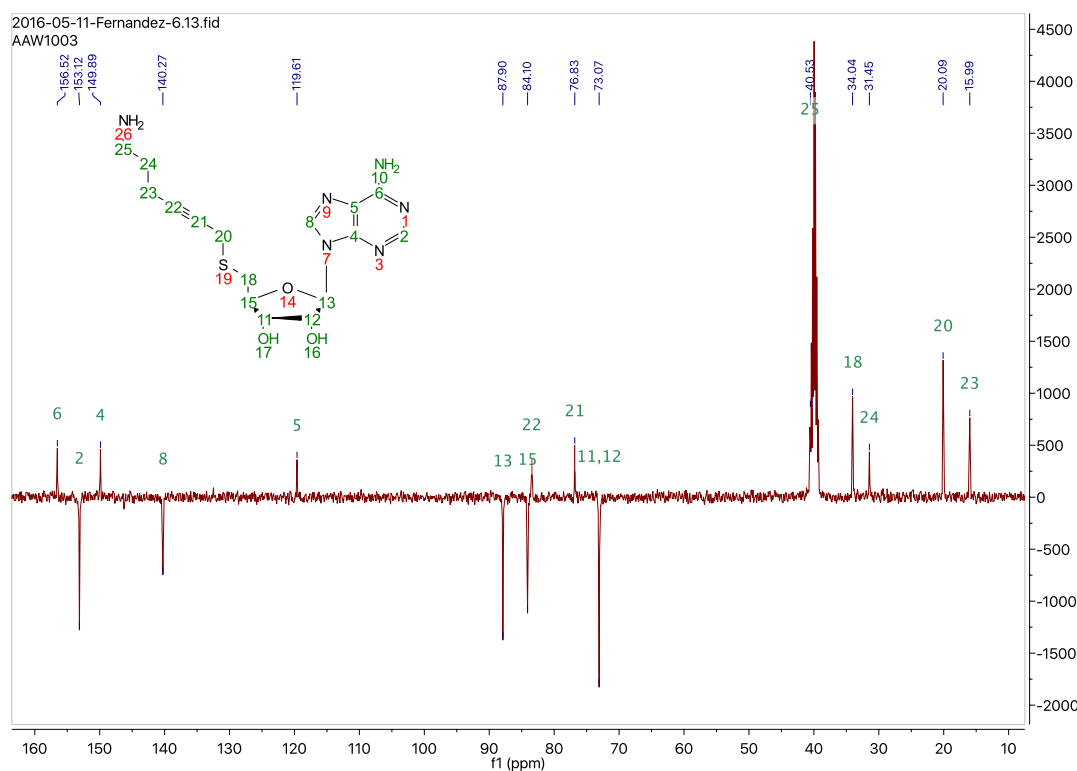


Sup. Fig. S6.42 Mass spectrum of *S*-6-azido-hex-2-yne-5'-thioadenosine (**9a**)

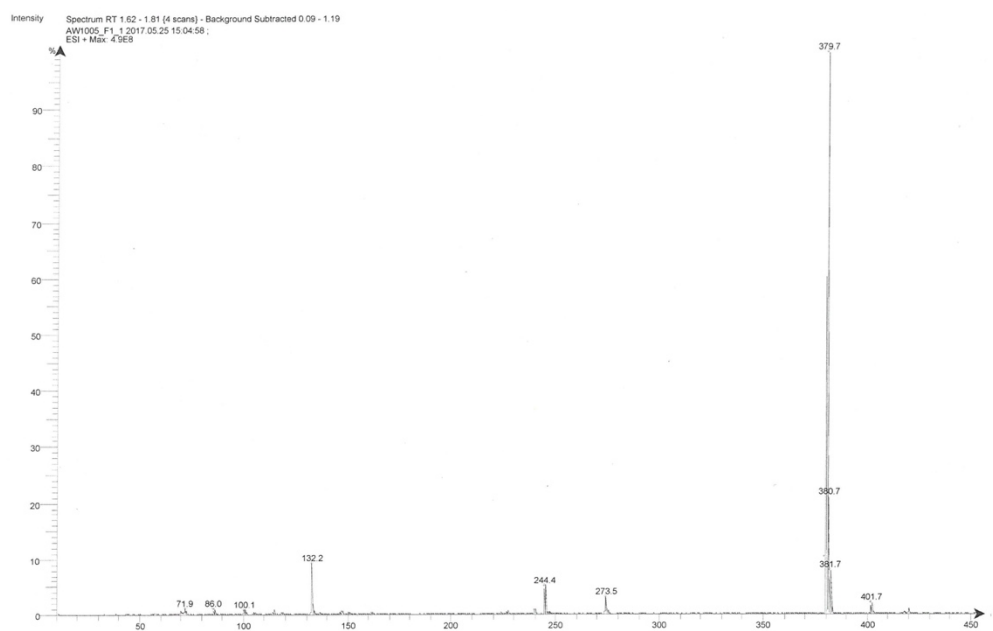
S-6-amino-hex-2-yne-5'-thioadenosine (**9b**)



Sup. Fig. S6.43 ¹H NMR spectrum of S-6-amino-hex-2-yne-5'-thioadenosine (**9b**)



Sup. Fig. S6.44 Pendant ¹³C NMR spectrum of S-6-amino-hex-2-yne-5'-thioadenosine (**9b**)



Sup. Fig. S6.45 Mass spectrum of S-6-amino-hex-2-yne-5'-thioadenosine (**9b**)

Site-selective writing, modifying, erasing and rewriting of nucleic acids with dynamic covalent S-adenosyl-L-methionine analogues.

Andrew A. Wilkinson, Elodie Jagu, Krystian Ubych, Ashleigh E. Rushton, Jack Kennefick, Qiang Su, Robert K. Neely,* Francisco Fernandez-Trillo.*

School of Chemistry, University of Birmingham, Edgbaston, Birmingham, UK, B15 2TT

r.k.neely@bham.ac.uk, f.fernandez-trillo@bham.ac.uk

ABSTRACT: Until recently, single-cell studies have largely focused on a single facet of the cell, such as genome sequence or transcription levels using independent, targeted experiments. However, recent developments have begun to link these approaches, to derive unique and more holistic understanding of the forces driving a given cell state or behavior. Despite the potential for new discovery in this field, precious few chemical tools are readily available to enable these studies. Here, we present a new method to repeatedly and site-selectively *write*, *modify*, *erase* and *rewrite* functionality to an biomolecular target. Underpinning this method is the synthesis of two novel cofactor analogues, carrying chemically-dynamic hydrazone or oxime linkers. We demonstrate, using methyltransferase-directed DNA modification, the site-selective *writing* of chemical functionality onto (unmethylated) DNA. The written modifications can be subsequently *modified*, *erased* and *rewritten* or replaced by some permanent, *further functionality*. We demonstrate the versatility of the methodology by *writing*, *erasing* and *rewriting* fluorescent dyes to DNA and by developing a protocol to capture and release DNA using magnetic beads, with subsequent amplification of the sorted DNA sample for further analysis. We show that imaging, capture, sorting, release and amplification of the same DNA target (unmethylated DNA in this case) is possible. Our approach enables the application of multiple analytical techniques to the same sample, and will underpin new integrative analyses of targeted biomolecular processes.

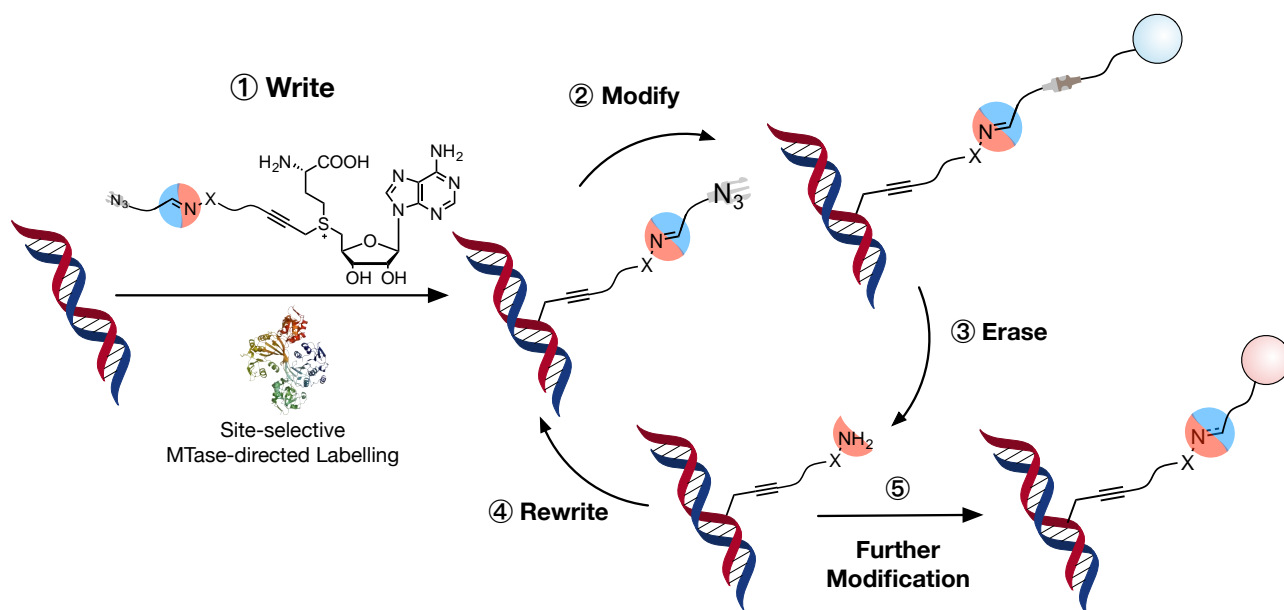
In an ideal experiment, it would be possible to completely understand the behavior of a cell by observing every aspect of its response to a given stimulus. However, methodologies, such as sequencing, have necessarily been developed in isolation, focused on improving understanding of specific biological problem. From these experiments, we now understand, for example, that knowledge of the genome sequence alone is not sufficient to understand a cell's disease state. Since gene expression levels can depend on the epigenetic state of the genome, its accessibility and even a gene's location in the cell nucleus an integrated approach to studying the genome is likely to yield improved understanding. To address this complexity, recent efforts have begun to tie together multiple methods and analyses, in a series of integrated analyses, such that the links between, for example, sequence, architecture and expression levels can be drawn, at the single cell level.

The site-selective, chemical modification (or functionalization) of proteins and nucleic acids has contributed to much of the methodology that underpins integrative single-cell analyses.¹⁻⁶ Selective introduction of non-natural chemical moieties to targeted biomolecules has enabled their isolation and purification, their imaging in cells and the study of other features of the system, such as post-translational modifications.^{1,4,7-13} In

recent years, there has been a growing interest in the use of chemoenzymatic methods to label biomolecules.^{14–16} These methods not only introduce site-selective chemical modifications but can label these biomolecules in complex mixtures, such as living cells or their lysates.

The methyltransferases (MTases) are emerging as a key class of enzymes for the site-selective modification of biomolecules because of their array of targets and their ability to add a wide variety of functional groups to these targets.^{17–20} MTases catalyse the transfer of a methyl group from the cofactor, S-adenosyl-L-methionine (AdoMet), to targets as diverse as small molecules, proteins and nucleic acids. Most importantly, a wide range of MTases can accommodate larger groups in their binding pocket, so that by manipulating the chemical structure of AdoMet it is possible to hijack MTase machinery to introduce functional groups to biomolecules.^{17,19–24} Commonly, MTase-directed ligation is used to introduce “clickable” groups for further modification,^{21–23,25–27} and this technology is now finding application in imaging, and genomic and metabolomic analysis. However, this strategy relies on the covalent attachment of often bulky functional groups, that can compromise the chemical and physical properties of the targeted biomolecule.²⁸ Recent examples have explored the use of disulfide linkers²⁸ or the use of light-cleavable moieties^{29,30} that can be erased under the right circumstances. However, these approaches still lack the required versatility to perform multiple modifications of the targeted DNA.

We present a new methodology for the modification of nucleic acids. The selectivity and robustness of MTase-directed labelling is combined with the versatility of dynamic covalent chemistry, to introduce chemical functionality to nucleic acids that can be *written, modified, erased* and *rewritten*. To this end, a new class of dynamic AdoMet derivatives, containing hydrazone or oxime linkers at the sulfonium center, was prepared. We demonstrate that the presence of these Schiff-bases does not affect the ability of representative MTases to label DNA selectively. The potential to *write, modify, erase* and *rewrite* or *further functionalize* DNA is demonstrated using a range of analytical techniques and exploited to develop a new protocol to capture and release fragments of DNA, using magnetic beads. Using qPCR, we show that the released DNA retains the functionality of an unmodified template molecule for the PCR.



Scheme 1 – Schematic representation of the reversible and rewritable modification of DNA. Site-selective MTase-directed *writing* of DNA ① allows for *modification* via azide-alkyne cycloaddition ②. The introduced functionality can then be *erased* via dynamic exchange ③, to give an intermediate DNA that can be *rewritten* via Schiff-base formation ④ to give the original functionality introduced via MTase-labelling. Alternatively, this intermediate DNA can be *further functionalized* via standard conjugation techniques ⑤ to give functional DNAs for follow-on research.

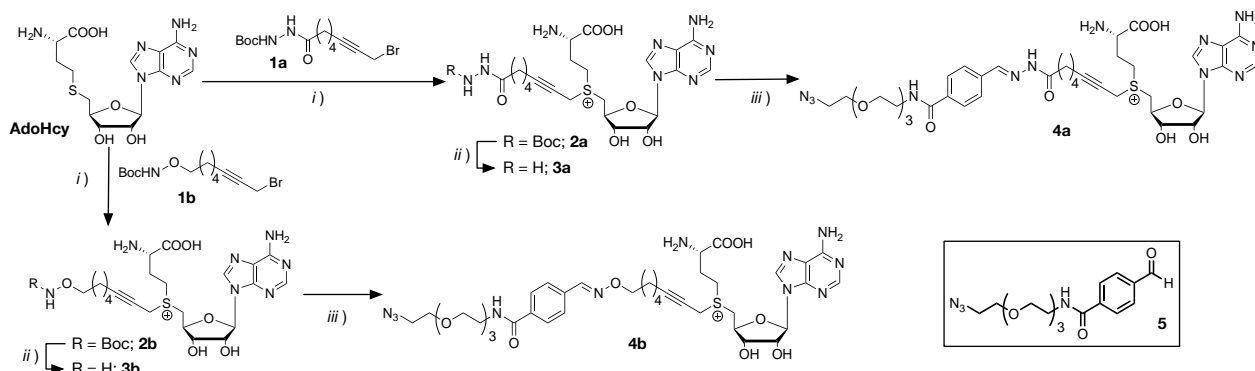
Results & Discussion

Cofactor Synthesis

We aimed to develop a reversible and rewritable AdoMet analogue. While several strategies are available to introduce reactive moieties to AdoMet derivatives, few of these are reversible,^{28–30} and allows this reversibility under mild conditions and with the flexibility for further functionalization. Ideally, this reactive linker should carry a chemical functionality orthogonal to common biological moieties (e.g. hydroxyl, amino and carboxyl). In nature, Schiff bases are commonly used to this end and chemists have now exploited the versatility of the C=N bond to develop a plethora of applications for this dynamic chemistry.^{31–36} The stability of the C=N bond can be tailored as a function of the “amine” used, with hydrazides and alkoxyamines demonstrating the largest ranges of stability.

With these principles in mind, linkers **1a** and **1b**, carrying a protected hydrazide or a protected alkoxyamine respectively, were prepared and reacted with S-adenosyl-L-homocysteine (**AdoHcy**) under acidic conditions to give Boc-protected AdoMet derivatives **2** (Scheme 2). Cofactors **2** were purified to remove the excess of linker **1** and deprotected under acidic conditions, to reveal the hydrazide or alkoxyamine moieties needed at later stages for further functionalization. Both diastereomers of cofactors **3** could be separated via HPLC (**Error! Reference source not found.** and **Error! Reference source not found.**), a separation that was not possible at later stages. These deprotected AdoMet derivatives slowly degraded (**Error! Reference source not found.**–**REF_Ref510985020 \h * MERGEFORMAT Error! Reference source not found.**), in particular following freeze-drying (**Error! Reference source not found.**). For instance, HPLC analysis of cofactor **3b** after freeze-drying revealed the presence of an additional peak at higher retention times (~31 min). MS analysis of this peak suggested a mass of

536.61 Da, very close to that of the alkoxy AdoMet derivative (537.62 Da). This difference in molecular mass, together with the increase in retention time, suggested that degradation of **3b** was also occurring via nucleophilic attack of the terminal amine of the linker to the sulfonium center (**Error! Reference source not found.**).



Scheme 2 – Synthesis of dynamic AdoMet derivatives 4. Conditions: i) HCO_2H , AcOH, 35 °C; ii) TFA; iii) 1.- TFA, 2.- 5, 20 mM ammonium acetate pH 5.5 Water/MeOH. Full details can be found in the Supporting information.

While degradation of AdoMet derivatives is common, and does not compromise their application, degradation is normally suppressed upon storage of these cofactors at low temperature and in a mildly acidic buffer (in our case 0.1% acetic acid 15-20 mM in cofactor).^{37,38} However, we decided to react AdoMet derivatives **3** immediately after purification by HPLC, to minimize these side reactions due to the nucleophilic nature of the hydrazide and alkoxyamine moieties. Commercially available benzaldehyde **5** was selected for this purpose (Scheme 2), as it incorporated a terminal azide. This way, the formed cofactors **4** contain a reactive moiety suitable for modification, while the condensation of the aldehyde with the hydrazide or the alkoxyamine incorporates a dynamic functionality, suitable for reversible and further functionalization. A slight excess of aldehyde **5** (1.2 equiv.) was employed to ensure full functionalization of the intermediate **3**, and the obtained cofactors **4** freeze-dried and stored in 0.1% acetic acid solution without further purification. No degradation of these AdoMet derivatives **4** was observed following this protocol (**Error! Reference source not found.**, **4b** as a representative example).

Writing chemical functionality – MTase-directed labelling

Following successful synthesis of AdoMet derivatives **4**, a restriction assay was used to demonstrate activity of the MTases with these new cofactor analogues.²⁷ M.TaqI, an N6-adenine DNA MTase, was incubated with cofactor analogue and pUC19 DNA, which has four recognition sites (TCGA) for this enzyme (Figure 2).³⁹ Successful transfer the functional group by M.TaqI results in protection of the plasmid from restriction digestion by R.TaqI, an endonuclease with the same target site as M. TaqI.⁴⁰

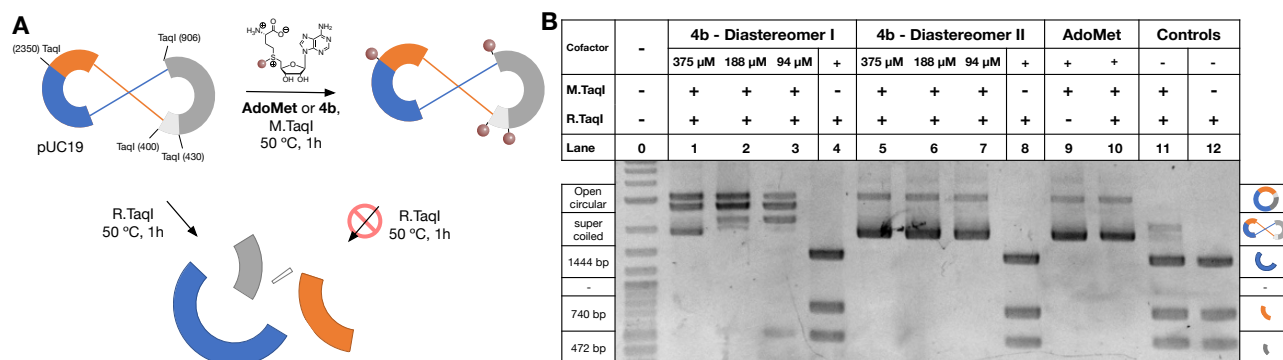


Figure 2 – MTase-directed *writing* of DNA. (A) Schematic representation of restriction assay and (B) gel electrophoresis of pUC19 following enzymatic treatment with M.TaqI and/or R.TaqI in the presence and absence of AdoMet (375 μ M) or AdoMet derivative **4b**. In the absence of M.TaqI-mediated alkylation (lanes 4, 8 and 12), pUC19 is cut into fragments, of which the largest three can be identified by gel electrophoresis. M.TaqI-mediated alkylation with AdoMet (lane 10) or derivative **4b** (lanes 1-3 and 5-6) results in partial to full protection from restriction by R.TaqI, with mainly open circular and supercoiled plasmid DNA being observed by gel electrophoresis. Controls in the absence of AdoMet derivatives (lanes 11 and 12), in the absence of M.TaqI (lanes 4, 8 and 12) and in the absence of R.TaqI (lane 9) are included. The gel clearly shows that catalysis of DNA transalkylation by M.TaqI with diastereomer I is less efficient than diastereomer II.

The native M.TaqI substrate, AdoMet, was used as a positive control for these experiments. In the absence of cofactor, M.TaqI is unable to alkylate pUC19 and bands corresponding to the three biggest DNA fragments formed upon digestion with R.TaqI were observed (Figure 2, lane 11). A similar effect is observed in the absence of M.TaqI (Figure 2, lane 12), demonstrating that neither isomer of AdoMet derivative **4b** interferes with the ability of R.TaqI to digest plasmid DNA (Figure 2, lane 4 and 8). More importantly, in the presence of both diastereomers of **4b**, M.TaqI was able to functionalize pUC19, with limited evidence of DNA digestion. In this case, only bands corresponding to open circular or supercoiled DNA were observed (Figure 2, lanes 1-3 and 5-7), in agreement with those observed when AdoMet was used (Figure 2, lane 10) or when no digestion was performed (Figure 2, lane 9). Dilution of the amount of cofactor used revealed that the second fraction had a higher activity. A similar effect was observed for the hydrazone derivative **4a** (**Error! Reference source not found.**), demonstrating that both analogues had the potential to be employed for the dynamic labelling of biomolecules. The reactivity of both diastereomers was unexpected, as it is commonly assumed that MTases are stereoselective,²² and thus we considered that isomerization around the sulfonium center was occurring in the assay conditions. However, no isomerization was observed by HPLC following incubation at 50 °C of cofactors **4b** (**Error! Reference source not found.**). Hence, the activity of M.TaqI with both isomers of **4b** is likely due to a lack of specificity (for this stereocentre) of this MTase.^{19,41}

Having demonstrated that M.TaqI was able to catalyze transalkylation of plasmid DNA with the dynamic cofactors **4**, we then decided to test whether other relevant DNA MTases could use these **AdoMet** derivatives as cofactors. M.MpeI is a cytosine-C5 MTase, that not only belongs to a different class, but targets the CpG dinucleotide, whose methylation is involved in the mechanism of gene regulation in vertebrates.^{28,42,43} Thus, pUC19 was incubated with mutant M.MpeI (Q136A, N347A) and **4a**, and then challenged with R.HaeII, a restriction enzyme that targets a subset of the CpG dinucleotides.⁴⁴ To

our delight, we observed efficient transalkylation of plasmid DNA (**Error! Reference source not found.**).

We sought direct evidence of the ability of MTases to alkylate DNA with AdoMet derivatives **4** by targeting a 14 base pair (bp) oligonucleotide with one copy of M.TaqI sequence (TCGA) for transalkylation. Labelling of the oligo was monitored directly using HPLC. This analysis was performed above the melting temperature of the DNA so that both strands could be clearly identified in the chromatogram (Figure 3). A clear shift in the retention time was seen upon labelling with AdoMet (Figure S3.2), **4a** or **4b** (Figure 3, middle) when compared to the retention times of the unmodified DNA (Figure 3, top and Figure S3.1). This shift was observed for both peaks, demonstrating that M.TaqI was able to label both strands as a consequence of the palindromic nature of the sequence this MTase recognises. Moreover, the shift was proportional to the size and nature of the linker transferred, with the AdoMet methylation resulting in a small shift in retention time (Figure S3.2) and the oxime derivate **4b** giving the biggest shift (Figure 3). In these chromatographs, we could also see what we expected to be, and later confirmed, free hydrazide (Figure 3, left, peaks at ~ 13 mins) and hydroxylamine (Figure 3, right, peaks at ~ 19 mins). The presence of this small amount of *erased* oligo DNA is likely due to hydrolysis under the HPLC conditions. Analysis of the individual peaks was carried out using MS which confirmed labelling was successful and the nature of the sidechain functionality introduced following incubation (Figure S3.1-Figure S3.10).

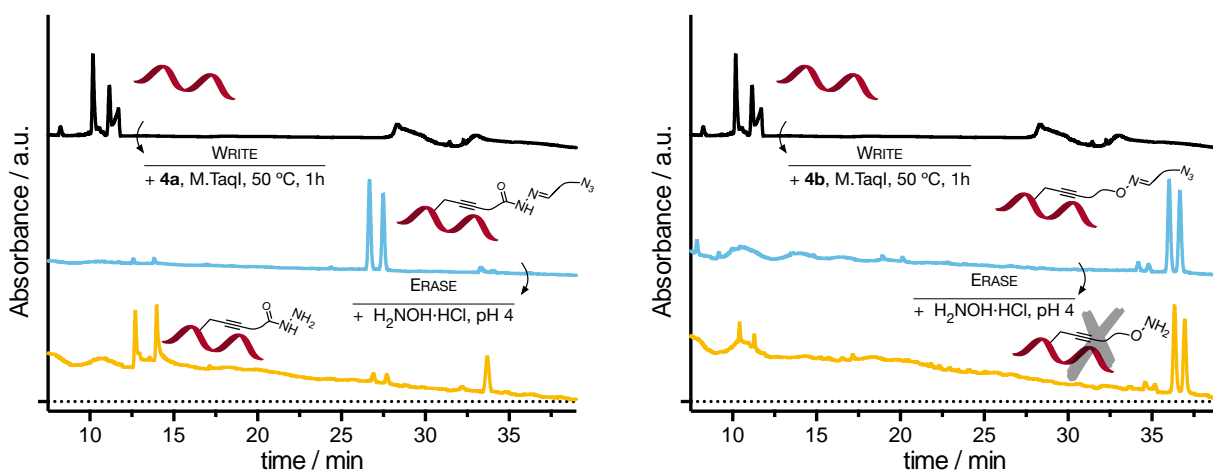


Figure 3 – Writing and erasing the introduced functionality. Analytical HPLC chromatograms of oligo DNA (top) and oligo DNA following incubation with M.TaqI and **4** (middle), and with M.TaqI and **4**, followed by incubation with 10 equiv. of H₂NOH·HCl in 10 mM ammonium acetate pH4 (bottom). HPLC conditions: 0.1 M triethylammonium acetate buffer, pH 7.0 (A)/MeCN (B) gradient at 60 °C. Under these conditions, oligo DNA melts and both strands of DNA can be observed independently.

Erasing the introduced functionality – dynamic cleaving of the Schiff-base

Following successful labelling of the oligo DNA, these samples were then used to demonstrate the reversible nature of the Schiff-base introduced and its potential to be efficiently cleaved. Aliquots of labelled oligo DNA were incubated at 50 °C for 1.5 h with 10 eq. of H₂NOH·HCl. This competing reagent was introduced to facilitate the cleavage of the Schiff-base and the pH of the samples was adjusted to pH 4.0.^{45,46} HPLC analysis showed a clear shift in the retention time of the hydrazone-labelled DNA following treatment with the competing reagent (Figure 3, left). Over 85% of the functional linker was cleaved under these conditions. The new peak shifted to lower retention time, as

expected following the loss of the potentially hydrophobic aldehyde **5**. MS analysis confirmed that these peaks (~13 min) corresponded to the hydrazide labelled oligo DNA (Figure S4.15-Figure S4.16). Conversely, under these conditions the oxime-labelled DNA remained intact (Figure 3, right and **Error! Reference source not found.**), consistent with the higher stability of this type of Schiff-bases.⁴⁷

Rewriting the original functionality via Schiff-base formation

To demonstrate rewritability of **4a**, a hydrazide functionalized oligo DNA (Figure 4B) was incubated in the presence of an excess of aldehyde **5**. As predicted, a clear shift in the retention time of the main peaks associated with oligo DNA was observed (Figure 4C). Comparison of this chromatogram to that obtained following incubation of oligo DNA with M.TaqI and cofactor **4a** (Figure 4D), shows good overlap of the peaks associated with azide-functionalized DNA at ~6.2 and 6.4 mins, and a similar ratio of this peak to that of the free hydrazide (at 3.8 and 4.0 mins). The presence of a small amount of hydrazide-functionalized oligo DNA is likely due to hydrolysis of the Schiff base under the HPLC conditions. Three additional peaks were observed following *rewriting* with aldehyde **5**, which overlapped with those observed when this aldehyde was incubated with H₂NOH·HCl (Figure 4E). Analysis of the individual peaks was carried out by LC-MS which confirmed *rewriting* was successful and the nature of the chemical functionality on the oligo DNA (**Error! Reference source not found.**-**Error! Reference source not found.** und.).

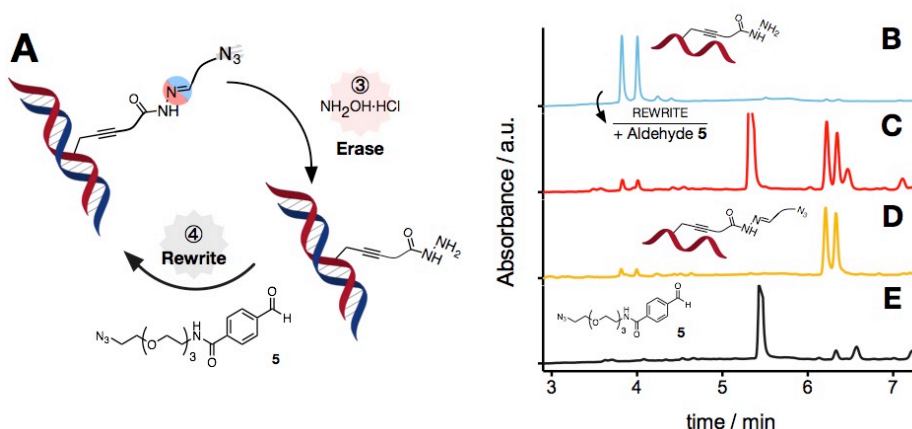


Figure 4 – Re-writing DNA. (A) Schematic representation of *erasing* and *rewriting* of chemical functionality on oligo DNA using the dynamic AdoMet derivative **4a** and aldehyde **5**. Analytical HPLC chromatograms of oligo DNA following incubation with M.TaqI and **4a**, followed by incubation with 10 equiv. of H₂NOH·HCl in 10 mM ammonium acetate pH4 (B), and oligo DNA from B followed by incubation with aldehyde **5** (C). Chromatogram of oligo DNA following incubation with M.TaqI and **4a** (D) and of aldehyde **5** incubated with 10 equiv. of H₂NOH·HCl in 10 mM ammonium acetate pH4 (E) shown for comparison. HPLC conditions: 0.1 M triethylammonium acetate buffer, pH 7.0 (A)/MeCN (B) gradient at 60 °C. Under these conditions, oligo DNA melts and both strands of DNA can be observed independently.

Further Modification following Schiff-base cleavage.

One common limitation of the current chemistries used for site selective modification of biomolecules is their lack of versatility. Once a chemical moiety is introduced to, for instance, facilitate the purification of the biomolecule (e.g. biotin), these moieties remain attached to the targeted biomolecule. More importantly, these moieties can not be further functionalized under mild and straightforward conditions, to introduce new functionality (e.g. fluorophores, targeting ligands) often required for further research. Our strategy for

MTase-directed labelling of DNA using dynamic AdoMet derivative **4a** results in DNA that now carries a Schiff-base that can be efficiently cleaved. *Erasing* the chemical functionality this way results in a hydrazide-modified DNA (Figure 3), that should be easily functionalized using standard bioconjugation techniques. To this end, fragments of DNA generated by PCR, containing 17 CpG sites,⁴⁸ were site-selectively labelled (*write*) with M.Mpel. Labelling was followed by incubation with H₂NOH·HCl (*erase*) and reaction with a commercially available NHS-activated fluorophore Atto 647N **7** (*further functionalization*) (**Error! Reference source not found.A**). The reaction was monitored via gel electrophoresis (**Error! Reference source not found.B**) and shows specific conjugation of Atto647N to DNA modified with **4a**. While no red-fluorescence was observed in the absence of Atto-647N (**Error! Reference source not found.B**, lanes 1, 3 and 3), this dye is positively charged and was able to non-specifically associate with the DNA in the control samples (**Error! Reference source not found.B**, lanes 2, 4 and 6). Comparison of the intensities of the red and green (Sybr® Green) channels, to evaluate the degree of labelling with Atto 647N **7** per unit DNA (**Error! Reference source not found.B**, lanes 6) showed a degree of labelling 4.8 times higher for the DNA that has undergone the *write-erase* than in the absence of the *erase* step, H₂NOH·HCl treatment (**Error! Reference source not found.B**, lanes 4), and over 23 times higher than in the absence of the *write* step, MTase-directed labelling (**Error! Reference source not found.B**, lanes 2).

Dual Modification – Sequential modification with complementary fluorescent dyes

To further demonstrate the utility of our strategy and the potential to perform multiple modifications on the same sample of DNA, we decided to explore the consecutive labelling of short DNA fragments with two different fluorescent dyes (*write-modify-erase-further modification*). To this end, DNA fragments were first incubated with M.Mpel and dynamic cofactor **4a**, to yield azide-functionalized DNA (Figure 5A, step ①). This modification resulted in a small shift in the migration time of the DNA on gel (Figure 5B, step ①) but, as expected, no fluorescence was observed (Figure 5C, step ①). A further shift in the migration time was observed when the azide-functionalized DNA was modified with TAMRA-DBCO **6** (Figure 5B, step ②) but, more importantly, emission from DNA-associated TAMRA fluorophore was clearly observed (Figure 5C, step ②). *Erasing* the TAMRA modification was achieved by incubation with an excess of H₂NOH·HCl. No fluorescence was observed from the resulting DNA fragments (Figure 5C, step ③) and a shift back to the original migration time was observed (Figure 5B, step ③), suggesting that this hydrazide linker had little impact on the physical properties of the DNA. Incubation of this hydrazide-functionalized DNA with NHS-activated Atto 647N **7** resulted in a new shift in migration time (Figure 5B, step ④) and corresponding appearance of fluorescence, now visible under red illumination (Figure 5C, step ④).

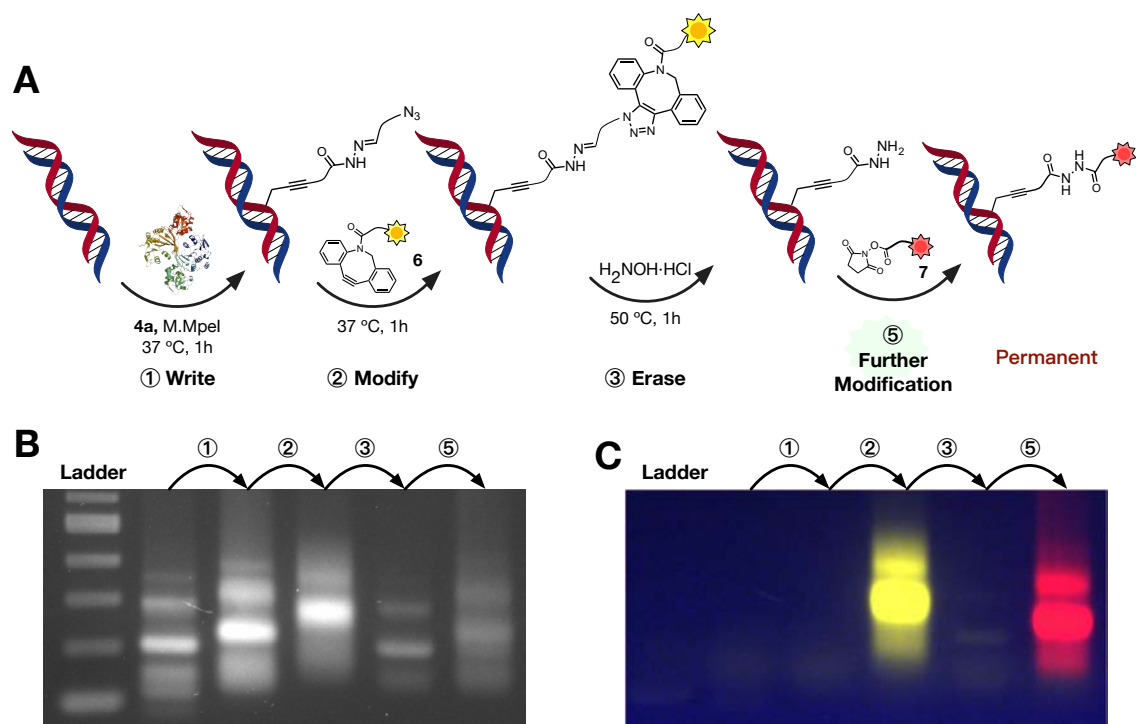


Figure 5 – Dual functionalization. (A) Schematic representation of the dual functionalization of DNA: DNA was first *written* via site-selective MTase-directed labelling using dynamic cofactor **4a** ①. The obtained azide-functionalized DNA fragments were then *modified* with TAMRA-DBCO **6** via azide-alkyne cycloaddition ②. The introduced TAMRA was then *erased* via dynamic exchange ③, to give hydrazide-functionalized DNA fragments that were further functionalized with NHS activated Atto 647N **7** ⑤. Modification was monitored using gel electrophoresis. Conditions: DNA concentration; 7 ng/ μ L, release buffer; 10 mM Ammonium Acetate, pH 6.8, 1 M NaCl, 0.01% SDS. DNA stained with GelRed®. Gel was visualized using a Bio-Rad Pharos FX (GelRed®: excitation, trans-UV; emission filter, 590/110 nm; TAMRA: excitation, epi-green illumination; emission filter: 602/50 nm; Atto 647N **7**: excitation, epi-red illumination; emission filter: 700/50 nm). TAMRA channel was colored yellow and Atto 647N **7** was colored red for visualization. (B) GelRed® channel and (C) Composite image of TAMRA and Atto647n channels. Full chemical structure of the fluorescent dyes are available in the Supporting Information.

Application - Reversible capture of DNA and amplification via PCR

Having demonstrated that dynamic AdoMet derivative **4a** not only let us site-selectively *write* DNA with a “clickable” tag but, more importantly, it allows us to *modify*, *erase* and *rewrite* the introduced functionality, we decided to test the potential of this methodology to underpin the isolation and purification of DNA. Isolation is normally achieved by labelling the DNA with affinity tags such as biotin, followed by capture with magnetic beads or via affinity chromatography. Release of the captured nucleic acid is commonly achieved using either denaturing conditions, that disrupt the biotin-(strept)avidin binding, or by cleaving a dithio (S-S) moiety placed in the linker between the biomolecule and the bead/resin. However, in both cases, the conditions used for release can result in low recovery of the targeted biomolecule, *vide infra*. We anticipated that a similar strategy could be employed using the dynamic AdoMet derivative **4a**, but the presence of the hydrazone linker would allow us to efficiently release the captured DNA under mild conditions.

Our studies thus focused on the capture and release of the 203 bp DNA fragment containing 17 CpG sites targeted for alkylation by M.Mpel.⁴⁸ After labelling with cofactor **4a**, the fragments were further modified by reaction with a bi-functional linker molecule,

carrying biotin and dibenzocyclooctyne (DBCO) groups (Figure 6A). MTase-directed labelling was also performed with AdoMet derivative **8** that, while including an azide moiety for further functionalization using DBCO chemistry, did not contain the dynamic hydrazone linker. Capture experiments, using streptavidin-coated beads, were performed using a high salt tris buffer. We consistently captured in excess of 80% of the DNA (Figure 6C, **Error! Reference source not found.** and **Error! Reference source not found.**), while almost no unlabeled DNA was adsorbed onto the magnetic bead under these conditions (Figure 6C, **■**).

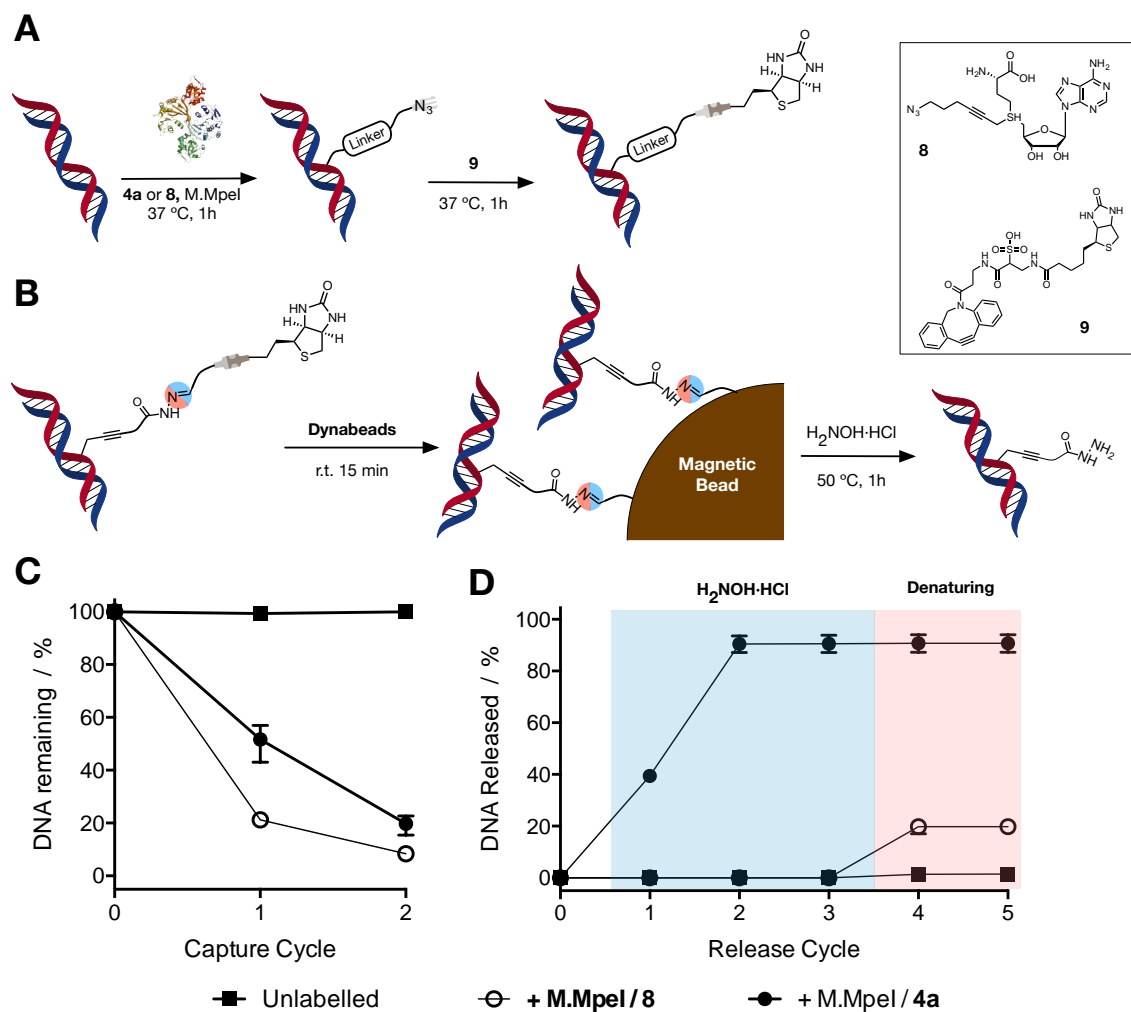


Figure 6 – Capture and release of DNA using MTase-directed labelling and dynamic exchange. (A) Schematic representation of functionalization of DNA fragments with a biotin tag using both MTase-directed labelling and subsequent DBCO conjugation. (B) Schematic representation of the capture and release of DNA fragments using the dynamic AdoMet derivative **4a**. (C) Percentage of DNA remaining following capture with magnetic beads and (D) percentage DNA released from the magnetic beads following treatment with H₂NOH·HCl (cycles 1-3) and denaturing conditions (reflux in denaturing buffer, cycles 4 and 5). DNA was either unlabeled (**■**), incubated with M.Mpel and **4a** (**●**), or incubated with M.Mpel and azide containing cofactor **8** (**○**). Conditions: DNA concentration; 7 ng/μL; capture buffer: 10 mM Tris, 1 M NaCl, pH 8.5; dynamic release buffer: 20 mM H₂NOH·HCl, 10 mM ammonium acetate, pH 6.8, 1 M NaCl, 0.01% SDS; denaturing buffer.

Following capture, the DNA coated beads were washed and then suspended in an ammonium acetate buffer solution with H₂NOH·HCl at 50 °C, to release the captured DNA via cleavage of the dynamic hydrazone linker. Initial experiments with 3 equiv. of

H₂NOH·HCl per CpG site on the DNA failed to release significant amounts of captured DNA (**Error! Reference source not found.**). Increasing the amount of this competing agent significantly increased the amount of released DNA up to ~50% of the original DNA captured. We believe the need to increase the amount of H₂NOH·HCl is probably due to a combination of factors including moving from solution to solid phase, the relative high number and density of labelling sites present in the captured DNA, and the length of the DNA (203 bp in PCR fragments vs 14 bp for oligo DNA). The effect of pH on the release efficiency was also evaluated (**Error! Reference source not found.**). While the optimal pH for hydrazone exchange is around pH 4.5,^{45,46} the degree of ionization of the captured DNA would be higher at higher pH, minimizing non-specific interactions between the bead and the DNA. In our case, no significant effect of pH was observed on the amount of DNA released, with similar levels obtained for all three-pH's tested (4.1, 4.9 and 7) (**Error! Reference source not found.**). Further optimization of this protocol included reducing the concentration of DNA used during the capture experiments, with almost quantitative release obtained under these conditions (Figure 6D, ●).

To demonstrate that release was the result of the cleavage of the dynamic hydrazone linker and not the result of non-specific interactions of the excess of H₂NOH·HCl, a sample of DNA was modified using M.Mpel and azide-containing AdoMet derivative **8**. This cofactor lacks the required dynamic chemistry for reversible capture of the DNA (Figure 6B) but reacts in a similar fashion to **4a** in the capture of MTase-labelled DNA (Figure 6A). In fact, similar levels of DNA capture were obtained using either of the cofactor analogues, **4a** or **8**. Incubating the captured DNA with an excess of H₂NOH·HCl gave no observable release of DNA modified with the non-dynamic cofactor **8**, even after four cycles of incubation (Figure 6D, ○). This lack of release was in sharp contrast to the almost quantitative release observed for DAN labelled with the dynamic AdoMet derivative **4a** (Figure 6D, ●). DNA labelled with AdoMet derivative **8** was only released from the magnetic beads using standard denaturing conditions (boiling in 0.1% SDS), although only 20% of the captured DNA was recovered. Moreover, no additional release of DNA was observed from those beads used to capture the DNA modified with dynamic cofactor **4a** (Figure 6D, ○).

To demonstrate that the presence of the hydrazide moiety did not affect the functionality of the released DNA, we tested the ability of a polymerase enzyme (Sso7d, BioRad) to further amplify DNA that was captured and released from beads. Real-time (RT) PCR (SsoAdvanced Universal SYBR Green Supermix, Biorad) showed an amplification curve that evolved in a similar fashion for both the modified and control DNA samples (**Error! Reference source not found.**). Furthermore, the concentration determined using RT-PCR (35 ± 11 ng/μL) was in good agreement with that obtained using a complementary, fluorescence-based measurement (Qubit fluorimeter) of the eluted DNA, (concentration for the DNA 38 ng/μL). These results demonstrate the potential of our *write*, *modify* and *erase* approach, and of AdoMet derivative **4a** to underpin the development of a mild and selective method for the purification of DNA with high recovery efficiencies.

Conclusion

We have presented a new versatile strategy to site-selective *write*, *modify*, *erase* and *rewrite* chemical functionality onto DNA. We envisage this flexible, efficient and straightforward approach finding widespread application in biotechnology, and particularly in the area of integrative analysis of biomolecules, where our chemical

handles can be used to isolate, label or functionalise a molecule for a series of complementary studies.

As an exemplar of the application of this chemistry, two new dynamic S-adenosyl-L-methionine (AdoMet) derivatives incorporating Schiff-bases in the transferable linker have been prepared, and the ability of two relevant methyltransferases (MTases) to label DNA using these cofactors has been demonstrated. The versatility of this technology has been explored by labelling DNA across a range of sizes, including oligo and plasmid DNA. The combination of MTase-directed labelling and dynamic covalent chemistry let us site-selectively *write* functionality into the DNA that can be *modified* using common bioconjugations techniques and then *erased* (i.e. via dynamic exchange) under mild conditions. The resulting hydrazide-functionalized DNA can not only be *further functionalized* (eg. via labelling with a fluorescent tag) to, for instance, facilitate follow-on experiments, but also *rewritten* to recover the original Schiff-base, and the chemical functionality introduced via MTase-directed labelling. Sequential labelling with two complementary fluorescent dyes has been demonstrated, via *writing*, *modification*, *erasing* and *further functionalization*. The potential of the presented methodology was explored by developing a new protocol to capture PCR fragments of plasmid DNA, that could be quantitatively recovered under mild conditions, and further amplified using RT-PCR.

In summary, we have reported a versatile chemistry to enable consecutive analyses on the same target biomolecule (e.g. imaging → capture → sequence). Several MTases that target other biomolecules (e.g. RNA, proteins) have been reported (REFS 14, 17, 20), hence, we believe the presented methodology can have a significant impact on the site-selective labeling of these targets. Similarly, this methodology should be easily adapted to incorporate chemical moieties beyond those demonstrated here, to explore alternative chemistries for functionalization, and applications that go beyond bioconjugation, capture, release and fluorescently labelling of these biomolecules. Our efforts in these directions will be reported in due course.

Supplementary information

Full experimental details can be obtained in the Supporting Information.

AUTHOR INFORMATION

Author Contributions

A.W., E. J., R.K.N. and F.F.-T. conceived and designed the experiments. A.W., K.U and E.J. synthesized and purified cofactors **4**. A.E.R. and J.K. produced and confirmed the activity of the M.TaqI and M.MpeI enzymes. A.W. and E.J. the further modification experiment. E.J. performed the dual-modification experiment. A.W., K.U., E. J. and Q. S. designed and performed capture and release of DNA experiment. Q.S. performed RT-PCR and A.W. performed all other experiments. F.F.-T. and R.K.N. secured funding. A.W., E. J., R.K.N. and F.F.-T. analyzed the data and wrote the paper, with all other authors contributing to the final version of the manuscript.

Conflicts of Interest

The authors declare no conflicts of interest.

ACKNOWLEDGMENTS

The authors thank the University of Birmingham for the John Evans Fellowship to F.F.-T. and scholarships to A.W., K.U., A.E.R. and J.K. This work was supported by the

European Union 634890, 'BeyondSeq', Royal Society U.K (RG140613), the EPSRC (EP/N020901/1, EP/M506461/1), the Wellcome Trust (177ISSFPP) and the Birmingham Science City and the European Regional Development Fund.

REFERENCES

1. Krall, N., da Cruz, F. P., Boutureira, O. & Bernardes, G. J. L. Site-selective protein-modification chemistry for basic biology and drug development. *Nat. Chem.* **8**, 103 (2015).
2. Haimovich, A. D., Muir, P. & Isaacs, F. J. Genomes by design. *Nat. Rev. Genet.* **16**, 501 (2015).
3. Wright, T. H., Vallée, M. R. J. & Davis, B. G. From Chemical Mutagenesis to Post-Expression Mutagenesis: A 50 Year Odyssey. *Angew. Chem. Int. Ed.* **55**, 5896–5903 (2016).
4. Hu, Q.-Y., Berti, F. & Adamo, R. Towards the next generation of biomedicines by site-selective conjugation. *Chem. Soc. Rev.* **45**, 1691–1719 (2016).
5. Pugh, G. C., Burns, J. R. & Howorka, S. Comparing proteins and nucleic acids for next-generation biomolecular engineering. *Nat. Rev. Chem.* **2**, 113–130 (2018).
6. Spicer, C. D., Pashuck, E. T. & Stevens, M. M. Achieving Controlled Biomolecule–Biomaterial Conjugation. *Chem. Rev.* **118**, 7702–7743 (2018).
7. Cobo, I., Li, M., Sumerlin, B. S. & Perrier, S. Smart hybrid materials by conjugation of responsive polymers to biomacromolecules. *Nat. Mater.* **14**, 143–159 (2014).
8. Chen, Y.-J., Groves, B., Muscat, R. A. & Seelig, G. DNA nanotechnology from the test tube to the cell. *Nat. Nanotechnol.* **10**, 748–760 (2015).
9. Lapa, S. A., Chudinov, A. V. & Timofeev, E. N. The Toolbox for Modified Aptamers. *Mol. Biotechnol.* **58**, 79–92 (2016).
10. Li, J., Green, A. A., Yan, H. & Fan, C. Engineering nucleic acid structures for programmable molecular circuitry and intracellular biocomputation. *Nat. Chem.* **9**, 1056–1067 (2017).
11. Raiber, E.-A., Hardisty, R., van Delft, P. & Balasubramanian, S. Mapping and elucidating the function of modified bases in DNA. *Nat. Rev. Chem.* **1**, 0069 (2017).
12. Wang, L. Engineering the Genetic Code in Cells and Animals: Biological Considerations and Impacts. *Acc. Chem. Res.* **50**, 2767–2775 (2017).
13. Young, D. D. & Schultz, P. G. Playing with the Molecules of Life. *ACS Chem. Biol.* **13**, 854–870 (2018).
14. Anhäuser, L. & Rentmeister, A. Enzyme-mediated tagging of RNA. *Chem. Biotechnol. • Pharm. Biotechnol.* **48**, 69–76 (2017).
15. Li, C. & Wang, L.-X. Chemoenzymatic Methods for the Synthesis of Glycoproteins. *Chem. Rev.* **118**, 8359–8413 (2018).
16. Zhang, Y., Park, K.-Y., Suazo, K. F. & Distefano, M. D. Recent progress in enzymatic protein labelling techniques and their applications. *Chem. Soc. Rev.* (2018). doi:10.1039/C8CS00537K
17. Zhang, J. & Zheng, Y. G. SAM/SAH Analogs as Versatile Tools for SAM-Dependent Methyltransferases. *ACS Chem. Biol.* **11**, 583–597 (2016).
18. Lyko, F. The DNA methyltransferase family: a versatile toolkit for epigenetic regulation. *Nat. Rev. Genet.* **19**, 81–92 (2017).
19. Deen, J. *et al.* Methyltransferase-Directed Labeling of Biomolecules and its Applications. *Angew. Chem. Int. Ed.* **56**, 5182–5200 (2016).

20. Tomkuvienė, M., Mickutė, M., Vilkaitis, G. & Klimašauskas, S. Repurposing enzymatic transferase reactions for targeted labeling and analysis of DNA and RNA. *Anal. Biotechnol.* **55**, 114–123 (2019).
21. Dalhoff, C., Lukinavičius, G., Klimašauskas, S. & Weinhold, E. Direct transfer of extended groups from synthetic cofactors by DNA methyltransferases. *Nat. Chem. Biol.* **2**, 31–32 (2005).
22. Dalhoff, C., Lukinavičius, G., Klimašauskas, S. & Weinhold, E. Synthesis of S-adenosyl-L-methionine analogs and their use for sequence-specific transalkylation of DNA by methyltransferases. *Nat. Protoc.* **1**, 1879–1886 (2006).
23. Wang, R., Zheng, W., Yu, H., Deng, H. & Luo, M. Labeling Substrates of Protein Arginine Methyltransferase with Engineered Enzymes and Matched S-Adenosyl-L-methionine Analogues. *J. Am. Chem. Soc.* **133**, 7648–7651 (2011).
24. Luo, M. Chemical and Biochemical Perspectives of Protein Lysine Methylation. *Chem. Rev.* **118**, 6656–6705 (2018).
25. Islam, K. *et al.* Defining efficient enzyme–cofactor pairs for bioorthogonal profiling of protein methylation. *Proc. Natl. Acad. Sci.* **110**, 16778–16783 (2013).
26. Vranken, C. *et al.* Super-resolution optical DNA Mapping via DNA methyltransferase-directed click chemistry. *Nucleic Acids Res.* **42**, e50–e50 (2014).
27. Lauer, M. H. *et al.* Methyltransferase-directed covalent coupling of fluorophores to DNA. *Chem. Sci.* **8**, 3804–3811 (2017).
28. Kriukienė, E. *et al.* DNA unmethylome profiling by covalent capture of CpG sites. *Nat. Commun.* **4**, 2190 (2013).
29. Anhäuser, L., Muttach, F. & Rentmeister, A. Reversible modification of DNA by methyltransferase-catalyzed transfer and light-triggered removal of photo-caging groups. *Chem. Commun.* **54**, 449–451 (2018).
30. Heimes, M., Kolmar, L. & Brieke, C. Efficient cosubstrate enzyme pairs for sequence-specific methyltransferase-directed photolabile caging of DNA. *Chem. Commun.* **54**, 12718–12721 (2018).
31. Adams, J. P. Imines, enamines and oximes. *J. Chem. Soc. Perkin 1* 125–139 (2000).
32. Ramström, O. & Lehn, J.-M. Drug discovery by dynamic combinatorial libraries. *Nat. Rev. Drug Discov.* **1**, 26–36 (2002).
33. Wojtecki, R. J., Meador, M. A. & Rowan, S. J. Using the dynamic bond to access macroscopically responsive structurally dynamic polymers. *Nat. Mater.* **10**, 14–27 (2010).
34. Belowich, M. E. & Stoddart, J. F. Dynamic imine chemistry. *Chem. Soc. Rev.* **41**, 2003–2024 (2012).
35. Rosales, A. M. & Anseth, K. S. The design of reversible hydrogels to capture extracellular matrix dynamics. *Nat. Rev. Mater.* **1**, 15012 (2016).
36. Zhang, W. & Jin, Y. Dynamic covalent chemistry : principles, reactions, and applications. (John Wiley & Sons, 2018).
37. Lukinavičius, G., Tomkuvienė, M., Masevičius, V. & Klimašauskas, S. Enhanced Chemical Stability of AdoMet Analogues for Improved Methyltransferase-Directed Labeling of DNA. *ACS Chem. Biol.* **8**, 1134–1139 (2013).
38. Huber, T. D. *et al.* Functional AdoMet Isosteres Resistant to Classical AdoMet Degradation Pathways. *ACS Chem. Biol.* **11**, 2484–2491 (2016).
39. Goedecke, K., Pignot, M., Goody, R. S., Scheidig, A. J. & Weinhold, E. Structure of the N6-adenine DNA methyltransferase M•TaqI in complex with DNA and a cofactor analog. *Nat. Struct. Biol.* **8**, 121 (2001).
40. Sato, S., Hutchinson, C. A. & Harris, J. I. A thermostable sequence-specific endonuclease from *Thermus aquaticus*. *Proc. Natl. Acad. Sci.* **74**, 542 (1977).

41. Pljevaljčić, G., Schmidt, F. & Weinhold, E. Sequence-specific Methyltransferase-Induced Labeling of DNA (SMILing DNA). *ChemBioChem* **5**, 265–269 (2004).
42. Lister, R. *et al.* Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* **462**, 315–322 (2009).
43. Jeltsch, A. Phylogeny of Methylomes. *Science* **328**, 837–838 (2010).
44. Slatko, B. E., Croft, R., Moran, L. S. & Wilson, G. G. Cloning and analysis of the HaeIII and HaeII methyltransferase genes. *Gene* **74**, 45–50 (1988).
45. Dirksen, A., Dirksen, S., Hackeng, T. M. & Dawson, P. E. Nucleophilic Catalysis of Hydrazone Formation and Transimination: Implications for Dynamic Covalent Chemistry. *J. Am. Chem. Soc.* **128**, 15602–15603 (2006).
46. Mahon, C. S., Jackson, A. W., Murray, B. S. & Fulton, D. A. Templating a polymer-scaffolded dynamic combinatorial library. *Chem. Commun.* **47**, 7209–7211 (2011).
47. Kalia, J. & Raines, R. T. Hydrolytic Stability of Hydrazones and Oximes. *Angew. Chem. Int. Ed.* **47**, 7523–7526 (2008).
48. Wojciechowski, M., Czapinska, H. & Bochtler, M. CpG underrepresentation and the bacterial CpG-specific DNA methyltransferase M.MpeI. *Proc. Natl. Acad. Sci.* **110**, 105–110 (2013).

8 References

- (1) Mout, R.; Moyano, D. F.; Rana, S.; Rotello, V. M. *Chem Soc Rev* **2012**, 41 (7), 2539–2544.
- (2) Kaboord, B.; Perr, M. *Methods Mol. Biol.* **2008**, 424 (Chapter 27), 349–364.
- (3) Bumgarner, R. *Overview of DNA Microarrays: Types, Applications, and Their Future*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2001; Vol. 6, pp 829–17.
- (4) Cutler, J. I.; Auyeung, E.; Mirkin, C. A. *J Am Chem Soc* **2012**, 134 (3), 1376–1391.
- (5) Nicholas, M. P.; Rao, L.; Gennerich, A. *Methods Mol. Biol.* **2014**, 1136 (Chapter 9), 137–169.
- (6) Chinen, A. B.; Guan, C. M.; Ferrer, J. R.; Barnaby, S. N.; Merkel, T. J.; Mirkin, C. A. *Chem. Rev.* **2015**, 115 (19), 10530–10574.
- (7) Moter, A.; Göbel, U. B. *J. Microbiol. Methods* **2000**, 41 (2), 85–112.
- (8) Toseland, C. P. *J Chem Biol* **2013**, 6 (3), 85–95.
- (9) Niemeyer, C. M. *Angew. Chem. Int. Ed. Engl.* **2010**, 49 (7), 1200–1216.
- (10) Adler, M.; Wacker, R.; Niemeyer, C. M. *Analyst* **2008**, 133 (6), 702–718.
- (11) Jungmann, R.; Avendaño, M. S.; Woehrstein, J. B.; Dai, M.; Shih, W. M.; Yin, P. *Nat. Methods* **2014**, 11 (3), 313–318.
- (12) Sano, T.; Smith, C. L.; Cantor, C. R. *Science* **1992**, 258 (5079), 120–122.
- (13) Chang, L.; Li, J.; Wang, L. *Anal. Chim. Acta* **2016**, 910, 12–24.
- (14) Elzahhar, P.; Belal, A. S. F.; Elamrawy, F.; Helal, N. A.; Nounou, M. I. *Methods Mol. Biol.* **2019**, 2000 (2), 125–182.
- (15) Benizri, S.; Gissot, A.; Martin, A.; Vialet, B.; Grinstaff, M. W.; Barthélémy, P. *Bioconjugate Chem* **2019**, 30 (2), 366–383.
- (16) Ye, J.; Liu, E.; Gong, J.; Wang, J.; Huang, Y.; He, H.; Yang, V. C. *Theranostics* **2017**, 7 (9), 2495–2508.
- (17) Rowan, S. J.; Cantrill, S. J.; Cousins, G. R. L.; Sanders, J. K. M.; Stoddart, J. F. *Angew. Chem. Int. Ed. Engl.* **2002**, 41 (6), 898–952.
- (18) Kolb, H. C.; Finn, M. G.; Sharpless, K. B. *Angew. Chem. Int. Ed. Engl.* **2001**, 40 (11), 2004–2021.
- (19) Kolb, H. C.; Sharpless, K. B. *Drug Discov. Today* **2003**, 8 (24), 1128–1137.
- (20) Moses, J. E.; Moorhouse, A. D. *Chem Soc Rev* **2007**, 36 (8), 1249–1262.
- (21) Hogg, P. J. *Trends Biochem. Sci.* **2003**, 28 (4), 210–214.
- (22) Vranken, C.; Deen, J.; Dirix, L.; Stakenborg, T.; Dehaen, W.; Leen, V.; Hofkens, J.; Neely, R. K. *Nucleic Acids Res.* **2014**, 42 (7), –e50.
- (23) Rostovtsev, V. V.; Green, L. G.; Fokin, V. V.; Sharpless, K. B. *Angew. Chem. Int. Ed. Engl.* **2002**, 41 (14), 2596–2599.
- (24) Presolski, S. I.; Hong, V. P.; Finn, M. G. *Curr Protoc Chem Biol* **2011**, 3 (4), 153–162.
- (25) El-Sagheer, A. H.; Brown, T. *Chem Soc Rev* **2010**, 39 (4), 1388–1405.

- (26) Cervantes-Cervantes, M. P.; Calderón-Salinas, J. V.; Albores, A.; Muñoz-Sánchez, J. L. *Biological Trace Element Research* **2005**, *103* (3), 229–248.
- (27) Chan, T. R.; Hilgraf, R.; Sharpless, K. B.; Fokin, V. V. *Org. Lett.* **2004**, *6* (17), 2853–2855.
- (28) Abel, G. R.; Calabrese, Z. A.; Ayco, J.; Hein, J. E.; Ye, T. *Bioconjugate Chem* **2016**, *27* (3), 698–704.
- (29) Jewett, J. C.; Sletten, E. M.; Bertozzi, C. R. *J Am Chem Soc* **2010**, *132* (11), 3688–3690.
- (30) Agard, N. J.; Prescher, J. A.; Bertozzi, C. R. *J Am Chem Soc* **2004**, *126* (46), 15046–15047.
- (31) Lauer, M. H.; Vranken, C.; Deen, J.; Frederickx, W.; Vanderlinden, W.; Wand, N.; Leen, V.; Gehlen, M. H.; Hofkens, J.; Neely, R. K. *Chem Sci* **2017**, *8* (5), 3804–3811.
- (32) Dommerholt, J.; Rutjes, F. P. J. T.; van Delft, F. L. *Top Curr Chem (Cham)* **2016**, *374* (2), 16.
- (33) Ning, X.; Guo, J.; Wolfert, M. A.; Boons, G.-J. *Angew. Chem. Int. Ed. Engl.* **2008**, *47* (12), 2253–2255.
- (34) Jewett, J. C.; Bertozzi, C. R. *Chem Soc Rev* **2010**, *39* (4), 1272–1279.
- (35) Debets, M. F.; van Berkel, S. S.; Schoffelen, S.; Rutjes, F. P. J. T.; van Hest, J. C. M.; van Delft, F. L. *Chem. Commun. (Camb.)* **2010**, *46* (1), 97–99.
- (36) Lee, H.; Rho, J.; Messersmith, P. B. *Adv. Mater. Weinheim* **2009**, *21* (4), 431–.
- (37) Agten, S. M.; Dawson, P. E.; Hackeng, T. M. *J. Pept. Sci.* **2016**, *22* (5), 271–279.
- (38) Wolcott, A.; Gerion, D.; Visconte, M.; Sun, J.; Schwartzberg, A.; Chen, S. W.; Zhang, J. Z. *J Phys Chem B* **2006**, *110* (11), 5779–5789.
- (39) Marchan, V.; Ortega, S.; Pulido, D.; Pedroso, E.; Grandas, A. *Nucleic Acids Res.* **2006**, *34* (3), –e24.
- (40) Schoch, J.; Wiessler, M.; Jäschke, A. *J Am Chem Soc* **2010**, *132* (26), 8846–8847.
- (41) Kriukiene, E.; Labrie, V.; Khare, T.; Urbanaviciute, G.; Lapinaite, A.; Koncivicius, K.; Li, D.; Wang, T.; Pai, S.; Ptak, C.; Gordevicius, J.; Wang, S.-C.; Petronis, A.; Klimasauskas, S. *Nat Commun* **2013**, *4*.
- (42) Gong, H.; Holcomb, I.; Ooi, A.; Wang, X.; Majonis, D.; Unger, M. A.; Ramakrishnan, R. *Bioconjugate Chem* **2016**, *27* (1), 217–225.
- (43) Rozkiewicz, D. I.; Gierlich, J.; Burley, G. A.; Gutsmedl, K.; Carell, T.; Ravoo, B. J.; Reinhoudt, D. N. *ChemBiochem* **2007**, *8* (16), 1997–2002.
- (44) Banerjee, A.; Gazon, C.; Nadal, B.; Pons, T.; Krishnan, Y.; Dubertret, B. *Bioconjugate Chem* **2015**, *26* (8), 1582–1589.
- (45) Fischler, M.; Sologubenko, A.; Mayer, J.; Clever, G.; Burley, G.; Gierlich, J.; Carell, T.; Simon, U. *Chem. Commun. (Camb.)* **2008**, *16* (2), 169–171.
- (46) Matteucci, M. D.; Caruthers, M. H. *Synthesis of deoxyoligonucleotides on a polymer support. 1981.; 1992; Vol. 24, pp 92–98.*
- (47) Caruthers, M. H. *Acc. Chem. Res.* **2002**, *24* (9), 278–284.
- (48) Verma, S.; Eckstein, F. *Annu. Rev. Biochem.* **1998**, *67* (1), 99–134.

- (49) Slattum, P. S.; Loomis, A. G.; Machnik, K. J.; Watt, M.-A.; Duzeski, J. L.; Budker, V. G.; Wolff, J. A.; Hagstrom, J. E. *Mol Ther* **2003**, *8* (2), 255–263.
- (50) Rombouts, K.; Martens, T. F.; Zagato, E.; Demeester, J.; De Smedt, S. C.; Braeckmans, K.; Remaut, K. *Mol. Pharm.* **2014**, *11* (5), 1359–1368.
- (51) Daniel, S. G.; Westling, M. E.; Moss, M. S.; Kanagy, B. D. *BioTechniques* **1998**, *24* (3), 484–489.
- (52) Grosse, S.; Thévenot, G.; Aron, Y.; Duverger, E.; Abdelkarim, M.; Roche, A.-C.; Monsigny, M.; Fajac, I. *J Control Release* **2008**, *132* (2), 105–112.
- (53) Klepel, F.; Ravoo, B. J. *Org. Biomol. Chem.* **2017**, *15* (18), 3840–3842.
- (54) van Gijlswijk, R. P.; Talman, E. G.; Janssen, P. J.; Snoeijers, S. S.; Killian, J.; Tanke, H. J.; Heetebrij, R. J. *Expert Rev. Mol. Diagn.* **2001**, *1* (1), 81–91.
- (55) Alers, J. C.; Rochat, J.; Krijtenburg, P. J.; van Dekken, H.; Raap, A. K.; Rosenberg, C. *Genes Chromosomes Cancer* **1999**, *25* (3), 301–305.
- (56) Tsangaris, G. T.; Botsonis, A.; Politis, I.; Tzortzatou-Stathopoulou, F. *Toxicology* **2002**, *178* (2), 135–160.
- (57) Hohjoh, H.; Fukushima, T. *Gene* **2007**, *391* (1-2), 39–44.
- (58) Goodchild, J. *Bioconjugate Chem* **1990**, *1* (3), 165–187.
- (59) Trainor, G. L.; Jensen, M. A. *Nucleic Acids Res.* **1988**, *16* (24), 11846.
- (60) Kumar, A.; Tchen, P.; Roullet, F.; Cohen, J. *Anal. Biochem.* **1988**, *169* (2), 376–382.
- (61) Studier, F. W.; Moffatt, B. A. *J. Mol. Biol.* **1986**, *189* (1), 113–130.
- (62) Milligan, J. F.; Groebe, D. R.; Witherell, G. W.; Uhlenbeck, O. C. *Nucleic Acids Res.* **1987**, *15* (21), 8783–8798.
- (63) Zhu, Z.; Chao, J.; Yu, H.; Waggoner, A. S. *Nucleic Acids Res.* **1994**, *22* (16), 3418–3422.
- (64) Schowalter, D. B.; Sommer, S. S. *Anal. Biochem.* **1989**, *177* (1), 90–94.
- (65) Schuelke, M. *Nat. Biotechnol.* **2000**, *18* (2), 233–234.
- (66) Roychoudhury, R.; Jay, E.; Wu, R. *Nucleic Acids Res.* **1976**, *3* (4), 863–877.
- (67) Schaefer, B. C. *Anal. Biochem.* **1995**, *227* (2), 255–273.
- (68) Igloi, G. L.; Schieffmayr, E. *BioTechniques* **1993**, *15* (3), 486–8–490–2–494–7.
- (69) Gorczyca, W.; Gong, J.; Darzynkiewicz, Z. *Cancer Res.* **1993**, *53* (8), 1945–1951.
- (70) Cosstick, R.; McLaughlin, L. W.; Eckstein, F. *Nucleic Acids Res.* **1984**, *12* (4), 1791–1810.
- (71) Barrio, J. R.; Barrio, M. C.; Leonard, N. J.; England, T. E.; Uhlenbeck, O. C. *Biochemistry* **1978**, *17* (11), 2077–2081.
- (72) Phillips, D. H. *Mutat. Res.* **1997**, *378* (1-2), 1–12.
- (73) Oshevski, S. I. *FEBS Lett.* **1982**, *143* (1), 119–123.
- (74) Rigby, P. W.; Dieckmann, M.; Rhodes, C.; Berg, P. *J. Mol. Biol.* **1977**, *113* (1), 237–251.
- (75) Werner, D.; Rest, R. *Biochem Bioph Res Co* **1987**, *147* (1), 340–345.
- (76) Kato, A.; Albert, P. S.; Vega, J. M.; Birchler, J. A. *Biotech Histochem* **2006**, *81* (2-3), 71–78.

- (77) Schubert, H. L.; Blumenthal, R. M.; Cheng, X. *Trends Biochem. Sci.* **2003**, 28 (6), 329–335.
- (78) Siedlecki, P.; Zielenkiewicz, P. *Acta Biochim. Pol.* **2006**, 53 (2), 245–256.
- (79) Esteller, M.; Corn, P. G.; Baylin, S. B.; Herman, J. G. *Cancer Res.* **2001**, 61 (8), 3225–3229.
- (80) Herman, J. G.; Baylin, S. B. *N. Engl. J. Med.* **2003**, 349 (21), 2042–2054.
- (81) Rao, S. T.; Rossmann, M. G. *J. Mol. Biol.* **1973**, 76 (2), 241–256.
- (82) Katz, J. E.; Dlakić, M.; Clarke, S. *Mol. Cell Proteomics* **2003**, 2 (8), 525–540.
- (83) Goedecke, K.; Pignot, M.; Goody, R. S.; Scheidig, A. J.; Weinhold, E. *Nat. Struct. Biol.* **2001**, 8 (2), 121–125.
- (84) Pignot, M.; Siethoff, C.; Linscheid, M.; Weinhold, E. *Angew. Chem. Int. Ed. Engl.* **1998**, 37 (20), 2888–2891.
- (85) Pljevaljcic, G.; Schmidt, F.; Weinhold, E. *Chembiochem* **2004**, 5 (3), 265–269.
- (86) Weller, R. L.; Rajski, S. R. *Org. Lett.* **2005**, 7 (11), 2141–2144.
- (87) Comstock, L. R.; Rajski, S. R. *Nucleic Acids Res.* **2005**, 33 (5), 1644–1652.
- (88) Pljevaljcic, G.; Pignot, M.; Weinhold, E. *J Am Chem Soc* **2003**, 125 (12), 3486–3492.
- (89) Osborne, T.; Roska, R. L. W.; Rajski, S. R.; Thompson, P. R. *J Am Chem Soc* **2008**, 130 (14), 4574–4575.
- (90) Hymbaugh Bergman, S. J.; Comstock, L. R. *Bioorgan Med Chem* **2015**, 23 (15), 5050–5055.
- (91) Dalhoff, C.; Lukinavicius, G.; Klimasauskas, S.; Weinhold, E. *Nat. Chem. Biol.* **2006**, 2 (1), 31–32.
- (92) Dalhoff, C.; Lukinavicius, G.; Klimasauakas, S.; Weinhold, E. *Nat Protoc* **2006**, 1 (4), 1879–1886.
- (93) Deen, J.; Vranken, C.; Leen, V.; Neely, R. K.; Janssen, K. P. F.; Hofkens, J. *Angew. Chem. Int. Ed. Engl.* **2016**, 56 (19), 5182–5200.
- (94) Lukinavicius, G.; Tomkuvienė, M.; Masevicius, V.; Klimasauskas, S. *ACS Chem. Biol.* **2013**, 8 (6), 1134–1139.
- (95) Thomsen, M.; Vogensen, S. B.; Buchardt, J.; Burkart, M. D.; Clausen, R. P. *Org. Biomol. Chem.* **2013**, 11 (43), 7606–7610.
- (96) Lipson, J. M.; Thomsen, M.; Moore, B. S.; Clausen, R. P.; La Clair, J. J.; Burkart, M. D. *Chembiochem* **2013**, 14 (8), 950–953.
- (97) Huber, T. D.; Wang, F.; Singh, S.; Johnson, B. R.; Zhang, J.; Sunkara, M.; Van Lanen, S. G.; Morris, A. J.; Phillips, G. N., Jr.; Thorson, J. S. *ACS Chem. Biol.* **2016**, acschembio.6b00348–acschembio.6b00348.
- (98) Iwig, D. F.; Booker, S. J. *Biochemistry* **2004**, 43 (42), 13496–13509.
- (99) Zhang, J.; Zheng, Y. G. *ACS Chem. Biol.* **2016**, 11 (3), 583–597.
- (100) Neely, R. K.; Deen, J.; Hofkens, J. *Biopolymers* **2011**, 95 (5), 298–311.
- (101) Anhäuser, L.; Muttach, F.; Rentmeister, A. *Chem. Commun. (Camb.)* **2018**, 54 (5), 449–451.
- (102) Muttach, F.; Muthmann, N.; Reichert, D.; Anhäuser, L.; Rentmeister, A. *Chem Sci* **2017**, 8 (12), 7947–7953.
- (103) Motorin, Y.; Burhenne, J.; Teimer, R.; Koynov, K.; Willnow, S.; Weinhold, E.; Helm, M. *Nucleic Acids Res.* **2011**, 39 (5), 1943–1952.

- (104) Artyukhin, A. B.; Woo, Y.-H. *Anal. Biochem.* **2012**, *425* (2), 169–174.
- (105) Peters, W.; Willnow, S.; Duisken, M.; Kleine, H.; Macherey, T.; Duncan, K. E.; Litchfield, D. W.; Lüscher, B.; Weinhold, E. *Angew. Chem. Int. Ed. Engl.* **2010**, *49* (30), 5170–5173.
- (106) Wang, R.; Zheng, W.; Yu, H.; Deng, H.; Luo, M. *J Am Chem Soc* **2011**, *133* (20), 7648–7651.
- (107) Singh, S.; Zhang, J.; Huber, T. D.; Sunkara, M.; Hurley, K.; Goff, R. D.; Wang, G.; Zhang, W.; Liu, C.; Rohr, J.; Van Lanen, S. G.; Morris, A. J.; Thorson, J. S. *Angew. Chem. Int. Ed. Engl.* **2014**, *53* (15), 3965–3969.
- (108) Bothwell, I. R.; Islam, K.; Chen, Y.; Zheng, W.; Blum, G.; Deng, H.; Luo, M. *J Am Chem Soc* **2012**, *134* (36), 14905–14912.
- (109) Lee, B. W. K.; Sun, H. G.; Zang, T.; Kim, B. J.; Alfaro, J. F.; Zhou, Z. S. *J Am Chem Soc* **2010**, *132* (11), 3642–3643.
- (110) Grunwald, A.; Dahan, M.; Giesbertz, A.; Nilsson, A.; Nyberg, L. K.; Weinhold, E.; Ambjornsson, T.; Westerlund, F.; Ebenstein, Y. *Nucleic Acids Res.* **2015**, *43* (18), –e117.
- (111) Wojciechowski, M.; Czapinska, H.; Bochtler, M. *P Natl Acad Sci USA* **2013**, *110* (1), 105–110.
- (112) Rombouts, K.; Braeckmans, K.; Remaut, K. *Bioconjugate Chem* **2016**, *27* (2), 280–297.
- (113) Neely, R. K.; Dedecker, P.; Hotta, J.-I.; Urbanavičiūtė, G.; Klimasauskas, S.; Hofkens, J. *Chem Sci* **2010**, *1* (4), 453–459.
- (114) Feinberg, A. P.; Tycko, B. *Nat. Rev. Cancer* **2004**, *4* (2), 143–153.
- (115) Pastor, W. A.; Aravind, L.; Rao, A. *Nat. Rev. Mol. Cell Biol.* **2013**, *14* (6), 341–356.
- (116) Schumacher, A.; Kapranov, P.; Kaminsky, Z.; Flanagan, J.; Assadzadeh, A.; Yau, P.; Virtanen, C.; Winegarden, N.; Cheng, J.; Gingeras, T.; Petronis, A. *Nucleic Acids Res.* **2006**, *34* (2), 528–542.
- (117) Weber, M.; Davies, J. J.; Wittig, D.; Oakeley, E. J.; Haase, M.; Lam, W. L.; Schübeler, D. *Nature Genetics* **2005**, *37* (8), 853–862.
- (118) Bock, C.; Walter, J.; Paulsen, M.; Lengauer, T. *Nucleic Acids Res.* **2008**, *36* (10), e55–e55.
- (119) Harris, R. A.; Wang, T.; Coarfa, C.; Nagarajan, R. P.; Hong, C.; Downey, S. L.; Johnson, B. E.; Fouse, S. D.; Delaney, A.; Zhao, Y.; Olshen, A.; Ballinger, T.; Zhou, X.; Forsberg, K. J.; Gu, J.; Echipare, L.; O'Geen, H.; Lister, R.; Pelizzola, M.; Xi, Y.; Epstein, C. B.; Bernstein, B. E.; Hawkins, R. D.; Ren, B.; Chung, W. Y.; Gu, H.; Bock, C.; Gnirke, A.; Zhang, M. Q.; Haussler, D.; Ecker, J. R.; Li, W.; Farnham, P. J.; Waterland, R. A.; Meissner, A.; Marra, M. A.; Hirst, M.; Milosavljevic, A.; Costello, J. F. *Nat. Biotechnol.* **2010**, *28* (10), 1097–1105.
- (120) Lister, R.; Pelizzola, M.; Downen, R. H.; Hawkins, R. D.; Hon, G.; Tonti-Filippini, J.; Nery, J. R.; Lee, L.; Ye, Z.; Ngo, Q.-M.; Edsall, L.; Antosiewicz-Bourget, J.; Stewart, R.; Ruotti, V.; Millar, A. H.; Thomson, J. A.; Ren, B.; Ecker, J. R. *Nature* **2009**, *462* (7271), 315–322.
- (121) Suzuki, M. M.; Bird, A. *Nat. Rev. Genet.* **2008**, *9* (6), 465–476.
- (122) Robinson, M. D.; Stirzaker, C.; Statham, A. L.; Coolen, M. W.; Song, J. Z.; Nair, S. S.; Strbenac, D.; Speed, T. P.; Clark, S. J. *Genome Res.* **2010**, *20* (12), 1719–1729.

- (123) Nair, S. S.; Coolen, M. W.; Stirzaker, C.; Song, J. Z.; Statham, A. L.; Strbenac, D.; Robinson, M. D.; Clark, S. J. *Epigenetics* **2011**, 6 (1), 34–44.
- (124) Cantoni, G. L. *J Am Chem Soc* **1952**, 74 (11), 2942–2943.
- (125) Lukinavicius, G.; Klimasauskas, S.; Dalhoff, C.; Weinhold, E. *Febs Journal* **2006**, 273, 336–337.
- (126) Jin, Y.; Yu, C.; Denman, R. J.; Zhang, W. *Chem Soc Rev* **2013**, 42 (16), 6634–6654.
- (127) Nishikaze, T.; Kawabata, S.-I.; Iwamoto, S.; Tanaka, K. *Analyst* **2013**, 138 (23), 7224–7232.
- (128) Gauthier, M. A.; Gibson, M. I.; Klok, H.-A. *Angew. Chem. Int. Ed. Engl.* **2009**, 48 (1), 48–58.
- (129) Kalia, J.; Raines, R. T. *Angew. Chem. Int. Ed. Engl.* **2008**, 47 (39), 7523–7526.
- (130) Furukawa, J. I.; Shinohara, Y.; Kuramoto, H.; Miura, Y.; Shimaoka, H.; Kurogochi, M.; Nakano, M.; Nishimura, S. I. *Anal. Chem.* **2008**, 80 (4), 1094–1101.
- (131) Lukinavicius, G.; Lapiene, V.; Stasevskij, Z.; Dalhoff, C.; Weinhold, E.; Klimasauskas, S. *J Am Chem Soc* **2007**, 129 (10), 2758–2759.
- (132) Appel, R. *Angewandte Chemie International Edition in English* **1975**, 14 (12), 801–811.
- (133) Coombs, J. R.; Zhang, L.; Morken, J. P. *J Am Chem Soc* **2014**, 136 (46), 16140–16143.
- (134) Belyk, K.; Rozema, M. J.; Knochel, P. *Journal of Organic Chemistry* **1992**, 57 (15), 4070–4074.
- (135) Seyferth, D.; Welch, D. E.; Heeren, J. K. *J Am Chem Soc* **1963**, 85 (5), 642–643.
- (136) Karmakar, R.; Yun, S. Y.; Chen, J.; Xia, Y.; Lee, D. *Angew. Chem. Int. Ed. Engl.* **2015**, 54 (22), 6582–6586.
- (137) Wei, H. X.; Truitt, C. L.; Pare, P. W. *Tetrahedron Letters* **2003**, 44 (4), 831–833.
- (138) Vázquez-Romero, A.; Verdaguer, X.; Riera, A. *European Journal of Organic Chemistry* **2013**, 2013 (9), 1716–1725.
- (139) Quinton, J.; Kolodych, S.; Chaumonet, M.; Bevilacqua, V.; Nevers, M.-C.; Volland, H.; Gabillet, S.; Thuery, P.; Creminon, C.; Taran, F. *Angew. Chem. Int. Ed. Engl.* **2012**, 51 (25), 6144–6148.
- (140) Cossy, J.; Pete, J. P. *Tetrahedron Letters* **1986**, 27 (5), 573–574.
- (141) Harrison, K. A.; Davies, S. S.; Marathe, G. K.; McIntyre, T.; Prescott, S.; Reddy, K. M.; Falck, J. R.; Murphy, R. C. *Journal of Mass Spectrometry* **2000**, 35 (2), 224–236.
- (142) Fishman, J. M.; Kiessling, L. L. *Angew. Chem. Int. Ed. Engl.* **2013**, 52 (19), 5061–5064.
- (143) Priegue, J. M.; Crisan, D. N.; Martinez-Costas, J.; Granja, J. R.; Fernandez-Trillo, F.; Montenegro, J. *Angew. Chem. Int. Ed. Engl.* **2016**, 55 (26), 7492–7495.
- (144) Watson, A. J. A.; Williams, J. M. J. *Science* **2010**, 329 (5992), 635–636.
- (145) Pignot, M.; Pljevaljic, G.; Weinhold, E. *European Journal of Organic Chemistry* **2000**, 2000 (3), 549–555.

- (146) MITSUNOBU, O.; YOSHIDA, N. *Tetrahedron Letters* **1981**, 22 (24), 2295–2296.
- (147) **2018**, 1–68.
- (148) Staudinger, H.; Meyer, J. *Helvetica Chimica Acta* **1919**, 2 (1), 635–646.
- (149) Hoffman, J. L. *Biochemistry* **1986**, 25 (15), 4444–4449.
- (150) Stecher, H.; Teng, M.; Ueberbacher, B. J.; Remler, P.; Schwab, H.; Griengl, H.; Gruber-Khadjawi, M. *Angew. Chem. Int. Ed. Engl.* **2009**, 48 (50), 9546–9548.
- (151) Islam, K.; Bothwell, I.; Chen, Y.; Sengelaub, C.; Wang, R.; Deng, H.; Luo, M. *J Am Chem Soc* **2012**, 134 (13), 5909–5915.
- (152) Bentley, R. *Chem Soc Rev* **2005**, 34 (7), 609–624.
- (153) Schluckebier, G.; Kozak, M.; Bleimling, N.; Weinhold, E.; Saenger, W. *J. Mol. Biol.* **1997**, 265 (1), 56–67.
- (154) Lenz, T.; Bonnist, E. Y. M.; Pljevaljcic, G.; Neely, R. K.; Dryden, D. T. F.; Scheidig, A. J.; Jones, A. C.; Weinhold, E. *J Am Chem Soc* **2007**, 129 (19), 6240–6248.
- (155) Luchowski, R.; Matveeva, E. G.; Gryczynski, I.; Terpetschnig, E. A.; Patsenker, L.; Laczko, G.; Borejdo, J.; Gryczynski, Z. *Curr Pharm Biotechnol* **2008**, 9 (5), 411–420.
- (156) Lakowicz, J. R.; Malicka, J.; D'Auria, S.; Gryczynski, I. *Anal. Biochem.* **2003**, 320 (1), 13–20.
- (157) Yang, S. J.; Zhang, H. *Anal. Chem.* **2012**, 84 (5), 2232–2238.
- (158) Liu, D.; Zou, X.; Zhong, L.; Lou, Y.; Bin Yang; Yin, Y. *International Journal of Mass Spectrometry* **2014**, 374, 20–25.
- (159) Carey, F. A.; Sundberg, R. J. *Advanced Organic Chemistry*, 5 ed.; 2008; Vol. A, pp 650–651.
- (160) Dahlgren, M. K.; Zetterström, C. E.; Gylfe, Å.; Linusson, A.; Elofsson, M. *Bioorgan Med Chem* **2010**, 18 (7), 2686–2703.
- (161) Song, C.-X.; Szulwach, K. E.; Fu, Y.; Dai, Q.; Yi, C.; Li, X.; Li, Y.; Chen, C.-H.; Zhang, W.; Jian, X.; Wang, J.; Zhang, L.; Looney, T. J.; Zhang, B.; Godley, L. A.; Hicks, L. M.; Lahn, B. T.; Jin, P.; He, C. *Nat. Biotechnol.* **2010**, 29 (1), 68–72.
- (162) Guibert, S.; Weber, M. *Curr. Top. Dev. Biol.* **2013**, 104, 47–83.
- (163) Hirsch, J. D.; Eslamizar, L.; Filanoski, B. J.; Malekzadeh, N.; Haugland, R. P.; Beechem, J. M.; Haugland, R. P. *Anal. Biochem.* **2002**, 308 (2), 343–357.
- (164) Vu, T. H.; Li, T.; Nguyen, D.; Nguyen, B. T.; Yao, X. M.; Hu, J. F.; Hoffman, A. R. *Genomics* **2000**, 64 (2), 132–143.