Selective Chemical Reactions for Nucleic Acid Sequencing and DNA-Encoded Library Synthesis

Yasaman Mahdavi-Amiri

A DISSERTATION SUBMITTED TO THE FACULTY OF GRADUATE STUDIES
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

GRADUATE PROGRAM IN CHEMISTRY
YORK UNIVERSITY
TORONTO, ONTARIO
December 2022

Abstract

Section One: The ability to map methylation sites in the human genome and epitranscriptome has transformed our understanding of how these modifications govern and influence a host of cellular processes and human diseases. Amongst the most widely studied methylations is N^6-methyladenine, known as 6mA in DNA and m^6A in RNA. While traditional methods to sequence these modifications have depended on antibody pulldowns, chemistry-based approaches are often less sequence dependent, can work on either DNA or RNA, and thus can provide a robust, inexpensive, and universal sequencing approach. In part one of this thesis, the first chemistry-based single-nucleotide resolution sequencing method for the detection of N^6-adenine methylation sites in DNA and RNA is presented. This method takes advantage of the chemoselective deamination of unmodified adenines under acidic nitrite conditions, resulting in a (d)A to (d)G transition, while leaving methylated adenine sites unaffected. As changes in N^6-adenine methylation of RNA and DNA have been implicated in a range of human diseases, especially cancers, the method has been rapidly adopted by researchers globally as an affordable and straightforward sequencing approach to assist in understanding the role and impact of the epigenome and epitranscriptome on human health. The ability of this method to detect other nucleotide modifications was also evaluated and described.

Section Two: DNA-encoded libraries (DELs) comprise millions to billions of small-molecules covalently linked to a unique DNA barcode that can be read using standard next-generation sequencing (NGS). This technology has revolutionized the field of drug discovery as a method to rapidly identify small molecules that can serve as novel leads

for drug development. The success of a drug discovery campaign involving a DEL depends on the chemical diversity presented within the DEL; methods that can generate DELs with new molecular architectures and with greater chemical diversity are critically needed to advance drug discovery efforts both within industry and academia. To this end, the use of photoredox chemistry as a facile method to generate DELs with drug-like properties is presented as part two of this thesis. An efficient approach for the photoredox-catalysed hydroaminoalkylation between on-DNA secondary *N*-substituted (hetero)arylamines and vinylarenes is explored as a method to generate DELs with known bioactive architectures. The developed reaction proceeds efficiently with a broad and well-explored substrate scope, working best with electron poor to neutral vinylarenes. This method is well suited for the construction of DELs enabling an expansion of drug-like chemical space.

To my parents, Nezam and Azita,
whose love and support made all this possible

Acknowledgment

First, and most importantly, I would like to thank my supervisor Professor Ryan Hili for providing me with this incredible opportunity. Thank you for your patience, guidance, and support at every step of this journey. Working under your supervision has been very enjoyable. You have been a great mentor.

I would also like to thank my committee members, Professor Sergey Krylov and Professor Philip Johnson. I really appreciate your valuable advice and feedback throughout my PhD studies.

I am grateful to Nayanthara Asok and the Baumgartner Lab for providing access to their glovebox and for various forms of assistance they have provided over the past several years.

I also wish to thank my partner, Omid, for being supportive throughout the good and bad times. Thanks for being so understanding and for being such a great friend.

A big thanks also goes out to my amazing friends, Nayanthara and An. For all the fun we have had, and the moral support you have given me.

I must also thank my wonderful labmates, Morgan, Kimberley, Matina, Natalie, Molly, Ian, Emily, Alex, Prakriti, Karanveer, Tristan, Nicole, Areeba and Hetvi. It has been an amazing experience working with you. Thank you all for your help, support and friendship.

Finally, my deepest gratitude goes to my family, Zari, Nezam, Azita, Anita, Lili and Maziar. Thank you for always standing by me. Your belief in me has kept my motivation high during this process.

Table of Contents

List of Figures

List of Schemes

List of Tables

Section 1

Chapter 1. Introduction to Modified Nucleic Acids

1.1. Nucleic acids

Nucleic acids are biopolymers essential to all known forms of life. The two main extant classes of nucleic acids are deoxyribonucleic acid (DNA) and ribonucleic acid (RNA). The genetic and hereditary information of eukaryotic organisms is encoded in a genome comprising molecules of DNA. Genetic information flows from DNA to RNA, which is then expressed as proteins.

1.1.1. Basic structure of nucleic acids

These naturally occurring biopolymers are made of units called nucleotides which are made of three components: a pentose (5-carbon) sugar, a phosphate group, and a nitrogenous base. There are five canonical nitrogenous bases: purines (adenine (A) and guanine (G)) and pyrimidines (cytosine (C), thymine (T) and uracil (U)). All nucleic acids contain the bases A, C, and G; However, T is only found in DNA, and U is only found in RNA. The sugar is 2′-deoxyribose in DNA and ribose in RNA. The phosphate group connects sugar residues by bridging the 5′-hydroxyl group on one sugar to the 3′-hydroxyl group of the next sugar in the chain (Figure 1-1). Without an attached phosphate group, the sugar attached to one of the bases is together known as a nucleoside.

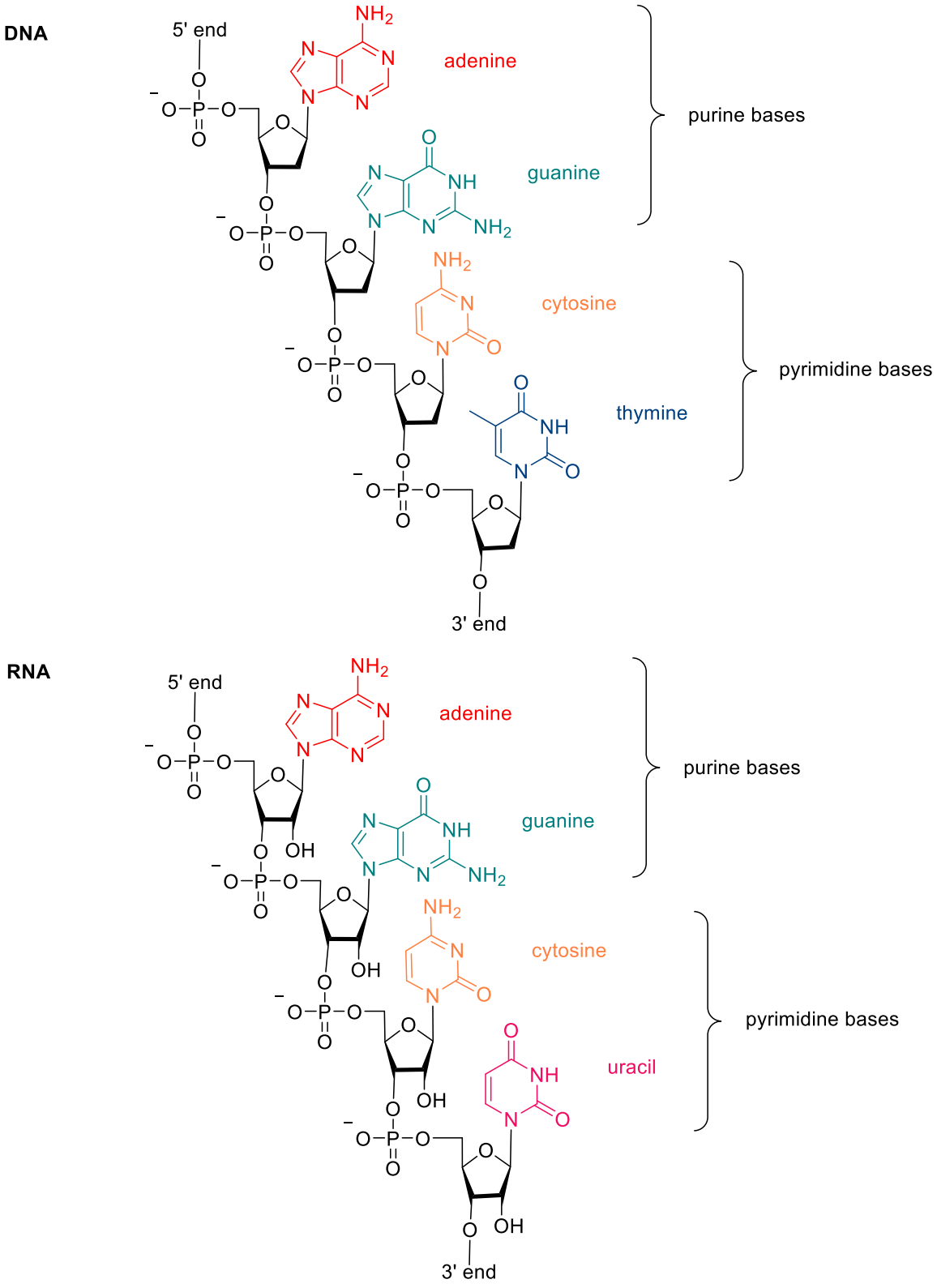


Figure 1-1: Portion of polynucleotide chain of DNA and RNA.

1.1.2. Deoxyribonucleic acid (DNA)

DNA was discovered by Friedrich Miescher in 1869 at the University of Tübingen, Germany. He named it “nuclein” at the time, since he had isolated it from cellular nuclei[1]. In 1944 Oswald T. Avery, Colin MacLeod, and Maclyn McCarty published their landmark paper suggesting that DNA was the carrier of genetic information[2]. Erwin Chargaff later discovered that within each species the DNA are always present in fixed ratios: the same number of adenine as thymine bases and the same number of cytosine as guanine bases[3,4]. In 1952, Raymond Gosling who was a graduate student working under the supervision of Rosalind Franklin, took an X-ray diffraction image of DNA, named as "Photo 51" (Figure 1-2)[5]. The image was tagged "photo 51" because it was the 51st diffraction photograph that Franklin and Gosling had taken.

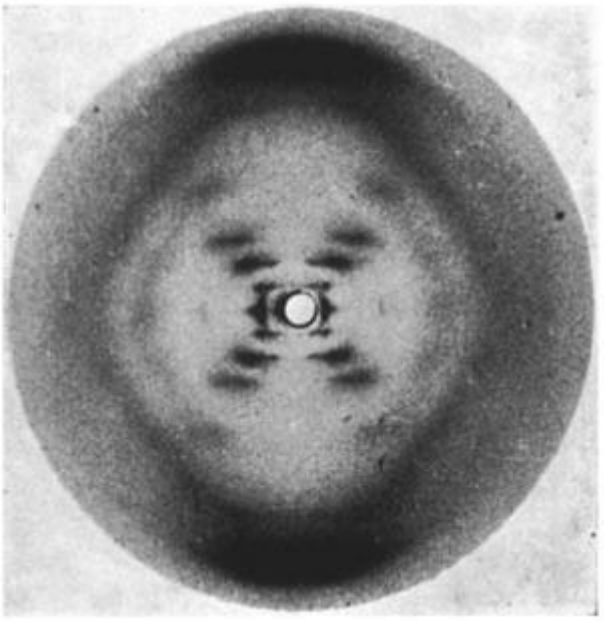


Figure 1-2: Photo 51, showing X-ray diffraction pattern of DNA[5].

Along with Francis Crick, James Watson used characteristics and features of Photo 51, together with Erwin Chargaff's observations, to develop the chemical model of the DNA molecule[6], for which they won the Nobel Prize in Physiology or Medicine in 1962. They suggested that DNA consists of two strands of polynucleotides coiled around each other forming a double helix. The two strands are connected through hydrogen bonds. T bases are paired with A's, and C's are paired with G's. The two strands run in antiparallel directions with the sugar-phosphate chains running along the outside of the helix and the nucleobases on the inside, where they engage in hydrogen bonding to complementary bases on the opposing strand (Figure 1-3).

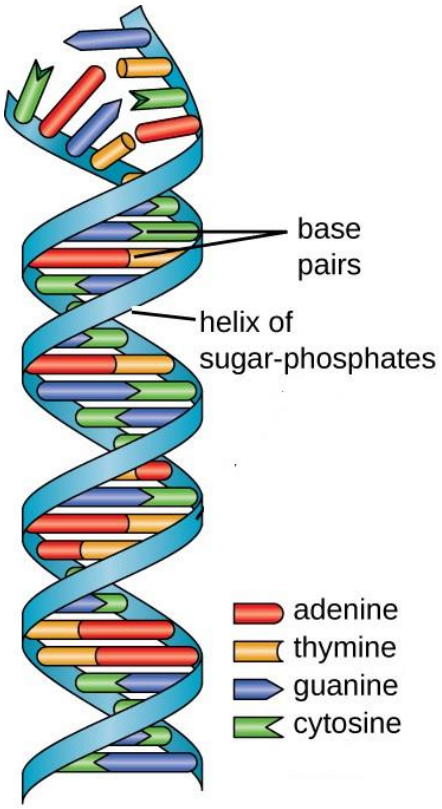


Figure 1-3: The double helical structure of DNA.
Figure from: Parker et al. Microbiology (2019) from Openstax

DNA molecules are amongst the largest naturally occurring biopolymers known. They are packaged into thread-like structures called chromosomes which are stored in the nucleus of each cell and make up an organism's genome. The human genome has approximately 3 billion base pairs of DNA arranged into 46 chromosomes[7]. DNA provides the genetic information for the cell and is called the blueprint of life since it contains the instructions to construct other components of the cell, such as proteins and RNA molecules.

1.1.3. Ribonucleic acid (RNA)

The complete nucleotide sequence of an RNA isolated from yeast, was found by Robert W. Holley in 1965 which was the first nucleic acid for which the structure is known[8]. The structure of RNA is very similar to that of DNA; however, RNA is usually single-stranded, and thus nucleobases pair with other bases within the same molecule, leading to complex three-dimensional structures (Figure 1-4).

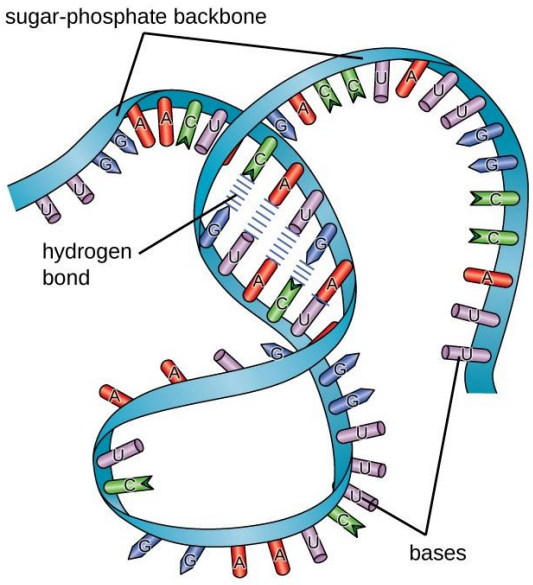


Figure 1-4: Example of an RNA structure.
Figure from: Parker et al. Microbiology (2019) from Openstax

RNA is made by RNA polymerase catalyzed DNA transcription, which copies the base sequence of a section of double-stranded DNA into a single-stranded RNA. While there are various forms of RNA in nature, each with unique biological roles, the three major classes include messenger RNA (mRNA), transfer RNA (tRNA), and ribosomal RNA (rRNA). These major classes are integral to the translation of proteins from RNA[9].

Examples of different types of RNA and their roles are listed in Table 1-1.

Table 1-1: Different types of RNA and their functions

Type of RNA	**Function**
Messenger RNA (mRNA)	Involved in translation: carries the genetic information copied from DNA in the form of a series of codons each of which specifies a particular amino acid
Transfer RNA (tRNA)	Involved in translation: brings specific amino acids to the ribosome, which are then matched up to the mRNA blueprints
Ribosomal RNA (rRNA)	Involved in translation: serves critical roles in forming the catalytic sites of translation of mRNA
Signal recognition particle RNA (SRP RNA)	Involved in translocation: is a part of the signal recognition particle (SRP) ribonucleoprotein complex which recognizes the signal peptide and binds to the ribosome

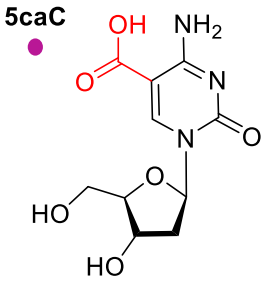
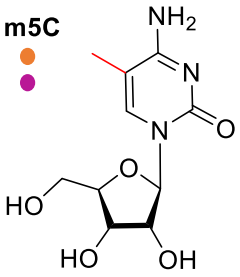
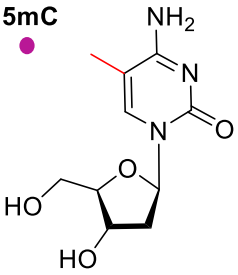
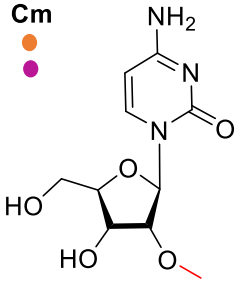
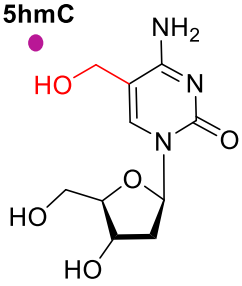
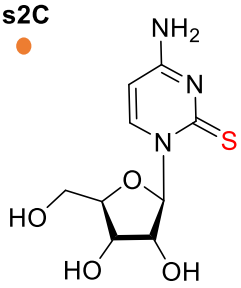
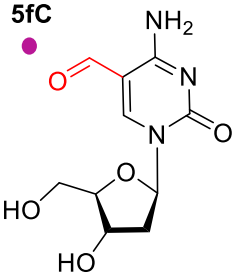
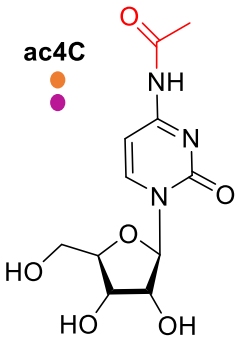
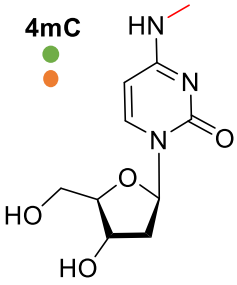
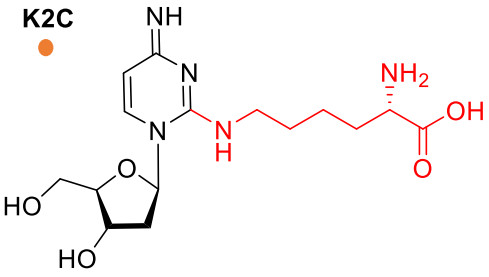
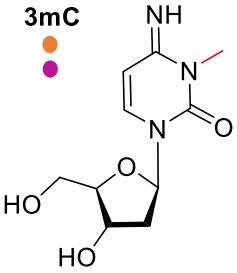
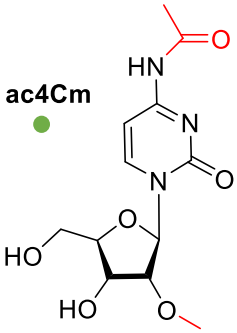
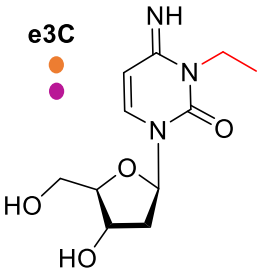
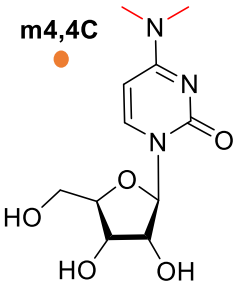
Small nuclear RNA (snRNA)	Form part of the splicing mechanism that is involved in the processing of pre-messenger RNA (pre-mRNA) into mature mRNA
Heterogeneous nuclear RNA (hnRNA)	Are large pre-mRNAs of various nucleotide sequences that are processed in the nucleus to become cytoplasmic mRNAs
Small nucleolar RNA (snoRNA)	Guide chemical modifications of other RNAs, mainly ribosomal RNAs, transfer RNAs and small nuclear RNAs

1.2. Epigenetic and epitranscriptomic markers

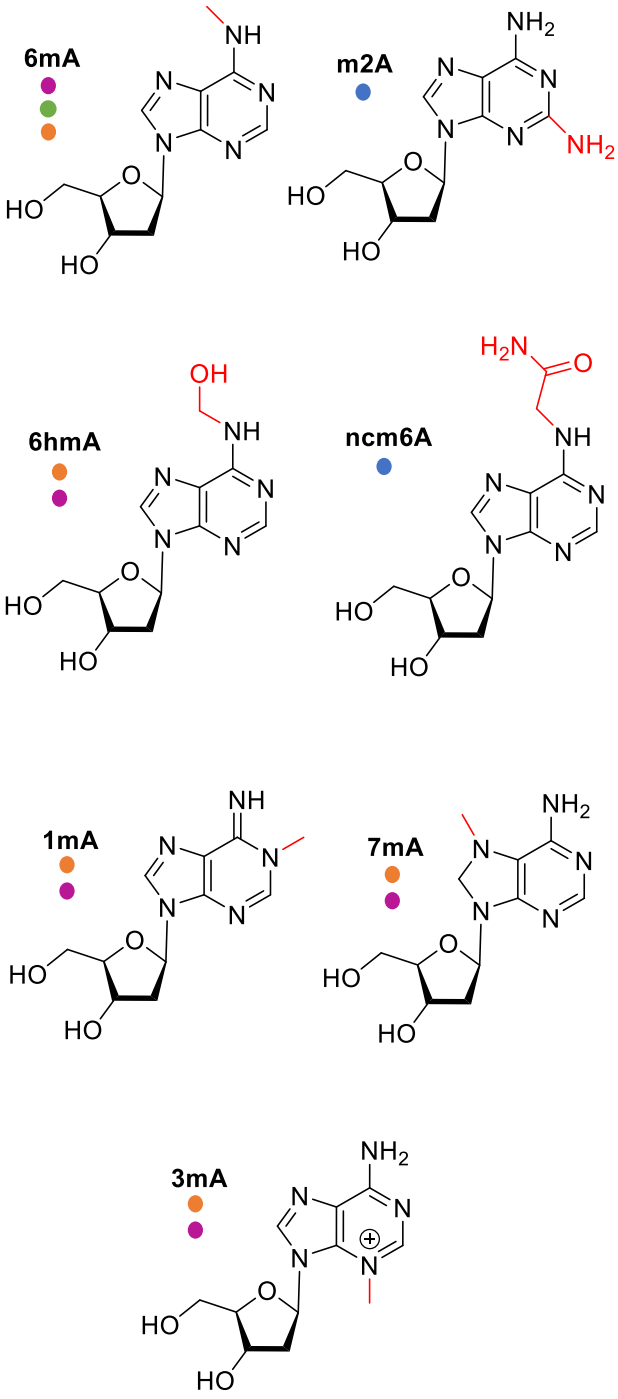
Since the discovery of (deoxy)adenosine (dA, A), (deoxy)cytidine (dC, C), (deoxy)guanosine (dG, G), deoxythymidine (dT) and uridine (U) as the information carrying building blocks which form the basis for RNA and DNA, various modifications of these nucleosides have been discovered. Beyond the genetic code, there is another hidden layer of complexity, usually mediated by these chemical modifications[10]. Recently, there has been an increased interest in understanding DNA and RNA modifications and their importance in human health and disease. To date, dozens of DNA and RNA modifications have been characterized[11,12], some examples of which are illustrated in Table 1-2.

Table 1-2: Examples of modifications found in DNA and RNA

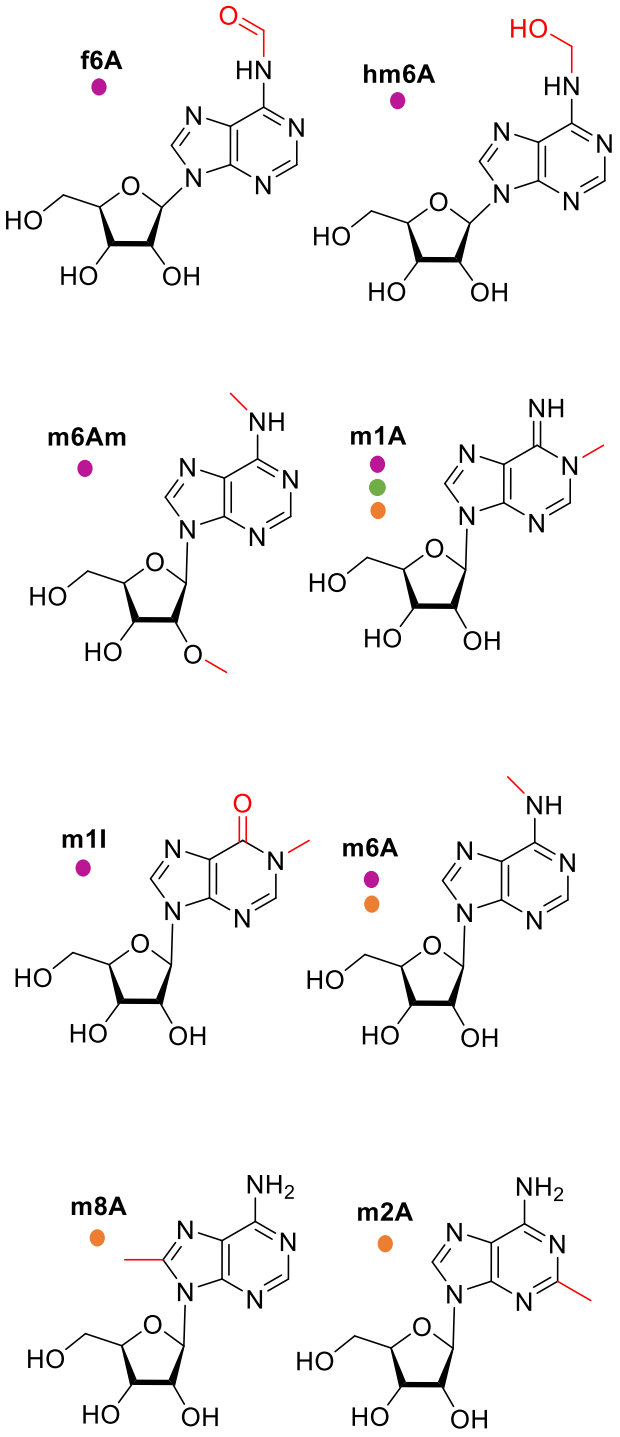
DNA modifications from: DNAmod database[11], RNA modifications from: Modomics database[12]

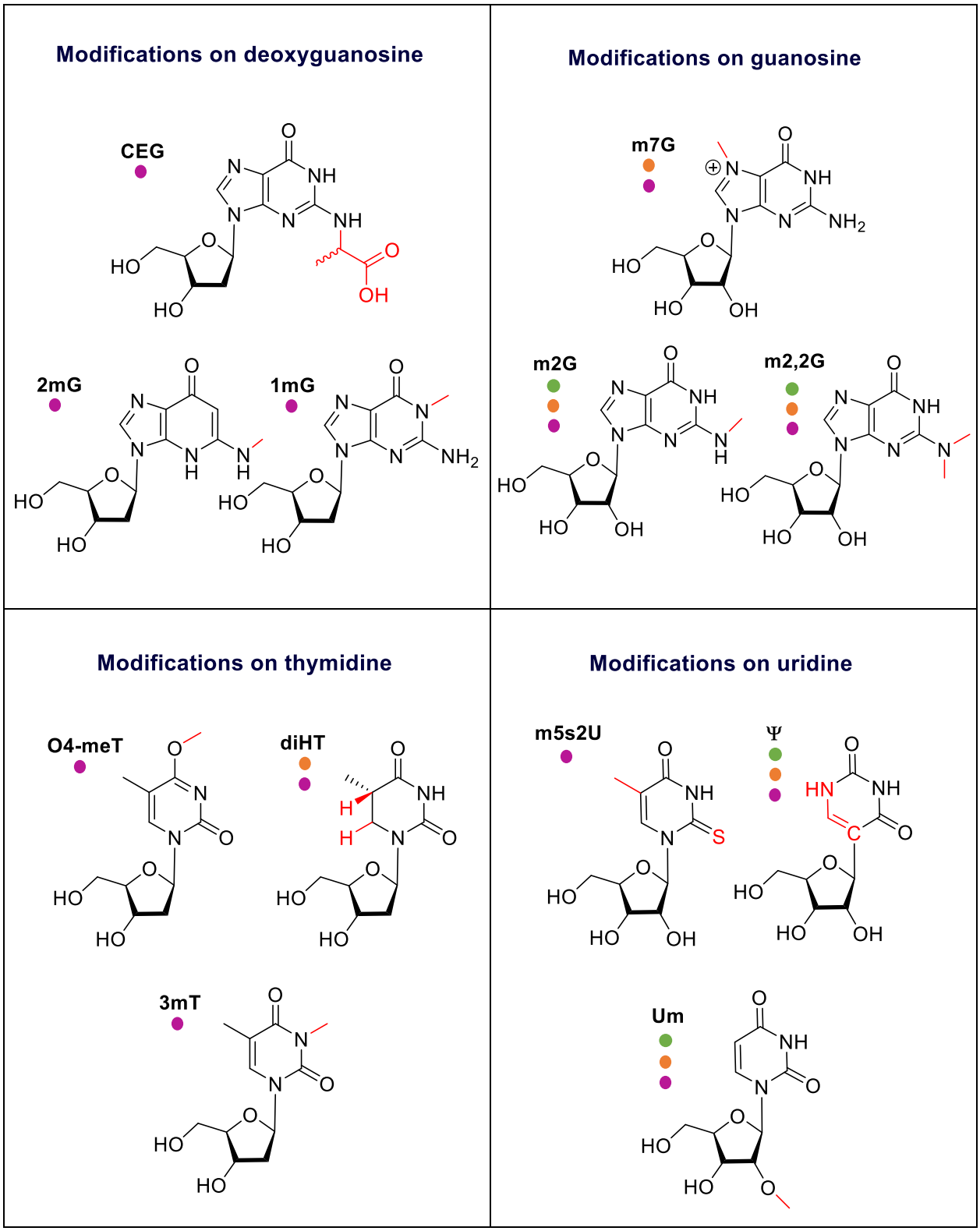
DNA modifications	RNA modifications
Modifications on deoxycytidine	**Modifications on cytidine**
5caC 	**m5C** 
5mC 	**Cm** 
5hmC 	**s2C** 
5fC 	**ac4C** 
4mC 	**K2C** 
3mC 	**ac4Cm** 
e3C 	**m4,4C** 

Modifications on deoxyadenosine



Modifications on adenosine




Modifications on deoxyguanosine
CEG
2mG
1mG
Modifications on guanosine
m7G
m2G
m2,2G
Modifications on thymidine
O4-meT
diHT
3mT
Modifications on uridine
m5s2U
Ψ
Um
Modification found in phylogeny: Archaea Eubacteria Eukaryota Viruses

1.2.1. DNA modifications

Each of the four A, C, G and T bases may have modifications, among which 5-methylcytosine (5mC) is the most abundant one in mammals and therefore is also known as the fifth base of DNA. About 4% of the cytosine residues in the human genome have been found to be methylated[13]. The existence of cytosine methylation in genomic DNA was first reported by Wyatt in 1951[14]. In mammals it is involved in various biological processes including the silencing of transposable elements, regulation of gene expression, genomic imprinting, and X-chromosome inactivation[15]. The enzymes responsible for 5mC deposition and removal have been well defined in mammals, including DNMT1/3A/3B proteins as methyltransferases also called “writers”, and TET1/2/3 as demethylases also called “erasers”[15]. Other modifications such as 5-hydroxymethylcytosine (5hmC), 5-formyl cytosine (5fC) and 5-carboxyl cytosine (5caC) have also been identified, though much less abundant[16]. Another kind of DNA modification, N^6-methyladenine (6mA) has been widely known as a DNA modification in bacteria[17,18]. Its presence in eukaryotes has only been recently established, including in humans where it represents ~0.051% of the genome[19]. 6mA is thought to play an epigenetic role in embryonic development[20], tumorigenesis[19], response to stress, neuropsychiatric disorders[21], and embryonic stem cell function[22].

1.2.2. RNA modifications

In addition to the canonical A, C, G, and U residues, modified nucleotides were discovered in abundant cellular RNAs as early as 1960[23]. >170 types of RNA modification are known to date, however, the functions of most of these modifications are largely unknown.

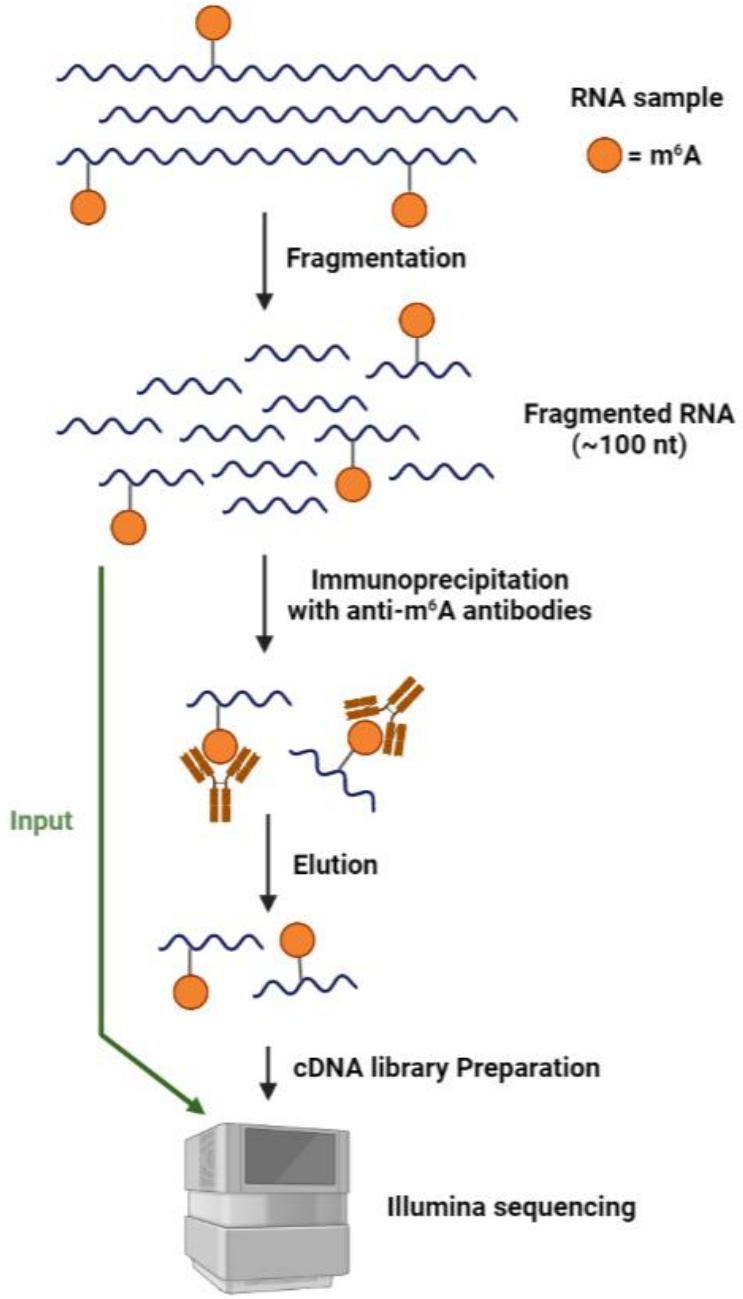
Transfer RNA (tRNA) is the most heavily modified RNA species in cells containing between 10% and 20% modified residues[24]. N^6-methyladenosine (m^6A) is the most common modification observed in RNA, where it constitutes 0.1–0.4% of adenosines, and accounts for approximately 50% of total methylations in RNA[25]. The dynamics of m^6A incorporation into RNA are regulated by "writers" (*i.e.*, methyltransferases) and "erasers" (*i.e.*, demethyltransferases). To date, the key participants involved in m^6A pathway have been well-characterized in mammals, including METTL3/14 and supporting proteins as "writers", FTO/ALKBH5 as "erasers", and YTH family proteins and associated proteins as "readers" to decode the modifications[26]. m^6A can directly affect processes such as nuclear RNA export, splicing, and RNA stability[27]. The dysregulation of these dynamics and resulting aberrant levels of m^6A has been linked to obesity, immunoregulation, and cancer[28]. In addition to m^6A, other RNA modifications have begun to catch attention in recent years. For example, N^6 ,2′-O-dimethyladenosine (m^6Am) is reported to destabilize mRNA and is controlled by the writer protein PCIF1 and eraser protein FTO[29].

1.3. Methods to map nucleotide modifications

As the first step to understand the role of DNA and RNA modifications, researchers have sought to develop methods to locate the precise locations of these modifications across the genome and transcriptome. However, each type of modification has its unique properties that have precluded a universal approach to their detection. Existing methods for the detection of nucleic acid modifications can be roughly divided into three classes: *i.* antibody-based methods; *ii.* enzyme-based methods; and *iii*. chemical reaction-based methods.

1.3.1. Antibody-based methods

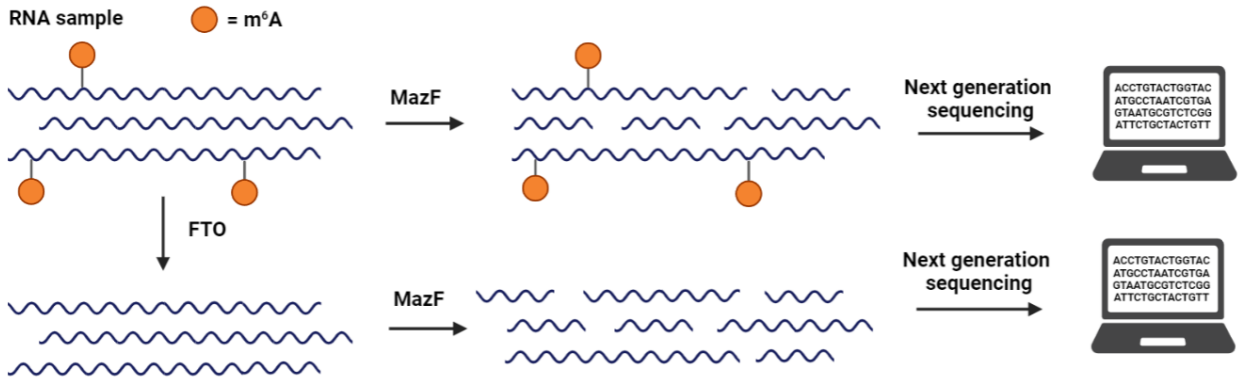
Antibodies that have high specificity and affinity to bind DNA or RNA fragments with specific modifications have been widely used for the initial profiling of various modifications. This method allows identification of sequences containing modified nucleotides and their potential consensus motifs. For example, antibodies generated against m^6A led to the identification of the DRACH (D=A, G or U; R=A or G; H=A, C or U) motif flanking m^6A[30–33]. m^6A-Seq[31] was one of the first approaches based on antibodies for the detection of m^6A modification landscape in a transcriptome-wide manner. In this method, RNA is first fragmented into ~100-nucleotide-long oligonucleotides (input) and immunoprecipitated using an anti-m^6A affinity purified antibody. After immunoprecipitation, the fragments are enriched in modifications compared to the background sequences (input). The modification sites can be estimated using high-throughput sequencing by bioinformatic comparison between fragments from the enriched portion versus the input (Scheme 1-1). Researchers further optimized the immunoprecipitation-based detection methods to improve resolution. m^6A individual-nucleotide-resolution cross-linking and immunoprecipitation (miCLIP)[30], involves the UV-induced cross-linking of the m^6A antibody to RNA. Subsequent reverse transcription (RT) of RNA cross-linked to antibodies results in a specific pattern of mutations or truncations in the cDNA that enabled precise identification of m^6A residues. Other examples of antibodies specific for modified nucleosides are those for 5mC, m^1A and m^7G[34–37].



Scheme 1-1: Schematic diagram of the m^{6}A-seq protocol. RNA is fragmented into ~100-nucleotide-long pieces and subjected to immunoprecipitation using an m^{6}A-specific antibody. Eluted m^{6}A-containing fragments and untreated input control fragments are then converted to cDNA, followed by high-throughput sequencing.
Created with BioRender.com

1.3.2. Enzyme-based methods

Some naturally occurring enzymes recognize nucleotide modifications to function *in vivo*. For instance, bacteria are able to protect themselves against foreign DNA through an enzymatically catalyzed defense mechanism in which restriction enzymes use certain modifications to distinguish their native DNA from the unmodified DNA from foreign invaders. Taking advantage of this, along with downstream enzyme evolution and screening, several methods to map modifications have been established based on certain modification-sensitive enzymes. m^6A-sensitive RNA-Endoribonuclease–Facilitated sequencing (m^6A-REF-seq)[38] utilizes an RNA endoribonuclease enzyme, MazF, which is sensitive to m^6A modification within ACA motif, specifically cleaving the unmethylated ACA motif, leaving methylated (m^6A)CA motifs intact. In this method, RNA is first digested into fragments by MazF. After end repair and purification, the RNA fragments are ligated to next-generation sequencing (NGS) adaptors and amplified by polymerase chain reaction (PCR). Comparing the resulting sequencing data with a parallel sample batch demethylated by fat mass and obesity-associated protein (FTO) in advance as the negative control, reveals the modification sites (Scheme 1-2).



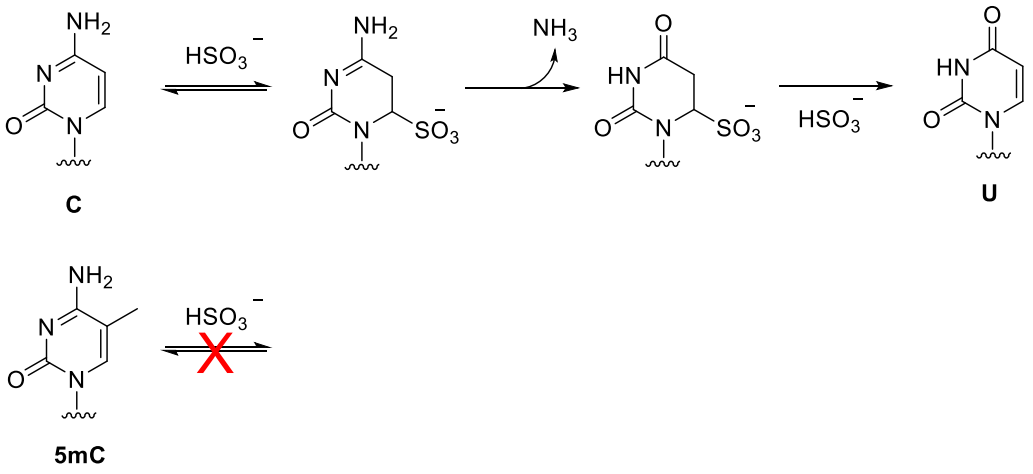
Scheme 1-2: Schematic diagram of the m^{6}A-REF-seq protocol. RNA sample and demethylated RNA control are first digested by mazF which specifically cleaves the unmethylated ACA motifs. Digested sample and digested control are then subjected to high-throughput sequencing for comparison.
Created with BioRender.com

Polymerases have also been used to detect m^{6}A in RNA by either increased mutation frequency[39,40] or decreased rate of incorporation[41] across from m^{6}A.

1.3.3. Chemical reaction-based methods

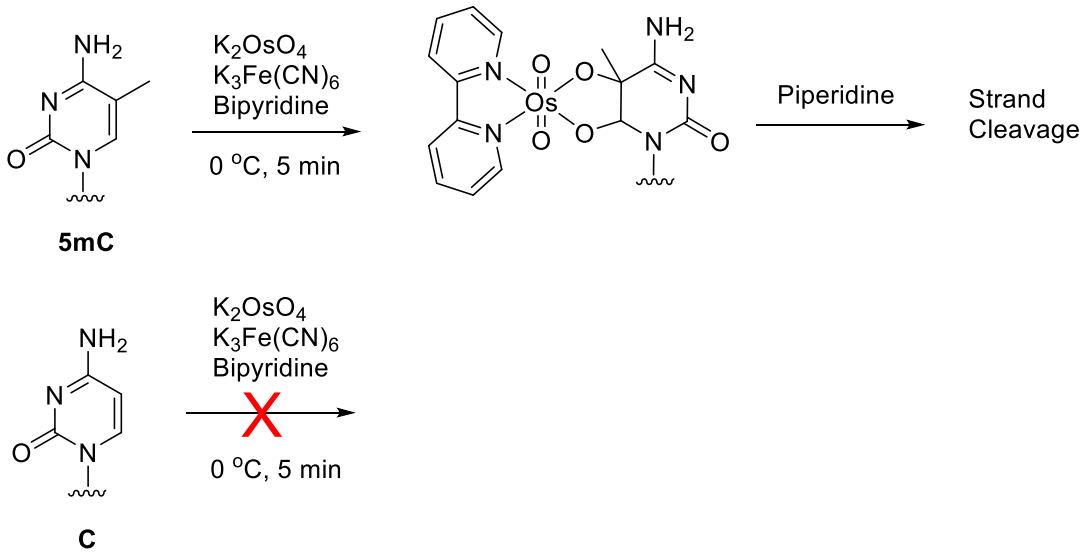
Due to the unique physical properties of some nucleobase modifications, certain chemical reactions can be specifically applied to enable their detection. During PCR or reverse transcription, the chemically reacted positions will cause polymerase or reverse-transcriptase (RT) to terminate, or introduce nucleotide transitions/transversion, and can thus be detected by sequencing. The most successful and extensive application of this strategy for modification mapping is bisulfite sequencing (BS-seq), which has been extensively used to map the sites of 5-methylcytosine (5mC) residues in DNA and RNA[42]. The method involves the bisulfite-catalyzed chemoselective deamination of cytosine resulting in a cytosine to uracil transition, while leaving 5mC largely unaffected by the process (Scheme 1-3). Thus, comparative sequencing analysis against a no-reaction

control can be used to readily identify the locations of 5mC within a DNA or RNA sequence.



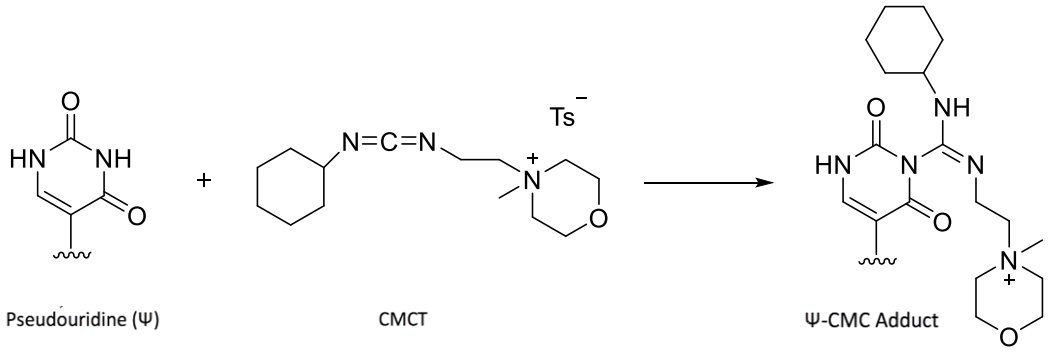
Scheme 1-3: Chemical reaction that underlies bisulfite sequencing. In the presence of bisulfite, C bases chemoselectively deaminate to U, leaving 5mC bases unaffected.

Another method for chemical discrimination between 5-methylcytosine and cytosine utilizes the large difference between their rate of oxidation with Osmium(VIII) (Scheme 1-4), which has been successfully employed in the discrimination of these bases in DNA[43]. The oxidized strand can be cleaved at the oxidized methylated base with hot piperidine, and then analyzed as a band for a shortened strand using polyacrylamide gel electrophoresis. Although OsO_4 reacts selectively with the more electrophilic C5-C6 double bond of 5-methylcytosine, and leaves the same bond in cytosine intact, it is highly toxic and mutagenic.



Scheme 1-4: Chemical reaction that underlies selective oxidation of 5-methylcytosine. OsO_4 reacts selectively with the more electrophilic C5-C6 double bond of 5-methylcytosine bases, leaving the same bond in cytosine bases intact.

The oxidation of methylated cytosine by vanadium species or sodium periodate was also reported in which the combination of V_2O_5–lithium bromide or sodium periodate–lithium bromide was used[44]. Similarly, the strand breaks in methylated cytosines were observed by denaturing polyacrylamide gel electrophoresis after treatment of the reaction product with hot piperidine. Another example of a chemistry-based method is the pseudouridine site identification sequencing (PSI-seq) for the detection of pseudouridine (Ψ)[45]. This method uses 1-cyclohexyl-(2-morpholinoethyl)carbodiimide metho-*p*-toluene sulfonate (CMCT) to selectively modify pseudouridines and generate RT stop sites. A CMC-adduct forms by the reaction of pseudouridine with CMCT followed by alkaline treatment and the bulky CMC group stops reverse transcription of RNA (Scheme 1-5). Thiouridines have also been reported to react with carbodiimides like CMCT, and the resulting conjugates can be identified by mass spectroscopy[46].



Scheme 1-5: Chemical reaction that underlies PSI-Seq.
1-cyclohexyl-(2-morpholinoethyl)carbodiimide metho-*p*-toluene sulfonate (CMCT) selectively modifies pseudouridine bases generating an adduct that can stop reverse transcription of RNA

1.4. Limitations of current techniques for the detection of nucleotide modifications

Even though several methods exist to probe nucleotide modifications, each of these suffer from limitations. For example, immunoprecipitation (IP) of short RNA fragments using m^{6}A-specific antibodies, MeRIP-seq[31,32], followed by sequencing provides low resolution mapping. Similarly, 6mA-DIP-seq[47,48], antibody-based method for the detection of 6mA, is also limited by the low resolution. miCLIP[30], which involves the UV-induced cross-linking of the m^{6}A antibody to RNA, increases the resolution of m^{6}A signature, but requires a cytosine residue at the +1-position, rendering a potentially large number of m^{6}A sites undetectable. 6mA-crosslinking-exonuclease-sequencing (6mACE-seq)[49], which utilizes 6mA-specific antibodies cross-linked to 6mA sites to protect 6mA-DNA fragments from subsequent exonuclease treatment, enables single-nucleotide resolution, but suffers from an extensive workflow. m^{6}A-sensitive RNA endoribonuclease-facilitated sequencing (m^{6}A-REF-seq) detects only at the ACA motif, which reduces sequence space[38]. PacBio

single molecule real-time (SMRT) sequencing technology[50] for the detection of 6mA in DNA, enhances the resolution down to the single-nucleotide level, but suffers from false positives and struggles with genomes high in 5mC. Polymerases that have been used to detect m^6A in RNA[39,40] have yet to find wide-scale use, and can give false positives of adenosines that are in close proximity downfield from the m^6A site[39]. As opposed to enzyme and antibody-based methods, chemical reactions are often less sequence dependent, can work on either DNA or RNA, and thus can provide a robust, inexpensive and universal sequencing approach to probe DNA and RNA modifications. Despite development of specific chemical reagents, the spectrum of available chemicals capable of selective reaction with modified oligonucleotides remains rather limited.

1.5. Thesis project

One limitation in studying the functional, physiological, and pathological roles of the epigenome and epitranscriptome is the availability of methods for the precise mapping of individual DNA and RNA modifications throughout the genome and transcriptome. The goal of the work presented in the first part of this thesis is to develop novel chemistry-based methods for the detection of certain epigenetic and epitranscriptomic modifications.

Chapter 2. Single-nucleotide resolution of N^6-adenine methylation sites in DNA and RNA by nitrite sequencing[1]

[1] The material presented in this chapter is reproduced from “Mahdavi-Amiri, Y.; Chung Kim Chung, K.; Hili, R. *Chem. Sci.* **2021**, *12*, 606-612” with permission from the Royal Society of Chemistry.

2.1. Abstract

A single-nucleotide resolution sequencing method of N^6-adenine methylation sites in DNA and RNA is described. Using sodium nitrite under acidic conditions, chemoselective deamination of unmethylated adenines readily occurs, without competing deamination of N^6-adenine sites. The deamination of adenines results in the formation of hypoxanthine bases, which are read by polymerases and reverse transcriptases as guanine; the methylated adenine sites resist deamination and are read as adenine. The approach, when coupled with high-throughput DNA sequencing and mutational analysis, enables the identification of N^6-adenine sites in RNA and DNA within various sequence contexts.

2.2. Introduction

The ability to map methylation sites in the human genome and epitranscriptome has transformed our understanding of how these modifications govern and influence a host of cellular processes and diseases[51,52]. Amongst the most widely studied methylations is N^6-methyladenine, known as 6mA in DNA and m^6A in RNA. m^6A is the most common methylation observed in RNA, where it constitutes 0.1–0.4% of adenosines, and accounts for approximately 50% of total methylations in RNA[25]. The dynamics of m^6A incorporation into RNA are regulated by “writers” (i.e., methyltransferases) and “erasers” (i.e., demethyltransferases), and can directly affect processes such as nuclear RNA export, splicing, and RNA stability[27]. Not surprisingly, the deregulation of these dynamics and resulting aberrant levels of m^6A has been linked to obesity, immunoregulation, and cancer[28]. While 6mA has been widely known as a DNA modification in prokaryotes, its presence in eukaryotes has only been recently established, including in humans where it

represents ~0.051% of the genome[19]. 6mA is thought to play an epigenetic role in embryonic development[20], tumorigenesis[19], response to stress, neuropsychiatric disorders[21], and embryonic stem cell function[22], and it can be inherited[53]. Understanding the role of *N*^{6}-methyladenine in RNA and DNA requires robust single-nucleotide sequencing methods. Due to the similar Watson–Crick–Franklin hydrogen-bonding nature of adenine and *N*^{6}-methyladenine with thymine (Figure 2-1), direct high-throughput sequencing has been challenging using conventional methods.

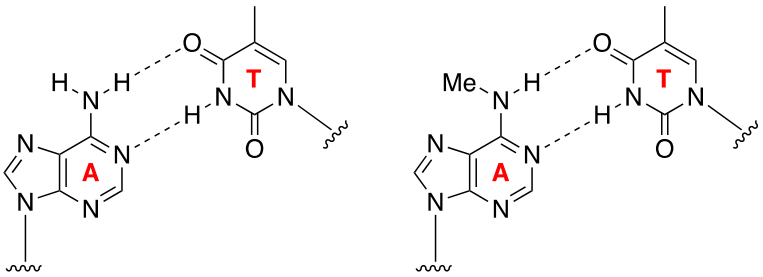


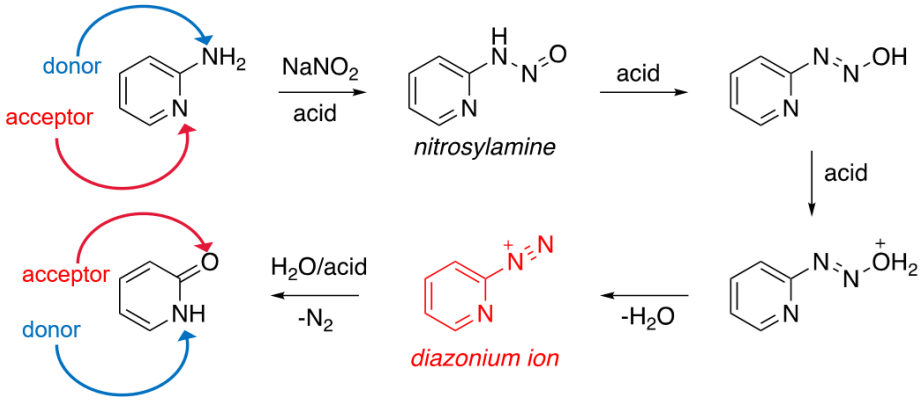
Figure 2-1: Similar Watson-Crick-Franklin base-pairing observed between adenine and thymine and *N*^{6}-methyladenine and thymine.

This notwithstanding, several existing methods have been developed to probe the m^{6}A and 6mA methylomes; however, each of these suffer from limitations. Immunoprecipitation (IP) of short RNA fragments using m^{6}A-specific antibodies, MeRIP-seq[31,32], followed by sequencing provides low resolution mapping; miCLIP[30], which involves the UV-induced cross-linking of the m^{6}A antibody to RNA, requires a cytosine residue at the +1-position, rendering a potentially large number of m^{6}A sites undetectable; m^{6}A-sensitive RNA endoribonuclease- facilitated sequencing (m^{6}A-REF-seq) detects only at the ACA motif, which reduces sequence space; polymerases have also been used to detect m^{6}A in RNA by either increased mutation frequency[39,40],or decreased rate of incorporation[41] across from m^{6}A; however, these have yet to find wide-scale use, and can

give false positives of adenosines that are in close proximity downfield from the m^{6}A site[39]. Similarly, while several 6mA sequencing methods are available, many of them suffer from issues. Traditional IP-based methods, such as 6mA-DIP-seq[47,48], suffer from low resolution; IP methods coupled with restriction digest, such as DA-6mA-seq[54], improve resolution at the expense of sequence space; PacBio single molecule real-time (SMRT) sequencing technology[50], enhances the resolution down to the single-nucleotide level, but suffers from false positives[55] and struggles with genomes high in 5mC[56] and 6mA-crosslinking-exonuclease-sequencing (6mACE-seq), enables single-nucleotide resolution, but suffers from an extensive workflow. New single-nucleotide sequencing methods for both m^{6}A and 6mA continue to be needed to provide access to probe the complete sequence space of RNA and DNA, enabling in-depth functional studies of these methylomes.

As opposed to enzyme-mediated sequencing methods, chemical reactions are often less sequence dependent, can work on either DNA or RNA, and thus can provide a robust, inexpensive, and universal sequencing approach to probe the 6mA and m^{6}A methylomes. To this end, I was inspired by the simplicity of bisulfite sequencing[42], which has been extensively used to map the sites of 5-methylcytosine (5mC) residues in DNA and RNA. The method involves the bisulfite-catalysed chemoselective deamination of cytosine resulting in a cytosine to uracil (C → U) transition, while leaving 5mC largely unaffected by the process. Thus, comparative sequencing analysis against a no-reaction control can be used to readily identify the locations of 5mC within a DNA or RNA sequence. I was inspired to use a similar approach to enable the single nucleotide resolution of m^{6}A and 6mA in RNA and DNA, respectively. To achieve this, a chemical reaction was required

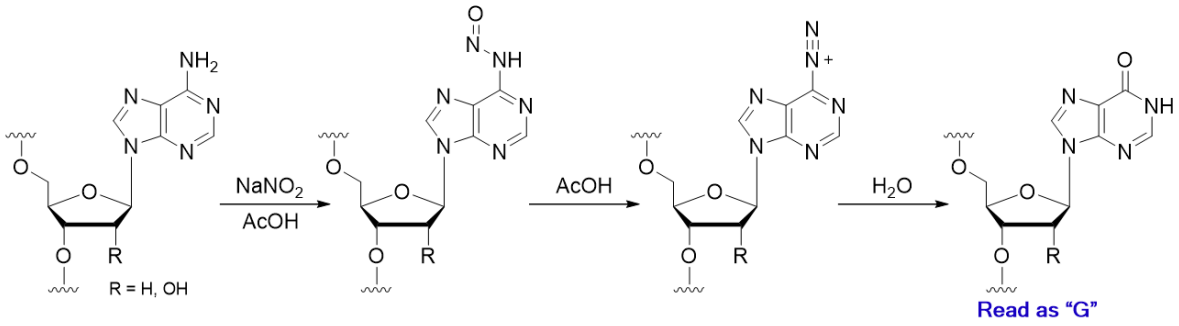
that (i) was water tolerant; (ii) did not degrade DNA or RNA; (iii) was chemoselective for either *N*^{6}-methyladenine or unmethylated adenine; and (iv) resulted in a change in how the nucleobase was read by a polymerase or reverse transcriptase. I was drawn to the nitrite-mediated diazotisation of aromatic amines, first described by Griess, as a possible reaction that would satisfy the four criteria – in particular the process later described on 2-aminopyridines[57] (Scheme 2-1).



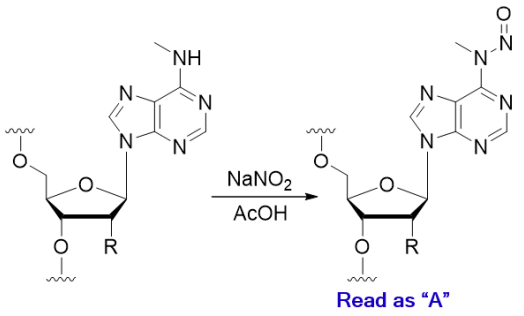
Scheme 2-1: Reaction of 2-aminopyridine with sodium nitrite under acidic aqueous conditions resulting in exchange of hydrogen bond donors and acceptors.

In the presence of acid under aqueous conditions, nitrite forms reactive nitrosonium ion, which reacts with aromatic amines to form nitrosamines. Subsequent dehydration to form the diazonium ion can only proceed with primary aromatic amines, as secondary aromatic amines lack the additional N–H required for dehydration. Hydrolysis of the diazonium yields the deaminated product. Accordingly, the process should be chemoselective for the primary exocyclic amine of adenine over the secondary exocyclic amine of *N*^{6}-methyladenine seen in m^{6}A and 6mA (Scheme 2-2 and Scheme 2-3). Thus, only unmethylated adenine will be hydrolysed under these conditions to form hypoxanthine – an exchange of a hydrogen bond donor for a hydrogen bond acceptor. Inosine

nucleobases are known to base pair with cytidine (Figure 2-2) and polymerases read them as guanosine[58], resulting in an A → G transition, which can be detected by high-throughput DNA sequencing.



Scheme 2-2: Reaction of nitrite with adenine bases in RNA or DNA.



Scheme 2-3: Reaction of nitrite with *N*^{6}-methyladenine bases in RNA or DNA.

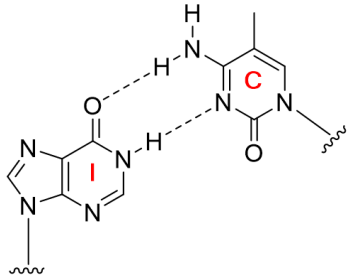
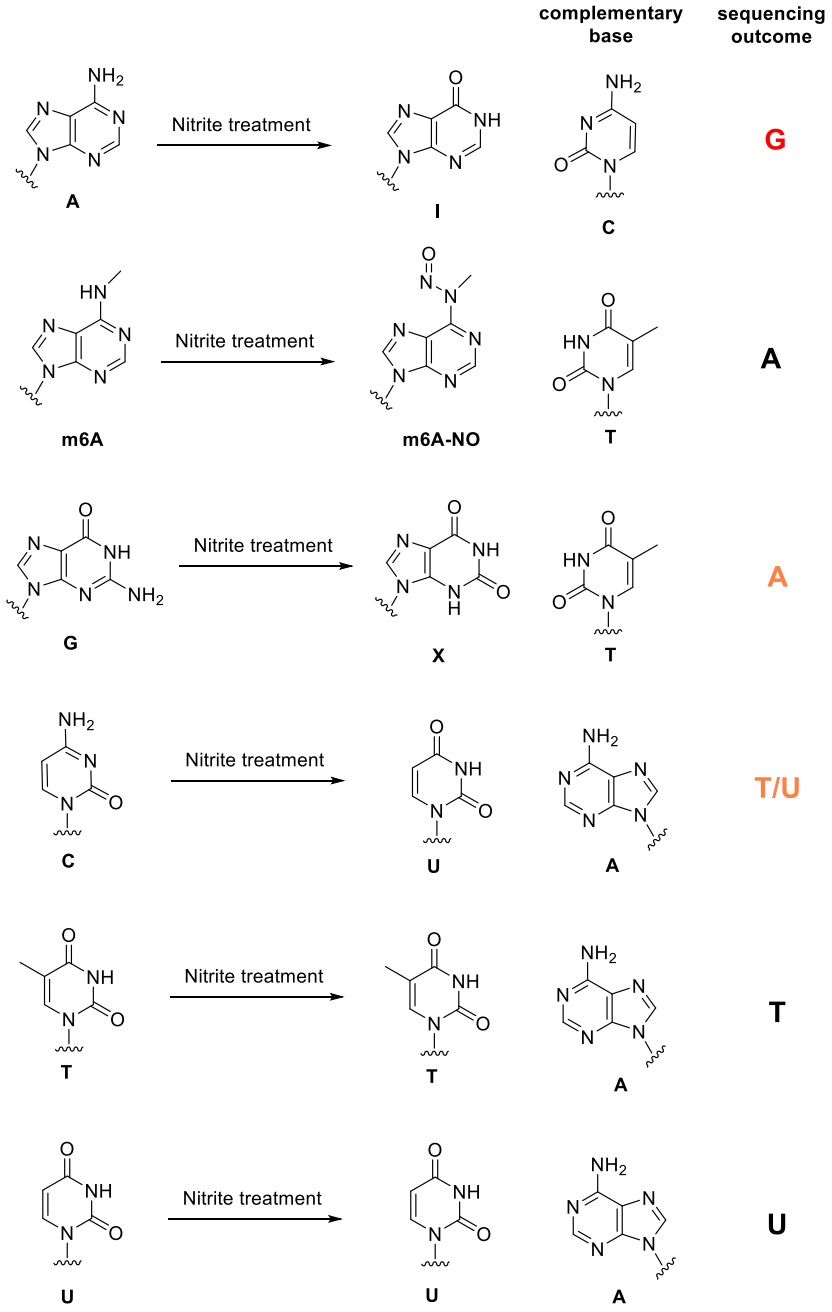


Figure 2-2: base-pairing between Inosine and cytidine.

Other exocyclic amines in DNA and RNA will also be susceptible to nitrite mediated deamination, including those on guanine and cytosine, which will result in G → A transitions and C → T/U transitions (Scheme 2-4); however, these can be handled during sequencing data analysis.



Scheme 2-4: Base conversion after nitrite treatment and following influence on sequencing.

2.3. Results and discussion

2.3.1. Nitrite-mediated deamination on single nucleosides

The nitrite-mediated deamination process was first examined on free adenosine. Using a 1 M aqueous $NaNO_2$ in the presence of 1.7% AcOH at 22 ^oC, complete consumption of adenosine into inosine was observed by HPLC analysis over a 12 h period (Figure 2-3a). Deamination of guanosine into xanthosine (Figure 2-3b) and cytidine into uridine (Figure 2-3c) was largely completed over a 12 h period under similar conditions. This suggests that nitrosylation and subsequent diazotisation of adenosine could be achieved using conditions that are compatible with nucleic acids. It was observed that deamination of adenosine into inosine was over 50% completed within 5 h. In order not to scramble the alignment of DNA and RNA sequences against a genome, it was decided that 5 h would be sufficient for detecting difference in deamination at methylated sites.

When subjecting N^6-methyladenosine to the same conditions, full conversion into N^6-nitroso-m^6A was observed within 3.5 h, with no trace amounts of inosine formed over a 12 h period (Figure 2-4). The lack of conversion of m^6A-NO into inosine highlights the resistance to hydrolysis under the tested experimental conditions. Interestingly, m^6A becomes nitrosylated significantly faster than adenosine owing to its increased nucleophilicity at the N^6 position.

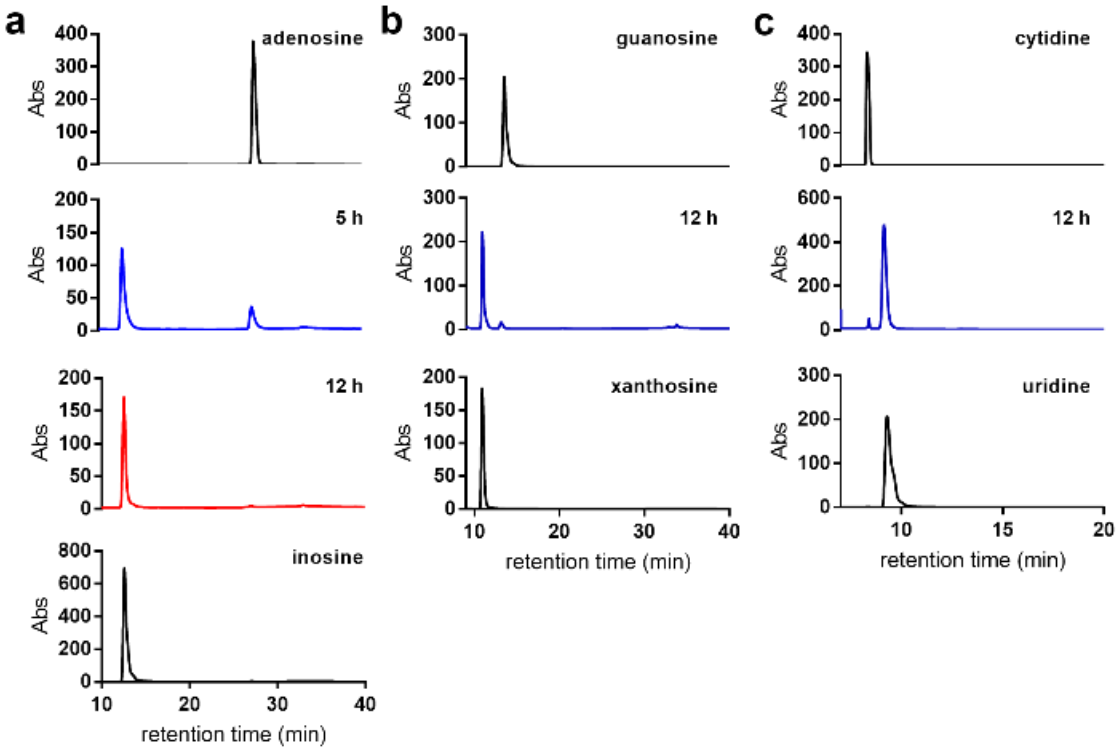


Figure 2-3: HPLC analysis of the conversion of (a) adenosine into inosine (b) guanosine into xanthosine and (c) cytidine into uridine using $NaNO_2$ and 1.7% aqueous AcOH at 22 ^oC over 12h.

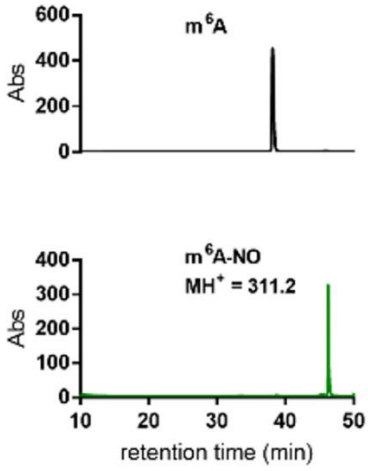


Figure 2-4: HPLC analysis of the conversion of m^6A into nitrosylated m^6A (m^6A-NO was confirmed by ESI-MS) using $NaNO_2$ and 1.7% aqueous AcOH, at 22 ^oC over 3.5 h. Note that inosine (retention time = 11 min) was not detected.

Other examined methylated nucleosides, including m^1A and m^3C were unreactive under the tested conditions (Figure 2-5). This is due to the decrease in electron density of these positively charged nucleobases.

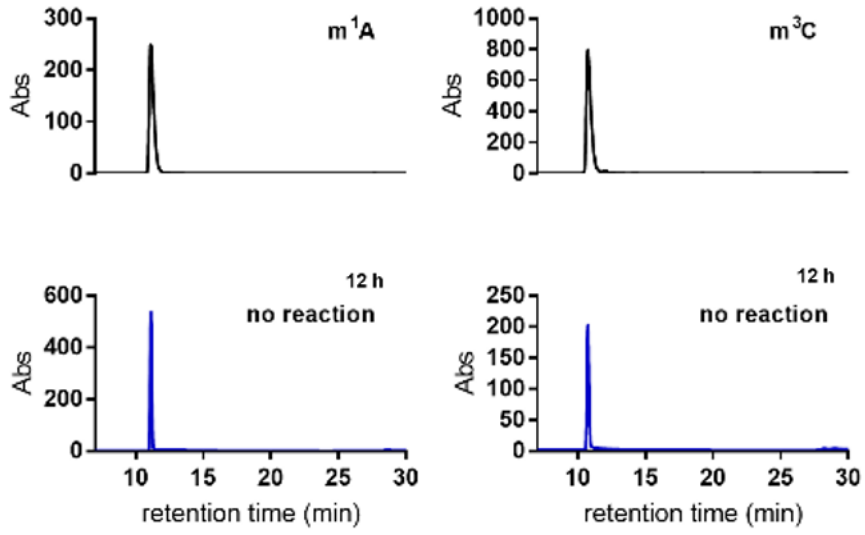


Figure 2-5: HPLC analysis showing unreactivity of m^1A and m^3C under nitrite mediated deamination conditions. Using $NaNO_2$ and 1.7% aqueous AcOH, at 22 ^oC, no reaction was observed over a 12 h period.

2.3.2. Optimisation of nitrite-mediated deamination on DNA and RNA

Prior to evaluating the performance of the nitrite-mediated deamination process on sequencing, the stability of RNA and DNA in the reaction conditions was determined while optimising variables (tested conditions are listed in Table 2-1 in Experimental detail section). It was found that acid had the most profound effect on the stability of RNA and DNA during the process. Using a ssDNA and ssRNA as models, the degradation of the sequences was monitored with increasing acid concentration using 1 M $NaNO_2$ for 5 h at 22 ^oC (Figure 2-6). It was observed that DNA was far more sensitive than RNA

under the acidic conditions used. The degradation was attributed due to acid-catalysed depurination and backbone cleavage, albeit cationic intermediates during the diazotisation process could also play a role. RNA, with its electronegative 2'-OH group is less susceptible to this depurination/cleavage process.

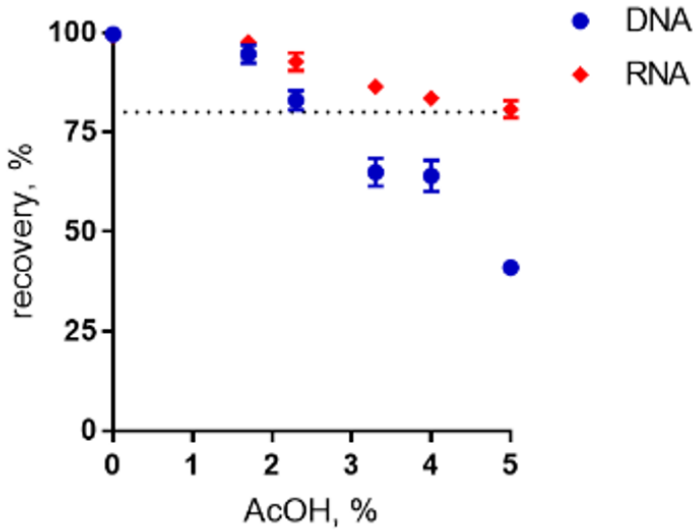


Figure 2-6: Recovery of DNA and RNA with respect to acid concentration during the nitrite reaction. Error based on assessment in duplicates. Dotted line represents 80% threshold of recovery.

To facilitate isolation and the study of low amounts of DNA and RNA, it was decided to place an 80% recovery threshold on the process, which limited acid concentration for RNA to 5% and DNA to 2.3%.

I next sought to study and optimise the A → G transition reaction on a model 60 nt RNA sequence containing one instance of m^6A. The sequence was subjected to 1 M $NaNO_2$ for 5 h at 22 ^oC with acetic acid concentrations ranging from 0 to 5%. As anticipated, it was observed that increasing the percentage of AcOH increased the A → G transitions

from background error rates of less than 0.1% transitions per adenosine to 14% when using 5% AcOH (Figure 2-7), which is attributed to acid promoted increase in nitrosonium ion concentration. Importantly, these data demonstrate no change in the frequency of A → C and A → U transversions caused by the reaction. As expected, deamination at cytosine and guanosine was observed, resulting in C → U and G → A mutations (Figure 2-7).

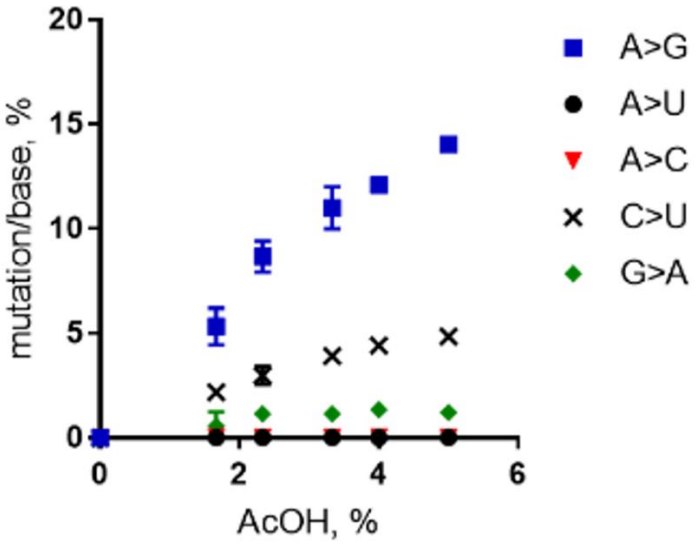


Figure 2-7: High-throughput sequencing of RNA after nitrite reaction at varying acid concentrations. ">" denotes corresponding transition or transversion. Mutations are represented in legend and correspond to the specific type of mutation per expected nucleobase.

Fortuitously, nitrosylated m^6A was read as adenosine by reverse transcriptase, and had a similar frequency of A → G transitions from adenosines in the no-reaction control. This result was unexpected due to the loss of canonical hydrogen-bonding to thymine during reverse transcription; however, alternative non-canonical interaction with thymine might be at play that give preference to thymine incorporation.

Due to the lower stability of DNA under the AcOH-promoted nitrite reaction, only the acid concentrations yielding >80% recovery were examined. Similar to the RNA experiments, increasing mutation frequencies of dA → dG, dC → dT, and dG → dA were observed with increasing AcOH concentrations (Figure 2-8).

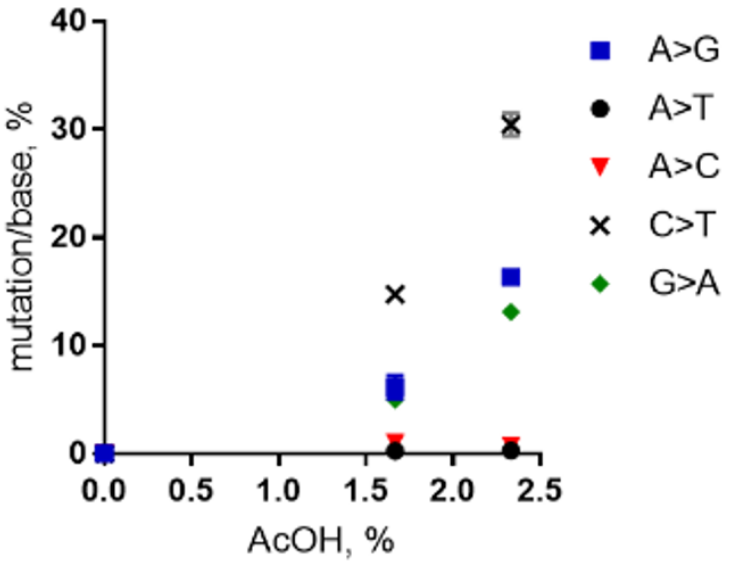


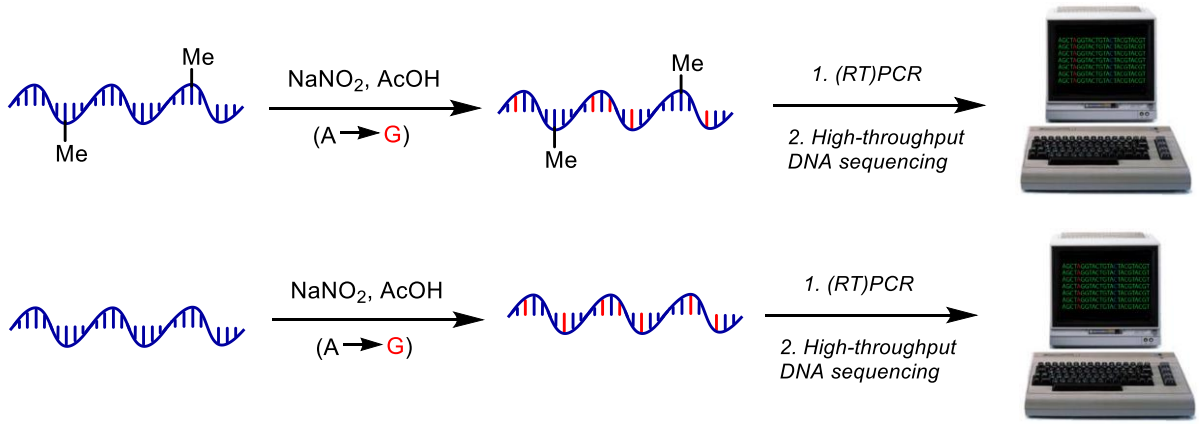
Figure 2-8: High-throughput sequencing of DNA after nitrite reaction at varying acid concentrations. ">" denotes corresponding transition or transversion. Mutations are represented in legend and correspond to the specific type of mutation per expected nucleobase. Note that high-throughput DNA analysis above 2.3% AcOH was not processed due to undesirably low isolation (per Figure 2-6).

Curiously, dC → dT mutations were greater than those of dA → dG – the opposite of which was observed in RNA (Figure 2-7). The higher propensity for deamination of cytosine in DNA over that of RNA has been previously observed in activation-induced deaminase processing of nucleic acids[59]. The increase in dG → dA mutation in DNA over RNA is unclear and compounded by the fact that deamination of the adenine base results in xanthine, which may be read with different error frequencies and propensities by DNA

polymerases and reverse transcriptases. After concluding the optimisation studies, it was found that the recovery boundary concentrations of 5% AcOH for RNA and 2.3% AcOH for DNA represented the best conditions for deamination activity. While, in principle, these mutations could be increased by further optimisation, it was decided not to push the process too far so as to avoid issues in sequence alignment during high-throughput sequencing analysis.

2.3.3. Evaluation of nitrite-mediated sequencing of *N*^{6}-methyladenine sites in DNA and RNA

With the optimised system in hand, the sequencing method was examined for its ability to detect *N*^{6}-methyladenine within DNA and RNA (Scheme 2-5). A 99 nt DNA sequence containing a single 6mA site was incubated with 1 M $NaNO_2$ and 2.3% aqueous acetic acid, and subsequently analysed by high-throughput DNA sequencing, trimmed for length and quality, and aligned to the reference sequence using bowtie 1 to enable induced SNP calling[60]. The demethylated sequence was also subjected to the same process for comparative analysis.



Scheme 2-5: Nitrite sequencing workflow.

As expected, extensive deamination was observed, with dA → dG transitions increasing >50-fold against the no-reaction control. The normalised ratio (R) of the dA → dG transitions at each nucleotide position compared to that of the demethylated sequence was plotted:

$$R = \frac{Freq.\ A \rightarrow G\ demethylated\ control}{Freq.A \rightarrow G\ methylated\ sample}$$

This afforded a convenient way to visualise the nitrite sequencing data (Figure 2-9). High A → G transition ratios are observed only at the 6mA sites, which is consistent with the nucleoside reaction data.

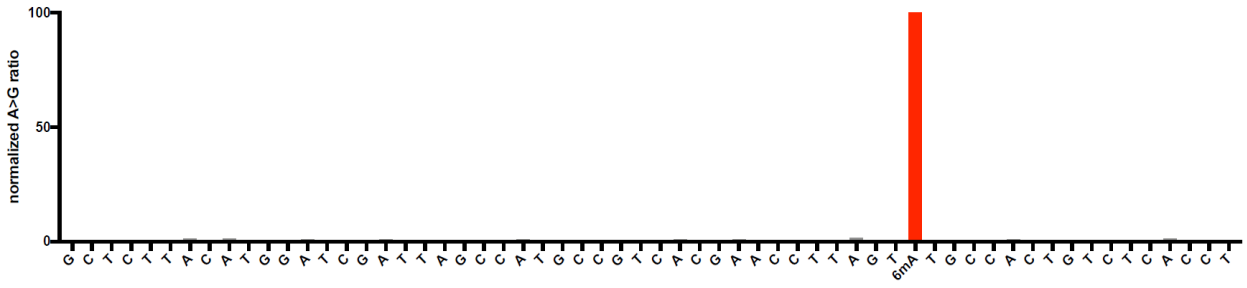


Figure 2-9: Normalised sequencing representation of the ratio of A → G mutation at each nucleobase following nitrite treatment of a ssDNA containing a single 6mA site at position 63. Primer sequence regions are not shown for clarity.

Encouraged by these findings, 6mA sequencing was attempted on a more challenging template – one comprising two dAs flanking a 6mA site, and also a double 6mA site, which would be overlooked by most existing sequencing methods should such motifs occur in nature. The method readily detected the flanked 6mA site, highlighting the single-nucleotide resolution. The contiguous 6mA sites were more challenging, yet still distinguished from unmethylated adenine sites (Figure 2-10). This slightly lower response

may be due to neighbouring group effects during diazotisations of adjacent nitrosylated adenines.

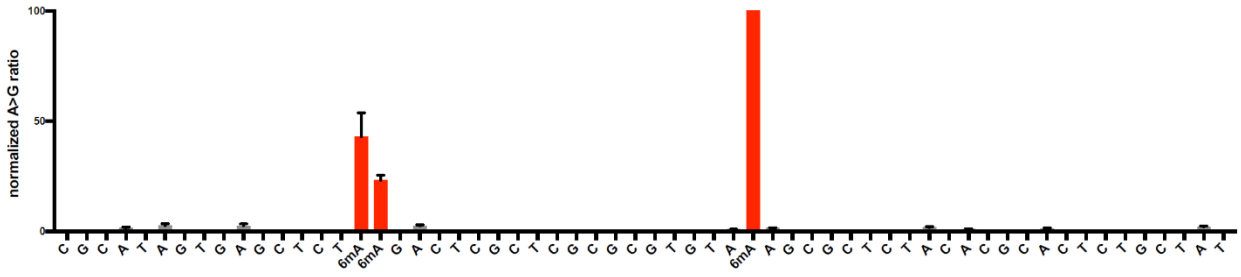


Figure 2-10: Normalised sequencing representation of the ratio of A → G mutation at each nucleobase following nitrite treatment of a ssDNA containing three 6mA sites at positions 35, 36, and 55. Primer sequence regions are not shown for clarity.

The method was also compatible with duplex DNA and readily detected the 6mA site (Figure 2-11), albeit with an expected decrease in response likely resulting from amplification of the non-target strand.

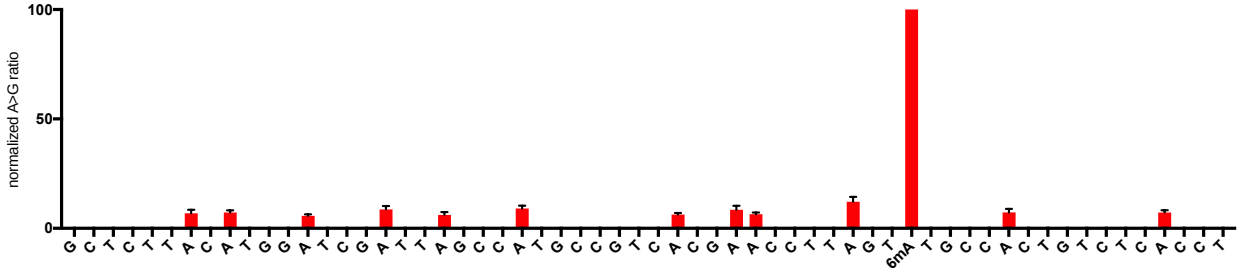


Figure 2-11: Normalised sequencing representation of the ratio of A → G mutation at each nucleobase following nitrite treatment of a dsDNA containing a single 6mA site at position 63. Duplex DNA was generated from a sequence containing a single 6mA site at position 63 using Klenow extension. Primer sequence regions are not shown for clarity.

Next, nitrite sequencing method was explored to detect m^6A in RNA using similar conditions as those used for DNA. One 60 nt sequence comprised a single m^6A flanked by two adenosines, which yielded good differentiation amongst other adenosines in the

sequence (Figure 2-12), again highlighting the single-nucleotide discrimination of the nitrite sequencing method.

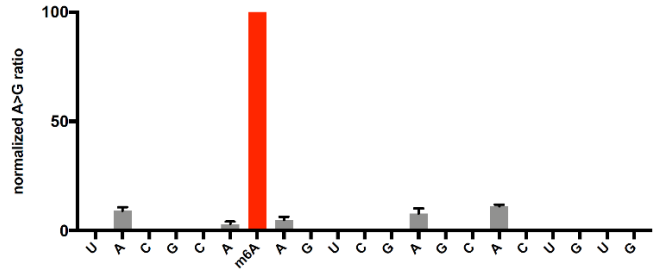


Figure 2-12: Normalised sequencing representation of the ratio of A → G mutation at each nucleobase following nitrite treatment of a ssRNA containing a single m^6A site at position 26. Primer sequence regions are not shown for clarity.

I also attempted the sequencing method on a contiguous instance of m^6A within a 60 nt RNA. Good detection above background was observed (Figure 2-13); however, issues with potential neighbouring group interference of nitrosylation were similarly noted.

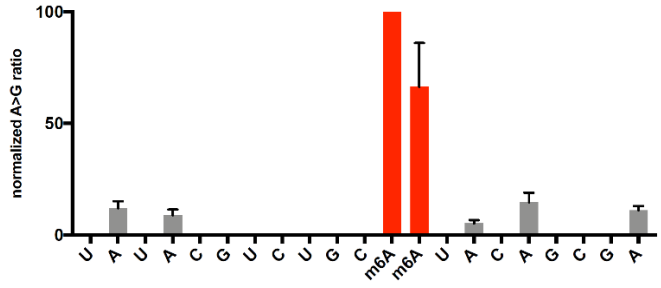


Figure 2-13: Normalised sequencing representation of the ratio of A → G mutation at each nucleobase following nitrite treatment of a ssRNA containing two m^6A sites at positions 31 and 32. Primer sequence regions are not shown for clarity.

Due to the importance of quantifying the methylation fraction at potential m^6A sites, a spike-in experiment that assessed the response for varying fractions of m^6A at a specific

adenosine site in RNA was performed. It was found that the nitrite sequencing method was able to quantify m^6A fractions down to 50%, below which the response was not significant above background levels (Figure 2-14).

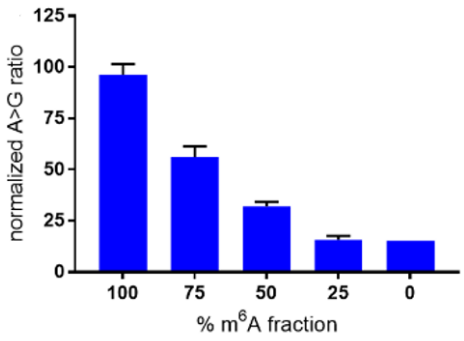


Figure 2-14: Quantification of methylation fraction of an adenosine site within an RNA sequence.

The sequencing method was further applied to detect naturally occurring m^6A in isolated RNA. To this end, *E. coli* rRNA, which is known to have an m^6A site at position 2030 of the 23S subunit[61], was purified and subjected to nitrite sequencing (Figure 2-15). The m^6A site at position 2030 was readily detected, with approximately 10-fold increase in signal over neighbouring unmodified adenosines. It was observed that peptides interfered with the desired nitrite chemistry on RNA, and thus should be thoroughly removed from samples.

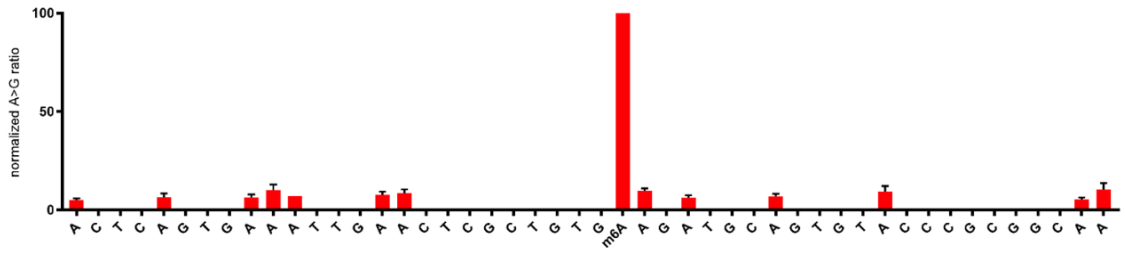


Figure 2-15: Normalised sequencing representation of the ratio of A → G mutation at each nucleobase following nitrite treatment of rRNA from *E. coli*. The 23S rRNA from *E. coli* contains a single m^6A site at position 2030. Primer sequence regions are not shown for clarity.

In all sequencing experiments, slightly higher background noise with RNA nitrite sequencing was observed compared with that of DNA. This could potentially be related to greater folding of single-stranded RNA versus DNA. Potential avenues around this would be the addition of mild denaturants and solvents. Such optimizations may also boost the quantification range for the level of methylation at putative m^6A sites and enable detection of low abundance m^6A sites in biological samples.

2.3.4. Potential applications and limitations of nitrite sequencing toward other modifications

The nitrite-mediated deamination process on DNA and RNA is anticipated to have further applications but also limitations in resolving other related methylation and alkylation sites. For instance, N^6,2'-O-dimethyladenosine (m^6Am), which is located in certain RNA transcripts at the first position following the 7-methylguanosine cap, would also not be able to undergo deamination in the presence of nitrite, resulting in a high R value similar to m^6A. While this could potentially yield false positives for m^6A sequencing, m^6Am is primarily located at the adenosine of the first encoded nucleotide in mRNA and could be handled through post-sequencing analysis. Furthermore, in principle, the nitrite

sequencing method could be used to identify such m^6Am sites in transcripts. Other common modified nucleosides that would give high R values during sequencing analysis were also identified. m^1A and m^3C, both of which are too electron poor to react with nitrosonium ion under the examined conditions (Figure 2-5), do not deaminate, thus this method could potentially be used for m^1A and m^3C sequencing to complement other burgeoning methods[62]. Also, very recently, our method has successfully been used by another group for genome-wide mapping of 6mA in *Escherichia coli* and *Helicobacter pylori*[63], which can boost the study of bacterial epigenetics.

2.4. Conclusions

In conclusion, the first chemistry-based method to facilitate the sequencing of both m^6A in RNA and 6mA in DNA was demonstrated. The chemistry takes advantage of the acid-mediated nitrite reaction that chemoselectively deaminates adenine in the presence of N^6-methyladenine. This results in a large increase in (d)A → (d)G transitions only at unmethylated sites. When coupled to high-throughput DNA sequencing, nitrite sequencing enables the identification of m^6A and 6mA sites at single-nucleotide resolution. This sequencing method is already finding broad use in labs across the globe as a straightforward and affordable approach to detect N^6-adenine methylation sites in RNA and DNA. Its application to RNA and DNA sequencing is expected to facilitate important discoveries in fundamental epigenetics and epitranscriptomic research and as a diagnostic detection tool for methylation sites implicated in human disease.

2.5. Experimental details

2.5.1. General information

Unless otherwise noted, water was purified with the MilliQ Direct Q3. DNA and RNA oligonucleotides were purchased from Integrated DNA Technologies, with HPLC purification. Nucleoside analysis was performed by reverse-phase high-performance liquid chromatography (HPLC, Agilent 1260 Infinity II) using a C18 stationary phase (Phenomenex, Luna® 5 µm C18(2) 100 Å, 250 x 4.6 mm) and an acetonitrile/100 mM triethylammonium acetate gradient. Oligonucleotide concentrations were determined using a Qubit 4.0 Fluorometer. High-throughput DNA sequencing samples were quantified using a Qubit 4 Fluorometer, prepared on an IonChef instrument and sequenced on an Ion Torrent GeneStudio S5 Plus using Ion 530 Chips.

2.5.2. DNA Sequences

Note: /iN6Me-dA/ = N^6-methyldeoxyadenosine

The sequences below are written from 5'→3':

temp1_unmeth GCT AGT TAT TGC TCA GCG GGC TCT TAC ATG GAT CGA TTA GCC ATG CCG TCA CGA ACC TTA GTA TGC CAC TGT CTC ACC TCC CTA TAG TGA GTC GTA TTA

temp2_unmeth GCT AGT TAT TGC TCA GCG GCG CAT AGT GAG CTC TAA GAC TCG CTC GCG CGT GTA AAG CGC TCT ACA CGC ACT CTG CTA TCC CTA TAG TGA GTC GTA TTA

temp1_6mA_pos63 GCT AGT TAT TGC TCA GCG GGC TCT TAC ATG GAT CGA TTA GCC ATG CCG TCA CGA ACC TTA GT/iN6Me-dA/ TGC CAC TGT CTC ACC TCC CTA TAG TGA GTC GTA TTA

temp2_6mA_pos35-36-55 GCT AGT TAT TGC TCA GCG GCG CAT AGT GAG CTC T/iN6Me-dA//iN6Me-dA/ GAC TCG CTC GCG CGT GTA /iN6Me-dA/AG CGC TCT ACA CGC ACT CTG CTA TCC CTA TAG TGA GTC GTA TTA

2.5.3. RNA Sequences

Note: /iN6Me-rA/ = N^6-methyladenosine

The sequences below are written from 5'→3':

temp3_unmeth rGrCrU rArGrU rUrArU rUrGrC rUrCrA rGrCrG rGrUrA rUrArC rGrUrC rUrGrC rArArU rArCrA rGrCrG rArCrC rCrUrA rUrArG rUrGrA rGrUrC rGrUrA rUrUrA

temp4_unmeth rGrCrU rArGrU rUrArU rUrGrC rUrCrA rGrCrG rGrUrA rCrGrC rArArA rGrUrC rGrArG rCrArC rUrGrU rGrCrC rCrUrA rUrArG rUrGrA rGrUrC rGrUrA rUrUrA

temp3_m^{6}A_pos31-32 rGrCrU rArGrU rUrArU rUrGrC rUrCrA rGrCrG rGrUrA rUrArC rGrUrC rUrGrC /iN6Me-rA//iN6Me-rA/rU rArCrA rGrCrG rArCrC rCrUrA rUrArG rUrGrA rGrUrC rGrUrA rUrUrA

temp4_m^{6}A_pos26 rGrCrU rArGrU rUrArU rUrGrC rUrCrA rGrCrG rGrUrA rCrGrC rA/iN6Me-rA/rA rGrUrC rGrArG rCrArC rUrGrU rGrCrC rCrUrA rUrArG rUrGrA rGrUrC rGrUrA rUrUrA

2.5.4. IonCode adapters

T7_P1 CCA CTA CGC CTC CGC TTT CCT CTC TAT GGG CAG TCG GTG ATT AAT ACG ACT CAC TAT AGG G

T7_term_IC_0101 CCA TCT CAT CCC TGC GTG TCT CCG ACT CAG CTAAGGTAAC GGTGAT GCT AGT TAT TGC TCA GCG G

T7_term_IC_0102 CCA TCT CAT CCC TGC GTG TCT CCG ACT CAG TAAGGAGAAC GGTGAT GCT AGT TAT TGC TCA GCG G

T7_term_IC_0103 CCA TCT CAT CCC TGC GTG TCT CCG ACT CAG AAGAGGATTC GGTGAT GCT AGT TAT TGC TCA GCG G

T7_term_IC_0104 CCA TCT CAT CCC TGC GTG TCT CCG ACT CAG TACCAAGATC GGTGAT GCT AGT TAT TGC TCA GCG G

T7_term_IC_0105 CCA TCT CAT CCC TGC GTG TCT CCG ACT CAG CAGAAGGAAC GGTGAT GCT AGT TAT TGC TCA GCG G

23S_+2004_IC_0106 CCA TCT CAT CCC TGC GTG TCT CCG ACT CAG CTGCAAGTTC GGTGAT AGG CTG TCT CCA CCC GAG

23S_+2004_IC_0108 CCA TCT CAT CCC TGC GTG TCT CCG ACT CAG TTCCGATAAC GGTGAT AGG CTG TCT CCA CCC GAG

23S_rRNA_-2053_P1 CCA CTA CGC CTC CGC TTT CCT CTC TAT GGG CAG TCG GTG ATG TTC ACG GGG TCT TTC CGT C

T7: TAA TAC GAC TCA CTA TAG GG

2.5.5. Nitrite reaction on nucleosides

Reaction on adenosine:

In a PCR tube was added 6 µL of 1 mM Adenosine (Sigma-Aldrich, A9251-1G), 37.5 µL milli-Q water and 1.5 µL acetic acid (Fisher Scientific, A38-212). Then, 45 µL of freshly-prepared 2 M sodium nitrite (Sigma-Aldrich, 237213-5G) was added, mixed thoroughly, and incubated on a thermal cycler (Biorad, T100) at 22 °C. After 12 h incubation, 10 µL

of 2 M TEAA was added to the reaction mixture and the resulting 100 µL solution was injected to HPLC.

Reaction on N^6-methyladenosine:

In a PCR tube was added 6 µL of 1 mM N^6-methyladenosine (Carbosynth, NM32281), 37.5 µL milli-Q water and 1.5 µL acetic acid (Fisher Scientific, A38-212). Then, 45 µL of freshly-prepared 2 M sodium nitrite (Sigma-Aldrich, 237213-5G) was added, mixed thoroughly, and incubated on a thermal cycler (Biorad, T100) at 22 °C. After 5 h incubation, 10 µL of 2 M TEAA was added to the reaction mixture and the resulting 100 µL solution was injected to HPLC.

2.5.6. HPLC analysis

High performance liquid chromatography (HPLC) analyses were conducted on a 1260 Infinity II LC System from Agilent.

HPLC method:

Flow rate: 0.5 mL/min

Detection wavelength: 260 nm

Column: Phenomenex, Luna® 5 µm C18(2) 100 Å, 250 x 4.6 mm

Mobile phase A: 3% acetonitrile in 0.1 M triethylammonium acetate (92:5:3 deionized water:2 M TEAA:acetonitrile)

Mobile phase B: 90% acetonitrile (9:1 acetonitrile:water)

Elapsed time (min)	%B
0	0
15	0
35	10
65	100
70	100
72	0
85	0

2.5.7. Optimisation of reaction conditions for deamination of DNA and RNA

A range of variables such as acid source and concentration, nitrite source, temperature, incubation time, solvent and pH values were tested to identify the optimal conditions. Conditions tested are summarized in Table 2-1.

Table 2-1: Tested conditions for deamination of DNA and RNA

Experiment	Acid source and final concentration	Nitrite source and final concentration	Incubation time (h)	Temp (°C)	Solvent
1	250 mM Acetate Buffer pH: 4.3	1 M $NaNO_2$	16	22	H_2O
2	200 mM Acetate Buffer pH: 4.3	1 M $NaNO_2$	5	25	H_2O
3	200 mM Acetate Buffer pH: 4.3	1 M $NaNO_2$	5	70	H_2O
4	43% AcOH	1 M $NaNO_2$	16	22	H_2O
5	33% AcOH	1 M $NaNO_2$	5	22	H_2O
6	23% AcOH	1 M $NaNO_2$	5	22	H_2O
7	13% AcOH	1 M $NaNO_2$	5	22	H_2O

8	7% AcOH	1 M $NaNO_2$	5	22	H_2O
9	5.3% AcOH	1 M $NaNO_2$	5	22	H_2O
10	4.3% AcOH	1 M $NaNO_2$	5	22	H_2O
11	3.3% AcOH	1 M $NaNO_2$	5	22	H_2O
12	3% AcOH	1 M $NaNO_2$	5	22	H_2O
13	2.7% AcOH	1 M $NaNO_2$	5	22	H_2O
14	2.3% AcOH	1 M $NaNO_2$	5	22	H_2O
15	2% AcOH	1 M $NaNO_2$	5	22	H_2O
16	1.7% AcOH	1 M $NaNO_2$	5	22	H_2O
17	1.3% AcOH	1 M $NaNO_2$	5	22	H_2O
18	1% AcOH	1 M $NaNO_2$	5	22	H_2O
19	0.3% AcOH	1 M $NaNO_2$	5	22	H_2O
20	5% AcOH	1 M $NaNO_2$	5	70	H_2O
21	5% AcOH	1 M $NaNO_2$	5	25	H_2O
22	10% AcOH	1 M $NaNO_2$	5	70	H_2O
23	10% AcOH	1 M $NaNO_2$	5	25	H_2O
24	15% AcOH	1 M $NaNO_2$	5	70	H_2O
25	15% AcOH	1 M $NaNO_2$	5	25	H_2O
26	20% AcOH	1 M $NaNO_2$	5	70	H_2O
27	20% AcOH	1 M $NaNO_2$	5	25	H_2O
28	25% AcOH	1.15 M $NaNO_2$	24	25	H_2O
29	20% AcOH	1.15 M $NaNO_2$	24	25	H_2O
30	80% AcOH	4 M amyl nitrite	16	25	H_2O

31	-	1 M amyl nitrite	16	25	MeOH/H_2O 50:50
32	-	1 M amyl nitrite	16	25	DMSO/H_2O 50:50
33	-	4 M amyl nitrite	16	60	H_2O
34	-	4 M amyl nitrite	16	25	H_2O
35	-	1 M amyl nitrite	16	60	DMSO/H_2O 50:50
36	-	1 M amyl nitrite	16	60	MeOH/H_2O 50:50

2.5.8. Nitrite sequencing for DNA

In a PCR tube was added 20 pmol (2 µL, 10 µM) of ssDNA or dsDNA, 12.3 µL milli-Q water and 0.7 µL acetic acid (Fisher Scientific, A38-212). Then, 15 µL of freshly-prepared 2 M sodium nitrite (Sigma-Aldrich, 237213-5G) was added, mixed thoroughly, and incubated on a thermal cycler (Biorad, T100) at 22 °C for 5 h. The reaction was then purified using E.Z.N.A. Cycle Pure Kit (Omega Bio-tek, D6492). The purified DNA was prepared for sequencing by PCR using IonCode adapters and Q5 High-Fidelity 2× Master Mix (New England Biolabs, M0492) using the following PCR cycles:

1. 95 ^{o}C, 30 s
2. 95 ^{o}C, 30 s
3. 50 ^{o}C, 30 s

4. 68 ^{o}C, 60 s
5. GOTO step 2 (variable)
6. 68 ^{o}C, 5 min
7. 4 ^{o}C, ∞

The amplified DNA was purified using E.Z.N.A. Cycle Pure Kit (Omega Bio-tek, D6492), and then purified using 10% native polyacrylamide gel. After staining the gel for 15 minutes with SYBR safe DNA gel stain (Invitrogen, 33100), the gel was visualized on BluPAD Dual LED Blue/White Light Transilluminator (bio-helix, BP001CU), and the desired DNA amplicon was excised from the gel. The excised band was crushed into a slurry, 100 µL of 0.3 M NaCl was added to the slurry and incubated overnight at 37 °C. The DNA was then purified from slurry using a CENTRI-SEP spin column (Princeton Separation, CS-901) pre-hydrated with milli-Q water. The concentration of the DNA was then measured using a Qubit 4.0 Fluorometer (Thermo Fisher Scientific) using the dsDNA HS Assay Kit (Invitrogen, Q32851) and then diluted to 50 pM. The prepped and pooled DNA libraries were loaded onto an Ion Chef with Ion 530 Chips (Thermo Fisher Scientific, A27764). The prepared chips were then sequenced on an Ion GeneStudio™ S5 Plus DNA sequencing system (Thermo Fisher Scientific).

Klenow extension procedure

temp1_unmeth GCT AGT TAT TGC TCA GCG GGC TCT TAC ATG GAT CGA TTA GCC ATG CCG TCA CGA ACC TTA GTA TGC CAC TGT CTC ACC TCC CTA TAG TGA GTC GTA TTA

temp1_6mA_pos63 GCT AGT TAT TGC TCA GCG GGC TCT TAC ATG GAT CGA TTA GCC ATG CCG TCA CGA ACC TTA GT/iN6MedA/ TGC CAC TGT CTC ACC TCC CTA TAG TGA GTC GTA TTA

Single stranded **temp1-unmeth** and **temp1-6mA-pos63** strands were converted to double stranded using the conditions below:

12 µL of 100 µM T7 primer, 10 µL of 100 µL ssDNA, 10 µL of 10X Klenow buffer and 0.5 µL H_2O were mixed and incubated at 95 ^oC for 5 minutes. The solution was then slow cooled to 25 ^oC (0.1 ^oC/s) and 8 µL of 10 mM dNTPs, 3 µL of 5000 U/mL Klenow exo- and 57 µl H_2O was added. The mixture was then incubated at 37 ^oC for 1 hour. The resulting dsDNA was then purified using E.Z.N.A. Cycle Pure Kit (Omega Bio-tek, D6492).

2.5.9. Nitrite sequencing for RNA

In a PCR tube was added 20 pmol (2 µL, 10 µM) of ssRNA, 11.5 µL nuclease free water (Ambion, AM9937) and 1.5 µL acetic acid (Fisher Scientific, A38-212). Then, 15 µL of freshly-prepared 2 M sodium nitrite (Sigma-Aldrich, 237213-5G) was added, mixed thoroughly, and incubated on a thermal cycler (Biorad, T100) at 22 °C for 5 h. The reaction was then purified using Monarch RNA cleanup kit (NEW ENGLAND BioLabs, T2030L). The purified RNA was prepared for sequencing by reverse transcription PCR using IonCode adapters and SuperScript III One-Step RT-PCR System with Platinum *Taq* DNA Polymerase (Invitrogen, Thermo Fisher Scientific, 12574-018) using the following PCR cycles:

1. 55 ^oC, 30 min
2. 94 ^oC, 2 min
3. 94 ^oC, 15 s
4. 50 ^oC, 30 s
5. 68 ^oC, 60 s

6. GOTO step 3 (9X)
7. 68 ^{o}C, 5 min
8. 4 ^{o}C, ∞

The reverse transcribed DNA was purified using E.Z.N.A. Cycle Pure Kit (Omega Bio-tek, D6492), and then purified using 10% native polyacrylamide gel. After staining the gel for 15 minutes with SYBR safe DNA gel stain (Invitrogen, 33100), the gel was visualized on BluPAD Dual LED Blue/White Light Transilluminator (bio-helix, BP001CU), and the desired DNA amplicon was excised from the gel. The excised band was crushed into a slurry, 100 µL of 0.3 M NaCl was added to the slurry and incubated overnight at 37 °C. The DNA was then purified from slurry using a CENTRI-SEP spin column (Princeton Separation, CS-901) pre-hydrated with milli-Q water. The concentration of the DNA was then measured using a Qubit 4.0 Fluorometer (Thermo Fisher Scientific) using the dsDNA HS Assay Kit (Invitrogen, Q32851) and then diluted to 50 pM. The prepped and pooled DNA libraries were loaded onto an Ion Chef with Ion 530 Chips (Thermo Fisher Scientific, A27764). The prepared chips were then sequenced on an Ion GeneStudio™ S5 Plus DNA sequencing system (Thermo Fisher Scientific).

2.5.10. Nitrite sequencing of *E. coli* 23S rRNA

rRNA was isolated from *E. coli* ribosome (NEW ENGLAND BioLabs, P0763S) using Monarch RNA cleanup kit (NEW ENGLAND BioLabs, T2030L). In a PCR tube was added 20 pmol (2 µL, 10 µM) of rRNA and 11.5 µL nuclease free water (Ambion, AM9937). The sample was denatured at 70 ^{o}C for 5 minutes and snap-cooled on ice. Then 1.5 µL acetic acid (Fisher Scientific, A38-212) and 15 µL of freshly-prepared 2 M sodium nitrite (Sigma-Aldrich, 237213-5G) were added, mixed thoroughly, and incubated on a thermal cycler

(Biorad, T100) at 22 ^{o}C for 5 h. The reaction was then purified using Monarch RNA cleanup kit (NEW ENGLAND BioLabs, T2030L). The purified RNA was prepared for sequencing by reverse transcription PCR using IonCode adapters and SuperScript III One-Step RT-PCR System with Platinum Taq DNA Polymerase (Invitrogen, Thermo Fisher Scientific, 12574-018) using the following PCR cycles:

1. 55 ^{o}C, 30 min
2. 94 ^{o}C, 2 min
3. 94 ^{o}C, 15 s
4. 50 ^{o}C, 30 s
5. 68 ^{o}C, 60 s
6. GOTO step 3 (12X)
7. 68 ^{o}C, 5 min
8. 4 ^{o}C, ∞

The reverse transcribed DNA was purified using E.Z.N.A. Cycle Pure Kit (Omega Bio-tek, D6492), and then purified using 10% native polyacrylamide gel. After staining the gel for 15 minutes with SYBR safe DNA gel stain (Invitrogen, 33100), the gel was visualized on BluPAD Dual LED Blue/White Light Transilluminator (bio-helix, BP001CU), and the desired DNA amplicon was excised from the gel. The excised band was crushed into a slurry, 100 µL of 0.3 M NaCl was added to the slurry and incubated overnight at 37 °C. The DNA was then purified from slurry using a CENTRI-SEP spin column (Princeton Separation, CS-901) pre-hydrated with milli-Q water.

Demethylation of rRNA

In a PCR tube was added 3 µL of 13.3 µM Ribosome (40 pmol) and 11.64 µL nuclease free water and incubated at 70 ^{o}C for 5 minutes, then snap-cooled on ice. 20 µL of *2X FTO buffer and 5.36 µL of 22.4 µM FTO (CAYMAN CHEMICAL, 26340) was added, and the mixture was incubated at 37 ^{o}C for 4 hours. The reaction was then purified using Monarch RNA cleanup kit (NEW ENGLAND BioLabs, T2030L). The purified RNA was then used for nitrite sequencing.

*2X FTO Buffer: 100 mM HEPES (Fisher Scientific, BP310) [pH 7.5], 600 µM α-Ketoglutaric acid (Sigma-Aldrich, 75890), 600 µM ammonium iron (II) sulfate hexahydrate (Sigma-Aldrich, 09719), and 4 mM L-Ascorbic acid (SIGMA Life Science, 95209).

2.5.11. Sequencing analysis

FastQ files generated from the IonTorrent system were trimmed to the expected length using the single-end read function Trimmomatic 0.36[64]. Bowtie 1[60] was used to build the template index and generate the map file for each experiment. Map files were analyzed for transitions and transversion at each nucleobase. Graphs were plotted from each adenosine as the ratio of the frequency of A → G transitions for the demethylated experiment over the frequency of A → G transitions for the methylated experiment.

Example of sequencing data

Sequence: GCT AGT TAT TGC TCA GCG GGC TCT TAC ATG GAT CGA TTA GCC ATG CCG TCA CGA ACC TTA GT/iN6MedA/ TGC CAC TGT CTC ACC TCC CTA TAG TGA GTC GTA TTA

Position (A)	A>G (methylated)	freq A>G (methylated)	A>G (demethylated)	freq A>G (demethylated)	**ratio**
26	577	0.008133634	181	0.002110221	**0.259444**
28	791	0.011150268	251	0.002926329	**0.262445**
32	1009	0.014223287	251	0.002926329	**0.205742**
36	2496	0.035184663	537	0.006260711	**0.177939**
39	241	0.003397237	47	0.000547958	**0.161295**
43	789	0.011122075	190	0.002215149	**0.199167**
51	772	0.010882436	207	0.002413347	**0.221765**
54	388	0.005469411	83	0.00096767	**0.176924**
55	1969	0.02775585	400	0.004663472	**0.168018**
60	426	0.006005075	198	0.002308419	**0.384411**
63	13	0.000183253	332	0.003870682	**21.12201**
68	1319	0.018593177	294	0.003427652	**0.184350**
76	449	0.006329292	143	0.001667191	**0.263409**

Chapter 3. Identification of Cytosine Modifications in DNA by Nitrite Sequencing

3.1. Abstract

Nitrite sequencing, initially designed for the detection of N^6-adenine methylation sites in DNA and RNA, was evaluated for its ability to detect cytosine modifications in DNA. It was demonstrated that this method could be used for the selective detection of the most reactive modified nucleoside (5-hydroxymethyl cytosine). However, the higher native error rates that was observed at 5hmdC sites compared to those of normal C sites is sufficient for distinguishing 5hmdC and C, without the need for nitrite treatment.

3.2. Introduction

It has long been recognized that cytosine can undergo chemical modifications at the fifth carbon of its pyrimidine ring. 5-methylcytosine (5mC) has been found to be the most abundant DNA modification in mammalian cells and it has important roles in development and disease[65]. 5mC in mammalian genomes plays essential roles in maintaining cellular function and genomic stability and is involved in biological processes including chromosome inactivation, genomic imprinting, and transposon silencing[66]. Methyl group is added at the carbon-5 position of the cytosine base primarily in cytosine-guanine dinucleotides (CpG) through the action of the DNA methyltransferase (DNMT) family of enzymes, and of ~28 million CpGs in the human genome, 60–80% are generally methylated[65]. It has previously been established that the methylation process is reversible and DNA demethylation could be achieved through the ten-eleven translocation (TET) protein mediated oxidization of 5mC with the generation of intermediates of 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC), and 5-carboxylcytosine (5caC)[67] (Figure 3-1). The identification of distinct reader proteins specific for 5hmC, 5fC

and 5caC and emerging experimental findings suggested that the oxidized derivatives of 5mdC should have individual functions besides serving as the intermediates in the active DNA demethylation pathway and may serve as stable epigenetic marks and possess unique regulatory functions[68].

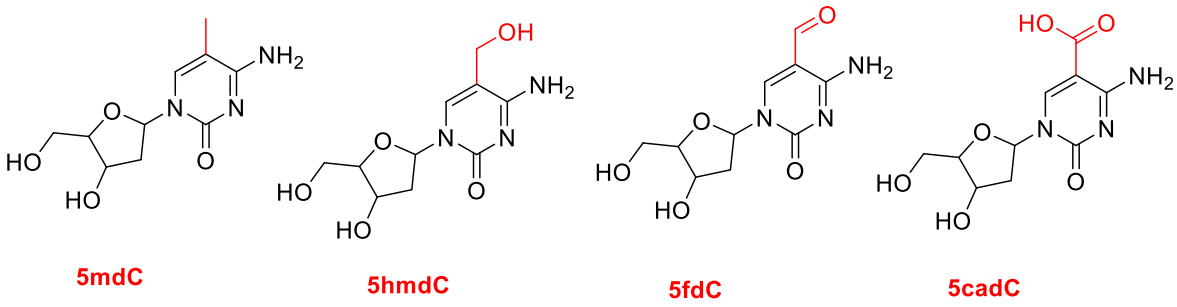


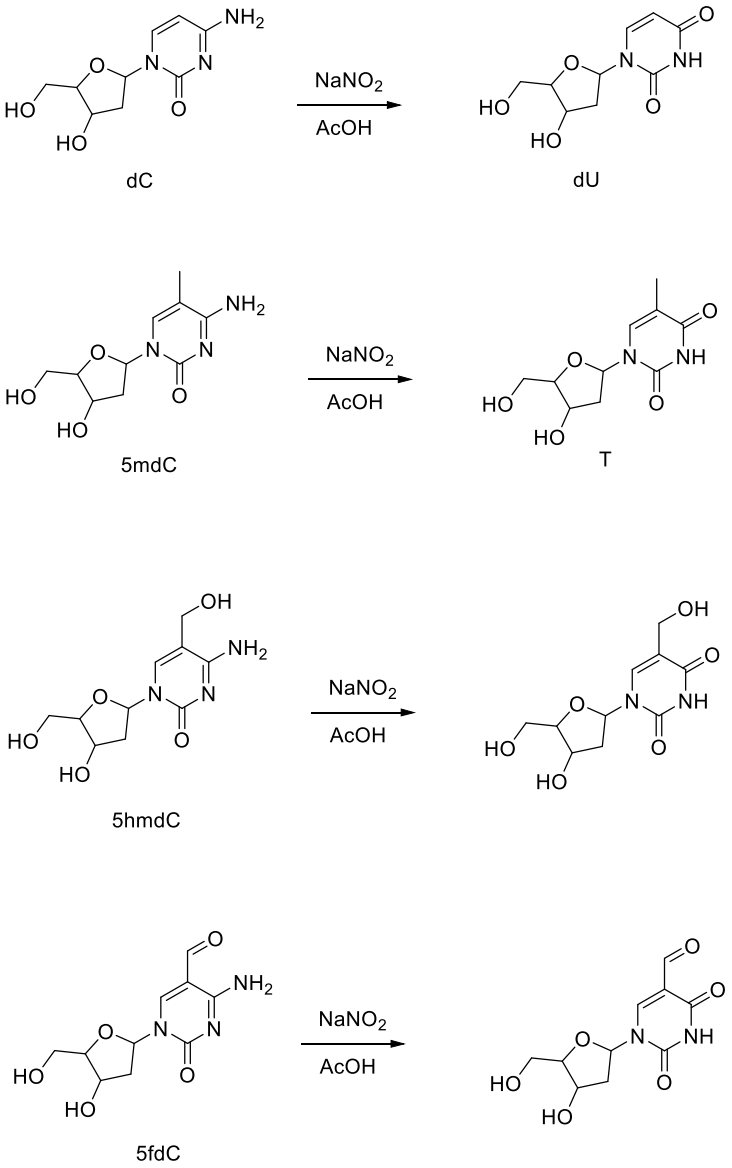
Figure 3-1: Chemical structures of modified cytosine derivatives found in mammalian DNA.

Nitrite sequencing was initially designed for the detection of m^6A and 6mA sites in RNA and DNA, which takes advantage of the chemoselective deamination of unmethylated adenines, without competing deamination of N^6-adenine sites. As cytosine bases contain an exocyclic amine, they will also be susceptible to nitrite mediated deamination resulting in C → T transitions (Figure 2-7, Figure 2-8 and Scheme 2-4). Therefore, it was speculated that nitrite mediated deamination with further optimization can potentially be used for the detection of C modifications, which can modulate the electron density of the cytosine base.

3.3. Results and discussion

Nitrite-mediated deamination process on free nucleosides was first examined, using 1 M aqueous $NaNO_2$ in the presence of AcOH. The deaminated products (Scheme 3-1) were observed by HPLC and confirmed by mass analysis. It was found that these

nucleosides deaminate with different rates under our optimised reaction conditions (1 M sodium nitrite, 2.3% acetic acid, incubation at 4 °C), with 5fdC being the least reactive nucleoside (Figure 3-2). This data suggested that nitrite sequencing could be used for selective detection of 5fdC over dC, 5hmdC, 5mdC and 5cadC.



Scheme 3-1: Reaction of dC, 5mdC, 5hmdC and 5fdC under nitrite treatment and their corresponding reaction products.

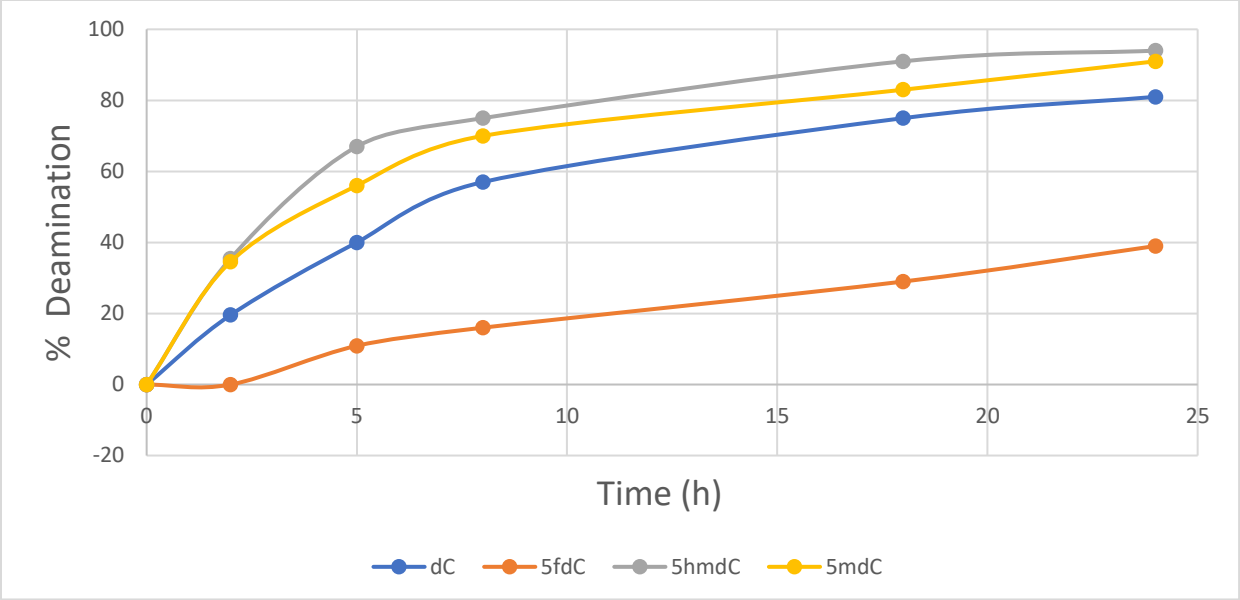


Figure 3-2: Percent deamination of 2'-deoxycytidine derivatives following treatment with 1 M sodium nitrite in the presence of 2.3% acetic acid at 4 °C at varying incubation times.

To test nitrite conditions on DNA, the 9mer DNA strand C(5mdC)(5hmdC)(5mdC)(5cadC)(5fdC)(5cadC)(5hmdC)C was first synthesized and used to optimise nitrite conditions for DNA. It was observed that nitrite-mediated deamination is slower on DNA compared to the free nucleosides. After testing several reaction conditions, digestion, and HPLC analysis, it was found that incubation of DNA with 1M $NaNO_2$ and 2.3% AcOH at 4 °C for 19 h provides the optimal condition in which dC, 5mdC, 5hmdC and 5cadC undergo partial deamination while leaving 5fdC unaffected by the process (Figure 3-3).

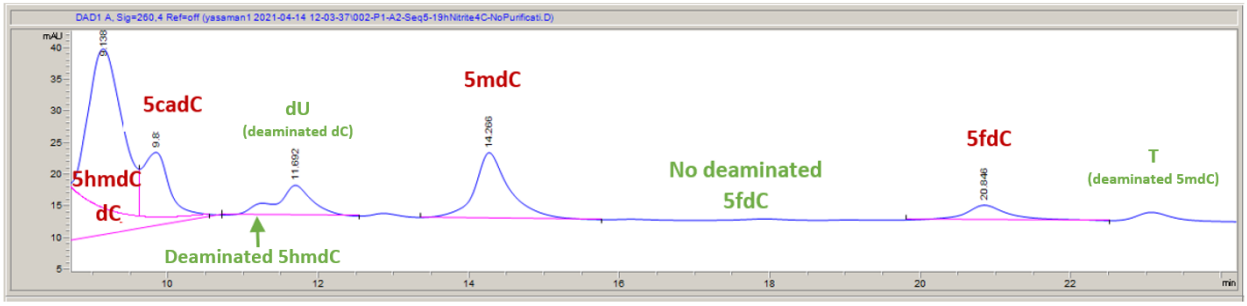
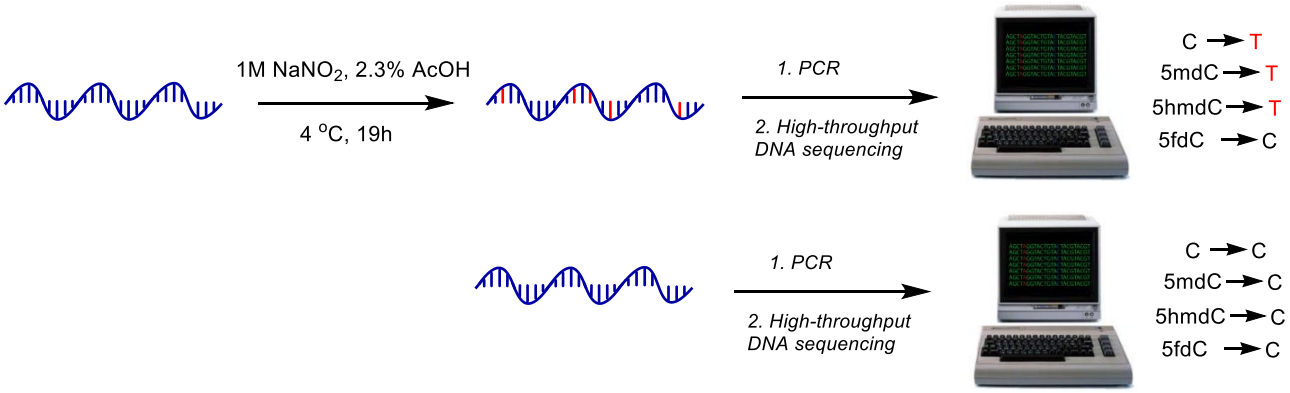


Figure 3-3: HPLC analysis after treatment of the 9mer DNA C(5mdC)(5hmdC)(5mdC)(5cadC)(5fdC)(5cadC)(5hmdC)C with 1M $NaNO_2$ and 2.3% AcOH at 4 °C for 19 h followed by enzymatic digestion.

5fdC, 5mdC, 5hmdC and 5cadC are known to be read as C by polymerases. The deaminated nucleosides, however, are expected to be read differently due to the change in the hydrogen-bonding relationship on the Watson-Crick face. Thus, it was proposed that amplification after nitrite treatment and comparative sequencing analysis against a no-reaction control can be used to readily identify the locations of 5fdC within a DNA sequence (Scheme 3-2).



Scheme 3-2: Proposed nitrite sequencing workflow for the detection of 5fdC.

Encouraged by initial findings, the following DNA strands containing the 2'-deoxycytidine modifications were synthesized:

Sequence1 5’-GCTAGTTATTGCTCAGCGGAGACTA(5fC)GGAAGGT(C)GACTCGCCCTATA GTGAGTCGTATTA-3’

Sequence2 5’-GCTAGTTATTGCTCAGCGGAGTGTA(5fC)GGAAGGT(5mC)GACTCGCCCT ATAGTGAGTCGTATTA-3’

Sequence3 5’-GCTAGTTATTGCTCAGCGGATCATA(5fC)GGAAGGT(5hmC)GACTCGCCCT ATAGTGAGTCGTATTA-3’

Sequence4 5’-GCTAGTTATTGCTCAGCGGATTATA(5fC)GGAAGGT(5caC)GACTCGCCCTA TAGTGAGTCGTATTA-3’

The 60mer sequences 1-4 were subjected to the optimised nitrite-mediated deamination conditions. Nitrite-treated and untreated sequences were prepared for sequencing by PCR using IonCode adapters and were submitted for sequencing. The sequencing results were not able to distinguish 5fdC from other C modifications. This might be due to the fact that these conditions were optimized on a 9mer and the reaction is probably slower on longer DNA strands. However, the data suggested that the optimised conditions could be used for the selective detection of the most reactive nucleoside (5hmdC). Encouraged by these findings, nitrite sequencing method was then tested on the following DNA strands:

Sequence5-hmdC-pos57 5’-GCTAGTTATTGCTCAGCGGTTACCTCCCTACCCTTCTCC AAGCAGTAGTCACTCAG(5hmC)GCCAACCCCTCTCAGGCGTCGGCCCTATAGTGAGTCGT ATA-3’

Sequence5-C-pos57 5’-GCTAGTTATTGCTCAGCGGTTACCTCCCTACCCTTCTCCAAGCA GTAGTCACTCAG(C)GCCAACCCCTCTCAGGCGTCGGCCCTATAGTGAGTCGTATA-3’

Sequence6-hmdC-pos32 5’-GCTAGTTATTGCTCAGCGGGCGCCGGAGCAG(5hmC)GTGGGAAAGAAGGGAAGAGTGTCGTTAAGTTTACGGCCAACGGTGGACCCTATAGTGAGTCGTATTA-3’

Sequence7-hmdC-pos59 5’-GCTAGTTATTGCTCAGCGGGGAGGCGCTGAAGTCGGGGCCCGCCCTGTGGGCCCCGCC(5hmC)GGCCCGCGCTTGCTAGGGCCCCCTATAGTGAGTCGTATTA-3’

Sequence8-hmdC-pos60 5’-GCTAGTTATTGCTCAGCGGCACTATCCCATTCCTAACCAGGCTCGGCTTCCGATCAGAC(5hmC)GCGGACGTAGCACATGGTCCCCTATAGTGAGTCGTATTA-3’

Sequence9-hmdC65-mdC73 5’-GCTAGTTATTGCTCAGCGGGAAGGTCGACTCGGTCGTTAAGTTACGGCCAACGGTGGAAGACTA(5hmC)GGAAGGT(5mC)GACTCGCCCTATAGTGAGTCGTATTA-3’

First, a spike-in experiment was performed that assessed the response for varying fractions of 5hmC at nucleotide 57 in sequence5. It was observed that the %T/(C+T) at position 57 is proportional to 5hmC content at this position (Figure 3-4).

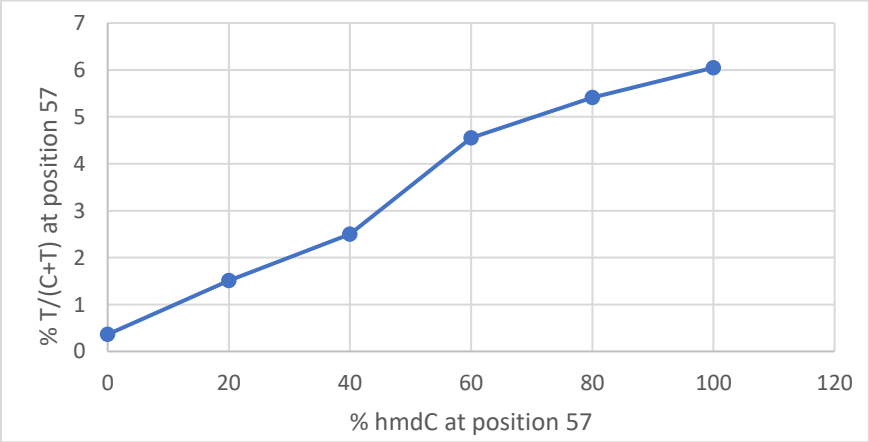


Figure 3-4: Dependence of %T/(C+T) to 5hmC content within a DNA sequence.

Sequencing analysis after nitrite treatment and amplification for all the sequences showed that the % T/(C+T) at the hmdC position was higher than the positions with normal C. However, unexpectedly, it was observed that this difference is even greater in the absence of nitrite reaction (Table 3-1, Table 3-2, Table 3-3, Table 3-4 and Table 3-5).

Table 3-1: Sequencing data for DNA Sequence5-hmdC-pos57 (The DNA sequence contains a single 5hmC site at position 57, cytosines in the primer regions are not shown)

		After nitrite reaction					No reaction				
Position	Base	A	C	G	T	% T/(C+T)	A	C	G	T	% T/(C+T)
23	C	351	10407202	65	61819	0.590495	467	14699860	25	2901	0.019730988
24	C	5400	10390093	71	73873	0.705975	4756	14695943	48	2506	0.017049418
26	C	1447	10439920	227	27843	0.265988	3444	14697376	705	1728	0.011755819
27	C	1325	10439566	70	28476	0.272028	1343	14700646	65	1199	0.008155439
28	C	14448	10380608	531	73850	0.706397	13882	14687213	623	1535	0.010450176
31	C	389	10420274	315	48459	0.462893	632	14700392	370	1859	0.012644322
32	C	796	10445705	82	22854	0.218311	897	14701135	52	1169	0.007951135
33	C	1320	10450520	341	17256	0.164849	890	14700778	550	1035	0.007039948
36	C	2182	10333865	304	133086	1.271488	3350	14698128	570	1205	0.008197651
38	C	1355	10449386	68	18628	0.177952	1250	14692508	66	9429	0.064134406
39	C	2095	10446320	126	20896	0.199633	2290	14699204	102	1657	0.011271449
43	C	1213	10409903	92	58229	0.55625	1406	14699666	93	2088	0.014202387
50	C	8806	10441350	63	19218	0.183719	13075	14687928	85	2165	0.014737824
52	C	6508	10452055	50	10824	0.103451	10597	14690649	27	1980	0.013476145
54	C	891	10413775	113	54658	0.522122	823	14700776	114	1540	0.01047454
57	hmdC	654	9534289	1369	933125	8.91457	793	14427523	2189	272748	1.855394367
59	C	1177	10443792	1020	23448	0.224013	1610	14699079	1316	1248	0.008489607
60	C	2454	10454624	85	12274	0.117265	3561	14698180	71	1441	0.009802974
63	C	1868	10451802	88	15679	0.149788	2610	14697856	42	2745	0.018672706
64	C	524	10456188	115	12610	0.120453	520	14700943	42	1748	0.01188898
65	C	1625	10459901	77	7834	0.074839	1569	14700657	61	966	0.006570703

66	C	7946	10438453	232	22806	0.218004	6738	14691357	209	4949	0.033675129
68	C	3995	10403082	587	61773	0.59029	4309	14695589	945	2410	0.01639679
70	C	2946	10392964	420	73107	0.698514	2620	14696944	457	3232	0.021986131
74	C	927	10436902	920	30688	0.293172	1139	14697594	1066	3454	0.023494924
77	C	5880	10445415	947	17195	0.164347	8453	14687569	1804	5427	0.036935966

Table 3-2: Sequencing data for DNA Sequence6-hmdC-pos32 (The DNA sequence contains a single 5hmC site at position 32, cytosines in the primer regions are not shown)

		After nitrite reaction					No reaction				
Position	**Base**	**A**	**C**	**G**	**T**	**% T/(C+T)**	**A**	**C**	**G**	**T**	**% T/(C+T)**
21	C	132	7704317	383	243543	3.064258807	211	8379191	539	8357	0.099635793
23	C	222	7711778	1012	235363	2.961605941	252	8384585	1575	1886	0.022488601
24	C	298	7870424	534	77119	0.970350208	417	8384682	326	2873	0.034253129
29	C	657	7898323	593	48802	0.61408371	883	8380133	491	6791	0.080971283
32	hmdC	624	6819241	1379	1127131	14.18422143	430	8249323	1476	137069	1.634421572
55	C	7290	7914676	319	26090	0.328557724	13152	8372041	561	2544	0.030377625
66	C	607	7911473	124	36171	0.455116007	522	8363151	174	24451	0.291513594
69	C	256	7909205	481	38433	0.483577637	247	8385917	475	1659	0.019779254
70	C	6857	7906155	116	35247	0.443838506	4409	8382316	96	1477	0.017617324
73	C	454	7938889	146	8886	0.111804876	789	8381619	176	5714	0.068126543

Table 3-3: Sequencing data for DNA Sequence7-hmdC-pos59 (The DNA sequence contains a single 5hmC site at position 59, cytosines in the primer regions are not shown)

		After nitrite reaction					No reaction				
Position	**Base**	**A**	**C**	**G**	**T**	**% T/(C+T)**	**A**	**C**	**G**	**T**	**% T/(C+T)**
25	C	153	2877343	44	14664	0.507052715	351	5119570	56	2438	0.04759852
27	C	92	2873700	13	18399	0.636181542	95	5121820	24	476	0.009292708
34	C	1199	2871596	794	18615	0.644070623	3663	5113799	1723	3230	0.063122566
39	C	47	2884981	316	6860	0.237219128	101	5121057	368	889	0.017356684
40	C	41	2879979	135	12049	0.416628055	73	5121570	29	743	0.014505166

41	C	92	2872461	881	18770	0.649204439	158	5119752	1533	972	0.018981691
43	C	24	2880292	549	11339	0.392131638	32	5120087	1167	1129	0.022045545
44	C	63	2853243	71	38827	1.342533203	122	5121446	83	764	0.014915437
45	C	138	2833439	42	58585	2.025743908	224	5120877	34	1280	0.024989472
52	C	138	2882312	1250	8504	0.294172995	318	5117139	2543	2415	0.047172078
53	C	52	2884601	22	7529	0.260327164	73	5121076	35	1231	0.02403214
54	C	139	2884731	70	7264	0.251176091	221	5121259	63	872	0.017024164
55	C	365	2880680	1788	9371	0.324250333	1187	5116017	3994	1217	0.023782379
57	C	39	2876612	160	15393	0.532260491	65	5121144	370	836	0.016321813
58	C	71	2879486	18	12629	0.436670049	130	5121159	19	1107	0.021611529
59	hmdC	610	2688395	166	203033	7.02189368	1770	5019338	454	100853	1.969711677
62	C	180	2887033	481	4510	0.155972088	414	5120007	1249	745	0.014548644
63	C	95	2886100	541	5468	0.189101553	335	5120368	846	866	0.016909987
64	C	444	2879014	636	12110	0.418868233	1480	5117185	1910	1840	0.035944345
66	C	89	2873116	315	18684	0.646102773	232	5119461	1467	1255	0.024508291
68	C	41	2880961	448	10754	0.371890038	80	5109297	1728	11310	0.220872252
72	C	125	2874146	1054	16879	0.583841371	191	5117135	2633	2456	0.047972582
78	C	190	2866989	10419	14606	0.506872062	280	5093659	24114	4362	0.085562613
79	C	67	2885472	304	6361	0.219964293	147	5116368	1257	4643	0.090665691

Table 3-4: Sequencing data for DNA Sequence8-hmdC-pos60 (The DNA sequence contains a single 5hmC site at position 60, cytosines in the primer regions are not shown)

		After nitrite reaction					No reaction				
Position	**Base**	**A**	**C**	**G**	**T**	**% T/(C+T)**	**A**	**C**	**G**	**T**	**% T/(C+T)**
20	C	120	5434512	4257	6619	0.121647503	59	5921855	5790	123	0.002077009
22	C	301	5420318	6175	18714	0.344068577	81	5919599	7457	690	0.011654836
26	C	1229	5416984	122	27173	0.49912227	1287	5925594	123	823	0.013886974
27	C	187	5440684	1899	2738	0.050299242	86	5925964	1667	110	0.001856204
28	C	3529	5437099	21	4859	0.089287716	2970	5924679	11	167	0.002818639
32	C	5731	5414474	59	25244	0.464068174	6520	5920965	9	333	0.005623767
33	C	92	5410167	6075	29174	0.536351738	30	5920542	6412	843	0.014236534
37	C	167	5429390	9932	6019	0.110736837	169	5921617	5622	419	0.007075269
38	C	1619	5424463	8867	10559	0.194277042	1486	5917686	6704	1951	0.032958102
42	C	329	5412481	267	32431	0.595620278	279	5926238	183	1127	0.019013508

44	C	263	5319477	191	125577	2.30625812	268	5924194	137	3228	0.054458751
47	C	120	5380033	649	64706	1.188413255	82	5918763	567	8415	0.141973128
50	C	553	5425868	5907	13180	0.242321818	601	5918827	6551	1848	0.031212657
51	C	1034	5431231	9167	4076	0.074991164	859	5912738	13791	439	0.007424097
55	C	4359	5396155	315	44679	0.821179253	6591	5919492	228	1516	0.025603749
59	C	10434	5416458	520	18096	0.332980406	14440	5907653	444	5290	0.089464756
60	hmdC	1067	5204451	2551	237439	4.363171619	1111	5773853	2950	149913	2.530704285
62	C	990	5399301	3984	41233	0.757885163	1491	5909775	8236	8325	0.140670148
66	C	2894	5414129	709	27776	0.510409498	5936	5917996	1221	2674	0.045163807
71	C	707	5370038	861	73902	1.357509451	713	5923945	1125	2044	0.034492133
73	C	648	5425624	37	19199	0.352610177	939	5925953	65	870	0.014679028
79	C	253	5405482	275	39498	0.725402114	214	5922544	416	4653	0.078502537

Table 3-5: Sequencing data for DNA Sequence9-hmdC65-mdC73 (The DNA sequence contains a 5hmC site at position 65 and a 5mC site at position 73, cytosines in the primer regions are not shown)

		After nitrite reaction					**No reaction**				
Position	**Base**	**A**	**C**	**G**	**T**	**% T/(C+T)**	**A**	**C**	**G**	**T**	**% T/(C+T)**
26	C	488	2622486	214	4219	0.160619483	416	2762757	245	830	0.030033431
29	C	219	2621229	140	5819	0.221503376	447	2762587	83	1131	0.040923133
31	C	125122	2485014	517	16754	0.669686398	118500	2643464	731	1553	0.058714178
35	C	17661	2381740	215820	12186	0.509038291	19667	2447839	292275	4467	0.182155082
45	C	208	2610250	142	16807	0.639765334	185	2758321	185	5557	0.201058079
48	C	105	2624464	190	2648	0.100795094	84	2763687	160	317	0.011468869
49	C	992	2621836	53	4526	0.172329633	1061	2762342	74	771	0.027903311
52	C	304	2590839	214	36050	1.372345767	238	2760390	224	3396	0.122874926
62	C	309	2624746	90	2262	0.086105562	314	2763562	23	349	0.012627035
65	hmdC	794	2407881	62	218670	8.325366612	952	2714018	97	49181	1.779857332
73	mdC	7315	2532254	632	87206	3.329159445	7779	2670517	757	85195	3.09157851
76	C	111	2625305	87	1904	0.072472346	164	2762289	110	1685	0.060962947
78	C	2885	2621473	314	2735	0.104221921	3534	2759022	558	1134	0.041084634

For sequence5, the ratio (R) of the % T/(C+T) at each C position compared to that of the dehydroxymethylated sequence was also calculated (Table 3-6):

$$R = \frac{\%\ \frac{T}{C+T}\ dehydroxymethylated\ control}{\%\ \frac{T}{C+T}\ hydroxymethylated\ sample}$$

This afforded a convenient way to visualise the nitrite sequencing data, as high R values are only observed at the hmdC site.

Table 3-6: Sequencing data for DNA Sequence5-hmdC-pos57 (The DNA sequence contains a single 5hmC site at position 57) and DNA Sequence5-C-pos57, and calculated R values

		Sequence5-hmdC-pos57 after nitrite reaction						Sequence5-C-pos57 after nitrite reaction					
Position	Base	A	C	G	T	% T/(C+T)	Base	A	C	G	T	% T/(C+T)	R
23	C	379	7094567	45	29596	0.42	C	757	10675365	157	30223	0.28	1.47
24	C	7885	7080598	47	36057	0.51	C	9983	10657778	152	38589	0.36	1.40
26	C	1750	7108156	121	14560	0.20	C	7006	10681853	232	17411	0.16	1.26
27	C	1666	7107041	70	15810	0.22	C	2254	10686607	162	17479	0.16	1.36
28	C	14796	7076032	335	33424	0.47	C	21222	10645128	284	39868	0.37	1.26
31	C	396	7098514	213	25464	0.36	C	800	10674487	119	31096	0.29	1.23
32	C	1198	7111434	61	11894	0.17	C	1907	10692012	153	12430	0.12	1.44
33	C	1543	7113592	267	9185	0.13	C	2171	10692486	203	11642	0.11	1.19
36	C	1690	7051552	207	71138	1.00	C	3696	10606965	119	95722	0.89	1.12
38	C	1649	7110860	59	12019	0.17	C	2172	10686425	125	17780	0.17	1.02
39	C	2112	7110915	88	11472	0.16	C	3222	10689464	249	13567	0.13	1.27
43	C	1053	7089486	66	33982	0.48	C	1462	10661014	298	43728	0.41	1.17
50	C	13850	7102921	105	7711	0.11	C	19251	10677080	225	9946	0.09	1.17
52	C	8472	7111167	38	4910	0.07	C	12103	10688700	105	5594	0.05	1.32
54	C	1498	7093791	112	29186	0.41	C	1883	10671631	217	32771	0.31	1.34
57	hmdC	520	6690360	3032	430675	6.05	C	770	10666203	685	38844	0.36	16.67
59	C	1511	7112550	1049	9477	0.13	C	2180	10690747	685	12890	0.12	1.10
60	C	4663	7114722	35	5167	0.07	C	3730	10696934	122	5716	0.05	1.36
63	C	2058	7114954	41	7534	0.11	C	2856	10696319	128	7199	0.07	1.57

64	C	872	7117547	75	6093	0.09	C	1341	10699342	172	5647	0.05	1.62
65	C	1789	7119278	85	3435	0.05	C	2197	10700421	135	3749	0.04	1.38
66	C	7213	7107580	47	9747	0.14	C	10783	10684833	145	10741	0.10	1.36
68	C	3536	7097087	377	23587	0.33	C	5305	10666592	261	34344	0.32	1.03
70	C	6547	7087551	125	30364	0.43	C	6892	10654083	395	45132	0.42	1.01
74	C	737	7109241	512	14097	0.20	C	1183	10688182	589	16548	0.15	1.28
77	C	8530	7107338	653	8066	0.11	C	14070	10681251	887	10294	0.10	1.18

3.4. Conclusions

Nitrite sequencing that was initially designed for the detection of N^6-adenine methylation sites in DNA and RNA, with minor modifications can be adapted to distinguish hydroxymethyl cytosine sites from normal cytosines in various sequence contexts in DNA. However, it was observed that the native error rates at 5hmdC sites are significantly higher than those of C, and this alone could enable discrimination between 5hmdC and C, without the need for nitrite treatment. Exploration of native error rates for 5hmdC sequencing is on-going.

3.5. Experimental details

3.5.1. General information

Water was purified with the MilliQ Direct Q3. DNA oligonucleotides 1-4 and the 9mer were synthesized on DNA/RNA synthesizer (Model 394, Applied Biosystems), followed by HPLC purification. DNA oligonucleotides 5-9 were purchased from Integrated DNA Technologies. HPLC analysis on nucleosides and oligonucleotide purifications were performed by reverse-phase high-performance liquid chromatography (HPLC, Agilent 1260 Infinity II). Oligonucleotide concentrations were determined using a Qubit 4.0

Fluorometer. High-throughput DNA sequencing samples were quantified using a Qubit 4 Fluorometer, prepared on an IonChef instrument and sequenced on an Ion Torrent GeneStudio S5 Plus using Ion 540 Chips.

3.5.2. IonCode adapters

T7_P1 CCA CTA CGC CTC CGC TTT CCT CTC TAT GGG CAG TCG GTG ATT AAT ACG ACT CAC TAT AGG G

T7_term_IC_0101 CCA TCT CAT CCC TGC GTG TCT CCG ACT CAG CTAAGGTAAC GGTGAT GCT AGT TAT TGC TCA GCG G

T7_term_IC_0102 CCA TCT CAT CCC TGC GTG TCT CCG ACT CAG TAAGGAGAAC GGTGAT GCT AGT TAT TGC TCA GCG G

T7_term_IC_0103 CCA TCT CAT CCC TGC GTG TCT CCG ACT CAG AAGAGGATTC GGTGAT GCT AGT TAT TGC TCA GCG G

T7_term_IC_0104 CCA TCT CAT CCC TGC GTG TCT CCG ACT CAG TACCAAGATC GGTGAT GCT AGT TAT TGC TCA GCG G

T7_term_IC_0105 CCA TCT CAT CCC TGC GTG TCT CCG ACT CAG CAGAAGGAAC GGTGAT GCT AGT TAT TGC TCA GCG G

T7_term_IC_0106 CCA TCT CAT CCC TGC GTG TCT CCG ACT CAG CTGCAAGTTC GGTGAT GCT AGT TAT TGC TCA GCG G

T7_term_IC_0107 CCA TCT CAT CCC TGC GTG TCT CCG ACT CAG TTCGTGATTC GGTGAT GCT AGT TAT TGC TCA GCG G

T7_term_IC_0108 CCA TCT CAT CCC TGC GTG TCT CCG ACT CAG TTCCGATAAC GGTGAT GCT AGT TAT TGC TCA GCG G

3.5.3. Oligonucleotide synthesis

Sequences 1-4 and the 9mer were synthesized on DNA synthesizer (Model 394, Applied Biosystems) using a DMT-ON protocol on a 0.2 µmol scale. Following synthesis, the beads were transferred to a vial and 1 mL of 0.4 M NaOH in MeOH/water 4:1 (v/v) was added to them and incubated at room temperature for 17 hours. The vial was then briefly sonicated to breakup the beads and the supernatant was transferred to a clean vial. The beads were rinsed with 250 uL water and the water was combined with the cleaved oligos. 250 uL of 2 M TEAA was added to the cleaved oligos and the sample was concentrated under reduced pressure using speedvac.The residue was then taken up into 100 µL of water and purified using the following HPLC method:

Flow rate: 4 mL/min

Detection wavelength: 260 nm

Column: Agilent 5 µm Eclipse XDB-C18 9.4 x 250 mm

Mobile phase A: 0.1 M triethylammonium acetate (TEAA)

Mobile phase B: Acetonitrile

Elapsed time (min)	%B
0	10
10	20
20	80
25	80
28	10
30	10

The purified oligonucleotides were then lyophilized and the residues were incubated at room temperature in 1 mL of 80% aqueous acetic acid for 6 hours, and then frozen and lyophilized. The dried product was dissolved into 100 µL water and subjected to HPLC purification using the following HPLC method:

Flow rate: 4 mL/min

Detection wavelength: 260 nm

Column: Agilent 5 µm Eclipse XDB-C18 9.4 x 250 mm

Mobile phase A: 0.1 M triethylammonium acetate (TEAA)

Mobile phase B: Acetonitrile

Elapsed time (min)	%B
0	10
10	20
23	45
26	80
28	80
29	10
31	10

3.5.4. Nitrite reaction on nucleosides

In PCR tubes were added 6 µL of 1 mM nucleoside (6 nmol), 36.9 µL milli-Q water and 2.1 µL acetic acid (Fisher Scientific, A38-212). Then, 45 µL of freshly-prepared 2 M sodium nitrite (Sigma-Aldrich, 237213-5G) was added, mixed thoroughly, and incubated on a thermal cycler (Biorad, T100) at 4 °C. After 2 h, 5 h, 8 h, 18 h, 24 h incubation, 10 µL of 2 M TEAA was added to the reaction mixture and the resulting 100 µL solution was injected to HPLC.

3.5.5. DNA enzymatic digestion

DNA (in water) was mixed with 1 µL of *stock solution and 10 µL of **buffer to the final volume of 40 µL and incubated at 37 ^{o}C for 14 hours.

*stock solution: (Benzonase® Nuclease, 0.1 U·µL-1; Phosphodiesterase I from *Crotalus adamanteu s venom*, 0.08 mU·µL-1 and Antarctic Phosphatase, 0.08 U·µL-1)

**buffer: (Tris-HCl, 20 mM, pH = 8, [$MgCl_2$] = 20 mM, [NaCl] =100 mM)

The digested sample was prepared for HPLC analysis using the following procedure:

1. 5.7 µL of 3 M sodium acetate and 143 µL of ethanol was added to the 40 µL sample

2. Sample was vortexed and chilled on dry ice for 10 min, then centrifuged at 22 ^{o}C at full speed (25000 x g) for 5 min

3. Supernatant was transferred to a new tube and the original tube containing the pellets was discarded

4. 429 µL ethanol was added to the tube containing the supernatant, Vortexed and chilled on dry ice for 10 min, then centrifuged at 22 ^{o}C at full speed (25000 x g), for 5 min

5. Supernatant was removed with a pipet, being careful not to disturb the pellet, transferred to a new tube, evaporated to complete dryness and then dissolved in 100 µL water and injected to HPLC

3.5.6. HPLC method for analysis of nucleosides

Flow rate: 0.5 mL/min

Detection wavelength: 260 nm

Column: Phenomenex, Luna® 5 µm C18(2) 100 Å, 250 x 4.6 mm

Mobile phase A: 3% acetonitrile in 0.1 M triethylammonium acetate (92:5:3 deionized water:2 M TEAA:acetonitrile)

Mobile phase B: 90% acetonitrile (9:1 acetonitrile:water)

Elapsed time (min)	%B
0	0
15	0
35	10
65	100
70	100
72	0
85	0

3.5.7. Nitrite sequencing for DNA

In a PCR tube was added 20 pmol (2 µL, 10 µM) of DNA , 12.3 µL milli-Q water and 0.7 µL acetic acid (Fisher Scientific, A38-212). Then, 15 µL of freshly-prepared 2 M sodium nitrite (Sigma-Aldrich, 237213-5G) was added, mixed thoroughly, and incubated on a thermal cycler (Biorad, T100) at 4 °C for 19 h. The reaction was then purified using Monarch PCR and DNA cleanup kit (NEW ENGLAND BioLabs, T1030S). The purified DNA was prepared for sequencing by PCR using IonCode adapters and Q5 High-Fidelity 2x Master Mix (New England Biolabs, M0492) using the following PCR cycles:

1. 95 ºC, 30 s
2. 95 ºC, 30 s
3. 50 ºC, 30 s
4. 68 ºC, 60 s
5. GOTO step 2 (x14)

6. 68 °C, 5 min
7. 4 °C, ∞

The amplified DNA was purified using E.Z.N.A. Cycle Pure Kit (Omega Bio-tek, D6492), and then purified using 10% native polyacrylamide gel. After staining the gel for 15 minutes with SYBR safe DNA gel stain (Invitrogen, 33100), the gel was visualized on BluPAD Dual LED Blue/White Light Transilluminator (bio-helix, BP001CU), and the desired DNA amplicon was excised from the gel. The excised band was crushed into a slurry, 100 µL of 0.3 M NaCl was added to the slurry and incubated overnight at 37 °C. The DNA was then purified from slurry using E.Z.N.A. Cycle Pure Kit (Omega Bio-tek, D6492). The concentration of the DNA was then measured using a Qubit 4.0 Fluorometer (Thermo Fisher Scientific) using the dsDNA HS Assay Kit (Invitrogen, Q32851) and then diluted to 50 pM. The prepped and pooled DNA libraries were loaded onto an Ion Chef with Ion 540 Chips (Thermo Fisher Scientific, A27764). The prepared chips were then sequenced on an Ion GeneStudio™ S5 Plus DNA sequencing system (Thermo Fisher Scientific).

3.5.8. Sequencing analysis

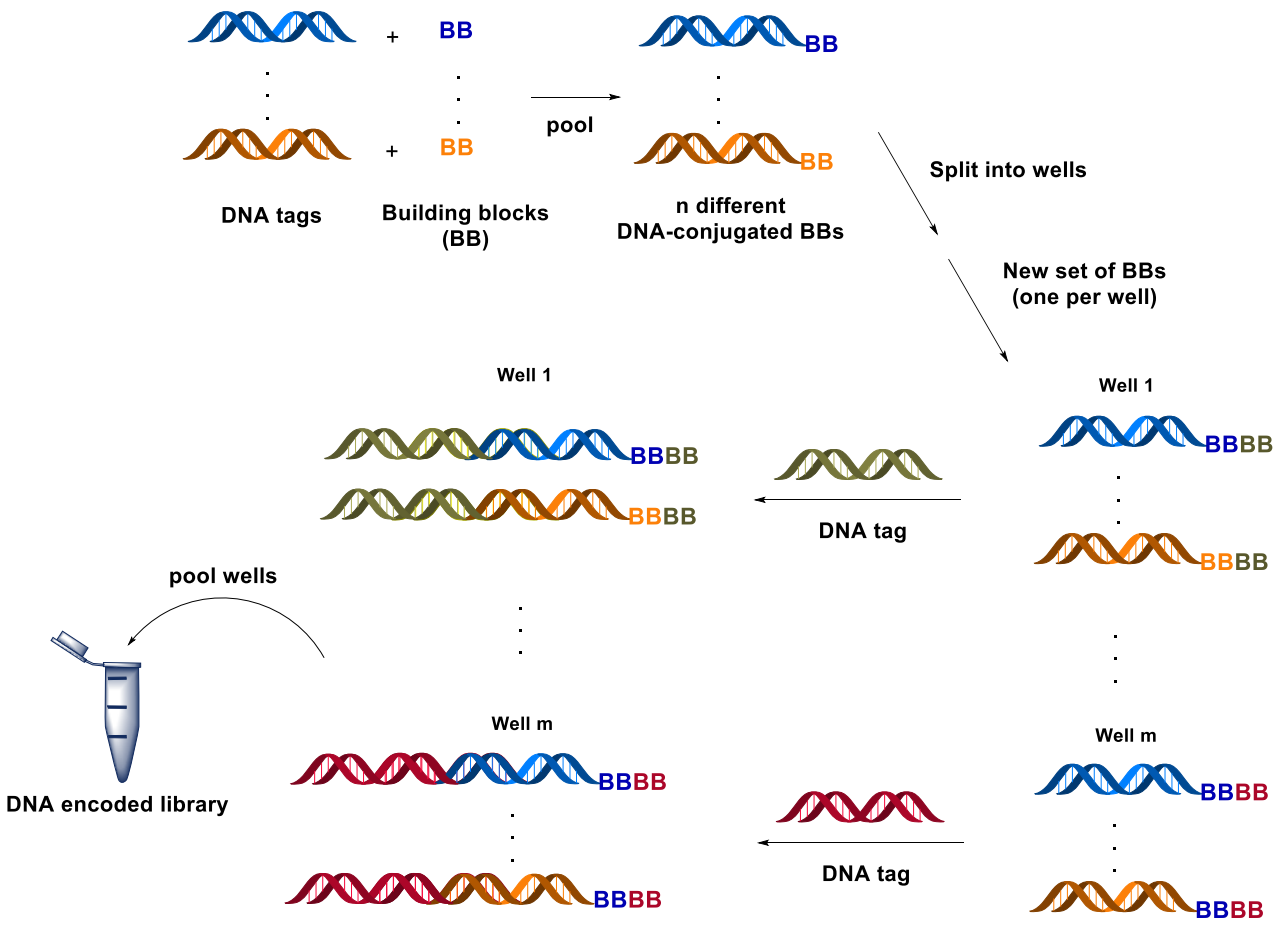
Following high-throughput sequencing, fastQ files generated from the IonTorrent system were trimmed for length and quality and aligned to the reference sequence using Bowtie 1[60] to generate the map file for each experiment. Map files were then analyzed for transitions at each nucleobase using snp-counter python program[69].

Section 2

Chapter 4. Introduction to DNA-Encoded Libraries

4.1. DNA-encoded libraries (DELs)

DNA-encoded libraries (DELs) comprise large libraries of small molecules, routinely >10^6 members, that are individually barcoded with a unique DNA sequence. This concept was first introduced by Brenner and Lerner in 1992[70]. As presented in Scheme 4-1 a multistep combinatorial approach (split and pool synthesis) is used to construct these libraries.

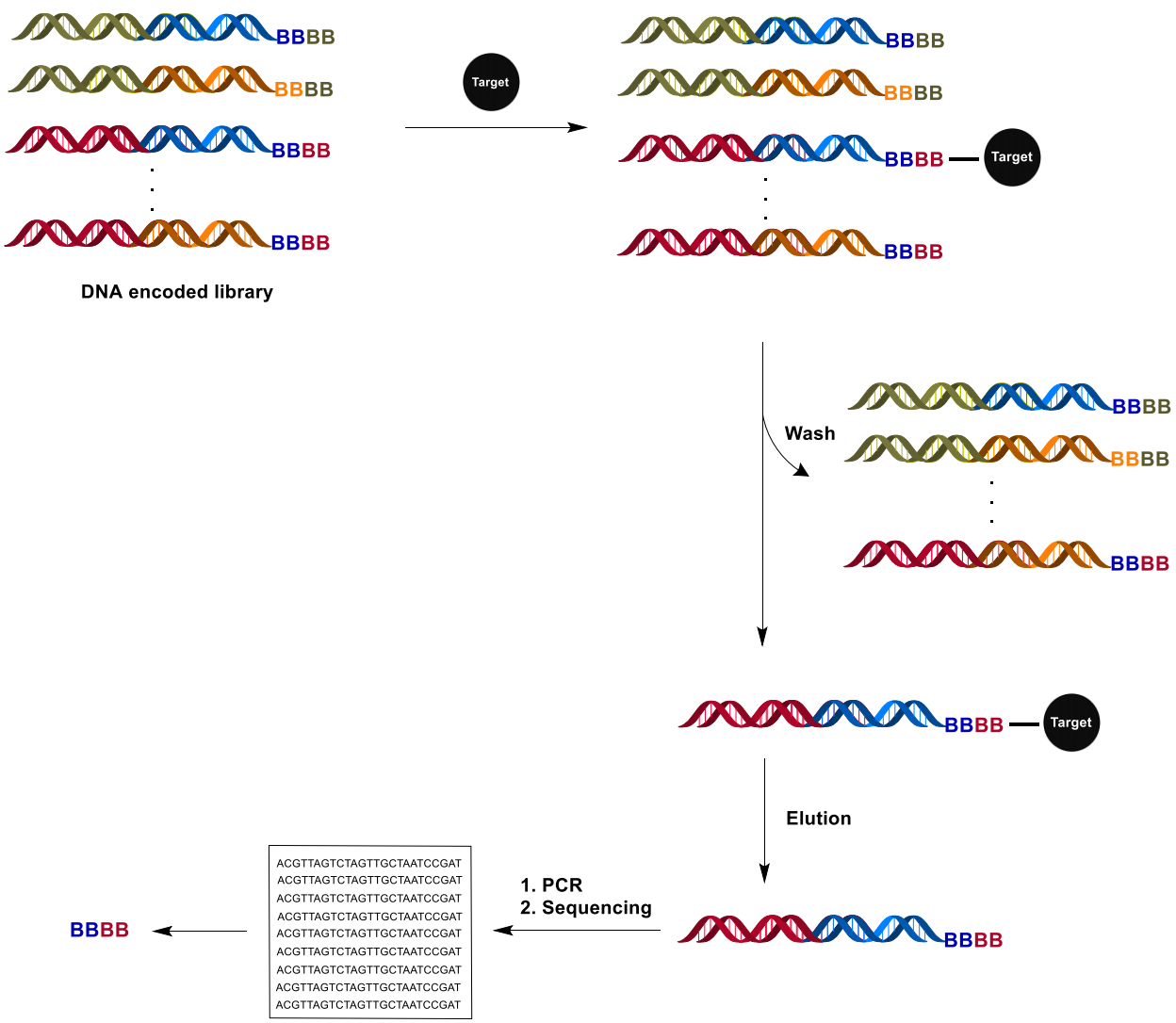


Scheme 4-1: Split and pool strategy for the synthesis of DNA-encoded libraries.

In a typical split and pool procedure, in the first step, *n* different small molecules (building blocks) are encoded using *n* different DNA tags which serve as a barcode. Then all the individual DNA-conjugated building blocks are pooled together and subsequently split into *m* different wells. Next, a second cycle of chemical reactions that attaches second set of building blocks is performed on each well and a new set of DNA strands specific to each new building block is ligated to previous DNA tags. This procedure yields $m \times n$ library compounds attached to their unique DNA barcodes. The split and pool can be repeated multiple times (usually 2−3 cycles).

4.2. DELs and drug discovery

The isolation of specific ligands against targets of pharmaceutical relevance is one of the most important steps in drug discovery. To meet the need of novel small-molecule lead candidates, many researchers have relied upon high-throughput screening (HTS) methods[71]. Notwithstanding the success of traditional HTS methods, the burden of operation costs, labour, infrastructure, and access to high quality diverse chemical libraries has prompted the development of more accessible drug discovery platforms. DELs have gained considerable traction as a cornerstone of contemporary drug discovery campaigns[72]. The encoding of each library member allows these libraries to be tested simultaneously in the same vessel. Usually, the target of interest is displayed on solid supports and then incubated with the encoded library. After incubation, non-binding members of the library are washed away, and binders (hits) are subsequently eluted and identified *via* high-throughput DNA sequencing and post-sequencing analysis (Scheme 4-2).



Scheme 4-2: General strategy for DEL selections.

Over the past few years, a growing number of hit compounds identified using DEL technology have been reported in the literature[73–81], with several candidates already undergoing phase trials in the US[72] (Figure 4-1).

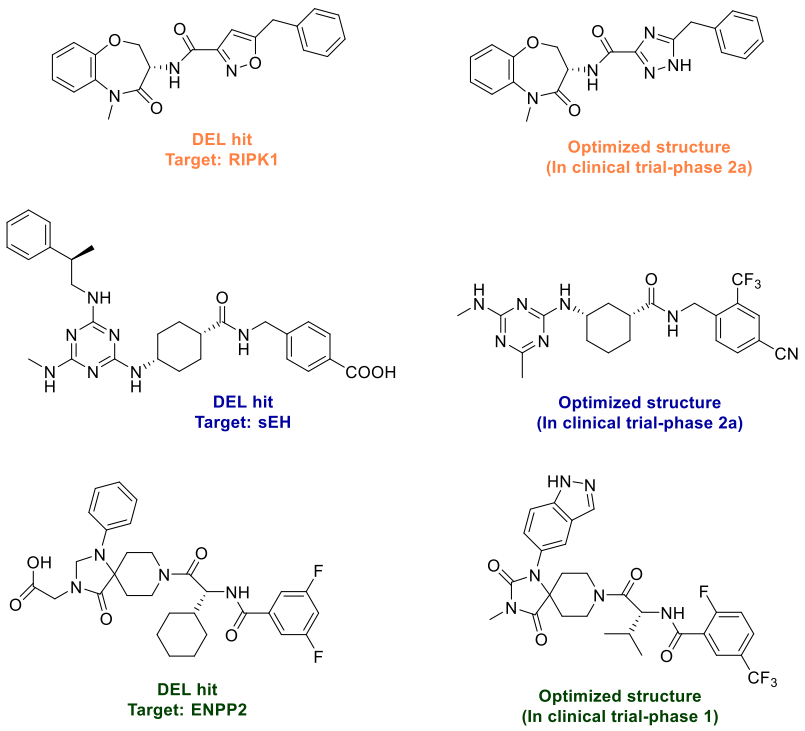


Figure 4-1: Hits derived from DEL screens and optimized structures currently in clinical trials.

4.3. DNA-compatible chemistries for DEL synthesis

New chemical reactions that can construct structurally diverse and pharmaceutically relevant molecules can expand the chemical diversity in DELs and increase the success of DEL selections. However, chemistries suitable for DEL synthesis are restricted by (*i*) insolubility of DNA in organic solvents and (*ii*) degradation of DNA under harsh conditions (*e.g.*, strongly acidic or oxidizing conditions, long reaction times and high temperatures). As a result, DELs continue to be limited by their chemical space and molecular properties, particularly when compared against traditional screening libraries[82]. Notwithstanding, the range of chemistries that can incorporate a variety of building blocks and functional groups into DELs is expanding rapidly[83–93], and the last decade has

witnessed an explosive growth in available DEL chemistry, including medicinal chemistry staples, such as Suzuki-Miyaura coupling[94] and Buchwald-Hartwig amination[90].

4.4. Employing Photoredox Catalysis for DNA-Encoded Chemistry

4.4.1. Photoredox Catalysis

Photocatalysis involves a series of oxidation and reduction (redox) reactions with the help of photo-emitted charge pairs under visible light. This approach relies on the ability of a photocatalyst (PC) to engage in single-electron transfer (SET) processes with organic substrates upon photoexcitation. Most of the commonly employed PCs are transition metal complexes or organic dyes (Figure 4-2 and Figure 4-3). While organic photoredox catalysts were dominant in the past, transition-metal complexes are more commonly used today, ruthenium and iridium polypyridyl complexes standing at the forefront of this class.

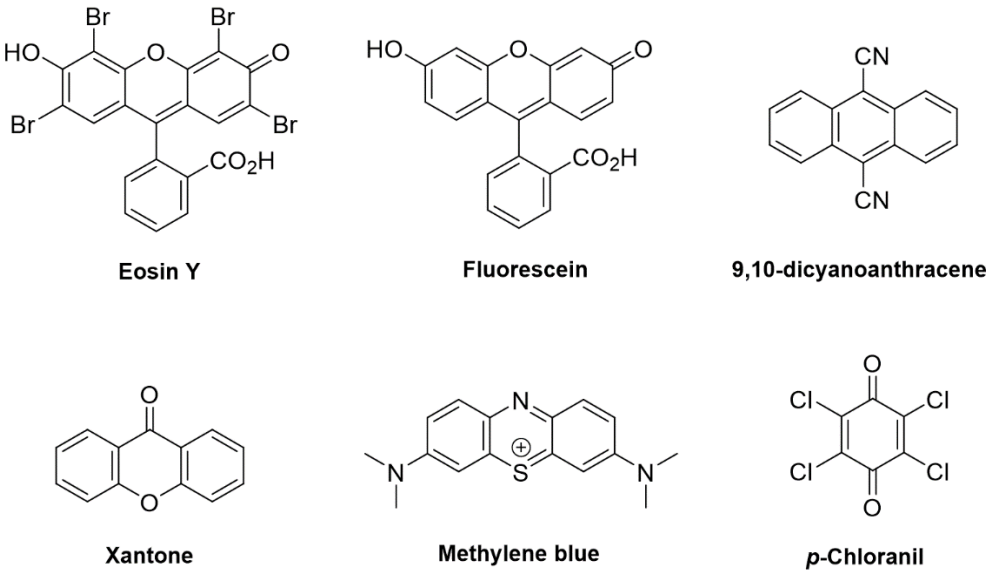


Figure 4-2: Common organic photocatalysts.

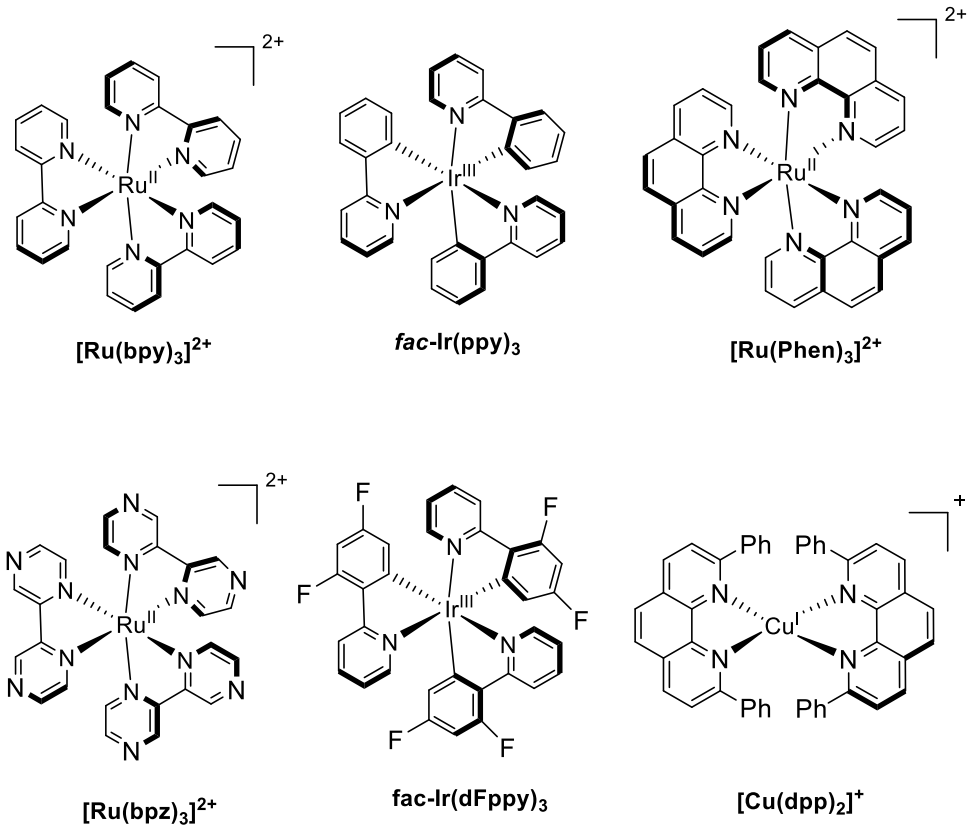
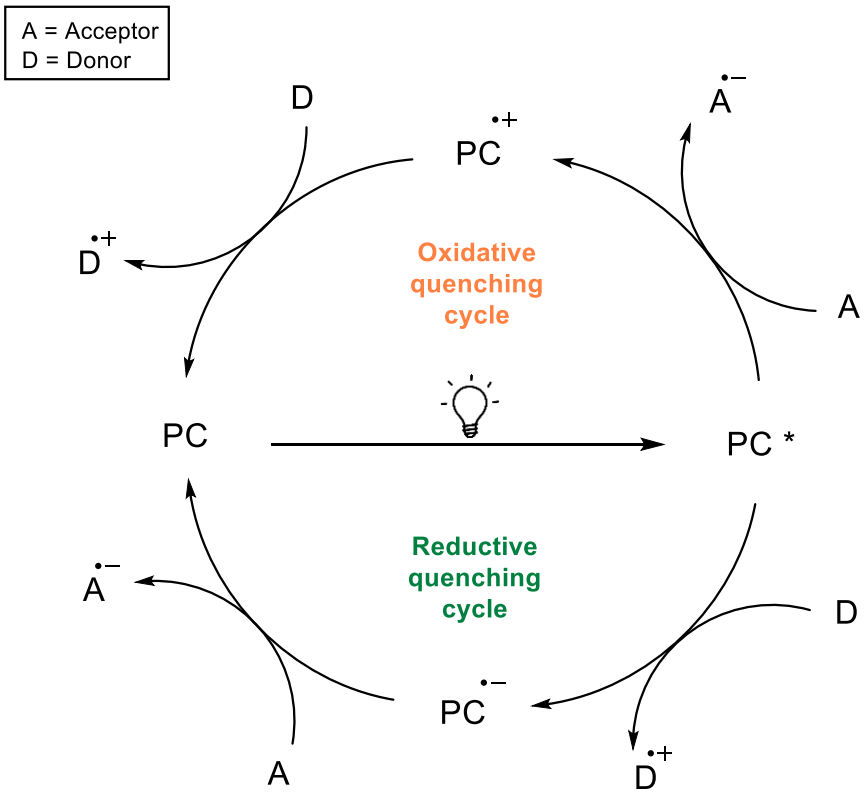


Figure 4-3: Common transition metal-based photocatalysts.

4.4.2. General photocatalysis pathways for transition metal complexes

Upon irradiation, a transition metal-based photocatalyst can absorb a photon and form a long-lived excited species (PC*) through a metal-to-ligand charge transfer. The photoexcited species is both more oxidizing and more reducing than the ground-state PC and, depending on the substrates and reagents that are present in the reaction mixture, can accept (reductive quenching cycle) or donate (oxidative quenching cycle) a single electron (Scheme 4-3). The path of the PC is often made based on the substrate, and by referring to standard reduction potentials.



Scheme 4-3: General pathways of photoredox catalysis.

4.4.3. Photocatalytic hydrogen atom transfer (HAT)

Hydrogen atom transfer (HAT) is a chemical reaction in which a hydrogen radical is abstracted from a substrate according to the following general equation:

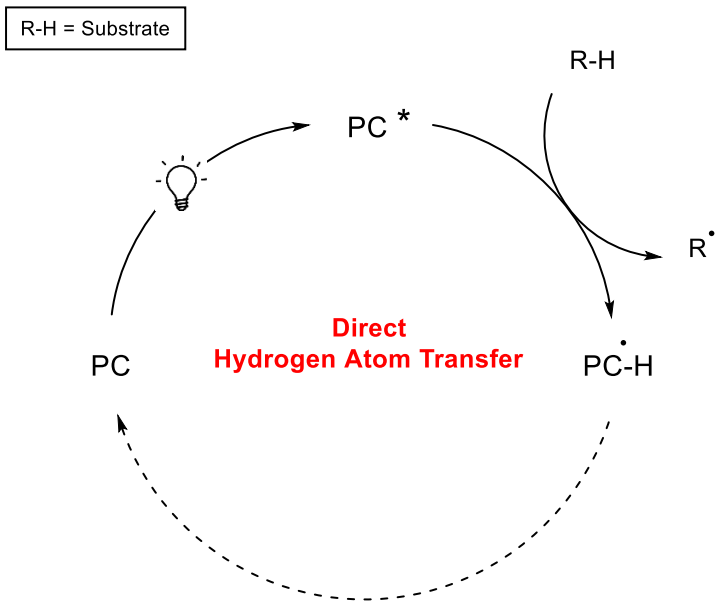
$$A^{\bullet} + B\text{-}H \rightarrow B^{\bullet} + A\text{-}H$$

This concerted movement of a proton and an electron between two compounds in a single step represents unique opportunities in organic synthesis, allowing the direct and selective functionalization of C–H bonds. Recently, there has been significant

development of photocatalytic approaches that activate a C-H bond through a HAT step[95]. These methods can be classified into two categories, direct and indirect approaches.

4.4.3.1. Direct HAT approaches

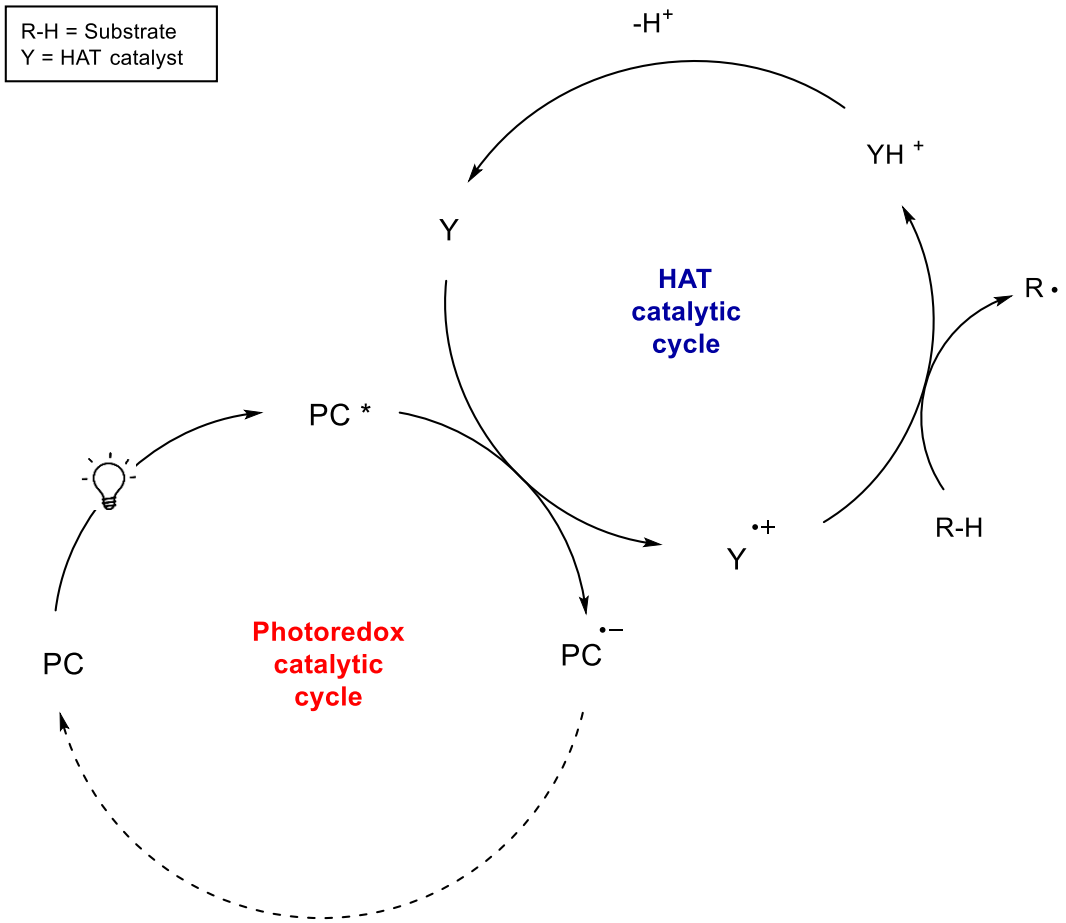
This approach is based on the capability of the excited photocatalyst (PC*) to induce a direct HAT process where it initiates the process by homolytic cleavage of a C-H bond in the organic substrate and abstracts a hydrogen atom directly from it (Scheme 4-4). So far, only a limited number of photocatalysts that are able to perform this chemistry have been reported[95,96]. The choice of competent catalysts able to perform direct HAT is mostly limited to several organic dyes – in particular oxo complexes. Upon light excitation, photoactive carbonyl compounds can reach an excited state capable of abstracting an H atom directly from C–H bonds of certain substrates, which possess redox potentials in a specific range that allows the electron transfer to be thermodynamically feasible.



Scheme 4-4: Reaction mechanism involving a direct photoredox HAT approach.

4.4.3.2. Indirect HAT approaches

In recent years, the limited number of photocatalysts able to perform direct HAT have led to the development of indirect HAT processes that enable activation of desired R–H bonds with high selectivity and efficiency[95,97,98]. In these reactions, a single electron transfer between a photocatalyst and a suitable purposely added HAT cocatalyst generates a radical intermediate capable of abstracting a hydrogen atom from the organic substrate resulting in a dual catalytic approach (Scheme 4-5).

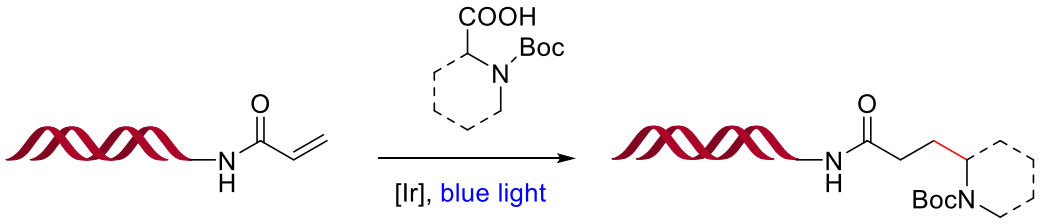


Scheme 4-5: Typical reaction mechanism involving an indirect photoredox HAT approach.

4.4.4. DEL-compatible photochemical reactions

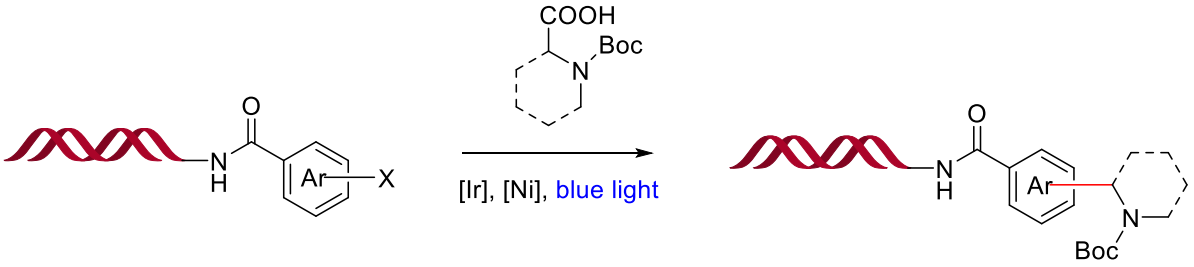
Photochemical reactions have been previously used in nucleic acid-based technologies and are a powerful tool that enables researchers to design novel biological systems and conduct various molecular manipulations[99]. These reactions have only recently been used for the construction of chemical bonds on nucleic acids and hold great potential for DEL preparations.

In 2018, inspired by a Giese-type reaction that was initially reported in 2014[100], Flanagan and co-workers used an Iridium photocatalyst for the addition of *N*-protected α-amino acids to DNA-tagged Michael acceptors to produce the C(sp3)–C(sp3)-coupled product (Scheme 4-6)[101].



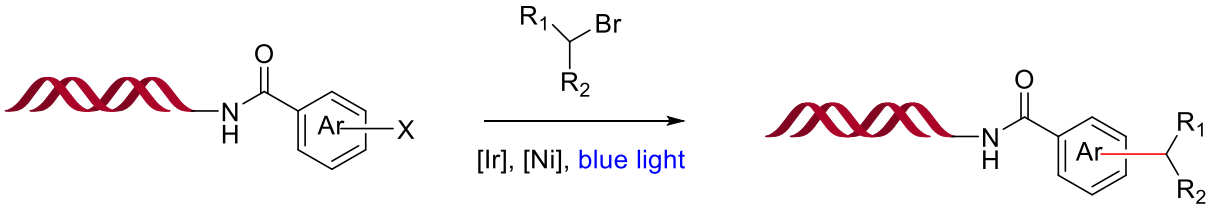
Scheme 4-6: Photoredox-mediated decarboxylative alkylation of α-amino acids.

In 2019, the same group reported a dual catalysis approach, using Iridium and Nickel catalysts for the decarboxylative arylation of DNA-tagged aryl halides and α-amino acids (Scheme 4-7).



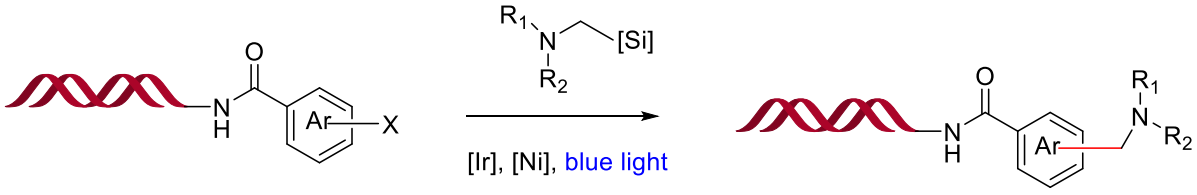
Scheme 4-7: Photoredox-mediated decarboxylative arylation of α-amino acids.

In 2020, Molander and coworkers reported three photoredox reactions to expand chemical space in DELs[102]. First, a photoredox/nickel-mediated reductive C(sp2)-C(sp3) coupling of alkyl bromides to DNA-tagged aryl halides (Scheme 4-8).



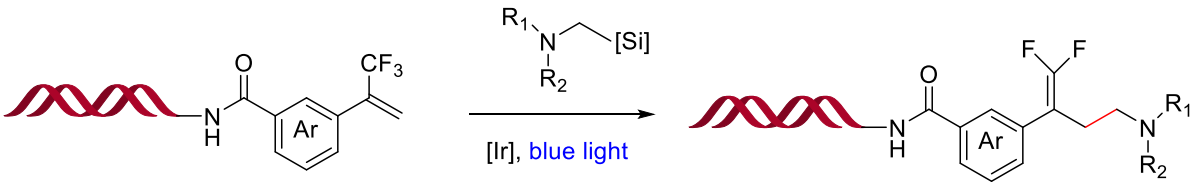
Scheme 4-8: C(sp2)-C(sp3) cross-coupling of alkyl bromides to DNA-tagged aryl halides.

Then aminomethyl cross coupling of on-DNA aryl halides and α-silylamines was reported as the second approach (Scheme 4-9).



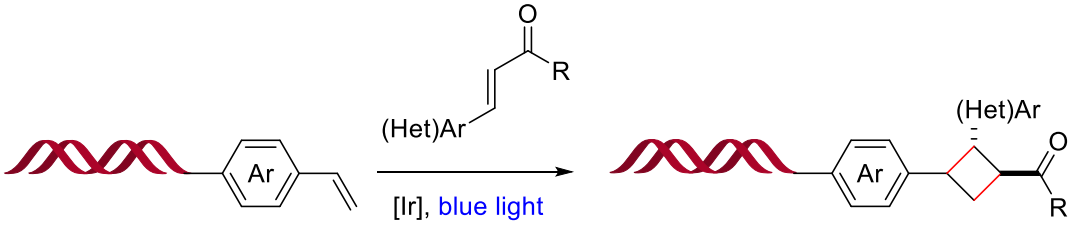
Scheme 4-9: Aminomethyl cross-coupling of on-DNA aryl halides and α-silylamines.

Finally, photoredox-mediated radical/polar crossover reactions using α-silylamines were employed for defluorinative alkylation (Scheme 4-10).



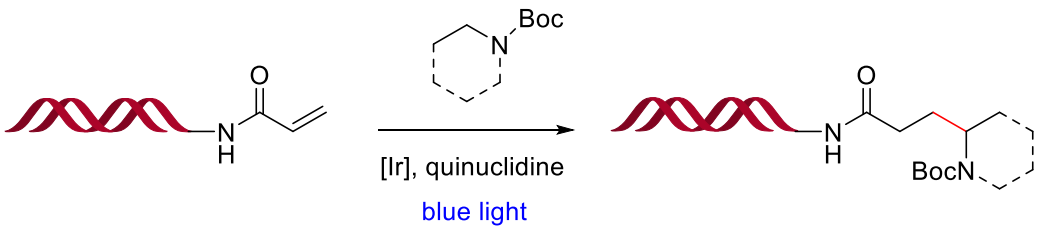
Scheme 4-10: Deflourinative aminomethylation using α-silylamines.

In the same year, Kölmel and co-workers also used an Iridium photocatalyst in a [2 + 2] cycloaddition reaction for the synthesis of highly substituted cyclobutanes[103]. This reaction was carried out using DNA-tagged styrene derivatives and cinnamates (Scheme 4-11).



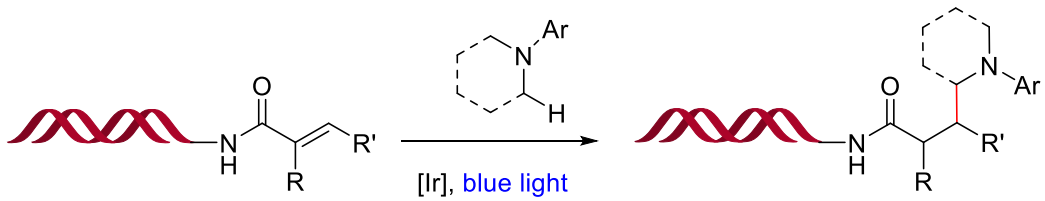
Scheme 4-11: [2+2] cycloaddition catalyzed by an iridium photocatalyst.

In 2021, Lu and co-workers described α-alkylation of *N*-Boc protected heterocyclic rings with DNA-tagged acrylamides using an Iridium complex in the presence of quinuclidine as a hydrogen atom transfer (HAT) catalyst[104] (Scheme 4-12).



Scheme 4-12: α-alkylation of *N*-Boc protected heterocycles with on-DNA acrylamides.

Moreover, in a recent study by Gao, Wan, Liu and co-workers, a photoredox reaction between on-DNA alkenes and *N*-aryl-substituted tertiary amines using an Iridium catalyst was described for C(sp3)−C(sp3) bond construction[105] (Scheme 4-13).



Scheme 4-13: Photoredox catalysis between on-DNA alkenes and *N*-aryl-substituted tertiary amines.

4.5. Thesis project

The success of DEL selections is largely related to the chemical diversity assembled by DNA-compatible reactions. Although much progress was made over the last decade, the number of chemical reactions compatible with DNA is comparatively limited, therefore DELs are still limited by their chemical space and molecular properties. Inspired by the

previous work done in the field, in the second part of this thesis a new DEL-compatible protocol aimed at constructing C(sp3)−C(sp3) bonds via photoredox catalysis is explored, which enables the production of structurally diverse small molecules and could increase the success of DEL selections.

Chapter 5. Photoredox-catalysed Hydroaminoalkylation on DNA-encoded Secondary *N*-Arylamines

5.1. Abstract

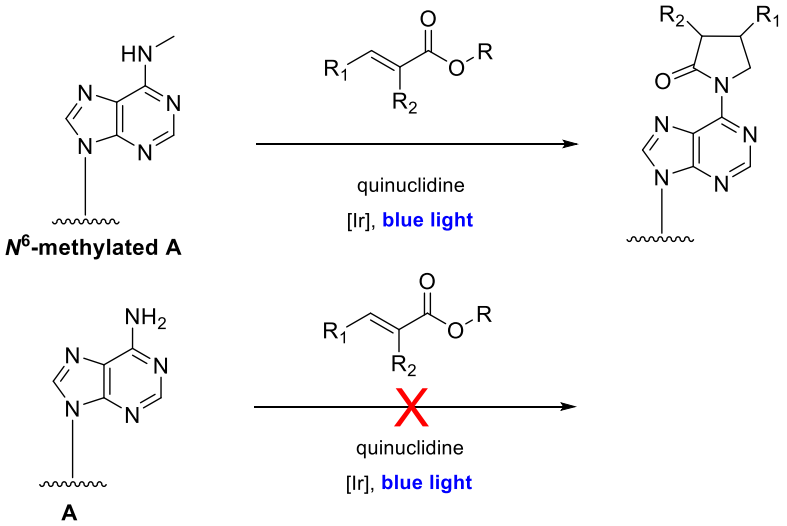
An efficient approach to the photoredox-catalysed hydroaminoalkylation between on-DNA secondary *N*-substituted (hetero)arylamines and vinylarenes has been developed and explored. The methodology was examined with a broad scope of vinylarenes and secondary arylamines to establish a preferred building block profile for the process. Compatible substrates furnished the desired derivitised amine products in modest to excellent yields and with minimal or no detectable by-products.

5.2. Introduction

Recently, advances in photoredox-catalysed reactions have been ported into DEL-compatible processes. Of particular note, photoredox catalysis has enabled broader access to new bond-formation reactions to generate complex amine-containing DELs. An evaluation of marketed drugs in the US demonstrates the prevalence of alkyl-substituted aromatic amines, which comprise approximately 20% of the top 100 drugs by sales. Not surprisingly, the synthesis of this class of molecules within the context of DEL synthesis has been extensively explored, particularly *via* metal-catalysed C-N cross-coupling reactions[90,106–111].

Initially inspired by the blue light mediated strategy reported in the literature for the construction of poly-substituted pyrrolidones[112], I was exploring blue light mediated reactions for the detection of N^6-methyladenine sites on oligonucleotides. Under the optimized conditions, N^6-methylated sites would react selectively, leaving normal

adenines intact (Scheme 5-1). Selectively modifying and blocking H-bonding sites on methylated nucleobases would result in high error rate on these sites during sequencing.



Scheme 5-1: Selective modification of N^6-methylated adenine bases using blue light reactions.

Later, I speculated that we might be able to use these reactions for DNA encoded library constructions. Due to the combinatorial nature of DEL synthesis, chemistries that can derivatise upon installed building blocks can expand the chemical space of the encoded library. Therefore, the focus of my research was shifted to optimizing the blue light mediated reactions for DEL synthesis.

I was drawn to the photoredox-catalysed hydroaminoalkylation[113], as it could combinatorially stack with well-established aryl C-N coupling chemistry. Hydroaminoalkylation via Giese reaction in a DEL context has been previously demonstrated using photoredox chemistry[101,102,104,105,114–117]; however, given its potential for library development, several aspects of this chemistry remain underexplored. First, α-aminoalkyl radicals are invariably generated as the off-DNA reactant, which then adds

to the DNA-tagged alkene. This limits the accessible library architectures and building blocks. Second, and of particular note, generation of the α-aminoalkyl proceeds *via* decarboxylation of α-aminoacids or their corresponding activated esters[101,114–116], α-silylamines[102], or using readily oxidisable symmetrical tertiary amines[104,105]. These substrate restrictions considerably limit the supply of available amine- containing building blocks. To address these issues and expand upon the scope of this chemistry, photoredox-catalysed hydroaminoalkylation of vinylarenes using DNA-tagged secondary *N*-alkylaniline and arylamine derivatives was evaluated. While secondary aniline substrates are known to be challenging substrates[112] for Giese-type chemistry, the products of such processes are highly represented in bioactive molecules (Figure 5-1) and will open the scope of complex amine libraries for DELs.

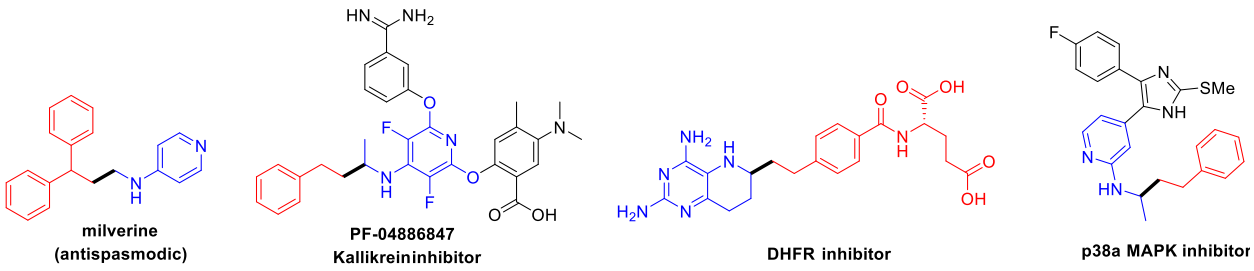


Figure 5-1: Examples of biologically active molecules synthetically accessible through hydroaminoalkylation with an alkene.

5.3. Results and discussion

The reaction between DNA-tagged *n*-butylaniline **1a** and 1,1-diphenylethylene **2a** was first evaluated (Table 5-1). Using the photoredox catalyst Ir[*p*-F(Me)ppy]$_2$(dtbbpy) (**PC1**) in the presence of HAT catalyst quinuclidine and blue light, resulted in 75% yield of **3a** as

determined by LC-MS. Various photocatalysts (**PC1**−**PC6**) (Figure 5-2) were evaluated; however, no increase in yield was observed (Table 5-1, entries 1-6).

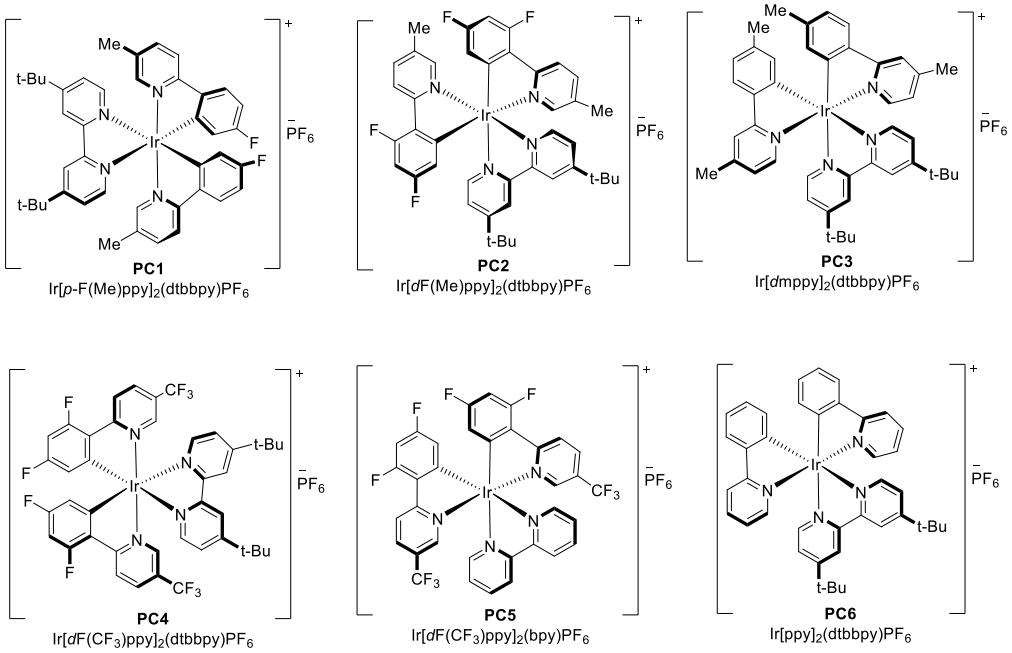
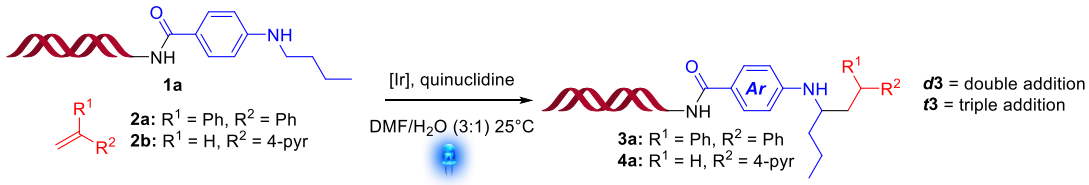


Figure 5-2: Structures of photocatalysts evaluated for photoredox-mediated hydroaminoalkylation reaction.

Extending the reaction time to 3 h using **PC1** resulted in quantitative yield of the desired product (Table 5-1, entry 8). As anticipated, the inclusion of quinuclidine as a HAT catalyst was necessary and significantly improved the yield of the reaction (Table 5-1, entry 9). Since the rate of α-alkylamino radical addition to alkene substrates will be governed by the electronics of the alkene, the process on electron poor alkene **2b** was evaluated. As expected, reaction rates increased, with quantitative yield observed after 1 h (Table 5-1, entry 10). When reactions were left for extended periods, double (d3) and triple (t3) addition adducts were observed with a concomitant decrease in desired product (Table 5-1, entries 11-12). To enable broad access to alkene substrates with variable electronics,

I opted for a 1.5 h reaction, which yielded 75% and 73% for **2a** and **2b** substrates respectively. Importantly, reactions were found to require rigorous degassing under an inert atmosphere to minimize undesired *N*-dealkylation.

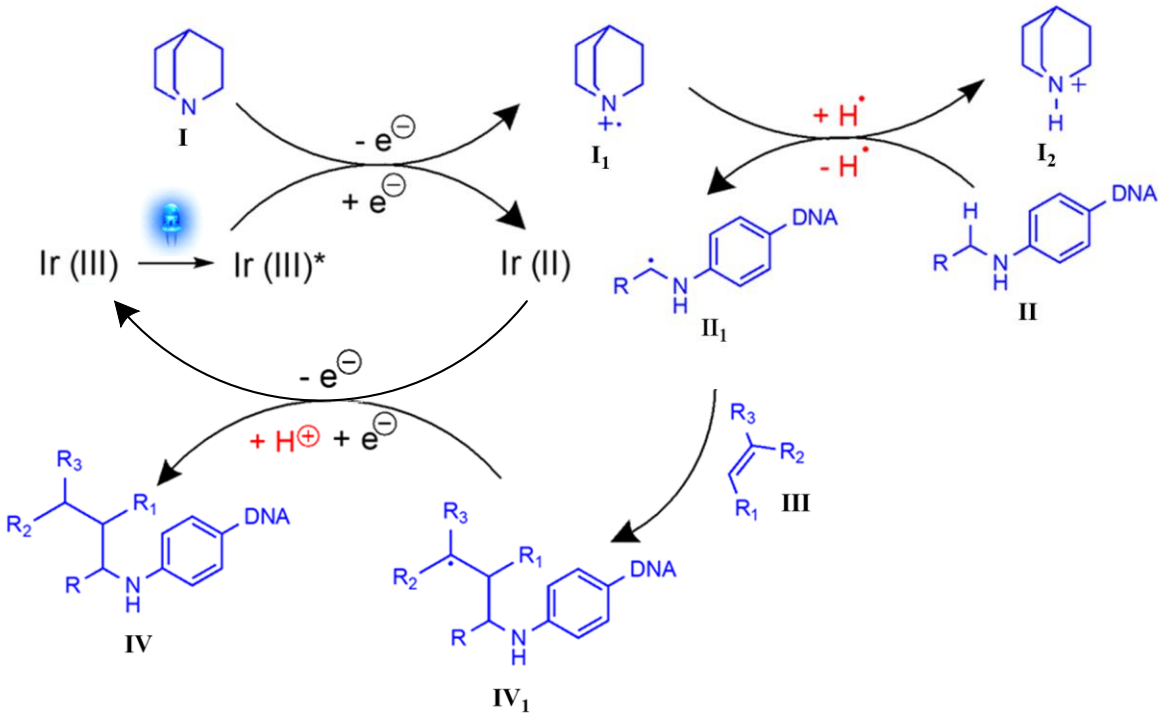
Table 5-1: Optimisation of on-DNA hydroaminoalkylation[a]



entry	alkene	catalyst	time (h)	3a or 4a (%)	*d3*[b] (%)	*t3*[c] (%)
1	2a	PC1	1.5	75	0	0
2	2a	PC2	1.5	58	0	0
3	2a	PC3	1.5	17	0	0
4	2a	PC4	1.5	49	0	0
5	2a	PC5	1.5	39	0	0
6	2a	PC6	1.5	75	0	0
7	2a	PC1	2	80	0	0
8	2a	PC1	3	100	0	0
9[d]	2a	PC1	1.5	59	0	0
10	2b	PC1	1	100	0	0
11	2b	PC1	1.5	73	27	0
12	2b	PC1	2	55	32	13
13[d]	2b	PC1	1.5	77	10	0

[a] Reaction conditions: degassed mixture of DNA tagged **1a** (10 nmol), quinuclidine (500 eq.), alkene **2a** or **2b** (250 eq.), Iridium photocatalyst PC1−PC6 (1 eq.), DMF/H_2O (3:1), blue light (10 cm, max intensity, Kessil A160WE), fan cooled. [b] double addition product. [c] triple addition product. [d] no quinuclidine.

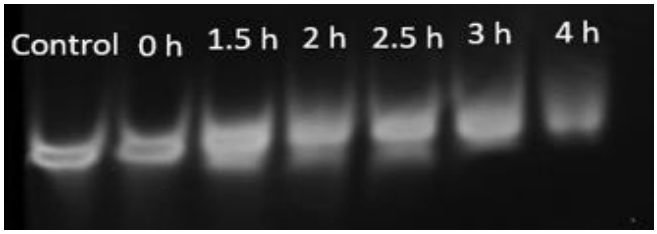
A proposed mechanistic cycle for the hydroaminoalkylation *via* photoredox-mediated HAT catalysis is outlined in Scheme 5-2.



Scheme 5-2: Proposed mechanism for photoredox-mediated hydroaminoalkylation.

The catalytic cycle is thought to begin with the excitation of the photocatalyst by blue light. The excited photocatalyst is a strong oxidant and can readily get reduced by quinuclidine *via* single electron transfer (SET) to form Ir (II) and quinuclidine in its radical cation form (**I_1**), which can abstract a hydrogen atom from the α-carbon of the *N*-alkylated arylamines (**II**) to form quinuclidinium ion (**I_2**) and radical **II_1** *via* hydrogen atom transfer (HAT). The nucleophilic addition of radical **II_1** to the acceptor alkene **III** would furnish alkyl radical **IV_1**. in the final step, the catalytic cycle is closed by single electron transfer and protonation to generate the desired product **IV**.

The stability of DNA tag under photoredox conditions was evaluated using gel analysis. (Figure 5-3) The band corresponding to the DNA tag was left clear and intact after 1.5 h reaction time and no degradation traces were observed.

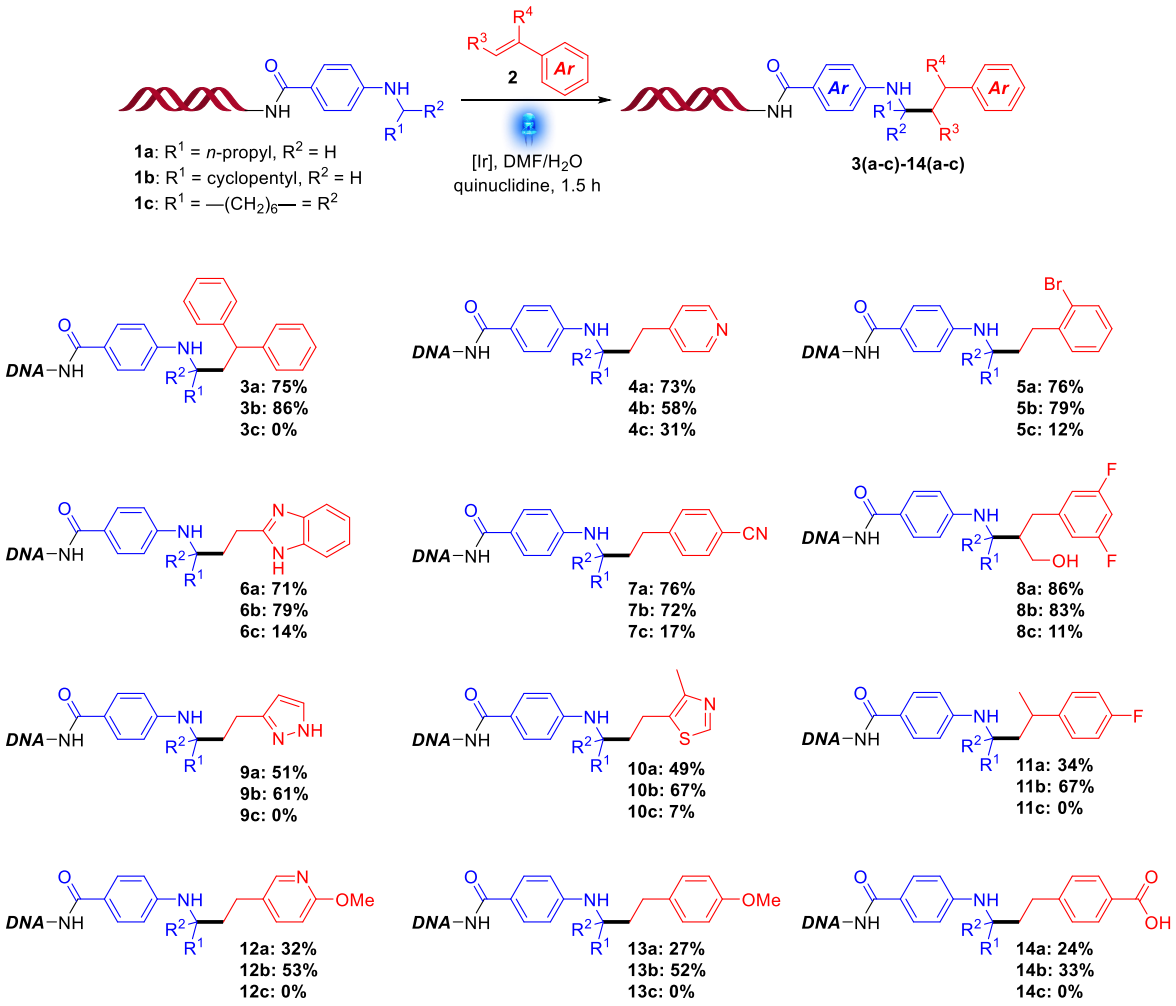


Photocatalysis reaction was performed on a DNA conjugate with 4-vinyl pyridine for 0, 1.5, 2, 2.5, 3, 4 h and the DNA stability was assessed using non-denaturing gel analysis.

Figure 5-3: Stability of DNA tag after photoredox-mediated hydroaminoalkylation reaction evaluated by gel electrophoresis.

The optimised hydroaminoalkylation process was subsequently applied to a more diverse substrate scope of vinylarenes (Table 5-2). Using DNA-tagged aniline **1a**, vinylarenes with diverse electronics were surveyed to afford products **3a**-**14a**. Disubstituted alkenes were compatible with the reaction and resulted in products (**3a**, **8a**, **11a**) in good to excellent yields. As expected, electron rich vinylarenes were less tolerated, resulting in products in modest yields (**12a**, **13a**, **14a**). Importantly, vinylheteroarenes were within the scope of the process (**6a**, **9a**, **10a**, **12a**) and did not result in the detection of unknown by-products. Other substituents on the aniline nitrogen were explored, including more sterically encumbered alkyl substrates (**1b**, **1c**). It was observed that having a trisubstituted α-carbon for the DNA-tagged amine, **1c**, severely hampered the reaction, resulting in low yield and produced products only for a handful of substrates. This is likely due to a sterically inhibited attack of the α-alkylamino radical on the vinylarene substrates.

Table 5-2: Hydroaminoalkylation of various vinylarenes with variably substituted DNA-tagged secondary *N*-alkyl anilines[a]

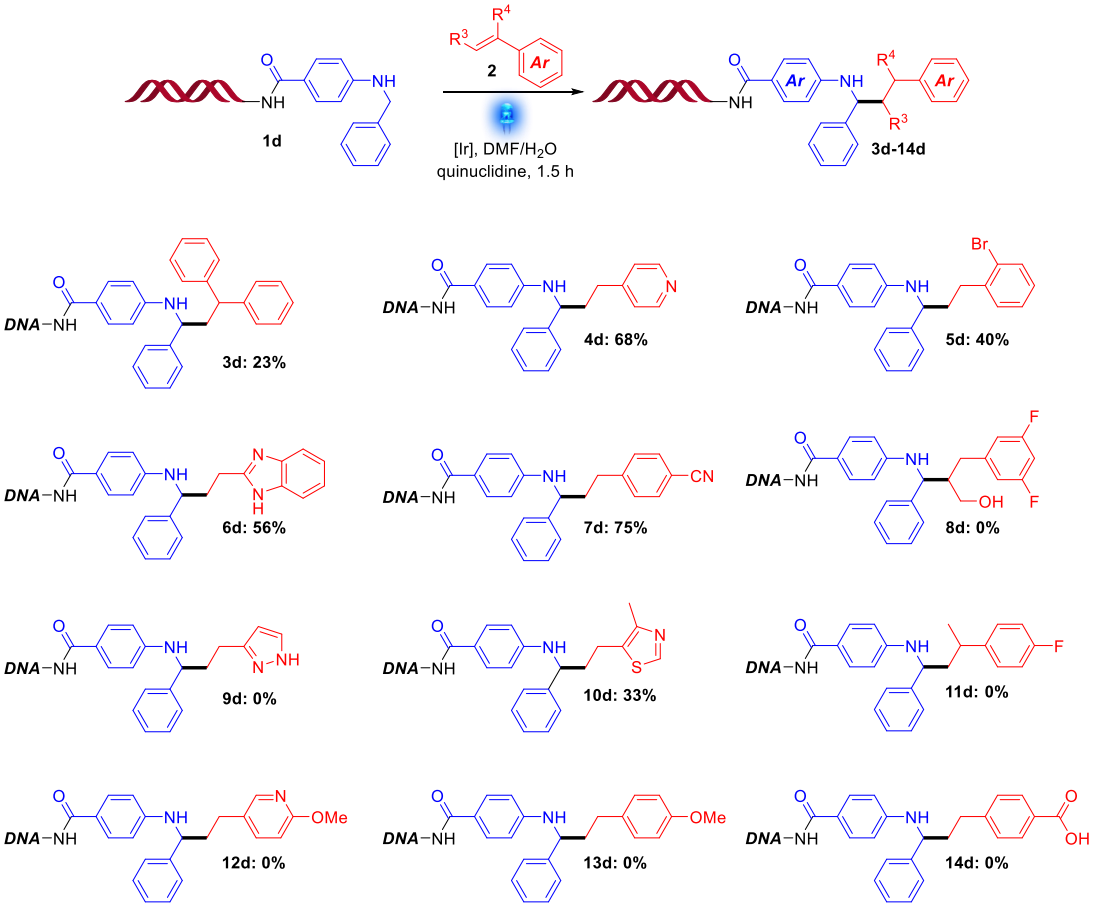


[a]Reaction conditions: degassed mixture of DNA tagged **1a-1c** (10 nmol), quinuclidine (500 eq.), alkene **2** (250 eq.), Iridium photocatalyst PC1 (1 eq.), DMF/H_2O (3:1), blue light (10 cm, max intensity, Kessil A160WE), fan cooled. The yield was determined by LC-MS analysis.

N-benzyl substituted aniline **1d** was also explored with the optimised hydroaminoalkylation process (Table 5-3). Due to the lower oxidation potential of *N*-benzyl anilines compared with their alkyl counterparts, a HAT catalyst was not necessary, but was found to increase the rate of reaction and minimise reaction by-products including *N*-dealkylation (Table 5-5). This enables the incorporation of such *N*-benzylic substrates during the combinatorial preparation of DELs without the

requirements for separate plate conditions. Electron-deficient vinylarenes were well tolerated during hydroaminoalkylation with **1d**, producing the desired products in good to high yield; however, sterically demanding alkenes and electron-rich alkenes resulted in no observable product formation. Curiously, while most failed reactions resulted in quantitative recovery of starting material, reactions for **9d**, **13d**, and **14d**, resulted in 10-20% of the desired product as a covalent adduct with HAT catalyst quinuclidine (Table 5-9).

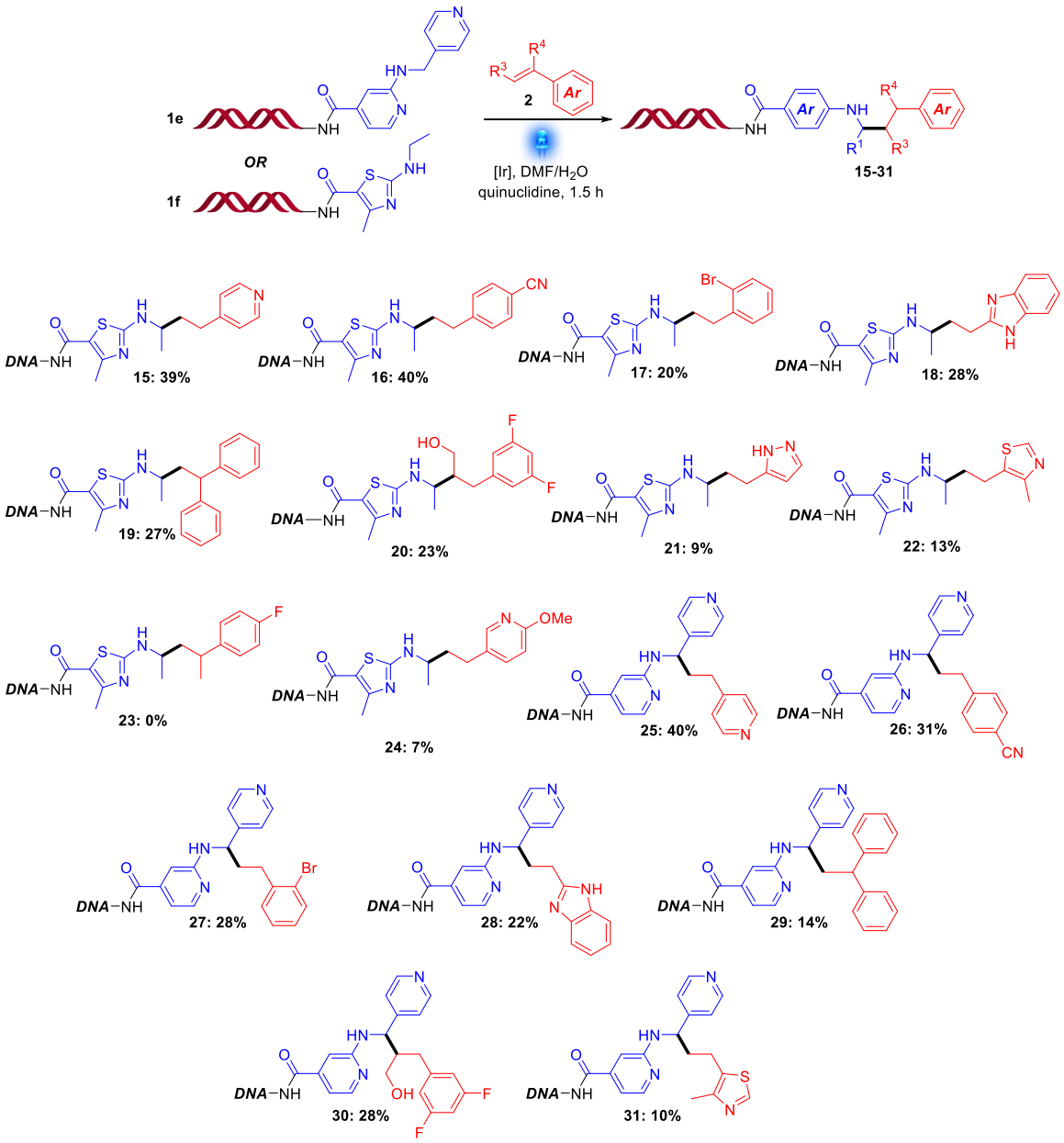
Table 5-3: Hydroaminoalkylation of various vinylarenes with DNA-tagged secondary *N*-benzyl aniline[a]



[a]Reaction conditions: degassed mixture of DNA tagged **1d** (10 nmol), quinuclidine (500 eq.), alkene **2** (250 eq.), Iridium photocatalyst PC1 (1 eq.), DMF/H_2O (3:1), blue light (10 cm, max intensity, Kessil A160WE), fan cooled. The yield was determined by LC-MS analysis.

To expand the scope of the chemistry, the tolerance of the optimised hydroaminoalkylation reaction using *N*-substituted heteroarylamines on DNA was also explored (Table 5-4).

Table 5-4: Hydroaminoalkylation of various vinylarenes with DNA-tagged secondary *N*-substituted heteroarylamines[a]



[a]Reaction conditions: degassed mixture of DNA tagged **1e-1f** (10 nmol), quinuclidine (500 eq.), alkene **2** (250 eq.), Iridium photocatalyst PC1 (1 eq.), DMF/H_2O (3:1), blue light (10 cm, max intensity, Kessil A160WE), fan cooled. The yield was determined by LC-MS analysis.

Consistent with the previous scope, electron-poor alkene substrates performed the best, with most giving modest yields. The yield of strongly electron deficient alkenes, such as those with 4-pyridine and 4-cyanobenzene, suffered from significant double and triple addition adducts (Table 5-10 and Table 5-11). By and large, the modest yields were mostly attributed to low reactivity, as the mass balance of the reactions were starting material, with the exception of **20**, which was partially contaminated with 20% *N*-dealkylated **1f**. A more comprehensive heteroarylamine scope may reveal other limitations or preferences for the chemistry; however, alkyl substituents on the amine paired with electron-deficient alkenes appear to be the best tolerated and highest yielding, while heteroarylamine substrates underperformed compared with substituted anilines.

It is envisaged that the optimised hydroaminoalkylation will find use in DEL workflows where secondary *N*-alkyl or *N*-benzyl arylamines are generated on DNA. A hypothetical library synthesis may involve an initial coupling of a library of (hetero)aryl halides for cycle 1, followed by palladium-mediated C-N coupling of a library of primary alkylamines/benzylamines for cycle 2[111]. Both steps would take advantage of large and readily available building block libraries. For cycle 3, the photoredox-catalysed hydroaminoalkylation with a library of electron poor/neutral vinylarenes would furnish the resulting derivatised amine library. The envisaged DEL could feasibly provide a library with 15 million encoded members (Figure 5-4).

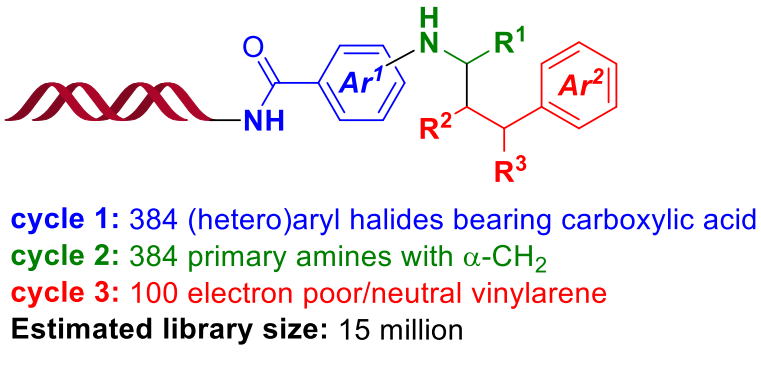


Figure 5-4: Proposed library design of 15 million members.

5.4. Conclusions

In conclusion, a photoredox-catalysed hydroaminoalkylation reaction between on-DNA secondary *N*-alkylated (hetero)arylamines and vinylarenes was developed. Despite secondary arylamines serving as poor surrogates for α-alkylamino radicals, reactions proceeded well with a broad scope of alkene acceptors. The most effective vinylarene reactants were found to be electron poor to neutral; strongly electron poor substrates were observed to undergo undesired double-addition, while electron-rich substrates resulted in poor yields. Ideal on-DNA secondary *N*-arylamine reactants contained an α-methylene; substrates with α-methines resulted in poor conversions. *N*-alkyl and *N*-benzyl substrates, along with heteroaryl amines were within the scope of the process. As the process can build from a secondary amine molecule on-DNA, products will contain a free NH group, which may serve as a hydrogen-bond donor when conducting selections against proteins, but may also serve a site for further combinatorial elaboration. This efficient approach for C(sp3)−C(sp3) bond construction is DNA-compatible, tolerates variety of different functional groups, and when

combinatorially coupled with readily available aryl C-N coupling chemistries, it has great potential to find broad use as an affordable approach in the design and construction of DNA-encoded libraries of drug-like molecules for drug discovery campaigns.

5.5. Experimental details and supporting data

5.5.1. General information

Purifications were performed by reverse-phase high-performance liquid chromatography (HPLC, Agilent 1260 Infinity II) using a C18 stationary phase (5µm Eclipse XDB-C18 9.4 x 250 mm). Liquid chromatography–mass spectrometry (LC–MS) analyses were performed using Agilent Infinity Lab LC/MSD system on a C18 stationary phase (HALO 400 A, ES-C18, 3.4 µM, 2.1 x 30 mm). ^{1}H NMR spectra were recorded at 400 MHz on a Bruker spectrometer. Processing of the spectra was performed with TopSpin software. Analytical thin-layer chromatography (TLC) was performed on aluminum plates pre-coated with silica gel 60F_{254} as the adsorbent (Sigma-Aldrich, 1.05554). The developed plates were air-dried and exposed to UV light.

5.5.2. DNA headpiece

DNA headpiece was prepared according to literature methods[118].

$P_{17}O_{106}N_{52}C_{165}H_{234}-NH_2$

Molecular Weight: 5184 D

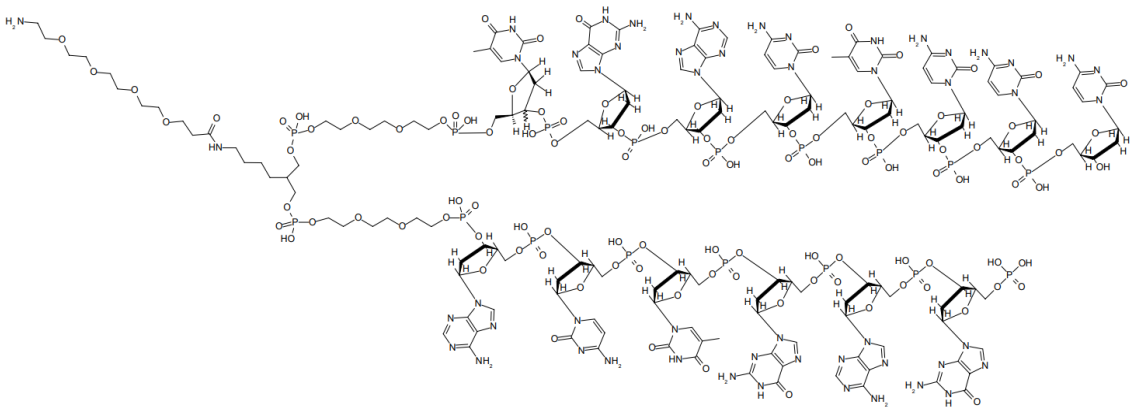


Figure 5-5: Structure of the DNA headpiece used for the preparation of DNA conjugates.

5.5.3. Photocatalysts

PC1 [(4,4′-di-tert-butyl-2,2′-bipyridine)-bis-(5-methyl-2-(5-fluoro-phenyl)-pyridine) iridium(III)] hexafluorophosphate (Sigma-Aldrich, 908703)

PC2 [4,4'-Bis(1,1-dimethylethyl)-2,2'-bipyridine-κN,κN]bis[3,5-difluoro-2-(5-methyl-2-pyridinyl) phenyl] iridium hexafluorophosphate (Strem Chemicals, 77-0330)

PC3 4,4'-Bis(t-butyl-2,2'-bipyridine]bis[5-methyl-2-(4-methyl-2-pyridinyl-kN)phenyl-kC]iridium hexafluorophosphate (Strem Chemicals, 77-0218)

PC4 (4,4'-Di-t-butyl-2,2'-bipyridine)bis[3,5-difluoro-2-[5-trifluoromethyl-2-pyridinyl-kN)phenyl-kC]iridium(III) hexafluorophosphate (Strem Chemicals, 77-0425)

PC5 (2,2'-Bipyridine)bis[3,5-difluoro-2-[5-(trifluoromethyl)-2-pyridinyl-kN] [phenyl-kC]iridium(III) hexafluorophosphate (Strem Chemicals, 77-0220)

PC6 (4,4'-Di-t-butyl-2,2'-bipyridine)bis[2-(2-pyridinyl-kN)phenyl-kC]iridium(III) hexafluorophosphate (Strem Chemicals, 77-0410)

5.5.4. Vinylarenes

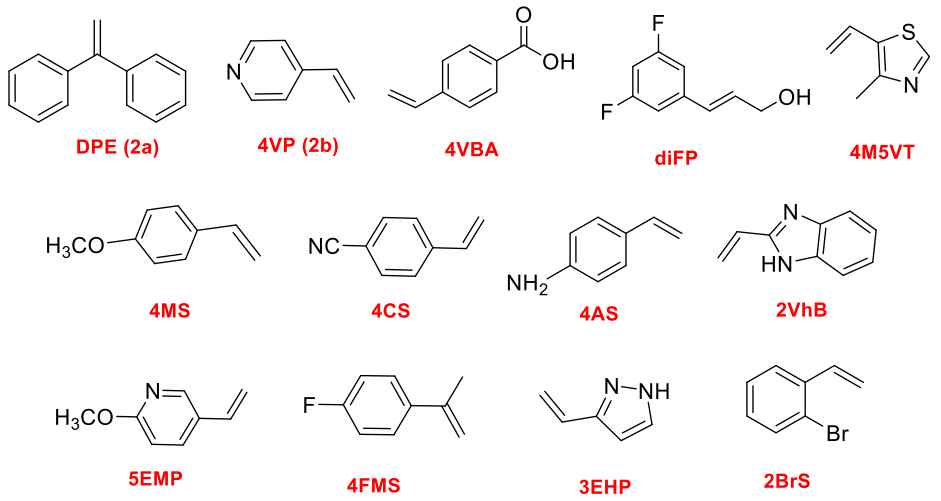


Figure 5-6: Structures of the vinylarenes explored for hydroaminoalkylation.

DPE (2a) 1,1-Diphenylethylene (Sigma-Aldrich, D206806)

4VP (2b) 4-Vinylpyridine (Sigma-Aldrich, V3204-5ML)

5EMP 5-Ethenyl-2 methoxy-pyridine (Combi Blocks, QE-5274)

4M5VT 4-Methyl-5-vinylthiazole (Combi Blocks, OR-0987)

diFP 3-(3,5-Difluorophenyl)propenol (Combi Blocks, SS-9410)

4MS 4-Methoxystyrene (Combi Blocks, QB-0479)

4AS 4-Aminostyrene (Combi Blocks, 4640)

4CS 4-Cyanostyrene (Combi Blocks, QF-7194)

2VhB 2-Vinyl-1h-benzimidazole (Combi Blocks, OR-7720)

2BrS 2-Bromostyrene (Combi Blocks, OT-0650)

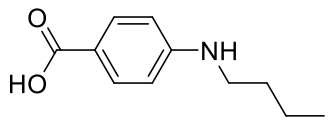
4VBA 4-Vinylbenzoic acid (Combi Blocks, ST-3506)

3EHP 3-Ethenyl-1h-pyrazole (Combi Blocks, QE-0558)

4FMS 4-Fluoro-alpha-methylstyrene (Combi Blocks, QC-4533)

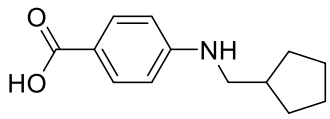
5.5.5. Synthetic procedures

4-(*N*-Butylamino)benzoic acid



4-(*N*-Butylamino)benzoic acid was made by a procedure adapted from literature[119]: 4-Aminobenzoic acid (Sigma-Aldrich, A9878) (0.5 g, 3.65 mmol), butyraldehyde (Sigma-Aldrich, 8.01555.0100) (0.428 mL, 4.75 mmol, 1.3 eq) and 2-Methylpyridine borane complex (Sigma-Aldrich, 654213) (0.411 g, 3.76 mmol, 1.03 eq) were stirred at room temperature in methanol (5 mL) for 14 h. TLC showed that the reaction was complete (TLC system: 10% MeOH/DCM). The reaction mixture was then concentrated and partitioned between EtOAc (7 mL) and aqueous acid (1 N HCl, 2 x 5 mL). The organic fractions were combined, dried over $MgSO_4$ (Sigma-Aldrich, MX0075-1) and concentrated to yield the product as a white powder. NMR spectrum matched literature data: ^{1}H NMR (400 MHz, CDCl3): δ 7.92 (d, J = 8.8 Hz, 2H), 6.55 (d, J = 8.8 Hz, 2H), 3.18 (t, J = 7.2 Hz, 2H), 1.63 (m, 2H), 1.44 (m, 2H), 0.97 (t, J = 7.4 Hz, 3H). HRMS Calcd for $C_{11}H_{16}NO_2$ (M+H): 194.1181 Found: 194.1158.

4-[(Cyclopentylmethyl)amino]benzoic acid



4-Aminobenzoic acid (Sigma-Aldrich, A9878) (0.25 g, 1.823 mmol), cyclopentanecarboxaldehyde 95% (Sigma-Aldrich, 526037) (0.24 mL, 2.188 mmol, 1.2 eq) and 2-Methylpyridine borane complex (Sigma-Aldrich, 654213) (0.22 g, 2 mmol, 1.1 eq) were stirred at room temperature in methanol (10 mL) for 14 hours. TLC of the top liquid showed that the reaction was complete (TLC system: 40% EtOAc/Hex). The resulting precipitate was collected, and the filtrate was acidified with 1 N hydrochloric acid to induce further precipitation. The solids were combined and dried under high vacuum to yield target material as a white powder. ^{1}H NMR (400 MHz, $CDCl_3$): δ 7.92 (d, J = 8.9 Hz, 2H), 6.56 (d, J = 8.9 Hz, 2H), 3.10 (d, J = 7.2 Hz, 2H), 2.17 (sep, J = 7.5 ,1H), 1.88-1.79 (m, 2H), 1.71-1.52 (m, 4H), 1.32-1.21 (m, 2H). ^{13}C NMR (MHz, $CDCl_3$): δ 172.08, 152.96, 132.47, 117.05, 111.44, 48.81, 39.46, 30.73, 25.38. HRMS Calcd for $C_{13}H_{18}NO_2$ (M+H): 220.1337 Found: 220.1326 Calcd for $C_{13}H_{17}NaNO_2$ (M+Na): 242.1156 Found: 242.1146.

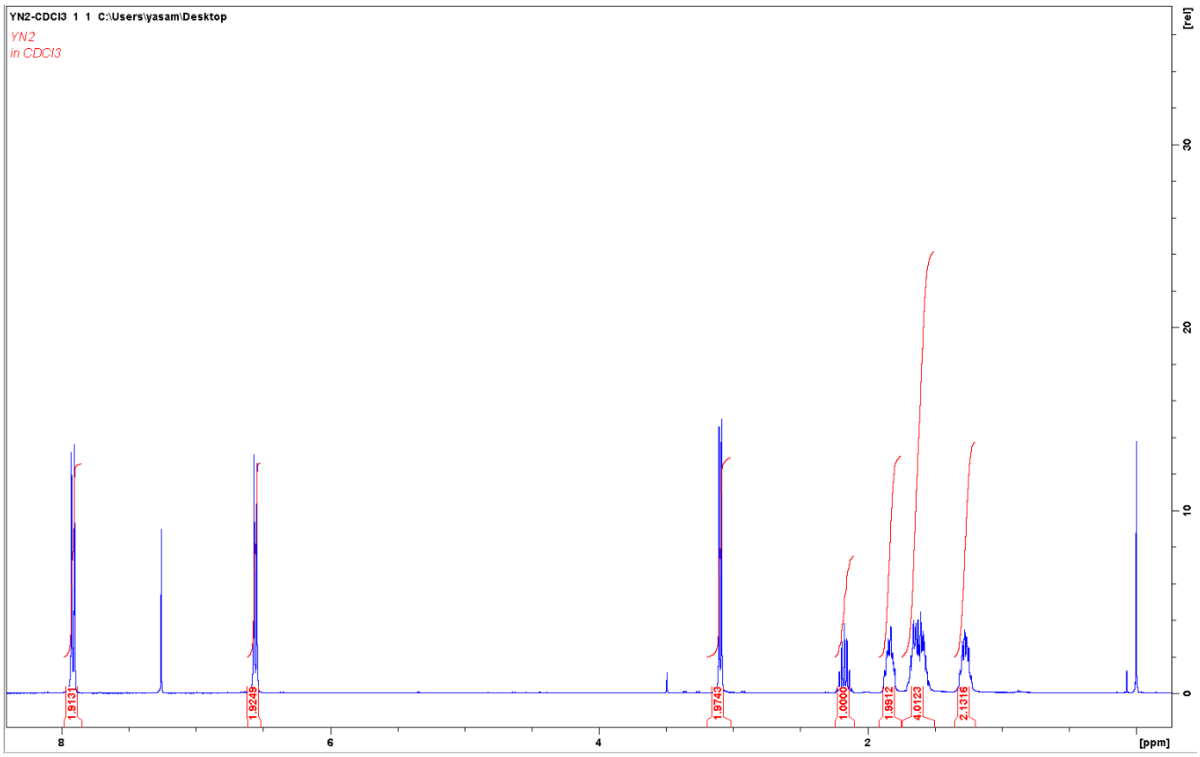


Figure 5-7: ^{1}H NMR spectrum of 4-[(Cyclopentylmethyl)amino]benzoic acid.

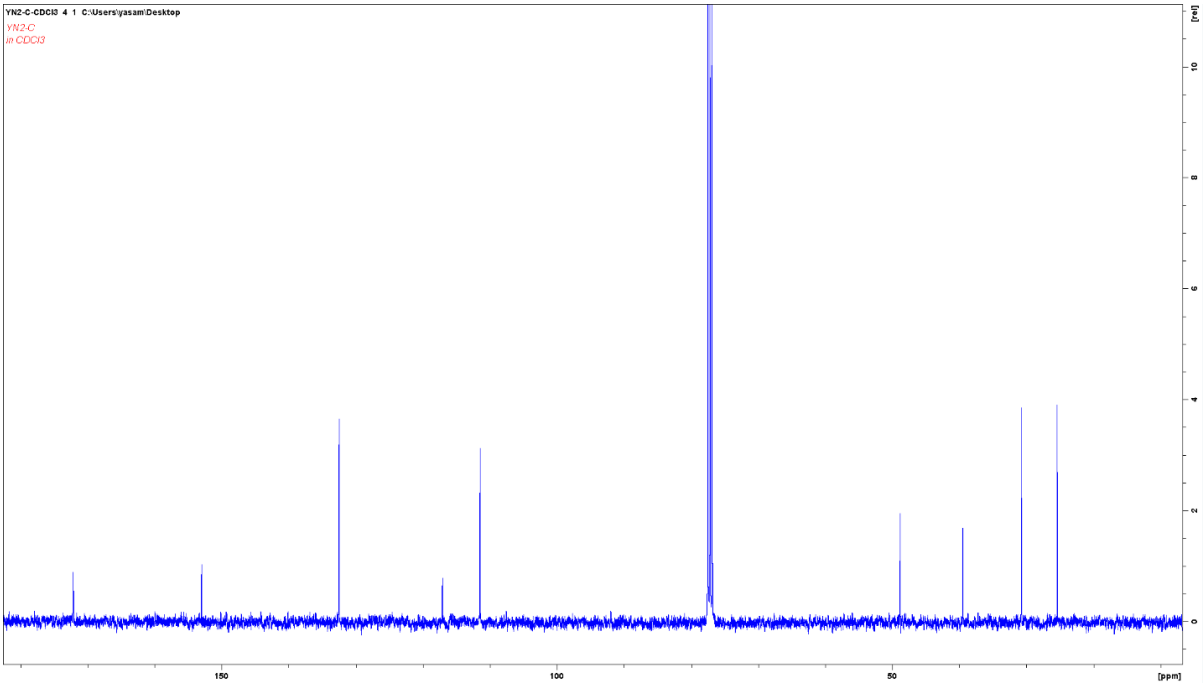
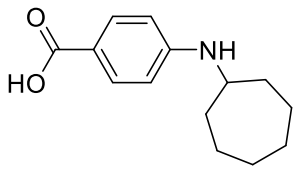


Figure 5-8: ^{13}C NMR spectrum of 4-[(Cyclopentylmethyl)amino]benzoic acid.

4-(Cycloheptylamino)benzoic acid



4-(Cycloheptylamino)benzoic acid was made by a procedure adapted from literature[120]: 4-amino benzoic acid (Sigma-Aldrich, A9878) (0.137 g, 1 mmol), cycloheptanone (Sigma-Aldrich, C99000) (236 uL, 2 mmol), and glacial AcOH (Fisher Scientific, A38-212) (300 uL, 5 mmol) were mixed in 1,2-dichloroethane (4.5 mL). Sodium triacetoxyborohydride (Sigma-Aldrich, 316393) (0.6 g, 2.8 mmol) was added to the above solution and the reaction mixture stirred at room temperature for 27 h. Then cycloheptanone (59 uL, 0.5 mmol), glacial AcOH (75 uL, 1.25 mmol), 1,2-dichloroethane (1.5 mL) and sodium triacetoxyborohydride (0.15 g, 0.7 mmol) were again added to the reaction mixture and the reaction stirred at room temperature for another 5 h after which TLC showed that the reaction was complete (TLC system: 40% Hex/EtOAc). The reaction was quenched with saturated aqueous $NaHCO_3$ (Fisher Chemical, S233-500), then the product was extracted with EtOAc (3 × 7.5 mL). The EtOAc extracts were combined, dried over $MgSO_4$ (Sigma-Aldrich, MX0075-1) and concentrated to yield the crude product as a white powder. The product was triturated with ether/hexane (7:3) and the solid was filtered. The pure sample was then recrystallized from EtOAc/hexane. NMR spectrum matched literature data: ^{1}H NMR (DMSO-d_6): δ 11.95 (s, 1H), 7.68 (d, J = 8.6 Hz, 2H), 6.52 (d, J = 8.6 Hz, 2H), 6.29 (d, J = 7.4 Hz, 1H), 3.45 (bs, 1H), 1.94-1.82 (m, 2H), 1.71-

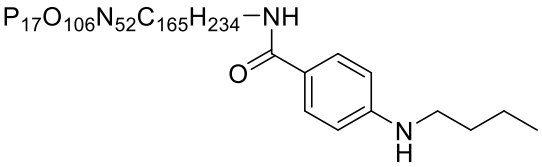
1.39 (m, 10H). HRMS Calcd for $C_{14}H_{20}NO_2$ (M+H): 234.1494 Found: 234.1494. Calcd for $C_{14}H_{19}NaNO_2$ (M+Na): 256.1314 Found: 256.1317.

5.5.6. General procedure for the preparation of DNA conjugates

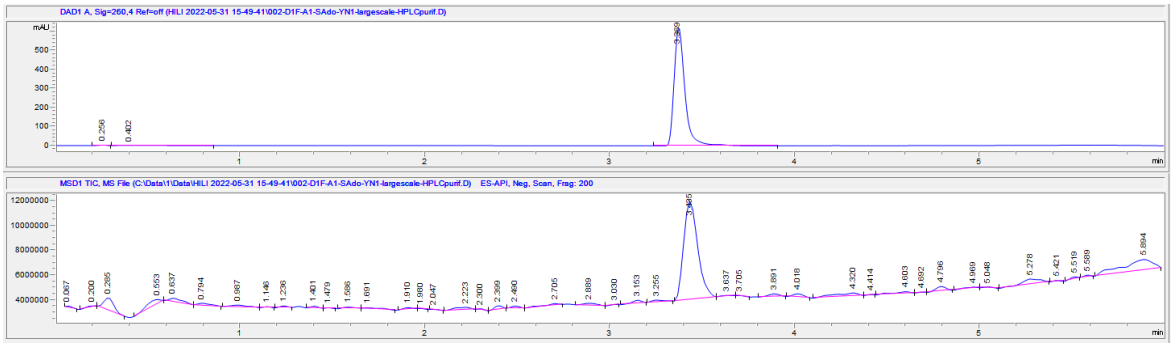
HATU (Combi Blocks, OR-0618) (400 uL of 0.2 M in DMF), DIPEA (Alfa Aesar, A11801) (400 uL of 0.2 M in DMF) and the respective carboxylic acid (400 uL of 0.2 M in DMF) were mixed. The stock was cooled at 4 ^{o}C for 10 minutes then transferred to 1000 uL of 1 mM solution of DNA headpiece in 250 mM sodium phosphate buffer (pH=9.4). The resulting solution was shaken at room temperature. After 16 h the DNA was recovered from the mixture by ethanol precipitation and then purified by HPLC.

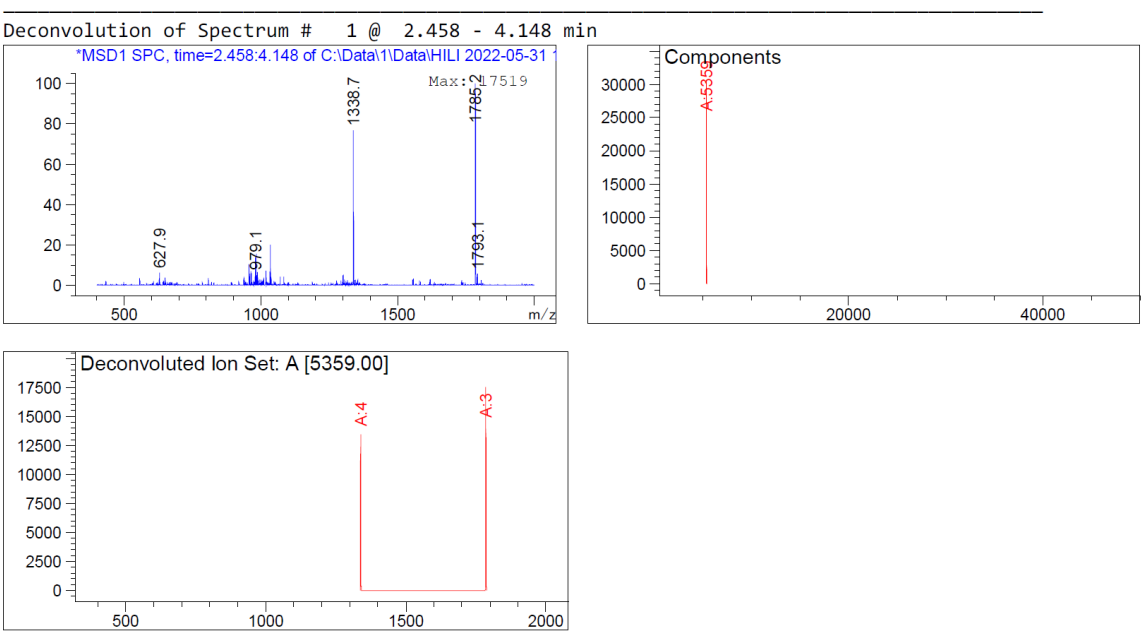
DNA conjugate 1a:

1a was synthesized according to the general procedure using 4-(*N*-Butylamino) benzoic acid.



Molecular Weight: 5359.7530 D



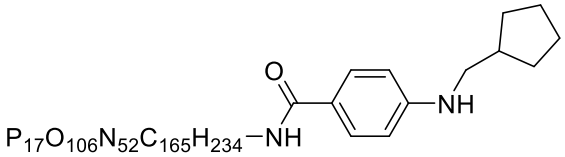


Component	Molecular Weight	Absolute Abundance	Relative Abundance
A	5359.00	30540	100.00

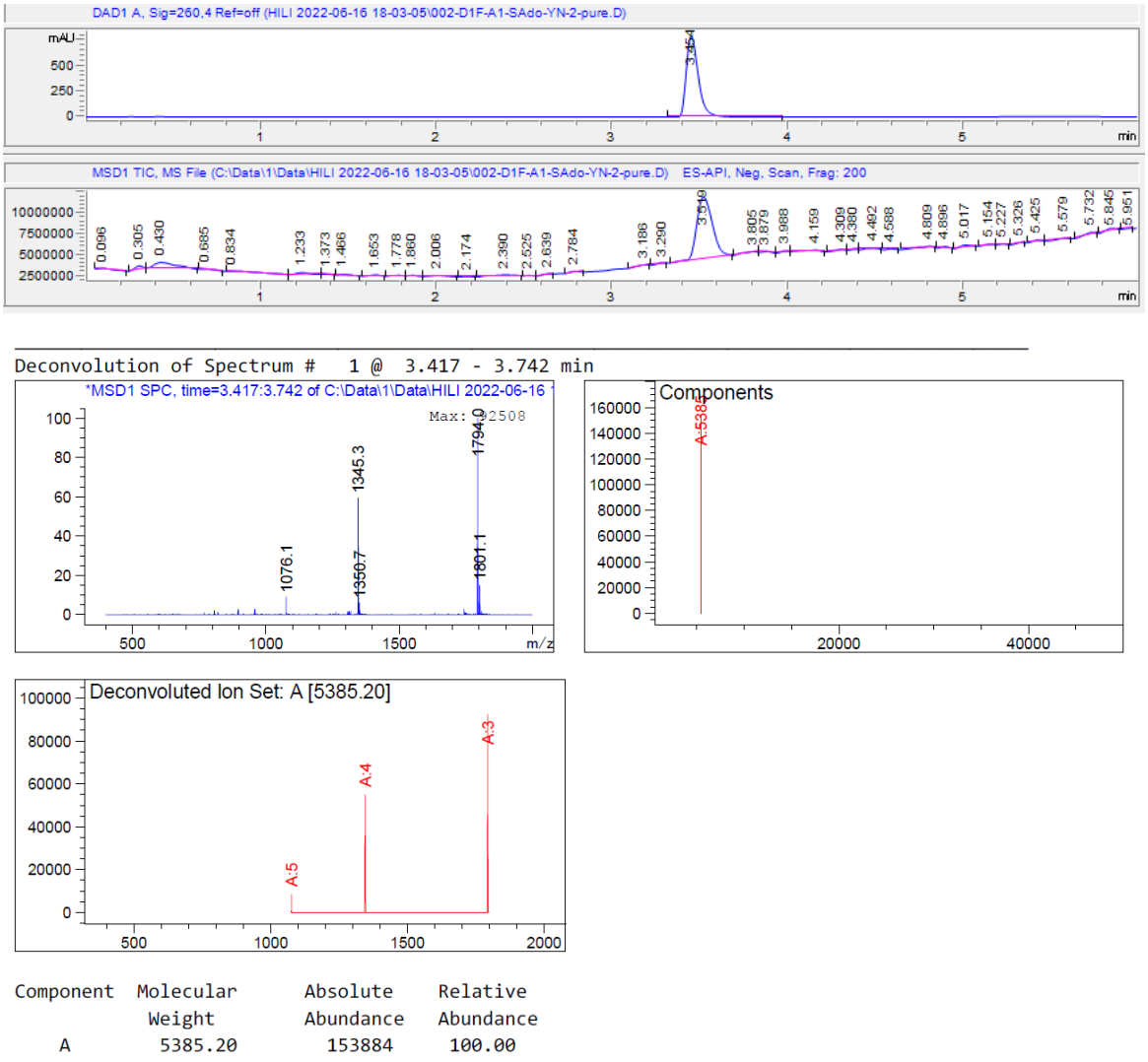
Figure 5-9: Deconvoluted LCMS data for DNA conjugate **1a**.

DNA conjugate 1b:

1b was synthesized according to the general procedure using 4-[(Cyclopentyl methyl) amino] benzoic acid.



Molecular Weight: 5385.7910 D

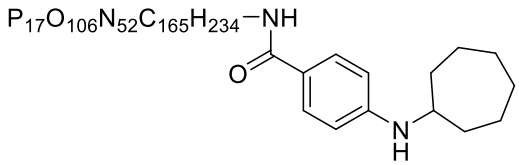


Component	Molecular Weight	Absolute Abundance	Relative Abundance
A	5385.20	153884	100.00

Figure 5-10: Deconvoluted LCMS data for DNA conjugate **1b**.

DNA conjugate 1c:

1c was synthesized according to the general procedure using 4-(Cycloheptylamino) benzoic acid.



Molecular Weight: 5399.8180 D

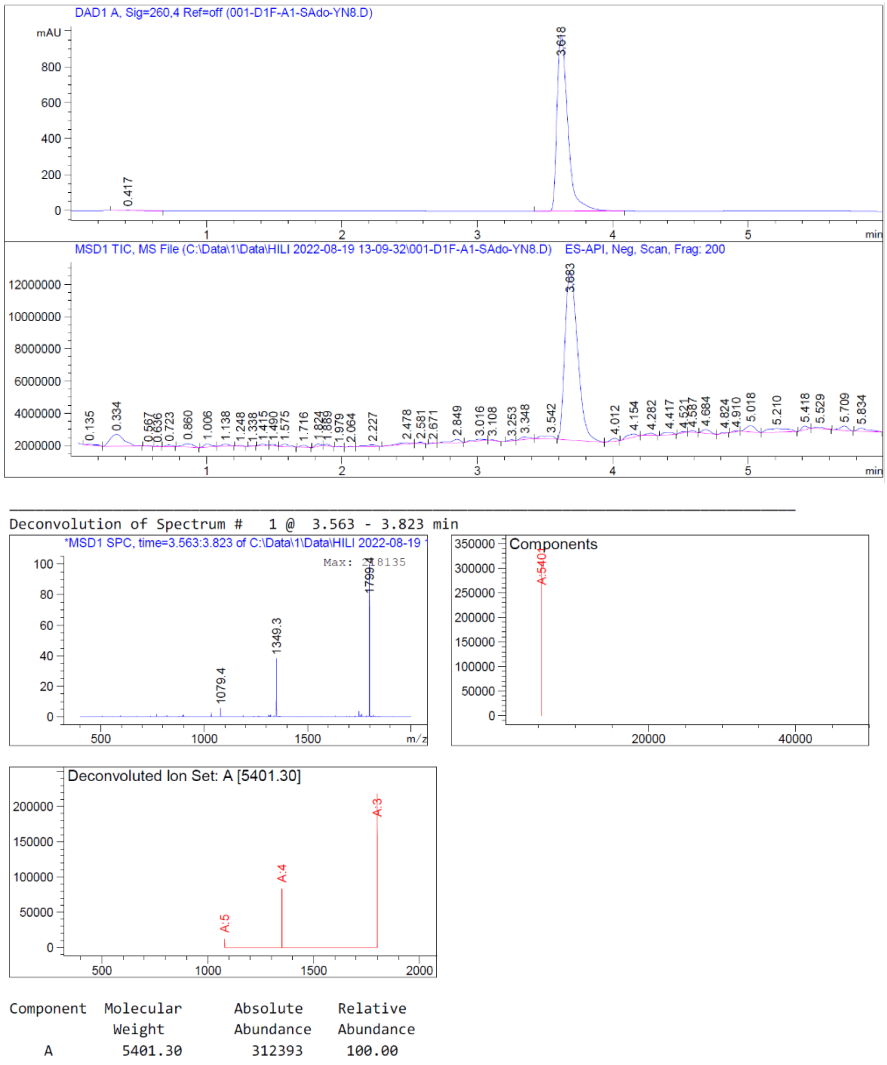
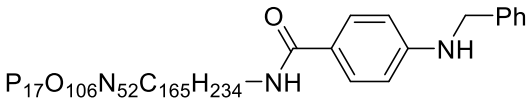


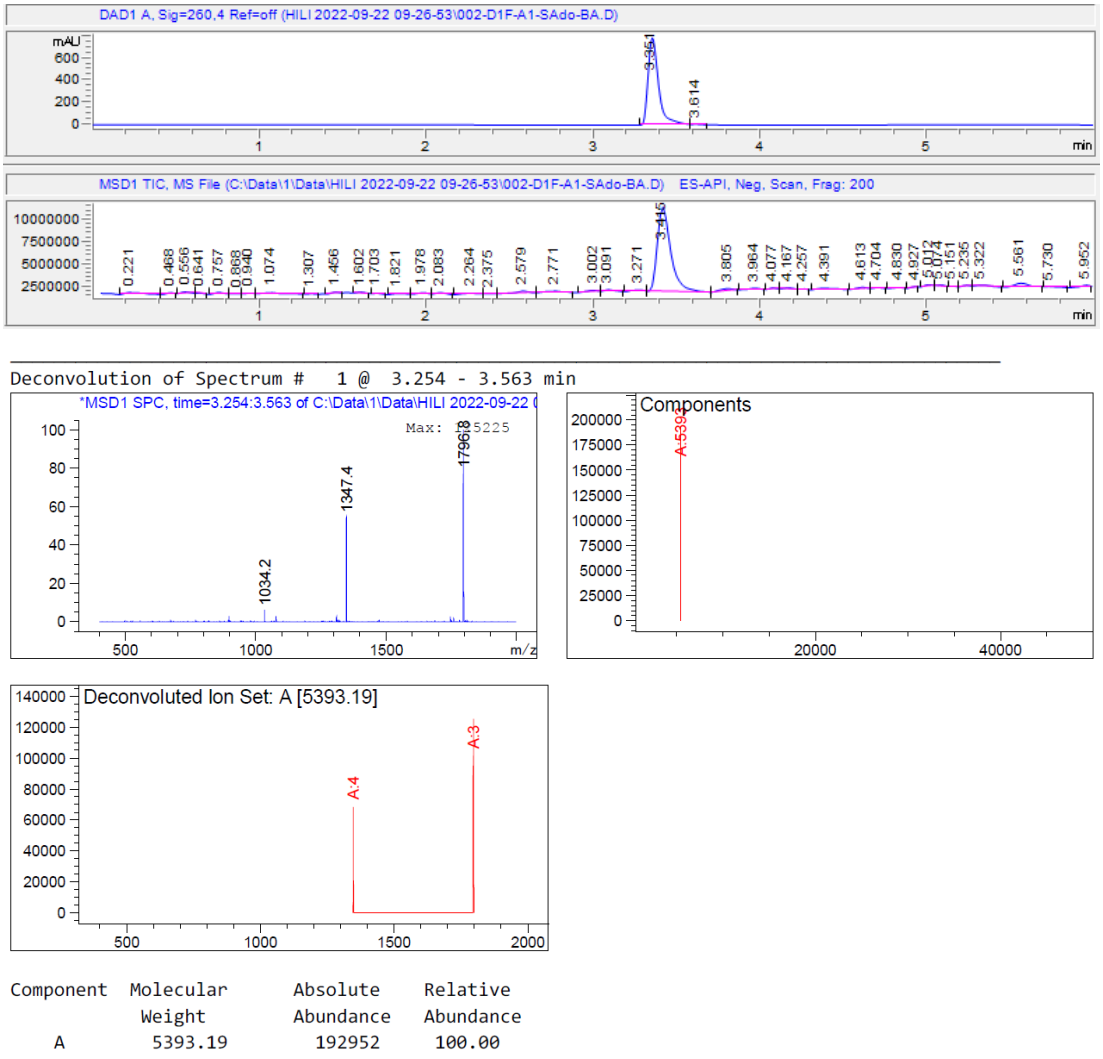
Figure 5-11: Deconvoluted LCMS data for DNA conjugate **1c**.

DNA conjugate 1d:

1d was synthesized according to the general procedure using 4-(Benzylamino) benzoic acid (Sigma Aldrich, L127728).



Molecular Weight: 5393.7700 D

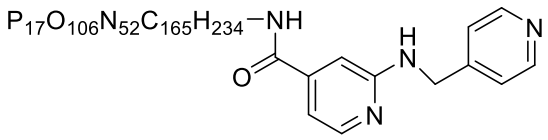


Component	Molecular Weight	Absolute Abundance	Relative Abundance
A	5393.19	192952	100.00

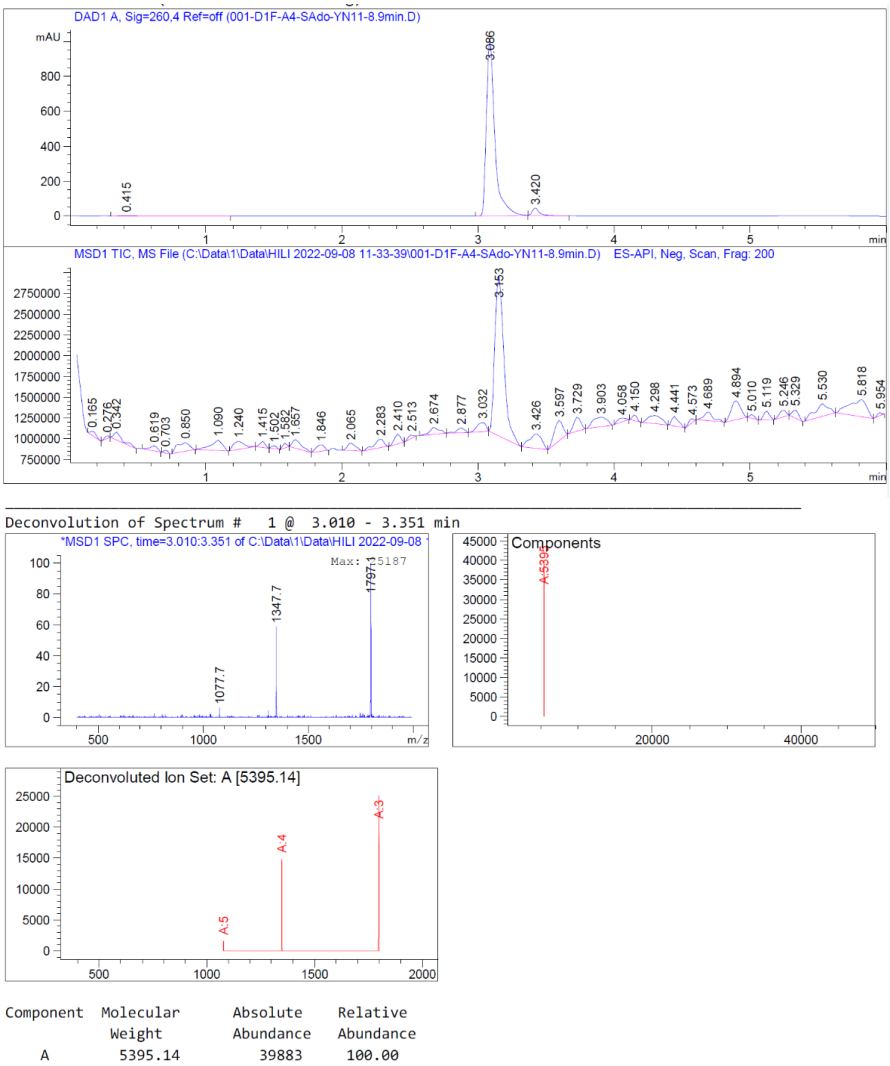
Figure 5-12: Deconvoluted LCMS data for DNA conjugate **1d**.

DNA conjugate 1e:

1e was synthesized according to the general procedure using 2-[(4-Pyridinylmethyl) amino] isonicotinic acid (Sigma-Aldrich, CDS021130).



Molecular Weight: 5395.7460 D

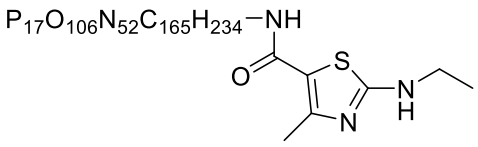


Component	Molecular Weight	Absolute Abundance	Relative Abundance
A	5395.14	39883	100.00

Figure 5-13: Deconvoluted LCMS data for DNA conjugate **1e**.

DNA conjugate 1f:

1f was synthesized according to the general procedure using 2-(Ethylamino)-4-methyl-1,3-thiazole-5-carboxylic acid (Sigma-Aldrich, CBR00568).



Molecular Weight: 5352.7360 D

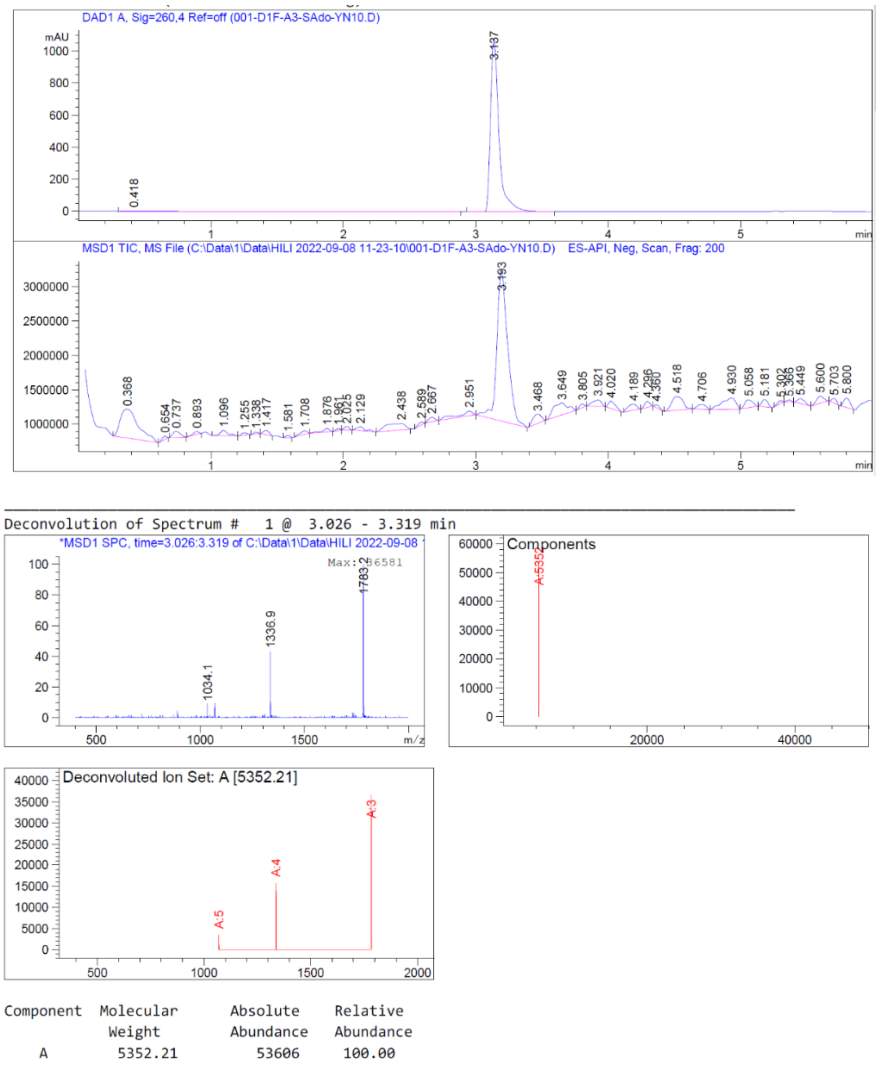


Figure 5-14: Deconvoluted LCMS data for DNA conjugate **1f**.

5.5.7. HPLC purification

HPLC purifications were conducted on a 1260 Infinity II LC System from Agilent.

HPLC method:

flow rate: 4 mL/min

Detection wavelength: 260 nm

mobile phase A: 0.1 M triethylammonium acetate (TEAA)

mobile phase B: Acetonitrile

Elapsed time (min)	%B
0	10
10	20
23	45
26	80
28	80
29	10
31	10

Column: Agilent 5µm Eclipse XDB-C18 9.4 x 250 mm

5.5.8. General procedure for ethanol precipitation

To the reaction mixture containing DNA, was added 10% (V/V) 4 M NaCl and 3 times the volume ethanol. The solution was placed on dry ice for 1 hour and then centrifuged at full speed (25000 x g), at 4 ºC for 30 minutes. the supernatant was removed, and the pellet was washed with 75% aq. ethanol and then air-dried.

5.5.9. Photocatalysis reaction setup

In a PCR tube was added 10 nmol of DNA conjugate (in 10 µL H_2O), quinuclidine (TCI America, Q0062) (10 µL of 500 mM in DMF), alkene (10 µL of 250 mM in DMF), and Iridium catalyst (10 µL of 1 mM in DMF). The solution was degassed in glove box for 2 hours and then placed approximately 10 cm from blue light (highest intensity) with cooling. After 1.5 h, the DNA was recovered from the reaction mixture by Ethanol precipitation. Pellet was air-dried and resuspended in 100 µL water and 5 µL of the resulting solution was injected to LCMS.

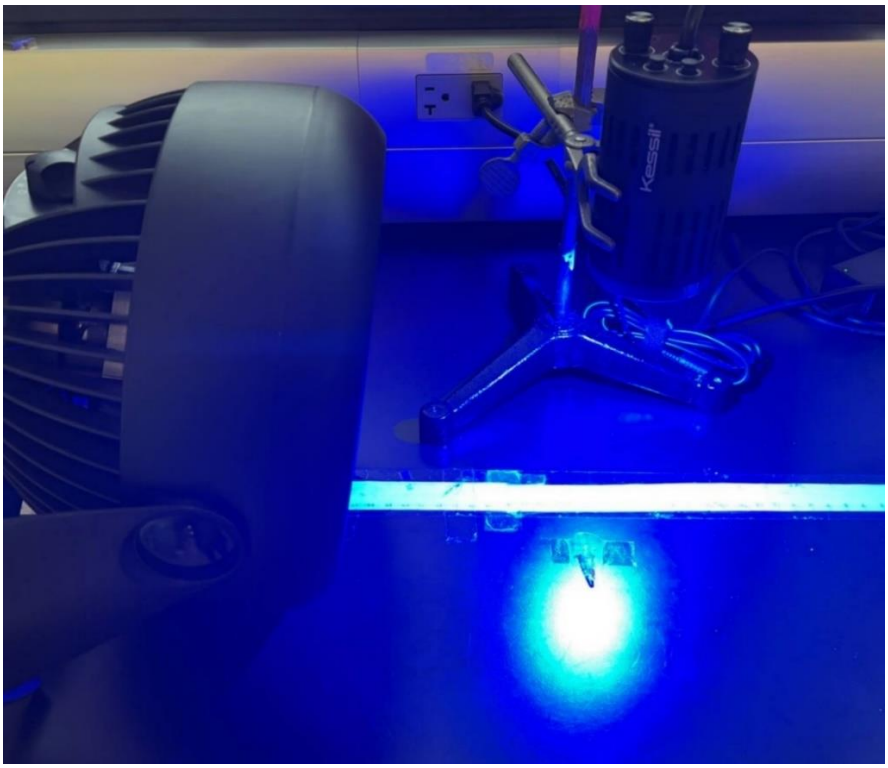


Figure 5-15: Reaction setup. Sample was secured 10 cm from Kessil Tuna Blue A160WE lamp set to the highest intensity. A fan was situated directly behind the reaction vessel to dissipate heat.

5.5.10. LCMS analysis

LCMS analyses were performed using Agilent Infinity Lab LC/MSD system.

LCMS method:

Flow rate: 0.5 mL/min

Detection wavelength: 260 nm

mobile phase A: 10 µM EDTA, 0.38% TEAA pH 7, 0.75% HFIP, in 90:10 methanol:water

Mobile phase B: 10 µM EDTA, 0.38% TEAA pH 7, 0.75% HFIP, in water

Elapsed time (min)	%B
0	90
4	10
5	90
6	90

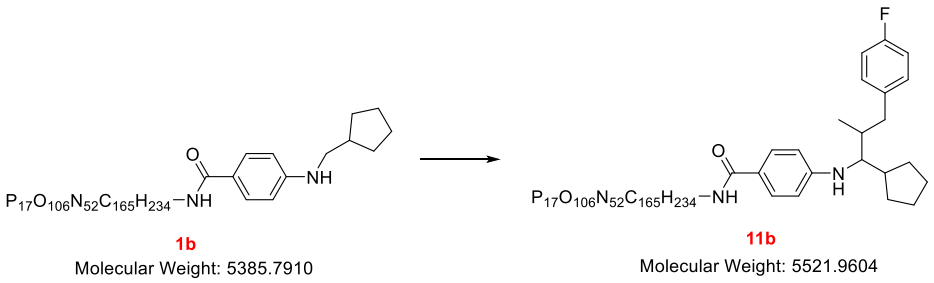
Column: HALO 400 A, ES-C18, 3.4 uM, 2.1 x 30 mm

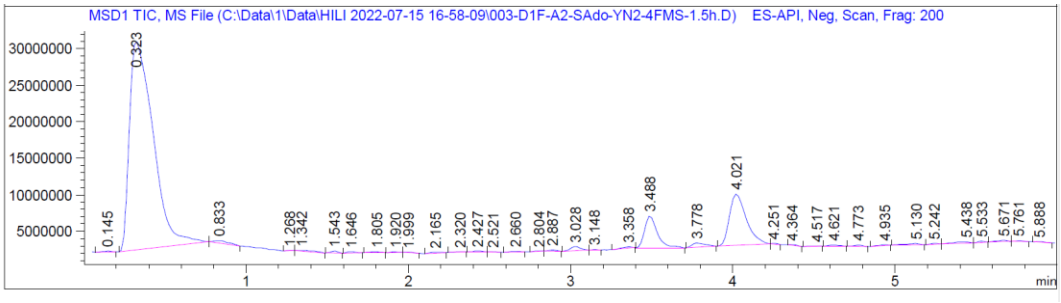
Conversion calculations for on-DNA reactions through LCMS:

Yields are reported as a % conversion as determined from LCMS analysis by comparing the abundance of all DNA-derived compounds.

$$\%\ \text{Conversion} = \frac{\text{Total abundance of target material}}{\text{Total abundance of DNA material}}\ \text{x}\ 100$$

Example of LCMS data calculations:





Component	Molecular Weight	Absolute Abundance	Relative Abundance
A	5521.30	65013	100.00
B	5385.22	32602	50.15

A (%) = percent single addition = 100 / (100+50.15) = 67%

B (%) = percent starting material = 50.15 / (100+50.15) = 33%

Figure 5-16: An example of conversion calculations.

5.5.11. Analysis of HAT catalyst requirement

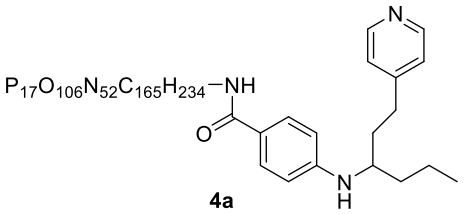
Table 5-5: Examination of HAT catalyst (quinuclidine) dependence on reaction

		SM	Single Addn	Double Addn	Triple Addn	Unknown	Dealkylation
1	**1d** + 4VP no quinuclidine	-	90%	-	-	-	10%
2	**1d** + DPE no quinuclidine	59%	16%	-	-	5482.31: 10%	15%
3	**1d** + 4VP with quinuclidine	-	68%	25%	7%	-	-
4	**1d** + DPE with quinuclidine	77%	23%	-	-	-	-

5.5.12. LCMS spectra and deconvolution results for 1a derivatives

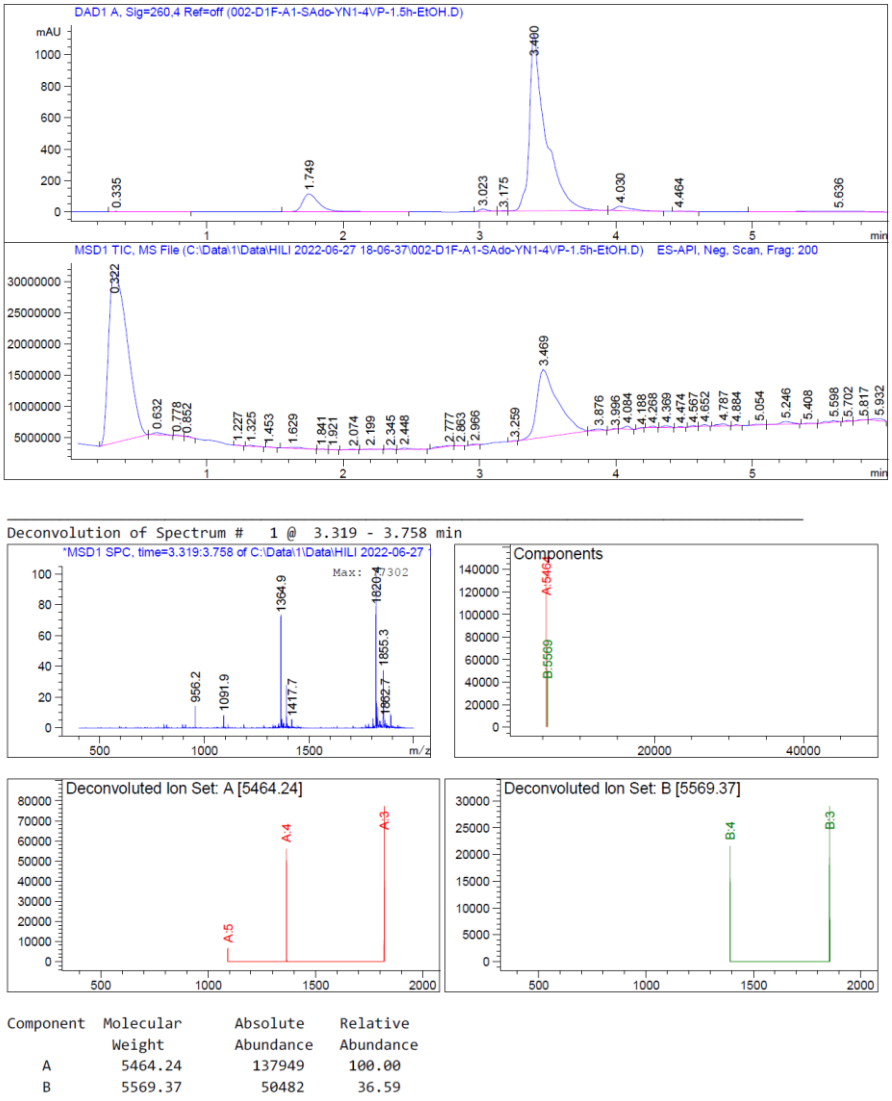
Table 5-6: Hydroaminoalkylation of various vinylarenes with DNA conjugate **1a**

	Starting Material (**1a**)	Single Addition	Double Addition	Triple Addition
1a + 4VP		**4a**: 73%	27%	-
1a + 4CS	8%	**7a**: 76%	16%	-
1a + 2VhB	15%	**6a**: 71%	14%	-
1a + 2BrS	15%	**5a**: 76%	9%	-
1a + diFP	14%	**8a**: 86%	-	-
1a + DPE	25%	**3a**: 75%	-	-
1a + 3EhP	49%	**9a**: 51%	-	-
1a + 4M5VT	51%	**10a**: 49%	-	-
1a + 4FMS	66%	**11a**: 34%	-	-
1a + 5EMP	68%	**12a**: 32%	-	-
1a + 4MS	73%	**13a**: 27%	-	-
1a + 4VBA	76%	**14a**: 24%	-	-
1a + 4AS	100%	-	-	-



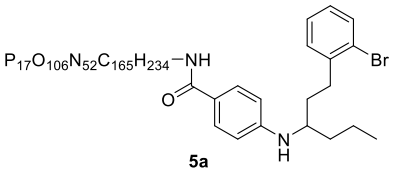
4a

Molecular Weight: 5464.8930



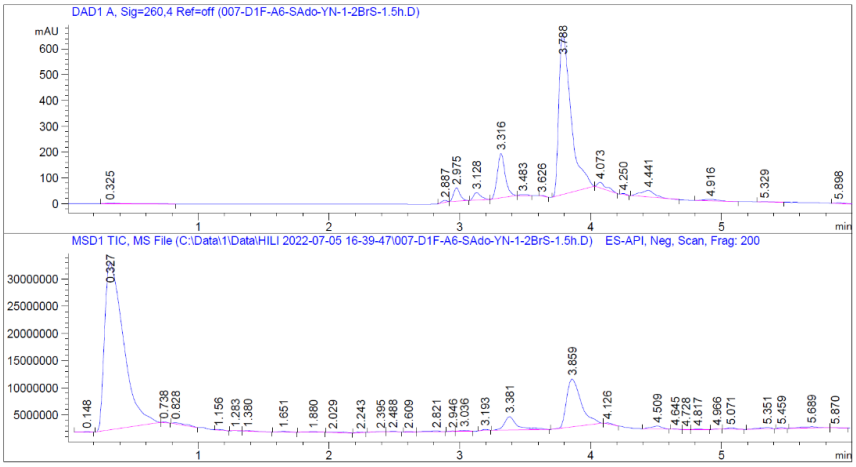
Component	Molecular Weight	Absolute Abundance	Relative Abundance
A	5464.24	137949	100.00
B	5569.37	50482	36.59

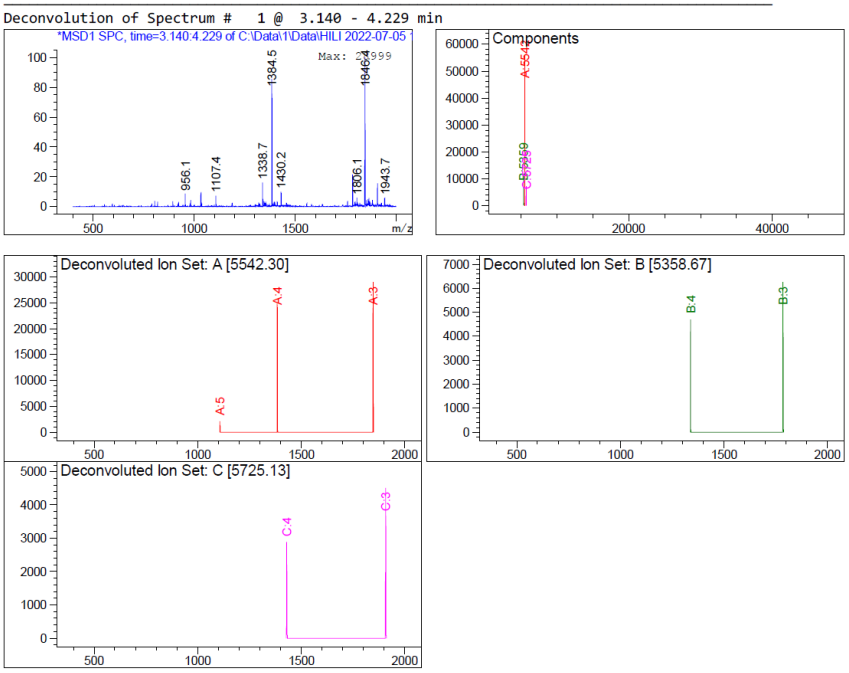
Figure 5-17**:** Deconvoluted LCMS data for **4a**.



5a

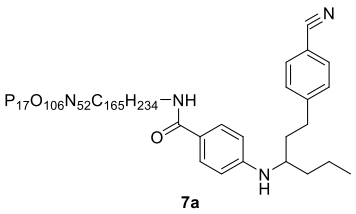
Molecular Weight: 5542.8010





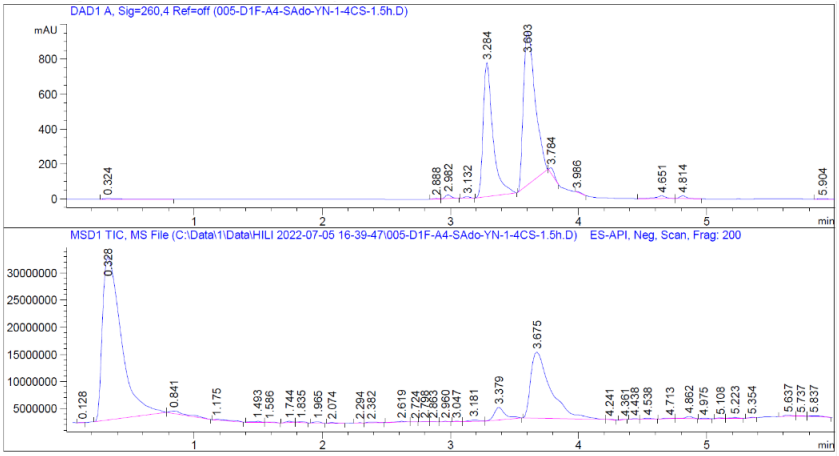
Component	Molecular Weight	Absolute Abundance	Relative Abundance
A	5542.30	55761	100.00
B	5358.67	10836	19.43
C	5725.13	7163	12.85

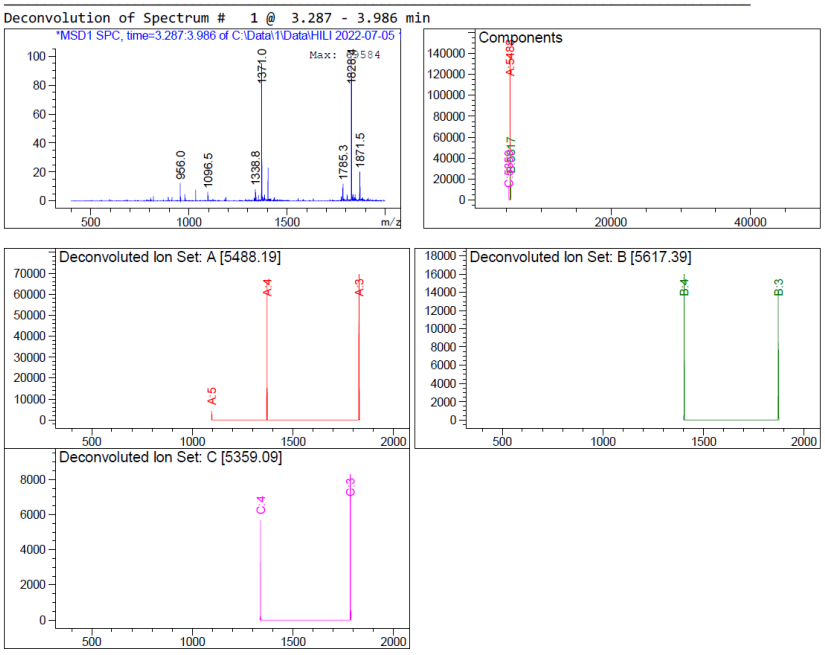
Figure 5-18: Deconvoluted LCMS data for **5a**.



7a

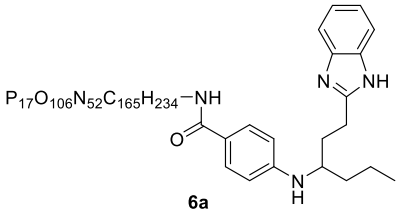
Molecular Weight: 5488.9150





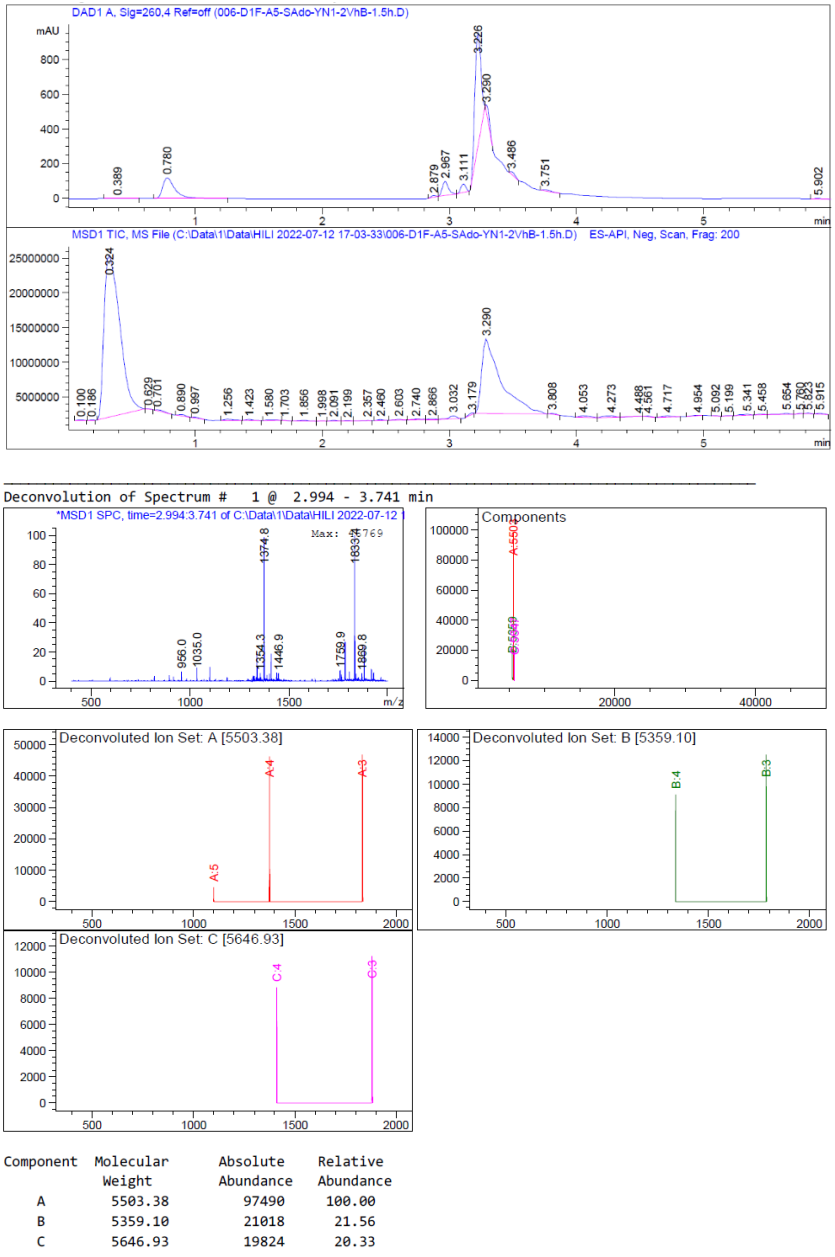
Component	Molecular Weight	Absolute Abundance	Relative Abundance
A	5488.19	139142	100.00
B	5617.39	30006	21.57
C	5359.09	13902	9.99

Figure 5-19: Deconvoluted LCMS data for **7a**.



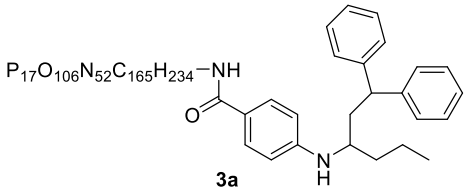
6a

Molecular Weight: 5503.9300



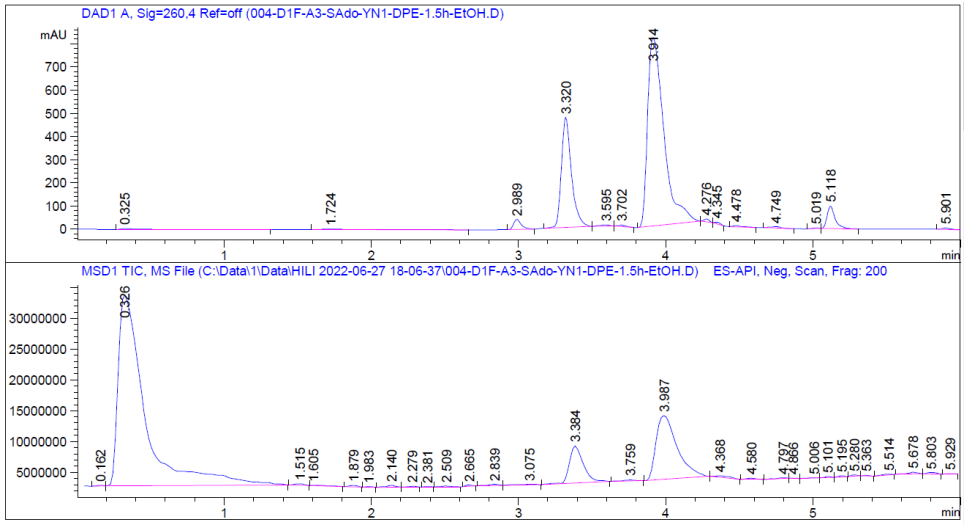
Component	Molecular Weight	Absolute Abundance	Relative Abundance
A	5503.38	97490	100.00
B	5359.10	21018	21.56
C	5646.93	19824	20.33

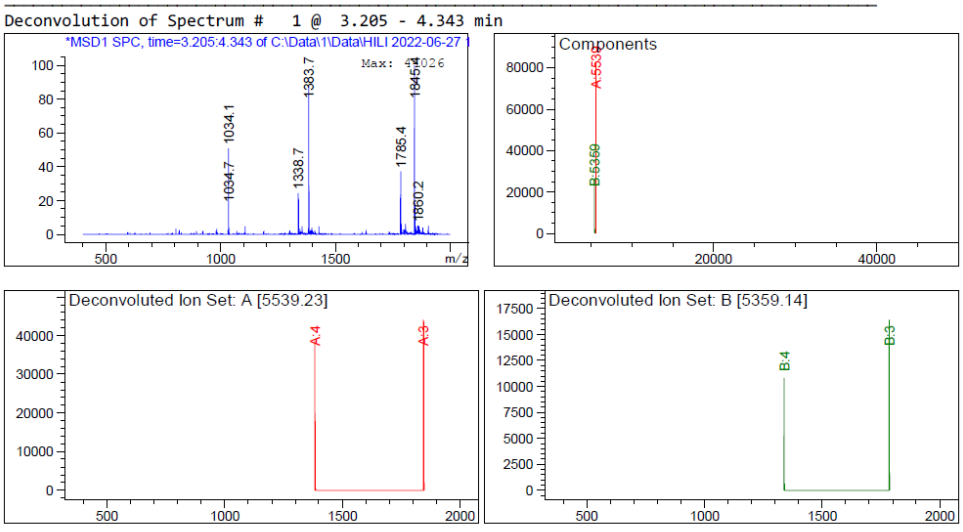
Figure 5-20: Deconvoluted LCMS data for **6a**.



3a

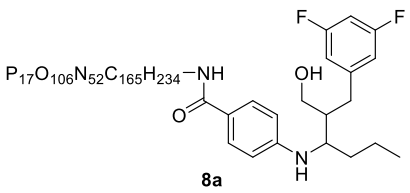
Molecular Weight: 5540.0030





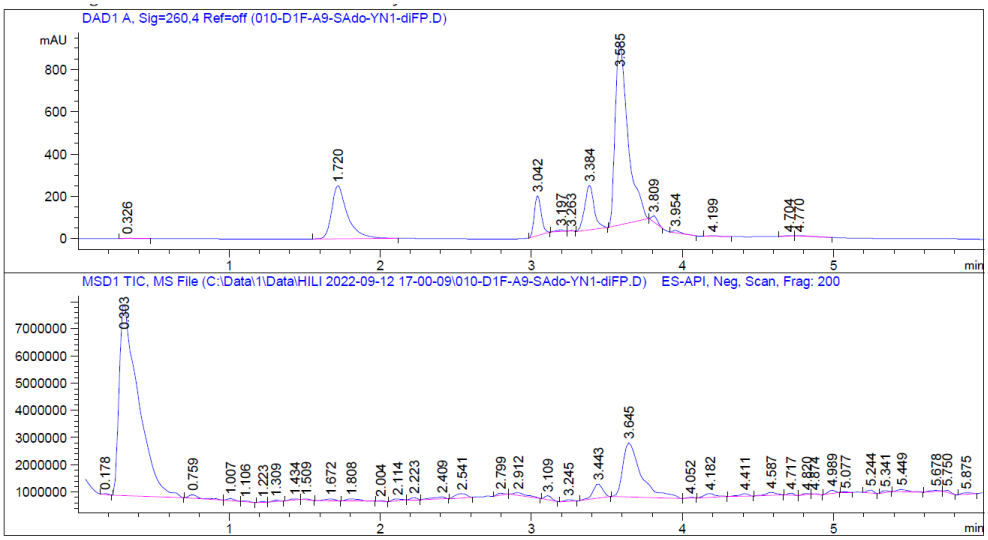
Component	Molecular Weight	Absolute Abundance	Relative Abundance
A	5539.23	82233	100.00
B	5359.14	26802	32.59

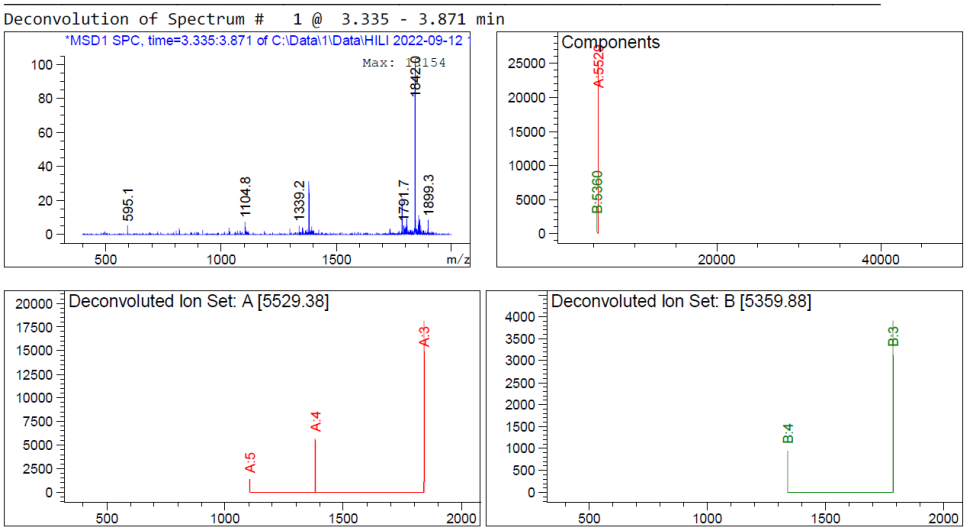
Figure 5-21: Deconvoluted LCMS data for **3a**.



8a

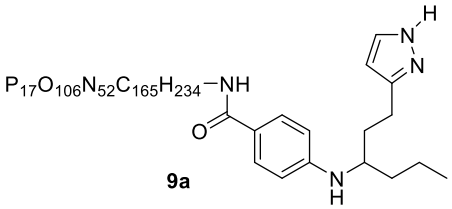
Molecular Weight: 5529.9118



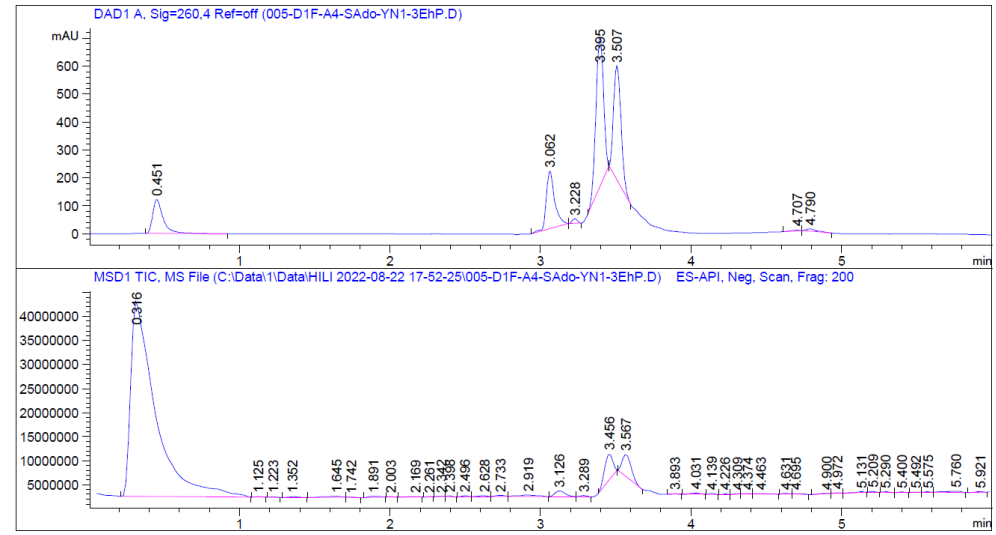


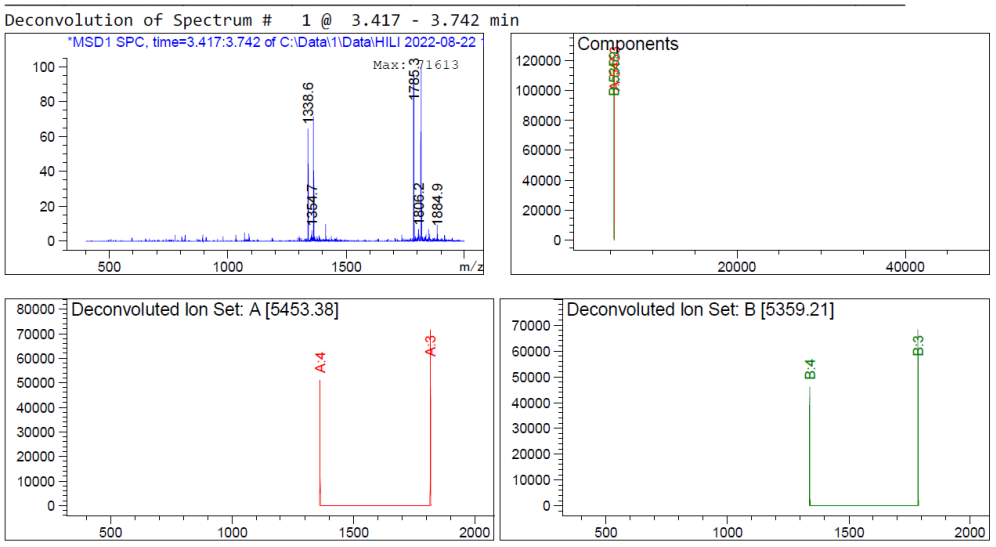
Component	Molecular Weight	Absolute Abundance	Relative Abundance
A	5529.38	25190	100.00
B	5359.88	4027	15.99

Figure 5-22: Deconvoluted LCMS data for **8a**.



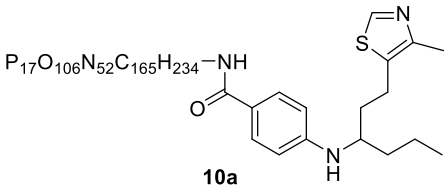
Molecular Weight: 5453.8700





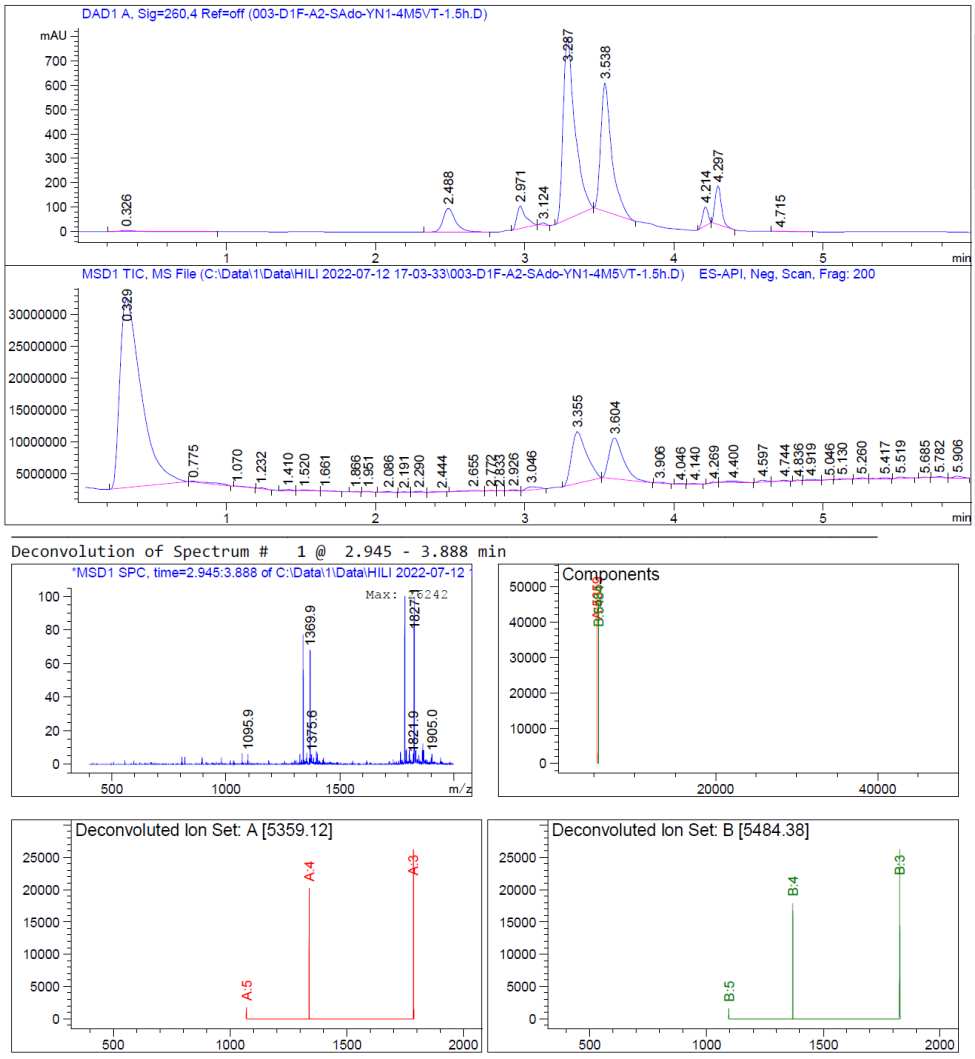
Component	Molecular Weight	Absolute Abundance	Relative Abundance
A	5453.38	117571	100.00
B	5359.21	112927	96.05

Figure 5-23: Deconvoluted LCMS data for **9a**.



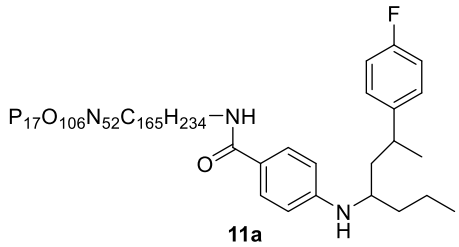
10a

Molecular Weight: 5484.9420



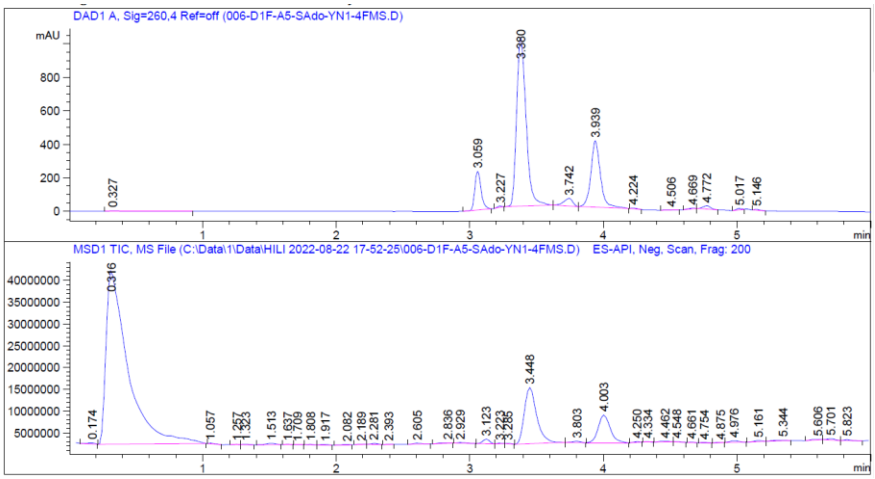
Component	Molecular Weight	Absolute Abundance	Relative Abundance
A	5359.12	47998	100.00
B	5484.38	45285	94.35

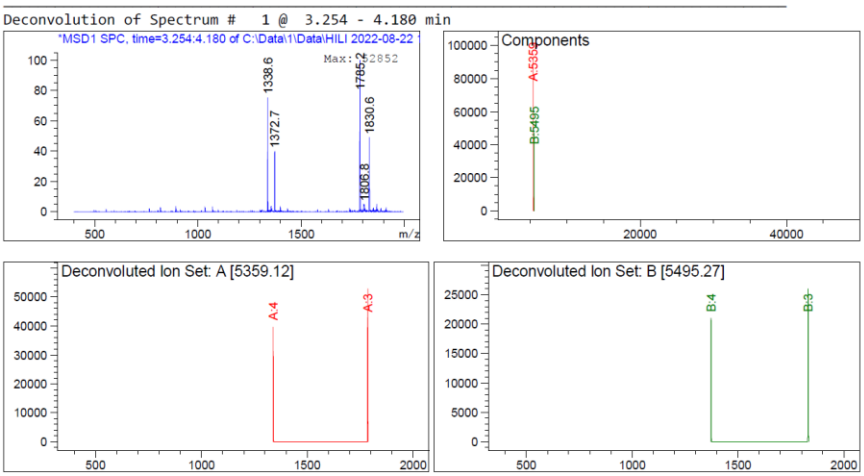
Figure 5-24: Deconvoluted LCMS data for **10a**.



11a

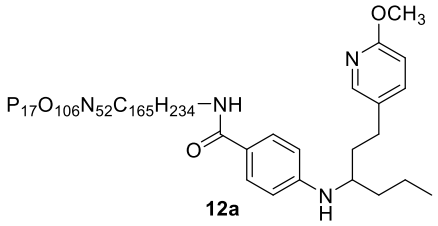
Molecular Weight: 5495.9224





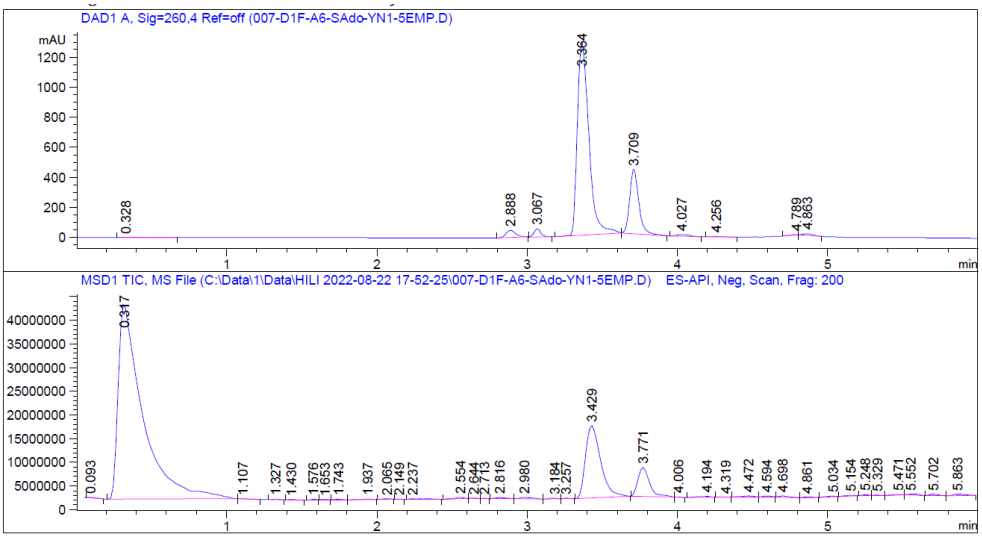
Component	Molecular Weight	Absolute Abundance	Relative Abundance
A	5359.12	92400	100.00
B	5495.27	47097	50.97

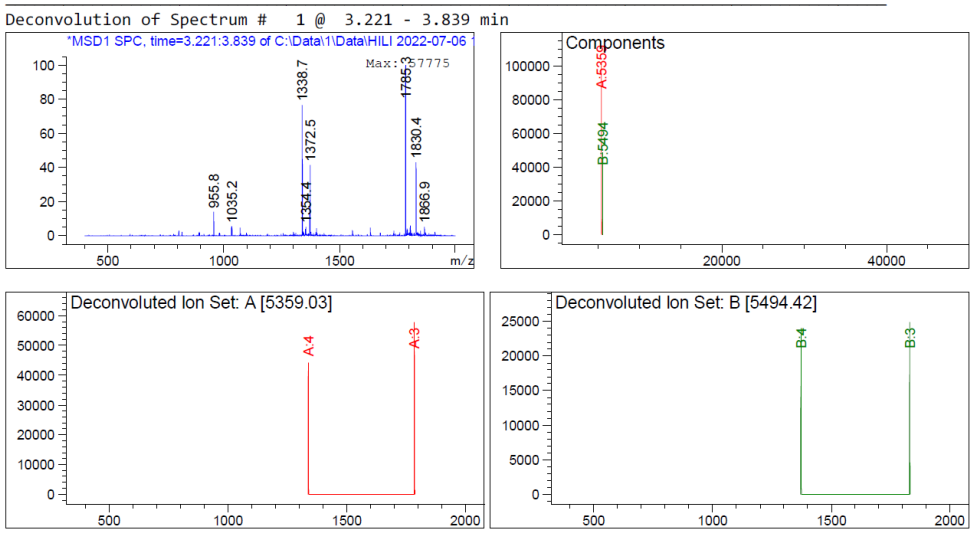
Figure 5-25: Deconvoluted LCMS data for **11a**.



12a

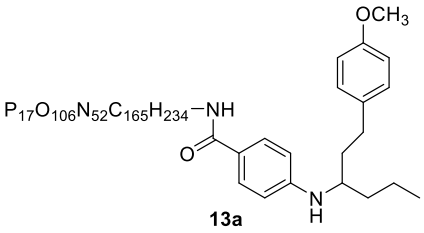
Molecular Weight: 5494.9190





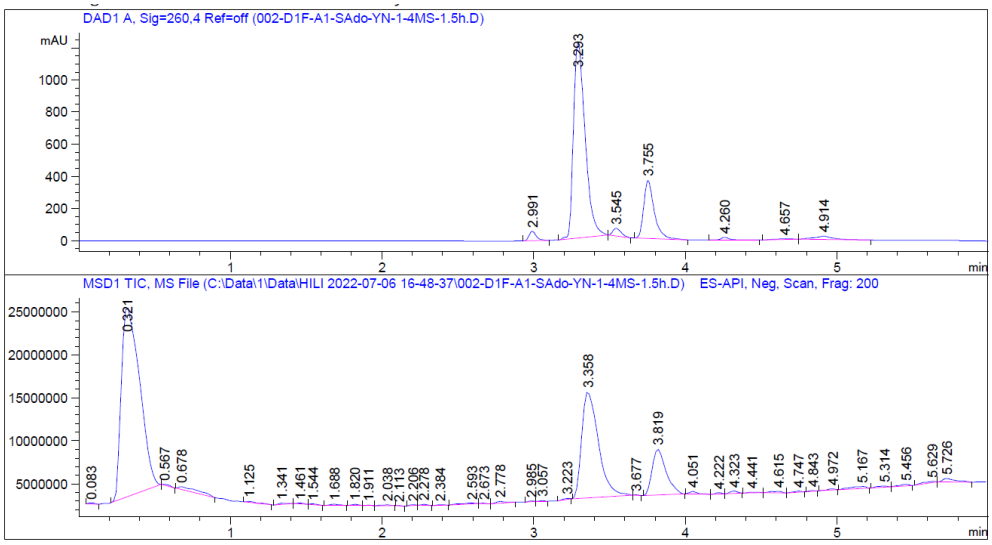
Component	Molecular Weight	Absolute Abundance	Relative Abundance
A	5359.03	101842	100.00
B	5494.42	48620	47.74

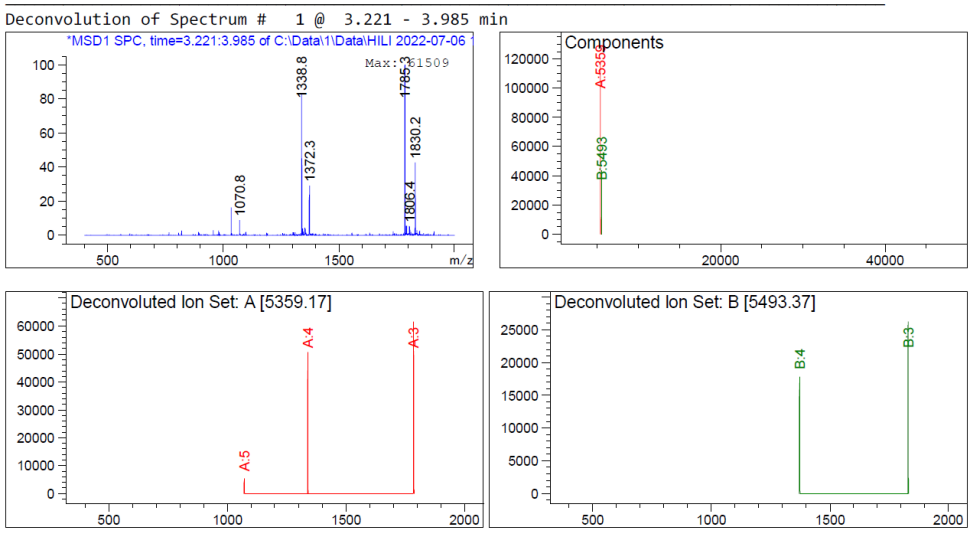
Figure 5-26: Deconvoluted LCMS data for **12a**.



13a

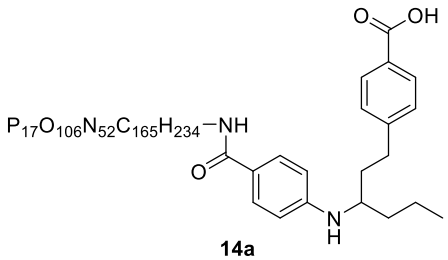
Molecular Weight: 5493.9310





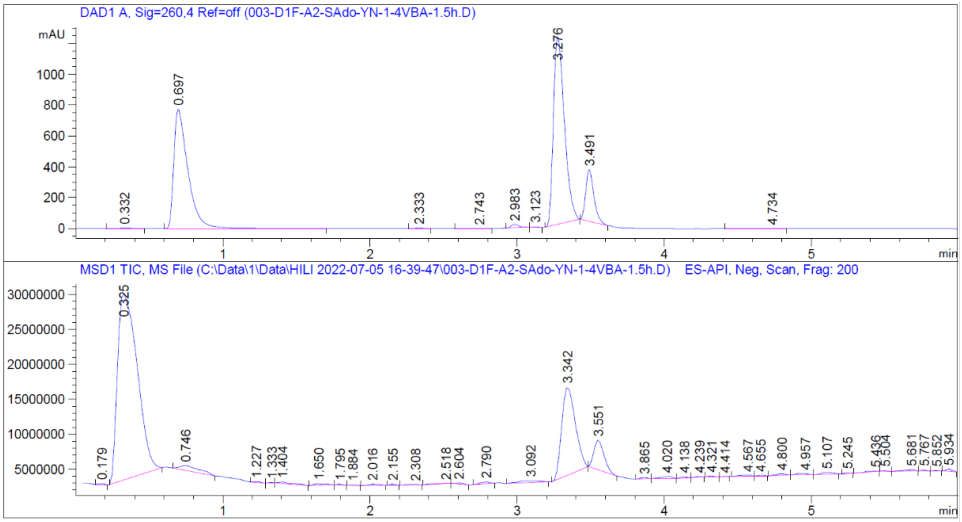
Component	Molecular Weight	Absolute Abundance	Relative Abundance
A	5359.17	117613	100.00
B	5493.37	43497	36.98

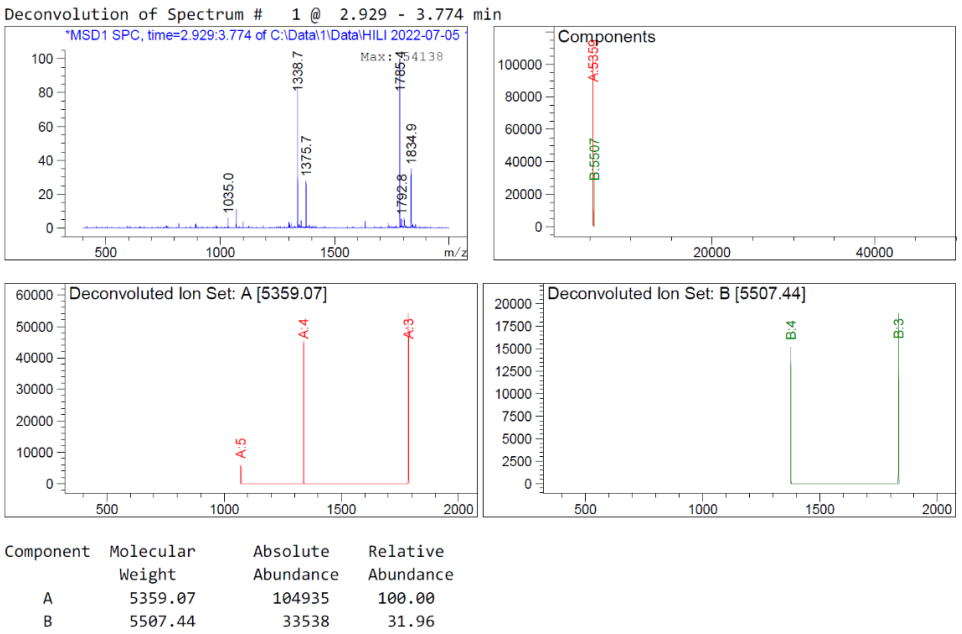
Figure 5-27: Deconvoluted LCMS data for **13a**.



14a

Molecular Weight: 5507.9140





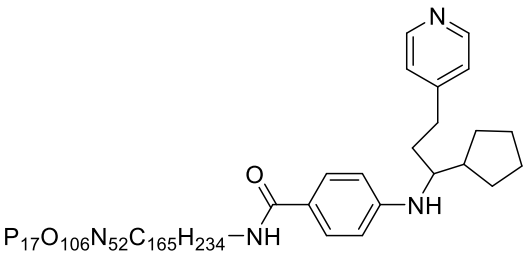
Component	Molecular Weight	Absolute Abundance	Relative Abundance
A	5359.07	104935	100.00
B	5507.44	33538	31.96

Figure 5-28: Deconvoluted LCMS data for **14a**.

5.5.13. LCMS spectra and deconvolution results for 1b derivatives

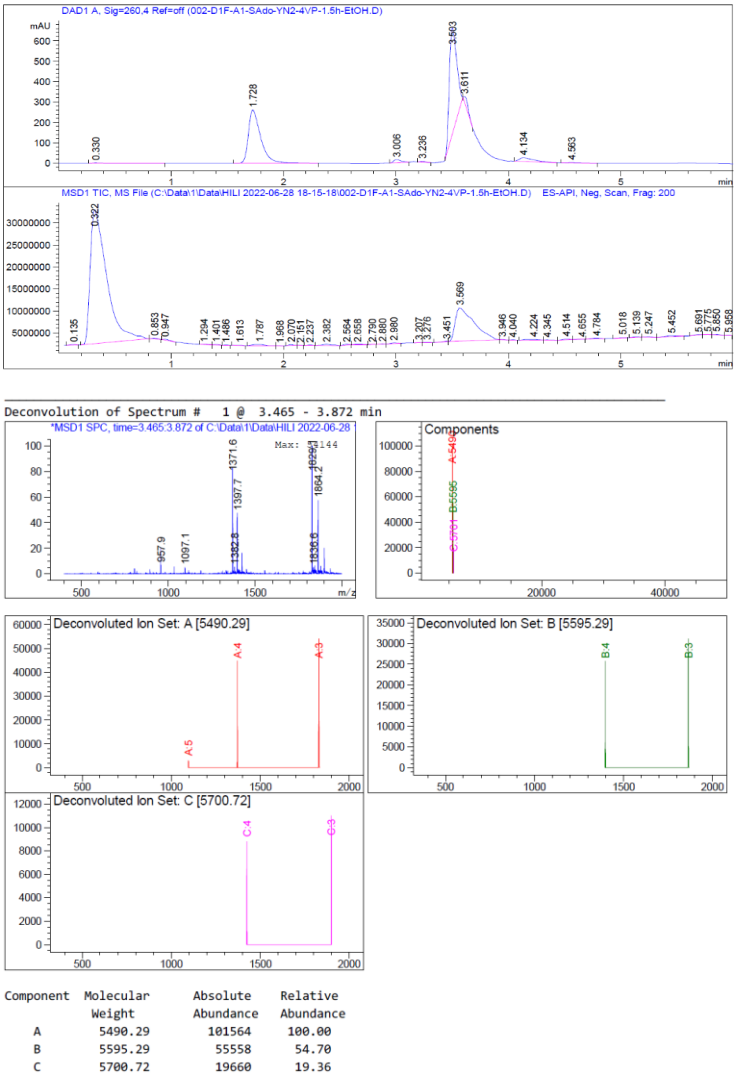
Table 5-7: Hydroaminoalkylation of various vinylarenes with DNA conjugate **1b**

	Starting Material (**1b**)	Single Addition	Double Addition	Triple Addition
1b + 4VP	-	**4b**: 58%	31%	11%
1b + 4CS	-	**7b**: 72%	28%	-
1b + 2BrS	-	**5b**: 79%	21%	-
1b + 2VhB	-	**6b**: 79%	21%	-
1b + diFP	-	**8b**: 83%	17%	-
1b + DPE	-	**3b**: 86%	14%	-
1b + 4M5VT	27%	**10b**: 67%	6%	-
1b + 4FMS	33%	**11b**: 67%	-	-
1b + 3EhP	39%	**9b**: 61%	-	-
1b + 5EMP	47%	**12b**: 53%	-	-
1b + 4MS	48%	**13b**: 52%	-	-
1b + 4VBA	67%	**14b**: 33%	-	-
1b + 4AS	100%	-	-	-



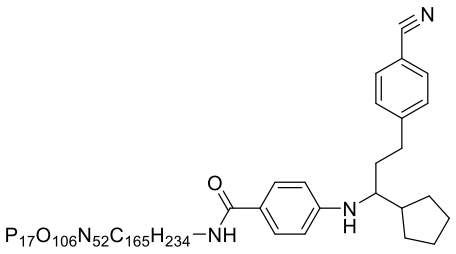
4b

Molecular Weight: 5490.9310



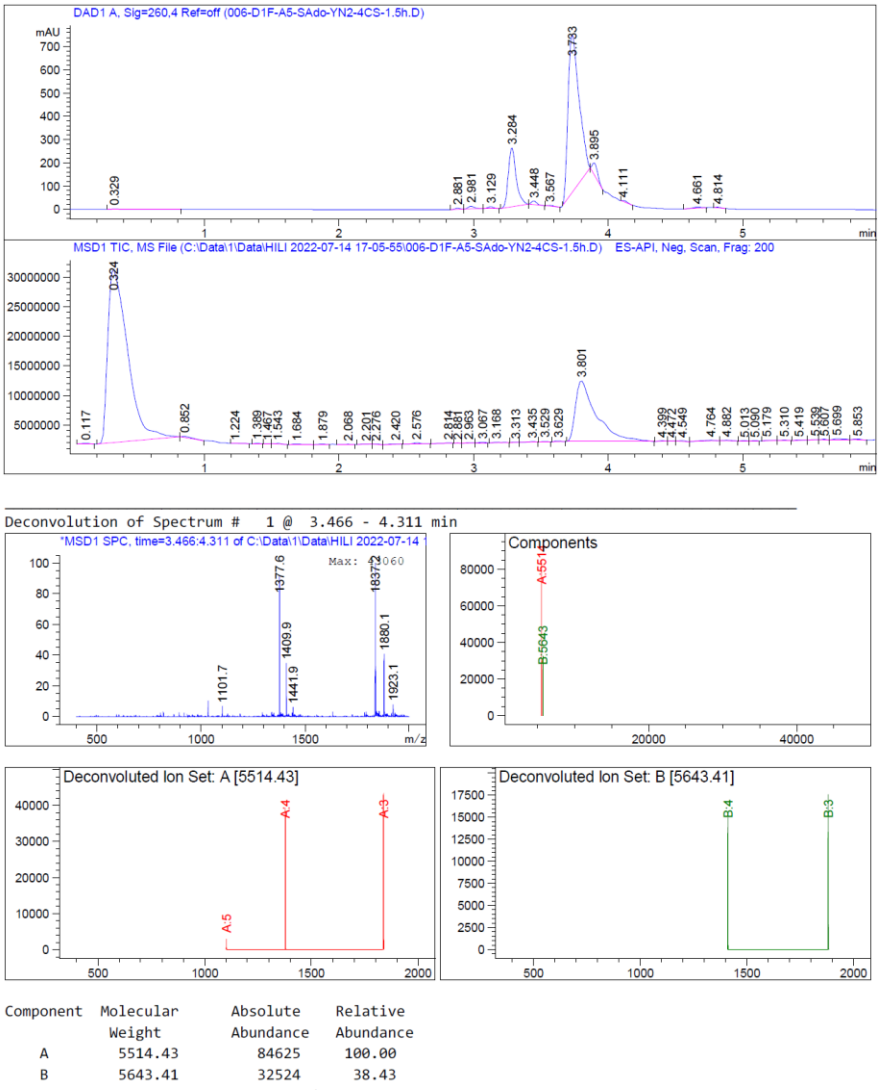
Component	Molecular Weight	Absolute Abundance	Relative Abundance
A	5490.29	101564	100.00
B	5595.29	55558	54.70
C	5700.72	19660	19.36

Figure 5-29: Deconvoluted LCMS data for **4b**.



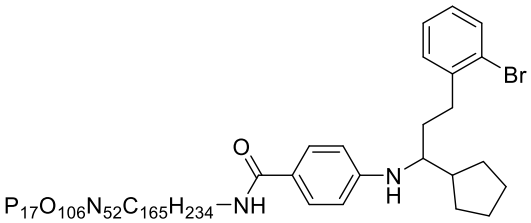
7b

Molecular Weight: 5514.9530



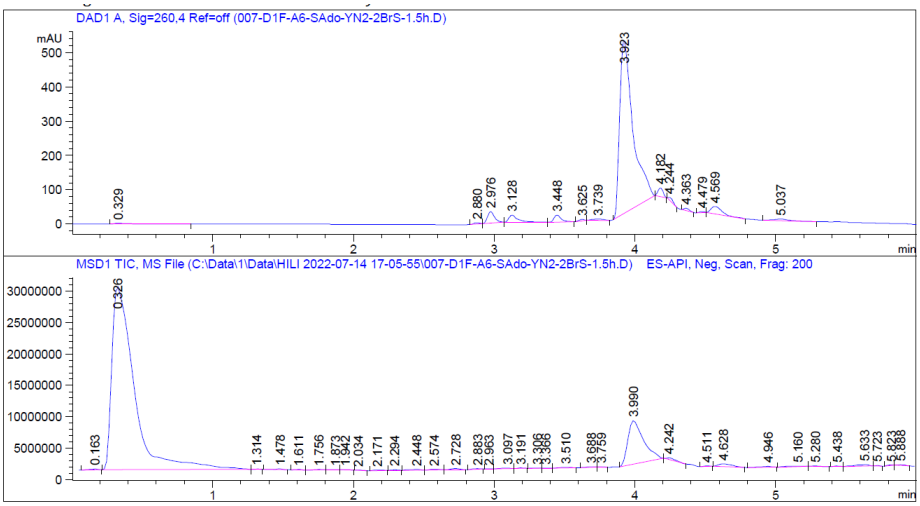
Component	Molecular Weight	Absolute Abundance	Relative Abundance
A	5514.43	84625	100.00
B	5643.41	32524	38.43

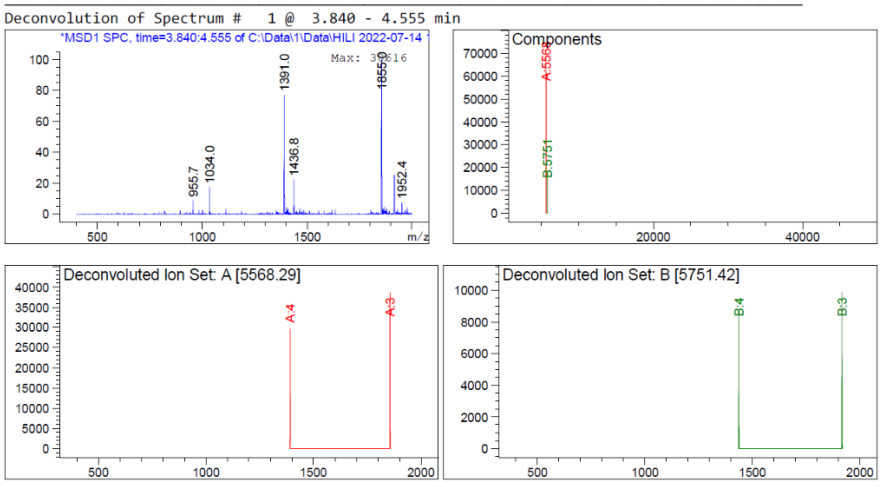
Figure 5-30: Deconvoluted LCMS data for **7b**.



5b

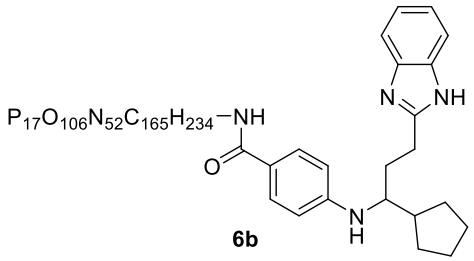
Molecular Weight: 5568.8390





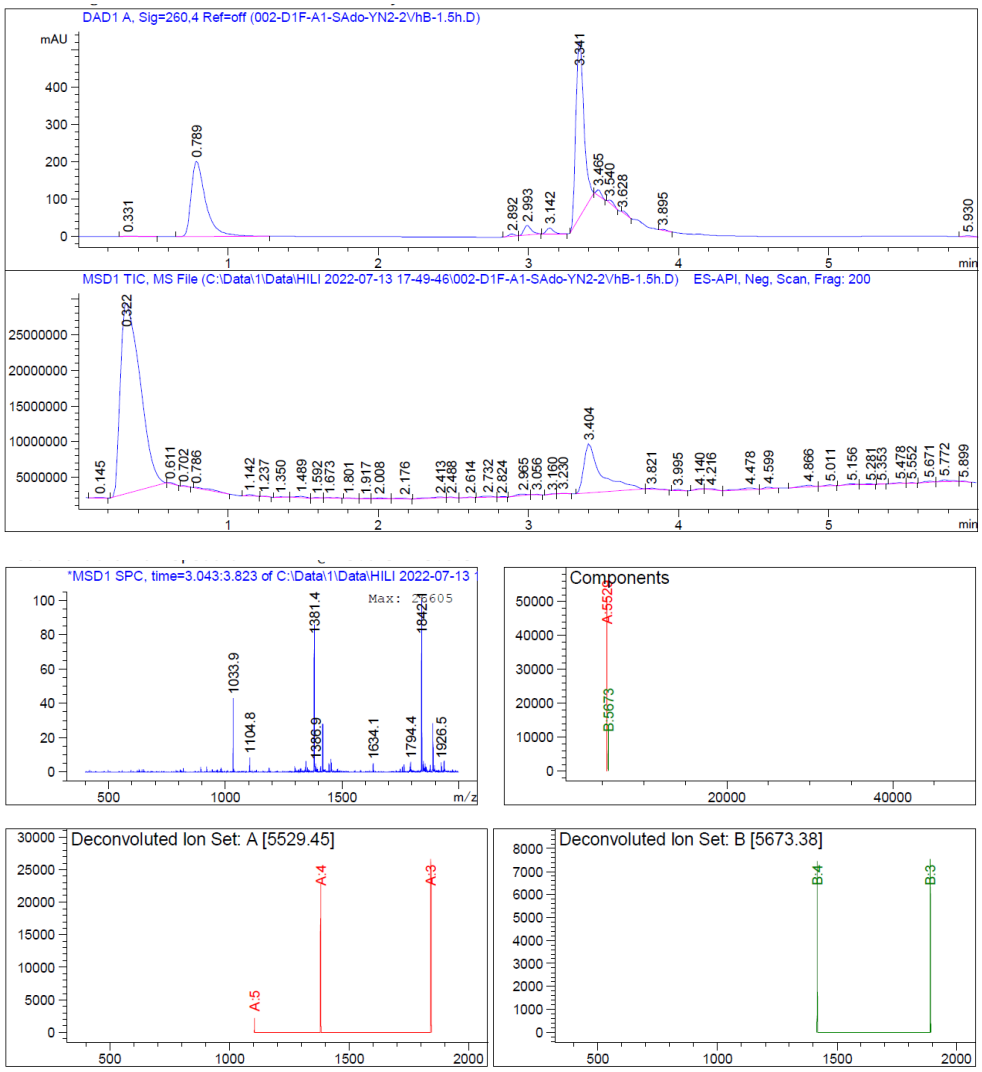
Component	Molecular Weight	Absolute Abundance	Relative Abundance
A	5568.29	68333	100.00
B	5751.42	18060	26.43

Figure 5-31: Deconvoluted LCMS data for **5b**.



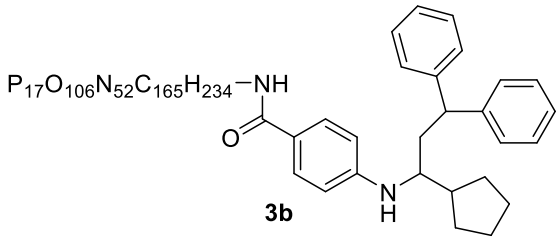
6b

Molecular Weight: 5529.9680

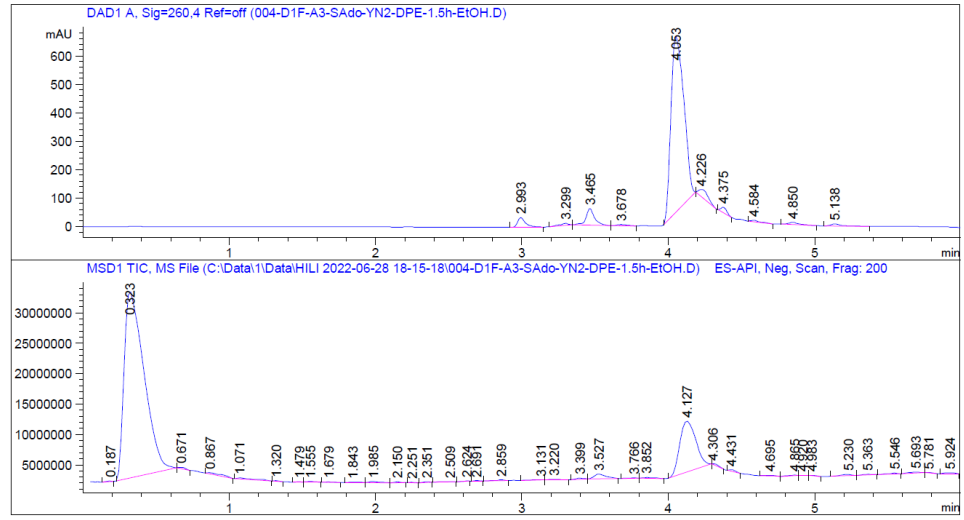


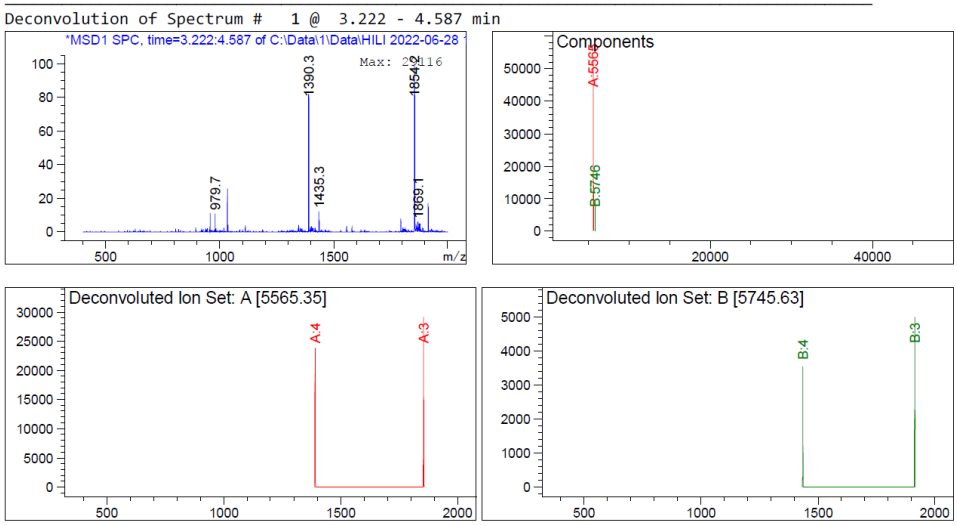
Component	Molecular Weight	Absolute Abundance	Relative Abundance
A	5529.45	51037	100.00
B	5673.38	13581	26.61

Figure 5-32: Deconvoluted LCMS data for **6b**.



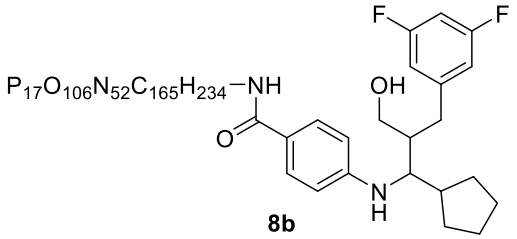
Molecular Weight: 5566.0410





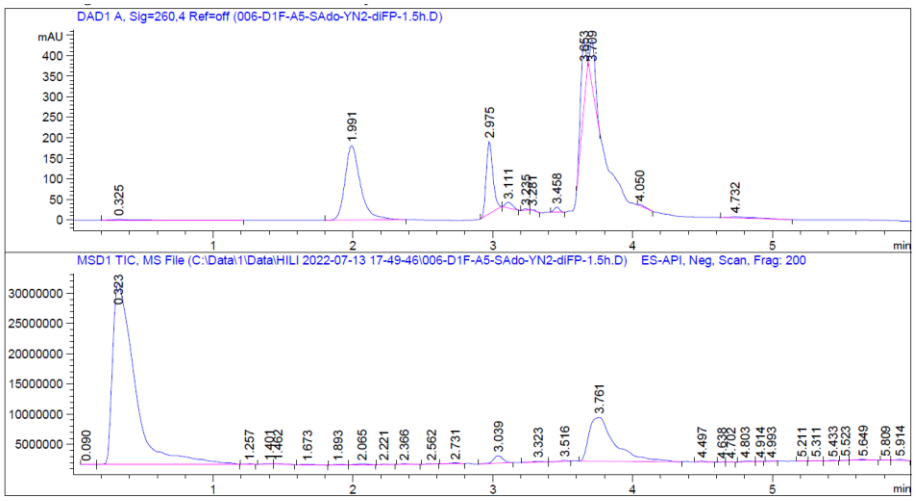
Component	Molecular Weight	Absolute Abundance	Relative Abundance
A	5565.35	52284	100.00
B	5745.63	8516	16.29

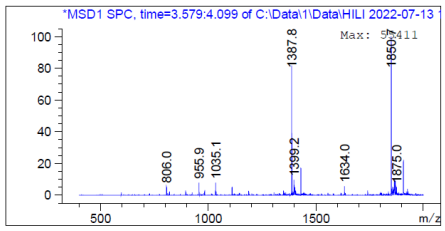
Figure 5-33: Deconvoluted LCMS data for **3b**.

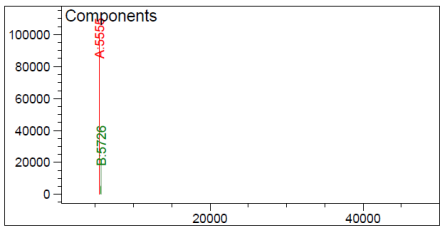


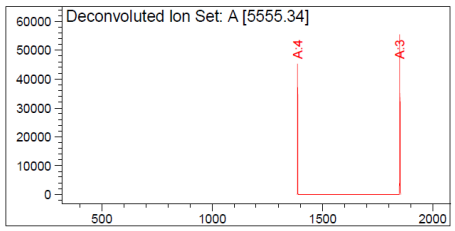
8b

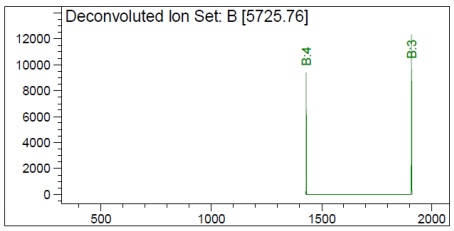
Molecular Weight: 5555.9498





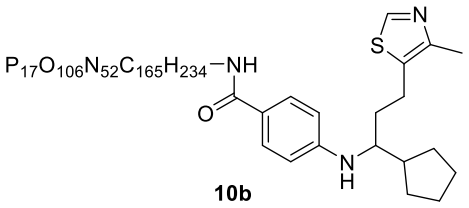






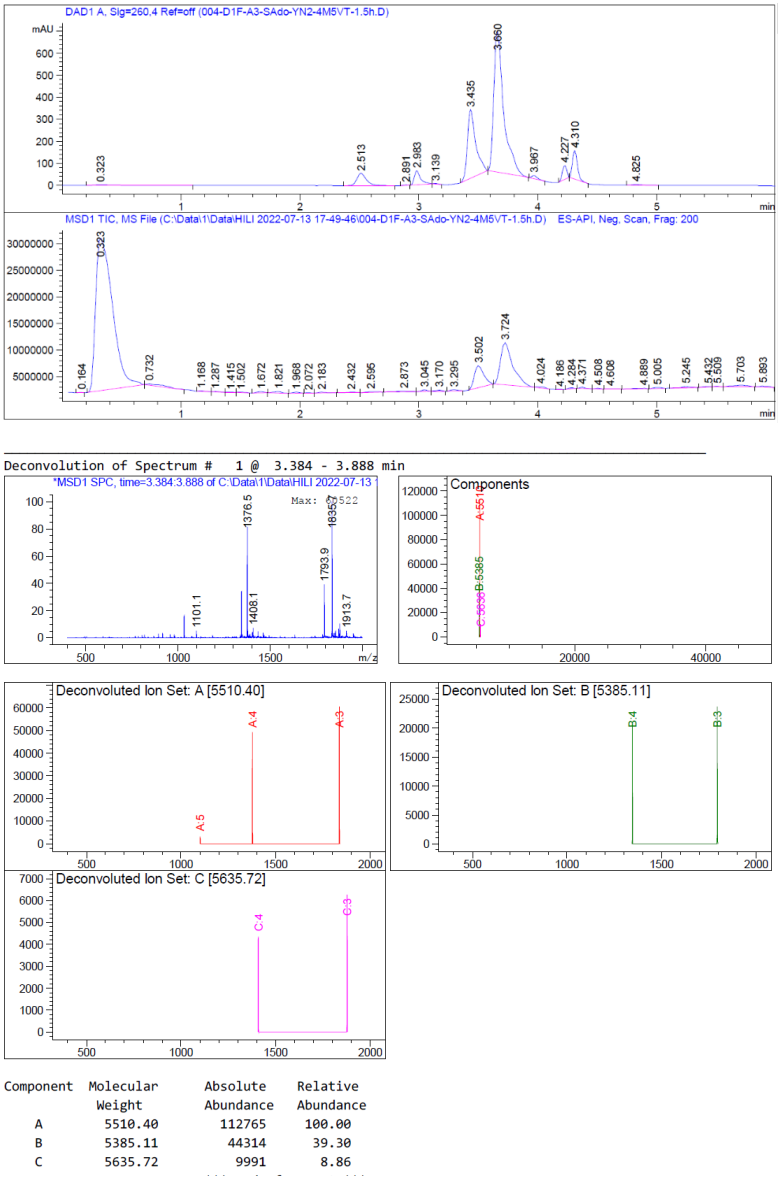
Component	Molecular Weight	Absolute Abundance	Relative Abundance
A	5555.34	100210	100.00
B	5725.76	20673	20.63

Figure 5-34: Deconvoluted LCMS data for **8b**.



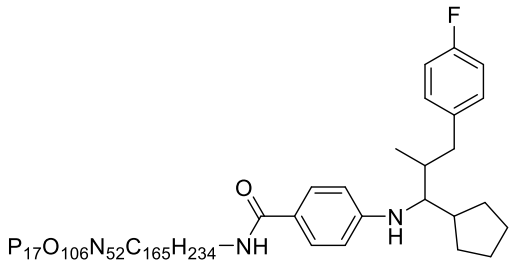
10b

Molecular Weight: 5510.9800



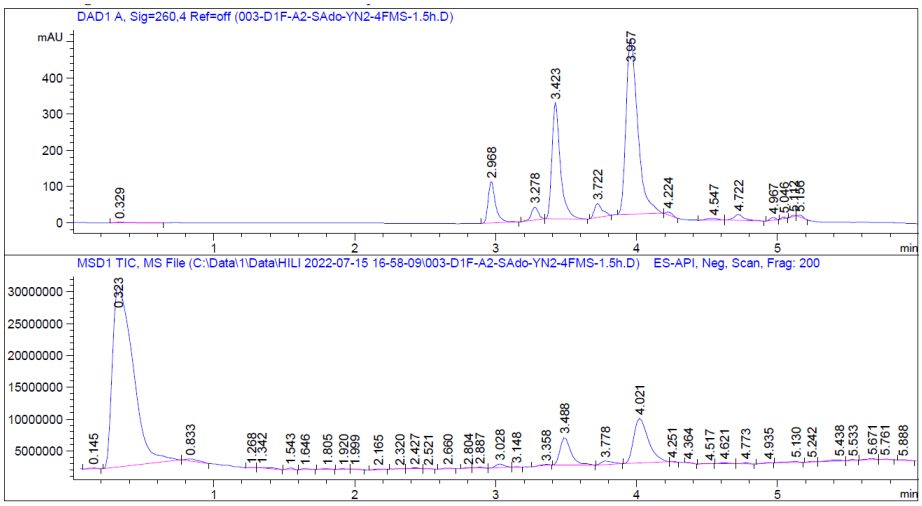
Component	Molecular Weight	Absolute Abundance	Relative Abundance
A	5510.40	112765	100.00
B	5385.11	44314	39.30
C	5635.72	9991	8.86

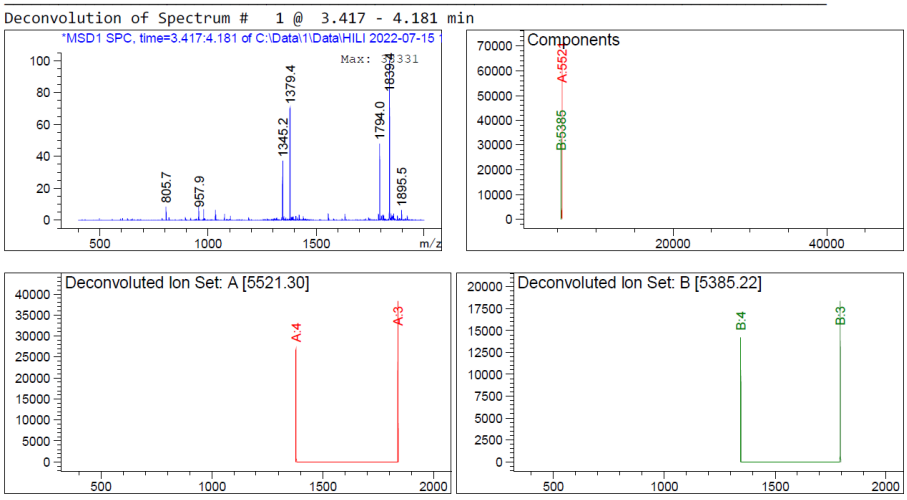
Figure 5-35: Deconvoluted LCMS data for **10b**.



11b

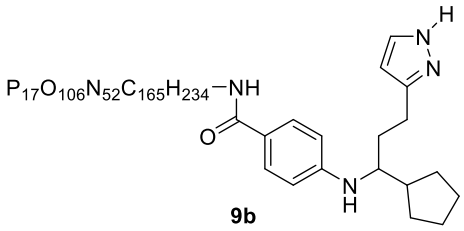
Molecular Weight: 5521.9604





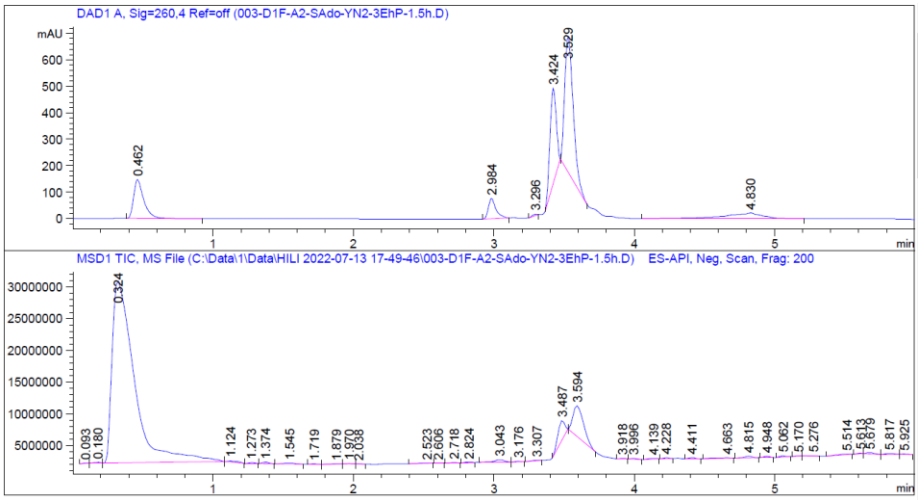
Component	Molecular Weight	Absolute Abundance	Relative Abundance
A	5521.30	65013	100.00
B	5385.22	32602	50.15

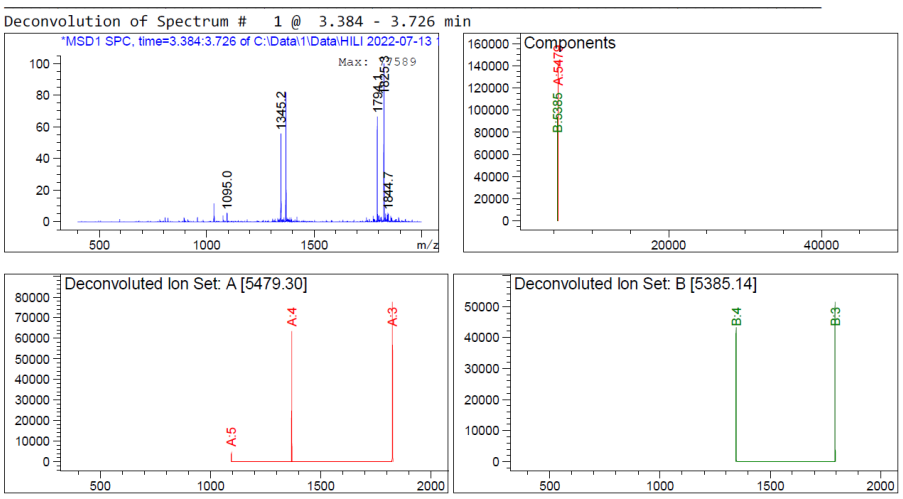
Figure 5-36: Deconvoluted LCMS data for **11b**.



9b

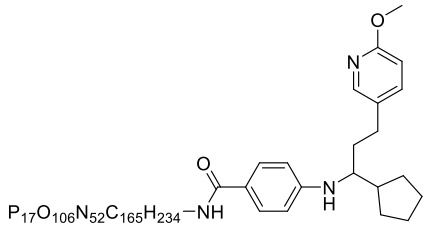
Molecular Weight: 5479.9080





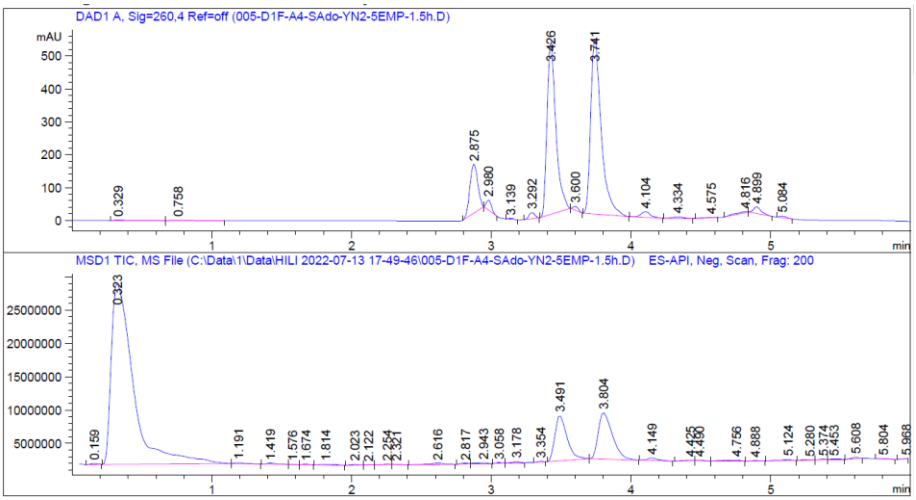
Component	Molecular Weight	Absolute Abundance	Relative Abundance
A	5479.30	143998	100.00
B	5385.14	93740	65.10

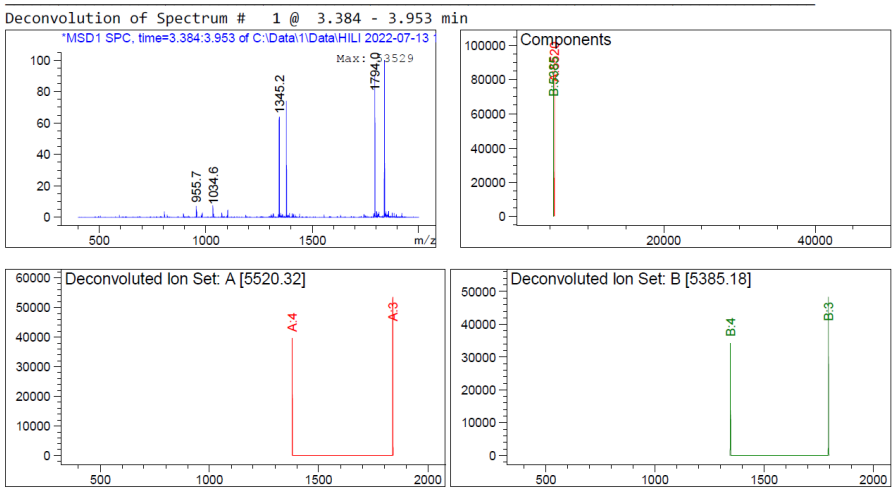
Figure 5-37: Deconvoluted LCMS data for **9b**.



12b

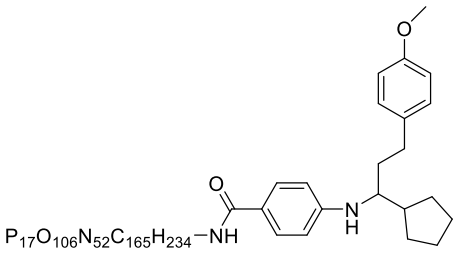
Molecular Weight: 5520.9570





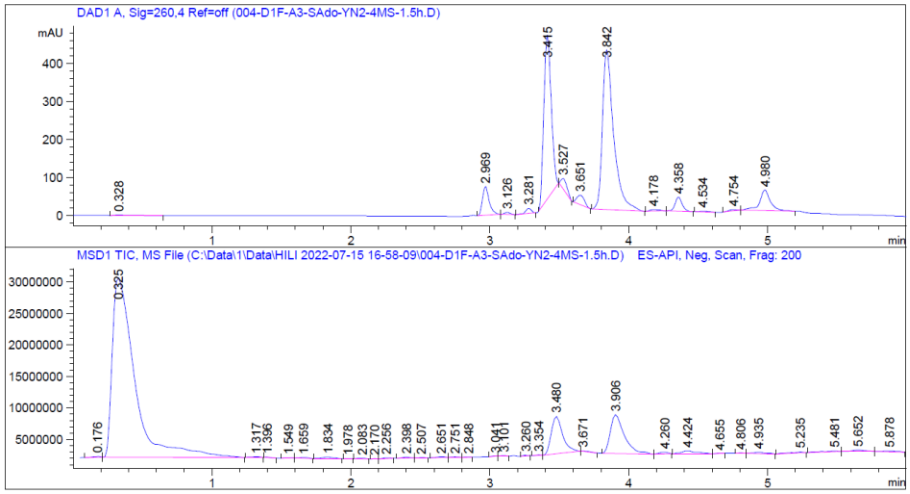
Component	Molecular Weight	Absolute Abundance	Relative Abundance
A	5520.32	92687	100.00
B	5385.18	82469	88.98

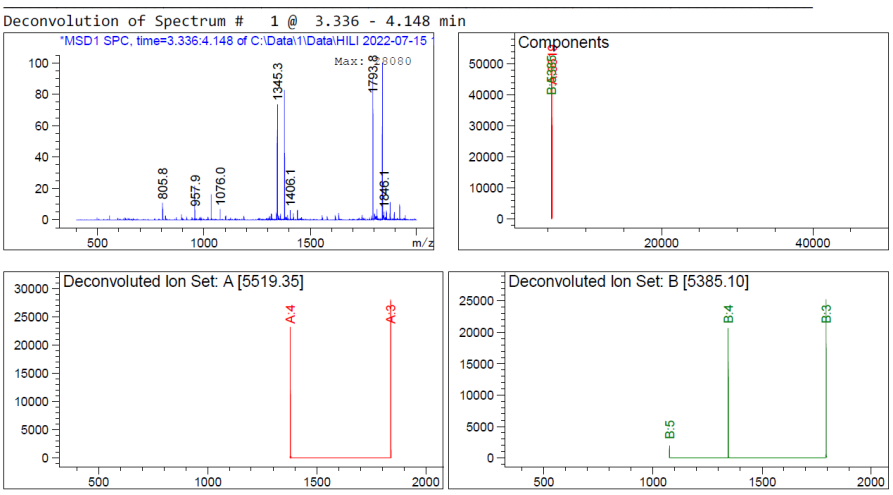
Figure 5-38: Deconvoluted LCMS data for **12b**.



13b

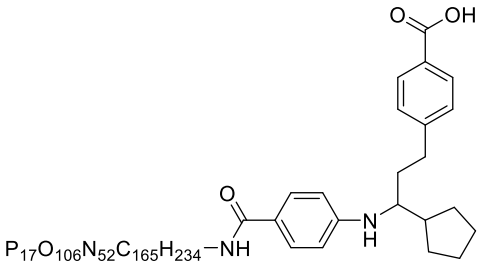
Molecular Weight: 5519.9690





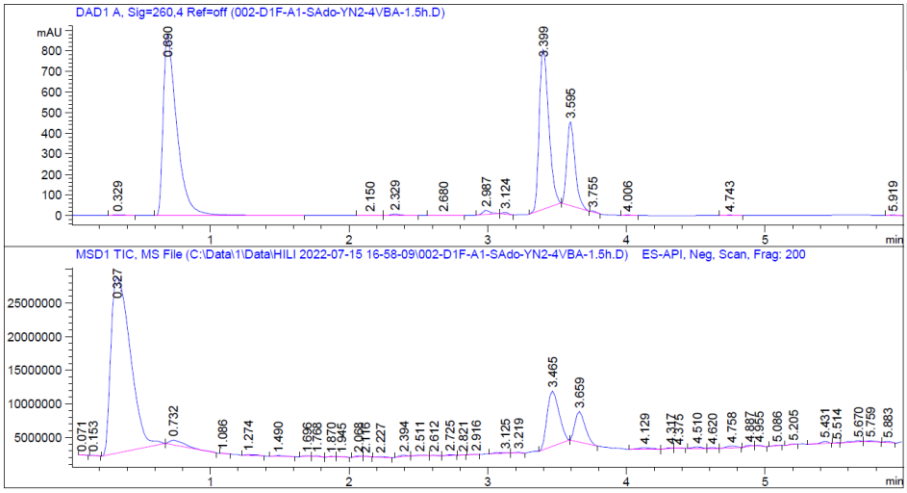
Component	Molecular Weight	Absolute Abundance	Relative Abundance
A	5519.35	51096	100.00
B	5385.10	47038	92.06

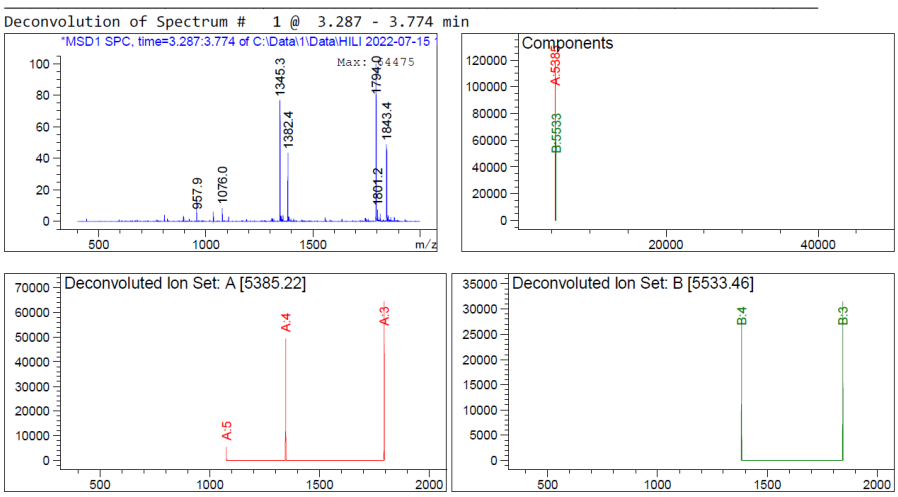
Figure 5-39: Deconvoluted LCMS data for **13b**.



14b

Molecular Weight: 5533.9520





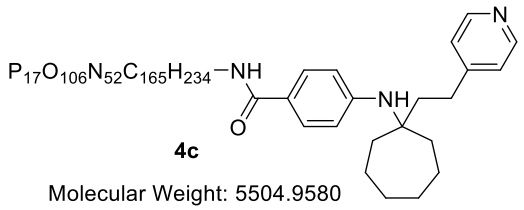
Component	Molecular Weight	Absolute Abundance	Relative Abundance
A	5385.22	118981	100.00
B	5533.46	59173	49.73

Figure 5-40: Deconvoluted LCMS data for **14b**.

5.5.14. LCMS spectra and deconvolution results for 1c derivatives

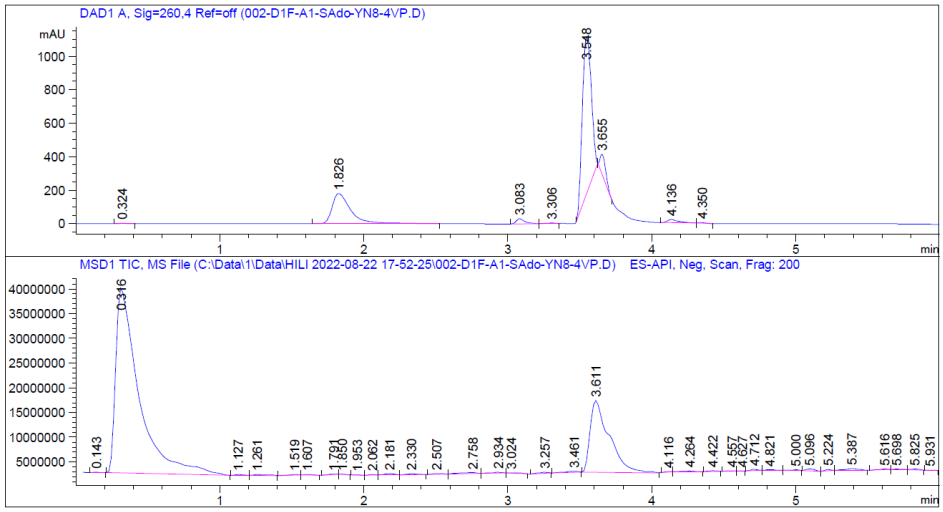
Table 5-8: Hydroaminoalkylation of various vinylarenes with DNA conjugate **1c**

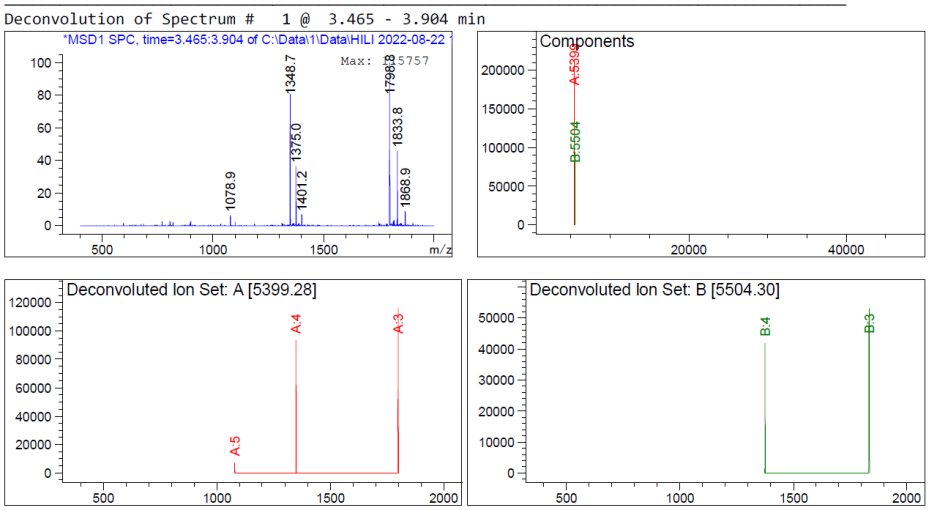
	Starting Material (**1c**)	Single Addition	Double Addition	Triple Addition
1c + 4VP	69%	**4c**: 31%	-	-
1c + 4CS	83%	**7c**: 17%	-	-
1c + 2VhB	86%	**6c**: 14%	-	-
1c + 2BrS	88%	**5c**: 12%	-	-
1c + diFP	89%	**8c**: 11%	-	-
1c + 4M5VT	93%	**10c**: 7%	-	-
1c + DPE	100%	**3c**: 0%	-	-
1c + 4FMS	100%	**11c**: 0%	-	-
1c + 3EhP	100%	**9c**: 0%	-	-
1c + 5EMP	100%	**12c**: 0%	-	-
1c + 4MS	100%	**13c**: 0%	-	-
1c + 4VBA	100%	**14c**: 0%	-	-
1c + 4AS	100%	-	-	-



4c

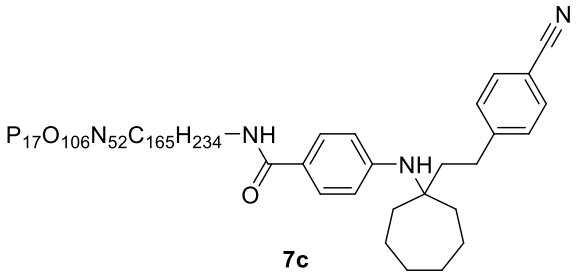
Molecular Weight: 5504.9580





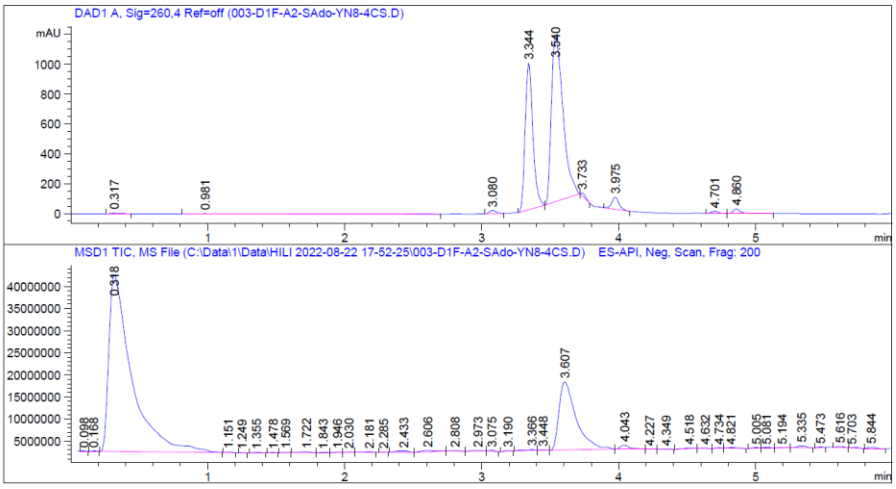
Component	Molecular Weight	Absolute Abundance	Relative Abundance
A	5399.28	212905	100.00
B	5504.30	94072	44.18

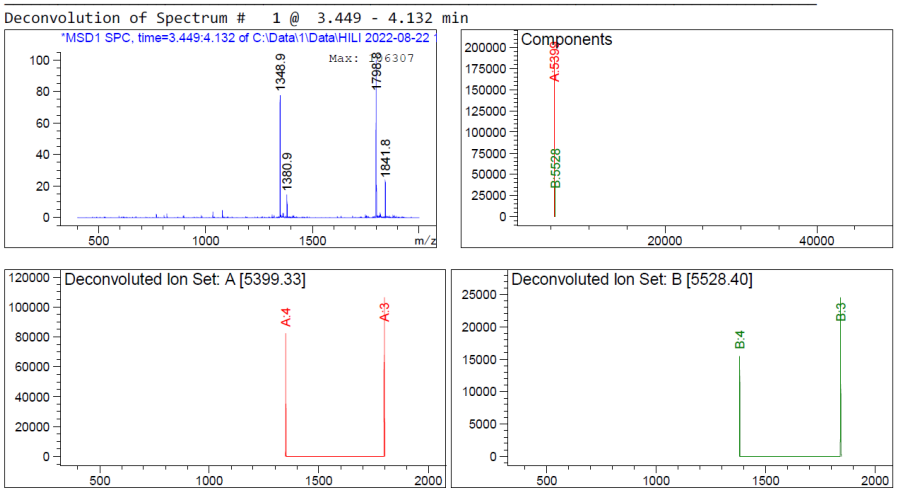
Figure 5-41: Deconvoluted LCMS data for **4c**.



7c

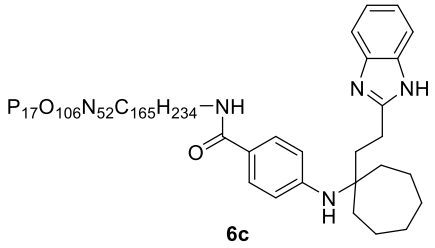
Molecular Weight: 5528.9800





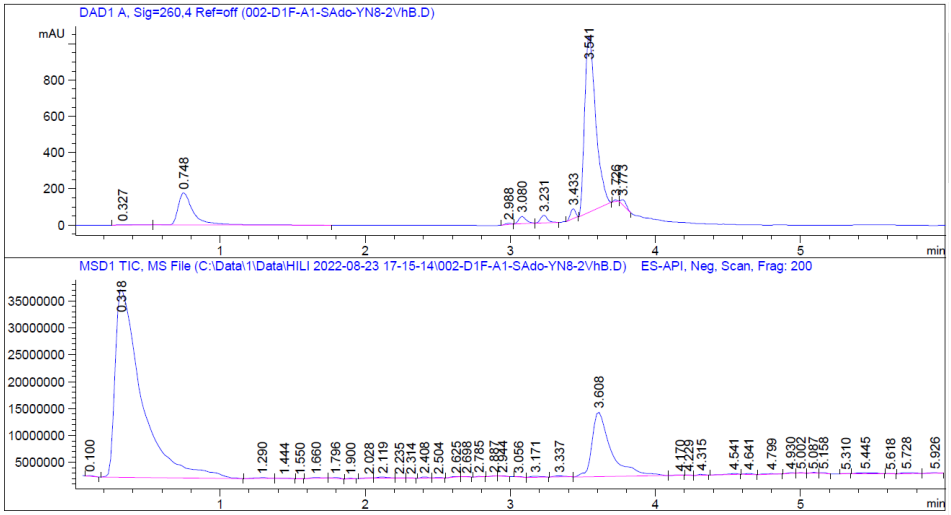
Component	Molecular Weight	Absolute Abundance	Relative Abundance
A	5399.33	187971	100.00
B	5528.40	39590	21.06

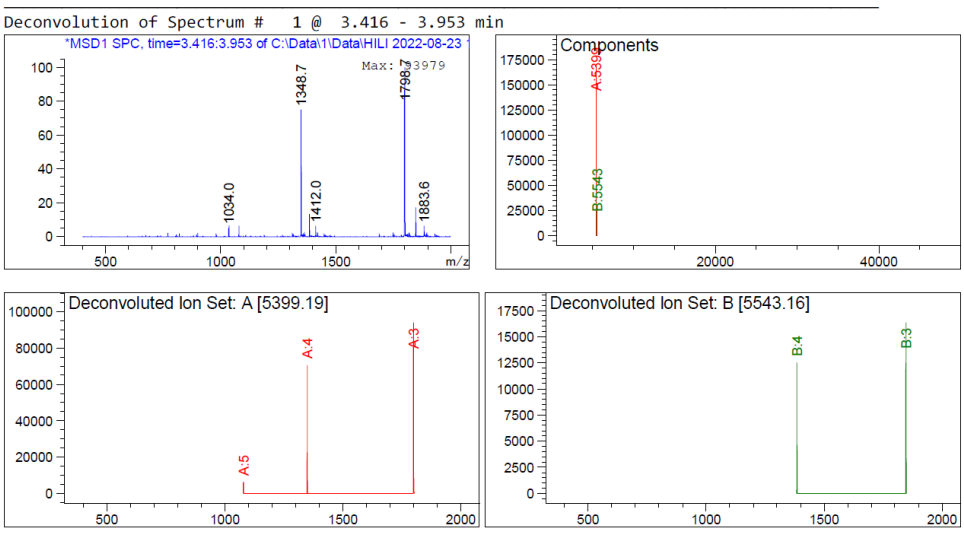
Figure 5-42: Deconvoluted LCMS data for **7c**.



6c

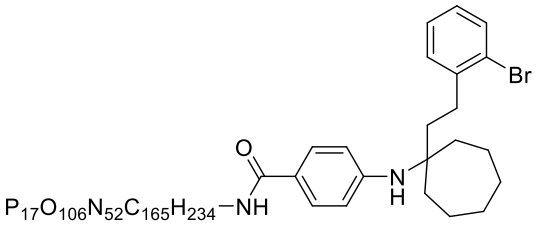
Molecular Weight: 5543.9950





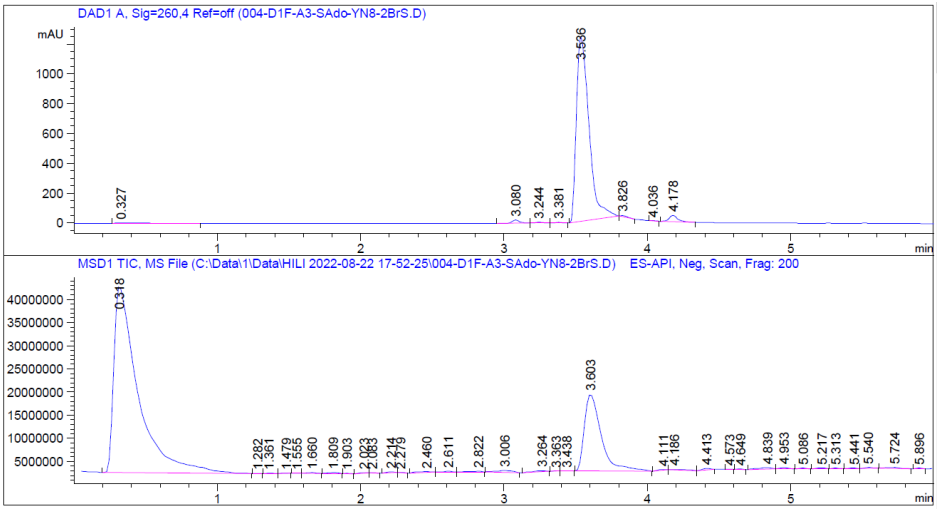
Component	Molecular Weight	Absolute Abundance	Relative Abundance
A	5399.19	169927	100.00
B	5543.16	28136	16.56

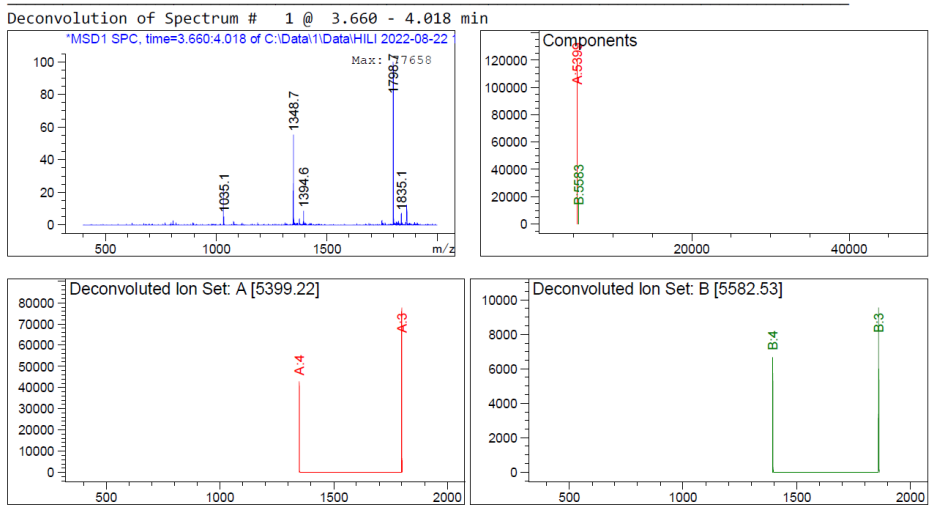
Figure 5-43: Deconvoluted LCMS data for **6c**.



5c

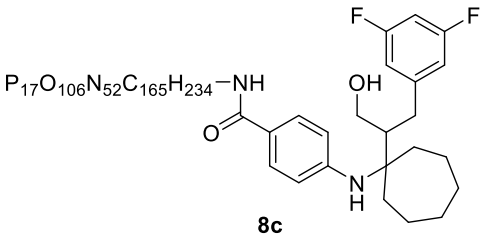
Molecular Weight: 5582.8660





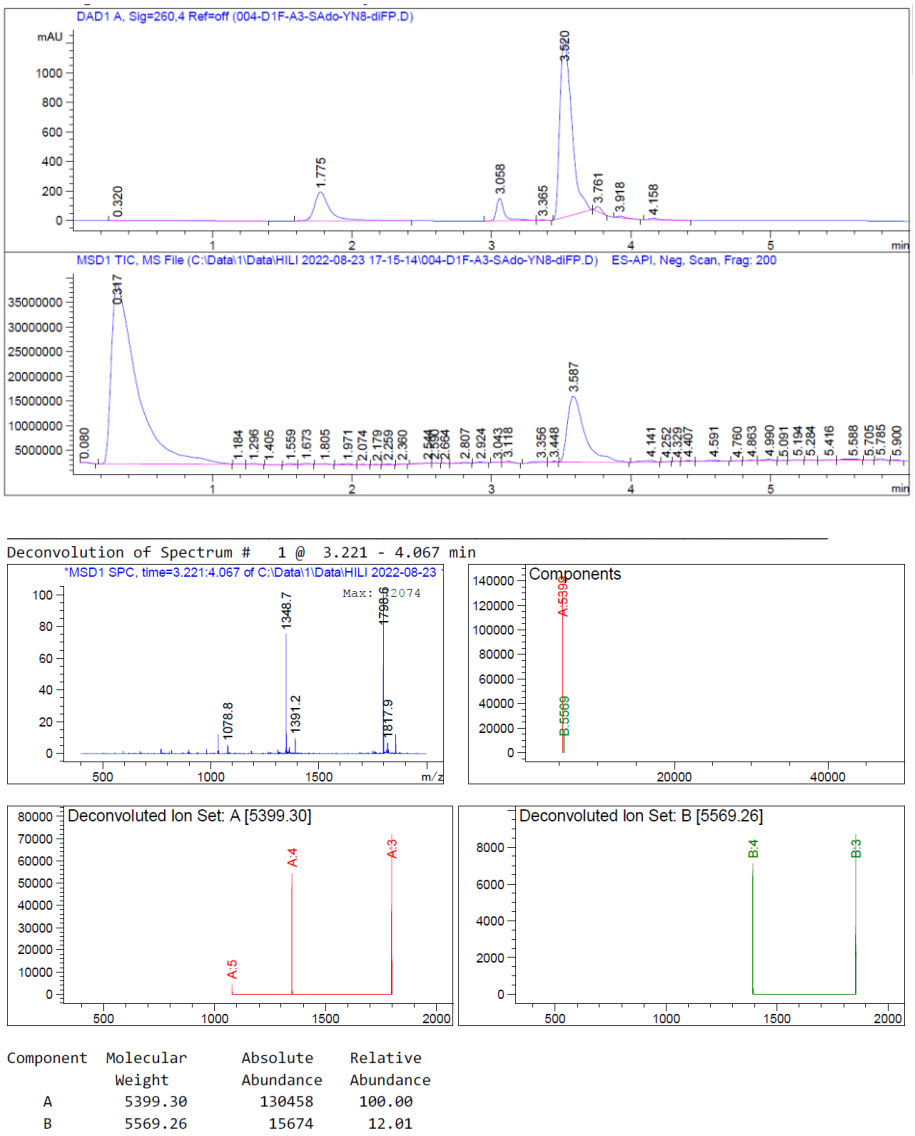
Component	Molecular Weight	Absolute Abundance	Relative Abundance
A	5399.22	120466	100.00
B	5582.53	15979	13.26

Figure 5-44: Deconvoluted LCMS data for **5c**.



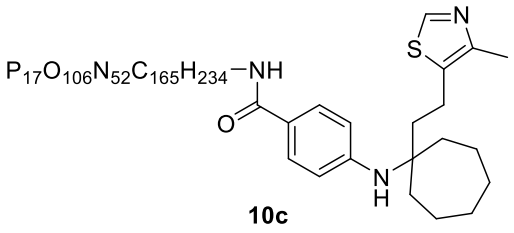
8c

Molecular Weight: 5569.9768



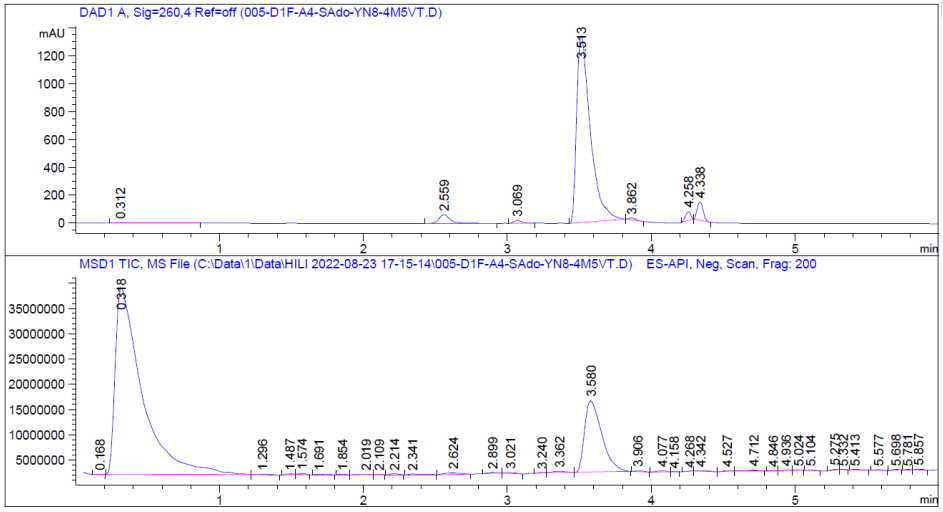
Component	Molecular Weight	Absolute Abundance	Relative Abundance
A	5399.30	130458	100.00
B	5569.26	15674	12.01

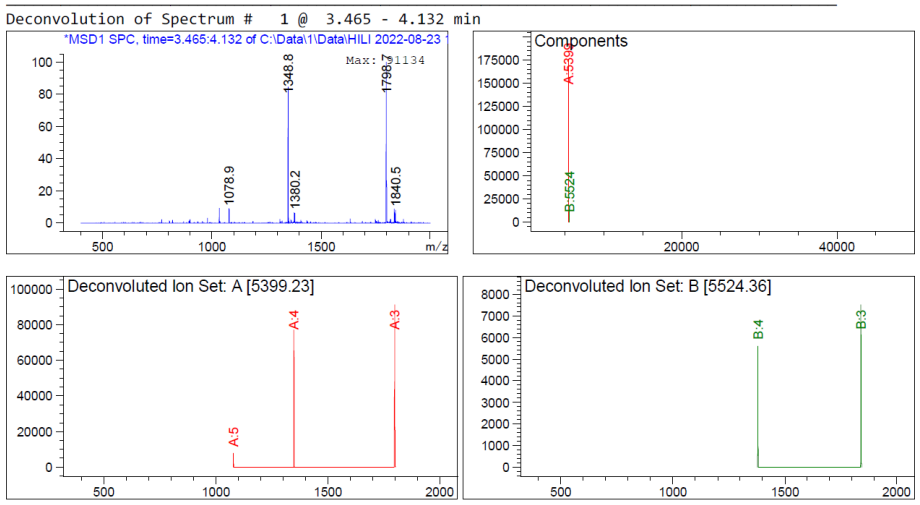
Figure 5-45: Deconvoluted LCMS data for **8c**.



10c

Molecular Weight: 5525.0070





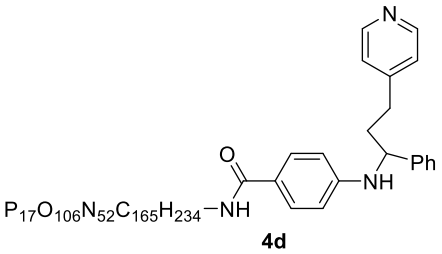
Component	Molecular Weight	Absolute Abundance	Relative Abundance
A	5399.23	175435	100.00
B	5524.36	13121	7.48

Figure 5-46: Deconvoluted LCMS data for **10c**.

5.5.15. LCMS spectra and deconvolution results for 1d derivatives

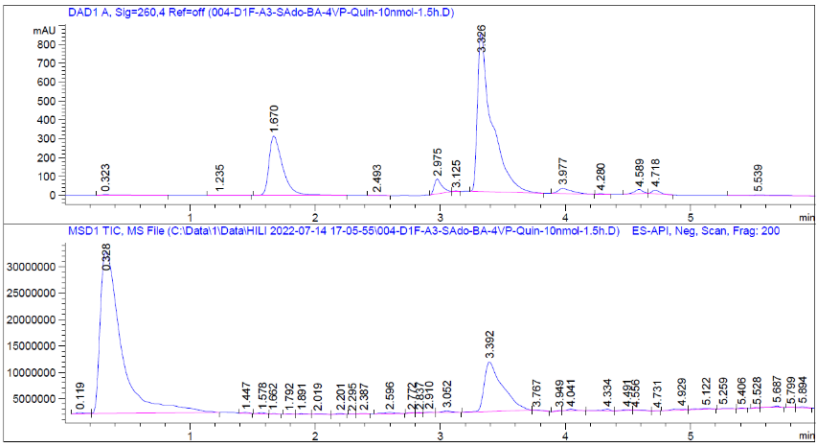
Table 5-9: Hydroaminoalkylation of various vinylarenes with DNA conjugate **1d**

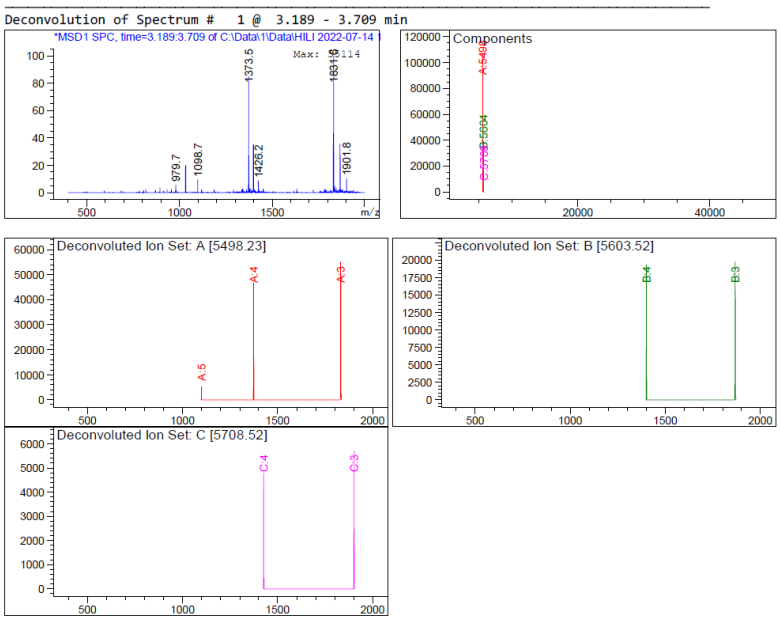
	Starting Material (**1d**)	Single Addition	Double Addition	Triple Addition	Other
1d + 4VP	-	**4d**: 68%	25%	7%	-
1d + 4CS	13%	**7d**: 75%	12%	-	-
1d + 2VhB	44%	**6d**: 56%	-	-	-
1d + 2BrS	60%	**5d**: 40%	-	-	-
1d + 4M5VT	67%	**10d**: 33%	-	-	-
1d + DPE	77%	**3d**: 23%	-	-	-
1d + diFP	100%	**8d**: 0%	-	-	-
1d + 4FMS	100%	**11d**: 0%	-	-	-
1d + 3EhP	82%	**9d**: 0%	-	-	**9d**-quinuclidine adduct: 18%
1d + 5EMP	100%	**12d**: 0%	-	-	-
1d + 4MS	91%	**13d**: 0%	-	-	**13d**-quinuclidine adduct: 9%
1d + 4VBA	90%	**14d**: 0%	-	-	**14d-**quinuclidine adduct: 10%
1d + 4AS	100%	-	-	-	-



4d

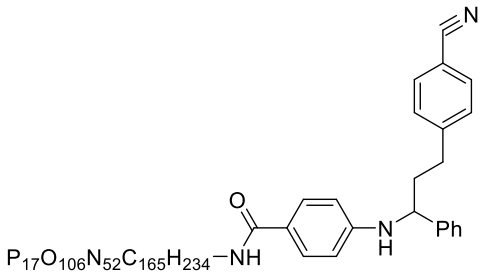
Molecular Weight: 5498.9100





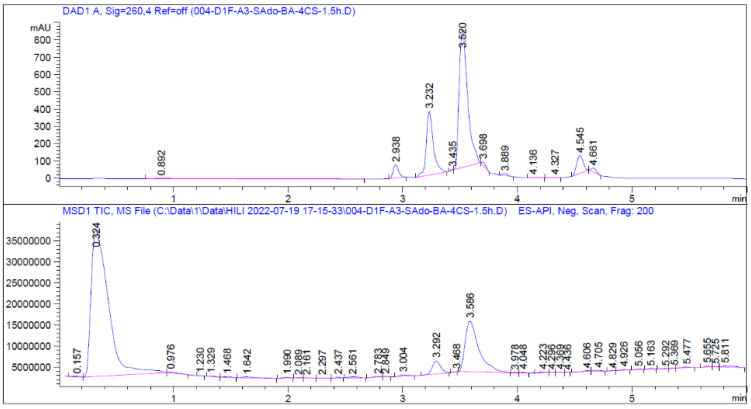
Component	Molecular Weight	Absolute Abundance	Relative Abundance
A	5498.23	106623	100.00
B	5603.52	38941	36.52
C	5708.52	10481	9.83

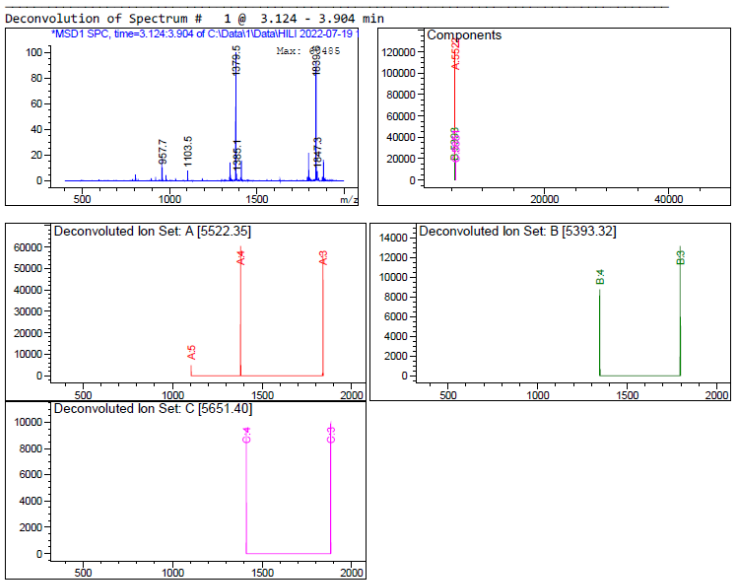
Figure 5-47: Deconvoluted LCMS data for **4d**.



7d

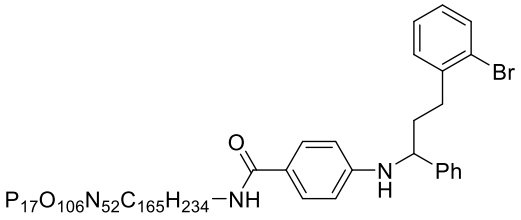
Molecular Weight: 5522.9320





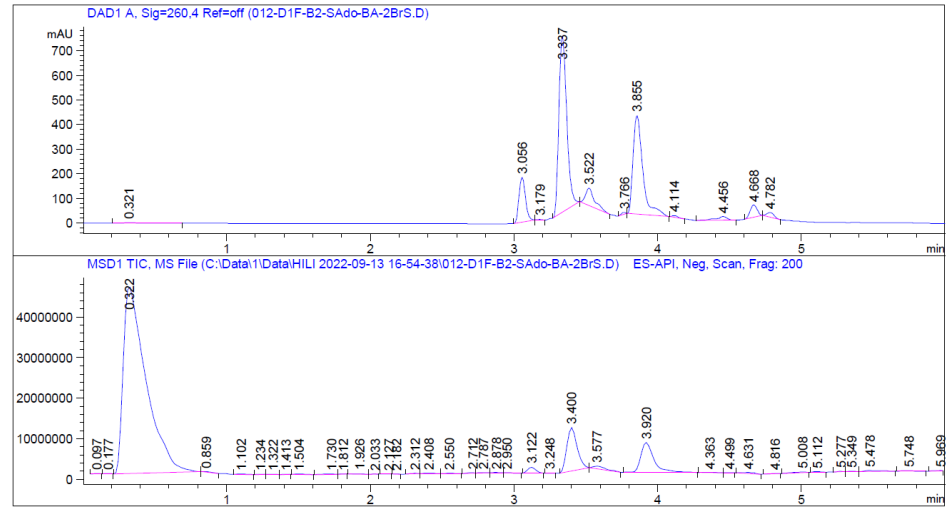
Component	Molecular Weight	Absolute Abundance	Relative Abundance
A	5522.35	121284	100.00
B	5393.32	21814	17.99
C	5651.40	19242	15.87

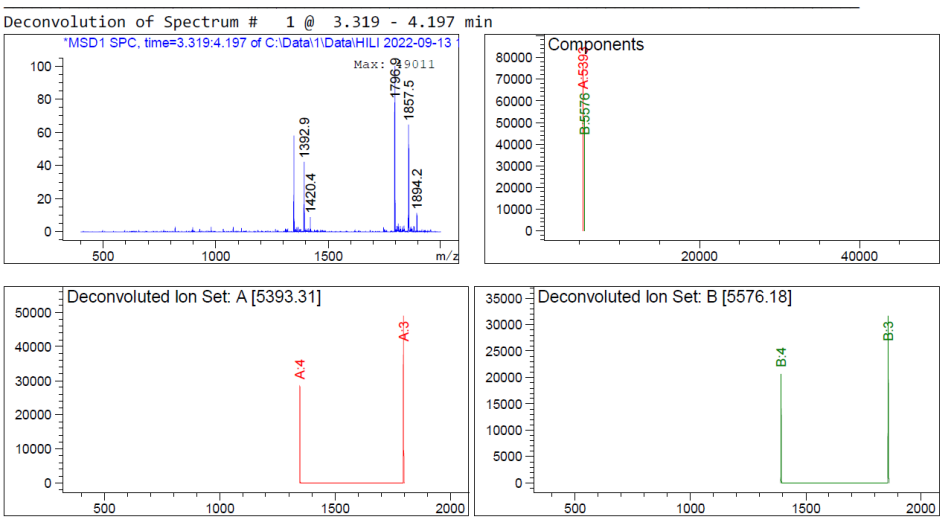
Figure 5-48: Deconvoluted LCMS data for **7d**.



5d

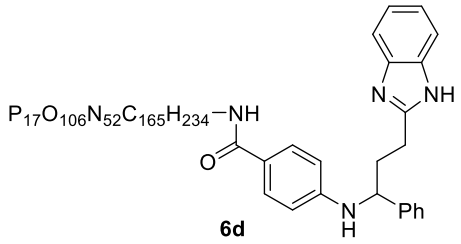
Molecular Weight: 5576.8180





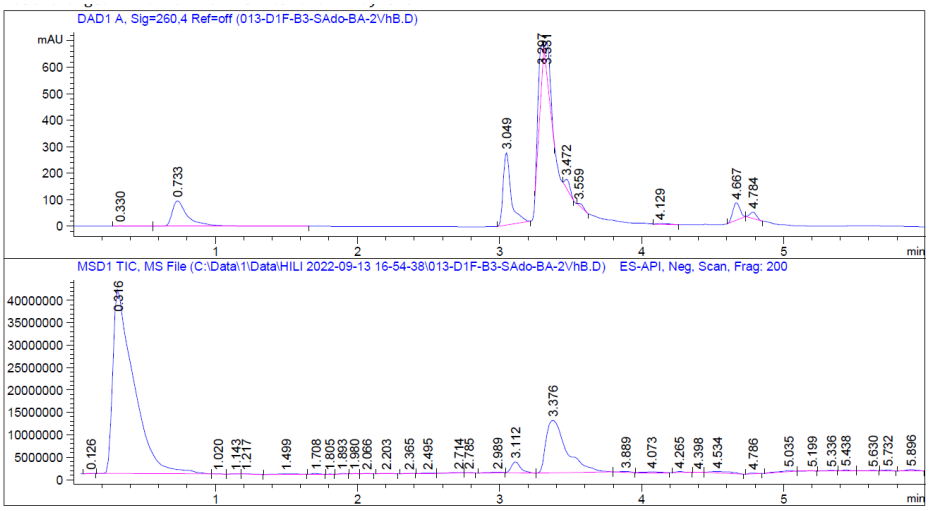
Component	Molecular Weight	Absolute Abundance	Relative Abundance
A	5393.31	77033	100.00
B	5576.18	52045	67.56

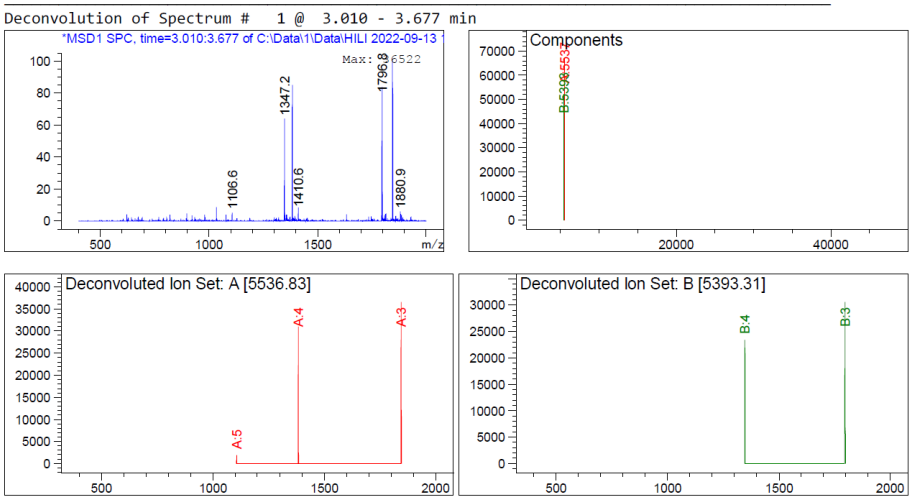
Figure 5-49: Deconvoluted LCMS data for **5d**.



6d

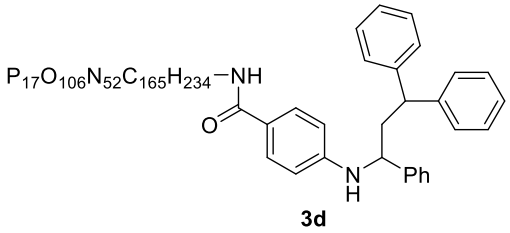
Molecular Weight: 5537.9470





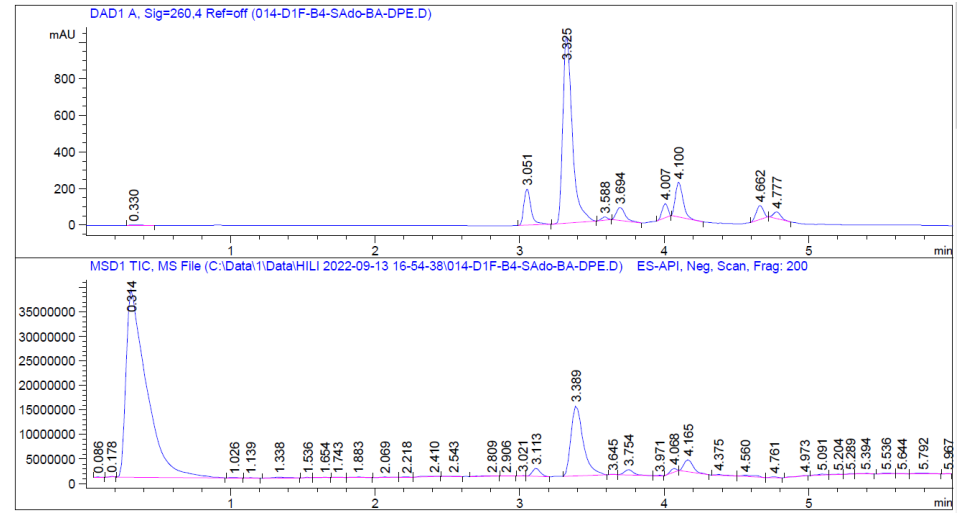
Component	Molecular Weight	Absolute Abundance	Relative Abundance
A	5536.83	66853	100.00
B	5393.31	52213	78.10

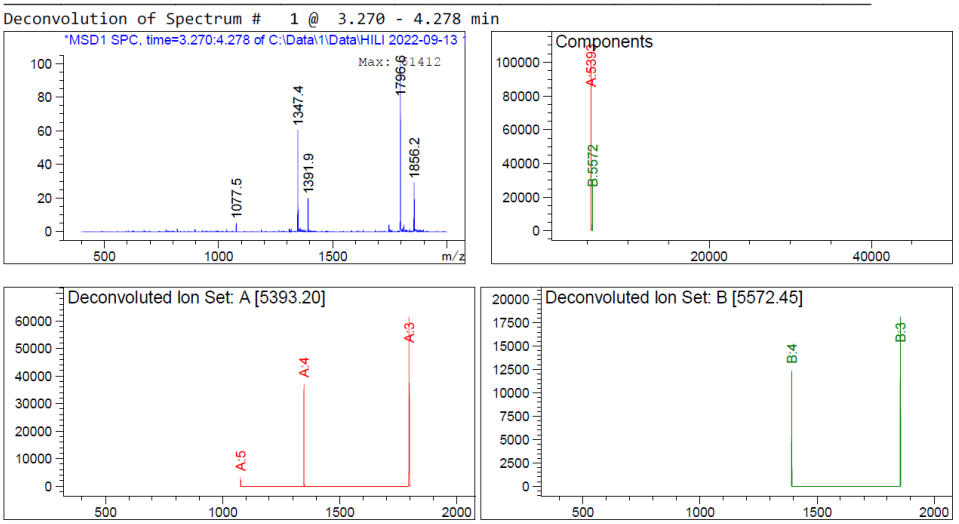
Figure 5-50: Deconvoluted LCMS data for **6d**.



3d

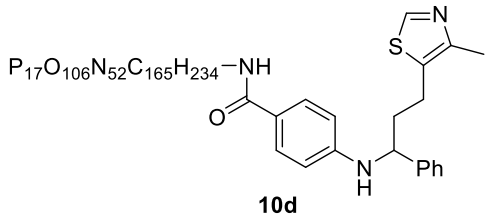
Molecular Weight: 5574.0200





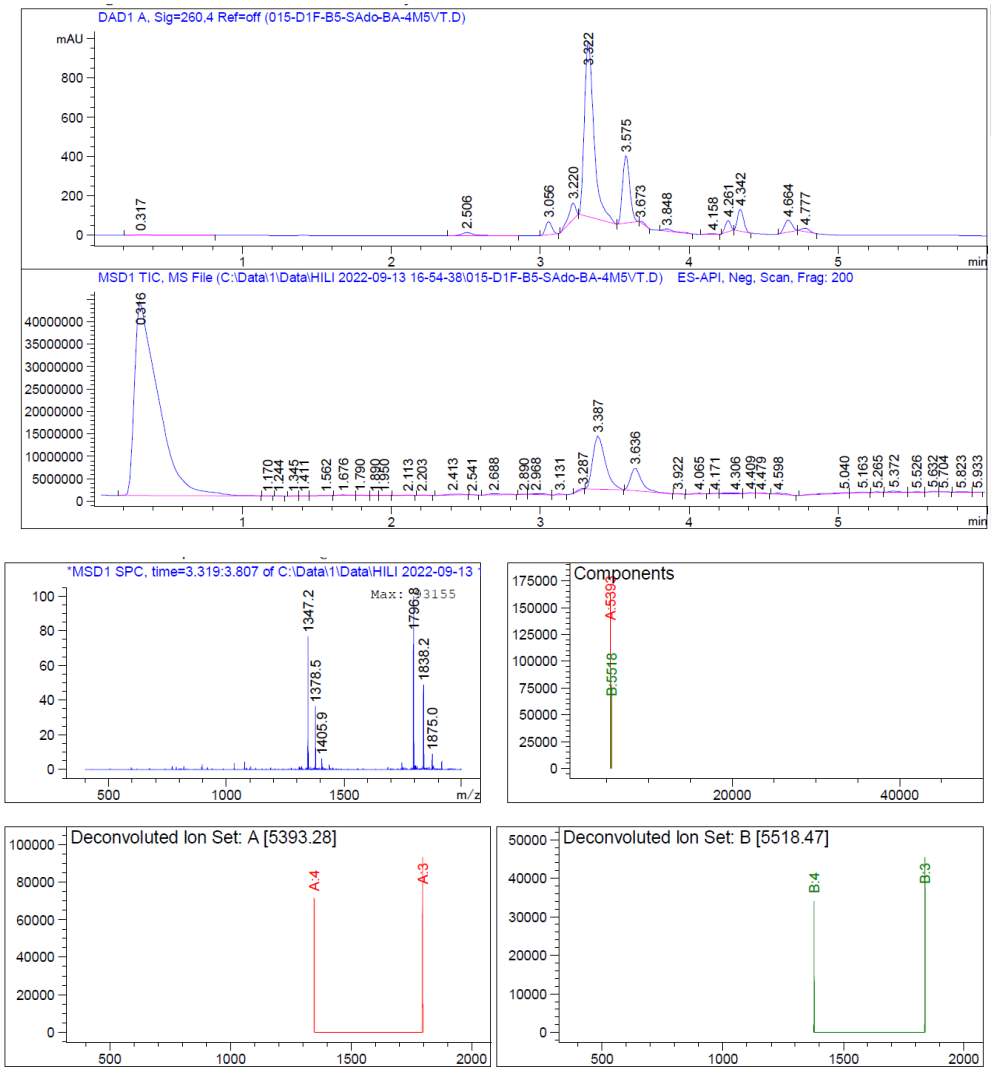
Component	Molecular Weight	Absolute Abundance	Relative Abundance
A	5393.20	100616	100.00
B	5572.45	30459	30.27

Figure 5-51: Deconvoluted LCMS data for **3d**.



10d

Molecular Weight: 5518.9590



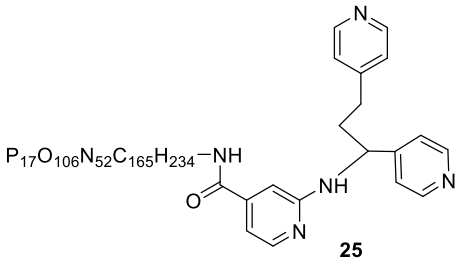
Component	Molecular Weight	Absolute Abundance	Relative Abundance
A	5393.28	162858	100.00
B	5518.47	78733	48.34

Figure 5-52: Deconvoluted LCMS data for **10d**.

5.5.16. LCMS spectra and deconvolution results for 1e derivatives

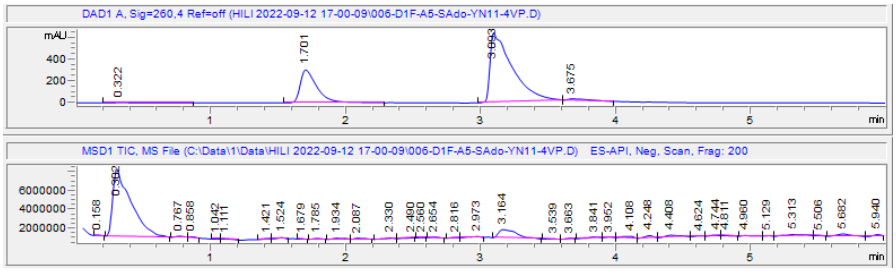
Table 5-10: Hydroaminoalkylation of various vinylarenes with DNA conjugate **1e**

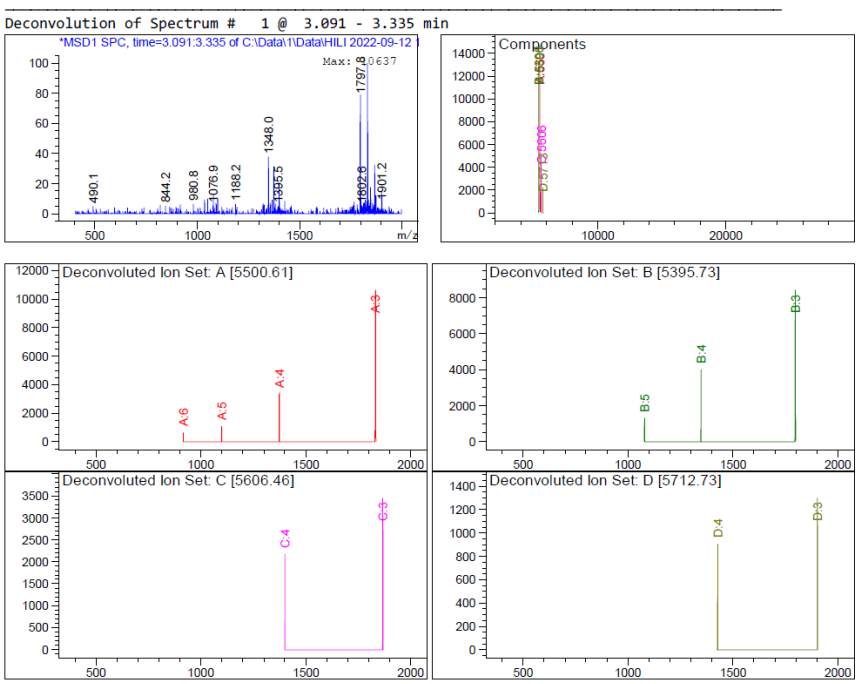
	Starting Material (**1e**)	Single Addition	Double Addition	Triple Addition
1e + 4VP	39%	**25**: 40%	15%	6%
1e + 4CS	63%	**26**: 31%	6%	-
1e + 2BrS	72%	**27**: 28%	-	-
1e + 2VhB	78%	**28**: 22%	-	-
1e + DPE	86%	**29**: 14%	-	-
1e + diFP	72%	**30**: 28%	-	-
1e + 4M5VT	90%	**31**: 10%	-	-
1e + 3EhP	100%	-	-	-
1e + 4FMS	100%	-	-	-
1e + 5EMP	100%	-	-	-
1e + 4MS	100%	-	-	-
1e + 4VBA	100%	-	-	-
1e + 4AS	100%	-	-	-



25

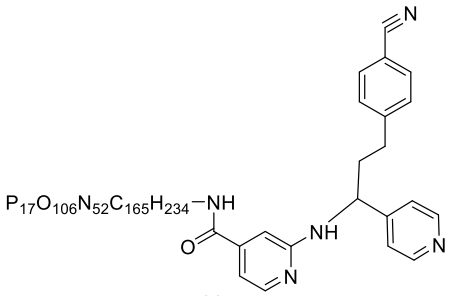
Molecular Weight: 5500.8860





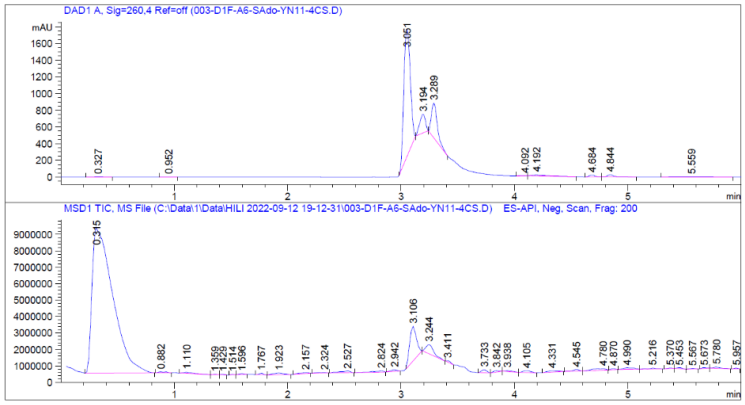
Component	Molecular Weight	Absolute Abundance	Relative Abundance
A	5500.61	13318	100.00
B	5395.73	13279	99.71
C	5606.46	5036	37.81
D	5712.73	2139	16.06

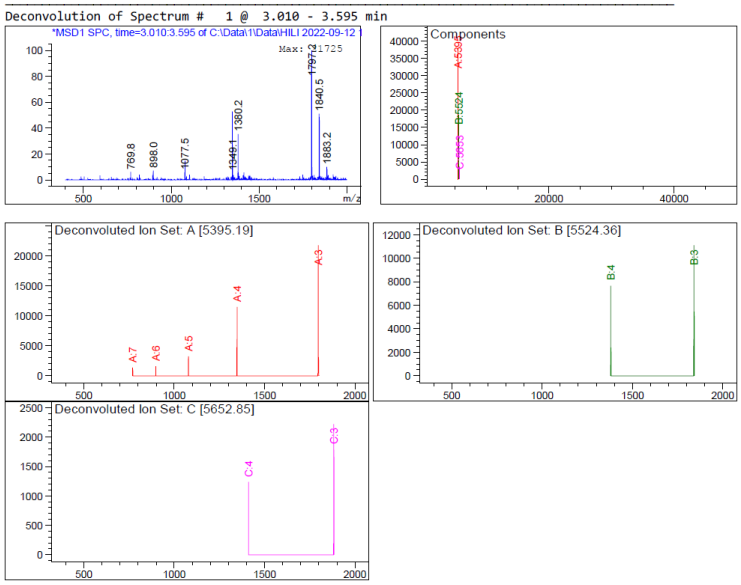
Figure 5-53: Deconvoluted LCMS data for **25**.



26

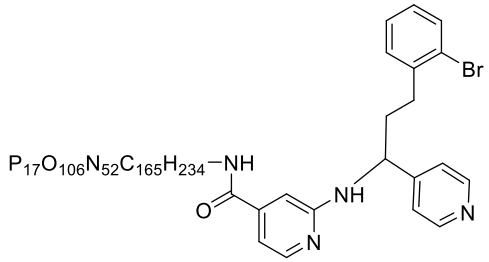
Molecular Weight: 5524.9080





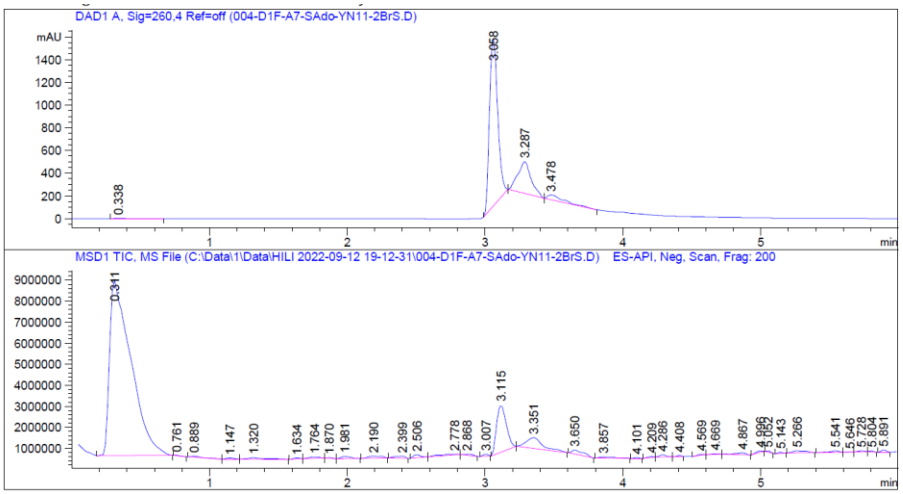
Component	Molecular Weight	Absolute Abundance	Relative Abundance
A	5395.19	37767	100.00
B	5524.36	18597	49.24
C	5652.85	3453	9.14

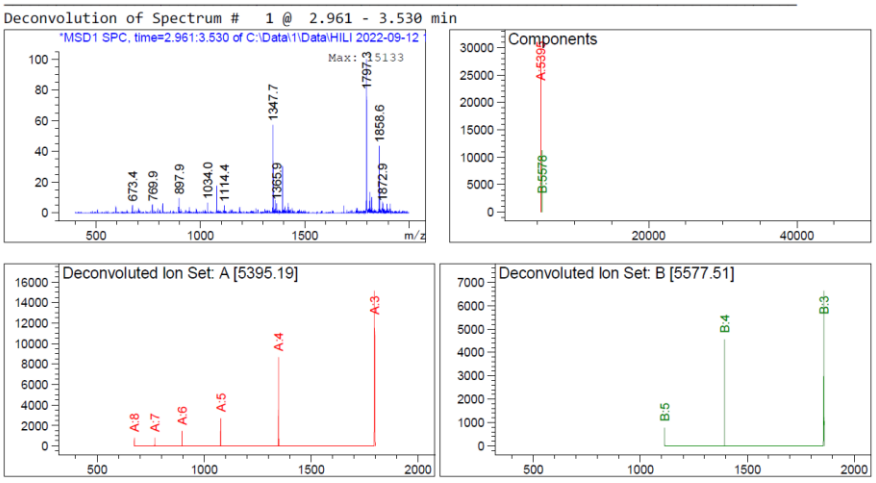
Figure 5-54: Deconvoluted LCMS data for **26**.



27

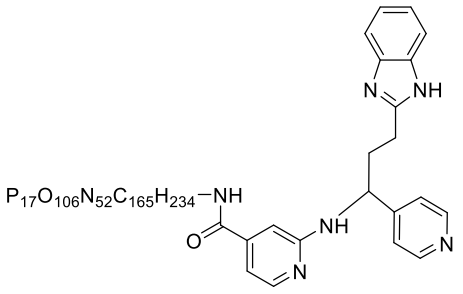
Molecular Weight: 5578.7940





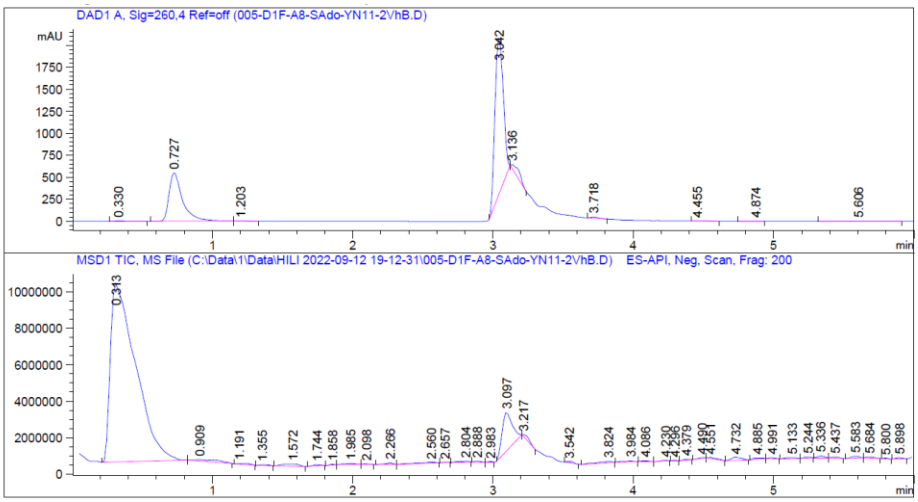
Component	Molecular Weight	Absolute Abundance	Relative Abundance
A	5395.19	28271	100.00
B	5577.51	11176	39.53

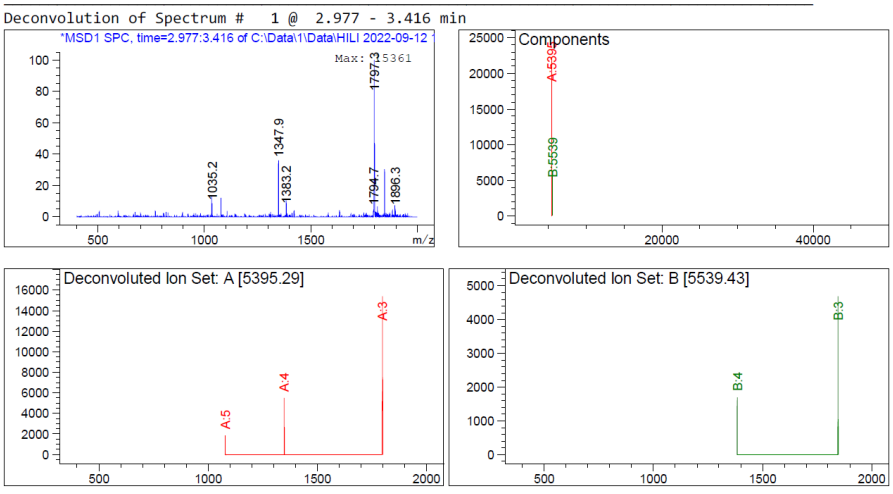
Figure 5-55: Deconvoluted LCMS data for **27**.



28

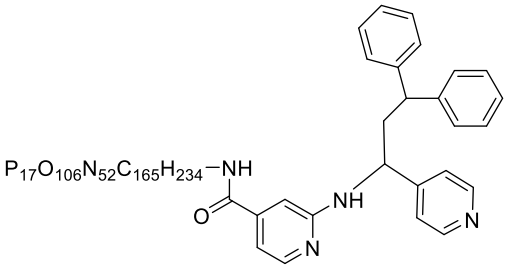
Molecular Weight: 5539.9230





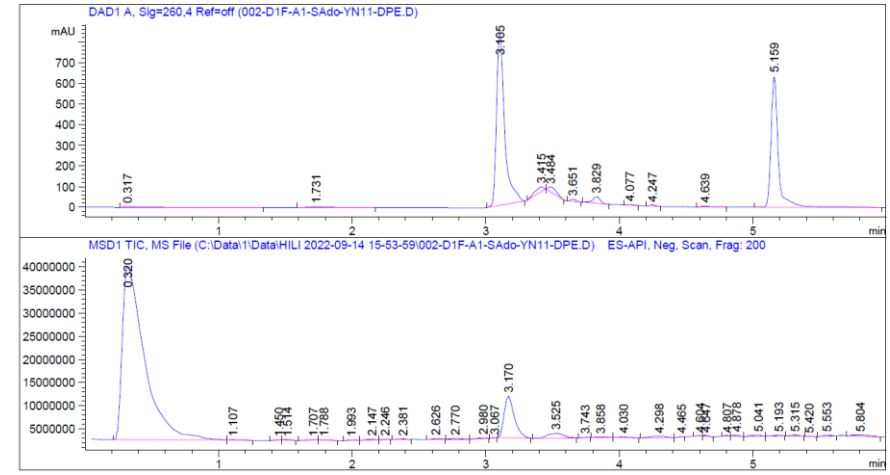
Component	Molecular Weight	Absolute Abundance	Relative Abundance
A	5395.29	22199	100.00
B	5539.43	6370	28.69

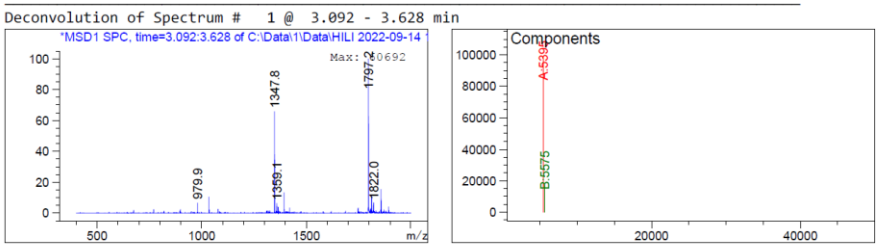
Figure 5-56: Deconvoluted LCMS data for **28**.

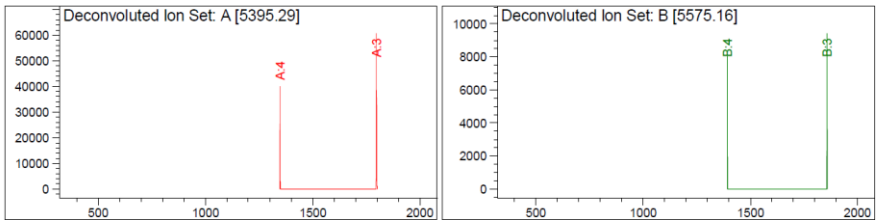


29

Molecular Weight: 5575.9960

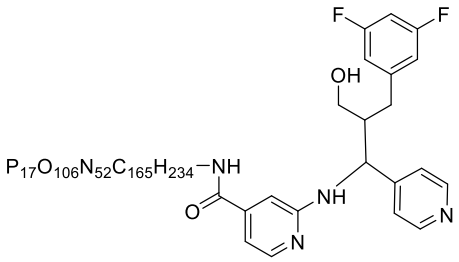






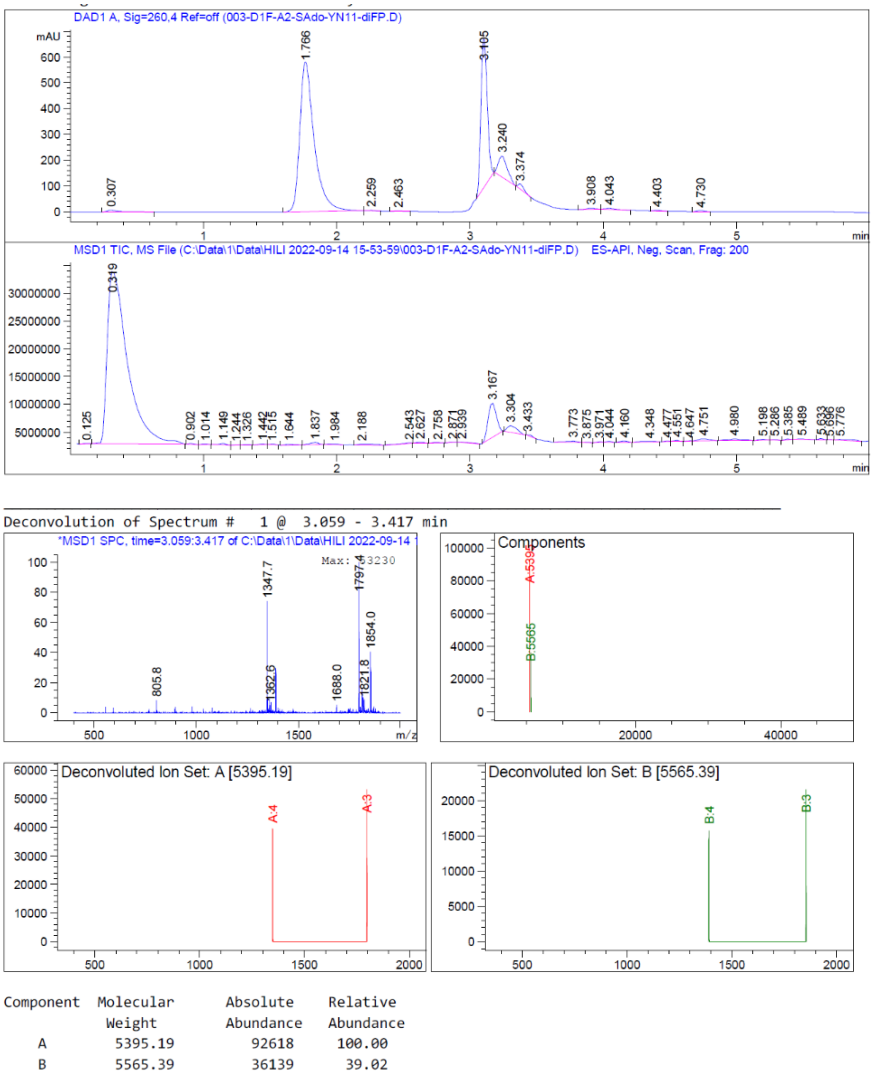
Component	Molecular Weight	Absolute Abundance	Relative Abundance
A	5395.29	98910	100.00
B	5575.16	16746	16.93

Figure 5-57: Deconvoluted LCMS data for **29**.



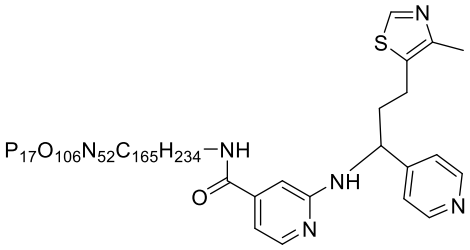
30

Molecular Weight: 5565.9048



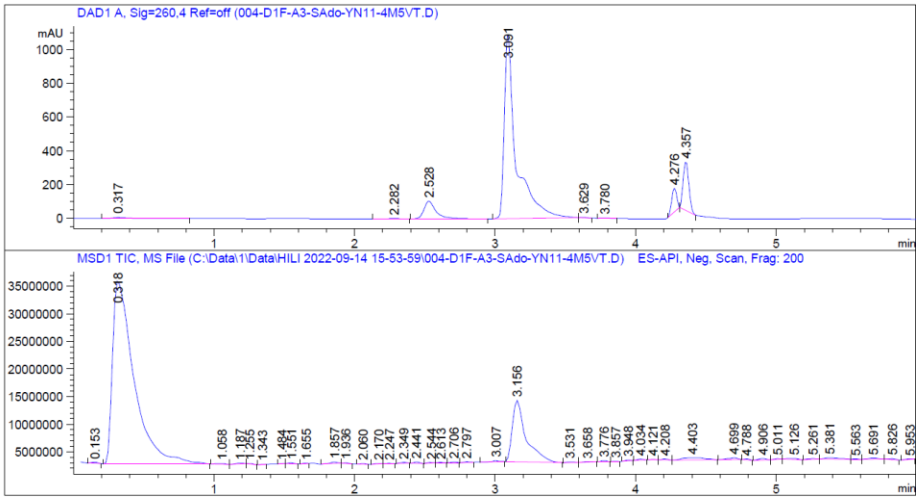
Component	Molecular Weight	Absolute Abundance	Relative Abundance
A	5395.19	92618	100.00
B	5565.39	36139	39.02

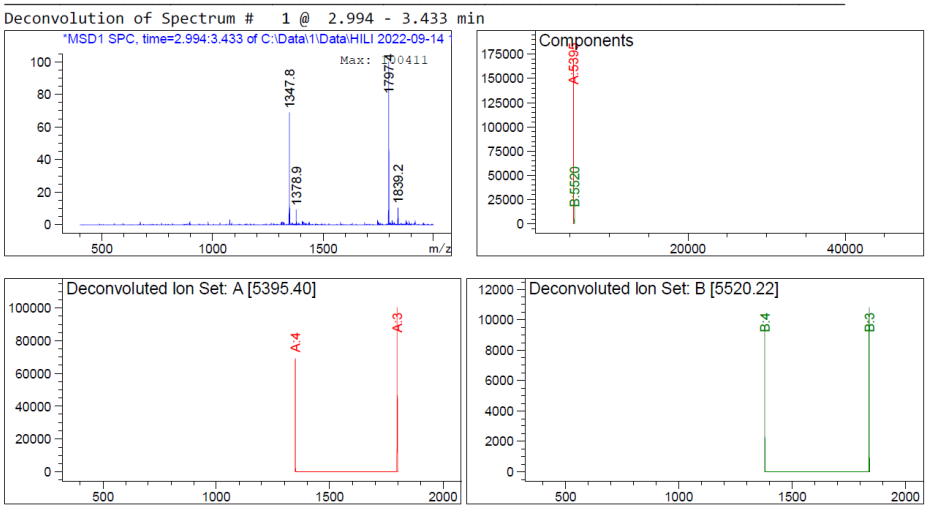
Figure 5-58: Deconvoluted LCMS data for **30**.



31

Molecular Weight: 5520.9350





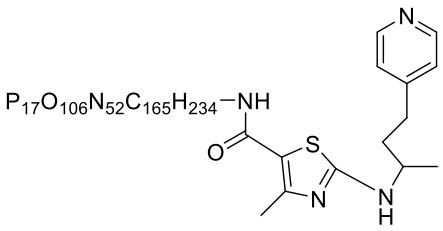
Component	Molecular Weight	Absolute Abundance	Relative Abundance
A	5395.40	169561	100.00
B	5520.22	19791	11.67

Figure 5-59: Deconvoluted LCMS data for **31**.

5.5.17. LCMS spectra and deconvolution results for 1f derivatives

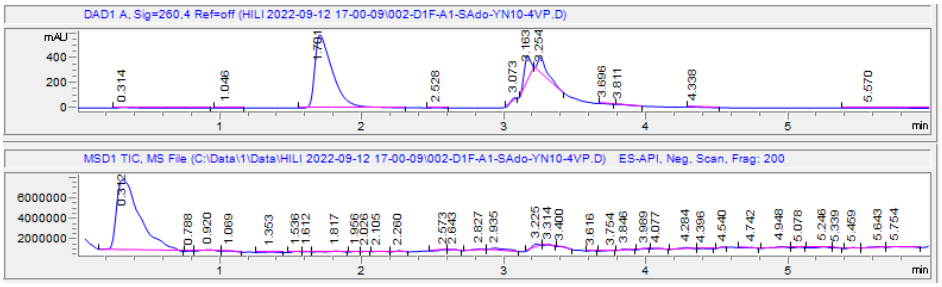
Table 5-11: Hydroaminoalkylation of various vinylarenes with DNA conjugate **1f**

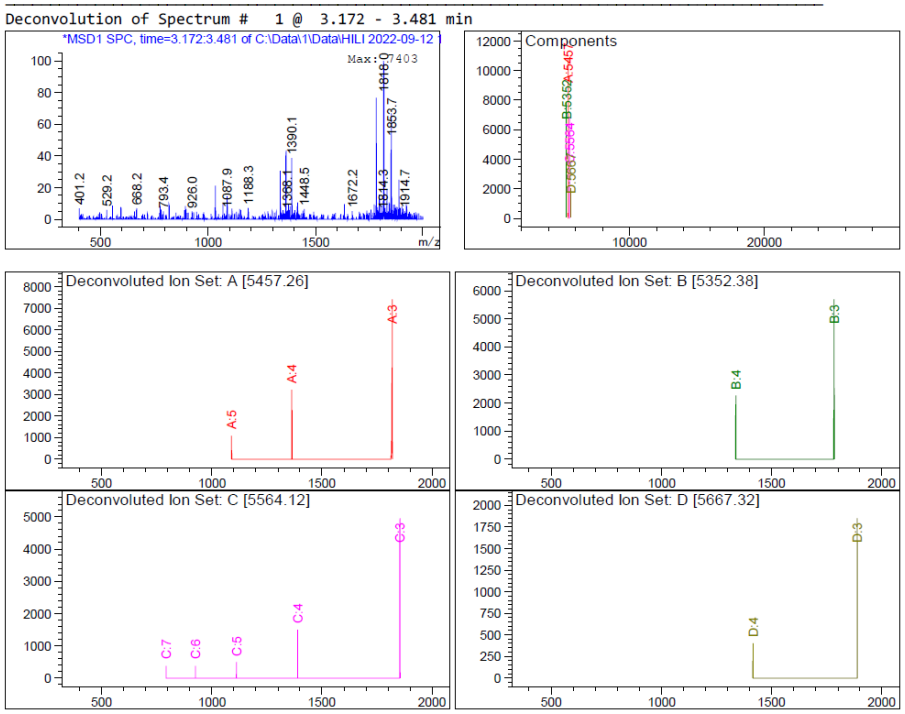
	Starting Material (**1f**)	Single Addition	Double Addition	Triple Addition	Other
1f + 4VP	29%	**15**: 39%	25%	7%	-
1f + 4CS	47%	**16**: 40%	13%	-	-
1f + 2VhB	72%	**18**: 28%	-	-	-
1f + DPE	73%	**19**: 27%	-	-	-
1f + 2BrS	80%	**17**: 20%	-	-	-
1f + diFP	57%	**20**: 23%	-	-	Dealkylated **1f**: 20%
1f + 4M5VT	87%	**22**: 13%	-	-	-
1f + 3EhP	91%	**21**: 9%	-	-	-
1f + 5EMP	93%	**24**: 7%	-	-	-
1f + 4FMS	100%	**23**: 0%	-	-	-
1f + 4MS	100%	-	-	-	-
1f + 4VBA	100%	-	-	-	-
1f + 4AS	100%	-	-	-	-



15

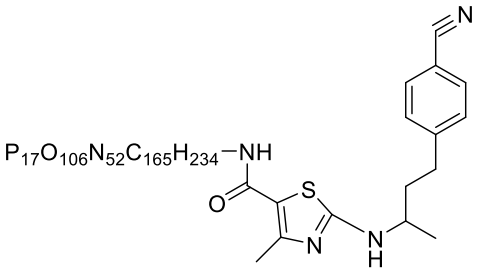
Molecular Weight: 5457.8760





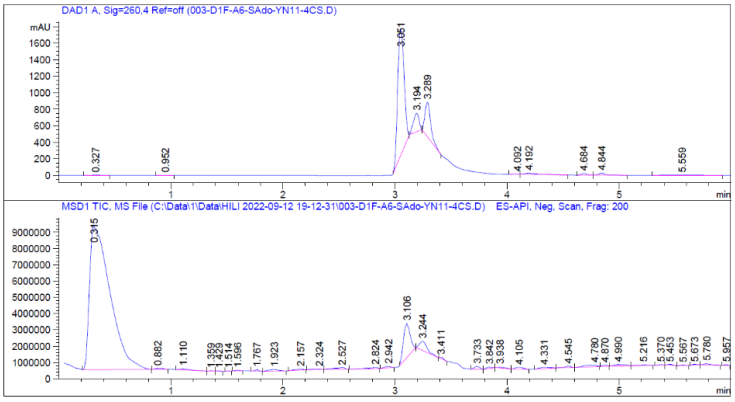
Component	Molecular Weight	Absolute Abundance	Relative Abundance
A	5457.26	10806	100.00
B	5352.38	7884	72.96
C	5564.12	6808	63.00
D	5667.32	2005	18.55

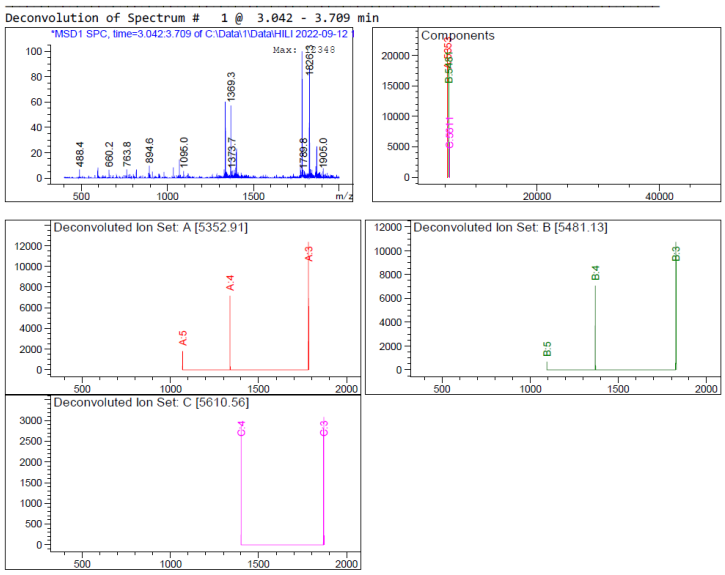
Figure 5-60: Deconvoluted LCMS data for **15**.



16

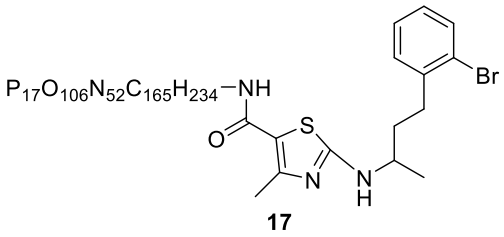
Molecular Weight: 5481.8980





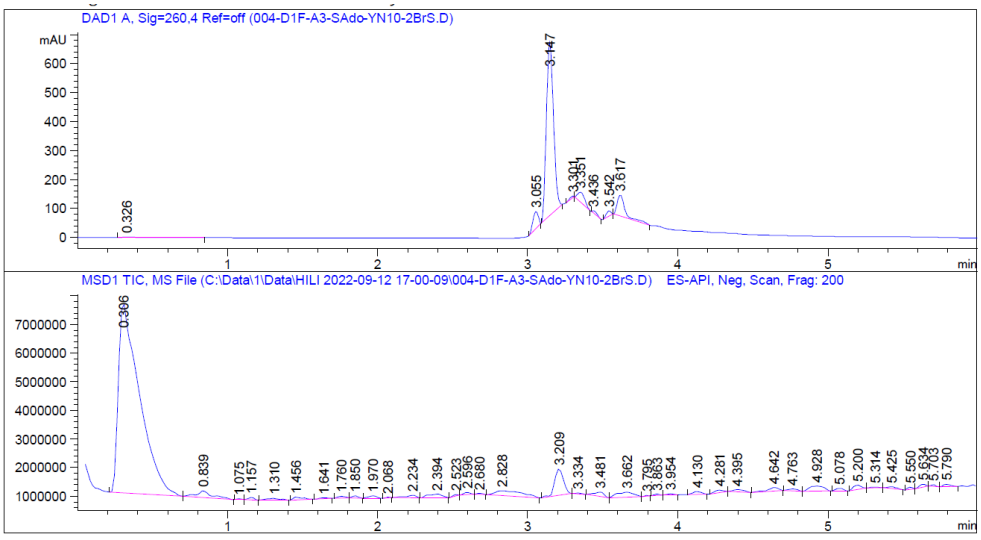
Component	Molecular Weight	Absolute Abundance	Relative Abundance
A	5352.91	20957	100.00
B	5481.13	18169	86.70
C	5610.56	5687	27.14

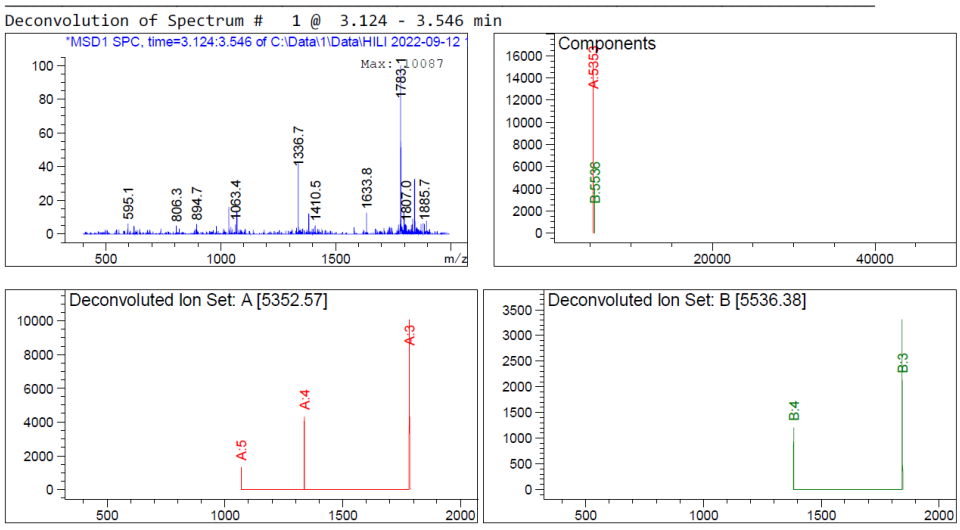
Figure 5-61: Deconvoluted LCMS data for **16**.



17

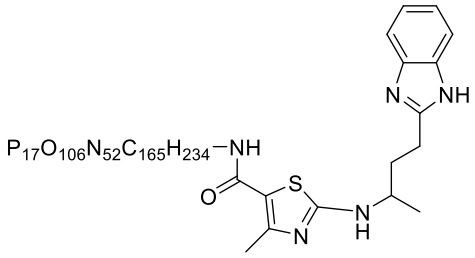
Molecular Weight: 5535.7840





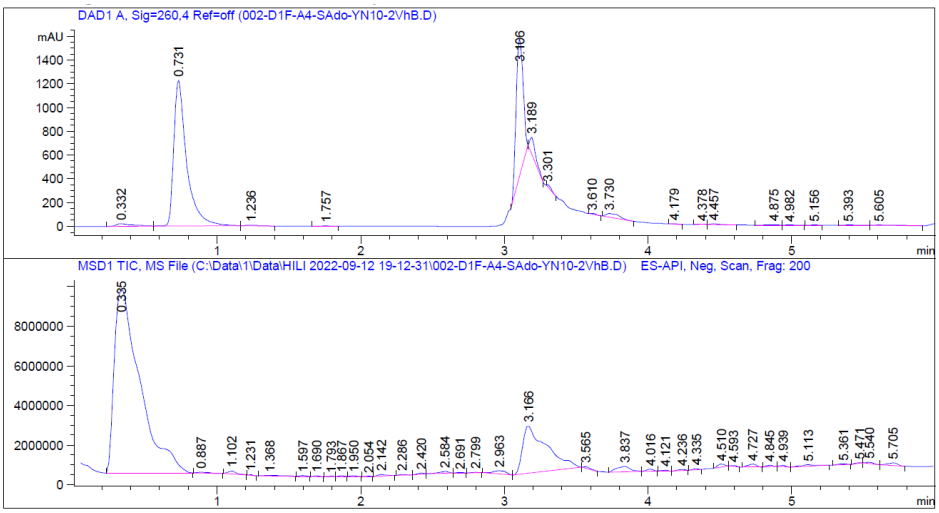
Component	Molecular Weight	Absolute Abundance	Relative Abundance
A	5352.57	15344	100.00
B	5536.38	3823	24.92

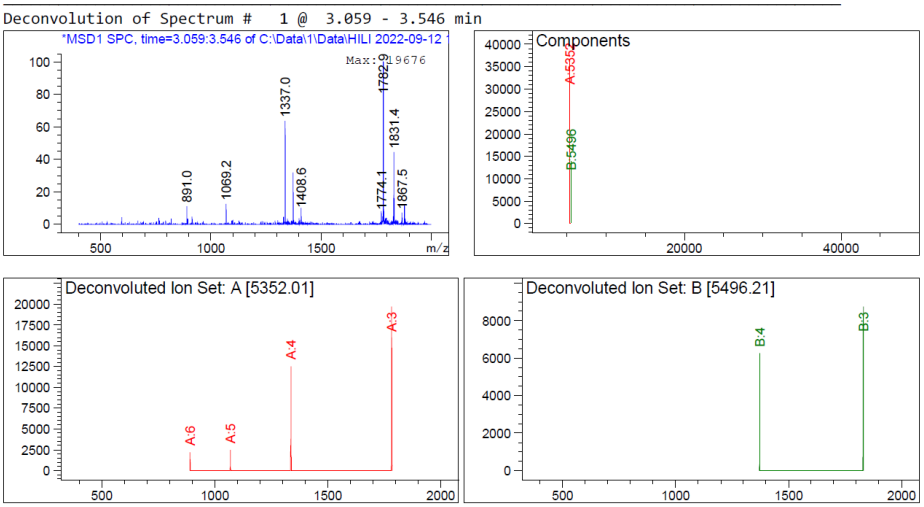
Figure 5-62: Deconvoluted LCMS data for **17**.



18

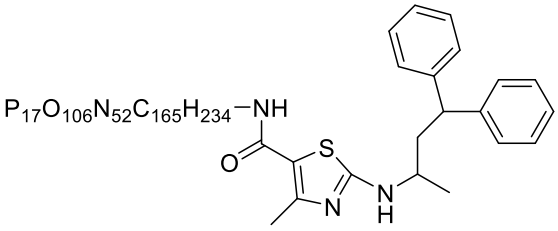
Molecular Weight: 5496.9130





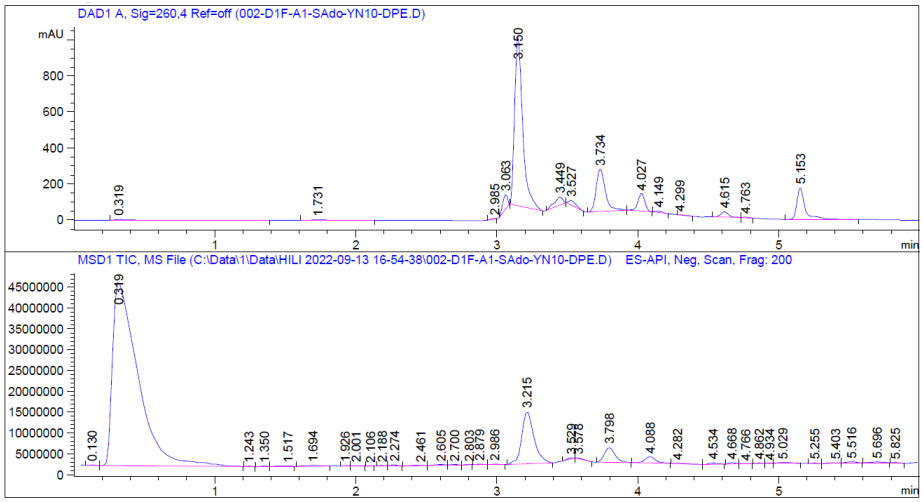
Component	Molecular Weight	Absolute Abundance	Relative Abundance
A	5352.01	36598	100.00
B	5496.21	13925	38.05

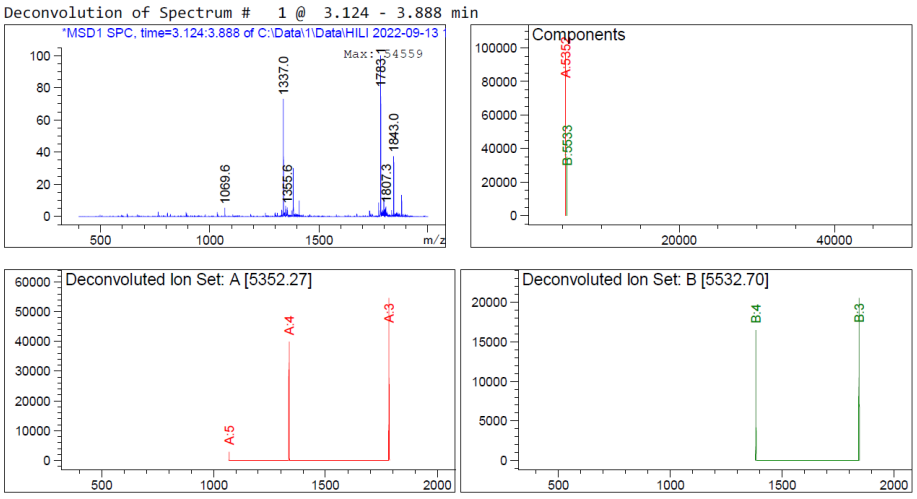
Figure 5-63: Deconvoluted LCMS data for **18**.



19

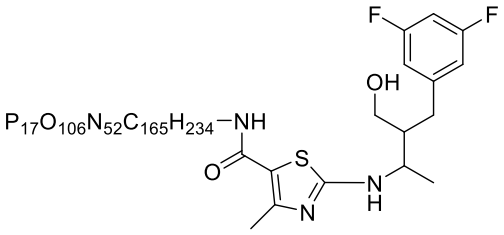
Molecular Weight: 5532.9860





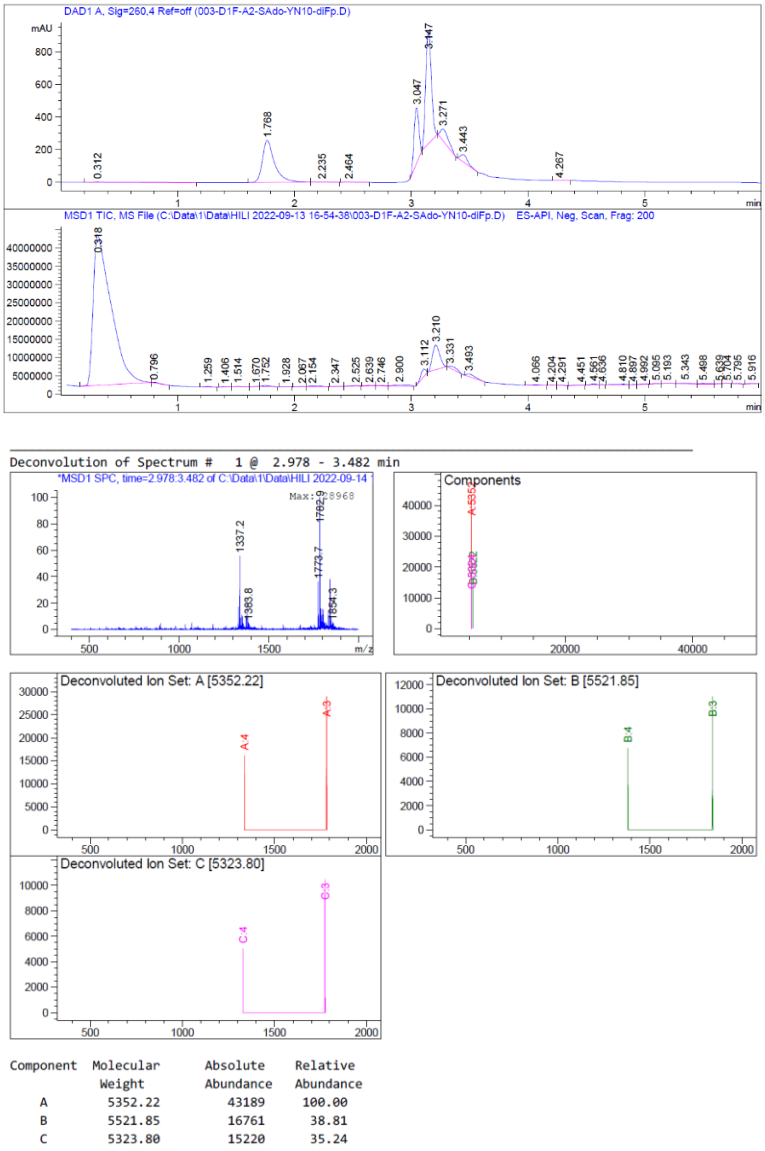
Component	Molecular Weight	Absolute Abundance	Relative Abundance
A	5352.27	96726	100.00
B	5532.70	34981	36.17

Figure 5-64: Deconvoluted LCMS data for **19**.



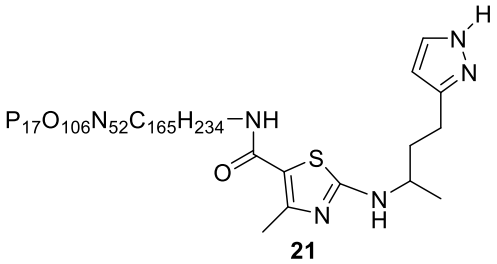
20

Molecular Weight: 5522.8948



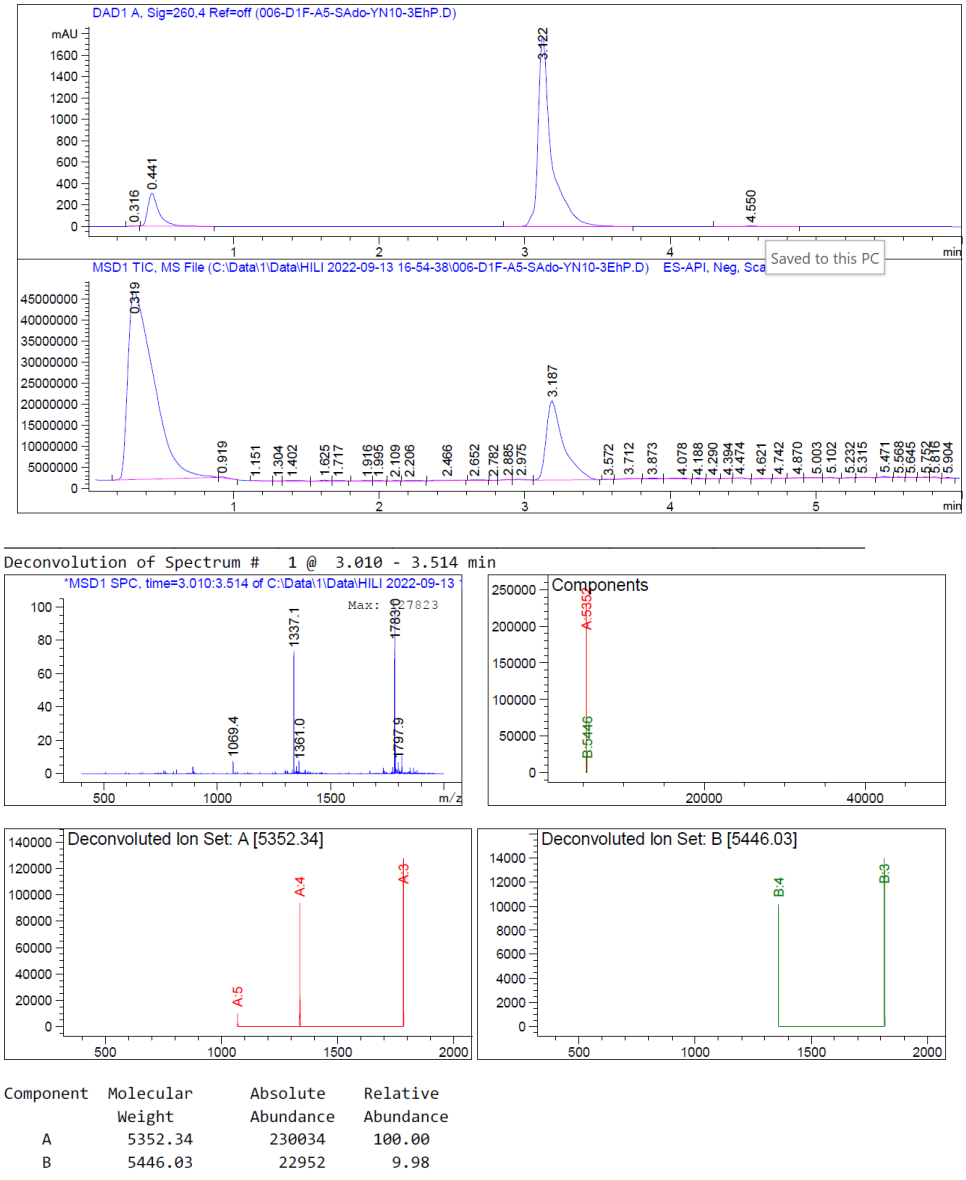
Component	Molecular Weight	Absolute Abundance	Relative Abundance
A	5352.22	43189	100.00
B	5521.85	16761	38.81
C	5323.80	15220	35.24

Figure 5-65: Deconvoluted LCMS data for **20**.



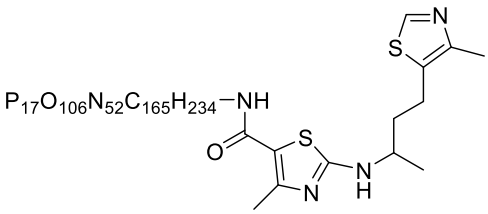
21

Molecular Weight: 5446.8530



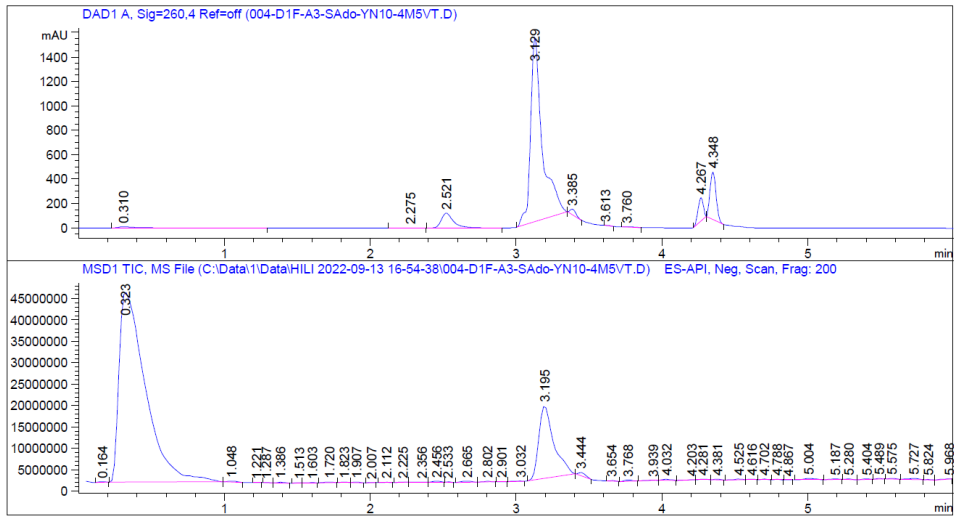
Component	Molecular Weight	Absolute Abundance	Relative Abundance
A	5352.34	230034	100.00
B	5446.03	22952	9.98

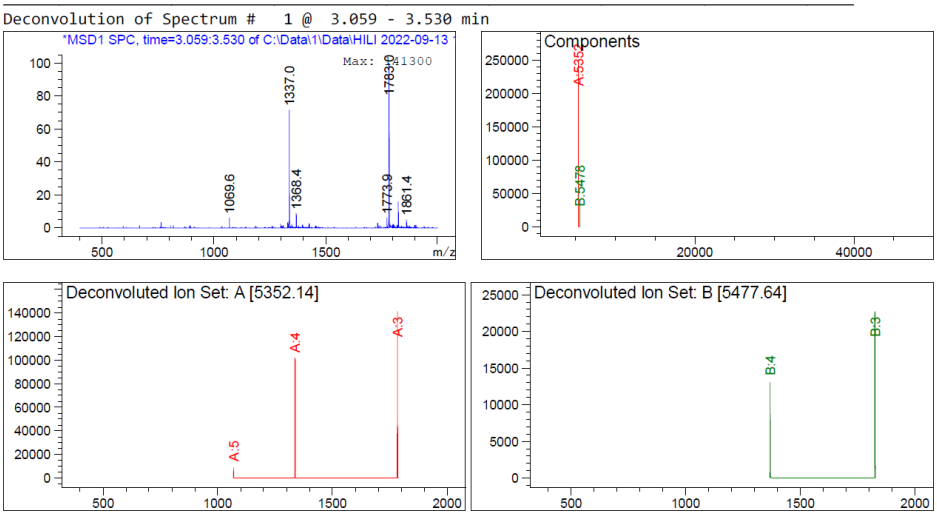
Figure 5-66: Deconvoluted LCMS data for **21**.



22

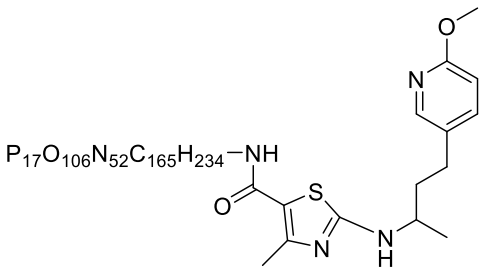
Molecular Weight: 5477.9250





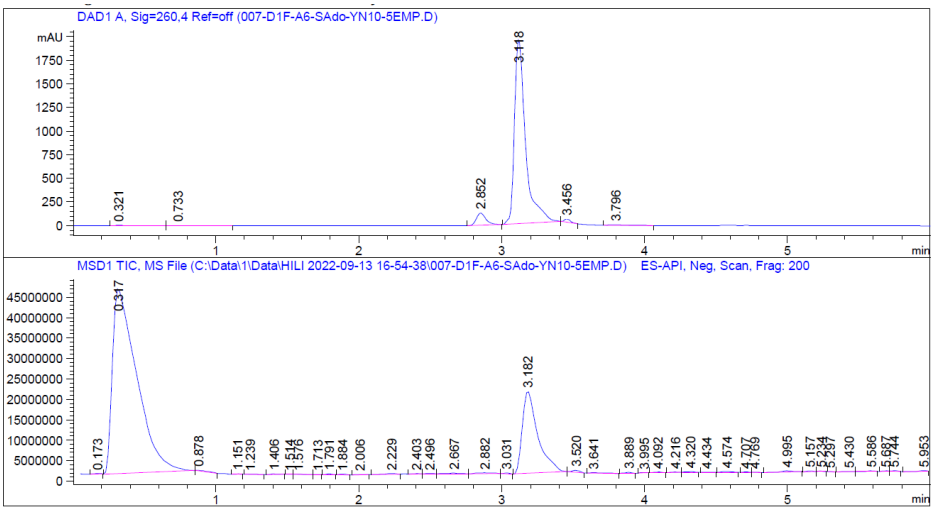
Component	Molecular Weight	Absolute Abundance	Relative Abundance
A	5352.14	249573	100.00
B	5477.64	35657	14.29

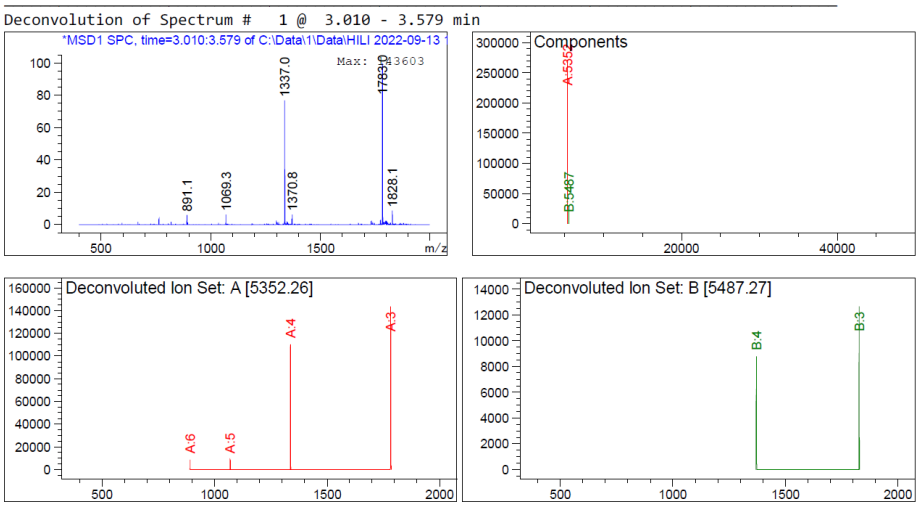
Figure 5-67: Deconvoluted LCMS data for **22**.



24

Molecular Weight: 5487.9020





Component	Molecular Weight	Absolute Abundance	Relative Abundance
A	5352.26	270283	100.00
B	5487.27	21224	7.85

Figure 5-68: Deconvoluted LCMS data for **24**.

References

(1) Dahm, R. Discovering DNA: Friedrich Miescher and the Early Years of Nucleic Acid Research. *Hum. Genet.* **2008**, *122* (6), 565–581. https://doi.org/10.1007/s00439-007-0433-0.

(2) Avery, O. T.; MacLeod, C. M.; McCarty, M. Studies on the Chemical Nature of the Substance Inducing Transformation of Pneumococcal Types. *J. Exp. Med.* **1944**, *79* (2), 137–158.

(3) Chargaff, E. Structure and Function of Nucleic Acids as Cell Constituents. *Fed. Proc.* **1951**, *10* (3), 654–659.

(4) Chargaff, E.; Vischer, E. The Composition of the Desoxypentose Nucleic Acids of Thymus and Spleen. *J. Biol. Chem.* **1949**, *177* (1), 405–416.

(5) Franklin, R. E.; Goslino, X. R. G. The Structure of Sodium Thymonucleate Fibres. I. The Influence of Water Content. 6.

(6) Watson, J. D.; Crick, F. H. C. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature* **1953**, *171* (4356), 737–738. https://doi.org/10.1038/171737a0.

(7) Venter, J. C.; Adams, M. D.; Myers, E. W.; Li, P. W.; Mural, R. J.; Sutton, G. G.; Smith, H. O.; Yandell, M.; Evans, C. A.; Holt, R. A.; Gocayne, J. D.; Amanatides, P.; Ballew, R. M.; Huson, D. H.; Wortman, J. R.; Zhang, Q.; Kodira, C. D.; Zheng, X. H.; Chen, L.; Skupski, M.; Subramanian, G.; Thomas, P. D.; Zhang, J.; Miklos, G. L. G.; Nelson, C.; Broder, S.; Clark, A. G.; Nadeau, J.; McKusick, V. A.; Zinder, N.; Levine, A. J.; Roberts, R. J.; Simon, M.; Slayman, C.; Hunkapiller, M.; Bolanos, R.; Delcher, A.; Dew, I.; Fasulo, D.; Flanigan, M.; Florea, L.; Halpern, A.; Hannenhalli, S.; Kravitz, S.; Levy, S.; Mobarry, C.; Reinert, K.; Remington, K.; Abu-Threideh, J.; Beasley, E.; Biddick, K.; Bonazzi, V.; Brandon, R.; Cargill, M.; Chandramouliswaran, I.; Charlab, R.; Chaturvedi, K.; Deng, Z.; Francesco, V. D.; Dunn, P.; Eilbeck, K.; Evangelista, C.; Gabrielian, A. E.; Gan, W.; Ge, W.; Gong, F.; Gu, Z.; Guan, P.; Heiman, T. J.; Higgins, M. E.; Ji, R.-R.; Ke, Z.; Ketchum, K. A.; Lai, Z.; Lei, Y.; Li, Z.; Li, J.; Liang, Y.; Lin, X.; Lu, F.; Merkulov, G. V.; Milshina, N.; Moore, H. M.; Naik, A. K.; Narayan, V. A.; Neelam, B.; Nusskern, D.; Rusch, D. B.; Salzberg, S.; Shao, W.; Shue, B.; Sun, J.; Wang, Z. Y.; Wang, A.; Wang, X.; Wang, J.; Wei, M.-H.; Wides, R.; Xiao, C.; Yan, C.; Yao, A.; Ye, J.; Zhan, M.; Zhang, W.; Zhang, H.; Zhao, Q.; Zheng, L.; Zhong, F.; Zhong, W.; Zhu, S. C.; Zhao, S.; Gilbert, D.; Baumhueter, S.; Spier, G.; Carter, C.; Cravchik, A.; Woodage, T.; Ali, F.; An, H.; Awe, A.; Baldwin, D.; Baden, H.; Barnstead, M.; Barrow, I.; Beeson, K.; Busam, D.; Carver, A.; Center, A.; Cheng, M. L.; Curry, L.; Danaher, S.; Davenport, L.; Desilets, R.; Dietz, S.; Dodson, K.; Doup, L.; Ferriera, S.; Garg, N.; Gluecksmann, A.; Hart, B.; Haynes, J.; Haynes, C.; Heiner, C.; Hladun, S.; Hostin, D.; Houck, J.; Howland, T.; Ibegwam, C.; Johnson, J.; Kalush, F.; Kline, L.; Koduru, S.; Love, A.; Mann, F.; May, D.; McCawley, S.; McIntosh, T.; McMullen, I.; Moy, M.; Moy, L.; Murphy, B.; Nelson, K.; Pfannkoch, C.; Pratts, E.; Puri, V.; Qureshi, H.; Reardon, M.; Rodriguez, R.; Rogers, Y.-H.; Romblad, D.; Ruhfel, B.; Scott, R.; Sitter, C.; Smallwood, M.; Stewart, E.; Strong, R.; Suh, E.; Thomas, R.; Tint, N. N.; Tse, S.; Vech, C.; Wang, G.; Wetter, J.; Williams, S.; Williams, M.; Windsor, S.; Winn-Deen, E.; Wolfe, K.; Zaveri, J.; Zaveri, K.; Abril, J. F.; Guigó, R.; Campbell, M. J.; Sjolander, K. V.; Karlak, B.; Kejariwal, A.; Mi, H.; Lazareva, B.; Hatton, T.; Narechania, A.; Diemer, K.; Muruganujan, A.; Guo, N.; Sato, S.; Bafna, V.; Istrail, S.; Lippert, R.; Schwartz, R.; Walenz, B.; Yooseph, S.; Allen, D.; Basu, A.; Baxendale, J.; Blick, L.; Caminha, M.; Carnes-Stine, J.; Caulk, P.; Chiang, Y.-H.; Coyne, M.; Dahlke, C.; Mays, A. D.; Dombroski, M.; Donnelly, M.; Ely, D.; Esparham, S.; Fosler, C.; Gire, H.; Glanowski, S.; Glasser, K.; Glodek, A.; Gorokhov, M.; Graham, K.; Gropman, B.; Harris, M.; Heil, J.; Henderson, S.; Hoover, J.; Jennings, D.; Jordan, C.; Jordan, J.; Kasha, J.; Kagan, L.; Kraft, C.; Levitsky, A.; Lewis, M.; Liu, X.; Lopez, J.; Ma, D.; Majoros, W.; McDaniel, J.; Murphy, S.; Newman, M.; Nguyen, T.; Nguyen, N.; Nodell, M.; Pan, S.; Peck, J.; Peterson, M.; Rowe, W.; Sanders, R.; Scott, J.; Simpson, M.; Smith, T.; Sprague, A.;

Stockwell, T.; Turner, R.; Venter, E.; Wang, M.; Wen, M.; Wu, D.; Wu, M.; Xia, A.; Zandieh, A.; Zhu, X. The Sequence of the Human Genome. *Science* **2001**, *291* (5507), 1304–1351. https://doi.org/10.1126/science.1058040.

(8) Holley, R. W.; Apgar, J.; Everett, G. A.; Madison, J. T.; Marquisee, M.; Merrill, S. H.; Penswick, J. R.; Zamir, A. Structure of a Ribonucleic Acid. *Science* **1965**, *147* (3664), 1462–1465. https://doi.org/10.1126/science.147.3664.1462.

(9) Lodish, H.; Berk, A.; Zipursky, S. L.; Matsudaira, P.; Baltimore, D.; Darnell, J. The Three Roles of RNA in Protein Synthesis. *Mol. Cell Biol. 4th Ed.* **2000**.

(10) Fu, Y.; He, C. Nucleic Acid Modifications with Epigenetic Significance. *Curr. Opin. Chem. Biol.* **2012**, *16* (5–6), 516–524. https://doi.org/10.1016/j.cbpa.2012.10.002.

(11) Sood, A. J.; Viner, C.; Hoffman, M. M. DNAmod: The DNA Modification Database. *J. Cheminformatics* **2019**, *11* (1), 30. https://doi.org/10.1186/s13321-019-0349-4.

(12) Machnicka, M. A.; Milanowska, K.; Osman Oglou, O.; Purta, E.; Kurkowska, M.; Olchowik, A.; Januszewski, W.; Kalinowski, S.; Dunin-Horkawicz, S.; Rother, K. M.; Helm, M.; Bujnicki, J. M.; Grosjean, H. MODOMICS: A Database of RNA Modification Pathways—2013 Update. *Nucleic Acids Res.* **2013**, *41* (D1), D262–D267. https://doi.org/10.1093/nar/gks1007.

(13) Kumar, S.; Chinnusamy, V.; Mohapatra, T. Epigenetics of Modified DNA Bases: 5-Methylcytosine and Beyond. *Front. Genet.* **2018**, *9*, 640. https://doi.org/10.3389/fgene.2018.00640.

(14) Wyatt, G. R. Recognition and Estimation of 5-Methylcytosine in Nucleic Acids. *Biochem. J.* **1951**, *48* (5), 581–584. https://doi.org/10.1042/bj0480581.

(15) Li, E.; Zhang, Y. DNA Methylation in Mammals. *Cold Spring Harb. Perspect. Biol.* **2014**, *6* (5), a019133. https://doi.org/10.1101/cshperspect.a019133.

(16) Breiling, A.; Lyko, F. Epigenetic Regulatory Functions of DNA Modifications: 5-Methylcytosine and Beyond. *Epigenetics Chromatin* **2015**, *8* (1), 24. https://doi.org/10.1186/s13072-015-0016-6.

(17) Campbell, J. L.; Kleckner, N. E. Coli OriC and the DnaA Gene Promoter Are Sequestered from Dam Methyltransferase Following the Passage of the Chromosomal Replication Fork. *Cell* **1990**, *62* (5), 967–979. https://doi.org/10.1016/0092-8674(90)90271-F.

(18) Messer, W.; Noyer-Weidner, M. Timing and Targeting: The Biological Functions of Dam Methylation in E. Coli. *Cell* **1988**, *54* (6), 735–737. https://doi.org/10.1016/S0092-8674(88)90911-7.

(19) Xiao, C.-L.; Zhu, S.; He, M.; Chen, D.; Zhang, Q.; Chen, Y.; Yu, G.; Liu, J.; Xie, S.-Q.; Luo, F.; Liang, Z.; Wang, D.-P.; Bo, X.-C.; Gu, X.-F.; Wang, K.; Yan, G.-R. N6-Methyladenine DNA Modification in the Human Genome. *Mol. Cell* **2018**, *71* (2), 306-318.e7. https://doi.org/10.1016/j.molcel.2018.06.015.

(20) Liu, J.; Zhu, Y.; Luo, G.-Z.; Wang, X.; Yue, Y.; Wang, X.; Zong, X.; Chen, K.; Yin, H.; Fu, Y.; Han, D.; Wang, Y.; Chen, D.; He, C. Abundant DNA 6mA Methylation during Early Embryogenesis of Zebrafish and Pig. *Nat. Commun.* **2016**, *7* (1), 13052. https://doi.org/10.1038/ncomms13052.

(21) Yao, B.; Cheng, Y.; Wang, Z.; Li, Y.; Chen, L.; Huang, L.; Zhang, W.; Chen, D.; Wu, H.; Tang, B.; Jin, P. DNA N6-Methyladenine Is Dynamically Regulated in the Mouse Brain Following Environmental Stress. *Nat. Commun.* **2017**, *8* (1), 1122. https://doi.org/10.1038/s41467-017-01195-y.

(22) Wu, T. P.; Wang, T.; Seetin, M. G.; Lai, Y.; Zhu, S.; Lin, K.; Liu, Y.; Byrum, S. D.; Mackintosh, S. G.; Zhong, M.; Tackett, A.; Wang, G.; Hon, L. S.; Fang, G.; Swenberg, J. A.; Xiao, A. Z. DNA Methylation on N6-Adenine in Mammalian Embryonic Stem Cells. *Nature* **2016**, *532* (7599), 329–333. https://doi.org/10.1038/nature17640.

(23) Cohn, W. E. Pseudouridine, a Carbon-Carbon Linked Ribonucleoside in Ribonucleic Acids: Isolation, Structure, and Chemical Characteristics. *J. Biol. Chem.* **1960**, *235*, 1488–1498.

(24) Jühling, F.; Mörl, M.; Hartmann, R. K.; Sprinzl, M.; Stadler, P. F.; Pütz, J. tRNAdb 2009: Compilation of tRNA Sequences and tRNA Genes. *Nucleic Acids Res.* **2009**, *37* (Database issue), D159-162. https://doi.org/10.1093/nar/gkn772.

(25) Wei, C.-M.; Gershowitz, A.; Moss, B. Methylated Nucleotides Block 5′ Terminus of HeLa Cell Messenger RNA. *Cell* **1975**, *4* (4), 379–386. https://doi.org/10.1016/0092-8674(75)90158-0.

(26) Shi, H.; Wei, J.; He, C. Where, When, and How: Context-Dependent Functions of RNA Methylation Writers, Readers, and Erasers. *Mol. Cell* **2019**, *74* (4), 640–650. https://doi.org/10.1016/j.molcel.2019.04.025.

(27) Boo, S. H.; Kim, Y. K. The Emerging Role of RNA Modifications in the Regulation of mRNA Stability. *Exp. Mol. Med.* **2020**, *52* (3), 400–408. https://doi.org/10.1038/s12276-020-0407-z.

(28) Wei, W.; Ji, X.; Guo, X.; Ji, S. Regulatory Role of N ^{6} -Methyladenosine (m ^{6} A) Methylation in RNA Processing and Human Diseases: R EGULATORY R OLE O F m ^{6} A. *J. Cell. Biochem.* **2017**, *118* (9), 2534–2543. https://doi.org/10.1002/jcb.25967.

(29) Boulias, K.; Toczydłowska-Socha, D.; Hawley, B. R.; Liberman, N.; Takashima, K.; Zaccara, S.; Guez, T.; Vasseur, J.-J.; Debart, F.; Aravind, L.; Jaffrey, S. R.; Greer, E. L. Identification of the m6Am Methyltransferase PCIF1 Reveals the Location and Functions of m6Am in the Transcriptome. *Mol. Cell* **2019**, *75* (3), 631-643.e8. https://doi.org/10.1016/j.molcel.2019.06.006.

(30) Linder, B.; Grozhik, A. V.; Olarerin-George, A. O.; Meydan, C.; Mason, C. E.; Jaffrey, S. R. Single-Nucleotide-Resolution Mapping of m6A and m6Am throughout the Transcriptome. *Nat. Methods* **2015**, *12* (8), 767–772. https://doi.org/10.1038/nmeth.3453.

(31) Dominissini, D.; Moshitch-Moshkovitz, S.; Schwartz, S.; Salmon-Divon, M.; Ungar, L.; Osenberg, S.; Cesarkas, K.; Jacob-Hirsch, J.; Amariglio, N.; Kupiec, M.; Sorek, R.; Rechavi, G. Topology of the Human and Mouse m6A RNA Methylomes Revealed by m6A-Seq. *Nature* **2012**, *485* (7397), 201–206. https://doi.org/10.1038/nature11112.

(32) Meyer, K. D.; Saletore, Y.; Zumbo, P.; Elemento, O.; Mason, C. E.; Jaffrey, S. R. Comprehensive Analysis of mRNA Methylation Reveals Enrichment in 3′ UTRs and near Stop Codons. *Cell* **2012**, *149* (7), 1635–1646. https://doi.org/10.1016/j.cell.2012.05.003.

(33) Chen, K.; Lu, Z.; Wang, X.; Fu, Y.; Luo, G.-Z.; Liu, N.; Han, D.; Dominissini, D.; Dai, Q.; Pan, T.; He, C. High-Resolution N6-Methyladenosine (m6A) Map Using Photo-Crosslinking-Assisted m6A Sequencing. *Angew. Chem. Int. Ed.* **2015**, *54* (5), 1587–1590. https://doi.org/10.1002/anie.201410647.

(34) Reynaud, C.; Bruno, C.; Boullanger, P.; Grange, J.; Barbesti, S.; Niveleau, A. Monitoring of Urinary Excretion of Modified Nucleosides in Cancer Patients Using a Set of Six Monoclonal Antibodies. *Cancer Lett.* **1992**, *61* (3), 255–262. https://doi.org/10.1016/0304-3835(92)90296-8.

(35) Achwal, C. W.; Iyer, C. A.; Chandra, H. S. Immunochemical Evidence for the Presence of 5mC, 6mA and 7mG in Human, Drosophila and Mealybug DNA. *FEBS Lett.* **1983**, *158* (2), 353–358. https://doi.org/10.1016/0014-5793(83)80612-7.

(36) Sano, H.; Royer, H. D.; Sager, R. Identification of 5-Methylcytosine in DNA Fragments Immobilized on Nitrocellulose Paper. *Proc. Natl. Acad. Sci.* **1980**, *77* (6), 3581–3585. https://doi.org/10.1073/pnas.77.6.3581.

(37) Itoh, K.; Mizugaki, M.; Ishida, N. Preparation of a Monoclonal Antibody Specific for 1-Methyladenosine and Its Application for the Detection of Elevated Levels of 1-Methyladenosine in Urines from Cancer Patients. *Jpn. J. Cancer Res.* **1988**, *79* (10), 1130–1138. https://doi.org/10.1111/j.1349-7006.1988.tb01536.x.

(38) Zhang, Z.; Chen, L.-Q.; Zhao, Y.-L.; Yang, C.-G.; Roundtree, I. A.; Zhang, Z.; Ren, J.; Xie, W.; He, C.; Luo, G.-Z. Single-Base Mapping of m6A by an Antibody-Independent Method. *Sci. Adv.* **2019**, *5* (7), eaax0250. https://doi.org/10.1126/sciadv.aax0250.

(39) Aschenbrenner, J.; Werner, S.; Marchand, V.; Adam, M.; Motorin, Y.; Helm, M.; Marx, A. Engineering of a DNA Polymerase for Direct m ^{6} A Sequencing. *Angew. Chem. Int. Ed.* **2018**, *57* (2), 417–421. https://doi.org/10.1002/anie.201710209.

(40) Harcourt, E. M.; Ehrenschwender, T.; Batista, P. J.; Chang, H. Y.; Kool, E. T. Identification of a Selective Polymerase Enables Detection of N ^{6} -Methyladenosine in RNA. *J. Am. Chem. Soc.* **2013**, *135* (51), 19079–19082. https://doi.org/10.1021/ja4105792.

(41) Wang, S.; Wang, J.; Zhang, X.; Fu, B.; Song, Y.; Ma, P.; Gu, K.; Zhou, X.; Zhang, X.; Tian, T.; Zhou, X. N ^{6} -Methyladenine Hinders RNA- and DNA-Directed DNA Synthesis: Application in Human rRNA Methylation Analysis of Clinical Specimens. *Chem. Sci.* **2016**, *7* (2), 1440–1446. https://doi.org/10.1039/C5SC02902C.

(42) Frommer, M.; McDonald, L. E.; Millar, D. S.; Collis, C. M.; Watt, F.; Grigg, G. W.; Molloy, P. L.; Paul, C. L. A Genomic Sequencing Protocol That Yields a Positive Display of 5-Methylcytosine Residues in Individual DNA Strands. *Proc. Natl. Acad. Sci. U. S. A.* **1992**, *89* (5), 1827–1831.

(43) Okamoto, A.; Tainaka, K.; Kamei, T. Sequence-Selective Osmium Oxidation of DNA: Efficient Distinction between 5-Methylcytosine and Cytosine. *Org. Biomol. Chem.* **2006**, *4* (9), 1638–1640. https://doi.org/10.1039/B600401F.

(44) Bareyt, S.; Carell, T. Selective Detection of 5-Methylcytosine Sites in DNA. *Angew. Chem. Int. Ed.* **2008**, *47* (1), 181–184. https://doi.org/10.1002/anie.200702159.

(45) Lovejoy, A. F.; Riordan, D. P.; Brown, P. O. Transcriptome-Wide Mapping of Pseudouridines: Pseudouridine Synthases Modify Specific mRNAs in S. Cerevisiae. *PloS One* **2014**, *9* (10), e110799. https://doi.org/10.1371/journal.pone.0110799.

(46) Durairaj, A.; Limbach, P. A. Improving CMC-Derivatization of Pseudouridine in RNA for Mass Spectrometric Detection. *Anal. Chim. Acta* **2008**, *612* (2), 173–181. https://doi.org/10.1016/j.aca.2008.02.026.

(47) Greer, E. L.; Blanco, M. A.; Gu, L.; Sendinc, E.; Liu, J.; Aristizábal-Corrales, D.; Hsu, C.-H.; Aravind, L.; He, C.; Shi, Y. DNA Methylation on N6-Adenine in C. Elegans. *Cell* **2015**, *161* (4), 868–878. https://doi.org/10.1016/j.cell.2015.04.005.

(48) Zhang, G.; Huang, H.; Liu, D.; Cheng, Y.; Liu, X.; Zhang, W.; Yin, R.; Zhang, D.; Zhang, P.; Liu, J.; Li, C.; Liu, B.; Luo, Y.; Zhu, Y.; Zhang, N.; He, S.; He, C.; Wang, H.; Chen, D. N6-Methyladenine DNA Modification in Drosophila. *Cell* **2015**, *161* (4), 893–906. https://doi.org/10.1016/j.cell.2015.04.018.

(49) Goh, W. S. S. Single-Nucleotide-Resolution Sequencing of N<Superscript>6</Superscript>-Methyldeoxyadenosine. *DNA Modif.* **2021**, 369–377. https://doi.org/10.1007/978-1-0716-0876-0_28.

(50) Clark, T. A.; Murray, I. A.; Morgan, R. D.; Kislyuk, A. O.; Spittle, K. E.; Boitano, M.; Fomenkov, A.; Roberts, R. J.; Korlach, J. Characterization of DNA Methyltransferase Specificities Using Single-Molecule, Real-Time DNA Sequencing. *Nucleic Acids Res.* **2012**, *40* (4), e29–e29. https://doi.org/10.1093/nar/gkr1146.

(51) Greenberg, M. V. C.; Bourc'his, D. The Diverse Roles of DNA Methylation in Mammalian Development and Disease. *Nat. Rev. Mol. Cell Biol.* **2019**, *20* (10), 590–607. https://doi.org/10.1038/s41580-019-0159-6.

(52) Zaccara, S.; Ries, R. J.; Jaffrey, S. R. Reading, Writing and Erasing mRNA Methylation. *Nat. Rev. Mol. Cell Biol.* **2019**, *20* (10), 608–624. https://doi.org/10.1038/s41580-019-0168-5.

(53) Ma, C.; Niu, R.; Huang, T.; Shao, L.-W.; Peng, Y.; Ding, W.; Wang, Y.; Jia, G.; He, C.; Li, C.-Y.; He, A.; Liu, Y. N6-Methyldeoxyadenine Is a Transgenerational Epigenetic Signal for Mitochondrial Stress Adaptation. *Nat. Cell Biol.* **2019**, *21* (3), 319–327. https://doi.org/10.1038/s41556-018-0238-5.

(54) Fu, Y.; Luo, G.-Z.; Chen, K.; Deng, X.; Yu, M.; Han, D.; Hao, Z.; Liu, J.; Lu, X.; Doré, L. C.; Weng, X.; Ji, Q.; Mets, L.; He, C. N6-Methyldeoxyadenosine Marks Active Transcription Start Sites in Chlamydomonas. *Cell* **2015**, *161* (4), 879–892. https://doi.org/10.1016/j.cell.2015.04.010.

(55) Schiffers, S.; Ebert, C.; Rahimoff, R.; Kosmatchev, O.; Steinbacher, J.; Bohne, A.-V.; Spada, F.; Michalakis, S.; Nickelsen, J.; Müller, M.; Carell, T. Quantitative LC-MS Provides No Evidence for m ^{6} DA or m ^{4} DC in the Genome of Mouse Embryonic Stem Cells and Tissues. *Angew. Chem. Int. Ed.* **2017**, *56* (37), 11268–11271. https://doi.org/10.1002/anie.201700424.

(56) Schadt, E. E.; Banerjee, O.; Fang, G.; Feng, Z.; Wong, W. H.; Zhang, X.; Kislyuk, A.; Clark, T. A.; Luong, K.; Keren-Paz, A.; Chess, A.; Kumar, V.; Chen-Plotkin, A.; Sondheimer, N.; Korlach, J.; Kasarskis, A. Modeling Kinetic Rate Variation in Third Generation DNA Sequencing Data to Detect Putative Modifications to DNA Bases. *Genome Res.* **2013**, *23* (1), 129–141. https://doi.org/10.1101/gr.136739.111.

(57) Caldwell, W. T.; Tyson, F. T.; Lauer, L. Substituted 2-Sulfonamido-5-Aminopyridines. II. *J. Am. Chem. Soc.* **1944**, *66* (9), 1479–1484. https://doi.org/10.1021/ja01237a018.

(58) Basilio, C.; Wahba, A. J.; Lengyel, P.; Speyer, J. F.; Ochoa, S. SYNTHETIC POLYNUCLEOTIDES AND THE AMINO ACID CODE, V. *Proc. Natl. Acad. Sci.* **1962**, *48* (4), 613–616. https://doi.org/10.1073/pnas.48.4.613.

(59) Nabel, C. S.; Lee, J. W.; Wang, L. C.; Kohli, R. M. Nucleic Acid Determinants for Selective Deamination of DNA over RNA by Activation-Induced Deaminase. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110* (35), 14225–14230. https://doi.org/10.1073/pnas.1306345110.

(60) Langmead, B.; Trapnell, C.; Pop, M.; Salzberg, S. L. Ultrafast and Memory-Efficient Alignment of Short DNA Sequences to the Human Genome. *Genome Biol.* **2009**, *10* (3), R25. https://doi.org/10.1186/gb-2009-10-3-r25.

(61) Punekar, A. S.; Liljeruhm, J.; Shepherd, T. R.; Forster, A. C.; Selmer, M. Structural and Functional Insights into the Molecular Mechanism of rRNA m6A Methyltransferase RlmJ. *Nucleic Acids Res.* **2013**, *41* (20), 9537–9548. https://doi.org/10.1093/nar/gkt719.

(62) Zhou, H.; Rauch, S.; Dai, Q.; Cui, X.; Zhang, Z.; Nachtergaele, S.; Sepich, C.; He, C.; Dickinson, B. C. Evolution of a Reverse Transcriptase to Map N1-Methyladenosine in Human Messenger RNA. *Nat. Methods* **2019**, *16* (12), 1281–1288. https://doi.org/10.1038/s41592-019-0550-4.

(63) Li, X.; Guo, S.; Cui, Y.; Zhang, Z.; Luo, X.; Angelova, M. T.; Landweber, L. F.; Wang, Y.; Wu, T. P. NT-Seq: A Chemical-Based Sequencing Method for Genomic Methylome Profiling. *Genome Biol.* **2022**, *23* (1), 122. https://doi.org/10.1186/s13059-022-02689-9.

(64) Bolger, A. M.; Lohse, M.; Usadel, B. Trimmomatic: A Flexible Trimmer for Illumina Sequence Data. *Bioinforma. Oxf. Engl.* **2014**, *30* (15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170.

(65) Smith, Z. D.; Meissner, A. DNA Methylation: Roles in Mammalian Development. *Nat. Rev. Genet.* **2013**, *14* (3), 204–220. https://doi.org/10.1038/nrg3354.

(66) Jones, P. A. Functions of DNA Methylation: Islands, Start Sites, Gene Bodies and Beyond. *Nat. Rev. Genet.* **2012**, *13* (7), 484–492. https://doi.org/10.1038/nrg3230.

(67) Wu, X.; Zhang, Y. TET-Mediated Active DNA Demethylation: Mechanism, Function and Beyond. *Nat. Rev. Genet.* **2017**, *18* (9), 517–534. https://doi.org/10.1038/nrg.2017.33.

(68) Song, C.-X.; He, C. Potential Functional Roles of DNA Demethylation Intermediates. *Trends Biochem. Sci.* **2013**, *38* (10), 480–484. https://doi.org/10.1016/j.tibs.2013.07.003.

(69) ckorfmann. Snp-Counter, 2021. https://github.com/ckorfmann/snp-counter (accessed 2022-04-05).

(70) Brenner, S.; Lerner, R. A. Encoded Combinatorial Chemistry. *Proc. Natl. Acad. Sci. U. S. A.* **1992**, *89* (12), 5381–5383. https://doi.org/10.1073/pnas.89.12.5381.

(71) Macarron, R.; Banks, M. N.; Bojanic, D.; Burns, D. J.; Cirovic, D. A.; Garyantes, T.; Green, D. V. S.; Hertzberg, R. P.; Janzen, W. P.; Paslay, J. W.; Schopfer, U.; Sittampalam, G. S. Impact of High-

Throughput Screening in Biomedical Research. *Nat. Rev. Drug Discov.* **2011**, *10* (3), 188–195. https://doi.org/10.1038/nrd3368.

(72) Gironda-Martínez, A.; Donckele, E. J.; Samain, F.; Neri, D. DNA-Encoded Chemical Libraries: A Comprehensive Review with Succesful Stories and Future Challenges. *ACS Pharmacol. Transl. Sci.* **2021**, *4* (4), 1265–1279. https://doi.org/10.1021/acsptsci.1c00118.

(73) Yu, Z.; Ku, A. F.; Anglin, J. L.; Sharma, R.; Ucisik, M. N.; Faver, J. C.; Li, F.; Nyshadham, P.; Simmons, N.; Sharma, K. L.; Nagarajan, S.; Riehle, K.; Kaur, G.; Sankaran, B.; Storl-Desmond, M.; Palmer, S. S.; Young, D. W.; Kim, C.; Matzuk, M. M. Discovery and Characterization of Bromodomain 2–Specific Inhibitors of BRDT. *Proc. Natl. Acad. Sci.* **2021**, *118* (9), e2021102118. https://doi.org/10.1073/pnas.2021102118.

(74) Kazmierski, W. M.; Xia, B.; Miller, J.; De la Rosa, M.; Favre, D.; Dunham, R. M.; Washio, Y.; Zhu, Z.; Wang, F.; Mebrahtu, M.; Deng, H.; Basilla, J.; Wang, L.; Evindar, G.; Fan, L.; Olszewski, A.; Prabhu, N.; Davie, C.; Messer, J. A.; Samano, V. DNA-Encoded Library Technology-Based Discovery, Lead Optimization, and Prodrug Strategy toward Structurally Unique Indoleamine 2,3-Dioxygenase-1 (IDO1) Inhibitors. *J. Med. Chem.* **2020**, *63* (7), 3552–3562. https://doi.org/10.1021/acs.jmedchem.9b01799.

(75) Cuozzo, J. W.; Clark, M. A.; Keefe, A. D.; Kohlmann, A.; Mulvihill, M.; Ni, H.; Renzetti, L. M.; Resnicow, D. I.; Ruebsam, F.; Sigel, E. A.; Thomson, H. A.; Wang, C.; Xie, Z.; Zhang, Y. Novel Autotaxin Inhibitor for the Treatment of Idiopathic Pulmonary Fibrosis: A Clinical Candidate Discovered Using DNA-Encoded Chemistry. *J. Med. Chem.* **2020**, *63* (14), 7840–7856. https://doi.org/10.1021/acs.jmedchem.0c00688.

(76) Dawadi, S.; Simmons, N.; Miklossy, G.; Bohren, K. M.; Faver, J. C.; Ucisik, M. N.; Nyshadham, P.; Yu, Z.; Matzuk, M. M. Discovery of Potent Thrombin Inhibitors from a Protease-Focused DNA-Encoded Chemical Library. *Proc. Natl. Acad. Sci.* **2020**, *117* (29), 16782–16789. https://doi.org/10.1073/pnas.2005447117.

(77) Deng, Y.; Peng, J.; Xiong, F.; Song, Y.; Zhou, Y.; Zhang, J.; Lam, F. S.; Xie, C.; Shen, W.; Huang, Y.; Meng, L.; Li, X. Selection of DNA-Encoded Dynamic Chemical Libraries for Direct Inhibitor Discovery. *Angew. Chem. Int. Ed.* **2020**, *59* (35), 14965–14972. https://doi.org/10.1002/anie.202005070.

(78) Li, Y.; De Luca, R.; Cazzamalli, S.; Pretto, F.; Bajic, D.; Scheuermann, J.; Neri, D. Versatile Protein Recognition by the Encoded Display of Multiple Chemical Elements on a Constant Macrocyclic Scaffold. *Nat. Chem.* **2018**, *10* (4), 441–448. https://doi.org/10.1038/s41557-018-0017-8.

(79) Harris, P. A.; Berger, S. B.; Jeong, J. U.; Nagilla, R.; Bandyopadhyay, D.; Campobasso, N.; Capriotti, C. A.; Cox, J. A.; Dare, L.; Dong, X.; Eidam, P. M.; Finger, J. N.; Hoffman, S. J.; Kang, J.; Kasparcova, V.; King, B. W.; Lehr, R.; Lan, Y.; Leister, L. K.; Lich, J. D.; MacDonald, T. T.; Miller, N. A.; Ouellette, M. T.; Pao, C. S.; Rahman, A.; Reilly, M. A.; Rendina, A. R.; Rivera, E. J.; Schaeffer, M. C.; Sehon, C. A.; Singhaus, R. R.; Sun, H. H.; Swift, B. A.; Totoritis, R. D.; Vossenkämper, A.; Ward, P.; Wisnoski, D. D.; Zhang, D.; Marquis, R. W.; Gough, P. J.; Bertin, J. Discovery of a First-in-Class Receptor Interacting Protein 1 (RIP1) Kinase Specific Clinical Candidate (GSK2982772) for the Treatment of Inflammatory Diseases. *J. Med. Chem.* **2017**, *60* (4), 1247–1261. https://doi.org/10.1021/acs.jmedchem.6b01751.

(80) Samain, F.; Ekblad, T.; Mikutis, G.; Zhong, N.; Zimmermann, M.; Nauer, A.; Bajic, D.; Decurtins, W.; Scheuermann, J.; Brown, P. J.; Hall, J.; Gräslund, S.; Schüler, H.; Neri, D.; Franzini, R. M. Tankyrase 1 Inhibitors with Drug-like Properties Identified by Screening a DNA-Encoded Chemical Library. *J. Med. Chem.* **2015**, *58* (12), 5143–5149. https://doi.org/10.1021/acs.jmedchem.5b00432.

(81) Maianti, J. P.; McFedries, A.; Foda, Z. H.; Kleiner, R. E.; Du, X. Q.; Leissring, M. A.; Tang, W.-J.; Charron, M. J.; Seeliger, M. A.; Saghatelian, A.; Liu, D. R. Anti-Diabetic Activity of Insulin-Degrading Enzyme Inhibitors Mediated by Multiple Hormones. *Nature* **2014**, *511* (7507), 94–98. https://doi.org/10.1038/nature13297.

(82) Castan, I. F. S. F.; Graham, J. S.; Salvini, C. L. A.; Stanway-Gordon, H. A.; Waring, M. J. On the Design of Lead-like DNA-Encoded Chemical Libraries. *Bioorg. Med. Chem.* **2021**, *43*, 116273. https://doi.org/10.1016/j.bmc.2021.116273.

(83) Liu, S.; Qi, J.; Lu, W.; Wang, X.; Lu, X. Synthetic Studies toward DNA-Encoded Heterocycles Based on the On-DNA Formation of α,β-Unsaturated Ketones. *Org. Lett.* **2021**, *23* (3), 908–913. https://doi.org/10.1021/acs.orglett.0c04118.

(84) Bao, Y.; Deng, Z.; Feng, J.; Zhu, W.; Li, J.; Wan, J.; Liu, G. A B2(OH)4-Mediated Synthesis of 2-Substituted Indazolone and Its Application in a DNA-Encoded Library. *Org. Lett.* **2020**, *22* (16), 6277–6282. https://doi.org/10.1021/acs.orglett.0c02032.

(85) An, Y.-L.; Li, K.; Shen, Y.; Hong, Z.; Chen, L.; Hu, Y.; Zhou, L.; Wang, D.; Shi, X.; Liu, S.; Su, W.; Cui, W.; Kuai, L.; Yang, H.; Peng, X. DNA Compatible Intermolecular Wittig Olefination for the Construction of α, β-Unsaturated Carbonyl Compounds. *Org. Lett.* **2020**, *22* (10), 3931–3935. https://doi.org/10.1021/acs.orglett.0c01215.

(86) Qu, Y.; Wen, H.; Ge, R.; Xu, Y.; Gao, H.; Shi, X.; Wang, J.; Cui, W.; Su, W.; Yang, H.; Kuai, L.; Satz, A. L.; Peng, X. Copper-Mediated DNA-Compatible One-Pot Click Reactions of Alkynes with Aryl Borates and TMS-N3. *Org. Lett.* **2020**, *22* (11), 4146–4150. https://doi.org/10.1021/acs.orglett.0c01219.

(87) Westphal, M. V.; Hudson, L.; Mason, J. W.; Pradeilles, J. A.; Zécri, F. J.; Briner, K.; Schreiber, S. L. Water-Compatible Cycloadditions of Oligonucleotide-Conjugated Strained Allenes for DNA-Encoded Library Synthesis. *J. Am. Chem. Soc.* **2020**, *142* (17), 7776–7782. https://doi.org/10.1021/jacs.9b13186.

(88) Gerry, C. J.; Yang, Z.; Stasi, M.; Schreiber, S. L. DNA-Compatible [3 + 2] Nitrone–Olefin Cycloaddition Suitable for DEL Syntheses. *Org. Lett.* **2019**, *21* (5), 1325–1330. https://doi.org/10.1021/acs.orglett.9b00017.

(89) Wang, X.; Sun, H.; Liu, J.; Dai, D.; Zhang, M.; Zhou, H.; Zhong, W.; Lu, X. Ruthenium-Promoted C–H Activation Reactions between DNA-Conjugated Acrylamide and Aromatic Acids. *Org. Lett.* **2018**, *20* (16), 4764–4768. https://doi.org/10.1021/acs.orglett.8b01837.

(90) Lu, X.; Roberts, S. E.; Franklin, G. J.; Davie, C. P. On-DNA Pd and Cu Promoted C–N Cross-Coupling Reactions. *MedChemComm* **2017**, *8* (8), 1614–1617. https://doi.org/10.1039/C7MD00289K.

(91) Li, Y.; Gabriele, E.; Samain, F.; Favalli, N.; Sladojevich, F.; Scheuermann, J.; Neri, D. Optimized Reaction Conditions for Amide Bond Formation in DNA-Encoded Combinatorial Libraries. *ACS Comb. Sci.* **2016**, *18* (8), 438–443. https://doi.org/10.1021/acscombsci.6b00058.

(92) Satz, A. L.; Cai, J.; Chen, Y.; Goodnow, R.; Gruber, F.; Kowalczyk, A.; Petersen, A.; Naderi-Oboodi, G.; Orzechowski, L.; Strebel, Q. DNA Compatible Multistep Synthesis and Applications to DNA Encoded Libraries. *Bioconjug. Chem.* **2015**, *26* (8), 1623–1632. https://doi.org/10.1021/acs.bioconjchem.5b00239.

(93) Franzini, R. M.; Samain, F.; Abd Elrahman, M.; Mikutis, G.; Nauer, A.; Zimmermann, M.; Scheuermann, J.; Hall, J.; Neri, D. Systematic Evaluation and Optimization of Modification Reactions of Oligonucleotides with Amines and Carboxylic Acids for the Synthesis of DNA-Encoded Chemical Libraries. *Bioconjug. Chem.* **2014**, *25* (8), 1453–1461. https://doi.org/10.1021/bc500212n.

(94) Ding, Y.; Clark, M. A. Robust Suzuki–Miyaura Cross-Coupling on DNA-Linked Substrates. *ACS Comb. Sci.* **2015**, *17* (1), 1–4. https://doi.org/10.1021/co5001037.

(95) Capaldo, L.; Ravelli, D. Hydrogen Atom Transfer (HAT): A Versatile Strategy for Substrate Activation in Photocatalyzed Organic Synthesis. *Eur. J. Org. Chem.* **2017**, *2017* (15), 2056–2071. https://doi.org/10.1002/ejoc.201601485.

(96) Protti, S.; Fagnoni, M.; Ravelli, D. Photocatalytic C□H Activation by Hydrogen-Atom Transfer in Synthesis. *ChemCatChem* **2015**, *7* (10), 1516–1523. https://doi.org/10.1002/cctc.201500125.

(97) Xiao, W.; Wang, X.; Liu, R.; Wu, J. Quinuclidine and Its Derivatives as Hydrogen-Atom-Transfer Catalysts in Photoinduced Reactions. *Chin. Chem. Lett.* **2021**, *32* (6), 1847–1856. https://doi.org/10.1016/j.cclet.2021.02.009.
(98) Shan, J.; Ling, X.; Liu, J.; Wang, X.; Lu, X. DNA-Encoded CH Functionality via Photoredox-Mediated Hydrogen Atom Transformation Catalysis. *Bioorg. Med. Chem.* **2021**, *42*, 116234. https://doi.org/10.1016/j.bmc.2021.116234.
(99) Koo, B.; Yoo, H.; Choi, H. J.; Kim, M.; Kim, C.; Kim, K. T. Visible Light Photochemical Reactions for Nucleic Acid-Based Technologies. *Molecules* **2021**, *26* (3), 556. https://doi.org/10.3390/molecules26030556.
(100) Chu, L.; Ohta, C.; Zuo, Z.; MacMillan, D. W. C. Carboxylic Acids as A Traceless Activation Group for Conjugate Additions: A Three-Step Synthesis of (±)-Pregabalin. *J. Am. Chem. Soc.* **2014**, *136* (31), 10886–10889. https://doi.org/10.1021/ja505964r.
(101) Kölmel, D. K.; Loach, R. P.; Knauber, T.; Flanagan, M. E. Employing Photoredox Catalysis for DNA-Encoded Chemistry: Decarboxylative Alkylation of α-Amino Acids. *ChemMedChem* **2018**, *13* (20), 2159–2165. https://doi.org/10.1002/cmdc.201800492.
(102) Badir, S. O.; Sim, J.; Billings, K.; Csakai, A.; Zhang, X.; Dong, W.; Molander, G. A. Multifunctional Building Blocks Compatible with Photoredox-Mediated Alkylation for DNA-Encoded Library Synthesis. *Org. Lett.* **2020**, *22* (3), 1046–1051. https://doi.org/10.1021/acs.orglett.9b04568.
(103) Kölmel, D. K.; Ratnayake, A. S.; Flanagan, M. E.; Tsai, M.-H.; Duan, C.; Song, C. Photocatalytic [2 + 2] Cycloaddition in DNA-Encoded Chemistry. *Org. Lett.* **2020**, *22* (8), 2908–2913. https://doi.org/10.1021/acs.orglett.0c00574.
(104) Shan, J.; Ling, X.; Liu, J.; Wang, X.; Lu, X. DNA-Encoded CH Functionality via Photoredox-Mediated Hydrogen Atom Transformation Catalysis. *Bioorg. Med. Chem.* **2021**, *42*, 116234. https://doi.org/10.1016/j.bmc.2021.116234.
(105) Wu, R.; Du, T.; Sun, W.; Shaginian, A.; Gao, S.; Li, J.; Wan, J.; Liu, G. Functionalization of DNA-Tagged Alkenes Enabled by Visible-Light-Induced C–H Activation of N-Aryl Tertiary Amines. *Org. Lett.* **2021**, *23* (9), 3486–3490. https://doi.org/10.1021/acs.orglett.1c00924.
(106) Ruff, Y.; Berst, F. Efficient Copper-Catalyzed Amination of DNA-Conjugated Aryl Iodides under Mild Aqueous Conditions. *MedChemComm* **2018**, *9* (7), 1188–1193. https://doi.org/10.1039/C8MD00185E.
(107) de Pedro Beato, E.; Priego, J.; Gironda-Martínez, A.; González, F.; Benavides, J.; Blas, J.; Martín-Ortega, M. D.; Toledo, M. Á.; Ezquerra, J.; Torrado, A. Mild and Efficient Palladium-Mediated C–N Cross-Coupling Reaction between DNA-Conjugated Aryl Bromides and Aromatic Amines. *ACS Comb. Sci.* **2019**, *21* (2), 69–74. https://doi.org/10.1021/acscombsci.8b00142.
(108) Flood, D. T.; Asai, S.; Zhang, X.; Wang, J.; Yoon, L.; Adams, Z. C.; Dillingham, B. C.; Sanchez, B. B.; Vantourout, J. C.; Flanagan, M. E.; Piotrowski, D. W.; Richardson, P.; Green, S. A.; Shenvi, R. A.; Chen, J. S.; Baran, P. S.; Dawson, P. E. Expanding Reactivity in DNA-Encoded Library Synthesis via Reversible Binding of DNA to an Inert Quaternary Ammonium Support. *J. Am. Chem. Soc.* **2019**, *141* (25), 9998–10006. https://doi.org/10.1021/jacs.9b03774.
(109) Chen, Y.-C.; Faver, J. C.; Ku, A. F.; Miklossy, G.; Riehle, K.; Bohren, K. M.; Ucisik, M. N.; Matzuk, M. M.; Yu, Z.; Simmons, N. C–N Coupling of DNA-Conjugated (Hetero)Aryl Bromides and Chlorides for DNA-Encoded Chemical Library Synthesis. *Bioconjug. Chem.* **2020**, *31* (3), 770–780. https://doi.org/10.1021/acs.bioconjchem.9b00863.
(110) Graham, J. S.; Hunter, J. H.; Waring, M. J. Micellar Buchwald–Hartwig Coupling of Aryl and Heteroarylamines for the Synthesis of DNA-Encoded Libraries. *J. Org. Chem.* **2021**, *86* (23), 17257–17264. https://doi.org/10.1021/acs.joc.1c02325.

(111) Chheda, P. R.; Simmons, N.; Schuman, D. P.; Shi, Z. Palladium-Mediated C–N Coupling of DNA-Conjugated (Hetero)Aryl Halides with Aliphatic and (Hetero)Aromatic Amines. *Org. Lett.* **2022**, *24* (18), 3401–3406. https://doi.org/10.1021/acs.orglett.2c01175.
(112) Zhao, H.; Leonori, D. Minimization of Back-Electron Transfer Enables the Elusive Sp3 C−H Functionalization of Secondary Anilines. *Angew. Chem. Int. Ed.* **2021**, *60* (14), 7669–7674. https://doi.org/10.1002/anie.202100051.
(113) Wu, Z.; Gockel, S. N.; Hull, K. L. Anti-Markovnikov Hydro(Amino)Alkylation of Vinylarenes via Photoredox Catalysis. *Nat. Commun.* **2021**, *12* (1), 5956. https://doi.org/10.1038/s41467-021-26170-6.
(114) Phelan, J. P.; Lang, S. B.; Sim, J.; Berritt, S.; Peat, A. J.; Billings, K.; Fan, L.; Molander, G. A. Open-Air Alkylation Reactions in Photoredox-Catalyzed DNA-Encoded Library Synthesis. *J. Am. Chem. Soc.* **2019**, *141* (8), 3723–3732. https://doi.org/10.1021/jacs.9b00669.
(115) Badir, S. O.; Lipp, A.; Krumb, M.; Cabrera-Afonso, M. J.; Kammer, L. M.; Wu, V. E.; Huang, M.; Csakai, A.; Marcaurelle, L. A.; Molander, G. A. Photoredox-Mediated Hydroalkylation and Hydroarylation of Functionalized Olefins for DNA-Encoded Library Synthesis. *Chem. Sci.* **2021**, *12* (36), 12036–12045. https://doi.org/10.1039/D1SC03191K.
(116) Wen, H.; Ge, R.; Qu, Y.; Sun, J.; Shi, X.; Cui, W.; Yan, H.; Zhang, Q.; An, Y.; Su, W.; Yang, H.; Kuai, L.; Satz, A. L.; Peng, X. Synthesis of 1,2-Amino Alcohols by Photoredox-Mediated Decarboxylative Coupling of α-Amino Acids and DNA-Conjugated Carbonyls. *Org. Lett.* **2020**, *22* (24), 9484–9489. https://doi.org/10.1021/acs.orglett.0c03461.
(117) Kölmel, D. K.; Meng, J.; Tsai, M.-H.; Que, J.; Loach, R. P.; Knauber, T.; Wan, J.; Flanagan, M. E. On-DNA Decarboxylative Arylation: Merging Photoredox with Nickel Catalysis in Water. *ACS Comb. Sci.* **2019**, *21* (8), 588–597. https://doi.org/10.1021/acscombsci.9b00076.
(118) Clark, M. A.; Acharya, R. A.; Arico-Muendel, C. C.; Belyanskaya, S. L.; Benjamin, D. R.; Carlson, N. R.; Centrella, P. A.; Chiu, C. H.; Creaser, S. P.; Cuozzo, J. W.; Davie, C. P.; Ding, Y.; Franklin, G. J.; Franzen, K. D.; Gefter, M. L.; Hale, S. P.; Hansen, N. J. V.; Israel, D. I.; Jiang, J.; Kavarana, M. J.; Kelley, M. S.; Kollmann, C. S.; Li, F.; Lind, K.; Mataruse, S.; Medeiros, P. F.; Messer, J. A.; Myers, P.; O'Keefe, H.; Oliff, M. C.; Rise, C. E.; Satz, A. L.; Skinner, S. R.; Svendsen, J. L.; Tang, L.; van Vloten, K.; Wagner, R. W.; Yao, G.; Zhao, B.; Morgan, B. A. Design, Synthesis and Selection of DNA-Encoded Small-Molecule Libraries. *Nat. Chem. Biol.* **2009**, *5* (9), 647–654. https://doi.org/10.1038/nchembio.211.
(119) Andrade, A. L.; Melich, K.; Whatley, G. G.; Kirk, S. R.; Karpen, J. W. Cyclic Nucleotide-Gated Channel Block by Hydrolysis-Resistant Tetracaine Derivatives. *J. Med. Chem.* **2011**, *54* (13), 4904–4912. https://doi.org/10.1021/jm200495g.
(120) Abdel-Magid, A. F.; Carson, K. G.; Harris, B. D.; Maryanoff, C. A.; Shah, R. D. Reductive Amination of Aldehydes and Ketones with Sodium Triacetoxyborohydride. Studies on Direct and Indirect Reductive Amination Procedures1. *J. Org. Chem.* **1996**, *61* (11), 3849–3862. https://doi.org/10.1021/jo960057x.