Diazo Compounds for the Customization of Important Biomolecules

Inauguraldissertation

zur

Erlangung der Würde eines Doktors der Philosophie

vorgelegt der

Philosophisch-Naturwissenschaftlichen Fakultät

der Universität Basel

von

Na Fei

Aus Rizhao, China

Basel, 2017

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät

auf Antrag von

Prof. Dr. Dennis Gillingham

Prof. Dr. Florian Seebeck

Basel, 24. Mai 2016

Prof. Dr. Jörg Schibler

Dekan

Dedicated to my mom

真者，精誠之至也

《莊子·雜篇·漁父第三十一》

Contents

Summary

The selective modification of native biomolecules such as nucleic acid, signaling molecules, and peptides offers new tools for investigation biological processes. These chemically modified analogs, radiolabeled analogs, fluorescent or affinity tags can be used to label, modify, or pull-down a molecule of interest to probe its function. Among these chemical modifications, diazo compounds are the most versatile building blocks for the customization of diverse biomolecules.

In chapter 1, I discuss briefly the properties of diazo compounds and their application in chemistry and chemical biology. The transition-metal catalyzed XHI reaction of stabilized diazo compounds is used later in chapter 2 and the alkylation reaction of unstabilized diazo compounds with Brønsted acids is used in chapter 3.

In chapter 2, I first give a short introduction of the biological role of CDNs (c-di-nucleotides) which are very important signaling molecules for bacterial growth and human innate immune system stimulation. The rhodium catalyzed XHI reaction with stabilized diazo compounds proved to be a simple method for selective modification of the exo-amine of CDNs. The method tolerates all members of the cyclic dinucleotide family and could be used to modulate their function or introduce useful side-chains such as fluorophores and photo-crosslinking groups. Cu catalyzed NHI for short hairpin oligo and the auto-tandem catalytic process of NHI and Cu-catalyzed azide–alkyne cycloaddition (CuAAC) are also reported here.

In chapter 3, I present the O-alkylation of phosphate groups by alkyl diazo compounds in a range of small molecules and biopolymers. The relatively high pK_a of phosphate in comparison to the other naturally occurring Brønsted acids can be exploited to control alkylation selectivity. I describe here a simple protocol for chemical modification of some of the most important instances of phosphates in natural compounds including small molecule metabolites, nucleic acids, and peptides.

In chapter 4, I talk about the base modifications in gene regulation networks. I was motivated to investigate dynamic RNA methylation as a regulation signal because we had hoped to use some of our chemistry to prove the process. Although this project is still in its infancy, I have already accomplished a convertible nucleotide RNA synthesis and succeed in the expression of FTO protein.

Chapter 1 Introduction

1.1 Diazo chemistry

Diazo compounds are versatile building blocks and used prevalently in synthetic chemistry and chemical biology.[1, 2] Ethyl diazoacetate (EDA) was the first synthetic diazo compound which was preapred by Curtius in 1883.[3] The structure of the diazo group was confirmed to be in a linear form by isotopic labeled experiment (Scheme 1.1).[4] The terminal ^{15}N labeled EDA was obtained from ethyl glycinate diazotisition. After hydrogenation of the labeled diazo compound, the terminal ^{15}N was released in the form of $^{15}NH_3$, the unlabeled nitrogen stayed in the glycine ester. This experiment excludes possibility of the cyclic diazirine isomer.
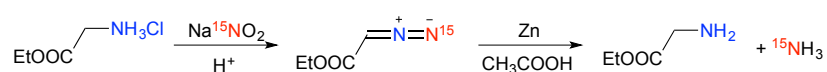
Scheme 1.1 Experimental evidence for the linear structure of ethyl diazoacetate

The general form of diazo compounds, R_2N_2, shows high negative polarity on the carbon atom adjacent to nitrogen. It can be represented as two resonance structures (Scheme 1.2).[5]
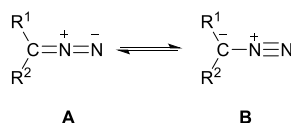
Scheme 1.2 Resonance structures of diazo compounds

The stability of diazo compounds is strongly determined by its neighboring substitutions (Scheme 1.2). Alkyl diazo compounds such as diazoethane and diazopropane are very unstable; their application in organic synthesis is therefore rather limited because they tend to decompose before they are applied in the reaction mixture. Diazo compounds with electron withdrawing groups are more stable since the negative charge can be further delocalized to the substituted group such as aromatic, carbonyl, phosphoryl, and sulfonyl substituents. Thus the stabilized ones are more likely existing in the diazonium form (Scheme 1.1, B). This trend is confirmed by IR measurements where increasing electron-withdrawing leads to a strong N≡N bond.[5]
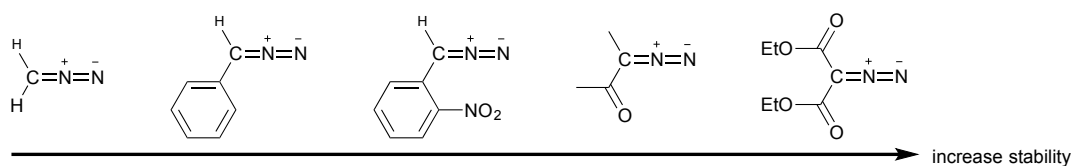
Scheme 1.3 The stability of diazo compounds increase with the electron acceptor substitution groups

Among stabilized diazo compounds, diazoacetates and diazoketones are the most widely applied diazo species in modern organic chemistry due to their easy accessibility. They undergo a wide variety of transformations with high chemoselectivity under mild conditions. [2]
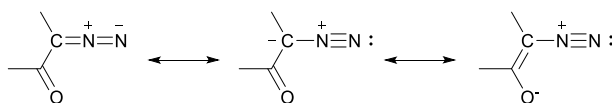
Scheme 1.4 Resonance structures of stabilized α‑diazocarbonyl compounds

The release of nitrogen gas is the driving force for the decomposition of diazo compounds. Acidic conditions favor the thermodynamic stable nitrogen formation by adding a proton to the nucleophilic carbon of the diazo compound. The instability of diazo compounds toward acids also models their reactivity with Lewis acidic transition metal catalysts. The transition-metal-catalyzed transformations of diazo compounds to stabilized carbenoids are ubiquitous in diazo chemistry.[6-9] The metal-carbenoid intermediates are capable of undergoing a range of reactions with high chemo-, regio-, and stereoselectivity such as cyclopropanation of alkenes; C−H activation; X-H (X= heteroatom) bond insertion reactions; and ylide formation.[2, 10-13] Furthermore, selective bioconjugation in proteins and nucleic acids and carbohydrates exploiting diazo compounds has also been demonstrated, offering new tools for application in chemical biology.[2]

The synthesis and application of diazo compounds are well summarized in many reviews.[1, 2] Here I am going to focus the discussion on the transition metal catalyzed α−diazocarbonyl compounds in XHI insertion reaction (X= N, O, S); and unstabilized diazo esterification with acid in chemical biology which are going be employed in chapter 2 and chapter 3.

1.2 Carbenoid formation based on α–diazocarbonyl compounds

Catalyzed X–H insertion reactions into diazo compounds (where X is any heteroatom) are a powerful yet underutilized class of transformations. X can be nitrogen, oxygen, sulfur, selenium, phosphorus, or a halogen.[14] It is a ubiquitous method for carbon-heteroatom bond generation due to the mild reaction conditions and the availability of diverse diazo compounds. However, the successful development of C-H insertion into diazo compounds should not be overlooked.[10]

In these reactions a metal–carbenoid is typically generated in situ from a diazo precursor and then reacts with an X–H bond to deliver, either in concerted or stepwise fashion (scheme 1.5A), to afford the X-H insertion product. Coordination of the negatively polarized carbon of the diazo substrate to the Lewis acidic metal (Scheme 1.5C, step I) is the first productive step in the catalytic cycle. Loss of nitrogen from the resulting intermediate, as shown in step II, then delivers the metal–carbenoid. A few metal–carbenoids have been characterized spectroscopically (Cu, Rh).[15] It is generally accepted that the metal-carbenoids proceed by a concerted mechanism with non-polar bonds (Scheme 1.5B).[12] With polar X-H bond where X bears lone-pair electrons, the reaction most likely undergoes a stepwise mechanism (Scheme 1.5C).
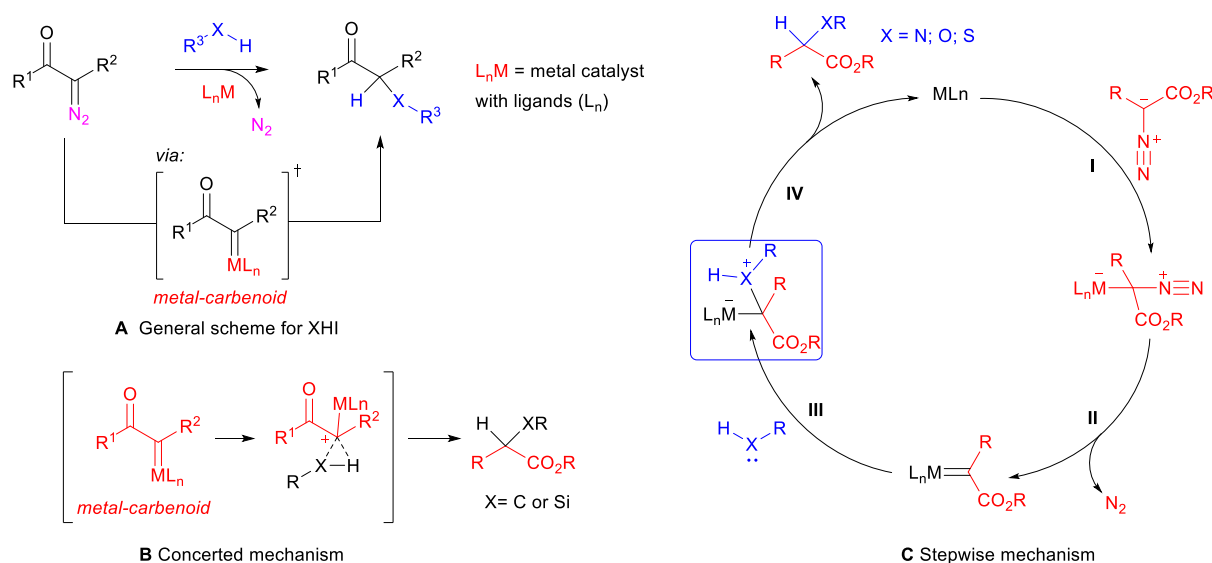
Scheme 1.5 The mechanism of XHI reaction

1.2.1 Catalyst comparison for carbenoid chemistry

$Rh_2(OAc)_4$ was first used by Teyssie in 1973 to decompose EDA to produce a rhodium–carbenoid intermediate, subsequent O–H insertion then delivered ethers.[16] Later they reported that rhodium carboxylates were also efficient catalysts for N–H and S–H insertion. These seminal observations set the stage for forty years of intensive study with rhodium (II) catalysts. As a result the substrate scope has been widely expanded to include aliphatic amine, aniline, amide, alcohols, phenols, thiols and silanes. The discovery of the incredible catalytic abilities of rhodium (II) acetate leads the attention to creating asymmetric catalysts based on rhodium (II). Over the ensuing years, the development of asymmetric catalysts based on rhodium (II) has achieved a great success in enantioselective C–H insertion. However the XHI products were obtained in rather low enantioselectivity.

The above mentioned issues with rhodium have led copper catalysts back to the spot light for high enantioselectivity in XHI. Although first reported to be useful for XHI reactions in the early fifties,[17, 18] XHI chemistry with copper remained underutilized until recently–likely as a result of the harsh reaction conditions, low insertion yields, and sparing solubility of the copper complexes. Research with copper was therefore largely abandoned for rhodium(II) catalysts. In 2002, a report from Pérez and co-workers demonstrated that copper(I) complexes with homoscorpionate ligands could catalyze the insertion of EDA into N–H bonds of amines and amides in high yields under mild conditions.[19] The electronic interaction between the copper and the heterocyclic ligand not only enhanced its stability, but also improved its reactivity and selectivity in the XHI reaction as a result of its unique structure. Especially the development of spiro ligands in highly enantioselective XHI reactions has led copper to a renaissance in XHI applications (Scheme 1.6).
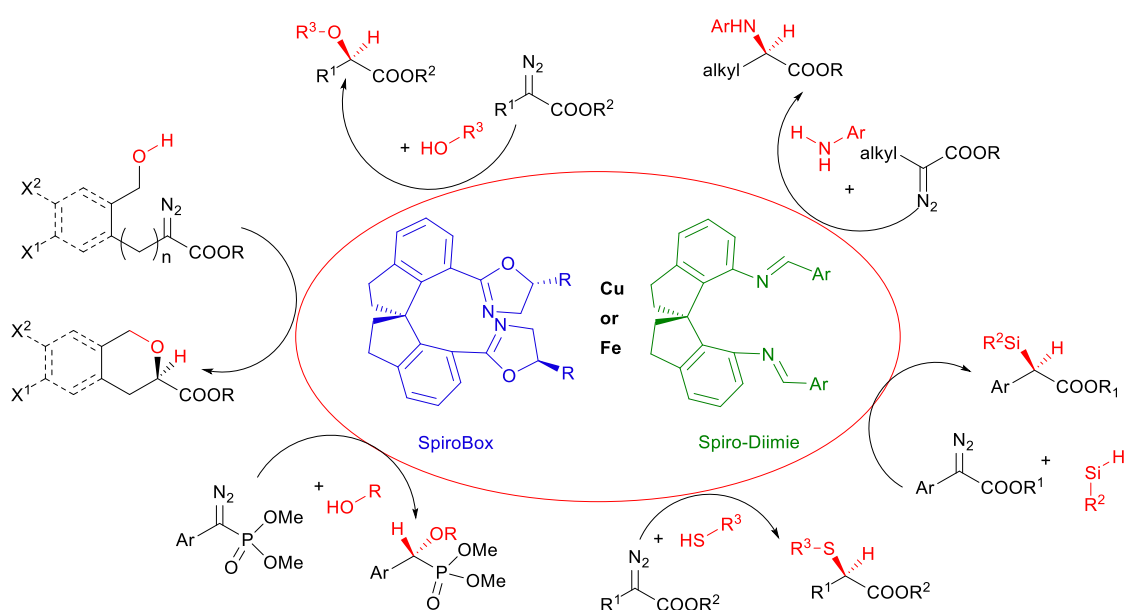
Scheme 1.6 Cu and Fe catalyzed asymmetric XHI by using spiro ligands developed by Zhou

A tenable explanation for the poor ability of rhodium(II) complexes in promoting enantioselectivity in O–H insertion has been put forth by the Yu group based on DFT calculations of both the copper(I) and rhodium(II)-catalyzed reaction paths (Scheme 1.7).[20] They found that a crucial difference in the stereogenesis is that whereas copper complexes prefer to remain bound to the substrate during protonation, the rhodium catalysts favor dissociation leading to a metal-free ylide pathway and hence cannot transfer chiral information. Shown in Scheme 1.7 are the calculated free energies of the proton transfer transition states that led to this mechanistic hypothesis. The copper-associated structure is nearly 10 kcal mol^{-1} lower in energy than the metal-free transition state, which in turn is approximately 6 kcal mol^{-1} lower in energy than the rhodium associated transition state. Even though the free ylide intermediate has not been observed in the rhodium catalyzed NHI process, it could be inferred from a few other examples. Oxonium ylide has been trapped with imine electrophiles.[21] Ally alcohol facilities the tandem ylide intermediate formation during the process of rhodium-catalyzed [2,3]-Sigmatropic Rearrangement.[22] Besides O-H insertion, the spiro-bisoxazoline ligands developed by Zhou strongly suggest Cu(I) catalyzed N-H insertion also proceeds by the metal associated ylide pathway.
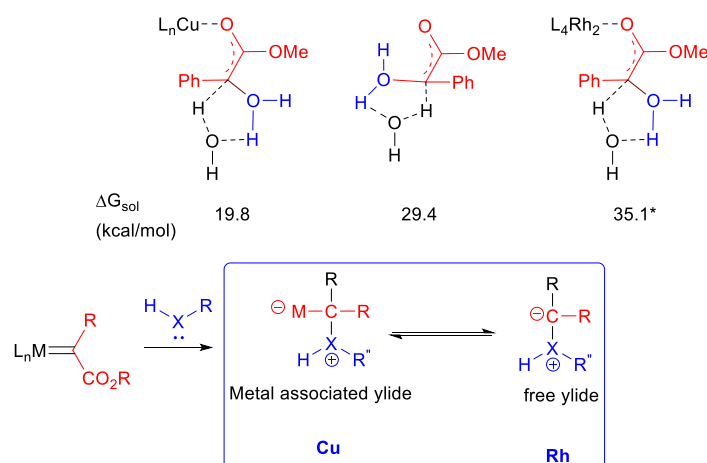
Scheme 1.7 Comparison of Cu and Rh catalyzed XHI

Given the costs associated with the rarer transition metals, processes based on iron would be well-received by potential users. The Woo and Gross groups have independently shown that iron(III)–corrole and iron(III)–porphyrin complexes are excellent catalysts for N–H insertion into a variety of amines and diazo substrates.[23, 24] Furthermore, the recent application of iron–spirobisoxaline complexes in highly efficient enantioselective X–H insertion should stimulate further developments with this practical alternative to the precious metal catalysts.[9]

Despite the preference for the metal-free ylide pathway in XHI, enantioselective processes based on rhodium(II) may yet be realized. A recent report demonstrating that an achiral rhodium(II) carbene can be intercepted by an imine activated with a chiral Brønsted acid indicates that cooperative rhodium(II)–

Brønsted acid catalysis may offer an alternative approach towards enantioselective XHI.[25] Indeed Saito and coworkers have found that cinchona alkaloid additives can deliver enantiomeric excesses of up to 50% in the O–H insertion of α-diazoesters with water catalyzed by achiral rhodium(II) catalysts.[26] The Zhou group has also recently discovered that BINOL-based chiral phosphoric acids (CPA) deliver high levels of enantioinduction in N–H insertion reactions (Scheme 1.8)[27]. These promising initial findings suggest that XHI coupled with enantioselective protonation is an area poised for further development.
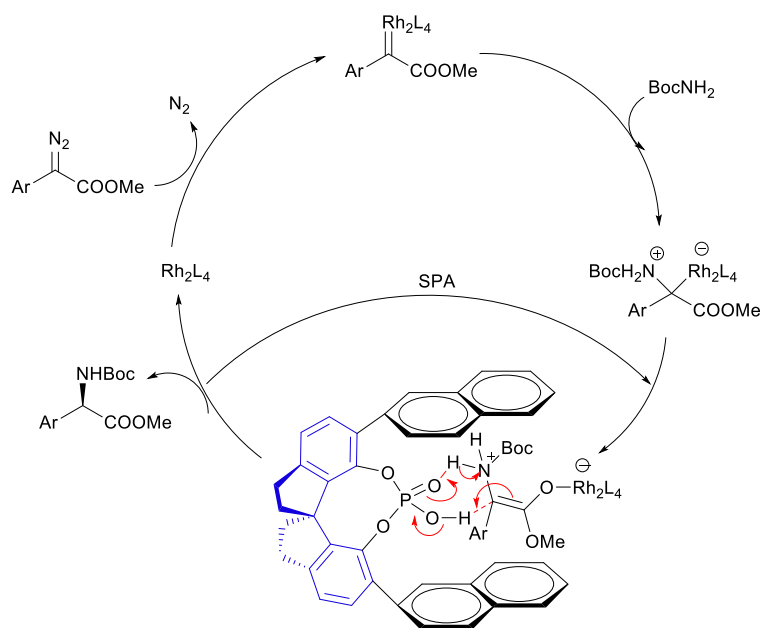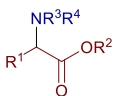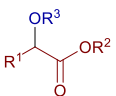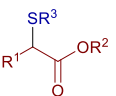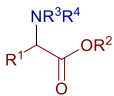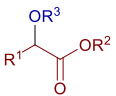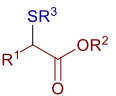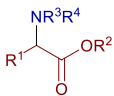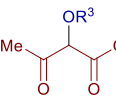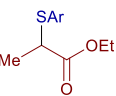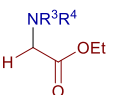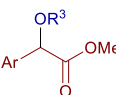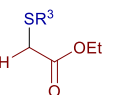
Scheme 1.8 CAP ligand mediated rhodium catalyzed asymmetric NHI

Ruthenium, one of rhodium's direct neighbours in the periodic table, was first introduced to catalyze EDA insertion into S–H and N–H bonds in 1997.[28] Recent results on ruthenium catalysed N–H[29] and O–H insertion[30] reactions, reported by the Che and Lacour groups respectively, demonstrate that ruthenium complexes can sometimes offer unique reactivity in comparison with other catalysts. Although a relative newcomer in XHI, the favourable properties of ruthenium (its similar reactivity to rhodium, lower cost, more available oxidation states, and rich coordination chemistry) suggest a bright future.

Over the past two decades the catalysis field has proven in countless cases that the properties of templating and turn-over offered by a catalyst can deliver benefits that no stoichiometric process can. Tuning catalyst structure can offer the prospect of controlling chemo-, diastereo-, enantio-, regio-, and site-selectivity. An overview was summarized here in tabular format of the breadth of substrates and reaction types each catalyst has so far been applied to (see Table 1.1). The rhodium, copper, ruthenium, and iron catalyzed versions of this reaction are tolerant to a range of reaction conditions

and functional groups, and can accommodate a variety of heteroatom donors. The assessment of their strengths and weaknesses are also included here for comparison.

Table 1.1 Comparison of metals for XHI reactions

metal	N-H insertion	O-H insertion	S-H insertion	strengths	weaknesses
Rh	R^3R^4NH = amides, amines, anilines R^1 = aryl, Me, H R^2 = Me, Et, Bn	R^3OH = phenols, alcohols, water R^1 = aryl, Me, H R^2 = Me, Et, Bn	R^3SH = aliphatic and aromatic thiols R^1 = aryl, Me, H R^2 = Me, Et, Bn	• High turn-over numbers and frequencies • Broad substrate scope • Biomolecule modification	• Expensive • Competing C-H insertio and β-elimination • Moderate stereocontrol
Cu	R^3R^4NH = amides, amines, anilines R^1 = aryl, Me, H R^2 = Me, Et, Bn	R^3OH = phenols, alcohols, water R^1 = aryl, Me, H R^2 = Me, Et, Bn	R^3SH = aliphatic, aromatic, and benzylic thiols R^1 = aryl, Me, H R^2 = Me, Et, Bn	• Cheap • Chemoselective • Enatioselective variant available • Reactivity highly sensitive to ligands	• Susceptible to inhibition by Lewis bases
Ru[a]	R^3R^4NH = amides, amines, anilines R^1 = aryl, Me, H R^2 = Me, Et, Bn	R^3OH = phenols, alcohols		• Chemoselective • Rich coordination chemistry	• Expensive • Narrow substrate scope[a]
Fe[a]	R^3R^4NH = amides, amines, anilines	R^3OH = phenols and alcohols	R^3SH = aliphatic, aromatic, and benzylic thiols	• Cheap • Low toxity • Reactivity highly sensitive to ligands • Enantioselective variant available	• Narrow substrate scope

[a] the narrow substrate range with these metals may simply be a reflection of lack of development

1.2.2 Modifications in chemical biology with α-diazocarbonyl compounds based on metal-carbenoid XHI

Nucleophilic heteroatoms are highly abundant in nucleic acid, protein, carbohydrates and signaling molecules. Selective modification of the native biomolecular structures by targeting the nucleophilic heteroatom could gain great benefit from XHI reactions. Given the right catalysts and ligands it should be possible for chemists to imitate the kind of selectivity achieved by enzymes, but progress in creating artificial catalysts to selectively target intact biomolecules has thus far been modest. Nevertheless, much of the progress has been achieved to date with catalysts for XHI.

Examination of protein active sites by diazo modification is nearly as old as the discovery of copper-carbenoid chemistry. Rajagopalan and coworkers[31] as well as Delpierre and Fruton[32] discovered almost simultaneously that the active-site carboxyl residue of pepsin is modified with diazo compounds at least 103 fold faster in the presence of large excesses of copper(II) salts. Although the mechanistic underpinnings of the reaction were unknown at the time, such reactivity with carboxylates is reasonable considering that XHI would prefer the protonated form of the acid. These authors also observed a short induction period before protein modification began, consistent with the necessity to generate copper(I) from a sacrificial diazo-coupling before effective catalysis can ensue.[33]
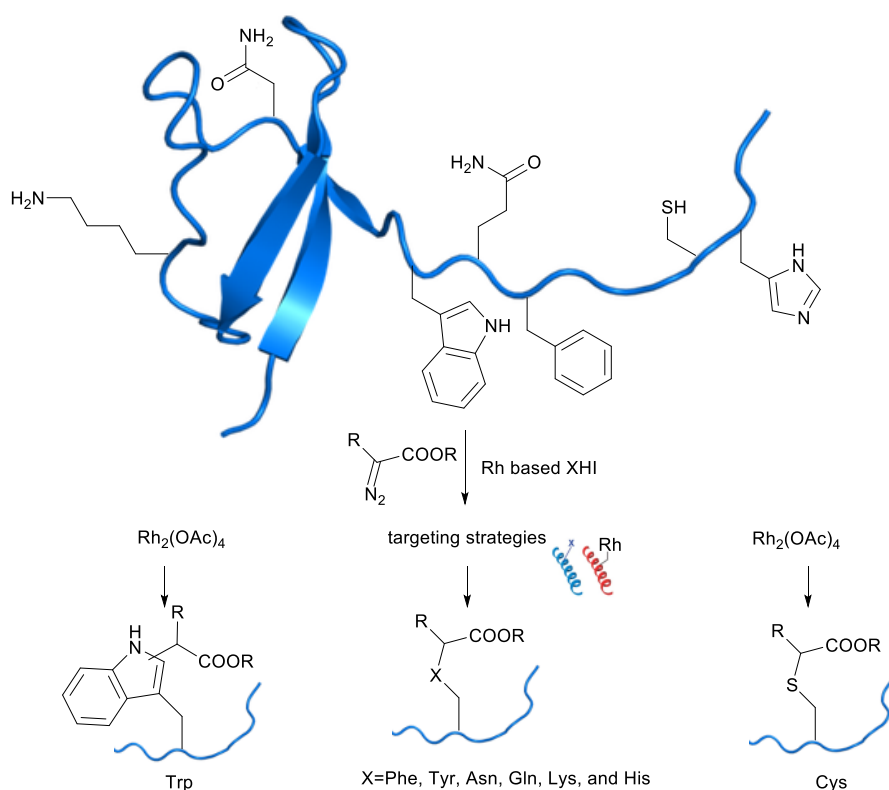
Scheme 1.9 XHI in protein labeling

It was many years later, before the potential of rhodium(II) complexes to promote XHI reactions for applications in chemical biology was recognized. Antos and Francis reported in 2004 that tryptophans in myoglobin and subtilisin could be modified at low pH using α-diazo esters and dirhodium tetraacetate (see Scheme 1.9).[34] The reaction seems to proceed by a mixture of XHI and cyclopropanation of the tryptophan indole moiety. Their refined method using *tert*-butyl hydroxylamine hydrochloride as the buffer condition leads to smooth alkylation of tryptophan residues on a variety of proteins. For the metal-mediated protein labelling to proceed efficiently its nucleophilic residues must be exposed to the carbenoid intermediate. Surface-exposed cysteine residues in cystic fibrosis transmembrane conductance regulator-associated ligand (CAL) and its PDZ domain CALP react selectively with diazo reagents via S-H insertion using rhodium acetate as the catalyst.[35] The selectivity of cysteine over tryptophan is remarkably high after screening a few other proteins. This could be explained that tryptophan residues are buried in the hydrophobic pocket thus they are not solvent accessible. Besides tryptophan and cysteine, it would be ideal to be able to modify nearly any desired amino acid side-chain with the site-selectivity controlled by the catalyst. Targeting strategies to deliver a rhodium–carbenoid intermediate to specific amino acids were achieved by the combination of peptide molecular recognition and residue-selective dirhodium catalysis. XHI modification of nucleophilic residues such as Phe, Tyr, Asn, Gln, Lys was achieved through applying the well understood binding rules of helix-interaction, where the helix carried a rhodium catalyst.[36]
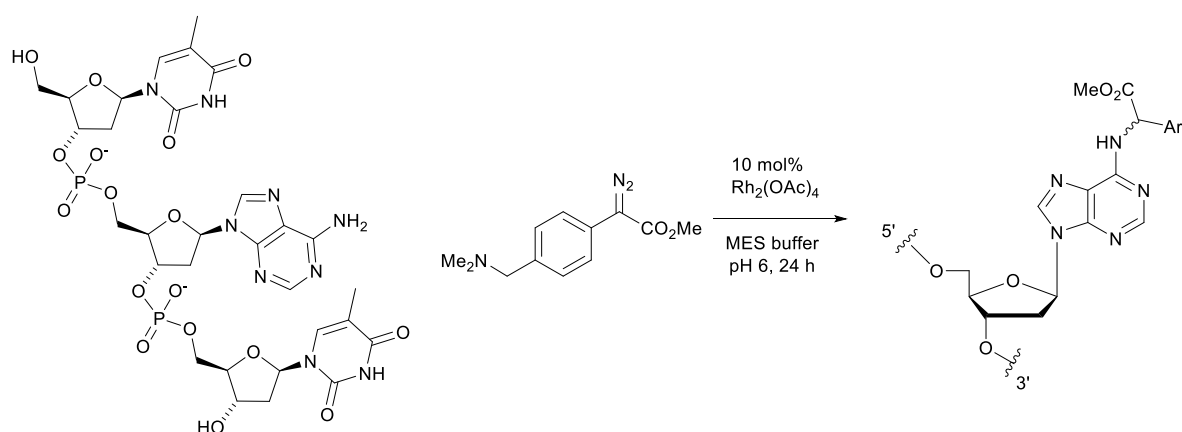
Scheme 1.10 Rh catalyzed alkylation of nucleic acid

While customized nucleic acids could be obtained by oligonucleotide solid support synthesis, comprehensive laborious work is needed. Our group recently reported that rhodium-carbenoids can target the exocyclic N-H group of various native nucleic acids. The $Rh_2(OAc)_4$ reacts with the diazo compound to create a rhodium carbenoid which can further react with nucleophilic groups on nucleic acids (Scheme 1.10).[37] Although a number of reaction pathways of rhodium carbenoids with nucleic

acids are conceivable, only N–H insertions into exocyclic amines of nucleobases were observed; thymidines, uridines, and the ribophosphate backbone were unreactive. The site and mode of reactivity in the alkylation was precisely determined through a combination of NMR studies with a ^{13}C-labeled diazo substrate, as well as tandem MS experiments. In contrast studies on hairpin sequences revealed that double-stranded stretches were unreactive. It seems that if the N-H bonds are engaged in Watson-Crick base- pairing they are unavailable for reaction with the rhodium.

The previous development of diazo compounds based modification of protein and nucleic acid greatly helps the further studies and substrate and reaction scope. The recent developments applying XHI to the modification of intact proteins and nucleic acids certainly give us hints for further application in this field. We could apply these theories and methods to expand the application in many other interesting biomolecules which are shown in the chapter 2.

1.3 Alkylation with unstabilized diazo compounds in chemical biology

Alkyl diazo compounds such as diazomethane, diazoethane, diazopropane are explosive compounds and the difficulty of preparation limited their use in organic synthesis. For example, diazomethane is an extremely sensitive yellow gas. It might explode even passing over sharp surfaces. Its permissible exposure limit is 0.2 ppm due to its high toxicity. Diazomethane generated in situ from N-nitroso-derivatives by base-promoted decomposition can be managed at a safe level.[38] Recently, the developments of flow systems combined with micro-reactor technology to generate diazomethane in situ inside the reactor, greatly diminishing the chances of accidents.[38, 39] On the other hand, trimethylsilyl diazomethane is more thermally stable and less explosive, used frequently in organic synthesis as a safer, commercially available alternative to diazomethane. However, precaution to its toxicity should be raised after two fatal incidents.[40]

Despite these safety concerns, alkyl diazo compounds are very useful precursors in synthetic chemistry. They undergo carboxylic acids alkylation,[41] homologation of carbonyl compounds,[42] cyclopropanation of alkenes,[43] 3-dipolar cycloadditions[44, 45] to name a few.

In aqueous condition, the alkyl diazo compounds are extreme acid labile (Scheme 1.11). Brønsted acids catalyze reactions by protonation of the partially negatively charged carbon of diazo compounds and forming the unstable diazocarbonoid to be attacked by nucleophiles. The equilibrium favoring the carbocation (and thus the rate of subsequent reactions) is determined by the strength of the Brønsted acid and the stability of the carbocation produced (Scheme 1.11). Carbocations generated from alkyl dizao compound are very unstable, immediately get attacked by water leading to hydrolysis. The extreme acid sensitivity and the lack of selectivity limit alkyl diazo compounds' utility with biomolecule in aqueous solution.
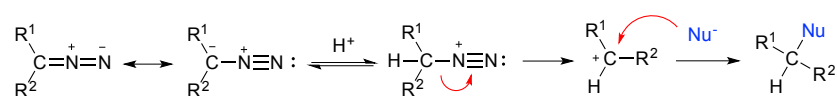
Scheme 1.11 The acid catalyzed decomposition of diazo compounds.

In stabilized diazo compounds, the carbon adjacent to the nitrogen group is weakly basic. The pronation is much slower than the carbon in alkyl diazo compounds. For example, diazomalonates have very long life time in water. However, the stability renders them of little value in Brønsted acid alkylation. The strength and nature of Brønsted acids also play a determinant role in the fate of the protonated diazo compound, which can form a tight ion pair with the carbocation expelled from the solvent, yielding esterification products. Half stabilized diazo compounds such as TMS, aryl substituted

diazo methanes are in the middle ground between these two extremes which is essential for controlling selectivity in diazo-type esterifications (Scheme 1.12).
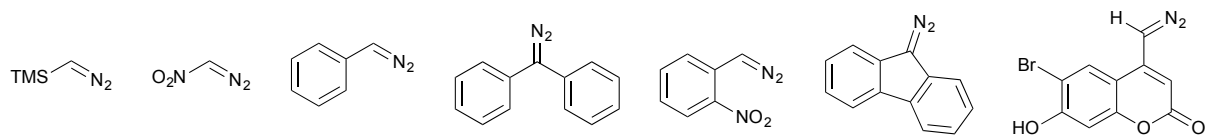
Scheme 1.12 Selected half stabilized diazo compounds used as alkylating reagents

Alkyl diazo compounds are potent alkylating reagents. They modify DNA at many different sites, thereby producing lethal and mutagenic lesions (Scheme 1.13). The alkylation only occurs at different sites of bases, including exocyclic oxygens and most ring nitrogens, but also happens to phosphates of the DNA backbone, thereby generating phosphotriesters (Scheme 1.13 lower, modification site is corresponding to the nucleophilicity of the nucleic acids).[46] Diazo precursors, such as N-nitroso-derivatives or triazene are often used in the cancer chemotherapy to alkylate DNA in the cancer cell inhibiting DNA replication and leading to cell death (Scheme 1.13 upper).[46]
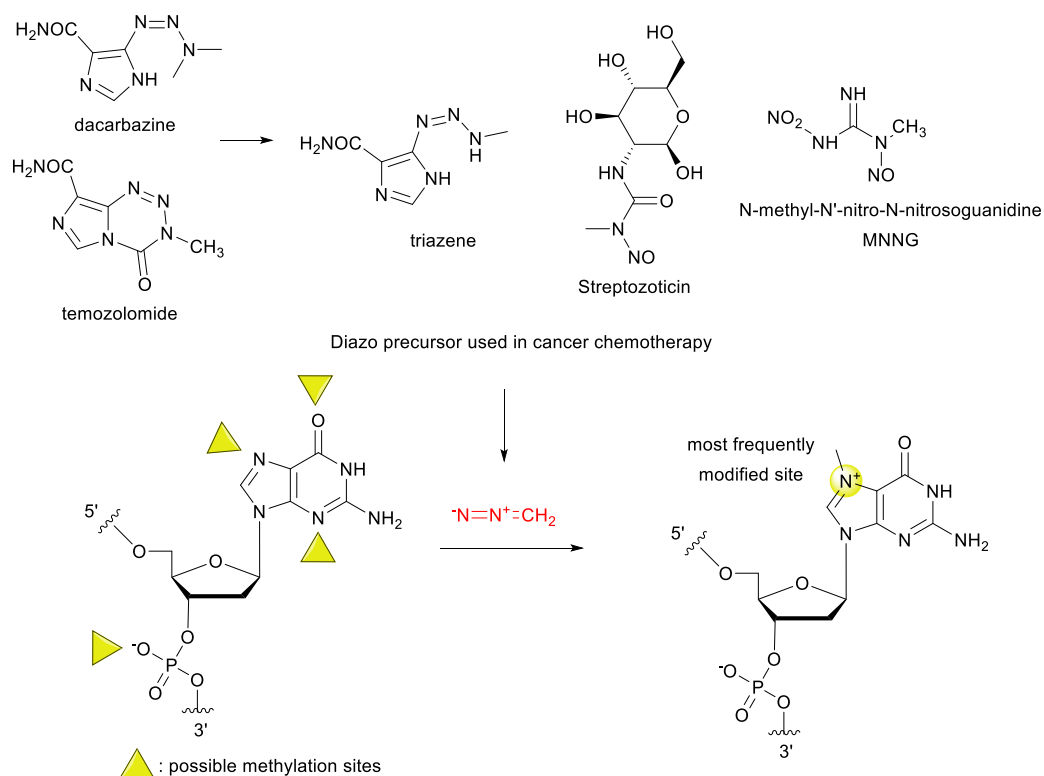
Scheme 1.13 Diazo precursor used in cancer chemotherapy

Alkylation of the phosphate group of nucleic acids with aryl diazo compounds has a long history. Dibenzyl uridine 3'-mono-phosphate has been prepared in anhydrous MeOH using corresponding diazoalkanes in 1955 (Scheme 1.14).[47] The phosphotriester was decomposed in aqueous solution to a

mixture of phosphodiester. It is considered that the occurrence of phosphotriester group in ribonucleic acids is unlikely.
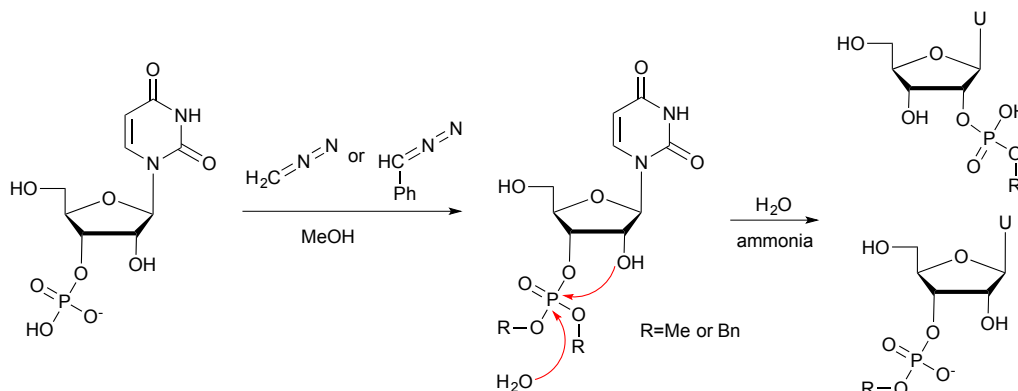
Scheme 1.14 Alkylation of UMP to phosphotriester and its stability in aqueous solution

Later, an expanded series of aryl diazo compounds were investigated in alkylation of the phosphate group of nucleic acids.[48] Typically, excess amounts of diazo compounds are employed in anhydrous solvents to achieve phosphotriesters. Among them, coumarine diazo[49] and 2-nitro-phenyl diazo methane[50, 51] modified nucleic acids are capable to release the native nucleic acid by photolysis, thus these are widely used as nucleic acid caging agents. 6-Bromo-4-diazomethyl-7-hydroxycoumarin (Bhc-diazo) has been employed to react with the phosphate moiety of the backbone of RNA, modifying approximately 30 sites on the phosphate moieties per 1 kb of RNA sequence.[52] The Bhc-caged green fluorescent protein (GFP) mRNA has severely reduced translational activity in vitro, whereas illumination of Bhc-caged mRNA with ultraviolet light leads to partial recovery of translational activity. In embryos injected with Bhc-caged GFP mRNA at the one-cell stage, GFP protein expression and fluorescence is specifically induced by ultraviolet light. However, Friedman's lab demonstrated that the diazo compounds react very poorly with backbone phosphates. Instead they show remarkable specificity for terminal phosphates and very modest modification of nucleobase.[53, 54]
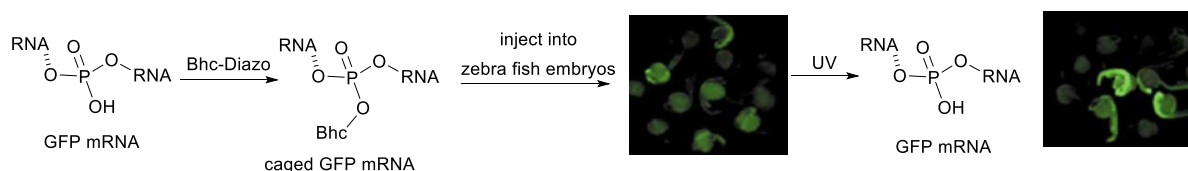
Scheme 1.15 Inactivation of mRNA by caging with Bhc-diazo and reactivation by photo-illumination.

While the alkyl diazo compounds are too reactive and difficult to handle, their application in aqueous buffer is unselective. Half stabilized diazo compounds which are between too reactive alkyl diazo and stabilized carbonyl diazo moieties, are employed in frequently in alkylation of carboxylic acids and phosphate esters.

Chapter 2 Catalytic NHI of nucleic acids by diazo compounds

2.1 Cyclic-di-nucleotides as important signaling molecules

The 3', 5' linked cyclic dinucleotides (CDNs, Figure 2.1) play diverse and important roles as bacterial second messengers in response to environmental changes. Bis-(3'-5')-cyclic dimeric guanosine monophosphate (c-di-GMP) is a ubiquitous signaling molecule in bacteria.[55] Fluctuating levels of c-di-GMP controls their lifestyle, from a swarming motile phase to sessile biofilm formation, which is an essential step for virulence and pathogenesis.[56] A c-di-GMP analog, c-di-AMP, is involved in the regulation of bacterial cell wall metabolism, osmotic stress responses and in pathogen-host interaction.[57-59] C-di-GAMP as is the newest member of second messengers, its biological role is still unknown.[60] Understanding how bacteria thrive and spread their population is important for environmental protection and human health. Once mammalian cells have been infected by bacteria, these exogenous 3', 5'-linked CDNs will be recognized by STING of the host's innate immune system, thus CDNs become the host's immunity alarmins.[61] STING (the name deriving from stimulator of interferon genes), is a transmembrane protein in mammalian cell that acts as an innate immune sensor of CDNs. STING promotes the induction of type I interferons (IFN-α and IFN-β).[62-64] Stimulation of host immune response with CDNs could be an alternative approach in combating antibacterial infection. Such alternative approaches are becoming more valuable due to the increase of antibiotics resistant in bacterial pathogens. The large scale overuse of antibiotics in livestock farming and fish farming becomes a big issue and alternative strategies are in urgent need instead of developing stronger antibiotics.

Until recently CDNs were not thought to be produced by mammalian cells and should appear in the cytosol only because of the presence of invading bacteria.[65] Only very recently was c-di-GAMP discovered in mammalian cell in an unusual isomeric form of previous CDNs.[66, 67] This unique secondary messenger c-G(2',5')pA(3',5')p is produced endogenously, as the result of an mammalian innate immune response to pathogen-derived cytoplasmic double stranded DNA (dsDNA).[67, 68] The c-G(2',5')pA(3',5')p synthase (cGAS) has been shown to be an innate immune sensor of retroviruses including HIV and Herpes virus.[68] All 3', 5'-and 2'-3'-linked CDNs are high-affinity ligands for STING, thereby ensuring successful host defenses against pathogen infection.[69]
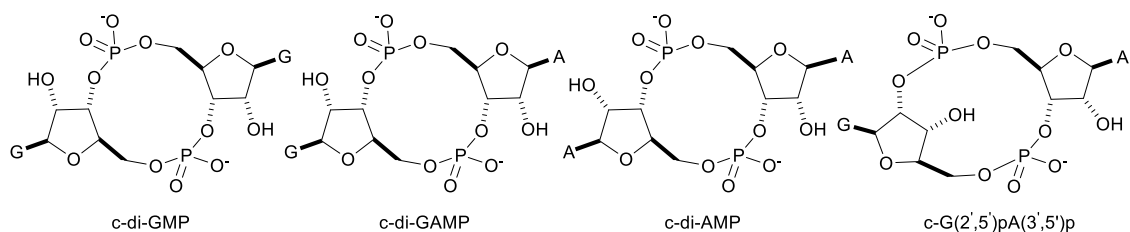
Figure 2.1 Structures of the natural CDNs

2.1.1 c-di-GMP as a ubiquitous second messenger

C-di-GMP was first discovered in 1987 by Benziman's group[70] as a cofactor needed to activate cellulose biosynthesis in *Gluconacetobacter xylinus*. The level of c-di-GMP is controlled by specific diguanylate cyclases (DGCs) and phosphodiesterases (PDEs). DGC enzymes contain a GGDEF amino acid sequence motif in the active domains and use two GTP molecules to synthesize one c-di-GMP. While PDEs contain either an EAL or an HD-GYP amino acid motif, and they degrade c-di-GMP to GMP. The cyclases and phosphodiesterases regulate the c-di-GMP at the appropriate level according to environmental changes (Figure 2.2).[56, 71, 72] C-di-GMP is a prevalent signaling second messenger in a wide variety of bacteria, but not found in archaea yet. It has only been observed in eukaryotes in Dictyostelium very recently.[73] Increasing levels of c-di-GMP lead to biofilm formation by enhancing the synthesis of adhesive structures and biofilm matrix components. While decreasing levels of c-di-GMP stimulate dispersal of biofilms and promotes a planktonic lifestyle. Fluctuating levels of c-di-GMP initiate adaptive responses in bacterial cells when they are exposed to various sources of stress, including antibiotics, nutrient limitation, anaerobiosis, heat shock. The c-di-GMP signaling is not only at the transcriptional level by binding specific regulator protein or riboswitches, but also at a post-translational level via PilZ domain containing proteins. C-di-GMP binds to its effector components and receptors regulating a wide range of bacteria cell functions including mobility, adhesion, cell to cell communication, quorum sensing, exopolysaccharide synthesis, biofilm formation and virulence.[55, 56, 72]
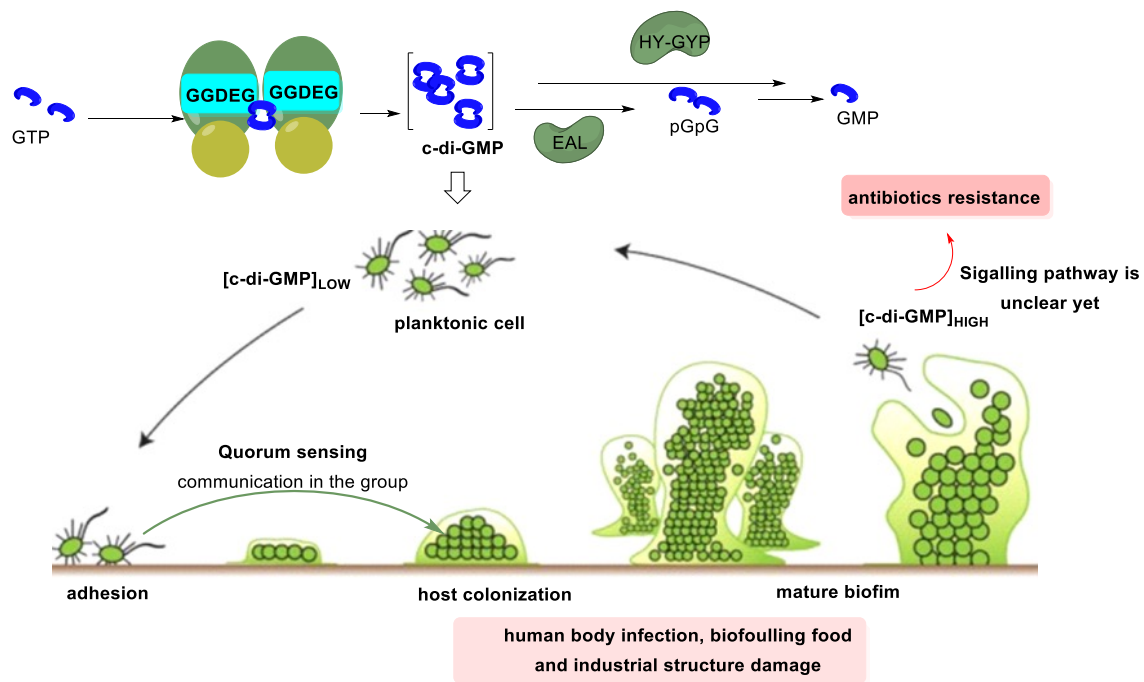
Figure 2.2 Biosynthesis of c-di-GMP and its role in regulating cell responses to environment changes

The c-di-GMP mediated cell to cell communication is integrated with quorum sensing helping bacteria in the host cell colonization, against cellular immunity attack and antimicrobial treatments.[74] An association between high levels of c-di-GMP and antibiotic resistance may be expected because c-di-GMP regulates biofilm formation. The highly structured surface of the biofilm as the physical barrier of bacteria is extraordinary resistant to antimicrobial agents.[75, 76] However, a clear understanding of this correlation has not been established. Chua showed that P. aeruginosa cells with low c-di-GMP levels were found to be more resistant to colistin than P. aeruginosa cells with high c-di-GMP levels.[77] At the same time, Sauer's group demonstrated elevated level of c-di-GMP in the planktonic cell of Pseudomonas aeruginosa benefits the brIR expression (brIR is a c-di-GMP responsive transcription regulator). The downstream response will increase the survival rate against antimicrobial agents.[78, 79] Nevertheless, understanding the increased antibiotics tolerance is very important for bacterial infection treatment in long term. Especially now, the increased drug tolerance turns some biofilm infections in human into a chronic problem.[80, 81] Interfering with the c-di-GMP signaling pathway has a great potential application in new therapeutics that specifically target biofilms.

The formation of biofilms not only significantly harms human health, but also damage a large amount of industrial structures, especially in aquatic environments. In contaminated marine engineering systems, tremendous efforts are needed to remove biofilm colonization. The cleaning combined with antibiotics is losing its efficiency in treating industrial biofilm damage due to the increased antibiotic resistance. Manipulation of bacteria from biofilm lifestyle to dispersed motile phase by controlling c-

di-GMP level is a new trend in research which could lead to a new strategy in the treatment of biofilm formation in industry.[82]

2.1.2 c-di-AMP as a newly discovered CDN draws attention

Cyclic di-adenosine monophosphate (c-di-AMP) was discovered in the DNA damage checkpoint initiation system in 2008.[83] In nutrition limited environments, the soil bacteria *Bacillus subtilis* will go through a sporulation process in which the cell will divide unequally to produce a spore cell which is more resistant to the unfavorable environmental conditions. The spore then regains its viability when conditions become favorable. During the sporulation process, if a cell is introduced with unexpected DNA double-strand break[83] or DNA-damaging agents[84], the DNA integrity scanning protein (DisA) will bind tightly to the unexpected DNA damage, initiating a DNA repair response. As a result, decreased levels of c-di-AMP are observed, and the sporulation also slows down (Figure 2.3). In the absence of DNA damage, the level of c-di-AMP is restored and the sporulation is continued. DisA forms an octameric complex (Figure 2.3), one tetramer is arranged in a parallel, interacting head to head with another DisA tetramer according to its structural and biochemical analyses.[83] Each monomer consists of three distinct domains. The HhH domain is responsible for checking DNA damage, while the DAC (diadenylyl cyclase) domain is responsible for c-di-AMP synthesis. DisA slides smoothly on intact DNA. Its activity is unaffected, thereby producing c-di-AMP and later initiating sporulation.
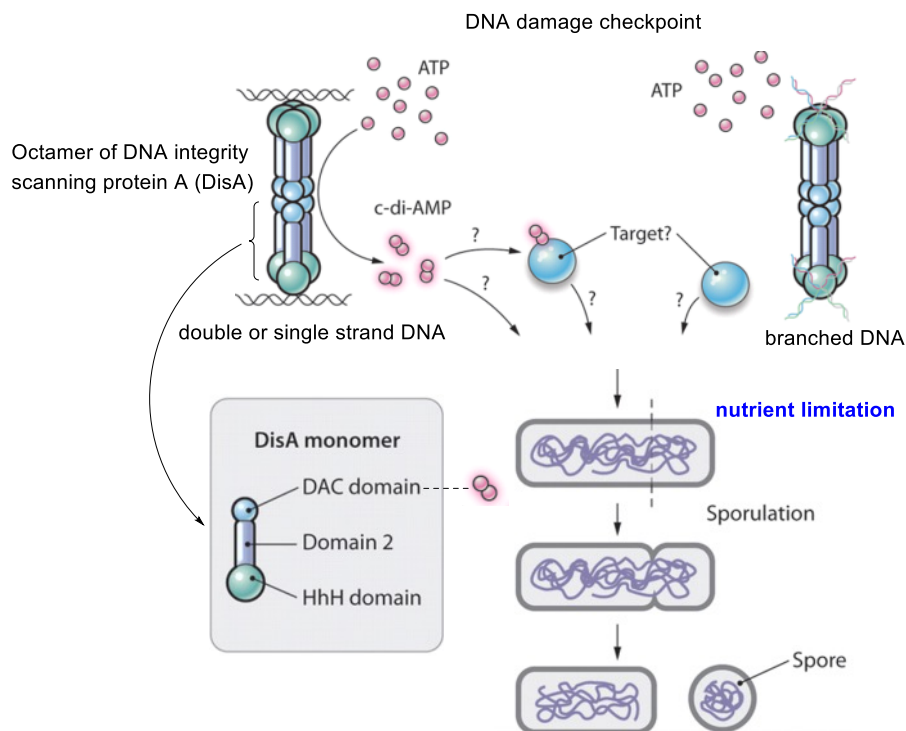
Figure 2.3 Biosynthesis of c-di-AMP and its role in cell sporulation

20

C-di-AMP is likely to be widely distributed among bacteria and is also found in a subset of archaea[85]. Like c-di-GMP, c-di-AMP is synthesized from ATP by DAC enzymes and is hydrolyzed by phosphodiesterase (PDE) enzymes. Recently, several c-di-AMP receptors were identified to be involved in the regulation of bacterial cell wall metabolism,[86] osmotic stress responses[87]. However the understanding of how these signals are transduced is still limited.

2.1.3 CDN signaling in eukaryotes domain
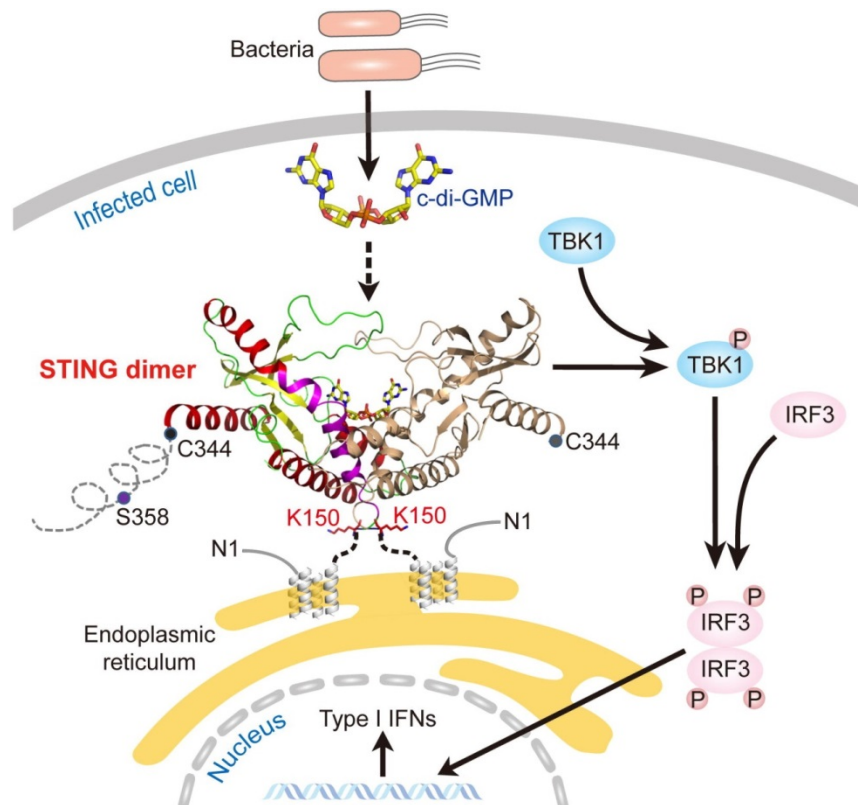
Figure 2.4 STING is a direct innate immune sensor of cyclic di-GMP. [Reprinted with permission from Elsevier Publishers Ltd: [Immunity] (S. Ouyang, X. Song, Y. Wang, H. Ru, N. Shaw, Y. Jiang, F. Niu, Y. Zhu, W. Qiu, K. Parvatiyar, Y. Li, R. Zhang, G. Cheng and Z.-J. Liu, Immunity, 2012, 36, 1073-1086), copyright (2012)].

The innate immune system, also known as in-born immunity system, recognizes and responds to pathogens in a generic way, defending the host from infection. STING has proved to be a key player in recognition of pathogens, triggering type I interferons (IFNs) protection, especially critical in detecting foreign nucleic acids like cytosolic DNA and CDNs.[62, 63, 88] X ray structures showed that a V shaped dimer of STING's C-terminal domain binds to a V-shaped c-di-GMP monomer at the interface of the STING dimer (Figure 2.4). Isothermal titration calorimetry showed c-di-GMP binds to STING with high affinity (K_d 2.5–5 μM).[63] Additionally, Parvatiyar et. al proved that another protein DDX41 has greater binding affinity for CDNs (Figure 2.6). It was hypothesized that DDX41 was the main sensor of c-di-GMP and c-di-AMP and operated upstream of STING.[89] STING's recognition of CDNs as the danger signal and stimulation of immune response against bacterial infection suggested that CDNs could be used as

innate immune agonists. Encouragingly, in the preclinical development, c-di-GMP proved to be a potential vaccine adjuvant against mucosal infections.[90]

Very recently, STING activation was found to be triggered by a new type of CDN, c-GAMP, the first known CDN produced in animals.[66] When the pathogen enters the host cell, its dsDNA will be detected by c-GAMP synthase (cGAS) in the cytosol. Upon binding dsDNA cGAS produces c-GAMP which then binds to STING, turning on a downstream immune response.[68] The structure of c-GAMP was presumed to be similar to previous CDNs by Wu,[66] but soon it was revised to be c-G(2',5')pA(3',5')p by Gao.[67] The 2'-5' phosphodiester linkage is very unusual. Very few nucleases are known to be able to hydrolyze such a linkage. The exact structure was carefully co-examined by X-ray, HPLC and NMR analysis (Figure 2.5). In the NMR spectrum of c[G(2'-5')pA(3'-5')p], the 1' position proton of guanosine (the proton in blue, Figure 2.5C) is a doublet ($^3J_{HH}$ = 9 Hz), but 1' position proton of adenosine a singlet (the proton in red, Figure 2.5C) when the phosphate ester is attached to the 3' position. Combined with X-ray analysis, the structure of this unusual CDN was confirmed. This unusual 2'-5' linkage has greater affinity and stronger activation towards human STING than exogenous 3'-5' linkage CDNs, thus strongly boosting antigen-specific T cell activation and antibody production. This antiviral property of this non-canonical CDN attracts great interest as a drug adjuvant in treating HIV and Herpes virus.[68]
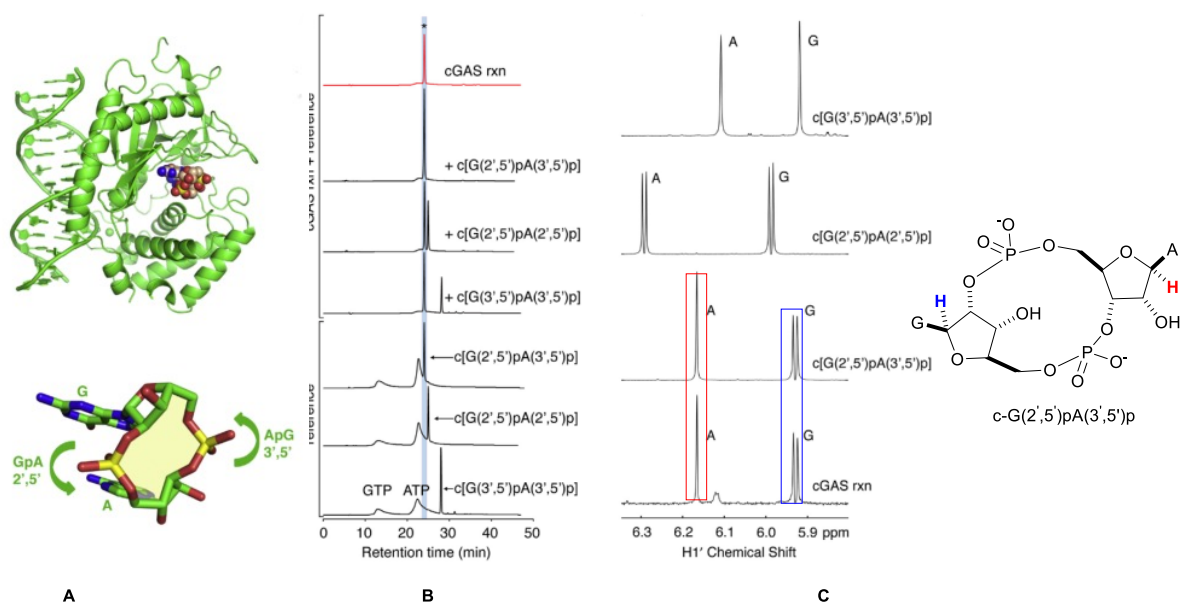
Figure 2.5 Structure study of c-G(2',5')pA(3',5')p by X-ray, HPLC and proton NMR analysis

22
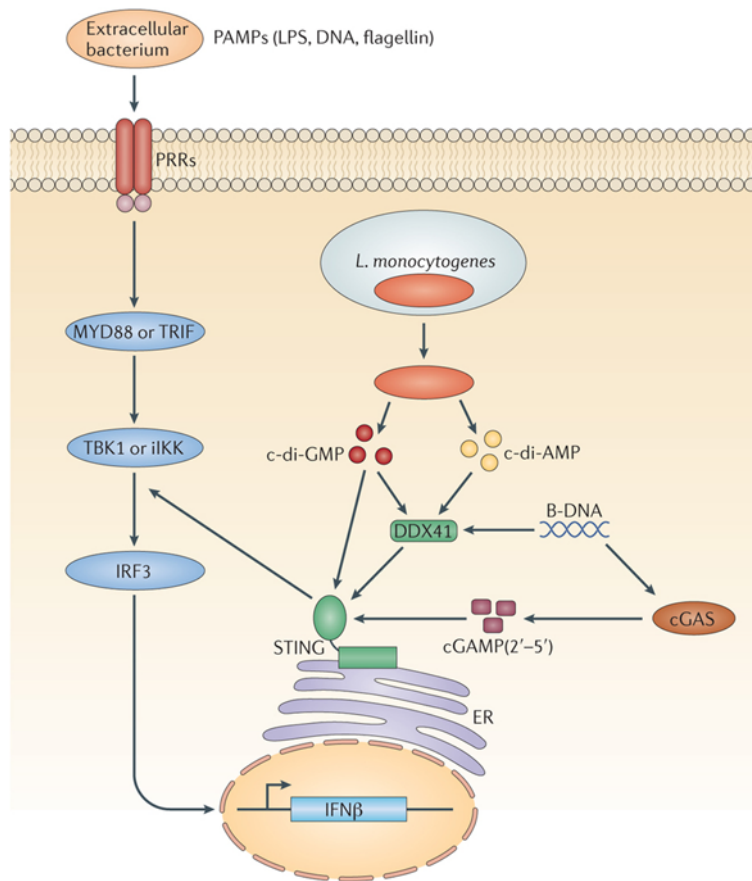
Figure 2.6 CDNs signaling in innate immune system. [Reprinted with permission from Nature Publishers Ltd: [Nat Rev Micro (R. M. Corrigan and A. Grundling, Nat Rev Micro, 2013, 11, 513-524), copyright (2013)].

CDNs are good factors for activating STING (Figure 2.6), thus stimulating the expression of various interferons, cytokines and T cell recruitment factors. They could be employed in cacer immunotherapy to amplify and strengthen immune activity. Aduro Biotech has proved that CDNs and their derivatives generate a potent immune response in preclinical models that specifically attacks tumor cells.[91-93] The development of next generation cancer immunotherapies targeting the STING pathway has raised a great hope for cancer patients.[91, 94]

In summary, 3'-5' CDNs produced by bacteria act as key factor in regulating their lifestyle. They are recognized in the host cell as exogenous alarmins by the innate immune sensor STING. Endogenous production of 2'-3' c-di-GAMP is stimulated after detection of pathogens' dsDNA leading to STING activation (Figure 2.6). The exogenous or endogenous CDNs interacting with STING study provides a great perspective in stimulating innate immune system for treatment of chronic infection, inflammatory states, and cancer immunotherapy.

2.2 Rh catalyzed NHI for customization of CDNs

2.2.1 Previous synthesis and modification of CDNs

Given the diverse roles of CDNs in prokaryotes and eukaryotes, great efforts have been devoted to the synthesis of CDNs and their derivatives. A One-Pot synthesis of 3'-5'-c-di-GMP and its thiophosphate analogues developed by Gaffney et al made large scale CDNs accessible (Scheme 2.1).[95] They started the total synthesis with the most common commercially available phosphoramidite. After 8 step in one flask, gram scale of CDNs could be obtained within 2 days. This method is also applicable in the synthesis of c-G(2',5')pA(3',5')p and its thiophosphate derivative utilizing DMT-3'O-TBDMS- 2'O-phosphoramidite.[67]
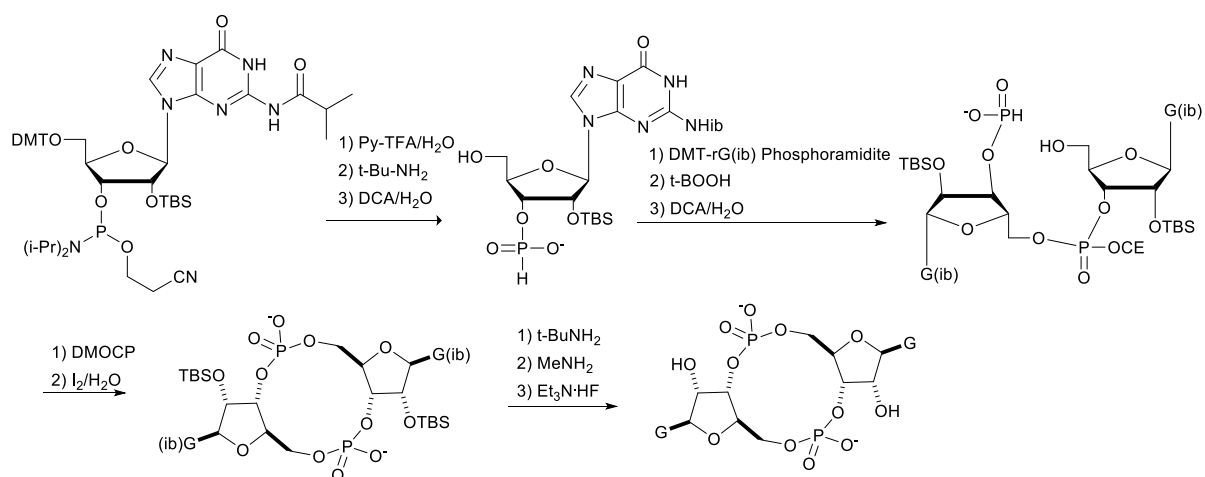
Scheme 2.1 One-Flask Syntheses of c-di-GMP

Levels of CDNs are regulated by synthases and phosphodiesterases in biological conditions. Phosphorothioate was introduced to prevent enzymatic hydrolysis because this artificial linkage can not be recognized by phosphodiesterases. The thioester derivatives have similar binding affinities to natural occurring CDNs. This has made the study of CDNs in artificial biological models possible (Scheme 2.2). The thioester derivatives of CDNs have been tested in many medical trials due to CDNs' potent impact on the innate immune system. For example, the structural analogue dithio-(R_P, R_P)-[c-A(2',5')pA(3',5')p] was shown to generate a potent anti-tumor response in mouse tumor models.[93]
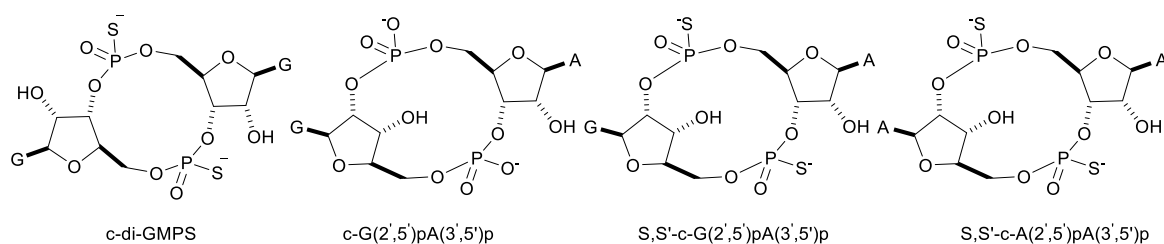
Scheme 2.2 CDNs and their phosphorothioate derivatives engineered for phosphodiesterase resistance

When we started our project in 2013, the modification of previous work on CDN modification had focused on making changes to the phosphate linkage or 2' position of the ribose.[96] For example, a propargyl group was introduced at 2' position of the phosphoramidite before the total synthesis. The propargylated CDNs provide the possibility for conjugating other functional groups like biotin or fluorescent tags.[97] Direct modification of CDNs was reported with more than a thousand equivalents of N-methylisatoic anhydride in the presence of 1M NaOH, however the yield was not clear, and the structure of the product was only confirmed by tandem MS.[98]
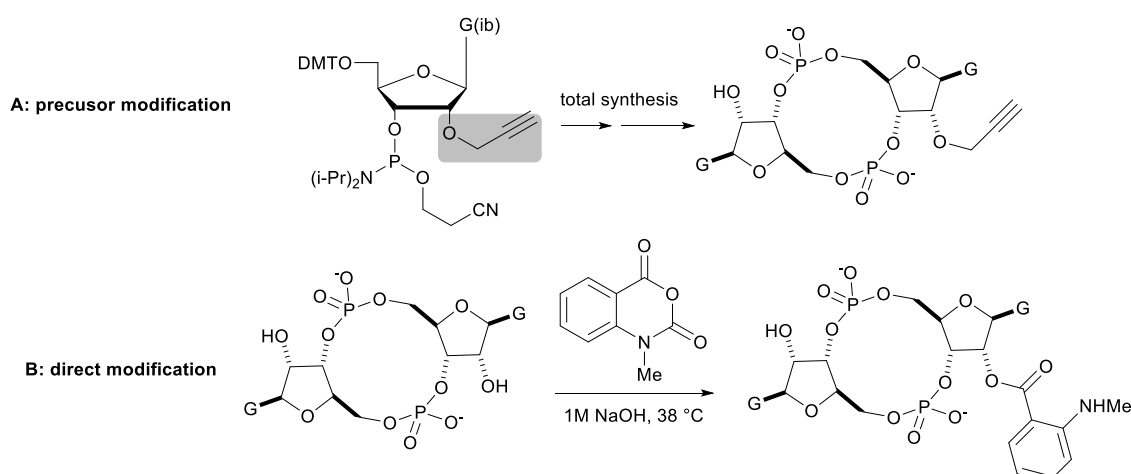
Scheme 2.3 previous modification of CDNs

Unravelling and reprogramming the complex biology of CDNs hinges on the ready availability of chemically tailored variants. While total chemical synthesis gives access to any variation, substantial expertise and labor are required. Direct modification is simpler, but until now these have focused on changes in the phosphate linkage or 2' position of the ribose with very low efficiency.

2.2.2 Preparative synthesis of exo-amine modified CDNs

In this chapter, our development of a one-step method to target the exocyclic amine of nucleobases in all types of natural 3'-5'-linked CDNs utilizing catalytic rhodium-based carbene transfer will be discussed. Direct nucleobase modification has never explored, but would provide an important

complement to previous approaches. As we discussed in chapter 1, metal catalyzed carbene reactions have emerged as a powerful technology in chemical biology. Our group recently has developed Rh-catalyzed NHI with diazo as a new tool for modifying nucleobases in DNA and RNA. While the technique is unselective in long single-stranded oligos, CDNs presented the more tractable problem of selecting between only two reactive N–H groups.

The test reaction started with c-di-GMP and methyl 2-(4-azidophenyl)-2-diazoacetate (Scheme 2.4). While the initial reaction started with 10mM c-di-GMP, no modification was observed. C-di-GMP forms interesting polymorphism of various oligomeric forms at millimolar concentrations, which causes base stacking and G-quartet interactions.[99, 100] When the c-di-GMP concentration was reduced to 2mM, within 20min, 58% of c-di-GMP was converted to the desired product. After preparative HPLC purification, 39% isolated yield was obtained (see also in Table 2.1, entry 4). The modified position was confirmed by 2D-NMR spectra. NMR assignments are available in the experimental part. The microscale thermophoresis confirms that the modified CDN bounds to its receptor protein DgrA at a similar level as the natural CDN, validating the exocyclic amine as a viable site for modification in the study of CDN biology.
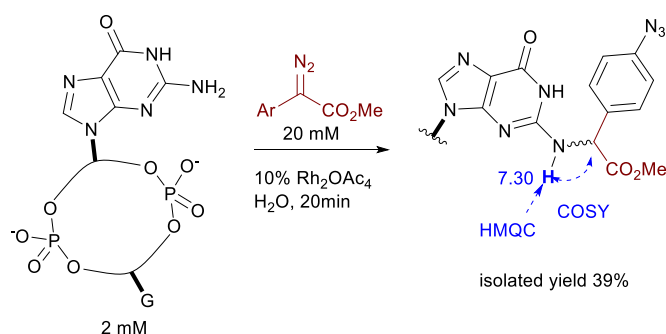
Scheme 2.4 Direct purine exocyclic amine modification of c-di-GMP

A handful of diazo compounds bearing common functional tags (amine for water solubility, azide for photo-crosslinking, alkyne for click chemistry) were synthesized and tested with CDNs (see Table 2.1). The reactions deliver mainly mono-modified CDNs along with some unproductive O-H insertion of the diazo starting material (hence 10 equivalents are required); in some cases minor double-modified products are also observed (0-18%). As shown in Table 2.1 the conversions range from 33-80% depending upon the precise substrate and diazo compound. The reactions are fast, requiring at most 2 hours to reach completion (entry 1-3, Table 2.1). We also investigated the phosphodiesterase resistant phosphorothioate derivative c-di-GMPS. Sulphur derivatives often hinder carbene transfer reactions, but in this case phosphorothioates are well-tolerated, delivering 80% conversion and 41% isolated yield (entry 6, Table 2.1).

Table 2.1 Rh-catalysed N-H insertion of CDNs with diazo compound
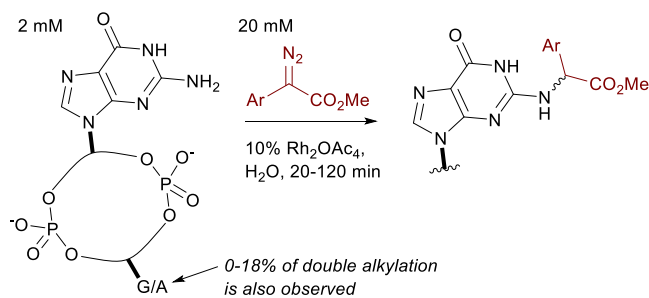
Entry	CDN	Ar	Main product	Time (min)	Conv (%) (yield)[a]
1	c-di-GMP			120	51 (39)
2	c-di-AMP			120	33[b]
3	c-GAMP			120	73 (19)
4	c-di-GMP			20	58 (40)
5	c-GAMP			60	67 (25)
6	c-di-GMPS			50	80 (41)
7	c-di-GMP			120	30[b]

[a] yield after prep-HPLC; [b] yield not determined due to low conversion.
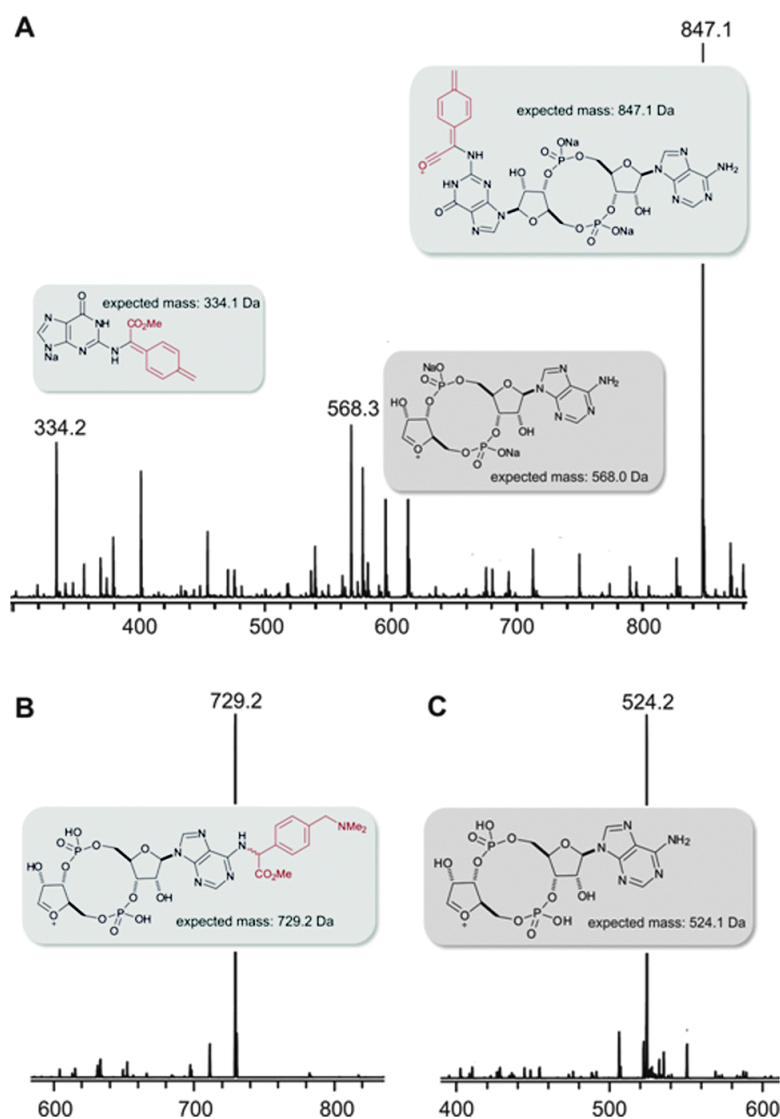
Figure 2.7 MS–MS analysis of alkylation products. Panel A: the major product from entry 3 in Table 2.1 delivered a daughter ion consistent with guanine alkylation; panel B: the minor product was consistent with adenine alkylation; panel C: the only isolated product from entry 5 in Table 2.1 delivered a daughter ion consistent with guanine alkylation.

One unexpected observation in Table 2.1 relates to the mixed dinucleotide c-GAMP: in the case of the dimethylamino derived diazo compound a mixture of guanine and adenine modified products were obtained in a 2 : 1 mixture according to HPLC analysis (entry 3, Table 2.1), but with the azide containing diazo substrate targeting of the guanine was far more selective (entry 5, Table 2.1 >80% selectivity for G alkylation according to integration of HPLC). The structures of the products were gleaned independently from NMR (ROESY, HMQC, and HMBC, see the experimental part for details) and MS–MS fragmentation (Figure 2.7). With the sample from entry 3 in Table 2.1, the major mass peaks matched the guanine-modified structure (see panel A, Figure 2.7). In contrast, the minor product from entry 3 in Table 1 delivered the modified adenine fragment (panel B, Figure 2.7); while the azide-containing diazo substrate (entry 5, Table 2.1) almost exclusively targeted the guanine (panel C, Figure

2.7). Alkylation on the phosphate was ruled on the basis that there was a strong HMBC correlation between the α-hydrogen derived from the diazo substrate and the nearest carbon on the nucleobase in each case. We had expected that the unsymmetrical CDN c-GAMP would represent a substantial challenge in chemoselectivity since adenine and guanine display similar functional groups to the catalyst. However, as entries 3 and 5 demonstrate, the substrate can play a role in controlling the site-selectivity. The source of the change could be related to the charge of the dimethylamino group, or the propensity of certain CDNs to form higher-order aggregates in solution. Although the product mixture obtained from entry 3 of Table 2.1 was more complex, we were able to separate each component (the 39% reported yield corresponds to the mixture) and therefore c-GAMP derivatives are available with alterations at either base through one protocol.

In prokaryotes CDNs are involved in a complicated regulatory network involving a multitude of individual protein components and several riboswitches. The modified CDN derivatives shown in Table 1 are versatile starting points for exploring the biology of these second messenger molecules. For example, the azide motif can be converted to a fluorescent CDN derivative through a catalytic azide-alkyne cycloaddition (Scheme 2.5). Furthermore, the aryl azide itself is a common photo-crosslinking group and therefore compounds such as those found in Table 2.1 (entries 4-6) could be used to probe binding sites of c-di-GMP receptors. Although 2'-hydroxyl derived probes are known, a family of photo-crosslinkers is important since different receptors will have different binding constraints.
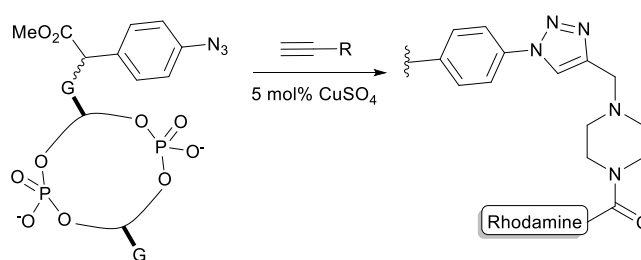
Scheme 2.5 Synthesis of a fluorescent CDN derivative through a catalytic azide–alkyne cycloaddition

2.2.3 Photocrosslinking of modified CDN with its receptor protein

To explore the photo-crosslinking of azide-modified CDNs we selected the known c-di-GMP receptor DgrA, a PilZ homolog that mediates c-di-GMP-dependent control of *Caulobacter crescentus* cell motility. Its high affinity and specificity towards c-di-GMP and the availability of binding mutants make this protein an ideal test bed. The aryl azide modified c-di-GMP (c-di-GMP-N$_3$) was incubated with the protein for 15 h under 366 nm irradiation (see top of Figure 2.8). The mixtures were analyzed by high-resolution ESI mass spectrometry, gel electrophoresis, MALDI-TOF, and finally the site of modification

was determined by a trypsin digest (see experimental for details). Even with a single equivalent of c-di-GMP-N$_3$ (lane 3, Figure 2.8) DgrA was covalently modified in a yield of 17% (according to integration of the gel bands). At 10 equivalents of c-di-GMP-N$_3$ there was complete conversion (lane 8, Figure 2.8), but the diffuse bands suggested competitive unspecific modification. Previous work has shown that Arg11, Arg12, and Trp75 are important residues for c-di-GMP binding of dgrA: A W75A mutant decreased binding 102-103-fold, while binding was completely abrogated in the R11A/R12A double mutant. Consistent with the reported binding studies, reaction of c-di-GMP-N3 with the W75A mutant gave reduced crosslinking (cf. lane 5 versus lane 3, Figure 2.8) and the R11A/R12A was not detectably modified (lane 7). A trypsin digest of the photo-crosslinking reaction revealed one new peak in the LC-MS whose mass was consistent with modification of the GGR peptide fragment shown in red in Figure 2.8. This tripeptide sits directly in the region of the purported c-di-GMP binding site (blue in Figure 2.8). Taken together these results demonstrate that c-di-GMP-N$_3$ is a selective cross-linking probe efficient enough to determine binding sites in CDN receptors.
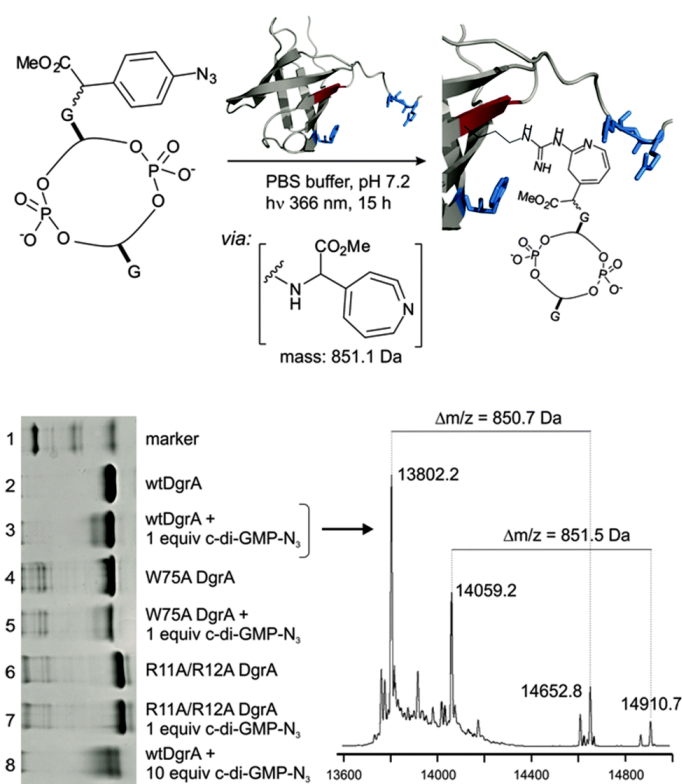
Figure 2.8 Modified c-di-GMP-N$_3$ maintains binding to DgrA and can be used for photo-crosslinking. The peak at 14059.2 is likely a post-translational modification which is also cross-linked with the azide compound (14910.7). The baseline impurities in the ESI stem from gradual photolytic degradation of DgrA

2.2.4 Conclusion

In summary, we describe a direct method for the synthesis of CDN derivatives modified at the exocyclic amine of the purine bases. This method tolerates all members of the cyclic dinucleotide family and could be used to modulate their function or introduce useful side-chains such as fluorophores and photo-crosslinking groups. The reaction is trivial to execute, making it accessible to non-experts in synthesis and catalysis. The most synthetically challenging aspect of the approach is in the synthesis of the diazo compounds, which typically require 3-5 operations.

New aspects of CDN biology are continually being unveiled. A challenge for chemical biologists is to provide a selective probe for each natural CDN receptor. The process we have described adds a new method for such bespoke probe development.

Further study could be continued in two directions due to the important role of CDNs in chemical biology. How modifications of the exocyclic amine of CDNs behave with innate immune sensor? Will the covalent linked CDNs and receptor (CDN-STING for example) stimulate the immune response since the receptor is in the constant active form? Answering these questions might offer new insights for CDN inspired therapeutics.

2.3 Copper catalyzed NHI with diazo compound

I was involved in this project with Kiril Tishnov, and contribute my work in double-stranded DNA motif alkylation and the auto-tandem catalytic reaction.

Our group recently reported Rh-carbenoids derived from α-diazo carbonyl compounds can target the exocyclic N–H group of various nucleic acids including short ssDNA, hairpin DNA, ssRNA, hairpin RNA and CDNs. The dominance of rhodium-based catalysts has left Cu(I) systems underdeveloped even though they were reported earlier than Rh (See the discussion in Chaper 1). Tracing through the past years of developments, there are a few reactions employing cooper as the XHI catalyst in chemical biology. In the sixties large excesses of Cu(II) salts were used to effect protein modification with diazo peptides.[31, 32] A very recent report has shown that intramolecular Cu-catalyzed cyclopropanation can be carried out in aqueous media using DNA as a chiral control element. These observations hint at unrealized potential for catalytic Cu(I)-carbenoid chemistry in water.[101] Combined with recently development of Cu(I)-carbenoids with diazo compounds in XHI reactions, it could offer another possibility for introducing unnatural chemical motifs into native nucleic acid. Additionally the Cu(I) is also a well-known catalyst in click reaction which enable the efficient and straight-forward introduction of a variety of functionally important tags and reporter groups. Thus the Cu(I) in the NHI reaction system could be also used for click reaction.

2.3.1 Copper (I) catalyzed NHI in dsDNA alkylation

Given our interest in catalytic methods for NA alkylation, our initial examination of the catalytic NHI started with the short oligonucleotide d(ATGC) in aqueous MES buffer at pH 6. Further refinement of the reaction conditions, including testing different ligands and copper sources, led to a convergence with the CuAAC conditions for bioconjugation developed by Finn. The key components of the Finn conditions are sodium ascorbate and the tris(3-hydroxypropyltriazolyl- methyl)amine (THPTA) ligand. Ascorbate insures the Cu(I) oxidation state is maintained even in the presence of oxygen. A consequence of Cu(I) in the presence of oxygen is the formation of reactive oxygen species (ROS). The role of the THPTA ligand seems to be as a sacrificial substrate to scavenge ROS before they damage nucleic acid. The modification of the model oligonucleotide d(ATGC) was a lot more efficient than with Rh, as the conversion of up to 70%, accompanied by formation of over twelve alkylation products were observed. Other short ssDNA oligomers were also tested within this system. It gave higher conversion and more alkylation sites compared to rhodium catalyzed system. Based on LC-MS date, the modification is mainly on the base, could be exo and endocyclic nitrogen, but we could not completely characterize these products due to insufficient material.

The propensity of double-stranded DNA motifs to react was tested on three hairpin structures (Table 2.2). All three hairpins were significantly less reactive than the single-stranded NAs tested, delivering only modest yields of modified DNA at extended reaction times. As expected the hairpin containing only Ts in the loop proved to be the least reactive with only 9% conversion. In comparison the remaining two hairpins exhibit a two-fold increase in reactivity. These contain an unpaired A as either a 3'-overhang or in the loop region, indicating that copper preferentially targets nitrogens that are not involved in Watson–Crick base-pairing.

Table 2.2 hairpin DNA modification

Hairpin oligonucleotide + Dz (Me₂N–CH₂–C₆H₄–C(=N₂)CO₂Me) → Cu(I)[a]

Entry	Sequence	Structure	number of alykation products	Conv.[c]
1	d(CGAACGTTT TTCGTTCG)		2	9%
2	d(CGAACGTTT TTCGTTCGA)		4	16%
3	d(ACGGAAT TCCGTTTTT CGGAATTCCG)		6	16%

[a]Reaction conditions : 5 mM oligonucleotide, 0.5 mM CuSO₄ 2.5 mM THPTA, 50 mM diazo substrate, 10 mM sodium ascorbate, 100 mM MES, pH 6, 48 h, room temperature. [b]Amount of the modified species as a percentage of the total amount of NA.

Based on the fact that our NHI process is fully compatible with the Cu-catalyzed azide–alkyne cycloaddition (CuAAC), therefore we attempted to investigate the possibility of a simultaneous CuAAC/NHI process. The concept was tested with a collection of different amines, alkynes, diazo compounds, and azides as shown in Table 2.3. Our initial experiments with small molecule substrates and in all cases the N-aryl triazole/NHI products were obtained in good yields (53–70%). Furthermore, changing the position of the participating functional groups had little impact on the reaction. The system was later tested with more challenging DNA-based substrates and as we expected it gave two consecutive modifications in one pot reaction.

Table 2.3 Auto-tandem catalytic CuAAC/NHI

amine	diazo	azide or alkyne	product	yield
				57%
				70%
				63%
				53%

These reactions were performed with 5 mol% $CuSO_4$ and 50% (v/v) t-BuOH.

2.3.2 Conclusion

Our work outlines a novel approach for molecular conjugation based on the discovery that Cu(I)-carbenoid chemistry is viable in water. The efficacy of Cu(I) for catalyzing both CuAAC and NHI drove us to combine both reactions in a one-pot operationally simple process. The substrate range includes simple arylamines as well as the nucleobases in DNA and RNA. This multi-component catalytic process not only reduces the set up cost but also saves purification labor work. It should prove useful for medicinal and combinatorial chemists since it provides a robust strategy to quickly assemble complex molecular scaffolds.

2.4 Experimental part

2.4.1 General

All reagents and solvents used were of analytical grade. Buffers were prepared with ultrapure water. All chemicals were purchased from Sigma-Aldrich, Fluka or Acros and used as received. Analytical TLC was performed on Silica gel 60 F_{254} pre-coated aluminium sheets. Flash chromatography was performed on Silica gel 60 40-63 μm (230-400 mesh) (SiliCycle, Quebec). ^{1}H and ^{13}C NMR and 2D spectra of modified CDN were acquired on a Bruker AvanceIII+ 600 MHz using Shigemi NMR tube. Other compounds were recorded on 400 MHz proton frequency spectrometer at 298 K. Chemical shifts relative to TMS were referenced to the solvent's residual peak and are reported in ppm. ESI MS-MS spectra were obtained on a Bruker Esquire3000plus spectrometer by direct injection in positive polarity of the ion trap detector. High resolution mass spectra were acquired on a Bruker maXis 4G QTOF ESI mass-spectrometer. MALDI TOF analyses were carried out on a Bruker Microflex mass-spectrometer in linear positive mode using sinapic acid as matrice. HPLC procedures were carried out on an Agilent 1100 LC system equipped with Eclipse XDB-C8 5μm 4.6 x 100 mm column (Agilent) for analytic analysis. Shimadzu preparative HPLC (LC-20AP) equipped with phenomenex column (Gemini® 10 μm C18 110 Å, LC Column 250 x 21.2 mm, AXIA™ Packed) was employed for preparative purification. 100 mM triethylammonium acetate (pH 7.2) and acetonitrile was used as a mobile phase. For analytical measurement: 1 mL/min: 0-35 % acetonitrile in 12 min, 35-80 % acetonitrile in 3 min, 80% acetonitrile in 2 min. For preparative separation: flow rate: 20 mL/min, 0% acetonitrile in 2 min, 0-60 % acetonitrile in 27 min. Detection was carried out by monitoring the absorbance of the column effluent at 254 nm. UV cross linking reaction was carried out with CAMAG TLC UV lamp at 366 nm. The sample was placed 2 cm away from the UV lamp.

2.4.2 General procedure for $Rh_2(OAc)_4$-catalyzed CDN modification with diazocarbonyl compounds.

For analytical reactions, typically 20 μL reaction mixtures containing 2 mM oligonucleotide, 200 μM $Rh_2(OAc)_4$ and 20 mM α-diazocarbonyl compound in 100 mM MES buffer, pH 6.0 were reacted at room temperature. Only aryl azide diazo modification of CDN was carried out in 50% water and 50% t-BuOH due to the solubility of aryl azide diazo compound. The reaction was traced by analytic HPLC and the identity of the product in each fraction was confirmed by HR-ESI.

To obtain enough products for NMR characterization preparative reactions were run on a 24 mL reaction scale. For example 4.8 mL 10 mM c-di-GMP, 2.4 mL 2 mM $Rh_2(OAc)_4$ and 12 mL 40 mM aryl azide diazo in t-BuOH and additional 4.8 mL water were mixed together. The reaction finished in 30 minutes. 3 mL ethyl acetate was added to the reaction mixture to remove the organic side product,

and aqueous layer was freezed dried and applied to preparative HPLC for purification. The modified compounds were freeze dried and confirmed by HRMS (ESI) and NMR. In some cases, pure diastereomers could be separated. However epimerization occurred very faster after purification. Due to epimerization and low conversion, only c-di-GMP-N$_3$ and C-GAMP-N$_3$ pure diastereomers were successfully characterized by 2D NMR, and the modification on the exocyclic amine of guanine was confirmed by COSY, HMQC, and separately by MS-MS analysis.
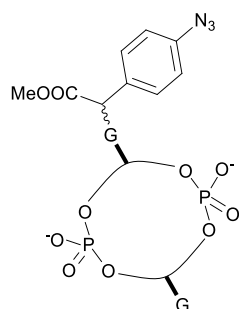
12 mL reaction mixtures containing 2 mM c-di-GMP, 200 μM Rh$_2$(OAc)$_4$ and 20 mM α-diazocarbonyl compound in 100 mM MES buffer, pH 6.0 were kept at room temperature for 3h. The reaction was traced by analytic HPLC (conv. 51%) and the mixture of diasteromer was isolated by preparative HPLC (39%). ^{1}H NMR (600 MHz, DMSO) δ 8.01 (d, *J* = 7.4 Hz, 1H), 7.99 – 7.93 (m, 2H), 7.43 (t, *J* = 8.2 Hz, 2H), 7.35 (dd, *J* = 13.6, 8.0 Hz, 3H), 7.25 (d, *J* = 7.9 Hz, 1H), 6.62 (s, 2H), 6.29 (s, 1H), 5.77 (t, *J* = 8.0 Hz, 1H), 5.67 (d, *J* = 8.2 Hz, 1H), 5.41 (dd, *J* = 19.0, 6.2 Hz, 1H), 4.89 (s, 1H), 4.65 (dd, *J* = 38.2, 17.6 Hz, 4H), 4.18 (dd, *J* = 10.5, 5.8 Hz, 2H), 4.00 (s, 2H), 3.79 (dd, *J* = 21.2, 11.1 Hz, 3H), 3.69 (d, *J* = 14.0 Hz, 3H), 2.45 (q, *J* = 7.1 Hz, 6H), 1.87 (s, 6H), 0.94 (t, *J* = 7.1 Hz, 9H).HRMS (ESI): calcd for [C$_{32}$H$_{40}$N$_{11}$O$_{16}$P$_2$]$^+$ 896.2051, found 896.2121.

S 2.1 HPLC trace

37

diastereomer 1, DMSO, 313 K, 600 MHz
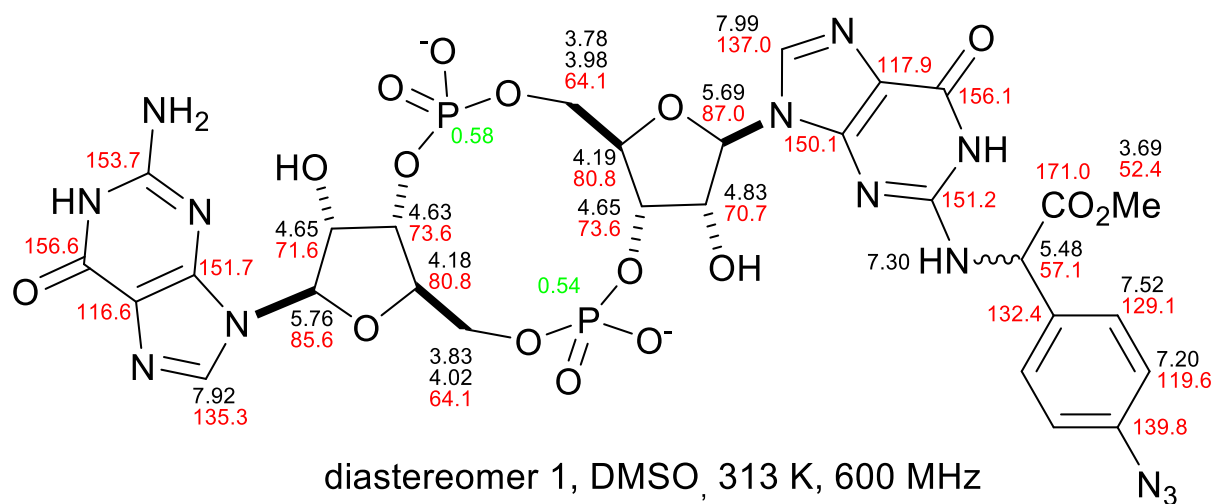
S 2.2 NMR assignment

S 2.3 ^{1}H NMR in DMSO, 313 K, 600 MHz
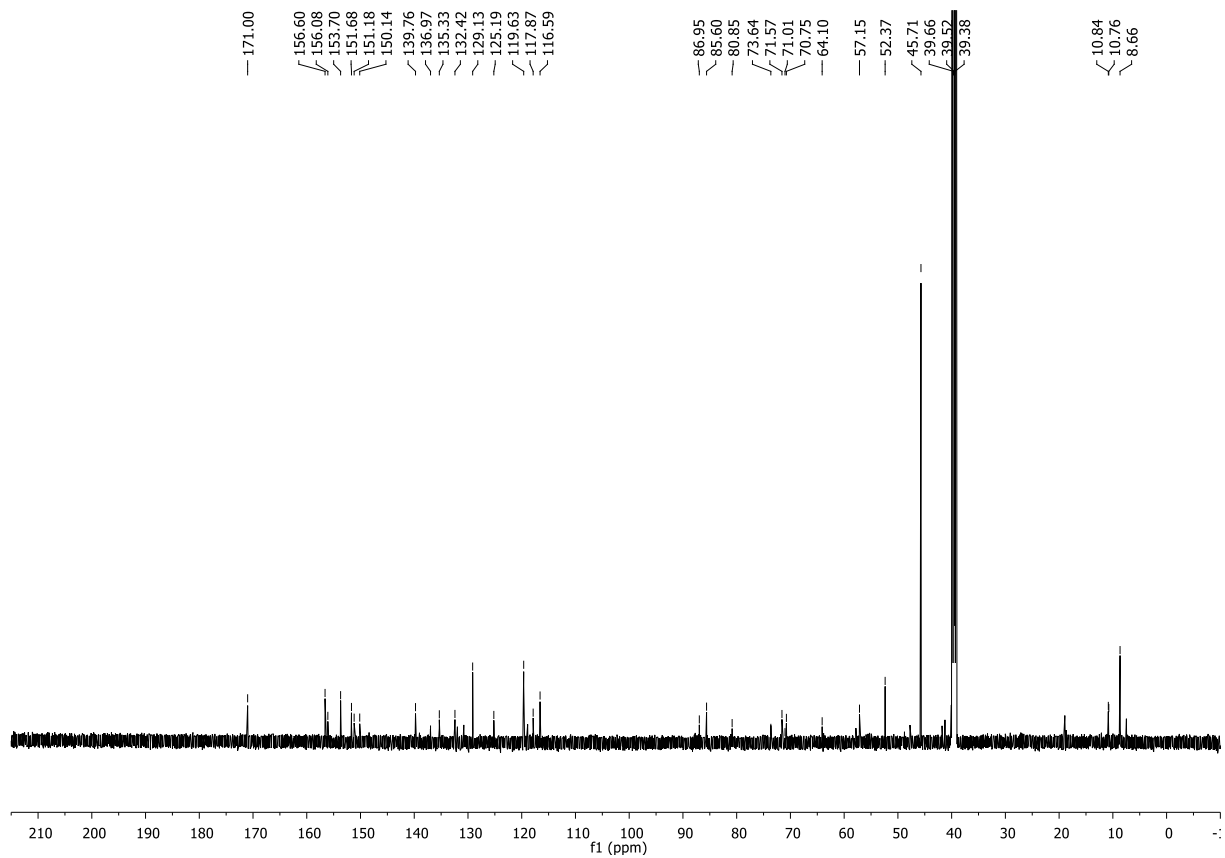
S 2.4 ¹³C NMR in DMSO, 313 K, 600 MHz
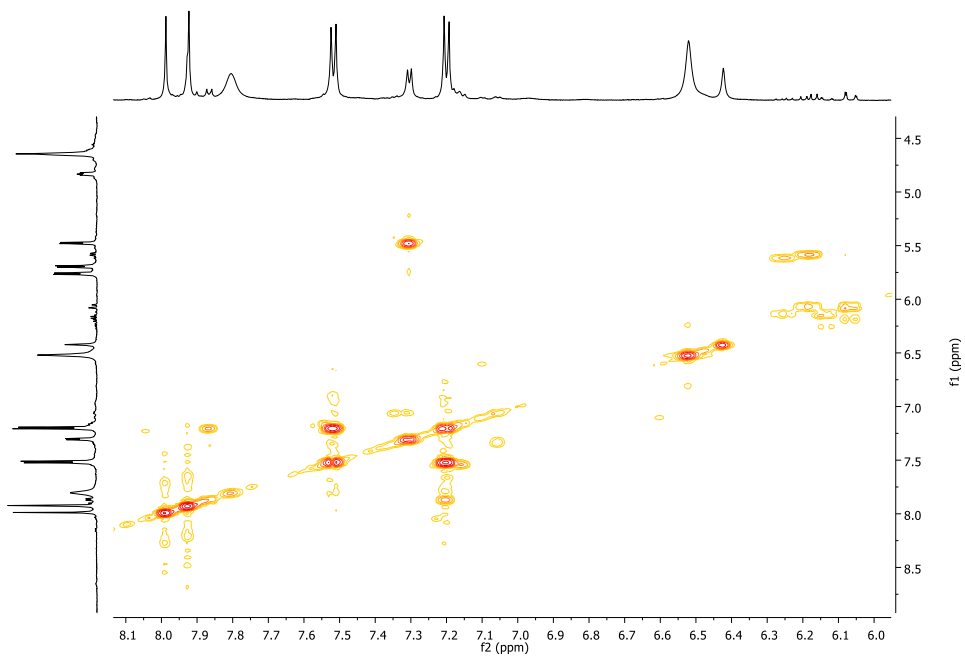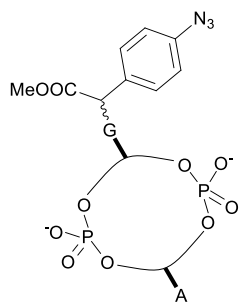
Peaks: 171.00, 156.60, 156.08, 153.70, 151.68, 151.18, 150.14, 139.76, 136.97, 135.33, 132.42, 129.13, 125.19, 119.63, 117.87, 116.59, 86.95, 85.60, 80.85, 73.64, 71.57, 71.01, 70.75, 64.10, 57.15, 52.37, 45.71, 39.66, 39.52, 39.38, 10.84, 10.76, 8.66

S5 ³¹P NMR in DMSO, 313 K, 600 MHz

Peaks: 0.59, 0.55

39

S 2.6 HMQC spectrum in DMSO, 313 K, 600 MHz

S 2.7 COSY spectrum in DMSO, 313 K, 600 MHz

S6

12 mL reaction mixtures containing 2 mM c-GAMP, 200 μM Rh$_2$(OAc)$_4$ and 20 mM aryl azide diazocarbonyl compound in 50% water and 50% *t*-BuOH were kept at room temperature for 30 min. The reaction was traced by analytic HPLC (conv. 67%) and isolated by preparative HPLC (iso. 22%). ^{1}H NMR (600 MHz, MeOD) δ 8.44 (s, 1H), 8.15 (s, 1H), 8.04 (s, 1H), 7.55 (d, *J* = 8.4 Hz, 2H), 7.10 (d, J = 8.5 Hz, 2H), 6.08 (s, 1H), 5.82 (s, 1H), 5.61 (s, 1H), 5.00 (s, 1H), 4.95 (d, *J* = 4.7 Hz, 1H), 4.41 – 4.29 (m, 5H), 4.13 – 4.09 (m, 2H), 3.78 (s, 3H), 3.20 (q, *J* = 7.3 Hz, 3H), 1.30 (t, *J* = 7.3 Hz, 5H). ^{13}C NMR (151 MHz, MeOD) δ 173.10, 159.23, 156.91, 153.10, 150.92, 150.00, 142.03, 141.10, 138.05, 135.32, 131.05, 120.67, 91.77, 91.43, 82.30, 82.06, 75.96, 75.77, 72.91, 72.70, 64.10, 59.62, 53.77, 49.44, 49.30, 49.15, 9.51. 31P NMR (243 MHz, MeOD) δ -0.67, -0.84. HRMS (ESI): calc'd for [C$_{29}$H$_{32}$N$_{13}$O$_{15}$P$_2$]$^+$ = 864.5821; found 864.1605.
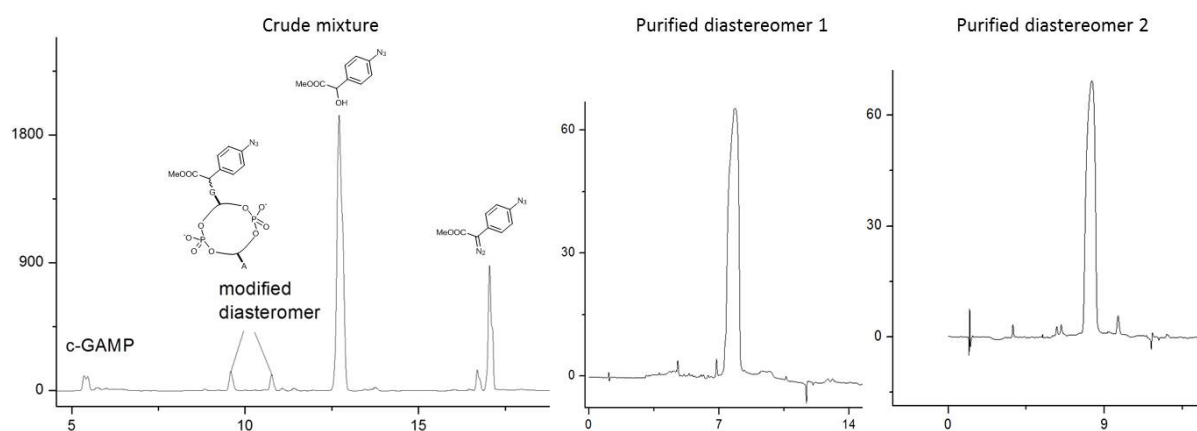
S 2.8 HPLC trace
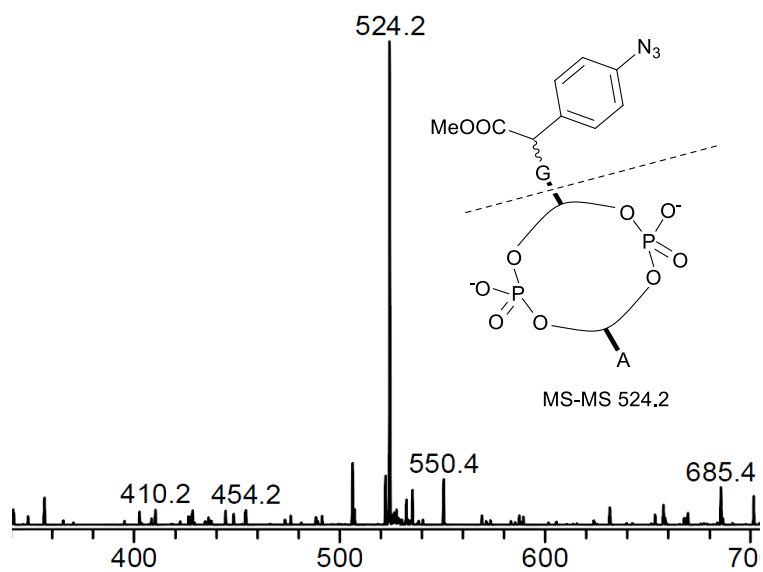
S 2.9 MS/MS fragment of product

diastereomer 2, CD$_3$OD, 318 K, 600 MHz

S 2.10 NMR assignment
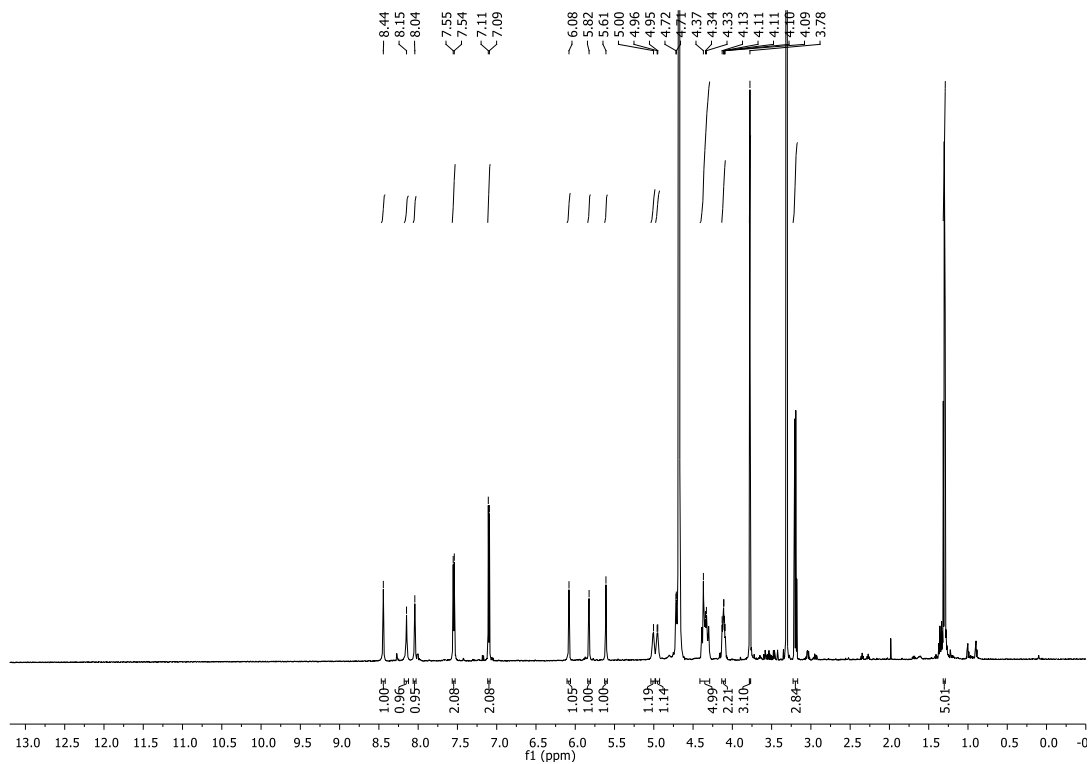
S 2.11 ^{1}H NMR in CD$_3$OD, 318 K, 600 MHz
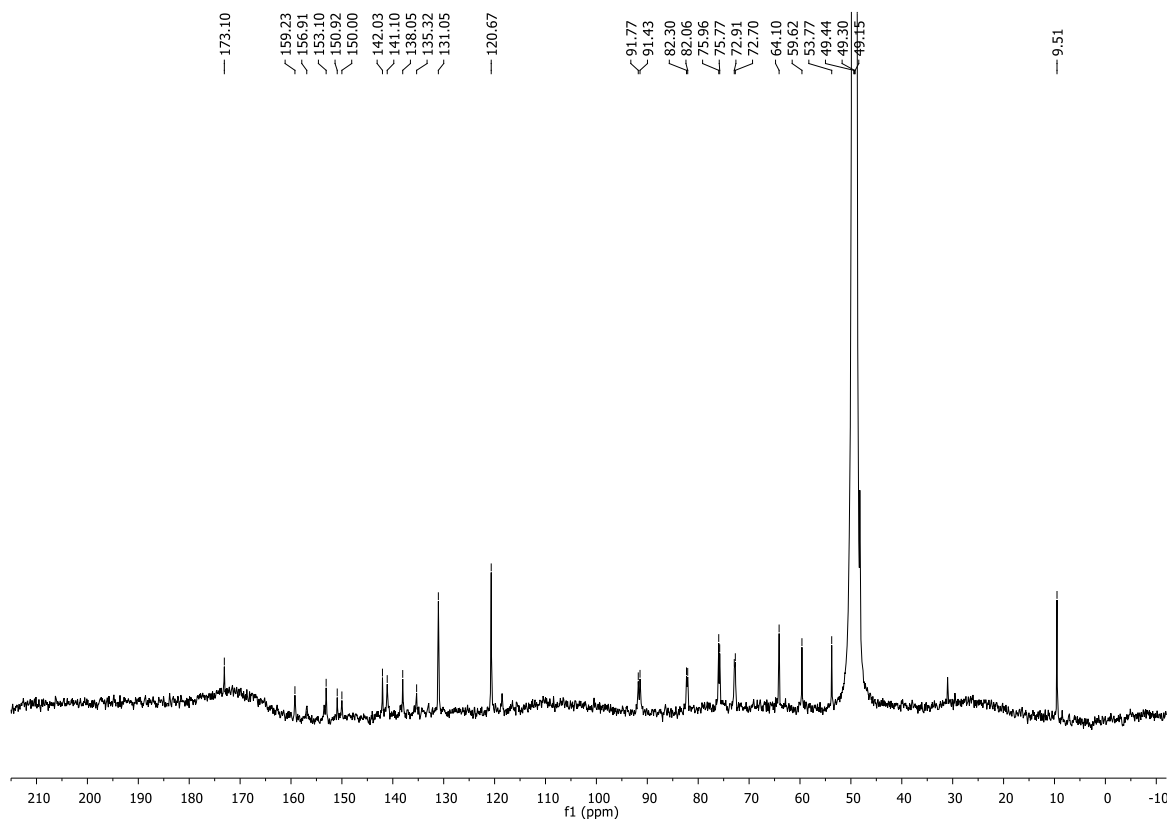
S 2.12 ^{13}C NMR in CD$_3$OD, 318 K, 600 MHz
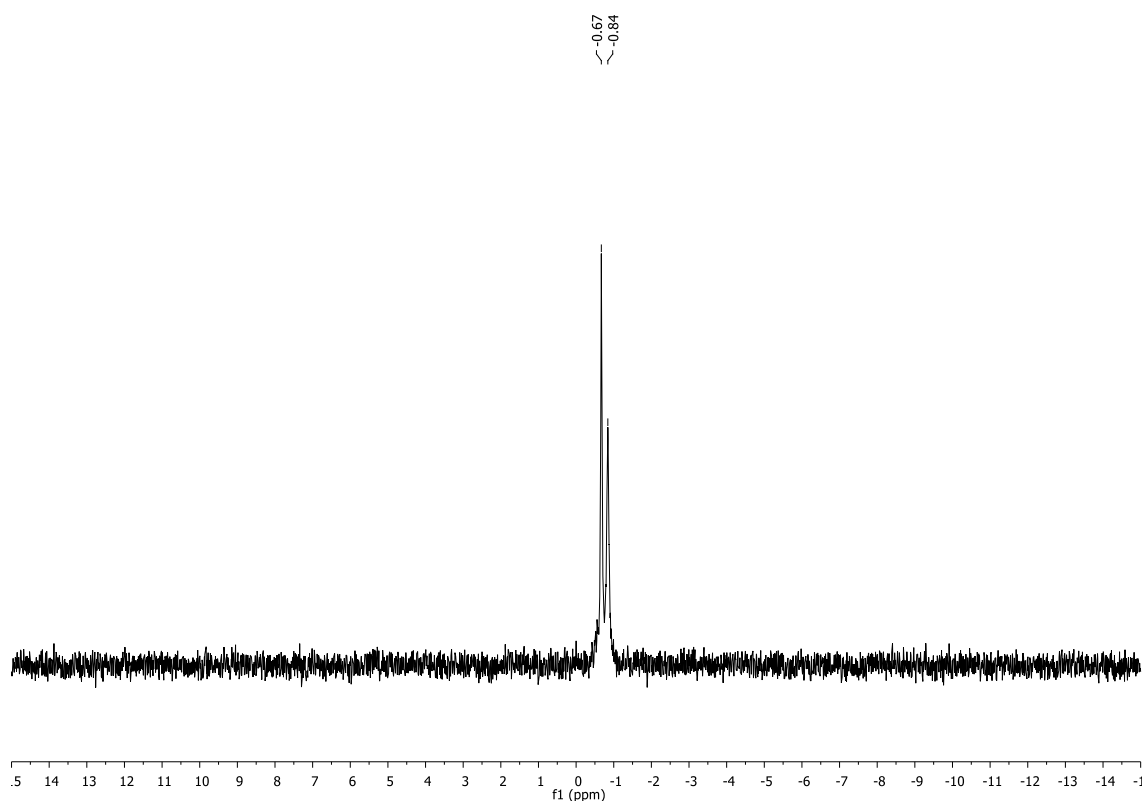
S 2.13 ^{13}C NMR in CD$_3$OD, 318 K, 600 MHz
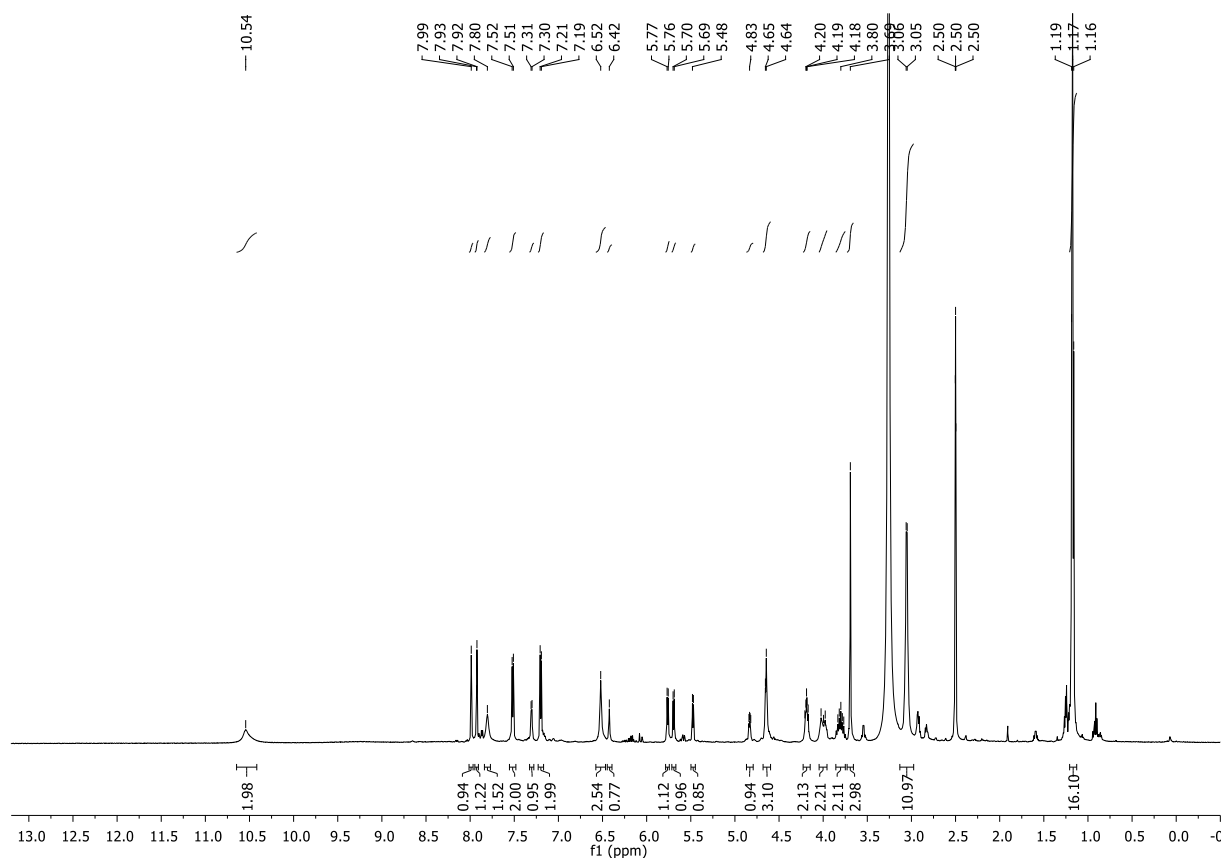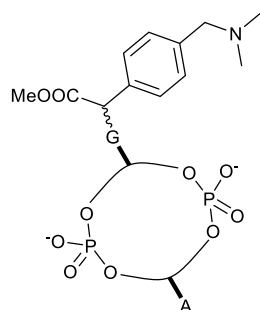

12 mL reaction mixtures containing 2 mM c-GAMP, 200 μM Rh$_2$(OAc)$_4$ and 20 mM α-diazocarbonyl compound in 100 mM MES buffer, pH 6.0 were kept at room temperature for 2h. The reaction was traced by analytic HPLC (73%) and the mixture of diasteromer was isolated by preparative HPLC (29%). ^{1}H NMR (600 MHz, DMSO) δ 10.67 – 10.45 (m, 2H), 8.42 (d, *J* = 10.8 Hz, 1H), 8.15 (s, 1H), 8.00 (s, 1H), 7.96 (d, *J* = 17.9 Hz, 1H), 7.44 (t, *J* = 8.8 Hz, 2H), 7.40 – 7.31 (m, 2H), 7.27 (d, *J* = 11.3 Hz, 2H), 6.63 (s, 1H), 6.54 (s, 1H), 5.88 (d, *J* = 7.7 Hz, 1H), 5.74 (d, *J* = 8.4 Hz, 1H), 5.66 (s, 1H), 5.43 (d, *J* = 6.1 Hz, 1H), 5.39 (s, 1H), 4.95 (s, 1H), 4.86 (s, 1H), 4.75 (d, *J* = 34.2 Hz, 1H), 4.65 (d, *J* = 32.0 Hz, 2H), 4.20 (d, *J* = 19.9 Hz, 2H), 4.02 (s, 1H), 3.93 (d, *J* = 26.0 Hz, 2H), 3.70 (d, *J* = 8.8 Hz, 3H), 1.91 (s, 6H). HRMS (ESI): calc'd for [C$_{32}$H$_{40}$N$_{11}$O$_{15}$P$_2$·Et$_3$N]$^+$ = 980.2102; found 980.2210.
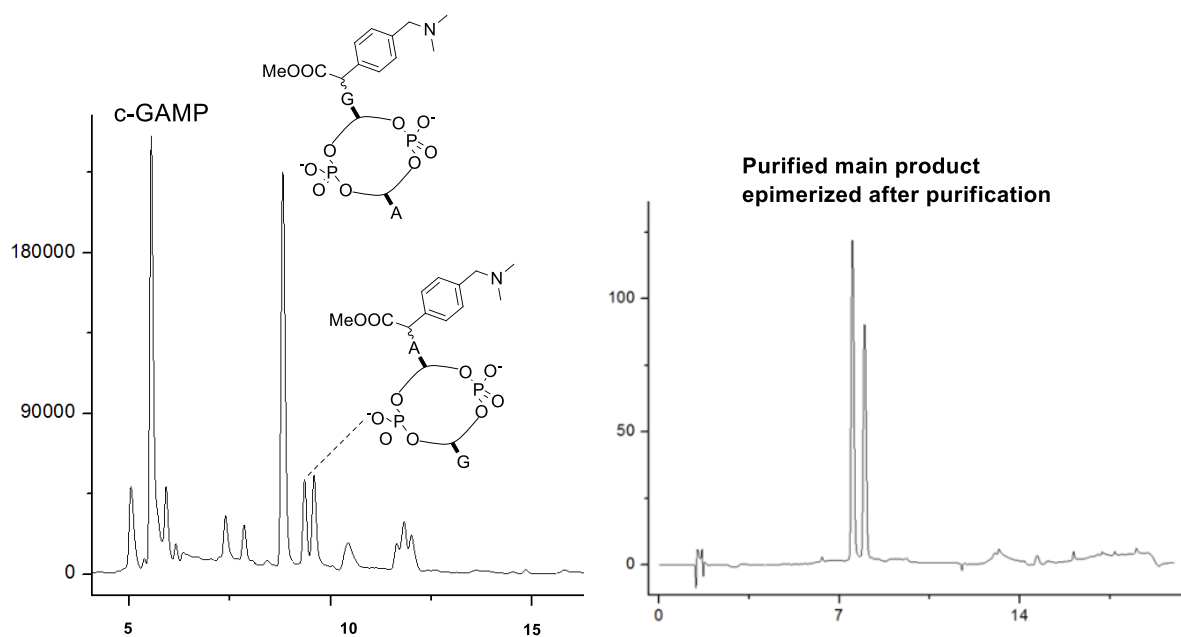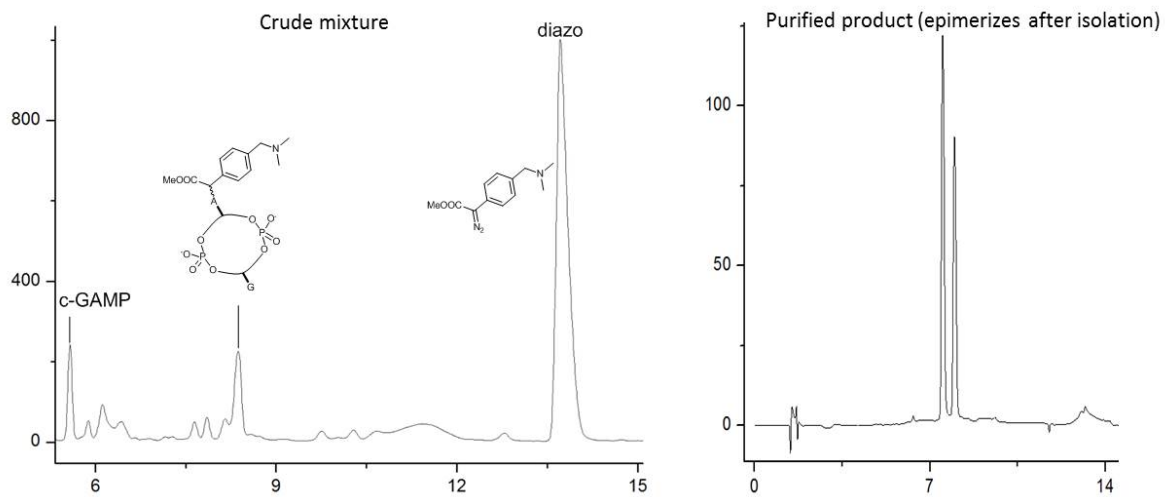
S 2.14 HPLC trace
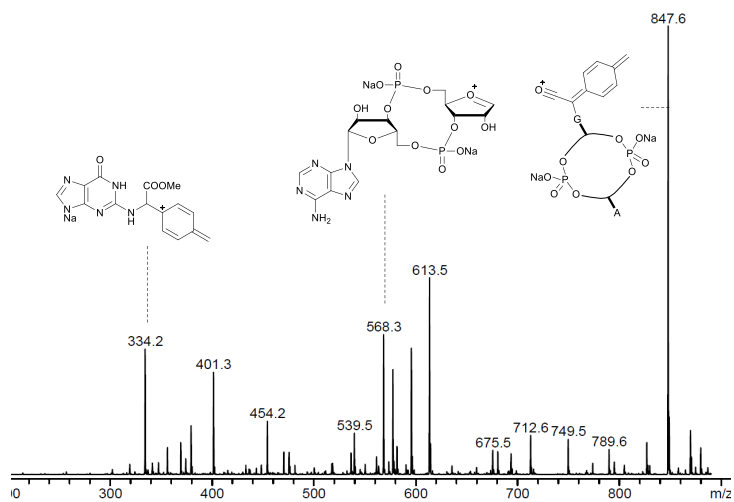
S 2.15 MS/MS fragment of main product
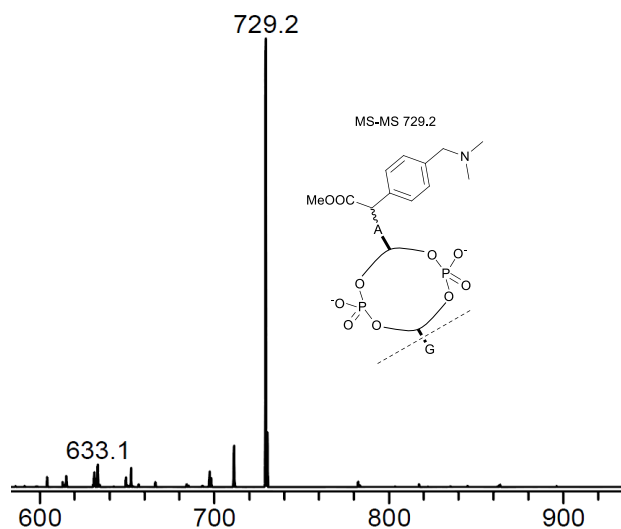
S 2.16 MS/MS fragment of side product

c-GAMP-NMe$_2$ in CD$_3$OD, 318 K, major isomer
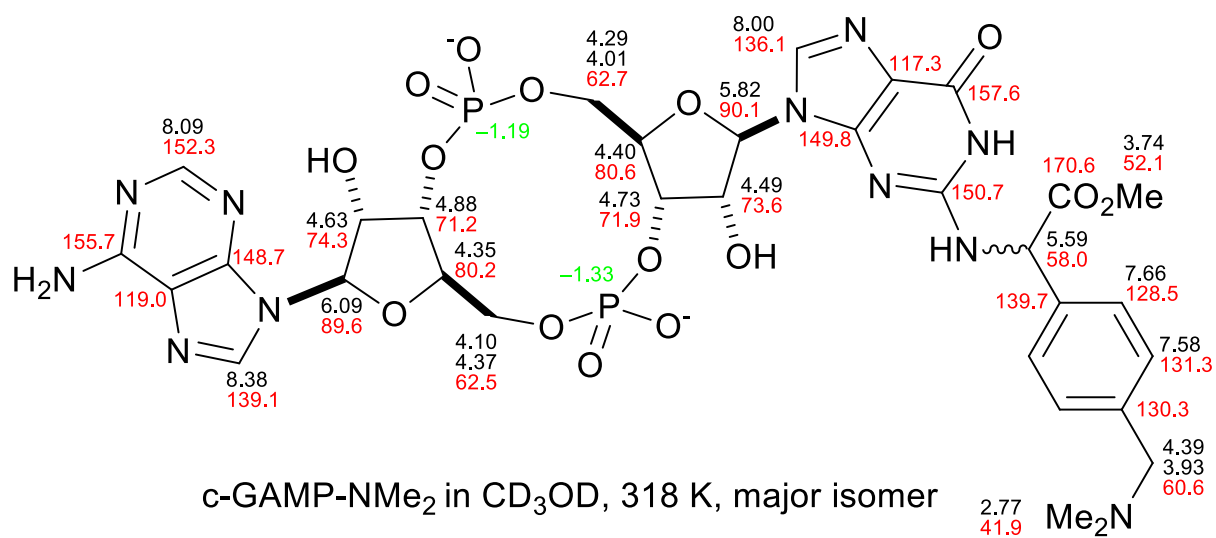
S 2.17 NMR assignment of main product: major diastereomer

c-GAMP-NMe₂ in CD₃OD, 318 K, minor isomer
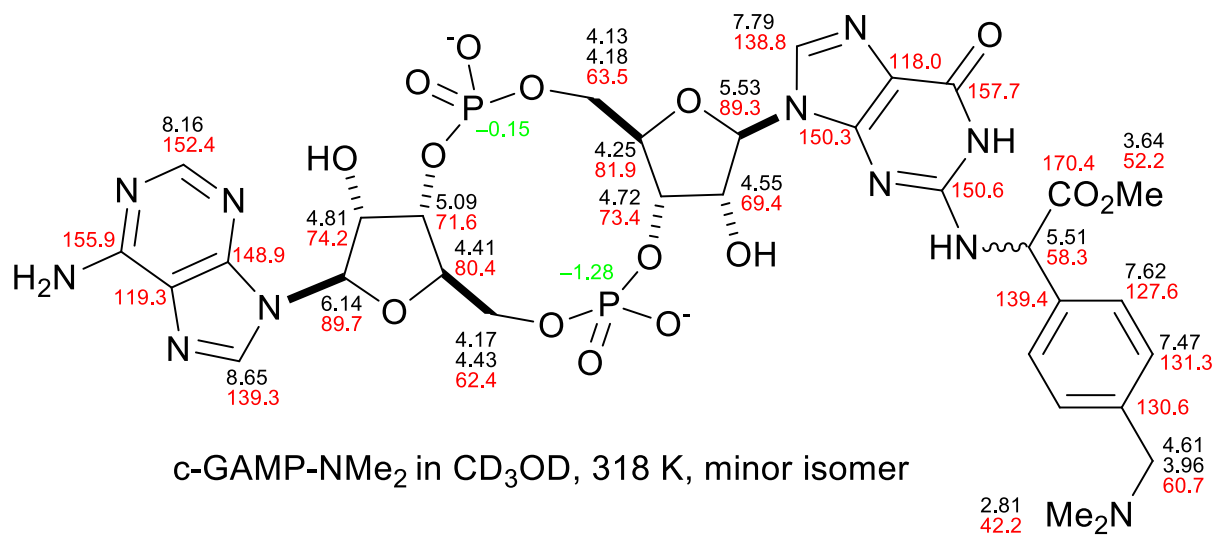
S 2.18 NMR assignment of main product: minor diastereomer
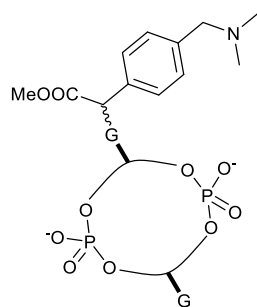
12 mL reaction mixtures containing 2 mM c-di-GMP, 200 μM Rh$_2$(OAc)$_4$ and 20 mM α-diazocarbonyl compound in 100 mM MES buffer, pH 6.0 were kept at room temperature for 3h. The reaction was traced by analytical HPLC (conv. 51%) and the mixture of diasteromers was isolated by preparative HPLC (39%). ^{1}H NMR (600 MHz, DMSO) δ 8.01 (d, *J* = 7.4 Hz, 1H), 7.99 – 7.93 (m, 2H), 7.43 (t, *J* = 8.2 Hz, 2H), 7.35 (dd, *J* = 13.6, 8.0 Hz, 3H), 7.25 (d, *J* = 7.9 Hz, 1H), 6.62 (s, 2H), 6.29 (s, 1H), 5.77 (t, *J* = 8.0 Hz, 1H), 5.67 (d, *J* = 8.2 Hz, 1H), 5.41 (dd, *J* = 19.0, 6.2 Hz, 1H), 4.89 (s, 1H), 4.65 (dd, *J* = 38.2, 17.6 Hz, 4H), 4.18 (dd, *J* = 10.5, 5.8 Hz, 2H), 4.00 (s, 2H), 3.79 (dd, *J* = 21.2, 11.1 Hz, 3H), 3.69 (d, *J* = 14.0 Hz, 3H), 2.45 (q, *J* = 7.1 Hz, 6H), 1.87 (s, 6H), 0.94 (t, *J* = 7.1 Hz, 9H). HRMS (ESI): calc'd for [C$_{32}$H$_{40}$N$_{11}$O$_{16}$P$_2$]$^+$ = 896.2051; found 896.2121.

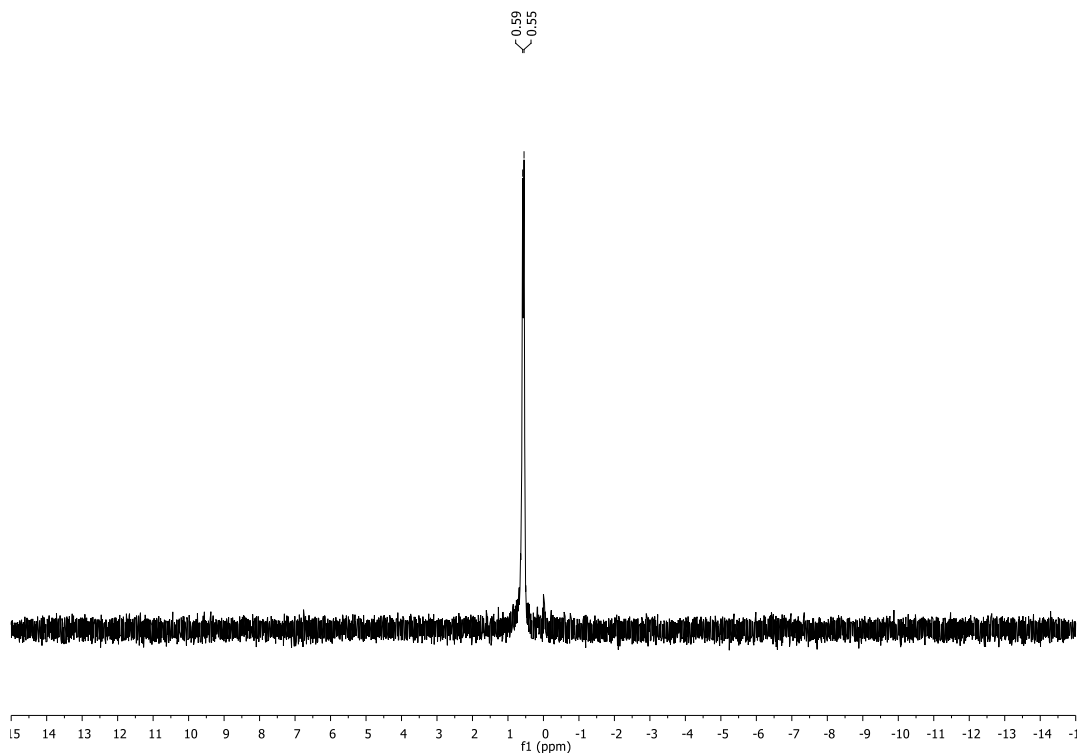**S 2.20 HPLC trace of reaction mixture and purified product**
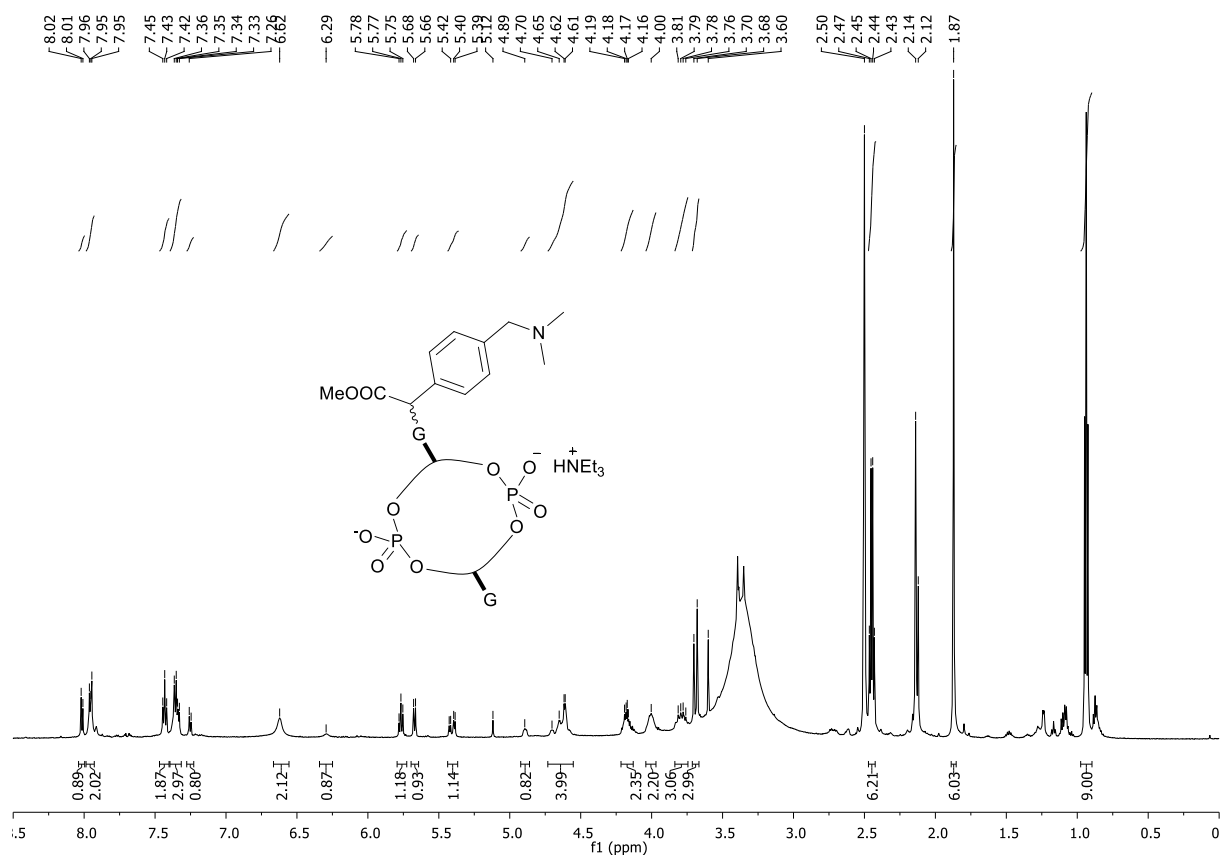
S 2.21 ¹HNMR in DMSO 600MHz
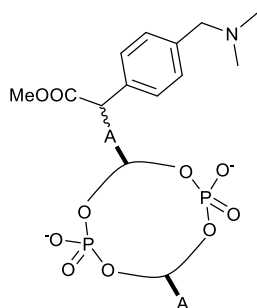
20 μL reaction mixtures containing 2 mM c-di-GMP, 200 μM Rh$_2$(OAc)$_4$ and 20 mM α-diazocarbonyl compound in 100 mM MES buffer, pH 6.0 were kept at room temperature for 2h. The reaction was monitored by analytical HPLC (conv. 33%). HRMS (ESI): calc'd for [C$_{32}$H$_{40}$N$_{11}$O$_{14}$P$_2$·Et$_3$N]$^+$ = 965.3358; found 965.3430.
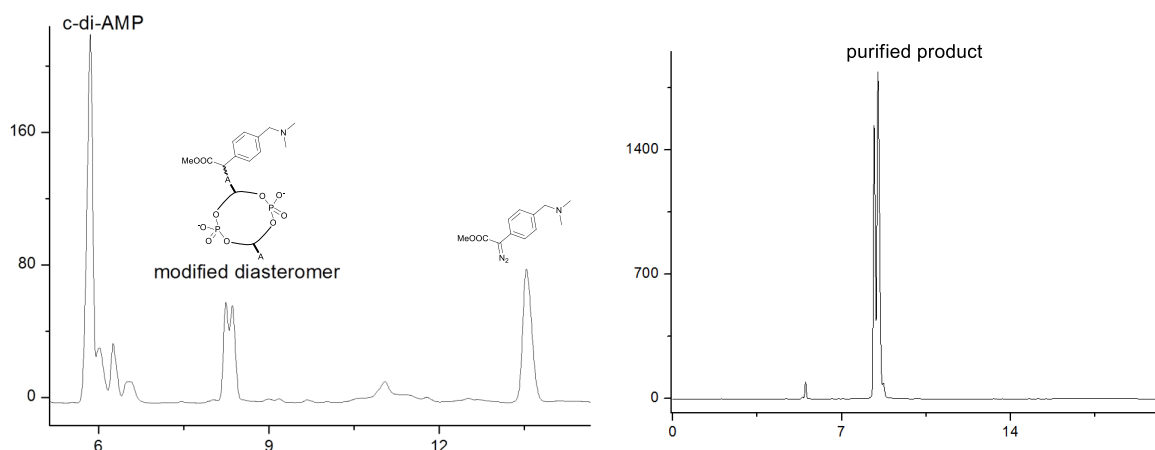
S 2.22 HPLC trace of reaction mixture and purified product
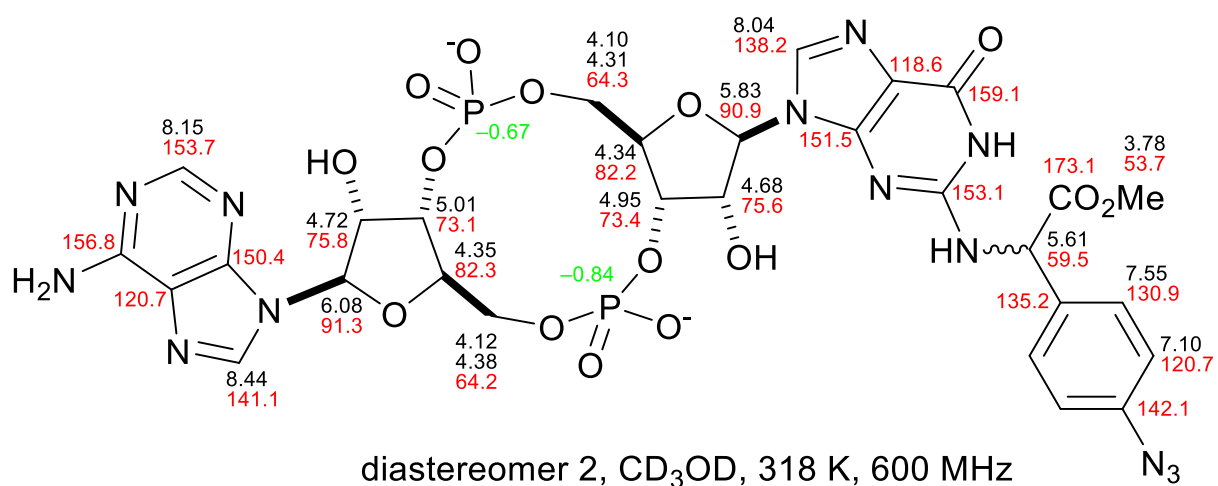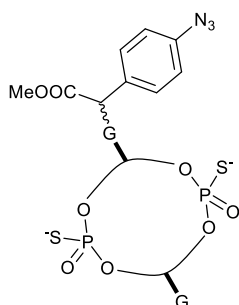

2.4 mL reaction mixtures containing 2 mM c-di-GMPS, 200 μM Rh$_2$(OAc)$_4$ and 20 mM aryl azide diazocarbonyl compound in 50% water and 50% *t*-BuOH were kept at room temperature for 50 min. The reaction was monitored by analytical HPLC (80%) and the product isolated by semi-preparative HPLC (41% yield). The pure diastereomer was separated by HPLC, however it epimerized after adding DMSO-d$_6$. ^{1}H NMR (600 MHz, DMSO) δ 7.98 (d, *J* = 3.1 Hz, 1H), 7.81 (dd, *J* = 33.5, 7.7 Hz, 1H), 7.65 – 7.57 (m, 1H), 7.51 (d, *J* = 8.8 Hz, 2H), 7.34 (s, 1H), 7.19 (d, *J* = 8.4 Hz, 2H), 6.99 (s, 1H), 6.83 (s, 1H), 6.71 (d, *J* = 22.8 Hz, 2H), 5.75 (d, *J* = 6.1 Hz, 1H), 5.65 (s, 1H), 5.47 (s, 1H), 5.40 (s, 1H), 4.80 (s, 1H), 4.72 (s, 1H), 4.29 (s, 2H), 4.10 (s, 2H), 3.94 (dd, *J* = 24.2, 12.7 Hz, 2H), 3.69 (d, *J* = 18.7 Hz, 3H). HRMS (ESI): calc'd for [C$_{29}$H$_{31}$N$_{13}$O$_{16}$P$_2$·2Et$_3$N]$^+$ = 1082.3896; found 1082.3973.
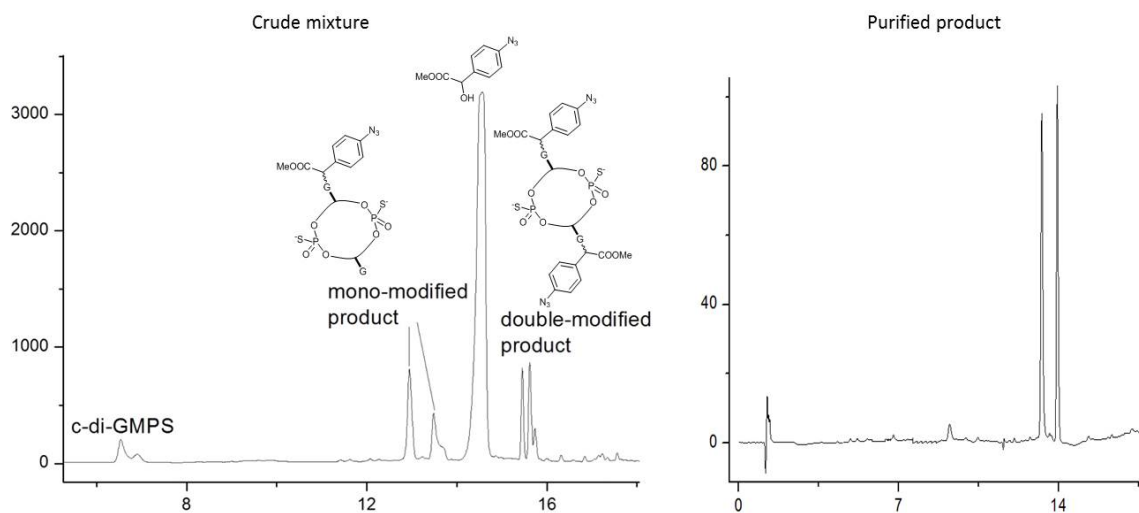
S 2.23 HPLC trace of reaction mixture and pure product
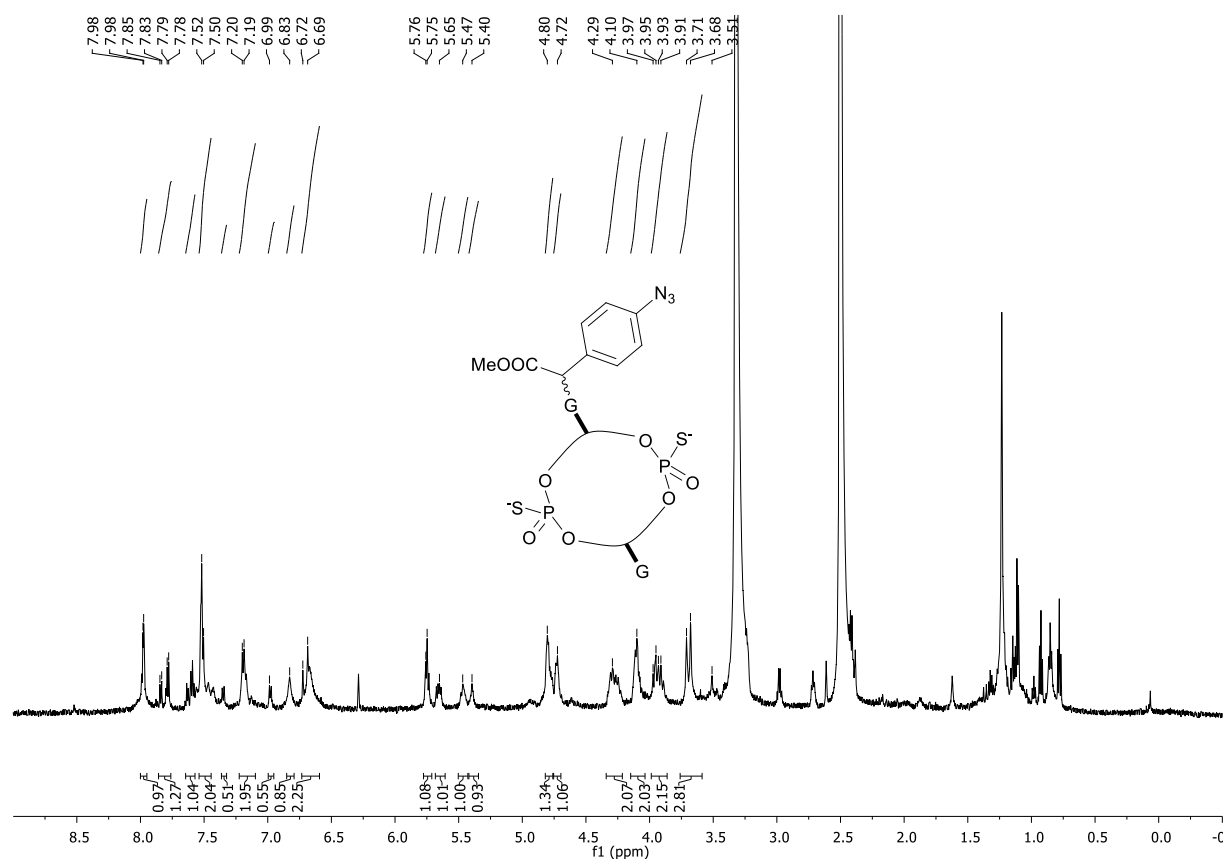
S 2.24 ^{1}H NMR in DMSO 600MHz
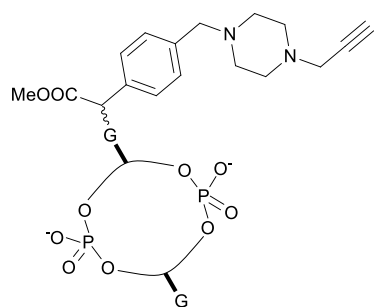
20 μL reaction mixtures containing 2 mM c-di-GMP, 200 μM Rh$_2$(OAc)$_4$ and 20 mM α-diazocarbonyl compound in 100 mM MES buffer, pH 6.0 were kept at room temperature for 2h. The reaction was traced by analytic HPLC (conv. 30%). HRMS (ESI): calc'd for [C$_{37}$H$_{45}$N$_{12}$O$_{16}$P$_2$]$^+$ = 975.2473; found 975.2488.
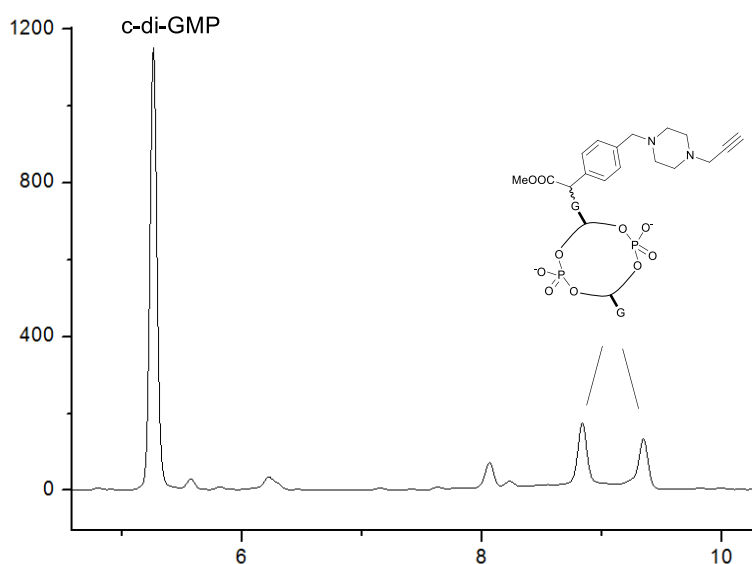
S 2.25 HPLC trace of reaction mixture

Click reaction to synthesize fluorescent tag on c-di-GMP
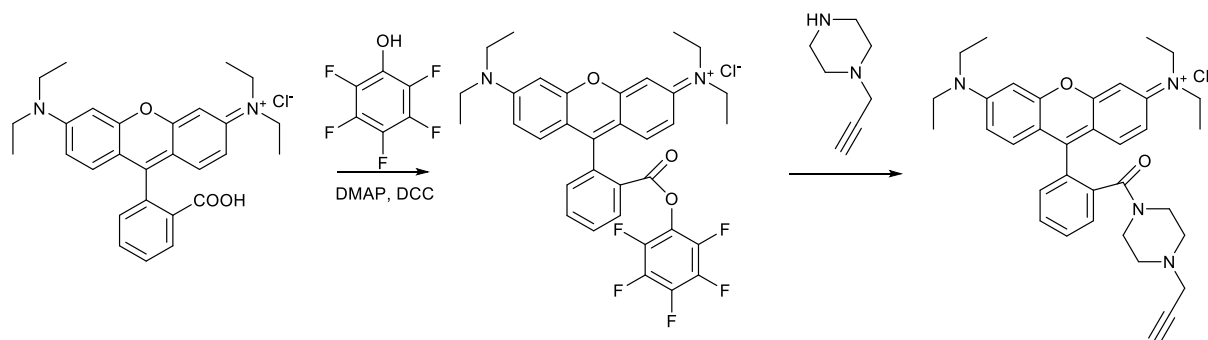
S 2.26 preparation of fluorescent tag

Rhodamine B (0.600 g, 1.35 mmol), 2,3,4,5,6-pentafluorophenol (0.250 g, 1.36 mmol) and 4-(N,N-dimethylamino)pyridine (0.021 g, 0.17 mmol) were dissolved in 20 ml of dry DCM under nitrogen and N,N'-dicyclohexylcarbodiimide (0.327 g, 1.59 mmol) as a solution in 2 ml of DCM was added with continuous stirring. The mixture was stirred at room temperature until complete as judged by TLC (2 h). It was then evaporated and the residue purified by column chromatography on Si60 in DCM/methanol to afford 0.46 g target product as a dark purple solid (55%). The obtained rhodamine-pentafluoroester (0.092 g, 0.14 mmol) was dissolved in 11 ml of dry acetonitrile and N-propargyl piperazine (0.018 g, 0.17 mmol) as a solution in 2.5ml of acetonitrile was added with continuous stirring under nitrogen. The mixture was stirred for 24 h at room temperature, then evaporated under vacuum and the residue purified by column chromatography on Si60 in DCM/methanol (20:1 to 10:1) to afford a deep purplish-violet solid 0.4 g (40%). ^{1}H NMR (400 MHz, CDCl$_3$) δ 7.69 – 7.60 (m, 2H), 7.53 (dd, J = 6.7, 2.1 Hz, 1H), 7.34 – 7.29 (m, 1H), 7.21 (d, J = 9.5 Hz, 2H), 6.91 (dd, J = 9.5, 2.2 Hz, 2H), 6.77 (d, J = 2.4 Hz, 2H), 3.60 (dt, J = 11.9, 7.3 Hz, 8H), 3.37 (d, J = 37.7 Hz, 4H), 3.25 (d, J = 2.3 Hz, 2H), 2.68 (s, 2H), 2.36 (s, 2H), 2.23 (s, 1H), 1.29 (t, J = 7.1 Hz, 13H). ^{13}C NMR (101 MHz, CDCl$_3$) δ 167.33, 157.65, 155.60, 155.54, 135.30, 131.99, 130.35, 130.10, 129.82, 127.56, 113.97, 113.61, 96.29, 77.32, 77.00, 76.68, 73.96, 51.42, 50.87, 47.16, 46.46, 46.01, 41.36, 12.52. HRMS (ESI): calc'd for [C$_{35}$H$_{41}$N$_4$O$_2$]$^{+}$ = 549.3224; found 549.3225.
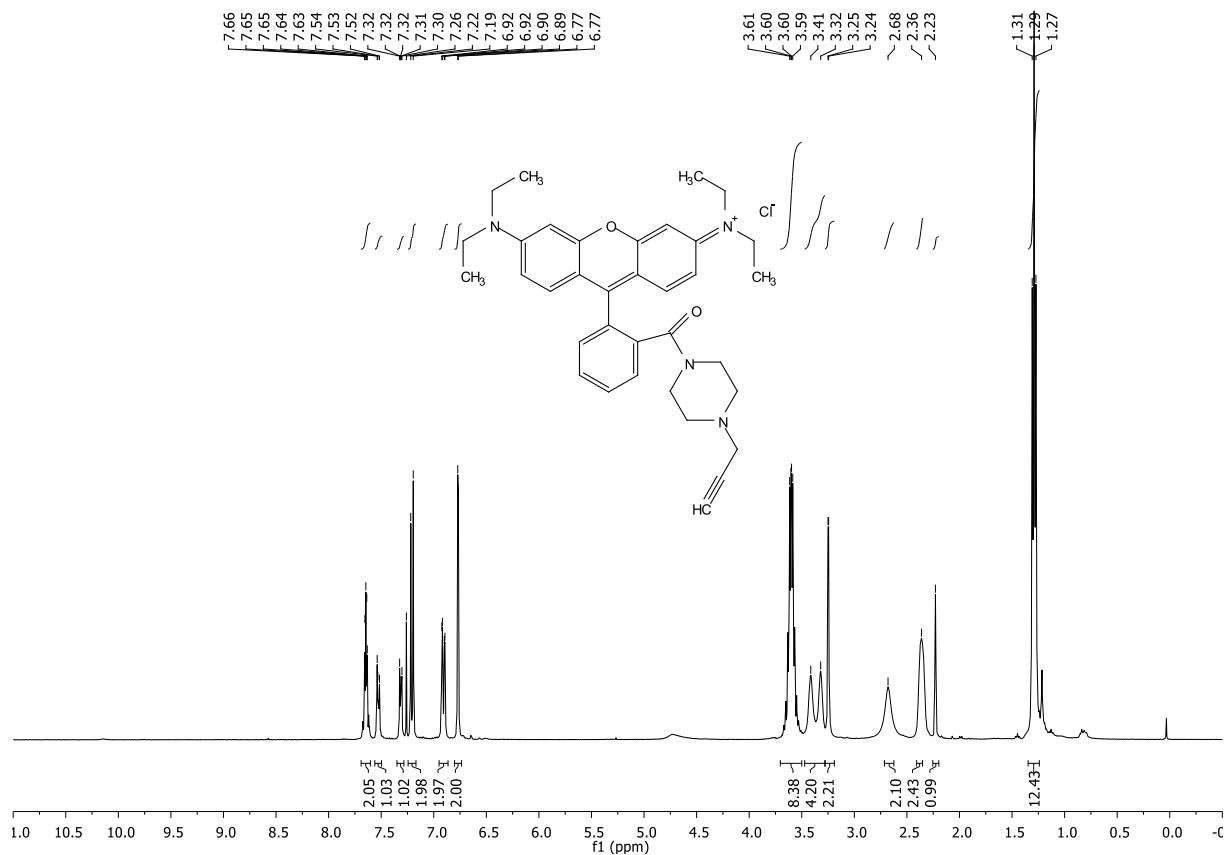
S 2.27 ^{1}HNMR in CDCl$_3$ 400MHz

167.33
157.65
155.60
155.54
135.30
131.99
130.35
130.10
129.82
127.56
113.97
113.61
96.29
77.32
77.00
76.68
73.96
51.42
50.87
47.16
46.46
46.01
41.36
12.52
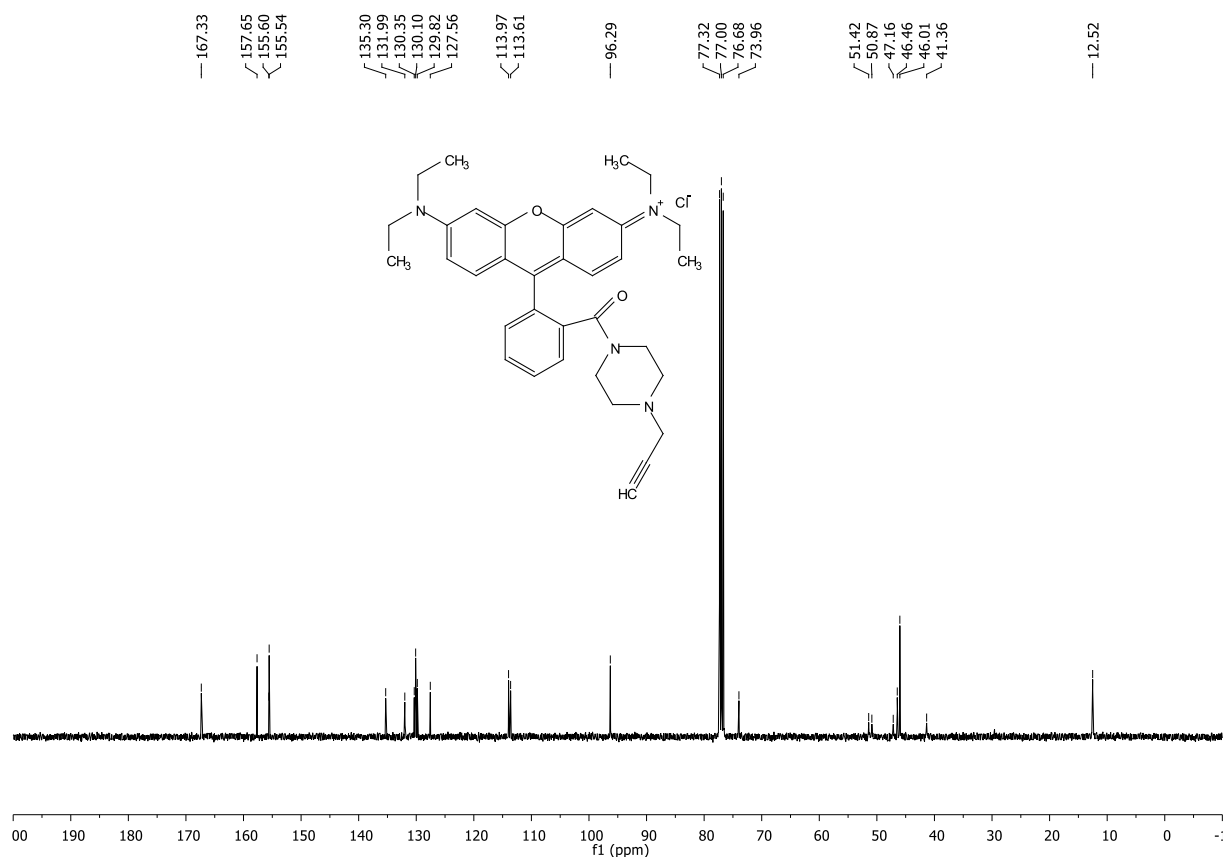
S 2.28 ^{13}CNMR in CDCl$_3$ 400MHz

A reaction mixture of 0.4 mM c-di-GMP-N$_3$, 4 mM Rhodamine alkyne, 0.2 mM CuSO$_4$ and 0.16 mM sodium ascorbate was left over night and check by HPLC and the product confirmed by HRMS (ESI): Calc'd for [C$_{64}$H$_{72}$N$_{17}$O$_{18}$P$_2$]$^+$ = 1428.4711; found 1428.4726.
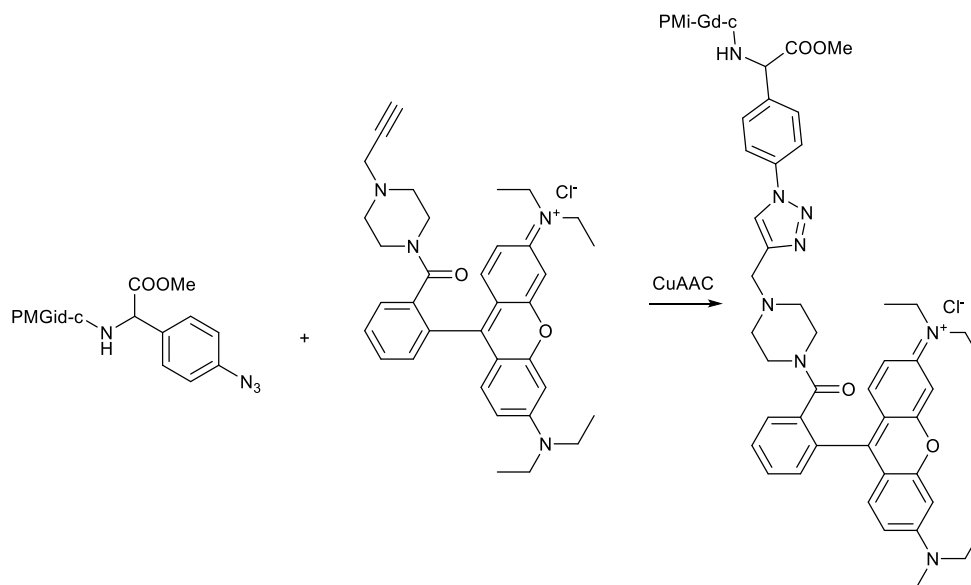
S 2.29 Click reaction of fluorescent c-di-GMP
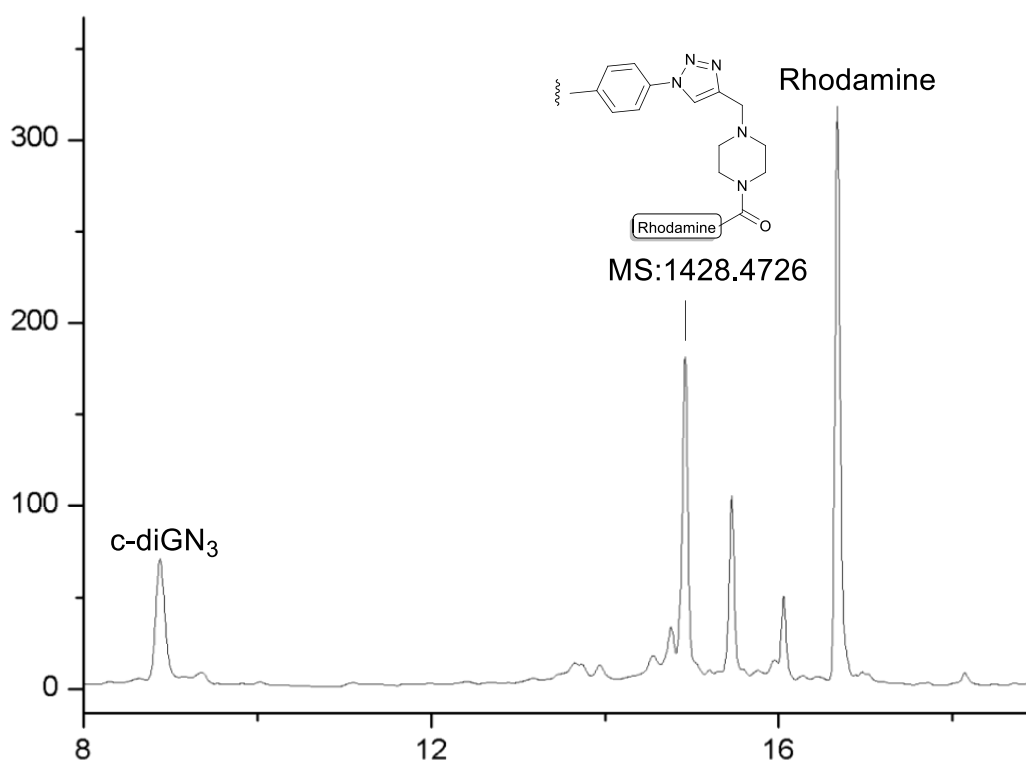
S 3.30 HPLC trace of click reaction of fluorescent c-di-GMP

2.4.3 Protein modification

The production of DgrA and its mutant have been reported previously.[102] DgrA was subcloned into pET42b vector to include an N-terminal His6 tag for purification of the protein. The construct was then transformed into Escherichia coli BL21 (DE3). Cells were grown in LB medium at 37 °C to an OD600 of approximately 0.6–0.9. Protein expression was induced by adding IPTG to a final concentration of 0.5 mM and left to grow for another 2-3 hours. Cells were collected by centrifugation at 6,500 r.p.m. and resuspended in 15 ml of lysis buffer (20 mM PBS, pH 7.0) per liter of culture. Resuspended cells were then lysed by sonication, and cell debris was removed by centrifugation at 22,000 r.p.m. 0.7 mL NTA solution was added to the clear lysis solution and stirred for half an hour then loading onto a 5-mL HisTrap column (GE Healthcare). Impurities were washed away with 20 mM imidazole in lysis buffer, and target protein was eluted with 250 mM imidazole. The peak fractions were checked by SDS-PAGE, and fractions containing the target protein were pooled together and dialysis in 10 mM PBS buffer overnight.

Sequence of DgrA

WT:

MVMVETSGAERRAHPRMPAARKIYIVDDPRSWKASLLDVAEKGGRISIAGIASPPDTFVFVDAGGRRVHLANVVW
RSGTEVGVQFAATQRIGPRAGGAAGALEIARRFLATLPAEDDALEHHHHHHHH

W75A:

MVMVETSGAERRAHPRMPAARKIYIVDDPRSWKASLLDVAEKGGRISIAGIASPPDTFVFVDAGGRRVHLANVVA
RSGTEVGVQFAATQRIGPRAGGAAGALEIARRFLATLPAEDDALEHHHHHHHH

R11A/R12A:

MVMVETSGAEAAAHPRMPAARKIYIVDDPRSWKASLLDVAEKGGRISIAGIASPPDTFVFVDAGGRRVHLANVVW
RSGTEVGVQFAATQRIGPRAGGAAGALEIARRFLATLPAEDDAHHHHHH

DgrA modification

The reaction was carried out by mixing protein and c-diGMP-N3 together in following step and the protein concentration was kept at 115 µM. The reaction was put under UV light 366 nm for 15 hours and later the sample was applied to 18% SDS gel and HR-ESI and MALDI-TOF analysis.

1: 2 µL691 MM DgrA-wt and 10 µL water

2: 1.5 µL 1 mM c-diGMP-N_3 was added to 2 µL 691 µM DgrA-wt and 8.5 µL water protein

3: 1.2 µL 1 mM c-diGMP-N_3, 8 µL 148 µM DgrA-R11/R12A protein and 1 µL water

4: 8 µL 148 µM DgrA-R11A/R12A protein and 2.2 µL water

5: 1.6 µL 1 mM c-diGMP-N_3, 6 µL 270 µM DgrA-W75A protein and 6.4 µL water

6: 6 µL 270 µM DgrA-W75A protein and 7.6 µL water

7: 1.5 µL 10 mM c-diGMP-N_3 was added 2 µL 691 µM DgrA-wt protein 8.5 µL water

Modification of DgrA-wt
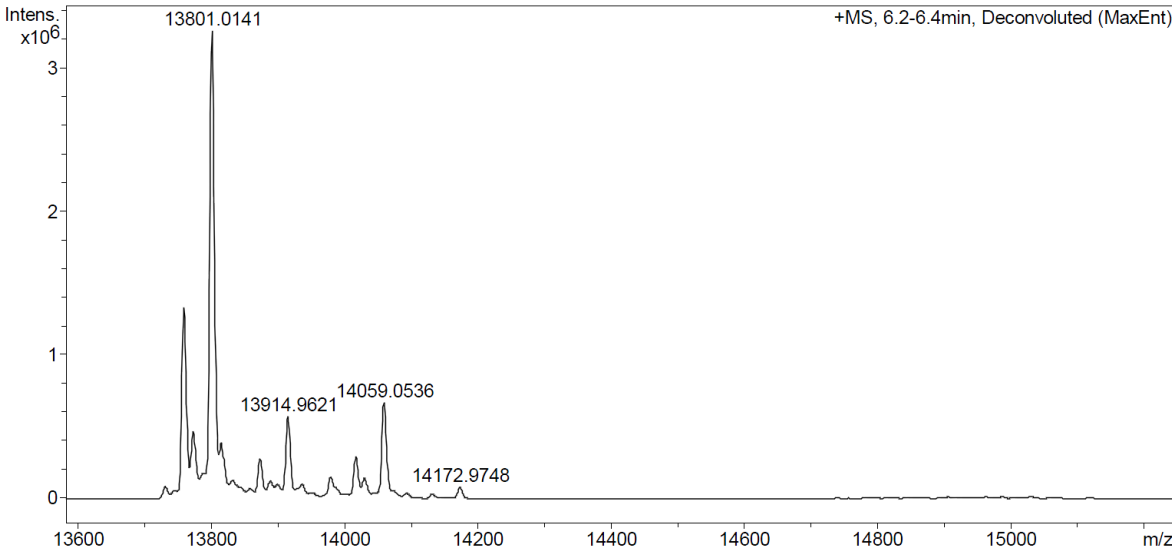
S 3.31 HR-ESI of DgrA-wt after UV irradiation (reaction 1)
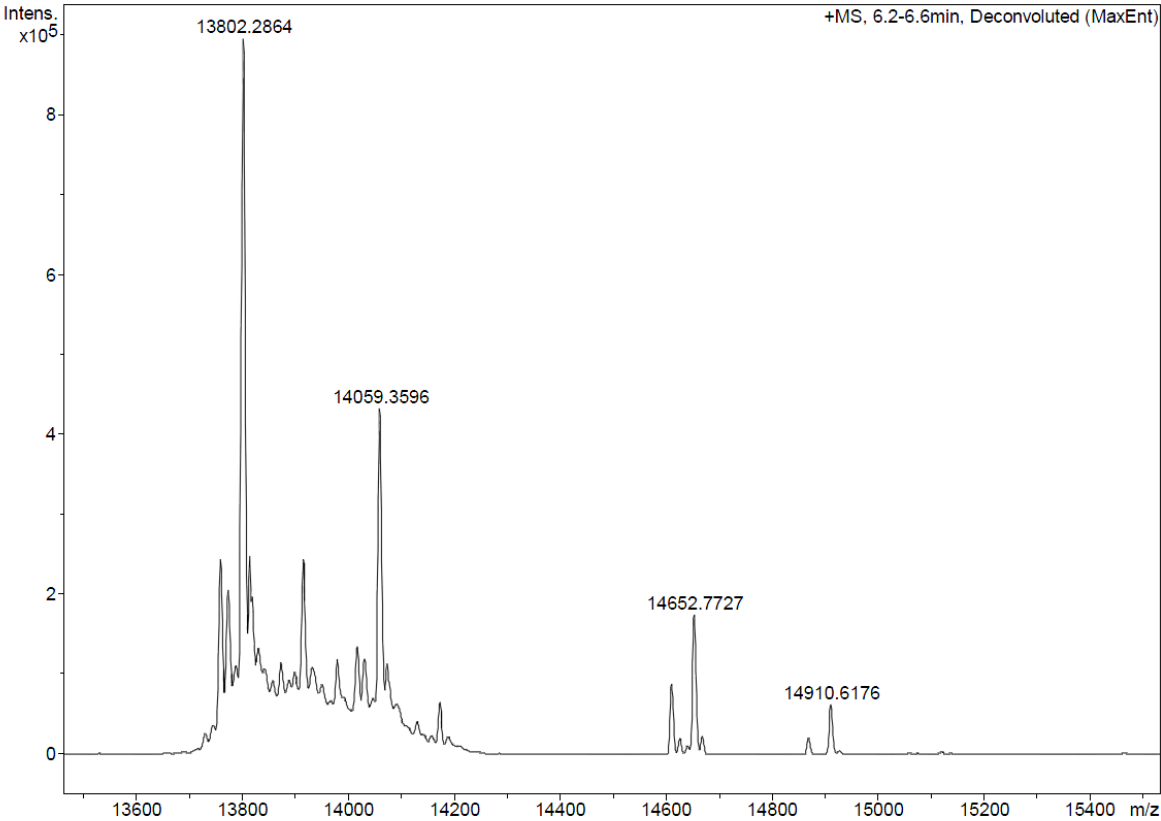
S 2.32 HR-ESI of modified DgrA after UV irradiation (reaction 2)
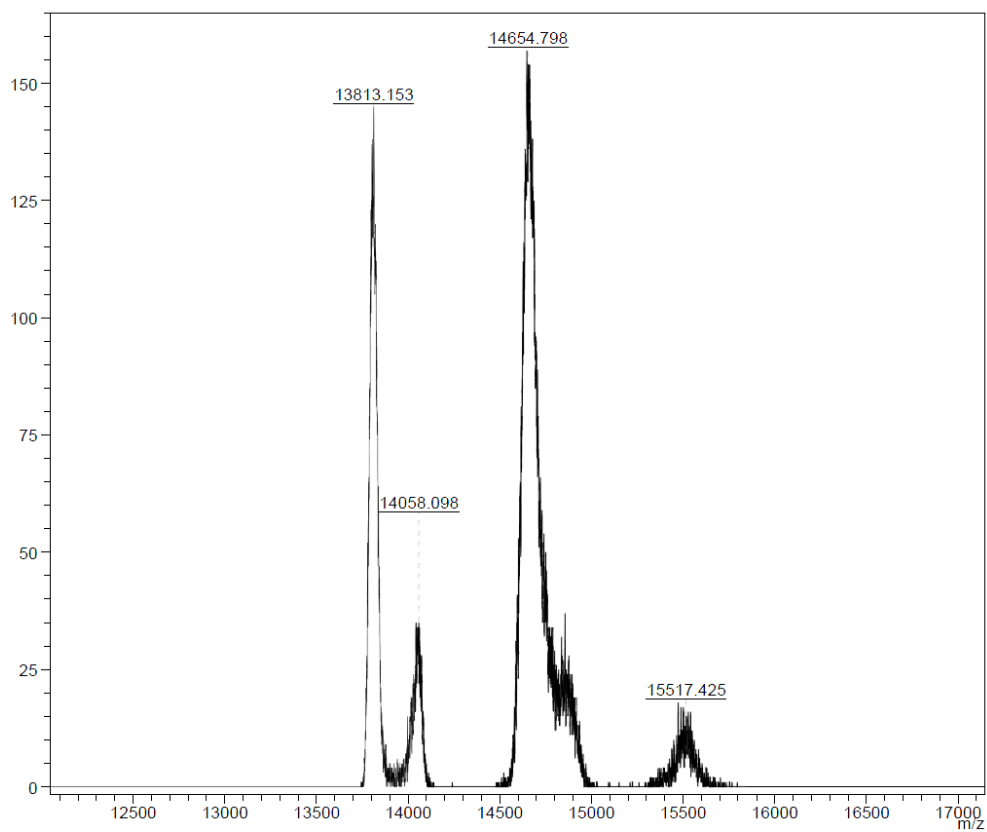
S 2.33 MALDI-TOF of modification of DgrA (reaction 2)
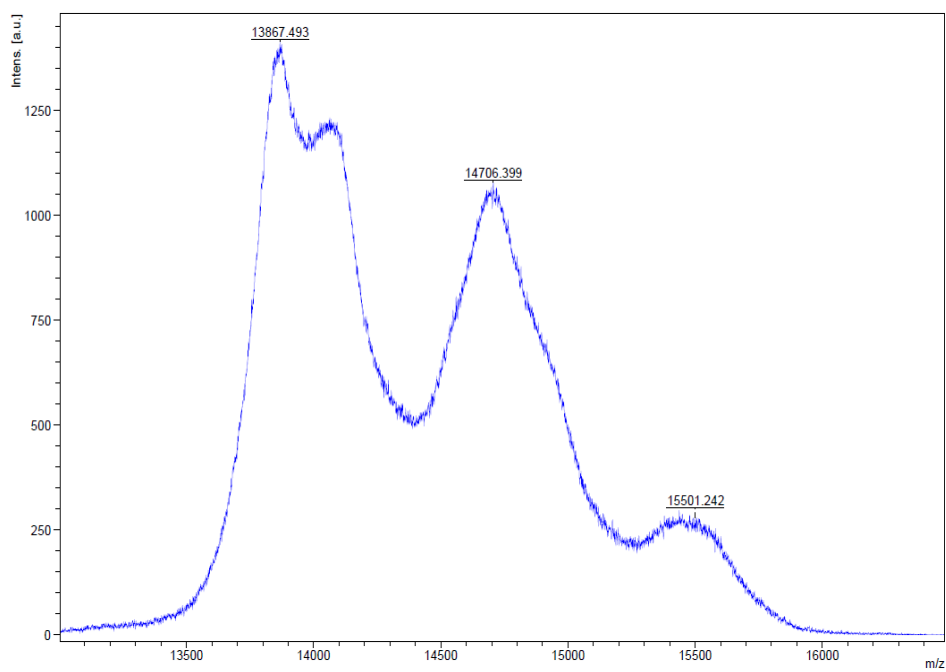
S 2.34 HR-ESI of modification DgrA-wt with 10 equivalents of c-di-GMP-N$_3$ (reaction 7)

Modified MS of mutant dgrA-W75 and DgrA-R11A/R12A was not observed in HR-ESI or in MALDI-TOF when treated with 1 equiv. of c-di-GMP-N3. Even when c-di-GMP-N_3 was increased to 10 equiv. the MS spectrum showed only broad indistinct peaks due to the high concentration of c-di-GMP-N_3. These observations are consistent with the gel data suggesting low or no modification with these mutants and with the low binding constant measured in the microscale thermophoresis.

Trypsin digestion for DgrA-wt modification (digestion of reaction 2)

	final concentration	stock solution	volume
protein	107 μM	115 μM	15 μL
Trypsin	2.1 μM	85 μM	0.4 μL
CaCl$_2$	4.2 μM	100 μM	0.7 μL

The trypsin digestion reaction mixture was left at 37 °C for 3-4 h and quenched by 10% formic acid and then injected to UPLC-MS to analyze the digested fractions. The mass of the fragment shown corresponds to the GGR peptide fragment plus the azide cross-linker. This fragment is the only species not seen in the control digestion of DgrA. The peak at 1118.4 Da is a dehydration product of the initial fragment, which likely forms from the acidic formic acid quench.
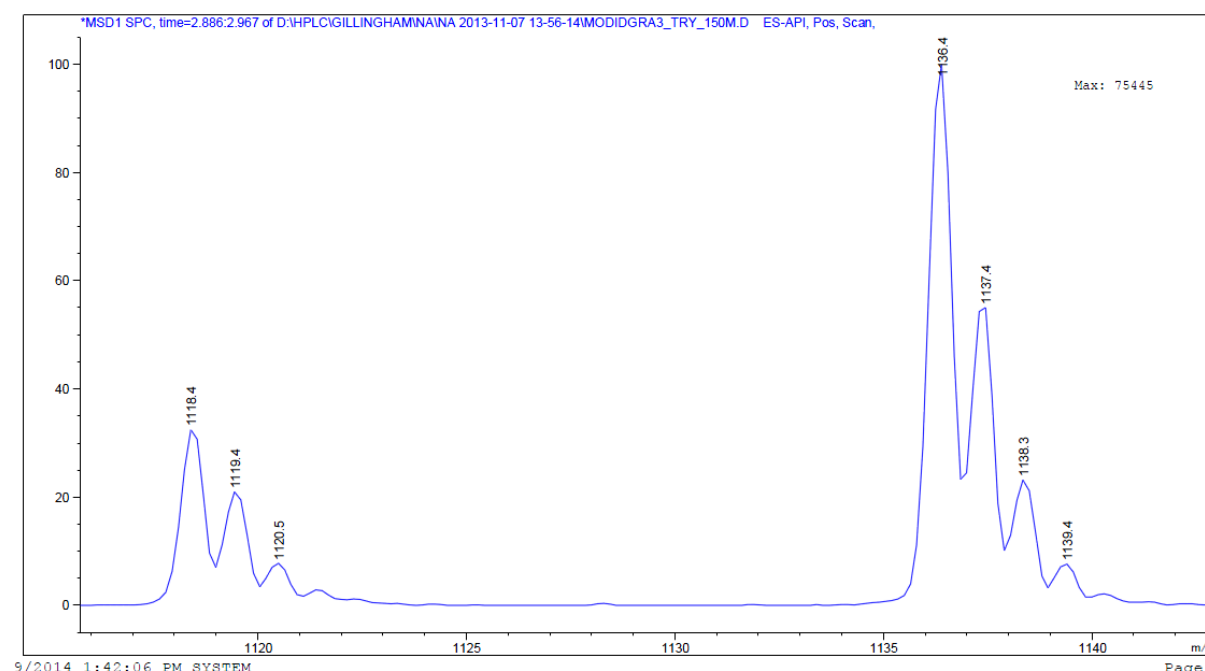
S 2.35 UPLC-MS of trypsin gigestion of modified DgrA-c-d-GMP-N$_3$

Microscale thermophoresis

DgrA and DgrA W75A were labeled with the fluorescent dye DyLight488 NHS Ester according to the manufacturer's protocol (ref. 46402, ThermoScientific, Rockford, IL-USA), with a degree of labeling of 80% and 33%, respectively. The excess of free dye was removed by dialysis. DyLight488-labelled protein concentration was kept constant at 50 nM for DgrA and 100 nM for DgrA W75A, while c-di-GMP and c-di-GMP-N_3 were titrated in PBS-T (PBS, 0.1% v/v Tween-20, pH 7.4). The samples were loaded into Monolith^{TM} standard-treated capillaries and thermophoresis was measured at 25 °C after 15 and 30 min incubation using a Monolith NT.115 instrument (NanoTemper Technologies, München, Germany). Laser power was set to 100% using 30 seconds on-time. The LED power was set to 100%. The K_d values were fitted from triplicates with Prism 6.00 (GraphPad Software) and expressed as mean ± SD.
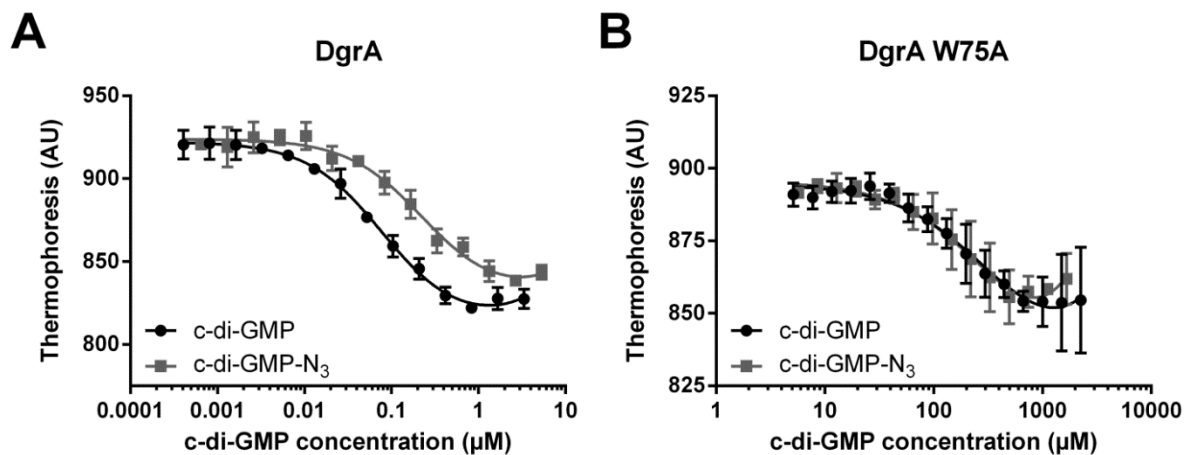
S 2.36 MST spectrum of modified c-di-GMP-N_3 and c-di-GMP

Microscale thermophoresis of the DyLight488-labelled DgrA (A) and DgrA W75A (B) shows binding to c-di-GMP (black circles) and c-di-GMP-N_3 (grey squares). The thermophoresis change upon addition of c-di-GMP is plotted as normalized fluorescence. DgrA and DgrA W75A showed a decreased fluorescence in the bound state. The Kd of DgrA was fitted to 79 ± 12 nM for c-di-GMP and 231 ± 48 nM for c-di-GMP-N_3 (n = 3). The Kd of DgrA W75A was fitted to 372 ± 190 µM for c-di-GMP and 397 ± 202 µM for c-di-GMP-N_3 (n = 3). The error bars represent SD (n = 3). The Kd of DgrA was fitted to 79 nM for c-di-GMP and 231 nM for c-di-GMP-N3, and the K_d of DgrA W75A was fitted to 372 µM for c-di-GMP and 397 µM for c-di-GMP-N_3 (n = 3). The aryl azide modification of c-di-GMP has a very minor effect on the binding affinity for DgrA and DgrA W75A (2.9 and 1.1 fold Kd increase respectively).

2.4.4 General procedure for Cu(I)-catalyzed hairpin oligonucleotide modification using α-diazocarbonyl compounds.

Solid-phase oligonucleotide synthesis was carried out on 1 μmol CPG columns using standard phosphoramidite chemistry with 0.3 M 5-benzylthio-1-H-tetrazole as activator. The DNA oligonucleotides were cleaved from the support with 32 % (v/v) aqueous ammonia for 2 h at room temperature and deprotected for 18 h at 55 °C, then freeze-dried and purified by micropreparative HPLC with TEAA buffer pH 7.2 (Method: 5 mL/min, 0-2 min, 0% CH_3CN, 2-40 min, 70% CH_3CN).

Typically 10 or 20 μL reaction mixtures containing 5 mM oligonucleotide, 500 μM $CuSO_4$, 2.5 mM THPTA and 50 mM α-diazocarbonyl compound, and 10 mM sodium ascorbate in 100 mM MES buffer, pH 6.0 were kept at 20 °C for 48 h. Analysis of the reaction products was carried out by HPLC-separation of 5 μL aliquots of the reaction. The collected peak fractions were further confirmed by ESI and MALDI TOF. The designations of the chromatographic peak fractions indicate there are only mono-alkylated products at different position of the hairpin.

Entry	Sequence	Structure	number of alykation products	Conv.[c]
1	d(CGAACGTTT TTCGTTCG)		2	9%
2	d(CGAACGTTT TTCGTTCGA)		4	16%
3	d(ACGGAAT TCCGTTTTT CGGAATTCCG)		6	16%

[a]Reaction conditions : 5 mM oligonucleotide, 0.5 mM $CuSO_4$ 2.5 mM THPTA, 50 mM diazo substrate, 10 mM sodium ascorbate, 100 mM MES, pH 6, 48 h, room temperature. [b]Amount of the modified species as a percentage of the total amount of NA.

2.4.5 General procedure for auto-tandem catalytic CuAAC/NHI
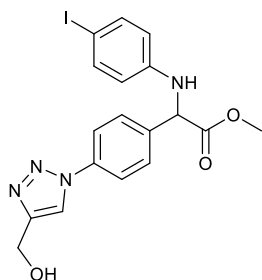
Methyl 2-(4-azidophenyl)-2-diazoacetate (14 mg, 0.064 mmol), 4-iodoaniline (15.5 mg, 0.071 mmol) and propargyl alchol (3.6 mg, 0.064 mmol) were mixed in tert-butanol (1.37 mL) under nitrogen, and 500 mM MES buffer, pH 6.0 (0.514 mL), 100 mM $CuSO_4$ in water (0.032 mL) were then added, the mixture was stirred, and 400 mM sodium ascorbate in water (0.32 mL), 0.5 mL water was introduced. The mixture was stirred for overnight at room temperature. Then an equal volume of 10 % (w/v) K_2CO_3 was added, and the mixture was extracted with 3 × 6 mL of dichloromethane. The combined organic layers were dried with anhydrous sodium sulfate, evaporated under vacuum and the residue purified by column chromatography on Si60 in dichloromethane/methanol to yield the desired product as a white solid (70 %). TLC (DCM-methanol 30:1 v/v) R_f =0.23. IR (film) 3368, 3151, 1734, 1497, 1184, 1059, 1005, 928, 749 cm^{-1}. ^1H NMR (500 MHz, $CDCl_3$) δ/ppm: 8.18 (s, 1H), 7.80 (d, J = 8.7 Hz, 2H), 7.71 (dd, J = 7.8, 1.5 Hz, 1H), 7.66 (d, J = 8.4 Hz, 2H), 7.12 – 7.05 (m, 1H), 6.49 – 6.37 (m, 2H), 5.68 (d, J = 5.8 Hz, 1H), 5.35 (d, J = 5.8 Hz, 1H), 4.69 (d, J = 5.5 Hz, 2H), 3.73 (s, 3H), 3.30 (t, J = 5.7 Hz, 1H). ^{13}C NMR (126 MHz, $CDCl_3$) δ/ppm: 172.28, 149.92, 145.93, 140.17, 138.84, 138.15, 130.41, 129.67, 121.86, 121.62, 120.66, 112.97, 85.99, 60.68, 56.52, 53.87. HRMS (ESI) calc. for $C_{18}H_{18}IN_4O_3^+$: $[M + H]^+$, 465.0345; Found: $[M + H]^+$, 465.0418.
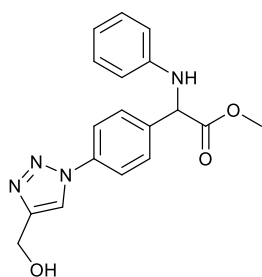
Methyl 2-(4-azidophenyl)-2-diazoacetate (14 mg, 0.064 mmol), aniline (6.57 mg, 0.071 mmol) and propargyl alchol (3.6 mg, 0.064 mmol) were mixed in tert-butanol (1.37 mL) under nitrogen, and 500 mM MES buffer, pH 6.0 (0.514 mL), 100 mM $CuSO_4$ in water (0.032 mL) were then added, the mixture was stirred, and 400 mM sodium ascorbate in water (0.32 mL), 0.5 mL water was introduced. The mixture was stirred for overnight at room temperature. Then an equal volume of 10 % (w/v) K_2CO_3 was

added, and the mixture was extracted with 3 × 6 mL of dichloromethane. The combined organic layers were dried with anhydrous sodium sulfate, evaporated under vacuum and the residue purified by column chromatography on Si60 in dichloromethane/methanol to yield the desired product as white solid, (70 %) TLC (DCM-methanol 30:1 v/v) R_f =0.31. IR (film) 3279, 3119, 1731, 1516, 1191, 1011, 765, 693 cm^{-1}. 1H NMR (500 MHz, CDCl$_3$) δ/ppm: 8.18 (s, 1H), 8.19 (s, 1H), 7.80 (d, J = 8.7 Hz, 2H), 7.68 (d, J = 8.5 Hz, 2H), 7.11 (dd, J = 8.5, 7.5 Hz, 2H), 6.65 (dt, J = 8.5, 1.1 Hz, 3H), 5.30 (d, J = 5.7 Hz, 1H), 4.70 (s, 2H), 3.70 (s, 3H). ^{13}C NMR (126 MHz, CDCl$_3$) δ/ppm: δ 172.74, 149.89, 147.26, 139.58, 137.98, 130.10, 129.81, 122.04, 121.72, 121.63, 118.88, 114.42, 60.58, 56.52, 53.40. HRMS (ESI) calc. for C$_{18}$H$_{19}$N$_4$O$_3$$^+$: [M + H]$^+$, 339.1379; Found: [M + H]$^+$, 339.1452.

Methyl 2-(4-azidophenyl)-2-diazoacetate (0.100 mmol), aniline (0.110 mmol) and N,N-dimethylprop-2-yn-1-amine (0.1 mmol) were mixed in tert-butanol (2 mL) under nitrogen, and 500 mM MES buffer, pH 6.0 (0.8 mL), 100 mM CuSO$_4$ in water (0.05 mL) were then added, the mixture was stirred, and 400 mM sodium ascorbate in water (0.5 mL), 0.63 mL water was introduced. The mixture was stirred for overnight at room temperature. Then an equal volume of 10 % (w/v) K$_2$CO$_3$ was added, and the mixture was extracted with 3× 6 mL of dichloromethane. The combined organic layers were dried with anhydrous sodium sulfate, evaporated under vacuum and the residue purified by column chromatography on Si60 in dichloromethane/methanol to yield a white product white solid (63%). TLC (DCM-methanol 20:1 v/v) R_f = 0.31. IR (film) 3366, 3292, 1728, 1525, 1223, 860, 830, 515 cm^{-1}. ^{1}H NMR (500 MHz, CDCl$_3$) δ/ppm: 8.14 (s, 1H), 7.80 (d, J = 8.7 Hz, 2H), 7.67 (d, J = 8.6 Hz, 2H), 7.14 – 7.07 (m, 2H), 6.68-6.64 (m, 3H), 5.39 (d, J = 7.1 Hz, 1H), 5.29 (d, J = 7.2 Hz, 1H), 3.69 (s, 3H), 3.59 (s, 2H), 2.23 (s, 6H). ^{13}C NMR (126 MHz, CDCl$_3$) δ/ppm: 172.74, 147.26, 146.66, 139.48, 138.03, 130.09, 129.79, 122.49, 121.62, 121.55, 118.87, 118.32, 114.42, 60.58, 54.77, 53.39, 45.22. HRMS (ESI) calc. for C$_{20}$H$_{24}$N$_5$O$_2$$^+$: [M + H]$^+$, 366.1852; Found: [M + H]$^+$, 366.1925
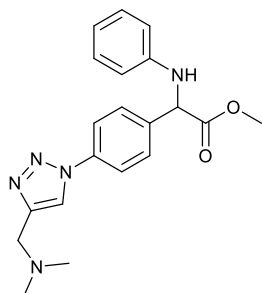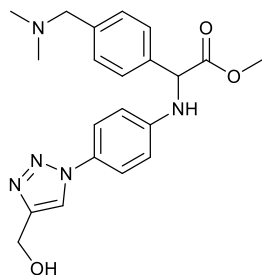
Methyl 2-diazo-2-(4-((dimethylamino)methyl)phenyl) acetate (30 mg, 0.128 mmol), aniline (20.61 mg, 0.15 mmol) and propargyl alcohol (7.2 mg, 0.128 mmol) were mixed in tert-butanol (0.8 mL) under nitrogen, and 500 mM MES buffer, pH 6.0 (0.93 mL), 100 mM $CuSO_4$ in water (0.064 mL) were then added, the mixture was stirred, and 400 mM sodium ascorbate in water (0.64 mL), 0.12mL water were introduced. The mixture was stirred for overnight at room temperature. Then an equal volume of 10 % (w/v) K_2CO_3 was added, and the mixture was extracted with 3× 6 mL of dichloromethane. The combined organic layers were dried with anhydrous sodium sulfate, evaporated under vacuum and the residue purified by column chromatography on Si60 in dichloromethane/methanol to yield the desired product as a colorless oil (62 %). TLC (DCM-methanol 10:1 v/v) R_f = 0.45. IR (film) 3366, 3292, 1728, 1525, 1223, 860, 830, 515 cm^{-1}. ^{1}H NMR (500 MHz, CDCl$_3$) δ/ppm: 8.00 (s, 1H), 7.59 (d, J = 8.2 Hz, 2H), 7.49 (dd, J = 18.1, 8.6 Hz, 4H), 6.75 (d, J = 9.0 Hz, 2H), 5.30 (s, 1H), 4.66 (s, 2H), 4.18 (s, 2H), 3.70 (s, 3H), 2.71 (s, 6H). ^{13}C NMR (126 MHz, CDCl$_3$) δ/ppm: 172.43, 147.64, 140.31, 132.44, 131.29, 129.21, 129.09, 122.98, 118.33, 114.55, 60.93, 60.67, 56.56, 53.47, 42.71. HRMS (ESI) calc. for $C_{21}H_{26}N_5O_3^+$: [M + H]$^+$, 396.1957; Found: [M + H]$^+$, 396.2030.

Chapter 3 Selective alkylation of biological phosphates by unstabilized diazo compounds

3.1 phosphate is one of the most important functional groups in chemical biology

The phosphate group and its ester are present in almost all biological process. In primary metabolism, once glucose goes into the cell by the glucose transporter protein, it immediately gets phosphorylated to form glucose 6-phosphate (G 6-P). The negatively charged phosphate group prevents the diffusion of glucose outside of the membrane. It also helps glucose to be broken down into pyruvate in the glycolysis pathway yielding ATP (adenosine triphosphate) and NADH (reduced nicotinamide adenine dinucleotide). Later different phosphorylated intermediates are involved in each step of the glucolysis pathway. These phospho-metabolites are also very important substrates, starting materials or cofactors in many other biological pathways (Figure 3.1). For example, the intermediate dihydroxyacetone phosphate (DHAP) is the source of the glycerol which can couple with fatty acids to form body fat. Furthermore, G 6-P also supplies ribulose 5-phosphate in the pentose phosphate pathway when the ratio of $NADP^+$: NADPH increases. Later it is converted to phosphoribosyl pyrophosphate (PRPP) by diphosphokinase for nucleotide and nucleic acid synthesis. PRPP is also used as the ribose supplier for nicotinic acid and quinolinic acid in NAD^+ biosynthetic pathway.[103] NAD^+ is the signaling molecule and key factor in the poly ADP ribose polymerase (PARP) and Sirtuins regulation process which means NADH analogs could be used as drug for anti-aging and life span prolongation.[104] Thus the phosphate group not only forms the backbone of genetic material, but also impacts gene expression.
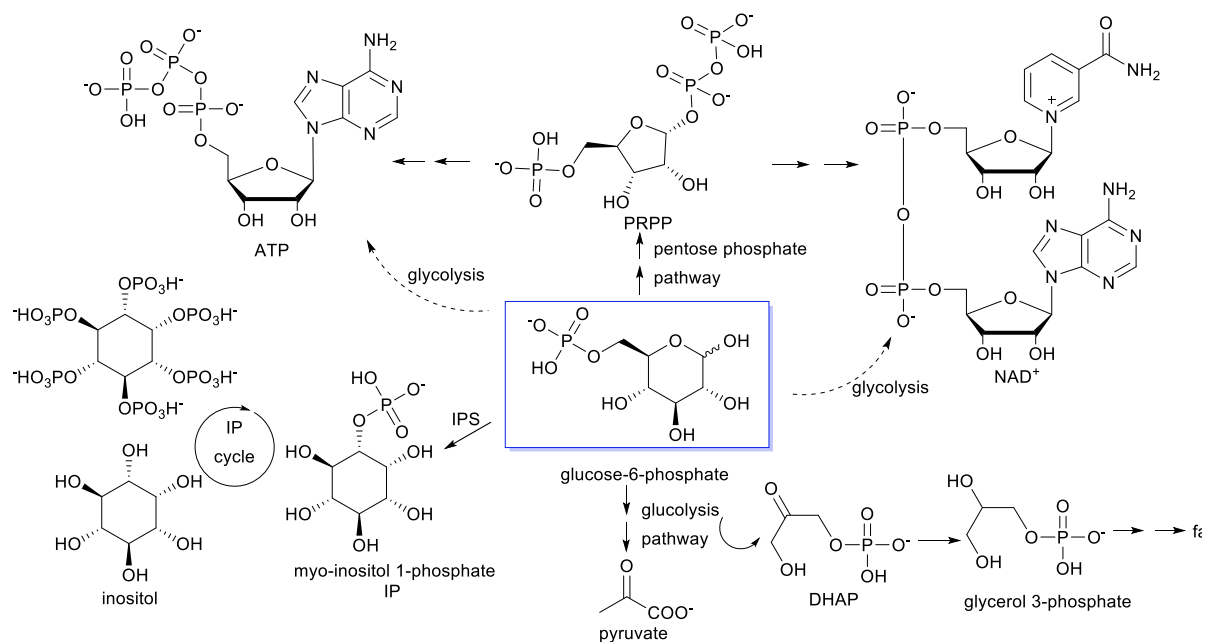
Figure 3.1 Glucose metabolic pathway, phosphorylation is present in most of its pathway.

After the discovery of inosityl hexaphosphate in green plants in 1919[105], inositol phosphates (IPs) and Phosphoinositides (PIs, inositol phospholipids) have drawn enormous attention. Inositol-3-phosphate synthase (IPS), a putative target for mood-stabilizing drugs such as lithium and valproate, converts G 6-P to *myo*-inositol 1-phosphate which later gets dephosphorylated to give inositol. Altered level of inositol affects brain signaling events leading to severe psychiatric and neurological problems like Alzheimer, suicide and stroke.[106-108] Inositol isomers could be phosphorylated at different levels leading to more than 60 different inositol phosphates. The complexity of low concentrations of IPs hinders the mapping of their functions.[109, 110] Nevertheless, more than 4000 references have contributed in the field due to their important roles in cell signaling and gene expression.[110]

Besides in the sugar metabolic pathway, phosphate also plays an important role in many other biomolecules' biosynthetic pathways (See Figure 3.1). Choline, which serves as the hydrophilic head in the cell membrane, is linked to fatty acids via a phosphate ester (Figure 3.2). Choline and its metabolites are important nutrients for humans, and they play a great role in structural integrity, cell signaling and cholinergic neurotransmission. They are rich in milk and egg as free choline, phosphocholine, glycerophosphocholine and phosphatidylcholine. Sufficient intake of choline is especially important for pregnant women since it reduces the risk of neural tube defects (spinal cord or brain defects) in infants.[111] It is also a major biological source of methyl groups through its role in S-adenosylmethionine (SAMe) synthesis.
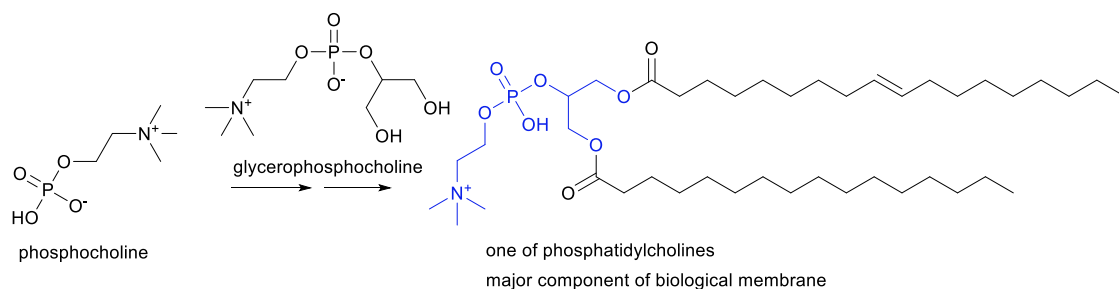
Figure 3.2 biosynthetic pathway of phosphatidylcholine

Triglyceride is the main component of body fat. Its biosynthesis starts with glycerol 3-phosphate. So does steroids biosynthesis with phosphorylated terpene. Terpene pyrophosphate, also called terpenoid serves as an important intermediate in this process. However the leaving group of these two biosynthetic pathways is different (See Figure 3.3). In the steroid biosynthetic way, pyrophosphate is the good leaving group, which can be attacked by another terpenoid. For body fat biosynthesis, the alcohol of the glycerol 3-phsphate attacks the fatty acid which is already activated by coenzyme A. After phosphatidic acid is obtained, the phosphate needs to be removed by phosphatase for the later steps. Then the free alcohol of the diacylglyceride can attack another acyl-CoA to afford triglyceride. Hence in the triglyceride pathway the phosphate monoester does not serve as a good leaving group, but rather as a special recruiting signal for glycerol's binding to different enzymes in the triglyceride biosynthetic pathway.
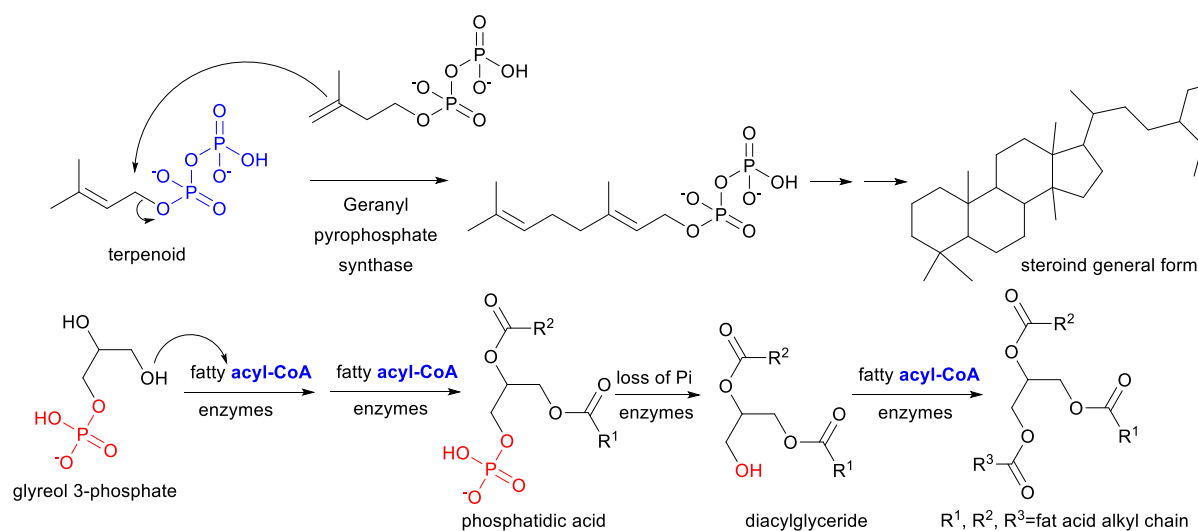
Figure 3.3 pyrophosphate as the nature leaving group

Different phosphate esters play different roles in biological process. Purine tri-phosphates are well known as the energy currency. Pyrophosphate is one of Nature's best leaving groups. Phosphodiesters form the backbone of genetic materials in all organisms. But the phosphate monoester has a role in many types of transformations in biology (Figure 3.1). Phosphate installation leads to such a dramatic change in chemical properties that it is often used as a chemical signal, for example in protein

phosphorylation. We have discussed how the phosphorylated form of glucose or glycerol is very important in interacting with enzymes to start their pathway. Their ubiquity has led many to speculate as to why nature chose phosphates as components of many of its processes.[112-114]
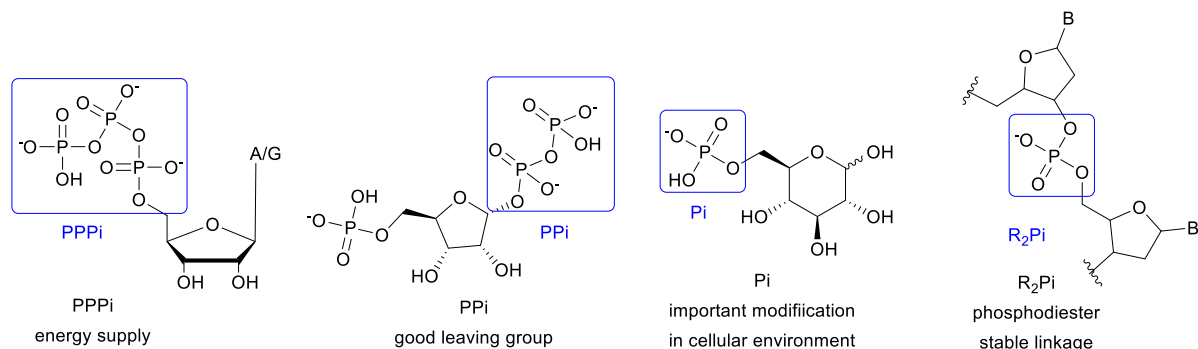
Figure 3.4 different phosphate group in chemical biology

3.1.1 Revealing the mysterious of reversible protein phosphorylation

The first example of phosphorylation was discovered in the late 1930s. During the study of sugar metabolisms, Carl and Gerty Cori discovered G 6-P as the first intermediate of the glycogen breakdown process (glycogenolysis, not glucolysis). The enzyme that catalyzes this process was named as phosphorylase since it added a phosphate group to glucose. They found there were two forms of this phosphorylase (A form and B form) which is under the regulation of another enzyme, the prosthetic-group-removing' (or PR) enzyme.[115] They believed phosphorylase a containing a covalently linked 5'-AMP (incorrect) is the active form in glycogenolysis, PR enzyme releases 5'-AMP and converts phosphorylase A to phosphorylase B. They were awarded the Nobel Prize in 1947 for discovering the course of "the catalytic conversion of glycogen". However many years passed, the Coris were stuck in a situation that they could show no formation of AMP in the reaction of phosphorylase A to B, further more they could show no AMP in phosphorylase a. The problem was then passed to Fischer and Krebs who later discovered that phosphorylase b could be converted to a in the presence of Mg-ATP by phosphorylase kinase.[116] They also proved that prosthetic-group-removing' (or PR) enzyme is actually a phosphate removing (PR) enzyme, not 5'-AMP removing enzyme. The transformation of phosphorylase A to B established the concept of reversible protein phosphorylation (Figure 3.5). At the same time, Sutherland made the unexpected observation that phosphorylase activation increased with the addition of 5'-AMP, which is a precursor of 3',5'-cyclic AMP.[117] That was why Coris logically assumed 5'-AMP linked phosphorylase is the active form in 1930s. After two decades of strenuous trial and error, the isolation and purification of protein kinase A (PKA) by Fisher and Krebs[118] revealed the cascade of phosphorylation.[119]
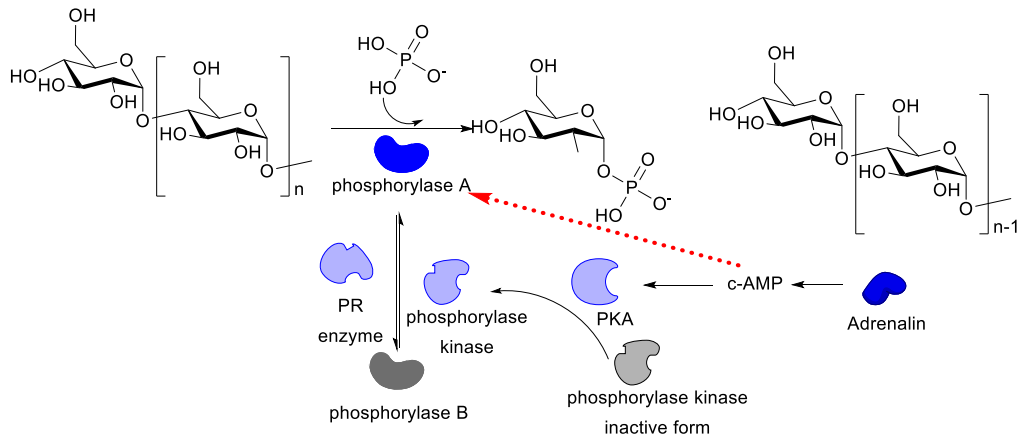
Figure 3.5 The glycogenolytic cascade[119]: Adrenalin stimulates the production of (c-AMP) leading to the sequential activation of c-AMP-dependent protein kinase and phosphorylase kinase.

PKA is known as cyclic AMP-dependent protein kinase which phosphorylates phosphorylase kinase (they called it kinase kinase when they sent the draft to the editors). Their concept of reversible protein phosphorylation in this phosphorylation cascade is embarrassingly simple according to their own humble comments, yet today this theory proves to be an absolute crucial phenomenon in every aspect of cell life. Protein kinases act like the messengers in the signal transduction pathways where one kinase activates a second kinase, which may act on yet another kinase (a kinase kinase kinase according to Fisher and Krebs). The sequential action of a series of protein kinases can transmit a signal received at the cell surface to target proteins within the cell nucleus, resulting in gene expression regulation, change of cell behavior in response to environmental stimuli. One commonly cited estimate is that approximately 30% of proteins are phosphorylated, with tens of thousands of distinct phosphorylation sites.[119] In HeLa cells, changes in 6,600 phosphorylation sites on 2,244 proteins were detected after stimulating the cell with epidermal growth factor (EGF).[120] This sounds formidable to non-experts who would like to know how exactly the reversible phosphorylation (especially the phosphorylation stoichiometries level) regulates cell life.
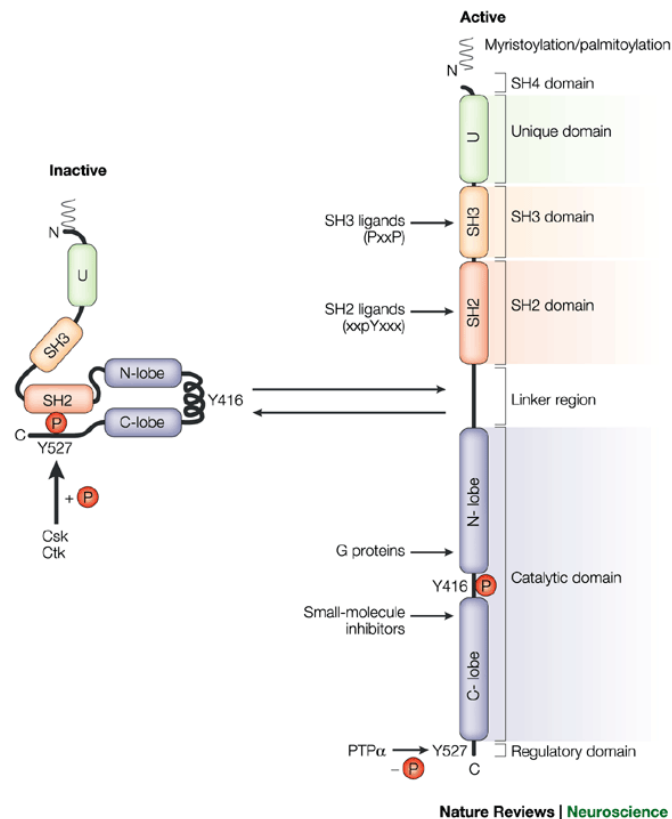
Figure 3.6 Structure and regulation of Src family kinases (SFKs). [Reprinted with permission from Macmillan Publishers Ltd: [Nat Rev Neurosci] (M. W. Salter and L. V. Kalia, Nat Rev Neurosci, 2004, 5, 317-328), copyright (2004)].

One of the oldest and most investigated protein kinase, c-Src, will be used here to exemplify the complexity of this process.[121-124] C-Src (cellular Src kinase) is a proto-oncogene tyrosine-protein kinase which means it phosphorylates specific tyrosine residues in other proteins. A different protein C-terminal Src kinase (CSK) phosphorylates c-Src at its C-terminal Y527 position, the resulting phosphorylated Y527 then folds over and binds to its own SH2 domain (Figure 3.6, left). In this form, c-Src is folded and its enzymatic activity is turned off. C-Src can be activated by many transmembrane proteins that include: adhesion receptors, receptor tyrosine kinases, G-protein coupled receptors and cytokine receptors. Once the stimuli passes down to turn on c-Src, Y527 is dephosphorylated (Figure 3.6 right) and the catalytic domain (SH1 domain, N- and C-lobes) which contains a conserved tyrosine residue (Y416), will be autophosphorylated and the enzyme opens to its active conformation and passes its signal to a downstream receptor like epidermal growth factor receptor (EGFR). Recently, it has been reported that c-Src also defines a new pool of EGFR substrates (like Mig6 and Shc1, they are signal adapters in mitogen-activated protein kinase pathway). Peptides containing a tandem motif of two tyrosines in a roll (YY) will be an optimized EGFR substrate when the second tyrosine is primed by c-Src mediated phosphorylation. Because the primed tyrosine sits in a 'priming pocket' of the EGFR kinase increasing the affinity and specificity of the EGFR kinase for binding to the tandem YY motif, the first tyrosine is subsequently efficiently phosphorylated by EGFR kinase.[125]

As the example above demonstrate, protein phosphorylation stoichiometry and the variety of diverse phosphate modification sites make characterization very difficult. Nevertheles, detection and sequencing of phosphorylated proteins (or phosphoproteome) has made a significant progress after a great number of scientists spent tremendous effort working on this problem.[126-129] Sequencing analysis could be done by tandem MS after the enrichment of phosphoprotein or phosphopeptide. Enrichment of the phosphorylated proteins could be achieved by immunoprecipitation (but highly dependant on the availability of antibodies). Large scale enrichment of phosphopeptide can be achieved by affinity chromatography purification. Immobilized metal affinity chromatography (IMAC) is the most extensively used method for global phosphopeptide enrichment.[85] It is rather "cheap and easy", but there is no selectivity compared to immunoprecipitation. The negatively charged phosphopeptide is bound on the positively charged immobilized metal ion. A review of recent developments of different metal ligands in the IMAC technique indicates the state-of-the-art.[130] IMAC's high sensitivity and selectivity for phosphopeptide over non-phosphorylated peptides have made measurement of absolute protein phosphorylation stoichiometries in human cells possible (Figure 3.7).[131] Two identical aliquots of tryptic peptides are either mock or phosphatase treated followed by isotopic tagging. Then, the mixing fraction is purified by IMAC. In the flow through of IMAC purification, the dephosphorylated peptides from phosphatase-treated aliquot (blue, light isotope labelled) will represent the total peptides, while the unphosphorylated counterpart in the untreated aliquot (red, heavy isotope labelled) will represent the fraction of initial unphosphorylated amount. The mixture of dephosphorylated and unphosphorylated peptides in the flow through is subjected to phosphorylation via a kinase reaction. Then the phosphopeptides are purified by IMAC. The ratio of heavy/light will represent the fraction of initially unphosphorylated amount and the phosphorylation stoichiometry can be calculated by formula **a** as shown in Figure 3.7a.
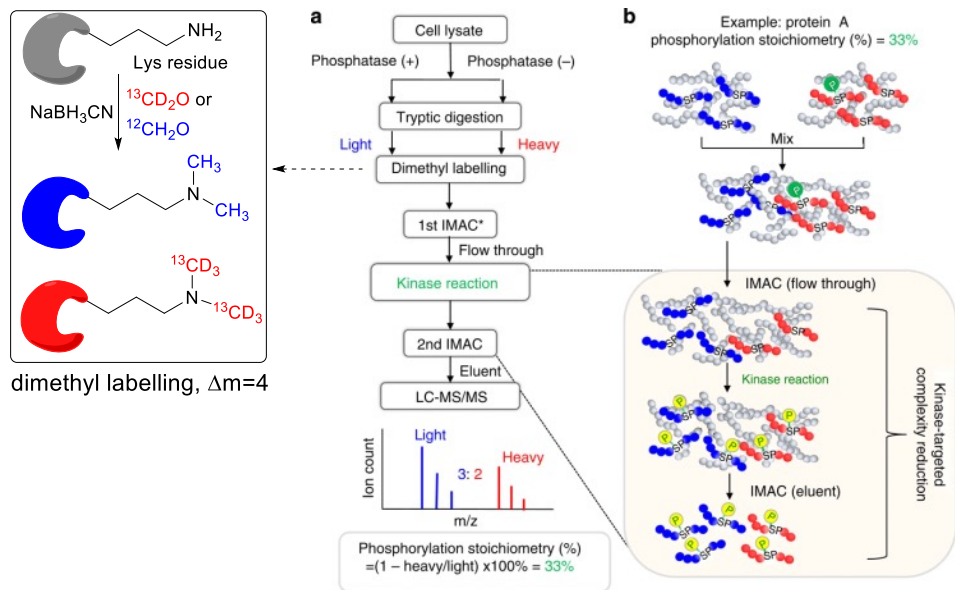
Figure 3.7 The basic principle of the motif-targeting quantitative proteomic approach. [Reprinted with permission from Macmillan Publishers Ltd: [Nat Commun] (C.-F. Tsai, Y.-T. Wang, H.-Y. Yen, C.-C. Tsou, W.-C. Ku, P.-Y. Lin, H.-Y. Chen, A. I. Nesvizhskii, Y. Ishihama and Y.-J. Chen, Nat Commun, 2015, 6) copyright (2015)]

3.1.2 Limited methods of chemical labelling of phosphopeptides

The chemical modification of the phosphoprotein or peptide has made little progress except for reports by Oda and Zhou developed a parallel of chemical derivatization strategies (Figure 3.8). Oda's protocol[132] started with a pool of protein, then serine and threonine phosphate could be eliminated to alkenes, making them susceptible to nucleophilic attack by a thiol, biotin could then be linked to these modified proteins. Zhou et al[133] started with a pool of peptides, then they first protected all amines, then carboxylic acids, then released phosphate by TFA for the cystamine attachment via carbodiimide-catalyzed condensation. Reduction of the internal disulfide of cystamine generated a free sulfhydryl group which could be captured by iodoacetyl groups immobilized on glass beads. Both methods required a series of chemical reaction steps to have an affinity tag which discourage people to try chemical modification. In comparison, IMAC is two steps and the protocol is easy to implement.
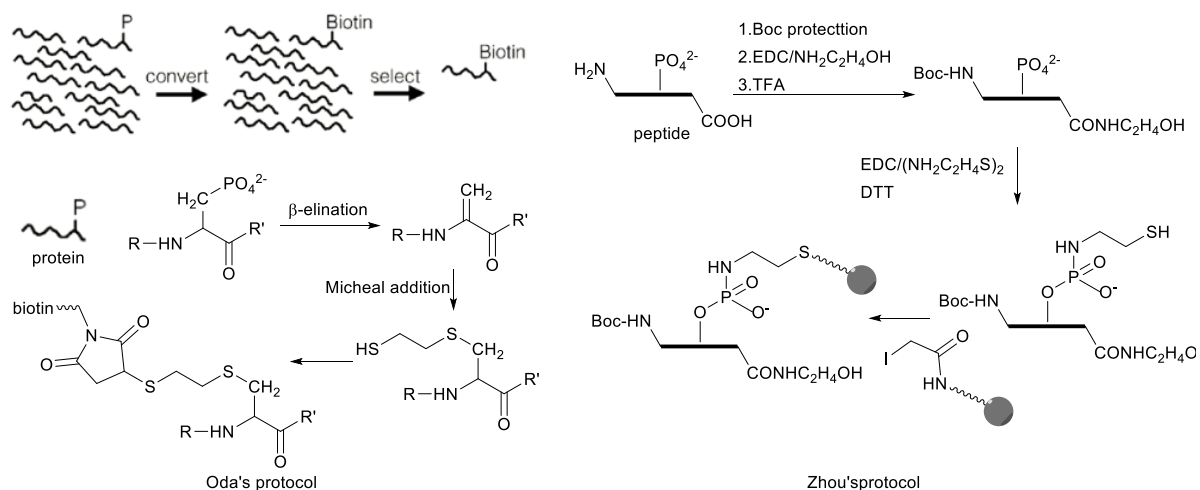
Figure 3.8 The chemical enrichment of phosphoprotein.

3.1.3 Previous phosphate modification

The phosphate group and its esters are ubiquitous in nature, serving critical roles in metabolism, nucleic acids and cell membranes. Especially protein phosphorylation is the most important post-translation modification of proteins, and yet, despite more than sixty years of study, it remains poorly understood since signal transduction networks in the phosphorylation cascade are complicated to trace. Chemical modification of phosphate biomolecules is in a urgent need due to its important role in biological process. There are a handful of cases of phosphate ester alkylation by diazo compounds as I discussed in Chapter 1. For example, nucleotide monophosphates were first modified by phenyldiazomethane in 1955.[47] Diazo-substituted coumarin has been employed to react with the phosphate moiety of the backbone of RNA, modifying approximately 30 sites on the phosphate moieties per 1 kb of RNA sequence.[52] The modification of seconD messager c-AMP and c-GMP with diazo compounds has been managed albeit in low yield (average 10-15%).[134] In most of the cases, DMSO or methanol is chosen rather than aqueous buffer for phosphate esterfication with diazo compounds since the non-ionized form of phosphoric acids woud dominante in organic solvents, whereas most of the phosphates are expected to be ionized in aqueous buffer at approximately pH 7. However no detailed look at the reactivity difference between phosphate mono- and diesters in water has been attempted.

3.2 Selective modification of phosphate in aqeous buffer

Many bioactive molecules have both carboxylate and phosphate group in presence of many other functional groups like amines or alcohols. Selectively modification of one group inn a sea of other functional groups is very difficult. Zhou and Oda's modification of phosphopeptide also showed how complicated it is to modify just the phosphate group in a peptide (Figure 3.8).[127] Despite phosphate esters and carboxylates' known reactivity with diazo compounds, no head-to-head comparison has

been investigated. In our own work on metal-catalysed DNA and RNA alkylation by diazo compounds, we sometimes found low levels of phosphate alkylation in product streams. In this chapter we take a close look at phosphate monoester alkylation and compare it to carboxylate and phosphate diester alkyation under conditions relevant to chemical biologists.

3.2.1 Competition alkylation between carboxylic acid and phosphate
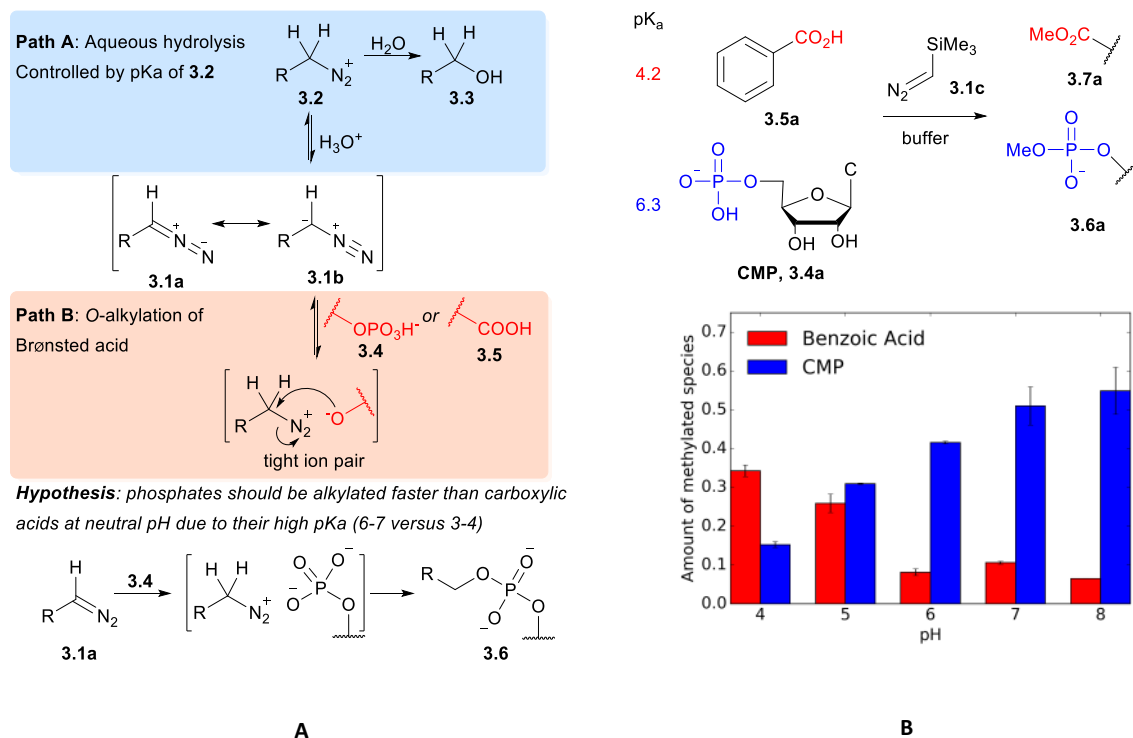
Figure 3.9 A: Diazo stability in water is controlled by its pKa (Path A); Brønsted acids that can protonate diazo compounds lead to O-alkylation (Path B); The pKa of phosphates (6-7) should facilitate selective reactions with diazo compounds at neutral pH (bottom). B: Competition experiments at varying pH support the role of pK_a in reactions of Brønsted acids with diazo compounds

For reactions of diazo compounds in water the pK_a of its protonated form (the diazonium ion) is critical. If the conjugate base (**3.1** in Figure 3.9A) is too basic it will be protonated quickly by water and decompose unselectively through hydrolysis (path A in Figure 3.9A). When used to perform transformations in water diazo compounds akin to diazomethane must therefore be used in enormous excess since hydrolysis will always predominate (for example diazomethane has a pKa of 10). If too weakly basic diazo compounds will not deprotonate the Brønsted acids (required for Path B in Figure 3.9) to initiate reaction – compounds in this class include diazomalonates and donor-acceptor substituted diazo esters. These stable diazo compounds have long lifetimes in water and hence can be used to perform aqueous metal-carbene reactions. Their weak basicity, however, renders them of little value in Brønsted acid alkylation. Quantifying the middle ground between these two extremes is

74

essential for controlling selectivity in diazo-type esterifications. The first step in a productive diazo-type O-alkylation is protonation by the Brønsted acid (see Path B in Figure 3.9A). It should then be possible to exploit pKa differences between acids to control O-alkylation selectivity. To test this hypothesis we examined product ratios of CMP versus benzoic acid alkylation in aqueous buffer at different pHs using commercially available trimethylsilyldiazomethane **3.1c** as the diazo compound. Indeed at low pH benzoic acid alkylation is preferred, but as the pH is raised phosphate alkylation begins to dominate (see Figure 3.9B). This is consistent with benzoic acid being largely benzoate at pHs higher than 6, while a large fraction of the phosphate remains protonated. Although **3.1c** was a convenient, highly basic, diazo compound for preliminary hypothesis testing, its extreme reactivity (comparable to diazomethane) would limit its practical application. The ideal diazo compound would have pKb that matches the pKa of the Bronsted acid. Unfortunately there has been no tabulation of pKas for diazo compounds. Raines[135] selected the pKa of the parent alkyl group as a proxy for diazo acidity in their discussion of ester alkylation (See Figure 3.10 right). In the measurement of pKa of diazomethane, McGarrity[136] found the acidity of methanediazonium ion in THF-water is similar to that of nitromethane in water (pKa = 10.2). So we could have a better idea of the acidity of diazo compounds using McGarrity's result (Figure 10 left) than Raines theory. In this manner, phenyl diazomethane (**3.1d**) and 2-nitro-phenyl diazomethane (**3.1e**) with increase acidity will be good substrates for O-alkylation of Brønsted acids.
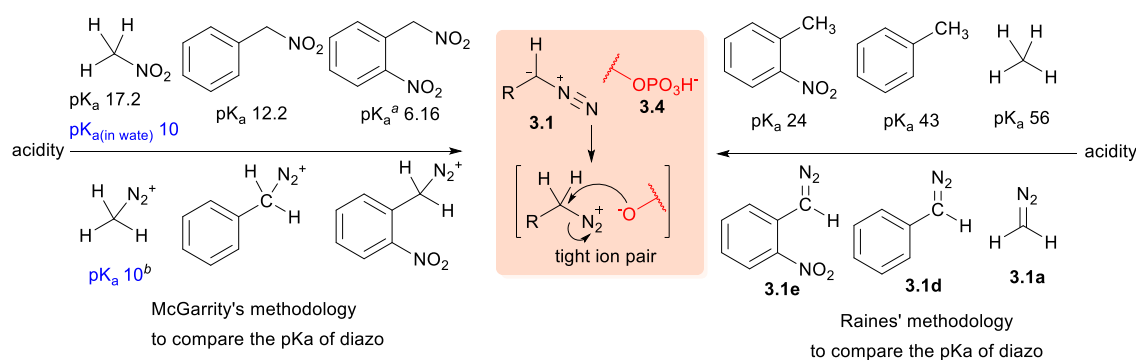
Figure 3.10 pKa of relevent diazo parent alkyl compounds, pKa (in DMSO) is from Bordwell pKa Table

a **pKa calculated using Advanced Chemistry Development (ACD/Labs) Software V11.02 (© 1994-2016 ACD/Labs),** *b* **pKa in THF/water**

3.2.2 Substrate scope of modification of bioactive phosphate compound

We next selected a range of biologically relevant monophosphates and looked at their reaction with diazo compounds **3.1d** and **3.1e**. Although yields are moderate (21-74% depending on the substrate), the ability to obtain milligram quantities of each compound from simple starting materials is a valuable innovation. The yields could likely be improved further if necessary by examining cosolvents (limited

solubility is one of the causes of low conversion) or by tweaking the structure of the diazo compound to make it more stable in water. The modification of serine and tyrosine phosphate only give the phosphate product as we desired. But in the case of serine phosphate (**3.6f**), side product carboxylic esterfication was observed when diazo is increase to 10 equivalents. This could be a result of the proximity of the serine carboxylate to the formed tight ion pair intermediate. In general we consider **3.1e** the superior reagent because it is more stable and the adduct, ortho-nitrobenzyl (ONB), is photolabile.
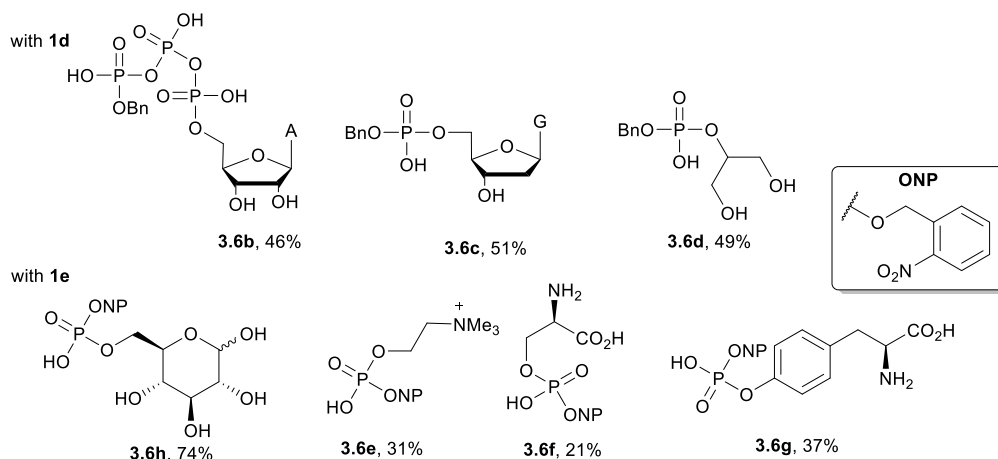
Figure 3.11 Alkylations of diverse naturally occurring monophosphates

Modifications of phosphates within large biopolymers would be a valuable application of diazo alkylation. This reaction has previously been employed in the alkylation of nucleoside monophosphates[47] and nucleic acid oligomers[48] but in those cases substantial amounts of diester alkylation were observed. It is likely that pH buffering is the decisive difference for selectivity and hence we examined the reactivity of two common phosphate diesters c-AMP and c-di-GMP under the same conditions employed in monophosphate alkylation. As shown in Figure 3.12A , both show very limited reactivity with the diazo compound under conditions that deliver smooth reaction of the molecules shown in Figure 3.11A. There are cases where phosphate diester alkylation has been used in practice, but an enormous excess is always employed in DMSO. With 30 equivalents of **3.1e** we begin to see phosphate diester alkylation, but under these forcing conditions substantial amounts of sulfonate alkylation of the MES buffer salt were also observed (experimental part). The dramatic reactivity difference between mono- and diesters augured well for a selective 5'-phosphate alkylation in larger oligonucleotides. Oligonucleotides bearing terminal phosphates are alkylated selectively. A simple hexanucleotide (**8**, see trace a in Figure 3.12B) bearing a 5'-phosphate was treated with the diazo compound **3.1d** in varying amounts at pH 7 (see traces b-g in Figure 3.12B). Optimal conversion was observed at 20 equivalents of **3.1e** with further increases leading to little improvement. The reactions are clean with most by-products derived from side-reactions of the diazo compound with itself or

water (see trace j for control sample where oligo **8** is not added to reaction) and not with the backbone or the bases of the oligonucleotide (see trace h for control experiment of **3.1e**'s reaction with dephosphorylated oligo **8,** figure 3.12B).
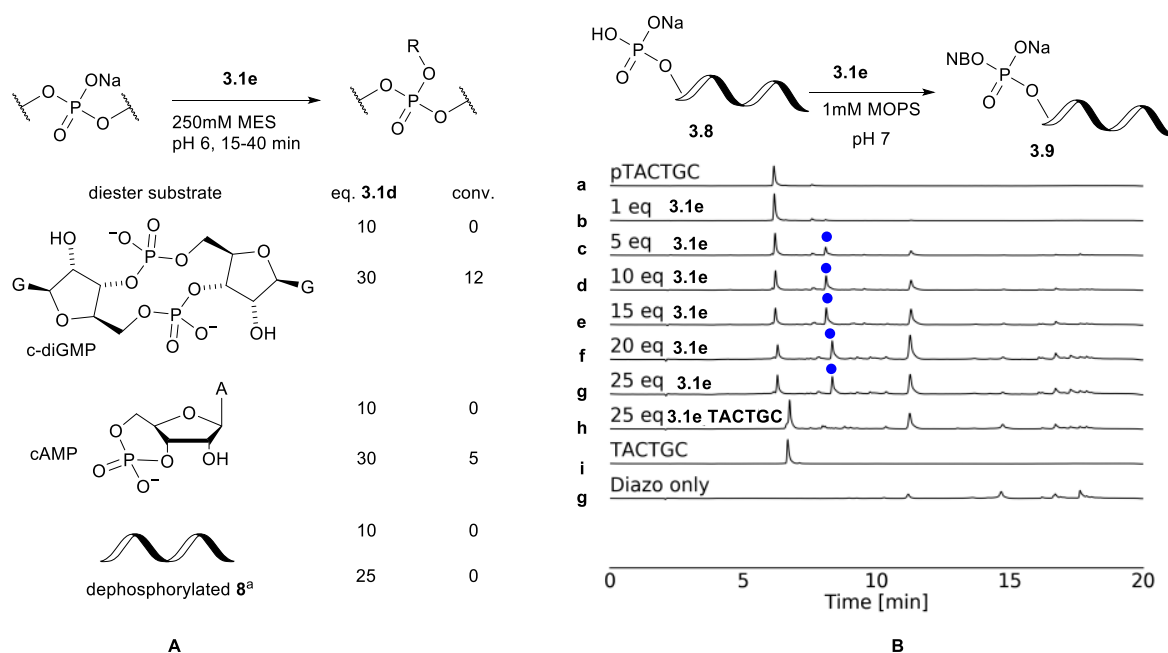
Figure 3.12 A: phosphate diesters are poorly alkylated in aqueous buffer, B: 5'-phosphates in DNA are alkylated cleanly by diazo compound 3.1e.

3.2.3 Src (521-533) and its inhibitor 5mer-peptide modification

Protein phosphorylation is a widespread post-translational modification that regulates numerous functions. The ability to selectively modify phosphates in a native protein would be a powerful tool for studying protein phosphorylation networks. In Zhou's work (left side of Figure 3.13), phosphate labeling requires multi-step synthesis. We are tempted to try selective alkylation of phosphate group in the phosphopeptide based on the competition reaction of carboxylate and phosphate (Figure 3.10), even though Raines group employed diazo compound for labelling of carboxylate groups in proteins.[135] Alkylation to create products **3.6f** and **3.6g** in Figure 3.11 already suggested that a selective phosphate alkylation in a peptide or protein might be feasible.
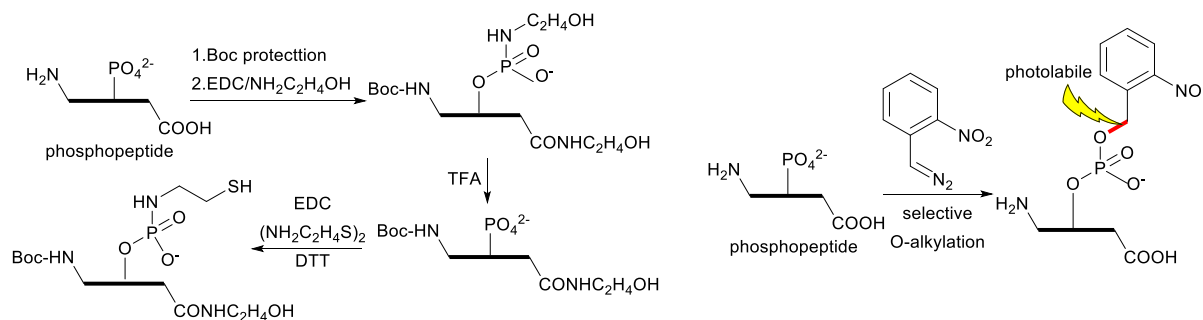
Figure 3.13 selective alkylation of phosphopeptide

Although whole proteins with a precisely defined phosphorylation state are difficult to obtain, many phosphorylated model peptides are commercially available. Phosphorylation of Y527 in the C-terminal region of c-Src causes this domain to fold over on its SH2 domain, holding the protein in an inactive state. A 13-mer phosphorylated peptide of the Src C-terminal domain (residues 521-533, see **3.10** in Tabel 3.1) has therefore become an important model peptide in studying Src signalling. If the commercially available TFA salt of phosphopeptide **3.10** is dissolved in unbuffered water (final pH 4.2) and treated directly with diazo, alkylation of glutamic acid 531 (E531) is the major product (see table 3.1, entry 2). Consistent with the pK$_a$s controlling selectivity, performing the same reaction in pH 7 MOPS buffer switches the alkylation preference to Y527 (table 3.1, entries 2 & 3). Although there are three carboxylates available for alkylation (C-terminus, E524 and E531), phosphate alkylation is the major product along with a small amount of carboxylate alkylation (**3.13a** & **3.13b**, sites of alkylation are determined by MS/MS sequencing, see the experimental part for details). In addition, while 10 equivalents of **3.1e** leads to 75% conversion of phosphopeptide **3.10**, dephosphorylated peptide **3.11** is only partially converted (approximately 30%) to alkylated products under the same conditions. Entries 3-7 of table 3.1 investigate the effect of adding more diazo compound and the ideal seems to be 3 equivalents; Further addition of **3.1e** increases consumption of phosphopeptide **3.10**, but also leads to substantial amounts of double alkylation, thereby reducing the yield of **3.12** (see number in parenthesis in the conversion column of Tabel 3.1).

Tabel 3. 1 selective alkylation of phosphopeptide Src

entry	substrate	pH[a]	equiv. diazo	Major product	conv.[b]	Y527: (E524+531)[c]
1	**3.10**	4	20	**3.13a**	40	1:9[d]
2	**3.10**	7	2	**3.12**	43 (35)	13:1
3	**3.10**	7	3	**3.12**	52 (42)	15:1
4	**3.10**	7	5	**3.12**	68(44)	11:1
5	**3.10**	7	10	**3.12**	75(47)	11:1
6	**3.10**	7	15	**3.1**	95	11:1
7	**3.11**	7	10	**3.14**	30	n. a.

[a]Run in 5mM MOPS buffer at pH 7, pH 4 experiment is run in unbuffered water and the pH is measured after the reaction; [b]first number is conversion of starting peptide; number in parenthesis refers to amount of alkylated phosphate; [c]ratios are determined by HPLC ion extraction intergration; [d]a small side product obscures phosphate slakyltion hence we can only report in a lower limit.

A second study on a model 5-mer phosphorylated peptide supports the generality of the observations with 13mer peptide. This results again indicate how pH can be exploited to control selectivity in Brønsted acid alkylation, even in the presence of competing acids.

Tabel 3. 2 selective alkylation of 5-mer phosphopeptide

entry	substrate	pH	eq. diazo	Major protduct	conv.[a]	Y:E[b]
1	**3.15**	7	3	**3.17**	43 (43)	99:1
2	**3.15**	7	10	**3.17**	70 (56)	5:1
3	**3.15**	7	100	**3.17**	81 (56)	3:1
4	**3.16**	7	20	**3.16**	0	

[a]first number is conversion of starting peptide; number in parenthesis refers to amount of alkylated phosphate; [b]ratios are determined by HPLC intergration

3.2.4 Photo-cleavage of modified product

2-Nitro benzyl ester is a photolabile tag. The alkylated tyrosine phosphate and phosphopeptide can be removed by shining 366nm UV light overnight. Tyrosine phosphate 2-nitro benzyl ester was dissolved in D_2O in the NMR tube. NMR spectrum was measured before and after the photo cleavage reaction (Figure 3.14). The reaction mixture was spiked with another 5 mg of tyrosine phosphate to confirm the photo cleavage result. Triethyl amine salt inside of phosphate 2-nitro benzyl ester was there from the purification buffer and here it could be used as internal standard. The photo-cleavage side product is insolvable in water, and thus is not visible in the NMR spectrum.
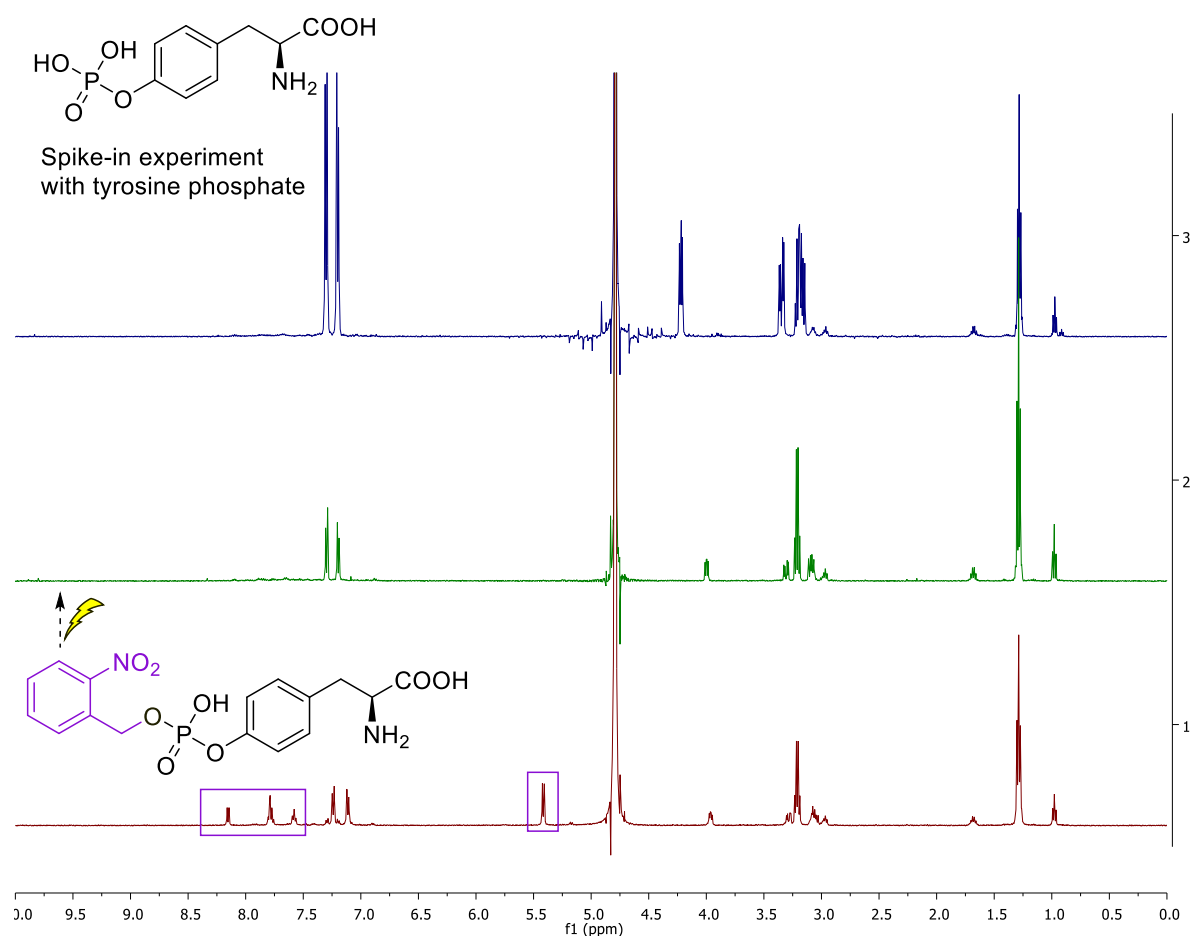
Figure 3.14 photo-cleavage of alkylated tyrosine phosphate

The alkylated peptide Src (527) **3.12** recoveried 89% original phosphorylated peptide with 11% dephosphorylated peptide after photo cleavage by UV radiation The reaction was monitored by UPLC-MS (Figure 3.15).
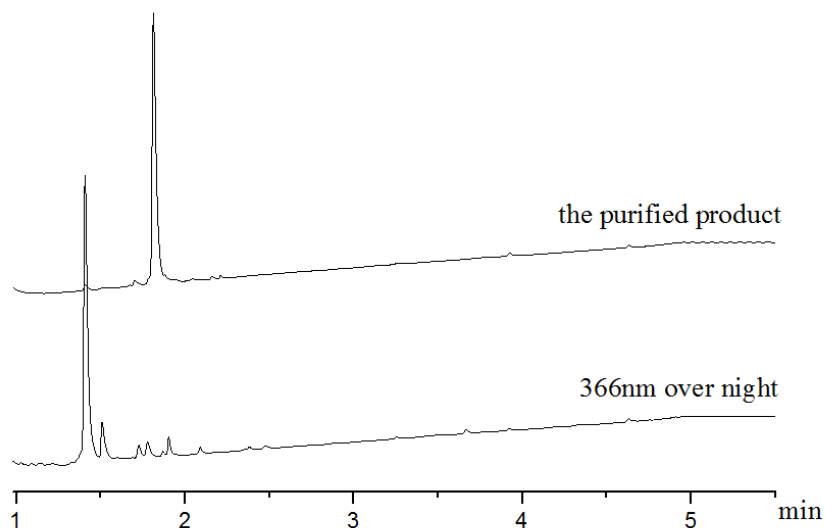
Figure 3.15 photocleavage of caged phosphopeptide 3.12

3.3 Conclusion

We have validated and extended the classical mechanistic view of how diazo compounds accomplish O-alkylation. The accepted mechanism predicts that the pH of the medium could be used to steer O-alkylation preference among Brønsted acids. We have validated this prediction and shown how this property can be used to selectively target phosphates in a mixture of other Brønsted acids including carboxylic acids and sulfonic acids.

In its present form the method is most useful for creating photo-caged variants of peptides, nucleic acids, and phosphate bearing metabolites. Perhaps the most exciting future application, however, would be in global phosphoproteomics, an area that suffers from a lack of selective pull-down reactions. While there remain a number of challenges to be addressed before the method can be applied in phosphoproteomics, a critical hurdle has been overcome: a phosphate selective chemical reaction that works under biological conditions.

The future work on selective esterification of peptide could be continued in two directions:

a) Adding biotin tag to the diazo precursor to pull down phosphopeptides for sequencing;

b) Esterification of phosphopeptides in the cell lysate (with phosphatase inhibitor) where the carboxylic ester could be hydrolyzed by esterase, then pull down the peptide phosphate ester by affinity assay.

3.4 Experimental part

3.4.1 General Methods

All reagents and solvents used were of analytical grade. Buffers were prepared with nanopure water. All chemicals were purchased from Sigma-Aldrich or TCI and used as received. Peptides were purchased from Bachem, oligonucleotides from Microsynth and used as received without further purification. ^{1}H and ^{13}C NMR were acquired on a Bruker 400 MHz, 500 MHz or AvanceIII+ 600 MHz depending on availability of the machine because photolabile compounds needed to be measured once obtained. Chemical shifts relative to TMS were referenced to the solvent's residual peak and are reported in ppm. ESI and ESI-MS-MS spectra were obtained on a Bruker Esquire3000+ spectrometer by direct injection in positive or negative polarity of the ion trap detector. High resolution mass spectra were acquired on a Bruker maXis 4G QTOF ESI mass-spectrometer. MALDI TOF analyses were carried out on a Bruker Microflex mass-spectrometer in linear negative mode using trihydroxyacetophenone as matrix. Shimadzu preparative HPLC (LC-20AP) equipped with phenomenex column (Gemini® 10 μm C18 110 Å, LC Column 250 x 21.2 mm, AXIA™ Packed) was employed for purification. 100 mM triethylammonium acetate (pH 7.2-7.3) and acetonitrile were used as mobile phases. Flow rate: 20 mL/min, 0% acetonitrile in 2 min, 0-30 % acetonitrile in 22 min. Detection was carried out by monitoring the absorbance of the column effluent at 254 nm. The identity of the product peaks was confirmed by ESI-MS. The UV photocleavage reaction was carried out with a CAMAG TLC UV lamp at 366 nm with a sample-lamp distance of ~2 cm. Alkylation of oligonuceotides was purified and analyzed by HPLC (Agilent 1100 LC system equipped with phenomenex ®Jupiter 3μ C18 300Å, 150x4.6mm column) using 100 mM triethylammonium acetate (pH 7.2-7.3) and acetonitrile as the mobile phase. Flow rate: 1 mL/min. 0-35 % acetonitrile in 12 min, 15 min 80 %. The fractions were analyzed by ESI-MS. Compeptition alkylations of CMP and benzoic acid was analyzed by UPLC-MS (method A). Peptide alkylation experiments were analyzed by UPLC-MS (method B) using an Agilent 1290 Infinity system, equipped with a Zorbax Eclipse Plus C18 2.1x50 mm column (Agilent), coupled to an Agilent 6130 Quadrupole LC/MS. Elution was performed using 0.1 % formic acid in water/acetonitrile. Method A: flow rate 0.45 mL/min, 5-95 % acetonitrile in 3.5 min; method B: 0% acetonitrile for 0.2 min, 0-54 % acetonitrile in 4 min, 54-95 % acetonitrile in 1 min.

3.4.2 Synthesis of diazo compounds

Diazo substrates were synthesized according to published procedures.[137-139] The corresponding aryl aldehyde was first converted into a hydrazone which was then oxidized to the final diazo compound. We tested two oxidation methods: Swern Oxidation and MnO_2 oxidation. They both work well. Oxidations using MnO_2 are fast and simple (~20 min), however during work up the reactive diazo product was found to decompose easily. Swern Oxidation was carried out at -78 °C, and there is less

decomposition afterwards, especially for benzyl diazo which is very unstable at room temperature. Once formed, benzyl diazo is used for modification reactions directly after work up and without further purification. 2-nitro-phenyl diazomethane was found to have a higher stability compared to benzyl diazo and thus can be stored for two weeks at -20°C.
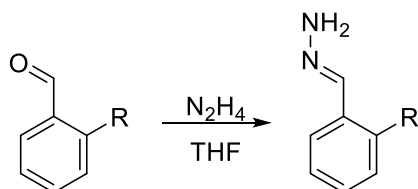
In a round-bottomed flask containing 12 mL of N_2H_4 (1 M in THF, 12 mmol), a solution of aldehyde (1 M in THF, 10 mmol) was slowly added. The mixture was stirred for 30 min at room temperature. The mixture was evaporated under reduced pressure to affording the desired hydrazone (>95% yield). All hydrazones were used without further purification for the generation of diazo compounds.
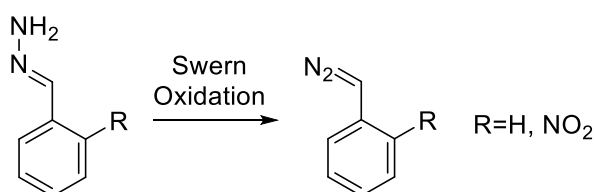
Under a nitrogen atmosphere oxalyl chloride (0.225 mL, 1.05 equiv.) was added drop wise to a stirred solution of Et_2O (21 mL) containing DMSO (0.195, 1.10 equiv.) at -55°C. After gas evolution ceased (~20 min), the reaction was cooled down to -78°C. Behind a blast shield, a mixture of Et_3N (0.732 mL, 2.10 equiv.) and hydrazone (2.50 mmol, 1.00 equiv.) in THF (6 mL) was added drop wise over a period of 5-7 min to the activated DMSO solution. An immediate color change and concomitant formation of a white precipitate were observed. The reaction mixture was maintained at -78°C for ~1 h and allowed to gradually warm up to 0°C followed by the addition of and ice cold half saturated NH_4Cl solution (20 mL) was added. The reaction mixture was extracted with cold Et_2O (dry ice was added to reach -30°C, 2 x 50 mL), and the combined organic layers were dried over magnesium sulfate, filtered and concentrated at 15 °C under reduced pressure.
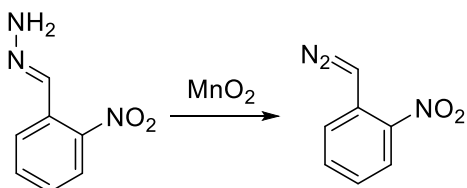
In a 100 mL round bottom flask, Hydrazone (661 mg, 4.0 mmol, 1.00 eq.) dissolved in 50 mL of $CHCl_3$, was stirred with MnO_2, (2.85 g, 33 mmol, 8.25 equiv.) for 5 min in the dark. In the absence of light the

mixture was filtered over celite and the filtrate washed with 0.1 M NaHCO$_3$ until pH 7 was reached, dried over MgSO$_4$, filtered and stored at -20 °C in a brown glass vial wrapped in black foil.

3.4.3 CMP methylation

All aqueous buffers were prepared freshly with final concentrations of 200 mM. Buffer pH 4 and 5: CH$_3$COOH/CH$_3$COONa, pH 6: MES/NaOH, pH 7 and 8: MOPS/NaOH.

CMP (20 mg, 54 µmol) was dissolved in the appropriate buffer (2.72 mL) to give a 20 mM stock solution. Then, different equivalents of TMSCHN$_2$ (2.0 M in hexane) were added. Upon disappearance of the yellow layer, UPLC-MS analysis (Method A) was performed. The results are summarized in table S 3.1.

Table S3.1 CMP methylation conversions at different pH with TMSCHN$_2$

pH	Equiv. Diazo	Conv.
5	1	2%
	5	4%
	10	9%
	20	27%
6	1	21%
	5	9%
	10	16%
	20	36%
	25	37%
	35	46%
7	1	41%
	5	16%
	10	31%
	20	57%
	25	64%
8	1	41%
	5	16%
	10	19%

20	49%
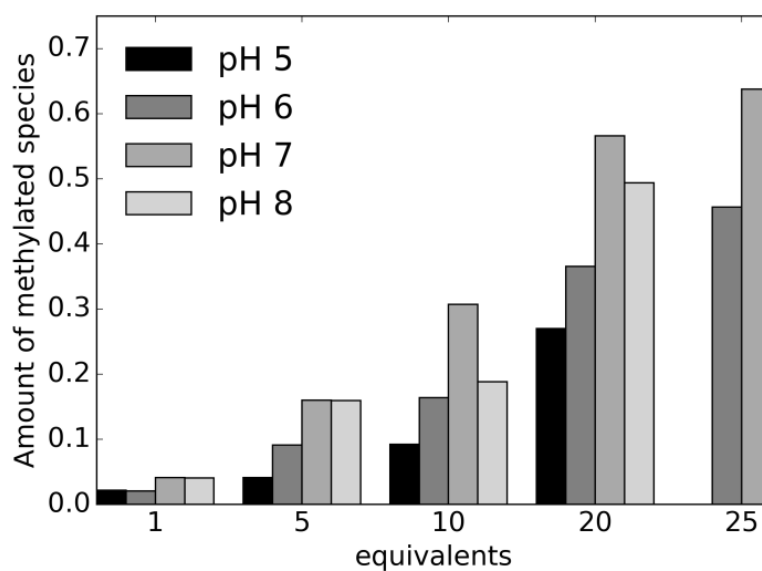
3.4.4 Methylation selectivity towards CMP and benzoic acid with TMSCHN$_2$
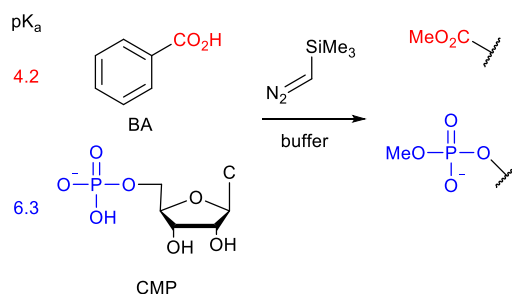
CMP (20 μL, 100 mM in water), benzoic acid (20 μL, 100 mM in CH$_3$CN), buffer (60 μL, 200 mM) were added into an Eppendorf vial. The vial was closed, vortexed for 5 s and then centrifuged for 10 s. Then TMSCHN$_2$ (2.0 M in hexane, 20 μL) was added and the mixture allowed to react without agitation at room temperature. After the yellow color disappeared, the reaction was analyzed by UPLC-MS (method A). The peak areas of CMP and methylated CMP were measured at 254 nm. The peak areas of benzoic acid and methyl benzoate were measured at 220 nm. The results are summarized in Table S3.2.

Table S3.2 Competition experiments at varying pH towards diazo compounds

pH	CMP conv.	ø	SD	BA conv.	ø	SD
				[%]		
	16.1			33.0		
4	15.2	15.2	0.767	34.9	34.3	1.49
	14.5			33.9		
	31.0			23.4		
5	31.1	30.9	0.154	26.2	25.9	2.37
	30.8			28.1		
	41.8			7.7		
6	41.4	41.6	0.321	7.6	8.1	0.86
	41.4			9.1		
	50.9			10.1		
7	56.0	51.0	4.93	10.9	10.6	0.356
	46.1			10.7		
	53.8			6.5		
8	61.4	54.9	6.05	6.4	6.5	0.056
	49.2			6.6		
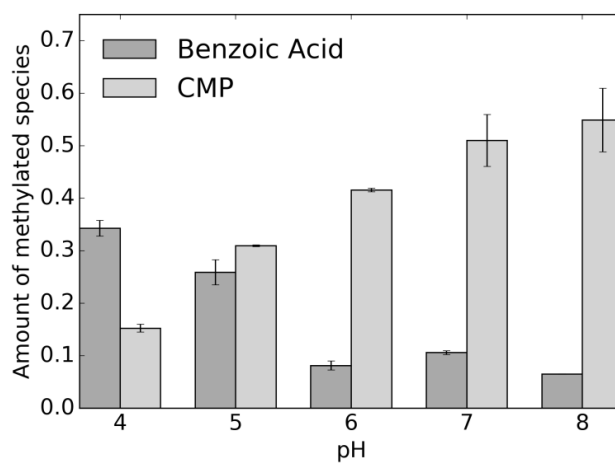
3.4.5 Modification of bioactive phosphate compounds

In a round bottomed flask a 50 mM stock solution of phosphate derivative (~50 mg dissolved in 250 mM MES buffer pH 6) was prepared. To the solution was added freshly prepared phenyl diazomethane (as a solid, 10 equiv.) whereupon gas bubbles were formed immediately. The reaction mixture was left for 30 min at room temperature until gas evolution ceased. The side-products were very sticky to the glass wall and the clear solution was purified by preparative HPLC to afford the pure product.

In some cases, the starting material phosphate monoester was only supplied as $RPO_4 \cdot xH_2O$ hydrate compounds. The amount of water is unclear. They were co-evaporated with CH_3CN three times to remove the water (100 mg phosphate monoester in 10 mL CH_3CN), afterwards they were used directly as anhydrous pure products.
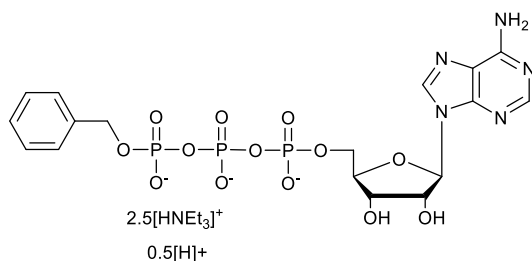
40 mg ATP sodium salt was dissolved in 1.45 mL of 250 mM MES buffer pH 6, and freshly prepared benzyl diazo (90 mg) was added to the reaction solution. Gas was formed immediately. The reaction mixture was left for 30 min until the gas formation ceased. The side-products stuck to the glassware and the clear solution was purified by preparative HPLC. The desired product eluted at 15 min. The product fractions were combined and lyophilized to afford the desired product (43 mg, 46%). ^{1}H NMR (500 MHz, DMSO-d_6) δ 10.84 (s, 2H), 8.48 (s, 1H), 8.14 (s, 1H), 7.37 − 7.26 (m, 6H), 7.24 (m, 1H), 5.91 (d, J_{HH} = 5.1 Hz, 1H), 4.88 (dd, J_{HP} = 6.3, 1.9 Hz, 2H), 4.58 (t, J_{HH} = 5.0 Hz, 1H), 4.34 (dd, J_{HH} = 17.7, 13.4 Hz, 1H), 4.10 − 3.99 (m, 3H), 3.04 − 2.88 (m, 16H), 1.12 (t, J_{HH} = 7.2 Hz, 25H). ^{31}P NMR (202 MHz, DMSO-d_6) δ -12.33, -12.43, -23.47. ^{13}C NMR (101 MHz, DMSO-d_6) δ 155.94, 152.56, 149.51, 139.23, 127.95, 127.03, 118.79, 86.97, 83.35 (d, $^3J_{CP}$ = 8.08 Hz), 73.92, 69.97, 66.28 (d, $^2J_{CP}$ = 5.35 Hz), 64.42 (d, $^2J_{CP}$ = 3.23 Hz), 45.00, 8.61. HR-ESI (Et$_3$N salt): Exact Mass calcd. for $C_{23}H_{37}N_6O_{13}P_3$ [M·NEt$_3$+H]$^+$: 699.1631, found: 699.1701.
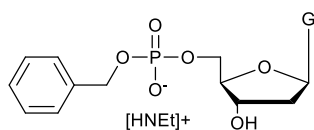
50 mg Deoxyguanosine monophosphatephospahte disodium salt was dissolved in 2.5 mL of 250 mM MES buffer pH 6, and freshly prepared benzyl diazo (151 mg) was added to the reaction solution, and gas was formed immediately. The reaction mixture was left for half an hour until the gas formation

ceased. The side-products stuck to the glassware and the clear solution was purified by preparative HPLC. The desired product eluted at 15 min. The product fractions were combined and lyophilized to afford the desired product (30 mg, 51%). ^{1}H NMR (500 MHz, DMSO-d_6) δ 10.69 (s, 1H), 7.92 (s, 1H), 7.36-7.27 (m, 4H), 7.24 – 7.20 (m, 1H), 6.60 (s, 2H), 6.12 (t, J_{HH} = 6.9 Hz, 1H), 5.49 (d, J_{HH} = 2.8 Hz, 1H), 4.71 (d, J_{HP} = 6.6 Hz, 2H), 4.42 (s, 1H), 3.89-3.92 (m, 2H), 3.83 – 3.72 (m, 1H), 2.90 (s, 7H), 2.61 – 2.52 (m, 1H), 2.19 (ddd, J_{HH} = 13.0, 6.1, 3.1 Hz, 1H), 1.11 (t, J_{HH} = 7.0 Hz, 11H). ^{31}P NMR (202 MHz, DMSO-d_6) δ -0.70 (s). ^{13}C NMR (101 MHz, DMSO-d_6) δ 156.77, 153.67, 150.88, 139.44 (d, $^3J_{CP}$ = 7.57 Hz) , 135.45, 128.01, 126.98, 116.68, 86.07 (d, $^3J_{CP}$ = 7.37 Hz) 82.77, 71.22, 65.90 (d, $^2J_{CP}$ = 5.35 Hz), 64.80 (d, $^2J_{CP}$ = 5.15 Hz), 45.36, 8.78. HR-ESI (Et$_3$N salt): Exact Mass calcd. for C$_{23}$H$_{36}$N$_6$O$_7$P [M$^-$Et$_3$N+H]$^+$: 539.2378, found: 539.2378.
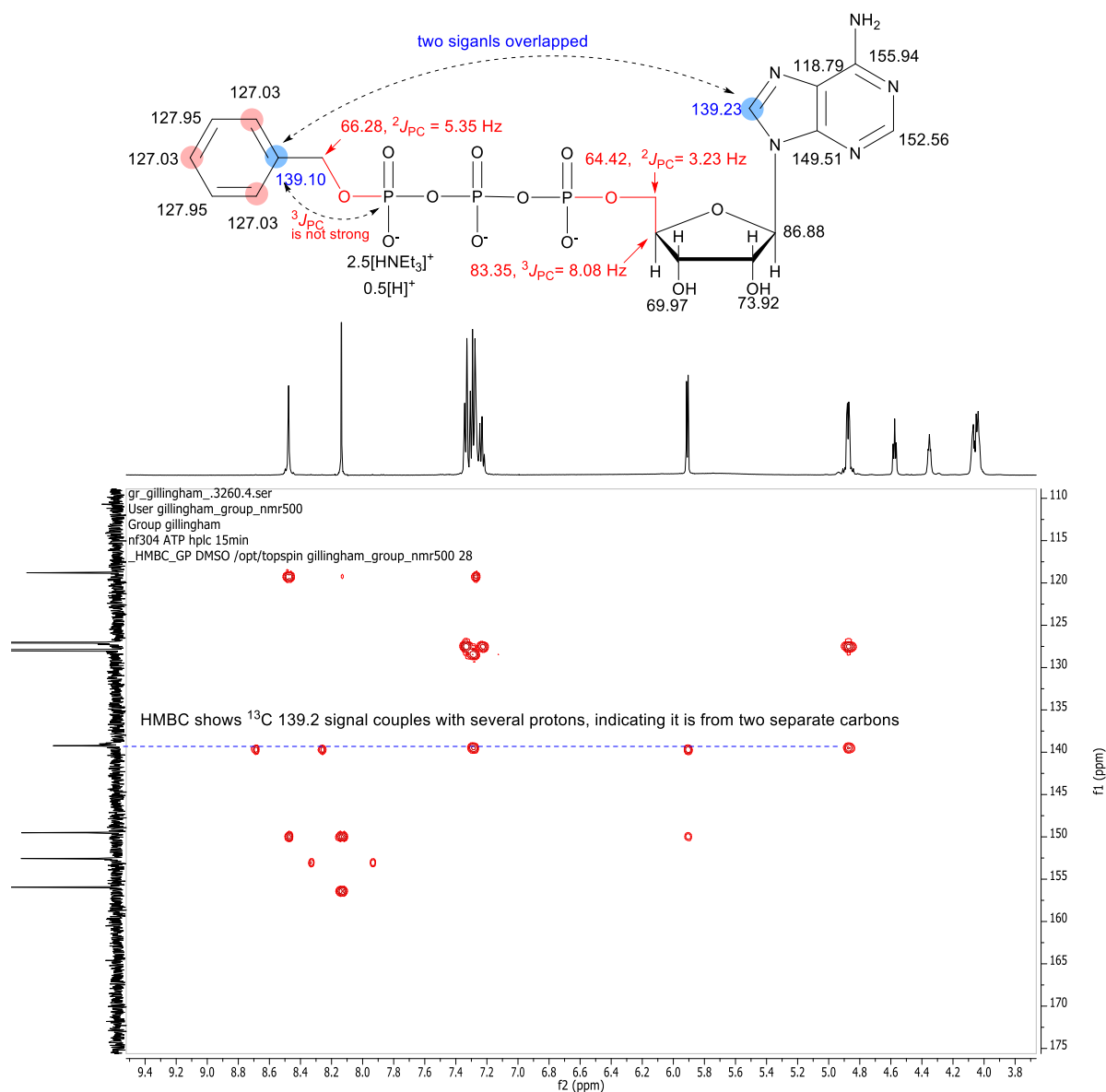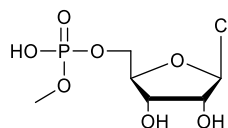
CMP (22 mg, 60 μmol) was dissolved in 1.2 mL of 250 mM MES buffer pH 6. TMS diazomethane was added (100 μL, 200 μmol, 2 M in hexane, 3.33 equiv.). It was stirred for 20 min at room temperature. According to HPLC, only a minor amount of product has formed. Additional TMS-diazomethane was added (200 μL, 400 μmol, 2 M in hexane, 6.67 equiv.) and the reaction was stirred until the yellow diazo color disappeared. The reaction mixture was lyophilized and purified by preparative HPLC. The product fractions were combined and lyophilized to afford the desired methylated product (5 mg, 25%) as a white solid. ^{1}H-NMR (500 MHz, D_2O) δ = 7.83 (d, J = 7.6Hz, 1H), 6.01 (d, J = 7.6Hz, 1H), 5.90 (d, J = 3.9Hz, 1H), 4.25 to 4.19 (m, 2H), 4.19 – 4.13 (m, 1H), 4.09 (ddd, J = 4.3, 12.0, 2.3Hz, 1H), 3.99 (ddd, J = 5.1, 12.1, 3.1Hz, 1H), 3.51 (d, J = 11.0Hz, 3H). ^{13}C-NMR (125 MHz, D_2O) δ 166.2, 157.70, 141.22, 96.43, 89.42, 82.52, 74.11, 69.22, 64.17, 52.87. ^{31}P-NMR (202 MHz, D_2O) δ 1.47. HR-ESI: Exact mass calcd. for $C_{10}H_{15}N_3O_8P$ [M][-]: 336.0602, found: 336.0605.
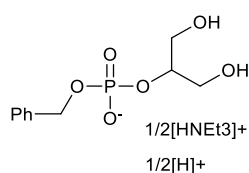
50 mg glycerol phosphate disodium salt was dissolved in 4.6 mL of 250 mM MES buffer pH 6 followed by the addition of freshly prepared benzyl diazo (90 mg). The reaction mixture was left for 30 min until the gas formation ceased. The side-products were stuck to the glassware and the clear solution was purified by preparative HPLC. The desired product eluted at 15 min. The product fractions were combined and lyophilized to afford the desired product (30 mg, 49%). ^{1}H NMR (500 MHz, DMSO-d_6) δ 9.85 (s, 0.5H), 7.42 – 7.32 (m, 4H), 7.30 (m, 1H), 4.85 (d, J_{HP} = 6.3 Hz, 2H), 4.04 (m, 1H), 3.46 (dd, J_{HH} = 5.4, 2.0 Hz, 4H), 3.05 (dd, J_{HH} = 7.5, 3.2 Hz, 3H), 1.17 (t, J_{HH} = 7.3 Hz, 4H). ^{31}P NMR (202 MHz, DMSO-d_6) δ -0.71. ^{13}C NMR (101 MHz, DMSO-d_6) δ 138.44 (d, $^3J_{CP}$ = 8.48 Hz), 128.16, 127.44, 127.26, 77.91 (d, $^2J_{CP}$ = 5.76 Hz), 66.60 (d, $^2J_{CP}$ = 5.35 Hz), 61.66 (d, $^3J_{CP}$ = 4.64 Hz), 45.49, 8.50. HR-ESI: Exact mass calcd. for $C_{10}H_{14}Na_2O_6P$ [M-H+2Na][+]: 307.0318, found: 307.0320.
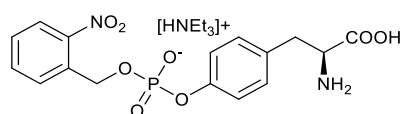
53 mg tyrosine phosphate was dissolved in 250 mM of MES buffer pH 6 followed by the addition of 2-nitro-phenyl diazomethane (350 mg). CH_3CN (1 mL) was added to rinse down the remaining solid particles. After 40 min the gas formation ceased and the aqueous solution was purified by preparative

HPLC. The desired product eluted at 17 min. The product fractions were combined and lyophilized to afford the desired product (25 mg, 37%). ^{1}H NMR (500 MHz, DMSO-d_6) δ 8.07 (dd, J_{HH}= 8.2, 0.9 Hz, 1H), 7.82 (d, J_{HH} = 7.6 Hz, 1H), 7.79 – 7.73 (m, 1H), 7.54 (t, J_{HH} = 7.2 Hz, 1H), 7.10 (d, J_{HH} = 8.5 Hz, 2H), 7.01 (d, J_{HH} = 8.4 Hz, 2H), 5.16 (d, J_{HP} = 7.3 Hz, 2H), 3.39 (dd, J_{HH} = 7.7, 4.8 Hz, 1H), 3.05 (d, J_{HH} = 4.7 Hz, 1H), 2.87 (q, J_{HH} = 7.2 Hz, 7H), 2.83 – 2.78 (m, 1H), 1.10 (t, J_{HH} = 7.2 Hz, 10H). ^{31}P NMR (202 MHz, DMSO-d_6) δ -6.07 (s). ^{13}C NMR (101 MHz, DMSO-d_6) δ 170.14, 152.64 (d, $^{2}J_{CP}$ = 6.46 Hz), 146.47, 135.51 (d, $^{3}J_{CP}$ = 7.58 Hz), 133.97, 130.30, 129.77, 128.47, 128.08, 124.35, 119.70-119.65 88 (d, $^{3}J_{CP}$ = 4.75 Hz), 63.25 (d, $^{2}J_{CP}$ = 4.75 Hz), 55.19, 45.31, 35.99, 8.75. HR-ESI: Exact mass calcd. for $C_{16}H_{15}N_2Na_2O_8P$ [M-H+2Na]$^{+}$: 441.0361, found:441.0432.
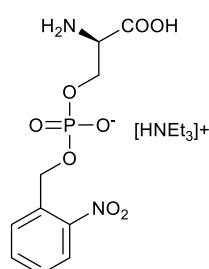
The reaction for serine phosphate modification at 50 mM resulted in the formation of a carboxylate ester as a side product. Therefore the reaction was carried out at 10 mM serine phosphate which resulted only in phosphate modification. The freshly prepared 2-nitro phenyl diazomethane (220 mg, 5 equiv.) was immediately added to the serine phosphate solution (50 mg in 27 mL 100 mM MES buffer pH 6). After 30 min the gas formation ceased. The side-products stuck to the glassware and the clear solution was purified by preparative HPLC. The desired product eluted at 14 min. The product fractions were combined and lyophilized to afford the desired product (18 mg, 21%). ^{1}H NMR (500 MHz, DMSO-d_6) δ 8.08 (d, J = 8.1 Hz, 1H), 7.90 – 7.73 (m, 2H), 7.56 (t, J = 7.5 Hz, 1H), 5.09 (d, J_{HP} = 6.8 Hz, 2H), 4.04 (t, J_{HH} = 12.1 Hz, J_{HP} = 12.1 Hz, 1H), 3.89 (dd, J_{HP} = 11.0 Hz, J_{HH} = 7.5 Hz, 1H), 3.41 (d, J_{HH} = 5.4 Hz, 1H), 3.00 (q, J_{HH} = 7.2 Hz, 5H), 1.15 (t, J_{HH} = 7.0 Hz, 8H). ^{31}P NMR (202 MHz, DMSO-d_6) δ 0.61.13C NMR (101 MHz, DMSO-d_6) δ 167.77, 146.68, 135.21, 135.17 (d, $^{3}J_{CP}$ = 8.08 Hz), 134.00, 128.63, 128.28, 124.42, 64.10 (d, $^{2}J_{CP}$ = 5.45 Hz), 63.17 (d, $^{2}J_{CP}$ = 4.04 Hz), 55.73, 45.31, 8.67.
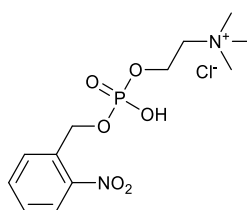
66 mg of choline phosphate chloride salt was dissolved in 4 mL of MES buffer pH 6. Then 327 mg of 2-nitro-phenyl diazomethane was added to phosphate aqueous solution. The reaction proceeded

smoothly and the aqueous solution was purified by preparative HPLC. The desired product eluted at 15 min. The product fractions were combined and lyophilized to afford the desired product (22 mg, 31%). ^{1}H NMR (400 MHz, DMSO-d_6) δ 8.07 (d, J_{HH} = 8.0 Hz, 1H), 7.86 (d, J_{HH} = 7.5 Hz, 1H), 7.79 (t, J_{HH} = 7.4 Hz, 1H), 7.55 (t, J_{HH} = 7.5 Hz, 1H), 5.07 (d, J_{HP} = 7.0 Hz, 2H), 4.06 (m, 2H), 3.52 (m, 2H), 3.13 (s, 9H). ^{31}P NMR (202 MHz, DMSO-d_6) δ -0.99. ^{13}C NMR (101 MHz, DMSO-d_6) δ 146.62, 135.89 (d, $^3J_{CP}$ = 7.37 Hz), 133.94, 128.58, 128.09, 124.34, 65.46, 62.92 (d, $^2J_{CP}$ = 4.44 Hz), 58.36 (d, $^2J_{CP}$ = 5.25 Hz), 53.10, 53.07, 53.04, 39.94, 39.73, 39.52, 39.31, 39.10. HR-ESI: Exact mass calcd. for $C_{12}H_{20}N_2O_6P^+$ [M-Cl]$^+$: 319.1053, found: 319.1056.
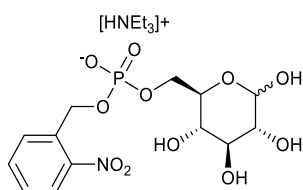
48 mg of glucose monophosphate dipotassium salt was dissolved in 2.85 mL of MES buffer pH 6 followed by the addition of 233 mg of 2-nitro-phenyl diazomethane. The reaction proceeded smoothly and the aqueous solution was purified by preparative HPLC. The desired product eluted at 17 min. The product fractions were combined and lyophilized to afford the desired product (42 mg, 74%). The glucose modified product is a mixture of antomer (ratio=1:1.2), but some of the NMR peaks are overlapped. ^{1}H NMR (400 MHz, D$_2$O) δ 8.19 (d, J_{HH} = 8.1 Hz, 1H), 7.95 – 7.75 (m, 2H), 7.60 (t, J_{HH} = 7.6 Hz, 1H), 5.35 (d, J_{HP} = 7.3 Hz, 2H), 5.14 (d, J_{HH} = 3.7 Hz, 0.49H), 4.61 (d, J_{HH} = 8.0 Hz, 0.59H), 4.20 – 4.00 (m, 2H), 3.88 (d, J_{HH} = 9.6 Hz, 0.5H), 3.69 (t, J_{HH} = 9.5 Hz, 0.5H), 3.57 – 3.39 (m, 3H), 3.22 (q, J_{HH} = 7.3 Hz, 6H), 1.29 (t, J_{HH} = 7.3 Hz, 9H). ^{13}C NMR (101 MHz, D$_2$O) δ 146.56, 146.51, 134.54, 133.72-133.65 (d, $^3J_{CP}$ = 7.37 Hz), 133.62-133.54(d, $^3J_{CP}$ = 7.77 Hz), 128.73, 128.70, 128.68, 128.61, 124.96, 124.92, 95.95, 92.08, 75.54, 74.67-74.59 (d, $^2J_{CP}$ = 7.77 Hz), 74.03, 72.54, 71.40, 70.35-70.28 (d, $^2J_{CP}$ = 7.67 Hz), 69.18, 64.57, 64.53 (d, $^2J_{CP}$ = 3.23 Hz), 64.49-64.44 (d, $^2J_{CP}$ = 4.84 Hz), 46.66, 8.21. ^{31}P NMR (202 MHz, D$_2$O) δ 0.61, 0.53. HR-ESI (Et$_3$N salt): Exact mass calcd. for $C_{19}H_{34}N_2O_{11}P^+$ [M·NEt$_3$+H]$^+$: 497.1900, found: 497.1902.
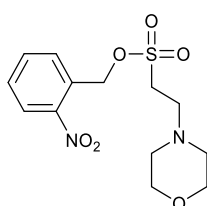
Because MES salt was used as the buffer, it is ten times concentrated than phosphate monoester, so there is also MES esterification reaction observed as the side reaction. But the conversion is very low

(Conv. < 5%). In the next chapter, when diazo was tested with phosphodiester (c-AMP and c-di-GMP), since there is no phosphate monoester available, thus the MES reacted with diazo and the sulfate esterification product could be purified. (in the c-di-GMP reaction, MES ester, isolated yield 3%). ^{1}H NMR (600 MHz, DMSO-d_6) δ 8.21 (dd, J = 7.9, 1.2 Hz, 1H), 7.97 – 7.81 (m, 3H), 5.04 (s, 2H), 3.97 – 3.89 (m, 2H), 3.85 (dd, J = 14.5, 9.2 Hz, 4H), 3.54 (d, J = 12.5 Hz, 2H), 3.30 – 3.26 (m, 2H), 3.16 – 3.08 (m, 2H).^{13}C NMR (151 MHz, DMSO-d_6) δ 150.74, 136.81, 133.95, 132.45, 126.06, 120.49, 60.71, 59.91, 56.31, 53.54, 43.41. HR-ESI (Et$_3$N salt): Exact mass calcd. for C$_{19}$H$_{34}$N$_3$O$_6$S [M·Et$_3$N+H]$^+$: 432.2163, found: 432.2170.

3.4.6 Photo cleavage of Tyrosine phosphate 2-nitro benzyl ester

0.8 mg of tyrosine phosphate 2-nitro benzyl ester was dissolved in D$_2$O in a NMR tube and kept in the dark for measuring NMR spectrum, then it was put under 366 nm UV lamp overnight (NMR spectrum was measured in between). After 15 h, the phosphate 2-nitro-benzyl ester was found to be cleaved according to ^{1}H-NMR spectral data. Afterwards 5 mg of tyrosine phosphate were added to the NMR tube in order to confirm the product peaks. The photo-cleaved side product of the reaction is insoluble in water and therefore not visible in the ^{1}H-NMR spectrum. Please note the presence of triethylamine salt which resulted from the HPLC purification buffer of the phosphate 2-nitro benzyl ester. In these experiments it was used as an internal standard.
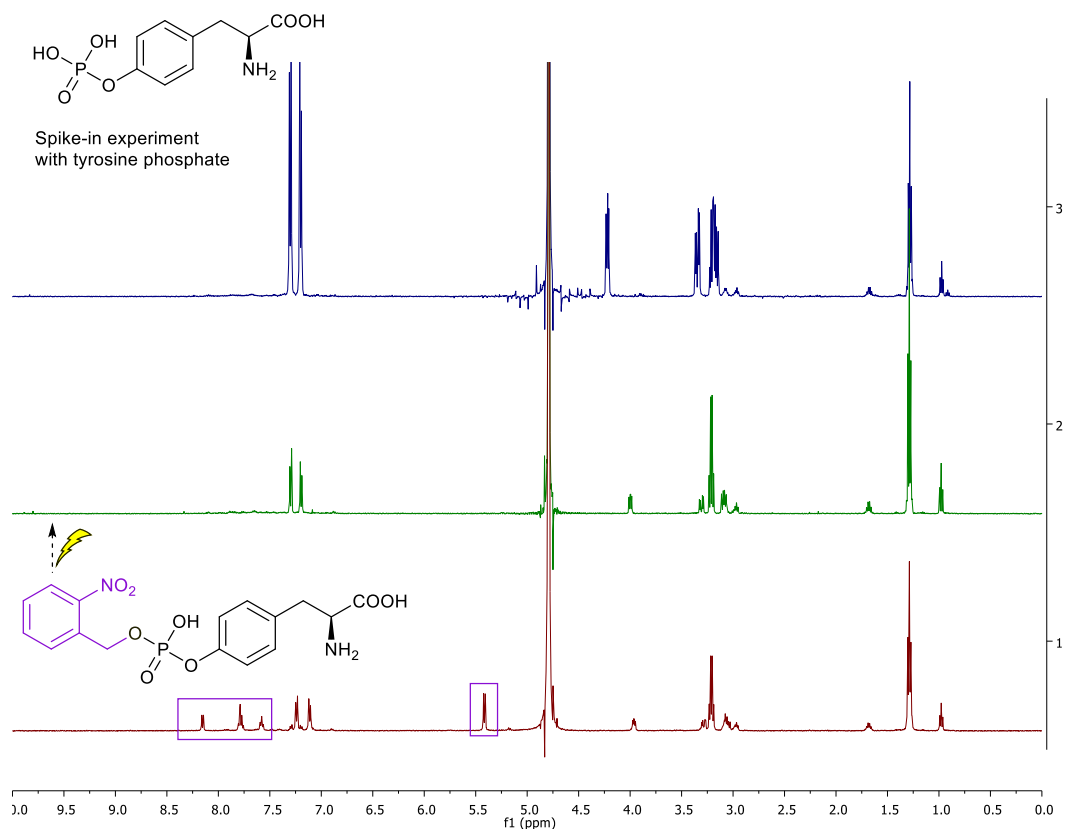
3.4.7 Alkylation of phosphate monoester vs phosphodiester

The c-AMP and c-di-GMP were tested under the same condition as before, pH 6, MES buffer. It was found that phosphodiesters remained unconverted employing 10 equiv. of diazo substrate, low conversion with 30 equiv. diazo. The 5'-phosphorylated hexanucleotide is alkylated with high conversion compared to non-phosphorylated hexanucleotide (table S3.3). Hexanucleotides with various amount of diazo were also tested later at pH 7 and summarized in table S3. 4.

Table S3.3 screening of alkylation of phosphodiester

		Conversion of nulceic acid	
	Reaction concentration	10 equiv. diazo	30 equiv. diazo
c-AMP	50 mM	0	5%
c-di-GMP	50 mM	0	12%
TACTGC	100 μM	0	0
pTACTGC	100 μM	42%	52%

c-AMP

70 mg c-AMP was dissolved in 4 mL of 250 mM MES buffer pH 6 to give a 50 mM c-AMP stock solution. Then 0.5 mL of this stock solution was incubated with 10 equiv. of 2-nitro-phenyldiazomethane whereupon gas bubbles formed slowly. After 1.5 h the gas formation ceased and UPLC-MS analysis (Method A) indicated no conversion of c-AMP. Another 0.5 mL of c-AMP solution was treated with 30 equiv. of 2-nitro-phenyldiazomethane, however due to the large excess of diazo the reaction became heterogeneous. Although the reagent couldn't interact with all the c-AMP solution, gas formation stopped after 2 h. The aqueous solution was transfer to a new flask and washed with diethyl ether (2 mL) to remove diazo side products. The aqueous solution was submitted to UPLC-MS analysis indicating 5% c-AMP conversion to the corresponding product. When using a huge excess of diazo reagent (50 equiv.) for c-AMP modification, conversion didn't exceed 6%.

c-di-GMP

22 mg c-di-GMP was dissolved in 1 mL of 250 mM MES buffer pH 6 to give a 50 mM stock solution. Then 0.5 mL of this stock solution was incubated with 10 equiv. of 2-nitro-phenyldiazomethane, whereupon gas bubbles formed slowly. After 1.5 h the gas formation ceased and UPLC-MS analysis

93

indicated no conversion of c-di-GMP. Another 0.5 mL of c-di-GMP solution was treated with 30 equiv. diazo and gas formation stopped after 2 h. The aqueous solution was transfer to a new flask and washed with diethyl ether to remove the diazo side products. The aqueous solution was submitted to UPLC-MS analysis where 12% of modified product was observed.

Hexanucleotide TACTGC

The purchased oligomers were dissolved in water to give approximately 2 mM stock solutions. The amount of water was calculated according to Microsynth's quality certification. The accurate concentration was determined by NanoDrop measurements. 5'-Phosphorylated oligo (pTACTGC) was determined to be at 1.4 mM and TACTGC to be at 1.7 mM. 10 mg of 2-nitro-phenyl diazomethane was dissolved in CH_3CN to form a 10 mM stock solution.

Reactions

Oligonucleotide alkylations were tested at 100 μM concentration. The oligomer was first mixed with buffer (MOPS, pH 7.15, 5 mM), diluted with water followed by the addition of the diazo. The reaction details and results are summarized in Table S3.4. The 5'-phospharylated oligomer pTACTGC was found to give a decent conversion using 5 equivalents of diazo and can be increased to 50% using up to 20 equivalents of diazo. Further increase of the stoichiometry didn't have a positive effect on the conversion. The modification of 5'-phospharylated oligomer pTACTGC at 50 μM was also tested with various amounts of diazo compound. Once the diazo stoichiometry was above 10 equivalents, the reaction gave similar conversions when compared to the modification at 100 μM. With the normal synthetic oligo TACTGC, no conversion was noted even when using 40 equiv. of diazo as determined by HPLC analysis.

In general the major HPLC peaks were collected manually and subjected to ESI-MS and MS/MS analysis. The desired alkylated oligo was further confirmed by HR-ESI: exact mass calcd. for $C_{65}H_{79}N_{21}O_4P_6^{2-}$, $[M]^{2-}$: 989.6687, found: 989.6603. Using MS-MS the exact alkylation site could not be determined due to the limitation of the MS-MS method. The reaction mixture was also submitted to MALDI TOF analysis. The pTACTGC modification showed mainly mono-alkylation product and a small fraction of double modification. This result is in good agreement with the MS-MS data (Figure S4). The minor peak observed (MS 867.07, I/I_{max} 10%) may be [T*.A.C-b(T)]⁻, [T.A*.C-b(T)]⁻ or [T.A.C*-b(T)]⁻, where * is the site of modification, **.** is the phosphate linkage and -b(T) is the ribose without base. For the TACTGC, no modification was observed by HPLC analysis. Instead MALDI TOF analysis indicated that besides the major oligo peak, a less intense mass peak for the mono-modified oligomer could be identified. This states that a very small portion of oligo is alkylated which could only be observed by MALDI TOF since a new HPLC peak could not be detected under these conditions.

Table S3.4 Reaction details and alkylation results of 5'-phospahte hexanucleotide

pTACTGC [mM]	TACTGC [mM]	Buffer [mM]	Diazo [mM]	Total [µL]	ACN [%]	equiv.	Conversion [%]
0.1		1	0.1	30	1	1	3.1
	0.1	1	0.1	30	1		0
0.1		1	0.5	30	5	5	26.9
0.05		1	0.25	30	2.5		19.3
	0.1	1	0.5	30	5		0
0.1		1	1.0	30	10	10	42.3
0.05		1	0.5	30	5		38.1
	0.1	1	1.0	30	10		0
0.1		1	1.5	30	15	15	49.5
0.05		1	0.75	30	7.5		49.6
	0.1	1	1.5	30	15		0
0.1		1	2	30	20	20	56.3
0.05		1	1	30	10		54.3
	0.1	1	2	30	20		0
0.1		1	2.5	30	25	25	49.2
0.05		1	1.25	30	12.5		60
	0.1	1	2.5	30	25		0
0.1		1	3.0	30	30	30	52.7
0.05		1	1.5	30	10		52.7
	0.1	1	3.0	30	30		0
0.1		1	4.0	30	40	40	47.6
	0.1	1	4.0	30	40		0
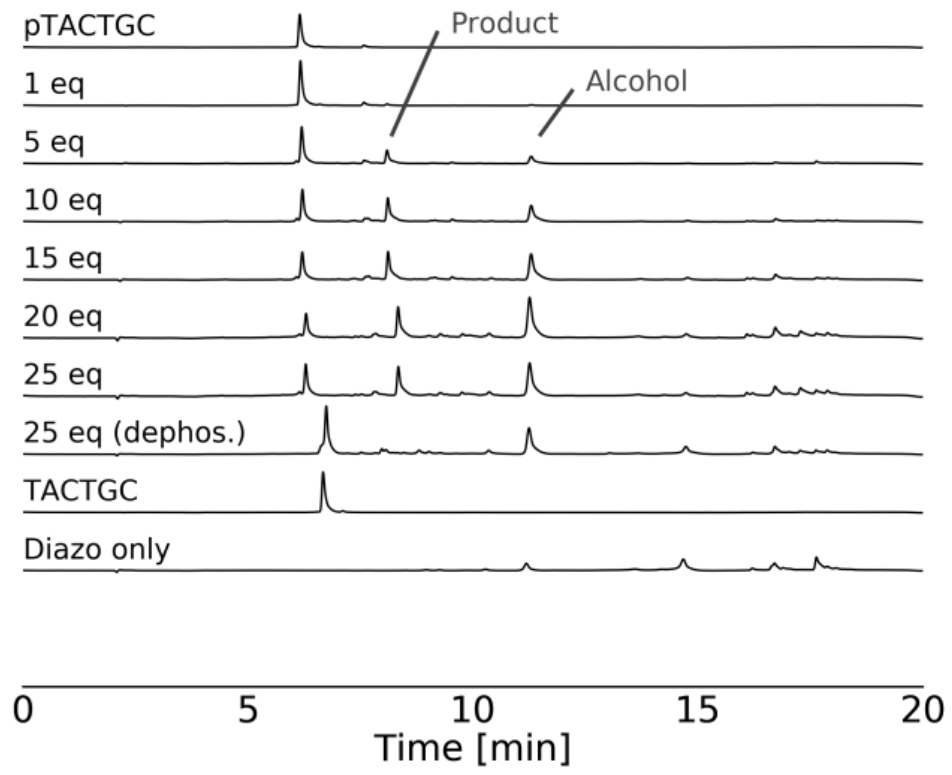
Figure S3.1 HPLC trace of 0.1 mM 5'-phospahte hexanucleotide modification
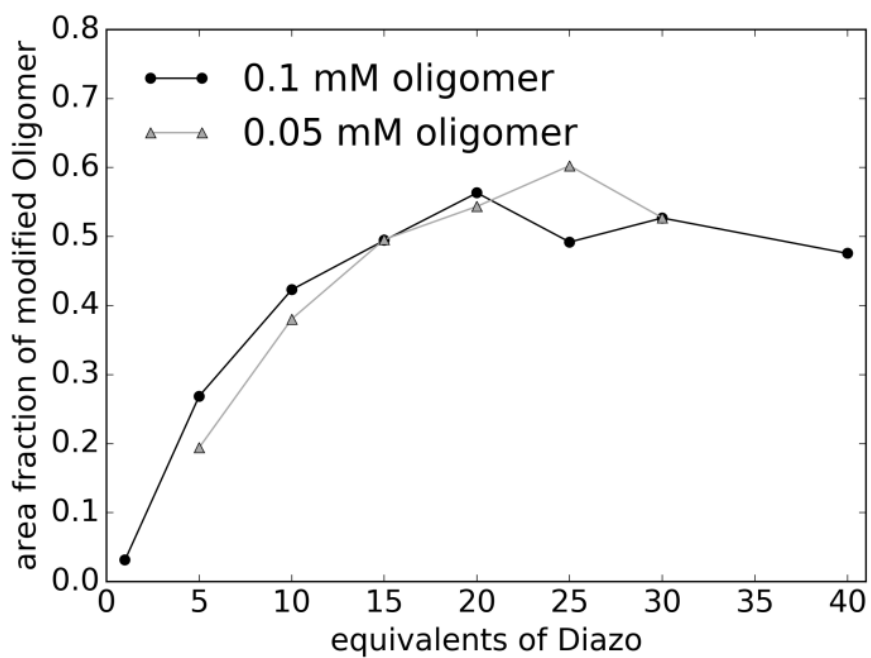
Figure S3.2 Alkylation of 5'-phospahte hexanucleotide with varying equiv. of diazo reagent

MS-MS 867 could be [T*.A.C-b(T)]⁻ or [T.A*.C-b(T)]⁻ or [T.A.C*-b(T)]⁻
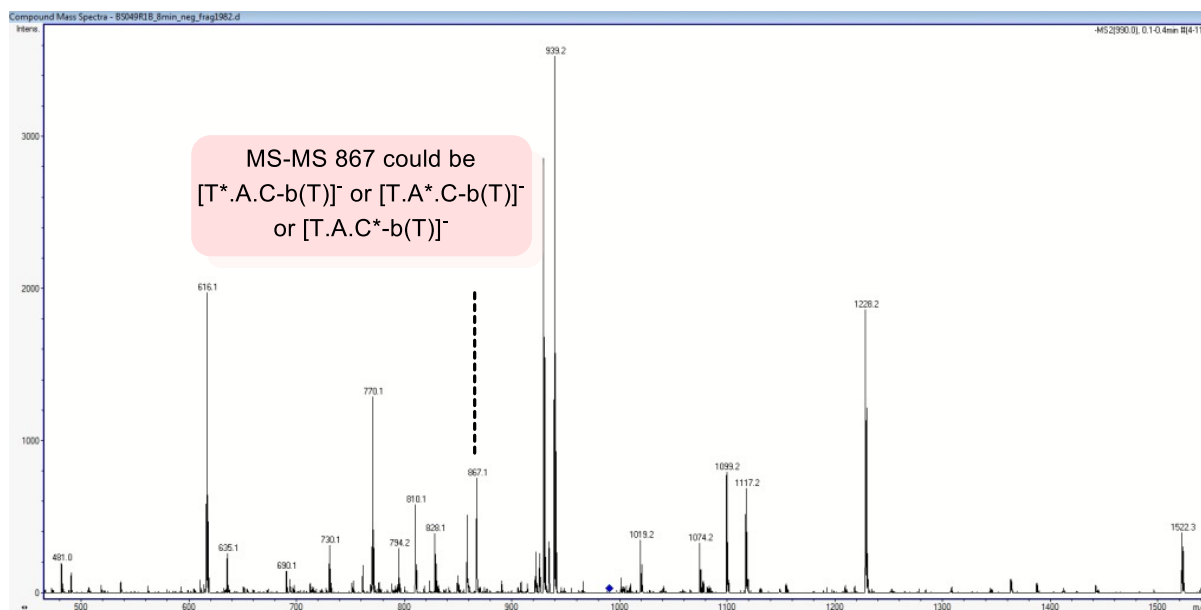
Figure S3.3 5'-PO$_4$-TACTGC modification
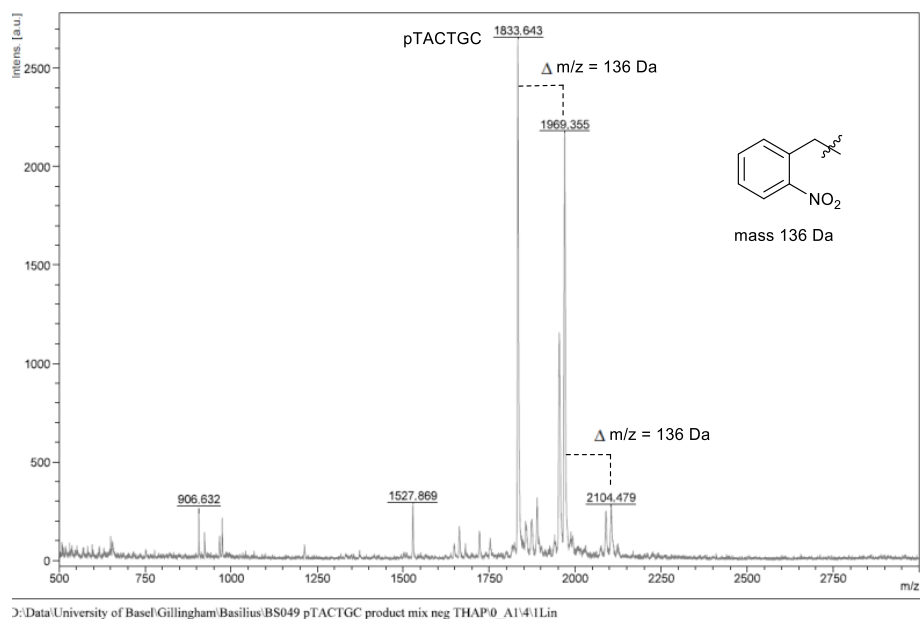
Figure S3.4 MALDI TOF spectrum of 5'-PO$_4$-TACTGC modification with 40 equiv. of diazo reagent
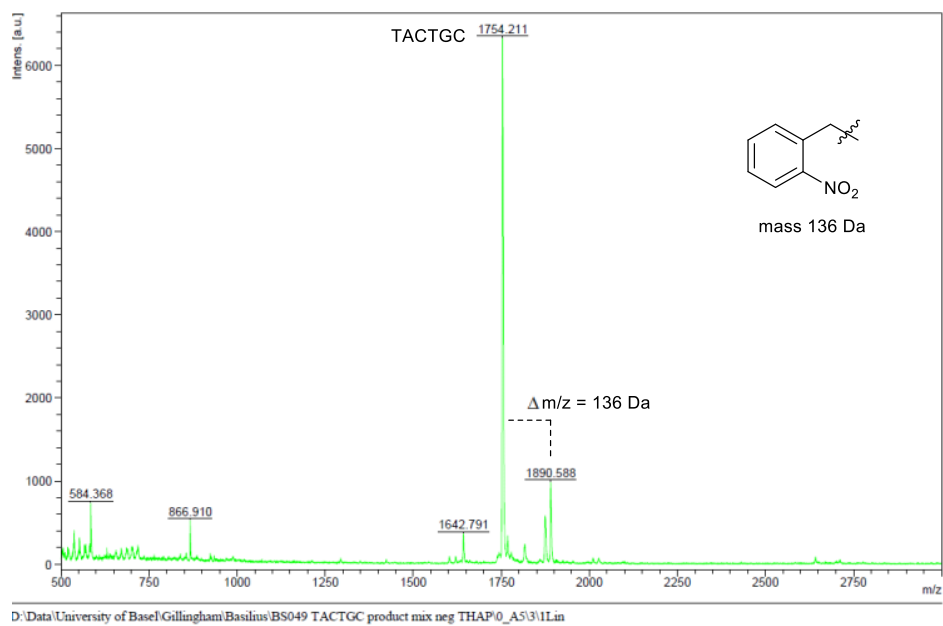
Figure S3.5 MALDI TOF spectrum of TACTGC modification with 40 equiv. of diazo reagent

3.4.8 Modification of phosphopeptide

All the reactions were carried out by adding the diazo compound to the peptide aqueous solution. The yellow reaction mixture turned colorless upon complete diazo consumption. It took 15-30 min for the following reactions. Since the diazo side products were found to be insoluble in water, diethyl ether (2 x 0.2 mL) was added to wash away the side products. Afterwards the aqueous solution was subjected to a gentle nitrogen flow for a minute to remove residual diethyl ether followed by UPLC-MS analysis (Method B) of the aqueous reaction mixture. The modification position was confirmed by collecting the corresponding HPLC peaks followed by MS-MS analysis. The collected fractions were resubmitted to UPLC-MS to confirm the identity of the purified peak. Sequencing of the modified peptide was carried out by standared MS/MS method.[140] The calculation was done by mMass software (http://www.mmass.org/).

13mer peptide pp60 Src (521-533)
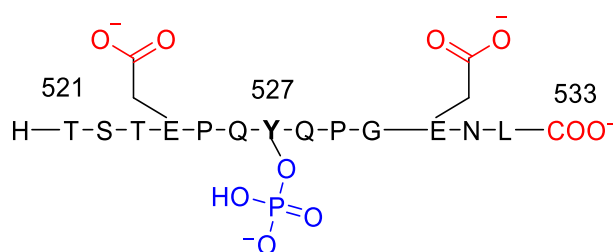
Phosphorylated peptide pp60 Src (521-533) was purchased from Bachem as TFA salt. It was dissolved in water to give a 5 mM stock solution (pH = 3.3).

The modification of the peptide is carried out at pH 6, however it didn't show selectivity of the phosphate modification. So the reaction was investigated at different pH.

Stock solution: 5 mM peptide in water, 50 mM diazo in CH_3CN, 5 mM MES buffer pH 6.05, 5 mM MOPS buffer pH 7.05.

1: 190 μL water, 50 μL MES buffer, pH 6.05, 5 μL peptide, 5 μL diazo

2: 190 μL water, 50 μL MOPS buffer, pH 7.05, 5 μL peptide, 5 μL diazo

3: 240 μL water, 5 μL peptide, 5 μL diazo, (pH = 4.2 due to TFA salt of the peptide)
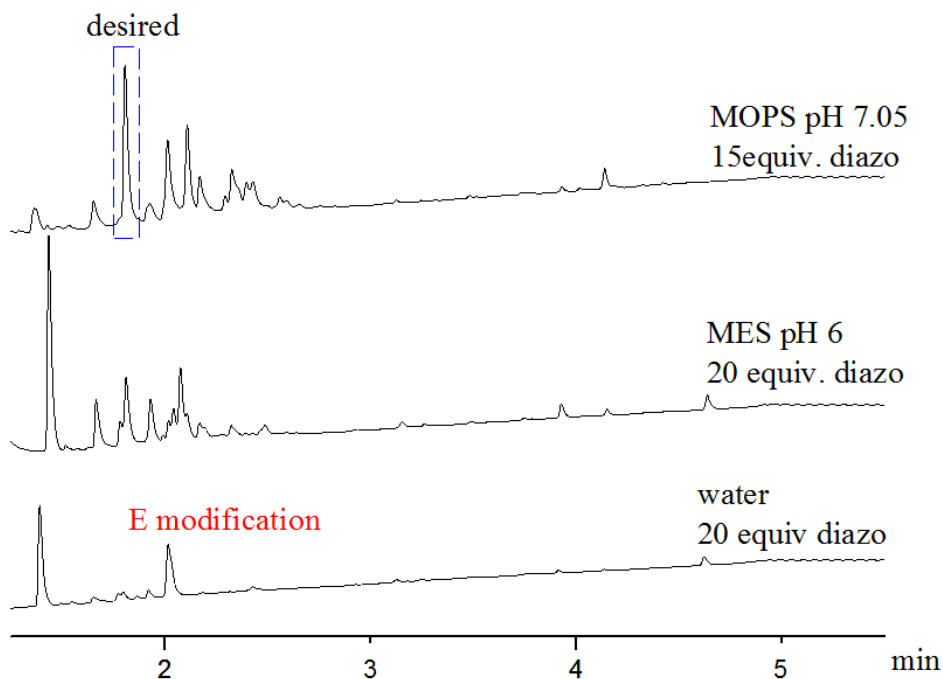
Figure S3.6 UPLC trace of the Src (521-533) modification at different pH

If the commercially available Src(521-533) TFA salt is dissolved in unbuffered water (final pH 4.2) and treated directly with diazo, alkylation of glutamic acid 531 (E531) was found to be the major product (see Figure S3.7). Consistent with the pK_a's controlling selectivity, performing the same reaction at pH 6 and 7, the alkylation preference switches to tyrosine 527 (Y527). However the addition of 15 equivalents diazo reagent leads to substantial amounts of double alkylation, thereby reducing the yield of the desired alkylation. The reaction was re-investigated with less equivalents of diazo reagent. (The sites of alkylation have been determined by MS/MS sequencing, see Table S3.9-S3.11 for details).

Stock solutions: 5 mM peptide in water, 10 mM diazo in CH_3CN, 5 mM MOPS buffer pH 7.05.

Final reaction mixture: 100 μM peptide, 1 mM MOPS pH 7.05, and diazo were added into the peptide solution one equiv. per time for the first 5 reactions.

1: 192 μL water, 50 μL MOPS buffer, pH 7.05, 5 μL peptide, 2.5 μL 10 mM diazo

2: 190 μL water, 50 μL MOPS buffer, pH 7.05, 5 μL peptide, 5 μL 10 mM diazo

3: 188 μL water, 50 μL MOPS buffer, pH 7.05, 5 μL peptide, 7.5 μL 10 mM diazo

4: 185 μL water, 50 μL MOPS buffer, pH 7.05, 5 μL peptide, 10 μL 10 mM diazo

5: 182 μL water, 50 μL MOPS buffer, pH 7.05, 5 μL peptide, 12.5 μL 10 mM diazo

6: 190 μL water, 50 μL MOPS buffer, pH 7.05, 5 μL peptide, 5 μL 50 mM diazo

7: 188 μL water, 50 μL MOPS buffer, pH 7.05, 5 μL peptide, 7.5 μL 50 mM diazo

Table S3.5 Src (521-533) alkylation with varying equivalents of diazo reagent

Entry	Diazo Equiv.	Phosphate modification	Conv.
1	1	19%	23%
2	2	35%	43%
3	**3**	**42%**	**52%**
4	4	47%	73%
5	5	44%	68%
6	10	47%	75%
7	15	40%	95%
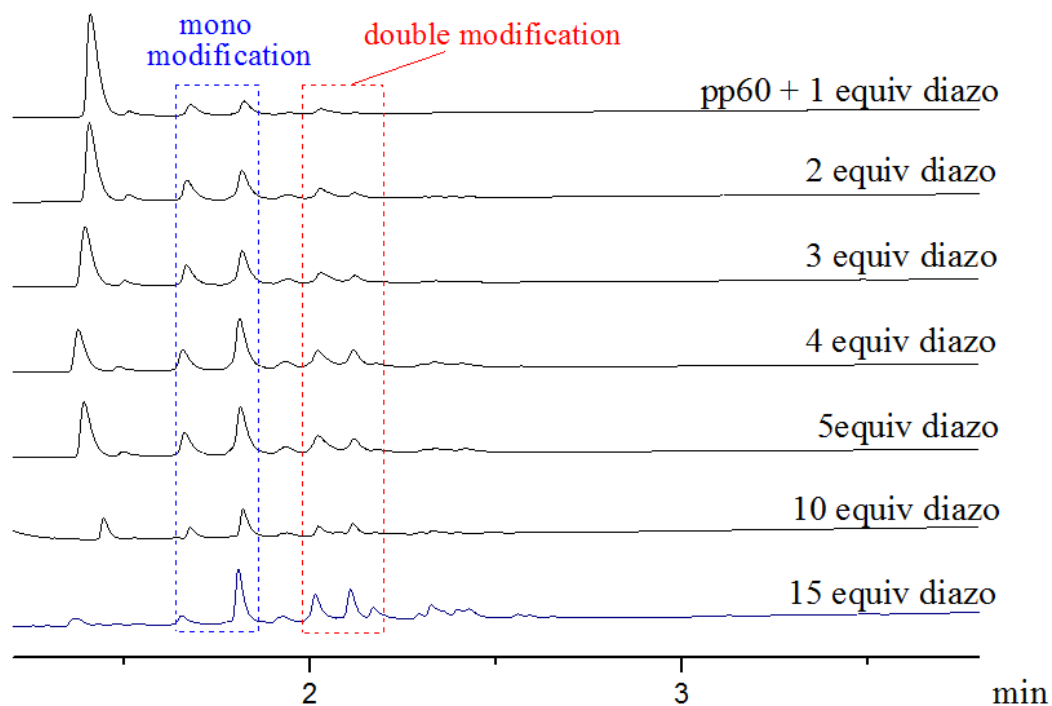
Figure S3.7 UPLC trace of the Src (521-533) modification with varying equiv. of diazo reagent

Table S3.6 summary of selective phosphopeptide modification

entry	substrate	pH[a]	equiv. diazo	Major protduct	conv.[b]	Y527: (E524+531)[c]
1	10	4.3	20	13	40	1:9[d]
2	10	7	2	12	43(35)	13:1
3	10	7	3	12	52(42)	15:1
4	10	7	5	12	68(44)	11:1
5	10	7	10	12	75(47)	11:1
6	10	7	15	12	95(40)	11:1
7	11	7	10	14	30	n.d.

[a]Run in 5 mM MOPS buffer at pH 7, pH 4 experiment is run in unbuffered water and the pH is measured after the reaction; [b]first number is conversion of starting peptide; number in parenthesis refers to amount of alkylated phosphate; [c]ratios are determined by HPLC intergration; [d]a small side product obscures phosphate alkylation hence we can only report in a lower limit.

The result of peptide modification with lower amount of diazo reagent is very interesting. With one equivalent diazo, only phosphate modification was achieved. Using three equivalents of diazo, it mainly modifies the phosphate of Y527 with minor double modification (see Figure S3.7 and Table S3.6). Although there are three carboxylates available for potential alkylation (C-terminal, E524 and E531), phosphate alkylation is the major product along with small portions of carboxylate modification (Y527: E=15:1, Table S3.6 entry 2-6). Since the pK_a of the C-terminal carboxylic acid is around 2-3, no alkylation of this position was observed. The increase of diazo equivalents leads to double alkylation as we observed before. The selectivity of phosphate and carboxylic acid alkylation was calculated according to the extracted ion abundance (See Figure S3.9-S3.13).

In addition, while 10 equiv. of diazo reagent leads to 75% conversion of the phosphorylated peptide the non-phosphorylated peptide is only partially modified under the same conditions (table S3.7).
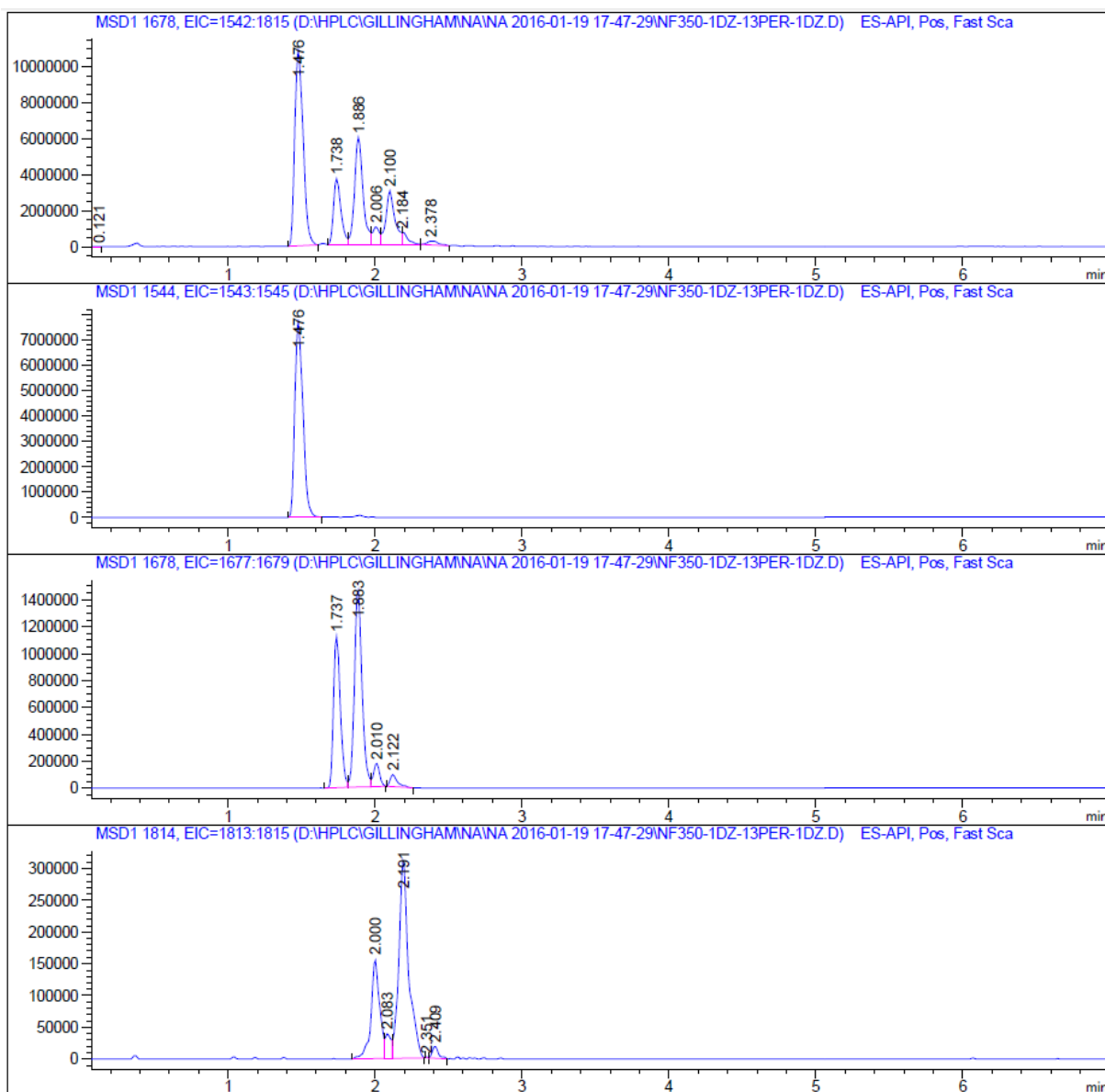
Figure S3.8 ion extraction of the reaction of Src (521-533) with 2 equiv. of diazo reagent
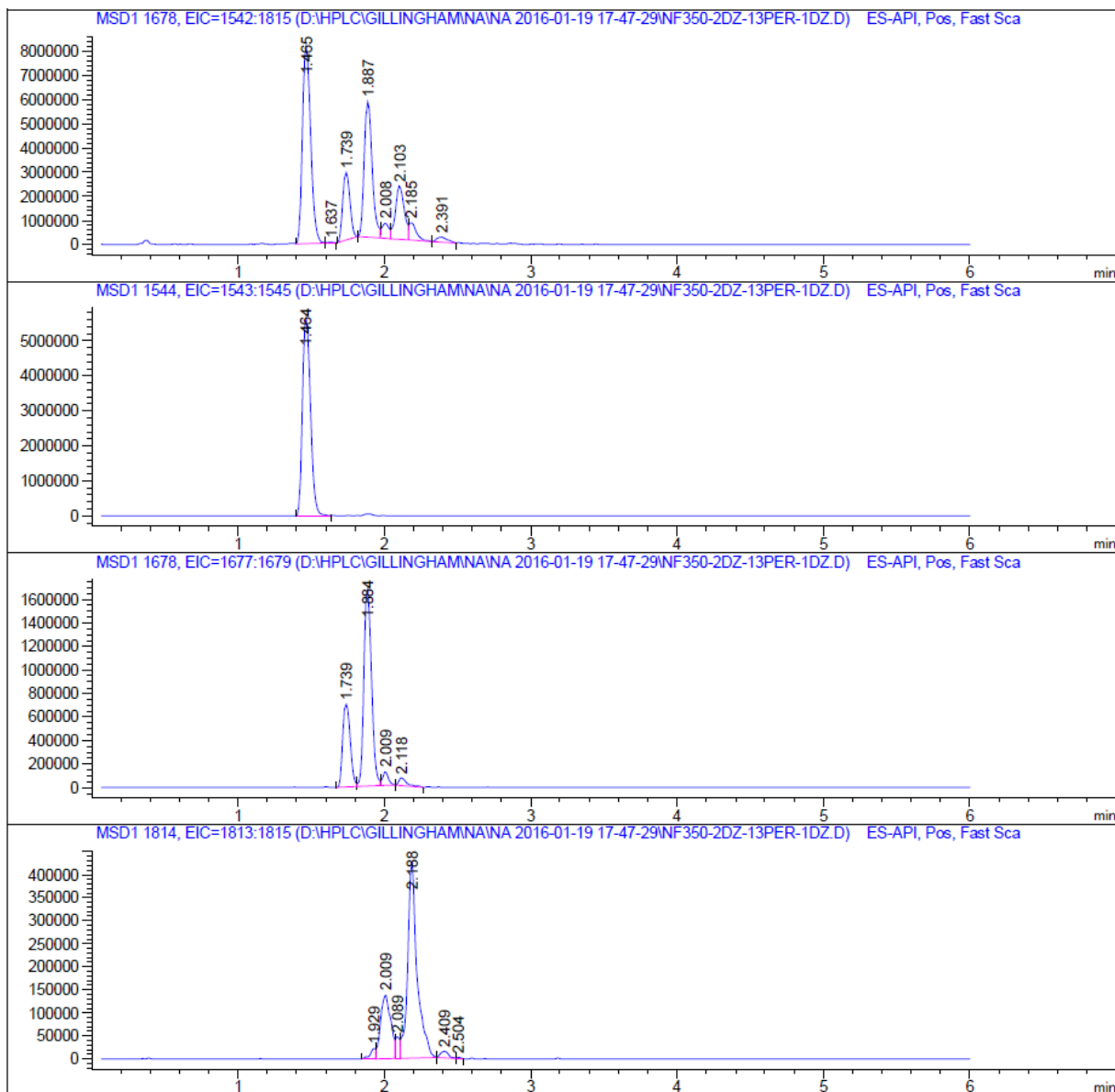
Figure S3.9 ion extraction of the reaction of Src (521-533) with 3 equiv. of diazo reagent
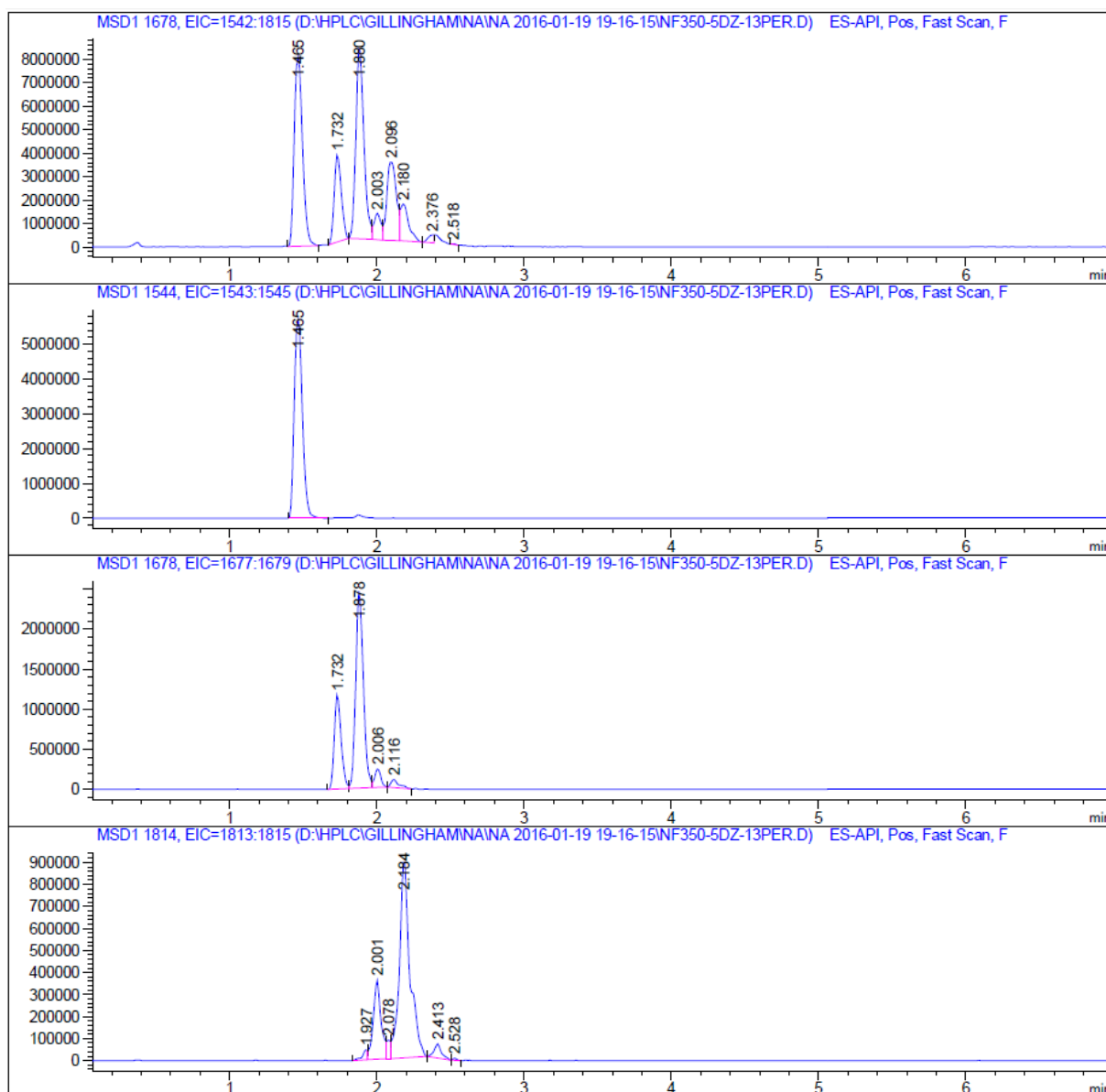
Figure S3.10 ion extraction of the reaction of Src (521-533) with 5 equiv. of diazo reagent
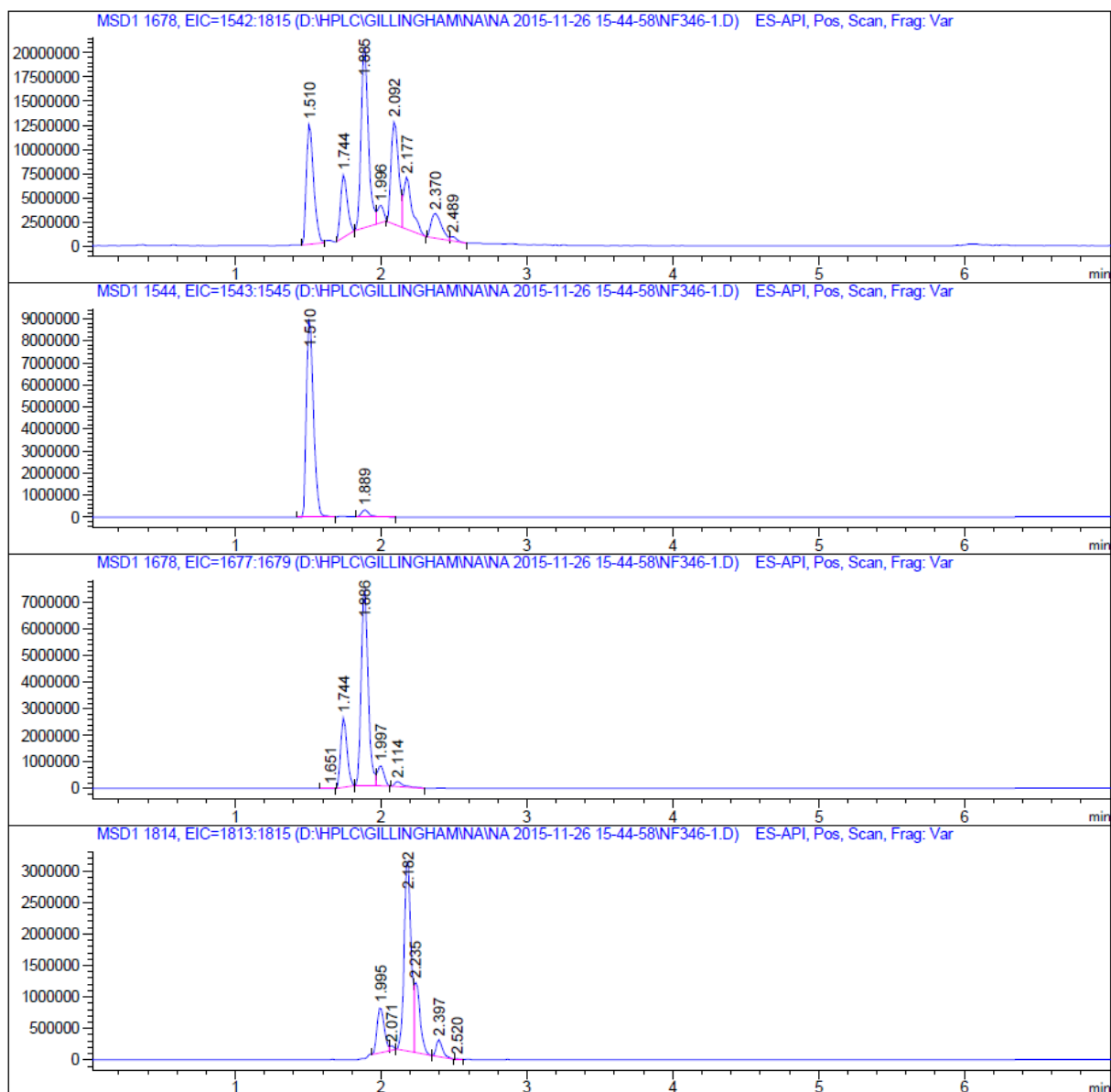
Figure S3.11 ion extraction of the reaction of Src (521-533) with 10 equiv. of diazo reagent
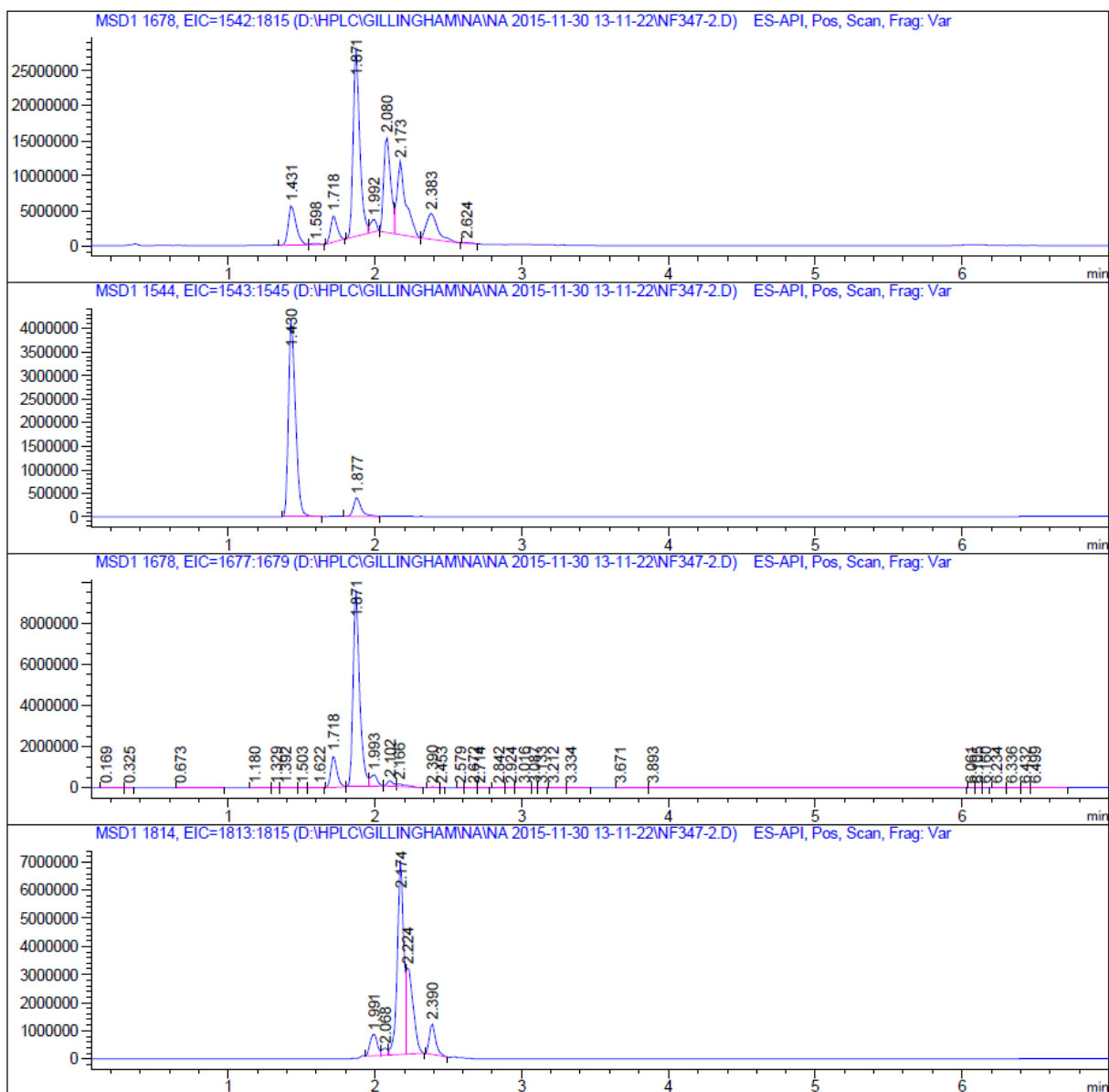
Figure S3.12 ion extraction of the reaction of Src (521-533) with 15 equiv. of diazo reagent

Peptide pp60 Src (521-533) non-phosphorylated
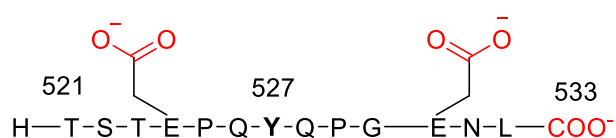
The original peptide without phosphate group was also investigated. The 13mer peptide could be modified with 10 equiv. of diazo reagent. The reaction was carried out as described before, 100 μM peptide, 1 mM MOPS buffer pH 7.05 followed by the addition of the diazo.

TSTEPQYQPGENL

2 equiv. diazo

10 equiv. diazo

20 equiv. diazo

| 1 | 2 | 3 | 4 | 5 | min |
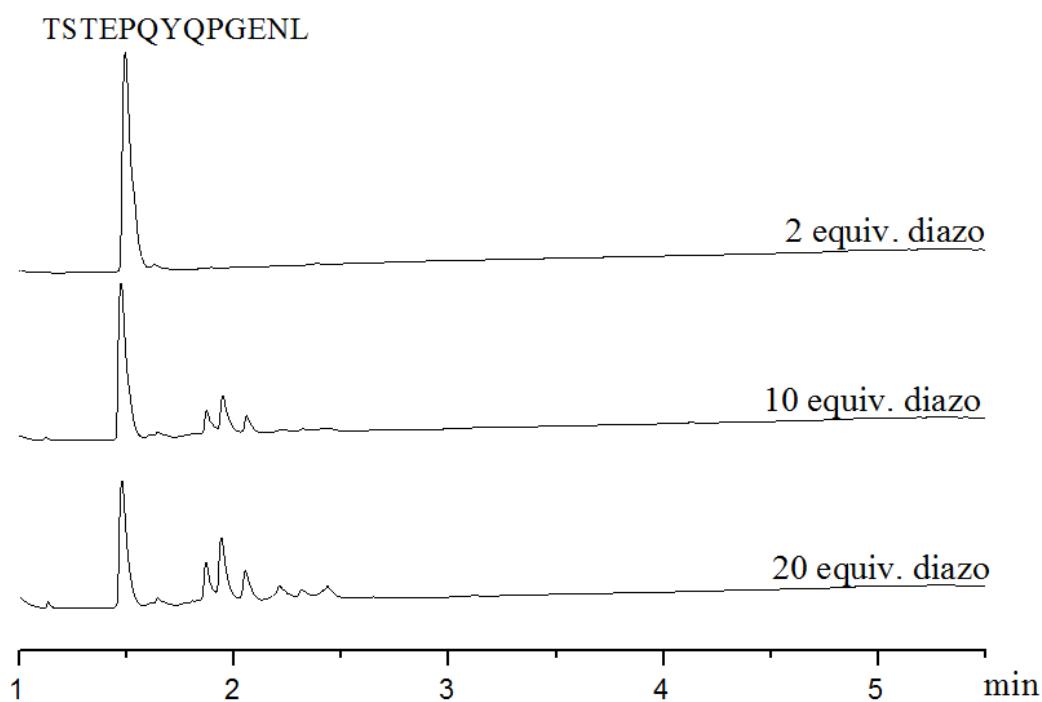
Figure S3.13 UPLC trace of Src (521-533) of modification

Table S3.7 Src (521-533 non phosphorylated form) modification

Entry	equiv. diazo	Conv.
1	2	0
2	10	30%
3	20	44%

MS-MS of modified pp60 Src products

The MS-MS was the fragmentation of the double charged products. In the following table, red is the observed MS which matches the theory MS/MS, the black is the caculated value.

Table S3.8 MS-MS fragments of phosphate modification TSTEPQYQPGENL
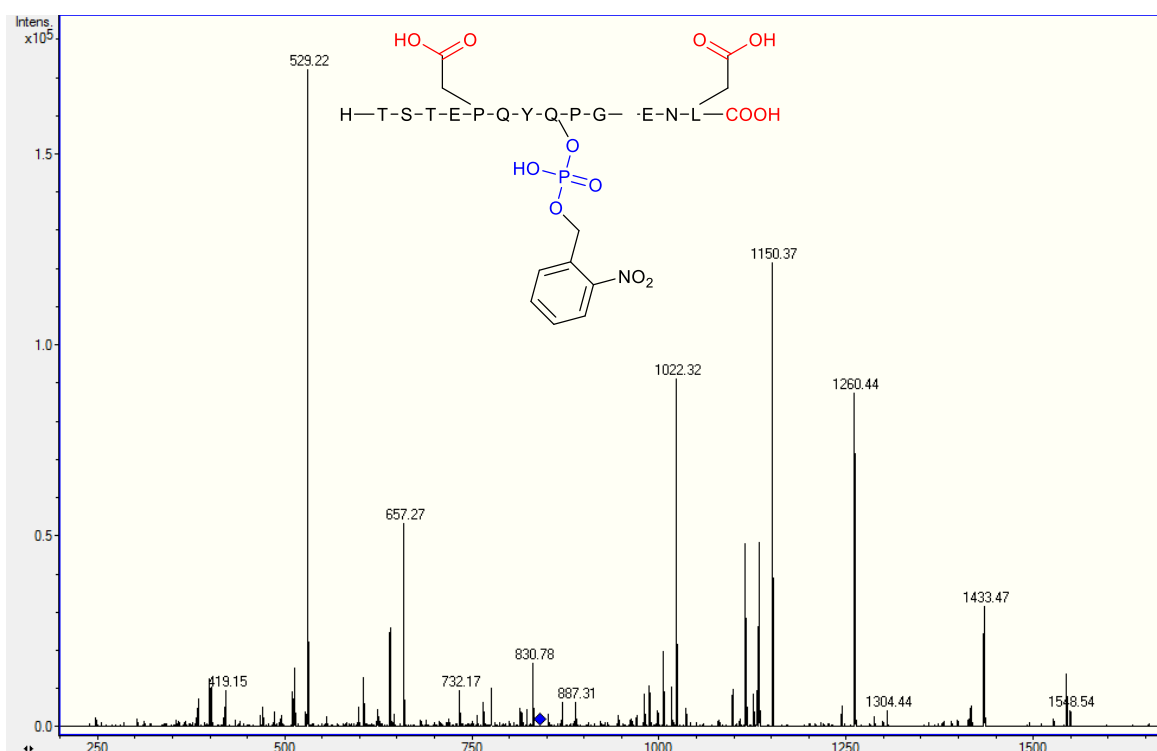
ion	slice	m/z	z	sequence
M	[1-13]	1678.663	1	.TSTEPQYQPGENL. [1x NO$_2$-Benzyl phosphate]
b2	[1-2]	189.087	1	.TS.t
b3	[1-3]	290.1347	1	.TST.e
b4	[1-4]	419.1773	1	.TSTE.p
b5	[1-5]	516.23	1	.TSTEP.q
b6	[1-6]	644.2886	1	.TSTEPQ.y
b7	[1-7]	1022.35	1	.TSTEPQY.q [1x NO$_2$-Benzyl phosphate]
b8	[1-8]	1150.409	1	.TSTEPQYQ.p [1x NO$_2$-Benzyl phosphate]
b9	[1-9]	1247.462	1	.TSTEPQYQP.g [1x NO$_2$-Benzyl phosphate]
b10	[1-10]	1304.483	1	.TSTEPQYQPG.e [1x NO$_2$-Benzyl phosphate]
b11	[1-11]	1433.526	1	.TSTEPQYQPGE.n [1x NO$_2$-Benzyl phosphate]
b12	[1-12]	1547.569	1	.TSTEPQYQPGEN.l [1x NO$_2$-Benzyl phosphate]
y12	[2-13]	1577.616	1	t.STEPQYQPGENL. [1x NO$_2$-Benzyl phosphate]
y11	[3-13]	1490.584	1	s.TEPQYQPGENL. [1x NO$_2$-Benzyl phosphate]
y10	[4-13]	1389.536	1	t.EPQYQPGENL. [1x NO$_2$-Benzyl phosphate]
y9	[5-13]	1260.493	1	e.PQYQPGENL. [1x NO$_2$-Benzyl phosphate]
y8	[6-13]	1163.441	1	p.QYQPGENL. [1x NO$_2$-Benzyl phosphate]
y7	[7-13]	1035.382	1	q.YQPGENL. [1x NO$_2$-Benzyl phosphate]
y6	[8-13]	657.3202	1	y.QPGENL.
y5	[9-13]	529.2617	1	q.PGENL
y4	[10-13]	432.2089	1	p.GENL.
y3	[11-13]	375.1874	1	g.ENL.
y2	[12-13]	246.1448	1	e.NL.
y1	[13-13]	132.1019	1	n.L.

Table S3.9 MS-MS of glutamic acid modification TSTEPQYQPGENL

ion	slice	m/z	z	sequence
M	[1-13]	1678.663	1	.TSTEPQYQPGENL. [1xPhosphate; 1x NO$_2$-benzyl]
b2	[1-2]	189.087	1	.TS.t
b3	[1-3]	290.1347	1	.TST.e
b4	[1-4]	419.1773	1	.TSTE.p
b5	[1-5]	516.23	1	.TSTEP.q
b6	[1-6]	644.2886	1	.TSTEPQ.y
b7	[1-7]	887.3183	1	.TSTEPQY.q [1xPhosphate]
b8	[1-8]	1015.377	1	.TSTEPQYQ.p [1xPhosphate]
b9	[1-9]	1112.43	1	.TSTEPQYQP.g [1xPhosphate]
b10	[1-10]	1169.451	1	.TSTEPQYQPG.e [1xPhosphate]
b11	[1-11]	1433.526	1	.TSTEPQYQPGE.n [1xPhosphate; 1x NO$_2$-benzyl]
b12	[1-12]	1547.569	1	.TSTEPQYQPGEN.l [1xPhosphate; 1x NO$_2$-benzyl]
y12	[2-13]	1577.616	1	t.STEPQYQPGENL. [1xPhosphate; 1x NO$_2$-benzyl]
y11	[3-13]	1490.584	1	s.TEPQYQPGENL. [1xPhosphate; 1x NO$_2$-benzyl]
y10	[4-13]	1389.536	1	t.EPQYQPGENL. [1xPhosphate; 1x NO$_2$-benzyl]
y9	[5-13]	1260.493	1	e.PQYQPGENL. [1xPhosphate; 1x NO$_2$-benzyl]
y8	[6-13]	1163.441	1	p.QYQPGENL. [1xPhosphate; 1x NO$_2$-benzyl]
y7	[7-13]	1035.382	1	q.YQPGENL. [1xPhosphate; 1xNO$_2$-benzyl]
y6	[8-13]	792.3523	1	y.QPGENL. [1x NO$_2$-benzyl]
y5	[9-13]	664.2937	1	q.PGENL. [1x NO$_2$-benzyl]
y4	[10-13]	567.2409	1	p.GENL. [1x NO$_2$-benzyl]
y3	[11-13]	510.2195	1	g.ENL. [1x NO$_2$-benzyl]
y2	[12-13]	246.1448	1	e.NL.
y1	[13-13]	132.1019	1	n.L.

Table S3.10   MS-MS of double modification TSTEPQYQPGENL
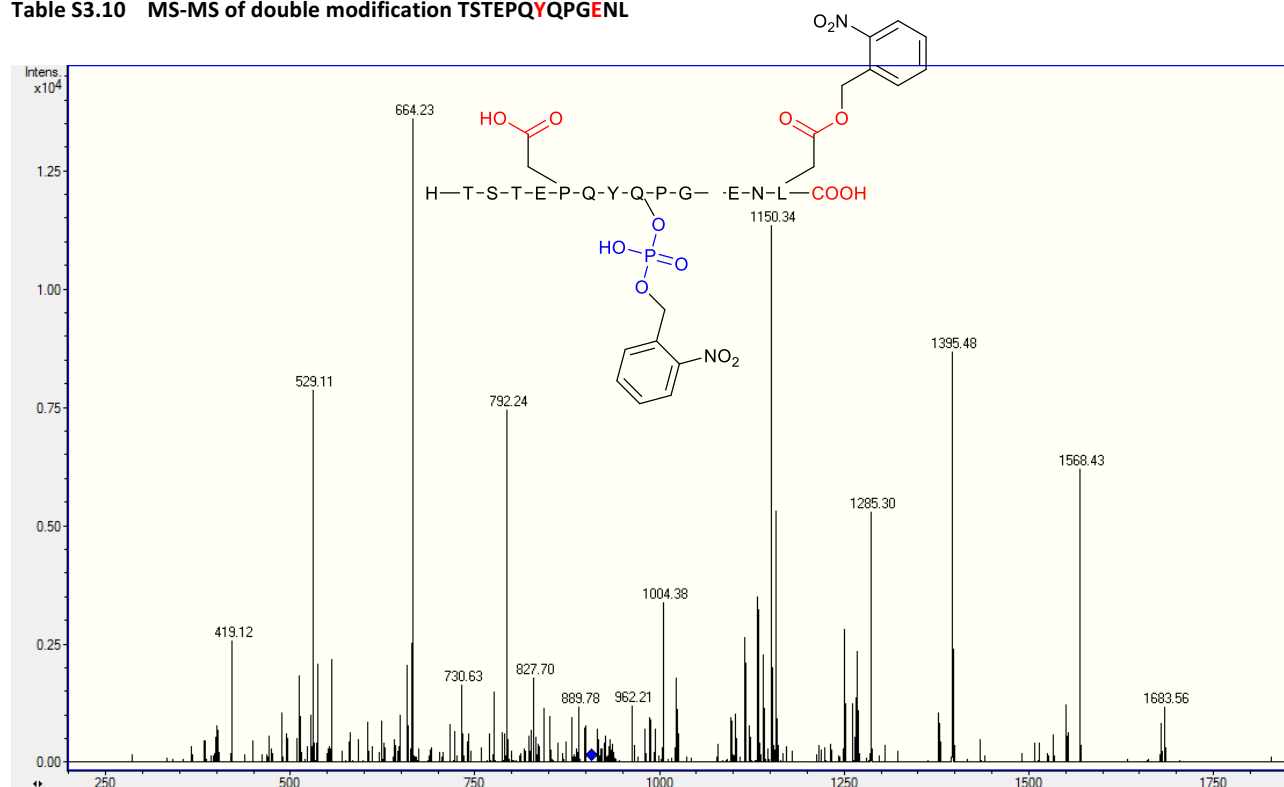
ion	slice	m/z	z	sequence
M	[1-13]	1813.695	1	.TSTEPQYQPGENL. [1xBenzyl phosphate; 1x NO$_2$-benzyl]
b2	[1-2]	189.087	1	.TS.t
b3	[1-3]	290.1347	1	.TST.e
b4	[1-4]	419.1773	1	.TSTE.p
b5	[1-5]	516.23	1	.TSTEP.q
b6	[1-6]	644.2886	1	.TSTEPQ.y
b7	[1-7]	1022.35	1	.TSTEPQY.q [1x NO$_2$-Benzyl phosphate]
b8	[1-8]	1150.409	1	.TSTEPQYQ.p [1x NO$_2$-Benzyl phosphate]
b9	[1-9]	1247.462	1	.TSTEPQYQP.g [1x NO$_2$-Benzyl phosphate]
b10	[1-10]	1304.483	1	.TSTEPQYQPG.e [1x NO$_2$-Benzyl phosphate]
b11	[1-11]	1568.558	1	.TSTEPQYQPGE.n [1x NO$_2$-Benzyl phosphate; 1xNO$_2$-benzyl]
b12	[1-12]	1682.601	1	.TSTEPQYQPGEN.l [1x NO$_2$-Benzyl phosphate;1xNO$_2$benzyl]
y12	[2-13]	1712.648	1	t.STEPQYQPGENL. [1x NO$_2$-Benzyl phosphate;1xNO$_2$-benzyl]
y11	[3-13]	1625.616	1	s.TEPQYQPGENL. [1x NO$_2$-Benzyl phosphate; 1x NO$_2$-benzyl]
y10	[4-13]	1524.568	1	t.EPQYQPGENL. [1x NO$_2$-Benzyl phosphate; 1x NO$_2$-benzyl]
y9	[5-13]	1395.525	1	e.PQYQPGENL. [1x NO$_2$-Benzyl phosphate; 1x NO$_2$--benzyl]
y8	[6-13]	1298.473	1	p.QYQPGENL. [1x NO$_2$-Benzyl phosphate; 1x NO$_2$-benzyl]
y7	[7-13]	1170.414	1	q.YQPGENL. [1x NO$_2$-Benzyl phosphate; 1x NO$_2$-benzyl]
y6	[8-13]	792.3523	1	y.QPGENL. [1x NO$_2$-benzyl]
y5	[9-13]	664.2937	1	q.PGENL. [1x NO$_2$-benzyl]
y4	[10-13]	567.2409	1	p.GENL. [1x NO$_2$-benzyl]
y3	[11-13]	510.2195	1	g.ENL. [1x NO$_2$-benzyl]
y2	[12-13]	246.1448	1	e.NL.
y1	[13-13]	132.1019	1	n.L.

Photo-cleavage of modified pp60 Src (527)

The purified Y527 alkylation product (which is manually collected by HPLC from a 50 nmol reaction setup) was re-dissolved in 100 μL water in an Eppendorf tube. The sample was irradiated with a UV lamp (366 nm) over night. Under this condition the modified product was cleaved affording the original phosphorylated Src (521-533) along with 11% of dephosphorylated peptide. The reaction was monitored by UPLC-MS (method B).

the purified product

366nm over night
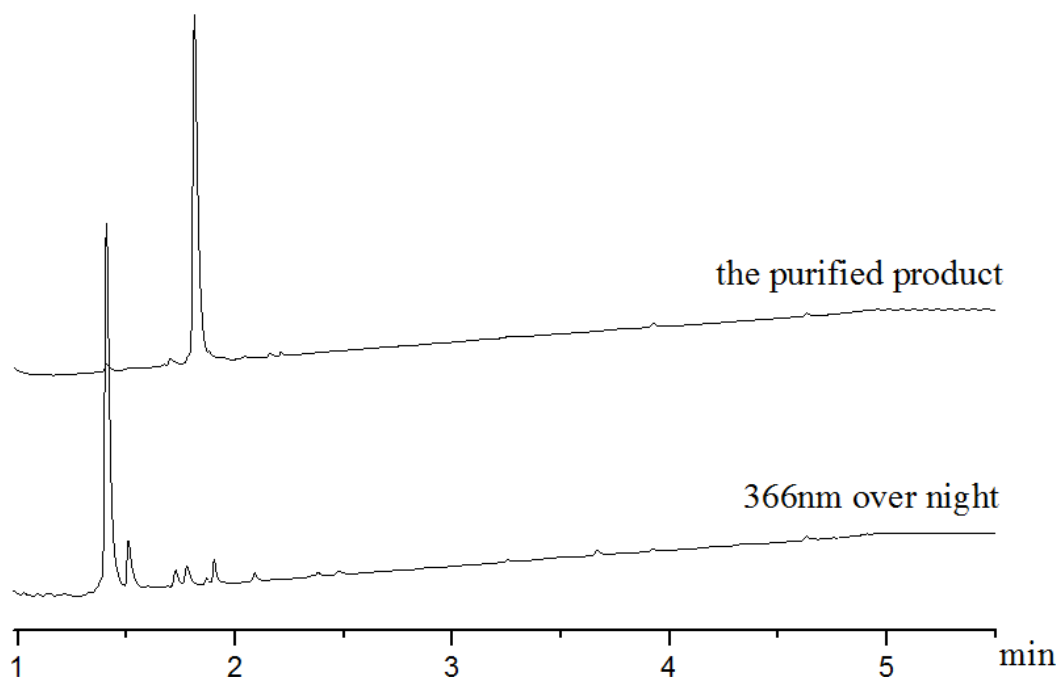
Figure S3.14 Photocleavage of caged Src (521-533)

5mer peptide
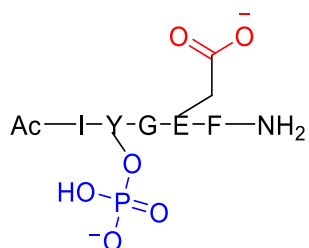
5mer peptide Ac-Ile-Tyr(PO$_3$H$_2$)-Gly-Glu-Phe-NH$_2$ was supplied from Bachem as ammonium salt. It is a phosphorylated synthetic peptide which can be used as an inhibitor of Src. The peptide was dissolved in water to give a 10 mM stock solution. The modification with different equivalents of diazo reagent was carried out as follows:
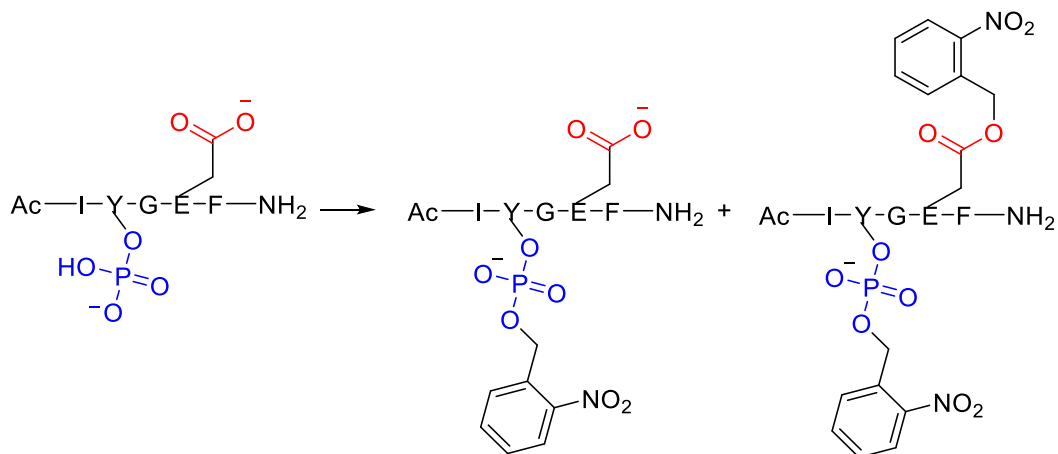
Stock solution

10 mM peptide, 5 mM MOPS pH 7.1, 10 mM or 100 mM diazo in CH_3CN.

It turned out that, with 3 equiv. diazo, phosphorylated peptide was selectively modified at its phosphate position 43% of conversion (see Figure S3.16). The desired product was manually collect by HPLC and confirmed using peptide MS-MS sequencing methods (See table S3.11-12).

1

200 μL reaction volume

4 μL peptide, 80 μL MOPS buffer pH 7.10, 6 μL diazo, 110 μL water

200 μM peptide, 2 mM MOPS, 3 equiv. Diazo

43% of the peptide converted to the desired monophosphate modified product

2

200 μL reaction volume

4 μL peptide, 80 μL buffer pH 7.10, 4 μL diazo, 112 μL water

200 μM peptide, 2mM MOPS, 10 equiv. diazo

56% of peptide converted to desired monophosphate modified product

14% of the peptide converted to double modified product, (both phosphate and glutamic acid)

3

200 μL reaction volume

4 μL peptide, 80 μL buffer pH 7.10, 40 μL diazo, 76 μL water

200 μM peptide, 2 mM MOPS, 100 equiv. Diazo

56% of peptide converted to desired monophosphate modified product

27% of the peptide converted to double modified product, (both phosphate and glutamic acid)
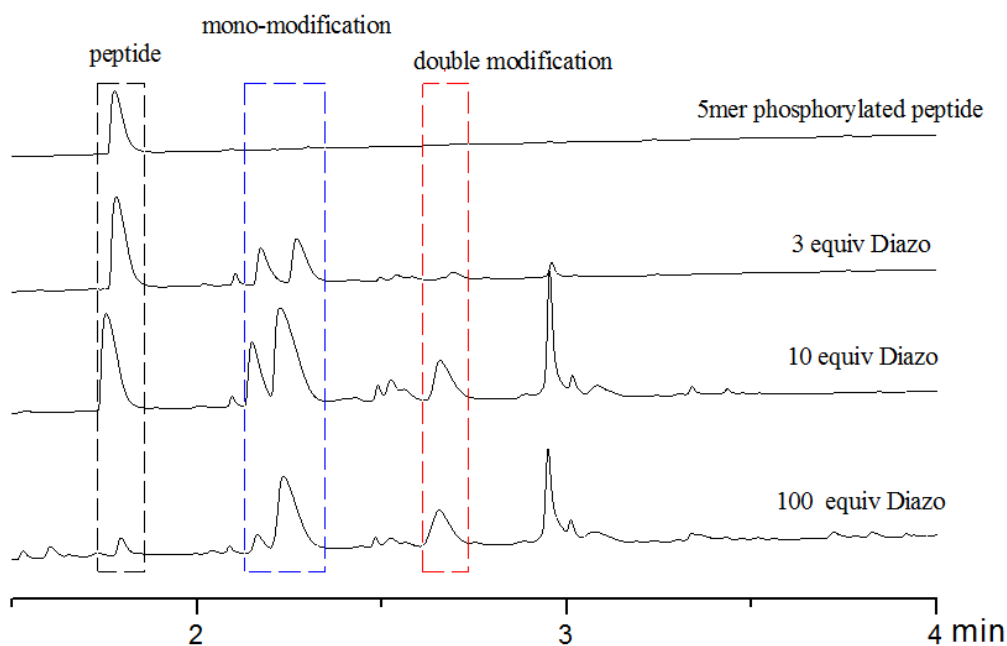
Figure S3.15 UPLC trace of the modification reaction using the 5mer peptide

5mer peptide dephosphorylated form Ac-IYGEF-NH$_2$
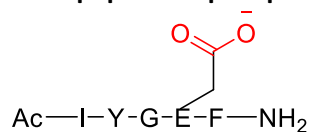
5mer peptide was obtained from Bachem as ammonium salt and was dissolved in DMSO as stock solution due its low solubility in water. The modification with different equiv. of diazo reagent was carried out in analogy to the setup for the phosphorylated peptide, however no modification was observed even when treated with 20 equiv. of diazo reagent.
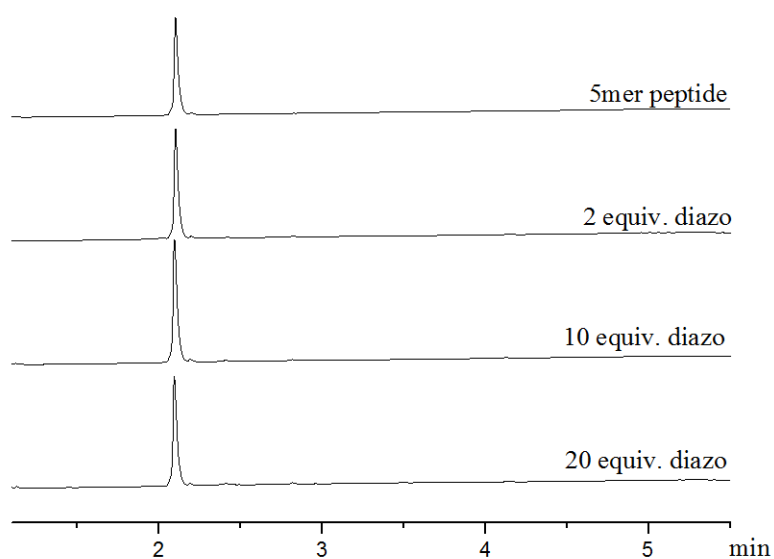
Figure S3.16 UPLC trace of 5mer peptide modification

MS/MS of modified products

Table S3.11 MS-MS of phosphate modified product AcIYGEF
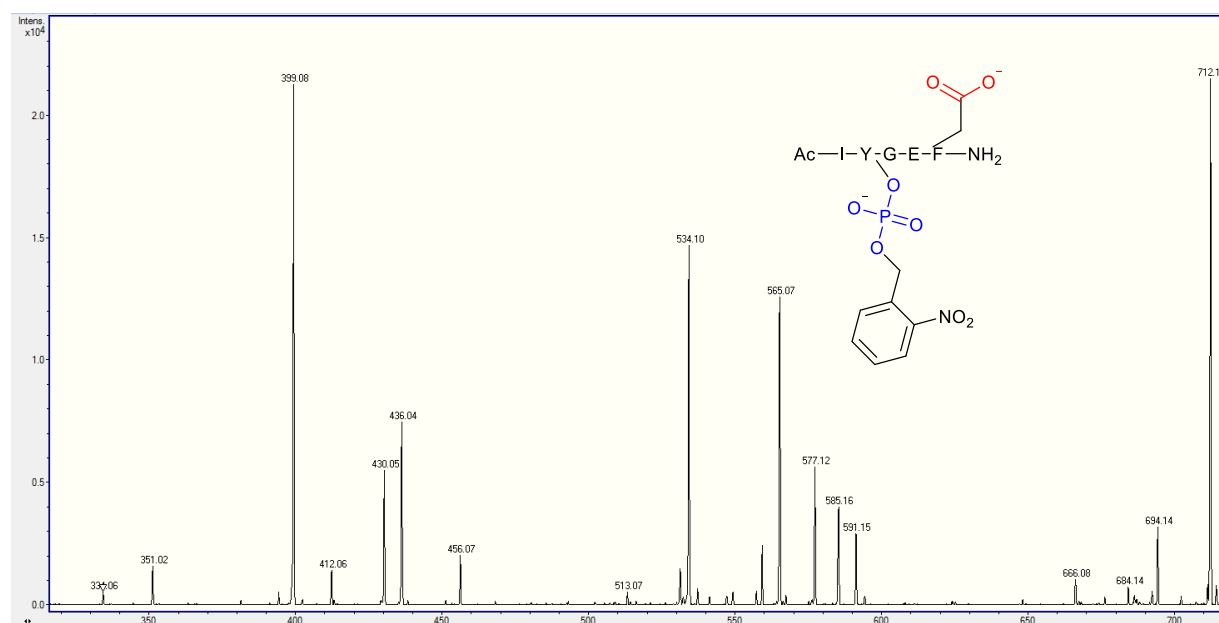
ion	slice	m/z	z	sequence
b2	[1-2]	534.1636	1	.IY.g [1xAcetyl; 1xNO$_2$-Benzyl phosphate]
b3	[1-3]	591.1851	1	.IYG.e [1xAcetyl; 1xNO$_2$-Benzyl phosphate]
b4	[1-4]	720.2276	1	.IYGE.f [1xAcetyl; 1xNO$_2$-Benzyl phosphate]
y4	[2-5]	729.228	1	i.YGEF. [1xAmide; 1xNO$_2$-Benzyl phosphate]
z4	[2-5]	712.2014	1	i.YGEF. [1xAmide; 1xNO$_2$-Benzyl phosphate]
y3	[3-5]	351.1663	1	y.GEF. [1xAmide]
z3	[3-5]	334.1397	1	y.GEF. [1xAmide]
y2	[4-5]	294.1448	1	g.EF. [1xAmide]
z2	[4-5]	277.1183	1	g.EF. [1xAmide]
y1	[5-5]	165.1022	1	e.F. [1xAmide]
z1	[5-5]	148.0757	1	e.F. [1xAmide]

Table S3.12 MS-MS of double modified product AcI*YGE*F
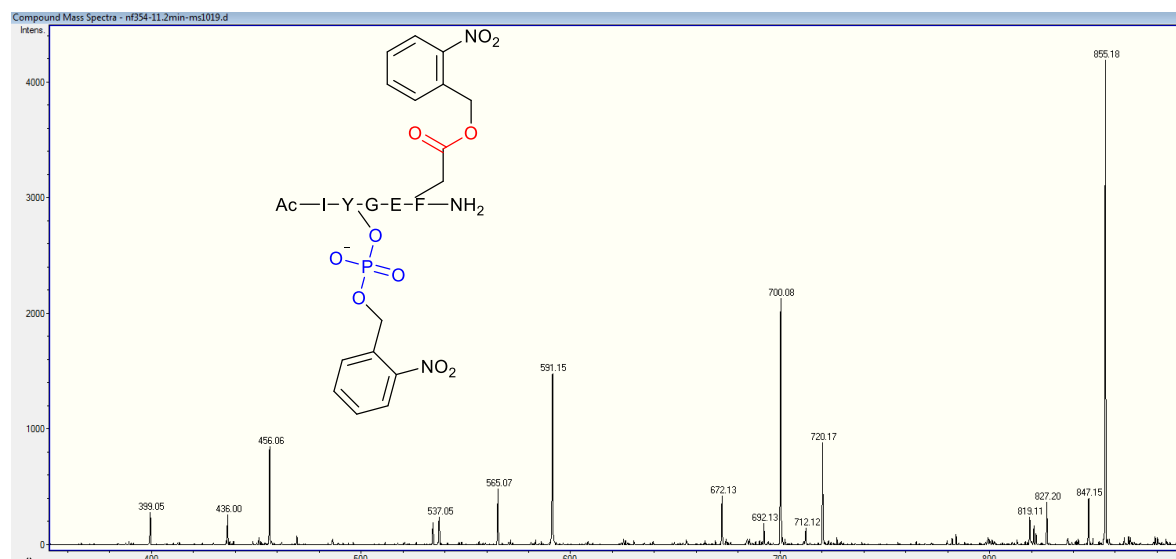
ion	slice	m/z	z	sequence
b2	[1-2]	534.1636	1	.IY.g [1xAcetyl; 1xNO$_2$-Benzyl phosphate]
b3	[1-3]	591.1851	1	.IYG.e [1xAcetyl; 1xNO$_2$-Benzyl phosphate]
b4	[1-4]	855.2597	1	.IYGE.f [1xAcetyl; 1xNO$_2$-Benzyl phosphate; 1xNO$_2$-benzyl]
y4	[2-5]	864.26	1	i.YGEF. [1xAmide; 1xNO$_2$-Benzyl phosphate; 1xNO$_2$-benzyl]
z4	[2-5]	847.2335	1	i.YGEF. [1xAmide; 1xNO$_2$-Benzyl phosphate; 1xNO$_2$-benzyl]
y3	[3-5]	486.1983	1	y.GEF. [1xAmide; 1xNO$_2$-benzyl]
z3	[3-5]	469.1718	1	y.GEF. [1xAmide; 1xNO$_2$--benzyl]
y2	[4-5]	429.1769	1	g.EF. [1xAmide; 1xNO$_2$-benzyl]
z2	[4-5]	412.1503	1	g.EF. [1xAmide; 1xNO$_2$-benzyl]
y1	[5-5]	165.1022	1	e.F. [1xAmide]
z1	[5-5]	148.0757	1	e.F. [1xAmide]

Chapter 4 Investigation of substrate tolerance of RNA demethylase FTO

4.1 Reversible DNA methylation in gene regulation

Methylation is a widespread chemical modification in DNA and RNA which plays a critical signaling role in gene regulation.[141] In mammalian cells, DNA methylation typically occurs mainly at the five positon of cytosine of CpG dinucleotides which can impede the binding of transcriptional proteins leading to gene silencing (Figure 4.1).[142] A similar process with adenine methylation has been discovered on mRNA giving the weight to the possibility of regulation of transcriptome.[143] The reversible methylation of DNA and RNA in gene expression regulation is responsible for cellular differentiation, which is not only critical in embryonic development, but also in aging related diseases. Understanding how gene expression is regulated and maintained at a stable level is essential for developing therapeutic strategies.

The mechanism of how DNA methylation patterns regulate tissue-specific gene expression and large-scale gene silencing is not yet fully understood.[144, 145] But fundamentally, most chromatin exists in a condensed, transcriptionally silent state as heterochromatin which has high levels of DNA methylation (Figure 4.1, upper). Euchromatin, the lightly packed form of chromatin, usually has low DNA methylation levels and less condensed. It contains most actively transcribed genes and promotes the tissue-specific gene expression (Figure 4.1, lower). Genome-wide studies demonstrated 5-methylcytosine (m^5C) is widespread throughout the mammalian genome and 70%-80% mammalian CpG sites are methylated.[146] A global loss of DNA methylation was first found in different organs of rats during aging,[147] and was later also confirmed in human genome by a growing body of literature.[148] The decrease of DNA methylation during aging is thought to be responsible for the activation of transcription and gene overexpression, as one mechanism leads to a higher incidence of age-related diseases including cancer.[149]
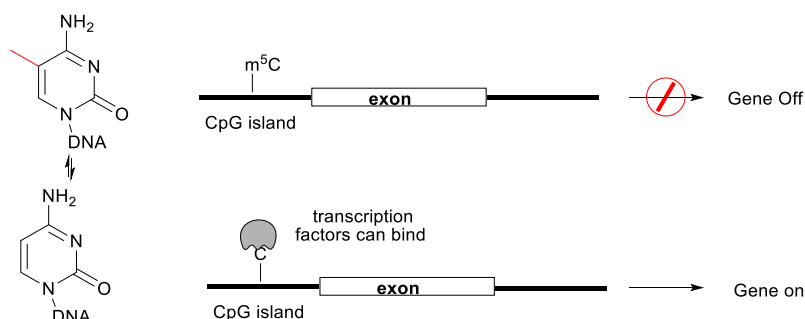
Figure 4.1 DNA methylation regulates gene expression. Upper: high levels of DNA methylation at gene promoters may inhibit the binding of transcription factors or recruiting transcriptional corepressors resulting in inactive genes. Lower: low levels of DNA methylation have been proposed to generate active genes through increased binding of transcriptio factors.

4.2 The discovery of FTO as m⁶A RNA demethylase

Although the reversible methylation of DNA as epigenetic marks has been intensively studied and well understood, the similar chemical modification of RNA starts to gain increased appreciation in gene regulation only recently. But different than m^5C in DNA, N^6-Methyladenosine (m^6A) is the most abundant modification in RNA.[150] Even though this methylation was first found in virus in 1974,[151] later it was detected in all higher eukaryotes, and is present ~3 times per mRNA on average in mammalian cells.[152, 153] The biological function of m^6A remained elusive for decades. Only very recently a major breakthrough of discovering the first m^6A RNA demethylase FTO, which is a fat mass and obesity-associated protein, revealed the reversible regulatory mechanism of m^6A at the molecular level for the first time.[154]

The FTO gene was first found in mice by positional cloning in 1999[155] and later drew attention because its misregulation is associated with obesity (Figure 4.2).[156] Genome-wide association studies in multiple populations demonstrated that FTO variant alleles (homozygous = AA and heterozygous = AT) are predisposed to greater adiposity than are those with 2 wild-type alleles (TT).[156-159] Carriers of one copy of the allele (AT) weighed on average 1.2 kilograms more than people with no copies (TT). The 16% of European adults who are homozygous (AA) weighed about 3 kilograms more and had 1.67-fold increased odds of obesity when compared with those not inheriting a copy (TT). Thus the FTO gene was identified as the first obesity risk allele.

Due to the important role of FTO in obesity, it triggered a great interest towards its enzymatic activity. The FTO protein was first shown to demethylate N^3-methylthymidine in single-stranded DNA[160] and N^3-methyluridine in single-stranded RNA[161] *in vitro*. Later the crystal structure[162] of the FTO protein revealed an active domain that is similar to other proteins of the AlkB family which are DNA repair enzymes protecting the genome from alkylation damage. Additionally, the C-terminal domain of the FTO protein showed the preference of binding ssRNA and ssDNA over dsDNA, indicating the demethylation target might be RNA under physiological conditions. It was not until 2011 that the FTO protein was characterized as a m^6A demethylase in RNA expression regulation.[154] But the regulatory mechanism of m^6A RNA in protein translation and the FTO mediated development of obesity in humans remains poorly understood. A positive relationship between the number of FTO risk alleles and plasma ghrelin levels was found which might indicate, the 'hunger hormone' ghrelin's mRNA is one of the FTO substrates (Figure 4.2).[163] People with the obesity-risk FTO variant alleles have higher circulating levels of ghrelin in their blood due to the high level of ghrelin mRNA, which could be a result of demethylation of methylated ghrelin mRNA by high levels of the FTO protein. Therefore people with FTO gene are prone to hunger due to the high level of ghrelin and prefer an energy dense diet which in the end leads to obesity.[163]
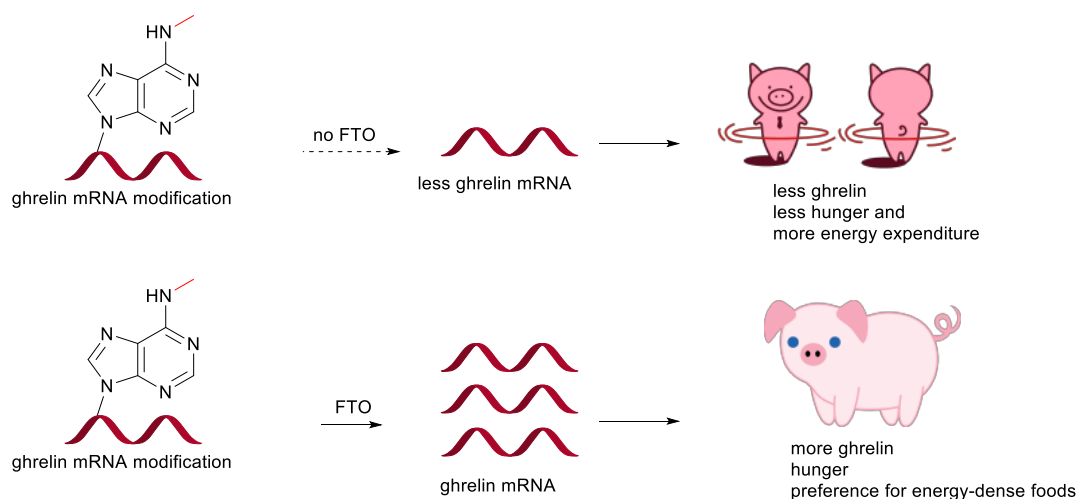
Figure 4.2 FTO confers an increased risk of obesity by subtly changing food intake and preference.

In the past 5 years, the development of chemical methods made the specific modified nucleic acid monomers available. Combined with advanced chemical analysis techniques, He's group could confirm that an oxidative RNA demethylation process is catalyzed by the FTO protein generating N^6-hydroxymethyladenosine (hm^{6}A) as an intermediate modification, and N^6-formyladenosine (f^{6}A) as a further oxidized product (Scheme 4.1)[154]. Chemically synthesized hm^{6}A and f^{6}A standards co-elute with intermediates generated in the FTO catalyzed oxidation process in the HPLC trace and have the same MS/MS fragmentation pattern. Furthermore, they showed hm^{6}A and f^{6}A are new RNA modifications in mammalian cells. The discovery of hydroxymethylated and formylated adenosines in mRNA further expanded the repertoire of chemical diversity of RNA modifications.
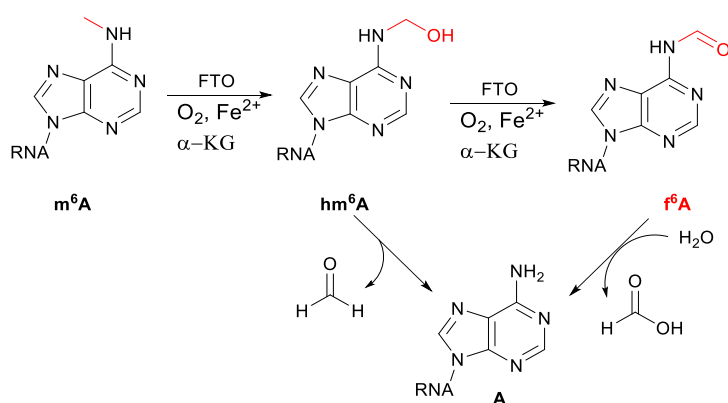
Scheme 4.1 FTO demethylates m^{6}A through oxidization of m^{6}A to hm^{6}A, with f^{6}A as a further oxidized product.

The recent advances[164,165] in sequencing of methylation at single-base resolution provide a comprehensive view of genomic methylation patterns, and should promote progress in understanding the chemistry of reversible methylation.[166] Lastest studies have highlighted three DNA/RNA demethylases families which are responsible for gene regulation and repair (Figure 4.3). The FeII/α-

ketoglutarate (α-KG)-dependent AlkB family demethylases mainly catalyze oxidative demethylation of N-alkylated nucleic acid bases including monoalkyl and exocyclic bridged adducts.[167] For simple alkyl substrates, AlkB oxidizes the carbon attached to a nucleobase nitrogen atom (i.e. ring nitrogen or exocyclic amine). The ten-eleven translocation (TET) proteins are AlkB family like DNA demethylases that mediate oxidation of methylcytosines to generate 5-hydroxymethylcytosine (hm^5C), 5-formylcytosine (f^5C), and 5-carboxylcytosine (ca^5C). Another very important TYPE of demethylase is O^6-G-DNA methyltransferase (MGMT) which repairs specifically O^6-methylguanine alkylation through an S_N2 mechanism. The cysteine residue in MGMT is the nucleophile which removes the methyl group on O^6 position of the guanosine. The various demethylases show different substrate recognition specificity through several demethylation mechanisms (Figure 4.3). They also indicate that the chemical diversity of RNA synthetic modifications is invaluable for elucidating the functional consequences of such demethylation.
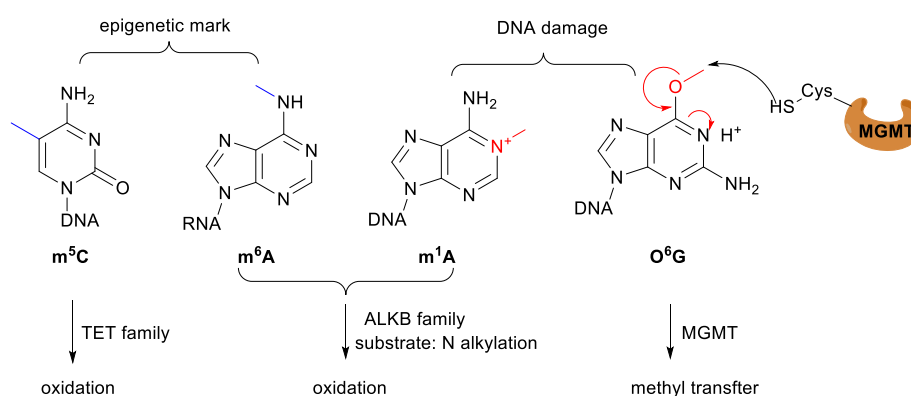
Figure 4.3 Selected demethylases show high recognition of methylated nucleobases

The DNA methylation is one of the best characterized epigenetic marks, which is critical for diverse biological processes, especially the embryonic development and cell differentiation. Despite dramatic progress in epigenetics during the past decade, the dynamic process of DNA demethylation, which has a profound impact on gene expression, is poorly understood.[168-170] Early efforts focused on testing various known DNA modifying proteins, such as DNA repair enzymes, cytosine deaminases, and DNA glycosylases achieving active DNA demethylation through different mechanisms.[171-175] Only the recent discovery of TET proteins as an m^5C oxidase has provided major insights into mechanisms of active DNA demethylation.[176]

The output of gene regulation is gene expression, which is often interpreted in terms of protein levels. The reversible methylation of RNA can directly affect protein production.[177] Therefore regulation at the RNA level is more direct than DNA regulation, a point which might be critical in the where transient gene expression is desirable. However, deciphering the regulatory mechanisms of reversible RNA

methylation was rather underdeveloped because RNAs are known as short-lived species. Furthermore, transcriptome-wide identification of the various modified bases at single-base resolution still remains a challenge.[178] Under these circumstances, it is often difficult to decide whether nucleobase methylation is a lesion or an epigenetic mark. FTO is the homolog of AlkB family proteins which are known for DNA repair 5 years ago. It was logical to think that the substrate of FTO, m^{6}A, is rather lesion repair than a novel regulatory mechanism. Nevertheless, with the help of nucleic acid chemistry, the importance and ubiquitous nature of RNA in gene regulation became widely recognized.[150] Very recent studies suggested that m^{6}A may pose an important mark for recruiting components of the splicing machinery and the FTO protein increases the rate of splicing by regulating the RNA binding ability.[179] Understanding of the link between FTO and energy homeostasis by unraveling mechanisms of splicing regulation of its methylated transcript targets will be beneficial for the FTO gene related therapies including obesity, diabetes, Alzheimer.[180]

Together, chemical modifications in DNA and RNA are diverse and abundant. With our expertise in nucleic acid modification, we were motivated to develop chemical methodologies for deciphering the full substrate tolerance of FTO. If large functional rich substrates could be processed by FTO, then enzymatic uncaging might allow activation of chemically modified RNA in a controlled manner (scheme 4.2).
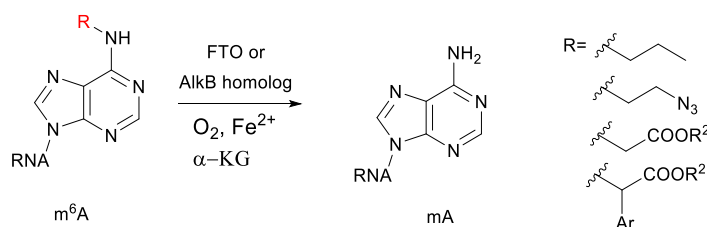
Scheme 4.2 Investigation of the tolerance of the FTO protein toward N^6-alkylated adenosine RNA target

4.3 Results and discussion

4.3.1 Synthesis of convertible RNA monomers

The availability of diverse modified RNA derivatives can greatly facilitate the understanding of their biological activities. However site-specific modification of single RNA monomer is laborious, single base alkylation of a long mRNA strand on the other hand, would be too challenging. Thus, we started with a 15-mer RNA synthesis.

There are several possibilities to obtain site-specifically modified RNA. Incorporating pre-modified RNA phosphoramidite monomer in the oligonucleotide solid-phase synthesis can be the most straight

forward procedure. However this method requires comlicated synthesis of different monomers for each ssRNA. An alternative approach to use the convertible nucleoside derivatives which can be treated with different alkyl amines after solid-supported synthesis (Scheme 4.3).
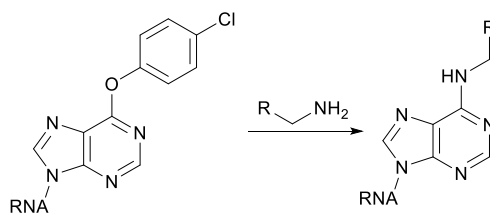
Scheme 4.3 The convertible nucleoside approach to afford different alkylated RNA

After careful consideration, we decided to use the convertible nucleoside approach to obtain a gram scale convertible phosphoramidite monomer according to a reported method.[181] For the preparation of the convertible nucleoside used to generate functionalized adenosines, we chose inosine as the substrate to initiate our project with a detailed literature procedure. The synthesis of 4-chlorophenol modified inosine phosphate proceeds in seven steps from inosine (Scheme 4.4). After tri-O-acetylation of the ribose hydroxyl groups, activation of the O^6 position for nucleophilic substitution was achieved through formation of the O^6-[(2,4,6-triisopropylphenyl)sulfonyl] ester. A major side product of this reaction, the N-sulfonated inosine, was easily separated from the desired product by flash chromatography. The sulfonate ester was again displaced in a trimethylamine-mediated reaction with 4-chlorophenol to produce O^6 modified inosine. Aminolysis of the acetate esters in methanol yielded the free nucleoside which further reacted under standard RNA phosphoramidite synthesis protocol to give the convertible RNA monomer.
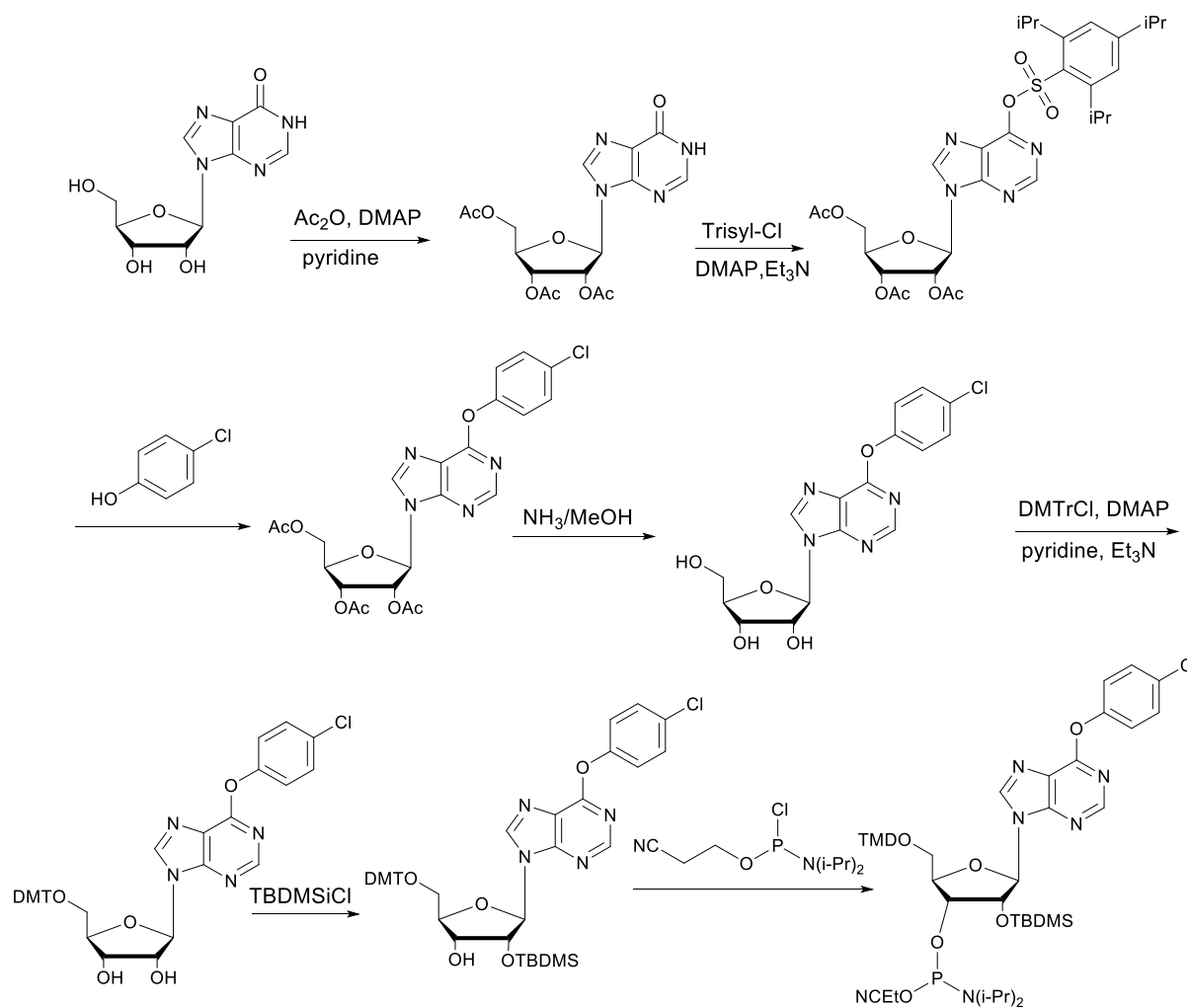
Scheme 4.4 Reported synthetic method of convertible O^6-4-chlorophenol substituted inosine monomer

After several months' struggle, we could obtain 0.6 gram of the finally product and 0.3 gram was incorporated into the solid-phase synthesizer for the 15mer-RNA synthesis with the sequence of 5'-AUU-GUC-A(X)C-AGC-AGC-3', where X is the convertible monomer. However, due to technical problems, the purification of RNA was unsuccessful, giving a mixture of unidentified products.

At the same time, other strategies were considered to obtain N^6 alkylated adenosine phosphoramidite. However there is no straight forward procedure due to the similar activity of three alcohol groups of the nucleoside monomer. One pre-modified adenosine monomer is applicable (Scheme 4.5). Starting with N^6 chlorine substituted adenosine, after nucleophilic attack with alkyl amines, N^6 alkylated adenosine was obtained with high yield. Followed by standard RNA phosphoramidite synthesis, the desired monomer could be obtained, but the method is still laborious.
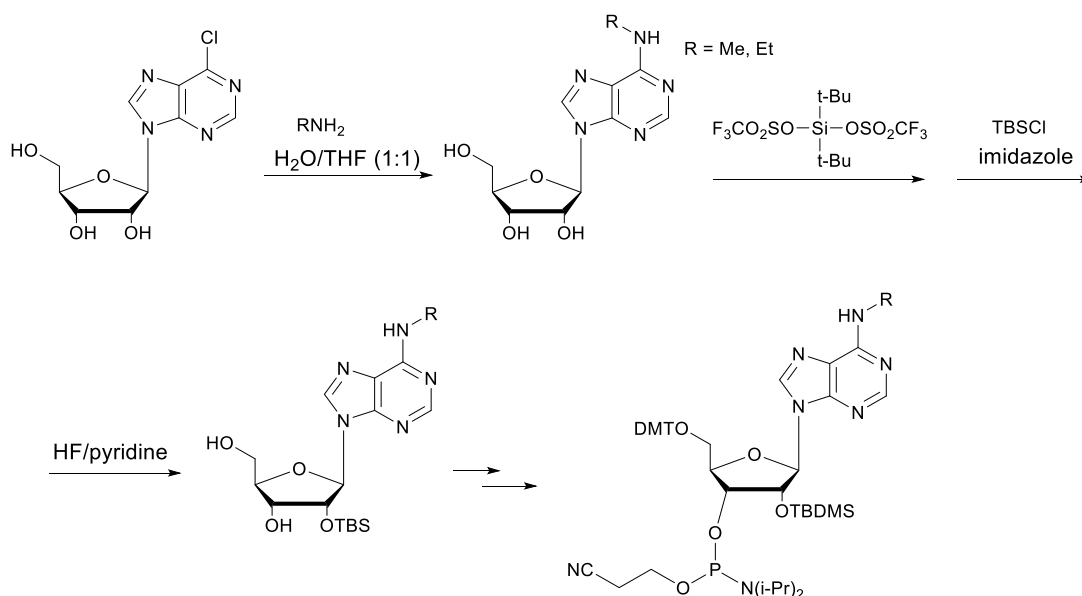
Scheme 4.5 Possible N^{6}A alkylated phosphoramidite synthesis

4.3.2 The FTO protein production

N-terminal his-tagged truncated human FTO was designed and produced as previously reported.[162] After lysis of the cell pellets, the filtered supernatant was purified by Ni-NTA chromatography. The fractions were analyzed through denaturing SDS–PAGE gel (Figure 4.4, A). However the obtained protein after purification showed two main bands on the gel (Figure 4.4, B lane 1). Later the repeat experiment didn't give the desired protein.
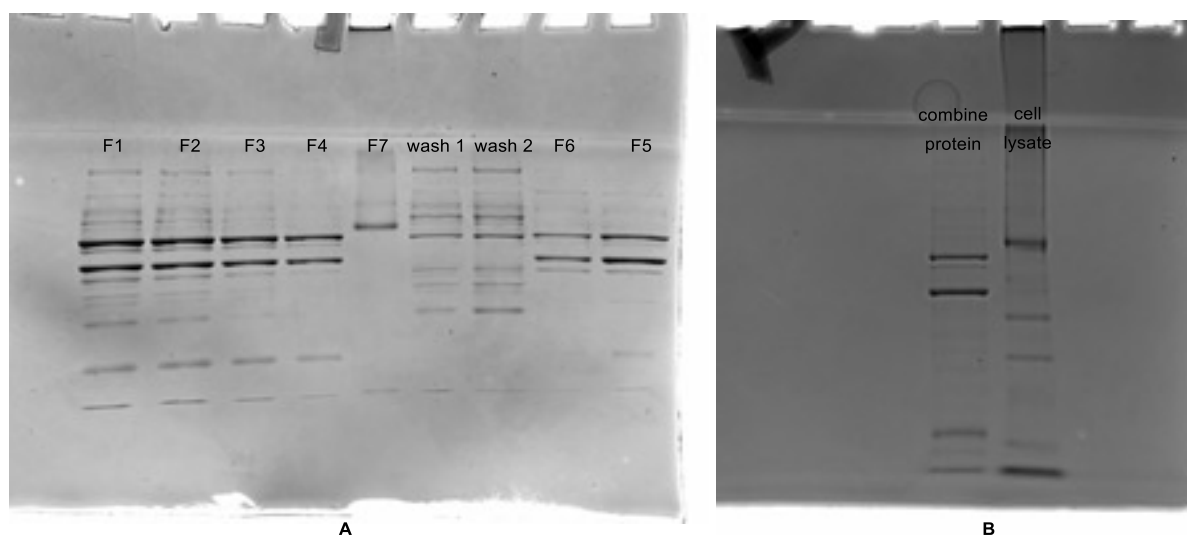
Figure 4.4 A: Protein purification of FTO by Ni-NTA, F1-F7 are fraction after adding the imidazole elute solution, wash 1 and 2 are the washing fractions before the elution; B: Purified FTO, lane 1 protein after dialysis; lane 2 cell lysate

4.4 Conclusion and outlook

The synthesis of convertible RNA phosphoramidite monomer demonstrated that tremendous amount of work and materials are needed. We need better approaches to have site-specific modified RNA which certainly is still a challenge for chemists. The FTO protein production is also difficult and requires further optimization. Mammalian protein production in *E. Coli* could be a problem because sometimes the bacteria can not tolerate the foreign protein as it might be toxic to its own growth.

In a short time frame, we couldn't solve all current problems, but the trials we have made could be valuable for our later research. Commercial available convertible RNA phosphoramidite could be a better solution in this case which can help us push the project further and faster. Cooperation with biologists might help chemists to solve the mammalian protein production in bacteria.

Nevertheless, this project is very interesting. New discoveries in the area are exciting and updating very fast. The presence of an N^1-methyl group on adenine bases in DNA and RNA was thought to be a form of damage (see also in Figure 4.3), now is proposed to be a new epigenetic mark in mRNA.[182] According to He's group, the new found m^1A is enriched around the start codon upstream of the first splice site, is dynamic in response to physiological conditions and correlates positively with protein production, strongly indicating a functional role for m^1A in promoting translation of methylated mRNA.

We should be aware of recent discoveries as these provide motivation and direction for our chemistry research. With our XHI methods and versatile diazo compounds, future work could test whether our exo-amine modified oligonucleotides are tolerated by the FTO protein, AlkB protein, and other demethylases. In the future, we could add functional groups selectively to nucleotides and uncage them with site specific recognition enzymes.

4.5 Experimental part

4.5.1 General

All reagents and solvents used were of analytical grade. Buffers were prepared with ultrapure water. All chemicals were purchased from Sigma-Aldrich, Fluka or Acros and used as received. Analytical TLC was performed on Silica gel 60 F_{254} pre-coated aluminium sheets. Flash chromatography was performed on Silica gel 60 40-63 μm (230-400 mesh) (SiliCycle, Quebec). ^{1}H and ^{13}C NMR and 2D spectra of modified CDN were acquired on a Bruker AvanceIII+ 600 MHz using Shigemi NMR tube. Other compounds were recorded on 400 MHz proton frequency spectrometer at 298 K. Chemical shifts relative to TMS were referenced to the solvent's residual peak and are reported in ppm. ESI MS-MS spectra were obtained on a Bruker Esquire3000plus spectrometer by direct injection in positive polarity of the ion trap detector. High resolution mass spectra were acquired on a Bruker maXis 4G QTOF ESI mass-spectrometer. MALDI TOF analyses were carried out on a Bruker Microflex mass-spectrometer in linear positive mode using sinapic acid as matrice. HPLC procedures were carried out on an Agilent 1100 LC system equipped with Eclipse XDB-C8 5μm 4.6 x 100 mm column (Agilent) for analytic analysis. Shimadzu preparative HPLC (LC-20AP) equipped with phenomenex column (Gemini® 10 μm C18 110 Å, LC Column 250 x 21.2 mm, AXIA™ Packed) was employed for preparative purification. 100 mM triethylammonium acetate (pH 7.2) and acetonitrile was used as a mobile phase. For analytical measurement: 1 mL/min: 0-35 % acetonitrile in 12 min, 35-80 % acetonitrile in 3 min, 80% acetonitrile in 2 min. For preparative separation: flow rate: 20 ml/min, 0% acetonitrile in 2 min, 0-60 % acetonitrile in 27 min.

4.5.2 Convertible RNA synthesis

The RNA monomer synthesis was carried out according to literature procedure.[181] The product of each step was confirmed by NMR and compared with literature. The seven-step synthesis took me several months to obtain 0.6 g of the finally phosphoramidite and 0.3 g was incorporated into the solid-phase synthesizer to synthesize a 15mer-RNA with the sequence of 5'-AUU-GUC-A(X)C-AGC-AGC-3', where X is the convertible monomer. The standard 1 μmol RNA synthesis cycle was modified to have an extended coupling time of 12 min. All phosphoramidites, including the convertible nucleoside phosphoramidites were dissolved to a concentration of 0.1 M in anhydrous CH_3CN. However during the cycle, the G channel was indicated to be blocked from the trityl monitor even though all the channels were thoroughly wash with CH_3CN before. Nevertheless, the solid phase was submitted to the deprotection step.

The resin-bound oligonucleotide was first cleavaged from the resin with 1.5 mL of ethanolic methylamine solution (8 M) for 2 hours, then heated at 42 ° C for 18 hours. The resulting solution was concentrated under vacuum. Removal of the 2'-O-silyl ether was afforded by treating the oligonucleotide with 0.6 mL of 1 M tetrabutylammonium fluoride (TBAF) in THF for 20 hours at room temperature. These reactions were quenched by the addition of 0.8 mL of 1 M TEAA, pH 7.5. The oligonucleotides were then purified by preparative HPLC. However the HPLC trace didn't give a clean peak, but rather a mixture of many fractions. Further purification was not continued and the repeat synthesis of RNA was postponed until the arrival of the new DNA synthesizer.

4.5.3 FTO protein production

N-terminal his-tagged truncated human FTO protein (his 6-FTO-ND 31) was designed and produced as previously reported.[162] The gene synthesis and subcloning were done by genscript in the following detail:

Gene name: FTO, Length: 1440 bp, Vector name: pUC57.

Sequence of the interest

CATATGACCCCCAAAGATGATGAATTCTATCAGCAGTGGCAGCTGAAATATCCTAAACTAATTCTCCGAGAAGC
CAGCAGTGTATCTGAGGAGCTCCATAAAGAGGTTCAAGAAGCCTTTCTCACACTGCACAAGCATGGCTGCTTA
TTTCGGGACCTGGTTAGGATCCAAGGCAAAGATCTGCTCACTCCGGTATCTCGCATCCTCATTGGTAATCCAGG
CTGCACCTACAAGTACCTGAACACCAGGCTCTTTACGGTCCCCTGGCCAGTGAAAGGGTCTAATATAAAACACA
CCGAGGCTGAAATAGCCGCTGCTTGTGAGACCTTCCTCAAGCTCAATGACTACCTGCAGATAGAAACCATCCA
GGCTTTGGAAGAACTTGCTGCCAAAGAGAAGGCTAATGAGGATGCTGTGCCATTGTGTATGTCTGCAGATTTC
CCCAGGGGTTGGGATGGGGTTCATCCTACAACGGACAAGATGAAGTGGACATTAAGAGCAGAGCAGCATACAAC
GTAACTTTGCTGAATTTCATGGATCCTCAGAAAATGCCATACCTGAAAGAGGAACCTTATTTTGGCATGGGGAA
AATGGCAGTGAGCTGGCATCATGATGAAAATCTGGTGGACAGGTCAGCGGTGGCAGTGTACAGTTATAGCTG
TGAAGGCCCTGAAGAGGAAAGTGAGGATGACTCTCATCTCGAAGGCAGGGATCCTGATATTTGGCATGTTGG
TTTTAAGATCTCATGGGACATAGAGACACCTGGTTTGGCGATACCCCTTCACCAAGGAGACTGCTATTTCATGC
TTGATGATCTCAATGCCACCCACCAACACTGTGTTTTGGCCGGTTCACAACCTCGGTTTAGTTCCACCCACCGAG
TGGCAGAGTGCTCAACAGGAACCTTGGATTATATTTTACAACGCTGTCAGTTGGCTCTGCAGAATGTCTGTGAC
GATGTGGACAATGATGATGTCTCTTTGAAATCCTTTGAGCCTGCAGTTTTGAAACAAGGAGAAGAAATTCATA
ATGAGGTCGAGTTTGAGTGGCTGAGGCAGTTTTGGTTTCAAGGCAATCGATACAGAAAGTGCACTGACTGGT
GGTGTCAACCCATGGCTCAACTGGAAGCACTGTGGAAGAAGATGGAGGGTGTGACAAATGCTGTGCTTCATG
AAGTTAAAAGAGAGGGGCTCCCCGTGGAACAAAGGAATGAAATCTTGACTGCCATCCTTGCCTCGCTCACTGC
ACGCCAGAACCTGAGGAGAGAATGGCATGCCAGGTGCCAGTCACGAATTGCCCGAACATTACCTGCTGATCA

GAAGCCAGAATGTCGGCCATACTGGGAAAAGGATGATGCTTCGATGCCTCTGCCGTTTGACCTCACAGACATC

GTTTCAGAACTCAGAGGTCAGCTTCTGGAAGCAAAACCCTAGTAGCTCGAG

The cDNA sequence encoding truncated hFTO was subcloned into pET-28a to generate a His-tagged fusion protein (done by Genscript). The plasmids were transformed into E. coli BL21 Star (DE3) and bacteria were grown on LB-agar plates containing 50 mg/l of kanamycin. Over-night pre-cultures, which were grown aerobically at 37 °C with a shaking speed of 190 rpm, were used to inoculate 1 liter LB medium with 50 mg/l kanamycin and grown at 37 °C and 250 rpm until OD600 reached 1.0. Then the bacterial cells were induced by isopropyl-b-D-thiogalactopyranoside (IPTG) (0.5 mM) at 15 °C and grown over-night at 15 °C and 250 rpm. The cells were harvested by centrifugation, frozen by liquid nitrogen, and stored at -80 °C. All subsequent steps were performed at 4 °C. The cell pellets were resuspended in buffer A (50mM imidazole, 300mM NaCl, 50mM sodium phosphate, pH 8.0), sonicated on ice and centrifuged at 12,000 rpm for 22 min. The filtered supernatant was purified by Ni-NTA chromatography. The fractions were analyzed through denaturing SDS–PAGE (Figure 4.3). However the obtained protein after purification showed two main bands on the gel (Figure 4.3, B lane 1). Later the repeat experiment didn't give the desired protein.
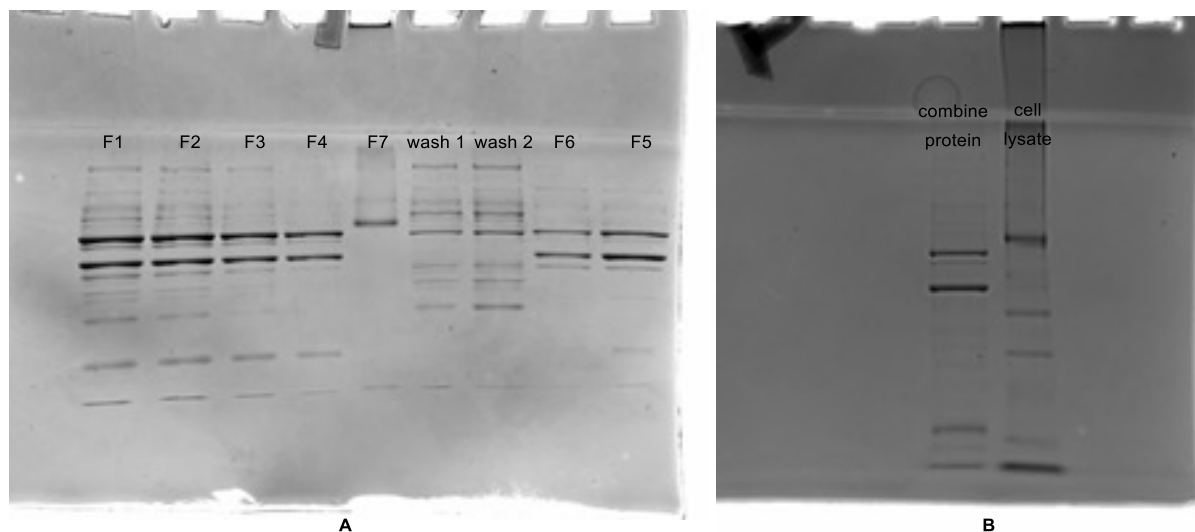
Figure S4.1 A: Protein purification of FTO by Ni-NTA, F1-F7 are fraction after adding the imidazole elute solution, wash 1 and 2 are the washing fractions before the elution; B: Purified FTO, lane 1 protein after dialysis; lane 2 cell lysate

List of abbreviations

2D NMR	two-dimensional nuclear magnetic resonance spectroscopy
A	adenine
AlkB	Alkylated DNA repair protein
ATP	adenosine triphosphate
Bhc	6-Bromo-4-diazomethyl-7-hydroxycoumarin (Bhc-diazo)
brlR	c-di-GMP responsive transcription regulator
C	cytosine
ca^{5}C	5-carboxylcytosine
c-di-AMP	bis-(3'-5')-cyclic dimeric adenosine monophosphate
c-di-GMP	bis-(3'-5')-cyclic dimeric guanosine monophosphate
CDN	cyclic dinucleotides
c-GAMP	cyclic guanosine monophosphate–adenosine
c-GAS	c-GAMP synthase
COSY	correlation spectroscopy
CpG	controlledpore glass
CpG	CpG dinucleotides
C-Src	proto-oncogene tyrosine-protein kinase
CuAAC	Cu-catalyzed azide–alkyne cycloaddition
DAC	diadenylyl cyclase
DFT	Density functional theory
DGC	diguanylate cyclase
DgrA	c-di-GMP receptor protein
DHAP	dihydroxyacetone phosphate
DIPA	N,N-diisopropylamine
DIPEA	N,N-diisopropylethylamine
DisA	the DNA integrity scanning protein
DMF	N,N-dimethylformamide
DMSO	dimethylsulfoxide
DNA	deoxyribonucleic acid
EDA	Ethyl diazoacetate
EDC	N-(3-Dimethylaminopropyl)-N'-ethylcarbonate
EFGR	epidermal growth factor receptor
EGF	epidermal growth factor
equiv.	equivalent(s)
f^{5}C	5-formylcytosine
f^{6}A	N^6-formyladenosine
FTO	fat mass and obesity-associated protein
G	guanine
G 6-P	glucose-6-phosphate
GFP	green fluorescent protein
GTP	guanosine-5'-triphosphate
HIV	human immunodeficiency virus
hm^{5}C	5-hydroxymethylcytosine
hm^{6}A	hydroxymethyladenosine

HMBC	heteronuclear multiple-bond correlation spectroscopy
HMQC	heteronuclear multiple-quantum correlation spectroscopy
HPLC	high pressure liquid chromatography
IFN	interferons
IMAC	Immobilized metal affinity chromatography
IP	inositol phosphate
IPS	inositol 3-phosphate synthase
m^5C	5-methylcytosine
m^6A	N^6-Methyladenosine
MES	2-(N-morpholino)ethanesulfonic acid
MGMT	O^6-methylguanine-DNA methyltransferase
MOPS	3-(N-morpholino)propanesulfonic acid
Ms	mesylate
NAD^+	Nicotinamide adenine dinucleotide
NHI	N-H bond insertion
NMR	nuclear magnetic resonance (spectroscopy)
ONB	ortho-nitrobenzyl
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate-buffered saline
PDE	phosphodiesterase
PI	inositol phospholipids
Pi	PO_4^{2-}
PKA	cyclic AMP-dependent protein kinase
PR	prosthetic-group-removing enzyme or phosphate-removing-enzyme
PRPP	phosphoribosyl pyrophosphate
RFP	red fluorescent protein
RNA	ribonucleic acid
ROESY	rotating frame nuclear Overhauser effect spectroscopy
ROS	reactive oxygen species
SAMe	S-adenosylmethionine
SDS	Sodium dodecyl sulfate
ssDNA	Sing strand DNA
ssRNA	Sing strand RNA
STING	stimulator of interferon genes
STING	the name deriving from stimulator of interferon genes
T	thymine
TBS	*t*-butyldimethyl silyl (ether, chloride, triflate,…)
TET	The ten-eleven translocation Pproteins
TFA	trifluoroacetic acid
THF	tetrahydrofuran
THPTA	tris(3-hydroxypropyltriazolyl- methyl)amine
TMS	trimethylsilyl (ether, chloride, triflate,…)
U	uracil
UPLC-MS	ultra pressure liquid chromatography, mass spectroscopy
XHI	X-H bond insertion, here X= N, O, S.

References

1.   G. Maas, New syntheses of diazo compounds. *Angew. Chem. Int. Ed.*, 2009, **48**, 8186-8195.

2.   A. Ford, H. Miel, A. Ring, C. N. Slattery, A. R. Maguire and M. A. McKervey, Modern organic synthesis with α-diazocarbonyl compounds. *Chem.Rev.*, 2015, **115**, 9981-10080.

3.   T. Curtius, Ueber die Einwirkung von salpetriger Säure auf salzsauren Glycocolläther. *Berichte der deutschen chemischen Gesellschaft*, 1883, **16**, 2230-2231.

4.   K. Clusius and U. Lüthi, Reaktionen mit15N. XXIV. Zur Bildungsweise und Struktur des Diazoessigesters. *Helv. Chim. Acta*, 1957, **40**, 445-456.

5.   R. W. Saalfrank, Diazo Compounds. Properties and Synthesis. Von M. Regitz und G. Maas. *Angew. Chem.*, 1987, **99**, 1335-1335.

6.   F. M. Wong, J. Wang, A. C. Hengge and W. Wu, Mechanism of rhodium-catalyzed carbene formation from diazo compounds. *Org. Lett.*, 2007, **9**, 1663-1665.

7.   X. Zhao, Y. Zhang and J. Wang, Recent developments in copper-catalyzed reactions of diazo compounds. *Chem. Comm.*, 2012, **48**, 10162-10173.

8.   D. Gillingham and N. Fei, Catalytic X-H insertion reactions based on carbenoids. *Chem. Soc. Rev.*, 2013, **42**, 4918-4931.

9.   S.-F. Zhu and Q.-L. Zhou, Enantioselective transition-metal-catalyzed heteroatom-hydrogen bonds insertion reactions. *Acc. Chem. Res.*, 2012, **45**, 1365-1377.

10.   M. P. Doyle, R. Duffy, M. Ratnikov and L. Zhou, Catalytic carbene insertion into C-H bonds. *Chem. Rev.*, 2010, **110**, 704-724.

11.   M. P. Doyle and D. C. Forbes, Recent advances in asymmetric catalytic metal carbene transformations. *Chem. Rev.*, 1998, **98**, 911-936.

12.   H. M. L. Davies and R. E. J. Beckwith, Catalytic enantioselective C–H activation by means of metal–carbenoid-Induced C–H insertion. *Chem. Rev.*, 2003, **103**, 2861-2904.

13.   A. Padwa and M. D. Weingarten, Cascade processes of metallo carbenoids. *Chem. Rev.*, 1996, **96**, 223-269.

14.   M. A. M. M. P. Doyle, T. Ye, *Modern Catalytic Methods for Organic Synthesis with Diazo Compounds*, Wiley, New York, 1998.

15.   I. V. Shishkov, F. Rominger and P. Hofmann, Remarkably stable copper(I) α-carbonyl carbenes: synthesis, structure, and mechanistic studies of alkene cyclopropanation reactions. *Organometallics*, 2009, **28**, 1049-1059.

16.   R. Paulissen, H. Reimlinger, E. Hayez, A. J. Hubert and P. Teyssié, Transition metal catalysed reactions of diazocompounds - II insertion in the hydroxylic bond. *Tetrahedron Lett.*, 1973, **14**, 2233-2236.

17.   R. Casanova and T. Reichstein, Methoxyketone aus Diazoketonen. Steroide, 5. Mitteilung. *Helv. Chim. Acta*, 1950, **33**, 417-422.

18.   P. Yates, The copper-catalyzed decomposition of diazoketones. *J. Am. Chem. Soc.*, 1952, **74**, 5376-5381.

19.   M. E. Morilla, M. M. Diaz-Requejo, T. R. Belderrain, M. C. Nicasio, S. Trofimenko and P. J. Perez, Catalytic insertion of diazo compounds into N–H bonds: the copper alternative. *Chem. Comm.*, 2002, 2998-2999.

20.   Y. Liang, H. Zhou and Z.-X. Yu, Why Is Copper(I) Complex more competent than dirhodium(II) complex in catalytic asymmetric O–H insertion reactions? A computational study of the metal carbenoid O–H insertion into water. *J. Am. Chem. Soc.*, 2009, **131**, 17783-17785.

21.   H. Huang, X. Guo and W. Hu, Efficient trapping of oxonium ylides with imines: A highly diastereoselective three-component reaction for the synthesis of β-amino-α-hydroxyesters with quaternary stereocenters. *Angew. Chem. Int. Ed.*, 2007, **46**, 1337-1339.

22.   Z. Li, B. T. Parr and H. M. L. Davies, Highly stereoselective C–C bond formation by rhodium-catalyzed tandem ylide formation/[2,3]-sigmatropic rearrangement between donor/acceptor carbenoids and chiral allylic alcohols. *J. Am. Chem. Soc.*, 2012, **134**, 10942-10946.

23. L. K. Baumann, H. M. Mbuvi, G. Du and L. K. Woo, Iron porphyrin-catalyzed olefination of carbonyl compounds with ethyl diazoacetate *Organometallics*, 2007, **26**, 3995-4002.

24. I. Aviv and Z. Gross, Iron(III) corroles and porphyrins as superior catalysts for the reactions of diazoacetates with nitrogen- or sulfur-containing nucleophilic substrates: synthetic uses and mechanistic insights. *Chem. Eur. J.*, 2008, **14**, 3995-4005.

25. B. Xu, S.-F. Zhu, X.-L. Xie, J.-J. Shen and Q.-L. Zhou, Asymmetric N[BOND]H insertion reaction cooperatively catalyzed by rhodium and chiral spiro phosphoric acids. *Angew. Chem. Int. Ed.*, 2011, **50**, 11483-11486.

26. H. Saito, R. Iwai, T. Uchiyama, M. Miyake and S. Miyairi, Chiral induction by cinchona alkaloids in the rhodium(II) catalyzed O–H insertion reaction, *Chem. Pharm. Bull.*, 2010, **58**, 872-874.

27. B. Xu, S.-F. Zhu, X.-L. Xie, J.-J. Shen and Q.-L. Zhou, Asymmetric N[BOND]H insertion reaction cooperatively catalyzed by rhodium and chiral spiro phosphoric acids. *Angew. Chem. Int. Ed.*, 2011, **50**, 11483-11486.

28. E. Galardon, P. Le Maux and G. Simonneaux, Insertion of ethyl diazoacetate into N–H and S–H bondscatalyzed by ruthenium porphyrin complexes. *J. Chem. Soc., Perkin Trans. 1*, 1997, 2455-2456.

29. Q.-H. Deng, H.-W. Xu, A. W.-H. Yuen, Z.-J. Xu and C.-M. Che, Ruthenium-catalyzed one-pot carbenoid N–H insertion reactions and diastereoselective synthesis of prolines *Org. Lett.*, 2008, **10**, 1529-1532.

30. M. Austeri, D. Rix, W. Zeghida and J. r. m. Lacour, CpRu-catalyzed O–H insertion and condensation reactions of α-diazocarbonyl compounds. *Org. Lett.*, 2011, **13**, 1394-1397.

31. T. G. Rajagopalan, W. H. Stein and S. Moore, The inactivation of pepsin by diazoacetyl-norleucine methyl ester. *J. Biol. Chem.*, 1966, **241**, 4295-4297.

32. G. R. Delpierre and J. S. Fruton, Specific inactivation of pepsin by a diazo ketone. *Proc. Natl. Acad. Sci. USA*, 1966, **56**, 1817-1822.

33. R. G. Salomon and J. K. Kochi, Copper(I) catalysis in cyclopropanations with diazo compounds. Role of olefin coordination. *J. Am. Chem. Soc.*, 1973, **95**, 3300-3310.

34. J. M. Antos and M. B. Francis, Selective tryptophan modification with rhodium carbenoids in aqueous solution. *J. Am. Chem. Soc.*, 2004, **126**, 10256-10257.

35. M.-A. Bind, J. Lepeule, A. Zanobetti, A. Gasparrini, A. Baccarelli and B. A. Coull, Air pollution and gene-specific methylation in the Normative Aging Study. *Epigenetics*, 2014, **9**. 448-458.

36. Z. Chen, F. Vohidov, J. M. Coughlin, L. J. Stagg, S. T. Arold, J. E. Ladbury and Z. T. Ball, Catalytic protein modification with dirhodium metallopeptides: specificity in designed and natural systems. *J. Am. Chem. Soc.*, 2012, **134**, 10138-10145.

37. K. Tishinov, K. Schmidt, D. Häussinger and D. G. Gillingham, Structure-selective catalytic alkylation of DNA and RNA. *Angew. Chem. Int. Ed.*, 2012, **51**, 12000-12004.

38. B. J. Deadman, S. G. Collins and A. R. Maguire, Taming hazardous chemistry in flow: The continuous processing of diazo and diazonium compounds, *Chem. Eur. J.*, 2015, **21**, 2298-2308.

39. F. Mastronardi, B. Gutmann and C. O. Kappe, Continuous flow generation and reactions of anhydrous diazomethane using a teflon AF-2400 tube-in-tube reactor, *Org. Lett.*, 2013, **15**, 5590-5593.

40. P. A. Bray and R. K. Sokas, Delayed respiratory fatality from trimethylsilyldiazomethane: What do workers need to know about potentially hazardous exposures? *J. Occup. Env. Med. ,* 2015, **57**, e15-e16.

41. E. Kühnel, D. D. P. Laffan, G. C. Lloyd-Jones, T. Martínez del Campo, I. R. Shepperson and J. L. Slaughter, Mechanism of methyl esterification of carboxylic acids by trimethylsilyl-diazomethane. *Angew. Chem. Int. Ed*, 2007, **46**, 7075-7078.

42. N. R. Candeias, R. Paterna and P. M. P. Gois, Homologation reaction of ketones with diazo compounds. *Chem. Rev.*, 2016, **116**, 2937-2981.

43. V. K. Aggarwal, J. de Vicente and R. V. Bonnert, Catalytic cyclopropanation of alkenes using diazo compounds generated in situ. A novel route to 2-arylcyclopropylamines. *Org. Lett.*, 2001, **3**, 2785-2788.

44.	J. Geittner, R. Huisgen and R. Sustmann, Kinetics of 1,3-dipolar cycloaddition reactions of diazomethane; A correlation with homo-lumo energies. *Tetrahedron Lett.*, 1977, **18**, 881-884.

45.	R. Huisgen, Kinetics and mechanism of 1,3-dipolar cycloadditions. *Angew. Chem. Int. Ed*, 1963, **2**, 633-645.

46.	D. Fu, J. A. Calvo and L. D. Samson, Balancing repair and tolerance of DNA damage caused by alkylating agents. *Nat. Rev. Cancer*, 2012, **12**, 104-120.

47.	D. M. Brown, D. I. Magrath and A. R. Todd, Nucleotides. Part XXXIV. The hydrolysis of dialkyl esters of uridine-3' phosphate and its relevance to the question of phosphotriester linkages in ribonucleic acids. *J. Chem. Soc. (Resumed)*, 1955, 4396-4401.

48.	C. Bourget, E. Trévisiol, I. Bridon, M. Kotera, J. Lhomme and A. Laayoun, Biotin-phenyldiazomethane conjugates as labeling reagents at phosphate in mono and polynucleotides. *Bioorganic & Medicinal Chemistry*, 2005, **13**, 1453-1461.

49.	R. S. Givens, M. Rubina and J. Wirz, Applications of p-hydroxyphenacyl (pHP) and coumarin-4-ylmethyl photoremovable protecting groups *Photochem. Photobiol. Sci.*, 2012, **11**, 472-488.

50.	J. A. McCray, L. Herbette, T. Kihara and D. R. Trentham, A new approach to time-resolved studies of ATP-requiring biological systems: Laser flash photolysis of caged ATP. *Proc. Natl. Acad. Sci. USA,* 1980, **77**, 7237-7241.

51.	Y. V. Il'ichev, M. A. Schwörer and J. Wirz, Photochemical Reaction Mechanisms of 2-Nitrobenzyl Compounds: Methyl Ethers and Caged ATP. *J. Am. Chem. Soc.*, 2004, **126**, 4581-4595.

52.	H. Ando, T. Furuta, R. Y. Tsien and H. Okamoto, Photo-mediated gene activation using caged RNA/DNA in zebrafish embryos. *Nat Genet*, 2001, **28**, 317-325.

53.	S. Shah, P. K. Jain, A. Kala, D. Karunakaran and S. H. Friedman, Light-activated RNA interference using double-stranded siRNA precursors modified using a remarkable regiospecificity of diazo-based photolabile groups. *Nucleic Acids Res.*, 2009, **37**, 4508-4517.

54.	P. K. Jain, S. Shah and S. H. Friedman, Patterning of gene expression using new photolabile groups applied to light activated RNAi. *J. Am. Chem. Soc.*, 2011, **133**, 440-446.

55.	U. Römling, M. Y. Galperin and M. Gomelsky, Cyclic di-GMP: the first 25 years of a universal bacterial second messenger. *Microbiol. Mol. Biol. Rev.,* 2013, **77**, 1-52.

56.	U. Jenal and J. Malone, Mechanisms of Cyclic-di-GMP signaling in bacteria. *Annual Review of Genetics*, 2006, **40**, 385-407.

57.	R. M. Corrigan, J. C. Abbott, H. Burhenne, V. Kaever and A. Grundling, c-di-AMP is a new second messenger in staphylococcus aureus with a role in controlling cell size and envelope stress. *Plos Pathogens*, 2011, **7**. e1002217.

58.	K. Sureka, Philip H. Choi, M. Precit, M. Delince, D. A. Pensinger, TuAnh N. Huynh, Ashley R. Jurado, Young A. Goo, M. Sadilek, Anthony T. Iavarone, J.-D. Sauer, L. Tong and Joshua J. Woodward, The cyclic dinucleotide c-di-AMP is an allosteric regulator of metabolic enzyme function. *Cell*, 2014, **158**, 1389-1401.

59.	R. M. Corrigan and A. Grundling, Cyclic di-AMP: another second messenger enters the fray. *Nat Rev Micro*, 2013, **11**, 513-524.

60.	B. W. Davies, R. W. Bogard, T. S. Young and J. J. Mekalanos, Coordinated regulation of accessory genetic elements produces cyclic di-nucleotides for V. cholerae virulence. *Cell*, 2012, **149**, 358-370.

61.	M. Tosolini, F. Pont, D. Bétous, E. Ravet, L. Ligat, F. Lopez, M. Poupot, M. Poirot, É. Pérouzel, G. Tiraby, E. Verhoeyen and J.-J. Fournié, Human monocyte recognition of adenosine-based cyclic dinucleotides unveils the A2a Gαs protein-coupled receptor tonic inhibition of mitochondrially induced cell death. *Mol. Cell. Biol.*, 2015, **35**, 479-495.

62.	D. L. Burdette, K. M. Monroe, K. Sotelo-Troha, J. S. Iwig, B. Eckert, M. Hyodo, Y. Hayakawa and R. E. Vance, STING is a direct innate immune sensor of cyclic di-GMP. *Nature*, 2011, **478**, 515-518.

63.	C. Shu, G. Yi, T. Watts, C. C. Kao and P. Li, Structure of STING bound to cyclic di-GMP reveals the mechanism of cyclic dinucleotide recognition by the immune system. *Nat Struct Mol Biol*, 2012, **19**, 722-724.

64.   P. Schaap, Cyclic di-nucleotide signaling enters the eukaryote domain. *IUBMB Life*, 2013, **65**, 897-903.

65.   A. G. Bowie, Innate sensing of bacterial cyclic dinucleotides: more than just STING. *Nat. Immunol.*, 2012, **13**, 1137-1139.

66.   J. Wu, L. Sun, X. Chen, F. Du, H. Shi, C. Chen and Z. J. Chen, Cyclic GMP-AMP is an endogenous second messenger in innate immune signaling by cytosolic DNA. *Science*, 2013, **339**, 826-830.

67.   P. Gao, M. Ascano, Y. Wu, W. Barchet, B. L. Gaffney, T. Zillinger, A. A. Serganov, Y. Z. Liu, R. A. Jones, G. Hartmann, T. Tuschl and D. J. Patel, Cyclic [G(2',5')pA(3',5')p] is the metazoan second messenger produced by DNA-activated cyclic GMP-AMP synthase. *Cell*, 2013, **153**, 1094-1107.

68.   D. X. Gao, J. X. Wu, Y. T. Wu, F. H. Du, C. Aroh, N. Yan, L. J. Sun and Z. J. J. Chen, Cyclic GMP-AMP synthase is an innate immune sensor of HIV and other retroviruses. *Science*, 2013, **341**, 903-906.

69.   X. D. Li, J. X. Wu, D. X. Gao, H. Wang, L. J. Sun and Z. J. J. Chen, Pivotal roles of cGAS-cGAMP signaling in antiviral defense and immune adjuvant effects. *Science*, 2013, **341**, 1390-1394.

70.   P. Ross, H. Weinhouse, Y. Aloni, D. Michaeli, P. Weinberger-Ohana, R. Mayer, S. Braun, E. de Vroom, G. A. van der Marel, J. H. van Boom and M. Benziman, Regulation of cellulose synthesis in Acetobacter xylinum by cyclic diguanylic acid. *Nature*, 1987, **325**, 279-281.

71.   T. Schirmer and U. Jenal, Structural and mechanistic determinants of c-di-GMP signalling. *Nat. Rev. Micro.*, 2009, **7**, 724-735.

72.   R. Hengge, Principles of c-di-GMP signalling in bacteria. *Nat.Rev. Micro.*, 2009, **7**, 263-273.

73.   Z.-h. Chen and P. Schaap, The prokaryote messenger c-di-GMP triggers stalk cell differentiation in Dictyostelium. *Nature*, 2012, **488**, 680-683.

74.   D. Srivastava and C. M. Waters, Cyclic diguanylate is a ubiquitous signaling molecule in bacteria: insights into biochemistry of the GGDEF protein domain. *J. Bacteriol.*, 2012, **194**, 4485-4493.

75.   P. S. Stewart and J. William Costerton, Antibiotic resistance of bacteria in biofilms. *The Lancet*, 2001, **358**, 135-138.

76.   C. de la Fuente-Núñez, F. Reffuveille, L. Fernández and R. E. W. Hancock, Bacterial biofilm development as a multicellular adaptation: antibiotic resistance and new therapeutic strategies. *Curr. Opin. Microbiol.*, 2013, **16**, 580-589.

77.   S. L. Chua, S. Y.-Y. Tan, M. T. Rybtke, Y. Chen, S. A. Rice, S. Kjelleberg, T. Tolker-Nielsen, L. Yang and M. Givskov, Synergistic activities of an efflux pump inhibitor and iron chelators against pseudomonas aeruginosa growth and biofilm formation. *Antimicrob. Agents Chemother.*, 2013, **57**, 2066-2075.

78.   J. R. Chambers, J. Liao, M. J. Schurr and K. Sauer, BrlR from Pseudomonas aeruginosa is a c-di-GMP-responsive transcription factor. *Mol. Microbiol.*, 2014, **92**, 471-487.

79.   K. Gupta, J. Liao, O. E. Petrova, K. E. Cherny and K. Sauer, Elevated levels of the second messenger c-di-GMP contribute to antimicrobial resistance of Pseudomonas aeruginosa. *Mol. Microbiol.*, 2014, **92**, 488-506.

80.   M. Fazli, T. Bjarnsholt, K. Kirketerp-Møller, A. Jørgensen, C. B. Andersen, M. Givskov and T. Tolker-Nielsen, Quantitative analysis of the cellular inflammatory response against biofilm bacteria in chronic wounds. *Wound Repair and Regeneration*, 2011, **19**, 387-391.

81.   T. Bjarnsholt, P. Ø. Jensen, M. J. Fiandaca, J. Pedersen, C. R. Hansen, C. B. Andersen, T. Pressler, M. Givskov and N. Høiby, Pseudomonas aeruginosa biofilms in the respiratory tract of cystic fibrosis patients. *Pediatr Pulmonol*, 2009, **44**, 547-558.

82.   D.-g. Ha and G. A. O'toole, in *Microbial Biofilms, Second Edition*, Am. Soc.Micro., 2015.

83.   G. Witte, S. Hartung, K. Büttner and K.-P. Hopfner, Structural biochemistry of a bacterial checkpoint protein reveals diadenylate cyclase activity regulated by DNA recombination intermediates. *Mol. Cell*, 2008, **30**, 167-178.

84.   Y. Oppenheimer-Shaanan, E. Wexselblatt, J. Katzhendler, E. Yavin and S. Ben-Yehuda, c-di-AMP reports DNA integrity during sporulation in Bacillus subtilis. *EMBO reports*, 2011, **12**, 594-601.

85. S. B. Ficarro, M. L. McCleland, P. T. Stukenberg, D. J. Burke, M. M. Ross, J. Shabanowitz, D. F. Hunt and F. M. White, Phosphoproteome analysis by mass spectrometry and its application to Saccharomyces cerevisiae. *Nat. Biotech.*, 2002, **20**, 301-305.

86. C. E. Witte, A. T. Whiteley, T. P. Burke, J.-D. Sauer, D. A. Portnoy and J. J. Woodward, Cyclic di-AMP is critical for Listeria monocytogenes growth, cell wall homeostasis, and establishment of infection. *mBio*, 2013, **4**. e00282-13.

87. Y. Bai, J. Yang, T. M. Zarrella, Y. Zhang, D. W. Metzger and G. Bai, Cyclic di-AMP impairs potassium uptake mediated by a cyclic di-AMP binding protein in Streptococcus pneumoniae. *J. Bacteriol.*, 2014, **196**, 614-623.

88. S. Ouyang, X. Song, Y. Wang, H. Ru, N. Shaw, Y. Jiang, F. Niu, Y. Zhu, W. Qiu, K. Parvatiyar, Y. Li, R. Zhang, G. Cheng and Z.-J. Liu, Structural analysis of the STING adaptor protein reveals a hydrophobic dimer interface and mode of cyclic di-GMP binding. *Immunity*, 2012, **36**, 1073-1086.

89. K. Parvatiyar, Z. Q. Zhang, R. M. Teles, S. Y. Ouyang, Y. Jiang, S. S. Iyer, S. A. Zaver, M. Schenk, S. Zeng, W. W. Zhong, Z. J. Liu, R. L. Modlin, Y. J. Liu and G. H. Cheng, The helicase DDX41 recognizes the bacterial secondary messengers cyclic di-GMP and cyclic di-AMP to activate a type I interferon immune response. *Nat. Immun.*, 2012, **13**, 1155-1161.

90. W. X. Chen, R. KuoLee and H. B. Yan, The potential of 3',5'-cyclic diguanylic acid (c-di-GMP) as an effective vaccine adjuvant. *Vaccine*, 2010, **28**, 3080-3085.

91. L. Corrales, L. H. Glickman, S. M. McWhirter, D. B. Kanne, K. E. Sivick, E. Lemmens, J. J. Leong, K. Metchette, T. W. Dubensky and T. Gajewski, Direct activation of STING in the tumor microenvironment with synthetic cyclic dinucleotide derivatives leads to potent and systemic tumor-specific immunity. *J. Immunother. Cancer*, 2014, **2**, 1-1.

92. J. Fu, D. B. Kanne, M. Leong, L. H. Glickman, S. M. McWhirter, E. Lemmens, K. Mechette, J. J. Leong, P. Lauer, W. Liu, K. E. Sivick, Q. Zeng, K. C. Soares, L. Zheng, D. A. Portnoy, J. J. Woodward, D. M. Pardoll, T. W. Dubensky and Y. Kim, STING agonist formulated cancer vaccines can cure established tumors resistant to PD-1 blockade. Science translational medicine. *Sc. Transl. Med.*, 2015, **7**, 283ra252-283ra252.

93. L. Corrales, Laura H. Glickman, Sarah M. McWhirter, David B. Kanne, Kelsey E. Sivick, George E. Katibah, S.-R. Woo, E. Lemmens, T. Banda, JustinJ. Leong, K. Metchette, Thomas W. Dubensky Jr and Thomas F. Gajewski, Direct activation of STING in the tumor microenvironment leads to potent and systemic tumor regression and immunity. *Cell Reports*, 2015, **11**, 1018-1030.

94. H. Miyabe, M. Hyodo, T. Nakamura, Y. Sato, Y. Hayakawa and H. Harashima, A new adjuvant delivery system 'cyclic di-GMP/YSK05 liposome'for cancer immunotherapy. *J. Control. Release* 2014, **184**, 20-27.

95. B. L. Gaffney, E. Veliath, J. W. Zhao and R. A. Jones, One-flask syntheses of c-di-GMP and the [R p, R p] and [R p, S p] thiophosphate analogues. *Org. Lett.*, 2010, **12**, 3269-3271.

96. P. Clivio, S. Coantic-Castex and D. Guillaume, (3′-5′)-cyclic dinucleotides: synthetic strategies and biological potential. *Chem. Rev.*, 2013, **113**, 7354-7401.

97. A. Grajkowski, J. Cieślak, A. Gapeev, C. Schindler and S. L. Beaucage, Convenient synthesis of a propargylated cyclic (3′-5′) diguanylic acid and its "click" conjugation to a biotinylated azide. *Bioconjugate Chemistry*, 2010, **21**, 2147-2152.

98. I. M. Sharma, T. Dhanaraman, R. Mathew and D. Chatterji, Synthesis and characterization of a fluorescent analogue of cyclic di-GMP. *Biochemistry*, 2012, **51**, 5443-5453.

99. Z. Y. Zhang, S. Kim, B. L. Gaffney and R. A. Jones, Polymorphism of the signaling molecule c-di-GMP. *J. Am. Chem, Soc.*, 2006, **128**, 7015-7024.

100. M. Gentner, M. G. Allan, F. Zaehringer, T. Schirmer and S. Grzesiek, Oligomer formation of the bacterial second messenger c-di-GMP: reaction rates and equilibrium constants indicate a monomeric state at physiological concentrations. *J. Am. Chem. Soc.*, 2012, **134**, 1019-1029.

101. B. E. Bernstein, T. S. Mikkelsen, X. Xie, M. Kamal, D. J. Huebert and J. Cuff, A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell*, 2006, **125**. 315–326.

102. M. Christen, B. Christen, M. G. Allan, M. Folcher, P. Jenö, S. Grzesiek and U. Jenal, DgrA is a member of a new family of cyclic diguanosine monophosphate receptors and controls flagellar motor function in Caulobacter crescentus. *Proc. Natl. Acad. Sci. USA*, 2007, **104**, 4112-4117.

103. K. W. Ryu, D.-S. Kim and W. L. Kraus, New facets in the regulation of gene expression by ADP-ribosylation and poly (ADP-ribose) polymerases. *Chem. Rev.*, 2015, **115**, 2453-2481.

104. L. Mouchiroud, Riekelt H. Houtkooper, N. Moullan, E. Katsyuba, D. Ryu, C. Cantó, A. Mottis, Y.-S. Jo, M. Viswanathan, K. Schoonjans, L. Guarente and J. Auwerx, The NAD+/Sirtuin pathway modulates longevity through activation of mitochondrial UPR and FOXO signaling. *Cell*, 2013, **154**, 430-441.

105. S. Posternak, Sur la synthése de l'ether hexaphosphorique de l'inosite avec le principe phospho-organique de réserve des plantes vertes. *Compt. Rend. Acad. Sci.*, 1919, **169**, 138-140.

106. R. S. Seelan, J. Lakshmanan, M. F. Casanova and R. N. Parthasarathy, Identification of myo-Inositol-3-phosphate synthase isoforms: characterization, expression, and putative role of a 16-kDa γc isoform. *J. Biol. Chem.*, 2009, **284**, 9443-9457.

107. A. M. Gardell, J. Yang, R. Sacchi, N. A. Fangue, B. D. Hammock and D. Kültz, Tilapia (Oreochromis mossambicus) brain cells respond to hyperosmotic challenge by inducing myo-inositol biosynthesis. *J. Experi. Biol.*, 2013, **216**, 4615-4625.

108. C. Ye and M. L. Greenberg, Inositol synthesis regulates the activation of GSK-3α in neuronal cells. *J. Neurochem.*, 2015, **133**, 273-283.

109. R. F. Irvine and M. J. Schell, Back in the water: the return of the inositol phosphates. *Nat. Rev. Mol. Cell. Biol.*, 2001, **2**, 327-338.

110. M. P. Thomas, S. J. Mills and B. V. L. Potter, The "other" inositols and their phosphates: synthesis, biology, and medicine (with recent advances in myo-Inositol chemistry). *Angew. Chem. Int. Ed.*, 2016, **55**, 1614-1650.

111. S. H. Zeisel, The fetal origins of memory: the role of dietary choline in optimal brain development. *J.Pediatr.*, 2006, **149**, S131-S136.

112. F. Westheimer, Why nature chose phosphates? *Science*, 1987, **235**, 1173-1178.

113. T. Hunter, Why nature chose phosphate to modify proteins? *Philos. Trans. R. Soc. Lond., B, Biol. Sci.,* 2012, **367**, 2513-2516.

114. S. C. L. Kamerlin, P. K. Sharma, R. B. Prasad and A. Warshel, Why nature really chose phosphate? *Q. Rev. Biophys.*, 2013, **46**, 1-132.

115. G. T. Cori and A. A. Green, Crystalline muscle phosphorylase II prosthetic group. *J. Biol. Chem.*, 1943, **151**, 31-38.

116. E. H. Fischer and E. G. Krebs, Conversion of phosphorylase-b to phosphorylase-a in muscle extracts. *J. Biol. Chem.*, 1955, **216**, 121-132.

117. E. W. Sutherland and W. D. Wosilait, The relationship of epinephrine and glucagon to liver phosphorylase. I. Liver phosphorylase; preparation and properties. *J.Biol. Chem.*, 1956, **218**, 459-468.

118. D. A. Walsh, J. P. Perkins and E. G. Krebs, An adenosine 3', 5'-monophosphate-dependant protein kinase from rabbit skeletal muscle. *J. Biol. Chem.*, 1968, **243**, 3763-3765.

119. P. Cohen, The origins of protein phosphorylation. *Nat Cell Biol*, 2002, **4**, E127-E130.

120. J. V. Olsen, B. Blagoev, F. Gnad, B. Macek, C. Kumar, P. Mortensen and M. Mann, Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. *Cell*, 2006, **127**, 635-648.

121. M. W. Salter and L. V. Kalia, Src kinases: a hub for NMDA receptor regulation. *Nat Rev Neurosci*, 2004, **5**, 317-328.

122. T. J. Yeatman, A renaissance for SRC. *Nat Rev Cancer*, 2004, **4**, 470-480.

123. S. J. Parsons and J. T. Parsons, Src family kinases, key regulators of signal transduction. *Oncogene*, 2004, **23**, 7906-7909.

124. S. Bagrodia, I. Chackalaparampil, T. E. Kmiecik and D. Shalloway, Altered tyrosine 527 phosphorylation and mitotic activation of p60 (c-src) *Nature*, 1991, **349**, 172-175.

125. N. Michael and N. Jura, Src defines a new pool of EGFR substrates. *Nat. Struct. Mol. Biol.*, 2015, **22**, 945-947.
126. M. Mann and O. N. Jensen, Proteomic analysis of post-translational modifications. *Nat. Biotech.*, 2003, **21**, 255-261.
127. N. G. Ahn and K. A. Resing, Toward the phosphoproteome. *Nat. Biotech.*, 2001, **19**, 317-318.
128. E. S. Witze, W. M. Old, K. A. Resing and N. G. Ahn, Mapping protein post-translational modifications with mass spectrometry. *Nat Meth*, 2007, **4**, 798-806.
129. N. M. Riley and J. J. Coon, Phosphoproteomics in the age of rapid and deep proteome profiling. *Anal. Chem.*, 2016, **88**, 74-94.
130. Y. Zhang, C. Zhang, H. Jiang, P. Yang and H. Lu, Fishing the PTM proteome with chemical approaches using functional solid phases. *Chem. Soc. Rev.*, 2015, **44**, 8260-8287.
131. C.-F. Tsai, Y.-T. Wang, H.-Y. Yen, C.-C. Tsou, W.-C. Ku, P.-Y. Lin, H.-Y. Chen, A. I. Nesvizhskii, Y. Ishihama and Y.-J. Chen, Large-scale determination of absolute phosphorylation stoichiometries in human cells by motif-targeting quantitative proteomics. *Nat. Commun.*, 2015, **6**.
132. Y. Oda, T. Nagasu and B. T. Chait, Enrichment analysis of phosphorylated proteins as a tool for probing the phosphoproteome. *Nat. Biotech.*, 2001, **19**, 379-382.
133. H. Zhou, J. D. Watts and R. Aebersold, A systematic approach to the analysis of protein phosphorylation. *Nat. Biotech.*, 2001, **19**, 375-378.
134. T. Furuta, H. Takeuchi, M. Isozaki, Y. Takahashi, M. Kanehara, M. Sugimoto, T. Watanabe, K. Noguchi, T. M. Dore, T. Kurahashi, M. Iwamura and R. Y. Tsien, Bhc-cNMPs as either water-soluble or membrane-permeant photoreleasable cyclic nucleotides for both one- and two-photon excitation. *Chem. Bio. Chem*, 2004, **5**, 1119-1128.
135. N. A. McGrath, K. A. Andersen, A. K. F. Davis, J. E. Lomax and R. T. Raines, Diazo compounds for the bioreversible esterification of proteins. *Chem. Sci.*, 2015, **6**, 752-755.
136. J. F. McGarrity and T. Smyth, Hydrolysis of diazomethane-kinetics and mechanism. *J. Am. Chem. Soc.*, 1980, **102**, 7303-7308.
137. M. I. Javed and M. Brewer, Diazo preparation via dehydrogenation of hydrazones with "activated" DMSO. *Org. Lett.*, 2007, **9**, 1789-1792.
138. D. N. Tran, C. Battilocchio, S.-B. Lou, J. M. Hawkins and S. V. Ley, Flow chemistry as a discovery tool to access sp 2–sp 3 cross-coupling reactions via diazo compounds. *Chem. Sci.*, 2015, **6**, 1120-1125.
139. J. W. Walker, G. P. Reid, J. A. McCray and D. R. Trentham, Photolabile 1-(2-nitrophenyl) ethyl phosphate esters of adenine nucleotide analogs. Synthesis and mechanism of photolysis. *J. Am. Chem. Soc.*, 1988, **110**, 7170-7177.
140. J. Seidler, N. Zinn, M. E. Boehm and W. D. Lehmann, De novo sequencing of peptides by MS/MS. *PROTEOMICS*, 2010, **10**, 634-649.
141. T. Phillips, The role of methylation in gene expression. *Nature Education*, 2008, **1**, 116.
142. P. A. Jones and D. Takai, The role of DNA methylation in mammalian epigenetics. *Science*, 2001, **293**, 1068-1070.
143. C. He, Grand challenge commentary: RNA epigenetics? *Nat. Chem. Biol.*, 2010, **6**, 863-865.
144. S. Beri, N. Tonna, G. Menozzi, M. C. Bonaglia, C. Sala and R. Giorda, DNA methylation regulates tissue-specific expression of Shank3. *J. Neurochem.*, 2007, **101**, 1380-1391.
145. J. Wan, V. F. Oliver, G. Wang, H. Zhu, D. J. Zack, S. L. Merbs and J. Qian, Characterization of tissue-specific differential DNA methylation suggests distinct modes of positive and negative gene expression regulation. *BMC Genomics*, 2015, **16**, 1-11.
146. E. Li and Y. Zhang, DNA methylation in mammals. *Cold Spring Harbor Perspectives in Biology*, 2014, **6**.
147. B. F. Vanyushin, L. E. Nemirovsky, V. V. Klimenko, V. K. Vasiliev and A. N. Belozersky, The 5-methylcytosine in DNA of rats. *Gerontology*, 1973, **19**, 138-152.
148. M. Jung and G. P. Pfeifer, Aging and DNA methylation. *BMC Biology*, 2015, **13**, 1-8.

149. M. Zampieri, F. Ciccarone, R. Calabrese, C. Franceschi, A. Bürkle and P. Caiafa, Reconfiguration of DNA methylation in aging. *Mechanisms of Ageing and Development*, 2015, **151**, 60-70.

150. I. A. Roundtree and C. He, RNA epigenetics—chemical messages for posttranscriptional gene regulation. *Curr. Opin. Chem.Biol.*, 2016, **30**, 46-51.

151. R. Desrosiers, K. Friderici and F. Rottman, Identification of methylated nucleosides in messenger RNA from Novikoff hepatoma cells. *Proc. Natl. Acad. Sci. USA*, 1974, **71**, 3971-3975.

152. S. Horowitz, A. Horowitz, T. W. Nilsen, T. W. Munns and F. M. Rottman, Mapping of N6-methyladenosine residues in bovine prolactin mRNA. *Proc. Natl. Acad. Sci. USA*, 1984, **81**, 5667-5671.

153. J. E. Harper, S. M. Miceli, R. J. Roberts and J. L. Manley, Sequence specificity of the human mRNA N6-adenosine methylase *in vitro. Nucleic Acids Res.*, 1990, **18**, 5735-5741.

154. G. Jia, Y. Fu, X. Zhao, Q. Dai, G. Zheng, Y. Yang, C. Yi, T. Lindahl, T. Pan, Y.-G. Yang and C. He, N6-methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. *Nat. Chem. Biol.*, 2011, **7**, 885-887.

155. T. Peters, K. Ausmeier and U. Rüther, Cloning of Fatso (Fto), a novel gene deleted by the Fused toes (Ft) mouse mutation. *Mammalian Genome*, **10**, 983-986.

156. C. Church, L. Moir, F. McMurray, C. Girard, G. T. Banks, L. Teboul, S. Wells, J. C. Bruning, P. M. Nolan, F. M. Ashcroft and R. D. Cox, Overexpression of Fto leads to increased food intake and results in obesity. *Nat. Genet.*, 2010, **42**, 1086-1092.

157. T. M. Frayling, N. J. Timpson, M. N. Weedon, E. Zeggini, R. M. Freathy, C. M. Lindgren, J. R. B. Perry, K. S. Elliott, H. Lango, N. W. Rayner, B. Shields, L. W. Harries, J. C. Barrett, S. Ellard, C. J. Groves, B. Knight, A.-M. Patch, A. R. Ness, S. Ebrahim, D. A. Lawlor, S. M. Ring, Y. Ben-Shlomo, M.-R. Jarvelin, U. Sovio, A. J. Bennett, D. Melzer, L. Ferrucci, R. J. F. Loos, I. Barroso, N. J. Wareham, F. Karpe, K. R. Owen, L. R. Cardon, M. Walker, G. A. Hitman, C. N. A. Palmer, A. S. F. Doney, A. D. Morris, G. D. Smith, A. T. Hattersley and M. I. McCarthy, A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. *Science*, 2007, **316**, 889-894.

158. K. A. Fawcett and I. Barroso, The genetics of obesity: FTO leads the way. *Trends Genet.*, 2010, **26**, 266-274.

159. R. J. F. Loos and G. S. H. Yeo, The bigger picture of FTO [mdash] the first GWAS-identified obesity gene. *Nat. Rev. Endocrinol.*, 2014, **10**, 51-61.

160. T. Gerken, The obesity-associated FTO gene encodes a 2-oxoglutarate-dependent nucleic acid demethylase. *Science*, 2007, **318**, 1469-1472.

161. G. Jia, Oxidative demethylation of 3-methylthymine and 3-methyluracil in single-stranded DNA and RNA by mouse and human FTO. *FEBS Lett.*, 2008, **582**, 3313-3319.

162. Z. Han, T. Niu, J. Chang, X. Lei, M. Zhao, Q. Wang, W. Cheng, J. Wang, Y. Feng and J. Chai, Crystal structure of the FTO protein reveals basis for its substrate specificity. *Nature*, 2010, **464**, 1205-1209.

163. C. Benedict, T. Axelsson, S. Söderberg, A. Larsson, E. Ingelsson, L. Lind and H. B. Schiöth, Fat mass and obesity-associated gene (FTO) is linked to higher plasma levels of the hunger hormone ghrelin and lower serum levels of the satiety hormone leptin in older adults. *Diabetes*, 2014, **63**, 3955-3959.

164. I.M. Wilson, J. J. Davies, M. Weber, C. J. Brown, C. E. Alvarez, C. MacAulay, D. Schübeler, W. L. Lam, Mapping the Methylome. *Cell Cycle*, 2006, **5**, 155-158.

165. T. Sibbritt, H. R. Patel and T. Preiss, Mapping and significance of the mRNA methylome. *Wiley Interdiscip Rev RNA* , 2013, **4**, 397-422.

166. L. Shen, C.-X. Song, C. He and Y. Zhang, Mechanism and function of oxidative reversal of DNA and RNA methylation. *Annu. Rev. Biochem.*, 2014, **83**, 585-614.

167. B. I. Fedeles, V. Singh, J. C. Delaney, D. Li and J. M. Essigmann, The AlkB family of Fe (II)/α-ketoglutarate-dependent dioxygenases: repairing nucleic acid alkylation damage and beyond. *J. Biol. Chem.*, 2015, **290**, 20734-20742.

168. X. Lu, B. S. Zhao and C. He, TET family proteins: oxidation activity, interacting molecules, and functions in diseases. *Chem. Rev.*, 2015, **115**, 2225-2239.

169. L. Hu, J. Lu, J. Cheng, Q. Rao, Z. Li, H. Hou, Z. Lou, L. Zhang, W. Li, W. Gong, M. Liu, C. Sun, X. Yin, J. Li, X. Tan, P. Wang, Y. Wang, D. Fang, Q. Cui, P. Yang, C. He, H. Jiang, C. Luo and Y. Xu, Structural insight into substrate preference for TET-mediated oxidation. *Nature*, 2015, **527**, 118-122.

170. L. Tan and Y. G. Shi, Tet family proteins and 5-hydroxymethylcytosine in development and disease. *Development*, 2012, **139**. 1895-1902.

171. P. A. Aas, M. Otterlei, P. O. Falnes, C. B. Vagbo, F. Skorpen, M. Akbari, O. Sundheim, M. Bjoras, G. Slupphaug, E. Seeberg and H. E. Krokan, Human and bacterial oxidative demethylases repair alkylation damage in both RNA and DNA. *Nature*, 2003, **421**, 859-863.

172. J. C. Delaney and J. M. Essigmann, Mutagenesis, genotoxicity, and repair of 1-methyladenine, 3-alkylcytosines, 1-methylguanine, and 3-methylthymine in alkB Escherichia coli. *Proc. Natl. Acad. Sci. USA*, 2004, **101**, 14051-14056.

173. P. Ø. Falnes, A. Klungland and I. Alseth, Repair of methyl lesions in DNA and RNA by oxidative demethylation. *Neuroscience*, 2007, **145**, 1222-1232.

174. M. Gehring, W. Reik and S. Henikoff, DNA demethylation by DNA repair. *Trends Genet.*, **25**, 82-90.

175. J.-K. Zhu, Active DNA demethylation mediated by DNA glycosylases. *Annu. Rev. Genet.*, 2009, **43**, 143-166.

176. Y.-F. He, B.-Z. Li, Z. Li, P. Liu, Y. Wang, Q. Tang, J. Ding, Y. Jia, Z. Chen, L. Li, Y. Sun, X. Li, Q. Dai, C.-X. Song, K. Zhang, C. He and G.-L. Xu, Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science*, 2011, **333**, 1303-1307.

177. Y. Yue, J. Liu and C. He, RNA N6-methyladenosine methylation in post-transcriptional gene expression regulation. *Genes & Development*, 2015, **29**, 1343-1355.

178. C. He and P. Cole, Introduction: Epigenetics. *Chem. Rev.*, 2015, **115**, 2223-2224.

179. Y. Yang, B.-F. Sun, W. Xiao, X. Yang, H.-Y. Sun, Y.-L. Zhao and Y.-G. Yang, Dynamic m6A modification and its emerging regulatory role in mRNA splicing. *Sci. Bull.*, 2015, **60**, 21-32.

180. M. S. Ben-Haim, S. Moshitch-Moshkovitz and G. Rechavi, FTO: linking m6A demethylation to adipogenesis. *Cell. Res.*, 2015, **25**, 3-4.

181. C. R. Allerson, S. L. Chen and G. L. Verdine, A chemical method for site-specific modification of RNA: the convertible nucleoside approach. *J. Am. Chem. Soc.*, 1997, **119**, 7423-7433.

182. D. Dominissini, S. Nachtergaele, S. Moshitch-Moshkovitz, E. Peer, N. Kol, M. S. Ben-Haim, Q. Dai, A. Di Segni, M. Salmon-Divon, W. C. Clark, G. Zheng, T. Pan, O. Solomon, E. Eyal, V. Hershkovitz, D. Han, L. C. Doré, N. Amariglio, G. Rechavi and C. He, The dynamic N1-methyladenosine methylome in eukaryotic messenger RNA. *Nature*, 2016, **530**, 441-446.

Acknowledgement

I would like to thank Prof. Dr. Dennis Gillingham for taking me as his PhD student and all his support during my study. I am deeply grateful for all the knowledge he has taught me, the encouragement and inspiration he has given to me. He strongly impacts my way of thinking and attitude to be a better researcher. I am also very grateful for all his kind support to help me through many difficult situations in my life. I will always benefit from his advice and encouragement.

I would like to thank Prof. Dr. Florian Seebeck for accepting the co-examination of my PhD thesis.

Furthermore, I would like to thank Prof. Dr. Christof Sparr for chairing my PhD defense.

I would like to thank students who worked under my supervision and contributed to my thesis. Master student Basilius Sauter; Wahlpraktikum students: Basilius Sauter, Seraina Blümli, Lorenzo D. Bizzini.

I would like to thank Matthias Knop, Anja Stampfli, Daniel Bachmann, Cedric Stress and Pascal Schmidt for proofreading my thesis. I would like to thank Laetitia Misson, Matthias Knop and Benoît-Joseph Laventie for their kind help with my protein production.

I would like to thank Dr. Daniel Häussinger and his group for their excellent NMR services, Dr. Heinz Nadig for express HRMS service and the Markus Ast with his Werkstatt-Team for taking care of immediate fixes in the lab.

I would like to thank my lab mates Daniel Bachmann, Pascal Schmidt, Kiril Tishinov, Stefanie Geigle, Linna Zhou, Cedric Stress, Vijay Shanker and Saule Zhanybekova for all the help and support they gave to me, making me feeling at home in Basel. The precious time we had together is so memorable that it always warms my heart.

I would like to thank all the friends that I have met in Basel. I really cherish the great time we spent together: the girls nights, the third floor parties, metal festivals, Fasnachts, summer barbecues and Chinese food. Among them I especially thank Laetitia Mission, Steffi Geigle, Ewa Milopolska, Sylvie Drayss, Almudena Gallego for all the girly time we spent together and my Chinese friends Ruixia Chao, Aping Niu, Xingwei Guo, Jian Gao, Dalin Wu, Fengfeng Cai for all the delicious Chinese food we enjoyed together.

I would like to thank Bachmann Family for all the great time we spent together, and their kindness of taking care of me like a daughter.

I would like to thank Daniel Bachmann for being there with me and cheering me up during the hardest time of my life. I also really appreciate that he helped me a lot to set up a new life in Basel in the beginning of my PhD study and later on all the support and encouragement during the four years study.

I would like to thank my uncle Zhaofu Fei and my aunt Yingzi Lu for always taking care of me and their strong support for my PhD study.

Foremost, I would like to thank my family for their unlimited love and support. There is no word that could express the gratitude that I owe to them.