Dissertation zur Erlangung des Doktorgrades
der Fakultät für Chemie und Pharmazie
der Ludwig-Maximilians-Universität München

Construction and Functionalization of Oligonucleotides

Meng Su
aus
Peking, Volksrepublik China

2016

Erklärung

Diese Dissertation wurde im Sinne von § 7 der Promotionsordnung vom 28. November 2011 von Herrn Prof. Dr. Thomas Carell betreut.

Eidesstattliche Versicherung

Diese Dissertation wurde eigenständig und ohne unerlaubte Hilfe erarbeitet.

München, den

Meng Su

Dissertation eingereicht am 26.04.2016

1. Gutachter: Prof. Dr. Thomas Carell

2. Gutachter: Prof. Dr. Ernst Wagner

Mündliche Prüfung am 28.06.2016

In Dankbarkeit
meinen Eltern gewidmet.

In me the tiger sniffs the rose.

---- Siegfried Sassoon

Parts of this thesis were published:

1. M. Su, M. T.-Gamasa, S. Serdjukow, P. Mayer, T. Carell

Chem. Commun., **2014**, *50*, 409-411.

Synthesis and properties of a Cu^{2+} complexing pyrazole ligandoside in DNA

2. M. Su, M. T. -Gamasa, T. Carell

Chem. Sci., **2015**, *6*, 632-638.

DNA based multi-copper ions assembly using combined pyrazole and salen ligandosides

3. M. Su, A. Kirchner, S. Stazzoni, M. Muller, M. Wagner, A. Schroder and T. Carell

Submitted

Analysis of single 5-formylcytosines at defined genomic sites reveals semi-permanent character of the base

Further publications:

1. M. T.–Gamasa, S. Serdjukow, M. Su, M. Müller, T. Carell

Angew. Chem. Int. Ed. **2015**, *54*, 796-800.

"Post-It" Type Connected DNA Created with a Reversible Covalent Cross-Link

Parts of this thesis were presented at conferences:

"Synthesis and characterization of a novel metal-pyrazole base pair."
Symposium des Sonderforschungsbereichs 623 (May 2013, Heidelberg, Germany)

"Metal complexing pyrazole base pair."
2nd International Symposium on Integration of Molecular Components (January 2014, Hannover, Germany)

"Pyrazole ligandoside: preparation, properties and prospection."
29th Annual Congress of Chinese Chemical Society (August 2014, Beijing, China),
Challenges in Chemical Biology: International Symposia on Advancing the Chemical Sciences 16 (June 2015, Zurich, Switzerland)

"Site-specific Formylcytosine Probe for the Loci-specific Quantification in the Genome"
Symposium des Sonderforschungsbereichs 1032 (January 2016, Munich, Germany)

Oral presentations on:

“DNA as ligand.”

26. Jahrestagung der Gesellschaft Chinesischer Chemiker und Chemieingenieure in der Bundesrepublik Deutschland (November 2014, Frankfurt am Main, Germany)

“From canonical nucleoside to functional oligonucleotides.”

27. Jahrestagung der Gesellschaft Chinesischer Chemiker und Chemieingenieure in der Bundesrepublik Deutschland (August 2015, Jena, Germany)

“Epigenetic base identification at specific loci of the genome.”

ClickGene: Web conference (January 2016)

Danksagung

Viele haben durch ihre Hilfe und Unterstützung zum Gelingen dieser Dissertation beigetragen. Ihnen allen gilt mein herzlicher Dank!

Allen voran möchte ich mich besonders bei meinem Doktorvater Herrn Prof. Dr. Thomas Carell für seine Einladung, die spannende Aufgabenstellung, die finanzielle Unterstützung, seine stetige Diskussionsbereitschaft, den ungebremsten Ideenfluss und seinen Optimismus und Enthusiasmus bedanken. Durch die Aufnahme in seinem hervorragend ausgestatteten Arbeitskreis wurde das Gelingen dieser Arbeit erst ermöglicht.

Den Mitgliedern meiner Prüfungskommission danke ich für ihre Bemühungen bei der Evaluierung meiner Arbeit. Insbesondere bedanke ich mich bei Prof. Ernst Wagner für die Übernahme des zweiten Gutachters.

Weiterhin danke ich auch Frau Slava Gärtner für die große Hilfsbereitschaft und Unterstützung in allen organisatorischen Belangen. Außerdem geht mein Dank an Frau Sabine Voß für ihre ausgezeichnete Organisation im Labor. Frau Kerstin Kurz und Kristof Hufnagel danke ich für ihre Hilfsbereitschaft Chemikalien in kürzester Zeit zu besorgen. Den Mitarbeitern der analytischen Serviceabteilungen der LMU München danke ich für ihre große Hilfe. Insbesondere gilt mein Dank Dr. Werner Spahl für seine Hilfe bei der Durchführung der massenspektrometrischen Untersuchungen.

Ganz herzlich bedanke ich mich bei Dr. Markus Müller für seine zahlreichen Ratschläge im Bereich der Biochemie für weitere Experimente. Ohne ihn wäre diese Dissertation nicht möglich gewesen.

Dr. Maria Tomas-Gamasa und Sascha Serdjukow danke ich für die erfolgreiche Zusammenarbeit. Dr. Karola Gasteiger und Barbara Steigenberger, danke ich für das gelegentliche Aufreinigen der Bst Pol I. Arne Schröder danke ich für die Synthese von Formylcytosinphosphoramidit. Dr. Sandra Koch, Dr. Verónica López-Carrillo, Dr. Stefan Schießer, Felix Gnerlich, Michael Gattner danke ich für die Hilfe und Unterstützung im alltäglichen Leben.

Ich danke Dr. Markus Müller, Dr. Iacovos N. Michaelides, Dr. Le Han, Samuele Stazzoni and Xiaolei Feng für das Korrekturlesen der Dissertation.

Ich bedanke mich bei meinen fleißigen Bachelorstudenten Shuo Li und Stephan Blum für ihre Mithilfe und Unterstützung, sowie die Korrekturlesen des deutschen Teils meiner Dissertation. Ich wünsche euch für eure Zukunft alles Gute.

Den restlichen Mitgliedern unseres Arbeitskreises danke ich für die angenehme Arbeitsatmosphäre und den großen Zusammenhalt innerhalb der Gruppe der vergangenen Jahre. Insbesondere bedanke ich mich auch bei Dr. Stefan Schießer and Dr. Michael Ehrlich von denen ich viele Kenntnisse in der organische Chemie gelernt habe.

Allen meinen chinesischen Freunden in München möchte ich für viele anregende Gespräche, schöne Abende und Wochenenden danken. Ich danke meinen Freunden in Deutschland, China und weltweit, die mich nicht vergessen haben. Insbesondere danke ich Fan Yang (Uni. Stockholm), Dr. Yongxiang Zheng (Uni. Sichuan) für die Diskussion und den Austausch von Meinungen und Ideen. Dr. Hongdong Hao (AK Trauner), Dongsheng He (AK Wagner), Kai Chen (TUM), Anqi Zhu (LMU) und Yiwen Zhan (Uni. Leipzig) danke ich auch für unsere Freundschaft und ihre Hilfe während meines Aufenthalts in München.

Zuletzt gilt mein größter Dank meiner Familie. Meiner Mutter und meinem Vater danke ich von Herzen für ihre grenzenlose Unterstützung und ihr Vertrauen in mich.

Kurz gesagt sind diese glückseligen und unvergesslichen Jahre Impulse meiner weiteren Karriere. Ich wünsche Ihnen allen ein wunderbares Leben und eine rosige Zukunft.

Vielen Dank!

Meng Su

Ludwig-Maximilians-Universität München

Table of contents

1. Abstract

Oligonucleotides are significant tools in chemical biology and show wide applications. This thesis contains two projects concerning the construction of ligandosides and functionalization of oligonucleotides for formylcytosine detection.

In the first project, a novel metal-base pair based on the pyrazole ligand was developed. The synthesis of the pyrazole ligandoside comprised the preparation of a protected base building block and a cuprate mediated *C*-glycosylation as the key step (Figure 1-1). The correct β-configuration of the nucleoside was confirmed by X-ray crystallography. The ligandoside precursor was incorporated into numerous oligonucleotides by automated DNA synthesis (Chapter 4.2).

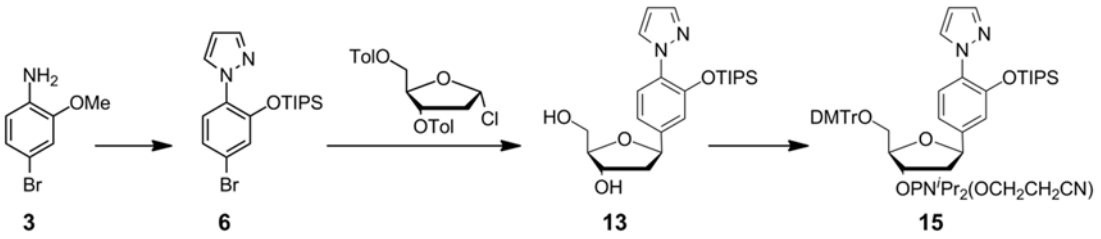


Figure 1-1 Synthesis of the protected pyrazole nucleoside **13** and the phosphoramidite **15**.

With a homo-pyrazole base pair inserted, duplex stability increased by 9°C after incooperation of one copper ion. The chelating performance depends on deprotonation of the phenol group of the ligandoside. Up to ten copper ions can be coordinated inside the duplex. Compared with the bridging salen base pair, the non-bridging pyrazole base pair shows a kinetical preference for complexation (Figure 1-2, Chapter 4.3).

The pyrazole triphosphate is accepted by the Therminator polymerase which extends the primer. The unsatisfactory efficiency, however, hinders its application in PCR (Chapter 4.4.1). A duplex with five pyrazole-copper pairs was applied as a chiral catalyst in a model Diels-Alder reaction, which allowed to reach an ee value of 39% (Chapter 4.4.2). The novel ligandoside sheds light on how DNA may be used as a catalyst in organic reactions and enlightens further design and optimization of

ligandosides.

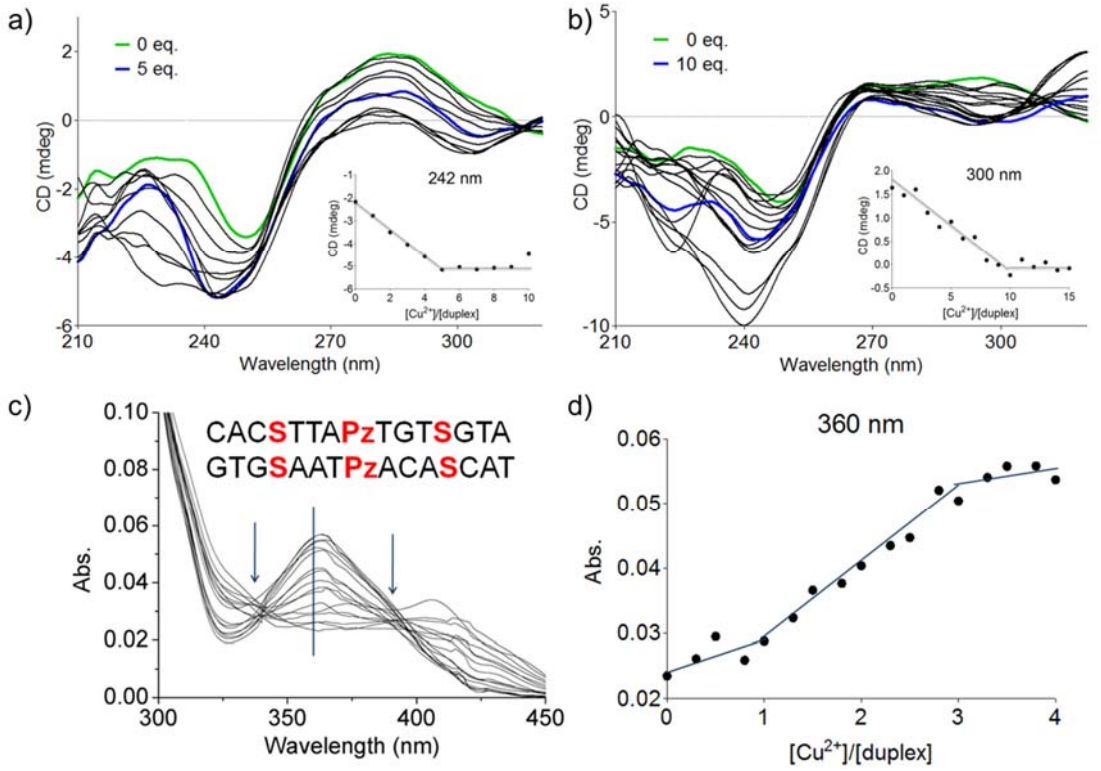


Figure 1-2 a) and b) CD titration of the duplexes containing 5 or 10 Pz pairs; c) and d) UV titration of duplex showing Cu^{2+} kinetical preference.

The second project is focused on developing a new fdC sequencing method based on an oligonucleotide probe connected to a hydroxylamine linker. The linker was selected using a combinational chemistry strategy. The most suitable linker at an n+4 position of the probe strand was known to react with the fdC in the target strand irreversibly. Because the probe and the target strand hybridized to form a duplex, the probe reacted with the target fdC with high positional specificity (Figure 1-3, Chapter 4.2). The reaction is limited to fdC and can tolerant single nucleotides polymorphisms in the target. Multiple fdC probes can be applied together.

Enzymatic digestion and primer extension experiments were performed on the cross-linked oligonucleotide towards LC-MS and PCR detection (Chapter 4.3). After oxime formation, the duplex can be digested into dinucleotides but they cannot be detached. The probe strands hinders the Taq polymerase to pass through the target strand.

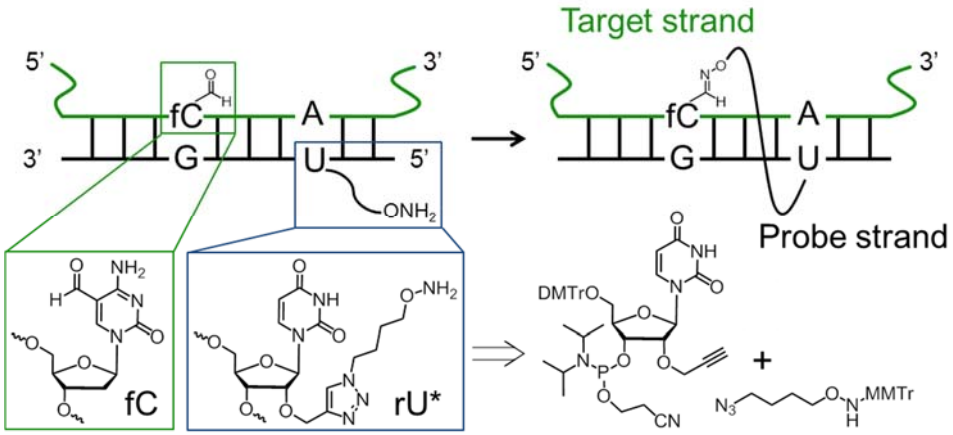


Figure 1-3 Schematic representation of the fdC probe and target crosslinking.

A method for relative fdC quantification was developed using the described probe (Figure 1-4 Chapter 4.4). After crosslinking, the fdC probe was ligated to an adapter strand, wrapped in nanodroplets and replicated using PCR. The signals were counted and compared to a reference amplicon. 10-fold increase of fdC was observed at one position in an exon in $Tdg^{-/-}$ mES cells compared to Dnmt TKO cells, and 2-fold compared to $Tdg^{+/-}$ cells. The method can be applied to other targets of interest in order to track the dynamic change of the epigenetic fdC bases.

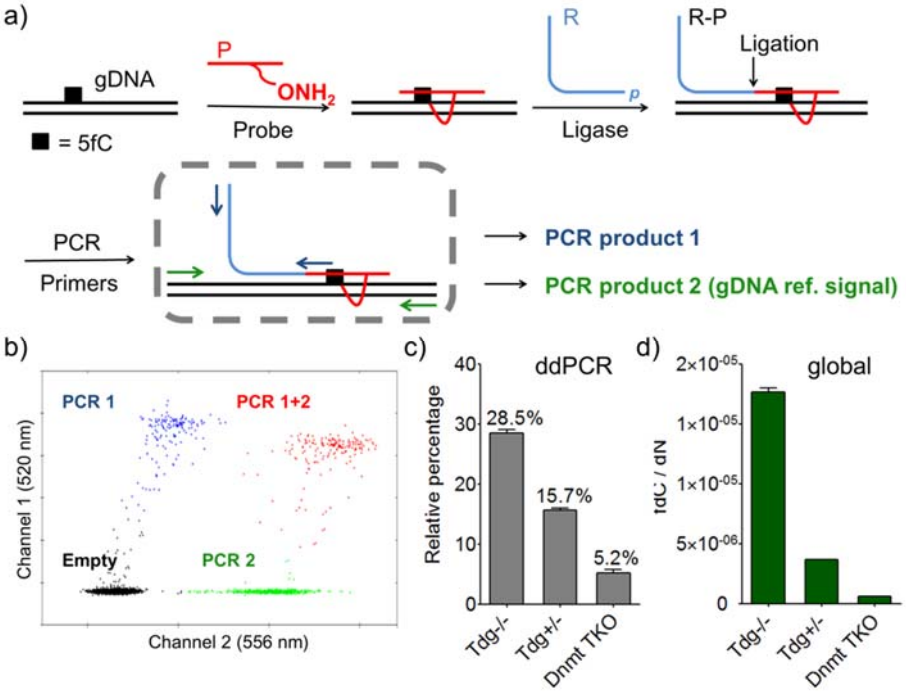


Figure 1-4 Schematic representation of the relative fdC quantification method and results.

2. Zusammenfassung

Oligonukleotide sind wichtige Werkzeuge in der chemischen Biologie und erlauben breite Anwendungen. Diese Arbeit gliedert sich in zwei Projekte. Den Einbau von Metallo-Basen in DNA und die Funktionalisierung von Oligonukleotiden zur Detektion von 5-Formylcytosin.

Im Zuge des ersten Projekts wird ein neues Metall-Basenpaar entwickelt, welches auf dem Pyrazolliganden basiert. Die Synthese des Pyrazolliganden erforderte die Darstellung eines geschützten Basenbausteins und eine Kuprat-vermittelte *C*-Glykosylierung als Schlüsselschritt (Abb.1-1). Die gewünschte β-Konfiguration des Nukleosids wurde durch eine Kristallstrukturanalyse bestätigt. Der Ligandosid-Vorläufer wurde in zahlreiche Oligonukleotide mittels automatisierter DNA-Synthese eingebaut (Kapitel 4.2).

Liegt ein Homo-Pyrazol-Einzelbasenpaar in einem Strang vor, so erhöht die Zugabe von Cu^{2+} den Schmelzpunkt des Doppelstrangs um 9°C. Die Komplexierung hängt von der Deprotonierung des Phenols im Ligandosid ab. Duplexe mit 10 Pyrazol-Basenpaaren ermöglichte das Stapeln von bis zu zehn Kupferionen im Inneren eines Doppelstranges. Verglichen mit dem überbrückten Salen-Basenpaar, zeigte das nicht überbrückten Pyrazole-Basenpaar eine kinetische Präferenz zur Komplexierung (Abb.1-2, Kapitel 4.3).

Die enzymatische DNA-Synthese erfolgt durch die Therminator Polymerase. Diese katalysiert die Polymerisation des Pyrazolnukleosid-triphosphats zu Polynukleotiden. Die unzureichende Effizienz steht weiteren Anwendungen in der PCR jedoch im Wege (Kapitel 4.4.1). Ein Duplexstrang mit fünf Pyrazol-Kupferpaaren wurde als chiraler Katalysator in einer Diels-Alder-Reaktion eingesetzt und erreicht einen ee-Wert von 39% (Kapitel 4.4.2). Der neue Ligandosid kann als DNA- Katalysator in organischen Reaktionen genutzt werden.

Das zweite Projekt konzentriert sich auf die Entwicklung einer neuen

Sequenzierungsmethode für fdC. Diese basiert auf einer Oligonukleotidsonde verbunden mit einem Hydroxylamin-Linker. Der Linker wurde durch kombinatorische Chemie entwickelt. Der am besten geeignete Linker an einer geeigneten Position auf dem Sondenstrang kann mit dem fdC des Zielstrangs irreversibel reagieren. Da die Sonde- und der Zielstrang zu einem Duplex hybridisieren, reagierte die Sonde mit dem erwünschten fdC mit Positionsspezifität (Abb.1-3, Kapitel 4.2). Die Reaktion ist spezifisch für fdC und toleriert Einzelnukleotid-Polymorphismus im Zielstrang. Mehrere fdC Sonden können gemeinsam angewendet werden.

Enzymatische Hydrolyse und Primer-Extension-Experimente wurden an den vernetzten Oligonukleotiden durchgeführt. Dadurch konnten die Sonde durch LC-MS- und PCR-Methoden analysiert werden (Kapitel 4.3). Nach der Oximreaktion können die Doppelstränge in Dinukleotide verdaut aber nicht voneinander getrennt werden. Durch die kovalente Bindung zwischen den Sondensträngen wird die Taq-Polymerase gestoppt. Dies unterbindet die Vervielfältigung.

Eine relative fdC Quantifizierungsmethode wurde mit dieser Sonde entwickelt (Abb.1-4, Kapitel 4.4). Nach der Vernetzung, wurden die fdC Sondestränge mit einem Adapterstrang ligiert, in Nanotröpfchen verteilt und repliziert. Die Signale in den Tröpfchen wurden gezählt und mit einem Referenzamplicon verglichen. So wurde in Tdg knokcout mES Zellen in Vergleich zu einem Dnmt Dreifachknockout die zehnfache Menge an fdC beobachtet und im Vergleich zu $Tdg^{+/-}$ Zellen die zweifache Menge. Das Verfahren kann auf eine Vielzahl von Zielen angewandet werden.

3. Introduction

3.1. Nucleosides and ligandosides

3.1.1. Nucleosides and their artificial modifications

3.1.1.1. Deoxyribonucleic acid

Nucleic acids are essential material for the storage and transmission of the genetic information. Due to their difference in the pentose, nucleic acids are divided into deoxyribonucleic acid (DNA) and ribonucleic acid (RNA).

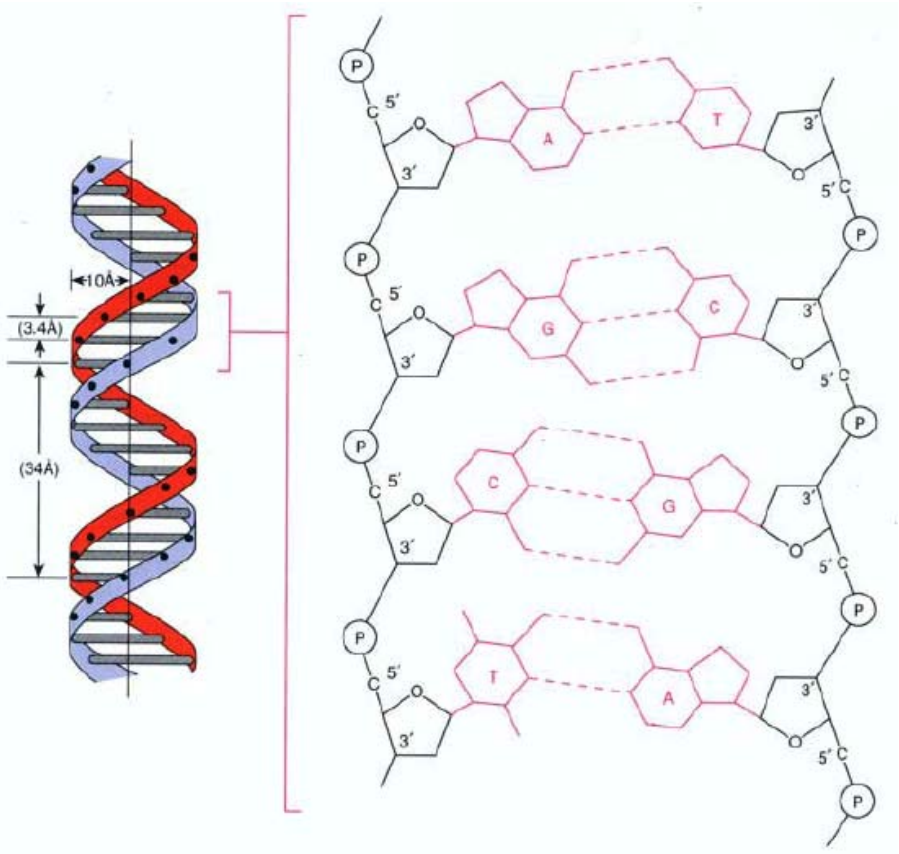


Figure 3-1 Double helix structure and chemical structure of DNA.

In general, DNA molecules are made up of four types of nucleotides linked together by phosphodiester bonds. Each nucleotide in DNA consists of a nitrogenous base, a sugar deoxyribose, and a phosphate residue. Two types of bases are found in DNA: purines, adenine (A) and guanine (G); and pyrimidines, cytosine (C) and thymine (T).

Besides, DNA may contain several additional modified bases, namely 5-methylcytosine (mdC) and its oxidized derivatives, N^6-methyladenine (m^6A),[1, 2] and N^4-methylcytosine (m^4A),[3] which result from the post-replicative modification of DNA by DNA methyltransferases. The former, mdC and oxidized derivatives, will be discussed in Chapter 3.2.

The nucleosides are linked by phosphodiester bonds, joining the 5' carbon of one deoxyribose to the 3' carbon of the next, to form a polynucleotide chain, as depicted in Figure 3-1. An oligonucleotide is a short strand of DNA, and a nucleoside refers to the nitrogenous base and the pentose.

Polynucleotides exist as single strand molecule or as double strand structure, in which two strands are held together by hydrogen bonds between the bases. The antiparallel double helix model was put forward in 1953 by J. D. Watson and F. H. C. Crick based X-ray diffraction measurements.[4, 5] The hydrogen bonds between A and T, G and C inside the double helix are the molecular basis for the biological function of DNA. Phosphate backbones with negative charges outside the helix are neutralized by metal ions or proteins inside the cell.

The B-DNA form is the most conventional nucleic acid helix. With a diameter of 2 nm, it has a major groove and a minor groove. Each base in B-DNA is separated by 0.34 nm along the length of the double helix; each turn consists of 10.5 bases. A-DNA form is a right-handed double helix fairly similar to B-DNA form, but with a shorter, more compact helical structure whose base pairs are not perpendicular to the helix axis. Z-DNA from is a left-handed double helical structure, deemed as one of three biologically active double helical structures along with A- and B-DNA.

DNA duplexes wind on histones and are packaged in nucleosomes in eukaryotic cell nuclei. The superhelix unwinds to replicate itself in S-phase of the cell cycle with the help of helicase, single-strand DNA-binding protein, DNA polymerase I/III and DNA ligase. The information contained in a section of DNA is transcribed in the form of a

newly assembled piece of messenger RNA (mRNA). This process is called as transcription. Transcription factors are proteins that bind to specific DNA sequences and control the rate of transcription. Furthermore, RNA polymerase also plays a role. In prokaryotic cells, there is no spatial separation of transcription and translation. In eukaryotic cells, mature mRNA finds its way to the ribosome and gets translated into polypeptides. The above process is named as the central dogma of molecular biology. Figure 3-2 provides a graphical elucidation.

In the natural information flow, 64 triplet codons are possible in the standard genetic lexicon with four bases (A, C, G, T). As the genetic code is degenerate, most natural amino acids are encoded by more than one triplet codon. As a result, only 20 proteinogenic amino acids are encoded by all 64 codons, leaving thousands conceivable amino acids behind.

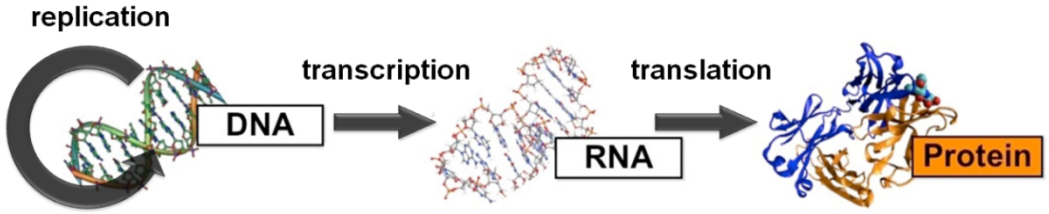


Figure 3-2 Schematic representation of the central dogma of molecular biology. Unusual flows of information, i.e. reverse transcription and RNA replication, are not showed.

3.1.1.2. DNA reconstruction: XNA world

Since the discovery of the molecular structure of DNA, its specific recognition and the potential for information storage urged the chemists to broaden the knowledge of the DNA structure and to endow them with new outstanding properties. In the past decades, modern organic synthesis methods, solid-phase synthesis, and polymerase chain reaction (PCR) technology made chemically modified nucleosides and oligonucleotides technically feasible. Consequently, a collection of sugar-, phosphate- and base- modified nucleic acid mimics have been synthesized and evaluated. These synthetic genetic polymers are called "*xeno*- nucleic acids" (XNA), proposed by P. Herdewijn and P. Marliere,[6] and as shown in Figure 3-3. As in all other endeavors that

to synthesize analogs of natural products, the goal of XNA investigation is to learn how DNA and RNA work, and to provide them with new properties. Nowadays, "XNA" constantly gain wider applicability in biotechnology, nanotechnology, molecular biology, and pharmaceutical sciences.

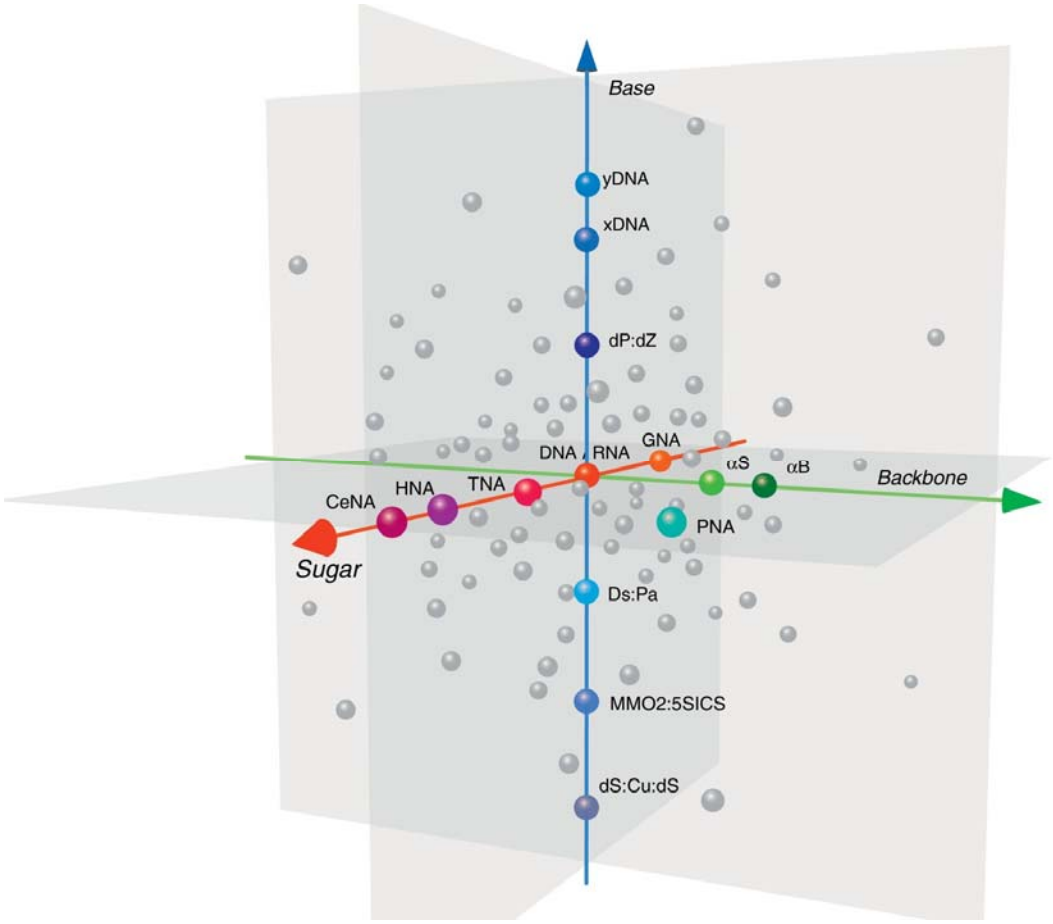


Figure 3-3 *Xeno*- nucleic acids (XNA) chemical space. The x-, y- and z-axes represent sugar, base and backbone modifications, respectively. The figure is adapted from reference.[15]

In practice, oligonucleotides with non-canonical nucleotides can be synthesized with the solid-phase synthesizer. The concept "solid-phase synthesis", originally invented for peptide synthesis, was put forward in 1963 by R. B. Merrifield.[7] It was of such importance that Merrifield was awarded the Nobel Price in Chemistry in 1984 for the development of the field. In short, the amino acids are immobilized on a solid phase and undergo repeated cycles of deprotection-wash-coupling-wash until the sequence terminates. The peptide is then cleaved from the solid-phase and purified.

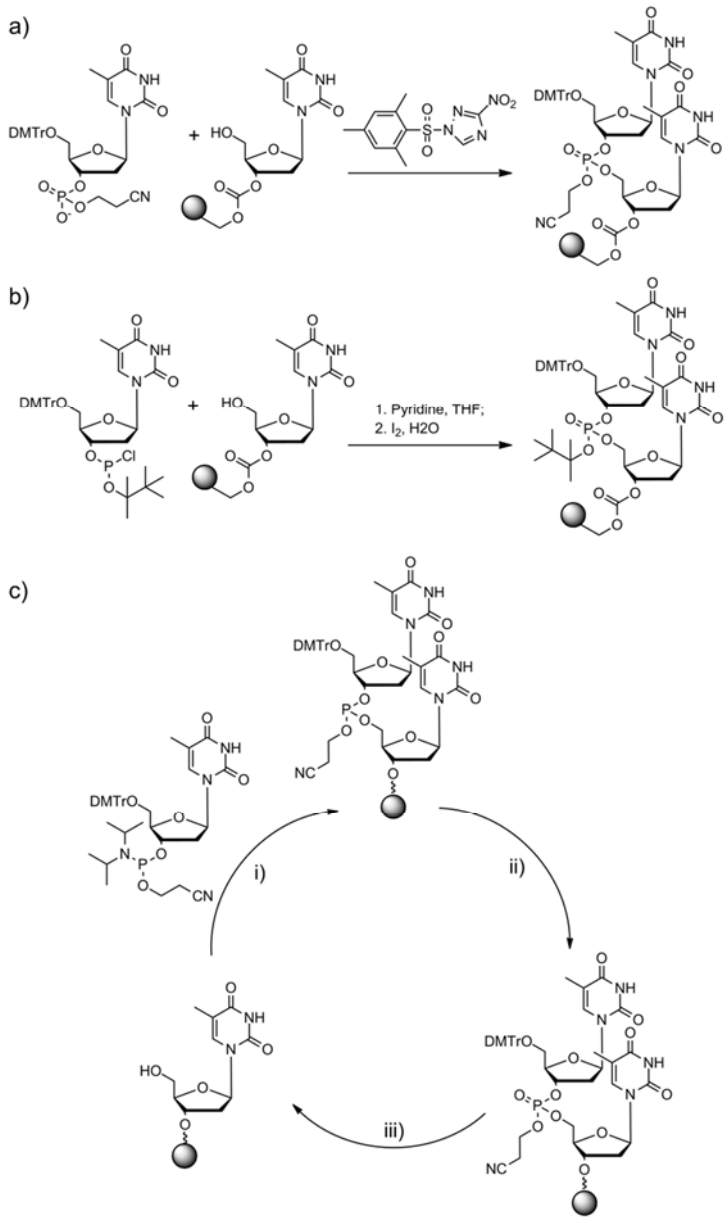


Figure 3-4 Chemistry for oligonucleotide solid-phase synthesis: a) phosphotriester and b) phosphite-triester approach invented by Letsinger et al.; c) phosphoramidite approach synthetic cycle, i) coupling, ii) capping and oxidation, iii) deprotection. Gray spheres represent the solid phase.

Another scientist, R.L. Letsinger, who was also working on the peptide solid phase synthesis,[8,9] after Merrifield's publication in 1963, decided to focus on oligonucleotide solid phase synthesis.[10] Phosphotriester[11-13] and phosphite-triester[14, 15] chemistry were invented (Figure 3-4a, b) by him, but both methods suffered from serious drawbacks such as low coupling efficiency, moisture sensitivity, and undesired 3'-3' by-products. In early 1980's, M. H. Caruthers, a former student of Letsinger, invented the phosphoramidite approach inspired by the phosphite-triester

method.[16, 17] The phosphoramidite monomers are more stable and can be activated before coupling by adding tetrazole, a weak acid. After coupling, the solid-phase is oxidized as in phosphite-triester chemistry. The unreacted strands which are left with a free 5'-hydroxy group are capped using a mixture of acetic anhydride and *N*-methylimidazole. In this way, a failed strand's synthesis is terminated. Deprotection the dimethoxytrityl (DMTr) group at 5' terminus with trichloroactate acid leads the cycle into the next round (Figure 3-4c). The synthesis cycle, ingenious, simple and feasible, has remained virtually unchanged since then.

Thanks to the automated solid-phase synthesizer later developed by Caruthers and co-workers, oligonucleotides containing canonical nucleotides and chemical modifiers became increasingly accessible. Chemical modifications corresponding to the three parts of the nucleotide are discussed below.

Modification of the phosphate diester

As a part of the backbone, the polyanionic sugar-phosphodiester unit force the nucleotides into a suitable distance for Watson-Crick pairing. On the same strands, Coulombic repulsion between the phosphodiesters warrants that the single strand can adopt an extended structure in the buffer system. The stretch-out conformation is crucial for the pre-organization process needed for the strand to bind its complement. Furthermore, the repeated anionic groups maintain the DNA and RNAs unified physicochemical properties. Unlike peptides, the properties of oligonucleotides are independent of the nucleobase sequence so that DNA mutations do not change the chemical and physical properties of the molecular. Figure 3-5 summarizes the reported modified phosphodiester linkage.

The phosphate diester backbone is the natural substrate of phosphodiesterases. Replacing the diester bond will alter the biological stability of the DNA duplex. The most common modification is formation of thiophosphate esters or phosphorothioates (Figure 3-5b).[18, 19] With the oxygen replaced by a sulfur atom, the phosphorothioate is

still toleranted by DNA polymerases. It causes only a slight change of the thermostability. But the stability toward phosphodiesterase is significantly enhanced.

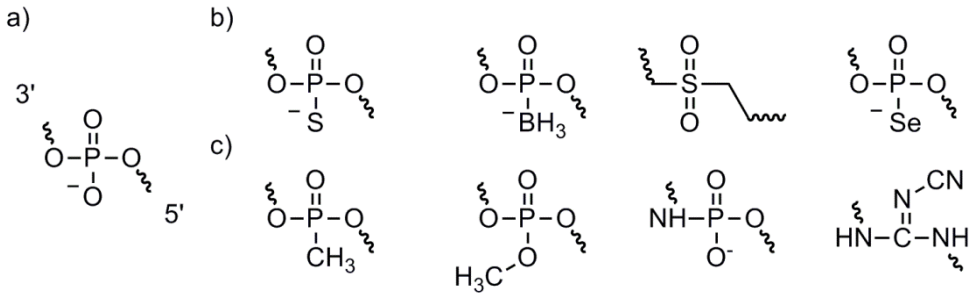


Figure 3-5 Examples of phosphodiester mimics: a) canonical phosphodiester; b) phosphorothioate, boranephosphonate, dimethylene sulfone,[20] and phosphoselenoate;[21] c) methyl-phosphonate,[22] methylated phosphodiester,[23] phosphoramidite,[24] and *N*-cyanoguanidine linkage.[25]

In 1990, a novel boranephosphonate nucleotide analog (bpDNA) was first reported by the Shaw lab.[26, 27] These nucleotides contain a borane moiety, which replaces the non-bridging phosphate oxygen, resulting in a hydrolytically stable phosphodiester analog. Taking advantage of the reductive BH_3 group, M. H. Caruthers and co-worker demonstrated that bpDNA was capable of reducing Au^{3+} and Ag^+ *in situ*.[28] While boranephosphonate solvolyzed to phosphate diesters, the oligomers generated the corresponding silver and gold nanoparticles. Soon afterward, the bpDNA oligomers were incorporated into a two-dimensional DNA array. *In situ* reduction of Ag^+ on the structure generated a well-defined array of silver nanoparticles.[29]

Modification of the pentose sugar

Tremendous efforts have been made to replace the pentose sugar with various structures. There were different sugar rings or ring-opening backbone structures.

Substitution of the 2' hydroxyl group to fluoro (2'-F) and the hydroxymethyl group (2'-OCH_3), as depicted in Figure 3-6, are known and have numerous beneficial effect, e.g. in antisense,[30] siRNA[31, 32], CRISPR-Cas9[33, 34] systems.

2'-SCF_3 substitution allows monitoring nucleotides by ^{19}F-NMR at micromolar concentration.[35] However, this novel modification exerts a high degree of destabilization in a double helix.[36] Synthetic methods towards 2'-$SeCH_3$ [37] and 2'-N_3

nucleotides[38, 39] were established by the same group. Normal phosphoramidite chemistry combined with alternate phosphotriester chemistry enabled azido group incorporation. 2-Azide nucleic acids later found its applications in RNA interference and oligonucleotide labeling.[39, 40]

Apart from 2' modifications, 4' modifications, i.e. replacing the oxygen atom with CH_2,[41] NAc,[42] S,[43, 44] Se,[45, 46] (Figure 3-6) were found useful in gene-based therapy and crystallography.

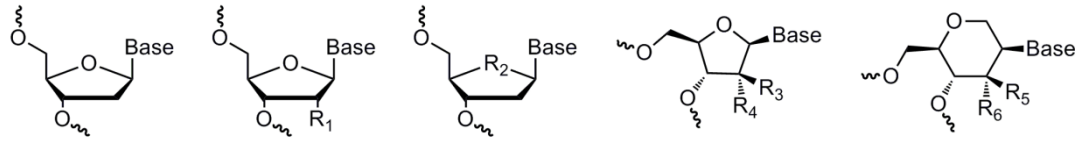


Figure 3-6 Examples of pentose sugar mimic: modification at 2', 4' and a sugar with an expanded ring. The Structure on the left shows the canonical deoxyribose. R_1 = F, OMe, SCF_3, SeMe, N_3. R_2 = CH_2, NAc, S, Se. R_3 = OH, F, R_4 = F. R_5 = H, R_6 = H, OH, OMe, F; R_5 = OH, F, R_6 = H.

A major consideration for the design is to improve the affinity of the synthetic oligonucleotide to the coresponding canonical counter strand. Generally, conformational restriction of the sugar-phosphate backbone is a very successful idea because this change does not interfere with the specificity of Watson-Crick base pairing. Revisiting the structure of the nucleotide (Figure 3-7a and b), there are at least three strategies to incorporate covalent conformational constraints: 1) rotation restriction around torsion angles γ and δ, e.g. tricyclo-DNA;[47] 2) backbone restriction around torsion angles, e.g. α,β-CNA;[48] 3) furanose conformation restriction in a northern type sugar pucker with 2',4'-bridge, e.g. locked nucleic acids (LNA).[49-53] Oligonucleotides consisting of only LNA bases have no double helix binding capacity.[54,55] On the other hand, LNA-containing oligonucleotides adopt Watson-Crick base pairing, resulting in thermostable and sequence-specific double helix interaction with DNA and RNA.[56, 57] The S. Hanessia group combined the first and the third strategy in the example of triNA.[58] (Figure 3-7b) Although the synthesis takes more than 20 steps, triNA stabilizes the double helix even better than LNA.

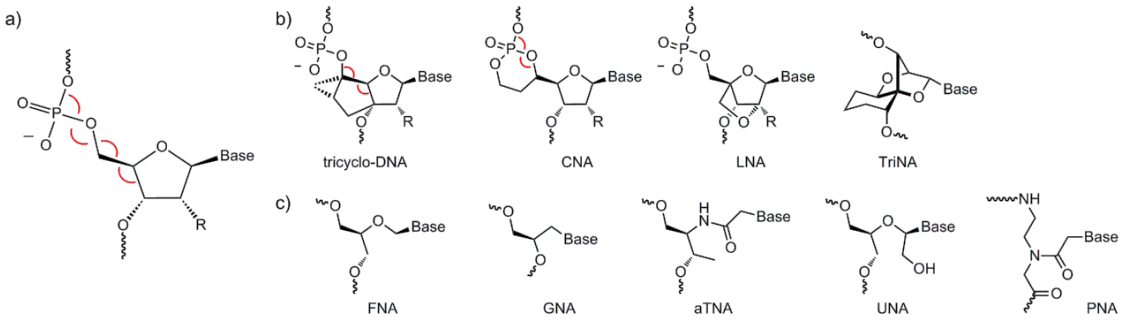


Figure 3-7 Examples of pentose sugar mimic: a) canonical nucleotide with highlighted torsion angles; b) conformational restricted and c) flexible backbones modifications.

Parallel to pentose sugar mimics, acyclic nucleotide mimics obtained the interest of the community (Figure 3-7c). Acyclic nucleosides that are based on glycerol,[59, 60] propanediol,[61-63] butanediol,[64] threninol,[65] serinol,[66] and unlocked nucleic acid[67] have been invented in the past two decades. In most cases, DNA double strands containing the flexible linker are less thermostable than canonical duplexes, supporting the idea that the rigid ribose ring is crucial for duplex preorganization rather than simply holding the nucleobases together.

Among these acyclic nucleoside mimics, peptide nucleic acid (PNA) that is endowed with remarkable hybridization properties has attracted considerable attention. PNA is an achiral and neutral DNA analog, in which the canonical backbone is substituted by *N*-(2-aminoethyl)-*N*-(methylenecarbonyl)glycine.[68] The amide linkage equips PNA with a dual character of a peptide and an oligonucleotide. PNA is chemically stable at a broad range of temperatures and pH, and it is not a substrate of protease, peptidase, or nuclease.[69] The neutral backbone structure leads to a PNA/DNA or PNA/RNA interaction which is stronger than homologous DNA or RNA duplexes. However, lacking charges on the backbone makes PNA less water soluble and PNA easily aggregates.[70] The fidelity of PNA-pairing to canonical DNA is very high.

In a previous report,[71] PNA modifiers served as the backbone to study the photochemical TpT products. The difference of the photoproducts between PNA and phosphate dimer is attributed to the backbone flexibility.

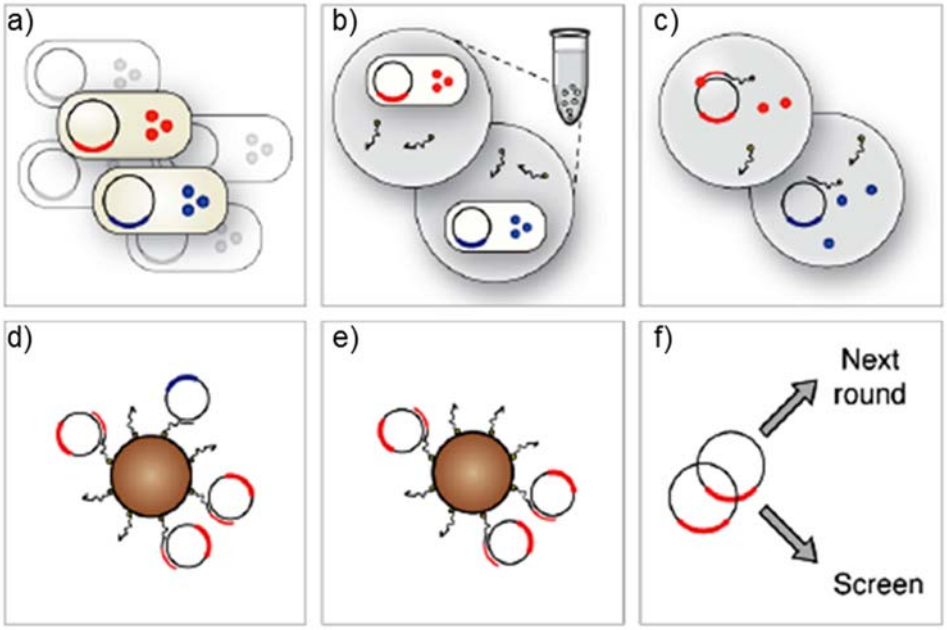


Figure 3-8 Schematic representation of compartmentalized self-tagging (CST): a) and b) polymerase are compartmentalized with primers and modified nucleotides in emulsions to ensure genotype-phenotype linkage; c) CST is based on a positive feedback loop whereby a polymerase tags the plasmid containing its gene by extending a biotinylated primer; d) primer extension stabilizes the metastable primer-plasmid complex allowing capture; e) selection can be further modulated through stringent washing of the streptavidin beads; f) recovered plasmid DNA is amplified and used to start a new round of selection or screening. The Figure is adapted from reference.[72]

To store and process genetic information in DNA analogues, and to discover polymerases that are able to process XNA synthesis, P. Holliger and co-workers developed a selection strategy called compartmentalized self-tagging (CST) (Figure 3-8).[72] In CST, a library of mutant polymerases is compartmentalized with biotinylated primers and modified nucleotides in water-in-oil emulsions. Primer extension stabilizes the metastable primer-plasmid complex allowing capture in proportion to its stability. Recovered plasmid DNA is amplified and used to start a new round of selection and screening. Among the six XNA modifiers, 1,5-anhydrohexitol (HNA) yielded rapid adaptation toward HNA polymerase activity for both replication from DNA and reversed transcription from RNA. Transcription possibilities for flexible nucleic acids (FNA, Figure 3-7c)[73] and threose nucleic acids (TNA)[74] were discovered later. However, the efficiency and fidelity are still precarious, and future work must envolve development of other mutant polymerases

for successful PCR replication.

Metal ions can be incorporated into the nucleic acid backbone. For example, altering the ribose ring to ferrocene made the metal an integral component of the backbone.[75,76] Although ferronucleoside derivatives were later found to have anticancer activity,[77] further benefits of the oligonucleotide are still awaiting discovery.

Modification of the base pair

Four strategies for designing artificial base pairs have been published. As demonstrated in Figure 3-9, bases can pair via hydrogen bonds, hydrophobic interaction, coordination bonds, and reversible covalent bonds.[78] The third pattern will be discussed in detail in section 3.1.2.

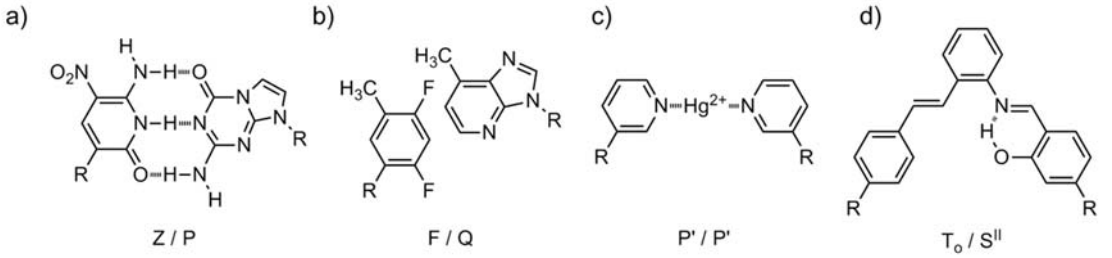


Figure 3-9 Interaction patterns of artificial base pairs: a) altered hydrogen bonds (Z/P pair), b) hydrophobic interactions (F/Q pair), c) nmetal coordination bond (P'/P' pair), and d) reversible covalent bond (T_o/S^{II} pair). R represents deoxyribose. Hydrogen bonds and bonds formed by coordination are shown as hashed bonds.

Different from the straightforwardly designed and accessible phosphate diester and pentose sugar modified bases, artificial hetrocycle base pair patterns are more difficult to compose and synthesize. The dilemma in design lies in mimicking the natural interaction strength in a limited space inside the DNA duplex. The key step for the synthesis is to couple the nucleobase with the deoxyribose sugar moiety.

Formation of *C*-glycosides or *N*-glycosides seems unavoidable, and several methods are available.[79-81] After the glycosidation, the β-anomer has to be separated from the α-anomer. In addition, suitable protecting strategies for the artificial base building blocks need be adopted. The protecting groups should tolerate solid-phase synthesis

and grant mild and efficient deprotection on the column or in aqueous solution after the automated synthesis.

The essence of mimicking Watson-Crick base pairs is size complementarity. Large purines pair (pu) with small pyrimidines (py) baned on hydrogen bonding complementarity between donors (D) from one nucleobase and acceptors (A) from the other. When three hydrogen bonds bridge the base pair, eight bridging patterns can be envisaged. Among them, pyDAA/puADD, i.e. guanine-cytosine, is a canonical base pair; patterns with three proton donors and acceptors in a row (pyDDD/puAAA and pyAAA/puDDD) are excluded because of overstability,[82, 83] but they displayed their prowess in constructing tetrameric macrocycles and novel supermolecules.[84]

In 1989, S. A. Benner and co-workers pioneered the development of base-pairs with five different hydrogen bonding patterns (Figure 3-10a).[85-88] Although the nucleosides with the pyADD heterocycles undergo slow acid-catalyzed epimerization,[89] iso-guanine (pyAAD) could be incorporated into oligonucleotides opposite iso-cytosine (pyDDA) on a DNA template.[85, 90] By translation of the (iso-C)AG codon opposite to CU(iso-G) anticodon on mRNA, the Benner group succeeded in incorporating L-3-iodotyrosine, a non-natural amino acid, into a peptide.[91]

Unfortunately, iso-G has a minor tautomeric form that allows it also to pair with T as shown in Figure 3-10b. Furthermore, iso-C undergoes deamination during solid-phase synthesis.[90] Realizing that the iso-canonical pair seems not to be an ideal candidate to enlarge the genetic alphabet, Benner and co-workers later replaced iso-C with 2-thio-T.[92] Because of the bulky thione moiety, minor tautomeric iso-G could not pair with 2-thio-T efficiently. Thus eliminated the tautomer percentage and increased the amplification fidelity.

At the same time, using T7 RNA polymerase, they incorporated a new base pair "κ/χ", as shown in Figure 3-10a.[86, 93, 94] Although the incorporation of the artificial pair is less efficient compared to that of the natural base pairs, to our best knowledge, these are the earliest examples of artificial base pairs in replication and transcription.

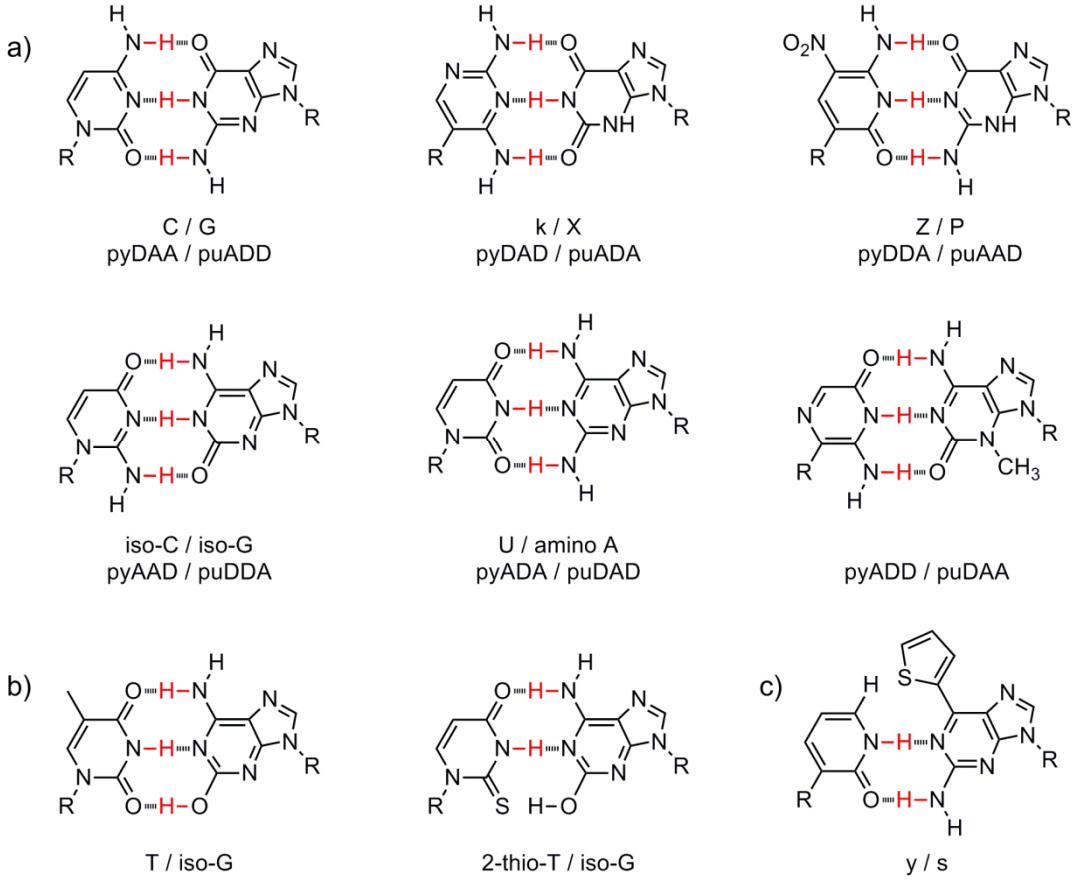


Figure 3-10 Examples of alternative hydrogen-bonding base pair mimics with canonical size: a) via variant hydrogen bonds from Benner *et al.*, pyrimidines are designated by the prefix "py", purines by the prefix "pu", follow the prefix is the order, from the major groove to the minor groove, of acceptor (A) and donor (D) group, letters, and bonds in red highlight the hydrogen donor; b) thymine/tautomeric iso-G pair and 2-thio-T/iso-G pair from Benner *et al.*; c) via variant hydrogen bonds and shape complementarity from Hirao *et al.* R represents pentose sugar. Hydrogen bonds are shown as hashed bonds.

Watson-Crick base pairs are not the only system that can store and transmit genetic information, and certain nucleoside mimics also play a role in the genetic information flow.[95] Chemical biologists are encouraged to question if other base pair can be constructed for the encoding of unnatural amino acids. By expanding the genetic alphabet, researchers aim at an enlarged genetic lexicon, increasing the numbers of non-natural amino acids that can be incorporated into peptides. These efforts do in principle not only provide the accessibility to engineered proteins but also shed light on the transcription and translation mechanisms.

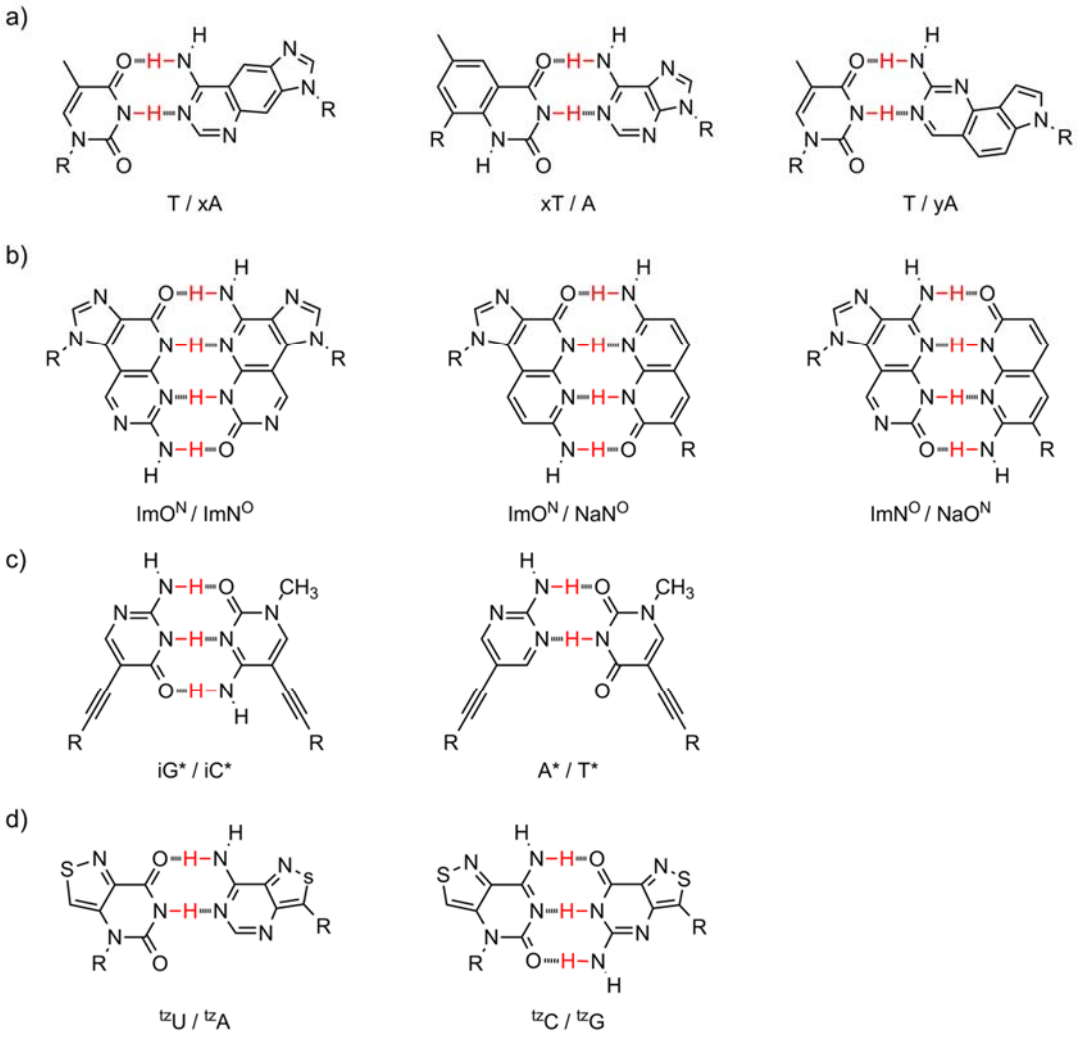


Figure 3-11 Examples of hydrogen-bonding base pair mimics with extended size: a) xDNA and yDNA pair patterns from Kool et at.; b) expanded bases with four hydrogen bridges from Matsuda *et al.*; c) ethynyl trestle DNA from Inouye *et al.*; d) emissive RNA from Tor et al;. R represents deoxyribose. Hydrogen bonds are shown as hashed bonds.

Slightly after Benner, I. Hirao and co-workers synthesized the base pair y/s (Figure 3-10c) with two hydrogen bonds in between. In the y/s pair, shape complementarity was also a factor in base recognition. A yAG codon in the *ras* mRNA recognized the CUs anticodons in aminoacylated tRNA, thereby, incorporation of an unnatural amino acid into the Ras protein was achieved.[96] In addition, the nucleoside "s" (Figure 3-10c) is a strong fluorophore. The translation process enabled the site-specific fluorescent labeling of the RNA molecule.[97] However, the y/s base pair exhibited a lower transcription efficiency compared to the canonical base pair.

In 1977, N. J. Leonard brought forward the idea of size-expanded bases.[98] 25 years

later, E. T. Kool and co-workers embellished this concept with hydrogen-bonding patterns as in xDNA (xA, xT) and yDNA (yA, yT) (Figure 3-11a).[99-101] The enlarged xDNA differs from widened yDNA in the extension direction. Both base pairs stabilized the canonical duplex but replicated with low efficiency and fidelity.[102]

As the same time, A. Matsuda *et al.* described another size-extended base pair with four hydrogen bonds in between (Figure 3-11b). Due to the excessive distance between the two ribose units in ImO^{N}/ImN^{O} base pair,[103] ImO^{N}/NaN^{O} and ImN^{O}/NaO^{N} were put forward as the second generation.[104, 105] Together with the ethynyl trestle base pairs from M. Inouye group (Figure 3-11c),[106] these examples imply the possibility to maintain the duplex stability using aggrandized bases, which suggests the space between the two phosphate backbone is flexible to some extent.

The Y. Tor group synthesized two sets of RNA bases derived from hieno[3,4-d]-pyrimidine[107] and isothiazolo[4,3-d]pyrimidine (Figure 3-11d).[108] As fluorescent nucleoside analogs retaining Watson-Crick pairing patterns, the new alphabet should be useful in nucleotide labeling and detection.

Although improvements of these analogs with H-bonding patterns continuously surprise the community, works of the E. T. Kool group proved that hydrogen bridges were not essential for polymerase-mediated base pair synthesis.[109-111] Packing and hydrophobic interactions, the forces which are prevalent and well-documented in protein folding, assist the base pairing inside DNA duplex as well. Unlike to design the hydrogen-bonding and electrostatic interaction patterns, to predict the impact of aromatic stacking, hydrophobic or CH/π interactions, and shape complementarity is more challenging.

Kool *et al.* started with difluorotoluene base F, the simplest isosteric mimic of thymine (Figure 3-12a).[112, 113] Although F is meager in stabilizing the DNA duplex, it is a satisfactory counter base for adenine in replication with the Klenow fragment.[109] In respect of adenine, its non-hydrogen-bonding shape mimics Z and Q are better in extension efficiency when pairing with thymine or F.[114] Q is superior to Z because H3

facilitates the minor groove H-bonding interaction.[115] This indicates that the minor groove interaction of base pairs and the polymerase, besides shape complementarity, is also a major factor for extension beyond the artificial base pair.

Hirao designed a similar hydrophobic base pair Dss/Pn as shown in Figure 3-12b. Pn was derived from Pa, a base designed for pairing Kool's Q base, to prevent Pa/A pairing. On the other side, as Q can mispair with thymine, the methyl group of Q was replaced by Ds with a thienyl group,[116] or Dss with a double-thienyl group. Dss has an excitation centered at 385 nm and an emission at 456 nm so that it acts as fluorophore and Pn as a quencher.[117] When the basepair comes to PCR, the efficiency and fidelity are crucial issues because any chance of mispairing will accumulate in the PCR process. They achieved amplification of Dss/Pa as a molecular beacon in real-time PCR.[118] Undoubtedly, the unnatural fluorophore/quencher base pair will be a useful tool for sensing, imaging and diagnostic applications.

Since 1999, F. E Romesberg group has done tremendous work in the realm of the hydrophobic base pair (Figure 3-12c). Starting from a homo-base pair PICS/PICS,[119] they substituted the isocarbostyril moiety with various heteroatoms. Homo NICS, SICS, and SNICS pairs destabilize the duplex compared to homo ICS; SNICS/SNICS extended more efficient than ICS with Klenow fragment.[120] Benzenes that contain methoxy groups at different positions were subsequently studied. Although the methoxy group is a weak hydrogen bond acceptor, *ortho*-situated methoxy group forms hydrogen bonds with the polymerase in the minor groove and promotes the replication process.[121] A screening assay among a library of artificial nucleotides revealed the promising 5SICS/MMO2 pair. 5SICS/MMO2 replicates more efficient than the previous but is still slightly inferior to the canonical pairs.[122, 123] Replacing of the methyl group with fluorine or a fusion ring furnished 5FM and NaM.[124] Both pairs, 5SICS/5FM and 5SICS/NaM, replicate and transcribe efficiently with high fidelity.[125-128] These experiments indicated that all of the properties required of a functional unnatural base pair can be optimized within predominantly hydrophobic

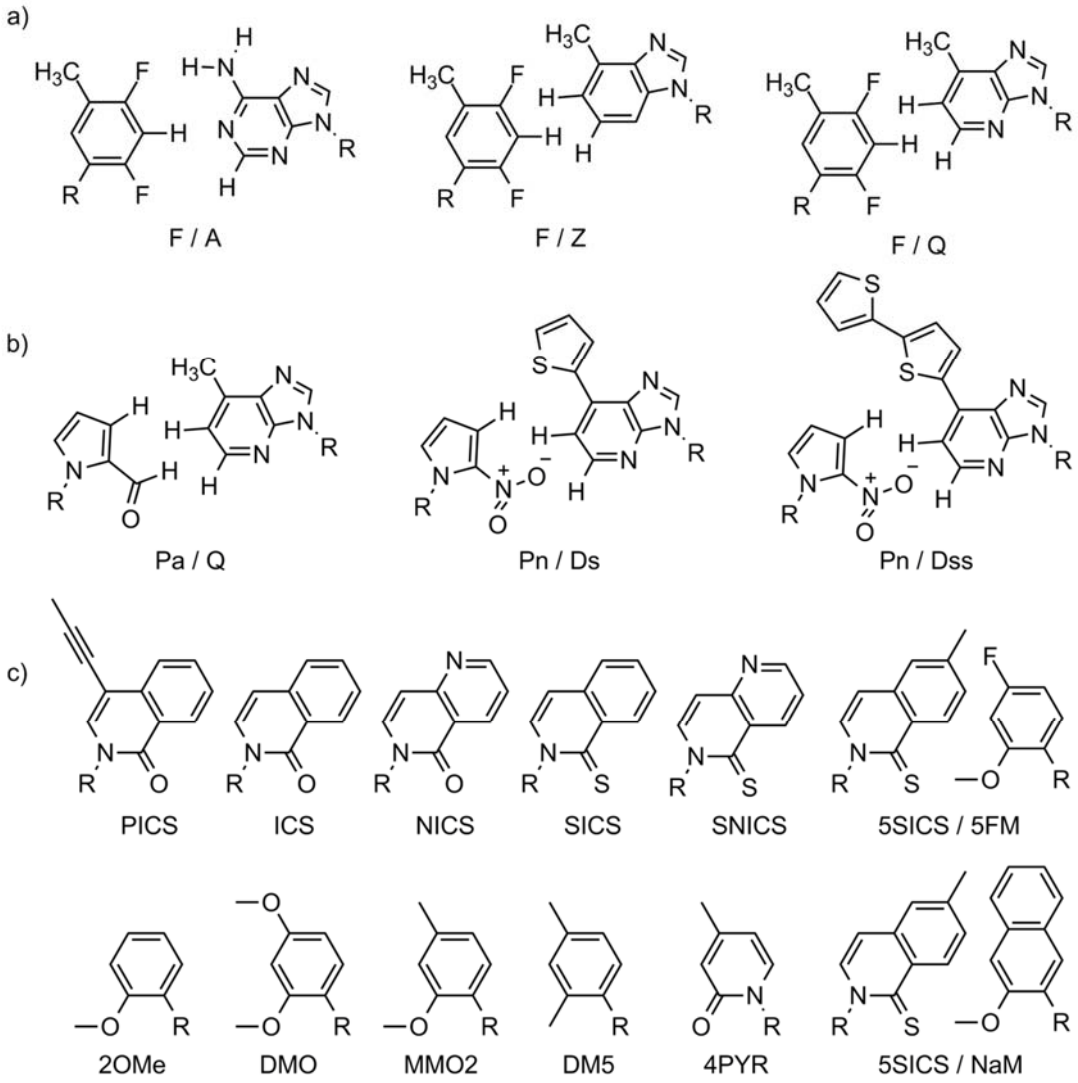


Figure 3-12 Examples of hydrophobic base pair mimics: a) F, Z, and Q from Kool *et al.*; b) Pa, Pn, Ds, and Dss from Hirao *et al.*; c) selected bases from Romesberg *et al.* R represents deoxyribose.

nucleobases that bear no homology to the natural nucleobases. In 2014, Romesberg and co-workers demonstrated that *E. coli* can stably maintain a plasmid containing a 5SICS/NaM pair with only the addition of an exogenously expressed algal nucleotide triphosphate transporter and the respective triphosphates.[129] The base pair can escape the excision in DNA repair pathways which made it the first expanded genetic alphabet propagated stably in the organism.

Recently, T. Carell and co-workers described a salen structure based base pair, as shown in Figure 3-13a, the first example of a base pair with a reversible covalent bond.[130] The covalent pattern is more a linker than a base pair. The T_o/S^{II} pair is

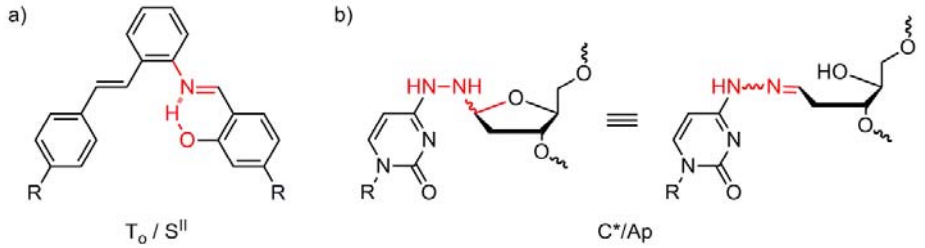


Figure 3-13 Examples of reversible covalent bond base pair mimics: a) T_o/S^{II} from Carell *et al.*; b) C* from Gates *et al.* R represents deoxyribose. The hydrogen bond is shown as hashed bond.

connected with a reversible imine bond, which stabilizes the duplex and facilitates replication with the DNA polymerase. A co-crystal structure with Bst Pol I showed that S^{II} slightly twisted from the T_o plane while the C1'-C1' distance was the same as in the canonical base pair. Shortly afterward, K. S. Gates described a simplified version, a C*/Ap crosslink containing a similar hydrazone linkage[131] based on their previous study (Figure 3-13b).[132] Both patterns should be beneficial in biochemistry and material science as the authors claimed.

3.1.2. Ligandoside: metal incorporated nucleoside

3.1.2.1. Metal incorporated canonical base pairs

Ligandosides are canonical or artificial nucleosides that can chelate metal ions inside the DNA duplex. In 1952, S. Katz reported a substantial decrease in the viscosity of natural DNA upon addition of $HgCl_2$.[133] The author explained the observation by an overall decrease size of DNA molecular. Two years later, C. A. Thomas clarified that the nucleobases can bind Hg^{2+} by UV spectroscopy.[134] Hence, Katz brought forward the model of Hg^{2+}-thymine (1:2) complex, as depicted in Figure 3-14, which was proved to be corrected.[135]

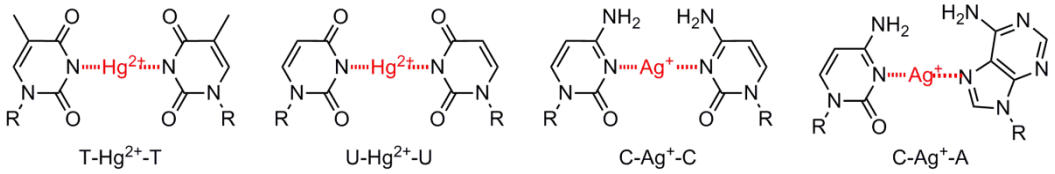


Figure 3-14 Examples of metal incorporated canonical base pairs. R represents pentose sugar. Bonds formed by coordination and bonds formed by proton displacement are shown as hashed bonds without further differentiation.

To form a stable metal-ion-mediated base pair, the nucleosides need to coordinate metal ions linearly in a 2:1 or 2:N ratio. In this respect, Ag^+ and Hg^{2+} are chosen because they are known to adopt linear coordination geometry but at the same time are flexible enough to accommodate more than two ligands.

The complexation is studied using the measuring the melting temperature of the DNA duplex. As the temperature rises, the double helix will get loosen and finally become two independent strands. The base with absorbance at around 260 nm will be no longer wrapped in a phosphate diester backbone. Therefore, the absorbance of the solution will increase. In general, the function of absorbance at 260 nm against the temperature is a typical S-shape curve for DNA-DNA duplexes. The maximum of the first derivative of this function corresponds to the temperature at which the duplex denatures at the maximum rate. This temperature is designated as melting temperature (T_M). The addition of metal ions leads to an increased melting temperature, which is the evidence for complexation. Nevertheless, comparison of T_M values across literature is difficult. Strand context, buffer system, temperature ramping, etc. affect the melting curves to a certain degree.

Apart from melting temperature measurements, ESI mass spectrometry, and NMR spectroscopy study provided further clues for complexation. Circular dichroism spectrum is helpful in studying the duplex conformation before and after the complexation. Among all, the authentic verification of base pair formation should be X-ray diffraction. However, difficulties in crystallization resulted only a few X-ray structures reported.[136, 137]

The past decade has seen a prosperous development of mismatched natural base pairs with metal coordination properties. A. Ono and co-workers applied the T-Hg^{2+}-T pair in a hairpin structure for selective detection of Hg^{2+} ions in aqueous solution.[138] Later they presented the biophysical data of the T-Hg^{2+}-T[139] and C-Ag^+-C[140] containing duplexes (Figure 3-14). Based on these results, the Ono group further studied the 5'-substituted homo-uracil pair.[141] 5-bromo-, fluoro-, and cyanide- uracils were

synthesized. Although all these modified uracils could chelate Ag^{+} and Hg^{2+}, melting profiles indicated that the duplex stabilizing ability was pH-dependent. The feature can be utilized for controlling metal arrangement by precisely placing uracil modifications along the strand. A systematic crystallographic study in the HIV-1 RNA dimerization initiation site revealed the formation of a C-Au^{3+}-G pair,[142] while most divalent cations as Mg^{2+} and Zn^{2+}, bind at Hoogsteen sites.

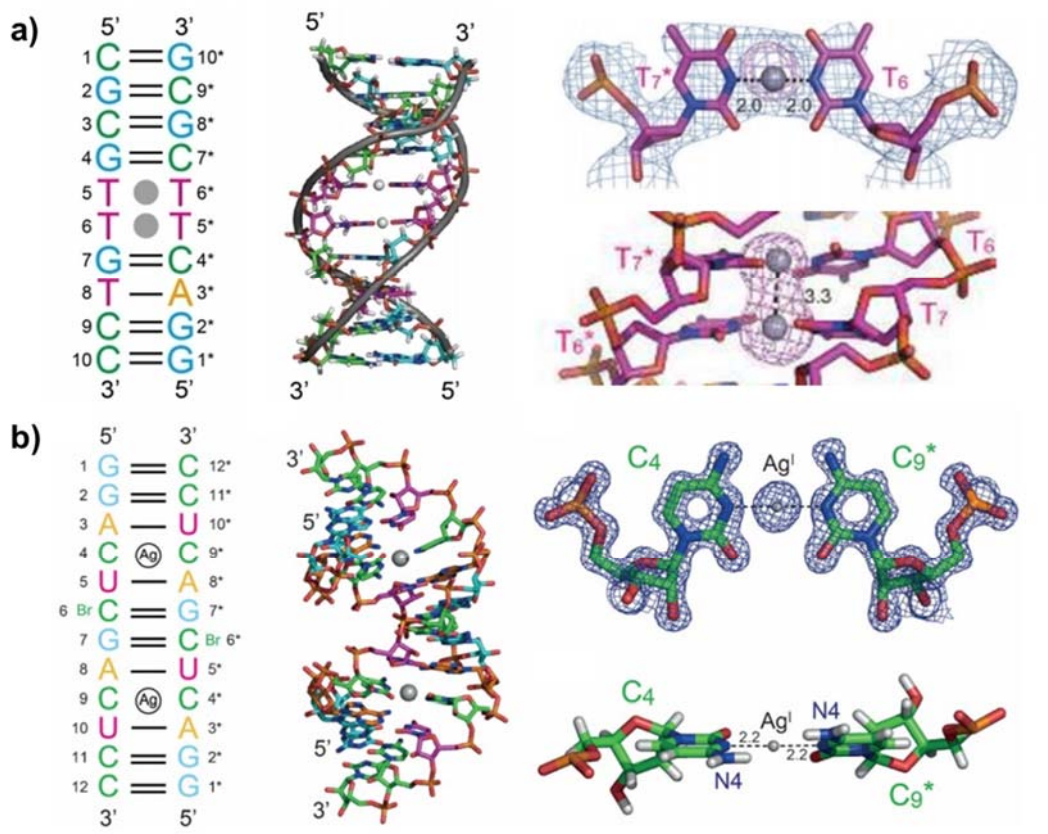


Figure 3-15 X-ray structure of a) T-Hg^{2+}-T and b) C-Ag^{+}-C inside DNA and RNA duplex. For each illustration: schematic representation of the complex; complete crystal structure; top review and side review of local omit map. Metal ions are shown as gray spheres. The figure is adapted from reference.[143, 144]

Recently, Ono *et al.* presented the X-ray crystal structure of T-Hg^{2+}-T in a DNA duplex and C-Ag^{+}-C pair in RNA duplexes (Figure 3-15),[143, 144] giving evidence that these pairs did not disturb the standard form of the duplex. In the absence of $Hg^{2+,}$ the DNA duplex was significantly distorted and adopts an unusual non-helical conformation while the RNA duplex was not distorted at all without Ag^{+} in mismatch C-C pair. The results provide the basis for the structure-based design of

metal-conjugating nucleic acid nanostructure.

Acknowledging that the T/T pair prefers Hg^{2+} to Ag^{+} and C/C mismatch bind Ag^{+} rather than Hg^{2+}, I. Willner *et al.* constructed a logic AND and OR gate based on metal chelating oligonucleotides.[145]

Metal ions not only stabilize mispairs in the DNA duplex but also assist DNA triplex formation. T. Ihara demonstrated a $CGCAg^{+}$ complex stabilized parallel-motif triplexes at neutral pH,[146] an alternative method to stabilize triplexes parallel to using auxiliary molecules and modified nucleosides.

To check whether DNA polymerase can recognize the mismatched pairs and elongate the primer, Urata *et al.* involved T-Hg^{2+}-T[147] and C-Ag^{+}-A[148] mismatches in primer extension experiments. Meanwhile, *in vitro* transcription of mercury-intermediated U/U pairs (U-Hg^{2+}-U) with T7 RNA polymerase was achieved by Müller *et al.*[149] These works are promising towards the construction of metal ion-triggered replication and transcription system as well as enzymatic preparation of metal-containing DNA nanodevices. Conceivably, in the prebiotic era, these base metal coordination bonds replaced hydrogen bridges, and the complexes acted as catalysts for initial reactions.

The binding capability of the canonical nucleobases to metal ions is limited to certain combinations. Watson-Crick base pairing may interfere with metal binding in sequence-dependent situations. Therefore, an orthogonal metal base pair system will have a larger arena in the synthetic biology and nanotechnology realm.

3.1.2.2. Metal incorporated artificial base pairs

Artificial DNA base pairing through metal complexation has attracted great interest in the development of functionalized biopolymers and in the expansion of the genetic alphabet. Since the pioneering work published in 1999 by M. Shionoya *et al.*, numerous novel structures came out. Figure 3-16 summarizes the published artificial base metal complexes.

The first example, *o*-phenylenediamine ligandoside (Figure 3-16a), was proposed by K. Tanaka and M. Shionoya.[150] The complex was believed to be square-planar, but evidence from NMR and ESI mass spectrometry is only limited to the nucleoside level. Further derivate nucleosides, based on the *o*-aminophenol[151] and catechol,[152, 153] were published without modified oligonucleotide properties. Lack of proper protection strategy in solid-phase synthesis impeded incorporation into oligonucleotides.

Apart from the *ortho*-disubstituted phenyl nucleoside, Shionoya reported the homo-hydroxypyridone base pair H/H, as well as homo-pyrimidine nucleotide Py/Py with incorporated Cu^{2+} inside the DNA duplex.[154, 155] Surprisingly, three Py ligandosides formed a base triplet in a triple-strand DNA, making the complex more stable than a duplex with Py-Cu^{2+}-Py. With the H/H pair, arranging five Cu^{2+} ions inside a DNA duplex was feasible, confirmed by UV, CD titration, electron paramagnetic resonance (EPR) spectroscopy and ESI mass spectrometry.[156] Based on the EPR results, the distance between the copper centers were estimated to be slightly larger than the base-to-base distance in canonical B-type DNA. To our best knowledge, it was the first biological approach of arranging metal ions in solution in a predictable and discrete manner. Later, Shionoya and co-workers described that three strands containing four hydroxypyridone bases coordinated with a Fe^{3+} ion formed a triplex.[157] As DNA arrays of the octahedral transition metal complex were seldom investigated; this approach provided an excellent way to construct rows of metal centers with magnetic and conductive properties.

Soon afterward, P. G. Schultz *et al.* synthesized a hetro- tridentate ligandoside pair Dipic/Py (Figure 3-16b).[158] Upon Cu^{2+} addition, the complex was similar to an A/T pair in the duplex while other metal ions destabilized the system. The crystal structure using X-ray showed the Dipic/Py pair is well-behaved as a Z-DNA like conformation, and C1'-C1' distance here is shorter than in the canonical base pair.[136] Later, derivate Dipim/Py coordinated with Cu^{2+} [159] and SPy/SPy with Ag^{+} [160] displayed a greater

stabilization effect. These 3+1 unsymmetrical base pairs were novel systems orthogonal to Watson-Crick base pairs. Although the enzymatic incorporation has not been realized yet, the unsymmetrical pattern enlightens the future design of ligandosides.

Similar to Shionoya's monodentate Py/Py pair, in 2007, J. Müller described a 1,2,4-trizaole based homo-monodentate ligandoside (Figure 3-16c).[161] A hairpin structure with ligandosides located in the loop was formed when three triazole bases were placed continuously. With properly labeled fluorescence and quencher groups, the oligonucleotide architecture turned out to be a promising metal ion sensor. Later, Müller *et al.* innovatively combined the Dipic base from the Schultz group with a glycol backbone and imidazole with a deoxyribose (Im) backbone.[162] Although Ag^+ increases the stability of the modified duplex the most compared to the other metal ions, it is still less stable than a canonical duplex. Another ligandoside based on the simplified backbone C_3HQ/C_3HQ was proposed by E. Meggers, stabilized the duplex similar to canonical base pairs upon Cu^{2+} complexation.[163]

The homo Py base pair and the 3+1 unsymmetrical pattern enlightened C. Switzer to develop a series of triaza-heterocyclic and pyridine ligandoside pairs (Figure 3-16d).[164, 165] Notably, Pu-2,6-Py/3-Py and Pu-6-bipy/4-Py with Ag^+ are as stable as G/C and A/T pairs in the DNA context. While hetro 3+1 aza-heterocyclic chelate Ag^+ preferably, homo 2+2 aza-heterocyclic ligandoside complexed with Ni^{2+} were described by Switzer as well (Figure 3-16e).[166, 167] Homo PurP and PyrP pairs are even more stable than canonical base pairs when coordinated with Ni^{2+}. The pyridine pairs adopted a square-planar geometry and the distance in between approached those in G/C and A/T of natural B-DNA helices according to Gaussian simulation, which was different from earlier reported homo-bipyridine Bipy and Bp pairs that tended to stack with each other along the strand simulated by Insight II.[168, 169]

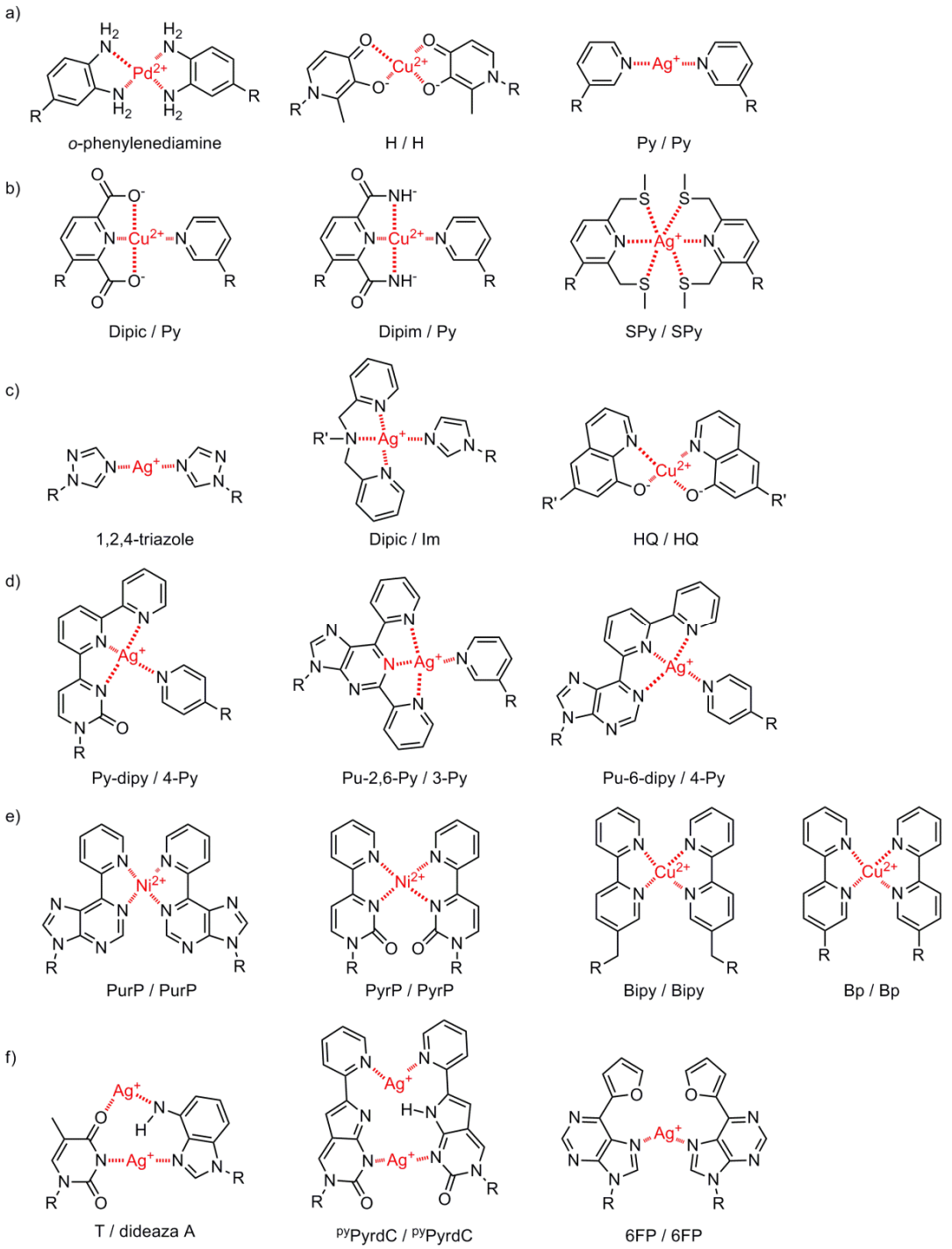


Figure 3-16 Examples of metal incorporated artificial base pairs: a) homo- bidentate and monodentate ligandosides from Shionoya *et al.*; b) hetro 3+1 ligandosides and derivatives from Schultz *et al.*; c) triazole liganoside and acyclic backbone derivates from Müller *et al.* and Meggers *et al.*; d) hetro poly-pyridine liganosides and e) homo poly-pyridine liganosides from Switzer *et al.*; f) multi-Ag^+ ligandosides from Müller *et al.* R represents pentose sugar. R' represents the acyclic backbone. Bonds formed by coordination and bonds formed by proton displacement are shown as hashed bonds without further differentiation.

Incorporation of one metal ion is difficult enough; complexation of two ions in one base pair remains more challenging. The Müller group reported Ag^+-mediated Hoogsteen-type base pairs comprising 1,3-dideaza-2'-deoxyadenosine and thymidine (Figure 3-16f).[170] Because the N3 nitrogen atom of the adenine was substituted by a methine moiety, the basicity of the exocyclic amino group increased so that two Ag^+ could be incorporated. The complexation was studied by UV, CD titration, dynamic light scattering, and mass spectrometry. Another homo-PyrdC pairs from F. Seela complexed up to three silver ions.[171] These novel complexations allow the construction of DNA structures with heavy metal ion loading. Most surprisingly, in the latest example, Müller and co-workers reported a highly stable Ag^+-mediated 6FP/6FP base pair in parallel stranded DNA.[172]

Regretfully, most of the reported research concerning the non-canonical ligandosides only provides biophysical characterization. Until now there are few reports regarding their structural information and their potential application in other fields of study.

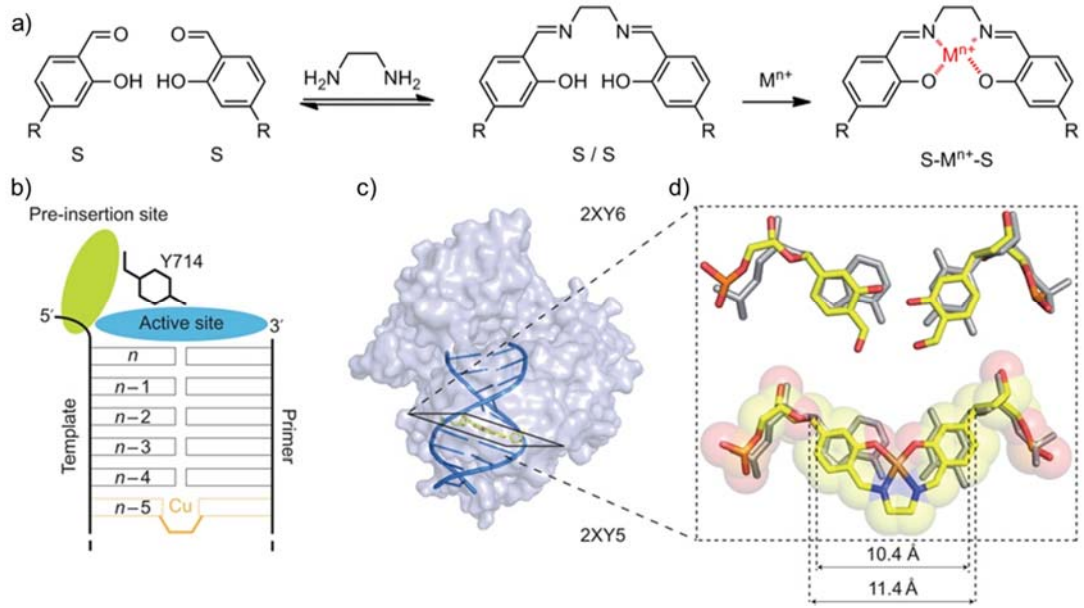


Figure 3-17 Salen ligandoside S/S: a) Reversible formation of the complex in the duplex. $Mn^+ = Cu^{2+}$, Mn^{3+}, etc. Bonds formed by coordination and bonds formed by proton displacement are shown as hashed bonds without further differentiation. b) Schematic representation of the complex; c) complete crystal structure; d) two salicylic aldehyde nucleosides are shown facing each other, and the fully assembled salen complex is at the n-5 position in the DNA duplex with the Cu^{2+} ion in a square planar geometry, overlaid with A/T base pair. The figure is adapted from reference.[137]

In 2005, T. Carell *et al.* reported a homo bridged ligandoside S/S based on the *N*,*N*'-bis(salicylidene) ethylenediamine (salen) ligand (Figure 3-17a).[173, 174] The metal-salen complex showed planar geometry, which matched well with Watson-Crick base pair. Due to the ethylenediamine linker lying inside the duplex, the modified duplex became extremely stable upon metal ion complexation. In the absence of ethylenediamine, the S/S pair with Cu^{2+} stabilized the duplex to the same degree as H/H pair (Figure 3-17a) from Shionoya in the same duplex context. The salen base pair was versatile in binding a variety of metal ions, e.g. Cu^{2+}, Mn^{2+}, Ni^{2+}, Fe^{2+} and VO^{2+}. As Mn^{2+} coordinated by a salen ligand is known to be oxidized to Mn^{3+} under aerobic conditions, each resulting Mn^{3+}-mediated salen base pair carries one positive charge, compensated by the negative charge on the backbone. The continually stacked salen base pairs can arrange ten Cu^{2+} or Mn^{3+} ions in a row, confirmed by UV titration and ESI mass spectrometry.[175] Mixing five S-Cu^{2+}-S with five T-Hg^{2+}-T pairs along the strand, Carell and co-workers achieved a mixed metal ion stack inside the DNA duplex.[176] Hence, Cu^{2+} and Hg^{2+} ions are arranged in a programmable fashion at the atomic level.

Carell group later reported that salen ligandoside could be replicated using polymerase Bst Pol I and amplified in PCR cycles with KOD DNA polymerase.[137] Co-crystal structures of the salen ligandoside inserted duplex inside Bst Pol I showed that the reversible imine bond sat directly inside the polymerase, enabling the efficient copy of the inorganic crosslink (Figure 3-17b-d).

Since the first structure of metal base pair complexation was hypothesized in 1952, investigation and development of this concept have caught the attention from the chemical and biological communities. The main benefit of the DNA-metal complexation is the enhancement of thermal stability. Combined with the canonical sequence based design and construction, more complex DNA nano-architectures with metal ions spatially arranged can be foreseen. The metal-containing double helices and nanostructures are expected to display intriguing chemical and physical properties,

e.g. magnetic and electric properties, metal-sequence based information storage ability, metal-ligandoside chiral complex catalytic activity.

Based on the structure of nucleotides and the DNA double helix, a collection of artificial nucleotides is reviewed above. Generally, modifications of the phosphate diester bond and the sugar moiety are mainly designed to interfere with the nucleic acid metabolism and sequence-specific labeling due to their specificity and strong nucleic acid-binding capacity; reconstructions on base pair patterns are aimed at expansion of genetic alphabets and mechanistic insights of the central dogma.[177, 178] Base pair with metal coordination bonds may have more applications in nanotechnology and material science. To sum up, originated from the evolved canonical, the collection of artificial nucleotides, the "XNA" world, displays the desire of man to manipulate the nature. The "XNA" world is a powerful toolkit in the broad realm of chemistry and biology.

In Chapter 4, a metal base pair based on pyrazole structure is introduced. The organic synthesis and biophysical properties is discussed as well as the efforts to apply the complexes as a chiral catalyst and to involve the ligandoside in DNA replication.

3.2. Epigenetic cytosine modifications and sequencing

3.2.1. Epigenetic bases: discovery and function

"Epigenetic" is a compound word of "epi-" and "genetic". "Epi" means an information layer above the pure genetic information, Watson-Crick base pairing, which is inheritable or transmittable to the next cell or generation. Methylation of cytosine in the mammalian genome, as a key epigenetic modification, changes the functional state of regulatory regions while do not alter the Watson-Crick base pairing of cytosine.

3.2.1.1. 5-Methylcytosine

Discovery. 5-Methylcytosine (mdC) was first reported as a constituent of nucleic acids in 1925.[179] Study of the methylation mechanism attracted more and more interest in the 1970s.[180]

In mammals, cytosine methylation patterns are established by the DNA methyltransferase 3 (DNMT3) family of *de novo* methyltransferases and maintained by the maintenance methyltransferase DNMT1.[181, 182] DNMT catalyze the covalent linkage of the methyl group from the methyl donor *S*-adenosyl-methionine (SAM) to the 5' position of the cytosine-pyrimidine ring.[183, 184] In plants, *de novo* methylation is catalyzed by domains rearranged methyltransferase 2 (DRM2), a homolog of the DNMT3, and maintained by different pathways.[185]

Abundance. mdC is found almost exclusively in CpG dinucleotide.[186, 187] Methylation of non-CpG context is common in plants, but rare in most mammalian cell types except pluripotent embryonic stem cells (ES cells),[188] oocytes,[189] and mature neurons.[190, 191] In total, mdC contributes 3.6~7.1% of the bases in higher plants and 0.7~3.8% in the vertebrate.[192] Global DNA hypomethylation and locus-specific DNA hypermethylation have been identified as key features in various cancers.[193]

Function. In mammals, mdC participate in various biological processes, including genomic imprinting, chromosome X-inactivation, and silencing of repetitive

elements.[184] In gene promoter and enhancer regions, CpG methylation is associated with repressed transcription.[194] Proteins with methyl-binding domains, e.g. MeCP2, bind to mdC and recruit repressor complexes containing histone deacetylases (HDAC), allowing the histones to wrap the DNA more tightly. On the other side, gene body methylation is usually correlated with gene expression.[195-197]

Fate. Global erasure of mdC is known to occur in development. mdC can be lost through deficient maintenance or replication-dependent dilution, namely passive demethylation.[198] However, passive demethylation cannot explain all cellular demethylation events during development or differentiation, many of which seem to occur in the absence of DNA replication.[199, 200]

In plants, active erasure of mdC is mediated by the DNA glycosylases DEMETER (DME)/repressor of silencing (ROS1) and the base excision repair (BER) machinery.[201] In contrast, mammalian orthologs of DME/ROS1 enzymes have not been reported. Several mechanisms of active DNA demethylation have been proposed over the last several decades, including a nucleotides excision repair (NER) pathway,[202] a pathway in which RNA is a key component,[203] and a pathway involving a demethylase.[204]

Although activation-induced deaminase (AID)/APOBEC displays lower activity on mdC than C,[205] studies support its potential role in mdC deamination.[206, 207] Moreover, DNMTs have been identified to possess mdC deaminase activity *in vitro* in the absence of SAM.[208]

3.2.1.2. 5-Hydoxymethylcytosine

Discovery. 5-Hydroxymethylcytosine (hmdC) was first identified in bacteriophage nucleic acids in 1952 and later in animal cells.[209, 210] The occurrence of hmdC in animal cells was debated until 2009 when two independent laboratories found that 5-hydroxymethylcytosine (hmdC) indeed existed in the mammalian genome.[211, 212] Their experiments convinced that hmdC was the oxidation product of mdC, mediated

by the ten-eleven translocation (TET) enzyme.

Abundance. N. Heintz *et al.* found that hmdC constitutes 0.6% of all bases in cerebellar Purkinje neurons, the highest level of any mammalian cell type.[211] A. Rao *et al.* reported in mES cells hmdC corresponded to 4% of all cytosine species in MspI cleavage sites (C^CGG) in ES cells, and 0.032% of all bases.[212] Bisulfite- and affinity-based sequencing methods (see Chapter 3.2.3.3 and 3.2.3.4) proved that in mouse and human ES cells, hmdC was enriched at intragenic, low-CpG-density CpG islands (CGIs),[213, 214] promoters where genes expressed at medium-to-low levels,[215,216] protein-DNA interaction sites,[217] and histone modifications associated enhancers.[218] Concerning the tissues, the percentage of hmdC in the brain is as high as 0.3% to 0.6%, increasing with age of the mice.[219] In liver, bladder and heart tissues, 0.15-0.17% relative to dG are observed while in liver, spleen and testes the level is 0.03-0.06% of dG.[220] Abundance of hmdC in the cancerous colorectal tissues was significantly reduced (0.02–0.06%) compared to that in normal colorectal tissues.[221]

Function. Although the biological role of mdC has been well studied, the function of hmdC, as well as further oxidized bases, remains elusive. Similar to mdC, hmdC is linked to numerous biological processes including embryonic stem cell maintenance and differentiation, zygote development and cancer.[222] The hmdC located in the gene body appears to correlative with increased expression and transcription.[223, 224]

Fate. Apart from further oxidation reaction, Song *et al.* claimed that hmdC was more sensitive than mdC to activation-induced deaminase. 5-Hydroxymethyluracil (5hmU), the deamination product of hmdC, activates BER pathway-mediated demethylation.[225] *In vitro* biochemical studies suggest that DNMT3A/3B can remove the hydroxymethyl group of hmdC to generate unmodified cytosine under oxidizing conditions in the absence of SAM.[226] Given that SAM is present at relatively high levels in all cell types, the physiological relevance of this reaction remains unclear.

3.2.1.3. 5-formylcytosine and 5-carboxycytosine

Discovery. 5-Formylcytosine (fdC) and 5-carboxycytosine (cadC) were discovered in 2011 by Y. Zhang *et al.*,[227] G. Xu *et al.*[228] and T. Carell *et al.*[229] fdC and cadC are consecutively oxidation products of hmdC, in which Fe(II)- and α-KG-dependent TET family of dioxygenases are employed as catalysts.

Abundance. Quantification by mass spectrometry of DNA digested into nucleosides showed that the fdC level in genomic DNA of ES cells was around 0.06-0.6% of hmdC. cadC was found 10-fold less abundant than fdC, corresponding to 0.01% of hmdC.[229] In mammalian brain tissues, fdC was found to be 2-3 and cadC 3-4 orders of magnitude lower than hmdC.[230]

Based on a PvuRts1I-assisted mapping experiment, fdC in non-CpG context was found lower in abundance and more dynamic than those in CpG context. fdC tend to distribute at regulatory protein-DNA binding sites, active and poised enhancers,[231] and actively transcribed gene bodies.[232]

Function. The mechanism whether fdC and cadC are merely intermediates in the active demethylation pathway, or they interact with genomic readers and carry out their functions is still not clear. The formyl group of fdC and carboxyl group of cadC provide unique chemical anchors for the protein recognition especially the former is a relatively reactive group that can react with various cellular components.

In wild-type mES cells, 90% of fdC-marked regions reside in hmdC-enriched regions, which correspond to ~30% of the hmdC regions, suggesting fdC and cadC define only a part of the hydroxymethylome.[233]

S. Balasubramanian recently speculated that hmdC and fdC were stable DNA modifications rather than demethylation intermediates. They fed the mouse with $^{13}CD_3$-SAM to generate labeled mdC and observed that globally labeled hmdC levels did not change during the cell cycle. The labeled hmdC remained present in DNA for more than five days, and the labeling ratios only decreased in cell proliferation. This indicates that hmdC is a stable modification. At the same time, fdC labeling ratio was

always lower than that of mdC or hmdC, implying fdC was also a stable DNA epigenetic base.[234]

Fate. Although thymine-DNA glycosylase (TDG) has no significant mdC or hmdC excision activity *in vitro*, it excises fdC and cadC that properly pair to G in duplex DNA.[235] Therefore, it is believed that fdC and cadC could be recognized and excised by TDG to give abasic sites *in vivo* that are subsequently converted to cytosines through base excision repair (BER).[228, 236] Knockdown of the gene encoding TDG led to elevated fdC and cadC levels in ES cells[228] while overexpression of TDG in HEK293 cell line led to fdC and cadC depletion.[237] These discoveries confirmed that TDG is able to excise fdC and cadC *in vivo*, completing the pathway for active DNA demethylation (Figure 3-18).

Apart from the TDG/BER pathway, direct decarboxylation of cadC by a putative decarboxylase may revert the oxidized bases to unmodified cytosines. However, the responsible enzyme waits to be identified.[238, 239]

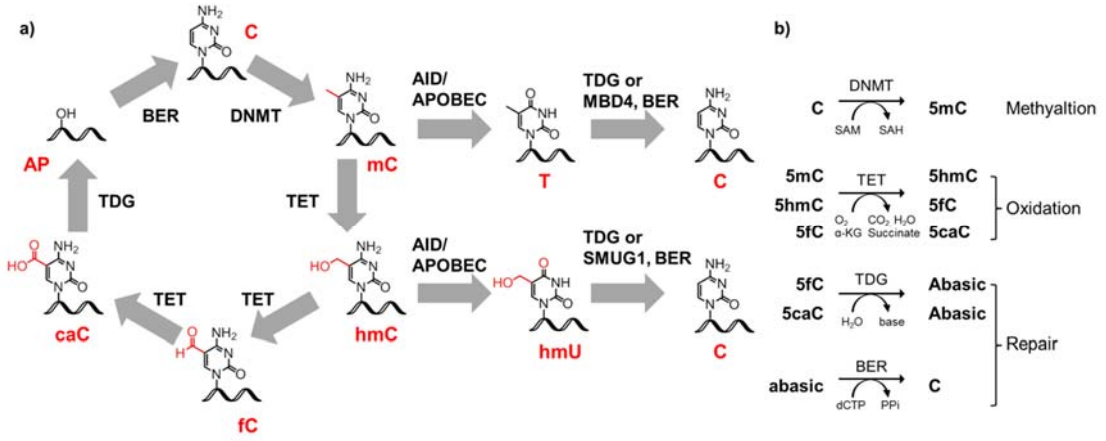


Figure 3-18 Pathway for dynamic modification of cytosine: a) biochemically validated pathway for cytosine modifications; b) individual reactions in the pathway are shown with all reactants involved. SMUG 1, strand-selective monofunctional uracil-DNA glycosylase 1; SAM, *S*-adenosylmethionine; SAH, *S*-adenosylhomocysteine; α-KG, α-ketoglutarate, PPi, pyrophosphate.

In short, our growing appreciation for epigenetic cytosine bases implies that these oxidized derivatives are more than transitional states, and DNA methylation-demethylation cycles are dynamic and intricately regulated.

3.2.2. Genome sequencing: methods and significance

For most of the 20th century, the central problem of genetics was to create genomic maps of entire chromosomes. These maps were crucial for researchers to understand gene structure, function, and evolution.

As early as 1983, the first genetic disease, Huntington's Disease, was mapped.[240] In 1984, US government proposed and funded the Human Genome Project (HGP). HGP was an international project for determining the sequence of chemical base pairs which make up human DNA, and identifying and mapping all of the genes of the human physically and functionally. HGP was undertaken in 1990 and declared complete in 2003.

In short, the target genome is fragmented to construct a library. The fragment is cloned into a large-fragment cloning vector, e.g. bacterial artificial chromosome (BAC) vectors. The genomic DNA fragments represented in the library are then organized into a physical map; individual clones are selected and sequenced.[241] Finally, the clone sequences are assembled to reconstruct the sequence of the genome.

To sequence short DNA strands, A. Maxam and W. Gilbert developed a sequencing method based on nucleobase-specific modification followed by backbone cleavage at sites adjacent to the modified nucleotides in the mid-1970s.[242] Guanine is methylated at N7 position by dimethyl sulfate;[243] adenine and guanine are depurinated using formic acid after methylation; cytosine and thymine are depyrimidinated using hydrazine;[244] 2M NaCl suppresses the reaction of thymines with hydrazine thus only cytosines are hydrolysed. After the cleavage in hot piperidine, the fragment is analyzed by PAGE and visualized by autoradiography. The replication ends in four parallel reactions indicate the bases at responding positions.

At the same time, Sanger sequencing, a chain termination method, was invented by F. Sanger and colleagues.[245,246] Sanger sequencing is based on the selective incorporation of chain-terminating dideoxynucleotides (ddNTP) during *in vitro* DNA replication. PAGE assay shows the replication termination in four parallel reactions,

each employing a different terminator nucleotide. Later, in 1985, dye-labeled ddNTPs were introduced, eliminating the need for radioactive labeling.[247,248] All of the four dideoxynucleotides were labeled with fluorescent dyes and emitted light at different wavelengths. Because of its extreme expediency and high speed, dye-terminator sequencing became the prevalent method of automated sequencing.

However, chain termination sequencing is limited to the single strands of 100 to 1000 base pairs. It remains challenging to separate longer strands and maintain one base pair resolution. Longer strands have to be fragmented to small ones first and reassembled after sequencing. Two methods are available here: primer walking sequencing,[249] using continuous priming of the sequencing reaction; and shotgun sequencing,[250, 251] using randomly cloned strands.

In the primer walking sequencing, a primer that matches the beginning of the region of interest is used to synthesize a short DNA strand adjacent to the unknown sequence. Afer the new short DNA strand is sequenced using Sanger sequencing, the end of this strand is used as a new primer for the next part of the long strand sequencing.

In the shotgun method, multiple overlapping reads for the target are obtained by performing several rounds of shotgun fragmentation and sequencing. Assembling the overlapped fragments into a continuous sequence can be fulfilled by software.

Shotgun sequencing generates high redundancy and consumes much computer time, but sequencing itself takes less time and minimal cost. Although the primer walking is superior in redundancy, it is extravagant to synthesize a great amount of primers. As time goes on, programs for sequence assembly become more sophisticated and computing power becomes cheaper, shotgun sequencing became the prevalent method for large-scale sequencing.

The International Human Genome Sequencing Consortium published the first draft of the human genome in February 2001 with the sequence of the entire genome's 3×10^9 base pairs.[252] In 2003, researchers completed a high-quality sequence of the entire human genome. Apart from the human genome, HGP finished sequencing other

organisms, including the mouse genome.[253]

Genome sequencing is not an end in itself. For us humans, it improves the understanding of inheritance and individuality. Genome sequencing has been critical for the identification of the genetic risk factors associated with complex human diseases and continues to have an emerging role in therapeutics and personalized medicine. For a larger picture, it paves the way towards a comprehensive understanding of all the biological functions encoded in the genome and their evolution. By studying the similarities and differences between human genes and those of other organisms, we discover the functions of particular genes and identify the genes that are vital for life.

3.2.3. Genome sequencing of epigenetic bases

By sequencing the Watson-Crick bases, we answered the question: What are these bases? By sequencing the epigenetic bases, we try to answer: how much are precent of these bases? Where are they located in the genome?

The first task is to quantify the level of mdC, hmdC, fdC and cadC at the genomic level, i.e. global quantification. The next question need to be reached by sequencing with single base resolution, i.e. genome mapping, to learn about the genomic sites that contain mdC, hmdC, fdC and cadC together with the corresponding probability.

3.2.3.1. Maxam-Gilbert sequencing based mapping methods for mdC

Before we encountered other oxidized bases, only mdC has been of interest for genome sequencing for a long time.

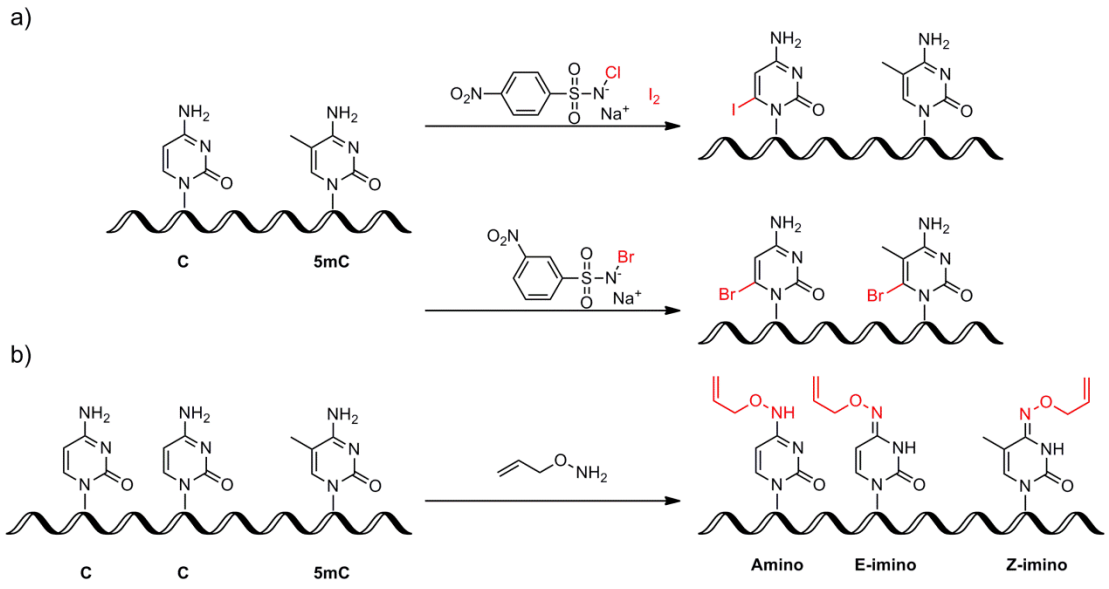


Figure 3-19 Chemical modification methods to discriminate mdC and cytosine: a) derivative Maxam-Gilbert sequencing from Zhou *et al.*;[254] b) pre-transformation method from Carell *et al.*.[255]

Although the Maxam-Gilbert method is no longer used for conventional sequencing of the genome, it found its application in mdC sequencing. The reaction of mdC with N_2H_4 is inefficient so that mdC does not introduce a strand cleavage. Compared with the recognition of G on the counter strand, mdC positions can be discriminated among C. A modified Maxam-Gilbert method reported by Zhou *et al.*[254] described that *N*-sodio-*N*-chloro-*m*-nitrobenzenesulfonamide with iodine preferred to react with cytosine rather than mdC and caused cleavage only at cytosine sites. Later it was known that *N*-sodio-*N*-bromo-*m*-nitrobenzenesulfonamide reacts with both cytosine and mdC but not with glycosylated hmdC thus resulting in more cleavage sites. In this way, mdC could be distinguished from cytosine and hmdC (Figure 3-19a).

Alternatively, cleavage at uracil sites by uracil DNA glycosylase (UDG) after bisulfite treatment of cytosine,[256] and selective cleavage at fdC and cadC sites by hot alkali,[257] were used but this method was only applied to synthetic oligonucleotides. Given the fact that Maxam-Gilbert sequencing is complicated, time-consuming and required much input material, it may not be suitable for epigenetic studies at the genomic level. M. Münzel *et al.* described *O*-allyhydroxylamine method to discriminate dC and mdC (Figure 3-19b).[255] Cytosine stayed as its amino isomer and was read as dC.

Alternativly, it could be shifted with the E-imino isomer which is read as T when it was reacted with hydroxylamine. mdC in contract was converted to the Z-imino isomer and blocked the polymerase. This method is unable to distinguish mdC and hmdC because both bring larger steric strain and stop the polymerase.

3.2.3.2. Global genome quantification

Tahilinani *et al.* discovered hmdC using radioactively labeled thin-layer chromatography (TLC).[212] TLC allows detection of hmdC within particular dinucleotide contexts. A similar method, called nearest neighbor analysis, allows global quantification of hmdC in specific sequence contexts as well.[211] These methods are useful techniques because of their sequence specificity detection; on the other hand, the inaccuracy is the drawback.

M. Münzel *et al.* synthesized isotope labeled mdC and hmdC as intern standards and quantified their levels in various tissues from mice of different age groups using liquid chromatography–mass spectrometry (LC-MS).[219] They concluded that hmdC level varied while mdC was stable among these tissues. Since then the LC-MS method constitutes the gold standard for the global quantification of epigenetic bases.

Given the fact that hmdC reacts with sodium bisulfite to yield cytosine 5-methylenesulphonate, Rao *et al.* developed an antibody against the mdC derivative,[223] which allows hmdC global quantification and Next Generation Sequencing (NGS) by dot-blot or enzyme-linked immunosorbent assay (ELISA). Although the antibody was claimed to be more sensitive and less density-dependent than antibodies against mdC,[258] it still bore the serious defect of density dependence and is less accurate compared to LC-MS. Relative mapping of mdC derivatives within samples can be achieved by methylated DNA immunoprecipitation (**MeDIP-Seq**) and chromatin immunoprecipitation (**ChIP-Seq**), but they are impotent in absolute quantification.

Glucosylation can be applied to assist global and local quantification. hmdC is

glucosylated with β-glucosyltransferase (β-GT). Quantification of genomic hmdC levels is performed by radioactive UDP-glucose labeling with ß-GT and subsequent scintillation counting.[259] Glucosylation also alters the activity of certain restriction enzymes. Therefore local hmdC levels can be measured by methylation-sensitive restriction followed by rtPCR. The method itself is convenient, except it is a restriction to the enzymatic cleavage site.[260, 261]

Concisely, TLC-, LC-MS-, antibody-, and enzyme- based techniques have been developed to determine the total amount of epigenetic bases in the genome. Although they vary from quite easy to very sophisticated, these methods provide comparable information.

3.2.3.3. Affinity enrichment based mapping methods

Affinity-enrichment based mapping methods profile the regions that contain epigenetic bases. The bases of interest are first attached to a biotin tag. After enrichment, the fragments of the genome are ligated to forward and backward strands for PCR amplification followed by NGS. In this way, the question "where are these bases?" can be solved.

To connect hmdC to a biotin, He and co-workers transferred azide glucose to the hydroxyl group of hmdC using β-glucosyltransferase and UDP-6-N_3-Glu. After copper-free click chemistry, hmdC was labeled with a biotin for enrichment (Figure 3-20a).[262] Similarly, Rao *et al.* attached the glucose to hmdC,[223] whose hydroxyl groups were oxidized to aldehydes with sodium periodate and connected to biotin linkers (Figure 3-20b).

The genome can be first cleaved with hmdC-dependent restriction endonucleases, i.e. PvuRts 1I. After hmdC connected to biotin-labeled glucose, amplicons containing hmdC are enriched and sequenced. Because of the sensitivity of the endonuclease, gDNA fragments are shorter and representative, and sequencing is more economical.[231]

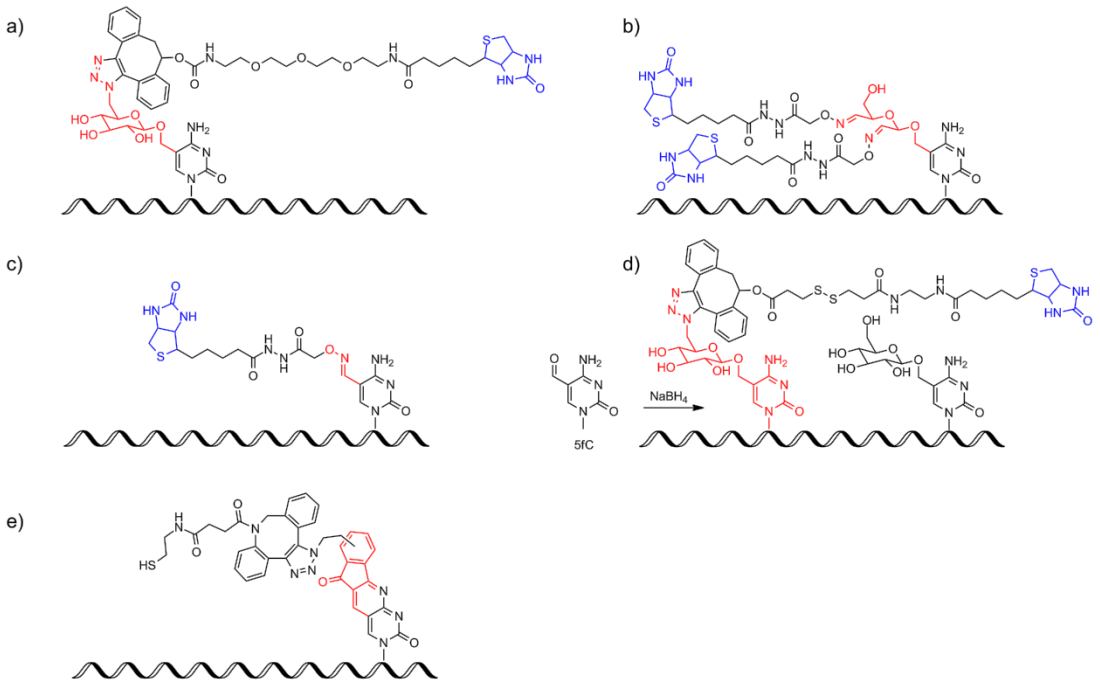


Figure 3-20 Pull-down enrichment sequencing methods for a), b) hmdC and c), d), e) fdC. The transformation moieties are highlighted in red, biotin highlighted in blue.

Balasubramanian took advantage of an aldehyde reactive probe (ARP), which was first applied by Pfaffeneder et al.,[229] to pull-down the fdC containing regions for genome-wide fdC mapping (Figure 3-20c).[263] Later, He *et al.* reported a similar fdC pulldown method (Figure 3-20d).[233] The genome was first treated with glucose and β-glucosyltransferase to block the hmdC. Then fdC is reduced to hmdC,[264] and labeled with the an azide-modified glucose. A biotin linker was clicked on the newly generate hmdC, after pull-down and sequencing fdC enriched regions were profiled. According to the author, this method reduced the nonspecific DNA capture compared to the former.

fdC can be converted to quinoline derivatives via Friedländer synthesis (Figure 3-20e). After biotin labeling, enrichment, disulfide bond cleavage, fdC is read as T during PCR amplification. Next Generation Sequencing maps fdC genomic-wide at single-base resolution.[265] The Friedländer conversion can be compared as the bisulfite transformation, but more efficient and precise.

In pull-down enrichment sequencing strategies, several disadvantages have to be considered. First, the base in question may not be labeled or converted completely;

the labeling can be unspecific. Some regions of the genome are more reactive, and the biotin tag may react with other components in the genome or reaction system as well. Besides, 8-oxoG is a good biotin mimic and would bind to streptavidin as well, making the methods more doubtful. Second, the enrichment efficiency should be questioned and checked. Third, the number of reads should be large enough to reach sufficient sequencing depth for high accuracy.

3.2.3.4. Bisulfite sequencing based mapping methods

Bisulfite sequencing (**BS-Seq**) is a sequencing method to determine the methylation pattern using bisulfite-mediated deamination (Figure 3-21). Bisulfite deamination reaction was first reported in 1970.[266] The bisulfite anion adds across the C5-6 double bond of cytosine at acidic pH, resulting in cytosine losing its aromaticity and undergo hydrolysis with the loss of ammonia. The uracil bisulfite adducts re-aromatise to form uracil when pH increases. In this way, cytosine is deaminated to uracil, but mdC remains intact. In Sanger sequencing, all cytosines in the genome that are converted to uracil will read as T, so any remaining bases that are read as C are assumed coming from mdC. The conversion requires the single strand DNA because bisulfite is inaccessible to C5-6 double bond in the double strand. [267]

Bisulfite sequencing fails to discriminate mdC from hmdC because both resist deamination thus are read as C. On the contrary, fdC and cadC are converted to uracil and read as T under bisulfite condition, which is similar to cytosine.

Since the discovery of the epigenetic bases, several methods based on bisulfite sequencing have been developed to discriminate cytosine, mdC, hmdC, fdC, and cadC. The target base is first converted to its adduct or another oxidation state that generates readout different from the original base after bisulfite conversion. Comparing the sequencing results before and after chemical treatment gives sequence information of the target bases. **Fehler! Verweisquelle konnte nicht gefunden werden.** summarizes these methods.

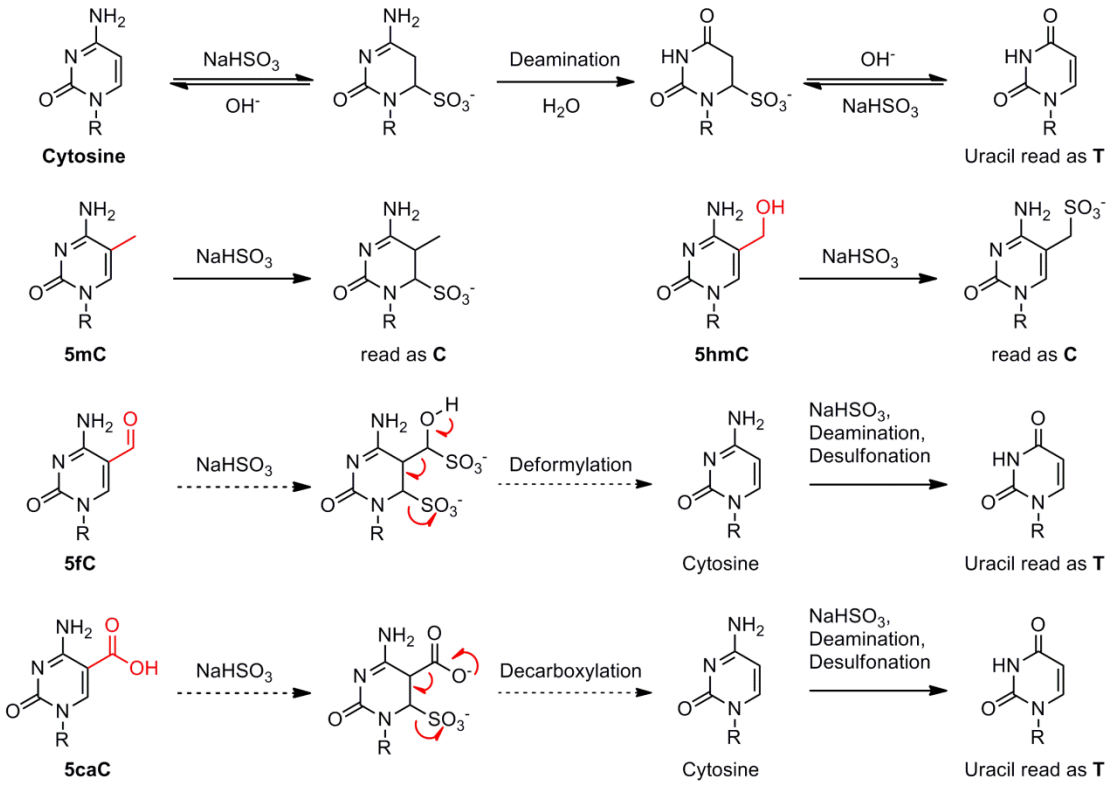


Figure 3-21 Bisulfite sequencing proceeding of cytosine, mdC, and further oxidization bases. Dash arrows show potential mechanisms of deformylation and decarboxylation.

Table 3-1 Summary of bisulfite sequencing and its derivatives for epigenetic bases. C indicates the product after the convention. R indicates bisulfite sequencing readout results. Bases under conversions are highlighted in red.

	5mC	5hmC				5fC				5caC	
	BS-Seq	**TAB-Seq**		**oxBS-Seq**		**redBS-Seq**		**fCAB-Seq**		**CAB-Seq**	
	R	**C**	**R**	**C**	**R**	**C**	**R**	**C**	**R**	**C**	**R**
C	T	C	T	C	T	C	T	C	T	C	T
5mC	C	**5caC**	**T**	5mC	C	5mC	C	5mC	C	5mC	C
5hmC	C	5hmC	C	**5fC**	**T**	5hmC	C	5hmC	C	5hmC	C
5fC	T	**5caC**	**T**	5fC	T	**5hmC**	**C**	**5fC-EtONH$_2$**	**C**	5fC	T
5caC	T	5caC	T	5caC	T	5caC	T	5caC	T	**5caC-Ar**	**C**

He *et al.* reported Tet-assisted bisulfite sequencing (**TAB-Seq**) for mapping hmdC at base resolution.[213] In TAB-Seq, hmdC is attached to glucose using β-glucosyltransferase to protect hmdC from TET oxidation. TET oxidizes mdC to cadC, which is read as T in BS-Seq while sugar protected hmdC is read as C. The difference in the readout indicated the hmdC sites.

Parallel methods for quantitive mapping hmdC and fdC, called as **oxBS-Seq**,[214]

oxidative bisulfite sequencing, and **redBS-Seq**,[268] reductive bisulfite sequencing. hmdC is first oxidized to fdC using $KRuO_4$ and read as T in BS-Seq while untreated hmdC is read as C. Or fdC is first reduced to hmdC using $NaBH_4$,[264] and read as C while fdC is read as T in untreated samples. Combined these two methods with reduced representation bisulfite sequencing (**RRBS**),[269, 270] which selectively enriches for CpG dinucleotides, provides a genomic map of mdC, hmdC, and fdC in mouse embryonic stem cells. These experiments revealed that fdC is present at comparable levels to hmdC and mdC in certain genomic locations.

Later, the so-called chemical assisted bisulfite sequencing **fCAB-Seq**[233] for fdC and **CAB-Seq**[271] for cadC were reported by He group. fdC will be not deaminated and read as C if the aldehyde is blocked by ethyl hydroxylamine. Similarly, cadC will not be deaminated and read as C after connected to (4-aminoethyl)-benzylazide and clicked with a biotin. Therefore, fdC and cadC are identified.

Although these BS-Seq variations allow us to distinguish cytosine oxidized bases, they share common defects of bisulfite sequencing. First of all, the original BS-Seq relies on the conversion of every single unmethylated cytosine residue to uracil and every single epigenetic base to the corresponding derivatives. Since bisulfite converts single-stranded but not double-stranded DNA, incomplete denaturation or reannealing leads to insufficient conversion. The subsequent analysis will incorrectly interpret, taken the cytosine, fdC, and cadC as mdC or hmdC for example. Not only the cytosine may be left unreacted, but also mdC react with bisulfite though with a lower rate than cytosine.[272]

The second challenge in BS-Seq is the genomic degradation that takes place contemporaneously with the conversion,[273] which caused by cytosine 5-methylenesulphonate depyrimidination.[274] No degradation can be observed in A, G and T as no bisulfite adducts form. Protracted bisulfite incubation produces abasic sites and leads to strand scission under basic condition. Therefore, conditions benefiting complete conversion, e.g., lengthy incubation time, elevated temperature, high bisulfite

concentration, can result in genome degradation up to 90%.[273] Given that the starting amount of DNA is usually limited, such extensive degradation is problematic.

Moreover, considering that most proofreading enzymes stall at uracil residues in the template strand, non-proofreading Taq polymerase has to be employed for the second-strand synthesis and PCR amplification, which may lead to sequencing errors.

Finally, high sequencing costs are required to provide sufficient sequencing depth for each oxidized cytosine due to their low abundance.[275] For mapping fdC, the whole genome-scale mapping is impossible given the requirement of unusually high sequencing coverage. So, only partial or reduced representation of the genome has been interrogated, e.g. RRBS. The above limitations above are less disturbing for single-copy loci but are significant in genome-wide sequencing.

3.2.3.5. Single molecular mapping methods

At single molecular level, although discrimination methods of mdC and hmdC with peptide[276] and oligonucleotides[277] were reported, they can hardly be applied in the genomic level mapping. The reasons are multiple: the complexity of the genomic DNA, the signal amplification, and the detection methods, etc.

SMRT

The single molecular real-time (**SMRT**) sequencing method.[278, 279] is designed for sequencing mdC and hmdC (Figure 3-22a). In SMRT sequencing, DNA polymerases catalyze the incorporation of fluorescently labeled nucleotides into complementary nucleic acid strands. The arrival times and durations of the resulting fluorescence pulses yield information about polymerase kinetics and allow direct detection of modified nucleotides in the DNA template, including mdC and hmdC.

Nanopore Sequencing

Nanopore originates from membrane channels that contain a gated pore for nutrients and ions to cross. Using nanopores as polymer sensors date back to the early 1990s.[280] Figure 3-22b illustrates the principle of the system. When a single pore is inserted into

the support device, a characteristic current appears. Single-stranded nucleic acid moves through the nanopore and results in a transient blockade of the ionic current. In 1996, Kasianowicz *et al.* first reported the electrophoretic transport of individual ssDNA and ssRNA molecules through α-haemolysin.[281]

Further developments have shown that DNA molecules can be read at single base level using nanopores coupled to processive enzyme.[282-284] Latest studies demonstrate that the nanopore device can discriminate C, mC, hmC,[285-287] as well as fC and caC.[288] Generally, the optimal speed for DNA sequencing is 1,000-50,000 nucleotide (nt)/s. Nanopore sequencer based on biological pores, e.g. α-haemolysin and MspA, provide a translocation speeds of 2.5-70 nt/s, while the solid-state nanopores, e.g. SiN and Al_2O_3 membranes, detect the bases with a speed of 3,000-50,000 nt/s.[289] Recently, a viscosity gradient system based on the ionic liquid $BminPF_6$ was reported to slow down DNA translocation through monolayer MoS_2 nanopores towards an optimal time resolution and signal resolution.[290]

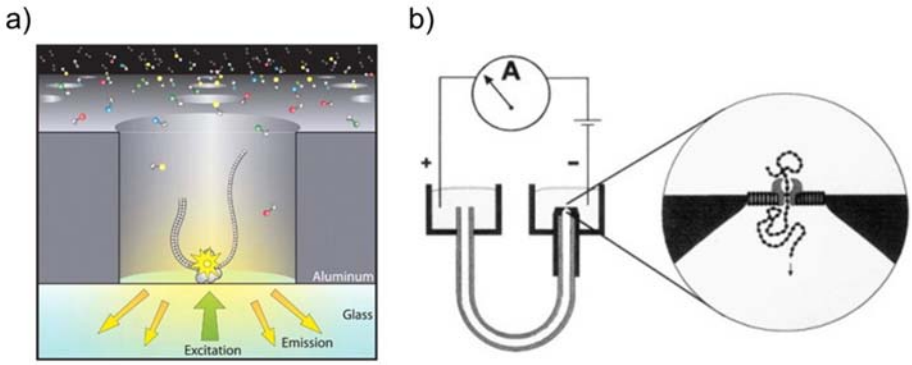


Figure 3-22 Illustration of single molecular sequencing methods: a) Principle of SMRT sequencing. A single molecule of DNA template-bound Φ29 DNA polymerase is immobilized on nanophotonic structure, which enables detection of individual phospholinked nucleotide substrates as they are incorporated into the DNA strand by the polymerase. The figure is adapted from reference.[291] b) Nanopore support device. U-tube supports a lipid bilayer membrane. Hemolysin subunits are added to the cis chamber facing the bilayer, and a voltage is applied positive on the trans side. When a single pore inserts into the bilayer, a characteristic current immediately appears. An amplifier and an analog/digital converter are connected, stored the signals in a computer. The figure is adapted from open source.

For the interest of time and money, sequencing methods based on NGS, bisulfite genome-level sequencing, and single molecular methods can be, but rarely, applied as

conventional clinical diagnostic means at this moment. The growing need for cheaper and faster genome sequencing prompts the development of new technologies. A direct, locus-level detection method for epigenetic bases in the context related to specific diseases would be poised to meet this challenge.

In Chapter 5, this task is attempted to fulfill in a simple way. An oligonucleotide probe to react with fdC on the target strand is described. Quantification of the fdC content in a specific locus in the genome from *Mus musculus* embryonic stem cells using the probe is discussed.

4. Part I: Pyrazole Ligandoside

4.1. Aims of the project and rationale

The first part of the thesis was aimed in developing a novel metal incorporated base pair and to broaden its application.

In 2012, Wang and co-workers genetically incorporated a metal-binding non-natural amino acid pyTyr (**1**) into green fluorescent protein for studying the electron transfer mechanism.[292] This work inspired the design of the pyrazole ligandoside (Figure 4-1).

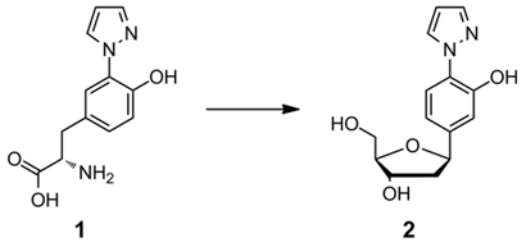


Figure 4-1 Origin of the idea: from pyrazole non-natural amino acid **1** to pyrazole noncanonical nucleoside **2**.

Identical to pyTyr **1**, 1-(2-hydroxyphenyl)pyrazolate based nucleoside **2** should be able to chelate copper ion with one oxygen atom and one nitrogen atom solely. Homo pyrazole ligandosides recognize each other with a metal ion in the center that consists of two oxygen atoms and two nitrogen atoms.

Characteristic differences resulted from changing the hydroxyl group to a fluorine atom, to a hydrogen, or replacing C with N atom, as shown in Figure 4-2. Although attempts were made toward these derivatives and indeed the bone building blocks have been obtained. However, the synthesis of the nucleosides failed.

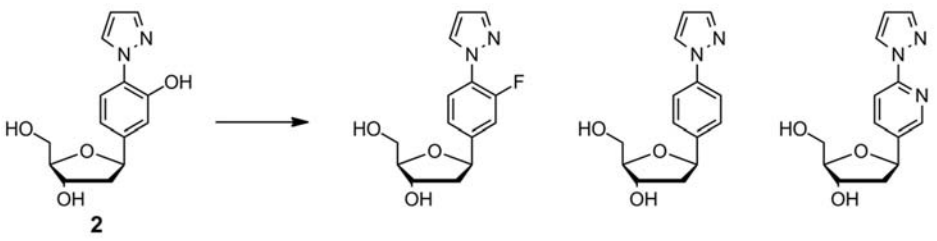


Figure 4-2 Further design of pyrazole-based nucleoside.

The synthesis of salen ligandoside inspired the synthesis in this work.[173] By first introducing a bromine atom at the *para*-position in the phenol, a cuprate-based coupling of the base building blocks with toloyl-protected 1-chlororiboside, Hoffer's α-chloro sugar,[293] furnished a mixture of α/β-configured protected nucleosides (Figure 4-3). The base building blocks could be obtained either by substitution of iodo- and fluoro-benzene or azotization and cyclization of aniline, depending on the substrates. In the next step, a suitable protecting group was chosen to facilitate the C-C bond formation for the base building block and the Hoffer's α-chloro sugar.

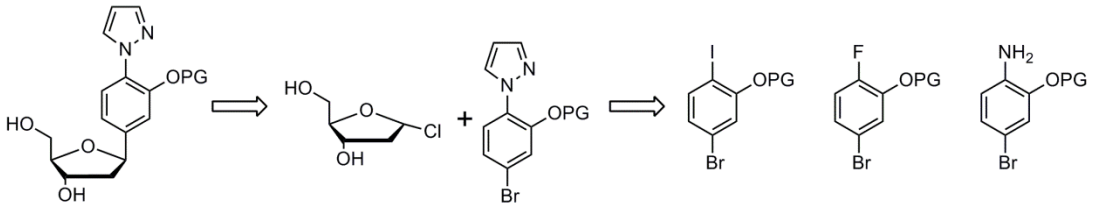
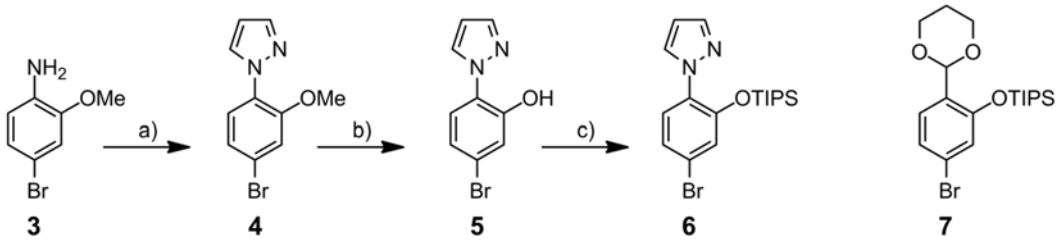


Figure 4-3 Retrosynthetic analysis of the pyrazole nucleoside. PG = protecting group.

4.2. Synthesis of the ligandoside containing oligonucleotide

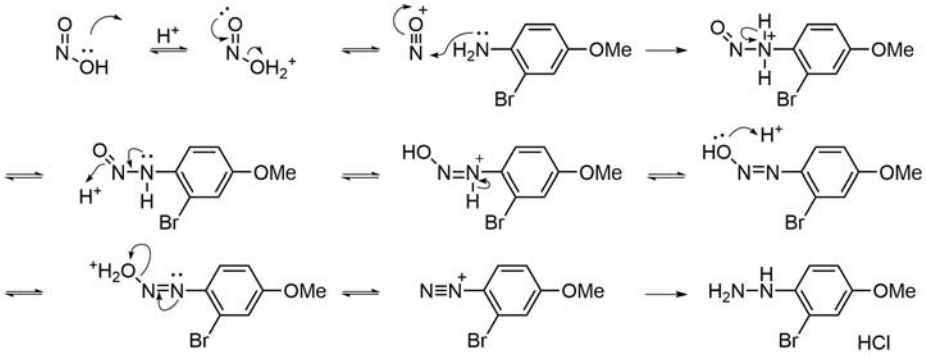
4.2.1. Synthesis of pyrazole ligandoside

The synthesis of pyrazole ligandoside was started from the commercially available reagent 4-bromo-2-methoxyaniline (**3**) as depicted in Scheme 4-1.



Scheme 4-1 Synthesis of the protected ligand building block **6**. a) 1) $NaNO_2$, $SnCl_2$, HCl, 0 ºC, 66%, 2) 1,1,3,3-tetramethoxypropane, EtOH, HCl, reflux, 85%; b) Et_2O-BBr_3, DCM, -78 ºC, 68%; c) TIPSOTf, DIPEA, DCM, 0 ºC, 99%. TIPS = triisopropylsilyl.

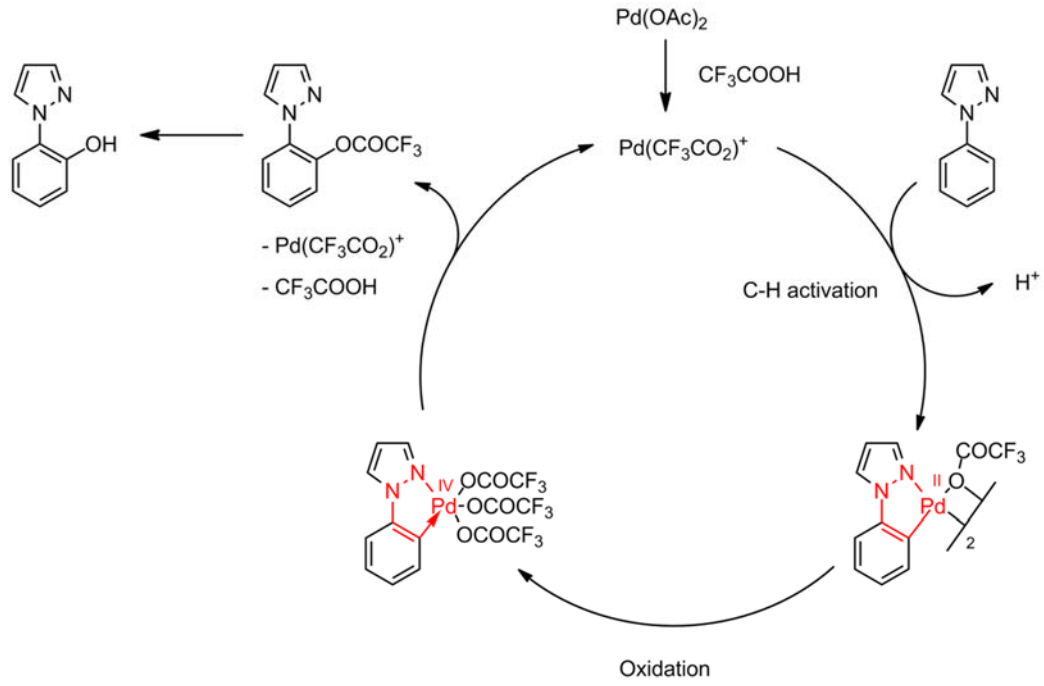
Diazotization of compound **3** with sodium nitrite gave the diazonium salt (Scheme 4-2),[294] followed by reduction with stannous chloride afforded the corresponding substituted phenyl hydrazine hydrochlorides intermediate.



Scheme 4-2 Mechanism of diazonium salt formation and the hydrazine intermediate.

Without further purification, the hydrazine intermediate went on to cyclize with excessive malonaldehyde bis(dimethyl acetal) in ethanol to form the pyrazole ring in **4**.[295] Cleavage of the methyl group with Lewis acid borane at -78°C provided the desired phenol **5**. The demethylation reaction proceeds via a bimolecular mechanism involving two Et_2O-BBr_3 adducts with the formation of phenol and bromomethane.[296] (Scheme 4-4) Low temperature is necessary because of the high energy of the B-O bond.

After this project was finished, H. Batchu *et al.* reported that the *ortho*-hydroxy group is accessible by *N*-phenylpyrazoles using lead(II) acetate as catalyst.[297] In a plausible mechanism, Pd(II) activates the *ortho*-H to form a five-membered cyclopalladium intermediate. After oxidation and elimination, a C-O bond can be formed. In our case, Cu(I) was applied when coupling the base building block with the sugar moiety, thus, a similar cyclo intermediate may form with H or OH and the N on the pyrazole rings (Scheme 4-3). This explains why the hydroxyl group should be protected before coupling to the sugar and why *N*-phenylpyrazole and other base building blocks in Figure 4-2 failed to generate the desired product when reacted with the sugar.

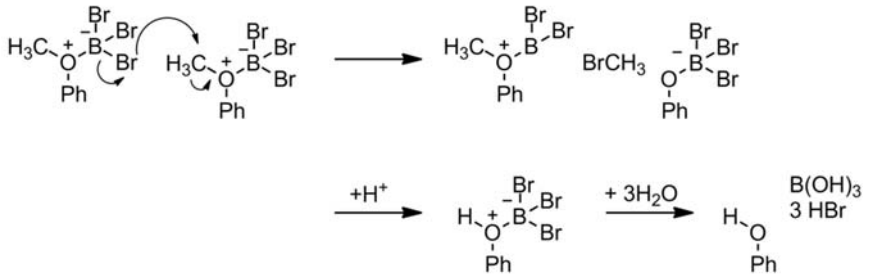


Scheme 4-3 Proposed mechanism of *ortho*-hydroxylation of *N*-phenylpyrazole.

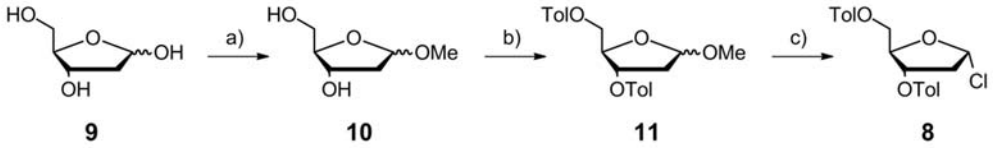
Subsequent reaction of **5** with triisopropylsilyl triflate provided the key intermediate **6**. Because the electronic density of phenyl in 1-phenyl-1*H*-pyrazole is higher than 2-phenyl-1,3-dioxane in the case of salen base **7**, the hydroxyl group here is much more acidic. As a result, the reaction rate of **5** is faster than the salen base building block **7**.

With the key building block **6** in hand, we started working on the sugar moiety. The Hoffer's α-chloro sugar (**8**) was synthesized from 2-deoxyribose (**9**). (Scheme 4-5) First, the ribose was methylated at 1'-hydroxyl group (**10**), followed by

4-methylbenzoyl protection giving **11**. Then, chloride substitution of the methoxyl group furnished the sugar building block **8** after precipitation in anhydrous ether.



Scheme 4-4 Biomolecular mechanism of BBr_3-assisted cleavage of methyl ethers.



Scheme 4-5 Synthesis of the protected ligand building block **8**. a) AcCl, MeOH, Ag_2CO_3, rt; b) 4-methylbenzoyl chloride, pyridine, 45 ºC, 87% for 2 steps; c) AcCl, HCl, AcOH, 0 ºC, 55%. Tol = 4-methylbenzoyl.

Due to their insolubility or failure to form a copper complex with the substrates, the base building blocks of benzene, pyridine, fluorobenzene and anisole (Figure 4-3) all failed to form a C-C bond with the sugar. The Corey-House synthesis was achieved with intermediate **6**.[298, 299] The reaction occurred in three steps (Scheme 4-6a). First, bromobenzene was treated with *t*-butyllithium to convert it into a phenyl lithium species. **6** was reluctant to undergo a lithium halogen exchange. In practice, neither *n*- nor *s*-butyllithium could generate the organic lithium species efficiently. So an excessive and sufficiently strong base, *tert*-butyllithium, was used. Because the product *t*-BuBr can react with *t*-BuLi, 2 equivalents *t*-BuLi must be used (Scheme 4-6b). Second, the phenyl lithium species was treated with copper bromide to create a lithium diphenyl cuprate intermediate, called "Gilman reagent" in honor of H. Gilman's contribution to the field.[300] In the Gilman reagent, the cuprate and lithium ions directly link to the carbon atom, forming an 8-membered ring with two lithium and two cuprate ions coordinating between four phenyl groups (Scheme 4-6c),[301] thus negative charges accumulate on the coordinate carbon atom and make the C-Cu and

C-Li more reactive. The reaction has to be conducted in ethyl ether to avoid side reactions. Finally, the diphenyl cuprate was reacted with the protected chloride ribose **8** to furnish the cross-product ligandoside. However, the coupling of the Gilman reagent with bromobenzene resulted in the undesired dibenzene by-product (Scheme 4-6b). β- and α-conformers are very close to each other on TLC, but they were still separable by column chromatography.

a) PhBr + *t*-BuLi → PhLi + *t*-BuBr

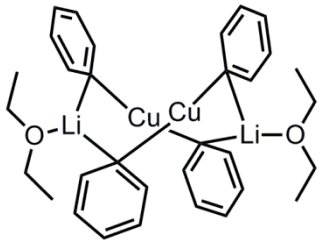
2 PhLi + CuI → Ph_2CuLi + LiI

Ph_2CuLi + R-Cl → PhR + PhCu + LiCl

b) *t*-BuBr + *t*-BuLi → $(CH_3)_2C{=}CH_2$ + LiBr + *t*-BuH

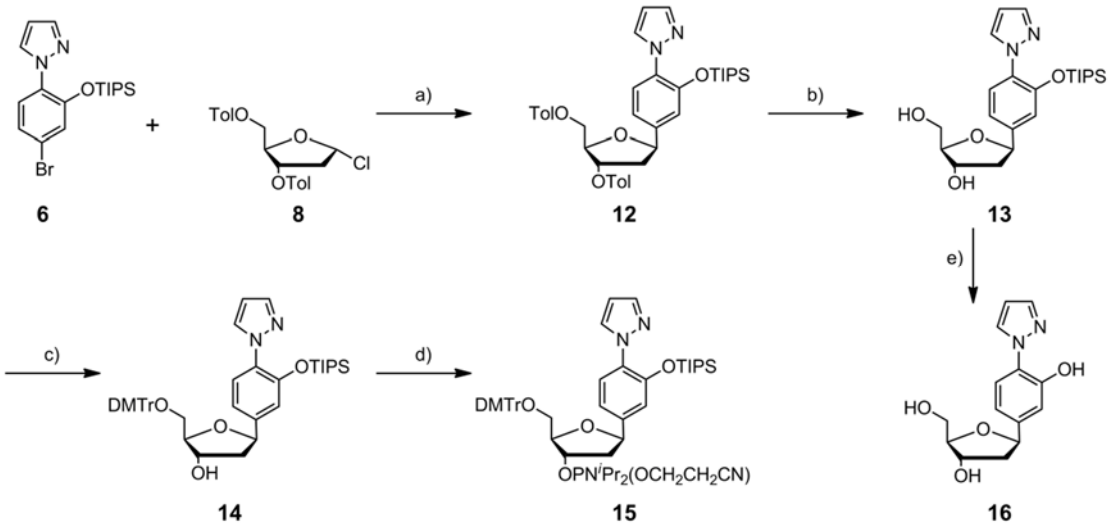
Ph_2CuLi + Ph-Br → Ph-Ph + PhCu + LiBr

c)



Scheme 4-6 Mechanism of Corey-House synthesis: a) reaction steps towards nucleotide **12**. PhBr represents **6**. R-Cl represents Hoffer's α-chloro sugar **8**. b) side reaction in the Corey-House synthesis; c) structure of 8-membered ring Gilman reagent.

The yield of **12** depended on several factors: 1) Moisture and oxygen can degrade the organolithium intermediate or oxidize the Cu(I) complex. 2) Insufficient reaction time and improper temperature control for the second step can leave a certain amount of organolithium unreacted. In practice, increasing the temperature rapidly at the final step generated the undesired α-conformer product. 3) The hydrochloric acid generated from the chloride ribose can destroy the Gilman reagent.



Scheme 4-7 Synthesis of the building block **15** and of the nucleoside **16**. a) 1) *t*-BuLi, Et_2O, -78 °C; 2) CuBr-Me_2S, -30 °C; 3) **8**, rt, 53%; b) K_2CO_3, MeOH, rt, 65%; c) DMTr-Cl, DIPEA, DCM, rt, 74%; d) P(OCH$_2$CH$_2$CN) (N*i*Pr$_2$)$_2$, diisopropylammonium tetrazolide, DCM, rt, quant.; e) TBAF, THF, rt, 87%.

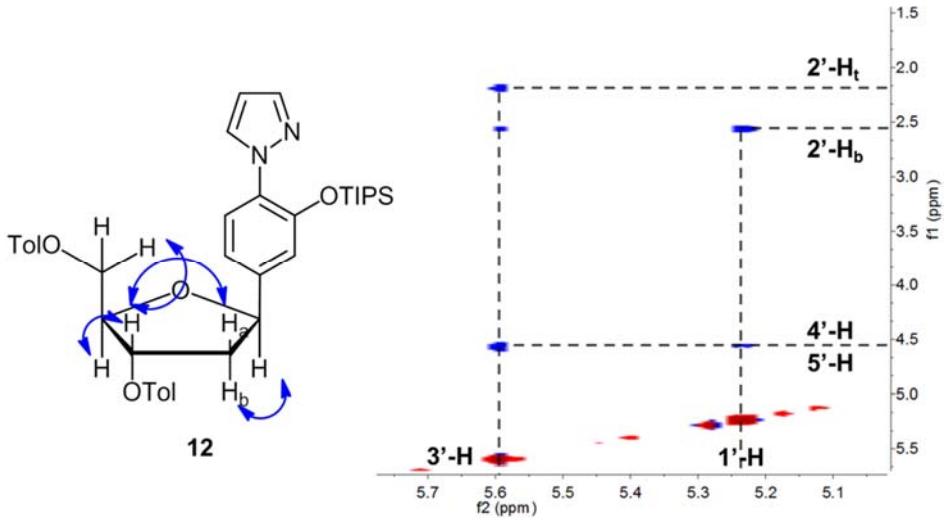


Figure 4-4 Pentose region of the 400 MHz NOESY spectrum of **12** in $CDCl_3$.

Next, deprotection of **12** to **13** was achieved with potassium carbonate in methanol (Scheme 4-7). Appropriate reaction time and concentration must be attained to ensure that only the 4-methylbenzoyl but not TIPS group is cleaved. NOESY spectra for both conformers were measured to confirm the ligandoside configuration. NOESY allows us to correlate nuclei through space smaller than 5Å. From the spectrum of **12** (Figure 4-4), it can be seen that in the β-ligandoside the 1'-H (5.25 ppm) is correlated with the 2'-H bottom (2.58 ppm). 2'-H top (2.20 ppm) is coupled with 3'-H (5.61 ppm). 3'-H is

further correlated with 4'-H (4.57 ppm) and 5'-H (4.60 ppm). No correlation between 1'-H and 3'-H can be observed. In contrast, α-conformer shows a correlation peak between the 1'-H and the 3'-H. The 3'-H is coupled to 2'-H top; 2'-H bottom correlates with the aromatic protons. Based on the data, the configuration of the obtained ligandoside is verified.

Finally, **13** was protected with dimethoxytrityl chloride (DMTr-Cl) at the 5'-hydroxyl group to give **14** and **14** was further reacted with the phosphoramidite reagent to provide **15**, needed for oligonucleotide solid-phase synthesis.

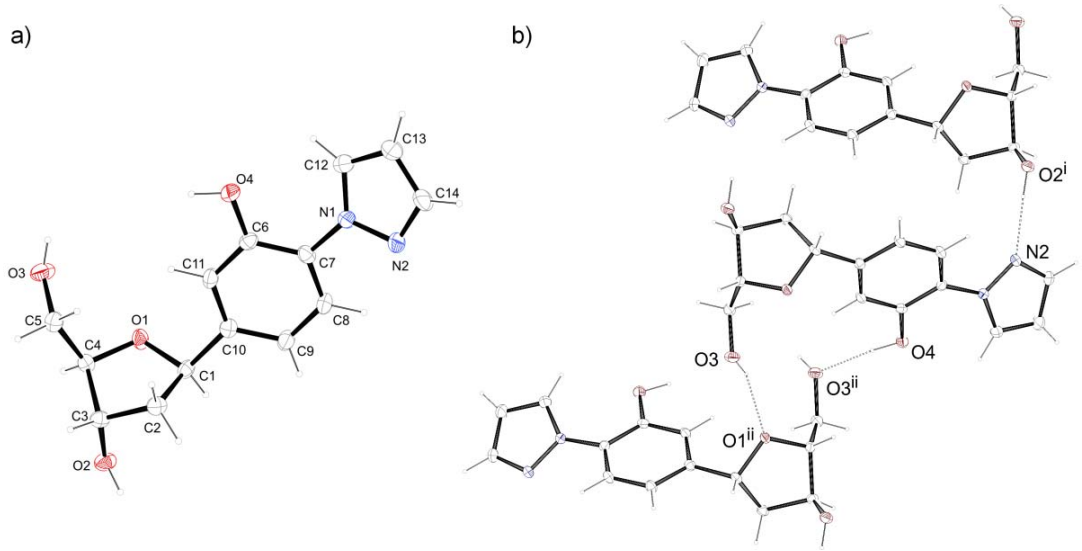


Figure 4-5 X-ray crystal structure of nucleoside **16**: a) crystal structure of single molecular; b) neighboring interaction in a unit cell.

Completely deprotected **13** furnished the pyrazole ligandoside **16**. Small crystals of **16** were attained by slow evaporation of the unpolar ethyl acetate solution. As presented in Figure 4-5, the result of an X-ray diffraction analysis further proved the β-configuration. In the X-ray crystal, the phenyl ring and the pyrazole ring are slightly tilted against each other by 23.9°~24.5°. The dihedral angle between two planes is narrow which facilitates copper chelating with the nitrogen and oxygen atom. Unexpectedly, the nitrogen atom (N2) and the hydroxyl group (O4) of the pyrazole are not on the same side. In the unit cell (Figure 4-5b), an interactions between N2 and $O2^{i}$ of one adjacent sugar, and O4 and $O3^{ii}$ of another adjacent sugar are observed, which could be responsible for the "anti" conformation around the anti anti bond.

Next, the UV-Vis spectrum of ligandoside **16** was measured. The highest peak is at 246 nm; the second highest peak is found at 284 nm, similar to the methylated monomer **19** (see Chapter 4.3.1.2) resulting from the benzyl-pyrazole bicycle system. As canonical bases have an average peak at 260 nm, the molar absorbance of **16** was determined at 260 nm, to calculate the concentration of the oligonucleotides.

4.2.2. Synthesis of oligonucleotides with ligandoside

With the ligandoside **15** in hand, solid-phase synthesis of DNA was conducted. The coupling yield of **15** was equally good compared to the natural bases even without extending the coupling time. Pyrazole base and TIPS protecting group did not interrupt the synthesis cycle.

As stated in the previous report,[174] TIPS groups on the DNA strands can be easily deprotected during ordinary solid-phase cleavage conditions with a mixture of NH_3 aq. / $MeNH_2$ aq. (AMA solution)[302] at 65°C for 5 min. For the strand with more than one pyrazole base, the cleavage time was extended to 10 min to ensure complete deprotection.

The DMTr-off mode was applied for the solid-phase synthesis so DMTr deprotection was not necessary after cleavage from the solid phase, and only one round of HPLC purification was executed. Typically, the highest peak referred to the desired product. Lyophilization after collecting the corresponding eluate gave the anhydrous powder of the oligonucleotide, which was then analysed by MALDI-TOF.

In the case of insertion of pyrazole and salen ligandosides (**ODN 8a/b**, **9a/b,** and **10a/b**), the DNA containing solid-phase material was first exposed to dichloroacetic acid solution to remove the cyclic acetal group, then furthere cleavage was performed with AMA solution as described above. Prior to preparative HPLC purification, the crude oligonucleotide sample was mixed with 20% acetic acid aq. to avoid formation of imines from ammonia and the salen aldehyde group.

A summary of oligonucleotides prepared in this work is listed in Table 4-1.

Table 4-1 Synthesized and purchased oligonucleotides in chapter 4. Bold and red letters X represent pyrazole ligandoside (Pz), M represent methylated pyrazole ligandoside (Pm), D represent 1',2'-dideoxyribose and S represent salen ligandoside. Found values indicate the mass results measured by MALDI-TOF. Mass values of purchased oligonucleotides are not shown.

Entry	5' ---------- 3'	mer	calc.	found
0a	CAC ATT AGT GTT GTA	15	/	/
0b	TAC AAC ACT AAT GTG	15	/	/
1a	CAC ATT AXT GTT GTA	15	4588.8	4585.0
1b	TAC AAC AXT AAT GTG	15	4606.8	4602.9
1c	TAC AAC AAT AAT GTG	15	/	/
1d	TAC AAC ACT AAT GTG	15	/	/
1e	TAC AAC AGT AAT GTG	15	/	/
1f	TAC AAC ATT AAT GTG	15	/	/
2a	CAC ATT ADT GTT GTA	15	4430.8	4428.8
2b	TAC AAC ADT AAT GTG	15	4448.8	4447.3
3a	CAC ATT AMT GTT GTA	15	4602.8	4597.9
3b	TAC AAC AMT AAT GTG	15	4620.9	4616.3
4a	ATG CGA CCX TCC CT	14	4206.7	4208.8
4b	AGG GAX GGT C	10	3140.6	3139.9
5a	CAC ATT XXT GTT GTA	15	4613.8	4611.6
5b	TAC AAC AXX AAT GTG	15	4640.9	4636.1
6a	GCGCG XXXXX GGCCG	15	4758.9	4757.6
6b	CGGCC XXXXX CGCGC	15	4678.9	4676.6
7a	GCGCG XXXXX XXXXX GGCCG	20	6449.2	6452.4
7b	CGGCC XXXXX XXXXX CGCGC	20	6369.2	6370.8
8a	CAC STT AXT GTS GTA	15	4572.7	4570.0
8b	TAC SAC AXT AAS GTG	15	4590.7	4586.8
9a	CAC STT AMT GTS GTA	15	4586.7	4586.8
9b	TAC SAC AMT AAS GTG	15	4604.7	4604.2
10a	CAT GST XGS AXC STX CST GCA	21	6452.9	6450.7
10b	TGC ASG XAS GXT SCX ASC ATG	21	6501.9	6500.1
11a	GGA TCG XTC TCC CTA TAG TGA GTC GTA TTA	30	9232.6	9228.5
11b	FI-TAA TAC GAC TCA CTA TAG GGA GA	23	/	/
11c	GGA TCG ATC TCC CTA TAG TGA GTC GTA TTA	30	/	/
11d	GGA TCG CTC TCC CTA TAG TGA GTC GTA TTA	30	/	/
11e	GGA TCG GTC TCC CTA TAG TGA GTC GTA TTA	30	/	/
11f	GGA TCG TTC TCC CTA TAG TGA GTC GTA TTA	30	/	/

4.3. Evaluation of the pyrazole ligandoside

4.3.1. Oligonucleotides with single-inserted ligandoside

4.3.1.1. Complexation properties

Once the single ligandoside base pair was inserted into two strands (**ODN 1a/b**), the biophysical properties of the duplex were investigated. The denaturing profiles of the double strands with and without the present of metal ions in different buffers were measured. Before the measurement, the two complementary strands were mixed with the respective metal ion in the corresponding buffer. The solution was denatured at 90°C for 5 min and then slowly cooled down to room temperature for reannealing. The modified duplex was formed in the annealing process together with metal ions.

For the control strands, the Pz/Pz pair were replaced a with G/C pair as in duplex **0a/b**. Here, the melting temperature is relatively stable in the system with varied pH from 6.0 to 9.0. Another control experiment duplex **2a/b** (Table 4-1) was made with the abasic nucleoside mimic, 1',2'-dideoxyribose was used instead of the G/C pair. This gave a less stable duplex with a melting temperature as low as 29°C at pH 9.0.

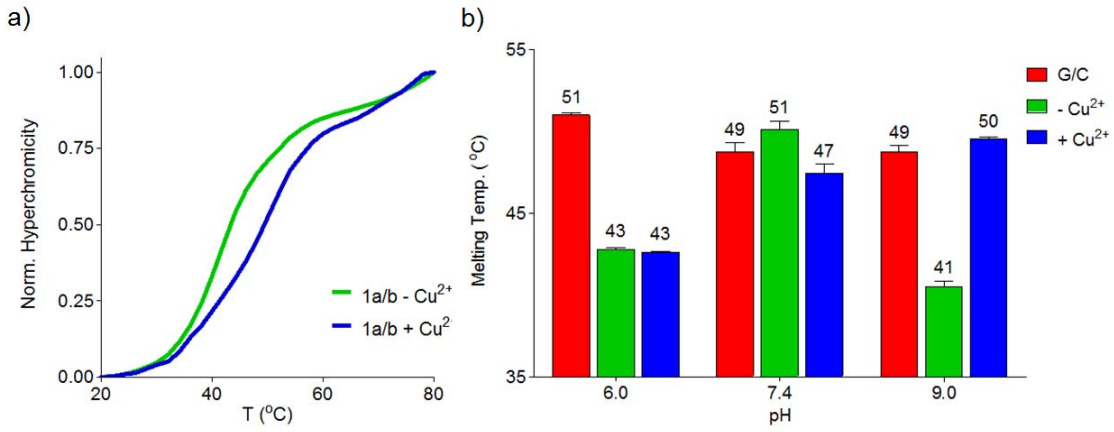


Figure 4-6 Melting profiles of duplex **1a/b**: a) melting curves of **1a/b** with and without Cu^{2+} at pH 9.0; b) Plot of melting temperatures of duplex **1a/b** with and without Cu^{2+} at varied pH, compared with **0a/b**. Data are mean ± SD of three replicates. Condition: 150 mM NaCl, 10 mM Na_2HPO_4/NaH_2PO_4 buffer pH 6.0 / 7.4 or CHES buffer pH 9.0, 1 µM oligonucleotides, with and without 1 µM Cu^{2+}, the final volume of 200 µL.

In the case of the Pz containing duplex **1a/1b**, the melting temperature was around 43°C in NaH_2PO_4-Na_2HPO_4 buffer (pH = 6.0), independent on Cu^{2+} (Figure 4-6).

When the pH was raised to 7.4, the duplex without copper reached 51°C, which was higher than when Cu^{2+} was added (47°C). When the buffer was changed to CHES buffer (pH = 9.0), the melting point for duplex without copper decreased to 41°C, while a melting temperature of 50°C was detected after adding the Cu^{2+}. This is a little higher than the control strands under the same condition, indicating that Cu^{2+} was probably incorporated into the strands in the basic environment. The chelating mechanism will be described in 4.3.1.2.

Compared with the homo-salen base pair, Pz-Cu^{2+}-Pz was less stable. In the same sequence context and buffer system but at higher duplex concentration, S-Cu^{2+}-S with ethylenediamine resulted in a melting temperature of 55°C. The melting temperature of Pz-Cu^{2+}-Pz is close to a methylated salen pair (52°C) or a homo hydroxypyridone pair (H-Cu^{2+}-H, 50°C),[154] Therefore, the geometry of the Pz-Pz is not the best for complexation inside the duplex.

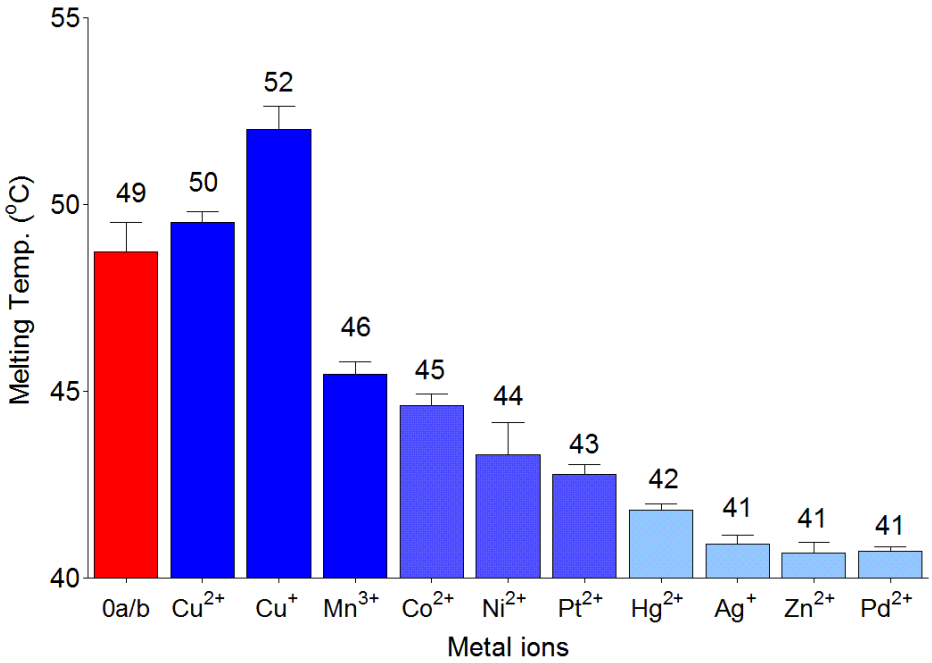


Figure 4-7 Plot of melting temperatures of duplex **1a/b** with varied metal ions, compared with **0a/b**. Condition: 150 mM NaCl, 10 mM CHES buffer pH 9.0, 1 µM oligonucleotides, with 1 µM metal ion, the final volume of 200 µL. Data are mean ± SD of three replicates.

Apparently, Cu^{+} exhibits a greater stabilization effect in the basic environment (Figure 4-7). At pH 9.0, Cu^{+} stabilized the duplex from 41°C to 52°C while a value of 50°C

was formed with Cu^{2+}. No impact on the stability was seen when the pH was raised from 6.0 to 7.4.

Manganese ions stabilized the Pz substituted duplex as well, increasing the melting temperature by 5°C in CHES buffer. According to the previous report,[303] Mn^{2+} will be oxidized to Mn^{3+} when chelated inside a salen ligand. Mn^{2+} is supposed to show a similar behavior here, adding one positive charge inside the double helix.

Scanning other metal ions for thermostability showed no improvement, as illustrated in Figure 4-7. Metal ions like Co^{2+}, $Ni^{2+,}$ and Pt^{2+} ions have only weak chelating effects with the pyrazole ligandoside, raising the melting point 2~4°C, while Hg^{2+}, Ag^{+}, Zn^{2+} and Pd^{2+} ions displayed no measurable interactions. The melting points stayed mostly unchanged.

In a single experiment of denaturing and annealing with Cu^{2+} and Mn^{3+}, a strong hysteresis between the heating and the cooling curves was observed (Figure 4-8a). For the copper ion, 50°C was measured for heating, while cooling provided 47°C for reverse. In the case of Mn^{3+} ion, the two values are 46°C and 41°C. This difference shows kinetic difference bewteen complex formation and destruction during heating and cooling of the duplex.

Titration experiments with copper and manganese ions were subsequently conducted. The experiments were carried out for each ion from 0.0 to 2.0 equivalents in CHES buffer (Figure 4-8b, c). Clearly, the complexation is finished after adding the first equivalent of Cu^{2+}, and further titration did not stabilize the duplex to a higher degree. This result supported the duplex:Cu^{2+}=1:1 stoichiometry. Though titration of manganese ion was not as clear as that of copper, 1:1 stoichiometry can be deduced from the plot (Figure 4-8d). It can be attributed to the impact of the CHES buffer and that excessive Mn^{3+} ion destabilizes the ligand inserted duplex.

Calculation of thermodynamic parameters is possible based on the data of the melting curves obtained above (Table 4-2). At low pH, Cu^{2+} incorporation made a greater enthalpy change in the annealing process, which decreased from -119.3 kJ/mol to

-131.9 kJ/mol. At pH 9.0, ΔH increased from -121.5 kJ/mol to -111.0 kJ/mol. The trend is same for the entropy change. At pH 9.0, the loss of the disorder between with and without copper became smaller. Cu^{2+} obviously contributes to ΔH and ΔS term particularly at higher pH.

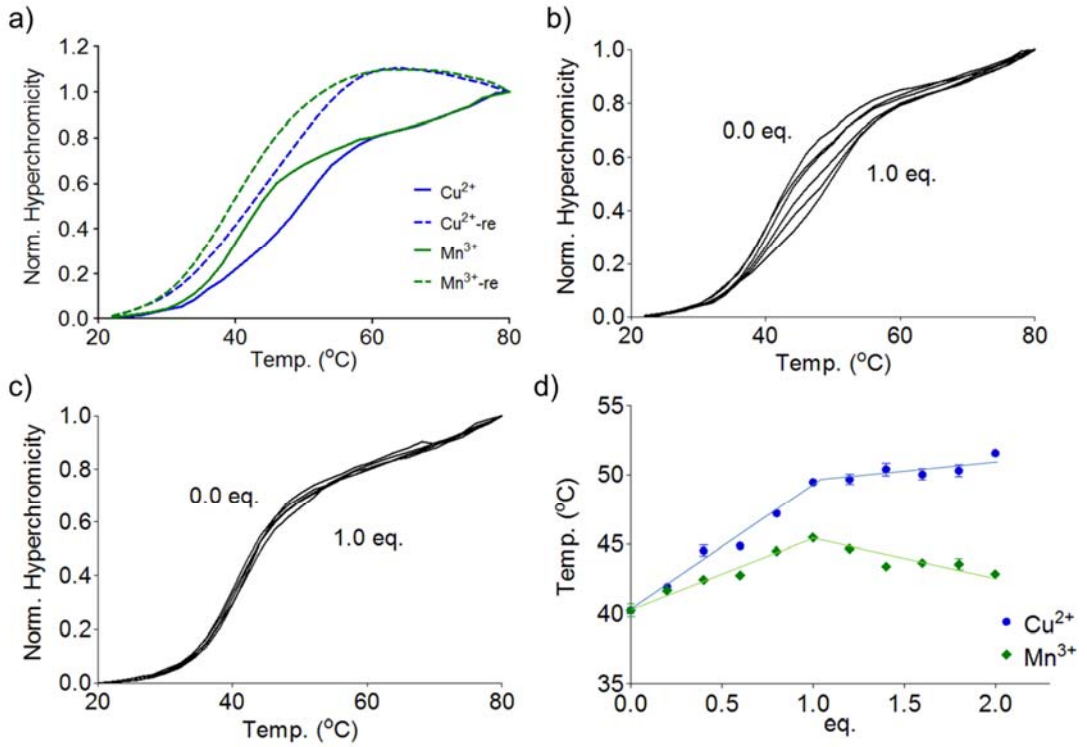


Figure 4-8 Metal ion titration experiments of duplex **1a/b**: a) denaturing and reannealing curves of **1a/b** with 1.0 eq. Cu^{2+} and Mn^{3+}; b) melting curves of titration with Cu^{2+} from 0.0 to 1.0 eq., step of 0.2 eq.; c) melting curves of titration with Mn^{3+} from 0.0 to 1.0 eq., step of 0.2 eq.; d) Plot of the melting temperatures against the metal ions equivalent. Data are mean ± SD of three replicates. Conditions: 150 mM NaCl, 10 mM CHES buffer pH 9.0, 1 µM oligonucleotide, 0.0-2.0 µM Cu^{2+}/Mn^{3+}, the final volume of 200 µL.

Table 4-2 Thermodynamic data for duplex **1a/b**. These values are determined by van't Hoff plots from the melting profiles.

	- Cu^{2+}			+ Cu^{2+}		
pH	ΔH° [kcal mol^{-1}]	ΔS° [cal K^{-1}mol^{-1}]	ΔG°$_{310K}$ [kcal mol^{-1}]	ΔH° [kcal mol^{-1}]	ΔS° [cal K^{-1}mol^{-1}]	ΔG°$_{310K}$ [kcal mol^{-1}]
6.0	-119.3	-373	-3.7	-131.9	-414	-3.6
7.4	-128.5	-397	-5.4	-125.9	-390	-4.8
9.0	-121.5	-388	-1.1	-111.0	-345	-4.0

Circular dichroism (CD) spectroscopy was employed to investigate the duplex conformation. In general, the CD spectrum of B-form DNA has a positive band near 270 nm, a negative band near 250 nm and a more intense negative band at 210 nm. A typical curve was obtained in the control experiment (Figure 4-9a). As expected, the duplex containing the pyrazole ligandoside gave a similar curve as the canonical duplex. When copper or manganese ions were added, no apparent changes in duplex conformation were observed. The DNA duplexes keep their canonical B-form when metal ions are incorporated into the ligandoside pairs. With the temperature rising from 20°C to 80°C, the spectrum shift is similar to the change in natural duplex denaturing process (Figure 4-9b,c,d). It is concluded that the duplex conformation is unchanged upon metal ion complexation.

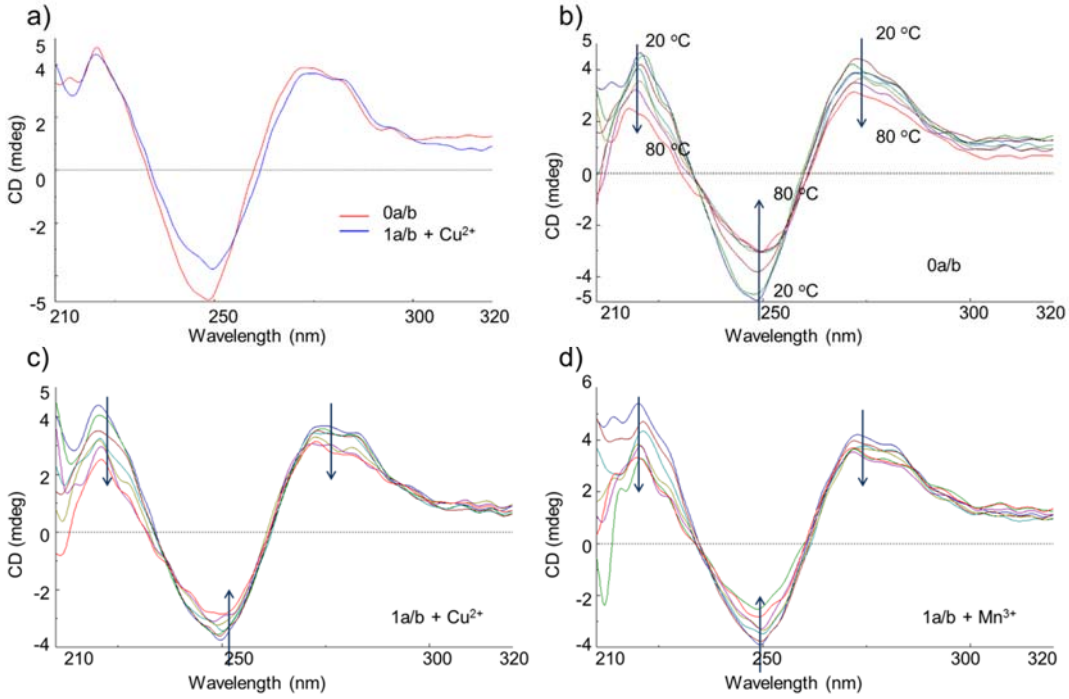
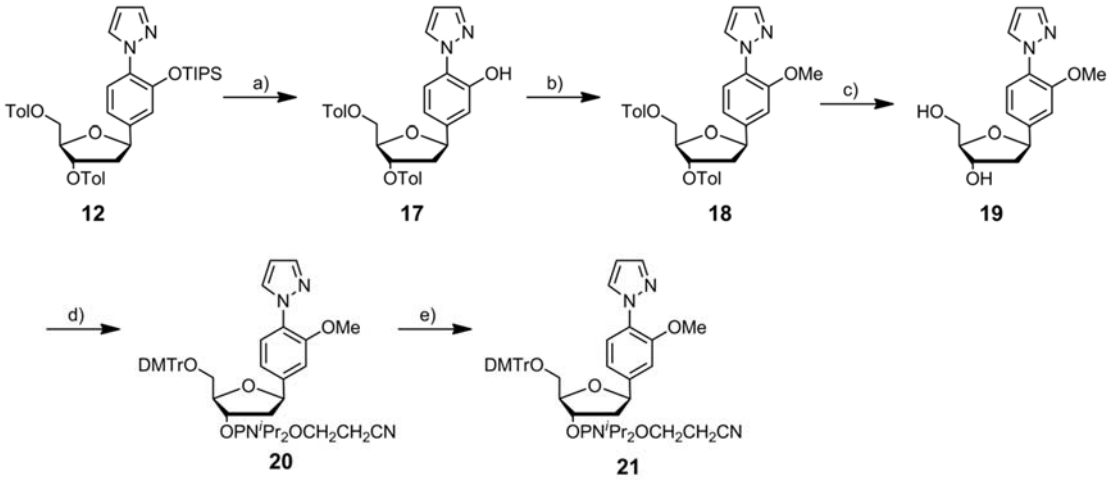


Figure 4-9 Circular dichroism spectrum of **1a/b** with Cu^{2+} or Mn^{3+}, compared with **0a/b**. a) CD spectrum of duplex **1a/b** with equal equivalent of Cu^{2+} at 20 °C; b) overlaid CD spectrum of duplex **0a**/**b** from 20 °C to 80 °C; c) and d) overlaid CD spectrum of duplex **1a**/**b** with equal equivalent of Cu^{2+} or Mn^{3+}, from 20 °C to 80 °C. Conditions: 150 mM NaCl, 10 mM CHES buffer pH 9.0, 1 μM oligonucleotides, 1 μM Cu^{2+}/Mn^{3+}, the final volume of 200 μL.

4.3.1.2. Complexation mechanism

As described above, the thermostability of duplex **1a/b** was pH-dependent. To clarify this issue, we synthesized methylated pyrazole ligandoside **19** (Scheme 4-8). The hydroxyl group was transformed into methoxyl group, deprived of any copper chelating ability. The methylated building block **19** provided us therefore with the possibility to study the coordination mechanism in the DNA duplex.

We experienced of that, **4** (Scheme 4-1) cannot form a C-C bond with the 1'-chloro ribose **8** (Scheme 4-7). Therefore, it was necessary to synthesize **18** from intermediate **12**. Selective deprotection of the TIPS group followed by deprotonation and methylation provided compound **18**. Basic deprotection of **18** furnished compound **19** with excellent yields. **19** is charactized with UV spectroscopy by absortion at the maximum wavelength of 249 nm. Succeeding tritylation and convertion of **19** into the phosphoramidite gave the building block **21** (Pm) for solid-phase synthesis.



Scheme 4-8 Synthesis of the nucleoside **19** and building block **21**. a) TBAF, THF, rt, 83%; b) NaH, CH_3I, DMF, 0 °C, 76%; c) K_2CO_3, MeOH, rt, 90%; d) DMTr-Cl, DIPEA, DCM, rt, 75%; e) $P(OCH_2CH_2CN)(N{i}Pr_2)_2$, diisopropylammonium tetrazolide, DCM, rt, quant.

After solid-phase synthesis, strands **ODN 3a/b** with Pm instead of Pz was obtained. The methoxy group cannot be deprotonated, and coordination with metal ions is impossible. Indeed, we found that pH effect were now negligible (Figure 4-10). The data proves that the hydroxyl group is crucial for the interaction with copper and the duplex. A proposed mechanism is depicted in Figure 4-11.

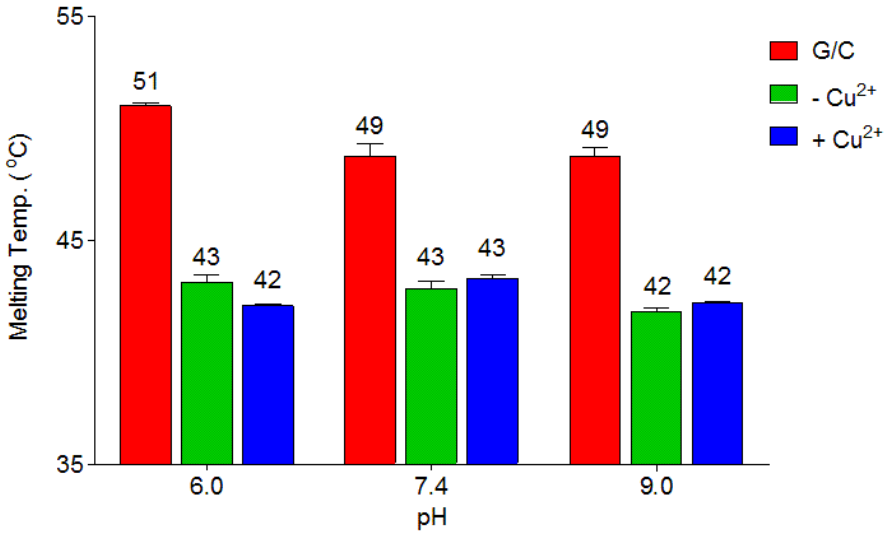


Figure 4-10 Plot of melting temperatures of duplex **3a/b** with and without Cu^{2+} at varied pH, compared with **0a/b**. Condition: 150 mM NaCl, 10 mM Na_2HPO_4/NaH_2PO_4 buffer pH 6.0 / 7.4 or CHES buffer pH 9.0, 1 µM oligonucleotides, with and without 1 µM Cu^{2+}, the final volume of 200 µL. Data are mean ± SD of three replicates.

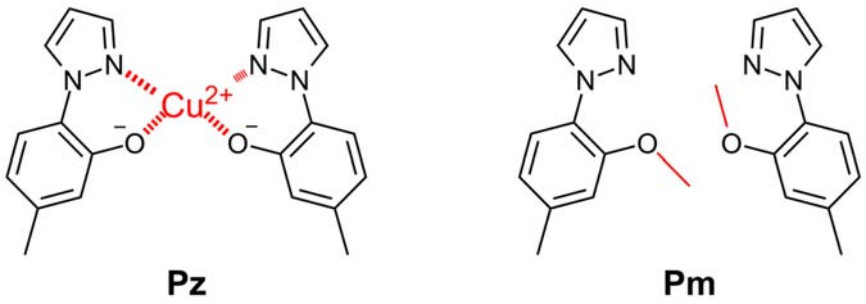


Figure 4-11 Proposed mechanisms of Pz and Pm coordinating with Cu^{2+} under basic condition.

4.3.1.3. Complexation orthogonality

The intention for DNA replication experiment involving the ligandoside prompted us to investigate the mismatch profile of pyrazole ligandosides at the biophysical level. Single Pz-inserted strand **1a** showed a lower melting temperature compared to the reference duplex at pH 9.0. This indicated a loose interaction of Pz with cytosine and faint interaction with adenine and guanine (Figure 4-12). In this property the pyrazole base is similar to a pyrimidine, so it shows more favorable interaction with the purine base. However, at pH 9.0, Pz-A pair (39°C) is very close regarding duplex stability to a

non-copper homo-Pz pair (41°C), suggesting that in the absence of Cu^{2+}, the spatial pattern of Pz-A and Pz-Pz are similar to some extent; the two bases intract with each other in the duplex.

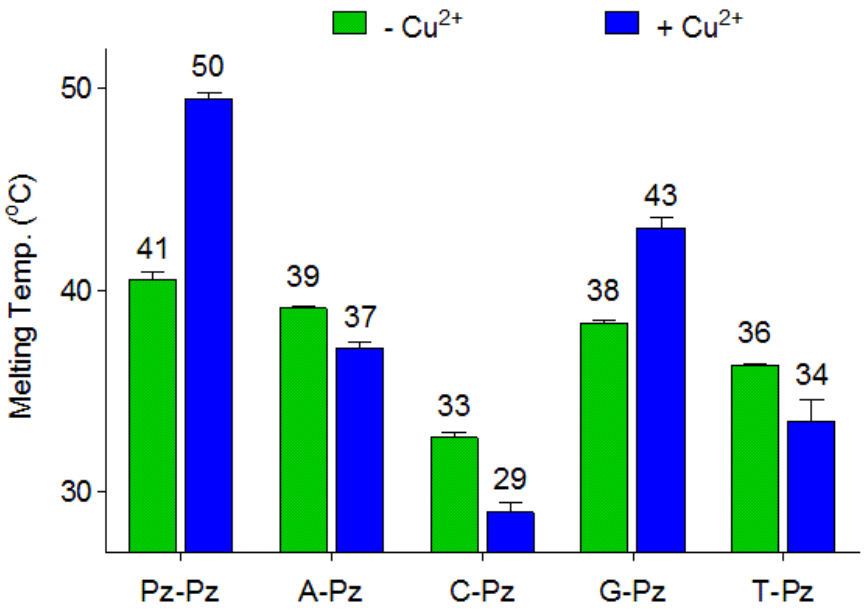


Figure 4-12 Plot of melting temperatures of **1a** combined with different counter strands containing Pz (**1b**), A (**1c**), C (**1d**), G (**1e**), and T (**1f**) as the counter base. Conditions: 150 mM NaCl, 1 mM CHES buffer pH 9.0, 1 µM oligonucleotide, with and without 1 µM Cu^{2+}, the final volume of 200 µL. Data are mean ± SD of three replicates.

Adding Cu^{2+} to these mismatched pairs stabilizes the duplex slightly. G-Cu^{2+}-Pz becomes the most stable pair (43°C) next to the Pz-Cu^{2+}-Pz pair, implying that some Cu^{2+}-mediated interaction between Pz and guanine occurs.

4.3.1.4. Cocrystallization of modified duplex and Bst Pol I

To investigate how pyrazole bases interact within each other in a DNA duplex, we crystallized the modified duplex **4a/b** together with polymerase Bst Pol I as a matrix. The Pz-Pz base pair was placed outside the protein binding site in the duplex forming regime. Cocrystalls were obtained using the hanging-drop vapor diffusion method in 4-morpholinoethanesulfonic acid (MES) aq. The crystal structure, as shown in Figure 4-13, shows that within in the duplex two Pz bases are stacked on top of each other with the oxygen and the nitrogen atom of the ligand pointing in the same direction.

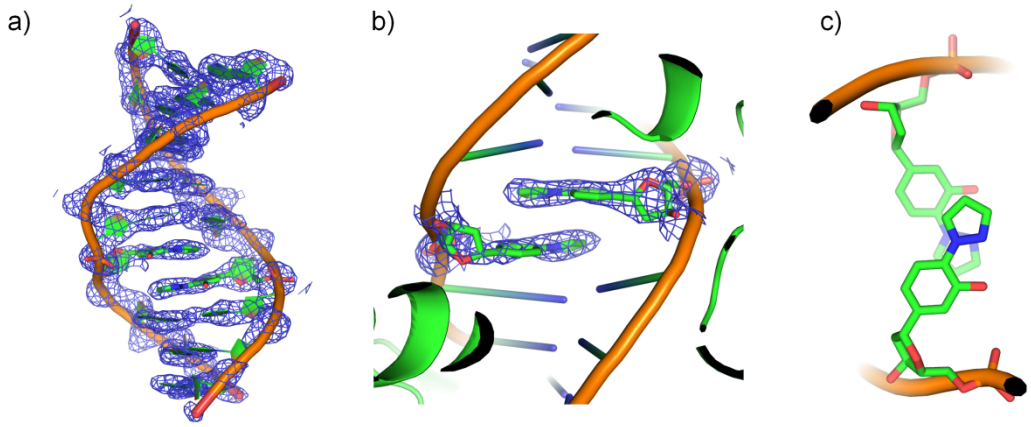


Figure 4-13 Cocrystal structures of duplex **4a/b** with Bst Pol I: a) complete crystal structure; b) side view of the local omit map; c) top view of local omit map.

Apparently, without the metal ion, the ligandoside is not arranged in a planar base pairing pattern but prefers a stacking interaction similar to the hydrophobic artificial base pairs from Leumann, Romesberg, and others.[127, 168, 169, 304] The stacking of the Pz ligandosides also disrupts the neighboring base pairs. The base pair is no longer coplanar but tilts and shifts out of the helix axis with only one inter-base hydrogen bond.

Cocrystal of **4a/b** with Cu^{2+} ions in the matrix of Bst Pol I was not obtained.

4.3.2. Oligonucleotides with multi-inserted ligandosides

4.3.2.1. Insertion of two ligandosides

In duplex **5a/5b**, two Pz pairs are placed in close proximity. In the absence of a copper ion, the T_M is 11°C lower than in the canonical strands (38°C vs. 49°C, Figure 4-14). When 1 eq. of Cu^{2+} is incorporated, the curve presents a two-phase S shape; the melting point of the first phase is 38°C, which is similar to the value without the metal ion. The second phase shows an inflection point at 70°C. Thus, the melting process is divided into two parts and Cu^{2+} is dissociated in the second stage. Melting profiles for further addition of Cu^{2+} display the normal S-shape curves. Both melting points are now at 70°C, clearly showing that now all Cu^{2+} binding sites are occupied.

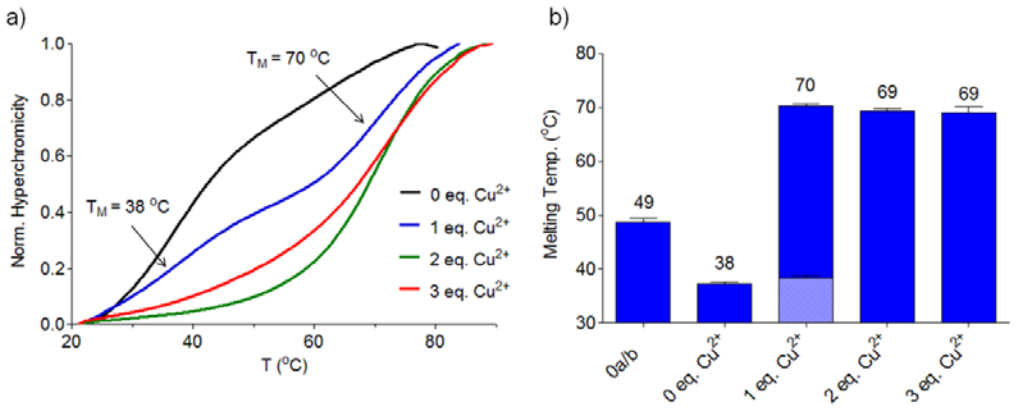


Figure 4-14 Melting profiles of duplex **5a/b**: a) melting curves of **5a/b** with 0-2 eq. Cu^{2+}, arrows show two inflection points; b) plot of melting temperatures of **5a/b** with 0-2 eq. Cu^{2+}, compared with **0a/b**. Data are mean ± SD of three replicates. Conditions: 150 mM NaCl, 10 mM CHES buffer pH 9.0, 1 µM oligonucleotides, the final volume of 200 µL.

The result indicates the presence of two well-defined species in solution. With 1 eq. Cu^{2+} half of the duplexes contain two metal ions while the other duplexes feature no metal ion. We concluded that Cu^{2+} complexes within the Pz ligandoside in a cooperative manner. This is in agreement with the observation of the isosbestic points at 246 nm in the UV and 256 nm in CD titration experiments (Figure 4-15). The isosbestic point is a specific wavelength, where the total absorbance of a sample does not change during complexation, indicating that only two species that vary in concentration contribute to the absorption, i.e. duplex **5a**/**b** without Cu^{2+} ion and **5a**/**b** with two Cu^{2+} ions.

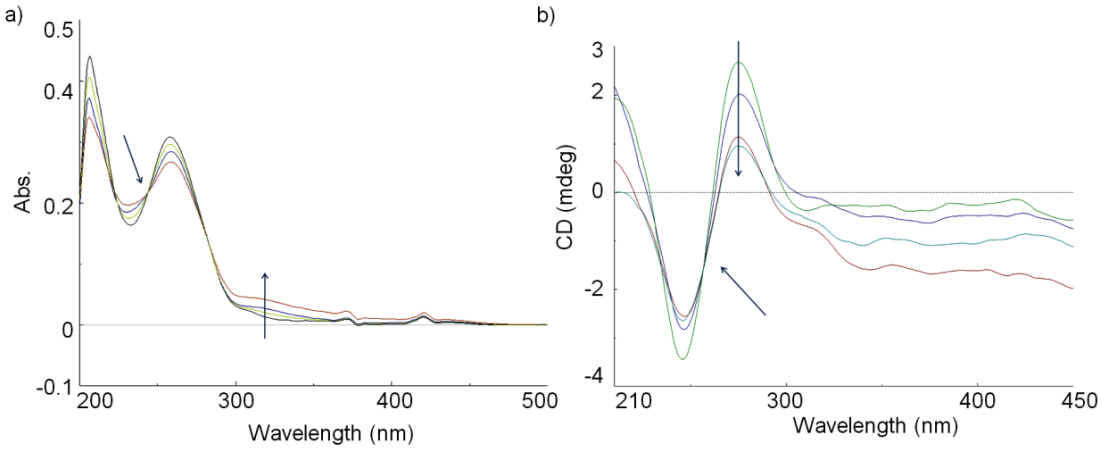


Figure 4-15 Overlaid UV (a) and CD (b) spectrum of duplex **5a/b** at various concentrations of Cu^{2+} (from 0 to 3 eq., step of 1 eq.). Conditions: 150 mM NaCl, 10 mM CHES buffer pH 9.0, 1 µM oligonucleotides, the final volume of 200 µL.

4.3.2.2. Insertion of consecutive five and ten ligandosides

Stacking metal ions in a duplex will provide the complex more property and function in constructing nanoobjects and nanocatalysts. Duplex **6a/b** with five and **7a/b** with ten consecutive Pz ligandoside pairs were prepared (Table 4-1). The two duplexes possess GC-rich wings at both termini for the interest of the duplexes thermostability; despite five consecutive ligandosides these wings stabilize the duplex **6a/b** with melting temperature of 61°C. After incorporation of copper ions, the melting temperature was so high that denaturing experiments were no longer possible.

CD spectra of copper titration with **7a/b** showed a clear transfer from 0 eq. to 5 eq. Cu^{2+} (Figure 4-16). After 5 equivalents Cu^{2+}, further titration did not change the signal at 242 nm. An isosbestic point at 316 nm is observed in the overlaid titration curves, which is a sign for two different species obtained during titration. This result is in accord with data obtained for the salen complex[175] and with data of Shionoya's hydroxyl pyridine ligandoside.

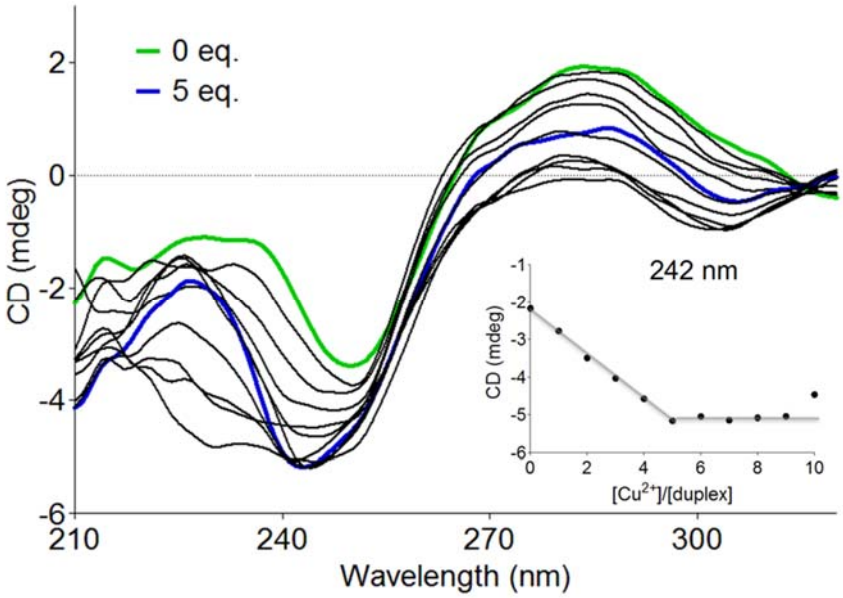


Figure 4-16 CD spectral changes of the duplex **6a/b** with Cu^{2+} from 0 to 10 eq., step of 1 eq., spectra of 0 eq. is shown in green and 5 eq. in blue. Inset: plot of circular dichroic changes at 242 nm against the ratio of [Cu^{2+}]/[**6a/b**]. Conditions: 150 mM NaCl, 10 mM CHES buffer pH 9.0, 1 µM oligonucleotide, the final volume of 200 µL.

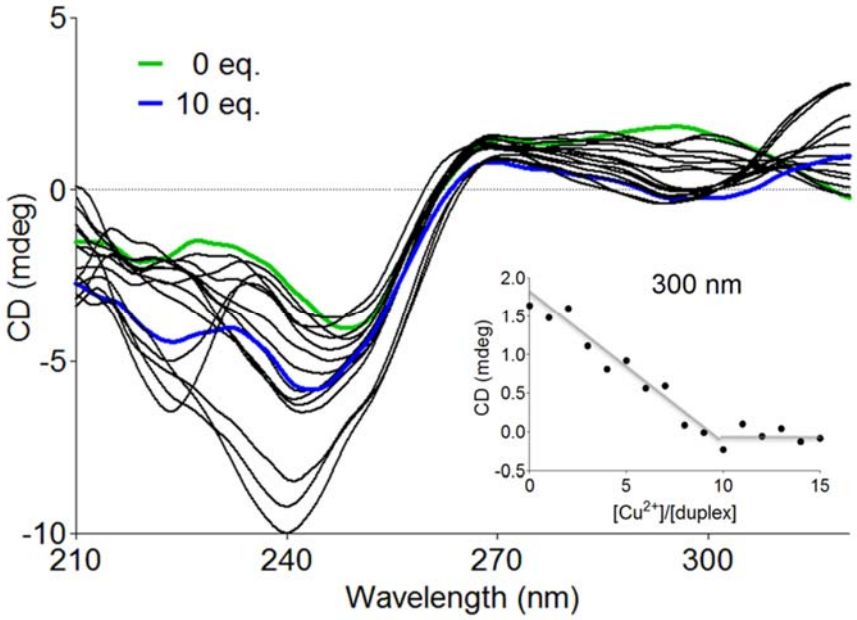


Figure 4-17 CD spectral changes of the duplex **7a/b** with Cu^{2+} from 0 to 15 eq., step of 1 eq., spectra of 0 eq. is shown in green and 10 eq. in blue. Inset: plot of circular dichroic changes at 300 nm against the ratio of [Cu^{2+}]/[**7a/b**]. Conditions: 150 mM NaCl, 10 mM CHES buffer pH 9.0, 1 µM oligonucleotide, the final volume of 200 µL.

For duplex **7a/b**, the overlaid CD spectra revealed that the structure of the duplex changes significantly upon Cu^{2+} complexation (Figure 4-17). The ellipticity at 300 nm shows a linear decrease which then comes to a constant after titration of 10 equivalents Cu^{2+}. An authentic isosbestic point is not present.

The results may be explained by the structural flexibility. Stacked Pz pairs are able to complex up to 10 metal ions, but then join in hand with structural changes of the duplex.

4.3.2.3. Insertion of mixed ligandosides: kinetics study

Modified duplexes displayed more intriguing properties when the pyrazole basepair was mixed with salen basepairs. **8a/b** contains two salen (S) pairs and one Pz pair. Duplex **9a/b** has two S and one Pm pair.

Copper ion titrations of duplexes **8a/b** and **9a/b** up to 4 equivalents exhibit remarkable featured on UV-Vis spectrum. For **8a/b**, the absorbances at 235 nm and 360 nm remain stable during the first equivalents of copper, then the absorbance increases for the next

two equivalents and become stable again after adding the third equivalent of copper (Figure 4-18a, b). Because S-Cu^{2+}-S complex has a specific wavelength at 360 nm,[174] and pyrazole pair absorbance is not responsive upon copper complexation at 360 nm, the result suggests that the copper ions first incorporate into the pyrazole base pairs and then inserting into the salen pairs taken place.

UV-Vis titration curves of duplex **9a/b** additionally support this hypothesis (Figure 4-18c, d). Because the Pm base pair does not chelate Cu^{2+}, the first two equivalents of copper ions must go directly into the salen pairs to form S-Cu^{2+}-S. Accordingly, the titration spectra shows a positive slope from 0 to 2 equivalents for the absorbance at 360 nm, which increases for about 0.04 abs., in accord with the observed increase in duplex **8a/b**. The curve then flattens out.

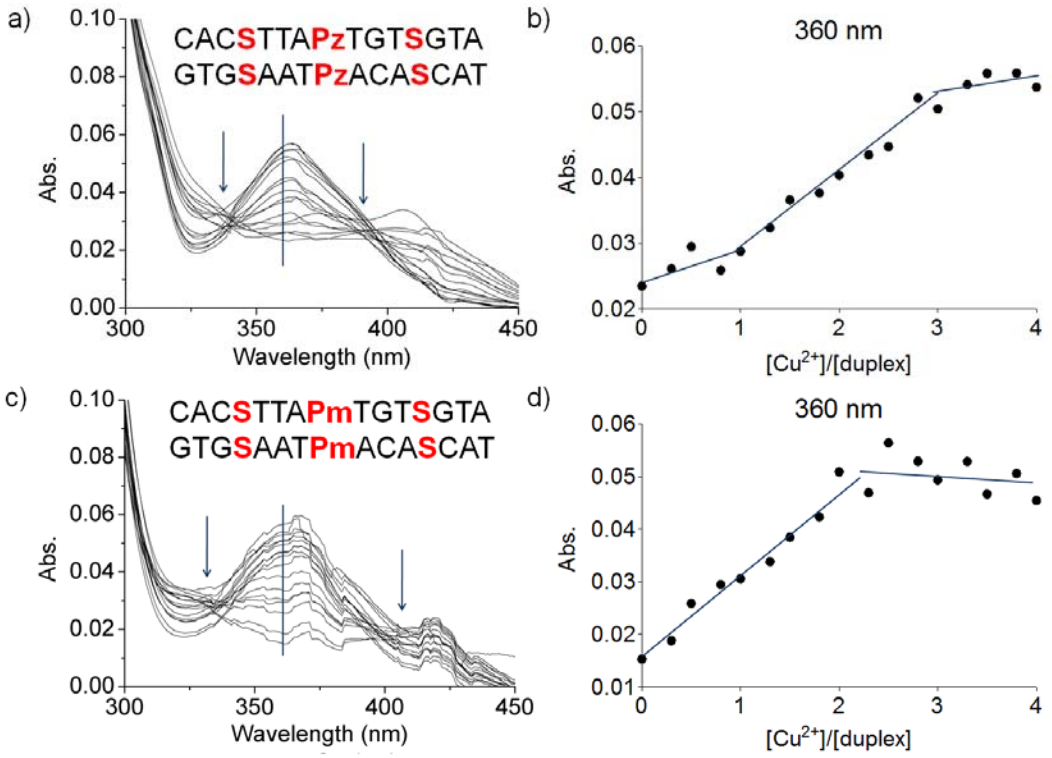


Figure 4-18 UV titration experiment of duplex **8a/b** and **9a/b**: a) Overlaid UV spectrum with different concentrations of Cu^{2+} (from 0 to 4 eq.) of the duplex **8a/b**; b) Plot of UV absorbance at 360 nm against the ration of [Cu^{2+}]/[**8a/b**]; c) Overlaid UV spectrum with different concentrations of Cu^{2+} (from 0 to 4 eq.) of the duplex **9a/b**; d) Plot of UV absorbance at 360 nm against the ration of [Cu^{2+}]/[**9a/b**]. Conditions: 150 mM NaCl, 10 mM CHES buffer pH 9.0, 3 µM oligonucleotides, 30 eq. ethylene-diamine, the final volume of 200 µL.

It is worth emphasizing that the absolute absorbance increases from 0.02 to 0.06, which is low in respect of what can be safely detected range. Despite the low signal to noise ratio, the general trend in the titration is evident.

A more complicated case was found in duplex **10a/b**. This duplex contains 3 pyrazole and 4 salen pairs. The complexation of copper ions was confirmed by CD titration, showing incorporation of 7 equivalents of Cu^{2+} into the duplex (Figure 4-19a, b). Overlaid spectrum show two isosbestic points, i.e. at 246 nm and 265 nm. The plot of the absorbance at 360 nm shows a plateau at 4 equivalents of Cu^{2+} followed by a steep slope upon addition of the subsequent 4 equivalents. We believe that because of the crowded situation and the flexible nature of the complexing unit that not all Cu^{2+} go into the Pz pairs in a stoichiometric manner. The plot of absorbance at 405 nm (Figure 4-19c, d) presents two different trends of the absorbance changes. The first three equivalents of Cu^{2+} seem to cause rather sharp decrease in the absorbance while the following four equivalents yield a slight decrease of absorbance.

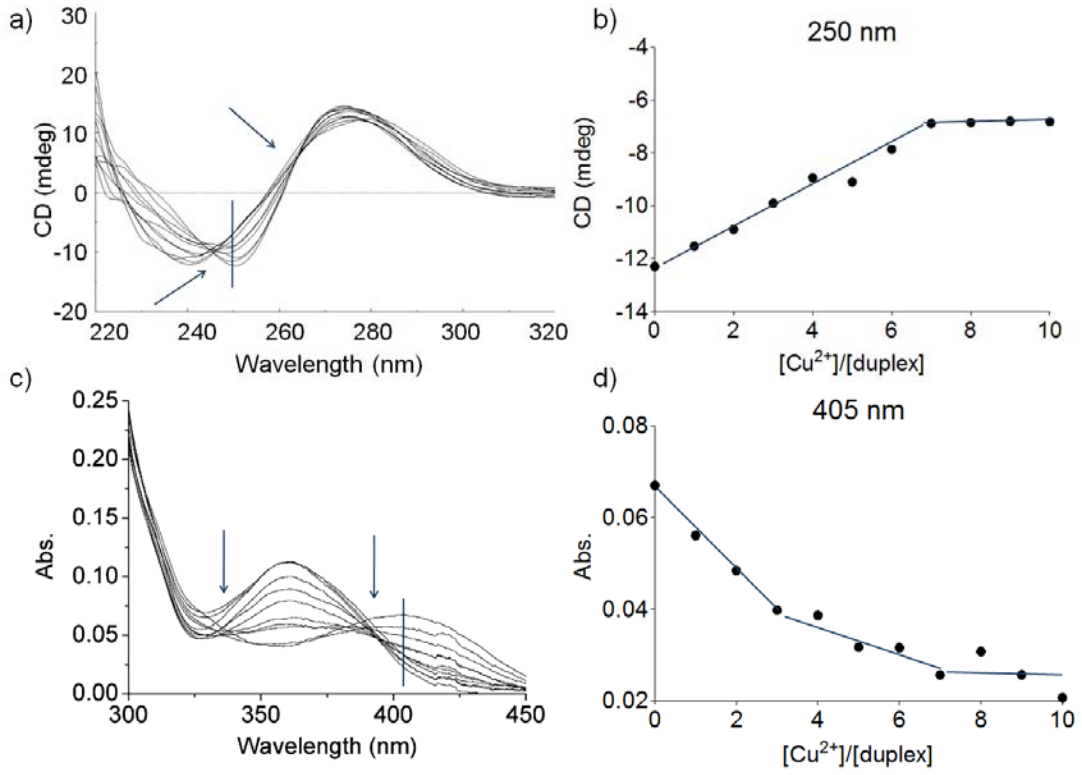


Figure 4-19 CD and UV titration experiment of duplex **10a/b**: a) Overlaid CD spectrum with different concentrations of Cu^{2+} (from 0 to 10 eq.) of the duplex **10a**/**b**; b) Plot of circular dichroic changes at 250 nm against the ration of [Cu^{2+}]/[**10a**/**b**]; c) Overlaid UV spectrum under the same conditions as in

a); d) Plot of UV absorbance at 405 nm against the ratio of [Cu^{2+}]/[**10a**/**b**]. Conditions: 150 mM NaCl, 10 mM CHES buffer pH 9.0, 3 μM oligonucleotides, 30 eq. ethylenediamine, the final volume of 200 μL.

It is important to note that the above samples were kept at 4°C and were measured at 20°C within a short period of time. When the sample stayed at room temperature longer than 1 h, the absorbance at 260 nm of duplex **10a/b** could rise even with less than 1 equivalent Cu^{2+} ion. This indicates that the copper ions jump from Pz-Pz pair to the Sa-Sa pair inside the DNA duplex at higher temperature. The instability is meanwhile evidence for the thermodynamic preference for the S-Cu^{2+}-S complex.

With all this evidence in hand, it is safe to conclude that there is a kinetic preference for copper coordination with the pyrazole pair at 4°C compared to the salen base pair. This conclusion could be the foundation of making logic gates using bridged and non-bridged ligandoside.

4.3.2.4. Insertion of mixed ligandosides: thermodynamic study

Intriguing results were obtained by ESI-Mass measurement. Although the ESI analysis of duplex **1a/b**, hybridized with Cu^{2+}, always presented two signals corresponding to the two single strands ODN **1a** and **1b**, desired mass peak of the duplex containing a copper ion inside could not be obtained, which suggests that the decomplexation of Pz-Cu^{2+}-Pz ↔ Pz-Pz is fast in the gas phase.

ESI measurement of S/Pz inserted duplex **8a/b** and **10a/b** were thereafter processed. Although 3 equivalents Cu^{2+} was incubated with duplex **8a/b** before ESI-MS measurement, the mass result showed signals corresponding to a duplex containing only two copper ions (Figure 4-20).

We reasoned that the ethylenediamine linkage of the salen base pairs make the complex so stable that the Cu^{2+} complex survives during the ionization while complexes with the Pz base pair are kinetically less stable.

Being curious in this unique property and considering the possibility of metal ion exchange after the coordination, we designed the following experiment. Duplex **8a/b**

was first incubated with ethylenediamine and 3 equivalents of Cu^{2+}; then another 2 equivalents of Mn^{2+} was added. ESI-Mass gave the main peak corresponding to the dsDNA+2Cu^{2+}. Sub-peaks corresponding to dsDNA+Cu^{2+}+Mn^{2+} were discovered along with the main peak (Figure 4-21).

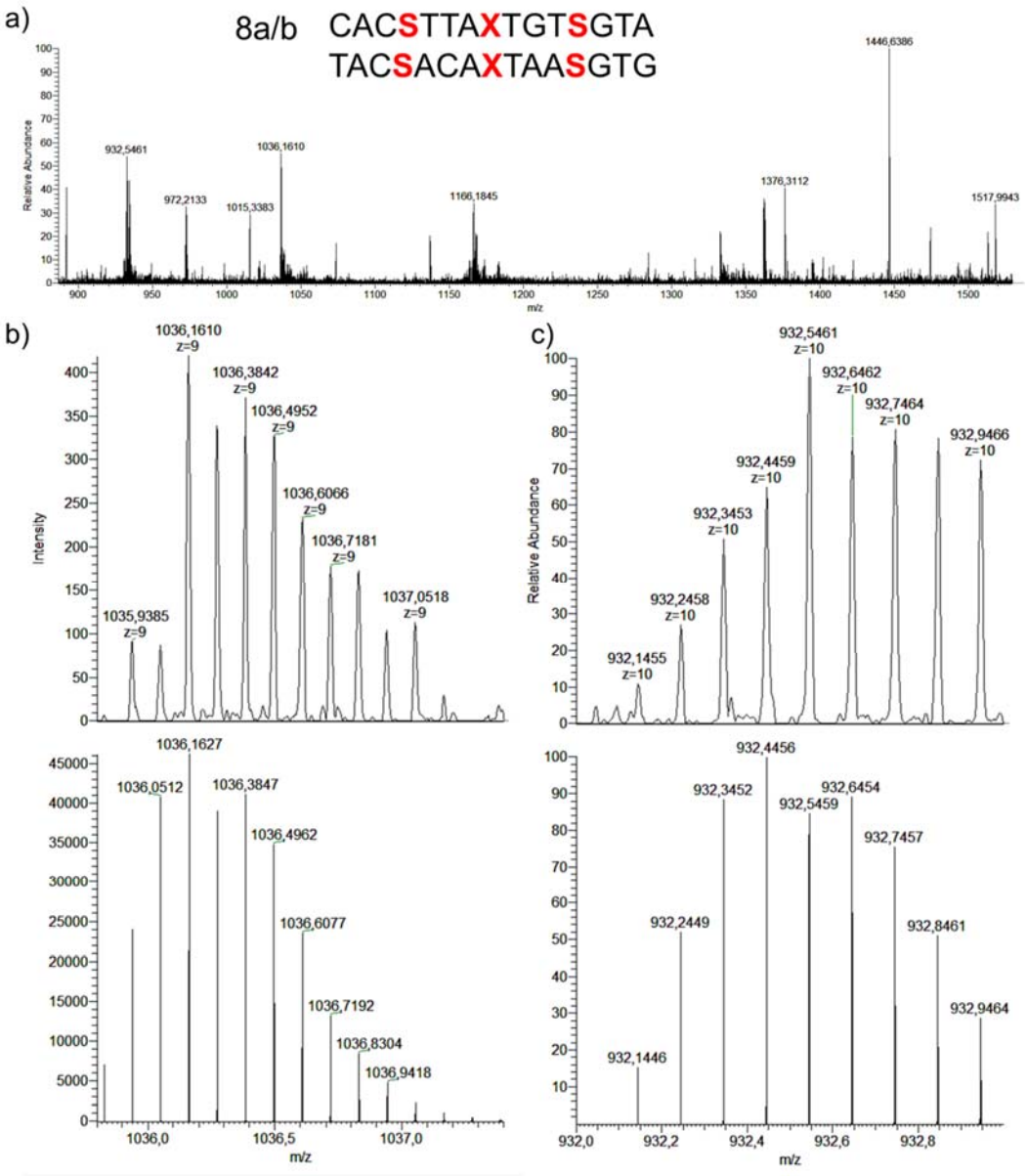


Figure 4-20 Experimental ESI-Mass spectrum data compared with calculated molecular weights of duplex **8a**/**b** with 2 eq. Cu^{2+}: a) Overall ESI spectrum; b) Peaks corresponding to duplex contains 9 charges (top) compared with calculation (bottom), $C_{316}H_{375}O_{176}N_{96}P_{28}Cu_2^{9-}$, [M-9H$^+$], calc. 1036.1627, found 1036.1610; c) Peaks corresponding to duplex contains 10 charges (top) compared with calculation (bottom), $C_{316}H_{374}O_{176}N_{96}P_{28}Cu_2^{10-}$, [M-10H$^+$], calc. 932.4456, found 932.4459. Conditions: 150 mM NH_4OAc, 30 µM oligonucleotides, 30 eq. ethylenediamine.

The manganese ions were added after the reannealing process. The solution was not heated before the ESI measurement.

In contract, the melting profile of duplex **1a/b** incubated with 1 eq. Cu^{2+} and 1 eq. Mn^{2+} shows the same temperature as **1a/b** with 1 eq. Cu^{2+}. Thus, it is safe to claim that the exchange between S-Cu^{2+}-S + Mn^{3+} ↔ S-Mn^{3+}-S + Cu^{2+} can take place but only to a small extent, while the exchange of Pz-Cu^{2+}-Pz + Mn^{3+} ↔ Pz-Mn^{3+}-Pz + Cu^{2+} does not occur. This shows that S-S perfers complexation of Mn^{3+} while Pz-Pz perfers Cu^{2+}.

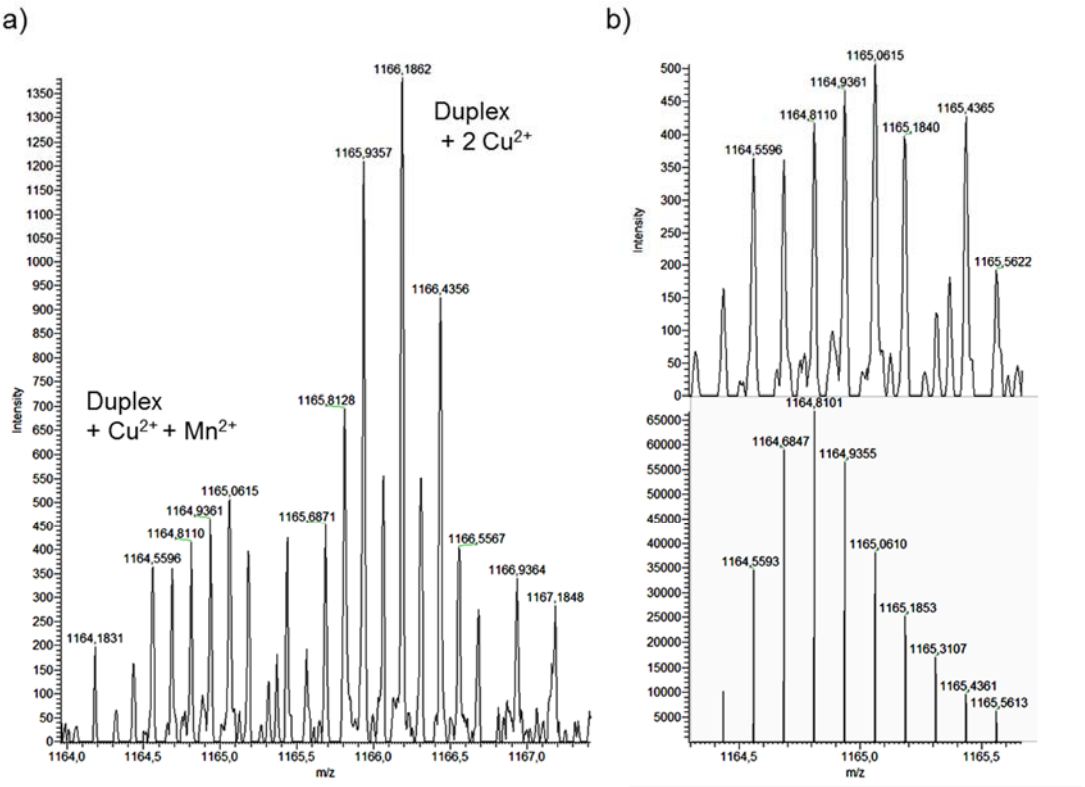


Figure 4-21 ESI-Mass spectrum and comparison of experimental data with calculated molecular weights of duplex **8a/b** with 3 eq. Cu^{2+} and 2 eq. Mn^{2+}: a) Peaks corresponding to **8a/b** + $2Cu^{2+}$ and **8a/b** + Cu^{2+} + Mn^{2+}; b) Peaks contain 8 charges (top) compared with calculation (bottom), $C_{316}H_{376}O_{176}N_{96}P_{28}CuMn^{8-}$, [M-8H^{+}], calc. 1165.0610, found 1165.0615. Conditions: 150 mM NH_4OAc, 30 µM oligonucleotides, 30 eq. ethylenediamine.

4.4. Application of the pyrazole ligandoside

4.4.1. Towards genetic code expansion

The intention to design ligandosides is to expand the genetic alphabet. Therefore, the following paragraphs will focus of ligandoside replication within a DNA template.

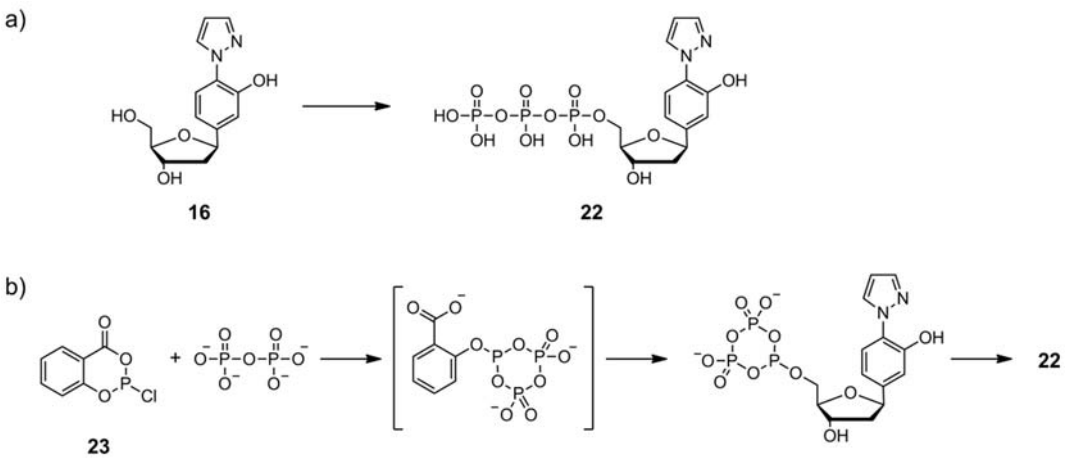
Despite being essential for signaling between cells and for the energy household of the cell, high concentrations of copper ions can be toxic to cells. Free copper ions generate reactive oxygen species such as superoxide O_2^{-}, hydrogen peroxide H_2O_2, and hydroxyl radical •OH, which damage proteins, lipids, and DNA.[305-307]

Based on these considerations, primer extension experiments were first executed with pyrazole ligandoside in the DNA template.

4.4.1.1. Synthesis of the replication components

To perform primer extension experiment one needs the following components: primer, template, deoxynucleoside triphosphates, DNA polymerase, a respective reaction buffer, and, if necessary, other factors. First of all, the pyrazole triphosphate was synthesized with a minor modified protocol,[308] as shown in Scheme 4-9. Further triphosphorylation methods were reviewed by K. Burgess and D. Cook.[309]

The one-pot synthesis of triphosphate is achieved by generating first a phosphorylation intermediate that reacts selectively with the 5'-hydroxyl groups of the unprotected nucleosides. In the case of pyrazole triphosphate, we found that the phenol group does not interfere with the process.



Scheme 4-9 Synthesis of triphosphate **22**: a) Condition: 1) bis-tetrabutylammonium pyrophosphate, tributylamine, 2-chloro-1,3,2-benzodioxaphosphorin-4-one **23**, DMF; 2) I_2, pyridine, H_2O, RP-HPLC purification, 11%. b) Mechanism of the triphosphate addition.

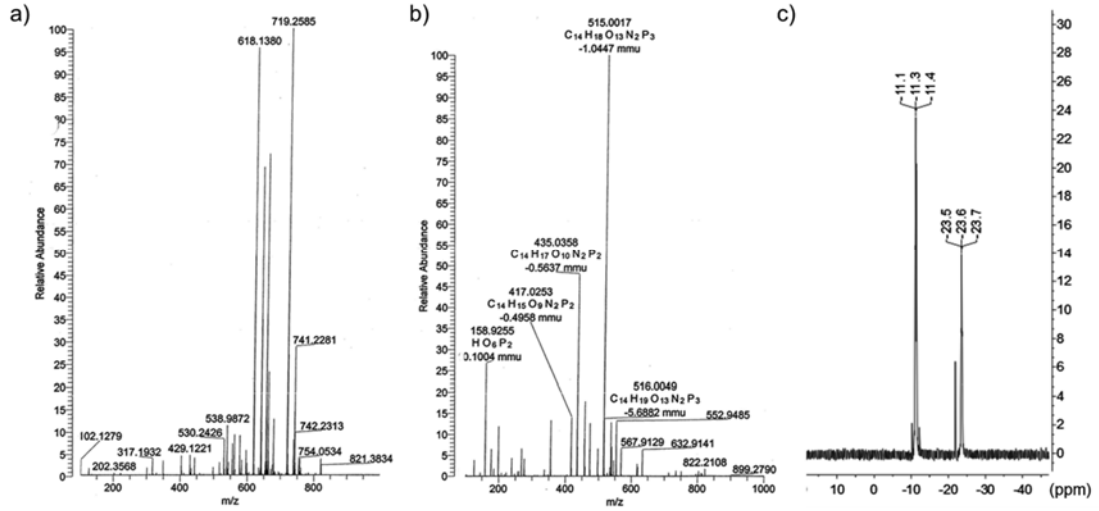


Figure 4-22 Characterization of triphosphate **22**: a) ESI+, $C_{20}H_{35}N_3O_{13}P_3^+$ [M+H$^+$+TEA], calc. 618.1383, found 618.1380, $C_{26}H_{50}N_4O_{13}P_3^+$ [M+H$^+$+2TEA], calc. 719.2587, found 719.2585; and b) ESI- of **22**, $C_{14}H_{18}N_2O_{13}P_3^-$ [M-H$^+$], calc. 515.0022, found 515.0017; c) ^{31}P NMR of **22**.

After HPLC purification, the triphosphate product was characterized by ^{31}P-NMR and ESI-Mass spectrum. ESI-Mass in the positive charge mode showed two high peaks at 618.1380 and 719.2585 (Figure 4-22a), which equal to the mass of the triphosphate plus one and two molecules of triethylamine (calc. 618.1383 and 719.2587), while ESI-Mass in the negative detection mode gave the exact triphosphate molecular weight of 515.0017 (calc. 515.0022, Figure 4-22b). ^{31}P-NMR showed two set of peaks, one set of peaks at around -11.3 ppm, and a set of the triplet at -23.6 ppm (Figure 4-22c). α- and γ- phosphorus atoms were crowded into the low field, leaving the β-phosphorus atom at

high field. Thus, we affirmed that the pyrazole triphosphate was obtained.

As a template, a 30 mer oligonucleotide (**ODN 11a**) with Pz at the 24th position from 3' terminus was synthesized using solid-phase synthesis (Table 4-1). A 5'-fluorescein-labeled 23 mer oligonucleotide was designed as a primer. Certainly, the primer can be labeled with radioactive ^{32}P using polynucleotide kinase for direct detection.

4.4.1.2. Replication with pyrazole triphosphate

With all the three components in hand, primer extension experiments were performed. First, a collection of polymerases were screened to find out which allowed the incorporation of dPzTP (**22**) in the 24th position of the primer. Taq, Q5, Pwo, Phusion, T4, T7, Φ29, KF (exo$^-$), Bst Pol large fragment, Deep Vent and Therminator were tried. In the reaction system, an excessive amount of template (2 μM) and primer (1.5 μM) as well as 100-fold excess of **22** (200 μM) was employed. Experiment with and without Cu^{2+} (200 μM) at both 37°C and 50°C were performed.

Deep vent (exo-) polymerase and Therminator polymerase, which belong to B family polymerase and which are isolated from archaea, are able to incorporate dPzTP with the primer opposite a Pz base on the template. Deep Vent achieved the goal at 50°C in 24 h, while Therminator DNA Polymerase succeeded in 2 h at 37°C or in 5 min at 75°C. It found out to be the best polymerase able to incorporate the ligandoside **22**. Considering the incorporation efficiency and temperature consistency, 75°C was fixed as the reaction temperature for all further investigations.

Therminator polymerase is a mutant variant of the 9^{o}N exo$^-$ polymerase (Thermococcus species 9^{o}N-7), in which the alanine 485 has been replaced by a leucine residue. This provides the polymerase with an enhanced ability to incorporate modified substrates, e.g. dideoxynucleotides, ribonucleotides, and acyclic-nucleotides.[310-313]

Next, we studied the **dPzTP** incorporation efficiency under different conditions. At 75°C within 5 min, the incorporation was achieved to a full extend, independent of

present of copper or magnesium ions (Figure 4-23).

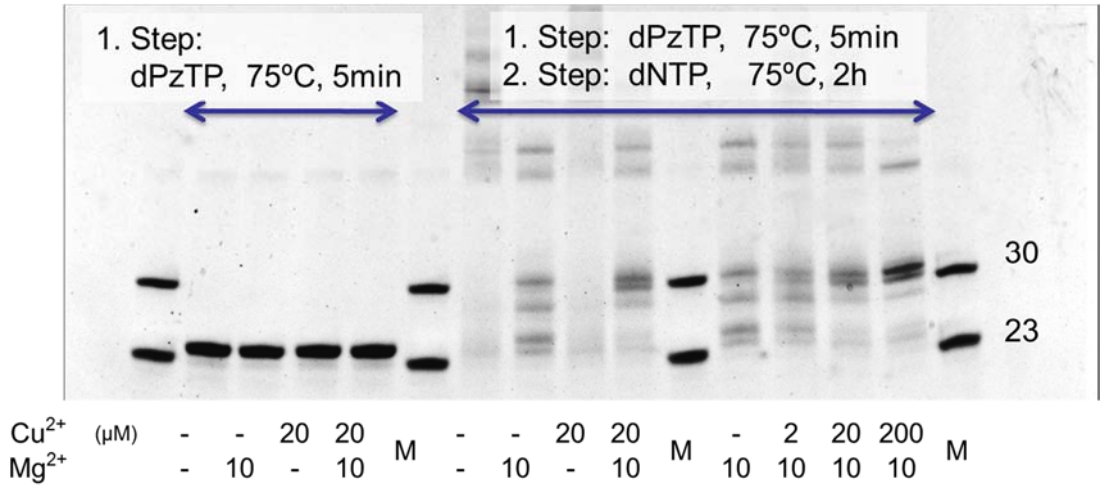


Figure 4-23 PAGE assay shows dPzTP incorporation and extension. Condition: 2 μM template **11a**, 1.5 μM primer **11b**, 200 μM dPzTP **22**, 200 μM each dNTP, 1 U Therminator, 1× Thermopol buffer, 10 mM $MgSO_4$, the final volume of 20 μL.

The extension towards full length seemed more challenging than the dPzTP incorporation. For polymerase Therminator, at 50°C, the yield of full length was less than 50%. When the reaction temperature was raised to 60°C with the additive Cu^{2+} (200 μM) or Mg^{2+} (10 μM), the yield reached a maximum value of 65% in 9 h and then further decayed. At 75°C, the reaction could be finished in 2 h with additional Cu^{2+} and Mg^{2+} (Figure 4-23). The co-enzyme Mg^{2+} is a requisite for the extension, especially at the high temperature. Extension experiments with variable Cu^{2+} concentration showed that 200 μM of copper ions give the best condition.

Further studies showed that the yield reached 80% in 90 min (Figure 4-24). Due to the dissatisfied efficiency, PCR amplification is not practical because the failed synthesis will be accumulated exponentially in PCR cycles.

After PAGE analysis, the corresponding gel band was cut out and the DNA contatining material was dissolved and desalted. MALDI-TOF analysisi gave a clear signal (Figure 4-27), corresponding to the 31 mer without the fluorescein tag but with a Pz base at the 3'-overhang. We speculated this is due to the lack of a 3'-5' exonuclease activity of the polymerase. We think hat the loss of 5'-fluorescein is happening during work up.

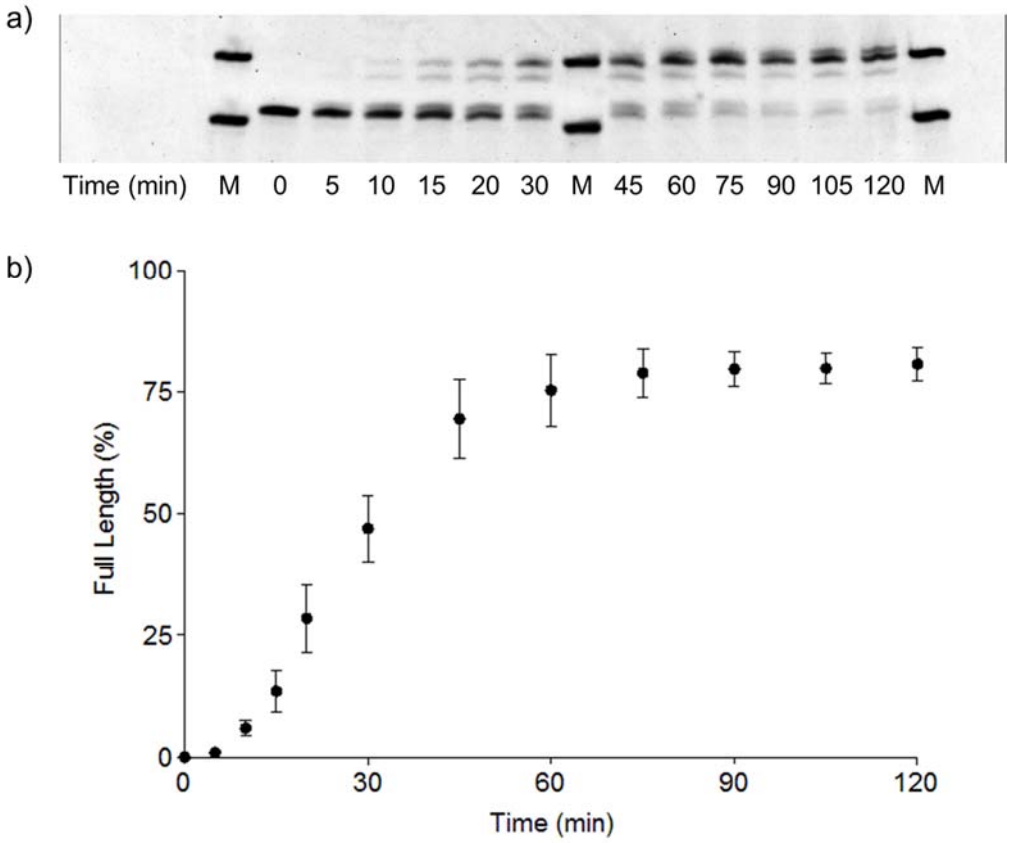


Figure 4-24 Time dependent study of dPzTP extension: a) PAGE assay shows the primer extends to full length in 2 h at 75°C after dPzTP full incorporated in 5 min at 75°C; b) Plot of the percentage of the full-length product by the integration from PAGE assay. Data are mean ± SD of three replicates. Condition: 2 μM template **11a**, 1.5 μM primer **11b** extended with **22**, 200 μM each dNTP, 1 U Therminator, 1× Thermopol buffer, 10 mM $MgSO_4$, the final volume of 20 μL.

To study if the Pz base was orthogonal to the natural bases, we added four natural triphosphates to Pz-containing template **ODN 11a**, and reversely, dPzTP to four different templates with A/C/G/T at the 24th position (**ODN 11c/d/e/f**).

At 75°C for 5 min, all the four natural bases were incorporated into the primer complementary to the Pz. However, no 25 mer product was visible showing again that elongation was blocked. dPzTP could also be complementarily incorporated opposite the natural bases on the template. Extension afterward was not discernible either (Figure 4-25), which could be good news for the plasmid replication experiments.

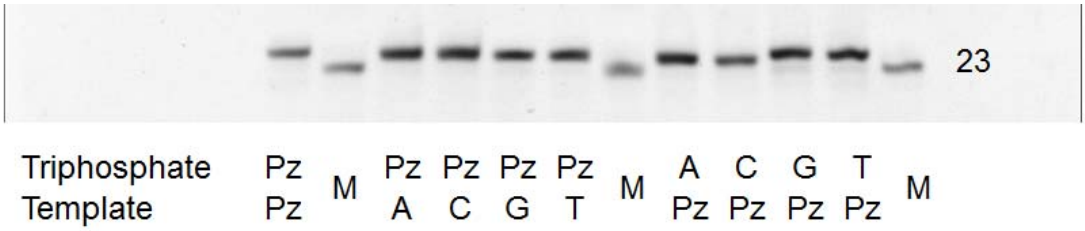


Figure 4-25 PAGE assay shows incorporation of dPzTP on **11c/d/e/f** with primer **11b** and incorporation of dATP/dCTP/dGTP/dTTP with template **11a** on primer **11b**. Condition: 2 μM template, 1.5 μM primer **11b**, 200 μM dPzTP or dNTP, 1 U Therminator, 1× Thermopol buffer, 10 mM $MgSO_4$, the final volume of 20 μL.

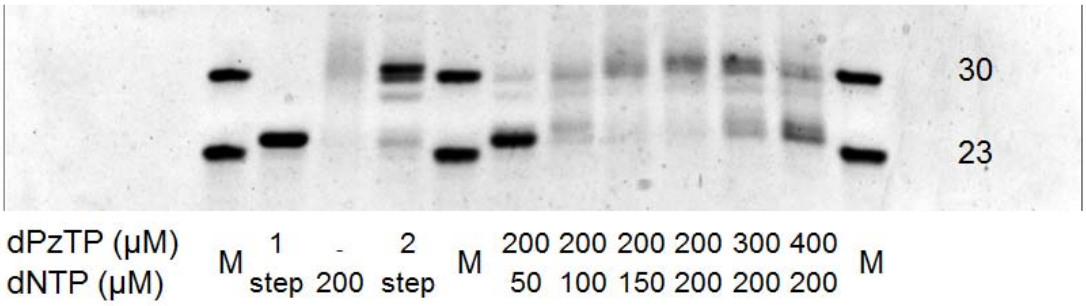


Figure 4-26 PAGE assay shows incorporation of dPzTP mixed with dNTP extend with template **11a** on primer **11b**. Condition: 2 μM template, 1.5 μM primer **11b**, 200 μM dPzTP and dNTP, 1 U Therminator, 1× Thermopol buffer, 10 mM $MgSO_4$, the final volume of 20 μL.

In all furture studies we fixed the reaction conditions (75°C, 2 h, 1U Therminator DNA polymerase, 200 μM Cu^{2+} and additive Mg^{2+}) constant and investigate the effect of the different ratio of dPzTP and dNTP was investigated (Figure 4-26). The best result were obtained of a ratio of 1/1, i.e. 200 μM dPzTP, and 200 μM dNTP. Although the primary product was a 31 mer oligonucleotide, the yield was lower than when it was done in two separate steps. MALDI-TOF failed to give a signal after analysis of band cut out of the the PAGE gel. To solve the problem, we cut the gel and extracted the DNA with phenol and chloroform, desalted, and measured again. The spectrum, shown in Figure 4-27, indicated formation of the same product formed in two steps.

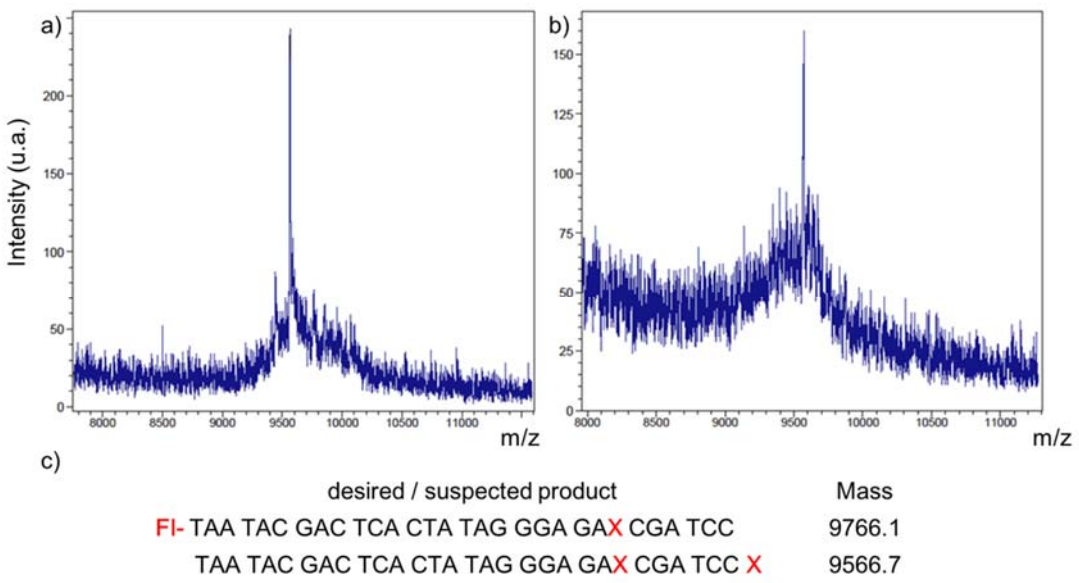


Figure 4-27 MALDI-TOF spectrum for a) extended primer in two steps, found 9567.3; b) extended primer in one pot, found 9569.5; calc. 9566.7, corresponding to the complementary template strand with Pz at 3'-overhang without 5'-fluorescein, listed in c).

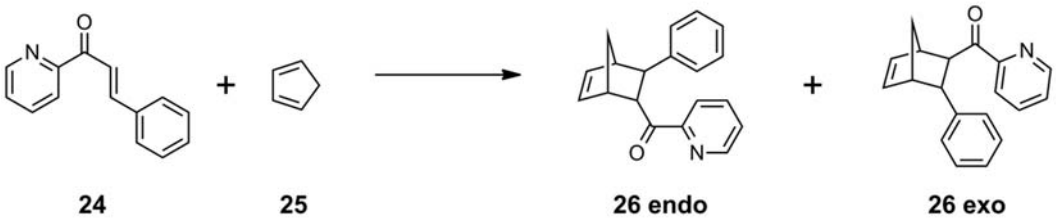
In summary, we showed the possibility to incorporate the dPzTP into a primer opposite a templating dPz base using the Therminator polymerase. Even for such a powerful polymerase, neither the efficiency nor the fidelity was sufficient to continue with PCR experiments.

4.4.2. Towards chiral catalyst

Ligandosides make it practicable to complex copper ions or other metal ions inside the DNA double helix. This promoted us to think of applying the complex as water soluble or solid-phase based chiral catalysts. Whether the duplex could transfer its helicity to the substrates and whether the metal ion inside the duplex could approach the substrates to act as a catalytic center are worthwhile and intriguing questions to investigate.

Previously, two strategies using the nucleic acid structure as a chiral catalyst were developed. The first example used a copper ligand with a linker that could suspend the ligand outside the natural DNA duplex.[314] This strategy was further improved as an interstrand linker hanging copper ion outside the duplex or inside the duplex.[315, 316] In the second strategy, a G-quadruplex with a copper ion in the center was

employed.[317] The complex showed excellent stereoselectivity in a model Diels-Alder reaction. The parallel and anti-parallel conformations made a difference on the stereoisomerism.



Scheme 4-10 Model Diels-Alder reaction for testing the catalytic effect of duplex **6a/b**. Condition: 0.5 mol% **6a/b**, 2.5 mol% Cu^{2+}, 50 mM NaCl, 20 mM buffer, 4°C, 24h.

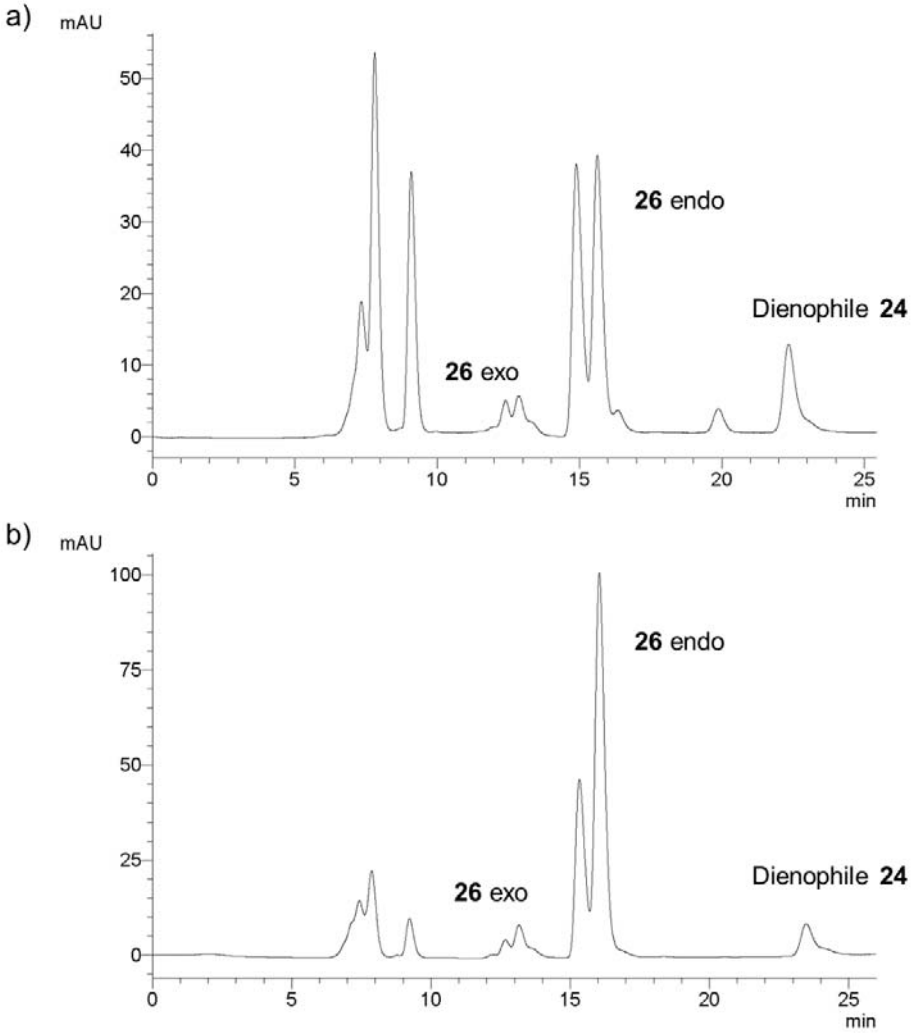


Figure 4-28 Typical HPLC traces of the crude products from Cu^{2+} catalyzed Diels-Alder reaction. a) Cu^{2+} without duplex **6a**/**b**, negative control, entry 2 in Table 2. b) Cu^{2+} with duplex **6a**/**b**, entry 1 in Table 4-3. Condition: heptane/2-propanol 99:1, 0.5 mL/min, column temperature 40 °C.

Here, duplex DNA with pyrazole ligandoside was attempted in the same model Diels-Alder reaction with aza-chalcone (**24**) and cyclopentadiene (**25**) as reported (Scheme 4-10).[314] To reach a relatively high concentration of copper ions in the system, a duplex catalyst with five continuously stacking pyrazole bases (duplex **6a/b**) was applied. At high temperature, the Diels-Alder reaction yields the predominantly the exo product, while at lower temperature needed to ensure the present of the intact DNA complex, the endo isomer turns out to be the primary product. Both the exo and endo products contain a pair of enantiomers (Figure 4-28). In the absence of the catalyst, acidic environment alone accelerates the reaction.[318]

The Cu^{2+} ion was first coordinated inside the duplex in basic ammonium acetate buffer, an evaporable mild alkaline solution, to ensure the complex formation. After evaporating to dryness, the residual material was redissolved in the corresponding reaction buffer. The reaction mixture was subjected to normal phase chiral HPLC to determine the enantiomeric excess. A standard curve was made to calculate the molar ratio of the product **26** and starting dienophile **24** (Figure 4-29).

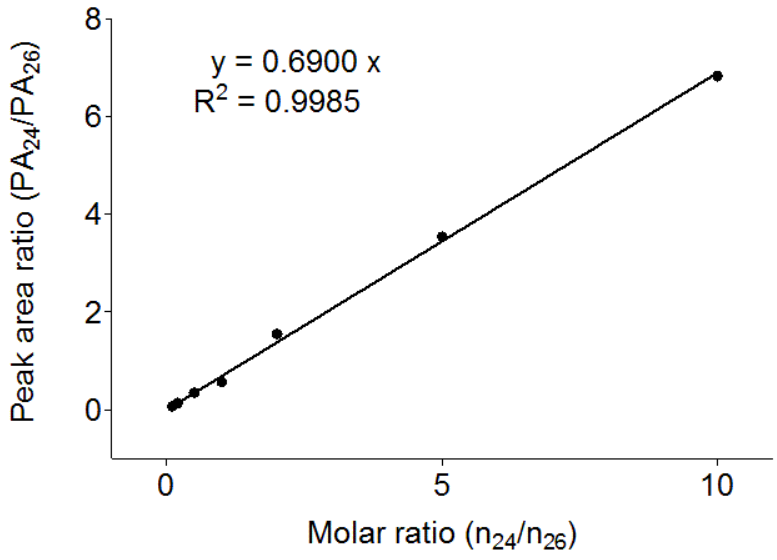


Figure 4-29 Determination of the correction factor. The HPLC ratio of peak area (PA_{24}/PA_{26}) were determined with the standard molar ratio (n_{24}/n_{26}) of 1/10, 1/5, 1/2, 1, 2, 5 and 10. The correction factor (f = 0.69) was estimated from the fitting curve (R^2 = 0.9985).

The results are summarized in Table 4-3. Both single and double strands with the Pz-ligands accelerated the reaction and induced chirality in the presence of Cu^{2+}.

Copper counter ions and surrounding pH also plays a role in the catalytic process.

Table 4-3 Results of the catalytic Diels-Alder reaction with duplex **6a/b**.

entry	ligand	Cu^{2+}	pH	ee	endo/exo	conv.
1	ds	NO_3^-	6.5	39	93	97%
2	ds	SO_4^{2-}	6.5	32	87	52%
3	ds	OTf^-	6.5	29	84	46%
4	ds	-	6.5	3	76	19%
5	-	NO_3^-	6.5	1	87	76%
6	ss	NO_3^-	6.5	29	93	96%
7	ds	NO_3^-	7.4	36	86	79%
8	ds	NO_3^-	9.0	23	80	47%

As generally believed, the metal ions inside the duplex could hardly access the substrate and serve as a catalytic center. The results here prove that the liganoside-containing DNA strand can act as a chiral catalyst. Further optimization of the ee values is now required as well as studies on the relationship between duplex sequence and catalytic effect and expansion of the substrate scope and reactions. In case that the ligandoside can also coordinate metal ions like palladium, gold or ruthenium this would open up new opportunity for recyclable noble metal chiral catalysts.

4.5. Summary: Part I

Concerning the reconstruction of oligonucleotides, this chapter focuses on the controlled binding of multiple metal ions inside the DNA duplex. Herein, a non-bridged pyrazole ligandoside is designed, synthesized and evaluated.

The organic synthesis started with **3** to furnish the ligand building block **6**. After connecting to the ribose derivate **8**, deprotection and protection gave the monomer **15** for solid-phase synthesis. In the solid-phase synthesis, the pyrazole ligandoside phosphoramidite coupled as efficiently as canonical nucleosides.

Among the metal ions that have been tested, Cu^{2+} and Mn^{3+} stabilized duplex **ODN 1a/b** containing a homo pair of pyrazole ligandoside equal to canonical duplex. The system maintains the B-form helix conformation. Pyrazole base showed self-recognition behavior orthogonal to natural bases in the present or absent of metal ions. A co-crystal structure of duplex **ODN 4a/b** and Bst Pol I was obtained, suggesting that, without metal ions, the pyrazole bases are stacking with each other instead of having a planar base pair type interaction.

When two pyrazole ligandosides were in neighboring places in the duplex, copper ions complexed cooperatively. The duplexes **ODN 6a/b** and **7a/b** were capable of arranging five and ten Cu^{2+} ions in the duplex. In the case of duplex **ODN 8a/b**, UV titration and ESI-MS suggests that the Pz-Pz pair perfers to complex Cu^{2+} even relative to the S-S pair.

With Thermopol polymerase, a pyrazole ligandoside triphosphate is placed opposite a dPz in the template and extended to full length. When applied as a catalyst, the metal-DNA complexation accelerated the Diels-Alder reaction and reached moderate stereoselectivity.

5. Part II: Oligonucleotide fdC probe

5.1. Aims of the project and rationale

The aim of the second project was to develop a novel squence specific quantification method for 5-formyl cytosine. Considering the disadvantages of the previous methods, i.e. reduce-bisulfite sequencing, and biotin-enriched sequencing, this method should be able to detect fdC content in a given gene position or loci. Sequencing depth can be alleviated; detection labor-hour and expense should be saved.

Detecting epigenetic bases is similar to detecting single nucleotide polymorphisms (SNP).[319] Most SNP detection strategies take advantage of amplification technologies like PCR,[320] or isothermal amplification,[321] to amplify and quantify the signal with probes.

However, epigenetic bases, i.e. mdC, hmdC, fdC, and cadC are similar to cytosine and pair with guanine. So, the discrimination among these bases will be impossible in the typical PCR reaction. As stated in Chapter 3.2.3.5, there are peptide[276] and oligonucleotides[277] to discriminate dC, mdC, and hmdC using inter-molecular interaction. Since they failed to amplify the sequence, these probes may not be suitable for genomic detection. Therefore, a covalent bond between the target strand and the probe strand must be established.

To tackle the problem, one idea was to design a complementary oligonucleotide containing a linker with a functional group that can react with fdC on the target strand. The reaction should be chemically specific with a favorable reaction rate. The hybridization of the probe strand with the target strand ensures a sequence-specific detection (Figure 5-1).

As fdC is found at levels below 0.002% of all cytosines in the genomic DNA, the method should be able to either amplify the fdC signal or detect the signal with an extremely high sensitivity. PCR, PAGE assay, antibody test are conventional methods

for gene sequencing. Taking advantage of these methods could be beneficial for the fdC profiling. LC-MS, orbitrap or tridruaplex, is another option.

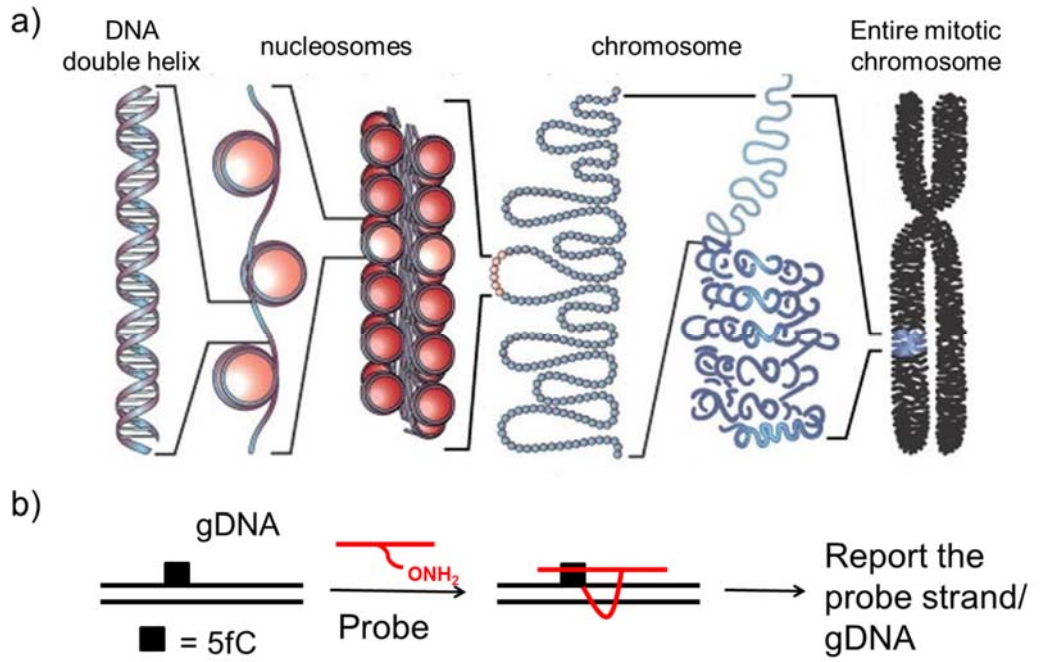


Figure 5-1 a) Structure organization of eukaryotic DNA, the figure is adapted from the reference;[322] b) schematic representation of the project workflow.

5.2. fdC probe strand: evolution and evaluation

5.2.1. Verso-evolution of probe strands

According to the previous reports, both hydroxylamine[229, 323] (R-ONH_2, **27**, **28**) and hydrazine derivatives[324] (R-$NHNH_2$, **29**) irreversibly reacted with fdC (Figure 5-2). 1-Pyrenemethylamine[325] (**30**) reaction with fdC for formylcytosine labeling was also brought to our attention, but due to the absence of α-effect, the Schiff base will be less stable in such an aliphatic system. Additionally, a fluorescein tag containing hydrazide group (R-$CONHNH_2$) was applied to react with a ketone for protein labeling.[326-328]

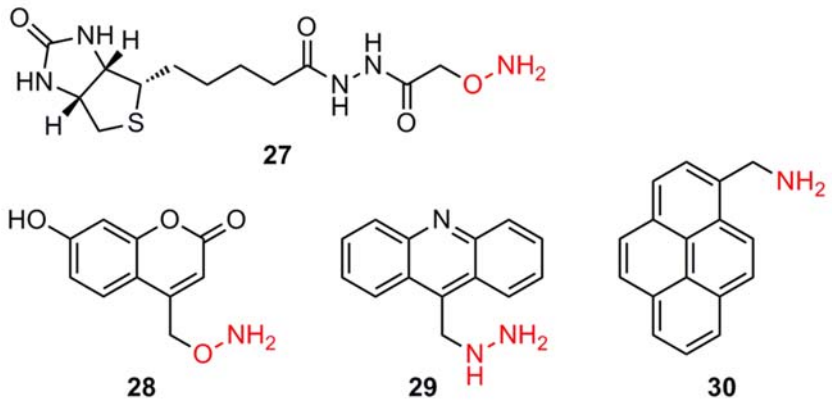


Figure 5-2 Examples of small molecules that acted as fdC probes: aldehyde reactive probe (ARP, **27**), coumarin-HA (**28**), 9-(hydrazinylmethyl)acridine (**29**), 1-pyrenemethylamine (**30**).

An orthogonal bioconjugation reaction should be employed to attach the amine linker on the probe strand. Apart from the hydroxylamine aldehyde condensation, Cu-catalyzed Huisgen cycloaddition, tetrazine trans-cyclooctene and cyclopropene Diels-Alder reaction,[329-331] acyl trifluoroborate amide formation,[332] etc. are all conventional reactions for bioconjugates that could be employed.[333] The trans-cyclooctene reaction shows a high rate, 10^3~10^6 $M^{-1}s^{-1}$, however, the large steric hindrance impedes its application. The Huisgen cycloaddition and acyl trifluoroborate amide formation have similar reaction rate, 10~20 $M^{-1}s^{-1}$,[332] and the hydroxylamine aldehyde condensation is relatively slow with a rate of 10^{-3}~10^{-1} $M^{-1}s^{-1}$.[334] In this thesis, the Huisgen cycloaddition was chosen due to its elegance and robustness. Thus,

an azide group was placed on one end of the linker, and an alkynyl modifier was placed in the probe strand.

Therefore, two sets of linkers with hydroxylamine (-ONH_2) or hydrazide (-$CONHNH_2$) at one end and an azide group at the other end were synthesized. Each set of linker consists of alkyl chains with three different lengths, i.e. propyl, butyl, and pentyl (Figure 5-3).

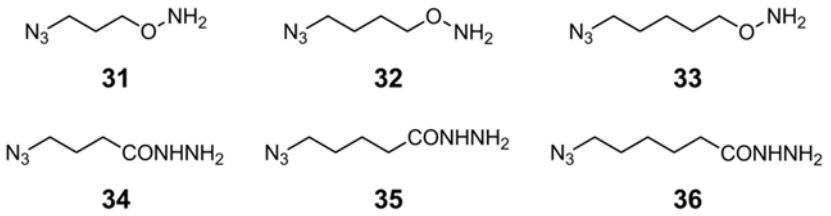
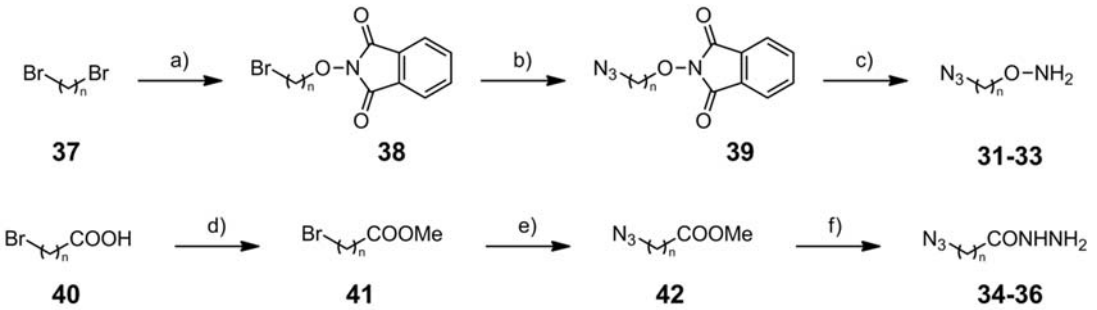


Figure 5-3 Structures of aliphatic linkers synthesized in this study.

Synthesis of the hydroxylamine linker **31**-**33** started with dibromoalkane (Scheme 5-1). Substitution with *N*-hydroxyphthalmide followed by conversion of bromide into azide furnished **39**. After hydrazine solution treatment the hydroxylamine linkers were obtained in overall yields of 52-67%. For the hydrazide linkers **34**-**36**, methyl esterification followed by azide substitution gave the ω-azide methyl ester **42**. Conversion of **42** into the hydrazide was achieved by refluxing in a methanol solution of hydrazine.



Scheme 5-1 Synthesis of the hydroxylamine linkers **31**-**33** and hydrazide linkers **34**-**36**. a) *N*-hydroxyphthalimide, triethylamine, DMF, rt; b) NaN_3, DMF, rt; c) N_2H_4•H_2O, DCM, rt, 52-67% over three steps; d) H_2SO_4, MeOH, reflux; e) NaN_3, DMF, rt; f) N_2H_4•H_2O, MeOH, reflux, 64-69% for three steps.

Aliphatic linkers with hydrazide were also prepared, but their instability at room temperature, at least in our hand, limited their further application.

Next, a 13 mer single DNA strand with one fdC and a complementary probe strand with alkynyl modifiers were prepared, hybridized and studied. To find the best combination of alkyl chain length, reactive moiety on the linker, and position of the linker on the probe strand, six linkers (**31**-**33** and **34**-**36**) were first connected to the fdC target strand (Table 5-1). 20 equivalents of the linker to **ODN 13** was used to guarantee a complete conversion. After HPLC purification, six fdC contained oligonucleotides with different azide linkers were prepared. They are stable enough to survive the concentration and redissolving process.

Table 5-1 Synthesized fdC contained oligonucleotides for duplex model study. Bold and red fdC represent uncrosslinked and crosslinked formyl cytosine. Found values indicate the mass results measured by MALDI-TOF. (L: linker)

Entry	5' ---------- 3'	mer	calc.	found
12a	GTA ATG CGC TAG G	13	/	/
12b	CCT AGC GCA TTA C	13	/	/
13	GTA ATG **fC**GC TAG G	13	4040.9	4036.2
13-**L31**	GTA ATG **fC**GC TAG G	13	4139.0	4137.4
13-**L32**	GTA ATG **fC**GC TAG G	13	4153.0	4149.8
13-**L33**	GTA ATG **fC**GC TAG G	13	4167.0	4163.8
13-**L34**	GTA ATG **fC**GC TAG G	13	4166.0	4166.2
13-**L35**	GTA ATG **fC**GC TAG G	13	4180.0	4175.7
13-**L36**	GTA ATG **fC**GC TAG G	13	4194.0	4190.2

To find out the best anchor position on the probe stand, we placed different alkynyl-modifiers on the probe strands. Reported alkynyl modifiers are summarized in Figure 5-4. 2'-*O*-alkyne modifier (**43**) and 5-ethynyl-2'-deoxycytidine (**45**) were chosen because of their accessibility.

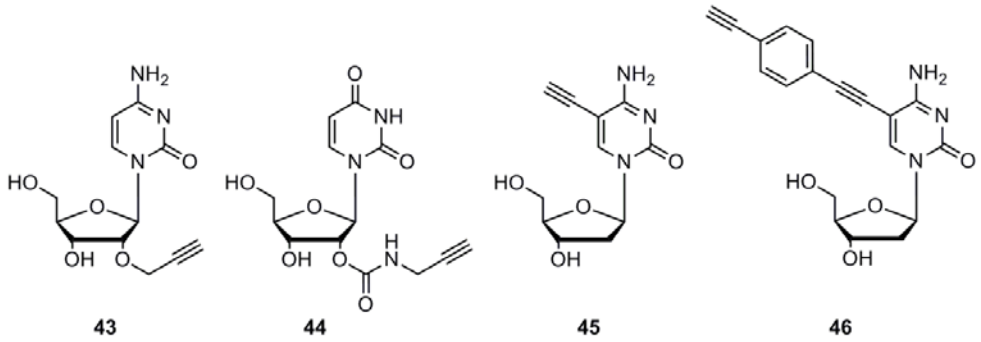


Figure 5-4 Summary of published alkynyl nucleoside modifiers. 2'-(*O*-propargyl) cytidine (**43**), uridin-2-yl carbamate (**44**),[335] 5-ethynyl-2'-deoxycytidine (**45**)[336] and 3'-*O*-Acetyl-5-(2-(4-ethynyl phenyl)ethynyl)-2'-deoxycytidine (**46**).[337]

The modifier **43** was placed every four bases in the counter strand for the consideration that too crowded alkynyl groups at 2'-OH may disturb the normal duplex form. **ODN 14-17**, each containing three alkynyl modifiers, were mixed with the six fdC-linker target strands. A library of 24 duplexes was thereby obtained. If the azide and the alkynyl group are in an appropriate position, Huisgen 1,3-dipolar cycloaddition will occur. The cross-linked oligonucleotides will show a band in the gel electrophoresis that is different from the single strand on a denaturing PAGE.

Table 5-2 Synthetized oligonucleotides containing 2'-*O*-propargyl bases for duplex model studies. Bold and red letters represent the alkyne modifier **37**. Found values indicate alkyne cytosine modifiers **39**. Mass results measured by MALDI-TOF.

Entry	5' ---------- 3'	mer	calc.	found
14	**C**CT A**G**C GC**A** TTA C	13	4054.7	4052.6
15	C**C**T AG**C** GCA **U**TA C	13	4040.7	4037.6
16	CC**U** AGC **G**CA T**U**A C	13	4026.7	4024.1
17	CCT **A**GC G**C**A TT**A** C	13	4054.7	4052.0
16a	CC**U** AGC GCA TTA C	13	3932.7	3930.5
16b	CCT AGC **G**CA TTA C	13	3946.7	3945.8
16c	CCT AGC GCA T**U**A C	13	3932.7	3931.5
18	CCT AGC GCA UTA C	13	3916.7	3913.3

Percentages of the duplex formed in 4 hours for each trial are listed in Table 5-3. In general, the hydroxylamine linkers (**L31-33**) gave better crosslink than hydrazide linkers (**L34-36**). As reactions were processed, hydrazide linkers displayed more

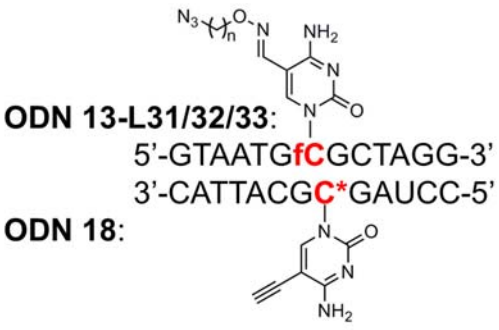
bands, suggesting unspecific reaction or instability under the reaction conditions. To obtain optimal crosslinking condition, freshly prepared CuBr-TBTA solution and 4 h reaction time are necessary. The best crosslink yield within duplexes (94%) is obtained using **ODN 16** and **ODN 13-L32**, suggesting that a hydroxylamine group with a $(CH_2)_4$ chain gives the most suitable approaching towards one of the three alkynyl groups on the probe strand. In the second round of evolution, **ODN 16a-c** were prepared, each containing one alkynyl group corresponding to **ODN 16**. This time, **ODN 16a** reacted most efficiently with **ODN 13-L32**: 93% of the fdC reacted in this duplex, similar as **ODN 16** and **ODN 13-L32**. This reuslt was then providing the optimal probe strand containing a hydroxylamine group to catch fdC aldehyde group on a given target strand.

A similar strategy, using the alkyne cytosine modifiers **45** (Figure 5-4) in the probe strand, was explored next one. In the genome, fdC and other epigenetic cytosine modification always appear in CpG sites. If the cytosine opposite the guanine can be used to target fdC in this CpG site, one can in principle detect fdC in any sequence context with the same probe base (Figure 5-5).

Probe strand **ODN 18** containing modifier **45** was kindly obtained from Dr. Milan Dejmek. **ODN 18** was mixed with **ODN 13-L31/32/33** and CuBr-TBTA under the same condition mentioned above. PAGE assays showed however that this linkage mode is not optimal. The maximal yield of the duplex was 54% with linker **33**, but multiple bands appeared. Therefore, this strategy was not investigated any further.

Table 5-3 Percentage of the aldoxime mediated duplex formation after 1, 2 and 4 h. Percentages in bold indicate the best result in each round; Percentages in red indicate the present of multiple bands.

1 h	13-**L31**	13-**L32**	13-**L33**	13-**L34**	13-**L35**	13-**L36**
ODN 14	25%	34%	43%	11%	26%	9%
ODN 15	17%	51%	39%	12%	16%	5%
ODN 16	38%	**75%**	52%	28%	37%	9%
ODN 17	38%	33%	44%	4%	11%	7%
ODN 16a		**69%**				
ODN 16b		14%				
ODN 16c		14%				
2 h	13-**L31**	13-**L32**	13-**L33**	13-**L34**	13-**L35**	13-**L36**
ODN 14	29%	34%	46%	22%	42%	10%
ODN 15	24%	55%	40%	14%	21%	5%
ODN 16	42%	**83%**	53%	29%	38%	12%
ODN 17	38%	36%	46%	6%	12%	8%
ODN 16a		**73%**				
ODN 16b		17%				
ODN 16c		15%				
4 h	13-**L31**	13-**L32**	13-**L33**	13-**L34**	13-**L35**	13-**L36**
ODN 14	47%	49%	76%	30%	55%	22%
ODN 15	52%	88%	74%	22%	49%	18%
ODN 16	70%	**94%**	85%	41%	67%	28%
ODN 17	71%	60%	77%	13%	26%	21%
ODN 16a		**93%**				
ODN 16b		52%				
ODN 16c		68%				



h	13-L31	13-L32	13-L33
1	27%	36%	49%
2	34%	40%	50%
4	39%	48%	54%

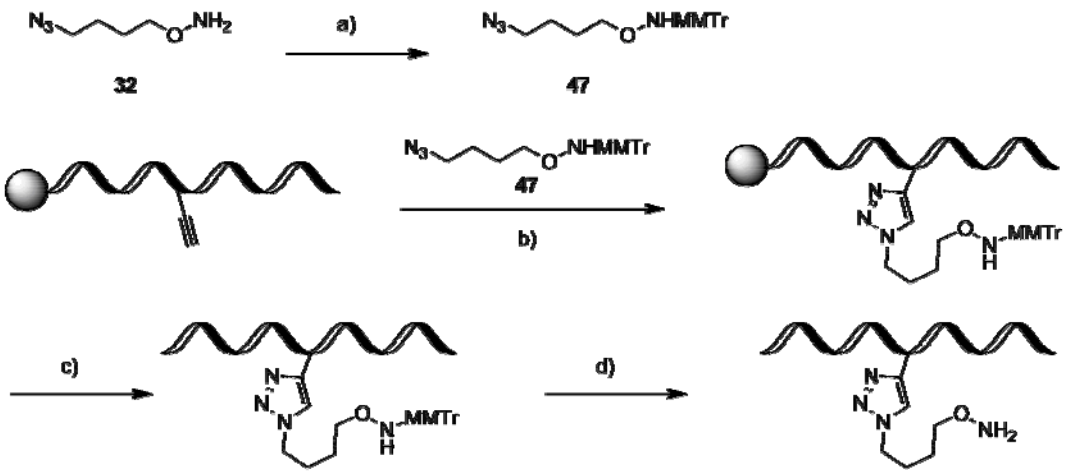
Figure 5-5 Schematic representation of probe **ODN 18** containing modifier **39**, and percentage of the aldoxime mediated duplex formation after 1, 2 and 4 h. Percentages in red indicate the present of multiple bands.

5.2.2. Evaluation of the probe strand

5.2.2.1. Confirmation of the crosslink

After the best linker had been obtained, we questioned whether this combination of the linker and modified position on the probe strand essentially worked. The above protocol was done reversely; i.e. the linker was first clicked on to the probe strand, and then hybridized with the counter strand containing a fdC.

Great efforts have been made to click the linker **32** on the deprotected probe strands, with and without the hydroxylamine protecting group. It is reported that 1,3-dipolar cycloaddition succeeded when using the uridin-2'-yl carbamate alkynyl monomer **44** in solution.[335] In the case of 2'-*O*-propargyl modifier, "on column" cycloaddition[338] and in solution click labeling[339] have been reported. However, in our hand, all the in solution reaction gave complicated product mixture according to HPLC analysis.



Scheme 5-2 Synthesis of oligonucleotides based probe with linker **32**. a) MMTr-Cl, DIPEA, DCM, rt, 84%; b) $CuSO_4$, sodium ascorbate, **47**, DIPEA, MeCN, DMSO, H_2O, rt; b) $NH_3 \bullet H_2O$, 25°C; c) CH3COOH aq., 25°C, quant.

So the solid-phase bound alkyne DNA was first reacted with MMTr protected hydroxylamine linker **47**, as depicted in Scheme 5-2. After cleavage from the solid phase, a sharp peak corresponding to the desired **ODN 16d** was observed by HPLC (Figure 5-6). After treating the product with aqueous acetic acid, a single peak with shorter retention time appeared in the HPLC trace, indicating MMTr was deprotected. The present of the product **ODN 16e** with the hydroxylamine linker ($R\text{-}ONH_2 + Na^+$)

was confirmed by MALDI-TOF, as shown in Table 5-4.

The reason why the peak corresponding to deprotected strands without sodium ion was absent is not known, even if no sodium ion was used in the deprotection and purification steps. Oligonucleotides with one linker were synthesized smoothly while effort towards strand with two linkers failed.

Table 5-4 Synthesized oligonucleotides containing hydroxylamine linker and fdC shifted strand. Bold and red letters represent alkyne modifiers. *: contain one sodium ion.

Entry	5' ---------- 3'	mer	calc.	found
16a	CC**U** AGC GCA TTA C	13	3932.7	3930.5
16d	CC**U-L-MMTr** AGC GCA TTA C	13	4335.2	4335.1
16e	CC**U-L-ONH$_2$** AGC GCA TTA C	13	4084.8*	4084.2
19	GTA AT**fC** CGC TAG G	13	4000.9	3999.6

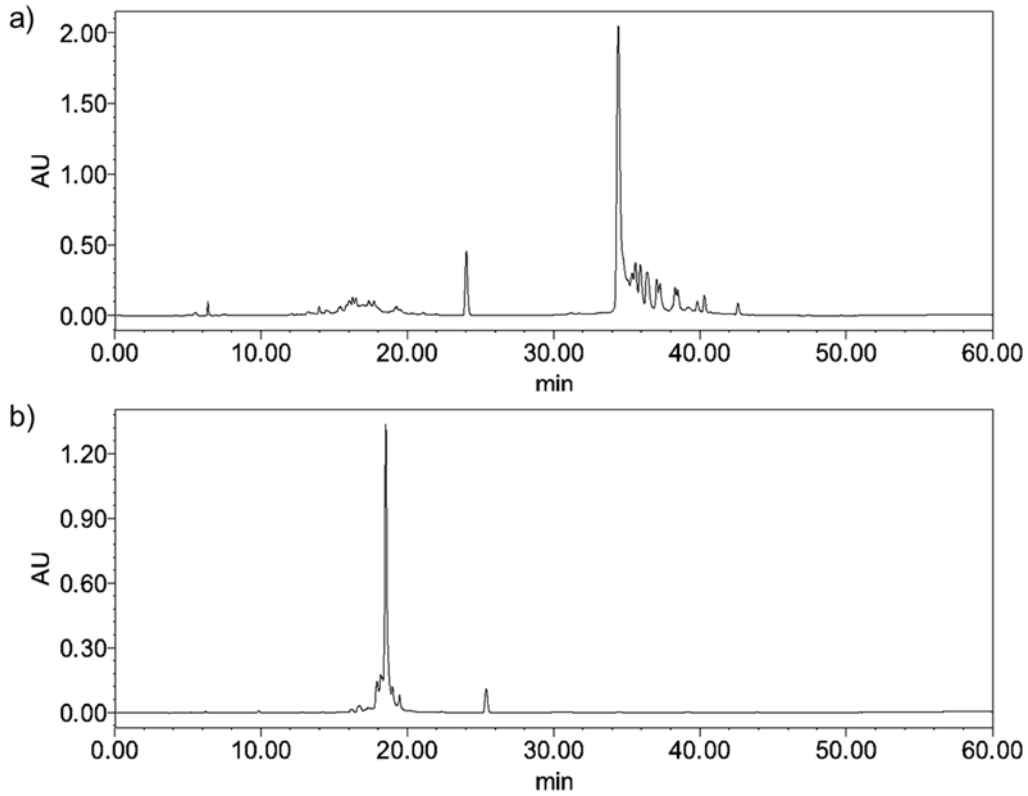


Figure 5-6 Typical HPLC trace of crude product: a) **ODN 16d** (**ODN 16a** attached MMTr protected linker 32) and b) **ODN 16e** (**ODN 16a** attached deprotected linker **32**). Conditions: buffer A, 0.1 M TEAA; buffer B, 0.1 M TEAA in 80% acetonitrile, linear gradient from 0% to 70% B over 45 min. Retention time: (a) 34.4 min, (b) 18.5 min. AU = arbitrary unit.

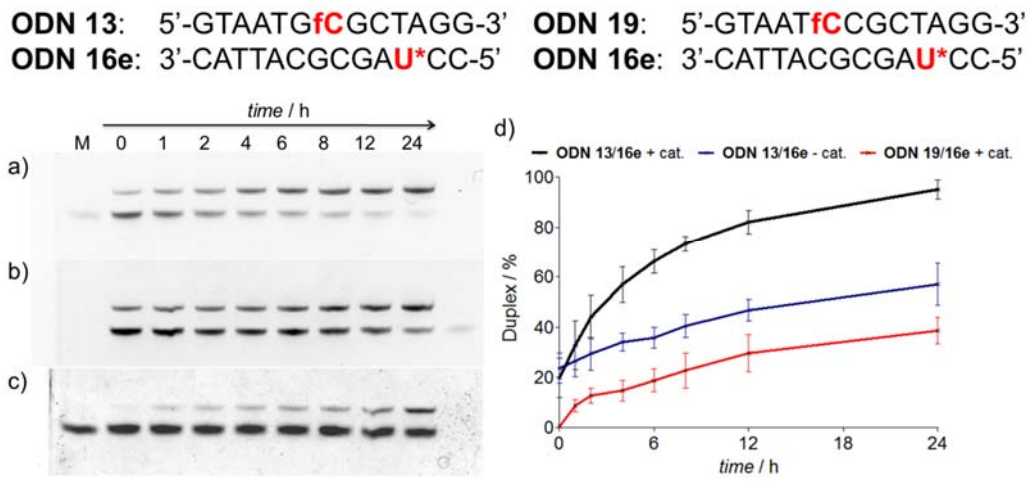


Figure 5-7 Kinetic of crosslink formation. Denaturing PAGE assay shows the duplex formation of a) **ODN 13**/**16e** with catalyst 4-methoxyaniline b) **ODN 13**/**16e** without a catalyst, c) **ODN 19**/**16e** with the catalyst at 25 °C for 24 h and d) quantification of DNA duplex formation along the reaction. Error bars represent the standard error of mean calculated from three replicates. Conditions: 2 μM oligonucleotides, 100 mM NaCl, 10 mM NaOAc buffer pH 6.0, 4-methoxyaniline as catalyst 10 mM. M = 13 mer single strand marker.

When **ODN 16e** was mixed with **ODN 13**, the aldoxime formation was monitored by denaturing PAGE assay within 24 hours. It was found that 50% of the single strand pairs became indivisible duplexes even without the addition of any catalyst after 24 hours. In the present of 4-methoxyaniline as the catalyst,[229] the reaction yield reached 95%. In the negative control experiment, **ODN 16e** and **ODN 19**, in which fdC was shifted from +4 to +5 position, only 17% of aldoxime was formed after 24 hours (Table 5-4 and Figure 5-7). These experiments confirm that the **ODN 16e** has the optimum position, proper linker length, and reactivity to form an irreversible covalent bond with fdC on the target strand.

In this reaction, fdC first reacts with the catalyst to form a conjugate. The 4-methoxyaniline has an electron-withdrawing group on the *para*-position of aniline. W. Michaela *et al.* studied the catalyst effects of different substituted anilines in their model protein-PEG conjugation system and deemed the 1,4-diaminobenzene as the best.[340] In practice, 1,4-diaminobenzene worked better in our case and completed the crosslink reaction in 4 hours (Figure 5-8).

To further confirm the oxime crosslink formation, MALDI-TOF mass spectrum were

measured (Table 5-5 and Figure 5-9). Mass signals corresponding to the duplex of two 13 mer oligonucleotides were observed (calc. 8084.7, found 8081.9). Signals corresponding to the two single strands **ODN 13**, **ODN 16e**, and **ODN 13**/ 4-methoxyaniline conjugate also appeared.

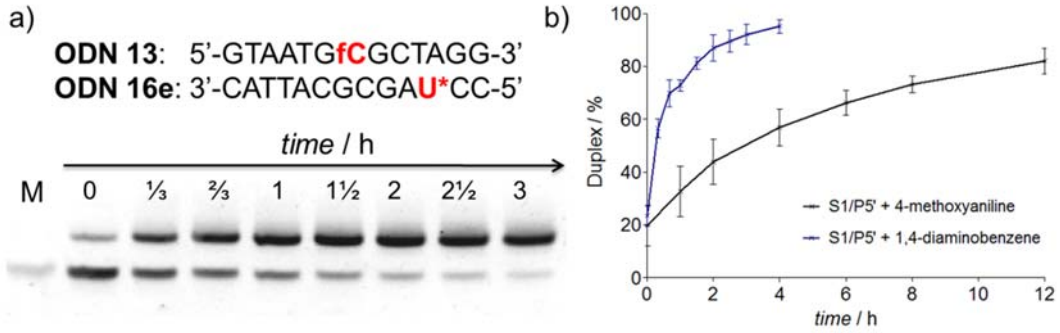


Figure 5-8 Comparison of aniline catalyst: a) denaturing PAGE assay shows the duplex formation using 1,4-diaminobenzene as a catalyst; b) quantification of DNA duplex formation along the reaction compared with reaction using 4-methoxyniline as a catalyst. Error bars represent the standard error of mean calculated from three replicates. Conditions: 2 μM oligonucleotides, 100 mM NaCl, 10 mM NaOAc buffer pH 6.0, 1 mM 1,4-diaminobenzene or 10 mM 4-methoxyniline. M = 13 mer single strand marker.

The meling profile of duplex **13**/**16e** was measured to provide additional evidence. (Figure 5-10) Because 4-methoxyaniline displays a strong absorbance at 200-300 nm, the catalyst was not added in this case. Two inflection points are observed after 24 h incubation of duplex **13**/**16e**. One point (44°C) corresponds to the duplex melting point measured after reannealing. The other from the crosslinked duplex (79°C) is 28 °C higher than the canonical duplex **13**/**12b**. It is consistent with our assumption that the linkage between the two single strands will stabilize the duplex. It also confirms that the catalyst is necessary if complete crosslinking is aimed, in accord with PAGE assay (Figure 5-7b).

To sum up, the above results verify that the probe containing a butyl hydroxylamine linker at the n+4 position can react with fdC on the counter target strand efficiently and in a sequence-specific manner.

Table 5-5 MALDI-TOF results of duplex **13**/**16e**. Bold and red letters represent alkyne modifiers.

Entry	5' ---------- 3'	mer	calc.	found
13	GTA ATG **fC**GC TAG G	13	4040.9	4038.4
13+cat.	GTA ATG **fC**GC TAG G	13	4146.0	4143.9
16e	CC**U-L-ONH$_2$** AGC GCA TTA C	13	4061.8	4060.5
	duplex 13/15e		8084.7	8081.9

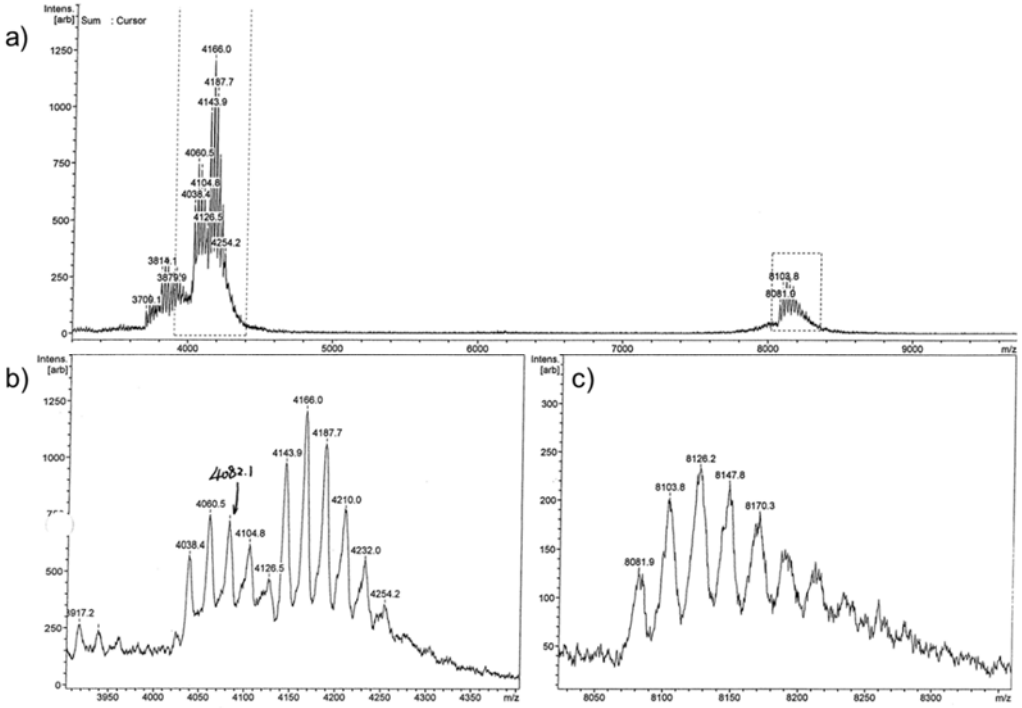


Figure 5-9 MALDI-TOF mass spectrum of linked duplex **13**/**16e** and single strands: a) overall MALDI-TOF spectrum; b) peaks corresponding to **ODN 13** and **ODN 13** conjugated 4-methoxyaniline; c) peaks corresponding to linked duplex **13**/**16e** Conditions: 10 µM oligonucleotides, 100 mM NaCl, 10 mM NaOAc buffer pH 6.0.

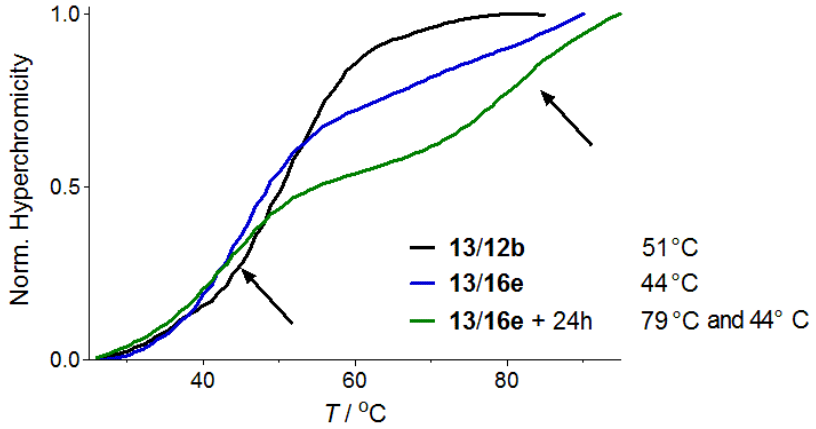


Figure 5-10 Melting curves of duplex **13**/**16e** after reannealing or after 24 h incubation without 4-methoxyaniline compared with duplex **13**/**12b**. Condition: 1 µM oligonucleotides, 100 mM NaCl, 10 mM NaOAc buffer pH 6.0, the final volume of 200 µL.

5.2.2.2. Chemical selectivity

When turned to the chemical selectivity, it is found that the probe strand does not react with DNA templates containing a site-specific mdC, hmdC, or cadC (**ODN 12a**, **13a-d**, Table 5-6 and Figure 5-11). Because genomic DNA may contain abasic sites (apurinic/apyrimidinic, AP sites) with an aldehyde function, a control experiment with oligonucleotides **ODN 13e** containing an AP site was performed. In contrast to epigenetic bases, AP sites did react with probe strand with a yield of 90% under the same condition.

Table 5-6 Synthesized oligonucleotides for specificity study. Bold letter T indicates mismatch position. Bold and red letters represent epigenetic, AP or hydroxylamine modifiers. Fl represents fluorescein tag (6-FAM). *: contain one sodium ion.

Entry	5' ---------- 3'	mer	calc.	found
12a	GTA ATG CGC TAG G	13	/	/
13a	GTA ATG **T**GC TAG G	13	/	/
13b	GTA ATG **mC**GC TAG G	13	4026.7	4021.6
13c	GTA ATG **hmC**GC TAG G	13	4042.7	4039.6
13d	GTA ATG **caC**GC TAG G	13	4056.7	4054.9
13e	GTA ATG **AP**GC TAG G	13	3916.4	3919.6
16e	CC**U-L-ONH$_2$** AGC GCA TTA C	13	4084.8*	4082.2
16f	**Fl**-CC**U-L-ONH$_2$** AGC GCA TTA C	13	4622.3*	4621.2

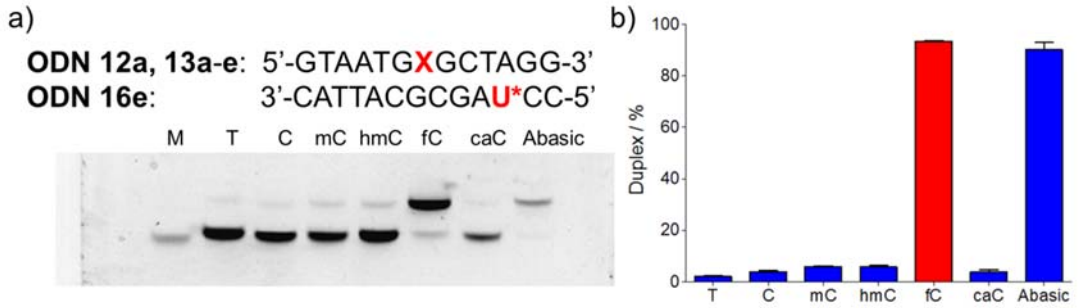


Figure 5-11 Specificity study of probe strand **ODN 16e**: a) denaturing PAGE assay shows the duplex formation of **ODN 12a**, **13**, **13a-e** and **ODN 16e** at 25 °C for 24 h; b) quantification of duple **12a**, **13**, **13a-e** and **16e** formed after 24 h. Error bars represent the standard error of mean calculated from three replicates. **X** indicates modified site in the test strand (T, C, mdC, hmdC, cadC, and abasic site). Conditions: 2 μM oligonucleotides, 100 mM NaCl, 10 mM NaOAc buffer pH 6.0. M = 13 mer single strand marker.

The sensitivity of this detection method, at this stage, was investigated by PAGE quantification. **ODN 16f**, the probe **ODN 16e** with a fluorescein tag was prepared for imaging. **ODN 13** (contains fdC) and **12a** (C instead of fdC) were mixed in different ratios and reacted with **ODN 16e**. In both methods, photographed directly for 6-FAM and photographed after *SYBR* Green dying, 10% of fdC can be observed, which equals to 0.5 pmol and 200 nM. The fdC probe reaction can of course be coupled to other techniques, e.g. PCR, LC-MS, etc., to reach a even better detection limit.

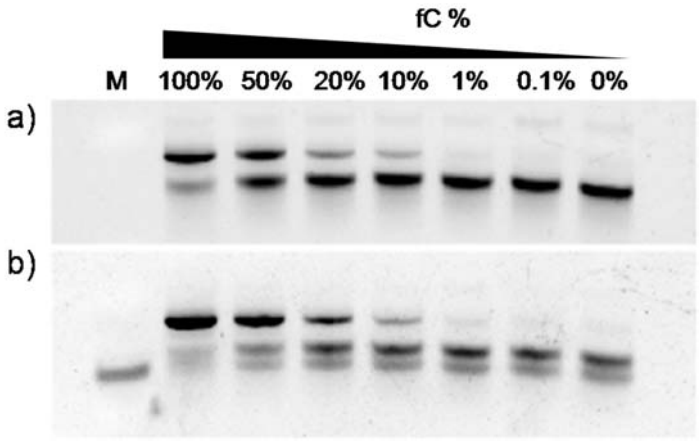


Figure 5-12 Denaturing PAGE assay shows the duplex formation of **ODN 16e** with different ratios of **ODN 13** and **ODN 12**: a) photographed directly for 6-FAM; b) photographed after *SYBR* Green I dyeing. Conditions: 2 μM oligonucleotides, 100 mM NaCl, 10 mM NaOAc buffer pH 6.0. Loading volume: 2.5 μL+ loading buffer. M = single strand marker **13**.

5.2.2.3. Sequence specificity

Next, to check if the probe could tolerate any mismatch pairs inside the duplex, three duplexes were synthesized as negative control experiment: one with T-C pair outside the N to N+4 area (mismatch duplex 1, **ODN 13**/**20c**, Table 5-7); one with G-T pair inside the N to N+4 area (mismatch duplex 2, **ODN 21**/**16e**). The third mismatch duplex consists of **ODN 22** with fdC at the same position as in **ODN 13.**

At 25°C, both mismatched duplexes 1 and 2 showed high aldoxime formation efficiencies of 92% and 88% respectively, compare to the positive experiment (95%, Figure 5-13). This means the fdC probe cannot discriminate a single mismatch at 25°C. Random mismatch duplex 3 showed a yield of 7% after 24h, suggested that the

two single strand could hardly approach each other to provide a suitable spatial location for aldoxime formation.

Table 5-7 Synthesized oligonucleotides for sequence selectivity study. Bold and red letters represent alkyne modifiers. *: contain one sodium ion.

Entry	5' ---------- 3'	mer	calc.	found
20a	CCU ATC GCA TTA C	13	/	/
20b	CCU-L-MMTr ATC GCA TTA C	13	4310.2	4310.3
20c	CCU-L-ONH$_2$ ATC GCA TTA C	13	4059.8*	4064.3
21	GTG ATG fCGC TAG G	13	4056.9	4055.5
22	CTA GCA fCGT CTC G	13	3936.8	3933.6

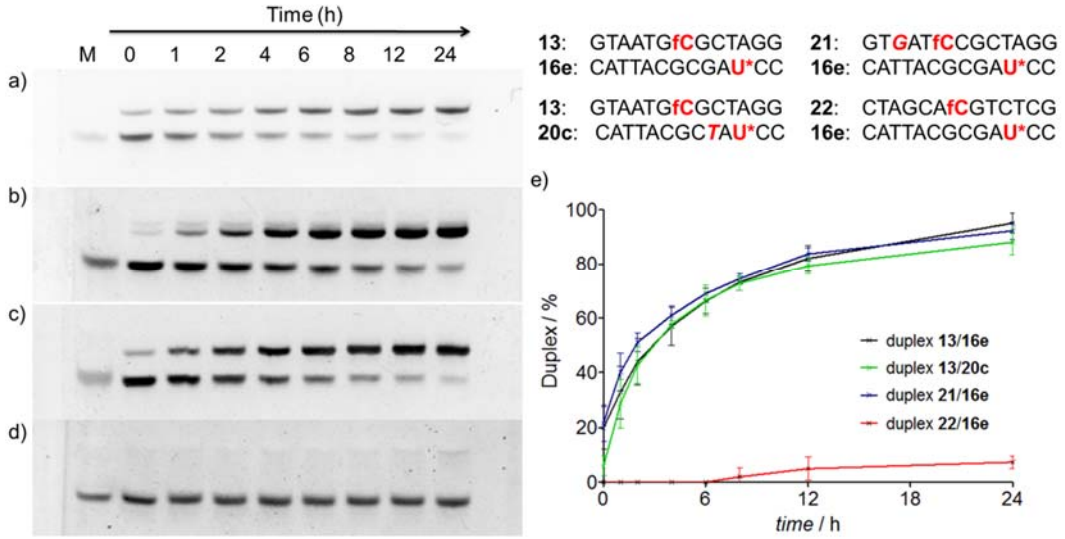
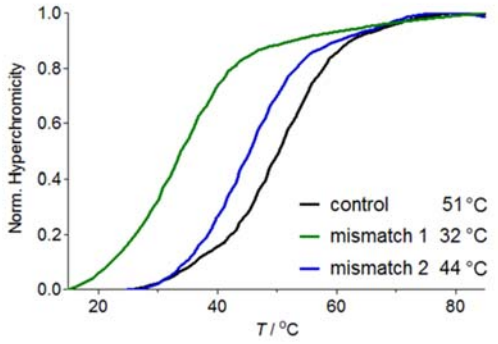


Figure 5-13 Denaturing PAGE shows the mismatched duplex crosslinking: a) positive control, **13/16e**; b) mismatch 1, **13/20c**; c) mismatch 2, **21/16e**; d) mismatch 3, **22/16e** at 25 °C for 24 h; e) quantification of DNA duplex formation along the reaction. Error bars represent the standard error of mean calculated from three replicates. Conditions: 2 μM oligonucleotides, 100 mM NaCl, 10 mM NaOAc buffer pH 6.0, 4-methoxyaniline as catalyst. M = single strand marker **13**.

The low discriminability evoked us to study the thermostability of these duplexes. Duplex **13**/**12b** (T instead of modified U in **16e**) showed a melting temperature of 51°C while mismatch duplex 1 and 2 had a T_M of 32°C and 44°C respectively (Figure 5-14). It was therefore found that at a higher reaction temperature, e.g. 45°C instead of 25°C, duplexes with mismatch pairs should be less stable so that aldoximes could potentially not form.



Entry		5' ---------- 3'	T_M
control	13	GTA ATG fCGC TAG G	51 °C
	12b	CCT AGC GCA TTA C	
mismatch 1	13	GTA ATG fCGC TAG G	32 °C
	20d	CCT A*T*C GCA TTA C	
mismatch 2	21	GT*G* ATG fCGC TAG G	44 °C
	12b	CCT AGC GCA TTA C	
mismatch 3	22	CTA GCA fCGT CTC G	--
	12b	CCT AGC GCA TTA C	

Figure 5-14 UV melting curves of duplexes **13**/**12b** (control), **13**/**20d** (mismatch 1), and **21**/**12b** (mismatch 2). Italic letters G and T indicate mismatch position. Condition: 1 µM oligonucleotides, 100 mM NaCl, 10 mM NaOAc buffer pH 6.0, the final volume of 200 µL.

The same experiments were therefore repeated with the difference that the incubation temperature was raised to 45°C. The yield of positive experiment and mismatch duplex 1 are similar to that at 25°C, i.e. 94% and 88% while the yield now obtained for the mismatched duplex 2 decreases to 59%. Random mismatched duplex 3 shows little crosslinking, 7% (Figure 5-15). As 45°C is similar to the melting temperature of duplex 2 (44°C) but higher than duplex 1 (32 °C), the mismatched situation can therefore be discriminated at an elevated temperature.

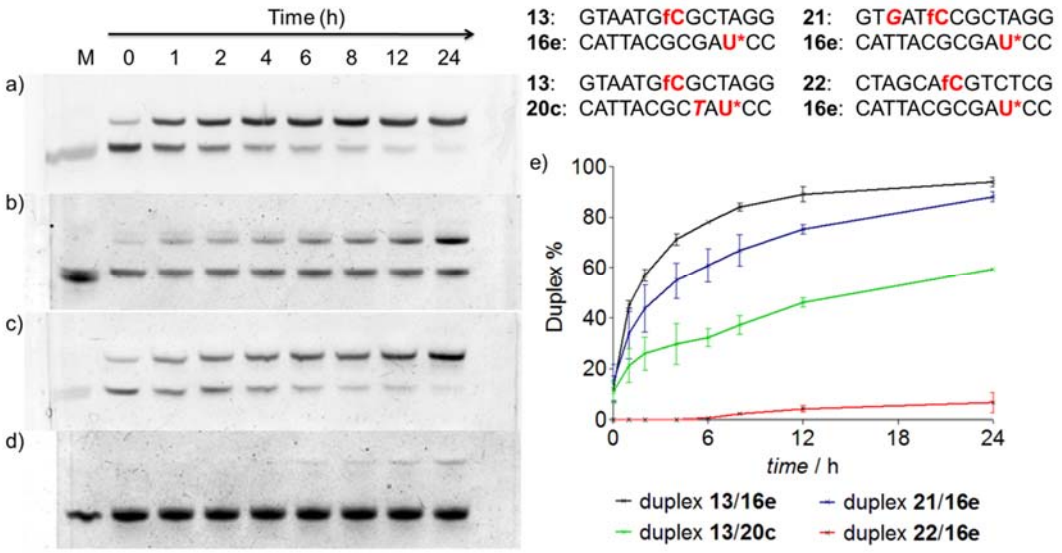


Figure 5-15 Denaturing PAGE assay shows the duplex formation of a) control: **13/16e** b) mismatch 1: **13/20c** c) mismatch 2: **21/16e** and d) mismatch 3: **22/16e** at 45 °C for 24 h. e) Quantification of DNA duplex formation along the reaction. Error bars represent the standard error of the mean calculated from three replicates. Conditions: 2 µM oligonucleotides, 100 mM NaCl, 10 mM NaOAc buffer pH 6.0, *p*-methoxyamine as catalyst. M = single strand marker **13**.

5.2.2.4. Detection selectivity with multiple probes

In order to test the method in a more complicated scenario, the target strand **ODN 23** containing three fdCs was prepared. **ODN 24**, **25** and **26** were prepared as the three probes of 12, 13, and 14 mer respectively. They are designed for reaction with the three different fdCs on **ODN 23**. Follow the procedure mentioned above, these three probes were reacted with **ODN 23**, to give slower migrating bands than **ODN 23** alone on PAGE due to specific crosslink formation (Figure 5-16). However, these three duplexes, which were different in only one nucleoside, showed quite similar migration distance. To clarify the difference of crosslinking between the target and each probe strand, the primer extension experiment with Klenow Fragment (exo-) was performed. As the catalyst in DMSO solution and acidic buffer may not suitable for the enzyme, a working solution of 20 μM was made for aldoxime formation, followed by dilution to 2 μM with polymerase reaction buffer. After the extension, clear distinction of three kinds of crosslink is observed, i.e. duplexes of 30 mer template strand with 15 mer (**ODN 24**), 24 mer (**ODN 25**) and 30 mer counter strand (**ODN 26**). Here, the position specificity of the method is proved.

When **ODN 24** is connected to **ODN 23**, position 5-15, there would be enough room to place another probe **ODN 26** on 17-30 base position. The situation should be same if **ODN 26** is placed firstly. However, if **ODN 25** occupies the 9-20 base position, neither **ODN 24** nor **ODN 26** could be able to attach to the sample strand. The experimental results proved this idea (Figure 5-17). Line 3 and line 5 are of the same height, indicating that a single strand is connected with two counter strands of 12 and 14 bases. No clear difference is shown when **ODN 25** is bound to **ODN 23** followed by addition of **ODN 24** or **ODN 26**. Notably, lines 7 and 8 show a double band around the shift position of the sample and one probe strand, which is caused by strands replacement. The uppermost band, slower than a 30 mer duplex in the gel shift, is presumably due to unspecific reaction or aggregation.

Once an oligonucleotide with two linkers is synthesized, together with a strand

containing two fdC or abasic sites, a self-assembled duplex structure will come out. More complicated nano-object can be constructed based on the mixture of covalent bond and hydrogen bridge, which makes the structure resistant to denaturing. Further investigation should be fascinating although it is not the research purpose here.

a)

Entry	5' ---------- 3'	mer	calc.	found
23	GGGGGfCGGTGfCCAGTAAGAAACTGfCAAGGT	30	9466.1	9460.0
24a	TACTGGC-L-MMTrACCGC	12	4061.2	4058.8
24	TACTGGC-L-ONH$_2$ACCGC	12	3810.9*	3812.0
25a	TTTCTTA-L-MMTrCTGGCA	13	4370.2	4367.5
25	TTTCTTA-L-ONH$_2$CTGGCA	13	4119.8*	4120.7
26a	AC-L-MMTrCTTGCAGTTTCT	14	4659.3	4657.7
26	AC-L-ONH$_2$CTTGCAGTTTCT	14	4408.9*	4407.6

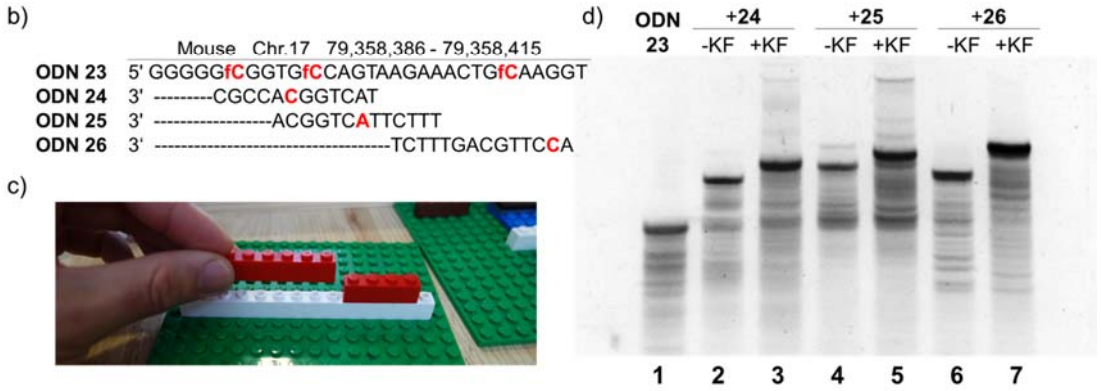


Figure 5-16 Study of detection selectivity I. a) synthesized oligonucleotides for this study. Bold and red letters represent alkyne modifiers and fdC. Found values indicate the mass results measured by MALDI-TOF. *: contain one sodium ion. b) Sequences of the **ODN 23-26** and schematic representation of primer extension experiment. c) Lego building block shows the primer extension experiment. d) Denaturing PAGE assay shows crosslink between **ODN 23** and three probes before (-KF) and after extension with Klenow Fragment (+KF). Condition: 2 μM crosslink, 200 μM dNTP, 6 U Klenow Fragment, 1× NEB buffer 2, the final volume of 30 μL, dyed with SYBR Green.

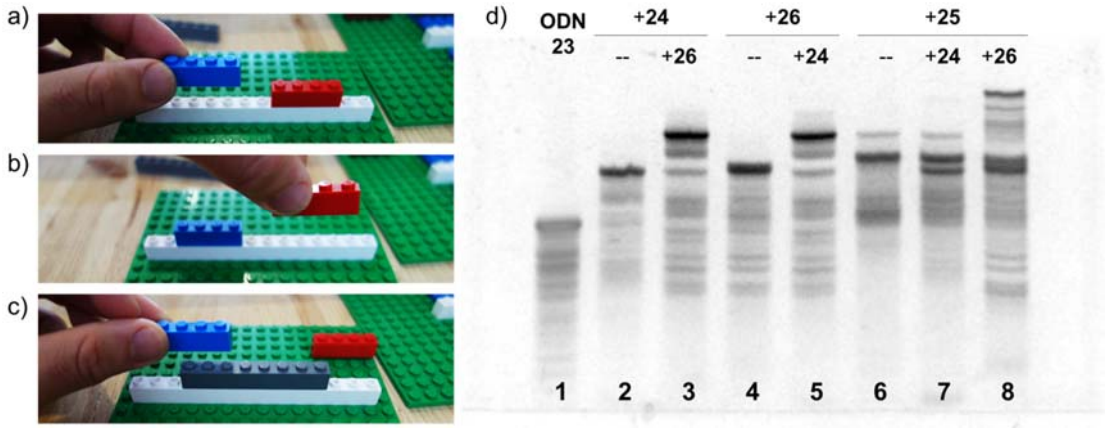


Figure 5-17 Study of detection selectivity II. a), b), and c) Logo building block shows the probe blocking experiments. The white block represents **ODN 23**, blue **ODN 24**, gray **ODN 25**, and red **ODN 26**. d) Denaturing PAGE assay shows crosslink of **ODN 23** with double probes: 1) **ODN 23** (30 mer), 2) duplex **23**/**24** (30+12 mer), 4) duplex **23**/**26** (30+14 mer), 6) duplex **23**/**25** (30+13 mer). Line 3 and 5 show **ODN 23** with two non-overlapping probes **24** and **26** corresponding to a) and b); line 7 and 8 show **ODN 23**/**25** with two non-overlapping probes **24** and **26**, corresponding to c). Double bands of line 7 and 8 are due to incomplete reaction of **ODN 24** and **26**. Condition: 2 µM **ODN 23**, **24**, **25**, and **26**, 100 mM NaCl, 10 mM NaOAc buffer pH 6.0, dyed with SYBR Green.

5.3. fdC profiling in synthetic DNA duplexes

The fdC-probe method is intended to be implemented later in a workflow coupling to LC-MS, PCR, etc. for genomic DNA detection. LC-MS directly detects the crosslinked nucleosides whereas PCR amplifies the signal in an exponential way. We first explored the MS-based detection possibility.

5.3.1. LC-MS study with crosslinked duplex

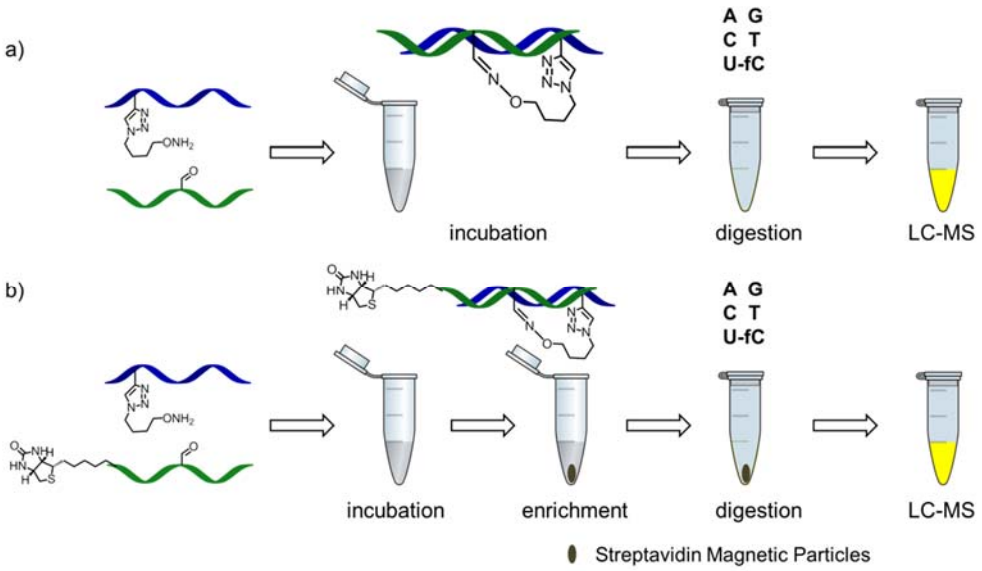


Figure 5-18 Schematic representation of LC-MS workflow of crosslinked duplex.

In the MS-mode, the 15 mer duplex **ODN 13/16e** was first digested with nuclease S1, antarctic phosphatase, and snake venom phosphodiesterase I using the conventional protocol from our group.[229] Afterward, the mixture was analyzed with oribitrap LC-MS (Figure 5-18a). Indeed, in model studies, the four canonical bases, fdC, U with hydroxylamine linker, and also the U crosslinked dinucleoside were observed (Figure 5-19). Additionally, we observed 0.5% of fdC-U dinucleoside, 0.2% of fdC, and 0.1% of deaminated dC relative to dC. The results confirm that the cross-linked oligonucleotides can be fully digested to give the dinucleoside together with others Slight deamination is also detected. This problem could be solved by adding tetrahydrouridine (THU), a deamination inhibitor known since 1968.[341]

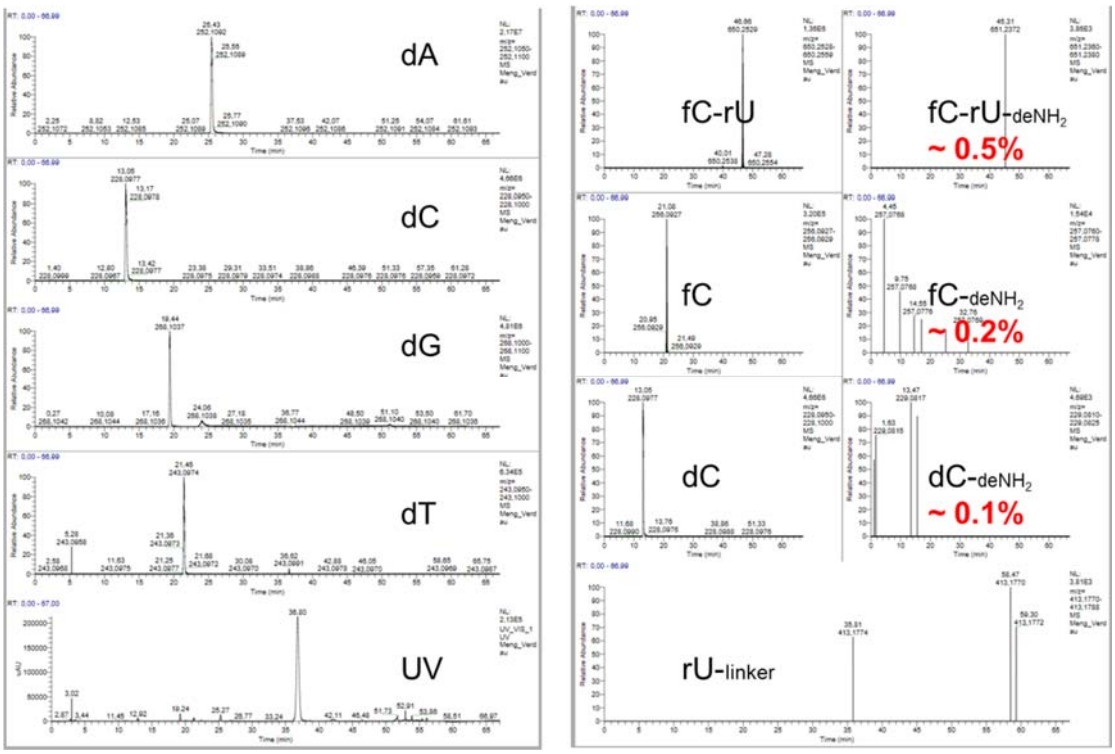


Figure 5-19 Orbitrap LC-MS spectroscopy of digested duplex **ODN 13/16e**. Condition: 0-20% buffer B in 45 min. Buffer A: 2 mM $HCOONH_4$ aq., buffer B: 2 mM $HCOONH_4$ in 80% MeCN. Retention time: dA 24.5 min, dC 13.1 min, dG 19.4 min, dT 21.5 min, fdC-U 46.7 min, fdC 21.1 min, U-linker 35.81 min.

Next, we used the biotin-avidin interaction to enrich the cross-linked probe strand and the target strand. Avidin is a protein derived from both avians and amphibians and shows a considerable affinity for biotin, or vitamin H, a co-factor that plays a role in multiple eukaryotic biological processes.[342] The interaction between biotin and avidin is highly specific with an affinity of Ka = 10^{15} M^{-1},[343] the strongest known non-covalent interaction between a protein and a ligand. Avidin and other biotin-binding proteins, including streptavidin, can bind up to four biotin molecules, making it ideal for both purification and detection strategies.

ODN 13 containing a fdC was connected to a biotin as **ODN 27**; **ODN 12a** containing a dC was connected with a biotin as **ODN 28** (Table 5-8). Streptavidin magnetic particles were added to the solution of duplex **ODN 27/16e** and **28/16e**. Digestion was carried out directly on the particles after the enrichment, with THU as additive. The mixture was then injected for orbitrap measurement (Figure 5-18b, Figure 5-20).

fdC-U dinucleoside and fdC is present in the case of **ODN 27/16e**, and absent in duplex **28/16e**, where U with linker can be observed. In both situations, deamination of fdC-U crosslinked dinucleoside and fdC are undetectable while deaminated cytosine is still present.

Table 5-8 Synthesized oligonucleotides for LC-MS study. Red letters represent to alkyne modifiers and cytosine modification. *: contain one sodium ion.

Entry	5' ---------- 3'	mer	calc.	found
27	GTA ATG **fC**GC TAG G-Biotin	13	4610.5	4608.0
28	GTA ATG CGC TAG G-Biotin	13	4582.3	4578.8
16e	CC**U-L-ONH$_2$** AGC GCA TTA C	13	4084.8*	4082.2

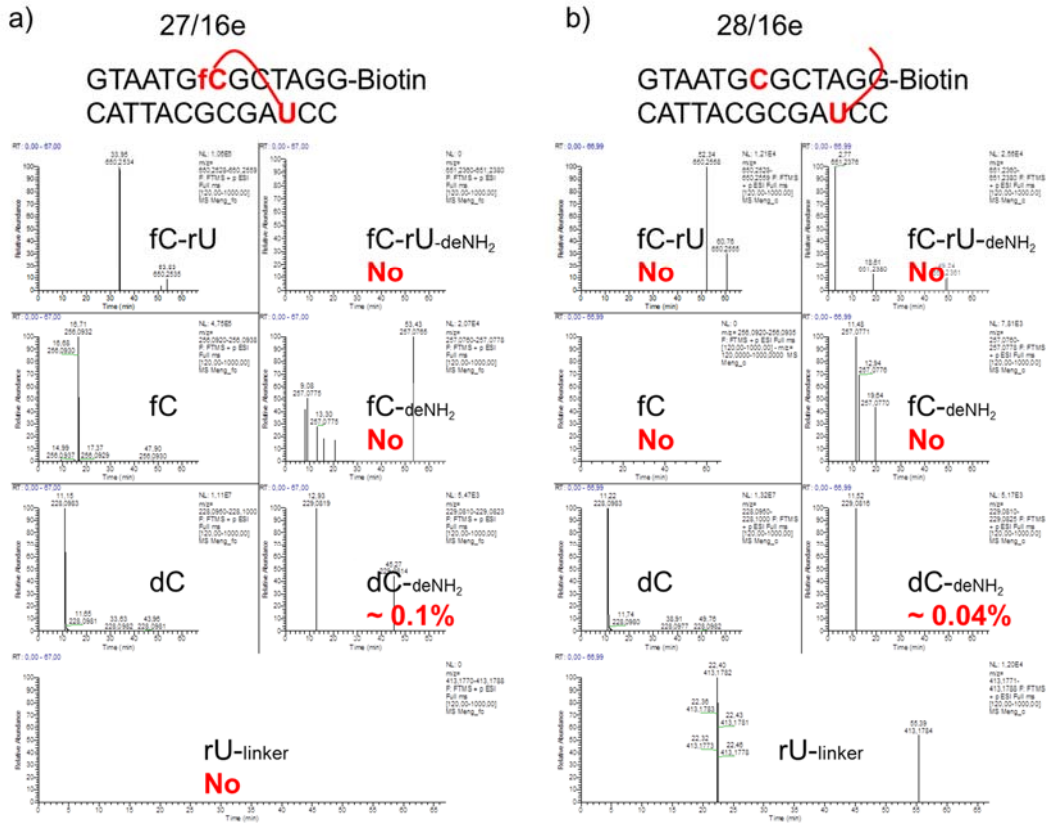


Figure 5-20 Oribtrap LC-MS spectrometry of digested duplex a) **ODN 27/16e** and b) **28/16e**. Condition: 0-30% buffer B in 45 min. Buffer A: 2 mM $HCOONH_4$ aq., buffer B: 2 mM $HCOONH_4$ in 80% MeCN. Retention time: fdC-U 34.0 min, fdC 16.8 min, U-linker 22.40 min, dC 11.2 min, deaminated dC 11.5 min.

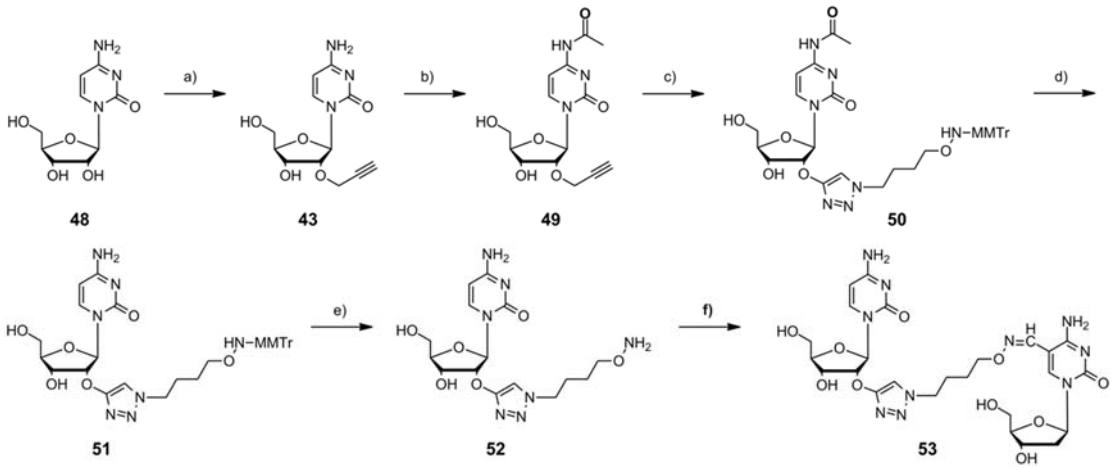
Here, we confirmed that the crosslinked duplex can be digested into the reacted dinucleoside; connected with an aliphatic linker; the dinucleoside cannot be further

digested into nucleosides. The orbitrap LC-MS has as expected an insufficient detection limit, which is not suitable for the project. Therefore the MS method was adpoted for Triple-Quad quantification.

5.3.2. Synthesis of the standard compounds

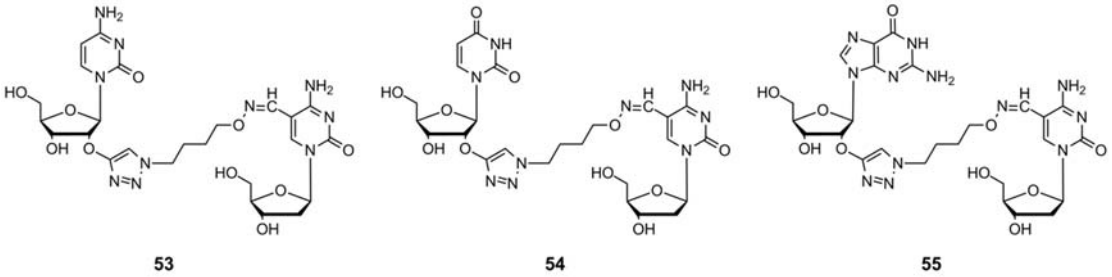
For the triple-quad MS standards, the corresponding dinucleotides and their isotope labeled counterparts had to be synthesized as standard compounds. Scheme 5-3 shows a typical synthetic route where cytosine is taken as a representative base.

The synthesis toward the crosslinked cytosine-fdC dinucleoside started from commercially available cytidine (**48**). The 5'-hydroxy group was deprotonated with sodium hydride. Intramolecular proton exchange gave a mixture of the deprotonated 2' and 3' hydroxy group intermediates. Substitution with propargyl bromide gave the 2'- and 3'-*O*-propargyl compound as a mixture, which was separated followed by acetylation and crystallization. Pure 2'-*O*-propargyl precursor **49** was in this way obtained. The 1,3-dipolar cycloadditon reaction connected the MMTr protected linker **47** to **49**. After deacetylation and detritylation, **52** was obtained after reverse phase HPLC. Connection of **52** and fdC nucleoside with 4-methoxyaniline gave the crosslinked dinucleoside **53**. Crosslinked uridine-fdC **54** and guanine-fdC **55** dinucleotides were synthesized similarly (Figure 5-21).



Scheme 5-3 Synthesis of the crosslinked cytidine-uridine dinucleoside **53**. a) tetrabutylammonium iodide, NaH, propargyl bromide, DMF, 0 °C; b) acetic anhydride, DMF, rt, 33% over 2 steps; c) **47**,

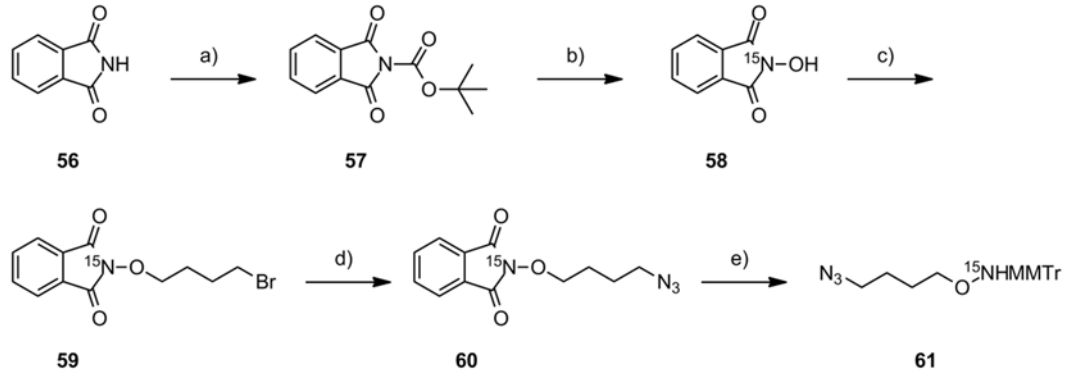
$CuSO_4$ $5H_2O$, sodium ascorbate, DIPEA, H_2O, MeCN, rt, 99%; d) NH_3, MeOH, rt; e) HOAc, H_2O, 25 °C, quant.; f) fdC nucleoside, 4-methoxyaniline, MeCN, H_2O, 25 °C, quant.



53 54 55

Figure 5-21 Synthesized canonical base-crosslinked fdC dinucleotides.

In order to obtain the isotope labeled dinucleotides, an aliphatic linker with ^{15}N was first synthesized. As depicted in Scheme 5-4, phthalimide **56** was first amidated to **57**. With ^{15}N labeled hydroxylamine hydrochloride, the nitrogen atom was subsituted with its heavy isotope in **58**. Isotope-labelled linker **61** was obtained following a procedure shown in Scheme 5-1. When sodium azide-1-^{15}N (Scheme 5-3, step d) is used for substitution, an aliphatic linker with double isotope labeling can be obtained.



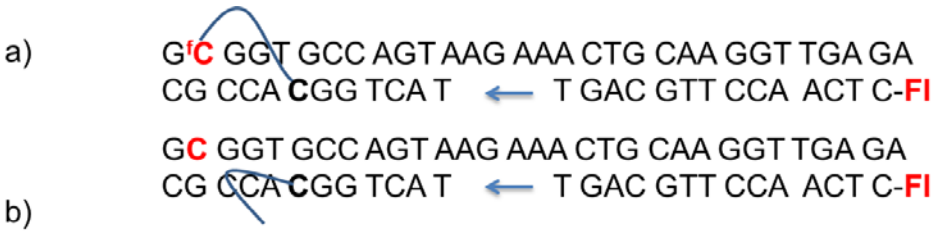
56 57 58 59 60 61

Scheme 5-4 Synthesis of isotope labeled linker **61**. a) di-tert-butyldicarbonate, DMAP, MeCN, rt, quant.; b) hydroxylamine-^{15}N hydrochloride, NaOH, H_2O, MeCN, rt, quant.; c) *N*-hydroxyphthalimide, triethylamine, DMF, rt; d) NaN_3, DMF, rt, 64% for 2 steps; e) i) $N_2H_4 \bullet H_2O$, DCM, rt, ii) MMTr, DIPEA, DCM, rt, 72%.

5.3.3. Primer extension study with crosslinked duplex

Towards coupling the fdC probe with PCR, we started with a model duplex to check if the crosslinker will block the primer extension reaction. A 31 mer template **ODN 29a** containing a fdC base, or **ODN 29b** with a cytosine, a 12 mer probe **ODN 24**, and a

14 mer primer **ODN 30** with a fluorescein at the 5' terminus were prepared (Figure 5-22).



Entry	5' ---------- 3'	mer	calc.	found
24	TACTGGC-L-ONH$_2$ACCGC	12	3810.9*	3812.0
29a	GfCG GTG CCA GTA AGA AAC TGC AAG GTT GAG A	31	9681.8	9679.2
29b	GCG GTG CCA GTA AGA AAC TGC AAG GTT GAG A	31	/	/
30	Fl- CTC AAC CTT GCA GT	14	/	/

Figure 5-22 Primer extension study with crosslinked template: a) schematic representation of the primer extension experiment; b) synthesized and purchased oligonucleotides for this study. Bold and red letters represent alkyne modifiers, fC, and fluorescein. Found values indicate the mass results measured by MALDI-TOF. *: contain one sodium ion.

First，the **ODN 29a** was crosslinked with the probe **ODN 24** and used as a template for primer extension. An extended primer can be seen from the PAGE (Figure 5-23, mixture 2 and 10 min). Because of the incomplete crosslinking of the probe and the template, the extended primer may result from the uncrosslinked **ODN 29a** or the **ODN 29a-**anisidine conjugate. The crosslinked template was then purified using RP-HPLC; the pure crosslinked duplex was obtained without anisidine and the fdC-template remained. No extended primer could be observed after 10 min at 60°C with Taq polymerase (Figure 5-23, purified 2 and 10 min). In contrast, on the template containing C (**ODN 29b**) mixed with fdC-probe, the primer extended to full length after 2 min.

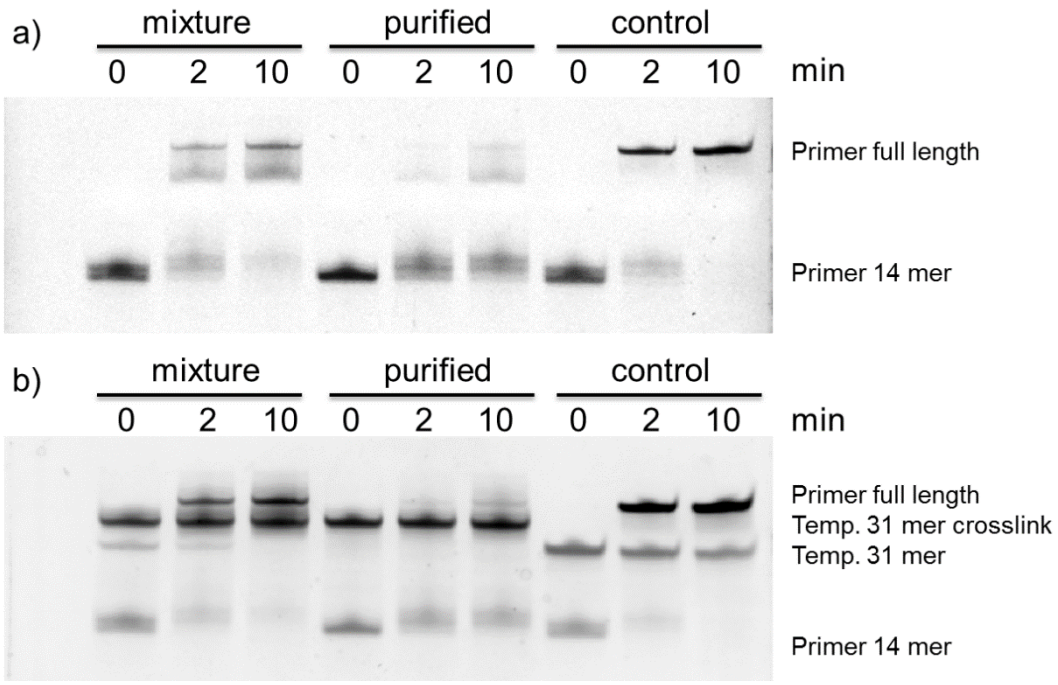


Figure 5-23 Primer extension study on the model duplex. After PAGE assay, the gel was photographed a) directly for 6-FAM or b) after *SYBR* Green I dyeing. Condition: 1.5 μM crosslinked **ODN 24**/**29a** before and after purification and **24**/**29b**, 1.0 μM primer **30**, 200 μM dNTP, 1× Thermopol buffer, and 5 U Taq polymerase, the final volume 40 μL, 60°C.

While these experiments were performed, K. Gates showed Bacteriophage Φ29 DNA polymerase was blocked by a counter-strand tied by dG-Ap and dA-Ap crosslinks.[131] They believed that the active site entry tunnel in Φ29 is created by a dynamic flap that can open to allow substrate access. If the tunnel is a rigid structure, the interstrand crosslinks will have stalled primer extension several nucleotides prior to the cross-link lesion, rather than at the last unmodified nucleotide. These explain the reason that the fdC probe is blocked at 15 or 16 mer.

The crosslinking efficiency can never be 100%. However, in the PCR reaction, an excessive of probe will be used. The unreacted template, if any, should be negligible.

Second, we tested if the fdC-anisidine conjugate and the free anisidine will inhibit the Taq polymerase. **ODN 29a** was incubated with catalytic 4-anisidine to form the conjugation. As shown in Figure 5-24 line 3, fdC-anisidine alone does not block replication, which is in accordance with results from Wang *et al.*[324] and He *et al.*,[271] who showed that DNA pol I can go through 9-(hydrazinylmethyl)-acridine-fdC and (4-aminomethyl)-benzyl-azide-cadC adducts on DNA template. Moreover, with a

concentration less than 500 μM, 4-anisidine makes little effect on the polymerase (Figure 5-24, line 4-7).

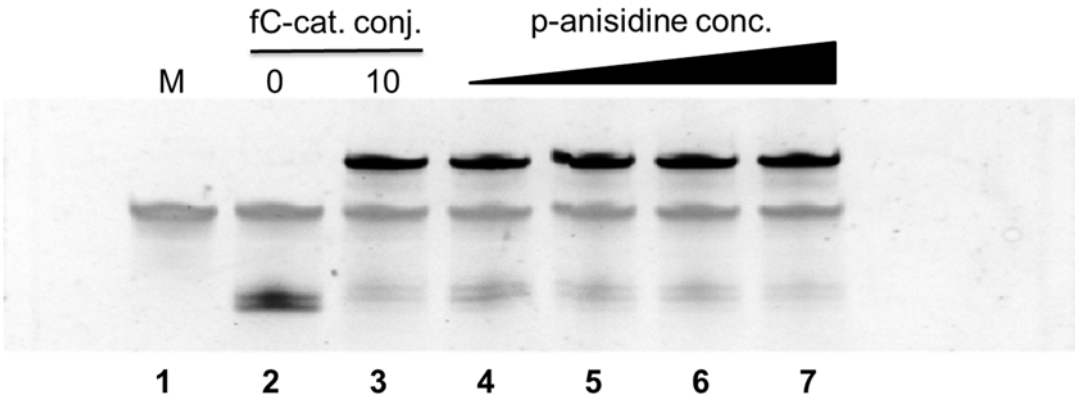


Figure 5-24 Primer extension study with fdC-conjugated and free 4-anisidine. Line 1: **ODN 29b** as a marker. Line 2, 3: **ODN 29a-cat.** as a template for replication in 10 min. Line 4-7: using **ODN 29a** as a template for extension for 10 min with 0, 5, 50, and 500 μM 4-anisidine. The gel was photographed after *SYBR* Green I dyeing. Condition: 1.5 μM template, 1.0 μM primer **30**, 200 μM dNTP, 1× Thermopol buffer, and 2.5 U Taq polymerase, the final volume 20 μL, 60°C.

Third, in the PCR cycle, the genome is heated up to 95°C for 2~10 min to denature the genome and activate the polymerase. Linker stability was therefore tested to see if the covalent oxime bond will be disrupted under this harsh condition. Purified crosslinked **ODN 14-29a** was dissolved in Thermopol buffer, and incubated at 95°C for 20 min. PAGE assay showed no sign of degradation (Figure 5-25). Therefore, it is believed that once the probe is crosslinked to the fdC sites in the target strand, the oxime bond is irreversible and will not detach in the PCR cycle.

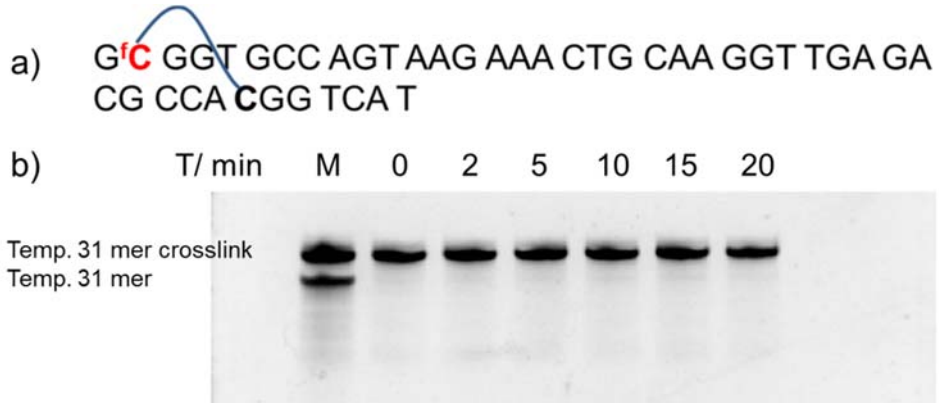


Figure 5-25 Crosslinker stability experiment: a) schematic representation of the crosslink **ODN 14-29a**; b) denaturing PAGE assay shows crosslink along the time. Mark (M) shows crosslink **ODN 14-29a** and **29a**. Condition: 1.5μM crosslink **ODN 14-29a**, 1× Thermopol buffer, 95°C

5.4. fdC profiling in genomic DNA

5.4.1. Profiling Strategy: design and validation

Quantiative detection of fdC in genomic DNA is much more challenging than in the synthetic DNA duplex. The task is divided into two steps. Figure 5-26 shows the mind map.

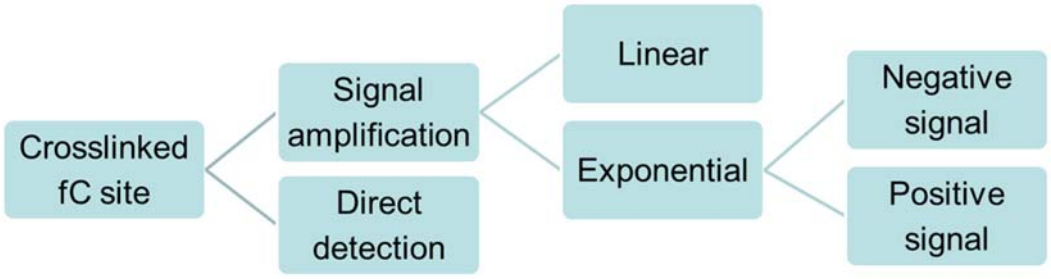


Figure 5-26 Mindmap of trial and error towards a probe-assistant loci-specific fdC detection strategy.

The first step is to react the probe with the gDNA, a "mega-biomolecule". To our best knowledge, only a few examples can be compared to our case: in PCR and fluorescence in situ hybridization (FISH), oligonucleotides, primers or probes, are hybridized with the gDNA. Padlock probes[344] require the probe to first hybridize to the target strands. Then the head and the tail are joined with ligase, creating circular DNA molecules complementary to the target sequence.

In theory, counter-strand oligonucleotides that recognize DNA sequences of at least 16 mer would have no off-target effects in the human genome or other higher eukaryotic genomes because the complexity of 16-bp sequences ($4^{16} = 4.3\times10^{9}$) is greater than the size of the haploid human genome (3.2×10^{9}). Therefore, here we designed the fdC probes longer than a 20 mer.

The second step is to distinguish the fdC from the canonical cytosine base with the help of the probe. As mentioned in Chapter 5.3, LC-MS detects fdC signal without amplification. In a volume of 20 μL solution containing 100 ng gDNA, assuming that 10% dC at one site has changed to fdC, the fdC concentration would be 250 aM (2.5×10^{-16} mol/L), coresponding to less than 50,000 molecules in total. While the Triple-Quadupole-MS method provide in principle that sensitivity, we thought that

amplification by PCR would ba prefect strategy.

Many linear amplification strategies for DNA detection and other biomolecules have been developed and are summarized by I. Willner *et al.* Among these strategies, the lowest detection limit is 10 fM of target DNA.[345] But still, this detection level is higher than what is need in our project.

At this stage, the fdC signals should be amplified in an exponential way, i.e. PCR. Either the fdC-containing strand is amplified, which provides a positive signal, or the non-fdC-containing strand is amplified, which given in total negative signal. After extensive optimization, a system is designed as depicted in Figure 5-27 a-c with sequence details. The synthetic oligonucleotides used in this part are listed in Table 5-9.

Table 5-9 Synthesized and purchased oligonucleotides for position 1 detection. Bold and red letters represent nucleoside modifications or functional group. G, alkyne modifier with the linker; p, phosphate group at 5' terminus; F, FAM; H, HEX; Q, BHQ-1. MALDI-TOF: **31a**, calc. 5505.9, found 5500.3; **32**, calc.7951.5, found 7948.5, contain one sodium ion.

ODN	Description	5' ---------- 3'	mer
31	fdC target	TAAAAAACTAGGAGCCAGGGGCCTCAAAACCAGCCCTGCGATTTTATfCGGTCATC	55
31a	Part I	GCGATTTTATfCGGTCATC	18
31b	Part II	TAAAAAACTAGGAGCCAGGGGCCTCAAAACCAGCCCT	37
31c	ligation bridge	TAAAATCGCAGGGCTGG	17
32	fdC probe	GAGATGACCGATAAAATCGCAGGGC	25
33	reporter strand	pTGGTTTTGAGGCCCCTGGTGTAGCTTGCTGTGGCGCAGGGACTCTGTGATGTCTGGAGCGATTCATTCTC	70
34	reporter-probe	ACCGATAAAATCGCAGGGCTGGTTTTGAGGTAGCTTGCTGTGGCGCAGGGACTCTGTGATGTCTGGAGCGATTCATTCTC	80
35	detection forward primer	CCGATAAAATCGCAGGGCTG	20
36	detection reverse primer	GAGAATGAATCGCTCCAGAC	20
37	detection TaqMan probe	F-ACAGAGTCCCTGCGCCACAGCAAG-Q	24
38	reference forward primer	GGCTGGACTCAAGCAACTAA	20
39	reference reverse primer	ACACACCTCTGTCCTCAGAT	20
40	reference TaqMan probe	H-CGGTTTAGTTTGGAGATGACCGAT-Q	24

We focused on the 30,020,539th nucleoside of chromosome 16 (position 1) from *Mus musculus* genome (MM9) in the exon 3 of 632428C04Rik, where it is claimed to contain 23% of fdC in a previous report using the redBS-Seq method.[268] A 25 mer fdC probe (**ODN 32**) was first incubated with the genomic DNA. A ratio of 1.2 ng gDNA and 2 pmol probe in a volume of 20 µL ensured complete hybridization; *p*-phenylenediamine was chosen as a catalyst for the benefit of crosslinking yield.

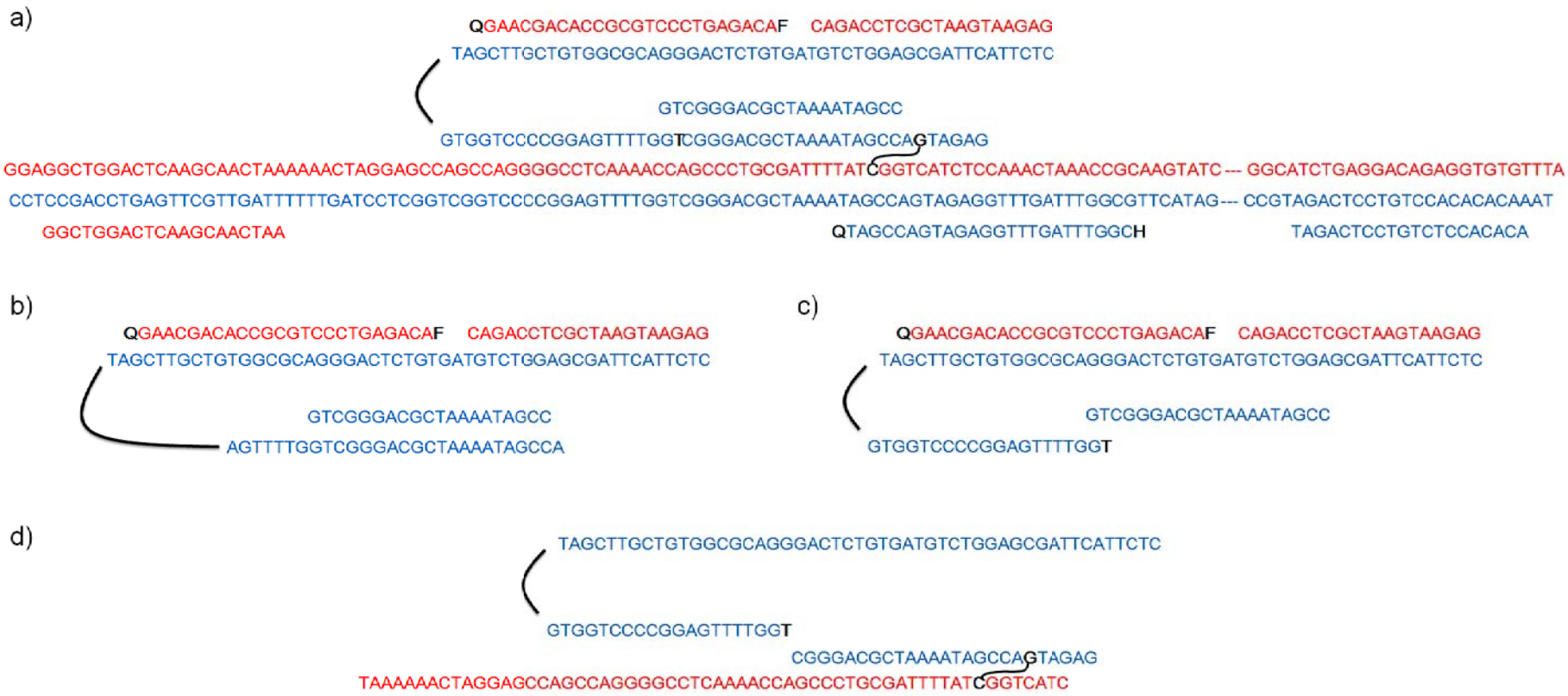


Figure 5-27 Illustration of the detection strategy for 30,020,539th site of chromosome 16 (MM9): a) sequence detail of the selected genomic DNA, fdC probe, reporter strands, detection primer pair, reference primer pair, and the respective TaqMan probes; b) sequence detail of postive control experiment; c) sequence detail of negative control experiment. Forward and reverse strands or primers are differentiated by red and blue; d) sequence detail of ligation optimization experiment. Letters in black and bold repersent nucleoside modifications or function groups: F, FAM; H, Hex; Q, BHQ-1; T, 5' terminus T with additional phosphate group; G, alkyne modifier with hydroxylamine linker; C, fdC. Hyphen represent the sequence omitted.

Subsequently, a 70 mer reporter strand (**ODN 33**) with a phosphate group at the 5'-terminus was added together with a Taq Ligase or Ampligase to connect to the fdC probe. A part of the reporter strand is complementray to the genome sequence nearby the probe while the other part is not hybridizing to the genome and facilitates signal amplification including a reverse primer and a TaqMan probe sequence. Apart from the TaqMan probes, molecular beacons, scorpions, dual-labeled probes, etc. could be used for sequence detection in PCR as well.[178] TaqMan is chosen here for design simplicity and increased sensitivity.

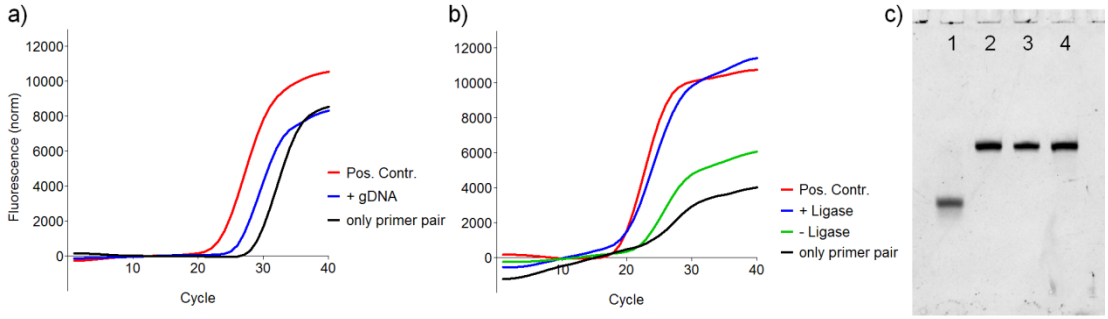


Figure 5-28 Real-time PCR curves and crosslink stability for idea validation: a) real-time PCR of detection primer pair with ligated reporter strand as positive control, with or without gDNA as negative control, using SYBR Green Supermix; b) rtPCR of detection primer pair with ligated reporter strand as positive control, with Taq ligase as experiment entry, without Taq Ligase or gDNA as negative control, using TaqMan probe. c) crosslinked duplex **31**/**32** after ligation cycle (line 4); line 1, **ODN 32**; line 2, template duplex **31**/**32**; line 3, **ODN 31**. Ligation condition: 95°C for 3 min followed by 10 cycles of 94°C for 1 min and 60°C for 1 h; 10 U Taq Ligase, 1 pM template duplex **31**/**32**, 265 nM reporter strand **ODN 33**, the final volume of 20 µL. PCR condition: 0.2 pM template duplex **31**/**32**, 0.2 pM ligated reporter strand **ODN 34**, 900 nM each primers **ODN 35**,**36**, 80 ng gDNA, 250 nM TaqMan probe **ODN 37**, 5 U Taq polymerase, 200 µM each dNTP, final volume of 20 µL; annealing temperature 60°C for SYBR Green, 64°C for TaqMan probes.

After ligation, unreacted fdC probe, catalyst, salts, etc. were removed using a PCR purification Kit from *NEB*. A forward primer (**ODN 35**) which binds to the fdC probe with 18 nt and left 2 nt overlap on the reporter strand, and a reverse primer (**ODN 36**) which is orthogonal to the genomic DNA and pairs with the reporter strand are used. The two primers without or with gDNA showed little amplification on real-time PCR (Figure 5-28a, black and blue line), neither with the reporter strand (Figure 5-28b,

green line). The 2-nt overlap of the reporter strand is not enough for the forward primer to anneal. Only if the reporter strand ligates with the fdC probe, the forward primer can hybridize and work together with the reverse primer and the TaqMan probe (**ODN 36**, **37**). The primers and TaqMan probes designs are based on Primer3Plus,[346] Mfold and general rules.

Ligation conditions were scanned and optimized using a synthetic strand (**ODN 31)** containing one fdC (Figure 5-27d). Different conditions were tried with 1 pM crosslinked dsDNA and 265 nM reporter strand. After ligation, rtPCR (Figure 5-28b) was conducted to check the product compared with the positive control (**ODN 34,** Figure 5-27c). To reach a complete conversion, denaturing and annealing/ligation cycles should be executed. With Taq DNA Ligase or Ampligase, 20 cycles of 94°C denaturation for 1 min followed by 60°C reannealing and ligation for 1 h are sufficient for completely ligation according to rtPCR results, similar to reported protocols.[321, 347] Figure 5-28b shows that the experiment entry gave a C_t number comparable to the positive control. Besides, the crosslinker is stable enough to endure the ligation cycles (Figure 5-31c).

The ligated reporter strand generates the detection signal while the amplicon on gDNA itself provides the reference signal (**ODN 38-40**, Figure 5-27a). Although the probe will block the reverse primer in the first cycle, the forward primer will synthesize an uncrosslinked template to accommodate the reverse primer in the second cycle. Detection signals and reference signals will be quantified using droplet digital PCR (ddPCR). Figure 5-29 shows the workflow of the method.

Before discussing the signal quantification in gDNA, it is necessary to introduce the ddPCR.

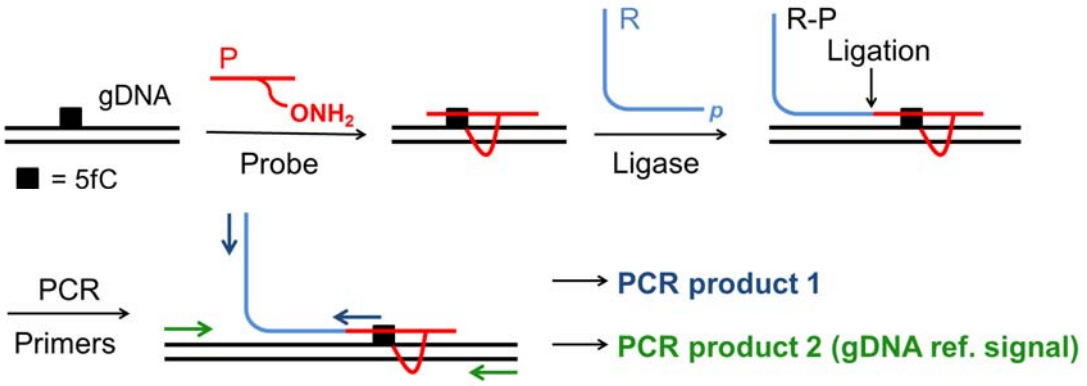


Figure 5-29 Schematic representation of the probe-based fdC detection workflow. Two strands of the duplex are differentiate by red and black lines; black segments represent fdC probe; gray and purple segment represent reporter strand; arrows represent primers; segments with end points represent TaqMan probes.

5.4.2. Droplet digital PCR: introduction and application

5.4.2.1. ddPCR outperform rtPCR

Real-time PCR (rtPCR), is the method that is used for quantification in the PCR experiments.[178] In real-time PCR, the amplicons are analysed after each cycle. Dyes or hydrolysis probe is added into the ordinary PCR mixture. By monitoring fluorescence signals during the exponential amplification phase of the reaction, the initial quantity of the target can be determined with high precision.

The threshold cycle (C_t) is the cycle number at which the fluorescent signal of the reaction crosses a certain threshold. Amplification occurs in an exponential manner with a base of 2 if the amplification efficiency is assumed to be 100%. So a difference of 10-fold in copy number results in a C_t value difference of $\log_2 10 = 3.32$. However, if only 20% of the template contains fdC while 80% contains non-fdC at the same position, the copy ratio of detection and reference would be 0.2 to 1. Thus, C_t difference will be $\log_2 (1/0.2) = 0.30$. In fact, according to S. Balasubramanian,[268], the maximum probability of a particular fdC site is less than 40% compared to cytosine, and the level should be below 0.002% corresponding to cytosine in the whole genome. A C_t difference of 0.30, corresponding to 20% fdC at one locus, cannot be resovled in rtPCR. Therefore, a novel PCR method is necessary.

Droplet digital PCR (ddPCR) came up in 2011.[348] Instead of analysing the PCR process, ddPCR measures the absolute quantity by counting nucleic acid molecules encapsulated in discrete water-in-oil droplet partitions. Figure 5-30 shows the ddPCR workflows. First, the reaction mixture is prepared as in rtPCR, either with EvaGreen or using TaqMan probe contains FAM, HEX or VIC. Second, uniform droplets are generated from the reaction mixture. Each microliter (μL) generates 1000 droplets. The target and background DNA samples are distributed randomly into the droplets together with the primers, the TaqMan probe, dNTP, and the polymerase. The process obeys the Poisson distribution. Third, the droplet is transferred to a PCR plate, and PCR cycle is conducted in the same manner as conventional PCR. Last, the PCR plate is placed in a droplet reader machine, which analyzes each droplet using a double wavelength detector.

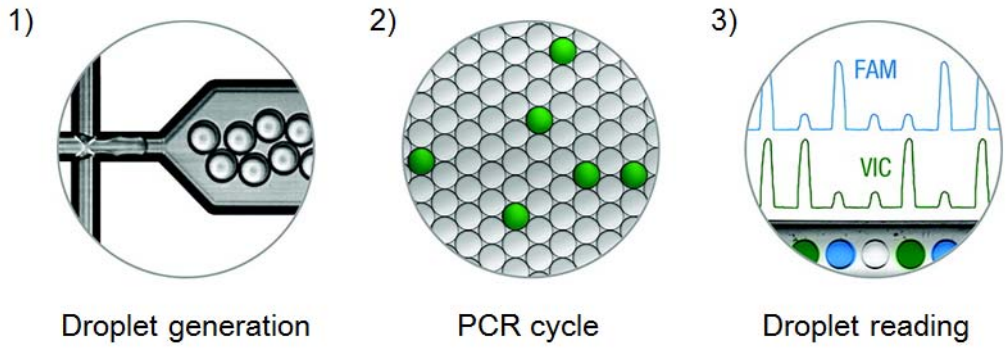


Figure 5-30 Schematic representation of the ddPCR workflow. The figure is adapted from reference.[348]

More and more research taking advantage of the ddPCR principle came out in the past three years. A recent report applied the drop-based microfluides for the isolation and sequencing of the viral genome.[349] Compared to the previous viral detection and sequencing method, droplet-based microfluidics followed by whole-genome amplification is a simple and applicable method to unknown viruses. This platform enables efficient isolation of single viral species from a mixture of other viruses and DNA contaminants, which was not feasible before.

Compared to rtPCR, ddPCR achieves absolute quantification of the amplicons with high accuracy. In this project, ddPCR is applied to develop a probe-based fdC

detection method. The ligation condition will be further optimized for the ddPCR method.

5.4.2.2. ddPCR: optimization and quantification modeling

The encapsulation maximum of one target amplicon in one droplet to generate a positive or negative signal is the ideal scenario for our situation. If a droplet contains more than one detection amplicon, for example, one contains one fdC and one cytosine, it will show a positive signal, and the negative cytosine signal vanishes.

The probability that two or more detection amplicons getting into one droplet can be calculated according to the Poisson distribution, i.e. a discrete random variable X complies the Poisson distribution with parameter $\lambda > 0$, if, for k = 0, 1, 2, …, the probability mass function of X is given by:

$$f\ (k;\lambda) = \Pr\ (\mathrm{X} = k) = \frac{\lambda^k e^{-\lambda}}{k!}$$

where e is Euler's number and *k*! is the factorial of *k*.

Table 5-10 Poisson distribution probabilities of genome copies in the droplet.

input ng	λ	*k*				
		1	2	3	4	5
3	0.05	4.8%	0.1%	0.0%	0.0%	0.0%
6	0.10	9.0%	0.5%	0.0%	0.0%	0.0%
9	0.15	12.9%	1.0%	0.0%	0.0%	0.0%
10	0.17	14.1%	1.2%	0.1%	0.0%	0.0%
15	0.25	19.5%	2.4%	0.2%	0.0%	0.0%
20	0.33	23.9%	4.0%	0.4%	0.0%	0.0%
30	0.50	30.3%	7.6%	1.3%	0.2%	0.0%
40	0.67	34.2%	11.4%	2.5%	0.4%	0.1%

For example, the mass of a mouse genome is approximately 3.0 pg (3.0×10^{-12} g). If 30 ng for a 20 μL reaction is used, 10,000 genomes will be distributed into 20,000 droplets. So, λ equals to 10,000 / 20,000 = 0.50. Let X = 1, then f (1; 0.50) = 0.303; let X = 2, then f (2; 0.50) = 0.076. This means 30.3% of the droplets, instead 50%, of the

droplet, contain a single copy while 7.6% of the droplets contain two copies. Extensive distribution probabilities are listed in Table 5-10. If less than 9 ng is settled in a 20 µL reaction, the probability to have two copies inside one droplet will be lower than 1%. For the ease of calculation, 6 ng gDNA is input in each reaction of 20 µL, corresponding to ca. 90 copy/µL.

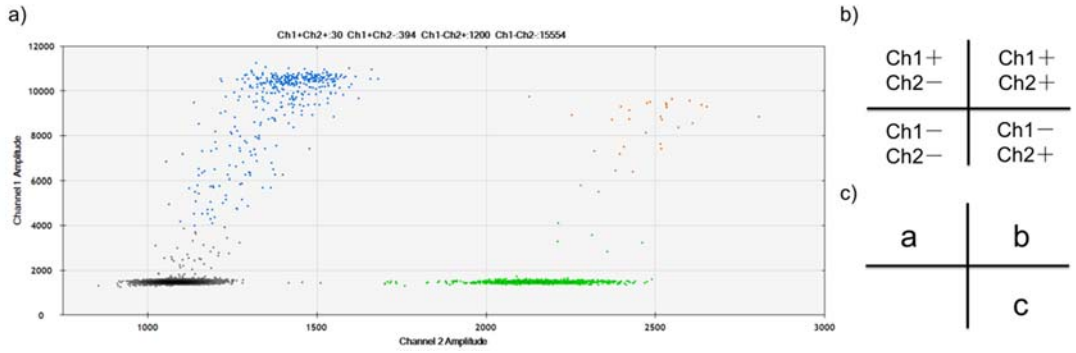


Figure 5-31 ddPCR output and modelling: a) 2-D plot of droplet fluorescence for illustration; b) clusters separation in four quadrants; c) algebraic simplification of counting numbers of the clusters.

As shown in Figure 5-31, for ease of nomination, Ch1+Ch2+ (yellow) refers to droplets with both positive signals; Ch1+Ch2− (blue) refers to droplets with only detection (reporter strand) signal; Ch1−Ch2+ (green) refers to droplets with only reference (gDNA) signal; Ch1−Ch2− (black) refers to droplets without target locus; AD refers to all the droplets accepted; resolution refers to the separation of the clusters.

In principle, Ch1+Ch2+ shows the droplets that containing fdC site in the target locus; Ch1+Ch2− indicates false-positive due to unspecific amplification; Ch1−Ch2+ shows the droplets contain non-fdC sites. Ideally, the ratio Ch1+Ch2+ to Ch2+ should be dependent on the gDNA sample. In reality, however, how the gDNA is treated and how the PCR is conducted impacts the resolution and signal to noise ratio. Therefore, it is crucial to optimize crosslinking, ligation and PCR conditions to separate the four clusters well.

A series of fdC probe concentrations were tested with the same reporter strand concentration using $Tdg^{-/-}$ cells as a positive control and Dnmt TKO cell where there should be no fdC as a negative control. The results are judged by the ratio Ch1+Ch2+

to Ch2+. The background noise arose with increasing amount of fdC probe. To get a lower noise and good differentiation between the positive and negative control; also consider the amount of probe available, a 100 nM concentration was chosen for the further experiment.

From these experiments, it is concluded that ligation will not break the crosslinker (Figure 5-28c); Ampligase is superior to Taq Ligase in both specificity, i.e. Ch1+Ch2+ / Ch1+Ch2−, and efficiency, i.e. Ch1+Ch2+ / Ch2+; ligation temperature should approach melting temperature of the reporter strand hybridization part; dilution of the ligation reaction mixture makes no difference in ligation specificity and efficiency.

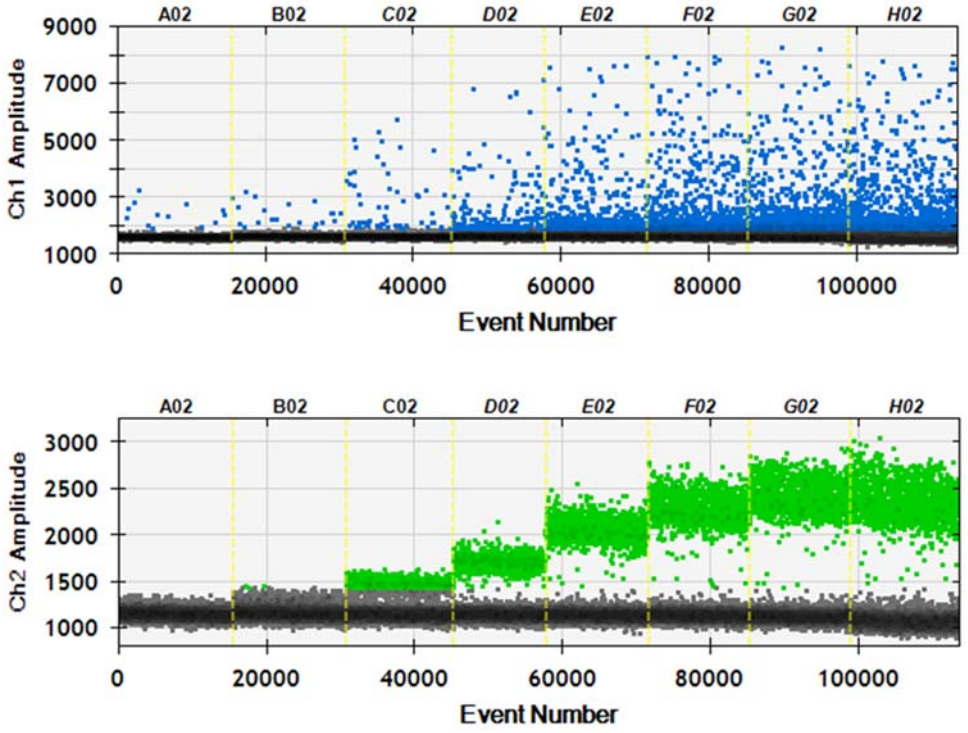


Figure 5-32 1-D plot of droplet fluorescence of detection primer pair (Ch1) and reference primer pair (Ch2) at 68-60°C. Well A 68.0°C, B 67.4°C, C 66.4°C, D 64.9°C, E 63.1°C, F 61.6°C, G 60.6°C, H 60.0°C. PCR was conducted for 35 cycles.

The lower the PCR annealing temperature was set, the more background noise in Ch1 but the better resolution in Ch2 (Figure 5-32) was obtained. 64°C was finally chosen as annealing temperature. When increasing the numbers of PCR cycles, the higher the Ch2 resolution; the faster the PCR ramps, the higher the resolution in Ch1 and Ch2 (Figure 5-33). 2°C/s and 1 nM reporter strand were chosen for an ideal outcome. Ch1

primer amplification promotes Ch2 primer replication in droplet numbers slightly and lowers Ch2 resolution (Figure 5-34). For the same sample resulted from the two-step reaction from replicates the Ch1+Ch2+ / Ch2+ ratios are stable and reproducible.

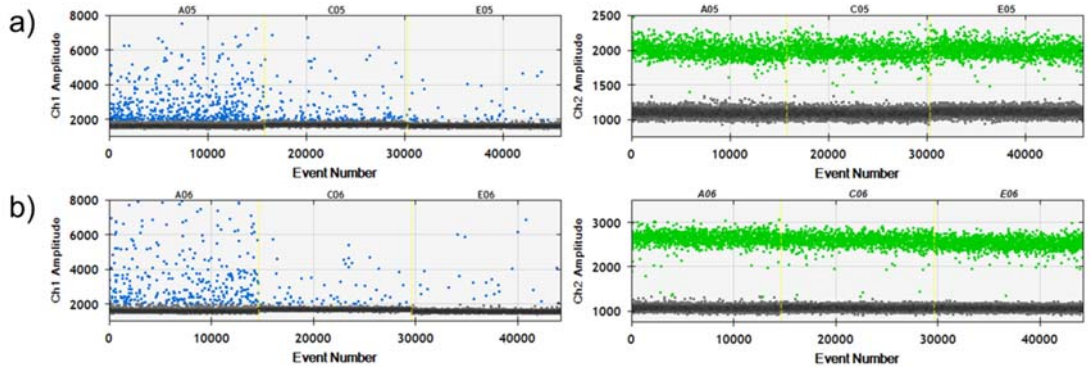


Figure 5-33 1-D plot of droplet fluorescence of different temperature ramping with varied reporter strand concentration: a) 0.4°C/s; b) 2°C/s. Reporter strand concentration: well A, 2 nM, well C, 1 nM; well E, 0.5 nM. PCR was conducted at 64°C for annealing and 35 cycles.

a)

	Ch1	Ch2	1+2+	1+2-	1-2+	1-2-	AD
Ch1,2	4.8	91	30	33	1128	14323	15514
	4.2	91	33	22	1110	14139	15304
Ch1	4.2	0	0	49	0	13544	13593
	3.4	0	0	43	0	14841	14884
Ch2	0	80.6	0	0	1088	15342	16430
	0	80.2	0	0	1145	16236	17381

b)

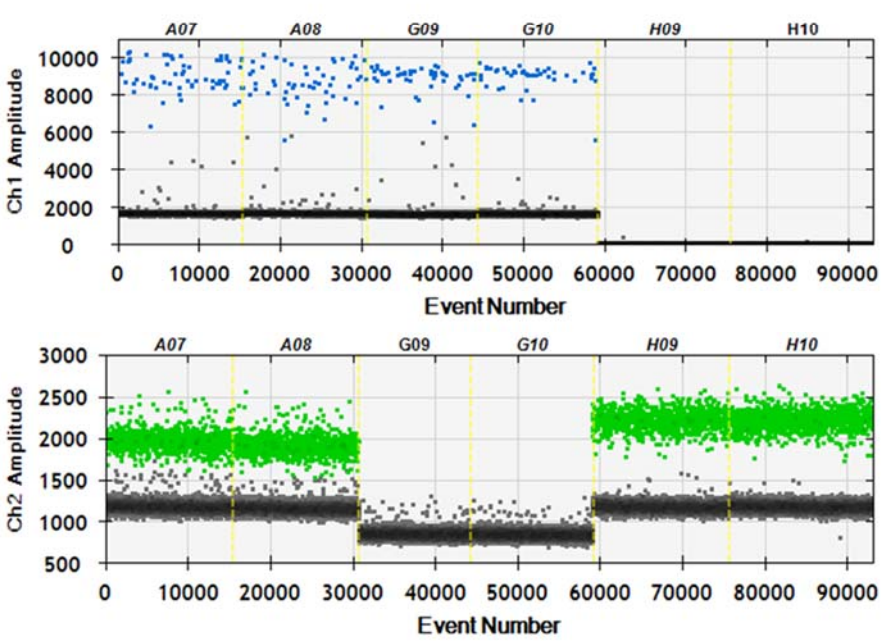
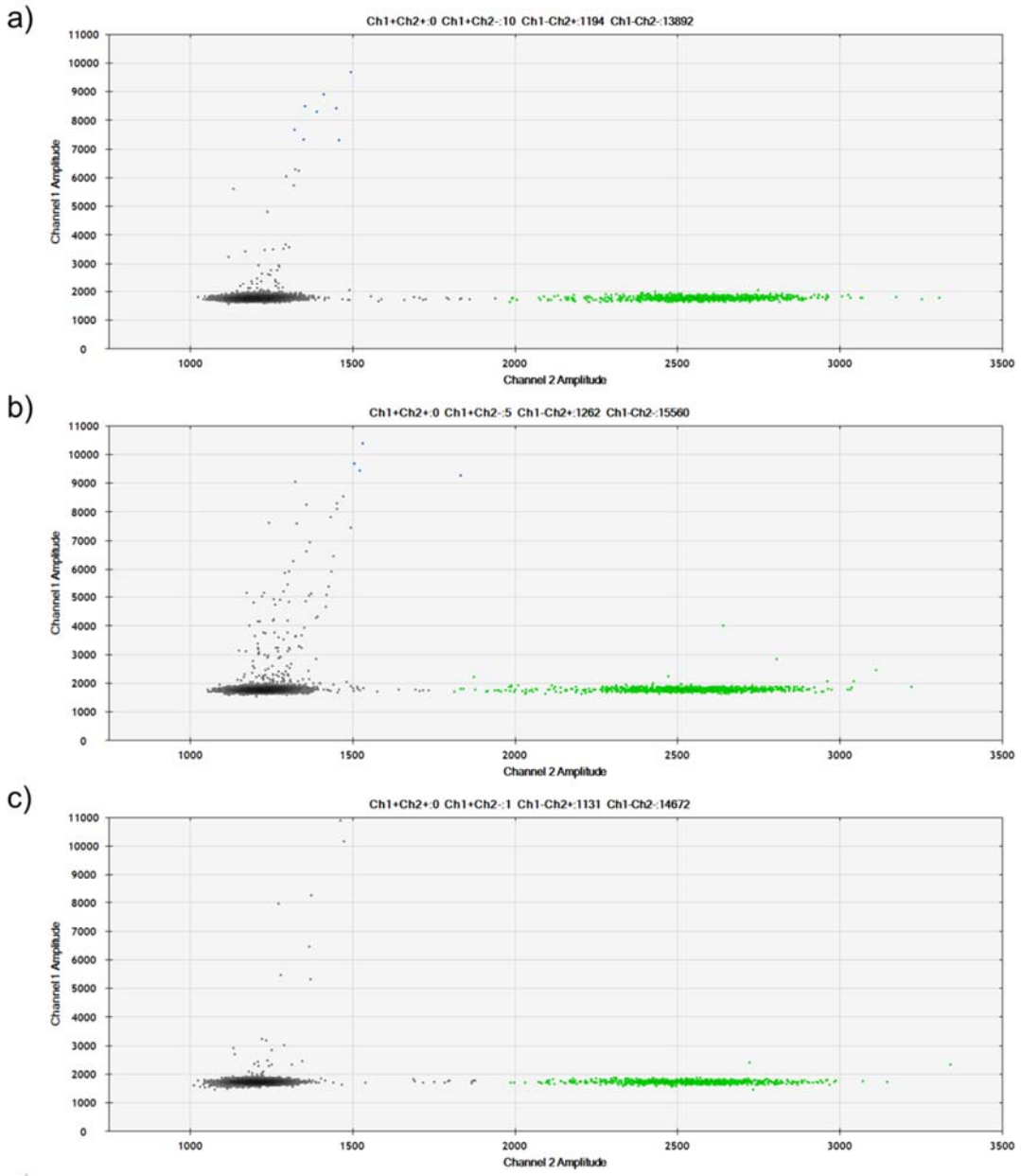


Figure 5-34 Double and single primer pair test: a) raw numbers of ddPCR results; b) 1-D plot of droplet fluorescence showing double primer pairs (well A), only detection primer pair **ODN 35**/**36** (well G), only reference primer pair **ODN 38**/**39** (well H).

Positive and negative controls were preformed on ddPCR. For negative control, a probe of position 2 (**ODN 41**, Chr 15, 8,846,677th, MM9, see Chapter 5.4.3) instead of probe **ODN 32** was used (Figure 5-35a); no reporter strand **ODN 33** (Figure 5-35b) or no ligase (Figure 5-35c) was added. None of the three negative controls showed a Ch1+Ch2+ cluster, confirming that the correct combination of probe for the right position, the reporter strand and the ligase are the prerequisites for the detection.

When a ligated reporter strand **ODN 34** was added to the reaction mixture, which served as a positive control, Ch1+Ch2+ and Ch1+Ch2− clusters appeared.



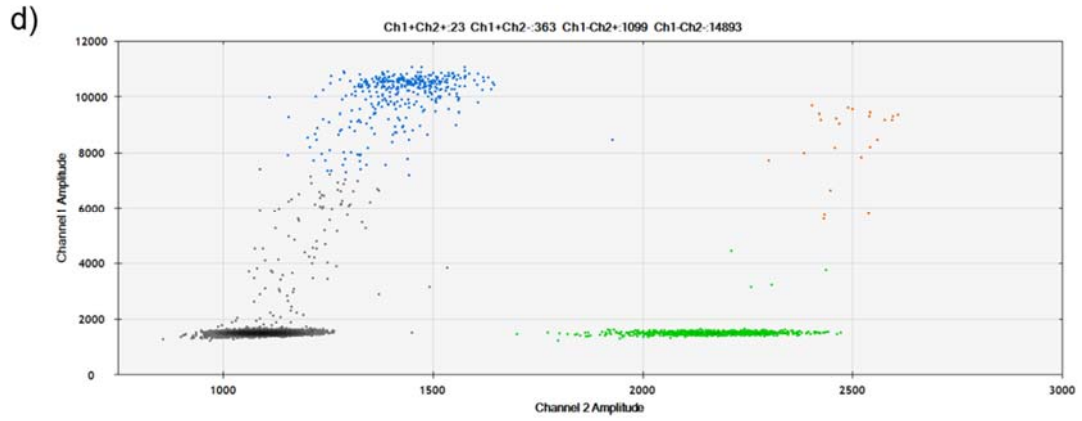


Figure 5-35 2-D plot of droplet fluorescence for negative and positive control of position 1: a) probe **ODN 41** was used; b) no reporter strand **ODN 33**; c) no ligase; d) positive control. Condition: 80 U Taq Ligase, PCR extended at 64°C, 40 cycles.

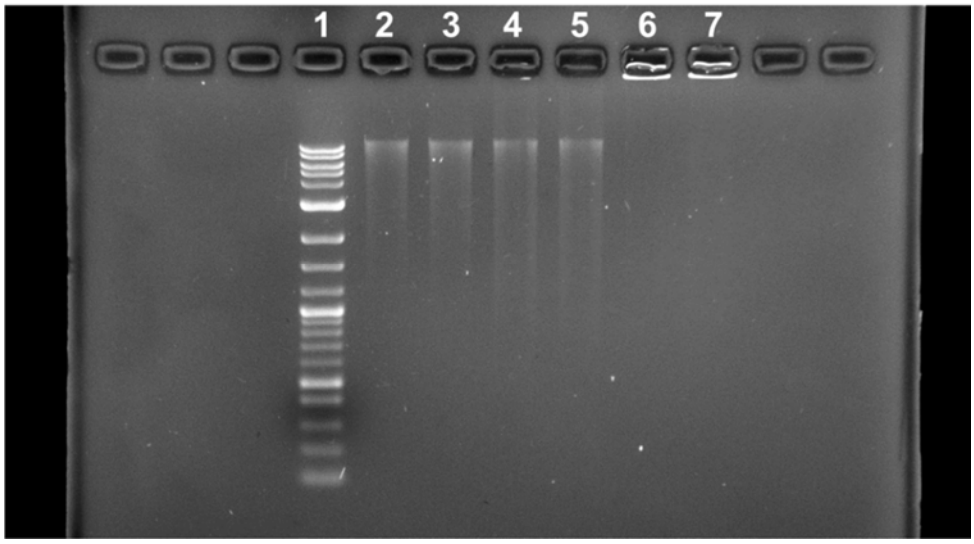


Figure 5-36 Agarose gel shows gDNA degradation: Line 1, log 2 marker; line 2,3, gDNA (150 ng) after crosslinking; line 4,5, gDNA (150 ng) after ligation cycle, 95°C for 3min, then 10 cycles of 94°C for 1min and 60°C for 1h; liner 6,7, gDNA (150 ng).

Without considering the dissociation of the ligated products, the unreacted probe, which remained in the system, will cause unspecific amplification, i.e. Ch1+Ch2− signals. Catalyst, acid buffer, and ligation cycles will cause gDNA degradation (Figure 5-36), giving more Ch1+Ch2− false-negative signals. However, Ch1+Ch2− and Ch1+Ch2− / Ch1−Ch2− resolution do not play a role in the mathematical modelling.

Assuming that all fdC at the target site is converted to the reporter strand via crosslinking and ligation, the yield is 100%. Assume that there are less than 150

copies in 1 µL so that the Poisson distribution is exclusive in our model.

Let a = Ch1+Ch2−, b = Ch1+Ch2+, c = Ch1−Ch2+, (Figure 5-31c) A = Accepted droplet for the experiment entry, respectively, a', b', c', and A' for the control, i.e. TET knockout cell line.

Let η = fdC content of the target site.

Then,

$$\eta = \frac{\mathrm{a} + \mathrm{b} - \frac{A}{A'}(a' + b')}{b + c + (a - \frac{A}{A'}a')/\eta}$$

where $\mathrm{a} - \frac{\mathrm{A}}{\mathrm{A}'}a'$ refers to the degraded gDNA copy containing fdC at the target site that does not show in Ch2, $(\mathrm{a} - \frac{\mathrm{A}}{A'}a')/\eta$ refers to all the degraded gDNA copies.

So,

$$\eta = \frac{(\mathrm{b} - \frac{A}{A'}\mathrm{b}')}{b + c}$$

Herein, in this ideal model, without considering the dissociation of the ligated products, η is independent of a, a', and c', i.e. genome degradation and unspecific ligation do not affect fdC percentage. Ch1+Ch2- only be resulted from the dissociation of the ligated products. Also, As shown in Figure 5-35 a-c, b' can be omitted. Simplified η'

$$\eta' = \frac{\mathrm{a} + \mathrm{b}}{b + c}$$

is calculated to indicate relative abundance of fdC at the target site.

To this point, the ddPCR-based profiling method was verified and optimized; a mathematical model was established.

5.4.3. fdC profiling: results and discussion

We next optimized the procedure, the experimental parameters and conditions are analysis to achieve the best resolution and lowest background. Finally, $Tdg^{-/-}$, $Tdg^{+/-}$, and Dnmt triple knockout (Dnmt TKO) ES cells were investigated using the method.

After extraction from the cells, gDNA was denatured and reannealed with the fdC

probe. Crosslinking was catalyzed by *p*-phenylenediamine. After crosslinking, catalyst and excessive probe were removed using the DNA cleanup kit and then ligated with the reporter strand. After a second purification, the gDNA was used as template in ddPCR and the fdC content was calculated as decribed in Chapter 5.4.3. Cluster ratios are summarized in Figure 5-37a; the typical 2-D plot of droplet fluorescence for each cell sample are shown in Figure 5-38.

Different droplets distribution from different cell samples can be observed in Figure 5-38. $Tdg^{-/-}$ cells contain 28.5% of fdC at the target site, 2-fold higher than $Tdg^{+/-}$ cells which gave a value of 15.7%, and 5-fold higher than Dnmt TKO cells (5.2%). The global mdC, hmdC, fdC, and cadC levels in the sample were analyzed using the known LC-MS method (Chapter 3.2.3.2) as shown in Figure 5-39.

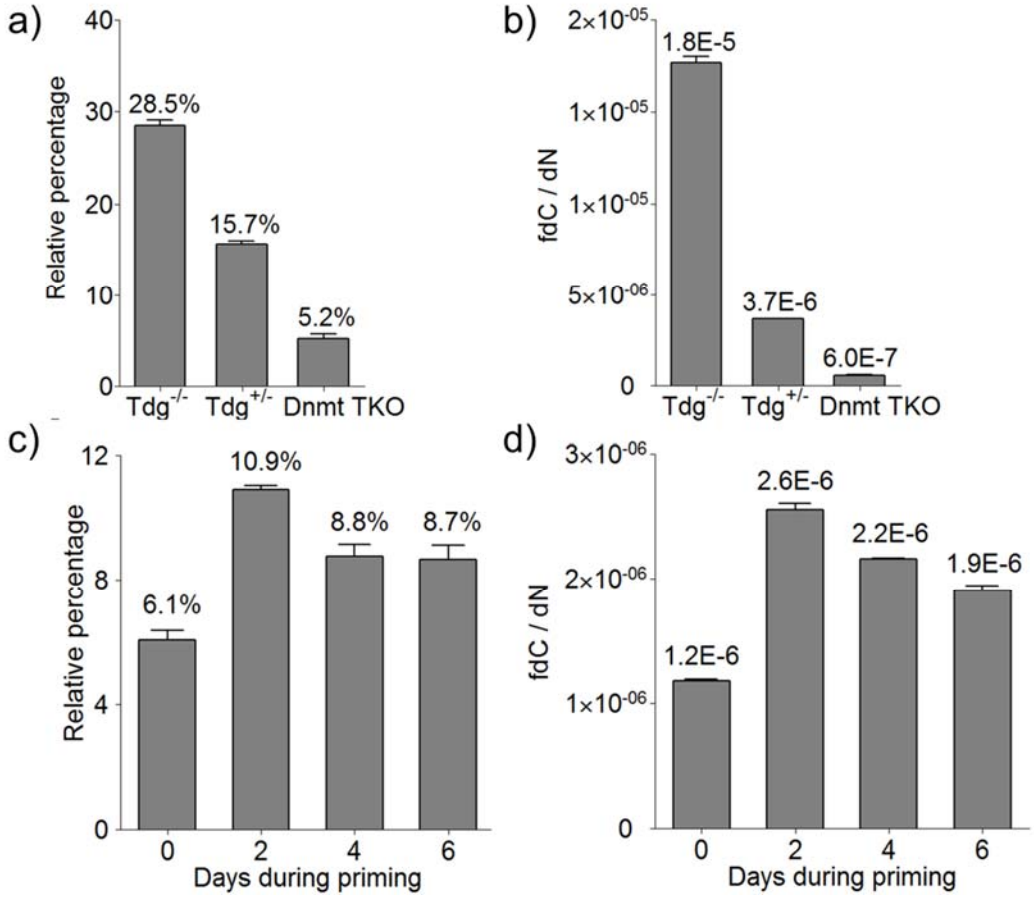


Figure 5-37 Cluster ratios of ddPCR 2-D plot for the position 1: a) $Tdg^{-/-}$, $Tdg^{+/-}$, and Dnmt TKO mES cells; b) 0, 2, 4, 6 days during priming of wild-type mES cells.

Next, five samples from 0 to 6 days of mESC priming were tested. It was found that the fdC content first increases to peak then decreases a bit along the days during

priming from 0 to 6 days to become finally stable (Figure 5-37b). For each cell sample, at least two "crosslink-ligation" replicates were conducted. After ligation and cleanup, at least four measurements were performed for each sample. The standard error of measurements and samples are 0.1~0.8% with a median of 0.36%.

Table 5-11 Synthesized and purchased oligonucleotides for position 2 detection. Bold and red letters represent nucleoside modifications or functional group.A, alkyne modifier with the linker; p, phosphate group at 5' terminus; F, FAM; H, HEX; Q, BHQ-1. MALDI-TOF: **41**, calc. 7672.5, found 7671.0, contains one sodium ion.

ODN	Description	5' ---------- 3'	mer
41	fdC probe	CAAGCGAAGCAGGGCAAATGGCGA	24
42	reporter strand	pTCTCGAACCTCTGCCAGCCTAGCTTGCTGTGGCGCAGGGACT CTGTGATGTCTGGAGCGATTCATTCTC	69
43	detection forward primer	AAGCAGGGCAAATGGCGA	19
36	detection reverse primer	GAGAATGAATCGCTCCAGAC	20
37	detection TaqMan probe	F-ACAGAGTCCCTGCGCCACAGCAAG-Q	24
44	reference forward primer	CCAGGGAGCATCTGTGAAAA	20
45	reference reverse primer	AAAAGCCCATCTGGGAAACA	20
46	reference TaqMan probe	H-CCCCTTCAGACGCAAGCGAAGCAG-Q	24

Next, the detection method was applied to a record position (8846677th of chromosome 15, MM9). The site sits in a non-coding DNA context where it is claimed to contain 32% of fdC according to the redBS-Seq method.[268] Probes, reporter strand, and primers were designed similarly and are listed in Table 5-11. 19.8% of fdC in $Tdg^{-/-}$ and 9.2% of fdC in $Tdg^{+/-}$ mES cells were observed.

In both sites, 1.8~2.2-fold in mean of fdC in $Tdg^{-/-}$ compared to $Tdg^{+/-}$ was obtained. Globally, $Tdg^{-/-}$ contains 4.9-fold more fdC than $Tdg^{+/-}$. The preferred explanation is that the content is not identical among different sites. Tdg may work differently at different sites. Another is that the detection method cannot reach a quantitative accuracy without calibration.

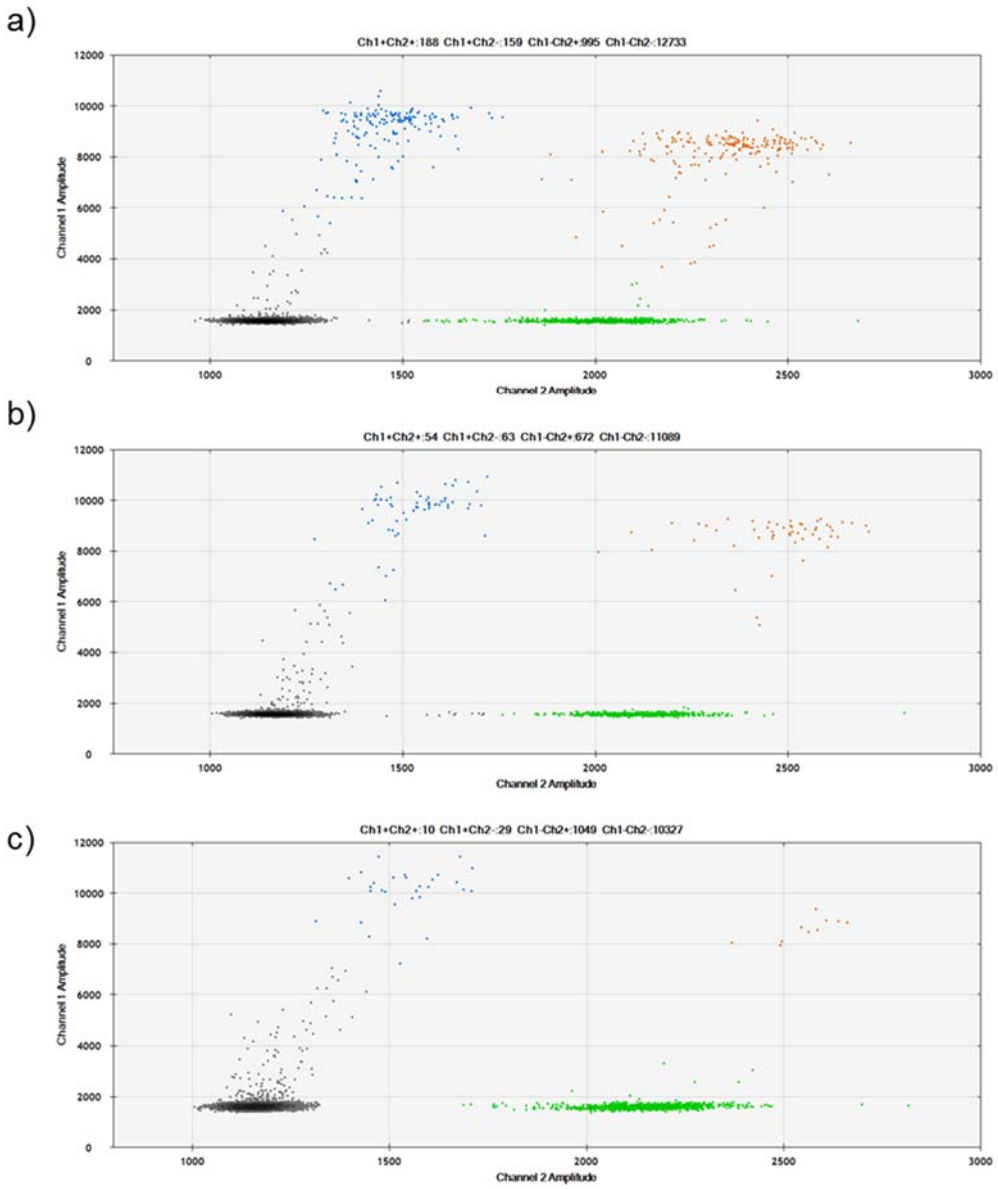
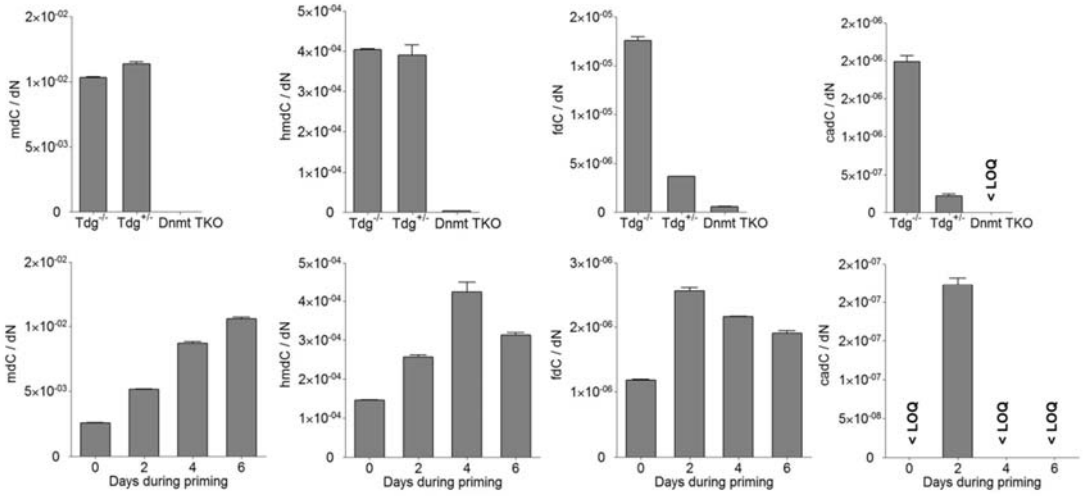


Figure 5-38 2-D plot of droplet fluorescence for the position 1: a) $Tdg^{-/-}$; b) $Tdg^{+/-}$; c) Dnmt TKO mES cells. Condition: 10 U Ampliagse, PCR extended at 64°C, 35 cycles.

Absolute quantification can be realized if the method is calibrated with a synthesized reference amplicon containing cytosine or fdC at the specific site. The method should also be available for mdC and hmdC site-specific detection if they are treated towards fdC.[214] According to the loci of interest, the method can track the epigenetic changes of cellular development, phenotype differentiation, memory consolidation,[350] etc. Once DNA array and multi-channel droplet counting technology are combined with ddPCR, multiple sites detection is insight.[351]

Compared to bisulfite variant NGS, the probe-ddPCR-based sequencing method

shortens and simplifies the sample preparation time and saves data analysis effort. Results can be read directly after droplet counting. Different from redRS-Seq and fdCAB-Seq,[233, 268] this method only provides information on the specific site, so it is not a substitution of BS-derived methods. Different probes, reporter strands, and primers have to be designed for different positions.



	mdC		hmdC		fdC		cadC	
	pro dN	STABW%	pro dN	STABW%	pro dN	STABW%	pro dN	STABW%
Tdg-/-	1.0E-02	1.1	4.0E-04	1.1	1.8E-05	3.6	2.0E-06	7.3
Tdg+/-	1.1E-02	2.0	3.9E-04	11.3	3.7E-06	0.8	2.2E-07	22.3
Dnmt TKO	3.0E-06	28.0	3.8E-06	18.8	6.0E-07	6.8	*	*
WT0	2.5E-03	1.7	1.5E-04	0.6	1.2E-06	1.9	*	*
WT2	5.1E-03	1.6	2.6E-04	3.6	2.6E-06	3.3	2.2E-07	6.8
WT4	8.7E-03	2.4	4.2E-04	11.2	2.2E-06	0.8	*	*
WT6	1.1E-02	1.7	3.1E-04	3.1	1.9E-06	3.4	*	*

Figure 5-39 Global mdC, hmdC, fdC, and cadC quantification: $Tdg^{-/-}$, $Tdg^{+/-}$, and Dnmt TKO mES cells and 0, 2, 4, 6 days during priming of wild-type mES cells. LC-MS analysis was performed by co-worker Mirko Wagner. *: < LOQ, below the limit of quantification.

5.5. Summary: Part II

Concerning on oligonucleotide functionalization, in this chapter, an oligonucleotide-probe was developed and applied for fdC position-specific detection and relative quantification.

The most suitable combination of the linker **32** and modifier position (n+4) was selected out of a library of 6 azide linkers and 12 positions. The linker was attached to the probe strand using solid-phase synthesis via Cu-catalyzed Huisgen cycloaddition. The crosslinking was completed in 24 h with 4-anisidine as the catalyst and accelerated to 4 h with the catalyst *p*-phenylenediamine. Tolerance to a mismatch in the duplex depends on the crosslinking temperature. The probe is specific for fdC does not react with other epigenetic C-base. Reaction occurs specific at the loci targeted by the probe strand. Multiple probes can react with different targets simultaneously.

Crosslinked dinucleoside do not inhibit nuclease and phosphatase. So digestionto nucleoside level is possible, but the conjugate itself cannot be digested which given a specific dinucleotide. Once the probe reacted with fdC on the target strand, primer extension using the target as a template is blocked. The crosslink is stable enough to survive denaturing and reannealing cycles.

With the help of a reporter strand, the probe can react with genomic DNA for relative fdC quantification in the genome. Optimal ddPCR parameters and procedure can differentiate fdC- and non-fdC-base at specific genome site. In $Tdg^{-/-}$ mouse ES cells, a 5-fold increase of fdC at one exon position is observed compared to Dnmt TKO cells, and 2-fold compared to $Tdg^{+/-}$ cells. Percentages of fdC at this site follow the same trend with global fdC content along priming.

6. Experimental part

6.1. General methods and materials

Organic synthesis

All experiments involving air and/or moisture-sensitive compounds were performed using flame- or oven-dried glassware under nitrogen or argon atmosphere. Chemicals were purchased from *Sigma-Aldrich, TCI,* and *Alfa-Aesar*. Solvents for organic synthesis were of reagent grade and purified by distillation.

Solutions were concentrated *in vacuu*m a *Heidolph* rotary evaporator with a *Vario PC2001* diaphragm pump by *Vacuubrand*. All mixed solvent systems are reported as v/v solutions. All reactions were monitored by thin-layer chromatography (TLC), performed on *Merck* 60 (silica gel F_{254}) plates. Chromatographic purification of products was accomplished using flash column chromatography on silica gel (230-400 mesh) purchased from Merck.

^{1}H-, ^{13}C-, ^{15}N-, ^{31}P-NMR and further 2D spectra were recorded in deuterated solvents on *Bruker ARX 300*, *Varian VXR400S*, *Varian Inova 400* and *Bruker AMX 600* spectrometers and calibrated to the residual solvent peak. Chemical shifts (δ, ppm) are adapted relative to the residual solvent peak as internal standard and coupling constants (J) are corrected and adapted to the nearest 0.1 Hz. Multiplicities are abbreviated as follows: s = singlet, d = doublet, t = triplet, m = multiplet, br = broad.

Oligonucleotides solid-phase synthesis

Commerical unavailable oligonucleotide synthesis was performed on an *Applied Biosystems* Incorporated 394 automated synthesizer, otherwise were purchased from *Metabion*. Phosphoramidites and solid supports columns were purchased from *Glen Research*, *Link Technology*, and *ChemGene Corporation*. Oligodeoxynucleotides were synthesized on 1 μmol scale with standard DNA synthesis cycles (trityl off mode). Coupling time for modified nucleosides was extended to 10 min. The oligonucleotide was cleaved using 1:1 mixture of ammonium hydroxide aq. (28%) and methylamine aq. (40%) at 65 °C for 10 min. The aqueous layer was then collected and removed in a *SpeedVac* concentrator, and the pellet redissolved in ddH_2O. Oligonucleotides contained mdC, hmdC, fdC, caC and abasic site were synthesized according to the

previous report.[352]

HPLC analysis and purification

Analytical RP-HPLC was performed using a *Macherey-Nagel* Nucleodure 100-3 C18ec column on *Waters Alliance* 2996 Photodiode Array Detector, 2695 Separation Module using a flow of 0.5 mL/min. Semi-preparative RP-HPLC was performed using a *Macherey-Nagel* C18 column (5 mm, 9.4×250 mm) on *Waters* Breeze 2487 Dual λ Array Detector, 1525 Binary HPLC Pump. Conditions: Buffer A, 0.1M TEAA (triethylammonium acetate) in water; Buffer B, 0.1M TEAA in 80% acetonitrile, flow rate 5 mL/min. HPLC grade acetonitrile was purchased from *VWR*.

The purified fractions were concentrated in a *SpeedVac* concentrator and characterized by Matrix Assisted Laser Desorption Ionization Time of Flight (MALDI-TOF) on *Bruker Daltonics Autoflex II*. MALDI matrix was made up of diammonium hydrogen citrate (10 mg), 3-hydroxypicolinic acid (50 mg), 15-crone-5 (10 μL) in ddH_2O (1 mL).

Concentration of oligonucleotide solutions was calculated from the UV absorbance at 260 nm on a *Nanodrop* ND-1000 spectrophotometer. Extinction coefficients of the oligonucleotides at 260 nm were calculated by addition of the extinction coefficients of the individual nucleobases: dA 15.0 L/(mmol·cm), dC 7.1 L/(mmol·cm), dG 12.0 L/(mmol·cm), dT 8.4 L/(mmol·cm), Pz 7.6 L/(mmol·cm), Pm 5.7 L/(mmol·cm), mdC 7.8 L/(mmol·cm), hmdC 8.7 L/(mmol·cm), fdC 11.3 L/(mmol·cm), cadC 7.1 L/(mmol·cm), 6-FAM 13.7 L/(mmol·cm). The 1,2,3-triazole, abasic monomer, and biotin are transparent at 260 nm.

Molecular biology

Enzymes and specific buffers were purchased from *New England BioLab* unless otherwise stated. H_2O for dilution and solution was obtained from *Sartorius Stedim arium pro* system.

Denaturing urea polyacrylamide gel electrophoresis (PAGE) assay was processed on 20 cm long or 10 cm long, 0.75 mm thick gel in *Bio-Rad Protein II* chamber. The gel consisted of sequencing gel (8 mL), TBE buffer (10×, 1 mL), urea aq. (8 M, 1 mL), ammonium persulfate (APS) aq. (10% w/v, 200 μL), and tetramethylethylenediamine

(TEMED, 20 µL). The loading buffer (2×) consisted of TBE (10×, 6 mL), Ficoll (3.60 g), bromophenol blue (3 mg), Xylene-Cyanol-FF (6 mg), urea (12.6 g) filled with ddH_2O up to 30 mL. The gel was dyed with *SYBR* Green I (0.1‰, i.e. 1 µL in 10 mL 1× TBE buffer) for 10 min, photographed by *FUJIFILM LAS-3000* and quantified by built-in software Ai*da Image Analyzer v.450*.

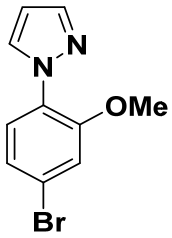
Agarose gel loading buffer (6×) consisted of Ficoll (1.50 g), Xylene-Cyanol-FF (25 mg) filled with ddH_2O up to 10 mL. Agarose gel was made of agarose (500 mg) and (2 µL) in TAE buffer (1×, 40 mL, i.e. 1.25%) and run in TAE buffer (0.5×) with 100 V for ca. 1 h. The gel was photographed and visualized by *E-Box VX5*.

PCR was performed on *Bio-Rad T100 Thermal Cycler* and *Eppendorf Mastercycler personal*. Real-time PCR was performed on *Eppendorf Realplex*. *iTaq* Universal SYBR Green Supermix was purchased from *Bio-Rad*. Cycle threshold numbers (C_t values) were calculated by CalQplex method.

Buffers: ***10×TBE:*** 2-amino-2-hydroxymethyl-propane-1,3-diol (Tris, 108.0 g), boric acid (55.0 g), EDTA-Na_2 (9.3 g), add H_2O to 1 L. ***50×TAE:*** Tris (242.3 g), EDTA-2Na_2 (37.2 g), adjust with acetic acid until pH = 8.3, and then add H_2O to 1 L. ***20×SSC:*** sodium chroloride (175.3 g), trisodium citrate 2H_2O (83.2 g), add HCl nutrulized to pH = 7.0, and then add H_2O to 1 L.

6.2. Part I: Synthesis of pyrazole ligandoside and its derivates

1-(4-Bromo-2-methoxyphenyl)-1*H*-pyrazole (4)



4-Bromo-2-methoxyaniline (**3**) (5.00 g, 24.7 mmol) was dissolved in 20% HCl (25 mL). The solution was stirred and cooled in an ice bath while a solution of sodium nitrite (1.79 g, 26.0 mmol) in 4 mL of H_2O was added dropwise. The mixture was stirred for further 1 h at 0°C. Stannous chloride dihydrate (11.2 g, 49.5 mmol) in 35% HCl (15 mL) was then added slowly. The mixture was allowed to warm to room temperature whilst stirring for another 2 h before being filtered. The brown residue was washed with 35% HCl (20 mL) twice and cold ethanol (20 mL) dried under vacuum to give (4-bromo-2-methoxyphenyl)hydrazine as light blue solid (4.12 g, 16.25 mmol, 66% yield). The resulted light blue diazonium salt (4.12 g, 16.25 mmol) was dissolved in methanol (100 mL) as a dark green solution and was dropwise added to a solution of 1,1,3,3-tetramethoxypropane (8.15 mL, 49.5 mmol) in methanol (50 mL) and 35% HCl (5.0 mL) at room temperature. The reaction was refluxed at 65°C for 12 h and then concentrated under vacuum. The residue was dissolved in water (100 mL).The aqueous phase was extracted with dichloromethane (3×100 mL). The combined organic layers were washed with $NaHCO_3$ (3×100 mL), brine (100 mL), and then dried over $MgSO_4$ and filtered. The solvent was removed under reduced pressure. The crude was purified by flash chromatography on silica gel (isohexane / ethyl acetate 10:1) to give **4** as light yellow oil (3.51 g, 13.9 mmol, 85% yield).

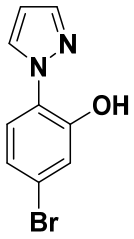
R_f (isohexane / dichloromethane = 1/1) = 0.11.

^{1}H-NMR (300 MHz, $CDCl_3$): δ = 8.02 (dd, J = 2.5, 0.6 Hz, 1H, C(5)***H***), 7.74 – 7.67 (m, 1H, C(3)***H***), 7.63 (dd, J = 8.3, 0.4 Hz, 1H, C_{ar}***H***), 7.19 (dt, J = 2.8, 1.9 Hz, 2H, C_{ar}***H***), 6.43 (dd, J = 2.5, 1.8 Hz, 1H, C(2)***H***), 3.89 (s, 3H, $C_{ar}OC$***H***$_3$).

^{13}C-NMR (75 MHz, $CDCl_3$): δ = 151.6 (***C***$_{ar}OCH_3$), 140.3 (***C***(3)H), 131.4 (***C***(5)H), 128.9 (***C***$_{ar}$N), 126.2 (***C***$_{ar}$H), 124.2 (***C***$_{ar}$H), 120.7 (***C***$_{ar}$Br), 115.8 (***C***$_{ar}$H), 106.5 (***C***(2)H), 56.3 ($C_{ar}O$***C***H_3).

HRMS (ESI+): calculated for $C_{10}H_{10}BrN_2O^+$ $[M+H]^+$: 252.9971, found: 252.9972.

1-(4-Bromo-2-hydroxyphenyl)-1*H*-pyrazole (5)



1-(4-Bromo-2-methoxyphenyl)-1H-pyrazole (**4**) (5.08 g, 20.1 mmol) was dissolved in anhydrous dichloromethane (100 mL) at -78°C. BBr_3 (1 M solution in dichloromethane, 24.1 mL, 24.1 mmol) was added dropwise in 1 h and then stirred for 12 h. The reaction was quenched with methanol and warmed to room temperature. The solution was diluted with dichloromethane (300 mL), washed with water (2×300 mL), and then dried over $MgSO_4$ and filtered. The solvent was removed under reduced pressure. The crude was purified by flash chromatography on silica gel (isohexane / dichloromethane 3:1) to give **5** as a white powder (3.25 g, 13.6 mmol, 68% yield).

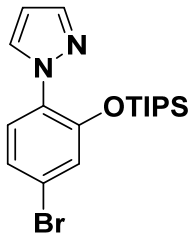
$\mathbf{R_f}$ (isohexane / dichloromethane = 1/1) = 0.51.

^{1}H-NMR (300 MHz, $CDCl_3$): δ = 11.58 (br, 1H, O***H***), 7.97 (d, J = 2.6 Hz, 1H, C(5)***H***), 7.73 (d, J = 1.9 Hz, 1H, C(3)***H***), 7.27 (d, J = 2.2 Hz, 1H, C$_{ar}$***H***), 7.23 (d, J = 8.6 Hz, 1H, C$_{ar}$***H***), 7.06 (dd, J = 8.6, 2.2 Hz, 1H, C$_{ar}$***H***), 6.53 (t, J = 2.2 Hz, 1H, C(4)***H***).

^{13}C-NMR (75 MHz, $CDCl_3$): δ = 150.2 (***C***$_{ar}$OH), 139.0 (***C***(3)H), 126.6 (***C***(5)H), 124.1 (***C***$_{ar}$N), 122.4 (***C***$_{ar}$H), 122.1 (***C***$_{ar}$H), 120.1 (***C***$_{ar}$Br), 118.7 (***C***$_{ar}$H), 107.1 (***C***(4)H).

HRMS (ESI+): calculated for $C_9H_8BrN_2O^+$ $[M+H]^+$: 238.9815, found: 238.9815.

1-[4-Bromo-2-[(triisopropylsilyl)oxy]phenyl]2-1*H*-pyrazole (6)



1-(4-Bromo-2-hydroxyphenyl)-1H-pyrazole (**5**) (2.41 g, 10.1 mmol) was dissolved in dichloromethane (50 mL) and *N*,*N*-diisopropylethylamine (4.18 mL, 25.2 mmol) at 0°C. Triisopropylsilyl trifluoromethanesulfonate (4.06 mL, 15.1 mmol) was added dropwise. The mixture was allowed to warm to room temperature while stirring for 1 h, diluted with dichloromethane (150 mL) and washed with water (2×300 mL), and then dried over $MgSO_4$ and filtered. Solvent was removed under reduced pressure. The crude was purified by flash chromatography on silica gel (isohexane / ethyl acetate 20:1) to give **6**

as yellowish oil (4.39 g, 10.0 mmol, 99% yield).

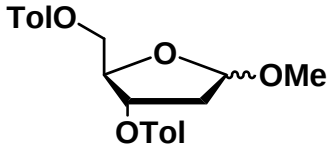
R_f (isohexane / ethyl acetate = 10/1) = 0.66.

^{1}H-NMR (300 MHz, $CDCl_3$): δ = 7.91 (dd, J = 2.4, 0.6 Hz, 1H, C(5)***H***), 7.67 (dd, J = 1.8, 0.6 Hz, 1H, C(3)***H***), 7.51 (d, J = 8.6 Hz, 1H, C_{ar}***H***), 7.21 – 7.15 (m, 1H, C_{ar}***H***), 7.13 (d, J = 1.9 Hz, 1H, C_{ar}***H***), 6.41 (dd, J = 2.4, 1.8 Hz, 1H, C(4)***H***), 1.30 – 1.11 (m, 3H, C***H***$(CH_3)_2$), 1.09 – 0.92 (m, 18H, CH(C***H***$_3)_2$).

^{13}C-NMR (75 MHz, $CDCl_3$): δ = 149.0 (***C***$_{ar}$OSi), 140.2 (***C***(3)H), 131.3 (***C***(5)H), 131.0 (***C***$_{ar}$N), 127.2 (***C***$_{ar}$H), 124.7 (***C***$_{ar}$H), 123.2 (***C***$_{ar}$H), 120.5 (***C***$_{ar}$Br), 106.3 (***C***(4)H), 17.7 (6×CH(***C***$H_3)_2$), 12.85 (3×***C***H$(CH_3)_2$).

HRMS (ESI+): calculated for $C_{18}H_{28}BrN_2OSi^+$ $[M+H]^+$: 395.1154, found: 395.1148.

1-*O*-Methyl-3,5-di-o-toluoyl-2-deoxy-L-riboside (11)



2-Deoxy-D-ribose (**9**) (13.7 g, 102 mmol), acetyl chloride (0.5 mL) were dissolved in anhydrous methanol (300 mL) and stirred at room temperature for 30 min. Then silver carbonate (4.50 g, 16.3 mmol) was added, the mixture was filtered, and the alcohol solution was concentrated. Anhydrous pyridine (2×3 mL) was added for co-evaporation. The residue (**10**) was dissolved in pyridine (60 mL) in the ice bath and stirred for 20 min before 4-methylbenzoyl chloride (26.5 mL, 204 mmol) was dropped in. The mixture was then raised to 45°C and further stirred for 2 h. H_2O (300 mL) was added and extracted with ethyl acetate (3×300 mL). The combined organic layer was washed with HCl (2 M, 2×1000 mL) and $NaHCO_3$ aq. (1000 mL) and dried, filtered, concentrated and purified by flash column chromatography (isohexane / ethyl acetate 4:1) to give **11** as a colorless oil (33.9 g, 88.3 mmol, 87% yield, α/β (R/S) mixture).

R_f (isohexane / ethyl acetate = 2/1) = 0.86.

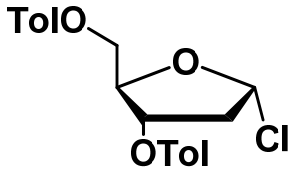
^{1}H-NMR (400 MHz, $CDCl_3$): δ = 8.08 – 7.83 (m, 8H, C_{ar}***H***, α/β), 7.39 – 7.08 (m, 8H, C_{ar}***H***, α/β), 5.61 – 5.58 (m, 1H, C(1')***H***, α), 5.41 (ddd, J = 8.2, 3.7, 2.2 Hz, 1H, C(1')***H***, β), 5.23 (dd, J = 5.4, 2.2 Hz, 1H, C(4')***H***, β), 5.19 (dd, J = 5.3, 1.0 Hz, 1H, C(4')***H***, α), 4.71 – 4.35 (m, 6H, C(3')***H***, C(5')***H***$_2$, α/β), 3.42 (s, 3H, OC***H***$_3$, β), 3.36 (s, 3H, OC***H***$_3$, α), 2.59 – 2.17 (m, 16H, C(2')***H***$_2$, 2×C_{ar}C***H***$_3$, α/β).

^{13}C-NMR (100 MHz, $CDCl_3$): δ = 166.5 – 166.1 (***C***OO), 144.0 – 143.7 (***C***$_{ar}$$CH_3$), 129.8 – 129.0 (***C***$_{ar}$H), 127.2 – 126.9 (***C***$_{ar}$CO), 105.6 (***C***(1')), 105.0 (***C***(1')), 81.9 (***C***(4')),

80.9 (***C***(4')), 75.4 (***C***(3')), 74.6 (***C***(3')), 65.1 (***C***(5')), 64.3 (***C***(5')), 55.2 (***C***(2')), 55.1 (C(1')O***C***H_3), 39.3 (***C***(2')), 21.7 (C_{ar}***C***H_3).

HRMS (ESI+): calculated for $C_{22}H_{24}O_6Na^+$ $[M+Na]^+$:407.1471, found: 407.1463.

α-3',5'-Biostoluoyl-1'-ribosyl chloride (8)



1-*O*-Methyl-3,5-di-*O*-toluoyl-2-deoxy-L-riboside (**11**) (33.9 g, 88.3 mmol) was dissolved in acetic acid (60 mL) and H_2O (5 mL) on an ice bath. Acetyl chloride (18.7 mL, 265 mmol) was added to the mixture dropwise. As the system solidified, the flask was warmed to room temperature. The mixture was then filtered, and the filter cake was washed with anhydrous ether (2×30 mL) to give **8** as a gray solid (19.0 g, 48.8 mmol, 55% yield).

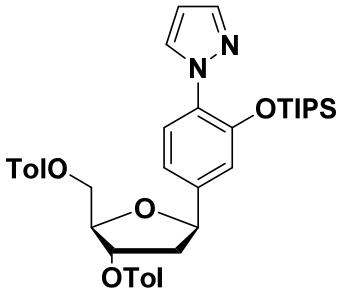
R_f (isohexane / ethyl acetate = 2/1) = 0.41.

^{1}H-NMR (200 MHz, $CDCl_3$): δ = 8.01 – 7.87 (m, 4H, C_{ar}***H***), 7.27 – 7.21 (m, 4H, C_{ar}***H***), 6.48 – 6.45 (m, 1H, C(1')***H***,), 5.73 – 5.32 (m, 1H, C(3')***H***), 4.84 (m, 1H, C(4')***H***), 4.64 (m, 2H, C(5')***H***), 2.84 (m, 2H, C(2')H_2), 2.41 (s, 6H, C_{ar}CH_3).

^{13}C-NMR (75 MHz, $CDCl_3$): δ = 166.4 – 166.0 (***C***OO), 144.6 – 143.2 (***C***$_{ar}$$CH_3$), 130.2 – 129.0 (***C***$_{ar}$H), 127.1 – 126.5 (***C***$_{ar}$$CH_3$), 110.6 (**C**(1')), 84.7 (***C***(4')), 73.5 (***C***(3')), 63.5 (***C***(5')), 44.5 (***C***(2')), 21.7 (C_{ar}***C***H_3).

HRMS (ESI+): calculated for $C_{21}H_{21}O_5^+$ $[M-Cl]^+$:353.1394, found: 353.1367.

5-[4-(1*H*-Pyrazol-1-yl)-3-[(triisopropylsilyl)oxy]phenyl]-2-[[(4-methylbenzoyl)oxy]methyl]tetrahydrofuran-3-yl 4-methylbenzoate (12)



1-[4-Bromo-2-[(triisopropylsilyl)oxy]phenyl]-1H-pyrazole (**6**) (3.52 g, 8.04 mmol) was dissolved in anhydrous diethyl ether (20 mL) and cooled to -78°C. *t*-BuLi in pentane (1.6 M, 10.1 mL, 16.1 mmol) was added dropwise over 30 min. The reaction

was kept at -78°C for another 2 h and then transferred to a precooled (-78°C) suspension of copper(I) bromide-disulfide complex (826 mg. 4.20 mmol) in ether (20 mL). The mixture was slowed warmed to -30°C and stirred for 20 min as all the copper dissolved. The resulted clear yellow solution was then cooled down to -78°C and transferred to a precooled gray solution of α-3',5'-biostoluoyl-1'-ribosyl chloride (**8**) (1.04 g, 2.68 mmol) in anhydrous dichloromethane (20 mL). The reaction mixture was allowed to warm up to room temperature for 12 h. The reaction was quenched by saturated NH_4Cl and ammonia aq. (2 M, 1 mL). The aqueous phases were extracted three times with ether (3×100 mL). The combined organic phases were washed with NH_4Cl aq. (2×300 mL), brine (300 mL), then dried over $MgSO_4$ and filtered. Solvent was removed under reduced pressure. The crude was purified by flash chromatography on silica gel (isohexane / ethyl acetate 10:1) to give **12** as yellowish oil (944 mg, 1.41 mmol, 53% yield).

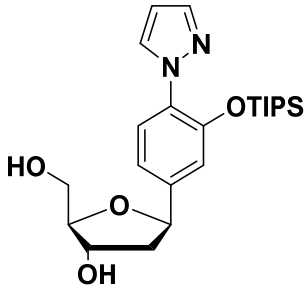
R_f (isohexane / ethyl acetate = 10/1) = 0.18 (β), 0.12 (α).

^{1}H-NMR (600 MHz, $CDCl_3$): δ = 7.99 (d, J = 8.4 Hz, 2H, C_{ar}***H***), 7.92 (d, J = 2.4 Hz, 1H, C(5)***H***), 7.91 (d, J = 8.6 Hz, 2H, C_{ar}***H***), 7.67 (dd, J = 1.8, 0.5 Hz, 1H, C(3)***H***), 7.59 (d, J = 8.4 Hz, 1H, C_{ar}***H***), 7.31 – 7.27 (m, 2H, C_{ar}***H***), 7.22 – 7.19 (m, 2H, C_{ar}***H***), 7.09 (d, J = 1.7 Hz, 1H, C_{ar}***H***), 7.05 – 7.01 (m, 1H, C_{ar}***H***), 6.41 – 6.37 (m, 1H, C(4)***H***), 5.61 (dd, J = 6.0, 0.8 Hz, 1H, C(3')***H***), 5.25 (dd, J = 10.9, 5.0 Hz, 1H, C(1')***H***), 4.60 (dd, J = 4.2, 2.7 Hz, 2H, C(5')***H***$_2$), 4.57 (dd, J = 4.2, 1.9 Hz, 1H, C(4')***H***), 2.58 (dd, J = 13.7, 5.1 Hz, 1H, C(2')***H***H), 2.43 (s, 3H, OC***H***$_3$), 2.41 (s, 3H, OC***H***$_3$), 2.20 (ddd, J = 13.8, 11.0, 6.0 Hz, 1H, C(2')***H***H), 1.18 – 1.10 (m, 3H, C***H***$(CH_3)_2$), 1.00 – 0.91 (m, 18H, CH(C***H***$_3)_2$).

^{13}C-NMR (151 MHz, $CDCl_3$): δ = 166.3 (***C***OO), 166.1 (***C***OO), 148.3 (***C***$_{ar}$OSi) , 144.2 (***C***$_{ar}$C(1')), 143.8 (***C***$_{ar}CH_3$), 141.0 (***C***$_{ar}CH_3$), 139.9 (***C***(3)H), 131.4 (***C***(5)H), 131.0 (***C***$_{ar}$), 129.7 (4×***C***$_{ar}$H), 129.2 (4×***C***$_{ar}$H), 129.0 (2×***C***$_{ar}$CO), 127.0 (***C***$_{ar}$H), 126.0 (***C***$_{ar}$H), 118.7 (***C***$_{ar}$H), 117.1 (***C***$_{ar}$H), 105.9 (***C***$_{ar}$(4)H), 82.9 (***C***(4')), 80.0 (***C***(1')), 77.1 (***C***(3')), 65.0 (***C***(5')), 41.7 (***C***(2')), 21.7 (C_{ar}***C***H_3), 21.6 (C_{ar}***C***H_3), 17.7 (6×CH(***C***$H_3)_2$), 12.9 (3×***C***$H(CH_3)_2$).

HRMS (ESI+): calculated for $C_{39}H_{49}N_2O_6Si^+$ $[M+H]^+$: 669.3360, found: 669.3339.

5-[4-(1*H*-Pyrazol-1-yl)-3-[(triisopropylsilyl)oxy]phenyl]-2-(hydroxymethyl)tetrahydrofuran-3-ol (13)



5-[4-(1*H*-Pyrazol-1-yl)-3-[(triisopropylsilyl)oxy]phenyl]-2-[[(4-methylbenzoyl)oxy]methyl]tetrahydrofuran-3-yl 4-methylbenzoate (**12**) (1.68 g, 2.52 mmol), and potassium carbonate (764 mg, 5.53 mmol) were dissolved in methanol (20 mL), stirred at room temperature for 1 h, then diluted with water (100 mL). The aqueous phases were extracted three times with chloroform (3×100 mL). The combined organic phases were dried over $MgSO_4$ and filtered. Solvent was removed under reduced pressure. The crude was purified by flash chromatography on silica gel (chloroform / methanol 20:1) to give **13** as colorless oil (712 mg, 1.65 mmol, 65% yield).

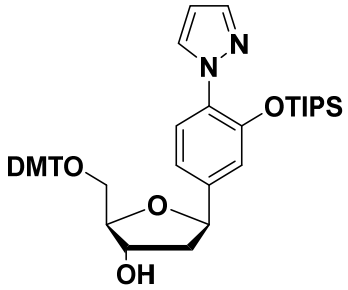
R_f (chloroform / methanol = 10/1) = 0.32.

^{1}H-NMR (400 MHz, MeOH-*d4*): δ = 7.96 – 7.87 (m, 1H, C(5)***H***), 7.67 (dd, *J* = 1.9, 0.6 Hz, 1H, C(3)***H***), 7.43 (d, *J* = 8.1 Hz, 1H, C_{ar}***H***), 7.15 (d, *J* = 1.8 Hz, 1H, C_{ar}***H***), 7.08 (ddd, *J* = 8.1, 1.8, 0.5 Hz, 1H, C_{ar}***H***), 6.48 (dd, *J* = 2.4, 2.0 Hz, 1H, C(4)***H***), 5.14 (dd, *J* = 10.5, 5.5 Hz, 1H, C(1')***H***), 4.34 (dt, *J* = 3.9, 1.7 Hz, 1H, C(3')***H***), 4.03 – 3.95 (m, 1H, C(4')***H***), 3.69 (dd, *J* = 11.4, 5.1 Hz, 1H, C(5')***H***H), 3.60 (ddd, *J* = 9.2, 6.4, 3.9 Hz, 1H, C(5')***H***H), 2.24 (ddd, *J* = 13.1, 5.5, 1.7 Hz, 1H, C(2')***H***H), 1.96 – 1.84 (m, 1H, C(2')***H***H), 1.28 – 1.12 (m, 3H, 3×C***H***$(CH_3)_2$), 1.01 (dt, *J* = 7.8, 3.9 Hz, 18H, 6×CH(C***H***$_3)_2$).

^{13}C-NMR (100 MHz, MeOH-*d4*): δ = 150.5 (***C***$_{ar}$OSi), 145.2 (***C***$_{ar}$C(1')), 141.0 (***C***(3)H), 133.4 (***C***(5)H), 131.7 (***C***$_{ar}$), 127.5 (***C***$_{ar}$H), 120.2 (***C***$_{ar}$H), 118.5 (***C***$_{ar}$H), 107.2 (***C***(4)H), 89.2 (***C***(4')), 80.7 (***C***(1')), 74.5 (***C***(3')), 64.2 (***C***(5')), 45.0 (***C***(2')), 18.3 (6×CH(***C***$H_3)_2$), 14.1 (3×***C***$H(CH_3)_2$).

HRMS (ESI+): calculated for $C_{23}H_{37}N_2O_4Si^+$ $[M+H]^+$: 433.2523, found: 433.2513.

5-[4-(1*H*-Pyrazol-1-yl)-3-[(triisopropylsilyl)oxy]phenyl]-2-[[bis(4-methoxyphenyl)(phenyl)methoxy]methyl]tetrahydrofuran-3-ol (14)



5-[4-(1*H*-Pyrazol-1-yl)-3-[(triisopropylsilyl)oxy]phenyl]-2-(hydroxymethyl)tetrahydrofuran-3-ol (**13**) (712 mg, 1.65 mmol) and 4,4'-dimethoxytrityl chloride (613 mg, 1.81 mmol) were dissolved in anhydrous dichloromethane (30 mL) and stirred at 0°C. Then *N, N*-diisopropylethylamine (683 µL, 4.11 mmol) was added dropwise. The reaction was allowed to reach room temperature for 1 h and diluted with dichloromethane (200 mL). The organic phase was washed two times with sat aqueous $NaHCO_3$ (2×200 mL), then dried over $MgSO_4$ and filtered. The solvent was removed under reduced pressure. The crude was purified by flash chromatography on silica gel (isohexane / ethyl acetate 2:1 +3% triethylamine) to give **7** as a yellowish foam (893 mg, 1.22 mmol, 74% yield).

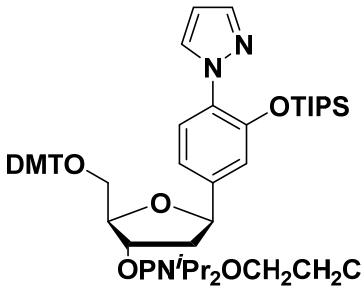
$\mathbf{R_f}$ (isohexane / ethyl acetate = 2/1) = 0.38.

^{1}H-NMR (400 MHz, $CDCl_3$): δ = 7.91 (dd, J = 2.4, 0.7 Hz, 1H, C(5)***H***), 7.66 (dd, J = 1.8, 0.6 Hz, 1H, C(3)***H***), 7.55 (d, J = 8.2 Hz, 1H, C_{ar}***H***), 7.48 – 7.41 (m, 2H, C_{ar}***H***, DMTr), 7.38 – 7.15 (m, 7H, C_{ar}***H***, DMTr), 7.09 – 6.98 (m, 1H, C_{ar}***H***), 6.91 (d, J = 1.8 Hz, 1H, C_{ar}***H***), 6.87 – 6.78 (m, 4H, C_{ar}***H***, DMTr), 6.39 (dd, J = 2.4, 1.8 Hz, 1H, C(4)***H***), 5.12 (dd, J = 10.1, 5.6 Hz, 1H, C(1')***H***), 4.40 (dd, J = 5.4, 2.8 Hz, 1H, C(3')***H***), 4.06 (ddd, J = 7.1, 4.6, 2.6 Hz, 1H, C(4')***H***), 3.79 (d, J = 2.4 Hz, 6H, $C_{ar}OC\boldsymbol{H}_3$, DMTr), 3.42 (dd, J = 9.6, 4.7 Hz, 1H, C(5')***H***H), 3.19 (dd, J = 9.6, 6.7 Hz, 1H, C(5')***H***H), 2.28 – 2.15 (m, 1H, C(2')***H***H), 1.92 (ddd, J = 13.1, 10.1, 6.0 Hz, 2H, C(2')***H***H), 1.20 – 1.07 (m, 3H, 3×C***H***$(CH_3)_2$), 1.01 – 0.83 (m, 18H, 6×CH(C***H***$_3)_2$).

^{13}C-NMR (100 MHz, $CDCl_3$): δ = 158.5 (2×***C***$_{ar}OCH_3$, DMTr), 148.1 (***C***$_{ar}$OSi), 144.7 (**C**$_{ar}$, DMTr), 142.1 (**C**$_{ar}$C(1')), 139.8 (***C***(5)H), 135.9 (2×***C***$_{ar,}$, DMTr), 131.4 (***C***(3)H), 130.8 (***C***$_{ar}$N), 130.0 (2×***C***$_{ar}$, DMTr), 128.1 (2×***C***$_{ar}$, DMTr), 127.8 (4×***C***$_{ar}$, DMTr), 126.8 (***C***$_{ar}$H, DMTr), 126.0 (***C***$_{ar}$H), 118.7 (***C***$_{ar}$H), 117.5 (***C***$_{ar}$H), 113.1 (4×***C***$_{ar}$H, DMTr), 105.8 (***C***(2)H), 86.3 (***C***(4')), 86.2 (O***C***C_{ar}, DMTr), 79.2 (***C***(1')), 74.7 (***C***(3')), 64.4 (***C***(5')), 55.2 (2×$C_{ar}O\boldsymbol{C}H_3$, DMTr), 43.5 (***C***(2')), 17.7 (6×CH(***C***$H_3)_2$), 12.8 (3×***C***$H(CH_3)_2$).

HRMS (ESI+): calculated for $C_{44}H_{55}N_2O_6Si^+$ $[M+H]^+$: 735.3829, found: 735.3820.

5-[4-(1*H*-Pyrazol-1-yl)-3-[(triisopropylsilyl)oxy]phenyl]-2-[[bis(4-methoxyphenyl)(phenyl)methoxy]methyl]tetrahydrofuran-3-yl-2-cyanoethyl)diisopropyl-phosphoramidite (15)



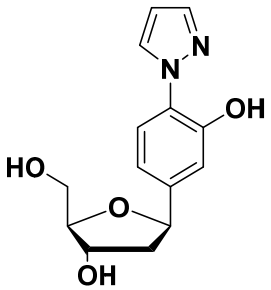
In a Schlenck tube, the DMTr-protected nucleoside **14** (237 mg, 322 µmol) and diisopropyl ammonium tetrazolide (221 mg, 1.29 mmol) were dissolved in rigorously degassed dichloromethane (3 mL). The solution was then degassed three more times. Bis(diisopropylamino)(2-cyanoethoxy)phosphine (205 µL, 645 µmol) was added to the solution. The reaction mixture was stirred at room temperature for 40 min under argon atmosphere, then directly purified by flash chromatography on silica gel (isohexane / ethyl acetate 2:1 +3% triethylamine) to give **15** as a yellowish foam (301 mg, 322 µmol, quant.).

R_f (isohexane / ethyl acetate = 2/1) = 0.71.

^{31}P-NMR (81 MHz, $CDCl_3$): δ = 149.25, 149.07.

HRMS (ESI+): calculated for $C_{53}H_{72}N_4O_7PSi^+$ $[M+H]^+$: 935.4908, found: 935.4902.

5-[3-Hydroxy-4-(1*H*-pyrazol-1-yl)phenyl]-2-(hydroxymethyl)tetrahydrofuran-3-ol (16)



5-[4-(1*H*-Pyrazol-1-yl)-3-[(triisopropylsilyl)oxy]phenyl]-2-[[(4-methylbenzoyl)oxy]methyl]tetrahydrofuran-3-yl 4-methylbenzoate (**3**) (393 mg, 588 µmol) and potassium carbonate (233 mg, 1.69 mmol) were dissolved in methyl alcohol (10 mL). A solution of tetra-n-butyl ammonium fluoride in THF (1 M, 0.60 mL) was added to the mixture

and then stirred at room temperature for 2 h. Solvents were removed under reduced pressure. The residue was purified by flash chromatography on silica gel (chloroform / methanol 20:1) to give **16** as a white amorphous solid (141 mg, 511 µmol, 87% yield). Further crystallization was achieved in an ethyl acetate solution of **16**.

R_f (chloroform / methanol = 10/1) = 0.23.

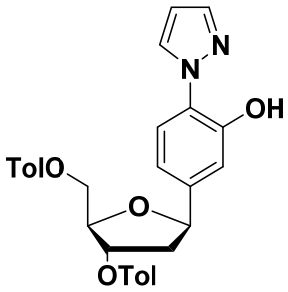
^{1}H-NMR (400 MHz, MeOH-*d4*): δ = 8.24 (dd, *J* = 2.5, 0.6 Hz, 1H, C(5)***H***), 7.72 (dd, *J* = 1.9, 0.5 Hz, 1H, C(3)***H***), 7.52 (d, *J* = 8.3 Hz, 1H, C_{ar}***H***), 7.11 – 7.03 (m, 1H, C_{ar}***H***), 6.97 (ddd, *J* = 8.3, 1.9, 0.5 Hz, 1H, C_{ar}***H***), 6.51 (dd, *J* = 2.5, 2.0 Hz, 1H, C(4)***H***), 5.10 (dd, *J* = 10.4, 5.4 Hz, 1H, C(1')***H***), 4.32 (dt, *J* = 4.1, 1.9 Hz, 1H, C(3')***H***), 3.96 (dt, *J* = 5.2, 2.4 Hz, 1H, C(4)***H***), 3.79 – 3.59 (m, 2H, C(5')***H***), 2.21 (ddd, *J* = 13.1, 5.5, 1.7 Hz, 1H, C(2')***H***H), 1.95 (ddd, *J* = 13.1, 10.4, 5.9 Hz, 1H, C(2')***H***H).

^{13}C-NMR (100 MHz, MeOH-*d4*)：δ = 150.5 (***C***$_{ar}$OH), 143.6 (***C***$_{ar}$C(1')), 140.3 (***C***(3)H), 130.7 (***C***(5)H), 126.8 (***C***$_{ar}$N), 122.4 (***C***$_{ar}$H), 118.4 (***C***$_{ar}$H), 116.2 (***C***$_{ar}$H), 107.4 (***C***$_{ar}$H), 89.2 (***C***(4')), 80.9 (***C***(1')), 74.4 (***C***(3')), 64.1 (***C***(5')), 44.8 (***C***(2')).

HRMS (ESI+): calculated for $C_{14}H_{17}N_2O_4^+$ $[M+H]^+$: 277.1183, found: 277.1186.

UV λ_{max} = 285 nm, 246 nm.

5-[3-Hydroxy-4-(1*H*-pyrazol-1-yl)phenyl]-2-[[(4-methylbenzoyl)oxy]methyl]tetrahydrofuran-3-yl 4-methylbenzoate (17)



5-[4-(1*H*-Pyrazol-1-yl)-3-[(triisopropylsilyl)oxy]phenyl]-2-[[(4-methylbenzoyl)oxy]-methyl]tetrahydrofuran-3-yl 4-methylbenzoate (**12**) (1.19 g, 1.78 mmol) was dissolved in THF (10 mL) and tetrabutylammonium fluoride solution (1.0 M in THF, 2 mL) was added dropwise at room temperature. The mixture was stirred for 30 min and then diluted with dichloromethane (200 mL). The organic phase was washed three times with water (3×200 mL), then dried over $MgSO_4$ and filtered. Solvent was removed under reduced pressure. The crude was purified by flash chromatography on silica gel (isohexane / ethyl acetate 4:1) to give **17** as white foam (760 mg, 1.48 mmol, 83% yield).

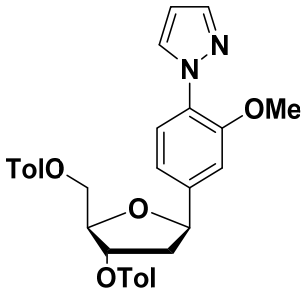
R_f (isohexane / ethyl acetate = 2/1) = 0.61.

^{1}H-NMR (600 MHz, $CDCl_3$): δ = 7.99 (m, 3H, 2×C_{ar}***H***, C(3)***H***), 7.92 (m, 2H, C_{ar}***H***), 7.72 (dd, *J* = 1.8, 0.5 Hz, 1H, C(5)***H***), 7.31 (d, *J* = 8.4 Hz, 1H, C_{ar}***H***), 7.29 – 7.27 (m, 2H, C_{ar}***H***), 7.22 – 7.21 (m, 2H, C_{ar}***H***), 7.14 (d, *J* = 1.7 Hz, 1H, C_{ar}***H***), 6.98-6.97 (m, 1H, C_{ar}***H***), 6.50 (dd, 1H, C(4)***H***), 5.60 (m, 1H, C(3')***H***), 5.23 (dd, *J* = 10.9, 5.0 Hz, 1H, C(1')***H***), 4.64 (dd, *J* = 4.2, 2.7 Hz, 2H, C(5')***H***$_2$), 4.54 (dd, *J* = 4.2, 1.9 Hz, 1H, C(4')***H***), 2.55 (dd, *J* = 13.8, 5.2, 0.9 Hz, 1H, C(2')***H***H), 2.43 (s, 3H, OC***H***$_3$), 2.41 (s, 3H, OC***H***$_3$), 2.24 (ddd, *J* = 13.8, 11.0, 6.0 Hz, 1H, C(2')***H***H).

^{13}C-NMR (151 MHz, $CDCl_3$): δ = 166.4 (***C***OO), 166.1 (***C***OO), 149.3 (***C***$_{ar}$OH), 144.1 (***C***$_{ar}$CH$_3$), 143.8 (***C***$_{ar}$CH$_3$), 140.6 (***C***$_{ar}$C(1')), 138.8 (***C***(5)H), 129.7 (4×***C***$_{ar}$H), 129.2 (4×***C***$_{ar}$H), 127.0 (2×***C***$_{ar}$CO), 126.6 (***C***(3)H), 124.3 (***C***$_{ar}$N), 118.0 (***C***$_{ar}$H), 116.7 (***C***$_{ar}$H), 116.4 (***C***$_{ar}$H), 106.8 (***C***(4)H), 83.0 (***C***(4')), 80.1 (***C***(1')), 77.1 (***C***(3')), 64.7 (***C***(5')), 41.5 (***C***(2')), 21.7 (C$_{ar}$***C***H$_3$), 21.6 (C$_{ar}$***C***H$_3$).

HRMS (ESI+): calculated for $C_{30}H_{29}N_2O_6^+$ $[M+H]^+$: 513.2026, found: 513.2028.

5-[3-Methoxy-4-(1*H*-pyrazol-1-yl)phenyl]-2-[[(4-methylbenzoyl)oxy]methyl]tetrahydrofuran-3-yl 4-methylbenzoate (18)



5-[3-Hydroxy-4-(1*H*-pyrazol-1-yl)phenyl]-2-[[(4-methylbenzoyl)oxy]methyl]tetrahydrofuran-3-yl-4-methylbenzoate (**17**) (741 mg, 1.45 mmol) was dissolved in anhydrous DMF (10 mL) at 0°C and sodium hydride (60 % dispersion in mineral oil, 116 mg, 2.89 mmol) was added slowly. The mixture was stirred for 30 min before the addition of CH_3I (108 µL, 1.74 mmol). The system was allowed to reach room temperature gradually for further 2 h and then diluted with dichloromethane (150 mL). The organic phase was washed with brine (150 mL) and water (2×100 mL), and then dried over $MgSO_4$ and filtered. The solvent was removed under reduced pressure. The crude was purified by flash chromatography on silica gel (isohexane / ethyl acetate 5:1) to give **18** as a colorless oil (582 mg, 1.10 mmol, 76% yield).

R_f (isohexane / ethyl acetate = 2/1) = 0.85.

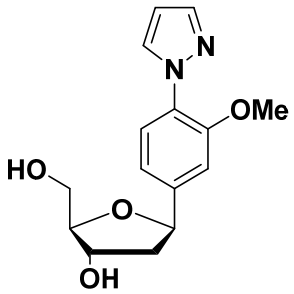
^{1}H-NMR (600 MHz, $CDCl_3$): δ = 7.99 (m, 2H, C_{ar}***H***), 7.92 (m, 2H, C_{ar}***H***), 7.70 (dd, *J* =

1.8, 0.5 Hz, 1H, C(5)***H***), 7.66 (s, 1H, C(3)***H***), 7.65 (s, 1H, C(5)***H***), 7.29 – 7.27 (m, 3H, C_{ar}***H***), 7.22 – 7.21 (m, 2H, C_{ar}***H***), 7.11 (d, *J* = 1.7 Hz, 1H, C_{ar}***H***), 7.02 (m, 1H, C_{ar}***H***), 6.41 (dd, *J* = 2.4, 1.9 Hz, 1H, C(4)***H***), 5.65 (m, 1H, C(3')***H***), 5.28 (m, 1H, C(1')***H***), 4.79 (dd, *J* = 11.8, 3.5 Hz, 1H, C(5')***H***H), 4.63 (dd, *J* = 11.8, 3.7 Hz, 1H, C(5')***H***H), 4.16 (dt, *J* = 3.6, 1.9 Hz, 1H, C(4')***H***), 3.70 (s, 3H, $C_{ar}OC\boldsymbol{H}_3$), 2.59 (dd, *J* = 13.8, 5.1 Hz, 1H, C(2')***H***H), 2.44 (s, 3H, $C_{ar}C\boldsymbol{H}_3$), 2.39 (s, 3H, $C_{ar}C\boldsymbol{H}_3$), 2.24 (ddd, *J* = 13.8, 11.0, 6.0 Hz, 1H, $C_{ar}C\boldsymbol{H}_3$).

^{13}C-NMR (151 MHz, $CDCl_3$): δ = 166.3 (***C***OO), 166.2 (***C***OO), 151.4 ($\boldsymbol{C}_{ar}OCH_3$), 144.2 ($\boldsymbol{C}_{ar}CH_3$), 144.0 ($\boldsymbol{C}_{ar}CH_3$), 141.0 ($\boldsymbol{C}_{ar}$C(1')), 139.9 (***C***(5)H), 131.5 (***C***(3)H), 129.7 (2×$\boldsymbol{C}_{ar}$H), 129.6 (2×$\boldsymbol{C}_{ar}$H), 129.2 (4×$\boldsymbol{C}_{ar}$H), 128.9 ($\boldsymbol{C}_{ar}$N), 127.0 (2×$\boldsymbol{C}_{ar}$CO), 125.1 ($\boldsymbol{C}_{ar}$H), 118.5 ($\boldsymbol{C}_{ar}$H), 109.3 ($\boldsymbol{C}_{ar}$H), 106.1 (***C***(4)H), 83.3 (***C***(4')), 80.4 (***C***(1')), 77.3 (***C***(3')), 64.6 (***C***(5')), 55.7 ($C_{ar}O\boldsymbol{C}H_3$), 42.1 (C(2')), 21.7 (2×$C_{ar}\boldsymbol{C}H_3$).

HRMS (ESI+): calculated for $C_{31}H_{31}N_2O_6^+$ $[M+H]^+$: 527.2182, found: 527.2185.

2-(Hydroxymethyl)-5-[3-methoxy-4-(1*H*-pyrazol-1-yl)phenyl]tetrahydrofuran-3-ol (19)



5-[3-Methoxy-4-(1*H*-pyrazol-1-yl)phenyl]-2-[[(4-methylbenzoyl)oxy]methyl]tetrahydrofuran-3-yl- 4-methylbenzoate (**18**) (570 mg, 1.08 mmol) and potassium carbonate (329 mg, 2.38 mmol) were dissolved in menthol (5 mL). The mixture was stirred at room temperature for 2 h then diluted with water (50 mL). The aqueous phase was extracted with chloroform (4×50 mL). The combined organic layer was dried over $MgSO_4$ and filtered. Solvent was removed under reduced pressure. The crude was purified by flash chromatography on silica gel (chloroform / methanol 20:1) to give **19** as white foam (283 mg, 975 µmol, 90% yield).

R_f (chloroform / methanol = 10/1) = 0.30.

^{1}H-NMR (600 MHz, MeOH-*d4*): δ = 8.01 (dd, *J* = 2.5, 0.6 Hz, 1H, C(3)***H***), 7.67 (dd, *J* = 1.9, 0.5 Hz, 1H, C(5)***H***), 7.50 (d, *J* = 8.1 Hz, 1H, C_{ar}***H***), 7.29 (d, *J* = 1.6 Hz, 1H, C_{ar}***H***), 7.09 (ddd, *J* = 8.2, 1.7, 0.6 Hz, 1H, C_{ar}***H***), 6.46 (dd, *J* = 2.4, 2.0 Hz, 1H, C(4)***H***), 5.19 (dd, *J* = 10.3, 5.6 Hz, 1H, C(1')***H***), 4.34 (dt, *J* = 5.9, 2.1 Hz, 1H, C(3')***H***), 3.98 (td, *J* =

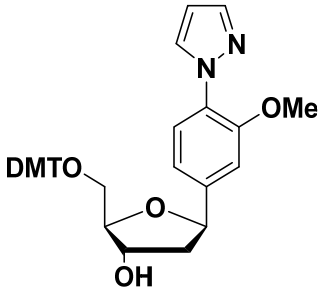
4.9, 2.6 Hz, 1H, C(4')***H***), 3.89 (s, 3H, $C_{ar}OC\boldsymbol{H}_3$), 3.71 (d, *J* = 4.9 Hz, 2H, C(5')$\boldsymbol{H}_2$), 2.26 (ddd, *J* = 13.1, 5.6, 1.8 Hz, 1H, C(2')***H***H), 2.07 – 1.94 (m, 1H, C(2')***H***H).

^{13}C NMR (151 MHz, MeOH-*d4*): δ = 153.4 ($\boldsymbol{C}_{ar}OCH_3$), 145.0 ($\boldsymbol{C}_{ar}$C(1')), 140.9 (***C***(5)H), 133.4 (***C***(3)H), 129.7 ($\boldsymbol{C}_{ar}$N), 126.4 ($\boldsymbol{C}_{ar}$H), 119.4 ($\boldsymbol{C}_{ar}$H), 111.3 ($\boldsymbol{C}_{ar}$H), 107.2 (***C***(4)H), 89.3 (***C***(4')), 81.1 (***C***(1')), 74.3 (***C***(3')), 64.0 (***C***(5')), 56.5 ($C_{ar}O\boldsymbol{C}H_3$), 45.1 (***C***(2')).

HRMS (ESI+): calculated for $C_{15}H_{19}N_2O_4^+$ $[M+H]^+$: 291.1339, found: 291.1345.

UV λ_{max} = 284.5, 248.7 nm.

2-[(Bis(4-methoxyphenyl)(phenyl)methoxy]methyl]-5-[3-methoxy-4-(1*H*-pyrazol-1-yl)phenyl]tetrahydrofuran-3-ol (20)



2-(Hydroxymethyl)-5-[3-methoxy-4-(1*H*-pyrazol-1-yl)phenyl]tetrahydrofuran-3-ol (**19**) (265 mg, 913 µmol) and 4,4'-dimethoxytrityl chloride (340 mg, 1.00 mmol) were dissolved in anhydrous dichloromethane (10 mL) and stirred at 0°C. Then *N*, *N*-diisopropylethylamine (378 µL, 2.28 mmol) was added dropwise. The reaction was allowed to reach room temperature for 3 h and then diluted with dichloromethane (150 mL). The organic phase was washed twice with saturated $NaHCO_3$ (2×150 mL), and then dried over $MgSO_4$ and filtered. The solvent was removed under reduced pressure. The crude was purified by flash chromatography on silica gel (isohexane / ethyl acetate 2:1 +3% triethylamine) to give **20** as a white foam (408 mg, 688 µmol, 75% yield).

R_f (isohexane / ethyl acetate = 2/1) = 0.32.

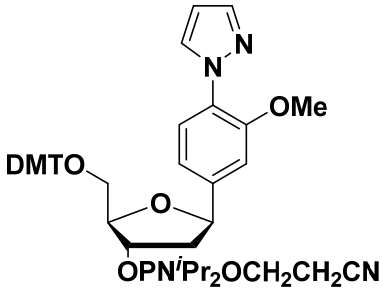
^{1}H-NMR (600 MHz, $CDCl_3$): *δ* = 8.00 (dd, *J* = 2.4, 0.6 Hz, 1H, C(3)***H***), 7.69 (dd, *J* = 1.8, 0.5 Hz, 1H, C(3)***H***), 7.66 (d, *J* = 8.1 Hz, 1H, $C_{ar}\boldsymbol{H}$), 7.49 – 7.43 (m, 2H, $C_{ar}\boldsymbol{H}$, DMTr), 7.39 – 7.31 (m, 3H, $C_{ar}\boldsymbol{H}$, DMTr), 7.30 – 7.25 (m, 3H, $C_{ar}\boldsymbol{H}$, DMTr), 7.21 (dt, *J* = 9.2, 4.3 Hz, 1H, $C_{ar}\boldsymbol{H}$), 7.15 (d, *J* = 1.6 Hz, 1H, $C_{ar}\boldsymbol{H}$), 7.06 – 6.99 (m, 1H, $C_{ar}\boldsymbol{H}$), 6.86 – 6.76 (m, 4H, DMTr), 6.41 (dd, *J* = 2.4, 1.9 Hz, 1H, C(4)***H***), 5.20 (dd, *J* = 10.2, 5.5 Hz, 1H, C(1')***H***), 4.67 – 4.36 (m, 1H, C(3')***H***), 4.11 – 4.06 (m, 1H, C(4')***H***), 3.78 (s, 6H, $C_{ar}OC\boldsymbol{H}_3$, DMTr), 3.68 (s, 3H, $C_{ar}OC\boldsymbol{H}_3$), 3.36 (ddd, *J* = 13.8, 9.9, 4.5 Hz, 2H, C(5')$\boldsymbol{H}_2$), 2.30 – 2.24 (m, 1H, C(2')***H***H), 2.11 (ddd, *J* = 13.1, 10.2, 6.0 Hz, 1H,

C(2')***H***H).

^{13}C-NMR (151 MHz, $CDCl_3$): δ = 158.5 (2×***C***$_{ar}$$OCH_3$, DMTr), 151.4 (***C***$_{ar}$$OCH_3$), 144.8 (O***C***$_{ar}$, DMTr), 142.2 (***C***$_{ar}$C(1')), 140.0 (***C***(5)H), 136.0 (2×O***C***$_{ar}$, DMTr), 131.4 (***C***(3)H), 130.1 (4×***C***$_{ar}$H, DMTr), 128.9 (***C***$_{ar}$N), 128.2 (2×***C***$_{ar}$H, DMTr), 127.8 (2×***C***$_{ar}$H, DMTr), 126.8 (***C***$_{ar}$H, DMTr), 125.0 (***C***$_{ar}$H), 118.7 (***C***$_{ar}$H), 113.1 (4×***C***$_{ar}$H, DMTr), 109.7 (***C***$_{ar}$H), 106.1 (***C***(2)H), 86.5 (***C***(4')), 86.2 (O***C***C_{ar}, DMTr), 79.8 (***C***(1')), 74.6 (***C***(3')), 64.2 (***C***(5')), 55.8 (C_{ar}O***C***H_3), 55.2 (2×C_{ar}O***C***H_3, DMTr), 44.3 (***C***(2')).

HRMS (ESI+): calculated for $C_{36}H_{37}N_2O_6^+$ $[M+H]^+$: 593.2652, found: 593.2653.

2-[[Bis(4-methoxyphenyl)(phenyl)methoxy]methyl]-5-[3-methoxy-4-(1*H*-pyrazol-1-yl)phenyl]tetrahydrofuran-3-yl (2-cyanoethyl) diisopropylphosphoramidite (21)



In a Schlenck tube, the DMTr-protected nucleoside **20** (93 mg, 157 µmol) and diisopropylammonium tetrazolide (107 mg, 628 µmol) were dissolved in rigorously degassed dichloromethane (3 mL). The solution was then degassed three more times. Bis(diisopropylamino)(2-cyanoethoxy)phosphine (100 µL, 314 µmol) was added to the solution. The reaction mixture was stirred at room temperature for 20 min under argon atmosphere, then directly purified by flash chromatography on silica gel (isohexane / ethyl acetate 2:1 +3% triethylamine) to give **21** as a white foam (123 mg, 157 µmol, quant.).

R_f (isohexane / ethyl acetate = 2/1) = 0.65.

^{31}P-NMR (81 MHz, $CDCl_3$): δ = 149.01, 148.87.

HRMS (ESI+): calculated for $C_{45}H_{53}N_4O_7P^+$ $[M+H]^+$:793.3730, found: 793.3731.

(3-Hydroxy-5-(3-hydroxy-4-(1*H*-pyrazol-1-yl)phenyl)tetrahydrofuran-2-yl)methyl tetrahydrogen triphosphate (dPzTP, 22)

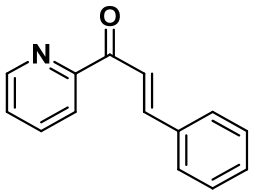
To a stirring solution of bis-tetrabutylammonium pyrophosphate (239 mg, 362 µmol) in anhydrous DMF (0.25 mL) under argon, anhydrous tributylamine (0.3 mL, 1.26 mmol) was added. This solution was transferred to a flask containing a solution of

2-chloro-1,3,2-benzodioxaphosphorin-4-one (73 mg, 362 µmol) in anhydrous DMF (0.25 mL) and stirred for 30 min at room temperature. Then, this mixture was added to the anhydrous DMF solution of **16** (50 mg, 181 µmol) and stirred for 1 h. When TLC (dichloromethane / methanol = 5/1) indicated complete conversion of the nucleoside, the mixture was oxidized with an iodine solution (3% I_2 in pyridine / H_2O 9:1) until a brown permanent color was observed (about 1.6 mL) and stirred for 15 min. Water (2 mL) was added and stirring continued for 1.5 h. The resulting solution was transferred to a 50 mL centrifuge tube, and NaCl (3 M, 0.5 mL) was added followed by ethanol (95%, 15 mL) and then vortexed. After 1 h stand still at -80 °C, the precipitate was collected by centrifugation (20 min, 3200 xg) and the resulting pellet was purified twice by preparative RP-HPLC (0-30% Buffer B in 45 min). This afforded the **dPzTP 22** (11% yield according to concentration calculated from UV absorbance) as a slight yellow tris- to tetrakis-triethylammonium salt after freeze-drying.

^{31}P-NMR (162 MHz, D_2O): δ = -11.3 (m, 2P, α-, γ-***P***), -23.6 (t, 1P, J = 19.3 Hz, β-***P***).

HRMS (ESI-): calculated for $C_{14}H_{18}N_2O_{13}P_3^-$ $[M-H]^-$: 515.0027; found: 515.0017.

3-Phenyl-1-(2-pyridyl)prop-2-en-1-one (24)



To a stirred aqueous solution of sodium hydroxide (10%, 1.0 mL) and ethanol solution (20 mL) of benzaldehyde (1.68 mL, 16.5 mmol), 2-acetylpyridine (1.91 mL, 17.0 mmol) was added dropwise over 2h. The temperature was kept at 0 °C. After being stirred for another 2 h, the reaction mixture was filtered to give **24** as a faint yellow solid (2.97 g, 14.2 mmol, 86%).

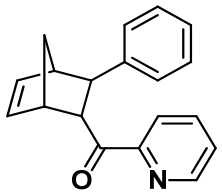
R_f (isohexane / ethyl acetate = 10/1) = 0.53.

^{1}H NMR (300 MHz, $CDCl_3$): δ = 8.75 (m, 1H), 8.32 (d, J = 15 Hz, 1H), 8.20 (m, 1H), 8.00 (d, J = 15 Hz, 1H), 7.89 (m, 1H), 7.74 (m, 2H), 7.49 (m, 1H), 7.42 (m, 3H).

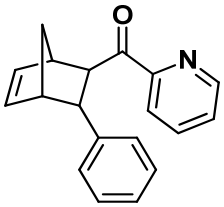
^{13}C NMR (75 MHz, $CDCl_3$): δ = 189.3, 154.1, 148.7, 144.8, 137.1, 135.1, 130.6, 128.8 (4C), 126.9, 122.9, 120.9.

HRMS (ESI+): calculated for $C_{14}H_{12}NO^+$ $[M+H]^+$: 210.0913; found: 210.0912.

(3-Phenylbicyclo[2.2.1]hept-5-en-2-yl)(pyridin-2-yl)methanone (26)



26 endo



26 exo

3-Phenyl-1-(2-pyridyl)prop-2-en-1-one (**24**) (209 mg, 1.0 mmol, 1 eq.), freshly distilled cyclopentadiene (**25**) (84 µL, 1.0 mmoL, 1 eq.), and $Cu(NO_3)_2 \cdot 2.5H_2O$ (23.2 mg, 0.1 mmol, 0.1 eq.) were dissolved in a mixture of H_2O and acetonitrile (v/v 1/1, 5 mL). The mixture was stirred at room temperature for 4 h, diluted with H_2O (10 mL). exacted twice with ethyl acetate (2×15 mL), dried over $MgSO_4$ and filtered. The solvent was removed under reduced pressure. The crude was purified by flash chromatography on silica gel (isohexane / ethyl acetate 10:1) to give endo/exo mixture **26** as a colorless oil (269 mg, 0.98 mmol, 98%).

R_f (isohexane / ethyl acetate = 10/1) = 0.65.

^{1}H NMR (600 MHz, $CDCl_3$): δ = 8.69 (m, 1H), 8.02 (d, *J* = 7.9 Hz, 1H), 7.84 (t, *J* = 7.9 Hz, 1H), 7.47 (m, 1H), 7.30 (m, 4H), 7.16 (t, *J* = 7.2 Hz, 1H), 6.50 (s, 1H), 5.83 (m, 1H), 4.55 (m, 1H), 3.55 (s, 1H), 3.45 (d, *J* = 4.8 Hz, 1H), 3.09 (s, 1H), 2.08 (d, *J* = 8.4 Hz, 1H), 1.60 (d, *J* = 8.4 Hz, 1H).

^{13}C NMR (150 MHz, $CDCl_3$): δ = 200.8, 153.3, 148.6, 144.5, 139.5, 137.1, 132.8, 128.4 (2C), 127.6 (2C), 126.9, 125.8, 122.3, 54.3, 49.4, 48.7, 48.2, 45.6.

HRMS (ESI+): calculated for $C_{19}H_{18}NO^+$ $[M+H]^+$: 276.1388; found: 276.1382.

6.3. Part I: Experimental details for evaluation and application

Sample preparation and melting point experiments

Melting profiles were measured on a *JASCO* V-650 spectrometer using quartz glass cuvettes with 10.00 mm path length. The samples contained 150 mM NaCl, 10 mM CHES buffer and 1 μM of each strand in a final volume of 200 μL. For strands **ODN 1**, **3-7a**/**b** the oligonucleotides with Cu^{2+} were hybridized by slowly cooling down the samples from 90 °C to room temperature overnight. Each sample was prepared and measured at least three times in parallel. The solutions were covered with silicon oil and tightly plugged. UV Absorbance was recorded in the forward and reverse direction at temperatures from 20 °C to 80 °C with a slope of 1 °C/min. Three denaturing and renaturing ramps were performed for each sample and averaged for evaluation of the melting point. T_M values were calculated as the zero-crossing of 2nd derivate of the 349 nm background corrected change in hyperchromicity at 260 nm.

UV / CD spectra measurements

UV and CD spectra were measured on a *JASCO* V-650 or J-810 spectrometer using quartz glass cuvettes with 10.00 mm path length. For the strands with mixed salen and pyrazole bases, **ODN 8a/b**, **9a/b** and **10a/b**, the duplex (3 μM of each strand) was first reannealed in 150 mM NaCl, 10 mM CHES buffer without Cu^{2+} ions or ethylenediamine overnight. Then an excess of ~30 fold ethylenediamine, diluted in degassed ddH_2O, was added, and the solution was incubated at 4 °C overnight. Finally, different amounts of Cu^{2+} were added into individual samples and the final volume was constant (200 μL). After Cu^{2+} had been added, the samples were kept at 4 °C and measured in 2 h. Each sample was prepared and measured at least twice in parallel at 20 °C. For UV measurements, five spectra were accumulated. For CD measurements, ten spectra from 320 nm to 210 nm were accumulated at 20°C with a scanning speed of 500 nm/min. A blank correction was made of an aqueous solution of buffer and salt

and measured for each series of titration measurement.

ESI-Mass measurements

For duplex **ODN 8a/b**, 30 μM of the duplex was hybridized in 150 mM NH_4OAc as above. Here, no CHES buffer and NaCl aq. were used. After addition of ethylenediamine, the mixture was incubated at 4 °C overnight; then Cu^{2+} (3 eq.) was added. Before ESI measurement, a mixture (10 μL) of imidazole (250 mM) and piperidine (250 mM) in 80 % aq. acetonitrile was added to the sample (40 μL).

MALDI-TOF Mass measurements

For confirmation of the primer extension product, the corresponding gel was cut into pieces and immersed in 1×TBE buffer at 35°C for 4 hrs. The solution was then concentrated, desalted on *Sep-Pak* C18 column, and further concentrated before MALDI measurement. Alternatively, at the end of the reaction, the primer extension assays were combined (8×20 μL) with H_2O (40 μL). Phenol (200 μL) was added and mixed. Then the aqueous phase was transferred to a fresh tube and extracted twice with chloroform (200 μL). The aqua phase was concentrated, desalted via *Zip Tip* C18 and further concentrated before MALDI measurement.

Primer extension experiments and PAGE assay

The reaction system consisted of 1.5 μM fluorescein-labeled primer, 2 μM template, 1× corresponding buffer, additive copper ion and magnesium ion and water. The mixture was first heated to 90°C for 4 min and then cooled down to 15°C for 1 h. Then, DNA polymerase (2 U/μL, 0.5 μL, 1 U) was added; the mixture was incubated at 37°C for another 15 min before the triphosphate added to start the reaction. For competitive extension experiments, different amount of dPzTP and dNTP were mixed prior to the addition. After the reaction had been stopped, the sample was immediately cool down in an ice bath, centrifuged and mixed with 1:1 with 2× loading buffer. The

mixture was heated to 90°C for 10 min for denaturing the polymerase. PAGE assay was run at a constant electric current of 35 mA at 40°C.

Procedure of Diels-Alder reaction using pyrazole ligandoside based catalyst

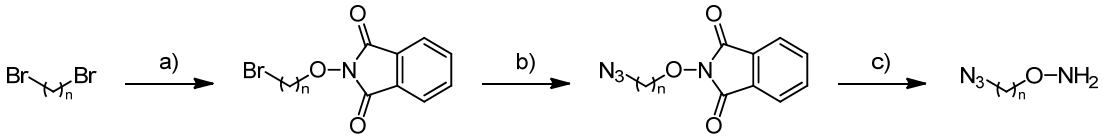
Duplex **ODN 6a**/**b** (5 nmol), Cu^{2+} salts (1 mM, 25 μL, 25 nmol, 5 eq., 2.5 mol%), ammonium acetate aq. (pH = 9.0, 0.1 M, 30 μL), NaCl aq. (0.5 M, 100 μL) and Milli-Q water (145 μL) were mixed to a final volume of 300 μL. The duplex was hybridized by heating at 90 °C for 5 min, then slowly cooled down to room temperature in a water bath. The solution was lyophilized to dryness and redissolved in MOPS buffer (pH = 6.5, 7.4 or 9.0, 20 mM, 1 mL). Acetonitrile solution of dienophile **24** (0.1 M, 10 μL, 1 μmol) was added. Freshly distilled cyclopentadiene **25** (7.5 μL, excessive), was added to initiate the reaction. The mixture was shaken for 24 hours at 4 °C, followed by the extraction with diethyl ether (3×4 mL), dried over $MgSO_4$, filtered and concentrated. The crude product was dissolved in heptane and 2-propanol (v/v 99/1) and analyzed by normal phase chiral HPLC directly, i.e. *Shimadzu* CBM-20A communication bus module, SPD-M20A prominence diode array detector and *Daicel Chiralpak* IB 81325, 250 mm × 4.6 mm. Conditions: 1% 2-propanol in heptanes, 0.5 mL/min, column temperature 40 °C, detected from 190 nm to 500 nm, integration at 212 nm. Conversion of **24** was calculated using the following formula: (PA: peak area)

$$\text{Conversion of } \mathbf{24}\ (\%) = \frac{PA_{\mathbf{26}}}{PA_{\mathbf{26}} + PA_{\mathbf{24}}/_{f}},$$

where PA_{24} and PA_{26} are the HPLC peak areas of **24** and **26**, respectively, and *f* is the correction factor determined to be 0.69 from the fitting curve.

6.4. Part II: Synthesis of azide linkers and related compounds

General procedure of hydroxylamine linker 31-33



37 **38** **39** **31-33**

Dibromine alkane **37** (1.5 eq.) was added to the solution of *N*-hydroxy-phthalimide (PhthNOH, 1.0 eq.) and triethylamine (2.5 eq.) in anhydrous dimethylformamide. The mixture was stirred at room temperature for 24 h. The reaction was diluted with water, and the aqueous phases were extracted three times with ethyl acetate. The combined organic phases were dried over $MgSO_4$, filtered and concentrated to give the crude product **38** as a white solid, which was dissolved in anhydrous dimethylformamide and sodium azide (1.2 eq.) was added. The mixture was stirred at room temperature for 2 h, diluted with water and extracted with ethyl acetate three times. The combined organic phases were dried over $MgSO_4$, filtered and concentrated. The crude was purified by flash chromatography on silica gel (isohexane/ethyl acetate 10:1→ 2:1) to give **39** as a yellow oil. The oil was redissolved in hydrazine monohydrate (1.5 eq.) and dichloromethane. The mixture was stirred at room temperature for 24 h and then filtered. The solution was diluted with dichloromethane and washed with NaCl aq. three times. The combined organic phases were dried over $MgSO_4$, filtered and concentrated to give **31-33** as a colorless oil.

***O*-(3-Azidopropyl)hydroxylamine (31)**

67% yield for three steps.

R_f (DCM / MeOH = 10/1) = 0.42.

^{1}H-NMR (400 MHz, $CDCl_3$): δ = 3.78 (t, J = 6.4 Hz, 2H, C(1)***H*₂***), 3.38 (t, J = 6.8 Hz, 2H, C(3)***H*₂***), 1.91−1.83 (m, 2H, C(2)***H*₂***).

^{13}C-NMR (100 MHz, $CDCl_3$): δ = 72.4 (***C***(1)H_2), 48.3 (***C***(3)H_2), 27.8 (***C***(2)H_2).

HRMS (ESI+): calculated for $C_3H_9N_4O^+$ $[M+H]^+$: 117.0771, found: 117.0771.

***O*-(4-Azidobutyl)hydroxylamine (32)**

57% yield for three steps.

R_f (DCM / MeOH = 10/1) = 0.42.

^{1}H-NMR (400 MHz, $CDCl_3$): δ = 3.70 (t, J = 5.6 Hz, 2H, C(1)***H₂***), 3.38 (t, J = 6.4 Hz, 2H, C(4)***H₂***), 1.68−1.64 (m, 4H, C(2)***H₂***-C(3)***H₂***).

^{13}C-NMR (100 MHz, $CDCl_3$): δ = 75.1 (***C***(1)H_2), 51.2 (***C***(4)H_2), 25.6 (***C***H_2), 25.5 (***C***H_2).

HRMS (ESI+): calculated for $C_4H_{11}N_4O^+$ $[M+H]^+$: 131.0927, found: 131.0928.

***O*-(5-Azidopentyl)hydroxylamine (33)**

52% yield for three steps.

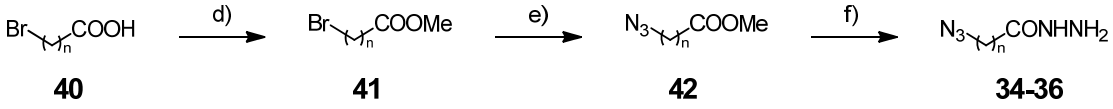
R_f (DCM / MeOH = 10/1) = 0.38.

^{1}H-NMR (400 MHz, $CDCl_3$): δ = 3.63 (t, J = 6.4 Hz, 2H, C(1)***H₂***), 3.21 (t, J = 6.5 Hz, 2H, C(5)***H₂***), 1.59−1.51 (m, 4H, C(2)***H₂***, C(4)***H₂***), 1.40−1.34 (m, 2H, C(3)***H₂***).

^{13}C-NMR (100 MHz, $CDCl_3$): δ = 75.6 (***C***(1)H_2), 51.2 (***C***(5)H_2), 28.6 (***C***(4)H_2), 27.8 (***C***(2)H_2), 23.1 (***C***(3)H_2).

HRMS (ESI+): calculated for $C_5H_{13}N_4O^+$ $[M+H]^+$: 145.1084, found: 145.1084.

General procedure of hydrazide linker 34-36



40 **41** **42** **34-36**

ω-Bromine aliphatic carboxylic acid **40** (1.0 eq.) was dissolved in methanol. Sulfuric acid (98%, 2.5 eq.) was then added to the solution slowly. The mixture was refluxed for 4 h. After cooled down to room temperature, the solvent was evaporated, and the residue was neutralized with 10% NaOH aq. and extracted with dichloromethane. The combined organic phase was dried over $MgSO_4$, filtered and concentrated to give the respective methyl ester **41** as a yellowish oil. The crude ester and sodium azide (2.0 eq.) were then redissolved in anhydrous dimethylformamide. The mixture was stirred at room temperature for 24 h, diluted with water and extracted with ethyl acetate three times. The combined organic phases were dried over $MgSO_4$, filtered and concentrated. The crude was purified by flash chromatography on silica gel (isohexane / ethyl acetate 15:1) to give **42** as a yellowish oil. Hydrazine monohydrate (10 eq.) was added to the methanol solution of ester **42**. The mixture was reflux for 4 h. After cooled down to room temperature, the mixture was diluted with H_2O and extracted with

dichloromethane. The combined organic phase was washed with NaCl aq. and dried over $MgSO_4$, filtered and concentrated to give the **34-36** as a colorless oil.

4-Azidobutanhydrazide (**34**)

69% yield for three steps.

$\mathbf{R_f}$ (isohexane / ethyl acetate = 10/1) = 0.38.

^{1}H-NMR (400 MHz, $CDCl_3$): δ = 8.21 (br, 1H, N***H***-NH_2), 3.37 (t, J = 4.8 Hz, 2H, C(2)***H₂***), 2.73 (t, J = 4.8 Hz, 2H, C(4)***H₂***), 1.98−1.94 (m, 2H, C(3)***H₂***).

^{13}C-NMR (100 MHz, $CDCl_3$): δ = 174.1 (***C***O), 50.9 (***C***(4)H_2), 29.5 (***C***(2)H_2), 23.7 (***C***(3)H_2).

HRMS (ESI-): calculated for $C_5H_{10}N_5O_3^-$ $[M+HCO_2]^-$: 188.0789, found: 188.0791.

5-Azidopentanhydrazide (**35**)

65% yield for three steps.

$\mathbf{R_f}$ (isohexane / ethyl acetate = 10/1) = 0.30.

^{1}H-NMR (400 MHz, $CDCl_3$): δ = 8.7 (br, 1H, N***H***-NH_2), 3.6 (br, 2H, NH-N***H₂***), 3.29 (t, J = 7.0 Hz, 2H, C(2)***H₂***),2.18 (t, J = 7.0 Hz, 2H, C(5)***H₂***), 1.77−1.69 (m, 2H, C(3)***H₂***), 1.65−1.58 (m, 2H, C(4)***H₂***).

^{13}C-NMR (100 MHz, $CDCl_3$): δ = 173.2 (***C***O), 51.1 (***C***(5)H_2), 33.7 (***C***(2)H_2), 28.3 (***C***(4)H_2), 22.6 (***C***(3)H_2).

HRMS (ESI-): calculated for $C_6H_{12}N_5O_3^-$ $[M+HCO_2]^-$: 202.0946, found: 202.0947.

6-Azidohexanhydrazide (**36**)

64% yield for three steps.

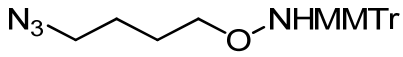
$\mathbf{R_f}$ (*iso*-hexane/ ethyl acetate = 10/1) = 0.25.

^{1}H-NMR (400 MHz, $CDCl_3$): δ = 8.50 (br, 1H, N***H***-NH_2), 3.27 (t, $^3J_{H,H}$ = 8.8 Hz, 2H, C(2)***H₂***), 2.16 (t, $^3J_{H,H}$ = 10.0 Hz, 2H, C(6)***H₂***), 1.73−1.56 (m, 4H, C(5)***H₂***, C(3)***H₂***), 1.45−1.37 (m, 2H, C(4)***H₂***).

^{13}C-NMR (100 MHz, $CDCl_3$): δ = 173.4 (***C***O), 51.1 (***C***(6)H_2), 34.2 (***C***(2)H_2), 28.5 (***C***(5)H_2), 26.3 (***C***(4)H_2), 24.9 (***C***(3)H_2).

HRMS (ESI-): calculated for $C_7H_{14}N_5O_3^-$ $[M+HCO_2]^-$: 216.1102, found: 216.1103.

O-(4-Azidobutyl)-N-[(4-methoxyphenyl)diphenylmethyl]hydroxylamine (47)



O-(4-Azidobutyl)hydroxylamine **32** (1.88 g, 14.4 mmol, 1.0 eq.) was solved in anhydrous dichloromethane (40 mL). 4-Monomethoxytritylchloride (MMTr-Cl, 4.91 g, 15.9 mmol, 1.1 eq.) and diisopropylethylamine (5.0 mL, 28.9 mmol, 2.0 eq.) was added to the mixture at 0°C. The reaction was stirred at room temperature for 2 h, diluted with dichloromethane and washed with $NaHCO_3$ aq., dried over $MgSO_4$, filtered and concentrated. The crude was purified by flash chromatography on silica gel (isohexanes/ethyl acetate 15:1 + 3% triethylamine) to give **47** as a yellowish oil (4.89 4g, 12.2 mmol, 84%).

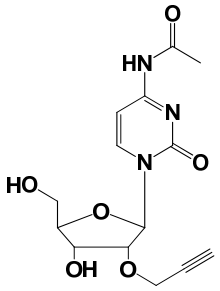
R_f (isohexane / ethyl acetate = 10/1 + 3% triethylamine) = 0.64.

^{1}H-NMR (400 MHz, $CDCl_3$): δ = 7.31−7.18 (m, 12H, 12 × C_{Ar}***H***), 6.79 (d, *J* = 8.8 Hz, 2H, 2 × CH_3-O-C-C***H***), 3.76 (s, 3H, O-C***H₃***),3.65 (t, *J* = 6.0 Hz, 2H, C(1)***H₂***), 3.05 (t, *J* = 6.8 Hz, 2H, C(4)***H₂***), 1.52−1.37 (m, 4H, C(2)***H₂***, C(3)***H₂***).

^{13}C-NMR (100 MHz, $CDCl_3$): δ = 158.3 (CH_3-O-***C***$_{Ar,}$ MMTr), 144.6 (2 × ***C***$_{Ar,}$ MMTr), 136.6 (2 × ***C***$_{Ar}$), 130.2 (2 × ***C***$_{Ar}$), 129.0 (4 × ***C***$_{Ar}$), 127.6 (4 × ***C***$_{Ar}$), 126.7 (2 × ***C***$_{Ar}$), 112.9 (2 × ***C***$_{Ar}$), 77.2 (O-NH-***C***), 73.2 (***C***(1)H_2), 55.2 (O-***C***H_3), 51.1 (***C***(4)H_2), 25.8 (***C***(3)H_2), 25.5 (***C***(2)H_2).

HRMS (ESI-): calculated for $C_{25}H_{27}N_4O_4^-$ $[M+HCO_2]^-$: 447.2038, found: 447.2040.

2'-*O*-Propargyl-N^4-acetyl-cytidine (49)



To a solution of cytidine (**48**) (4.00 g, 16.4 mmol) in DMF (160 mL) at 0 °C were added tetrabutylammonium iodide (1.21 g, 3.28 mmol, 0.2 eq.), propargyl bromide (80% in toluene, 2.30 mL, 21.3 mmol, 1.3 eq.) and sodium hydride (60% in mineral oil, 984 mg, 24.6 mmol, 1.5 eq.). The reaction mixture was warmed up to room temperature and stirred for 18 h. After evaporated of the volatiles, the residue was purified by flash chromatography on silica gel (dichloromethane/ methanol 15:1) to

give a mixture of 2'-*O*-propargyl-cytidine **43** and 3'-*O*-propargylcytidine (3.61 g, 12.8 mmol, 78% yield) as a yellow foam. The mixture of isomers (1.13 g, 4.00 mmol, 1 eq.) was dissolved in DMF (20 mL) and treated with acetic anhydride (0.57 mL, 6.00 mmol, 1.5 eq.). After 16 h stirring at room temperature, the reaction mixture was evaporated to dryness and dissolved in MeOH. The white solid formed was filtered and washed twice with cold methanol. The filter cake was collected and dried under high vacuum to give **49** as a white solid (420 mg, 1.30 mmol, 33% yield). COSY experiment in DMSO-*d6* confirmed the nature of the isomer by the coupling of HO with H3'.

R_f (chloroform / methanol = 10/1) = 0.31.

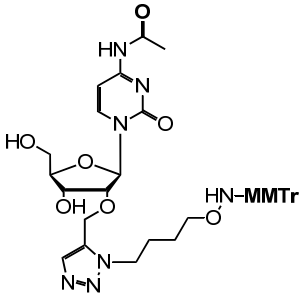
For the byproduct 3'-*O*-propargyl-N^4-acetyl-cytidine, **R_f** = 0.34.

^{1}H-NMR (400 MHz, DMSO-*d6*): δ = 10.91 (br., 1H, N***H***) , 8.36 (d, *J* = 3.7 Hz, 1H, C(6)***H***), 7.20 (d, *J* = 3.7 Hz, 1H, C(5)***H***), 5.86 (d, *J* = 14.6 Hz, 1H, C(1')***H***), 5.22-5.19 (m, 2H, O***H***), 4.39 (d, *J* = 1.2 Hz, 2H, C***H₂***CCH), 4.09 (m, 1H, C(3')***H***), 3.99 (dd, *J* = 14.6, 4.9 Hz, 1H, C(2')***H***), 3.92-3.89 (m, 1H, C(4')***H***), 3.78-3.73 (m, 1H, C(5')***H***H), 3.63-3.58 (m, 1H, C(5')***H***H), 3.42 (t, *J* = 2.4 Hz, 1H, CC***H***), 2.10 (s, 3H, C***H₃***CO).

^{13}C-NMR (100 MHz, DMSO-*d6*): δ = 171.1 (***C***OCH$_3$), 162.5 (***C***(4)), 154.5 (***C***(2)), 145.1 (***C***(6)H), 95.3 (***C***(5)H), 87.9 (***C***(1')), 84.4 (C(4'), 81.0 (***C***(2')), 80.2 (***C***CH), 77.4 (C***C***H), 67.5 (***C***(3')), 59.5 (***C***(5')), 57.0 (***C***H$_2$CCH) , 24.4 (***C***H$_3$CO)

HRMS (ESI+): calculated for $C_{14}H_{18}N_3O_6^+$ $[M+H]^+$: 324.1190, found: 324.1192.

2'-*O*-Propargyl-N^4-acetyl-cytidine connected to MMTr protected linker 32 (50)



49 (403 mg, 1 mmol, 1 eq.), **47** (281 mg, 1 mmol, 1 eq.), $CuSO_4 \cdot 5H_2O$ (25 mg, 0.1 mmol, 0.1 eq.), sodium ascorbate (39 mg, 0.2 mmol, 0.2 eq.) and DIPEA (88 µL, 0.5 mmol, 0.5 eq.) were dissolved in H_2O (3 mL) and MeCN (3 mL) and stirred at room temperature. After 2 h, another $CuSO_4 \cdot 5H_2O$ (25 mg, 0.1 mmol, 0.1 eq.), sodium ascorbate (39 mg, 0.2 mmol, 0.2 eq.) were added and stirred for further 2 h. Then the

mixture was directly purified by flash chromatography on silica gel (chloroform / methanol 10:1 + 3% triethylamine) to give **50** (715 mg, 0.99 mmol, 99%) as a light yellow foam.

R_f (chloroform/ methanol = 10/1 + 3% triethylamine) = 0.50.

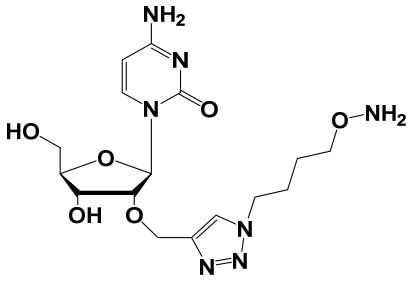
^{1}H-NMR (400 MHz, $CDCl_3$): δ = 8.48 (d, *J* = 7.4 Hz, 1H, C(6)***H***), 7.42 (d, *J* = 7.4 Hz, 1H, C(5)***H***), 7.39 (s, 1H, C_{azole}***H***), 7.35-7.22 (m, 12H, C_{Ar}***H***), 6.85-6.82 (m, 2H, C_{Ar}***H***), 5.90 (d, *J* = 1.5 Hz, 1H, C(1')***H***), 5.02 (dd, 2H, *J* = 6.4, 11.5 Hz, C(5')***H₂***), 4.32-4.11 (m, 5H, C(2')***H***, C(3')***H***, C(4')***H***, $C_{azole}NC$***H₂***CH_2), 4.00 (dd, *J* = 5.0, 62 Hz, 2H, OC***H₂***C_{azole}), 3.80 (s, 3H, OC***H₃***), 3.73 (t, *J* = 6.0 Hz, 2H, CH_2C***H₂***ON), 2.24 (s, 3H, C***H₃***CO), 1.77 (dt, *J* = 7.0, 14.2 Hz, 2H, $C_{azole}NCH_2C$***H₂***), 1.47 (dt, *J* = 6.4, 12.9 Hz, 2H, C***H₂***CH_2ON).

^{13}C-NMR (100 MHz, $CDCl_3$): δ = 170.5 (***C***OCH_3), 162.5 (***C***(4)), 158.3 (CH_3-O-***C***$_{Ar,}$ MMTr), 155.4 (***C***(2)), 146.1 (OCH_2***C***$_{azole}$H), 144.6 (2 × ***C***$_{Ar,}$ MMTr), 143.8 (***C***(6)H), 136.5 (2 × ***C***$_{Ar}$, MMTr), 130.2 (2 × ***C***$_{Ar}$, MMTr), 129.0 (4 × ***C***$_{Ar}$, MMTr), 127.7 (4 × ***C***$_{Ar}$, MMTr), 126.8 (2 × ***C***$_{Ar}$, MMTr), 123.1 (***C***$_{azole}$HNCH_2CH_2), 113.0 (2 × ***C***$_{Ar}$, MMTr), 96.5 (***C***(5)H), 90.9 (***C***(1')), 84.9 (C(4'), 81.4 (***C***(2')), 77.2 (O-NH-***C***, MMTr), 73.5 (C_{azole}HN***C***H_2***C***H_2), 73.3 (CH_2***C***H_2ON), 67.6 (***C***(3')), 63.6 (O***C***H_2C_{azole}), 60.1 (***C***(5')), 55.2 (O-***C***H_3, MMTr), 27.4 (***C***H_3CO), 25.1 ($C_{azole}NCH_2$***C***H_2), 25.0 (***C***H_2CH_2ON).

HRMS (ESI+): calculated for $C_{38}H_{44}N_7O_8^+$ $[M+H]^+$: 726.3251, found: 726.3255.

UV λ_{max} = 273 nm.

2'-*O*-Propargyl-cytidine connected to linker 32 (52)



50 (700 mg, 0.97 mmol) was dissolved in methanol solution of ammonia (7 M, 5 mL) and stirred at room temperature overnight. TLC showed a quantitative deprotection to give **51**. The solution was then concentrated to dryness and a part of the remains (10 mg) redissolved in acetate acid aq. (v/v 50%, 1 mL) in a 2 mL *Eppendorf* tube. The tube was shaken at 25 °C overnight until the mixture became a milky turbid solution.

The mixture was extracted with ethyl acetate twice then concentrated to dryness to give **52** as a while solid. Analytical HPLC showed a quantitative deprotection. 0% - 70% Buffer B in 45 min.

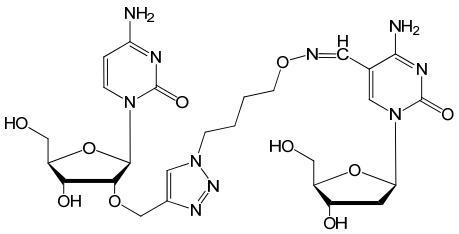
Retention time: 14.9 min.

^{1}H-NMR (800 MHz, D_2O): δ = 7.98 (s, 1H, C_{azole}***H***), 7.69 (d, *J* = 7.6 Hz, 1H, C(6)***H***), 5.88 (d, *J* = 5.2 Hz, 1H, C(5)***H***), 4.81 (m, 1H, C(1')***H***), 4.78 (dd, 2H, *J* = 6.6, 11.0 Hz, C(5')***H$_2$***), 4.40 (t, *J* = 7.0 Hz, 2H, CH_2C***H$_2$***ON), 4.31 (t, *J* = 5.0 Hz, 1H, C(3')***H***), 4.21 (t, *J* = 5.4 Hz, 1H, C(4')***H***), 4.10 (dd, *J* = 2.0, 7.4 Hz, 1H, C(2')***H***), 3.88 (t, *J* = 6.0 Hz, 2H, OC***H$_2$***C_{azole}), 3.81 (ddd, *J* = 1.4, 12.7, 60.8 Hz, 2H, $C_{azole}NC$***H$_2$***CH_2), 1.94 (dt, *J* = 11.2, 3.5 Hz, 2H, $C_{azole}NCH_2C$***H$_2$***), 1.57 (dt, J = 7.5, 3.2 Hz, 2H, C***H$_2$***CH_2ON).

HRMS (ESI+): calculated for $C_{16}H_{26}N_7O_6^+$ $[M+H^+]^+$:412.1939, found: 412.1942.

UV λ_{max} = 271 nm.

Crosslinker dinucleoside rC-fdC (53)



52 (12.0 µmol, 1.02 eq.) and fdC nucleoside (3.00 mg, 11.75 µmol, 1 eq.) were dissolved in acetonitrile (400 µL) and H_2O (200 µL) in 2 mL *Eppendorf* tube. To the solution, 4-methoxyaniline solution in ddH_2O/DMSO (v/v 9/1, acidified to pH = 5.5 with acetate acid, 9.6 µL, 250 mM, 2.4 µmol, 0.2 eq.) was added. The tube was shaken at 25 °C, 400 rpm for 12 h, then purified with HPLC: 0% - 70% Buffer B in 45 min, quant.

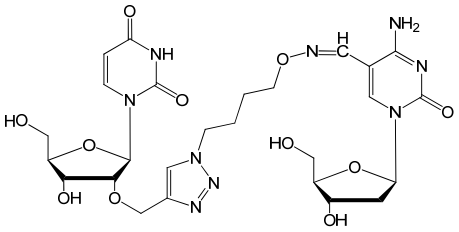
Retention time: 21.0 min.

^{1}H-NMR (800 MHz, D_2O): δ = 8.14 (s, 1H), 7.98 (d, *J* = 23.6 Hz, 2H), 7.58 (d, *J* = 7.6 Hz, 1H), 6.87 (m, 2H), 6.82 (m, 2H), 6.22 (t, *J* = 6.4 Hz, 1H), 5.91 (d, *J* = 7.6 Hz, 1H), 5.83 (d, *J* = 5.4 Hz, 1H), 4.63 (d, *J* = 13.3 Hz, 1H), 4.41 (m, 3H), 4.29 (t, *J* = 5.0 Hz, 1H), 4.15 (t, *J* = 5.4 Hz, 1H), 4.10 (m, 4H), 3.85 (m, 2H), 3.76 (m, 2H), 2.48 (m, 1H), 2.31 (m, 1H), 1.98 (m, 2H), 1.62 (m, 2H).

HRMS (ESI+): calculated for $C_{26}H_{37}N_{10}O_{10}^+$ $[M+H]^+$: 649.2689, found: 649.2689.

UV λ_{max} = 267 nm.

Uridine and fdC conjugation via linker 32 (54)



2'-*O*-Propargyl-uridine (2.32 mg, 8.22 µmol, 1 eq.) was dissolved in DMSO (200 µL) in a 2 mL *Eppendorf* tube. To the solution, DIPEA (acetonitrile solution, 200 mM, 60 µL, 12.0 µmol, 1.5 eq.), ascorbic sodium (aqueous solution, 500 mM, 40 µL, 20.0 µmol, 2.4 eq.), and $CuSO_4$ (aqueous solution, 100 mM, 100 µL, 10.0 µmol, 1.2 eq.) were added. The mixture was shaken at 25°C for 30 min, before **32** (2.50 mg in 100 µL DMSO, 8.90 µmol, 1.1 eq.) was added. The mixture was shaken for another 2 h at 25°C, and then H_2O (1.0 mL) was added. The tube was vortexed, centrifuged. The water layer was discarded. Acetonitrile (500 µL) was added to the residue, and the solution was divided into five tubes (100 µL each). To each tube, acetic acid (6% aqueous solution, 900 µL) was added. The tube was then shaken at 25°C for 40 min and lyophilized to dryness. In each tube, the dry powder was then redissolved in fdC nucleoside DMSO/H_2O solution (v/v 1/1, 5 mM, 263 µL, 1.31 µmol, 0.8 eq.) with the above 4-methoxyaniline solution (250 mM, 5.2 µL, 1.30 µmol, 0.8 eq.). The mixture was shaken at 25°C for 2 h, and then purified by HPLC: 0% - 70% Buffer B in 45 min, quant.

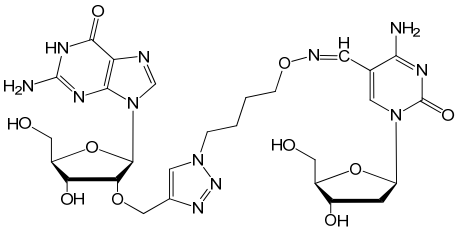
Retention time: 21.3 min.

^{1}H-NMR (800 MHz, D_2O): δ = 8.17 (s, 1H), 8.02 (d, J = 15.8 Hz, 2H), 7.61 (d, J = 8.0 Hz, 1H), 6.25 (d, J = 6.4 Hz, 1H), 5.90 (dd, J = 5.6, 46 Hz, 1H), 5.78 (dd, J = 8.0, 18 Hz, 1H), 4.66 (d, J = 13.2 Hz, 1H), 4.44 (m, 4H), 4.34 (t, J = 5.0 Hz, 1H), 4.20 (t, J = 5.4 Hz, 1H), 4.15 (t, J = 6.2 Hz, 2H), 4.12 (m, 3H), 3.88 (m, 2H), 3.79 (m, 2H), 2.52 (m, 1H), 2.33 (m, 1H), 2.02 (m, 2H), 0.84 (m, 1H), 1.65 (m, 2H),

HRMS (ESI+): calculated for $C_{26}H_{36}N_9O_{10}^+$ $[M+H]^+$: 650.2534, found: 650.2531.

UV λ_{max} = 264 nm.

Guanosine and formylcytosine conjugation via linker 32 (55)



2'-*O*-Propargyl-guanosine (3.28 mg, 10.2 µmol, 1 eq.) was dissolved in DMSO (200 µL) in a 2 mL *Eppendorf* tube. To the mixture, DIPEA (acetonitrile solution, 200 mM, 76 µL, 15.3 µmol, 1.5 eq.), sodium ascorbate aq. (500 mM, 49 µL, 24 µmol, 2.4 eq.), and $CuSO_4$ aq. (100 mM, 122 µL, 12.2 µmol, 1.2 eq.) were added. The mixture was shaken at 25°C for 30 min, before **32** (3.50 mg in 100 µL DMSO, 12.2 µmol, 1.2 eq.) was added. The mixture was shaken for another 6 h at 25°C, and then H_2O (1.0 mL) was added. The tube was vortexed, centrifuged. The water layer was discarded. Acetonitrile (500 µL) was added to the residue, and the suspension was divided into five tubes (100 µL each). To each tube, acetic acid (6% aqueous solution, 900 µL) was added. The tube was then shaken at 25°C for 40 min and lyophilized to dryness. In each tube, the dry powder was redissolved in fdC nucleoside solution (DMSO/H_2O: v/v 1/1, 5 mM, 245 µL, 1.22 µmol, 0.6 eq.) with the above 4-methoxyaniline solution (250 mM, 4.9 µL, 1.22 µmol, 0.6 eq.). The mixture was shaken at 25°C for 2 h, then purified by HPLC: 0% - 70% Buffer B in 45 min.

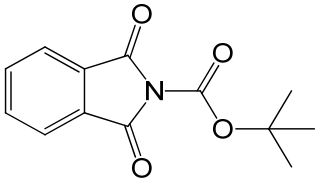
Retention time: 20.2 min.

^{1}H-NMR (800 MHz, D_2O): δ = 8.15 (s, 1H), 7.95 (s, 1H), 7.79 (s, 1H), 7.73 (s, 1H), 6.26 (t, J = 6.4 Hz, 1H), 6.78 (d, J = 7.3 Hz, 1H), 4.66 (m, 1H), 4.56 (m, 2H), 4.45 (m, 1H), 4.30 (m, 3H), 4.10 (m, 3H), 3.84 (m, 5H), 3.21 (m, 1H), 2.52 (m, 1H), 2.34 (m, 1H), 1.92 (m, 2H), 1.57 (m, 2H), 1.29 (m, 2H).

HRMS (ESI+): calculated for $C_{27}H_{37}N_{12}O_{10}^+$ $[M+H]^+$: 689.2756, found: 689.2754.

UV λ_{max} = 257 nm.

tert-Butyl 1,3-dioxoisoindoline-2-carboxylate (57)



To a suspension of phthalimide (**56**) (1.47 g, 10 mmol) in anhydrous acetonitrile (5 mL) at room temperature, di-tert-butyldicarbonate (2.29 g, 10.5 mmol) was added,

followed by 4-dimethylaminopyridine (DMAP) (12 mg, 0.1 mmol). After 30 min at room temperature, a homogeneous solution was obtained. The solvent was removed under reduced pressure. The remaining solid was recrystallized from isohexane / ethyl acetate (9:1) to give **57** as slightly beige plates (2.47 g, 10 mmol, quant.).

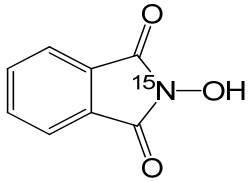
R_f (chloroform / methanol = 10/1) = 0.77.

^{1}H-NMR (400 MHz, $CDCl_3$): δ = 7.94-7.92 (m, 2H, C_{Ar}***H***), 7.81-7.79 (m, 2H, C_{Ar}***H***), 1.62 (s, 9H, C(C***H*****$_3$**)$_3$).

^{13}C-NMR (100 MHz, MeOH-*d4*): δ = 164.1 (2 × C_{Ar}***C***O), 146.7 (N***C***OO), 135.1 (2 ×***C***$_{Ar}$), 131.1 (2 × ***C***$_{Ar}$CO), 124.2 (2 × ***C***$_{Ar}$), 85.3 (***C***$(CH_3)_3$), 27.9 (3 × C(***C***$H_3)_3$).

HRMS (ESI+): calculated for $C_{13}H_{13}NO_4Na^+$ $[M+Na]^+$: 270.0737, found: 270.0741.

^{15}*N*-Hydroxyhothalimide (58)



To a solution of **57** (742 mg, 3 mmol) in acetonitrile (5 mL) at room temperature was added an aqueous solution (4 mL) of hydroxylamine-^{15}*N* hydrochloride (423 mg, 6 mmol, 2eq.) and NaOH (240 mg, 6 mmol, 2 eq.). After stirring the mixture overnight at room temperature, the mixture was concentrated under vacuum. The residue was diluted with water and naturalized with HCl (2 M) until pH 1. The aqueous medium was extracted with ethyl acetate three times. The combined organic layers were dried over $MgSO_4$, filtered and concentrated to give **58** as a white solid (492 mg, 3 mmol, quant.).

R_f (chloroform / methanol = 10:1) = 0.56.

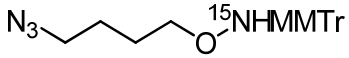
^{1}H-NMR (400 MHz, MeOH-*d4*): δ = 7.82−7.76 (m, 4H, C_{Ar}***H***).

^{13}C-NMR (100 MHz, MeOH-*d4*): δ = 164.5 (2 × ***C***O), 134.1 (2 × ***C***$_{Ar}$), 129.1 (2 × ***C***$_{Ar}$CO), 122.6 (2 × ***C***$_{Ar}$).

^{15}N-NMR (40 MHz, MeOH-*d4*): δ = -181.9.

HRMS (ESI-): calculated for $C_8H_4{}^{15}NO_3^-$ $[M-H]^-$: 163.0197, found: 163.0167.

***O*-(4-Azidobutyl)-^{15}*N*-[(4-methoxyphenyl)diphenylmethyl]hydroxylamine (61)**



61 was synthesized according to the protocol of **47**, except **58** was applied instead of *N*-hydroxy-phthalimide.

R_f (isohexane / ethyl acetate = 10/1 + 3% triethylamine) = 0.64.

^{1}H-NMR (400 MHz, $CDCl_3$): δ = 7.31−7.18 (m, 12H, 12 × C_{Ar}***H***), 6.79 (d, *J* = 8.8 Hz, 2H, 2 × CH_3-O-C-C***H***), 3.76 (s, 3H, O-C***H₃***),3.65 (t, *J* = 6.0 Hz, 2H, C(1)***H₂***), 3.05 (t, *J* = 6.8 Hz, 2H, C(4)***H₂***), 1.52−1.37 (m, 4H, C(2)***H₂***, C(3)***H₂***).

^{13}C-NMR (100 MHz, $CDCl_3$): δ = 158.3 (CH_3-O-***C***$_{Ar}$), 144.6 (2 × ***C***$_{Ar}$), 136.6 (2 × ***C***$_{Ar}$), 130.3 (2 × ***C***$_{Ar}$), 129.2 (4 × ***C***$_{Ar}$), 127.6 (4 × ***C***$_{Ar}$), 126.8 (2 × ***C***$_{Ar}$), 112.9 (2 × ***C***$_{Ar}$), 77.3 (O-NH-***C***), 73.2 (***C***(1)H_2), 55.2 (O-***C***H_3), 51.1 (***C***(4)H_2), 25.7 (***C***(3)H_2), 25.5 (***C***(2)H_2).

^{15}N-NMR (40 MHz, MeOH-*d4*): δ = -211.4.

HRMS (ESI-): calculated for $C_{25}H_{27}{}^{15}NN_3O_4^-$ $[M+HCO_2]^-$: 448.2008, found: 448.2012.

6.5. Part II: Experimental details for fdC probe

Formation of Schiff base between fdC-oligonucleotides and azide linkers

A mixture of **ODN 13** (80 μM, 150 μL, 12 nmol, 1 eq.), azide linker (**L31-36**) in acetonitrile solution (24 mM, 10 μL, 240 nmol, 20 eq.), 4-methoxyaniline solution in ddH_2O/DMSO (v/v 9/1, acidified to pH = 5.5 with acetate acid, 250 mM, 12 μL), NaOAc aq. (pH = 6.0, 400 mM, 30 μL) and ddH_2O (98 μL) was prepared to make a final volume of 300 μL. The reaction was incubated at 37 °C for 24 h followed by HPLC analysis and purification.

Click reaction of fdC-linker oligonucleotides and alkynes contained strands

A mixture of **ODN-13-L31-36** (15 μM, 20 μL, 0.3 nmol, 1 eq.), probe strand aqueous solution (**ODN 14-17**, **16a/b/c**, 15 μM, 20 μL, 0.3 nmol, 1 eq.), NaCl aq. (1 M, 6 μL), NaOAc aq. (pH = 6.0, 100 mM, 6 μL) and ddH_2O (8 μL) was prepared to make a final volume of 60 μL (oligonucleotide working concentration 5 μM). The mixture was heated to 85 °C for 5 min then slowly cooled down to room temperature. CuBr-TBTA (Tris[(1-benzyl-1*H*-1,2,3-triazol-4-yl)methyl]amine) solution in DMSO/*t*-BuOH (v/v 3/1, 1.5 mM, 2 μL, 3 nmol, 10 eq.) was added, and the reaction was conducted at 20 °C. After 2 h, another 1 μL (5 eq.) of the freshly prepared above CuBr-TBTA solution was added. Aliquots (15 μL) were taken at 1, 2, 4 h and quenched by addition of loading buffer. The samples were heated at 85 °C for 3 min followed by denaturing urea PAGE (100 V, ca. 120 min). Percentage of duplex DNA was calculated based on the integrations: dsDNA % = Int.(dsDNA)/[Int.(dsDNA)+ Int.(ssDNA)].

Click reaction on the solid support with azide linkers and deprotection

After solid-phase synthesis (0.2 μmol scale, ca. 50% yield for 13 mer, calculated as 0.1 μmol), the solid support was submerged in dimethyl sulfoxide (80 μL) and acetonitrile (25 μL). To the mixture, $CuSO_4$ aq. (100 mM, 50 μL, 5.0 μmol, 50 eq.), sodium ascorbate aq. (500 mM, 20 μL, 10 μmol, 100 eq.), *N*,*N*-diisopropylethylamine solution in acetonitrile (200 mM, 75 μL, 15 μmol, 150 eq.), azide solution (**L32**) in acetonitrile (100 mM, 50 μL, 5.0 μmol, 50 eq.) were added. The reaction was conducted at 25 °C for 24 h. Afterward, the solid support was washed with dimethyl sulfoxide (2×1 mL), dilute $NaHCO_3$ aq. (2×1 mL), acetonitrile (2×1 mL), ether (1 mL) and air-dried to a

powder. The solid-phase was then cleaved with conc. aqueous NH_3 at 25 °C for 17 h, HPLC purified and lyophilized. Deprotection of MMTr group was carried out by dissolving the oligonucleotide in acetic acid aq. (20%, 200 µL) at 25°C for 30 min, precipitated by addition of sodium acetate solution (3 M, 60 µL) and EtOH (1040 µL), and then purified again with HPLC.

Formation of Schiff base between fdC-oligonucleotides and probe strands

A mixture of fC containing oligonucleotides **ODN 13**, or **ODN 13a-e** in respective control experiments (15 µM, 20 µL, 0.3 nmol, 1 eq.), probe strands **ODN 16e** (15 µM, 20 µL, 0.3 nmol, 1 eq.), NaCl aq. (1 M, 15 µL), NaOAc aq. (pH = 6.0, 100 mM, 15 µL) and ddH_2O (80 µL) was prepared to make a final volume of 150 µL (oligonucleotide working concentration 2 µM). The mixture was heated to 85 °C for 5 min then slowly cooled down to 25 °C for 3 h or 45 °C for 1 h 45 min. A First aliquot (15 µL) was taken and quenched before the 4-methoxyaniline solution (5.4 µL, working concentration 10 mM) was added. The reaction was performed at 25 °C, 300 rpm for 24 h. Aliquots (15 µL) were taken at 1, 2, 4, 6, 8, 12, 24 h and quenched by addition of loading buffer. All the samples were then heated at 85 °C for 3 min followed by PAGE assay as mentioned above.

When using 1,4-benzenediamine as a catalyst, a stock solution of 1,4-benzenediamine (10 mM, 0.5% acetate acid aq.) was prepared.

Melting point experiments

Melting profiles were measured on a *JASCO* V-650 spectrometer using quartz glass cuvettes with 10.00 mm path length. The samples contained NaCl aq. (100 mM), NaOAc aq. (10 mM, pH 6.0) and each strand (1 µM) in a final volume of 200 µL. The measurements were repeated three times with independent samples. Before the measurement, the oligonucleotides were hybridized by slowly cooling down the samples from 85 °C to room temperature. The solutions were covered with silicon oil and tightly plugged. Absorbance was recorded in the forward and reverse direction at temperatures from 25 °C (or 15 °C) to 85 °C with a slope of 1 °C/min. T_M values were calculated as the zero-crossing of 2nd derivate of the hyperchromicity change at 260 nm corrected with 339 nm background.

Probe extension with Klenow Fragment

A mixture of **ODN 23** and **ODN 24/25/26** (60 μM, 10 μL, 0.6 nmol, 1 eq.), NaCl aq. (1 M, 3 μL), NaOAc aq. (pH = 6.0, 100 mM, 3 μL) and ddH_2O (4 μL) was prepared to make a final volume of 30 μL (oligonucleotide working concentration 20 μM). The mixture was heated to 85 °C for 5 min and then slowly cooled down to 25 °C in 3 h. Then 1.2 μL of the mentioned 4-methoxyaniline solution was added to give a catalyst working concentration of 10 mM. The reaction was conducted at 25 °C, 500 rpm for 24 h. Then, to 3 μL of the reaction mixture (60 pmol in theory), *NEB* buffer 2 (10×, 3 μL), dNTP (2 mM, 3 μL, 6 nmol), ddH_2O (18 μL), and Klenow Fragment (3'→5' exo-) (5000 U/mL, 1.2 μL, 6 U) were added. The mixture was incubated at 37 °C for 1 h. The reaction was quenched by addition of PAGE loading buffer. The samples were heated at 85 °C for 3 min followed by denaturing urea polyacrylamide gel electrophoresis (150 V, 3 h).

Two probes react with ODN 23 in turn

ODN 23 and the first probe **ODN 24/25/26** (20 μM, 6 μL), NaCl aq. (1 M, 6 μL), NaOAc aq. (pH = 6.0, 100 mM, 6 μL) and ddH_2O (42 μL) was prepared to make a final volume of 60 μL (oligonucleotide working concentration 2 μM). The mixture was heated to 85 °C for 5 min then slowly cooled down to 25 °C in 3 h. 2.4 μL of mentioned 4-methoxyaniline solution was added to give a catalyst working concentration of 10 mM. The reaction was conducted at 25 °C for 24 h. An aliquot (20 μL) was then taken out and the second probe (20 μM, 2 μL) was added. The mixture was further incubated for 24 h. The remaining mixture was quenched with loading buffer. All the samples were heated at 85 °C for 3 min followed by denaturing urea polyacrylamide gel electrophoresis (150 V, 3 h).

6.6. Part II: Experimental details for fdC detection

LC-MS study with crosslinked duplex or gDNA

Duplex crosslinking between fdC-oligonucleotides and probe strands is described in Chapter 7.5. Biotin enrichment using Streptavidin Magnetic Particles from *Roche* was conducted according to the manufacturer's instructions. The reaction mixture (variable amount of dsDNA, 35 µL), with or without magnetic particles, was mixed with $ZnSO_4$ aq. (4 mM, 0.9 µL), S1 nuclease (42 U/µL, 1.0 µL), Antarctica Phosphatase (5 U/µL, 1.0 µL), and ddH_2O (4.6 µL) and incubated at 37°C for 3 h. EDTA-Na_2 aq. (1 mM, 3.9 µL), snake venom phosphatase (0.1 U/µL, 1.0 µL), and ddH_2O (2.6 µL) were then added to the mixture and incubated for additional 3 h. The mixture was then filtered and submitted to LC-MS analysis.

Primer extension study with crosslinked duplex

Duplex crosslinking between fdC-oligonucleotides and probe strands is described in Chapter 7.5. Primer extension study is stated in Chapter 7.3.

Synthesis of the target strand ODN 31 and crosslinked duplex

ODN 31a containing one fdC was synthesized on solid-phase. **ODN 31b** and **31c** were purchased from *Metabion*. **ODN 31a** (5 nmol), T4 DNA ligase buffer (10×, 5 µL), T4 polynucleotide kinase (10 U/µL, 4 µL, 40 U) and ddH_2O were mixed to a final volume of 50 µL and incubated at 37°C for 4 h. To the mixture, **ODN 31b** (5 nmol), **ODN 31c** (5.5 nmol), and ddH_2O were added to a final volume of 96 µL. The reaction mixture was denatured at 85°C for 3 min and cooled down to room temperature in 2 h, and then T4 DNA ligase (400 U/µL, 4 µL, 1600 U) was added and further incubated at 16°C for 12 h. Afterwards, the mixture (100 µL) was mixed with loading buffer (100 µL, 90% formamide, 10% 0.1M EDTA, pH = 8.0) and purified with PAGE at 40°C, using power of 20 W for 2 h. The gel containing **ODN 31** was then cut out, minced and extracted with TBE buffer (1×, 1 mL) at room temperature overnight, desalted with *Waters Sep-Pak* C18 cartridge and lyophilized to dryness. Duplex crosslinking between **ODN 31** and probe strands is described in Chapter 7.5. Salts and catalyst were removed by *NEB Monarch* PCR DNA Cleanup Kit.

Genomic fdC detection study

Isolation Cell lysates, kindly obtained from co-worker Angie Kirchner, were homogenized with Tissue Lyzer for 1 min at 30 Hz and centrifuged for 5 min at 21000 xg. The lysates were then isolated with Zymo-Spin V column according to the manufacturer's instructions. The concentration was measured using the *Nanodrop*, obtained 200-600 ng/µL.

Probe crosslinking The gDNA solution obtained above (1.2 µg), fdC probe (1 µM, 2 µL), NaH_2PO_4-Na_2HPO_4 buffer (200 mM, pH = 6.0, 2 µL), NaCl aq. (1.5 M, 2µL), and ddH_2O were mixed to a final volume of 18 µL. The mixture was heated to 95 °C for 3 min, and then cooled down rapidly to 25 °C. 1,4-Benzenediamine aq. (10 mM, 2 µL) was added and the reaction vial shaken (300 rpm) for 6 h at 25 °C. The mixture was neutralized with Na_2HPO_4 aq. (200 mM, 40 µL) before purified with *NEB Monarch* PCR DNA Cleanup Kit using the binding buffer (120 µL), eluted with elution buffer (30 µL). The eluted solution was quantified with Nanodrop, obtained 22-32 ng/µL. UV spectra confirmed the main peak centered at 260 nm.

Ligation The above gDNA solution (300 ng), arch strands (20 nM, 1 µL), Ampligase reaction buffer (10×, 2 µL), Ampligase from *Epicenter* (5 U/µL, 2 µL, 10 U) and ddH_2O were mixed to a final volume of 20 µL. The mixture was heated to 95 °C for 3 min, and then 94 °C for 1 min, 60 °C for 1 h and back to 94 °C for 10 cycles. Then, the reaction mixture was diluted with Tris-HCl buffer (200 mM, pH = 7.6, 50 µL) before purified with *NEB Monarch* PCR DNA Cleanup Kit using binding buffer (140 µL), eluted with elution buffer (10 µL). The eluted solution was quantified with Nanodrop, obtained 20-32 ng/µL. UV spectra confirmed the main peak centered at 260 nm.

Real-time PCR Real-time PCR was conducted on an *Eppendorf* realplex. For a typical duplicate reaction, primers (18 µM each, 1 µL), *iTaq* Universal SYBR Green Supermix (2×, 10 µL) were used to a final volume of 20 µL. PCR cycle: 95°C 2 min, 95°C 15 sec and 60°C 1 min for 40 cycles. Threshold cycle numbers were calculated using the

CalQplex method.

Droplet digital PCR ddPCR was conducted on a *Bio-Rad* QX100 ddPCR System. For one reaction, gDNA (6 ng), four primers (18 μM each, 1 μL), two TaqMan probes (5 μM each, 1 μL), digital PCR Supermix for Probes (no dUTP, 2×, 10 μL), and ddH_2O were mixed to a final volume of 20 μL with primer working concentration of 900 nM and TaqMan probe working concentration of 250 nM.

PCR cycle was conducted on a *Bio-Rad* T100 Thermal cycler. PCR cycle: 95°C for 10 min, 94°C for 30 sec and specific annealing temperature (64°C) for 1 min for 35 or 40 cycles, then 98°C for 10 min and cooled down to 12°C. A temperature ramp of 2°C/s was used. Droplet generation and counting were conducted according to the manufacturer's instructions, i.e. reaction mixture prepared as above (20 μL) and ddPCR droplet generate oil (70 μL) were used for one reaction. the Accounted droplet number was retained in 10000-18000. FAM for detection amplicon was set to channel 1; HEX for reference amplicon was set to channel 2.

Appendix

Part I: Crystallographic data

5-[3-Hydroxy-4-(1*H*-pyrazol-1-yl)phenyl]-2-(hydroxymethyl)tetrahydrofuran-3-ol (16)

Net formula	$C_{14}H_{16}N_2O_4$
Mr /g mol^{-1}	276.288
Crystal size /mm	0.195 × 0.127 × 0.051
T /K	173(2)
Radiation	'Mo Kα
Diffractometer	'Bruker D8Venture'
Crystal system	Monoclinic
Space group	P21
a /Å	11.8630(13)
b /Å	4.8437(7)
c /Å	12.6119(16)
α /°	90
β /°	116.802(3)
γ /°	90
V /Å^3	646.84(14)
Z	2
Calc. density /g cm^{-3}	1.4186(3)
μ /mm^{-1}	0.105
Absorption correction	multi-scan
Transmission factor range	0.9043–0.9580
Refls. measured	8079
Rint	0.0407
Mean σ(I) /I	0.039
θ range	3.18–25.41
Refls. observed	2117
x, y (weighting scheme)	0.0376, 0.1254
Hydrogen refinement	Mixed
Flack parameter	1.7(11)
Refls. in refinement	2358
Parameters	193
Restraints	1
R(Fobs)	0.0327
Rw(F2)	0.0809
S	1.059

shift/errormax	0.001
Max electron density /e Å^{-3}	0.154
Min electron density /e Å^{-3}	−0.168

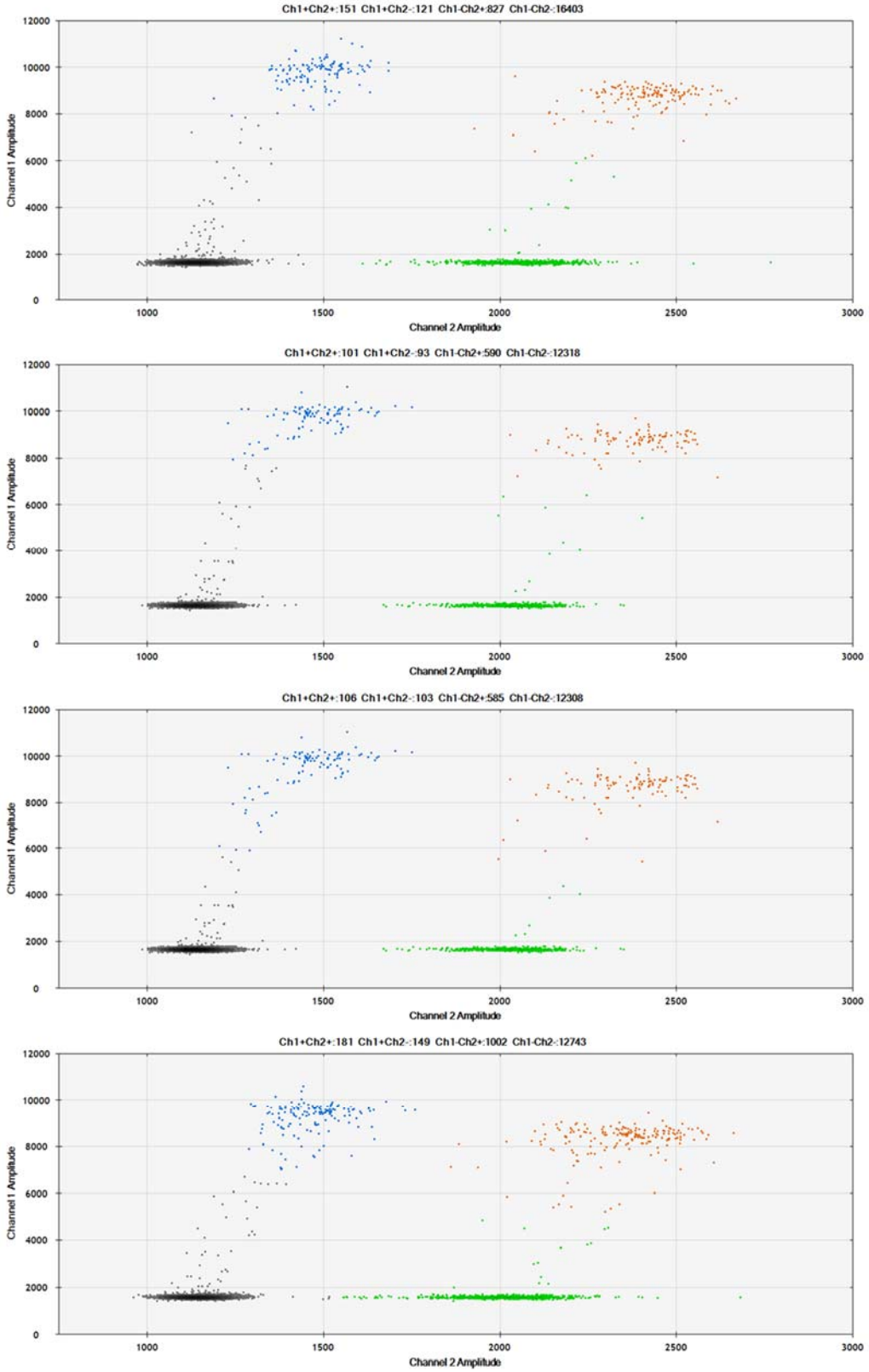
C-bound H: constr, O-bound H: refall.

Part II: Selected ddPCR results

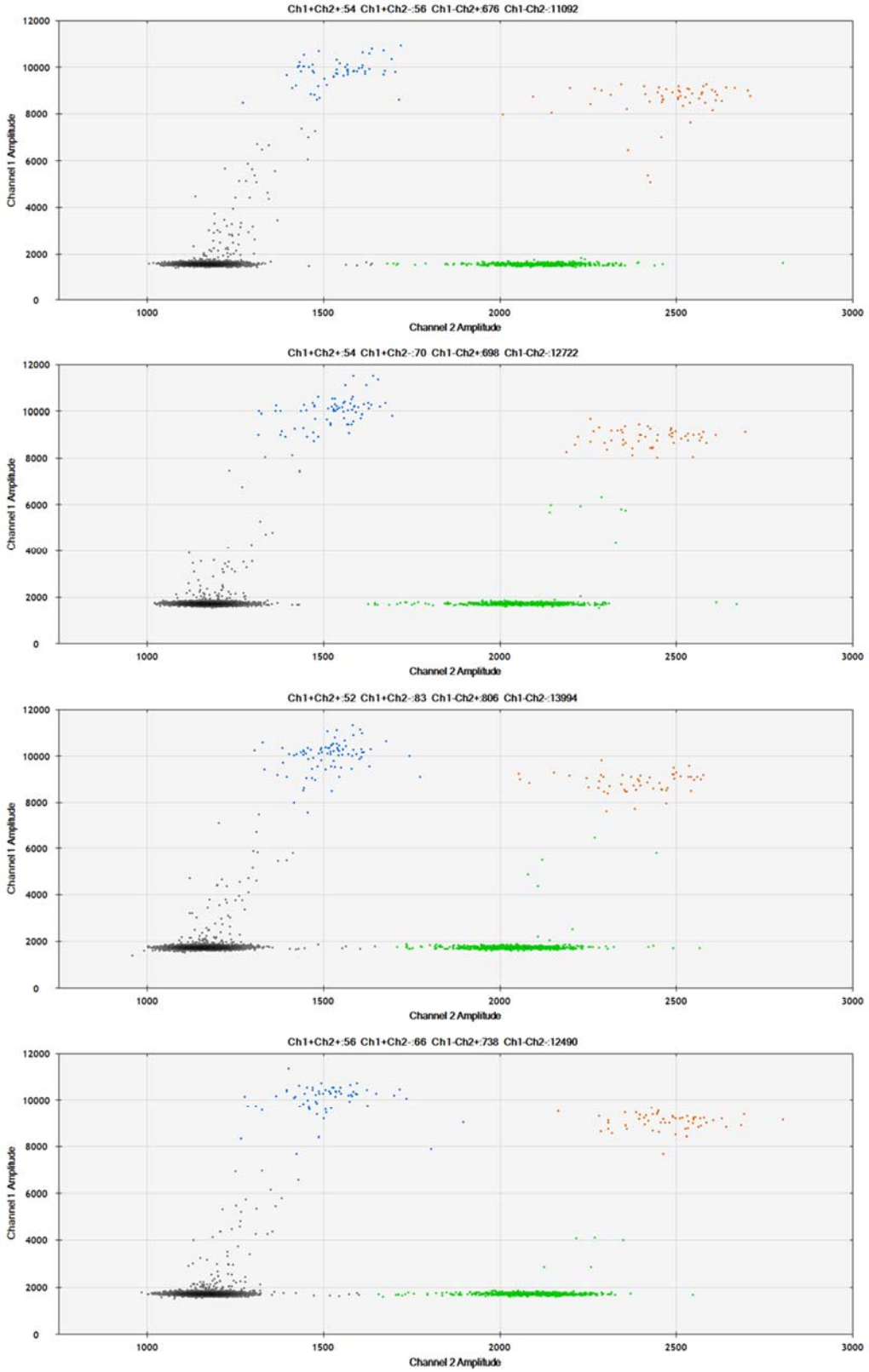
Raw data of fdC detection of position 1 in three cells types.

	Ch1	Ch2	1+2+	1+2-	1-2+	1-2-	AD	η	Average
Tdg-/-	17.6	67.6	151	121	827	16399	17502	27.8%	28.5%
Tdg-/-	17.6	63.7	101	93	590	12318	14896	28.1%	
Tdg-/-	18.9	63.7	106	103	585	12308	13102	30.2%	
Tdg-/-	28.8	103	181	149	1002	12743	14075	27.9%	
Tdg+/-	11.6	74.0	54	56	676	11085	11878	15.1%	15.7%
Tdg+/-	11.3	67.2	54	70	698	12722	13544	16.5%	
Tdg+/-	10.7	69.6	52	83	806	13994	14935	15.7%	
Tdg+/-	10.8	72.0	56	66	738	12490	13350	15.4%	
Dnmt TKO	4.0	115.0	10	29	1049	10327	11415	3.7%	5.2%
Dnmt TKO	3.1	61.1	11	17	625	11913	12566	4.4%	
Dnmt TKO	4.1	58.0	14	30	620	12508	13174	6.9%	
Dnmt TKO	2.8	60.6	8	21	600	11484	12119	4.8%	
Dnmt TKO	5.7	96.0	35	28	986	11998	13047	6.2%	

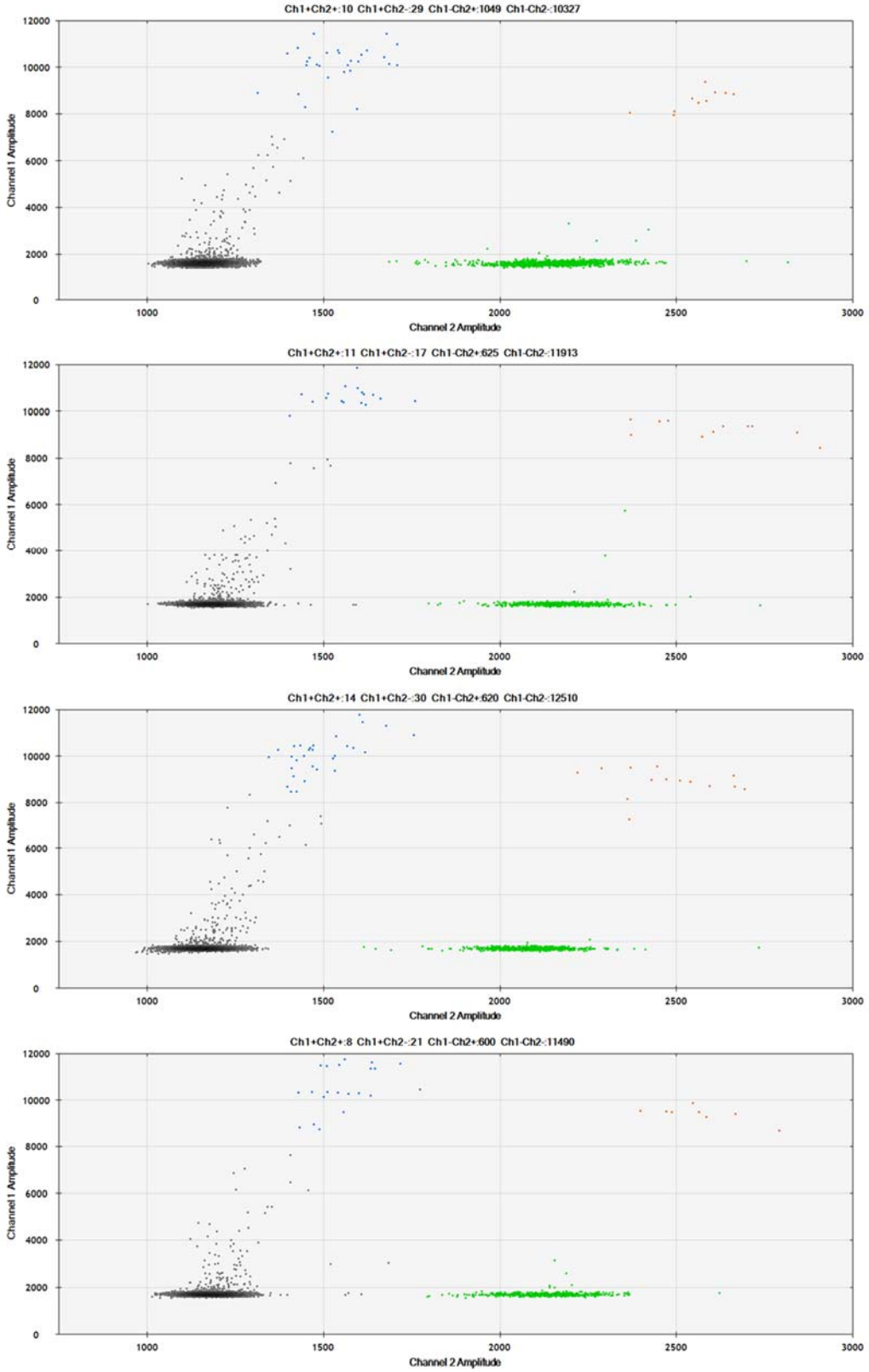
Tdg $^{-/-}$ ES cell sample



$Tdg^{+/-}$ ES cell sample.



Dnmt TKO ES cell sample.



Raw data of fdC detection of position 1 in wild -type cell along days during priming.

	Ch1	Ch2	1+2+	1+2-	1-2+	1-2-	AD	η	Average
WT0	5.4	108	20	34	1017	10787	11858	5.21%	6.06%
WT0	6.8	110	30	58	1336	13841	15265	6.44%	
WT0	6.6	112	26	53	1255	12738	14072	6.17%	
WT0	7.8	108	23	58	1241	13060	14382	6.41%	
WT2	9.7	89.2	40	98	1193	15552	16883	11.19%	10.90%
WT2	10.2	88.7	35	105	1249	15948	17337	10.90%	
WT2	10.8	93.0	35	96	1168	14476	15771	10.89%	
WT2	10.2	96.0	54	94	1279	15637	17064	11.10%	
WT2	6.3	56.2	19	65	744	15582	16410	11.01%	
WT2	6.0	57.3	27	54	759	15780	16620	10.31%	
WT4	6.2	68.0	19	30	501	8702	9252	9.42%	8.76%
WT4	5.4	68.4	26	46	811	13942	14821	8.60%	
WT4	4.5	59.0	14	24	477	9449	9964	7.74%	
WT4	5.8	64.5	23	43	688	12579	13333	9.28%	
WT6	11.1	117	42	67	1081	10691	11881	9.71%	8.67%
WT6	8.2	114	28	41	888	9000	9957	7.53%	
WT6	9.8	120	32	41	820	7909	8802	8.57%	
WT6	8.5	114	43	57	1086	11010	12196	8.86%	

Raw data of fdC detection of in $Tdg^{-/-}$, $Tdg^{+/-}$cells for locus 2.

	Ch1	Ch2	1+2+	1+2-	1-2+	1-2-	AD	η	Average
Tdg-/-	10.5	52	55	92	650	15671	16468	20.9%	19.8%
Tdg-/-	20.8	115	88	133	1084	11285	12590	18.9%	
Tdg-/-	22.5	114	98	173	1226	12805	14302	20.5%	
Tdg-/-	20.1	109	94	149	1178	12927	14348	19.1%	
Tdg-/-	21.8	116	90	159	1179	12141	13569	19.6%	
Tdg+/-	5.5	66	28	49	884	15696	16657	8.4%	9.2%
Tdg+/-	5.9	59	13	29	400	8002	8444	10.2%	
Tdg+/-	8.3	96	43	78	1299	15749	17169	9.0%	
Tdg+/-	8.1	93	39	73	1205	15070	16387	9.0%	

Abbreviations

cadC	5-carboxycytosine
fdC	5-formylcytosine
hmdC	5-hydroxymethylcytosine
mdC	5-methylcytosine
AID	activation-induced cytidine deaminase
APS	ammonium persulfate
aq.	aqueous
ARP	aldehyde reactive probe
BAC	bacterial artificial chromosome
BER	base excision repair
β-GT	β-glucosyltransferase
BS-Seq	bisulfite sequencing
CD	circular dichroism
CGIs	CpG islands
CHES	2-(cyclohexylamino)ethanesulfonic acid
ChIP	chromatin immunoprecipitation
DME	DEMETER
DMF	dimethylformamide
DMTr	4,4'-dimethoxytrityl
DNA	deoxyribonucleic acid
DNMT	DNA methyltransferase
dNTP	deoxyribonucleoside triphosphate
DRM	domains rearranged methyltransferase
EDTA	ethylenediaminetetraacetic acid
ELISA	enzyme-linked immunosorbent assay
EPR	electron paramagnetic resonance
ES cells	embryonic stem cells
ESI	electrospray ionisation mass spectrometry
FAM	fluorsescein
FISH	fluorescence in situ hybridization
FNA	flexible nucleic acids
gDNA	genomic DNA
HDAC	histone deacetylases
HGP	Human Genome Project
HNA	1,5-anhydrohexitol nucleic acids
HRMS	high resolution mass spectrometry
LV-PS	low volume polystyrene
MALDI-TOF	matrix assisted laser desorption/ionization - time of flight
MeDIP	methylated DNA immunoprecipitation

MES	4-morpholinoethane-sulfonic acid
MMTr	4-methoxytrityl
NER	nucleotides excision repair
NMR	nuclear magnetic resonance
NOESY	nuclear Overhauser effect spectroscopy
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PhthN-OH	*N*-hydroxy-phthalimide
RNA	ribonucleic acid
ROS	repressor of silencing
RP-HPLC	reverse phase high performance liquid chromatography
RRBS	reduced representation bisulfite sequencing
SAH	*S*-adenosylhomocysteine;
SAM	*S*-adenosyl-methionine
SMRT	single molecular real-time
SMUG	strand-selective monofunctional uracil-DNA glycosylase
TDG	thymine-DNA glycosylase
TEMED	tetramethylethylenediamine
TET	ten-eleven translocation
TIPS	triisopropylsilyl
TLC	thin layer chromatography
TNA	threose nucleic acids
Tol	toluene
Tris	2-amino-2-hydroxymethyl-propane-1,3-diol
UDP	uridine 5′-diphosphate
UV	ultraviolate

Reference

1. Dunn, D. B.; Smith, J. D., The Occurrence of 6-Methylaminopurine in Deoxyribonucleic Acids. *Biochem. J.*, **1958**, *68*, 627-636.
2. Wion, D.; Casadesus, J., N^6-Methyl-adenine: An Epigenetic Signal for DNA-protein Interactions. *Nat. Rev. Micro.*, **2006**, *4*, 183-192.
3. Ehrlich, M.; Gama-Sosa, M. A.; Carreira, L. H.; Ljungdahl, L. G.; Kuo, K. C.; Gehrke, C. W., DNA Methylation in Thermophilic Bacteria: N^4-Methylcytosine, 5-Methylcytosine, and N^6-Methyladenine. *Nucleic Acids Res.*, **1985**, *13*, 1399-1412.
4. Watson, J. D.; Crick, F. H. C., Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature*, **1953**, *171*, 737-738.
5. Watson, J. D.; Crick, F. H. C., Genetical Implications of the Structure of Deoxyribonucleic Acid *Nature*, **1953**, *171*, 964-967.
6. Herdewijn, P.; Marlière, P., Toward Safe Genetically Modified Organisms through the Chemical Diversification of Nucleic Acids. *Chem. Biodivers.*, **2009**, *6*, 791-808.
7. Merrifield, R. B., Solid Phase Peptide Synthesis. I. The Synthesis of a Tetrapeptide. *J. Am. Chem. Soc.*, **1963**, *85*, 2149-2154.
8. Letsinger, R. L.; Kornet, M. J., Popcorn Polymer as a Support in Multistep Syntheses. *J. Am. Chem. Soc.*, **1963**, *85*, 3045-3046.
9. Letsinger, R. L.; Kornet, M. J.; Mahadevan, V.; Jerina, D. M., Reactions on Polymer Supports. *J. Am. Chem. Soc.*, **1964**, *86*, 5163-5165.
10. Letsinger, R. L.; Mahadevan, V., Oligonucleotide Synthesis on a Polymer Support *J. Am. Chem. Soc.*, **1965**, *87*, 3526-3527.
11. Letsinger, R. L.; Ogilvie, K. K., Nucleotide Chemistry. XIII. Synthesis of Oligothymidylates via Phosphotriester Intermediates. *J. Am. Chem. Soc.*, **1969**, *91*, 3350-3355.
12. Letsinger, R. L.; Miller, P. S., Nucleotide Chemistry. XIV. Protecting Groups for Nucleosides Used in Synthesizing Oligonucleotides. *J. Am. Chem. Soc.*, **1969**, *91*, 3356-3359.
13. Letsinger, R. L.; Ogilvie, K. K.; Miller, P. S., Nucleotide Chemistry. XV. Developments in Syntheses of Oligodeoxyribonucleotides and their Organic Derivatives. *J. Am. Chem. Soc.*, **1969**, *91*, 3360-3365.
14. Letsinger, R. L.; Finnan, J. L.; Heavner, G. A.; Lunsford, W. B., Nucleotide Chemistry. XX. Phosphite Coupling Procedure for Generating Internucleotide Links. *J. Am. Chem. Soc.*, **1975**, *97*, 3278-3279.
15. Letsinger, R. L.; Lunsford, W. B., Synthesis of Thymidine Oligonucleotides by Phosphite Triester Intermediates. *J. Am. Chem. Soc.*, **1976**, *98*, 3655-3661.
16. Matteucci, M. D.; Caruthers, M. H., Synthesis of Deoxyoligonucleotides on a Polymer Support. *J. Am. Chem. Soc.*, **1981**, *103*, 3185-3191.
17. McBride, L. J.; Caruthers, M. H., An Investigation of Several Deoxynucleoside Phosphoramidites Useful for Synthesizing Deoxyoligonucleotides. *Tetrahedron Lett.*,

1983, *24*, 245-248.

18. Eckstein, F., A Dinucleoside Phosphorothioate. *Tetrahedron Lett.*, **1967**, *8*, 1157-1160.
19. Wilson, C.; Keefe, A. D., Building Oligonucleotide Therapeutics using Non-natural Chemistries. *Curr. Opin. Chem. Biol.* , **2006**, *10*, 607-614.
20. Huang, Z.; Schneider, K. C.; Benner, S. A., Building Blocks for Oligonucleotide Analogs with Dimethylene Sulfide, Sulfoxide, and Sulfone Groups Replacing Phosphodiester Linkages. *J. Org. Chem.*, **1991**, *56*, 3869-3882.
21. Mori, K.; Boiziau, C.; Cazenave, C.; Matsukura, M.; Subasinghe, C.; Cohen, J. S.; Broder, S.; Toulmé, J. J.; Stein, C. A., Phosphoroselenoate Oligodeoxynucleotides: Synthesis, Physico-chemical Characterization, Anti-sense Inhibitory Properties and Anti-HIV Activity. *Nucleic Acids Res.*, **1989**, *17*, 8207-8219.
22. Vyazovkina, E. V.; Rife, J. P.; Lebedev, A. V.; Wickstrom, E., Preparation of Trimers and Tetramers of Mixed Sequence Oligodeoxynucleoside Methylphosphonates and Assignment of Configurations at the Chiral Phosphorus. *Nucleic Acids Res.*, **1993**, *21*, 5957-5963.
23. Koole, L. H.; Moody, H. M.; Broeders, N. L. H. L.; Quaedflieg, P. J. L. M.; Kuijpers, W. H. A.; Van Genderen, M. H. P.; Coenen, A. J. J. M.; Van der Wal, S.; Buck, H. M., Synthesis of Phosphate-methylated DNA Fragments using 9-Fluorenyl-methoxycarbonyl as Transient Base Protecting Group. *J. Org. Chem.*, **1989**, *54*, 1657-1664.
24. Kutateladze, T.; Beabealashvili, R.; Azhayev, A.; Krayevsky, A., 3′-Deoxy-3′-aminonucleoside 5′-Triphosphates-Terminators of RNA Synthesis, Catalyzed by DNA-dependent RNA Polymerase from Escherichia coli. *FEBS Lett.*, **1983**, *153*, 420-426.
25. Pannecouque, C.; Wigerinck, P.; Van Aerschot, A.; Herdewijn, P., Dimeric Building Blocks with *N*-cyanoguanidine Linkage for Oligonucleotide Synthesis. *Tetrahedron Lett.*, **1992**, *33*, 7609-7612.
26. Sood, A.; Shaw, B. R.; Spielvogel, B. F., Boron-containing Nucleic Acids. 2. Synthesis of Oligodeoxynucleoside Boranophosphates. *J. Am. Chem. Soc.*, **1990**, *112*, 9000-9001.
27. Li, P.; Sergueeva, Z. A.; Dobrikov, M.; Shaw, B. R., Nucleoside and Oligonucleoside Boranophosphates: Chemistry and Properties. *Chem. Rev.*, **2007**, *107*, 4746-4796.
28. Roy, S.; Olesiak, M.; Padar, P.; McCuen, H.; Caruthers, M. H., Reduction of Metal Ions by Boranephosphonate DNA. *Org. Bio. Chem.*, **2012**, *10*, 9130-9133.
29. Roy, S.; Olesiak, M.; Shang, S.; Caruthers, M. H., Silver Nanoassemblies Constructed from Boranephosphonate DNA. *J. Am. Chem. Soc.*, **2013**, *135*, 6234-6241.
30. Kawasaki, A. M.; Casper, M. D.; Freier, S. M.; Lesnik, E. A.; Zounes, M. C.; Cummins, L. L.; Gonzalez, C.; Cook, P. D., Uniformly Modified 2'-Deoxy-2'-fluoro-phosphorothioate Oligonucleotides as Nuclease-resistant Antisense Compounds with High Affinity and Specificity for RNA Targets. *J. Med. Chem.*, **1993**, *36*, 831-841.
31. Allerson, C. R.; Sioufi, N.; Jarres, R.; Prakash, T. P.; Naik, N.; Berdeja, A.; Wanders, L.; Griffey, R. H.; Swayze, E. E.; Bhat, B., Fully 2'-Modified Oligonucleotide Duplexes with Improved in Vitro Potency and Stability Compared to Unmodified Small

Interfering RNA. *J. Med. Chem.*, **2005**, *48*, 901-904.

32. Prakash, T. P.; Allerson, C. R.; Dande, P.; Vickers, T. A.; Sioufi, N.; Jarres, R.; Baker, B. F.; Swayze, E. E.; Griffey, R. H.; Bhat, B., Positional Effect of Chemical Modifications on Short Interference RNA Activity in Mammalian Cells. *J. Med. Chem.*, **2005**, *48*, 4247-4253.
33. Hendel, A.; Bak, R. O.; Clark, J. T.; Kennedy, A. B.; Ryan, D. E.; Roy, S.; Steinfeld, I.; Lunstad, B. D.; Kaiser, R. J.; Wilkens, A. B.; Bacchetta, R.; Tsalenko, A.; Dellinger, D.; Bruhn, L.; Porteus, M. H., Chemically Modified Guide RNAs Enhance CRISPR-Cas Genome Editing in Human Primary Cells. *Nat. Biotech.*, **2015**, *33*, 985-989.
34. Rahdar, M.; McMahon, M. A.; Prakash, T. P.; Swayze, E. E.; Bennett, C. F.; Cleveland, D. W., Synthetic CRISPR RNA-Cas9-guided Genome Editing in Human Cells. *Proc. Natl. Acad. Sci. USA*, **2015**, *112*, E7110-E7117.
35. Fauster, K.; Kreutz, C.; Micura, R., 2′-SCF_3 Uridine: A Powerful Label for Probing Structure and Function of RNA by ^{19}F NMR Spectroscopy. *Angew. Chem., Int. Ed.*, **2012**, *51*, 13080-13084.
36. Košutić, M.; Jud, L.; Da Veiga, C.; Frener, M.; Fauster, K.; Kreutz, C.; Ennifar, E.; Micura, R., Surprising Base Pairing and Structural Properties of 2′-Trifluoro-methylthio-Modified Ribonucleic Acids. *J. Am. Chem. Soc.*, **2014**, *136*, 6656-6663.
37. Höbartner, C.; Micura, R., Chemical Synthesis of Selenium-Modified Oligoribonucleotides and Their Enzymatic Ligation Leading to an U6 SnRNA Stem-loop Segment. *J. Am. Chem. Soc.*, **2004**, *126*, 1141-1149.
38. Aigner, M.; Hartl, M.; Fauster, K.; Steger, J.; Bister, K.; Micura, R., Chemical Synthesis of Site-specifically 2′-Azido-modified RNA and Potential Applications for Bioconjugation and RNA Interference. *ChemBioChem*, **2011**, *12*, 47-51.
39. Fauster, K.; Hartl, M.; Santner, T.; Aigner, M.; Kreutz, C.; Bister, K.; Ennifar, E.; Micura, R., 2'-Azido RNA, a Versatile Tool for Chemical Biology: Synthesis, X-ray Structure, siRNA Applications, Click Labeling. *ACS Chem. Bio.*, **2012**, *7*, 581-589.
40. Büttner, L.; Javadi-Zarnaghi, F.; Höbartner, C., Site-specific Labeling of RNA at Internal Ribose Hydroxyl Groups: Terbium-assisted Deoxyribozymes at Work. *J. Am. Chem. Soc.*, **2014**, *136*, 8131-8137.
41. Shealy, Y. F.; O'Dell, C. A., The Carbocyclic Analog of Cytidine, Synthesis and Antineoplastic Activity. *J. Heterocyclic Chem.*, **1976**, *13*, 1353-1354.
42. Varaprasad, C. V.; Averett, D.; Ramasamy, K. S.; Wu, J., Synthesis and Structural Studies of Monocyclic 4′-Aza-L-nucleosides. *Tetrahedron*, **1999**, *55*, 13345-13368.
43. Hancox, E. L.; Connolly, B. A.; Walker, R. T., Synthesis and Properties of Oligodeoxynucleotides Containing the Analogue 2′-Deoxy-4′-thiothemidine. *Nucleic Acids Res.*, **1993**, *21*, 3485-3491.
44. Hoshika, S.; Minakawa, N.; Matsuda, A., Synthesis and Physical and Physiological Properties of 4′-thioRNA: Application to Post-modification of RNA Aptamer toward NF-κB. *Nucleic Acids Res.*, **2004**, *32*, 3815-3825.
45. Jeong, L. S.; Tosh, D. K.; Kim, H. O.; Wang, T.; Hou, X.; Yun, H. S.; Kwon, Y.; Lee, S. K.; Choi, J.; Zhao, L. X., First Synthesis of 4'-Selenonucleosides Showing Unusual

Southern Conformation. *Org. Lett.*, **2007**, *10*, 209-212.

46. Watts, J. K.; Johnston, B. D.; Jayakanthan, K.; Wahba, A. S.; Pinto, B. M.; Damha, M. J., Synthesis and Biophysical Characterization of Oligonucleotides Containing a 4'-Selenonucleotide. *J. Am. Chem. Soc.*, **2008**, *130*, 8578-8579.
47. Renneberg, D.; Leumann, C. J., Watson-Crick Base-pairing Properties of Tricyclo-DNA. *J. Am. Chem. Soc.*, **2002**, *124*, 5993-6002.
48. Dupouy, C.; Iché-Tarrat, N.; Durrieu, M.-P.; Rodriguez, F.; Escudier, J.-M.; Vigroux, A., Watson-Crick Base-Pairing Properties of Nucleic Acid Analogues with Stereocontrolled α and β Torsion Angles (α,β-D-CNAs). *Angew. Chem., Int. Ed.*, **2006**, *45*, 3623-3627.
49. Wengel, J., Synthesis of 3'-*C*- and 4'-*C*-Branched Oligodeoxynucleotides and the Development of Locked Nucleic Acid (LNA). *Acc. Chem. Res.*, **1998**, *32*, 301-310.
50. Imanishi, T.; Obika, S., BNAs: Novel Nucleic Acid Analogs with a Bridged Sugar Moiety. *Chem. Commun.*, **2002**, *38*, 1653-1659.
51. Seth, P. P.; Pallan, P. S.; Swayze, E. E.; Egli, M., Synthesis, Duplex Stabilization and Structural Properties of a Fluorinated Carbocyclic LNA Analogue. *ChemBioChem*, **2013**, *14*, 58-62.
52. Morihiro, K.; Kodama, T.; Kentefu; Moai, Y.; Veedu, R. N.; Obika, S., Selenomethylene Locked Nucleic Acid Enables Reversible Hybridization in Response to Redox Changes. *Angew. Chem., Int. Ed.*, **2013**, *52*, 5074-5078.
53. Kaur, H.; Babu, B. R.; Maiti, S., Perspectives on Chemistry and Therapeutic Applications of Locked Nucleic Acid (LNA). *Chem. Rev.*, **2007**, *107*, 4672-4697.
54. Obika, S.; Uneda, T.; Sugimoto, T.; Nanbu, D.; Minami, T.; Doi, T.; Imanishi, T., 2'-*O*,4'-*C*-Methylene Bridged Nucleic Acid (2',4'-BNA): Synthesis and Triplex-forming Properties. *Bioorg. Med. Chem.*, **2001**, *9*, 1001-1011.
55. Koizumi, M.; Morita, K.; Daigo, M.; Tsutsumi, S.; Abe, K.; Obika, S.; Imanishi, T., Triplex Formation with 2'-*O*,4'-*C*-ethylene-bridged Nucleic Acids (ENA) having C3′-endo Conformation at Physiological pH. *Nucleic Acids Res.*, **2003**, *31*, 3267-3273.
56. K. Singh, S.; A. Koshkin, A.; Wengel, J.; Nielsen, P., LNA (locked nucleic acids): Synthesis and High-affinity Nucleic Acid Recognition. *Chem. Commun.*, **1998**, 455-456.
57. Brunet, E.; Alberti, P.; Perrouault, L.; Babu, R.; Wengel, J.; Giovannangeli, C., Exploring Cellular Activity of Locked Nucleic Acid-modified Triplex-forming Oligonucleotides and Defining its Molecular Basis. *J. Biol. Chem.*, **2005**, *280*, 20076-20085.
58. Hanessian, S.; Schroeder, B. R.; Giacometti, R. D.; Merner, B. L.; Østergaard, M.; Swayze, E. E.; Seth, P. P., Structure-based Design of a Highly Constrained Nucleic Acid Analogue: Improved Duplex Stabilization by Restricting Sugar Pucker and Torsion Angle γ. *Angew. Chem., Int. Ed.*, **2012**, *51*, 11242-11245.
59. Schneider, K. C.; Benner, S. A., Oligonucleotides Containing Flexible Nucleoside Analogs. *J. Am. Chem. Soc.*, **1990**, *112*, 453-455.
60. Wang, R.; Wang, C.; Cao, Y.; Zhu, Z.; Yang, C.; Chen, J.; Qing, F.-L.; Tan, W.,

Trifluoromethylated Nucleic Acid Analogues Capable of Self-assembly through Hydrophobic Interactions. *Chem. Sci.*, **2014**, *5*, 4076-4081.

61. Nielsen, P.; Dreiøe, L. H.; Wengel, J., Synthesis and Evaluation of Oligodeoxynucleotides Containing Acyclic Nucleosides: Introduction of Three Novel Analogues and a Summary. *Bioorg. Med. Chem.*, **1995**, *3*, 19-28.
62. Zhang, L.; Peritz, A.; Meggers, E., A Simple Glycol Nucleic Acid. *J. Am. Chem. Soc.*, **2005**, *127*, 4174-4175.
63. Karri, P.; Punna, V.; Kim, K.; Krishnamurthy, R., Base-pairing Properties of a Structural Isomer of Glycerol Nucleic Acid. *Angew. Chem., Int. Ed.*, **2013**, *52*, 5840-5844.
64. Li, P.; Sun, J.; Su, M.; Yang, X.; Tang, X., Design, Synthesis and Properties of Artificial Nucleic Acids from (R)-4-amino-butane-1,3-diol. *Org. Bio. Chem.*, **2014**, *12*, 2263-2272.
65. Asanuma, H.; Toda, T.; Murayama, K.; Liang, X.; Kashida, H., Unexpectedly Stable Artificial Duplex from Flexible Acyclic Threoninol. *J. Am. Chem. Soc.*, **2010**, *132*, 14702-14703.
66. Kashida, H.; Murayama, K.; Toda, T.; Asanuma, H., Control of the Chirality and Helicity of Oligomers of Serinol Nucleic Acid (SNA) by Sequence Design. *Angew. Chem., Int. Ed.*, **2011**, *50*, 1285-1288.
67. Campbell, M. A.; Wengel, J., Locked vs. Unlocked Nucleic Acids (LNA vs.UNA): Contrasting Structures Work Towards Common Therapeutic Goals. *Chem. Soc. Rev.*, **2011**, *40*, 5680-5689.
68. Nielsen, P.; Egholm, M.; Berg, R.; Buchardt, O., Sequence-selective Recognition of DNA by Strand Displacement with a Thymine-substituted Polyamide. *Science*, **1991**, *254*, 1497-1500.
69. Braasch, D. A.; Corey, D. R., Synthesis, Analysis, Purification, and Intracellular Delivery of Peptide Nucleic Acids. *Methods*, **2001**, *23*, 97-107.
70. Gildea, B. D.; Casey, S.; MacNeill, J.; Perry-O'Keefe, H.; Sørensen, D.; Coull, J. M., PNA Solubility Enhancers. *Tetrahedron Lett.*, **1998**, *39*, 7255-7258.
71. Clivio, P.; Guillaume, D.; Adeline, M.-T.; Fourrey, J.-L., A Photochemical Approach to Highlight Backbone Effects in PNA. *J. Am. Chem. Soc.*, **1997**, *119*, 5255-5256.
72. Pinheiro, V. B.; Taylor, A. I.; Cozens, C.; Abramov, M.; Renders, M.; Zhang, S.; Chaput, J. C.; Wengel, J.; Peak-Chew, S.-Y.; McLaughlin, S. H.; Herdewijn, P.; Holliger, P., Synthetic Genetic Polymers Capable of Heredity and Evolution. *Science*, **2012**, *336*, 341-344.
73. Heuberger, B. D.; Switzer, C., A Pre-RNA Candidate Revisited: Both Enantiomers of Flexible Nucleoside Triphosphates are DNA Polymerase Substrates. *J. Am. Chem. Soc.*, **2007**, *130*, 412-413.
74. Yu, H.; Zhang, S.; Dunn, M. R.; Chaput, J. C., An Efficient and Faithful in vitro Replication System for Threose Nucleic Acid. *J. Am. Chem. Soc.*, **2013**, *135*, 3583-3591.
75. Nguyen, H. V.; Sallustrau, A.; Male, L.; Thornton, P. J.; Tucker, J. H. R.,

1,1′-Homodisubstituted Ferrocenes Containing Adenine and Thymine Nucleobases: Synthesis, Electrochemistry, and Formation of H-Bonded Arrays. *Organometallics*, **2011**, *30*, 5284-5290.

76. Nguyen, H. V.; Zhao, Z.-y.; Sallustrau, A.; Horswell, S. L.; Male, L.; Mulas, A.; Tucker, J. H. R., A Ferrocene Nucleic Acid Oligomer as an Organometallic Structural Mimic of DNA. *Chem. Commun.*, **2012**, *48*, 12165-12167.
77. Nguyen, H. V.; Sallustrau, A.; Balzarini, J.; Bedford, M. R.; Eden, J. C.; Georgousi, N.; Hodges, N. J.; Kedge, J.; Mehellou, Y.; Tselepis, C.; Tucker, J. H. R., Organometallic Nucleoside Analogues with Ferrocenyl Linker Groups: Synthesis and Cancer Cell Line Studies. *J. Med. Chem.*, **2014**, *57*, 5817-5822.
78. Hirao, I.; Kimoto, M.; Yamashige, R., Natural versus Artificial Creation of Base Pairs in DNA: Origin of Nucleobases from the Perspectives of Unnatural Base Pair Studies. *Acc. Chem. Res.*, **2012**, *45*, 2055-2065.
79. Jaramillo, C.; Knapp, S., Synthesis of *C*-Aryl Glycosides. *Synthesis*, **1994**, *1994*, 1-20.
80. Adamo, M. F. A.; Pergoli, R., Synthesis and Medicinal Properties of 2'-deoxyribose and Ribose *C*-Nucleosides. *Curr. Org. Chem.*, **2008**, *12*, 1544-1569.
81. Štambaský, J.; Hocek, M.; Kočovský, P., C-Nucleosides: Synthetic Strategies and Biological Applications. *Chem. Rev.*, **2009**, *109*, 6729-6764.
82. Jorgensen, W. L.; Pranata, J., Importance of Secondary Interactions in Triply Hydrogen Bonded Complexes: Guanine-cytosine vs Uracil-2,6-diaminopyridine. *J. Am. Chem. Soc.*, **1990**, *112*, 2008-2010.
83. Pranata, J.; Wierschke, S. G.; Jorgensen, W. L., OPLS Potential Functions for Nucleotide Bases. Relative Association Constants of Hydrogen-bonded Base Pairs in Chloroform. *J. Am. Chem. Soc.*, **1991**, *113*, 2810-2819.
84. Papmeyer, M.; Vuilleumier, C. A.; Pavan, G. M.; Zhurov, K. O.; Severin, K., Molecularly Defined Nanostructures Based on a Novel AAA-DDD Triple Hydrogen-Bonding Motif. *Angew. Chem., Int. Ed.*, **2016**, *55*, 1685-1689.
85. Switzer, C.; Moroney, S. E.; Benner, S. A., Enzymatic Incorporation of a New Base Pair into DNA and RNA. *J. Am. Chem. Soc.*, **1989**, *111*, 8322-8323.
86. Piccirilli, J. A.; Krauch, T.; Moroney, S. E.; Benner, S. A., Enzymatic Incorporation of a New Base Pair into DNA and RNA Extends the Genetic Alphabet. *Nature*, **1990**, *343*, 33-37.
87. Piccirilli, J. A.; Krauch, T.; MacPherson, L. J.; Benner, S. A., A Direct Route to 3-(D-Ribofuranosyl)pyridine Nucleosides. *Helv. Chim. Acta*, **1991**, *74*, 397-406.
88. Voegel, J. J.; Benner, S. A., Nonstandard Hydrogen Bonding in Duplex Oligonucleotides. The Base Pair between an Acceptor-Donor-Donor Pyrimidine Analog and a Donor-Acceptor-Acceptor Purine Analog. *J. Am. Chem. Soc.*, **1994**, *116*, 6929-6930.
89. von Krosigk, U.; Benner, S. A., pH-Independent Triple Helix Formation by an Oligonucleotide Containing a Pyrazine Donor-Donor-Acceptor Base. *J. Am. Chem. Soc.*, **1995**, *117*, 5361-5362.
90. Switzer, C. Y.; Moroney, S. E.; Benner, S. A., Enzymic Recognition of the Base Pair Between Isocytidine and Isoguanosine. *Biochemistry*, **1993**, *32*, 10489-10496.

91. Bain, J. D.; Switzer, C.; Chamberlin, R.; Bennert, S. A., Ribosome-mediated Incorporation of a Non-standard Amino Acid into a Peptide through Expansion of the Genetic Code. *Nature*, **1992**, *356*, 537-539.
92. Sismour, A. M.; Benner, S. A., The Use of Thymidine Analogs to Improve the Replication of an Extra DNA Base Pair: A Synthetic Biological System. *Nucleic Acids Res.*, **2005**, *33*, 5640-5646.
93. Horlacher, J.; Hottiger, M.; Podust, V. N.; Hübscher, U.; Benner, S. A., Recognition by Viral and Cellular DNA Polymerases of Nucleosides Bearing Bases with Nonstandard Hydrogen Bonding Patterns. *Proc. Natl. Acad. Sci. USA*, **1995**, *92*, 6329-6333.
94. Lutz, M. J.; Held, H. A.; Hottiger, M.; Hübscher, U.; Benner, S. A., Differential Discrimination of DNA Polymerases for Variants of the Non-standard Nucleobase Pair Between Xanthosine and 2,4-Diaminopyrimidine, Two Components of an Expanded Genetic Alphabet. *Nucleic Acids Res.*, **1996**, *24*, 1308-1313.
95. Benner, S. A., Understanding Nucleic Acids Using Synthetic Chemistry. *Acc. Chem. Res.*, **2004**, *37*, 784-797.
96. Hirao, I.; Ohtsuki, T.; Fujiwara, T.; Mitsui, T.; Yokogawa, T.; Okuni, T.; Nakayama, H.; Takio, K.; Yabuki, T.; Kigawa, T.; Kodama, K.; Yokogawa, T.; Nishikawa, K.; Yokoyama, S., An Unnatural Base Pair for Incorporating Amino Acid Analogs into Proteins. *Nat. Biotech.*, **2002**, *20*, 177-182.
97. Hirao, I.; Harada, Y.; Kimoto, M.; Mitsui, T.; Fujiwara, T.; Yokoyama, S., A Two-unnatural-base-pair System toward the Expansion of the Genetic Code. *J. Am. Chem. Soc.*, **2004**, *126*, 13298-13305.
98. Scopes, D.; Barrio; Leonard, N., Defined Dimensional Changes in Enzyme Cofactors: Fluorescent "Stretched-out" Analogs of Adenine Nucleotides. *Science*, **1977**, *195*, 296-298.
99. Liu, H.; Gao, J.; Lynch, S. R.; Saito, Y. D.; Maynard, L.; Kool, E. T., A Four-Base Paired Genetic Helix with Expanded Size. *Science*, **2003**, *302*, 868-871.
100. Liu, H.; Gao, J.; Maynard, L.; Saito, Y. D.; Kool, E. T., Toward a New Genetic System with Expanded Dimensions: Size-expanded Analogues of Deoxyadenosine and Thymidine. *J. Am. Chem. Soc.*, **2004**, *126*, 1102-1109.
101. Lu, H.; He, K.; Kool, E. T., yDNA: A New Geometry for Size-expanded Base Pairs. *Angew. Chem., Int. Ed.*, **2004**, *43*, 5834-5836.
102. Delaney, J. C.; Gao, J.; Liu, H.; Shrivastav, N.; Essigmann, J. M.; Kool, E. T., Efficient Replication Bypass of Size-Expanded DNA Base Pairs in Bacterial Cells. *Angew. Chem., Int. Ed.*, **2009**, *48*, 4524-4527.
103. Minakawa, N.; Kojima, N.; Hikishima, S.; Sasaki, T.; Kiyosue, A.; Atsumi, N.; Ueno, Y.; Matsuda, A., New Base Pairing Motifs. The Synthesis and Thermal Stability of Oligodeoxynucleotides Containing Imidazopyridopyrimidine Nucleosides with the Ability to Form Four Hydrogen Bonds. *J. Am. Chem. Soc.*, **2003**, *125*, 9970-9982.
104. Hikishima, S.; Minakawa, N.; Kuramoto, K.; Fujisawa, Y.; Ogawa, M.; Matsuda, A., Synthesis of 1,8-Naphthyridine *C*-Nucleosides and Their Base- pairing Properties in Oligodeoxynucleotides: Thermally Stable Naphthyridine:Imidazopyridopyrimidine

Base-pairing Motifs. *Angew. Chem., Int. Ed.*, **2005**, *44*, 596-598.

105. Minakawa, N.; Ogata, S.; Takahashi, M.; Matsuda, A., Selective Recognition of Unnatural Imidazopyridopyrimidine:Naphthyridine Base Pairs Consisting of Four Hydrogen Bonds by the Klenow Fragment. *J. Am. Chem. Soc.*, **2009**, *131*, 1644-1645.
106. Doi, Y.; Chiba, J.; Morikawa, T.; Inouye, M., Artificial DNA Made Exclusively of Nonnatural *C*-Nucleosides with Four Types of Nonnatural Bases. *J. Am. Chem. Soc.*, **2008**, *130*, 8762-8768.
107. Shin, D.; Sinkeldam, R. W.; Tor, Y., Emissive RNA Alphabet. *J. Am. Chem. Soc.*, **2011**, *133*, 14912-14915.
108. Rovira, A. R.; Fin, A.; Tor, Y., Chemical Mutagenesis of an Emissive RNA Alphabet. *J. Am. Chem. Soc.*, **2015**, *137*, 14602-14605.
109. Moran, S.; Ren, R. X. F.; Rumney; Kool, E. T., Difluorotoluene, a Nonpolar Isostere for Thymine, Codes Specifically and Efficiently for Adenine in DNA Replication. *J. Am. Chem. Soc.*, **1997**, *119*, 2056-2057.
110. Moran, S.; Ren, R. X.-F.; Kool, E. T., A Thymidine Triphosphate Shape Analog Lacking Watson-Crick Pairing Ability is Replicated with High Sequence Selectivity. *Proc. Natl. Acad. Sci. USA*, **1997**, *94*, 10506-10511.
111. Matray, T. J.; Kool, E. T., A Specific Partner for Abasic Damage in DNA. *Nature*, **1999**, *399*, 704-708.
112. Schweitzer, B. A.; Kool, E. T., Aromatic Nonpolar Nucleosides as Hydrophobic Isosteres of Pyrimidines and Purine Nucleosides. *J. Org. Chem.*, **1994**, *59*, 7238-7242.
113. Schweitzer, B. A.; Kool, E. T., Hydrophobic, Non-hydrogen-bonding Bases and Base Pairs in DNA. *J. Am. Chem. Soc.*, **1995**, *117*, 1863-1872.
114. Morales, J. C.; Kool, E. T., Efficient Replication between Non-hydrogen-bonded Nucleoside Shape Analogs. *Nat. Struct. Mol. Biol.*, **1998**, *5*, 950-954.
115. Morales, J. C.; Kool, E. T., Minor Groove Interactions between Polymerase and DNA: More Essential to Replication than Watson-Crick Hydrogen Bonds? *J. Am. Chem. Soc.*, **1999**, *121*, 2323-2324.
116. Hirao, I.; Kimoto, M.; Mitsui, T.; Fujiwara, T.; Kawai, R.; Sato, A.; Harada, Y.; Yokoyama, S., An Unnatural Hydrophobic Base Pair System: Site-specific Incorporation of Nucleotide Analogs into DNA and RNA. *Nat. Methods*, **2006**, *3*, 729-735.
117. Kimoto, M.; Mitsui, T.; Yokoyama, S.; Hirao, I., A Unique Fluorescent Base Analogue for the Expansion of the Genetic Alphabet. *J. Am. Chem. Soc.*, **2010**, *132*, 4988-4989.
118. Kimoto, M.; Mitsui, T.; Yamashige, R.; Sato, A.; Yokoyama, S.; Hirao, I., A New Unnatural Base Pair System between Fluorophore and Quencher Base Analogues for Nucleic Acid-based Imaging Technology. *J. Am. Chem. Soc.*, **2010**, *132*, 15418-15426.
119. McMinn, D. L.; Ogawa, A. K.; Wu, Y.; Liu, J.; Schultz, P. G.; Romesberg, F. E., Efforts toward Expansion of the Genetic Alphabet: DNA Polymerase Recognition of a Highly Stable, Self-Pairing Hydrophobic Base. *J. Am. Chem. Soc.*, **1999**, *121*, 11585-11586.
120. Yu, C.; Henry, A. A.; Romesberg, F. E.; Schultz, P. G., Polymerase Recognition of Unnatural Base Pairs. *Angew. Chem., Int. Ed.*, **2002**, *41*, 3841-3844.

121. Matsuda, S.; Leconte, A. M.; Romesberg, F. E., Minor Groove Hydrogen Bonds and the Replication of Unnatural Base Pairs. *J. Am. Chem. Soc.*, **2007**, *129*, 5551-5557.
122. Hwang, G. T.; Romesberg, F. E., Unnatural Substrate Repertoire of A, B, and X Family DNA Polymerases. *J. Am. Chem. Soc.*, **2008**, *130*, 14872-14882.
123. Seo, Y. J.; Romesberg, F. E., Major Groove Derivatization of an Unnatural Base Pair. *ChemBioChem*, **2009**, *10*, 2394-2400.
124. Seo, Y. J.; Hwang, G. T.; Ordoukhanian, P.; Romesberg, F. E., Optimization of an Unnatural Base Pair toward Natural-like Replication. *J. Am. Chem. Soc.*, **2009**, *131*, 3246-3252.
125. Seo, Y. J.; Matsuda, S.; Romesberg, F. E., Transcription of an Expanded Genetic Alphabet. *J. Am. Chem. Soc.*, **2009**, *131*, 5046-5047.
126. Malyshev, D. A.; Seo, Y. J.; Ordoukhanian, P.; Romesberg, F. E., PCR with an Expanded Genetic Alphabet. *J. Am. Chem. Soc.*, **2009**, *131*, 14620-14621.
127. Betz, K.; Malyshev, D. A.; Lavergne, T.; Welte, W.; Diederichs, K.; Dwyer, T. J.; Ordoukhanian, P.; Romesberg, F. E.; Marx, A., KlenTaq Polymerase Replicates Unnatural Base Pairs by Inducing a Watson-Crick Geometry. *Nat. Chem. Biol.*, **2012**, *8*, 612-614.
128. Malyshev, D. A.; Dhami, K.; Quach, H. T.; Lavergne, T.; Ordoukhanian, P.; Torkamani, A.; Romesberg, F. E., Efficient and Sequence-independent Replication of DNA Containing a Third Base Pair Establishes a Functional Six-letter Genetic Alphabet. *Proc. Natl. Acad. Sci. USA*, **2012**, *109*, 12005-12010.
129. Malyshev, D. A.; Dhami, K.; Lavergne, T.; Chen, T.; Dai, N.; Foster, J. M.; Correa, I. R.; Romesberg, F. E., A Semi-synthetic Organism with an Expanded Genetic Alphabet. *Nature*, **2014**, *509*, 385-388.
130. Tomás-Gamasa, M.; Serdjukow, S.; Su, M.; Müller, M.; Carell, T., "Post-It" Type Connected DNA Created with a Reversible Covalent Crosslink. *Angew. Chem., Int. Ed.*, **2015**, *54*, 796-800.
131. Yang, Z.; Price, N. E.; Johnson, K. M.; Gates, K. S., Characterization of Interstrand DNA-DNA Cross-links Derived from Abasic Sites Using Bacteriophage Φ29 DNA Polymerase. *Biochemistry*, **2015**, *54*, 4259-4266.
132. Johnson, K. M.; Price, N. E.; Wang, J.; Fekry, M. I.; Dutta, S.; Seiner, D. R.; Wang, Y.; Gates, K. S., On the Formation and Properties of Interstrand DNA-DNA Cross-Links Forged by Reaction of an Abasic Site with the Opposing Guanine Residue of 5′-CAp Sequences in Duplex DNA. *J. Am. Chem. Soc.*, **2013**, *135*, 1015-1025.
133. Katz, S., The Reversible Reaction of Sodium Thymonucleate and Mercuric Chloride. *J. Am. Chem. Soc.*, **1952**, *74*, 2238-2245.
134. Thomas, C. A., The Interaction of $HgCl_2$ with Sodium Thymonucleate. *J. Am. Chem. Soc.*, **1954**, *76*, 6032-6034.
135. Katz, S., The Reversible Reaction of Hg (II) and Double-stranded Plynucleotides a Step-function Theory and its Significance. *Biochim. Biophys. Acta*, **1963**, *68*, 240-253.
136. Atwell, S.; Meggers, E.; Spraggon, G.; Schultz, P. G., Structure of a Copper-mediated Base Pair in DNA. *J. Am. Chem. Soc.*, **2001**, *123*, 12364-12367.

137. Kaul, C.; Müller, M.; Wagner, M.; Schneider, S.; Carell, T., Reversible Bond Formation Enables the Replication and Amplification of a Crosslinking Salen Complex as an Orthogonal Base Pair. *Nat. Chem.*, **2011**, *3*, 794-800.
138. Ono, A.; Togashi, H., Highly Selective Oligonucleotide-based Sensor for Mercury(II) in Aqueous Solutions. *Angew. Chem., Int. Ed.*, **2004**, *43*, 4300-4302.
139. Miyake, Y.; Togashi, H.; Tashiro, M.; Yamaguchi, H.; Oda, S.; Kudo, M.; Tanaka, Y.; Kondo, Y.; Sawa, R.; Fujimoto, T.; Machinami, T.; Ono, A., Mercury(II)-mediated Formation of Thymine-Hg(II)-Thymine Base Pairs in DNA Duplexes. *J. Am. Chem. Soc.*, **2006**, *128*, 2172-2173.
140. Ono, A.; Cao, S.; Togashi, H.; Tashiro, M.; Fujimoto, T.; Machinami, T.; Oda, S.; Miyake, Y.; Okamoto, I.; Tanaka, Y., Specific Interactions between Silver(I) Ions and Cytosine-cytosine Pairs in DNA Duplexes. *Chem. Commun.*, **2008**, *44*, 4825-4827.
141. Okamoto, I.; Iwamoto, K.; Watanabe, Y.; Miyake, Y.; Ono, A., Metal-Ion Selectivity of Chemically Modified Uracil Pairs in DNA Duplexes. *Angew. Chem., Int. Ed.*, **2009**, *48*, 1648-1651.
142. Ennifar, E.; Walter, P.; Dumas, P., A Crystallographic Study of the Binding of 13 Metal Ions to Two Related RNA Duplexes. *Nucleic Acids Res.*, **2003**, *31*, 2671-2682.
143. Kondo, J.; Yamada, T.; Hirose, C.; Okamoto, I.; Tanaka, Y.; Ono, A., Crystal Structure of Metallo DNA Duplex Containing Consecutive Watson-Crick-like T-Hg(II)-T Base Pairs. *Angew. Chem., Int. Ed.*, **2014**, *53*, 2385-2388.
144. Kondo, J.; Tada, Y.; Dairaku, T.; Saneyoshi, H.; Okamoto, I.; Tanaka, Y.; Ono, A., High-resolution Crystal Structure of Silver(I)-RNA Hybrid Duplex Containing Watson-Crick-like C-Silver(I)-C Metallo-Base Pairs. *Angew. Chem., Int. Ed.*, **2015**, 13323-13326.
145. Freeman, R.; Finder, T.; Willner, I., Multiplexed Analysis of Hg^{2+} and Ag^{+} Ions by Nucleic Acid Functionalized CdSe/ZnS Quantum Dots and Their Use for Logic Gate Operations. *Angew. Chem., Int. Ed.*, **2009**, *48*, 7818-7821.
146. Ihara, T.; Ishii, T.; Araki, N.; Wilson, A. W.; Jyo, A., Silver Ion Unusually Stabilizes the Structure of a Parallel-Motif DNA Triplex. *J. Am. Chem. Soc.*, **2009**, *131*, 3826-3827.
147. Urata, H.; Yamaguchi, E.; Funai, T.; Matsumura, Y.; Wada, S.-i., Incorporation of Thymine Nucleotides by DNA Polymerases through T-Hg(II)-T Base Pairing. *Angew. Chem., Int. Ed.*, **2010**, *49*, 6516-6519.
148. Funai, T.; Miyazaki, Y.; Aotani, M.; Yamaguchi, E.; Nakagawa, O.; Wada, S.-i.; Torigoe, H.; Ono, A.; Urata, H., AgI Ion Mediated Formation of a C-A Mispair by DNA Polymerases. *Angew. Chem., Int. Ed.*, **2012**, *51*, 6464-6466.
149. Johannsen, S.; Paulus, S.; Düpre, N.; Müller, J.; Sigel, R. K. O., Using in vitro Transcription to Construct Scaffolds for One-dimensional Arrays of Mercuric Ions. *J. Inorg. Biochem.*, **2008**, *102*, 1141-1151.
150. Tanaka, K.; Shionoya, M., Synthesis of a Novel Nucleoside for Alternative DNA Base Pairing through Metal Complexation. *J. Org. Chem.*, **1999**, *64*, 5002-5003.
151. Tasaka, M.; Tanaka, K.; Shiro, M.; Shionoya, M., A Palladium-mediated DNA Base Pair of a β-*C*-Nucleoside Possessing a 2-Aminophenol as the Nucleobase. *Supramol.*

Chem., **2001**, *13*, 671-675.

152. Cao, H.; Tanaka, K.; Shionoya, M., An Alternative Base-pairing of Catechol-bearing Nucleosides by Borate Formation. *Chem. Pharm. Bull.*, **2000**, *48*, 1745-1748.
153. Tanaka, K.; Tasaka, M.; Cao, H.; Shionoya, M., An Approach to Metal-assisted DNA Base Pairing: Novel β-*C*-Nucleosides with a 2-Aminophenol or a Catechol as the Nucleobase. *Eur. J. Pharm. Sci.*, **2001**, *13*, 77-83.
154. Tanaka, K.; Tengeiji, A.; Kato, T.; Toyama, N.; Shiro, M.; Shionoya, M., Efficient Incorporation of a Copper Hydroxypyridone Base Pair in DNA. *J. Am. Chem. Soc.*, **2002**, *124*, 12494-12498.
155. Tanaka, K.; Yamada, Y.; Shionoya, M., Formation of Silver(I)-Mediated DNA Duplex and Triplex through an Alternative Base Pair of Pyridine Nucleobases. *J. Am. Chem. Soc.*, **2002**, *124*, 8802-8803.
156. Tanaka, K.; Tengeiji, A.; Kato, T.; Toyama, N.; Shionoya, M., A Discrete Self-assembled Metal Array in Artificial DNA. *Science*, **2003**, *299*, 1212-1213.
157. Takezawa, Y.; Maeda, W.; Tanaka, K.; Shionoya, M., Discrete Self-assembly of Iron(III) Ions inside Triple-stranded Artificial DNA. *Angew. Chem., Int. Ed.*, **2009**, *48*, 1081-1084.
158. Meggers, E.; Holland, P. L.; Tolman, W. B.; Romesberg, F. E.; Schultz, P. G., A Novel Copper-mediated DNA Base Pair. *J. Am. Chem. Soc.*, **2000**, *122*, 10714-10715.
159. Zimmermann, N.; Meggers, E.; Schultz, P. G., A Second-generation Copper(II)-mediated Metallo-DNA-base pair. *Bioorg. Chem.*, **2004**, *32*, 13-25.
160. Zimmermann, N.; Meggers, E.; Schultz, P. G., A Novel Silver(I)-mediated DNA Base Pair. *J. Am. Chem. Soc.*, **2002**, *124*, 13684-13685.
161. Böhme, D.; Düpre, N.; Megger, D. A.; Müller, J., Conformational Change Induced by Metal-ion-bnding to DNA Containing the Artificial 1,2,4-Triazole Nucleoside. *Inorg. Chem.*, **2007**, *46*, 10114-10119.
162. Seubert, K.; Guerra, C. F.; Bickelhaupt, F. M.; Muller, J., Chimeric GNA/DNA Metal-mediated Base Pairs. *Chem. Commun.*, **2011**, *47*, 11041-11043.
163. Zhang, L.; Meggers, E., An Extremely Stable and Orthogonal DNA Base Pair with a Simplified Three-Carbon Backbone. *J. Am. Chem. Soc.*, **2004**, *127*, 74-75.
164. Shin, D.; Switzer, C., A Metallo Base-pair Incorporating a Terpyridyl-like Motif: Bipyridyl-pyrimidinone-Ag(I)-4-pyridine. *Chem. Commun.*, **2007**, *43*, 4401-4403.
165. Heuberger, B. D.; Shin, D.; Switzer, C., Two Watson-Crick-Like Metallo Base-Pairs. *Org. Lett.*, **2008**, *10*, 1091-1094.
166. Switzer, C.; Sinha, S.; Kim, P. H.; Heuberger, B. D., A Purine-like Nickel(II) Base Pair for DNA. *Angew. Chem., Int. Ed.*, **2005**, *44*, 1529-1532.
167. Switzer, C.; Shin, D., A Pyrimidine-like Nickel(II) DNA Base Pair. *Chem. Commun.*, **2005**, *41*, 1342-1344.
168. Weizman, H.; Tor, Y., 2,2‘-Bipyridine Ligandoside: A Novel Building Block for Modifying DNA with Intra-duplex Metal Complexes. *J. Am. Chem. Soc.*, **2001**, *123*, 3375-3376.
169. Brotschi, C.; Häberli, A.; Leumann, C. J., A Stable DNA Duplex Containing a Non-

hydrogen-bonding and Non-shape-complementary Base Couple: Interstrand Stacking as the Stability Determining Factor. *Angew. Chem., Int. Ed.*, **2001**, *40*, 3012-3014.

170. Megger, D. A.; Fonseca Guerra, C.; Hoffmann, J.; Brutschy, B.; Bickelhaupt, F. M.; Müller, J., Contiguous Metal-mediated Base Pairs Comprising Two AgI Ions. *Chem. Eur. J.*, **2011**, *17*, 6533-6544.
171. Mei, H.; Röhl, I.; Seela, F., Ag^{+}-Mediated DNA Base Pairing: Extraordinarily Stable Pyrrolo-dC–Pyrrolo-dC Pairs Binding Two Silver Ions. *J. Org. Chem.*, **2013**, *78*, 9457-9463.
172. Sinha, I.; Fonseca Guerra, C.; Müller, J., A Highly Stabilizing Silver(I)-mediated Base Pair in Parallel-Stranded DNA. *Angew. Chem., Int. Ed.*, **2015**, *54*, 3603-3606.
173. Clever, G. H.; Polborn, K.; Carell, T., A Highly DNA-duplex-stabilizing Metal–salen Base Pair. *Angew. Chem., Int. Ed.*, **2005**, *44*, 7204-7208.
174. Clever, G. H.; Söltl, Y.; Burks, H.; Spahl, W.; Carell, T., Metal-Salen-Base-Pair Complexes Inside DNA: Complexation Overrides Sequence Information. *Chem. Eur. J.*, **2006**, *12*, 8708-8718.
175. Clever, G. H.; Carell, T., Controlled Stacking of 10 Transition-metal Ions inside a DNA Duplex. *Angew. Chem., Int. Ed.*, **2007**, *46*, 250-253.
176. Tanaka, K.; Clever, G. H.; Takezawa, Y.; Yamada, Y.; Kaul, C.; Shionoya, M.; Carell, T., Programmable Self-assembly of Metal Ions inside Artificial DNA Duplexes. *Nat. Nano*, **2006**, *1*, 190-194.
177. Jain, M. L.; Bruice, P. Y.; Szabó, I. E.; Bruice, T. C., Incorporation of Positively Charged Linkages into DNA and RNA Backbones: A Novel Strategy for Antigene and Antisense Agents. *Chem. Rev.*, **2012**, *112*, 1284-1309.
178. Navarro, E.; Serrano-Heras, G.; Castaño, M. J.; Solera, J., Real-time PCR Detection Chemistry. *Clin. Chim. Acta*, **2015**, *439*, 231-250.
179. Johnson, T. B.; Coghill, R. D., Researches on Pyrimidines. C111. The Discovery of 5-Methylcytosine in Tuberculinic Acid, the Nucleic Acid of the Tubercle Bacillus 1. *J. Am. Chem. Soc.*, **1925**, *47*, 2838-2844.
180. Holliday, R.; Pugh, J., DNA Modification Mechanisms and Gene Activity During Development. *Science*, **1975**, *187*, 226-232.
181. Goll, M. G.; Bestor, T. H., Eukaryotic Cytosine Methyltransferases. *Ann. Rev. Biochem.*, **2005**, *74*, 481-514.
182. Cheng, X.; Blumenthal, R. M., Mammalian DNA Methyltransferases: A Structural Perspective. *Structure*, **2008**, *16*, 341-350.
183. Chiang, P. K.; Gordon, R. K.; Tal, J.; Zeng, G. C.; Doctor, B. P.; Pardhasaradhi, K.; McCann, P. P., *S*-Adenosylmethionine and Methylation. *FASEB J.*, **1996**, *10*, 471-80.
184. Bird, A., DNA Methylation Patterns and Epigenetic Memory. *Genes Dev.*, **2002**, *16*, 6-21.
185. Chan, S. W. L.; Henderson, I. R.; Jacobsen, S. E., Gardening the Genome: DNA Methylation in Arabidopsis Thaliana. *Nat. Rev. Genet.*, **2005**, *6*, 351-360.
186. Law, J. A.; Jacobsen, S. E., Establishing, Maintaining and Modifying DNA Methylation Patterns in Plants and Animals. *Nat. Rev. Genet.*, **2010**, *11*, 204-220.

187. Deaton, A. M.; Bird, A., CpG Islands and the Regulation of Transcription. *Genes Dev.*, **2011**, *25*, 1010-1022.
188. Ramsahoye, B. H.; Biniszkiewicz, D.; Lyko, F.; Clark, V.; Bird, A. P.; Jaenisch, R., Non-CpG Methylation is Prevalent in Embryonic Stem Cells and may be Mediated by DNA Methyltransferase 3a. *Proc. Natl. Acad. Sci. USA*, **2000**, *97*, 5237-5242.
189. Shirane, K.; Toh, H.; Kobayashi, H.; Miura, F.; Chiba, H.; Ito, T.; Kono, T.; Sasaki, H., Mouse Oocyte Methylomes at Base Resolution Reveal Genome-wide Accumulation of Non-CpG Methylation and Role of DNA Methyltransferases. *PLoS Genet.*, **2013**, *9*, e1003439.
190. Lister, R.; Mukamel, E. A.; Nery, J. R.; Urich, M.; Puddifoot, C. A.; Johnson, N. D.; Lucero, J.; Huang, Y.; Dwork, A. J.; Schultz, M. D.; Yu, M.; Tonti-Filippini, J.; Heyn, H.; Hu, S.; Wu, J. C.; Rao, A.; Esteller, M.; He, C.; Haghighi, F. G.; Sejnowski, T. J.; Behrens, M. M.; Ecker, J. R., Global Epigenomic Reconfiguration During Mammalian Brain Development. *Science*, **2013**, *341*, 629.
191. Xie, W.; Barr, Cathy L.; Kim, A.; Yue, F.; Lee, Ah Y.; Eubanks, J.; Dempster, Emma L.; Ren, B., Base-resolution Analyses of Sequence and Parent-of-origin Dependent DNA Methylation in the Mouse Genome. *Cell*, **2011**, *148*, 816-831.
192. Ehrlich, M.; Wang, R., 5-Methylcytosine in Eukaryotic DNA. *Science*, **1981**, *212*, 1350-1357.
193. Feinberg, A. P.; Ohlsson, R.; Henikoff, S., The Epigenetic Progenitor Origin of Human Cancer. *Nat. Rev. Genet.*, **2006**, *7*, 21-33.
194. Razin, A.; Riggs, A., DNA Methylation and Gene Function. *Science*, **1980**, *210*, 604-610.
195. Lister, R.; Pelizzola, M.; Dowen, R. H.; Hawkins, R. D.; Hon, G.; Tonti-Filippini, J.; Nery, J. R.; Lee, L.; Ye, Z.; Ngo, Q.-M.; Edsall, L.; Antosiewicz-Bourget, J.; Stewart, R.; Ruotti, V.; Millar, A. H.; Thomson, J. A.; Ren, B.; Ecker, J. R., Human DNA Methylomes at Base Resolution Show Widespread Epigenomic Differences. *Nature*, **2009**, *462*, 315-322.
196. Rauch, T. A.; Wu, X.; Zhong, X.; Riggs, A. D.; Pfeifer, G. P., A Human B Cell Methylome at 100-Base Pair Resolution. *Proc. Natl. Acad. Sci. USA*, **2009**, *106*, 671-678.
197. Laurent, L.; Wong, E.; Li, G.; Huynh, T.; Tsirigos, A.; Ong, C. T.; Low, H. M.; Kin Sung, K. W.; Rigoutsos, I.; Loring, J.; Wei, C.-L., Dynamic Changes in the Human Methylome during Differentiation. *Genome Res.*, **2010**, *20*, 320-331.
198. Chen, T.; Ueda, Y.; Dodge, J. E.; Wang, Z.; Li, E., Establishment and Maintenance of Genomic Methylation Patterns in Mouse Embryonic Stem Cells by Dnmt3a and Dnmt3b. *Mol. Cell Bio.*, **2003**, *23*, 5594-5605.
199. Wu, S. C.; Zhang, Y., Active DNA Demethylation: Many Roads Lead to Rome. *Nat. Rev. Mol. Cell Biol.*, **2010**, *11*, 607-620.
200. Bhutani, N.; Burns, David M.; Blau, Helen M., DNA Demethylation Dynamics. *Cell*, **2011**, *146*, 866-872.
201. Zhu, J.-K., Active DNA Demethylation Mediated by DNA Glycosylases. *Ann. Rev.*

Genet., **2009**, *43*, 143-166.

202. Bhattacharya, S. K.; Ramchandani, S.; Cervoni, N.; Szyf, M., A Mammalian Protein with Specific Demethylase Activity for mCpG DNA. *Nature*, **1999**, *397*, 579-583.
203. Zhu, B.; Zheng, Y.; Hess, D.; Angliker, H.; Schwarz, S.; Siegmann, M.; Thiry, S.; Jost, J.-P., 5-Methylcytosine-DNA Glycosylase Activity is Present in a Cloned G/T Mismatch DNA Glycosylase Associated with the Chicken Embryo DNA Demethylation Complex. *Proc. Natl. Acad. Sci. USA*, **2000**, *97*, 5135-5139.
204. Barreto, G.; Schafer, A.; Marhold, J.; Stach, D.; Swaminathan, S. K.; Handa, V.; Doderlein, G.; Maltry, N.; Wu, W.; Lyko, F.; Niehrs, C., Gadd45a Promotes Epigenetic Gene Activation by Repair-mediated DNA Demethylation. *Nature*, **2007**, *445*, 671-675.
205. Nabel, C. S.; Jia, H.; Ye, Y.; Shen, L.; Goldschmidt, H. L.; Stivers, J. T.; Zhang, Y.; Kohli, R. M., AID/APOBEC deaminases disfavor modified cytosines implicated in DNA demethylation. *Nat Chem Biol*, **2012**, *8*, 751-758.
206. Popp, C.; Dean, W.; Feng, S.; Cokus, S. J.; Andrews, S.; Pellegrini, M.; Jacobsen, S. E.; Reik, W., Genome-wide Erasure of DNA Methylation in Mouse Primordial Germ Cells is Affected by AID Deficiency. *Nature*, **2010**, *463*, 1101-1105.
207. Kumar, R.; DiMenna, L.; Schrode, N.; Liu, T.-C.; Franck, P.; Munoz-Descalzo, S.; Hadjantonakis, A.-K.; Zarrin, A. A.; Chaudhuri, J.; Elemento, O.; Evans, T., AID Stabilizes Stem-cell Phenotype by Removing Epigenetic Memory of Pluripotency Genes. *Nature*, **2013**, *500*, 89-92.
208. Metivier, R.; Gallais, R.; Tiffoche, C.; Le Peron, C.; Jurkowska, R. Z.; Carmouche, R. P.; Ibberson, D.; Barath, P.; Demay, F.; Reid, G.; Benes, V.; Jeltsch, A.; Gannon, F.; Salbert, G., Cyclical DNA Methylation of a Transcriptionally Active Promoter. *Nature*, **2008**, *452*, 45-50.
209. Wyatt, G. R.; Cohen, S. S., A New Pyrimidine Base from Bacteriophage Nucleic Acids. *Nature*, **1952**, *170*, 1072-1073.
210. Penn, N. W.; Suwalski, R.; O'Riley, C.; Bojanowski, K.; Yura, R., The Presence of 5-Hydroxymethylcytosine in Animal Deoxyribonucleic Acid. *Biochem. J.*, **1972**, *126*, 781-790.
211. Kriaucionis, S.; Heintz, N., The Nuclear DNA Base 5-Hydroxymethylcytosine is Present in Purkinje Neurons and the Brain. *Science*, **2009**, *324*, 929-930.
212. Tahiliani, M.; Koh, K. P.; Shen, Y.; Pastor, W. A.; Bandukwala, H.; Brudno, Y.; Agarwal, S.; Iyer, L. M.; Liu, D. R.; Aravind, L.; Rao, A., Conversion of 5-Methylcytosine to 5-Hydroxymethylcytosine in Mammalian DNA by MLL Partner TET1. *Science*, **2009**, *324*, 930-935.
213. Yu, M.; Hon, Gary C.; Szulwach, Keith E.; Song, C.-X.; Zhang, L.; Kim, A.; Li, X.; Dai, Q.; Shen, Y.; Park, B.; Min, J.-H.; Jin, P.; Ren, B.; He, C., Base-resolution Analysis of 5-Hydroxymethylcytosine in the Mammalian Genome. *Cell*, **2012**, *149*, 1368-1380.
214. Booth, M. J.; Branco, M. R.; Ficz, G.; Oxley, D.; Krueger, F.; Reik, W.; Balasubramanian, S., Quantitative Sequencing of 5-Methylcytosine and 5-Hydroxymethylcytosine at Single-Base Resolution. *Science*, **2012**, *336*, 934-937.

215. Szulwach, K. E.; Li, X.; Li, Y.; Song, C.-X.; Han, J. W.; Kim, S.; Namburi, S.; Hermetz, K.; Kim, J. J.; Rudd, M. K.; Yoon, Y.-S.; Ren, B.; He, C.; Jin, P., Integrating 5-Hydroxymethylcytosine into the Epigenomic Landscape of Human Embryonic Stem Cells. *PLoS Genet.*, **2011**, *7*, e1002154.
216. Xu, Y.; Wu, F.; Tan, L.; Kong, L.; Xiong, L.; Deng, J.; Barbera, A. J.; Zheng, L.; Zhang, H.; Huang, S.; Min, J.; Nicholson, T.; Chen, T.; Xu, G.; Shi, Y.; Zhang, K.; Shi, Yujiang G., Genome-wide Regulation of hmdC, mdC, and Gene Expression by Tet1 Hydroxylase in Mouse Embryonic Stem Cells. *Mol. Cell*, **2011**, *42*, 451-464.
217. Stroud, H.; Feng, S.; Morey Kinney, S.; Pradhan, S.; Jacobsen, S. E., 5-Hydroxymethylcytosine is Associated with Enhancers and Gene Bodies in Human Embryonic Stem Cells. *Genome Biol.*, **2011**, *12*, R54.
218. Shen, Y.; Yue, F.; McCleary, D. F.; Ye, Z.; Edsall, L.; Kuan, S.; Wagner, U.; Dixon, J.; Lee, L.; Lobanenkov, V. V.; Ren, B., A Map of the cis-Regulatory Sequences in the Mouse Genome. *Nature*, **2012**, *488*, 116-120.
219. Münzel, M.; Globisch, D.; Brückl, T.; Wagner, M.; Welzmiller, V.; Michalakis, S.; Müller, M.; Biel, M.; Carell, T., Quantification of the Sixth DNA Base Hydroxymethylcytosine in the Brain. *Angew. Chem., Int. Ed.*, **2010**, *49*, 5375-5377.
220. Globisch, D.; Münzel, M.; Müller, M.; Michalakis, S.; Wagner, M.; Koch, S.; Brückl, T.; Biel, M.; Carell, T., Tissue Distribution of 5-Hydroxymethylcytosine and Search for Active Demethylation Intermediates. *PLoS ONE*, **2010**, *5*, e15367.
221. Li, W.; Liu, M., Distribution of 5-Hydroxymethylcytosine in Different Human Tissues. *J. Nucleic Acids*, **2011**, *2011*, 870726.
222. Wu, H.; Zhang, Y., Reversing DNA Methylation: Mechanisms, Genomics, and Biological Functions. *Cell*, **2014**, *156*, 45-68.
223. Pastor, W. A.; Pape, U. J.; Huang, Y.; Henderson, H. R.; Lister, R.; Ko, M.; McLoughlin, E. M.; Brudno, Y.; Mahapatra, S.; Kapranov, P.; Tahiliani, M.; Daley, G. Q.; Liu, X. S.; Ecker, J. R.; Milos, P. M.; Agarwal, S.; Rao, A., Genome-wide Mapping of 5-Hydroxymethylcytosine in Embryonic Stem Cells. *Nature*, **2011**, *473*, 394-397.
224. Mellén, M.; Ayata, P.; Dewell, S.; Kriaucionis, S.; Heintz, N., MeCP2 Binds to hmdC Enriched within Active Genes and Accessible Chromatin in the Nervous System. *Cell*, **2012**, *151*, 1417-1430.
225. Guo, Junjie U.; Su, Y.; Zhong, C.; Ming, G.-l.; Song, H., Hydroxylation of 5-Methylcytosine by TET1 Promotes Active DNA Demethylation in the Adult Brain. *Cell*, **2011**, *145*, 423-434.
226. Chen, C.-C.; Wang, K.-Y.; Shen, C.-K. J., The Mammalian de Novo DNA Methyltransferases DNMT3A and DNMT3B are also DNA 5-Hydroxymethylcytosine Dehydroxymethylases. *J. Biol. Chem.*, **2012**, *287*, 33116-33121.
227. Ito, S.; Shen, L.; Dai, Q.; Wu, S. C.; Collins, L. B.; Swenberg, J. A.; He, C.; Zhang, Y., Tet Proteins Can Convert 5-Methylcytosine to 5-Formylcytosine and 5-Carboxylcytosine. *Science*, **2011**, *333*, 1300-1303.
228. He, Y.-F.; Li, B.-Z.; Li, Z.; Liu, P.; Wang, Y.; Tang, Q.; Ding, J.; Jia, Y.; Chen, Z.; Li, L.; Sun, Y.; Li, X.; Dai, Q.; Song, C.-X.; Zhang, K.; He, C.; Xu, G.-L., Tet-mediated

Formation of 5-Carboxylcytosine and Its Excision by TDG in Mammalian DNA. *Science*, **2011**, *333*, 1303-1307.

229. Pfaffeneder, T.; Hackner, B.; Truss, M.; Munzel, M.; Muller, M.; Deiml, C.; Hagemeier, C.; Carell, T., The Discovery of 5-Formylcytosine in Embryonic Stem Cell DNA. *Angew. Chem., Int. Ed.*, **2011**, *50*, 7008 - 7012.
230. Liu, S.; Wang, J.; Su, Y.; Guerrero, C.; Zeng, Y.; Mitra, D.; Brooks, P. J.; Fisher, D. E.; Song, H.; Wang, Y., Quantitative Assessment of Tet-induced Oxidation Products of 5-Methylcytosine in Cellular and Tissue DNA. *Nucleic Acids Res.*, **2013**, *41*, 6421-6429.
231. Sun, Z.; Dai, N.; Borgaro, Janine G.; Quimby, A.; Sun, D.; Corrêa, Ivan R., Jr.; Zheng, Y.; Zhu, Z.; Guan, S., A Sensitive Approach to Map Genome-wide 5-Hydroxy methylcytosine and 5-Formylcytosine at Single-base Resolution. *Mol. Cell*, **2015**, *57*, 750-761.
232. Shen, L.; Wu, H.; Diep, D.; Yamaguchi, S.; D'Alessio, Ana C.; Fung, H.-L.; Zhang, K.; Zhang, Y., Genome-wide Analysis Reveals TET- and TDG-dependent 5-Methyl-cytosine Oxidation Dynamics. *Cell*, **2013**, *153*, 692-706.
233. Song, C.-X.; Szulwach, Keith E.; Dai, Q.; Fu, Y.; Mao, S.-Q.; Lin, L.; Street, C.; Li, Y.; Poidevin, M.; Wu, H.; Gao, J.; Liu, P.; Li, L.; Xu, G.-L.; Jin, P.; He, C., Genome-wide Profiling of 5-Formylcytosine Reveals Its Roles in Epigenetic Priming. *Cell*, **2013**, *153*, 678-691.
234. Bachman, M.; Uribe-Lewis, S.; Yang, X.; Burgess, H. E.; Iurlaro, M.; Reik, W.; Murrell, A.; Balasubramanian, S., 5-Formylcytosine can be a Stable DNA Modification in Mammals. *Nat. Chem. Biol.*, **2015**, *11*, 555-557.
235. Maiti, A.; Drohat, A. C., Thymine DNA Glycosylase can Rapidly Excise 5-Formyl-cytosine and 5-Carboxylcytosine. *J. Biol. Chem.*, **2011**, *286*, 35334-35338.
236. Cortazar, D.; Kunz, C.; Selfridge, J.; Lettieri, T.; Saito, Y.; MacDougall, E.; Wirz, A.; Schuermann, D.; Jacobs, A. L.; Siegrist, F.; Steinacher, R.; Jiricny, J.; Bird, A.; Schar, P., Embryonic Lethal Phenotype Reveals a Function of TDG in Maintaining Epigenetic Stability. *Nature*, **2011**, *470*, 419-423.
237. Nabel, C. S.; Jia, H.; Ye, Y.; Shen, L.; Goldschmidt, H. L.; Stivers, J. T.; Zhang, Y.; Kohli, R. M., AID/APOBEC Deaminases Disfavor Modified Cytosines Implicated in DNA Demethylation. *Nat. Chem. Biol.*, **2012**, *8*, 751-758.
238. Schiesser, S.; Hackner, B.; Pfaffeneder, T.; Müller, M.; Hagemeier, C.; Truss, M.; Carell, T., Mechanism and Stem-cell Activity of 5-Carboxycytosine Decarboxylation Determined by Isotope Tracing. *Angew. Chem., Int. Ed.*, **2012**, *51*, 6516-6520.
239. Schiesser, S.; Pfaffeneder, T.; Sadeghian, K.; Hackner, B.; Steigenberger, B.; Schröder, A. S.; Steinbacher, J.; Kashiwazaki, G.; Höfner, G.; Wanner, K. T.; Ochsenfeld, C.; Carell, T., Deamination, Oxidation, and C-C Bond Cleavage Reactivity of 5-Hydroxymethylcytosine, 5-Formylcytosine, and 5-Carboxycytosine. *J. Am. Chem. Soc.*, **2013**, *135*, 14593-14599.
240. Tanzi, R.; Watkins, P. C.; Ottina, K.; Wallace, M. R.; Sakaguchi, A.; Young, A. B.; Shoulson, I.; Bonilla, E.; Martin, J., A Polymorphic DNA Marker Genetically Linked

to Huntingtons Disease. *Nature*, **1983**, *306*, 234-238.

241. Venter, J. C.; Smith, H. O.; Hood, L., A New Strategy for Genome Sequencing. *Nature*, **1996**, *381*, 364-366.
242. Maxam, A. M.; Gilbert, W., A New Method for Sequencing DNA. *Proc. Natl. Acad. Sci. USA*, **1977**, *74*, 560-564.
243. Lawley, P. D.; Brookes, P., Further Studies on the Alkylation of Nucleic Acids and their Consitituent Nucleotides. *Biochem. J.*, **1963**, *89*, 127-138.
244. Temperli, A.; Türler, H.; Rüst, P.; Danon, A.; Chargaff, E., Studies on the Nucleotide Arrangement in Deoxyribonucleic Acids IX. Selective Degradation of Pyrimidine Deoxyribonucleotides. *Biochim. Biophys. Acta*, **1964**, *91*, 462-476.
245. Sanger, F.; Coulson, A. R., A Rapid Method for Determining Sequences in DNA by Primed Synthesis with DNA Polymerase. *J. Mol. Biol.*, **1975**, *94*, 441-448.
246. Sanger, F.; Nicklen, S.; Coulson, A. R., DNA Sequencing with Chain-terminating Inhibitors. *Proc. Natl. Acad. Sci. USA*, **1977**, *74*, 5463-5467.
247. Smith, L. M.; Fung, S.; Hunkapiller, M. W.; Hunkapiller, T. J.; Hood, L. E., The Synthesis of Oligonucleotides Containing an Aliphatic Amino Group at the 5′ Terminus: Synthesis of Fluorescent DNA Primers for Use in DNA Sequence Analysis. *Nucleic Acids Res.*, **1985**, *13*, 2399-2412.
248. Smith, L. M.; Sanders, J. Z.; Kaiser, R. J.; Hughes, P.; Dodd, C.; Connell, C. R.; Heiner, C.; Kent, S. B. H.; Hood, L. E., Fluorescence Detection in Automated DNA Sequence Analysis. *Nature*, **1986**, *321*, 674-679.
249. Craig Chinault, A.; Carbon, J., Overlap Hybridization Screening: Isolation and Characterization of Overlapping DNA Fragments Surrounding the Leu2 Gene on Yeast Chromosome III. *Gene*, **1979**, *5*, 111-126.
250. Staden, R., A Strategy of DNA Sequencing Employing Computer Programs. *Nucleic Acids Res.*, **1979**, *6*, 2601-2610.
251. Anderson, S., Shotgun DNA Sequencing using Cloned DNase I-generated Fragments. *Nucleic Acids Res.*, **1981**, *9*, 3015-3027.
252. Consortium, I. H. G. S., Initial Sequencing and Analysis of the Human Genome. *Nature*, **2001**, *409*, 860-921.
253. Consortium, M. G. S., Initial Sequencing and Comparative Analysis of the Mouse Genome. *Nature*, **2002**, *420*, 520-562.
254. Wang, T.; Hong, T.; Tang, T.; Zhai, Q.; Xing, X.; Mao, W.; Zheng, X.; Xu, L.; Wu, J.; Weng, X.; Wang, S.; Tian, T.; Yuan, B.; Huang, B.; Zhuang, L.; Zhou, X., Application of *N*-Halogeno-*N*-sodiobenzenesulfonamide Reagents to the Selective Detection of 5-Methylcytosine in DNA Sequences. *J. Am. Chem. Soc.*, **2013**, *135*, 1240-1243.
255. Münzel, M.; Lercher, L.; Müller, M.; Carell, T., Chemical Discrimination between dC and 5MedC via their Hydroxylamine Adducts. *Nucleic Acids Res.*, **2010**, *38*, e192.
256. Huang, R.; Wang, J.; Mao, W.; Fu, B.; Xing, X.; Guo, G.; Zhou, X., A Novel Combined Bisulfite UDG Assay for Selective 5-Methylcytosine Detection. *Talanta*, **2013**, *117*, 445-448.
257. Tian, T.; Zhang, X.; Fu, B.; Long, Y.; Peng, S.; Wang, S.; Zhou, X.; Zhou, X.,

Systematic Investigation of DNAs with Modified Cytosines under Hot Alkali Treatment. *Chem. Commun.*, **2013**, *49*, 9968-9970.

258. Ko, M.; Huang, Y.; Jankowska, A. M.; Pape, U. J.; Tahiliani, M.; Bandukwala, H. S.; An, J.; Lamperti, E. D.; Koh, K. P.; Ganetzky, R.; Liu, X. S.; Aravind, L.; Agarwal, S.; Maciejewski, J. P.; Rao, A., Impaired Hydroxylation of 5-Methylcytosine in Myeloid Cancers with Mutant TET2. *Nature*, **2010**, *468*, 839-843.
259. Szwagierczak, A.; Bultmann, S.; Schmidt, C. S.; Spada, F.; Leonhardt, H., Sensitive Enzymatic Quantification of 5-Hydroxymethylcytosine in Genomic DNA. *Nucleic Acids Res.*, **2010**, *38*, e181-e181.
260. Ficz, G.; Branco, M. R.; Seisenberger, S.; Santos, F.; Krueger, F.; Hore, T. A.; Marques, C. J.; Andrews, S.; Reik, W., Dynamic Regulation of 5-Hydroxymethylcytosine in Mouse ES Cells and during Differentiation. *Nature*, **2011**, *473*, 398-402.
261. Kinney, S. M.; Chin, H. G.; Vaisvila, R.; Bitinaite, J.; Zheng, Y.; Estève, P.-O.; Feng, S.; Stroud, H.; Jacobsen, S. E.; Pradhan, S., Tissue-specific Distribution and Dynamic Changes of 5-Hydroxymethylcytosine in Mammalian Genomes. *J. Biol. Chem.*, **2011**, *286*, 24685-24693.
262. Song, C.-X.; Szulwach, K. E.; Fu, Y.; Dai, Q.; Yi, C.; Li, X.; Li, Y.; Chen, C.-H.; Zhang, W.; Jian, X.; Wang, J.; Zhang, L.; Looney, T. J.; Zhang, B.; Godley, L. A.; Hicks, L. M.; Lahn, B. T.; Jin, P.; He, C., Selective Chemical Labeling Reveals the Genome-wide Distribution of 5-Hydroxymethylcytosine. *Nat. Biotech.*, **2011**, *29*, 68-72.
263. Raiber, E.-A.; Beraldi, D.; Ficz, G.; Burgess, H.; Branco, M.; Murat, P.; Oxley, D.; Booth, M.; Reik, W.; Balasubramanian, S., Genome-wide Distribution of 5-Formylcytosine in Embryonic Stem Cells is Associated with Transcription and Depends on Thymine DNA Glycosylase. *Genome Biol.*, **2012**, *13*, R69.
264. Dai, Q.; He, C., Syntheses of 5-Formyl- and 5-Carboxyl-dC Containing DNA Oligos as Potential Oxidation Products of 5-Hydroxymethylcytosine in DNA. *Org. Lett.*, **2011**, *13*, 3446-3449.
265. Xia, B.; Han, D.; Lu, X.; Sun, Z.; Zhou, A.; Yin, Q.; Zeng, H.; Liu, M.; Jiang, X.; Xie, W.; He, C.; Yi, C., Bisulfite-free, Base-resolution Analysis of 5-Formylcytosine at the Genome Scale. *Nat. Meth.*, **2015**, *12*, 1047-1050.
266. Shapiro, R.; Servis, R. E.; Welcher, M., Reactions of Uracil and Cytosine Derivatives with Sodium Bisulfite. *J. Am. Chem. Soc.*, **1970**, *92*, 422-424.
267. Shapiro, R.; Braverman, B.; Louis, J. B.; Servis, R. E., Nucleic Acid Reactivity and Conformation: II. Reaction of Cytosine and Uracil with Sodium Bisulfite. *J. Biol. Chem.*, **1973**, *248*, 4060-4064.
268. Booth, M. J.; Marsico, G.; Bachman, M.; Beraldi, D.; Balasubramanian, S., Quantitative Sequencing of 5-Formylcytosine in DNA at Single-Base Resolution. *Nat. Chem.*, **2014**, *6*, 435-440.
269. Meissner, A.; Gnirke, A.; Bell, G. W.; Ramsahoye, B.; Lander, E. S.; Jaenisch, R., Reduced Representation Bisulfite Sequencing for Comparative High-resolution DNA Methylation Analysis. *Nucleic Acids Res.*, **2005**, *33*, 5868-5877.
270. Meissner, A.; Mikkelsen, T. S.; Gu, H.; Wernig, M.; Hanna, J.; Sivachenko, A.; Zhang,

X.; Bernstein, B. E.; Nusbaum, C.; Jaffe, D. B.; Gnirke, A.; Jaenisch, R.; Lander, E. S., Genome-scale DNA Methylation Maps of Pluripotent and Differentiated Cells. *Nature*, **2008**, *454*, 766-770.

271. Lu, X.; Song, C.-X.; Szulwach, K.; Wang, Z.; Weidenbacher, P.; Jin, P.; He, C., Chemical Modification-Assisted Bisulfite Sequencing (CAB-Seq) for 5-Carboxylcytosine Detection in DNA. *J. Am. Chem. Soc.*, **2013**, *135*, 9315-9317.
272. Wang, R. Y.-H.; Gehrke, C. W.; Ehrlich, M., Comparison of Bisulfite Modification of 5-Methyldeoxycytidine and Deoxycytidine Residues. *Nucleic Acids Res.*, **1980**, *8*, 4777-4790.
273. Grunau, C.; Clark, S. J.; Rosenthal, A., Bisulfite Genomic Sequencing: Systematic Investigation of Critical Experimental Parameters. *Nucleic Acids Res.*, **2001**, *29*, e65.
274. Tanaka, K.; Okamoto, A., Degradation of DNA by Bisulfite Treatment. *Bioorg. Med. Chem. Lett.*, **2007**, *17*, 1912-1915.
275. Rivera, Chloe M.; Ren, B., Mapping Human Epigenomes. *Cell*, **2013**, *155*, 39-55.
276. Nomura, A.; Sugizaki, K.; Yanagisawa, H.; Okamoto, A., Discrimination Between 5-Hydroxymethylcytosine and 5-Methylcytosine by a Chemically Designed Peptide. *Chem. Commun.*, **2011**, *47*, 8277-8279.
277. Duprey, J.-L. H. A.; Bullen, G. A.; Zhao, Z.-y.; Bassani, D. M.; Peacock, A. F. A.; Wilkie, J.; Tucker, J. H. R., Single Site Discrimination of Cytosine, 5-Methylcytosine, and 5-Hydroxymethylcytosine in Target DNA Using Anthracene-tagged Fluorescent Probes. *ACS Chem. Bio.*, **2015**.
278. Flusberg, B. A.; Webster, D. R.; Lee, J. H.; Travers, K. J.; Olivares, E. C.; Clark, T. A.; Korlach, J.; Turner, S. W., Direct Detection of DNA Methylation during Single-molecule, Real-time Sequencing. *Nat. Methods*, **2010**, *7*, 461-465.
279. Clark, T.; Lu, X.; Luong, K.; Dai, Q.; Boitano, M.; Turner, S.; He, C.; Korlach, J., Enhanced 5-Methylcytosine Detection in Single-molecule, Real-time Sequencing via Tet1 Oxidation. *BMC Biol.*, **2013**, *11*, 4.
280. Deamer, D. W.; Branton, D., Characterization of Nucleic Acids by Nanopore Analysis. *Acc. Chem. Res.*, **2002**, *35*, 817-825.
281. Kasianowicz, J. J.; Brandin, E.; Branton, D.; Deamer, D. W., Characterization of Individual Polynucleotide Molecules using a Membrane Channel. *Proc. Natl. Acad. Sci. USA*, **1996**, *93*, 13770-13773.
282. Cherf, G. M.; Lieberman, K. R.; Rashid, H.; Lam, C. E.; Karplus, K.; Akeson, M., Automated Forward and Reverse Ratcheting of DNA in a Nanopore at 5-A Precision. *Nat. Biotech.*, **2012**, *30*, 344-348.
283. Manrao, E. A.; Derrington, I. M.; Laszlo, A. H.; Langford, K. W.; Hopper, M. K.; Gillgren, N.; Pavlenok, M.; Niederweis, M.; Gundlach, J. H., Reading DNA at Single-nucleotide Resolution with a Mutant MspA Nanopore and Φ29 DNA Polymerase. *Nat. Biotech.*, **2012**, *30*, 349-353.
284. Lieberman, K. R.; Cherf, G. M.; Doody, M. J.; Olasagasti, F.; Kolodji, Y.; Akeson, M., Processive Replication of Single DNA Molecules in a Nanopore Catalyzed by Φ29 DNA Polymerase. *J. Am. Chem. Soc.*, **2010**, *132*, 17961-17972.

285. Wallace, E. V. B.; Stoddart, D.; Heron, A. J.; Mikhailova, E.; Maglia, G.; Donohoe, T. J.; Bayley, H., Identification of Epigenetic DNA Modifications with a Protein Nanopore. *Chem. Commun.*, **2010**, *46*, 8195-8197.

286. Laszlo, A. H.; Derrington, I. M.; Brinkerhoff, H.; Langford, K. W.; Nova, I. C.; Samson, J. M.; Bartlett, J. J.; Pavlenok, M.; Gundlach, J. H., Detection and Mapping of 5-Methylcytosine and 5-Hydroxymethylcytosine with Nanopore MspA. *Proc. Natl. Acad. Sci. USA*, **2013**, *110*, 18904-18909.

287. Schreiber, J.; Wescoe, Z. L.; Abu-Shumays, R.; Vivian, J. T.; Baatar, B.; Karplus, K.; Akeson, M., Error Rates for Nanopore Discrimination among Cytosine, Methylcytosine, and Hydroxymethylcytosine along Individual DNA Strands. *Proc. Natl. Acad. Sci. USA*, **2013**, *110*, 18910-18915.

288. Wescoe, Z. L.; Schreiber, J.; Akeson, M., Nanopores Discriminate among Five C5-Cytosine Variants in DNA. *J. Am. Chem. Soc.*, **2014**, *136*, 16582-16587.

289. Venkatesan, B. M.; Bashir, R., Nanopore Sensors for Nucleic Acid Analysis. *Nat. Nano*, **2011**, *6*, 615-624.

290. Feng, J.; Liu, K.; Bulushev, R. D.; Khlybov, S.; Dumcenco, D.; Kis, A.; Radenovic, A., Identification of Single Nucleotides in MoS_2 Nanopores. *Nat. Nano*, **2015**, *10*, 1070-1076.

291. Eid, J.; Fehr, A.; Gray, J.; Luong, K.; Lyle, J.; Otto, G.; Peluso, P.; Rank, D.; Baybayan, P.; Bettman, B.; Bibillo, A.; Bjornson, K.; Chaudhuri, B.; Christians, F.; Cicero, R.; Clark, S.; Dalal, R.; deWinter, A.; Dixon, J.; Foquet, M.; Gaertner, A.; Hardenbol, P.; Heiner, C.; Hester, K.; Holden, D.; Kearns, G.; Kong, X.; Kuse, R.; Lacroix, Y.; Lin, S.; Lundquist, P.; Ma, C.; Marks, P.; Maxham, M.; Murphy, D.; Park, I.; Pham, T.; Phillips, M.; Roy, J.; Sebra, R.; Shen, G.; Sorenson, J.; Tomaney, A.; Travers, K.; Trulson, M.; Vieceli, J.; Wegener, J.; Wu, D.; Yang, A.; Zaccarin, D.; Zhao, P.; Zhong, F.; Korlach, J.; Turner, S., Real-time DNA Sequencing from Single Polymerase Molecules. *Science*, **2009**, *323*, 133-138.

292. Liu, X.; Li, J.; Dong, J.; Hu, C.; Gong, W.; Wang, J., Genetic Incorporation of a Metal-chelating Amino Acid as a Probe for Protein Electron Transfer. *Angew. Chem., Int. Ed.*, **2012**, *51*, 10261-10265.

293. Hoffer, M., α-Thymidin. *Chem. Ber.*, **1960**, *93*, 2777-2781.

294. Gu, W.; Wang, S., Synthesis and Antimicrobial Activities of Novel 1H-dibenzo[a,c] carbazoles from Dehydroabietic Acid. *Eur. J. Med. Chem.*, **2010**, *45*, 4692-4696.

295. de Brito, A. F.; Martins, J. L. R.; Fajemiroye, J. O.; Galdino, P. M.; De Lima, T. C. M.; Menegatti, R.; Costa, E. A., Central Pharmacological Activity of a New Piperazine Derivative: 4-(1-Phenyl-1*H*-pyrazol-4-ylmethyl)-piperazine-1-carboxylic Acid Ethyl ester. *Life Sci.*, **2012**, *90*, 910-916.

296. Sousa, C.; Silva, P. J., BBr_3-Assisted Cleavage of Most Ethers Does Not Follow the Commonly Assumed Mechanism. *Eur. J. Org. Chem.*, **2013**, *2013*, 5195-5199.

297. Batchu, H.; Bhattacharyya, S.; Kant, R.; Batra, S., Palladium-catalyzed Chelation-assisted Regioselective Oxidative Dehydrogenative Homocoupling/ *Ortho*-Hydroxylation in *N*-Phenylpyrazoles. *J. Org. Chem.*, **2015**, *80*, 7360-7374.

298. Corey, E. J.; Posner, G. H., Selective Formation of Carbon-carbon Bonds between Unlike Groups using Organocopper Reagents. *J. Am. Chem. Soc.*, **1967**, *89*, 3911-3912.
299. House, H. O.; Respess, W. L.; Whitesides, G. M., The Chemistry of Carbanions. XII. The Role of Copper in the Conjugate Addition of Organometallic Reagents1. *J. Org. Chem.*, **1966**, *31*, 3128-3141.
300. Gilman, H.; Jones, R. G.; Woods, L. A., The Preparation of Methylcopper and some Observations on the Decomposition of Organocopper Compounds. *J. Org. Chem.*, **1952**, *17*, 1630-1634.
301. Lorenzen, N. P.; Weiss, E., Synthesis and Structure of a Dimeric Lithium Diphenylcuprate: $[\{Li(OEt)_2\}(CuPh_2)]_2$. *Angew. Chem., Int. Ed.*, **1990**, *29*, 300-302.
302. Reddy, M. P.; Farooqui, F.; Hanna, N. B., Methylamine Deprotection Provides Increased Yield of Oligoribonucleotides. *Tetrahedron Lett.*, **1995**, *36*, 8929-8932.
303. Larrow, J. F.; Jacobsen, E. N.; Gao, Y.; Hong, Y.; Nie, X.; Zepp, C. M., A Practical Method for the Large-scale Preparation of [*N*,*N*'-Bis(3,5-di-tertbutylsalicylidene)-1,2-cyclohexanediaminato(2-)]manganese(III) chloride, a Highly Enantioselective Epoxi- dation Catalyst. *J. Org. Chem.*, **1994**, *59*, 1939-1942.
304. Betz, K.; Malyshev, D. A.; Lavergne, T.; Welte, W.; Diederichs, K.; Romesberg, F. E.; Marx, A., Structural Insights into DNA Replication without Hydrogen Bonds. *J. Am. Chem. Soc.*, **2013**, *135*, 18637-18643.
305. Agarwal, K.; Sharma, A.; Talukder, G., Effects of Copper on Mammalian Cell Components. *Chem. Biol. Interact.*, **1989**, *69*, 1-16.
306. Valko, M.; Rhodes, C. J.; Moncol, J.; Izakovic, M.; Mazur, M., Free Radicals, Metals and Antioxidants in Oxidative Stress-induced Cancer. *Chem. Biol. Interact.*, **2006**, *160*, 1-40.
307. Hureau, C.; Faller, P., Aβ-mediated ROS Production by Cu Ions: Structural Insights, Mechanisms and Relevance to Alzheimer's Disease. *Biochimie*, **2009**, *91*, 1212-1217.
308. Caton-Williams, J.; Smith, M.; Carrasco, N.; Huang, Z., Protection-free One-pot Synthesis of 2′-Deoxynucleoside 5′-Triphosphates and DNA Polymerization. *Org. Lett.*, **2011**, *13*, 4156-4159.
309. Burgess, K.; Cook, D., Syntheses of Nucleoside Triphosphates. *Chem. Rev.*, **2000**, *100*, 2047-2060.
310. Gardner, A. F.; Jack, W. E., Determinants of Nucleotide Sugar Recognition in an Archaeon DNA Polymerase. *Nucleic Acids Res.*, **1999**, *27*, 2545-2553.
311. Gardner, A. F.; Jack, W. E., Acyclic and Dideoxy Terminator Preferences Denote Divergent Sugar Recognition by Archaeon and Taq DNA Polymerases. *Nucleic Acids Res.*, **2002**, *30*, 605-613.
312. Horhota, A.; Zou, K.; Ichida, J. K.; Yu, B.; McLaughlin, L. W.; Szostak, J. W.; Chaput, J. C., Kinetic Analysis of an Efficient DNA-dependent TNA Polymerase. *J. Am. Chem. Soc.*, **2005**, *127*, 7427-7434.
313. Ichida, J. K.; Zou, K.; Horhota, A.; Yu, B.; McLaughlin, L. W.; Szostak, J. W., An in vitro Selection System for TNA. *J. Am. Chem. Soc.*, **2005**, *127*, 2802-2803.
314. Roelfes, G.; Feringa, B. L., DNA-based Asymmetric Catalysis. *Angew. Chem., Int. Ed.*,

2005, *44*, 3230-3232.

315. Park, S.; Ikehata, K.; Watabe, R.; Hidaka, Y.; Rajendran, A.; Sugiyama, H., Deciphering DNA-based Asymmetric Catalysis Through Intramolecular Friedel-Crafts Alkylations. *Chem. Commun.*, **2012**, *48*, 10398-10400.
316. Park, S.; Zheng, L.; Kumakiri, S.; Sakashita, S.; Otomo, H.; Ikehata, K.; Sugiyama, H., Development of DNA-Based Hybrid Catalysts through Direct Ligand Incorporation: Toward Understanding of DNA-Based Asymmetric Catalysis. *ACS Catal.*, **2014**, *4*, 4070-4073.
317. Wang, C.; Jia, G.; Zhou, J.; Li, Y.; Liu, Y.; Lu, S.; Li, C., Enantioselective Diels-Alder Reactions with G-Quadruplex DNA-based Catalysts. *Angew. Chem., Int. Ed.*, **2012**, *51*, 9352-9355.
318. Otto, S.; Bertoncin, F.; Engberts, J. B. F. N., Lewis Acid Catalysis of a Diels−Alder Reaction in Water. *J. Am. Chem. Soc.*, **1996**, *118*, 7702-7707.
319. Kim, S.; Misra, A., SNP Genotyping: Technologies and Biomedical Applications. *Annu. Rev. Biomed. Eng.*, **2007**, *9*, 289-320.
320. Li, P.; He, H.; Wang, Z.; Feng, M.; Jin, H.; Wu, Y.; Zhang, L.; Zhang, L.; Tang, X., Sensitive Detection of Single-nucleotide Mutation in the BRAF Mutation Site (V600E) of Human Melanoma Using Phosphate-pyrene-labeled DNA Probes. *Anal. Chem.*, **2016**, *88*, 883-889.
321. Banér, J.; Isaksson, A.; Waldenström, E.; Jarvius, J.; Landegren, U.; Nilsson, M., Parallel Gene Analysis with Allele-specific Padlock Probes and Tag Microarrays. *Nucleic Acids Res.*, **2003**, *31*, e103.
322. Felsenfeld, G.; Groudine, M., Controlling the Double Helix. *Nature*, **2003**, *421*, 448-453.
323. Guo, P.; Yan, S.; Hu, J.; Xing, X.; Wang, C.; Xu, X.; Qiu, X.; Ma, W.; Lu, C.; Weng, X.; Zhou, X., Selective Detection of 5-Formyl-2′-deoxycytidine in DNA Using a Fluorogenic Hydroxylamine Reagent. *Org. Lett.*, **2013**, *15*, 3266-3269.
324. Xu, L.; Chen, Y.-C.; Nakajima, S.; Chong, J.; Wang, L.; Lan, L.; Zhang, C.; Wang, D., A Chemical Probe Targets DNA 5-Formylcytosine Sites and Inhibits TDG Excision, Polymerases Bypass, and Gene Expression. *Chem. Sci.*, **2014**, *5*, 567-574.
325. Hu, J.; Xing, X.; Xu, X.; Wu, F.; Guo, P.; Yan, S.; Xu, Z.; Xu, J.; Weng, X.; Zhou, X., Selective Chemical Labelling of 5-Formylcytosine in DNA by Fluorescent Dyes. *Chem. Eur. J.*, **2013**, *19*, 5836-5840.
326. Zhang, Z.; Smith, B. A. C.; Wang, L.; Brock, A.; Cho, C.; Schultz, P. G., A New Strategy for the Site-Specific Modification of Proteins in vivo. *Biochemistry*, **2003**, *42*, 6735-6746.
327. Wang, L.; Zhang, Z.; Brock, A.; Schultz, P. G., Addition of the Keto Functional Group to the Genetic Code of Escherichia coli. *Proc. Natl. Acad. Sci. USA*, **2003**, *100*, 56-61.
328. Chen, I.; Howarth, M.; Lin, W.; Ting, A. Y., Site-specific Labeling of Cell Surface Proteins with Biophysical Probes using Biotin Ligase. *Nat. Methods*, **2005**, *2*, 99-104.
329. Taylor, M. T.; Blackman, M. L.; Dmitrenko, O.; Fox, J. M., Design and Synthesis of Highly Reactive Dienophiles for the Tetrazine-trans-cyclooctene Ligation. *J. Am. Chem.*

Soc., **2011**, *133*, 9646-9649.

330. Patterson, D. M.; Nazarova, L. A.; Xie, B.; Kamber, D. N.; Prescher, J. A., Functionalized Cyclopropenes As Bioorthogonal Chemical Reporters. *J. Am. Chem. Soc.*, **2012**, *134*, 18638-18643.
331. Liang, Y.; Mackey, J. L.; Lopez, S. A.; Liu, F.; Houk, K. N., Control and Design of Mutual Orthogonality in Bioorthogonal Cycloadditions. *J. Am. Chem. Soc.*, **2012**, *134*, 17904-17907.
332. Noda, H.; Erős, G.; Bode, J. W., Rapid Ligations with Equimolar Reactants in Water with the Potassium Acyltrifluoroborate (KAT) Amide Formation. *J. Am. Chem. Soc.*, **2014**, *136*, 5611-5614.
333. Saito, F.; Noda, H.; Bode, J. W., Critical Evaluation and Rate Constants of Chemoselective Ligation Reactions for Stoichiometric Conjugations in Water. *ACS Chem. Bio.*, **2015**, *10*, 1026-1033.
334. Dirksen, A.; Hackeng, T. M.; Dawson, P. E., Nucleophilic Catalysis of Oxime Ligation. *Angew. Chem., Int. Ed.*, **2006**, *45*, 7581-7584.
335. Duprey, J.-L. H. A.; Takezawa, Y.; Shionoya, M., Metal-locked DNA Three-way Junction. *Angew. Chem., Int. Ed.*, **2013**, *52*, 1212-1216.
336. Guan, L.; van der Heijden, G. W.; Bortvin, A.; Greenberg, M. M., Intracellular Detection of Cytosine Incorporation in Genomic DNA by Using 5-Ethynyl-2′-deoxycytidine. *ChemBioChem*, **2011**, *12*, 2184-2190.
337. Obeid, S.; Bußkamp, H.; Welte, W.; Diederichs, K.; Marx, A., Interactions of Non-polar and "Clickable" Nucleotides in the Confines of a DNA Polymerase Active Site. *Chem. Commun.*, **2012**, *48*, 8320-8322.
338. Griffith, J. A.; Withers, J. M.; Martin, D. J.; Rowlands, G. J.; Filichev, V. V., Ligand Assembly and Chirality Transfer Guided by DNA Modified with Enantiomerically Pure [2.2]Paracyclophanes. *RSC Adv.*, **2013**, *3*, 9373-9380.
339. Wenge, U.; Ehrenschwender, T.; Wagenknecht, H.-A., Synthesis of 2′-*O*-Propargyl Nucleoside Triphosphates for Enzymatic Oligonucleotide Preparation and “Click” Modification of DNA with Nile Red as Fluorescent Probe. *Bioconjugate Chem.*, **2013**, *24*, 301-304.
340. Wendeler, M.; Grinberg, L.; Wang, X.; Dawson, P. E.; Baca, M., Enhanced Catalysis of Oxime-based Bioconjugations by Substituted Anilines. *Bioconjugate Chem.*, **2014**, *25*, 93-101.
341. Camiener, G. W., Studies of the Enzymatic Deamination of Ara-cytidine-V: Inhibition in vitro and in vivo by Tetrahydrouridine and Other Reduced Pyrimidine Nucleosides. *Biochem. Pharmacol.*, **1968**, *17*, 1981-1991.
342. Green, N. M., Avidin. 3. The Nature of the Biotin-binding Site. *Biochem. J.*, **1963**, *89*, 599-609.
343. Green, N. M., Avidin. 1. The Use of ^{14}C Biotin for Kinetic Studies and for Assay. *Biochem. J.*, **1963**, *89*, 585-591.
344. Nilsson, M.; Malmgren, H.; Samiotaki, M.; Kwiatkowski, M.; Chowdhary, B.; Landegren, U., Padlock Probes: Circularizing Oligonucleotides for Localized DNA

Detection. *Science*, **1994**, *265*, 2085-2088.

345. Liu, S.; Wang, C.; Zhang, C.; Wang, Y.; Tang, B., Label-free and Ultrasensitive Electrochemical Detection of Nucleic Acids Based on Autocatalytic and Exonuclease III-assisted Target Recycling Strategy. *Anal. Chem.*, **2013**, *85*, 2282-2288.
346. Untergasser, A.; Nijveen, H.; Rao, X.; Bisseling, T.; Geurts, R.; Leunissen, J. A. M., Primer3Plus, An Enhanced Web Interface to Primer3. *Nucleic Acids Res.*, **2007**, *35*, W71-W74.
347. van Eijk, M. J. T.; Broekhof, J. L. N.; van der Poel, H. J. A.; Hogers, R. C. J.; Schneiders, H.; Kamerbeek, J.; Verstege, E.; van Aart, J. W.; Geerlings, H.; Buntjer, J. B.; van Oeveren, A. J.; Vos, P., SNPWave^{TM}: A Flexible Multiplexed SNP Genotyping Technology. *Nucleic Acids Res.*, **2004**, *32*, e47.
348. Hindson, B. J.; Ness, K. D.; Masquelier, D. A.; Belgrader, P.; Heredia, N. J.; Makarewicz, A. J.; Bright, I. J.; Lucero, M. Y.; Hiddessen, A. L.; Legler, T. C.; Kitano, T. K.; Hodel, M. R.; Petersen, J. F.; Wyatt, P. W.; Steenblock, E. R.; Shah, P. H.; Bousse, L. J.; Troup, C. B.; Mellen, J. C.; Wittmann, D. K.; Erndt, N. G.; Cauley, T. H.; Koehler, R. T.; So, A. P.; Dube, S.; Rose, K. A.; Montesclaros, L.; Wang, S.; Stumbo, D. P.; Hodges, S. P.; Romine, S.; Milanovich, F. P.; White, H. E.; Regan, J. F.; Karlin-Neumann, G. A.; Hindson, C. M.; Saxonov, S.; Colston, B. W., High-throughput Droplet Digital PCR System for Absolute Quantitation of DNA Copy Number. *Anal. Chem.*, **2011**, *83*, 8604-8610.
349. Han, H.-S.; Cantalupo, P. G.; Rotem, A.; Cockrell, S. K.; Carbonnaux, M.; Pipas, J. M.; Weitz, D. A., Whole-genome Sequencing of a Single Viral Species from a Highly Heterogeneous Sample. *Angew. Chem., Int. Ed.*, **2015**, *54*, 13985-13988.
350. Lubin, F. D.; Roth, T. L.; Sweatt, J. D., Epigenetic Regulation of bdnf Gene Transcription in the Consolidation of Fear Memory. *J. Neurosci.*, **2008**, *28*, 10576-10586.
351. Hardenbol, P.; Baner, J.; Jain, M.; Nilsson, M.; Namsaraev, E. A.; Karlin-Neumann, G. A.; Fakhrai-Rad, H.; Ronaghi, M.; Willis, T. D.; Landegren, U.; Davis, R. W., Multiplexed Genotyping with Sequence-tagged Molecular Inversion Probes. *Nat. Biotech.*, **2003**, *21*, 673-678.
352. Schröder, A. S.; Steinbacher, J.; Steigenberger, B.; Gnerlich, F. A.; Schiesser, S.; Pfaffeneder, T.; Carell, T., Synthesis of a DNA Promoter Segment Containing All Four Epigenetic Nucleosides: 5-Methyl-, 5-Hydroxymethyl-, 5-Formyl-, and 5-Carboxy-2′-Deoxycytidine. *Angew. Chem., Int. Ed.*, **2014**, *53*, 315-318.

Postscript

Frank Herbert, an American writer, said “There is no real ending. It’s just the place where you stop the story.” So, here is the best place to close this thesis.

Four years ago, it was the first time for the author, a little boy, to go abroad. He was alone, with two pieces of baggage. Facing a new and fantastic world, he did not know what would be the future. He had imagined the most difficult scenarios. There is no way back.

Since then, he learned to manage everything himself in this strange land. He tried to rent a flat. He took the ankle surgery in the hospital. He was a Marathon runner more than a football player in the lab. He drew up the itinerary and traveled alone to see the world. He talked to himself to drive away loneliness. He went through the days when his friend slid off the mountain and passed away. He met his girlfriend from Kyoto in Munich by coincidence. He strived hard to find out where should be his next stop, and what the future will be. He followed his heart in making decisions because he knew that this is his own life. He tries to live his life to the fullest.

Little by little, he learned to be brave, to be independent, to be grateful. He thanks all the person he encountered, kind or severe. He thanks all the things happened to him, good or bad. He learned to be himself. He learned to love the world, a world of laughter and tears, a world of prosperity and poverty, a world of light and darkness. He learned that there is no definite answer for the future because the future itself means infinite possibilities.

He thanks himself sincerely, for the unparalleled courage, for the curiosity, for the enthusiasm and persistent in chasing his dreams.

Every story has an end, and every ending is just a new beginning. Now, the young man is eager for a larger world, for the future, even more than ever, as the boy four years ago, but with less fear.