



**HAL**  
open science

# Enzymatic Synthesis of Base-Modified Nucleic Acids

Marcel Hollenstein

► **To cite this version:**

Marcel Hollenstein. Enzymatic Synthesis of Base-Modified Nucleic Acids. Handbook of Chemical Biology of Nucleic Acids, Springer Nature, pp.1-39, 2023, 978-981-16-1313-5 [ebook]. 10.1007/978-981-16-1313-5\_23-1 . pasteur-04104830

**HAL Id: pasteur-04104830**

**<https://pasteur.hal.science/pasteur-04104830>**

Submitted on 24 May 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## Enzymatic Synthesis of Base-Modified Nucleic Acids

### A. Author's name and details.

Marcel Hollenstein, Institut Pasteur, Université de Paris Cité, Department of Structural Biology and Chemistry, Laboratory for Bioorganic Chemistry of Nucleic Acids, CNRS UMR3523, 28, rue du Docteur Roux, 75724 Paris Cedex 15, France

**E-mail:** [marcel.hollenstein@pasteur.fr](mailto:marcel.hollenstein@pasteur.fr)

### B. Abstract

*Nucleic acids collectively play primordial functions, some of which are still barely understood, in all living organisms. DNA and RNA oligonucleotides have also advanced as pivotal tools in numerous applications such as therapeutic technologies, diagnostics, or material sciences. Chemical modifications, particularly at the level of the nucleobases, are ubiquitous. Indeed, base-modifications in natural DNA and RNA are nearly as important as the primary sequences and are involved in numerous biological roles and functions, mainly via epigenetic regulation mechanisms. In synthetic nucleic acids, chemical modifications provide oligonucleotides with improved binding or catalytic activities as well as exogenous reactivities, all of which are largely non-accessible to canonical DNA or RNA. Nucleobase-modifications can be incorporated into oligonucleotides either by standard solid-phase synthesis or by application of a chemoenzymatic method. In this Chapter, we will describe how the combination of polymerases and nucleoside triphosphate analogs can be harnessed to synthesize oligonucleotides containing nucleobase-modifications. We will highlight the potency of this approach through the description of relevant examples including synthesis of oligonucleotides containing naturally occurring modifications and the selection of functional nucleic acids with improved properties. In a last section of this Chapter, more recent applications of this method will be discussed with an emphasis on mRNA-based vaccines and on controlled enzymatic synthesis.*

### C. Please highlight keywords (in Green) in the MS to be used in the index

Nucleobase-modified nucleotides; polymerases; epigenetic modifications; SELEX; aptamers; DNAzymes; ribozymes; de novo DNA synthesis; mRNA vaccines

### D. Introduction

Nucleic acids (DNA and RNA) are one of the most important biopolymers in living organisms since they collectively serve as storage of genetic information, regulation of gene expression, and represent key intermediates in protein synthesis. Like all biopolymers, DNA and particularly RNA are subjected to post-transcriptional modification mechanisms which introduce covalent, chemical modification patterns that significantly expand the chemical repertoire as well as the functional and structural diversity of nucleic acids well beyond to that offered by standard, canonical purine and pyrimidine containing oligonucleotides. In addition to these ubiquitous, naturally occurring alterations of the nucleosidic scaffold, there is an increasing demand for the chemical modification of RNA and DNA nucleosides, nucleotides, and oligonucleotides for numerous practical applications. This is particularly the case since nucleic acids have advanced as highly potent therapeutic agents in various forms (e.g. mRNA vaccines, antisense oligonucleotides, DNA enzymes, RNA editing tools, siRNA and miRNAs) which calls in the need for chemical modifications to improve their intrinsic biostability, target specificity, *in vivo* residence times, and promote cell penetration and delivery. Chemical modifications can be brought to any of the constituting parts of nucleotides, namely the phosphate backbone, the (deoxy-)ribose sugar, or the nucleobase. However, the emphasis of this Chapter will be on base modifications since these represent the main alterations found in natural DNA and RNA and the nucleobases represent popular sites for the introduction of modifications to improve binding, catalysis, target recognition, and duplex stabilities of nucleic acids (for more information, the interested reader is directed to references S1-S3 of the Further readings Section).

Traditionally, chemical modifications are incorporated into nucleic acids using standard automated solid-phase synthesis in which activated building blocks (i.e. phosphoramidites) are coupled to nascent oligonucleotide chains. This approach is particularly powerful for the synthesis of short, highly modified

sequences and is readily amenable to large scale production of oligonucleotides. However, this methodology is restricted to rather short oligonucleotides (< 120 nucleotides) and various functional groups such as reactive aldehydes are not compatible with the rather harsh conditions imposed by solid-phase synthesis. Polymerase-mediated synthesis represents a valuable alternative to solid-phase synthesis that grants access to long, modified oligonucleotides and the mild conditions permit the introduction of a broad array of functional groups. Besides alleviating the limitations of solid-phase synthesis, enzymatic synthesis of modified nucleic acids also enables *in vitro* Darwinian evolution experiments which yield nucleic acid based affinity ligands (aptamers) or catalysts (DNAzymes and ribozymes). In addition, polymerases typically insert hundreds of nucleotides per second on nascent chains and this by maintaining low error rates (in the to range  $10^{-7}$  to  $10^{-8}$ ) even with modified nucleotides.<sup>1</sup> Polymerase-mediated synthesis is therefore unrivaled in terms of rates and fidelity especially when compared to chemical synthesis. In this Chapter, polymerase-mediated synthesis of base-modified DNA and RNA oligonucleotides will be described. In a first section, emphasis will be laid on application of this method for the crafting of tailor-made sequences containing epigenetic marks, either with partial or complete exchange of canonical nucleotides. Moreover, the application of base-modified nucleotides in Darwinian evolution methods for the identification of potent binding sequences (aptamers) and catalysts (DNAzymes and ribozymes) will be discussed. Lastly, emerging applications of enzymatic synthesis of base-modified nucleic acids will be discussed with a particular emphasis on improving the stability and intrinsic properties of mRNA-based vaccines and on controlled enzymatic synthesis of oligonucleotides.

## E. Main Text

### 1) General concepts on polymerase-mediated synthesis of modified nucleic acids:

*Polymerases are responsible for replication of the genetic material and the concomitant transcription of this information into RNA. These enzymes collectively catalyze the same chemical reaction, namely the nucleophilic attack of an activated 3'-OH residue located on the growing DNA or RNA chain to the  $\alpha$  phosphorous atom of an incoming nucleoside triphosphate. This reaction yields a new phosphodiester linkage and leads to the release of pyrophosphate. Activation of the hydroxyl moiety and stabilization of negative charge on the different reaction partners and in the transition state implicate various divalent metal cations but the exact mechanisms and even number of metal cations varies substantially and appears to be dependent on the nature of each individual polymerase.(see references S4 and S5 of the Further readings section) Polymerases are capable of incorporating nucleotides at astounding kinetic rates and fidelities which make them highly versatile tools for many biotechnological applications. Importantly, despite their high fidelity, polymerases tolerate the presence of chemical modifications provided they are affixed at precise locations of the nucleosidic scaffold. This acceptance of chemical modifications is of biological relevance due to the numerous, naturally occurring modifications on nucleobases and sugar units in DNA and RNA. This tolerance for chemically altered nucleotides combined with progress in protein engineering which enables the identification of polymerases with broad substrate specificities also allows for facile enzymatic synthesis of modified oligonucleotides. In this Section, we will describe how polymerases can be used as tools for the incorporation of base-modified nucleotides to generate sequences equipped with naturally occurring modifications or with functional groups alien to biology.*

### 2) Natural modifications of nucleobases: synthesis and polymerase-mediated incorporation

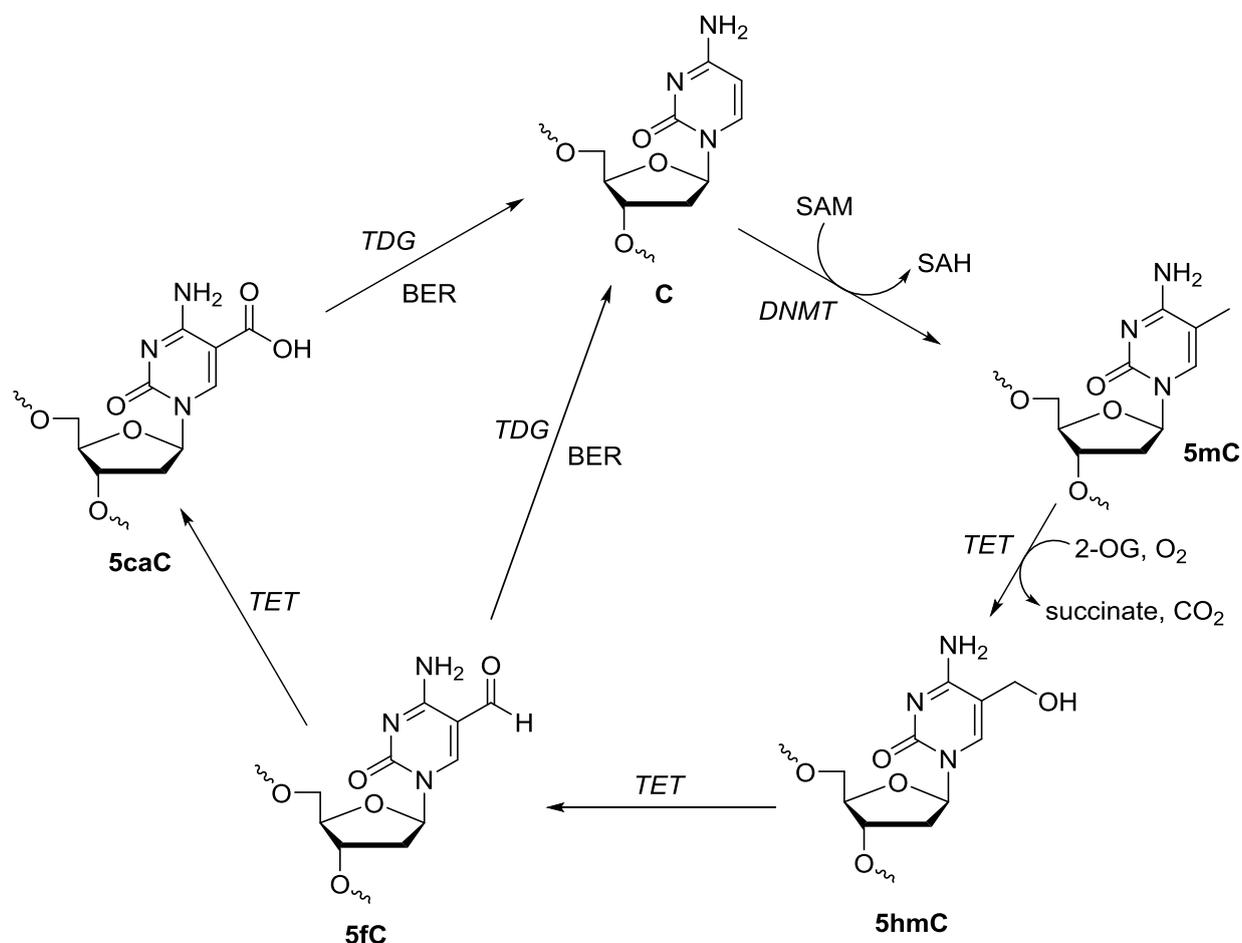
*Advances in high-throughput and next generation sequencing (NGS) technologies grant easy and affordable access to the primary sequences of entire genomes, including the human genome. Hence, these technologies have permitted to gather tremendous knowledge on genomic features and fundamental insights into the information encrypted in the primary sequences consisting of the four, canonical nucleotides(see references S3 and S6 of the Further readings section) More importantly, NGS combined with liquid chromatography-mass spectrometry (LC-MS/MS) led to the realization that nucleic acids carry a number of modifications, mainly on the nucleobases, and that these modifications play pivotal roles in numerous biological processes such as gene regulation or modulation of specific chromatin regions but also serve as structural components which directly shape and influence the secondary and tertiary structures adopted by RNA and DNA oligonucleotides. These chemical marks are collectively coined as epigenetic modifications of nucleic acids. The synthesis of nucleoside analogues bearing such epigenetic marks is of obvious importance for their identification and quantification where they can serve for instance as internal standards for LC-MS/MS analysis. However, the crafting of sequences containing these natural modifications is equally necessary in order to shed light into their biological functions but also to identify proteins that interact with such epigenetic modifications. While solid-phase DNA and RNA synthesis*

permits the incorporation of modified nucleosides into short sequences, long oligonucleotides containing epigenetic marks mimicking natural systems are only accessible by combining the corresponding nucleoside triphosphates combined with enzymatic synthesis methods such as PCR or primer extension (PEX) reactions.

#### a) Epigenetic base-modification patterns on DNA

So far, less than 20 different types of natural modification patterns have been identified in DNA, most of which are located at position C5 of pyrimidine nucleobases. This particular location of pyrimidine, also a common choice for the synthesis of nucleotide analogs for other applications, allows the introduction of chemical modifications into the major groove of duplex DNA without interfering with Watson-Crick-Franklin base pairing. In this context, the most abundant and best characterized epigenetic modifications in DNA is 5-methyl-cytosine (5mC) which stems from a direct methylation of cytosine. This DNA methylation is important for epigenetic gene regulation and expression and is involved in maintenance of cellular functions and genomic integrity particularly during mammalian embryonic development. (see references S6 and S7 of the Further readings section) Importantly, this epigenetic modification does not act alone but is involved in cross-talks with chromatin components and histone modifications.

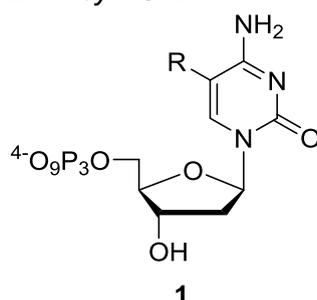
Methylation of cytosine is mainly mediated by DNA methyltransferases (DNMTs) which use S-adenosylmethionine (SAM) as a cofactor (Scheme 1). Among the known DNMTs, DNMT1 is responsible for maintaining methylation patterns during DNA replication while other enzymes including DNMT3a and DNMT3b install new methylation sites, predominantly in CG dinucleotides (CpGs) (see references S3 and S6 of the Further readings section) The reverse reaction, i.e. active, enzyme-mediated demethylation, is believed to occur by successive oxidation of the methyl moiety of 5mC to the corresponding alcohol (5hmC), aldehyde (5fC), and carboxylic acid (5caC). These oxidation events are catalyzed by enzymes called ten-eleven translocation (TET) enzymes. The oxidized products 5fC and 5caC are first excised from DNA by a DNA glycosylase such as thymine DNA glycosylase (TDG) to yield an abasic site and base excision repair (BER) then restores unmodified cytosine (Scheme 1).



**Scheme 1.** Schematic representation of methylation, oxidation and oxidative demethylation of cytosine in mammalian DNA. DNA methyltransferases (DNMT) use S-adenosylmethionine (SAM) as a cofactor to methylate C at position 5 of the nucleobase to the corresponding 5-methylC (5mC). The methylated C is subjected to a first formal oxidation step mediated by ten-eleven translocation (TET) enzymes using both oxygen and 2-oxoglutarate (2-OG) to yield 5-hydroxymethylC (5hmC). Subsequent TET-mediated, oxidation steps will yield the corresponding formyl (5fC) and carboxylic acid (5caC) modified cytosine

derivatives. Canonical C can be restored from both 5fC and 5caC following excision by thymine DNA glycosylase (TDG) and substitution by the base excision repair (BER) pathway.

Due to the relatively high abundance of 5mC (over 4% of all cytosines of the human genome are methylated) as well as the lack of clear and conclusive information on the biological significance of oxidized 5mC analogues, all the corresponding nucleoside triphosphates have been synthesized (Figure 1), and are even commercially available. A recent article reports a facile synthesis of the triphosphates of 5mC, 5fC, and 5caC using unprotected 5-methyl-cytosine as common precursor.<sup>2</sup> After application of the “one-pot, three-step” strategy developed by Ludwig (see reference S9 of the Further readings section) on 5mC, the resulting 5mCTP was further oxidized to the aldehyde by menadione-mediated photosensitization or to the carboxylic acid by treatment with TEMPO. A different approach was followed by Steigenberger et al., where the corresponding unprotected aldehyde and carboxylic acid containing nucleosides were converted directly into nucleoside triphosphates using either a two-step protocol or the Ludwig-Eckstein version of the triphosphorylation method, respectively.<sup>3</sup> Of all four modified cytosine analogues, 5hmC presents the biggest synthetic challenge due to the higher reactivity of the hydroxymethyl residue compared to the primary 5'-OH moiety on the sugar. Steigenberger et al. circumvented this differential reactivity by introducing a transient carbamate protecting group on 5hmC and the resulting triphosphate. Alternatively, Sun et al., acetyl-protected the hydroxymethyl moiety to install a monophosphate moiety on the 5'-OH position.<sup>4</sup> 5hmC monophosphate was then converted into the corresponding 5'-phosphoropiperidate derivative which readily reacts with pyrophosphate after activation with 4,5-dicyanoimidazole to yield 5hmCTP. All these modified dCTP analogues are readily accepted as substrates by the KOD XL polymerase and can be incorporated into DNA by PCR.<sup>3</sup>



**Figure 1.** Chemical structures of base-modified dCTP bearing methyl ( $R = \text{CH}_3$ ), hydroxyl ( $R = \text{CH}_2\text{OH}$ ), formyl ( $R = \text{CHO}$ ), and carboxylic acid ( $R = \text{COOH}$ ) residues.<sup>2, 3, 4</sup>

Other modification patterns beyond 5-methylated and related oxidized derivatives have been reported for the cytosine nucleobase. Amongst such modifications, methylation at the exocyclic amine resulting in the formation of the N4-methyl cytosine (4mC) analog is certainly the best known example. 4mC is relatively abundant in bacteria and has recently been reported to act as a true epigenetic mark in eukaryotic genomes. (see reference S10 of the Further readings section) The presence of an additional methyl group at position N4 of the nucleobase does not preclude formation of a third hydrogen bond to complementary dG. Also, the methyl substituent protrudes into the major groove of dsDNA upon formation of a N4C-dG base pair and only marginally perturbs the overall structure of duplex DNA. However, presence of N4C-dG base pairs induce a destabilization of duplex DNA ( $\Delta T_m$  of  $-4.5^\circ\text{C}$  per modification<sup>5</sup>) because of the reduction of the hydration of the double helical structure.<sup>6</sup> The corresponding triphosphate (**2** in Figure 2) has been synthesized and shown to be accepted by various DNA polymerases and incorporated into DNA via PCR.<sup>6</sup> Other dCTP analogues bearing small alkyl substituents (see reference S11 of the Further readings section) or even reactive functional groups<sup>7</sup> at position N4 of cytosine have been reported and shown to be accepted by polymerases even though with reduced efficiency compared to the corresponding canonical nucleotides. Hence, the N4 position of cytosine appears to be a favorable location for the introduction of chemical modifications, both naturally occurring or synthetic, in nucleoside triphosphates without impairing their substrate capacities.

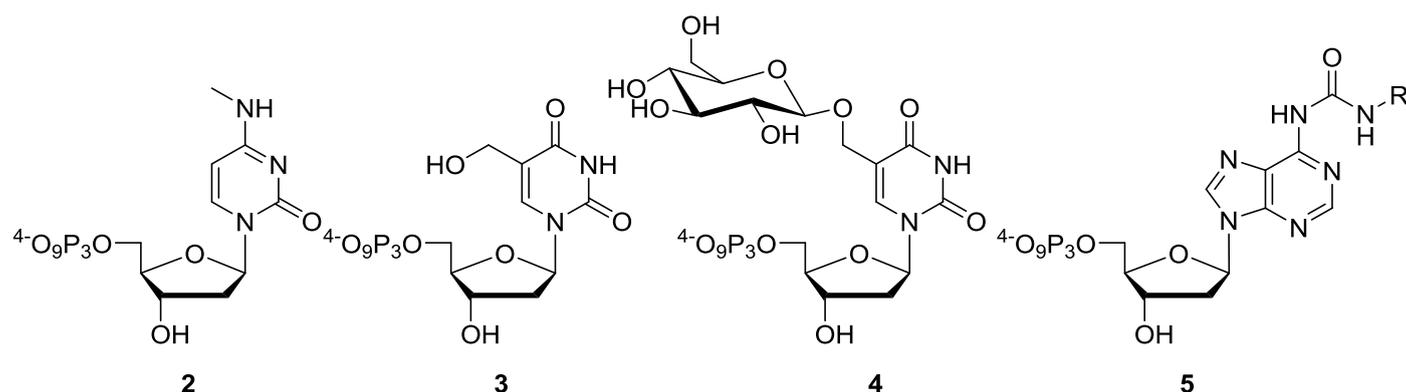
Eukaryotic cells also contain the thymidine nucleotides 5-hydroxymethyluracil (hmU) and 5-formyluracil (fU) which bear identical oxidative patterns than 5hmC and 5fC, respectively, but are present at much lower concentrations than the corresponding cytosine analogues. (see references S6 and S12 of the Further readings section) Even though hmU and fU stem from oxidative DNA damage events, these nucleotide analogues might also have specific functions, particularly in stem cells. In order to facilitate structural and functional studies, the nucleoside triphosphate building block of hmU has been reported.<sup>8</sup> Hydroxymethyl-modified dUTP (**3** in Figure 2) acts as a very good substrate for various polymerases under PEX reaction conditions and can also be used to construct longer DNA fragments by combining PCR conditions and the KOD XL DNA polymerase.<sup>8</sup> Glycosylated 5hmU, thymine  $\beta$ -D-glucosyl-5hmU (also known as base J), has been identified in trypanosomatid species (to a ratio of 0.5% base J per canonical T) and plays a function in transcriptional regulation. Recently, the corresponding nucleoside triphosphate (**4** in Figure 2) has been

synthesized and shown to be compatible with enzymatic DNA synthesis, both under PEX reactions and PCR conditions.<sup>9</sup>

Other, hypermodified DNA pyrimidine nucleotides such as glucosyl-phosphoglucuronolactone-dihydropentauracil and 5-(methylamino)methyluridine (5-mNmU) have been identified, mainly in the genomes of bacteriophages, but none of the corresponding nucleoside triphosphates have yet been reported.

DNA methylation affects not only pyrimidine but also purine nucleotides with N6-methyl-deoxyadenosine (6mA) being the most prevalent at least in prokaryotes where it plays a pivotal role to distinguish host from foreign pathogenic DNA. This methylated adenine analogue has also been detected in various higher eukaryotes where it appears to be involved in gene expression and cell cycle regulation albeit the mechanisms and exact functions remain elusive. This lack of understanding of the exact biological roles and functions of 6mA might be ascribed to its low abundance (approximately ~0.4-0.8% of all adenines in unicellular protists are methylated at position N6) which calls in for highly specific and sensitive analytical methods. Hence, enzymatic synthesis of DNA fragments containing this modification could facilitate biological studies. The corresponding N6-methyl-dATP has been synthesized and is even commercially available. The methyl unit on the exocyclic amine can adopt two rotational isomers: the *cis* isomer adopted by monomeric nucleosides and nucleotides is thermodynamically favored but interferes with Watson-Crick-Franklin base pairing and the *trans* conformation where the methyl group protrudes in the major groove of dsDNA. This dichotomy explains the destabilization caused by the insertion of this modification into dsDNAs.<sup>10</sup> Moreover, the presence of this methyl group at position N6 might also account for the rather poor capacity of N6-methyl-dATP at serving as a substrate for DNA polymerases under PCR conditions.<sup>11</sup> On the other hand, incorporation of 6mA into DNA can be improved when polymerases involved in translesion replication and by-pass such as pol  $\eta$  are used. Interestingly, pol  $\eta$  also efficiently misincorporates 6mA nucleotides opposite templating guanosine nucleotides oxidized at position 8 of the nucleobase (8-oxo-G).<sup>12</sup>

Methylation of position N1 of adenine nucleotides (1mA) is an important posttranscriptional RNA modification (*vide infra*) and represents a major toxic and mutagenic lesion caused by alkylation events of DNA. At the level of DNA, the 1mA modification has not been identified as an epigenetic mark. Nonetheless, the introduction of a methyl group at position N1 and hence a permanent positive charge, substantially alters the overall properties of the adenine nucleobase in terms of its protonation/deprotonation state (i.e. the  $pK_a$  substantially increases from 3.6-3.8 to 8.2-8.8<sup>13</sup>), Watson-Crick-Franklin hydrogen bonding capacity, and reactivity due to its susceptibility to be converted into the corresponding 6mA modification via an intramolecular Dimroth rearrangement.<sup>14</sup> So far, only phosphoramidite building blocks for DNA<sup>13</sup> and RNA<sup>14</sup> solid-phase synthesis exist and no nucleotide for enzymatic synthesis has been reported yet. Lastly, as for 1mA, N6-carbamoylated adenine represents a class of important, highly conserved tRNA modifications (e.g. N<sup>6</sup>-threonylcarbamoyladenosine (t<sup>6</sup>A)) and are only known in DNA as lesions occurring via exposition of DNA to isocyanate or isothiocyanate derivatives.<sup>15</sup> The triphosphate of N6-(N'-phenylcarbamoyl)-dATP (**5** in Figure 2) was synthesized and shown to be incorporated opposite templating T nucleotides (but not opposite abasic sites) with lower but non-negligible rate constants compared to unmodified dATP by the Klenow fragment of DNA polymerase I (Kf (exo)).<sup>15</sup>

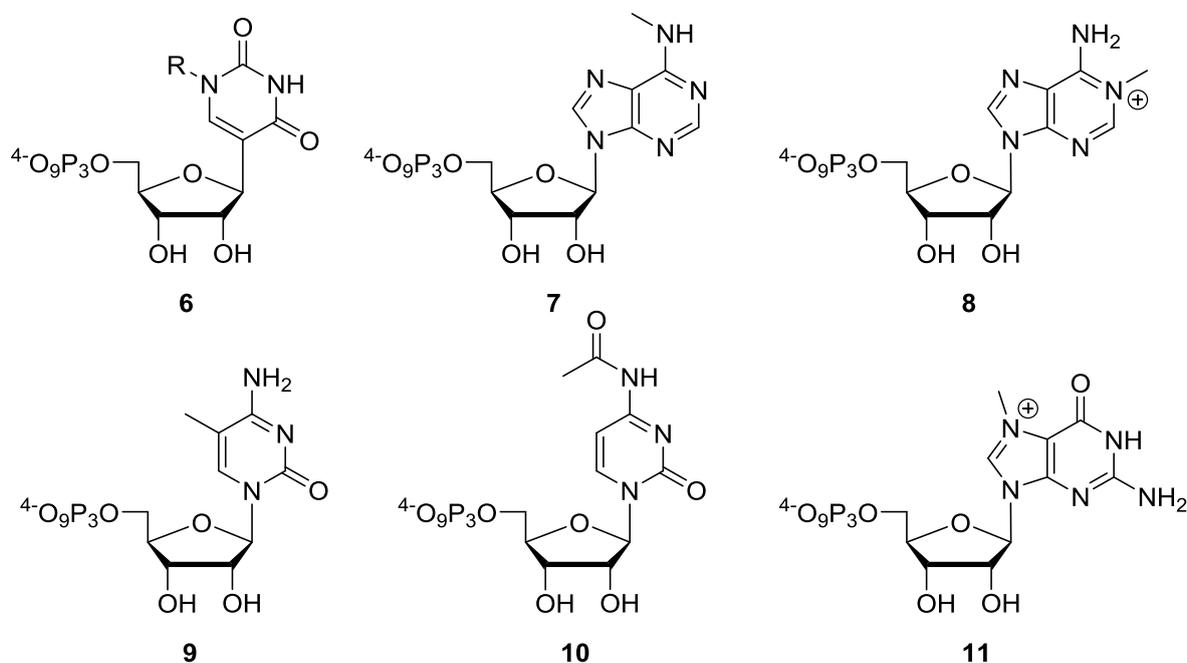


**Figure 2.** Chemical structures of nucleoside triphosphate bearing naturally occurring DNA base modifications.

Numerous modification patterns of guanosine such as 8-oxo-G, O<sup>6</sup>-alkylguanine (O<sup>6</sup>-alkylG), and N7-alkylated guanine DNA adducts have been identified.<sup>16, 17</sup> However, unlike methylated and alkylated A, C, or T nucleotides, all these modified guanosine nucleotides result from chemical damage to DNA rather than as epigenetic marks. Modified nucleoside triphosphates bearing modification at the O<sup>6</sup> and N7 are well tolerated by polymerases<sup>12, 16, 17</sup> and can be used to generate modified DNA by enzymatic synthesis even though misincorporation events are rather frequent.

## b) Epigenetic base-modifications on RNA

More than 100 post-transcriptional modifications are present in RNA and affect all four nucleobases but also the ribose sugar. Amongst the different types of RNA, ribosomal RNA (rRNA) and transfer RNA (tRNA) are the most decorated with chemical modifications, both in terms of frequency and diversity of functionalities. Since it is beyond the scope of this present Chapter to describe all RNA modifications that have been identified so far, only the most prominent examples will be described.



**Figure 3.** Chemical structures of nucleoside triphosphate bearing naturally occurring RNA base modifications.

**Pseudouridine (Ψ)** is the most abundant modification in total RNA and is often considered as the “5<sup>th</sup> base” of RNA. This C-nucleotide analog is present in all forms of RNA and hence is involved in numerous functional roles which are directly connected to its distinctive chemical structure. Even though Ψ and uridine share similar topologies in RNA, Ψ rigidifies RNA backbones and stabilizes U-A base pairs. The stabilizing effect of Ψ compared to canonical U is of around 1 kcal/mol and arises due to a strong preference of Ψ for the C3'-endo sugar conformation imposed by the C-C glycosidic bond combined with favorable, sequence-dependent stacking interactions. Rigidification of RNA scaffolds is explained by the presence of the additional N1H hydrogen bond donor which can mediate the formation of water bridges between the 5' phosphates of either Ψ or other RNA residues with N1H. (see reference S13 of the Further readings section) Due to its high abundance in total RNA and its involvement in numerous biological roles and functions, the nucleoside triphosphate of Ψ (**6** in Figure 3 with R = H) has been prepared.<sup>18</sup> This nucleotide analog is expectedly well-tolerated as a substrate by RNA polymerases such as the T7 RNA polymerase and represents an important chemical tool in numerous biological studies. For instance, synthesis of Ψ-containing RNAs by *in vitro* transcription allowed to demonstrate that this modification improves protein production<sup>19</sup> and markedly reduces the immunostimulatory potential of mRNAs presumably by disrupting interactions with immune sensors such as the Toll-like receptor TLR7. Triphosphate **6** has also been used to probe the function of Multicopy-Associated Filamentation (Maf) proteins which are believed to be involved in the regulation of RNA synthesis. Interestingly, the appendage of a single methyl group on the nucleobase of Ψ, leading to N1-methylpseudouridine (m1Ψ, **6** with R = CH<sub>3</sub>), strongly improves translation efficiency<sup>20</sup> without affecting the low immunogenicity of *in vitro* transcribed mRNAs.<sup>21</sup> These favorable assets have propelled m1Ψ into a key element of modern mRNA vaccines, particularly of those against the SARS-CoV-2 disease (see Section 4a) of this Chapter). (see reference S15 of the Further readings section) Since such long RNA fragments (e.g. the Pfizer-BioNTech vaccine consists of an over 4000 nucleotide-long RNA sequence) cannot be produced by solid-phase synthesis, production of mRNA vaccines heavily relies on the tolerance of the T7 RNA polymerase for modified triphosphate **6** during *in vitro* transcription reactions.

### **Methylated adenosine analogs m1A (8) and m6A (7)**

N6-methyladenosine (m6A; **7** in Figure 3) is one of the most abundant, internal modification found in all

forms of RNA since it represents 0.1 to 0.4% of all adenosines in total cellular RNA and around 50% of all methylated nucleobases. (see reference S16 of the Further readings section) Due to its high prevalence, m6A modifications are involved in numerous biological functions. Amongst these, the best-established is destabilization of mRNA to control RNA decay in a methylation-dependent manner. This destabilization is mainly caused by the presence of the methyl group on the exocyclic amine which needs to adopt a high energy anti orientation as mentioned in the case of the DNA counterpart (*vide supra*). In addition, m6A is involved in translation enhancement of mRNAs and splicing events. This modification also plays roles in pre-mRNA and pri-miRNA processing and has been shown to affect alternative polyadenylation. Various pathways have been devised for the synthesis of modified nucleoside triphosphate **7** (see for instance <sup>22</sup>) which has been demonstrated to be compatible with enzymatic RNA synthesis.<sup>23, 24</sup> Interestingly, when combined with other modified nucleotides such as m5C or 2-thiouridine (s2U), m6A containing mRNA transcripts can block the activation of various TLRs including TLR3, TLR7, and TLR8. (see reference S17 of the Further readings section)

The methyl substituent can also be located at position N1 of the purine nucleobase (m1A; **8** in figure 3) instead of the exocyclic amine as in m6A. The resulting, positively charged modified nucleotide m1A plays an important structural role particularly for the maintenance of the three dimensional structure of mRNA and transfer RNA (tRNA). Indeed, the presence of the methyl group interferes with Watson-Crick-Franklin base pairing but concomitantly interacts with the negatively charged phosphate backbone and can induce the formation of stable hydrogen bonding interactions. Nucleotide **8** has been synthesized and is even commercially available, but no studies on its compatibility with RNA polymerases have been reported yet. Nucleotide **8**, if compatible with *in vitro* transcription methods, could be an important tool for the elucidation of the role played by this modification in mRNAs especially since m1A sites are difficult to detect due to rapid conversion to m6A via the Dimroth rearrangement.

#### **5-methyl-cytosine m5C (9)**

In the previous section, we have highlighted the importance of methylation at position C5 of the nucleobase of cytosine nucleobases in DNA as an epigenetic mark and its involvement in numerous essential cellular functions. The RNA version of m5C (**9** in Figure 3) is mainly found in ribosomal RNA (rRNA) and tRNA, but has also been identified (albeit in low abundance) in most forms of RNA including mRNA, long non-coding RNAs, as well as viral RNAs. As for its DNA counterpart, m5C plays important biological roles for instance in ribosomal assembly and RNA export and its presence increases both translation efficiency and the stability of RNA structures. The corresponding base-modified nucleotide **9** has been synthesized and shown to be compatible with T7 RNA polymerase-mediated *in vitro* transcription reactions.<sup>18</sup> While not as efficient as m1Ψ, the presence of m5C in synthetic mRNA transcripts still allows to improve translation efficiency and to inhibit innate immune response by blocking activation of the TLR7 and TLR8 receptors.<sup>21</sup> The compatibility of nucleotide **9** is expected to provide useful information to further elucidating the structural and functional roles played by this modification, particularly in the context of mRNAs. Nucleotide **9** might also be included in future mRNA vaccines with improved stability and/or efficiencies.

#### **N4-acetylcytidine ac4C (10)**

Next to m5C, N4-acetylcytidine (ac4C; **10** in Figure 3) is undoubtedly the most important and abundant cytosine modification since it has been observed in the main forms of RNA (*i.e.* mRNA, tRNA, and rRNA). The ac4C modification stabilizes the C3'-endo conformation and the bulky substituent is placed in the major groove of duplex RNA via a weak hydrogen bonding interaction between the oxygen atom of the acetyl moiety and the hydrogen atom at position C5.<sup>25</sup> This highly conserved modification is believed to help in correct codon reading and improving both the stability of mRNAs and translation efficiency. The corresponding nucleotide **10** has been synthesized and shown to be compatible with *in vitro* transcription methods.<sup>26</sup> Recently, ac4C **10** was used to produce synthetic mRNAs and the presence of this modification on the resulting mRNAs was compatible with protein synthesis even though translation efficiency did not improve compared to unmodified mRNA. On the other hand, presence of ac4C allowed to reduce the immunogenicity of mRNA mainly by altering protein-mRNA interactions.<sup>27</sup> Hence, the usefulness of nucleotide **10** in the preparation of synthetic mRNA analogs with enhanced properties has clearly been established and is expected to be used in the future.

#### **7-Methylguanosine m7G (11)**

Methylation at position N7 of guanosine yields the positively charged, highly stable 7-methylguanosine nucleotide (m7G; **11** in Figure 3). The presence of the methyl moiety on the nucleobase affects its hydrogen bonding capacity on its Hoogsteen edge but not on the Watson-Crick face as is the case for instance in m1A **8**.<sup>28</sup> m7G is an essential modification of the 5'-cap of eukaryotic mRNA where it reinforces the stability of transcripts against nuclease-mediated degradation and is involved in nearly all stages of the life cycle of mRNAs. This modification has also been observed on internal positions mainly of tRNA and rRNA where it is believed to affect secondary structures and their processing.<sup>28</sup> Due to its high biological

significance, the triphosphate of m7G<sup>16</sup> as well as various derivatives<sup>29</sup> have been synthesized and mainly used in cap forming assays. However, the compatibility of m7G **11** along with its DNA analog with enzymatic synthesis has been evaluated with RNA and DNA nucleotidyl transferases.<sup>16</sup> This study revealed that the RNA nucleotide **11** was readily accepted as a substrate by the RNA nucleotidyl transferases unlike the corresponding DNA polymerase which did not support enzymatic synthesis with this modified nucleotide. Moreover, N7-cyanoborane-modified dGTP, a synthetic analog of **11**, was very well-tolerated by DNA polymerases such as Taq and Sequenase,<sup>17</sup> and N7-deaza-N7-alkylated GTP analogs can be used as substrates for the T7 RNA polymerase in *in vitro* transcription reactions.<sup>30</sup> Hence, nucleotides such as **11** serve for the enzymatic construction of vital 5'-caps of mRNA but can potentially be used for the enzymatic synthesis of modified mRNA transcripts.

### 3) Base modifications for aptamer and catalytic nucleic acids generation via SELEX

The natural modifications highlighted in the previous Section served as inspiration for the construction of synthetic analogs of nucleoside triphosphates. These modified nucleotides present alluring features including facile introduction of (complex) chemical modifications into nucleic acids, synthesis of long sequences, and compatibility with Darwinian *in vitro* selection methods provided they are recognized as substrates by polymerases. Hence, careful design of the modified nucleotides, mainly based on empirical as well as structural information, is required prior to synthesis of these analogs. In this context, unnatural substrates are best tolerated by polymerases when the functional groups are tethered to the C5 of pyrimidines and N7 of N7-deaza-purines because of favorable interactions of the modifications with the polymerase (e.g. cation- $\pi$  or hydrogen bonding) and positioning into the major groove of dsDNA. (see references S18 and S19 of the Further readings section) Introduction of functional groups at these sites has revealed particularly efficient and nucleotides bearing modifications as large as oligonucleotides or proteins are readily incorporated into DNA. The usefulness of this approach for enzymatic synthesis of unnatural DNA and RNA will be highlighted by describing recent progress made in *in vitro* selection of modified functional nucleic acids, i.e. oligonucleotides displaying either binding or catalytic activity.

#### a) Aptamer selection with base-modified nucleotides

Aptamers are single-stranded nucleic acids capable of binding to a broad range of targets with both high affinity and specificity. Due to their intrinsic properties, aptamers are often considered as the nucleic acids version of proteinaceous antibodies. Aptamers, and functional nucleic acids in general, are obtained by Systematic Evolution of Ligands by EXponential enrichment (SELEX) and related combinatorial methods of *in vitro* selection.<sup>31, 32</sup> In conventional SELEX, large libraries of randomized oligonucleotides (consisting typically of  $10^{13}$  individual molecules) are subjected to iterative rounds of target incubation, binding and separation, and PCR amplification with increasing selection pressure or stringency to isolate sequences with high binding affinity. At the end of the SELEX process, enriched pools are subjected to NGS (or cloning-sequencing protocols) and aptamer candidates are identified after a bioinformatics analysis of the resulting sequencing data. After chemical or enzymatic synthesis, the corresponding sequences are then evaluated for their binding capacity by applying various biophysical methods such as surface plasmon resonance (SPR).

Aptamers present a number of advantages compared to antibodies including ease of synthesis and facile chemical modification, limited batch-to-batch variation, and low immunogenicity. Hence, aptamers have been applied in numerous fields and practical applications including drug delivery, diagnostic and sensing, medical imaging, therapeutics, and affinity purification. However, despite these numerous favorable aspects, aptamers also present certain limitations. Indeed, many existing aptamers consist of unmodified DNA or RNA and are thus prone to rapid, nuclease-mediated degradation which is a severe impediment for their direct translation to *in vivo* applications. Although their small size is believed to facilitate tissue penetration, it also drastically reduces their *in vivo* residence times and bioavailability due to rapid renal filtration. Lastly, the arsenal of functional groups present in natural DNA and RNA is rather limited (mainly to exocyclic amines and aromatic moieties), especially when compared to antibodies. Aptamers mediate binding to targets by using this limited array of functional groups combined with  $\pi$ - $\pi$  stacking, hydrophobic interactions (both mainly via the aromatic nucleobases), hydrogen bonding, or van der Waals forces. Hence, due to this lack of functional group variety it is often difficult to identify aptamers against certain targets such as glycoproteins, acidic proteins with low isoelectric points (i.e.  $pI < 7$ ), proteins with no or little defined three-dimensional conformation, or small, hydrophobic molecules. The introduction of chemical modifications alleviates these limitations (at least partially) since they stabilize aptamers for biological and *in vivo* applications and bridge the gap with antibodies by facilitating target recognition via alternate binding mechanisms. Modifications at all levels of the nucleosidic scaffold can be introduced either directly during SELEX using unnatural nucleotides or after SELEX (post-SELEX method) using solid-phase synthesis. The post-SELEX method is very useful for adding chemical alterations to the termini (5' and/or 3') of aptamers

which can increase the bioavailability (e.g. polyethylene glycol (PEG) units) or convey reactivity (e.g. fluorescence, drug delivery, cross-linking). On the other hand, internal modification of aptamers via the post-SELEX approach is also possible but requires tedious and uncertain structure-activity relationship (SAR) studies since the introduction of chemical alterations often comes at the expense of significant depletions in binding affinity. Inclusion of modified nucleotides during SELEX avoids further fine-tuning of the resulting aptamers via SAR studies and is not limited to terminal sections of the aptameric sequences. Sugar and phosphate modified nucleotides, when accepted as substrates by polymerases, are involved in the stabilization of aptamers<sup>33</sup> while nucleobase modifications contribute to increases in binding affinity and specificity, epitope coverage, convey additional reactivity, and are involved to a certain extent in metabolic stabilization.<sup>34</sup> In essence, finely tuned, chemically modified aptamers with improved properties can be obtained when nucleotides equipped with diversity-enhancing functional groups are used directly in SELEX. In this Section, we will describe the modus operandi of this method and highlight recent examples of base-modified aptamers.

SELEX protocols with modified nucleotides do not markedly deviate from those established in conventional SELEX. However, some (subtle) differences exist: one or multiple canonical nucleotides are substituted with the corresponding modified analogs; these unnatural nucleotides need to be accepted by polymerases and their polymerization into randomized libraries needs to be high yielding and specific. Also, after the selection step populations of base-modified oligonucleotide need to be converted, mainly via PCR, into unmodified sequences. This is important prerequisite for SELEX with modified nucleotides since the modifications might be partially read or not recognized at all during NGS analysis or might induce quantification biases. On the other hand, the binding capacity of base-modified aptamer (candidates) can be evaluated by standard biophysical methods such as surface plasmon resonance (SPR) or Octet technology.

The concept of employing base-modified nucleotides in SELEX to bestow DNA with additional functional groups to modulate binding affinity was pioneered by Latham et al. who used 5-(1-pentynyl)-dUTP (**12** in Figure 4A) to prepare naïve ssDNA libraries.<sup>35</sup> The randomized library equipped with a hydrophobic side-chain was then screened for strong binders of human thrombin. This selection experiment yielded anti-thrombin aptamers with  $K_D$  values in the high nanomolar range and clearly demonstrated that i) chemical modifications could easily be introduced by the polymerization of functionalized nucleotides, ii) the presence of the modification altered the outcome of the SELEX process, and iii) the sequences strongly relied on the presence of the hydrophobic side chains for binding interaction and clotting inhibition. Hydrophobic side chains are also the fundament of the class of aptamers called SOMAmers (Slow Off-rate Modified Aptamers). In SOMAmers, pyrimidine nucleotides are equipped with small, amino acid-like, mainly hydrophobic side chains with the intent of reducing the rather high  $k_{off}$  rates often displayed by unmodified aptamers.<sup>36</sup> This strategy is particularly efficient since modified aptamers with  $K_D$  values in the low picomolar range are regularly identified. For instance, when 5-phenylpropyl-1-aminocarbonyl-dUTP (PPdU; **13** in Figure 4A) was used to construct a modified library with  $10^{12}$  individual variants, an aptamer was selected against the RIG-I receptor which displayed an impressive 19 pM binding affinity. SOMAmers with not only one but two modified nucleotides have also been reported.<sup>34</sup> The simultaneous introduction of two unnatural pyrimidine nucleotides (e.g. **13** and 5-[N-(1-naphthylmethyl)carboxamide]-dCTP **14** in Figure 4A) in randomized libraries resulted in the identification of aptamers that displayed an improved binding affinity and target recognition compared to aptamers equipped with only single modification patterns.<sup>34</sup> So far, more than 5000 SOMAmers have been selected against human proteins and are currently used in a multiplexed proteomic analysis platform called SOMAscan.<sup>37</sup> The SOMAscan uses the dual nature of aptamers, namely their capacity at binding to specific targets and that to form duplexes with complementary sequences, to capture and identify individual proteins in complex biological samples.<sup>38</sup> Decorating aptamers with hydrophobic residues improves not only binding affinity and specificity against protein targets but also enables their cellular internalization capacity. Indeed, Tanaka et al. have used 5-((3-indolyl)propionamide-N-allyl)-dUTP (trp; **15** in Figure 4A) to construct a modified, randomized library which was applied in a modified cell-SELEX protocol against a human lung cancer cell line: after incubation, cell penetrating aptamers were recovered by cell lysis.<sup>39</sup> The capacity of the resulting aptamer at specifically internalizing these cancer cells was confirmed by a marked reduction in the levels of the long noncoding RNA (lncRNA) MALAT1 (Metastasis-associated lung adenocarcinoma transcript 1) due to conjugation to an antisense oligonucleotide specific for MALAT1.

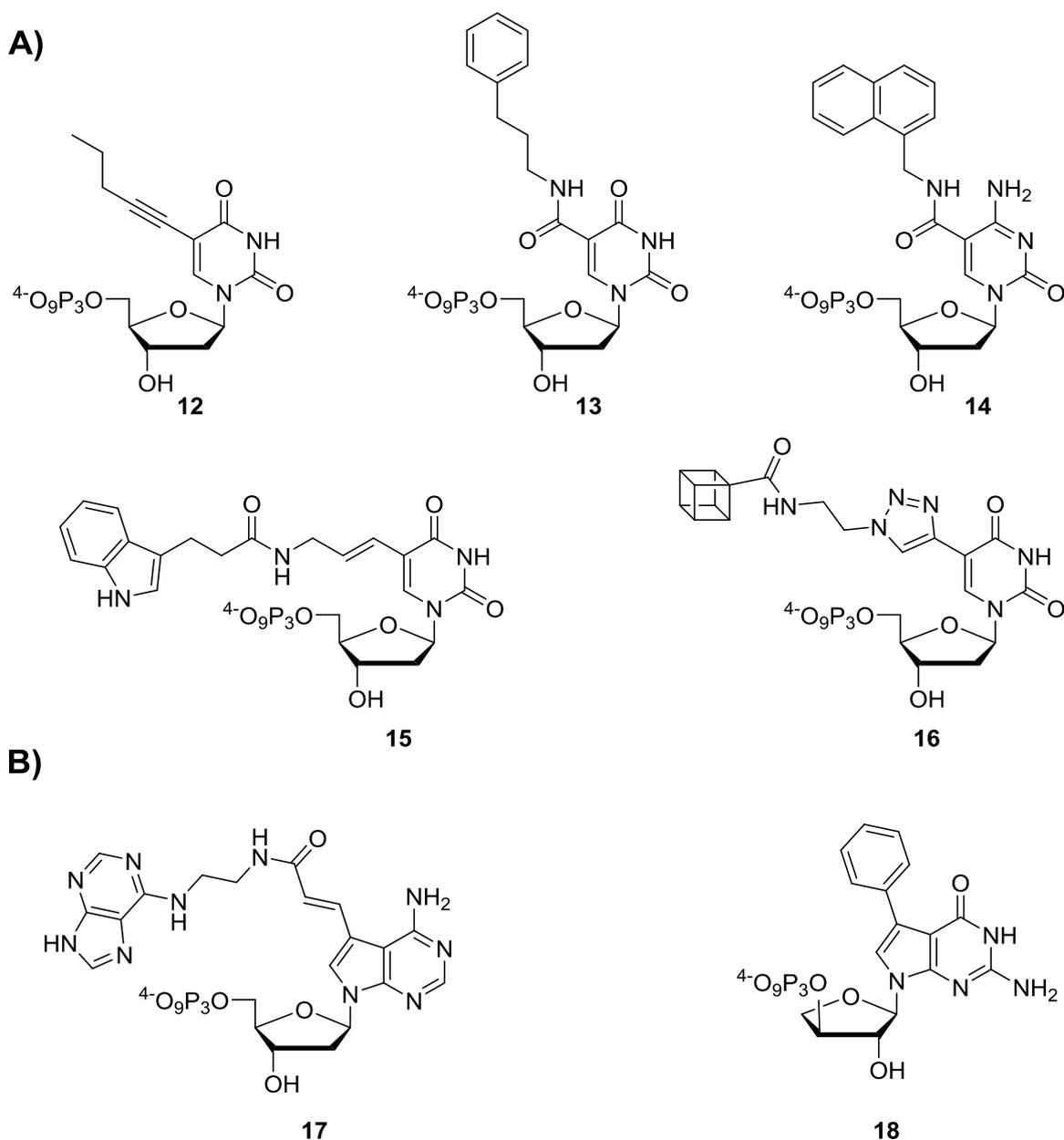
Most chemically modified aptamers are decorated with small, mostly aromatic and hydrophobic moieties or residues reminiscent of the side chains of naturally occurring amino acids. However, a notorious counterexample is the cubane-modified aptamer (cubamer).<sup>40</sup> In the SELEX of cubamers, a dUTP analog bearing the benzene isostere cubane (**16** in Figure 4A) replaces natural dTTP during the generation of the modified libraries. Cubane displays unique properties including biological stability and limited toxicity, disruption of stacking interactions and higher water solubility, and possible formation of unusual C-H $\cdots$ O hydrogen bonding interactions caused by the important ring strain. In addition, a cubane modification deviates from “flatland” chemical space covered by aromatic moieties found in other aptamer selection

experiments. Using the modified dUTP **16** in SELEX, a cubamer against the malaria biomarker *Plasmodium vivax* lactate dehydrogenase (PvLDH) was recently isolated. Even though the  $K_D$  value of the cubamer falls in the higher nanomolar range, the cubamer successfully distinguishes the LDH from *P. vivax* from that of *P. falciparum* despite a high (i.e. 90%) sequence homology between both proteins. This feature is remarkable since aptamers made of unmodified DNA bind to both LDH proteins indistinctively. This distinction was also harnessed to specifically detect the PvLDH malaria biomarker in mimetic clinical samples.

Nucleobase modifications of aptamers are mainly located on (but by no means restricted to) pyrimidine nucleotides due to ease of synthesis of such analogs. (see references S2 and S18 of the Further readings section) Nonetheless, some base-modified purine nucleotides have been incorporated into aptamers. A prime example is the base-appended base (BAB) modification pattern such as (E)-5-(2-(N-(2-(N6-adeninyl)ethyl))carbamylvinyl)-dATP (Ad; **17** in Figure 4B) where an additional adenine nucleobase is attached to position N7 of the deazapurine nucleobase. This functionalization was designed to increase the probability of additional hydrogen bonding and stacking interactions with protein targets. Inclusion of such BAB-modified nucleotides in SELEX has permitted the identification of various, high affinity binding aptamers against relevant biomarkers.<sup>41</sup>

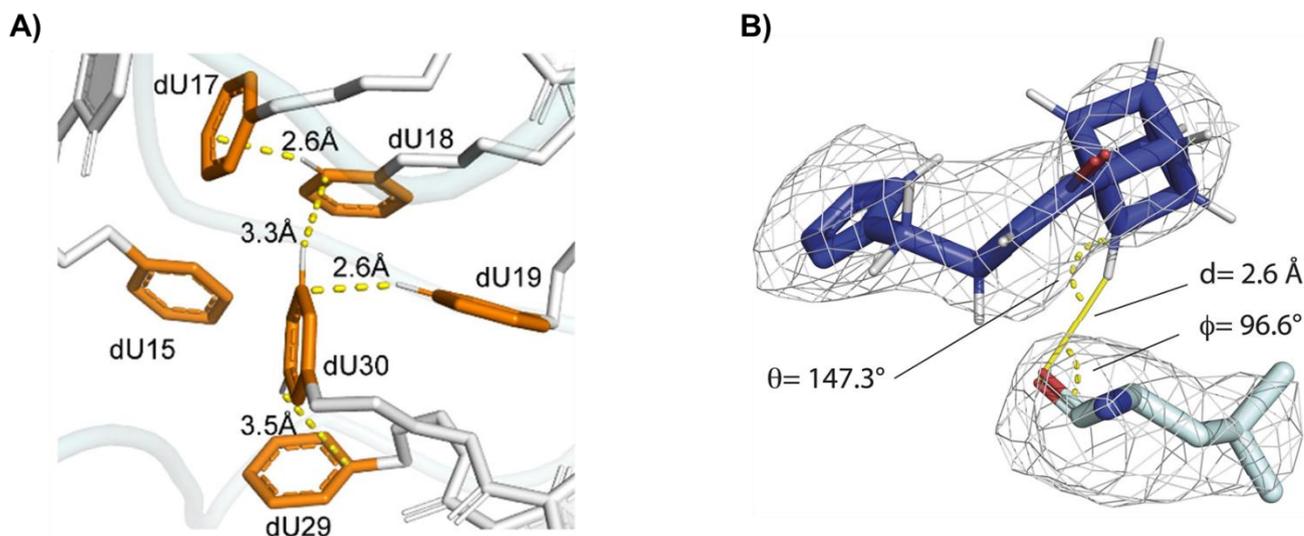
Modified nucleobases can also be appended on nucleotides with sugar chemistries deviating from canonical deoxyribose. For instance, a base-modified RNA aptamer was recently identified by Kabza et al.<sup>42</sup> In this example, 5-aminoallyl-D-UTP was used in lieu of its natural counterpart UTP in SELEX to identify RNA aptamers against the mirror image (i.e. L-RNA) version of the pre-miR-19a hairpin. The enantiomer of the natural RNA target was strictly required in this SELEX experiment to avoid Watson-Crick-Franklin base pairing between oligonucleotides of the randomized library and the solid-support bound RNA. In addition, the positively charged side-chain located on the nucleobase was surmised to favor target recognition via formation of additional hydrogen bonding and electrostatic interactions. After conversion of the base-modified D-RNA aptamer into its mirror image via solid-phase synthesis, a highly potent ( $K_D$  value of 2.2 nM) L-RNA aptamer against native D-pre-miR-19a hairpin was obtained.<sup>42</sup> This first example certainly paves the way for the identification of other base-modified RNA aptamers in the near future.

Nucleoside triphosphate analogs with altered aldose and nucleobase chemistries have also been shown to be compatible with aptamer selection protocols. For instance, N7-deaza-7-phenyl-tGTP (**18** in Figure 4) displays an N7-deaza-guanine nucleobase equipped with a small aromatic moiety affixed on a  $\alpha$ -L-threofuranosyl nucleic acid (TNA) scaffold.<sup>43</sup> The presence of the N7-deaza-7-phenyl modification not only increases the possibility of hydrophobic interactions with target proteins but also avoids the unfavorable syn conformation adopted by the unmodified tGTP nucleotide and hence improves the fidelity of replication during enzymatic TNA synthesis. Similarly, the presence of the threose sugar enhances resistance against nuclease-mediated degradation of the ensuing aptamers but also offers the possibility of exploring altered helical geometry compared to natural DNA or RNA. The modified tGTP nucleotide **18** was used along with the three corresponding TNA triphosphates to construct a randomized library containing  $\sim 10^{13}$  individual TNA sequences which was subjected to SELEX. After three rounds of SELEX against the HIV reverse transcriptase, a high binding ( $K_D$  value of 3.0 nM), TNA modified aptamer was identified.<sup>43</sup>



**Figure 4.** Chemical structures of base-modified nucleoside triphosphates used in aptamer selection. A) Modified pyrimidine and B) modified purine nucleotides.

All aptamers decorated with modifications at the level of the nucleobase strongly depend on the presence of the functional groups for binding activity and even minute alterations to these side chains can cause important decreases in binding affinity. In order to gain insight into the mechanism and the role played by chemical modifications in the formation of aptamer-target complexes several structural biology methods can be employed. The most potent, yet not the most forthcoming, method is the resolution of crystal structures of binary aptamer-target complexes. Several co-crystal structures of modified aptamers, mainly SOMAmers, have been reported and reveal structural motifs (Figure 5A) that complement those adopted by natural aptamers. Indeed, common structural motifs adopted by unmodified DNA and RNA aptamers include portions of double helices majorly capped with hairpin loops, formation of non-canonical base pairs, and G-quartet domains. The presence of base-modifications leads to the formation of hydrophobic clusters, both within aptamer sequences and at the target interface. (see reference S21 of the Further readings section) This additional structural motif found in pseudo-knots forms a surface that is capable of interacting with hydrophobic pockets on the target protein. The main force behind such a hydrophobic cluster motif is the build-up of an extensive array of  $\pi$ - $\pi$  stacking interactions between canonical and functionalized nucleotides. Interestingly, such motifs are also present in aptamers equipped with non-planar, non-aromatic side-chains such as the cubamer where a cluster of four cubane residues was found to interact with a hydrophobic pocket on the protein partner.<sup>40</sup> In addition, the modifications can induce backbone distortions to favor non-classical base pairing, edge-to-face instead of more conventional face-to-face interactions between aromatic residues (Figure 5A),<sup>44</sup> or the establishment of unusual C-H $\cdots$ O bonds which might be involved in the discriminatory specificity of the cubamer (Figure 5B).<sup>40</sup>



**Figure 5.** Structural key features of novel binding interactions of base-modified aptamers. A) array of edge-to-face interactions of benzyl moieties of the SOMAmer obtained with nucleotide **13**. Reproduced with permission from reference <sup>44</sup>; B) unusual C-H...O hydrogen bonding observed in the interaction of the cubane-modified aptamer in complex with its protein target. Adapted from reference <sup>40</sup>.

Overall, the inclusion of base-modifications in SELEX offers the possibility of screening and exploring both sequence and chemical space but also generates a structural plasticity which is not available to unmodified systems. This combination massively improves the general properties of aptamers. Additional structural elucidations will allow to refine our understanding of the involvement of modifications in binding interaction and the design of nucleotides to be used in future SELEX for the identification of yet more potent aptamers.

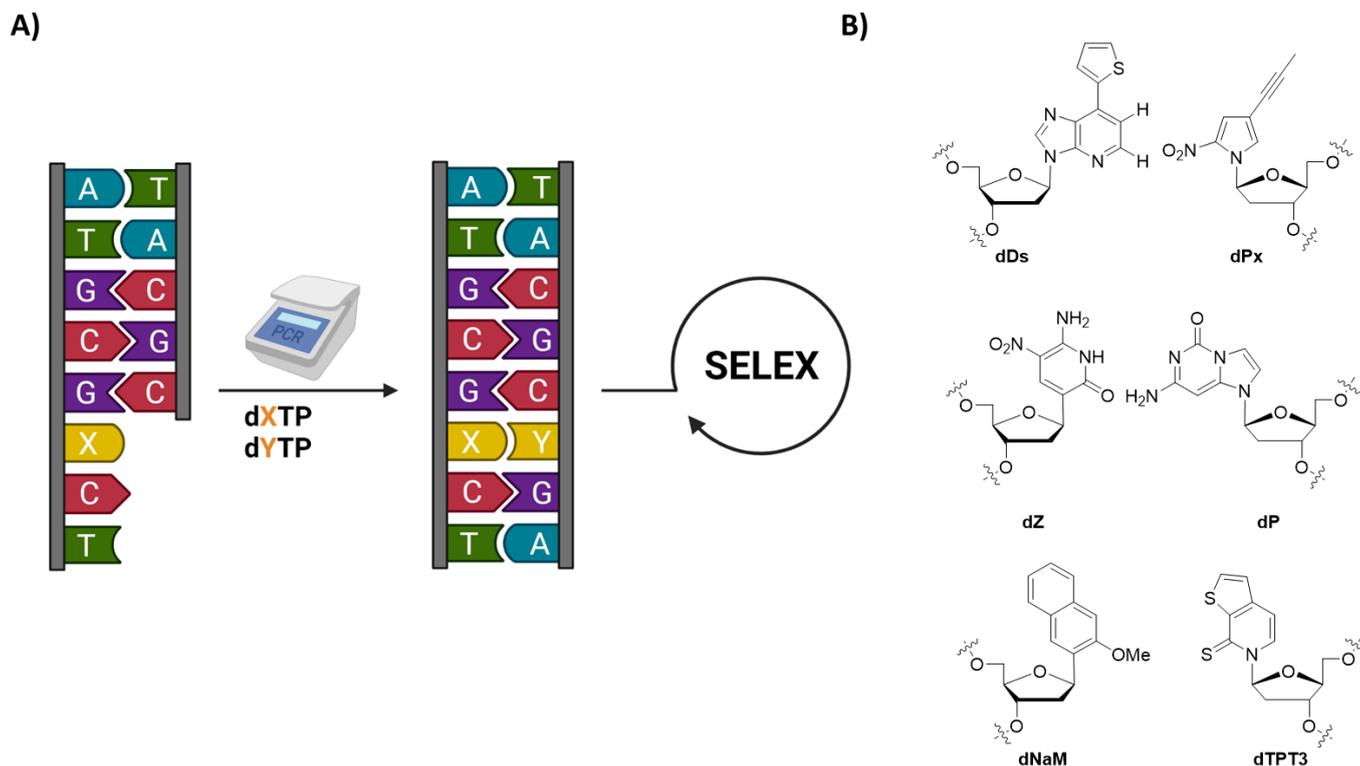
#### b) Expansion of the genetic alphabet and aptamers

Expanding the genetic alphabet by one or multiple unnatural base pairs (UBPs) represents another method to diversify the chemical landscape of randomized libraries amenable to SELEX (Figure 6A). These artificial nucleotides complement the canonical A-T and C-G pairs and rely on alternate hydrogen-bonding patterns, shape complementarity, or packing forces to favor pairing and concomitantly prevent mispairing with natural nucleotides. Successful examples of UBPs are scarce due to the complexity and the high demands of such systems which usually call in for lengthy SAR studies prior to identification.<sup>45</sup> Nonetheless, one of the most prominent examples is the **dDs-dPx** base pair (Figure 6B) developed by the Hirao laboratory and which was inspired by one of the first ever discovered UBP consisting of a difluorotoluene nucleotide (F) base pairing with an analog of deoxyadenosine (Z) devoid of hydrogen-bonding capacity.<sup>46</sup> In the pairing partner of **dDs** in the UBP, the six-membered aromatic moiety of the original base F was substituted with a pyrrole unit and the presence of nitro and alkyne groups further reduced pairing with canonical dA and improved the selectivity of incorporation, respectively. In the unnatural base **dDs**, a nitrogen atom was restored in base Z to permit interactions with the polymerase while an additional thienyl moiety was installed to avoid pairing with natural nucleotides. This molecular engineering resulted in the creation of the highly orthogonal **dDs-dPx** base pair that displays an impressive fidelity in PCR amplification (>99.9% per cycle). (see reference S22 of the Further readings section) The **dDs-dPx** UBP follows a similar shape complementarity to that of natural DNA where large purines pair with small pyrimidines and the cohesion of both partners is assured only by hydrophobic interactions. Due to the alluring intrinsic properties of the **dDs-dPx** UBP, this additional base pair was used in a modified SELEX protocol called ExSELEX (standing for genetic alphabet Expansion SELEX) to identify high-affinity aptamers with an expanded genetic alphabet. A particular feature of ExSELEX is the protocol for the sequencing of enriched populations due to the incompatibility of the modified nucleotides (i.e. **dDs** and **dPx**) with conventional NGS or related sequencing methods. To circumvent this limitation, **dDs** nucleotides are positioned at pre-defined positions of the randomized library which can be identified by the presence of small (2-3 nucleotide long), specific bar-code sequences at the end of the 5'-primer binding region. In addition to this recognition bar-coding system, enriched populations are subjected to replacement PCR amplification where all **dDs** analogs are eventually substituted with natural dA nucleotides prior to NGS analysis. (see reference S22 of the Further readings section) Application of the ExSELEX protocol has led to the identification of aptamers that bind to their protein targets with low picomolar affinities.<sup>47</sup> ExSELEX was recently extended to the identification of aptamers capable of recognizing various breast cancer cell lines.<sup>48</sup> Interestingly, the affinity, exonuclease-resistance, and thermal stability of **dDs**-modified aptamers can be further increased by including small structures called mini-hairpin DNA at their 3'-termini.

By shuffling and relocating hydrogen bond donors and acceptors on purine and pyrimidine nucleobases, Benner and co-workers have identified the **dZ-dP** UBP (Figure 6B). In this UBP, **dP** is an isoguanosine analog where position 1 of the purine scaffold was modified to act as a hydrogen bond acceptor rather than donor and to simultaneously block the keto-enol tautomerism observed isoguanosine that would make the base complementary to thymidine. Moreover, atom shifting of the N7 nitrogen atom of a purine scaffold to position 5 alters the  $pK_a$  and avoids the emergence of charges on the nucleobase. The desired hydrogen bonding pattern of the pseudoisocytidine analog **dZ**, matching that of the pairing partner **dP**, can only be achieved with C-glycosides. In addition, the presence of an electron-withdrawing nitro substituent is strictly required to prevent efficient epimerization from occurring.<sup>49</sup> The **dZ-dP** UBP is fully orthogonal to canonical nucleotides due to the careful design of the hydrogen bonding pattern and displays very low levels of mutation (less than 0.2% per cycle) during PCR amplification.<sup>50</sup> The intrinsic properties of the **dZ-dP** UBP can be leveraged to generate artificially expanded genetic information systems (AEGIS) and to identify aptamers using in vitro evolution.<sup>51, 52</sup> As for the Ex-SELEX, using AEGIS in selection experiments calls for a protocol that deviates from the conventional in vitro selection method on specific points. First, the small level of mutation observed during PCR amplification is exploited for the sequencing of the enriched pools (**dZ** and **dP** are also not compatible with sequencing methods) obtained after several rounds of SELEX. Essentially, the **dZ** is transformed into natural pyrimidines (C and T), while **dP** is converted into canonical purines (A and G) during bar-coded DNA amplification. Second, **dZ-dP** UBPs do not necessarily need to be present in the starting library and can be gained or lost during the SELEX process. This AEGIS-SELEX protocol was successfully applied and aptamers with  $K_D$  values in the low nanomolar range were identified against a range of cell lines and proteins.<sup>51, 52</sup> Rearrangement of the hydrogen bonding pattern permitted the development of the **ACGTZP** AEGIS but is, in theory, not limited to only two additional base pairs and can be expanded to a total of four additional UBPs. In a careful design, the **dZ-dP** UBP can be complemented by the **dS-dB**, **dV-dJ**, and the **dK-dX** which would constitute the **ACGTZPSBVJKX** AEGIS. In a recent effort by the Benner laboratory, the possibility of further extending AEGIS to a system consisting of the two natural base pairs (A-T and C-G) and two modified base pairs (**dZ-dP** and **dS-dB**) was demonstrated. The information encoded in the resulting **ACGTZPSB** AEGIS, coined hachimoji DNA, could be transmitted to RNA by T7 RNA polymerase mediated in vitro transcription using the corresponding, base-modified nucleoside triphosphates.<sup>53</sup> So far, the **ACGTZPSB** AEGIS has not been used in SELEX for the identification of aptamers.

Following a more medicinal chemistry type of approach, the Romesberg laboratory developed other potent UBPs based on hydrophobic and packing forces.<sup>45</sup> Screening of hundreds of UBP candidates, followed by SAR studies and optimization steps, partially guided by structural investigations led to the identification of the **dNaM-dTPT3** base pair (Figure 6B). In the C-glycoside nucleotide **dNaM**, the nucleobase is substituted with an aromatic surrogate consisting of two fused rings that offers a larger surface area for stacking interactions during enzymatic synthesis and protrudes into the major groove of DNA. Similarly, **dTPT3** is the binding partner of **dNaM** and consists also of two fused rings. Additional structural features include a thio N-glycoside structure which is strictly required for UBP synthesis and replication and a thieno ring which is believed to increase intrastrand packing interactions and concomitantly promote deintercalation of the **dNaM-dTPT3** during the extension step.<sup>54</sup> The **dNaM-dTPT3** base pair is replicated with high fidelity during in vitro synthesis and replication (> 99.98% in PCR) which in principle makes it directly amenable to SELEX (even though no aptamers have been identified with this system yet). Importantly, this **dNaM-dTPT3** UBP is recognized in vivo by the replication and translation machineries and is the fundament of the development of the first semi-synthetic organisms with expanded genetic alphabets.<sup>55</sup> These semi-synthetic organisms efficiently store the additional genetic information provided by a **dNaM-dTPT3** UBP since replication proceeds with sufficient retention of the UBP. The encoded information can then be relayed to RNA and harnessed to synthesize modified proteins.<sup>45</sup>

Instead of hydrophobic and packing interactions, shape complementarity, or alternate hydrogen bonding patterns, metal coordination of nucleobases can be used in the development of orthogonal UBPs. Conceptually, the formation of artificial metal base pairs requires two synthetic nucleotides that do not recognize canonical nucleotides altogether and self-pair in the strict presence of specific metal cations. The resulting metal base pairs display high levels of orthogonality and improved thermal stabilities compared to canonical Watson-Crick-Franklin base pairs. Generally, metal base pairs do not consist of large, hydrophobic and aromatic moieties and hence do not suffer from the drawbacks associated with existing UBPs such as poor solubility and production of ROS due to near-visible light absorption.<sup>56</sup> The presence of metal cations is also expected to provide nucleic acids with additional reactivity which in turn could be used for the development of catalysts (vide infra) and might provide valuable help for structural elucidation studies.

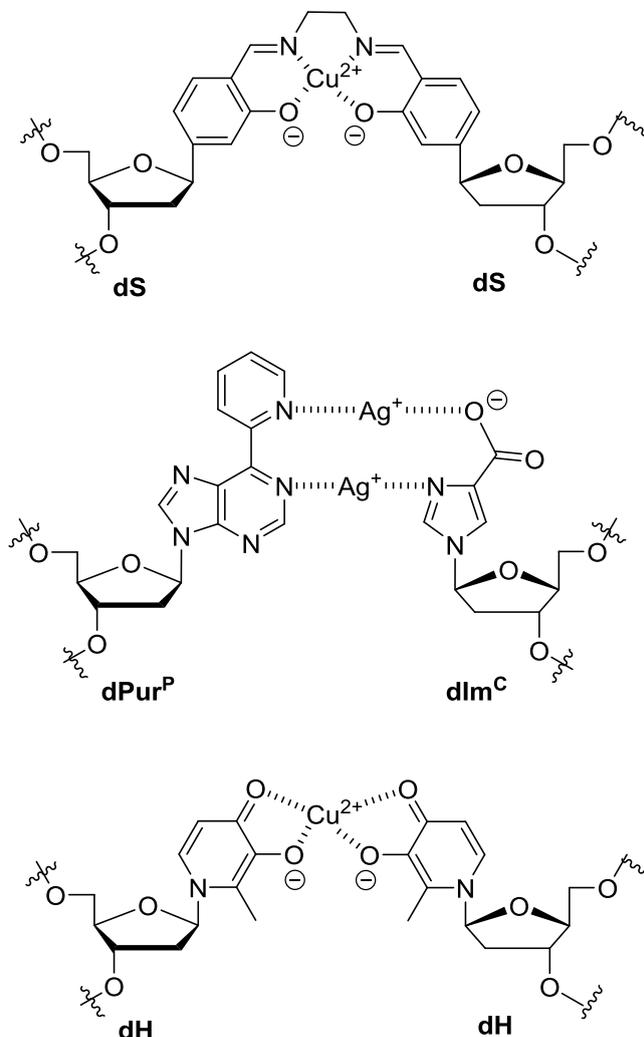


**Figure 6.** A) Schematic representation of the generation of modified libraries with unnatural base pairs for use in SELEX (where  $dXTP$  and  $dYTP$  represent the corresponding modified nucleoside triphosphates); B) chemical structures of representative unnatural base pairs systems.

A large variety of efficient metal-mediated base pairs have been reported which broaden the scope of such artificial nucleic acids. Most studies aiming at identifying artificial metal base pairs focus on small, synthetic oligonucleotides and metal UBP formation is generally monitored by biophysical methods such as UV melting experiments or by structural studies. However, enzymatic synthesis of metal-mediated base pairs needs to fulfill several criteria that cannot be addressed by short synthetic oligonucleotides and biophysical or structural studies. Indeed, besides high thermal stability of the resulting UBP, the modified nucleotides need to be incorporated specifically opposite the respective binding partner on the template strand in the strict presence of specific metal cations. In addition, once the metal-driven UBP is installed in the duplex, DNA synthesis needs to resume. This prerequisite is rather difficult to achieve, not only for metal-based systems but in all AEGIS systems. Lastly, the concentration and nature of the metal cation need to be finely tuned so as not to compromise the efficiency of the polymerase. Despite these rather drastic requirements, some successful examples of enzymatic synthesis of metal-mediated base pairs have been reported (Figure 7). In a prime example, the group of Carell used transition metal-based template synthesis to form salen complexes that connect two artificial  $dS$  nucleotides together.<sup>57</sup> More specifically, the corresponding nucleoside triphosphate of  $dS$  is incorporated opposite a templating  $dS$  nucleotide with little cross-reactivity with natural nucleotides. The addition of ethylene diamine and a suitable transition metal cation (in this case  $Cu^{II}$ ) leads to the formation of the complex displayed in Figure 7. Crystal structures of  $dS$ -modified DNA located in the polymerase, in absence or presence of metal cations and/or modified triphosphates allowed to refine the mechanism of formation of such an artificial metal base pair. In another example, the silver-mediated base pair formed between the artificial nucleotides  $dIm^C$  and  $dPur^P$  (Figure 7) can be constructed using a two-polymerase protocol.<sup>58</sup> The nucleotide analog  $dIm^C$  consists of an imidazole nucleobase surrogate equipped with an additional carboxylic acid moiety for further metal coordination while  $dPur^P$  contains a purine analog expanded with a pyridine moiety which installs a bipyridine-type of ligand structure into the nucleobase. When both nucleotide analogs are employed in conjunction with  $Ag^I$  and DNA polymerases including Vent (exo<sup>-</sup>) and Kf (exo<sup>-</sup>), the corresponding metal base pair is formed in high yields. Since most polymerases do not bypass such a UBP, an engineered polymerase (Therminator) is necessary to resume DNA synthesis after installation of the silver-mediated base pair. Interestingly, UV melting experiments with short synthetic oligonucleotides containing  $dPur^P$  revealed strong  $Ni^{II}$ -based metal homo base pair formation<sup>59</sup> which was not observed in enzymatic synthesis studies with this nucleotide analog. Lastly, the artificial nucleotide ( $dH$ ; Figure 7) bearing a ligand-type hydroxypyridone moiety, forms very stable,  $Cu^{II}$ -mediated base pairs.<sup>60</sup> The corresponding nucleoside triphosphate,  $dHTP$ , acts as a good substrate for Kf (exo<sup>-</sup>) and single incorporation events can be achieved in high yields when the templating nucleotides are either unmodified  $dA$  or  $dT$ . Similarly to the  $dIm^C$ - $Ag^I$ - $dPur^P$  system, after incorporation of single  $dH$  nucleotides, DNA synthesis can be resumed by using the translesion DNA polymerase *Sulfolobus solfataricus* DNA Polymerase IV.<sup>60</sup> Alternatively, multiple  $dH$  nucleotides can be incorporated

into DNA by using an elegant version of gap filling synthesis.<sup>61</sup> Chemical diversity at the level of the sugar combined with ligand-type nucleobases would grant access to the enzymatic formation of metal base pairs in xenonucleic acids (XNAs). In a first step towards this important aim, rU-Hg<sup>II</sup>-dT hetero-base pairs could be formed by enzymatic synthesis.<sup>62</sup>

Despite these impressive results, misincorporation of natural nucleotides opposite templating modified nucleotides (as in the case of dS) or vice versa (as for dH and dPur<sup>P</sup>) or requirement for multi-enzyme protocols (as for dH and dPur<sup>P</sup>) still hampers the application of enzymatic synthesis of metal base pairs to SELEX for the identification of functional nucleic acids.



**Figure 7.** Chemical structures of metal-mediated base pairs synthesized with polymerases and modified nucleoside triphosphates.

### c) Selection of base-modified catalytic nucleic acids

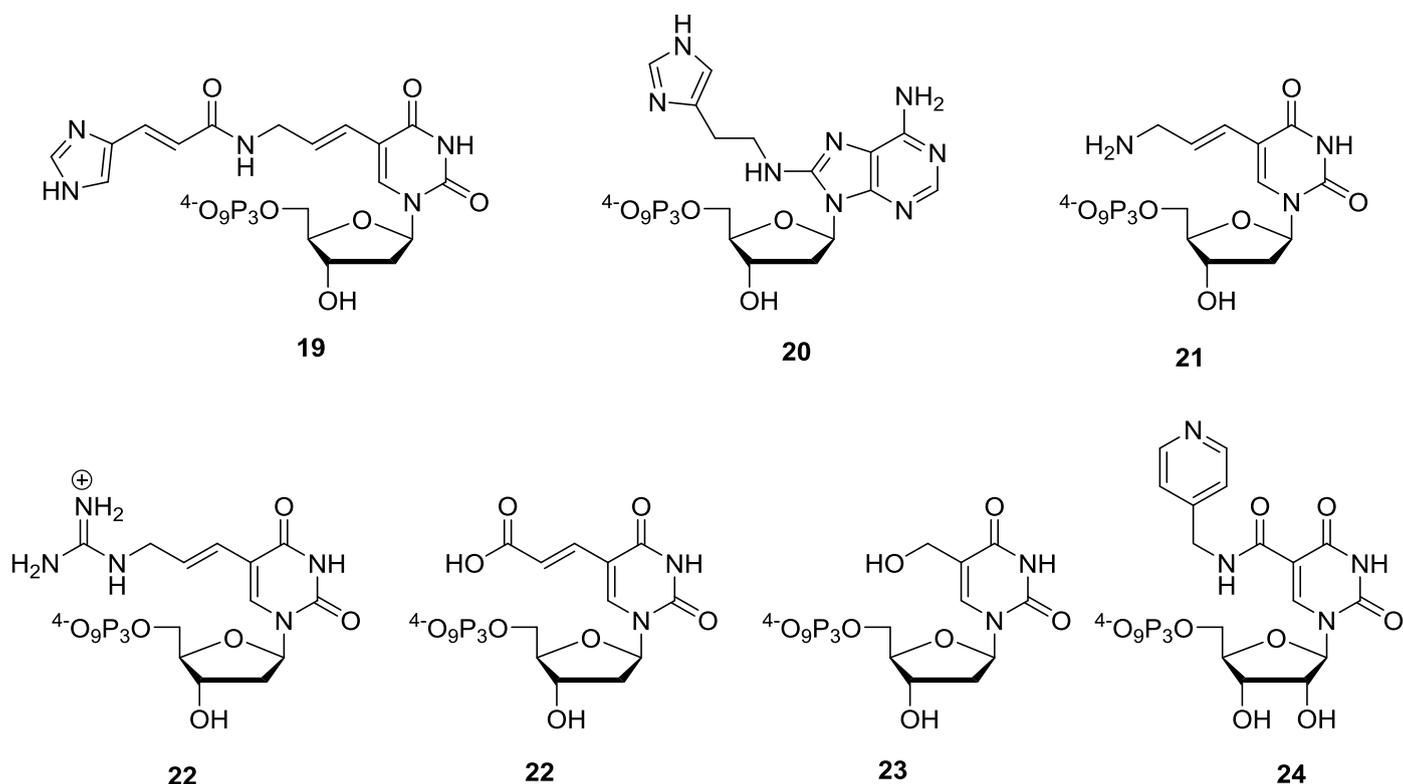
Besides binding to specific targets, nucleic acids can also catalyze chemical transformations and reactions. Such nucleic acids catalysts are called DNAzymes or deoxyribozymes in the case of DNA and ribozymes for RNA. While ribozymes are found in natural settings, DNAzymes are man-made entities. All catalytic nucleic acids can be obtained by a similar selection process as that described for the identification of aptamers. In aptamer SELEX, the selection step involves partitioning between binders and non-binders. At this stage of the SELEX protocol, non-binding species mainly remain in solution while oligonucleotides are recovered from the immobilized target mainly by heat treatment. In contrast, during DNAzyme/ribozyme SELEX randomized libraries containing the substrate are immobilized on solid-support and the catalytically active species are recovered in solution after magneto-separation and cleavage. Cleavers can also be separated from species devoid of catalytic activity by gel electrophoresis or a combination of both methods. (see reference S25 of the Further readings section) Application of this altered SELEX protocol resulted in the identification of catalysts with efficiencies reaching near kinetic perfection (i.e.  $k_{cat}/K_M$  of  $10^9 \text{ min}^{-1} \cdot \text{M}^{-1}$ ).<sup>63</sup> However, since canonical nucleic acids are functionality deprived biopolymers, DNAzymes and ribozymes require divalent metal cations ( $M^{2+}$ ) to mediate catalytic activity – and this often at very high and non-physiological concentrations. The requirement for  $[M^{2+}]$  often exceeding those found in cells combined with the relative fragility of oligonucleotides in the presence of nucleases limits the *in vivo* usefulness of DNAzymes and ribozymes. In addition, nucleic acids are rather refractory to catalyze certain reactions such as amide bond scission, mainly due to the lack of functional groups capable of sustaining

such processes. As for aptamers, chemical modifications can (at least partially) remediate these shortcomings. For instance, the introduction of sugar modifications, either by application of post-SELEX SAR studies or by using modified nucleotides during the *in vitro* selection process, enhances the stability of nucleic acid catalysts in biological media.<sup>64</sup> However, sugar modification do not abrogate the strong  $M^{2+}$ -dependence or expand the repertoire of chemical reactions accessible to DNAzymes and ribozymes. On the other hand, appendage of chemical modifications on the nucleobases provides nucleic acids with additional functionalities that can mediate or participate in catalytic activity. Such an approach can potentially lead to  $M^{2+}$ -independent catalytic activity and/or lead to the catalysis of reactions that are not accessible to canonical systems. This versatile approach will be highlighted through various examples of DNAzymes and ribozymes decorated with artificial side-chains on the nucleobases.

In a first example, Santoro and Joyce used the modified dUTP **19** (Figure 8) which is equipped with an imidazole moiety capable of chelating  $Zn^{2+}$  to prepare randomized libraries with an expanded chemical repertoire.<sup>65</sup> When these modified libraries were subjected to SELEX, a highly potent RNA-cleaving DNAzyme coined 16.2-11 was isolated. This  $Zn^{2+}$ -dependent catalyst hydrolyzed all-RNA substrates with impressive catalytic efficiencies ( $k_{cat}/K_M$  of  $10^8 \text{ min}^{-1} \cdot M^{-1}$ ) under simulated physiological conditions. In terms of mechanism, three conserved nucleotides are believed to place a  $Zn^{2+}$  cofactor into close proximity of the 2'-OH via coordination through the imidazole moieties. In turn, the metal cation might then activate the hydroxyl residue for the *in-line* attack on the phosphorous center of the scissile phosphodiester linkage in a mechanism reminiscent to that adopted by certain metalloproteinases. Hence, DNAzyme 16.2-11 highlights the benefits from endowing DNA with exogenous functional groups to bestow reactivity and catalytic prowess that are difficult to reach for unmodified, canonical nucleic acids. Related to this seminal work, Perrin and co-workers used two rather than one modified nucleotides in SELEX to identify DNAzymes capable of catalyzing the  $M^{2+}$ -independent hydrolysis of RNA linkages.<sup>66</sup> By equipping dATP with an imidazole residue and dUTP with a cationic amine (**20** and **21** in Figure 8, respectively), catalytic species were expected to arise during SELEX that used residues reminiscent of those of the lysine and histidine side chains that are implicated in the general acid–base mechanism adopted by the protein enzyme RNase A. The simultaneous polymerization of the modified nucleotides **20** and **21** and the concomitant SELEX experiment resulted in the identification of DNAzyme 9<sub>25</sub>-11 which hydrolyzed RNA substrates in the absence of  $M^{2+}$  with high rate constants ( $k_{obs}$  of  $0.20 \text{ min}^{-1}$ ) and which displayed multiple-turnover kinetics. The overall properties of DNAzyme 9<sub>25</sub>-11, including higher optimal reaction temperatures, cleavage of all-RNA substrates, and increased rate constants and catalytic efficiencies, were further improved by including a third modified nucleotide. Indeed, presence of a permanently positively charged guanidium residue (dUTP analog **22** in Figure 8) in the scaffold of DNAzymes ensures stabilization of the surrounding negatively charged phosphate units and might be implicated in the catalytic step (e.g. by stabilization of the build-up of negative charge in the transition state).<sup>67, 68</sup>

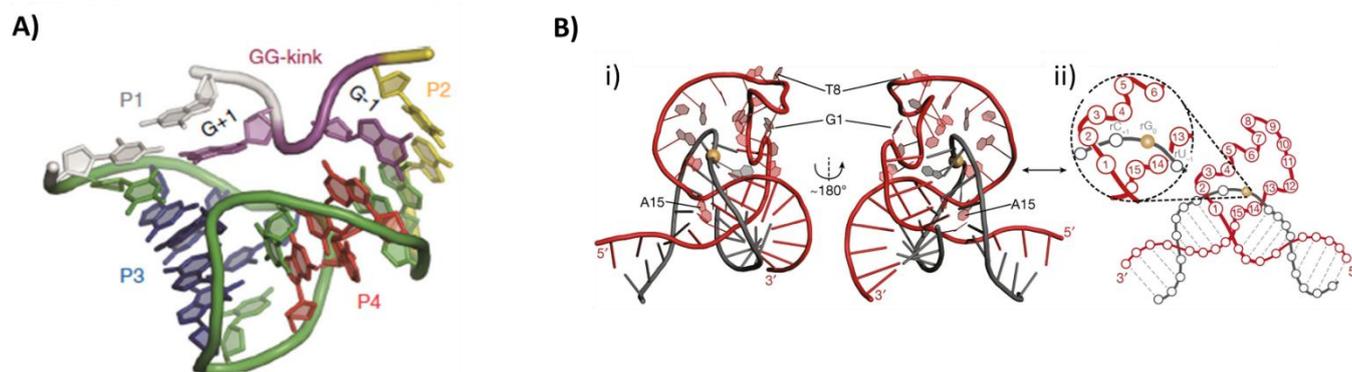
The presence of modified side-chains also conveys reactivity to DNA that cannot be attained with canonical nucleic acids. In this context, amide bond hydrolysis is a reaction that has eluded nucleic acids catalysts based on natural scaffolds despite numerous isolation attempts. When nucleotides equipped with amine, carboxylic acid, and hydroxyl side chains (**21**, **22**, and **23** in Figure 8, respectively) were used in parallel SELEX experiments, DNA catalysts capable of hydrolyzing amide bond linkages could be identified.<sup>69</sup> While the catalytic efficiencies remained rather modest ( $k_{obs}$  values in the  $10^{-3} \text{ min}^{-1}$  range with conversion yields not exceeding 70%), this seminal work clearly highlights the benefits nucleic acids can gain from the addition of chemical functional groups. Reselection and substitution to other, more reactive functional groups might yield more proficient amide-bond cleaving DNA-based enzymes in the near future.

Examples of ribozymes bearing exogenous functionalities on the nucleobases are rather scarce. The Eaton laboratory used the modified NTP **24** (Figure 8) in SELEX to isolate ribozymes capable of catalyzing the Diels-Alder cycloaddition. Uridine analog **24** bears a pyridyl moiety on the nucleobase which was deemed to form Lewis acid complexes with transition metals and interact with constituting elements of the RNA scaffold via stacking and/or hydrogen bonding interactions. The resulting ribozyme displayed important catalytic efficiencies ( $k_{cat}/K_M$  of  $3.95 M^{-1} \cdot s^{-1}$ ), representing an 800-fold rate acceleration compared to the uncatalyzed Diels-Alder cycloaddition.<sup>70</sup> Expectedly, the ribozyme strictly required the presence of both the modified nucleotide and  $Cu^{2+}$  as cofactor for catalytic activity. Hence, RNA can also benefit from the inclusion of chemical modifications to mediate catalysis, especially for more demanding or complex reactions such as C-C bond formation.



**Figure 8.** Chemical structures of base-modified nucleotides used in SELEX for the identification of DNazymes.

So far, post-SELEX methods aiming at improving the catalytic properties of DNazymes mainly relied on SAR studies where non-conserved nucleotides were replaced by modified counterparts equipped with functional groups capable of interacting with the substrate, chelate metal cations, or convey stabilization of the catalytic core by stacking and hydrogen bonding interactions. Guidance of these SAR studies was limited due to the lack of structural and detailed mechanistic information on DNazymes. However, the recent resolution of X-ray (Figure 9A)<sup>71</sup> and NMR (Figure 9B)<sup>72</sup> structures of the most proficient, unmodified DNazymes 8-17 and 10-23, respectively will certainly reduce the complexity of such an undertaking and help engineering modified, more potent versions of these catalysts. In addition, synthesis of such modified constructs is not restricted to solid-phase synthesis but can be complemented by chemoenzymatic methods involving either multiple polymerases (see e.g. for the construction of metal base pairs) or temporary blocking groups affixed on the sugar of nucleotides (see Section 4b).



**Figure 9.** A) Close-up view of the catalytic core of the RNA-cleaving DNzyme 8-17. Adopted from reference <sup>71</sup>; B) Representative structure (i) of the precatalytic complex formed by DNzyme 10-23 and its RNA substrate obtained from NMR-based *ab initio* calculations and the corresponding, simplified representation (ii). Reproduced with permission from reference <sup>72</sup>.

#### 4) Overview of other applications

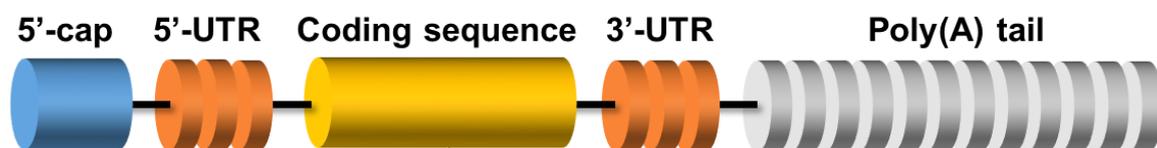
Polymerase-mediated synthesis of nucleobase-modified nucleic acids is a mild and highly versatile method that reaches numerous scientific fields and has resulted in a broad range of practical applications such as functional tagging, electrochemical sensing, or *in vivo* tracking and detection of oligonucleotides.<sup>73</sup> This method is now routinely used for DNA and RNA and starts to be extended to orthogonal, sugar-modified nucleic acids called XNAs. While it is well beyond the scope of this Chapter to cover all types of

applications of enzymatic synthesis of base-modified nucleic acids, this last Section will highlight some recent, emerging examples where this method plays an increasingly important role.

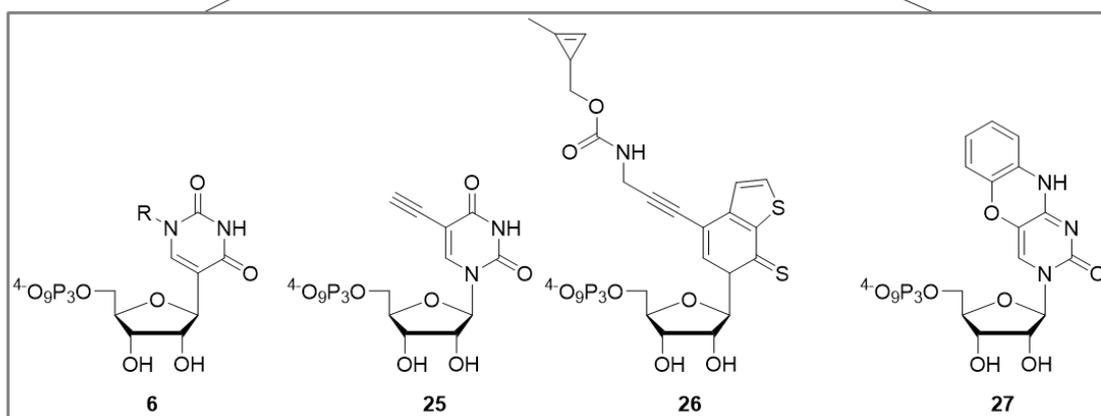
a) Generation of chemically modified mRNA vaccines:

Conceptually, in mRNA vaccines, the amino acid sequence composition of specific antigens is encoded into RNA which then relays this information to the ribosomal machinery of the host cells and thus serves as an indirect vector for triggering an immune response.<sup>74</sup> The recent SARS-CoV2 pandemic spurred the industrial and commercial development of mRNA-based vaccines and propelled these nucleic acids into the forefront of protective measures against this viral disease. mRNA vaccines are linear sequences of over 4000 nucleotides that contain five main elements (Figure 10A): a 5'-cap which stabilizes the mRNA and permits recruitment of the ribosome, a coding sequence which contains the information required for antigen production, 3'- and 5'-untranslated regions (UTR) that flank the coding sequence and increase protein synthesis efficacy, and a 5'-terminal poly(A) tail which further stabilizes mRNAs. Besides the essential lipid nanoparticle formulation which prevents mRNAs from rapid nuclease degradation and facilitates cellular delivery, mRNA vaccines contain a base-modified nucleoside triphosphate: N1-methyl-pseudouridine (**6** with R = CH<sub>3</sub> in Figures 3 and 10B). (see reference S15 of the Further readings section) N1-methyl-pseudouridine has a dual role since its presence in mRNAs significantly increases protein production efficiency and reduces the uncontrolled immune activation that synthetic mRNAs might trigger in host cells. Importantly, industrial production of mRNA vaccines predominantly relies on T3, T7, or SP6 phage RNA polymerase mediated synthesis where canonical uridine triphosphate is substituted with the corresponding N1-methyl-pseudouridine nucleotide. While N1-methyl-pseudouridine is a key constituent of mRNA vaccines, other base-modified nucleotides have been considered for the production of efficient candidates (Figure 10B).

A)



B)



**Figure 10.** A) Schematic representation of the key structural elements of *in vitro* transcribed mRNA vaccines and B) chemical structures of modified nucleotides used in the context of mRNAs.

There are several reasons why chemical modifications can be incorporated into mRNA scaffolds. Indeed, chemical modifications such as N1-methyl-pseudouridine, pseudouridine (R = H in nucleotide **6**), 2-thiouridine, or 5-methylcytosine (**9** in Figure 3) reduce the immunogenicity of the resulting transcripts by eluding pattern recognition receptors (*vide supra*). (see reference S26 of the Further readings section) In addition, modified nucleotides can enhance the rate of transcription and more importantly that of translation which ensures high protein production. Potentially, modified nucleotides can also enhance the stability of mRNAs in nanoparticle formulations and in biological media, particularly in the cytosol, and to functionally tag the resulting sequences with fluorophores for mechanistic and structural investigation purposes. Stabilization of therapeutic oligonucleotides is generally achieved by introducing a combination of backbone alterations such as phosphorothioates and sugar modifications including 2'-OMe or 2'-fluoro ribose units. A

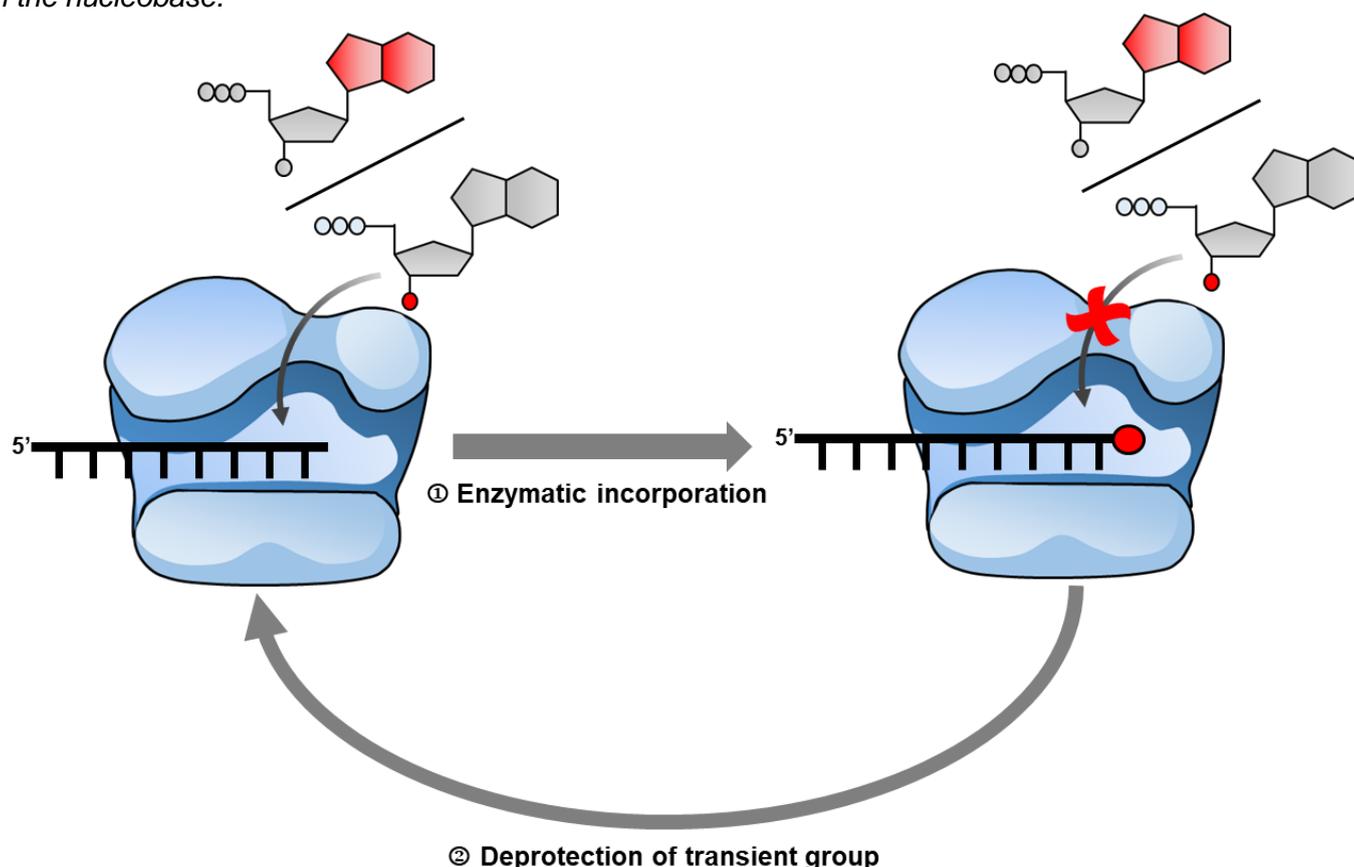
similar strategy can be adopted for mRNA vaccines however these modifications should not come at the expense of reduced translation efficiencies. Since the effect of modifications on nuclease resistance, translation capacity, and immune response is often unpredictable this needs to be evaluated in a case-to-case manner. For instance, the presence of phosphorothioate moieties (stereopure with all- $R_p$  configuration at all sites) enhances both the resistance against nuclease degradation and protein production,<sup>75</sup> while 2'-*O*-methylation stabilizes the resulting mRNAs but massively reduces translation efficiency.<sup>76</sup> So far, base-modified nucleotides differing from naturally occurring modification patterns have not been considered for improving the stability of mRNA/mRNA vaccines but it is to be expected that known nucleotide analogs, for instance bearing cationic amine side-chains, will be evaluated in the near future. On the other hand, various base-modified nucleotides have been incorporated into RNA for imaging and detection studies. In this context, 5-ethynyl uridine triphosphate (EUTP **25** in Figure 10B) bears a small alkyne moiety that acts as a synthetic handle for the grafting of additional functional groups by click chemistry. The presence of this substituent causes minimal perturbation in RNA and within the active site of RNA polymerases and has been employed to label mRNAs both *in vitro* and *in vivo*. For instance, EU-labelled mRNAs provided a means to investigate tail-length dynamics.<sup>77</sup> Once integrated in mRNA, EU-modified species were subjected to a click reaction with a corresponding azide-containing biotin moiety and subsequently captured by streptavidin-coated magnetic beads. The length of the poly(A) tail of trapped mRNA species could then be measured and quantified. This study revealed that the rate of deadenylation spans over three orders of magnitude and deadenylation-linked mechanisms appear to be the main contributors to this decay. Similarly, EUTP was shown to be compatible with *in vitro* transcription production of synthetic mRNAs and hence alkyne moieties can be incorporated into the various constituting elements and further labelled by click chemistry based methods. Interestingly, the corresponding 7-ethynyl-7-deazaadenosine 5'-triphosphate (EATP) is compatible with the yeast poly(A) polymerase and hence poly(A) tails can also be labelled with alkyne moieties using a similar approach. While labelling with ethynyl-containing nucleotides has met some success, also for the *in vivo* monitoring of mRNAs, it is partially limited by the inherent metabolic instability of azide functionalities, the steric hindrance surrounding ethynyl moieties which lowers the efficiency of the click reaction, and the need for cytotoxic copper to catalyze the click reaction. One approach to circumvent these limitations involves an extension of the genetic alphabet in RNA. Indeed, incorporation of an unnatural nucleotide into RNA provides a means to insert synthetic handles for further modification without compromising sites occupied by one of the canonical or naturally occurring modified nucleotides as is the case with EU. In addition, a careful choice of the synthetic moiety grafted on the unnatural nucleotide can also reduce the dependence to cytotoxic metal cations. To do so, one or two artificial dNaM-dTPT3 base pairs (Figure 6) can be incorporated into a dsDNA template using a six-letter PCR protocol.<sup>78</sup> Once installed in dsDNA, the dNaM partner of the modified base pairs will be specifically recognized by an RNA counterpart of dTPT3 during *in vitro* transcription mediated by the T7 RNA polymerase. Indeed, nucleotide **26** (Figure 10B) consists of a ribose sugar equipped with the thio N-glycoside containing base surrogate of dTPT3 along with a cyclopropane group which can undergo inverse electron demand Diels-Alder (IEDDA) cycloaddition reactions. Besides nucleotide **26**, pseudouridine **6**, and 5mC **9** (Figure 3) were simultaneously incorporated into mRNA during transcription. Once incorporated in mRNA molecules, the cyclopropane moiety of nucleotide **26** can undergo an iEDDA reaction with a tetrazine-conjugated fluorophore which enables *in-cell* mRNA visualization. Alternatively, the two-step protocol involving incorporation of a modified nucleotide followed by post-*in vitro* transcription derivatization can be avoided by using inherently fluorescent nucleotide such as the tricyclic cytosine analogue **27** (Figure 10B).<sup>73</sup> Nucleotide **27** contains a 1,3-diaza-2-oxophenoxazine base surrogate which displays both very high quantum yields ( $\Phi_F = 0.24$ ) and average brightness. Importantly, nucleotide **27** is an excellent substrate for the T7 RNA polymerase and its presence does not cause many transversions (i.e. A to G) during reverse transcription of modified RNA. Importantly, upon transfection of modified transcripts, protein synthesis is maintained suggesting that nucleotide **27** is compatible with translation. Due to its intrinsic fluorescent properties, mRNAs modified with nucleotide **27** can easily be detected in live cells by confocal microscopy.

These examples clearly highlight the potency of polymerizing base-modified nucleotides to label and detect mRNAs and the momentum created by the SARS-CoV2 pandemic will certainly catalyze the emergence of other modified mRNAs with enhanced properties.

## b) Controlled enzymatic synthesis

The main methods currently employed for the synthesis of modified nucleic acids, i.e. solid-phase synthesis and polymerization of modified nucleoside triphosphates, display non negligible limitations. Indeed, phosphoramidite-based chemistry is restricted in terms of sequence length and compatibility of functional groups, while depurination events might affect the quality of the resulting oligonucleotides. On the other hand, during polymerase-mediated synthesis, modified analogs will be guided by all complementary templating nucleotides which results in a homogenous modification pattern. Heterogenous modification

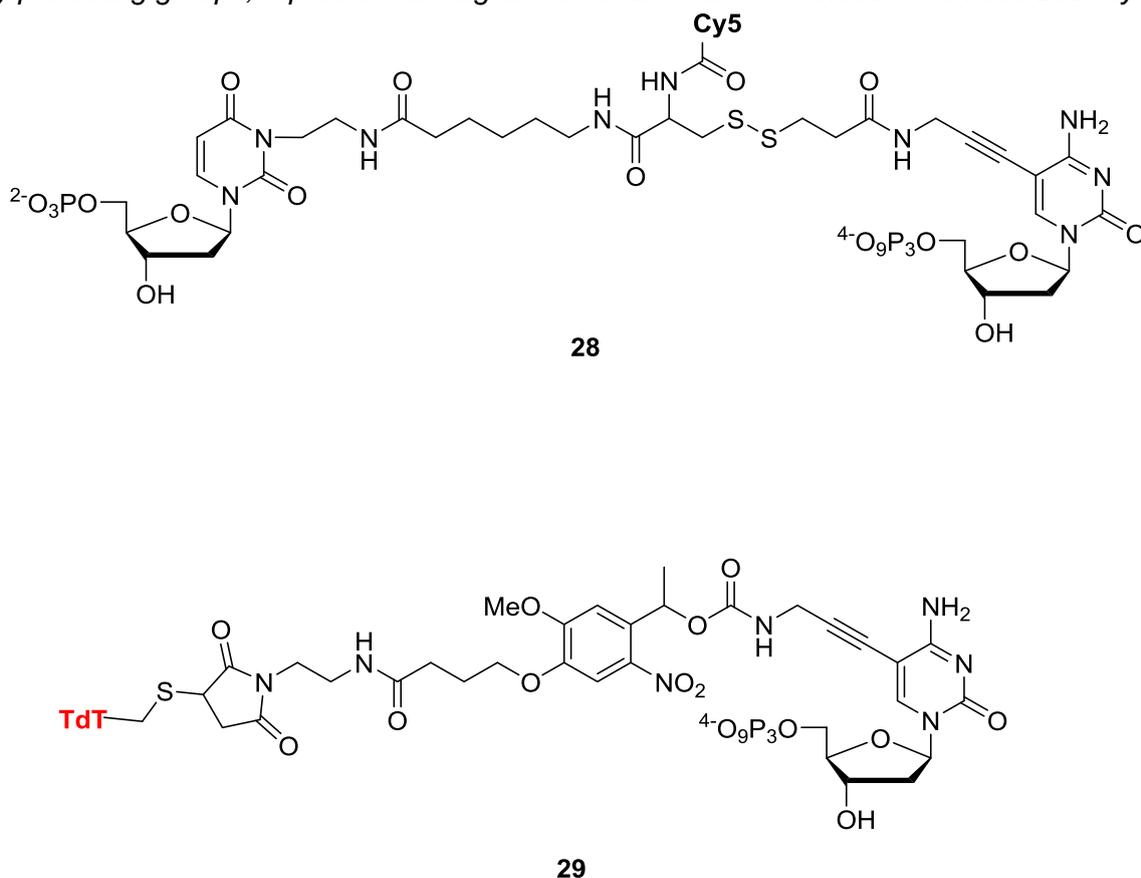
patterns might be obtained by using ratios of modified and/or canonical nucleotides but the resulting distribution will depend on the incorporation efficiencies of individual triphosphates and analysis by sequencing methods might be difficult and accessible only to short oligonucleotides. Controlled enzymatic synthesis of nucleic acids is an alternative method that combines elements of both strategies: temporarily blocked nucleoside triphosphates are incorporated into DNA, mainly by template-independent polymerases such as the terminal deoxynucleotidyl transferase (TdT).<sup>79</sup> The transient protecting groups ensure that only one nucleotide is incorporated during one synthetic cycle. After removal of excess polymerase and nucleotide, the transient group is deprotected resulting in a DNA oligonucleotide that can undergo subsequent synthetic cycles (Figure 11). Single incorporation of nucleotides can be achieved by either transiently blocking the 3'-OH moiety of the incoming triphosphate or by attaching bulky, reversible groups on the nucleobase.



**Figure 11.** Schematic representation of the working principle of controlled enzymatic synthesis. Red color indicates the location of the transient protecting group.

Appendage of reversible terminators at the 3'-OH of a nucleotide is the most common method, most likely because of the compatibility of such nucleotides with sequencing-by-synthesis methods.<sup>80, 81, 82</sup> However, such an approach is delicate since the presence of the 3'-O-protecting group might interact and directly interfere with side chains of the active site of polymerases which calls in for the necessity of engineered polymerases or SAR studies for the identification of suitable protecting groups. On the other hand, finely tuned, base-modified nucleotides are rather well tolerated by polymerases and several examples of such analogs have been reported for controlled enzymatic synthesis (Figure 12). For instance, a series of virtual terminator nucleotides was synthesized and shown to be particularly efficient at preventing additional incorporation events despite the presence of a free 3'-OH moiety (**28** in figure 12).<sup>83</sup> In these virtual terminators, modifications consisting of a Cy5 dye and nucleoside monophosphate inhibitor are attached to the C5 position of pyrimidines and the N7 location of N7-deaza-purine nucleotides through rigid propargylamino linker arms and a cleavable disulfide unit. The efficiency of the nucleotide inhibitor (deoxyuridine or deoxycytosine) attached to the nucleobase is proportional to the number of phosphate units present on the sugar moiety (i.e. no inhibition without phosphate and highest inhibition with a diphosphate moiety). Moreover, the presence of the large modification pattern reduced the catalytic efficiency of Kf (exo<sup>-</sup>) but did not lead to any misincorporation of the modified nucleotides. Importantly, due to the presence of the bulky side chain and the inhibiting nucleotide on the nucleobases, the TdT only incorporated a single modified nucleotide at the time. Once the modified nucleotide incorporated into DNA by standard PEX reactions, inhibition was reversed by reductive cleavage of the disulfide linkage and capping of the resulting thiol group. Removal of the bulky side chain and inhibiting nucleotide permitted controlled synthesis over numerous cycles with high fidelity. The remaining molecular scar on the nucleobase does not interfere with sequencing methods but this approach has not been extended to de

*novo* synthesis of DNA. In a related approach, the TdT polymerase itself is directly tethered to the nucleobase of a modified nucleotide (**29** in Figure 12) via a rigid linker arm and a photocleavable connector. The rationale behind such a construct is that the polymerase incorporates the nucleotide attached to the TdT via a surface-accessible cysteine residue into a DNA primer and remains covalently bound to the extended DNA. This covalent fixation of the TdT onto DNA prevents other polymerase-bound nucleotides from incorporating into DNA and thus strictly restricts the reaction to single incorporation events. Light irradiation (365 nm) cleaves the linker moiety and liberates the primer from the TdT-primer complex and the resulting free primer can be further extended by other TdT-nucleotides.<sup>84</sup> Repetition of such a two-step protocol permits controlled enzymatic synthesis of DNA oligonucleotides. The resulting DNA oligonucleotides can serve as templates for synthesis of complementary DNA by PCR despite the presence of the propargylamino scars. Hence, base-modified nucleotides, potentially complemented with 3'-O-temporary protecting groups, represent alluring and versatile tools for controlled *de novo* DNA synthesis.



**Figure 12.** Chemical structures of base-modified nucleotides used in controlled enzymatic synthesis.

## F. Conclusion

The chemical arsenal available to nucleic acids is rather limited and this negatively impacts on the scope of their application potential. For instance, nucleic acid therapeutics need to be stable in biological settings, aptamers have to be capable of recognizing their targets, and mRNA vaccines should display high translation efficiencies without concomitantly triggering an undesired immune response. Without introducing chemical modifications such feats cannot (or only with low efficiency) be accomplished by nucleic acids. Fortunately, versatile and robust, chemical and chemoenzymatic methods are available and permit an expansion of the chemical repertoire available to nucleic acids. Decorating nucleobases with alien functionalities is a particularly interesting approach due to its compatibility with polymerase-mediated synthesis. The exogenous groups can also be positioned so as to cause minimal interference with Watson-Crick-Franklin base pair formation and little perturbation of the overall three-dimensional structure of nucleic acids. As highlighted in this Chapter, this method has been applied to generating oligonucleotides equipped with epigenetic chemical modifications but also to improve the general properties of functional nucleic acids. Such a method has also been integrated in other emerging applications such as the functional tagging and chemical modification of mRNAs and in the development of controlled enzymatic synthesis.

As tools such as engineered polymerases and nanopore sequencing become more available, the diversity of chemistries available to nucleic acids will expand rapidly. Such an expansion might be further assisted by exploring other enzymes such as primases, transferases, replicases, and ligases to catalyze the polymerization of base-modified nucleic acids. It is therefore predictable that base-modified nucleic acids will be applied to other research directions such as the construction of metal nanowires, for the storage of

digital information, or for the development of synthetic light-harvesting supramolecular polymers which could be integrated in next-generation solar cells.

## G. References

1. Cahová H, Panattoni A, Kielkowski P, Fanfrlík J, Hocek M. 5-Substituted Pyrimidine and 7-Substituted 7-Deazapurine dNTPs as Substrates for DNA Polymerases in Competitive Primer Extension in the Presence of Natural dNTPs. *ACS Chem Biol* 2016, **11**(11): 3165-3171.
2. Kore AR, Yang B, Srinivasan B. Concise synthesis of 5-methyl-, 5-formyl-, and 5-carboxy analogues of 2'-deoxycytidine-5'-triphosphate. *Tetrahedron Lett* 2013, **54**(39): 5325-5327.
3. Steigenberger B, Schiesser S, Hackner B, Brandmayr C, Laube SK, Steinbacher J, *et al.* Synthesis of 5-Hydroxymethyl-, 5-Formyl-, and 5-Carboxycytidine-triphosphates and Their Incorporation into Oligonucleotides by Polymerase Chain Reaction. *Org Lett* 2013, **15**(2): 366-369.
4. Sun Q, Sun J, Gong S-S, Wang C-J, Pu S-Z, Feng F-D. Efficient synthesis of 5-hydroxymethyl-, 5-formyl-, and 5-carboxyl-2'-deoxycytidine and their triphosphates. *RSC Adv* 2014, **4**(68): 36036-36039.
5. Nguyen H-K, Asseline U, Dupret D, Nguyen TT. Studies towards the design of a modified GC base pair with stability similar to that of the AT base pair. *Tetrahedron Lett* 1997, **38**(23): 4083-4086.
6. Flores-Juárez CR, González-Jasso E, Antaramian A, Pless RC. Capacity of N4-methyl-2'-deoxycytidine 5'-triphosphate to sustain the polymerase chain reaction using various thermostable DNA polymerases. *Anal Biochem* 2013, **438**(1): 73-81.
7. Jakubovska J, Tauraitė D, Birštonas L, Meškys R. N4-acyl-2'-deoxycytidine-5'-triphosphates for the enzymatic synthesis of modified DNA. *Nucleic Acids Res* 2018, **46**(12): 5911-5923.
8. Vaníková Z, Hocek M. Polymerase Synthesis of Photocaged DNA Resistant against Cleavage by Restriction Endonucleases. *Angew Chem Int Ed* 2014, **53**(26): 6734-6737.
9. Chakrapani A, Ruiz-Larrabeiti O, Pohl R, Svoboda M, Krásný L, Hocek M. Glucosylated 5-hydroxymethylpyrimidines as epigenetic DNA bases regulating transcription and restriction cleavage. *Chem Eur J* 2022, **28**(31): e202200911.
10. Engel JD, von Hippel PH. Effects of methylation on the stability of nucleic acid conformations. Studies at the polymer level. *J Biol Chem* 1978, **253**(3): 927-934.
11. Kuwahara M, Hososhima S, Takahata Y, Kitagata R, Shoji A, Hanawa K, *et al.* Simultaneous incorporation of three different modified nucleotides during PCR. *Nucleic Acids Res Suppl* 2003(3): 37-38.
12. Choi J-S, Dasari A, Hu P, Benkovic SJ, Berdis AJ. The use of modified and non-natural nucleotides provide unique insights into pro-mutagenic replication catalyzed by polymerase  $\epsilon$ . *Nucleic Acids Res* 2015, **44**(3): 1022-1035.
13. Timofeev EN, Mikhailov SN, Zuev AN, Efimtseva EV, Herdewijn P, Somers RL, *et al.* Oligodeoxynucleotides Containing 2'-Deoxy-1-methyladenosine and Dimroth Rearrangement. *Helv Chem Acta* 2007, **90**(5): 928-937.
14. Mikhailov SN, Rozenski J, Efimtseva EV, Busson R, Van Aerschot A, Herdewijn P. Chemical incorporation of 1-methyladenosine into oligonucleotides. *Nucleic Acids Res* 2002, **30**(5): 1124-1131.
15. Nakano S-i, Uotani Y, Sato Y, Oka H, Fujii M, Sugimoto N. Conformational changes of the phenyl and naphthyl isocyanate-DNA adducts during DNA replication and by minor groove binding

molecules. *Nucleic Acids Res* 2013, **41**(18): 8581-8590.

16. Hendler SS, Fuerer E, Srinivasan PR. Synthesis and chemical properties of monomers and polymers containing 7-methylguanine and an investigation of their substrate or template properties for bacterial deoxyribonucleic acid or ribonucleic acid polymerases. *Biochemistry* 1970, **9**(21): 4141-4153.
17. Porter KW, Tomasz J, Huang F, Sood A, Shaw BR. N7-Cyanoborane-2'-Deoxyguanosine 5'-Triphosphate Is a Good Substrate for DNA Polymerase. *Biochemistry* 1995, **34**(37): 11963-11969.
18. Shanmugasundaram M, Senthilvelan A, Xiao Z, Kore AR. An Efficient Protection-Free One-Pot Chemical Synthesis of Modified Nucleoside-5'-Triphosphates. *Nucleosides, Nucleotides & Nucleic Acids* 2016, **35**(7): 356-362.
19. Karikó K, Muramatsu H, Welsh FA, Ludwig J, Kato H, Akira S, *et al.* Incorporation of pseudouridine into mRNA yields superior nonimmunogenic vector with increased translational capacity and biological stability. *Mol Ther* 2008, **16**(11): 1833-1840.
20. Andries O, Mc Cafferty S, De Smedt SC, Weiss R, Sanders NN, Kitada T. N1-methylpseudouridine-incorporated mRNA outperforms pseudouridine-incorporated mRNA by providing enhanced protein expression and reduced immunogenicity in mammalian cell lines and mice. *J Control Release* 2015, **217**: 337-344.
21. Svitkin YV, Cheng YM, Chakraborty T, Presnyak V, John M, Sonenberg N. N1-methylpseudouridine in mRNA enhances translation through eIF2 $\alpha$ -dependent and independent mechanisms by increasing ribosome density. *Nucleic Acids Res* 2017, **45**(10): 6023-6036.
22. Gabbai A, Marcus I, Falbriard JG, Posternak H. Dérivés de nucléotides d'intérêt biologique VIII. Synthèse de quelques analogues de l'acide adénosine-triphosphorique (ATP). *Helv Chim Acta* 1971, **54**(7): 2133-2141.
23. Karikó K, Buckstein M, Ni H, Weissman D. Suppression of RNA Recognition by Toll-like Receptors: The Impact of Nucleoside Modification and the Evolutionary Origin of RNA. *Immunity* 2005, **23**(2): 165-175.
24. Kim KL, van Galen P, Hovestadt V, Rahme GJ, Andreishcheva EN, Shinde A, *et al.* Systematic detection of m6A-modified transcripts at single-molecule and single-cell resolution. *Cell Rep Methods* 2021, **1**(5): 100061.
25. Bartee D, Nance KD, Meier JL. Site-Specific Synthesis of N4-Acetylcytidine in RNA Reveals Physiological Duplex Stabilization. *J Am Chem Soc* 2022, **144**(8): 3487-3496.
26. Sinclair WR, Arango D, Shrimp JH, Zengeya TT, Thomas JM, Montgomery DC, *et al.* Profiling Cytidine Acetylation with Specific Affinity and Reactivity. *ACS Chem Biol* 2017, **12**(12): 2922-2926.
27. Nance KD, Gamage ST, Alam MM, Yang A, Levy MJ, Link CN, *et al.* Cytidine acetylation yields a hypoinflammatory synthetic messenger RNA. *Cell Chem Biol* 2022, **29**(2): 312-320.e317.
28. Pandolfini L, Barbieri I, Bannister AJ, Hendrick A, Andrews B, Webster N, *et al.* METTL1 Promotes let-7 MicroRNA Processing via m7G Methylation. *Mol Cell* 2019, **74**(6): 1278-1290.e1279.
29. Ren J, Goss DJ. Synthesis of a Fluorescent 7-Methylguanosine Analog and a Fluorescence Spectroscopic Study of its Reaction with Wheat germ Cap Binding Proteins. *Nucleic Acids Res* 1996, **24**(18): 3629-3634.
30. Milisavljevič N, Perlíková P, Pohl R, Hocek M. Enzymatic synthesis of base-modified RNA by T7 RNA polymerase. A systematic study and comparison of 5-substituted pyrimidine and 7-substituted 7-deazapurine nucleoside triphosphates as substrates. *Org Biomol Chem* 2018, **16**(32): 5800-5807.
31. Tuerk C, Gold L. Systematic Evolution of Ligands by Exponential Enrichment: RNA Ligands to Bacteriophage T4 DNA Polymerase. *Science* 1990, **249**(4968): 505-510.

32. Ellington AD, Szostak JW. In vitro selection of RNA molecules that bind specific ligands. *Nature* 1990, **346**(6287): 818-822.
33. Arangundy-Franklin S, Taylor AI, Porebski BT, Genna V, Peak-Chew S, Vaisman A, *et al.* A synthetic genetic polymer with an uncharged backbone chemistry based on alkyl phosphonate nucleic acids. *Nat Chem* 2019, **11**(6): 533-542.
34. Gawande BN, Rohloff JC, Carter JD, Carlowitz Iv, Zhang C, Schneider DJ, *et al.* Selection of DNA aptamers with two modified bases. *Proc Natl Acad Sci USA* 2017, **114**(11): 2898-2903.
35. Latham JA, Johnson R, Toole JJ. The application of a modified nucleotide in aptamer selection: novel thrombin aptamers containing -(1 -pentynyl)-2'-deoxyuridine. *Nucleic Acids Res* 1994, **22**(14): 2817-2822.
36. Gold L, Ayers D, Bertino J, Bock C, Bock A, Brody EN, *et al.* Aptamer-Based Multiplexed Proteomic Technology for Biomarker Discovery. *PLoS One* 2010, **5**(12): e15004.
37. Wolk SK, Mayfield WS, Gelinis AD, Astling D, Guillot J, Brody EN, *et al.* Modified nucleotides may have enhanced early RNA catalysis. *Proc Natl Acad Sci USA* 2020, **117**(15): 8236-8242.
38. Sun BB, Maranville JC, Peters JE, Stacey D, Staley JR, Blackshaw J, *et al.* Genomic atlas of the human plasma proteome. *Nature* 2018, **558**(7708): 73-79.
39. Tanaka K, Okuda T, Kasahara Y, Obika S. Base-modified aptamers obtained by cell-internalization SELEX facilitate cellular uptake of an antisense oligonucleotide. *Mol Ther Nucleic Acids* 2021, **23**: 440-449.
40. Cheung Y-W, Röthlisberger P, Mechaly AE, Weber P, Levi-Acobas F, Lo Y, *et al.* Evolution of abiotic cubane chemistries in a nucleic acid aptamer allows selective recognition of a malaria biomarker. *Proc Natl Acad Sci USA* 2020, **117**(29): 16790-16798.
41. Minagawa H, Kataoka Y, Kuwahara M, Horii K, Shiratori I, Waga I. A high affinity modified DNA aptamer containing base-appended bases for human  $\beta$ -defensin. *Anal Biochem* 2020, **594**: 113627.
42. Kabza AM, Sczepanski JT. An L-RNA Aptamer with Expanded Chemical Functionality that Inhibits MicroRNA Biogenesis. *ChemBioChem* 2017, **18**(18): 1824-1827.
43. Mei H, Liao J-Y, Jimenez RM, Wang Y, Bala S, McCloskey C, *et al.* Synthesis and Evolution of a Threose Nucleic Acid Aptamer Bearing 7-Deaza-7-Substituted Guanosine Residues. *J Am Chem Soc* 2018, **140**(17): 5706-5713.
44. Ren X, Gelinis AD, Linehan M, Iwasaki A, Wang W, Janjic N, *et al.* Evolving A RIG-I Antagonist: A Modified DNA Aptamer Mimics Viral RNA. *J Mol Biol* 2021, **433**(21): 167227.
45. Manandhar M, Chun E, Romesberg FE. Genetic Code Expansion: Inception, Development, Commercialization. *J Am Chem Soc* 2021, **143**(13): 4859-4878.
46. Morales JC, Kool ET. Efficient replication between non-hydrogen-bonded nucleoside shape analogs. *Nat Struct Biol* 1998, **5**(11): 950-954.
47. Kimoto M, Yamashige R, Matsunaga K-i, Yokoyama S, Hirao I. Generation of high-affinity DNA aptamers using an expanded genetic alphabet. *Nat Biotechnol* 2013, **31**(5): 453-457.
48. Futami K, Kimoto M, Lim YWS, Hirao I. Genetic Alphabet Expansion Provides Versatile Specificities and Activities of Unnatural-Base DNA Aptamers Targeting Cancer Cells. *Mol Ther Nucleic Acids* 2019, **14**: 158-170.
49. Hutter D, Benner SA. Expanding the Genetic Alphabet: Non-Epimerizing Nucleoside with the pyDDA Hydrogen-Bonding Pattern. *J Org Chem* 2003, **68**(25): 9839-9842.
50. Yang Z, Chen F, Alvarado JB, Benner SA. Amplification, Mutation, and Sequencing of a Six-Letter Synthetic Genetic System. *J Am Chem Soc* 2011, **133**(38): 15105-15112.

51. Sefah K, Yang Z, Bradley KM, Hoshika S, Jiménez E, Zhang L, *et al.* In vitro selection with artificial expanded genetic information systems. *Proc Natl Acad Sci USA* 2014, **111**(4): 1449-1454.
52. Zhang L, Yang Z, Sefah K, Bradley KM, Hoshika S, Kim M-J, *et al.* Evolution of Functional Six-Nucleotide DNA. *J Am Chem Soc* 2015, **137**(21): 6734-6737.
53. Hoshika S, Leal NA, Kim M-J, Kim M-S, Karalkar NB, Kim H-J, *et al.* Hachimoji DNA and RNA: A genetic system with eight building blocks. *Science* 2019, **363**(6429): 884-887.
54. Li L, Degardin M, Lavergne T, Malyshev DA, Dhimi K, Ordoukhanian P, *et al.* Natural-like Replication of an Unnatural Base Pair for the Expansion of the Genetic Alphabet and Biotechnology Applications. *J Am Chem Soc* 2014, **136**(3): 826-829.
55. Malyshev DA, Dhimi K, Lavergne T, Chen T, Dai N, Foster JM, *et al.* A semi-synthetic organism with an expanded genetic alphabet. *Nature* 2014, **509**(7500): 385-388.
56. Wang H, Wang L, Ma N, Zhu W, Huo B, Zhu A, *et al.* Access to Photostability-Enhanced Unnatural Base Pairs via Local Structural Modifications. *ACS Synth Biol* 2022, **11**(1): 334-342.
57. 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**(10): 794-800.
58. Flamme M, Röthlisberger P, Levi-Acobas F, Chawla M, Oliva R, Cavallo L, *et al.* Enzymatic Formation of an Artificial Base Pair Using a Modified Purine Nucleoside Triphosphate. *ACS Chem Biol* 2020, **15**(11): 2872-2884.
59. Switzer C, Sinha S, Kim PH, Heuberger BD. A Purine-like Nickel(II) Base Pair for DNA. *Angew Chem Int Ed* 2005, **44**(10): 1529-1532.
60. Takezawa Y, Nakama T, Shionoya M. Enzymatic Synthesis of Cu(II)-Responsive Deoxyribozymes through Polymerase Incorporation of Artificial Ligand-Type Nucleotides. *J Am Chem Soc* 2019, **141**(49): 19342-19350.
61. Nakama T, Takezawa Y, Sasaki D, Shionoya M. Allosteric Regulation of DNAzyme Activities through Intrastrand Transformation Induced by Cu(II)-Mediated Artificial Base Pairing. *J Am Chem Soc* 2020, **142**(22): 10153-10162.
62. Levi-Acobas F, McKenzie LK, Hollenstein M. Towards polymerase-mediated synthesis of artificial RNA–DNA metal base pairs. *New J Chem* 2022, **46**(10): 4871-4876.
63. Santoro SW, Joyce GF. A general purpose RNA-cleaving DNA enzyme. *Proc Natl Acad Sci USA* 1997, **94**(9): 4262-4266.
64. Wang Y, Nguyen K, Spitale RC, Chaput JC. A biologically stable DNAzyme that efficiently silences gene expression in cells. *Nat Chem* 2021, **13**(4): 319-326.
65. Santoro SW, Joyce GF, Sakthivel K, Gramatikova S, Barbas CF. RNA Cleavage by a DNA Enzyme with Extended Chemical Functionality. *J Am Chem Soc* 2000, **122**(11): 2433-2439.
66. Perrin DM, Garestier T, Hélène C. Bridging the Gap between Proteins and Nucleic Acids: A Metal-Independent RNaseA Mimic with Two Protein-Like Functionalities. *J Am Chem Soc* 2001, **123**(8): 1556-1563.
67. Hollenstein M, Hipolito CJ, Lam CH, Perrin DM. A self-cleaving DNA enzyme modified with amines, guanidines and imidazoles operates independently of divalent metal cations ( $M^{2+}$ ). *Nucleic Acids Res* 2009, **37**(5): 1638-1649.
68. Wang YJ, Liu EK, Lam CH, Perrin DM. A densely modified  $M^{2+}$ -independent DNAzyme that cleaves RNA efficiently with multiple catalytic turnover. *Chem Sci* 2018, **9**(7): 1813-1821.

69. Zhou C, Avins JL, Klauser PC, Brandsen BM, Lee Y, Silverman SK. DNA-Catalyzed Amide Hydrolysis. *J Am Chem Soc* 2016, **138**(7): 2106-2109.
70. Tarasow TM, Tarasow SL, Eaton BE. RNA-catalysed carbon-carbon bond formation. *Nature* 1997, **389**(6646): 54-57.
71. Liu H, Yu X, Chen Y, Zhang J, Wu B, Zheng L, *et al.* Crystal structure of an RNA-cleaving DNAzyme. *Nat Commun* 2017, **8**(1): 2006.
72. Borggräfe J, Victor J, Rosenbach H, Viegas A, Gertzen CGW, Wuebben C, *et al.* Time-resolved structural analysis of an RNA-cleaving DNA catalyst. *Nature* 2022, **601**(7891): 144-149.
73. Baladi T, Nilsson JR, Gallud A, Celauro E, Gasse C, Levi-Acobas F, *et al.* Stealth Fluorescence Labeling for Live Microscopy Imaging of mRNA Delivery. *J Am Chem Soc* 2021, **143**(14): 5413-5424.
74. Martinon F, Krishnan S, Lenzen G, Magné R, Gomard E, Guillet J-G, *et al.* Induction of virus-specific cytotoxic T lymphocytes in vivo by liposome-entrapped mRNA. *Eur J Immunol* 1993, **23**(7): 1719-1722.
75. Kawaguchi D, Kodama A, Abe N, Takebuchi K, Hashiya F, Tomoike F, *et al.* Phosphorothioate Modification of mRNA Accelerates the Rate of Translation Initiation to Provide More Efficient Protein Synthesis. *Angew Chem Int Ed* 2020, **59**(40): 17403-17407.
76. Choi J, Indrisiunaite G, DeMirici H, Jeong K-W, Wang J, Petrov A, *et al.* 2'-O-methylation in mRNA disrupts tRNA decoding during translation elongation. *Nat Struct Mol Biol* 2018, **25**(3): 208-216.
77. Eisen TJ, Eichhorn SW, Subtelny AO, Lin KS, McGeary SE, Gupta S, *et al.* The Dynamics of Cytoplasmic mRNA Metabolism. *Mol Cell* 2020, **77**(4): 786-799.e710.
78. Bornewasser L, Domnick C, Kath-Schorr S. Stronger together for in-cell translation: natural and unnatural base modified mRNA. *Chem Sci* 2022, **13**(17): 4753-4761.
79. Lee HH, Kalhor R, Goela N, Bolot J, Church GM. Terminator-free template-independent enzymatic DNA synthesis for digital information storage. *Nat Commun* 2019, **10**(1): 2383.
80. Hutter D, Kim M-J, Karalkar N, Leal NA, Chen F, Guggenheim E, *et al.* Labeled nucleoside triphosphates with reversibly terminating aminoalkoxyl groups. *Nucleosides Nucleotides Nucleic Acids* 2010, **29**(11): 879-895.
81. Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG, *et al.* Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* 2008, **456**(7218): 53-59.
82. Flamme M, Hanlon S, Marzuoli I, Püntener K, Sladojevich F, Hollenstein M. Evaluation of 3'-phosphate as a transient protecting group for controlled enzymatic synthesis of DNA and XNA oligonucleotides. *Commun Chem* 2022, **5**: 68.
83. Bowers J, Mitchell J, Beer E, Buzby PR, Causey M, Efcavitch JW, *et al.* Virtual terminator nucleotides for next-generation DNA sequencing. *Nat Methods* 2009, **6**(8): 593-595.
84. Palluk S, Arlow DH, de Rond T, Barthel S, Kang JS, Bector R, *et al.* De novo DNA synthesis using polymerase-nucleotide conjugates. *Nat Biotechnol* 2018, **36**(7): 645-650.

## H. Further Readings

- S1. Barbieri I, Kouzarides T. Role of RNA modifications in cancer. *Nat Rev Cancer* 2020, **20**(6): 303-322.
- S2. McKenzie LK, El-Khoury R, Thorpe JD, Damha MJ, Hollenstein M. Recent progress in non-native nucleic acid modifications. *Chem Soc Rev* 2021, **50**(8): 5126-5164.
- S3. Hofer A, Liu ZJ, Balasubramanian S. Detection, Structure and Function of Modified DNA Bases. *J Am*

*Chem Soc* 2019, **141**(16): 6420-6429.

S4. Raper AT, Reed AJ, Suo ZC. Kinetic Mechanism of DNA Polymerases: Contributions of Conformational Dynamics and a Third Divalent Metal Ion. *Chem Rev* 2018, **118**(12): 6000-6025.

S5. Houlihan G, Arangundy-Franklin S, Holliger P. Engineering and application of polymerases for synthetic genetics. *Curr Opin Biotechnol* 2017, **48**: 168-179.

S6. Raiber E-A, Hardisty R, van Delft P, Balasubramanian S. Mapping and elucidating the function of modified bases in DNA. *Nat Rev Chem* 2017, **1**(9): 0069.

S7. Kumar S, Chinnusamy V, Mohapatra T. Epigenetics of Modified DNA Bases: 5-Methylcytosine and Beyond. *Front Genet* 2018, **9**: 640.

S8. Kohli RM, Zhang Y. TET enzymes, TDG and the dynamics of DNA demethylation. *Nature* 2013, **502**(7472): 472-479.

S9. Ludwig J. A new route to nucleoside 5'-triphosphates. *Acta Biochim Biophys Acad Sci Hung* 1981, **16**(3-4): 131-133.

S10. Rodriguez F, Yushenova IA, DiCorpo D, Arkhipova IR. Bacterial N4-methylcytosine as an epigenetic mark in eukaryotic DNA. *Nat Commun* 2022, **13**(1): 1072.

S11. Nguyen H-K, Auffray P, Asseline U, Durand M, Maurizot J-C, Thuong NT, *et al.* The stability of duplexes involving AT and/or G4EtC base pairs is not dependent on their AT/G4EtC ratio content. Implication for DNA sequencing by hybridization. *Nucleic Acids Res* 1998, **26**(18): 4249-4258.

S12. Pfaffeneder T, Spada F, Wagner M, Brandmayr C, Laube SK, Eisen D, *et al.* Tet oxidizes thymine to 5-hydroxymethyluracil in mouse embryonic stem cell DNA. *Nat Chem Biol* 2014, **10**(7): 574-581.

S13. Borchardt EK, Martinez NM, Gilbert WV. Regulation and Function of RNA Pseudouridylation in Human Cells. *Annu Rev Genet* 2020, **54**(1): 309-336.

S14. Pastor F, Berraondo P, Etxeberria I, Frederick J, Sahin U, Gilboa E, *et al.* An RNA toolbox for cancer immunotherapy. *Nat Rev Drug Discov* 2018, **17**(10): 751-767.

S15. Nance KD, Meier JL. Modifications in an Emergency: The Role of N1-Methylpseudouridine in COVID-19 Vaccines. *ACS Cent Sci* 2021, **7**(5): 748-756.

S16. Zhang C, Fu J, Zhou Y. A Review in Research Progress Concerning m6A Methylation and Immunoregulation. *Front Immunol* 2019, **10**: 922.

S17. Fang E, Liu X, Li M, Zhang Z, Song L, Zhu B, *et al.* Advances in COVID-19 mRNA vaccine development. *Sig Transduct Target Ther* 2022, **7**(1): 94.

S18. Hocek M. Enzymatic Synthesis of Base-Functionalized Nucleic Acids for Sensing, Cross-linking, and Modulation of Protein–DNA Binding and Transcription. *Acc Chem Res* 2019, **52**(6): 1730-1737.

S19. Hottin A, Marx A. Structural Insights into the Processing of Nucleobase-Modified Nucleotides by DNA Polymerases. *Acc Chem Res* 2016, **49**(3): 418-427.

S20. Chaput JC. Redesigning the Genetic Polymers of Life. *Acc Chem Res* 2021, **54**(4): 1056-1065.

S21. Gelin AD, Davies DR, Janjic N. Embracing proteins: structural themes in aptamer–protein complexes. *Curr Opin Struct Biol* 2016, **36**: 122-132.

S22. Kimoto M, Hirao I. Genetic alphabet expansion technology by creating unnatural base pairs. *Chem Soc Rev* 2020, **49**(21): 7602-7626.

S23. Karalkar NB, Benner SA. The challenge of synthetic biology. Synthetic Darwinism and the aperiodic crystal structure. *Curr Opin Chem Biol* 2018, **46**: 188-195.

S24. Seo YJ, Hwang GT, Ordoukhanian P, Romesberg FE. Optimization of an Unnatural Base Pair toward Natural-Like Replication. *J Am Chem Soc* 2009, **131**(9): 3246-3252.

S25. Silverman SK. Catalytic DNA: Scope, Applications, and Biochemistry of Deoxyribozymes. *Trends Biochem Sci* 2016, **41**(7): 595-609.

S26. Chaudhary N, Weissman D, Whitehead KA. mRNA vaccines for infectious diseases: principles, delivery and clinical translation. *Nat Rev Drug Discov* 2021, **20**(11): 817-838.