Non-natural Nucleic Acids for Synthetic Biology

Daniel H. Appella
Laboratory of Bioorganic Chemistry, NIDDK, NIH, DHHS, Bethesda, MD 20892
Daniel H. Appella: appellad@niddk.nih.gov

Abstract

Genetic manipulation is an important facet of synthetic biology but can be complicated by undesired nuclease degradation. Incorporating non-natural nucleic acids into a gene could convey resistance to nucleases and promote expression. The compatibility of non-natural nucleosides with polymerases is reviewed with a focus on results from the past two years. Details are provided about how the different systems could be useful in synthetic biology.

Synthetic biology aims to fabricate living systems and at the same time achieve a better understanding of life in general [1-3]. A number of recent reviews develop the finer points of synthetic biology [4-6], which include bioengineering, biotechnology, molecular evolution, and systems biology [7]. The deletion and insertion of genes is often an important component of synthetic biology, and the use of non-natural nucleic acids could facilitate genetic manipulation. Over the past few decades, a number of chemistry-based research groups have clearly demonstrated that DNA and RNA are not the only molecular systems capable of storing genetic information [8-10]. At the same time, studies on polymerase replication and transcription of oligonucleotide sequences reveal that chemical entities other than the typical nucleoside triphosphates can be tolerated to varying degrees [11]. Therefore, a non-natural genetic system could be considered as the carrier of genetic information in a synthetic cell. It is intriguing to consider the possibility that a cell could be designed to rely on a genome in which DNA and RNA are replaced by alternative chemical entities.

Most synthetic biology borrows and/or hijacks the existing components of cells while reprogramming the hosts genome, but this process can be complicated by degradation of the non-natural gene. Some of the recent problems in accomplishing gene transfer between species of bacteria and yeast were due to restriction endonuclease activity that needed to be suppressed before success was realized [12]. A possible advantage to using non-natural alternatives to DNA and RNA in synthetic biology would be added stability to endogenous proteins that are eager to degrade natural nucleic acids. Therefore, a gene partially composed of non-natural nucleic acids could have a higher chance of being expressed due to its resistance to degradation. The challenge of incorporating non-natural nucleosides has typically stemmed from their incompatibility with biological machinery that replicates nucleic acids, namely DNA and RNA polymerases. Recent advances, however, point to the possibility of overcoming such impediments.

For application in synthetic biology a non-natural nucleic acid must be recognized and properly used by polymerases while not being degraded by nucleases. The majority of nucleic acid mimics developed over the past few decades were made for applications in the field of medicine and drug discovery, and such molecules were designed to inhibit cellular machinery associated with progression of a disease [13]. For synthetic biology, however, only a handful of nucleic acid mimics are tolerated in gene replication and transcription in bacteria. This review specifically focuses on the examples of DNA and RNA mimics that satisfy one or more of the criteria below:
1. There is some type of base pairing between the non-natural nucleic acid and a complementary oligonucleotide sequence.

2. There is a mechanism to incorporate the non-natural nucleic acid across from a template.

3. The information copied from the synthetic gene can be converted into a RNA message.

Before describing the specific examples of the different analogs, there are a few caveats to consider. Most of the examples below satisfy criteria 1; in some cases Watson-Crick hydrogen bonding between canonical base pairs is maintained but in other examples there are non-natural base pairs that rely on alternative hydrogen bonding arrangements or on hydrophobic interactions. Criteria 2 would be important in engineering genes that need to be copied as cells divide. This criterion is achieved with varying degrees of efficiency, some very good, some mediocre, and others are too preliminary to thoroughly evaluate. There are only a few cases where a chemically different form of messenger RNA is translated by ribosomes (which satisfies criteria 3). Additional complications of selecting the proper cell membrane and delivering the non-natural nucleosides across a cell membrane are not evaluated in this review, but recent work of Szostak and Luisi addresses some of these concerns in the context of synthetic biology [14,15]. This review focuses primarily on incorporation of analogs into DNA and RNA; the stability of non-natural oligonucleotides to nuclease degradation can be found in recent reviews covering the areas of antisense and siRNA research [16-18]. Finally, this article should be regarded as a partial survey of the chemical systems that could satisfy some of the requirements important to synthetic biology.

**Phosphate Analogs**

Synthetic changes to the phosphodiester backbone of DNA and RNA are good initial options to explore for applications in synthetic biology [16]. Numerous alternatives to this chemical linkage have been investigated with the explicit intention of stabilizing the modified oligonucleotide to nuclease degradation [19]. In some cases, the effects of such changes on polymerase activity have also been studied. The most common phosphodiester analog is the phosphorothioate in which one phosphate oxygen is replaced with a sulfur atom (Figure 1A). These derivatives have been extensively examined for antisense activity [13]. A number of DNA polymerases accept a triphosphate version of this derivative in which the sulfur is present at the α-position of the triphosphate [11]. Phosphorothioate linkages are chiral, with a stereogenic center present at the phosphorous atom. Mechanistic studies have demonstrated that several polymerases selectively use one stereoisomer of a sulfur-modified nucleoside triphosphate (the \( S^\alpha \) configuration) and that polymerization proceeds with inversion of configuration at phosphorous [20]. Recent examination of this mechanism by Benner indicates that the stereochemical requirements may not be so rigid for all polymerases, especially when a sulfur-modified nucleoside triphosphate contains a non-natural nucleobase (which in Banners work was the nucleobase P, Figure 2C) [21]. Ribosomes also accept mRNA messages with the phosphorothioate modification [22]. Recent reports have even revealed that certain bacteria naturally incorporate phosphorothioates into their genomes [23], and the genes encoding this process have been determined [24,25]. An alternative to phosphorothioates are the boranophosphates, a linkage that is similarly tolerated by polymerases and ribosomes [26]. The ability of polymerases to incorporate phosphonate nucleosides has been recently examined [27]. These derivatives contain an extra methylene group between the 5′ oxygen and the phosphorous atom and are highly resistant to nuclease degradation. A mutant DNA polymerase (Therminator) was able to incorporate 20 consecutive adenine-bearing phosphonate units across from a DNA duplex that had a thymine overhang of the same number. A subsequent study indicated that
adenine- and cytosine-phosphonate derivatives are favored by polymerases for incorporation while the thymine and uracil versions are incorporated to a limited extent. In a competition study, the natural adenine nucleoside triphosphate is still preferred compared to the phosphonate version [28]. Continued studies with these analogues, along with efforts of polymerase engineering [29], could provide optimum conditions for extended polymerization of phosphonates.

**Aptamers**

The field of aptamers has also tackled the incorporation of synthetic nucleotides, focusing on both phosphodiester replacements as well as modified bases [30,31]. Aptamers are single-stranded oligonucleotides that are generated through an *in vitro* selection and evolution process called SELEX (which stands for Systematic Evolution of Ligands by Exponential enrichment). The evolution process for SELEX depends on employing several rounds of enzymatic amplification of oligonucleotide sequences that are selected to have good binding properties to a particular target. Since aptamers composed of DNA and RNA are themselves susceptible to nuclease degradation, the field has made significant efforts to find synthetic derivatives that are sufficiently stable for therapeutic applications. Therefore, techniques to perform SELEX with modified nucleosides has been developed, and also recently reviewed [30,31]. To briefly summarize, *in vitro* evolution using phosphorothioates has been successfully accomplished [32]. In addition, a number of mutant T7 RNA polymerases have been developed that are tolerant of substitutions at the 2′ position of the furanose ring (Figure 1B). Substitutions that may be attached to C2′ include: fluorne, an amine, and a methoxy group [33]. One of the most impressive accomplishments from aptamer development is the extensive number of modifications that are tolerated at the 5-position of uracil (Figure 1C) [34]. A wide variety of sidechains may be attached to this position, ranging from simple hydrophobic groups to extended sidechains terminated by polar groups. These modified nucleotides are tolerated by a variety of T7 RNA polymerases [35].

**Nucleobase Analogs**

The development of non-natural base pairs for incorporation into DNA and RNA has been the focus of several research groups for a number of years [36]. There are two basic strategies in this area: the first aims to maintain a Watson-Crick-like network of hydrogen bonding between bases, while the second strategy uses hydrophobic groups that mimic the shape and polarity of the natural bases. In the first of these strategies, Benner is developing synthetic biology based on “artificially expanded genetic information systems (AEGIS) [37].” These systems contain six or more nucleotide base pairs, each one still conforming to the basic Watson-Crick geometry but differing in the arrangement of hydrogen-bond donors and acceptors (compare the natural T:A and C:G (Figure 2A) to the synthetic isoC:isoG and Z:P base pairs (Figure 2C)). Benner initially investigated the ability of an isoC:isoG base pair to be incorporated and replicated by DNA polymerases, but the simultaneous use of 2-thiothymine (2-thioT, Figure 2B) was essential to maintain high isoC:isoG selectivity [38]. Recently, Benner and his group have shown that one of these new base-pairs (Z:P) [39] can be readily incorporated by DNA polymerases with 97.5% retention of the non-natural base pair in a PCR amplification [40]. In addition, Benner has shown that an α-thiotriphosphate containing the nucleobase P can be used by polymerases to make a phosphorothioate-containing oligonucleotide [21]. Benner has also recently investigated the ability of pseudothymidine (ψT, Figure 2B) to be incorporated into DNA as a substitute for thymidine [41].
Kool has pioneered the development of expanded DNA bases to create xDNA and yDNA [36,42]. A recent paper by Kool has shown that two nucleotides (xA and xC, Figure 2B) of his extended nucleotide system can be replicated by polymerases in E. Coli to develop a normal message [43]. Similarly, two purines from yDNA (yT and yC) can be incorporated into E. Coli, but fidelity of replication is less efficient [44].

The hydrophobic approach to develop synthetic base pairs aims to mimic the size, shape, and polarity (in some cases) of a natural base pair, but without the hydrogen bonding component. Seminal work by Kool showed that the hydrophobic group difluorotoluene mimics thymine and is efficiently incorporated by polymerases on an oligonucleotide template [45]. Since this discovery, other hydrophobic base pairs have been developed. Hirao and coworkers recently reported a Ds:Px base pair (Figure 2D). This system allows multiple functional groups to be incorporated on the propargyl position of Px. Sequences with this pair can be amplified by 107 fold with 30 cycles of PCR and >99.9% fidelity per cycle [46]. Romesburg has developed his own system of hydrophobic base pairs and has recently shown that both replication and translation of the 5SICS:NaM pair (Figure 2D) proceeds efficiently [47]. Romesburg’s bases also can be present in both the sense and antisense strand of DNA without loss in replication or translation efficiency [48].

Substitutions of the Furanose Ring

One of the most fundamental changes that could be made in a synthetic cell would be to alter the structure of the furanose ring that supports both the nucleotide base and the phosphodiester backbone. Three specific examples highlighted below are promising candidates for advancing into synthetic biology.

Threose Nucleic Acid (TNA)

Among the numerous nucleic acid mimics studied by Eschenmoser, the version called TNA (an abbreviation for α-(L)-threofuranosyl-(3′-2′) nucleic acid, Figure 3B) could have the most immediate application to synthetic biology [49]. TNA differs from DNA in that the phosphates are linked to oxygens on the 3′ and 2′ positions of the furanose and there is no methylene group present between the sugar ring and an oxygen atom on the phosphate [50]. TNA forms hydrogen-bonded duplexes to complementary TNA sequences (called intrasystem cross-pairing) and similarly interacts with complementary sequences of DNA or RNA (intersystem cross-pairing). Structural details on such complexes reveal that a TNA-TNA duplex bears structural resemblance to A-form DNA or RNA with a slightly shorter distance between adjacent phosphorous atoms [51,52]. Perhaps the most intriguing aspect of TNA is the ability with which genetic information transfer can be catalyzed by polymerases. In particular, Szostak has demonstrated that Therminator DNA polymerase is adept at incorporation of TNA triphosphates using a DNA template [53]. Under the optimal conditions, it was estimated that TNA's of about 200 nucleotides could be generated with acceptable fidelity for in vitro selection experiments [54]. Another important facet of this study is the development of a TNA sequencing scheme to transfer the message from TNA back to DNA followed by DNA sequencing. A recent study of phosphonate analogs of TNA also indicate that Therminator polymerase has some potential to incorporate the adenine analogs [27].

Locked Nucleic Acid (LNA)

A nucleic acid analog that has seen an enormous amount of development across numerous fields of nucleic acid-based applications is LNA (an abbreviation for Locked Nucleic Acid, Figure 3C) [55,56]. These molecules were developed by Wengel and first reported about 10 years ago. The LNA unit derives from RNA in that it has a 2′-oxygen, but the special feature
of LNA is the extra methylene group that is covalently linked to C4' and the oxygen of C2' of the furanose ring. This linkage of LNA locks the sugar ring into an N-type or C3'-endo conformation that is favorable for binding natural nucleic acids, RNA in particular. LNAs engage in both intra- and intersystem cross-pairing, and LNA-LNA duplexes are extremely stable [57]. Structural studies have also shown that an LNA-RNA duplex has an A-form double helical structure while an LNA-DNA duplex resembles the structure seen in RNA-DNA duplexes [58,59]. Small numbers of LNA units have been used to replace portions of natural RNA units in aptamers or ribozymes to improve resistance to degradation by nucleases and even improve activity [60-62]. Investigations into the ability of polymerases to use LNA triphosphates began a couple of years ago and so far the results are preliminary but encouraging [63]. To date, incorporation of LNA units on a DNA or RNA template is limited to around 8 consecutive bases using KOD DNA polymerase or T7 RNA polymerase [64]. More work is needed to optimize such conditions, but it should be noted that there are a number of other LNA stereoisomers that show similar nucleic acid binding properties [65-70]. One of these other isomers could be more compatible with polymerases.

### Hexitol Nucleic Acid (HNA)

The work of Herdewijn has extensively contributed to the general field of nucleic acid mimics, especially for medicinal applications and bioengineering [71-75]. Among the numerous nucleic acid analogues Herdewijn has developed, HNA (an abbreviation for hexitol nucleic acid) in which the nucleobase is present in the β position (β-HNA, Figure 3D) is a promising candidate for introduction into synthetic biology applications [76]. HNAs engage in both intra- and intersystem cross-pairing, and bind very tightly to RNA [77]. Similar to LNA, structures of HNA-HNA and HNA-RNA duplexes adopt an A-form helix [78]. There are a few polymerases that demonstrate good activity for incorporating HNA, namely the Vent(exo') DNA polymerase, and the kinetics of incorporation are similar to the natural deoxynucleotide triphosphates [79]. However, the fidelity of HNA incorporation with Vent(exo') can be maintained only if no more than two consecutive HNAs are inserted.

Another study of HNA explored the ability of HIV Reverse Transcriptase (RT) to insert HNA, and a mutant form of HIV-RT (M184V) was adept at this process, specifically for the adenine derivative [80]. Taking this further, Herdewijn also demonstrated that a 33 nucleotide mRNA containing six consecutive HNAs (encompassing the AUG start codon plus the adjacent UUC codon) did not interfere with ribosome binding, tRNA binding, or translation of the message in an in vitro expression system [81]. More recently, a series of HNA:DNA duplexes (in which 1 to 6 consecutive HNAs were present in one strand of the duplex) were incorporated into a functional and selectable *E. coli* gene for thymidylate synthase [82]. Following ligation to a vector and transformation into *E. coli*, functional transformants were obtained, although the yield of such transformants decreased with increasing number of HNA residues (possibly due to the inefficient translation of the synthetic plasmid by bacterial polymerases). More recent results of Herdewijn demonstrate polymerase incorporation of other isomers related to HNA [83].

### Acyclic Analogues

Acyclic nucleic acid analogs are conformationally more flexible compared to most of the other nucleic acid mimics. At the same time, the synthetic routes to make such analogs are typically much simpler compared to the methods to make the cyclic mimics of nucleic acids. Some of these flexible backbones still have the ability to cross-pair with nucleic acids as well as, in some cases, to be utilized by polymerases.
**Flexible Nucleic Acid (FNA)**

One of the earliest proposals for an evolutionary precursor of DNA and RNA is a nucleic acid based on formyl glycerol (FNA, Figure 4A) [84]. This system was initially described by Leslie Orgel (d. 2007); one of the first scientists to propose that an RNA-world preceded the molecular evolution of DNA and proteins [85,86]. While FNA shows no tendency to cross-pair with DNA, a recent investigation by Switzer has demonstrated that either FNA enantiomer (as well as a racemic mixture) can be used by DNA polymerases, with a preference for the $R$ enantiomer [87]. With Terminator DNA polymerase, extensions of up to 7 nucleotides are possible with good fidelity.

**Glycerol Nucleic Acid (GNA)**

Meggers has introduced an alternative acyclic nucleic acid mimic based on commercially available glycidol [88]. The resulting propylene glycol phosphate diester backbone (GNA, Figure 4B) is highly flexible, yet the oligomer with the $S$ enantiomer cross-pairs well with RNA. GNA also cross-pairs with itself [89]. Recently, Szostak investigated the ability of GNA triphosphates to be used by DNA polymerases [90]. Terminator DNA polymerase successfully incorporates 2 GNA triphosphates on a DNA template, but efficiency decreases substantially with 5 GNAs and the pyrimidine GNA triphosphates are weak binders to the polymerase. An interesting solution to improve the activity of pyrimidine GNA triphosphates was to make thymine-substituted versions with a propyne group at the 5 position in an effort to improve base stacking with the terminal nucleotide of the primer. This strategy worked for cytosine but not for thymine. In a separate study by Szostak, polymerases were also able to use GNA as a template to make DNA [91]. Aminopropyl versions of GNA have also been reported recently, but compatibility with polymerases has not yet been investigated [92].

**Peptide Nucleic Acid (PNA)**

One of the most successful strategies to mimic nucleic acids has been to graft nucleobases onto a peptide backbone to make, “Peptide Nucleic Acids,” usually abbreviated PNA. A number of different PNAs have been explored over the past few decades, including aminoethylglycine PNA (aegPNA) [93] and very recently thioester PNA (tPNA) [94]. Ghadiri and Orgel have recently reported a very intriguing thioester PNA (tPNA, Figure 4C) in which the nucleobases self-assemble onto the peptide backbone in the presence of an oligonucleotide template [94]. The tPNA is constructed from a sequence of cysteine residues at every other position along a polypeptide backbone. Residues at the non-cysteine locations can be variable, but are typically amino acids with polar, ionic sidechains. The formation of the tPNA itself is the product of combining the cysteine-bearing polypeptide with individual nucleobases (derivatized with a thioester) and an oligonucleotide template. After incubation for a couple of hours, the tPNA enriched with the sequence of nucleobases complementary to the oligonucleotide template is obtained. This unique self-assembly process could potentially become the basis of a replication mechanism in a synthetic biology construct.

Nielsen and Buchardt developed what has become the most commonly used PNA [93]. The aminoethyl glycine version of PNA (sometimes called aegPNA, Figure 4D) displays both intra- and intersystem cross-pairing with remarkably high thermal stability and sequence specificity [95]. As a result, aegPNA-based applications have been developed in areas of biomedical research [96-105]. The simplicity of the aegPNA backbone has also contributed to proposals that such a molecule could have been a prebiotic precursor to nucleic acids [106,107]. While polymerase-based synthesis of aegPNA has not been reported, a number of template-based strategies have been developed. Nielsen and Orgel, for example, have reported that oligonucleotides can template the formation of aegPNA and vice versa, although chemically activated monomers have to be used [108]. A recent report has
indicated the potential of a dipeptide to promote aegPNA formation, but not in a templated manner [109]. Other recent reports by Liu and coworkers demonstrate that short aegPNA segments can assemble onto a DNA template and then be chemically ligated together [110,111]. This technique was developed for the in vitro evolution of aegPNAs that bear proteinogenic sidechains of amino acids. At the moment, there is no PNA system that would be an immediate choice for applications in synthetic biology since there is no compatibility with existing biological machinery for replication. While the template-based systems represent important steps forward, the fidelity and efficiency of these systems is not quite at the level that would likely be needed in synthetic biology.

Conclusions

Nucleic acid analogs that are useful in synthetic biology are close to becoming widely applicable. Currently, the most promising analogues that could be applied include alterations to the phosphodiester backbone and to the nucleobases, as well as some of the alternatives to the ribofuranosyl sugar (namely TNA, LNA, HNA, and GNA). It is interesting to note that relatively few attempts have been made to simultaneously incorporate different strategies for mimicking nucleic acids into a single monomeric unit. For example, what are the properties of an oligomer made with a Z:P base pair attached to a threofuranosyl sugar where each monomer is linked together by phosphorothioate units? The numerous possible combinations of the existing components of synthetic nucleic acid mimics constitute a promising area for further development.

Acknowledgments

Support from the intramural research program at NIDDK, NIH is gratefully acknowledged. I would also like to sincerely thank Dr. Ethan Englund for critical reading and comments on this manuscript.

References


35. Knudsen SM, Robertson MP, Ellington AD. In vitro selection using modified or unnatural nucleotides. CurrProtoc Nucleic Acid Chem. 2002; Chapter 9(Unit 9 6)


Figure 1.
A. Natural Base Pairs:

\[
\begin{align*}
\text{T} & \quad \text{A} \\
\text{dR} & \quad \text{dR} \\
\text{Me} & \quad \text{dR} \\
\text{N} & \quad \text{N} \\
\end{align*}
\]

B. Single Base Replacements:

\text{Thymine Replacements:}

\[
\begin{align*}
\text{2-thioT} & \\
\text{yT} & \\
\text{yT} & \\
\text{dR} & \\
\text{dR} & \\
\text{dR} & \\
\text{dR} & \\
\end{align*}
\]

\text{Adenine Replacement:}

\[
\begin{align*}
\text{xA} & \\
\text{dR} & \\
\text{dR} & \\
\text{dR} & \\
\text{dR} & \\
\text{dR} & \\
\text{dR} & \\
\end{align*}
\]

\text{Cytosine Replacements:}

\[
\begin{align*}
\text{yC} & \\
\text{xC} & \\
\text{dR} & \\
\text{dR} & \\
\text{dR} & \\
\text{dR} & \\
\text{dR} & \\
\end{align*}
\]

\text{Guanine Replacement:}

\[
\begin{align*}
\text{xG} & \\
\text{dR} & \\
\text{dR} & \\
\text{dR} & \\
\text{dR} & \\
\text{dR} & \\
\text{dR} & \\
\end{align*}
\]

C. Orthogonal Base Pairs: Watson-Crick-like

\[
\begin{align*}
\text{IsoC} & \\
\text{IsoG} & \\
\text{dR} & \\
\text{dR} & \\
\text{dR} & \\
\text{dR} & \\
\text{dR} & \\
\end{align*}
\]

\[
\begin{align*}
\text{Z} & \\
\text{P} & \\
\text{dR} & \\
\text{dR} & \\
\text{dR} & \\
\text{dR} & \\
\end{align*}
\]

D. Orthogonal Base Pairs: Hydrophobic

\[
\begin{align*}
\text{Ds} & \\
\text{Px} & \\
\text{dR} & \\
\text{dR} & \\
\text{dR} & \\
\text{dR} & \\
\end{align*}
\]

\[
\begin{align*}
\text{5SICS} & \\
\text{NaM} & \\
\text{dR} & \\
\text{dR} & \\
\text{dR} & \\
\text{dR} & \\
\text{dR} & \\
\end{align*}
\]

\[ R = \text{aminohexanamide or a fluorescein-acetylated hexanamide sidechain} \]

Figure 2.
Figure 3.
Figure 4.