

# Beyond DNA and RNA: The Expanding Toolbox of Synthetic Genetics

Alexander I. Taylor, Gillian Houlihan, and Philipp Holliger

Medical Research Council Laboratory of Molecular Biology, Cambridge CB2 0QH, United Kingdom

Correspondence: [ph1@mrc-lmb.cam.ac.uk](mailto:ph1@mrc-lmb.cam.ac.uk)

## SUMMARY

The remarkable physicochemical properties of the natural nucleic acids, DNA and RNA, define modern biology at the molecular level and are widely believed to have been central to life's origins. However, their ability to form repositories of information as well as functional structures such as ligands (aptamers) and catalysts (ribozymes/DNAzymes) is not unique. A range of nonnatural alternatives, collectively termed xeno nucleic acids (XNAs), are also capable of supporting genetic information storage and propagation as well as evolution. This gives rise to a new field of "synthetic genetics," which seeks to expand the nucleic acid chemical toolbox for applications in both biotechnology and molecular medicine. In this review, we outline XNA polymerase and reverse transcriptase engineering as a key enabling technology and summarize the application of "synthetic genetics" to the development of aptamers, enzymes, and nanostructures.

## Outline

- |   |  |
|---|--|
| <ul style="list-style-type: none"><li>1 Introduction</li><li>2 Polymerase engineering for synthetic genetics</li><li>3 Evolving function using synthetic genetic polymers</li></ul> | <ul style="list-style-type: none"><li>4 Nanoscale engineering using synthetic genetic polymers</li><li>5 Concluding remarks: Worlds beyond RNA?</li><li>References</li></ul> |
|---|--|

## 1 INTRODUCTION

The capacity of DNA and RNA for accessible high-density information storage and propagation separates nucleic acids from other sequence-defined biopolymers (including proteins and peptides). This provides both the medium and the mechanism for Darwinian evolution and is as yet unmatched by any other polymer or other molecular system. Indeed, DNA can store as much as 200 petabytes of information per gram with high chemical stability. This function is underpinned by a unique chemistry including the polyanionic phosphodiester backbone, which dominates the physicochemical behavior and thereby decouples base sequence (i.e., information content) from molecular properties and Watson–Crick base-pairing (a combination of hydrogen-bonding and stacking interactions), which allows encoding and decoding of information in a redundant fashion.

### 1.1 Natural Chemical Variation

Given the stringency of molecular requirements for function, one might expect the chemical makeup of natural nucleic acids to be completely uniform. However, considerable chemical variation from the cognate DNA and RNA chemistry is found in nature. Such variation is both diverse and widespread, including a range of epigenetic and other markers in both prokaryotic and eukaryotic DNA. Although most occur sporadically, some modifications can completely replace the canonical bases in some bacteriophage genomes (reviewed in Weigle and Raleigh 2016). Arguably, such bacteriophages represent genetics based on alternative, genetic polymers, suggesting that integration of expanded chemistry sets into living systems (xenobiology) may be possible (Herdewijn and Marlière 2009).

In RNA, in addition to a diverse array of modifications found in transfer RNAs (tRNAs) (reviewed in Schimmel 2018), an expanding range of epigenetic messenger RNA (mRNA) markers modulating translation initiation, RNA half-life, and even decoding by the ribosome have been discovered (reviewed in Meyer and Jaffrey 2017).

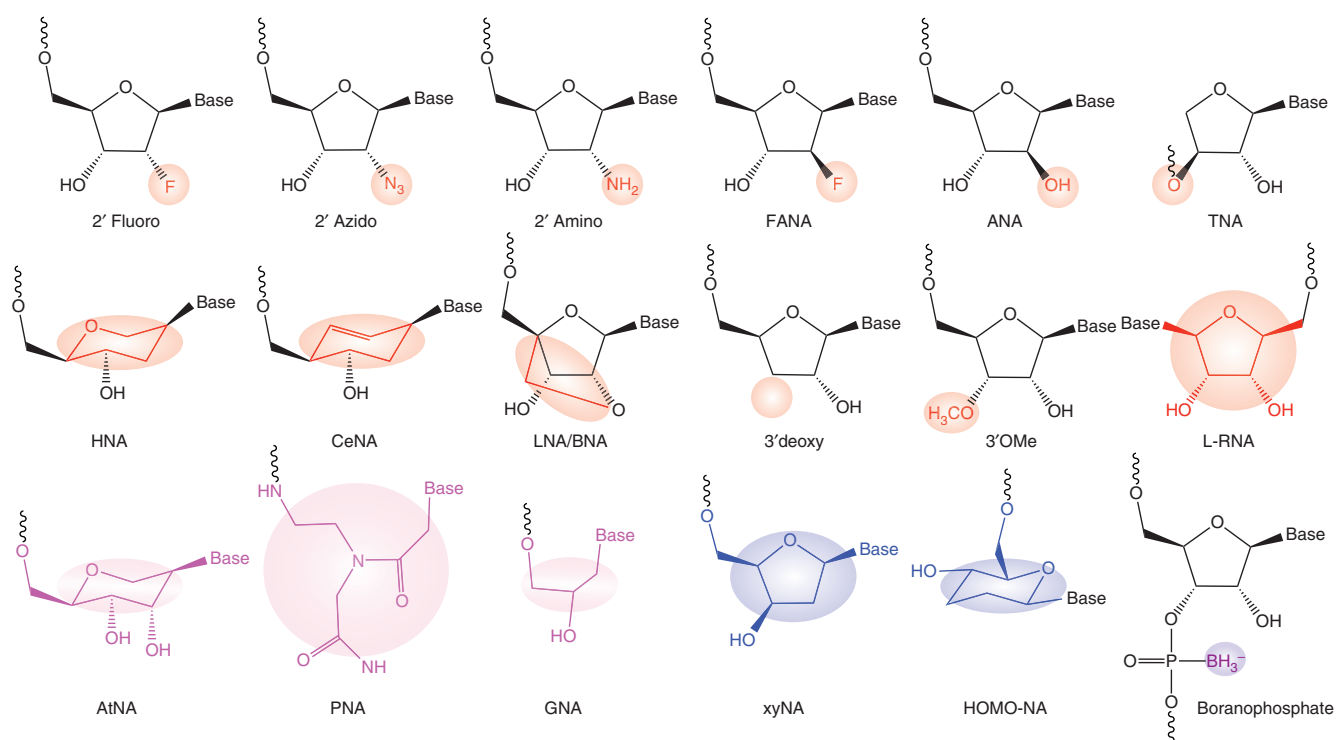
### 1.2 Artificial Chemical Variation and Synthetic Genetics

Beyond the natural variations, organic chemistry has explored a much wider range of alternative backbones, sugar congeners, and base chemistries with the aim to better define the key molecular parameters required for nucleic acid function (Eschenmoser 1999). This has recently been extended to their potential for genetic information storage, propagation, and evolution. Termed “synthetic genetics,”

this approach promises both new insights into the chemical boundary conditions of genetics as well as new tools to study and modify biological processes. A prime example is the modification of nucleobases, utilizing alternative hydrogen bonding patterns, hydrophobic and/or geometric compatibility, or metal ion chelation, toward an expansion of the genetic alphabet *in vitro* and *in vivo*, as well as yielding secondary structure motifs beyond those of DNA and RNA (reviewed in Lee et al. 2018). Furthermore, *in vitro* replication and evolution of nucleic acid variants comprising backbone chemistries not found in nature (here referred to collectively as xeno nucleic acids [XNAs]) yields ligands (XNA aptamers) and enzymes (XNAzymes), as well as simple nanostructures with novel properties including increased biostability (Pinheiro et al. 2012; Yu et al. 2012; Taylor et al. 2015; 2016). Similarly, expansion of DNA molecular diversity (primarily through pyrimidine C5 substituent chemistry) has accelerated aptamer discovery (Gold et al. 2010) and improved nucleic acid catalysts (Zhou et al. 2016; Wang et al. 2018). Here, we provide an overview of the synthetic genetics toolbox, with a special focus on the impact of replacing the canonical ribofuranose sugar of DNA and RNA with synthetic congeners, informing broader chemical possibilities for biotechnology and genetic function (Fig. 1).

## 2 POLYMERASE ENGINEERING FOR SYNTHETIC GENETICS

High-fidelity polymerase enzymes catalyze synthesis and decoding of DNA sequence information with error rates as low as  $1 \times 10^{-7}$  coupled with an exquisite substrate specificity enforced by both minor groove readout and active site geometrical constraints. This provides for both accurate substrate selection and exclusion of damaged and/or non-cognate nucleotides. Indeed, DNA polymerases must stringently select cognate deoxyribonucleotide triphosphates (dNTPs) from a pool that may contain between 10- and 100-fold higher concentration of ribonucleotide triphosphates (NTPs) as incorporation of NTPs into the genome leads to genomic instability. The molecular basis of NTP discrimination lies in bulky amino acid side chains (e.g., tyrosine in the case of the archaeal polB family polymerases like *Thermococcus gorgonarius* [Tgo]: Y409) (Cozens et al. 2012), which excludes nucleotides with 2'-OH substituents from the polymerase active site. This “steric gate” residue has analogs in all DNA polymerase families and is paradigmatic of the mechanisms polymerases use to ensure substrate specificity, understanding of which will be key for progress in synthetic genetics (Fig. 1).



**Figure 1.** Chemical structures of unnatural nucleic acids investigated for synthetic genetics. Sugar modifications incorporated by polymerases are shown in red, whereas nucleic acids that currently cannot be synthesized by polymerases are shown in pink. Unnatural chemistries that do not hybridize to natural nucleic acids are shown in blue, and modifications to the phosphodiester backbone are shown in purple.

## 2.1 Engineering Primer-Dependent RNA Polymerases

The comparable simplicity of such molecular checkpoints has inspired efforts to engineer DNA polymerases into primer-dependent RNA polymerases (Houlihan et al. 2017a). Replacement of the steric gate with less bulky amino acids can improve NTP incorporation kinetics, up to levels comparable to dNTP incorporation kinetics by DNA polymerases, but still precludes RNA synthesis beyond short oligomers. Indeed, a second checkpoint in the thumb domain was recently discovered that enabled primer-dependent synthesis of longer RNAs (Cozens et al. 2012). Primer-dependent RNA synthesis has a number of applications, as it obviates the need to initiate transcription with a 5'-pppG and allows a wide range of chemical groups to be added during solid-phase synthesis of primers (e.g., divergent 5'-cap chemistries and cap-proximal epigenetic modification such as  $N^6,2'$ -O-dimethyladenosine [ $m^6A_m$ ] and 7-methylguanosine [ $m^7G$ ] [reviewed in Meyer and Jaffrey 2017]). Further engineering enabled incorporation of regioisomeric 2'-5' linkages into both DNA and RNA by replacement of purine dNTPs with 3'-deoxy or 3'-O-methyl (3'OMe) analogs (Cozens et al. 2015).

## 2.2 Engineering Modified DNA and RNA Polymerases

To incorporate 2' modified nucleotide substrates, much work has gone into engineering T7 RNA polymerase (T7RP) (e.g., for incorporation of dNTPs as well as 2'-F- and 2'-N<sub>3</sub>-pyrimidine triphosphates [YTPs]). One such T7RP is available commercially as DuraScribe (Lucigen Corp.); it enables incorporation of 2'OMe-, 2'-F-, and 2'-N<sub>3</sub>-YTPs or all four 2'OMe-NTPs provided GTP is also present to enable the initiation step (Burmeister et al. 2005).

DNA polymerases from the polA family have also been extensively engineered for 2'-modified substituents. Phage display was adapted to evolve the Stoffel fragment of *Taq* polymerase for 2'OMe-CTP and -ATP incorporation (Chen et al. 2016a). Limited synthesis was possible with a polymerase (SFM19) with a mutation at residue 615 (E615G), which had previously been identified as a "steric gate" mutation enabling NTP incorporation (Ong et al. 2006). Further evolution resulted in polymerase SFM4-3 (SFM19: V518A, N583S, D655N, E681K, E742Q, M747R), which was able to transcribe fully modified 2'OMe 60-mer templates (Chen et al. 2016a).

The success of this approach suggests that mutations in the finger domain can provide efficiency gains possibly by stabilizing its interaction with the thumb subdomain, thus favoring the catalytically competent closed complex.

### 2.3 Engineering XNA Polymerases and Reverse Transcriptases

Going beyond 2' substitution on the RNA ribofuranose ring, wholesale replacement with alternative ring structures (congeners) has also been explored. Such nucleic acids are not found in nature and have therefore been termed xeno nucleic acids (XNAs). Such modifications can engender profound changes in the nucleic acid helix structure, stability, hybridization kinetics, and biological and chemical behavior (Anosova et al. 2016). Engineering polymerases compatible with these structurally and conformationally divergent chemistries from the canonical nucleic acids required the development of novel polymerase selection and screening methods (reviewed in Houlihan et al. 2017a,b). These enabled the discovery of polymerases for a range of XNAs, including 1,5-anhydrohexitol nucleic acid (HNA), cyclohexene nucleic acid (CeNA), 2'-O,4'-C-methylene- $\beta$ -D-ribonucleic acids (locked or bridged nucleic acid, LNA/BNA), and arabino- (ANA) and 2'-fluoro-arabinonucleic acid (FANA) (Pinheiro et al. 2012). Another XNA chemistry,  $\alpha$ -L-threofuranosyl nucleic acid (TNA), based on a threose scaffold (a possible "simpler" precursor of RNA) has also been explored, with mutants of the archaeal DNA polymerases Tgo and 9°N both able to synthesize TNA polymers (Pinheiro et al. 2012; Larsen et al. 2016).

Genetic information encoded in RNA can be retrieved using reverse transcriptases (RTs) derived from retroviruses. However, natural polymerases and RTs generally cannot reverse transcribe XNA strands into DNA. Notable exceptions include Bst (the large fragment of the *Bacillus stearothermophilus* DNA polymerase), which was recently shown to faithfully reverse transcribe TNA polymers containing TNA-A, -G, and -T and a fluorescent TNA-C nucleotide (Mei et al. 2017), and RTs from avian myeloblastosis virus (AMV) and Moloney murine monkey virus (M-MLV), which are able to reverse transcribe templates containing 2'OMe (Pan et al. 1995; Burmeister et al. 2005).

Although not as developed as polymerase engineering for forward synthesis, XNA-RT activity has also been discovered by screening or in vitro evolution strategies. Statistical coupling analysis (SCA) has been used to map polymerase positions involved in function; screening SCA-informed mutants of the Tgo polymerase led to the

discovery of the critical active site- and C-motif-proximal 521 position. A subtle but crucial mutational change (I521L) enabled reverse transcription of a range of XNAs including HNA, ANA, FANA, and TNA (Pinheiro et al. 2012); introduction of E664K into the Tgo pol (shown to increase primer-template affinity [Cozens et al. 2012]) further extended XNA reverse transcription to LNA and CeNA. Recently, an emulsion-based selection method was adapted to evolve *Thermococcus kodakaraensis* (KOD) DNA polymerase, yielding a mutant (RTX) capable of high-fidelity RNA reverse transcription (Ellefson et al. 2016), with a threefold reduction in error rate compared with the M-MLV RT. RTX (with 3'-5' exonuclease activity inactivated) could also reverse transcribe a fully modified 2'OMe oligomer.

Although the above XNAs hybridize specifically to DNA or RNA (albeit with variable affinity), a number of chemistries have been described that hybridize very weakly or not at all to natural nucleic acids (Anosova et al. 2016). These include glycol nucleic acid (GNA): R-GNA and S-GNA enantiomers pair with themselves, but not each other, and neither is able to cross-pair with DNA. Similarly, other orthogonal self-pairing systems that do not pair with DNA include xylo nucleic acid (xyNA),  $\beta$ -Homo-DNA, enantio-DNA, and the mirror image L-forms (L-DNA, L-RNA) of the natural D-form nucleic acids. Such chemistries cannot currently be replicated via a DNA intermediate and have proved challenging for polymerase engineering. However, S-GNA synthesis by Bst polymerase has been reported, suggesting that the polymerase active site can enforce and stabilize weak primer-template interactions (Tsai et al. 2007).

In summary, polymerase engineering has progressed rapidly, and both the synthesis and reverse transcription of a range of unnatural or modified nucleic acids have become routine. With multiple selection formats available (Houlihan et al. 2017a,b), the prospects for adapting polymerases to ever more diverse chemistries are good. Conversely, medicinal chemistry approaches (adapting nucleotide substrate to the polymerase) have proven remarkably successful, most notably in the discovery of unnatural base pairs (UBPs) (reviewed in Feldman and Romesberg 2017). However, a number of challenges remain. Catalytic efficiency and fidelity of polymerases with unnatural triphosphates is generally lower than with cognate dNTPs or NTPs. A recent crystal structure of the ternary complex of KOD-RI with a DNA primer-template duplex and a TNA triphosphate (tATP) bound revealed severe distortion of the nascent base pair when compared with the binary structure (Chim et al. 2017), suggesting an important role for structural investigations to guide future efforts in synthetic genetics.





### 3 EVOLVING FUNCTION USING SYNTHETIC GENETIC POLYMERS

#### 3.1 Functional Nucleic Acids

Beyond the classical double-helix conformation, nucleic acids (and in particular RNA) are able to form a wide range of three-dimensional folds and structures (including internal hairpin loops, internal bulges, pseudoknots, and G-quartets), underpinned by canonical Watson–Crick as well as noncanonical intramolecular bonds (e.g., Hoogsteen pairs, base triples, GU wobble pairs). Such nucleic acid structures can, in some cases, match the ability of proteins to form sophisticated interaction surfaces (paratopes), comprising shape–charge complementarity, salt bridges, hydrogen bond networks, van der Waals forces, and metal ion binding pockets (Gelinas et al. 2016). The diverse ranges of structures and functions in modern biology have been formulated in the concept of the cell as an “RNA machine.”

#### 3.2 In Vitro Evolution of Functional Nucleic Acids

Nucleic acids are unique in their capacity to encompass both genotype (i.e., sequence information) and phenotype (i.e., positioning functional chemical groups in a defined 3D fold), thus enabling Darwinian evolution at the molecular level, as first shown in vitro by Spiegelman (reviewed in Joyce 2007). In vitro evolution experiments to identify synthetic functional nucleic acids by selection from diverse repertoires are typically termed “systematic evolution of ligands by exponential enrichment” (SELEX) methods (reviewed in Darmostuk et al. 2015). Fundamentally, SELEX relies on the generation of very large ( $>10^{14}$ ) repertoires of single-stranded, random-sequence RNA or DNA repertoires that are panned for binding to a target of choice, allowing functional library members to be enriched from the bulk population through iterative cycles of selection and amplification.

Using this approach, a wide variety of synthetic functional nucleic acids have been developed, including RNA and DNA ligands (aptamers) (Dunn et al. 2017) or catalysts—ribozymes/DNAzymes (Wachowius et al. 2017).

#### 3.3 Aptamers

Nucleic acid aptamers can be selected to bind a wide range of molecular targets including small molecules, proteins, or RNA motifs immobilized on solid phase, or extracellular structures on bacteria, viruses, or eukaryotic cells. Selections for cell-penetrating aptamers (Tawiah et al. 2017) and even selections directly in live animals have also been performed (Mi et al. 2010). In recent years, the field has grown exponentially and is increasingly recognized as a

powerful biotechnology for biosensing, proteomics, diagnostics, imaging, and targeted drug delivery (Zhou and Rossi 2017; Hori et al. 2018).

In comparison with antibodies or other protein-based affinity reagents, aptamers offer a number of advantages, including automation of discovery and solid-phase synthesis. Unlike protein biologics, aptamers tolerate repeated cycles of heating, drying, and freezing without irreversible aggregation and have a very long shelf life without dependence on a temperature-controlled cold chain. Furthermore, aptamers show low immunogenicity (with the exception of CpG DNA motifs) and their small size (typically 5–20 kDa) offers improved tissue penetration compared with whole antibodies.

Aptamers can be incorporated into multimeric complexes (such as multispecific constructs, concatemers, dendrimers, and nanoparticles) in a relatively straightforward manner, allowing directed cross-linking of target molecules, or design of systems for hybridization-mediated signal amplification, as well as integration with nanotechnology designs for assembly of biosensors, enzyme arrays, and nanorobots (Jo and Ban 2016). However, a number of significant challenges to their widespread use in biotechnology have been identified. For one, the functional scope of aptamers is constrained by systemic shortcomings inherent in nucleic acid chemistry, including a comparatively uniform chemical makeup with the steric, electronic, and physicochemical properties of the four canonical bases spanning a rather limited range. This has spurred efforts to expand the functionality of aptamers (and functional nucleic acids in general) by expanding their chemical diversity through the addition of extraneous substituent chemistries.

#### 3.4 Modifying Aptamers by Postselection Medicinal Chemistry

Despite setbacks in the clinical progression of aptamers as therapeutic agents, key hurdles for in vivo applications have begun to be addressed; many of the lessons learned from the antisense/short-interfering RNA (siRNA) community (reviewed in Wittrup and Lieberman 2015) have improved the prospects for aptamers—in particular, including diversification of backbone and substituent chemistry to enhance and fine-tune functionality as well as chemical conjugation to improve organ-specific or intracellular delivery and reduce renal clearance.

One of the most significant roadblocks to in vivo applications for all types of functional nucleic acid remains their poor biostability. Exo- and endonucleases that rapidly degrade naked DNA and RNA polymers are common components of all extracellular as well as intracellular (both endosomal and cytoplasmic) compartments. Typically,



the half-life of single-stranded DNA or RNA aptamers in serum is minutes to hours, although a combination of adding small terminal DNA hairpins and engineering increased G-C content of stem structures can (at least in some cases) offer improved stability. Traditionally, medicinal chemistry cycles have been used to test each nucleotide position for substitution with chemical analogs that enhance biostability (Röthlisberger and Hollenstein 2018). Although such modifications can stabilize secondary structures and increase binding affinity in some contexts, extensive substitution can disrupt aptamer 3D structure or folding and consequently reduce target binding affinity.

### 3.5 Partially Modified DNA and RNA Aptamer Selections (modSELEX)

To expand beyond the limitations of the natural chemistries and bypass costly and time-consuming medicinal chemistry cycles, much effort has been expended on incorporating modified nucleotides into the SELEX process. Selections based on direct partitioning can allow discovery from modified libraries prepared by solid-phase synthesis, and for close steric and electronic analogs of the natural nucleotides, natural polymerases may be sufficient for both synthesis and reverse transcription. However, more divergent modifications, such as 2'-modified sugars or ring congeners have generally required the engineering of polymerases and RTs tolerant to alternative NTP analogs (see Sec. 2).

Although the value of expanded chemistries has been shown in antisense and siRNA therapeutics (improving target binding, increasing resistance to nucleases, and reducing immunogenicity) (Khvorova and Watts 2017), comparatively few of these chemistries have so far been incorporated into successful aptamer selections. Furthermore, the majority of reports to date have involved selections with partially modified libraries (modSELEX), with some DNA or RNA components retained during the selection step. Despite this, modSELEX has proven to be a powerful approach for selecting aptamers with improved properties for in vivo applications (Dellafiore et al. 2016; Lipi et al. 2016); although phosphorothioate ( $\alpha$ S)- and boranophosphate ( $\alpha$ BH<sub>3</sub>)-linked RNA backbones have been explored, modified sugars have become a preferred chemistry, such as 2'-amino- and 2'-fluoro-pyrimidine analogs (2'-F-dYTPs) in particular, because of the ease of chemical synthesis of both triphosphates and oligonucleotides. DNA aptamer selection using the analogous purines (2'-F-dRTPs) has recently been shown (Thirunavukarasu et al. 2017), and polymerases capable of using all four 2'-F analogs (2'-F-dNTPs) have been developed (Cozens et al. 2012), thus aptamer selections at full-2'-F substitution should now be feasible. Polymerase engineering (see Sec. 2) has

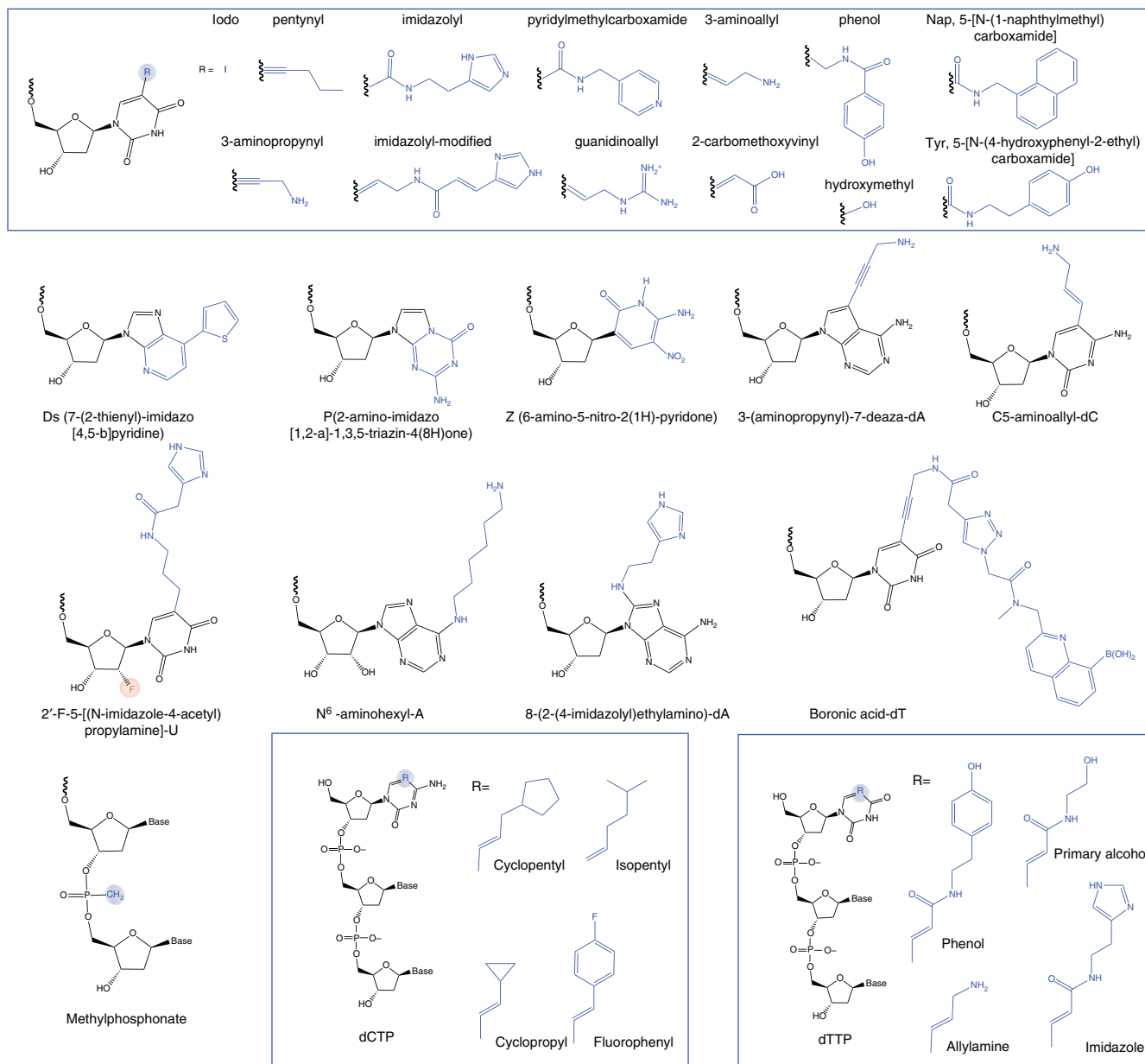
also allowed another standard chemistry, 2'-O-methyl (2'OMe) RNA, to be used in selections. 2'OMe is an attractive substitution as it is biostable as well as nontoxic, occurring naturally as a modification of ribosomal RNA and tRNAs. A significant milestone involved a selection for anti-VEGF aptamers using a T7RP mutant to incorporate a mixture of 2'OMe-A, -C, -G, -U, spiked with GTP to enable polymerase initiation (Burmeister et al. 2005).

Although the first crop of 2'OMe-modified therapeutic aptamers proved insufficiently effective in clinical trials (Zhou and Rossi 2017), recent successes of 2'-modifications in the context of designed antisense reagents has generated renewed interest in the therapeutic potential of nucleic acids. More than 30 antisense therapeutics are in clinical trials, close to approval or about to enter the clinic, most notably those containing 2'-methoxyethyl- (2'MOE) RNA such as Spinraza/Nusinersen and IONIS-HTTRx (Ionis Pharmaceuticals, Inc., in partnership with Biogen/Hoffman-La Roche AG, respectively), as well as uncharged phosphorodiamidate morpholino (PMO) oligonucleotides, such as ExonDys 51/Eteplirsen (Sarepta Therapeutics, Inc.).

Other modifications that have been explored in modSELEX include 4'-thioribose pyrimidines (4'S-YTPs), which conferred a 50-fold increase in resistance to RNase A degradation. Locked or bridged nucleic acids (LNA/2'4'BNA), in which the 2'O and 4'C of the deoxyribose ring are linked by a (oxy)methylene bridge, are known to be even more resistant to nuclease degradation, and confer exceptional duplex stability (Dellafiore et al. 2016; Lipi et al. 2016). So far, only one instance has been reported in which an LNA TTP replaced dTTP in the randomized region of a DNA library for aptamer selection (Kasahara et al. 2013). Engineered polymerases capable of synthesis and reverse transcription of fully substituted LNA polymers have been described (Pinheiro et al. 2012), suggesting that in principle all-LNA aptamer selection should also be possible.

### 3.6 Aptamer Selections Using Modified Bases and Novel Base Pairs

Although modified sugars can improve target binding as well as biostability, diversification of the nucleobase chemistry (Fig. 2) has proven to be a more effective strategy for expansion of the chemical diversity of target binding surfaces (the paratope). Various side chains were explored early on; however, the majority of studies have focused on the attachment of hydrophobic groups to the C5 position of uridine or deoxyuridine. A key advance in this area are slow off-rate modified aptamers (SOMAMers) comprising an amide bond linkage between the pyrimidine C5 position and the substituent side chain, which enhanced aptamer



**Figure 2.** Chemical structures of a range of base-modified nucleic acids used in aptamer and nucleic acid catalyst selections and in nanostructures. Base modifications are indicated in blue, whereas sugar modifications are shown in red.

discovery success rate, especially against previously challenging targets (Gold et al. 2010). Unlike the freely rotating alkyne linkages used previously, the amide is planar, thereby restricting free rotation and limiting entropic costs upon binding to target. SOMAmers look especially promising for diagnostic proteomics, in which the hydrophobic substituents (e.g., benzyl derivatives) provide extensive nonpolar interactions with target molecules, expanding the range of targetable epitopes beyond that of unmodified DNA and RNA (Gelinas et al. 2016). This is an interesting approach

as tyrosine side chains are overrepresented in the complementarity-determining regions (CDRs) of antibodies, and high-affinity antibodies have been isolated from repertoires comprising a simple binary Tyr/Ser binding code (Koide and Sidhu 2009).

Although incorporation of a range of C5-modified pyrimidine triphosphates has been achieved using natural (or minimally engineered) polymerases, high-density modification remains challenging. An alternative approach involves postsynthetic modification, whereby only a single

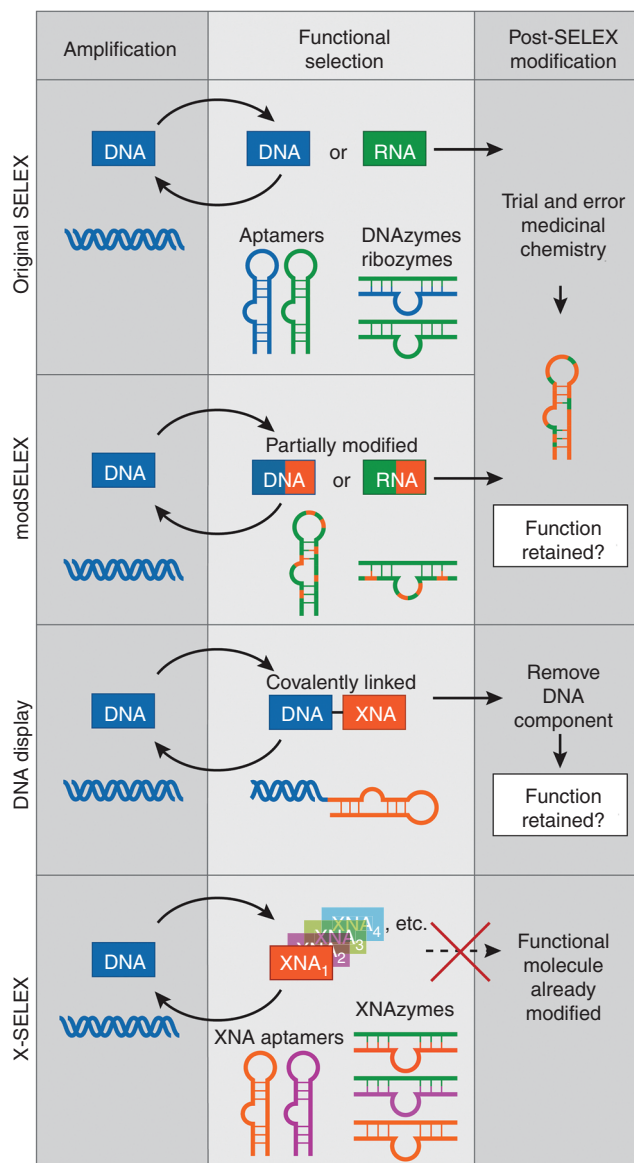
modified nucleotide (for each backbone chemistry) comprising the requisite chemical handle need be synthesized. In the recently developed click-SELEX, 5-ethynyl-substituted dUTP is incorporated and postsynthetically reacted with azido group-containing substituents using click chemistry (Pfeiffer et al. 2018). Again, as described above, this approach may also benefit from the restricted rotation of the resulting 1,2,3-triazole ring.

In a different approach, “synthetic translation” has the potential to offer an even wider range of substituent chemistries: DNA triplets bearing one of eight different side chains can be aligned on DNA templates and ligated (by T4 DNA ligase) to prepare heavily modified DNA, which was recently used to select aptamers against IL-6 and PCSK9 (Chen et al. 2018).

Distinct from the modification of the natural bases, the development of novel orthogonal base-pairing systems can be exploited for aptamer technology. Indeed, an expanded alphabet should allow site-specific incorporation of defined functional groups or an expansion of the side chain repertoire in aptamers (Benner et al. 2016). Five- and six-letter systems (A,T,C,G plus one or two orthogonal base analogs) have been independently developed, termed “genetic alphabet expansion SELEX” (ExSELEX) and “artificially expanded genetic information systems” (AEGIS), respectively. Both systems have been used to select expanded-alphabet DNA aptamers specific for targets that include whole cells and toxins (Lee et al. 2018). The AEGIS pair (Z and P) exploits alternative patterns of hydrogen bond donor and acceptor groups and can be sequenced by specific differential chemical conversion (akin to bisulfite sequencing). Even in the absence of nonnatural functional groups (apart from the nitro group on Z; 6-amino-5-nitro-3(1'-β-D-2'-deoxyribofuranosyl)-2(1H)-pyridone), the six-letter system appears to increase the number of functional aptamers within a repertoire, presumably because of an increase in the diversity of possible aptamer structures, as well as a reduction of alternative folds of each sequence (i.e., limiting the conformational ambiguity). The ExSELEX system uses a different pair (Ds and Px), which achieves orthogonality through steric complementarity, and only one of the letters (Ds) is incorporated into DNA libraries; these positions were initially fixed, but recent aptamer selections have allowed them to vary (Matsunaga et al. 2017).

### 3.7 Fully Modified Xeno Nucleic Acid (XNA) Selections (X-SELEX)

The modSELEX approaches described above still involve DNA or RNA components and thus retain vulnerability to nucleases, but complete replacement of all four natural nucleotides during aptamer selections is possible (Fig. 3).



**Figure 3.** Overview of methods for directed evolution of functional nucleic acids (aptamers and enzymes) composed of modified DNA or RNA, or xeno nucleic acids (XNAs), using systematic evolution of ligands by exponential enrichment (SELEX) strategies.

Polymerase engineering by in vitro evolution and screening (see Sec. 2) have recently enabled DNA-templated synthesis and reverse transcription of a range of nonnatural nucleotide analogs composed of modified or alternative backbones, sugars, or ring congeners not found in nature (xeno nucleic acids [XNAs]). Systems that allow XNA replication (via a DNA intermediate [DNA→XNA→DNA]) enable selections for functional molecules directly from diverse repertoires of fully modified nucleic acid analogs. Transcribing between DNA and XNA in a selection cycle (cross-chemistry SELEX [X-SELEX]) (Pinheiro et al. 2012)



allows selection steps to be performed on XNA libraries, whereas cDNA is used for amplification, cloning, and sequencing as XNAs cannot yet be directly sequenced, although the advent of nanopore sequencing may allow this in the near future.

To date, XNA aptamer selections have been reported using a range of chemistries bearing 2'-modified ribose or alternative ring structures. Before the development of an appropriate RT for TNA, DNA display was used to select antithrombin TNA aptamers from a three-letter TNA alphabet (synthesized from tA-, tC-, and tTTP) (Yu et al. 2012). Using engineered polymerases for synthesis and reverse transcription, HNA aptamers against hen egg lysozyme (HEL) or a truncated version of the TAR RNA motif from HIV were selected directly from diverse repertoires (up to  $10^{14}$  variants) (Pinheiro et al. 2012). The anti-TAR HNA aptamers were shown to be specific for different RNA structural motifs (and not binding TAR RNA by antisense complementarity), whereas the anti-HEL HNA aptamers could stain mouse plasmacytoma cells expressing HEL as a transmembrane protein. Crucially, HNA aptamers could be exposed to acid, base, and nuclease treatment with no measurable effect on activity, showing that such XNAs can yield aptamers with both extreme biostability and chemical inertness.

2'OMe-RNA, although technically not an XNA (as it is found as a posttranslational modification of cellular RNA), has also continued to be explored. Engineering of T7RP allowed selections to be performed on a mixture of all four 2'OMe nucleotides plus GTP (Burmeister et al. 2005). Subsequent efforts have shown that a mixture of 2'OMe-dA, -dC-, -dUTPs, and 2'F-GTP can be used for selection of fully substituted aptamers with extended half-lives in 10% serum (Friedman et al. 2015). Recently, mutants of the Stoffel fragment of *Taq* DNA polymerase were described that could synthesize and reverse transcribe fully 2'OMe-RNA in the context of a selection for binders to human neutrophil elastase (Liu et al. 2017).

An alternative approach, adapted from the mirror-image phage selection strategy originally described by Kim (Schumacher et al. 1996) is the selection of RNA or DNA aptamers against enantiomeric target molecules, followed by chemical synthesis of the nonnatural L-RNA or L-DNA isomers of the aptamers, termed "Spiegelmers." Although Spiegelmers are composed of unmodified DNA or RNA, the L-forms are not recognized by natural nucleases, and so show greatly extended half-lives in serum. Various Spiegelmers have been developed, including (remarkably) L-RNA aptamers able to recognize natural D-RNA motifs such as microRNAs (Sczepanski and Joyce 2015), and various candidate therapeutics are in clinical trials (Vater and Klußmann 2015). One drawback of this approach is that

mirror-image targets must first be prepared (by chemical synthesis of D-(poly)peptides or L-nucleic acid targets), precluding the use of whole cells or large proteins in Spiegelmer selections. However, the preparation of L-DNA polymerases composed of D-amino acids (Wang et al. 2016) suggests selection of Spiegelmers directly from L-form aptamer libraries will be possible.

Another strategy in development seeks to leverage DNA template-directed nonenzymatic assembly of chemical entities to assemble sequence-defined polymers, such as peptide nucleic acid (PNA), an uncharged nonpolar backbone chemistry with very high resistance to both nucleases and proteases. Approaches such as this have the potential to extend in vitro evolution into chemistries increasingly distant from the natural nucleic acids (Lutz et al. 2013).

### 3.8 Modified Ribozymes and DNazymes

As described above, the diverse folded structures of nucleic acids not only engender specific ligand binding but can form catalytic motifs analogous to protein enzymes, which can be evolved de novo from random sequence repertoires. Many such ribozymes and DNA catalysts (DNazymes) have been isolated (Silverman 2016; Wachowius et al. 2017), including ribozymes with RNA polymerase activity. Initially limited to short templates, years of directed evolution and RNA engineering have enhanced RNA polymerase ribozyme activity toward synthesis of ever longer RNAs, although mostly on favorable RNA templates. Nevertheless, the most advanced polymerase ribozymes are now able to synthesize RNA aptamers (Horning and Joyce 2016), synthesize simple ribozymes (Wochner et al. 2011), and even support exponential amplification (ribo-PCR) (Horning and Joyce 2016). Although functional nucleic acids are able to catalyze a wide spectrum of chemical transformations (Silverman 2016), they appear to be most adept at phosphoester chemistry, with the most powerful catalysts being RNA endonucleases and ligase ribozymes, including the class I ligase ribozyme from which the RNA polymerase ribozymes derive (Wachowius et al. 2017).

Although both ribozymes and DNazymes have begun to find applications in synthetic biology as biosensors, molecular computational systems or synthetic regulators of gene expression (e.g., synthetic riboswitches), the majority of research involving synthetic nucleic acid catalysts has taken two avenues: (1) the expansion of chemical functionality to broaden the scope of their enzymatic activity (Silverman 2016); and (2) the improvement of biostability and activity under physiological conditions for potential usage in precision molecular medicine, such as sequence-specific

targeting of viral or disease-associated RNAs (Fokina et al. 2015).

From these perspectives, several groups have explored an expansion of the chemical functionality beyond the standard nucleotides. Typically, as with aptamers, postselection medicinal chemistry has been a major focus. A considerable body of work has also been performed on the classic “10–23” and “8–17” RNA-endonuclease DNAzymes (Santoro and Joyce 1997). A range of variants with modified bases, sugars, and backbones have been prepared by solid-phase synthesis, with a view to augment in vivo stability, substrate binding, or catalytic power.

A fascinating approach has also been to explore simpler evolutionary systems with reduced chemical (and informational) diversity. Ligase ribozymes were selected from genetic alphabets consisting of just three (by deaminating Cs to Us using bisulphite), and subsequently two (replacing A with 2,6-diaminopurine), chemical letters (Reader and Joyce 2002). The observed reductions in catalytic rate suggest that catalytic power (at least for these simple ribozymes) does scale with informational and chemical complexity, boding well for synthetic genetic approaches to expanding the chemical diversity of nucleic acid backbones, base substituent chemistries as well as informational diversity through expansion of the genetic alphabet itself (Benner et al. 2016).

Substitutions with modified nucleotides are necessary for in vivo applications, but generally it has not been possible to replace all RNA or DNA residues in pre-existing ribozymes or DNAzymes (particularly their catalytic cores) without compromising activity. The challenge of replacing all residues with robust analogs while retaining catalytic activity has posed a major obstacle to the development of effective endogenously delivered ribozyme and DNAzyme therapeutics. Although several ribozymes have progressed to clinical trials, such as “angiozyme” (an RNA endonuclease targeting mRNA encoding the vascular endothelial growth factor [VEGF] receptor), and appeared to be well-tolerated in vivo (Morrow et al. 2012), in general efficacy was limited, in part due to reduced biostability resulting from unmodified ribonucleotide positions. To date, no nucleic acid catalyst has achieved clinical approval.

The majority of research to advance DNAzymes for biological applications has centered on the 10–23 RNA endonuclease DNAzyme, due to its modularity and exceptional catalytic efficiency. Variants targeting disease-associated RNAs have been reported to show activity in vivo, and clinical applications for knockdown of pro-inflammatory RNAs in the lung and skin are under investigation (Garn and Renz 2017). However, inactive or photoactivatable versions of 10–23 show similar effects suggesting that knockdown activity may be due to antisense-mediated

mechanisms rather than catalytic activity of the DNAzyme itself (Young et al. 2010).

The use of alternative nucleotide chemistries in ribozyme or DNAzyme selections has the potential to overcome some of these difficulties, offering the prospect of more biostable backbones, formation of more stable secondary structures and improved activities under physiological conditions. Base modifications have also been explored (Fig. 2) including C5-imidazolyl-dU, which yielded a DNAzyme with an impressive catalytic rate in the presence of  $Zn^{2+}$  (Santoro et al. 2000). Extending this approach, RNA endonuclease DNAzymes with up to three nucleotide analogs in which the bases display amino acid-like side chains have been isolated, including the recent report of an RNase A mimic that functions independently of  $Mg^{2+}$  (Wang et al. 2018). A DNAzyme capable of amide bond hydrolysis was recently realized, again using protein-like side chains, a chemical transformation that could not be performed by an unmodified nucleic acid catalyst (Zhou et al. 2016).

As with Spiegelmer aptamers, modification of catalysts with L-RNA or L-DNA offers an alternative route to biostability. However, this requires (for now) selection of D-RNA/D-DNA libraries on mirror-image substrates unless these are achiral. Such cross-chiral catalysis, specifically a cross-chiral RNA polymerase ribozyme, has also been explored to overcome chiral inhibition of prebiotic RNA-catalyzed RNA polymerization (Szczepanski and Joyce 2014).

### 3.9 Catalysts from Synthetic Genetic Polymers (XNAzymes)

The advent of synthetic genetics presents new opportunities to extend nucleic acid catalyst selections to fully modified libraries composed of XNA chemistries (i.e., X-SELEX for catalysts [Taylor and Holliger 2015]). Using broadly similar selection strategies used for ribo- and DNAzyme discovery, XNA catalysts (XNAzymes) were isolated from four different backbone chemistries: ANA, CeNA, FANA, and HNA (Taylor et al. 2015). XNAzymes were identified with RNA endonuclease activity, as well as RNA and XNA ligase activity (in the FANA chemistry). Prediction of XNA secondary structures using standard software are fraught with difficulty as the requisite force fields and structural and conformational parameters are largely unknown (Anosova et al. 2016). However, secondary structures of several of the FANAzymes in complex with their RNA substrates could be determined using a combination of SHAPE (on non-cleavable variants of their RNA substrates in the case of an RNA endonuclease FANAzyme) and mutational profiling (RING-MaP) by reacting FANA (as well as substrate RNA) nucleotides with DMS and reverse transcribing to DNA for deep sequencing using an engineered polymerase

(Taylor et al. 2015). Using this information, catalytic core residues and substrate-binding guide sequences could be unambiguously identified and subsequently reengineered. That catalysts can be discovered by in vitro evolution processes in the sequence and chemical spaces beyond those used in biology has broad implications for our understanding of the boundary conditions necessary for life or its hallmark processes.

## 4 NANOSCALE ENGINEERING USING SYNTHETIC GENETIC POLYMERS

### 4.1 DNA and RNA Nanotechnology

Watson–Crick hybridization rules allow hierarchical, programmable formation of duplexes into various 2D and 3D nanoscale objects and devices. Many studies have shown the potential usage of DNA (and RNA) nanotechnology as, for example, in vivo logic devices, cargo delivery vehicles, or topological guides for synthesis of polymers (Seeman and Sleiman 2017). However, DNA and RNA impose constraints on in vivo applications such as poor biostability and immunogenicity. Although several studies have focused on the impact of alternative design strategies using unmodified DNA and RNA, a range of chemical modification to 2D and (in some cases) 3D designs have also been explored (such as GNA, PNA, methylphosphonate backbones, and alternative bases capable of photo-cross-linking or metal coordination), either by solid-phase synthesis of modified oligonucleotides or by appending functional groups following nanostructure assembly (Gothelf 2017).

One recent advance of note is the application of mirror-image L-DNA. L-DNA versions of a tetrahedral structure were more resistant to nucleases and could be used to deliver a toxic doxorubicin payload to tumor cells in a xenograft mouse model (Kim et al. 2016). In an impressive expansion of nanostructure function using alternative chemistries, hydrophobic moieties and phosphorothioate-linked oligonucleotides (reacted postassembly with ethyliodide) could be used to add nonpolar motifs to DNA barrel-like architectures, allowing them to insert into lipid bilayers, a strategy that has provided a range of engineered DNA nanopores (Howorka 2017).

### 4.2 Incorporating Synthetic Genetic Polymers into DNA Nanotechnology

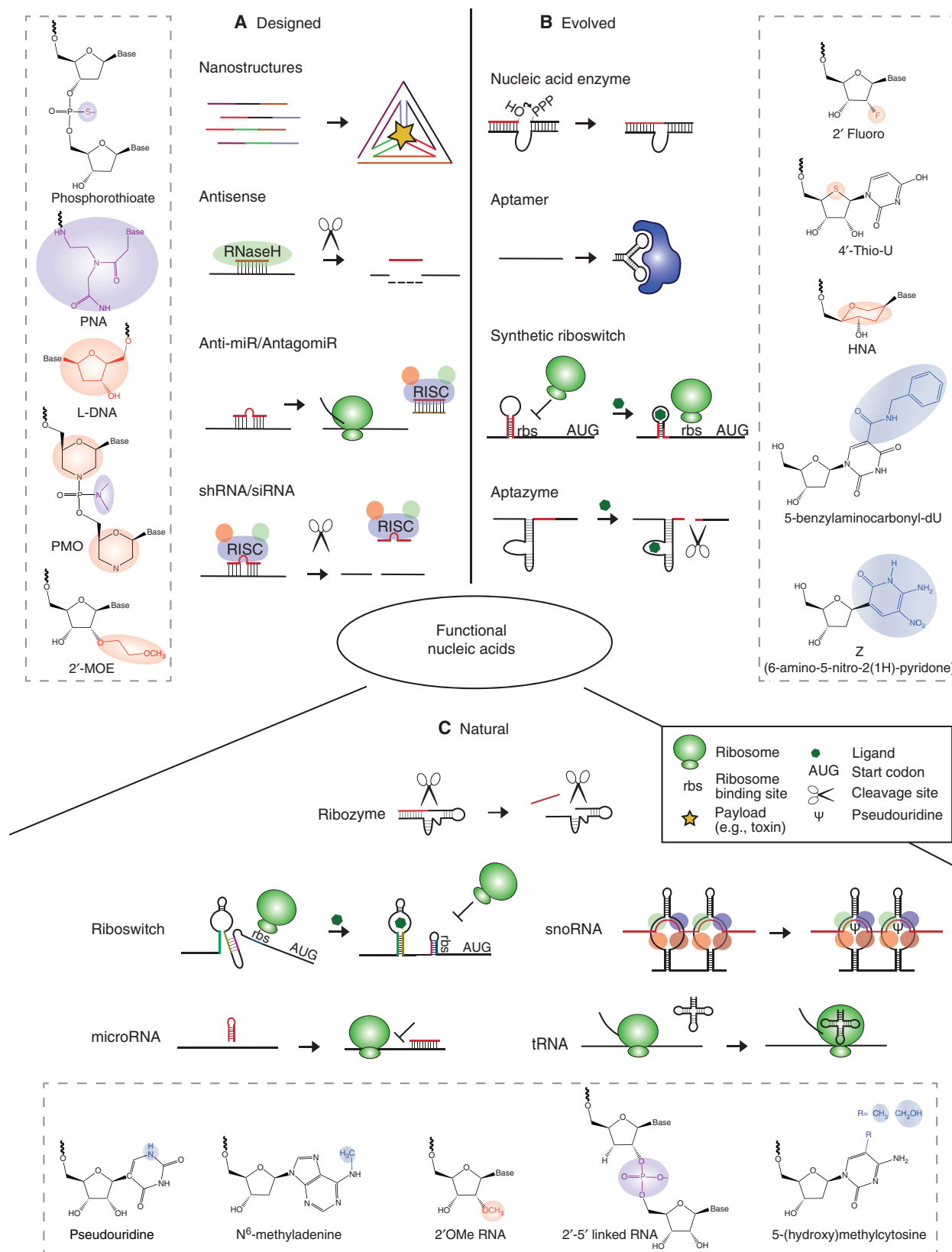
Recently, an alternative approach has sought the wholesale replacement of natural nucleic acids as the building material with XNAs. Encouragingly, partially 2'-F-substituted  $\Phi$ 29 bacteriophage ~80-nt packaging RNA (pRNA) nanomotors were found to retain their ability to assemble into a

functional complex, while acquiring increased resistance to RNase degradation (Liu et al. 2011).

Engineered polymerases (see Sec. 2) can enable access to XNA strands that cannot be obtained by solid-phase synthesis (e.g., a 1.7-kb FANA strand [Taylor et al. 2016]). This has enabled assembly of four-strand tetrahedra composed of 100% 2'-F-DNA, FANA, HNA, and CeNA, as well as hybrid tetrahedral structures comprising DNA/FANA, DNA/2'-F-DNA, and 2'-F-DNA/HNA components. In the case of HNA, tetrahedra were shown to be completely resistant to degradation in serum-containing media at 37°C for more than a week. However, it was notable that tetrahedral designs presenting two-nucleotide single-stranded vertices were more amenable to XNA assembly, presumably allowing a relaxation of conformational strain. In a similar vein, a fully modified 1.7-kb FANA strand, together with five 40-nt FANA staple strands, was shown to assemble into an octahedral design that could be resolved by transmission electron microscopy (TEM) single-particle reconstruction (Taylor et al. 2016). Deviations from the DNA design were apparent, presumably caused by the divergent FANA helical geometries, although the overall topology of the structure was maintained. A deeper understanding of the structural and energetic parameters of the various XNAs may be required to access more complex XNA nanostructure designs.

## 5 CONCLUDING REMARKS: WORLDS BEYOND RNA?

There is considerable merit to the argument that natural nucleic acids appear to be near-optimal solutions for information storing molecules and may hint at general principles likely to be common among functional genetic systems (Benner et al. 2004). However, the ability of a range of unnatural nucleic acid polymers (XNAs), in addition to natural DNA and RNA variants, to form functional molecules (Fig. 4), including aptamers, enzymes, and nanostructures as described, suggests that for nucleic acids (and biopolymers in general) to evolve function in biology, they may fundamentally only require (1) a molecular framework capable of precisely positioning chemically reactive groups, and (2) a link between this phenotypic arrangement and heritable information. Consequently, a whole range of alternative sequence-defined polymers not found in nature might in principle be capable of storing information and forming structures that underpin biological processes. Indeed, as we have discussed, efforts are ongoing to explore the sequence space of ever more divergent sequence defined polymers for the discovery of new materials and drugs. One might therefore ask if “nature’s choice” of RNA (and DNA) is based on clear functional advantages over simple alter-



**Figure 4.** A summary schematic of (A) designed, (B) evolved, and (C) natural functional nucleic acids with expanded chemistry and their respective functions as discussed in this review.



natives or rather on historical contingencies such as the constraints of terran prebiotic chemistry. The question of the degree to which nucleic acids present unique solutions to the problem of genetic information storage has been an ongoing subject of investigation in organic and medicinal chemistry, which has synthesized a wide range of backbone, sugar ring, and base variants. Synthetic genetics is now beginning to probe their functional and genetic potential, and their interaction with the natural genetic system (Chaput et al. 2012; Benner et al. 2016; Chen et al. 2016b; Houlihan et al. 2017a,b).

Expansion of the genetic alphabet probes fundamental aspects of molecular information encoding and the approaches taken and the difficulties encountered are illustrative. Efforts to expand the genetic alphabet by one (or several) novel, orthogonal base pairs were initiated by Benner and colleagues by shuffling of H-bond donor and acceptor pairs on the natural purine and pyrimidine rings to potentially yield four novel pairs (Benner et al. 2016). Alternative approaches have involved exploiting hydrophobic interactions (i.e., base stacking) and shape complementarity. Some of these novel pairs have already allowed the in vitro evolution of higher affinity aptamers to protein or cell-surface targets (see Sec. 3.6). Even more striking, some of the hydrophobic pairs can be both replicated and transcribed in a living organism, enabling new triplet codons for in vivo genetic code expansion (Ledbetter et al. 2018). Although the current generation of artificial base pairs works best when integrated sporadically into DNA (or RNA), future iterations may be able to function at higher multiplicity. However, the manifold challenges that had to be overcome to arrive at functional, orthogonal pairs may hint at possible causes for the “sparse” natural four-letter alphabet. For example, the original base-pair designs with modified H-bond donor–acceptor patterns suffered from unfavorable tautomerization or epimerization equilibria that required further chemical modification (Benner et al. 2016). Likewise, the design of successful hydrophobic pairs lacking H-bonds required systematic screening of a large number of candidate compounds. Thus, a four-letter alphabet may not just reflect constraints of prebiotic chemistry or mutual orthogonality (Szathmáry 2003), but also the rarity of suitable heterocycles that support stable pairing patterns.

Furthermore, stable pairing appears to be an insufficient criterion by itself as some simple heterocycles capable of doing so are not used (or used only sparingly) in biology. For example, 2,6-diaminopurine would appear to have some advantages over adenine, notably providing a more stable pairing with uridine or thymidine, which can aid RNA-catalyzed RNA synthesis by RNA polymerase ribozymes (Attwater et al. 2013) and might have simplified

some genomic transactions by equalizing base-pairing energetics among different base pairs. Likewise, many A- and U-rich anticodons bear chemical modifications to stabilize interactions and enable more efficient readout of the genetic code (Grosjean and Westhof 2016). Nevertheless, 2,6-diaminopurine appears to be only found in the genome of a cyanophage (Weigle and Raleigh 2016). This and other modifications may reflect vestiges of a range of prebiotic base-pairing systems, which did not become stably incorporated into biology.

As discussed above, XNAs bearing a range of variations to the sugar ring chemistry (e.g., 2′F, 2′OMe) including nonglycosidic ring congeners (e.g., HNA, FANA) as well as acyclic analogs (GNA) have been found to be compatible with at least rudimentary genetic function. Broadly, this finding could be interpreted as an indication that life’s genetic system based on phosphodiester and ribofuranose backbones may reflect constraints of prebiotic chemistry rather than stringent functional requirements. However, at least some XNAs, including TNA and ANA, appear compatible with some prebiotic synthesis schemes suggesting that a hypothetical “pre-RNA World” could have had multiple chemical inhabitants from which only the most functional survive in biology. Furthermore, it cannot be excluded that perhaps XNAs could support genetic systems (XNA Worlds) elsewhere in the cosmos.

However, the ribofuranose chemistry of RNA may be special in at least one respect, the vicinal *cis*-diol configuration. This vicinal diol configuration appears to aid import of ribose sugar into protocellular vesicles (Sacerdote and Szostak 2005) and, in the context of the RNA backbone, activates the proximal phosphodiester linkage for transesterification (either cleavage or recombination). This structure underlies the instability of RNA oligomers at high (alkaline) pH and elevated temperatures especially in the presence of bivalent metal ions. Although such hydrolytic instability may be considered undesirable for a genetic material, it accelerates recombination and splicing reactions among RNA strands, and this may have aided early RNA evolution. Thus, RNA’s “Achilles’ heel” that renders it susceptible to hydrolysis might be an important aspect of its functional potential as a genetic polymer.

## REFERENCES

- Anosova I, Kowal EA, Dunn MR, Chaput JC, Van Horn WD, Egli M. 2016. The structural diversity of artificial genetic polymers. *Nucleic Acids Res* 44: 1007–1021.
- Attwater J, Tagami S, Kimoto M, Butler K, Kool ET, Wengel J, Herdewijn P, Hirao I, Holliger P. 2013. Chemical fidelity of an RNA polymerase ribozyme. *Chem Sci* 4: 2804–2814.
- Benner SA, Ricardo A, Carrigan MA. 2004. Is there a common chemical model for life in the universe? *Curr Opin Chem Biol* 8: 672–689.

- Benner SA, Karalkar NB, Hoshika S, Laos R, Shaw RW, Matsuura M, Fajardo D, Moussatche P. 2016. Alternative Watson–Crick synthetic genetic systems. *Cold Spring Harb Perspect Biol* **8**: a023770.
- Burmeister PE, Lewis SD, Silva RF, Preiss JR, Horwitz LR, Pendergrast PS, McCauley TG, Kurz JC, Epstein DM, Wilson C, et al. 2005. Direct in vitro selection of a 2'-O-methyl aptamer to VEGF. *Chem Biol* **12**: 25–33.
- Chaput JC, Yu H, Zhang S. 2012. The emerging world of synthetic genetics. *Chem Biol* **19**: 1360–1371.
- Chen T, Hongdilokkul N, Liu Z, Adhikary R, Tsuen SS, Romesberg FE. 2016a. Evolution of thermophilic DNA polymerases for the recognition and amplification of C2'-modified DNA. *Nat Chem* **8**: 556–562.
- Chen T, Hongdilokkul N, Liu Z, Thirunavukarasu D, Romesberg FE. 2016b. The expanding world of DNA and RNA. *Curr Opin Chem Biol* **34**: 80–87.
- Chen Z, Lichtor PA, Berliner AP, Chen JC, Liu DR. 2018. Evolution of sequence-defined highly functionalized nucleic acid polymers. *Nat Chem* **10**: 420–427.
- Chim N, Shi C, Sau SP, Nikoomanzar A, Chaput JC. 2017. Structural basis for TNA synthesis by an engineered TNA polymerase. *Nat Commun* **8**: 1810.
- Cozens C, Pinheiro VB, Vaisman A, Woodgate R, Holliger P. 2012. A short adaptive path from DNA to RNA polymerases. *Proc Natl Acad Sci* **109**: 8067–8072.
- Cozens C, Mutschler H, Nelson GM, Houlihan G, Taylor AI, Holliger P. 2015. Enzymatic synthesis of nucleic acids with defined regioisomeric 2'-5' linkages. *Angew Chem Int Ed* **54**: 15570–15573.
- Darmostuk M, Rimpelova S, Gbelcova H, Ruml T. 2015. Current approaches in SELEX: An update to aptamer selection technology. *Bio-technol Adv* **33**: 1141–1161.
- Dellafore MAA, Montserrat JM, Iribarren AM. 2016. Modified nucleoside triphosphates for in-vitro selection techniques. *Front Chem* **4**: 18.
- Dunn MR, Jimenez RM, Chaput JC. 2017. Analysis of aptamer discovery and technology. *Nat Rev Chem* **1**: 0076.
- Ellefson JW, Gollihar J, Shroff R, Shivram H, Iyer VR, Ellington AD. 2016. Synthetic evolutionary origin of a proofreading reverse transcriptase. *Science* **352**: 1590–1593.
- Eschenmoser A. 1999. Chemical etiology of nucleic acid structure. *Science* **284**: 2118–2124.
- Feldman AW, Romesberg FE. 2017. Expansion of the genetic alphabet: A chemist's approach to synthetic biology. *Acc Chem Res* **51**: 394–403.
- Fokina AA, Stetsenko DA, François J-C. 2015. DNA enzymes as potential therapeutics: Towards clinical application of 10–23 DNazymes. *Expert Opin Biol Ther* **15**: 689–711.
- Friedman AD, Kim D, Liu R. 2015. Highly stable aptamers selected from a 2'-fully modified fGmH RNA library for targeting biomaterials. *Bio-materials* **36**: 110–123.
- Garn H, Renz H. 2017. GATA-3-specific DNazyme—A novel approach for stratified asthma therapy. *Eur J Immunol* **47**: 22–30.
- Gelinas AD, Davies DR, Janjic N. 2016. Embracing proteins: Structural themes in aptamer–protein complexes. *Curr Opin Struct Biol* **36**: 122–132.
- Gold L, Ayers D, Bertino J, Bock C, Bock A, Brody EN, Carter J, Dalby AB, Eaton BE, Fitzwater T, et al. 2010. Aptamer-based multiplexed proteomic technology for biomarker discovery. *PLoS ONE* **5**: e15004.
- Gothelf KV. 2017. Chemical modifications and reactions in DNA nanostructures. *MRS Bulletin* **42**: 897–903.
- Grosjean H, Westhof E. 2016. An integrated, structure- and energy-based view of the genetic code. *Nucleic Acids Res* **44**: 8020–8040.
- Herdewijn P, Marlière P. 2009. Toward safe genetically modified organisms through the chemical diversification of nucleic acids. *Chem Bio-divers* **6**: 791–808.
- Hori S-I, Herrera A, Rossi JJ, Zhou J. 2018. Current advances in aptamers for cancer diagnosis and therapy. *Cancers* **10**: 9.
- Horning DP, Joyce GF. 2016. Amplification of RNA by an RNA polymerase ribozyme. *Proc Natl Acad Sci* **113**: 9786–9791.
- Houlihan G, Arangundy-Franklin S, Holliger P. 2017a. Engineering and application of polymerases for synthetic genetics. *Curr Opin Biotechnol* **48**: 168–179.
- Houlihan G, Arangundy-Franklin S, Holliger P. 2017b. Exploring the chemistry of genetic information storage and propagation through polymerase engineering. *Acc Chem Res* **50**: 1079–1087.
- Howorka S. 2017. Building membrane nanopores. *Nat Nanotech* **12**: 619–630.
- Jo H, Ban C. 2016. Aptamer–nanoparticle complexes as powerful diagnostic and therapeutic tools. *Exp Mol Med* **48**: e230.
- Joyce GF. 2007. Forty years of in vitro evolution. *Angew Chem Int Ed Engl* **46**: 6420–6436.
- Kasahara Y, Irisawa Y, Fujita H, Yahara A, Ozaki H, Obika S, Kuwahara M. 2013. Capillary electrophoresis-systematic evolution of ligands by exponential enrichment selection of base- and sugar-modified DNA aptamers: Target binding dominated by 2'-O,4'-C-methylene-bridged/locked nucleic acid primer. *Anal Chem* **85**: 4961–4967.
- Khvorova A, Watts JK. 2017. The chemical evolution of oligonucleotide therapies of clinical utility. *Nat Biotechnol* **35**: 238–248.
- Kim K-R, Kim HY, Lee Y-D, Ha JS, Kang JH, Jeong H, Bang D, Ko YT, Kim S, Lee H, et al. 2016. Self-assembled mirror DNA nanostructures for tumor-specific delivery of anticancer drugs. *J Control Release* **243**: 121–131.
- Koide S, Sidhu SS. 2009. The importance of being tyrosine: Lessons in molecular recognition from minimalist synthetic binding proteins. *ACS Chem Biol* **4**: 325–334.
- Larsen AC, Dunn MR, Hatch A, Sau SP, Youngbull C, Chaput JC. 2016. A general strategy for expanding polymerase function by droplet microfluidics. *Nat Commun* **7**: 11235.
- Ledbetter MP, Karadeema RJ, Romesberg FE. 2018. Reprogramming the replisome of a semisynthetic organism for the expansion of the genetic alphabet. *J Am Chem Soc* **140**: 758–765.
- Lee KH, Hamashima K, Kimoto M, Hirao I. 2018. Genetic alphabet expansion biotechnology by creating unnatural base pairs. *Curr Opin Biotechnol* **51**: 8–15.
- Lipi F, Chen S, Chakravarthy M, Rakesh S, Veedu RN. 2016. In vitro evolution of chemically-modified nucleic acid aptamers: Pros and cons, and comprehensive selection strategies. *RNA Biol* **13**: 1232–1245.
- Liu J, Guo S, Cinier M, Shlyakhtenko LS, Shu Y, Chen C, Shen G, Guo P. 2011. Fabrication of stable and RNase-resistant RNA nanoparticles active in gearing the nanomotors for viral DNA packaging. *ACS Nano* **5**: 237–246.
- Liu Z, Chen T, Romesberg FE. 2017. Evolved polymerases facilitate selection of fully 2'-OMe-modified aptamers. *Chem Sci* **8**: 8179–8182.
- Lutz J-F, Ouchi M, Liu DR, Sawamoto M. 2013. Sequence-controlled polymers. *Science* **341**: 1238149.
- Matsunaga K-I, Kimoto M, Hirao I. 2017. High-affinity DNA aptamer generation targeting von Willebrand factor A1-domain by genetic alphabet expansion for systematic evolution of ligands by exponential enrichment using two types of libraries composed of five different bases. *J Am Chem Soc* **139**: 324–334.
- Mei H, Shi C, Jimenez RM, Wang Y, Kardouh M, Chaput JC. 2017. Synthesis and polymerase activity of a fluorescent cytidine TNA triphosphate analogue. *Nucleic Acids Res* **45**: 5629–5638.
- Meyer KD, Jaffrey SR. 2017. Rethinking m<sup>6</sup>A readers, writers, and erasers. *Annu Rev Cell Dev Biol* **33**: 319–342.
- Mi J, Liu Y, Rabbani ZN, Yang Z, Urban JH, Sullenger BA, Clary BM. 2010. In vivo selection of tumor-targeting RNA motifs. *Nat Chem Biol* **6**: 22.
- Morrow PK, Murthy RK, Ensor JD, Gordon GS, Margolin KA, Elias AD, Urba WJ, Weng DE, Rugo HS, Hortobagyi GN. 2012. An open-label, phase 2 trial of RPL4610 (angiozyme) in the treatment of metastatic breast cancer. *Cancer* **118**: 4098–4104.

- Ong JL, Loakes D, Jaroslowski S, Too K, Holliger P. 2006. Directed evolution of DNA polymerase, RNA polymerase and reverse transcriptase activity in a single polypeptide. *J Mol Biol* **361**: 537–550.
- Pan W, Craven RC, Qiu Q, Wilson CB, Wills JW, Golovine S, Wang JF. 1995. Isolation of virus-neutralizing RNAs from a large pool of random sequences. *Proc Natl Acad Sci* **92**: 11509–11513.
- Pfeiffer F, Tolle F, Rosenthal M, Brändle GM, Ewers J, Mayer G. 2018. Identification and characterization of nucleobase-modified aptamers by click-SELEX. *Nat Protoc* **13**: 1153–1180.
- Pinheiro VB, Taylor AI, Cozens C, Abramov M, Renders M, Zhang S, Chaput JC, Wengel J, Peak-Chew S-Y, McLaughlin SH, et al. 2012. Synthetic genetic polymers capable of heredity and evolution. *Science* **336**: 341–344.
- Reader JS, Joyce GF. 2002. A ribozyme composed of only two different nucleotides. *Nature* **420**: 841–844.
- Röthlisberger P, Hollenstein M. 2018. Aptamer chemistry. *Adv Drug Deliv Rev* doi: 10.1016/j.addr.2018.04.007.
- Sacerdote MG, Szostak JW. 2005. Semipermeable lipid bilayers exhibit diastereoselectivity favoring ribose. *Proc Natl Acad Sci* **102**: 6004–6008.
- Santoro SW, Joyce GF. 1997. A general purpose RNA-cleaving DNA enzyme. *Proc Natl Acad Sci* **94**: 4262–4266.
- Santoro SW, Joyce GF, Sakthivel K, Gramatikova S, Barbas CF. 2000. RNA cleavage by a DNA enzyme with extended chemical functionality. *J Am Chem Soc* **122**: 2433–2439.
- Schimmel P. 2018. The emerging complexity of the tRNA world: Mammalian tRNAs beyond protein synthesis. *Nat Rev Mol Cell Biol* **19**: 45–58.
- Schumacher T, Mayr LM, Minor DL. 1996. Identification of D-peptide ligands through mirror-image phage display. *Science* **271**: 1854–1857.
- Szczepanski JT, Joyce GF. 2014. A cross-chiral RNA polymerase ribozyme. *Nature* **515**: 440–442.
- Szczepanski JT, Joyce GF. 2015. Specific inhibition of microRNA processing using l-RNA aptamers. *J Am Chem Soc* **137**: 16032–16037.
- Seeman NC, Sleiman HF. 2017. DNA nanotechnology. *Nat Rev Mater* **3**: 17068.
- Silverman SK. 2016. Catalytic DNA: Scope, applications, and biochemistry of deoxyribozymes. *Trends Biochem Sci* **41**: 595–609.
- Szathmáry E. 2003. Why are there four letters in the genetic alphabet? *Nat Rev Genet* **4**: 995–1001.
- Tawiah KD, Porciani D, Burke DH. 2017. Toward the selection of cell targeting aptamers with extended biological functionalities to facilitate endosomal escape of cargoes. *Biomedicines* **5**: 51.
- Taylor AI, Holliger P. 2015. Directed evolution of artificial enzymes (XNAzymes) from diverse repertoires of synthetic genetic polymers. *Nat Protoc* **10**: 1625–1642.
- Taylor AI, Pinheiro VB, Smola MJ, Morgunov AS, Peak-Chew S, Cozens C, Weeks KM, Herdewijn P, Holliger P. 2015. Catalysts from synthetic genetic polymers. *Nature* **518**: 427–430.
- Taylor AI, Beuron F, Peak-Chew S-Y, Morris EP, Herdewijn P, Holliger P. 2016. Nanostructures from synthetic genetic polymers. *ChemBioChem* **17**: 1107–1110.
- Thirunavukarasu D, Chen T, Liu Z, Hongdilokkul N, Romesberg FE. 2017. Selection of 2'-fluoro-modified aptamers with optimized properties. *J Am Chem Soc* **139**: 2892–2895.
- Tsai C-H, Chen J, Szostak JW. 2007. Enzymatic synthesis of DNA on glycerol nucleic acid templates without stable duplex formation between product and template. *Proc Natl Acad Sci* **104**: 14598–14603.
- Vater A, Klufmann S. 2015. Turning mirror-image oligonucleotides into drugs: The evolution of Spiegelmer® therapeutics. *Drug Discov Today* **20**: 147–155.
- Wachowius F, Attwater J, Holliger P. 2017. Nucleic acids: Function and potential for abiogenesis. *Q Rev Biophys* **50**: e4.
- Wang Z, Xu W, Liu L, Zhu TF. 2016. A synthetic molecular system capable of mirror-image genetic replication and transcription. *Nat Chem* **8**: 698–704.
- Wang Y, Liu E, Lam CH, Perrin DM. 2018. A densely modified M<sup>2+</sup>-independent DNAzyme that cleaves RNA efficiently with multiple catalytic turnover. *Chem Sci* **9**: 1813–1821.
- Weigle P, Raleigh EA. 2016. Biosynthesis and function of modified bases in bacteria and their viruses. *Chem Rev* **116**: 12655–12687.
- Wittrup A, Lieberman J. 2015. Knocking down disease: A progress report on siRNA therapeutics. *Nat Rev Genet* **16**: 543–552.
- Wochner A, Attwater J, Coulson A, Holliger P. 2011. Ribozyme-catalyzed transcription of an active ribozyme. *Science* **332**: 209–212.
- Young DD, Lively MO, Deiters A. 2010. Activation and deactivation of DNAzyme and antisense function with light for the photochemical regulation of gene expression in mammalian cells. *J Am Chem Soc* **132**: 6183–6193.
- Yu H, Zhang S, Chaput JC. 2012. Darwinian evolution of an alternative genetic system provides support for TNA as an RNA progenitor. *Nat Chem* **4**: 183–187.
- Zhou J, Rossi J. 2017. Aptamers as targeted therapeutics: Current potential and challenges. *Nat Rev Drug Discov* **16**: 181–202.
- Zhou C, Avins JL, Klauser PC, Brandsen BM, Lee Y, Silverman SK. 2016. DNA-Catalyzed amide hydrolysis. *J Am Chem Soc* **138**: 2106–2109.



# Cold Spring Harbor Perspectives in Biology

## Beyond DNA and RNA: The Expanding Toolbox of Synthetic Genetics

Alexander I. Taylor, Gillian Houlihan and Philipp Holliger

*Cold Spring Harb Perspect Biol* 2019; doi: 10.1101/cshperspect.a032490

Subject Collection [RNA Worlds](#)

### Beyond DNA and RNA: The Expanding Toolbox of Synthetic Genetics

*Alexander I. Taylor, Gillian Houlihan and Philipp Holliger*

### Structural Basis of Nuclear pre-mRNA Splicing: Lessons from Yeast

*Clemens Plaschka, Andrew J. Newman and Kiyoshi Nagai*

### Coming Together: RNAs and Proteins Assemble under the Single-Molecule Fluorescence Microscope

*Ameya P. Jaliha, Paul E. Lund and Nils G. Walter*

### Mapping the dsRNA World

*Daniel P. Reich and Brenda L. Bass*

### Structural Insights into Nuclear pre-mRNA Splicing in Higher Eukaryotes

*Berthold Kastner, Cindy L. Will, Holger Stark, et al.*

### U1 snRNP Telescripting: Suppression of Premature Transcription Termination in Introns as a New Layer of Gene Regulation

*Christopher C. Venters, Jung-Min Oh, Chao Di, et al.*

### Molecular Mechanisms of pre-mRNA Splicing through Structural Biology of the Spliceosome

*Chuangye Yan, Ruixue Wan and Yigong Shi*

### Imaging the Life and Death of mRNAs in Single Cells

*Jeffrey A. Chao and Timothée Lionnet*

### Discovering and Mapping the Modified Nucleotides That Comprise the Epitranscriptome of mRNA

*Bastian Linder and Samie R. Jaffrey*

### Illuminating Genomic Dark Matter with RNA Imaging

*Arjun Raj and John L. Rinn*

### Group II Intron RNPs and Reverse Transcriptases: From Retroelements to Research Tools

*Marlene Belfort and Alan M. Lambowitz*

### The Origin of Biological Homochirality

*Donna G. Blackmond*

### Principles of Stress Granules Revealed by Imaging Approaches

*Briana Van Treeck and Roy Parker*

### RNA Switches for Synthetic Biology

*Calvin M. Schmidt and Christina D. Smolke*

### The RNA Base-Pairing Problem and Base-Pairing Solutions

*Zhipeng Lu and Howard Y. Chang*

### The Story of RNA Folding, as Told in Epochs

*Daniel Herschlag, Steve Bonilla and Namita Bisaria*

For additional articles in this collection, see <http://cshperspectives.cshlp.org/cgi/collection/>



**All Modifications and  
Oligo Types Synthesized**

Long Oligos • Fluorescent • Chimeric • DNA • RNA • Antisense

*Oligo Modifications?*

Your wish is our command.





For additional articles in this collection, see <http://cshperspectives.cshlp.org/cgi/collection/>



**All Modifications and  
Oligo Types Synthesized**

Long Oligos • Fluorescent • Chimeric • DNA • RNA • Antisense

*Oligo Modifications?*  
Your wish is our command.

