The Emerging World of Synthetic Genetics

John C. Chaput,1,* Hanyang Yu,1 and Su Zhang1
1Center for Evolutionary Medicine and Informatics in the Biodesign Institute, and Department of Chemistry and Biochemistry, Arizona State University, Tempe, AZ 85287-5301, USA
*Correspondence: john.chaput@asu.edu
http://dx.doi.org/10.1016/j.chembiol.2012.10.011

For over 20 years, laboratories around the world have been applying the principles of Darwinian evolution to isolate DNA and RNA molecules with specific ligand-binding or catalytic activities. This area of synthetic biology, commonly referred to as in vitro genetics, is made possible by the availability of natural polymerases that can replicate genetic information in the laboratory. Moving beyond natural nucleic acids requires organic chemistry to synthesize unnatural analogues and polymerase engineering to create enzymes that recognize artificial substrates. Progress in both of these areas has led to the emerging field of synthetic genetics, which explores the structural and functional properties of synthetic genetic polymers by in vitro evolution. This review examines recent advances in the Darwinian evolution of artificial genetic polymers and their potential downstream applications in exobiology, molecular medicine, and synthetic biology.

Merging Chemistry and Biology

In a recent article marking the 40th anniversary of the first in vitro evolution experiment, Gerald Joyce compared Soll Spiegelman’s directed evolution of bacteriophage Qβ RNA to Friedrich Wöhler’s synthesis of urea (Joyce, 2007). Joyce noted that, just as Wöhler’s synthesis of urea broke the false distinction between biological chemistry and organic synthesis (Wöhler, 1828), Spiegelman’s experiments on Qβ RNA broke a similar false distinction between Darwinian evolution as a biological process and Darwinian evolution as a chemical process (Mills et al., 1967). This timely comparison reminds us of the seminal role that both individuals played in the history of science and how their pioneering work gave rise to the prodigious fields of organic chemistry and in vitro genetics (Appella, 2010; Joyce, 2012).

Now, after many decades, these two disciplines are uniting to form a new field of science called synthetic genetics that aims to explore the structural and functional properties of synthetic genetic polymers. Researchers working in this area are developing in vitro selection systems that allow unnatural genetic polymers to evolve in response to imposed selection constraints (Pinheiro et al., 2012; Yu et al., 2012). Previously, the only genetic polymers capable of undergoing Darwinian evolution were DNA and RNA, because these were the only molecules with polymerases available that could replicate, transcribe, and reverse-transcribe genetic information in the laboratory (Joyce, 2004; Wilson and Szostak, 1999). This paradigm is now changing as a new generation of polymerases becomes available that can copy sequence information back and forth between DNA and XNA. Here, we use the term XNA, sometimes referred to as xenonucleic acids, to describe a general class of nucleic acid molecules in which the natural ribose and 2′-deoxygen ribose sugars found in RNA and DNA have been replaced by an unnatural moiety (X). While X usually refers to a genetic polymer with an alternative sugar, nonsugar moieties are also possible, albeit less common.

Herewjin and Marlière (2009) introduced the term XNA in a theoretical paper on the development of genetically altered organisms that use alternative nucleic acid polymers with backbones that do not interfere with normal DNA and RNA biosynthesis. In this review article, we adhere to the traditional view of XNA as synthetic genetic polymers with unnatural nucleic acid backbones. We do, however, recognize that the last 2 decades have witnessed tremendous growth in all areas of nucleic acid chemistry. To date, more than 100 different chemical modifications have been made to DNA and RNA, and together, these modifications have improved our understanding of nucleic acid structure and function. Notable accomplishments include the following: (1) expansion of the genetic alphabet beyond the four natural bases of A, C, T, and G (Henry and Romesberg, 2003; Kool, 2002b; Piccirilli et al., 1990); (2) formation of Watson-Crick base pairs in the absence of complementary hydrogen bond donor and acceptor groups (Krueger and Kool, 2007; Lavergne et al., 2012; Mitsui et al., 2003; Moran et al., 1997); (3) formation of size and strand expanded DNA helices (Chaput and Switzer, 1999; Kang et al., 2012; Liu et al., 2003); (4) replication of modified bases using the polymerase chain reaction (PCR) (Malyshev et al., 2009; Yang et al., 2010); (5) development of in vitro selection strategies that support modified bases (Hollenstein et al., 2009a, 2009b; Vaish et al., 2000; Vaught et al., 2010); and (6) replication and translation of unnatural codons in living bacteria cells (Delaney et al., 2009; Krueger et al., 2011). These and many other accomplishments have profoundly impacted our knowledge of the chemical and physical properties of DNA and RNA and led to the development of new sensors, diagnostics, and therapeutic agents for biotechnology and molecular medicine. Since a thorough discussion of nucleic acid chemistry is beyond the scope of this article, we direct readers interested in learning more about modified nucleic acids and their recognition by DNA and RNA polymerases to several excellent reviews on this topic (Appella, 2009; Brudno and Liu, 2009; Kool, 2002a; Zhang et al., 2010).

In this article, we examine recent advances in the in vitro evolution of synthetic genetic polymers with novel functional properties. These developments represent promising new strategies for creating nuclelease-resistant molecules that can be made to function in a variety of natural and unnatural environments. We begin with a review of molecular evolution and the...
transition from natural genetic systems composed of DNA and RNA to synthetic genetic polymers with unnatural nucleic acid backbones. Next, we discuss examples where in vitro evolution has been used to isolate XNA molecules with discrete ligand-binding properties. This discussion covers the complete spectrum of XNA polymers from early studies on molecules that closely resemble natural genetic systems to more diverse backbone structures that have begun to explore new regions of chemical space far beyond the local structural neighborhood of DNA and RNA. Finally, we conclude with a discussion of synthetic genetics and the potential impact that this new field will have on the future of exobiology, molecular medicine, and synthetic biology. It is our hope that this article will stimulate others to think about new ways to explore the functional properties of synthetic genetic polymers and push the field of synthetic genetics into mainstream molecular biology.

In Vitro Evolution

Darwinian evolution describes how a population changes overtime to optimize the fitness of individual members for a particular environment (Darwin, 1859). Evolution is driven by natural selection, which allows certain inherited traits to become more or less common as a result of differential survival. Individuals who adapt to a particular environment pass their genes on to future generations, while members who are unable to do so go extinct. Although evolution and natural selection are most often discussed in terms of living organisms, these concepts can also be applied to individual molecules (Joyce, 1994; Szostak, 1992). In this regard, genetic polymers are ideally suited for this purpose, because they can fold into shapes with defined functional properties (phenotypes) and their sequences (genotypes) can be replicated in vitro to produce progeny molecules. The ability to amplify individual molecules with desired phenotypes and optimize their functions by directed evolution distinguishes genetic polymers, like DNA and RNA, from all other organic molecules, which are incapable of replication because they lack a genotype-phenotype connection.

To apply the principles of Darwinian evolution to nucleic acids, a large population of nonidentical sequences is created and this pool is assayed en masse for individual molecules that can fold into shapes that can bind to a particular target or catalyze a specific chemical reaction (Ellington and Szostak, 1990; Robertson and Joyce, 1990; Tuerk and Gold, 1990). Molecules with desired phenotypes are separated from the nonfunctional pool, and their genotypes are amplified to generate a new population of progeny molecules that has become enriched in a particular trait. Amplification is usually done at the DNA level using PCR (Saiki et al., 1985, 1988), but this process can also be performed at the RNA level by isothermal RNA amplification (Guatelli et al., 1990). Random mutations that occur during the amplification process are critical to the selection outcome, because these genetic changes introduce new diversity into the sequence population. In general, mutations made to highly conserved positions tend to have a negative impact on biopolymer function, while mutations made to permissive sites either accumulate as a series of neutral mutations or improve biopolymer function by correcting suboptimal contacts in the tertiary structure or by forming new interactions to their substrate or target ligand (Garothers et al., 2004, 2006). Of course, the real power of in vitro evolution comes through iterative cycles of selection and amplification, which makes it possible to search vast regions of “sequence space” for individual molecules that arose from one of 10^{15} different sequences present in the starting population. This level of combinatorial power far surpasses anything that can be accomplished using even the most sophisticated high throughput screening methods, which is one reason why in vitro evolution has emerged as a powerful tool in the biotechnology arsenal.

While in vitro evolution has been used for more than 20 years to isolate functional DNA and RNA molecules from large pools of random sequences (Ellington and Szostak, 1990; Robertson and Joyce, 1990; Tuerk and Gold, 1990), applying the principles of Darwinian evolution to artificial genetic polymers remains a challenging problem. Two major obstacles exist: (1) how to obtain unnatural nucleic acid substrates that are not commercially available; and (2) how to identify polymerases that can convert genetic information back and forth between DNA and XNA. The first problem can be overcome with organic chemistry; however, this requires expertise in chemical synthesis or access to collaborators with knowledge of organic chemistry. Indeed, the chemical synthesis of modified nucleic acid triphosphates remains a difficult and labor-intensive process. Nucleoside triphosphates are not trivial to synthesize or purify, and many analogs require 10 or more synthetic steps. While these challenges are not insurmountable, they do represent a significant bottleneck in the discovery process and one that could slow the pace at which new genetic polymers are examined. Presumably, this problem will become less severe over time as the practical applications of XNA technology motivate chemical and biotechnology companies to construct these substrates as commercial products. Until this happens, the focus will remain on chemistry and the use of chemical synthesis to generate new types of artificial genetic systems with a diverse repertoire of chemical functionality.

The second problem is potentially more serious and involves identifying polymerases that can either copy libraries back and forth between DNA and XNA or copy XNA directly into XNA without using DNA as an intermediate. Discovering enzymes that are suitable for a given XNA system typically involves screening commercial enzymes, but going forward polymerase engineering likely will be more important. Unfortunately, natural polymerases like T7 RNA polymerase and Taq DNA polymerase rarely accept modified substrates with unnatural backbones. Even nucleic acid substrates bearing subtle changes to the ribose or deoxyribose sugar can inhibit these enzymes. This stringent level of substrate specificity makes it difficult to identify polymerases that will recognize modified substrates, either as nucleoside triphosphates or as templates. Fortunately, many variants of wild-type polymerases have been developed for DNA sequencing and other biotechnology applications, and these enzymes are often more permissive of unnatural substrates. Determining which enzymes are best suited for a given XNA polymer requires screening commercial enzymes in a primer-extension assay where each polymerase is challenged to extend a DNA primer annealed to a DNA template with XNA. Since XNA tends to disrupt critical contacts in the enzyme-substrate complex, most polymerases terminate primer extension after one or two nucleotide incorporations. However,
some polymerases will synthesize longer XNA products and these enzymes then become candidates for further optimization. Optimization can proceed by identifying conditions that allow the polymerase to function with enhanced efficiency or by engineering the enzyme itself for improved activity. Both of these approaches have been used successfully to identify polymerases that will transcribe and reverse-transcribe XNA polymers with high efficiency and fidelity (Betz et al., 2012; Loakes et al., 2009; Pinheiro et al., 2012; Yu et al., 2012). Certain engineered polymerases are even able to copy limited stretches of XNA into XNA, suggesting that further polymerase engineering could lead to DNA independent pathways for XNA evolution.

Darwinian Evolution of Artificial Genetic Systems

In vitro selection has allowed for the isolation of nucleic acid sequences that can bind to a wide range of targets from small molecules to whole cells. In common nucleic acid parlance, these molecules are referred to as aptamers, which are single-stranded sequences identified by in vitro evolution that mimic antibodies by folding into shapes that can bind to a desired target with high affinity and high specificity (Ellington and Szostak, 1990). Aptamers represent a valuable class of affinity reagents as ligand binding can alter or inhibit the biological properties of their target proteins (Famulok et al., 2007; Keefe et al., 2010; Mayer, 2009; Nimjee et al., 2005). Unfortunately, genetic polymers composed of natural DNA or RNA are poor candidates for diagnostic and therapeutic applications as these molecules are readily degraded by nucleases—enzymes present in biological samples that cleave phosphodiester bonds. Introducing chemical modifications into the backbone structure that stabilize the molecule against nuclease degradation can enhance the potential utility of aptamers as diagnostic and therapeutic agents. In particular, substitution of the 2‘ hydroxyl position of ribonucleotides with amino (NH₂), fluoro (F), and methoxy (OCH₃) groups confers resistance to nucleases that utilize the 2‘ hydroxyl group for cleavage of the phosphodiester bond. Consequently, the first in vitro selection experiments focused on subtle chemical modifications that were tolerated by natural polymerases and rendered the aptamer more resistant to nuclease degradation (Keefe and Cload, 2008). As increasing numbers of mutant polymerases became available, the field gradually shifted toward nucleic acid polymers with more diverse chemical structures (Figure 1). In the following sections, we divide our examples into two sections: XNA polymers with backbone structures related to RNA and XNA polymers with novel backbone structures. This distinction is designed to highlight the progress of different artificial genetic systems and their contribution to the field of synthetic genetics.

XNA Polymers Related to RNA

In 1994, Jayasena and coworkers reported the first example where Darwinian evolution was applied to pools of modified genetic polymers (Lin et al., 1994). Using recombinant T7 RNA polymerase, a library of RNA molecules bearing 2‘-amino-modified pyrimidines in place of the natural 2‘ OH group and all natural purine nucleotides was constructed by in vitro transcription (Figure 2). Cell-free transcription gave the modified RNA with reduced yield (~50%) relative to natural RNA. This library was used to isolate modified aptamers that bound to human neutrophil elastase (HNE) by performing iterative rounds of in vitro selection and amplification against HNE molecules immobilized on nitrocellulose filters. HNE is implicated in a number of respiratory diseases, including pulmonary emphysema, chronic bronchitis, and cystic fibrosis, and as such represents an interesting therapeutic target. After 15 rounds of in vitro selection and amplification, amino-modified aptamers were identified that bound to HNE with low nanomolar affinity and showed only weak affinity to porcine pancreatic elastase—a closely related protease. Comparative binding assays performed using DNA and RNA molecules with the same nucleotide sequence showed that the aptamer required the amino modification to achieve high affinity and high specificity binding. The authors discovered that the amino-modified aptamer has a half-life of ~20 hr in human serum, which is considerably longer than the serum half-life of RNA (seconds to minutes, depending on sequence and structure). While the success of the HNE aptamer provided a clear demonstration that modified RNA aptamers could be created by direct selection methods (as opposed to postselection modification), the amino modification was quickly discarded due to problems associated with the solid-phase synthesis of amino-modified RNA and the unpredictable nature of the 2‘ amino group (pKa 6.2) at physiological pH.

Recognizing the limitations of 2‘-amino-modified RNA, other functional groups were explored as substrates for T7 RNA polymerase. Of the possible substitutions, 2‘ fluoro RNA emerged as a popular RNA analog (Figure 2). Relative to the 2‘ amino group, 2‘ fluoro leads to increased coupling efficiency during solid-phase synthesis and avoids the need for an additional protection step when constructed as a phosphoramidite monomer. In addition, model studies on sequence-defined synthetic oligonucleotides indicate that the 2‘ fluoro modification enhances the thermal stability of a duplex (Lesnik et al., 1993), while the 2‘ amino modification destabilizes the helical structure (Aurup et al., 1994). The best early example of a 2‘ fluoro aptamer is Macugen (pegaptanib sodium), the first therapeutic aptamer clinically approved by the Food and Drug Administration. Macugen, which is used to treat age-related macular degeneration,
functions as a drug by inhibiting the binding of vascular endothelial growth factor (VEGF)-165 to its target receptor. In clinical trials, 80% of the patients treated with this aptamer show stable or improved vision 3 months after treatment (Eyetech Study Group, 2003). Macugen was isolated from a random sequence RNA library bearing 2′-fluoro modified pyrimidines and natural purine nucleotides (2′ OH) (Ruckman et al., 1998). After the selection, the 28-nucleotide RNA aptamer was further modified by replacing the 2′ OH group at most of the purine positions with 2′-methoxy groups (OCH₃) and appending polyethylene glycol moieties onto the structure (Ruckman et al., 1998).

The simplicity of the 2′ fluoro modification coupled with its recognition by T7 RNA polymerase have helped make 2′ fluoro RNA a widely used derivative for in vitro selection. More recent examples of 2′-fluoro substituted RNA aptamers include work by the Rossi and Sullenger labs. In an exciting new development, Rossi and coworkers evolved 2′ fluoro aptamers that bind to HIV glycoprotein 120 (gp120) with nanomolar affinity (Zhou et al., 2009). In this study, a commercial T7 RNA polymerase mutant called DuraScribe T7 RNA polymerase was used due to its improved ability to incorporate 2′-F CTP and 2′-F UTP substrates as well as ATP and GTP. Remarkably, the aptamer is internalized into HIV-infected cells expressing the HIV envelope protein, suggesting that this aptamer could be used as a drug delivery system. To test this possibility, the anti-gp120 aptamer was conjugated with anti-HIV siRNA to form a chimeric aptamer-siRNA molecule that targets HIV-infected cells and leads to inhibition of HIV replication and infectivity. In another exciting development (Mi et al., 2010), the Sullenger lab designed an in vivo selection method to target tumor-specific moieties. In this system, an in vivo selection was applied to an animal model of intrahepatic colorectal cancer metastases by injecting a nuclease-resistant library of 2′ fluoro pyrimidine-modified RNA into mice bearing previously implanted hepatic tumors. Following an incubation period, the livers were isolated and RNA molecules from the library were recovered and amplified by RT-PCR. After multiple cycles of selection and amplification, an RNA motif emerged that bound to p68 RNA helicase with high affinity. This aptamer colocalizes with p68 and inhibits p68 induced ATPase activity in tumor cells.

Another promising 2′ RNA substitution often used for the production of therapeutic aptamers is 2′-methoxy RNA (OCH₃). The 2′ methoxy substitution (Figure 2) is a naturally occurring RNA analog found in ribosomal RNA that arises through the posttranscriptional machinery of a modified T7 RNA polymerase bearing the mutations K378R, Y639F, and H784A, Burmeister et al., reported the first example of a fully 2′-O-methyl-substituted aptamer isolated by in vitro selection (Burmeister et al., 2005). This unnatural nucleic acid aptamer binds to VEGF₁₆₅ with low nanomolar affinity and shows superior stability in blood with a clearance half-life of 23 hr in mice models. The same authors used a similar strategy to select 2′-O-methyl substituted pyrimidine aptamers that bind to human thrombin and human interleukin (IL)-23 (Burmeister et al., 2006). These aptamers also showed high nuclease resistance in model assays.

**Modified Phosphodiester Linkages**

Phosphorothioate (PS) oligonucleotides differ from natural oligonucleotides (DNA and RNA) by substitution of a single nonbridging oxygen atom in the phosphodiester linkage for a sulfur atom (Figure 2). The markedly lowered electronegativity and greater polarizability of the sulfur atom provides a useful probe for mechanistic studies involving reactions at the phosphodiester linkage (Li et al., 2011). This substitution also renders PS oligonucleotides more sensitive to nuclease degradation (Xu and Kool, 1998); hence, this substitution is commonly found in antisense technology (Dias and Stein, 2002). Nucleoside and deoxynucleoside triphosphates bearing sulfur at the α-phosphate position are recognized by a number of DNA and RNA polymerases, although their incorporation efficiency varies depending on the enzyme and the nature of nucleic acid polymer. In a very early example, Ellington and colleagues showed that fully substituted PS RNA can be generated by in vitro transcription using standard T7 RNA polymerase and all four (αS) NTPs (Jhaveri et al., 1998). Enzymatic incorporation of (αS) dNTPs by DNA polymerases proceeds less efficiently and only oligonucleotides with partial substitution are possible using standard DNA polymerases. Recognizing this problem, Holliger and colleagues
mutant taq DNA polymerases that can support the PCR amplification of fully substituted phosphorothioate DNA polymers up to 2 kb in length (Ghadessy et al., 2004). While thiopatamers have received broad interest, this backbone substitution is less common than substitutions made to the 2’ ribo-position. One reason for this could be that PS oligonucleotides are sticky molecules that can recognize off-target proteins with nonspecific binding interactions. Nevertheless, in vitro evolution has produced thiopatamers with affinity to a number of proteins, including molecules that bind to human basic fibroblast growth factor (Jhaveri et al., 1998), nuclear factor for human IL6 (King et al., 1998), the RNase H domain of HIV RT (Somasunderam et al., 2005), the Venezuelan equine encephalitis virus capsid protein (Kang et al., 2007), and transforming growth factor-b1 (Kang et al., 2008). The thiopatamer selected to bind the RNase H domain of HIV RT is interesting in that it inhibited RNase H activity in a dose dependent manner across a broad range of virus inoculum. At an m.o.i. of < 0.005, viral suppression was comparable to the HIV drug AZT (Somasunderam et al., 2005).

Another interesting substitution that has been examined by in vitro evolution is boronophosphate RNA, or bRNA (Figure 2). Similar to PS RNA, bRNA is constructed using synthetic chemistry that replaces the nonbridging oxygen atom in the phosphodiester linkage with BH3. This substitution, which is isosteric and isoelectric with respect to the natural phosphodiester linkage, leads to improved nuclease stability (Hall et al., 2004, 2006). In the context of a DNA backbone, borano-DNA supports RNase H activity, suggesting a possible role for this modification in antisense technology (Li et al., 2007). Nucleotide triphosphates bearing the (xB) NTP substitution are accepted as substrates by T7 RNA polymerase. bRNA represents an interesting class of therapeutic aptamers that could find practical application in an experimental therapy called boron neutron capture therapy (BNCT) (Hawthorne, 1993). In BNCT, a patient is first injected with a boron compound that localizes to a cancerous tumor. The patient is then exposed to neutron irradiation, which causes the boron compound to emit alpha particles that destroy nearby cancer cells. In theory, borano-aptamers could function as smart therapeutics by delivering the boron atom to the cancer cell, while avoiding healthy cells. In a proof-of-principle demonstration, Burke and coworkers evolved a series of borano-aptamers that bound to the small molecule target, adenosine 5’-triphosphate (ATP) (Lato et al., 2002). ATP is a common target for new in vitro selection technologies as the ATP-binding RNA aptamer has a conserved motif that is easy to rediscover by in vitro selection (Sassanfar and Szostak, 1993). In this case, partially substituted bRNA was constructed by in vitro transcription using T7 RNA polymerase and NTP mixtures containing either (xB) UTP or (xB) GTP. In addition to identifying the conserved ATP-binding RNA motif, sometimes referred to as the Sassanfar aptamer, named after its discoverer Mandana Sassanfar, the authors found several aptamers whose ligand-binding activity required the borano modification. This example suggests that it should be possible to evolve borano-aptamers to other targets of biological interest.

Mirror Image RNA

Mirror image design provides a clever strategy for generating stable aptamers with high affinity and specificity for a given biological molecule. This approach is based on the principle of reciprocal chiral substrate specificity, which states that polymers of one stereoconfiguration will exhibit reciprocal chiral substrate specificity when constructed in the opposite stereoconfiguration. In simple terms, this means that when the L-RNA version of an RNA aptamer is synthesized, the new L-RNA aptamer will bind the mirror image of the target (Figure 2). Thus, using mirror image symmetry, L-RNA aptamers can be generated by carrying out in vitro selection experiments against the mirror image of natural biological targets. Fürste and colleagues first demonstrated this principle by identifying RNA sequences that could bind to L-adenosine and D-arginine (the enantiomers of the natural D-adenosine and L-arginine metabolites) (Klussmann et al., 1996; Nolette et al., 1996). Nucleic acid sequences present in the selection output were constructed by solid-phase synthesis as both L- and D-RNA. The sequences constructed of L-RNA were given the name “spiegelmers”—from the German word “spiegel,” meaning mirror—to distinguish them from the traditional D-RNA aptamers. Functional analysis studies revealed that spiegelmers and aptamers with the same sequence have identical binding affinity but reciprocal binding specificity. In this case, spiegelmers recognized the natural metabolites to the near exclusion of the unnatural enantiomers, while the aptamers exhibited the opposite binding pattern and bound the unnatural D-adenosine and L-arginine metabolites. Because L-RNA is an unnatural nucleic acid backbone, the evolved spiegelmers remain stable in human serum after 60 hr, while the natural RNA aptamers rapidly degrade.

The ability for spiegelmers to withstand nuclease degradation suggests that these molecules could have potential therapeutic value. The first biological assay involving a spiegelmer was reported in 1997 by Kim and Bartel (Williams et al., 1997). Here, a spiegelmer to the peptide hormone L-vasopressin was produced using the unnatural D-peptide analog as bait. After several rounds of in vitro selection and amplification followed by directed evolution, a spiegelmer was generated that binds to vasopressin. This molecule was shown to antagonize the vasopressin response in cultured kidney cells. Similar studies have since been performed on other pharmacological targets, including the peptide hormone gonadotropin-releasing hormone (GnRH) (Leva et al., 2002). Spiegelmers generated to GnRH were shown to inhibit GnRH binding to its cognate receptor in Chinese hamster ovary cells with half maximal inhibitory concentration (IC50) values of 50–200 nM. While the concept of mirror image design provides a convenient strategy for generating nuclease-resistant ligands, the requirement for mirror image designs limits this approach to small molecules and short proteins that are accessible by chemical synthesis. Recognizing this problem, Klussmann and colleagues developed a domain-based approach in which the peptide segment of a surface-accessible domain was used as a target in place of the whole protein (Purschke et al., 2003). In this case, the spiegelmer, which was selected to bind a 25-amino-acid segment of staphylococcal Enterotoxin B, exhibited nanomolar affinity to the full-length protein.

Joyce and colleagues recently extended the concept of mirror image symmetry to include catalytic RNA (Olea et al., 2012). In this case, a self-replicating all L-RNA enzyme and its all L-RNA substrate were prepared by solid-phase synthesis from L-nucleoside phosphoramidites. The whole system was based on
a previously discovered RNA ribozyme that had been developed for ligand sensing but was limited in its application due to its susceptibility to ribonuclease degradation. The L-ribozyme was shown to undergo isothermal, theophylline-dependent exponential amplification, in a manner analogous to the natural RNA ribozyme system. However, unlike the RNA ribozyme, the L-RNA system was impervious to nuclease degradation. This demonstration suggests that it should be possible to create other nuclease-resistant sensors with aptamer recognition domains that function under a wide range of natural conditions.

**Mosaic Nucleic Acids**

Motivated by a desire to understand chemical steps that led to the emergence of early genetic systems, Szostak and coworkers have demonstrated that nucleic acid molecules with heterogeneous backbones can evolve ligand-binding functions despite the repeated shuffling of backbone chemistry between iterative cycles of selection (Trevino et al., 2011). In a simplified model system, a mutant T7 RNA polymerase (Y639F) capable of transcribing nucleic acids that contain both deoxy- and ribonucleotides was used to construct a library of oligonucleotides that contained equal amounts of DNA and RNA residues. These DNA-RNA chimeric polymers were given the name mosaic nucleic acids (MNA). The library was used to isolate nucleotide-binding MNA aptamers by in vitro selection. After each round of selection, functional sequences were reverse-transcribed into DNA and amplified by RT-PCR. The molecules were then forward-transcribed back into MNA, which retained the nucleotide sequence but randomized the sugar-phosphate backbone. As a consequence, sequences that depended on specific sugar arrangements were selected against, while sequences that retained function despite variation at the sugar positions were enriched. Following multiple rounds of selection and amplification, MNA aptamers were identified from two parallel selections that targeted ATP and GTP, respectively. The selected aptamers bound specifically to their cognate ligands, demonstrating that nonheritable backbones may not have posed an insurmountable obstacle to the emergence of early functional molecules. Whether similar selections could yield MNA molecules with catalytic activity is an interesting question that has yet to be determined.

**XNA Polymers with Novel Backbone Structures**

**Peptide Nucleic Acid**

One of the most interesting molecules developed as a nucleic acid analog is peptide nucleic acid (PNA; Figure 3). Unlike all DNA and RNA analogs described thus far, which have a sugar-phosphate backbone, PNA has a backbone composed of repeating N-(2-aminoethyl)-glycine units that are linked by peptide bonds (Nielsen et al., 1991a). Since PNA has an uncharged backbone, hybridization with DNA is stronger than DNA/DNA hybridization due to the absence of electrostatic repulsion. PNA is also more specific for a given target sequence than DNA or RNA, as base pair mismatches are more destabilizing for PNA than a natural oligonucleotide of the same sequence (Nielsen, 1995). These properties, coupled with enhanced nuclease stability, make PNA a useful molecule for antisense therapy (Demidov et al., 1994). Although PNA cannot easily cross the cellular membrane, conjugation of PNA antisense oligonucleotides to cell-penetrating peptides has been used successfully to improve cellular delivery (Shiraishi and Nielsen, 2011; Wittung et al., 1995).

Recognizing the potential that synthetic polymers could have in expanding the functional utility of genetic polymers, Liu and colleagues recently described an in vitro translation, selection, and amplification system for PNA (Brudno et al., 2010). Since PNA is not a substrate for any known polymerase, conditions were developed that allowed for the nonenzymatic sequence-specific oligomerization of tetramer and pentamer PNA building blocks on DNA templates. These building blocks collectively represent a “genetic code” for PNA that can be faithfully “translated” on a DNA template using reductive amination chemistry to link multiple PNA units together. For example, 10 consecutive coupling reactions of pentamer codons lead to the synthesis of a 50-mer PNA molecule containing secondary amine linkages between every fifth nucleotide. Translation on a library of self-primer DNA templates produced a pool of PNA-DNA hairpins in which one strand was composed of PNA and the other strand was composed of DNA. As illustrated in Figure 4, the DNA portion of these molecules was then made double-stranded by extending a DNA primer annealed to the stem-loop region of the hairpin with dNTPs. This difficult step required Herculase II, a strong thermophilic DNA polymerase to copy the DNA strand by displacing the PNA strand. The resulting PNA-dsDNA molecules contained the genotype-phenotype linkage necessary to perform in vitro selection since the PNA molecules could be selected on the basis of their chemical function and their encoding DNA information could be amplified by their attached DNA sequence. The authors validated this approach in a proof-of-principle demonstration by performing five iterative cycles of translation, selection, and amplification for PNA molecules that could incorporate a biotinylated codon into their sequence. The selection showed an overall enrichment of >105-fold of PNA-encoding DNA templates on streptavidin-coated beads. This example suggests that it should be possible to isolate functional PNA aptamers by in vitro selection.

**Threose Nucleic Acid**

Among the many different alternative nucleic acid molecules developed by Eschenmoser and colleagues, 1-L-threose nucleic acid (TNA) has emerged as the first genetic polymer with relevance in synthetic genetics (Figure 3) (Schöning et al., 2000). TNA is an artificial nucleic acid polymer in which the natural five-carbon ribose sugar found in RNA has been replaced with an unnatural four-carbon threose sugar and phosphodieste

---

**Figure 3. XNA Polymers with Diverse Nucleic Acid Backbone Structures**

HNA and TNA are XNA systems in which the natural ribose sugar found in RNA has been replaced with a nonribose sugar moiety. HNA and TNA require engineered polymerases for their replication. PNA is a nucleic acid system based on repeating amide bonds that replicates by nonenzymatic template-directed polymerization.
linkages are connected at the 2' and 3’ vicinal positions. This substitution shortens the backbone repeat unit by one bond as compared to natural DNA and RNA. Eschenmoser found that TNA undergoes informational Watson-Crick base pairing in an antiparallel strand orientation and also cross-pairs opposite complementary strands of DNA and RNA (Schöning et al., 2000). This latter feature is remarkable given the backbone difference between TNA and the natural genetic polymers of DNA and RNA.

The ability for TNA to exchange genetic information with RNA, coupled with the chemical simplicity of threose relative to ribose, have prompted careful consideration of TNA as a possible RNA progenitor (Orgel, 2000). To explore this possibility, we and others have attempted to identify polymerases that could be used to examine the functional properties of TNA by in vitro selection. Numerous primer-extension assays were performed to identify DNA polymerases that could copy DNA templates into TNA (Chaput and Szostak, 2003; Kempeneers et al., 2003) and other polymerases that could copy TNA back into DNA (Chaput et al., 2003). By identifying enzymes that could perform these functions, we aimed to develop a replication system for TNA that allowed pools of TNA molecules to be reverse-transcribed into DNA, amplified by PCR, and forward-transcribed back into TNA. From this body of work, it was discovered that therminator DNA polymerase, an engineered variant of the 9 nA485L mutation functions as an efficient DNA-dependent TNA polymerase (Horhota et al., 2005; Ichida et al., 2005). Under optimal conditions, therminator was found to copy a long DNA template into TNA with high efficiency and fidelity but failed to copy a pool of DNA sequences into TNA (Ichida et al., 2005). We subsequently determined that stretches of G-nucleotides in the DNA template sequence lead to polymerase pausing and chain termination. By designing DNA libraries that either eliminated all G-residues in the template or presented G nucleotides at sufficiently low frequency so they would occur as isolated residues, we were able to efficiently transcribe DNA libraries into TNA (Yu et al., 2012). Consequently, the resulting TNA sequences were either devoid of cytidine or contained cytidine at reduced frequency.

To determine whether TNA could fold into structures with discrete ligand-binding properties, a DNA display strategy similar to the hairpin strategy developed by Liu and colleagues was used to evolve TNA aptamers with affinity to human thrombin (Figure 4) (Brudno et al., 2010). This approach provides a solution to the problem of how to evolve XNA polymers when an RT is not available to convert the pool of XNA molecules back into DNA for amplification by PCR. Accordingly, a pool of self-priming DNA templates was extended with TNA using therminator DNA polymerase to copy the DNA library into TNA. The DNA region of the TNA-DNA chimeric hairpins was made double-stranded by extending a separate DNA primer annealed to the step-loop structure with dNTPs. The resulting TNA-dsDNA molecules produced the genotype-phenotype linkage necessary to perform iterative rounds of in vitro selection and amplification since TNA molecules could be selected on the basis of function and amplified on the basis of their attached DNA sequence. After several rounds of in vitro selection and amplification, TNA aptamers emerged with high binding affinity and specificity to human thrombin. The best aptamer identified in the selection bound to thrombin with an equilibrium dissociation constant of 200 nM, which is similar to DNA and RNA aptamers previously evolved to bind human thrombin. The isolation of TNA aptamers from a pool of random TNA molecules demonstrated, for the first time, that the problem of ligand binding is not unique to the natural polymers of RNA and DNA or close structural analogs thereof (Yu et al., 2012). Indeed, one could imagine that purely chemical constraints, like a shorter backbone repeat unit, might preclude the ability for TNA to fold into structures with a desired function. Since TNA does not appear to be limited in this regard, it seems reasonable that similar selections for catalytic function could be used to isolate novel XNA enzymes or “threozymes” by in vitro evolution. As these studies progress, it will be interesting to see how the functional properties of TNA compare to other XNA molecules.

Figure 4. In Vitro Selection Strategies for XNA Evolution
(A) DNA display provides a general strategy for evolving XNA polymers when an RT is not available to convert XNA sequences back into DNA for amplification by PCR. In this technique, a genotype-phenotype link is established by extending a self-priming DNA library with XNA to link each XNA molecule to its encoding DNA sequence. (B) When an XNA transcriptase and RT are available, in vitro selection can be performed using a traditional two-enzyme replication strategy.
Hexitol Nucleic Acid

Hexitol nucleic acid (HNA) was the second XNA polymer to be studied in the context of synthetic genetics (Figure 3) (Pinheiro et al., 2012). Developed by Herdewijn and colleagues as an antisense reagent (Verheggen et al., 1993, 1995), HNA has a backbone structure composed of 1,5-anhydrohexitol, which is a six-membered pyranosyl ring structure. The nucleobase is located at the 2' position and phosphodiester linkages occur between the 4' and 6' carbon atoms. Unlike most other RNA analogs with six-membered carbohydrate moieties (Herdewijn, 2010), such as homo-DNA and pyranosyl-RNA (Krishnamurthy et al., 1996), HNA is capable of base pairing opposite complementary strands of itself and RNA and DNA (Hendrix et al., 1997a, 1997b). The rigid anhydrohexitol ring adopts a chair conformation that causes HNA to form an A-type helical structure (Declercq et al., 2002; Lescrinier et al., 2000; Maier et al., 2005). Like PNA, HNA is also more specific for a given target sequence than DNA or RNA and HNA is able to distinguish base pair mismatches more easily than a natural oligonucleotide of the same sequence (Hendrix et al., 1997b).

To explore the functional properties of HNA, Holliger and coworkers used a compartmentalized self-tagging strategy to evolve DNA polymerases capable of copying DNA templates into HNA (Pinheiro et al., 2012). The authors constructed a library of polymerase variants based on the DNA enzyme Thermococcus gorgonarius (Tgo) polymerase by randomly incorporating genetic mutations into the parent gene. The library was then used to identify mutant polymerases capable of synthesizing HNA polymers on DNA templates. The selection produced several enzymes that exhibited broad substrate specificity, including polymerases that could transcribe HNA, cyclohexenyl nucleic acids, locked nucleic acids (Campbell and Wengel, 2011), TNA, arabinose nucleic acid (Noronha et al., 2000) and 2'-fluoro arabinose nucleic acid (Wilds and Damha, 2000). To reverse-transcribe the XNA polymers back into DNA, the authors evolved several novel RTs that could copy HNA and other XNA polymers back into DNA. Their starting library utilized statistical correlation analysis to identify an allosteric network responsible for template recognition and RNA-RT activity in the polB family of DNA polymerases. Neighboring residues in the vicinity of possible hits were mutated and screened in a polymerase activity assay. This approach identified a new TgoT mutant, called RT521, that was a proficient HNA RT. Together, the two enzymes allowed for the in vitro evolution of HNA aptamers (Figure 4). HNA aptamers were generated against HIV trans-activating response RNA and the protein hen egg lysozyme. Both aptamers were found to bind their targets with high affinity and specificity, consistent with the interpretation that HNA, like TNA, can fold into shapes with very specific ligand binding sites (Yu et al., 2012). Perhaps even more importantly, however, the Holliger study provided a general approach for creating XNA polymerases with broad application in synthetic genetics.

Moving Beyond DNA and RNA: Implications for Exobiology, Molecular Medicine, and Synthetic Biology

Armed with a new generation of polymerases and a large collection of nucleic acid modifications, the field of synthetic genetics plans to expand the frontier of in vitro genetics to include a growing list of structurally diverse XNA polymers (Figure 5). Motivation for this work comes, in part, from a strong desire to understand basic physical properties that may have contributed to nature’s choice of ribofuranosyl nucleic acids (DNA and RNA) as life’s genetic material (Eschenmoser, 1999). Since it is impossible to rewind the evolutionary clock of time, experiments of this type provide an important model for evaluating the fitness of a given genetic polymer for a particular chemical function. Although all forms of life (including those that are known to have existed but are now extinct) are based on organisms that

Figure 5. Promising XNA Systems with Potential Application in Synthetic Genetics

The future of synthetic genetics will likely include a wide range of alternative genetic systems that are each capable of undergoing Darwinian evolution. These examples highlight a few promising candidates for future XNA polymers.
stored genetic information in DNA genomes and catalyzed reactions with protein enzymes, earlier forms of life may have relied on one biopolymer, not DNA or protein per se, but maybe something else (Joyce, 2002). One possibility is that life evolved from simple organisms that were based on RNA (Crick, 1968; Orgel, 1968; Woese, 1967). This postulate, commonly referred to as the RNA world hypothesis (Gilbert, 1986), received widespread popularity following the discovery that RNA catalysts exist in nature (Kruger et al., 1982). Since then, many other examples of RNA enzymes, or ribozymes, have been discovered and others with functions more relevant to the RNA world have been created by in vitro selection (Joyce, 2004; Wilson and Szostak, 1999). Despite these observations, it is difficult to envision how a molecule as complicated as RNA could have emerged on the early Earth and remained in the environment long enough to replicate and produce offspring. This concern has led some to suggest that RNA was not the first genetic material, but rather an important evolutionary intermediate in the path to extant life (Engelhart and Hud, 2010; Joyce et al., 1987).

If RNA was not the first genetic material, then what genetic polymers came before RNA? Whatever prebiotic chemistry gave rise to RNA would have almost certainly produced other RNA analogs, some of which could have preceded or competed directly with RNA. This hypothetical period in evolutionary history is referred to as the pre-RNA world—a time when chemistry gave rise to simple self-replicating polymers that stored genetic information and catalyzed chemical reactions (Lazzaro and Miller, 1996). What these structures looked like and how they were produced is a mystery that has challenged the minds of many great chemists. Suggestions include PNAs (Nielsen et al., 1996), an acyclic derivative of TNA. These alternative genetic systems are interesting because they are: (1) physically realistic, meaning that they can be constructed by chemical synthesis; (2) capable of storing genetic information via Watson-Crick base pairing; and (3) chemically simpler than RNA, meaning that prebiotic chemistry could, in theory, have favored their synthesis over RNA. With the advent of synthetic genetics, it should be possible to evaluate the functional properties for at least a subset of synthetic genetic polymers. Since some of these molecules are also strong candidates for early RNA progenitors, these studies provide a basis for comparing molecular functions across a range of biopolymer systems. As these studies progress, an important goal will be to develop XNA enzymes that can catalyze their own synthesis. If successful, these self-replicating molecules could be used to develop synthetic life forms with unnatural genomes.

In addition to addressing basic questions related to life’s first genetic system, synthetic genetics also provides an opportunity to develop nuclease-resistant molecules for biotechnology and molecular medicine. Nucleic acid molecules with specific target recognition and catalytic properties have become valuable tools in many areas of basic and applied research (Krishnan and Simmel, 2011; Mascini et al., 2012; Tombelli and Mascini, 2009). Unfortunately, most of the molecules created thus far are composed of DNA and RNA, which are susceptible to nuclease degradation and therefore not suitable for most biological applications. In some cases, this problem can be overcome by chemically modifying the nucleic acid backbone after the selection is complete; however, structural changes of this type are often expensive to implement and come with uncertain outcomes. Synthetic genetics offers an alternative solution to this problem by producing functional molecules with nucleic acid backbones that are not recognized by natural enzymes. This property of molecular invisibility represents a possible paradigm shift, as functional molecules will now be cloaked behind an unnatural nucleic acid backbone. Whether XNA polymers will be able to supplant their DNA and RNA counterparts will depend on how well these molecules function in biological environments and the availability of XNA triphosphates and enzymes needed to produce these molecules by in vitro selection.

Beyond therapeutic and diagnostic applications, synthetic genetics could find widespread use in future synthetic biology projects. Many practitioners of synthetic biology treat gene networks found in living organisms as electronic circuitry that can be “rewired” to produce engineered organisms with new metabolic pathways embedded in their genomes (Benner and Sismour, 2005). This process of reading and writing genetic information relies on the fact that DNA is a universal language recognized by all organisms. Consequently, synthetic biologists often view DNA, as a plug-and-play technology where genetic pathways are seamlessly transplanted from one organism into another. Unfortunately, in practice, DNA does not always function in a predictable manner, and unexpected problems can arise due to such genetic properties as epistasis and pleiotropy, which occur when the effects of one gene are modified by other genes or when a single gene exhibits multiple phenotypic traits, respectively. Another disadvantage of using DNA is that safety measures are difficult to engineer, as wild-type and engineered genes cannot be distinguished by the cellular machinery responsible for transcribing and translating genetic information. This problem raises the concern that aberrant mutations in engineered organisms could have unintended consequences that are harmful to society if the organism escapes into the wild. One way to mitigate the risks of bioengineering is to use genetic polymers that are not recognized by natural enzymes (Herdewijn and Mariére, 2009). Using this approach, engineered organisms that escape into the wild would pose a minimal risk to society because XNA substrates and XNA enzymes do not exist naturally. In this light, XNA polymers represent an attractive genetic material for synthetic biology as their molecular invisibility provides the orthogonality needed to establish a genetic firewall between nature and synthetic biology (Schmidt, 2010). Although widespread use of synthetic genetic polymers may be years away, the potential for XNA polymers to function as safe and effective tools in synthetic biology warrants continued technological development.

**Conclusions**

The emerging field of synthetic genetics provides a rich opportunity to explore the functional landscape of artificial genetic polymers by in vitro evolution. Currently, the most promising xenonucleic acids are TNA and HNA—two unnatural genetic polymers capable of undergoing Darwinian evolution in the laboratory. We anticipate that the continued development of nucleic acid chemistry and molecular biology techniques will enable the
creation of additional examples of XNA polymers that exhibit heredity and evolution (DeLeavey and Damha, 2012; Herdewijn and Marlière, 2012). Further pursuit of these polymers provides an exciting opportunity to develop novel genetic polymers with significant downstream applications in exobiology, molecular medicine, and synthetic biology.

ACKNOWLEDGMENTS

The authors thank G. Joyce, P. Herdewijn, and members of the Chaput lab for helpful comments and suggestions. This work was supported by the Biodesign Institute at Arizona State University.

REFERENCES


