# Orthogonal Genetic Systems

John Chaput, Piet Herdewijn, Marcel Hollenstein

# Orthogonal Genetic Systems

John C. Chaput, Piet Herdewijn, and Marcel Hollenstein

Xenobiology is an area of synthetic biology that aims to create model cellular organisms in which all synthetic biology information is encoded in artificial genetic polymers, commonly referred to as xeno-nucleic acids or XNAs.[1] Xenobiology cells created in this way would safeguard synthetic biology efforts by establishing a firewall separating synthetic biology information from natural biological information.[2] This challenging endeavor requires chemical synthesis to produce nucleic acid monomers (xNTPs) that are not commercially available and protein engineering to generate the requisite enzymatic machinery needed to synthesize and propagate genetic information encoded in strands of XNA polymers. To date, a significant number of advances have been made to support the development of XNA systems, including the generation of laboratory-evolved polymerases that have enabled the isolation of aptamers and catalysts with backbone structures that are distinct from those found in nature.[3]

Extending this work to cellular systems that can maintain an XNA chromosome (episome) in actively dividing cells is a daunting task that will require new technological advances. Among the various problems facing those wishing to pursue this challenging endeavor is the need to establish an orthogonal genetic system that can replicate separately from the endogenous genome. This means that the orthogonal genome and enzymatic machinery required to replicate the synthetic genome cannot interfere with the natural biosynthetic pathways of the cell and, vice versa, the natural system cannot interfere with the replication of the synthetic system. Here the term ‘orthogonality’ refers to genetic orthogonality, which describes the separation of genetic information between two different classes of nucleic acid molecules (e.g., XNA and DNA). This use of the term ‘orthogonality’ differs from its chemical biology usage, which describes chemical or enzymatic reactions that occur with high specificity, such as the site-specific labeling of a biological macromolecule or the introduction of a non-canonical residue into DNA or protein.[4]

Thus far, most of the work in this area has focused on the evolution of synthetic genetic polymers by in vitro selection.[5] The XNAs evaluated in these systems represent a special subset of synthetic genetic polymers because they have the unique ability to form stable antiparallel Watson-Crick duplex structures with DNA and RNA, thus allowing information to pass between different classes of nucleic acid molecules. This feature of biocompatibility simplifies the selection process by allowing genetic information present in the starting DNA library to be transcribed into XNA using an engineered DNA-dependent XNA polymerase and XNA members captured during the selection step to be reverse transcribed back into DNA using an XNA-dependent DNA polymerase.[3c, 6] For each cycle of selection and amplification, the recovered cDNA molecules are amplified by PCR, made single-stranded, and used as templates to construct a new population of XNA molecules that has become enriched in molecules that exhibit a desired functional property.[7] An important aspect of this strategy is that users can take advantage of infrastructure that is already available for the writing and reading DNA sequences by solid-phase synthesis and next-generation sequencing, respectively.[8]

The ability for certain XNAs to cross-pair with DNA and RNA depends on the sugar moiety. By definition, nucleic acids consist of three main parts: the nucleobase, phosphodiester linkage, and sugar moiety. Although each group plays an important role in the structural and

functional properties of nucleic acid polymers, it is the sugar moiety that determines the helical geometry of the duplex.[9] DNA, for example, favors a B-form helix with deoxyribose adopting a 2'-*endo* sugar pucker, while RNA prefers an A-form helix with ribose adopting a 3'-*endo* sugar pucker. As expected, changes to the helical conformation occur when the natural sugar is replaced with a different type of sugar. For example, arabino nucleic acid (ANA), an RNA analog in which the stereochemistry of the 2' carbon atom positions the 2' hydroxyl in the upward configuration, adopts a B-form helical geometry rather than the standard A-form geometry commonly observed for RNA.[10] Similarly, when a 4' thio group is introduced into the deoxyribose sugar, the corresponding DNA analog adopts an A-form helix rather than the standard B-form geometry commonly observed for DNA.[11] Figure 1 provides a comparison of the helices formed by these natural and slightly modified nucleic acid systems.

The structural parameters that define the helical geometry of DNA and RNA duplexes have been characterized in detail using information obtained by NMR and X-ray crystallography.[12] The parameters that describe the helicity of a double stranded nucleic acid structure in the most straightforward way are the average helical rise and the average helical twist (Fig. 2).[13] These parameters provide the average stacking distances between bases and the number of nucleotides per helical turn. Standard B-form DNA, for example, has an average helical twist of 36° and an average helical rise of 3.37 Å, which leads to a duplex with 10 nucleotides per helical turn.[9] By comparison, standard A-form RNA has an average helical twist of 31° and an average helical rise per residue of 2.9 Å, which produces a duplex with 11.5 nucleotides per helical turn.[9]

Although a reasonably large number of XNAs have been reported in the literature, most are incapable of cross-pairing with DNA and RNA.[14] These systems have structural parameters that lie outside the chemical space defined by A- and B-form helices.[15] However, two prominent cases have been studied in detail where XNA cross-pairing with DNA and RNA occurs with high efficiency. The first is threose nucleic acid (TNA), which is an artificial genetic polymer in which the natural five-carbon ribose sugar has been replaced with a synthetic four-carbon threose sugar.[16] TNA forms a B-like helical conformation that is similar to DNA with an average helical twist of 26° and an average helical rise of 3.2 Å, which leads to a duplex with 10 nucleotides per helical turn.[17] The second case is anhydrohexitol (HNA), which derives from a six-carbon pyranosyl sugar.[18] HNA forms an A-like helical conformation that is similar to RNA with an average helical twist of 31° and an average helical rise of 2.6 Å, which leads to a duplex with 11 nucleotides per helical turn.[19]

Interestingly, both TNA and HNA show signs of structural heterogeneity depending on whether they are base paired with themselves or cross-paired with a natural genetic polymer.[17, 19-20] This property of structural flexibility suggests that TNA and HNA might be capable of exchanging genetic information with nucleic acid polymers other than DNA and RNA. In the event that such cases could be identified, TNA and HNA might serve as a bridge (Fig. 3) by allowing information to flow from natural genetic polymers into synthetic genetic polymers that are not capable of base pairing with DNA and RNA. Preliminary data has already shown that HNA can base pair with β-homo-DNA, α-homo-DNA, and cyclohexyl nucleic acid,[21] suggesting that certain types of genetic polymers may have a greater propensity for XNA-XNA cross-pairing than others. Other experiments reveal that some polymerases can synthesize limited stretches of XNA on DNA templates, and vice versea, certain polymerases can copy short XNA templates into DNA.[22]

Orthogonal genetic polymers are incapable of cross-pairing with natural genetic polymers, because they have structural parameters that are incompatible with A- and B-form helices. Deoxyxylo nucleic acid (dXyloNA), for example, is an epimer of DNA in which the 3' hydroxyl group is in the upward configuration and the sugar adopts a 3'-*endo* conformation rather than the 2'-*endo* sugar pucker typically observed for DNA.[23] This relatively subtle chemical change leads to an enormous structural change with self-pairing duplexes adopting left-handed helices that are far removed from the chemical space of DNA and RNA.[24] NMR analysis reveals the existence of a ladder like structure (Fig. 3) with an average helical twist of only 2.7°. A similar case is observed for xylo nucleic acid (XyloNA), an equivalent RNA analog that adopts a stable self-pairing system that is also orthogonal to DNA and RNA.[25] Pyranosyl RNA (pRNA) is another self-pairing system that adopts a ladder-like structure that is orthogonal to DNA and RNA.[26] pRNA, like HNA, have sugar moieties based on a six-membered pyranosyl carbohydrate ring rather than the more natural five-membered furanosyl carbohydrate ring.

The relationship between the base pairing properties and helical geometry of furanosyl and pyranosyl nucleic acids led to the false assumption that five-membered furanosyl systems are helical, while six-member pyranosyl systems are ladder-like. However, structural insights into the helical conformations of XyloNA and HNA have changed the nucleic acid dogma by showing that five-membered systems (XyloNA) can be ladder-like and six-membered systems (HNA) can be helical (Fig. 3). These observations raise interesting questions about the conformational space of nucleic acid polymers and the potential for new duplex structures. Are helical and linear conformations the only way that XNAs can self-assemble into duplex structures or do other examples of physically realistic structures exist that have yet to be discovered? Similarly, to what extent can these systems base pair with other XNA systems, and is there a mechanism for the transfer of functional information when replication is complete?

The linear duplexes observed for XyloNA and pRNA make them logical candidates for orthogonal genetic materials, as there is no clear way for them to base pair with DNA and RNA. However, this same feature will also make it difficult to create the enzymatic machinery necessary to replicate these systems in vitro and eventually in cells. Perhaps the structural flexibility of TNA and HNA could serve as a conduit by providing a pathway for transmitting genetic information from helical structures that are compatible with DNA to linear structures that are orthogonal to DNA and RNA. Such systems provide an interesting strategy for establishing genetic orthogonality and would undoubtedly lead to new fundamental knowledge about the ability for nucleic acid polymers to store and transmit genetic information.

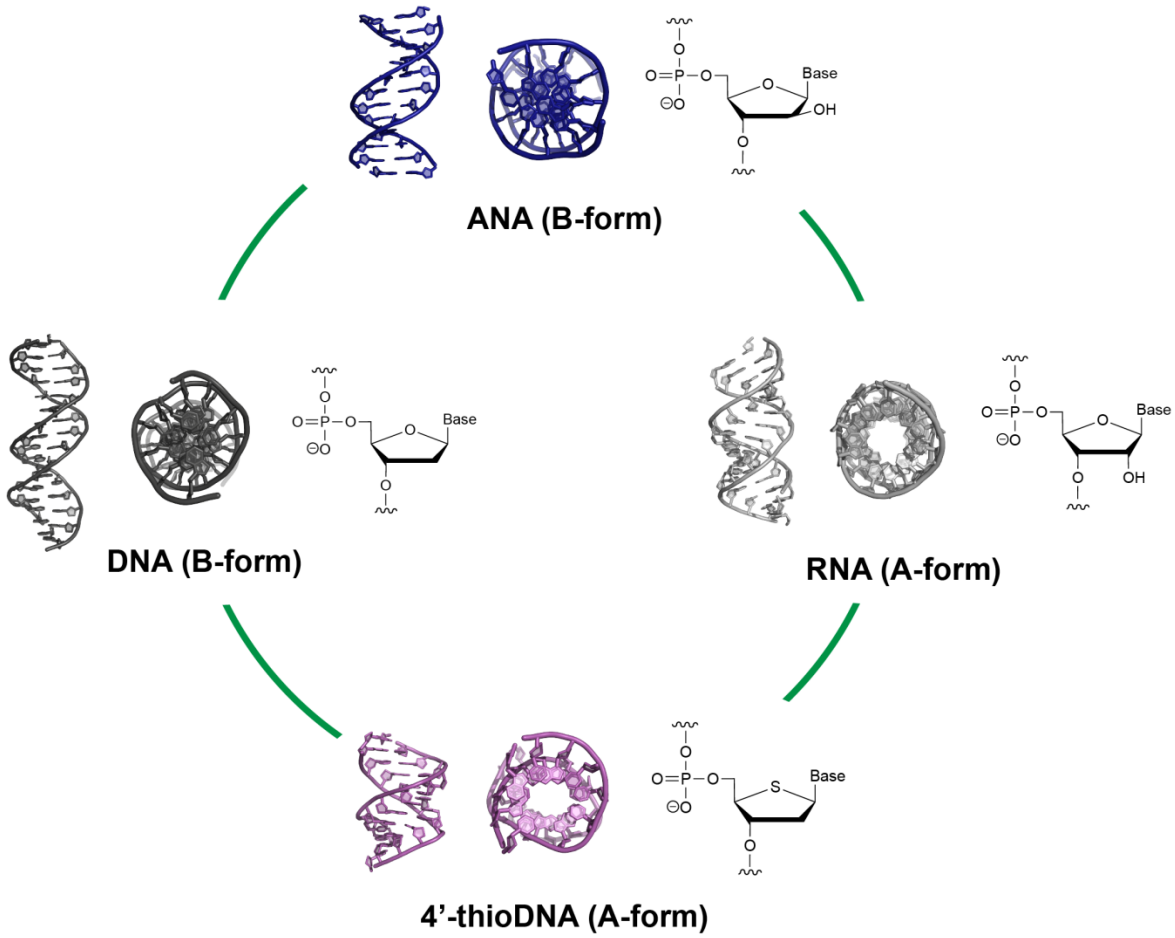


**Figure 1.** Helical structures of natural and closely related synthetic genetic systems. DNA and ANA adopt B-form helical geometries, while RNA and 4'-thioDNA adopt A-form helical structures.

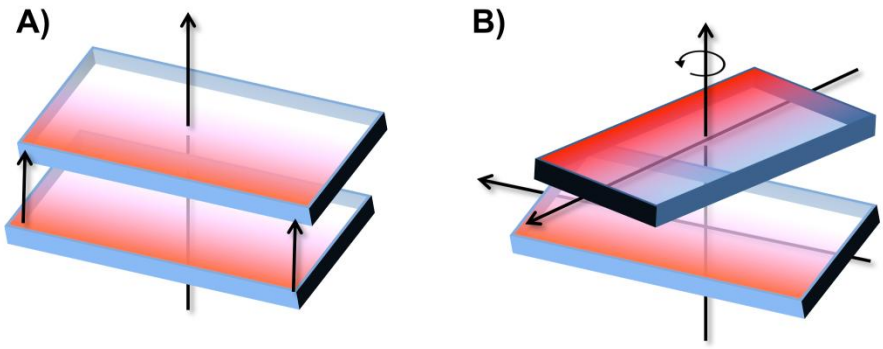


**Figure 2.** Schematic representation of A) helical rise and B) helical twist in nucleic acid duplexes.

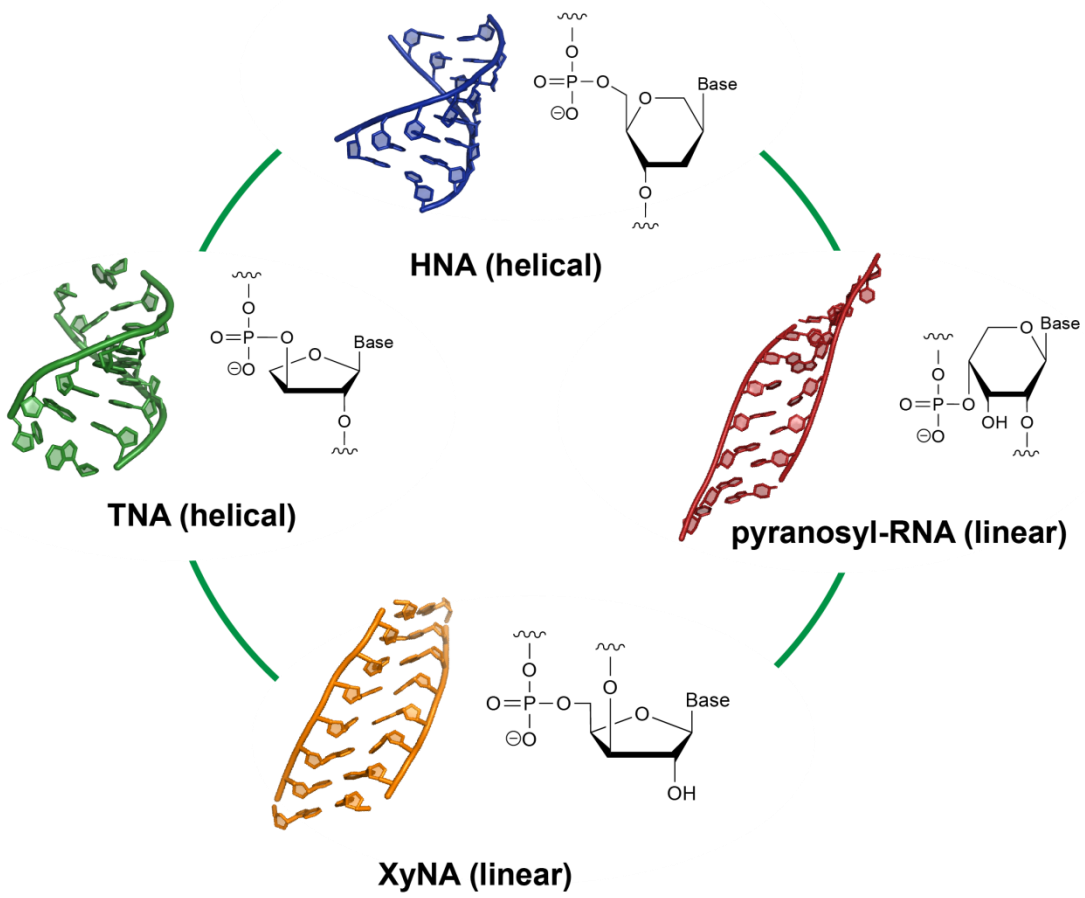


**Figure 3.** Helical and linear structures of synthetic genetic systems (XNAs). TNA and HNA have the capacity to exchange information with DNA and RNA, while XyNA and pyranosyl-RNA are orthogonal genetic systems.

## References


[1] P. Herdewijn, P. Marliere, *Chem Biodivers* **2009**, *6*, 791.
[2] M. Schmidt, *BioEssays* **2010**, *32*, 322.
[3] a)A. C. Larsen, et al., *Nature Communications* **2016**, *7*, 11235; b)A. Nikoomanzar, D. Vallejo, J. C. Chaput, *ACS Synth Biol* **2019**, *8*, 1421; c)V. B. Pinheiro, et al., *Science* **2012**, *336*, 341.
[4] a)K. Lang, J. W. Chin, *Chem Rev* **2014**, *114*, 4764; b)J. W. Chin, *Annu Rev Biochem* **2014**, *83*, 379.
[5] a)J. C. Chaput, H. Yu, S. Zhang, *Chemistry and Biology* **2012**, *19*, 1360; b)G. Houlihan, S. Arangundy-Franklin, P. Holliger, *Acc Chem Res* **2017**, *50*, 1079; c)M. R. Dunn, R. M. Jimenez, J. C. Chaput, *Nat. Rev. Chem.* **2017**, *1*, 0076.
[6] a)H. Mei, et al., *J. Am. Chem. Soc.* **2018**, *140*, 5706; b)E. Eremeeva, et al., *Nucleic Acids Res.* **2019**, *47*, 4927; c)S. Arangundy-Franklin, et al., *Nat. Chem.* **2019**, *11*, 533.
[7] F. Levi-Acobas, et al., *Org. Biomol. Chem.* **2019**, *17*, 8083.
[8] a)S. Kosuri, G. M. Church, *Nature Methods* **2014**, *11*, 499; b)S. Goodwin, J. D. McPherson, W. R. McCombie, *Nat Rev Genet* **2016**, *17*, 333.
[9] W. Saenger, *Principles of Nucleic Acid Chemistry*, Springer-Verlag, New York, **1984**.
[10] N. Martin-Pintado, et al., *Nucleic Acids Res* **2012**, *40*, 9329.
[11] A. Matsugami, et al., *Nucleic Acids Res* **2008**, *36*, 1805.
[12] a)B. Schneider, S. Neidle, H. M. Berman, *Biopolymers* **1997**, *42*, 113; b)X. J. Lu, Z. Shakked, W. K. Olson, *J Mol Biol* **2000**, *300*, 819; c)X. J. Lu, W. K. Olson, *Nucleic Acids Res* **2003**, *31*, 5108.
[13] M. Ghorbani, F. Mohammad-Rafiee, *Nucleic Acids Res.* **2011**, *39*, 1220.
[14] A. Eschenmoser, *Science* **1999**, *284*, 2118.
[15] I. Anosova, et al., *Nucleic Acids Research* **2016**, *44*, 1007.
[16] K. U. Schöning, et al., *Science* **2000**, *290*, 1347.
[17] L. N. Jackson, N. Chim, C. Shi, J. C. Chaput, *Nucleic Acids Res* **2019**, *47*, 6973.
[18] C. Hendrix, et al., *Chemistry European Journal* **1997**, *3*, 110.
[19] T. Maier, I. Przylas, N. Strater, P. Herdewijn, W. Saenger, *J Am Chem Soc* **2005**, *127*, 2937.
[20] a)M.-O. Ebert, C. Mang, R. Krishnamurthy, A. Eschenmoser, B. Jaun, *J Am Chem Soc* **2008**, *130*, 15105; b)R. Declercq, A. Van Aerschot, R. J. Read, P. Herdewijn, L. Van Meervelt, *J Am Chem Soc* **2002**, *124*, 928.
[21] L. Kerremans, et al., *Org. Lett.* **2001**, *3*, 4129.
[22] a)M. Renders, M. Abramov, M. Froeyen, P. Herdewijn, *Chemistry* **2009**, *15*, 5463; b)C.-H. Tsai, J. Chen, J. W. Szostak, *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 14598.
[23] M. Maiti, et al., *Chemistry* **2012**, *18*, 869.
[24] A. Ramaswamy, M. Froeyen, P. Herdewijn, A. Ceulemans, *J. Am. Chem. Soc.* **2010**, *132*, 587.
[25] M. Maiti, et al., *Nucleic Acids Res.* **2015**, *43*, 7189.
[26] R. Krishnamurthy, et al., *Angew Chem Int Edit* **1996**, *35*, 1537.