Published in final edited form as:
Science. 2012 April 20; 336(6079): 341–344. doi:10.1126/science.1217622.

Synthetic genetic polymers capable of heredity and evolution

Vitor B. Pinheiro[1], **Alexander I. Taylor**[1], **Christopher Cozens**[1], **Mikhail Abramov**[2], **Marleen Renders**[2,†], **Su Zhang**[3], **John C. Chaput**[3], **Jesper Wengel**[4], **Sew-Yeu Peak-Chew**[1], **Stephen H. McLaughlin**[1], **Piet Herdewijn**[2], and **Philipp Holliger**[1,*]

[1]MRC Laboratory of Molecular Biology, Hills Road, Cambridge, CB2 0QH, UK

[2]REGA Institute, Katholieke Universiteit Leuven, Minderbroederstraat 10, B 3000, Leuven, Belgium

[3]Center for Evolutionary Medicine and Informatics, The Biodesign Institute at Arizona State University, 1001 S. McAllister Avenue, Tempe, AZ 85287-5301, USA

[4]Nucleic Acid Center, Department of Physics and Chemistry, University of Southern Denmark, Campusvej 55, DK-5230 Odense M, Denmark

Abstract

Genetic information storage and processing rely on just two polymers, DNA and RNA. Whether their role reflects evolutionary history or fundamental functional constraints is unknown. Using polymerase evolution and design, we show that genetic information can be stored in and recovered from six alternative genetic polymers (XNAs) based on simple nucleic acid architectures not found in nature. We also select XNA aptamers, which bind their targets with high affinity and specificity, demonstrating that beyond heredity, specific XNAs have the capacity for Darwinian evolution and for folding into defined structures. Thus, heredity and evolution, two hallmarks of life, are not limited to DNA and RNA but are likely to be emergent properties of polymers capable of information storage.

The nucleic acids DNA and RNA provide the molecular basis for all life through their unique ability to store and propagate information. To better understand these singular properties and discover relevant parameters for the chemical basis of molecular information encoding, nucleic acid structure has been dissected by systematic variation of nucleobase, sugar and backbone moieties (1-7).

These studies have revealed the profound influence of backbone, sugar and base chemistry on nucleic acid properties and function. Crucially, only a small subset of chemistries allows information transfer through base pairing with DNA or RNA, a prerequisite for crosstalk with extant biology. However, base pairing alone cannot conclusively determine the capacity of a given chemistry to serve as a genetic system, as hybridization need not preserve information content (8). A more thorough examination of candidate genetic polymers' potential for information storage, propagation and evolution requires a system for replication which would allow a systematic exploration of the informational, evolutionary and functional potential of synthetic genetic polymers and open up applications ranging from biotechnology to material science.

In principle, informational polymers can be synthesized and replicated chemically (9) with advances in the non-enzymatic polymerization of mononucleotides (10) and short oligomers

*Correspondence to: ph1@mrc-lmb.cam.ac.uk.
†Current address: Chemistry Department, University of British Columbia, Vancouver, V6T 1Z1, Canada.

Europe PMC Funders Author Manuscripts

(11, 12) enabling model selection experiments (13). Nevertheless, chemical polymerization remains relatively inefficient. On the other hand, enzymatic polymerization has been hindered by the stringent substrate selectivity of polymerases. Despite progress in understanding the determinants of polymerase substrate specificity and in engineering polymerases with expanded substrate spectra (7), most unnatural nucleotide analogues are poor polymerase substrates at full substitution, both as nucleotides for polymer synthesis and as templates for reverse transcription. Notable exceptions are 2′OMe-DNA and TNA. 2′OMe-DNA is present in eukaryotic rRNAs, is well-tolerated by natural reverse transcriptases (RTs) and has been shown to support heredity and evolution at near full substitution (14). TNA allowed polymer synthesis and evolution in a three letter system (15) but only limited reverse transcription (16) .

Here we describe a general strategy to enable enzymatic replication and evolution of a broad range of synthetic genetic polymers based on: 1) a chemical framework (XNA) capable of specific base-pairing with DNA, 2) the engineering of polymerases that can synthesize XNA from a DNA template, as well as 3) the engineering of polymerases that can reverse transcribe XNA back into DNA. We chose six different XNAs in which the canonical ribofuranose ring of DNA and RNA is replaced by five- or six-membered congeners comprising HNA (1,5 anhydrohexitol nucleic acids), CeNA (cyclohexenyl nucleic acids), LNA (2′-O,4′-C-methylene-β-D-ribonucleic acids; locked nucleic acids), ANA (arabinonucleic acids), FANA (2′-fluoro-arabinonucleic acid) and TNA (α-L-threofuranosyl nucleic acids) (4-6, 17, 18).

To enable discovery of polymerases capable of processive XNA synthesis, we developed a selection strategy called compartmentalized self-tagging (CST) (Fig. S1). CST selections were performed on libraries of TgoT, a variant of the replicative polymerase of *Thermococcus gorgonarius* comprising mutations to the uracil-stalling (V93Q) (19, 20) and 3′-5′ exonuclease (D141A, E143A) functions, as well as a “;Therminator” mutation (A485L) (21). TgoT libraries were created from both random and phylogenetic diversity targeted to 22 short sequence motifs within a 10 Å-shell of the nascent strand (Fig. S2).

CST selections with HNA and CeNA nucleotide triphosphates (hNTPs / ceNTPs) yielded rapid adaptation towards HNA and CeNA polymerase activity. One polymerase, Pol6G12 (TgoT: V589A, E609K, I610M, K659Q, E664Q, Q665P, R668K, D669Q, K671H, K674R, T676R, A681S, L704P, E730G) (Fig. 1A), displayed general DNA-templated HNA polymerase activity dependent on the presence of all four hNTPs (Fig. S4) and enabled the synthesis of HNAs long enough to encode meaningful genetic information such as tRNA genes. HNA synthesis was further investigated by mass spectrometry (MS) confirming expected molecular mass, composition and sequence of HNA polymers (Figs. 2C, Fig. S6).

Having established HNA synthesis, we sought to discover a reverse transcriptase for HNA (HNA-RT), capable of synthesizing complementary DNA from an HNA template, to retrieve the genetic information encoded in HNA and enable both analysis and evolution. As no available polymerase displayed this activity, we engineered an HNA-RT *de novo*. Since HNA adopts RNA-like A-form helical conformations (5), we hypothesized that an HNA-RT might be found in the structural neighborhood of an RNA-RT. Starting from TgoT, we used Statistical Correlation Analysis (SCA) (22) of the polB family (Fig. S7) to uncover potential allosteric interaction networks, involved in template recognition. Random mutagenesis and screening by a polymerase activity assay (PAA, Fig. S3) of four SCA “hits” (F405, Y520, I521, L575) in the vicinity of L408 (a residue implicated in RNA RT activity in the related Pfu DNA polymerase (23)) identified a mutant, TgoT: E429G, I521L, K726R (RT521), as a proficient HNA RT (Fig. 2D). Together with Pol6G12, the evolved HNA polymerase,

RT521 enables the transfer of genetic information from DNA to HNA and its retrieval back into DNA (Fig. S11).

Next, we explored if other polymerases derived by CST and SCA might enable synthesis and reverse transcription of other synthetic genetic polymers. Screening identified PolC7 (TgoT: E654Q, E658Q, K659Q, V661A, E664Q, Q665P, D669A, K671Q, T676K, R709K) and PolD4K (L403P, P657T, E658Q, K659H, Y663H, E664K, D669A, K671N, T676I) (Fig. 1) as efficient synthetases for CeNA (C7), LNA (C7), ANA (D4K) and FANA (D4K) (Fig. 3A-C, E, F). Therminator" (9°N *exo*-: A485L) polymerase has previously been shown to support TNA synthesis (16), but TNA RTs were lacking. RT521 proved capable of both efficient TNA synthesis and reverse transcription (Fig. 3D). In addition, RT521 is an efficient RT for both ANA and FANA (Fig. 3B and 3C). Another polymerase variant, RT521K (RT521: A385V, F445L, E664K), was found to enhance CeNA-RT activity and enable reverse transcription of LNA (Figs. 3A, 3E, S8). Together, these engineered polymerases support the synthesis and reverse transcription of six synthetic genetic polymers, and thus enable replication of the information encoded therein (Fig. 3G).

Mutations enabling DNA-templated XNA synthesis were found to cluster at the periphery of the primer-template interaction interface in the polymerase thumb subdomain, >20 Å from the active site (Fig. 1B) and in one case allowed direct XNA-templated XNA replication (FANA, Fig. S9). In contrast, broad XNA-RT activity was mostly effected by a mutation (I521L) in proximity to a catalytic aspartate (D542) and the polymerase active site. Its identification by SCA points to potential allosteric interaction networks involved in template recognition.

As previously observed for TNA (16), non-cognate polymer synthesis can come at a cost of reduced fidelity as polymerase structures are poorly adapted to detect mismatches or aberrant geometry in the non-canonical XNA•DNA (or DNA•XNA) duplexes. We determined aggregate fidelities (as the probability of errors per position) of a full DNA → XNA → DNA replication cycle ranging from 4.3×10^{-3} (CeNA) to 5.3×10^{-2} (LNA), with HNA, CeNA, ANA and FANA superior to LNA and TNA (Table S8, Figs. S11, S12).

Synthesis and reverse transcription establishes heredity (defined as the ability to encode and pass on genetic information) in all six XNAs. We next sought to explore the capacity of such genetic polymers for Darwinian evolution. As a stringent test for evolution and for acquisition of higher order functions such as folding and specific ligand binding, we initiated aptamer selections directly from diverse HNA sequence repertoires. We used a modification of the standard aptamer selection protocol comprising magnetic beads for capture and isolation of all-HNA aptamers against two targets that had been previously used to generate both DNA and RNA aptamers (24, 25): the HIV trans-activating response RNA (TAR) and hen egg lysozyme (HEL).

After eight rounds (R8) of selection using a biotinylated (27 nt) version of the TAR RNA motif (sTAR) as bait, clear consensus motifs emerged (Fig. S13) from which we identified an HNA aptamer (T5–S8-7) that bound specifically to sTAR with a dissociation constant (K_D) between 28-67 nM as determined by Surface Plasmon Resonance (SPR), Bio-Layer Interferometry (BLI) and ELONA titration (Figs. 4C, S14 and Table S6). Other anti-TAR HNA aptamers, from the same selection experiment, displayed similar affinities but distinctive fine specificities with regards to binding TAR loop or bulge regions (Figs. 4A and S14). We initiated selection against HEL from an N_{40} random sequence repertoire and again observed emergence of a consensus motifs after R8 (Fig. S15). We identified specific HEL binders with K_D of 107-141 nM as determined by SPR, BLI and Fluorescence Polarization (FP) (Figs. 4C, S16 and Table S7). Anti-HEL HNA aptamers cross-reacted with

human lysozyme and to a minor degree (<10%) with the highly-positively charged cytochrome C (pI = 9.6), but did not show binding to unrelated proteins such as BSA and streptavidin (Fig. 4B). Fluorescently-labeled HNA aptamers allowed direct detection of surface HEL expression by flow cytometry (FACS) in a transfected cell line demonstrating specificity in a complex biological environment (Fig. 4D).

Our work establishes strategies for the replication and evolution of synthetic genetic polymers not found in nature, providing a route to novel sequence space. The capacity of synthetic polymers for both heredity and evolution also shows that DNA and RNA are not functionally unique as genetic materials. The methodologies developed herein are readily applied to other nucleic acid architectures and have the potential to enable the replication of genetic polymers of increasingly divergent chemistry, structural motifs and physicochemical properties, as shown here by the acid resistance of HNA aptamers (Fig. S17). Thus, aspects of the correlations between chemical structure, evolvability and phenotypic diversity may become amenable to systematic study. Such “synthetic genetics” (26), i.e. the exploration of the informational, structural and catalytic potential of synthetic genetic polymers, should advance our understanding of the parameters of chemical information encoding, and provide a source of ligands, catalysts and nanostructures with tailor-made chemistries for applications in biotechnology and medicine.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by MRC U105178804 (PH, VBP, CC) and by grants from the European Union Framework (FP6-STREP-029092 NEST (PH, VBP, MA, MR, P Herdewijn)), the European Science Foundation (ESF), BBSRC (09-EuroSYNBIO-OP-013) (AT), the European Research Council (ERC-2010-AdG_20100317) (JW), and KU Leuven (GOA/IDO programs)(P Herdewijn). MRC has filed a patent continuation in part (US 2010/018407 A1) and a patent application (WO 2011/135280 A2) on the CST selection system and the polymerases for XNA synthesis and reverse transcription. Polymerases are available for non-commercial purposes from PH upon request subject to MTA.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

One Sentence Summary

The capacity of synthetic polymers for heredity and evolution shows that DNA and RNA are not unique as genetic materials.

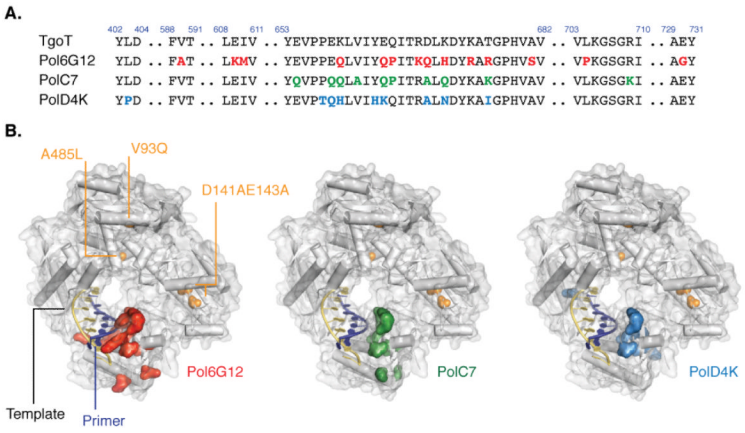

Figure 1. Engineering XNA polymerases

(A) Sequence alignments showing mutations from Tgo consensus in polymerases Pol6G12 (red), PolC7 (green) and PolD4K (blue). (B) Mutations are mapped on the structure of Pfu (PDB: 4AIL). Template is shown in yellow, primer in dark blue. Mutations present in the parent polymerase TgoT are shown in orange.

Europe PMC Funders Author Manuscripts

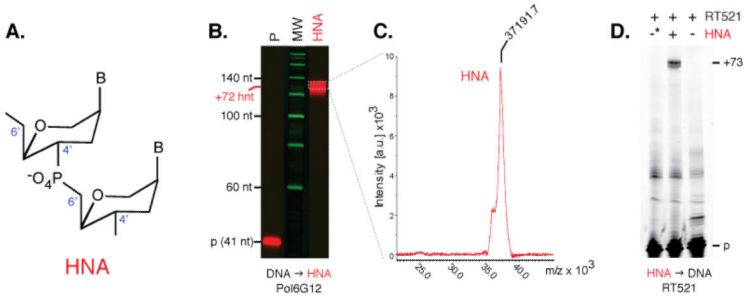

Figure 2. HNA synthesis, mass spectrometry (MS) analysis and reverse transcription

(**A**) Structure of 1,5-anhydrohexitol (HNA) nucleic acids (B: nucleobase). (**B**) Pol6G12 extends the primer (p) incorporating 72 hNTP against template T1 (Table S3) to generate a full-length hybrid molecule of 37,215 Da expected molecular mass (27). MW (ILS 600 molecular weight marker. (**C**) MALDI-ToF MS spectrum of full-length HNA molecule showing a measured HNA mass of 37,190 ± 15 Da (n = 3). (**D**) HNA reverse transcription (DNA synthesis from an HNA template). Polymerase-synthesised HNA (from template YtHNA4; Table S3) is used as template by RT521 for HNA-RT (-* denotes a no HNA synthesis control to rule out template contamination).

Europe PMC Funders Author Manuscripts

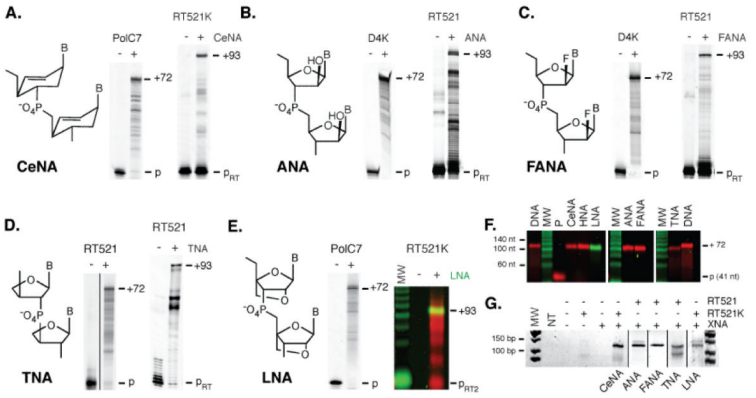

Figure 3. XNA genetic polymers

Structures (B: nucleobase), PAGE of synthesis (+72 xnt) and reverse transcription (+93 nt) of **(A)** CeNA, **(B)** ANA, **(C)** FANA, **(D)** and TNA. **(E)** PAGE of LNA synthesis (primer (41 nt) + 72 lnt) and LNA RT (red) resolved by alkali agarose gel electrophoresis (AAGE). LNA synthesis (green) migrates at its expected size (113 nt) and co-migrates with reverse transcribed DNA (red) synthesized from primer P_{RT2} (20 nt) (Table S3, Fig. S8). **(F)** AAGE of XNA and DNA polymers of identical sequence (PAGE: Fig. S5) (MW: ILS 600 molecular weight markers). **(G)** XNA RT-PCR (MW: NEB Low Molecular weight marker, NT: no template control). Amplification products of expected size (133 bp) are only obtained with both XNA forward synthesis and RT (RT521 or RT521K) (Fig. S12).

Europe PMC Funders Author Manuscripts

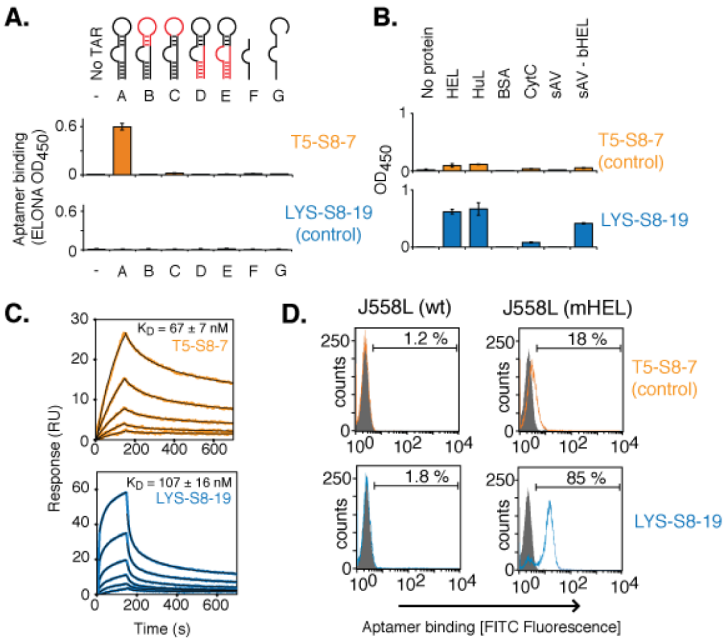

Figure 4. Characterization of HNA aptamers

Anti-TAR aptamer T5-S8-7 (HNA: 6′-AGGTAGTGCTGTTCGTTCATCTCAAATCTAGTTCGCT ATCCAGTTGGC-4′) and anti-HEL aptamer LYS-S8-19 (HNA: 6′-AGGTAGTGCTGTTCGTTTAAATGTGTGTCGTCGTTCGCTATCCAGTTGGC-4′) were characterized by ELONA (27). (**A** and **B**) Aptamer binding specificity against TAR variants (red: sequence randomized but with base-pairing patterns maintained) and different protein antigens (hen egg lysozyme, HEL; human lysozyme, HuL; bovine serum albumin, BSA; cytochrome C, CytC; streptavidin, sAV; and biotinylated-HEL bound to streptavidin, sAV-bHEL). (**C**) Affinity measurements of aptamer binding by SPR. **(D)** FACS analysis of FITC-labeled aptamers binding to plasmacytoma line J558L with and without expression of membrane-bound HEL (mHEL) (27).

Europe PMC Funders Author Manuscripts