REVIEW ARTICLE
Tingjian Chen et al.
From polymerase engineering to semi-synthetic life: artificial expansion of the central dogma
Introduction

Natural nucleic acids, DNA and RNA, have been employed for storing, retrieving, and transmitting genetic information by all living organisms. Both DNA and RNA are composed of four kinds of nucleotide monomers that are linked by 3'-5' phosphodiester linkages, and each typical nucleotide unit is composed of nitrogenous base (nucleobase), pentose, and phosphodiester moieties. The Watson–Crick pairing between the nucleobases lays the foundation of storage, replication, retrieval, and transmission of genetic information via nucleic acids and also confers highly programmable structures, properties, and functions to nucleic acids, which lead to broad application of nucleic acids in biotechnology and biomedicine. However, the limited number of natural nucleobases significantly constrains sequence space; genetic information capacity, and chemical diversity of DNA and RNA. Deoxyribose/ribose-phosphate backbones also restrict properties of DNA and RNA and possess relatively poor biological and chemical stability, which severely limits the functions and practical applications of DNA and RNA. Extensive efforts on the design and synthesis of DNA and RNA analogs with unnatural moieties have been made to address these problems in recent years, which greatly expanded the scope and application of genetic materials, and led to the emergence of a thriving field that has been named xenobiology. To support the efficient synthesis, replication, and evolution of unnatural nucleic acids, which are crucial for making full use of them, unnatural nucleic acid polymerases are essential. Unfortunately, for many of the exotic unnatural nucleic acids, especially those with modified sugar-phosphate backbones, natural DNA polymerases (DNAPs) and RNA polymerases (RNAPs) are unable to synthesize them efficiently. Approaches of protein engineering have thus been applied for tailoring polymerases for the efficient synthesis of various unnatural nucleic acids, and a number of strategies specifically effective for screening or selecting polymerase mutants have also been developed to facilitate these efforts. Numerous polymerase mutants with varied activities of synthesizing, reverse transcribing, and amplifying different unnatural nucleic acids have been obtained and broadly applied in the production and evolution of functional unnatural nucleic acids, including aptamers, catalysts, and nanomaterials. Other than in vitro application of unnatural nucleic acids, efforts have also been made to introduce unnatural nucleic acid components into living organisms, including creation of semi-synthetic organisms (SSOs) that are able to replicate, transcribe and translate unnatural base pairs (UBPs), which greatly expanded the genetic alphabet in vivo.
Development of unnatural nucleic acids

Modifications or alterations have been introduced into different moieties of nucleic acids via chemical synthesis of the structural units, which leads to broad expansion of structures, properties, functions, and applications of nucleic acids. The introduction of chemical modifications into natural nucleobases adds new functionalities to DNA and RNA, while the introduction of UBP pairing orthogonally to natural base pairs increases the genetic information capacity of DNA and RNA. Modification or replacement of sugar-phosphate backbones with unnatural components usually leads to a significant change of the overall properties of nucleic acids, including an increase of chemical or biological stabilities and a change of electronegativity. Combination of modifications on different moieties led to the successful production of nucleic acid analogs with combined properties and functions added.

Unnatural nucleic acids with modified nucleobases

Chemical modifications of nucleobases are achieved via direct replacement of atoms in the nucleobases or attachment of functional groups onto the atoms in the nucleobases (Fig. 1). Atoms C5 in pyrimidines and N7 in purines are usually picked as the atoms to be modified, due to their positions in the major groove of the DNA duplex and the correspondingly less steric hindrance for the modifications. A broad range of modifications, including hydrophobic groups, side chains of amino acids, fluorophores, and reactive groups for later labelling or coupling, have been introduced into nucleobases to confer desired functionalities, properties, or reactivities to DNA or RNA molecules. The recognition of nucleoside triphosphates containing modified nucleobases by polymerases is crucial for the practical application of these modified nucleobases, and many of the nucleoside triphosphates containing C5-modified pyrimidines or N7-modified purines have proven to be efficient substrates for polymerases, and broadly used in the field of biotechnology. Recently, Herdewijin and co-workers demonstrated efficient polymerase replication of DNA when the natural deoxyribonucleoside triphosphates (dNTPs) were extensively replaced with up to four nucleoside triphosphates containing various non-canonical nucleosides, including 7-deaza-2′-deoxyadenosine, 7-deaza-2′-deoxyguanosine, 5-fluoro-2′-deoxycytidine, and 5-chloro-2′-deoxyuridine. They termed the DNA with all four nucleobases replaced by 5-substituted pyrimidines and 7-deazapurines DZA, and demonstrated its function for carrying genetic information in living cells and potential use in biotechnology and synthetic biology.

Unnatural base pairs (UBPs)

Other than modified nucleobases, various UBPs that can pair orthogonally to natural base pairs have been developed in the past few decades, which add not only novel properties and functions, but also a significant increase of the genetic information capacity to DNA and RNA. Key factors to be considered during the design of UBPs include shape complementarity and forces responsible for base pairing. The most predominant UBPs that can be efficiently replicated similar to natural base pairs have been developed by the Benner, Romesberg, and Hirao groups (Fig. 2). The pairing of UBPs developed by Benner’s group, including isoG–isoC, V–J, K–X, Z–P, and S–B, is mainly based on hydrogen-bonding with rearranged patterns, while the pairing of UBPs developed by Romesberg’s group, including MMO2–5SICS, NaM–5SICS, NaM–TPT3, PTMO–TPT3, CNMO–TPT3 and NaM–TAT1, is based on hydrophobic packing and stacking forces. Hirao and co-workers constructed a series of UBP pairing mainly based on hydrophobic interactions, such as Ds–Pa and Ds–Px, as well as several UBP pairing based on...
hydrogen bonding with designed steric exclusion, such as x–y, s–y and Q–Pa.39–43

Unnatural nucleic acids with an altered sugar backbone

Modifications or alterations of sugar backbones have also been extensively introduced into nucleic acids, which usually leads to a dramatic change of the overall properties of nucleic acids, including a change of melting temperature, an increase of chemical stability, resistance against nucleases, and a decrease or abolition of electronegativity34 (Fig. 1). Modification on the 2′-position of the ribose or the deoxyribose is one of the most explored modifications, and the hydrogen or hydroxyl group has been replaced with various atoms or groups, including 2′-fluoro (2′-F), 2′-azido (2′-Az), 2′-amino (2′-Am), and 2′-methoxy (2′-OMe). Although only one atom or group in the pentose is substituted, some of these modifications lead to significant changes in the properties of DNA or RNA. For example, 2′-F and 2′-OMe modifications greatly increase the melting temperatures, the duplex stabilities, and nucleic resistances of the nucleic acids.45 Replacement of the oxygen in the sugar ring with sulfur leads to the production of 4′-thiol-modified DNA and RNA, which have been prepared and investigated by Matsuda and co-workers.46,47 They demonstrated that 4′-thiol-modified RNA can be transcribed from a DNA template with 4′-thiol-CTP and 4′-thiol-UTP by T7 RNAP, and 4′-thiol-modified DNA can be efficiently amplified by KOD dash DNAP under appropriate conditions and transcribed by T7 RNAP in vitro and by mammalian RNA polymerases in mammalian cells.48,49 Other than simple substitution of atoms or groups in the pentose, replacement of the entire deoxyribose or ribose with other sugars has also been explored to construct unnatural nucleic acids with varied structures and properties (Fig. 1). For example, replacement of the pentose with arabinose or 2′-fluoro-modified arabinose leads to the production of arabinose nucleic acid (ANA) and 2′-fluoro-arabinose nucleic acid (FANA). Although the structures of ANA and FANA are similar, the thermal stability of FANA is very different from that of ANA, due to the smaller size and bigger electronegativity of 2′-F, which favors the formation of a pseudo hydrogen bond between 2′-F and purine H8.50 Locked nucleic acid (LNA) was constructed by “locking” the conformation of the ribofuranose with a 2′-O and 4′-C methylene bridge, which increases the stability of the duplex.51 In α-L-threofuranosyl nucleic acid (TNA), the sugar backbone is built with a tetrose instead of a pentose, which does not affect the capability of TNA to hybridize with DNA, RNA, or another strand of TNA, as well as to form G-quadruplexes.52,53 Hexitol nucleic acid (HNA) and cyclohexenyl nucleic acid (CeNA) harbor six-membered rings instead of five-membered rings in their backbones, and retain the ability to hybridize with complementary strands of DNA or RNA.54,55 The sugar rings of different sugar-modified nucleic acids possess varied pucker conformations, which are closely related with the overall helical structures and stabilities of the nucleic acid duplexes.56–58 Replacement of the entire sugar-phosphate backbone of DNA with a peptide backbone produces peptide nucleic acids (PNAs), which can hybridize with DNA or RNA with enhanced melting temperatures, and are highly resistant to nuclease degradation.59,60

Unnatural nucleic acids with an altered phosphate moiety

The phosphate moiety links the structural units of DNA and RNA together to form genetic polymers, and is the major contributor to the hydrophilicity and charge of DNA and RNA molecules. Accordingly, chemical modification or substitution of the phosphate moiety in principle will lead to a dramatic change of the physicochemical properties of nucleic acids, and thus has also been extensively explored (Fig. 1). For example, the most straightforward modifications of the phosphate have been done by replacing the non-bridging oxygen with another atom or group, such as a sulfur or a borane, which led to the production of phosphorothioate (PS) DNA or boranophosphate (PB) DNA.61,62 These modifications introduce chirality into the phosphate moieties of the nucleoside triphosphates and the nucleic acid backbones, and the polymerase incorporation efficiencies of the nucleoside triphosphates are affected by their configurations.63 For example, polymerases, including Escherichia coli (E. coli) DNAP, E. coli RNAP, and T7 RNAP, can polymerize the Sp diastereomers of nucleoside 5′-(1-thiotriphosphates), and inversion of configuration during the polymerization leads to the production of diesters with an Rp configuration.63 Recently, phosphorothioate modification of mRNA in proper patterns was reported to dramatically increase the efficiency of protein synthesis in an E. coli cell-free translation system.64 Holliger and co-workers reported the synthesis of backbone-uncharged phosphonate nucleic acid (phNA) by replacing the non-bridging oxygen with an alkyl group.65 In some other cases, other atoms in the phosphate backbone were replaced to construct nucleic acid analogs. For example, N3′-P5′ phosphoramidate (PN) DNA was synthesized via replacement of the bridging 3′ oxygen with an amino group.66 Specially, triazole DNA was constructed by linking the structural units via click chemistry.66

Unnatural nucleic acids with combination of modifications on different moieties

Introduction of two or more modifications into different moieties of one nucleic acid at the same time is supposed to introduce
Polymerase engineering for the synthesis, reverse transcription, and amplification of unnatural nucleic acids

Natural nucleic acids are efficiently replicated, transcribed, or reverse transcribed by different natural polymerases, including DNAPs, RNAPs, reverse transcriptases, or RNA-dependent RNAPs, which enables efficient replication and transmission of genetic information among these nucleic acids, and also lays the foundation for broad application of DNA and RNA. Although many unnatural nucleic acids have been designed and synthesized, natural polymerases are unable to synthesize a great number of them efficiently, which has severely limited their potential contribution to the expansion of the central dogma, as well as their immediate use in practical applications. To address this problem, polymerases have to be engineered to be able to recognize and polymerize the building blocks of unnatural nucleic acids efficiently. Among the efforts for engineering polymerases for an expanded substrate spectrum, directed evolution with various strategies has proven to be very effective, and a lot of polymerase mutants with excellent unnatural activities were obtained via directed polymerase evolution. Directed evolution mimics the natural evolution process in the laboratory, and yet the mutation rate of the target biomolecule is greatly increased and the selection pressure is artificially set to direct the evolution for the desired properties and functions. A typical directed polymerase evolution process includes two major steps: diversification of the polymerase gene and screening or selection for the polymerase mutants with desired activities.

Methods for the creation of polymerase libraries

Many strategies for artificial gene diversification have been developed since the early efforts of protein evolution (Table 1), and some of them have been successfully applied, either independently or in combination, in the creation of DNAP or RNAP libraries to be selected for enhanced activities against unnatural substrates.

Error-prone PCR introduces random mutations into the genes of target proteins, and is one of the most frequently employed methods for creating protein libraries. In a typical error-prone PCR reaction, random mutagenesis is carried out simply by increasing the mutation rate of the gene during PCR amplification, which is achieved by using polymerases of low fidelity, unbalancing the concentrations of the dNTPs, using analogs of some of the dNTPs, increasing PCR cycles, enhancing the concentration of magnesium ions, and adding manganese ions. It is very important to control the mutation rate of the target gene in an error-prone PCR experiment, since the library size is restricted and only able to cover a small portion of all possible mutants due to limited transformation efficiency and screening or selection throughput, and an excessively high mutation rate usually leads to a rapid loss of protein activity during directed evolution. Error-prone PCR is especially useful for protein library construction when the structural information of the target protein is not available or sufficient to predict which exact residues are crucial for the desired activity and should be randomized or directly mutated to be certain amino acids. Also, beneficial mutations that are far away from the active site of a protein are frequently revealed from a completely randomized library. Error-prone PCR has already been successfully used for constructing libraries of many DNAPs, including Tth DNAP, Klentaq DNAP, Stoffel fragment (SF) of Taq DNAP, full-length Taq DNAP and phi29 DNAP.

The DNA shuffling technique mimics natural hybridization or recombination processes for rapid molecular breeding of proteins by recombining the genes of homologous proteins in vitro. In traditional DNA shuffling experiments, the genes of two or more homologous proteins, or mutants of the same protein, are first segmented by DNase I, and then assembled to be recombined full-length genes by PCR to generate the libraries. Desired protein mutants with beneficial mutations accumulated and deleterious mutations reduced are then screened or selected from the DNA shuffling libraries. DNA shuffling of a set of homologous genes from different species is called family shuffling. The application of family shuffling on genes with relatively low homology may result in less efficient recombination, which could be improved by using restriction endonucleases, instead of DNase I, for DNA fragmentation. In another method called nucleotide exchange and excision technology (NEXT), DNA fragmentation was achieved by dosing uridine triphosphate (dUTP) into the PCR reaction of the target DNA, excising uracil bases in the PCR product with uracil-DNA-glycosylase, and then cleaving the DNA at the positions where uracil bases were excised with piperidine. In this method, the size distribution of the DNA fragments could be easily controlled by the concentration of dosed dUTP. Rather than recombining the target genes by DNA fragmentation and reassembly, Arnold and co-workers developed another strategy called the staggered extension process (STEP) for DNA shuffling. In this strategy, the target genes were mixed and subjected to the PCR reaction with a shortened extension time in each PCR cycle, which led to frequent template switching for primer elongation.
before the elongation reached full-length of the genes every time. In another method for generating recombination libraries, random-priming recombination (RPR), random-priming synthesis is used to generate short gene fragments containing low levels of

| Method                                      | Advantages                                      | Disadvantages                                      | Application examples in polymerase evolution | Ref. |
|---------------------------------------------|-------------------------------------------------|---------------------------------------------------|---------------------------------------------|------|
| Error-prone PCR                             | Simple and easy to implement                    | Base bias of mutagenesis                           | Klentaq M1, Klentaq M2, SFM4-3, SFM4-6, SFM4-9, Taq T8, Taq H15, Taq M1, Taq M4, Tth SA7b/Cs12RsEx pol mutants | 71 and 76–81 |
| DNA shuffling with fragmentation by DNase I  | Simple and easy to implement                    | Requirement for high sequence homology of the parental proteins | Taq/Tth/Tfl 5D4, Bst LF/Klentaq v5.9, v7.16 | 83, 83, 93 and 95 |
| Family shuffling                            | Parental proteins can be from different species | Requirement for high sequence homology of the parental proteins | Taq/Tth/Tfl 5D4, Bst LF/Klentaq v5.9, v7.16 | 83, 83, 93 and 95 |
| Nucleotide exchange and excision technology (NExT) | Good controllability of the DNA fragment sizes during fragmentation of the parental sequences | Requirement for high sequence homology of the parental proteins | SFM4-3, SFM4-6, SFM4-9, Taq T8, Taq H15, Taq M1, Taq M4, Pfu DNAP E10 | 77–79 and 80, 88, 89 and 94 |
| Staggered extension process (StEP)          | Simple and easy to implement                    | Requirement for high sequence homology of the parental proteins | SFM4-3, SFM4-6, SFM4-9, Taq T8, Taq H15, Taq M1, Taq M4, Pfu DNAP E10 | 77–79 and 80, 88, 89 and 94 |
| Synthetic shuffling/assembly of designed oligonucleotides (ADO) | Highly combinatorial DNA library Libraries without limits to the length and number of the parental sequences Increased recombination resolution | Requirement for carefully designed oligonucleotides | KlenTaq Mut_ADL, KlenTaq Mut_RT | 90 and 96 |
| Random chimeragenesis on transient templates (RACHITT) | Lower sequence homology of the parental proteins is required Higher recombination frequency | Requirement for synthesized templates | KlenTaq Mut_ADL, KlenTaq Mut_RT | 90 and 96 |
| Incremental truncation for the creation of hybrid enzymes (ITCHY) | No requirement for sequence homology Numbers of parental sequences and fragments for recombination are limited | Requirement for the structural information of the target protein | SFM4-3, SFM4-6, SFM4-9, KlenTaq Mut_ADL, KlenTaq Mut_RT | 69, 77, 88, 96, 97, 103, 106, 107 and 163 |
| Site-saturated mutagenesis (SM)              | Simple and easy to implement                    | Requirement for the structural information of the target protein | — | — |
| Combinatorial active-site saturation test (CAST) | Potential synergistic conformational effects can be taken into account Small but focused high-quality mutant library for each round of evolution | Requirement for the structural information of the target protein | — | 99 |
| Iterative saturation mutagenesis (ISM)       | Consecutive point mutations Controllable mutational bias Controllable fragment distribution of the DNA library | Requirement for the structural information of the target protein | — | 100 |
| Sequence saturation mutagenesis (SeSaM)      | — | Requirement for the structural information of the target protein | Only one randomized site in each mutant | — | 101 |

—: no application example in polymerase evolution yet.
point mutations to be assembled.\textsuperscript{87} Synthetic shuffling, in which degenerate oligonucleotides encoding all the variations in the parental genes are used to assemble the mutants, has been demonstrated to be an effective library creation method for evolving highly chimeric enzymes.\textsuperscript{87,88} In another study, Reetz and co-workers created recombination libraries of proteins by the assembly of designed oligonucleotides (ADO), in which the oligonucleotides for assembly were designed based on sequence information to control the overlapping process and increase the recombination frequency.\textsuperscript{89} Random chimeragenesis on transient information to control the overlapping process and increase the probability of obtaining protein mutants with desired properties is another approach.\textsuperscript{90} It is achieved by designing oligonucleotides for assembly, in which the assembly of designed oligonucleotides (ADO), in which the overlapping information was controlled to generate the hybrid library.\textsuperscript{91} In this method, fragments of homologous genes were incrementally truncated with exonuclease III first, and then the fragments were then 3’ tailed with universal nucleobase using terminal transferase, and elongated to full-length genes.

Methods for the screening or selection of polymerase mutants

To efficiently identify mutants with desired properties from a polymerase library, a well-designed screening or selection method is essential. The key point of constructing a screening or selection method for mutants of a protein is to establish a linkage between the genotype and the phenotype, i.e., the gene and the activity or other functions, of a protein mutant, which can be achieved by many strategies. For example, this linkage can be built by displaying a protein on a carrier of its gene, such as a cell, a phage particle, a magnetic bead, or an mRNA strand.\textsuperscript{1,108–113} Other strategies focus on spatially separating the gene and the expressed protein of each mutant from those of other mutants in a confined space, such as a well of multi-well plates, or a cell-like emulsion compartment, before checking the activity or other properties of each mutant.\textsuperscript{77,102–104,114,115} Based on these original strategies broadly used in protein evolution and taking advantage of the unique nucleotide polymerization activity of polymerases, various methods have been developed for the screening or selection of polymerase libraries\textsuperscript{2–4} (Table 2).

Multi-well plate screening methods for screening polymerase variants were developed by immobilizing a primer/template complex on the bottom surface of the wells, and extending the primer with certain nucleoside triphosphate substrates using cell lysate of each polymerase mutant in each well.\textsuperscript{77,104} The success primer extension led to the incorporation of fluorescent, biotinylated, or digoxigenin (DIG)-labelled nucleotides or the annealing of the extension product with labelled oligonucleotides, which could then be detected by reading the fluorescence or by binding with a DIG antibody or streptavidin-coupled enzyme and assaying the activity of this enzyme. Although single clones of active polymerases can be directly identified with these methods, the throughput is limited, which makes these methods more useful for screening pre-enriched or small focused polymerase libraries.
For example, variants of Stoffel fragment, Taq DNAP and Tgo DNAP, have been identified with these methods from focused libraries or libraries pre-enriched with other high-throughput selection methods, which will be introduced below.

Emulsion or microfluidic system-based compartmentalization technology has been extensively used to develop novel methods for polymerase evolution (Fig. 3). For example, Holliger and co-workers developed a compartmentalized self-replication (CSR) method, in which a water-in-oil emulsion system was employed to confine PCR amplification of the gene of each polymerase mutant by the expressed protein of itself in an emulsion compartment, which led to rapid enrichment of polymerase mutants with good activities (Fig. 3A). Using this system, they successfully evolved mutants of Taq DNAP with enhanced thermostability or resistance to inhibitor heparin. Later, they developed a modified version of CSR, short-patch compartmentalized self-replication (spCSR). In this method, only a short region of the polymerase gene was diversified and amplified during the evolution, which reduced the requirements for catalytic activity and processivity of polymerases in the early stage of evolution, and thus made

### Table 2 Methods for the screening or selection of polymerase mutants

| Method       | Advantages                                      | Disadvantages                                      | Application examples | Ref. |
|--------------|-------------------------------------------------|----------------------------------------------------|----------------------|------|
| Multi-well plate screening | Simple Direct identification of single active mutants | Time consuming Limited screening throughput | SFM4-3 SFM4-6 SFM4-9 Tgo RT-TKK Tgo RT-C8 Tgo Pol6G12 Tgo PolC7 Tgo PolD4K Tgo RTS21 | 77, 102, 104, 106 and 115 |
| CSR          | High throughput Simple Target polymerase needs to replicate the full-length of its own gene | High temperature is usually needed to break the emulsified cells | Taq T8 Taq H15 Tth SA7fCs12RsEx pol mutants Phi29 DNAP Mut Bst v5.9 Bst v7.16 KOD RTX KOD RTX-Ome v6 Pfu DNAP E10 Taq AA40 | 78, 80, 81, 95, 116 and 167 |
| spCSR        | High throughput Target polymerase only needs to replicate a part of its own gene | High temperature is usually needed to break the emulsified cells | Tgo Pol6G12 Tgo PolC7 Tgo PolD4K Tgo RTS21 | 94 and 115 |
| CST          | High throughput Allows the selection for activities towards difficult nucleoside triphosphate substrates and under challenging conditions | High temperature is usually needed to break the emulsified cells Plasmid DNA has to be used as the extension template for the tagging primer | Tgo RT-TKK | 104 |
| CPR          | High throughput Expanded scope of proteins to be evolved Mitigated effect on host fitness | High temperature is usually needed to break the emulsified cells Challenging to design genetic circuits | T7 RNAP CGG-R7-8 T7 RNAP CGG-R12-KIRV | 117 |
| CBL          | High throughput Suitable for evolving various reverse transcription activities | High temperature is usually needed to break the emulsified cells Experiment complexity | Tgo RT-TKK Tgo RT-C8 | 102 |
| Phage display | High throughput Kinds of the nucleic acid template, primer and nucleoside triphosphates for selection can all be well controlled Adjustable selection stringency Rapid reproduction of phage | The target polymerase needs to be actively displayed on phage | SFM4-3 SFM4-6 SFM4-9 SFR1 SFR2 SFR3 Phi29 DNAP | 77, 125 and 129 |
| PACE         | High throughput Rapid reproduction of phage Continuous evolution Minimal researcher intervention Rapid evolutionary cycle | Experiment complexity Expensive facilities Challenging to design genetic circuits | T7 RNAP A6-36.4 | 131 |
| Cell surface display | High throughput Expanded scope of polymerases to be displayed for selection | The target polymerase needs to be actively displayed on cell surface | KF I709E E710G | 106 |
this method suitable for the evolution of challenging activities. A variant of Taq DNAP, AA40, which possessed replication, transcription and reverse transcription activities, as well as an expanded substrate spectrum for 2'-modified nucleoside triphosphates, was successfully evolved with this method. Ellington and co-workers developed a modified version of CSR, high-temperature isothermal compartmentalized self-replication (HTI-CSR), in which the self-replication of the polymerase gene was realized via rolling circle amplification (RCA) instead of PCR. This method was successfully used to evolve a thermostable strand-displacing polymerase mutant from a shuffled library of Bst LF and Klentaq DNAP. They also developed another modified CSR method, reverse transcription-compartmentalized self-replication (RT-CSR), to evolve reverse transcription activity of a DNAP. In the design of this method, to realize self-replication of the polymerase gene, the polymerase mutant had to reverse transcribe several RNA nucleotides in a flank primer, which partially annealed to the polymerase gene, to produce a full-length template that could be PCR amplified with outer primers. A high-fidelity thermostable
reverse transcriptase, which they called reverse transcription xenopolymerase (RTX), was then evolved from KOD DNAP with this method. To expand the CSR method for the evolution of more proteins, a compartmentalized partnered replication (CPR) method was developed by the same group117 (Fig. 3B). In the CPR method, the activity of a partner protein that needed to be evolved was linked to the expression of Taq DNAP, which in turn PCR amplified the gene of the partner protein. This method was successfully applied on the evolution of several proteins, including T7 RNAP mutants for the recognition of orthogonal promoters.118

In CSR or the derivatives of CSR introduced above, full-length or a short region of the polymerase or the partner protein gene needs to be replicated by the polymerase to fulfill the evolution process. However, in some cases, especially when the desired activities are too exotic or challenging to evolve, replication of a gene or part of it during the evolution is unrealistic or hard to be correlated with the desired activities. To address this problem, several other compartmentalization-based strategies for polymerase evolution have been developed. For example, compartmentalized self-tagging (CST) was developed to evolve polymerases for the synthesis of xenobiotic nucleic acids (XNAs)104,119 (Fig. 3C). In this method, the selection of active polymerase mutants did not rely on self-replication in the compartment, but relied on the extension of a short biotinylated primer with unnatural nucleoside triphosphates using the plasmid harboring the polymerase gene as a template. Success extension of the primer resulted in tight binding of the primer and the plasmid, and thus enabled streptavidin bead separation of the active mutants. With this method, TgoT DNAP mutants have been evolved for efficient synthesis and reverse transcription of various XNAs. Recently, Holliger and co-workers developed a compartmentalized bead labelling (CBL) method for the evolution of RNA and XNA reverse transcriptases from a DNAP mutant102 (Fig. 3D). This method employed streptavidin-coated beads to co-display two kinds of oligonucleotides, one of which was responsible for the capture of the plasmid harboring polymerase gene, and another served as the primer for the reverse transcription of an XNA/RNA template. When a polymerase mutant successfully reverse transcribed the XNA/RNA template in a compartment, the reverse transcription product would later trigger a hybridization chain reaction (HCR), resulting in intensive fluorescent labelling of the bead, which then allowed fluorescent-activated bead sorting of the beads carrying plasmids of the active polymerase mutants. Polymerase mutants efficient for the reverse transcription of 2'-Ome-RNA, HNA, 6-ditolyl nucleic acid (AtNA), 2'-methoxethyl-RNA (2'-MOE-RNA), and P-β-S-phosphorothioate 2'-MOE-RNA (PS 2'-MOE-RNA) were obtained using this method.

In recent years, the rapidly developing microfluidic technology has also been employed in the design of compartmentalization-based methods for polymerase evolution. In these methods, the generation of the compartments was more controllable, and the process of sorting for the active polymerase mutants could also be directly integrated into the system. For example, Chaput and co-workers developed microfluidic-based protein evolution methods, such as droplet-based optical polymerase sorting (DrOPS) and fluorescence-activated droplet sorting (FADS)-based methods, and used them for evolving polymerases with expanded function.103,114 In these methods, polymerase mutants were encapsulated in water-in-oil-in-water or water-in-oil droplets generated by microfluidics. Polymerase-catalyzed primer extension led to the removal of a fluorescent quencher DNA annealed to the fluorophore-labelled template by strand displacement. The generated fluorescence was then used as an optical signal for the sorting of active polymerase droplets.

Phage display technology was initially developed for the evolution of small peptides or proteins, including antibodies, with high affinity towards the targets, and later proved to be a powerful tool for developing methods of polymerase evolution3,111,120–124 (Fig. 3E and F). For example, Romesberg and co-workers developed a phage-display-based method for polymerase evolution, in which a polymerase mutant was displayed on one of the p3 proteins of an M13 phage particle, while the primer/template substrate was attached to other p3 proteins.77,125 The substrate attachment was accomplished either by the coupling of an acidic peptide displayed on a p3 protein with a basic peptide conjugated to the primer, or by click reaction of an unnatural amino acid p-azidophenylalanine (pAzaF) displayed on a p3 protein and a cycloalkyne conjugated to the primer. When the primer was extended with unnatural nucleoside triphosphates by the polymerase mutant displayed on the same phage, biotinylated-UTPs were incorporated to the end of the extension product, which allowed subsequent streptavidin bead separation of the active polymerase mutants. Using this method, mutants of SF of Taq DNAP that efficiently synthesize and amplify various 2'-modified nucleic acids have been obtained.77,126–128 Other strategies have also been used to attach the primer/template onto the phage. For example, Delesaul and co-workers co-displayed phi29 DNAP and a modified haloalkane dehalogenase, HaloTag, on M13 phage, which allowed the attachment of a DNA substrate coupled with a haloalkane ligand.129

Other than phage-display-based methods, bacteriophages have also been used to develop other methods for directed protein evolution. For example, Liu and co-workers developed a phage-assisted continuous evolution (PACE) strategy, in which the activity of a protein to be evolved was coupled to the propagation of a bacteriophage, and used it to rapidly evolve a variety of proteins with different traits.130–135 Variants of T7 RNAP with altered promoter specificity were successfully evolved with this method by coupling M13 phage propagation with T7 RNAP-mediated transcription of the phage p3 protein.131

Cell surface display technology has been used for the evolution of numerous proteins for either enhanced affinities against certain targets or increased catalytic activities.112,136–138 Recently, the application of an E. coli cell display system for polymerase evolution was also demonstrated by Schwaneberg and co-workers.106 The Klenow fragment (KF) of E. coli DNAP was displayed on the outer membrane of E. coli cells by fusing with autotransporter proteins, and the polymerase mutant-displaying cells were directly used for screening. The activity of each polymerase mutant was checked by monitoring the fluorescence of a
fluorescent dye binding with double-stranded primer-extension product in multi-well plates. With this method, a KF mutant with enhanced activity against 2′-O-methyl nucleoside triphosphates (2′-OMe-NTPs) was successfully evolved.

Polymersases for the synthesis, reverse transcription, and replication of nucleic acids with unnatural moieties

With the library creation methods and screening or selection strategies described above, various polymerases have been engineered to be efficient for synthesizing, reverse transcribing, and even replicating nucleic acids with unnatural moieties. The wild type or engineered polymerases that demonstrated activities with unnatural substrates are summarized in Table 3, and the distribution of predominant mutations in the structures of representative engineered unnatural nucleic acid polymerases is illustrated in Fig. 4.

Although usual modifications of nucleobases, especially those at the C5 position of pyrimidines and the C7 position of deazapurines, are well tolerated by natural polymerases and broadly used in the labelling and functionalization of DNA and RNA,\textsuperscript{119,140} engineering of polymerases can help further increase the enzymatic incorporation efficiency of the nucleotides with base modifications, and even achieve efficient PCR amplification of DNA extensively modified on nucleobases. For example, using the CSR method, Holliger and co-workers evolved a mutant of Taq DNAP, M1, that had an expanded substrate spectrum, and could perform efficient PCR amplification of DNAs with 7-deaza-dGTP, FITC-12-dATP, Biotin-16-dUTP or 2S-dNTPs replacing the corresponding dNTP(s).\textsuperscript{79} In another study, they applied the spCSR method on the evolution of a family B DNAP, Pfu, and successfully obtained a mutant, E10, which could PCR amplify a DNA fragment up to 1 kb with dCTP completely substituted by Cy3- or Cy5-modified dCTP.\textsuperscript{94} Fujita and co-workers reported the enzymatic synthesis of DNA containing high-density amphiphilic functionalities attached to the nucleobases with 7-substituted 7-deazapurine nucleoside triphosphates, dG\textsuperscript{am}TP and dA\textsuperscript{am}TP, and 5-substituted pyrimidine nucleoside triphosphates, dT\textsuperscript{am}TP and dC\textsuperscript{am}TP, using KOD Dash DNAP (KOD XL DNAP).\textsuperscript{141} Efficient PCR amplification of a 500-bp DNA fragment with the mixture of these nucleobase-modified nucleoside triphosphates and natural dNTPs using the same polymerase was also demonstrated. Later, Hoshino and co-workers demonstrated that a mutant of KOD DNAP, KOD exo \textsuperscript{A485L}, could synthesize longer DNA products containing nucleobases modified with these amphiphilic functionalities faithfully and more efficiently.\textsuperscript{142} T7 RNAP has also been proven to be efficient for incorporating nucleoside triphosphates with various modified nucleobases, including N1-methylpsuedouridine (m1\textsuperscript{A}) triphosphate, which allows the \textit{in vitro} transcription of mRNA vaccines with modified bases against various diseases, such as COVID-19.\textsuperscript{143,145}

UBPs for the expansion of genetic alphabet have been developed by Benner’s group have been demonstrated with various DNAPs and RNAPs or their mutants. For example, DNA containing the isoG-isoC pair was successfully PCR amplified with a truncated mutant of Tq DNA, TtTaq.\textsuperscript{146} \textit{In vitro} replication of the K-X pair was also carried out with DNAP I from \textit{E. coli}.\textsuperscript{147} Although Taq DNAP can replicate the Z-P pairs,\textsuperscript{148} for enhanced replication efficiency of the Z-P pair, directed evolution of Taq DNAP was carried out with CSR method, in which oligonucleotides containing multiple P nucleobases were used as the primers for self-replication.\textsuperscript{149} The evolved Taq DNAP mutants Taq (N580S, L628V, and E832V) and Taq (M444V, P527A, D551E, and E832V) demonstrated a much less pause when incorporating DZTP against P nucleobases in a template. T7 RNAP mutant FAL has been shown to be able to efficiently transcribe DNA containing both the Z-P and S-B pairs (Hachimoji DNA), resulting in the production of RNA containing P, Z, B and S nucleobases (Hachimoji RNA).\textsuperscript{148,149} The \textit{in vitro} replication of some of the representative hydrophobic UBPs developed by Ronesberg’s group, such as MMOO–5SICS, NaM–5SICS and NaM–TPT3, has been shown to be efficient with various family A or B DNAPs, including Klenow fragment of \textit{E. col}i DNAP I, Taq DNAP, Deep Vent DNAP, Phusion DNAP, and OneTaq DNAP (a mixture of Deep Vent DNAP and Taq DNAP).\textsuperscript{27–29,31} Directed evolution of polymerases has also proven effective to increase their replication performance for hydrophobic UBPs. For example, a mutant of SF of Taq DNAP, P2, which could synthesize DNA containing PICS self-pair more efficiently than wild type SF, was successfully obtained by directed evolution using the phage-display-based selection method.\textsuperscript{150} In another approach, with the CSR method, Holliger and co-workers evolved a Taq/Tth/Tfl DNAP mutant, SD4, for the ability of forming and extending other self-pairs of hydrophobic nucleobase analogs, including 5NI and 5NC.\textsuperscript{95} The transcription of some of these UBPs has been demonstrated with several well-studied RNAPs, including T7 RNAP and eukaryotic RNAP II.\textsuperscript{30,151} Recent \textit{in vivo} experiments suggested that \textit{E. coli} RNAP could also transcribe DNA containing some of these UBPs.\textsuperscript{152} Recently, reverse transcription of RNA containing TPT3 or NaM has been investigated with several reverse transcriptases or DNAP mutants, including avian myeloblastosis virus (AMV) reverse transcriptase, Moloney murine leukemia virus (MMLV) reverse transcriptase, SuperScript II reverse transcriptase, SuperScript III reverse transcriptase, SuperScript IV reverse transcriptase, and an engineered Taq DNAP with reverse transcription activity, Volcano2G (V2G).\textsuperscript{153,154} It was found that the UBP reverse transcription efficiencies of different reverse transcriptases were sharply different. The replication or transcription efficiencies of different reverse transcriptases or DNAP mutants, including avian myeloblastosis virus (AMV) reverse transcriptase, Moloney murine leukemia virus (MMLV) reverse transcriptase, SuperScript II reverse transcriptase, SuperScript III reverse transcriptase, SuperScript IV reverse transcriptase, and an engineered Taq DNAP with reverse transcription activity, Volcano2G (V2G).\textsuperscript{153,154} It was found that the UBP reverse transcription efficiencies of different reverse transcriptases were sharply different. For UBPs developed by Hirao’s group, it has been shown that UBPs Q-Pa, s-z, Ds-Pa and Ds–Pa could be recognized by the Klenow fragment of \textit{E. coli} DNAP I.\textsuperscript{156,157,158} and remarkably, Ds–Px, Ds–Pn and Ds–Pz pairs could be PCR amplified by Deep Vent DNAP efficiently and faithfully.\textsuperscript{157,158} PCR amplification of DNA containing the Ds–Pa pair has also been carried out with Vent DNAP.\textsuperscript{156} It has been shown that the x–y, s–y, s–z, s–Pa, Dss–Pa, Dss–Px and Ds–Pz pairs could be transcribed by T7 RNAP or its mutant VERS-M5,\textsuperscript{159,160} and the Q–Pa pair could be reverse transcribed by AMV reverse transcriptase.\textsuperscript{43}
Table 3  Summary of natural and engineered polymerases for the synthesis/transcription, reverse transcription or amplification of nucleic acids with unnatural moieties

| Polymerase Mutation sites | Activities | Ref. |
|---------------------------|------------|-----|
| DNAP I from E. coli       | Incorporation of K–X | 147 |
| KF of DNAP I from E. col  | Incorporation of MMO2–5SICS, NaM–5SICS, NaM–TPT3 | 27, 28, 31, 43, 153 and 156 |
| Taq                      | Incorporation of NaM–5SICS, NaM–TPT3 and Z–P | 29, 31 and 148 |
| TITaq                    | Incorporation of isoG–isoC | 146 |
| Taq DNAP mutant M444V, P527A, D551E, E832V | Incorporation of Z–P | 105 |
| Taq DNAP mutant N580S, L628V, E832V | Incorporation of Z–P | 105 |
| Taq/Tth/Tfl 5D4 V62L, Y78H, T88S, P114Q, P264S, E303V, G389V, E424G, E432G, E602G, A606V, I614M, M761T, M77ST | Incorporation of 5NI and 5NIC | 93 |
| Taq M1 G84A, D144G, K314R, E520G, F598L, A608V, E742G | PCR with 7-deaza-dGTP, FITC-12-dATP, Biotin-16-dNTPs | 79 |
| SF P2 F598I, I614F, Q489H | Incorporation of PICS–PICS | 150 |
| OneTaq DNAP V62I, Y78H, T88S, P114Q, P264S, E303V, G389V, E424G, E432G, E602G, A606V, I614M, M761T, M775T | Incorporation of NaM–5SICS, NaM–TPT3, x–y, s–y, s–z, s–Pa, Ds–Pa and Ds–Px | 29, 157 and 158 |
| Vent DNAP V93Q, D141A, E143A, V337I, E399D, N400D, R407I, Y546H | Incorporation of Ds–Pa and Ds–Px | 36 |
| Phusion DNAP V93Q, D141A, E143A, V337I, E399D, N400D, R407I, Y546H | Incorporation of Cy3- or Cy5-modified dCTP | 94 |
| Pfu E10 F Y639F, H784A, S430P, N433T, S633P, F849I, F880Y | Incorporation of m1^P triphosphate | 141, 142, 30, 34, 36, 39, 40, 145, 154–156, 160 and 238 |
| KOD Dash DNAP F Y639F, H784A, S430P, N433T, S633P, F849I, F880Y | Incorporation of dN^{32}TPs | 141, 142, 30, 34, 36, 39, 40, 145, 154–156, 160 and 238 |
| KOD DNAP mutant D141A, E143A, A185L | Incorporation of dN^{32}TPs | 141, 142, 30, 34, 36, 39, 40, 145, 154–156, 160 and 238 |
| T7 RNAP F Y639F | Transcription of MMO2–5SICS, NaM–5SICS, NaM–TPT3, PTMO–TPT3, CNMO–TPT3, x–y, s–y, s–z, s–Pa, Ds–Pa, Dss–Pa and Ds–Px | 159 |
| T7 RNAP F–M5 Y639F, S430P, N433T, S633P, F849I, F880Y | Transcription of Ds–Pa and Ds-modified Pa | 159 |
| T7 RNAP FA–M5 Y639F, G784A, S430P, N433T, S633P, F849I, F880Y | Transcription of Ds–Pa and Ds-modified Pa | 159 |
| T7 RNAP VRS–M5 G542V, H772R, H784S, S430P, N433T, S633P, F849I, F880Y | Transcription of Ds–Pa and Ds-modified Pa | 159 |
| T7 RNAP FAL G784A, H784A, S430P, N433T, S633P, F849I, F880Y | Transcription of Z–P and S–B | 149 |
| T7 RNAP FAL–M5 Y639F | Transcription of Z–P and S–B | 149 |
| T7 RNAP F AL–M5 Y639F, H784A, S430P, N433T, S633P, F849I, F880Y | Transcription of Z–P and S–B | 149 |
| T7 RNAP F AL–M5 Y639F, H784A, S430P, N433T, S633P, F849I, F880Y | Transcription of Z–P and S–B | 149 |
| T7 RNAP F AL–M5 Y639F, H784A, S430P, N433T, S633P, F849I, F880Y | Transcription of Z–P and S–B | 149 |
| T7 RNAP F AL–M5 Y639F, H784A, S430P, N433T, S633P, F849I, F880Y | Transcription of Z–P and S–B | 149 |
| T7 RNAP F AL–M5 Y639F, H784A, S430P, N433T, S633P, F849I, F880Y | Transcription of Z–P and S–B | 149 |
| AMV reverse transcriptase V518A, N583S, D655N, E681K, E742Q, M747R | Reverse transcription of NaM–TPT3 and Q–Pa | 43, 153 and 154 |
| MMLV reverse transcriptase V518A, N583S, D655N, E681K, E742Q, M747R | Reverse transcription of NaM–TPT3 | 153 |
| SuperScript II reverse transcriptase V518A, N583S, D655N, E681K, E742Q, M747R | Reverse transcription of NaM–TPT3 | 153 |
| SuperScript III reverse transcriptase V518A, N583S, D655N, E681K, E742Q, M747R | Reverse transcription of NaM–TPT3 | 153 |
| SuperScript IV reverse transcriptase V518A, N583S, D655N, E681K, E742Q, M747R | Reverse transcription of NaM–TPT3 | 153 and 154 |
| Taq Volcano2G I614E, E615G, V318A, N583S, D655N, E681K, E742Q, M747R | Reverse transcription of NaM–TPT3 | 153 and 154 |
| Bst DNAP I614E, E615G, V318A, N583S, D655N, E681K, E742Q, M747R | Reverse transcription of FANA and TNA | 143, 153 and 154 |
| Deep Vent DNAP I614E, E615G, V318A, N583S, D655N, E681K, E742Q, M747R | Reverse transcription of FANA and TNA | 143, 153 and 154 |
| Deep Vent-RI I614E, E615G, V318A, N583S, D655N, E681K, E742Q, M747R | Reverse transcription of FANA and TNA | 143, 153 and 154 |
| Tgo DNAP I614E, E615G, V318A, N583S, D655N, E681K, E742Q, M747R | Reverse transcription of FANA and TNA | 143, 153 and 154 |
| Tgo-RI I614E, E615G, V318A, N583S, D655N, E681K, E742Q, M747R | Reverse transcription of FANA and TNA | 143, 153 and 154 |
| Tgo Pol6G12 I614E, E615G, V318A, N583S, D655N, E681K, E742Q, M747R | Reverse transcription of FANA and TNA | 143, 153 and 154 |
| Tgo RT521 I614E, E615G, V318A, N583S, D655N, E681K, E742Q, M747R | Reverse transcription of FANA and TNA | 143, 153 and 154 |
| Tgo-6G12-1521L I614E, E615G, V318A, N583S, D655N, E681K, E742Q, M747R | Reverse transcription of FANA and TNA | 143, 153 and 154 |
### Table 3 (continued)

| Polymerase      | Mutation sites                          | Activities                                                                 | Ref. |
|-----------------|-----------------------------------------|-----------------------------------------------------------------------------|------|
| Tgo RT521K      | RT521I: A385V, F445L, E664K             | Reverse transcription of HNA, ANA, FANA, TNA and tPhoNA                     | 104  |
| Tgo RT-TKK      | RT521K: I114T, S383K, N735K             | Reverse transcription of LNA and CeNA                                       | 102  |
| Tgo RT-C8       | RT-TKK: F493V, Y496N, Y497L, Y499A, A500Q, K501H | Reverse transcription of 2'-OMe-RNA, AtNA                                    | 102  |
| Tgo PolC7       | TgoT: E653Q, E658Q, K659Q, V661A, E664Q, G663P, D669A, K671Q, T676K, R709K | Synthesis of CeNA and LNA                                                   | 104  |
| Tgo PolD4K      | TgoT: L403P, P657T, E658Q, K659H, Y663H, E664K, D669K, K716N, T767I | Synthesis of FANA, ANA, TNA, HNA, PMT and RNA                               | 104  and 177 |
| Tgo QGLK        | V93Q, D141A, E143A, Y409G, A485L, E664K | Synthesis of FANA, FANA, ANA and HNA                                        | 177  |
| Tgo EPFLH       | V93Q, D141A, E143A, H147E, L403P, L408F, A485L, I521L, E664H | Synthesis of RNA, FANA, ANA and HNA                                        | 69 and 177 |
| KOD DNAP        |                                          | Reverse transcription of RNA and DNA                                         | 116  |
| KOD Dash DNAP   |                                          | Reverse transcription of 2'-OMe-RNA                                         | 167  |
| KOD DLNKK       | N210D, Y409G, A485L, D614N, E664K       | Reverse transcription of tPhoNA                                             | 69   |
| KOD DLK         | N210D, A485L, E664K                     | Synthesis of FANA, ANA, HNA and TNA                                        | 177  |
| KOD Ri          | D141A, E143A, A485R, E664I              | Synthesis of TNA                                                            | 98   |
| KOD RS          | D141A, E143A, A485R, N491S              | Synthesis of TNA                                                            | 174  |
| KOD QS          | D141A, E143A, L489Q, N491S              | Synthesis of TNA                                                            | 174  |
| KOD RSGA        | D141A, E143A, A485R, N491S, R606G, T723A| Synthesis of RNA, FANA, ANA, HNA, TNA, C5-modified TNA, RNA and PMT          | 68, 175 and 177 |
| KOD RTX         | F38L, R97M, K118I, M137L, R381H, Y384H, V389I, K466R, Y493L, T514I, T521L, F587L, E664K, G711V, N735K, W768R | Reverse transcription of RNA and DNA                                         | 116  |
| KOD RTX-Ome v6  | RTX: A40V, E251K, S340P, G350V, V353L, H381R, K384Y, W387R, D388Y, K384Y, F387L, K486A, K486A, E486A, A485R, I485R, E521L, E664K | Reverse transcription of 2'-OMe-RNA                                         | 167  |
| KOD RT521K      | V93E, D141A, E143A, A485L, I521L, E664K | Reverse transcription of tPhoNA                                             | 69   |
| 9'N DNAP         | D141A, E143A, A485L                     | Synthesis of FANA, ANA, HNA and TNA                                        | 177  |
| 9'N-Thermatator  | D141A, E143A, A485R, E664I              | Synthesis of TNA                                                            | 172 and 173 |
| 9'N-YRI         | D141A, E143A, A485R, N491S              | Synthesis of TNA                                                            | 174  |
| 9'N-NVA         | D141A, E143A, Y409N, D432G, A485V, V636A, E664A | Synthesis of TNA                                                            | 174  |
| Phi29 DNAP mutant | D12A                                    | Synthesis of HNA, FANA and 2'-F-DNA                                         | 180  |
| Tgo PGV2        | RT521I: D455P, K487G, R606V, R613V       | Synthesis of phNA                                                           | 64   |
| DNAP from E. coli |                                        | Polymerization of the Sp diastereomers of nucleoside 5’-(1-thiotriphosphates) | 63   |
| T7 RNAP         |                                          | PolymORIZATION of the Sp diastereomers of nucleoside 5’-(1-thiotriphosphates) | 63   |
| T7 RNAP mutant  | Y639F                                   | Transcription of RNA from 4’-thiol-modified DNA from DNA                    | 46, 63 and 194 |
| T7 RNAP mutant  | Y639F, H784A                            | Transcription of 2’-F, 2’-Am and 2’-F-EdU-modified RNA                       | 48, 49, 67, 168 and 169 |
| T7 RNAP mutant  | Y639W, V698A                            | Transcription of 2’-OMe and 2’-Az-modified RNA                              | 170  |
| T7 RNAP mutant  | Y639W, V698A                            | Transcription of 2’-OMe-modified RNA                                        | 107  |
| T7 RNAP mutant  | Y639W, E593G, V685A                     | Transcription of 2’-OMe-RNA                                                | 171  |
| T7 RNAP mutant  | RGVG: S430P, N433T, S633P, F849I, F880Y | Transcription of 2’-OMe-RNA                                                | 171  |
| T7 RNAP mutant  | RGVG: P266L, S430P, N433T, S633P, F849I, F880Y | Transcription of RNA from 4’-thiol-modified DNA                           | 48 and 49 |
| T7 RNAP mutant  | RGVG-M5                                 | Primary transcription of 2’-OMe-modified nucleic acids from DNA              | 166  |
| T7 RNAP mutant  | RGVG-M6                                 | Partially 2’-modified nucleic acids                                         | 166  |
| T7 RNAP mutant  | RGVG-M5                                 | Recently, Ellington and co-workers employed the RT-CSR method to further evolve a previously evolved mutant of KOD DNAP, RTX, which could reverse transcribe RNA faithfully, and obtained mutant RTX-Ome v6 that could reverse transcribe 2’-OMe-RNA efficiently. | 167  |

Efficient synthesis of sugar-modified nucleic acids with polymerases is usually more challenging, and thus much effort has been made on engineering polymerases to achieve this goal. Using the phage-display-based method for polymerase evolution, SF of Taq DNAP has been evolved to efficiently synthesize, reverse transcribe, and even amplify nucleic acids with various 2’-modifications, including 2’-OMe, 2’-F, 2’-Cl, 2’-Az, 2’-Am and 2’-arabino-modifications. Among the evolved SF mutants, SFM4-3 and SFM4-6 demonstrated good activity for the synthesis of 2’-modified nucleic acids, SFM4-9 was more efficient for the reverse transcription of 2’-modified nucleic acids, and SFM4-3 could PCR amplify partially 2’-modified nucleic acids. Recently, Ellington and co-workers employed the RT-CSR method to further evolve a previously evolved mutant of KOD DNAP, RTX, which could reverse transcribe RNA faithfully, and obtained mutant RTX-Ome v6 that could reverse transcribe 2’-OMe-RNA efficiently. Mutants of T7 RNAP have been extensively investigated for the activity of incorporating 2’-modified nucleotides. T7 RNAP mutant Y639F has been found to be able to use various 2’-substituted-NTPs, including dNTPs, 2’-F-dNTPs, and 2’-Am-dNTPs, as substrates during transcription, and the mutant with one more mutation, T7 RNAP (Y639F, H784A), displayed higher activity against NTPs with bulkier 2’-substitutions.
including 2′-OMe and 2′-Az. Later, Ellington and co-workers carried out directed evolution of T7 RNAP for enhanced activity towards 2′-modified NTPs by randomizing residues R425, G542, Y639 and H784. Active mutants were selected using the autogene selection method, in which the activity of T7 RNAP was coupled with the transcription of an antibiotic resistance gene. The activity towards 2′-modified NTPs of each selected active mutants was then checked. Evolved mutants ‘RGFA’, (‘RGVG’, E593G, and V685A), ‘RGFH’ and ‘RGLH’ showed good activity when 2′-OMe UTP was used as a substrate, and mutants (‘RGVG’, E593G, and V685A) showed the best activity when more kinds of 2′-OMe-NTPs were used as substrates. Further engineering of mutants (‘RGVG’, E593G, and V685A) by introducing more reported mutations responsible for increased activities and thermostability of other T7 RNAP mutants led to the generation of mutants RGVG-M5 and RGVG-M6, which could synthesize 2′-OMe-modified RNA much more efficiently. T7 RNAP mutant VR5-M5 has also been demonstrated to be able to efficiently transcribe RNA containing modified unnatural base Pa from a DNA template containing UBP Ds–Px, and allowed the production of functional RNA molecules with both 2′-modification and an expanded genetic alphabet.

Polymerases for the efficient synthesis of nucleic acids in which the entire pentose is replaced with unnatural sugars have also been developed. For example, a mutant of replicative family B DNAP from Thermococcus gorgonarius, TgoT, has been evolved for the efficient synthesis and reverse transcription of various XNAs with the CST method. Among the evolved TgoT mutants, Pol6G12 showed good activity for the synthesis of HNA. PolC7 showed good activities for the syntheses of CeNA and LNA. PolD4K showed good activities for the syntheses of ANA and FANA. RT521 showed good activities for the syntheses of TNA and reverse transcription of HNA, ANA, FANA and TNA. RT521K showed good activities for the reverse transcription of CeNA and LNA. Recently, mutant RT521K was further evolved with the CBL RT selection method and reverse transcription activity screening to be reverse transcriptases for 2′-OMe-RNA, HNA, AtNA, 2′-MOE-RNA and PS 2′-MOE RNA with varied efficiencies. Natural 9°N, Deep Vent, and Vent DNAP were shown to be able to synthesize a short stretch of TNA from a DNA template with tNTPs, and several mutants of 9°N DNAP, A485L (Therminator), Y409V, and Y409V, A485L double mutant, demonstrated enhanced activity to extend a primer with tNTPs. Among these mutants, Therminator has the highest

---

**Fig. 4** Distribution of mutations in engineered polymerases with synthesis, reverse transcription, or replication activities for unnatural nucleic acids. Key mutations in engineered (A) Taq DNA polymerase (yellow); (B) T7 RNA polymerase (cyan); (C) KOD DNA polymerase (red); and (D) Tgo DNA polymerase (green). The DNA templates and DNA primers are shown in orange and blue, respectively, while RNA product is shown in purple.
activity for TNA synthesis, and has been used for the construction of a TNA selection system.173 Using the DrOPS strategy, Chaput and co-workers carried out directed evolution of a mutant of 9′N DNAP, 9n-GLK (Y409G, A485L and E664K), and obtained mutant 9n-YRI harboring mutations A485R and E664I and mutant 9n-NVA harboring mutations Y409N, D432G, A485V, V636A and E664A.163 Both of the mutants could efficiently synthesize TNA in the absence of manganese, and thus increase the fidelity of TNA synthesis. By sampling mutations A485R and E664I in other homologous polymerase scaffolds, efficient TNA polymerases, Kod-R, Tgo-R, DV-R, which are mutants of KOD, Tgo and Deep Vent DNAPs harboring mutations, have been identified.98 Combining the microfluidic screening method and deep mutational scanning, two other mutants of KOD DNAP with enhanced TNA synthesis activity, Kod-RS and Kod-QS, both of which harbored two epistatic mutations, have been identified.174 Mutant Kod-RS also demonstrated inverted substrate specificity towards tNTPs and dNTPs, compared with wild type KOD DNAP. Further screening of Kod-RS variants with mutations in tiles 6 and 8 of the thumb subdomain led to the discovery of mutant Kod-RSGA, which demonstrated enhanced activity, high fidelity, and low template sequence bias for TNA synthesis.175 KOD DNAP has also been engineered for efficient synthesis of other unnatural nucleic acids. Obika and co-workers developed KOD DNAP mutants KOD DGLNK and KOD DLK, which could efficiently synthesize LNA or 2′-Ome-RNA from DNA templates and reverse transcribe LNA or 2′-Ome-RNA to DNA, respectively.176 Recently, Chaput and co-workers systematically compared the activities of some natural and evolved polymerases for the synthesis and reverse transcription of different XNAs.177 Natural 9′N, Deep Vent, Tgo and KOD DNAPs showed the ability to synthesize full-length FANA and limited activity for the syntheses of other XNAs. Laboratory-evolved polymerases, including Tgo-QGLK, Tgo-6G12, TgoD4K, Tgo-6G12-I521L, Tgo-EPLFH, and Kod-RSGA demonstrated varied activities for the syntheses of RNA, FANA, ANA, HNA, TNA, and 3′-2′ phosphonomethyl-threosyl nucleic acid (PMT). Full-length products of different XNAs could be produced by different polymerase mutants. In another study, Tgo-EPLFH was demonstrated to be a tPhoNA synthase, while Tgo RT521 and KOD RT521K showed efficient ability to reverse transcribe tPhoNA into DNA.180 Bst DNAP displayed good activities for the reverse transcription of FANA and TNA, but much lower activity for the reverse transcription of ANA.177–179 Other than the extensively explored polymerases described above, some other natural or mutated polymerases have also been investigated for the activities towards unnatural substrates. For example, production of HNA, FANA, and 2′-F-DNA with phi29 DNAP mutant D12A has been reported.180

Polymerases can be evolved to be efficient for the synthesis of nucleic acids with bulky modifications on the phosphate moiety as well. For example, Holliger and co-workers further engineered a Tgo DNAP mutant, RT521L, which was previously evolved to be a reverse transcriptase for several XNAs, for efficient synthesis of pHNA.64 After screening of a site-saturation mutagenesis library, evolution with the CST method, and reverse introduction of a single point mutation, they successfully obtained a mutant, PGV2, with enhanced activity for the synthesis of fully modified pHNAs.

Engineered polymerase mutants have also found application in the synthesis of nucleic acids containing combined modifications on different moieties. For example, KOD DNAP mutant Kod-RSGA, which was evolved for efficient TNA synthesis, also demonstrated good activities against tUTP containing various C5-modifications, and was used to synthesize TNA containing functionalized nucleobases.68 In another example, T7 RNAP mutant Y639F was shown to be efficient for the transcription of RNA with 2′-deoxy-2′-fluoro (5-ethyl) uridine triphosphate and other natural NTPs, and used for the evolution of 2′-F-modified RNA-scaffolded carbohydrate clusters.67,168 Very recently, Niu and co-workers synthesized C8-allyne-FANA UTP, and demonstrated its enzymatic incorporation into FANA by Tgo DNAP.181 This work further enriched the XNA toolbox with components containing clickable handles.

Besides proteinaceous polymerases, Z RNA polymerase ribozyme, which is an RNA replicase generated via in vitro evolution, has also been investigated for its activity of incorporating unnatural nucleoside triphosphates.182 It was found that this ribozyme was able to incorporate different sugar or base-modified nucleoside triphosphates with varied efficiencies, as well as efficiently replicate UBP isoG–isoC under appropriate conditions.

In vitro application of unnatural nucleic acids and tailored polymerases

By combining use of the unnatural nucleic acids and their polymerases, either natural or engineered ones, novel aptamers, biocatalysts, and biomaterials can be produced. The acquirement of aptamers or biocatalysts composed of unnatural nucleic acids is usually achieved via evolution of unnatural nucleic acids from a pool of randomized sequences.183 The evolution can be carried out mainly with two procedures when proper polymerases are available.15 In one procedure, a randomized DNA pool is first transcribed into an unnatural nucleic acid pool, and after the selection, the unnatural nucleic acid pool is reverse transcribed back into a DNA pool, amplified, and then subjected into next round of evolution.184 In another procedure, the unnatural nucleic acid pool is directly amplified, and subjected into next round of evolution.128 Novel functionalities and properties can also be incorporated into nucleic-acid-based materials by using nucleic acid components containing unnatural moieties.

Selection of improved aptamers

Systematic Evolution of Ligands by Exponential Enrichment (SELEX) technology has been broadly used to evolve aptamers with high affinity and specificity towards specific targets, ranging from small molecules to cells.185 Using unnatural nucleic acid triphosphates and polymerases that can recognize them, unnatural nucleotides can be introduced into the sequences in the pools for selection in SELEX, and serve to integrate novel functionalities and improved properties into the selected aptamers.
Modifications on nucleobases can help expand the chemical diversity of and add novel functionalities to aptamers. For example, SELEX experiments have been carried out with DNA containing hydrophobic groups or amino-acid-like modifications attached to uracil or both uracil and cytosine nucleobases, resulting in the generation of protein-targeting high-affinity aptamers, which were called slow off-rate modified aptamers (SOMAmers).\textsuperscript{183,186} Due to its good acceptance of base-modified triphosphates, KOD (exo\textsuperscript{+}) DNAP has been employed to generate SELEX libraries with 5-modified dC and dU.\textsuperscript{183} Application of DNA containing UBPs in SELEX significantly expands the sequence diversity of the pools to be selected, and incorporates the properties of the unnatural nucleobases into the selected aptamers, which has proven effective to increase the probability of obtaining aptamers with higher affinities. Hirao and co-workers carried out SELEX with DNA containing Ds, and successfully obtained high-affinity aptamers for vascular endothelial cell growth factor-165 (VEGF-165) and interferon-γ (IFN-γ), with the $K_d$ values in the subnanomolar to astonishing subpicomolar range.\textsuperscript{6} Later, DNA aptamers containing Ds or both Ds and Pa with high affinity towards von Willebrand factor A1-domain (vWF) or dengue non-structural protein 1 (DEN-NS1) serotypes were reported.\textsuperscript{187,188} Hydrophobic UBP Z-P developed by Benner’s group has also been extensively applied in the SELEX of aptamers with expanded genetic information for various targets, including different cell lines and proteins.\textsuperscript{189–192}

Development of aptamers with unnatural sugar backbones has drawn even more attention, since modification of the sugar backbone can lead to a dramatic improvement of the overall properties of the aptamers, such as obtaining good chemical or biological stabilities, which are properties that natural DNA and RNA aptamers lack the most for practical applications. By employing evolved SF mutants to transcribe, reverse transcribe, or amplify 2′-modified DNAs, Romesberg and co-workers selected fully 2′-OMe-modified or partially 2′-modified aptamers against human neutrophil elastase (HNE), which displayed good biological stability and retained high affinity in a high concentration of salt.\textsuperscript{190} Recently, they reported the selection of HNE and factor IXa aptamers with large hydrophobic groups attached to the 2′-position of the sugar backbone by producing 2′-Az-DNA with SF mutant SFM4-3 and coupling alkyne modified molecules to the 2′-azido group via click chemistry.\textsuperscript{128} It was found that these 2′-hydrophobic groups significantly increased not only the binding affinity, but also the serum stability of the selected aptamers. With the assistance of T7 RNAP, Matsuda and co-workers successfully selected 4′-thiol-modified RNA aptamers against human α-thrombin, which have not only high binding affinity, but also superior stability toward RNase A.\textsuperscript{194,195} Holliger and co-workers demonstrated the application of Tgo DNAP mutants that they evolved in SELEX experiments for HNA aptamers against different targets, including hen egg lysozyme (HEL) and HIV trans-activating response RNA (TAR).\textsuperscript{104} Later, using one mutant of Tgo DNAP, D4K, to transcribe and reverse transcribe FANA, DeStefano and co-workers selected FANA aptamers against HIV-1 reverse transcriptase, HIV-1 integrase, and very recently receptor binding domain of SARS-CoV-2 S protein.\textsuperscript{196–198} TNA aptamers against various targets, including small molecules and proteins, were selected either with a DNA display strategy, in which the polymerase-synthesized TNA was attached to the template DNA annealed with its complementary strand during selection, or through cycles of the transcription–selection–reverse transcription–amplification process.\textsuperscript{199–201} Different polymerases, including Therminator DNAP and a mutant of KOD DNAP, Kod-RI, have been used for the synthesis/transcription of TNA, and Bst DNAP has been used for the reverse transcription of TNA in these studies. Recently, using TNA polymerase in combination with nucleobase-modified tNTPs, a stable TNA aptamer with functionalized nucleobases has also been selected.\textsuperscript{202,203} Mirror-image DNA has drawn broad interest in recent years, since it possesses good resistance to nucleases while retaining similar properties and functions of DNA.\textsuperscript{204} In a very recent study, Zhu and co-workers carried out SELEX experiment with a chemically synthesized mirror-image DNAP α-DpO-5m, and successfully obtained bistable 1- DNA aptamers against human thrombin.\textsuperscript{205}

**Generation of new catalysts**

The fact that similar to natural DNA and RNA, nucleic acids with unnatural moieties can fold into structures defined by their sequences makes it possible to build novel ribozyme-like catalysts with these unnatural nucleic acids, and the unnatural moieties can presumably provide these catalysts with much more possibilities for the catalytic activities, properties, and applications compared with natural ribozymes.

Nucleobase modification can be used to attach functional groups, including amino acid-like side chains, to nucleic acid catalysts, and thus confer novel activities, such as protein enzyme-like activities, to these catalysts. For example, recently, Perrin and co-workers used dCTP and dUTP modified with arginine and lysine-like side chains for the selection of DNAzymes that could cleave RNA in a divalent metal cation-independent manner.\textsuperscript{206}

Development of nucleic acid catalysts with unnatural sugar backbones not only expands the scope of macromolecular biocatalysts out of DNA, RNA and protein, but also has great potential to provide practically valuable catalysts with superior biostability. Using evolved TgoT DNAP mutants, Holliger and co-workers successfully selected ANA, FANA, HNA, and CeNA enzymes (XNAzymes) that could cleave or ligate RNA substrates, as well as a FANA enzyme with XNA–XNA ligase activity.\textsuperscript{7} Later, Chaput and co-workers evolved a general RNA-cleaving FANA enzyme with both strong catalytic activity and good nuclease-resistance, which could be further engineered to target different RNA sequences.\textsuperscript{178} They also reported the introduction of XNA modifications, including FANA and TNA nucleotides, into an existing DNAzyme scaffold for the construction of a novel enzyme, X10–23, with enhanced biological stability and good catalytic activity, and demonstrated the application of X10–23 in gene knockdown and pathogen detection.\textsuperscript{207–209} Recently, selection of TNA enzymes with RNA cleavage or ligation activity has been reported by Yu and co-workers.\textsuperscript{210,211}
**Construction of novel biomaterials**

Biomaterials composed of nucleic acids have been extensively developed and used in numerous applications in recent years, due to their fascinating properties, including high programmability and good biocompatibility. Integration of unnatural nucleic acid components is helpful to further improve the properties, increase the functions and expand the applications of these materials.14

Introduction of nucleobases modified with functional groups into nucleic acid materials immediately enables the coupling of various molecules onto these materials and thus expands the functionalities of these materials. For example, Brown and co-workers employed RCA with base-modified dUTP and dCTP to construct modified DNA nanoflowers, to which various cargos, including fluorophores and functional peptides, could be densely attached, and demonstrated their potential use in diagnostics and therapeutics. UBPs can be employed to increase the number of possible DNA or RNA sequences used for the assembly of nucleic acid nanostructures, and also to make these nanostructures invadable to natural DNAs or RNAs. For example, Tan and co-workers recently used DNA sequences containing unnatural bases Z and P to construct an aptamer-nanotrain assembly, and demonstrated its application in drug delivery.212

Modifications on the sugar-phosphate backbones of nucleic acid frameworks are valuable for augmentation of nucleic-acid-based materials with enhanced thermal, chemical and biological stabilities. Taylor and co-workers demonstrated the assembly of different nanostructures, including tetrahedron and octahedron, with various XNAs, including 2'-F-DNA, FANA, HNA or CeNA.213 Recently, Li and co-workers constructed FANA-based double crossover nanotiles with increased thermal and biological stability, and demonstrated their potential in cellular delivery of small molecules under physiological conditions.214 Other than improving the properties of nucleic acid materials, modification of the sugar-phosphate backbone can also be used for functionalizing the frameworks of, and even providing new strategies for the construction of, nucleic acid materials. For example, Chen and Romesberg used 2'-Az-DNA produced by SFM4-3 polymerase for the construction of a novel DNA hydrogel, in which the 2'-Az group was coupled with ssDNA primers for PCR crosslinking of the 2'-Az-DNA scaffolds.127

**Construction of semi-synthetic organisms (SSOs) with unnatural nucleic acids**

With the fast development of unnatural nucleic acids that can efficiently serve as carriers of genetic information, it is possible to apply them for increasing the diversity of or replacing some of the genetic materials in living cells, which leads to the production of SSOS.11 Expansion of the genetic alphabet and even the central dogma in living organisms with unnatural nucleic acid components is attractive, not only because it may help deepen our understanding of life, but also because the development of novel functional biomacromolecules, especially biopharmaceuticals, will greatly benefit from it.4,215 After years of efforts, UBPs developed by different groups have been used to efficiently expand the genetic alphabet to different extents in vitro, and the UBPs developed by Romesberg’s group were used to build the first SSO with an expanded genetic alphabet, which was then optimized, and successfully used to express proteins with several unnatural amino acids incorporated at the same time.8–11,216 Meanwhile, several successful preliminary attempts of introducing XNAs into living cells have also been demonstrated. All these efforts have initiated the construction and application of an expanded central dogma.

**In vivo expansion of the genetic alphabet**

Development of UBPs and their introduction into DNA and RNA have significantly expanded the genetic alphabet in vitro. Replication, transcription, and reverse transcription of UBPs developed by different groups have been achieved with various polymerases as described above, and the most representative examples are summarized in Fig. 5. To thoroughly complete the retrieval of increased genetic information encoded by UBPs in DNA, the RNA transcribed from UBP-containing DNA has to be further translated into protein, with the unnatural nucleobase-containing codons properly decoded. In 2002, Hirao and co-workers demonstrated the use of UBP s-y to incorporate unnatural amino acids, exemplified by 3-chlorotyrosine, into proteins in vitro.40

To expand the genetic alphabet with UBPs in vivo is much more challenging, and key issues that need to be addressed include the availability of unnatural nucleoside triphosphates in...
the cells, recognition of UBPs by endogenous replication, transcription, and translation machineries, and stability of UBPs in the cells during cell growth and propagation. In 2014, Romesberg and co-workers reported the first SSO with an expanded genetic alphabet, in which an initial information plasmid was constructed with UBP NaM–TPT3, and then replicated with dNaMTP and d5SICSTP.11 Nucleoside triphosphate transporter from Phaeodactylum tricornutum (PNTT2) was employed to import dNaMTP and d5SICSTP into the cytoplasm of E. coli cells, allowing the in vivo replication of NaM–5SICS. Later, by using chemically optimized UBP NaM–TPT3 instead of NaM–5SICS for in vivo replication, engineering the PtNTT2 transporter, introducing the CRISPR/Cas system to eliminate DNA sequences that had lost the UBP, the SSO was optimized for robust growth, constitutive unnatural nucleoside triphosphate uptake, and much better UBP retention.9 In 2017, in vivo transcription and translation of UBP to incorporate non-canonical amino acids (ncAAs) into proteins was accomplished with the SSO.10

Since the successful development of SSOs for the storage and retrieval of increased genetic information, lots of efforts have been made to further explore and optimize the SSOs. For example, exploration of the contributions of different endogenous polymerases on UBP replication and the effects of cellular DNA repair mechanisms on UBP retention led to replisome reprogramming of the SSO for increased UBP retention, and subsequently allowed the incorporation of UBP into the chromosome of the SSO.12 Other than chassis cells for SSO construction, UBPs and unnatural triphosphates can also be continuously optimized for higher efficiencies of triphosphate uptake, in vivo replication, transcription, and translation. Early efforts of constructing SSOs used UBPs and unnatural triphosphates that have been screened and optimized based on in vitro SAR analysis, and thus might be less optimal for in vivo performance. The successful construction of SSOs enabled in vivo SAR analysis of UBPs, which led to the identification of more optimal UBPs and unnatural triphosphates for in vivo applications, exemplified by the combination of UBP CNM0–TPT3 and triphosphates NaMTP and TAT1TP, the use of which gave a high yield of a protein with high-fidelity incorporation of an ncAA.14,15

Expansion of the genetic alphabet with UBPs led to a great increase in the number of genetic codons, allowing the incorporation of much more kinds of amino acids into a protein at the same time. However, translation efficiencies of different unnatural base-containing codons can be dramatically different, and selective use of these codons for incorporating ncAAs into proteins is thus important for good protein yields, as well as high translation fidelity. Romesberg and co-workers have systematically analyzed unnatural codons, and identified nine most promising ones for efficient incorporation of ncAAs.16 Using three orthogonal ones of these codons, they successfully constructed an SSO with 67 codons, which includes 64 conventional codons and 3 new codons with unnatural bases. SSOs with additional sense codons containing unnatural bases have immediate application in producing novel protein products, including proteins site-specifically conjugated with other molecules for therapeutic use.17 For example, employing an SSO, human cytokine IL-2 variants, in which a modifiable unnatural amino acid was incorporated by decoding an unnatural base-containing codon, were produced, site-specifically modified with PEG polymers, and screened for altered receptor binding specificities and improved pharmacological properties.18

Expansion of the genetic alphabet in eukaryotes is also attractive, since it will not only allow the incorporation of various ncAAs into proteins that can only be well produced by eukaryotic cells, but also enable the development of molecular tools, including nucleic acid sequences containing unnatural nucleotide derivatives or proteins containing functional ncAAs, for regulating cellular functions or even behaviors of the entire organisms. As an initial effort for constructing eukaryotic SSOs with an expanded genetic alphabet, Romesberg and co-workers carried out translation experiment with unnatural codon–anticodon pairs containing NaM and TPT3 in HEK293 and CHO cells.19 The results suggested that eukaryotic ribosome could decode unnatural codons, and appeared more tolerant to different unnatural codons than prokaryotic ribosomes. Recently, Bormewasser et al. demonstrated the application of functionalized TPT3 for the labeling and visualization of mRNA in living cells.20

Development of methods for the sequencing of UBP-containing DNAs will significantly facilitate the ever-increasing efforts on expanding the genetic alphabet and constructing SSOs. Benner and Hirao groups have developed sequencing methods for their UBP s, respectively, in which the UBPs were first converted into different natural base pairs under different conditions and sequenced, and subsequent alignment and analysis of the resulting sequences revealed the positions of the UBPs.21,22 Hirao’s group also developed a method for UBP sequencing, termed Sanger gap sequencing, in which the sequencing processivity was increased and modified Px analogs were used to generate clear gap patterns in the sequencing spectrum, which indicated the UBP positions.23 Recently, Romesberg and co-workers reported the application of nanopore sequencing for the thorough analysis of DNA containing UBP NaM–TPT3.24

Introduction of other unnatural nucleic acid components into living organisms

Other than constructing SSOs with an expanded genetic alphabet, a lot of efforts have also been made on introducing other unnatural nucleic acid components into living cells to expand the scope of genetic materials in vivo. Mutzel and co-workers demonstrated the replacement of a large proportion of thymine with artificial base 5-chlorouracil in the E. coli genome by evolving a thymidylate synthase-deficient E. coli strain with a gradually increased ratio of 5-chlorouracil to thymine in the medium.25 In most of other approaches, the nucleic acids containing unnatural moieties served as the initial replication templates that carried the genetic information, and were converted into natural DNAs after in vivo replication of the carrier plasmids. For example, Herdewijn and co-workers PCR amplified the gene of a dihydrofolate reductase (DHFR) with 5-chloro-2’-deoxyuridine, 7-deaza-2’-deoxycytidine, 5’-fluoro-2’-deoxyuridine, and 7-deaza-2’-deoxyguanosine triphosphates, cloned and transformed the PCR product into E. coli cells, and...
found that the fully-modified DNA product could serve as a replication template of \textit{E. coli} DNAP, and confer the phenotype encoded by the DHFR gene to the cells.\textsuperscript{21} Kool and co-workers demonstrated that DNA containing single to multiple size-expanded nucleobases (xDNA bases) could be read by \textit{E. coli} DNAP, and \textit{in vivo} converted into fully natural DNA that is functional with high fidelity.\textsuperscript{225} Other than templates containing nucleobase analogs, templates containing unnatural sugar-phosphate backbones have also been explored for encoding genetic information \textit{in vivo}. For example, Pezo and co-workers found that short G/T sequences of CeNA and I/U sequences of ANA or HNA could guide the faithful biosynthesis of DNA sequences in \textit{E. coli} cells.\textsuperscript{226} Liu and co-workers demonstrated that the \textit{thyA} gene containing a few tPhoNA or dPhoNA oligonucleotides was still able to encode the prototrophic phenotype in \textit{thyA}-deficient \textit{E. coli} cells, although DNA propagation was severely diminished due to the addition of these nucleotides.\textsuperscript{69} Recently, \textit{in vivo} transliteration of (S)-ZNA, which has an acyclic phosphonate backbone, to DNA was also demonstrated in \textit{E. coli} cells.\textsuperscript{227} As an example of introducing unnatural nucleic acid components as genetic information carriers into eukaryotes, Matsuda and co-workers reported the transcription of 1’-thiol-modified DNA into natural RNA in mammalian cells.\textsuperscript{48,49}

**Conclusion and perspective**

With the fast development of nucleic acid chemistry and organic synthesis, various unnatural moieties have been introduced into nucleic acids to expand their sequences, structures, properties, functions, and applications. To make full use of these unnatural nucleic acids, traditional and newly developed approaches for polymerase engineering have been employed to empower DNAPs and RNAPs with the ability to efficiently synthesize, reverse transcribe, and even amplify these unnatural nucleic acids. By combining application of unnatural nucleic acids and their polymerases, a series of novel aptamers, biocatalysts, and materials have been produced. Meanwhile, the UBPs have been successfully introduced into living cells to expand the genetic alphabet \textit{in vivo}, and successful attempts of introducing unnatural nucleic acids with modified nucleobases or backbones into living cells have also been demonstrated.\textsuperscript{69,224,227} With the introduction of these artificially synthesized moieties into the genome, \textit{E. coli} has been made into the first set of SSOs. All these efforts summarized above are gradually and yet rapidly expanding the central dogma in different dimensions.

To further expand the central dogma, more genetic polymers with novel modifications or combination of modifications can be designed and synthesized, and their efficient polymerases also need to be discovered or engineered, with the development and employment of novel polymerase evolution strategies, as well as the assistance of computational tools, including novel machine-learning methods.\textsuperscript{228,229} For existing unnatural nucleic acids, transcription of short stretches of them from a DNA template and reverse transcription of them back into DNA processes are already relatively efficient and sufficient for various \textit{in vitro} applications, including SELEX for aptamers and XNAzymes, after years of efforts on engineering their polymerases. However, to achieve direct replication and even efficient amplification of the unnatural nucleic acids, the polymerases have to be engineered to be able to synthesize a strand of unnatural nucleic acid from an unnatural nucleic acid template. Although efficient amplification of partially sugar-modified short unnatural nucleic acids has been demonstrated with evolved DNAPs,\textsuperscript{77,127} further engineering of these polymerases is still needed to achieve efficient replication and amplification of fully sugar-modified long unnatural nucleic acids, which is the prerequisite of actually using these unnatural nucleic acids as full-function and augmented alternatives of DNA for the storage and transmission of genetic information, and will obviously lead to more efficient use of these unnatural nucleic acids, for example, SELEX of unnatural aptamers with less steps. Also, engineering polymerases for efficient transcription of different fully-sugar modified unnatural nucleic acids with big length will enable the full use of these unnatural nucleic acids as RNA alternatives with altered properties and expanded functions, not only for the production of larger biocatalysts or assembled nanomaterials, but also for the transmission of genetic information from the original carrier, such as DNA, to the function performer, say, protein or another genetic polymer. Moreover, in order to translate proteins from an unnatural nucleic acid, efforts have to be made to engineer the translational machinery to well adopt this unnatural nucleic acid, as well as to efficiently decode its genetic information with tRNAs or even other unnatural tRNA alternatives, the efficient charge of which with amino acids again may need extensive engineering of aminoacyl-tRNA synthetases (aaRSs).\textsuperscript{230,231} In an
ideal world, all of the unnatural nucleic acids have efficient polymerases to replicate them, and to transmit genetic information from arbitrary one to another, which will lead to the expansion of the central dogma to higher dimensions (Fig. 6).

For expanding the central dogma in vivo, synthesis or replication of unnatural nucleic acids in living cells needs efficient polymerases as well. Moreover, to be used in vivo, unnatural nucleic acid polymerases need to be further engineered for good substrate specificity immediately, since all of these polymerases were derived from natural DNAPs or RNAPs, and may still possess good activities against dNTPs or NTPs, which are abundant in living cells, and will obviously interfere the synthesis of unnatural nucleic acids from unnatural nucleoside triphosphates. To make the mutant polymerases function better in vivo, their optimal working temperatures and ionic strengths may also need to be engineered to adapt to the internal environment of the hosts. Efficient pathways for cellular polymerases to acquire various unnatural nucleoside triphosphates, either direct import from the medium or step-by-step synthesis via metabolic pathways, also need to be further exploited and optimized immediately. For example, kinases for the phosphorylation of nucleosides, nucleoside monophosphates, and nucleoside diphosphates can be engineered for higher activities against the unnatural substrates, and then employed to produce unnatural nucleoside triphosphates in vivo, as well as to regenerate unnatural triphosphates that have been dephosphorylated by endogenous phosphatases. Initial efforts on engineering the phosphorylation pathways to produce unnatural nucleoside triphosphates have already been made by several groups, including Benner’s group and Romesberg’s group.232–235 Long-term efforts for expanding the central dogma in vivo may include construction of replicable XNA plasmids or chromosomes, establishment of in vivo XNA transcription systems, and engineering of the host cells to balance

Fig. 7 Future expansion of the central dogma in vivo. XNAs can serve as carriers of genetic information, together with DNAs and RNAs, or as functional polymers, together with RNAs and proteins in the cells. UBPs are fully optimized for the expansion of the genetic alphabet, and working in different nucleic acids. Unnatural nucleoside triphosphates for the synthesis of unnatural nucleic acids can be acquired in vivo either via the triphosphate import by nucleoside triphosphate transporters, or through the metabolic pathways for triphosphate synthesis, exemplified by the cascade phosphorylation of unnatural nucleosides.
Review RSC Chemical Biology to further expand the central dogma not only an expanded genetic alphabet, but also an increased structural genetic systems. With all these efforts, organisms with not polymerases with good substrate specificity to build the unna
gonal replication origins or promoters and corresponding and employing replication or transcription systems with ortho
gonal replication origins or promoters and corresponding polymerases with good substrate specificity to build the unnat
ural genetic systems. With all these efforts, organisms with not only an expanded genetic alphabet, but also an increased number of fully functional genetic polymers may be developed to further expand the central dogma in vivo, and find unprecedentedly broad application in the fields of biotechnology and biomedicine in the future (Fig. 7).

Conflicts of interest
There are no conflicts to declare.

Acknowledgements
We thank National Key R&D Program of China (2019YFA0904102), National Natural Science Foundation of China (21978100), Guangdong Provincial Pearl River Talents Program (2019QN01Y228), and the Program for Guangdong Introducing Innovative and Entrepreneurial Teams (2019ZT08Y318) for the financial support of this work.

References
1 M. Schmidt, BioEssays, 2010, 32, 322–331.
2 A. Nikoomanzar, N. Chin, E. J. Yik and J. C. Chaput, Q. Rev. Biophys., 2020, 53, 1–31.
3 T. Chen and F. E. Romesberg, FEBS Lett., 2014, 588, 219–229.
4 K. Duffy, S. Arangundy-Franklin and P. Holliger, BMC Biol., 2020, 18, 112.
5 Z. Ouaray, S. A. Benner, M. M. Georgiadis and N. G. J. Richards, J. Biol. Chem., 2020, 295, 17046–17059.
6 M. Kimoto, R. Yamashige, K. Matsunaga, S. Yokoyama and I. Hirao, Nat. Biotechnol., 2013, 31, 453.
7 A. I. Taylor, V. B. Pinheiro, M. J. Smola, A. S. Morganov, S. Peak-Chew, C. Cozens, K. M. Weeks, P. Herdewijn and P. Holliger, Nature, 2015, 518, 427–430.
8 Y. R. Baker, L. Yuan, J. Chen, R. Belle, R. Carlisle, A. H. El-Sagheer and T. Brown, Nucleic Acids Res., 2021, 49, 9042–9052.
9 Y. Zhang, B. M. Lamb, A. W. Feldman, A. X. Zhou, T. Lavergne, L. Li and F. E. Romesberg, Proc. Natl. Acad. Sci. U. S. A., 2017, 114, 1317–1322.
10 Y. Zhang, J. L. Ptacin, E. C. Fischer, H. R. Aerni, C. E. Caffaro, K. S. Jose, A. W. Feldman, C. R. Turner and F. E. Romesberg, Nature, 2017, 551, 644.
11 D. A. Malyshev, K. Dhami, T. Lavergne, T. Chen, N. Dai, J. M. Foster, Jr., I. R. Correa and F. E. Romesberg, Nature, 2014, 509, 385.
12 E. Eremeeva and P. Herdewijn, Curr. Opin. Biotechnol., 2019, 57, 25–33.
13 K. H. Lee, K. Hamashima, M. Kimoto and I. Hirao, Curr. Opin. Biotechnol., 2018, 51, 8–15.
14 V. B. Pinheiro and P. Holliger, Trends Biotechnol., 2014, 32, 321–328.
15 T. Chen, N. Hongdilokkul, Z. Liu, D. Thirunavukarasu and F. E. Romesberg, Curr. Opin. Chem. Biol., 2016, 34, 80–87.
16 K. G. Devine and S. Jheeta, Life, 2020, 10, 346.
17 M. Kwak and A. Herrmann, Angew. Chem., Int. Ed., 2010, 49, 8574–8587.
18 W. B. Wan and P. P. Seth, J. Med. Chem., 2016, 59, 9645–9667.
19 A. Espinasse, H. K. Lembke, A. A. Cao and E. E. Carlson, RSC Chem. Biol., 2020, 1, 333–351.
20 S. Ochoa and V. T. Milam, Molecules, 2020, 25, 4659.
21 E. Eremeeva, M. Abramov, L. Margamuljana, J. Rozenski, V. Pezo, P. Marliere and P. Herdewijn, Angew. Chem., Int. Ed., 2016, 55, 7515–7519.
22 E. Eremeeva, M. Abramov, L. Margamuljana and P. Herdewijn, Chem. – Eur. J., 2017, 23, 9560–9576.
23 M. Manandhar, E. Chun and F. E. Romesberg, J. Am. Chem. Soc., 2021, 143, 4859–4878.
24 F. E. Romesberg, J. Mol. Biol., 2021, 434, 167331.
25 M. P. Ledbetter, J. M. Craig, R. J. Karadeema, M. T. Noakes, H. C. Kim, S. J. Abell, J. R. Huang, B. A. Anderson, R. Krishnamurthy, J. H. Gundlach and F. E. Romesberg, J. Am. Chem. Soc., 2020, 142, 2110–2114.
26 M. Kimoto and I. Hirao, Chem. Soc. Rev., 2020, 49, 7602–7626.
27 A. M. Leconte, G. T. Hwang, S. Matsuda, P. Capek, Y. Hari and F. E. Romesberg, J. Am. Chem. Soc., 2008, 130, 2336–2343.
28 Y. J. Seo, G. T. Hwang, P. Ordoukhian and F. E. Romesberg, J. Am. Chem. Soc., 2009, 131, 14596.
29 D. A. Malyshev, Y. J. Seo, P. Ordoukhian and F. E. Romesberg, J. Am. Chem. Soc., 2009, 131, 14620.
30 Y. J. Seo, S. Matsuda and F. E. Romesberg, J. Am. Chem. Soc., 2009, 131, 5046.
31 L. Li, M. Degardin, T. Lavergne, D. A. Malyshev, K. Dhami, P. Ordoukhian and F. E. Romesberg, J. Am. Chem. Soc., 2014, 136, 826–829.
32 V. T. Dien, M. Holcomb, A. W. Feldman, E. C. Fischer, T. J. Dwyer and F. E. Romesberg, J. Am. Chem. Soc., 2018, 140, 16115–16123.
33 K. Dhami, D. A. Malyshev, P. Ordoukhian, T. Kubelka, M. Hocek and F. E. Romesberg, Nucleic Acids Res., 2014, 42, 10235–10244.
34 A. W. Feldman, V. T. Dien, R. J. Karadeema, E. C. Fischer, Y. You, B. A. Anderson, R. Krishnamurthy, J. S. Chen, L. Li and F. E. Romesberg, J. Am. Chem. Soc., 2019, 141, 10644–10653.
35 A. W. Feldman and F. E. Romesberg, J. Am. Chem. Soc., 2017, 139, 11427–11433.
36 I. Hirao, M. Kimoto, T. Mitsui, T. Fujiwara, R. Kawai, A. Sato, Y. Harada and S. Yokoyama, Nat. Methods, 2006, 3, 729–735.
234 F. Chen, Y. Zhang, A. B. Daugherty, Z. Yang, R. Shaw, M. Dong, S. Lutz and S. A. Benner, PLoS One, 2017, 12, e174163.
235 Y. Wu, M. Fa, E. L. Tae, P. G. Schultz and F. E. Romesberg, J. Am. Chem. Soc., 2002, 124, 14626–14630.
236 J. C. Chaput, P. Herdewijn and M. Hollenstein, ChemBioChem, 2020, 21, 1408–1411.
237 D. A. Malyshev, K. Dhami, H. T. Quach, T. Lavergne, P. Ordoukhanian, A. Torkamani and F. E. Romesberg, Proc. Natl. Acad. Sci. U. S. A., 2012, 109, 12005–12010.
238 V. T. Dien, E. M. Sydney, J. K. Rebekah and E. Floyd Romesberg, Curr. Opin. Chem. Biol., 2018, 46, 196–202.