Quarterly Reviews of Biophysics

cambridge.org/qrb

Engineering polymerases for applications in synthetic biology

Ali Nikoomanzar[1], Nicholas Chim[1], Eric J. Yik[1] and John C. Chaput[1,2,3]

[1]Department of Pharmaceutical Sciences, University of California, Irvine, CA 92697-3958, USA; [2]Department of Chemistry, University of California, Irvine, CA 92697-3958, USA and [3]Department of Molecular Biology and Biochemistry, University of California, Irvine, CA 92697-3958, USA

Review

Cite this article: Nikoomanzar A, Chim N, Yik EJ, Chaput JC (2020). Engineering polymerases for applications in synthetic biology. *Quarterly Reviews of Biophysics* **53**, e8, 1–31. https://doi.org/10.1017/S0033583520000050

Received: 6 April 2020
Revised: 24 June 2020
Accepted: 28 June 2020

Key words:
Aptamers; catalysts; polymerase engineering; SELEX; synthetic biology; xeno-nucleic acid (XNA); XNAzymes

Author for correspondence:
John C. Chaput,
E-mail: jchaput@uci.edu

Abstract

DNA polymerases play a central role in biology by transferring genetic information from one generation to the next during cell division. Harnessing the power of these enzymes in the laboratory has fueled an increase in biomedical applications that involve the synthesis, amplification, and sequencing of DNA. However, the high substrate specificity exhibited by most naturally occurring DNA polymerases often precludes their use in practical applications that require modified substrates. Moving beyond natural genetic polymers requires sophisticated enzyme-engineering technologies that can be used to direct the evolution of engineered polymerases that function with tailor-made activities. Such efforts are expected to uniquely drive emerging applications in synthetic biology by enabling the synthesis, replication, and evolution of synthetic genetic polymers with new physicochemical properties.

Table of contents

Introduction 1

Natural polymerases 2
- Fundamentals of DNA synthesis 2
- Visualizing DNA synthesis through snapshots of trapped intermediates 3
- Capturing phosphodiester bond formation by time-resolved crystallography 6

Promiscuous activities of natural polymerases 7
- Natural DNA polymerases that function with reverse transcription activity 7
- Expanding the genetic alphabet with new hydrogen-bonding base pairs 7
- Expanding the genetic alphabet with hydrophobic base pairs 8
- Expanding the genetic alphabet with metal-mediated base pairs 9
- Replicating six-letter genetic alphabets with increased efficiency and fidelity 9
- Testing hypotheses about polymerase recognition 10
- Recognizing chemical modifications made to nucleobase positions 10
- Propagation and evolution of an artificial genetic system 11
- Structural insights into Bst DNA polymerase as an XNA reverse transcriptase 12

Engineering polymerase functions by rational design 13
- Structural permutations of natural DNA polymerases 13
- Exonuclease silencing 15
- Accelerating DNA synthesis with non-specific DNA-binding domains 15
- Determinants of sugar recognition 16
- Improving DNA polymerase performance for PCR 17

Engineering polymerases by directed evolution 18
- Phage display 18
- Compartmentalized self-replication 19
- Compartmentalized self-tagging 20
- Droplet-based optical polymerase sorting 20

Applications in synthetic biology 22
- Synthetic antibodies 22
- Catalysts for RNA modifying reactions 23

Future directions 24

Conclusion 25

© The Author(s), 2020. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

CAMBRIDGE UNIVERSITY PRESS

Introduction

DNA polymerases are an ancient family of enzymes responsible for replicating the genomes of organisms during cell division. Their movement on a template associates them to molecular motors that are powered by the free energy of nucleotide polymerization (Gelles and Landick, 1998). However, in contrast to most molecular motors, which are largely responsible for transporting cargo along protein tracks (Schliwa and Woehlke, 2003), polymerases are a class of enzymes whose movements facilitate the transfer of information from parent to daughter strands using well-established Watson–Crick base pairing rules (Watson and Crick, 1953). This impressive feat of chemical synthesis is accomplished through a complicated reaction pathway where each cycle of nucleotide addition involves a set of carefully orchestrated conformational changes that allow the enzyme to form a covalent bond between the growing primer strand and the correct incoming nucleotide (Steitz, 1999). The superiority of these motors is further demonstrated through the use of accessory domains that enable the enzyme to recognize and correct mistakes that arise due to misincorporation events. Thus, polymerases can be thought of as biological scribes capable of forward and reverse motions that allow for the writing and editing of genetic messages with unparalleled speed and accuracy.

The process of nucleotide selection, insertion, and extension is regulated by a series of checkpoints that control the efficiency and fidelity of nucleotide synthesis. Elegant biochemical, kinetic, and structural studies reveal the importance of induced fit in distinguishing correct nucleotides from incorrect nucleotides (Bryant *et al.*, 1983; Johnson, 2008; Ludmann and Marx, 2016), noting that Watson–Crick hydrogen bonds are not always necessary for the replication of a DNA base pair (Moran *et al.*, 1997*b*). Other factors that affect nucleotide recognition include hydrogen bonding to minor groove heteroatoms, base stacking, solvent exclusion, and shape (Kool, 2002). Once present in the active site, chemical bond formation requires the substrate to adopt a productive geometry that leads to phosphodiester bond formation. In all cases, this involves a combination of side chain and divalent metal ion interactions that orient the substrate in a position that is suitable for in-line nucleophilic attack by the terminal 3′ hydroxyl group on the primer strand (Steitz *et al.*, 1994; Genna *et al.*, 2016). In cases where a polymerase is able to incorporate a modified nucleotide, additional molecular recognition events are available to detect changes in the duplex geometry, which often leads to polymerase stalling (Miller and Grollman, 1997). Although these parameters can vary between individual polymerases, the checkpoints of nucleotide selection, chemical bond formation, and primer extension place severe limitations on the synthesis of unnatural nucleic acid polymers by natural DNA polymerases.

One striking example of substrate specificity is the ability for polymerases to discriminate between DNA and RNA substrates inside the cell. The molecular difference between 2′-deoxyribonucleoside triphosphates (dNTPs) and ribonucleoside triphosphates (NTPs) is the presence of a 2′-hydroxyl group on the ribose sugar, which causes the furanose ring to adopt a different sugar pucker (C2′-endo *versus* C3′-endo for DNA and RNA, respectively) (Anosova *et al.*, 2016). Even though intracellular NTP levels are elevated relative to dNTP levels (>10-fold) (Traut, 1994), DNA polymerases, such as *Escherichia coli* (*E. coli*) DNA polymerase I, are able to discriminate against NTPs by a factor of up to 10^5-fold (Astatke *et al.*, 1998). This remarkable level of substrate specificity is achieved by a single bulky amino acid residue, referred to as the 'steric gate', which packs against the 2′ sugar position, preventing the insertion of NTPs into the enzyme active site. The steric gate is now recognized as a common feature of most DNA polymerases (Bonnin *et al.*, 1999; Brown and Suo, 2011).

In this review, we examine the impact of polymerase engineering on the field of synthetic biology. Special emphasis is placed on examples in which engineered polymerases have enabled the synthesis, replication, and evolution of synthetic genetic polymers with new physicochemical properties, such as enhanced ligand binding, catalysis, and biological stability. Such activities represent the forefront of polymerase engineering, as functional non-natural polymers are expected to drive future applications in synthetic biology, biotechnology, and healthcare. We begin with a review of polymerase function and structure, illustrating the latest techniques that have been used to answer fundamental questions about the mechanism of DNA synthesis. Next, we discuss examples where natural polymerases are able to recognize non-cognate and synthetic congeners as substrates either in the template or as nucleoside triphosphates. We then examine several techniques that have been applied to engineer polymerases with desired functional properties. Here we focus our attention on avant-garde strategies that are rapidly advancing the field of polymerase engineering. Finally, we conclude with examples of synthetic biology applications that have arisen due to the availability of engineered polymerases.

Natural polymerases

Fundamentals of DNA synthesis

DNA polymerases follow a primer extension mechanism in which a single strand of parental DNA is used as a template to synthesize the complementary daughter strand. In this reaction, the growing daughter strand is recognized as a primer that is extended in the 5′-3′ direction by sequentially adding the corresponding dNTP to the terminal 3′-hydroxyl group. As illustrated in Fig. 1, the template dictates the sequence of nucleotide addition following the classic Watson–Crick base pairing rules of adenine (A) pairing with thymine (T) and guanine (G) pairing with cytosine (C). Because the polymerase moves down the template in the 3′-5′ direction and the new DNA strand is generated in the 5′-3′ direction, the resulting product is an antiparallel DNA duplex.

Phylogenetic analysis reveals that DNA polymerases organize into seven different highly homologous sequence families (A, B, C, D, X, Y, and RT) (Table 1) (Ito and Braithwaite, 1991) that allow the activity of one member to predict the activity of another member. For example, the mutations required to imbue a natural DNA polymerase with RNA synthesis activity have similar activity when transferred to homologous enzymes (Dunn *et al.*, 2016). As expected, some polymerase families have been more widely studied than others. Thermostable DNA polymerases belonging to the A- and B-family categories have been extensively studied due to their importance in DNA synthesis and sequencing applications. For example, A-family DNA polymerase I isolated from the thermophilic bacterial species *Thermus aquaticus* (Taq) is widely used in quantitative polymerase chain reaction (qPCR) applications due to its 5′-3′ exonuclease activity, which allows for the digestion of a downstream donor–quencher fluorescent probe that quantitatively measures DNA synthesis during polymerase extension (Holland *et al.*, 1991). Taq DNA polymerase is also routinely used for T–A ligation and cloning strategies due to its proclivity

https://doi.org/10.1017/S0033583520000050 Published online by Cambridge University Press

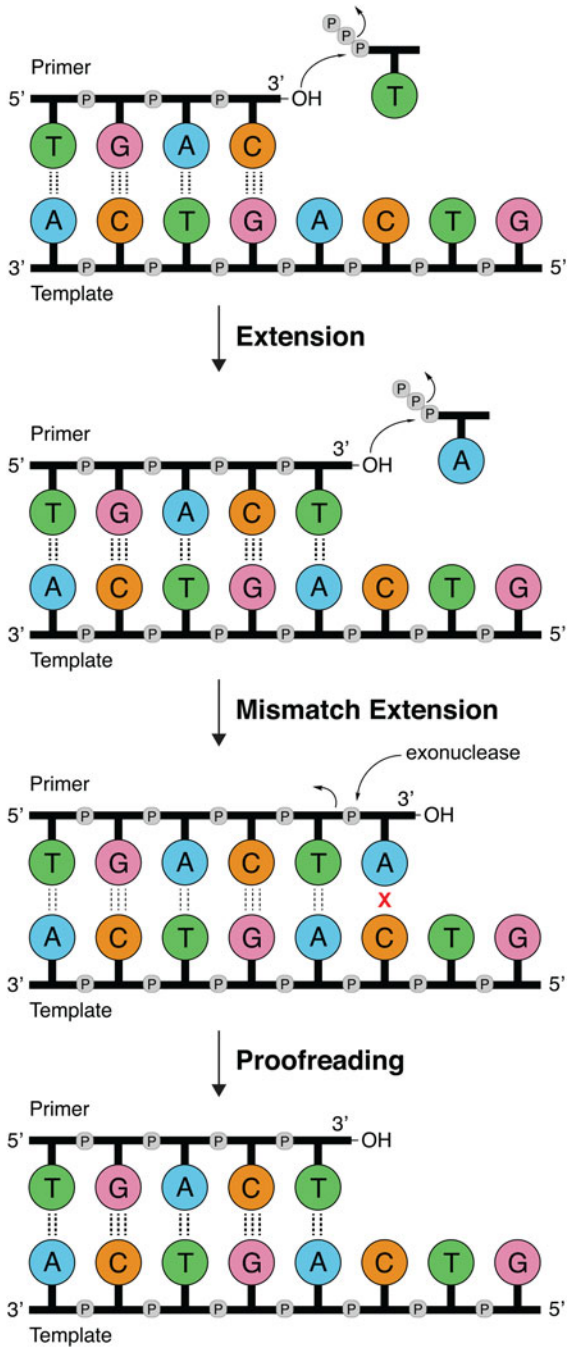


Fig. 1. DNA synthesis and mismatch repair. Natural DNA polymerases extend a DNA primer in the 5′-3′ direction using the template to determine the sequence of the growing strand. Polymerases with 3′-5′ exonuclease activity have the ability to correct mistakes by removing terminal nucleotides that are incorrectly paired with the template.

for adding a single-untemplated adenosine residue to the 3′ end of the daughter strand (Clark, 1988). Hyperthermophilic archaeal B-family DNA polymerases, which include such members as Tgo (*Thermococcus gorgonarius*), Kod (*Thermococcus kodakarensis*), Pfu (*Pyrococcus furiosus*), and 9°N (*Thermococcus 9°N-7*), are the basis of several DNA-sequencing applications (Zhang *et al.*, 2015*a*). These enzymes are known to function with enhanced fidelity due to the presence of a strong 3′-5′ exonuclease proofreading domain. They are also reported to be more resistant than standard Taq polymerase to the inhibitory effects of blood components and detergents (Miura *et al.*, 2013). Interestingly, B-family polymerases have the ability to recognize and stall DNA replication when they encounter uracil residues in the template (Greagg *et al.*, 1999). Structural studies indicate that uracil discrimination is caused by a binding pocket in the amino-terminal domain of the polymerase that accommodates uracil but prevents binding to the four natural DNA bases (Fogg *et al.*, 2002).

Despite extensive sequence diversity, X-ray crystal structures reveal that nearly all polymerases adopt a catalytic domain that closely resembles a human right hand (Steitz, 1999). The one exception is X-family polymerases, which adopt a left-handed polymerase domain (Beard and Wilson, 2000). The catalytic domain is further divided into three subdomains that are commonly referred to as the palm, fingers, and thumb (Fig. 2*a*). The palm subdomain is composed of a β-sheet that forms the base of a deep cleft containing the catalytic residues responsible for promoting phosphodiester bond formation. The fingers subdomain is an α-helical structure lining one side of the cleft, while the thumb subdomain is another α-helical structure lining the opposite side of the cleft. The fingers are responsible for recognizing the incoming nucleoside triphosphate, while the thumb positions the DNA primer–template duplex in the cleft and plays a role in translocation and processivity (Brautigam and Steitz, 1998).

Speed and fidelity are critical parameters for DNA synthesis in rapidly dividing cells. For each nucleotide incorporation, a polymerase must distinguish the correct nucleoside triphosphate from an excess of incorrect and non-cognate (NTP) substrates. Due to their functional roles, the rate and fidelity of DNA synthesis can vary widely between different DNA polymerases (Fig. 2*b* and *c*). Replicative DNA polymerases found in A- and B-families have rates that can exceed 100 nt s^{-1} and intrinsic fidelities in the range of one error in 10^5–10^6 incorporation events (Kunkel, 2004). For example, Kod polymerase functions with a rate of ~200 nt s^{-1}, making it one of the fastest B-family DNA polymerases (Griep *et al.*, 2006). In addition, many polymerases have 3′-5′ exonuclease proofreading activity that exists as a separate domain or a tightly bound subunit, which can remove non-complementary nucleotides after phosphodiester bond formation (Fig. 1) (Kunkel and Bebenek, 2000). These domains increase the fidelity of DNA synthesis by 10-fold (10^6–10^7 nt s^{-1}) relative to polymerases lacking a proofreading domain (Loeb and Monnat, 2008). By comparison, repair polymerases, such as pol β (X-family), are much slower at DNA synthesis and less faithful than replicative DNA polymerases, often functioning with rates in the range of ~10 nt s^{-1} and fidelities on the order of 1 error in 10^2–10^4 incorporation events (Fig. 2*b* and *c*) (Wu *et al.*, 2017). However, the reduced activity of repair polymerases is expected given their functional role in repairing damaged sites in genomic DNA by various cellular repair mechanisms.

Visualizing DNA synthesis through snapshots of trapped intermediates

Since its discovery in 1958, DNA polymerase I has been viewed as a model system for DNA synthesis in cells (Lehman *et al.*, 1958). Structural insights into the mechanism of DNA synthesis have been obtained from crystal structures that trap the enzyme at different stages of the catalytic cycle (Fig. 3) (Chim *et al.*, 2018). Some of the most insightful data have been obtained from high-

https://doi.org/10.1017/S0033583520000050 Published online by Cambridge University Press

Table 1. Properties of natural DNA polymerases

Family	Polymerase	Exonuclease		Strand displacement	Rate	Error rate	Reference
		3′ to 5′	5′ to 3′				
A	DNA Pol I	Yes	Yes	No		9.0×10^{-6}	Kunkel *et al.* (1984)
	Taq	No	Yes	No	2800 min^{-1}	1.5×10^{-4}	Potapov and Ong (2017), Hogrefe *et al.* (2001)
	Bst	No	Yes	No			
	Tth	No	Yes	No			
	T7 DNA Pol	Yes	No	No	v_{max} 100 s^{-1}	15.0×10^{-6}	Mattila *et al.* (1991), Wuite *et al.* (2000)
	Bst large fragment	No	No	Strong	11 500 min^{-1}		Hogrefe *et al.* (2001)
	Klenow fragment	Yes	No	Strong	310 min^{-1}	18.0×10^{-6}	Bebenek *et al.* (1990), Hogrefe *et al.* (2001)
	KlenTaq1	No	No	No		1.5×10^{-4}	Kermekchiev *et al.* (2003)
	Stoffel fragment	No	No	No	v_{max} 130 s^{-1}		Lawyer *et al.* (1993)
	Tth	No	Yes	No			
B	Tgo	Yes	No	Weak	385 min^{-1}		Nikoomanzar *et al.* (2017)
	Kod	Yes	No	Weak	11 777 min^{-1}	1.2×10^{-5}	Nikoomanzar *et al.* (2017), Potapov and Ong (2017)
	Pfu	Yes	No	Weak	560 min^{-1}	5.1×10^{-6}	Hogrefe *et al.* (2001), Potapov and Ong (2017)
	9N	Yes	No	Weak		5.7×10^{-5}	Hogrefe *et al.* (2001)
	Deep vent	Yes	No	Weak		4.0×10^{-6}	Potapov and Ong (2017)
	Vent	Yes	No	Weak	1000 min^{-1}	2.8×10^{-6}	Hogrefe *et al.* (2001)
X	Human pol η	No		Weak	0.4 s^{-1}	3.5×10^{-3}	Takata *et al.* (2006), Arana *et al.* (2007)
	Human pol β	No	No	Weak	v_{max} 1.5 s^{-1}	5×10^{-4}	Yamtitch and Sweasy (2010)
	TdT	No	Yes	No	180 min^{-1}		Boulé *et al.* (2001)
RT	HIV-RT	No	No	Weak	1.9 s^{-1}	7.5×10^{-5}	Reardon and Miller (1990), Hottiger *et al.* (1994), Boyer *et al.* (2007)
	MMLV-RT	No	No	Weak	29 s^{-1}	3.6×10^{-5}	Potter *et al.* (2003), Ji and Loeb (1992)
	Superscript II	No	No	Weak		6.7×10^{-5}	Arezi and Hogrefe (2007)
	Superscript III	No	No	Weak		6.7×10^{-5}	Potter *et al.* (2003)

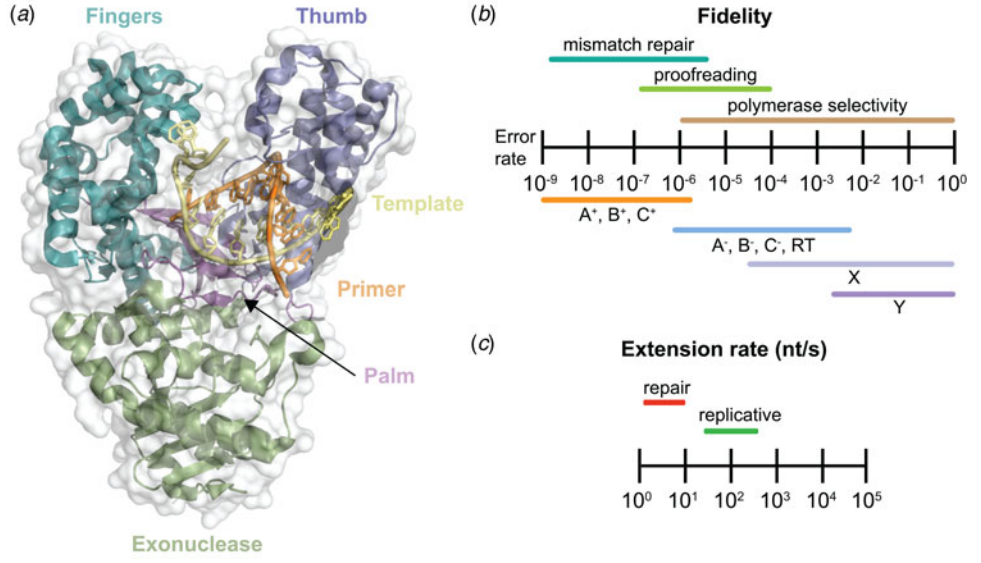


Fig. 2. DNA polymerase structure and catalysis. (*a*) Structure of the binary complex of Bst DNA polymerase bound to the DNA duplex (PDB: 6DSY). (*b*) Fidelity of DNA replication summarized according to different steps that enhance the fidelity of DNA synthesis and polymerase family. +/− Indicates the presence or absence of exonuclease activity. (*c*) Differences in the extension rate between repair and replicative DNA polymerases under *in vitro* conditions independent of accessory proteins that lead to faster rates in the cellular environment.

https://doi.org/10.1017/S0033583520000050 Published online by Cambridge University Press

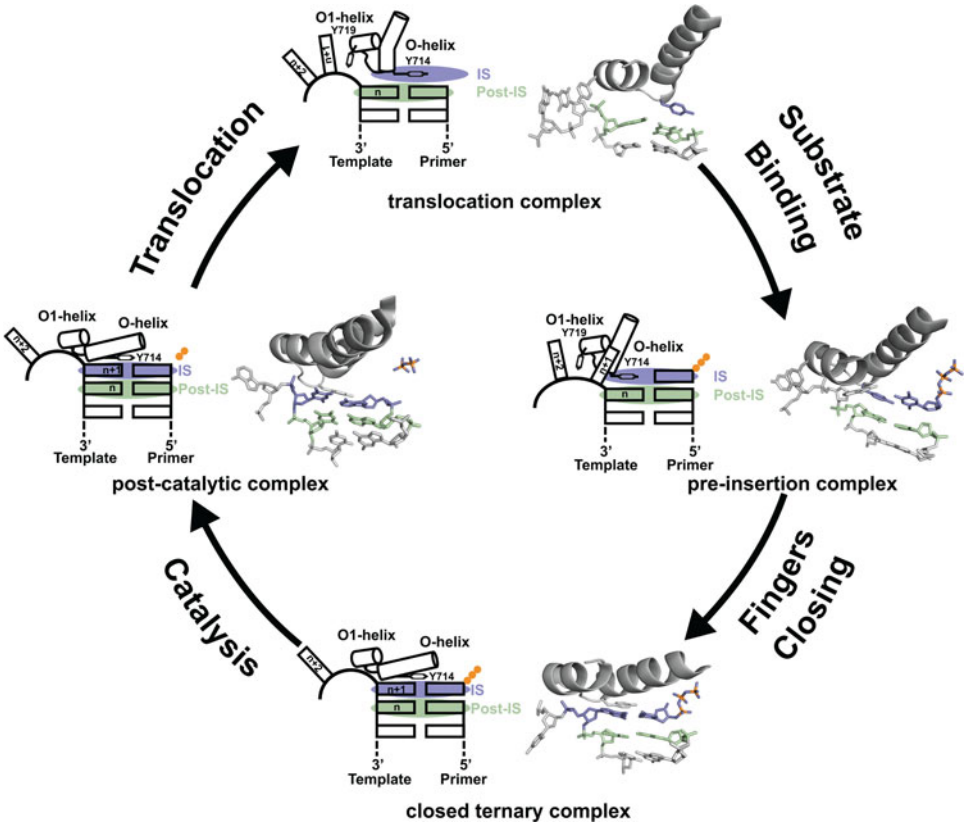


Fig. 3. Mechanism of DNA synthesis. The four key mechanistic steps depict a replication cycle for DNA synthesis. The translocation complex (top) is stabilized by π-stacking interactions between Tyr^{719} and the $n+1$ templating base and between Tyr^{714} and the primer strand. Tyr^{714} occupies the insertion site (IS, purple) while a newly formed base pair is located in the post insertion site (post-IS, green). In the pre-insertion complex (right), the O-helix adjusts to accommodate the incoming dNTP substrate, which binds opposite Tyr^{714} in the IS. In the closed ternary complex (bottom), the polymerase undergoes a major conformational change to allow the $n+1$ templating base to form a nascent base pair with the dNTP substrate in pre-catalytic state. Following catalysis, the finger subdomain remains closed with a trapped pyrophosphate moiety observed in the active site of the post-catalytic complex (left). To complete the cycle, the finger subdomain opens, pyrophosphate is released, and the enzyme translocates to the next position on the template. The translocation (6DSY), pre-insertion (6DSU), and closed ternary complexes (1LV5) are based on crystal structures of Bst DNA polymerase. The post-catalytic complex is T7 RNA polymerase (1S77), a homolog of Bst DNA polymerase. Adapted from Chim *et al*. (2018).

resolution structures of a thermostable bacterial DNA polymerase I member isolated from *Geobacillus stearothermophilus* (Bst, Fig. 2*a*) and its bacteriophage homolog T7 RNA polymerase (Kiefer *et al*., 1998; Yin and Steitz, 2002, 2004). Starting from the Bst binary complex produced from one round of dNTP addition, Tyr^{714} on the O-helix occupies the insertion site, stacking above the newly added nucleotide on the growing primer strand (Chim *et al*., 2018). In this same structure, Tyr^{719} on the O1-helix forms a second stacking interaction with the $n+1$ templating base, thereby preventing the next templating base from entering the active site. In step 2, the polymerase undergoes a conformational change to adopt a pre-insertion complex with the incoming nucleotide paired opposite Tyr^{714} in the enzyme active site (Chim *et al*., 2018). This intermediate, commonly referred to as the open ternary complex, is achieved by releasing the $n+1$ templating base from its stacking interaction with Tyr^{719} and retracting Tyr^{714} to a position above the n templating base in the post-insertion site. In step 3, the enzyme undergoes a more significant conformational change to adopt a closed ternary complex, which defines the pre-catalytic state of the enzyme (Johnson *et al*., 2003). Here, the $n+1$ templating base finally enters the insertion site and forms a Watson–Crick base pair with the incoming nucleotide. In this structure, the fingers have rotated ~40° to allow several lysine and arginine residues on the O-helix to contact the triphosphate moiety of the dNTP substrate. In step 4, the enzyme adopts a post-catalytic complex in which chemical bond formation has occurred and the primer has been extended by one nucleotide (Yin and Steitz, 2004). Close examination of the enzyme active site reveals the presence of the pyrophosphate leaving group, suggesting that pyrophosphate departure coincides with opening of the fingers. To complete the cycle, the polymerase must translocate to the next position on the template to reform the binary complex. Together, crystal structures of the binary complex, pre-insertion site, closed ternary complex, and post-catalytic complex provide a structural view of DNA synthesis by a replicative DNA polymerase.

Structural and kinetic data reveal that DNA polymerase fidelity is governed by subtle local rearrangements that are distinct from the major conformational domain movements observed in the binding and catalysis of cognate Watson–Crick base pairs. In particular, researchers have identified a distinct conformation in A-family polymerases that was suggested to be a fidelity checkpoint for correct nucleotide selection. X-ray crystal structures of Bst DNA polymerase containing mismatched substrates reveal a kink in the O-helix of the finger subdomain that results in a partially closed ternary complex, termed the 'ajar' conformation (Fig. 4*a*) (Wu and Beese, 2011). The wobble base pair between the templating G nucleotide and the incoming TTP substrate

https://doi.org/10.1017/S0033583520000050 Published online by Cambridge University Press

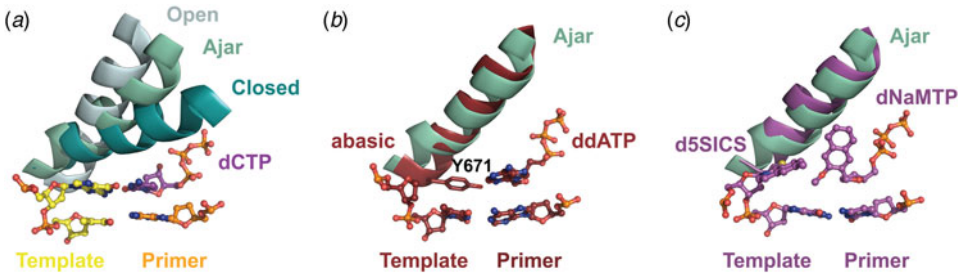


Fig. 4. X-ray crystal structures capturing the ajar conformation. (*a*) Structural comparison of the open (1L3S), ajar (3HP6), and closed (1LV5) ternary conformations of Bst DNA polymerase. The Bst ajar conformation was obtained using a mismatch template-dNTP combination. Structural comparisons showing KlenTaq with an abasic template (3LWL, red, panel *b*) and unnatural 5SICS-NaMTP base pair (4C8K, purple, panel *c*), both structures superimpose on the ajar conformation of Bst DNA polymerases (green).

places the α-phosphate at a distance that is too far from the 3′-OH group of the primer to facilitate efficient in-line attack on the dNTP substrate. This observation is supported by a reduction of at least 100-fold in the rate of nucleotide addition compared to the complementary dCTP substrate (Wu and Beese, 2011).

Förster resonance energy transfer (FRET) studies performed on the large (Klenow) fragment of *E. coli* DNA polymerase I provide additional evidence for the existence of the ajar conformation (Berezhna *et al.*, 2012; Hohlbein *et al.*, 2013). Here, an intermediate FRET species, which appears to be a distinct conformation between the open and closed structures, was found to persist in the presence of mismatched substrates but only transiently exists when the complementary dNTP is present. Interestingly, structures of KlenTaq (Klenow-fragment analog of Taq DNA polymerase) containing an abasic site in the template reveal that the conserved gating tyrosine residue (Fig. 4*b*) can pair opposite an incoming substrate to allow for primer extension, albeit at significantly reduced rates due to the formation of a sub-optimal enzyme active site (Obeid *et al.*, 2010, 2012). More recently, the ajar conformation was witnessed in a ternary structure of KlenTaq bound to the unnatural d5SICS:dNaMTP base pair (Fig. 4*c*) (Betz *et al.*, 2013). Collectively, these data suggest that the ajar conformation plays a functional role in nucleotide discrimination in which base pair mismatches stabilize an intermediate conformation that is not catalytically active.

Capturing phosphodiester bond formation by time-resolved crystallography

Data acquired from structural studies into the mechanism of DNA synthesis confirm the prediction that all polymerases catalyze the same nucleotide-transfer reaction, which involves the formation of a phosphodiester bond through nucleophilic attack of the 3′-OH group of the primer on the α-phosphate of the incoming nucleoside triphosphate with concomitant displacement of the pyrophosphate leaving group (Steitz *et al.*, 1994). The reaction is pH-dependent and analogous to acid-base catalysis, where the nucleophile (3′-OH) needs to be deprotonated and the leaving group (pyrophosphate) needs to be protonated. It requires two-metal ions that stabilize a pentacoordinate transition state in a bimolecular substitution (S_N2) reaction mechanism (Fig. 5). Metal ion A activates the 3′-OH for nucleophilic attack, while metal ion B stabilizes the buildup of negative charge on the pyrophosphate leaving group via coordination to the β- and γ-phosphates. The reaction may be further activated through the formation of an intramolecular hydrogen bond between the 3′-hydroxyl and β-phosphate groups of the incoming dNTP substrate (Genna *et al.*, 2016).

Despite the accumulation of significant structural data showing polymerases from all domains of life trapped in various

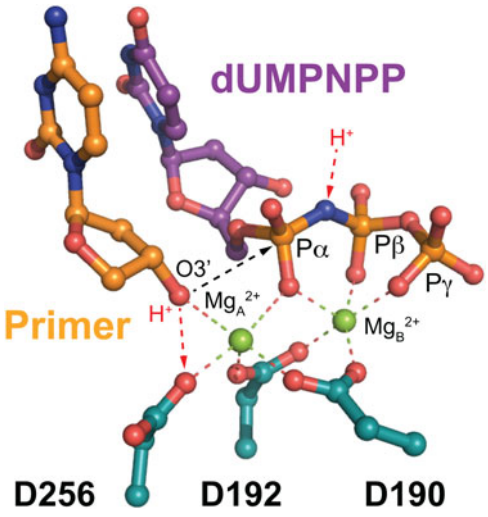


Fig. 5. Conventional two-metal mechanism for DNA synthesis. Mg_A^{2+} assists in deprotonation of the 3′ OH and Mg_B^{2+} stabilizes the transition state and protonation of the pyrophosphate leaving group (2FMS).

stages of DNA synthesis, the actual step of chemical bond formation has long remained elusive. This problem was elegantly solved when Yang and colleagues applied the technique of time-resolved X-ray crystallography to follow the course of phosphodiester bond formation by human polymerase η (pol η) (Nakamura *et al.*, 2012). In this study, inactive pol η crystals were obtained by crystallizing a ternary complex of the polymerase bound to DNA and dATP in the presence of Ca^{2+} ions, a catalytically inactive divalent metal ion. The nucleotide-transfer reaction was then initiated *in crystallo* by transferring individual crystals first to a wash buffer and then to a reaction buffer containing Mg^{2+} ions, which displace the Ca^{2+} ions and allow phosphodiester bond formation to proceed. The reaction was stopped at defined times by freezing the crystals in liquid nitrogen for structural analysis. Electron density maps reveal that Mg^{2+} ions displace the Ca^{2+} ion within the first 40 s, forming the two-metal ion complex required for nucleotide transfer. From 40 to 230 s, the structures show a steady increase in the nucleotide addition product, thus capturing the chemical bond forming step (Fig. 6). Transient densities identified the rate limiting step of the reaction as deprotonation of the 3′-OH group, which is accompanied by a change in the sugar pucker conformation of the terminal nucleotide from C2′ endo to a C3′ endo. Interestingly, a third Mg^{2+} ion was found to be essential for DNA synthesis (Gao and Yang, 2016). Similar results have also been observed for time-resolved reactions performed on DNA polymerase β (pol β) (Freudenthal *et al.*, 2013).

https://doi.org/10.1017/S0033583520000050 Published online by Cambridge University Press

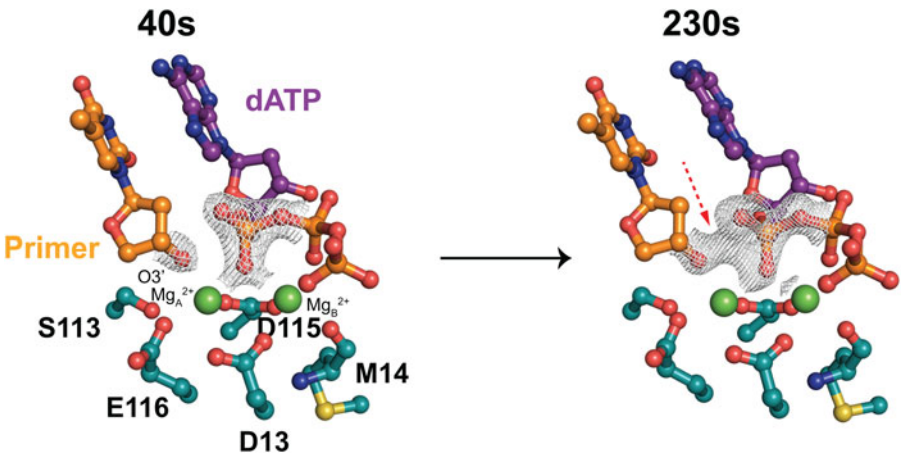


Fig. 6. Phosphodiester bond formation visualized by time-resolved X-ray crystallography. Two views of $2F_o–F_c$ maps (1.5σ) of 40 s (4ECR) and 230 s (4ECV) structures reveal progression of catalysis by pol η.

Promiscuous activities of natural polymerases

Although DNA polymerases are generally thought of as remarkably specific catalysts, many examples now exist where natural DNA polymerases are able to incorporate limited numbers of non-cognate or unnatural substrates into an otherwise natural DNA strand. The catalytic activity and fidelity of these reactions varies significantly depending on the type of chemical modification and the number of chemically modified nucleotides incorporated into the growing strand. In general, natural polymerases are more accepting of base modifications made to the 5-position of pyrimidines and the 7-position of purines than modifications made to the sugar moiety. Reactions of this type are typically performed using polymerases that are either naturally or intentionally deficient in exonuclease activity (exo-), which prevents removal of the modified residue after nucleotide incorporation. Most of the examples cataloged to date involve the incorporation of one or a small number of modified nucleotides into an otherwise natural DNA strand. However, a few cases are known where the template or extension product is composed entirely of non-natural nucleotides.

The varying degrees of tolerance exhibited by natural polymerases for unnatural substrates have played an important role in elucidating the mechanistic underpinnings behind how polymerases recognize their substrates. These details are not easily discerned from crystal structures obtained for polymerases caught at a specific step in the DNA synthesis cycle. Instead, they require chemical analogs that probe the enzyme active site in ways that are not possible purely with natural substrates. As illustrated in the section below, such experiments demonstrate that: (1) Watson–Crick hydrogen bonding groups can be rearranged or removed altogether, and (2) substrate tolerance varies considerably depending on the type of polymerase and chemical modification. Such information has provided insights into the limits of substrate specificity and identified starting points for evolving new variants with improved activity.

Natural DNA polymerases that function with reverse transcription activity

In 1973, Loeb and colleagues were the first to discover the promiscuous activities of natural polymerases by demonstrating that natural RNA templates can be copied into DNA using *E. coli* DNA polymerase I (Loeb *et al.*, 1973). This activity, commonly known as reverse transcription (RT), makes it possible to synthesize the cDNA products of RNA sequences. In nature, reverse transcription is mediated by reverse transcriptases, a class of polymerases that are responsible for replicating the genomes of RNA viruses (Coffin and Fan, 2016). Nearly two decades later, other laboratories recognized that Taq and *Thermus thermophilus* (Tth) DNA polymerases have measurable RT activity with Tth exhibiting 100-fold greater activity than Taq (Jones and Foulkes, 1989; Tse and Forget, 1990; Myers and Gelfand, 1991). Although this activity helped establish the first examples of a coupled RT-PCR process for detecting and quantifying cellular RNAs, Tth's requirement for manganese ions results in higher error rates during cDNA synthesis. More recently, Bergquist and coworkers identified polymerases from other thermophilic organisms that exhibit RT-PCR activity under standard magnesium conditions (Shandilya *et al.*, 2004).

Expanding the genetic alphabet with new hydrogen-bonding base pairs

In 1987, Benner and coworkers suggested that the functional activity of nucleic acid catalysts could be improved by incorporating additional chemical diversity into DNA and RNA (Benner *et al.*, 1987). Toward this goal of augmenting nature's genetic alphabet, several non-natural base pairs were envisioned that would allow for novel hydrogen-bonding schemes between the various hydrogen-bond donor and acceptor groups found on the Watson–Crick face of designer nucleobases (Fig. 7). In 1989, Switzer and Benner demonstrated that this concept was physically possible by enzymatically synthesizing natural genetic polymers containing an unnatural iso-guanine:iso-cytosine (iso-G:iso-C) base pair (Switzer *et al.*, 1989). These experiments were performed using Klenow DNA polymerase and T7 RNA polymerase to synthesize DNA and RNA, respectively. Although iso-G was found to suffer from a minor enol tautomer that leads to mispairing opposite T and iso-C was susceptible to deamination, this foundational study paved the way for what would eventually become an artificially expanded genetic information system that includes the four canonical bases found in DNA plus four additional genetic letters that make up the S:B and Z:P base pairs (Piccirilli *et al.*, 1990; Benner, 2004; Hoshika *et al.*, 2019).

Being the first unnatural base pair, iso-G:iso-C was widely studied in a variety of different contexts. Tor and Dervan used N^6-(6-aminohexyl)isoguanine (6-AH-isoG) to establish a general

https://doi.org/10.1017/S0033583520000050 Published online by Cambridge University Press

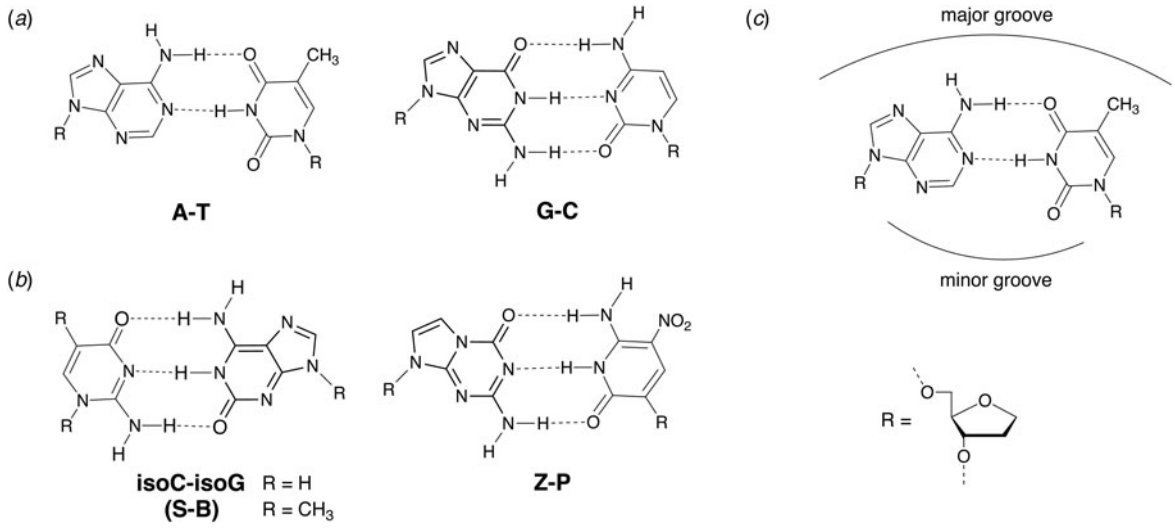


Fig. 7. Chemical structures of natural and alternative DNA base pairs with complementary hydrogen bonding groups. (*a*) The natural A:T and G:C base pairs. (*b*) Examples of alternative base pairs obtained by shuffling the hydrogen bond donor and acceptor groups. (*c*) Schematic view of the major and minor groove regions of a Watson–Crick base pair.

protocol for site-specifically labeling RNA (Tor and Dervan, 1993). Accordingly, T7 RNA polymerase is used to transcribe RNA molecules that contain the 6-AH-isoG nucleotide at a defined position that is then post-transcriptionally modified by coupling biotin or a fluorescent dye to the primary amino group attached to the iso-G nucleobase. Whereas iso-C was originally reported to be prone to deamination, Horn and colleagues found that the 5-methyl iso-C (iso-C^{Me}) analog ameliorates this problem (Horn *et al.*, 1995). The ability to chemically synthesize oligonucleotides containing iso-C and iso-G led to thermal and thermodynamic studies on duplexes arranged in both the antiparallel and parallel strand configurations (Roberts *et al.*, 1997*a*; Seela *et al.*, 1999) as well as the formation of iso-G tetraplex and pentaplex motifs that self-assemble around monovalent cations (Roberts *et al.*, 1997*b*; Chaput and Switzer, 1999; Kang *et al.*, 2012). Moreover, iso-G:iso-C base pairing has been visualized inside the duplex of DNA crystals (Robinson *et al.*, 1998), evaluated in the context of the hammerhead ribozyme (Ng *et al.*, 1994), shown to replicate nonenzymatically (Chaput and Switzer, 2000), and found to be a viable substrate for RecA-mediated DNA recombination (Rice *et al.*, 2000). More recently, the iso-G:iso-C^{Me} base pair has been renamed the S:B pair in honor of its inventors, Switzer and Benner (Hoshika *et al.*, 2019).

Expanding the genetic alphabet with hydrophobic base pairs

An alternative approach to generating unnatural base pairs began in 1997 when Kool and colleagues made the surprising discovery that hydrogen bonding is not an absolute requirement for DNA synthesis (Moran *et al.*, 1997*b*). Steady-state kinetic measurements showed that Klenow DNA polymerase recognizes difluorotoluene (F), a non-hydrogen bonding isostere of thymine (Fig. 8*a*), only ~4-fold less efficiently than natural TTP (Moran *et al.*, 1997*a*). Additional polymerase studies revealed that selectivity for the insertion of A opposite F rather than C, T, or G was strikingly similar to that of T, making F a strong shape mimic of T (Moran *et al.*, 1997*a*). Subsequent study on 4-methylbenzimidazole (Z), a nonpolar analog of adenine (Fig. 8*a*), led to the first demonstration in which a hydrophobic base pair was replicated by a DNA polymerase (Morales and Kool, 1998). This study showed that the unnatural Z:F base pair exhibits strong selectivity against natural nucleotides, with the noted exception of dATP mispairing opposite F in the template. Nevertheless, the ability to replicate hydrophobic base pairs *in vitro* cultivated the notion that hydrophobicity and shape complementarity contribute to the recognition of DNA substrates (Kool, 2002).

Inspired by the success of the Z:F base pair, Schultz and Romesberg applied a more traditional medicinal chemistry approach to identify an array of nonpolar molecules that are recognized as base pairs by natural DNA polymerases (Ogawa *et al.*, 2000). One of the more successful early examples was 7AI, an indole ring system that is capable of self-pairing (7AI:7AI) in duplex DNA (Fig. 8*a*) (Tae *et al.*, 2001). Using Klenow DNA polymerase, 7AI showed modest incorporation efficiency (~200-fold less efficient than natural bases), but high selectivity against natural nucleotides (Tae *et al.*, 2001). However, 7AI is poorly extended after nucleotide incorporation, which limits its utility as an orthogonal third base pair. This problem was partially solved using a second DNA polymerase, mammalian polymerase β (pol β), which allows DNA synthesis to continue from a primer that has been extended with 7AI (Tae *et al.*, 2001).

In 2003, Hirao and colleagues extended the number of nonpolar base pairs that are recognized by DNA polymerases by demonstrating strong shape complementarity between the adenosine analog Q and a new pyrimidine analog pyrrole-2-carbaldehyde (Pa) (Mitsui *et al.*, 2003). The Q:Pa base pair (Fig. 8*a*) was designed to be more selective than the original Q:F base pair, which permits modest to high levels of misincorporation opposite A and T nucleotides (Morales and Kool, 1999). Unlike the 7AI:7AI base pair, Klenow DNA polymerase is able to efficiently incorporate and extend the Q:Pa base pair in both sequence contexts with Q or Pa present in the template strand. Mispairing experiments reveal that dATP inserts opposite either Q or Pa, but that the resulting mispair leads to chain termination in the subsequent extension step. However, PaTP is inserted and extended with low efficiency opposite A, indicating that the geometry of the terminal Pa nucleotide is not a complete impediment to further extension. These data indicate that the Q:Pa pair is an

https://doi.org/10.1017/S0033583520000050 Published online by Cambridge University Press

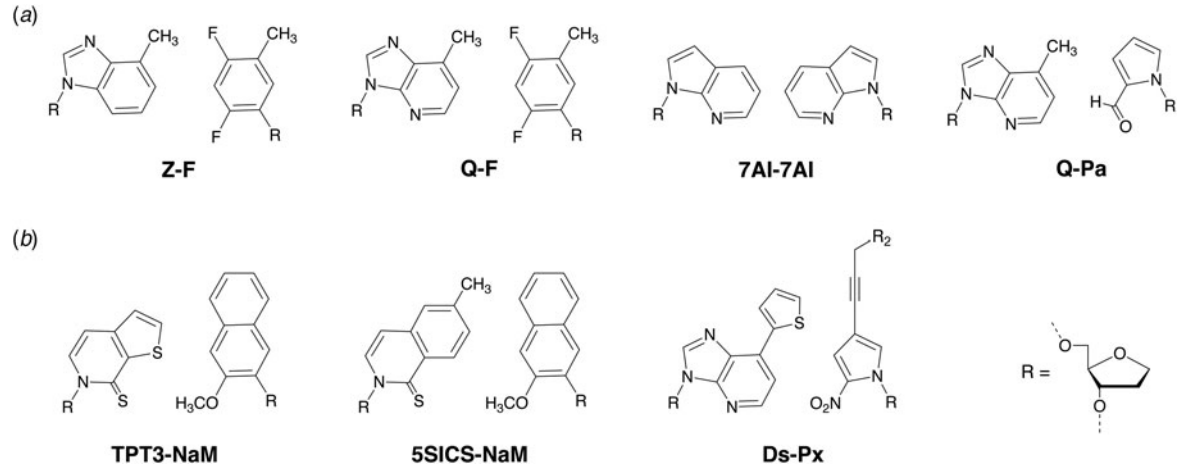


Fig. 8. Chemical structures of hydrophobic base pairs developed for DNA replication. (*a*) First-generation base pairs designed as hydrophobic shape mimics (isosteres) of natural Watson–Crick base pairs. (*b*) Second-generation analogs established for higher replication efficiency and fidelity using an iterative chemical optimization approach of design, synthesize, and test.

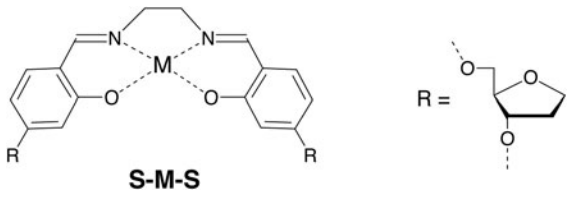


Fig. 9. Chemical structure of a metal-mediated DNA base pair. Metal-mediated base pairs consist of two ligands in the DNA nucleobase position that coordinate a metal ion.

improvement over the original F:Q pair in terms of selectivity and extension efficiency but that further engineering would be required to achieve true orthogonality.

Expanding the genetic alphabet with metal-mediated base pairs

Metal-mediated base pairs represent a third approach for expanding the genetic alphabet beyond the four bases found in nature. Metal-mediated base pairs consist of two artificial bases that coordinate a suitable metal ion in the Watson–Crick base pairing region of a natural base pair (Jash and Muller, 2017). Dozens of examples have been described that coordinate metal ions, such as Cu^{2+}, Ag^{1+}, Hg^{2+}, Pd^{2+}, and Cd^{2+}, in synthetic DNA produced by solid-phase synthesis. Although metal-mediated base pairs have been described in the architectures of several DNA nanostructures (Jash and Muller, 2017), significantly less is known about their recognition properties in the context of DNA replication. One of the more successful examples is the dS–Cu–dS base pair (Fig. 9), which is fully orthogonal and can be PCR amplified in the presence of the canonical A:T and G:C base pairs (Kaul *et al.*, 2011). However, the requirement for an organic co-factor (ethylene diamine) in addition to the inorganic co-factor (Cu^{2+}) may limit the application of the dS–Cu–dS base pair relative to other pairs that rely on an inorganic co-factor alone (Kim and Switzer, 2013; Kobayashi *et al.*, 2016; Rothlisberger *et al.*, 2017). Despite this minor weakness, the ability to design unnatural base pairs based on metal ion coordination chemistry provides ample room for further development. For example, Shionoya and colleagues recently found that Cu^{2+}-mediated artificial base pairing offers a novel approach for controlling the allosteric regulation of catalytic DNA molecules (Nakama *et al.*, 2020). One could imagine applying similar design principles toward the development of metal-responsive materials and logic circuits.

Replicating six-letter genetic alphabets with increased efficiency and fidelity

Early efforts toward the development of orthogonal base pairs led to the realization that many first-generation base pairs suffer from problems that limit their use in practical applications. In some cases, the efficiency of nucleotide incorporation was low when compared to natural bases, while other cases witnessed poor extension kinetics with the polymerase pausing after nucleotide insertion (Hamashima *et al.*, 2018). Another common problem was nucleotide selectivity in the enzyme active site with unnatural bases mispairing to varying degrees with natural bases (Hamashima *et al.*, 2018). To overcome these problems, organic chemistry was used to design new versions of unnatural base pairs that replicate with higher catalytic efficiency and fidelity. Benner and colleagues, for example, developed the Z:P base pair (Fig. 7), which is more stable than a conventional G:C base pair (Wang *et al.*, 2017). In the context of a six-letter genetic alphabet, the Z:P base pair is sufficiently robust that it can be enzymatically synthesized (Yang *et al.*, 2007), amplified by PCR and sequenced (Yang *et al.*, 2011), transcribed into RNA and reverse transcribed back into DNA (Leal *et al.*, 2015), subjected to iterative rounds of *in vitro* selection, and used to evolve aptamers, a type of synthetic antibody (Dunn *et al.*, 2017), that bind to breast and liver cancer cell lines (Sefah *et al.*, 2014; Zhang *et al.*, 2015*b*). In subsequent study, DNA aptamers containing Z and P were generated with high specificity to mammalian cells overexpressing glypican 3, a known biomarker for liver cancer (Zhang *et al.*, 2016).

Similarly, Romesberg and Hirao also developed second generation unnatural base pairs that faithfully replicate using natural DNA polymerases (Malyshev and Romesberg, 2015; Hamashima *et al.*, 2018). For example, the hydrophobic TPT3:NAM base pair (Fig. 8*b*) generated by Romesberg and coworkers achieves 99.98% selectivity per doubling by PCR using OneTaq DNA polymerase (Li *et al.*, 2014), and the hydrophobic Ds:Px

https://doi.org/10.1017/S0033583520000050 Published online by Cambridge University Press

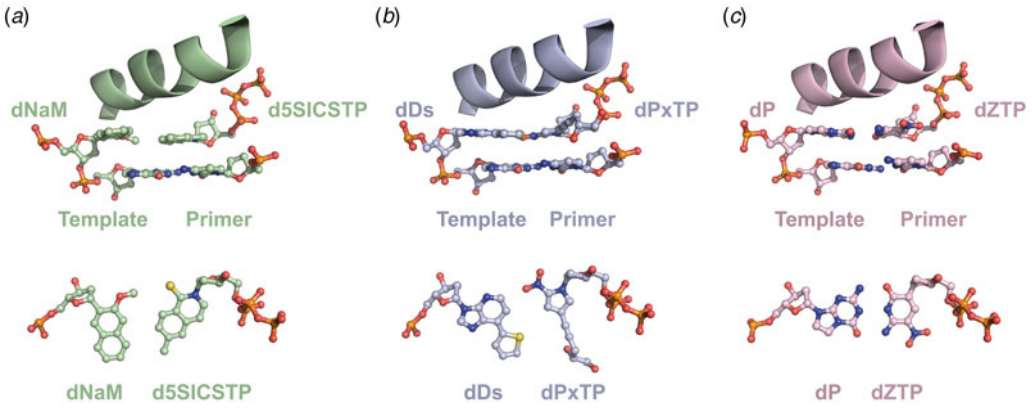


Fig. 10. Crystal structures of KlenTaq ternary complexes with unnatural Watson–Crick base pairs. Closed ternary structures of KlenTaq complexed with (*a*) NaM:5SICSTP (3SV3), (*b*) Ds:PxTP (5NKL), and (*c*) P:ZTP (5W6K).

base pair (Fig. 8*b*) produced by Hirao and colleagues achieves 99.97% selectivity per doubling by PCR using Deep Vent DNA polymerase (Okamoto *et al.*, 2016). The Ds:Px base pair was used to evolve high affinity DNA aptamers containing five genetic letters (A,C,G,T,Ds) to the protein targets vascular endothelial growth factor 165 ($VEGF_{165}$), interferon-γ (INFγ), and von Willebrand factor A1 domain (vWF) (Kimoto *et al.*, 2013; Matsunaga *et al.*, 2017). The increased chemical diversity of these libraries led to the production of aptamers with significantly higher affinity for their targets than comparable libraries using only natural bases. In subsequent study, the Ds-containing DNA aptamers were shown to inhibit $VEGF_{165}$ and INFγ binding to their cognate cellular receptors (Matsunaga *et al.*, 2015; Kimoto *et al.*, 2016), which advances the use of aptamers as synthetic affinity reagents.

Given the propensity for natural polymerases to replicate unnatural base pairs, structural studies were undertaken to compare the geometry of unnatural base pairs to those found in nature. Three different ternary structures have now been solved with an unnatural base pair occupying the insertion site of a KlenTaq DNA polymerase. The examples (Fig. 10) feature the unnatural base pairs of NaM–5SICS, Ds–Px, and P–Z in which the nucleotides NaM, Ds, and P occupy the templating position and 5SICS, Px, and Z are the incoming substrates, respectively (Betz *et al.*, 2012, 2017; Singh *et al.*, 2018). The collection of structures shows the artificial base pairs adopting planar geometries that are structurally similar to natural base pairs. Interestingly, a solution structure of duplex DNA containing a NaM–5SICS base pair unconstrained by a DNA polymerase reveals an intercalated structure rather than the more normal coplanar structure with edge-on-edge packing (Malyshev *et al.*, 2010). Similar structures have also been witnessed for other hydrophobic base pairs (Brotschi *et al.*, 2001; Matsuda *et al.*, 2007; Wojciechowski and Leumann, 2011), indicating that the polymerase induces a Watson–Crick geometry required for DNA replication.

Testing hypotheses about polymerase recognition

Beyond the immediate implications of establishing new hydrophobic base pairs, the ability to construct synthetic analogs of natural bases provides a unique opportunity to test hypotheses about how polymerases recognize their substrates (Jung and Marx, 2005). In the mid-1990s, some of the first crystal structures of polymerases bound to their substrates were solved to high resolution (Pelletier *et al.*, 1994; Doublie *et al.*, 1998). These structures, which include Bst DNA polymerase (Kiefer *et al.*, 1998), a close structural analog of Klenow DNA polymerase, reveal the presence of hydrogen bonding interactions between polar side chains and hydrogen bond acceptor atoms (N3 of purines and O2 of pyrimidines) found on the minor groove side of A:T and G:C base pairs. The observation of these interactions in the enzyme active site suggested that minor groove hydrogen bonding is an important aspect of DNA substrate recognition. To test this hypothesis, DNA synthesis reactions were performed using hydrophobic bases that either contain or lack minor groove hydrogen-bonding acceptor atoms (Morales and Kool, 1999). The resulting data clearly show that minor groove hydrogen bonding is critical for base pair recognition. Moreover, these interactions are more prevalent at the nucleotide extension step than the nucleotide insertion step and are stronger for the growing primer strand than the templating strand (Morales and Kool, 1999). Interestingly, each of the second-generation unnatural base pairs described above (Z:P, 5SICS:NaM, and Ds:Px, see Fig. 8*b*) have hydrogen bond acceptor atoms on the minor groove side of the Watson–Crick base pair to facilitate polymerase recognition.

Recognizing chemical modifications made to nucleobase positions

Structure–activity studies indicate that thermophilic DNA polymerases exhibit broad tolerance for chemical modifications made to the C5 position of pyrimidines and the C7 position of 7-deazapurines (Fig. 11) (Jager and Famulok, 2004; Jager *et al.*, 2005; Hollenstein, 2012; Kielkowski *et al.*, 2014; Cahova *et al.*, 2016). Notable examples include the use of Kod and Vent DNA polymerases to evolve slow off-rate modified aptamers (SOMAmers) from diversity-enhancing libraries containing C5-modified deoxyuridine residues (Vaught *et al.*, 2010; Gawande *et al.*, 2017). This strategy led to the development of an array-based platform for monitoring protein levels in human serum (Ostroff *et al.*, 2010; Williams *et al.*, 2019). Interestingly, the ability to synthesize DNA strands with multiple consecutive modifications uncovered strong substrate preferences between thermophilic A- and B-family DNA polymerases. Famulok and coworkers, for example, found that archaeal B-family DNA polymerases are more accepting of base-modified nucleotides than

https://doi.org/10.1017/S0033583520000050 Published online by Cambridge University Press

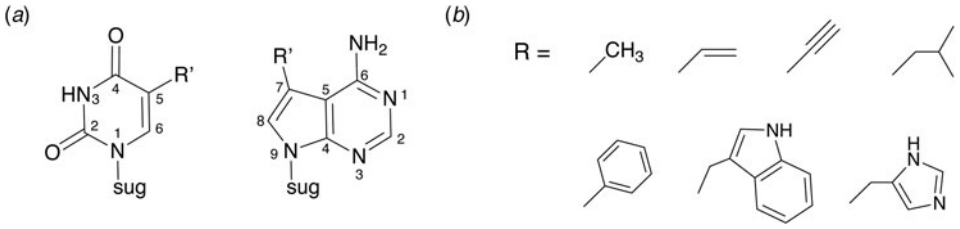


Fig. 11. Chemical structure of nucleobase-modified DNA. (*a*) Numbering of pyrimidine and purine ring aromatic systems. (*b*) Examples of common aliphatic and aromatic side chains.

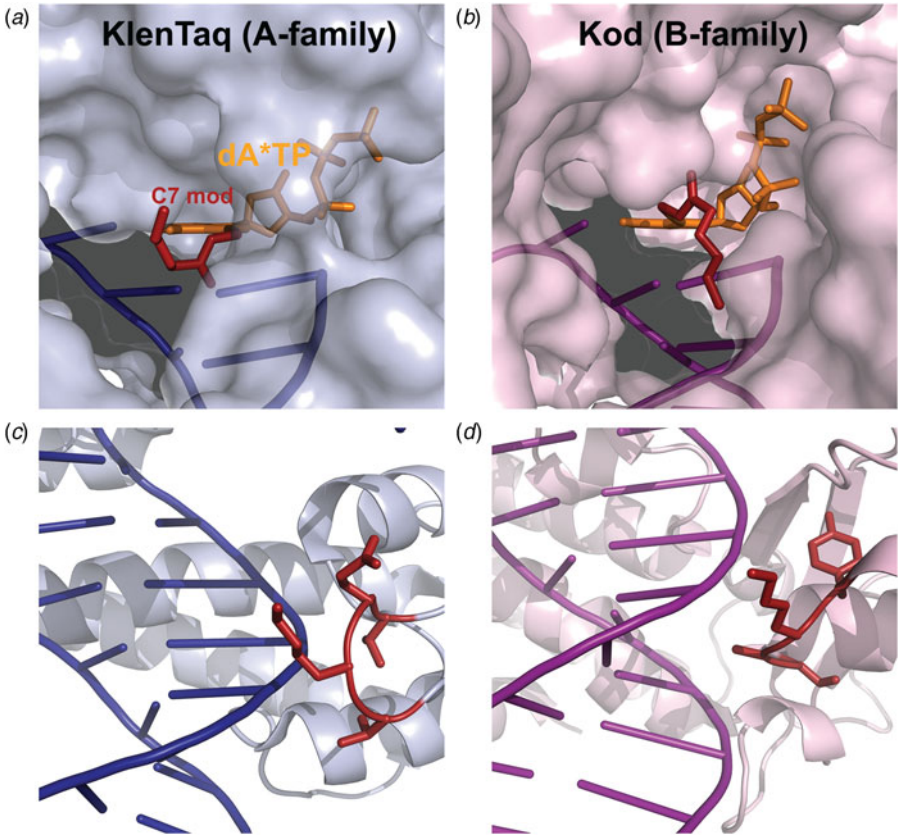


Fig. 12. Structural comparison of A- and B-family polymerases toward modified nucleotides. Surface representation of the closed ternary structures of (*a*) KlenTaq (6Q4U) and (*b*) Kod (6Q4T) in complex with a C7-modified dATP (denoted dA*TP) reveals a larger cavity in Kod. Consequently, the C7 modification (red) is well-resolved in KlenTaq but highly flexible in Kod. The thumb subdomains from (*c*) KlenTaq extends into the major groove, while the analogous region in (*d*) Kod interacts with the phosphodiester backbone.

thermophilic A-family DNA polymerases (Jager *et al.*, 2005). Sawai and colleagues made similar observations for C5-modified pyrimidines (Kuwahara *et al.*, 2006). Together, these observations suggest that A- and B-family polymerases have different structural constraints in the major groove region of the polymerase active site.

Marx and coworkers investigated the substrate specificity of A- and B-family DNA polymerases by solving high resolution crystal structures of KlenTaq and Kod DNA polymerases bound to natural and base-modified substrates (Bergen *et al.*, 2012, 2013; Kropp *et al.*, 2018; Kropp *et al.*, 2019). The structures indicate that bulky modifications pass through a cavity that extends outside the enzyme active site. This cavity enables members of both polymerase families to incorporate C5-modified pyrimidines and C7-modified purines into the growing DNA strand and to continue DNA synthesis afterward. Consistent with polymerase activity observed by Famulok and Sawai (Jager *et al.*, 2005; Kuwahara *et al.*, 2006), the cavity is larger and more accessible for Kod DNA polymerase than KlenTaq DNA polymerase (Fig. 12). In addition, the structures also show that substrate specificity is impacted by the location of the thumb subdomain. In the case of KlenTaq, the tip of the thumb (residues 506–509) extends into the major groove region of the DNA duplex, whereas the analogous region in Kod (residues 668–675) interacts with the phosphodiester backbone.

Propagation and evolution of an artificial genetic system

In a striking example of enzyme promiscuity, we recently discovered two naturally occurring DNA polymerases that will faithfully replicate 2′-fluoroarabino nucleic acid (FANA) (Wang *et al.*, 2018*b*), which is an unnatural genetic polymer that contains 2′-fluoroarabino residues in place of natural ribose or deoxyribose nucleotides (Damha *et al.*, 1998). Kinetic measurements collected using polymerase kinetic profiling (PKPro), a technique that monitors nucleotide synthesis using high-resolution fluorescent dyes that intercalate into the growing duplex (Nikoomanzar *et al.*, 2017), reveal that Tgo DNA polymerase catalyzes the synthesis of FANA polymers on DNA templates with a rate of ~15 nt min^{-1}, while Bst DNA

https://doi.org/10.1017/S0033583520000050 Published online by Cambridge University Press

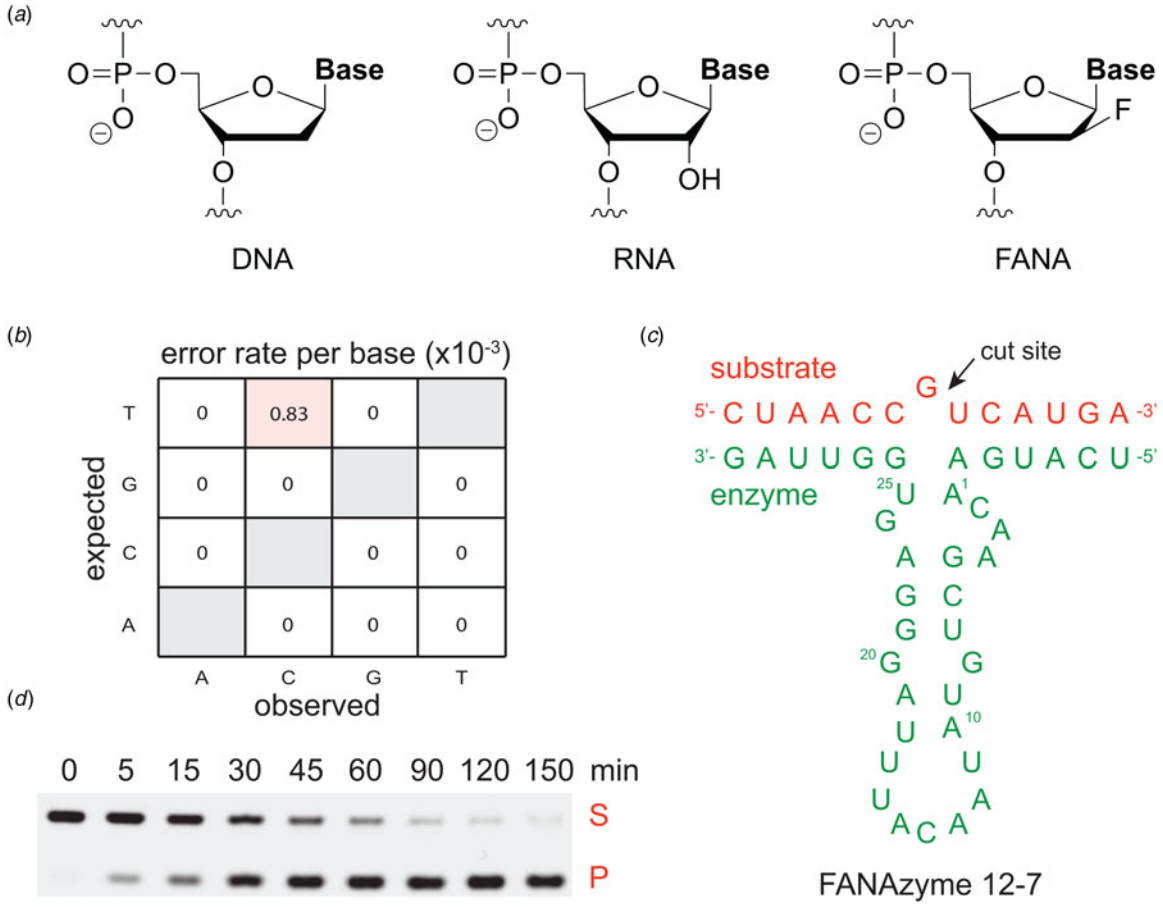


Fig. 13. A general RNA-cleaving FANA enzyme. (*a*) Molecular structures of DNA, RNA, and FANA. (*b*) Fidelity of FANA replication after a cycle of Tgo transcription and Bst reverse-transcription. (*c*) Predicted secondary structure showing the FANA enzyme (green) and RNA substrate (red) complex. (*d*) Time course of RNA cleavage by FANAzyme 12-7. S, RNA substrate; P, 5′-cleavage product. Adapted from Wang *et al.* (2018*b*).

polymerase promotes DNA synthesis on FANA templates with a rate of ~1 nt min^{-1} (Wang *et al.*, 2018*b*). The replication process occurs with a mutational rate of $\sim 8 \times 10^{-4}$ and an overall fidelity of 99.9% (Fig. 13*a*), making it the most faithful replication system for a xeno-nucleic acid (XNA) polymer (Chaput and Herdewijn, 2019).

An obvious application of the FANA replication system is the evolution of XNA aptamers and catalysts with enhanced nuclease resistance for diagnostic and therapeutic applications (Houlihan *et al.*, 2017). Toward this goal, an efficient RNA-cleaving FANA enzyme (FANAzyme, Fig. 13*b*) was generated that functions at a rate of $>10^6$-fold over the uncatalyzed reaction and achieves substrate saturation with Michaelis–Menten kinetics (Fig. 13*c*) (Wang *et al.*, 2018*b*). The enzyme comprises a small 25 nt catalytic domain that is flanked by substrate-binding arms that can be engineered to recognize diverse RNA targets. Divalent metal ion, pH profiles, and mass spectrometry analyses indicate that the reaction follows a metal and pH-dependent transesterification mechanism to produce an upstream cleavage product carrying a cyclic 2′,3′-monophosphate and a downstream strand with a 5′-OH group. In addition to expanding the chemical space of nucleic acid enzymes, this example provides a framework for evolving new types of FANA enzymes that can be generated using commercially available reagents, which is not the case for other XNA systems (Wang *et al.*, 2018*b*).

Structural insights into Bst DNA polymerase as an XNA reverse transcriptase

Bst DNA polymerase is unusual among naturally occurring replicative DNA polymerases, as it exhibits innate reverse transcriptase activity on nucleic acid templates of diverse chemical composition. Primer-extension studies reveal that Bst will copy templates composed of non-cognate RNA (Shi *et al.*, 2015), and the synthetic congeners of glycerol nucleic acid (Tsai *et al.*, 2007), FANA (Wang *et al.*, 2018*b*), and threose nucleic acid (TNA) (Dunn and Chaput, 2016), into full-length DNA products. We obtained the first structural insights into an enzyme with XNA reverse transcriptase activity by solving crystal structures of Bst DNA polymerase that capture the post-translocated product of DNA synthesis on templates composed entirely of FANA and TNA (Fig. 14) (Jackson *et al.*, 2019). Comparison of these structures with Bst DNA polymerase bound to the natural DNA primer–template duplex (Chim *et al.*, 2018) reveals differences, particularly at the enzyme active site as well as in protein interactions with the duplexes (Jackson *et al.*, 2019). The DNA/FANA and DNA/TNA duplexes within the active site adopt distinct conformations from the natural system (Fig. 14*a*), whereas the number of protein contacts to the phosphodiester backbone increase by 8 and 13, respectively, presumably to better position the template for DNA synthesis. Interestingly, despite strikingly different backbone conformations, both FANA and TNA adopt B-type

https://doi.org/10.1017/S0033583520000050 Published online by Cambridge University Press

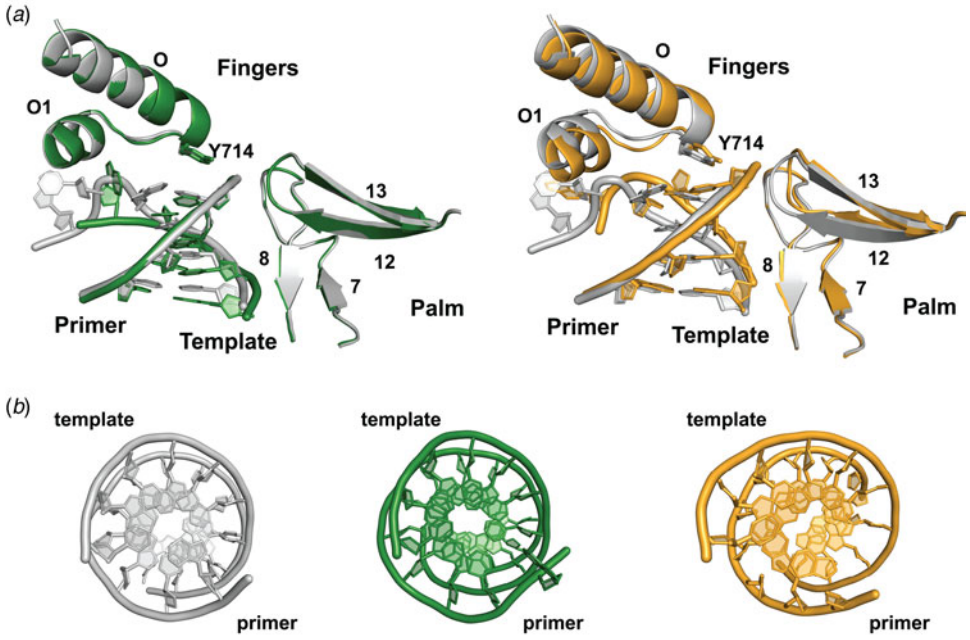


Fig. 14. X-ray crystal structures of natural Bst DNA polymerase exhibiting XNA reverse transcription activity. (*a*) Active site region with the polymerase bound to a DNA (gray), FANA (6MU4, green), and TNA (6MU5, orange) primer hybridized to a DNA template. Strands 7, 8, 12, and 13 comprise a portion of the antiparallel *β*-sheet of the palm subdomain. (*b*) Top-down view of bound the duplex structures. Adapted from Jackson *et al.* (2019).

helical structures when hybridized to DNA (Fig. 14*b*). Taken together, these data suggest the importance of structural plasticity as a possible mechanism for XNA-dependent DNA synthesis and offers preliminary rationale for designing variants with improved functional activity. However, it should be stressed that further structural studies are needed to fully understand how gain-of-function mutations are changing the active site conformation of engineering polymerases.

Engineering polymerase functions by rational design

Rational design has been used to discover new polymerase activities without resorting to molecular evolution. Early strategies utilized natural sequence diversity and residue or domain swapping to change the substrate specificity or biological stability of a polymerase. Structure-guided approaches have also been used to predict specific amino acid changes that would lead to a desired activity. Together, these strategies provide insight into the mechanism of DNA synthesis, the functional role of accessory domains, and the potential for new or improved activities to arise from natural sequence variation. The following section illustrates a number of cases where the deletion or transfer of residues between DNA polymerases leads to enhancements in enzyme performance (Table 2). Other cases, however, show the limitations of rational design and the need for more advanced approaches to enzyme engineering.

Structural permutations of natural DNA polymerases

In 1970, Klenow and Henningsen were the first to use limited proteolysis as a way to evaluate the mechanism of a DNA polymerase (Klenow and Henningsen, 1970). Using affinity chromatography to purify DNA polymerase I from crude cellular extracts of *E. coli* lysate, two distinct polymerase elution profiles emerged with different enzymatic properties and molecular weights. Although both enzymes retained their cognate polymerase and 3′-5′ exonuclease proofreading activities, only the larger enzyme (~150 kDa) exhibited 5′-3′ exonuclease activity. Speculating that the 5′-3′ exonuclease domain had been removed by proteolytic digestion, the larger DNA polymerase was treated with subtilisin to produce a smaller fragment (~70 kDa) with the same size and enzymatic properties observed for the smaller polymerase isolated by affinity chromatography. This version of DNA polymerase I, now commonly known as Klenow DNA polymerase, is routinely used to form blunt ended DNA by filling in 5′ overhangs and removing 3′ overhangs (Sambrook *et al.*, 1989) and for second strand cDNA synthesis after reverse transcription of RNA back into DNA (Gubler, 1987). Klenow DNA polymerase holds a special place among synthetic biologists, as it was the first DNA polymerase used to replicate an unnatural base pair in DNA (Switzer *et al.*, 1989).

Following the invention of PCR (Saiki *et al.*, 1988), a significant effort was made to improve the isolation of Taq DNA polymerase by recombinant protein expression in *E. coli* so that the enzyme could be used as a tool for molecular biology research. In addition to optimizing the promoter sequence (Lawyer *et al.*, 1989; Engelke *et al.*, 1990), researchers sought to increase protein expression levels by truncating the enzyme. In two separate cases, shorter versions of Taq DNA polymerase (94 kDa) were produced by removing segments of the gene encoding the 5′-3′ exonuclease domain (Fig. 15). The first example was a 705 bp 5′-truncation that yielded a 67 kDa variant called KlenTaq (67 kDa), which is the Klenow-fragment analog of Taq DNA polymerase (Barnes, 1992). The second example was a truncation that removed the first 867 bp region to yield a 61 kDa derivative known as Stoffel (Lawyer *et al.*, 1993). Although full-length Taq DNA polymerase is widely used in quantitative real-time PCR applications due to its 5′-3′ exonuclease activity, the smaller KlenTaq and Stoffel polymerases are often used to amplify DNA containing modified nucleotides

https://doi.org/10.1017/S0033583520000050 Published online by Cambridge University Press

Table 2. Engineered polymerases and applications

Engineered polymerase	Application	Reference
Pfu + Sso7d	Improved enzyme processivity and PCR performance	Wang *et al.* (2004)
Stoffel + Sso7d	Improved enzyme processivity and PCR performance	Wang *et al.* (2004)
Taq + Sso7d	Improved enzyme processivity and PCR performance	Wang *et al.* (2004)
Stoffel + HhH	Increased salt tolerance	Pavlov *et al.* (2002)
Pfu + HhH	Increased salt tolerance	Pavlov *et al.* (2002)
'Helimerase'	Couples the activities of polymerase and helicase from noncovalent coiled-coil protein interactions for improved HDA performance	Motre *et al.* (2008)
T7 DNA Pol + Y526F	Increases discrimination against ddNTPs by >2000-fold	Tabor and Richardson (1995)
Tgo-QGLK	Increased RNA synthesis up to 1.2 kb length	Cozens *et al.* (2012)
Tgo C7	LNA/CeNA synthesis	Pinheiro *et al.* (2012)
Tgo D4K	FANA/ANA synthesis	Pinheiro *et al.* (2012)
Tgo RT521	TNA synthesis/HNA RT/FANA RT/TNA RT/ANA RT	Pinheiro *et al.* (2012)
Tgo RT521K	LNA RT/CeNA RT	Pinheiro *et al.* (2012)
Pol6G12	HNA synthesis	Pinheiro *et al.* (2012)
Tgo-T	XNA polymerase backbone	Pinheiro *et al.* (2012)
Taq + LVL motif	Increased discrimination against mismatch extension	Summerer *et al.* (2005)
Taq + E708K,L,W	Increased resistance to inhibitors	Kermekchiev *et al.* (2009)
KlenTaq1 + E708K,L,W	Increased resistance to inhibitors	Kermekchiev *et al.* (2009)
Taq + E742/A743 mutations	Improved PCR performance	Yamagami *et al.* (2014)
Pol 5D4	Increased recognition of hydrophobic base analogs and concomitant ability to PCR amplify DNA containing the bisulfate intermediate 5,6-dihydrouracil 6-sulfonate (dhU6S)	Loakes *et al.* (2009), Millar *et al.* (2015)
Taq + I614K	Error-prone polymerase with >20-fold decreased fidelity as compared to wild-type Taq and can bypass damaged or abasic sites in template DNA	Patel *et al.* (2001)
SM19	Improved activity for 2′-methoxy (OCH_3) nucleoside triphosphates, 2′ fluoro (F), 2′ azido (N_3), and 2′ amino (NH_2) modifications	Schultz *et al.* (2015)
SM4-9	Ability to PCR amplify DNA containing the 2′ OCH_3 and 2′ F modifications on pyrimidine residues	Chen *et al.* (2016)
RTX	Proofreading reverse transcriptase with ultrahigh thermostability	Ellefson *et al.* (2016)
Kod-RI	1st generation manganese independent TNA polymerase	Larsen *et al.* (2016)
Kod-RS	2nd generation TNA polymerase with nearly equal substrate specificities of DNA *versus* TNA	Nikoomanzar *et al.* (2019)
Kod-RSGA	3rd generation TNA polymerase with improved fidelity, template sequence bias, and DNA *versus* TNA specificity ratio	Nikoomanzar *et al.* (2020)
Therminator + biotinylated peptide legs	Increased Therminator processivity from <20 per binding event to several thousands	Williams *et al.* (2008)
Therminator (NEB)	Increased ability to tolerate a wide variety of modified nucleotide substrates	Gardner *et al.* (2019)
OneTaq (NEB)	Optimized blend of natural Taq and deep vent polymerase for improved PCR performance and fidelity	Li *et al.* (2014)
Pol 2D9	Increased resistance to inhibitors	Baar *et al.* (2011)
T8	Increased thermal stability Taq variant with 11-fold greater half-life at 97.5 °C	Ghadessy *et al.* (2001)
H15	130-fold increased resistance to heparin Taq variant	Ghadessy *et al.* (2001)
3A10	Increased ability to extend mismatches, process non-canonical templates, and bypass lesions found in ancient DNA	d'Abbadie *et al.* (2007)
3D1	Increased ability to extend mismatches, process non-canonical templates, and bypass lesions found in ancient DNA	d'Abbadie *et al.* (2007)
M1	Increased ability to extend mismatches	Ghadessy *et al.* (2004)

(*Continued*)

https://doi.org/10.1017/S0033583520000050 Published online by Cambridge University Press

Table 2. (*Continued.*)

Engineered polymerase	Application	Reference
M4	Increased ability to extend mismatches	Ghadessy *et al.* (2004)
'Split finger' Pfu	Improved incorporation of a bulky γ-phosphate-*O*-linked-dabcyl nucleotide	Hansen *et al.* (2011)
AA40	Taq variant with DNA, RNA, and RT activity capable of producing mixed DNA–RNA PCR amplification products	Ong *et al.* (2006)
E10	Cy3 or Cy5 dC incorporation for creating fluorescent PCR products	Ramsay *et al.* (2010)
SFM4-3	PCR amplifies substrates with 2′ F and 2′ azido (N_3) groups	Chen *et al.* (2016)
SFM4-6	Ability to PCR amplify fully 2′-OCH_3 modified aptamers to HNE	Liu *et al.* (2017)
SFM4-9	Ability to PCR amplify fully 2′-OCH_3 modified aptamers to HNE	Liu *et al.* (2017)

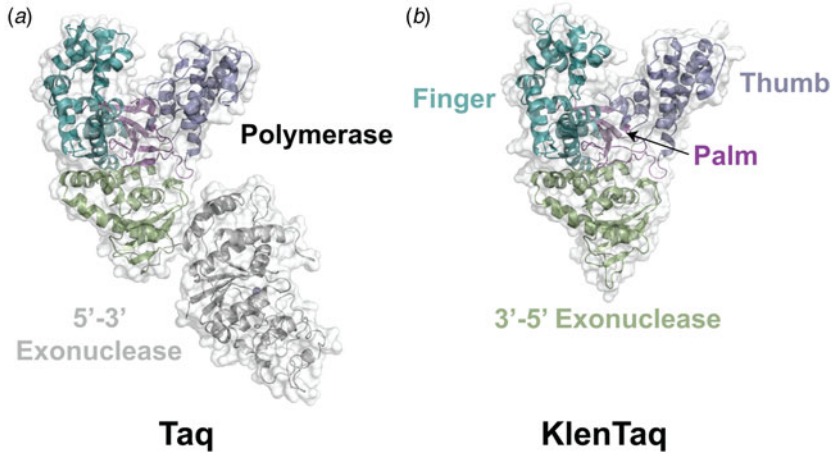


Fig. 15. Structural variants of Taq DNA polymerase. Crystal structures of (*a*) Taq (1TAQ) and (*b*) KlenTaq (1KTQ) DNA polymerase. KlenTaq is the Klenow DNA polymerase analog of Taq DNA polymerase in which the 5′-3′ exonuclease domain has been removed.

and as starting points for directed evolution (Malyshev *et al.*, 2009; Yamashige *et al.*, 2012). In recent years, KlenTaq has become a favorite polymerase among X-ray crystallographers wishing to capture the structures of DNA polymerases synthesizing non-cognate and synthetic congeners of natural nucleotides (Betz *et al.*, 2012, 2017; Singh *et al.*, 2018).

Exonuclease silencing

The 3′-5′ exonuclease proofreading activity associated with many DNA polymerases is designed to correct single-nucleotide mismatches that occur during the course of normal DNA synthesis. Mutations that silence this activity are often advantageous for synthetic biology applications that require polymerases to incorporate unnatural nucleotides into the growing strand. In the absence of these mutations, modified nucleotides are often difficult to incorporate as the rate of nucleotide addition must compete with the rate of DNA editing. Early attempts at silencing the 3′-5′ exonuclease domains led to the surprising discovery that certain exonuclease-silent (exo-) polymerases can function with enhanced activity. Tabor and Richardson, for example, discovered that T7 DNA polymerase (exo-) functions with ~10-fold higher activity than natural T7 DNA polymerase, which enables the enzyme to read through difficult hairpins (Tabor and Richardson, 1989*b*). Similar activity silencing mutations led to the production of a Bst DNA polymerase variant that functions with elevated thermal stability (Riggs *et al.*, 1996).

Accelerating DNA synthesis with non-specific DNA-binding domains

Improving the performance of thermophilic DNA polymerases that are capable of PCR amplification was an important early goal in molecular biology. Efforts to study this problem led to the realization that replicative DNA polymerases often use complicated mechanisms that cannot be applied in a general way to *in vitro* assays. For example, many replicative DNA polymerases rely on accessory proteins, such as thioredoxin (Das and Fujimura, 1979) or ring-shaped protein complexes that make up the 'sliding clamp' (Baker and Bell, 1998), which are highly specific to individual polymerases. One exception is the double-stranded DNA-binding protein Sso7d isolated from *Sulfolobus solfataricus*, which provides general enzyme enhancing activity when fused to standard DNA polymerases (Wang *et al.*, 2004). Examples where DNA polymerases have been fused to the Sso7d DNA-binding domain include Taq and Stoffel (both A-family members) and Pfu, a hyperthermophilic archaeal B-family DNA polymerase isolated from *P. furiosus*. Activity assays show a ~5–20-fold increase in processivity for the three enzymes tested with the greatest increase observed for Stoffel (2.9 nt *versus* 51 nt per binding event). Importantly, addition of the Sso7d DNA-binding domain to the polymerase did not alter the catalytic properties of the enzyme, which is critical for high-fidelity DNA synthesis. Polymerases engineered with the Sso7d domain reduced the cycle times required for DNA amplification, generated amplicons of increased length, and provided

https://doi.org/10.1017/S0033583520000050 Published online by Cambridge University Press

increased tolerance against salt inhibition. Phusion DNA polymerase is an example of a DNA polymerase (Pfu fused to Sso7D) that was engineered for rapid, high-fidelity DNA synthesis of long amplicons.

Helix–hairpin–helix (HhH) motifs found in DNA modifying enzymes, including nucleases, ligases, polymerases, and helicases are a second example where a general DNA-binding motif has been used to enhance the activity of a DNA polymerase. In nature, two-thirds of DNA topoisomerase V is comprised of HhH motifs (Slesarev *et al.*, 1993). When these motifs are removed from the enzyme, topoisomerase retains activity but is more sensitive to salt inhibition than the full-length version. Guided by this observation, variants of Stoffel and Pfu DNA polymerases were constructed with HhH motifs fused to their N- and C-terminal regions (Pavlov *et al.*, 2002). The engineered polymerases exhibit increased resistance to inhibition under high salt conditions with more HhH motifs providing greater protection. As an example, an engineered Stoffel polymerase remains active in the presence of 250 mM NaCl, whereas the natural polymerase was found to be completely inactive. One hypothesis drawn from this result is that the lack of enzymatic activity observed under high salt conditions is not due to the presence of monovalent ions interacting with the enzyme active site, but rather an inability to form the complex needed for DNA synthesis.

The addition of helicase to the reaction mixture represents a third approach for improving polymerase activity. Helicase-dependent amplification (HDA) uses the energy from ATP and helicase to produce a single-stranded template that can be copied under ambient conditions (Vincent *et al.*, 2004). As such, HDA has become an attractive technique for point of care diagnostics that require minimal instrumentation. Versions of HDA have been performed using UvrD helicase, Klenow DNA polymerase, MutL, and single-stranded binding protein (An *et al.*, 2005). Thermophilic versions of this technique do not require accessory proteins but are limited to short amplicons of only 200 bp in length (An *et al.*, 2005). To produce longer amplicons, a non-covalent system termed 'helimerase' was developed that relies on a coiled-coil motif to synchronize the activities of the helicase and polymerase (Motre *et al.*, 2008). The complex forms *in vitro* as well as *in vivo* and can be used to produce amplicons that exceed 1 kb in length.

Determinants of sugar recognition

Natural DNA polymerases are significantly less tolerant toward chemical modifications made to the sugar moiety than the nucleobase. One of the few early reports on sugar recognition is an acyclic peptide nucleic acid derivative that functions as a chain terminator of DNA synthesis (Martinez *et al.*, 1997). However, despite facile preparation, this analog is less efficient compared to 2′,3′-dideoxyribonucleoside triphosphates, which is the standard reagent set used for Sanger sequencing (Sanger *et al.*, 1977). A slightly different example is C4′-acylated thymidine triphosphates developed to study DNA strand repair (Marx *et al.*, 1997, 1999). Other prominent examples where sugar modifications have been evaluated in DNA polymerase reactions include the recognition of: 2′,5′-isomeric DNA by Klenow and HIV RT (Sinha *et al.*, 2004); acyclic nucleotides by Vent (Gardner *et al.*, 2004); glucose nucleotides by Vent (Renders *et al.*, 2009); flexible nucleic acids by Klenow (Heuberger and Switzer, 2008); locked nucleic acid (LNA) by Superscript III (Crouzier *et al.*, 2012); cyclohexynyl nucleic acid (CeNA) by HIV RT and Vent (Kempeneers *et al.*, 2005); hexose nucleic acid (HNA) by Klenow and Taq (Pochet *et al.*, 2003); and TNA by Superscript II and MMLV RT (Chaput and Szostak, 2003; Chaput *et al.*, 2003). However, the activity observed with these substrates is significantly less than the wild type activity observed with natural substrates. In some cases, manganese ions are used to loosen the enzyme active site, which is a common technique for increasing the tolerance of a DNA polymerase for unnatural nucleoside triphosphates (Dube and Loeb, 1975; Tabor and Richardson, 1989*a*). However, as noted previously, supplementing the reaction with manganese ions often leads to higher rates of nucleotide misincorporation.

Tabor and Richardson were among the first to explore the determinants of substrate specificity by a DNA polymerase (Tabor and Richardson, 1995). Recognizing that bacteriophage T7 DNA polymerase incorporates chain terminating ddNTPs into DNA more efficiently than DNA polymerases from *E. coli* and Taq, polymerases bearing hybrid sequences in the enzyme active site were constructed and tested to determine the molecular basis of substrate specificity. The mutational study uncovered a single hydroxyl group on Tyr526 that was responsible for the observed substrate specificity. Substitution of Tyr526 in T7 DNA polymerase with phenylalanine increases the discrimination against ddNTPs by >2000-fold, while replacing the analogous Phe residue in either *E. coli* DNA polymerase I or Taq DNA polymerase with Tyr decreases discrimination against ddNTPs up to 8000-fold. Since *E. coli* DNA polymerase I binds ddTTP and dTTP with equal affinity, the source of discrimination likely occurs at a subsequent step in the catalytic cycle.

Related studies on Vent DNA polymerase (exo-) isolated from *Thermococcus litoralis* demonstrated that mutating the active site residue Ala488 to a larger side chain increases the incorporation of sugar-modified nucleotides, including ddNTPs, NTPs, and 3′-dNTPs (Cordycepin) (Gardner and Jack, 1999). The pattern of relaxed specificity at this position roughly correlates with the size of the amino acid substitution with larger residues showing a higher tolerance for sugar-modified substrates. Similar effects were observed when the Vent Ala488 mutation was transferred to other archaeal DNA polymerases, including Pfu (exo-) (Evans *et al.*, 2000) and Kod (exo-) (Hoshino *et al.*, 2016). Addition of the Vent A488L mutation to 9°N produced a commercial enzyme known as Therminator polymerase, which found early widespread use as a research tool for DNA sequencing using acyclic nucleotide analogs (Gardner and Jack, 2002).

Since its discovery, Therminator DNA polymerase has become the most widely studied and experimentally utilized engineered polymerase for synthesizing modified nucleotides (Gardner *et al.*, 2019). Derived from a hyperthermophilic euryarcheon *Thermococcus* sp. 9°N, this B-family polymerase carries an A485L mutation in the O-helix of the finger subdomain along with the 3′-5′ exonuclease silencing mutations D141A and E143A. Despite the fact that position 485 faces away from the polymerase active site and does not directly interact with the incoming nucleoside triphosphate, this mutation imparts strong gain-of-function activity for a wide variety of sugar, base, and backbone modified substrates (Bergen *et al.*, 2013; Kropp *et al.*, 2017). This observation is thought to be due to a change in the dynamics between the open and closed state of the fingers, which increases the occupancy of the closed conformation necessary for chemical catalysis. This relatively straightforward mechanism could explain the ability for Therminator to accept a broad range of substrates, including noncognate substrates (NTPs)

https://doi.org/10.1017/S0033583520000050 Published online by Cambridge University Press

(McCullum and Chaput, 2009), base-modified substrates (dN*TPs), sugar-modified substrates (xNTPs) (Lapa *et al.*, 2016), unnatural base pair substrates (dXTP and dYTP) (Hwang and Romesberg, 2008), and terminator substrates (ddNTP, acyclic, 3′-blocked dNTP, and lightning terminators) (Gardner and Jack, 2002; Litosh *et al.*, 2011).

Recognizing the importance of the A485L mutation as a critical determinant of substrate specificity, a significant effort has been made to further improve the activity of this mutation through rational design (Gardner *et al.*, 2019). In the case of RNA synthesis, combining the A485L mutation with Y409G and E664K, the steric gate and so-called second steric-gate, respectively, enabled Tgo DNA polymerase to synthesize RNA strands up to 1.7 kb in length (Cozens *et al.*, 2012). The attachment of two biotinylated 'peptide legs' to Therminator led to a polymerase complex with streptavidin that increased the processivity of DNA synthesis from less than 20 nucleotides to several thousand nucleotides per binding event (Williams *et al.*, 2008). The A485L mutation has also been used to improve XNA synthesis wherein an engineered polymerase named TgoT (V93Q, D141A, E143A, and A485L) provided the backbone for generating new polymerase variants that can synthesize a variety of artificial nucleic acids, including CeNA, ANA, FANA, HNA, TNA, and LNA (Pinheiro *et al.*, 2012). In the case of next-generation sequencing (NGS), Therminator was used as the starting point for generating a polymerase that facilitates the synthesis of fluorescently-tagged nucleotides (Gardner *et al.*, 2012). If the past is any indication of the future, it would seem likely that the next generation of engineered polymerases will benefit from further exploration of the Therminator position.

Improving DNA polymerase performance for PCR

PCR has had a major impact on molecular biology by providing a simple method for amplifying DNA (Saiki *et al.*, 1988). Early experiments required fresh polymerase to be added during each extension cycle due to the high temperatures >95 °C required for denaturing the DNA strands prior to the start of another round of synthesis. This arduous task greatly reduced the speed of amplification, as it not only required the physical presence of a researcher to add new enzyme but also lowered the theoretical limit of DNA replication due to the presence of increasing quantities of inactive enzyme. A solution to this problem came when a thermophilic DNA polymerase was isolated from the bacterium species *T. aquaticus* (Chien *et al.*, 1976). Taq DNA polymerase was harnessed for its intrinsic thermal stability, which allows for uninterrupted cycles of DNA replication. PCR has since found widespread use in DNA cloning, NGS, criminal forensics, molecular diagnostics, epigenetic mapping, and pathogen detection (Garibyan and Avashia, 2013). However, as the demand for PCR amplification has grown, so has the need for new variants that can function under more demanding conditions.

Genotyping biological samples require precise DNA amplification to distinguish single-nucleotide polymorphisms from random mutations. Recognizing that motif C in A- and B-family DNA polymerases may contribute to mismatch extension through indirect H-bonding between the minor groove and a histidine side chain (Franklin *et al.*, 2001), Marx and coworkers applied a structure-guided approach to identify variants of Taq DNA polymerase that function with increased fidelity. An automated fluorescent screen was established to evaluate 1316 variants of Klenow DNA polymerase (exo-) bearing mutations at positions 879–881 (Summerer *et al.*, 2005). Protein expression was conducted in 96-well plates and crude lysate was queried for activity in 384-well format. Fidelity values were assigned based on the ratio of extension from primers containing matched and mismatched 3′-terminal residues. A Klenow variant with LVL at positions 879–881 exhibited strong kinetic discrimination against mismatch extension. Transferring the LVL mutations to analogous positions in wild-type Taq DNA polymerase produced an engineered version of Taq DNA polymerase with increased discrimination against transitions and transversions (Summerer *et al.*, 2005).

Taq DNA polymerase is readily inactivated by hemoglobin and humic acid present in blood and soil samples used for DNA analysis. Surprisingly, Klentaq1, a truncated version of Taq DNA polymerase with a 278 aa N-terminal deletion, can amplify single-copy genomic DNA in the presence of 5–10% whole blood (Abu Al-Soud and Radstrom, 1998; Abu Al-Soud and Radstrom, 2000). To generate an enzyme with improved activity, Barnes and coworkers screened a library of 40 arbitrary but functional variants with mutations at positions 626 and 706–708 for improved PCR performance under increasing amounts of whole blood (Kermekchiev *et al.*, 2009). The screen revealed that mutation of E708 to K, L, or W resulted in enhanced resistance to various inhibitors, including plasma, hemoglobin, lactoferrin, serum IgG, soil extracts, and humic acid. The resulting polymerase facilitates the amplification of single-copy human genomic targets from whole blood, which eliminates the need for a sample treatment step.

Archaeal B-family DNA polymerases are widely used enzymes for PCR because of their high thermal stability and presence of a strong 3′-5′ proofreading exonuclease domain. However, despite high sequence and structural homology, the *Thermococcales* order of archaeal DNA polymerases exhibits strikingly different kinetic properties that affect their PCR performance. Kod DNA polymerase, for example, possesses higher processivity (defined as the number of dNTP incorporations per binding event) than its related homologs but is 10 °C less stable (83 *versus* 93 °C) than Pfu DNA polymerase, which limits its utility as an enzyme for PCR. To improve the processivity of Pfu DNA polymerase, Connolly and coworkers transferred residues from the forked-point (polymerase junction between the template-binding and editing cleft consisting of seven arginine residues) and entire thumb regions of Kod DNA polymerase to Pfu (Elshawadfy *et al.*, 2014). The resulting polymerase with the combined forked-point and thumb regions from Kod DNA polymerase retained the high thermal stability of Pfu while gaining an increased capacity for PCR performance.

Similar efforts to explore the natural diversity of DNA polymerases were performed by recombining gene fragments of A-family DNA polymerases taken from soil samples of microorganisms found near thermal hot springs (Yamagami *et al.*, 2014). Corresponding regions of the *pol* gene for Taq DNA polymerase were substituted with the amplified gene fragments and the chimeric variants were tested for activity. Biochemical analysis led to the identification of two mutations, E742R and A743R, that impart higher DNA-binding affinity and faster primer extension activity on Taq DNA polymerase. Both factors resulted in improved PCR performance, suggesting that natural diversity is a promising strategy for finding new amino acid positions with strong gain of function activity.

The ability to sequence epigenetic modifications is an important goal of genomic research. Of all possible epigenetic

https://doi.org/10.1017/S0033583520000050 Published online by Cambridge University Press

modifications, none is more prevalent than 5-methylcytosine (5mC). This subtle chemical change has far reaching implications for normal cellular growth and development as well as several neurological diseases and cancer (Allis and Jenuwein, 2016). Bisulfite treatment converts natural cytosine bases to a 5,6-dihydrouracil 6-sulfonate (dhU6S) intermediate that is subsequently hydrolyzed to deoxyuracil (dU). Because 5mC is resistant to bisulfite treatment, this approach can be used to identify 5mC epigenetic markers by mapping the conversion of bases that are read as dC before and after bisulfite treatment. Unfortunately, this approach leads to significant degradation of the genomic DNA sample, which hampers genome-wide association studies. Holliger and coworkers recently discovered that the engineered polymerase 5D4, previously developed to recognize hydrophobic base analogs, is able to amplify DNA carrying the bisulfite intermediate (Millar *et al.*, 2015). This discovery greatly improves the workflow and sensitivity of 5mC detection in genomic DNA samples.

Mutagenic DNA polymerases that function with low fidelity have value as reagents for creating degenerate libraries for directed evolution studies and offer clues into the mechanistic underpinnings of substrate recognition during DNA synthesis. To investigate this phenomenon, Loeb and coworkers created a library of ~ 200 000 mutant Taq DNA polymerase variants comprising random mutations in the dNTP binding pocket of motif A (residues 605–617) (Patel *et al.*, 2001). The library was screened for activity using a temperature-sensitive complementation assay in *E. coli* and a subset of active variants were tested for fidelity using a mismatch primer extension assay. Taq polymerase variants with strong mismatch extension activity each contain substitutions at I614, indicating that a single, highly mutable, active amino acid is critical for DNA polymerase fidelity. A Taq DNA polymerase variant bearing the I614K mutation was shown to function with a 20-fold higher error rate than wild-type Taq DNA polymerase and can bypass damaged and abasic sites in DNA templates. This example provides an approach for producing polymerases that function with error-prone activity during PCR.

Engineering polymerases by directed evolution

In the last 20 years, the field of polymerase engineering has benefited from the growth of new technologies that make it possible to generate custom polymerases by directed evolution. Whether searching designer libraries that carefully sample all possible single-point mutations at defined positions or less sophisticated libraries that contain random mutations at unknown positions, the technologies available today allow users to rapidly search large combinatorial libraries ($>10^7$ unique members) in timeframes ranging from days to weeks. These efforts have been aided by the development of clever strategies for establishing genotype–phenotype linkages that make it possible to determine the sequence of active variants with valuable gain-of-function mutations. The most common approaches perform the activity step *in vitro*, which allows for greater control over the reaction conditions and substrate chemistries, including the use of synthetic congeners that bear little or no resemblance to natural nucleotides. In addition to establishing new enzymes with practical applications in biotechnology and medicine, these studies also provide a wealth of information about how polymerases function. As these studies continue, sufficient knowledge may be gained that will enable future generations to one day bypass the need for directed evolution and allow computational methods to predict individual sequences with desired activities. However, realizing these dreams will require a greater understanding of the determinants that govern substrate specificity, which is a major goal of most polymerase-engineering efforts.

Phage display

Phage display is one of the oldest and most successful methods for evolving peptides and proteins with ligand binding activity (Smith and Petrenko, 1997). With this technique, a gene encoding a protein of interest is inserted into a phage coat protein gene, which causes the phage to display the protein on its outside surface while retaining the encoding genetic information inside the bacteriophage. A modified version of phage display was originally developed by Jestin, and subsequently refined by Romesberg, to facilitate the evolution of polymerases with new activities (Jestin *et al.*, 1999; Xia *et al.*, 2002). In this method, phage particles are engineered to display the DNA primer–template duplex and polymerase variant in close proximity. The polymerase library is expressed as an N-terminal fusion of the minor M13 phage coat protein pIII in such a way that the phage surface contains one copy of the polymerase and four copies of a short acidic peptide. Separately, a complementary basic peptide is conjugated to the DNA primer, annealed to a DNA template, and combined with the phage particle to form a coiled-coil linking the DNA primer–template duplex to the phage surface. Activity screens are then performed in-*cis* by enriching for polymerase variants that can incorporate a biotin-tagged nucleotide into the growing DNA strand, which is used to capture the phage particle on streptavidin-coated beads (Fig. 16). The beads are washed to remove inactive variants and the genes encoding functional polymerases are recovered by eluting the bacteriophage with DNase I. The population of enriched phage particles is then amplified by infecting a fresh *E. coli* culture. Recently, the technique has been improved by incorporating *p*-azidophenylalanine into the pIII protein, which allows for an alkynyl-modified primer–template duplex to be conjugated to the phage surface using click-chemistry (Chen *et al.*, 2016). The revised protocol avoids the need to synthesize and purify peptide–DNA conjugates comprising the basic peptide and DNA primer.

Phage display was used by Jestin and coworkers to evolve a population of Taq DNA polymerase variants with thermostable reverse transcriptase activity (Vichier-Guerre *et al.*, 2006). Romesberg and coworkers have used this technique to identify variants of the Stoffel fragment (SF) of Taq DNA polymerase that function with improved activity for ribonucleoside triphosphates (Xia *et al.*, 2002), 2′-methoxy (OCH_3) nucleoside triphosphates (Fa *et al.*, 2004), and the unnatural PICS:PICS self-pair (Leconte *et al.*, 2005). Further characterization of the polymerase with 2′ OCH_3 activity revealed that this variant (SM19) could also recognize substrates with 2′-fluoro (F), 2′-azido (N_3), and 2′-amino (NH_2) modifications (Schultz *et al.*, 2015). Using the click-chemistry version of phage display, SM19 was evolved to yield SM4-9, which is a thermostable polymerase able to PCR amplify DNA containing the 2′-OCH_3 and 2′-F modifications on pyrimidine residues (Chen *et al.*, 2016).

The major benefit of the phage display approach is the ability to detect a single nucleotide incorporation event using biotinylated substrates. Anticipated weaknesses include complications of phage particle assembly, the potential for low multiple turnover activity caused by the in-*cis* selection strategy, and the possibility for high background due to non-specific binding to the solid

https://doi.org/10.1017/S0033583520000050 Published online by Cambridge University Press

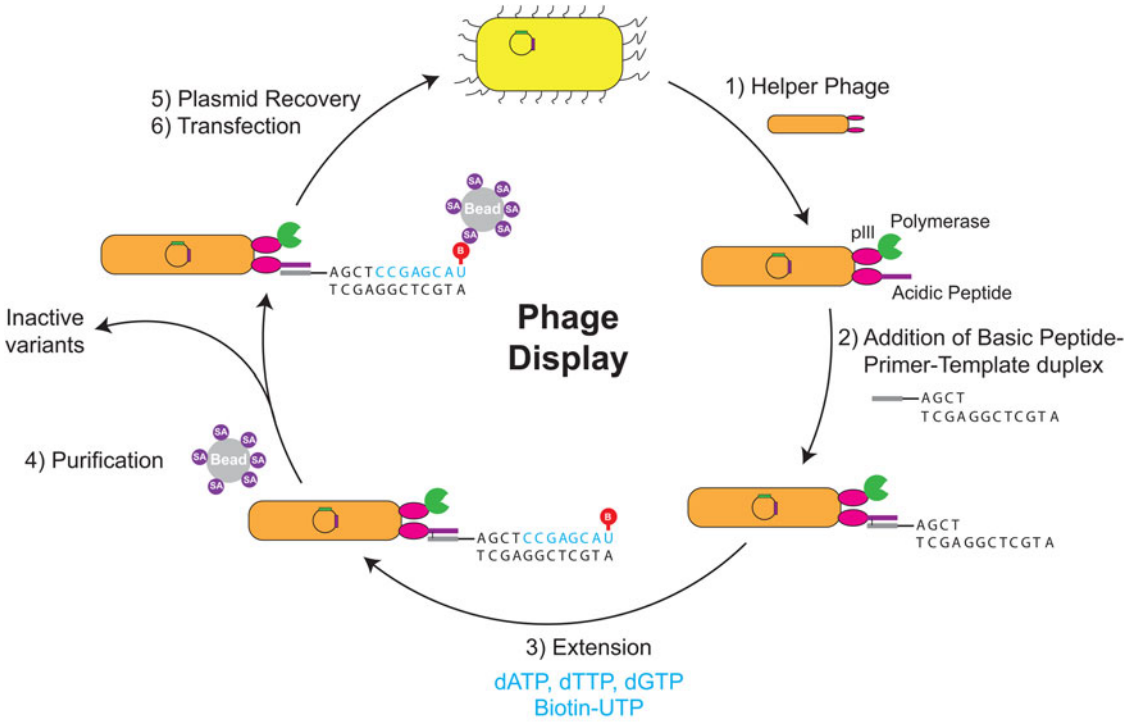


Fig. 16. Phage display. Bacteriophage particles are constructed using a proximity strategy that places the polymerase and DNA primer–template duplex in close proximity on the phage surface. Activity screening leads to the identification of polymerase variants that incorporate a biotin-tagged substrate that is captured on streptavidin-coated beads. Functional variants are recovered by eluting the beads with DNase and amplified by infecting a fresh *E. coli* culture.

support. In the case of SM4-9, for example, the selection required the screening of 500–1000 individual variants between each of the four rounds of selection (Chen *et al.*, 2016).

Compartmentalized self-replication

In 2001, Holliger and coworkers developed a polymerase evolution strategy called compartmentalized self-replication (CSR) that is based on a simple feedback loop in which a polymerase replicates its own gene by PCR (Ghadessy *et al.*, 2001). With this technique (Fig. 17), a population of *E. coli* expressing different polymerase variants is encapsulated along with the reaction buffer, dNTPs, and primers into emulsions that are produced by vigorous bulk mixing of aqueous and organic phases. During thermocycling, *E. coli* lysis occurs, releasing the polymerase and encoding plasmid into the surrounding solution. The emulsion serves as a barrier separating each polymerase extension assay into an individual reaction compartment. If the polymerase is able to amplify its own gene using the gene-specific primers supplied in the aqueous phase, then adaptive gains are made that directly and proportionately translate to an increase in the number of amplicons present that encode the active polymerase variant. Through iterative rounds of selective amplification, active polymerases will outcompete the inactive variants.

CSR has proven useful for generating polymerases with enhanced thermostability and increased resistance to a range of blood and other environmental inhibitors that prevent DNA samples from being amplified using natural polymerases (Ghadessy *et al.*, 2001; Baar *et al.*, 2011). Molecular breeding experiments performed on thermophilic polymerases led to the isolation of a chimeric polymerase with an increased ability to amplify DNA from ice-age specimens (d'Abbadie *et al.*, 2007). CSR has been used to generate polymerases that can recognize a broad range of nucleoside triphosphates, including α-phosphorothioate dNTPs (Ghadessy *et al.*, 2004), dNTPs with hydrophobic base analogs (Loakes *et al.*, 2009), and γ-modified dNTPs for sequencing and kinetic assays (Hansen *et al.*, 2011). More recently, Benner and coworkers used CSR to evolve a polymerase that could amplify DNA with a six-letter genetic alphabet that includes the unnatural base pair P:Z (Laos *et al.*, 2013).

Modified versions of CSR have been developed to reduce the adaptive burden of amplifying the entire polymerase gene (>2 kb). The first modified version, termed short-patch CSR (spCSR), focuses the amplification step on a narrow segment of the polymerase gene, which is then incorporated into the full-length gene when the plasmid is reconstructed between rounds of selection (Ong *et al.*, 2006). spCSR enables the isolation of Taq DNA polymerase variants with enhanced activity for 2′-modified nucleotides including NTPs (Ong *et al.*, 2006) as well as Pfu variants capable of replacing dCTP with fluorescent Cy3- and Cy5-labeled dCTP substrates in PCR reactions (Ramsay *et al.*, 2010). Ellington and coworkers developed another version of CSR called reverse-transcription CSR (RT-CSR), which enables the screening of up to 10^9 polymerase variants for RT activity (Ellefson *et al.*, 2016). RT-CSR was used to produce a thermostable polymerase that actively proofreads DNA synthesis during RT-PCR.

CSR benefits from a strong feedback loop that enables the identification of new polymerase variants that are capable of PCR. However, the PCR reactions take place in polydisperse droplets, which could lead to uneven levels of PCR amplification. CSR is also limited to the range of polymerase functions that promote DNA, or RNA in the case of RT-CSR, templated synthesis.

https://doi.org/10.1017/S0033583520000050 Published online by Cambridge University Press

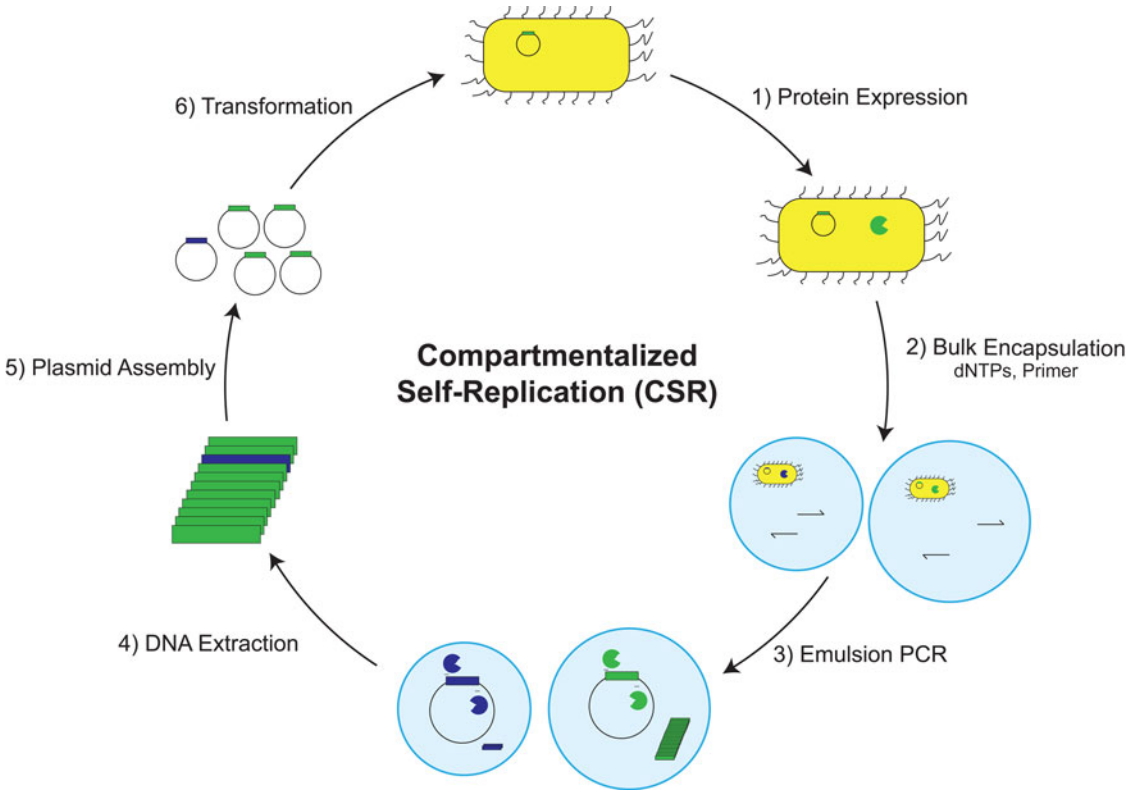


Fig. 17. Compartmentalized self-replication. A library of polymerase genes expressed in *E. coli* is encapsulated in bulk emulsions. Following PCR amplification inside the droplet, active polymerases generate multiple copies of their own gene, while inactive variants fail to replicate the gene. The degree of amplification is directly proportional to the activity of the polymerase. Through iterative rounds of selective amplification, polymerases with desired activity outcompete the population of inactive variants.

Compartmentalized self-tagging

Efforts to establish engineered polymerases with increased tolerance for challenging substrates with highly modified sugars led to the development of compartmentalized self-tagging (CST) (Pinheiro *et al.*, 2012). CST is based on a positive selection loop where a polymerase tags its encoding DNA plasmid with a biotinylated primer that hybridizes to a complementary region of the plasmid (Fig. 18). The initial primer–plasmid complex is a weak affinity interaction that becomes stabilized when the primer is extended by the polymerase. After extension, the primer–plasmid complexes are captured on streptavidin-coated beads, which are washed with mild denaturants to remove the unextended primer–plasmid pairs. Plasmids encoding active library members are then recovered from the beads, PCR amplified, and used to initiate another round of selection and amplification.

CST enabled the discovery of engineered polymerases that could synthesize XNA polymers with backbone structures that are distinct from those found in DNA and RNA (Pinheiro *et al.*, 2012). By exploring diverse library repertoires of Tgo DNA polymerase that sampled mutations within a 10 Å shell of the polymerase active site, novel polymerase variants were identified that could copy DNA templates into HNA, CeNA, TNA, FANA, and ANA. In this same study, a statistical coupling analysis was used to identify polymerases that could copy the XNA strands back into DNA. Together, these polymerase pairs demonstrate the capacity for artificial genetic polymers to replicate using engineered polymerases to facilitate the passage of genetic information back and forth between DNA and XNA. CST is widely recognized as a major advance in synthetic genetics, a field which aims to explore the structural and functional properties of XNA by *in vitro* selection (Joyce, 2012).

The major advantage of CST is that it allows for the evolution of polymerases that can synthesize nucleic acid polymers with diverse sugar-phosphate backbones. However, the range of functions is limited to DNA-templated reactions (i.e. DNA-dependent XNA polymerases), as the selection strategy uses the plasmid DNA as the template for the primer-extension reaction. CST also requires affinity purification on a solid support matrix, which lowers the partitioning efficiency of the selection due to unwanted non-specific binding of DNA to the matrix. Finally, the reliance on a metastable primer–plasmid complex requires fine-tuning of the denaturing conditions to ensure proper separation of the plasmids encoding active and inactive variants.

Droplet-based optical polymerase sorting

To overcome some of the weaknesses of previous *in vitro* selection technologies, our laboratory established a general strategy for evolving new polymerase functions called droplet-based optical polymerase sorting (DrOPS) (Larsen *et al.*, 2016). DrOPS is a high-throughput approach that combines the ultrafast screening power of microfluidics with the high sensitivity of optical sorting. With this technique, a library of polymerase variants is expressed in *E. coli* and single cells are encapsulated in microfluidic droplets containing a fluorescent sensor that is responsive to polymerase activity (Fig. 19). As with CSR and CST, the surrounding oil

https://doi.org/10.1017/S0033583520000050 Published online by Cambridge University Press

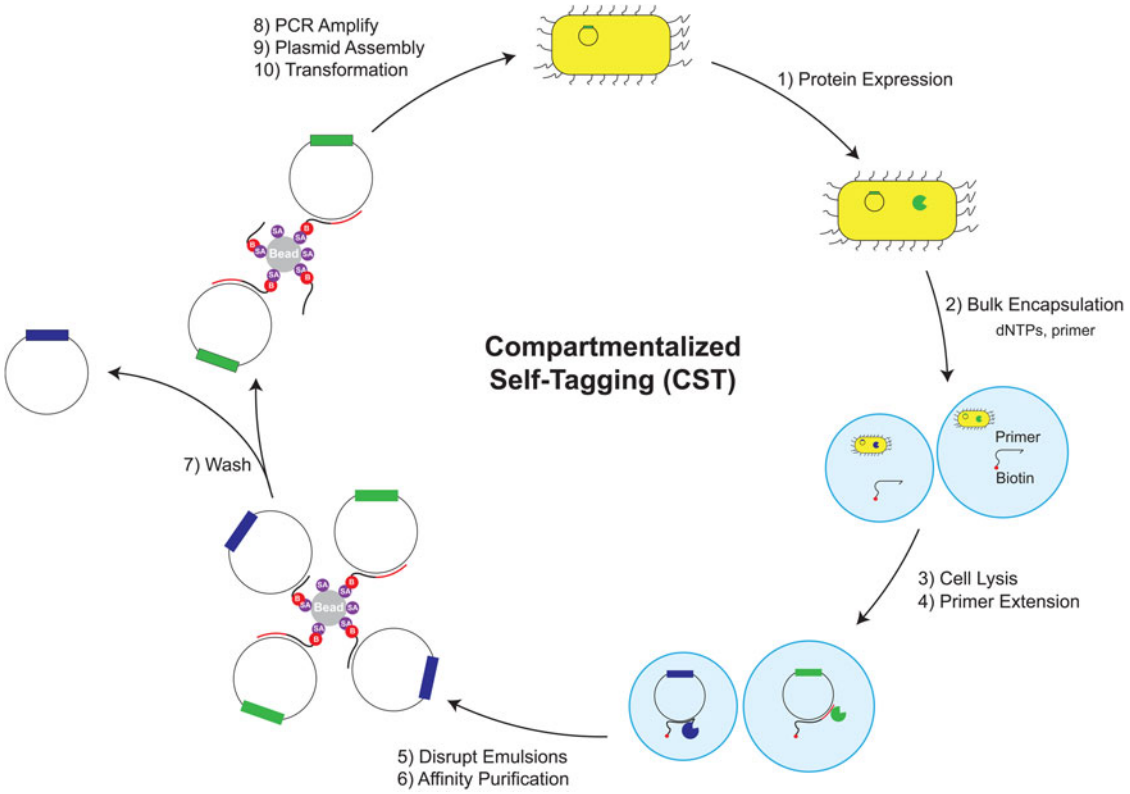


Fig. 18. Compartmentalized self-tagging. *E. coli* cells expressing different polymerase variants are encapsulated in bulk emulsions. Following *E. coli* lysis, the polymerase is challenged to extend a biotinylated primer annealed to the plasmid. Active polymerases that extend the primer increase the stability of the primer–plasmid complex. After disruption of the emulsion, the primer–plasmid complexes are captured on streptavidin beads, and plasmids annealed to unextended primers are removed with washing. Plasmids annealed to extended primers are recovered, PCR amplified, and used to initiate another round of selection.

acts as a barrier preventing the contents of one droplet from mixing with the contents of another droplet. However, unlike CSR and CST, microfluidic devices are used to generate a uniform population of droplets. The latest microfluidic designs are capable of generating 18 µm droplets at a rate of 30 000 per second, which allows for the production of $>10^8$ droplets in 1 h (Vallejo *et al.*, 2019). Following droplet production, the polymerase and encoding plasmid are released into the droplet by lysing the *E. coli* with heat. Polymerases that successfully copy the template into full-length product produce a fluorescent signal by disrupting a donor–quencher pair located at the 5′-end of the template strand. The population of droplets can then either be sorted directly using a custom microfluidic fluorescence-activated droplet sorting (FADS) device or converted to double emulsion droplets that are compatible with a traditional fluorescence-activated cell sorting instrument (Vallejo *et al.*, 2019).

Despite being a relatively new technique for polymerase engineering, DrOPS has been used to evolve polymerase variants that can synthesize TNA, an artificial genetic polymer in which the natural ribose sugar found in RNA has been replaced with an unnatural threose sugar (Schöning *et al.*, 2000). In its first demonstration, DrOPS was used to identify a manganese-independent TNA polymerase from a site-saturation library of 8000 unique variants after a single round of high-throughput screening (Larsen *et al.*, 2016). More recently, DrOPS was combined with the protein-engineering approach of deep mutational scanning (Araya and Fowler, 2011), to map the sequence function relationships of a replicative DNA polymerase, Kod, isolated from the thermophilic archeae *T. kodakarensis* (Nikoomanzar *et al.*, 2019). The resulting enrichment profile provided an unbiased view of the ability of each single-point mutant to synthesize TNA. From a single high-throughput screen, two cases of epistasis were discovered, where double-mutant variants functioned with higher activity than the sum of the contributions from either of the individual mutations. This new polymerase, termed Kod-RS, recognizes TNA substrates with nearly the same efficiency as DNA substrates, suggesting that the mutations are beginning to reshape the enzyme active site. An engineered variant with even greater TNA polymerase activity was discovered by performing deep mutational scanning across the entire polymerase domain (Nikoomanzar *et al.*, 2020).

The DrOPS technique compares favorably with other polymerase-engineering technologies in several important ways. First, it provides enormous control over the composition of the primer, template, and nucleoside triphosphates, which should make it possible to select for any type of polymerase activity (i.e. transcription, reverse transcription, and replication). Second, it relies on physical methods for identifying and sorting individual droplets with active polymerases, which greatly increases the partitioning efficiency of the selection and reduces the occurrence of background DNA contamination relative to bead binding assays. Third, microfluidic approaches provide a more economical approach to library screening by allowing researchers to screen $\sim 10^8$ variants per day using $\sim 10^6$-fold less sample volume than is typically required for automated screening approaches (Price and Paegel, 2016). The economy of scale is especially important when using unnatural nucleic acid substrates that can only be obtained by chemical synthesis and are not readily available from a commercial supplier.

https://doi.org/10.1017/S0033583520000050 Published online by Cambridge University Press

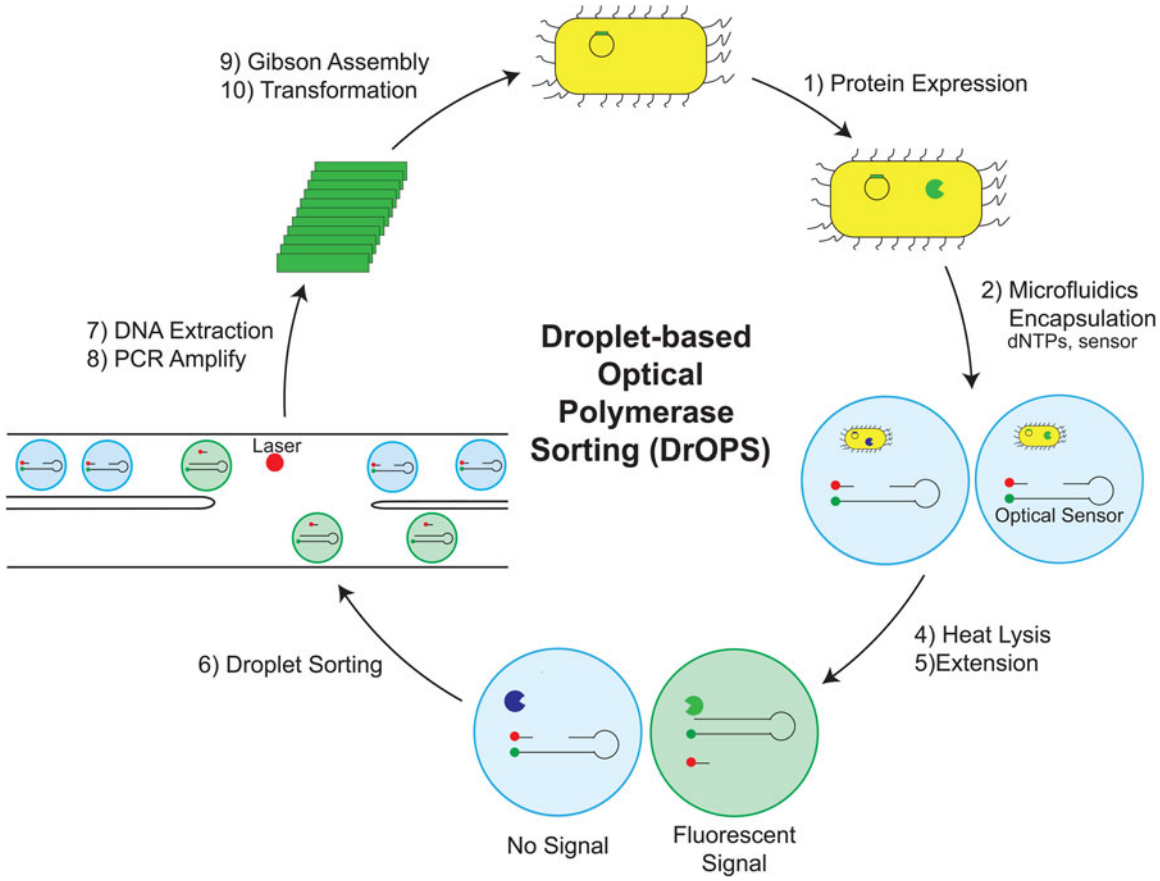


Fig. 19. Droplet-based optical polymerase sorting. *E. coli* cells expressing different library members are encapsulated in water-in-oil droplets using a microfluidic device. The droplets are collected and lysed off-chip to release the polymerase and encoding plasmid into the solution. Polymerases that extend the primer to full-length product trigger a fluorescent sensor by disrupting a fluorescent donor–quencher pair. Fluorescent droplets are sorted using a custom FADS device. Recovered DNA is PCR amplified and used to initiate another round of selection.

Applications in synthetic biology

The slow but growing availability of engineered polymerases that can synthesize artificial genetic polymers (XNAs) with high efficiency and fidelity has already started to make an impact on applications in synthetic biology, biotechnology, and medicine. The following section summarizes major achievements that have been accredited to the discovery of engineered polymerases. Most notably, these examples focus on the generation of biologically stable versions of synthetic antibodies (aptamers) and catalytic enzymes that are composed entirely of XNA. Such efforts have made it possible to bypass the arduous task of introducing modifications post-selection in which medicinal chemists painstakingly modify the backbone structure for improved biological stability while carefully avoiding chemical changes that lead to losses in activity. Since many XNAs are resistant, if not recalcitrant to nuclease digestion, research efforts have focused on establishing methods for the discovery of XNAs with desired functional properties (Culbertson *et al.*, 2016). Mastering the production of these reagents by *in vitro* selection will lead to a new generation of diagnostic and therapeutic agents for the detection and treatment of human diseases.

Synthetic antibodies

Aptamers are nucleic acid molecules that mimic antibodies by folding into tertiary structures that can bind to a broad range of targets from ions and small molecules to proteins and whole cells (Dunn *et al.*, 2017). Although some aptamers exist naturally as the binding domain of riboswitches (Doudna and Cech, 2002), most are generated by *in vitro* selection or SELEX (systematic evolution of ligands by exponential enrichment) (Wilson and Szostak, 1999). Similar to natural selection, *in vitro* selection is a Darwinian evolution process in which a large population of nucleic acid molecules (typically $>10^{14}$ unique sequences) is challenged to bind a target (Joyce, 2004). Molecules that bind to the target are recovered and amplified to generate a new population of molecules that has become enriched in members with the desired activity. The process of selection and amplification is continued until the pool becomes dominated by members that bind the target with high affinity. The ability to amplify individual molecules with desired properties and to optimize their functions by directed evolution is a defining feature that separates nucleic acid molecules from other types of organic compounds that cannot replicate because they lack a genotype–phenotype connection (Szostak, 1992).

Aptamers are often compared to antibodies due to their ability to function with high ligand binding affinity and specificity (Jayasena, 1999). However, unlike antibodies, aptamers do not require animals for their production, thus freeing them from the constraints of cellular biology and allowing for greater flexibility in their evolution under *in vitro* conditions. Once discovered, aptamers are produced through a chemical process rather than a

https://doi.org/10.1017/S0033583520000050 Published online by Cambridge University Press

biological process, which avoids the problem of viral or bacterial contamination and greatly reduces the potential for batch-to-batch variation. Aptamers developed for therapeutic purposes generally exhibit a lower immune response than proteins, and their small size (<30 *versus* ~150 kDa) provides access to biological areas that are inaccessible to antibodies (Nimjee *et al.*, 2005; Keefe *et al.*, 2010). Aptamers are able to fold reversibly, which overcomes the cold-chain problem that limits the shelf-life, reproducibility, and performance of antibodies. Therapeutic aptamers can also be deactivated with antisense oligonucleotides that recognize the binding domain, thereby providing a valuable antidote that can alleviate unwanted symptoms (Rusconi *et al.*, 2004). Finally, because aptamers are nucleic acid molecules, they can be seamlessly integrated into sensors, actuators, and other devices that are central to emerging technologies (Cho *et al.*, 2009).

Despite the many benefits of aptamers relative to antibodies, aptamers composed of natural DNA and RNA are poor candidates for diagnostic and therapeutic applications, as these molecules are rapidly degraded by nucleases present in biological samples. In one case, an unmodified DNA aptamer developed as an inhibitor of α-thrombin exhibited an *in vivo* half-life of <2 min when assayed in a primate animal model (Griffin *et al.*, 1993). Overcoming this problem led to the use of chemical modifications that protect the 2′-OH group against nucleases that utilize this position for cleavage of the phosphodiester bond. In particular, substitution of the 2′-OH group with amino (NH_2), fluoro (F), and methoxy (OCH_3) groups has led to enhanced nuclease stability (Keefe and Cload, 2008). For example, Macugen, the first FDA-approved aptamer, is an RNA sequence in which most of its 2′-OH groups have been replaced with 2′-F and 2′-OCH_3 groups (Ng *et al.*, 2006). However, it is important to note that these modifications are often still prone to nuclease digestion (Cummins *et al.*, 1995; Noronha *et al.*, 2000).

While numerous examples of 2′-modified aptamers have been described in the literature (Keefe and Cload, 2008), most were generated by transcription using T7 RNA polymerase. Recently, Romesberg and colleagues reported the directed evolution of variants of the Stoffel fragment of Taq DNA polymerase that accepts a broad range of 2′-modified substrates (Chen *et al.*, 2016). One variant, SFM4-3, was found to PCR amplify substrates with 2′-F and 2′-azido (N_3) groups (Chen *et al.*, 2016). This engineered polymerase was subsequently used to evolve aptamers that bind to human neutrophil elastase (HNE), a serine protease associated with inflammatory diseases, using libraries that are partially substituted with 2′-modified nucleotides (Thirunavukarasu *et al.*, 2017; Shao *et al.*, 2020). In a related study, Romesberg used the engineered polymerases, SFM4-6 and SFM4-9, to evolve fully 2′-OCH_3 modified aptamers to HNE (Liu *et al.*, 2017). Structure–activity assays reveal that the 2′ modifications are necessary for aptamer activity.

A fundamentally different approach to nuclease stability involves the use of XNAs, which are artificial genetic polymers in which the ribose and deoxyribose sugars found in RNA and DNA have been replaced with a different sugar moiety (Chaput and Herdewijn, 2019). TNA and HNA are particularly interesting as their backbone structures are recalcitrant to nuclease digestion, making them valuable systems for diagnostic and therapeutic applications (Hendrix *et al.*, 1997; Culbertson *et al.*, 2016). The first XNA aptamers appeared in 2012 with the evolution of TNA sequences that can bind to human α-thrombin and HNA sequences having affinity for the HIV trans-activating response element and hen egg lysozyme (Pinheiro *et al.*, 2012; Yu *et al.*, 2012). These results were viewed as a milestone in synthetic biology, as they demonstrated that heredity and evolution are no longer limited to DNA and RNA (Joyce, 2012).

The last few years have witnessed tremendous growth in the field of XNA aptamer research, with improved enzymes and selection techniques having given rise to higher quality aptamers whose oligonucleotide sequences derive entirely from building blocks with sugar moieties unrelated to those found in nature. Defestano and colleagues, for example, have developed a FANA aptamer that binds HIV RT with low picomolar affinity (Alves Ferreira-Bravo *et al.*, 2015). Similarly, Herdewijn and colleagues evolved an HNA aptamer to rat vascular endothelial growth factor 164 ($VEGF_{164}$) that distinguishes VEGF164 from the $VEGF_{120}$ isoform (Eremeeva *et al.*, 2019). In addition, new TNA aptamers have been discovered with affinity to the small molecule target ochratoxin A and the proteins thrombin and HIV RT (Dunn and Chaput, 2016; Mei *et al.*, 2018; Rangel *et al.*, 2018). More recently, our lab has developed a DNA-display strategy for evolving XNA aptamers in which each XNA strand is physically linked to its encoding DNA template (Dunn *et al.*, 2020). This strategy is analogous to protein display technologies, such as mRNA display that provide a covalent link between the encoding mRNA and translated protein (Roberts and Szostak, 1997). This approach is generalizable to any XNA system where an XNA polymerase is available to copy DNA templates into XNA. It also avoids the need for an XNA reverse transcriptase, which improves the recovery of functional sequences that are present in low abundance after stringent washing has been performed to remove weaker affinity binders. Using this strategy, a TNA aptamer to HIV RT was produced that rivals the best monoclonal antibodies in terms of binding affinity and thermal stability (Dunn *et al.*, 2020). As these studies continue, it will be interesting to see how effective XNA aptamers will be at disrupting extracellular targets, such as the interaction of the viral spike protein of SARS-CoV-2 with the ACE2 receptor of human lung cells.

Catalysts for RNA modifying reactions

Nucleic acid enzymes provide powerful tools for precision medicine by allowing viral or disease-associated RNAs to be cleaved at specific nucleotide positions. The most widely studied member of this family of enzymes is DNAzyme 10-23 (Santoro and Joyce, 1997), which has been evaluated in phases I and II clinical trials for a variety of diseases ranging from basal cell carcinoma to bronchial asthma (Fokina *et al.*, 2015). 10-23 is a magnesium-dependent enzyme that catalyzes the hydrolysis of a phosphodiester bond at a specific dinucleotide junction in the RNA substrate (Santoro and Joyce, 1998). Unlike other gene-silencing technologies (e.g. antisense, siRNA, and CRISPR), DNAzymes benefit from a mechanism that does not require the recruitment of endogenous enzymes. Instead, Watson–Crick base pairing directs the enzyme to a cleavage site that is cut via an in-line attack by a deprotonated form of the 2′-OH group to produce an upstream cleavage product carrying a 2′,3′-cyclic phosphate and a downstream strand with a 5′-OH group. The enzyme is made generalizable by designing the substrate binding arms to be complementary to the cleavage site. This property of chemical simplicity, coupled with its ease of synthesis, has allowed 10-23 to become a popular tool for clinical and basic research. Over the years, numerous chemical modifications have been made to protect 10-23 from nuclease digestion and increase its efficacy *in vivo*

https://doi.org/10.1017/S0033583520000050 Published online by Cambridge University Press

(Fokina *et al.*, 2017). These include the introduction of an inverted 3′-3′ nucleotide and substitution of the deoxyribose sugar for other sugar moieties (Fokina *et al.*, 2017).

Although 10-23 is known to function with high activity ($k_{cat}/K_m \sim 10^9\ M^{-1}\ min^{-1}$) under optimized *in vitro* conditions, biological assays show that its capacity for RNA cleavage activity is greatly diminished in cellular environments where Mg^{2+} ions are present in lower abundance (Young *et al.*, 2010). One approach to solving this problem involves developing catalysts that carry functional groups that augment the chemical functionality of natural bases. Perrin and colleagues, for example, have evolved a divalent-metal independent DNAzyme that cleaves RNA with multiple turnover activity using imidazole side chains that mimic the mechanism of RNase A (Wang *et al.*, 2018*a*). Another approach is to evolve XNA enzymes (XNAzymes) that are able to bind divalent metal ions with higher affinity. In 2015, Holliger and colleagues described the first examples of XNA catalysts with RNA cleavage and ligation activity (Taylor *et al.*, 2015). Using engineered polymerases, XNAzymes were isolated from four different backbone chemistries: FANA, ANA, CeNA, and HNA. However, despite the presence of high concentrations of Mg^{2+} ions, the XNAzymes produced from this study functioned with relatively weak activity. More recently, we have discovered a FANAzyme, termed FANAzyme 12-7, by *in vitro* evolution that functions at a rate enhancement of $\sim 10^6$-fold over the uncatalyzed reaction and exhibits substrate saturation kinetics typical of most enzymes (Wang *et al.*, 2018*b*). Remarkably, FANAzyme 12-7 cleaves chimeric DNA substrates (DNA substrates having a riboG residue at the cleavage site) under physiological conditions with an activity rivaling that of known DNAzymes that were intentionally selected to recognize such substrates (Wang *et al.*, 2020).

Future directions

Many synthetic biology applications are currently limited by a lack of polymerases that are available to perform a specific function with optimal activity. However, with the advent of new polymerase-engineering technologies, we anticipate that many of these limitations will likely be overcome in the near future. The following section provides some examples where polymerase-engineering technologies could help drive future innovations in synthetic biology, biotechnology, and medicine. Of course, many other examples are possible, including those that have not yet been envisioned.

Next-generation sequencing. Next-generation DNA-sequencing technologies that follow a sequencing-by-synthesis strategy require a DNA polymerase that can facilitate the incorporation of chemically modified nucleotides (Goodwin *et al.*, 2016). Illumina technology, for example, utilizes dNTP substrates that carry a fluorescent dye and reversible 3′-terminator that are removed following nucleotide incorporation (Chen, 2014). However, the removal chemistry leaves behind a portion of the linker connecting the fluorescent dye to the base, commonly referred to as a scar that can reduce the efficiency of polymerase-mediated primer extension. Evolving polymerases that can function with higher activity in the presence of scarred nucleotides may improve NGS technology by allowing for longer read lengths, faster turn-around times, and higher quality reads. A similar case exists for RNA-Seq applications where polymerases continually struggle to read through complex RNA structures. Polymerase evolution could overcome this problem by providing thermophilic reverse transcriptases that function at higher temperatures where larger RNA structures denature into single-stranded form.

Oligonucleotide synthesis. Solid-phase DNA synthesis based on phosphoramidite chemistry has driven major advances in the biological sciences by providing easy access to synthetic oligonucleotides (Caruthers, 1985). Examples where this technology has made a major impact include DNA nanotechnology (Seeman and Sleiman, 2018), digital data archiving (Ceze *et al.*, 2019), genome synthesis (Hutchison *et al.*, 2016), PCR (Saiki *et al.*, 1988), and SELEX (Wilson and Szostak, 1999). However, the process of industrial scale DNA synthesis produces large quantities of hazardous waste that require appropriate disposal mechanisms. Moving to an enzymatic DNA synthesis platform would eliminate or greatly reduce this problem by allowing the reactions to proceed in an aqueous environment. Toward this goal, several groups are working to develop terminal deoxynucleotide transferase (TdT) as a possible paradigm for template-independent enzymatic DNA synthesis (Palluk *et al.*, 2018; Lee *et al.*, 2019). One could imagine that such approaches would benefit from polymerase engineering by providing access to new TdT variants that allow for longer DNA synthesis lengths, higher yields, and access to diverse nucleotide chemistries. The application of this approach to XNA, for example, could provide access to synthetic XNAs that are difficult to generate by conventional solid-phase synthesis due to low nucleotide coupling yields.

Information storage. In an age of ever-increasing data, new mechanisms for data storage are in short supply. One paradigm that has attracted significant attention involves using DNA as a soft material for low energy, high-density information storage of digital data (Ceze *et al.*, 2019). At its maximum, 1 g of DNA can store 455 exabytes of information, which vastly exceeds the largest conventional devices (Church *et al.*, 2012). Information storage occurs in four main steps that involve encoding digital information (e.g. text, pictures, and movies) in DNA, writing the information by massively parallel DNA synthesis, reading the information by NGS analysis, and decoding the information back into digital format. However, because DNA is a naturally occurring molecule that is prone to nuclease digestion, information stored in DNA could be unintentionally lost through an accidental encounter with nucleases present in the environment. XNAs that are recalcitrant to nuclease digestion offer a solution to this problem by providing a biologically stable alternative to DNA (Culbertson *et al.*, 2016). For this to be possible, XNA polymerases will likely need to be optimized for higher fidelity and lower template-sequence bias, which could be achieved by directed evolution.

Engineering bacteria. One recent exciting example relevant to the field of synthetic biology is the generation of mutant *E. coli* strains derived from engineered bacteria that contain significant (~40–50%) ribonucleotide content in their genome (Mehta *et al.*, 2018). These systems have the potential to provide new insight into the origin of life by offering a better understanding of the transition from the RNA to the DNA world. Such studies would evaluate the relevance of chimeric DNA–RNA genomes with respect to the processes of replication, transcription, and repair. In this area of research, engineered polymerases will almost certainly be required to produce new mutant *E. coli* strains that contain >50% RNA content in their genomes. Although a daunting task, establishing an *E. coli* strain with an entirely RNA-derived genome would herald a new advance in synthetic biology. Other related areas where polymerase engineering could contribute to the development of engineered bacterial

https://doi.org/10.1017/S0033583520000050 Published online by Cambridge University Press

strains involves ongoing efforts to create bacterial cells in which all synthetic biology information is stored in XNA polymers (Schmidt, 2010). This field of study, commonly referred to as xenobiology, relies on the concept of genetic orthogonality in which synthetic biology information is stored in XNA chromosomes that are made to replicate independently from the natural host genome (Chaput *et al.*, 2020). In this way, the growing field of xenobiology promises to make synthetically engineered organisms safer by establishing a genetic firewall between natural biology and synthetic biology (Herdewijn and Marliere, 2009).

Conclusion

In summary, we provide a comprehensive review of polymerase engineering that travels the path from early exploratory studies to modern enzyme-engineering technologies where variants are sampled with incredible speed and accuracy. Though not explicitly discussed in this review, it should be noted that these endeavors have been supported by equally significant advances in nucleic acid chemistry, which provide access to chemical building blocks with new physical and chemical properties. This combination of nucleic acid chemistry with enzyme engineering will uniquely drive new applications in synthetic biology, medicine, and biotechnology.

Acknowledgement. We wish to thank members of the Chaput lab for helpful discussions and critical reading of the manuscript.

Financial support. This work was supported by the National Science Foundation (MCB: 1946312) and the W.M. Keck Foundation.

Conflict of interest. The authors declare no conflicts of interest.

References

Abu Al-Soud W and Radstrom P (1998) Capacity of nine thermostable DNA polymerases to mediate DNA amplification in the presence of PCR-inhibiting samples. *Applied and Environmental Microbiology* **64**, 3748–3753.

Abu Al-Soud W and Radstrom P (2000) Effects of amplification facilitators on diagnostic PCR in the presence of blood, feces, and meat. *Journal of Clinical Microbiology* **38**, 4463–4470.

Allis CD and Jenuwein T (2016) The molecular hallmarks of epigenetic control. *Nature Reviews Genetics* **17**, 487–500.

Alves Ferreira-Bravo I, Cozens C, Holliger P and Destefano JJ (2015) Selection of 2′-deoxy-2′-fluoroarabinonucleotide (FANA) aptamers that bind HIV-1 reverse transcriptase with picomolar affinity. *Nucleic Acids Research* **43**, 9587–9599.

An L, Tang W, Ranalli TA, Kim HJ, Wytiaz J and Kong H (2005) Characterization of a thermostable UvrD helicase and its participation in helicase-dependent amplification. *Journal of Biological Chemistry* **280**, 28952–28958.

Anosova I, Kowal EA, Dunn MR, Chaput JC, Van Horn WD and Egli M (2016) The structural diversity of artificial genetic polymers. *Nucleic Acids Research* **44**, 1007–1021.

Arana ME, Takata K, Garcia-Diaz M, Wood RD and Kunkel TA (2007) A unique error signature for human DNA polymerase nu. *DNA Repair* **6**, 213–223.

Araya CL and Fowler DM (2011) Deep mutational scanning: assessing protein function on a massive scale. *Trends in Biotechnology* **29**, 435–442.

Arezi B and Hogrefe HH (2007) Escherichia coli DNA polymerase III epsilon subunit increases Moloney murine leukemia virus reverse transcriptase fidelity and accuracy of RT-PCR procedures. *Analytical Biochemistry* **360**, 84–91.

Astatke M, Ng K, Grindley NDF and Joyce CM (1998) A single side chain prevents *Escherichia coli* DNA polymerase I (Klenow fragment) from incorporating ribonucleotides. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 3402–3407.

Baar C, d'Abbadie M, Vaisman A, Arana ME, Hofreiter M, Woodgate R, Kunkel TA and Holliger P (2011) Molecular breeding of polymerases for resistance to environmental inhibitors. *Nucleic Acids Research* **39**, e51.

Baker TA and Bell SP (1998) Polymerases and the replisome: machines within machines. *Cell* **92**, 295–305.

Barnes WM (1992) The fidelity of Taq polymerase catalyzing PCR is improved by an N-terminal deletion. *Gene* **112**, 29–35.

Beard WA and Wilson SH (2000) Structural design of a eukaryotic DNA repair polymerase: DNA polymerase beta. *Mutation Research* **460**, 231–244.

Bebenek K, Joyce CM, Fitzgerald MP and Kunkel TA (1990) The fidelity of DNA synthesis catalyzed by derivatives of Escherichia coli DNA polymerase I. *Journal of Biological Chemistry* **265**, 13878–13887.

Benner SA (2004) Understanding nucleic acids using synthetic chemistry. *Accounts of Chemical Research* **37**, 784–797.

Benner SA, Allemann RK, Ellington AD, Ge L, Glasfeld A, Leanz GF, Krauch T, MacPherson LJ, Moroney S, Piccirilli JA, Weinhold E*et al.* (1987) Natural selection, protein engineering, and the last riboorganism: rational model building in biochemistry. *Cold Spring Harbor Symposia on Quantitative Biology* **52**, 53–63.

Berezhna SY, Gill JP, Lamichhane R and Millar DP (2012) Single-molecule Forster resonance energy transfer reveals an innate fidelity checkpoint in DNA polymerase I. *Journal of the American Chemical Society* **134**, 11261–11268.

Bergen K, Steck AL, Strutt S, Baccaro A, Welte W, Diederichs K and Marx A (2012) Structures of KlenTaq DNA polymerase caught while incorporating C5-modified pyrimidine and C7-modified 7-deazapurine nucleoside triphosphates. *Journal of the American Chemical Society* **134**, 11840–11843.

Bergen K, Betz K, Welte W, Diederichs K and Marx A (2013) Structures of KOD and 9°N DNA polymerases complexed with primer template duplex. *ChemBioChem* **14**, 1058–1062.

Betz K, Malyshev DA, Lavergne T, Welte W, Diederichs K, Dwyer TJ, Ordoukhanian P, Romesberg FE and Marx A (2012) Klentaq polymerase replicates unnatural base pairs by inducing a Watson–Crick geometry. *Nature Chemical Biology* **8**, 612–614.

Betz K, Malyshev DA, Lavergne T, Welte W, Diederichs K, Romesberg FE and Marx A (2013) Structural insights into DNA replication without hydrogen bonds. *Journal of the American Chemical Society* **135**, 18637–18643.

Betz K, Kimoto M, Diederichs K, Hirao I and Marx A (2017) Structural basis for expansion of the genetic alphabet with an artificial nucleobase pair. *Angewandte Chemie* **56**, 12000–12003.

Bonnin A, Lazaro JM, Blanco L and Salas M (1999) A single tyrosine prevents insertion of ribonucleotides in the eukaryotic-type phi 29 DNA polymerase. *Journal of Molecular Biology* **290**, 241–251.

Boulé JB, Rougeon F and Papanicolaou C (2001) Terminal deoxynucleotidyl transferase indiscriminately incorporates ribonucleotides and deoxyribonucleotides. *Journal of Biological Chemistry* **276**, 31388–31393.

Boyer PL, Stenbak CR, Hoberman D, Linial ML and Hughes SH (2007) In vitro fidelity of the prototype primate foamy virus (PFV) RT compared to HIV-1 RT. *Virology* **367**, 253–264.

Brautigam CA and Steitz TA (1998) Structural and functional insights provided by crystal structures of DNA polymerases and their substrate complexes. *Current Opinion in Structural Biology* **8**, 54–63.

Brotschi C, Haberli A and Leumann CJ (2001) A stable DNA duplex containing a non-hydrogen-bonding and non-shape-complementary base couple: interstrand stacking as the stability determining factor. *Angewandte Chemie* **40**, 3012–3014.

Brown JA and Suo Z (2011) Unlocking the sugar 'steric gate' of DNA polymerases. *Biochemistry* **50**, 1135–1142.

Bryant FR, Johnson KA and Benkovic SJ (1983) Elementary steps in the DNA polymerase I reaction pathway. *Biochemistry* **22**, 3537–3546.

Cahova H, Panattoni A, Kielkowski P, Fanfrlik J and Hocek M (2016) 5-Substituted pyrimidine and 7-substituted 7-deazapurine dNTPs as substrates for DNA polymerases in competitive primer extension in the presence of natural dNTPs. *ACS Chemical Biology* **11**, 3165–3171.

Caruthers MH (1985) Gene synthesis machines: DNA chemistry and its uses. *Science (New York, N.Y.)* **230**, 281–285.

https://doi.org/10.1017/S0033583520000050 Published online by Cambridge University Press

Ceze L, Nivala J and Strauss K (2019) Molecular digital data storage using DNA. *Nature Reviews Genetics* **20**, 456–466.

Chaput JC and Herdewijn P (2019) What is XNA? *Angewandte Chemie* **58**, 11570–11572.

Chaput JC and Switzer C (1999) A DNA pentaplex incorporating nucleobase quintets. *Proceedings of the National Academy of Sciences of the United States of America* **96**, 10614–10619.

Chaput JC and Switzer C (2000) Non-enzymatic transcription of an isoG-isoC base pair. *Journal of the American Chemical Society* **122**, 12866–12867.

Chaput JC and Szostak JW (2003) TNA synthesis by DNA polymerases. *Journal of the American Chemical Society* **125**, 9274–9275.

Chaput JC, Ichida JK and Szostak JW (2003) DNA polymerase-mediated DNA synthesis on a TNA template. *Journal of the American Chemical Society* **125**, 856–857.

Chaput JC, Herdewijn P and Hollenstein M (2020) Orthogonal genetic systems. *ChemBioChem* **21**, 1408–1411.

Chen CY (2014) DNA polymerases drive DNA sequencing-by-synthesis technologies: both past and present. *Frontiers in Microbiology* **5**, 305.

Chen T, Hongdilokkul N, Liu Z, Adhikary R, Tsuen SS and Romesberg FE (2016) Evolution of thermophilic DNA polymreases for the recognition and amplification of C2′-modified DNA. *Nature Chemistry* **8**, 556–562.

Chien A, Edgar DB and Trela JM (1976) Deoxyribonucleic acid polymerase from the extreme thermophile *Thermus aquaticus*. *Journal of Bacteriology* **127**, 1550–1557.

Chim N, Jackson LN, Trinh AM and Chaput JC (2018) Crystal structures of DNA polymerase I capture novel intermediates in the DNA synthesis pathway. *Elife* **7**, e40444.

Cho EJ, Lee JW and Ellington AD (2009) Applications of aptamers as sensors. *Annual Review of Analytical Chemistry* **2**, 241–264.

Church GM, Gao Y and Kosuri S (2012) Next-generation digital information storage in DNA. *Science (New York, N.Y.)* **337**, 1628.

Clark JM (1988) Novel non-templated nucleotide addition reactions catalyzed by procaryotic and eucaryotic DNA polymerases. *Nucleic Acids Research* **16**, 9677–9686.

Coffin JM and Fan H (2016) The discovery of reverse transcriptase. *Annual Review of Virology* **3**, 29–51.

Cozens C, Pinheiro VB, Vaisman A, Woodgate R and Holliger P (2012) A short adaptive path from DNA to RNA polymerases. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 8067–8072.

Crouzier L, Dubois C, Edwards SL, Lauridsen LH, Wengel J and Veedu RN (2012) Efficient reverse transcription using locked nucleic acid nucleotides towards the evolution of nuclease resistant RNA aptamers. *PLoS ONE* **7**, e35990.

Culbertson MC, Temburnikar KW, Sau SP, Liao J-Y, Bala S and Chaput JC (2016) Evaluating TNA stability under simulated physiological conditions. *Bioorganic & Medicinal Chemistry Letters* **26**, 2418–2421.

Cummins LL, Owens SR, Risen LM, Lesnik EA, Freier SM, Mcgee D, Guinosso CJ and Cook PD (1995) Characterization of fully 2′-modified oligoribonucleotide hetero- and homoduplex hybridization and nuclease sensitivity. *Nucleic Acids Research* **23**, 2019–2024.

d'Abbadie M, Hofreiter M, Vaisman A, Loakes D, Gasparutto D, Cadet J, Woodgate R, Paabo S and Holliger P (2007) Molecular breeding of polymerases for amplification of ancient DNA. *Nature Biotechnology* **25**, 939–943.

Damha MJ, Wilds CJ, Noronha A, Brukner I, Borkow G, Arion D and Parniak MA (1998) Hybrids of RNA and arabinonucleic acids (ANA and 2′F-ANA) are substrates for ribonuclease H. *Journal of the American Chemical Society* **120**, 12976–12977.

Das SK and Fujimura RK (1979) Processiveness of DNA polymerases. A comparative study using a simple procedure. *Journal of Biological Chemistry* **254**, 1227–1232.

Doublie S, Tabor S, Long AM, Richardson CC and Ellenberger T (1998) Crystal structure of a bacteriophage T7 DNA replication complex at 2.2 a resolution. *Nature* **391**, 251–258.

Doudna JA and Cech TR (2002) The chemical repertoire of natural ribozymes. *Nature* **418**, 222–228.

Dube DK and Loeb LA (1975) Manganese as a mutagenic agent during in vitro DNA synthesis. *Biochemical and Biophysical Research Communications* **67**, 1041–1046.

Dunn MR and Chaput JC (2016) Reverse transcription of threose nucleic acid by a naturally occurring DNA polymerase. *ChemBioChem* **17**, 1804–1808.

Dunn MR, Otto C, Fenton KE and Chaput JC (2016) Improving polymerase activity with unnatural substrates by sampling mutations in homologous protein architectures. *ACS Chemical Biology* **11**, 1210–1219.

Dunn MR, Jimenez RM and Chaput JC (2017) Analysis of aptamer discovery and technology. *Nature Reviews Chemistry* **1**, 0076.

Dunn MR, McCloskey CM, Buckley P, Rhea K and Chaput JC (2020) Generating biologically stable TNA aptamers that function with high affinity and thermal stability. *Journal of the American Chemical Society* **142**, 7721–7724.

Ellefson JW, Gollihar J, Shroff R, Shivram H, Iyer VR and Ellington AD (2016) Synthetic evolutionary origin of a proofreading reverse transcriptase. *Science (New York, N.Y.)* **352**, 1590–1593.

Elshawadfy AM, Keith BJ, Ee Ooi H, Kinsman T, Heslop P and Connolly BA (2014) DNA Polymerase hybrids derived from the family-B enzymes of *Pyrococcus furiosus* and *Thermococcus kodakarensis*: improving performance in the polymerase chain reaction. *Frontiers in Microbiology* **5**, 224.

Engelke DR, Krikos A, Bruck ME and Ginsburg D (1990) Purification of *Thermus aquaticus* DNA polymerase expressed in *Escherichia coli*. *Analytical Biochemistry* **191**, 396–400.

Eremeeva E, Fikatas A, Margamuljana L, Abramov M, Schols D, Groaz E and Herdewijn P (2019) Highly stable hexitol based XNA aptamers targeting the vascular endothelial growth factor. *Nucleic Acids Research* **47**, 4927-4939.

Evans SJ, Fogg MJ, Mamone A, Davis M, Pearl LH and Connolly BA (2000) Improving dideoxynucleotide-triphosphate utilisation by the hyperthermophilic DNA polymerase from the archaeon *Pyrococcus furiosus*. *Nucleic Acids Research* **28**, 1059–1066.

Fa M, Radeghieri A, Henry AA and Romesberg FE (2004) Expanding the substrate repertoire of a DNA polymerase by directed evolution. *Journal of the American Chemical Society* **126**, 1748–1754.

Fogg MJ, Pearl LH and Connolly BA (2002) Structural basis for uracil recognition by archaeal family B DNA polymerases. *Nature Structural Biology* **9**, 922–927.

Fokina AA, Stetsenko DA and Francois JC (2015) DNA enzymes as potential therapeutics: towards clinical application of 10-23 DNAzymes. *Expert Opinion Biological Therapy* **15**, 689–711.

Fokina AA, Chelobanov BP, Fujii M and Stetsenko DA (2017) Delivery of therapeutic RNA-cleaving oligodeoxyribonucleotides (deoxyribozymes): from cell culture studies to clinical trials. *Expert Opinion on Drug Delivery* **14**, 1077–1089.

Franklin MC, Wang J and Steitz TA (2001) Structure of the replicating complex of a pol alpha family DNA polymerase. *Cell* **105**, 657–667.

Freudenthal BD, Beard WA, Shock DD and Wilson SH (2013) Observing a DNA polymerase choose right from wrong. *Cell* **154**, 157–168.

Gao Y and Yang W (2016) Capture of a third Mg(2)(+) is essential for catalyzing DNA synthesis. *Science (New York, N.Y.)* **352**, 1334–1337.

Gardner AF and Jack WE (1999) Determinants of nucleotide sugar recognition in an archaeon DNA polymerase. *Nucleic Acids Research* **27**, 2545–2553.

Gardner AF and Jack WE (2002) Acyclic and dideoxy terminator preferences denote divergent sugar recognitionby archaeon and TaqDNA polymerases. *Nucleic Acids Research* **30**, 605–613.

Gardner AF, Joyce CM and Jack WE (2004) Comparative kinetics of nucleotide analog incorporation by vent DNA polymerase. *Journal of Biological Chemistry* **279**, 11834–11842.

Gardner AF, Wang J, Wu W, Karouby J, Li H, Stupi BP, Jack WE, Hersh MN and Metzker ML (2012) Rapid incorporation kinetics and improved fidelity of a novel class of 3′-OH unblocked reversible terminators. *Nucleic Acids Research* **40**, 7404–7415.

Gardner AF, Jackson KM, Boyle MM, Buss JA, Potapov V, Gehring AM, Zatopek KM, Correa IR Jr, Ong JL and Jack WE (2019) Therminator

https://doi.org/10.1017/S0033583520000050 Published online by Cambridge University Press

DNA polymerase: modified nucleotides and unnatural substrates. *Frontiers in Molecular Biosciences* **6**, 28.

Garibyan L and Avashia N (2013) Polymerase chain reaction. *Journal of Investigative Dermatology* **133**, 1–4.

Gawande BN, Rohloff JC, Carter JD, Von Carlowitz I, Zhang C, Schneider DJ and Janjic N (2017) Selection of DNA aptamers with two modified bases. *Proceedings of the National Academy of Sciences of the United States of America* **114**, 2898–2903.

Gelles J and Landick R (1998) RNA polymerase as a molecular motor. *Cell* **93**, 13–16.

Genna V, Vidossich P, Ippoliti E, Carloni P and De Vivo M (2016) A self-activated mechanism for nucleic acid polymerization catalyzed by DNA/RNA polymerases. *Journal of the American Chemical Society* **138**, 14592–14598.

Ghadessy FJ, Ong JL and Holliger P (2001) Directed evolution of polymerase function by compartmentalized self-replication. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 4552–4557.

Ghadessy FJ, Ramsay N, Boudsocq F, Loakes D, Brown A, Iwai S, Vaisman A, Woodgate R and Holliger P (2004) Generic expansion of the substrate spectrum of a DNA polymerase by directed evolution. *Nature Biotechnology* **22**, 755–759.

Goodwin S, Mcpherson JD and Mccombie WR (2016) Coming of age: ten years of next-generation sequencing technologies. *Nature Reviews Genetics* **17**, 333–351.

Greagg MA, Fogg MJ, Panayotou G, Evans SJ, Connolly BA and Pearl LH (1999) A read-ahead function in archaeal DNA polymerases detects promutagenic template-strand uracil. *Proceedings of the National Academy of Sciences of the United States of America* **96**, 9045–9050.

Griep M, Kotera C, Nelson R and Viljoen H (2006) Kinetics of the DNA polymerase pyrococcus kodakaraensis. *Chemical Engineering Science* **61**, 3885–3892.

Griffin LC, Tidmarsh GF, Bock LC, Toole JJ and Leung LL (1993) In vivo anticoagulant properties of a novel nucleotide-based thrombin inhibitor and demonstration of regional anticoagulation in extracorporeal circuits. *Blood* **81**, 3271–3276.

Gubler U (1987) Second-strand cDNA synthesis: classical method. *Methods in Enzymology* **152**, 325–329.

Hamashima K, Kimoto M and Hirao I (2018) Creation of unnatural base pairs for genetic alphabet expansion toward synthetic xenobiology. *Current Opinion in Chemical Biology* **46**, 108–114.

Hansen CJ, Wu L, Fox JD, Arezi B and Hogrefe HH (2011) Engineered split in Pfu DNA polymerase fingers domain improves incorporation of nucleotide gamma-phosphate derivative. *Nucleic Acids Research* **39**, 1801–1810.

Hendrix C, Rosemeyer H, Verheggen I, Seela F, Van Aerschot A and Herdewijn P (1997) 1′,5′-Anhydrohexitol oligonucleotides: synthesis, base pairing and recognition by regular oligodeoxyribonucleotides and oligoribonucleotides. *Chemistry – A European Journal* **3**, 110–120.

Herdewijn P and Marliere P (2009) Toward safe genetically modified organisms through the chemical diversification of nucleic acids. *Chemistry and Biodiversity* **6**, 791–808.

Heuberger BD and Switzer C (2008) A pre-RNA candidate revisited: both enantiomers of flexible nucleoside triphosphates are DNA polymerase substrates. *Journal of the American Chemical Society* **130**, 412–413.

Hogrefe HH, Cline J, Lovejoy AE and Nielson KB (2001) DNA polymerases from hyperthermophiles. *Methods in Enzymology* **334**, 91–116.

Hohlbein J, Aigrain L, Craggs TD, Bermek O, Potapova O, Shoolizadeh P, Grindley ND, Joyce CM and Kapanidis AN (2013) Conformational landscapes of DNA polymerase I and mutator derivatives establish fidelity checkpoints for nucleotide insertion. *Nature Communications* **4**, 2131.

Holland PM, Abramson RD, Watson R and Gelfand DH (1991) Detection of specific polymerase chain reaction product by utilizing the 5′----3′ exonuclease activity of *Thermus aquaticus* DNA polymerase. *Proceedings of the National Academy of Sciences of the United States of America* **88**, 7276–7280.

Hollenstein M (2012) Nucleoside triphosphates – building blocks for the modification of nucleic acids. *Molecules* **17**, 13569–13591.

Horn T, Chang CA and Collins ML (1995) Hybridization properties of the 5-methyl-isocytidine/isoguanosine base-pair in synthetic oligodeoxynucleotides. *Tetrahedron Letters* **36**, 2033–2036.

Hoshika S, Leal NA, Kim MJ, Kim MS, Karalkar NB, Kim HJ, Bates AM, Watkins NE Jr, Santalucia HA, Meyer AJ, Dasgupta S, Piccirilli JA, Ellington AD, Santalucia J Jr, Georgiadis MM and Benner SA (2019) Hachimoji DNA and RNA: a genetic system with eight building blocks. *Science (New York, N.Y.)* **363**, 884–887.

Hoshino H, Kasahara Y, Fujita H, Kuwahara M, Morihiro K, Tsunoda SI and Obika S (2016) Consecutive incorporation of functionalized nucleotides with amphiphilic side chains by novel KOD polymerase mutant. *Bioorganic & Medicinal Chemistry Letters* **26**, 530–533.

Hottiger M, Podust VN, Thimmig RL, Mchenry C and Hübscher U (1994) Strand displacement activity of the human immunodeficiency virus type 1 reverse transcriptase heterodimer and its individual subunits. *Journal of Biological Chemistry* **269**, 986–991.

Houlihan G, Arangundy-Franklin S and Holliger P (2017) Engineering and application of polymerases for synthetic genetics. *Current Opinion in Biotechnology* **48**, 168–179.

Hutchison CA 3rd, Chuang RY, Noskov VN, Assad-Garcia N, Deerinck TJ, Ellisman MH, Gill J, Kannan K, Karas BJ, Ma L, Pelletier JF, Qi ZQ, Richter RA, Strychalski EA, Sun L, Suzuki Y, Tsvetanova B, Wise KS, Smith HO, Glass JI, Merryman C, Gibson DG and Venter JC (2016) Design and synthesis of a minimal bacterial genome. *Science (New York, N.Y.)* **351**, aad6253.

Hwang GT and Romesberg FE (2008) Unnatural substrate repertoire of A, B, and X family DNA polymerases. *Journal of the American Chemical Society* **130**, 14872–14882.

Ito J and Braithwaite DK (1991) Compilation and alignment of DNA polymerase sequences. *Nucleic Acids Research* **19**, 4045–4057.

Jackson LN, Chim N, Shi C and Chaput JC (2019) Crystal structures of a natural DNA polymerase that functions as an XNA reverse transcriptase. *Nucleic Acids Research* **47**, 6973–6983.

Jager S and Famulok M (2004) Generation and enzymatic amplification of high-density functionalized DNA double strands. *Angewandte Chemie* **43**, 3337–3340.

Jager S, Rasched G, Kornreich-Leshem H, Engeser M, Thum O and Famulok M (2005) A versatile toolbox for variable DNA functionalization at high density. *Journal of the American Chemical Society* **127**, 15071–15082.

Jash B and Muller J (2017) Metal-mediated base pairs: from characterization to application. *Chemistry* **23**, 17166–17178.

Jayasena SD (1999) Aptamers: an emerging class of molecules that rival antibodies in diagnostics. *Clinical Chemistry* **45**, 1628–1650.

Jestin JL, Kristensen P and Winter G (1999) A method for the selection of catalytic activity using phage display and proximity coupling. *Angewandte Chemie* **38**, 1124–1127.

Ji JP and Loeb LA (1992) Fidelity of HIV-1 reverse transcriptase copying RNA in vitro. *Biochemistry* **31**, 954–958.

Johnson KA (2008) Role of induced fit in enzyme specificity: a molecular forward/reverse switch. *Journal of Biological Chemistry* **283**, 26297–26301.

Johnson SJ, Taylor JS and Beese LS (2003) Processive DNA synthesis observed in a polymerase crystal suggests a mechanism for the prevention of frameshift mutations. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 3895–3900.

Jones MD and Foulkes NS (1989) Reverse transcription of mRNA by *Thermus aquaticus* DNA polymerase. *Nucleic Acids Research* **17**, 8387–8388.

Joyce GF (2004) Directed evolution of nucleic acid enzymes. *Annual Review of Biochemistry* **73**, 791–836.

Joyce GF (2012) Toward an alternative biology. *Science (New York, N.Y.)* **336**, 307–308.

Jung KH and Marx A (2005) Nucleotide analogues as probes for DNA polymerases. *Cellular and Molecular Life Sciences* **62**, 2080–2091.

Kang M, Heuberger B, Chaput JC, Switzer C and Feigon J (2012) Solution structure of a parallel-stranded oligoisoguanine DNA pentaplex formed by d(T(iG)(4) T) in the presence of Cs(+) ions. *Angewandte Chemie* **51**, 7952–7955.

https://doi.org/10.1017/S0033583520000050 Published online by Cambridge University Press

Kaul C, Muller M, Wagner M, Schneider S and Carell T (2011) Reversible bond formation enables the replication and amplification of a crosslinking salen complex as an orthogonal base pair. *Nature Chemistry* **3**, 794–800.

Keefe AD and Cload ST (2008) SELEX with modified nucleotides. *Current Opinion in Chemical Biology* **12**, 448–456.

Keefe AD, Pai S and Ellington AD (2010) Aptamers as therapeutics. *Nature Reviews Drug Discovery* **9**, 537–550.

Kempeneers V, Renders M, Froeyen M and Herdewijn P (2005) Investigation of the DNA-dependent cyclohexenyl nucleic acid polymerization and the cyclohexenyl nucleic acid-dependent DNA polymerization. *Nucleic Acids Research* **33**, 3828–3836.

Kermekchiev MB, Tzekov A and Barnes WM (2003) Cold-sensitive mutants of Taq DNA polymerase provide a hot start for PCR. *Nucleic Acids Research* **31**, 6139–6147.

Kermekchiev MB, Kirilova LI, Vail EE and Barnes WM (2009) Mutants of Taq DNA polymerase resistant to PCR inhibitors allow DNA amplification from whole blood and crude soil samples. *Nucleic Acids Research* **37**, e40.

Kiefer JR, Mao C, Braman JC and Beese LS (1998) Visualizing DNA replication in a catalytically active *Bacillus* DNA polymerase crystal. *Nature* **391**, 304–307.

Kielkowski P, Fanfrlik J and Hocek M (2014) 7-Aryl-7-deazaadenine 2′-deoxyribonucleoside triphosphates (dNTPs): better substrates for DNA polymerases than dATP in competitive incorporations. *Angewandte Chemie* **53**, 7552–7555.

Kim EK and Switzer C (2013) Polymerase recognition of a Watson–Crick-like metal-mediated base pair: purine-2,6-dicarboxylatecopper(II) pyridine. *ChemBioChem* **14**, 2403–2407.

Kimoto M, Yamashige R, Matsunaga K-I, Yokoyama S and Hirao I (2013) Generation of high affinity DNA aptamers using an expanded genetic alphabet. *Nature Biotechnology* **31**, 453–457.

Kimoto M, Nakamura M and Hirao I (2016) Post-ExSELEX stabilization of an unnatural-base DNA aptamer targeting VEGF165 toward pharmaceutical applications. *Nucleic Acids Research* **44**, 7487–7494.

Klenow H and Henningsen I (1970) Selective elimination of exonuclease activity of deoxyribonucleic acid polymerase from *Escherichia-coli*-B by limited proteolysis. *Proceedings of the National Academy of Sciences of the United States of America* **65**, 168–175.

Kobayashi T, Takezawa Y, Sakamoto A and Shionoya M (2016) Enzymatic synthesis of ligand-bearing DNAs for metal-mediated base pairing utilising a template-independent polymerase. *Chemical Communications* **52**, 3762–3765.

Kool ET (2002) Active site tightness and substrate fit in DNA replication. *Annual Review of Biochemistry* **71**, 191–219.

Kropp HM, Betz K, Wirth J, Diederichs K and Marx A (2017) Crystal structures of ternary complexes of archaeal B-family DNA polymerases. *PLoS ONE* **12**, e0188005.

Kropp HM, Durr SL, Peter C, Diederichs K and Marx A (2018) Snapshots of a modified nucleotide moving through the confines of a DNA polymerase. *Proceedings of the National Academy of Sciences of the United States of America* **115**, 9992–9997.

Kropp HM, Diederichs K and Marx A (2019) The structure of an archaeal B-family DNA polymerase in complex with a chemically modified nucleotide. *Angewandte Chemie* **58**, 5457–5461.

Kunkel TA (2004) DNA Replication fidelity. *Journal of Biological Chemistry* **279**, 16895–16898.

Kunkel TA and Bebenek K (2000) DNA Replication fidelity. *Annual Review of Biochemistry* **69**, 497–529.

Kunkel TA, Loeb LA and Goodman MF (1984) On the fidelity of DNA replication. The accuracy of T4 DNA polymerases in copying phi X174 DNA in vitro. *Journal of Biological Chemistry* **259**, 1539–1545.

Kuwahara M, Nagashima J, Hasegawa M, Tamura T, Kitagata R, Hanawa K, Hososhima S, Kasamatsu T, Ozaki H and Sawai H (2006) Systematic characterization of 2′-deoxynucleoside-5′-triphosphate analogs as substrates for DNA polymerases by polymerase chain reaction and kinetic studies on enzymatic production of modified DNA. *Nucleic Acids Research* **34**, 5383–5394.

Laos R, Shaw R, Leal NA, Gaucher E and Benner S (2013) Directed evolution of polymerases to accept nucleotides with nonstandard hydrogen bond patterns. *Biochemistry* **52**, 5288–5294.

Lapa SA, Chudinov AV and Timofeev EN (2016) The toolbox for modified aptamers. *Molecular Biotechnology* **58**, 79–92.

Larsen AC, Dunn MR, Hatch A, Sau SP, Youngbull C and Chaput JC (2016) A general strategy for expanding polymerase function by droplet microfluidics. *Nature Communications* **7**, 11235.

Lawyer FC, Stoffel S, Saiki RK, Myambo K, Drummond R and Gelfand DH (1989) Isolation, characterization, and expression in *Escherichia coli* of the DNA polymerase gene from *Thermus aquaticus*. *Journal of Biological Chemistry* **264**, 6427–6437.

Lawyer FC, Stoffel S, Saiki RK, Chang SY, Landre PA, Abramson RD and Gelfand DH (1993) High-level expression, purification, and enzymatic characterization of full-length *Thermus aquaticus* DNA polymerase and a truncated form deficient in 5′ to 3′ exonuclease activity. *PCR Methods and Applications* **2**, 275–287.

Leal NA, Kim H-J, Hoshika S, Kim M-J, Carrigan MA and Benner SA (2015) Transcription, reverse transcription, and analysis of RNA containing artificial genetic components. *ACS Synthetic Biology*, **4**, 407–413.

Leconte AM, Chen L and Romesberg FE (2005) Polymerase evolution: efforts toward expansion of the genetic code. *Journal of the American Chemical Society* **127**, 12470–12471.

Lee HH, Kalhor R, Goela N, Bolot J and Church GM (2019) Terminator-free template-independent enzymatic DNA synthesis for digital information storage. *Nature Communications* **10**, 2383.

Lehman IR, Bessman MJ, Simms ES and Kornberg A (1958) Enzymatic synthesis of deoxyribonucleic acid. I. Preparation of substrates and partial purification of an enzyme from *Escherichia coli*. *Journal of Biological Chemistry* **233**, 163–170.

Li L, Degardin M, Lavergne T, Malyshev DA, Dhami K, Ordoukhanian P and Romesberg FE (2014) Natural-like replication of an unnatural base pair for the expansion of the genetic alphabet and biotechnology applications. *Journal of the American Chemical Society* **136**, 826–829.

Litosh VA, Wu W, Stupi BP, Wang J, Morris SE, Hersh MN and Metzker ML (2011) Improved nucleotide selectivity and termination of 3′-OH unblocked reversible terminators by molecular tuning of 2-nitrobenzyl alkylated HOMedU triphosphates. *Nucleic Acids Research* **39**, e39.

Liu Z, Chen T and Romesberg FE (2017) Evolved polymerases facilitate selection of fully 2′-OMe-modified aptamers. *Chemical Science* **8**, 8179–8182.

Loakes D, Gallego J, Pinheiro VB, Kool ET and Holliger P (2009) Evolving a polymerase for hydrophobic base analogues. *Journal of the American Chemical Society* **131**, 14827–14837.

Loeb LA and Monnat RJ Jr (2008) DNA polymerases and human disease. *Nature Reviews Genetics* **9**, 594–604.

Loeb LA, Tartof KD and Travaglini EC (1973) Copying natural RNAs with *Escherichia-coli* DNA polymerase-I. *Nature New Biology* **242**, 66–69.

Ludmann S and Marx A (2016) Getting it right: how DNA polymerases select the right nucleotide. *Chimia* **70**, 203–206.

Malyshev DA and Romesberg FE (2015) The expanded genetic alphabet. *Angewandte Chemie* **54**, 11930–11944.

Malyshev DA, Seo YJ, Ordoukhanian P and Romesberg FE (2009) PCR with an expanded genetic alphabet. *Journal of the American Chemical Society* **131**, 14620–14621.

Malyshev DA, Pfaff DA, Ippoliti SI, Hwang GT, Dwyer TJ and Romesberg FE (2010) Solution structure, mechanism of replication, and optimization of an unnatural base pair. *Chemistry* **16**, 12650–12659.

Martinez CI, Ansari MA, Gibbs R and Burgess K (1997) Acyclic nucleoside triphosphate analogs as terminators in biocatalytic DNA replication. *Bioorganic & Medicinal Chemistry Letters* **7**, 3013–3016.

Marx A, Macwilliams MP, Bickle TA, Schwitter U and Giese B (1997) 4′-Acylated thymidines: a new class of DNA chain terminators and photocleavable DNA building blocks. *Journal of the American Chemical Society* **119**, 1131–1132.

Marx A, Spichty M, Amacker M, Schwitter U, Hubscher U, Bickle TA, Maga G and Giese B (1999) Probing interactions between HIV-1 reverse transcriptase and its DNA substrate with backbone-modified nucleotides. *Chemistry & Biology* **6**, 111–116.

Matsuda S, Fillo JD, Henry AA, Rai P, Wilkens SJ, Dwyer TJ, Geierstanger BH, Wemmer DE, Schultz PG, Spraggon G and Romesberg FE (2007)

https://doi.org/10.1017/S0033583520000050 Published online by Cambridge University Press

Efforts toward expansion of the genetic alphabet: structure and replication of unnatural base pairs. *Journal of the American Chemical Society* **129**, 10466–10473.

Matsunaga K, Kimoto M, Hanson C, Sanford M, Young HA and Hirao I (2015) Architecture of high-affinity unnatural-base DNA aptamers toward pharmaceutical applications. *Scientific Reports* **5**, 18478.

Matsunaga KI, Kimoto M and Hirao I (2017) High-affinity DNA aptamer generation targeting von Willebrand factor A1-domain by genetic alphabet expansion for systematic evolution of ligands by exponential enrichment using two types of libraries composed of five different bases. *Journal of the American Chemical Society* **139**, 324–334.

Mattila P, Korpela J, Tenkanen T and Pitkanen K (1991) Fidelity of DNA synthesis by the Thermococcus litoralis DNA polymerase-—an extremely heat stable enzyme with proofreading activity. *Nucleic Acids Research* **19**, 4967–4973.

McCullum EO and Chaput JC (2009) Transcription of an RNA aptamer by a DNA polymerase. *Chemical Communications*, 2938–2940.

Mehta AP, Wang Y, Reed SA, Supekova L, Javahishvili T, Chaput JC and Schultz PG (2018) Bacterial genome containing chimeric DNA-RNA sequences. *Journal of the American Chemical Society* **140**, 11464–11473.

Mei H, Liao J-Y, Jimenez RM, Wang Y, Bala S, Mccloskey C, Switzer C and Chaput JC (2018) Synthesis and evolution of a threose nucleic acid aptamer bearing 7-deaza-7-substituted guanosine residues. *Journal of the American Chemical Society* **140**, 5706–5713.

Millar D, Christova Y and Holliger P (2015) A polymerase engineered for bisulfite sequencing. *Nucleic Acids Research* **43**, e155.

Miller H and Grollman AP (1997) Kinetics of DNA polymerase I (Klenow fragment exo-) activity on damaged DNA templates: effect of proximal and distal template damage on DNA synthesis. *Biochemistry* **36**, 15336–15342.

Mitsui T, Kitamura A, Kimoto M, To T, Sato A, Hirao I and Yokoyama S (2003) An unnatural hydrophobic base pair with shape complementarity between pyrrole-2-carbaldehyde and 9-methylimidazo[(4,5)-b]pyridine. *Journal of the American Chemical Society* **125**, 5298–5307.

Miura M, Tanigawa C, Fujii Y and Kaneko S (2013) Comparison of six commercially-available DNA polymerases for direct PCR. *Revista Do Instituto De Medicina Tropical De São Paulo* **55**, 401–406.

Morales JC and Kool ET (1998) Efficient replication between non-hydrogen-bonded nucleoside shape analogs. *Nature Structural & Molecular Biology* **5**, 950–954.

Morales JC and Kool ET (1999) Minor groove interactions between polymerase and DNA: more essential to replication than Watson–Crick hydrogen bonds? *Journal of the American Chemical Society* **121**, 2323–2324.

Moran S, Ren RX and Kool ET (1997*a*) A thymidine triphosphate shape analog lacking Watson–Crick pairing ability is replicated with high sequence selectivity. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 10506–10511.

Moran S, Ren RX-F, Rumney S and Kool ET (1997*b*) Difluorotoluene, a nonpolar isostere for thymine, codes specifically and efficiently for adenine in DNA replication. *Journal of the American Chemical Society* **119**, 2056–2057.

Motre A, Li Y and Kong H (2008) Enhancing helicase-dependent amplification by fusing the helicase with the DNA polymerase. *Gene* **420**, 17–22.

Myers TW and Gelfand DH (1991) Reverse transcription and DNA amplification by a *Thermus thermophilus* DNA polymerase. *Biochemistry* **30**, 7661–7666.

Nakama T, Takezawa Y, Sasaki D and Shionoya M (2020) Allosteric regulation of DNAzyme activities through intrastrand transformation induced by Cu(II)-mediated artificial base pairing. *Journal of the American Chemical Society* **142**, 10153–10162.

Nakamura T, Zhao Y, Yamagata Y, Hua YJ and Yang W (2012) Watching DNA polymerase eta make a phosphodiester bond. *Nature* **487**, 196–201.

Ng MMP, Benseler F, Tuschl T and Eckstein F (1994) Isoguanosine substitution of conserved adenosines in the hammerhead ribozyme. *Biochemistry* **33**, 12119–12126.

Ng EWM, Shima DT, Calias P, Cunningham ET, Guyer DR and Adamis AP (2006) Pegaptanib, a targeted anti-VEGF aptamer for ocular vascular disease. *Nature Reviews Drug Discovery* **5**, 123–132.

Nikoomanzar A, Dunn MR and Chaput JC (2017) Evaluating the rate and substrate specificity of laboratory evolved XNA polymerases. *Analytical Chemistry* **89**, 12622–12625.

Nikoomanzar A, Vallejo D and Chaput JC (2019) Elucidating the determinants of polymerase specificity by microfluidic-based deep mutational scanning. *ACS Synthetic Biology* **8**, 1421–1429.

Nikoomanzar A, Vallejo D, Yik E and Chaput JC (2020) Programmed allelic mutagenesis of a DNA polymerase with single amino acid resolution. *ACS Synthetic Biology*. in press: doi.org/10.1021/acssynbio.0c00236.

Nimjee SM, Rusconi CP and Sullenger BA (2005) Aptamers: an emerging class of therapeutics. *Annual Review of Medicine* **56**, 555–583.

Noronha AM, Wilds CJ, Lok CN, Viazovkina K, Arion D, Parniak MA and Damha MJ (2000) Synthesis and biophysical properties of arabinonucleic acids (ANA): circular dichroic spectra, melting temperatures, and ribonuclease H susceptibility of ANA.RNA hybrid duplexes. *Biochemistry* **39**, 7050–7062.

Obeid S, Blatter N, Kranaster R, Schnur A, Diederichs K, Welte W and Marx A (2010) Replication through an abasic DNA lesion: structural basis for adenine selectivity. *The EMBO Journal* **29**, 1738–1747.

Obeid S, Welte W, Diederichs K and Marx A (2012) Amino acid templating mechanisms in selection of nucleotides opposite abasic sites by a family a DNA polymerase. *Journal of Biological Chemistry* **287**, 14099–14108.

Ogawa A, Wu Y, Mcminn D, Liu J, Schultz PG and Romesberg FE (2000) Efforts toward the expansion of the genetic alphabet: information storage and replication with unnatural hydrophobic base pairs. *Journal of the American Chemical Society* **122**, 3274–3287.

Okamoto I, Miyatake Y, Kimoto M and Hirao I (2016) High fidelity, efficiency and functionalization of Ds-Px unnatural base pairs in PCR amplification for a genetic alphabet expansion system. *ACS Synthetic Biology* **5**, 1220–1230.

Ong JL, Loakes D, Jaroslawski S, Too K and Holliger P (2006) Directed evolution of DNA polymerase, RNA polymerase and reverse transcriptase activity in a single polypeptide. *Journal of Molecular Biology* **361**, 537–550.

Ostroff RM, Bigbee WL, Franklin W, Gold L, Mehan M, Miller YE, Pass HI, Rom WN, Siegfried JM, Stewart A, Walker JJ, Weissfeld JL, Williams S, Zichi D and Brody EN (2010) Unlocking biomarker discovery: large scale application of aptamer proteomic technology for early detection of lung cancer. *PLoS ONE* **5**, e15003.

Palluk S, Arlow DH, De Rond T, Barthel S, Kang JS, Bector R, Baghdassarian HM, Truong AN, Kim PW, Singh AK, Hillson NJ and Keasling JD (2018) De novo DNA synthesis using polymerase-nucleotide conjugates. *Nature Biotechnology* **36**, 645–650.

Patel PH, Kawate H, Adman E, Ashbach M and Loeb LA (2001) A single highly mutable catalytic site amino acid is critical for DNA polymerase fidelity. *Journal of Biological Chemistry* **276**, 5044–5051.

Pavlov AR, Belova GI, Kozyavkin SA and Slesarev AI (2002) Helix-hairpin-helix motifs confer salt resistance and processivity on chimeric DNA polymerases. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 13510–13515.

Pelletier H, Sawaya MR, Kumar A, Wilson SH and Kraut J (1994) Structures of ternary complexes of rat DNA polymerase beta, a DNA template-primer, and ddCTP. *Science (New York, N.Y.)* **264**, 1891–1903.

Piccirilli JA, Krauch T, Moroney SE and Benner SA (1990) Enzymatic incorporation of a new base pair into DNA and RNA extends the genetic alphabet. *Nature* **343**, 33–37.

Pinheiro VB, Taylor AI, Cozens C, Abramov M, Renders M, Zhang S, Chaput JC, Wengel J, Peak-Chew SY, Mclaughlin SH, Herdewijn P and Holliger P (2012) Synthetic genetic polymers capable of heredity and evolution. *Science (New York, N.Y.)* **336**, 341–344.

Pochet S, Kaminski PA, Van Aerschot A, Herdewijn P and Marliere P (2003) Replication of hexitol oligonucleotides as a prelude to the propagation of a third type of nucleic acid in vivo. *Comptes Rendus Biologies* **326**, 1175–1184.

Potapov V and Ong JL (2017) Examining sources of error in PCR by single-molecule sequencing. *PLoS ONE* **12**, e0169774.

Potter J, Zheng W and Lee J (2003) Thermal stability and cDNA synthesis capability of SuperScript III reverse transcriptase. *Focus* **25**, 19–24.

https://doi.org/10.1017/S0033583520000050 Published online by Cambridge University Press

Price AK and Paegel BM (2016) Discovery in droplets. *Analytical Chemistry* **88**, 339–353.

Ramsay N, Jemth A-S, Brown A, Crampton N, Dear P and Holliger P (2010) CyDNA: synthesis and replication of highly Cy-dye substituted DNA by an evolved polymerase. *Journal of the American Chemical Society* **132**, 5096–5104.

Rangel AE, Chen Z, Ayele TM and Heemstra JM (2018) In vitro selection of an XNA aptamer capable of small-molecule recognition. *Nucleic Acids Research* **46**, 8057–8068.

Reardon JE and Miller WH (1990) Human immunodeficiency virus reverse transcriptase. Substrate and inhibitor kinetics with thymidine 5'-triphosphate and 3'-azido-3'-deoxythymidine 5'-triphosphate. *Journal of Biological Chemistry* **265**, 20302–20307.

Renders M, Abramov M, Froeyen M and Herdewijn P (2009) Polymerase-catalysed incorporation of glucose nucleotides into a DNA duplex. *Chemistry* **15**, 5463–5470.

Rice KP, Chaput JC, Cox MM and Switzer C (2000) Reca protein promotes strand exchange with DNA substrates containing isoguanine and 5-methyl isocytosine. *Biochemistry* **39**, 10177–10188.

Riggs MG, Tudor S, Sivaram M and McDonough SH (1996) Construction of single amino acid substitution mutants of cloned *Bacillus stearothermophilus* DNA polymerase I which lack 5′-->3′ exonuclease activity. *Biochimica et Biophysica Acta* **1307**, 178–186.

Roberts RW and Szostak JW (1997) RNA-peptide fusions for the in vitro selection of peptides and proteins. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 12297–12302.

Roberts C, Bandaru R and Switzer C (1997*a*) Theoretical and experimental study of isoguanine and isocytosine: base pairing in an expanded genetic system. *Journal of the American Chemical Society* **119**, 4640–4649.

Roberts C, Chaput JC and Switzer C (1997*b*) Beyond guanine quartets: cation-induced formation of homogeneous and chimeric DNA tetraplexes incorporating iso-guanine and guanine. *Chemistry & Biology* **4**, 899–908.

Robinson H, Gao YG, Bauer C, Roberts C, Switzer C and Wang AHJ (1998) 2′-Deoxyisoguanosine adopts more than one tautomer to form base pairs with thymidine observed by high-resolution crystal structure analysis. *Biochemistry* **37**, 10897–10905.

Rothlisberger P, Levi-Acobas F, Sarac I, Marliere P, Herdewijn P and Hollenstein M (2017) On the enzymatic incorporation of an imidazole nucleotide into DNA. *Organic and Biomolecular Chemistry* **15**, 4449–4455.

Rusconi CP, Roberts JD, Pitoc GA, Nimjee SM, White RR, Quick G, Scardino E, Fay WP and Sullenger BA (2004) Antidote-mediated control of an anticoagulant aptamer in vivo. *Nature Biotechnology* **22**, 1423–1428.

Saiki R, Gelfand D, Stoffel S, Scharf S, Higuchi R, Horn G, Mullis K and Erlich H (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science (New York, N.Y.)* **239**, 487–491.

Sambrook J, Fritsch EF, and Maniatis T (1989) *Molecular Cloning: A Laboratory Manual*, 2nd Edn. Cold Spring Harbor: Cold Spring Harbor Laboratory Press.

Sanger F, Nicklen S and Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America* **74**, 5463–5467.

Santoro SW and Joyce GF (1997) A general purpose RNA-cleaving DNA enzyme. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 4262–4266.

Santoro SW and Joyce GF (1998) Mechanism and utility of an RNA-cleaving DNA enzyme. *Biochemistry* **37**, 13330–13342.

Schliwa M and Woehlke G (2003) Molecular motors. *Nature* **422**, 759–765.

Schmidt M (2010) Xenobiology: a new form of life as the ultimate biosafety tool. *BioEssays* **32**, 322–331.

Schöning KU, Scholz P, Guntha S, Wu X, Krishnamurthy R and Eschenmoser A (2000) Chemical etiology of nucleic acid structure: the a-threofuranosyl-(3′-->2′) oligonucleotide system. *Science (New York, N.Y.)* **290**, 1347–1351.

Schultz HJ, Gochi AM, Chia HE, Ogonowsky AL, Chiang S, Filipovic N, Weiden AG, Hadley EE, Gabriel SE and Leconte AM (2015) Taq DNA polymerase mutants and 2′-modified sugar recognition. *Biochemistry* **54**, 5999–6008.

Seela F, He Y and Wei CF (1999) Parallel-stranded oligonucleotide duplexes containing 5-methylisocytosine-guanine and isoguanine-cytosine base pairs. *Tetrahedron* **55**, 9481–9500.

Seeman NC and Sleiman HF (2018) DNA nanotechnology. *Nature Reviews Materials* **3**, article number 17068.

Sefah K, Yang Z, Bradley KM, Hoshika S, Jimenez E, Zhang L, Zhu G, Shanker S, Yu F, Turek D, Tan W and Benner SA (2014) In vitro selection with artificial expanded genetic information systems. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 1449–1454.

Shandilya H, Griffiths K, Flynn EK, Astatke M, Shih PJ, Lee JE, Gerard GF, Gibbs MD and Bergquist PL (2004) Thermophilic bacterial DNA polymerases with reverse-transcriptase activity. *Extremophiles* **8**, 243–251.

Shao Q, Chen T, Sheng K, Liu Z, Zhang Z and Romesberg FE (2020) Selection of aptamers with large hydrophobic 2′-substituents. *Journal of the American Chemical Society* **142**, 2125–2128.

Shi C, Shen X, Niu S and Ma C (2015) Innate reverse transcriptase activity of DNA polymerase for isothermal RNA direct detection. *Journal of the American Chemical Society* **137**, 13804–13806.

Singh I, Laos R, Hoshika S, Benner SA and Georgiadis MM (2018) Snapshots of an evolved DNA polymerase pre- and post-incorporation of an unnatural nucleotide. *Nucleic Acids Research* **46**, 7977–7988.

Sinha S, Kim PH and Switzer C (2004) 2′,5′-Linked DNA is a template for polymerase-directed DNA synthesis. *Journal of the American Chemical Society* **126**, 40–41.

Slesarev AI, Stetter KO, Lake JA, Gellert M, Krah R and Kozyavkin SA (1993) DNA topoisomerase V is a relative of eukaryotic topoisomerase I from a hyperthermophilic prokaryote. *Nature* **364**, 735–737.

Smith G and Petrenko V (1997) Phage display. *Chemistry Reviews* **97**, 391–410.

Steitz TA (1999) DNA polymerases: structural diversity and common mechanisms. *Journal of Biological Chemistry* **274**, 17395–17398.

Steitz TA, Smerdon SJ, Jager J and Joyce CM (1994) A unified polymerase mechanism for nonhomologous DNA and RNA polymerases. *Science (New York, N.Y.)* **266**, 2022–2025.

Summerer D, Rudinger NZ, Detmer I and Marx A (2005) Enhanced fidelity in mismatch extension by DNA polymerase through directed combinatorial enzyme design. *Angewandte Chemie* **44**, 4712–4715.

Switzer C, Moroney SE and Benner SA (1989) Enzymatic incorporation of a new base pair into DNA and RNA. *Journal of the American Chemical Society* **111**, 8322–8323.

Szostak JW (1992) In vitro genetics. *Trends in Biochemical Sciences* **17**, 89–93.

Tabor S and Richardson CC (1989*a*) Effect of manganese ions on the incorporation of dideoxynucleotides by bacteriophage T7 DNA polymerase and *Escherichia coli* DNA polymerase I. *Proceedings of the National Academy of Sciences of the United States of America* **86**, 4076–4080.

Tabor S and Richardson CC (1989*b*) Selective inactivation of the exonuclease activity of bacteriophage T7 DNA polymerase by in vitro mutagenesis. *Journal of Biological Chemistry* **264**, 6447–6458.

Tabor S and Richardson CC (1995) A single residue in DNA-polymerases of the *Escherichia-coli* DNA-polymerase-I family is critical for distinguishing between deoxyribonucleotides and dideoxyribonucleotides. *Proceedings of the National Academy of Sciences of the United States of America* **92**, 6339–6343.

Tae EL, Wu Y, Xia G, Schultz PG and Romesberg FE (2001) Efforts toward expansion of the genetic alphabet: replication of DNA with three base pairs. *Journal of the American Chemical Society* **123**, 7439–7440.

Takata K, Shimizu T, Iwai S and Wood RD (2006) Human DNA polymerase N (POLN) is a low fidelity enzyme capable of error-free bypass of 5S-thymine glycol. *Journal of Biological Chemistry* **281**, 23445–23455.

Taylor AI, Pinheiro VB, Smola MJ, Morgunov AS, Peak-Chew S, Cozens C, Weeks KM, Herdewijn P and Holliger P (2015) Catalysts from synthetic genetic polymers. *Nature* **518**, 427–430.

https://doi.org/10.1017/S0033583520000050 Published online by Cambridge University Press

Thirunavukarasu D, Chen T, Liu Z, Hongdilokkul N and Romesberg FE (2017) Selection of 2′-fluoro-modified aptamers with optimized properties. *Journal of the American Chemical Society* **139**, 2892–2895.

Tor Y and Dervan PB (1993) Site-specific enzymatic incorporation of an unnatural base, N(6)-(6-aminohexyl)Isoguanosine, into RNA. *Journal of the American Chemical Society* **115**, 4461–4467.

Traut TW (1994) Physiological concentrations of purines and pyrimidines. *Molecular and Cellular Biochemistry* **140**, 1–22.

Tsai C-H, Chen J and Szostak JW (2007) Enzymatic synthesis of DNA on glycerol nucleic acid templates without stable duplex formation between the product and template. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 14598–14603.

Tse WT and Forget BG (1990) Reverse transcription and direct amplification of cellular RNA transcripts by Taq polymerase. *Gene* **88**, 293–296.

Vallejo D, Nikoomanzar A, Paegel BM and Chaput JC (2019) Fluorescence-activated droplet sorting for single-cell directed evolution. *ACS Synthetic Biology* **8**, 1430–1440.

Vaught JD, Bock C, Carter J, Fitzwater T, Otis M, Schneider D, Rolando J, Waugh S, Wilcox SK and Eaton BE (2010) Expanding the chemistry of DNA for in vitro selection. *Journal of the American Chemical Society* **132**, 4141–4151.

Vichier-Guerre S, Ferris S, Auberger N, Mahiddine K and Jestin JL (2006) A population of thermostable reverse transcriptases evolved from *Thermus aquaticus* DNA polymerase I by phage display. *Angewandte Chemie* **45**, 6133–6137.

Vincent M, Xu Y and Kong H (2004) Helicase-dependent isothermal DNA amplification. *EMBO Reports* **5**, 795–800.

Wang Y, Prosen DE, Mei L, Sullivan JC, Finney M and Vander Horn PB (2004) A novel strategy to engineer DNA polymerases for enhanced processivity and improved performance in vitro. *Nucleic Acids Research* **32**, 1197–1207.

Wang X, Hoshika S, Peterson RJ, Kim MJ, Benner SA and Kahn JD (2017) Biophysics of artificially expanded genetic information systems. Thermodynamics of DNA duplexes containing matches and mismatches involving 2-amino-3-nitropyridin-6-one (Z) and imidazo [1,2-a]-1,3,5-triazin-4(8H)one (P). *ACS Synthetic Biology* **6**, 782–792.

Wang Y, Liu E, Lam CH and Perrin DM (2018*a*) A densely modified M(2+)-independent DNAzyme that cleaves RNA efficiently with multiple catalytic turnover. *Chemical Science* **9**, 1813–1821.

Wang Y, Ngor AK, Nikoomanzar A and Chaput JC (2018*b*) Evolution of a general RNA-cleaving FANA enzyme. *Nature Communications* **9**, 5067.

Wang Y, Vorperian A, Shehabat M and Chaput JC (2020) Evaluating the catalytic potential of a general RNA-cleaving FANA enzyme. *ChemBioChem* **21**, 1001–1006.

Watson JD and Crick FH (1953) Genetical implications of the structure of deoxyribonucleic acid. *Nature* **171**, 964–967.

Williams JG, Steffens DL, Anderson JP, Urlacher TM, Lamb DT, Grone DL and Egelhoff JC (2008) An artificial processivity clamp made with streptavidin facilitates oriented attachment of polymerase-DNA complexes to surfaces. *Nucleic Acids Research* **36**, e121.

Williams SA, Kivimaki M, Langenberg C, Hingorani AD, Casas JP, Bouchard C, Jonasson C, Sarzynski MA, Shipley MJ, Alexander L, Ash J, Bauer T, Chadwick J, Datta G, Delisle RK, Hagar Y, Hinterberg M, Ostroff R, Weiss S, Ganz P and Wareham NJ (2019) Plasma protein patterns as comprehensive indicators of health. *Nature Medicine* **25**, 1851–1857.

Wilson DS and Szostak JW (1999) In vitro selection of functional nucleic acids. *Annual Review of Biochemistry* **68**, 611–647.

Wojciechowski F and Leumann CJ (2011) Alternative DNA base-pairs: from efforts to expand the genetic code to potential material applications. *Chemical Society Reviews* **40**, 5669–5679.

Wu EY and Beese LS (2011) The structure of a high fidelity DNA polymerase bound to a mismatched nucleotide reveals an 'ajar' intermediate conformation in the nucleotide selection mechanism. *Journal of Biological Chemistry* **286**, 19758–19767.

Wu WJ, Yang W and Tsai MD (2017) How DNA polymerases catalyse replication and repair with contrasting fidelity. *Nature Reviews Chemistry* **1**, 0068.

Wuite GJ, Smith SB, Young M, Keller D and Bustamante C (2000) Single-molecule studies of the effect of template tension on T7 DNA polymerase activity. *Nature* **404**, 103–106.

Xia G, Chen L, Sera T, Fa M, Schultz PG and Romesberg FE (2002) Directed evolution of novel polymerase activities: mutation of a DNA polymerase into an efficient RNA polymerase. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 6597–6602.

Yamagami T, Ishino S, Kawarabayasi Y and Ishino Y (2014) Mutant Taq DNA polymerases with improved elongation ability as a useful reagent for genetic engineering. *Frontiers in Microbiology* **5**, 461.

Yamashige R, Kimoto M, Takezawa Y, Sato A, Mitsui T, Yokoyama S and Hirao I (2012) Highly specific unnatural base pair systems as a third base pair for PCR amplification. *Nucleic Acids Research* **40**, 2793–2806.

Yamtitch J and Sweasy JB (2010) DNA polymerase family X: function, structure, and cellular roles. *Biochimica et Biophysica Acta* **1804**, 1136–1150.

Yang Z, Sismour AM, Sheng P, Puskar NL and Benner SA (2007) Enzymatic incorporation of a third nucleobase pair. *Nucleic Acids Research* **35**, 4238–4249.

Yang ZY, Chen F, Alvarado JB and Benner SA (2011) Amplification, mutation, and sequencing of a six-letter synthetic genetic system. *Journal of the American Chemical Society* **133**, 15105–15112.

Yin YW and Steitz TA (2002) Structural basis for the transition from initiation to elongation transcription in T7 RNA polymerase. *Science (New York, N.Y.)* **298**, 1387–1395.

Yin YW and Steitz TA (2004) The structural mechanism of translocation and helicase activity in T7 RNA polymerase. *Cell* **116**, 393–404.

Young DD, Lively MO and Deiters A (2010) Activation and deactivation of DNAzyme and antisense function with light for the photochemical regulation of gene expression in mammalian cells. *Journal of the American Chemical Society* **132**, 6183–6193.

Yu H, Zhang S and Chaput JC (2012) Darwinian evolution of an alternative genetic system provides support for TNA as an RNA progenitor. *Nature Chemistry* **4**, 183–187.

Zhang L, Kang M, Xu J and Huang Y (2015*a*) Archaeal DNA polymerases in biotechnology. *Applied Microbiology and Biotechnology* **99**, 6585–6597.

Zhang L, Yang Z, Sefah K, Bradley KM, Hoshika S, Kim MJ, Kim HJ, Zhu G, Jimenez E, Cansiz S, Teng IT, Champanhac C, Mclendon C, Liu C, Zhang W, Gerloff DL, Huang Z, Tan W and Benner SA (2015*b*) Evolution of functional six-nucleotide DNA. *Journal of the American Chemical Society* **137**, 6734–6737.

Zhang LQ, Yang ZY, Trinh TL, Teng IT, Wang S, Bradley KM, Hoshika S, Wu QF, Cansiz S, Rowold DJ, Mclendon C, Kim MS, Wu Y, Cui C, Liu Y, Hou WJ, Stewart K, Wan S, Liu C, Benner SA and Tan WH (2016) Aptamers against cells overexpressing glypican 3 from expanded genetic systems combined with cell engineering and laboratory evolution. *Angewandte Chemie* **55**, 12372–12375.

https://doi.org/10.1017/S0033583520000050 Published online by Cambridge University Press