Title:

Engineering-driven biological insights into DNA polymerase mechanism

Authors and affiliations:

Vitor B. Pinheiro[1,2]

[1]KU Leuven, Rega Institute for Medical Research, Herestraat 49, box 1030, 3000 Leuven, Belgium

[2]KU Leuven, Department of Pharmaceutical Sciences, Herestraat 49, bus 420, 3000 Leuven, Belgium

Corresponding author:

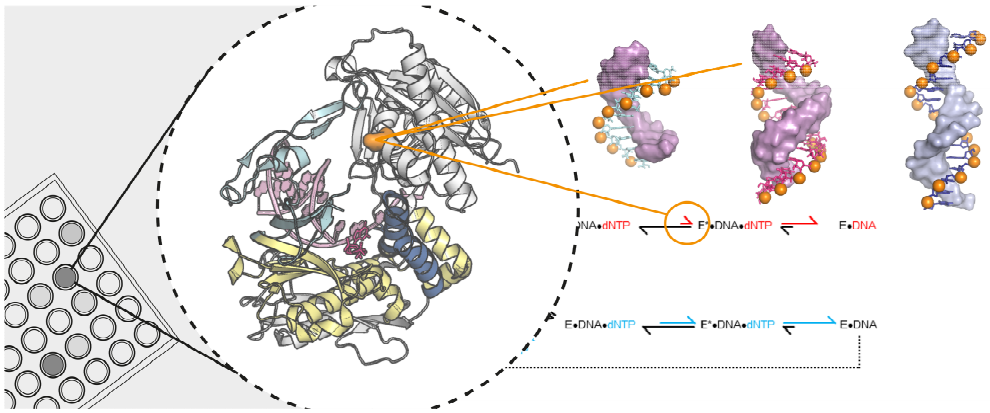
Vitor B. Pinheiro (vitor.pinheiro@kuleuven.be)

Telephone: +32 16 330 257

Abstract:

DNA-dependent DNA polymerases have been extensively studied for over 60 years and lie at the core of multiple biotechnological and diagnostic applications. Nevertheless, these complex molecular machines remain only partially understood. Here we present some evidence on how polymerase engineering for the synthesis and replication of xenobiotic nucleic acids (XNAs) have improved our understanding of these enzymes and how that can be used to gain further insight into their mechanism. Better understanding of the mechanisms of DNA polymerases can accelerate their engineering and we highlight how it is now feasible to use structure- and function-based approaches to systematically and iteratively develop XNA polymerases for increasingly divergent chemistries.

Graphical Abstract:



Highlights:

- Selection platforms used for polymerase engineering can be a powerful tool to probe DNA polymerase mechanism.
- The polymerase field has been shaped by bulk biophysical assays and detailed structures, making identification of mutations that affect polymerase dynamics non-trivial.
- Mutations that stabilize the closed conformation or that reduce the likelihood of finger re-opening decrease polymerase fidelity but greatly enhance the potential of polymerases to synthesise xenobiotic nucleic acids (XNAs).

Introduction

DNA-dependent DNA polymerases are essential to life on Earth. They are the key catalysts available for the molecular replication of DNA and, therefore, they have a central role in the propagation of genetic information. DNA polymerases have been extensively characterised for over 60 years [1], using a wide range of biochemical and biophysical approaches developed, and further supported by substantial structural information.

Some paradigms have become well-established in the field, such as the consensus that all DNA polymerases operate with two catalytic magnesium ions in their active site, and that polymerases can be best grouped by their sequence similarity and conserved motifs into the eight currently described families (A, B, C, D, E, X, Y and RT).

Other areas have remained more contentious, such as the molecular features in polymerases and in triphosphate substrates that account for fidelity of incorporation, or the identification of the rate limiting steps in catalysis.

While the structure of the polymerase active site seems remarkably similar across polymerases of different families and of different phylogenies (Figure 1), it is not an unreasonable assumption that enzymes may have been optimised through evolution to their local role (e.g. replication, repair) and context (e.g. other available polymerases in the cell, regulation). For instance, bacteriophage T4 and RB69 both harbour a β hairpin, whose mutation has direct measurable impact on polymerase fidelity. Nevertheless, similar mutations to disable the equivalent structure in *S. cerevisiae* DNA polymerase δ (also a B-family polymerase), show no significant impact on the overall polymerase fidelity [2].

It is also not unreasonable to argue that DNA polymerases carry out multiple different but related reactions, considering the identity of the incoming triphosphate, the sequence of the template and the nascent duplex. Consequently, the data (and the techniques needed to generate it) required to distinguish a universal mechanism from multiple local models on polymerase mechanism are not currently available.

In all, DNA polymerases remain partially understood enzymes. But partial (and potentially even incorrect) understanding of important aspects of polymerase mechanism has not been an obvious barrier to their engineering – with a wide range of reported examples, from enzymes with altered fidelity or dynamics, to enzymes capable of synthesising unnatural polymers.

This brief review highlights some of the recent advances in the engineering of DNA polymerases (particularly B-family polymerases) for the synthesis and reverse transcription of xenobiotic nucleic acids (XNAs) and how these mutations may yield greater insight into the mechanism of DNA polymerases. Other recent reviews are available that provide a detailed mechanistic and structural analysis of DNA polymerases [3–7].

Bypassing knowledge gaps with focused libraries – the second gate

Engineering approaches, such as directed evolution, are sufficiently powerful to bypass knowledge gaps [8]. They rely on establishing a strong link between phenotype and genotype, ensuring that isolation of the desired function retrieves the gene sequence responsible for it. Selection and screening strategies can be scaled up, with many selection platforms able to sample populations exceeding 10^8 variants. That remains however a fraction of the local sequence neighbourhood of even a small protein: sampling all possible triple mutants of a 100-residue protein leads to more than 10^9 variants.

Clear understanding of the function and mechanism of an enzyme enables the design of focused libraries, targeting diversity onto relevant residues. That approach is meant to increase the likelihood that a polymerase with the desired function is present in the library, and it also maximises the population of desired mutants as a fraction of the total library. Both processes increase the efficiency of the engineering cycle.

However, if the understanding of the enzyme is incomplete or incorrect, it creates a knowledge gap, which can lead to diversity introduced for selection not targeting relevant residues as well as targeting residues with no relevant role.

Early work from multiple groups on the engineering of XNA synthetases (based on thermostable A-family polymerases) capable of synthesising RNA or congeners such as 2'F-RNA and 2'OMe-RNA, focused on diversity introduced in the vicinity of the active site: optimising nucleotide binding and incorporation [9–11]. The resulting enzymes all displayed ribonucleotide incorporation efficiencies comparable to natural substrates but could not extend the nascent chain beyond a handful of incorporations.

Later efforts, using a different polymerase scaffold (B-family archaeal enzymes), identified a region in the thumb domain of the polymerase that was key to changing their substrate specificity [12,13]. Engineering of this second region enabled the synthesis of a range of different XNAs [13] while a single residue was sufficient to convert a DNA polymerase into an efficient RNA polymerase capable of accepting a range of 2' substitutions [12].

Termed a 'second gate' – in addition to the well-established 'steric gate' in the polymerase active site – residue E664 (in *Thermococcus gorgonarius*) has been hypothesised to have a role in the specificity of the polymerase for the nascent duplex. As the polymerase incorporates and elongates the primer strand with ribonucleotides (or other non-cognate substrates), the nascent hybrid duplex shifts away from the canonical B-form double-stranded DNA. Thus, different contacts are needed to ensure the nascent duplex remains a viable substrate for the polymerase.

For RNA synthesis [12], for synthesis of chimeric polymers containing 2'-5' linkages [14] and for reverse transcription activity [15], E664K has been the preferred mutation. Other mutations, targeting the same E664 residue, have been reported to enhance TNA (threosyl nucleic acids; E664I) [16] and tPhoNA (3′–2′ phosphonomethylthreosyl nucleic acid; E664H) synthesis [17]. In all, the repeated isolation of E664 variants with improved XNA synthesis activity suggests that this is an important determinant of substrate specificity among archaeal B-family DNA polymerases.

Because the polymerase thumb domain is not well conserved even among B-family DNA polymerases (Figure 2), if the same mechanism is present in other polymerases (including other B-family enzymes) the residues need not be structurally close. For instance, although the archaeal HNA synthetase required extensive mutation of the polymerase thumb [13], the DNA polymerase from the bacteriophage Phi29 (a mesophilic B-family polymerase) requires no additional mutations on the thumb to processively synthesise HNA [18].

On the other hand, Romesberg and colleagues, building on previous engineering efforts [9], reported recently the engineering of a 2'OMe-RNA synthetase based on the A-family DNA polymerase from *Thermus aquaticus* [19]. The mutations cluster in two regions, one at the interface between finger and thumb (see below), the other in the thumb in close contact to the nascent double helix – a region highly reminiscent of E664.

Challenges from bypassing knowledge gaps with random mutagenesis – The XNA RT

Another traditional strategy for creating a library for selection is through using error-prone PCR (epPCR) to introduce a low number of mutations across a defined (and usually long, i.e. > 100 bp) fragment of the target gene. While epPCR libraries are not designed to sample the whole of the available sequence space, they make no assumptions about the functional roles of the targeted residues. Consequently, epPCR libraries are a powerful tool to probe knowledge gaps potentially identifying previously missed relevant residues.

On the other hand, and common to all directed evolution approaches, the output of selection does not necessarily isolate the minimum set of essential mutations needed to generate the selected trait. Neutral mutations (or passenger mutations), which can have a significant impact in the evolvability of a library [20], tend to accumulate in selection and it is not always practical to deconvolute an isolated variant to trace the contribution of individual mutations.

Deconvolution can be carried out systematically, introducing individual mutations on a wild-type backbone, as carried out by Skirgaila and colleagues in engineering a more thermotolerant Phi29 DNA polymerase [21]. Skirgaila and colleagues created a library of Phi29 DNA polymerase variants by epPCR and used a variation on compartmentalised self-replication [22] to select for enzymes with increased stability – the wild-type Phi29 polymerase has an approximate half-life of 18 minutes at 30°C.

Sixteen variants isolated directly from selection were shown to harbour between seven and 13 mutations (average 9.4), but eight mutations dominated that dataset (present in at least 50% of the isolated variants). The eight-residue mutant was tested for activity confirming the selected phenotype – the engineered variant was shown to have a half-life of over eight hours at 30°C. Crucially, these 8 positions were then tested individually to measure their individual contribution to protein stability, with six of them contributing significantly to the stability of the enzyme in the absence or presence of a DNA substrate [21].

Ellington and colleagues in engineering a proof-reading RT based on the KOD (*Thermococcus kodakarensis*) DNA polymerase, faced a more daunting task, since the selected polymerase harboured 37 mutations [15]. Deep sequencing of the libraries combined with structural and modelling work was used to identify residues that could contribute to the selected function.

The individual contribution of each mutation on the engineered polymerase function was not determined since, as the authors suggest, the RT activity is probably an emergent property of reshaping the active site and duplex binding tunnel in the polymerase. Nevertheless, Ellington and colleagues showed that a variant, harbouring twelve mutations, did not have sufficient RT activity to perform an RT-PCR on a 0.5 kb fragment. Adding two mutations, M137L and K466R (in KOD), to that background enabled the polymerase to carry out RT-PCR of a 1.5 kb fragment. Two further mutations, I521L and N735K, further enhanced the efficiency of the process.

Of those, the I521L mutation is the best characterised to date. It was originally isolated from a screen for HNA reverse transcriptases from *T. gorgonarius* mutants [13] alongside two other variants, I521P and I521H [23]. It has been shown to be a key enabling mutation for the reverse transcription of a wide range of XNAs [13,24]. Structurally, I521 is in close proximity to a loop crucial to catalysis (C-motif: 538 YSDTDGF 544 in KOD) but how it affects catalysis is less clear. Notably, mutations at the structurally equivalent position in the Phi29 DNA polymerase (T441) did not improve the enzyme's XNA RT activity (Pinheiro, V.B.; unpublished), and could suggest that the effect of I521L is restricted to archaeal thermostable enzymes, where that residue is strictly conserved.

Testing I521L for incorporation of nucleotide analogues harbouring protecting groups in their 3' hydroxyl (as used in some platforms of next-generation sequencing), suggests that I521L increases template switching and reduces phosphorolysis [25]. In the context of XNA synthesis,

I521L significantly improved HNA synthesis on a 6G12 background [26], removing the need for forcing reaction conditions (combining Mn^{2+} and Mg^{2+}). That apparent enzyme increase in activity was also observed for the synthesis of tPhoNA and DNA (of an intermediate in the evolution of the tPhoNA synthetase) [17].

It is plausible that residue I521 affects the catalytic process through direct interaction with the C-motif loop (akin to previously reported A-motif variants of Pfu DNA polymerase [27]), however detailed incorporation kinetics assays of the I521L mutant have not yet been reported. It is also plausible that I521L could be affecting the polymerase dynamics, hence assays that can measure incorporation rates [28,29] and fidelity [30], will also be of extreme value to understand the role of I521 in polymerase function and how the mutant affects that.

Engineering polymerase dynamics – The 'Therminator' mutation

Our understanding of DNA polymerase mechanism has been shaped by bulk biophysical assays and it is supported by multiple detailed structures that capture snapshots of various steps along the catalytic cycle. The result is a well-established functional model that takes into account substrate binding and release, protein conformational changes and chemical steps (Figure 3).

A major limitation of such continuous mass action model is that it is an approximation of a quantised stochastic model. As a result, it can be difficult to relate bulk measurements (or properties) to the effect and mechanism of individual polymerases. Attempts have been made in developing agent-based models of DNA polymerase catalysis [31] but to date they have gained little traction in the field.

The mass action model covers one of the key dynamic processes in DNA polymerases: the switch between 'open' and 'closed' conformations. Once thought to be the rate limiting step in catalysis, the movement of the finger domain is more compatible with an 'induced fit' [32] or 'conformational diversity' [33] models, where the closed conformation is stabilized upon nucleotide triphosphate binding.

Crucially, incorrect nucleotides do not stabilize the closed conformation as efficiently. The result is not only a lower probability of the chemical reaction taking place but a higher likelihood of the polymerase to switch back to an open conformation, allowing nucleotide triphosphate exchange [32]. It remains unclear whether an incorrect nucleotide in the active site can access the catalytically relevant closed conformation efficiently, but it is accepted that residence time of a triphosphate substrate in the active site is the key determinant of incorporation.

This dynamic process represents a key bottleneck in the engineering of XNA polymerases, where the unnatural triphosphate substrates (or XNA templates, or both) are likely to have suboptimal fit in the active site, as recently reported for a KOD-based TNA synthase crystallised as a ternary complex with a nascent DNA duplex and TNA nucleotide triphosphate in its active site [34].

Hence, for engineering, if the conformational sampling is relevant to polymerase mechanism, it follows that there must be mutations capable of modulating that process. Mutations that stabilize the closed conformation or that slow the finger dynamics, would likely result in polymerases that can more readily synthesise XNA (or reverse transcribe it or both), but are less active (due to the longer catalytic cycle and lower probability of nucleotide or pyrophosphate release) and more prone to misincorporation (due to the longer residence of triphosphates in the active site of the closed conformation).

Therminator (9°N-7 DNA polymerase harbouring A485L mutation) has become one of the most important enzymes in the XNA field for its increased ability to incorporate base-, sugar- and phosphate-modified triphosphates. The A485L mutation was originally identified as a determinant of nucleotide sugar recognition in a related archaeal polymerase [35,36] but its

mechanism of action has never been determined. The mutation does not affect the structure of the finger domain significantly and it is not expected to create any steric clashes, however, its effect is clear: The mutant polymerase is less active than the wild-type enzyme (in U/mg), it has lower fidelity and it is more prone to template-independent synthesis.

This effect is not exclusive to B-family archaeal DNA polymerases. Stabilizing thumb interactions with the polymerase fingers has been shown to contribute to an engineered A-family 2'OMe-RNA synthetase activity [19]. Pacific Biosciences has, across multiple patents, also described Phi29 DNA polymerase mutants that are affected in the sampling rate of incoming nucleotides and on the stability of the closed conformation [37–40].

Nucleotide sampling in the active site is not the only dynamic process in a polymerase: the binding of the polymerase to the DNA duplex, the balance between exonuclease and polymerase active sites as well as the translocation process are all dynamic processes that can be modulated through polymerase engineering.

Of those, the balance between exonuclease and polymerase activity in B-family enzymes has been the best characterised to date. It is well-established that B-family enzymes operate in a balance between polymerase and exonuclease activity: that process is central to the enzyme's proof-reading and enhanced fidelity compared to A-family polymerases like Taq, which lack the 3'->5' exonuclease function. Mismatches or lack of nucleotide triphosphate substrates curb the polymerisation process allowing clearer observation of the exonucleolytic activity, which has long been exploited for DNA cloning [41].

Salas and colleagues were among the first to report that mutations targeting residues other than conserved exonuclease catalytic site motifs, which could also lead to loss of exonuclease function [42]. Mutations like N62D in Phi29 DNA polymerase (structurally and functionally equivalent to N210 in KOD [15]), have little impact on polymerase activity (and strand-displacing activity) but greatly reduce exonuclease function without disabling it. While residual exonuclease activity in Phi29 N62D is at the limit of detection for DNA synthesis, it remained a significant factor in engineering a mesophilic HNA polymerase [18] – presumably a consequence of lower XNA synthesis rates and distorted nascent duplex that are both expected to drive the probability towards exonuclease sampling. Other residues have also been implicated in modulating exonuclease sampling, such as H147 in KOD with some mutations increasing and others decreasing sampling (and with measurable impact on downstream fidelity) [43].

Crucially, exonuclease activity remains a separate process from exonuclease sampling. For the engineering of XNA polymerases, it is beneficial to target both processes, such that the nascent duplex cannot be degraded (exonuclease activity) and remains in a productive conformation (polymerase site) without futile cycles of switching between the two catalytic sites. In the engineering of a tPhoNA synthetase, modulating the exonuclease sampling through the introduction of an H147E mutation (which would be expected to reduce sampling to about 30% of the wild-type) had a small but measurable impact on the activity of an engineering intermediate [17].

Future of polymerase engineering

Next-generation sequencing platforms that rely (or are compatible with) a continuous polymerase reaction, such as Pacific Biosciences [44] or Oxford Nanopore [45,46], are powerful tools to characterise polymerases with the single-molecule real-time monitoring sure to enhance our mechanistic insight on these highly complex enzymes.

Polymerase structures also have significant value in the engineering process, particularly from related structures harbouring a range of substrates, sampling mismatches [47,48], unnatural

bases [49] and triphosphate analogues [50]. Those “families” of structures give a very detailed knowledge of spatial constraints that can affect polymerase function.

Deep understanding of polymerase mechanism will have a direct impact on their engineering, shifting selection platforms away from large libraries towards small knowledge-based ones, capable of delivering iterative improvements of function.

We have recently demonstrated the power of that approach in the engineering of a tPhoNA synthase, where initial *in vitro* selection efforts based on compartmentalised self-tagging [51] were not successful due to interference from components of the cell lysate. The systematic engineering bypassed that limitation helping deliver not only the most orthogonal XNA polymerase to date but also the first synthetic genetic material that differs from nature in two of its three chemical moieties [17].

Polymerases are at the core of multiple biotechnological and medical applications, from PCR to aptamer selections, from DNA sequencing to DNA *de novo* synthesis. The functional requirements that drove the natural evolution of these enzymes have created incredible molecular machines that can be further refined. Reverse engineering these machines to truly understand their function, will greatly accelerate their further development, allowing the precise fine tuning of function for each desired application.

References

1. Lehman IR, Bessman MJ, Simms ES, Kornberg A: **Enzymatic Synthesis of Deoxyribonucleic Acid: I. PREPARATION OF SUBSTRATES AND PARTIAL PURIFICATION OF AN ENZYME FROM ESCHERICHIA COLI** . *J Biol Chem* 1958, **233**:163–170.

2. Darmawan H, Harrison M, Reha-Krantz LJ: **DNA polymerase 3′ ⟶ 5′ exonuclease activity: Different roles of the beta hairpin structure in family-B DNA polymerases**. *DNA Repair (Amst)* 2015, **29**:36–46.

3. Beard WA, Wilson SH: **Structure and Mechanism of DNA Polymerase β**. *Biochemistry* 2014, **53**:2768–2780.

4. Houlihan G, Arangundy-Franklin S, Holliger P: **Exploring the Chemistry of Genetic Information Storage and Propagation through Polymerase Engineering**. *Acc Chem Res* 2017, **50**:1079–1087.

5. Laos R, Thomson JM, Benner SA: **DNA polymerases engineered by directed evolution to incorporate non-standard nucleotides** . *Front Microbiol* 2014, **5**:565.

6. Ishino S, Ishino Y: **DNA Polymerases and DNA Ligases**. In *Thermophilic Microbes in Environmental and Industrial Biotechnology*. Edited by Satyanarayana T, Littlechild J, Kawarabayasi Y. Springer Netherlands; 2013:429–457.

7. Xia S, Konigsberg WH: **RB69 DNA Polymerase Structure, Kinetics, and Fidelity**. *Biochemistry* 2014, **53**:2752–2767.

8. Tizei PAG, Csibra E, Torres L, Pinheiro VB: **Selection platforms for directed evolution in synthetic biology**. *Biochem Soc Trans* 2016, **44**:1165–1175.

9. Xia G, Chen L, Sera T, Fa M, Schultz PG, Romesberg FE: **Directed evolution of novel polymerase activities: mutation of a DNA polymerase into an efficient RNA polymerase.** *Proc Natl Acad Sci U S A* 2002, **99**:6597–602.

10. Ong JL, Loakes D, Jaroslawski S, Too K, Holliger P: **Directed evolution of DNA polymerase, RNA polymerase and reverse transcriptase activity in a single polypeptide**. *J Mol Biol*

2006, *J Mol Biol* 2006, **361**:537–550.

11. Patel PH, Loeb LA: **Multiple Amino Acid Substitutions Allow DNA Polymerases to Synthesize RNA**. *J Biol Chem* 2000, **275**:40266–40272.

12. Cozens C, Pinheiro VB, Vaisman A, Woodgate R, Holliger P: **A short adaptive path from DNA to RNA polymerases**. *Proc Natl Acad Sci U S A* 2012, **109**:8067–8072.

13. Pinheiro VB, Taylor AI, Cozens C, Abramov M, Renders M, Zhang S, Chaput JC, Wengel J, Peak-Chew S-Y, McLaughlin SH, et al.: **Synthetic Genetic Polymers Capable of Heredity and Evolution**. *Science (80-)* 2012, **336**:341–344.

14. Cozens C, Mutschler H, Nelson GM, Houlihan G, Taylor AI, Holliger P: **Enzymatic Synthesis of Nucleic Acids with Defined Regioisomeric 2'-5' Linkages**. *Angew Chemie - Int Ed* 2015, **54**:15570–15573.

15. Ellefson JW, Gollihar J, Shroff R, Shivram H, Iyer VR, Ellington AD: **Synthetic evolutionary origin of a proofreading reverse transcriptase.** *Science* 2016, **352**:1590–3.

16. Dunn MR, Otto C, Fenton KE, Chaput JC: **Improving Polymerase Activity with Unnatural Substrates by Sampling Mutations in Homologous Protein Architectures**. *ACS Chem Biol* 2016, **11**:1210–1219.

17 ••. Liu C, Cozens C, Jaziri F, Rozenski J, Maréchal A, Dumbre S, Pezo V, Marlière P, Pinheiro VB, Groaz E, et al.: **Phosphonomethyl Oligonucleotides as Backbone-Modified Artificial Genetic Polymers**. *J Am Chem Soc* 2018, **140**:6690–6699.

•• This is the first report of a synthetic genetic material that modifies the phosphate backbone, as well as the sugar. The polymerase reported was rationally engineered with specific mutations or positions being targeted for functional improvement. The result is the most chemically diverse XNA and the most orthogonal XNA synthetase reported to date.

18 •. Torres LL, Pinheiro VB: **Xenobiotic Nucleic Acid (XNA) Synthesis by Phi29 DNA Polymerase**. *Curr Protoc Chem Biol* 2018, **10**:e41.

• This is the first report of a mesophilic XNA synthetase.

19 •. Chen T, Hongdilokkul N, Liu Z, Adhikary R, Tsuen SS, Romesberg FE: **Evolution of thermophilic DNA polymerases for the recognition and amplification of C2-modified DNA**. *Nat Chem* 2016, **8**:556–562.

• This is the first report of an A-family polymerase capable of meaningful XNA synthesis (> 30 bases).

20. Bershtein S, Goldin K, Tawfik DS: **Intense Neutral Drifts Yield Robust and Evolvable Consensus Proteins**. *J Mol Biol* 2008, **379**:1029–1044.

21. Povilaitis T, Alzbutas G, Sukackaite R, Siurkus J, Skirgaila R: ***In vitro* evolution of phi29 DNA polymerase using isothermal compartmentalized self replication technique**. *Protein Eng Des Sel* 2016, **29**:1–12.

22. Ghadessy FJ, Ong JL, Holliger P: **Directed evolution of polymerase function by compartmentalized self-replication**. *Proc Natl Acad Sci U S A* 2001, **98**:4552–4557.

23. Holliger P, Cozens C, Pinheiro VB: **Enzymes**. 2013, US2013/0130320 A1, 2013. p. 166.

24. Cozens C, Pinheiro VB: **XNA Synthesis and Reverse Transcription by Engineered Thermophilic Polymerases**. *Curr Protoc Chem Biol* 2018, **10**:e47.

25. Chen C-Y, He M, Bomati E, Peisajovich S: **Modified Polymerases for Improved**

Incorporation of Nucleotide Analogues. 2014, WO/2014/142921, 2014. p. 137.

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

27. Kennedy EM, Hergott C, Dewhurst S, Kim B: **The mechanistic architecture of thermostable *Pyrococcus furiosus* family B DNA polymerase motif A and its interaction with the dNTP substrate**. *Biochemistry* 2009, doi:10.1021/bi9010122.

28 •. Nikoomanzar A, Dunn MR, Chaput JC: **Evaluating the Rate and Substrate Specificity of Laboratory Evolved XNA Polymerases**. *Anal Chem* 2017, doi:10.1021/acs.analchem.7b03807.

• The assay reported here extends fluorimetric assays developed for DNA polymerases to monitor DNA synthesis towards XNA synthetases through the systematic screening of fluorescent dyes.

29. Seville M, West AB, Cull MG, McHenry CS: **Fluorometric assay for DNA polymerases and reverse transcriptase**. *Biotechniques* 1996, **21**:664–672.

30. Potapov V, Ong JL: **Examining Sources of Error in PCR by Single-Molecule Sequencing**. *PLoS One* 2017, **12**:e0169774.

31. Keller DJ, Brozik JA: **Framework model for DNA polymerases**. *Biochemistry* 2005, doi:10.1021/bi0477079.

32. Johnson KA: **Role of induced fit in enzyme specificity: A molecular forward/reverse switch**. *J Biol Chem* 2008, **283**:26297–26301.

33. Hohlbein J, Kapanidis AN: **Probing the Conformational Landscape of DNA Polymerases Using Diffusion-Based Single-Molecule FRET**. In *Methods in Enzymology*. 2016:353–378.

34. Chim N, Shi C, Sau SP, Nikoomanzar A, Chaput JC: **Structural basis for TNA synthesis by an engineered TNA polymerase**. *Nat Commun* 2017, **8**:1810.

35. Gardner AF, Jack WE: **Determinants of nucleotide sugar recognition in an archaeon DNA polymerase**. *Nucleic Acids Res* 1999, **27**:2545–2553.

36. Gardner AF, Jack WE: **Acyclic and dideoxy terminator preferences denote divergent sugar recognition by archaeon and Taq DNA polymerases.** *Nucleic Acids Res* 2002, **30**:605–13.

37. Rank DR, Bibillo A, Hanzel DK, Peluso P, Wegener J: **Polymerase enzymes and reagents for enhanced nucleic acid sequencing.** 2008, WO 2008/051530 A2, 2008. p. 68.

38. Clark S, Bibillo A, Peluso P, Christians F, He M, Park I, Lee H, Bjornson K, Jia L: **Generation of modified polymerases for improved accuracy in single molecule sequencing**. 2009, WO 2009/145820 A2, 2009. p. 152.

39. Patel P, Korlach J, Bibillo A, Bjornson K, Hanes J: **Engineering polymerases and reaction conditions for modified incorporation properties**. 2010, WO 2010/027484 A2, 2010. p. 117.

40. Pinheiro VB, Ong JL, Holliger P: **Polymerase engineering: from PCR and sequencing to synthetic biology**. In *Protein Engineering Handbook*. . 2012:279–302.

41. Aslanidis C, de Jong PJ: **Ligation-independent cloning of PCR products (LIC-PCR)**. *Nucleic Acids Res* 1990, **18**:6069–6074.

42. de Vega M, Lazaro JM, Salas M, Blanco L: **Primer-terminus stabilization at the 3'-5' exonuclease active site of phi29 DNA polymerase. Involvement of two amino acid residues highly conserved in proofreading DNA polymerases.** *EMBO J* 1996, **15**:1182–92.

43. Kuroita T, Matsumura H, Yokota N, Kitabayashi M, Hashimoto H, Inoue T, Imanaka T, Kai Y: **Structural mechanism for coordination of proofreading and polymerase activities in archaeal DNA polymerases**. *J Mol Biol* 2005, **351**:291–298.

44. Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, Peluso P, Rank D, Baybayan P, Bettman B, et al.: **Real-time DNA sequencing from single polymerase molecules**. *Science* 2009 **323**:133–8.

45. Manrao EA, Derrington IM, Laszlo AH, Langford KW, Hopper MK, Gillgren N, Pavlenok M, Niederweis M, Gundlach JH: **Reading DNA at single-nucleotide resolution with a mutant MspA nanopore and phi29 DNA polymerase**. *Nat Biotechnol* 2012, **30**:349–353.

46. Lieberman KR, Cherf GM, Doody MJ, Olasagasti F, Kolodji Y, Akeson M: **Processive replication of single DNA molecules in a nanopore catalyzed by phi29 DNA polymerase**. *J Am Chem Soc* 2010, **132**:17961–17972.

47. Xia S, Konigsberg WH: **Mispairs with Watson-Crick base-pair geometry observed in ternary complexes of an RB69 DNA polymerase variant**. *Protein Sci* 2014, **23**:508–513.

48. Clausen AR, Murray MS, Passer AR, Pedersen LC, Kunkel TA: **Structure–function analysis of ribonucleotide bypass by B family DNA replicases**. *Proc Natl Acad Sci* 2013, **110**:16802 LP-16807.

49 ••. Kropp HM, Dürr SL, Peter C, Diederichs K, Marx A: **Snapshots of a modified nucleotide moving through the confines of a DNA polymerase**. *Proc Natl Acad Sci* 2018, **115**:9992–9997.

•• This is among the first systematic structural characterisations of XNA incorporation. It maps how a nucleobase-modified XNA triphosphate is incorporated and extended through the nascent duplex. It represents a new approach to structure characterisation of polymerases which has the potential to be of great value to engineering.

50 •. Batra VK, Oertell K, Beard WA, Kashemirov BA, McKenna CE, Goodman MF, Wilson SH: **Mapping Functional Substrate–Enzyme Interactions in the pol β Active Site through Chemical Biology: Structural Responses to Acidity Modification of Incoming dNTPs**. *Biochemistry* 2018, **57**:3934–3944.

• A detailed structural characterisation of the active site of Pol beta using a family of nucleotide triphosphate analogues to probe the effect on the polymerase structure and catalysis from changing acidity in the pyrophosphate group.

51. Pinheiro VB, Arangundy-Franklin S, Holliger P: **Compartmentalized Self-Tagging for in vitro-directed evolution of XNA polymerases**. *Curr Protoc Nucleic Acid Chem* 2014, **57**:9.9.1-9.9.18.

52. Tsai YC, Johnson KA: **A new paradigm for DNA polymerase specificity**. *Biochemistry* 2006, **45**:9675–9687.

Figure legends:

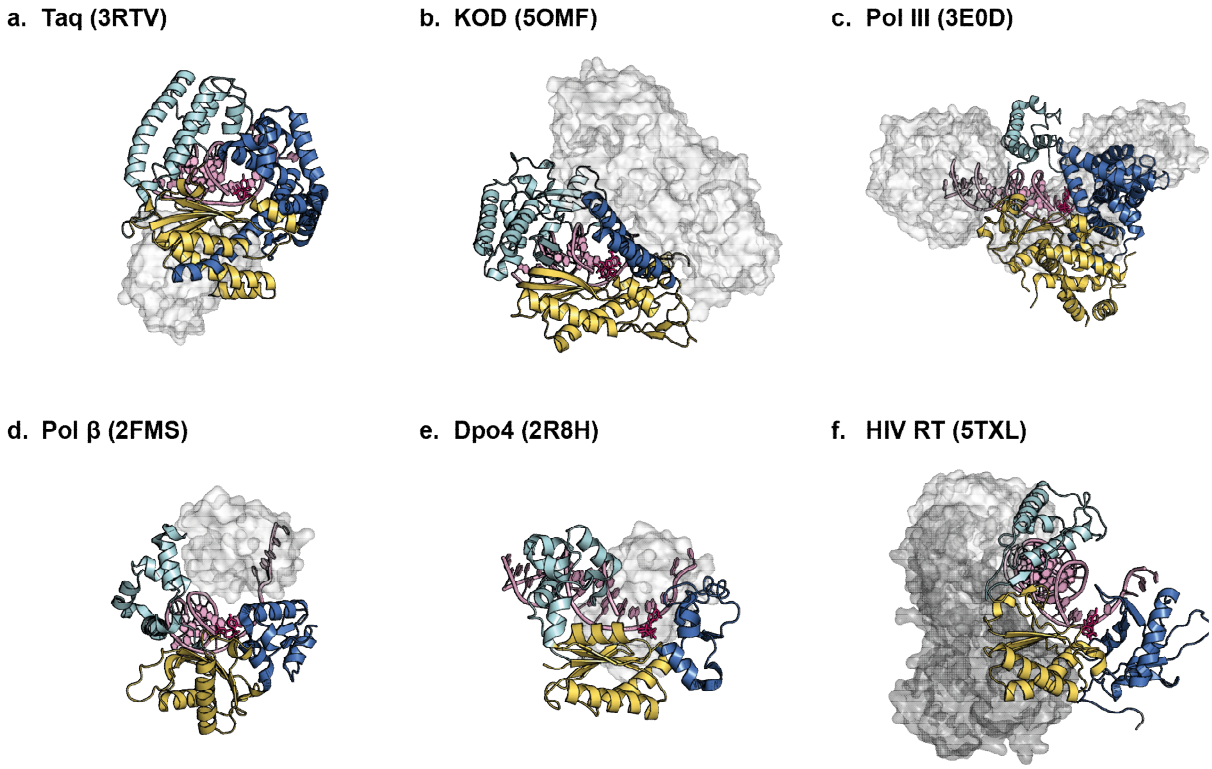


Figure 1: Closed ternary structure of DNA polymerases from different families. The polymerase domain (cartoon) is divided in three parts: palm (yellow), fingers (dark blue) and thumb (cyan). The palm domain is the most conserved structural element among DNA polymerases. Fingers are involved in the binding of the incoming triphosphate (magenta). The thumb makes multiple contacts to the nascent duplex (pink) and has a role in some of the dynamic processes of the enzymes. Domains other than the polymerase domain (e.g. exonuclease and PHP) are shown as surfaces for clarity. PDB entry of the structures is given next to the respective structure.

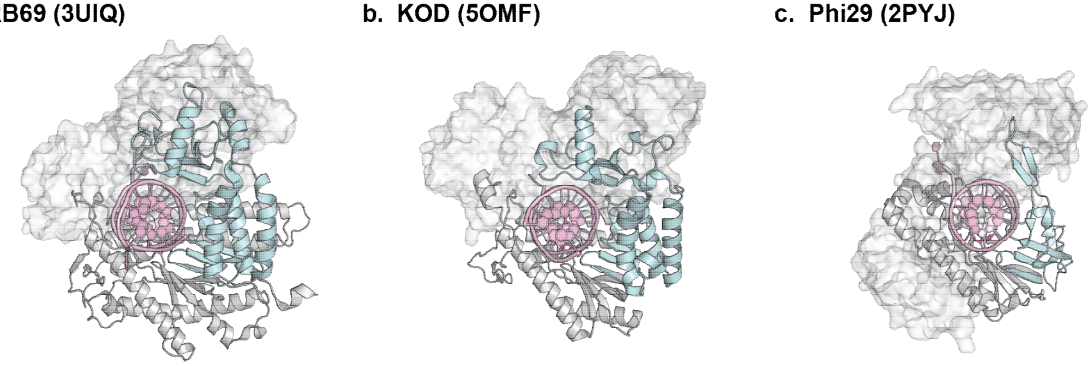


Figure 2: Diversity of thumb structures among B-family polymerases. Polymerase domains are shown as cartoons with thumb subdomains highlighted in cyan. Although it is clear that the thumb subdomains make extensive contact with each nascent chain (pink), the wide variation in structure makes generalization of a functional residue difficult – and that can have an impact on transferring mutations from an engineered enzyme to other polymerases of the same class,

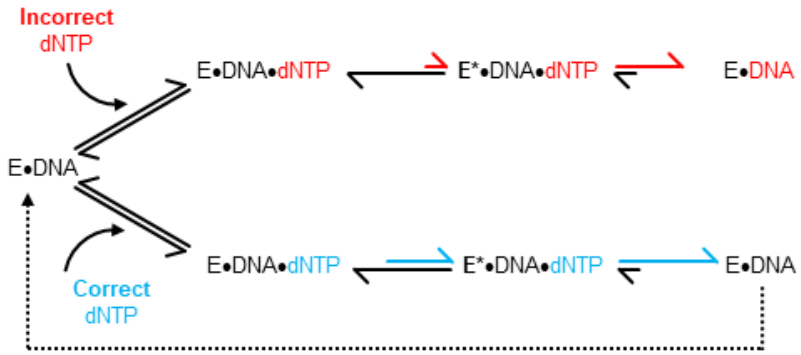


Figure 3: Simplified DNA polymerase reaction pathway. The residence time of a nucleotide in the closed polymerase complex (**E***•DNA•dNTP) is a key determinant for its incorporation. Mutations isolated from engineering that have been characterised (particularly for archaeal hyperthermophilic B-family polymerases) support the induced fit model proposed by Tsai and Johnson [52], where the polymerase binary complex (E•DNA) bind correct and incorrect nucleotides (blue and red dNTP respectively) with minimal discrimination. The correct nucleotide is more likely to stabilize the active closed conformation, which leads to incorporation. Distortions caused by the sampling of the incorrect nucleotide are less likely to stabilize the closed conformation and may also interfere with catalysis – both processes enhance polymerase fidelity. The polymerase thus behaves as a ratchet with the correct nucleotide maximising the probability that the (near) irreversible chemical step takes place. The more the system is skewed away from incorporation, the less likely it is for misincorporation to occur, with the enzyme sampling nucleotides for longer before each incorporation. If a misincorporation does take place it may lead to polymerase stalling and correction by an active exonuclease domain (in proofreading polymerases).

Acknowledgements

VBP thanks Dr. Leticia Torres and Dr. Chris Cozens for critical reading of the manuscript. VBP thanks the Biotechnology and Biosciences Research Council (grant numbers BB/N01023X/1, BB/N010221/1) and the European Research Council [ERC-2013-StG project 336936 (HNAepisome)] for their support.