Title
Structural basis for TNA synthesis by an engineered TNA polymerase.

Permalink
https://escholarship.org/uc/item/40h342x8

Journal
Nature communications, 8(1)

ISSN
2041-1723

Authors
Chim, Nicholas
Shi, Changhua
Sau, Sujay P
et al.

Publication Date
2017-11-27

DOI
10.1038/s41467-017-02014-0

License
https://creativecommons.org/licenses/by/4.0/ 4.0

Peer reviewed
Structural basis for TNA synthesis by an engineered TNA polymerase

Nicholas Chim1, Changhua Shi1, Sujay P. Sau1, Ali Nikoomanzar1 & John C. Chaput1

Darwinian evolution experiments carried out on xeno-nucleic acid (XNA) polymers require engineered polymerases that can faithfully and efficiently copy genetic information back and forth between DNA and XNA. However, current XNA polymerases function with inferior activity relative to their natural counterparts. Here, we report five X-ray crystal structures that illustrate the pathway by which α-(L)-threofuranosyl nucleic acid (TNA) triphosphates are selected and extended in a template-dependent manner using a laboratory-evolved polymerase known as Kod-RI. Structural comparison of the apo, binary, open and closed ternary, and translocated product detail an ensemble of interactions and conformational changes required to promote TNA synthesis. Close inspection of the active site in the closed ternary structure reveals a sub-optimal binding geometry that explains the slow rate of catalysis. This key piece of information, which is missing for all naturally occurring archaean DNA polymerases, provides a framework for engineering new TNA polymerase variants.

1 Departments of Pharmaceutical Sciences, Chemistry, and Molecular Biology and Biochemistry University of California, Irvine, CA 92697-3958, USA. Nicholas Chim and Changhua Shi contributed equally to this work. Correspondence and requests for materials should be addressed to J.C.C. (email: jchaput@uci.edu)
Synthetic genetics is an emerging field of science that aims to extend the principles of heredity and evolution to nucleic acid polymers with backbone structures that are distinct from those found in nature. Collectively referred to as xenonucleic acids or XNA, these polymers have unique physicochemical properties that often include strong resistance to degradative enzymes and duplex structures that adopt a range of helical geometries. By engineering polymerases to synthesize and recover genetic information encoded in XNA, researchers are developing complex molecular systems that are capable of undergoing Darwinian evolution in response to imposed selection constraints. These studies, which expand our ability to store, propagate, and evolve genetic information, have profound implications for biotechnology, molecular medicine, and the origin of life.

To date, five different XNA polymers with non-ribose sugars (1,5-anhydrohexitol nucleic acid (HNA), arabino nucleic acid (ANA), 2'-fluoro-arabino nucleic acid (FANA), cyclohexenyl nucleic acid (CeNA) and α-1-threose nucleic acid (TNA)) have achieved successful replication in a Darwinian evolution system. Remarkably, despite this difference, TNA is capable of forming duplexes with itself and complementary strands of DNA and RNA. Solution NMR studies reveal that duplex formation in either the self-helical geometries or TNA/RNA occurs through an A-like helical geometry that is templated by a rigid TNA backbone. More recently, stability constraints have been identified for an archaeal polymerase that derives from a replicative B-family evolved polymerase that can synthesize unnatural TNA polymers on natural DNA templates. The collection of five X-ray crystal structures details an ensemble of intermolecular interactions and conformational changes that allow Kod-RI to promote TNA synthesis. Close inspection of the enzyme-active site in the closed ternary structures would help evaluate the mechanism of natural DNA synthesis relative to other polymerase families. However, the crystal structures of XNA polymerases themselves are even more valuable, as these structures would provide insights into the limitations of existing XNA polymerases. In their absence, structural information about the ternary complex must be derived from distantly related viral (RB69 Pol and Phi29 Pol) and eukaryotic polymerases (Pols α, δ, and ε), which share only ~20% identity with archaeal B-family polymerase.

Here, we describe a structural approach that was taken to evaluate the pathway by which a laboratory-evolved polymerase is able to synthesize unnatural TNA polymers on natural DNA templates. The collection of five X-ray crystal structures details an ensemble of intermolecular interactions and conformational changes that allow Kod-RI to promote TNA synthesis. Close inspection of the enzyme-active site in the closed ternary structures would help evaluate the mechanism of natural DNA synthesis relative to other polymerase families. However, the crystal structures of XNA polymerases themselves are even more valuable, as these structures would provide insights into the limitations of existing XNA polymerases. In their absence, structural information about the ternary complex must be derived from distantly related viral (RB69 Pol and Phi29 Pol) and eukaryotic polymerases (Pols α, δ, and ε), which share only ~20% identity with archaeal B-family polymerase.

![Fig. 1 TNA synthesis by Kod-RI.](image)

- **Apo**: 3'-Template
- **Primer**: 5'-Primer
- **Binary complex**: 3'-Primer
- **Ternary complex (open conformation)**: 3'-Primer
- **Ternary complex (closed conformation)**: 3'-Primer
- **Translocated product**: 3'-Primer

**Fig. 1 TNA synthesis by Kod-RI.** a Molecular structures comparing TNA to DNA and RNA. b Denaturing PAGE showing TNA synthesis on a library of degenerate DNA templates. c Functional analysis of TNA substrate and primer template binding. Error bars represent the average of three independent trials. d Schematic view of the TNA synthesis pathway where each cartoon image represents an elucidated structure described in this study.
substrates (Fig. 1c). These values, which are within the range of natural archaeal B-family DNA polymerases, led us to speculate that the slow rate of TNA synthesis is due to an imperfect active site that positions tNTP substrates in a geometry that is sub-optimal for phosphodiester bond formation. A key piece of information explaining the slow rate of Kod-RI mediated TNA synthesis relative to DNA synthesis by natural Kod DNA polymerase is the presence of tATP with excess tATP and/or magnesium ions just bound triphosphate or magnesium ions. Similar results have been observed, such as no electron density was observed for either the bound triphosphate or magnesium ions. Critical to this effort was the chemical synthesis of a chain-terminating primer bearing a 2′-deoxy-α-L-threofuranosyl adenosine residue (tAδ) at the 3′ end. Controlled pore glass (CPG) functionalized with a tAd-analog was obtained in 8 synthetic steps (Fig. 2) from a known dimethoxytrityl-protected α-L-threofuranosyl adenosine nucleoside. Solid-phase synthesis was then used to construct the DNA primer (P2, Supplementary Table 1) bearing an unmodified TNA adenosine residue (tA) was similarly constructed using tA-modified CPG resin. A second primer (P1, Supplementary Table 1) bearing an unmodified TNA adenosine residue (tA) was similarly constructed using tA-modified CPG resin. The α-L-threofuranosyl adenosine 3′-triphosphate (tATP) required for crystallization was obtained in 12 steps from L-ascorbic acid using known methodology. Although Kod-RI crystallized in the presence of tATP and magnesium, no electron density was observed for either the bound triphosphate or magnesium ions. Similar results have been reported by others for crystallization trials conducted on the wild-type Kod polymerase and related homologs. In an effort to overcome this problem, we explored a variety of crystallization conditions, which included variations in tATP concentration, soaking times, and metal ion additives. After extensive optimization, we successfully obtained open and closed ternary structures with clear electron density for the TNA substrate and bound magnesium ions in the enzyme-active site. The condition that proved most successful involved soaking co-crystals grown in the presence of tATP with excess tATP and/or magnesium ions just

**Fig. 2** Chemical synthesis scheme of 2′-deoxy-α-L-threofuranosyl adenosine-modified CPG (Compound 7). Reagents: a PhNCS, NaH, THF, rt, 2.5 h, 80%; b AIBN, Bu3SnH, toluene, 100 °C, 45 m, 86%; c K2CO3, MeOH, 72%; d R1, PyBOP, DIPEA, DCE, 43%; e TBAF/THF, 70%; f 2-cyanoethyl-N,N-disopropylchlorophosphoramidite, DIPEA, DCM, g R2, ETS-tetrazole, MeCN; h 1Py-water

**Results**

**Function.** Kod-RI is a DNA-dependent TNA polymerase that is able to transcribe individual strands or large libraries of degenerate DNA sequences into TNA (Fig. 1b). This property, which is remarkable considering the backbone structure of TNA relative to DNA and RNA, has enabled the evolution of TNA aptamers from unbiased pools of random sequences. The enzyme functions by a primer-extension mechanism in which a primer strand (DNA or TNA) annealed to a DNA template is extended with tNTP substrates in a geometry that is sub-optimal for phosphodiester bond formation. Crystallization. To provide structural insights into the mechanism of TNA synthesis, protein crystals of Kod-RI were grown under conditions that were designed to capture four main steps in the TNA synthesis pathway (Fig. 1d), namely P/T binding, nucleoside triphosphate binding, catalysis, and translocation. The condition that proved most successful involved soaking co-crystals grown in the presence of tATP with excess tATP and/or magnesium ions just

**Fig. 2** Chemical synthesis scheme of 2′-deoxy-α-L-threofuranosyl adenosine-modified CPG (Compound 7). Reagents: a PhNCS, NaH, THF, rt, 2.5 h, 80%; b AIBN, Bu3SnH, toluene, 100 °C, 45 m, 86%; c K2CO3, MeOH, 72%; d R1, PyBOP, DIPEA, DCE, 43%; e TBAF/THF, 70%; f 2-cyanoethyl-N,N-disopropylchlorophosphoramidite, DIPEA, DCM, g R2, ETS-tetrazole, MeCN; h 1Py-water

**Results**

**Function.** Kod-RI is a DNA-dependent TNA polymerase that is able to transcribe individual strands or large libraries of degenerate DNA sequences into TNA (Fig. 1b). This property, which is remarkable considering the backbone structure of TNA relative to DNA and RNA, has enabled the evolution of TNA aptamers from unbiased pools of random sequences. The enzyme functions by a primer-extension mechanism in which a primer strand (DNA or TNA) annealed to a DNA template is extended with tNTP substrates in a geometry that is sub-optimal for phosphodiester bond formation.
prior to freezing. Specifically, the open ternary structure required a 20 min soak with 0.2 mM tATP, whereas the closed ternary structure required a 45 min soak with 2 mM tATP and 20 mM MgCl₂. Although the precise reason for why the ternary structures required additional soaking is unclear, we speculate that it could be due to an active site pocket that allows for the rapid exchange of nucleotide substrates.

**Structures.** Five structures spanning a resolution limit of 2.05–3.2 Å were solved by molecular replacement (Table 1 and Supplementary Fig. 2). We used the apo and binary forms of Kod and 9°N31. The second major conformational change involves formation of the TNA synthesis pathway, including the sequence motif KKKK (residues 591–594, Fig. 3b), which is thought to stabilize the B-form helix by bringing the primer and template closer together. The unpaired region of the template is stabilized by residues from the NTD and exo domains, which cause an abrupt kink in the template at the +1 position. These interactions are all consistent with the high resolution binary structures previously solved for Kod and 9°N31.

**Conformational changes.** Because we were able to solve all five of the polymerase structures that define the TNA synthesis pathway, it was possible to study the conformational changes that facilitate TNA synthesis. Comparative structural analyses identified three major conformational changes between the set of five Kod-RI structures. The first conformational change arises when the apo form of the polymerase binds the P/T duplex to form the binary complex (Fig. 3b). Upon P/T binding, the thumb subdomain transitions from an ensemble of poorly defined conformations to a well-ordered binary structure. One striking example of positional rearrangement is helix α20, which shifts ~23 Å to bind the minor groove face of the P/T duplex (Fig. 3b). In addition, several other secondary structural elements (e.g., β26–28 and 3ζ1ζ4) not visible in the apo structure become visible in the binary complex (Fig. 3b). These structural changes signify the importance of the thumb subdomain in P/T binding.

The second major conformational change involves formation of the ternary complex with the P/T duplex and tATP substrate bound in the enzyme-active site. This step involves insertion of the phosphodiester backbone, with only a small number of direct contacts being made to the sugar and nucleobase moieties (Supplementary Fig. 3a–d). All of the sugar and nucleobase contacts occur through the minor groove, which is consistent with the propensity for B-family polymerases to accept modified nucleotides bearing functional groups at the C5 pyrimidine and C7 deazapurine positions35. In addition, many residues responsible for recognizing the P/T duplex are highly conserved among B-family polymerases, including the sequence motif KKKK (residues 591–594, Fig. 3b), which is thought to stabilize the B-form helix by bringing the primer and template closer together.

The unpaired region of the template is stabilized by residues from the NTD and exo domains, which cause an abrupt kink in the template at the +1 position. These interactions are all consistent with the high resolution binary structures previously solved for Kod and 9°N31.

**Architecture.** Consistent with all known structures of B-family polymerases, Kod-RI adopts a disk-shaped architecture that encompasses the N-terminal (NTD), exonuclease (Exo), and catalytic domains (Fig. 3a)34. The catalytic domain is further divided into the palm, finger, and thumb subdomains. The P/T duplex is bound in a groove defined by the palm and thumb subdomains, making contacts to 9 base pairs in the P/T duplex. Interaction maps created for the binary and ternary structures reveal that the duplex is primarily recognized by contacts made to the phosphodiester backbone, with only a small number of direct contacts being made to the sugar and nucleobase moieties (Supplementary Fig. 3a–d). All of the sugar and nucleobase contacts occur through the minor groove, which is consistent with the propensity for B-family polymerases to accept modified nucleotides bearing functional groups at the C5 pyrimidine and C7 deazapurine positions. In addition, many residues responsible for recognizing the P/T duplex are highly conserved among B-family polymerases, including the sequence motif KKKK (residues 591–594, Fig. 3b), which is thought to stabilize the B-form helix by bringing the primer and template closer together.

The unpaired region of the template is stabilized by residues from the NTD and exo domains, which cause an abrupt kink in the template at the +1 position. These interactions are all consistent with the high resolution binary structures previously solved for Kod and 9°N31.

**Conformational changes.** Because we were able to solve all five of the polymerase structures that define the TNA synthesis pathway, it was possible to study the conformational changes that facilitate TNA synthesis. Comparative structural analyses identified three major conformational changes between the set of five Kod-RI structures. The first conformational change arises when the apo form of the polymerase binds the P/T duplex to form the binary complex (Fig. 3b). Upon P/T binding, the thumb subdomain transitions from an ensemble of poorly defined conformations to a well-ordered binary structure. One striking example of positional rearrangement is helix α20, which shifts ~23 Å to bind the minor groove face of the P/T duplex (Fig. 3b). In addition, several other secondary structural elements (e.g., β26–28 and 3ζ1ζ4) not visible in the apo structure become visible in the binary complex (Fig. 3b). These structural changes signify the importance of the thumb subdomain in P/T binding.

The second major conformational change involves formation of the ternary complex with the P/T duplex and tATP substrate bound in the enzyme-active site. This step involves insertion of...
the tNTP substrate into the active site pocket followed by a closing of the finger subdomain onto the thumb subdomain to form the catalytically relevant closed ternary structure. Structural alignment of the binary and open ternary complexes (rmsd 0.6 Å) reveals that the binary and open ternary structures are identical (Supplementary Fig. 4), implying that the binary complex can accommodate a tNTP substrate without the need for structural change. This observation is consistent with previous kinetic data showing that dNTPs diffuse directly into the active site of B-family polymerases37, rather than occupying a pre-insertion site as has been observed for some A-family polymerases38.

The initial tATP binding event is followed by a major conformational change (Fig. 3c) in which the finger subdomain closes upon the nucleoside triphosphate. Alignment of the open and closed ternary complexes (rmsd 2.5 Å) reveals that Kod-RI undergoes a substantial conformational change to achieve the catalytically relevant state. For this transition, helix α15 in the finger subdomain tilts 21° inward to form the closed ternary complex in which tATP is trapped in a pre-catalytic state by the 2′-deoxy-tA residue on the primer. In addition, numerous residues in the finger (e.g., Arg460, Lys464, Lys487, and Asn491) and palm (e.g., Asp404, Asp540, Asp542, Glu578, and Glu580) subdomains undergo repositioning to promote substrate recognition and catalysis (Fig. 4c).

Following catalysis, the polymerase experiences a third conformational change whereby the finger subdomain reopens and the polymerase translocates to the next position on the template (Fig. 3d). This last step completes the cycle of nucleotide addition by enabling the polymerase to move from position 0 to position −1 on the DNA template. The structural changes observed between the closed ternary complex and translocated product are similar to the changes observed between the open and closed ternary complexes (rmsd 2.5 Å), which is expected as the finger subdomain will open and close an equal distance between each cycle of nucleotide catalysis.

Active site analysis. The active site pocket, which encompasses the nascent T4:tATP base pair, is primarily formed by residues Ile488, Asn491, Ser492, Tyr494, Gly495, Gly498, and Tyr499 in the finger subdomain and residues Leu508 and Tyr509 in the palm subdomain (Fig. 4a, b). Three highly conserved carboxylate groups (Asp404, Asp540, and Asp542) mark the polymerase-active site34. Interestingly, the steric gate residue Tyr509 does not interact with the threose sugar of the incoming tATP substrate (~4.0 Å away) even though this position is known to discriminate against ribonucleoside triphosphates in analogous wild-type polymerases (Supplementary Fig. 5)39. Electron density maps indicate that tATP is tightly bound in the closed ternary conformation (Fig. 4b), but only weakly bound in the open conformation (Fig. 4d). This observation is expected based on the location of the finger subdomain in the open and closed ternary structures.

Three magnesium ions are observed in the active site of the closed ternary complex. Two of these (Mg2 and Mg3) adopt positions that are structurally identical to other B-family polymerases whose ternary structures have been solved by X-ray crystallography (Fig. 4b, c)21, 24. A third magnesium ion...
(Mg₃), located between the α-phosphate of tATP and the primer strand (Fig. 4b, c), is responsible for aligning the 2′-hydroxyl group on the TNA primer for nucleophilic attack on the tNTP substrate. This metal ion lies 3.2 Å from the C2’ atom of the primer strand and 2.5 Å from the α-phosphate on the incoming tATP substrate. A simulated annealing omit map for tATP reveals that the adenine base is highly flexible relative to the triphosphate tail and threose sugar (Fig. 4b). Nucleobase flexibility is due to an active site pocket that is not fully optimized for the smaller size of the tNTP substrate (Fig. 4a). Unlike DNA, TNA lacks a 5’-methylene carbon, which would facilitate stronger Watson–Crick base pairing by bringing the nucleoside triphosphate closer to the templating base. Presumably, these interactions could be strengthened by mutations in the thumb subdomain that better constrain the tNTP substrate in the active site pocket.

**Base pair geometry.** Computational analysis of the Kod-RI structures containing a P/T duplex revealed significant deviations in the planarity of the base pair at the active site. In particular, the buckle and propeller parameters for the T5:tA12 base pair of the binary complex are ~22° and ~18°, respectively (Fig. 4d, Supplementary Table 2). These distortions are recapitulated in the nascent T4:tATP base pair of the open and closed ternary complexes and the non-planar geometry propagates to the T5:tA12 and C6:G11 base pairs (Supplementary Table 2). However, following a single turnover event, the base pair geometry returns to a normal planar conformation as evidenced in the translocated structure (Fig. 4d), indicating the sub-optimal base pair geometry observed in the pre-catalytic state is corrected following the chemical bond forming step.

Similar base pair analyses performed on the ternary structures from known viral and eukaryotic B-family polymerases bound to dNTPs reveals that the base pair distortions observed for Kod-RI are distinct and severe relative to natural B-family polymerases (Supplementary Table 3). In all cases, base pair planarity is maintained throughout the duplex, including the incoming nucleoside triphosphate, which stacks directly on the 3′ end of the primer strand. One minor deviation from this trend is the binary structure of wild-type Kod DNA polymerase bound to an all-DNA P/T duplex, which exhibits buckle and propeller distortions at T5:A12 (Supplementary Table 2). Whether this distortion is typical for Kod DNA polymerase is difficult to assess in the absence of more structural information, most notably a ternary structure for the wild-type polymerase.

**TNA synthesis mutations.** Kod-RI differs from natural Kod DNA polymerase by the mutations A485R and E664I, which were identified by directed evolution as amino acid changes that confer TNA synthesis activity on the natural polymerase scaffold. Both mutations are located outside the active site pocket and mutational analysis suggests that the two residues perform independent functions. Arg⁴⁸⁵ is located on helix α15 in the finger subdomain (Fig. 5). In the closed ternary complex, Arg⁴⁸⁵ adopts an upward facing position with respect to the helix that interacts with residues Arg⁵⁰⁶, Glu⁵³⁰, and Leu⁵³³ located on helices α6 and α9 of the Exo domain while this residue in each of the open conformation structures (i.e., binary, open, and translocated), is oriented in the downward position along the helix as shown in Supplementary Fig. 6. Comparison of the binary structures for Kod-RI and natural Kod DNA polymerase, reveals that the bulky Arg⁴⁸⁵ mutation causes helix α15 to bend toward the primer-template duplex (Supplementary Fig. 7). We postulate that this structural perturbation enables the recognition of TNA substrates by altering the shape of the active site pocket.

By contrast, Ile⁶⁶⁴ is located on strand β27 in the thumb subdomain (Fig. 5). This mutation contacts the phosphodiester backbone at nucleotide positions +5 and +6 on the P/T duplex. Holliger and co-workers have identified position 664 as a key site for the processive synthesis of non-cognate nucleic acid polymers. In one example, variants of a DNA polymerase isolated from *Thermococcus gordonii* were found to promote efficient RNA synthesis on DNA templates. While the precise functional role of Ile⁶⁶⁴ remains unknown, we speculate that this mutation reduces stringent recognition of the natural P/T duplex.
Discussion

Archaea constitute one of the three major evolutionary lineages of life. These organisms exist in a broad range of habitats that include harsh environments, such as hot springs and salt lakes, as well as milder areas consisting of soils, oceans, and marshlands. Because of their ability to withstand high temperatures and organic solvents, enzymes isolated from thermophilic archaea have been exploited in many biotechnology applications.

Indeed, numerous examples now exist where archaean B-family DNA polymerases have been shown to accept chemically modified nucleotides bearing alternative functionality at the sugar or nucleobase moieties and these enzymes are often used as the starting point for the directed evolution of XNA polymerases.

Here, we address this shortcoming by providing a series of X-ray crystal structures that describe how a laboratory-evolved polymerase is able to synthesize unnatural TNA polymers on natural DNA templates. This body of work produced X-ray structures of the polymerase in different conformational states, allowing us to understand how the enzyme interacts with its substrate.

Fig. 5 TNA gain-of-function mutations. Closed ternary Kod-RI structure depicting the mutant residues Arg485 and Ile664 as space-filling models (center). Conformational differences observed between the open and closed forms of the ternary structure are depicted for Ile664 (left insets) and Arg485 (right insets). Residues observed in the open and closed conformations are drawn as yellow and green sticks, respectively.

Fig. 6 Structural comparison of the finger subdomain for B-family polymerases. Conserved finger subdomain residues across archaeal (Kod-RI), viral (RB69, PDB: 1IG9 and Phi29, PDB: 2PYL), and eukaryotic (Pol α PDB: 4FYD, Pol ε PDB: 3IAY, and Pol ε PDB: 4M8O) polymerases interacting with the nascent base pair. In Kod-RI, Mg1 is novel, whereas Mg2 and Mg3 are observed in other B-family polymerases.

by replacing a critical electrostatic interaction with a less discriminating hydrophobic side chain.
crystal structures of the apo, binary, open ternary, and closed ternary, and translocated product of a laboratory-evolved polymerase. Although apo and binary structures have been determined for a limited number of hyperthermophilic B-family polymerases, the open and closed ternary structures of distantly related viral and eukaryotic B-family polymerases exhibit a co-planar geometry for the nascent base pair with divalent metal ions that are ideally positioned for the subsequent chemical bond forming step (Fig. 6). However, this result is not surprising when one considers the limited evolutionary history of Kod-R1 relative to natural polymerases, which have been selected from billions of years of natural selection. The fact that engineered polymerases are able to synthesize artificial genetic polymers with backbone structures that are distinct from those found in nature is a remarkable achievement that can be further improved by directed evolution.

Moving forward, structural information available in the closed ternary complex provides an important framework for developing new synthetic polymerases variants that function with increased catalytic activity. Future efforts aimed at improving the geometry of the nascent base pair between studies that focus on primary and secondary shell residues in the enzyme-active site. This could, for example, include the mutagenesis of nearby residues in the thumb subdomain as well as possible residues in the finger and palm regions that may be required for compensatory reasons. As these studies continue, efforts aimed at developing a more comprehensive view of natural and engineered archael polymerases are warranted and would benefit from high-resolution X-ray crystal structures for three important classes of ternary polymerases that have yet to be solved: (i) wild-type polymerase bound to natural dNTP; (ii) wild-type polymerase bound to tNTP; and (iii) Kod-R1 bound to natural dNTP. In addition, these efforts would also benefit from binary and ternary structures in which the primer strand of the P/T duplex is composed entirely of TNA. Structural information of this type would make it possible to better evaluate the slow rate of TNA synthesis relative to natural DNA synthesis by the wild-type polymerase.

In summary, structural analysis of the TNA synthesis pathway provides critical insights into the constraints of a laboratory-evolved polymerase. This approach of directed evolution and structure determination provides important clues that can be used to guide the development of future XNA polymerases. In the future, it will be interesting to see how well molecular evolution, either alone or guided by computational design, can close the gap in catalytic activity between engineered polymerases and their natural counterparts. Such efforts open the door to a vast new world of synthetic genetics, where sequence-defined synthetic polymers can be used to create new tools for biotechnology and medicine, and possibly even improve our understanding of the origin of life.

### Methods

#### Synthetic procedures

All reagents and solvents of highest purity were purchased from commercial suppliers and used without further purification. Reactions were run under an inert atmosphere of nitrogen or argon, unless otherwise indicated. Thin-layer chromatography (TLC) was performed using Silica Gel 60 F254-coated glass plates (0.250 mm thickness) and visualization was performed with UV irradiation. Chromatography was accomplished with Sil-Aflash 60 (230–400 mesh) and THF (4 mL) was added 1 M TBAF in THF (1 mL). After stirring for 30 min at 24 °C, the solvents were evaporated and the crude material was purified by silica gel chromatography [50–100% EtOAc-Hexane containing 1% TEA].

Next, to a solution of 2 (0.34 g, 1.15 mmol) in anhydrous toluene (20 mL) was added 0.5 M AIBN in toluene (10 mL, 20 mmol) and 1 M BuSnH in cyclohexane (3.5 mL, 3.5 mmol). After heating for 45 min at 100 °C, the mixture was evaporated and purified by silica gel chromatography (50–100% EtOAc-Hexane containing 1% TEA). The pure product was obtained as a solid foam (0.86 g, 85%).

#### Chemical synthesis

**N**-Benzoyl-9-′[(4′,4′-dimethoxytriphenyl)methyl]-2′-deoxy-α-β-threo-furanosyladenosine (4). A solution of 9′-(4′,4′-dimethoxytriphenyl)methyl)-2′-deoxy-β-ribofuranosyladenosine (3) (0.20 g, 0.6 mmol) in methanol (15 mL) was added 0.5 M MgSO4 and evaporated to dryness. The pure product (2) was obtained as a white powder (0.17 g, 87%, 80%) after silica gel chromatography (50–100% EtOAc-Hexane containing 1% TEA).

**N**-[(tert-Butyldiphenylsiloxy)hexanoyl]-9-′[(3′,3′,4′)-4′-deoxy-α-β-threo-furanosyl]adenosine (5). A solution of 4 (0.24 g, 0.46 mmol) in dry THF (4 mL) was added 0.5 M DIPEA/THF (4 mL) and THF (4 mL). The pure product was obtained as a solid foam (0.23 g, 72%).

**N**-[(4′,4′-dimethoxytriphenyl)methyl]-2′-deoxy-α-β-threo-furanosyl]adenosine (6). A solution of 5 (0.175 g, 0.20 mmol) in THF (4 mL) was added 1 M TBAF in THF (1 mL). After stirring for 18 h at 70 °C, the solvents were evaporated and the crude material was purified by silica gel chromatography (20–80% EtOAc-Hexane containing 0.5% TEA). The pure product was obtained as a solid foam (0.20 g, 83%).

**N**-[(4′,4′-dimethoxytriphenyl)methyl]-2′-deoxy-α-β-threo-furanosyl]adenosine (7). A solution of 6 (0.175 g, 0.20 mmol) in THF (4 mL) was added 1 M TBAF in THF (1 mL). After stirring for 18 h at 70 °C, the solvents were evaporated and the crude material was purified by silica gel chromatography (20–80% EtOAc-Hexane containing 0.5% TEA). The pure product was obtained as a solid foam (0.175 g, 83%).

**N**-[(4′,4′-dimethoxytriphenyl)methyl]-2′-deoxy-α-β-threo-furanosyl]adenosine (8). A solution of 7 (0.175 g, 0.175 mmol) in THF (4 mL) was added 0.5 M DIPEA/THF (4 mL). The pure product was obtained as a solid foam (0.175 g, 83%).

**N**-[(4′,4′-dimethoxytriphenyl)methyl]-2′-deoxy-α-β-threo-furanosyl]adenosine (9). A solution of 8 (0.175 g, 0.175 mmol) in THF (4 mL) was added 0.5 M DIPEA/THF (4 mL). The pure product was obtained as a solid foam (0.175 g, 83%).

**N**-[(4′,4′-dimethoxytriphenyl)methyl]-2′-deoxy-α-β-threo-furanosyl]adenosine (10). A solution of 9 (0.175 g, 0.175 mmol) in THF (4 mL) was added 0.5 M DIPEA/THF (4 mL). The pure product was obtained as a solid foam (0.175 g, 83%).

**N**-[(4′,4′-dimethoxytriphenyl)methyl]-2′-deoxy-α-β-threo-furanosyl]adenosine (11). A solution of 10 (0.175 g, 0.175 mmol) in THF (4 mL) was added 0.5 M DIPEA/THF (4 mL). The pure product was obtained as a solid foam (0.175 g, 83%).

**N**-[(4′,4′-dimethoxytriphenyl)methyl]-2′-deoxy-α-β-threo-furanosyl]adenosine (12). A solution of 11 (0.175 g, 0.175 mmol) in THF (4 mL) was added 0.5 M DIPEA/THF (4 mL). The pure product was obtained as a solid foam (0.175 g, 83%).

**N**-[(4′,4′-dimethoxytriphenyl)methyl]-2′-deoxy-α-β-threo-furanosyl]adenosine (13). A solution of 12 (0.175 g, 0.175 mmol) in THF (4 mL) was added 0.5 M DIPEA/THF (4 mL). The pure product was obtained as a solid foam (0.175 g, 83%).

**N**-[(4′,4′-dimethoxytriphenyl)methyl]-2′-deoxy-α-β-threo-furanosyl]adenosine (14). A solution of 13 (0.175 g, 0.175 mmol) in THF (4 mL) was added 0.5 M DIPEA/THF (4 mL). The pure product was obtained as a solid foam (0.175 g, 83%).

**N**-[(4′,4′-dimethoxytriphenyl)methyl]-2′-deoxy-α-β-threo-furanosyl]adenosine (15). A solution of 14 (0.175 g, 0.175 mmol) in THF (4 mL) was added 0.5 M DIPEA/THF (4 mL). The pure product was obtained as a solid foam (0.175 g, 83%).

**N**-[(4′,4′-dimethoxytriphenyl)methyl]-2′-deoxy-α-β-threo-furanosyl]adenosine (16). A solution of 15 (0.175 g, 0.175 mmol) in THF (4 mL) was added 0.5 M DIPEA/THF (4 mL). The pure product was obtained as a solid foam (0.175 g, 83%).
Oligonucleotide synthesis. TNA modified oligonucleotides were synthesized on an Applied Biosystems 3400 DNA synthesizer using standard β-cyanoethyl phosphoramidite chemistry (Supplementary Table 1). The P1 primer was synthesized on a universal Support II CPG column (1 μm scale, Glen Research) using chemically synthesized 1α-P phosphoramidite. The P2 primer was synthesized on the chemically synthesized 2′-deoxythreofuranosyl adenosine-modified CPG (Fig. 2). TNA oligonucleotides were obtained after phosphoramidite on-solution deprotection on a universal Support II CPG column (1 μm scale using standard DNA coupling conditions10). Cleavage from the solid support by temperature equilibration. The data were then loaded into separate capillaries and loaded onto the Thermopol buffer. The reaction mixtures were then incubated for 60 min at 23 °C.

CPG was then placed into three DNA synthesis columns and oxidized and capped for 15 min at 24 °C. The CPG was then eluted from the column using DCM (100 μL). After stirring for 1 h at 24 °C, the reaction was added to dried detritylated dT-CPG (R2) (150 mg) and 0.25 M EDT in McCN (0.5 mL) and stirred for 15 min at 24 °C. The CPG was filtered, washed with DCM, and MeCN. The CPG was then placed into three DNA synthesis columns and oxidized and capped using standard DNA synthesis protocol at 1 μm scale.

Kod-Ri expression and purification. The Kod-Ri gene was PCR amplified from a previously constructed vector, pGDRI1-Kod-Ri10 which additionally harbors two mutations (D141A and E143A) to inactivate exonuclease activity, using Kod-Ri_for and Kod-Ri_rev primers (IDT) containing Ndel and NotI restriction enzyme sites, respectively (Supplementary Table 1). Purified PCR product and pET23 (Novagen) were digested with Ndel and NotI restriction enzymes (NEB) and ligated into the resulting pET21–Kod-Ri construct was sequence verified (Retrogen). Acellα® cells (Edge BioSystems) harboring pET21–Kod-Ri were grown aerobically at 37 °C in LB medium containing 100 μg·mL−1 ampicillin. At an OD600 of 0.8, expression of a tagless Kod-Ri was induced with 0.8 mM isopropyl β-D-thiogalactoside at 18 °C for 20 hr. The cells were harvested by centrifugation for 20 min at 3315 × g at 4 °C and lysed in 40 mL lysis buffer (10 mM Tris·Cl pH 7.5, 100 mM NaCl, 0.1 mM EDTA, 1 mM DTT, 10% glycerol, 5 mg egg hen lysozyme) by sonication. The cell lysate was centrifuged at 23,708 × g for 30 min and the clarified supernatant was loaded onto 5 mL Hitrap Q HP and heparin HP columns (GE) assembled in series with the efflux of the Q column loaded in front of the heparin column. After washing with lysis buffer, the Q column was removed and Kod-Ri was eluted from the heparin column with a high salt buffer (10 mM Tris·Cl pH 7.5, 1 M NaCl, 0.1 mM EDTA, 1 mM DTT, 10% glycerol) using a linear gradient. Eluted fractions containing Kod-Ri were visualized by SDS–PAGE, pooled, and concentrated using a 30 kDa cutoff Amicon centrifugal filter (Millipore). Further purification was achieved by size exclusion chromatography (Superdex 200 HiLoad 16/600, GE) pre-equilibrated with Kod-Ri buffer (30 mM Tris·Cl pH 8.5, 200 mM NaCl, 0.1 mM EDTA, 1 mM DTT). Purified Kod-Ri was concentrated to 10 mg·mL−1 for crystallization trials.

TNA synthesis. Primer-extension reactions were performed in a final volume of 10 μL using the PBS8 primer (5′-IR008-label-CTGCTCTGGTGGATCCACCC-3′) and the L11 library (5′-GGATGCTCGTTGATCTGAGATGGTTGATCCCGAGGGGAC-3′, where N is the random region). Each reaction contained 30 pmol primer/template complex, 1× ThermostoPol buffer [20 mM Tris·HCl, 10 mM (NH4)2SO4, 10 mM KCl, 2 mM MgSO4, 0.1% Triton X-100, pH 8.8], 1 μM Kod-Ri, 100 μM of each tNTP, and 1 mM MnCl2. Reactions were incubated for 30 min and the clarified supernatant was loaded onto 5 mL Hitrap Q HP and heparin HP columns (GE) assembled in series with the efflux of the Q column loaded in front of the heparin column. After washing with lysis buffer, the Q column was removed and Kod-Ri was eluted from the heparin column with a high salt buffer (10 mM Tris·Cl pH 7.5, 1 M NaCl, 0.1 mM EDTA, 1 mM DTT, 10% glycerol) using a linear gradient. Eluted fractions containing Kod-Ri were visualized by SDS–PAGE, pooled, and concentrated using a 30 kDa cutoff Amicon centrifugal filter (Millipore). Further purification was achieved by size exclusion chromatography (Superdex 200 HiLoad 16/600, GE) pre-equilibrated with Kod-Ri buffer (30 mM Tris·Cl pH 8.5, 200 mM NaCl, 0.1 mM EDTA, 1 mM DTT). Purified Kod-Ri was concentrated to 10 mg·mL−1 for crystallization trials.

Cryocrystallization procedures. All reagents purchased from commercial suppliers were of analytical grade. Stock solutions of sodium sulfate dehydrate (Fluka), polyethylene glycol 3000 and 3350 (Sigma Aldrich), and 2-(N-morpholino)ethanesulfonic acid (Calbiochem) were filtered before use. The additives, 1.6-Hexanediol and Silver Bullets Bio 50 (Hampton Research), were used without further purification. All crystallization conditions were prepared as 10 mM tATP was replaced by tA. The initial models were used as the incoming substrate; however, due to poor electron density, the final models were determined using iterative rounds of manual building through Coot34 and refinement with phenix.35 The final stages of refinement employed TLS parameters; unless specified, Kod-Ri was partitioned into 4 TLS groups (i.e., 1–156, 157–304, 305–532, 533–756) while the template and primer strands contributed an additional TLS group each. The stereochemistry and geometry of all structures were evaluated with Molprobity36, with the final acceptable stereochemistry parameters summarized in Table 1. Final coordinates and structure factors have been deposited in the Protein Data Bank. All molecular graphics were prepared with PyMOL37.

Apo Kod-Ri: Apo Kod-Ri (1 mg·mL−1) crystallized in 0.1 M 2 (N-morpholinio)ethanesulfonic acid pH 6.0, and 20% polyethylene glycol 3300. MR was performed using X-ray synchrotron data (Advanced Light Source and Stanford Synchrotron Radiation Lightsource) from single crystals. Unless specified, images were indexed, integrated, and merged using iMOSELM52. Data collection statistics are summarized in Table 1. Initial models were determined by molecular replacement (MR) using Phaser38 and all final models were determined using iterative rounds of manual building through Coot34 and refinement with phenix.35 The final stages of refinement employed TLS parameters; unless specified, Kod-Ri was partitioned into 4 TLS groups (i.e., 1–156, 157–304, 305–532, 533–756) while the template and primer strands contributed an additional TLS group each. The stereochemistry and geometry of all structures were evaluated with Molprobity36, with the final acceptable stereochemistry parameters summarized in Table 1. Final coordinates and structure factors have been deposited in the Protein Data Bank. All molecular graphics were prepared with PyMOL37.
% w/w 1.6-Hexanediol, and 19 % polyethylene glycol 3350 for 45 min. The images were indexed, integrated, and merged using HKL200058 and MR was performed using an early open ternary complex structure, with tATP as its substrate, as the search model. During initial model building, the finger domain was deleted and rebuilt based on finger domain of the structurally aligned yeast Pol 8 (PDB ID: 3JAYv2). The final closed ternary complex structure spans residues 1–759 and contains a total of 23 residues mutated to alanines (i.e., E150, E154, K200, K221, Q285, R346, E385, R394, R476, I528, K531, K638, R668, L704, K705, I710, R713, H725, K726, Y727, D728, Q736, and R751) as well as two missing residues (i.e., E658 and R689) due to poor electron densities and three magnesium ions. PDB ID: 5VU8.

Translocated product: The binary complex with the P1/T duplex was prepared using a similar protocol and 5 M excess tATP was added and incubated at 37 °C for 30 min before crystallization. Co-crystals of the translocated complex were grown in 0.2 M sodium sulfate decahydrate, 0.1 M 2-(N-morpholino)ethanesulfonic acid (pH 5.8), and 12 % polyethylene glycol 3350, supplemented with Silver Bullets Bio (Hampton Research) additive #56. MR was performed using an early binary complex (with P1/T duplex) structure as the search model. During initial model building, the P1/T duplex was translocated and a TNA A model, tA13, was added complex (with P1/T duplex) structure as the search model. During initial model building, the final closed ternary complex structure spans residues 1–759 and contains 363 water molecules. PDB ID: 5VU9.

Data availability. Coordinates and structure factors for all five Kdo-RI crystal structures have been deposited in the PDB with the accession codes: 5VU5, 5VU6, 5VU7, 5VU8, and 5VU9. Other data are available from the corresponding author upon reasonable request.

Received: 15 September 2017 Accepted: 1 November 2017
Published online: 27 November 2017

References

1. Chaput, J. C., Yu, H. & Zhang, S. The emerging world of synthetic genetics. Nature 469, 1360–1371 (2012).
2. Pinheiro, V. B. & Holliger, P. The Xna world: progress toward replication and evolution of synthetic genetic polymers. Curr. Opin. Chem. Biol. 16, 245–252 (2012).
3. Anosova, I. et al. The structural diversity of artificial genetic polymers. Nucleic Acids Res. 44, 1007–1021 (2016).
4. Pinheiro, V. B., Loakes, D. & Holliger, P. Synthetic polymers and their potential as genetic materials. Bioessays 35, 113–122 (2012).
5. Joyce, G. F. Toward an alternative biology. Nature Rev. Genet. 19, 1360–1371 (2012).
6. Alves Ferreira-Bravo, I., Cozens, C., Holliger, P. & DeStefano, J. J. Selection of unnatural substrates by sampling mutations in homologous amber suppressors. EMBO J. 32, 394–395 (2013).
7. Perera, R. L. et al. Mechanism for priming DNA synthesis by yeast DNA polymerase alpha. Elife 2, e00482 (2013).
8. Swam, M. S., Johnson, R. E., Prakash, L., Prakash, S. & Aggarwal, A. K. Structural basis of high-fidelity DNA synthesis by yeast DNA polymerase delta. Nat. Struct. Mol. Biol. 16, 978–986 (2009).
9. Hogg, M. et al. Structural basis for processive DNA synthesis by yeast DNA polymerase varepsilon. Nat. Struct. Mol. Biol. 21, 49–55 (2014).
10. Dunn, M. R. & Chaput, J. C. Reverse transcription of threose nucleic acid by a naturally occurring DNA polymerase. ChemBioChem 17, 1804–1808 (2016).
11. Grieb, M., Kotera, C., Nehson, R. & Viljoen, H. Kinetics of the DNA polymerase pyrococcus kodakaraensis. Chem. Eng. Sci. 61, 3893–3892 (2006).
12. Kennedy, E. M., Hergott, C., Dwurst, S. & Kim, B. The mechanistic architecture of thermostable pyrococcus furiosus B DNA polymerase motif A and its interaction with DNTP substrates. Biochemistry 48, 1116111168 (2009).
13. Sau, S. P., Fahmi, N. E., Liao, J. Y., Bala, S. & Chaput, J. C. A scalable synthesis of α-threose nucleic acid monomers. J. Org. Chem. 81, 2302–2307 (2016).
14. Sau, S. P. & Chaput, J. C. A one-pot synthesis of α-threofuranosyl nucleoside triphosphates. Bioorg. Med. Chem. Lett. 26, 3271–3273 (2016).
15. Bergen, K., Betz, K., Welte, W., Diederichs, K. & Marx, A. Structures of KOD and 9°N DNA polymerases complexed with primer template duplex. ChemBioChem 14, 1058–1063 (2013).
16. Wyne, S. A., Pinheiro, V. B., Holliger, P. & Leslie, A. G. Structures of an apo and a binary complex of an evolved archael B family DNA polymerase capable of synthesising high-cy-dye labelled DNA. PLoS ONE 8, e70982 (2013).
17. Hashimoto, H. et al. Crystal structure of DNA polymerase from hyperthermophilic archaeon Pyrococcus Kodakaraensis KOD1. J. Mol. Biol. 306, 469–477 (2001).
18. Rodriguez, A. C., Park, H.-W., Mao, C. & Reese, L. S. Crystal structure of a Pol A family DNA polymerase from the hyperthermophilic Archaeon Thermococcus sp. 9°N-7. J. Mol. Biol. 299, 447–462 (2000).
19. Hottin, A. & Marx, A. Structural insights into the processing of nucleobase-modified nucleotides by DNA polymerases. Acc. Chem. Res. 49, 418–427 (2016).
20. Brathwaite, D. K. & Ito, J. Compilation, alignment, and phylogenetic relationships of DNA polymerases. Nucleic Acids Res. 21, 787–802 (1993).
21. Yang, G., Franklin, M., Li, J. C. & Konigsberg, W. Correlation of the steric gate with processive DNA synthesis by B-family polymerases. Structure 18, 1058–1063 (2010).
22. Letvin, L. N., McArthur, L. A., Hayden, F. G., Spear, P. & Podolsky, R. K. Herpesvirus glycoprotein D as an early T helper cell mitogen. Science 247, 347–351 (1990).
23. Small, P. R., Letvin, N. L. & Hottz, J. C. The structure and function of a T-helper cell mitogen. Annu. Rev. Immunol. 11, 559–587 (1993).
24. Hough, M. et al. Structural studies of RT polymerase complexes by X-ray crystallography. J. Mol. Biol. 212, 135–148 (1990).
25. Hogg, M. et al. Structural studies of RT polymerase complexes by X-ray crystallography. J. Mol. Biol. 212, 135–148 (1990).
26. Brown, J. A. & Suo, Z. Unlocking the sugar “steric gate” of DNA polymerases. Biochemistry 50, 1135–1142 (2011).
27. Lu, X. J. & Olson, W. K. JDNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. Nucleic Acids Res. 31, 5108–5112 (2003).
28. Cozens, C., Pinheiro, V. B., Vaisman, A., Woodgate, R. & Holliger, P. A short adaptive path from DNA to RNA polymerases. Proc. Natl Acad. Sci. USA 109, 8067–8072 (2012).
29. Cavichioli, R. Archaea—a timeline of the third domain. Nat. Rev. Microbiol. 9, 51–61 (2011).
30. Pavlov, A. R., Pavlova, N. V., Kozyakin, S. A. & Slesarev, A. I. Recent developments in the optimization of thermostable DNA polymerases for efficient applications. Trends Biotechnol. 22, 253–260 (2004).
31. Gardner, A. F. & Jack, W. E. Determinants of nucleotide sugar recognition in an archaeon DNA polymerase. Nucleic Acids Res. 27, 2543–2553 (1999).
32. Hough, M., Aragundy-Franklin, S. & Holliger, P. Exploring the chemistry of genetic information storage and propagation through polymerase engineering. Acc. Chem. Res. 50, 1079–1087 (2017).
33. Hopfner, K. P. et al. Crystal structure of a thermostable type B DNA polymerase from Thermococcus gordonii. Proc. Natl Acad. Sci. USA 96, 3660–3665 (1999).
Acknowledgements

We thank members of the Chaput laboratory, T. Poulos, M. Egli, R. Spitale, and C. Switzer for helpful discussions and critical reading of the manuscript. This work was supported by the DARPA Folded Non-Natural Polymers with Biological Function Fold F (x) Program under award number N66001-16-2-4061 and the National Science Foundation (MCB-1607111). Data sets were collected at the Advanced Light Source (ALS) and the Stanford Synchrotron Radiation Lightsource (SSRL), which are supported by the DOE (Contract Nos. DE-AC02-05CH11231 and DE-AC02-76SF00515, respectively). The SSRL Structural Molecular Biology Program is further supported by the National Institutes of Health, National Institute of General Medical Sciences (including P4GM103393).

Author contributions

N.C., C.S., and J.C.C. conceived the experiments. S.S. synthesized the P1 and P2 primers. A.N. performed the kinetic and Kd analyses. N.C. and C.S. performed all the experiments that led to Kod-RI crystals. N.C. collected and processed all X-ray diffraction data sets. N.C. and J.C.C. wrote the manuscript. All authors reviewed and commented on the manuscript.

Additional information

Supplementary Information accompanies this paper at https://doi.org/10.1038/s41467-017-02014-0.

Competing interests: The authors declare no competing financial interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access

This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2017