Published online 9 July 2018

Nucleic Acids Research, 2018, Vol. 46, No. 15 ***7977–7988***
doi: 10.1093/nar/gky552

Snapshots of an evolved DNA polymerase pre- and post-incorporation of an unnatural nucleotide

Isha Singh[1], Roberto Laos[2], Shuichi Hoshika[2], Steven A. Benner[2,3,*] and Millie M. Georgiadis[1,*]

[1]Department of Biochemistry & Molecular Biology, Indiana University School of Medicine, Indianapolis, IN 46202, USA, [2]Foundation for Applied Molecular Evolution and the Westheimer Institute of Science & Technology, Alachua, FL 32615, USA and [3]Firebird Biomolecular Sciences LLC, Alachua, FL 32615, USA

Received February 20, 2018; Revised June 05, 2018; Editorial Decision June 06, 2018; Accepted June 15, 2018

ABSTRACT

The next challenge in synthetic biology is to be able to replicate synthetic nucleic acid sequences efficiently. The synthetic pair, 2-amino-8-(1-beta-D-2′- deoxyribofuranosyl) imidazo [1,2-a]-1,3,5-triazin-[8H]-4-one (trivially designated P) with 6-amino-3-(2′-deoxyribofuranosyl)-5-nitro-1H-pyridin-2-one (trivially designated Z), is replicated by certain Family A polymerases, albeit with lower efficiency. Through directed evolution, we identified a variant KlenTaq polymerase (M444V, P527A, D551E, E832V) that incorporates dZTP opposite P more efficiently than the wild-type enzyme. Here, we report two crystal structures of this variant KlenTaq, a post-incorporation complex that includes a template-primer with P:Z trapped in the active site (binary complex) and a pre-incorporation complex with dZTP paired to template P in the active site (ternary complex). In forming the ternary complex, the fingers domain exhibits a larger closure angle than in natural complexes but engages the template-primer and incoming dNTP through similar interactions. In the binary complex, although many of the interactions found in the natural complexes are retained, there is increased relative motion of the thumb domain. Collectively, our analyses suggest that it is the post-incorporation complex for unnatural substrates that presents a challenge to the natural enzyme and that more efficient replication of P:Z pairs requires a more flexible polymerase.

INTRODUCTION

DNA polymerases have evolved for billions of years to replicate natural DNA with error rates ranging from 10^{-6} to 10^{-8} misincorporations per cycle (1). One goal of synthetic biology has been to expand such replication processes to include modified deoxyribonucleotide triphosphates (dNTPs). For example, an entire class of modified nucleotides have fluorescent or other signaling groups on the C5 position of pyrimidines or the C7 position of deazapurines (2). These have many uses including next-generation sequencing (3). A basis for improved incorporation of cyanine-labeled dCTP has been proposed from an analysis of apo and binary structures of an evolved family B polymerase, Pfu-E10, which efficiently synthesizes Cy-labeled DNA (4).

The nucleobases in these examples retain the hydrogen bonding moieties in the same arrangement found in natural nucleobases. This similarity makes their number limited. Thus, synthetic biologists have also worked to expand the genetic alphabet by adding pairs that *change* the hydrogen bonding patterns that join them, seeking those that might be incorporated into DNA by polymerases. Polymerase development is required to enable a new direction in synthetic biology where investigators seek to create new genetic systems that support Darwinism, including those with both backbone modifications and nucleobase modification. Many hope to move these into living cells to create new forms of life that use alternative and/or expanded genetic alphabets (5–9).

Two strategies have been pursued in creating synthetic nucleobases resulting in two distinct kinds of added nucleotides. The first dispenses with hydrogen bonding entirely, relying on steric complementarity alone to force pairing specificity (10–13). The second are artificially expanded genetic information systems (AEGIS). These rearrange hydrogen bonding moieties on purine- and pyrimidine-like nucleobases, relying on hydrogen bonding complementarity to ensure faithful pairing (14,15). The AEGIS strategy adds as many as eight nucleotides forming as many as four orthogonal pairs.

One of the more useful AEGIS pairs to emerge joins 2-amino-8-(1-beta-D-2′- deoxyribofuranosyl) imidazo [1,2-

*To whom correspondence should be addressed. Tel: +1 317 278 8486; Fax: +1 317 274 4686; Email: mgeorgia@iu.edu
Correspondence may also be addressed to Steven A. Benner. Tel: +1 386 418 8085; Email: manuscripts@ffame.org

a]-1,3,5-triazin-[8H]-4-one (trivially designated **P**) with 6-amino-3-(2′-deoxyribofuranosyl)-5-nitro-1H-pyridin-2-one (trivially designated **Z**) (Figure 1). Proceeding from the major to the minor groove, **P** presents a hydrogen bond acceptor-acceptor-donor pattern that is Watson-Crick complementary to **Z**, which presents a hydrogen bond donor-donor-acceptor pattern. Like the G:C pair, the **P:Z** pair is joined by three hydrogen bonding interactions. However, to get its desired hydrogen bonding pattern, the heterocycle on **Z** must be joined to its deoxyribose derivative via a carbon-carbon bond (forming a C-glycoside), different from the nitrogen-carbon bond (N-glycoside) that holds the cytosine nucleobase to its deoxyribose derivative. To control its epimerization and redox reactivity, **Z** also contains a functional moiety not found in any encoded building block in biology, a nitro group at position 3 of the ring, analogous to position 5 in a standard pyrimidine nucleotide ring (15). The nitro group on **Z** may account for the ease with which six-letter GACT**ZP** DNA delivers especially tight binding receptors and ligands when driven by laboratory *in vitro* evolution (16–18).

DNA molecules containing single and multiple exemplars of **P:Z** pairs retain the essential 'aperiodic crystal' features of duplex DNA as demonstrated in structures of 16-mer duplexes that have a total of four (two consecutive) or six consecutive **P:Z** pairs. These are observed in B-form and A-form, respectively (19–21). However, **P:Z** pairs impart to the double helix some novel properties (19–21). In particular, the nitro group of **Z** imparts novel stacking interactions with pairs above and below it in the duplex. Further, **P:Z** pairs, while maintaining standard hydrogen-bonding capabilities in the minor groove, appear to give the duplex a wider major groove (19–21).

Not surprisingly, many DNA polymerases replicate DNA containing **P:Z** pairs. This has allowed the pair to contribute to many applications, including some in diagnostics and therapy (22,23). Further, it allows GACT**ZP** 6-letter DNA to evolve under selective pressure to create functional molecules (19–21,24). However, this replication is not perfect. Although mispairing between **P**:C or A:**Z** in the context of oligonucleotides results in significant decreases in melting temperature consistent with destabilization (25) and is therefore unlikely to occur during DNA replication, many natural polymerases misincorporate dGTP opposite template d**Z**, especially at higher pH, where deprotonation of d**Z** gives an anionic nucleobase that forms a complementary hydrogen bond to G (26,27). Further, incorporation stops after three or four consecutive **P:Z** pairs are incorporated (28). The polymerase itself, however, can be changed. For example, variants of the polymerase from *Thermus aquaticus* (*Taq*) have been generated that accept a range of natural and unnatural nucleotide analogs under a range of conditions. The *Taq* variant (E602V, A608A, I614M, E615G) incorporates both NTPs and dNTPs with the same efficiency (29). Another Taq variant (F73V, R205K, K219E, M236T, E434D, A608V) functions at higher temperatures (30). Variants called M1 (G84A, D144G, K314R, E520G, F598L, A608V, E742G) and M4 (D58G, R74P, A109T, L245R, R343G, G370D, E520G, N583S, E694K, A743P) extend substrates that have 3′-mismatches such as C:C or A:G (31).

Accordingly, compartmentalized self-replication (CSR) has been used as a directed evolution technology to create variants of KlenTaq DNA polymerase (which covers residues 290–832 of the native protein) that better incorporate d**Z**TP opposite template **P** were obtained. One variant that emerged from this process had four amino acid replacements (M444V, P527A, D551E and E832V) (32). It incorporates d**Z**TP opposite template **P** more efficiently than native KlenTaq, as judged by primer extension experiments (32). The variant also had special interest because some of the amino acid replacements observed in this variant and others recovered from similar experiments appeared to recapitulate natural evolution in the polymerase family. For example, in the evolutionary history leading to the polymerase from *Salmonella* species, an ancestral methionine was replaced at the site equivalent to 444 by a valine. Alanine likewise emerged in natural history at the site now holding proline 527 in the lineages giving rise to the T3 and T7 viral DNA polymerases. Further, various gram positive bacteria have evolved to have a glutamate at the site now holding Asp 551 in Taq polymerase (32).

Based on available crystal structures for KlenTaq-substrate complexes, none of the amino acids that are substituted in our selected polymerase are predicted to directly interact with the template/primer or incoming triphosphate. However, it was reasonable to ask whether these remote substitutions might have second order effects on the structural properties of the polymerase and to determine the structural basis for recognition of **P:Z** pairs.

MATERIALS AND METHODS

Protein production

DNA encoding amino acid residues 293–832 of the Klenow fragment of Taq polymerase with mutations leading to the following amino acid substitutions, M444V, P527A, D551E and E832V, (32) was PCR amplified using forward and reverse primers (F, 5′-ATATAGATCTGCTCTGGA AGAAGCGCCG-3′ and R, 5′-ATATAAGCTTTTATTA AACTTTCGCAGACAGCCAATCT-3′) containing BglII and HindIII restriction sites, respectively, and inserted between BamHI (compatible ends with BglII) and HindIII restriction sites of pSMT3 vector (gift from Dr. Christopher Lima, Sloan-Kettering Institute). The N-terminally hexahistidine-SUMO tagged protein was expressed in Rosetta cells in LB media for 4 h following induction with 0.5 mM IPTG at 37°C for 4 h. For purification, the harvested cells were resuspended in lysis buffer containing 50 mM phosphate pH 7.8, 0.3 M NaCl, and 20 mM imidazole and lysed using a French press (Aminco). Following lysis, the His-SUMO tagged protein was bound to Ni-NTA agarose and then subjected to on-column cleavage to remove the hexahistidine SUMO tag using the SUMO-specific Ulp1 protease. The cleaved protein was further purified using anion exchange chromatography (Q-Sepharose) in buffer containing 20 mM Tris–Cl pH 8.5, 1 mM EDTA and 1 mM beta-mercaptoethanol. The protein was eluted using a linear gradient of NaCl. Finally, the protein was purified by size exclusion chromatography and concentrated to 0.4 mM in buffer containing 20 mM Tris–Cl pH 7.5, 1 mM EDTA, 1 mM beta-mercaptoethanol and 150 mM NaCl. Purified

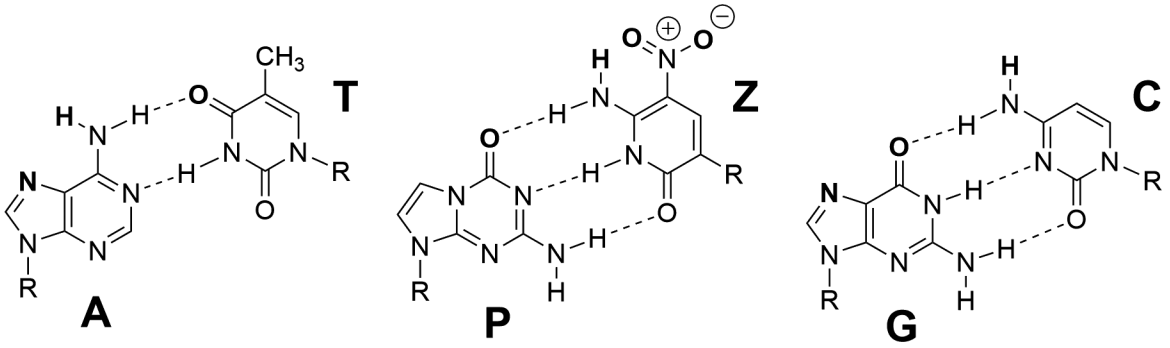

Figure 1. Chemical structures of standard nucleobase pairs, A:T and G:C, and the synthetic nucleobase pair **P**:**Z**.

protein was stored at 4°C. The E832V Taq polymerase variant was expressed, purified, and analyzed by PCR assays as previously reported (33).

Synthesis and purification of Z/P containing oligonucleotides

Standard phosphoramidites (Bz-dA, Ac-dC, dmf-dG and dT) and CPG having standard residues were purchased from Glen Research (Sterling, VA, USA), and AEGIS phosphoramidites (dZ and dP) and CPG having dZ residues were provided by Firebird Biomolecular Sciences LLC (Alachua, FL). All oligonucleotides containing d**Z** and d**P** were synthesized on an ABI 394 DNA Synthesizer following standard phosphoramidite chemistry and as previously reported (19). The CPGs having oligonucleotides were treated with 2.0 ml of 1 M DBU in anhydrous acetonitrile at room temperature for 24 h to remove the NPE group from the d**Z** nucleobase. Then the CPGs were filtered, dried, and treated with concentrated ammonium hydroxide at 55°C for 16 h. After removal of ammonium hydroxide, the oligonucleotides containing d**Z** and d**P** were purified on ion-exchange HPLC, and then desalted using Sep-Pac® Plus C18 cartridges (Waters) (19,25).

Crystallization of binary and ternary KlenTaq complexes

Templates and primers containing unnatural base pairs **Z** or **P** and d**Z**TP were resuspended in buffer containing 10 mM Tris–Cl pH 8.0, 10 mM NaCl and 5 mM $MgCl_2$. d**Z**TP was resuspended in 10 mM Tris–Cl pH 7.0 to give a final stock of 25 mM. Templates and primers were annealed at 70°C for 10 min at a final concentration of 1.5 mM prior to crystallization. For crystallization of the ZP binary complex, 0.12 mM KlenTaq was incubated with template (5′-AAAG**P**GCGCCGTGGTC-3′)–primer (5′-GAC CACGGCGC**Z**-3′) in a molar ratio of 1:1.5 with 20 mM $MgCl_2$ at 4°C for 1 h. The ternary complex on the other hand was obtained by incubating KlenTaq (0.12 mM) with template (5′-AAA**P**GGCGCCGTGGTC-3′)–primer (5′-G ACCACGGCGCC$_{dd}$) and d**Z**TP at a molar ratio of 1:2:10 along with 20 mM $MgCl_2$ at 4°C for 1 h. Crystals of both binary and ternary complexes were obtained by microseeding in vapor diffusion hanging drops incubated at 20°C and grew in 3 days. Crystals for the binary complex grew at 14% PEG 8000, 0.2 M magnesium formate, and 0.1 M MES pH 6.0 while the crystallization condition for the ternary complex included 19% PEG 4000, 0.2 M ammonium acetate, 0.01 M magnesium acetate, and 0.05 M sodium cacodylate pH 6.5. Both binary and ternary complex crystals were stabilized in 20% ethylene glycol with their respective reservoir solutions before flash freezing them in liquid nitrogen for storage prior to data collection. The ZP binary complex data (180°, 1° oscillations) were collected to a resolution of 2.66 Å at the LRL-CAT 31-ID beamline at a wavelength of 0.97931 Å and a temperature of 100 K. Data (180°, 0.2° oscillations) for the ZP ternary complex were collected at the SBC 19-ID beamline to a resolution of 2.35 Å at the Advanced Photon Source (APS) at Argonne, Illinois at a wavelength of 0.97926 Å and 100 K.

Structure determination and refinement

ZP binary crystals belonged to space group *C*2. The ZP binary data were processed, scaled and merged with *XDS* and *AIMLESS* (34–36). To obtain unbiased electron density for the DNA component, initial molecular replacement of the ZP binary complex was done using only the protein molecule from PDB ID: 3SZ2 (37) as the search model in *PHASER* (38). Following molecular replacement, the model was subjected to rigid body and restrained refinement in *REFMAC* (39). Initially, 10 bp of the template-primer were modeled in the electron density using *COOT* (40) and subsequently refined in *PHENIX* (41,42) using restrained least square refinement. To model the **P:Z** pairs, geometry restraints of these two unnatural base pairs were created using *eLBOW-PHENIX* (43). The remainder of the template-primer including modeling of the **P:Z** pairs was done using *COOT* followed by crystallographic refinement in *PHENIX*. TLS groups for *TLS* refinement in *PHENIX* were generated using TLSMD server (44). Crystallographic refinement statistics are reported in Table 1. The binary complexes of ZP polymerase include residues 296–644, 660–686 and 693–831 for chain A, 297–644 and 659–831 for chain C, and 294–645 and 657–831 for chain G. Each of the polymerase models is missing a loop in the fingers domain, model A is also missing part of an alpha helix.

ZP ternary complex data were processed using *HKL3000* (45), and data reduction was done in space group $P3_121$. Molecular replacement was done with *MOLREP* (46) using only the protein from PDB ID: 3RTV (37) as the search model. This ensured generation of unbiased electron den-

Table 1. Data collection and refinement statistics for ZP binary and ternary complex

	ZP binary	ZP ternary
Data collection		
Space group	*C*121	$P3_121$
Cell dimensions		
a, *b*, *c* (Å)	199.3, 114.6, 90.5	109.3, 109.3, 90.9
α, β, γ (°)	90, 90, 90	90, 90, 120
Resolution (Å)	99.34–2.66 (2.81–2.66)*	100–2.34 (2.39–2.35)*
R_{merge}	0.082 (0.452)	0.112 (0.399)
$I / \sigma I$	11.8 (2.2)	15.1 (2.5)
Completeness (%)	98.8 (97.8)	99.3 (96.5)
Redundancy	3.4 (2.7)	6.1 (3.5)
$CC_{1/2}$	0.73**	0.82**
Refinement		
Resolution (Å)	49.13–2.66	41.98–2.35
No. reflections	57 625	25 878
R_{work}/R_{free}	0.211/0.247	0.224/0.275
No. atoms		
Protein	12351	8335
DNA	1542	838
Water	295	169
B-factors		
Protein	40.3	46.3
DNA	38.3	32.3
Water	35	29
R.m.s. deviations		
Bond lengths (Å)	0.002	0.003
Bond angles (°)	0.703	0.590

* Both ZP binary and ternary datasets were collected for a single crystal.
* Values in parentheses are for highest resolution shell. **$CC_{1/2}$ values for the highest resolution shell.

sity for both template-primer duplex as well as the incoming nucleoside triphosphate (d**Z**TP for ternary complex). Initially, 10 bp of the template–primer DNA model were built using *COOT* followed by rigid body and restrained least squares refinement in *REFMAC*. d**Z**TP geometry restraints were created in *eLBOW-PHENIX,* and the restraint files were used to fit d**Z**TP and P in electron density in *COOT*. Multiple rounds of refinement were carried out using *PHENIX* and eventually a few rounds of *TLS* refinement using TLS groups identified by the TLMSD server. The final ZP polymerase model for the ternary complex includes residues 296–685, 700–831. An alpha helix comprising residues 686–699 is disordered in our structure. The data processing and refinement statistics for both ZP binary and ternary complexes are included in Table 1. Coordinate files for ZP binary and ternary complexes have been deposited with PDB, 5W6Q and 5W6K, respectively.

Structural analysis

The template-primer duplexes of both ZP binary and ternary complexes were analyzed using *3DNA* (47). Base pair and dinucleotide step parameters are calculated in 3DNA along with groove widths based on El Hassan and Calladine's algorithm. The contact distances between DNA and protein were calculated using *CONTACTS* in *CCP4* (48). Protein domain motion analysis was carried out using *DynDom* (49). Figures were created using *PYMOL* (50).

Modeling of natural KlenTaq with bound PZ-containing nucleic acid substrates

Hybrid binary and ternary models of natural KlenTaq and unnatural PZ nucleic acid substrates were created as follows. The unnatural KlenTaq (chain A in the unnatural binary/ternary complex) and natural KlenTaq (PDB IDs 3RTV, ternary and 3SZ2, binary) were superimposed using SSM in COOT. The transformation matrix obtained for superimposing the protein molecules was then applied to the unnatural PZ nucleic acid substrate in either the binary or ternary complex to create composite models.

RESULTS AND DISCUSSION

The evolved KlenTaq (M444V, P527A, D551E, E832V) used in these studies resulted from selection in a compartmentalized self-replication experiment; it was shown to more efficiently incorporate d**Z**TP opposite **P** in primer extension assays (32). Here, we report crystal structures of both binary and ternary complexes at 2.66 and 2.37 Å, respectively, the former referring to the evolved KlenTaq with productively bound template/primer representing a post-incorporation complex and the latter with bound template/primer and incoming nucleotide trisphosphate, a pre-incorporation complex (Figure 2). These structures provide insights on the overall properties of the variant KlenTaq as well as those of the variant amino acids M444V, P527A, and D551E; E832V at the carboxyl end of the protein is disordered but has been characterized independently. As expected, the variant KlenTaq retains the domain architecture of the natural enzyme, an N-terminal exonuclease domain, and polymerase fingers, palm, and thumb domains.

Variant KlenTaq exhibits increased domain movement

We consider first domain motion in the unnatural binary (PDB ID 5W6Q) versus natural binary (3SZ2 or 4KTQ) and unnatural ternary (5W6K) versus natural ternary complexes (3RTV or 3KTQ), unnatural referring to the evolved KlenTaq with unnatural template-primer (Table 2). For each pair, a slightly different strategy was used to obtain structures of binary and ternary complexes. In our study, the template/primer (5′-AAAG**P**GCGCCGTGGT C/5′-GACCACGGCGC**Z**) used for the binary complex was designed such that the **P:Z** pair would be bound in the active site of the enzyme, analogous to natural binary complexes (PDB IDs 3SZ2 and 4KTQ). In the ternary complex, the template/primer and nucleotide used were 5′-A AA**P**GGCGCCGTGGTC/5′-GACCACGGCGCC$_{dd}$ and d**Z**TP, which is paired to **P** in the active site, prior to incorporation. As was done for the natural ternary complex (PDB ID: 3RTV), the primer was terminated with a dideoxy nucleotide to trap the unnatural ternary complex. With the exception of the **P:Z** pairs, the template/primer used for our analysis is identical to that used for 3SZ2 and 3RTV. Furthermore, the sequence of the duplex portion of the template/primers for all of the structures included in this analysis is the same. The binary structure for 3SZ2 pairs C with ddG in the active site, while the analogous ternary complex 3RTV includes G:dCTP in the ac-

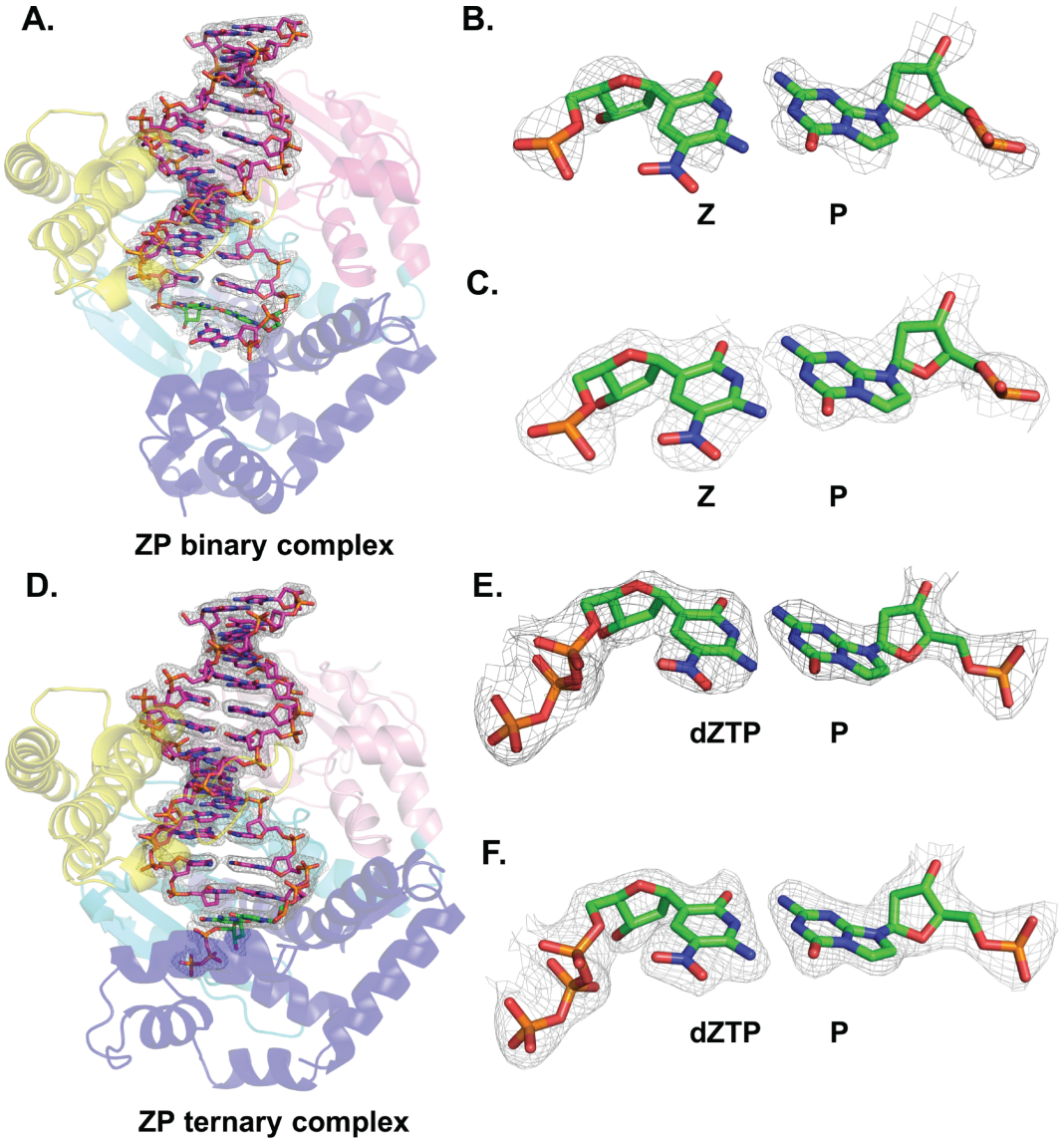

Figure 2. The crystal structure of the binary ZP KlenTaq polymerase complex is shown in (**A**) with the polymerase in a cartoon rendering (exonuclease, pink; fingers, blue; thumb, yellow and palm domain, cyan) and the template-primer as a stick rendering. The **Z:P** pair shown with carbons in green, all natural nucleobase pairs with carbons in magenta. The final $2F_o - F_c$ electron density map for the DNA contoured at 1 sigma is superimposed on the model. A close-up of **Z:P** in the binary complex is shown in (**B**) for the unbiased $F_o - F_c$ electron density map calculated prior to inclusion of DNA in the model contoured at 2 sigma and in (**C**) for the final $2F_o - F_c$ map in contoured at 1 sigma. The crystal structure of the ternary ZP KlenTaq polymerase complex (**D**) is shown in the same type of rendering as the binary complex in (A) with the final $2F_o - F_c$ map superimposed and contoured at 1 sigma. A close-up of the d**Z**TP:**P** in the ternary complex is shown in (**E**) for the initial $F_o - F_c$ map prior to inclusion of the DNA in the model contoured at 3 sigma and in (**F**) for the final $2F_o - F_c$ map contoured at 1 sigma.

tive site. For the other pair of structures, the binary structure (4KTQ) includes G:ddC paired in the active site, and the ternary structure includes the same duplex length for the template/primer with G:ddCTP paired in the active site (Table 2).

The ZP KlenTaq binary complex crystallized in a monoclinic lattice, space group *C*2, with three polymerase-template/primer complexes in the asymmetric unit rather than the trigonal lattice, space group $P3_121$, in which the natural binary structures have been reported with one polymerase–template/primer complex in the asymmetric unit (Supplementary Figure S1). For this analysis, ZP KlenTaq structures were compared to pairs of natural KlenTaq for which both binary and ternary complexes are available for the same template/primer substrates (3SZ2 (binary) versus 3RTV (ternary) and 4KTQ (binary) vs 3KTQ (ternary), see Table 2). The three ZP KlenTaq molecules are structurally similar to each other with rmsds of 0.31–0.42 Å and to the other binary KlenTaq models, rmsds of 0.69–0.86 Å to 3SZ2, and rmsds of 0.62–0.78 Å to 4KTQ (Supplementary Table S1). Relative domain movements were analyzed using DynDom (51) for binary ZP KlenTaq as compared to natural binary KlenTaq structures. Movement of ZP binary KlenTaq domains relative to 3SZ2 results from a rotation of

Table 2. Comparison of template/primers in binary and ternary complexes of ZP KlenTaq versus natural KlenTaq

PDB ID	Binary template/primer	Ternary template/primer/nucleotide triphosphate
5W6Q/5W6K (ZP KlenTaq)	AAAG**P**GCGCCGTGGTC **Z**CGCGGCACCAG	AAA**P**GGCGCCGTGGTC (**Z**=dZTP) **Z**CCGCGGCACCAG (C=ddC)
3SZ2/3RTV	AAAGCGCGCCGTGGTC GCGCGGCACCAG (G=ddG)	AAAGCGCGCCGTGGTC (**C**=dCTP) **C**GCGCGGCACCAG (G=ddG)
4KTQ/3KTQ	GGGCGCCGTGGTC CCGGGCACCAG (C=ddC)	GGGCGCCGTGGTC CCGCGGCACCAG (C=ddCTP)

Bold, unnatural nucleotides; underlined and bold, unnatural deoxyribonucleotide triphosphates; underlined, dideoxynucleotides.

3.1° to 4.4° for the moving domain, the thumb domain, vs. the fixed domain, including much of the palm and fingers domains, as identified by DynDom (Supplementary Figure S2, Table 3). There is essentially no translation observed but a calculated closure of 53% is consistent with domain closure as opposed to twisting (51). The movement is defined by a complex set of interactions involving residues within the fixed and moving domains. Although different strategies were used for trapping binary versus ternary complexes, no relative domain motion was detected for a comparison of ZP binary KlenTaq versus 4KTQ or for 3SZ2 versus 4KTQ.

The ZP KlenTaq ternary complex crystallized in a trigonal lattice, space group $P3_121$, with one ZP KlenTaq-template/primer:dZTP complex in the asymmetric unit and is very similar in structure to natural ternary counterparts as indicated by rmsds of 0.49 Å with 3RTV and 0.51 Å with 3KTQ. The natural KlenTaq structures are even more similar with an rmsd of 0.29 Å for superpositioning of 3RTV with 4KTQ structures (Supplementary Table S1). No relative domain motion was evident in pairwise comparisons of ZP KlenTaq to natural KlenTaq in ternary complexes.

Next, we analyzed domain motion of unnatural binary versus unnatural ternary complexes and natural binary versus natural ternary. A defining feature of the ternary complex is the closing down of the fingers domain positioning the O helix to interact with the substrate (Supplementary Figure S2). Similar to the natural complexes, the fingers domain closes down in the ternary complex of ZP KlenTaq with bound template/primer and d**Z**TP. Closure of the fingers domain in the ZP KlenTaq ternary versus the three ZP binary KlenTaq complex structural models results in an average rotation angle of 63.3° and translation of −1.7 to −2.0 Å, whereas a similar analysis for natural complexes results in rotation of the fingers domain by 59.6° for (3SZ2 (binary) versus 3RTV (ternary)) and 58.2° for (4KTQ (binary) versus 3KTQ (ternary)) with translations of −1.5 and −1.9 Å, respectively (Table 3).The fixed and moving domains as defined by DynDom (51) are similar for each of the pairwise comparisons as are the residues identified as 'bending'. The ZP KlenTaq in forming the ternary complex closes the fingers domain on average by 4.4° more than the analogous movement in the natural complexes (Table 3). As each of the three molecules of the binary complex has different packing interactions in the lattice, we conclude that the increased angle of closure for the fingers domain is a property of the ZP KlenTaq and not an artifact of lattice forces. This finding along with the observed domain motion in the binary complex is consistent with increased domain movement in ZP KlenTaq.

Effects of amino acid substitutions on the structure of KlenTaq

We next considered how individual amino acid substitutions might contribute to the domain motions observed in pairwise comparisons of polymerase complexes, which is potentially of broad interest for the future design of DNA polymerases. As shown in Figure 3A, M444V is located within the palm domain, while P527A and D551E are in the thumb domain. M444V is part of a closed pocket formed by two more methionine residues including M779 and M765. Although the M444V substitution might be viewed as fairly conservative, replacing Met by Val at position 444 introduces space for movement not present in the tightly packed region of the wild-type enzyme as illustrated in Figure 3B and C.

When considering the P527A replacement, it seemed likely that the substitution of *any* residue for Pro could impact the conformation of the main chain. Comparison of the ZP binary amino acid residues around 527 with 3SZ2 reveals a shift in position of the residues surrounding 527 (Figure 3D). This in turn alters interactions with the primer strand by the tip of the thumb.

No large impact was apparent from the substitution of D551 by E, comparing the ZP binary structure with other binary complexes and with natural ternary complexes. As the natural complex transitions from a binary to ternary structure, D551 of helix I maintains hydrogen bonding to R457 of helix H. In all of the natural KlenTaq structures, the positions of D/E 551 within helix I and R457 in helix H are very similar (Figure 3E). However, in a comparison of the ZP binary complex to ZP ternary complex, R457 adopts a slightly different conformation while still maintaining a hydrogen bonding interaction with E551 (Figure 3F). As the ZP binary complex transitions to the ternary complex, helix H including R457 is shifted slightly relative to the conformations observed in other pairwise comparisons. This shift may result from its proximity to V444, which, as noted above, fails to pack as tightly as M444 at the base of the palm. Thus, the D551E substitution might increase the length of the side chain just enough to maintain its hydrogen-bonding interaction with R457. This would allow helices H and I to stay linked while at the same time allowing for slight conformational differences that accom-

Table 3. Domain motion analysis

PDB File1	PDB File 2	Fixed domain	Moving domain	Rotation angle (°)	Translation (Å)	% Closure	Bending residues
Comparison of binary and ternary complexes							
ZP tern	ZP Bin A	298–636 672–829	637–671	62.0	−1.9	87.3	631–637 669–704
ZP tern	ZP Bin C	299–637 672–829	638–671	63.6	−1.7	88.8	632–638 670–704
ZP tern	ZP Bin G	298–637 672–829	638–671	64.5	−2.0	86.8	631–638 669–704
3RTV	3SZ2	295–637 672–830	638–671	59.6	−1.5	79.4	626–638 669–696
3KTQ	4KTQ	296–636 671–829	637–670	58.2	−1.9	67.8	626–637 670–686
Comparison of binary complexes							
ZP binary Chain A	3SZ2	306–453	454–457	3.8	−0.1	53.3	453–454
		458–461	462–541				457–458
		542–579	580–593				461–462
		594–597	598–599				541–542
		600–824	825–828				579–580
							593–594
							597–600
							824–825

(Note: Domain movement was analyzed using DynDom. There was no evidence of relative domain motion for the ZP ternary complex vs. other ternary complexes or the ZP binary versus 4KTQ, only 3SZ2. Analysis of the other two ZP binary complexes yielded similar results with rotation angles of 3.1° and 4.4°, respectively for the C and G chains).

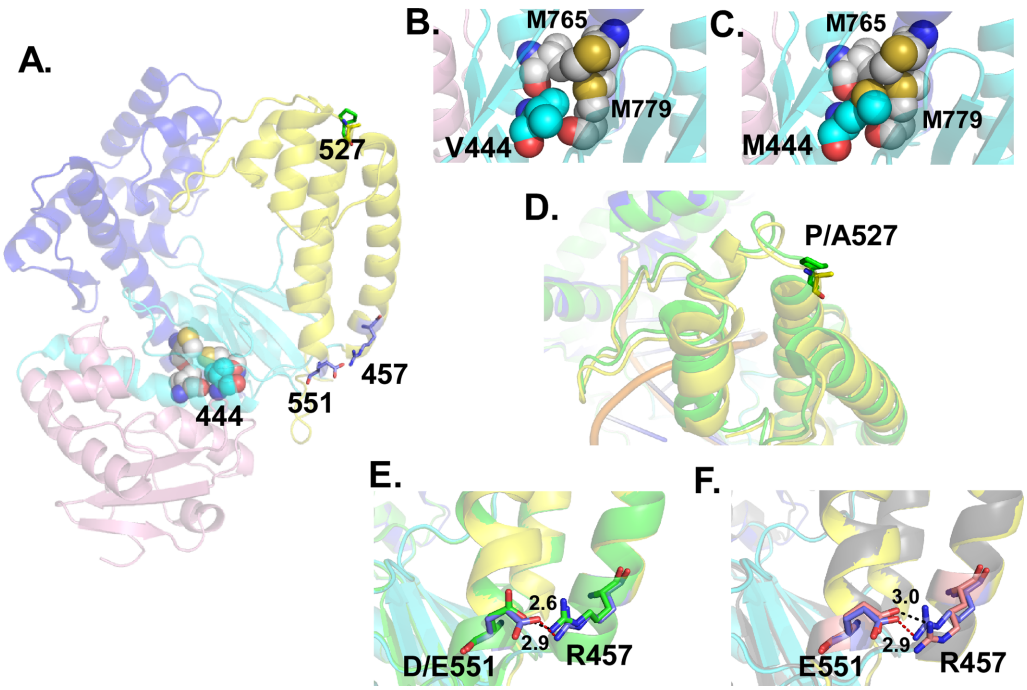

Figure 3. An overview of the evolved ZP KlenTaq is shown in (**A**) with the domains colored as in Figure 2 and the positions of the amino acid substitutions shown for 444, 551 and 527. Structural differences are shown for V444 (cyan) found in the ZP KlenTaq (**B**) versus M444 (cyan) in wild-type KlenTaq (**C**) with the residues shown as sphere models. ZP KlenTaq containing A527 (yellow stick model) is shown superimposed on wild-type KlenTaq with P527 (green stick model) in (**D**). (**E**) Hydrogen bonding interactions between D551E-R457 ZP KlenTaq (blue stick model, yellow cartoon thumb domain) are shown superimposed on wild-type KlenTaq D551-R457 (green stick model, green cartoon, PDB ID 3SZ2) in a comparison of binary complexes. (**F**) A comparison of E551–R457 in the ZP KlenTaq binary complex (blue stick model, yellow cartoon) versus E551–R457 in the ternary KlenTaq model (pink stick model, gray cartoon).

modate the unnatural **P**:**Z** or **P**:d**Z**TP pair. Further adaptation conferred by the P527A replacement at the tip of the thumb, which grips the primer strand, may also contribute to its ability to appropriately position the template/primer in the active site.

The C-terminal V832 in the evolved KlenTaq polymerase is disordered while the wild-type E832 is ordered in natural KlenTaq (4KTQ and 3RTV) structures, albeit modeled differently in each. E832 does not interact directly with the substrate in the natural structures. We can only speculate that V832 is disordered due to an inability to form specific interactions with neighboring structural elements as it does in the natural enzyme and may contribute to greater flexibility within the C-terminus. The E832V Taq polymerase characterized in the context of a single substitution was found to more efficiently incorporate d**Z**TP opposite **P** when compared to wild-type Taq as analyzed by PCR analysis, (Supplementary Figure S3 (33)). The E832V Taq polymerase does not incorporate d**Z**TP as well as the evolved variants but outperforms the native enzyme.

Substrate recognition by ZP KlenTaq

The nature of interactions found in unnatural complexes as compared to those in natural complexes is of broad interest in understanding how polymerases recognize unnatural substrates while retaining properties required for efficient DNA replication. In this study, we focus on determining whether specific interactions provide a basis for understanding how KlenTaq recognizes PZ-containing template-primers. Toward this goal, we first compared protein-nucleic acid interactions in the unnatural complexes with those in the natural complexes. Next, we considered whether the structures provide any insight on the ability of natural KlenTaq to incorporate dZTP less efficiently than the evolved polymerase (32).

In the natural and unnatural binary and ternary KlenTaq structures, the template/primer adopts B-form, with the exception of the two nucleobase pairs adjacent to the site of nucleotide incorporation in the active site of enzyme; those two pairs are A-form. The polymerase forms extensive contacts with the template-primer DNA, primarily with the sugar-phosphate backbone of DNA involving most of the same amino acid residues (Supplementary Figures S4 and S5, and Supplementary Tables S2 and S3) in both the natural and unnatural complexes. The terminal nucleobase pair positioned in the active site exhibits significant buckling with angles of 13.7° for G:ddC (4KTQ), 17.6° for C:ddG (3SZ2), and 22.6°, 22.5° and 16.8° for **P**:**Z** in the three complexes present in the ZP binary complex (Supplementary Figure S6). Thus, structural properties of the terminal nucleobase pair are conserved in both natural and unnatural binary complexes although the overall positioning of the template-primer in the ZP binary is shifted in the region outside of the active site as shown in a comparison with the 3SZ2 binary complex (Supplementary Figure S6).

Substrate recognition of the binary and ternary complexes involves distinct sets of interactions that position the template/primer or template/primer and dNTP appropriately in the active site of the polymerase. We compared the three **ZP** binary complexes present in our structure with each other and with two KlenTaq binary structures reported for natural complexes 3SZ2 (primer terminated by ddG) and 4KTQ (primer terminated by ddC). Minor groove hydrogen bonding interactions 'read' the terminal nucleobase pair in the active site. Q754 hydrogen bonds to either N3 or O2 of the template nucleobase, and R573 hydrogen bonds to N3 or O2 of the terminal primer nucleobase. In the ZP binary structure, the NH2 atom of R573 is 2.7 Å from O2 of **Z** in the minor groove, while the NE2 atom of Q754 is 3.6 Å from N3 of **P** in the minor groove (distances refer to the A chain). This difference is due to a slightly different positioning of the **P:Z** pair in the active site as compared to natural complexes (Figure 4A-C). Hydrogen bonding distances for Q754 to the template O2 of C or N3 of G in 3SZ2 and 4TKQ, respectively, were 3.3 and 3.2 Å.

For the structure with the ddG terminated primer (3SZ2), R587 hydrogen bonds directly to N7 in the major groove. In one of the three ZP binary complexes (A chain), R587 hydrogen bonds to an oxygen of the nitro group for the terminal **Z** (residue number 112) and to the phosphate O of **Z** (Figure 4B, refer to Supplementary Figure S4 for numbering scheme). For the ddC terminated complex (4KTQ), there is no hydrogen bond acceptor present in the major groove, and R587 adopts an alternate conformation and interacts with the phosphodiester backbone of the adjacent nucleotide (C 111, residue number in the coordinate file). With the template/primer appropriately positioned in the active site, Y671 stacks over **P** 205, which is Watson–Crick paired with **Z** in the active site of the binary structure. This prevents the next nucleobase in the 5′ single strand overhang of the template from stacking over the hydrogen-bonded template/primer pair.

To gain insight into the basis for recognition of d**Z**TP, we compared our ZP ternary complex to those of two natural ternary complexes, 3RTV and 3KTQ. d**Z**TP is positioned similarly to dNTP or ddNTP found in natural complexes, and its deoxyribose adopts a C3′ endo conformation as observed for incoming nucleotides of corresponding natural complexes, including 3RTV and 3KTQ. The triphosphate of d**Z**TP is coordinated by two magnesium ions, which are in turn directly coordinated by D610 and D785 of the polymerase. This is similar to the coordination of dCTP in the natural ternary complex 3RTV (Figure 5A and B).

Positioning of d**Z**TP involves hydrogen bonding of O2 in the minor groove to a water molecule, which is also hydrogen bonded to E615, N750 and Q754 as is seen in natural ternary complexes (Figure 5C and D). Similar to the ZP binary complexes, Q754 and R573 are positioned to hydrogen bond in the minor groove, in this case to the nucleobase pair adjacent to that formed by the incoming dNTP and complementary template nucleobase. For the natural ternary complexes, the NE2 atom of Q754 forms a water-mediated hydrogen bond to O2 of the incoming dNTP, and as observed in the ZP binary complex, the NE2 atom of Q754 is 3.6 Å from minor groove N3, this time of G 205 rather than **P**, while in the natural complexes, this distance is 3.4 Å. However, normal hydrogen bonding is maintained for the water-mediated interaction between Q754 and O2 of d**Z**TP and for R573, which is 2.8 Å from O2 of C 112. The positioning of Q754 relative to the template in both the binary and

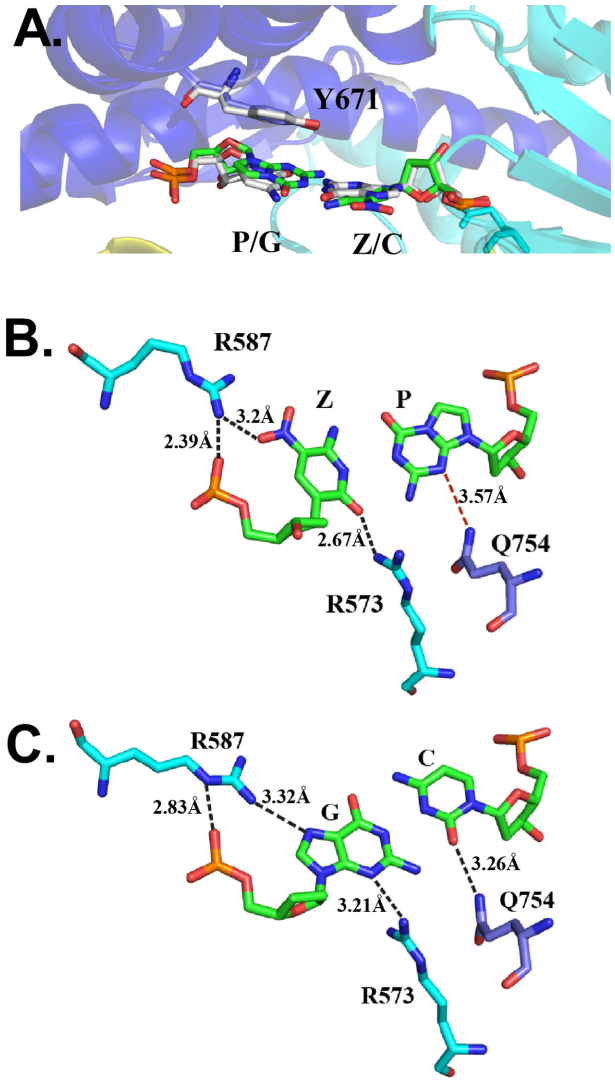

Figure 4. A comparison of superimposed ZP KlenTaq and wild-type KlenTaq binary complexes is shown in (**A**) with the two polymerase models shown as cartoons (fingers blue, palm cyan) and **P:Z** (green stick model), **G:C** (gray stick model), and Y671 (blue stick in ZP KlenTaq and gray stick in wild-type KlenTaq. Specific hydrogen bonding interactions of **Z:P** bound in the active site with R587 in the major groove and R573 (black dashes) along with Q754 (red dashes), which makes a longer range interaction, in the minor groove are shown in (**B**). All distances are listed in SI Supplementary Table S2. Analogous interactions are shown in (**C**) for the wild-type KlenTaq binary complex (PDB ID 3SZ2) with hydrogen bonding interactions shown as black dashes between the same residues.

ternary ZP complexes suggests that the architecture of the active site is slightly different in the variant KlenTaq than in the corresponding natural complexes.

In all of the ternary complexes, closing of the fingers domain positions F667 adjacent to the sugar moiety of the incoming dNTP providing a steric barrier in the region of the 2′-carbon. This feature allows DNA polymerases to distinguish between NTPs and dNTPs, as shown for the DNA polymerase Moloney murine leukemia virus reverse transcriptase (52,53). This same conformational change displaces Y671, located at the C-terminal end of the O-helix, allowing the next template nucleobase to pair with the incoming dNTP (Figures 4A and 5A, B). In the natural ternary complexes, the side chain of R660 hydrogen bonds to the phosphate of the 3′ nucleobase of the primer displacing R587, which adopts a conformation similar to that seen in two of the ZP binary complexes (chains C and G) and forms a hydrogen bond with the phosphate of nucleobase 111 (see Supplementary Figure S4 for numbering scheme). In the ZP ternary complex, the electron density for R660 is poorly ordered beyond the Cβ position but as currently modeled is in a conformation similar to that in 3RTV and 3KTQ structures with the NH2 atom of R660 positioned within hydrogen bonding distance of a nitro O atom and a phosphate oxygen of the 3′ terminal primer nucleobase (Figure 5C). This potential hydrogen bonding interaction is possible due to the fact that the nitro group extends out into the major groove. Similarly, R660 hydrogen bonds to ddGTP in the major groove to N7 and O8 atoms potentially contributing to the 10-fold higher incorporation rate observed for ddGTP over other ddNTPs (54).

In addition to the direct interactions of KlenTaq with the template/primer in these complexes, there are indirect interactions of the polymerase with DNA through water molecules that line the minor groove. Only two of the water-mediated contacts to the DNA that are found in all three **ZP** binary complex are also found in the natural DNA complex, namely those involving His 784 and Lys 540 (Supplementary Figure S3). A number of additional indirect contacts of polymerase with the **ZP** ternary complex DNA are also observed as highlighted in Supplementary Figure S4. Of potential interest is a water-mediated contact involving Q582 that is found in both binary and ternary ZP KlenTaq complexes but not in the natural complexes.

To address the question of how the natural polymerase interacts with unnatural PZ-DNA, hybrid models were created by superimposing our unnatural binary or ternary complex to the natural complex with the most similar template-primer sequence (PDB IDs 3SZ2 and 3RTV, respectively) and then applying the transformation matrix to either ZP binary or ZP ternary nucleic acid as described above. We recognize that this is a very simplistic approach but suggest that interactions predicted from the hybrid models are valid as a first approximation given that the natural and unnatural template-primer models alone superimpose with rmsds for all common atoms of 0.67 and 0.54 Å in binary and ternary complexes, respectively (Supplementary Figure S7), and as noted above, the KlenTaq models all superimpose with rmsds <1 Å (0.5–0.86 Å). Overall, this modeling exercise provided two critical observations. First, the nature of the interactions in the ternary hybrid complex, summarized in Supplementary Table S3, are very similar to those in the natural and unnatural complex with no observed clashes for atoms. The nucleic acid models are also extremely similar (Supplementary Figure S7). We further note that it has been possible to obtain crystal structures of ternary complexes of natural KlenTaq with other unnatural substrates including dNaM:5SICS (37) and Ds:Px (55).

Second, in the binary complex, although many of the interactions are conserved (Supplementary Table S2), the structure of the unnatural template-primer is sufficiently different that it results in two clashes, one between the side chain of R587 and the nitro group of **Z** on the primer strand

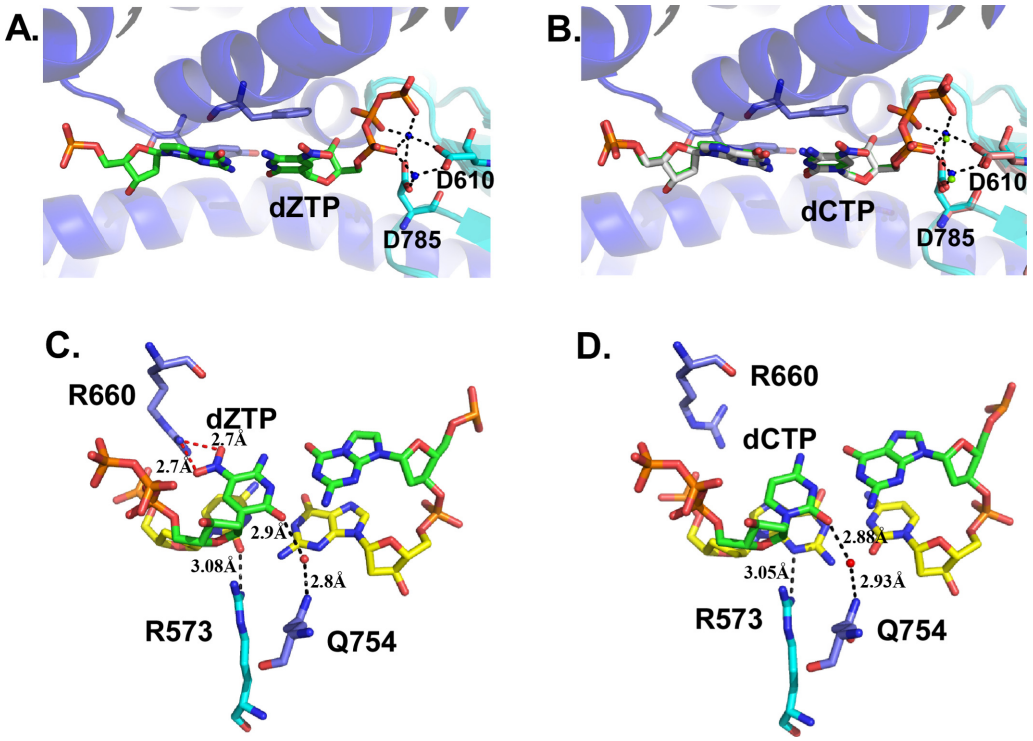

Figure 5. Coordination of the incoming d**Z**TP paired to template **P** by two Mg^{2+} ions (blue spheres) bound to D785 and D610 in the active site of ZP KlenTaq is shown in (**A**). (**B**) Wild-type KlenTaq ternary complex (PDB ID 3RTV) is shown with incoming dCTP paired to template G (gray stick model) with Mg^{2+} ions (green spheres) and superimposed with the ZP KlenTaq ternary complex. (**C**) Specific hydrogen bonding interactions in the minor groove to d**Z**TP are shown in the minor groove between R573 and a water mediated interaction with Q754 (black dashes). Similar hydrogen bonding interactions are shown in (**D**) for the wild-type KlenTaq ternary complex for dCTP. In the major groove, as modeled R660 is predicted to hydrogen bond to d**Z**TP in the major groove (red dashes) in (C) but not in the wild-type complex as dCTP lacks a hydrogen bond acceptor.

and a second between S674 and the backbone of the overhanging nucleotide of the template strand (Supplementary Figure S7). These clashes with both the primer and template strands potentially result from differences in the structure of the duplex DNA, which superimposes well in the middle at the expense of the ends (Supplementary Figure S7). We suggest that this is an inherent property of this template-primer given that the structures of the template-primers present in all three complexes within our binary structure all show the same differences but do not have the same lattice contacts (Supplementary Figure S7). This finding suggests that it is binding of the post-incorporation product, i.e. with **P:Z** in the active site, that presents a challenge to the natural polymerase. It also supports the finding that the natural polymerase pauses during incorporation of 4 d**Z**TPs in primer extension assays leading to intermediate products (see Supporting Information in (26)). We suggest that further investigation of this issue would best be accomplished through comprehensive molecular dynamics studies.

The post-incorporation complex is also a problem for other unnatural substrates. Although there is currently no available binary complex for Ds:Px bound in the active site for comparison, the synthetic nucleobase dNaM preferentially stacks with d5SICS in the active site of the polymerase in an intercalative mode in available binary structures resulting in an unusual intermediate form of the enzyme that is neither closed nor open (56).

CONCLUSIONS

Trapping the unnatural KlenTaq with unnatural substrates provides insights on how the evolved polymerase interacts with unnatural substrates in two different scenarios: following incorporation of d**Z**TP opposite **P** in a binary complex and with d**Z**TP poised for incorporation opposite **P** in a ternary complex. Here, laboratory evolution identified 'third shell' sites for amino acid replacements, sites that would not have been selected *a priori* by molecular modelers. These confer changes in the structure of the enzyme that collectively allow increased relative domain motion both in the binary and ternary complexes. Further, modeling of binary and ternary complexes for ZP-containing substrates with natural KlenTaq suggests that it is the post-incorporation step that benefits most from an evolved enzyme with increased relative domain motion.

It is quite remarkable that relatively subtle amino acid substitutions collectively confer increased relative domain motion of the thumb domain in the binary complex and fingers domain in the ternary complex for ZP KlenTaq. The broader impact of these findings is that identification of amino acid substitutions that have the potential to alter domain motion such as M444V should be considered in efforts to design polymerases or other multi-domain enzymes with altered properties. It is perhaps not so remarkable that the enzyme 'reads' **P:Z** pairs through hydrogen bonding in the minor groove given that this pair presents the expected

N3 and O2 atoms. Nonetheless, the observed minor groove interactions closely mirror those found in the natural complexes with subtle differences in the positioning of the template relative to Q754. Further, just as natural KlenTaq incorporated ddGTP with 10-fold higher efficiency than other ddNTPs (54), the potential for ZP KlenTaq to hydrogen bond in the major groove to the nitro group of **Z** either through R587 in the binary complex or R660 may confer preferred incorporation of d**Z**TP opposite template **P**.

DATA AVAILABILITY

Coordinates for ZP binary and ternary complex have been deposited with PDB, 5W6Q and 5W6K, respectively.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

This research used resources of the SBC and LRL beamlines at the Advanced Photon Source, a U.S. Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under Contract No. DE-AC02-06CH11357. We thank the staff at the SBC and LRL beamlines and Drs Thomas Hurley and Kishore Mahalingan for assistance with data collection.

FUNDING

National Institutes of Health [1R41GM119434, R01GM128186] (in part); National Aeronautics and Space Administration [NNX15AF46G]. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Aeronautics and Space Administration. Funding for open access charge: NIH (to S.B.).
Conflict of interest statement. None declared.

REFERENCES

1. Kunkel,T.A. (2004) DNA replication fidelity. *J. Biol. Chem.*, **279**, 16895–16898.
2. Chen,C.Y. (2014) DNA polymerases drive DNA sequencing-by-synthesis technologies: both past and present. *Front. Microbiol.*, **5**, 305.
3. Metzker,M.L. (2010) Sequencing technologies - the next generation. *Nat. Rev. Genet.*, **11**, 31–46.
4. Wynne,S.A., Pinheiro,V.B., Holliger,P. and Leslie,A.G. (2013) Structures of an apo and a binary complex of an evolved archeal B family DNA polymerase capable of synthesising highly cy-dye labelled DNA. *PLoS One*, **8**, e70892.
5. Neumann,H., Wang,K., Davis,L., Garcia-Alai,M. and Chin,J.W. (2010) Encoding multiple unnatural amino acids via evolution of a quadruplet-decoding ribosome. *Nature*, **464**, 441–444.
6. Pinheiro,V.B., Taylor,A.I., Cozens,C., Abramov,M., Renders,M., Zhang,S., Chaput,J.C., Wengel,J., Peak-Chew,S.Y., McLaughlin,S.H. *et al.* (2012) Synthetic genetic polymers capable of heredity and evolution. *Science*, **336**, 341–344.
7. Bande,O., Braddick,D., Agnellow,S., Jang,M., Pezo,V., Schepers,V., Rozenski,J., Lescrinier,E., Marliere,P. and Herdewijn,P. (2016) Base pairing involving artificial bases in vitro and in vivo. *Chem. Sci.*, **7**, 995–1010.
8. Malyshev,D.A., Dhami,K., Lavergne,T., Chen,T., Dai,N., Foster,J.M., Corrêa,I.R. and Romesberg,F.E. (2014) A semi-synthetic organism with an expanded genetic alphabet. *Nature*, **509**, 385–388.
9. Herdewijn,P. and Marliere,P. (2009) Toward safe genetically modified organisms through the chemical diversification of nucleic acids. *Chem. Biodivers.*, **6**, 791–808.
10. Johnson,S.C., Sherrill,C.B., Marshall,D.J., Moser,M.J. and Prudent,J.R. (2004) A third base pair for the polymerase chain reaction: inserting isoC and isoG. *Nucleic Acids Res.*, **32**, 1937–1941.
11. Kimoto,M., Kawai,R., Mitsui,T., Yokoyama,S. and Hirao,I. (2009) An unnatural base pair system for efficient PCR amplification and functionalization of DNA molecules. *Nucleic Acids Res.*, **37**, e14.
12. Malyshev,D.A., Dhami,K., Quach,H.T., Lavergne,T., Ordoukhanian,P., Torkamani,A. and Romesberg,F.E. (2012) Efficient and sequence-independent replication of DNA containing a third base pair establishes a functional six-letter genetic alphabet. *Proc. Natl. Acad. Sci. U.S.A.*, **109**, 12005–12010.
13. Morales,J.C. and Kool,E.T. (1999) Minor groove interactions between polymerase and DNA: more essential to replication than watson-crick hydrogen bonds? *J. Am. Chem. Soc.*, **121**, 2323–2324.
14. Geyer,C.R., Battersby,T.R. and Benner,S.A. (2003) Nucleobase pairing in expanded Watson-Crick-like genetic information systems. *Structure*, **11**, 1485–1498.
15. Benner,S.A. (2004) Understanding nucleic acids using synthetic chemistry. *Acc. Chem. Res.*, **37**, 784–797.
16. Sefah,K., Yang,Z., Bradley,K.M., Hoshika,S., Jimenez,E., Zhang,L., Zhu,G., Shanker,S., Yu,F., Turek,D. *et al.* (2014) In vitro selection with artificial expanded genetic information systems. *Proc. Natl. Acad. Sci. U.S.A.*, **111**, 1449–1454.
17. Zhang,L., Yang,Z., Sefah,K., Bradley,K.M., Hoshika,S., Kim,M.J., Kim,H.J., Zhu,G., Jimenez,E., Cansiz,S. *et al.* (2015) Evolution of functional six-nucleotide DNA. *J. Am. Chem. Soc.*, **137**, 6734–6737.
18. Zhang,L., Yang,Z., Le Trinh,T., Teng,I.T., Wang,S., Bradley,K.M., Hoshika,S., Wu,Q., Cansiz,S., Rowold,D.J. *et al.* (2016) Aptamers against cells overexpressing glypican 3 from expanded genetic systems combined with cell engineering and laboratory evolution. *Angew. Chem. Int. Ed. Engl.*, **55**, 12372–12375.
19. Georgiadis,M.M., Singh,I., Kellett,W.F., Hoshika,S., Benner,S.A. and Richards,N.G. (2015) Structural basis for a six nucleotide genetic alphabet. *J. Am. Chem. Soc.*, **137**, 6947–6955.
20. Molt,R.W. Jr., Georgiadis,M.M. and Richards,N.G.J. (2017) Consecutive non-natural PZ nucleobase pairs in DNA impact helical structure as seen in 50 mus molecular dynamics simulations. *Nucleic Acids Res.*, **45**, 3643–3653.
21. Richards,N.G.J. and Georgiadis,M.M. (2017) Toward an expanded genome: structural and computational characterization of an artificially expanded genetic information system. *Acc. Chem. Res.*, **50**, 1375–1382.
22. Benner,S.A., Kim,H.-J., Merritt,K.B., Yang,Z., McLendon,C., Hoshika,S. and Hutter,D. (2015) Next-generation DNA in pathogen detection, surveillance, and CLIA-waivable diagnostics. *SPIE Sens. Technol. Applic. SPIE Digital Library*, **9490**, 6.
23. Benner,S.A., Karalkar,N.B., Hoshika,S., Laos,R., Shaw,R.W., Matsuura,M., Fajardo,D. and Moussatche,P. (2016) Alternative Watson-Crick synthetic genetic systems. *Cold Spring Harbor Perspect. Biol.*, **8**, a023770.
24. Biondi,E., Lane,J.D., Das,D., Dasgupta,S., Piccirilli,J.A., Hoshika,S., Bradley,K.M., Krantz,B.A. and Benner,S.A. (2016) Laboratory evolution of artificially expanded DNA gives redesignable aptamers that target the toxic form of anthrax protective antigen. *Nucleic Acids Res.*, **44**, 9565–9577.
25. Yang,Z., Hutter,D., Sheng,P., Sismour,A.M. and Benner,S.A. (2006) Artificially expanded genetic information system: a new base pair with an alternative hydrogen bonding pattern. *Nucleic Acids Res.*, **34**, 6095–6101.
26. Yang,Z., Chen,F., Alvarado,J.B. and Benner,S.A. (2011) Amplification, mutation, and sequencing of a six-letter synthetic genetic system. *J. Am. Chem. Soc.*, **133**, 15105–15112.
27. Reichenbach,L.F., Sobri,A.A., Zaccai,N.R., Agnew,C., Burton,N., Eperon,L.P., de Ornellas,S., Eperon,I.C., Brady,R.L. and Burley,G.A. (2016) Structural basis of the mispairing of an artificially expanded genetic information system. *Chem*, **1**, 946–958.

28. Yang,Z., Sismour,A.M., Sheng,P., Puskar,N.L. and Benner,S.A. (2007) Enzymatic incorporation of a third nucleobase pair. *Nucleic Acids Res.*, **35**, 4238–4249.
29. Ong,J.L., 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. *J. Mol. Biol.*, **361**, 537–550.
30. Ghadessy,F.J., Ong,J.L. and Holliger,P. (2001) Directed evolution of polymerase function by compartmentalized self-replication. *Proc. Natl. Acad. Sci. U.S.A.*, **98**, 4552–4557.
31. Ghadessy,F.J., 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. *Nat. Biotechnol.*, **22**, 755–759.
32. Laos,R., Shaw,R., Leal,N.A., Gaucher,E. and Benner,S. (2013) Directed evolution of polymerases to accept nucleotides with nonstandard hydrogen bond patterns. *Biochemistry*, **52**, 5288–5294.
33. Laos,R. (2015) *PhD Thesis. Directed evolution of polymerases to accept nucleotides with non-strandard hydrogen bonding patterns*. University of Florida.
34. Kabsch,W. (2010) Xds. *Acta Crystallogr. D, Biol. Crystallogr.*, **66**, 125–132.
35. Evans,P.R. and Murshudov,G.N. (2013) How good are my data and what is the resolution? *Acta Crystallogr. D, Biol. Crystallogr.*, **69**, 1204–1214.
36. Kabsch,W.(1988) Automatic indexing of rotation diffraction patterns *J.Appl. Cryst.*, **21**, 916–924.
37. Betz,K., Malyshev,D.A., Lavergne,T., Welte,W., Diederichs,K., Dwyer,T.J., Ordoukhanian,P., Romesberg,F.E. and Marx,A. (2012) KlenTaq polymerase replicates unnatural base pairs by inducing a Watson-Crick geometry. *Nat. Chem. Biol.*, **8**, 612–614.
38. McCoy,A.J., Grosse-Kunstleve,R.W., Adams,P.D., Winn,M.D., Storoni,L.C. and Read,R.J. (2007) Phaser crystallographic software. *J. Appl. Crystallogr.*, **40**, 658–674.
39. Murshudov,G.N., Vagin,A.A. and Dodson,E.J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D. Biol. Crystallogr.*, **53**, 240–255.
40. Emsley,P, Lohkamp,B., Scott,W.G. and Cowtan,K. (2010) Features and development of coot. *Acta Crystallogr. D. Biol. Crystallogr.*, **486–501**, 486–501.
41. Adams,P.D., Afonine,P.V., Bunkoczi,G., Chen,V.B., Davis,I.W., Echols,N., Headd,J.J., Hung,L.W., Kapral,G.J., Grosse-Kunstleve,R.W. *et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D. Biol. Crystallogr.*, **66**, 213–221.
42. Afonine,P.V., Grosse-Kunstleve,R.W., Echols,N., Headd,J.J., Moriarty,N.W., Mustyakimov,M., Terwilliger,T.C., Urzhumtsev,A., Zwart,P.H. and Adams,P.D. (2012) Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D. Biol. Crystallogr.*, **68**, 352–367.
43. Moriarty,N.W., Grosse-Kunstleve,R.W. and Adams,P.D. (2009) Electronic ligand builder and optimization workbench (eLBOW): a tool for ligand coordinate and restraint generation. *Acta Crystallogr. D. Biol. Crystallogr.*, **65**, 1074–1080.
44. Painter,J. and Merritt,E.A. (2006) Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta Crystallogr. D. Biol. Crystallogr.*, **62**, 439–450.
45. Minor,W., Cymborowski,M., Otwinowski,Z. and Chruszcz,M. (2006) HKL-3000: the integration of data reduction and structure solution–from diffraction images to an initial model in minutes. *Acta Crystallogr. D. Biol. Crystallogr.*, **62**, 859–866.
46. Vagin,A. and Teplyakov,A. (2010) Molecular replacement with MOLREP. *Acta Crystallogr. D. Biol. Crystallogr.*, **66**, 22–25.
47. Lu,X.J. and Olson,W.K. (2003) 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res.*, **31**, 5108–5121.
48. Collaborative Computational Project, Number 4. (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D. Biol. Crystallogr.*, **50**, 760–763.
49. Hayward,S. and Lee,R.A. (2002) Improvements in the analysis of domain motions in proteins from conformational change: DynDom version 1.50. *J. Mol. Graph. Model.*, **21**, 181–183.
50. Kaina,B., Christmann,M., Naumann,S. and Roos,W.P. (2007) MGMT: key node in the battle against genotoxicity, carcinogenicity and apoptosis induced by alkylating agents. *DNA Repair (Amst.)*, **6**, 1079–1099.
51. Taylor,D., Cawley,G. and Hayward,S. (2014) Quantitative method for the assignment of hinge and shear mechanism in protein domain movements. *Bioinformatics*, **30**, 3189–3196.
52. Georgiadis,M.M., Jessen,S.M., Ogata,C.M., Telesnitsky,A., Goff,S.P. and Hendrickson,W.A. (1995) Mechanistic implications from the structure of a catalytic fragment of Moloney murine leukemia virus reverse transcriptase. *Structure*, **3**, 879–892.
53. Gao,G., Orlova,M., Georgiadis,M.M., Hendrickson,W.A. and Goff,S.P. (1997) Conferring RNA polymerase activity to a DNA polymerase: a single residue in reverse transcriptase controls substrate selection. *Proc. Natl. Acad. Sci. U.S.A.*, **94**, 407–411.
54. Li,Y., Mitaxov,V. and Waksman,G. (1999) Structure-based design of Taq DNA polymerases with improved properties of dideoxynucleotide incorporation. *Proc. Natl. Acad. Sci. U.S.A.*, **96**, 9491–9496.
55. 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. *Angew. Chem. Int. Ed. Engl.*, **56**, 12000–12003.
56. Betz,K., Malyshev,D.A., Lavergne,T., Welte,W., Diederichs,K., Romesberg,F.E. and Marx,A. (2013) Structural insights into DNA replication without hydrogen bonds. *J. Am. Chem. Soc.*, **135**, 18637–18643.