Title
Analysis of Two Widespread Versions of a Bacterial Replicative DNA Polymerase

Author
Guenther, Joel Michael

Publication Date
2010

Peer reviewed|Thesis/dissertation

Analysis of Two Widespread Versions of a Bacterial Replicative DNA Polymerase

by

Joel Michael Guenther

A dissertation submitted in partial satisfaction of the requirements for the degree of

Doctor of Philosophy

in

Chemistry

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor John Kuriyan, Chair

Professor Susan Marqusee

Professor David Wemmer

Fall 2010

Abstract

Analysis of Two Widespread Versions of a Bacterial Replicative DNA Polymerase

by

Joel Michael Guenther

Doctor of Philosophy in Chemistry

University of California, Berkeley

Professor John Kuriyan, Chair

In all known cellular organisms, from the smallest bacteria to the tallest of trees, DNA stores the instructions for maintenance, growth and — crucially — reproduction that have allowed life to proliferate to the far reaches of our planet. When a cell divides to form two equivalent cells, it must first copy its genome using an enzyme known as a replicative DNA polymerase. This protein functions as part of a replisome, which is a group of specialized proteins that work in concert to duplicate, faithfully, DNA molecules numbering several million bases in length. As a graduate student, I studied the atomic-level structure, biochemical properties and evolution of the replicative polymerase, and some of its helper proteins, in the replisome of the Gram-negative γ-proteobacterium *Escherichia coli.* Interestingly, although some replisome proteins are universally conserved, the replicative polymerases themselves are not, and those from Bacteria appear, on the basis of sequence and structure, unrelated to those from the two other domains of life, Eukaryota and Archaea. This lack of homology is evidence of the convergent evolution of two different types of polymerases and suggests that bacterial DNA replication may be a promising target for antibiotic development. Within bacteria, there are two main types of replicative polymerases — the PolC homologs and the DnaE1-pol homologs. PolC homologs appear to be limited to the low-GC Gram-positive bacteria that comprise the phyla *Firmicutes* and *Tenericutes,* while nearly all others use a DnaE1-pol.

In this thesis, I align a diverse set of DnaE1-homologs and identify among them two main versions, one like the polymerase from *E. coli* (Eco-like) and the other like that of *Thermus aquaticus* (Taq-like). Using a second alignment, I observe that the Eco-like polymerase almost always appears with an editing protein, DnaQ-exonuclease, which removes mismatch errors during DNA synthesis. Guided by these sequence data, I analyze previously solved structures of the *E. coli* and *T. aquaticus* polymerases and place the sequence and structure interpretations in the context of known bacterial phylogeny. Taq-like DnaE1-pol, which is a constitutively active polymerase that contains DNA-editing exonuclease activity in its PHP domain, is the ancestral version and, based on its presence in cyanobacteria, likely dates back roughly 3,500 million years, nearly to the origin of life on Earth. In contrast, Eco-like DnaE1-pol can adopt a

distorted conformation incompatible with DNA synthesis and has a PHP domain that lacks editing activity and, instead, binds a DnaQ-exonuclease for editing *in trans.* Because mitochondria are descended from an α-proteobacterium, the Eco-like polymerase predates the emergence of eukaryotes more than 1,500 million years ago.

In addition to identifying the two main types of DnaE1-pol, I describe significant progress toward crystal structures the *E. coli* replicative polymerase in complex with DNA. Furthermore, I demonstrate the precise structural conservation of the PHP domain in this polymerase by restoring metal binding to it using only three point-mutations.

for my family and friends, near and far, supporting

TABLE OF CONTENTS

LIST OF SYMBOLS & ABBREVIATIONS

3mPHP, 4mPHP, 5mPHP — EcoE1(917) with 3, 4 and 5 PHP metal binding mutations, respectively

βME — β-mercaptoethanol

βNT — Polβ-like nucleotidyltransferase (superfamily of enzymes)

Cm — denaturant concentration at which the folded and unfolded states of a protein are equally populated at equilibrium

DNA — deoxyribonucleic acid

dsDNA — double-stranded DNA

DTT — dithiothreitol

Eco — *Escherichia coli*

EcoE1(917) — *E. coli* DnaE1-pol residues 1 through 917 of 1160

Eco-Asp401 — *Escherichia coli* DnaE1-pol residue Asp401 (other residues are named using the same convention)

Gdn•HCl — guanidine hydrochloride

Gka — *Geobacillus kaustophilus*

Gka-Ile872 — *Geobacillus kaustophilus* PolC Ile872 (other residues are named using the same convention)

HEPES — 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid

His6 — hexahistidine

IPTG — isopropyl β-D-1-thiogalactopyranoside

nt — nucleotide(s)

PCR — polymerase chain reaction

PEG — polyethylene glycol

PHP — Polymerase and Histidinol-Phosphatase (protein domain)

Pol II — DNA polymerase II

pol III — DNA polymerase III holoenzyme

Polβ — DNA polymerase β

RNA — ribonucleic acid

ssDNA — single-stranded DNA

Taq — *Thermus aquaticus*

Taq-Ile464 — *Thermus aquaticus* DnaE1-pol Cys145 (other residues are named using the same convention)

Tm — melting temperature, the temperature at which the folded and unfolded states of a protein are equally populated at equilibrium

Tris — tris(hydroxymethyl)aminomethane

ACKNOWLEDGMENTS

My deep gratitude goes to John Kuriyan, whose support and encouragement made these studies possible. It has been a gift to work with those who strive for understanding and find beauty in the world.

Complementing John's strengths are those of the individuals he has chosen as collaborators and coworkers, and, to these people, I am similarly indebted. On his visits, Mike O'Donnell brought a passion for biochemistry and warm encouragement. Meindert Lamers helped guide me, with gusto, through a world of experimental details that I could not have imagined. Together with Brian Kelch, he also helped me grasp key narrative threads in our work. Tiago Barros brought a contagious mirth to our studies, and, along with Nick Levinson, Jon Winger and Jeff Iwig, he enthusiastically helped edit this document. Debora Makino shared an insight that set new standards for protein yield, and Jordan Anaya pushed me, hard.

The product of a group environment, this thesis is built from the work of many. Meindert Lamers and Brian Kelch performed significant portions of the structural and bioinformatic analyses in CHAPTER 3. Jordan Anaya completed all the steps leading to the structure solved by Meindert Lamers in CHAPTER 2 and also performed several assays and protein preparations. Xiaoxian Cao, Caleb Cassidy-Amstutz, Mary Coons, Lore Leighton and Tiffany Chou all cloned constructs. Corie Ralston and other ALS staff helped with X-ray experiments.

Brian Kelch introduced me to DNA-binding and polymerization assays, which Ken Dong and Richard Rymer helped me establish in high-throughput format. Jacob Corn provided guidance for crosslinking, and Arnold Falick, at the UC Berkeley HHMI Mass Spectrometry Facility, analyzed samples and advised on their preparation. Also at the facility, David King generated data on countless proteins, synthesized peptides, and provided sage advice. Markus Seeliger and Katherine Tripp helped analyze protein-folding data while Kath Ratcliff advised on sample preparation. Jeff Iwig provided critical analysis of metal binding.

In additional ways, scientific and non-, Marsha Henderson, Nick Levinson, Jon Winger, Tanya Freedman, Luke Chao and Sonja Lorenz well embodied the spirit of the Kuriyan laboratory — engaged and inviting, contemplative but quick to jest, and a tolerant, supportive community. Special thanks also go to Steven Jacques for setting a respectful tone while deftly managing a complex and fast-moving research environment.

CHAPTER 1

Replicative DNA Polymerase Diversity Across Bacterial Species

1.1 Summary

DNA stores the information of life and is transmitted from one generation to the next through the action of replicative DNA polymerases. Despite their diversity, these enzymes are united by (1) their dependence on the helper proteins with which they form a replisome and (2) a two-metal mechanism for DNA synthesis. Unlike eukaryotes or Archaea, bacteria use a C-family polymerase as their replicative enzyme. Within the bacterial replisome, a DNA polymerase III holoenzyme (pol III) contains both the replicative polymerase and an editing exonuclease for error removal. There are two main types of pol III, which respectively feature either a PolC or DnaE1-pol homolog as the replicative polymerase. The latter is more common and appears in many Gram-positive and nearly all Gram-negative bacteria. Combining sequence analysis with biochemical data and structure interpretation, this thesis describes the two main versions of DnaE1-pol (CHAPTER 3). The ancestral (Taq-like) version, represented by the *Thermus aquaticus* polymerase, can edit DNA using a metal-binding PHP domain, while the more recently evolved (Eco-like) version, as exemplified by *Escherichia coli* DnaE1-pol, can adopt an inactive conformation and has a non-enzymatic PHP domain that binds a separate DnaQ-exonuclease for editing DNA *in trans.*

1.2 Introduction to DNA polymerases and replisomes

Within every cellular organism, there are several different types of DNA polymerases [1, 2]. Most are specialists that have evolved to overcome specific types of DNA damage [3-6], and only one or two serve as replicative enzymes [7]. Regardless of their roles, all DNA-dependent (DNA-directed) DNA polymerases require primer-template DNA as a substrate and contain a Palm, Thumb and Fingers domain that, together, form a shape like a cupped right hand (Figure 1-1) [8]. During DNA synthesis [9], an unpaired base in the template DNA strand and the 3'-end of the primer strand bind in the active site of a polymerase. Upon further binding of a nucleotide complementary to the template base, the Fingers and Thumb come together and the polymerase adopts a closed conformation in which the incoming nucleotide is added to the 3'-end of the primer via a two-metal catalytic mechanism [8]. In addition to synthesizing DNA, some high fidelity polymerases, including most replicative enzymes (this work and [10]), contain an editing 3'-5' exonuclease domain that can remove mismatches from the primer strand by cleaving nucleotides from its 3'-end. In *E. coli,* the editing domain increases the fidelity of the replicative polymerase 25-fold, lowering the average synthesis error rate to roughly one incorrect nucleotide per 50 million DNA bases [11].

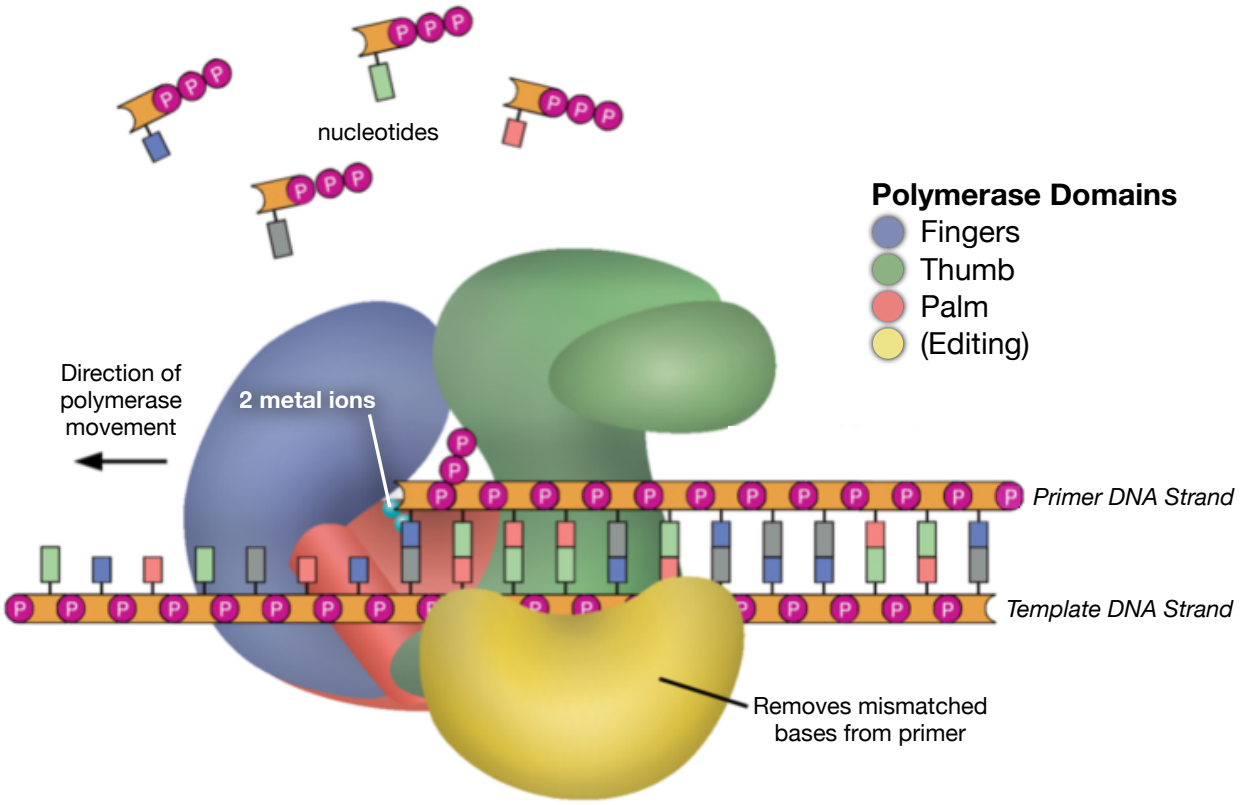


Figure 1-1. DNA-dependent polymerases add nucleotides to a primer DNA strand paired with a template.

Schematic representation of a DNA-dependent DNA polymerase synthesizing DNA. The polymerase domains are represented as globular shapes, the primer and template DNA strands are depicted as linear, and nucleotides (deoxynucleotide triphosphates, dNTPs) are floating above. All polymerases have a shape akin to that of a cupped right hand, but an editing domain is most frequently found in replicative polymerases, either as a domain within the polymerase or as a separate protein that binds to it. Figure adapted from Lore Leighton.

In the iconic DNA double helix, two complementary DNA strands are oriented in opposite directions [12]. This antiparallel arrangement presents a problem for replication, because polymerases can only synthesize DNA in one direction (from 5' to 3') [9], and both strands must be completely duplicated before cell division can occur. In the *E. coli* replisome, helper proteins solve this problem by linking together three copies of its DnaE1-pol (Figure 1-2) [13]. One moves in the same direction as the replication fork while continuously synthesizing a complementary primer strand for the "leading" DNA template. Since the other, "lagging" template, is oriented in the opposite direction, the two other polymerases synthesize its complement discontinuously in segments, generally 1,000 to 2,000 bases in length [14], called Okazaki fragments [15, 16].

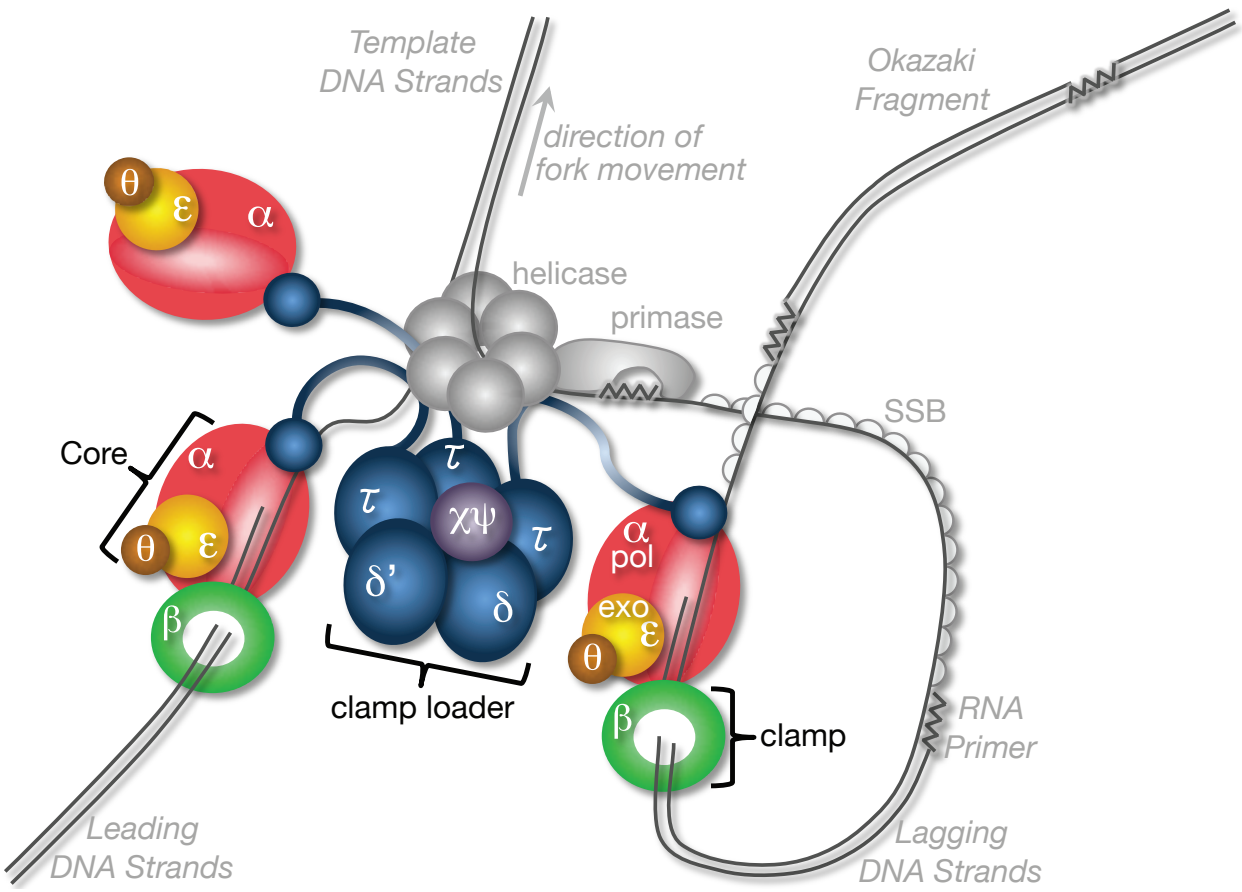


Figure 1-2. Replicative DNA polymerases function within a protein ensemble known as a replisome.

Schematic representation of the *E. coli* replisome synthesizing DNA at a replication fork. Subunits of the DNA polymerase III holoenzyme [14] are colored, while the other proteins of the replisome are in grayscale. The α subunits (replicative polymerase DnaE1-pol) synthesize DNA, while editing ε subunits (DnaQ-exonuclease, MutD) remove synthesis errors and two DNA sliding clamps (dimers of β, DnaN) tether the Core subassemblies to their DNA substrates. Besides arranging clamps on primed sites, the clamp loader also provides order to the replisome by binding the DnaE1-pol subunits and the helicase (DnaB). SSB stands for single-stranded DNA binding protein, and interactions been the clamp loader χ subunit and SSB have been omitted for clarity. Figure adapted from Meindert Lamers.

Although the physical linkage of polymerases in the *E. coli* replisome ensures that both DNA strands are simultaneously replicated, this creates a secondary problem, because linked polymerases cannot rotate to accommodate the DNA helix. Instead, each polymerase unwinds its template strand, which accumulates torsional strain. Eventually, this forces the polymerase to release the DNA so that it can rotate back into its preferred helical geometry. To permit this relaxation without halting processive (continuous) synthesis, each polymerase binds to a DNA sliding clamp [17, 18] that encircles the newly formed duplex (Figure 1-2).

Along with its polymerase and clamp subunits, the other proteins of the *E. coli* replisome (Figure 1-2) are also functionally conserved by all cellular life [7, 10, 14, 19, 20]. These include (1) an editing exonuclease for removal of mismatches that attaches to each polymerase, a (2) helicase for unwinding DNA into individual strands, which are, in turn, protected by a coating of (3) single-stranded DNA binding proteins, a (4) primase for RNA primer synthesis, and a (5) clamp loader that places clamps on primed sites and whose binding interactions organize the replisome. Interestingly, while the functions listed above are universally conserved, only the clamp loader and clamp have homologs across the three domains of life [7]. In contrast, comparison of the replicative polymerases from Bacteria (C-family) to those in Eukaryota and Archaea (B- or D-family) leads to the surprising conclusion that these two groups of enzymes may be an example of convergence from different evolutionary origins [21, 22].

1.3 Two non-homologous superfamilies share a two-metal mechanism for polymerization

At the most general level, DNA-dependent DNA polymerases can be grouped into two superfamilies between which there is no apparent sequence homology [1, 2]. Together, the C- and X-family polymerases are members of the Polβ-like nucleotidyltransferase (βNT) superfamily, which also contains many non-polymerase enzymes [23]. The other superfamily includes the "classical" polymerases [8], so named because they were the first to be studied structurally. Beyond homologs of the replicative B-family polymerases from eukaryotes and the replicative B- or D-family polymerases in archaeans, the classics include members of the A- and Y-families along with RNA polymerases, which synthesize RNA based on a DNA template (transcription), and reverse transcriptases, which make DNA using an RNA template.

Despite no homology between the two superfamilies, superposition of polymerase structures based on incoming nucleotide position suggests a universally shared two-metal catalytic mechanism (Figure 1-3c) [8, 24, 25]. In the vicinity of the nucleotide, the primer DNA strands superimpose, as do the templates, two divalent metal ions (usually magnesium) and a pair of aspartate residues whose carboxylate head groups bridge both metals. One of these metals binds to all three members of the nucleotide triphosphate tail (the triphosphate-binding metal). The other directly facilitates catalysis (the catalytic metal) by aligning the 3'-hydroxyl of the DNA primer for attack on the α-phosphate of the nucleotide.

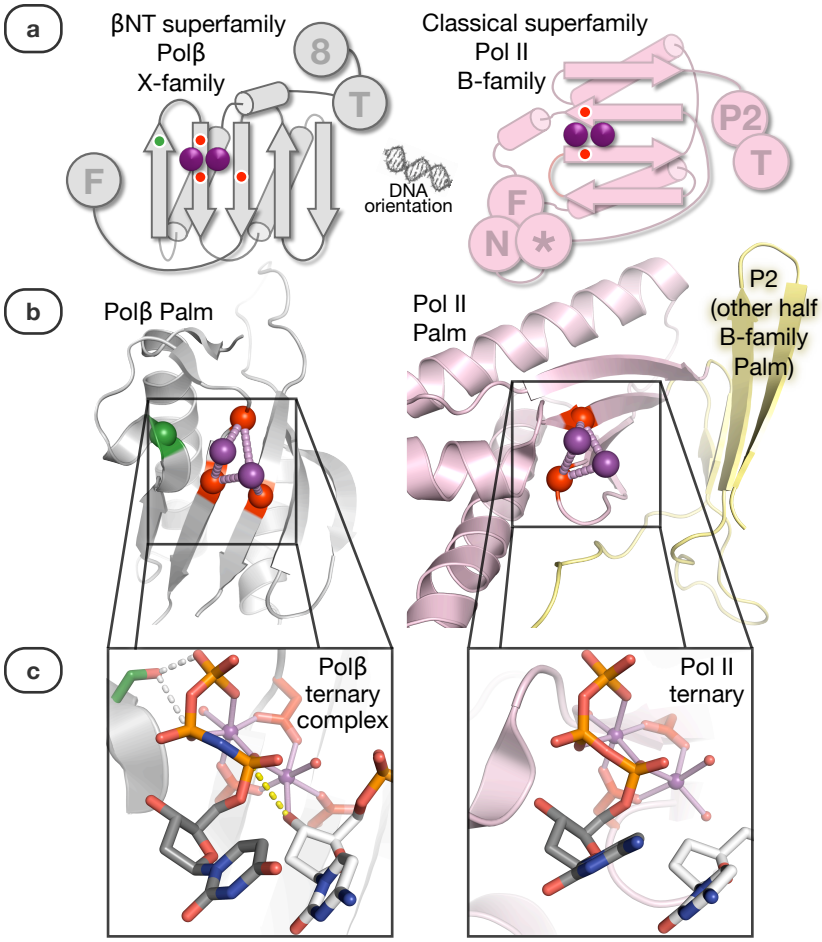


Figure 1-3. The two non-homologous superfamilies of polymerases share a two-metal catalytic mechanism.

(***a***) Schematic representations of the Palm domain of human Polβ and *E. coli* Pol II, respectively, the namesake of the Polβ-like nucleotidyltransferase (βNT) superfamily and a member of the classical polymerase superfamily. Relative to the positions of the two metals, the β-stands of Pol II are roughly perpendicular to those of Polβ. Using the Lamers *et al.* [26] notation for Polβ, F: Fingers, T: Thumb, 8: 8 kDa domain, N: N-terminal domain, ***:** exonuclease domain, P2: 2nd half of Palm unique to B-family. Circles or spheres: ●serine, ●aspartate, ●metal ion. The two aspartate residues indicated for Pol II are the only two specific structural features present in all DNA-dependent DNA polymerases. (***b***) Corresponding structure excerpts (PDB 2fms, 3k59). (***c***) For the Polβ ternary complex, a yellow dashed line (▮) indicates the path of attack on the α-phosphate of the incoming nucleotide (nonhydrolyzable analog dUpnpp) by the 3'-hydroxyl of the DNA primer strand. Bond formation cannot occur in the Pol II ternary complex because the primer lacks a 3'-hydroxyl.

Though the members of the two polymerase-superfamilies all have a *general* cupped-hand shape, their only *specific* universal structures are the two metal-bridging aspartates. This lack of homology is underscored by the nearly perpendicular orientation of the β-strands from the classical Palm relative to that of the C- and X-families (Figure 1-3) [8]. In the classical polymerases, the two aspartates are on adjacent β-strands, and, as a result, these strands are orthogonal to the plane formed

by the base of the incoming nucleotide. In the C- and X-family polymerases, however, the bridging aspartates occupy the same β-strand, which is parallel to the incoming base. In addition, the C- and X-families strictly conserve a glycine-serine pair (the GS-motif) and a third aspartate that are absent from the classical superfamily [8]. The third aspartate, like the two that are universally conserved, directly binds the catalytic metal ion, and the serine contacts the β- and γ-phosphates of the incoming nucleotide.

Given the lack of structural homology between the two superfamilies of DNA polymerases, their mutual use of a two-metal catalytic mechanism is likely a dramatic example of convergent evolution [1, 21, 27]. Furthermore, absence of structural homology between the eukaryotic B-family and bacterial C-family replicative polymerases suggests that DNA replication may pose a valuable target for antibiotics [28-32].

1.4 The C-family DNA polymerases

1.4.1 Universal and distinguishing features of the two main classes

Found only in bacteria [1, 2], C-family polymerases conserve a set of four principle domains (Figure 1-4) that are required for DNA synthesis [33]. These include, in sequential order, a PHP (from Polymerase and Histidinol-Phosphatase [34]), Palm, Thumb and Fingers. With a lone caveat for the dsDNA-binding HhH (helix-hairpin-helix) motif near the end of the Fingers, the last two of those domains share no sequence homology with any other known proteins. Each X-family polymerase contains two HhH motifs [5], but the regions flanking them are not homologous to any portion of the C-family polymerases. Among the other principle domains, the Palm does show clear structural homology [24-26] between the C- and X-families, but only six residues [24] (including those mentioned in the previous section) are conserved between them. In contrast, the N-terminal PHP of the C-family polymerases (Figure 1-4, Figure 1-5b) exhibits a significant structural difference that distinguishes it from canonical PHP domains, such as the one tethered to the C-terminus of some X-family members. While the central β-strands in the canonical domains are all parallel, one strand in the C-family PHP is anti-parallel, an arrangement that has only been shown for one other PHP protein (*Thermotoga maritima* tm0559, PDB 2anu). Given the significantly different structural location and threading of the C-family PHP [24-26, 35] relative to that of the X-family [36], the distant relatedness of these two families [23, 24], and the absence of PHP domains from the other members of the Polβ-like nucleotidyltransferase superfamily [23, 37], the presence of PHP domains in the C- and X-family polymerases is likely an example of convergent evolution.

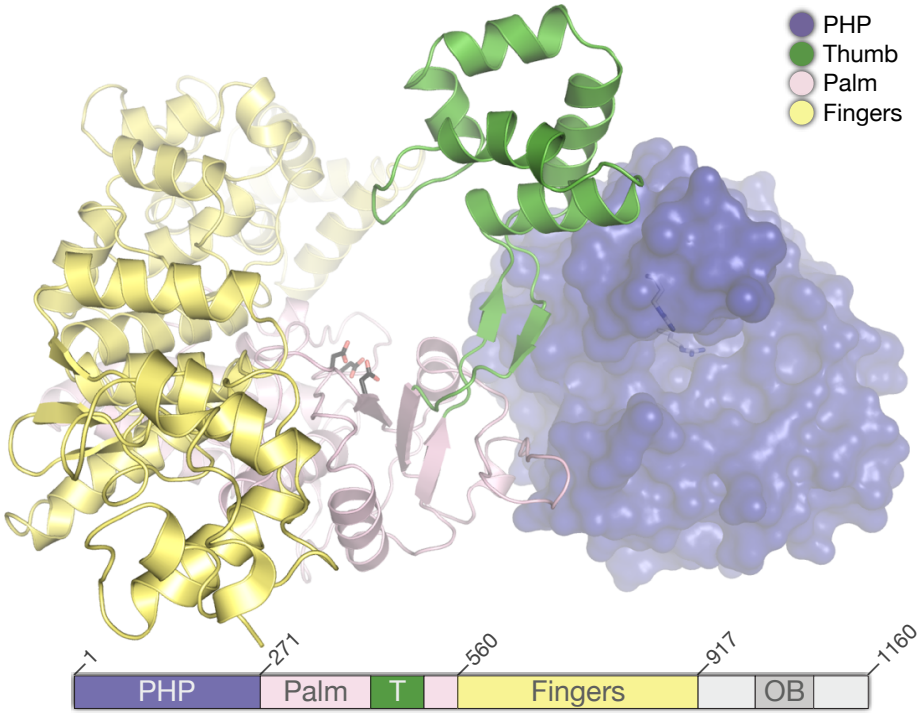


Figure 1-4. C-family polymerases conserve four principle domains.

EcoE1(917): residues 1 though 917 of 1160 (PDB: 2hnh) [26]. The PHP domain (blue) is a large, roughly spherical domain that makes extensive contacts with the Palm (pink) and structurally supports the Thumb (green). The three metal-binding, catalytic aspartates in the Palm (Figure 1-6) are shown as black sticks, and the two structural arginines that substitute for PHP metals (Figure 1-7, Table 1-1) are white. Although the guanidino head groups of the arginines are solvent exposed, they are located in a deep cleft, which places them near the center of the PHP. This cleft starts as a narrow groove where the Thumb extends upward from the PHP-side of the Palm. In *T. aquaticus* DnaE1-pol, the PHP has retained DNA editing activity, and the narrow groove may guide the DNA primer strand toward metal ions in the PHP cleft (Figure 1-7) for mismatch removal [26].

The C-family is comprised of PolC and DnaE-pol homologs, which are distinguished by their respective peripheral domains (Figure 1-5b). In the DnaE-pol homologs, the domains are C-terminal to the Fingers and include, in reverse sequence order, a peptide motif that binds to the clamp loader [38], a poorly conserved globular domain that presents the binding peptide, and an OB (oligonucleotide-binding) domain that sits up on the Fingers and guides the DNA template strand toward the Palm [24]. In contrast, there are no domains after the Fingers of the PolC homologs, which, instead, begin with a poorly conserved domain of unknown structure and function followed by an OB domain that forms the tip of the Thumb [25]. Inserted within the standard PolC PHP domain is a DnaQ-exonuclease that edits the DNA primer strand by removing mismatches in the 3' to 5' direction [25]. Interestingly, a separate DnaQ-exonuclease subunit binds to the PHP and performs editing *in trans* for some DnaE-pol homologs (see next section).

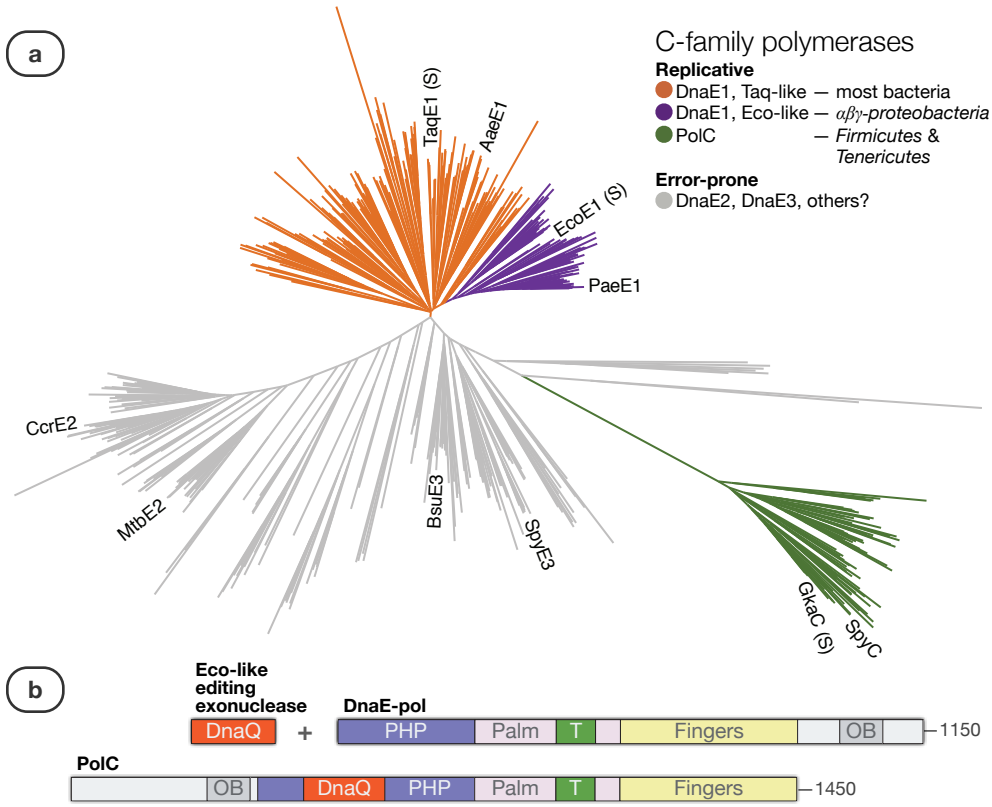


Figure 1-5. Diversity of sequence and function are correlated in the C-family polymerases.

(***a***) A phylogenetic tree of 585 sequences with less than 70% identity displays the known sequence diversity among C-family polymerases (excluding cyanobacterial split-inteins [39]). Approximate locations of biochemically or structurally (S) characterized polymerases are indicated (TaqE1: *T. aquaticus* DnaE1-pol, AaeE1: *Aquifex aeolicus* DnaE1-pol, EcoE1: *E. coli* DnaE1-pol, PaeE1: *Pseudomonas aeruginosa* DnaE1-pol, CcrE2: *Caulobacter crescentus* DnaE2-pol, MtbE2: *Mycobacterium tuberculosis* DnaE2-pol, BsuE3: *Bacillus subtilis* DnaE3-pol, SpyE3: *Streptococcus pyogenes* DnaE3-pol, GkaC: *Geobacillus kaustophilus* PolC, SpyC: *S. pyogenes* PolC). Taq- and Eco-like DnaE1-pol are identified and described in CHAPTER 3. Tree coloring is based on apparent relatedness to characterized polymerases. (***b***) Domains are colored to indicate homology (gray: nonhomologous, T: Thumb, OB: oligonucleotide-binding, DnaQ: exonuclease homologous to the protein that binds Eco-like DnaE1-pol).

PolC homologs are present in the low-GC Gram-positive bacteria that comprise the phyla *Firmicutes* and *Tenericutes* [39]. Although these species also conserve a DnaE-pol [40, 41], recent analysis of the *Bacillus subtilis* replisome [42] suggests that PolC is likely the sole replicative polymerase in the organisms that contain it. The *B. subtilis* PolC synthesizes the complements to both the leading and lagging template DNA strands, but it cannot initiate replication from sites primed with RNA. Instead, the slow, error-prone DnaE-pol [43] enables synthesis by the PolC by adding a DNA primer to the end of the RNA primer generated by the primase. In agreement with the data for *B. subtilis,* other bacteria (*Streptococcus pyogenes* [44] and *Staphylococcus aureus* [45]) also use their PolC homolog for leading strand synthesis, and the relatively high

sequence conservation among known PolC homologs suggests that most, if not all, are replicative polymerases.

With respect to synthesis rate and fidelity, DnaE-pol homologs [14, 43, 46, 47] appear far more variable than the PolC [42, 44, 45]. Consistent with this observation, Zhao *et al.* [40, 41] recently used sequence alignments to identify three subclasses of DnaE-pol, which they named DnaE1, DnaE2 and DnaE3, respectively. Using this nomenclature, the DnaE-pol homologs that have been characterized biochemically as error-prone belong to either the DnaE2- or DnaE3-pol subclasses. Among these, the *Mycobacterium tuberculosis* DnaE2-pol contributes to the emergence of drug resistance [47], and DnaE2-pol homologs may, in general, be translesion polymerases whose expression is induced in response to DNA damage [47-49]. The DnaE3-pol homologs, such as the one in *B. subtilis* (mentioned above), are almost always the accompanying DnaE-pol found in PolC-containing species [40, 41]. In addition to synthesizing DNA primers [42] (a role analogous to that of the B-family Pol1 subunit of Eukaryotic Pol α [10]), the low fidelity of DnaE3-pol homologs allows them to bypass DNA lesions [43, 46].

The most well studied C-family member is the high-speed, high-fidelity replicative polymerase from *E. coli.* As part of its replisome (Figure 1-2), this DnaE1-pol makes roughly one error per 50 million bases [11] while synthesizing DNA at nearly 1000 nucleotides per second [14]. Similarly, the DnaE1-based replisomes of *Pseudomonas aeruginosa* [50], *Aquifex aeolicus* [51] and *Thermus thermophilus* [52] have also been shown to synthesize DNA rapidly. Since none of these species contain a PolC homolog, the replisomes of most non-PolC bacteria, including essentially all Gram-negative species, are probably based on a DnaE1-pol (minor subtypes of replicative C-family polymerases are mentioned in CHAPTER 3).

The classification of bacteria into those with PolC- or DnaE1-based replisomes is also consistent with data on the relative susceptibility of different bacterial species to antibiotics. In 1964, a small molecule was discovered that selectively killed Gram-positive bacteria [53-55], and, in 1973, Cozzarelli and coworkers [56] used a related molecule to identify the replicative polymerase (a PolC) in the Gram-positive bacterium *Bacillus subtilis.* In contrast, the replicative polymerase (a DnaE1-pol) from *E. coli,* a Gram-negative bacterium, was largely unaffected by these molecules.

Despite the discovery of PolC inhibitors nearly a half-century ago, there are still no antibiotic drugs that target bacterial DNA polymerases, and there are no published accounts of comparable inhibitors of the DnaE1-pol homologs. The prospects for antibiotics targeting replication, however, appear good again, as recent years have witnessed the discovery of new PolC inhibitors [28-31] and the solving of high-resolution crystal structures of a PolC [25] and two DnaE1-pol homologs [24, 26].

1.4.2 Structural differences within the DnaE class govern polymerization and editing

Although Kornberg and colleagues discovered *E. coli* DNA polymerase III in 1971 [57, 58], nearly three and a half decades passed before the first structure of the synthesis subunit from this enzyme was solved [26]. Over the intervening years, this DnaE1-pol and its replisome became the model system for replicative DNA synthesis [14], yet the structure still managed to surprise.

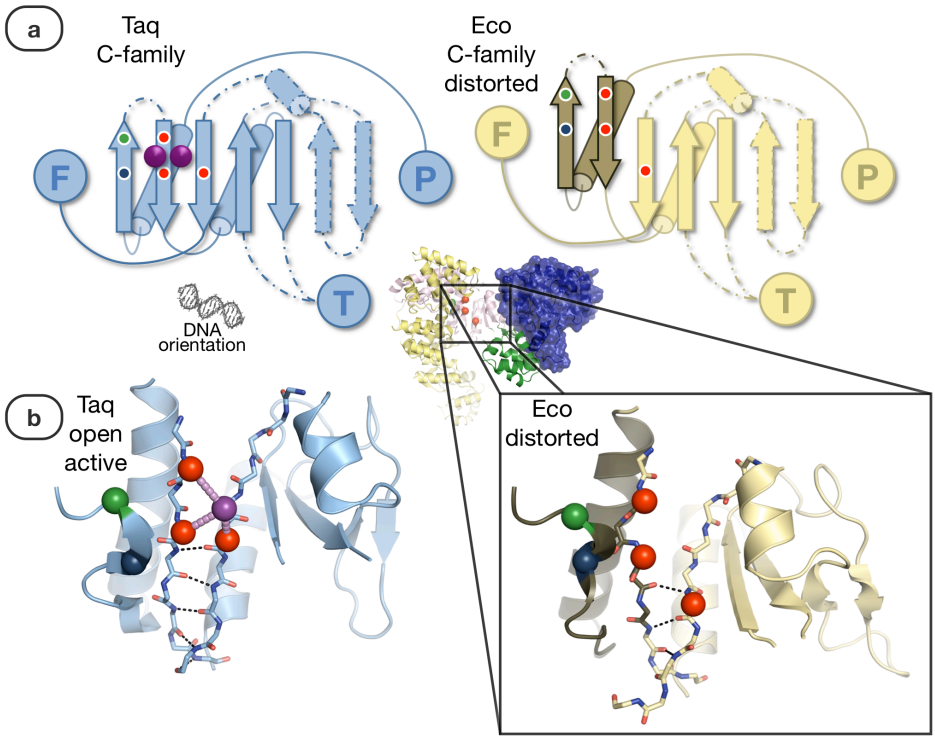


Figure 1-6. *E. coli* DnaE1-pol crystallized with its Palm in a distorted conformation.

(***a***) Schematic representation of the active C-family polymerase Palm geometry (Taq: *T. aquaticus* DnaE1-pol) and the inactive *E. coli* DnaE1-pol geometry (Eco). Other domains include F: Fingers, P: PHP, and T: Thumb. Circles or spheres: ●serine, ●aspartate, ●arginine, ●metal ion. The first two (Fingers-side) β-strands of the Eco Palm are shifted upward, disrupts the catalytic metal binding site between the aspartate-bearing β-strands. (***b***) Corresponding excerpted structures (PDB 2hpi [24], 2hnh [26]).

As predicted [59], *E. coli* DnaE1-pol was clearly [26] homologous to the X-family polymerases, but its three catalytic aspartate residues were not properly positioned (Figure 1-6) for DNA synthesis (Figure 1-3). An apparent rotational-sliding motion [26] between the Fingers- and PHP-sides of the Palm had moved the third aspartate out of position, thus abrogating binding of the catalytic metal ion. Underscoring the significance of this distortion, a simultaneously published structure of the *T. aquaticus*

DnaE1-pol [24] featured a Palm in which the three conserved aspartates both adopted the expected geometry and, together, bound a magnesium ion (Figure 1-6).

PHP metal-binding residues

	PDB	Topo	1	2	3	4	5	6	7	8	9	Me1	Me2	Me3
Consensus			**H**	**H**	**D/H**	**H**	**E**	**H**	**C/H**	**D/N**	**H**	**Zn**		
Eco DnaE1	2hnh	1-anti	R_{10}	H_{12}	D_{19}	F_{44}	D_{69}	H_{83}	G_{134}	D_{201}	R_{203}	R_{203}	R_{10}	-
3mPHP	••	"	**H_{10}**	"	"	**H_{44}**	"	"	"	"	**H_{203}**	**Zn**	**Zn**	**-**
4mPHP	**n.a.**	"	**H_{10}**	"	"	**H_{44}**	"	"	**C_{134}**	"	**H_{203}**	**n.a.**	**n.a.**	**n.a.**
5mPHP	**n.a.**	"	**H_{10}**	"	"	**H_{44}**	**E_{69}**	"	**C_{134}**	"	**H_{203}**	**n.a.**	**n.a.**	**n.a.**
Taq DnaE1	2hpi	1-anti	H_{11}	H_{13}	D_{20}	H_{47}	E_{72}	H_{95}	C_{145}	D_{212}	H_{214}	Zn	Zn	-
Gka PolC	3f2d	1-anti	H_{346}	H_{348}	D_{355}	H_{380}	E_{405}	H_{620}	C_{670}	N_{743}	H_{745}	Zn	Mn	Mn
Dra PolX	2w9m	all-par	H_{332}	H_{334}	D_{339}	H_{364}	E_{401}	H_{428}	H_{456}	D_{516}	H_{518}	Zn	Zn	Zn
Eco YcdX	1m68	all-par	H_{7}	H_{9}	H_{15}	H_{40}	E_{73}	H_{101}	H_{131}	D_{192}	H_{194}	Zn	Zn	Zn
Tth ppPHP	2yxo	all-par	H_{5}	H_{7}	H_{13}	H_{38}	E_{80}	H_{108}	H_{154}	D_{224}	H_{226}	Zn	Fe	Fe
Tma tm0559	2anu	1-anti	H_{13}	H_{15}	D_{20}	H_{45}	E_{97}	H_{106}	H_{140}	D_{193}	H_{195}	Zn	Zn	Zn

Table 1-1. *E. coli* DnaE1-pol conserves only four of nine PHP metal-binding residues.

Structurally (••CHAPTER 2) and biochemical (CHAPTER 3) analyzed PHP metal binding mutants generated of EcoE1(917) (Figure 1-4) are indicated in bold. *PDB* refers to the PDB code for the given structures. *Topo* indicates the β-sheet topology of the PHP domains, which are generally assumed to have an all-parallel β-strand arrangement (all-par) with the exception of the C-family polymerases and tm0559 that, instead, have one antiparallel β-strand (1-anti). Numbers *1* through *9* are the nine PHP metal-binding residue positions. *Me1, Me2* and *Me3* are the three canonical PHP metal binding sites. Structural data for 4mPHP and 5mPHP are not available (n.a.). Eco: *E. coli,* Taq: *T. aquaticus,* Gka: *G. kaustophilus,* Dra: *Deinococcus radiodurans,* Tth: *T. thermophilus,* Tma: *Thermotoga maritima.*

In addition to the unexpected Palm distortion, *E. coli* DnaE1-pol also had a significantly different PHP domain than that of the *T. aquaticus* polymerase. While the latter contained all nine canonical metal-binding residues (Table 1-1) and crystallized with two zinc ions bound in its PHP cleft (Figure 1-7), *E. coli* DnaE1-pol, as predicted [34], did not conserve several residues of the PHP motif and appeared incapable of metal binding. In place of the metals observed in the *T. aquaticus* DnaE1-pol structures, the head groups from two arginines are hydrogen bonded to the metal-binding residues that were still conserved by the PHP of *E. coli* EcoE1-pol (Figure 1-7).

Later crystallization efforts [35] revealed that, like its close homolog in *T. thermophilus* [60], the PHP domain of *T. aquaticus* DnaE1-pol is an active nuclease. Since *E. coli* DnaE1-pol was known to bind a DnaQ-exonuclease [14, 61, 62], some speculation occurred regarding the possibility that *Thermophilus* replisomes might contain both an *in cis* DNA editing activity within their DnaE1-pol and an *in trans* activity from a DnaQ-exonuclease [24, 35, 60].

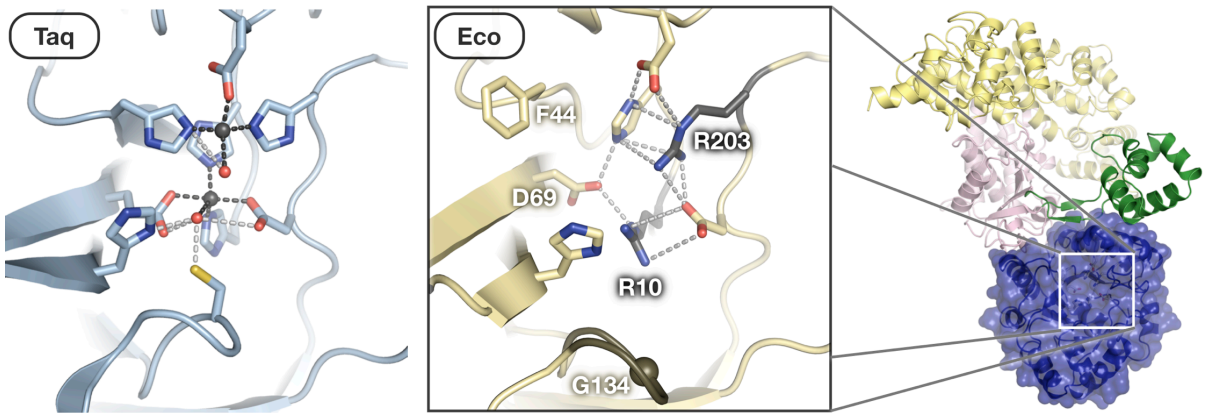


Figure 1-7. Two arginine residues replaced metals in the PHP domain of *E. coli* DnaE1-pol.

PHP cleft structures of *T. aquaticus* and *E. coli* DnaE1-pol. Black dashed lines correspond to liganding interactions with a zinc ion (<3 Å), and gray dashed lines indicate hydrogen bonding (<3.9 Å). (***Taq***) [24] (PDB 2hpi) Black spheres: two zinc ions, red spheres: two ordered waters. Although the *T. aquaticus* DnaE1-pol structure is missing the third canonical PHP metal, an ordered water is bound in roughly the expected position for this metal (Figure 2-1). The zinc ion in the first metal position (upper) has four ligands — two His, one Asp and an ordered water. The zinc ion in the second metal position (middle) has five ligands — two histidines, one aspartate, one glutamate and the ordered water (lower) that replaced the third metal. (***Eco***) [26] (PDB 2hnh) Five non-canonical residues are labeled (Table 1-1). The *E. coli* DnaE1-pol PHP cleft did not contain any metals, and, instead, the positively charged guanidino groups from two arginine residues (Arg10, Arg203) substituted for the two zincs seen in *T. aquaticus* DnaE1-pol. Four ordered waters and a phosphate ion that hydrogen bond to the two arginines in *E. coli* DnaE1-pol have been omitted for clarity.

1.4.3 Divergence within DnaE1-pol analyzed in this dissertation

To examine the Palm distortion of *E. coli* DnaE1-pol and the evolution of DNA editing in DnaE1-based replisomes, bioinformatic, structural and biochemical analyses were performed as described in CHAPTER 3. Briefly, I used sequence alignments to identify (Eco-like) motifs within *E. coli* DnaE1-pol and DnaQ-exonuclease that are characteristic of their homologs in α-, β- and γ-*proteobacteria.* In contrast, most other DnaE1-pol featured an alternative set of (Taq-like) motifs similar to those found in the *T. aquaticus* polymerase*.* I found no alternative sequence conservation among in the DnaQ-exonuclease homologs outside α-, β- and γ-*proteobacteria.*

Based on the alignments, I concluded that Eco- and Taq-like sequence motifs define the two main subtypes of DnaE1-pol. Considering the motifs in the context of the available structures, Eco-like DnaE1-pol generally conserve both an arginine residue (equivalent to Eco-Arg10), which replaces the middle PHP metal, and three Palm residues (equivalent to Eco-Phe402, Val552 and Phe554) that facilitate Palm distortion. Since DnaQ-exonuclease homologs containing a polymerase-binding tail [62] were only found in species with an Eco-like DnaE1-pol, I concluded that nearly all DnaE1-based replisomes derive their editing activity either from a PHP domain or a DnaQ-exonuclease. The taxonomic distribution of the two DnaE1-pol subtypes also allowed me to hypothesize the order and timing of their evolution. Before an α-

proteobacterium became the first mitochondrion more than 1,500 million years ago, the Eco-like polymerase had already diverged from the Taq-like subtype, which, itself, likely predates the emergence of cyanobacteria more than 3,500 million years ago (Figure 1-8) [63].

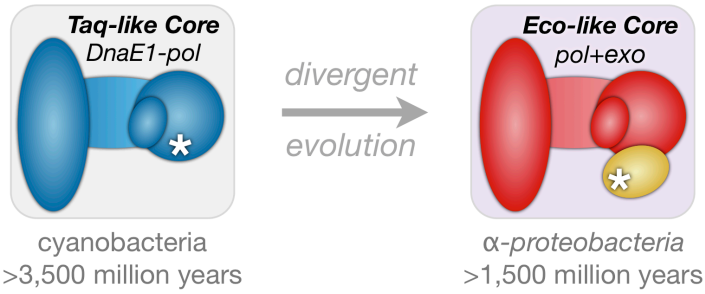


Figure 1-8. The Eco-like pol III Core evolved from a Taq-like ancestor more than 1,500 million years ago.

The ancestral Taq-like pol III Core is a single DnaE1-pol that is a constitutively active polymerase and contains a metal-binding PHP domain with 3'-5' exonuclease for editing (*) the DNA primer strand. In contrast, the Eco-like Core consists of three proteins: (1) a DnaE1-pol that lacks PHP metal binding and can adopt a distorted, inactive conformation, (2) an editing (*) DnaQ-exonuclease, and (3, not shown) a small protein (θ, HolE) that stabilizes the DnaQ (structure: Figure 3-10) [64, 65]. Ages of the Taq- and Eco-like Cores are based on their presence in modern-day cyanobacteria and *α-proteobacteria,* respectively [63].

Continuing the analysis of Eco-like DnaE1-pol, I present biochemical and bioinformatic data in CHAPTER 3 that are consistent with allosteric communication between the Palm domain and DnaQ-exonuclease subunit. In CHAPTER 2, I demonstrate that the PHP domain of *E. coli* DnaE1-pol is structurally highly conserved by restoring metal binding to it using only three point-mutations. Finally, I present advances toward DNA-bound structures of the *E. coli* polymerase in CHAPTER 4.

1.5 Conclusion

To pass information from one generation to the next, all cellular organisms rely on replicative polymerases that duplicate genomic DNA as part of a protein ensemble known as a replisome. Although few replisome proteins are conserved across the three domains of life, the functions they perform are, nevertheless, universal. Remarkably, the replicative polymerases, themselves, are not among the conserved proteins, and they instead belong to one of either two nonhomologous superfamilies. Eukaryota and Archaea use B- or D-family polymerases, which are members of the "classical" polymerase superfamily. In contrast, recent crystal structures of three C-family polymerases [24-26] show that the bacterial replicative enzymes belong to the Polβ-like nucleotidyltransferase (βNT) superfamily [23].

The C-family structures also served to highlight a division within in the DnaE1-pol homologs, which are the replicative polymerases from most known bacteria. Based on my analysis in CHAPTER 3, there appear to be two main subtypes of DnaE1-pol. Members of the Eco-like subtype are close homologs of the *E. coli* polymerase, which lacks PHP metal binding and contains a Palm that may adopt a distorted, inactive conformation. Compensating for their lack of PHP exonuclease activity, the Eco-like polymerases bind a DnaQ-exonuclease, which edits DNA *in trans.* The Taq-like DnaE1-pol homologs are more closely related to the *T. aquaticus* enzyme, which is a constitutively active polymerase whose metal-binding PHP domain edits DNA. Based on the phylogeny of the present-day bacteria that contain them, both the Taq- and Eco-like DnaE1-pol likely have ancient origins. The Eco-like polymerase predates the appearance of the first mitochondrion from an α-proteobacterium more than 1,500 million years ago, and the ancestral, Taq-like polymerase dates back over 3,500 million years to before the first cyanobacterium.

CHAPTER 2

Restored PHP Metal Binding Demonstrates Structural Conservation of the *E. coli* Replicative DNA Polymerase

2.1 Summary

Replisomes copy DNA with high fidelity, in part, by removing errors from the newly synthesized strand during polymerization. In many bacterial replisomes, this editing activity resides in the PHP domain [60] of the replicative polymerase (the α subunit of the DNA polymerase III holoenzyme) [40, 41]. The PHP domain of the *Escherichia coli* polymerase, however, lost metal binding more than 1,500 million years ago (CHAPTER 3) and therefore lacks editing activity. Nevertheless, the available crystal structures [26] of this polymerase suggest that the backbone of its PHP is highly conserved. Demonstrating the precise structural conservation of this domain, I present a crystal structure at 3.0-Å resolution of a triple point-mutant of the *E. coli* polymerase with restored PHP metal binding. X-ray fluorescence data collected on this mutant, in solution, suggests that the two metal-ions detected in the crystal structure are both zinc (and not manganese, iron or nickel). Given the strict distance and geometry requirements for zinc ligation, the restoration of metal binding to the *E. coli* replicative polymerase demonstrates the precise structural conservation of its PHP and suggests a central role for this domain in all bacterial replisomes.

2.2 Introduction

The C-family polymerases are found only in bacteria, which require these enzymes for DNA replication. Consequently, C-family members may prove to be an attractive target for antibiotic development [28-32]. In this context, both the universal and dissimilar features (Figure 1-5b) of PHP domains (Polymerase and Histidinol-Phosphatase [34]) from C-family members are largely uncharted terrain. Aside from the identification of DNA-editing exonuclease activity in *Thermus thermophilus* DnaE1-pol [60], C-family PHP domains have received little attention, and analysis of the *E. coli* DnaE1-pol PHP has been limited to noting that it structurally supports the Thumb (Figure 1-4) [26], lacks metal binding [24, 34], and binds the DNA-editing protein DnaQ-exonuclease [61].

PHP domains are α/β-barrels, similar to triosephosphate isomerase (TIM), but comprised of only seven β-strands $(\beta\alpha)_7$ [66] instead of eight. Canonical PHP domains catalyze the hydrolysis of phosphate moieties using a trinuclear metal center that is supported by nine amino acid residues. There are two types of PHP domains — those with an all-parallel β-strand arrangement and those with six parallel strands and one inverted, antiparallel strand (Table 1-1). The PHP domains of C-family polymerases are of the single-inversion type [24-26], as is the PHP in the protein tm0559 from *Thermotoga maritima* (PDB 2anu). All other PHP domains, including the ones occasionally fused to an X-family polymerase [36], are assumed to contain the all-parallel arrangement.

Because *E. coli* DnaE1-pol conserves less than half of the nine residues in the canonical PHP motif (Table 1-1), its PHP domain possesses neither metal binding (Figure 1-7) nor known enzymatic activity [26]. Despite its functional degeneracy, the overall architecture of this PHP closely resembles that of the metal-binding PHP domains from *Thermus aquaticus* DnaE1-pol [24] (PDB 2hpi) and *Geobacillus kaustophilus* PolC [25] (PDB 3f2d). The apparent structural conservation of the PHP in *E. coli* DnaE1-pol despite its loss of editing activity suggests the possibility of a universal, non-enzymatic role for the PHP domain found in each C-family polymerase.

To gauge the structural conservation of the PHP in *E. coli* DnaE1-pol, I supervised the solving of a crystal structure of 3mPHP, a triple point-mutant (Arg10His, Phe44His, Arg203His) of the crystallization construct EcoE1(917) (*E. coli* DnaE1-pol residues 1 through 917 of 1160). The three mutations correspond to canonical PHP metal binding residues, all of which are found in *T. aquaticus* DnaE1-pol and *G. kaustophilus* PolC (Table 1-1).

2.3 Materials & Methods

Work performed by Jordan Anaya, an undergraduate student, occurred under my supervision.

2.3.1 Polymerase cloning, expression and purification

Using standard restriction enzyme techniques, Xiaoxian Cao cloned the EcoE1(917) construct used by Lamers *et al.* [26] into the vector pET-28a-PP, which differs from the standard pET-28a (Novagen) by the replacement of a thrombin cleavage sequence with a PreScission Protease cleavage sequence immediately following the N-terminal His6-tag.

For constructs bearing point-mutations (3m-, 4m- and 5mPHP), J.A. performed mutagenesis with reagents equivalent to those of the QuikChange XL Site-Directed Mutagenesis Kit (Agilent Technologies).

J.A. expressed and purified 3m, 4m- and 5mPHP and determined protein concentrations by absorption at 280 nm based on theoretical molar extinction coefficients.

The T7 Express I^q (New England BioLabs) strain of phage-resistant *E. coli* served as the expression host for the three constructs, and growths occurred in culture flasks within incubator shakers. Cells were transformed by electroporation using a Gene Pulser II with a Pulse Controller Plus module (Bio-Rad), and, following a 1-h outgrowth at 37 °C with 400 µL unmodified Terrific Broth, selection for positive transformants occurred on agar plates containing 50 µg/mL kanamycin. After overnight growth for approximately 18 h, colonies were scraped from the plates *en masse,* resuspended at 37 °C in media consisting of phosphate-buffered Terrific Broth with 50 µg/mL

kanamycin, and grown to an optical density of roughly 1.0 (measured at 600 nm). At this point, 1-mL aliquots were taken to make -80 °C glycerol cell stocks before the cultures were diluted 1:1 with either room temperature (22 °C) or 4 °C media and moved to either 30° or 16 to 20 °C, respectively. Expression was induced by the addition of 1 mM IPTG once the cultures regrew to an optical density of 0.75 to 1.0. Expression of 3mPHP was conducted for either 3 h at 30 °C or roughly 18 h (overnight) at the lower temperatures. 4m- and 5mPHP did not express at 30 °C, so expression was performed exclusively at the lower temperatures for these polymerases.

J.A. prepared glycerol cell stocks by mixing 1 mL of cells, at an optical density of roughly 1.0, with 0.5 mL of 50% glycerol in a 2.0-mL tube. These samples were flash frozen in liquid nitrogen prior to long-term storage at -80 °C. Subsequent expressions were initiated by plating scrapings from the -80 °C stocks *in lieu* of repeat transformations.

Note that the following polymerase purification protocol has been updated as described in CHAPTER 4 for increased yield, purity and efficiency, but this updated version will not yield zinc-containing 3mPHP.

Following expression, 3mPHP cells were harvested by centrifugation, and the cells were resuspended in 25 mL of buffer I-500-25 (10% glycerol, 500 mM NaCl, 5 mM βME, 25 mM imidazole pH 8) per 1 L of 30 °C expression culture. For delayed purifications, the resuspended cells were flash frozen in liquid nitrogen and immediately moved to -80 °C for storage. When continuing directly to purification, the resuspended cells were lysed using an EmulsiFlex-C50 (Avestin) cell disruptor with an additional 50 mL of buffer I-500-25, followed by centrifugation for 40 min at 40,000 g and 4 °C to pellet the insoluble fraction. The supernatant was then filtered to 0.45 µm, loaded onto a 35-mL Ni Sepharose 6 Fast Flow (GE Healthcare) column, and washed with buffer I-500-25. With the polymerase still bound, the column was equilibrated in buffer β-A (10% glycerol, 20 mM HEPES pH 7.5, 5 mM βME). The polymerase was eluted from the Ni column by buffer I-0-500 (10% glycerol, 500 mM imidazole pH 8), loaded onto a 50-mL Heparin Sepharose 6 Fast Flow column (GE Healthcare), and eluted using a gradient from 0 to 40% buffer β-B (β-A plus 1 M NaCl) in 4 column volumes. Since the majority of the peak from the heparin column was contaminated with C-terminal degradation products of 3mPHP, only a small portion of the high-salt end of the peak was collected. This pool was digested with a version of the human rhinovirus 3C protease (HRV 3C) equivalent to PreScission Protease during an overnight dialysis with buffer β-A to a NaCl concentration of 100 mM. The dialyzed, digested sample was passed through the 35-mL Ni column, and the flow through was collected on a 5-mL HiTrap Q HP column (GE Healthcare). 3mPHP was eluted from the Q column using a 12 to 45% buffer β-B gradient over 16 column volumes, and the main peak was collected and concentrated to 6 mL using a Centriprep YM-50 concentrator (Millipore). The concentrated sample was run through a 16/60 HiLoad Superdex 200 prep grad size exclusion column (GE Healthcare) with isocratic elution in the buffer HKD (10 mM HEPES pH 7.5, 20 mM K glutamate, 10 mM DTT), and the peak was collected and concentrated using a YM-50 concentrator to 50.5 mg/mL.

4mPHP was purified using the same columns and gradients as 3mPHP but with slightly different buffers. In place of buffers I-500-25 and I-0-500, the buffers used were PI-500-25 (I-500-25 plus 25 mM Na/K PO_4 pH 8.5) and PI-0-500 (I-0-25 plus 25 mM Na/K PO_4 pH 8.5), respectively. Buffers β-A and β-B were replaced, respectively, by buffer T-A (β-A with HEPES replaced by 20 mM Tris pH 7.5) and buffer T-B (same as β-B with HEPES replaced by 20 mM Tris pH 7.5). The final buffer, HKD, was unchanged, and the endpoint 4mPHP concentration was 42.0 mg/mL.

5mPHP had very poor affinity for the heparin column, so this step was eliminated. The buffers used for 5mPHP were the same as those for 4mPHP with the exception of the final buffer, which was changed to HKD+ (K glutamate concentration increased to 30 mM, 5% glycerol added). 5mPHP was concentrated to 25.3 mg/mL.

2.3.2 Crystallization

J.A. conducted protein crystallization trials.

3mPHP crystallized in essentially the same condition as EcoE1(917) [26] (*protein:* 10 mg/mL, buffer HKD; *well solution:* 50 mM HEPES pH 7.5, 370 mM NaH_2PO_4, 14% PEG 3350). Crystallizations were set manually in hanging-drop vapor diffusion format above a reservoir of 400 µL well solution on siliconized glass cover slides using drops of 1+1+0.2 µL protein plus well solution plus diluted seed crystals of 3mPHP. Crystals were transferred to a cryoprotectant solution (well solution plus 15% glycerol and 2.5% higher PEG 3350) for flash freezing and storage in liquid nitrogen prior to data collection.

In addition to manual screening around the EcoE1(917) condition, 4mPHP and 5mPHP were subjected to extensive sitting-drop vapor diffusion crystallization trials at 100+100 nL (protein plus well solution) against commercial screens (e.g. the Index, PEG/Ion, and Natrix screens from Hampton Research and the JCSG+ and PACT screens from Qiagen) using a Phoenix crystallization robot (Art Robbins Instruments) in 3x 96-well Intelli-Plate crystallization trays (Art Robbins Instruments).

Misshapen spheres that may have contained 4mPHP appeared in a 100+100 nL protein to well solution drop (*protein:* 10 mg/mL, buffer HKD plus 1 mM $ZnCl_2$; *well solution:* PACT #94: 0.2 M Na/K PO_4, 0.1 M bis-tris propane pH 8.5, 20% PEG 3350).

2.3.3 Structure determination

J.A. generated diffraction from a single crystal using synchrotron radiation and collected data in frames of 1° oscillation per 20 sec exposure (Table 2-1). Meindert Lamers performed data reduction using HLK2000 [67] (HKL Research) and iMosflm [68], molecular replacement with Phaser [69] using the lower resolution EcoE1(917) structure (PDB 2hqa) as the input model, and refinement with phenix.refine [70]. M.L. carried out manual model building in Coot [71, 72].

2.3.4 X-ray fluorescence scanning

J.A. and I collected data at the Lawrence Berkeley National Laboratory, Advanced Light Source (ALS), Howard Hughes Medical Institute (HHMI) beamline 8.2.1. The 1-mM zinc standard was a standard crystallization loop coated with a $ZnCl_2$ solution and allowed to dry, and the 3mPHP sample was a similar loop holding a droplet of 3mPHP stock solution. Scans were collected in 1-eV steps using 2-sec exposures. Scans were individually baseline-corrected by subtracting the average intensity of the first 10 points from all points in the scan. I generated trend lines to the data using the "Weighted" algorithm with a smoothing factor of 40% in KaleidaGraph 4 (Synergy Software).

2.3.5 Molecular models and in silico mutagenesis

PyMOL 1.3 [73] (Schrödinger) was used to prepare all structure-based figures and to generate the Asp69Glu *in silico* mutants of 3mPHP (Figure 2-2).

2.3.6 Tryptophan fluorescence microscopy

J.A. captured tryptophan fluorescence images and accompanying visible light images using a PRS-1000 Protein Review Station microscope (Korima).

2.4 Results & Discussion

2.4.1 Extra electron density appeared in the structure of 3mPHP

Crystals of 3mPHP grew under the same conditions as EcoE1(917), were morphologically indistinguishably, and shared the same space group ($P2_12_12_1$). The phases for 3mPHP were solved by molecular replacement using the lower, rather than the higher, resolution EcoE1(917) structure (PDB code 2hqa) because its unit cell dimensions were a better match (higher = 2hnh: 82.6, 92.6, 130.2 Å; lower = 2hqa: 83.1, 98.5, 139.5 Å; 3mPHP: 82.4, 99.0, 140.1 Å, Table 2-1). Globally, there were no significant conformational differences between EcoE1(917) and 3mPHP (data not shown), so manual intervention was only performed during the building of the PHP domain of 3mPHP. Upon completion, this domain (residues 5 through 270) had a RMSD of 0.339 Å (1512 atoms) and 0.309 Å (1568 atoms) relative to PHP domains in the high and low resolution EcoE1(917) structures [26].

Data Collection	
Beamline	ALS 8.2.2
Wavelength (Å)	1.000
Space group	$P2_12_12_1$
Unit cell (Å)	82.4 99.0 140.1 90 90 90
Resolution (Å)	71-3.0 (3.15-3.0)
R_{merge}	14.5 (75.2)
$I/\sigma I$	11.3 (1.9)
Completeness (%)	99.7 (100.0)
Multiplicity	5.4 (5.6)

Refinement	
Resolution (Å)	71-3.0
No. reflections	23834
test set	5%
Rwork/Rfree	19.9/25.4
Monomers/a.u.	1
No. TLS domains	5
No. atoms	7183
protein	7148
waters	18
ligands	17
RMSD length	0.003
RMSD angle	0.752

Table 2-1. Crystals of 3mPHP and EcoE1(917) share a space group and similar cell dimensions.

3mPHP crystal structure: data collection and refinement.

After a round of refinement, difference density maps revealed peaks of 4.5 and 5.5 σ (Figure 2-1a) in the PHP cleft of 3mPHP. Based on their intensity and roughly spherical shape, the peaks did not appear to correspond to water or any of the components of the crystallization mother liquor. The peaks were also in positions corresponding to the two PHP metal sites that were shared by the *T. aquaticus* DnaE1-pol (Figure 1-7) [24] and *G. kaustophilus* PolC structures (Figure 2-1a) [25] (PDB 2hpi and 3f2d, respectively).

Because X-ray diffraction data for 3mPHP was only collected at a single wavelength, they could not be used to identify the species that gave rise to the extra density in the PHP cleft. To evaluate the possibility that the cleft density corresponded to bound metals, I performed X-ray wavelength scans, and, near the inflection point for zinc, I observed fluorescence peaks of similar shape, position and intensity for a solution of 3mPHP (50.5 mg/mL, 0.49 mM) and a 1-mM zinc standard (Figure 2-1c). This result is consistent with zinc binding by 3mPHP and does not support the presence of manganese, iron or nickel.

2.4.2 3mPHP binds two zinc ions with near-canonical geometry

Since the electron density peaks in the PHP cleft of 3mPHP likely both correspond to zinc ions, all three of the histidine point mutations appear to contribute to zinc binding (Figure 2-1a), as do several other residues at positions corresponding to the PHP motif. The Arg203His mutation created a pocket in the 3mPHP cleft, and, together with Phe44His, His203 ligated to one of the two zinc ions. Including an interaction with a bound phosphate (not shown), this zinc had roughly the same tetrahedral geometry observed for the corresponding zinc in the *T. aquaticus* and *G. kaustophilus* polymerase structures (Figure 1-7, Figure 2-1).

The third mutation, Arg10His, opened another pocket within the PHP cleft and participated in the binding of a second zinc ion (Figure 2-1a). The coordination of this metal, however, differed from that seen in the other polymerases, because, at a distance of 3.5 Å, the canonical residue, Asp201, at PHP position-8 was roughly 1.5 Å beyond optimal ligation range (Figure 2-1b) [74, 75]. As a result, the second zinc contacted only three protein side chains and one of oxygen atom from the bound phosphate.

In addition to lacking an interaction with Asp201, the second zinc also binds to the non-canonical residue, Asp69, at position-5 in the PHP motif (Figure 2-1a, Figure 2-2). The standard residue at this position is a glutamate, and, because the side chain of aspartate is shorter by one carbon than that of glutamate, binding by Asp69 may have pulled the zinc away from canonical Asp201 on the other side of the PHP cleft.

Comparing the *T. aquaticus* DnaE1-pol and *G. kaustophilus* PolC structures to that of 3mPHP shows that its second PHP metal is slightly closer to motif position-5 (Asp69) than the corresponding metal bound to the other two polymerases (Figure 2-2). The cross-cleft distance between the α-carbons of position-5 and -8, however, are nearly identical in the structures of *T. aquaticus* DnaE1-pol, *G. kaustophilus* PolC, and 3mPHP (Figure 2-2), which suggests that adding an additional Asp69Glu point-mutation to 3mPHP may restore canonical metal binding at the second PHP site.

Interestingly, although it has regained metal binding, the cross-cleft distance in 3mPHP is also the same, to within a tenth of an angstrom, as that of the wild type *E. coli* polymerase (position-5 to -8 α-carbon distance, 10.5 Å for both). Because hydrogen bonds are generally longer than liganding interactions with zinc, Asp69 may be necessary to accommodate the structural arginine (Arg10) at the second metal position in the wild type cleft (Figure 1-7). Were residue 69, instead, the canonical glutamate, it would likely suffer a major steric clash with Arg10. More importantly, the preservation of the wild type cross-cleft distance by 3mPHP, despite the restoration of metal binding, demonstrates that the wild type domain is the product of exacting structural conservation in the absence of enzymatic activity.

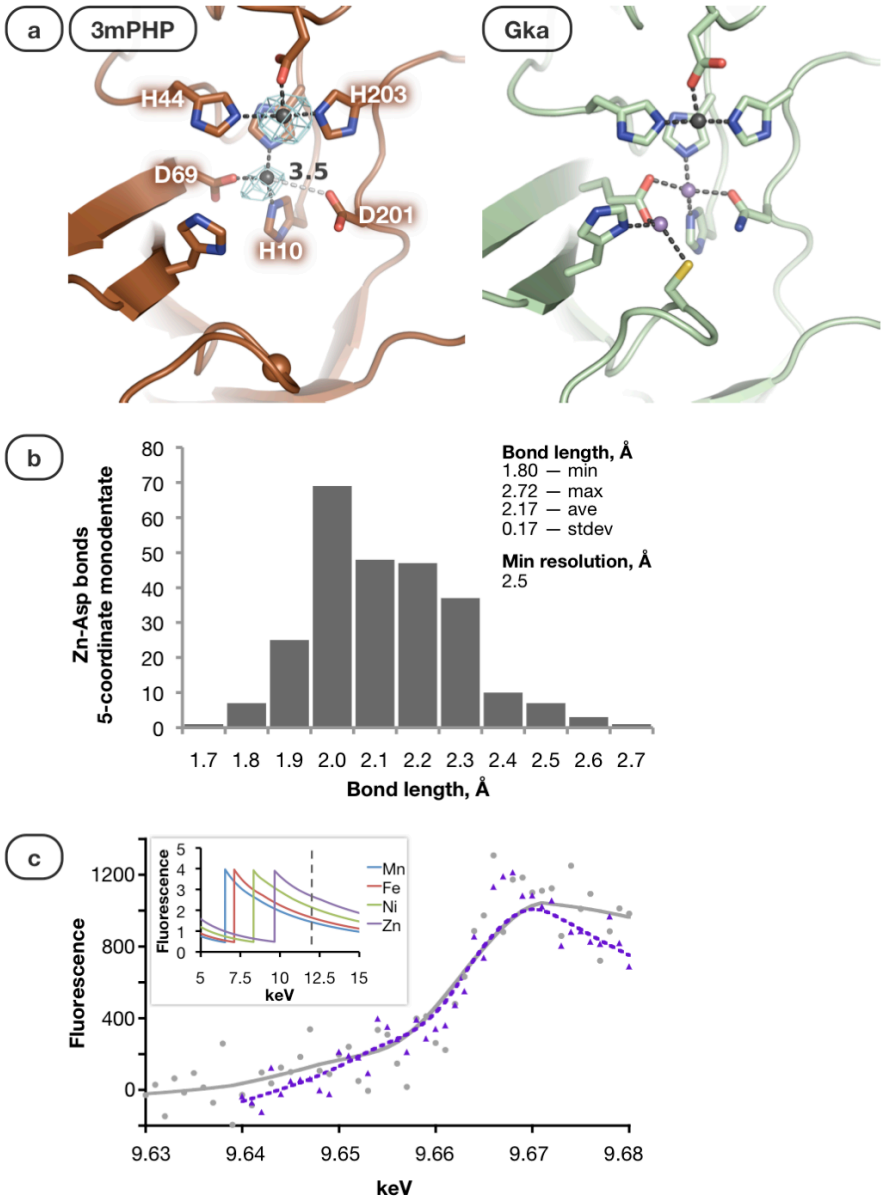


Figure 2-1. 3mPHP bound zinc ions at two of the three canonical PHP sites.

(***a***) PHP clefts of 3mPHP and *G. kaustophilus* (PDB undeposited and 3f2d, respectively). Cyan mesh: anomalous difference electron density contoured at 3.5 σ, black spheres: zinc ions, purple spheres: manganese ions, black dashed lines: bonds to metals, white dashed line: 3.5 Å distance. A phosphate bound to the metals in each structure was removed for clarity. (***b***) Interatomic distances observed for monodentate ligation of five-coordinate zinc by aspartate side chains (average: 2.17 ± 0.17, bonds analyzed: 255) [74, 75]. (***c***) X-ray fluorescence of a zinc standard compared to a 3mPHP solution. Inset: Theoretical f" scattering factors for manganese, iron, nickel and zinc [76-78] relative to the energy (dashed line, 12.4 keV = 1 Å wavelength) used for 3mPHP structure determination.

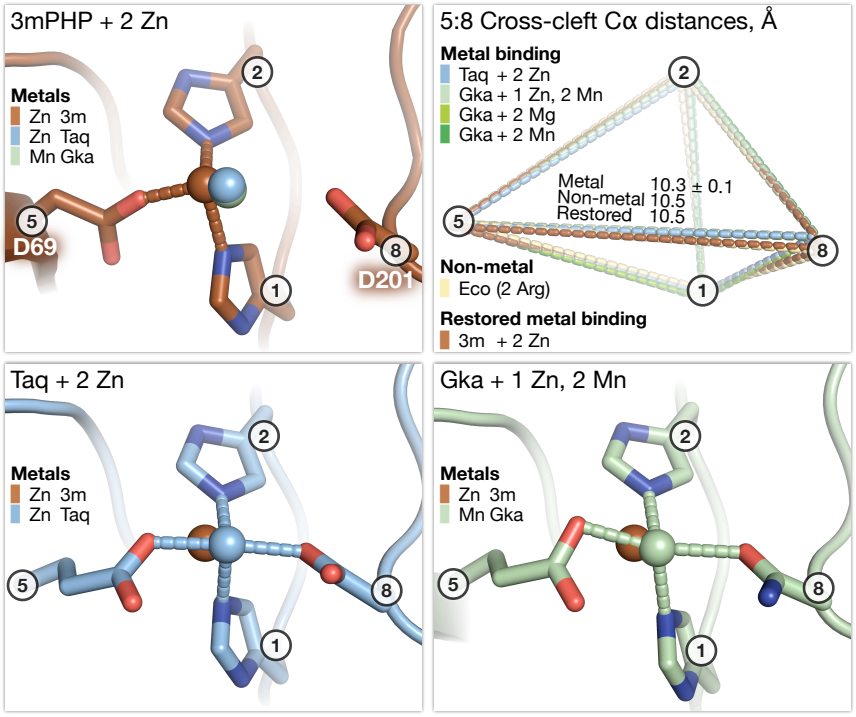


Figure 2-2. Adding Asp69Glu to 3mPHP may complete canonical coordination of the second zinc.

Using the α-carbons of PHP motif position-1, -2, -5 and -8 to superimpose the binding sites for the second PHP metal (spheres) in 3mPHP, *T. aquaticus* DnaE1-pol (Taq, PDB 2hpi) and *G. kaustophilus* PolC (Gka, PDB 3f2d) shows that the metal is exceptionally close to position-5 (and further from position-8) in 3mPHP. For comparison, the 3mPHP metal is included in the structures of Taq and Gka, and their metals are included for 3mPHP. Because the superimposed α-carbon positions are nearly identical in all the high-resolution C-family polymerase structures (upper left), an Asp69Glu mutation in 3mPHP might reposition the metal and restore binding with Asp201.

2.4.3 Additional PHP mutations impede crystallization

Building upon 3mPHP, I supervised the preparation of two mutant polymerases (4m- and 5mPHP, respectively, Table 1-1) in which all nine positions in the PHP metal-binding motif contained residues capable of metal ligation. Neither mutant yielded crystals in the standard EcoE1(917) condition, so an attempt was made to find alternative conditions for these polymerases. From this effort, one drop of 4mPHP with $ZnCl_2$ produced 10 to 20 small (perhaps 10 μm), misshapen spheres, which appeared to contain protein on the basis of their tryptophan fluorescence (Figure 2-3). These objects could not be reproduced.

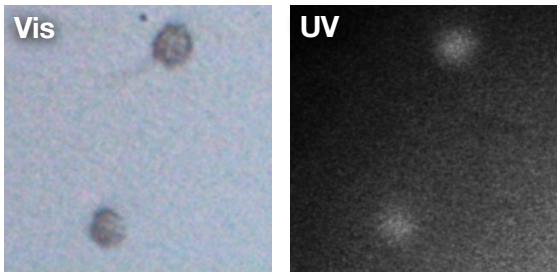


Figure 2-3. Tryptophan fluorescence suggests that the spheres contained protein.

Visible and tryptophan fluorescence (ultraviolet, UV) images of a crystallization trial of 4mPHP.

Since these crystallization attempts were made, I have developed techniques to monomerize both 4m- and 5mPHP (CHAPTER 4), which tend to form high molecular weight aggregates under conditions suitable for the storage of EcoE1(917) and 3mPHP. Furthermore, I demonstrated that maximal stabilization of 4mPHP occurred at substantially lower zinc concentrations (2.5 to 10 μM, CHAPTER 3) than I had previously used (up to 1 mM). Both of these technical advances may assist future crystallization efforts.

2.4.4 Why retain a PHP cleft that lacks metal binding?

In nature, there are several pairs of homologous proteins in which one binds a metal ion while the other has replaced the ion with a positively charged amino acid. Examples include both familiar enzymes, such as thrombin [79-81], pyruvate kinase [81, 82], ferritin [83-85] and tyrosyl-tRNA synthetase [86-89], and also more exotic proteins like snake venom myotoxins [90-95]. Among these, the venom phospholipase A_2 homologs are the only proteins that, like the *E. coli* DnaE1-pol PHP, appear to have lost catalytic activity. In all others, the metal and the positively charged amino acid serve the same catalytic roles in their respective homologs. In the venom, however, the non-metal-binding protein has a different function than the conventional homologs, so it is subject to different evolutionary pressures. A similar scenario may also exist for the non-metal-binding PHP domain in *E. coli* DnaE1-pol.

Looking beyond the C-family polymerases, the ribonuclease P protein Ph1877p from *Pyrococcus horikoshii* OT3 [66, 96] contains a PHP domain that lacks metal binding. Despite this loss of function, Ph1877p has been retained because it is one of five proteins that stabilize a catalytic RNA. The existence of this structurally divergent PHP domain, which lacks a cleft formed by the C-terminal ends of its β-barrel, shows that sites compatible with metal binding may be completely lost from PHP domains without selective pressure to retain them. Thus, precise structural conservation of the PHP domain of *E. coli* DnaE1-pol, despite its lack of metal binding, suggests either recent loss of binding or intense selective pressure for maintenance of the PHP cleft.

In CHAPTER 3 (Figure 3-7), I show that the *E. coli* polymerase is representative of a subtype of DnaE1-pol (Eco-like) that evolved from an ancestral subtype, exemplified by the *T. aquaticus* polymerase (Taq-like), more than 1,500 million years ago. Given its ancient nature, the structure of the PHP cleft in *E. coli* DnaE1-pol was probably conserved because it is important for the proper functioning of this enzyme.

Assuming that the C-family polymerase PHP domain plays a structural role in catalysis, the conservative replacement of metals by arginines would probably be favored over more heavily mutated PHP domains. As my data in CHAPTER 3 show, *E. coli* DnaE1-pol is highly sensitive to perturbations of its PHP, as evidenced by the significantly decreased polymerization rates and stabilities of the metal-binding mutants (3m-, 4m- and 5mPHP) relative to EcoE1(917).

When replicating genomic DNA, the *E. coli* polymerase functions as part of a replisome (Figure 1-2) [14], and its PHP domain serves as the binding site for the editing subunit, DnaQ-exonuclease [61, 62]. In the model proposed by Lamers *et al.* [26] for *E. coli* DnaE1-pol binding by DnaQ-exonuclease, this editing subunit is positioned over the "groove" that stretches from the base of the Thumb into the PHP cleft (Figure 1-4). In polymerases like *T. aquaticus* DnaE1-pol, the groove may serve as the binding site for the DNA primer strand as it is edited in the metal-binding PHP cleft, and, if Lamers *et al.* [26] are correct, the corresponding groove and cleft on *E. coli* DnaE1-pol could similarly direct the primer to the DnaQ-exonuclease. Evans *et al.* [25] proposed a similar model for PolC homologs (in which a DnaQ-exonuclease is inserted into the PHP domain, Figure 1-5b) based on their structures of the *G. kaustophilus* polymerase, such a conserved role for the PHP cleft would make it essential for faithful DNA replication.

2.5 Conclusion

As demonstrated by X-ray crystallography, three point-mutations were sufficient to restore metal binding to the PHP domain of *E. coli* DnaE1-pol. This result demonstrates an extraordinarily high degree of structural conservation for the PHP despite its ancient loss of editing exonuclease activity. Since such conservation is not an obligate property of PHP domains, the domain in *E. coli* DnaE1-pol is likely the product of intense selective pressure for its structural roles in the stability and activity of the polymerase (CHAPTER 3). Furthermore, as the binding site for a DnaQ-exonuclease, this PHP may also have importance in the context of DNA editing.

More broadly, the PHP domain of *E. coli* DnaE1-pol appears to be part of a larger evolutionary trend among DNA polymerases toward the accumulation of inactivated domains. While the replicative polymerases of bacteria and eukaryotes all contain or recruit an editing exonuclease (CHAPTER 1 and [10]), many also feature inactivated exonuclease domains. Four multi-subunit B-family polymerases function at the eukaryotic replication fork [10], and among them are several examples of inactivated

domains. In total, the four polymerases contain six inactivated exonucleases (three domains within synthesis subunits and three calcineurin-like domains in separate subunits [34, 97]), four inactivated uracil binding domains (one per catalytic subunit [98]), and also an inactivated, but otherwise fully intact, polymerase domain (C-terminal within the synthesis subunit of Pol ε [99]). Although inactivated polymerases may be generally rare, a family of highly conserved pseudo-polymerases was recently discovered in Archaea [100], and a polymerase homolog, ImuB [48], in *Mycobacterium tuberculosis* similarly lacks catalytic residues and is co-expressed with an error-prone DnaE2-pol.

Similar to that of the PHP in *E. coli* DnaE1-pol, the stories of other inactivated polymerase-associated domains will almost certainly be ones of stabilization and protein-protein interaction [97]. Accordingly, a rich understanding of DNA replication will require structural analysis of active polymerases in complex with their non-enzymatic partners.

CHAPTER 3

Coevolution and Structural Coupling of DNA Synthesis and Editing in Divergent Bacterial Replisomes

3.1 Summary

Using sequence alignments, I identify the two main versions, respectively exhibited by *Escherichia coli* and *Thermus aquaticus,* of the DnaE1-based replisome found in most bacterial species. In the more ancient replisome, a Taq-like DnaE1-pol (α subunit) [40, 41] constitutes the entire DNA polymerase III 'Core' and both synthesizes and edits DNA *in cis.* The Eco-like pol III Core (Figure 1-2) [14], which is found in *α-, β-* and *γ-proteobacteria,* contains a DnaE1-pol that diverged from a Taq-like ancestor when its PHP domain lost DNA editing activity and bound a DnaQ-exonuclease (ε subunit, MutD) for editing *in trans.* Interpreting the available high-resolution structures of C-family polymerases [24-26], I propose that a three-residue motif in the Palm of *E. coli* DnaE1-pol may facilitate its observed inactive conformation, which I define based on distortion of the aspartate triad required for DNA synthesis and burial of an arginine residue crucial for DNA proofreading. Building on these interpretations, Statistical Coupling Analysis (SCA) [101] suggests the possibility of allostery between synthesis and editing in both Taq- and Eco-like pol III Cores, which is consistent with the diminished rate of DNA synthesis for PHP point-mutants of *E. coli* DnaE1-pol.

3.2 Introduction

In 1964, Roy-Burman and Sen [55] working in Calcutta, India, showed that the growth of some Gram-positive bacteria could be selectively inhibited by a small molecule at 0.5 μM — a concentration 2,000 to 4,000 times lower than that affecting Gram-negative species. Unbeknownst to them [54], this was the first evidence for the most major schism within bacterial replisomes, the division between PolC- and DnaE1-based pol III. Over the next 10 years, Kornberg [57, 58], Cozzarelli [56] and their respective colleagues discovered pol III and demonstrated differences between versions found in Gram-negative and -positive species. Prominent among the differences was the susceptibility of Gram-positive pol III to another antimicrobial small molecule, which allowed the Cozzarelli group to prove that pol III was the replicative polymerase in *Bacillus subtilis* [53].

Since then, researchers have come to learn that the enzymes inhibited by the two molecules are PolC homologs and that they are the pol III α subunits of the low-GC Gram-positive bacteria comprising the phyla *Firmicutes* and *Tenericutes* [39]. Essentially all other bacteria, both Gram-positive and -negative, use a pol III based on a DnaE1-pol homolog. Consequently, these are the polymerases that replicate the genomes of the most diverse and abundant species on Earth [102, 103].

The first crystal structures of the DnaE1-pol subunits of pol III from *E. coli* [26] (residues 1 through 917 of 1160, EcoE1(917)) and *T. aquaticus* [24] threw into stark

relief the differences in their PHP (Polymerase and Histidinol-Phosphatase [34]) and Palm domains. Consistent with the possession of both DNA editing and synthesis activities, the *T. aquaticus* DnaE1-pol bound metals to its PHP domain (Table 1-1, Figure 1-7) [60], and, in its Palm, the three catalytically important aspartates bound a magnesium ion in a manner consistent with the two-metal catalytic mechanism (Figure 3-2) conserved by [104-106] members of the Polβ-like nucleotidyltransferase superfamily (βNT) [23]. These enzymes, which include the C- [24] and X-family [106] DNA polymerases, use three aspartate residues to bind directly to the catalytic metal that coordinates the primer DNA strand 3'-hydroxyl and the α-phosphate that it attacks during covalent bond formation to the incoming nucleotide [105] (Figure 3-2b).

In contrast, the *E. coli* polymerase appeared unfit for either editing or synthesis based on its non-metal-binding PHP (Table 1-1, Figure 1-7) and a Palm presenting its aspartate residues in an extended, obtuse triangular, arrangement (Figure 3-2). Although the loss of PHP metal binding was predicted [34], the Palm distortion of *E. coli* DnaE1-pol came as a surprise because it is one of the fastest known polymerases, synthesizing DNA at 750 bases per second in the context of its replisome [14].

By drawing connections between structure, sequence and biochemical data, I present definitions of two contrasting and widely abundant subtypes of pol III, as distinguished by the coevolution of DNA synthesis and editing activities within their respective Cores. Comparing the distribution of these two subtypes to the known phylogeny of bacteria [102], I show that the more recently evolved (Eco-like) pol III subtype, such as is found in *E. coli,* possesses a Core [14] containing a DnaE1-pol and a DnaQ-exonuclease. Respectively, these two enzymes perform DNA synthesis and error removal *in trans.* In contrast, the ancestral (Taq-like) pol III subtype, an example of which is found in *T. aquaticus,* has a Core consisting of only a single DnaE1-pol that combines synthesis and editing activities *in cis.* This finding directly contradicts previous speculation regarding of the possibility of multiple editing exonuclease activities in pol III from *T. aquaticus* and related species [24, 35, 60].

Encouraged by the apparent coevolution between the Palm and PHP domains of DnaE1-pol homologs and their DnaQ-exonuclease partners, I and a coworker, Brian Kelch, perform preliminary biochemical and bioinformatics analyses, the results of which are consistent with communication between DNA synthesis and editing in both Eco- and Taq-like pol III Cores. This is of particular interest with respect to *E. coli* DnaE1-pol, which synthesizes longer stretches of DNA in the presence of its partner DnaQ-exonuclease [107]. In this case, binding of the exonuclease might allosterically bias the polymerase toward an active conformation and away from the distorted state observed in crystal structures.

This chapter is organized in the following way. After the Materials & Methods, the Results & Discussion is divided into five parts. Part A — PHP & Palm — identifies three Eco-like Palm residues that coevolved with the loss of PHP metal binding in the DnaE1-pol homologs from *α-, β-* and *γ-proteobacteria.* In Part B — Palm Distortion, I propose how the three Eco-like Palm may residues contribute to the distorted conformation observed for *E. coli* DnaE1-pol. Conservation of an Eco-like DnaQ-

exonuclease within α-, β- and γ-*proteobacteria* is presented in Part C — DnaQ-exonuclease — along with a phylogenetic tree and timeline for the evolution of DnaE1-based pol III. Part D — Allostery & Bioinformatics — contains Statistical Coupling Analysis that suggests the possibility of allostery within the Eco-like pol III Core. This is consistent with synthesis and stability data for *E. coli* DnaE1-pol presented in Part E — Allostery & Biochemistry. Finally, the above results are briefly summarized in the Conclusion.

3.3 Materials & Methods

3.3.1 Selection of DnaE1-pol sequences

Using the PSI-BLAST [108] module of the Bioinformatics Toolkit [109] at the Max-Planck Institute for Developmental Biology, Tübingen, Germany, Brian Kelch initiated a query with the *E. coli* DnaE1-pol sequence and retrieved the sequences of 500 apparently homologous polymerases with less than 90% identity. From these, he manually removed the sequences that appeared to correspond to more distantly related, non-replicative homologs (e.g. DnaE2-pol and DnaE3-pol) by performing a sequence alignment, generating a phylogenetic tree using PHYLIP-NEIGHBOR [110], and deleting those sequences that failed to clade with known DnaE1-pol homologs. B.K. used this curated sequence set as the starting point for Statistical Coupling Analysis (defined later in Materials & Methods). Although a DnaE3-pol usually complements PolC, a DnaE1-pol is sometimes present, instead [40, 41], and B.K. included representatives of these homologs in the set used for Statistical Coupling Analysis.

B.K. manually selected a subset of roughly 50 DnaE1-pol sequences from those used for Statistical Coupling Analysis. From these, I chose 47 homolog sequences for more in-depth analysis.

3.3.2 Sequence alignments and phylogenetic trees

Brian Kelch used MUSCLE [111-113] (MUltiple Sequence Comparison by Log-Expectation) version 3.6 for preparation of the input alignments for Statistical Coupling Analysis.

I used MAFFT (Multiple Alignment by Fast Fourier Transform) L-INS-i version 6.8x [114-116] for sequence alignment and Jalview [117, 118] to sort alignments based on phylogenetic trees, for minor manual alignment adjustments, and for figure preparation.

For publication, I generated phylogenetic trees using the MAFFT web server [119] with the NJ (Neighbor Joining) [120] method on all ungapped sites, the WAG (Whelan And Goldman) [121] substitution model, and estimation of heterogeneity among sites. FigTree [122] was my program of choice for tree figure preparation.

3.3.3 Identification of characteristic Eco-like and non-Eco-like residues

The Two Sample Logo web server [123, 124] was used for quantitatively identifying residues specific to either Eco-like or non-Eco-like DnaE1-pol, DnaQ-exonuclease and the DNA sliding clamp. Provided with a sequence alignment that has been split into an ingroup (Positive Sample) and outgroup (Negative Sample), Two Sample Logo identifies the statistically significant residue differences between the two groups. DnaE1-pol and DnaQ-exonuclease from representative α-, β- and γ-*proteobacteria* were used as the Eco-like (Positive Sample) sequences and homologous proteins from other bacteria were used as the non-Eco-like (Negative Sample) sequences.

3.3.4 Statistical Coupling Analysis

Brian Kelch performed Statistical Coupling Analysis (brief description in Part D — Allostery & Bioinformatics) [101, 125, 126] using the Gerstein lab coevolution web server [127, 128], supplemental data processing in MATLAB (MathWorks), and heat map generation in HippoDraw [129]. For intermolecular Statistical Coupling Analysis, similar to that performed on histidine kinase / response regulator pairs [130-133], B.K. concatenated the sequences for two subunits into a single sequence (e.g. DnaE1-pol plus DnaQ-exonuclease) for each species prior to alignment and submission to the Gerstein server.

3.3.5 Modeling a Taq-like active conformation for E. coli *DnaE1-pol*

Meindert Lamers aligned the *E. coli* and *T. aquaticus* DnaE1-pol sequences with manual adjustment. With this alignment and the apo-structure of *T. aquaticus* DnaE1-pol (PDB 2hpi) [24] as inputs, M.L. used Modeller [134, 135] to model a Taq-like, active conformation for *E. coli* DnaE1-pol. To further the comparison of his model to the higher resolution EcoE1(917) structure, M.L. used the Yale Morphing Server [136] to generate an energy-minimized pathway between the structure and the model.

Using the input files from M.L., I produced 100 (data not shown) additional Taq-like models for *E. coli* DnaE1-pol, and I visually compared the position of Phe402 (Figure 3-3) in these models to that in the model from M.L. In 79 of my models, Phe402 had rotated into a position that I considered comparable to M.L. model position, while 17 models included what I interpreted as an unrotated Phe402 in a position like that observed in the EcoE1(917) structures. The four remaining models showed Phe402 in an intermediate position. Because the Phe402 rotation appeared in a significant majority of the models, I used the original model generated by M.L. for my structure analyses.

3.3.6 Polymerase cloning, expression and purification

As mentioned in CHAPTER 2, EcoE1(917) was cloned into pET-28a-PP, which was modified from the standard pET-28 by replacement of the thrombin cleavage site with

a PreScission Protease-compatible sequence. This EcoE1(917) with an N-terminal His6-tag was purified using the 3mPHP protocol from CHAPTER 2 with minor variations.

Protein concentrations were determined by absorption at 280 nm based on their theoretical molar extinction coefficients.

Following EcoE1(917) expression for 19.5 h (overnight) at 18 °C, the host cells were harvested and resuspended in buffer PI-500-25 (10% glycerol, 25 mM Na/K PO_4 pH 8.5, 500 mM NaCl, 5 mM βME, 25 mM imidazole pH 8). The cells were flash frozen in liquid nitrogen and stored at -80 °C. Polymerase purification was begun by partially thawing the stock of frozen, resuspended cells in a 30 °C water bath for 30 min. The thawed cells were lysed using an EmulsiFlex-C50 (Avestin) cell disruptor, and, after centrifugation and filtration, the pH of the supernatant was increased to 8.9 by the addition of 0.5 M tribasic Na/K PO_4 (stock pH 11.75). The sample was loaded onto 2x in-series 5-mL HisTrap HP columns (GE Healthcare) that had been equilibrated in buffer PI-500-0 (same as PI-500-25 without imidazole) with 4% buffer PI-40-500 (same as PI-500-25 with the NaCl concentration reduced to 40 mM and the imidazole concentration increased to 500 mM). After elution of the unbound sample components, the column was equilibrated in buffer β-A while EcoE1(917) was still bound. The polymerase was eluted off the column using buffer PI-40-500, directly collected on 2x in-series 5-mL HiTrap Q HP columns (GE Healthcare), and eluted from these columns using a gradient from 12 to 45% β-B in 12 combined-column volumes. Only the early portion of the main eluate peak was pooled, because the latter portion contained high levels of proteolytically degraded EcoE1(917). The pool was dialyzed overnight against buffer β-A to a NaCl concentration of roughly 100 mM in the presence of 400 μL of roughly 8-mg/mL human rhinovirus 3C protease (HRV 3C) equivalent to PreScission Protease. After dialysis, the sample was run over a 5-mL HisTrap HP column, and the flow through was collected on a 5-mL HiTrap Q HP column. The polymerase was eluted from the Q column using 100% buffer β-B and concentrated to roughly 6 mL using a Centriprep YM-30 concentrator (Millipore). The concentrated sample was loaded on a 16/60 HiLoad Superdex 200 prep grad size exclusion column (GE Healthcare) and isocratically eluted in the buffer TK100D (10 mM TAPS pH 8.5, 100 mM K glutamate, 10 mM DTT). The peak was concentrated to 40.8 mg/mL EcoE1(917).

Using the protocol described in CHAPTER 2, 3mPHP was expressed overnight (18.5 h) at 16 °C, and the host cells were harvested by centrifugation. Resuspension was performed in buffer PI-500-0 (10% glycerol, 25 mM Na/K PO_4 pH 8.5, 500 mM NaCl, 5 mM βME), and the cells were lysed using an EmulsiFlex-C50 (Avestin) cell disruptor with an additional 50 mL PI-500-0, followed by centrifugation for 40 min at 40,000 g and 4 °C to pellet the insoluble fraction. The supernatant was filtered to 5 μm and then to 0.45 μm after which its pH was raised to roughly 9 with the addition of approximately 50 mL of 0.5 M tribasic Na/K PO_4. An aliquot from a 2 M imidazole solution at pH 8 was added to the supernatant for a final concentration of 40 mM imidazole. This sample was loaded on 2x in-series 5-mL HisTrap HP columns (GE Healthcare), washed with a buffer containing 20 mM imidazole (96% buffer PI-500-0

plus 4% PI-0-500, which is PI-500-0 without NaCl and plus 500 mM imidazole). After washing, the column with the polymerase still bound was equilibrated in buffer β-A (10% glycerol, 20 mM HEPES pH 7.5, 5 mM βME) followed by elution with PI-0-500. The eluate was loaded on a 5-mL HiTrap Q column (GE Healthcare) and eluted using a gradient from 12 to 45% buffer β-B (same as β-A plus 1 M NaCl) in 12 column volumes. The early portion of the elution peak was pooled and dialyzed overnight against buffer β-A to a NaCl concentration of roughly 100 mM in the presence of 400 μL of roughly 8 mg/mL human rhinovirus 3C protease (HRV 3C) equivalent to PreScission Protease. After dialysis, the sample was rerun over 2x in-series 5-mL HisTrap HP columns, and the flow through was collected on a 5-mL HiTrap Q HP column. The polymerase was eluted from the Q column using 100% buffer β-B and concentrated to roughly 6 mL using a Centriprep YM-30 concentrator (Millipore). The concentrated sample was loaded on a 16/60 HiLoad Superdex 200 prep grad size exclusion column (GE Healthcare) and isocratically eluted in the buffer HK100D (10 mM HEPES pH 7.5, 100 mM K glutamate, 10 mM DTT). Note that the presence of chelators, glutamate and DTT, in this buffer strip bound metals from 3mPHP, thus delivering it in an apo, metal-free form. The peak was concentrated to 43.1 mg/mL 3mPHP.

Jordan Anaya prepared 4mPHP and 5mPHP as described in CHAPTER 2.

3.3.7 DNA polymerization assay

DNA polymerization was monitored by fluorescence intensity using a slightly modified version of the standard PicoGreen-based quench assay [137-139] in 96-well format. The substrate was a DNA primer-template complex generated by annealing two oligomers:

```
Tem50   3-  CCCTGCGTGCGCCGTAAGTTCCTGAATCCAGACATAAATAGATGGGTGTT  -5
Pri23   5-  GGGACGCACGCGGCATTCAAGGA                            -3
```

$$F = F_{\infty} - Ae^{-kt}$$

Equation 3-1. Exponential fit of DNA polymerization data.

Fluorescence, F, was monitored as a function of time, t. KaleidaGraph fit the three other parameters. F_{∞} is the fluorescence at time infinity (i.e. the fluorescence were the reaction to reach completion). A is the fluorescence amplitude increase from reaction initiation to completion. The parameter of interest, k, is the reaction constant, which is proportional to the rate of DNA polymerization.

Time point samples were taken until the reactions essentially reached completion. After quenching the last sample, 90 min were allowed for fluorescence development before data acquisition on a PerkinElmer Victor3 fluorescence plate reader using a 535/30 nm excitation filter, 595/60 nm emission filter, and an averaging time of 1 sec. To achieve stable readings, samples were scanned 9 times, and the last 5 reads were

used for curve fitting. Rate constants were fit in KaleidaGraph using a standard exponential function, Equation 3-1.

3.3.8 Denaturant melts

An individual sample (1 mL) was prepared for each data point, and all samples were allowed to equilibrate at 25 °C overnight (roughly 18 h). Samples were held in a 4-mL Hellma quartz cuvette during analysis. Circular dichroism was measured in kinetic mode on a Circular Dichroism Spec 410 (AVIV Biomedical) at 226 nm using 60 separate 1-sec reads. The reads for each sample were averaged and normalized by conversion into units of mean residue ellipticity using Equation 3-2.

After circular dichroism analysis, a FluoroMax-3 (Jorbin Yvon Horiba) was used to assay tryptophan fluorescence. Excitation occurred at 280 nm (slit width 2 nm), and emission scans were collected from 295 to 397 nm (slit width 4 nm) in 2 nm increments with 0.5 s of integration time. Each scan was reduced to its center of mass using Equation 3-3.

$$[\Theta] = \frac{\Psi}{10\ell cN}$$

Equation 3-2. Conversion of circular dichroism signal to mean residue ellipticity.

$[\Theta]$ is the mean residue ellipticity, which is a function of circular dichroism signal, Ψ, divided by path length in cm, ℓ, molar protein concentration, c, and the number of amino acid residues in the protein, N.

$$\langle\lambda\rangle = \frac{\sum_{i=1}^{n} F_i \lambda_i}{\sum_{i=1}^{n} F_i}$$

Equation 3-3. Scan center of mass determination.

$\langle\lambda\rangle$ is the scan center of mass (i.e. the fluorescence emission intensity-weighted average scan wavelength). F_i is the fluorescence intensity corresponding to wavelength λ_i [140].

Increasing Gdn•HCl caused a non-linear shift of the scan centers of mass toward longer wavelengths. To simplify curve fitting, the points corresponding to the highest Gdn•HCl concentration (3 M) samples were omitted during data analysis. In KaleidaGraph 4 (Synergy Software), I separately fit the circular dichroism and

tryptophan fluorescence data to a standard denaturant melt equation (Equation 3-4) [141] with variables for the native state baseline, unfolding transition and denatured state baseline.

$$S = \frac{\alpha_N + \beta_N[D] + (\alpha_D + \beta_D[D])e^{\frac{m([D]-c_m)}{RT}}}{1 + e^{\frac{m([D]-c_m)}{RT}}}$$

Equation 3-4. Raw denaturant melt curve fitting.

S is the protein signal, from either circular dichroism or tryptophan fluorescence analysis, as a function of $[D]$, denaturant concentration. α_N and α_D are the signals of the native and denatured states, respectively, of the protein in the absence of denaturant. β_N and β_D are the slopes of the baselines for the native and denatured states, respectively. m is the slope of the unfolding transition. R and T are the ideal gas constant and temperature, respectively. The parameter of interest, c_m, is the denaturant concentration at which the folded and unfolded states of a protein are equally populated at equilibrium.

To overlay the circular dichroism and fluorescence data, the output from fitting the raw denaturant melt data was normalized using Equation 3-5(4), which was generated by inserting the native and denatured baseline equations into the definition of the fraction denatured (Equation 3-5).

$$f_D = \frac{S_N - S}{S_N - S_D} \quad (1)$$

$$S_N = \alpha_N + \beta_N[D] \quad (2)$$

$$S_D = \alpha_D + \beta_D[D] \quad (3)$$

$$f_D = \frac{\alpha_N + \beta_N[D] - S}{\alpha_N + \beta_N[D] - (\alpha_D + \beta_D[D])} \quad (4)$$

Equation 3-5. Normalization of denaturant melt data.

(1) is the definition of the fraction of denatured protein, f_D, in terms of the observed signal, S, and the signal baselines for the native, S_N, and denatured, S_D, states of the protein. (2) and (3) are the equations for the folded and denatured baselines, respectively. Generated by inserting equations (2) and (3) into (1), equation (4) gives the fraction denatured, f_D, as a function of the observed signal. All symbols not defined here are the same as in Equation 3-4.

$$S = \frac{1}{1 + e^{\frac{-m([D]-c_m)}{RT}}}$$

Equation 3-6. Normalized denaturant melt curve fitting.

For normalized denaturant melt data, the fitting equation is simpler because the lower (native) and upper (denatured) baselines are 0 and 1, respectively. The symbols used are the same as in Equation 3-4.

The normalized denaturant melt data was fit in KaleidaGraph using a simplified version (Equation 3-6) of the raw denaturant melt equation (Equation 3-4).

3.3.9 Temperature melts

Samples (1.5 mL) were individually subjected to temperature titrations with 1 °C intervals separated by 1 min equilibration periods. Data collection occurred in a sealed 4-mL Hellma quartz cuvette using a FluoroMax-3 (Jorbin Yvon Horiba) fluorometer with a Wavelength Electronics Model LFI-3751 temperature controller. Excitation occurred at 280 nm (slit width 3.5 nm), and emission scans were collected from 295 to 397 nm (slit width 7 nm) in 2 nm increments with 0.5 s of integration time and converted into scan centers of mass using Equation 3-3.

KaleidaGraph was used for curve fitting to a standard temperature melt equation with variables for the native state baseline, unfolding transition and denatured state baseline (Equation 3-7).

$$S = \frac{\alpha_N + \beta_N T + (\alpha_D + \beta_D T)e^{\frac{m(T-T_m)}{RT}}}{1 + e^{\frac{m([D]-T_m)}{RT}}}$$

Equation 3-7. Raw temperature melt curve fitting.

The constant of interest, T_m, is the melting temperature. All other symbols are the same as those used in Equation 3-4.

To overlay multiple temperature melts, the output from the fitting of the raw temperature melt data was normalized using Equation 3-8(3).

As with the normalized denaturant melt data, the normalized temperature-melt data were fit using a simplified version (Equation 3-9) of the corresponding raw data equation (Equation 3-7).

$$S_N = \alpha_N + \beta_N T \quad (1)$$

$$S_D = \alpha_D + \beta_D T \quad (2)$$

$$f_D = \frac{\alpha_N + \beta_N T - S}{\alpha_N + \beta_N T - (\alpha_D + \beta_D T)} \quad (3)$$

Equation 3-8. Normalization of temperature melt data.

(1) and (2) are the equations for the folded and denatured baselines, respectively. Generated by inserting equations (1) and (2) into Equation 3-5(1), equation (3) gives the fraction denatured, f_D, as a function of the observed signal. All symbols not defined here are the same used in Equation 3-7.

$$S = \frac{1}{1 + e^{\frac{-m(T - T_m)}{RT}}}$$

Equation 3-9. Normalized temperature melt curve fitting.

For normalized temperature melt data, the fitting equation is simplified because the lower (native) and upper (denatured) baselines are 0 and 1, respectively. The symbols used are the same as in Equation 3-7.

3.3.10 Temperature shift assay for metal binding

The effects of zinc concentration on the melting temperatures of the mutant polymerases was used to determine their binding affinities for zinc. To compensate for a general decrease in melting temperature with increasing zinc concentration, an extra linear term was added to the standard hyperbolic equation for ligand binding to yield Equation 3-10.

$$T_m = \frac{A[M]}{K_d + [M]} + m[M] + T_{m,0}$$

Equation 3-10. Apparent dissociation constant from melting temperature shift.

Melting temperature of the polymerase construct, T_m, is a function of the divalent metal (zinc) concentration, $[M]$. KaleidaGraph fit the other terms, which include: A, the amplitude of the maximum increase in melting temperature upon metal binding; K_d, the apparent dissociation constant for the protein+metal complex; m, the slope term compensating for decreasing melting temperature with increasing metal concentration; and $T_{m,0}$, the melting temperature in the absence of divalent metal.

3.4 Results & Discussion

3.4.1 Part A — PHP & Palm

3.4.1.1 Coevolution of PHP metal binding with a Palm motif in DnaE1-pol

In 1998, when Aravind and Koonin [34] discovered the PHP domain using sequence alignments, they noted that *E. coli* and other species of the phylum *Proteobacteria* contained a DnaE1-pol that appeared to lack conserved metal-binding residues. This was later proven for the *E. coli* polymerase by X-ray crystallography [26], which revealed a PHP domain in which two arginine residues (Arg10, Arg203) had structurally replaced metal ions (Figure 1-7).

In the years since the discovery of the PHP domain, the number of sequenced bacterial genomes greatly expanded, and, from these, my colleague, Brian Kelch, aligned a set of roughly 500 DnaE1-pol (sequence selection, Materials & Methods) with less than 90% sequence identity.

Refining the distribution observed by Aravind and Koonin [34], this alignment suggested that loss of PHP metal binding exists mostly in DnaE1-pol homologs from the classes α-, β- and γ-*proteobacteria* (Figure 3-1). In these polymerases, the nine-residue PHP metal-binding motif was very poorly conserved, and, at PHP position-1, an arginine was often present. In *E. coli* DnaE1-pol, this residue (Arg10) replaces the second PHP metal (Figure 1-7).

Based on the presence of an Arg10-equivalent almost exclusively in α-, β- and γ-*proteobacteria,* I hypothesized that their DnaE1-pol homologs might form an Eco-like (similar to *E. coli)* clade that is distinct from the other, non-Eco-like homologs, which are often Taq-like (similar to the *T. aquaticus* polymerase). To detect, quickly and unambiguously, Eco-like residues in the DnaE1-pol homologs from α-, β- and γ-*proteobacteria,* I used the program Two Sample Logo (Materials & Methods) [123, 124], which identifies the most significant residue differences between two groups of sequences that have been aligned together.

In the Fingers and Palm, I observed positions that, like Arg10 versus the canonical, metal-binding histidine residue, are often different in α-, β- and γ-*proteobacteria* (Figure 3-1). Among these, I was particularly interested in a three-residue Palm motif (in *E. coli:* Phe402, Val552, Phe554) because its close proximity to the catalytic aspartates (Asp401, Asp403, Asp555) hinted that it might play a role in Palm distortion (Figure 3-2).

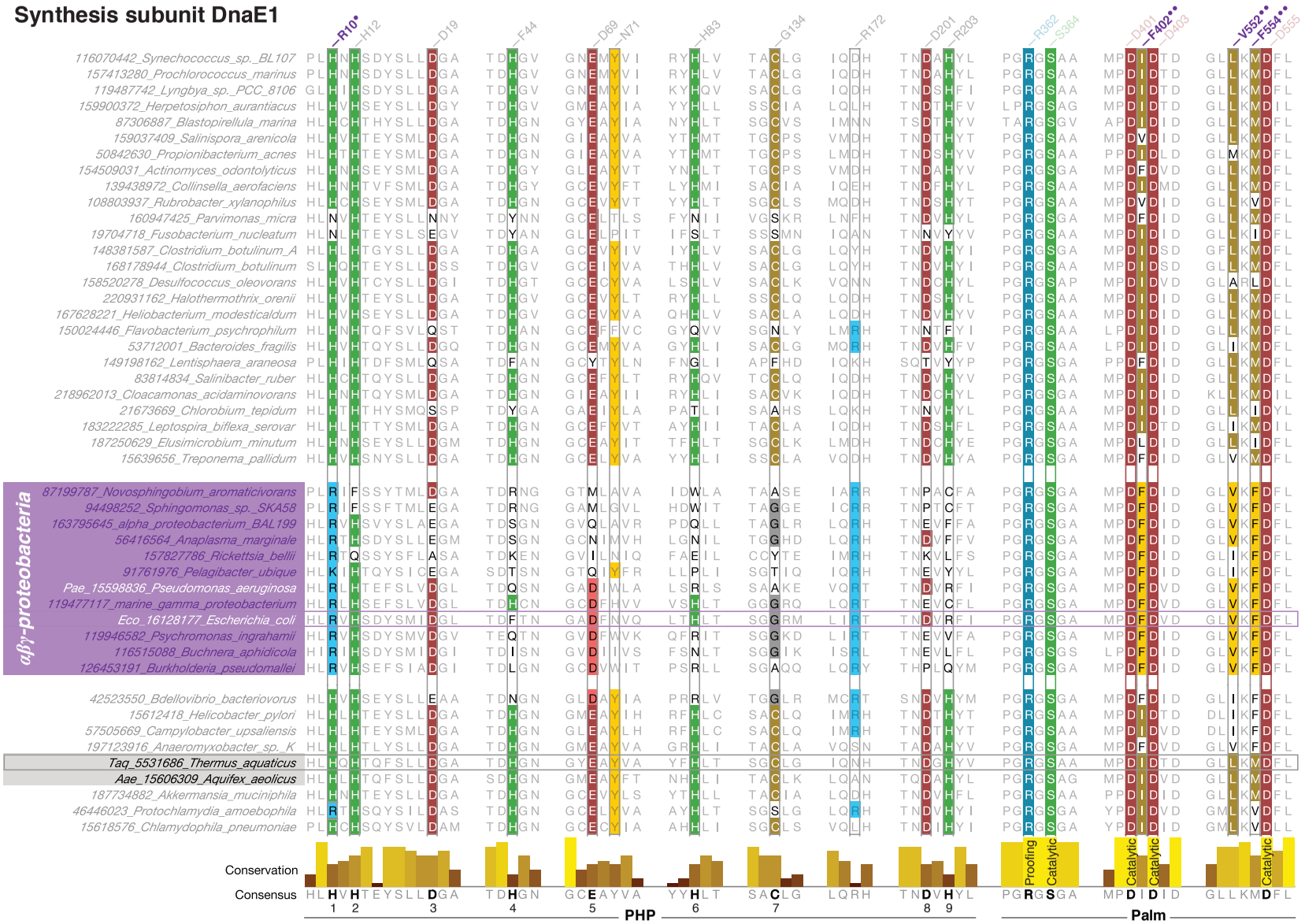

Synthesis subunit DnaE1
R10
H12
D19
F44
D69
N71
H83
G134
R172
D201
R203
R382
S384
D401
F402
D403
V552
F554
D555
αβγ-proteobacteria
87199787_Novosphingobium_aromaticivorans
94498252_Sphingomonas_sp._SKA58
163795645_alpha_proteobacterium_BAL199
56416564_Anaplasma_marginale
157827786_Rickettsia_bellii
91761976_Pelagibacter_ubique
Pae_15598836_Pseudomonas_aeruginosa
119477117_marine_gamma_proteobacterium
Eco_16128177_Escherichia_coli
119946582_Psychromonas_ingrahamii
116515088_Buchnera_aphidicola
126453191_Burkholderia_pseudomallei
42523550_Bdellovibrio_bacteriovorus
15612418_Helicobacter_pylori
57505669_Campylobacter_upsaliensis
197123916_Anaeromyxobacter_sp._K
Taq_5531686_Thermus_aquaticus
Aae_15606309_Aquifex_aeolicus
187734882_Akkermansia_muciniphila
46446023_Protochlamydia_amoebophila
15618576_Chlamydophila_pneumoniae
116070442_Synechococcus_sp._BL107
157413280_Prochlorococcus_marinus
119487742_Lyngbya_sp._PCC_8106
159900372_Herpetosiphon_aurantiacus
87306887_Blastopirellula_marina
159037409_Salinispora_arenicola
50842630_Propionibacterium_acnes
154509031_Actinomyces_odontolyticus
139438972_Collinsella_aerofaciens
108803937_Rubrobacter_xylanophilus
160947425_Parvimonas_micra
19704718_Fusobacterium_nucleatum
148381587_Clostridium_botulinum_A
168178944_Clostridium_botulinum
158520278_Desulfococcus_oleovorans
220931162_Halothermothrix_orenii
167628221_Heliobacterium_modesticaldum
150024446_Flavobacterium_psychrophilum
53712001_Bacteroides_fragilis
149198162_Lentisphaera_araneosa
83814834_Salinibacter_ruber
218962013_Cloacamonas_acidaminovorans
21673669_Chlorobium_tepidum
183222285_Leptospira_biflexa_serovar
187250629_Elusimicrobium_minutum
15639656_Treponema_pallidum
Conservation
Consensus
Proofing
Catalytic
Catalytic
Catalytic
PHP
Palm

Figure 3-1. Three Palm residues in DnaE1-pol coevolved with PHP metal binding.

Alignment of DnaE1-pol sequences, designated by GenInfo Identifier (GI) number [142, 143], genus and species. Column labels at top give the *E. coli* polymerase residues (•Eco-like Arg10 or PHP motif position-1, ••coevolved three-residue Palm motif). At bottom, the nine residues of the metal-binding PHP motif (Table 1-1) are in bold along with five strictly conserved Palm residues, including the catalytic aspartate triad.

3.4.2 Part B — Palm Distortion

3.4.2.1 *E. coli* DnaE1-pol was inactivated by a rotational-sliding motion within its Palm

To better understand the distorted, inactive Palm conformation of *E. coli* DnaE1-pol, my colleague, Meindert Lamers, modeled a Taq-like, active conformation for this polymerase using Modeller [134, 135] (Figure 3-2) and sampled an energy-minimized pathway for a hypothetical inactive-to-active conformational change between the higher resolution EcoE1(917) structure (PDB 2hnh) and the active model using the Yale Morphing Server [136] (Supplemental Movie 1).

Lamers *et al.* [26] observed a "buckling" of the β-sheet that spanned the EcoE1(917) Palm as part of a 20° "twisting apart" of the two aspartate-bearing β-strands. The active model reversed this twisting and brought the three catalytic aspartates into a geometry compatible with the two-metal mechanism for DNA polymerases (Figure 3-2). Along the interface between the aspartate-bearing β-strands, this repositioning was accompanied by a two-residue register-shift and an increase in the number of interstrand hydrogen-bonds from three to five. More generally, morphing between the structure and active model predicted a rotational-sliding motion of the Fingers- and PHP-sides of the Palm relative to each other along the interface between the aspartate-bearing β-strands (Figure 3-2).

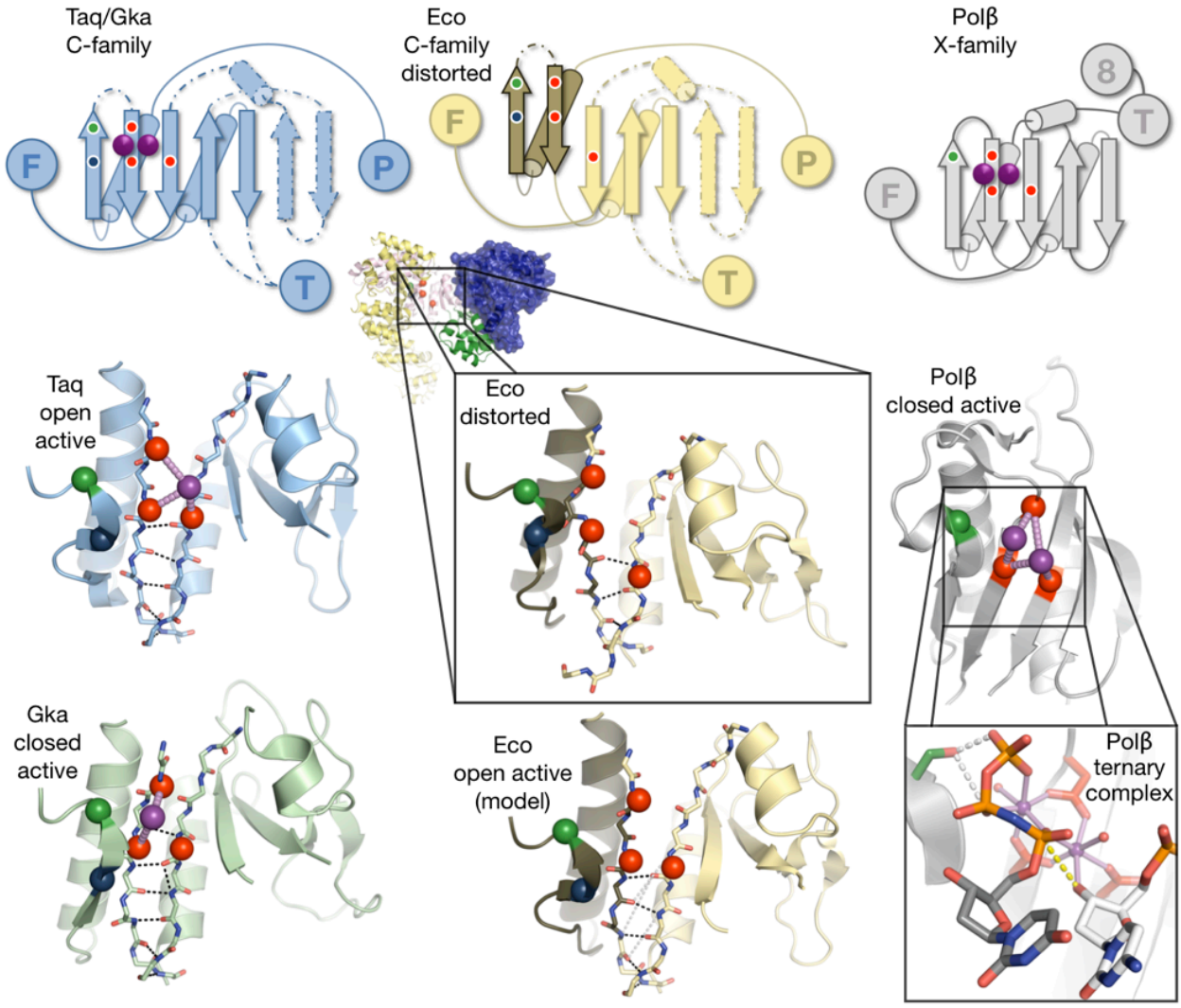


Figure 3-2. *E. coli* DnaE1-pol Palm distortion involves a register shift between the aspartate-bearing strands.

(***a***) Schematic representation of the active C-family polymerase Palm geometry (Taq/Gka), the inactive *E. coli* DnaE1-pol geometry (Eco), and X-family member Polβ in its active geometry. Using the Lamers *et al.* [26] notation, F: Fingers, P: PHP, T: Thumb, 8: 8 kDa domain, Taq: *T. aquaticus* DnaE1-pol, Gka: *Geobacillus kaustophilus* PolC. Circles or spheres: ●serine, ●aspartate, ●arginine, ●metal ion. Despite the other repeated domain names (Fingers, Thumb), the C- and X-family polymerases do not share structural homolog outside their Palms. The first two (Fingers-side) β-strands of the Eco Palm are shifted upward, which disrupts the catalytic metal binding site between the aspartate-bearing β-strands. Dashed lines indicated insertions that distinguish the C-family Palm from the simpler, contiguous X-family Palm. (***b***) Corresponding excerpted structures (PDB 2hpi [24], 2hnh [26], 2fms [105], 3fd2 [25]), a model of *E. coli* DnaE1-pol in a Taq-like active conformation, and the Polβ active site showing both the triphosphate-binding metal (left) and catalytic metal (right), which simultaneously binds all three catalytic aspartate residues. A yellow dashed line (▮) indicates the path of attack on the α-phosphate of the incoming nucleotide (nonhydrolyzable analog dUpnpp) by the 3'-hydroxyl of the DNA primer strand. For the *E. coli* Palm inactive-to-active transition, see Supplemental Movie 1.

3.4.2.2 Roles for Arg362 and Phe402 in proofreading and Palm distortion

Beneath the surface of the Palm, the model for active *E. coli* DnaE1-pol predicts numerous changes to its hydrophobic core, and chief among these is a movement of Phe402 within the Fingers-side of the Palm (Figure 3-3). In addition to being highlighted by Lamers *et al.* [26] as making important contacts in support of the distorted conformation, Phe402 is one of the three residues of the Eco-like Palm motif found in α-, β- and γ-*proteobacteria.* In the *E. coli* DnaE1-pol crystal structures, Phe402 shares prominent T-stacking (edge-face) interactions on the Fingers-side of the Palm with two other phenylalanine residues (Phe341 and Phe391, Figure 3-3b) that are highly conserved among all DnaE1-pol homologs. In the modeled active conformation, however, the favorable T-stacking interactions between the three phenylalanines are missing, which suggests that *E. coli* DnaE1-pol might be biased toward Palm distortion in the absence of its partner proteins and DNA.

In Taq-like DnaE1-pol, Phe402 is replaced by an isoleucine residue, which results in significantly different packing within the Fingers-side of the Palm. While Phe402 lies flat beneath the β-sheet of the Palm, the isoleucine in *T. aquaticus* DnaE1-pol (Taq-Ile464) stands upright, perpendicular beneath the plane of the β-sheet. The vertical orientation of this isoleucine residue allows the β-strand bearing the glycine-serine (GS-motif) to pack more closely against the two-aspartate β-strand than is permitted by the broad repose of Phe402 in *E. coli* DnaE1-pol (Figure 3-3b). A similarly broad conformation for isoleucine is highly unlikely, because, as a β-branched amino acid, it has only a limited set of available rotamers. Indeed, none of the nine backbone independent isoleucine conformations (rotamers, >99.9% cumulative prevalence) in the PyMOL database [73] are comparable to Phe402 in the EcoE1(917) structures.

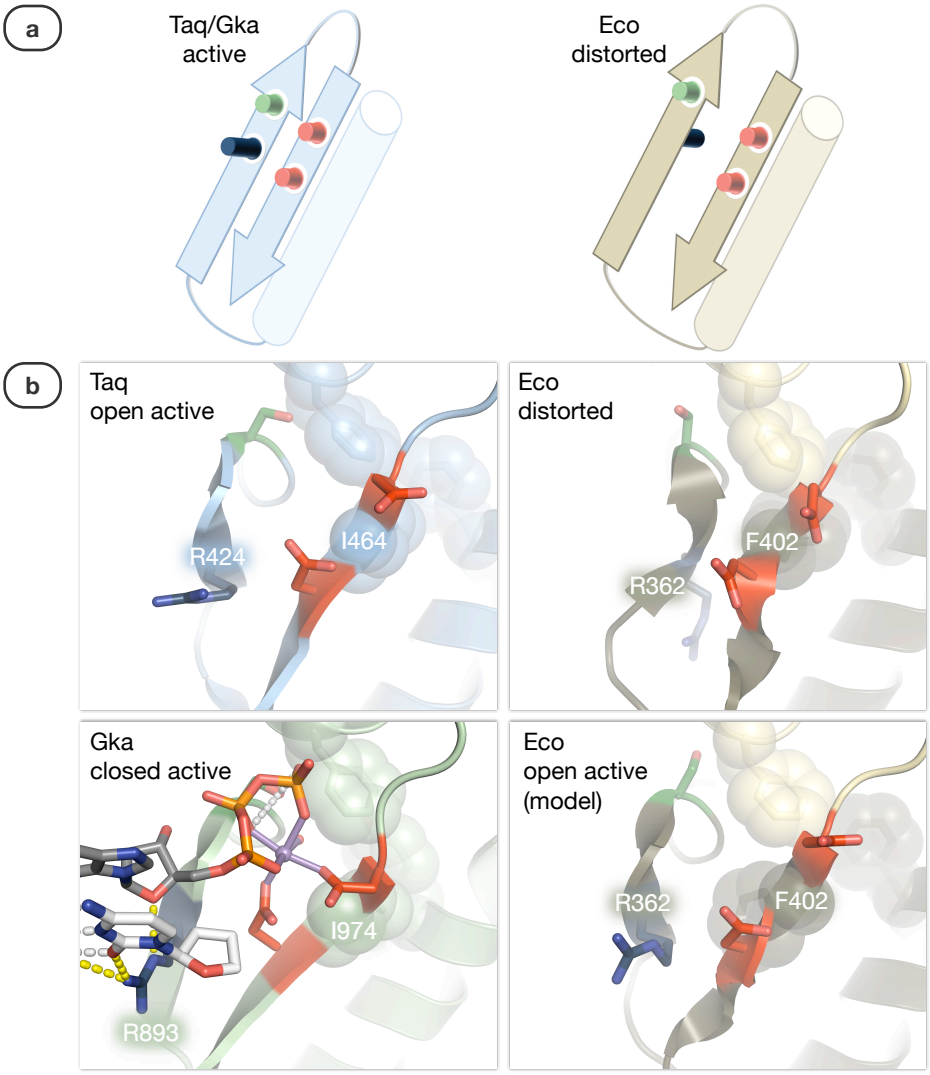


Figure 3-3. Phe402 stabilizes Palm distortion while proofreading Arg362 is buried and cannot bind DNA.

(***a***) Schematics comparing the Fingers-side of the Palms from polymerases *T. aquaticus* DnaE1-pol (Taq) / *G. kaustophilus* PolC (Gka) and *E. coli* DnaE1-pol (Eco). Circles: ●serine, ●aspartate, ●arginine. (***b***) Isolated Fingers-side Palm structures and a Taq-like model for the active conformation of Eco. Residues represented as sticks only are colored as in part (*a*), and hydrophobic residues within the interiors of the polymerases are indicated as both sticks and spheres. In the Gka structure, a portion of the 3'-base of the DNA primer, the incoming nucleotide and the triphosphate-binding metal have been included as points of reference. Two catalytic aspartate residues (Eco-Asp401 and Asp403) are adjacent to Eco-Phe402 and to the corresponding isoleucine residues in Taq and Gka. Gka-Arg893 forms one hydrogen bond each (yellow dashed line, ▌) with the template, primer and incoming nucleotide. See also: Supplemental Movie 2.

The interface between the two β-strands of the Fingers-side of the Palm and the rotation of Phe402 are of particular importance for Arg362, which is strictly conserved by C-family polymerases. In the *Geobacillus kaustophilus* PolC structures [25], the corresponding arginine makes hydrogen bonds with the base at the 3'-end of the DNA

primer strand, its partner in the template (Figure 3-3b), and also the ribose oxygen of the incoming nucleotide. In the X-family polymerase Polλ, a similarly located arginine (human Arg517, PDB 1xsn) hydrogen bonds to the corresponding template base and is crucial for fidelity [144, 145]. I, therefore, suspect that the Arg362-equivalent in C-family polymerases plays a key role in fidelity by proofreading for proper base pairing. In *T. aquaticus* DnaE1-pol and *G. kaustophilus* PolC, the arginine is surface exposed (Figure 3-3b) and supported by hydrogen bonds with two other residues that are also conserved, a tyrosine (Taq-Tyr821, Eco-Tyr764) and a non-metal-binding aspartate (Taq-Asp467, Eco-Asp405, asparagine in PolC homologs). In the structures of *E. coli* DnaE1-pol, however, Arg362 is buried deep within the Palm and unavailable to hydrogen bond with DNA (Figure 3-3b). In the active *E. coli* DnaE1-pol model, Arg362 occupies the expected position for proofreading, and its former burial site is filled by Phe402 (Figure 3-3b).

The repositioning of Phe402 and Arg362 in the active model suggests a possible connection between proofreading and Palm distortion in Eco-like DnaE1-pol. By extension, the formation of hydrogen bonds by Arg362 to correctly base-paired DNA in the active Palm conformation might counteract T-stacking by Phe402 with its partner phenylalanines in the inactive, distorted conformation. So long as a correct base-pair were present, the Arg362 interactions could prevail, and the polymerase would remain active, synthesizing DNA at a fast rate. If the polymerase were to add an incorrect base (or if no DNA were present, like in the structures), T-stacking by Phe402 would be unopposed, and the polymerase could enter the distorted conformation. In addition to halting polymerization, the loss of hydrogen bonds to Arg362 would likely decrease the affinity of the Palm for DNA, which could either dissociate or melt into single strands and be edited by the DnaQ-exonuclease subunit. After mismatch removal, correct base pairing could once again occur, and the polymerase would reset into the active conformation and resume synthesis upon DNA binding by Arg362.

3.4.2.3 Val552 and Phe554 form a smooth surface that facilitates Palm distortion

In contrast with the many Palm residues repositioned in the model for active *E. coli* DnaE1-pol, Phe554 stands out for its near total lack of movement. Identified both by Lamers *et al.* [26] as an important stabilizer of the inactive conformation and by sequence alignment as an Eco-like member of the coevolved Palm motif (Figure 3-1), Phe554 maintains its position beneath the catalytic aspartate triad and stays in contact with Val552 (also of the Palm motif). By packing closely together on the PHP-side of the Palm, the Val552:Phe554 pair forms a smooth surface along which the Fingers-side of the Palm slides in the inactive-to-active morph (Figure 3-4).

Both the valine and phenylalanine are limited to a small number of possible conformational rotamers within the Palm of *E. coli* DnaE1-pol. Like isoleucine (see previous section), valine is a β-branched amino acid, and none of the rotamers in the PyMOL library (not shown) would allow Val552 to break contact with Phe554. Although phenylalanine generally has more conformational freedom than valine, when generated

in PyMOL, the Phe554 rotamers appeared either to pack closely against Val552 or to suffer massive steric clashes with other Palm residues (not shown).

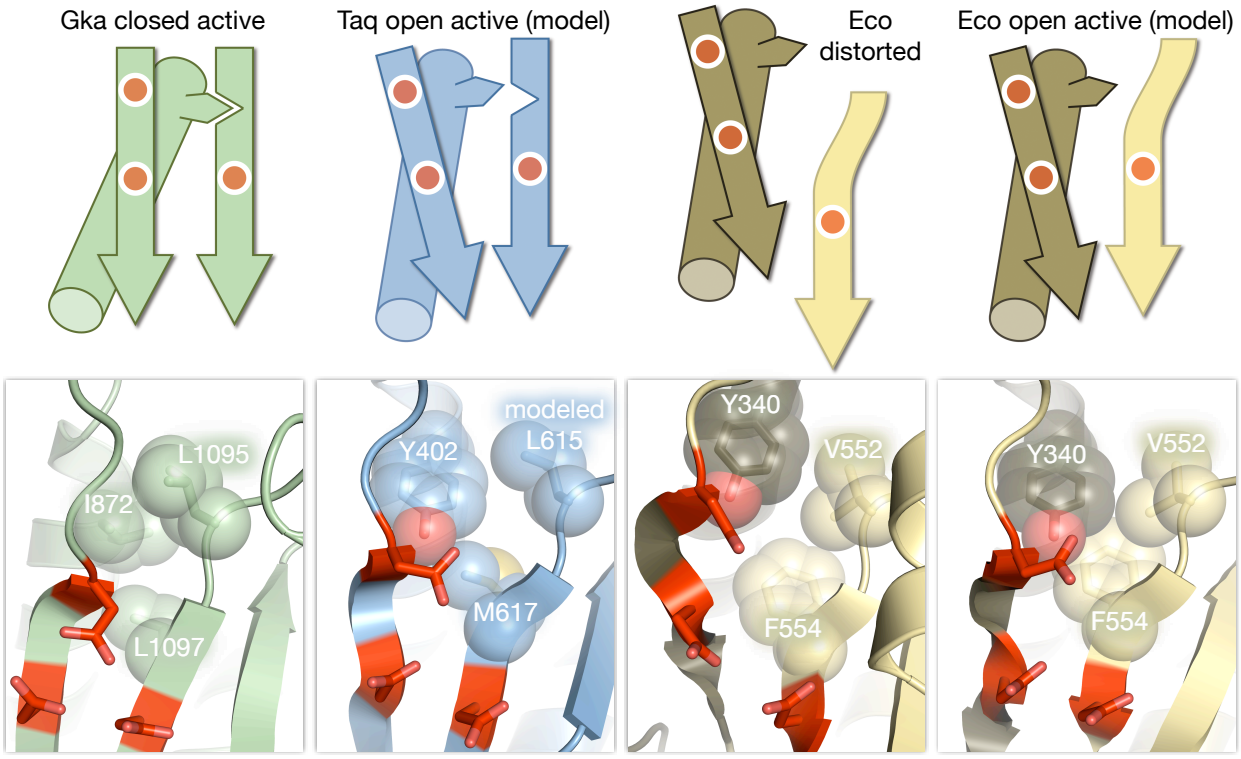


Figure 3-4. A smooth surface in *E. coli* DnaE1-pol blocks stabilizing cross-Palm intercalation.

Schematics, structures and models of a Palm region from three C-family polymerases (Gka: *G. kaustophilus* PolC structure (PDB 3f2d), Taq: a Leu615 rotamer model of *T. aquaticus* DnaE1-pol (generated in PyMOL from PDB 2hpi), Eco: *E. coli* DnaE1-pol structure (PDB 2hnh) and model). The region shown here is similar to the Polβ ternary complex at the lower right of Figure 3-2b. In the schematics, the three catalytically important aspartate residues are shown as red circles. In the structures and models, the aspartates are red sticks, and three buried hydrophobic residues are sticks with transparent spheres. See also: Supplemental Movie 3.

On the basis of the above modeling, I conclude that the smooth surface generated by Val552:Phe554 is likely a permanent feature of the PHP-side of the Palm in *E. coli* DnaE1-pol and, also, that this surface may be an important factor in Palm distortion. Since the Palm motifs of PolC and Taq-like DnaE1-pol homologs contain a leucine in place of Val552 (Taq-Leu615:Met617, Gka-Leu1095:Leu1097), I speculated that the conformational flexibility of this leucine might help explain why Palm distortion was not observed for *T. aquaticus* DnaE1-pol in the absence of DNA.

Unlike Val552:Phe554, the pair of leucine residues on the PHP-side of the Palm in *G. kaustophilus* PolC do not form a smooth surface. Instead, they are spread apart, and an isoleucine from the Fingers-side (Gka-Ile872, also conserved as leucine and methionine in PolC homologs) has inserted between them (Figure 3-4). This cross-Palm

intercalation may protect the polymerase from distortion like that observed for *E. coli* DnaE1-pol. Along with an additional (sixth) cross-Palm hydrogen bond (Figure 3-2), the cross-Palm intercalation also brings the aspartate triad into an even tighter arrangement than in the *T. aquaticus* DnaE1-pol structures (Figure 3-4). I suspect that this tight arrangement of the aspartate triad may be a general feature of the "closed" conformation for C-family polymerases in ternary complex with DNA and incoming nucleotide.

The available ternary structure of *T. aquaticus* DnaE1-pol [35] lacks sufficient resolution to assay cross-Palm intercalation, and, based on comparison to *G. kaustophilus* PolC, Evans *et al.* suggested that this DnaE1-pol structure is unlikely to represent a fully closed conformation. For these reasons, I chose to analyze the high resolution, DNA-free structure (PDB 2hpi) of *T. aquaticus* DnaE1-pol for compatibility with cross-Palm intercalation.

As evidenced by the simultaneous binding of a magnesium ion to all three catalytic aspartates (Figure 3-2), the high-resolution structure of *T. aquaticus* DnaE1-pol (PDB 2hpi) represents an active conformation, but, in comparison with the available ternary structure of this polymerase [35], it also represents an "open" conformation. If the closed conformation of *T. aquaticus* DnaE1-pol were similar to that of *G. kaustophilus* PolC, then a tyrosine on the Fingers-side of its Palm (Taq-Tyr402, conserved across DnaE1-pol homologs) would suffer a steric clash with Taq-Leu615 of the PHP-side pair. However replacing this clashing leucine, in PyMOL, with the most common leucine rotamer (62.5% prevalence) spreads apart the PHP-side pair (Taq-Leu615:Met617, Figure 3-4) and opens a gap for cross-Palm intercalation by the tyrosine. In this way, the leucine residue corresponding to Val552 in the Palm motifs of both *G. kaustophilus* PolC and *T. aquaticus* DnaE1-pol may serve to protect these polymerases from inactivation by an Eco-like Palm distortion.

Returning to *E. coli* DnaE1-pol, although it has a Fingers-side tyrosine (Tyr340), this residue appears unable to intercalate into the PHP-side of Palm because it is blocked by Val552, both in the crystal structures and in the model for a Taq-like open-active conformation (Figure 3-4). Although I have not attempted to model a ternary complex of this polymerase, I expect this state to resemble that of *G. kaustophilus* PolC. In such a closed conformation, the Val552:Phe554 pair will likely maintain its smooth surface on the PHP-side of the Palm, thus continuing to block intercalation by the Fingers-side tyrosine. This would leave the *E. coli* polymerase susceptible to distortion, perhaps in response to proofreading of a mismatch by Arg362, as described in the previous section.

3.4.3 Part C — DnaQ-exonuclease

3.4.3.1 Coevolution of DNA synthesis and editing activities within the pol III Core

Having identified apparent coevolution between loss of PHP metal binding and a predisposition toward Palm distortion in Eco-like polymerases, I turned my attention to homologs of DnaQ-exonuclease, which is the editing subunit of the *E. coli* pol III Core. In this subunit, exonuclease activity resides in the globular, N-terminal domain [146], and, preceded by a flexible linker [147], the C-terminal region binds to the PHP domain of DnaE1-pol [61, 62].

To gauge the likelihood that Eco- and Taq-like pol III Cores contain a DnaQ-exonuclease, I gathered those proteins homologous to the *E. coli* editing subunit from the species represented in a subset of the DnaE1-pol alignment. This was complicated by the presence of DEDDh 3'-5' exonuclease domains [148, 149] similar to DnaQ-exonuclease in other proteins such as REX, RNase T, PolC, a UvrC-like excinuclease and DinG- or UvrD-like helicases. To avoid these, I limited my selection to proteins in which the exonuclease domain was the only region functionally annotated in the National Center for Biotechnology Information (NCBI) Protein database [143]. At least one DnaQ-exonuclease meeting my criteria was found in each member of α- β- and γ-*proteobacteria,* but outside this clade, several bacterial species contained no such homolog and may, therefore, depend on DNA editing activity from a Taq-like DnaE1-pol or some other, unidentified, source.

Aligning the DnaQ-exonuclease sequences (Figure 3-5), I noticed strong conservation of an Eco-like C-terminal tail in one DnaQ-exonuclease homolog from each member of α- β- and γ-*proteobacteria* and no such tail and generally poor conservation of the N-terminal exonuclease domain of the homologs from other species. These findings suggest that a DnaQ-exonuclease providing *in trans* DNA editing is almost exclusively found in conjunction with an Eco-like DnaE1-pol homolog.

Editing subunit DnaQ

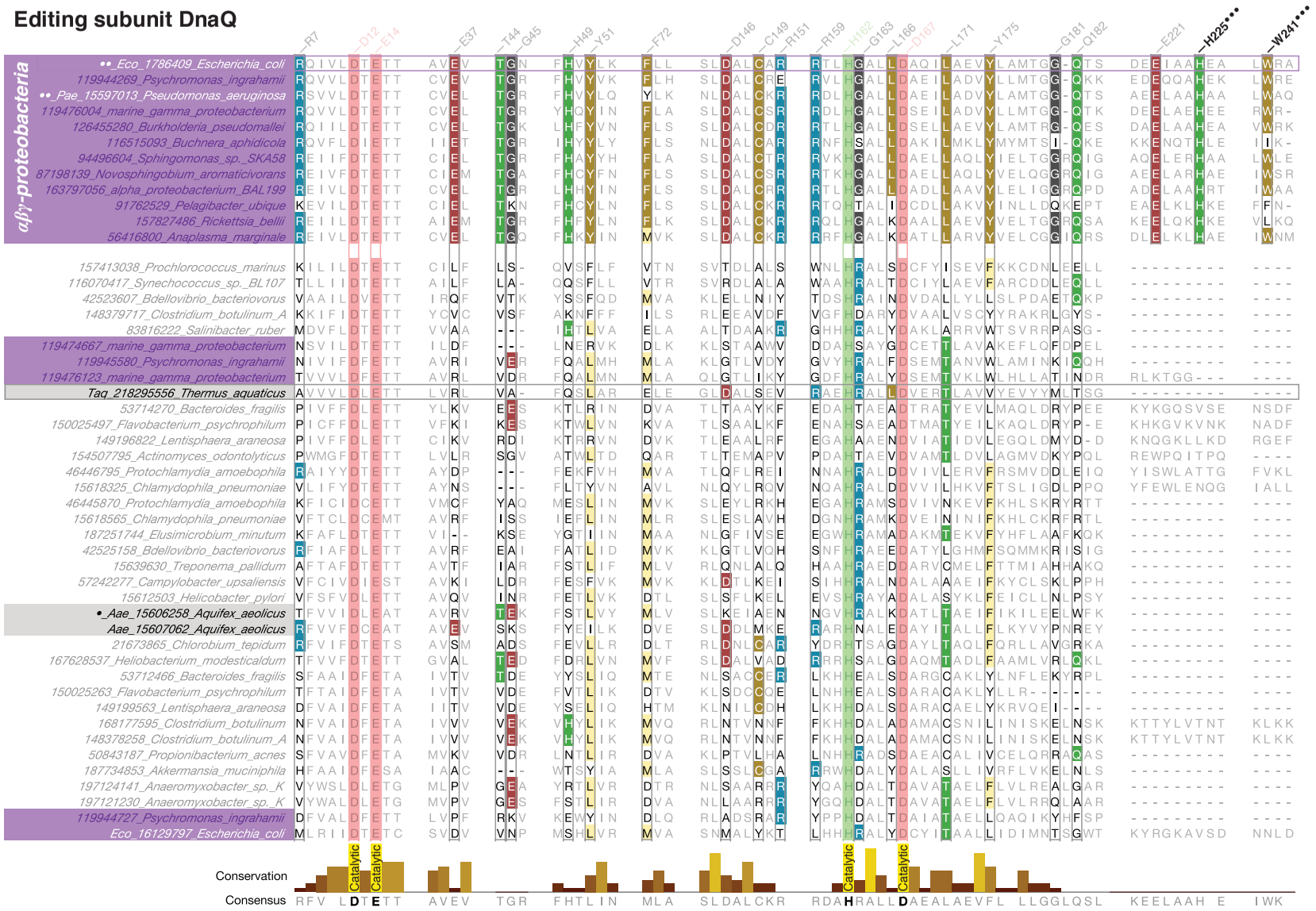


Figure 3-5. α-, β- and γ-*proteobacteria* conserved a DnaQ-exonuclease.

Sequence alignment of DnaQ-exonucleases (three with biochemically determined DnaE1-pol binding affinity: •low [51], ••high [14, 50]) from the bacteria represented in the DnaE1-pol alignment (Figure 3-1). Column labels at top give the *E. coli* DnaQ-exonuclease residues (•••two C-tail residues (His225 [150], Trp241 [151]) biochemically shown to be important for DnaE1-pol binding.) At bottom, four of the catalytic residues common to the DEDDh superfamily [148, 149] are in bold.

In addition to the *E. coli* subunits, a stable polymerase+exonuclease (DnaE1+DnaQ) complex was observed for another γ-proteobacterium, *Pseudomonas aeruginosa* [50], which has a DnaQ-exonuclease with a C-terminal tail (Figure 3-5) that conserves three residues commonly found in the editing subunits from of α- β- and γ-*proteobacteria.* Of these residues, His225 [150] and Trp241 [151] were shown to be important for DnaE1-pol binding in *E. coli.* More distantly related, the two DnaQ-exonuclease homologs from *Aquifex aeolicus* lack C-terminal tails, and the one that was tested biochemically showed low affinity [51] for the Taq-like DnaE1-pol (Figure 3-1) from this species.

My DnaQ-exonuclease sequence observations are also consistent with analysis performed earlier by Daniel H. Haft at the J. Craig Venter Institute (JCVI). Using the protein homology search tool HHpred [109, 152, 153], I compared the sequence of *E. coli* DnaQ-exonuclease to roughly 120,000 multiple sequence alignments represented in condensed form as Hidden Markov Model protein signatures. Among these, the TIGRFAMs [154, 155] entry TIGR01406 (dnaQ_proteo) [156], generated by Haft, was the highest scoring and possessed roughly the same features as the Eco-like DnaQ-exonuclease homologs in my alignment, including a C-terminal tail with residues implicated in DnaE1-pol binding. In defining TIGR01406, Haft mentioned that this signature identified the pol III editing subunits from most *Proteobacteria* while excluding other homologous proteins. Note, however, that the NCBI Conserved Domains Database [157] includes an outdated version of TIGR01406 [158], so dnaQ_proteo annotation (and annotation based on other insufficiently restrictive signatures such as cd06130 (DNA_pol_III_epsilon_like) [159] and cd06131 (DNA_pol_III_epsilon_Ecoli_like) [160]) is not necessarily indicative of an Eco-like DnaQ-exonuclease.

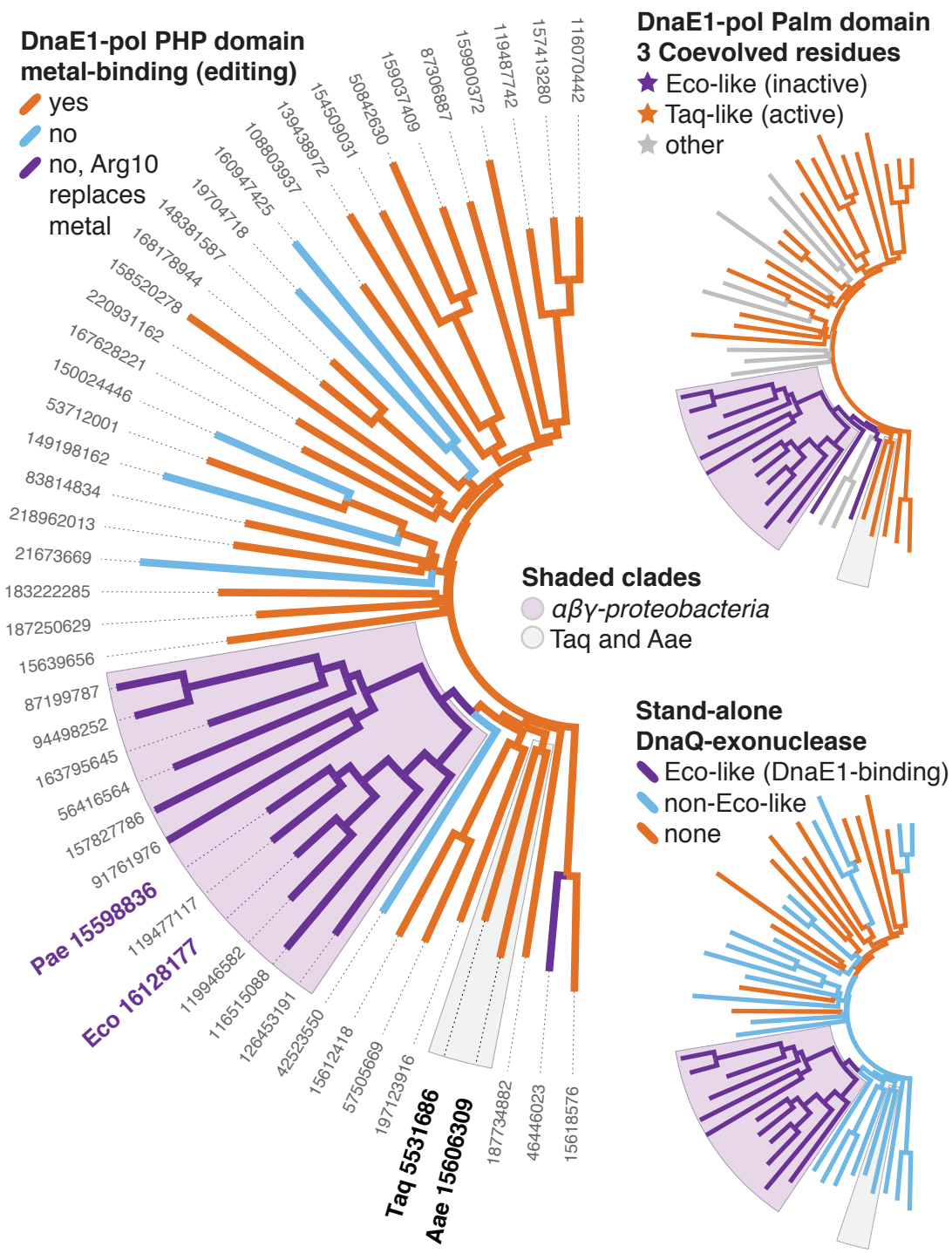


Figure 3-6. A DnaQ-exonuclease coevolved with Eco-like DnaE1-pol.

A tree constructed from my DnaE1-pol alignment (Figure 3-1) is shown three times with different node (line) coloring-schemes. A GenInfo Identifier (GI) number [142, 143] labels each sequence, and abbreviations indicate four species (Pae: *P. aeruginosa,* Eco: *E. coli,* Taq: *T. aquaticus,* Aae, *A. aeolicus).* Two clades, corresponding to (1) *α-, β-* and *γ-proteobacteria* and (2) Taq and Aae, are shaded consistently between the three versions of the tree.

Considering the features of the DnaE1-pol and DnaQ-exonuclease homologs, two complementary themes emerge (Figure 3-6). Firstly, a strong correlation exists between the Eco-like motif that appears to facilitate Palm distortion, loss of PHP metal binding, and the presence of an Eco-like DnaQ-exonuclease. This suggests that the Eco-like homologs might conserve an inactivation mechanism linked to DNA editing. Secondly, there is an even stronger anti-correlation between the existence of an Eco-like DnaQ-exonuclease and PHP metal binding. Taken together, these trends suggest that there are two widespread versions of DnaE1-based pol III. One subtype, as found in *T. aquaticus,* contains a Core comprised of only a DnaE1-pol homolog, and this single protein contains both DNA synthesis and editing activities *in cis.* The other subtype, which exists in *E. coli,* contains a DnaE1-pol whose Palm can be distorted and whose PHP lost metal binding and gained a tightly bound DnaQ-exonuclease to provide editing activity *in trans.*

3.4.3.2 Eco-like DnaE1-pol evolved from a Taq-like ancestor over 1,500 million years ago

The taxonomic distribution of Eco- and Taq-like DnaE1-pol subtypes allows me hypothesize their order of evolution and relative ages (Figure 3-7). Because Eco-like pol III is only found in α-, β- and γ-*proteobacteria,* this subtype is likely derived from the Taq-like version, which is found in nearly all other species that contain a DnaE1-pol. Furthermore, PHP domains generally feature metal binding [34], as is the case for the Taq-like, but not the Eco-like, polymerase.

Because modern cyanobacteria contain a Taq-like DnaE1-pol, their ancestors likely did as well. Since the earliest evidence for cyanobacteria (stromatolites and microfossils) dates back roughly 3,500 million years [161-165], Taq-like DnaE1-pol is almost certainly at least as ancient.

For the Eco-like pol III Core, a minimum age can be inferred from the evolution of eukaryotic mitochondria, which are descended from an α-proteobacterial ancestor [166-169]. Like current α-, β- and γ-*proteobacteria* (Figure 3-6), the forerunner of present-day mitochondria probably contained an Eco-like pol III Core, whose minimum age would then be 1,500 million years, based on the oldest fossils [170] widely accepted as eukaryotes [171]. However, since structures corresponding to mitochondria were not identified in these specimens, a dissenting argument could possibly be made that these were evidence of an intermediate form and not a true eukaryote. In that case, the earliest fossil evidence for plants, 1,200 million years before the present [172], gives a highly conservative minimum age for the Eco-like pol III subtype.

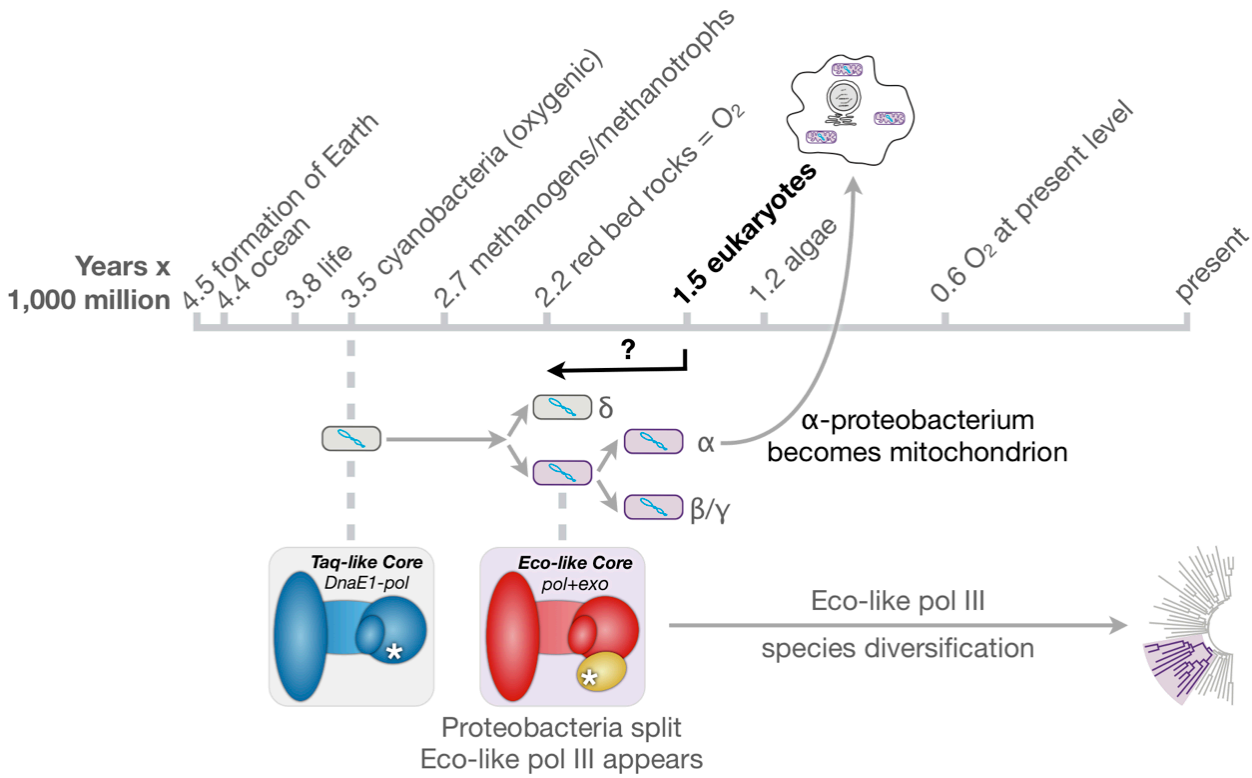


Figure 3-7. The Eco-like pol III Core evolved from a Taq-like ancestor more than 1,500 million years ago.

Timeline [63] for DnaE1-based pol III evolution relative to the history of life on Earth [63, 102]. Current cyanobacteria feature a Taq-like DnaE1-pol, and *α-proteobacteria* have a pol III Core containing an Eco-like DnaE1-pol and DnaQ-exonuclease (Figure 3-6).

In addition to the Eco- and Taq-like DnaE1-pol homologs, my sequence analysis also revealed several apparently rare versions of the pol III Core. These include PolC homologs (within the classes *Clostridia* and *Negativicutes* of the phylum *Firmicutes*) that, similar to the Taq-like Core, lack a DnaQ-exonuclease domain and might therefore use a PHP domain for editing. I also found examples of DnaE1-pol with an N-terminal DnaQ-exonuclease fusion (within the classes *Flavobacteria* and *Shingobacteria* of the phylum *Bacteroidetes*) that appear likely perform editing in these polymerases, most of which lack PHP metal binding. In cyanobacteria, DnaE1-pol often appears as a split-intein [39], and, although I did not include these polymerases in my analysis, other researchers have reported that these combine to form a standard DnaE1-pol. Also, there are several examples of non-Eco-like DnaE1-pol that have lost PHP metal binding (Figure 3-6). Since these species do not contain an Eco-like DnaQ-exonuclease, it is unclear how or if error removal is performed by pol III in these species.

3.4.4 Part D — Allostery & Bioinformatics

3.4.4.1 Coupled and conserved residues span the PHP and Palm of DnaE1-pol

Allostery is the process by which an event occurring at one site in a protein causes an effect elsewhere in that protein [173]. Using Statistical Coupling Analysis, which is all-against-all check for sequence correlations in a group of homologous proteins, researchers have identified paths of intramolecular allosteric communication through a variety of proteins [125, 126, 174-178]. Brian Kelch applied this technique using the Gerstein method [127], which differs from that of its developers, Ranganathan and colleagues, in that the results from multiple residue-pair perturbations are combined into one score for each pair of residue positions (a site-pair).

By far, the highest score in the Statistical Coupling Analysis (1.81) belonged to a pair that included the phenylalanine residue involved in burying the proofreading arginine (Figure 3-3) in *E. coli* DnaE1-pol (Phe402:Gly365), and a pair containing the other two residues of the Eco-like Palm motif (Val552:Phe554) likewise had an extraordinarily high score (1.36). Also consistent with the analysis described above, but no less remarkable, was the high coupling of PHP metal-binding motif positions, at least one of which was present in 15 of the top 50 highest scoring pairs.

Plotting a selection of the highest scoring site-pairs as a heat map (Figure 3-8) showed a high density of coevolving residues in the PHP and Palm domains. A lower density also appeared in the Fingers along with two sites in the oligonucleotide-binding domain. There were essentially no coupled residues in the C-terminal portion of the Fingers (referred to as the β-binding domain by Bailey *et al.* [24] and duplex-binding domain by Evans *et al.* [25]) or the remainder of the polymerase C-terminal tail. On the whole, strongly coupled site-pairs appeared in all regions within DnaE1-pol for which there was a relatively high degree of conservation among the homologs in the alignment.

To identify those coupled residues most likely to participate in allostery, a hierarchical clustering algorithm [125] was employed that groups coupled positions that exhibit similar scores and patterns. This analysis revealed multiple clusters whose members included residues in the PHP cleft, in the Palm near the catalytic aspartates, and at several points within the Fingers (Figure 3-9a). Combining the two strongest clusters of coupled residues with those residues conserved with greater than 96% identity, a path appears that stretches for roughly 40 Å between the aspartate triad in the Palm and Arg10 in the PHP cleft (Figure 3-9).

Because the Statistical Coupling Analysis was based on both Taq- and Eco-like DnaE1-pol homologs, the path between the PHP and Palm might indicate a general capacity for interdomain allostery in DnaE1-pol homologs. In a Taq-like DnaE1-pol, allosteric communication between the Palm and PHP could be analogous to movements occurring within other polymerases upon switching of DNA between their synthesis and editing active sites (e.g. the Klenow fragment of *E. coli* Pol I [179-181]). Despite the lack of a metal-binding PHP domain, interdomain allostery might serve a

similar function in an Eco-like DnaE1-pol homolog, because its PHP domain is probably the binding site for a editing DnaQ-exonuclease subunit [61].

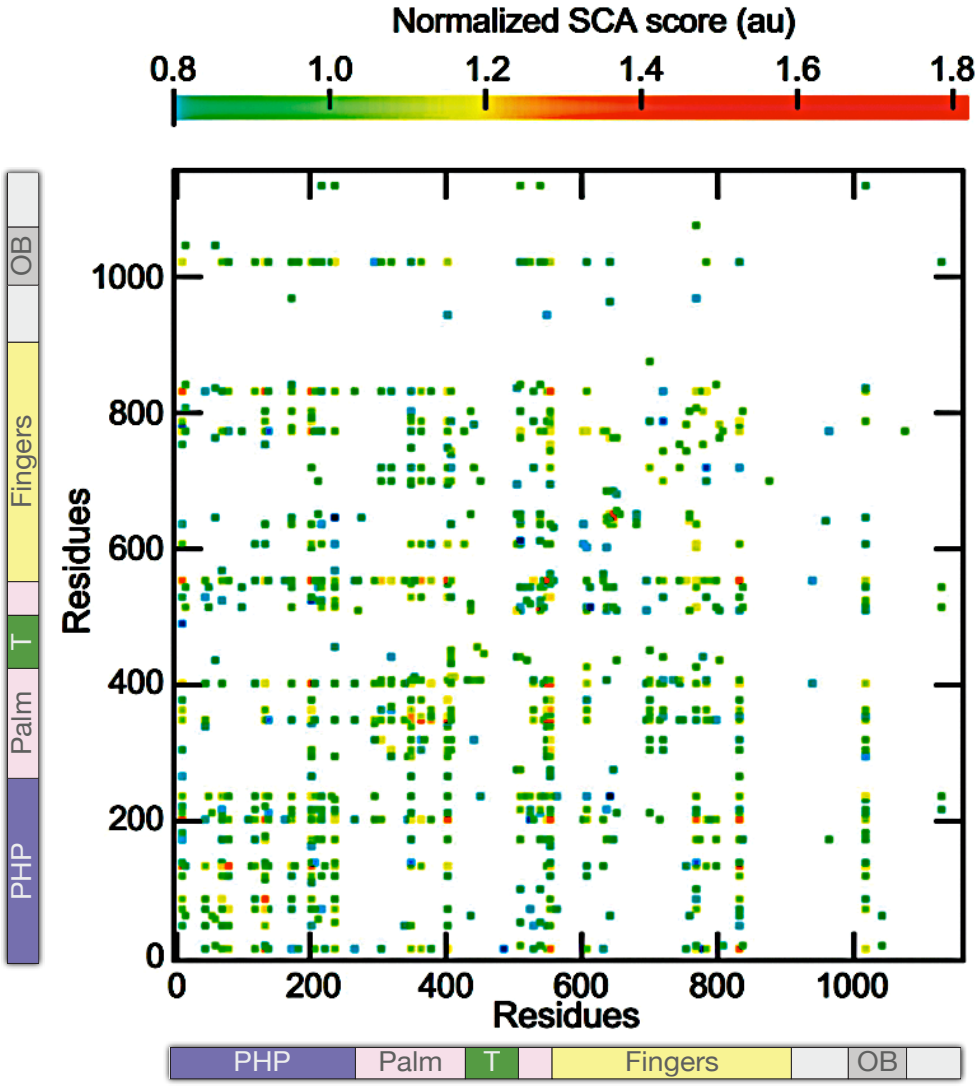


Figure 3-8. Evolutionarily coupled residues occur in the principle domains of DnaE1-pol.

Heat map showing site-pairs with Statistical Coupling Analysis scores in the top 55% of the observed range. At left and bottom, the polymerase domains are indicated as in Figure 1-4 (T: Thumb, OB: oligonucleotide binding). Figure adapted from Brian Kelch and used with permission.

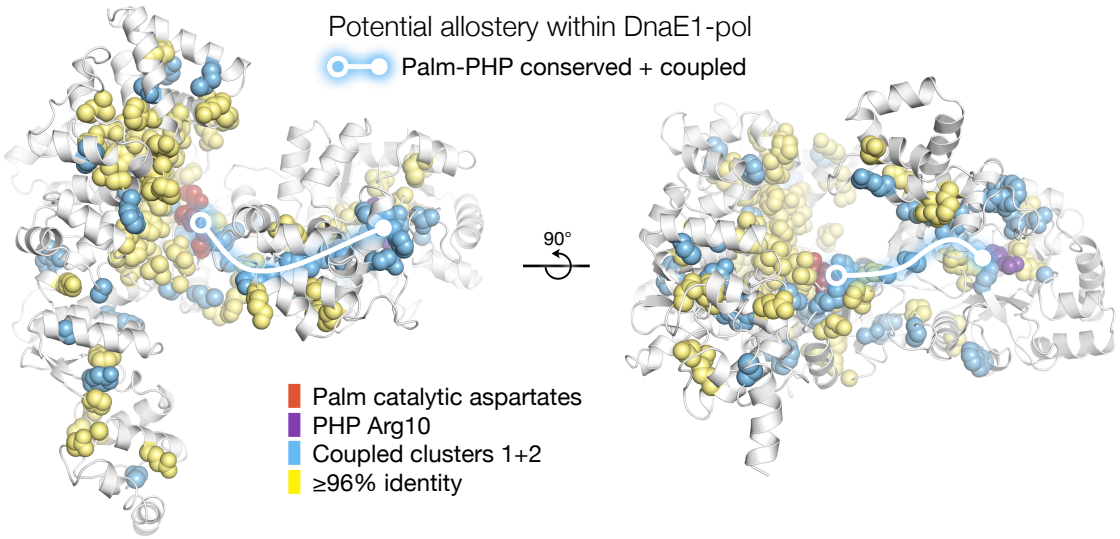


Figure 3-9. Statistical Coupling Analysis suggests PHP-Palm allostery in DnaE1-pol.

Conserved (≥96% identity, yellow spheres) and coupled residues (Statistical Coupling Analysis clusters 1 and 2, blue spheres) form a path across DnaE1-pol from the catalytic aspartates in the Palm (red spheres) to Eco-like residue Arg10 (purple spheres) in the PHP (structure: PDB 2hnh). See also: Supplemental Movies 4 and 5.

3.4.4.2 Conservation within Eco-like DnaQ-exonuclease dominates coupling to DnaE1-pol

To extend Statistical Coupling Analysis for interprotein studies, the DnaE1-pol and DnaQ-exonuclease were combined to yield a hybrid sequence for a total of approximately 250 species. The hybrids included both Eco-like homologs, in which the two proteins may bind each other tightly [14, 50] and also Taq-like homologs, for which there is scant biochemical data in support of polymerase-exonuclease (DnaE1-DnaQ) binding [51] and poor DnaQ-exonuclease conservation (Figure 3-5).

In addition to recapitulating coupling within DnaE1-pol, high scores resulted for positions within DnaQ-exonuclease, and, in stunning fashion, the highly coupled positions within each subunit also coupled strongly between the two (Figure 3-10a). Applying the hierarchical clustering algorithm, the positions identified in the polymerase (not shown) were largely the same as for the analysis of this subunit alone. In DnaQ-exonuclease, the clustered positions were abundant throughout the protein, with the exception of a small region distal to the enzyme active site (Figure 3-10a).

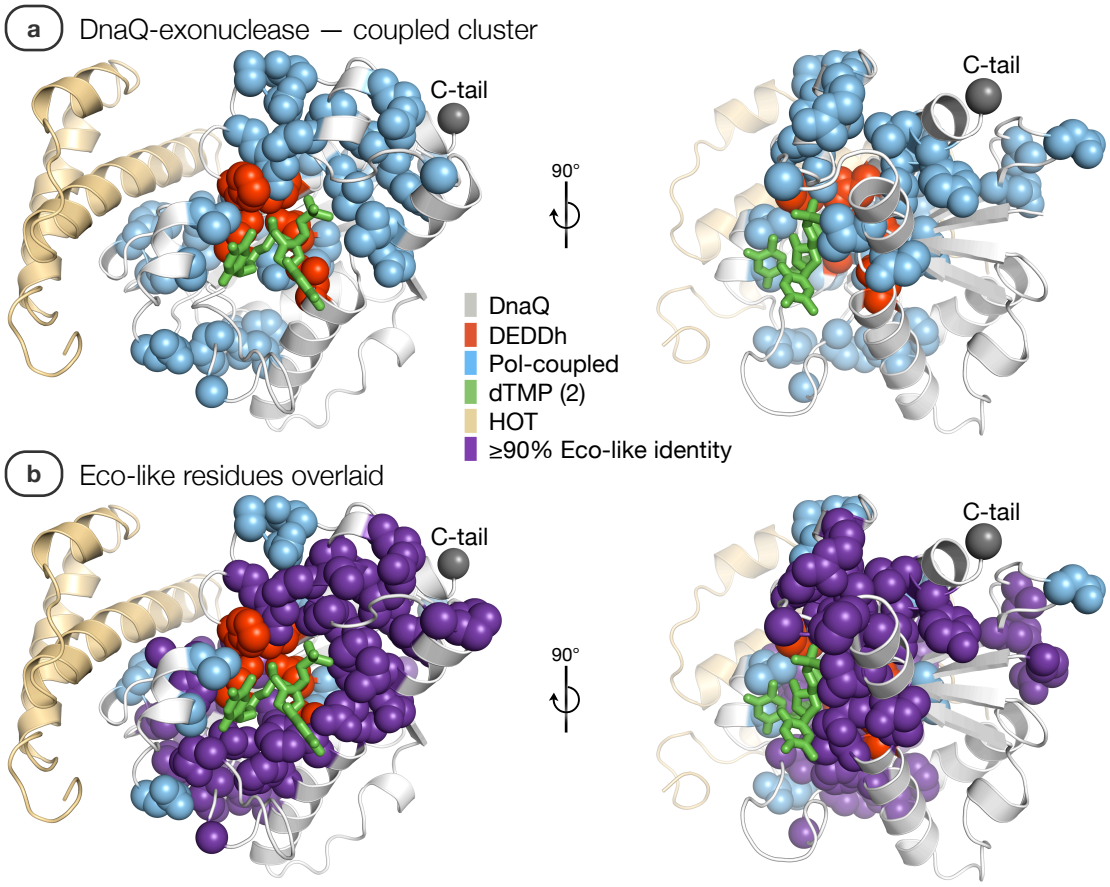


Figure 3-10. Intermolecular coupling analysis identified residues conserved by Eco-like DnaQ-exonuclease.

Structure of *E. coli* DnaQ-exonuclease in complex with HOT [182], the bacteriophage homolog of the θ subunit of pol III (PDB 1j53: DnaQ, 2ido(B): HOT, dTMP: monophosphate product of editing reaction, DEDDh: the five residue active-site motif [149]). The DnaQ-exonuclease structure contains residues 7-180 and is therefore missing the 63-residue C-tail that binds DnaE1-pol. (***a***) Mapping the high-scoring cluster (blue spheres) from the polymerase-exonuclease intermolecular Statistical Coupling Analysis onto DnaQ-exonuclease shows a high density of positions near the enzyme active site. (***b***) Overlaying the residues (purple spheres) conserved by ≥90% of a set of 72 Eco-like DnaE1-pol homologs with less than 70% global sequence identity shows that high conservation among these homologs accounts for most of the positions in the coevolved cluster.

To better interpret the significance of the DnaQ-exonuclease coupling data, I aligned a set of 72 Eco-like homologs with less than 70% global identity. Even within this highly diverse set, there was considerable sequence conservation, and comparison of those residues with 90% or higher identity to the positions from the coupling analysis revealed that the clustered positions were often highly conserved within the Eco-like homologs (Figure 3-10b). These conserved residues likely contribute to a variety of functions, potentially including allostery with DnaE1-pol. While some DnaQ-exonuclease surface residues probably contact the polymerase, others bind DNA for editing, and additional residues might be involved in allosteric

communication between these functions. In the absence of a polymerase+exonuclease (DnaE1+DnaQ) co-crystal structure, however, the participants in these functions remain a tantalizingly unknown, as does the mechanism by which DnaQ-exonuclease increases the length of the DNA fragments processively synthesized by the pol III holoenzyme [107].

3.4.4.3 DnaE1-pol and the DNA sliding clamp are evolutionarily coupled

Another pol III subunit that binds to DnaE1-pol is the DNA sliding clamp (DnaN, the β subunit, Figure 1-2), which is essential for processive DNA synthesis [14]. Following the procedure used to assay coupling between the polymerase and exonuclease, a set of roughly 125 DnaE1-pol+clamp hybrid sequences was generated for intermolecular Statistical Coupling Analysis.

In addition to DnaE1-pol intramolecular coupling, the polymerase-clamp analysis yielded coupled positions within the clamp and between it and the polymerase. Interestingly, when the hierarchical clustering algorithm was applied, most of the positions identified in the clamp were in close proximity to the conserved binding site for the five-to-six-residue peptide (β-binding motif) shared by DnaE1-pol, the clamp loader, and other polymerases [183, 184] (Figure 3-11a). This may be indicative of a larger binding surface on the clamp for DnaE1-pol, perhaps, perhaps similar in size to that of the error-prone, translesion synthesis (TLS) polymerase Pol IV [185].

Within the DnaE1-pol portion of the coupled cluster, the most prominent positions were, again, those seen for the polymerase alone (not shown). Because many of these positions contain residues that distinguish between Eco- and Taq-like DnaE1-pol, this finding suggests the possibility that there may be corresponding subtypes of the clamp. Further support for the existence of an Eco-like clamp comes from the locations of the most highly conserved (65% or greater identity, see Materials & Methods) residues that are unique to the clamp homologs from *α-, β-* and *γ-proteobacteria* (Figure 3-11b). Most of these residues are also near the peptide-binding site and several correspond to positions in the coupled clusters. Given the substantially lower conservation of the clamp, however, evidence for subtypes of this protein is much less robust than for the polymerase and exonuclease.

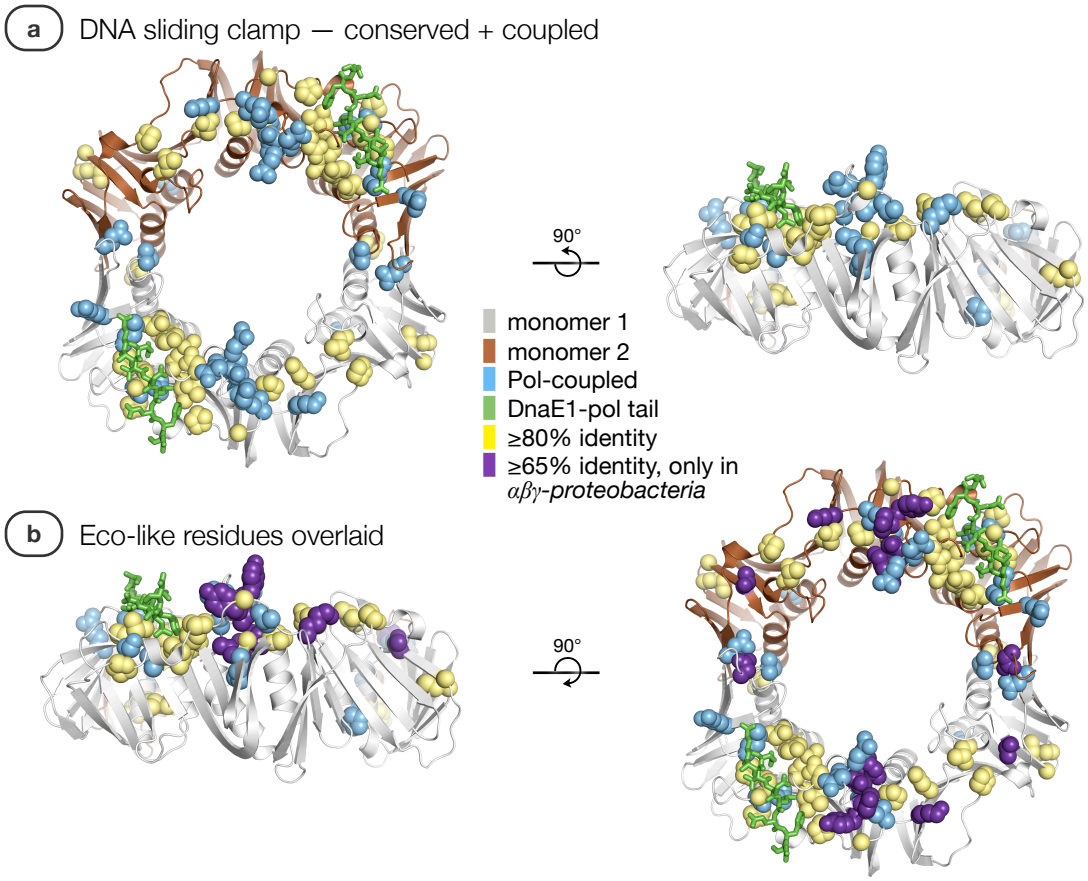


Figure 3-11. Conserved and coevolved residues cluster near the peptide binding-site on the clamp.

Structure of the *E. coli* clamp bound to a C-terminal peptide from DnaE1-pol (PDB 3d1f). (***a***) Residues at positions corresponding to the strongest coupling cluster in the DnaE1-pol+clamp intermolecular Statistical Coupling Analysis are indicated as blue spheres. To compensate for much weaker conservation of the clamp than was observed for DnaE1-pol or Eco-like DnaQ-exonuclease, identity as low as 80% was used for selecting conserved residues (yellow spheres). (***b***) Residues with 65% or greater identity that are specific to the clamp homologs from *α-, β-* and *γ-proteobacteria* are shown as purple spheres.

One potential source of differences between Eco- and Taq-like clamp subtypes could be the Eco-like DnaQ-exonuclease homologs. Although the *E. coli* clamp and DnaQ-exonuclease do not form a stable complex [186], Lamers *et al.* [26] suggest that the two probably contact each other while bound to the polymerase. Similarly, Evans *et al.* [25] depict apparent interaction of the DnaQ-exonuclease domain of *G. kaustophilus* PolC with a bound clamp. Because Taq-like DnaE1-pol homologs lack a partner DnaQ-exonuclease, they might interact differently their clamps, perhaps binding them more tightly than would an Eco-like polymerase on its own. Consistent with this hypothesis, preliminary small angle X-ray scattering (SAXS) data (Meindert Lamers, personal communication) show that the addition of *E. coli* DnaQ-exonuclease to a polymerase+clamp complex reduces its radius of gyration.

3.4.5 Part E — Allostery & Biochemistry

3.4.5.1 PHP mutations decrease polymerase activity

To evaluate the compatibility of *E. coli* DnaE1-pol with allostery, I used the mutants described in CHAPTER 2 (Table 1-1) as a proxy for physical perturbations of the PHP domain that might occur during allosteric communication between it and the Palm. Starting from the crystallization construct EcoE1(917) (residues 1 through 917 of 1160), the least heavily mutated protein was 3mPHP in which three non-metal-binding residues were reverted to the canonical histidines (Arg10His, Phe44His, Arg203His). The other two mutants (4m- and 5mPHP) had four and five point mutations, respectively, which collectively restored metal-binding residues to all nine positions in the PHP motif. For this assay, 3mPHP was purified in the presence of chelators to remove bound zinc (Materials & Methods).

Consistent with PHP-Palm allostery, DNA polymerization rates of the *E. coli* DnaE1-pol constructs decreased as the number of point mutations in the PHP domain was increased (Figure 3-12). Although the 50 to 60% decreases in polymerization rate were small, they were nevertheless significant (Figure 1-7, Figure 2-1).

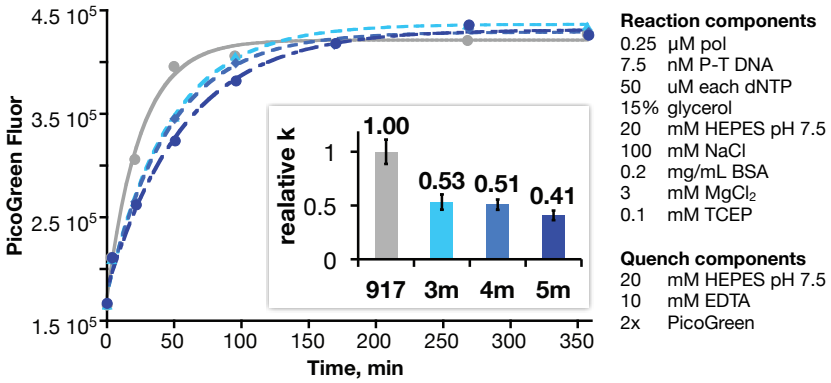


Figure 3-12. PHP mutations slow polymerization by DnaE1-pol.

Production of dsDNA monitored by the intercalating dye PicoGreen. Inset: Relative rates of DNA polymerization (917: EcoE1(917), 3m, 4m and 5m: 3m-, 4m- and 5mPHP, respectively).

Although slowing DNA synthesis by mutation of the EcoE1(917) PHP cleft does not prove the existence of allostery, it does show that changes to the PHP can affect the Palm. If binding of the DnaQ-exonuclease subunit to the PHP domain [61] changed the polymerase in a way favored the active Palm conformation, this would provide an allosteric explanation for the ability of the *E. coli* DnaQ-exonuclease to stimulate DNA synthesis [14, 107].

3.4.5.2 Zinc modulates polymerase stability

Since allostery [173] and protein folding [187] are both forms of information transfer within proteins, I analyzed the unfolding of constructs derived from *E. coli* DnaE1-pol as a means toward understanding the potential for allosteric regulation of DNA synthesis by this polymerase.

In the presence of Gdn•HCl, both the crystallization construct, EcoE1(917), and its triple point mutant, 3mPHP, unfolded via a two-state mechanism (Figure 3-13a). For these denaturation experiments, the two constructs were purified using protocols that minimized divalent metal binding (Materials & Methods), and, as monitored by circular dichroism and tryptophan fluorescence (Figure 3-13b), both had low, yet significantly different c_m values (vs. Gdn•HCl: EcoE1(917): 659 ± 6 mM, 3mPHP: 559 ± 3 mM). In comparison to EcoE1(917), the lower c_m for 3mPHP is consistent with the removal of the two structural arginine residues (Arg10 and Arg203) from its PHP cleft (Figure 1-7).

Using temperature melts, I showed that increasing the number of point mutations in the PHP domain (Table 1-1) progressively destabilized the polymerase (Figure 3-14a). EcoE1(917) was the most stable (T_m = 46.0 °C), while 3mPHP (T_m = 43.9 °C) and 4mPHP (T_m = 40.4 °C) were considerably less stable under these conditions. Although the melting temperature for 5mPHP (T_m = 39.4 °C) was only 1 °C lower than that of 4mPHP, this difference was significant, and the unfolding transition of 5mPHP was considerably broader than that of the other constructs (Figure 3-14a).

Adding zinc to my samples, I showed that this divalent metal globally stabilized two of the PHP mutants (Figure 3-14b). The constructs 3m- and 4mPHP both experienced large increases in their melting temperatures, whereas EcoE1(917) and 5mPHP both melted at lower temperatures as the zinc concentration was increased. Using a binding equation with a linear term to compensate for lower melting temperatures at the highest zinc concentrations (Equation 3-10), I determined apparent zinc dissociation constants for 3m- (K_d = 0.29 ± 0.02 μM) and 4mPHP (K_d = 2.1 ± 0.3 μM). Although not commonly used for metal binding, my temperature-shift assay is similar to Differential Scanning Fluorimetry, which has recently taken precedence in the field of high-throughput screening for receptor-ligand interactions [188, 189].

The apparent affinities of 3m- and 4mPHP for zinc are much weaker than those of naturally occurring zinc metalloproteins (typical dissociation constants: 1 to 100 picomolar [190]), but the continued appearance of only single unfolding transitions (Figure 3-14a), despite significantly different melting temperatures for the constructs in the presence of zinc, further supports the conclusion that they all unfold cooperatively.

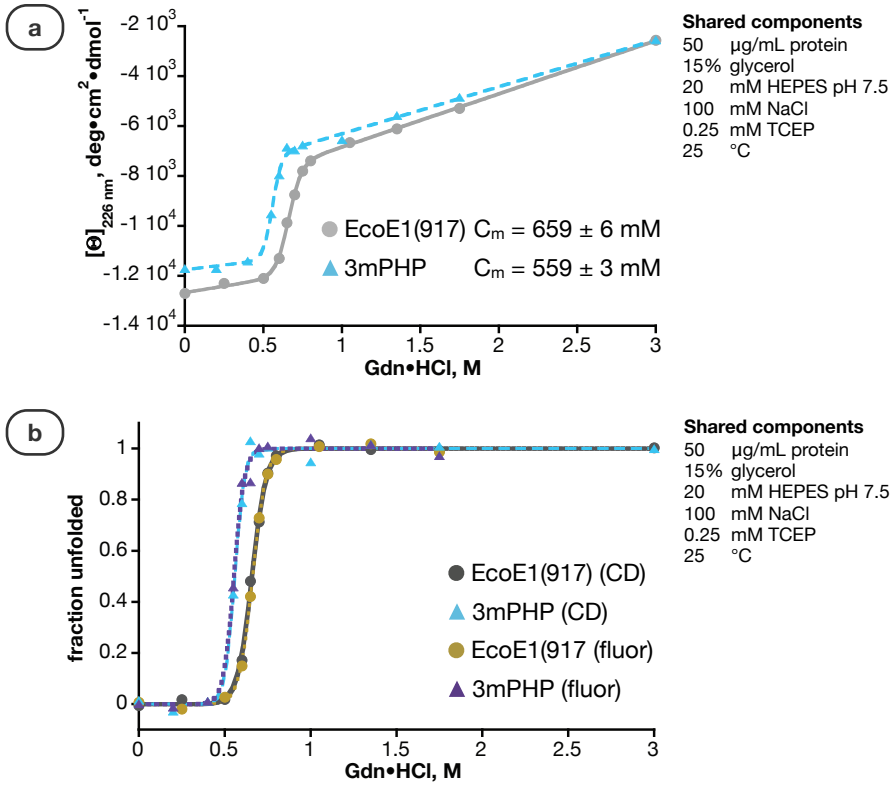


Figure 3-13. EcoE1(917) and 3mPHP show apparent two-state unfolding.

(***a***) Raw circular dichroism (CD) data and (***b***) normalized circular dichroism and tryptophan fluorescence data for chemical denaturation of the *E. coli* DnaE1-pol crystallization construct (EcoE1(917)) and a triple point mutant of this construct (3mPHP).

Although the Palm sites have evolved to bind magnesium with octahedral geometry (Figure 3-2), many similar sites in other proteins have zinc dissociation constants comparable to [191-193] the calculated values for 3m- and 4mPHP. Zinc ions are also often bound by cysteine residues [194, 195], and five of the 10 cysteines in EcoE1(917) are surface exposed (Figure 4-14), which makes them prone to intermolecular disulfide formation, and perhaps especially vulnerable to zinc binding. Consequently, the destabilization of EcoE1(917) and 5mPHP upon the addition of zinc and the more gradual loss of stability by 3m- and 4mPHP at higher zinc concentrations (Figure 3-14b) might be attributable to spurious binding of the metal by the catalytic aspartate triad in the Palm or by any of the 10 cysteine residues present in the constructs.

Though my denaturation experiments do not prove the existence of allostery, the cooperative unfolding of EcoE1(917) appears compatible with allosteric communication. By definition, a protein that folds via a cooperative, two-state mechanism [196] must pass through a transition state in which information is

simultaneously transmitted to all of its parts. Although cooperative folding is certainly not a prerequisite for allostery, latent allosteric networks may be a general feature of proteins with two-state folding mechanisms. This necessitates the assumption that those residues transferring information during a cooperative folding transition are likely to maintain their contacts in the folded state, and through these, facilitate allostery. Since the four constructs tested all exhibit two-state unfolding, *E. coli* DnaE1-pol might be equipped to respond, allosterically, to the binding of its DnaQ-exonuclease.

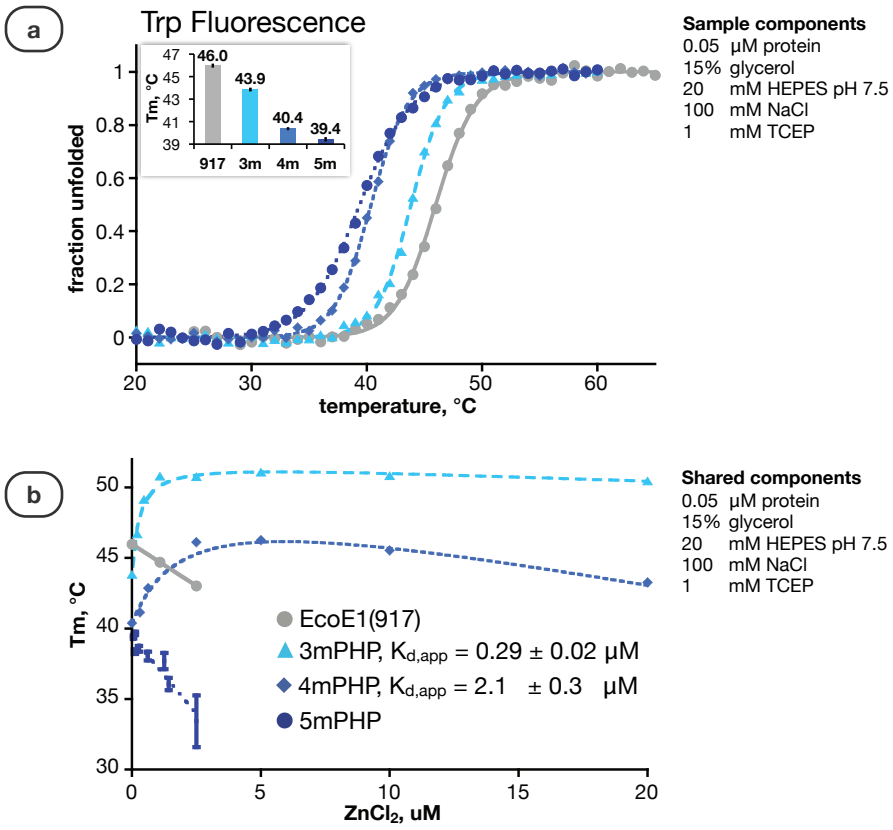


Figure 3-14. Two PHP mutants of *E. coli* DnaE1-pol bind zinc.

(***a***) Normalized thermal denaturation traces for *E. coli* DnaE1-pol constructs as monitored by tryptophan fluorescence. (***b***) Melting temperature, determined as in part (*a*) versus zinc concentration. Two constructs, 3m- and 4mPHP, bind zinc with the indicated apparent affinities.

Looking beyond allostery, the cooperative unfolding of EcoE1(917) is remarkable given its evolutionarily distinct domains (PHP, Palm and Fingers) and size in excess of 100 kDa. Surveying the literature, most proteins that fold and unfold via a two-state mechanism are small (generally not much more than 100 residues or roughly 10 kDa) and consist of a single domain [197-201]. Cooperative unfolding, however, has also been demonstrated for some multi-domain proteins near 50 kDa [202-204], and since

some of these did not refold after denaturation, comparable studies would be necessary to determine if EcoE1(917) can fold cooperatively *in vitro.*

Interestingly, the largest protein for which I found suggestion of two-state unfolding was the crystallization construct of *G. kaustophilus* PolC [25]. Evans *et al.* [25] mutated two of the residues in the PHP motif this protein and mentioned (but did not include data showing) that the mutants had the same thermal denaturation properties and level of polymerase activity observed for the crystallization construct. Such results are quite different from those for EcoE1(917), which, as shown above, became a slower, less stable polymerase upon mutation of its PHP. Also different for the two polymerases, removal of the DnaQ-exonuclease domain from *G. kaustophilus* PolC increased its DNA synthesis rate, while, in my studies (not shown) of EcoE1(917) and in published analysis of full-length *E. coli* polymerase [107], addition of the DnaQ-exonuclease subunit stimulated DNA synthesis.

3.5 Conclusion

Using sequence alignments and structural analysis, I have identified and characterized the two main subtypes of DnaE1-pol (Figure 3-15). The ancestral, Taq-like subtype likely dates back to before the appearance of cyanobacteria, more than 3,500 million years ago (Figure 3-7), and is a constitutively active polymerase that both synthesizes and, using its metal-binding PHP domain, edits DNA to remove mismatch errors. In contrast, the Eco-like subtype of DnaE1-pol has a Palm that can adopt a distorted, inactive conformation, and its non-metal-binding PHP domain lacks editing activity. In the absence of *in cis* editing, the Eco-like PHP binds to a DnaQ-exonuclease that edits DNA *in trans.* Because this polymerase-exonuclease pair is found only in α*-,* β- and γ*-proteobacteria,* a minimum age of 1,500 million years can be estimated for the Eco-like subtype based on the emergence of mitochondria, which are descended from an α-proteobacterium.

In addition, the results from Statistical Coupling Analysis and biochemical experiments suggest the possibility of PHP-Palm allostery in both Taq- and Eco-like DnaE1-pol, along with the potential for communication with the DnaQ-exonuclease by the Eco-like polymerase. Such intersubunit allostery would be a compelling explanation for the ability of DnaQ-exonuclease to boost DNA synthesis [14, 107], and PHP-Palm allostery, in general, might be a component of the mechanism by which the Palm domain of DnaE1-pol releases its grasp on the DNA primer strand to permit mismatch removal, either by an editing PHP domain or DnaQ-exonuclease subunit.

Beyond the pol III Core, Statistical Coupling Analysis suggests that the Eco- vs. Taq-like distinction may extend to other pol III subunits and other proteins that bind to them. Besides DnaE1-pol+clamp coupling mentioned in this chapter, unpublished data from Brian Kelch suggests coevolution of DnaE1-pol Pol II and Pol IV, but not Pol V or

the τ subunit of the clamp loader, despite its binding site at the C-terminus of DnaE1-pol.

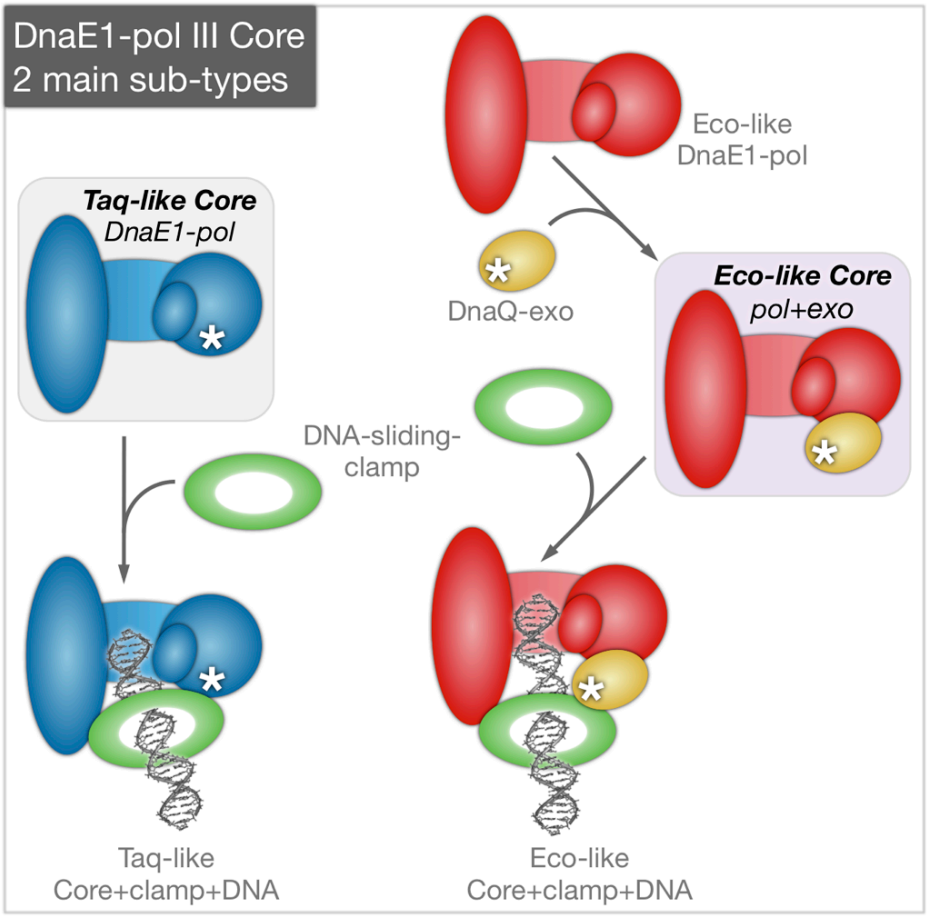


Figure 3-15. The two main subtypes of DnaE1-based pol III have different Cores.

Schematic showing the Taq- and Eco-like pol III Cores binding to the clamp and DNA. The Taq-like Core consists of a single DnaE1-pol that contains both DNA synthesis and editing (*) activities *in cis.* The Eco-like Core contains a DnaE1-pol that lacks editing activity and, instead, binds a DnaQ-exonuclease, which performs editing *in trans.*

All told, my analysis brings into focus the distinguishing features of two widespread versions of the bacterial replisome. This knowledge may assist in the development of novel antibiotics that are selective for species with dissimilar replicative polymerases, and it confirms the ancient origins of these enzymes. Besides the Eco- and Taq-like DnaE1-pol described here, I also found two less prevalent replicative polymerase subtypes that may not yet have been tested biochemically. One is a PolC lacking a DnaQ-exonuclease domain, the other is a DnaE1-pol with a DnaQ-exonuclease fused to its N-terminus, and both further demonstrate the evolution of surprisingly diverse bacterial replicative DNA polymerases from one synthesis and two editing domains.

CHAPTER 4

Crystallization of the *Escherichia coli* Replicative Polymerase with Primer-Template DNA

4.1 Summary

The recently solved structures of the replicative DNA polymerases from three bacterial replisomes [24-26, 35] illuminated features common to all C-family polymerases and others that distinguish the main homolog types — PolC and the Taq- and Eco-like DnaE1-pol (CHAPTER 3, [40, 41]). Although high-resolution data exists for a PolC in complex with DNA, there is no comparable data for either subtype of DnaE1-pol. Toward this end, I present two high-yield, high-purity expression and purification protocols for *Escherichia coli* DnaE1-pol; I exploit fluorescence anisotropy to identify conditions that enhance DNA binding affinity; and I present a protocol for lessening DNA degradation, as detected by a simple, gel-based assay. Leveraging these advances, I grow crystals with radiation-sensitive diffraction to 4-Å resolution that likely contain a complex of a truncated form of the *E. coli* polymerase plus DNA, as suggested by the properties of similarly grown crystals that showed tryptophan fluorescence indicative of protein and coloration consistent with the presence of dye-labeled DNA. As a complement to standard crystallization techniques, I also develop cysteine mutants of a truncated form of the *E. coli* polymerase that are suitable for site-specific attachment of DNA via disulfide exchange or Michael addition reactions.

4.2 Introduction

This chapter is organized in the following way. After the Materials & Methods, the Results & Discussion are presented in four parts. Part A — Polymerase Handling — covers protocols for improved purification and monomerization of *E. coli* DnaE1-pol constructs and identifies C-terminally truncated constructs suitable for crystallization trials. In Part B — Polymerase+DNA, I determine buffer conditions that favor DNA binding while reducing DNA degradation caused by a nuclease contaminant. Crystals of a polymerase+DNA complex are grown in Part C — Crystallization, and their diffraction is analyzed. Strategies for sulfur-based protein-DNA crosslinking are presented, and I generate cysteine mutants of a C-terminal truncation of the *E. coli* DnaE1-pol that are compatible with crosslinking in Part D — Crosslinking. Finally, these results are briefly summarized in the Conclusion.

4.3 Materials & Methods

Under my supervision, Jordan Anaya performed some aspects of the cloning, protein expression and purification, crystallization trials, and DNA binding experiments

described herein. Additional cloning was performed by Mary Coons and Tiffany Chou (moving constructs into the p3AT vector) and also by Caleb Cassidy-Amstutz, Xiaoxian Cao and Lore Leighton (removal and introduction of cysteine residues by QuikChange mutagenesis).

4.3.1 Polymerase cloning, expression and purification

Constructs in the pET-28a-PP vector were expressed and purified as described in CHAPTER 2 and CHAPTER 3 and as outlined in Table 4-1.

Constructs in the p3AT vector cloned using ligation independent cloning (LIC) [205] and N- and C-terminal tags were added as part of the primers used during the PCR amplification step. Expression and purification of constructs from the p3AT vector was performed as for the pET-28a-PP constructs using the steps outlined in Table 4-1. With the exception of the simplified HisTrap protocol, steps for constructs expressed from the p3AT vector were identical to those for the pET-28a-PP constructs. Protein concentrations were determined by absorption at 280 nm based on theoretical molar extinction coefficients.

4.3.2 Polymerase monomerization assay

Samples were prepared as described in Figure 4-4, spin filtered to remove foreign objects and large aggregates, and run through a Superdex 200 SMART analytic size-exclusion column (GE Healthcare).

4.3.3 DNA binding assay

Experiments were performed under the conditions described in Figure 4-6. Binding was monitored by observing the anisotropy of TAMRA-labeled DNA (Figure 4-1) using either a Fluoromax-3 fluorometer (HORIBA Jorbin Yvon) or a PerkinElmer Victor3 fluorescence plate reader with a 535/30 nm excitation filter and a 595/60 nm emission filter. Binding-curves were fit using a standard hyperbolic equation (Equation 4-1) in KaleidaGraph 4 (Synergy Software). Exponential and linear fitting of dissociation constant trends was performed in Excel 2008 (Microsoft) using the equations indicated in Figure 4-6d or a standard linear equation (Figure 4-6e).

```
Tem50   3-   CCCTGCGTGCGCCGTAAGTTCCTGAATCCAGACATAAATAGATGGGTGTT -TAMRA 5
Pri40   5-   GGGACGCACGCGGCATTCAAGGACTTAGGTCTGTATTTAT              -3
```

Figure 4-1. Fluorescently labeled DNA used for binding experiments.

The above DNA served as the primer-template for DNA binding experiments. For ssDNA binding, the template (5pTAMRA-Tem50) labeled with TAMRA on its 5'-end was used in the absence of the primer strand (Pri40).

$$a = \frac{A[pol]}{K_d + [pol]} + a_0$$

Equation 4-1. Anisotropy as a function of polymerase concentration yields DNA dissociation constant.

Anisotropy of the fluorophore attached to the DNA, a, is a function of polymerase concentration, $[pol]$. KaleidaGraph fit the other terms, which include: A, the amplitude of the maximum increase in anisotropy upon binding; K_d, the dissociation constant for the polymerase+DNA complex; and a_0, the anisotropy in the absence of polymerase.

4.3.4 Crystallization trials

Robotic crystallization trials were performed using a Phoenix crystallization robot (Art Robbins Instruments), commercial screens were purchased from Hampton Research and Qiagen, and custom screens were prepared using a PerkinElmer MultiPROBE II HT liquid handling robot.

Primer-template DNA (purchased from Integrated DNA Technologies, Figure 4-2) was prepared by dissolving the primer and template strands in water and annealing them together using a PCR machine programmed to begin a temperature gradient at 95 °C, cooling to 5 °C in 2 to 2.5 minute steps of 5 °C. Annealing was verified by PAGE (polyacrylamide gel electrophoresis, Figure 4-2) using 18%-acrylamide gels in 1x TBE buffer (Tris, borate, EDTA). The Cy5-labeled primer-templates (Integrated DNA Technologies) were a generous gift from Randall McNally.

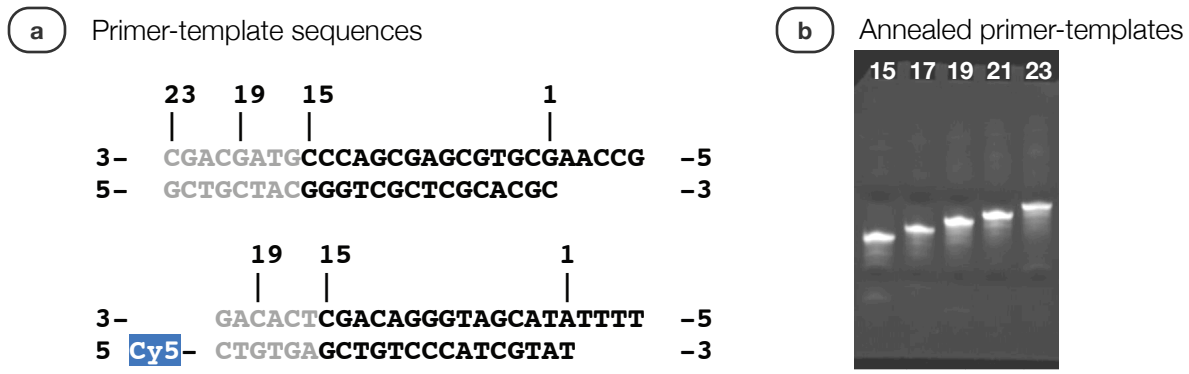


Figure 4-2. DNA used in crystallization trials.

Bases colored black were present in all primer-templates, while gray bases were only present in primer-templates with longer duplex regions. (***a***) The members of the series of unlabeled primer-templates were named according to the lengths of the duplex and the 5'-overhang on the template strand (e.g. PT17.5: 17 duplex bases with a 5-base overhang). The Cy5-labeled series was the fifth series used in crystallization trials, so their names begin with the number 5 (e.g. 5PT19.4: 5th primer-template series, 19 duplex bases with a 4-base overhang). (***b***) Annealing of the unlabeled primer-templates was verified by TBE-PAGE.

4.3.5 Crystal imaging

Tryptophan fluorescence images and accompanying visible light images were captured using a PRS-1000 Protein Review Station microscope (Korima). All other visible light images were recorded using a Nikon microscope and digital camera.

4.3.6 X-ray diffraction

Diffraction from single crystals was generated using synchrotron radiation at the Lawrence Berkeley National Laboratory, Advanced Light Source (ALS), Howard Hughes Medical Institute (HHMI) beamlines 8.2.1 and 8.2.2 and collected in frames of 1° oscillation per 20 second exposure.

4.3.7 Disulfide crosslinking assay

Cysteine-incorporation mutants and primer-template DNA were combined under the conditions indicated in Figure 4-15. After incubation, the samples were diluted with buffer containing MMTS (methyl-methanethiosulfonate, Thermo Scientific (Pierce)), which caps unreacted cysteine residues, and aliquots were analyzed by non-reducing SDS-PAGE (sodium dodecyl sulfate PAGE, Figure 4-15).

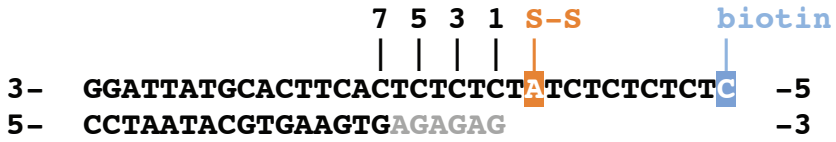


Figure 4-3. DNA for disulfide crosslinking trials.

Bases colored black were present in all primer-templates, while gray bases were only present in primer-templates with longer duplex regions. After reacting with cystamine (Figure 4-12a), the orange base contains the disulfide (S-S). A biotin moiety (biotin) is attached to the blue base for pull downs via the FASTDXL (focused array screening technique for disulfide X-linking) method of Corn and Berger [206]. These crosslinking primer-templates are referred to by the number of single-stranded bases between the disulfide-containing base and the first base of the duplex region (e.g. 5cPT1: member of the 5th series of crosslinking primer-templates, 1 single-stranded base between the duplex and the disulfide-containing base).

4.4 Results & Discussion

4.4.1 Part A — Polymerase Handling

4.4.1.1 High-yield polymerase purification with N- or C-terminal His6-tags

As a first step toward structural analysis with DNA of *E. coli* DnaE1-pol, I developed constructs of this polymerase that could be expressed and purified at high yield. The first *E. coli* DnaE1-pol structures were solved [26] (PDB codes 2hnh, 2hqa) using an EcoE1(917) construct prepared without the benefit of expression or purification tags, and, with reasonable purity, yields were on the order of 1 mg/L cell culture (Meindert Lamers, personal communication). By adding a cleavable N-terminal His6-tag (Materials & Methods), however, I increased this yield to 15 mg/L while greatly enhancing sample purity (purification protocol: Table 4-1).

Although EcoE1(917) is generally well behaved, there are other constructs based on *E. coli* DnaE1-pol that I found hard to purify due to extensive C-terminal proteolytic degradation. Following the suggestion of Debora Makino, I designed a full-length polymerase construct with a cleavable T7-tag on its N-terminus and a non-cleavable His6-tag on its C-terminus. Tiago Barros successfully used this construct to purity a complex of the polymerase plus the clamp, DnaQ-exonuclease and the C-terminal portion of the τ subunit from the clamp loader (personal communication). T.B. also placed these tags on EcoE1(917), and, with these in place, he was able to prepare (protocol: Table 4-1) this construct with high purity at a yield greater than 100 mg/L (personal communication).

a

N-His6-PP — C-term truncations

Step	Buffers
1 HisTrap 3x 5-mL	PI-500-0
load in and wash with 4%	PI-40-500
desalt on column with 100%	β-A
elute directly onto Q with 100%	PI-40-500
2 HiTrap Q 3x 5-mL	β-A
180 mL (12 CV) gradient 12-45%	β-B
3 • PreScission Protease cleavage	
dialyze to 100 mM NaCl with	β-A
4 HisTrap 1x 5-mL	β-A
capture flow-through directly on Q	PI-40-500
5 HiTrap Q 1x 5-mL	β-A
elute with 100%	β-B
6 • Concentrate	
7 S200 16/60	TK100D
8 • Concentrate	

N-T7-PP & C-His6 — full length

Step	Buffers
1 HisTrap 3x 5-mL	PI-500-0
load in and wash with 4%	PI-500-500
elute with 100%	PI-500-500
2 • PreScission Protease cleavage	
dialyze to 100 mM NaCl with	β-A
3 HiTrap Q 3x 5-mL	β-A
180 mL (12 CV) gradient 12-45%	β-B
4 • Concentrate	
5 S200 16/60	TK100D
6 • Concentrate	

b

Buffers

PI-500-0		stock solution, mL
10%	glycerol	50% 150
25	mM Na/K PO4 pH 8.5	500 37.5
500	mM NaCl	5000 75
5	mM βME	14300 0.262
		add water to 750

β-B		stock solution, mL
10%	glycerol	50% 100
20	mM HEPES pH 7.5	1000 10
1000	mM NaCl	5000 100
5	mM βME	14300 0.175
		add water to 500

PI-40-500		stock solution, mL
10%	glycerol	50% 50
25	mM Na/K PO4 pH 8.5	500 12.5
40	mM NaCl	5000 2
500	mM Imidazole pH 8	2000 62.5
		add water to 250

β-A		stock solution, mL
10%	glycerol	50% 200
20	mM HEPES pH 7.5	1000 20
0	mM NaCl	5000 0
5	mM βME	14300 0.350
		add water to 1000

PI-500-500		stock solution, mL
10%	glycerol	50% 50
25	mM Na/K PO4 pH 8.5	500 12.5
500	mM NaCl	5000 25
500	mM Imidazole pH 8	2000 62.5
		add water to 250

TK100D		stock solution, mL
10	mM TAPS pH 8.5	1000 4
100	mM K glutamate	2000 20
10	mM DTT	2000 2
		add water to 400

Table 4-1. Purification of *E. coli* DnaE1-pol constructs with either N- or C-terminal His6-tags.

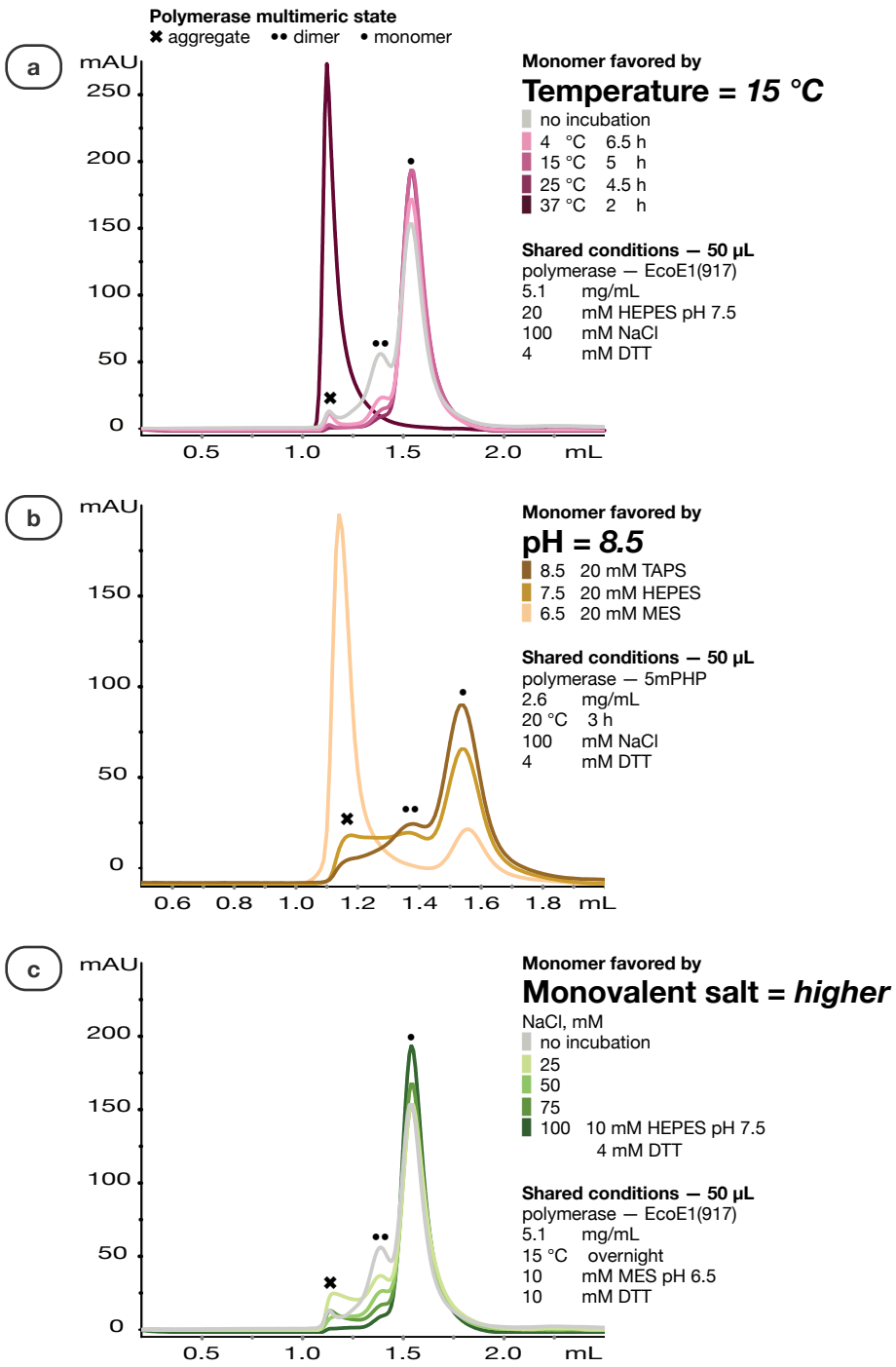
(***a***) Purification protocols. ***N-His6-PP*** — N-terminal His6-tag, cleavable with PreScission Protease (in pET-28a-PP plasmid, a modified pET-28a vector). This tag is only suitable for purification of C-terminally truncated constructs (e.g. EcoE1(917), EcoE1(956) and EcoE1(975) and some mutated derivatives thereof). ***N-T7-PP & C-His6*** — N-terminal T7-tag cleavable with PreScission Protease and a non-cleavable C-terminal His6-tag (in p3AT plasmid, tags added using PCR primers). This two-tag system is required for full-length polymerase, and, unless experimental conditions will not tolerate the presence of a C-terminal His6-tag, this system is better for all *E. coli* DnaE1-pol constructs, because it gives much higher yields and purities. Full length *E. coli* DnaE1-pol has fairly high intrinsic affinity for Ni-resin in low ionic strength buffers, so high concentrations of both NaCl and imidazole are necessary for the efficient elution of this polymerase from a HisTrap column. Non-column steps (•) do not require an FPLC. (***b***) Buffer recipes.

4.4.1.2 Regeneration of monomeric polymerase from higher-order multimers

Before formulating the storage buffer TK100D (Table 4-1) for *E. coli* DnaE1-pol constructs, I used several other buffers that were ultimately not suitable for long-term sample storage. Although the polymerase samples that were kept in these buffers appeared to be monomeric during initial size-exclusion chromatography (data not shown), after concentration and a freeze-thaw cycle, some of the samples appeared to contain higher molecular weight species (Figure 4-4) that were likely the result of multimerization and aggregation.

To convert the higher molecular weight species back into monomeric polymerase, I tested temperature and a series of buffer components by incubating the polymerase samples under various conditions and then running aliquots from these tests through an analytical-scale size-exclusion column. Based on the results (Figure 4-4), I chose a monomerization buffer (bMono: 15% glycerol, 20 mM TAPS pH 8.5, 100 mM NaCl (or K glutamate [207-210]), 1 mM TCEP and, optionally, 5 mM EDTA) and incubation conditions suitable for *E. coli* DnaE1-pol and constructs based upon it (Table 4-2).

Constructs of *E. coli* DnaE1-pol may tolerate other buffering reagents in addition to TAPS, but most will not tolerate prolonged exposure to lower pH. Exceptions include EcoE1(917) and 3mPHP, which are stable down to pH 7.5 (data not shown). NaCl is preferred for circular dichroism (CD) experiments, because it is less optically active than K glutamate. Chloride, however, inhibits DNA binding [207-210], so K glutamate is the preferred monovalent salt for crystallization trials with DNA. Reducing agent is required when working with *E. coli* DnaE1-pol constructs containing the full complement of endogenous cysteine residues, because several of these are liable to form intermolecular disulfide bonds (Figure 4-14). TCEP is a strong and non-volatile reducing agent that neither chelates metals nor contains sulfur. Consequently, TCEP is the reductant of choice for metal binding experiments, and it is likely to be an essential buffer component when forming polymerase-DNA crosslinks under reducing conditions via Michael addition (Figure 4-13). Because 3mPHP may copurify with bound metal ions (CHAPTER 2), and metal-depended nuclease contamination is a known issue for *E. coli* DnaE1-pol constructs (Figure 4-7), inclusion of the chelator EDTA is recommended during monomerization to yield polymerase samples with reduced nuclease activity and without bound divalent metals. Before performing experiments (such as crystallization trials with DNA, Figure 4-6b) that require divalents, however, a compensatory buffer exchange step is necessary to remove the EDTA after monomerization.


Polymerase multimeric state
✖ aggregate •• dimer • monomer
a
mAU
250
200
150
100
50
0
0.5
1.0
1.5
2.0
mL
Monomer favored by
Temperature = 15 °C
no incubation
4 °C 6.5 h
15 °C 5 h
25 °C 4.5 h
37 °C 2 h
Shared conditions — 50 µL
polymerase — EcoE1(917)
5.1 mg/mL
20 mM HEPES pH 7.5
100 mM NaCl
4 mM DTT
b
mAU
150
100
50
0
0.6
0.8
1.0
1.2
1.4
1.6
1.8
mL
Monomer favored by
pH = 8.5
8.5 20 mM TAPS
7.5 20 mM HEPES
6.5 20 mM MES
Shared conditions — 50 µL
polymerase — 5mPHP
2.6 mg/mL
20 °C 3 h
100 mM NaCl
4 mM DTT
c
mAU
200
150
100
50
0
0.5
1.0
1.5
2.0
mL
Monomer favored by
Monovalent salt = higher
NaCl, mM
no incubation
25
50
75
100 10 mM HEPES pH 7.5
4 mM DTT
Shared conditions — 50 µL
polymerase — EcoE1(917)
5.1 mg/mL
15 °C overnight
10 mM MES pH 6.5
10 mM DTT

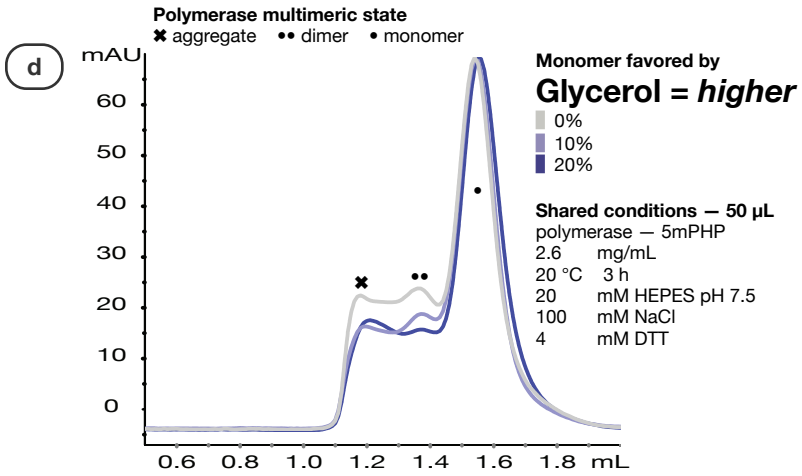


Figure 4-4. Monomeric polymerase is favored at 15 °C by pH 8.5 and higher monovalent salt and glycerol.

Size-exclusion chromatograms. 5mPHP is EcoE1(917) with five mutations in its PHP domain (CHAPTER 2). Within each test, the same amount of protein was used for each sample. Note that differences between the y-axis values between tests are not meaningful. Also, volume is measured from the start of the run and not from the injection point, so the x-axis values are accordingly larger. (***b***) Other data, not shown here, indicates that the polymerase constructs are stable at pH 9.5, but use of such high pH is not recommended because it is likely detrimental to DNA binding (Figure 4-6). (***d***) The chromatograms for the glycerol test are normalized to the monomer peak. As a consequence, the samples containing less monomer appear to contain an excess of the higher molecular weight species.

Overnight incubation at 15 °C served to monomerize most constructs, and shorter incubation times of only a couple hours at 25 °C were sufficient (data not shown) for EcoE1(917) and the PHP metal-binding mutant 3mPHP (CHAPTER 2, Table 4-2).

Monomerization buffer — bMono	DnaE1-pol construct	Monomeric tendency	Incubation temperature	Incubation time
Standard components				
15% glycerol 20 mM TAPS pH 8.5 100 mM NaCl or K Glu 1 mM TCEP	EcoE1(917) 3mPHP	high	25 °C	1 h
	E. coli DnaE1 4mPHP 5mPHP [Cys xL mutants]	low	15 °C	overnight (>8 h)
Optional components				
5 mM EDTA (2^+ metal stripping)				

Table 4-2. Recommended conditions for monomerization of *E. coli* DnaE1-pol constructs.

Variations are included for different applications. NaCl is less optically active than K glutamate, so it is suitable for circular dichroism (CD) experiments. Chloride, however, interferes with DNA binding, so K glutamate should be used for complex formation (Figure 4-6a) [207-210]. The chelator EDTA can be included if a divalent-metal-free state of a polymerase construct is desired.

4.4.1.3 Identification of monomeric C-terminal truncations of *E. coli* DnaE1-pol

EcoE1(917) crystallizes robustly (Figure 4-8) [26], but, lacking the wild type C-terminal domains, it cannot bind its clamp [211] or clamp loader [26, 211] and has decreased DNA binding affinity relative to the full length protein [212]. To improve the chances of crystallizing a polymerase+DNA complex, I prepared, with the assistance of Jordan Anaya, full-length *E. coli* DnaE1-pol and four alternative C-terminal truncation constructs. These purified to apparent homogeneity (as estimated by Coomassie-stained SDS-PAGE, not shown), but the longer two truncations formed soluble aggregates that could not be monomerized (Figure 4-5). The first (EcoE1(1072)) was truncated near the end of the oligonucleotide-binding (OB) domain, and the second (EcoE1(1154)) was missing only the C-terminal peptide motif that binds the clamp loader [26, 211]. The two shorter constructs (EcoE1(956) and EcoE1(975)), both of which were truncated before the oligonucleotide-binding domain, and the full-length polymerase all appeared monomeric and are therefore suitable for crystallization trials.

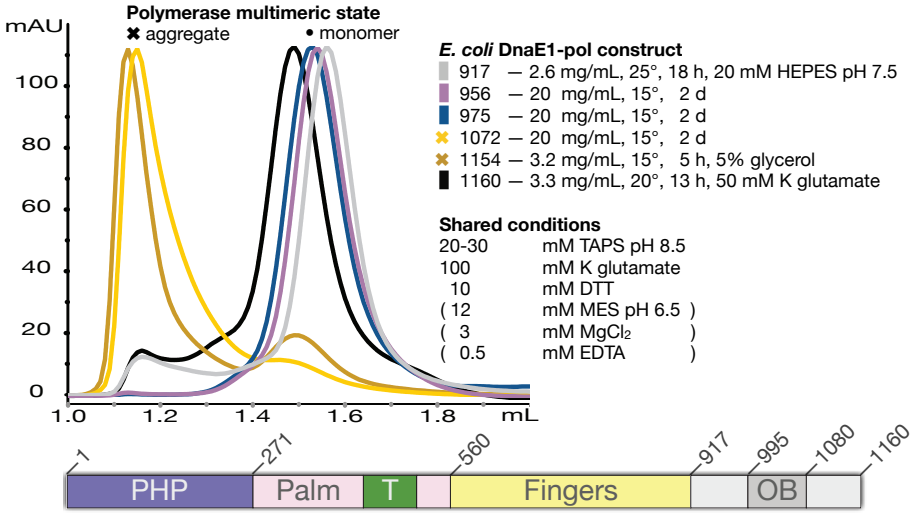


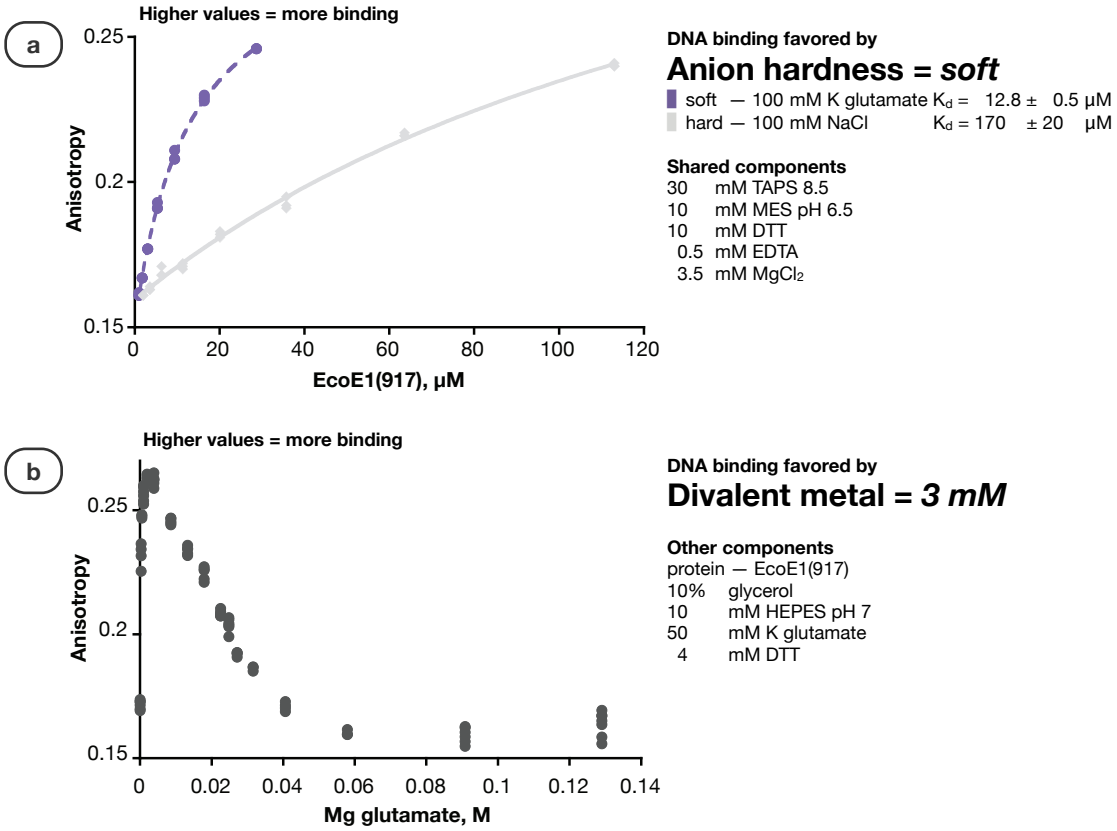
Figure 4-5. Two C-terminal truncations of *E. coli* DnaE1-pol aggregated irreversibly.

Size-exclusion chromatograms. Constructs are indicated by the number of the most C-terminal residue that they include (e.g. 917: EcoE1(917), 1160: full length). Conditions in parenthesis were not present in the EcoE1(917) or full length samples. Domain (T: Thumb) boundaries are approximate. Samples contained roughly the same amount of protein, and peak heights were normalized. Note that the elution volume is measured from the start of the run and not from the injection point, so the peaks appear to elute at correspondingly later volumes.

4.4.2 Part B — Polymerase+DNA

4.4.2.1 Buffer conditions favoring DNA binding

To increase the likelihood of crystallizing a polymerase+DNA complex, I surveyed buffer conditions affecting DNA binding (Figure 4-6) and compared the affinity trends with the polymerase monomerization data (Figure 4-4) to formulate a buffer (bMono, Table 4-2) that strikes a compromise between minimizing polymerase aggregation and facilitating DNA binding. Among the conditions tested, monomerization and DNA binding have opposing preferences for monovalent salt concentration and pH. Monomerization requires higher salt and higher pH (Figure 4-4), but DNA binding increases when both are lower (Figure 4-6).



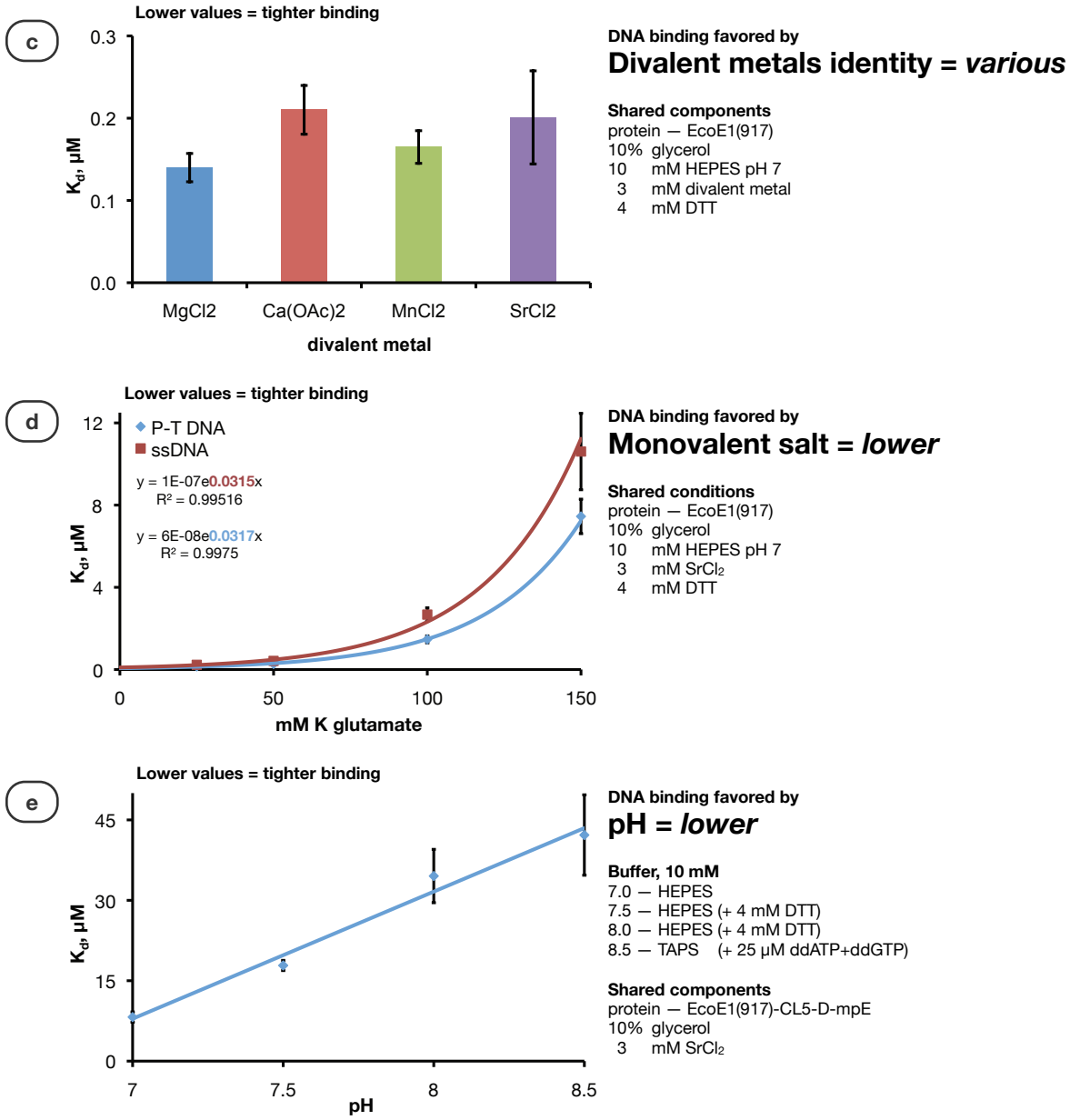


Figure 4-6. Soft anions, lower pH and monovalent salt, and 3 mM divalent metal favor DNA binding.

Raw binding data, as in (*a*) and (*b*), based on fluorescence anisotropy of DNA labeled with TAMRA was used to generate the plots in (*c*), (*d*) and (*e*). The protein sample used in (*e*) was a mixture of cysteine mutants of EcoE1(917). (***d***) The higher Kd values for ssDNA, which are in constant ratio with values for primer-template, indicate that the polymerase preferentially binds primer-template throughout the tested range of K glutamate concentrations.

Polymerase DNA binding affinity	++		+		–	
	rel K_d	1	rel K_d	7.5	rel K_d	150
•	10% glycerol		10% glycerol		10% glycerol	
	20 mM HEPES pH 7.5	1	20 mM HEPES pH 8	1.5	20 mM TAPS pH 8.5	2.25
	50 mM K glutamate	1	100 mM K glutamate	5	100 mM NaCl	65
••	3 mM $MgCl_2$		3 mM $MgCl_2$		3 mM $MgCl_2$	
•••	1 mM TCEP		1 mM TCEP		1 mM TCEP	

Table 4-3. Minor changes in buffer composition can dramatically affect DNA binding.

Replacing 50 mM K glutamate with 100 mM NaCl while increasing pH from 7.5 to 8.5 would result in a 150-fold increase in dissociation constant for DNA. •Glycerol can be added to enhance construct monomerization (Figure 4-4) without affecting DNA binding. ••Magnesium is included because I was unable to detect DNA binding in the absence of divalent metal ions (Figure 4-6). •••*E. coli* DnaE1-pol constructs are highly redox sensitive, so reducing agent is required. DTT is also effective, but it can chelate some metals and may therefore interfere with DNA binding (data not shown).

Having tested individual components, I combined the results to estimate the relative DNA binding affinities for *E. coli* DnaE1-pol constructs in three different buffers with varying pH, monovalent salt concentration, and salt identity (Table 4-3). This illustrates that making a series of relatively minor changes in buffer composition can dramatically decrease DNA binding affinity. Note that divalent metal ions are absolutely required for DNA binding (Figure 4-6b), and the optimal concentration is 3 mM. Considerable flexibility exists, however, with respect to the identity of the divalent metal, and magnesium, calcium, manganese and strontium appeared to support binding equally (Figure 4-6c). Note that these binding experiments were performed in the absence of nucleotide triphosphate, and formation of a ternary complex of polymerase+DNA+nucleotide may have different metal requirements than formation of a polymerase+DNA binary complex.

4.4.2.2 Slowing unwanted DNA degradation

Using a gel-based assay, I determined that my polymerase samples caused extensive DNA degradation over a length of time comparable to a standard crystallization trial (3 days, Figure 4-7). Pre-treating my samples with 0.5 mM EDTA for 30 min before the addition of divalent metal ions reduced DNA degradation, and decreasing the incubation temperature from 20 to 15 °C caused a further reduction.

The effectiveness of the EDTA pre-treatment is somewhat surprising, because it persists even after the addition of excess divalent metal ions (Figure 4-7). This could be explained, however, if the contaminating nuclease irreversibly denatured upon metal removal by EDTA or if the nuclease were unable to catalyze DNA degradation using the metal ions that were subsequently added (magnesium or strontium).

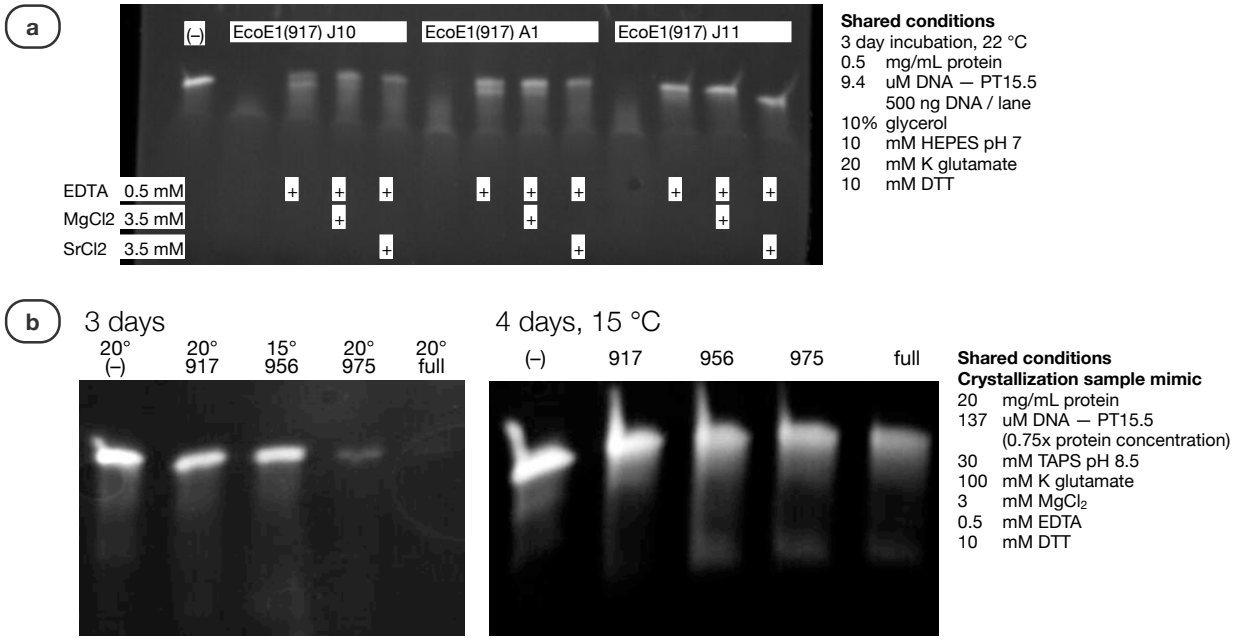


Figure 4-7. EDTA pre-treatment and lower temperature reduce DNA degradation.

Urea SDS-PAGE and ethidium bromide DNA staining. Decreased upper band intensity indicates degradation of DNA. Negative control samples (–) contained no protein. EDTA was added to samples at ambient temperature (22 °C) 30 min before the addition of metal. (***a***) Three separate preparations of EcoE1(917). (***b***) C-terminal truncation constructs (Figure 4-5) and full length *E. coli* DnaE1-pol. A greater amount of DNA degradation was observed for the longer polymerase constructs. Since the constructs were all purified using an N-terminal His6-tag, the longer constructs were also substantially less pure. Much higher purity is expected when using constructs containing a C-terminal His6-tag.

4.4.3 Part C — Crystallization

4.4.3.1 Characterization of crystals containing a polymerase+DNA complex

Having identified buffer components compatible with DNA binding (Figure 4-6) and steps to lower nuclease activity (Figure 4-7), I performed crystallization trials using EcoE1(917) and primer-template DNA with varying duplex lengths. Many crystals were observed, and, using the presence of tryptophan fluorescence (protein) and high birefringence (DNA only) as indicators, I scored these according to their likely contents (Figure 4-8). Both EcoE1(917) and the primer-templates had a tendency to crystallize on their own. DNA crystals often appeared in conditions with PCB buffer (sodium propionate, sodium cacodylate and bis-tris propane) or high concentrations of divalent metal ions (200 mM calcium or magnesium, well into the inhibitory range for polymerase binding, Figure 4-6b). Crystals that grew at low pH with PEG 3350 were often morphologically similar to the initial hits that led to the published EcoE1(917) structures [26] (PDB codes 2hnh, 2hqa) and occasionally appeared in the negative controls without DNA, so they likely contained only the polymerase.

Among the many crystal hits, only those in one condition (Index E7) did not appear to be either polymerase or DNA alone (Figure 4-8). Although the well solution for these

crystals contained magnesium, at 50 mM, this was the lowest concentration at which divalent metal was found in any of the commercial screens that I used. The crystals appeared at some point between three and nine days after the start of the trial and were misshapen spheres that exhibited tryptophan fluorescence (Figure 4-9a) but not birefringence.

After reproducing the crystals and identifying conditions that consistently produced single rods (100 x 10 x 10 μm, Figure 4-10), I applied the method of Georgescu *et al.* [213] to check for the presence of DNA in the crystals. When using a primer strand labeled with the blue dye Cy5 on its 5'-end, crystals grew with a blue color that was consistent with the presence of the dye and, therefore, DNA (Figure 4-9b).

Having found evidence of both protein and DNA within the crystals, I began collecting X-ray diffraction data on my largest single rods and redoubled my efforts to grow ones still larger. Using a combination of robotic and manual screening, including the use of various crystal seeding techniques, I performed more than 50,000 crystallization trials. Among these, one yielded a single crystal (Figure 4-10) that was substantially larger than the single rods that typically appeared, but this crystal did not have improved diffraction.

During X-ray analysis, some rods initially produced spots to near 4 Å (data not shown), but these higher resolution reflections disappeared after the first few frames, likely as the result of radiation damage. Typical diffraction for the rods went to 8 Å and was highly anisotropic (Figure 4-11).

At this point, there are many variables that could still be tested to improve this crystal form. For instance, the DNA could be modified in many ways, including increasing the duplex length, varying the 5'-overhang length, attaching different fluorophores to either the duplex or overhang ends, and adding a 3'-overhang to the duplex end. With respect to the protein, constructs with differing degrees of C-terminal truncation or tags on their N- and/or C-termini might make additional contacts within the crystals and thereby improve their diffraction. In addition, these and other DNA and polymerase permutations (e.g. homologs from other species) could be included when screening for as yet unidentified crystal forms.

xtal type	screen	well (grid)	DNA	NTP	crystal description	bi-refrin-gence	precip	Trp fluor	day first seen	well solution
1	Index	E7	PT19.5, PT21.5	-	poorly crystalline spheres	-	++	+	3 < x < 9	0.05 M MgCl2 0.1 M HEPES pH 7.5 30% v/v PEG 550 MME
1	Index	E7	PT17.5	-	poorly crystalline spheres	-	++	+	3 < x < 9	0.05 M MgCl2 0.1 M HEPES pH 7.5 30% v/v PEG 550 MME
2	JCSG+	A2	PT21.5, PT23.5	-	EcoE1(917) needle fan	-	+	+	1 < x < 4	0.1 M tri-Na citrate pH 5.5 20% withv PEG 3000
2	JCSG+	G8	PT19.5, PT21.5, PT23.5	-	EcoE1(917) needle fan	-	+	+	1 < x < 4	0.15 M DL-Malic acid pH 7.0 20% withv PEG 3350
2	JCSG+	G8	PT15.5, PT17.5	ddTTP	EcoE1(917) needle fan	-	+	+	1 < x < 4	0.15 M DL-Malic acid pH 7.0 20% withv PEG 3350
2	JCSG+	G8	PT17.5	-	EcoE1(917) needle fan	-	+		1 < x < 4	0.15 M DL-Malic acid pH 7.0 20% withv PEG 3350
2	Index	H2	PT23.5	-	EcoE1(917) needle fan	-	+		1 < x < 4	0.2 M K Na tartrate tetrahydrate 20% withv PEG 3350
2	Index	H7	PT19.5, PT21.5, PT23.5	-	EcoE1(917) needle fan	-	+		1 < x < 4	0.15 M DL-Malic acid pH 7.0 20% withv PEG 3350
2	Index	H7	PT17.5	-	EcoE1(917) needle fan	-	+		1 < x < 4	0.15 M DL-Malic acid pH 7.0 20% withv PEG 3350
2	Index	H8	-, PT17.5		EcoE1(917) needle fan		++		3 < x < 1.5 mo	0.1 M Mg formate dihydrate 15% withv PEG 3350
2	Index	H7	PT15.5, PT17.5	ddTTP	EcoE1(917) needle fan	-	+		x < 3	0.15 M DL-Malic acid pH 7.0 20% withv PEG 3350
2	Index	H6	PT17.5	ddTTP	EcoE1(917) needle fan	-	++		x < 3	0.2M Na formate 20% withv PEG 3350
2	Index	H8	-	ddTTP	EcoE1(917) needle fan		++		3 < x < 1.5 mo	0.1 M Mg formate dihydrate 15% withv PEG 3350
2	Index	H7	PT19.5, PT21.5	ddTTP	EcoE1(917) needle fan	-	+		x < 3	0.15 M DL-Malic acid pH 7.0 20% withv PEG 3350
2	Index	H2	PT23.5	ddTTP	EcoE1(917) needle fan	-	+		x < 3	0.2 M K Na tartrate tetrahydrate 20% withv PEG 3350
2	JCSG+	A2	PT15.5	ddTTP	EcoE1(917) single crystal	-	++		x < 3	0.1 M tri-Na citrate pH 5.5 20% withv PEG 3000
2	PACT	C7	PT19.5, PT21.5, PT23.5	-	EcoE1(917) spheres with spikes	-	++		x < 1	0.2M NaCl 0.1M Hepes pH 7 20% withv PEG 6000
2	PACT	C9	PT19.5, PT21.5, PT23.5	-	EcoE1(917) spheres with spikes	-	++		x < 1	0.2M LiCl 0.1M Hepes pH 7 20% withv PEG 6000
2	PACT	C7	PT15.5, PT17.5	-	EcoE1(917) spheres with spikes	-	+++		x < 4	0.2M NaCl 0.1M Hepes pH 7 20% withv PEG 6000
2	PACT	C8	PT15.5	-	EcoE1(917) spheres with spikes	-	++		x < 4	0.2M Ammonium Cl 0.1M Hepes pH 7 20% withv PEG 6000
2	PACT	C9	-, PT15.5, PT17.5	-	EcoE1(917) spheres with spikes	-	+++		x < 4	0.2M LiCl 0.1M Hepes pH 7 20% withv PEG 6000
2	PACT	E6	PT19.5	ddTTP	EcoE1(917) spheres with spikes	+	++		x < 3	0.2M Na formate 20% withv PEG 3350
2	PACT	C7	PT19.5, PT21.5, PT23.5	ddTTP	EcoE1(917) spheres with spikes	-	++		x < 3	0.2M NaCl 0.1M Hepes pH 7 20% withv PEG 6000
2	PACT	C8	PT19.5, PT21.5	ddTTP	EcoE1(917) spheres with spikes	-	++		x < 3	0.2M Ammonium Cl 0.1M Hepes pH 7 20% withv PEG 6000
2	PACT	C9	PT19.5, PT21.5, PT23.5	ddTTP	EcoE1(917) spheres with spikes	-	+++		x < 4	0.2M LiCl 0.1M Hepes pH 7 20% withv PEG 6000
2	PACT	C7	-, PT15.5, PT17.5	ddTTP	EcoE1(917) spheres with spikes	-	++		x < 3	0.2M Na Cl 0.1M Hepes pH 7 20% withv PEG 6000
2	PACT	C8	PT15.5, PT17.5	ddTTP	EcoE1(917) spheres with spikes	-	++		x < 3	0.2M Ammonium Cl 0.1M Hepes pH 7 20% withv PEG 6000
2	PACT	C9	-, PT15.5, PT17.5	ddTTP	EcoE1(917) spheres with spikes	-	+++		x < 4	0.2M LiCl 0.1M Hepes pH 7 20% withv PEG 6000
3	JCSG+	D2	PT23.5	-	block shards	++	+	-	x < 1	0.2 M MgCl2 0.1 M HEPES pH 7.5 30% v/v PEG 400
3	PACT	C2	PT23.5	-	block shards without sharp edges	+++	+++	-	x < 1	0.1M PCB buffer pH 5 25% withv PEG 1500
3	JCSG+	H11	PT21.5, PT23.5	-	block shards without sharp edges	+++	+++	-	x < 4	0.2 M MgCl2 0.1 M bis-Tris pH 5.5 25% withv PEG 3350
3	PACT	C3	PT21.5, PT23.5	-	rods with edges	+++	+++	-	x < 1	0.1M PCB buffer pH 6 25% withv PEG 1500
3	PACT	C4	PT23.5	-	rods with edges	+++	+++	-	x < 1	0.1M PCB buffer pH 7 25% withv PEG 1500
3	Index	E1	PT19.5, PT21.5, PT23.5	-	block shards	+	+		x < 4	0.2 M CaCl2 0.1 M BIS-TRIS pH 6.5 45% v/v MPD
3	Index	G10	PT19.5, PT21.5	ddTTP	block shards	+++	+++		x < 3	0.2 M MgCl2 0.1 M BIS-TRIS pH 5.5 25% withv PEG 3350
3	PACT	C2	PT23.5	ddTTP	block shards	+++	+++		x < 3	0.1M PCB buffer pH 5 25% withv PEG 1500
3	PACT	A10	PT23.5	ddTTP	block shards	+++	+++		x < 3	0.2M MgCl2 0.1M Na AcOH pH 5 20% withv PEG 6000
3	PACT	C3	PT23.5	ddTTP	rods with edges	+++	+++		x < 3	0.1M PCB buffer pH 6 25% withv PEG 1500
3	PACT	C4	PT23.5	ddTTP	rods with edges	+++	+++		x < 3	0.1M PCB buffer pH 7 25% withv PEG 1500

xtal type — likely crystal contents	
1	polymerase+DNA
2	polymerase
3	DNA

Figure 4-8. Initial crystallization trials yielded one hit containing both protein and DNA.

Screening of a complex of EcoE1(917) plus primer-template with varying duplex lengths (15 to 23 bases) and a 5-base 5'-overhang. Negative control (–) samples did not contain DNA. Abbreviations and acronyms — *xtal:* crystal, *NTP:* deoxynucleotide triphosphate, *precip:* precipitation, *Trp fluor:* tryptophan fluorescence detected using a UV fluorescence microscope. Crystals that appeared morphologically similar to those that led to the published EcoE1(917) structures (PDB 2hnh, 2hqa) were very common also appeared in negative control wells that did not contain DNA. In previous trials, crystals containing only DNA had very high birefringence (unpublished results). All trials were performed at 20 °C in 96-well sitting-drop format using 100+100 nL sample plus well solution.

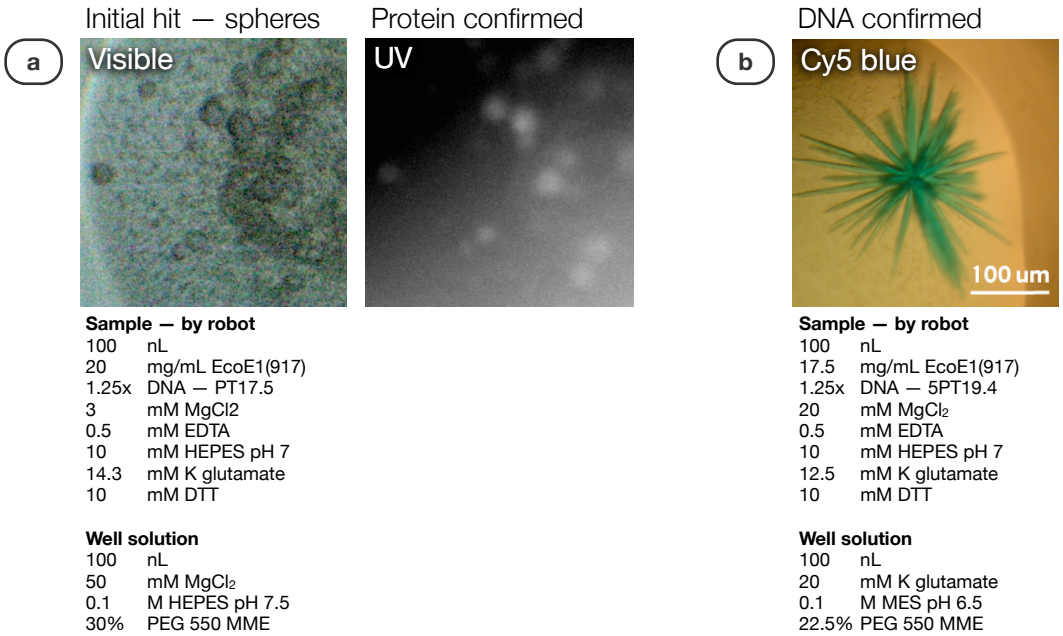
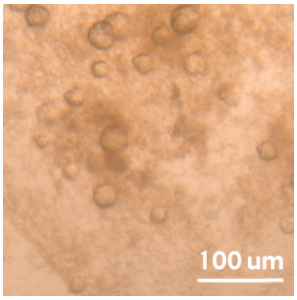


Figure 4-9. Crystals contained both protein and DNA.

(***a***) Visible light and UV fluorescence images of the initial crystals indicate the presence of protein. (***b***) Visible light image of improved, blue crystals grown starting from a complex of EcoE1(917) plus primer-template DNA labeled with the blue dye Cy5 on the 5'-end (duplex end) of the primer DNA strand.

Initial hit — spheres



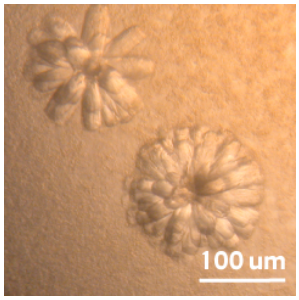
Sample — by robot

100	nL
20	mg/mL EcoE1(917)
1.25x	DNA — PT17.5
3	mM MgCl2
0.5	mM EDTA
10	mM HEPES pH 7
14.3	mM K glutamate
10	mM DTT

Well solution

100	nL
50	mM $MgCl_2$
0.1	M HEPES pH 7.5
30%	PEG 550 MME

Columnar clusters



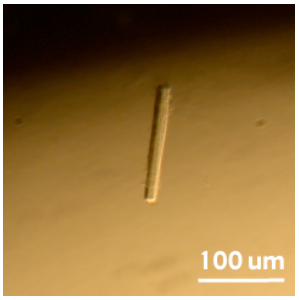
Sample — by hand

10	mg/mL EcoE1(917)
1.25x	DNA — PT17.5

Well solution

5	mM $MgCl_2$
0.1	M HEPES pH 6.5
30%	PEG 550 MME

Single rod



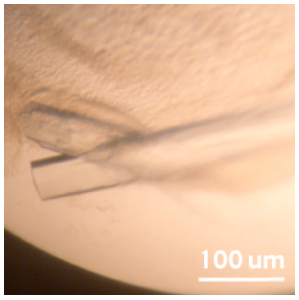
Sample — by hand + seeding

10	mg/mL EcoE1(917)
1.25x	DNA — PT17.5

Well solution

5	mM $MgCl_2$
0.1	M HEPES pH 6.5
30%	PEG 550 MME

Largest single crystal



Sample — by robot

100	nL
17.5	mg/mL EcoE1(917)
3x	DNA — PT17.5
20	mM $MgCl_2$
0.5	mM EDTA
10	mM HEPES pH 7
12.5	mM K glutamate
10	mM DTT

Well solution

100 nL

80%

- 32.5% PEG 550 MME
- 100 mM MES pH 6.5
- 20 mM K glutamate
- 20 mM $MgCl_2$

20%

- 0.2 M NH_4 formate pH 6.6
- 20% w/v PEG 3350

Figure 4-10. Reproduction and enlargement of crystals likely containing a polymerase+DNA complex.

Visible light images. Conditions for trials prepared by hand are approximate. Reducing the magnesium concentration from 50 to 5 mM in the well solution significantly improved the crystals, and seeding appeared to increase the frequency of single rod formation. Both the typical single rods and the uniquely large single crystal diffracted to roughly 8 Å. All crystallization trials shown here were performed at 20 °C. Robot trays were sitting-drop; manual trays were hanging-drop.

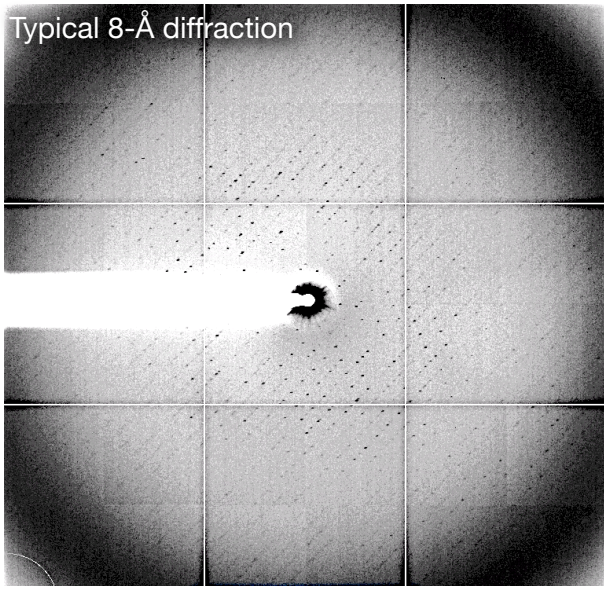


Figure 4-11. Single rods typically diffracted to 8 Å.

An X-ray diffraction pattern with spots to 8 Å.

4.4.4 Part D — Crosslinking

4.4.4.1 Strategies for cysteine-based crosslinking of protein to DNA

In recent years, the pioneering work of Verdine and coworkers has led to numerous crystal structures of low affinity protein+DNA complexes that were trapped [214-222] using disulfide crosslinking (Figure 4-12). This method for covalent bond formation requires DNA containing a disulfide and a protein that has a reactive, surface-exposed cysteine that can crosslink to that DNA. Generally, some endogenous cysteine residues must be removed from the protein to prevent unwanted crosslinking. Afterward, a cysteine residue is introduced into the protein at the desired crosslinking site. Reversible crosslinking occurs by disulfide exchange when the sulfur atom in a cysteine residue replaces one of the sulfur atoms in the disulfide attached to the DNA (Figure 4-12b).

Researchers typically use a convertible base (O6-phenyl-dI, available from Eurofins MWG Operon) to introduce the disulfide into the DNA by reaction with deprotonated cystamine (Figure 4-12a). For applications where attachment to the DNA 5'-end is acceptable (Figure 4-14), such as crosslinking to a polymerase, DNA can be ordered with a disulfide-containing group directly attached (e.g. 5' Thiol Modifier C6 S-S, available from Integrated DNA Technologies, Figure 4-12b), thus avoiding the losses of time and yield associated with performing the cystamine reaction.

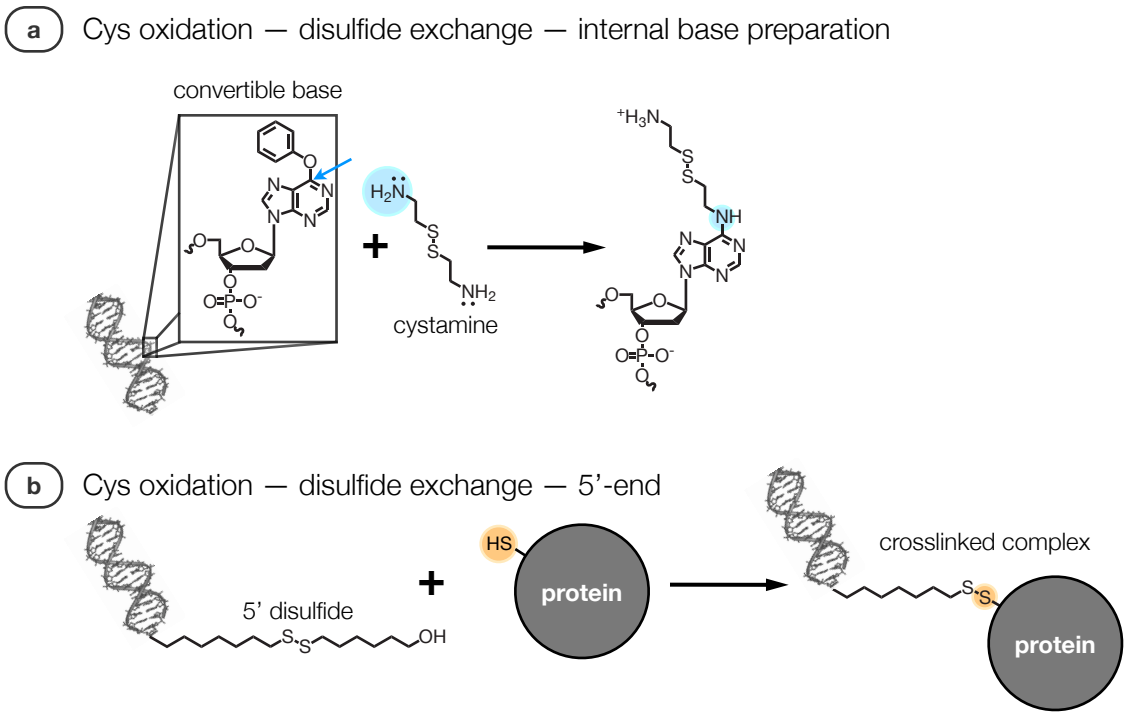


Figure 4-12. Disulfide exchange: Cysteine-DNA crosslinking under oxidizing conditions.

(***a***) To attach the disulfide to an internal site within the DNA, it can be synthesized with a convertible base and reacted with cystamine. (***b***) A disulfide can also be added to the 5'-end of a DNA strand during its synthesis. Crosslinking by disulfide exchange occurs when free thiol from a cysteine residue replaces for one of the sulfur atoms in the disulfide attached to the DNA.

During crosslinking by disulfide exchange, the sulfur atom from a reduced cysteine residue replaces one of the two sulfurs in the disulfide on the DNA (Figure 4-12b). Note that, if the cysteine were to replace the other sulfur in the disulfide, crosslinking to the DNA would not occur, and the protein would, instead, be attached to the other half of the original disulfide. However, since disulfide exchange is readily reversible, a crosslinking between protein and bound DNA is favored at equilibrium, because the small-molecule half of the original disulfide may diffuse away, whereas the DNA cannot.

The reversibility of the disulfide exchange also allows for tuning of this reaction to select, at equilibrium, for crosslinks formed as the result of tighter binding between the protein and DNA. Disulfide crosslinks can be broken by the addition of thiol-containing reducing agents such as βME or DTT. With increasing reducing agent concentrations, the formation of protein-DNA disulfides is restricted to crosslinks resulting from tight binding interactions that hold in close proximity the disulfide on the DNA and a cysteine on the protein.

In addition to reversible disulfide crosslinking, cysteine residues can also be irreversibly crosslinked to DNA by Michael addition (Figure 4-13) under reducing

conditions. This may be beneficial when working with a protein that, like *E. coli* DnaE1-pol, is prone to aggregation under oxidizing conditions (data not shown). During Michael addition, a soft nucleophile attacks the β-carbon of an α,β-unsaturated carbonyl, which is a soft electrophile. Because sulfur is the only soft nucleophile found in proteins, only cysteine residues readily participate in Michael additions, and the stability of the resulting crosslink permits the storage and crystallization of the resulting protein+DNA complex in the presence of high concentrations of reducing agent, if necessary for protein stability.

Preparation of cysteine mutants for crosslinking by Michael addition is the same as for disulfide crosslinking, and, likewise, DNA can be crosslinked at either an arbitrary position or its 5'-end. Because Michael additions are not reversible under conditions generally compatible with protein stability, the efficiency (and therefore selectivity) of this crosslinking reaction can only be tuned kinetically. Such tuning can be performed by altering buffer conditions or by using different Michael acceptors. There are two commonly used acceptors for biological crosslinking applications (maleimide and (meth)acrylamide, Figure 4-13), and both may be suitable for trapping of protein+DNA complexes.

Maleimide reacts rapidly with thiols [223, 224] and is commonly used to attach fluorescence dyes to cysteine residues. As part of the heterobifunctional crosslinker Sulfo-SMCC (sulfosuccinimidyl 4-[N-maleimidomethyl]cyclohexane-1-carboxylate, available from Thermo Scientific, Figure 4-13a), maleimide is commonly used in nonspecific crosslinking applications such as attaching enzymes to antibodies [225]. DNA can be modified to contain a primary amine (e.g. Int Amino Modifier C6 dT, available from Integrated DNA Technologies, Figure 4-13a) that will replace the N-hydroxysuccinimide ester (NHS ester) of Sulfo-SMCC under mildly basic conditions (pH 7 to 9). This results in a stable intermediate in which the maleimide of SMCC extends from the site of the reacted primary amine. Upon addition of protein, the maleimide group can react with a cysteine residue to form a protein-DNA crosslink.

The Sulfo-SMCC strategy presents the advantage of arbitrary positioning the crosslinking site within the DNA, but the steric bulk of the resulting linker (Figure 4-13a) may pose problems for some protein+DNA complexes. Furthermore, because the high reactivity of maleimide toward cysteine can facilitate crosslinking based on only weak or transient interactions, it may also result in covalent complexes that lack physiological relevance.

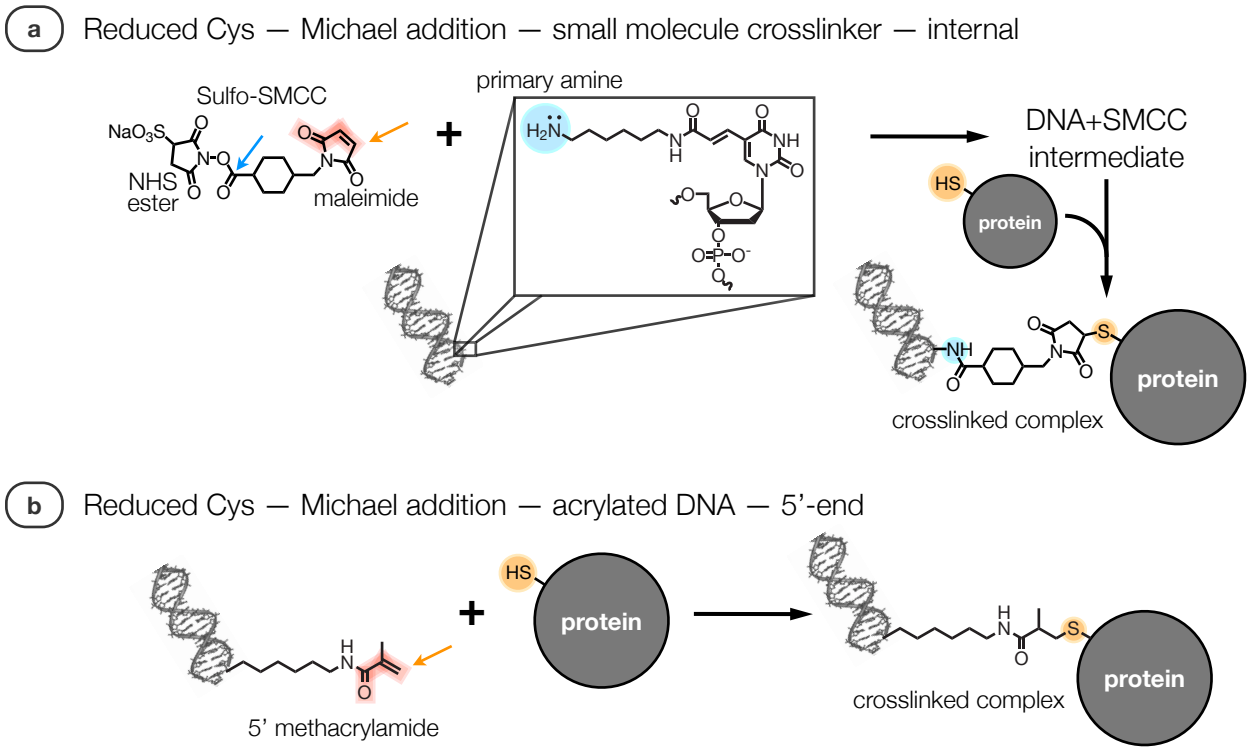


Figure 4-13. Michael addition: Cysteine-DNA crosslinking under reducing conditions.

Cysteine can crosslink with α,β-unsaturated carbonyls by Michael addition. (***a***) For crosslinking to internal sites, a Michael acceptor can be added to DNA by reacting a modified base containing a primary amine with the small molecule Sulfo-SMCC (Thermo Scientific). The intermediate product formed by this reaction can react with a cysteine residue in a protein to form a crosslinked complex. (***b***) DNA can also be synthesized with the Michael acceptor methacrylamide on its 5'-end.

For trapping polymerase+DNA complexes and other applications in which crosslinking to the DNA 5'-end would be sufficient, DNA containing a methacrylamide group (5' Acrydite, available from Integrated DNA Technologies, Figure 4-13b) could be used. Because methacrylamide is a much less reactive Michael acceptor than maleimide [223, 224], DNA containing Acrydite would not be suitable for very low affinity complexes, but it would also be less liable to form spurious, non-physiological crosslinks. Because Acrydite is added during synthesis, DNA with this Michael acceptor can be prepared faster and at higher yield than via the Sulfo-SMCC protocol.

4.4.4.2 Validation of crosslinking-compatible cysteine mutants of EcoE1(917)

I initially removed all 10 endogenous cysteine residues (Figure 4-14a) from EcoE1(917) and used this Cys-free background to generate a series of mutants with single cysteines introduced at different locations. These constructs were unusable, however, because they rapidly degraded under crosslinking conditions (Figure 4-15).

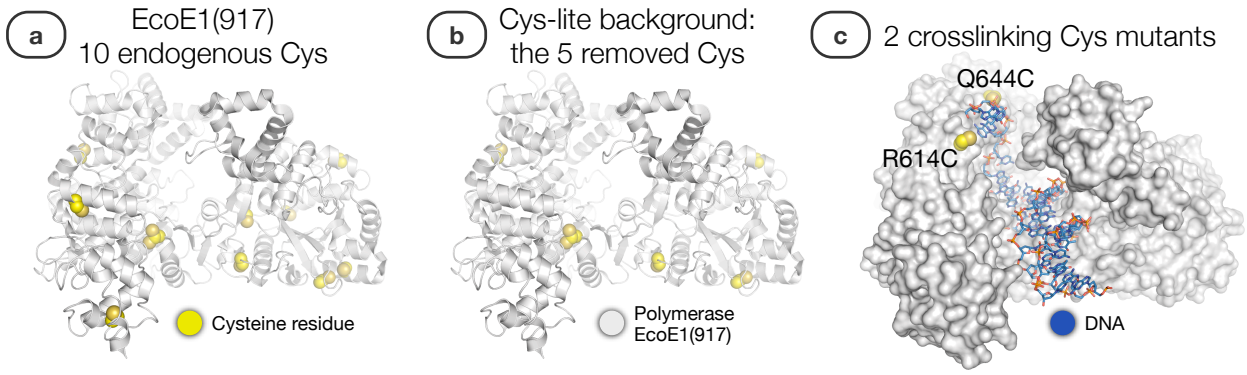
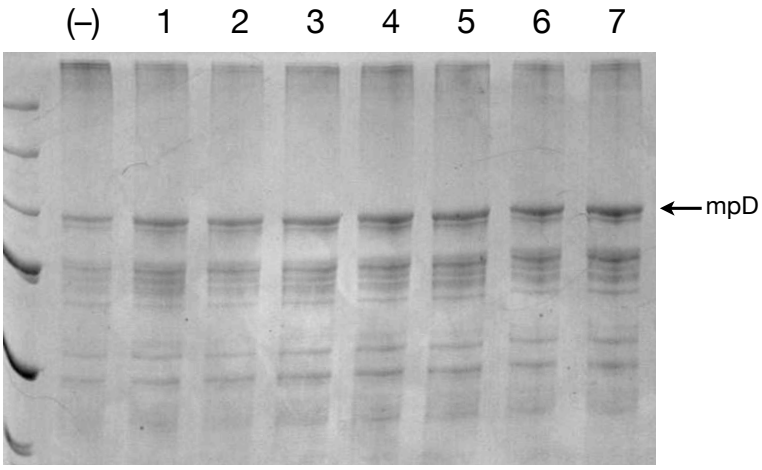


Figure 4-14. Five of 10 cysteine residues were removed from EcoE1(917) to generate a Cys-lite polymerase.

(***a***) EcoE1(917) contains 10 cysteine residues, some of which are highly reactive (e.g. Cys530, Figure 4-16) and needed to be removed to prevent protein-protein crosslinking. Because removing all 10 made the protein susceptible to proteolysis (Figure 4-15), I generated a (***b***) Cys-lite background construct from which half of the endogenous cysteines had been removed. (***c***) Using this background, I generated two crosslinking mutants with a single cysteine (Arg614Cys, Gln644Cys) added near the expected path of the template DNA strand. EcoE1(917) structure: PDB 2hnh, model with DNA: generated by Tiago Barros (from PDB 2hnh and 3f2d) and used with permission.



Shared conditions
20 hours @ 25 °C
5.4 mg/mL protein — EcoE1(917)-CF-mpD
(52.5 uM protein)
0.09 mM HEPES pH 7.5
0.18 mM MgCl2
9.70 mM Tris pH 8.5
16.67 mM K glutamate
0.67 mM MnCl2
0.27 mM ddATP+ddGTP
5.0 mM EDTA
3.7 mL reaction volume

DNA
5cPT1-5cPT7
0.9 mg/mL
(54.1 uM)

Figure 4-15. Cysteine crosslinking mutants based on the Cys-free background were prone to proteolysis.

Non-reducing SDS-PAGE. EcoE1(917)-CF-mpD (mpD) is a pool of roughly a dozen cysteine incorporation mutants generated from the Cys-free version of EcoE1(917). Bands below the one corresponding to the unreacted mutants (←arrow) are indicative of protein degradation. If higher bands had been observed, they might have corresponding to crosslinking products. Numbers 1 though 7: different disulfide-containing DNA primer-templates (see Materials & Methods), (–): negative control without DNA.

To overcome construct degradation, I designed a Cys-lite version of EcoE1(917) from which only half of the endogenous cysteine residues were removed (Figure

4-14b). Unlike the EcoE1(917) or the Cys-free construct, the Cys-lite polymerase neither multimerized nor degraded, as verified by non-reducing SDS-PAGE (Figure 4-16), during a 5-day, 4 °C incubation at greater than 40 mg/mL.

Having created a suitable Cys-lite background construct, I designed two crosslinking mutants, each with a single cysteine (Arg614Cys or Gln644Cys, Figure 4-14c) introduced at the top of its Fingers domain. Based on modeling the DNA from a structure of *Geobacillus kaustophilus* PolC into a structure of EcoE1(917) (Figure 4-14c), the introduced cysteines should be close to the path of the single-stranded DNA template as it approaches the polymerase active site. Together with the studies of polymerase monomerization and DNA binding described earlier, these mutants set the stage for covalent crosslinking of the *E. coli* replicative polymerase to primer-template DNA.

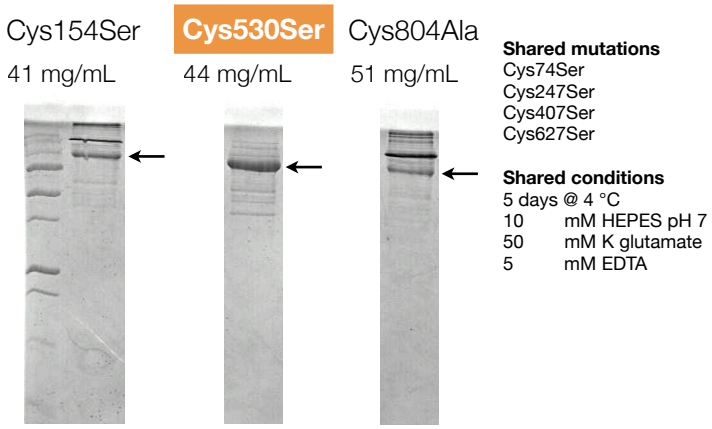


Figure 4-16. Identification of a reactive cysteine residue in EcoE1(917).

Non-reducing SDS-PAGE. As indicated by the absence of significant bands above or the monomer band (←arrow), adding the Cys530Ser mutation to a mutant from which four cysteines had already been removed resulted in a Cys-lite construct that suffered neither protein-protein disulfide crosslinking nor rampant degradation. The Cys154Ser and Cys804Ala constructs both suffered protein-protein crosslinking, as evidenced by multiple dark bands above their respective monomers (←arrow).

4.5 Conclusion

Using standard crystallization techniques, I grew crystals, which reliably diffracted to a resolution of 8 Å, of what was likely a complex of EcoE1(917) plus primer-template DNA. Although this diffraction was not sufficient for structure determination, I have made significant progress toward this aim.

1. I increased the yield of purified EcoE1(917) 100-fold and eliminated protein availability as a limiting factor of solving structures of full-length *E. coli* DnaE1-pol and other constructs based upon it.

2. Jordan Anaya and I prepared two additional C-terminal truncation constructs that, based on previous biochemical data [26], should be able to bind the DNA sliding clamp for tethering of these constructs to DNA.
3. I identified buffer conditions to maintain *E. coli* DnaE1-pol constructs in a monomeric state without sacrificing high DNA binding affinity.
4. I established criteria for distinguishing crystals of protein+DNA complexes from crystals of either protein or DNA alone.
5. For an EcoE1(917)+DNA complex, I established conditions that give rise to crystals that could potentially be improved for structure determination.
6. I designed and experimentally validated a Cys-lite construct of EcoE1(917) for use in covalent trapping of polymerase+DNA complexes.

Taken together, the above advances constitute significant progress toward solving structures of a bacterial replicative polymerase bound to DNA.

REFERENCES

1. Filée, J., et al., *Evolution of DNA polymerase families: evidences for multiple gene exchange between cellular and viral proteins.* J Mol Evol, 2002. **54**(6): p. 763-73.
2. Ito, J. and D.K. Braithwaite, *Compilation and alignment of DNA polymerase sequences.* Nucleic Acids Research, 1991. **19**(15): p. 4045-57.
3. Friedberg, E.C., *Suffering in silence: the tolerance of DNA damage.* Nat Rev Mol Cell Biol, 2005. **6**(12): p. 943-53.
4. Pata, J.D., *Structural diversity of the Y-family DNA polymerases.* Biochim Biophys Acta, 2010. **1804**(5): p. 1124-35.
5. Yamtich, J. and J.B. Sweasy, *DNA polymerase family X: function, structure, and cellular roles.* Biochim Biophys Acta, 2010. **1804**(5): p. 1136-50.
6. Yang, W. and R. Woodgate, *What a difference a decade makes: insights into translesion DNA synthesis.* Proc Natl Acad Sci USA, 2007. **104**(40): p. 15591-8.
7. Yao, N.Y. and M. O'Donnell, *SnapShot: The replisome.* Cell, 2010. **141**(6): p. 1088, 1088.e1.
8. Brautigam, C.A. and T.A. Steitz, *Structural and functional insights provided by crystal structures of DNA polymerases and their substrate complexes.* Curr Opin Struct Biol, 1998. **8**(1): p. 54-63.
9. Berdis, A.J., *Mechanisms of DNA polymerases.* Chem Rev, 2009. **109**(7): p. 2862-79.
10. Pavlov, Y.I. and P.V. Shcherbakova, *DNA polymerases at the eukaryotic fork-20 years later.* Mutat Res, 2010. **685**(1-2): p. 45-53.
11. Bloom, L.B., et al., *Fidelity of Escherichia coli DNA polymerase III holoenzyme. The effects of beta, gamma complex processivity proteins and epsilon proofreading exonuclease on nucleotide misincorporation efficiencies.* J Biol Chem, 1997. **272**(44): p. 27919-30.
12. Watson, J. and F. Crick, *Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid.* Nature, 1953. **171**(4356): p. 737-8.
13. Reyes-Lamothe, R., D.J. Sherratt, and M.C. Leake, *Stoichiometry and architecture of active DNA replication machinery in Escherichia coli.* Science, 2010. **328**(5977): p. 498-501.
14. Kelman, Z. and M. O'Donnell, *DNA polymerase III holoenzyme: structure and function of a chromosomal replicating machine.* Annu Rev Biochem, 1995. **64**: p. 171-200.
15. Ogawa, T. and T. Okazaki, *Discontinuous DNA replication.* Annu Rev Biochem, 1980. **49**: p. 421-57.
16. Okazaki, R., et al., *Mechanism of DNA Replication - Possible Discontinuity of DNA Chain Growth.* Japanese Journal of Medical Science & Biology, 1967. **20**(3): p. 255-&.
17. Krishna, T.S., et al., *Crystal structure of the eukaryotic DNA polymerase processivity factor PCNA.* Cell, 1994. **79**(7): p. 1233-43.
18. Kong, X.P., et al., *Three-dimensional structure of the beta subunit of E. coli DNA polymerase III holoenzyme: a sliding DNA clamp.* Cell, 1992. **69**(3): p. 425-37.
19. Langston, L.D., C. Indiani, and M. O'Donnell, *Whither the replisome: emerging perspectives on the dynamic nature of the DNA replication machinery.* Cell Cycle, 2009. **8**(17): p. 2686-91.
20. Johnson, A. and M. O'Donnell, *Cellular DNA replicases: components and dynamics at the replication fork.* Annu Rev Biochem, 2005. **74**: p. 283-315.
21. Koonin, E.V., *Temporal order of evolution of DNA replication systems inferred by comparison of cellular and viral DNA polymerases*, in *Biol Direct*. 2006. p. 39.
22. Leipe, D.D., L. Aravind, and E.V. Koonin, *Did DNA replication evolve twice independently?* Nucleic Acids Research, 1999. **27**(17): p. 3389-401.
23. Aravind, L. and E.V. Koonin, *DNA polymerase beta-like nucleotidyltransferase superfamily: identification of three new families, classification and evolutionary history.* Nucleic Acids Research, 1999. **27**(7): p. 1609-18.
24. Bailey, S., R.A. Wing, and T.A. Steitz, *The structure of T. aquaticus DNA polymerase III is distinct from eukaryotic replicative DNA polymerases.* Cell, 2006. **126**(5): p. 893-904.

25. Evans, R.J., et al., *Structure of PolC reveals unique DNA binding and fidelity determinants.* Proc Natl Acad Sci USA, 2008. **105**(52): p. 20695-700.

26. Lamers, M.H., et al., *Crystal structure of the catalytic alpha subunit of E. coli replicative DNA polymerase III.* Cell, 2006. **126**(5): p. 881-92.

27. Steitz, T.A., et al., *A unified polymerase mechanism for nonhomologous DNA and RNA polymerases.* Science, 1994. **266**(5193): p. 2022-5.

28. Guiles, J., et al., *Quinazolin-2-ylamino-quinazolin-4-ols as novel non-nucleoside inhibitors of bacterial DNA polymerase III*, in *Bioorg Med Chem Lett*. 2009. p. 800-2.

29. Svenstrup, N., et al., *New DNA polymerase IIIC inhibitors: 3-subtituted anilinouracils with potent antibacterial activity in vitro and in vivo*, in *ChemMedChem*. 2008. p. 1604-15.

30. Lange, R.P., et al., *The targets of currently used antibacterial agents: lessons for drug discovery*, in *Curr Pharm Des*. 2007. p. 3140-54.

31. Rose, Y., et al., *Novel non-nucleobase inhibitors of Staphylococcus aureus DNA polymerase IIIC*, in *Bioorg Med Chem Lett*. 2006. p. 891-6.

32. Tarantino, P.M., et al., *6-Anilinouracil-based inhibitors of Bacillus subtilis DNA polymerase III: antipolymerase and antimicrobial structure-activity relationships based on substitution at uracil N3*, in *J Med Chem*. 1999. p. 2035-40.

33. Kim, D.R., A.E. Pritchard, and C.S. McHenry, *Localization of the active site of the alpha subunit of the Escherichia coli DNA polymerase III holoenzyme.* J Bacteriol, 1997. **179**(21): p. 6721-8.

34. Aravind, L. and E.V. Koonin, *Phosphoesterase domains associated with DNA polymerases of diverse origins.* Nucleic Acids Research, 1998. **26**(16): p. 3746-52.

35. Wing, R., S. Bailey, and T. Steitz, *Insights into the Replisome from the Structure of a Ternary Complex of the DNA Polymerase III alpha-Subunit.* J Mol Biol, 2008.

36. Leulliot, N., et al., *The family X DNA polymerase from deinococcus radiodurans adopts a non-standard extended conformation.* J Biol Chem, 2009.

37. *Conserved Domain Database (CDD), cl11966: NT_Pol-beta-like Superfamily*. Available from: http://www.ncbi.nlm.nih.gov/Structure/cdd/cddsrv.cgi?uid=cl11966.

38. Georgescu, R., et al., *Mechanism of polymerase collision release from sliding clamps on the lagging strand.* EMBO J, 2009.

39. Huang, Y.P. and J. Ito, *DNA polymerase C of the thermophilic bacterium Thermus aquaticus: classification and phylogenetic analysis of the family C DNA polymerases.* J Mol Evol, 1999. **48**(6): p. 756-69.

40. Zhao, X., et al., *GC content variability of eubacteria is governed by the pol III alpha subunit.* Biochem Biophys Res Commun, 2007. **356**(1): p. 20-5.

41. Zhao, X.-Q., J.-F. Hu, and J. Yu, *Comparative analysis of eubacterial DNA polymerase III alpha subunits.* Genomics Proteomics Bioinformatics, 2006. **4**(4): p. 203-11.

42. Sanders, G.M., H.G. Dallmann, and C.S. Mchenry, *Reconstitution of the B. subtilis Replisome with 13 Proteins Including Two Distinct Replicases*, in *Mol Cell*. 2010. p. 273-281.

43. Le Chatelier, E., et al., *Involvement of DnaE, the second replicative DNA polymerase from Bacillus subtilis, in DNA mutagenesis*, in *J Biol Chem*. 2004. p. 1757-67.

44. Bruck, I. and M. O'Donnell, *The DNA replication machine of a gram-positive organism.* J Biol Chem, 2000. **275**(37): p. 28971-83.

45. Bruck, I., R.E. Georgescu, and M. O'Donnell, *Conserved interactions in the Staphylococcus aureus DNA PolC chromosome replication machine.* J Biol Chem, 2005. **280**(18): p. 18152-62.

46. Bruck, I., M.F. Goodman, and M. O'Donnell, *The essential C family DnaE polymerase is error-prone and efficient at lesion bypass*, in *J Biol Chem*. 2003. p. 44361-8.

47. Boshoff, H.I.M., et al., *DnaE2 polymerase contributes to in vivo survival and the emergence of drug resistance in Mycobacterium tuberculosis*, in *Cell*. 2003. p. 183-93.

48. Warner, D.F., et al., *Essential roles for imuA'- and imuB-encoded accessory factors in DnaE2-dependent mutagenesis in Mycobacterium tuberculosis*, in *Proc Natl Acad Sci USA*. 2010. p. 13093-8.

49. Galhardo, R.S., et al., *An SOS-regulated operon involved in damage-inducible mutagenesis in Caulobacter crescentus*, in *Nucleic Acids Research*. 2005. p. 2603-14.
50. Jarvis, T.C., et al., *Reconstitution of a minimal DNA replicase from Pseudomonas aeruginosa and stimulation by non-cognate auxiliary factors.* J Biol Chem, 2005. **280**(9): p. 7890-900.
51. Bruck, I., et al., *Analysis of a multicomponent thermostable DNA polymerase III replicase from an extreme thermophile.* J Biol Chem, 2002. **277**(19): p. 17334-48.
52. Bullard, J.M., et al., *DNA polymerase III holoenzyme from Thermus thermophilus identification, expression, purification of components, and use to reconstitute a processive replicase.* J Biol Chem, 2002. **277**(16): p. 13401-8.
53. Gass, K.B., R.L. Low, and N.R. Cozzarelli, *Inhibition of a DNA polymerase from Bacillus subtilis by hydroxyphenylazopyrimidines.* Proc Natl Acad Sci USA, 1973. **70**(1): p. 103-7.
54. Brown, N.C. and R.E. Handschumacher, *Inhibition of the synthesis of deoxyribonucleic acid in bacteria by 6-(p-hydroxyphenylazo)-2,4-dihydroxypyrimidine. I. Metabolic studies in Streptococcus fecalis.* The Journal of biological chemistry, 1966. **241**(13): p. 3083-9.
55. Roy-Burman, P. and D. Sen, *Effect of a Number of N-Pyrimidyl Amino Acids and of Some of Their 5-Arylazo Derivatives on the Growth of Certain Microorganisms.* Biochem Pharmacol, 1964. **13**: p. 1437-49.
56. Gass, K.B. and N.R. Cozzarelli, *Bacillus subtilis DNA polymerases.* Meth Enzymol, 1974. **29**(0): p. 27-38.
57. Kornberg, T. and M.L. Gefter, *Purification and DNA synthesis in cell-free extracts: properties of DNA polymerase II.* Proc Natl Acad Sci USA, 1971. **68**(4): p. 761-4.
58. Gefter, M.L., et al., *Analysis of DNA polymerases II and III in mutants of Escherichia coli thermosensitive for DNA synthesis.* Proc Natl Acad Sci USA, 1971. **68**(12): p. 3150-3.
59. Pritchard, A.E. and C.S. McHenry, *Identification of the acidic residues in the active site of DNA polymerase III.* Journal of Molecular Biology, 1999. **285**(3): p. 1067-80.
60. Stano, N.M., J. Chen, and C.S. McHenry, *A coproofreading Zn(2+)-dependent exonuclease within a bacterial replicase.* Nat Struct Mol Biol, 2006. **13**(5): p. 458-9.
61. Wieczorek, A. and C.S. Mchenry, *The NH2-terminal php domain of the alpha subunit of the Escherichia coli replicase binds the epsilon proofreading subunit.* J Biol Chem, 2006. **281**(18): p. 12561-7.
62. Taft-Benz, S. and R. Schaaper, *The C-Terminal Domain of DnaQ Contains the Polymerase Binding Site.* J Bacteriol, 1999.
63. Dyall, S.D., M.T. Brown, and P.J. Johnson, *Ancient invasions: from endosymbionts to organelles.* Science, 2004. **304**(5668): p. 253-7.
64. Keniry, M.A., et al., *Structure of the theta subunit of Escherichia coli DNA polymerase III in complex with the epsilon subunit.* J Bacteriol, 2006. **188**(12): p. 4464-73.
65. Taft-Benz, S. and R. Schaaper, *The θ subunit of Escherichia coli DNA polymerase III: a role in stabilizing the epsilon proofreading subunit.* J. Bacteriol, 2004.
66. Omi, R., et al., *Crystal structure of monofunctional histidinol phosphate phosphatase from Thermus thermophilus HB8.* Biochemistry, 2007. **46**(44): p. 12618-27.
67. Otwinowski, Z. and W. Minor, *Processing of X-ray diffraction data collected in oscillation mode.* Method Enzymol, 1997. **276**: p. 307-326.
68. Leslie, A.G.W., *Recent changes to the MOSFLM package for processing film and image plate data.* Joint CCP4 + ESF-EAMCB Newsletter on Protein Crystallography, No. 26, 1992.
69. McCoy, A.J., et al., *Phaser crystallographic software.* J Appl Crystallogr, 2007. **40**(Pt 4): p. 658-674.
70. Afonine, P.V., R.W. Grosse-Kunstleve, and P.D. Adams, CCP4 Newsl. 42, contribution 8., 2005.
71. Emsley, P., et al., *Features and development of Coot.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 4): p. 486-501.
72. Emsley, P. and K. Cowtan, *Coot: model-building tools for molecular graphics.* Acta Crystallogr D Biol Crystallogr, 2004. **60**(Pt 12 Pt 1): p. 2126-32.
73. *The PyMOL Molecular Graphics System, Version 1.3, Schrödinger, LLC.*; Available from: http://pymol.org/.
74. Hsin, K., et al. *Metal Sites in Proteins - MESPEUS database*. 2008; Available from: http://eduliss.bch.ed.ac.uk/MESPEUS/.

75. Hsin, K., et al., *MESPEUS: a database of the geometry of metal sites in proteins.* J Appl Crystallogr, 2008. **41**: p. 963-968.

76. Cromer, D. and D. Liberman, *Anomalous Dispersion Calculations Near to and on the Long-Wavelength Side of an Absorption-Edge.* Acta Crystallogr A, 1981. **37**(MAR): p. 267-268.

77. Cromer, D. and D. Liberman, *Relativistic Calculation of Anomalous Scattering Factors for X-Rays.* J Chem Phys, 1970. **53**(5): p. 1891.

78. Merrett, E.A. *Anomalous Scattering Coefficients*. 1996-2009; Available from: http://skuld.bmsc.washington.edu/scatter/AS_form.html.

79. Bush, L.A., R.W. Nelson, and E. Di Cera, *Murine thrombin lacks Na+ activation but retains high catalytic activity.* J Biol Chem, 2006. **281**(11): p. 7183-8.

80. Di Cera, E., *Thrombin.* Mol Aspects Med, 2008. **29**(4): p. 203-54.

81. Page, M.J. and E. Di Cera, *Role of Na+ and K+ in enzyme function.* Physiol Rev, 2006. **86**(4): p. 1049-92.

82. Oria-Hernández, J., H. Riveros-Rosas, and L. Ramírez-Sílva, *Dichotomic phylogenetic tree of the pyruvate kinase family: K+ -dependent and -independent enzymes.* J Biol Chem, 2006. **281**(41): p. 30717-24.

83. Wang, Z., et al., *Structure of human ferritin L chain.* Acta Crystallogr D Biol Crystallogr, 2006. **62**(Pt 7): p. 800-6.

84. Granier, T., et al., *Structural description of the active sites of mouse L-chain ferritin at 1.2 A resolution.* J Biol Inorg Chem, 2003. **8**(1-2): p. 105-11.

85. Santambrogio, P., et al., *Functional and immunological analysis of recombinant mouse H- and L-ferritins from Escherichia coli.* Protein Expr Purif, 2000. **19**(1): p. 212-8.

86. Yang, X.-L., et al., *Functional and crystal structure analysis of active site adaptations of a potent anti-angiogenic human tRNA synthetase.* Structure, 2007. **15**(7): p. 793-805.

87. Kuratani, M., et al., *Crystal structures of tyrosyl-tRNA synthetases from Archaea.* Journal of Molecular Biology, 2006. **355**(3): p. 395-408.

88. Yang, X.-L., et al., *Crystal structures that suggest late development of genetic code components for differentiating aromatic side chains.* Proc Natl Acad Sci USA, 2003. **100**(26): p. 15376-80.

89. Austin, J. and E.A. First, *Potassium functionally replaces the second lysine of the KMSKS signature sequence in human tyrosyl-tRNA synthetase.* J Biol Chem, 2002. **277**(23): p. 20243-8.

90. Linderoth, L., et al., *Mechanistic study of the sPLA2-mediated hydrolysis of a thio-ester pro anticancer ether lipid.* J Am Chem Soc, 2009. **131**(34): p. 12193-200.

91. Murakami, M.T., et al., *Crystal structure of a novel myotoxic Arg49 phospholipase A2 homolog (zhaoermiatoxin) from Zhaoermia mangshanensis snake venom: insights into Arg49 coordination and the role of Lys122 in the polarization of the C-terminus.* Toxicon, 2008. **51**(5): p. 723-35.

92. Murakami, M.T., et al., *Structure of myotoxin II, a catalytically inactive Lys49 phospholipase A2 homologue from Atropoides nummifer venom.* Acta Crystallogr Sect F Struct Biol Cryst Commun, 2006. **62**(Pt 5): p. 423-6.

93. Yamazaki, Y., et al., *Identification of vascular endothelial growth factor receptor-binding protein in the venom of eastern cottonmouth. A new role of snake venom myotoxic Lys49-phospholipase A2.* J Biol Chem, 2005. **280**(34): p. 29989-92.

94. Apiyo, D., et al., *X-ray structure of the R69D phosphatidylinositol-specific phospholipase C enzyme: insight into the role of calcium and surrounding amino acids in active site geometry and catalysis.* Biochemistry, 2005. **44**(30): p. 9980-9.

95. de Araujo, H.S., S.P. White, and C.L. Ownby, *Sequence analysis of Lys49 phospholipase A2 myotoxins: a highly conserved class of proteins.* Toxicon, 1996. **34**(11-12): p. 1237-42.

96. Takagi, H., et al., *Crystal structure of the ribonuclease P protein Ph1877p from hyperthermophilic archaeon Pyrococcus horikoshii OT3.* Biochem Biophys Res Commun, 2004. **319**(3): p. 787-94.

97. Baranovskiy, A.G., et al., *X-ray structure of the complex of regulatory subunits of human DNA polymerase delta.* Cell Cycle, 2008. **7**(19): p. 3026-36.

98. Wardle, J., et al., *Uracil recognition by replicative DNA polymerases is limited to the archaea, not occurring with bacteria and eukarya*, in *Nucleic Acids Research*. 2008. p. 705-11.

99. Tahirov, T.H., et al., *Evolution of DNA polymerases: an inactivated polymerase-exonuclease module in Pol epsilon and a chimeric origin of eukaryotic polymerases from two classes of archaeal ancestors.* Biol Direct, 2009. **4**: p. 11.

100. Rogozin, I., et al., *A highly conserved family of inactivated archaeal B family DNA polymerases*, in *Biol Direct*. 2008. p. 32.

101. Lockless, S.W. and R. Ranganathan, *Evolutionarily conserved pathways of energetic connectivity in protein families.* Science, 1999. **286**(5438): p. 295-9.

102. Wu, D., et al., *A phylogeny-driven genomic encyclopaedia of Bacteria and Archaea*, in *Nature*. 2009. p. 1056-60.

103. Hugenholtz, P., B.M. Goebel, and N.R. Pace, *Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity.* J Bacteriol, 1998. **180**(18): p. 4765-74.

104. Garcia-Diaz, M., et al., *Role of the catalytic metal during polymerization by DNA polymerase lambda.* DNA Repair (Amst), 2007. **6**(9): p. 1333-40.

105. Batra, V.K., et al., *Magnesium-induced assembly of a complete DNA polymerase catalytic complex.* Structure (London, England : 1993), 2006. **14**(4): p. 757-66.

106. Sawaya, M.R., et al., *Crystal structures of human DNA polymerase beta complexed with gapped and nicked DNA: evidence for an induced fit mechanism.* Biochemistry, 1997. **36**(37): p. 11205-15.

107. Studwell, P.S. and M. O'Donnell, *Processive replication is contingent on the exonuclease subunit of DNA polymerase III holoenzyme.* J Biol Chem, 1990. **265**(2): p. 1171-8.

108. Altschul, S.F., et al., *Basic local alignment search tool.* Journal of Molecular Biology, 1990. **215**(3): p. 403-10.

109. *Max-Planck Institute for Developmental Biology, Bioinformatics Toolkit*. 2008; Available from: http://toolkit.tuebingen.mpg.de/.

110. Felsenstein, J. *PHYLIP (Phylogeny Inference Package) version 3.6.* 2005; Available from: http://evolution.genetics.washington.edu/phylip.html, http://toolkit.tuebingen.mpg.de/phylip.

111. Edgar, R.C., *MUSCLE: multiple sequence alignment with high accuracy and high throughput.* Nucleic Acids Res, 2004. **32**(5): p. 1792-7.

112. Edgar, R.C., *MUSCLE: a multiple sequence alignment method with reduced time and space complexity.* BMC Bioinformatics, 2004. **5**: p. 113.

113. Edgar, R.C. *MUSCLE: MUltiple Sequence Comparison by Log-Expectation for phylogenetic tree estimation, secondary structure prediction and critical residue identification.*; Available from: http://www.ebi.ac.uk/Tools/muscle/index.html.

114. Katoh, K. and H. Toh, *Recent developments in the MAFFT multiple sequence alignment program.* Brief Bioinformatics, 2008. **9**(4): p. 286-98.

115. Katoh, K., et al., *MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform.* Nucleic Acids Research, 2002. **30**(14): p. 3059-66.

116. Katoh, K. *MAFFT: Multiple alignment program for amino acid or nucleotide sequences*. Available from: http://mafft.cbrc.jp/alignment/server/index.html.

117. Waterhouse, A.M., et al., *Jalview Version 2--a multiple sequence alignment editor and analysis workbench.* Bioinformatics, 2009. **25**(9): p. 1189-91.

118. Clamp, M., et al., *The Jalview Java alignment editor.* Bioinformatics, 2004. **20**(3): p. 426-7.

119. Katoh, K. *Molecular phylogeny by the NJ / UPG methods*. Available from: http://mafft.cbrc.jp/alignment/server/phylogeny.html.

120. Saitou, N. and M. Nei, *The neighbor-joining method: a new method for reconstructing phylogenetic trees.* Mol Biol Evol, 1987. **4**(4): p. 406-25.

121. Whelan, S. and N. Goldman, *A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach.* Mol Biol Evol, 2001. **18**(5): p. 691-9.

122. Rambaut, A. *FigTree Tree Figure Drawing Tool*. Available from: http://tree.bio.ed.ac.uk/software/figtree/.

123. Vacic, V., L.M. Iakoucheva, and P. Radivojac, *Two Sample Logo: a graphical representation of the differences between two sets of sequence alignments.* Bioinformatics, 2006. **22**(12): p. 1536-7.

124. Vacic, V., L.M. Iakoucheva, and P. Radivojac. *Two Sample Logos: web-based application*. Available from: http://www.twosamplelogo.org/.

125. Süel, G.M., et al., *Evolutionarily conserved networks of residues mediate allosteric communication in proteins.* Nat Struct Biol, 2003. **10**(1): p. 59-69.

126. Hatley, M.E., et al., *Allosteric determinants in guanine nucleotide-binding proteins.* Proc Natl Acad Sci USA, 2003. **100**(24): p. 14445-50.

127. Yip, K.Y., et al., *An integrated system for studying residue coevolution in proteins.* Bioinformatics, 2008. **24**(2): p. 290-2.

128. Gerstein, M.B. *Coevolution analysis of protein residues*. Available from: http://coevolution.gersteinlab.org/coevolution/.

129. Kunz, P.F. *HippoDraw*. Available from: http://www.slac.stanford.edu/grp/ek/hippodraw/.

130. Lee, B.-C. and D. Kim, *A new method for revealing correlated mutations under the structural and functional constraints in proteins.* Bioinformatics, 2009. **25**(19): p. 2506-13.

131. Skerker, J.M., et al., *Rewiring the specificity of two-component signal transduction systems.* Cell, 2008. **133**(6): p. 1043-54.

132. Szurmant, H., et al., *Co-evolving motions at protein-protein interfaces of two-component signaling systems identified by covariance analysis.* Biochemistry, 2008. **47**(30): p. 7782-4.

133. Weigt, M., et al., *Identification of direct residue contacts in protein-protein interaction by message passing.* Proc Natl Acad Sci USA, 2009. **106**(1): p. 67-72.

134. Fiser, A. and A. Sali, *MODELLER: Generation and refinement of homology-based protein structure models*, in *Macromolecular Crystallography, Pt D*. 2003, Academic Press Inc: San Diego. p. 461-+.

135. Sali, A. and T.L. Blundell, *Comparative protein modelling by satisfaction of spatial restraints.* Journal of Molecular Biology, 1993. **234**(3): p. 779-815.

136. Krebs, W.G. and M. Gerstein, *The morph server: a standardized system for analyzing and visualizing macromolecular motions in a database framework.* Nucleic Acids Research, 2000. **28**(8): p. 1665-75.

137. Dallmann, H.G., et al., *Parallel multiplicative target screening against divergent bacterial replicases: identification of specific inhibitors with broad spectrum potential.* Biochemistry, 2010. **49**(11): p. 2551-62.

138. Seville, M., et al., *Fluorometric assay for DNA polymerases and reverse transcriptase.* BioTechniques, 1996. **21**(4): p. 664, 666, 668, 670, 672.

139. Tveit, H. and T. Kristensen, *Fluorescence-based DNA polymerase assay.* Anal Biochem, 2001. **289**(1): p. 96-8.

140. Royer, C.A., C.J. Mann, and C.R. Matthews, *Resolution of the fluorescence equilibrium unfolding profile of trp aporepressor using single tryptophan mutants*, in *Protein Sci*. 1993. p. 1844-52.

141. Clarke, J. and A.R. Fersht, *Engineered disulfide bonds as probes of the folding pathway of barnase: increasing the stability of proteins against the rate of denaturation*, in *Biochemistry*. 1993. p. 4322-9.

142. *GenInfo Identifier (GI) definition*. Available from: http://www.ncbi.nlm.nih.gov/Sitemap/samplerecord.html#GInB.

143. NCBI. *Proteins: Translations of Life*. Available from: http://www.ncbi.nlm.nih.gov/protein.

144. Brown, J.A., et al., *Identification of Critical Residues for the Tight Binding of Both Correct and Incorrect Nucleotides to Human DNA Polymerase lambda.* Journal of Molecular Biology, 2010.

145. Foley, M.C. and T. Schlick, *Simulations of DNA pol lambda R517 mutants indicate 517's crucial role in ternary complex stability and suggest DNA slippage origin.* Journal of the American Chemical Society, 2008. **130**(12): p. 3967-77.

146. Hamdan, S., et al., *Structural basis for proofreading during replication of the Escherichia coli chromosome.* Structure, 2002. **10**(4): p. 535-46.

147. Ozawa, K., et al., *The proofreading exonuclease subunit epsilon of Escherichia coli DNA polymerase III is tethered to the polymerase subunit alpha via a flexible linker.* Nucleic Acids Research, 2008. **36**(15): p. 5074-82.

148. Marchler-Bauer, A., et al., *CDD: specific functional annotation with the Conserved Domain Database.* Nucleic Acids Research, 2009. **37**: p. D205-D210.

149. *Conserved Domain Database (CDD), cd06127: DEDDh*. Available from: http://www.ncbi.nlm.nih.gov/Structure/cdd/cddsrv.cgi?uid=176648.

150. Lehtinen, D.A. and F.W. Perrino, *Dysfunctional proofreading in the Escherichia coli DNA polymerase III core.* Biochem J, 2004. **384**(Pt 2): p. 337-48.

151. Taft-Benz, S.A. and R.M. Schaaper, *Mutational analysis of the 3'-->5' proofreading exonuclease of Escherichia coli DNA polymerase III.* Nucleic Acids Research, 1998. **26**(17): p. 4005-11.

152. Söding, J., A. Biegert, and A.N. Lupas, *The HHpred interactive server for protein homology detection and structure prediction.* Nucleic Acids Research, 2005. **33**(Web Server issue): p. W244-8.

153. Söding, J., *Protein homology detection by HMM-HMM comparison.* Bioinformatics, 2005. **21**(7): p. 951-60.

154. Haft, D.H., et al., *TIGRFAMs: a protein family resource for the functional identification of proteins.* Nucleic Acids Res, 2001. **29**(1): p. 41-3.

155. Institute, J.C.V., *TIRGFAMs.*

156. *JCVI Comprehensive Microbial Resource, TIGR01406*. Available from: http://cmr.jcvi.org/cgi-bin/CMR/HmmReport.cgi?hmm_acc=TIGR01406.

157. NCBI. *Conserved Domains*. Available from: http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi.

158. *Conserved Domain Database (CDD), TIGR01406: dnaQ_proteo*. Available from: http://www.ncbi.nlm.nih.gov/Structure/cdd/cddsrv.cgi?uid=TIGR01406.

159. *Conserved Domain Database (CDD), cd06130: DNA_pol_III_epsilon_like*. Available from: http://www.ncbi.nlm.nih.gov/Structure/cdd/cddsrv.cgi?uid=cd06130.

160. *Conserved Domain Database (CDD), cd06131: DNA_pol_III_epsilon_Ecoli_like*. Available from: http://www.ncbi.nlm.nih.gov/Structure/cdd/cddsrv.cgi?uid=cd06131.

161. Nisbet, E.G. and N.H. Sleep, *The habitat and nature of early life.* Nature, 2001. **409**(6823): p. 1083-91.

162. Schopf, J.W., *Microfossils of the Early Archean Apex chert: new evidence of the antiquity of life.* Science, 1993. **260**: p. 640-6.

163. Byerly, G., D. LOWER, and M. WALSH, *Stromatolites from the 3,300-3,500-Myr Swaziland Supergroup, Barberton Mountain Land, South-Africa.* Nature, 1986. **319**(6053): p. 489-491.

164. Walter, M., R. Buick, and J. Dunlop, *Stromatolites 3,400-3,500 Myr Old from the North-Pole Area, Western-Australia.* Nature, 1980. **284**(5755): p. 443-445.

165. Lowe, D.R., *Stromatolites 3,400-Myr Old from the Archean of Western-Australia.* Nature, 1980. **284**(5755): p. 441-443.

166. Esser, C., et al., *A genome phylogeny for mitochondria among alpha-proteobacteria and a predominantly eubacterial ancestry of yeast nuclear genes*, in *Mol Biol Evol*. 2004. p. 1643-60.

167. Andersson, S.G.E., et al., *On the origin of mitochondria: a genomics perspective*, in *Philos Trans R Soc Lond, B, Biol Sci*. 2003. p. 165-77; discussion 177-9.

168. Gray, M.W., G. Burger, and B.F. Lang, *Mitochondrial evolution*, in *Science*. 1999. p. 1476-81.

169. Andersson, S.G., et al., *The genome sequence of Rickettsia prowazekii and the origin of mitochondria*, in *Nature*. 1998. p. 133-40.

170. Javaux, E.J., A.H. Knoll, and M.R. Walter, *Morphological and ecological complexity in early eukaryotic ecosystems*, in *Nature*. 2001. p. 66-9.

171. Embley, T.M. and W. Martin, *Eukaryotic evolution, changes and challenges.* Nature, 2006. **440**(7084): p. 623-30.

172. Butterfield, N., *Bangiomorpha pubescens n. gen., n. sp.: implications for the evolution of sex, multicellularity, and the Mesoproterozoic/Neoproterozoic radiation of eukaryotes.* Paleobiology, 2000. **26**(3): p. 386-404.

173. Goodey, N.M. and S.J. Benkovic, *Allosteric regulation and catalysis emerge via a common route.* Nat Chem Biol, 2008. **4**(8): p. 474-82.

174. Lee, S.-Y., A. Banerjee, and R. MacKinnon, *Two separate interfaces between the voltage sensor and pore are required for the function of voltage-dependent K(+) channels.* PLoS Biol, 2009. **7**(3): p. e47.

175. Lee, J., et al., *Surface sites for engineering allosteric control in proteins.* Science, 2008. **322**(5900): p. 438-42.

176. Ferguson, A.D., et al., *Signal transduction pathway of TonB-dependent transporters.* Proc Natl Acad Sci USA, 2007. **104**(2): p. 513-8.

177. Peterson, F.C., et al., *Cdc42 regulates the Par-6 PDZ domain through an allosteric CRIB-PDZ transition.* Molecular Cell, 2004. **13**(5): p. 665-76.

178. Kass, I. and A. Horovitz, *Mapping pathways of allosteric communication in GroEL by analysis of correlated mutations.* Proteins, 2002. **48**(4): p. 611-7.

179. Datta, K., N.P. Johnson, and P.H. von Hippel, *DNA conformational changes at the primer-template junction regulate the fidelity of replication by DNA polymerase*, in *Proc Natl Acad Sci USA*. 2010.

180. Li, Y., S. Korolev, and G. Waksman, *Crystal structures of open and closed forms of binary and ternary complexes of the large fragment of Thermus aquaticus DNA polymerase I: structural basis for nucleotide incorporation.* EMBO J, 1998. **17**(24): p. 7514-25.

181. Beese, L.S. and T.A. Steitz, *Structural basis for the 3'-5' exonuclease activity of Escherichia coli DNA polymerase I: a two metal ion mechanism.* EMBO J, 1991. **10**(1): p. 25-33.

182. Kirby, T.W., et al., *Structure of the Escherichia coli DNA polymerase III epsilon-HOT proofreading complex.* J Biol Chem, 2006. **281**(50): p. 38466-71.

183. Georgescu, R., et al., *Structure of a small-molecule inhibitor of a DNA polymerase sliding clamp.* Proc Natl Acad Sci USA, 2008.

184. Dalrymple, B.P., et al., *A universal protein-protein interaction motif in the eubacterial DNA replication and repair systems.* Proc Natl Acad Sci USA, 2001. **98**(20): p. 11627-32.

185. Bunting, K.A., S.M. Roe, and L.H. Pearl, *Structural basis for recruitment of translesion DNA polymerase Pol IV/DinB to the beta-clamp.* EMBO J, 2003. **22**(21): p. 5883-92.

186. Stukenberg, P.T., P.S. Studwell-Vaughan, and M. O'Donnell, *Mechanism of the sliding beta-clamp of DNA polymerase III holoenzyme.* J Biol Chem, 1991. **266**(17): p. 11328-34.

187. Anfinsen, C.B., *Principles that govern the folding of protein chains.* Science, 1973. **181**(96): p. 223-30.

188. Lo, M.-C., et al., *Evaluation of fluorescence-based thermal shift assays for hit identification in drug discovery.* Anal Biochem, 2004. **332**(1): p. 153-9.

189. Pantoliano, M.W., et al., *High-density miniaturized thermal shift assays as a general strategy for drug discovery.* J Biomol Screen, 2001. **6**(6): p. 429-40.

190. Maret, W. and Y. Li, *Coordination dynamics of zinc in proteins.* Chem Rev, 2009. **109**(10): p. 4682-707.

191. Dudev, T. and C. Lim, *Metal binding affinity and selectivity in metalloproteins: insights from computational studies*, in *Annual review of biophysics*. 2008. p. 97-116.

192. Dudev, T. and C. Lim, *Principles governing Mg, Ca, and Zn binding and selectivity in proteins.* Chem Rev, 2003. **103**(3): p. 773-88.

193. Lipscomb, W.N. and N. Sträter, *Recent Advances in Zinc Enzymology.* Chem Rev, 1996. **96**(7): p. 2375-2434.

194. Glusker, J.P., *Structural aspects of metal liganding to functional groups in proteins.* Adv Protein Chem, 1991. **42**: p. 1-76.

195. Maret, W., *Zinc and sulfur: a critical biological partnership.* Biochemistry, 2004. **43**(12): p. 3301-9.

196. Bryngelson, J.D., et al., *Funnels, pathways, and the energy landscape of protein folding: a synthesis.* Proteins, 1995. **21**(3): p. 167-95.

197. Watters, A.L., et al., *The highly cooperative folding of small naturally occurring proteins is likely the result of natural selection.* Cell, 2007. **128**(3): p. 613-24.

198. Krantz, B.A., et al., *Fast and slow intermediate accumulation and the initial barrier mechanism in protein folding.* Journal of Molecular Biology, 2002. **324**(2): p. 359-71.

199. Krantz, B.A. and T.R. Sosnick, *Distinguishing between two-state and three-state models for ubiquitin folding.* Biochemistry, 2000. **39**(38): p. 11696-701.

200. Jackson, S.E., *How do small single-domain proteins fold?* Fold Des, 1998. **3**(4): p. R81-91.

201. Jackson, S.E. and A.R. Fersht, *Folding of chymotrypsin inhibitor 2. 1. Evidence for a two-state transition.* Biochemistry, 1991. **30**(43): p. 10428-35.

202. Fitter, J., *The perspectives of studying multi-domain protein folding.* Cell. Mol. Life Sci., 2009. **66**(10): p. 1672-81.

203. Strucksberg, K.H., T. Rosenkranz, and J. Fitter, *Reversible and irreversible unfolding of multi-domain proteins.* Biochim Biophys Acta, 2007. **1774**(12): p. 1591-603.

204. Feller, G., D. d'Amico, and C. Gerday, *Thermodynamic stability of a cold-active alpha-amylase from the Antarctic bacterium Alteromonas haloplanctis.* Biochemistry, 1999. **38**(14): p. 4613-9.

205. Aslanidis, C. and P.J. de Jong, *Ligation-independent cloning of PCR products (LIC-PCR).* Nucleic Acids Res, 1990. **18**(20): p. 6069-74.

206. Corn, J.E. and J.M. Berger, *FASTDXL: a generalized screen to trap disulfide-stabilized complexes for use in structural studies.* Structure, 2007. **15**(7): p. 773-80.

207. Deredge, D.J., et al., *The Glutamate Effect on DNA Binding by Pol I DNA Polymerases: Osmotic Stress and the Effective Reversal of Salt Linkage.* Journal of Molecular Biology, 2010.

208. Lundback, T. and T. Hard, *Salt dependence of the free energy, enthalpy, and entropy of nonsequence specific DNA binding.* J Phys Chem-US, 1996. **100**(44): p. 17690-17695.

209. Ha, J.H., et al., *Thermodynamic stoichiometries of participation of water, cations and anions in specific and non-specific binding of lac repressor to DNA. Possible thermodynamic origins of the "glutamate effect" on protein-DNA interactions.* Journal of Molecular Biology, 1992. **228**(1): p. 252-64.

210. Griep, M.A. and C.S. McHenry, *Glutamate overcomes the salt inhibition of DNA polymerase III holoenzyme.* J Biol Chem, 1989. **264**(19): p. 11294-301.

211. Dohrmann, P.R. and C.S. McHenry, *A bipartite polymerase-processivity factor interaction: only the internal beta binding site of the alpha subunit is required for processive replication by the DNA polymerase III holoenzyme.* J Mol Biol, 2005. **350**(2): p. 228-39.

212. McCauley, M., et al., *Distinct Double- and Single-Stranded DNA Binding of E. coli Replicative DNA Polymerase III alpha Subunit.* ACS Chem Biol, 2008.

213. Georgescu, R., et al., *Structure of a sliding clamp on DNA*, in *Cell*. 2008. p. 43-54.

214. Banerjee, A., W.L. Santos, and G.L. Verdine, *Structure of a DNA glycosylase searching for lesions.* Science, 2006. **311**(5764): p. 1153-7.

215. Banerjee, A. and G.L. Verdine, *A nucleobase lesion remodels the interaction of its normal neighbor in a DNA glycosylase complex.* Proc Natl Acad Sci USA, 2006. **103**(41): p. 15020-5.

216. Corn, J.E., J.G. Pelton, and J.M. Berger, *Identification of a DNA primase template tracking site redefines the geometry of primer synthesis.* Nat Struct Mol Biol, 2008. **15**(2): p. 163-9.

217. Fromme, J.C., et al., *Structural basis for removal of adenine mispaired with 8-oxoguanine by MutY adenine DNA glycosylase.* Nature, 2004. **427**(6975): p. 652-6.

218. He, C. and G.L. Verdine, *Trapping distinct structural states of a protein/DNA interaction through disulfide crosslinking.* Chemistry & Biology, 2002. **9**(12): p. 1297-303.

219. Huang, H., et al., *Structure of a covalently trapped catalytic complex of HIV-1 reverse transcriptase: implications for drug resistance.* Science, 1998. **282**(5394): p. 1669-75.

220. Qi, Y., et al., *Entrapment and structure of an extrahelical guanine attempting to enter the active site of a bacterial DNA glycosylase, MutM.* The Journal of biological chemistry, 2010. **285**(2): p. 1468-78.

221. Sarafianos, S.G., et al., *Trapping HIV-1 reverse transcriptase before and after translocation on DNA.* The Journal of biological chemistry, 2003. **278**(18): p. 16280-8.

222. Verdine, G.L. and D.P.G. Norman, *Covalent trapping of protein-DNA complexes.* Annu Rev Biochem, 2003. **72**: p. 337-66.

223. Lopachin, R.M., D.S. Barber, and T. Gavin, *Molecular mechanisms of the conjugated alpha,beta-unsaturated carbonyl derivatives: relevance to neurotoxicity and neurodegenerative diseases*, in *Toxicol Sci*. 2008. p. 235-49.

224. Lopachin, R.M. and T. Gavin, *Acrylamide-induced nerve terminal damage: relevance to neurotoxic and neurodegenerative mechanisms*, in *J Agric Food Chem*. 2008. p. 5994-6003.

225. Thermo Scientific. *Sulfo-SMCC*. Available from: http://www.piercenet.com/products/browse.cfm?fldID=02030378.